



Population Structure

Guillot, Gilles; Orlando, Ludovic

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Guillot, G., & Orlando, L. (2013). *Population Structure*.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

POPULATION STRUCTURE

Gilles Guillot

Technical University of Denmark

Ludovic Orlando

Natural History Museum, Copenhagen University

INTRODUCTION

When considering a population of organisms with sexual reproduction, assuming that mating takes place at random can be convenient from a modelling and computational point of view but it is often not realistic as this may be impeded by various pre- or post-zygotic barriers. The term population structure (or population subdivision) usually refers to the patterns in neutral genetic variation that result from the past or present departure from panmixia of a population. Its study is of interest to evolutionary biologists because it can provide insight into the causes and consequences of restricted gene flow in natural populations. The general populations genetics background is presented by Hartl and Clark 1988 and Nielsen and Slatkin 2013, connection to ecology is presented by Hanski and Gaggiotti 2004 while more detailed investigation of the role of geography is investigated by Epperson 2003 and Rousset 2004. Population structure plays also a central role in biomedical applications because it acts as a confounding factor in the study of the statistical association between a genetic variant and a phenotype. This (and many other aspects) are presented by Balding et al. 2007. Question specific to conservation biology are treated by Allendorf et a. 2012.

The topic of population structure relates to several other questions covered by the present series of commented bibliographies, in particular Landscape Ecology, Conservation Genetics, Population Genetics, Geographic variation, Phylogeography, Interpretation of Phylogenetic Trees, Metapopulations & Spatial Population Processes, Hybrid Zones, and lastly Dispersal.

General overviews

Hartl DL, Clark AG, Principles of Population Genetics. Sinauer Associates, Sunderland, Massachusetts, 1988.

A reference text book on basic population genetics, including on population subdivision. Suitable for any undergraduate students in evolutionary biology.

Epperson, B.K. Geographical Genetics, Princeton University Press, 2003.

Reviews methods and applications on spatially structured populations with a focus on simulation-based inference methods.

F. Rousset, *Genetic Structure and Selection in Subdivided Populations*, Princeton University Press 2004.

Summarizes theories used to describe geographically structure population. Written in a fairly mathematical style. Suitable for graduate students in evolutionary biology or statistics.

I.A Hanski, O. Gaggiotti Eds, *Ecology, Genetics and Evolution of Metapopulations*, 2004. Elsevier.

An overview of various aspects of population structure from an ecological perspective.

D. Balding, C. Cannings, M. Bishop, Eds. 2007. *Handbook of Statistical Genetics*, 3rd edition, 1392 pp., John Wiley & Sons.

An account of the state of the art on the many aspects of populations genetics, with statistical development. Includes a chapter on Inferences from Spatial Population Genetics by F. Rousset and a chapter on Analysis of Population Subdivision by L. Excoffier. The concise style imposed by the book format and the statistical content makes it best suited to graduate students and researchers.

Allendorf, F. W., G. H. Luikart, and S. N. Aitken. *Conservation and the genetics of populations*, 2nd edition, Wiley, 2012.

A mine of information in genetics and ecology in a clearly written style with a nice balance between data and methods. A lot of material about population structure, its causes, consequences and how it can be dealt with from a conservation perspective.

Nielsen, R and Slatkin, M., 2013 *An introduction to population genetics*, Sinauer.

A modern and clearly written introduction to population genetics with concrete examples. Provides introduction to key concepts in an intuitive yet rigorous way without mathematical details.

Defining population structure

The term population structure is defined negatively as a departure from random mating and can therefore correspond to different processes and patterns. Departure from panmixia can be due to the existence of demes (such as flocks, colonies or herds) with restricted within- and between-demes gene flow (Manel et al .2005, Wapples and Gaggiotti 2005). In a limiting case, a deme consists of a single individual, and reproduction occurs preferentially between geographically close individuals, a situation leading to a continuous population structure (Guillot et al. 2009). Modelling aspects are discussed by Hey and Machado 2003, and Lawson 2013.

Hey, J. and Machado, C.A., The study of structured populations – new hope for a difficult and divided science. *Nature Review Genetics*, 4, 535-543, 2003.

A review that outlines the main models of population structure and focuses on methods

for detecting subpopulations and their recent common history.

Manel S, Gaggiotti O, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, 20, 136–142.

Waples R. and O. Gaggiotti, What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity, *Mol Ecol*, 15, 1419-1439, 2006.

The two papers above discuss the various concepts associated to demic population structure, give an overview of models and algorithms for statistical inference and assess the accuracy they can achieve.

G. Guillot, R. Leblois, A. Coulon, A. Frantz 2009. Statistical methods in spatial genetics. *Molecular Ecology*.

A review with a focus on statistical inference of geographical structure.

D. Lawson, Populations in statistical genetics modeling and inference, book chapter for the interdisciplinary book "What is a Population", Oxford University Press, 2013 (submitted)

The most recent review paper. Includes a discussion of methods for the detection of structure for large SNP data.

Journals

The main journals in the fields are *Genetics*, *PLOS Genetics*, *Molecular Biology and Evolution*, *Molecular Ecology* and *Molecular Ecology Resources*. Other journals relevant to the topic of population structure include *Applied Ecology*, *The American Naturalist*, *Bioinformatics*, *Conservation Genetics*, *Evolution*, *Genome Research*, *Heredity*, *Methods in Ecology and Evolution*, *Nature Genetics*, *Nature Reviews Genetics*, *Systematic Biology* and *Trends in Ecology and Evolution*.

Genetics

Edited and published by the Genetics Society of America. Has published the seminal papers by the fathers of modern population genetics; has long been considered the leading journal in the field.

PLOS Genetics

A peer-reviewed, open-access, electronic-only journal that publishes contributions in all areas of genetics. Has played a key role in the publication of the most innovative method papers related to population structure lately.

Molecular Biology and Evolution

A monthly peer-reviewed scientific journal edited by the Society for Molecular Biology and Evolution and published by Oxford University Press. It publishes work in the intersection

of molecular biology and evolutionary biology.

Molecular Ecology and Molecular Ecology Resources

Molecular Ecology is a twice monthly scientific journal covering investigations that use molecular genetic techniques to address questions in ecology, evolution, behavior, and conservation. Its sister journal Molecular Ecology Resources publishes articles on technical methods, computer programs and genomic resource development.

Historical perspective

The seminal contributions are due to Wright 1922, 1931 and 1943.

Wright, S. (1922). Coefficients of inbreeding and relationship. American Naturalist 56, 330-338.

Introduces the fixation index F .

Wright S. (1931) Evolution in Mendelian populations. Genetics, 16, 97-159.

Introduces and analyses a set of assumptions now known as the Wright-Fisher model which became a corner stone of population genetics and on which more complex models can be built.

Wright, S. Isolation by distance. Genetics 28, 114-138 (1943).

Introduces the first spatially explicit model of geographically structure populations

STATISTICAL INFERENCE ABOUT POPULATION STRUCTURE

The tasks involved in the study of population structure can be split into (i) quantifying and testing the existence of a structure among various pre-defined sub-populations, (ii) detecting the presence of 'pure' or admixed sub-populations, (iii) inferring the evolutionary history of pre-defined populations (in form of a series of splitting and merging events, with possible variation of population sizes and amount of gene flow) and lastly, (iv) jointly estimating the presence of sub-populations and their history.

There has been a permanent tension in statistical genetics between the interest of developing increasingly complex models to extract as much information as possible from the data on the one hand, and on the other hand the recognition that complex models lead to computational difficulties, especially when implemented on datasets of ever growing sizes. This is reflected by the existence a various methods ranging from description and quantification (estimation of F statistics or factor analysis methods), to complex models based on coalescent theory. In between stand clustering models which achieve a trade-off between biological realism and inference complexity. They borrow to the mainstream statistics literature but also incorporate features of biologically-inspired

models.

Measuring and testing structure

This aspect is reviewed by Weir and Hill 2002 and Holsinger and Weir 2009.

Weir BS, Hill WG (2002) Estimating F statistics. Annual Review of Genetics, 36, 721–750.
Establishes the standard method to estimate F statistics from data.

Holsinger, K.E. & Weir, B.S. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat. Rev. Genet. 10, 639–650 (2009).

An old and useful concept reviewed with a modern perspective.

Isolation by distance models

There is a long tradition of using stochastic models to describe isolation by distance. Probability models of isolation by distance with analytical results include those introduced by Malécot 1975 and Slatkin and Maruyama 1975, Rousset 1997 and Rousset 2000. Epperson 1995 and Epperson and Li 1997 report results based on simulations. Overviews can be found in Hardy and Vekemans 1999 and Charlesworth et al. 2003.

Malécot, G. 1975. Heterozygosity and relationship in regularly subdivided populations. Theor. Popul. Biol. 8:212–241.

Slatkin, A. M., and T. Maruyama. 1975. The influence of gene flow on genetic distance. Am. Nat. 109:597–601.

Some of the earliest attempts to compute probabilities of identity of gene lineages or correlations in allele frequencies as a function of distance.

Epperson B (1995) Spatial distribution of genotypes under isolation by distance. Genetics, 140, 1431–1440.

Epperson B, Li T (1997) Gene dispersal and spatial genetic structure. Evolution, 51, 672–681.

On the basis of simulated data, describes the expected patterns in spatial genetic variation under some models of isolation by distance.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics, 145, 1219–1228.

Propose a method to infer dispersal parameters based on the analysis of variation of the pairwise population differentiation with geographic distance.

Hardy O, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity, 83, 145–154.

Investigate theoretically the link between population genetics model of isolation by distance and more descriptive approaches based on spatial auto-correlation functions.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, 13, 58-62.

Develops earlier results to make inference of dispersal parameters from data sampled 'continuously' in space instead of in form of sub-populations.

Charlesworth B., D. Charlesworth and N. H. Barton, 2003, The effects of genetic and geographic structure on neutral variation, *Annu. Rev. Ecol. Evol. Syst.* 34:99-125

A review article and key reference on the subject.

Detecting demic structure

The most powerful methods to detect demic structure do not attempt to resolve the population history, but seek instead either to estimate admixture proportions or to compute principal components which in both cases synthesizes the population history. The various methods differ by the assumption made about the dependency between loci (with or without linkage equilibrium), the possibility to consider admixed individuals, inbreeding and the technique used to estimate parameters.

Lawson, D. and D. Falush. Similarity matrices and clustering algorithms for population identification using genetic data. *Annual Review of Human Genomics*, 13: 337-361, 2012

The most recent paper to date that reviews and compare methods to perform inference of population structure.

Bayesian mixture models

Bayesian mixture models originate from mainstream statistics. They assume the existence of extant or past clusters at Hardy-Weinberg equilibrium that are characterized by cluster-specific allele frequencies. These models were pioneered by Pritchard et al. 2000, Dawson and Belkhir 2001 and Falush et al. 2003.

Pritchard, J.K., M. Stephens and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics* 155.2, 945-959, 2000.

A landmark paper that introduces a Bayesian model-based approach to cluster individuals into populations assuming Hardy-Weinberg equilibrium and linkage equilibrium. Method implemented in the widely used computer program Structure.

Dawson, K.J. and Belkhir, K. (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genet. Res.* 78, 59-77

Corander, J. et al. (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* 163, 367-374

These two papers tackle the same problem as the paper above by Pritchard et al., with a

slightly different computational strategy and the extra difficulty to estimate the number of clusters directly. Methods implemented in the programs Partition and BAPS.

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

Generalizes the model implemented in the Structure program to account for admixture linked disequilibrium and correlated allele frequencies.

Demic structure with inbreeding

The methods above have been modified by François et al. 2006 and Gao et al. 2007 to account for departure from Hardy-Weinberg equilibrium due to within-cluster inbreeding.

François O, Ancelet S, Guillot 2006. G. Bayesian clustering using hidden Markov random fields. *Genetics*, 174, 805–816.

Proposes an alternative to the spatial component of the Geneland program and also generalizes the likelihood model to take into account inbreeding.

Gao, H., S. Williamson and C. D. Bustamante, . A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, 176.3, 1635-1651, 2007.

An alternative approach to the method above (for genetic data only). Method implemented in the Instruct program.

Spatial models

The power of the methods above can be increased by embedding them in a spatially explicit model that makes use of in the information contained in the location of the individuals sampled. The non-admixture case is considered by Guillot et al. 2005 and an admixture model is presented by Durand et al. 2009. The various spatial models are discussed and compared by Safner et al. 2011 and Blair et al. 2012.

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

A spatially explicit model to make use of geo-referenced genetic data and enable increased power in statistical inferences. Method implemented in the Geneland program.

Durand, E., F. Jay, O. Gaggiotti, and O. Francois, Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution* 26: 1963–197, 2009.

Introduces a statistical model of spatial variation of admixture coefficients

Safner T, Miller MP , Mcrae B, Fortin MJ, Manel S (2011) Comparison of Bayesian

clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences*, 12:865-889.

Blair C, Weigel DE, Balazik M, Keeley AT, Walker FM, Landguth E, Cushman S, Murphy M, Waits L, Balkenhol N.. A simulation-based evaluation of methods for inferring linear barriers to gene flow, *Mol. Ecol. Resour.* 2012

Two comparisons of the various computer programs used in evolutionary biology to infer demic population genetic structure.

Likelihood-based inference

Many of the models for the detection of demic structure discussed above are based on a Bayesian model which is a statistical way to inject information about biological processes by way of probability distributions. This approach proves to be too computer demanding when the number of genetic markers exceeds a few hundreds. An alternative strategy consists in optimizing a likelihood function which is the approach taken by Tang et al. 2005, Tang et al. 2006 and Alexander et al. 2009.

Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology* 28: 289-301.

A model reminiscent of the approaches above, but based on likelihood than posterior distribution. Method implemented in the Frappe program.

Tang H, Coram M, Wang P, Zhu X, Risch N. Reconstructing genetic ancestry blocks in admixed individuals. *Am J Hum Genet.* 2006; 79 (1): 1-12.

A fast likelihood-based approach to infer ancestry blocks. Implemented in the Saber program.

Alexander, D. H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655- 1664.

Proposes an improved computational method for likelihood-based methods of inference of admixture coefficients. Method implemented in the Admixture program.

Phased data

Model based clustering algorithms described above work with un-phased data and do not make use of the information about linkage disequilibrium. This is tackled by the recent work by Lawson et al. 2012.

Lawson, D. J., G. Hellenthal, S. Myers, and D. Falush, 2012 Inference of population structure using dense haplotype data. *PLoS Genetics* 8: e100245.

Makes use of the information available in haplotypes obtained with phasing algorithm. Provides information about fine scale population structure with unprecedented accuracy

The new standard for massive SNP datasets.

Factor analysis models

Bayesian models have been the reference methods for the analysis of micro-satellite data until the advent of SNP datasets for which they showed some limitations due to computing times. This has been addressed by the use of principal component analysis (PCA) methods, for example by Patterson et al. 2006 and Engelhart and Stephens 2010. The decrease in computing is made at the expense of the ease of interpretation of analysis outputs, an aspect discussed by Reich et al. 2008, Novembre and Stephens 2008 and McVean 2009.

Patterson N, Price A, Reich D 2006 Population structure and eigen analysis. *PLoS Genetics*, 2, 2074–2093.

Shows how population structure can be investigated and tested through classical principal components methods at the era of large SNP datasets. Method implemented in the Eigensoft program.

Reich D, Price A, Patterson N. 2008 Principal component analysis of genetic data. *Nature Genetics*, 40, 491–492.

Novembre, J., and M. Stephens, 2008. Interpreting principal component analyses of spatial population genetic variation. *Nature genetics*, 40.5 646-649.

McVean, G., 2009 A Genealogical Interpretation of Principal Components Analysis. *PLoS Genetics* 5(10): e1000686.

Discuss the interpretation of principal components in population genetics in terms of admixture events or population migration events.

Engelhardt BE, Stephens M (2010) Analysis of population structure: a unifying framework and novel methods based on sparse factor analysis. *PLoS Genet* 6: e1001117. doi:10.1371/journal.pgen.1001117.

Develops a framework bridging the gap between cluster models and principal components approaches.

Inferring history of populations

The topic of population structure touches upon the important and vast topic of phylogenetics. Recent review articles include those of Knowles 2009 and Nielsen and Beaumont 2009. Estoup and Guillemaud 2009, Arenas et al. 2012, Excoffier et al. 2009, Francois et al. 2010 investigate questions related to population expansion while Choi and Hey 2011 and Pickrell and Pritchard 2012 develop new statistical methods to infer population history.

Knowles, L. L., 2009 Statistical Phylogeography. *Annual Review of Ecology, Evolution, and*

Systematics 40: 593-612.

A review of methods aimed at understanding the present and past structure of populations on a time and space scale coarser than that considered in landscape genetics.

Nielsen R, Beaumont MA. 2009 Statistical inference in phylogeography. *Mol. Ecol.* 18:1034-47 .

A more statistically oriented review with view to the nested clade analysis method controversy.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10) (2009):e1000695.

Although *δaδi* is primarily dedicated to demographic inference, this software can handle multi-population models and estimate from the allele frequency spectrum migration parameters and admixture stages across populations.

Estoup A, Guillemaud T 2010 Reconstructing routes of invasion using genetic data: why, how and so what? *Mol Ecol* 19: 4113-4130.

Description of distance-based and model-based inference methods for reconstructing the routes of introduction and expansion of invasive species, with reference to examples from the literature.

Choi, Sang Chul, and Jody Hey.2011. Joint inference of population assignment and demographic history. *Genetics* 189.2: 561-577.

The first method to date aimed at jointly detecting the presence of sub-population and estimating their common history.

Pickrell JK, Pritchard JK 2012 Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet* 8(11): e1002967.

An alternative approach to the method of Choi et al. above suitable for the analysis of large SNP datasets. Method implemented in the Treemix program.

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* 186(3) (2010):983-95.

Simulation studies showing the difficulty in disentangling population structure and demographic bottleneck signatures, providing guidelines and solution for correct inference in a conservation context.

Arenas, M, Ray, N, Currat, M, and Excoffier, L. 2012. Consequences of range contractions and range shifts on molecular diversity. *Mol Biol Evol* 29(1):207-18.

Investigates the consequences of range contractions and shifts on pattern of molecular

diversity. Discovers that the latter will be shaped by the species' dispersal abilities as well as by the speed of the environmental change.

Excoffier, L., M. Foll and R. J. Petit, 2009 Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* 40: 481-501.

Slatkin, M, and Excoffier, L. 2012 Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics* 191(1):171-81.

Francois, O, Currat, M, Ray, N, Han, E, Excoffier, L, Novembre, J. 2010. Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol* 27(6):1257-68.

Introduce allele surfing, a phenomenon that occurs at the wavefront of range expansion and that has consequences on geographic patterns of allele-frequency differentiation.

Detecting admixture and structure disruption

Admixture between differentiated populations can occur and disrupt pre-existing population boundaries. Recent interest in admixture has grown considerably with the discovery that non-African human genomes harbour detectable levels of admixture with archaic hominins such as Neanderthal and Denisovans.

Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28(8) (2011):2239-52

Introduces the D-statistics that exploits any excess of shared derived polymorphism between groups as a test for admixture.

Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., and Siepel, A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics* 43:1031-1034.

gPhoCS is a Bayesian software using Markov Chain Monte Carlo to jointly sample model parameters and genealogies for inferring ancestral population sizes, population divergence times, and migration rates from individual genome sequences.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. 2012. Ancient admixture in human history. *Genetics* 192(3):1065-1093.

This article presents the theories underlying the ADMIXTOOLS package and a number of statistical tests for detecting admixture from genome-wide data.

USING STRUCTURE TO DETECT THE ORIGIN OF INDIVIDUALS

If some information about allele frequencies of population of known origin is available, one can characterize existing structure and use this information to assign individuals of unknown ancestry. This is the approach taken by Rannala and Mountain 1997, Cornuet et al. 1999 and Anderson and Thompson 2002. Acknowledging that a sample of unknown of

origin may not have direct ancestry in any of the population of known geographic origin, Wasser et al 2004 developed a method for continuous spatial assignment. Yang et al. 2012 proposed an alternative models to deal with large SNP data. A review from an epidemiology perspective is provided by Remais et al. 2013.

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

The first method to use a formal Bayesian method to assign individuals to some known sub-populations.

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

A comparison showing that Bayesian methods outperforms other existing methods.

Anderson, E.C. and Thompson, E.A. (2002) A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160, 1217–1229

Generalizes existing methods to deal with hybrid individuals

Wasser S, Shedlock A, Comstock K, Ostrander E, Mutayoba B, Stephens M (2004) Assigning African elephants DNA to geographic region of origin: applications to the ivory trade. Proceedings of the National Academy of Sciences, 101, 14847– 14852.

A method to perform continuous spatial assignment of individuals.

Remais J.V., Xiao N., Akullian A., Qiu D., Blair D. Genetic Assignment Methods for Gaining Insight into the Management of Infectious Disease by Understanding Pathogen, Vector, and Host Movement. PLoS Pathogens 7(4): e1002013, 2013.

Reviews how genetic data can be used in epidemiology to infer the origin of pathogens.

Yang, W. Y., Novembre, J., Eskin, E., & Halperin, E. (2012). A model-based approach for analysis of spatial structure in genetic data. Nature genetics, 44(6), 725-731.

Shows how a simple logistic curve can discriminate between signal and noise in the variation of allele frequencies and enable fast inference of geographical origin without reference samples of known geographic origin.

INVESTIGATING ECOLOGICAL FACTORS CAUSING STRUCTURE

An important question in evolutionary biology consists in understanding the causes of population structure, a question related to the central goal of *phylogeography* but referred to as *landscape genetics* when considered at a smaller space and time scale. The topic has been reviewed by Manel et al. 2003, Storfer et al. 2006, Diniz-Filho et al. 2009, Storfer et al. 2010 and Manel and Holderegger. Notable modelling innovation to assess the role of geography have been made by McRae 2006, McRae and Beier 2007

and Bradburd et al. 2013. The impact of climate change is investigated by Moritz and Agudo 2013 and Devitt et al. 2013. Examples of studies on the effect of landscape on genetic structure can be found in Levy et al. 2013 and Trumbo et al. 2013.

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.

The first paper to review systematically the various questions and methods in the field.

McRae B 2006 Isolation by resistance. *Evolution*, 60, 1551-1561.

McRae B, Beier P 2007 Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences*, 104, 19885-19890.

An original model-based approach to go beyond isolation by distance models and assess how landscape affects gene flow.

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Waits, L. P. 2006. Putting the 'landscape' in landscape genetics. *Heredity*, 98(3), 128-142.

Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., & Waits, L. P. 2010. Landscape genetics: where are we now?. *Molecular Ecology*, 19(17), 3496-3514.

Review the exponentially growing method and applied literature in the field of landscape genetics.

Moritz, C., and Agudo, R. 2013. The future of species under climate change: resilience or decline. *Science* 341(6145):504-8.

Devitt, T.J., Devitt, S.E., Hollingsworth, B.D., McGuire, J.A., and Moritz, C. 2013. Montane refugia predict population genetic structure in the Large-blotched *Ensatina* salamander. *Mol Ecol* 22(6):1650-65.

Show case examples of population structure changes and discusses possible management guidelines in the face of current global warming.

Levy E, Tomkins JL, Lebas NR, Kennington WJ. 2013. Contrasting effects of landscape features on genetic structure in different geographic regions in the ornate dragon lizard, *Ctenophorus ornatus*. *Mol Ecol* 22(15):3904-15.

Demonstrates that land clearing following intense agriculture has affected the population structure of the ornate dragon lizard, increasing fragmentation, population differentiation and limiting local genetic diversity.

Trumbo DR, Spear SF, Baumsteiger J, Storfer A. 2013 Rangewide landscape genetics of an endemic Pacific northwestern salamander. *Mol Ecol*. 22(5):1250-66.

Investigates which landscape factors amongst climate, land cover and degree of anthropogenic disturbance, affect gene flow in three regions spanning the range of the Cope's giant salamander, with implication for conservation biology.

Diniz-Filho JA, Nabout JC, de Campos Telles MP, Soares TN, Rangel TF. 2009. A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genet Mol Biol* 32(2):203-11.

Reviews complex models in spatial analyses of genetic variation that account for autocorrelation issues.

Manel S, Holderegger R. 2013. Ten years of landscape genetics. *Trends Ecol Evol* 28(10):614-621

Reviews recent conceptual and methodological advances in landscape genomics, with emphasis on future challenges for understanding temporal and geographic patterns of gene-flow and population structure.

Bradburd, G.S, P.L. Ralph, G.M. Coop, 2013, Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, DOI: 10.1111/evo.12193.

Wang, I.J., R.E. Glor and J. B. Losos, 2013, Quantifying the roles of ecology and geography in spatial

genetic divergence, *Ecology Letters*, 16: 175-182

Two statistical models for quantifying the effect of distance relatively to that of other barriers on differentiation.

ADAPTATION AND NON-NEUTRAL CAUSES UNDERLYING STRUCTURE

Within a given species, populations facing different conditions can become adapted to local environments, creating a geographical structure at the genomic loci underlying adaptive phenotypic traits. This topic goes far beyond the present bibliography but interested readers can refer to Nielsen et al. 2005

Nielsen, R., 2005 Molecular signatures of natural selection. *Annu. Rev. Genet.* 39: 197-218.

Non-mathematical description of the issues involved in detecting selection from DNA sequences and genome-wide genetic datasets.

Pritchard J.K. 2010. How are we evolving? *Scientific American*. October 2010.

SELECTED APPLICATIONS

Virtually any single study involving population genetics data starts with an attempt to detect structure. Here are only mentioned studies that received a special attention due to the striking results reported or because of their methodological relevance. This includes studies in human genetics or in model species such as *A. thaliana* and also studies in large mammals or marine species for which factors that affect gene flow are poorly understood.

Humans

N.A. Rosenberg, J.K. Pritchard, J.L. Weber, H.M. Cann, K.K. Kidd, L.A. Zhivotovsky and M.W. Feldman, The genetic structure of human populations. 2002. *Science*, 298:

2381-2385.

The first application of the Structure program. Provides key insights about the organization of neutral genetic variation in humans at the synoptic scale. Contributed widely to demonstrate the utility of the program. Considered also as a milestone paper in medical genetics for the design of association studies.

Rosenberg N, Saurabh S, Ramachandran S, Zhao C, Pritchard J, Feldman M 2005 Clines, clusters, and the effect of study design on the influence of human population structure. Public Library of Science, Genetics, 1, 660-671.

A follow-up to to the previous reference that investigates the role of the spatial sampling design in the ascertainment of population structure and discusses the interpretation of differences of allele frequencies among populations.

Novembre J, Johnson T, Bryc K et al. 2008 Genes mirror geography within Europe. Nature, 456, 98-101.

Lao O, Lu T, Nothnagel M et al. 2008 Correlation between genetic and geographic structure in Europe. Current Biology, 18, 1241-1248

Two papers that received a considerable attention for the striking fine scale pattern of genetic variation observed on a large SNP dataset in Europe. The latter shows also how principal component analysis can be used to extract information about geographic origin of samples.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Pääbo S. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468(7327):1053-60.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. 2010. A draft sequence of the Neandertal genome. Science 328(5979):710-22.

Two seminal papers presenting draft genomes of archaic hominins and providing the first direct evidence for admixture with ancient modern humans.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan, Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, Zhou G, Tang M, Qin J, Wang T, Feng S, Li G, Huasang, Luosang J, Wang W, Chen F, Wang Y, Zheng X, Li Z, Bianba Z,

Yang G, Wang X, Tang S, Gao G, Chen Y, Luo Z, Gusang L, Cao Z, Zhang Q, Ouyang W, Ren X, Liang H, Zheng H, Huang Y, Li J, Bolund L, Kristiansen K, Li Y, Zhang Y, Zhang X, Li R, Li S, Yang H, Nielsen R, Wang J, Wang J. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329(5987) (2010):75-8.

Huerta-Sánchez E, Degiorgio M, Pagani L, Tarekegn A, Ekong R, Antao T, Cardona A, Montgomery HE, Cavalleri GL, Robbins PA, Weale ME, Bradman N, Bekele E, Kivisild T, Tyler-Smith C, Nielsen R. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol Biol Evol* 30(8) (2013):1877-88.

Detect signatures of adaptation to life at high-altitude in two independent human populations.

Apes

Becquet C, Patterson N, Stone AC, Przeworski M, Reich D 2007 Genetic Structure of Chimpanzee Populations. *PLoS Genet* 3(4): e66.

An investigation of the structure of our closest living relative with a study design similar to that of recent studies that have elucidated human structure.

Ghobrial, Lora, et al. 2010. Tracing the origins of rescued chimpanzees reveals widespread chimpanzee hunting in Cameroon. *BMC ecology* 10.1: 2.

Illustrates the usefulness of continuous assignment methods on genetically structured populations to detect illegal hunting.

Gonder, Mary Katherine, et al. Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations, 2011. *Proceedings of the National Academy of Sciences*. 108.12: 4766-4771.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, Cagan A, Theunert C, Casals F, Laayouni H, Munch K, Hobolth A, Halager AE, Malig M, Hernandez-Rodriguez J, Hernando-Herraez I, Prüfer K, Pybus M, Johnstone L, Lachmann M, Alkan C, Twigg D, Petit N, Baker C, Hormozdiari F, Fernandez-Callejo M, Dabad M, Wilson ML, Stevison L, Camprubí C, Carvalho T, Ruiz-Herrera A, Vives L, Mele M, Abello T, Kondova I, Bontrop RE, Pusey A, Lankester F, Kiyang JA, Bergl RA, Lonsdorf E, Myers S, Ventura M, Gagneux P, Comas D, Siegmund H, Blanc J, Agueda-Calpena L, Gut M, Fulton L, Tishkoff SA, Mullikin JC, Wilson RK, Gut IG, Gonder MK, Ryder OA, Hahn BH, Navarro A, Akey JM, Bertranpetit J, Reich D, Mailund T, Schierup MH, Hvilsom C, Andrés AM, Wall JD, Bustamante CD, Hammer MF, Eichler EE, Marques-Bonet T. Great ape genetic diversity and population history. *Nature* 499(7459):471-5.

Further results on the structure of the chimpanzee population in West Africa.

vonHoldt, B.M., Pollinger, J.P., Earl, D.A., Knowles, J.C., Boyko, A.R., Parker, H., Geffen, E., Pilot, M., Jędrzejewski, W., Jędrzejewska, B., Sidorovich, V., Greco, C., Randi, E., Musiani, M., Kays, R., Bustamante, C.D., Ostrander, E.A., Novembre, J., Wayne, R.K. 2011 A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res*. 21(8):1294-305.

Unravels the evolutionary history underlying population differentiation among wolves and coyotes.

Carnivore mammals

Cegelski, C.C. et al. 2003 Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Mol. Ecol.* 12, 2907–2918.

One of the first study to use a model-based clustering method to detect structure and barriers to gene flow in mammals.

Stenseth N, Shabbar A, Chan K et al. 2004 Snow conditions may create an invisible barrier for lynx. *Proceedings of the National Academy of Sciences*, 101, 10632–10634.

A study of the effect of climate and environmental conditions on genetic structure.

Pilot M, Jedrzejewski W, Branicki W et al. 2006 Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology*, 14, 4533–4553.

Those two studies show the effect of landscape on genetic structure and gene flow in two large mammals.

Tammeleht, E., Remm, J., Korsten, M., Davison, J., Tumanov, I., Saveljev, A., Mannil, P., Kojola, I, Saarma, U. 2010. Genetic structure in large, continuous mammal populations: the example of brown bears in northwestern Eurasia. *Mol Ecol.* 19(24):5359-70.

Describes how a range of landscape barriers has shaped the population structure of brown bears across North-Western Eurasia.

Sacks B, Bannasch DL, Chomel BB, Ernst H 2008 Coyotes demonstrate how habitat specialization by individuals of a generalist species can diversify populations in a heterogeneous ecoregion. *Molecular Biology and Evolution*, 25, 1354–1395.

One of the first studies to investigate the effect of habitat specialization at a local scale with model-based clustering methods.

Herbivore mammals

Coulon A, Guillot G, Cosson J et al. 2006 Genetics structure is influenced by landscape features. Empirical evidence from a roe deer population. *Molecular Ecology*, 15, 1669–1679.

Frantz AC, Tiget Pourtois J, Heuertz M et al. 2006 Genetic structure and assignment tests demonstrate illegal translocation of red deer (*Cervus elaphus*) into a continuous population. *Molecular Ecology*, 15, 3191–3203

Shows how model-based approach can be an aid for monitoring individuals with a view to populations management.

Pope L, Pope BR, Wilson G et al. 2007 Genetic evidence that culling increases badger

movement: implications for the spread of bovine tuberculosis. *Molecular Ecology*, 16, 4919–4929.

Shows how the analysis of population genetics structure can give insight into the spread of pathogens.

Other terrestrial animals

Gauffre B, Estoup A, Bretagnolle V, Cosson J 2008 Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology*, 17, 4616–4629.

Investigates the (absence of) effect of a potential barrier to gene flow and all suggests a general methodology to validate the output of clustering programs.

Frantz AC, Cellina S, Krier A, Schley L, Burke T 2009 Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, 46, 493–505.

A timely discussion of the interpretation of the outputs of clustering programs in light of isolation by distance.

Amphibians

Rowe G, Beebee T 2007 Defining population boundaries: use of three Bayesian approaches with the microsatellite data from British natterjack toads (*Buffo calamita*). *Molecular Ecology*, 16, 795–796.

Illustrates a common difficulty associated MCMC-based Bayesian clustering model, namely the frequent incongruence between programs outputs.

Birds

Rosenberg NA et al. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics*. 2001 159:699-713.

The first paper to evaluate the accuracy of the Structure program on real data in a controlled set-up (20 chicken breeds).

Coulon A, Fitzpatrick J, Bowman R et al. 2008 Congruent population structure inferred from dispersal behavior and intensive genetic surveys of the threatened Florida Scrub-Jay *Aphelocoma coerulescens*. *Molecular Ecology*, 17, 1685–1701.

Shows how the analysis of 20 microsatellite markers for 1000 individuals can provide insight into the effect of habitat fragmentation of an endangered bird species.

Rundel, C. W., Wunder, M. B., Alvarado, A. H., Ruegg, K. C., Harrigan, R., Schuh, A., Kelly, J. F., Siegel, R. B., DeSante, D. F., Smith, T. B. and Novembre, J. 2013, Novel statistical methods for integrating genetic and stable isotope data to infer individual-level migratory connectivity. *Molecular Ecology*, 22:4163–4176.

Illustrates the power of continuous assignment methods for the study of bird migration and also shows how they can be improved by combining genetic data to isotopic data.

Aquatic species

Palumbi S.R., 2003. Population Genetics, Demographic Connectivity, and the Design Of Marine Reserves, *Ecological Applications*, 13(1).

Discusses the use of genetic markers for the study of marine organisms in a conservation or management setting.

Galindo H, Olson D, Palumbi S 2006 A coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology*, 16, 1622–1626.

A key paper in 'seascape' genetics literature as it provides a method to predict gene flow in a marine environment.

Fontaine M, Baird S, Piry S et al. 2007 Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in old world waters. *BMC Biology*, 5, 30, doi:10.1186/1741-7007-5-30.

One of the first studies to undertake the analysis of a large dataset of a marine mammal. Provides also key insight into the interpretation of inferred clusters in presence of clinal variation.

Galarza J, Carreras-Carbonell J, Macpherson E et al. 2009 The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences*, 106, 1473–1478.

Explores the effect of variation in water density on seven fish species.

De Luna, C.J., Goodman, S.J., Thatcher, O., Jepson, P.D., Andersen, L., Tolley, K., and Hoelzel, A.R. 2012. Phenotypic and genetic divergence among harbour porpoise populations associated with habitat regions in the North Sea and adjacent seas. *J Evol Biol* 25(4):674-81.

Investigates population differentiation of harbour porpoise in relation to morphology, habitat and resources.

Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, Cresko WA. 2013. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Mol Ecol*. 2013 22(11):2864-83.

Uses a RAD-seq approach to investigate the population structure of the threespine stickleback fish, a species that has repeatedly undergone parallel phenotypic and genetic differentiation following the colonization of freshwater habitats from oceanic sources.

Insects

Fonseca DM, Keyghobadi N, Malcolm CA, Mehmet C, Schaffner F, Mogi M, Fleischer RC, Wilkerson RC. 2004. Emerging vectors in the *Culex pipiens* complex. *Science*. 303:1535-8.

Shows how the study of population structure can highlight the mode of propagation of a pathogen.

Emerson, KJ, Merz, CR, Catchen, JM, Hohenlohe, PA, Cresko, WA, Bradshaw, WE, Holzapfel, CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. Proc Natl Acad Sci U S A 107(37):16196-200.

Utilization of RAD-sequencing for unraveling the population structure and phylogeographic history of the pitcher plant mosquito.

Arias CF, Rosales C, Salazar C, Castaño J, Bermingham E, Linares M, McMillan WO. 2012. Sharp genetic discontinuity across a unimodal *Heliconius* hybrid zone. Mol Ecol. 21(23):5778-94.

Investigates one Columbian hybrid zone between two diverging *Heliconius* butterfly subspecies exhibiting different wing colour patterns.

Plants

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet. 28:286-9.

One of the first examples in plant genetics of how a clustering program can be used in an association study to control for the confounding effect of population structure.

Fenster C, Vekemans X, Hardy O 2003 Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (leguminosae). Evolution, 57, 995-1007.

A careful study of the important question of pollen dispersal investigated with molecular markers.

M. Delêtre, McKey D.B., Hodkinson T.R. 2011 . Marriage exchanges, seed exchanges, and the dynamics of manioc diversity. PNAS 108 (45) 18249-18254

Analyzes how kinship systems among farmers in Gabon impinge on the dynamics of crop genetic diversity by investigating the relationships between marriage exchanges and seed exchange networks.

Platt, A., M. Horton, Y. S. Huang, Y. Li, A. E. Anastasio, N.W. Mulyati, J. Aagren, O. Bossdorf, D. Byers, P. D. Kathleen, M. Dunning, E. B. Holub, A. Hudson, V. L. Corre, O. Loudet, F. Roux, N. Warthmann, D. Weigel, L. Rivero, R. Scholl, M. Nordborg, J. Bergelson, and J. O. Borevitz, 2010, The scale of population structure in *Arabidopsis thaliana*. PLoS Genetics 6: e1000843.

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, et al. 2012 Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. Nat Genet 44: 212-216.

Pico, F.X., Mendez-Vigo, B., Martinez-Zapater, J.M., and Alonso-Blanco, C. Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. Genetics 180 2008: 1009-2021.

Detailed studies of the population structure of *A. thaliana* at global and local scales.

Holderegger R, Buehler D, Gugerli F, Manel S. Landscape genetics of plants. *Trends Plant Sci.* 15(12) 2010:675-83.

Reviews plant studies in landscape genetics, with emphasis on selection-driven geographical patterns of genetic variation.

De Carvalho, D., P. K. Ingvarsson, J. Joseph, L. Suter, C. Sedivy et al., 2010 Admixture facilitates adaptation from standing variation in the European aspen (*P. tremula* L.), a widespread forest tree. *Mol. Ecol.* 19: 1638-1650.

Identifies Scandinavia as a postglacial contact zone between divergent lineages of *Populus tremula*, the widespread forest common aspen tree, and reveals signatures of adaptation within admixed populations.

Micro-organisms

D. Falush, T. Wirth, B. Linz, J.K. Pritchard, M. Stephens 2003 Traces of human migrations in *Helicobacter pylori* populations. and 13 others, *Science*, 299: 1582-1585

The first application of the Structure program to bacteria. Relates inferred genetic structure in bacteria to routes of human migration.

Rieux A., T. Lenormand, J. Carlier, L. de Lapeyre de Bellair, V. Ravigné 2013. Using neutral cline decay to estimate contemporary dispersal: a generic tool and its application to a major crop pathogen, *Ecology Letters*, 16(6):721-30..

Suggests a method to gain insights about the spatial scales involved in the dispersal process.

STATISTICAL AND COMPUTATIONNAL TOOLS

Statistical background

The modelling of population structure is a technical subject. In addition to probability theory (Durrett 2008), it has relied extensively on several recent statistical techniques such as Bayesian modelling (Beaumont and Rannala 2004), Monte Carlo simulation (Gilks et al. 1996) and approximate Bayesian computations (Beaumont 2010, Csillery et al. 2010).

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter, eds. 1996 *Markov chain Monte Carlo in practice*. Vol. 2. CRC press.

A reference book on the most widely used computational technique in the study of population structure. Does not contain specific material in genetics or evolutionary biology but addresses most of the questions encountered by users of MCMC-based clustering programs.

Beaumont, M. A., & Rannala, B. 2004. The Bayesian revolution in genetics. *Nature*

Reviews Genetics, 5(4), 251-261.

Introduces all the key concepts underlying the increasingly important field of Bayesian modelling, from a geneticist perspective.

R. Durrett. 2008. Probability models for DNA sequence evolution. Series in Probability and its Applications (New York). Springer, New York, second edition.

The mathematical background to many models and computational statistical methods in population genetics. Includes a chapter on spatial models of structure known as stepping stone models.

Beaumont, M., 2010 Approximate Bayesian Computation in Evolution and Ecology. Annual Review of Ecology, Evolution, and Systematics 41: 379-406.

K Csilléry, M. Blum, O. Gaggiotti, O. Francois 2010 Approximate Bayesian Computation (ABC) in practice. Trends in Ecology and Evolution 25: 410-418

Bertorelle, G., Benazzo, A., and Mona, S. 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Mol Ecol 19(13):2609-25.

These three review papers introduce ABC as the most important alternative to MCMC computation with mathematical background and connection to the evolutionary biology literature.

Coalescent theory

Theoretical results on populations genetics models are sometimes easier to obtain if one thin backward in time, which is the key idea behind coalescent theory (Schierup and Wiuf 2004, Wakeley 2009).

Gene genealogies, variation and evolution. A primer in coalescent theory. 2004. Hein, J., Schierup, M.H., and Wiuf, C. Oxford University Press.

An excellent introduction to coalescent theory, showing clear mathematical and theoretical derivations as well as a diversity of examples.

Wakeley J. 2009. Coalescent theory: an introduction. Roberts and company.

One of the most authoritative references on the models and computational methods used to infer evolutionary scenarios.

Computer resources

Excoffier, L., and G. Heckel. 2006 Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics* 7.10: 745-758.

A well-informed catalogue of computer programs.