Molecular Ecology (2008) **17**, 1685–1701 doi: 10.1111/j.1365-294X.2008.03705.x

Congruent population structure inferred from dispersal behaviour and intensive genetic surveys of the threatened Florida scrub-jay (*Aphelocoma cœrulescens*)

A. COULON,* J. W. FITZPATRICK,* R. BOWMAN,† B. M. STITH,‡ C. A. MAKAREWICH,* L. M. STENZLER* and I. J. LOVETTE*

**Laboratory of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850, USA, †Archbold Biological Station, PO Box 2057, Lake Placid, Florida 33862, USA, ‡IAP World Services, Inc., US Geological Survey, 2201 NW 40th Terrace, Gainesville, Florida 32601, USA*

Abstract

The delimitation of populations, defined as groups of individuals linked by gene flow, is possible by the analysis of genetic markers and also by spatial models based on dispersal probabilities across a landscape. We combined these two complimentary methods to define the spatial pattern of genetic structure among remaining populations of the threatened Florida scrub-jay, a species for which dispersal ability is unusually well-characterized. The range-wide population was intensively censused in the 1990s, and a metapopulation model defined population boundaries based on predicted dispersal-mediated demographic connectivity. We subjected genotypes from more than 1000 individual jays screened at 20 microsatellite loci to two Bayesian clustering methods. We describe a consensus method for identifying common features across many replicated clustering runs. Ten genetically differentiated groups exist across the present-day range of the Florida scrub-jay. These groups are largely consistent with the dispersal-defined metapopulations, which assume very limited dispersal ability. Some genetic groups comprise more than one metapopulation, likely because these genetically similar metapopulations were sundered only recently by habitat alteration. The combined reconstructions of population structure based on genetics and dispersal-mediated demographic connectivity provide a robust depiction of the current genetic and demographic organization of this species, reflecting past and present levels of dispersal among occupied habitat patches. The differentiation of populations into 10 genetic groups adds urgency to management efforts aimed at preserving what remains of genetic variation in this dwindling species, by maintaining viable populations of all genetically differentiated and geographically isolated populations.

Keywords: *Aphelocoma coerulescens*, Bayesian clustering, demographic units, dispersal distance, genetic structure, metapopulations

Received 7 September 2007; revision received 11 January 2008; accepted 11 January 2008

Introduction

In their review of the population concept, Waples & Gaggiotti (2006) showed that most definitions can be associated with one of two paradigms: the ecological paradigm, where the contacts between individuals are mainly due to demographic events; and the evolutionary paradigm, where the cohesive forces are mainly genetic. In the evolutionary paradigm, the criterion to define populations is the amount of gene flow — or effective dispersal — among locations. Following this approach, it is possible to define populations by estimating gene flow through the analysis of genetic markers or from field data on dispersal. In practice, however, genetic clustering often defies expectations based on dispersal biology or present-day geographical patterns (but see Berry *et al*. (2004), Vandewoestijne & Baguette (2004)). Molecular studies often reveal levels of gene flow exceeding those inferred from

Correspondence: Aurélie Coulon, Fuller Evolutionary Biology Program, Laboratory of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850, USA. Fax: +1 607 254 2486; E-mail: ac462@cornell.edu

field data, suggesting either elevated past gene flow or rare long-distance dispersal events (Crochet 1996; Koenig *et al*. 1996; but see Thompson & Goodman 1997; Nathan 2001). The combination of the two methods and the analysis of the discordances or concordances in their results may actually provide complimentary information (about, for example, the existence of long-distance dispersal events that are difficult to infer from field data, or on temporal changes in the levels of gene flow). Here, we used these two complimentary approaches to provide a comprehensive picture of the spatial organization of the endangered Florida scrub-jay (*Aphelocoma coerulescens*) over its entire current range.

The Florida scrub-jay (FSJ hereafter) is a nonmigratory bird species endemic to Florida and restricted to early successional, fire-maintained xeric oak scrub (Woolfenden & Fitzpatrick 1984). This unique, biologically diverse scrub habitat was widespread in the peninsula in the late Pleistocene, but became fragmented and reduced in area as humid climate prevailed during the Holocene (Myers & Ewel 1990). Over the past century, the rate of habitat loss and fragmentation of Florida's oak scrub increased by several orders of magnitude because of anthropogenic land-use conversion (especially to citrus plantations and housing developments) and suppression of the lightning-caused fires necessary for its persistence (Fernald 1989; Myers & Ewel 1990; Fitzpatrick *et al*. 1991; Woolfenden & Fitzpatrick 1996). As a result, over the last several decades FSJ populations have declined precipitously, or disappeared altogether, throughout much of the remaining range of the species. The FSJ was classified as a 'Threatened' species by the Florida Game and Fresh Water Fish Commission in 1975 and by the US Fish and Wildlife Service in 1987. By 1993, Florida scrub-jay populations had declined to an estimated 3% of their original number, to about 10 000 individuals (Pranty 1996). Since then, further severe declines have been documented in several populations, including those in Brevard (Breininger *et al*. 2003), Charlotte (Miller & Stith 2002), and Sarasota counties (Fig. 1). Facing these ongoing dramatic declines, robust information on the population structure of the species is required to enable agencies and regulators to prioritize conservation efforts.

Florida scrub-jays are permanently territorial, cooperatively breeding birds. Each territory (mean size = 10 ha) is defended by a family group consisting of one breeding pair and zero to six (mean = 1) helpers that contribute to defending the territory, as well as to feeding nestlings and fledglings (Woolfenden & Fitzpatrick 1984). Because some populations have been studied intensively for decades (e.g. Woolfenden & Fitzpatrick 1984, 1996; Breininger *et al*. 1995; Fitzpatrick *et al*. 1999), the species' dispersal biology is unusually well characterized. Individuals usually remain on their natal territory for at least 1 year before dispersing. Dispersal distances are mostly short, with a strongly leptokurtic frequency distribution and a modal dispersal distance of only one territory for both sexes. Dispersal events farther than five territories are notably rare. Females tend to disperse farther than males (Fitzpatrick *et al*. 1999): the maximum documented dispersal distance is 60 km for females vs. 13 km for males (unpublished data). The overall rate of dispersal off the natal territory is substantially higher for females (94%) than for males (44%) (Woolfenden & Fitzpatrick 1984). Once established as breeders, FSJs are essentially sedentary (Woolfenden & Fitzpatrick 1984).

An intensive survey of FSJs in 1992–1993 (Stith *et al*. 1996) yielded a fine-scale map of the species' distribution and estimates of its local population sizes in most of the then-occupied habitat patches throughout its range. These census data, together with field-documented dispersal data, were used to generate a map of metapopulations range-wide, in which a metapopulation was defined as a group of territories among which dispersal likely occurs, but which is probably separated from other such territory clusters by habitat gaps that impede or prevent dispersal. A 12-km isolation threshold was inferred from the landscape distribution of occupied territories: occupancy of suitable habitat patches drops to near zero when distance to the nearest occupied patch exceeds 12 km, suggesting that FSJ dispersal beyond this distance is rare and demographically unimportant. Assuming a 12-km dispersal limit across habitat gaps, Stith *et al*. (1996) defined current metapopulations by creating 12-km buffers around all extant jay territories, thereby clustering all occupied remnant habitat patches assumed to function as demographically connected units. Buffers were further modified to incorporate known hard barriers to dispersal (such as open water with forested margins). This process delineated 42 FSJ metapopulations, most of which were small and isolated.

Strong philopatry, short dispersal distance, natural patchiness of scrub habitat, and further isolation of most jay populations by recent habitat loss all suggest that the FSJ should show substantial genetic structuring. A preliminary study based on 10 microsatellite loci and 11 populations yielded a G_{ST} of 0.048 (McDonald *et al*. 1999), which was much greater than that observed for the widespread, congeneric western scrub-jay, supporting the prediction of strong genetic structuring in FSJs. To examine the question more comprehensively, we genotyped over 1000 individuals at 20 microsatellite loci, with populations sampled across the species' entire geographical range. We employ two Bayesian clustering methodologies: the widely used STRUCTURE (Pritchard *et al*. 2000; Falush *et al*. 2003), and the more recently developed GENELAND (Guillot *et al*. 2005b), which makes use of the information contained in the spatial locations of the individuals. We compare the inferred clusters from the Bayesian clustering analyses (hereafter 'genetic groups') with the metapopulations defined by

Stith *et al*. (1996). The general congruence we find between these two approaches supports the existence of genuine genetic boundaries among dispersal-defined metapopulations. This result has important implications for the conservation of this endangered species, since it provides information useful for prioritizing conservation efforts as most local FSJ populations rapidly decline toward extinction.

Materials and methods

Study area and sampling design

We sampled Florida scrub-jays across their entire current range (Fig. 1). We obtained genetic material from at least some FSJs from 21 of the 42 metapopulations defined by Stith *et al*. (1996). Nongenetically sampled metapopulations were generally small (one to seven pairs) populations found in isolated remnant scrub patches in the early 1990s, many of which are now extinct. From some demographically large or geographically broad populations, we obtained samples from multiple sites within the metapopulation, whereas at other sites our sampling was limited by the number of FSJs remaining in the metapopulation (Fig. 1). Sampling was most intensive in metapopulation 21, which included the FSJ population at Archbold Biological Station and surrounding areas of the Lake Wales Ridge, where jays have been intensively monitored for more than 30 years. Sample collection was undertaken between 1995 and 2001 (13% of the samples) and between 2003 and 2005 (87%). The first sampling period involved only the Lake Wales Ridge area, where many individuals were also sampled between 2003 and 2005. Genetic samples were obtained

Meta-population number	Sample size
1	16
2	38
3	19
4	48
5	47
6	61
7	36
8	16
9	9
10	47
11	29
12	54
13	48
14	41
15	38
16	3
17	48
18	19
19	80
20	70
21	261

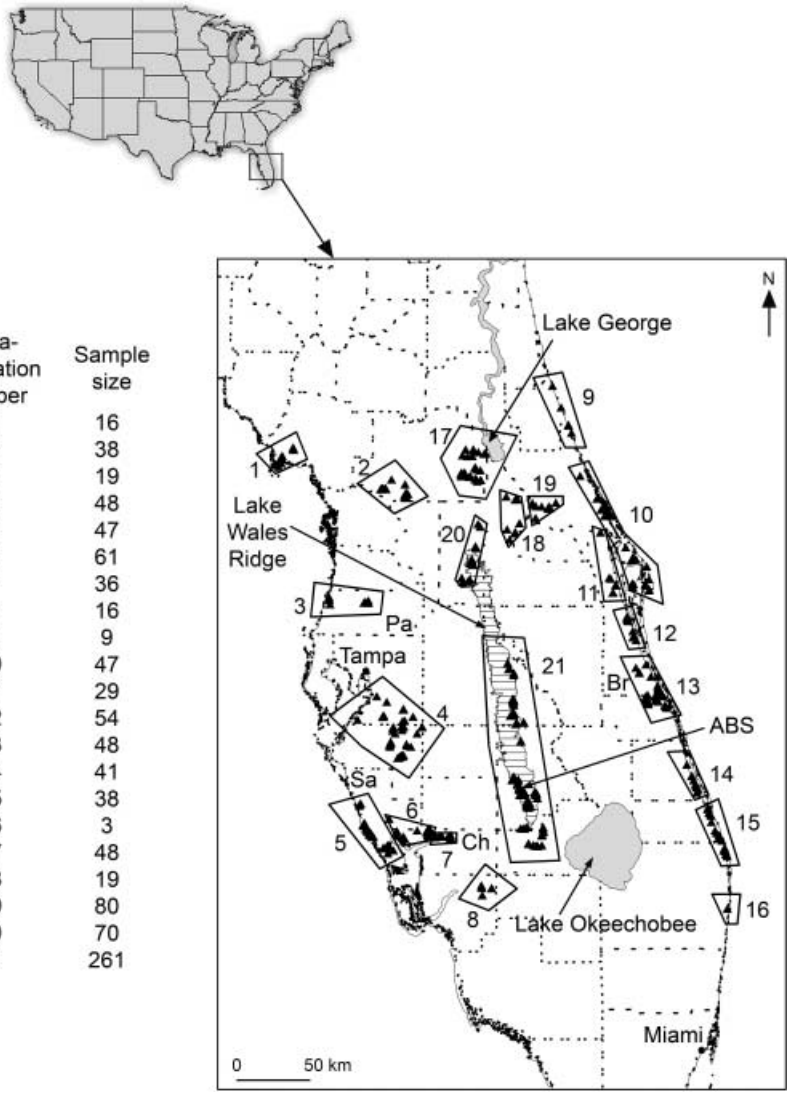

Fig. 1 Study site, sampling locations (black triangles) and sample sizes. Polygons surround individuals belonging to the same metapopulation as defined by Stith *et al*. (1996). Numbers are the metapopulation identifiers given by Stith *et al*. (1996). Dashed lines represent county limits. Br, Brevard County; Ch, Charlotte County; Sa, Sarasota County; Pa, Pasco County. ABS, Archbold Biological Station. The boundary of Lake Wales Ridge was drawn by Weekley *et al*. (in press).

opportunistically as part of other studies, or by targeted capture of birds using Potter traps, drop traps, bow nets and nylon mist-nets. Sampling locations were recorded as GPS (Global Positioning System) coordinates. From each individual, blood samples were taken via needle-pricks in the brachial vein of one or both wings, and collected into heparinized microhematocrit tubes. Blood samples were placed in 0.5 mL of lysis buffer (Hoelzel 1992) and stored thereafter at room temperature.

DNA extraction and genotyping

DNA was extracted from whole blood in lysis buffer (Hoelzel 1992) using Perfect gDNA Blood Mini Kits (Eppendorf). We genotyped each individual via polymerase chain reaction (PCR) at 20 microsatellite loci previously developed for the FSJ (Stenzler & Fitzpatrick 2002). During generation of PCR products, a multiplexing protocol was used to reduce the number of reactions per individual from 20 to six (Hailer *et al*. 2005). These reactions were optimized to 0.25 U Jumpstart *Taq* polymerase (Sigma-Aldrich) and 3.25 mM $MgCl_2$. The PCR cycling profile consisted of one cycle at 95 °C for 2 min, 35 cycles of 50 s at 95 °C, 1 min at 48 or 58 °C (specific to each locus), and 1 min at 72 °C, followed by a final extension cycle of 30 min at 72 °C. In addition to the *Taq* polymerase and $MgCl_2$, each reaction (10 μL) contained 10–100 ng of genomic DNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 200 μM of dNTPs (Invitrogen) and from 1.0 to 4.8 pM each of forward and reverse primers. Primer concentrations varied to obtain equal fluorescent signals across multiplexed loci. Genotypes were run on an ABI PRISM 3100 Genetic Analyser (Applied Biosystems). Multiplexes were designed to avoid marker overlap based on fluorescent labels and fragment sizes. Allele sizes were estimated using GENEMAPPER version 3.7 (Applied Biosystems).

Genetic clustering analyses

We performed Bayesian clustering analyses to infer spatial structure in the genetic data. Several methods are currently available that implement this general type of analysis, and alternative methods sometimes show somewhat different results (Latch *et al*. 2006). We employed the most widely used approach, which is implemented in STRUCTURE version 2.0 (Pritchard *et al*. 2000; Falush *et al*. 2003), and likewise utilized a second approach that incorporates spatial information, as implemented in GENELAND version 1.0.5 (Guillot *et al*. 2005b).

STRUCTURE *analyses*

To infer the number of genetic groups in our data set, we used the Δ*K* method of Evanno *et al*. (2005), which consists in finding the breakpoint in the slope of the distribution of lnP(D) for the different *K* values tested, where lnP(D) is an estimate of the posterior probability of the data for a given *K*. We then assigned each individual to the group for which its inferred ancestry was the highest, provided this value was higher than 0.6 (i.e. we considered that if more than half the genome of an individual is assigned to the same genetic group then this individual can be assigned to this group with reasonable confidence; the individuals with maximum inferred ancestry < 0.6 were not assigned to any group). Since Evanno *et al*. (2005) showed that this Δ*K* method detects the uppermost level of population structure when several hierarchical levels exist, we repeated the analyses (estimation of the number of groups with the Δ*K* method + assignment of the individuals to the groups) on each of the *K* groups inferred in the previous step. We repeated this process until the number of genetic groups inferred was 1, or the number of individuals was small (see Fig. S1, Supplementary material). The Δ*K* method is not appropriate when the true *K* is 1 (as there is no break in the slope in that case), so for each round of runs, we first tested for this possibility by examining whether lnP(D) was maximum for *K* = 1.

At each round of this process, runs were performed with the admixture model and the correlated allele frequency model, without prior population information and with alphapropsd set to 0.005 (higher values led to substantial variations of alpha along the runs). Each run was composed of a burn-in period of 50 000 MCMC (Markov chain Monte Carlo), followed by 1 million iterations. We checked that the length of the burn-in period was sufficient by ensuring that the lnP(D) and the likelihood of the runs had stabilized. We generated five runs per *K* value tested, with *K* ranging from 1 to 20 in the first round of runs (i.e. on the whole data set) and then from 1 to *K*max, values of which depend on the number of individuals on which the runs were performed and on the behaviour of lnP(D); if lnP(D) was still increasing for the highest values of *K*-tested, then *K*max was increased (see Fig. S1 for the *K*max values eventually used). At each round, the assignment step was performed on the outputs of the most likely run among the five replicates (i.e. with the highest lnP(D)) of the inferred *K*.

This hierarchical use of the Δ*K* method has, to our knowledge, not been applied previously. We were hence interested in comparing these results to those obtained under the alternate method, which, loosely speaking, consists in looking for the *K* value that gives the highest lnP(D), although we note that Pritchard *et al*. (2000) warned that the results obtained with this method should be treated with care and the simulations performed by Evanno *et al*. (2005) showed that the Δ*K* method is usually more reliable. For this highest lnP(D) method, we inferred the number of genetic groups with the estimator of posterior probabilities of *K* provided by Pritchard & Wen (2004), using, among the

Table 1 Summary of the goals and parameter sets of the different GENELAND runs

Parameters	Inference of *K*	Assignment of individuals to the *K* populations	Consistency check of the results with different run lengths	Consistency check of the results with different run lengths
No. of runs	10	100	3	1
K constant?	No	Yes	Yes	Yes
nit	100 000	100 000	250 000	500 000
npopmin	1	1	1	1
npopmax	50	12	12	12
delta.coord	200 m	200 m	200 m	200 m
nb.nuclei.max*	50	50	50	50
rate.max*	50	50	50	50
allfreq model	Dirichlet	Dirichlet	Dirichlet	Dirichlet
burn-in period length	—	20 000	37 500	75 000
nxdom	—	725	725	725
nydom	—	850	850	850

nit, number of MCMC iterations.
npopmin, minimum number of populations.
npopmax, maximum number of populations.
delta.coord, amount of uncertainty attached to the spatial coordinates.
nb.nuclei.max, maximum number of nuclei in the Poisson–Voronoi tessellation: each genetic group has a spatial subdomain which is the union of convex polygons (Voronoi tessellation) induced by a homogeneous Poisson process; each of these polygons is characterized by its nucleus.
rate.max , maximum rate of the Poisson process used to generate the Voronoi cells.
allfreq model = allele frequency model
burn-in period length, number of iterations, situated at the beginning of the runs, that were not used for the post-processing step.
nxdom, nydom, number of cells for discretization of the spatial domain in the horizontal and vertical directions respectively
*For these two parameters, we used lower values than those recommended in the manual of GENELAND (i.e. number equal to the number of individuals for rate.max and to three times the number of individuals for nb.nuclei.max). Indeed, when the number of individuals is high (> 100) it is not necessary to use such high values, which trigger very high computation times (G. Guillot, personal communication). Instead, one can try with a value of, say, 100, and then refine it from the behavior of the chain along the run; if the rate of the Poisson process and/or the number of nuclei take values close from 100, one should then increase rate.max and/or nb.nuclei.max. If, on the contrary, the chain stayed very far from 100, one can decrease them.

five replicate runs, the one with the highest lnP(D) value; we confirmed this diagnostic by a visual observation of the plot of the lnP(D) values of the different runs performed, as a function of the *K*-values tested. We then ran 100 supplementary runs with the same parameters as above and *K* set to the inferred value and examined the assignment results of the best of them (see below, 'Consensus analyses').

GENELAND analyses

In our GENELAND analyses, we first performed a series of runs to infer the number of genetic groups (*K*) in our sample. We conducted multiple preliminary runs to adjust input parameter values based on the behaviour of the MCMC, thereby ensuring that the maximum values we set for the parameters were large enough to allow the MCMC to explore all likely regions of the parameter space, and confirming that the chains converged by the end of the runs. We then performed 10 runs of 100 000 MCMC iterations, each with the selected set of parameters (see Table 1, column 2) and where *K* was allowed to vary. Because of their strong territorial behaviour (Woolfenden & Fitzpatrick 1984), we assumed birds were sampled in or very close to their natal or breeding territories. The uncertainty associated with the spatial coordinates was hence set to 200 m, the mean territory radius (Woolfenden & Fitzpatrick 1984). Two allele frequency models are available in GENELAND; we used the Dirichlet option (allele frequencies assumed to be independent among populations), as Guillot *et al*. (2005a) showed it performs better than the F-model (allele frequencies assumed to be nonindependent). *K* was inferred as the modal number of genetic groups estimated among the 100 000 iterations of the best of the 10 runs. To select this best run, we used the posterior density of the runs as an estimator of their quality: the posterior density is estimated for each state of parameters explored along the Markov chain, and represents the posterior probability of that current state of parameters.

The second step consisted of running the model to have it assign each individual to one of the *K* genetic groups. We performed 100 independent runs using the parameter set established in the first step, with *K* fixed to the number of populations inferred therein (Table 1, column 3). To assign individuals to genetic groups, GENELAND first calculates

the posterior probabilities of genetic group membership for each cell in a spatial domain. We divided the current range of the FSJ into a grid spanning 725 cells along the east–west axis, and 850 cells along the north–south axis, yielding a cell size of approximately 400 m on a side (= typical diameter of an FSJ territory). We ranked the 100 runs according to their mean logarithm of posterior density (the plot of the mean of the posterior density along the runs showed that the chains converged after the 20 000th iteration, so we calculated these means only on the iterations sampled after this initial burn-in). We looked at the results of assignments of the best of them (see below, 'consensus analyses').

To check that the results were not affected by the length of the runs (which was unlikely given that the runs reached convergence quickly), we also performed four longer runs: three with 250 000 iterations and one with 500 000 iterations. All other parameters were identical to those used for the 100 runs with 100 000 iterations, except the length of the burn-in periods, which had to be increased (Table 1, columns 1 and 5).

Consensus analyses

In Bayesian clustering, it is common that replicate runs give slightly different solutions (Jakobsson & Rosenberg 2007). We employed two different approaches to analyse the 100 runs performed with the highest lnP(D) implementation of STRUCTURE and the 100 runs performed with GENELAND: one that we developed, hereafter termed CONSANA; and one that was recently released, called CLUMPP (Jakobsson & Rosenberg, in press). Neither type of consensus analysis could be used with our hierarchical application of the Δ*K* method for STRUCTURE because of the high number of rounds on which we would have had to apply them.

The first step of each consensus analyses was to select the best of the 100 runs performed: we ranked the runs by their mean posterior densities and selected those that had the highest values, based on the pattern of decline of the mean posterior densities.

CONSANA: To better define the features common to all or most of the replicated reconstructions of genetic population structure, we devised a consensus approach that generates a composite map based on these commonalities (functions for the program R (Ihaka & Gentleman 1996); available by request to the corresponding author). This method creates a pairwise matrix that sums the number of times two individuals were assigned to the same genetic group in the replicate runs. The vector of assignments necessary to perform this step is one of the outputs of the software in the case of GENELAND; for STRUCTURE, we created it by assigning each individual to the group for which it had the highest inferred ancestry. The program then clusters together all such paired individuals assigned to the same genetic group in more than x% of the runs, x being defined by two criteria. First, for some clustering thresholds of x (usually when they are low), there are cases where clustering patterns are nontransitive (e.g. where individuals A and B are consistently assigned to the same genetic group, and likewise B and C, but not A and C). The presence of nontransitive groupings in the combined analyses suggests that these particular groupings are not consistent among runs, and we sought a threshold that minimizes the number of nontransitive cases. The second criterion is based on the observation that as the clustering stringency threshold is raised to 100%, an increasing number of individuals are assigned as singletons to their own genetic group, and likewise the number of very small (two to five individuals) genetic groupings increases. The threshold is hence best set to the lowest percentage value that avoids nontransitive assignments.

CLUMPP: In contrast to CONSANA's basis in the vector of assignments of individuals to genetic groups, CLUMPP works on the matrices of membership coefficients of each individual to the *K* genetic groups. It hence has to deal with the problem of label switching, i.e. the fact that among different runs, the same group may have been assigned different labels. CLUMPP first uses an algorithm to find the optimal alignment of the replicate matrices, and then it calculates the mean of the permuted matrices across replicates. Because of the high number of genetic groups we have inferred and the high number of replicate runs in our analyses (see Results), we used the LargeK-Greedy algorithm, with 30 000 random input orders. We used H as the estimate of average pairwise similarity between matrices. We then assigned each individual to the genetic group for which it had the highest mean membership coefficient.

Visualization of the population structure patterns

We plotted the results of the assignments of STRUCTURE and of GENELAND on maps generated with the GIS software ARCGIS 9.1. (ESRI). We surrounded the individuals assigned to the same genetic groups with minimum convex polygons (MCP) generated with the extension Hawth's analysis tools for ARCGIS (Beyer 2004). For improved clarity, we excluded some outlier individuals from these polygons, i.e. single individuals that were located within a cluster of individuals belonging to a different genetic group than their own. These outliers are illustrated by symbols in the maps. If more than one individual was assigned to a different group than those around them, we included them in the MCP drawing.

These maps were then compared to the map of the metapopulations provided by the territory-clustering approach of Stith *et al*. (1996).

Analyses of genetic groups

We estimated several characteristics of the inferred genetic groups — among the different results of the various methods we used, we based these analyses on the grouping pattern that we judged as being the most biologically plausible (see Results). For each group, we calculated observed heterozygosity and heterozygosity expected under Hardy–Weinberg equilibrium, the latter corrected for sampling bias, using GENETIX 4.05.2 (Belkhir *et al*. 1996–2004). We tested for the significance of heterozygote deficiency with the Markov chain method of GENEPOP 3.4 (Raymond & Rousset 1995). Default values of Markov chain parameters were increased in order to obtain standard deviations lower than 0.01 for *P*-value estimates. We performed 10 000 dememorizations (default value), 200 batches, and 2000 iterations of batches. We also tested for the presence of a significant deficiency of heterozygotes in the total sample (i.e. without any partitioning of individuals to populations), using 10 000 dememorizations, 200 batches, and 3000 iterations of batches. We estimated F_{IS} values within each group with GENETIX. We tested for the presence of linkage disequilibrium among loci with the Markov chain method of GENEPOP. Here also, default values were increased in order to obtain standard deviations for the *P*-value estimates lower than 0.01; we used 10 000 dememorizations, 1000 batches, and 10 000 iterations of batches. We also calculated Weir & Cockerham's (1984) F_{ST} and pairwise F_{ST} estimator with GENETIX and tested their significance with the permutation test (10 000 permutations).

When applicable, we controlled for multiple comparisons by calculating the *P* values adjusted for FDR (false discovery rate) with the function compute.fdr for R 2.3.1. The library of the function is available online at http://www.stjuderesearch.org/depts/biostats/documents/fdr-library.R. We used the method of Benjamini & Hochberg (1995), which controls the proportion of significant results that are in fact false positives (type I errors). It rejects the null hypotheses of the tests which *P* values are lower than or equal to alpha/m*i. Alpha is the desired level of type I errors, 'm' is the number of tests performed and 'i' is the number of tests that have lower *P* values than the test currently assessed. FDR control has the advantage of being less stringent than the widely used Bonferroni correction, and hence avoids the considerable loss of power triggered by the use of this criticized method (e.g. Moran 2003; Verhoeven *et al*. 2005).

Results

We genotyped 1028 FSJs (49 on average per metapopulation, min = 3, max = 261; see Fig. 1), each at 20 microsatellite loci. All 20 loci were polymorphic, with three to 22 alleles per locus and a mean of 9.6 alleles per locus (SD = 4.5) (Table S1, Supplementary material). Based on the high level of geographical structuring in this species, we expected the collective data set to exhibit signs of a Wahlund effect. This was indeed the case, with the global sample of microsatellite loci and of individuals showing a highly significant deficit of heterozygotes as compared to Hardy–Weinberg equilibrium expectations ($P < 0.0001$).

Genetic clustering

STRUCTURE analyses under the ΔK *method.* At each round of the process, the consistency of the lnP(D) values among the different runs with the same *K*, compared to the variability observed among runs with different *K*s, indicated that these runs were adequately long (not shown). The results of the different rounds of the process are summarized in Fig. S1. In total, 898 individuals were assigned to 22 genetic groups, and 130 individuals were not assigned to any group because their maximum estimated membership coefficients were below 0.6. The map of these assignments is shown in Fig. 2. The first round of runs inferred two groups, one including the individuals from the north and the east of the peninsula (groups with names beginning with '1'), and the other including the individuals from the southwest (names beginning with '2'). Group 1 was then subdivided into two groups, a and b, with b including the individuals from the east coast (with two exceptions) and a all others. Group 2 was also subdivided into two groups, a and b, a mostly including individuals in the west and b those in the east. Then, each of these groups (except 2b) was further subdivided (Fig. 2 and Fig. S1). Eventually, this method inferred genetic groups that were often spatially overlapping and, for some, that were defined at a surprisingly fine spatial scale (e.g. groups 1a1a, 1a1b and 1a1c).

STRUCTURE analyses under the highest lnP(D) method. The estimator of posterior probabilities of *K* indicated that the most likely *K* was 13: P(*K* = 13) = 1. However, the variability of lnP(D) among different runs performed with the same *K* increased for values of *K* higher than 10 (Fig. S2, Supplementary material). As there was more variability in the quality of the runs for *K* = 13, and the lnP(D) value of the runs of *K* = 10 was almost as high as for the best run for *K* = 13, we chose *K* = 10 as the most conservative value and performed the subsequent 100 runs with *K* set to 10. The quality of these runs, as estimated by their lnP(D), declined more sharply after the 72nd best run (Fig. S3a, Supplementary material). We hence performed the consensus analyses on the 72 best runs.

CONSANA: We explored threshold values ranging from 50 to 90%, and selected 72% as the value that minimized the number of transitive cases (to 1). The 1028 individuals

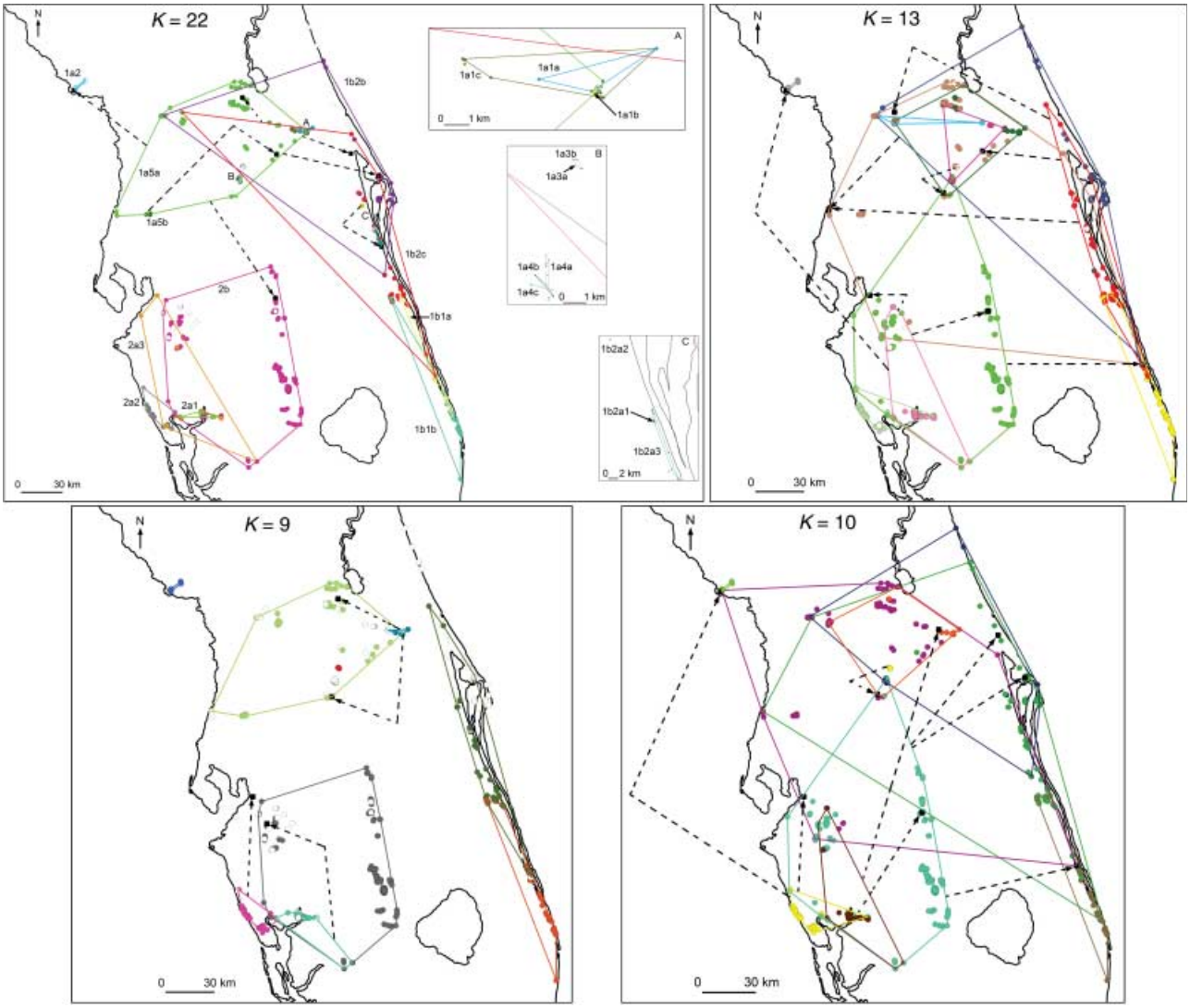

Fig. 2 Results of the genetic clustering analysis STRUCTURE. Polygons are the minimum convex polygons surrounding the individuals belonging to the same genetic group. White circles are individuals that were assigned to no genetic group. Dashed arrows indicate isolated individuals that were located within a cluster of individuals belonging to a different genetic group than their own; they originate on a boundary of the genetic group to which the individual was assigned. Top left: Δ*K* hierarchical method. A, B and C refer to the magnified areas, showing patterns visible at a fine spatial scale only; the other designations are the names of the genetic groups, as defined in Figure S1, Supplementary material. Top right: highest lnP(D) method, combined with CONSANA. Bottom left: highest lnP(D) method, combined with CLUMPP with a threshold of 0.6 for the assignment of individuals to genetic groups. Bottom right: highest lnP(D) method, combined with CLUMPP without any threshold.

were assigned to 13 genetic groups, composed of two to 327 individuals (Fig. 2). The main patterns were similar to those inferred with the Δ*K* method, with three distinct groups in the east coast, three in the southwest of the peninsula, and one in the northwest, but most of the groups were highly overlapping, some containing individuals scattered through most of the range of the species. The north-central part of the range showed a particularly confusing structure.

CLUMPP: We used a threshold of 0.6 to assign individuals to their most likely genetic group. This resulted in nine genetic groups, composed of 13–297 individuals, and 243 individuals without assignment (Fig. 2). These groups showed much less overlap than in the previous methods, and exhibited the same major patterns except that the east coast was comprised of only two groups. However, as many individuals were not assigned, we used the same method without a threshold. This produced a clustering

Table 2 Characteristics of the populations defined by the genetic clustering method GENELAND, in combination with CONSANA. See Fig. 4 for their location in the species' range. H_E n.b. is the expected heterozygosity under the Hardy-Weinberg hypothesis corrected for sampling bias, H_O is the observed heterozygosity. q is the FDR-adjusted P value (FDR, false discovery rate) of the Hardy-Weinberg tests

Genetic group	No. of individuals	H_E n.b.	H_O	F_{IS}	q
A	128	0.692	0.668	0.035	0.001***
B	262	0.686	0.663	0.033	< 0.001***
C	59	0.705	0.702	0.005	0.301
D	176	0.734	0.701	0.044	< 0.001***
E	47	0.592	0.617	–0.041	1
F	140	0.682	0.668	0.021	0.086
G	77	0.693	0.675	0.026	0.032*
H	15	0.579	0.637	–0.103	1
I	82	0.716	0.673	0.060	< 0.001***
J	22	0.616	0.666	–0.082	1
K	16	0.645	0.634	0.017	0.386
L	2	—	—	—	—

*: $q \leq 0.05$.
***: $q \leq 0.001$.

pattern of 10 genetic groups (14–333 individuals per group) that, although showing the same main clustering patterns, were much more overlapping (Fig. 2).

GENELAND analyses

In the 10 runs performed to estimate K, the posterior density and the log-likelihood levels reached a plateau well before the end of the MCMC runs, indicating that they had reached convergence. These replicate runs gave generally consistent results: in seven of 10 runs, the modal number of genetic groups (K) estimated along the MCMC was 12; in the remaining three runs, the modal values of K were 9, 11 and 13. The run with the highest mean posterior density likewise had a K of 12 (based on their mean posterior density the runs with 9, 11 and 13 as modal values were ranked 10th, eighth and third, respectively). Therefore, subsequent runs were performed with K set to 12.

The 100 runs of 100 000 iterations performed for the assignment step also reached convergence (not illustrated). Many runs assigned individuals to fewer than 12 groups, showing 'ghost populations' (Guillot *et al*. 2005a), i.e. populations that were modal to none of the individuals and to which no individual was assigned. The results of the three longest runs showed the same consistent patterns as the shorter runs. When the 100 runs were ranked by the mean posterior densities, their qualities (as estimated by these mean densities) showed only a slight decrease through the first 85 runs, and then a sharp decrease in the lowest-ranked 15 runs (Fig. S3b, Supplementary material). We therefore looked for consistent structuring features among the 85 best runs.

CONSANA: We explored thresholds ranging from 50 to 80%. There were nontransitive cases when these clustering thresholds were low (50–65%). Groupings at and above the 70% threshold were entirely transitive. We therefore chose the 70% threshold as a conservative value for defining general clustering patterns among the 85 runs. Twelve genetic groups were consistent across all 85 runs, ranging in size from two to 262 individuals (Table 2, Fig. 3). The clustering pattern was very similar to the one revealed by STRUCTURE/CLUMPP with the threshold of 0.6. The main differences apparent in the GENELAND/CONSANA results were: the existence of a third group on the east coast; the separation of the individuals from Lakes Wales Ridge from those located farther west; and the separate clustering, in the west, of the individuals from the southernmost part of the range. The other major difference is that with this method all but two individuals were assigned to a genetic group.

CLUMPP: The 1028 individuals were assigned to eight genetic groups, ranging in size from one to 274 (Fig. 3). The clustering pattern mirrored that obtained with CONSANA, with two differences. First, there was a biologically implausible grouping of individuals from the southernmost group of the east coast with individuals from the southwestern, the northwestern and the north-central part of the peninsula (in CONSANA these individuals were assigned to four distinct groups). Second, the individuals from the southernmost part of the west of the peninsula were assigned to the same group as those of Lake Wales Ridge, whereas in CONSANA they were a distinct entity.

Comparison of the genetic clustering pattern with the metapopulations defined by Stith et al. *(1996)*

Although differences are apparent, the results obtained with each genetic clustering method nevertheless showed some consistent patterns, with a more or less clearly defined separate clustering of the individuals of the east

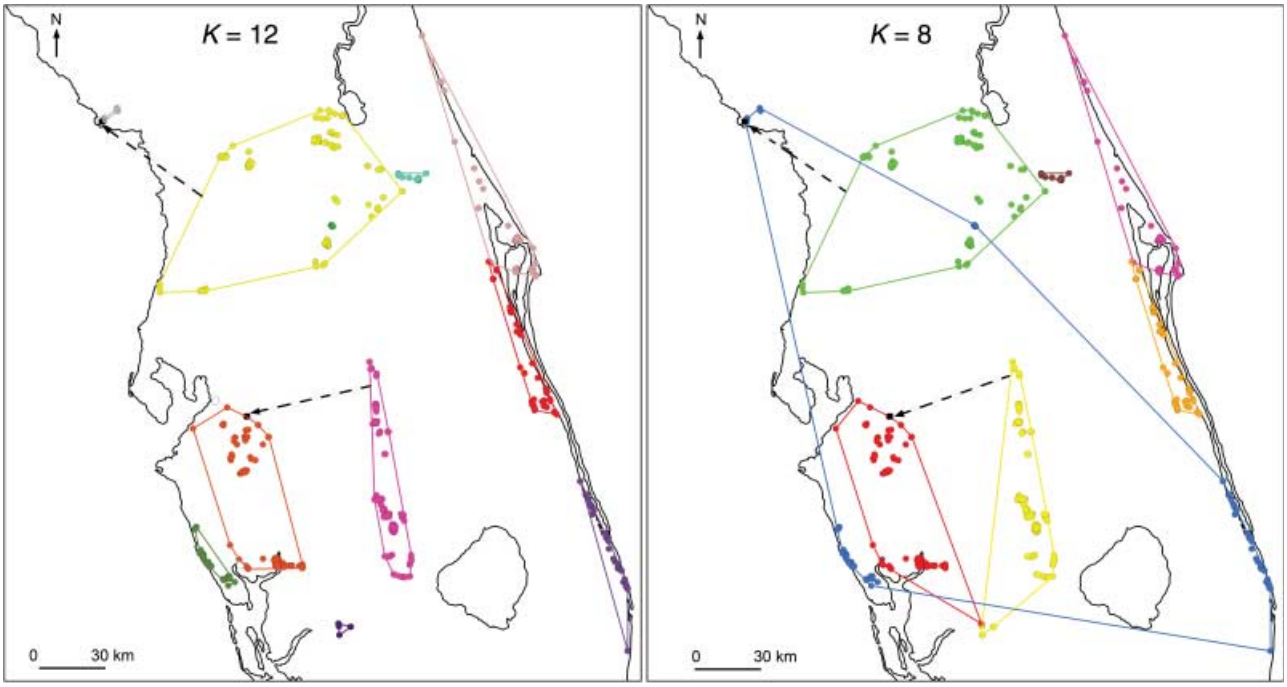

Fig. 3 Results of the genetic clustering analysis GENELAND. Polygons are the minimum convex polygons surrounding the individuals belonging to the same genetic group. Dashed arrows indicate isolated individuals that were located within a cluster of individuals belonging to a different genetic group than their own; they originate on a boundary of the genetic group to which the individual was assigned. Left: output from CONSANA. Right: output from CLUMPP.

coast (into two or three groups according to the methods), those of the southwest and those of the northeast. Among the different results, those obtained with STRUCTURE/CLUMPP/threshold, GENELAND/CONSANA and GENELAND/CLUMPP are the more biologically plausible, as they involve the least overlapping of groups. These overlapping patterns are not due to isolated individuals that could be migrants, as we excluded those potential isolated 'migrants' from the drawing of the polygons representing the genetic groups. Among those three methods, STRUCTURE/CLUMPP/threshold involved the dropping of 243 individuals, because of their low assignment level, and GENELAND/CLUMPP generated a very implausible clustering of individuals (southeast/southwest/northwest/north-central). We hence decided to compare the genetic clustering and the metapopulations of Stith *et al*. (1996) primarily to the results of GENELAND/CONSANA, the genetic analysis that we believe is the most accurate.

In the GENELAND/CONSANA clustering (Fig. 4), one group (L) was composed of only two individuals. They were assigned in half the runs to genetic group E and in the other half to genetic group F (except in one run where they were assigned to the genetic group H). This intermediate assignment explains why they were assigned to their own very small genetic group in the consensus analysis. The same phenomenon explains why two individuals were assigned as singletons to their own genetic groups (see Fig. 4): they were assigned alternately to two groups in the 10 best runs.

Boundaries of the genetic groups accorded well with metapopulation boundaries defined by Stith *et al*. (1996). Either the two types of units match exactly (i.e. genetic group/metapopulation: B/21, E/5, G/19, H/1 and K/8), or genetic groups cleanly comprise two or more metapopulations (i.e. A/11-12-13, C/9–10, D/2-3-17-18–20, F/4-6-7, I/14-15-16). In no instance was a single metapopulation split into multiple genetic groups, except in the surprising pattern occurring within metapopulation 20 (see Discussion). Only six of 1028 sampled individuals were assigned to a genetic group other than the one that included the remaining individuals from their spatial metapopulation (see Fig. 4).

Characteristics of the genetic groups obtained with GENELAND/CONSANA

We excluded the group L from these analyses because of its extremely small sample size of only two individuals. Heterozygosity levels were high, ranging from 0.617 to 0.702 (Table 2). Five genetic groups showed a significant deficiency of heterozygotes after FDR correction (groups A, B, D, G and I), and a sixth (group F) had a *P* value close to significance (Table 2). A small proportion of locus × genetic group pairs (23 such pairs of loci out of 2090

Table 3 Pairwise F_{ST} between the genetic groups

	B	C	D	E	F	G	H	I	J	K
A	0.092	0.035	0.045	0.110	0.079	0.063	0.128	0.047	0.120	0.120
B		0.085	0.061	0.094	0.029	0.086	0.141	0.078	0.155	0.048
C			0.045	0.138	0.087	0.072	0.132	0.065	0.124	0.123
D				0.097	0.063	0.025	0.092	0.045	0.101	0.085
E					0.062	0.123	0.167	0.111	0.214	0.113
F						0.088	0.136	0.069	0.155	0.041
G							0.136	0.069	0.110	0.115
H								0.130	0.228	0.171
I									0.115	0.084
J										0.172

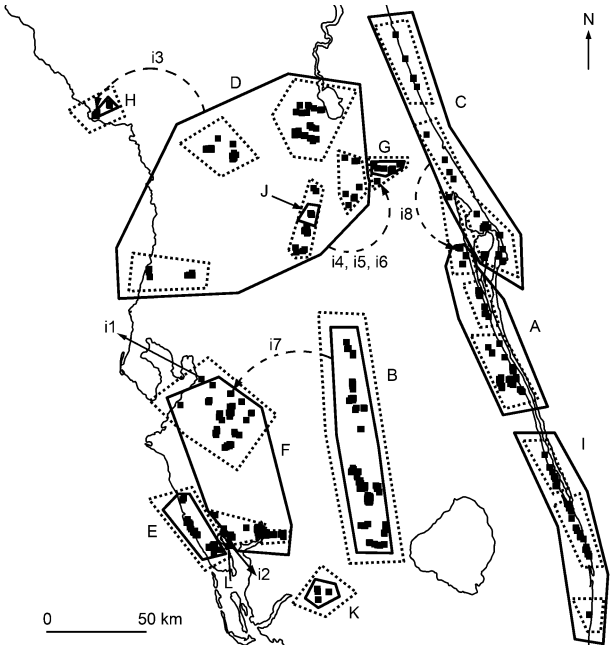

Fig. 4 Results of the genetic clustering analysis GENELAND obtained with CONSANA, and the metapopulations of Stith *et al*. (1996). Solid lines surround the individuals assigned to the same genetic group. Letters are identifiers of the genetic groups. Dashed lines surround the individuals belonging to the same metapopulation as defined by Stith *et al*. (1996). i1 and i2 are the only two individuals assigned as singletons to their own genetic groups. i3 to i8 are potential migrants; the arrows point toward the location in which they were sampled and originate in the genetic group to which they were assigned. The migrants are here defined based on the metapopulation design.

total) showed significant linkage disequilibrium, after correcting for multiple testing. In the majority of cases, one locus was involved in only one or two of these pairs. The most notable exceptions were ApCo68, ApCo29, ApCo36 and ApCo 41, involved, respectively, in 4, 5, 6 and 8 pairs. The pairs exhibiting significant linkage disequilibrium were not consistent among the genetic groups, except from one that appeared in four groups (ApCo29-ApCo36) and two that appeared in two groups (ApCo15-ApCo41 and ApCo41-ApCo68).

Global F_{ST} was 0.0735, and highly significant ($P < 0.0001$). All pairwise F_{ST} values between genetic groups also were highly significant ($P < 0.0001$), with values ranging from 0.0254 (between genetic groups D and G) to 0.228 (genetic groups H and J; Table 3).

To confirm that these results were not biased by the few individuals assigned to genetic groups not congruent with their geographical metapopulations, we performed the same analyses with the six individuals excluded. The results (not shown) were very similar to those reported above.

Discussion

Spatially explicit Bayesian clustering to infer genetic groups

As a consequence of their statistical power, the development of Bayesian clustering methods has expanded rapidly. The results obtained with different methods can differ, and a recommended approach is to try and compare several of them (Latch *et al*. 2006; Chen *et al*. 2007). However, it is not clear how best to deal with variability among the results. Here, we decided to base our choice (GENELAND, analysed with CONSANA) on biological plausibility.

Some of the runs of GENELAND exhibited 'ghost' populations to which no actual individual was assigned. This is a typical outcome of the GENELAND clustering process (Guillot *et al*. 2005a; Coulon *et al*. 2006; Pilot *et al*. 2006), which likely stems from the fact that any clustering model is a simplification of the many processes that naturally contribute to spatial patterns of genetic structuring. As a result, some of the assumptions (such as completely random mating within populations) of the method are not fully fulfilled, creating spurious classes in the runs with varying *K* (Guillot *et al*. 2005a; Coulon *et al*. 2006). Real FSJ

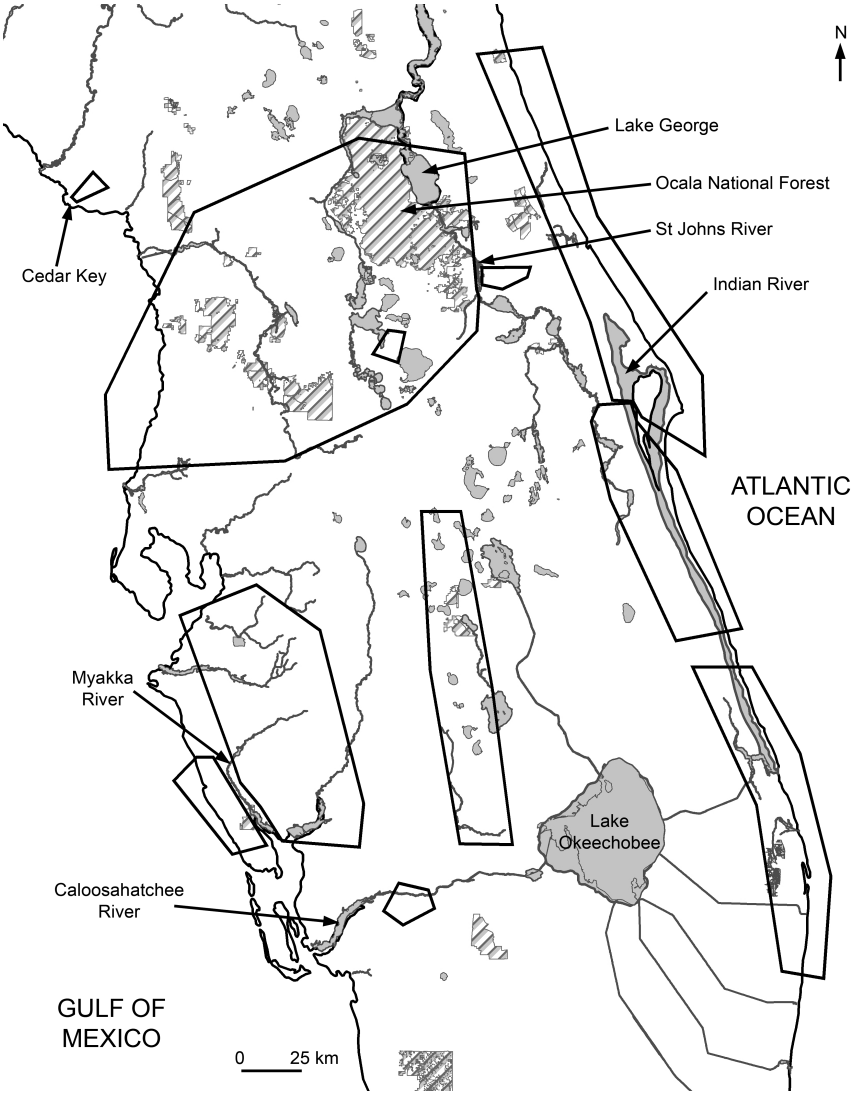

Fig. 5 Summary of relevant landscape features, including major forests (grey shading), lakes, and rivers. The genetic groups are represented by the same solid lines as in Fig. 4.

populations are unlikely to be truly panmictic, and they indeed exhibit within-population isolation by distance (unpublished data). No model exists that is able to deal fully with the presence of isolation by distance (Guillot *et al*. 2005a). The existence of within-population isolation by distance can also explain the Hardy–Weinberg disequilibrium exhibited by some of the genetic groups inferred by GENELAND.

To deal with the underlying variation that can appear in replicate runs of Bayesian clustering analyses, we replicated our clustering runs many times, and developed a consensus technique, CONSANA, to identify the well-supported features common to the entire set of replicates. Compared to another consensus technique recently released, CLUMPP, the disadvantage of CONSANA is that it requires the user to set a threshold in a 'trial-error' fashion and can sometimes lead to cases where there are not entirely transitive solutions (i.e. where there is no threshold for which there are no cases with individuals A and B consistently assigned to the same genetic group, and likewise B and C, but not A and C). However, in our data set, CONSANA, combined with GENELAND, provided the most biologically plausible clustering pattern.

Range-wide genetic structure of the Florida scrub-jay

The Florida scrub-jay is currently subdivided into 10 major genetic groups (Figs 4 and 5), not including two additional small peculiar groups (J and L) which are discussed below. Three of these 10 groups (A, C and I in Fig. 4) occur on Florida's east coast along a nearly continuous, narrow, 330 km long ancient beach ridge. Boundaries separating these east coast populations coincide with a broad water barrier (Indian River, separating A and C) and a region of sparsely distributed scrub patches long disturbed by human settlement (between A and I). In

central Florida, the Lake Wales Ridge constitutes a genetic group (B) bounded both east and west by wide prairies. In the southwest, three genetically differentiated populations (E, F, and K) are separated by two major river systems and their associated mesic hardwood and pine forests (Myakka and Caloosahatchee rivers). In contrast to these well-differentiated populations, those in the northern peninsula are more homogeneous. A single genetic group (D) encompasses individuals ranging from Lake George in the Ocala National Forest to northern Pasco County on the west coast. A distant isolated scrub along the northern Gulf Coast (Cedar Key, H) is separated from this large northern group by flatwoods and sandhill forests. A small genetic group (G) exists on ancient scrub soils remarkably close to the widespread northern population (D), but separated by the relatively wide St Johns River. All these clustering patterns confirm that certain landscape features, such as extensive forests and large expanses of open water, serve as barriers that have limited FSJ dispersal movements over long timescales. Nevertheless in a few cases, the areas encompassed by the genetic groups inferred in this study include large rivers (e.g. groups D and F), suggesting that water gaps do not always act as barriers to FSJ gene flow. The width of these rivers or the composition of the landscape surrounding them may influence their degree of permeability. More detailed landscape genetic analyses are necessary to understand the specific effects of the various landscape components on FSJ movements.

An 11th genetic group (J), comprised of 22 sampled individuals, is embedded within the geographical area occupied by the genetic group D. In the 10 most likely GENELAND runs, these 22 individuals always clustered together, generally with the individuals of some other genetic group (the identity of which differed across the 10 runs). In all other methods tried in this study, these individuals were also clustered apart from their neighbours. One reason could be that by chance, these 22 individuals might be closely related. However, this would result in a high F_{IS} value, which is not the case ($F_{IS} = -0.082$). A more likely explanation is that the very high density of lakes and orange groves surrounding this small area has limited FSJ movements and isolated the individuals of this group. Conversely, the presence of similar landscape features does not seem to have impeded FSJ gene flow in other areas, as in group B, or even between the group D individuals to the north and south of group J.

Finally, a 12th group (L) was composed of only two sampled individuals, exhibiting nearly equal probabilities of assignment to either groups E or F. This pattern could stem from admixture events between groups E and F (reproduction between individuals of these two groups). The same type of explanation also could hold for the two individuals assigned as singletons to their own genetic groups (i1, i2; see Fig. 4).

Six individuals were assigned to a different genetic group than were the other individuals from their geographical metapopulation (see Fig. 4). One possibility is that these individuals are immigrants from the genetic group to which they were assigned, although the long dispersal distances required under this scenario for two of them (male i3 and female i7, 70 and 75 km, respectively) are surprising. These six potential dispersal events link groups between which the levels of genetic differentiation are among the lowest (pairwise F_{ST} of 0.025, 0.029, 0.035 and 0.092).

High congruence between metapopulations and genetic groups

The metapopulations of Stith *et al*. (1996) in the 1990s were defined as discrete clusters of then-occupied territories within which movements can likely occur, but among which dispersal is impeded by distance or habitat features likely acting as barriers to FSJ movements. To the extent that such extant habitat barriers reflect ancient landscape features, these metapopulations might be expected to reflect genetic structuring across the species as a whole. The Florida scrub-jay demonstrates strong concordance between metapopulation boundaries defined by dispersal assumptions and those of the genetic groups defined on the basis of microsatellite variation. In all cases, either the genetic groups and the metapopulations had the same boundaries (H/1, G/19, B/21, K/8 and E/5), or the genetic groups cleanly encompassed several metapopulations (D/2-3-17-18–20, C/9–10, A/11-12-13, I/14-15-16, F/4-6-7). No metapopulation based on territory clustering appeared genetically subdivided, except for the peculiar group J (see above) and the few cases of potential immigrants or admixed individuals (involving eight of the 1028 genotyped individuals).

The fact that a number of the FSJ genetic groups include several putative metapopulations indicates that genetic structuring occurs at a somewhat larger spatial scale than predicted by the present-day territory clustering. Indeed, some of today's metapopulations probably were demographically connected in the recent past. Most importantly, the xeric landscapes inhabited by these jays have been severely reduced in area over the past century (Fernald 1989; Myers & Ewel 1990; Fitzpatrick *et al*. 1991; Woolfenden & Fitzpatrick 1996). For example, between the late 1800s and 1992, the area encompassed by the group F saw a reduction of 82% in potential FSJ habitat (62 689–11 353 ha) and the number of patches decreased from 1887 to 452. Similarly for group D, the area of potential scrub decreased of 79% (614 769–126 641 ha) and the number of patches decreased from 4260 to 329 (unpublished data). The current pattern of genetic structuring presumably retains a strong signal of the predevelopment

pattern of connectivity, whereas the territory clustering predicted metapopulation structure at a specific recent point in time when many populations had already shrunk or gone extinct. Thus, many of today's apparent metapopulations were substantially less isolated from one another prior to the recent fire suppression and fragmentation of their landscape.

It is also certain that some jays were missed during the 1992–1993 survey, and had they been included in the metapopulation analysis, the dispersal buffers applied by Stith *et al*. (1996) would likely have been larger and may have even more closely matched the genetic groupings documented here. This explanation is most likely in portions of the state that could not be surveyed because access to large private properties was denied, or because soil types did not correspond well with habitat types (Stith 1999, e.g. F, D). It also is certain that at least occasionally FSJs disperse farther than assumed by the 12-km threshold territory clustering method. However, the species' social system dictates that such long-distance dispersers even more rarely become breeders (Fitzpatrick *et al*. 1999). Moreover, although some overall dispersals longer than 12 km have been documented, to date, all have involved birds passing through a landscape mosaic that includes scrub-habitat 'stepping stone' patches no greater than 12 km apart. The territory clustering explicitly excluded currently unoccupied patches that could have been occupied in the past or which could continue to serve as stepping stones. Despite these assumptions, the strong congruence between the metapopulations and genetic groups provides yet another confirmation of the poor dispersal ability of these jays.

The combined methods of Stith *et al*. (1996) to define metapopulations and Bayesian genetic clustering to infer genetic groups provide complementary and congruent information regarding FSJ population structure, and a robust depiction of the current genetic and demographic organization of this species reflecting past and present levels of dispersal among occupied habitat patches. Assuming that the differences between the metapopulations and the genetic groups mainly reflect the past existence of now extinct FSJ populations, then the metapopulations of Stith *et al*. (1996) may be a good estimate of what the FSJ genetic structure will be in the future. However, this assumes that populations present in the early 1990s will persist, an assumption that is not being realized in present-day Florida.

This empirical study illustrates that the combination of direct and indirect estimates of gene flow can provide complementary information about the genetic structure in a data set and should be encouraged. In particular, these complementary approaches can give insight into recent changes in gene flow patterns, which, as in this case study, can be important for understanding how the species might respond to landscape changes. However, an almost universal difficulty is that reliable direct dispersal data are a challenge to obtain.

Conservation and management implications

Biologically sound conservation and management decisions about endangered species require knowledge of how the species is spatially structured and to what extent its populations interact demographically. Data on past and present connectivity provide vital information on the spatial scale at which conservation measures should be applied. This general idea has led to the definition of conservation and management 'units' via various criteria, but debate continues as how to best define such units using genetic markers and other approaches (e.g. Pennock & Dimmick 1997; Waples 1998; Paetkau 1999; Crandall *et al*. 2000; Fraser & Bernatchez 2001; Moritz 2002; Palsbøll *et al*. 2007). No general consensus has emerged, but as pointed out by Green (2005), many conservation problems are too urgent to wait for resolution of these debates. Such is the case for the Florida scrub-jay. For this species, a very large number of local populations have already gone extinct (Woolfenden & Fitzpatrick 1996). Most of the rest — remarkably genetically distinct from one another — are in imminent risk of the same fate.

The Florida scrub-jay can now be understood as representing no fewer than 10 genetically distinct groups, which today are further subdivided into at least 21 demographically isolated populations. Some genetic groups show substantial differentiation (e.g. A/E, H/B, H/E ...) of almost the same order of magnitude as the differentiation between two closely related species, the island scrub-jay (*Aphelocoma insularis*) and the western scrub-jay (*Aphelocoma californica*) (F_{ST} = 0.184, Delaney & Wayne 2005). The FSJ has likely lost a substantial but unknown amount of genetic variation already through the extinction of other isolated populations. Maintaining all remaining genetic variation now present will require conservation efforts that foster long-term persistence of each of the 10 distinctive groups, with special urgency attached to the smallest and most vulnerable of these groups. The small populations located at the perimeter of today's contracted range (group G, east of the St Johns River; group H, at Cedar Key; and group K, south of the Caloosahatchee River) demand immediate conservation action. Maintaining genetic variation within the larger genetic units (e.g. groups C, D and F; Fig. 4) will be facilitated by restoring suitable habitat to connect via corridors or stepping-stones the now-isolated populations within these groups. The spatial pattern of genetic variation identified here should assist in identifying genetically appropriate source populations for translocations of FSJs to formerly occupied habitat patches, as the all-important preserve networks are created and fire-suppressed scrub habitats are restored. Finally, the genetic

patterns reported here confirm the suspicion of Stith *et al*. (1996) that in order for preserve networks to maintain full genetic viability of FSJs within each metapopulation, individual tracts must be located within a landscape configuration promoting dispersal among them. In this respect, landscape genetic approaches are promising tools for inferring more precisely how landscape elements influence gene flow (Manel *et al*. 2003).

Acknowledgements

We thank the staff and interns at Archbold Biological Station, who contributed widely to the blood sampling effort, especially Robert Aldredge, Raoul Boughton, Larry Riopelle, and Christopher Valligny. We also thank the numerous collaborators, both private individuals and state and federal agency staff, with whom we have worked. Many of these collaborators tamed birds prior to trapping and provided us with important details about the populations from which we sampled. We also thank Jay Garcia, David Gordon, Lee Walton, Karl Miller, Bill Pranty, John Puschock and Glen Woolfenden for providing blood samples from their own research on FSJs, and Roberta Pickert for help with GIS data analysis. We thank Amanda Talaba for genotyping assistance, Roger Slothower for GIS advice, and Gilles Guillot and Sébastien Regnault for providing advice on GENELAND use and the ΔK method for STRUCTURE, respectively. Part of this work was carried out using the resources of the Computational Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation. We thank three anonymous reviewers for their insightful suggestions. This work was supported by the National Science Foundation (project no. DEB-0316292).

References

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) GENETIX *4.05, Logiciel sous Windows pour la Génétique des Populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B: Biological Sciences*, **57**, 289–300.

Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? *Molecular Ecology*, **13**, 551–561.

Beyer HL (2004) Hawth's analysis tools for ARCGIS. Available at http://www.spatialecology.com/htools.

Breininger DR, Larson VL, Duncan BW *et al*. (1995) Landscape patterns of Florida scrub jay habitat use and demographic success. *Conservation Biology*, **9**, 1442–1453.

Breininger DR, Toland B, Oddy D *et al*. (2003) Biological criteria for the recovery of Florida scrub-jay populations on public lands in Brevard County and Indian River County. *Final Report to Endangered Species Office, US Fish and Wildlife Service, Jacksonville, Florida*. Dynamac Corporation, DYN-2, Kennedy Space Center, Florida.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Coulon A, Guillot G, Cosson J-F *et al*. (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, **15**, 1669–1679.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, **15**, 290–295.

Crochet P-A (1996) Can measures of gene flow help to evaluate dispersal? *Acta Oecologia*, **17**, 459–474.

Delaney KS, Wayne RK (2005) Adaptive units for conservation: population distinction and historic extinctions in the island scrub-jay. *Conservation Biology*, **19**, 523–533.

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

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

Fernald RT (1989) Coastal xeric scrub communities of the Treasure Coast Region, Florida: a summary of their distribution and ecology, with guidelines for their preservation and management. *Nongame Wildlife Program. Technical Report No. 6*. Florida Game and Fresh Water Fish Commission, Tallahassee, Florida.

Fitzpatrick JW, Woolfenden GE, Kopeny MT (1991) Ecology and development-related habitat requirements of the Florida scrub-jay (*Aphelocoma coerulescens coerulescens*). *Nongame Wildlife Program Technical Report No. 8*. Florida Game and Freshwater Fish Commission. Tallahassee, Florida.

Fitzpatrick JW, Woolfenden GE, Bowman R (1999) Dispersal distance and its demographic consequences in the Florida scrub-jay. In: *Proceedings of the XXII International Ornithological Congress* (eds Adams NJ, Slotow RH), Durban, South Africa. Symposium S41.2, 2465–2479.

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**, 2741–2752.

Green DM (2005) Designable units for status assessment of endangered species. *Conservation Biology*, **19**, 1813–1820.

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 program for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Hailer F, Gautschi B, Helander B (2005) Development and multiplex PCR amplification of novel microsatellite markers in the White-tailed Sea Eagle, *Haliaeetus albicilla* (Aves: Falconiformes, Accipitridae). *Molecular Ecology Notes*, **5**, 938–940.

Hoelzel AR (1992) *Molecular Genetic Analysis of Populations: A Practical Approach*. IRL Press at Oxford University Press, Oxford, New York.

Ihaka R, Gentleman R (1996) R. A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Jakobbson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program with label switching and multilmodality in analysis of population structure. *Bioinformatics*, **3**, 1801–1806.

Jakobsson M, Rosenberg NA (2007*) CLUMPP: Cluster Matching and Permutation Program*. Version 1.1 (user's manual). URL: http://rosenberglab.bioinformatics.med.umich.edu/software/clumpp_manual.pdf.

Koenig WD, Van Vuren D, Hooge PN (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology & Evolution*, **11**, 514–517.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OEJ (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295–302.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.

McDonald DB, Potts WK, Fitzpatrick JW, Woolfenden GE (1999) Contrasting genetic structures in sister species of North American scrub-jays. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 1117–1125.

Miller KE, Stith BM (2002) *Florida Scrub-Jay distribution and habitat in Charlotte County*. Final report, contract no. 2001000116: scrub-jay survey December 2002. U.S. Fish and Wildlife Service, Vero Beach, Florida.

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, **100**, 403–404.

Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematics Biology*, **51**, 238–254.

Myers RL, Ewel JJ (1990) *Ecosystems of Florida*. University of Central Florida Press, Orlando, Florida.

Nathan R (2001) The challenges of studying dispersal. *Trends in Ecology & Evolution*, **16**, 481–483.

Paetkau D (1999) Using genetics to identify intraspecific conservation units: a critique of current methods. *Conservation Biology*, **13**, 1507–1509.

Palsbøll PJ, Bérubé M, Allendorf FW (2007) Identification of management units using population genetic data. *Trends in Ecology & Evolution*, **22**, 11–16.

Pennock DS, Dimmick WW (1997) Critique of the evolutionary significant unit as a definition for 'distinct population segments' under the US Endangered Species Act. *Conservation Biology*, **11**, 611–619.

Pilot M, Jedrzejewski W, Branicki W *et al*. (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology*, **15**, 4533–4553.

Pranty B (1996) Distribution of the Florida scrub-jay, 1992–1993. *Report on cooperative agreement #14-16-0004-91-950*. US Fish and Wildlife Service, Jacksonville, Florida.

Pritchard JK, Wen W (2004) Documentation for STRUCTURE software: version 2. Available online at http://pritch.bsd.uchicago.edu/software/readme_structure2_1.pdf.

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

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

Stenzler LM, Fitzpatrick JW (2002) Isolation of microsatellite loci in the Florida scrub-jay *Aphelocoma coerulescens*. *Molecular Ecology Notes*, **2**, 547–550.

Stith BM (1999) Metapopulation viability analysis of the Florida scrub-jay (*Aphelocoma coerulescens*): a statewide assessment. *Final Report to the Endangered Species Office, US Fish and Wildlife Service, Jacksonville, Florida*. Contract No. 1448-40181-98-M324. August 1999.

Stith BM, Fitzpatrick JW, Woolfenden GE, Pranty B (1996) Classification and conservation of metapopulations: a case study of the Florida scrub-jay. In: *Metapopulations and Wildlife Conservation* (ed. McCullough DR), pp. 187–215. Island Press, Washington, D.C./Covelo, California.

Thompson PM, Goodman S (1997) Direct and indirect estimates of dispersal distances. *Trends in Ecology & Evolution*, **12**, 195–196.

Vandewoestijne S, Baguette M (2004) Demographic versus genetic dispersal measures. *Population Ecology*, **46**, 281–285.

Verhoeven KJF, Simonsen KL, Mcintyre LM (2005) Implementing false discovery rate control: increasing your power. *Oikos*, **108**, 643–647.

Waples RS (1998) Evolutionary significant units, distinct population segments, and the Endangered Species Act: reply to Pennock and Dimmick. *Conservation Biology*, **12**, 718–721.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1440.

Weekley CW, Menges ES, Pickert RL (in press) An ecological map of Florida's Lake Wales Ridge: a new boundary delineation and an assessment of post-Columbian habitat loss. *Florida Scientist*.

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

Woolfenden GE, Fitzpatrick JW (1984) *The Florida Scrub Jay – Demography of a Cooperative-Breeding Bird*. Princeton University Press, Princeton, New Jersey.

Woolfenden GE, Fitzpatrick JW (1996) Florida Scrub-Jay (*Aphelocoma coerulescens*). In: *Birds of North America* (eds Poole A, Gill F), pp. 1–28. The Academy of Natural Sciences, The American Ornithologists' Union, Philadelphia, Washington, D.C.

Aurélie Coulon investigates movement ecology. She is particularly interested in the influence of landscape structure on animal movements, and on their consequences on population structure. John Fitzpatrick works on population ecology and social evolution of cooperative breeding birds, especially Florida scrub-jays, and the conservation implications of fire-maintained ecosystems within the context of increasingly human dominated landscapes. Reed Bowman is interested in the demography and behavioural ecology of cooperative breeding birds. Brad Stith's primary research interests include individual-based modeling, population viability analysis, landscape ecology, and GIS applications in conservation, with an emphasis on endangered species. Chris Makarewich is a research technician particularly interested in conservation genetics and avian behavioural ecology. Laura Stenzler is broadly interested in the genetic consequences of habitat fragmentation, especially in birds. Irby Lovette directs the Fuller Evolutionary Program at the Cornell Lab of Ornithology, where he works primarily on avian evolutionary ecology and phylogenetics.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Results of STRUCTURE ΔK hierarchical analyses

Fig. S2 Mean ± standard deviation of the estimated ln probability of the data (lnP(D)) for the different values of *K* tested with STRUCTURE under the highest lnP(D) method

Fig. S3 Evolution of the quality of the runs as a function of their ranking

Table S1 Characteristics of the 20 microsatellite loci used in this study

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03705.x
(This link will take you to the article abstract).

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