Heredity (2016) 116, 349–350

www.nature.com/hdy

NEWS AND COMMENTARY

Homoploid hybrid expectations

The devil is in the details: the effect of population structure on demographic inference

P Orozco-terWengel

Heredity (2016) **116,** 349–350; doi:10.1038/hdy.2016.9; published online 17 February 2016

Multiple ecological and evolutionary forces shape the genetic makeup of natural populations; one of which is population structure (Wright, 1951). Population structure refers to the division of a population's gene pool between groups of individuals such that random mating is more likely between individuals within a group and less likely between individuals from different groups. Owing to the ubiquity of population structure (and its associated varying levels of connectivity between subpopulations), it is common practice in molecular ecology to use statistical approaches to identify population structure in multi-individual datasets using molecular markers (for example, Corander and Martinen, 2006; Alexander *et al.*, 2009). However, developments in sequencing technologies over the last decade have enabled researchers to produce data covering most, if not all, of the complete genome of a species relatively cheaply (for example, Groenen *et al.*, 2012; Zhan *et al.*, 2013), and estimate important evolutionary parameters using as few as a single genome, for example, the trend in effective population size over time (for example, Li and Durbin, 2011; Schiffels and Durbin, 2014). However, such methodologies assume that the data used for the analysis represent that of an unstructured population, and as it has been shown previously, deviations from that model can inflate estimates of the effective population size (Leblois *et al.*, 2006; Heller *et al.*, 2013; Bosse *et al.*, 2014).

In this issue of *Heredity*, Mazet *et al.* develop a theoretical framework that illustrates the effect of otherwise ignored population structure on demographic inferences using genomic methods such as the Pairwise Sequentially Markovian Coalescent (PSMC; Li and Durbin, 2011). For their method, Mazet *et al.* derive the 'Inverse Instantaneous Coalescent Rate (IICR)': the inverse of the coalescent rate throughout time estimated from a sample consisting of two haplotypes as found in a diploid individual. The IICR in a genetic dataset from an unstructured population sample corresponds to the trajectory of the effective population size over time, analogous to the output of the PSMC method. However, if the data are from a structured population, the IICR function corresponds to the trajectory of the effective population size and the migration pattern between subpopulations (Figure 1). Interestingly, Mazet *et al.* show that when sampling a structured population, if the two genomic haplotypes sampled derive from the same subpopulation, the inferred demographic trend describes a reduction in the effective population size akin to a bottleneck signal (Figure 1, top right). In contrast, if each of the two haplotypes sampled derive from different subpopulations (for example, as when the sampled genome corresponds to an F1 individual descending from a migrant), the resulting inferred demographic trend is the opposite, namely a population expansion (Figure 1, bottom right). What is interesting is that in either case (that is, the signal of a bottleneck or that of an expansion), the trend described is spurious and independent of whether a real change in effective population size occurred. In other words, estimating the demographic history of a population using approaches that do not take into consideration the effect of population structure may show an inherent bias towards identifying changes in the demographic history of a population, even when the population has remained stable over time. In addition, changes in the migration rates between subpopulations (still without changes in the overall effective population size) result in wavy PSMC and IICR trends, like those typically interpreted as representing expansions and bottlenecks (Figure 5 of Mazet *et al.*, this issue). Using their IICR approach, Mazet *et al.* investigated the demographic history of humans previously analysed using PSMC (Li and Durbin, 2011), but assuming a structured population with changes in migration rate and either a stable or changing demography. For both scenarios Mazet *et al.* carried out neutral simulations of a structured population with the same parameters used by Li and Durbin (2011) for the human PSMC, but including up to three migration events between 2.52 and 0.24 million years ago. With their simulations Mazet *et al.* showed that the previously described demographic history of our species (Li and Durbin, 2011) can also be obtained from a model of changes in connectivity between subpopulations since the beginning of the Pleistocene, and irrespectively of whether there was a change in population size (Figure 6 and Figure S4 of Mazet *et al.*, this issue).

The risk of such spuriously inferred population size changes has been reported previously (for example, Leblois *et al.*, 2006; Stadler *et al.*, 2009; Heller *et al.*, 2013; Bosse *et al.*, 2014) and the developers of the PSMC method also pointed out that if a population splits in half and later on merges again, even in the absence of demographic changes, an increase in effective population size can be observed (Li and Durbin, 2011). Heller *et al.* (2013) showed how deviations from the random mating model affect demographic inferences using Bayesian skyline plots (BSP), a standard tool used for analysing mitochondrial DNA sequences. Using sampling schemes where all haploid samples derived from the same subpopulation or from

P Orozco-terWengel is at School of Biosciences, Cardiff University, Cardiff, UK
E-mail: orozco-terwengelpa@cardiff.ac.uk

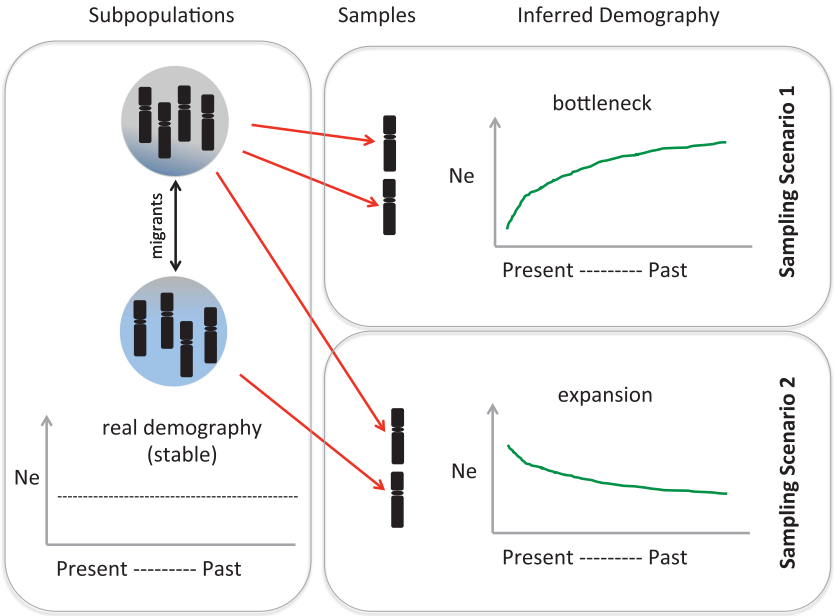

Figure 1 Schematic diagram of a structured population and the inferred demographic histories from two alternative sampling scenarios. Two subpopulations are shown (one grey and one blue) which exchange migrants at varying rates throughout time, and which have a stable demographic history. In scenario 1, a diploid individual is sampled whose haplotypes derive from the same subpopulation and a bottleneck type demographic trend was inferred; while in scenario 2, an individual is sampled whose haplotypes derive from different subpopulations (e.g. the F1 from a migrant) and an expansion type demographic trend was inferred.

more than one subpopulation, Heller *et al.* demonstrated that for simulations with a stable population size the various sampling regimes mostly resulted in inferred population bottlenecks (Figure 1 in Heller *et al.*, 2013). In addition, if simulations included a population expansion or a bottleneck, the former could not be identified if the samples derived form the same subpopulation, and the latter could not be detected if an even amount of samples from each subpopulation was used for the analysis (Figure 2 in Heller *et al.*, 2013).

The results described above are in line with those of Mazet *et al.* in this issue of *Heredity* and call for the development of approaches that allow the comparison of alternative models to the simple one based on effective population size changes over time (for example, population structure, varying levels of gene flow between subpopulations or a combination of the previous two and simultaneously occurring demographic changes). In that context, a maximum likelihood approach has recently been developed to attempt to distinguish between a structured model and a population demographic model with one single change using a single diploid genome dataset (Mazet *et al.*, 2015). Nevertheless, until better approaches are developed that allow researchers to disentangle the effect of structure from that of demographic change, the IICR results showed here call for a cautious interpretation of demographic trends like those inferred by PSMC and which have become a standard result in many publications using genomic data.

CONFLICT OF INTEREST

The author declares no conflict of interest.

Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**: 1655–1664.

Bosse M, Megens HJ, Madsen O, Frantz LA, Paudel Y, Crooijmans RP, Groenen MA (2014). Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent *Sus scrofa* populations. *Mol Ecol* **23**: 4089–4102.

Corander J, Martinen P (2006). Bayesian identification of admixture events using multilocus molecular markers. *Mol Ecol* **10**: 2833–2843.

Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF *et al.* (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**: 393–398.

Heller R, Chikhi L, Siegismund HR (2013). The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS ONE* **8**: e62992.

Leblois R, Estoup A, Streiff R (2006). Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? *Mol Ecol* **15**: 3601–3615.

Li H, Durbin R (2011). Inference of human population history from individual whole-genome sequences. *Nature* **475**: 493–496.

Mazet O, Rodriguez W, Chikhi L (2015). Demographic inference using genetic data from a single individual: separating population size variation from population structure. *Theor Popul Biol* **104**: 46–58.

Mazet O, Rodriguez W, Grusea S, Boitard S, Chikhi L (2016). On the importance of being structured: instantaneous coalescence rates and human evolution: lessons for ancestral population size inference? *Heredity* (this issue).

Schiffels S, Durbin R (2014). Inferring human population size and separation history from multiple genome sequences. *Nat Genet* **46**: 919–925.

Stadler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P (2009). The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* **182**: 205–216.

Wright S (1951). The Genetical structure of populations. *Ann Eugen* **15**: 323–354.

Zhan X, Pan S, Wang J, Dixon A, He J, Muller MG *et al.* (2013). Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nat Genet* **45**: 563–566.