MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2010) **10**, 797–805 doi: 10.1111/j.1755-0998.2010.02887.x

THE STATE OF THE FIELD

Detecting populations in the 'ambiguous' zone: kinship-based estimation of population structure at low genetic divergence

PER J. PALSBØLL,* M. ZACHARIAH PEERY† and MARTINE BÉRUBÉ*
Department of Genetics, Microbiology and Toxicology, Stockholm University, SE-106 91 Stockholm, Sweden, †Department of Forest and Wildlife Ecology, University of Wisconsin Madison, 1630 Linden Drive, Madison, WI 53706, USA

Abstract

Identifying population structure is one of the most common and important objectives of spatial analyses using population genetic data. Population structure is detected either by rejecting the null hypothesis of a homogenous distribution of genetic variation, or by estimating low migration rates. Issues arise with most current population genetic inference methods when the genetic divergence is low among putative populations. Low levels of genetic divergence may be as a result of either high ongoing migration or historic high migration but no current, ongoing migration. We direct attention to recent developments in the use of the tempo-spatial distribution of closely related individuals to detect population structure or estimate current migration rates. These 'kinship-based' approaches complement more traditional population-based genetic inference methods by providing a means to detect population structure and estimate current migration rates when genetic divergence is low. However, for kinship-based methods to become widely adopted, formal estimation procedures applicable to a range of species life histories are needed.

Keywords: coalescent, population genetics, Population structure, relatedness, Wright's F_{ST}

Received 10 January 2010; revision received 23 April 2010, 4 May 2010; accepted 5 May 2010

Introduction

Identifying population structure is one of the most common and important objectives of spatial analyses using population genetic data (Waples & Gaggiotti 2006). Interests in detecting population structure are many and diverse, such as understanding processes of local adaptation, effects of habitat fragmentation as well as delineating management units (MUs) (Hauser & Carvalho 2008).

Although there currently is no single generally applicable definition of what constitutes a population (Waples & Gaggiotti 2006), the absence of such a definition has not prevented the application of population genetic inference methods to identify population structure. Two different general approaches are employed to identify population structure: (i) the detection of discernable groupings of conspecifics (e.g., as a 'distinct population segment' under the US Endangered Species Act, Brosi & Biber 2009), or (ii) from estimation of (reduced) migration rates among putative populations.

In the first instance, the 'statistical' criterion is used to identify populations, essentially following the operational criterion for MUs originally proposed by Moritz (1994). Moritz defined populations as having '…significantly different allele frequencies.' (Moritz 1994), implying that the probability of drawing the observed samples from the same allele frequency distribution is <0.05 (Waples & Gaggiotti 2006; Palsbøll *et al.* 2007; Brosi & Biber 2009; Morin *et al.* 2009). The use of the statistical criterion has attracted criticism because of the fact that the statistical power to detect populations structure is, in part, a function of the amount of data and thus correlated to the number of loci and individuals analysed (Waples & Gaggiotti 2006; Fallon 2007; Palsbøll *et al.* 2007; Bernard *et al.* 2009). Consequently, population structure may go undetected when population genetic divergence is high

Correspondence: Per J. Palsbøll, Fax: +46 08 16 43 15; E-mail: per.palsboll@gmt.su.se

because of low statistical power, or genetic structure may be identified among genetically similar populations when the statistical power is high (see Waples & Gaggiotti 2006; Palsbøll *et al.* 2007). For example, Fallon (2007) found that populations of threatened species were more likely to be designated as 'distinct population segments' when large numbers of loci and several different kinds of genetic markers (e.g., mitochondrial DNA and nuclear DNA) were employed.

The second approach is to identify population structure from the degree of migration among putative populations. What exactly is implied by 'migration' depends upon the study objective as well as the data and inference method used. Migration may be gene flow, the movement of gene copies among population with or without permanent displacement of individuals. Alternatively, migration may be dispersal, the permanent translocation of individuals among populations with or without a lasting transfer of gene copies into the recipient population. In many instances, it is not clear what kind of migration is estimated when inferring migration rates from population genetic data. An example of the latter case is MUs. MUs are usually assumed to represent demographically independent populations (Brown & Ehrlich 1980; Ihssen *et al.* 1981; McElhany *et al.* 2000; Latta 2008), where demographic independence implies that abundance is mainly driven by local recruitment and mortality rates. In such cases, population genetic methods are used to estimate migration rates, which in turn may provide information regarding the degree of demographic independence among putative MUs (e.g., Palme *et al.* 2008).

The popularity of population genetic approaches to infer population structure lies primarily with the well-developed population genetic theory and the ease by which population genetic data are collected (DeSalle & Amato 2004; Sarre & Georges 2009). The most commonly used population genetic inference methods (e.g., Wright's F_{ST}, assignment tests or coalescent-based inference methods) to detect population structure perform well when the degree of genetic structure is high because the interpretation is relatively straightforward. It is, however, another matter when the degree of genetic divergence among putative populations is low because current low genetic divergence has two possible causes; ongoing migration, or recent historic migration but no (or low) current migration. In other words, low genetic divergence does not necessarily imply ongoing migration among populations. The inability of most population genetic inference methods to discern between ongoing or recent historic migration is because of the assumptions underlying the estimation of genetic divergence (and inferred migration rates) (see, Whitlock & McCauley 1999; Pearse & Crandall 2004; Pertoldi *et al.* 2007). The purpose of this paper is to direct attention towards this 'ambiguous zone' when genetic divergence is low and point to the use of an emerging, novel breed of kinship-based genetic inference methods.

Population and kinship-based genetic inference methods

For the purposes of this paper, we categorize the methods used to infer population genetic structure as either population-based or kinship-based. Examples of population-based inference methods are moment-based statistics, such as Wright's F_{ST} (Wright 1951) as well as coalescent methods (Kingman 1982; Hudson 1998). Even assignment methods (Paetkau *et al.* 1995; Pritchard *et al.* 2000), including spatial methods (Manel *et al.* 2007; Guillot *et al.* 2009) belong to this category as the individual multilocus genotypes are assigned to a 'population' of samples.

In contrast, kinship-based methods rely upon assigning individual multilocus genotypes to other individual multilocus genotypes (as opposed to a population of multilocus genotypes). One example of a kinship-based method is genetic tagging methods during which the multilocus genotypes obtained from different samples are assigned either to an individual or related individuals based upon the estimated degree of relatedness (Paetkau *et al.* 1995; Palsbøll *et al.* 1997; Taberlet *et al.* 1997; Garrigue *et al.* 2004; Peery *et al.* 2008; Broquet *et al.* 2009; Planes *et al.* 2009; Saenz-Agudelo *et al.* 2009).

Population-based and kinship-based approaches are complementary. Strongly structured populations (i.e., displaying a high degree of genetic divergence) are readily detected with high statistical power with modest amounts of data using population-based methods. Kinship-based methods, on the other hand, are well suited when the degree of genetic divergence is low, such as when current migration rates are high (Peery *et al.* 2008) or when populations have undergone a recent reduction in migration rates (e.g., Epps *et al.* 2005; Peery *et al.* 2010), which is when population-based methods tend to be uninformative. Kinship-based methods may be employed to estimate migration rates or to detect population structure using the statistical criterion.

The dubious 'ambiguous' zone; when genetic divergence is low

Most population-based population genetic inference methods (except for assignment methods) make the assumption that population size and migration rates remain constant through time (e.g., Wright 1951; Beerli & Felsenstein 2001). A few implementations allow for non-zero (linear or exponential) population growth rates, but such nonzero growth rates are still assumed to remain

constant through time (Hey *et al.* 2004; Kuhner 2006). In addition, most population-based methods assume ideal Wright–Fisher populations, where each population is panmictic with no reproductive skew amongst individuals and has discrete generations. The most commonly employed estimator of population genetic divergence, Wright's F_{ST} (Wright 1951), relies upon the simplest population model where all subpopulations are identically sized ideal populations with identical migration rates among all populations, and no change in population sizes and migration rates through time.

When these assumptions are valid then low genetic divergence (e.g., $F_{ST} < 0.05$–0.02) implies high migration rates and high genetic divergence implies low migration rates. However, most natural populations undergo changes in population size, and migration rates fluctuate (Whitlock & McCauley 1999; Pearse & Crandall 2004) in which case the correlation between current genetic divergence and current migration rates is less straightforward. Such deviations from the underlying assumptions do not constitute an issue, in terms of inferring population structure, when the degree of genetic divergence is high in which case current migration rates are unlikely to exceed the inferred rate, although it is in principle possible that a recent rapid increase in migration may not have resulted in a corresponding decrease in genetic divergence (Waples & Gaggiotti 2006). However, low genetic divergence among populations does not necessarily imply that *current* migration rates are high because such low genetic divergence structure could be because of high *historical* migration rates among populations that are now isolated (see Fig. 1). There are numerous examples of such recent changes in migration rates in natural populations inhabiting humanly altered ecosystems leading to populations that are not in drift-mutation-migration equilibrium. For example, logging of old-growth forests in the Pacific Northwest of North America some 60–120 years ago appears to have reduced the rate of gene flow into a peripheral population of marbled murrelets (*Brachyramphus marmoratus*) to near zero, yet the current degree of genetic divergence was estimated at only approximately 0.03 (Peery *et al.* 2010). Similarly, greatly reduced migration among mountain ranges by desert mountain sheep (*Ovis canadensis nelsoni*) in southeastern California because of urbanization, construction of canals for water supply, and fencing of highways during the last 40–70 years is not yet reflected in the degree of genetic divergence (Epps *et al.* 2005).

Assignment tests have been promoted as a means to obtain 'real-time' estimates of migration rates (Davies *et al.* 1999; Paetkau *et al.* 2004). Individual multilocus genotypes are assigned to putative populations based upon the relative likelihoods of multilocus genotypes given the population allele frequencies and assuming

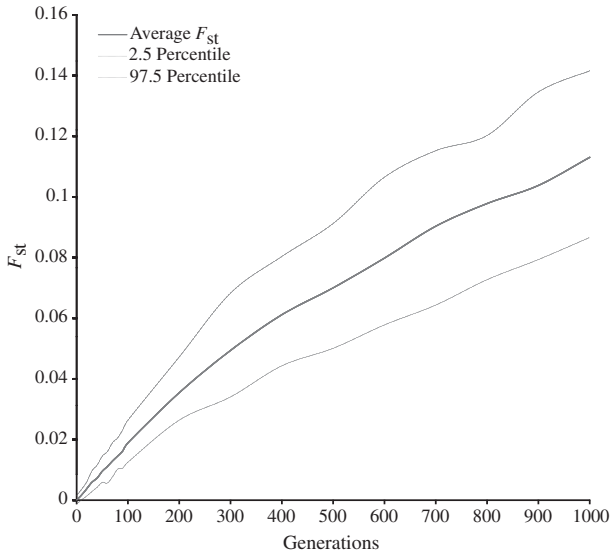

Fig. 1 The increase in divergence, estimated as F_{ST}, as a function of the number of generations at a zero migration rate after fragmentation. Generated using SimCoal2 (Laval & Excoffier 2004) with a total effective population size of 5000 (and genotypes from 20 microsatellite loci in 100 samples from each population) which is split into two equal-sized populations with zero migration.

panmictic populations (Paetkau *et al.* 1995; Pritchard *et al.* 2000; Corander *et al.* 2004). The proportion of individuals assigned to other populations than that from which they were sampled is positively correlated with the immigration rate and reflects migration during the last generation(s). Descendants from immigrant individuals may also be identified as being of mixed 'origin' (Rannala & Mountain 1997; Wilson & Rannala 2003). Identifying descendants of immigrants can extend the time frame over which migration rates are estimated beyond a single generation. This approach, or variations thereof, are implemented in BayesAss+ (Wilson & Rannala 2003) and BIMr (Faubet & Gaggiotti 2008), which estimate m in this manner and which do not assume panmixia either. Broquet *et al.* (2009) have developed an assignment-based method where postdispersal genotypes are assigned to predispersal source populations.

Successful implementation of assignment tests and the precision of resulting migration rate estimates require that individuals are assigned to their source population with statistical rigour, which in turn is correlated to the degree of genetic divergence among source populations. The confidence in assignments increases with the degree of genetic divergence among populations (Paetkau *et al.* 2004; Faubet *et al.* 2007). However, high levels of genetic divergence also imply low migration rates, and consequently a large proportion of each population needs to be sampled for immigrants to be among the collected samples, or the result will be a zero rate of migration rates. At

low levels of genetic divergence, confidence in assignments is low because the likelihood of a given genotype is similar across populations. Hence, it is not surprising that assignment methods perform best at intermediate levels of genetic divergence (F_{ST} above 0.05, Faubet *et al.* 2007) and at even higher levels of genetic divergence when the assumptions of panmixia are not met ($F_{ST} > 0.1$, Faubet *et al.* 2007). In conclusion, population-based genetic inference methods, including more recent coalescent-based approaches, rely upon a number of assumptions which creates difficulties in terms of drawing firm conclusions regarding the *current* population structure and *current* migration rates when the degree of genetic divergence is low.

Kinship-based methods

An alternative to population-based methods is kinship-based methods, where population structure or migration rates are estimated from the spatio-temporal distribution of dyads of related individuals. Kinship-based methods are in essence assignment methods, where a multilocus genotype is assigned to other multilocus genotypes rather than populations. The data then becomes the number (and distribution) of 'recaptures', which may either be the same individual identified from different samples or closely related individuals. The number of such recaptures in turn influences the precision of the final estimate of migration rates and the statistical power to detect population structure.

The key advantage of kinship-based methods is that the statistical rigour and power of each assignment (and thus any inferred population structure or estimated migration rates) depends upon overall level of genetic variation but *not* the degree of genetic divergence among populations (Palsbøll 1999; Peery *et al.* 2008). Consequently, kinship-based methods are well suited for elucidating population structure in the 'ambiguous' zone, i.e., when the genetic divergence among putative populations is low, complementing population-based methods (Palsbøll 1999; Hellberg 2009; Planes *et al.* 2009; Saenz-Agudelo *et al.* 2009). Kinship-based methods yield estimates of *current* population structure or migration rates. The exact time frame of kinship-based estimates depends upon several factors, such as generation time, dispersal characteristics, as well as the targeted degree of relatedness (i.e., 1st, 2nd or more distantly related dyads). Targeting dyads of distantly related individuals will increase the time frame to which the migration estimate applies because the probability of a migration event increases with time. However, because the number of loci that can be typed for each sample limits the degree of relationship that can be discerned from unrelated individuals (usually first or second-order relatives), the time frame covered by kinship-based methods will be (at most) the last few generations.

The main issue with kinship-based methods lies with the magnitude of the required sampling effort. Typically a larger proportion of the target populations needs to be sampled compared to most population-based methods, and a sufficient number of loci need to be genotyped to provide confidence in individual identifications or the assignment of dyads of related individuals (Blouin 2003; Wang 2006; Jones *et al.* 2010), which in turn makes this approach sensitive to genotyping errors (Pompanon *et al.* 2005). Detecting dyads of close relatives amongst large sample sizes is an additional issue with kinship approaches (Csillery *et al.* 2006). Each putative dyad of related individuals constitutes a statistical test leading to multitesting issues, which is further complicated by the fact that many tests are dependent (i.e., the same individual is part of n-1 dyads, where n is the number of sampled individuals). Applying common corrective measures, such as Sequential Bonferroni corrections (Holm 1979), is often too restrictive and results in a large number of Type II errors because of the desire to keep Type I errors at zero. The alternative is FDR (false discovery rates, Benjamini & Hochberg 1995) methods where a fraction of Type I errors is accepted to minimize the number of Type II errors (e.g., Hall *et al.* 2009; Skaug *et al.* 2010). The optimal FDR depends upon the objective. The FDR approach is informally implemented in one of the most commonly used computer programs used to infer parentage, CERVUS (Marshall *et al.* 1998). In CERVUS, the level of significance (i.e., the probability of the observed value of the test statistic Δ) may be adjusted to increase the number of detected paternities. It is common to use 0.2 instead of the 'usual' 0.05. By choosing 0.2, a higher rate of Type I errors is accepted, but the power to identify parents is increased (e.g., Pilot *et al.* 2010). Note that even a relatively low Type I error rate of 0.05 could result in misleading estimates of migration rates. For example, if 100 individuals are sampled in each of two populations unconnected by migration, a Type I error rate of 0.05 would be expected to result in 500 split 'parent–offspring' dyads that actually involved unrelated individuals (given 100 000 unique pairs where the two individuals occur in different populations). However, parent–offspring dyads constitute a special case, which, if a sufficiently large number of loci is genotyped, may be identified by exclusion (Jones *et al.* 2010). In this case, both Type I and II errors may be reduced to near zero provided that a sufficient number of highly polymorphic loci are used and that the genotyping error rate is sufficiently low.

The simplest case of kinship-based methods is 'genetic tagging' of individuals (Palsbøll *et al.* 1997; Taberlet *et al.* 1997). Here, multilocus genotypes obtained from

different samples are assigned to the same individual when the relatedness coefficient is estimated to one. The spatial distribution of samples assigned to the same individual may then be used to infer population structure and migration rates as with any other individual tagging method (Palsbøll *et al.* 1997; Taberlet *et al.* 1997). However, this method requires that the same individuals are sampled on multiple occasions, and the identification of migration events requires that same individual be sampled in at least two populations.

Only a very few studies of natural populations have utilized dyads of close relatives to estimate migration rates. One example is the study by Peery *et al.* (2008) of marbled murrelets (*Brachyramphus marmoratus*) in central California, for which abundance had been stable over a 5-year study period despite apparent reproductive failure, possibly high immigration rates from abundant neighbouring populations. The kinship-based approach was adapted to a single-population system because the number of dyads of close kin decreases with increasing immigration rates relative to a closed population. Individual-based simulations based upon estimates of demographic parameters, such as survival and birth rates, derived from field studies were employed to estimate the expected proportion of parent–offspring dyads occurring in the central California population at different immigration rates (incl. zero) assuming a constant local abundance of approximately 660 individuals. Genetic analyses identified 70 dyads of first-order relatives among the 271 individuals genotyped at 16 loci, which matched the expectations of a population with an annual immigration rate of approximately 4% (95% confidence interval: 3–5%) (Peery *et al.* 2008). This annual immigration rate translates into 32% per generation corresponding to *Nm* at approximately 50 assuming an effective population size at 155 individuals (Peery *et al.* 2010).

Kinship-based methods have also been developed to detect population structure using statistical criteria. In this application, the objective is to detect spatial heterogeneity in the distribution of dyads of related individuals. The null expectation (homogeneity) is estimated by randomizing sampled individuals among sampling positions. If the observed value of the test statistic (e.g., the number of dyads with members in two different populations) is outside the 95 percentile bounds of the null distribution, then the null hypothesis of spatial homogeneity is rejected and population structure inferred. The caveats of homogeneity tests in this context (in terms of statistical power and sample sizes) are the same as those mentioned earlier for population-based methods.

Økland *et al.* (2010) developed and tested a clustering procedure for genetically-identified dyads of close relative to define populations. Their approach was applied to individual multilocus genotypes obtained by individual-based simulations (Martien *et al.* 2009). Dyads of individuals estimated to be related as second or first-order relatives (excluding full siblings which were deemed unlikely for the generic target species, a baleen whale) were identified using standard likelihood methods. In a two-area population model, simulations used census population sizes at 3750 or 7500 individuals per population and sample sizes were varied from 100 to 600 individuals per population. The migration rate was varied from 5×10^{-6} to 5×10^{-3}, and the generation time set at 20 years. The number of migrants per generation was thus between 0.0375 and 37.5 migrants per generation for the lower population sizes and twice that for the larger population sizes, assuming an effective population size at 10% of census population size. Statistical power to reject homogeneity was estimated to be >95%, given a total sample size of 500–600 individuals. Notably, the statistical power did not increase linearly with the proportion of the population sampled. A sample of 300 individuals taken from a population of 7500 individuals yielded the same statistical power (approximately 65%) as a sample size of 400 individuals taken from a population of 15 000 individuals. In other words, panmixia was rejected with a statistical power at 65% with only approximately 2.6% of the individuals sampled and only 10 microsatellite loci genotyped at migration rates at or above approximately 75 migrants per generation.

Økland *et al.*(2010) targeted dyads of both first and second-order relatives to accommodate genotyping errors and increase the detection rate of dyads of related individuals. A genotyping error may result in a true parent and offspring dyad being deemed incompatible with being a parent and offspring dyad as they will not match at minimum one allele at all loci. In such instance, the dyad of first-order relatives is likely to appear in the analysis as a dyad of second-order relatives. With as few as 10 microsatellite loci, many first and second-order dyads are not identified, and dyads of unrelated individuals will erroneously be assigned as dyads of first or second-order relatives. Økland *et al.* (2010) used FDR (Benjamini & Hochberg 1995) to maximize the number of dyads of close relatives while accepting that some of these may be incorrect (i.e., not close relatives). The effect of accepting a large fraction of incorrect dyads (i.e., unrelated individuals inferred as dyads of first or second-order relatives) depends upon the degree of genetic divergence among populations; if genetic divergence is low, then an increased FDR is expected to reduce the statistical power to reject homogeneity, whereas high genetic divergence may result in erroneous rejection of homogeneity (because most erroneous dyads are likely to be from the same 'population').

The use of kinship-based genetic inference methods is still very much in its infancy. The two main obstacles to a more widespread adoption are the following: (i) the necessary 'scaling-up' in terms of samples and loci, and (ii) the integration of demographic and genetic theory and modelling to develop suitable estimation procedures. The first obstacle will soon be obsolete given the ongoing forward leap in genotyping technologies which will make robust genotyping of a large number of loci in 100s of individuals a trivial matter. The most promising markers are single nucleotide polymorphisms (SNPs), which may be detected by next generation sequencing technologies or the many highly efficient SNP genotyping platforms that have emerged over the last decade (Brumfield *et al.* 2003; Morin *et al.* 2004). SNP discovery in nonmodel species is still a significant challenge but this is very much the same situation at the onset of microsatellite genotyping in nonmodel species, and new efficient SNP discovery methods are developed at an increasing rate (Kerstens *et al.* 2009; Sanchez *et al.* 2009; Xu *et al.* 2009). In the longer perspective, the same kind of data may be achieved by simply next-generation sequencing of the same small fraction of the genome in a pool of many individuals (Erlich *et al.* 2009).

In terms of the second obstacle, a general estimation framework needs to be developed that may be applied to a range of species and specific life history rates. There are a few programs available that enable simulating individuals under a variety of demographic models, which in turn may be utilized to generate expected distribution of close kin given specific life history and vital rates (e.g., Peery *et al.* 2008). Perhaps more importantly, formal statistical frameworks for estimating migration rates (or population structure) from the observed number and spatial distribution of dyads of close kin have not been developed.

Conclusions and recommendations

The current and common use of population-based inference methods to detect population structure works well when the degree of genetic divergence among populations is high. However, at low genetic divergences inferring population structure becomes nontrivial because a low degree of genetic divergence may signify either high historic (but low current) or high current levels of migration. Here, we have presented kinship-based methods as a viable approach for characterizing current population structure and estimating migration rates in this 'ambiguous' zone of low genetic divergence. Although it has been suggested that kinship-based methods were infeasible in large populations (>1000 individuals, see Hellberg 2009), the work by Økland *et al.* (2010) shows this not to be the case. Indeed, there are no theoretical grounds for such limitation, but there are practical limitations to a successful implementation of kinship-based methods when population sizes become large in terms of the number of samples that needs to be collected and genotyped. In theory, such limitations may be overcome by extending the assessment to include more distant relationships (Palsbøll 1999). However, third-order, and more distant, relatives only share a small fraction (approximately 0.125) of their genome because of their common ancestry making it technically challenging to discern dyads of such distant relationships from dyads of unrelated individuals. The work by Økland *et al.* (2010) is encouraging by revealing that the same level of statistical power is achieved at lower sampling intensity as abundance increases. Saenz-Agudelo *et al.* (2009) compared the performance of assignments tests (i.e., a population-based method in the present context) and paternity tests (a kinship-based method) in detecting population structure by identifying movement among reef fish subpopulations at low and high population genetic divergence. As expected, and outlined earlier, Saenz-Agudelo *et al.* (2009) found that assignment tests and paternity analyses captured movements among subpopulations best at high and low levels of genetic divergence, respectively, in accordance with what is to be expected.

We conclude that kinship-based and population-based inference methods are complementary and which approach is the most suitable depends upon the specific

Table 1 Individual-based programs for simulating genetic data

Program name	Population model	Comments	Reference
EASYPOP	Wright–Fisher	Tracks parentage	(Balloux 2001)
SPIP	Demographic*	Limited to two populations	(Anderson & Dunham 2005)
RMetaSim	Demographic	No tracking of parentage	(Strand 2002)
SimuPop, Nemo	Demographic	Tracks parentage	(Peng & Kimmel 2005; Guillaume & Rougemont 2006)

*Enable the simulation of age structured populations with overlapping generations as well as reproductive and migratory skew among individuals and age classes.

context. Kinship-based inference methods, however, are in need of further development to become more widely adopted. Such development should include estimation procedures which may be tailored to the demographic parameter values for a specific target species (such as abundance, survival rates, reproductive skew, mating fidelity etc.) as well as other biologically plausible deviations from a Wright–Fisher population model and mutation-drift-migration equilibrium. There are several individual-based modelling programs available enabling simulation of kinship data under different migration rates and a specific demographic model (see Table 1). Only a few software packages permit temporal changes in migration rates and abundance which is required if the objective is to model a recent human-induced change, such as habitat fragmentation. However, none of these programs permit a formalized estimation of current migration rates from the spatial distribution of dyads of close kin, but are intended for generating simulated data. Although such software may be used to obtain an estimate of migration rates in an *ad hoc* manner (e.g., Peery *et al*. 2008), estimation procedures need to be formalized and in a rigorous statistical framework to be of more general use.

Acknowledgements

We thank Finn Larsen, Oscar Gaggiotti and two anonymous reviewers for their constructive comments.

References

Anderson EC, Dunham KK (2005) SPIP 1.0: a program for simulating pedigrees and genetic data in age-structured populations. *Molecular Ecology Notes*, **5**, 459–461.

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

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

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

Bernard AM, Ferguson MM, Noakes DLG, Morrison BJ, Wilson CC (2009) How different is different? Defining management and conservation units for a problematic exploited species. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 1617–1630.

Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution*, **18**, 503–511.

Broquet T, Yearsley J, Hirzel AH, Goudet J, Perrin N (2009) Inferring recent migration rates from individual genotypes. *Molecular Ecology*, **18**, 1048–1060.

Brosi BJ, Biber EG (2009) Statistical inference, Type II error, and decision making under the US Endangered Species Act. *Frontiers in Ecology and the Environment*, **7**, 487–494.

Brown IL, Ehrlich PR (1980) Population biology of the checkerspot butterfly, *Euphydryas-chalcedone* structure of the Jasper Ridge colony. *Oecologia*, **47**, 239–251.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

Corander J, Waldmann P, Marttinen P, Sillanpaa MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, **20**, 2363–2369.

Csillery K, Johnson T, Beraldi D *et al*. (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, **173**, 2091–2101.

Davies N, Villablanca FX, Roderick GK (1999) Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology & Evolution*, **14**, 17–21.

DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews Genetics*, **5**, 702–712.

Epps CW, Palsbøll PJ, Wehausen JD *et al*. (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters*, **8**, 1029–1038.

Erlich Y, Chang K, Gordon A *et al*. (2009) DNA Sudoku-harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome Research*, **19**, 1243–1253.

Fallon SM (2007) Genetic data and the listing of species under the US endangered species act. *Conservation Biology*, **21**, 1186–1195.

Faubet P, Gaggiotti OE (2008) A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics*, **178**, 1491–1504.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149–1166.

Garrigue C, Dodemont W, Steel D, Baker CS (2004) Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm low abundance and reproductive autonomy of humpback whales on the wintering grounds of New Caledonia. *Marine Ecology-Progress Series*, **274**, 251–262.

Guillaume F, Rougemont J (2006) Nemo: an evolutionary and population genetics programming framework. *Bioinformatics*, **22**, 2556–2557.

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734–4756.

Hall LA, Palsbøll PJ, Beissinger SR *et al*. (2009) Characterizing dispersal patterns in a threatened seabird with limited genetic structure. *Molecular Ecology*, **18**, 5074–5085.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333–362.

Hellberg ME (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology Evolution and Systematics*, **40**, 291–310.

Hey J, Won YJ, Sivasundar A, Nielsen R, Markert JA (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Molecular Ecology*, **13**, 909–919.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Hudson RR (1998) Island models and the coalescent process. *Molecular Ecology*, **7**, 413–418.

Ihssen PE, Booke HE, Casselman JM *et al.* (1981) Stock identification - materials and methods. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1838–1855.

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

Kerstens HHD, Crooijmans R, Veenendaal A *et al.* (2009) Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey. *BMC Genomics*, **10**, DOI:10.1186/1471-2164-10-479.

Kingman JFC (1982) On the Genealogy of large populations. *Journal of Applied Probability*, **19**, 27–43.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, **22**, 768–770.

Latta RG (2008) Conservation genetics as applied evolution: from genetic pattern to evolutionary process. *Evolutionary Applications*, **1**, 84–94.

Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, **20**, 2485–2487.

Manel S, Berthoud F, Bellemain E *et al.* (2007) A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Molecular Ecology*, **16**, 2031–2043.

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

Martien KK, Gregovich D, Bravington MV *et al.* (2009) tossm: an R package for assessing performance of genetic analytical methods in a management context. *Molecular Ecology Resources*, **9**, 1456–1459.

McElhany P, Ruckelshaus MH, Ford MJ, Wainwright TC, Bjorkstedt EP (2000) Viable salmonid populations and the recovery of evolutionarily significant units. *U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-42*, 156 pp.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, **9**, 66–73.

Moritz C (1994) Defining evolutionarily-significant-units for conservation. *Trends in Ecology & Evolution*, **9**, 373–375.

Økland JM, Haaland ØA, Skaug HJ (2010) A method for defining management units based on genetically determined close relatives. *ICES Journal of Marine Science*, **67**, 551–558.

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

Paetkau D, Slade R, Burden M, Estoup A (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.

Palme A, Laikre L, Utter F, Ryman N (2008) Conservation genetics without knowing what to conserve: the case of the Baltic harbour porpoise Phocoena phocoena. *Oryx*, **42**, 305–308.

Palsbøll PJ (1999) Genetic tagging: contemporary molecular ecology. *Biological Journal of the Linnean Society*, **68**, 3–22.

Palsbøll PJ, Allen J, Berube M *et al.* (1997) Genetic tagging of humpback whales. *Nature*, **388**, 767–769.

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

Pearse DE, Crandall KA (2004) Beyond F-ST: analysis of population genetic data for conservation. *Conservation Genetics*, **5**, 585–602.

Peery MZ, Beissinger SR, House RF *et al.* (2008) Characterizing source-sink dynamics with genetic parentage assignments. *Ecology*, **89**, 2746–2759.

Peery MZ, Hall LA, Sellas A *et al.* (2010) Genetic analyses of historic and modern marbled murrelets suggest decoupling of migration and gene flow after habitat fragmentation. *Proceedings of the Royal Society B-Biological Sciences*, **277**, 697–706.

Peng B, Kimmel M (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21**, 3686–3687.

Pertoldi C, Bijlsma R, Loeschcke V (2007) Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodiversity and Conservation*, **16**, 4147–4163.

Pilot M, Dahlheim ME, Hoelzel AR (2010) Social cohesion among kin, gene flow without dispersal and the evolution of population genetic structure in the killer whale (Orcinus orca). *Journal of Evolutionary Biology*, **23**, 20–31.

Planes S, Jones GP, Thorrold SR (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5693–5697.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.

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 of the United States of America*, **94**, 9197–9201.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2009) Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. *Molecular Ecology*, **18**, 1765–1776.

Sanchez CC, Smith TPL, Wiedmann RT *et al.* (2009) Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics*, **10**, DOI:10.1186/1471-2164-10-559.

Sarre SD, Georges A (2009) Genetics in conservation and wildlife management: a revolution since Caughley. *Wildlife Research*, **36**, 70–80.

Skaug HJ, Bérubé M, Palsbøll PJ (2010) Detecting dyads of related individuals in large collections of DNA-profiles by controlling the false discovery rate. *Molecular Ecology Resources*, **10**, 693–700.

Strand AE (2002) METASIM 1.0: an individual-based environment for simulating population genetics of complex population dynamics. *Molecular Ecology Notes*, **2**, 373–376.

Taberlet P, Camarra JJ, Griffin S *et al*. (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, **6**, 869–876.

Wang JL (2006) Informativeness of genetic markers for pairwise relationship and relatedness inference. *Theoretical Population Biology*, **70**, 300–321.

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**, 1419–1439.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). *Heredity*, **82**, 117–125.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Xu JY, Xu GB, Chen SL (2009) A new method for SNP discovery. *BioTechniques*, **46**, 201–208.