Molecular Ecology (2007) **16**, 2031–2043 doi: 10.1111/j.1365-294X.2007.03293.x

A new individual-based spatial approach for identifying genetic discontinuities in natural populations

S. MANEL,* F. BERTHOUD,† E. BELLEMAIN,* M. GAUDEUL,* G. LUIKART,* J. E. SWENSON,‡ L. P. WAITS,§ P. TABERLET* and INTRABIODIV CONSORTIUM[1]

**Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, F-38041 Grenoble Cedex 9, France, †Laboratoire de Physique et Modélisation des Milieux Condensés, Maison des Magistères, CNRS, BP 166, F-38042 Grenoble Cedex, France, ‡Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Post Box 5003, NO-1432 Ås, Norway, and Norwegian Institute for Nature Research Tungasletta 2, NO-7485 Trondheim, Norway, §Department of Fish and Wildlife Resources, University of Idaho, Moscow, ID 83844-1136, USA*

Abstract

The population concept is central in evolutionary and conservation biology, but identifying the boundaries of natural populations is often challenging. Here, we present a new approach for assessing spatial genetic structure without the *a priori* assumptions on the locations of populations made by adopting an individual-centred approach. Our method is based on assignment tests applied in a moving window over an extensively sampled study area. For each individual, a spatially explicit probability surface is constructed, showing the estimated probability of finding its multilocus genotype across the landscape, and identifying putative migrants. Population boundaries are localized by estimating the mean slope of these probability surfaces over all individuals to identify areas with genetic discontinuities. The significance of the genetic discontinuities is assessed by permutation tests. This new approach has the potential to reveal cryptic population structure and to improve our ability to understand gene flow dynamics across landscapes. We illustrate our approach by simulations and by analysing two empirical datasets: microsatellite data of *Ursus arctos* in Scandinavia, and amplified fragment length polymorphism (AFLP) data of *Rhododendron ferrugineum* in the Alps.

Keywords: assignment test, genetic discontinuity, moving windows, multilocus genotype, spatial genetics

Received 24 September 2006; revision accepted 9 January 2007

Introduction

Delineating populations is a crucial first step for assessing evolutionary processes (e.g. local adaptation, gene flow) and for the preservation of biodiversity (e.g. identification of management units). The definition of a population remains an area of active discussion (Waples & Gaggiotti 2006). Here, we present a new approach that combines large molecular datasets, assignment tests and spatial moving window analysis to identify population boundaries.

Classical population genetic analyses consist of sampling groups of individuals from predefined populations, and then estimating allele frequencies and parameters, such as genetic distance and *F*-statistics (Weir & Cockerham 1984; Nei 1987). For more continuously distributed populations, individuals are at risk of being grouped somewhat arbitrarily using unproven criteria, such as habitat characteristics, morphological differences, geographical distance, or political boundaries (Pritchard *et al*. 2000). Also, classical population genetic analyses do not utilize the full discriminatory power of multilocus data (Waser & Strobek 1998).

Aspatial Bayesian clustering methods (Pritchard *et al*. 2000; Corander *et al*. 2003) have emerged as a popular new tool for defining populations using multilocus genotype data. When the genetic structure coincides with the geography, using spatial information might increase the power of detecting genetic discontinuities. However, using spatial methods to localize genetic discontinuities among individuals remains rare in the field of population genetics (e.g. Barbujani *et al*. 1989; Bocquet-Appel & Bacro 1994; Sokal

[1]Annexe 1

Correspondence: Dr Stéphanie Manel, Fax: +33 476 51 42 79; E-mail: stephanie.manel@ujf-grenoble.fr

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

& Thomson 1998; Manel *et al*. 2003; Coulon *et al*. 2006). Perhaps one of the major reasons is the lack of available datasets with both genetic and geographical coordinate information, but this limitation has now been removed. New spatial Bayesian clustering algorithms (e.g. Guillot *et al*. 2005; Corander *et al*. 2006; François *et al*. 2006) indirectly identify genetic boundaries at an individual level. An alternative and complementary approach is to look directly for zones of sharp change in genetic data (Legendre & Legendre 2006). Two main approaches are available to achieve this goal: the Monmonier algorithm (Monmonnier 1973; Manni *et al*. 2004; Miller 2005) based on genetic distance analysis and the 'wombling' method (Womble 1951; Barbujani *et al*. 1989) based on the analysis of allele frequencies.

Here, we introduce a new approach that uses individual genotypes and their geographical co-ordinates to identify populations. This new approach has the potential to reveal genetic discontinuities and identify migrants without prior assumptions about population boundaries. Our method does not group individuals *a priori* into perceived populations, but rather adopts a spatial approach by using a 'moving window' placed across points of a grid map to identify population boundaries from the inferred genetic discontinuities.

We illustrate and validate this new approach using simulations and two empirical datasets from two types of molecular marker systems. Our first dataset encompasses 18 microsatellites loci from 964 brown bear (*Ursus arctos*) samples collected across Scandinavia. The second dataset contains 382 samples of the shrub *Rhododendron ferrugineum* sampled over the entire European Alps, and genotyped at 123 polymorphic AFLP markers.

Materials and methods

Description of the method

Our approach applies an assignment test locally within a moving window over the entire sampled area (Paetkau *et al*. 1995; Rannala & Mountain 1997). Thus, a probability map is constructed for each individual, showing the estimated probability of finding its multilocus genotype at each point of a grid across the landscape. The local variation of probability between adjacent points is used to calculate a slope at each point of the grid across the landscape. Each map reveals the most likely region of origin of the individual (i.e. area with the highest probabilities), and also highlights areas where the probability rapidly declines (greatest slope). The mean slopes calculated for all the individual probability maps are used to identify population genetic discontinuities and thus putative population boundaries.

Data analysis consisted of three main steps. First, for each individual, we constructed a map showing the probability of finding its multilocus genotype at each of many points across the landscape (Fig. 1). Points were distributed on a grid with a distance d between adjacent points. The algorithm consisted of: (i) removing the individual to be assigned; (ii) computing allele frequencies using individuals within a window of radius R around each point of the grid; (iii) computing the logarithm of the probability (log-likelihood) that the individual multilocus genotype occurs in the area around each point of the grid, based on the Bayesian assignment method of Rannala & Mountain (1997); and (iv) graphically representing the log-likelihood values in each point. The graphical representation consists of a circle, whose size corresponds to the magnitude of the log-likelihood. To avoid imprecise estimates of allele frequencies, we considered only windows with more than n_{min} individuals (n_{min} = 20 for the simulations and the brown bear dataset; n_{min} = 10 for the *R. ferrugineum* dataset). Missing data are allowed in the model by omitting missing loci when calculating individual probabilities (Pritchard *et al*. 2000). The appropriate values of d and R have to be determined by the user and will vary by species depending on density, home range size and dispersal ability.

Second, we identified the geographical zones of highest genetic change (discontinuity) for all individuals. Such discontinuities represent putative population boundaries. For each individual and for each point of the grid, we estimated the slope of the likelihood function. For each point of the grid, the slope was estimated as the mean absolute value of the difference between the likelihood value at the point considered and the eight adjacent points (if available). This resulted in a map of slopes of the likelihood function for each individual. In the final step, all individual slope maps were combined to produce the final map, where the slopes values at each point of the grid corresponds to the mean slopes values estimated over all individuals.

The significance of the mean slope values was estimated through a randomization test (Sokal & Rohlf 1981; Monte Carlo procedure). Null distributions of mean slopes were constructed for each point of the grid. Individual locations were randomised with respect to individual genotypes. Mean slopes for each point of the grid were estimated for each randomised data set. On the basis of the resulting distribution of mean slopes for each point of the grid, we decide whether the mean slope observed for the analysed data deviated less or more than the distribution at a level α (e.g. $\alpha = 0.05$).

The probability maps, the identification of genetic discontinuities and their significance were conducted using a C++ computer program (available upon request) and the GIS software ARCVIEW (version 3.1). A fully implemented, user-friendly program is in development.

Simulations

Microsatellite genotypes were simulated with the program EASYPOP (Balloux 2001) using the following parameters:

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

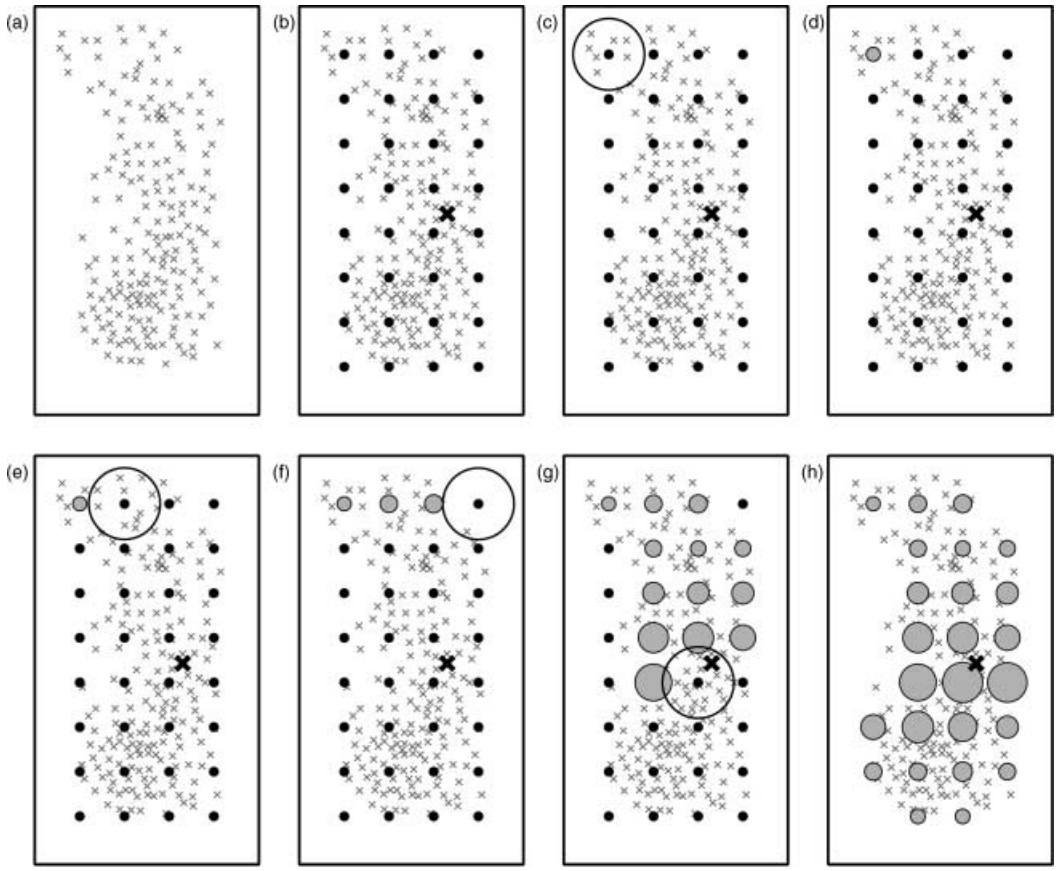

Fig. 1 Steps in the method used to generate one individual probability map. (a) Geographical distribution of the samples (small black crosses). (b) A grid (black dots) covers the sampling area, and the first individual to be analysed is chosen (large black cross). (c) Allele frequencies are estimated within the window (black circle) centred around the first point of the grid. (d) The log-likelihood of the chosen individual is estimated, and represented on the map by grey circles scaled to the size of the likelihood value. (e) The window moves to the next point of the grid. (f) (g) (h) log-likelihoods are estimated and represented on the map for the next points of the grid. If the moving window does not encompass enough individuals (n = 20) to precisely estimate allele frequencies, the log-likelihood is not estimated for that grid point.

random mating, same number of males and females in each population, same migration scheme, low migration rate (0.00001), free recombination between loci and KAM mutation model. Individual spatial coordinates were independently generated using the software R (R Development Core Team 2004) with the objective of creating three scenarios differing in spatial distribution and genetic structure (Fig. 2). Each scenario was replicated 10 times, but only one repetition is shown. First, one population (1000 individuals and 20 loci) was generated with a spatially continuous distribution and no genetic boundary (Fig. 2a). Second, the same population was then distributed into two spatially separated subpopulations but with no genetic boundary (Fig. 2b). Finally, two genetically differentiated populations (300 individuals each, 20 loci, F_{ST} = 0.1) were generated and uniformly distributed in space (Fig. 2c). The uniform distributions were chosen in order to obtain two populations with similar mean x-values, but different mean y-values.

Brown bears (Ursus arctos) in Scandinavia

We used a dataset consisting of 18 microsatellite loci genotyped for 964 brown bears distributed over *c*. 200 000 km^2 in Sweden (Waits *et al*. 2000; Bellemain 2004) (Fig. 3a). The 18 microsatellite loci (G1A, G1D, G10B, G10C, G10L, G10P, G10X, G10H, G10O, G10J, Mu05, Mu10, Mu15, Mu23, Mu50, Mu51, Mu59, Mu61) were described in Waits *et al*. (2000). The error rate per locus for this dataset was estimated to be 0.008 (Bonin *et al*. 2004).

We analysed three datasets: the entire dataset comprising the 964 individuals, the 517 males and the 447 females. The parameters d and R were set at 50 and 120 km, respectively. Only windows containing at least 20 individuals were considered. The value of d was chosen in relation to the size and the shape of the sampled area and the density of the sampling. The value of R was chosen to insure that each moving window would contain enough individuals to obtain reliable estimate allele frequencies. This value of

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

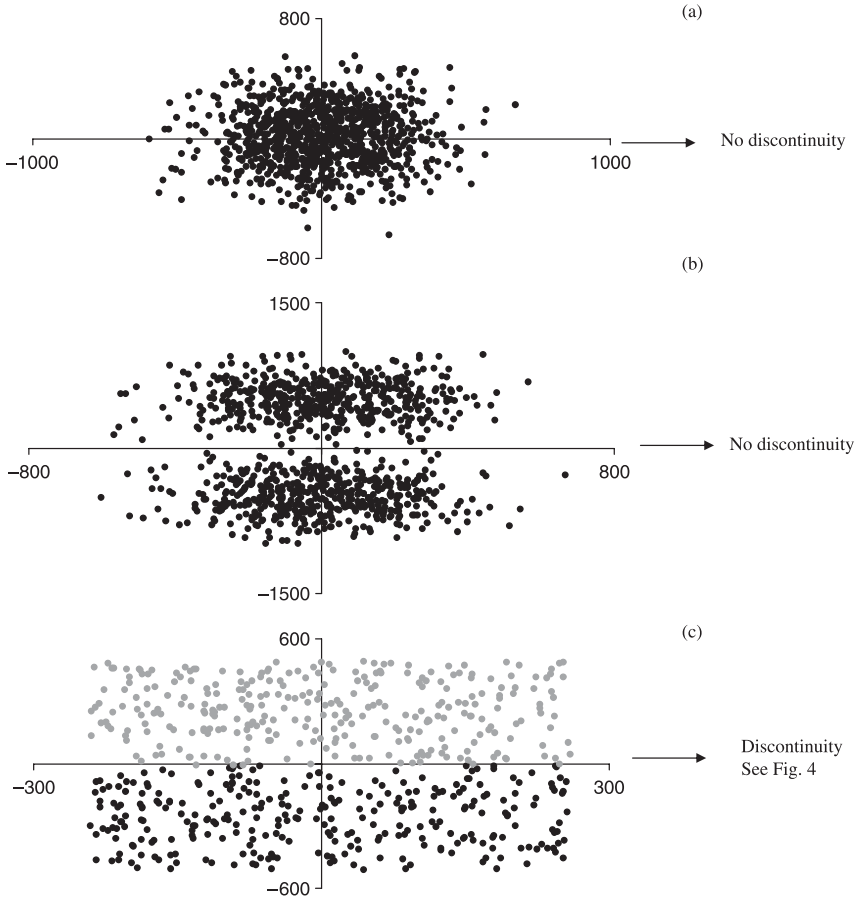

Fig. 2 Spatial representation of the three simulated scenarios. (a) One population uniformly distributed; 1000 individuals, 20 loci. (b) One population distributed uniformly in two separate patches; 1000 individuals, 20 loci. (c) Two genetically differentiated populations (20 loci, $F_{ST} = 0.1$) uniformly distributed. The uniform distributions were chosen in order to obtain two populations with similar mean x-values, but different mean y-values. Population one: black dots; population two: grey dots. The values on the axes follow an arbitrary coordinate system.

R produced 100 windows with a minimum of 20 individuals where the slope is calculated.

A subset of these data (366 bears) was previously analysed using a classical approach (F_{ST} and assignment tests), grouping individuals into four putative populations corresponding to areas with a high density of females (Waits *et al*. 2000). A second study based on the analysis of the same 366 multilocus genotypes, but using a Bayesian clustering approach without *a priori* population definition (Pritchard *et al*. 2000), identified only three main genetic groups (Manel *et al*. 2004).

Rhododendron ferrugineum in the Alps

Leaf samples of *R. ferrugineum* were collected during summer 2004 across the entire European Alps (latitude: 44°48′ to 48°36′; longitude: 5°20′ to 15°40′). A 12′ latitude × 20′ longitude (*c*. 23 km × 25 km) grid was adopted and three plants were sampled in every other cell (Fig. 3b), resulting in a total of 381 samples (127 cells) distributed over *c*. 171 350 km². Because vegetative growth is extensive in *R. ferrugineum*, sampled plants were chosen at least 10 m apart to avoid sampling several ramets of a single genet. Samples were immediately dried in silica gel and preserved at room temperature. AFLP data were generated using a protocol inspired from Vos *et al*. (1995). After electrophoresis on an ABI 3100 automated sequencer, 123 polymorphic AFLP markers were manually scored as present/absent using the GENOGRAPHER software (http://hordeum.oscs.montana.edu/genographer/). An error rate per locus of 0.013 was estimated by duplicating analyses for 46 random samples.

Binary AFLP data were then recoded as suggested by Evanno *et al*. (2005). Because dominant homozygotes and heterozygotes cannot be distinguished, marker presence was recoded as 1 –9 and the absence as 0 0, where 1 indicates presence of the band, –9 indicates a missing data and 0 absence of the band. *d* and *R* were set at 25 and 50 km, respectively. For the AFLP dataset, we used only windows containing at least 10 individuals. The value of *d* (25 km) is directly related to the size of the sampling grid (23km × 25km). The value of *R* (50 km) was chosen to insure that enough individuals (at least 10) are in the windows to calculate allele frequencies.

Results

Visualization of the results on maps

Our new method produces two types of maps. The first map represents the log-likelihood of observing an individual's

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

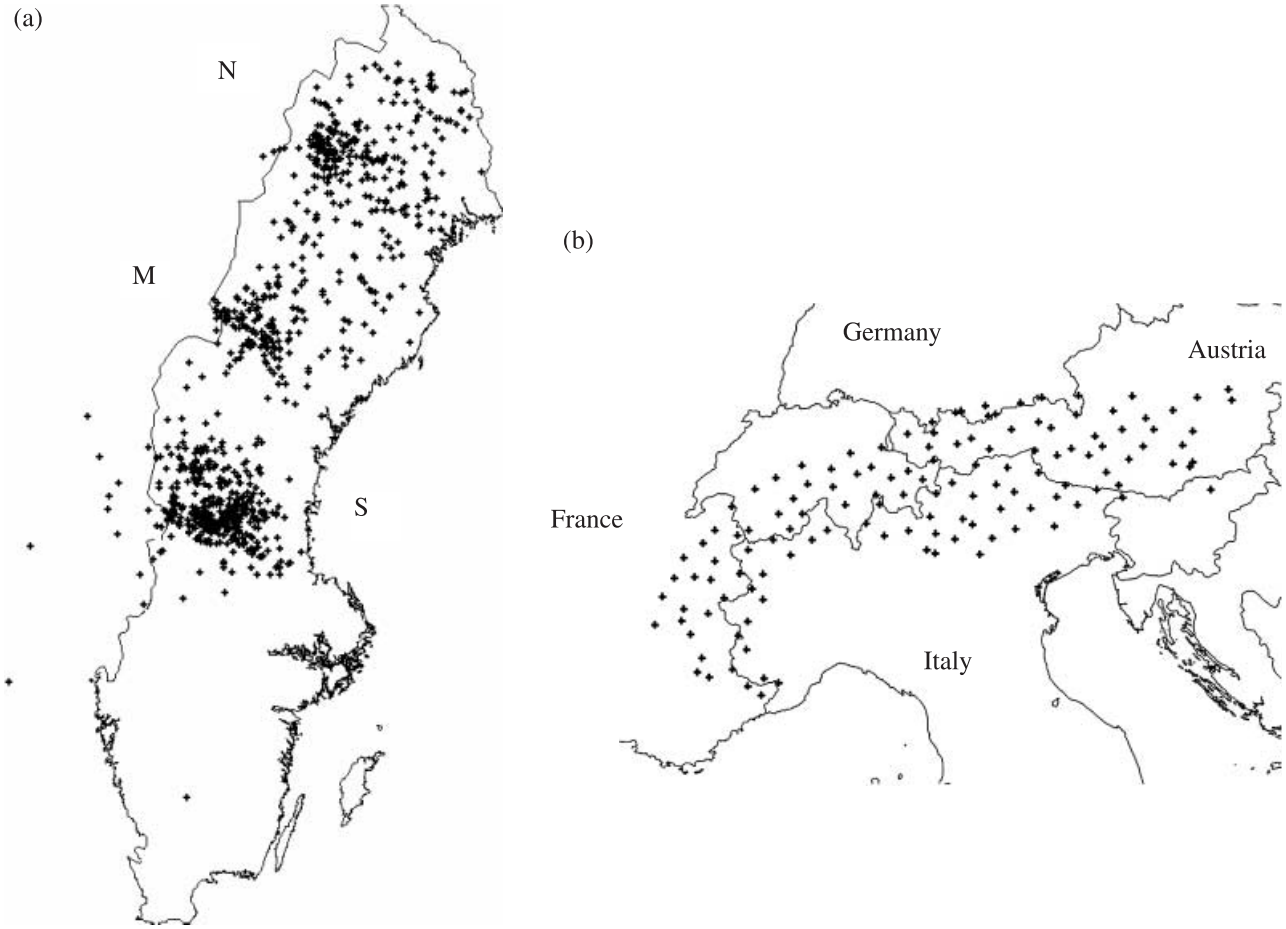

Fig. 3 Sampling maps. (a) *U. arctos* in Scandinavia (n = 964). Letters N, M and S indicate subpopulation regions identified by previous analysis (Manel *et al*. 2004). One cross represents one individual sampled. (b) *R. ferrugineum* in the Alps (n = 381). One cross represents three individuals sampled in the same locality.

multilocus genotype at each point of the grid. The second type of map is a single map of the mean slopes of the log-likelihood surfaces calculated over all individuals. For visualization purposes, we have chosen to represent the value of each parameter (log-likelihood and slopes) calculated at each point of the grid with circles of varying sizes. For the slopes, two alternative representations can be used, either for highlighting the discontinuities by representing the highest slopes with large circles (see, e.g. Figure 4 representing the results of the simulations), or for highlighting the populations themselves by representing the highest slopes with a small circles (see, e.g. Figures 5 and 7 representing the results of the empirical datasets).

Simulations

As expected, no genetic discontinuity was identified when we simulated a single population regardless of its spatial distribution (continuous or two patches; not shown). In the case of two simulated populations with F_{ST} = 0.1 and the distance between each point of the grid d = 40 m, Fig. 4 shows the significant genetic discontinuities (α = 0.05) according to different moving windows sizes (R = 80 m; R = 120 m; R = 200 m). The moving windows with less than 20 individuals (n_{min} = 20) were not considered. With R = 80 m, many points of the grid were not considered because the number of individuals within the moving windows were lower than n_{min} (Fig. 4a). As a consequence, the above combination of the parameters (d = 40; R = 80; n_{min} = 20) does not allow the identification of the genetic discontinuity. The analysis can be improved either by increasing R, or by decreasing n_{min}. With R = 120, the genetic discontinuity is properly identified (Fig. 4b), even with higher stringency (79, 75 and 68 slopes were significant with α = 0.05, 0.01 and 0.001, respectively). With R = 200, the genetic discontinuity is also identified, but with less spatial precision, i.e. the significant slopes cover a larger geographical area (Fig. 4c).

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

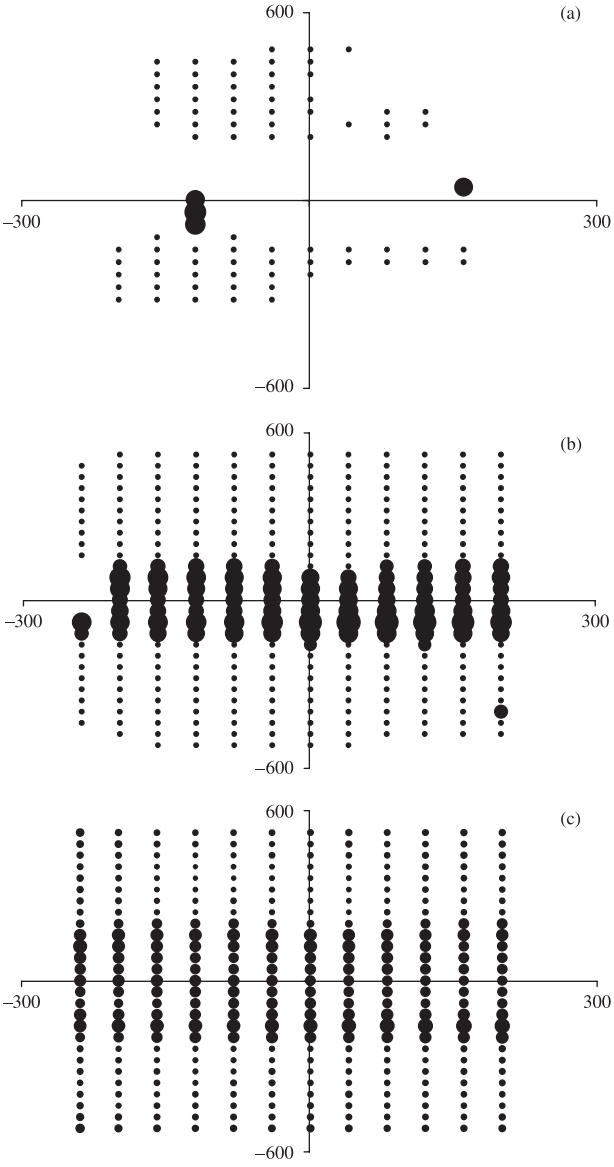

Fig. 4 Results of the third simulation scenario (described in Fig. 2c). Small circles indicate points of the grid taken into account in the analysis according to the chosen parameters. Large circles correspond to significant slopes at level $\alpha = 0.05$ and identify the genetic discontinuity between the two populations. Values on the axes are in meters. (a) $d = 40$, $R = 80$; these parameters do not allow the identification of the discontinuity. (b) $d = 40$, $R = 120$; these parameters allow a precise identification of the discontinuity. (c) $d = 40$, $R = 200$; these parameters also allow the identification of the discontinuity but with less spatial precision.

Scandinavian brown bear dataset. Three homogeneous regions (i.e. putative populations) were found (Fig. 5a), corresponding to the S, M and N regions (Fig. 3a) already identified by Manel *et al*. (2004). Nevertheless, permutation tests indicated that almost all slopes (97%) were significant at the 1% level, including small values in regions that we considered homogeneous (large grey circles, Fig. 5a). The strong population genetic discontinuity detected between the M and S regions (Fig. 3a) also corresponds to a female gene flow discontinuity detected using mitochondrial DNA data (Taberlet *et al*. 1995). A second discontinuity was detected between the M and N regions.

When the sexes were separated, the global genetic structures did not differ significantly (Fig. 5b, c). However, the mean slope values were significantly lower for male than for female samples (student *t*-test, $t = -2.05$, $P < 0.05$), indicating a stronger genetic structure for females than for males.

Using individual probability maps, it is also possible to determine whether each individual is a resident or a putative migrant. For example, a resident is an individual sampled in the region where its multilocus genotype is most likely to originate (Fig. 6a). In contrast, a migrant is an individual sampled in a region where its multiloci genotype is unlikely (Fig. 6b).

Rhododendron dataset. Figure 7 shows the existence of two strong genetic discontinuities isolating three putative populations. Two of these populations — one in the eastern Alps and one in the central Alps — were very homogeneous. The third, located in the southwestern Alps, was less homogeneous.

Discussion

Our new method for spatially locating genetic discontinuities based on local assignment tests applied in a moving window performed well in all evaluations. Using the simulated datasets, the method did not reveal a genetic discontinuity when applied to a single population, regardless of the geographical distribution of the samples (i.e. one vs. two patches). In the simulation scenario with two moderately differentiated populations ($F_{ST} = 0.1$), the spatial location of the genetic discontinuity was correctly identified. In addition, the method performed well on two empirical datasets using two different types of markers (microsatellite vs. AFLP).

Interpretation of the results on the two empirical datasets

Scandinavian brown bear dataset. The three populations identified by our new method are consistent with the known demographic history of this species and previous genetic analyses using other genetic analysis approaches (Waits *et al*. 2000; Manel *et al*. 2004). But the results obtained in this study add new information about the exact spatial location of the boundaries and their significance. To evaluate the ability of our method to detect different levels of population structure, we calculated F_{ST} values for the three identified populations (Fig. 5a). The lowest value, calculated between M and N areas, was 0.05 and demonstrated that

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

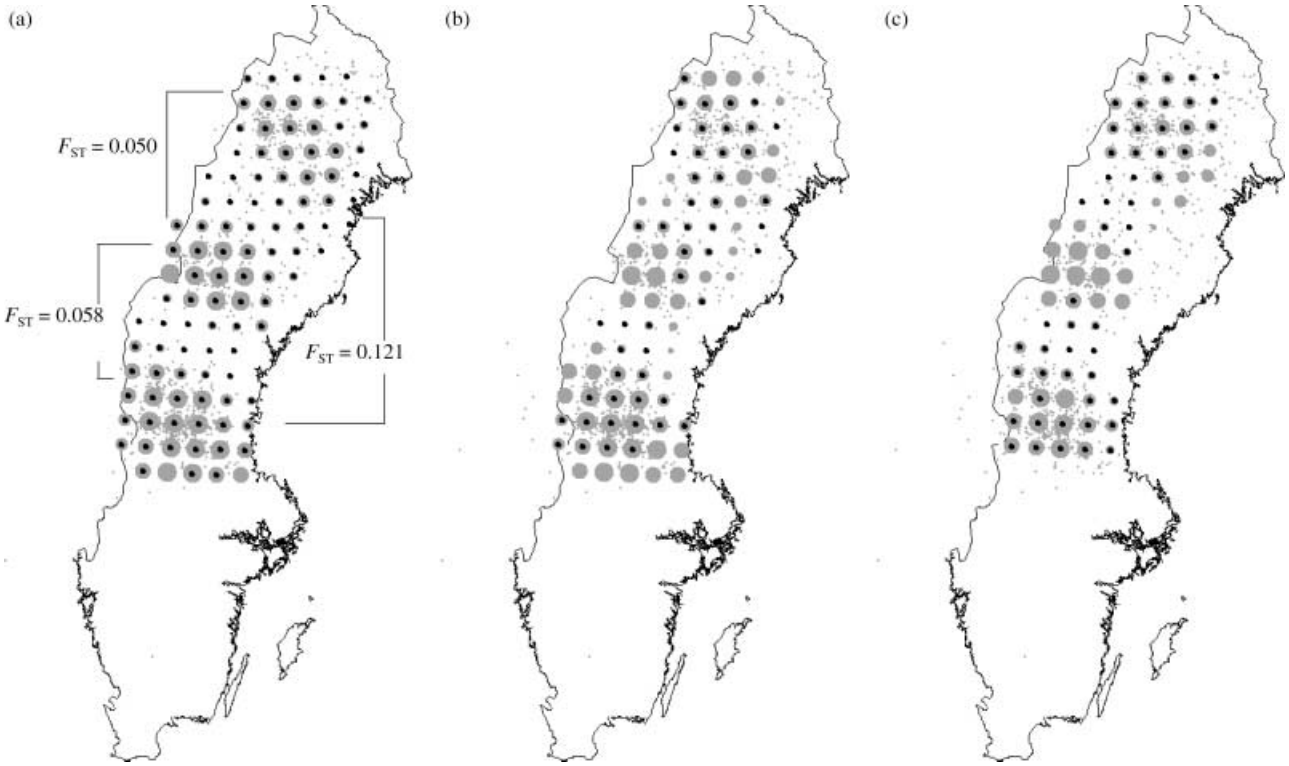

Fig. 5 Map showing the three putative populations N, M, and S (large grey circles) and genetic discontinuities (small grey circles) identified for the Scandinavian brown bears. In this case, we highlight the populations, not the discontinuities as in the Fig. 4, by large grey circles characterizing small slope values (homogeneous regions) and small grey circles large slope values (genetic discontinuities). Small grey crosses (+) represent sampled individuals. (a) 964 individuals. (b) 445 females. (c) 512 males. Overlaid small black circles indicate the slopes that are significant ($\alpha = 0.05$). When the slope is very high, only the black circle (indicating significance of the slope) is visible.

the approach was able to detect genetic discontinuities even at relatively low levels of divergence.

The presence of significant slopes within areas inferred to be homogeneous (i.e. populations) can be explained by the known matriarchal structure of bears, with females associating spatially with close relatives (Støen *et al*. 2005). This means that when moving a window with a radius of 120 km, the new individuals integrated in the window differ significantly from the removed individuals. The fact that male bears appeared to be less genetically structured than female bears is in accordance with behavioural data, i.e. males are the dispersing sex (McLellan & Hovey 2001; Støen *et al*. 2006). This method is also useful for identifying migrant individuals. For example, a male sampled in the south was identified as a migrant from the northernmost population (Fig. 6b). Mitochondrial DNA analysis confirmed this finding, since this bear has a mitochondrial DNA sequence from the eastern lineage which is only observed in Scandinavian from the M and N populations (Taberlet *et al*. 1995).

Rhododendron dataset. The classification of rhododendron samples into three populations agrees with the structure identified with the methods of Pritchard *et al*. (2000) (data not shown). We hypothesize that the genetic discontinuities in this species correspond to contact zones among populations that were previously isolated in different refugia during Quaternary cold periods (Schönswetter *et al*. 2005). Estimation of F_{ST} values between the three populations revealed moderate to high levels of genetic differentiation with the lowest F_{ST} (0.16) between the eastern and central regions.

Advantages and limitations of the method

Our new approach is simple in practice and based on a well-known method, assignment tests (Paetkau *et al*. 1995; Rannala & Mountain 1997), that is widely used and validated (Cornuet *et al*. 1999; Peatkau *et al*. 2004). Another advantage of the approach is that the output allows easy visual interpretation of the spatial location of genetic discontinuities. This individual-based approach can be applied across a wide range of geographical scales, from within local populations to the continent level, and to any organism showing genetic variation. Interestingly, there has been much controversy in projects such as the human

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

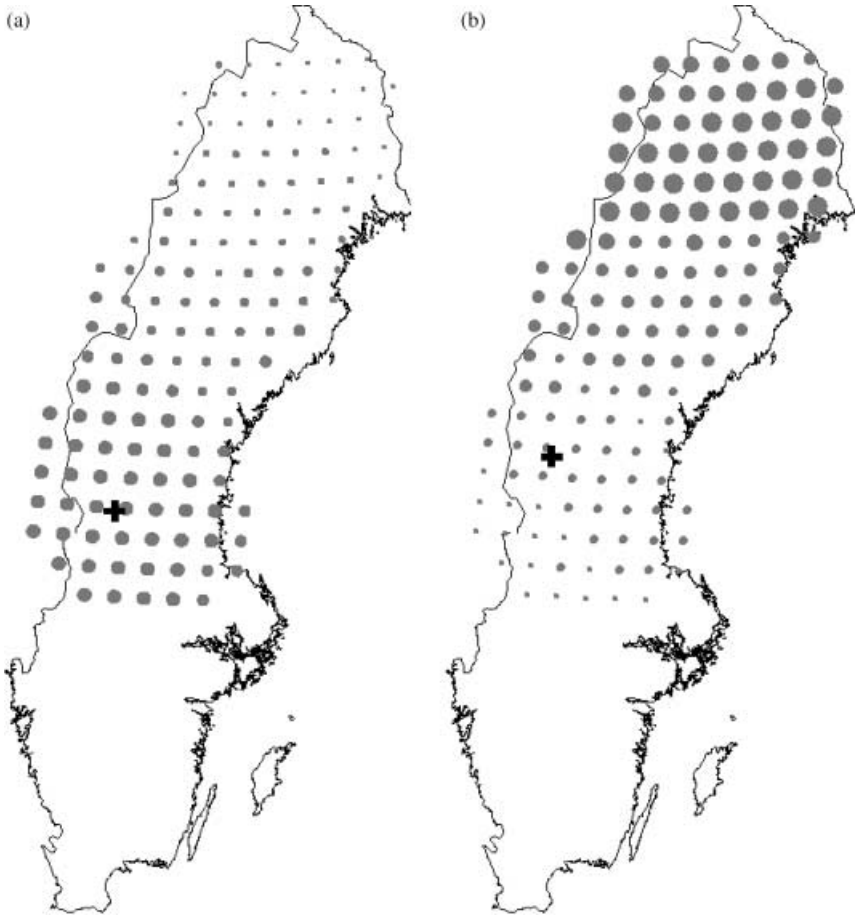

Fig. 6 Example of likelihood maps for brown bears. Large circles indicate a higher probability of finding the particular multilocus genotype. The cross represents the sampling location. (a) An individual assigned to its area of sampling origin (area M). (b) A migrant (from area N to area S).

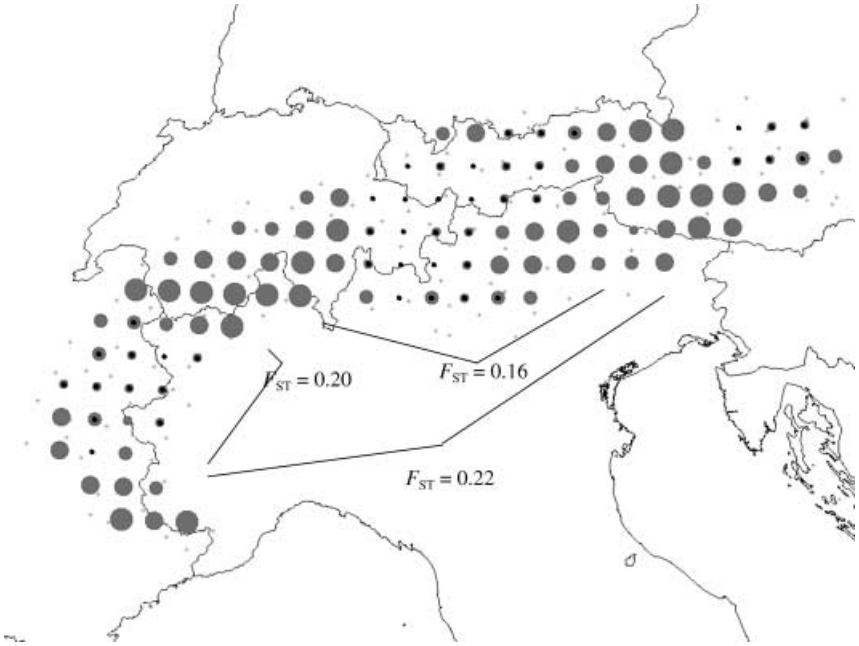

Fig. 7 Map showing the three putative populations for *R. ferrugineum* in the Alps. Large grey circles characterize small slope values (homogeneous regions) and small grey circles large slope values (genetic discontinuities). Overlaid small black circles indicate the slopes that are significant ($\alpha = 0.05$). Small crosses (+) represent sampled individuals. When the slope is very high, only the black circle (indicating significance of the slope) is visible.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

genome diversity project, as to which sampling strategy to adopt: a systematic sampling using a grid or a sampling only of populations identified *a priori* (Cavalli-Sforza *et al*. 1991; Greely 2001). Our results illustrate that a systematic sampling on a grid or evenly spaced sampling is efficient for identifying spatial patterns of genetic diversity. This approach will be especially useful when studying large continuous populations with no obvious boundaries or with cryptic barriers to gene flow. However it requires a continuous, evenly spaced sampling and multiple genetic markers to provide enough resolution to assign individuals to a sampling region with confidence. The importance of continuous, evenly spaced sampling in landscape studies that seek to detect the spatial location of population boundaries was also noted by Guillot *et al*. (2005).

The results of our method were mainly influenced by the size of the windows, *R*. For a fixed value of *d*, increasing *R* decreases the precision of the localization of the boundary: for example if $d = 40$ m and $R = 240$ m, the boundary would be identified at a precision of ± 260 m ($R + d/2$). Thus researchers will want to use the smallest possible value of *R*. However, as *R* decreases it becomes more difficult to maintain enough individuals in a window for accurately estimating allele frequencies. In the case of *R. ferrugineum*, the low minimum number of individuals per windows (10) was compensated by a relatively high number of markers (i.e. 123). This choice influences the accuracy of the location and the time of computing. Implementation of the method requires finding the optimal balance between *R (*radius of the windows) and *d* (distance between adjacent points of the grid) to identify zones of sharp discontinuities with meaningful geographical precision. Nevertheless, the method is robust when using different values for the moving window and grid sizes, because it always locates the discontinuities (i.e. the discontinuities are always at the same place, but are detected less accurately). Finally, the choice of *R* and *d* will depend on the study organism and the density of sampling and requires multiple iterations of attempts.

In summary, our method first produces likelihood maps for each individual. These maps allow the identification of migrants, corresponding to the individuals that were sampled far from the area where their multilocus genotypes are more likely. For example, the examination of the distribution of geographical distances between individuals and their 'windows' with largest assignment probabilities should provide valuable data about the dispersion distances. Second, the approach can identify the geographical location of genetic discontinuities from multilocus genotypes and individual geographical locations and adds a new tool for addressing the first step of landscape genetics (Manel *et al*. 2003). It can be used to detect conservation units by identifying cryptic genetic boundaries when individuals are continuously distributed across space. Its main advantages come from its ease of implementation and from the simplicity of the theory behind assignment tests. It is less sophisticated than the recent spatial Bayesian clustering method introduced by Guillot *et al*. (2005) and Corander *et al*. (2006) where the spatial organization of populations is modelled through the coloured Voronoi tesselation. Our new approach and the two previous methods differ from Wombling (Womble 1951) in that they use multilocus data instead of analysing each allele separately. They also differ from Monmonier algorithm (Monmonier 1973), because this is a distance-based method that is generally applied at the population level. These methods have never been thoroughly compared, and the next step will therefore be to assess their relative performance in identifying genetic discontinuities and population boundaries in a spatial context.

Once a genetic discontinuity has been detected, an exciting new research direction will be to determine the underlying explanatory variables by overlaying spatial maps of genetic variation onto maps of environmental variables and exploiting spatial statistics and GIS (e.g. Manel *et al*. 2003; Spear *et al*. 2005). Such approaches have many applications in expanding our fundamental knowledge of ecology, evolution, behaviour and landscape level processes. There is clearly great potential in this analytical approach and much need for further development and validation of individual-based geographical approaches in population genetics.

Acknowledgements

We thank J. Hogg and M. Schwartz for helpful discussions and comments on the manuscript. This study was supported by the University Joseph Fourier (Grenoble), the Centre National de la Recherche Scientifique, the Swedish Environmental Protection Agency, the Norwegian Directorate for Nature Management, the Research Council of Norway, the Swedish Association for Hunting and Wildlife Management, and the WWF-Sweden. Most of the computations were performed at the LPM2C.

References

Balloux F (2001) EASYPOP (version 1.7) a computer program for the simulation of population genetics. *Journal of Heredity*, **92**, 301–302.

Barbujani G, Oden NL, Sokal R (1989) Detecting regions of abrupt change in maps of biological variables. *Systematic Zoology*, **38**, 376–389.

Bellemain E (2004) Genetics of the Scandinavian Brown Bear: Implications for Biology and Conservation. PhD Dissertation, Agricultural University of Norway, Ås, Norway.

Bocquet-Appel JP, Bacro JN (1994) Generalized wombling. *Systematic Biology*, **43**, 442–448.

Bonin A, Bellemain E, Eidesen PB *et al*. (2004) How to track and assess genotyping errors in population genetic studies. *Molecular Ecology*, **13**, 3261–3273.

Cavalli-Sforza LL, Wilson AC, Cantor CR *et al*. (1991) Call for a world-wide survey of human genetic diversity: a vanishing

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

opportunity for the human genome project. *Genomics*, **11**, 490–491.

Corander J, Waldmann P, Sillanpa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163**, 367–374.

Corander J, Marttinen P, Sirén J, Tang J (2006) *BAPS: Bayesian Analysis of Population Structure, Manual v, 4.1*. Available at: http://www.rni.helsinki.fi/ ~jic/bapspage.html.

Cornuet J-M, Piry S, Luikart G *et al.* (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**, 1989–2000.

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

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.

François O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics*, **4**, 805–816.

Greely HT (2001) Human genome diversity: what about the other human genome project? *Nature Reviews Genetics*, **2**, 222–227.

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

Legendre P, Legendre L (2006) *Numerical Ecology*. Elsevier, Amsterdam, p. 853.

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

Manel S, Bellemain E, Swenson J, François O (2004) Assumed and inferred spatial structure of populations: the Scandinavian brown bears revisited. *Molecular Ecology*, **13**, 1227–1331.

Manni F, Guerard E, Heyer E (2004) Geographic patterns of genetic, morphologic and linguistic variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**, 173–190.

McLellan BN, Hovey FW (2001) Natal dispersal of grizzly bears. *Canadian Journal of Zoology*, **79**, 838–844.

Miller MP (2005) ALLELES IN SPACE (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, **96**, 722–724.

Monmonnier M (1973) Maximum–difference barriers: an alternative numerical regionalization method. *Geographical Analysis*, **3**, 245–261.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, **4**, 347–354.

Peatkau D, Slades R, Burdens M, Arnaud E (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

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

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA*, **94**, 9197–9201.

Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology*, **14**, 3547–3555.

Sokal RR, Rohlf FJ (1981) *Biometry. The Principles and Practice of Statistics in Biological Research*. Freeman WH and Company, San Fransisco.

Sokal RR, Thomson BA (1998) Spatial genetic structure of human populations in Japan. *Human Biology*, **70**, 1–22.

Spear SF, Peterson CR, Matocq MD, Storfer A (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology*, **14**, 2553–2564.

Støen OG, Bellemain E, Sæbø S, Swenson JE (2005) Kin-related spatial structure in brown bears *Ursus arctos*. *Behavioural Ecology and Sociobiology*, **59**, 191–197.

Støen O-G, Zedrosser A, Sæbo S, Swenson JE (2006) Natal dispersal in expanding brown bear *Ursus arctos* populations. *Oecologia DOI 1007*.

Taberlet P, Swenson JE, Sandegren F, Bjärvall A (1995) Localization of a contact zone between two highly divergent mitochondrial DNA lineages of the brown bear (*Ursus arctos*) in Scandinavia. *Conservation Biology*, **9**, 1255–1261.

Vos P, Hagers R, Bleeker M *et al.* (1995) AFLP: new technique for DNA fingerprinting. *Nucleic Acid Research*, **23**, 4407–4414.

Waits LP, Taberlet P, Swenson JE, Sandegren F, Franzén R (2000) Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Molecular Ecology*, **9**, 421–431.

Waples R, 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**, 1419–1439.

Waser PM, Strobek C (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution*, **13**, 43–44.

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

Womble WH (1951) Differential systematics. *Science*, **114**, 315–322.

This work has been done in collaboration with different researchers. Stéphanie Manel and Lisette Waits are interested in developing spatial analysis in population genetics (i.e. landscape genetics). Françoise Berthoud is a computer scientist that helped in the development of the C++ program. Pierre Taberlet and Gordon Luikart are interested in population genetics and molecular tools. Pierre Taberlet is the leader of the European project that provided the Rhododendron datasets. Jon Swenson is the scientific leader of the Scandinavian brown bear research project. Eva Bellermain is a postdoctoral researcher interested in population genetics and evolution. Myriam Gaudeul is interested in various aspects of plant evolutionary biology, including biogeography and population genetics.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Annexe 1

IntraBioDiv Consortium

The IntraBioDiv Consortium is composed of members of the IntraBioDiv project, as well as additional scientists, botanical experts, and technical assistants who participated to this project in relation with the official contractors and subcontractors.

The project has been supported financially by the European Commission, under the Six Framework Programme.
Acronym: IntraBioDiv
Title: Tracking surrogates for intraspecific biodiversity: towards efficient selection strategies for the conservation of natural genetic resources using comparative mapping and modelling approaches.
Reference: GOCE-CT-2003-505376
Duration: 1st January 2004–31st December 2006
Coordinator: Pierre Taberlet

Scientists

Wolfgang AHLMER[6]
Paolo AJMONE MARSAN[2]
Nadir ALVAREZ[3]
Enzo BONA[9]
Maurizio BOVIO[9]
El¿bieta CIESLAK[12]
Gheorghe COLDEA[11]
Licia COLLI[2]
Vasile CRISTEA[11]
Jean-Pierre DALMAS[8]
Thorsten ENGLISCH[4]
Luc GARRAUD[8]
Myriam GAUDEUL[1]
Ludovic GIELLY[1]
Felix GUGERLI[5]
Walter GUTERMANN[4]
Rolf HOLDEREGGER[5]
Nejc JOGAN[7]
Alexander A. KAGALO[18]
Gra¿yna KORBECKA[12]
Philippe KÜPFER[3]
Benoît LEQUETTE[14]
Dominik Roman LETZ[10]
Stéphanie MANEL[1]
Guilhem MANSION[3]
Karol MARHOLD[10]
Fabrizio MARTINI[9]
Zbigniew MIREK[12]
Riccardo NEGRINI[2]
Harald NIKLFELD[4]
Fernando NIÑO[13]
Massimiliano PATRINI[2]
Ovidiu PAUN[4]
Marco PELLECCHIA[2]
Giovanni PERICO[9]
Halina PIEKOS-MIRKOWA[12]
Peter POSCHLOD[6]
Filippo PROSSER[16]
Mihai PUSCAS[11]
Micha[3] RONIKIER[12]
Patrizia ROSSI[15]
Martin SCHEUERER[6]
Gerald SCHNEEWEISS[4]
Peter SCHÖNSWETTER[4]
Luise SCHRATT-EHRENDORFER[4]
Fanny SCHÜPFER[3]
Alberto SELVAGGI[17]
Katherina STEINMANN[5]
Pierre TABERLET[1]
Conny THIEL[5]
Andreas TRIBSCH[4]
Marcela VAN LOO[4]
Manuela WINKLER[4]
Thomas WOHLGEMUTH[5]
Tone WRABER[7]
Niklaus ZIMMERMANN[5]

Botanical experts

Gabriel ALZIAR[14]
Carlo ARGENTI[16]
Tinka BACIC[7]
Jean-Eric BERTHOUSE[14]
Alessio BERTOLLI[16]
Enrico BRESSAN[9]
François BRETON[14]
Massimo BUCCHERI[9]
Sonia D'ANDREA[17]
Sergio DANIELI[9]
Rosanna DE MATTEI[17]
Thierry DELAHAYE[8]
Roberto DELLA VEDOVA[17]
Cédric DENTANT[8]
Alessandra DI TURI[17]
Wolfgang DIEWALD[6]
Rolland DOUZET[1]
Constantin DRAGULESCU[11]
Philippe DRUART[3]
Siegrun ERTL[4]
Delphine FALLOUR-RUBIO[14]
Gino FANTINI[9]
Paolo FANTINI[15]
Germano FEDERICI[9]
Franco FENAROLI[9]
Viera FERÁKOVÁ[10]
Roberto FERRANTI[9]
Francesco FESTI1[6]
Bozo FRAJMAN[7]
Jean-Claude GACHET[14]
Bruno GALLINO[17]

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Federica GIRONI[9]
Gheorghe GROZA[11]
Andreas HILPOLD[19]
Catherine JOLLIBERT[14]
Denis JORDAN[8]
Thomas KIEBACHER[19]
Michael KLEIH[9]
Michel LAMBERTIN[14]
Cesare LASEN[9]
Petra MAIR[19]
Luca MANGILI[9]
Diego MARANGONI[17]
Carlo & Marisa MARCONI[9]
Hugues MERLE[8]
Marco MERSCHEL[6]
Henri MICHAUD[8]
Luca MISERERE[17]
Gian Paolo MONDINO[17]
Patrik MRÁZ[10]
Benoît OFFERHAUS[14]
Adrian OPREA[11]
Marziano PASCALE[17]
Roberto PASCAL[17]
Giorgio PERAZZA[16]
Marián PERNY[10]
Jean-Louis POLIDORI[14]
Guy REBATTU[14]
Jean-Pierre ROUX[8]
Ioan SÂRBU[11]
Silvio SCORTEGAGNA[16]
Paola SERGO[9]
Natalia SKIBITSKA[18]
Adriano SOLDANO[17]
Jean-Marie SOLICHON[14]
Simona STRGULC KRAJSEK[7]
Nadiya SYTSCHAK[18]
Zbigniew SZELAG[12]
Filippo TAGLIAFERRI[9]
Peter TURIS[10]
Tudor-Mihai URSU[11]
Jérémie VAN ES[8]
Jean-Charles VILLARET[8]
£ukasz WILK[12]

Technical assistants

Sarah BOUDON[13]
Sabine BRODBECK[5]
Véronique FINIELS[8]
Jean-Michel GENIS[8]
Hanna KUCIEL[12]
Philippe LAGIER-BRUNO[8]
Chritian MIQUEL[1]
Virgile NOBLE[8]
Tjasa POGACNIK LIPOVEC[7]
Delphine RIOUX[1]
Dirk SCHMATZ[5]
Ivan VALKO[10]
Stéphanie ZUNDEL[1]

Acknowledgments

Christian BOUCHER[14]
Jean-Marie CEVASCO[14]
Guillaume CHAUDE[14]
Dominique CHAVY[14]
Bruno CUERVAN[14]
Gil DELUERMOZ[14]
Daniel DEMONTOUX[14]
Laurence FOUCAULT[14]
Jean-Félix GANDIOLI[14]
Ernest GRENIER[14]
Emmanuel ICARDO[14]
Zoltan JABLONOVSKI[11]
Vincent KULESZA[14]
Mihai MICLÃUS[11]
Monique PERFUS[14]
Daniel REBOUL[14]
Alain ROCCHIA[14]
Jean-Pierre ROUX[14]
Robert SALANON[14]

Addresses of the Institutions (contractors and subcontractors)

[1] Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 43, F-38041 Grenoble Cedex 9, France
[2] Istituto di Zootecnica Università Cattolica del S. Cuore via E. Parmense, 84 I-29100 Piacenza, Italy
[3] Laboratoire de Botanique Evolutive Université de Neuchâtel 11, rue Emile-Argand CH-2007 Neuchâtel, Switzerland
[4] Department of Biogeography Institute of Botany, University of Vienna Rennweg 14, A-1030 Vienna, Austria
[5] Swiss Federal Research Institute WSL Ecological Genetics Zuercherstrasse 111 CH-8903 Birmensdorf, Switzerland
[6] University of Regensburg Institute of Botany D-93040 Regensburg, Germany
[7] Univerza v Ljubljani Oddelek za biologijo BF Vecna pot 111 SI-1000 Ljubljana, Slovenia
[8] Conservatoire Botanique National Alpin — CBNA Domaine de Charance F-05000 Gap, France

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

[9] Dipartimento di Biologia
Università di Trieste
Via L. Giorgieri 10
I-34127 Trieste, Italy
[10] Institute of Botany of Slovak Academy of Sciences, Department of Vascular Plant Taxonomy,
Dúbravská cesta 14,
SK-845 23 Bratislava, Slovakia
[11] Institutul de Cercetari Biologice
Str. Republicii nr. 48
400015 Cluj-Napoca, Romania
[12] Institute of Botany
Polish Academy of Sciences
Lubicz 46
PL-31-512 Kraków, Poland
[13] Medias-France/IRD
CNES – BPi 2102
18, Av. Edouard Belin
F-31401 Toulouse Cedex 9, France
[14] Parc national du Mercantour
23 rue d'Italie, BP 1316
F-06006 Nice Cedex 1, France
[15] Parco Naturale Alpi Marittime
Corso Bianco 5
I-12010 Valdieri (CN) Italy
[16] Museo Civico
Largo S. Caterina 41,
I-38068 Rovereto, Italy
[17] Istituto per le Piante da Legno e l'Ambiente
c.so Casale, 476
I-10132 Torino, Italy
[18] Institute of Ecology of the Carpathians N.A.S. of Ukraine, 4 Kozelnitska str.,
79026 Lviv, Ukraine
[19] Naturmuseum Südtirol, Museo Scienze Naturali Alto Adige
Bindergasse 1
I-39100 Bozen/Bolzano, Ital

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd