The scaling of genetic diversity in a changing and fragmented world

Miguel Arenas[1,2,*], Stefano Mona[1,2], Audrey Trochet[3], Anna Sramkova Hanulova[1,2], Mathias Currat[4], Nicolas Ray[5,6], Lounes Chikhi[7,8,9], Rita Rasteiro[7,10], Dirk S. Schmeller[3,11] and Laurent Excoffier[1,2]

[1] Computational and Molecular Population Genetics Lab (CMPG), Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Berne, Switzerland.
[2] Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.
[3] Station d'Ecologie Expérimentale du CNRS à Moulis, USR 2936, 09200 Saint Girons, France.
[4] Anthropology, Genetics and Peopling History Lab, Anthropology Unit – Department of Genetics & Evolution, University of Geneva, Geneva, Switzerland.
[5] EnviroSPACE Laboratory, Institute for Environmental Sciences, University of Geneva, Geneva, Switzerland.
[6] Forel Institute, University of Geneva, Geneva, Switzerland.
[7] Instituto Gulbenkian de Ciências, Oeiras, Portugal
[8] CNRS, Université Paul Sabatier, ENFA, UMR 5174 EDB (Laboratoire Evolution et Diversité Biologique), Toulouse, France
[9] Université de Toulouse, UMR 5174 EDB, Toulouse, France
[10] University of Leicester, University Road, Leicester, LE1 7RH, United Kingdom
[11] Helmholtz Center for Environmental Research, Department of Conservation Biology, Permoserstrasse 15, 04318 Leipzig, Germany.
** Corresponding author*

Current Email addresses:

Miguel Arenas: marenas@cbm.uam.es
Stefano Mona: mona@mnhn.fr
Audrey Trochet: audrey.trochet@ecoex-moulis.cnrs.fr
Anna Sramkova Hanulova: anna.sramkova@iee.unibe.ch
Mathias Currat: Mathias.Currat@unige.ch
Nicolas Ray: Nicolas.Ray@unige.ch
Lounes Chikhi: chikhi@igc.gulbenkian.pt
Rita Rasteiro: rr147@leicester.ac.uk
Dirk S. Schmeller: ds@die-schmellers.de
Laurent Excoffier: laurent.excoffier@iee.unibe.ch

Chapter for the SCALES book.

Species living in a changing world

Most species do not live in a constant environment over space or time. Their environment is often heterogeneous with a huge variability in resource availability and exposure to pathogens or predators, which may affect the local densities of the species. Moreover, the habitat might be fragmented, preventing free and isotropic migrations between local sub-populations (demes) of a species, making some demes more isolated than others. For example, during the last ice age populations of many species have migrated towards refuge areas from which re-colonization originated when conditions improved. However, population extinctions may have occurred for populations that could not move fast enough or could not adapt to the new environmental conditions. Populations living in these types of dynamic environments are often referred to and modeled as metapopulations. Several studies have focused on the description of their demography, they probability of extinction and expected patterns of diversity at different scales. Importantly, all these evolutionary processes may affect genetic diversity, which can affect the chance of populations to persist. In this chapter we overview the consequences of fragmentation, long-distance dispersal, range contractions and range shifts on genetic diversity. In addition, we describe new methodologies to detect and quantify underlying evolutionary processes from sampled genetic data.

1. Spatial and temporal genetic simulation using SPLATCHE2

Computer simulations mimic the processes that occur in the real world and allow us to study which patterns may affect systems. We have developed the program SPLATCHE2 (http://www.splatche.com) (Ray et al. 2010), which performs spatially explicit simulations of genetic data under environmental variable scenarios and accounting for recombination, complex migration and long-distance dispersal. As input, the program requires a map (specified by a grid of demes) where the carrying capacity (K) and the migration rate must be user-specified for each deme. Optionally, both K and migration rate can change with time (moreover, anisotropic migration is also implemented). Other important inputs are related with demography (e.g., initial

population size and geographic origin, growth rate, generation time, total number of generations and a number of demographic models). Then, SPLATCHE2 performs a demographic simulation over the map followed by a coalescent simulation based on user-defined samples (Figure 1). The coalescent simulation just traces the evolutionary history of the sampled genes going backwards in time until their most recent common ancestor. It is followed by a simulation of genetic data (DNA, STRs and SNPs) along the coalescent (gene) genealogy. Although the model makes several assumptions (like a molecular clock or non-overlapping generations) it is probably one of the most realistic software available and has been used in a variety of important publications. Below we describe some important applications of this program for scaling genetic diversity under complex evolutionary scenarios.

2. Influence of habitat fragmentation on genetic diversity

Previous studies have suggested that environmental heterogeneity can affect genetic diversity, but these effects were not evaluated at different spatial scales. By using the results from extensive simulations, we address here the influence of fragmented habitats at different scales on the species genetic diversity. Using SPLATCHE2, we simulated range expansions where demes were partitioned into groups (patches) by adding barriers to dispersal. We also included scenarios with long-distance dispersal events, where individuals can migrate to non-neighboring demes. Then, samples were collected within demes, patches, regions and at the global landscape level.

As expected, we found that strong levels of fragmentation result in a severe loss of genetic diversity in the population at a global scale, but we also found that the detection of this decreased diversity requires sampling at different scales. Moreover, we varied fragmentation intensity at specific time points and we found that local genetic diversity and population differentiation were markedly affected by ancient fragmentation, and much less by recent events. Our results explain why recent habitat fragmentation does not always lead to detectable signatures in the genetic structure of populations. Conversely, if habitat fragmentation is removed, it also takes a long time to recover lost diversity by natural processes, suggesting that long-term conservation measures (e.g., by restoring gene flow) should be implemented to locally restore previously lost genetic diversity. We also found that species with long-distance dispersal abilities can however

migrate out of the barriers. As a consequence, their diversity is less influenced by the fragmented landscape.

3. Influence of range contractions and range shifts on genetic diversity

Range contractions and range shifts may occur as a consequence of temporal climatic fluctuations, depending on the geographical structure of the landscape, the duration of the climatic changes, or the species dispersal abilities. Under such environmental changes, a common response of species is migration towards more suitable regions. Many studies have analyzed the migration behaviour and spatial distribution of range-contraction and -shifting species, nevertheless less attention has been paid to the influence of such processes on genetic diversity. By using SPLATCHE2, we simulated DNA sequence data in populations suffering diverse range shifts and contractions over a landscape constituted by a grid of demes (Arenas et al. 2012). Simulated scenarios of range shifts and range contractions varied according to dispersal abilities and migration patterns. For example fast range contractions (e.g., as a consequence of a fast climate change) may lead to the extinction of populations that do not move. We analyzed genetic diversity of the simulated data. Contrary to our expectations, we found that fast contractions had less effect on genetic diversity. Fast contractions preserve higher levels of diversity and induced lower levels of genetic differentiation among refuge areas than slow contractions towards refuge areas. Thus slow contractions have the highest negative impact on final (low) levels of diversity. Contrastingly, fast range shifts lead to lower levels of diversity than slow range shifts. Interestingly, we found that species actively migrating towards refuge areas can actually bring additional diversity to these areas, but only if the range contraction is rapid. When contractions or shifts are slow, we found that active migrations towards refuge areas could lead to a more pronounced loss of diversity than if migration was isotropic (Arenas et al. 2012). These results suggest that species with different generation times and different migration abilities should be differently affected by environment change.

BOX 1. Effect of range contractions on principal component analysis of European molecular diversity

The genetic signal of range contractions (RCs) can be also observed in genetic gradients estimated by principal component analysis (PCA). Initial studies that represented genetic relationships among human populations with PCA revealed the presence of a southeast–northwest (SE-NW) gradient of genetic variation in current European populations, which was interpreted as the result of a diffusion process of early Neolithic farmers during their expansion from the Middle East. However, this interpretation has been largely questioned, as PCA gradients may occur even when there is no expansion, and because the first PC axis is often orthogonal to the expansion axis. However, the effect of more complex evolutionary scenarios on PCA, such as those including both range expansions and contractions, had not been investigated.

In a recent study, we (Arenas et al. 2013) have performed simulations of range contractions that might have occurred during the last glacial maximum period to better understand the formation of genetic gradients across Europe. In particular, we have simulated range contractions of human Paleolithic populations and admixture between Paleolithic and Neolithic populations over Europe (see Figure 1). The simulations were performed for diverse levels of admixture and under two range contraction scenarios where the refuge areas were either over all southern Europe or only in the Iberian Peninsula (see Figure 1). We observed that the first PC (PC1) gradients were orthogonal to the expansion axis, but only when the expansion was recent (Neolithic). More ancient (Paleolithic) expansions altered the orientation of the PC1 gradient due to 1) a spatial homogenization of genetic diversity over time, and 2) the exact location of the LGM refugia. Overall we found that PC1 gradients consistently follow a SE-NW orientation if there is a large Paleolithic contribution to the current European gene pool, and if the main refuge area during the last ice age was in the Iberian Peninsula. Our study suggests that the observation of a SE-NW PC1 gradient is compatible with the view that range contractions have affected observed patterns of genetic diversity, and suggest that the genetic contribution of Neolithic populations to the current European gene pool may have been limited (Figure 2). Although this study was focused on humans, this framework could be applied to other species that might have experimented range contractions as a consequence of environmental changes.

END BOX 1.

4. Inference of fragmentation levels from genetic data gathered at different scales over the species range

Populations living in a heterogeneous environment usually show a large variance in local population densities and migration rates, and generally present less local genetic diversity and higher levels of population differentiation than populations of similar size living in a constant and uniform environment. This is because genetic diversity is more rapidly lost in small demes than it is gained in large demes, leading to higher rates of local genetic drift.

Patterns of genetic diversity have been used to assess many properties of a population, but no attempt has been made to directly estimate the degree of environmental heterogeneity directly from patterns of diversity at different scales. It would therefore be useful to be able to infer the degree of environmental heterogeneity directly from genetic data, especially for sparse and cryptic species, or for species for which the exact definition of the population is difficult to assess.

We have simulated environmental heterogeneity using SPLATCHE2 where local deme carrying capacities (K) are drawn from a Gamma distribution with mean $\overline{K}$ and shape parameter α. Note that small values of α (typically $\alpha < 1$) are indicative of strong environmental heterogeneity, where a few demes have very high population densities and most others have very low densities (even being zero, which correspond to uninhabitable regions). Therefore, because habitat fragmentation usually creates uninhabitable regions, it is also associated to high levels of environmental heterogeneity. On the other hand, large values of α (typically $\alpha > 5$) imply little environmental heterogeneity, such that most demes have a very similar carrying capacity. Previous studies have shown that both local genetic diversity and levels of population differentiation would strongly depend on α, suggesting that patterns of genetic diversity at different scales could be used to infer α, and therefore, indirectly, the level of environmental heterogeneity.

We used an Approximate Bayesian Computation framework to infer the shape parameter of a Gamma distribution directly from patterns of genetic diversity of several samples taken from a population having gone through a recent range expansion. Our results show that the degree of environmental heterogeneity (α) can be very well estimated if all other parameters of the model are known (Figures 3). When all other parameters need to be co-estimated, the estimation of α becomes difficult, and we can

mainly distinguish small from large α values (Figure 4). In other words, we have only power to distinguish very heterogeneous environments from more homogeneous ones, but little prospect to get accurate estimations of α.

BOX 2. Sex-biased dispersal

Population genetic structure is influenced by migration patterns. This includes sex-biased dispersal, which is frequent in many species, likely impacting life-history evolution, population genetic structure and metapopulation functioning. In population genetics, sex-biased dispersal may not only reflect a difference in the number of dispersing individuals of one sex in relation to the opposite sex, but also the unequal reproductive success of dispersers. Fine-scale genetic structure and adaptation to local environments might therefore be promoted by sex-biased dispersal. Sex-biased dispersal can be identified and quantified to some extent by comparing the genetic differentiation of females to that of males, and by looking at sex-related genetic systems. The sex with the highest dispersal frequency would have a lower genetic differentiation among different subpopulations (i.e. as measured by the genetic parameter F_{ST}). Similarly, sex-biased dispersal could be measured by comparing the level of genetic structure inferred from nuclear markers (inherited by both parents) to that indicated by mtDNA (which is transmitted only by females) or Y chromosome (transmitted only by males). If the level of genetic differentiation inferred from mtDNA is higher than that inferred from nuclear markers, male-biased dispersal may be assumed. Simulations, undertaken with a different program inspired by SPLATCHE2 (Rasteiro et al. 2012), clearly show that different patterns of genetic differentiation can be detected under three scenarios, 1) bilocality (no sex-biased dispersal), 2) matrilocality (male-biased dispersal), and 3) patrilocality (female-biased dispersal, Figure 5). Y-chromosome genetic diversity is very low, especially in the patrilocality scenario for which only one Y-haplotype often remain after 1000 simulated generations. Note that the same effect was not seen in simulated mtDNA, probably due to differences in mutation rates and types of markers (Rasteiro et al. 2012). Indeed, the authors showed that the simple difference in mutation rates between the two types of sex-related genetic systems is sufficient to create an asymmetry that could be mistaken for differences in migration rates, even under bilocality scenarios.

Accounting for sex-biased migration in population and conservation genetics studies is of great importance as significant differences in sex-biased dispersal have been demonstrated among different taxonomic groups. Dispersal of mammals, reptiles and fishes were more frequently male-biased whereas dispersal in birds was more frequently female-biased (Figure 6). Therefore, knowledge on sex-biased dispersal may prove essential to develop and assess habitat management and landscape planning strategies for different species.

In many species population decline has been linked directly to loss and fragmentation of habitats and indirectly to reduced inter-patch dispersal. Concerns about habitat fragmentation and landscape structure are usually based on the ability of wildlife to disperse between the blocks of habitat types that they require. Our simulations showed that patterns of sex-biased dispersal can have important consequences on some genetic markers and conversely they should inform us on the importance of sex-biased dispersal in natural systems that are difficult to study. Some studies have suggested that the different sexes may have a differing impact on demographic connectivity at different scales, the less dispersing sex more on local scales, while the more and farther dispersing sex on larger scales. Another consequence of sex-biased dispersal is that the rate of natural recolonization of locally extinct populations may be slower as it requires that both sexes disperse. Sex-biased dispersal may also act as a buffer against reduction of genetic variability due to high genetic drift in populations with small effective size (Schmeller and Merila 2007). Ultimately, explorations of the implication of unequal effective population size, migration rate and non-random individual dispersal will be necessary for synthesizing ecological and genetic theory on dispersal and population structure.

END OF BOX 2

6. Concluding remarks

In this chapter we described the strong influences that habitat fragmentation and dispersal heterogeneity can have on genetic diversity, at different geographical and temporal scales. To this purpose, we mainly used the SPLATCHE framework to perform spatially explicit simulations of genetic diversity under complex demographic models, also allowing for temporal heterogeneity. We found that fragmented habitats often have a significant loss of genetic diversity relative to homogeneous habitats. This

effect was reduced in species with long distance dispersal abilities. Similarly, range contractions led to a loss of genetic diversity, in particular when the contraction was slow. As a consequence, the generation time of species needs to be taken into account when considering genetic diversity after climatic changes. Species with shorter generation times should suffer from more diversity loss after a range contraction than long lived species (Arenas et al. 2012). We note however, that such species may also adapt faster to new environments. Fast range shifts, on the contrary, reduced genetic diversity more than slow range shifts where more individuals can track favorable environments. Indeed species with low migration rates and going through fast range shifts can easily become extinct (Arenas et al. 2012). In addition, we found signatures of range contractions on diversity by using PCA. In this case, a re-expansion after a range contraction introduces spatial genetic diversity gradients that depend on the location of refuge areas (Arenas et al. 2013). We also described a procedure to detect the level of habitat fragmentation from observed patterns of genetic diversity. Finally, we performed simulations incorporating sex-biased migration and found that such a bias could highly impact genetic data, which can therefore be used to infer sex-biased dispersal in species that are difficult to study in the field. The fact that habitat fragmentation, dispersal patterns, and range movements strongly alter genetic diversity of species implies that they need to be considered for biodiversity conservation strategies.

Acknowledgements

This work was supported by a EU/FP7 grant N° 226852 (SCALES). LC was supported by the Laboratoire d'Excellence (LABEX) entitled TULIP (ANR-10-LABX-41) and by the FCT project ref. PTDC/BIA-BEC/100176/2008.

References

ARENAS M, FRANCOIS O, CURRAT M, RAY N, EXCOFFIER L (2013) Influence of admixture and paleolithic range contractions on current European diversity gradients. *Mol. Biol. Evol.* 30**:** 57-61.

ARENAS M, RAY N, CURRAT M, EXCOFFIER L (2012) Consequences of Range Contractions and Range Shifts on Molecular Diversity. *Mol. Biol. Evol.* 29**:** 207-218.

PIAZZA A, RENDINE S, MINCH E, MENOZZI P, MOUNTAIN J *et al.* (1995) Genetics and the origin of European languages. *Proc Natl Acad Sci U S A* 92**:** 5836-5840.

RASTEIRO R, BOUTTIER PA, SOUSA VC, CHIKHI L (2012) Investigating sex-biased migration during the Neolithic transition in Europe, using an explicit spatial simulation framework. *Proc Biol Sci* 279**:** 2409-2416.

RAY N, CURRAT M, FOLL M, EXCOFFIER L (2010) SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* 26**:** 2993-2994.

SCHMELLER DS, MERILA J (2007) Demographic and genetic estimates of effective population and breeding size in the amphibian Rana temporaria. *Conserv. Biol.* 21**:** 142-151.

Figures

Figure 1. Timeline simulation of complex scenarios of range expansion, range contraction and posterior re-expansion. Each plot corresponds to a snapshot of the program SPLATCHE2. White areas indicate unoccupied demes while green areas represent occupied demes. Snapshots presented at each line differ in 50 generations, see detailed settings in (Arenas et al. 2013). At the top, we describe a range expansion over Europe from the Near East. Then, we show a range contraction from the north to the south, which mimic the LGM period and leads to two refuge areas (all south of Europe (A) and only the Iberian Peninsula (B)). At the bottom, we show a re-expansion from both refuge areas.

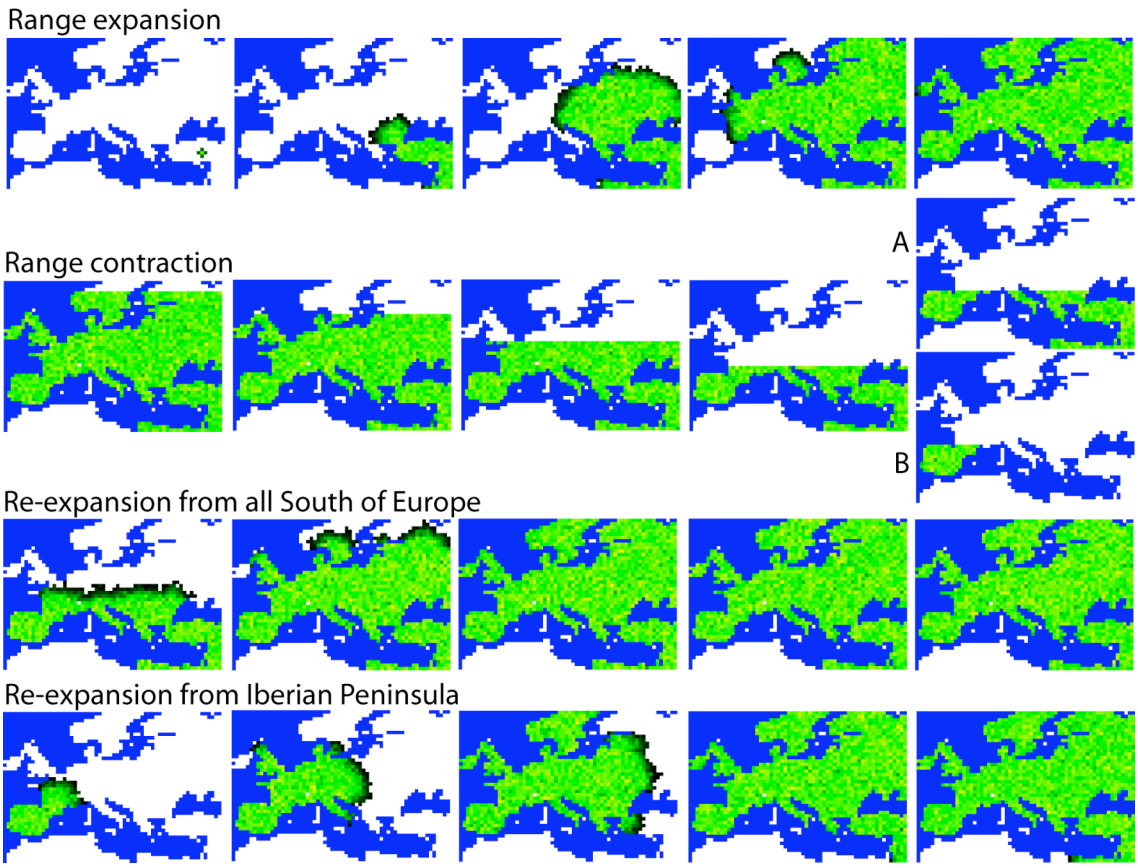


Figure 2. Influence of range contraction on PCA maps. SNP data PC maps for admixture of Neolithic (20%) and Paleolithic (80%) range expansions from Middle East. (A) Illustrative example of PCA derived from a range expansion. The PC1 gradient has a SW-NE orientation. (B) Illustrative example of PCA derived from range expansion followed by a RC towards all south of Europe, and posterior re-expansion. The PC1 gradient has an E-W orientation. (C) Illustrative example of PCA derived from range expansion followed by a RC towards only the Iberian Peninsula, and posterior re-expansion. The PC1 gradient has an NW-SE orientation. (D) Original PC1 map inferred from Piazza et al. (1995) [© 1995 National Academy of Sciences, USA] with a superimposed line connecting positive and negative PC1 centroids. The PC1 gradient shown in (C), which is the most similar to such finding from real data (D), was also found in scenarios with larger Paleolithic contribution and either pure range expansions or range expansions with RC towards the Iberian Peninsula (further details in Arenas et al. 2013).

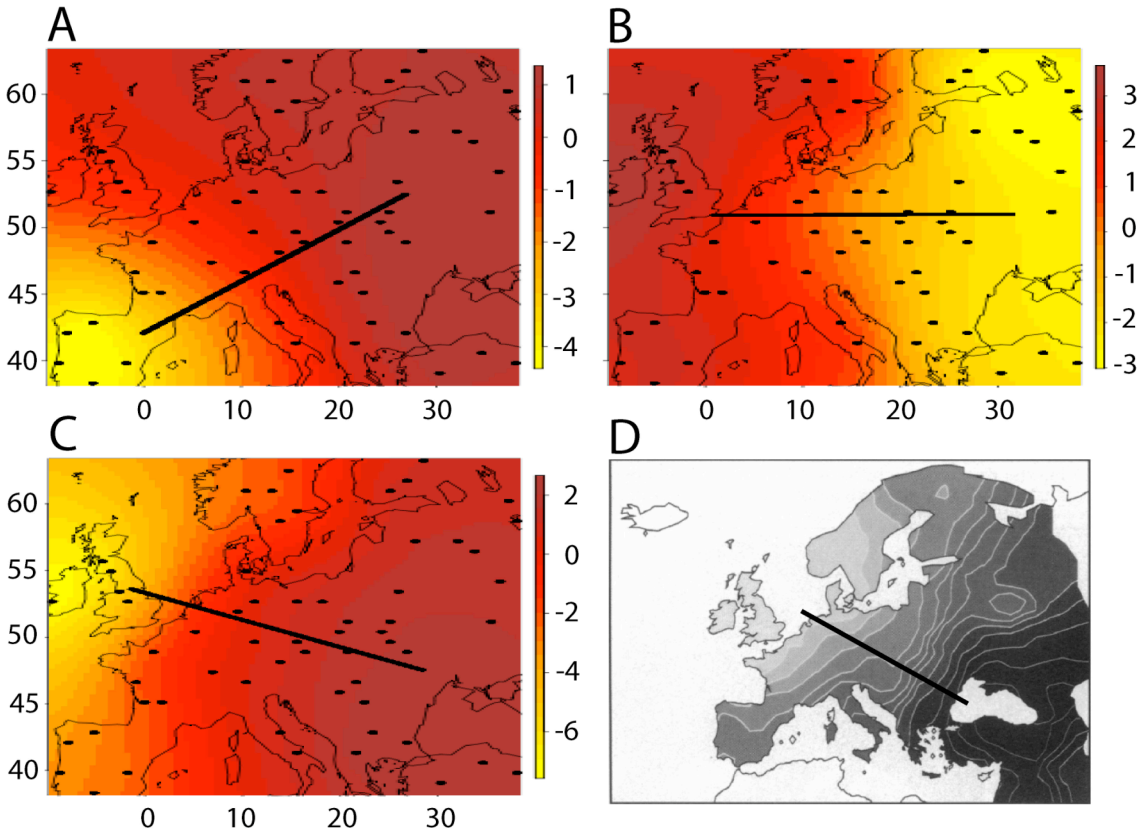


Figure 3. ABC estimation of α from genetic diversity simulated in species with small and large carrying capacity average when all other parameters of the model are known.

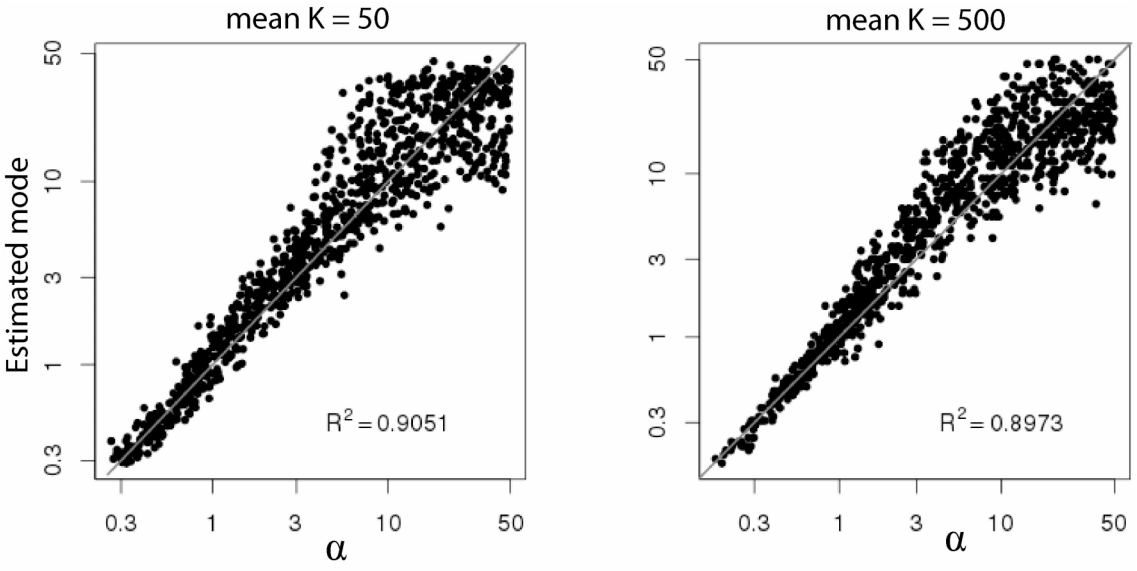


Figure 4. Optimal distinction between small and large α values when all parameters of the range expansion model need to be co-estimated with the environmental heterogeneity. The plot shows the estimated proportion of times where α was incorrectly estimated as below or above a threshold (a given true value). This incorrect assignment is minimized for α=0.63 (blue line), showing a maximal power to distinguish between values of α above and below this value. Here, the misclassification rate is inferred from an analysis of the plot of true (x-axis) vs. estimated (y-axis) α values shown in the central insert. Misclassification rate is obtained as the sum of the proportion of points in the blue regions relative to those in the orange regions on the left and right hand side of the blue line.

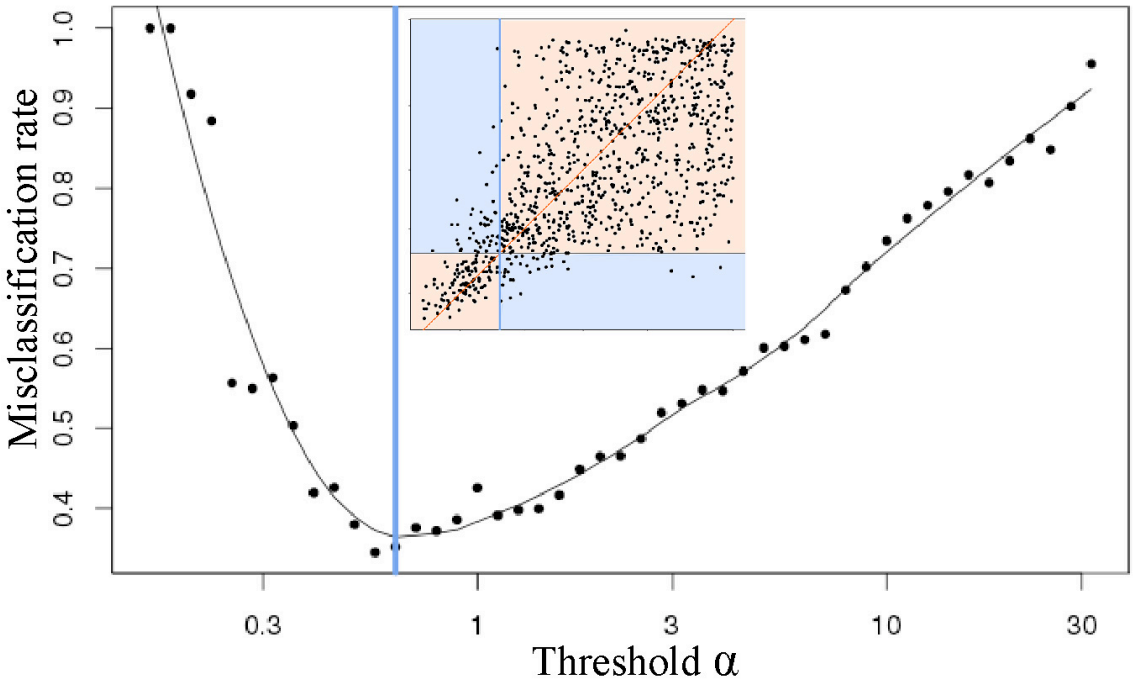


Figure 5: Genetic differentiation patterns under sex-biased migration patterns. Simulations were performed using a forward simulation program similar to SPLATCHE2. A square environment of 400 demes (20x20) was simulated under three scenarios, 1) bilocality (no sex-biased dispersal), 2) matrilocality (male-biased dispersal), and 3) patrilocality (female-biased dispersal). For each scenario we simulated independent autosomal loci, Y and X chromosome and mtDNA sequences. For each scenario and genetic marker type we computed a measure of genetic differentiation between demes at increasing distances. For simplicity only demes from the diagonal were used and compared to the same deme located in one of the corners (deme 19,19). As the panels show, sex-biased migration has a strong impact on the overall level of genetic differentiation, and on the differences between markers. The results also show that mtDNA and Y chromosome markers do not necessarily play symmetrical roles in the patrilocality and matrilocality scenarios because they differ also in mutations rates, as noticed by Rasteiro et al. (2012).

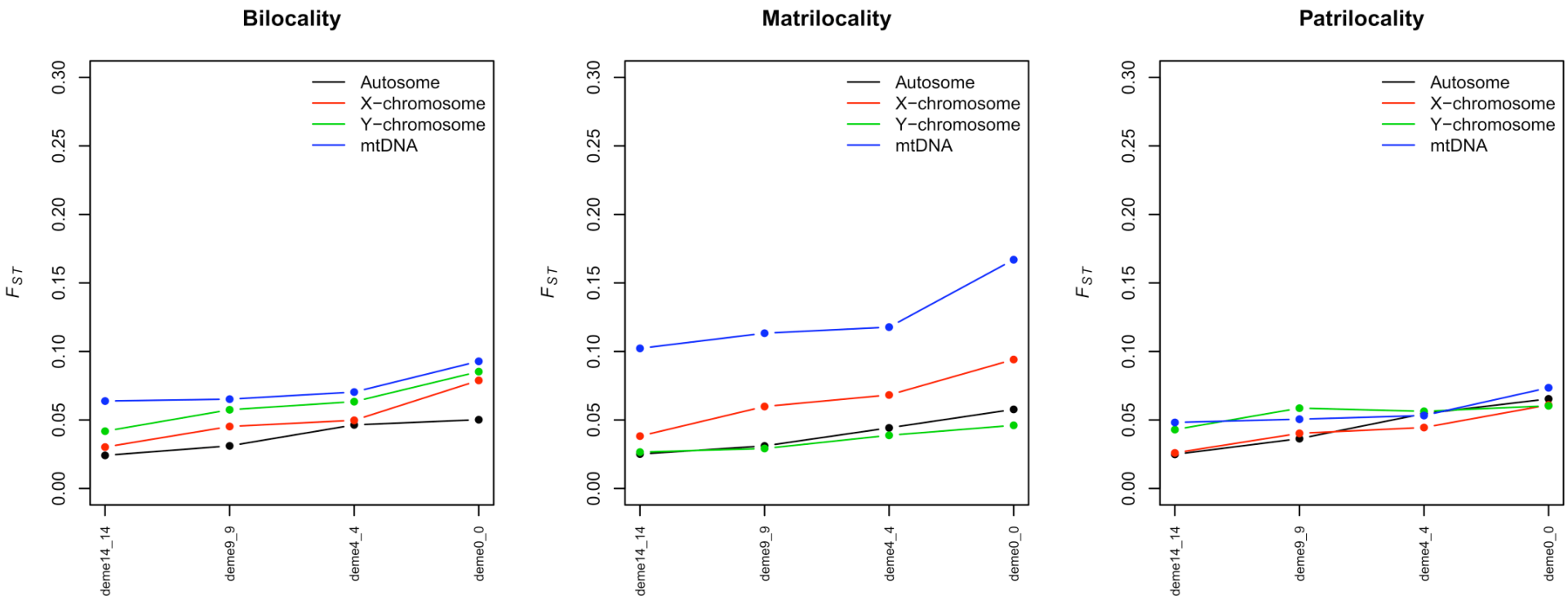
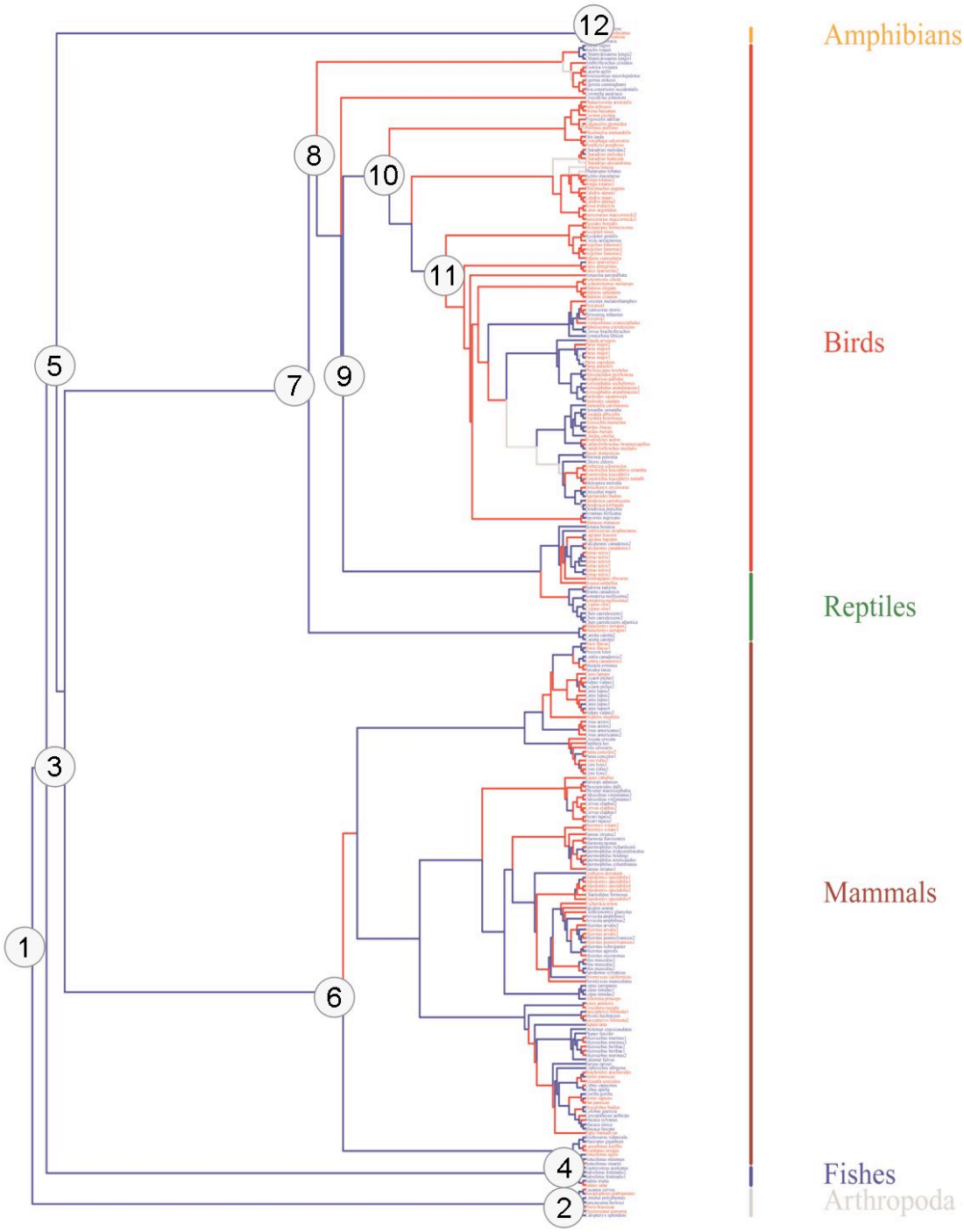

Bilocality
Matrilocality
Patrilocality
Autosome
X−chromosome
Y−chromosome
mtDNA
F_{ST}
0.00
0.05
0.10
0.15
0.20
0.25
0.30
deme14_14
deme9_9
deme4_4
deme0_0

Figure 6. Phylogenetic tree of the ancestral character states reconstruction of sex-biased dispersal based on a parsimonious method on the 216 species (275 *populations* from publications) used. Branches and tips are coloured in blue for a male biased dispersal state and in red for a female biased dispersal state. In grey, branches for which the reconstruction method did not allow one to choose between a male or a female bias. Numbers on nodes correspond to: 1. Bilateria, 2. Arthropoda, 3. Osteichthyes, 4. Fishes, 5. Tetrapoda, 6. Mammals, 7. Amniota, 8. Sauria, 9. Neognathae, 10. Neonaves, 11. Birds, 12. Batrachia.


12
8
10
11
5
7
9
3
1
6
4
2
Amphibians
Birds
Reptiles
Mammals
Fishes
Arthropoda