Statistical inference on evolutionary processes in Alpine ibex (Capra ibex): mutation, migration and selection

Simon Aeschbacher

PhD

The University of Edinburgh

2011

To my parents Susann and Hansjörg

To my sisters Corina and Sarah
and their husbands and families

To my brother Christof

Declaration

I confirm that I have composed and written this thesis myself, that a significant part of the work is my own, and that I have not submitted this thesis or parts of it for any other degree or professional qualification than the PhD degree of the University of Edinburgh. I have, to the best of my knowledge, declared and acknowledged contributions, ideas, resources and help provided to me by others.

Simon Aeschbacher, Klosterneuburg, Austria, 30 July 2011

. . . . . .

Contents

Abstract

The thesis begins with a general introduction to population genetics in chapter 1. I review the fundamental processes of evolution – mutation, recombination, selection, gene flow and genetic drift – and give an overview of Bayesian inference in statistical population genetics. Later, I introduce the studied species, Alpine ibex (*Capra ibex*), and its recent history. This history is intimately linked to the structured population in the Swiss Alps that provides the source of genetic data for this thesis.

A particular focus is devoted to approximate Bayesian computation (ABC) in chapter 2, a method of inference that has become important over the last 15 years and is convenient for complex problems of inference.

In chapter 3, the biological focus is on estimating the distribution of mutation rates across neutral genetic variation (microsatellites), and on inferring the proportion of male ibex that obtain access to matings each breeding season. The latter is an important determinant of genetic drift. Methodologically, I compare different methods for the choice of summary statistics in ABC. One of the approaches proposed by collaborators and me and based on boosting (a technique developed in machine learning) is found to perform best in this case. Applying that method to microsatellite data from Alpine ibex, I estimate the scaled ancestral mutation rate ($\theta_{\text{anc}} = 4N_e u$) to about 1.288, and find that most of the variation across loci of the ancestral mutation rate u is between $7.7 \cdot 10^{-4}$ and $3.5 \cdot 10^{-3}$. The proportion of males with access to matings per breeding season is estimated to about 21%.

Chapter 4 is devoted to the estimation of migration rates between a large number of pairs of populations. Again, I use ABC for inference. Estimating all rates jointly comes with substantial methodological problems. Therefore, I assess if, by dividing the whole problem into smaller ones and assuming that those are approximately independent, more accuracy may be achieved overall. The net accuracy of the second approach increases with the number of migration rates. Applying that approach to microsatellite data from Alpine ibex, and accounting for the possibility that a model without migration could also explain the data, I find no evidence for substantial gene flow via migration, except for one pair of demes in one direction.

While chapters 3 and 4 deal with neutral variation, in chapter 5 I investigate if an allele of the Major Histocompatibility Complex (MHC) has been under selection over the last ten generations. Short- and medium-term methods for detecting signals of selection are combined.

For the medium-term analysis, I adapt a matrix iteration approach that allows for joint estimation of the initial allele frequency, the dominance coefficient, and the strength of selection. The focal MHC allele is shared with domestic goat, and an interesting side issue is if this reflects an ancestral polymorphism or is due to recent introgression via hybridization. I find most evidence for asymmetric overdominance (selection coefficient s: 0.974; equilibrium frequency: 0.125) or directional selection against the 'goat' allele (s: 0.5) with partial recessivity. Both scenarios suggest a disadvantage of the 'goat' homozygote, but differ in the relative fitness of the heterozygotes.

Overall, two aspects play a dominating role in this thesis: the biological questions and the process of inference. They are linked, yet while the proximate motivation for the biological component is given by a specific system – the structured population of Alpine ibex in the Swiss Alps – the methods used and advanced here are fairly general and may well be applied in different contexts.

Acknowledgements

First of all, I would like to thank my supervisor Nick Barton for his help and support throughout. Nick, your intuition, your advice and your criticism were invaluable. Thanks for always commenting in detail and within a short time on drafts, results and ideas. Thanks for your patience, trust and the freedom you gave me. Thanks for offering me to follow you to the IST Austria. Although I had mixed feelings to start with, it turned out to be just the right decision. Thanks for covering the costs for workshops, conferences and trips between Edinburgh and Vienna – and many months of computation time. I liked our discussions about the PhD projects, but also about future plans.

I would also like to thank Josephine Pemberton, my second supervisor, as well as Andy Leigh Brown and Richard Ennos for serving as advisors and committee members.

Many thanks to my collaborators Andreas Futschik, Mark Beaumont, Iris Biebach, Lukas Keller and Christine Grossen. Thank you Andreas for the discussions on ABC, on the choice of summary statistics and on many more ideas still waiting for implementation. Thank you Mark for giving instructions that led to the design of the ABC studies and for comments that substantially improved them. Thanks for joining the meeting in Seraplana, which turned out to be crucial for my PhD. Thank you Iris, Lukas and Christine for all these years of collaboration in the Alpine ibex project. Thanks, Iris, for always sending updated genetic and demographic data, for going to the lab again when something was missing, and for performing some preliminary analyses. Thanks for hosting that meeting at your home in Udligenswil to co-ordinate the MHC project. Many thanks, Lukas, for the meetings in Zurich, for the retreat in Seraplana and for funding some of my trips to Zurich. Thanks for co-ordinating the ibex project. Thank you Christine for providing the genetic data on MHC diversity. Your discovery of that one allele has given the whole project such a boost. Thanks for giving feedback to drafts even at times when you were very busy with your own thesis.

Thanks a lot to the members of the Keller group in Zurich, especially to Thomas Bucher, Clauco Camenisch and Ursina Koller for their work in the lab, and Barbara Oberholzer for double-checking our data on the re-introduction of Alpine ibex.

I thank Markus Brülisauer, Erwin Eggenberger, Flurin Filli, Bernhard Nievergelt, Marc Rosset, Urs Zimmermann, Martin Zuber, the staff from Tierpark Langenberg, Tierpark Dählhölzli (Bern), Wildpark Peter and Paul (St. Gallen), Wildtier Schweiz and the Swiss Federal Office

for the Environment for providing information on population history and re-introduction of Alpine ibex.

Thanks to Walter Abderhalden for discussions about migration in Alpine ibex, Christian Willisch for sharing his results on matig behavior of male ibex, and Ferran Palero, Michael Blum and Kati Csilléry for advice on ABC. Thanks, Kati, for inviting me to that meeting in Chamonix. I thank Damien Zufferey for carrying out some parallel computations during the phase of code testing, and Bill Hill for advice on estimating effective population size from demographic data.

Many thanks to Reinhard Bürger, Christoph Lampert and Jitka Polechová for comments on earlier versions of manuscripts.

Thanks to the Barton research group, both in Edinburgh and at the IST, for always being helpful, for discussions and social contacts. Thanks, Jerome Kelleher, for advice on UNIX and coding – and for taking me to my first rugby game. Thanks, Konrad Lohse, for interesting discussions about gall wasps, oaks, beetles, ibex and the coalescent. Thanks, Konrad and Marie, for that chanterelle foray and for the lasagne thereafter. Thanks to my office mates at IST, Anne Kupczok, Pavel Payne, Jitka Polechová and Daniel Weissman for a good atmosphere and different kinds of humor. And for help in eating Swiss chocolate.

When I arrived at the IST in October 2008, I was the second scientist, and for a long time, there were only a hand full of us. It has therefore been essential for me to connect to people from various research institutes in Vienna. Thanks to all of them for having us take part in seminars, meetings and retreats, especially to the groups of Reinhard Bürger, Andreas Futschik, Joachim Hermisson and Christian Schlötterer. Thanks to Stephan Peischl and Ada Akerman for being great team mates during course work, and to Ludwig Geroldinger for discussions before exams. Many thanks to all the members of the Evolutionary Theory Club (etc) for discussions and teaching. Thanks to Reinhard for providing an office in town.

Many thanks to Andy Leigh Brown, Reinhard Bürger, Brian and Deborah Charlesworth, Joachim Hermisson, Peter Keightley, Andrew Rambaut and Claus Vogl for your teaching in courses, seminars and journal clubs. It has been a privilege to learn from you.

Thanks a lot to Julia Asimakis, Carole Ferrier, Elisabeth Hacker, Nicole Hotzy for being great secretaries and for support in everyday life as a PhD student. Thanks to Gerti Resch and Martina Doppler for help and advice at the beginning of my time at the IST. Many thanks to Orlando Richards and Franz Schäfer and their teams at the University of Edinburgh and the IST Austria for great IT service and support.

I am deeply thankful to my family and relatives, and my friends in Switzerland, Scotland and Austria. You are too numerous to be listed one by one. Thank you very much, mum and dad, for mental and financial support, for always being there when I needed you. I was always welcome at home. Thanks to my parents for visiting me in Edinburgh. Thanks to friends and relatives who visited me in Vienna; thanks for every post card, letter, e-mail and text message. They meant a lot to me. Thanks, Dave and Mirjam Morf, for that particular post card from Vienna. Thank you Lucia and Conor Hull for letting me stay with you during my visits in Edinburgh.

Thanks to Miles Carter and Damien Zufferey for being nice and supportive flat mates. Thanks, Miles and Jo, for that great trip to Ben Vorlich and Stùc a' Chroin, my first Munros.

During my PhD, I have made heavy use of the computational resources provided by IST Austria and the Edinburgh Compute and Data Facility (ECDF; `http://www.ecdf.ed.ac.uk`). The ECDF is partially supported by the eDIKT initiative (`http://www.edikt.org.uk`).

Finally, I would like to acknowledge financial support by IST Austria, the Janggen-Pöhn Foundation, St. Gallen, the Roche Research Foundation, Basel, and by the University of Edinburgh in form of a Torrance Studentship.

Chapter 1

Introduction

1.1 What duck ponds tell us about evolution

When I look out of the window of our office at the Institute of Science and Technology (IST) Austria, I see a pond with ducks. They belong to the species called mallard (*Anas plathyrhynchos*). It is July now, and the males are in molt. At that time, they look pale, similar to the females, and you need to look twice to tell them apart. Yet, there is one duck that is very different. It is white, taller and has an orange beak. It must belong to a domestic breed. I have been observing these ducks for a while now, so I know that this white one is a male. And if you look closely, you will find that some other ducks have white feathers at positions where mallards normally do not. These are the offspring of the white duck and a female mallard from the previous year. That story reminds me of another duck pond with mallards, the one on the Irchel life science campus of the University of Zurich. There was one duck which was taller and of a different color than the others. Its body shape was reminiscent of that of an Indian Runner duck, but more bulky. My colleagues and me used to call it Max. Max must also have been a hybrid between a mallard and a domestic duck. Some years later, when I returned to visit my former group, I could not find Max anymore. I was not sure whether to feel sorry for Max or to be glad that – at least for a certain time – 'nature' had been restored at the Irchel pond.

What do these duck stories tell us? Most domestic duck breeds descend from *A. plathyrhynchos*. Their appearance has diverged from that of the mallard, both as a direct consequence as well as a side effect of artificial selection during the process of domestication. Artificial selection may be rather strong – just think of the difference between a wolf and a Pekinese dog. Yet, in the case of ducks, this divergence has not gone as far as to prevent successful intercrossing and cause what is known as reproductive isolation. The latter would define them as two separate species. The duck example implies that the processes leading to reproductive isolation may be gradual. Where do we draw the line? Moreover, if hybridization between domestic and wild ducks occurs every now and then at duck ponds, why do we not see more hybrids out there in nature? Is there a limit to their spread? Do they have disadvantages in the wild? Another question comes up if I think of the white feathers of the hybrid ducks at the IST. Why are these not found all over the place, but only at specific positions? Why do the hybrids look similar? Is there a mechanism controlling how characteristics inherited from parents are distributed, arranged and expressed in the offspring? How are these characteristics transferred

at all from generation to generation? And: is there maybe more variation than we can see by eye? What is the importance of such hidden variation? These are questions that lead to the heart of evolutionary genetics. They are not exclusive to ducks or any organism, but concern life in general. In this introduction, I would like to mention the most important evolutionary processes and questions in population genetics, the subfield of evolutionary genetics to which this thesis may be assigned. I will give an overview on some of this 'hidden' variation mentioned above; on how it is stored and organized, on the way it is transmitted across generations, and on how it can be detected. I will then introduce Alpine ibex, the species that provides the biological motivation for this thesis and from which genetic data were used to understand the recent evolutionary past of a population in Switzerland. I hope to provide the broad context of this thesis and the questions that will be addressed in later chapters. Each chapter will have a more specific introduction of its own, where more details and references to relevant literature are given.

1.2 Evolution, inheritance and genetic variation

When Charles Darwin and Alfred R. Wallace formulated their theories on evolution in the middle of the 19$^{\text{th}}$ century (Darwin 1859; Provine 1971), they did not know about the details of the underlying mechanisms. Yet, their observations of the diversity of life, both contemporary and as reflected in historical records, the apparent changes over time, the geographic distribution, the common patterns and shapes led them to postulate the principle processes of evolution. They regarded evolution as a gradual process by which new variants develop from existing ones and, going back in time, by which all living organisms go back to a common origin. At about the same time, the Austrian/Czech scientist and monk Gregor Mendel conducted experiments on plants. He crossed different varieties of the Bean (*Phaseolus spp.*) and the Pea (*Pisum sativum*), and observed the color of flowers and the shape and color of the fruits in the offspring generations. By back-crossing and other variations of the breeding scheme, he postulated fundamental rules according to which characteristics of the parent generation are *inherited* by the offspring. His experiments suggested that *discrete units* were transmitted in certain proportions. Mendel also found that novel types could appear, but these did not lose the capability to re-establish the original types in their offspring. While the hypotheses by Darwin and Wallace were broadly received and discussed, Mendel's discoveries were much underappreciated. The second half of the 19$^{\text{th}}$ century brought a heated debate between proponents of different theories – I am tempted to say speculations – about the kind and mechanism of evolution (see Provine 1971). The debate was essentially about whether evolution happened in discrete steps as the *Mendelians* proposed, or gradually as the *Biometricians* argued.

In 1900, Hugo de Vries, Carl Correns and Erich von Tschermak rediscovered Mendel's laws of heredity. This, together with a crucial insight by the British mathematician and statistician George U. Yule that Mendelism was not necessarily associated with discontinuous evolution, and that Mendelian factors might themselves be variable in small but discontinuous steps (Yule 1902), anticipated what is nowadays called the *evolutionary synthesis* (Provine 1971). The evolutionary synthesis reconciled many of the opposing arguments that were building up at the turn of the century. It stated that natural selection and gradual evolution were not incompatible

with Mendelian inheritance in discrete (but potentially small) steps. The evolutionary synthesis was brought about in the 1930s and 40s by experimental evidence as well as mathematical theory. Important figures were, among others, Theodosius Dobzhansky, Ronald A. Fisher, John B. S. Haldane, Julian Huxley and Sewall Wright.

What Mendel had observed as discrete units and what Yule referred to as Mendelian factors are today called *genes*. The term goes back to a publication in 1909 by the Danish botanist Wilhelm Johannsen (Provine 1971). A gene is a unit of heredity. Going back to Mendel, the concept of a gene had been postulated about fifty years before the physical carriers of the genes in the nucleus of the cell, the *chromosomes*, were discovered in 1915 by Thomas H. Morgan (Provine 1971). It was not until 1952 that the *deoxyribonucleic acid* (DNA) was identified as the chemical substance that stores the genetic information (Hershey and Chase 1952). DNA as a molecule had been discovered much earlier, in 1869, by Friedrich Miescher, who called it "nuclein" (Dahm 2008). DNA has the structure of a double helix, as discovered by Watson and Crick (1953) and Rosalind Franklin. The DNA consists of units called nucleotides, each of which is made up of a sugar, a phosphate and one out of four base molecules (called *bases*) – adenine (A), cytosine (C), guanine (G), or thymine (T). The genetic information is stored in this four-letter alphabet on the DNA. The two strands of a DNA helix run in opposite directions, and the bases of the two strands are paired according to the rule that A binds with T, and G with C. They complement each other, and the information is therefore stored redundantly. This redundancy is crucial, because before cell division, the information has to be copied (*replicated*) and redistributed to the two daughter cells so that each of them has the same information. During replication, the two strands of the DNA helix are forced apart, an enzyme complex (including the DNA *polymerase*) walks along the fork between the strands and synthesizes a complementary strand to both of them. As a result, the double helix is copied and the two helices can be passed on to each of the daughter cells. Most mammals are *diploid*, meaning that each of their cells has *two* sets of chromosomes. In sexually reproducing organisms, one set comes from one parent, the other from the second. The chromosomes that correspond to each other are called *homologues*. For reproduction, diploid organisms produce a particular type of cells, the *gametes*. These have only one set of chromosomes and are therefore *haploid*. They are built by a special type of cell division, during which the homologue chromosomes are separated, such that each gamete contains only half the genetic information of the cell it was built from. When two gametes meet and fuse during fertilization, the resulting *zygote* has again two full sets of chromosomes; now, one set originates from one parent, and the other from the second parent. This is the mechanism underlying Mendelian inheritance. The fact that, during sexual reproduction, the combination of chromosomes is to some extent re-shuffled, is called *recombination*. These two mechanisms determine how genes are passed on from one generation to the next, and how statistical associations among genes are broken up.

DNA may be separated into *coding* and *non-coding* parts. The coding DNA is read by enzymes, transcribed to an intermediate molecule – the *ribonucleic acid* (RNA) – which is then translated by a macromolecule called *ribosome* into sequences of amino acids. This last step is accomplished according to the *genetic code* that maps to every possible triple of bases one or several amino acids. Amino acids are the building blocks of proteins. The proteins serve as enzymes in biochemical reactions or as structural units of the cells. They are therefore directly

linked to the function and development of different parts of an organism. This cascade of events from DNA to function reflects the way genes are *expressed*. It also links the *genotype* – the genetic constitution of an organism – to its *phenotype* – the set of characteristics and traits by which an organism interacts with its environment. The phenotype itself may be affected by the environment, not only by the genes. Understanding exactly how a genotype translates into a phenotype, accounting for the effects of the environment, is of great interest, but a difficult task. The non-coding DNA does not code for proteins, but may nevertheless have a function. For example, it may contain particular nucleotide sequences to which proteins bind. This way, non-coding DNA sequences can act as regulators of gene expression, enhancers, or promoters. Moreover, they may play a structural role in determining how far particular genes are from each other, which can again effect the expression of genes in the coding DNA. The ensemble of coding and non-coding DNA in an organism is called the *genome*.

Above, I have introduced the gene as the unit of heredity. To be more precise, a gene is a region on the DNA that is associated with some function (Pearson 2006). An obvious function is coding for a protein, but the function may also be to serve as a binding site where proteins bind and from which they regulate processes in the cell nucleus (see above). The order, arrangement and number of genes on the DNA varies greatly between species. A more general term than gene is *locus*. A locus refers to a particular position on the genome, be it part of a coding or non-coding region. Locus is often used to denote a gene *or* some non-functional DNA that is of particular interest. So, locus is a more general term than gene.

DNA is subject to processes that alter its chemical composition and rearrange parts of it. This is called *mutation*. Mutations may be caused by errors during replication, radiation, mutagenic chemicals, viruses or other pieces of DNA that can transpose themselves from one DNA molecule to another one. Mutations can affect single base pairs (point mutations), but also result in insertion, deletion or inversion of whole sections of DNA. Both coding and non-coding DNA can be affected by mutation. If mutations occur in protein-coding parts of the DNA, they may or may not be reflected in the protein, depending on the genetic code. The code is redundant, associating several triplets of bases to a given amino acid. Therefore, it is resistant to some mutations. Mutations that are reflected in the protein are called *non-synonymous*. Those that are not are called *synonymous*. Similarly, mutations in non-coding regions of the DNA may or may not have an effect, depending on whether they occur at functional or non-functional positions. Mutation is the process by which new genetic variation is caused. In contrast, recombination (see above) is the process by which existing variation is re-arranged during reproduction.

Due to mutations, organisms of the same species may differ at particular positions in their genome. Loci may occur in different variations, some of which are reflected in variation that is visible to the environment. Different variants of a locus are called *alleles*. In diploid organisms, the genotype at a particular locus for a given individual is made up of the two alleles it received from its parents. If the two alleles are identical, then the individual is *homozygous* for that locus, otherwise it is *heterozygous*. The two alleles do not necessarily contribute equally to the corresponding phenotype. The asymmetry in this contribution is called *dominance*. There is no dominance if the two alleles contribute equally. An allele that overrides the other to some degree is called *dominant*; the other allele is then called *recessive*.

The first molecular data became available in the form of *allozymes* (Hubby and Lewontin 1966). These are variants of a given protein that differ in their electric charge, and that are coded by different alleles of the gene that codes for that protein. When put onto a gel across which an electrical potential is established, allozymes move at a speed that depends on their electrical charge and on their size. Relative differences can then be detected. This process is called allozyme electrophoresis. It provided insight into levels of molecular diversity and allowed first comparisons between theoretical predictions and data. About ten years later, it became possible to sequence, *i.e.* read, RNA and DNA directly. In 1983, the *polymerase chain reaction* (PCR), was invented. It allowed for amplification of specific sections of DNA and made it possible to study variation at the DNA level. Before molecular and genetic data of this kind were available, only variation that was visible to the human eye could be detected. Genetic data revealed much more, previously hidden variation. Loci that are used to detect this variation are called genetic *markers*. One type of markers are the so called *microsatellites*, also known as short tandem repeats (STRs). These consist of a specific motif of one to six base pairs of length, which is repeated a certain number of times. The number of times the motif is repeated defines the different alleles. Microsatellites have a relatively high mutation rate compared to other types of loci; in mammals it is estimated to 10^{-4} to 10^{-2} per locus and generation (Di Rienzo et al. 1998; Estoup and Angers 1998). The predominant cause for mutations in microsatellites is slippage of the protein complex responsible for replication, resulting in additional motifs being added, or in motifs being lost. The resulting mutation process can be modelled by the stepwise model of mutation that also applies to allozymes (Kimura and Ohta 1978). Microsatellites mainly occur in non-coding DNA and have been used extensively as markers in studies on genetic variation in mammals. In chapter 3, 4 and 5, microsatellite data are used to indirectly estimate different evolutionary parameters of interest (see below). There are other types of markers, with corresponding models of mutation, and other methods for obtaining genetic data. Covering these would be beyond the scope of this introduction, however. Next, I will focus on the evolutionary processes and the questions that are of interest in population genetics.

1.3 Evolutionary processes and questions in population genetics

Organisms can be categorized into species. One definition of a species is that it comprises all organisms that are capable of interbreeding and producing fertile offspring. Within a species, however, organisms may be organized in further units. An important unit is that of a *population*. A population is made up of organisms that belong to the same species *and* live in the same area so that every individual can in principle mate with any other to produce offspring. More precisely, individuals within a population are considered *more likely* to mate with each other than are two individuals from two different populations. Both the species and population concept are vague (*e.g.* Barton et al. 2007). Going into details here would lead us too far off track, however. *Population genetics* is the study of the change in time and space of allele frequencies in populations. There are five fundamental evolutionary processes that cause such change. We have already encountered two – mutation and recombination. The others are *selection*, *gene flow* and *genetic drift*.

To understand selection, it is helpful to introduce the concept of *fitness*. Fitness can be defined with respect to a phenotype or a genotype. For simplicity, we assume here that the genotype translates directly into a phenotype. As mentioned above, this is in general not the case, but it simplifies the explanation. Fitness then describes the ability of a genotype to survive and produce viable offspring. If different genotypes have different fitnesses, the genotype frequencies will change from generation to generation. Because the genotypes are made up of alleles, allele frequencies are in general also affected by selection. The fitness of an allele (the so called *marginal* fitness) is defined as the mean fitness of all the genotypes that contain this allele, weighted by the probability that the allele occurs in the respective genotype. The change in allele frequencies due to fitness differences is called selection. Selection may be due to fitness differences felt in the natural environment, or due to fitness differences artificially imposed by humans (*e.g.* in a laboratory or during domestication), and is then called *natural selection* or *artificial selection*, respectively. A locus or gene that is not under selection is called *neutral*.

Gene flow describes the change in allele frequencies due to the displacement of genes in space. In most organisms, gene flow occurs via the physical movement of individuals, seeds or gametes between the place of birth (or the place where gametes were built) and the place of reproduction (or fertilization). This physical movement is called migration or dispersal. Strictly speaking, gene flow can also refer to the movement of genes between different genomic backgrounds within an organism, or the exchange of genetic material from cell to cell in bacteria, for example. Here, we focus on gene flow via migration or dispersal. The concept of gene flow implies a notion of space. Indeed, the natural environment of populations enforces some spatial organisation. For example, islands on the ocean constrain the spatial distribution of land animals and plants. Mountains may limit the spread of organisms that cannot pass them. Populations are therefore often subdivided into smaller units – subpopulations or *demes*. The rate at which *demes* exchange migrants is called *migration rate*. It is a demographic parameter with a direct impact on the strength of gene flow.

Genetic drift describes the random changes in allele frequencies from generation to generation. These changes are a consequence of the *finite* number of individuals in real populations. At reproduction, the genetic composition of the offspring generation is *sampled* from the gene pool (the ensemble of gametes) produced by the parental generation. Because the offspring generation is again finite, some alleles may be lost, others may increase in frequency just by chance. Genetic drift does not change the *expected* allele frequency in the next generation, but it increases the variance of the allele frequency. The effect of this random sampling decreases with increasing size of the population. For large enough populations, genetic drift has a negligible effect.

The genetic composition of natural populations is affected by a combination of these evolutionary forces. Population genetic theory studies the evolutionary processes individually, as well as jointly. Examples of questions that are addressed are the following:

1. How and when is genetic diversity maintained?
2. What level of gene flow is necessary to prevent two demes from diverging from each other?
3. How quickly is genetic diversity lost as a consequence of genetic drift?

4. What strength of selection is necessary to overcome the random effects of drift?
5. What is the effect of demography on genetic diversity?

Population genetic theory uses mathematics and statistics to answer these questions. Mathematical models play an important role in this process. They are used to formalize the evolutionary forces, but also the demography of populations. Models provide some abstraction and simplification of the real problem, while still capturing the features of interest. Building a model implies making assumptions. To study the model analytically, it is often necessary to make additional assumptions. For example, a rather simplistic model of a population could include the following assumptions:

- Individuals are diploid
- Inheritance occurs according to Mendel's laws
- There is sexual reproduction with random mating
- There is no mutation, no selection and no gene flow
- The population is infinitely large
- There is one locus with two alleles, A_1 and A_2

Denoting the frequency of the A_1 allele by p and that of the A_2 allele by $q = 1 - p$, we could then ask about the change of p from one generation to the next. Let us denote the frequency of the three possible genotypes, A_1A_1, A_1A_2 and A_2A_2 by P_{11}, P_{12} and P_{22}, respectively. Further, assume that the adults in the current generation contribute equally to an infinitely large pool of gametes, and that these gametes then unite randomly to form the zygotes of the next generation. Denoting the genotype frequencies in the next generation with a prime, we have

$$\begin{aligned} P'_{11} &= p^2 \\ P'_{12} &= 2pq \\ P'_{22} &= q^2. \end{aligned} \tag{1.1}$$

What is the allele frequency p' in the zygotes? Because A_1A_1 has two A_1 alleles, and A_1A_2 has one A_1 allele, and using (1.1) we obtain

$$p' = (2\,P'_{11} + P'_{12})/2 = (2p^2 + 2pq)/2 = p^2 + pq = p(p+q) = p. \tag{1.2}$$

We have just shown that the allele frequency does not change. Moreover, because the allele frequencies do not change, the genotype frequencies in further generations will also not be changed. Both will remain constant as long as the assumptions above hold. This is known as the Hardy-Weinberg principle (Halliburton 2004), and the proportions in (1.1) are the Hardy-Weinberg proportions. Any violation to the assumptions may cause this to break down. Obviously, this model is not very realistic. Moreover, if we go out, sample some genetic data and find that (1.1) holds, does this mean that all the assumptions apply to the population? No. We cannot exclude that a particular combination of evolutionary forces led to genotype and allele frequencies in accordance with the Hardy-Weinberg proportions. The only valid conclusion we could draw from this considerations is the following: If we do find deviations from Hardy-Weinberg

proportions in a sample, then at least one assumption must be violated. Overall, we cannot infer very much from this simple model. But the example illustrates the idea of using a model, of stating assumptions, and of using the model to obtain the answer to a question. This is the principle that goes throughout population genetic theory. Usually, the models are more complicated, sometimes the assumptions are more realistic.

The early mathematical treatment of population genetics was strongly influenced by the work of Fisher (1930), Haldane (1932) and Wright (1931). They laid the groundwork for the quantitative study of mutation, selection, gene flow and genetic drift. Wright and Fisher formulated a model of an idealized population, later called the *Wright-Fisher model*, to describe the effects of genetic drift under a set of assumptions (Fisher 1922a; Wright 1931). The Wright-Fisher model has since played a crucial role in theoretical studies. Wright (1931) introduced the *effective population size*, N_e, as the size of a Wright-Fisher population that would experience the same amount of genetic drift as the population under consideration. This concept became important, because it allowed to map a large number of more complicated models to the Wright-Fisher model, such that results obtained for the Wright-Fisher model could be generalized for these other models. In chapters 3, 4, and especially 5, I make use of this principle. Wright (1931, 1943, 1951) also studied the effects of inbreeding and population structure on genetic diversity. Fisher and Haldane focussed more on the theory of selection. Fisher had a strong influence on *quantitative genetics* and first applied the diffusion equations to approximate the distribution of allele frequencies among populations (Fisher 1922a). The *diffusion approximation* was later also applied by Wright (1937, 1945). Haldane analyzed selection in the context of various dominance schemes, modes of inheritance, mating patterns, mutation, multiple loci, non-overlapping generations or competition. His work showed that natural selection was a plausible mechanism for evolution (Haldane 1932), a question that had been strongly debated before (Provine 1971). Later important contributions to population genetic theory include the work by Motoo Kimura, Tomoko Ohta and Gustave Malécot. Kimura and Ohta extensively used the diffusion approximation, most importantly to study *fixation times* and *fixation probabilities* of mutations under a variety of conditions (Kimura and Ohta 1969; Ohta and Kimura 1972; Kimura and Ohta 1974). Fixation means that an allele reaches frequency $p = 1$, so that all other alleles at that locus are lost. Variation at that locus can only be re-established by mutation or gene flow into the population. Ohta and Kimura also postulated the (nearly) neutral theory of evolution as an attempt to reconcile theory with observed levels of genetic diversity (Kimura 1984). That theory was much debated; its opponents believed that natural selection played a much more important role in shaping genetic diversity than did Ohta and Kimura. Malécot, on the other hand, developed the concept of *identity by descent* (Malécot 1969), extending earlier, related work by Wright on inbreeding coefficients and genetic drift. Kimura and his collaborators, as well as Malécot, also had a strong focus on spatially structured populations (Kimura and Ohta 1978; Nagylaki 1989). Moreover, Malécot's work anticipated a shift in population genetic modelling from a forward to a retrospective view: In the early 80s, Kingman (1982) established a stochastic theory for the ancestral relationship of genes, the *coalescent* theory. The coalescent models genetic drift, as one follows the history of a sample of genes back into the past. This ancestry is reflected in a genealogy, a bifurcating tree. All lineages ultimately coalesce in the most recent common ancestor. The coalescent theory

provides information about the length of the tree and the distribution of coalescent times, and it allows for much more efficient simulation of populations compared to the forward perspective. Many classical results can be re-interpreted and rediscovered in the coalescent framework. The coalescent theory has been extended to incorporate spatial structure, population growth, alternative models of reproduction, and – to some extent – recombination and selection (Kaplan et al. 1988; Hudson and Kaplan 1988; Hein et al. 2005; Wakeley 2009).

1.4 Statistical population genetics

Within the field of population genetics, there is one branch which is concerned with the estimation of evolutionary or demographic parameters, given observed data. That branch may be called *statistical population genetics*. Some questions of interest in statistical population genetics are:

- What is the relative strength of evolutionary processes in shaping genetic diversity?
- What are the relative time scales over which the evolutionary forces act?
- What is the rate at which genes mutate?
- What is the migration rate between two or several demes?
- What is the extent of inbreeding in a population?
- What strength and mode of selection is compatible with an observed genetic composition?

Statistical population genetics uses general methods and principles of *inference* to answer such questions. Inference is the process of drawing conclusions about unobserved quantities of a system of interest, given some observed quantities. The unobserved quantities are often the parameters of a process in the system. The observed quantities may correspond to random variables or, more generally, to what we call the *data*. Inference is tightly linked to the concept of *probability*. Throughout this thesis, I am mainly taking a Bayesian viewpoint of inference. In *Bayesian statistics*, probabilities quantify a 'belief' in some fact, *e.g.* that a quantity of interest has a certain value, given some assumptions or previous knowledge (*e.g.* MacKay 2003). Another way of looking at it is to say that, in Bayesian statistics, probabilities are used to quantify uncertainty. In essence, Bayesian inference tries to estimate a probability distribution across all potential values of the parameter(s) of interest, given what is called *prior* knowledge or belief. A probability distribution is a function that assigns a probability to each value of its argument. Bayesian statistics goes back to Bayes' theorem, which expresses the probability of an event A given another event B in terms of the inverse conditional probability (Gelman et al. 2004):

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}, \tag{1.3}$$

where $P(A \mid B)$ means 'the probability of A conditioning on B' or '... given B'. In Bayesian statistics, as opposed to the frequentists' interpretation, a probability may not only be assigned to an observable quantity, but also to an unobservable parameter.

To make things concrete, suppose we have a 'system' – a population of ducks, say – and we want to know 'something' about that system we cannot observe directly. We might be interested

in the mutation rate at a certain locus. What are the steps to get to an answer? Mutations leave traces in the DNA, so we would first collect samples (blood, tissue, meat), extract DNA and amplify it at that locus, to then determine the sequences (or alleles) for every sampled individual. For this, we would need a lab. Suppose we do or we have a collaborator who does, and that we have obtained our raw data; from now on, we need paper, pencil and, perhaps, a computer. Second, we need a model that links the process of interest – mutation – with the observed data. That model must be probabilistic, meaning that it should provide a probability distribution of both the observed data and the parameter – in our case the mutation rate. The model also needs to make assumptions about demography and, potentially, other processes that affected the past of our duck population. In example 2 of chapter 2 we will encounter one model that might be appropriate in this case. Third, we need a methodology to condition on the observed data and compute the *posterior distribution*. Since we do Bayesian statistics, we also have to chose a *prior distribution*. That choice can be more or less informative. If we do have prior knowledge on the mutation rate, we should incorporate it; otherwise, it is common practice to choose a prior that covers the range of possible parameter values uniformly on an appropriate scale. The choice of the prior distribution can have a strong effect on the result, in particular if the data are not informative. With respect to (1.3), A corresponds to the mutation rate and B to the observed data. $P(A \mid B)$ is called the posterior distribution, $P(B \mid A)$ the *likelihood*, $P(A)$ the prior distribution and $P(B)$ is the total probability of the data, also called the *marginal likelihood* (see chapter 2). Having found the posterior distribution, we may compute *point* or *interval* estimates of the mutation rate. The appealing property of Bayesian inference is that its result, the posterior distribution, reveals intuitively the uncertainty attributed to our inference. As a fourth step, we want to evaluate our inference. Does the model fit the observed data? Is the result plausible? Evaluation is often an iterative process; we may have to go back, adjust our model, and re-calculate the posterior distribution (Gelman et al. 2004).

The above example illustrates a process that is common to many studies in statistical population genetics. It will appear in all the following chapters of this thesis. Steps two and three are probably the most demanding ones. In chapter 2, I will introduce an approximate method for step 3 – the computation of the posterior distribution – called *approximate Bayesian computation* (ABC). In chapters 3 and 4, ABC will be used to estimate evolutionary and demographic parameters in Alpine ibex (see below). In chapter 5, an exact method is used for inference about the mode and strength of selection on a particular gene in Alpine ibex.

1.5 Alpine ibex (Capra ibex) and its history

At the end of an undergraduate course at the University of Zurich, I had the chance to participate in a birdwatching trip across Scotland. On the coast near Aberdeen, Lukas Keller told me about his plans to engage in a research project on molecular ecology and population genetics of Alpine ibex. A few months later, I joined his group and started working in that project. That was the beginning of a process that led to this thesis – and the reason why birds will not play a role in it anymore from now on.

Alpine ibex (*Capra ibex*) is a wild goat species (genus *Capra*), belonging to the bovids (family *Bovidae*), and therefore to the even-toed ungulates (order *Artiodactyla*). Alpine ibex

are one of several ibex species occuring in Europe, Asia and northern Africa. The contemporary distribution of Alpine ibex is restricted to the European Alps, in the alpine zone at altitudes of 1,800 to 3,000 meters. Alpine ibex was almost extinct by the beginning of the 18^{th} century, most likely as a consequence of over-hunting since the 16^{th} century (Stuwe and Nievergelt 1991). It is speculated that climatic changes also played a role. Only one population of 100 to 300 individuals was left in the Gran Paradiso Mountains in the Italian Alps. After protection in 1858 by the Italian King, the Gran Paradiso population increased to approximately 3,000 individuals by the beginning of the 20^{th} century (Stuwe and Scribner 1989). Between 1906 and 1942, roughly 100 ibex captured in the Gran Paradiso population were brought to two zoos in Switzerland, where a breeding program was started (Stuwe and Scribner 1989). Since 1911, several former populations have been stocked with founders from these captive breeding program. Some of these re-established wild populations were later used as a reservoir for further translocations. Alpine ibex were also re-introduced to other countries along the European Alps. The efforts were successful; by 2005, the total ibex population in Switzerland was estimated as 14,000, and in Europe as 40,000 (Biebach and Keller 2009). The population in the Swiss Alps can be divided into more or less discrete *colonies*, called demes in the rest of this thesis. The re-introduction of ibex into the Swiss Alps has been documented in great detail by game keepers and hunters. For a large number of demes, census sizes have been recorded, and the number and sex of individuals transferred between demes have been listed. This information, although spread over different sources, could be gathered and the complete history reconstructed (Aeschbacher 2007; Biebach and Keller 2009).

Different ibex demes vary in the number of founder events and bottlenecks they experienced. They also differ in their dynamics and the number of generation since re-introduction. Moreover, some demes were affected by environmental effects such as diseases, avalanches and climate (Sæther et al. 2002; Grøtan et al. 2008, see also chapters 4 and 5 of this thesis). Since 1977, hunting has been imposed on the majority of demes in Switzerland to control population density. Annual culling rates range from 6 to 12% (Stuwe and Nievergelt 1991). Some demes declined in size at the end of the 1990s, and it was not clear for what reasons. After 2000, the Swiss Federal Office for the Environment (FOEN) initiated a research project to investigate potential causes, and to evaluate and improve the existing management strategy. This project had several modules, one on population demography and dynamics, one on diseases, another on (behavioral) ecology, and one on molecular ecology and population genetics. My thesis has its roots in the last module. Its central theme is to infer demographic and evolutionary parameters from genetic data, conditioning on the demographic information available. Along these lines, it became necessary to tailor, and further develop, existing methods of inference for that specific setting. That is the reason why, apart from the biological motivation from Alpine ibex, this thesis has a strong methodological focus. A more detailed outline is given in the following section.

1.6 Outline of thesis

In chapter 2, I give an introduction to approximate Bayesian computation (ABC), the method of inference used in two of the following chapters. Chapter 3 is concerned with the choice of

summary statistics in ABC. Since statistics are in most cases not sufficient (for a definition, see next chapter, section 2.2), that choice involves a trade-off between loss of information and reduction of dimensionality. The latter may increase the efficiency of ABC. Me and my collaborators propose a novel approach for choosing summary statistics basen on boosting, a technique developed in the machine learning literature. Different types of boosting are proposed and compared to partial least squares regression (PLS) as an alternative method. To mitigate the lack of sufficiency, we also propose an approach for choosing summary statistics locally, in the putative neighborhood of parameter values inferred from the observed data. We study a demographic model motivated by the re-introduction of Alpine ibex (*Capra ibex*) into the Swiss Alps. The parameters of interest are the mean and standard deviation across microsatellites of the scaled ancestral mutation rate ($\theta_{\text{anc}} = 4N_e u$), and the proportion of males obtaining access to matings per breeding season (ω). In a simulation study, we assess the accuracy and coverage properties of the various methods. We find that ABC with summary statistics chosen locally via boosting with the L_2-loss function performs best. Applying that method to the ibex data, we estimate $\hat{\theta}_{\text{anc}} \approx 1.288$, and find that most of the variation across loci of the ancestral mutation rate u is between $7.7 \cdot 10^{-4}$ and $3.5 \cdot 10^{-3}$. The proportion of males with access to matings per breeding season is estimated to $\hat{\omega} \approx 0.21$, which is in good agreement with recent independent estimates.

In chapter 4, my collaborators and I propose a two-step procedure for estimating multiple migration rates in the ABC framework, accounting for global nuisance parameters. We condition on a known, but complex demographic model of a spatially subdivided population, motivated by the re-introduction of Alpine ibex into Switzerland. In a first step, the global parameters ancestral mutation rate and male mating skew have been estimated for the whole population in chapter 3. In chapter 4, we estimate the migration rates independently for clusters of demes putatively connected by migration. For large clusters (many migration rates), ABC runs into the curse of dimensionality. We therefore assess by simulation if estimation per pair of demes is a valid alternative. We find that the trade-off between reduced dimensionality for the pairwise estimation on the one hand, and lower accuracy due to the assumption of pairwise independence on the other, depends on the number of migration rates to be inferred. The net accuracy of the pairwise approach increases with the number of migration rates. To distinguish between low and zero migration, we perform an ABC-type model comparison procedure between a model with migration and an alternative model without migration. We further confirm boosting as a valid method for choosing summary statistics in ABC. Applying the approach to microsatellite data from Alpine ibex, we find no evidence for substantial gene flow via migration, except for one pair of demes in one direction.

Chapter 5 is devoted to a gene of the Major Histocompatibility Complex (MHC), and to the question whether this gene has recently been under selection in Alpine ibex. MHC is likely to be under parasite-mediated balancing selection in many vertebrate taxa. However, empirical studies have not provided a univocal answer regarding the underlying mechanism (overdominance, spatio-temporally varying selection) and the strength of selection. My collaborators and I combine short- and medium-term evidence to infer the evolutionary fate of an MHC allele in a structured population of Alpine ibex in the Swiss Alps. The allele is shared with domestic goat. As a short-term signal of selection, we find a negative correlation between heterozygos-

ity and age at sampling, suggesting viability selection with underdominance or intermediate dominance. For the medium-term, we focus on the observed allele frequency distribution. Low variance across demes implies spatially homogeneous selection. To estimate the selection coefficient (s) we employ a drift-selection-migration model and develop a matrix iteration approach to compute likelihoods. We find most evidence for asymmetric overdominance (s: 0.974; equilibrium frequency: 0.125) or directional selection against the 'goat' allele (s: 0.5) with partial recessivity. Both scenarios suggest a disadvantage of the 'goat' homozygote, but differ in the relative fitness of the heterozygotes. We relate our results to MHC function and hypotheses on its evolution, and discuss the disparity between short- and medium-term evidence.

Chapters 3 and 4 are closely related, and it is best to read them one after another in that order. Chapter 5, on the other hand, may also be read separately.

1.7 Format, use of language and electronic resources

Chapters 3 to 5 have been written as journal papers. Subject to editorial changes, they will be submitted shortly after submission of this thesis. Since co-authors were involved as stated at the beginning of each chapter, I use 'we' throughout these chapters. Nevertheless, I have written the whole text as it appears in this thesis myself. Each of these chapters has its own introduction. In addition, appendices and supporting information are given at the end of each chapter. A website with electronic resources such as the simulation program `SPoCS` written by me and used in chapters 3 to 5, scripts for analysis and for parallelizing ABC on a cluster, and tables with additional information that could not be included in the main text can be accessed via `http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`.

Chapter 2

Approximate Bayesian Computation

2.1 Introduction

Approximate Bayesian computation (ABC) is a collective term for a family of inference methods in Bayesian statistics (Beaumont 2010). ABC uses Monte Carlo simulations and a rejection algorithm to condition on observed data. It does not depend on explicit calculation of the likelihood, and is therefore most often applied in contexts where computation of the likelihood is impossible or prohibitive. ABC was invented in a series of papers in evolutionary and population genetics in the late 1990s. It has since been further developed and applied in many studies, also in other fields than evolutionary genetics. In this chapter, I give a short introduction to ABC. I will explain the principle of ABC, discuss some of its advantages and limitations, and present strategies to overcome the latter. I will also illustrate ABC with two examples and give some hints for using ABC in practice. This introduction is not intended to be exhaustive, since excellent reviews already exist (see end of this chapter). Rather, I would like to introduce the concept and some notation. I hope to prepare the reader for chapters 3 and 4, in which ABC is used to infer mutation rates, male mating skew and migration rates in Alpine ibex, and where a methodological contribution to ABC is proposed.

2.2 The principle of ABC

In Bayesian statistics, the desired quantity is the posterior distribution of the parameter of interest, $\boldsymbol{\phi}$, given some observed data, D. Here, $\boldsymbol{\phi}$ is actually a vector of parameters, the components of which I denote by $\phi^{(k)}$ $(k = 1, \ldots, K)$. Let the data D also be multidimensional. For instance, they may represent the full allele frequency distribution or the joint site-frequency distribution from one or several populations, sampled at a certain number of loci. According to the Bayesian paradigm, the posterior distribution is proportional to the probability of the data given a certain parameter value, times the unconditional probability of that parameter value. The former is called the likelihood of the parameter, the latter is the prior distribution. More formally, we have

$$\pi(\boldsymbol{\phi} \mid D) \propto P(D \mid \boldsymbol{\phi})\, \pi(\boldsymbol{\phi}), \tag{2.1}$$

where $\pi(\boldsymbol{\phi} \mid D)$ is the posterior distribution, $P(D \mid \boldsymbol{\phi})$ the likeilihood, and $\pi(\boldsymbol{\phi})$ the prior distribution. The posterior is only proportional to the right-hand side of (2.1), because the latter

is not a proper probability density. To have equality, the right-hand side must be normalized by the total probability of the data, $P(D) = \int_\Phi P(D \mid \boldsymbol{\phi})\, \pi(\boldsymbol{\phi})\, \mathrm{d}\boldsymbol{\phi}$, so that

$$\pi(\boldsymbol{\phi} \mid D) = \frac{P(D \mid \boldsymbol{\phi})\, \pi(\boldsymbol{\phi})}{P(D)}. \tag{2.2}$$

$P(D)$ is also called the *marginal likelihood*, because the parameter(s) are marginalized over by integration. It is further referred to as the *prior predictive distribution*, emphasizing that no conditioning on the observed data has occurred yet. $P(D)$ is independent of the parameter value, and therefore a constant for a given prior range Φ.

For reasonably complex models – and therefore for most practical applications in evolutionary genetics – computation of $P(D)$ is challenging, because it involves that potentially complicated integration over the whole parameter space with prior support. In some situations, it is possible to compute $P(D \mid \boldsymbol{\phi})$ for a given value of $\boldsymbol{\phi}$, but just the integration is prohibitive. Then, Markov chain Monte Carlo (MCMC) or importance sampling (IM) techniques may be used to approximate the posterior distribution (see *e.g.* MacKay 2003). For these methods, the proportionality in (2.1) suffices, which is why $P(D)$ is not needed. MCMC and IS are well studied and established in the context of likelihood-based inference. However, they have a number of pitfalls and their application requires careful tuning (Marjoram and Tavaré 2006; Sisson et al. 2007; Kuhner 2009; Bertorelle et al. 2010). In cases where even computing $P(D \mid \boldsymbol{\phi})$ is prohibitive, alternative approaches are needed. ABC offers one by directly targeting the posterior distribution. Because it avoids calculation of the likelihood, ABC is sometimes referred to as a *likelihood-free* method of inference (Ratmann et al. 2007; Bazin et al. 2010; Sisson and Fan 2010). However, this is slightly misleading, since the likelihood does not *disappear* – it is just not explicitly calculated. The step of conditioning on the data – the conceptual meaning of a likelihood – is implicitly present in ABC, as will be seen below.

The central principle of ABC is that a large number of Monte Carlo simulations are performed under a model that is believed to explain how the observed data were generated. Each of the simulations takes as input a sample of parameter values $\boldsymbol{\phi}'$ from the prior distribution – one value for each of the components of $\boldsymbol{\phi}$ – and yields as output the simulated data D' with the same dimensionality as the observed data, D. The simulated data are then compared to the observed data, and those simulations that resulted in a close match between D' and D are accepted, the others rejected. The meaning of *close* in the previous sentence will be specified later. The parameter values associated with accepted simulations represent a direct sample from the posterior distribution of interest (Marjoram et al. 2003). The sample may be visualized in a histogram, or a continuous approximation to $\pi(\boldsymbol{\phi} \mid D)$ can be obtained via any density estimation method (*e.g.* Loader 1996). *Point estimates* such as the mode, mean or median are readily obtained, and *credible intervals* such as 95% highest posterior density (HPD) intervals can be calculated (Gelman et al. 2004). The choice and justification of the model under which simulations are performed is an interesting topic of its own, but beyond the scope of this text (see *e.g.* Gelman et al. 2004). We assume that the model is well chosen.

A generic rejection algorithm formalizing the above description is:

Generic rejection algorithm:

A.1 For $t = 1$ to $t = N$:

i Sample $\boldsymbol{\phi}'_t$ from $\pi(\boldsymbol{\phi})$.

ii Simulate D'_t from $P(D \mid \boldsymbol{\phi}'_t)$.

iii Accept $\boldsymbol{\phi}'_t$ if $D'_t = D$.

A.2 Estimate the posterior density $\pi(\boldsymbol{\phi} \mid D)$ from the accepted points.

In principle, this algorithm approximates the posterior distribution arbitrarily well for large enough N. It is further straightforward to parallelize it on a cluster computer, because the iterations are independent. This advantage carries over without limitation to some, but not all ABC algorithms (see below). However, the rejection algorithm above has limitations in practice. First, simulation under step A.1.ii may take some time, depending on the complexity of the model and the implementation. Therefore, there is a constraint on N and the approximation cannot be deliberately precise. Second, if D is high-dimensional, there is little chance for any simulations to be accepted in step A.1.iii. This renders posterior density estimation in A.2 problematic. To alleviate that second limitation, one may replace the rejection condition in step A.1.iii by:

A.1.iii' Accept $\boldsymbol{\phi}'_t$ if $\rho(D'_t, D) \leq \delta_\epsilon$,

where $\rho(\cdot)$ is some distance metric, and δ_ϵ a threshold defined on the same space as $\rho(\cdot)$ (see below). The threshold δ_ϵ is usually chosen implicitly such that a proportion ϵ of the N simulations is accepted. This adjustment implies a potential reduction of the dimensionality – $\rho(\cdot)$ may be lower-dimensional than D – and it allows for a control over the acceptance rate. Together, $\rho(\cdot)$ and δ_ϵ formalize what was meant by 'close' in the previous paragraph: a simulation is close to the observed target, if the distance as measured with the metric $\rho(\cdot)$ between the two is smaller than δ_ϵ. The result of this altered algorithm is a sample of independent and identically distributed observations from $\pi(\boldsymbol{\phi} \mid \rho(D'_t, D) \leq \delta_\epsilon)$, and hence an approximation to $\pi(\boldsymbol{\phi} \mid D)$ (Marjoram et al. 2003).

When D is high-dimensional or continuous, the above adjustment may still be inefficient. Therefore, the full data D are usually projected to a lower-dimensional set of summary statistics, $\mathbf{S}(D)$. Let $\mathbf{S}$ have p dimensions. This yields what is commonly referred to as the basic *ABC rejection algorithm*:

ABC rejection algorithm:

B.1 Compute $\mathbf{s} = \mathbf{S}(D)$.

B.2 For $t = 1$ to $t = N$:

i Sample $\boldsymbol{\phi}'_t$ from $\pi(\boldsymbol{\phi})$.

ii Simulate D'_t from $P(D \mid \boldsymbol{\phi}'_t)$, and compute the corresponding statistics $\mathbf{s}' = \mathbf{S}(D')$.

iii Accept $\boldsymbol{\phi}'_t$ if $\rho(\mathbf{s}', \mathbf{s}) \leq \delta_\epsilon$.

B.3 Estimate the posterior density $\pi(\boldsymbol{\phi} \mid D)$ from the accepted points.

This algorithm samples independent and identically distributed realisations of $\pi(\boldsymbol{\phi} \mid \rho(\mathbf{s}', \mathbf{s}) \leq \delta_\epsilon)$. If N were increased to infinity and δ_ϵ reduced to zero, the ABC rejection algorithm should converge to the generic rejection algorithm above, if the summary statistics are *sufficient*. A statistic is called sufficient, if the likelihood of the parameter of interest given the full data is the same as the likelihood of the parameter given the summary statistic. In other words, a summary statistic is sufficient, if it extracts all information that can be extracted from the full data on the parameter. However, in population genetics, hardly any commonly used summary statistic is sufficient. In practice, one therefore tries to choose an optimal combination of statistics. This choice is one of the main challenges in ABC (see below). Another choice that must be made is the one of the metric $\rho(\cdot)$. The Euclidean distance or a weighted version of it, *e.g.* the Mahalanobis distance (Mahalanobis 1936), is often used (Beaumont et al. 2002; Hamilton et al. 2005; Beaumont 2010). The rejection kernel – a function that assigns to each data point a weight according to which the point is considered for posterior estimation – may then be uniform as in Pritchard et al. (1999) or Blum and Tran (2010), or a Gaussian or an Epanechnikov kernel (Wilkinson 2008) (see example 1 below). These are all somewhat arbitrary *ad hoc* choices, and so far no explicit strategy for an optimal choice of $\rho(\cdot)$ has been suggested (but see Wilkinson 2008, for some guidance on this topic). In most applications, the summary statistics are scaled, for instance to have zero mean and unit variance (Beaumont 2010), prior to the computation of the metric. An alternative is to perform a principal component analysis to rotate and de-correlate the summary statistics (Leuenberger and Wegmann 2010). If the Mahalanobis distance is chosen, the scaling by the covariances is implicit. Such scaling makes the rejection condition less stringent along those summary statistics which are not very informative about the parameter, and more focussed on those that are. This makes sense, because the former mainly contribute noise that causes unjustified rejections and decreases the efficiency of the algorithm (see example 2 below). A third choice is that of the acceptance rate ϵ. This is the main tuning parameter, with a potentially strong influence on the accuracy of the posterior estimate. If ϵ is increased, more points are accepted, but these will on average be further away from the underlying truth and may introduce an error. If ϵ is chosen too small, few points will be accepted, such that the posterior estimate is affected by a large sampling variance.

To summarize, the ABC rejection algorithm is characterized by the following properties:

1. A *finite* number N of Monte Carlo simulations is performed and combined with a rejection step conditioning on the data to directly sample from the posterior distribution.
2. The full data D are *projected to a lower-dimensional set of summary statistics* $\mathbf{S}$ that are in most cases not sufficient.
3. Conditioning on the data is done with some *rejection tolerance* δ_ϵ.

These properties also represent the three approximations that coin the name of ABC. According to Beaumont (2010), the first rejection algorithm for Bayesian inference of population genetic parameters was proposed by Tavaré et al. (1997). Tavaré et al. (1997) also replaced the full

data by a summary statistic, but still relied on the likelihood being available analytically. Fu and Li (1997) and Weiss and von Haeseler (1998) replaced explicit calculation of likelihoods by a simulation step for one and multiple summary statistics, respectively. They sampled the parameters for their simulations from a grid of values, not from a prior distribution. The first 'real' ABC rejection algorithm was used in Pritchard et al. (1999) to study the demographic history of the human Y chromosome. As Marjoram et al. (2003) and Beaumont (2010) point out, some aspects that are now part of the ABC framework, such as the use of summary statistics instead of the full data, or the fitting of simulations to an observation, trace further back to Diggle and Gratton (1984) or Rubin (1984).

2.3 Three strategies to improve ABC

A limitation of the basic ABC rejection algorithm introduced in the previous section is its low efficiency: A large number of simulations must be performed, while only a small proportion can be accepted without substantial loss of precision. This tension increases with the number of summary statistics in $\mathbf{S}$, which is known as the *curse of dimensionality* (*e.g.* Blum and François 2010; Beaumont 2010). The curse of dimensionality describes the following phenomenon. Suppose we have performed $N = 10^5$ simulations, and that $\mathbf{S}_1$ has only $p = 1$ dimension. For the rejection step, we may then require that the one percent of simulations closest to the observed data are accepted, *i.e.* $\epsilon_0 = 0.01$. This results in 1'000 accepted simulations, enough for stable estimation of the posterior density. However, assume a different set of summary statistics, $\mathbf{S}_4$, with $p = 4$ dimensions. If we now apply the same rejection criterion as before to each of the four statistics individually, the overall acceptance rate drops to $\epsilon_0^p = 0.01^4$, or 10^{-6} percent. On average, no simulation will be accepted. The example is extreme, because usually, the statistics are not fully uncorrelated, and a less stringent rejection criterion is applied (see Pritchard et al. 1999; Beaumont 2010). Nevertheless, it reveals the need for some strategy to reduce the curse of dimensionality and improve the efficiency of ABC. Three strategies have been suggested, and I will discuss them in the following.

2.3.1 Post-rejection adjustment via regression

The first goes back to Beaumont et al. (2002), who proposed fitting a linear regression between the accepted parameter values and the corresponding summary statistics. The accepted parameter values are treated as response, and the corresponding values of the summary statistics as explanatory variables. Instead of estimating the posterior distribution from the accepted parameter values directly, one then estimates the posterior distribution from the values *predicted* by the linear regression, given the respective values of the summary statistics. The idea is that a linear relationship might hold at least in the vicinity of the observed data. To stress this, Beaumont et al. (2002) weighted the accepted points according to their distance from the observed data, using an Epanechnikov kernel (Fan and Gijbels 1996). The effect of the weighted local-linear regression is that accepted parameter values are projected along the line of the linear fit, which may compensate for the error introduced by accepting with some tolerance $\delta_\epsilon > 0$ (see example 2 below). Beaumont et al. (2002) were able to show this effect, suggesting that δ_ϵ may be substantially increased, and hence the acceptance rate improved, compared to the

basic ABC rejection algorithm. Increased acceptance rate is of interest because the posterior estimate becomes more robust as more points are available. Overall, this reduces the Monte Carlo error inherent to ABC (see Fearnhead and Prangle 2011). Beaumont et al. (2002) introduced their idea in a univariate context, but multivariate linear regressions can of course be fit if there is more than one parameter. A potential limitation of the approach by Beaumont et al. (2002) is that it assumes a linear relation and that the variance of the parameters is constant as the summary statistics change. Blum and François (2010) relaxed both assumptions, using a feed-forward neural network and showing improved performance compared to the original method by Beaumont et al. (2002). Leuenberger and Wegmann (2010) criticised the somewhat unnatural approach of regressing the parameters onto the summary statistics. Instead, they suggested fitting a general linear model with summary statistics as explanatory variables and parameters as response. The advantage is that this perspective allows for an approximation of the marginal likelihood, and hence for model comparison. Toghether, these approaches are often referred to as *post-rejection adjustment*, and ABC combined with them is called *ABC regression*, as opposed to *ABC rejection* (Beaumont 2010).

2.3.2 More efficient sampling in the ABC algorithm

The second strategy addresses the inefficiency of the proposal mechanism in the ABC rejection algorithm. There, in every iteration candidate parameter values $\boldsymbol{\phi}'$ are independently chosen from the prior distribution. If the prior distribution is broad compared to the (unknown) posterior distribution, this mechanism is very inefficient, because most of the time it proposes $\boldsymbol{\phi}'$ not anywhere near the range of acceptance (Beaumont 2010). It would be more efficient to adjust the proposal mechanism as – with an increasing number of iterations – more and more about the putative truth is being revealed. Two approaches have been devised. The first introduces a Metropolis-Hastings type MCMC step to ABC (Marjoram et al. 2003), the second enhances ABC with an adaptive sequential Monte Carlo (SMC) scheme (Sisson et al. 2007, 2009; Beaumont et al. 2009). I will briefly describe the two in turn. In the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970), a new candidate value $\boldsymbol{\phi}^*$ is proposed in every iteration according to a distribution that assigns a probability to the move from $\boldsymbol{\phi}'_{t-1}$ to $\boldsymbol{\phi}^*$. This distribution is called *proposal distribution* and often denoted by $q(\boldsymbol{\phi}'_{t-1} \to \boldsymbol{\phi}^*)$ (*e.g.* Marjoram et al. 2003). The proposed value $\boldsymbol{\phi}^*$ is then accepted according to the Metropolis-Hastings probability,

$$h = \min\left[1, \frac{P(D \mid \boldsymbol{\phi}^*)\,\pi(\boldsymbol{\phi}^*)\,q(\boldsymbol{\phi}^* \to \boldsymbol{\phi}'_{t-1})}{P(D \mid \boldsymbol{\phi}'_{t-1})\,\pi(\boldsymbol{\phi}'_{t-1})\,q(\boldsymbol{\phi}'_{t-1} \to \boldsymbol{\phi}^*)}\right]. \tag{2.3}$$

If accepted, $\boldsymbol{\phi}'_t$ is set to $\boldsymbol{\phi}^*$, otherwise $\boldsymbol{\phi}'_t = \boldsymbol{\phi}'_{t-1}$. In general, $q(\boldsymbol{\phi}'_{t-1} \to \boldsymbol{\phi}^*)$ is a function of $\boldsymbol{\phi}'_{t-1}$ and vice versa, which explains why consecutive values of $\boldsymbol{\phi}'$ are no longer independent – they form a Markov chain (MacKay 2003). In the context of ABC, the likelihoods in (2.3) are not available, and h cannot be computed. Hovewer, Marjoram et al. (2003) have devised an MCMC algorithm without the need of computing likelihoods, which has since become known as *MCMC-ABC*:

MCMC-ABC algorithm:

C.1 Compute $\mathbf{s} = \mathbf{S}(D)$.

C.2 Sample an initial value $\boldsymbol{\phi}'_0$ from $\pi(\boldsymbol{\phi})$.

C.3 For $t = 1$ to $t = N_{\mathrm{MCMC}}$:

i Propose $\boldsymbol{\phi}^*$ according to $q(\boldsymbol{\phi}'_{t-1} \rightarrow \boldsymbol{\phi}^*)$.

ii Simulate D'_t from $P(D \mid \boldsymbol{\phi}^*)$, and compute the corresponding statistics $\mathbf{s}' = \mathbf{S}(D')$.

iii If $\rho(\mathbf{s}', \mathbf{s}) \leq \delta_\epsilon$, go to C.3.iv, otherwise set $\boldsymbol{\phi}'_t = \boldsymbol{\phi}'_{t-1}$ and return to C.3.i.

iv Calculate

$$h_{\mathrm{ABC}} = \min\left[1, \frac{\pi(\boldsymbol{\phi}^*)\, q(\boldsymbol{\phi}^* \rightarrow \boldsymbol{\phi}'_{t-1})}{\pi(\boldsymbol{\phi}'_{t-1})\, q(\boldsymbol{\phi}'_{t-1} \rightarrow \boldsymbol{\phi}^*)}\right].$$

v Accept $\boldsymbol{\phi}^*$ with probability h_{ABC} and set $\boldsymbol{\phi}'_t = \boldsymbol{\phi}^*$; otherwise set $\boldsymbol{\phi}'_t = \boldsymbol{\phi}'_{t-1}$ and return to C.3.i.

C.4 Discard the first n_b accepted $\boldsymbol{\phi}^*$ values and estimate the posterior density $\pi(\boldsymbol{\phi} \mid D)$ from the remaining accepted $\boldsymbol{\phi}^*$.

In step C.4, the first accepted values are discarded to account for the so-called burn-in period, during which the trajectory of parameter values has not yet reached the stationary distribution. Overall, the hope is that the MCMC-ABC algorithm makes more efficient use of the available computation time, because it tends to suggest parameter values more likely to be accepted, compared to ABC rejection, where suggested parameter values are not correlated. Sacrificing uncorrelated sampling comes at a price, however: The MCMC-ABC algorithm is prone to the same issues as conventional MCMC. First, its mixing behavior can be bad, such that the chain becomes stuck in a region of low posterior probability, or the chain may move up to a local maximum of the posterior, but not to the global one (Sisson et al. 2007). Second, it is not obvious when the chain has converged and the algorithm can be stopped. A number of improvements have been suggested to address these issues (see Ratmann et al. 2007; Wegmann et al. 2009a).

As an alternative to MCMC-ABC, Sisson et al. (2007) proposed an adaptive version of ABC, embedding the rejection algorithm into a sequential Monte Carlo (SMC) framework. Their original version was biased, but has been corrected by Sisson et al. (2009) and Beaumont et al. (2009). The idea of this approach is twofold: Instead of having a fixed threshold δ_ϵ, one defines a sequence of decreasing tolerance thresholds $\delta_{(1)}, \ldots, \delta_{(T)}$ $(\tau = 1, \ldots, T)$. At each iteration, one chooses the next lower $\delta_{(\tau)}$ and re-samples $\boldsymbol{\phi}'$ from a weighted sample of parameters already accepted in the previous iteration. In the first iteration, parameter values are drawn from the prior, but in successive iterations, the posterior of the preceding iteration is used. Importance weights are used to correct for the fact that the values are no longer sampled from the prior. Because the rejection tolerance is reduced at every step, the weighted sets of parameters yield a gradually improved approximation to the posterior. Details and the full algorithm are given in Beaumont et al. (2009), Sisson et al. (2009) or Beaumont (2010), for instance. The advantage of *SMC-ABC* over MCMC-ABC is that it does not get stuck in a region of low acceptance probability. Compared to ABC rejection and ABC regression, SMC-ABC is more efficient, because it avoids drawing parameter values from regions with low posterior probability. This effect may be substantial, if the data are informative (Beaumont 2010).

2.3.3 Optimizing the choice of summary statistics

The third strategy to reduce the curse of dimensionality in ABC is to optimize the choice of summary statistics. Here, the goal is to select a set that is optimal in the sense that as much information is extracted from the original data as possible, with as few summary statistics as possible. The problem of summarizing large data sets is not unique to ABC (Nunes and Balding 2010) and commonly known as *variable selection* in statistics, or *feature selection* in machine learning (Hastie et al. 2011). As mentioned above, most summary statistics used by population geneticists are not sufficient. The number of different alleles observed under the infinite-alleles model of mutation (Kimura and Crow 1964) is a rare example of a sufficient statistic for the scaled mutation rate $\theta = 4N_e u$. Here, N_e is the effective population size and u the mutation rate per locus and generation (Ewens 1972). In this case, no additional information about a sample is needed to esimate θ (see example 2 below). In most other cases, however, the likelihood of the parameter given the full data is different from the likelihood given just a summary statistic. In principle, this hampers ABC completely, because the true posterior distribution is only approximated by the ABC-posterior if the statistics $\mathbf{S}$ are sufficient (*e.g.* Sisson and Fan 2010). In practice, however, this problem is not quite as drastic. Although not sufficient, most summary statistics have a theoretical justification; it can be shown that they are sensitive to the parameter of interest. Such statistics are good candidates for ABC, and empirical results seem to confirm this. A systematic approach for choosing summary statistics in ABC has long been missing, and statistics were usually chosen based on theory for simpler models, and on the researcher's intuition.

The first systematic approach was proposed by Joyce and Marjoram (2008). The authors used a sequential scheme and employed the concept of *approximate sufficiency*. The idea is to start with a set of candidate statistics $\mathbf{S} = (S^{(1)}, \ldots, S^{(p)})$, and to ask if adding a further candidate statistic, $S^{(p+1)}$, has an effect on the posterior that is larger than some threshold. If the effect is smaller than the threshold, adding $S^{(p+1)}$ is not needed; $\mathbf{S}$ is approximately sufficient. Otherwise, $S^{(p+1)}$ is added, and the procedure is repeated with a new candidate $S^{(p+2)}$. While the theoretical results motivating this procedure are straightforward, the implementation is somewhat tricky (see Appendix of Joyce and Marjoram 2008). One issue is that the result may depend on the order in which candidate statistics are added to $\mathbf{S}$. If the total number of statistics is large, testing all possible configurations is too expensive. Joyce and Marjoram (2008) suggested a forward-backward heuristic to tackle this; some belief is needed that this simpler strategy does not miss out on a relevant combination of statistics.

It is worth noting a point that was first brought to my attention by Andreas Futschik: Sufficiency is a global concept, whereas in practice a summary statistic may be informative in some part of the parameter space, but uninformative in the region of the parameter space that matters for the actual estimation problem. Therefore, rather than choosing statistics with respect to the whole prior range, one could imagine focussing the choice on the (putative) neighbourhood of the true value. The truth is of course not known in advance. In chapter 3, my collaborators and I propose a solution to this. Along the same lines, Nunes and Balding (2010) proposed the following two-step procedure: They first used a minimum-entropy algorithm to identify simulated data sets close to the observed ones. Then, they successively regarded these

simulated sets as observed data sets, computed the error for all possible sets of summary statistics, and then chose the set which minimized the mean error across the data sets. A potential limitation of this approach is that, as was the case for the approach by Joyce and Marjoram (2008), assessing *all* combinations of candidate statistics is expensive if there are many. Another, similar approach was proposed recently by Fearnhead and Prangle (2011). The authors first proved that, given a certain criterion by which the discrepancy between the true and inferred value is measured, an optimal summary statistic can be defined. For instance, if the criterion is the quadratic loss function, the optimal statistic is the posterior mean; if the criterion is the absolute error, the optimal statistic is the posterior median. In practice, these quantities are of course not known. Therefore, Fearnhead and Prangle (2011) devised a heuristic multi-step procedure, in which a pilot ABC study is used to define the putative vicinity of the true parameter value, a number of training data sets are simulated with known true values, and a linear regression is fit to these training data. For each parameter, one linear predictor is obtained from a set of candidate statistics, by regressing the parameter values linearly against a function of the candidate summary statistics. These predictors are then used as the summary statistics in the final ABC analysis. Fearnhead and Prangle (2011) call their approach *semi-automatic*, because the choice of summary statistics is based on simulations. However, there are still choices to be made by the user with respect to the set of candidate statistics, potential scaling of them, and the type of regression used.

As an alternative, Wegmann et al. (2009a) proposed performing a partial least squares (PLS) regression of the parameters on the summary statistics. PLS regression is similar to principal component analysis (PCA) in which the explanatory variables (summary statistics in this case) are de-correlated. In addition, however, PLS also takes into account the relation with the response variables (parameters in this case), therefore jointly optimizing both criteria. A leave-one-out cross-validation was then used by Wegmann et al. (2009a) to find the optimal number of PLS components to keep, based on the root mean squared error. Since PLS assumes a linear relationship between summary statistics and parameters, Wegmann et al. (2009a) applied a Box-Cox transformation (Box and Cox 1964) to the summary statistics, prior to PLS regression. The hope is that PLS results in a reduced, less correlated set of summary statistics compared to the original set of candidate statistics.

To summarize, the following three strategies have been proposed to reduce the curse of dimensionality and increase the efficiency of the basic ABC rejection algorithm:

1. Improved density estimation, allowing for larger rejection tolerance
2. Correlated sampling of parameter values to improve the acceptance rate
3. Optimal choice of summary statistics

2.4 Examples

In the following, I will show two examples that illustrate different aspects of ABC. The first is to show the error introduced by having a rejection tolerance $\delta_\epsilon > 0$. The second will illustrate the effect of non-sufficient statistics and the post-rejection adjustment with a local-linear regression as proposed by (Beaumont et al. 2002).

2.4.1 Example 1: Estimating the mean of a Gaussian distribution

I have borrowed this example from a draft for a book chapter by Sisson and Fan (2010). We will use ABC to infer the mean μ of a univariate Gaussian distribution with known variance $\sigma^2 = 1$, which I denote by $N(\mu, 1)$. To better distinguish between observed and simulated data, I will use y instead of D for the observed data, and x instead of D' for the simulated data. Since there is only one parameter, there is no need to use the bold-face symbol denoting a vector, so we have $\phi = \mu$. Moreover, for the mean of a Gaussian distribution, the observations x are sufficient summary statistics. We can therefore set $S(x) = x$. Notice that we are dealing with one single observation from a univariate Gaussian distribution. The goal of this example is to illustrate the error introduced in ABC when rejection is performed with some tolerance $\delta_\epsilon > 0$. For this, it is helpful to consider the following formalization of the marginal posterior distribution obtained with ABC (Sisson and Fan 2010):

$$\pi_{\text{ABC}}(\phi \mid y) \propto \pi(\phi) \int_{\mathcal{Y}} P(y \mid x, \phi)\, P(x \mid \phi)\, \mathrm{d}x, \tag{2.4}$$

where $P(y \mid x, \phi)$ is the error introduced by ABC in addition to the Monte Carlo error (see Fearnhead and Prangle 2011). $P(y \mid x, \phi)$ is determined by the rejection kernel and the tolerance δ_ϵ – two choices that have to be made when implementing ABC. A common choice for $P(y \mid x, \phi)$ is the uniform kernel density, such that

$$P_\epsilon(y \mid x, \phi) \propto \begin{cases} 1 & \text{if } \rho\big(S(x), S(y)\big) \leq \delta_\epsilon \\ 0 & \text{otherwise,} \end{cases} \tag{2.5}$$

where the subscript to P should emphasize the dependence on ϵ. Further, we will use the Euclidean distance for $\rho(\cdot)$. Recalling $S(x) = x$, we then have $\rho\big(S(x), S(y)\big) = ||x - y|| = \sqrt{(x-y)^2}$. Going back to our specific example, let us assume that the true posterior $\pi(\mu \mid y)$ is the univariate standard Gaussian, $N(0, 1)$, *i.e.* that $\mu = 0$. For the univariate Gaussian distribution, the likelihood $P(x \mid \mu)$ is available analytically and simply specified by $x \sim N(\mu, 1)$. We further set the observed data point $y = 0$ and choose a uniform prior $\pi(\mu) \propto 1$. With the rejection kernel given in (2.5), one can show that the ABC-posterior is

$$\pi_{\text{ABC}}(\mu \mid y) \propto \frac{\Phi(\epsilon - \mu) - \Phi(-\epsilon - \mu)}{2\epsilon}, \tag{2.6}$$

where $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function (Sisson and Fan 2010). One can further show that $\pi_{\text{ABC}}(\mu \mid y) \to N(0, 1)$ as $\epsilon \to 0$, as we would expect. Figure 2.1 shows the effect of δ_ϵ on the quality of the approximation $\pi_{\text{ABC}}(\mu \mid y)$ to $\pi(\mu \mid y)$. The smaller δ_ϵ, the closer the ABC posterior is to the true posterior. In this case, $N = 10^5$ ABC simulations were performed. Even with the smallest tolerance, $\delta_\epsilon = \sqrt{3}/10$, about $4,400$ points were accepted – enough for robust posterior density estimation.

2.4.2 Example 2: Estimating the parameter of the Ewens sampling formula

This example is motivated by Joyce and Marjoram (2008) who used it to give a proof of concept for their method for choosing summary statistics. As mentioned earlier, the number of different alleles observed in a sample is a sufficient statistic for the scaled mutation rate θ

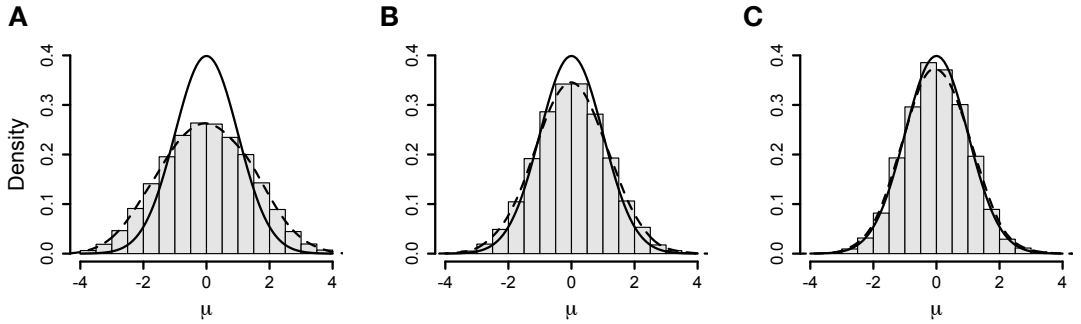

Figure 2.1: Effect of the rejection tolerance δ_ϵ on the posterior variance of ABC when estimating the mean μ of a univariate Gaussian as discussed in example 1. The thick line is the true posterior, a standard Gaussian distribution $N(0,1)$. The histogram represents the distribution of μ values accepted with ABC, and the dashed line is the fit of a continuous density to this distribution. (A)–(C) correspond to δ_ϵ values of $\sqrt{3}$, $\sqrt{3}/2$ and $\sqrt{3}/10$, with about $44,000$, $22,000$ and $4,400$ accepted points, respectively.

under the infinite-alleles model of mutation (Kimura and Crow 1964). This goes back to the Ewens sampling formula (Ewens 1972), which gives the probability that a sample of n gene copies contains k allele types and that in this sample, there are $a_1, a_2, \ldots, a_n$ alleles present $1, 2, \ldots, n$ times:

$$P\{k, a_1, a_2, \ldots, a_n\} = \frac{n!}{\theta_{(n)}} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{a_j} \frac{1}{a_j!}, \tag{2.7}$$

where $\theta_{(n)} = \theta(\theta+1)\cdots(\theta+n-1)$ (Hein et al. 2005; Wakeley 2009). Note that $\sum_{j=1}^{n} a_j = k$ and that (2.7) holds only for sampling configurations that satisfy $\sum_{j=1}^{n} j\, a_j = n$. Notice that the term θ^{a_j} in the product in (2.7) may be replaced by θ^k outside the product, because

$$\prod_{j=1}^{n} \theta^{a_j} = \theta^{\sum_{i=1}^{n} a_j} = \theta^k. \tag{2.8}$$

This makes the dependence of the probability in (2.7) on k explicit. The crucial property of the Ewens sampling formula is that *conditional* on some k, the probability of a sampling configuration does not depend on θ:

$$P\{a_1, a_2, \ldots, a_n \mid k\} = \frac{n!}{s_n^k} \prod_{j=1}^{n} \frac{1}{j^{a_j}\, a_j!}, \tag{2.9}$$

where s_n^k is the Sterling number of the first kind (Wakeley 2009). Therefore, k is a sufficient statistic for θ. In the following, we will infer θ using ABC, where k is an obvious choice for a summary statistic, $S^{(1)} = k$. To illustrate the effect of using non-sufficient statistics, we will add two more statistics. First, let us add the number of singletons $S^{(2)} = a_1$, *i.e.* the number of alleles that occur only once in the sample. This statistic is expected to contain at least some information about θ. Second, we will add as a third statistic $S^{(3)}$ random number drawn from a uniform distribution between 0 and 25. So, we have $\mathbf{S} = (S^{(1)}, S^{(2)}, S^{(3)})$. For ABC, we perform $N = 10^5$ simulations with θ drawn from a uniform prior between 0 and 10. Figure 2.2 shows the summary statistics as a function of θ. As expected, $S^{(1)}$ and $S^{(2)}$ depend on θ, while $S^{(3)}$ shows no correlation.

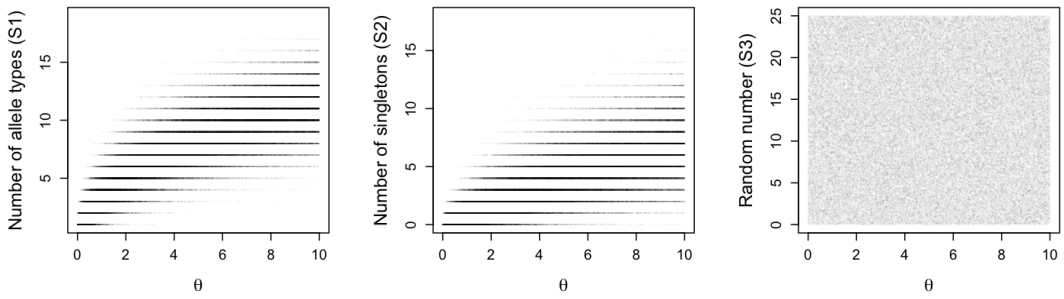

Figure 2.2: Summary statistics from example 2 as a function of the parameter, the scaled mutation rate θ. Notice that the summary statistics $S^{(1)}$ and $S^{(2)}$ only take discrete values, which is why the points appear on a grid along the y-axis (*cf.* Figure 2.3).

Let us assume that the true parameter was $\theta = 4$. A draw from the Ewens sampling formula then resulted in $S^{(1)} = 7$, $S^{(2)} = 3$ and $S^{(3)} = 1.539$; this is our observation. For rejection, we then use the Euclidean distance as metric $\rho(\cdot)$ and a uniform rejection kernel with $\delta_\epsilon = 0.4$. Moreover, to compare the effect of the number of dimensions and the non-sufficient statistics, we first condition on all three statistics, and then repeat rejection conditioning only on the sufficient statistic $S^{(1)}$. In both cases, we perform a weighted local-linear multivariate regression after rejection (see above). Figure 2.3 illustrates both the rejection and the regression step. The grey points represent all simulations, the blue ones are those which were accepted. Green and white arrows show the effect of regression for a set of points chosen at random. The arrows lead from the original position of the points to the position after the projection along the regression line. Figures 2.3A–2.3C show this for the case where we conditioned on all three statistics; Figure 2.3D applies to the case where we only used the sufficient statistic. The effect of more dimensions is that the accepted points are on average further apart from the observed value. This is because we conditioned on $\delta_\epsilon N$ points being accepted. The hope is that the post-rejection adjustment via the local-linear regression would to some degree correct for the error introduced by the large rejection tolerance. Figures 2.3A and 2.3B show that this was actually the case: accepted points relatively far away from the observed summary statistic are projected closer to the known true value. Figure 2.3B suggests that the effect of the random summary statistic $S^{(3)}$ was to add noise, since points really far from the observed statistic got accepted. These points were far from the observation only in that dimension; they were probably very close to the observation in the direction of $S^{(1)}$ and $S^{(2)}$. Otherwise, they would not have had a Euclidean distance from the observation small enough to be accepted. In Figure 2.4, the posterior distributions obtained with ABC are compared to the true posterior computed from equation (2.9). The posteriors from ABC without regression are further from the true posterior than those with regression. With regression, it did not make a big difference whether all or only the sufficient statistic was used. This confirms that the post-rejection adjustment did a good job in correcting for both the higher number of dimensions and the presence of non-sufficient summary statistics. It is worth pointing out that for three summary statistics, the curse of dimensionality is not yet that strong. In chapters 3 and 4 we will encounter cases where this is not necessarily the case, and where it becomes crucial to chose summary statistics well.

2.5 ABC in practice

In practice, the three strategies to increase the efficiency of ABC – post-rejection adjustment, correlated sampling, and optimization of the choice of summary statistics – may be combined in various ways. Therefore, one will find a variety of ABC subtypes in applied studies. In general, it is not obvious in advance what combination is optimal, and one should perform a simulation study in which the accuracy of alternative combinations is compared. Various measures of accuracy may be employed, such as the absolute, relative or root mean squared error of

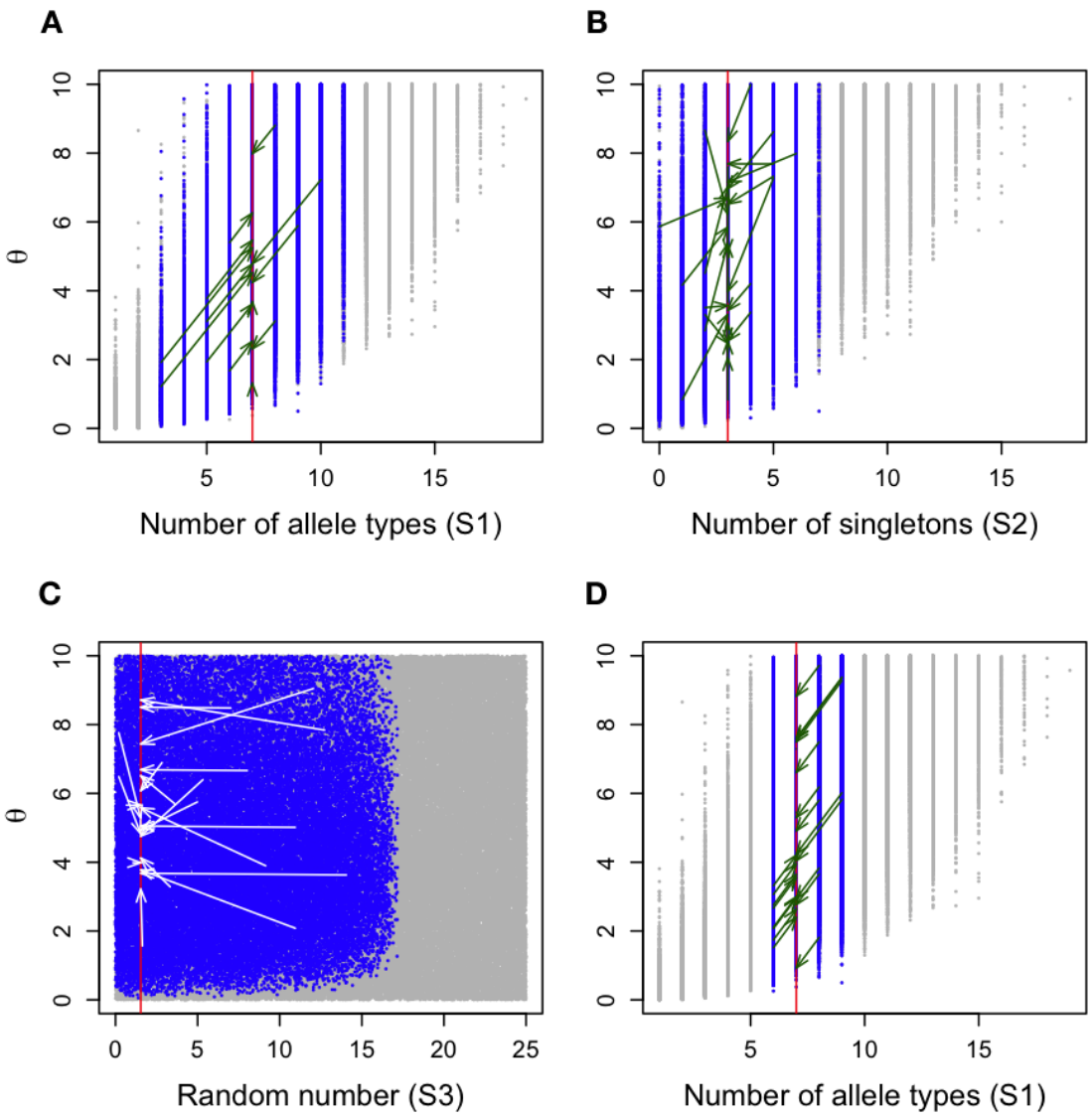

Figure 2.3: Illustration of the rejection and regression step in ABC. Grey points are ABC simulations, blue points are the accepted simulations. The vertical red line is the observed statistic, and green or white arrows show how accepted points were projected by the weighted local-linear regression. (A)–(C) ABC with all three summary statistics. One plot is shown for each dimension. (D) ABC with only the sufficient statistic. Notice that the summary statistics $S^{(1)}$ and $S^{(2)}$ only take discrete values, which is why the points in (A), (B) and (D) appear on a grid along the x-axis.

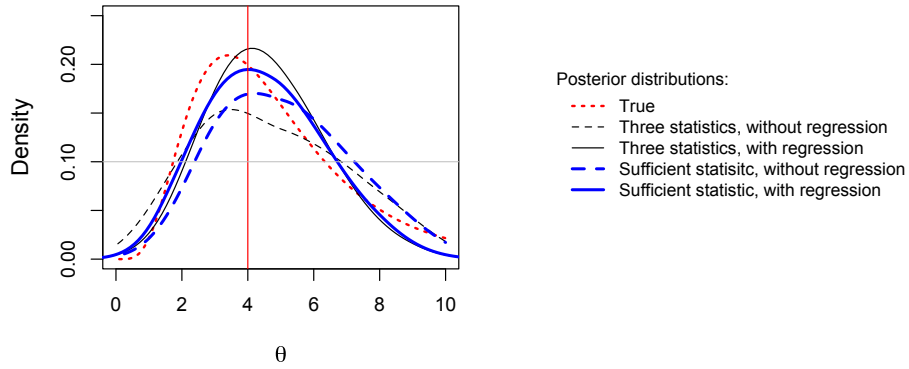

Figure 2.4: Posterior distributions inferred in example 2. The vertical red line denotes the true value, and the red dotted line is the true posterior (see text).

the posterior point estimate. Moreover, it is advisable to assess the coverage properties of the posterior distribution. This may be done by simulating a set of test data sets with known true parameter values sampled from the prior distribution, and to then compute the estimated posterior probabilities of the true values. By definition, for a proper probability density function, these probabilites should be uniformly distributed (Cook et al. 2006). This also holds for ABC posteriors (Wegmann et al. 2009a). The uniformity can be tested with a Kolmogorov-Smirnov test (Sokal and Rohlf 1981). Moreover, a histogram of the posterior probabilities reveals the kind of deviation from uniformity. For instance, a left-skewed distribution of probabilities, *i.e.* one with a long tail on the left and most of its mass on the right side, implies that the true parameter was on average underestimated, and vice versa (*cf.* Wegmann et al. (2009a) or chapter 3).

The possibility to combine alternative methods for the various steps in ABC is one reason for its versatility and certainly an advantage when it comes to tailoring ABC to a particular application. From the point of view of introducing new approaches for one of the ABC steps, the downside is that there is so far no 'standard' ABC procedure against which innovations are being compared. Moreover, defining a standard ABC procedure would not be enough, because the performance of a given strategy is likely to depend also on the model studied. In principle, one would therefore need a *standard ABC setting* – defining both the ABC steps and the model – as a reference. While this is desirable for comparison of alternative approaches to individual ABC steps, it contradicts the common practice and the *ad hoc* character of applied ABC.

It is also worth checking whether the model used to simulate the data is plausible, in other words, if it is possible at all to obtain $\mathbf{S}(D')$ in the range of the observed summary statistics $\mathbf{S}(D)$. This can be assessed by plotting the prior predictive distribution, *i.e.* the joint distribution of the summary statistics, together with the point $\mathbf{S}(D)$. If the cloud of simulated points $\mathbf{S}(D'_t)$, $(t = 1, \ldots, N)$ covers $\mathbf{S}(D)$ well, the model is well specified. In practice, it is hard to visualize this in more than two, perhaps three, dimensions. Then, pairs of components of $\mathbf{S}$ should at least be plotted. If any of the $p(p-1)/2$ pairwise plots reveals that $\mathbf{S}$ is not well covered by the simulated point cloud, one should be sceptical. Unfortunately, the inverse

conclusion is not justified: if everything is fine with the pairwise prior predictive plots, there is no guarantee that the same holds for triples, for instance.

Several software packages facilitating inference with ABC are available. For instance, `DIYABC` (Cornuet et al. 2008) is a user-friendly program with a graphical interface that allows for inference under a great variety of demographic models. A collection of simulation programs and scripts for various steps of the ABC workflow is offered by `ABCtoolbox` (Wegmann et al. 2010). The advantage of `ABCtoolbox` is its versatility; the user can design combinations of existing programs with her own code, or adjust previous versions to the particular needs of a project. In comparison to `DIYABC`, `ABCtoolbox` requires some familiarity with command line environments and coding. Further, the `abc` package (Csilléry et al. 2011) for `R` (R Development Core Team 2011) implements various methods for rejection and density estimation, once data have been simulated.

2.6 Further reading

The principles, history and different flavors of ABC are described in much more detail in an excellent and exhaustive review by Beaumont (2010). That review also includes a summary on where ABC has been applied so far. Csilléry et al. (2011) review some practical aspects and applications of ABC, while Bertorelle et al. (2010) describe its flexibility and discuss advantages and limitations. Moreover, Bertorelle et al. (2010) give a nicely illustrated step-by-step description of the workflow in a typical ABC project. Some more details, hints and pitfalls relevant for application of ABC may be found in the manual for `ABCtoolbox` (see above) by Wegmann et al. (2009b). ABC and its relation to other Bayesian methods of inference in genetics are reviewed by Beaumont and Rannala (2004). A broader review on modern computational approaches for analysis of genetic data, including short descriptions of the coalescent theory, importance sampling, Markov chain Monte Carlo, ABC as well as examples of application has been given by Marjoram and Tavaré (2006).

Chapter 3

Choice of summary statistics in ABC via boosting and application to the estimation of mutation rates and mating skew in Alpine ibex (Capra ibex)

The work presented in this chapter was influenced by discussions with Andreas Futschik and Mark Beaumont. Andreas has suggested to use boosting for the choice of summary statistics. The chapter is intended for publication in Genetics*, as a companion paper to the one resulting from chapter 4 of this thesis, with Andreas and Mark as co-authors.*

3.1 Introduction

Understanding the mechanisms leading to observed patterns of genetic diversity has been a central objective since the beginnings of population genetics (Fisher 1922b; Haldane 1932; Wright 1951; Charlesworth and Charlesworth 2010). Three recent trends keep advancing this undertaking: i) molecular data are becoming available at an ever higher pace (Rosenberg et al. 2002; Frazer et al. 2007); ii) new theory is being developed (Griffiths and Tavaré 1994a; Wakeley 2004, 2009); and iii) increased computational power allows solution of problems that were intractable just a few years ago. In parallel, the focus has shifted to inference under complex models (*e.g.* Fagundes et al. 2007; Blum and Jakobsson 2011), and to the joint estimation of parameters (*e.g.* Williamson et al. 2005). Usually, these models are stochastic. The increasing complexity of models is justified by the underlying processes: inheritance, mutation, chromosomes, modes of reproduction and spatial structure. On the other hand, complex models are often not amenable to inference based on exact analytical results. Instead, approximate methods such as Markov chain Monte Carlo (MCMC, Gelman et al. 2004) or approximate Bayesian computation (ABC, Beaumont and Rannala 2004) are used. A significant part of research in the field is currently devoted to the refinement and development of such methods (Wakeley 2004; Marjoram and Tavaré 2006). ABC (Fu and Li 1997; Tavaré et al. 1997; Weiss and von Haeseler 1998; Pritchard et al. 1999; Beaumont et al. 2002) is a Monte Carlo method of inference that emerged from the confrontation with models for which the evaluation of the likelihood is

computationally prohibitive or impossible. It may be viewed as a class of rejection algorithms (Marjoram et al. 2003; Marjoram and Tavaré 2006). The principle is to first simulate data under the model of interest, and to then accept simulations that produced data close to the observation. Parameter values belonging to accepted simulations yield an approximation to the posterior distribution, without the need to explicitly calculate the likelihood. The full data are usually compressed to summary statistics in order to reduce the number of dimensions. Formally, the posterior distribution of interest is given by

$$\pi(\boldsymbol{\phi} \mid D) = \frac{P(D \mid \boldsymbol{\phi})\, \pi(\boldsymbol{\phi})}{P(D)} = \frac{P(D \mid \boldsymbol{\phi})\, \pi(\boldsymbol{\phi})}{\int_{\Phi} P(D \mid \boldsymbol{\phi})\, \pi(\boldsymbol{\phi})\, \mathrm{d}\boldsymbol{\phi}}, \tag{3.1}$$

where $\boldsymbol{\phi}$ is a vector of parameters living in space Φ, D denotes the observed data, $\pi(\boldsymbol{\phi})$ the prior distribution, and $P(D \mid \boldsymbol{\phi})$ the likelihood. With ABC, (3.1) is approximated by

$$\pi(\boldsymbol{\phi} \mid \mathbf{s}) \;\propto\; P\big(\rho(\mathbf{s}', \mathbf{s}) \leq \delta_\epsilon \mid \boldsymbol{\phi}\big)\, \pi(\boldsymbol{\phi}), \tag{3.2}$$

where $\mathbf{s}$ and $\mathbf{s}'$ are abbreviations for realisations of $\mathbf{S}(D)$ and $\mathbf{S}(D')$, respectively, and $\mathbf{S}$ is a function generating a q-dimensional vector of summary statistics calculated from the full data. The prime denotes simulated points, in contrast to the summary statistics of the observed data. Further, $\rho(\cdot)$ is a distance metric and δ_ϵ the rejection tolerance in that metric space, such that a proportion ϵ of all simulated points is accepted. ABC, its position in the ensemble of model-based inference methods, and its application in evolutionary genetics are reviewed in Marjoram et al. (2003), Beaumont and Rannala (2004), Marjoram and Tavaré (2006), Beaumont (2010), Bertorelle et al. (2010) and Csilléry et al. (2010). Although the origin of ABC is generally assigned to Fu and Li (1997) and Tavaré et al. (1997), some aspects, such as the summary description of the full data, inference for implicit stochastic models and algorithms directly sampling from the posterior distribution trace further back (*e.g.* Diggle 1979; Diggle and Gratton 1984; Rubin 1984).

A fundamental issue with the basic ABC rejection algorithm (*e.g.* Marjoram et al. 2003) is its inefficiency: a large number of simulations is needed to obtain a satisfactory number of accepted runs. This problem becomes worse as the number of summary statistics increases and is known as the curse of dimensionality. Three solutions have been proposed: i) more efficient algorithms combining ABC with principles of MCMC (*e.g.* Marjoram et al. 2003; Wegmann et al. 2009a) or sequential Monte Carlo (*e.g.* Sisson et al. 2007; Beaumont et al. 2009; Sisson et al. 2009; Toni et al. 2009); ii) fitting a statistical model to describe the relationship of parameters and summary statistics after the rejection step, allowing for a larger tolerance δ_ϵ (Beaumont et al. 2002; Blum and François 2010; Leuenberger and Wegmann 2010); and iii) reduction of dimensions by sophisticated choice of summary statistics (*e.g.* Joyce and Marjoram 2008; Wegmann et al. 2009a). Point iii) is related to two further issues. First, most summary statistics in evolutionary genetics are not sufficient. A summary statistic is sufficient for a parameter, if the likelihood of that parameter given the summary statistic is proportional to the likelihood of the parameter given the full data. Second, the choice of summary statistics implies the choice of a suitable metric $\rho(\cdot)$ to measure the 'closeness' of simulations to observation. The Euclidean distance (or a weighted version, *e.g.* Hamilton et al. 2005) has been used in most applications, but it is not obvious why this should be optimal. The Euclidean distance is a

scale-dependent measure of distance – changing the scale of measurement changes the results. Since this scale is determined by the summary statistics, the choice of summary statistics is linked to the choice of the metric. For these reasons, the choice of summary statistics should not only aim at reducing the dimensions, but at extracting (combinations of) statistics that contain the essential information about the parameters of interest. Moreover, the choice of the metric should be considered. The first two problems are reminiscent of the classical problem of variable selection in statistics and machine learning (e.g. Hastie et al. 2011).

The choice of summary statistics in ABC has become a focus of research only recently. Joyce and Marjoram (2008) proposed a sequential scheme based on the principle of approximate sufficiency. Statistics are included if their effect on the posterior distribution is larger than some threshold. Their approach seems demanding to implement in practice, and it is not obvious how to define an optimal threshold. Wegmann et al. (2009a) used partial least squares (PLS) regression to choose summary statistics. In this context, PLS regression can be used to seek linear combinations of the original summary statistics that are maximally decorrelated and, at the same time, have high correlation with the parameters (Hastie et al. 2011). A reduction in dimensions is achieved by choosing only the first n PLS components. This choice is based on cross-validation. PLS is one out of several approaches for variable selection (Hastie et al. 2011), but it is an open question how it compares to alternative methods in any specific ABC setting. Moreover, the optimal choice of summary statistics may depend on the location of the true (but unknown) parameter values. By definition, this is to be expected whenever the summary statistics are not sufficient. Therefore, it is not obvious why methods that assess the relation between statistics and parameters on a global scale should be optimal. Instead, focussing on the correlation only in the (supposed) neighborhood of the true parameter values might be preferable. The problem is of course that this neighborhood is not known in advance – otherwise we would not need ABC. However, the neighborhood may be established approximately, as we will argue later. The idea of focussing the choice of summary statistics on some local optimization has recently also been followed in two papers by Nunes and Balding (2010) and Fearnhead and Prangle (2011). Nunes and Balding (2010) proposed to use a minimum-entropy algorithm to identify the neighborhood of the true value, and then chose the set of summary statistics that minimized the mean squared error across a test data set. Fearnhead and Prangle (2011), on the other hand, first proved that, for a given loss function, an *optimal* summary statistic may be defined; for the quadratic loss, the optimal summary statistic is the posterior mean. Since this is not available *a priori*, the authors devised a heuristic to estimate it, and were able to show good performance of their approach. The choice of the optimization criterion may include a more local or a global focus on the parameter range. Different criteria will lead to different optimal summary statistics. The approaches by Nunes and Balding (2010) and Fearnhead and Prangle (2011), and the one we will take here, have in common that they employ a two-step procedure, first defining 'locality', and then using standard methods from statistics or machine learning to select summary statistics in this restricted range. They differ in the details of these two steps.

Here, we propose a novel approach for choosing summary statistics in ABC. It is based on boosting, a method developed in machine learning to establish the relationship between predictors and response variables in complex models (Freund 1995; Freund and Schapire 1996,

1999; Schapire 1990). It has been argued that boosting is relatively robust to overfitting (Friedman et al. 2000), which would be an advantage with regard to high-dimensional problems as encountered in ABC. Different flavors of boosting exist, depending on assumptions about the error distribution, the loss function and the learning procedure. In a simulation study, we compare the performance of ABC with three types of boosting to ABC with summary statistics choosen via PLS, and to ABC with all candidate statistics. We further suggest an approach for choosing summary statistics locally, and compare the local variants of the various methods to their global versions. Throughout, we study a model that is motivated by the re-introduction of Alpine ibex (*Capra ibex*) into the Swiss Alps. The parameters of interest are the mean and standard deviation across microsatellites of the scaled ancestral mutation rate, and the proportion of males that obtain access to matings per breeding season. This model is used first in the simulation study for inference on synthetic data and assessment of accuracy. Later, we apply the best method to infer posterior distributions given genetic data from Alpine ibex.

3.2 Model and parameters

We study a neutral model of a spatially structured population with genetic drift, mutation and migration. The demography includes admixture, subdivision and changes in population size. This model is motivated by the recent history of Alpine ibex and their re-introduction into the Swiss Alps (Figures 3.1 and 3.2). By the end of the 18^{th} century, Alpine ibex had been extinct except for about 100 individuals in the *Gran Paradiso* area in Northern Italy (Figure 3.1). At the beginning of the 20^{th} century, a schedule was set up to re-establish former demes in Switzerland (Couturier 1962; Stuwe and Nievergelt 1991; Scribner and Stuwe 1994; Maudet et al. 2002). The re-introduction has been documented in great detail by game keepers and authorities. We could therefore reconstruct for 35 demes their census sizes between 1906 and 2006 (Supporting File 3.6 *census sizes*) and the number of females and males transferred between them, as well as the times of these founder/admixture events (Supporting File 3.7 *transfers*). Inference on mutation and migration can therefore be done conditional on this information. The signal for this inference comes from the distribution of allele frequencies across loci and across demes.

We constructed a forwards in time model starting with an ancestral gene pool d_{anc} of unknown effective size, N_e, representing the *Gran Paradiso* ibex deme. At times t_1 and t_2, two demes, d_1 and d_2, are derived from the ancestral gene pool. They represent the breeding stocks that were established in two zoological gardens in Switzerland in 1906 and 1911 (Figure 3.1; Stuwe and Nievergelt 1991). Further demes are then derived from these. In general, we let t_i be the time at which deme d_i is established. Once a derived deme has been established, it may contribute to the foundation of additional demes. The sizes of derived demes follow the observed census size trajectories (Supporting File 3.6 *census sizes*). We interpolated missing values linearly, if the gap was only one year, or exponentially, if values for two or more successive years were missing. Derived demes may exchange migrants if they are connected. This depends on information obtained from game keepers and on geography (Figure 3.1). Given a pair of connected demes d_i and d_j, we define the forward migration rates, $\tilde{m}_{i,j}$ and $\tilde{m}_{j,i}$. More precisely, $\tilde{m}_{i,j}$ is the proportion of potential emigrants (see Supporting Information (SI)) in deme d_i that

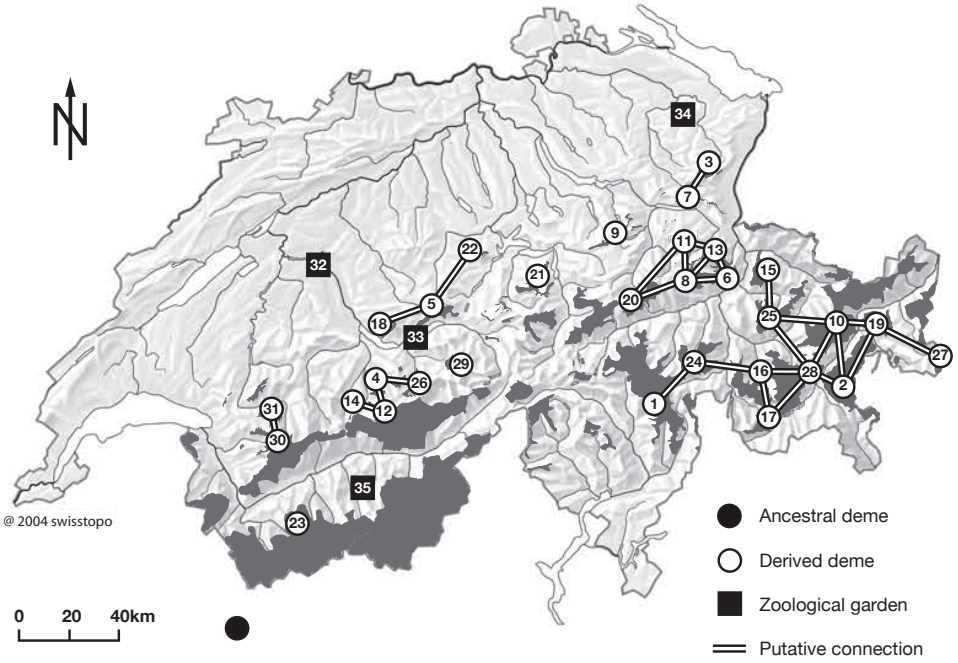

Figure 3.1: Location of Alpine ibex demes in the Swiss Alps. The dark shaded parts represent areas inhabited by ibex. The ancestral deme is located in the *Gran Paradiso* area in Northern Italy, close to the Swiss border. The two demes in the zoological gardens 33 and 34 were first established from the ancestral one. Further demes, including the two in zoological gardens 32 and 35, were derived from demes 33 and 34. Putative connections indicate the pairs of demes for which migration is considered possible. For a detailed record of the demography and the genealogy of demes see Figure 3.7 and Supporting File *transfers*. For deme names see Table 3.5. Map obtained via the Swiss Federal Office for the Environment (FOEN) and modified with permission.

migrate to deme d_j per year. We assume that $\tilde{m}_{i,j}$ is constant over time and the same for females and males. Migration is included in the model, although we do not estimate migration rates in this paper, but in a companion paper (see Aeschbacher et al. 2011b, or chapter 4). A schematic representation of the model is given in Figure 3.2.

Population history is split into two phases. The first started at some unknown point in the past and ended at $t_1 = 1906$, when the first ibex were brought from *Gran Paradiso* (d_{anc}) to d_1. For this ancestral phase, we assume constant, but unknown effective size N_e, and mutation following the single stepwise model (Ohta and Kimura 1973) at a rate u per locus and generation. Accordingly, we define the scaled mutation rate in the ancestral deme as $\theta_{\text{anc}} = 4N_e u$. Mutation rates may vary among microsatellites for several reasons (Estoup and Cornuet 1999). To account for this, we use a hierarchical model (*cf.* Bazin et al. 2010), assuming that θ_{anc} is normally distributed across loci on the $\log_{10}$-scale, with mean $\mu_{\theta_{\text{anc}}}$ and standard deviation $\sigma_{\theta_{\text{anc}}}$. In our case, $\mu_{\theta_{\text{anc}}}$ and $\sigma_{\theta_{\text{anc}}}$ are the hyperparameters (Gelman et al. 2004) of interest. Here, we make the implicit assumption that N_e is the same for all loci, so that variance in θ_{anc} may be attributed to u exclusively. In principle, however, variation in diversity across loci could also be due to selection at linked genes (Maynard Smith and Haigh 1974; Charlesworth et al. 1993; Barton 2000), rather than variable mutation rates. Most likely, we

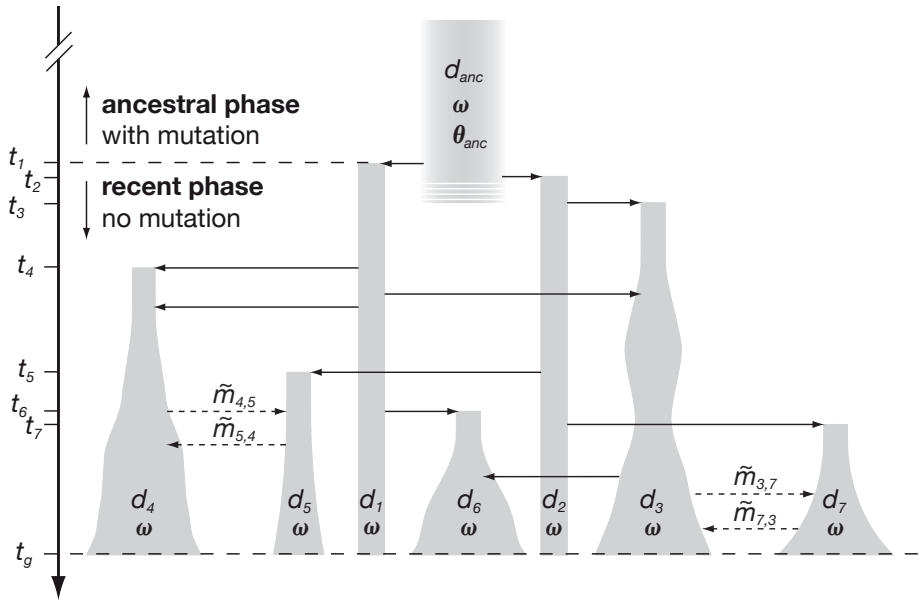

Figure 3.2: Schematic representation of the demographic model motivated by the re-introduction of Alpine ibex into the Swiss Alps. Gray shapes represent demes, indexed by d_i, and the width of the shapes reflects the census size. Time goes forward from top to bottom, and the point in time when deme d_i is established is shown as t_i; t_g is the time of genetic sampling. The total time is split by t_1 into an ancestral phase with mutation and a recent phase for which mutation is ignored (see text for details). Solid horizontal arrows represent founder/admixture events and dashed arrows migration. The parameters are i) the scaled mutation rate in the ancestral deme, $\theta_{\text{anc}} = 4N_e u$; ii) the proportion of males getting access to matings, ω; and iii) forward migration rates between putatively connected demes, $\tilde{m}_{i,j}$ (see text for details). The actual model considered in the study contains 35 derived demes (Figure 3.1 and Table 3.5). The exact demography is reported in Figure 3.7 and Supporting File 3.7 *transfers*.

cannot distinguish these alternatives with our data. The second, recent phase started at time t_1 and went up to the time of genetic sampling, $t_g = 2006$. During this phase, the number of males and females transferred at founder/admixture events and census population sizes are known and accounted for. Mutation is neglected in the recent phase, since, in the case of ibex, the phase spans only about eleven generations at most (Stuwe and Grodinsky 1987). At the transition from the ancestral to the recent phase, genotypes of the founder individuals introduced to demes d_1 and d_2 are sampled at random from the ancestral deme, d_{anc}. At the end of the recent phase (t_g), genetic samples are taken according to the sampling scheme under which the real data were obtained. Out of the total 35 demes, 31 were sampled (Table 3.5).

In Alpine ibex, male reproductive success is highly skewed towards dominant males. Dominance is correlated with age (Willisch et al. 2011), and ranks are established during summer. Only a small proportion of males obtain access to matings during the rut in winter (Aeschbacher 1978; Stuwe and Grodinsky 1987; Scribner and Stuwe 1994; Willisch and Neuhaus 2009; Willisch et al. 2011). To take this into account, we introduce the proportion of males obtaining access to matings, ω, as a parameter. It is defined relative to the number of potentially reproducing males (and therefore conditional on age; see SI), and has an impact on the strength of genetic drift. We assume that ω is the same in all demes and independent of deme size.

In principle, we would like to infer the joint posterior distribution $\pi(\boldsymbol{\alpha}, \tilde{\boldsymbol{m}} \mid D)$, where $\boldsymbol{\alpha} = (\mu_{\theta_{\text{anc}}}, \sigma_{\theta_{\text{anc}}}, \omega)$ and $\tilde{\boldsymbol{m}} = \{\tilde{m}_{i,j} : i \neq j,\, i \in \mathcal{J}_m,\, j \in \mathcal{J}_m\}$, with $\mathcal{J}_m$ denoting the set of all demes connected via migration to at least one other deme (Figure 3.1). This is a complex problem, mainly because there are so many parameters and even more candidate summary statistics; the curse of dimensionality is severe. Targeting the joint posterior with ABC naïvely would give a result, but it would be hard to assess its validity. It is more promising to address intermediate steps and assess them one by one. A first step is to focus on a subset of parameters and marginalize over the others. By marginalizing we formally mean that the joint posterior distribution is integrated with respect to the parameters that are not of interest. In this case, we integrate over the prior of the migration rates given in Table 3.1. In practice, marginal posteriors may be targeted directly with ABC (see below). A second step is to clarify what summary statistics should be chosen for the subset of focal parameters. A third one is to deal with the curse of dimensionality related to estimating $\tilde{\boldsymbol{m}}$. In this paper, we deal with steps one and two: We aim at estimating $\boldsymbol{\alpha}$ marginally to $\tilde{\boldsymbol{m}}$, and we seek a good method for choosing summary statistics with respect to $\boldsymbol{\alpha}$. The third step – estimating $\tilde{\boldsymbol{m}}$ – is treated in chapter 4. Notice that this division implies the assumption that priors of the migration rates and male mating success are independent. We make this assumption partly for convenience, and partly because we are not aware of any study that has shown a relation between the two in Alpine ibex.

3.3 Methods

The joint posterior distribution of our model may be factored as

$$\pi(\tilde{\boldsymbol{m}}, \boldsymbol{\alpha} \mid D) = \pi(\tilde{\boldsymbol{m}} \mid \boldsymbol{\alpha}, D)\, \pi(\boldsymbol{\alpha} \mid D). \tag{3.3}$$

As mentioned, here we only target the marginal posterior of $\boldsymbol{\alpha}$ on the right hand side. Formally, this is obtained as

$$\pi(\boldsymbol{\alpha} \mid D) = \int_{\mathcal{M}} \pi(\tilde{\boldsymbol{m}}, \boldsymbol{\alpha} \mid D)\, \mathrm{d}\tilde{\boldsymbol{m}}, \tag{3.4}$$

where $\mathcal{M}$ is the domain of possible values for $\tilde{\boldsymbol{m}}$. By the nature of our problem, $\pi(\tilde{\boldsymbol{m}}, \boldsymbol{\alpha} \mid D)$ is not available. However, with ABC we may target (3.4) directly by sampling from $\pi\big(\boldsymbol{\alpha} \mid \mathbf{s}_{\boldsymbol{\alpha}} = \mathbf{S}_{\boldsymbol{\alpha}}(D)\big)$, where we assume that $\mathbf{S}_{\boldsymbol{\alpha}}$ is a subset of summary statistics approximately sufficient for estimating $\boldsymbol{\alpha}$. Notice that $\mathbf{S}_{\boldsymbol{\alpha}}$ may not be sufficient to estimate the joint posterior (3.3), however (Raiffa and Schlaifer 1968). The following standard ABC algorithm provides an approximation to $\pi\big(\boldsymbol{\alpha} \mid \mathbf{s}_{\boldsymbol{\alpha}}\big)$ (*e.g.* Marjoram et al. 2003):

Algorithm A:

A.1 Calculate summary statistics $\mathbf{s}_{\boldsymbol{\alpha}} = \mathbf{S}_{\boldsymbol{\alpha}}(D)$ from observed data.

A.2 For $t = 1$ to $t = N$:

- i Sample $(\boldsymbol{\alpha}'_t, \tilde{\boldsymbol{m}}'_t)$ from $\pi(\boldsymbol{\alpha}, \tilde{\boldsymbol{m}}) = \pi(\boldsymbol{\alpha})\, \pi(\tilde{\boldsymbol{m}})$.
- ii Simulate data D'_t (at all loci and for all demes) from $P(D \mid \boldsymbol{\alpha}'_t, \tilde{\boldsymbol{m}}'_t)$.
- iii Calculate $\mathbf{s}'_{\boldsymbol{\alpha},t} = \mathbf{S}_{\boldsymbol{\alpha}}(D'_t)$ from simulated data.

A.3 Scale $\mathbf{s}_{\boldsymbol{\alpha}}$ and $\mathbf{s}'_{\boldsymbol{\alpha}}$ appropriately.

A.4 For each t, accept $\boldsymbol{\alpha}'_t$ if $\rho(\mathbf{s}'_{\boldsymbol{\alpha},t}, \mathbf{s}_{\boldsymbol{\alpha}}) \leq \delta_\epsilon$, using scaled summary statistics from A.3.

A.5 Estimate the posterior density $\pi(\boldsymbol{\alpha} \mid \mathbf{s}_{\boldsymbol{\alpha}})$ from the ϵN accepted points $\langle \mathbf{s}'_{\boldsymbol{\alpha},t}, \boldsymbol{\alpha}'_t \rangle$.

Step A.2 may be easily parallelized on a cluster computer. In doing so, one needs to store $\langle \mathbf{s}'_{\boldsymbol{\alpha},t}, \boldsymbol{\alpha}'_t \rangle$. Step A.5 may include post-rejection adjustment via regression (Beaumont et al. 2002; Blum and François 2010; Leuenberger and Wegmann 2010) and scaling of parameters. In general, the set of summary statistics $\mathbf{S}_{\boldsymbol{\alpha}}$ is not known in advance. Therefore, we propose algorithm B – a modified version of algorithm A – that includes an additional step for the empirical choice of summary statistics $\mathbf{S}_{\boldsymbol{\alpha}}$ informative on $\boldsymbol{\alpha}$ given a set of candidate statistics, $\mathbf{S}$ (for similar approaches, see Hamilton et al. 2005; Wegmann et al. 2009a):

Algorithm B:

B.1 Calculate candidate summary statistics $\mathbf{s} = \mathbf{S}(D)$ from observed data.

B.2 For $t = 1$ to $t = N$:

- i Sample $(\boldsymbol{\alpha}'_t, \tilde{\boldsymbol{m}}'_t)$ from $\pi(\boldsymbol{\alpha}, \tilde{\boldsymbol{m}}) = \pi(\boldsymbol{\alpha})\,\pi(\tilde{\boldsymbol{m}})$.
- ii Simulate data D'_t (at all loci and for all demes) from $P(D \mid \boldsymbol{\alpha}'_t, \tilde{\boldsymbol{m}}'_t)$.
- iii Calculate candidate summary statistics $\mathbf{s}'_t = \mathbf{S}(D'_t)$ from simulated data.

B.3 Sample without replacement $n \leq N$ simulated pairs $\langle \mathbf{s}'_t, \boldsymbol{\alpha}'_t \rangle$ and use them as a training data set to choose informative statistics $\mathbf{S}_{\boldsymbol{\alpha}}$.

B.4 According to B.3, obtain $\mathbf{s}_{\boldsymbol{\alpha}}$ from $\mathbf{s}$; for $t = 1$ to $t = N$, obtain $\mathbf{s}'_{\boldsymbol{\alpha},t}$ from $\mathbf{s}'_t$.

B.5 Scale $\mathbf{s}_{\boldsymbol{\alpha}}$ and $\mathbf{s}'_{\boldsymbol{\alpha}}$ appropriately.

B.6 For each t, accept $\tilde{\boldsymbol{\alpha}}'_t$ if $\rho(\mathbf{s}'_{\boldsymbol{\alpha},t}, \mathbf{s}_{\boldsymbol{\alpha}}) \leq \delta_\epsilon$, using scaled summary statistics from B.5.

B.7 Estimate the posterior density $\pi(\boldsymbol{\alpha} \mid \mathbf{s}_{\boldsymbol{\alpha}})$ from the ϵN accepted points $\langle \mathbf{s}'_{\boldsymbol{\alpha},t}, \boldsymbol{\alpha}'_t \rangle$.

Notice that $\mathbf{S}_{\boldsymbol{\alpha}}$ in steps B.3 and B.4 may either be a subset of $\mathbf{S}$ or some function (*e.g.* a linear combination) of $\mathbf{S}$ (details of implementation given below). In the following, we describe a novel approach based on boosting and recently proposed by Lin et al. (2011) for the choice of $\mathbf{S}_{\boldsymbol{\alpha}}$ in B.3.

3.3.1 Choice of summary statistics via boosting

Boosting is a collective term for meta-algorithms originally developed for supervised learning in classification problems (Schapire 1990; Freund 1995). Later, versions for regression (Friedman et al. 2000) and other contexts have been developed (Bühlmann and Hothorn 2007, and references therein). Assume a set of n observations indexed by i and associated with a one-dimensional response Y_i. For (binary) classification, $Y_i \in \{0, 1\}$, but in a regression context, Y_i may be continuous in $\mathbb{R}$. Further, each observation is associated with a vector of q predictors $\mathbf{X}_i = (X_i^{(1)}, \ldots, X_i^{(q)})$. Given a training data set $\{\langle \mathbf{X}_1, Y_1 \rangle, \ldots, \langle \mathbf{X}_n, Y_n \rangle\}$, the task of a boosting algorithm is to learn a function $F(\mathbf{X})$ that predicts Y. Boosting was invented to deal with cases where the relationship between predictors and response is potentially complex, for

example non-linear (Schapire 1990; Freund 1995; Freund and Schapire 1996, 1999). Establishing the relationship between predictors and response, and weighting predictors according to their importance, directly relates to the problem of choosing summary statistics in ABC: Given candidate statistics $\mathbf{S}$, we want to find a subset or combination of statistics $\mathbf{S}_{\boldsymbol{\alpha}^{(i)}}$ informative for a particular $\boldsymbol{\alpha}^{(i)}$. Taking the set of simulated pairs $\langle \mathbf{s}'_t,\, f(\boldsymbol{\alpha}'_t)\rangle$ from step B.3 of algorithm B as a training data set, this may be achieved by boosting. For this purpose, we interpret the summary statistics $\mathbf{S}$ as predictors $\mathbf{X}$ and the parameters $\boldsymbol{\alpha}$ as the response Y. Notice that we use $f(\boldsymbol{\alpha}'_t)$ to be generic in the sense that the response might actually be a function – such as a discretisation step (see below) – of $\boldsymbol{\alpha}'_t$.

The principle of boosting is to iteratively apply a *weak learner* to the training data, and then combine the ensemble of weak learners to construct a *strong learner*. While the weak learner predicts only slightly better than random guessing, the strong learner will usually be well correlated with the true Y. This is because the training data are re-weighted after each step according to the current error, such that the next weak learner will focus on those observations that were particularly hard to assign. However, too strong a correlation will lead to overfitting, so that in practice one defines an upper limit for the number of iterations (see below). The behavior of the weak learner is described by the base procedure $\hat{g}(\cdot)$, a real valued function. The final result (strong learner) is the desired function estimate $\hat{F}(\cdot)$. Given a loss function $L(\cdot,\cdot)$ that quantifies the disagreement between Y and $F(\mathbf{X})$, we want to estimate the function that minimizes the expected loss,

$$F^*(\cdot) = \arg\min_{F(\cdot)} \mathbb{E}\Big[L\big(Y, F(\mathbf{X})\big)\Big]. \tag{3.5}$$

This can be done by considering the empirical risk $n^{-1}\sum_{i=1}^{n} L\big(Y_i, F(\mathbf{X}_i)\big)$ and pursuing iterative steepest descent in function space (Friedman 2001; Bühlmann and Hothorn 2007). The corresponding algorithm is given in the Appendix. The generic boosting estimator obtained from this algorithm is a sum of base procedure estimates

$$\hat{F}(\cdot) = \nu \sum_{m=1}^{m_{\text{stop}}} \hat{g}^{[m]}(\cdot). \tag{3.6}$$

Both ν and m_{stop} are tuning parameters that essentially control the overfitting behavior of the algorithm. Bühlmann and Hothorn (2007) argue that the learning rate ν is of minor importance as long as $\nu \leq 0.1$. The number of iterations, m_{stop}, however, should be chosen specifically in any application via cross-validation, bootstrapping or some information criterion (*e.g.* AIC).

Base procedure

Different versions of boosting are obtained depending on the base procedure $\hat{g}(\cdot)$ and the loss function $L(\cdot,\cdot)$. Here, we let $\hat{g}(\cdot)$ be a simple component-wise linear regression (Bühlmann and Hothorn 2007, see Appendix). With this choice, the boosting algorithm selects in every iteration only one predictor, namely the one that is most useful in reducing the current loss. After each step, $\hat{F}(\cdot)$ is updated linearly according to

$$\hat{F}^{[m]}(\mathbf{x}) = \hat{F}^{[m-1]}(\mathbf{x}) + \nu\hat{\lambda}^{(\hat{\zeta}_m)}\mathbf{x}^{(\hat{\zeta}_m)}, \tag{3.7}$$

where $\hat{\zeta}_m$ denotes the index of the predictor variable selected in iteration m. Accordingly, in iteration m only the $\hat{\zeta}^{\text{th}}$ component of the coefficient estimate $\hat{\lambda}^{[m]}$ is updated. As m goes to infinity, $\hat{F}(\cdot)$ converges to a least squares solution. In practice, we stop at m_{stop}, and we denote the final vector of estimated coefficients as $\hat{\lambda} = \hat{\lambda}^{[m_{\text{stop}}]}$.

Loss functions

We employed boosting with three loss functions. The first two, L_1-loss and L_2-loss, are appropriate for a regression context with a continuous response $Y \in \mathbb{R}$. In this case, the parameters $\boldsymbol{\alpha}'_t$ are directly interpreted as y_i (*i.e.* $f(\boldsymbol{\alpha}'_t) = \boldsymbol{\alpha}'_t$). The L_1-loss is given by

$$L_{L_1}(y, F) = |y - F|\,, \tag{3.8}$$

and results in L_1Boosting. The L_2-loss is given by

$$L_{L_2}(y, F) = \frac{1}{2}\,|y - F|^2\,, \tag{3.9}$$

and results in L_2Boosting. The scaling factor $1/2$ in (3.9) ensures that the negative gradient vector U in the FGD algorithm (Appendix and SI) equals the residuals (Bühlmann and Hothorn 2007). L_1- and L_2Boosting result in a fit of a linear regression, similarly to ordinary regression using the least absolute deviation (L_1 norm) or the least squares criterion (L_2 norm), respectively. The difference, and a potential advantage of boosting, is that residuals are fitted multiple times depending on the importance of the components of $\mathbf{X}$. Moreover, boosting is less prone to overfitting than ordinary L_1 or L_2 fitting (Bühlmann and Hothorn 2007). In general, the L_1-loss is more robust to outliers, but it may produce multiple, potentially unstable solutions. Using L_1- and L_2Boosting to choose summary statistics means assuming a linear relationship between summary statistics and parameters. This is a strong assumption, and most likely not globally true. However, the advantage is that the resulting linear combination (and hence $\mathbf{S}_{\boldsymbol{\alpha}}$) has only one dimension, such that the curse of dimensionality in ABC may be strongly reduced. Moreover, the approach results in one linear combination per parameter. These linear combinations may end up being correlated across parameters, especially if parameters cannot be well separated. To motivate the third loss function, we propose to consider the choice of summary statistics as a classification problem. Imagine two classes of parameter values – say, high values in one class, and low values in the other. We may ask what summary statistics are important to assign simulations to one of these two classes. With $Y \in \{0, 1\}$ as the class label and $p(\mathbf{x}) := \Pr[Y = 1 \mid \mathbf{X} = \mathbf{x}]$, a natural choice is the negative binomial log-likelihood loss

$$L_{\text{log-lik}}(y, p) = -\big[y\,\log(p) + (1 - y)\,\log(1 - p)\big], \tag{3.10}$$

omitting the argument of p for ease of notation. If we parametrize $p = e^F/(1 + e^F)$ so that we obtain $F = \log\,[p/(1 - p)]$ corresponding to the logit-transformation, the loss in (3.10) becomes

$$L_{\text{log-lik}}(y, F) = \log\,\big[1 + e^{-(2y-1)F}\big]. \tag{3.11}$$

The corresponding boosting algorithm is called LogitBoost (or Binomial Boosting; Bühlmann and Hothorn 2007). An advantage is that it does not assume a linear relationship between summary statistics and parameters, as is the case for the L_1- and L_2Boosting versions used

here. Instead, LogitBoost fits a logistic regression model, which might be more appropriate. On the other hand, it requires choosing a discretization procedure $f(\cdot)$ to map $\boldsymbol{\alpha}_t \in \mathbb{R}$ to $y \in \{0, 1\}$ (see below). Since such a choice is arbitrary, it would be problematic to use the resulting fit (a linear combination on the logit-scale) directly as $\mathbf{S}_{\boldsymbol{\alpha}^{(i)}}$. In practice, we instead assigned a candidate statistic $\mathbf{S}^{(j)}$ $(j = 1, \ldots, q)$ to $\mathbf{S}_{\boldsymbol{\alpha}^{(i)}}$ if the corresponding boosted coefficient $\hat{\lambda}^{(j)}$ (*cf.* equation (3.7)) was different from zero, and omitted it otherwise. Therefore, compared L_1- and L_2Boosting, the reduction in dimensionality was on average lower, but the strong assumption of a linear $\boldsymbol{\alpha}^{(i)}$ and $\mathbf{S}_{\boldsymbol{\alpha}^{(i)}}$ was avoided. Notice that, in principle, non-linear relationships may be fitted with the L_1- and L_2-loss, too (Friedman et al. 2000). In the SI we provide explicit expressions for the population minimizers (3.5) and some more insight on the boosting algorithms under the three loss functions used here.

Partial Least Squares regression

Recently, Wegmann et al. (2009a) proposed to choose summary statistics in ABC via Partial Least Squares (PLS) regression (*e.g.* Hastie et al. 2011, and references therein). PLS is related to Principal Component regression. But in addition to maximizing the variance of the predictors $\mathbf{X}$, at the same time, it maximizes the correlation of $\mathbf{X}$ with the response $\mathbf{Y}$. Applied to the choice of summary statistics, it therefore not only decorrelates the summary statistics, but also chooses them according to their relation to $\boldsymbol{\alpha}$. Hastie et al. (2011) argue that the first aspect dominates over the latter, however. The number k of PLS components to keep is usually determined based on some cross-validation procedure (see below). In the context of ABC, the k components are multiplied by the corresponding statistics $\mathbf{S}^{(j)}$ $(j \leq k)$ to obtain $\mathbf{S}_{\boldsymbol{\alpha}^{(i)}}$ (Wegmann et al. 2009a).

3.3.2 Global versus local choice

We have so far suggested that $\mathbf{S}_{\boldsymbol{\alpha}}$ is close to sufficient for estimating $\boldsymbol{\alpha}$. This will hardly be the case in practice. By definition, the optimal choice of $\mathbf{S}_{\boldsymbol{\alpha}}$ then depends on the unknown true parameter value. Ideally, we would therefore like to focus the choice of $\mathbf{S}_{\boldsymbol{\alpha}}$ on the neighborhood of the truth. The latter is not known in practice. As a workaround, we propose to use the n simulated pairs $\langle \mathbf{s}'_t, \boldsymbol{\alpha}'_t \rangle$ from step B.3 in algorithm B and the observed summary statistics $\mathbf{s}$ to approximately establish this neighborhood as follows:

Local choice of summary statistics in B.3:

1. Consider the n pairs $\langle \mathbf{s}'_{t^*}, \boldsymbol{\alpha}'_{t^*} \rangle$ from step B.3 in algorithm B.
2. Mean center each component $\mathbf{s}'^{(j)}$ $(j = 1, \ldots, q)$ and scale it to have unit variance.
3. Rotate $\mathbf{s}'$ using Principal Component Analysis (PCA).
4. Apply the scaling from steps 2 and 3 to the observed summary statistics $\mathbf{s}$.
5. Mean center the PCA-scaled summary statistics obtained in step 3, and scale them to have unit variance. Do the same for the PCA-scaled observed statistics obtained in step 4. Denote the results by $\dot{\mathbf{s}}'$ and $\dot{\mathbf{s}}$, respectively.
6. For each $t^* \in n$, compute the Euclidean distance $\delta_{t^*} = \|\dot{\mathbf{s}}'_{t^*} - \dot{\mathbf{s}}_{t^*}\|$.

7. Keep the n' pairs $\langle \mathbf{s}'_{t^*}, \boldsymbol{\alpha}'_{t^*} \rangle$ for which $\delta_{t^*} \leq z$, where z is some threshold.
8. Use the n' points accepted in step 7 as a training set to choose statistics $\mathbf{S}_{\boldsymbol{\alpha}}$ with the desired method.
9. Continue with step B.4 in algorithm B.

In step 2 above, the original summary statistics are brought to the same scale. Otherwise, summary statistics with a high variance would on average contribute relatively more to the Euclidean distance than summary statistics with a low variance. However, whether a simulated data point is far or close to the target ($\mathbf{s}$) in multidimensional space may not only depend on the distance along the dimension of each statistic, but also on the correlation among statistics. This can be accounted for by decorrelating the statistics, as is done by PCA in step 3. In combination with the Euclidean distance in step 6, the procedure above essentially uses the Mahalanobis distance as metric (Mahalanobis 1936). Although we cannot prove the optimality of this approach, it seems to work well in our simulations. Notice that in steps 8 and 9, the summary statistics are used on their original scale again. This is because we want our method for choosing parameter-specific combinations of statistics to use the information comprised in the difference in scale among the original statistics – even in the vicinity of $\mathbf{s}$. The PCA-scaling in step 5 is only used temporarily to determine δ_{t^*} in step 6. Figure 3.8 visualizes the different scales and the effect of determining an approximate neighborhood around $\mathbf{s}$.

The scheme just described may be combined with any of the methods for choosing summary statistics described above. In our case, we considered ABC with global and local versions of PLS (called `pls.glob` and `pls.loc` in the following), LogitBoosting (`lgb.glob`, `lgb.loc`), L_1-Boosting (`l1b.glob`, `l1b.loc`), and L_2Boosting (`l2b.glob`, `l2b.loc`). Moreover, we performed ABC with all candidate statistics $\mathbf{S}$ (`all`) as a reference.

Candidate summary statistics

Our set $\mathbf{S}$ of candidate summary statistics consisted of the mean and standard devation across loci of the following statistics: the average within-deme variance of allele length, the average within-deme gene diversity (H_1), the average between-deme gene diversity (H_2), the total F_{IS}, the total F_{ST}, the total within-deme mean squared difference (MSD) in allele length (S_1), the total between-deme MSD in allele length (S_2), the total R_{ST}, and the number of allele types in the total population. This amounts to a total of 18 summary statistics. We computed H_1, H_2, F_{IS} and F_{ST} according to Nei and Chesser (1983), and S_1, S_2 and R_{ST} according to Slatkin (1995). Notice that all summary statistics are symmetrical with respect to the order of the loci, which is consistent with our hierarchical parametrization of the ancestral mutation rate.

Implementation

Throughout, we used the prior distributions given in Table 3.1. In algorithm B, we performed $N = 10^6$ simulations and in B.2i we assumed that $\pi(\boldsymbol{\alpha}, \tilde{\boldsymbol{m}}) = \pi(\boldsymbol{\alpha})\,\pi(\tilde{\boldsymbol{m}})$. In B.3, we used $n = 10^4$ simulations for the choice of summary statistics (both in the global and local versions). Moreover, we first chose sets of summary statistics for each parameter separately, and then took the union of the sets, *i.e.* $\mathbf{S}_{\boldsymbol{\alpha}} = \bigcup_i \mathbf{S}_{\boldsymbol{\alpha}^{(i)}}$, where each $\mathbf{S}_{\boldsymbol{\alpha}^{(i)}}$ is chosen according to one

Table 3.1: Parameters and prior distributions

Parameter	Description	Prior distribution
$\theta_{\text{anc},\,l}$	Scaled ancestral mutation rate at locus l, $4N_e u$	$\log_{10}(\theta_{\text{anc},l}) \sim N\left(\mu_{\theta_{\text{anc}}}, \sigma^2_{\theta_{\text{anc}}}\right)^a$
$\mu_{\theta_{\text{anc}}}$	Mean across loci of $\theta_{\text{anc},l}$ (on $\log_{10}$-scale)	$\mu_{\theta_{\text{anc}}} \sim N\,(0.5,\, 1)$
$\sigma_{\theta_{\text{anc}}}$	Standard deviation across loci of $\theta_{\text{anc},l}$ (on $\log_{10}$-scale)	$\sigma_{\theta_{\text{anc}}} \sim \log_{10}$-uniform $[0.01,\, 1]$
ω	Proportion of mature males with access to matings	$\omega \sim \log_{10}$-uniform $[0.01,\, 1]$
${\tilde{m}_{i,j}}^a$	Forward migration rate per year from deme i to deme j	$\tilde{m}_{i,j} \sim \log_{10}$-uniform $\left[10^{-3.5},\, 10^{-0.5}\right]$

[a] $N\left(\mu, \sigma^2\right)$, normal distribution with mean μ and variance σ^2.
[b] Although migration rates are not estimated here, they are drawn from the prior in all simulations (see main text).

of the methods proposed. This also applies to step 8 in the procedure for the local choice of summary statistics (see above). For the local choice, we kept the $n' = 1000$ pairs closest to the observation `s`, and we used the `pcrcomp` function in `R` version 2.11 (R Development Core Team 2011) for PCA. In B.5, we mean-centered the summary statistics and scaled them to have unit variance. In B.6, we chose the Euclidean distance as metric $\rho(\cdot)$. In B.7 we did post-rejection adjustment with a weighted local-linear regression with weights from an Epanechnikov kernel (Beaumont et al. 2002), without additional scaling of parameters. For steps B.6 and B.7 we used the `abc` package (Csilléry et al. 2011) for `R`.

For the PLS method, we used the `pls` package (Mevik and Wehrens 2007) for `R` and followed Wegmann et al. (2009a) and Wegmann et al. (2010). Specifically, we performed a Box-Cox transformation of the summary statistics prior to the PLS regression, and we chose the number of components to keep based on a plot of the root mean squared prediction error. We kept 10 components, both for `pls.glob` and `pls.loc` (Figure 3.9). For all methods based on boosting, we mean-centered the summary statistics before boosting and used the `glmboost` function of the `mboost` package (Bühlmann and Hothorn 2007; Hothorn et al. 2011) for `R`. For the LogitBoost methods, we chose the first and third quartile of the sample of $\boldsymbol{\alpha}$ drawn in step B.3 of algorithm B.3 as the centers of the two classes of parameter values. For `lgb.glob`, we then assigned the 500 $\boldsymbol{\alpha}$-values closest to the first quartile to the first class ($y = 0$) and the 500 values closest to the third quartile to the second class ($y = 1$). For `lgb.loc`, we analogously assigned the 100 $\boldsymbol{\alpha}$-values closest to the two quartiles to the two classes. For both `lgb.glob` and `lgb.loc`, we chose the optimal m_{stop} based on the Akaike information criterion (AIC, Akaike 1974; Bühlmann and Hothorn 2007), but set an upper limit for m_{stop} of 500 iterations. For `l1b.glob` and `l1b.loc`, we chose m_{stop} via 10-fold cross-validation with the `cvrisk` function of the `mboost` package, setting an upper limit of 100. Last, for `l2b.glob` and `l2b.loc`, we chose m_{stop} based on the AIC, with an upper limit of 100. Figures 3.10 to 3.12 further illustrate the booting procedure.

3.3.3 Simulation study and application to data

To assess the performance of the different methods for choosing summary statistics and to study the influence of the rejection tolerance ϵ, we carried out a simulation study. For each $\epsilon \in \{0.001, 0.01, 0.1\}$, we simulated 500 test data sets with parameter values sampled from the prior distributions and then inferred the posterior distribution for each set. Similar to Wegmann et al.

(2009a), we used as a measure of accuracy of the marginal posterior distributions the root mean integrated squared error (RMISE), defined as $\text{RMISE}_k = \sqrt{\int_{\Phi^{(k)}} (\phi^{(k)} - \mu_k)^2 \, \pi(\phi^{(k)} \mid \mathbf{s}) \, \mathrm{d}\phi^{(k)}}$, where μ_k is the true value of the k^{th} component of the parameter vector $\boldsymbol{\phi}$ and $\pi(\phi^{(k)} \mid \mathbf{s})$ is the corresponding estimated marginal posterior density. Recall that $\boldsymbol{\phi} = \boldsymbol{\alpha} = (\mu_{\theta_{\text{anc}}}, \sigma_{\theta_{\text{anc}}}, \omega)$ in our case. From this, we obtained the relative absolute RMISE (RARMISE) as $\text{RARMISE}_k = \text{RMISE}_k / |\mu_k|$. We also computed the absolute difference (AE_k) between three marginal posterior point estimates (mode, mean and median) and μ_k. Dividing by $|\mu_k|$, we obtained the relative absolute error (RAE_k). To directly compare the various methods to ABC with all summary statistics, we computed *standardized* variants of the RMISE and AE as follows: If $a_k^{\texttt{all}}$ is the measure of accuracy for ABC with all summary statistics, and a_k^* the one for ABC with the method of interest, the standardized measure was obtained as $a_k^* / a_k^{\texttt{all}}$. As a further criterion, we assessed the coverage property of the inferred posterior distributions. For this, we checked if the posterior probabilities of the true parameter values across the 500 test data sets were uniformly distributed in $[0, 1]$. This approach has been motivated by Cook et al. (2006) and applied in previous ABC studies (*e.g.* Wegmann et al. 2009a). However, notice that Cook et al. (2006) called these posterior probabilities 'posterior quantiles', which is somewhat misleading. We tested for a uniform distribution of the posterior probabilities using a Kolmogorov-Smirnov test (Sokal and Rohlf 1981).

For the application to Alpine ibex, we used microsatellite allele frequencies and repeat lengths as described in Biebach and Keller (2009) (*see* Figure 3.1 and Table 3.5). The data were provided to us by the authors. ABC simulations and inference were identical to those in the simulation study (see also SI). The program called `SPoCS` that we wrote and used for simulation of the ibex scenario, and a collection of `R` and shell scripts used for inference are available on the website `http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`.

3.4 Results and discussion

3.4.1 Comparison of methods for choice of summary statistics

We have suggested boosting with component-wise linear regression as a base procedure for choosing summary statistics in ABC. Three loss functions were considered: the L_1-, and L_2-loss, and the negative binomial log-likelihood. We have compared the performance of ABC with summary statistics chosen via boosting to ABC with statistics chosen via partial least squares (PLS, Wegmann et al. 2009a), and to ABC with all candidate summary statistics (Table 3.2). The relative absolute error (RAE) behaved similarly for the three point estimates (mode, mean, median), but the mode was less reliable in cases were the posterior distributions did not have a unique mode (Figure 3.13). We decided to focus on the median. For assessment of the methods, we sought a low RARMISE and a low RAE of the median ($\text{RAE}_{\text{median}}$ in the following), and we required that the distribution of posterior probabilities of the true value did not deviate from uniformity for any parameter.

ABC with all summary statistics (`all`) and ABC with LogitBoosting (`lgb.glob`) performed well in terms of RARMISE and $\text{RAE}_{\text{median}}$, especially when estimating $\mu_{\theta_{\text{anc}}}$ and ω (Figure 3.3A and 3.3B). However, the posteriors of $\mu_{\theta_{\text{anc}}}$ inferred with `all` and `lgb.glob` were biased

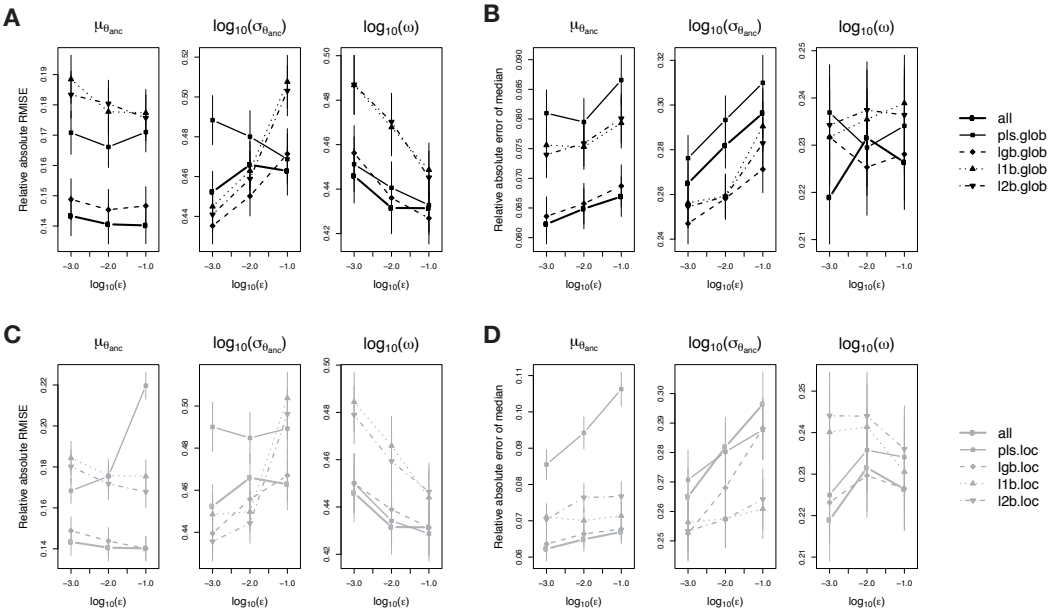

Figure 3.3: Accuracy of different methods for choosing summary statistics as a function of the acceptance rate (ϵ). (A) and (B) show results for different methods when applied to the whole parameter range (*global* choice). In (C) and (D), the methods were applied only in the neighborhood of the (supposed) true value (*local* choice). The performance resulting from using all candidate summary statistics is shown for comparison in both rows. (A) and (C) show the root mean integrated squared error (RMISE), relative to the absolute true value. (B) and (D) give the absolute error of the posterior median, relative to the absolute true value. Plotted are the medians across $n = 500$ independent test estimations with true values drawn from the prior (error bars denote the $\text{median} \pm \text{MAD}/\sqrt{n}$, where MAD is the median absolute deviation).

(coverage p-value in Table 3.2). Figure 3.14 implies that `all` yielded too narrow a posterior on average (U-shaped distribution of posterior probabilities of the true value), while `lgb.glob` tended to underestimate $\mu_{\theta_{\text{anc}}}$ (left-skewed distribution of posterior probabilities). This made us disfavor the two methods. Throughout, ABC with L_1- and L_2Boosting on the global scale (`l1b.glob` and `l2b.glob`) performed very similarly in terms of RARMISE and $\text{RAE}_{\text{median}}$ (Figure 3.3A and 3.3B). Because the L_2-loss is in general more sensitive to outliers, similarity in performance of `l1b.glob` and `l2b.glob` suggests that there were no problems with outliers, *i.e.* no simulations producing extreme combinations of parameters and summary statistics. The accuracy of the `pls.glob` method was intermediate, except for the $\text{RAE}_{\text{median}}$ of $\mu_{\theta_{\text{anc}}}$ and ω, where `pls.glob` performed worst (Figure 3.3B). For all methods, the RARMISE and the $\text{RAE}_{\text{median}}$ were considerably lower for $\mu_{\theta_{\text{anc}}}$ than for $\sigma_{\theta_{\text{anc}}}$ and ω. This implies that the latter two are more difficult to estimate with the data and model given here (see Figure 3.13). For an idea of how the data drive the parameter estimates, it is instructive to consider the correlation of individual summary statistics with the parameters (see Figures 3.17 to 3.19).

The accuracy of estimation is expected to depend on the acceptance rate ϵ in a way determined by a trade-off between bias and variance (*e.g.* Beaumont et al. 2002). While the RAE only measures the error of the point estimator, the RARMISE is a joint measure of bias and variance across the whole posterior distribution. The variance may be assigned to different sources. A first component – call it *simulation variance* – is a consequence of the finite number

Table 3.2: Accuracy of different methods for choosing summary statistics on a global scale

Method	ϵ	Parameter	RARMISE[a]	RAE[b] mode	RAE mean	RAE median	Cov. p[c]
all	0.001	$\mu_{\theta_{\text{anc}}}$	0.143 (0.147)	0.062 (0.074)	0.065 (0.075)	0.062 (0.075)	0.011*
		$\sigma_{\theta_{\text{anc}}}$	0.452 (0.231)	0.269 (0.213)	0.269 (0.222)	0.265 (0.218)	0.61
		ω	0.446 (0.272)	0.221 (0.225)	0.215 (0.218)	0.219 (0.22)	0.859
	0.01	$\mu_{\theta_{\text{anc}}}$	0.141 (0.145)	0.061 (0.072)	0.064 (0.074)	0.065 (0.075)	0.003*
		$\sigma_{\theta_{\text{anc}}}$	0.466 (0.257)	0.299 (0.21)	0.286 (0.225)	0.282 (0.226)	0.992
		ω	0.432 (0.259)	0.233 (0.232)	0.226 (0.23)	0.232 (0.232)	0.88
	0.1	$\mu_{\theta_{\text{anc}}}$	0.140 (0.134)	0.065 (0.075)	0.067 (0.078)	0.067 (0.075)	0.003*
		$\sigma_{\theta_{\text{anc}}}$	0.463 (0.272)	0.324 (0.238)	0.306 (0.248)	0.296 (0.243)	0.677
		ω	0.431 (0.263)	0.234 (0.229)	0.228 (0.22)	0.226 (0.223)	0.482
pls.glob	0.001	$\mu_{\theta_{\text{anc}}}$	0.171 (0.16)	0.077 (0.087)	0.083 (0.089)	0.081 (0.088)	0.466
		$\sigma_{\theta_{\text{anc}}}$	0.488 (0.276)	0.291 (0.223)	0.289 (0.252)	0.276 (0.228)	0.936
		ω	0.451 (0.275)	0.238 (0.221)	0.234 (0.224)	0.237 (0.227)	0.969
	0.01	$\mu_{\theta_{\text{anc}}}$	0.166 (0.152)	0.080 (0.09)	0.079 (0.09)	0.079 (0.089)	0.562
		$\sigma_{\theta_{\text{anc}}}$	0.480 (0.291)	0.307 (0.223)	0.295 (0.268)	0.293 (0.242)	0.473
		ω	0.441 (0.262)	0.241 (0.234)	0.230 (0.225)	0.229 (0.226)	0.562
	0.1	$\mu_{\theta_{\text{anc}}}$	0.171 (0.146)	0.083 (0.091)	0.086 (0.097)	0.087 (0.094)	0.497
		$\sigma_{\theta_{\text{anc}}}$	0.469 (0.283)	0.319 (0.237)	0.307 (0.286)	0.310 (0.276)	0.089
		ω	0.433 (0.265)	0.240 (0.226)	0.234 (0.224)	0.234 (0.23)	0.178
lgb.glob	0.001	$\mu_{\theta_{\text{anc}}}$	0.149 (0.152)	0.064 (0.074)	0.065 (0.076)	0.064 (0.074)	0.002*
		$\sigma_{\theta_{\text{anc}}}$	0.435 (0.204)	0.270 (0.231)	0.261 (0.214)	0.247 (0.205)	0.466
		ω	0.456 (0.275)	0.235 (0.23)	0.230 (0.237)	0.232 (0.224)	0.913
	0.01	$\mu_{\theta_{\text{anc}}}$	0.145 (0.15)	0.066 (0.076)	0.066 (0.078)	0.066 (0.076)	<0.001*
		$\sigma_{\theta_{\text{anc}}}$	0.450 (0.223)	0.281 (0.215)	0.269 (0.217)	0.258 (0.209)	0.238
		ω	0.436 (0.27)	0.235 (0.234)	0.222 (0.223)	0.225 (0.228)	0.916
	0.1	$\mu_{\theta_{\text{anc}}}$	0.147 (0.142)	0.068 (0.079)	0.067 (0.078)	0.069 (0.079)	<0.001*
		$\sigma_{\theta_{\text{anc}}}$	0.471 (0.284)	0.288 (0.209)	0.301 (0.249)	0.271 (0.233)	0.103
		ω	0.427 (0.259)	0.232 (0.222)	0.225 (0.216)	0.228 (0.22)	0.329
l1b.glob	0.001	$\mu_{\theta_{\text{anc}}}$	0.188 (0.178)	0.075 (0.087)	0.074 (0.087)	0.076 (0.088)	0.573
		$\sigma_{\theta_{\text{anc}}}$	0.445 (0.202)	0.271 (0.236)	0.261 (0.232)	0.256 (0.216)	0.954
		ω	0.487 (0.297)	0.251 (0.259)	0.226 (0.227)	0.232 (0.226)	0.723
	0.01	$\mu_{\theta_{\text{anc}}}$	0.178 (0.17)	0.075 (0.087)	0.075 (0.088)	0.075 (0.085)	0.711
		$\sigma_{\theta_{\text{anc}}}$	0.463 (0.217)	0.288 (0.24)	0.271 (0.238)	0.259 (0.221)	0.805
		ω	0.468 (0.288)	0.255 (0.262)	0.228 (0.222)	0.235 (0.233)	0.595
	0.1	$\mu_{\theta_{\text{anc}}}$	0.177 (0.173)	0.078 (0.092)	0.078 (0.094)	0.079 (0.094)	0.311
		$\sigma_{\theta_{\text{anc}}}$	0.508 (0.299)	0.307 (0.21)	0.304 (0.269)	0.290 (0.248)	0.144
		ω	0.449 (0.272)	0.238 (0.241)	0.237 (0.222)	0.239 (0.227)	0.716
l2b.glob	0.001	$\mu_{\theta_{\text{anc}}}$	0.183 (0.173)	0.075 (0.087)	0.074 (0.085)	0.074 (0.086)	0.794
		$\sigma_{\theta_{\text{anc}}}$	0.441 (0.202)	0.273 (0.229)	0.257 (0.228)	0.254 (0.212)	0.828
		ω	0.487 (0.296)	0.251 (0.257)	0.231 (0.226)	0.234 (0.229)	0.648
	0.01	$\mu_{\theta_{\text{anc}}}$	0.180 (0.173)	0.077 (0.087)	0.077 (0.088)	0.076 (0.087)	0.766
		$\sigma_{\theta_{\text{anc}}}$	0.459 (0.213)	0.278 (0.242)	0.262 (0.235)	0.259 (0.214)	0.815
		ω	0.470 (0.288)	0.253 (0.26)	0.231 (0.221)	0.237 (0.229)	0.497
	0.1	$\mu_{\theta_{\text{anc}}}$	0.176 (0.171)	0.080 (0.092)	0.080 (0.096)	0.080 (0.093)	0.365
		$\sigma_{\theta_{\text{anc}}}$	0.503 (0.281)	0.300 (0.213)	0.297 (0.249)	0.283 (0.253)	0.139
		ω	0.445 (0.267)	0.240 (0.24)	0.239 (0.227)	0.236 (0.225)	0.755

RARMISE and RAE (see below) are given as the median across 500 independent estimations with true values drawn from the prior (median absolute deviation in parentheses). $\sigma_{\theta_{\text{anc}}}$ and ω were estimated on the $\log_{10}$ scale.

[a] Relative absolute root mean integrated squared error (see text) with respect to the true value.

[b] Relative absolute error with respect to the true value.

[c] P-value from a Kolmogorov-Smirnov test for the uniformity of posterior probabilities (∗: $p < 0.05$; *cf.* Figure 3.14).

N of simulations. The lower ϵ, the fewer points are accepted in the rejection step (B.6 of algorithm B, see above). Posterior densities estimated from fewer points will be less stable than those inferred from more points. A second variance component – the *sampling variance* – is

due to the loss of information caused by using non-sufficient summary statistics. To illustrate the trade-off between simulation and sampling variance, assume ϵ fixed. If a large number of summary statistics is chosen, these may extract most of the information and thus limit the sampling variance. However, more summary statistics means more dimensions, and therefore a lower chance of accepting the same number of simulations than with fewer summary statistics, hence a higher simulation variance. In addition, accepting with $\delta_\epsilon > 0$ – which is characteristic of ABC – will introduce a systematic bias if the multi-dimensional density is not symmetric on the chosen metric with respect to the observation $\mathbf{s}$. On the other hand, increasing δ_ϵ reduces the simulation variance. Hence, there are in fact multiple trade-offs. It is not obvious in advance which one will dominate, and it is hard to make a prediction. This is reflected in our results: We found no uniform pattern for the dependence on ϵ of the RARMISE and the $\text{RAE}_{\text{median}}$. For instance, with `l2b.glob` the RARMISE increased as a function of ϵ for $\sigma_{\theta_{\text{anc}}}$, but decreased for ω (Figure 3.3A). Moreover, and typically for a trade-off, the relationship between accuracy and ϵ need not be monotonic (Figure 3.3; *cf.* Beaumont et al. 2002).

Attempting to mitigate the lack of sufficiency, we have proposed to choose summary statistics locally – in the putative neighborhood of the true parameter values – rather than globally over the whole prior range. As expected, the local choice led to different combinations of statistics, and it had an effect on the scaling of the statistics for `pls.loc`, `l1b.loc` and `l2b.loc` (Figure 3.20). However, the local versions of the different methods performed similarly to their global counterparts in terms of RARMISE and $\text{RAE}_{\text{median}}$ (Table 3.3 and Figure 3.3). Only with `pls.loc` the estimation error for $\mu_{\theta_{\text{anc}}}$ increased more strongly with ϵ than for `pls.glob`. More importantly, however, the coverage properties of the posteriors for $\mu_{\theta_{\text{anc}}}$ deteriorated for `pls.loc`, `l1b.loc` and `l2b.loc` (Table 3.3), compared to their global versions (Table 3.2). The effect was weakest for `l2b.loc`, and in general increased as a function of ϵ. Method `pls.loc` tended to overestimate $\mu_{\theta_{\text{anc}}}$, while `lgb.loc`, `l1b.loc` and `l2b.loc` tended to underestimate it (Figure 3.15).

For direct comparison of methods, before averaging across test sets, we standardized the measures of accuracy relative to those obtained with all summary statistics (Figure 3.4). The only local method that, for all parameters, led to lower RARMISE and $\text{RAE}_{\text{median}}$ than its global version was `l2b.loc`. In contrast, `lgb.glob` and `lgb.loc` performed very similarly; `pls.loc` did worse than `pls.glob` for $\mu_{\theta_{\text{anc}}}$, but better than `pls.glob` for $\sigma_{\theta_{\text{anc}}}$ and ω. Overall, we chose `l2b.loc` with $\epsilon = 0.01$ as our favored method. This configuration provided good coverage for all parameters (Table 3.3). At the same time, it had lower RARMISE and $\text{RAE}_{\text{median}}$ than `pls.glob`, the method that would also have had good coverage properties for $\mu_{\theta_{\text{anc}}}$. We disfavored `all`, `lgb.glob` and `lgb.loc` due to their weak coverage properties. Notice that all methods compared in Figure 3.4 performed worse in terms of RARMISE and $\text{RAE}_{\text{median}}$ than `all` when estimating $\mu_{\theta_{\text{anc}}}$. This might be due to the loss of information caused by leaving out some summary statistics. Apparently, this loss is not fully compensated in our setting by the potential gain from reducing the dimensions. In models with many more dimensions, this may be different.

It is worth recalling some of the characteristics of the methods compared here. The `pls` method is the only one that involves de-correlation of the statistics. Apparently, this did not lead to a net improvement compared to the other methods. One explanation is that

Table 3.3: Accuracy of different methods for choosing summary statistics on a local scale

Method	ϵ	Parameter	RARMISE	RAE mode	RAE mean	RAE median	Cov. p
pls.loc	0.001	$\mu_{\theta_{\text{anc}}}$	0.168 (0.136)	0.081 (0.091)	0.088 (0.095)	0.086 (0.091)	0.314
		$\sigma_{\theta_{\text{anc}}}$	0.490 (0.262)	0.283 (0.229)	0.277 (0.234)	0.271 (0.226)	0.314
		ω	0.450 (0.278)	0.232 (0.234)	0.225 (0.228)	0.225 (0.228)	0.723
	0.01	$\mu_{\theta_{\text{anc}}}$	0.175 (0.126)	0.088 (0.094)	0.098 (0.103)	0.094 (0.099)	0.023*
		$\sigma_{\theta_{\text{anc}}}$	0.485 (0.274)	0.287 (0.222)	0.287 (0.243)	0.280 (0.223)	0.232
		ω	0.434 (0.259)	0.240 (0.238)	0.235 (0.224)	0.236 (0.227)	0.655
	0.1	$\mu_{\theta_{\text{anc}}}$	0.220 (0.147)	0.101 (0.103)	0.113 (0.108)	0.106 (0.104)	0.001*
		$\sigma_{\theta_{\text{anc}}}$	0.489 (0.282)	0.294 (0.216)	0.275 (0.243)	0.288 (0.231)	0.078
		ω	0.429 (0.259)	0.239 (0.226)	0.239 (0.227)	0.234 (0.223)	0.273
lgb.loc	0.001	$\mu_{\theta_{\text{anc}}}$	0.149 (0.151)	0.061 (0.074)	0.067 (0.081)	0.064 (0.077)	0.006*
		$\sigma_{\theta_{\text{anc}}}$	0.440 (0.213)	0.271 (0.213)	0.259 (0.209)	0.253 (0.209)	0.5
		ω	0.450 (0.283)	0.229 (0.231)	0.223 (0.219)	0.223 (0.217)	0.794
	0.01	$\mu_{\theta_{\text{anc}}}$	0.144 (0.147)	0.065 (0.074)	0.068 (0.078)	0.066 (0.077)	0.001*
		$\sigma_{\theta_{\text{anc}}}$	0.456 (0.237)	0.292 (0.209)	0.277 (0.223)	0.268 (0.213)	0.576
		ω	0.439 (0.27)	0.235 (0.229)	0.228 (0.225)	0.230 (0.225)	0.862
	0.1	$\mu_{\theta_{\text{anc}}}$	0.140 (0.133)	0.068 (0.077)	0.069 (0.078)	0.068 (0.078)	<0.001*
		$\sigma_{\theta_{\text{anc}}}$	0.467 (0.275)	0.315 (0.233)	0.298 (0.24)	0.288 (0.234)	0.991
		ω	0.431 (0.264)	0.232 (0.22)	0.226 (0.219)	0.227 (0.222)	0.423
l1b.loc	0.001	$\mu_{\theta_{\text{anc}}}$	0.184 (0.183)	0.070 (0.081)	0.070 (0.083)	0.071 (0.082)	0.062
		$\sigma_{\theta_{\text{anc}}}$	0.449 (0.215)	0.263 (0.234)	0.254 (0.219)	0.256 (0.218)	0.61
		ω	0.484 (0.281)	0.246 (0.253)	0.232 (0.218)	0.240 (0.233)	0.61
	0.01	$\mu_{\theta_{\text{anc}}}$	0.176 (0.18)	0.072 (0.081)	0.070 (0.083)	0.070 (0.082)	0.012*
		$\sigma_{\theta_{\text{anc}}}$	0.450 (0.218)	0.268 (0.25)	0.263 (0.23)	0.257 (0.221)	0.651
		ω	0.466 (0.279)	0.255 (0.265)	0.234 (0.22)	0.241 (0.234)	0.791
	0.1	$\mu_{\theta_{\text{anc}}}$	0.175 (0.181)	0.076 (0.092)	0.072 (0.084)	0.071 (0.085)	<0.001*
		$\sigma_{\theta_{\text{anc}}}$	0.504 (0.276)	0.277 (0.234)	0.291 (0.251)	0.261 (0.227)	0.257
		ω	0.444 (0.267)	0.238 (0.236)	0.237 (0.227)	0.231 (0.225)	0.694
l2b.loc	0.001	$\mu_{\theta_{\text{anc}}}$	0.180 (0.18)	0.071 (0.08)	0.074 (0.084)	0.070 (0.081)	0.314
		$\sigma_{\theta_{\text{anc}}}$	0.436 (0.207)	0.249 (0.222)	0.251 (0.215)	0.253 (0.213)	0.759
		ω	0.479 (0.275)	0.257 (0.261)	0.233 (0.226)	0.244 (0.235)	0.5
	0.01	$\mu_{\theta_{\text{anc}}}$	0.172 (0.173)	0.075 (0.085)	0.077 (0.087)	0.076 (0.087)	0.084
		$\sigma_{\theta_{\text{anc}}}$	0.444 (0.211)	0.258 (0.246)	0.264 (0.225)	0.257 (0.215)	0.651
		ω	0.459 (0.276)	0.256 (0.276)	0.234 (0.228)	0.244 (0.236)	0.532
	0.1	$\mu_{\theta_{\text{anc}}}$	0.168 (0.169)	0.077 (0.091)	0.076 (0.09)	0.077 (0.091)	<0.001*
		$\sigma_{\theta_{\text{anc}}}$	0.496 (0.266)	0.277 (0.235)	0.289 (0.241)	0.264 (0.23)	0.284
		ω	0.446 (0.271)	0.239 (0.242)	0.237 (0.23)	0.236 (0.233)	0.579

Details as in Table 3.2 (*cf.* Figure 3.15).

there was not much correlation to start with. Another one is that this correlation did not substantially reduce the efficiency. Figure 3.16 implies that the latter was the case. The reduction of dimensions is strongest with the `l1b` and `l2b` methods, since they result in one linear predictor per parameter. On the other hand, these methods assume a linear relationship between parameters and statistics. Since the latter was clearly not the case (*e.g.* Figure 3.17), it seems that the reduction of dimensions compensated for that assumption. This effect might be more pronounced in problems with many more statistics.

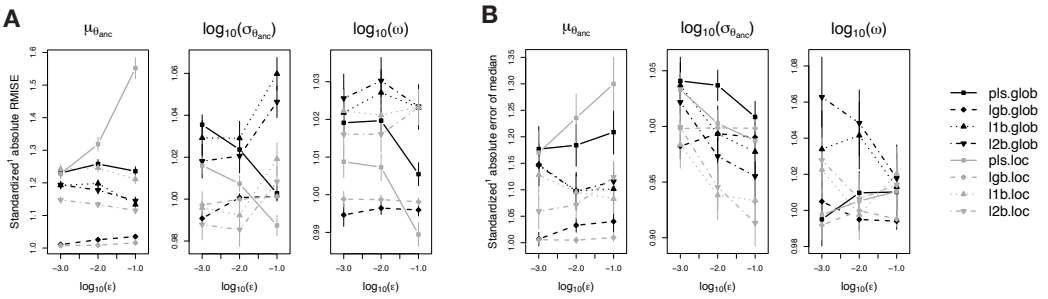

Figure 3.4: Standardized accuracy of different methods for choosing summary statistics as a function of the acceptance rate (ϵ). [1]Standaridized means that, before averaging across test sets, we divided the measures of accuracy for the respective method by the measure of accuracy obtained with all candidate summary statistics. (A) Root mean integrated squared error (RMISE), relative to the RMISE obtained with all summary statistics. (B) Absolute error of the posterior median, relative to the one obtained with all summary statistics. Further details as in Figure 3.3.

3.4.2 Application to Alpine ibex

Posterior distributions inferred for the ibex data with the various methods and $\epsilon = 0.01$ are shown in Figure 3.5. Point estimates and 95% highest posterior density (HPD) intervals for the method that performed best in the simulation study, `l2b.loc`, are given in Table 3.4. Recall that $\mu_{\theta_{\text{anc}}}$ and $\sigma_{\theta_{\text{anc}}}$ are hyperparameters of the distribution of $\theta_{\text{anc},l}$ across loci: $\log_{10}(\theta_{\text{anc},l}) \sim N\left(\mu_{\theta_{\text{anc}}}, \sigma^2_{\theta_{\text{anc}}}\right)$ (*cf.* Table 3.1). Inserting the estimates from Table 3.4, we obtained $\log_{10}(\theta_{\text{anc},l}) \sim N(0.110, 0.163^2)$, which implies a mean $\hat{\theta}_{\text{anc}}$ across loci of 1.288. The limits of the interval defined by $\hat{\mu}_{\theta_{\text{anc}}} \pm 2\hat{\sigma}_{\theta_{\text{anc}}}$ translate into (0.607, 2.735) on the scale of θ_{anc}. Recall that $\theta_{\text{anc}} = 4N_e u$; it measures the total genetic diversity present in the ancestral deme at time $t_1 = 1906$ (Figure 3.2), *i.e.* at the start of the reintroduction phase. Although we were able to estimate θ_{anc} with relatively high precision, that does not immediately tell us about N_e or u without knowing one of the two. However, given some rough, independent estimates of N_e and u, we may assess if our estimate $\hat{\theta}_{\text{anc}} \approx 1.288$ is plausible. On the one hand, historical records of the census size of the ancestral *Gran Paradiso* deme are available. In combination with an estimate of the ratio of effective to census size, we may therefore obtain a rough estimate of N_e. Specifically, the census size of the *Gran Paradiso* deme (Figure 3.1) was estimated as less than 100 for the early 19$^{\text{th}}$ century (Scribner and Stuwe 1994; Stuwe and Nievergelt 1991), as 3,000 for the early 20$^{\text{th}}$ century (Stuwe and Scribner 1989), and as 4,000 for the year 1913 (Maudet et al. 2002). In addition, Scribner and Stuwe (1994) estimated for eight ibex demes in the Swiss Alps the effective population size from census estimates of the numbers of adult males and females. Their estimates of N_e were about one third of the respective total census estimates. Together, these figures suggest that a realistic range for the ancestral effective size N_e might be between 30 and 1,300. On the other hand, estimates of the mutation rate u for microsatellites range from 10^{-4} to 10^{-2} per locus and generation (Di Rienzo et al. 1998; Estoup and Angers 1998). Combining these two ranges results in θ_{anc} ranging from $1.2 \cdot 10^{-2} \approx 10^{-2}$ to $5.2 \cdot 10 \approx 10^2$, suggesting that our estimate $\hat{\theta}_{\text{anc}} \approx 1.288$ is plausible. Perhaps more interestingly, we may ask about the range *across loci* of u that is compatible with the range of $\hat{\theta}_{\text{anc}}$ corresponding to $\hat{\mu}_{\theta_{\text{anc}}} \pm 2\hat{\sigma}_{\theta_{\text{anc}}}$, (0.607, 2.735). The underlying assumption is that N_e is roughly the same for all loci, so that variation in $\hat{\theta}_{\text{anc}}$ is exclusively due to variation of u across

loci. Taking the geometric mean of the extremes from above, $\hat{N}_e = (30 \cdot 1300)^{1/2} \approx 197$, as a typical value, the corresponding interval for $\hat{u}$ across loci is $(7.7 \cdot 10^{-4}, 3.5 \cdot 10^{-3})$. In other words, most of the variation in u across loci spans less than one order of magnitude.

Table 3.4: Posterior estimates for Alpine ibex data from ABC with summary statistics chosen locally via L_2Boosting and acceptance rate $\epsilon = 0.01$

Parameter	Mode	Mean	Median	95% HPD[a] interval
$\mu_{\theta_{\text{anc}}}$	0.1089	0.1081	0.1101	$(-0.0391, 0.2545)$
$\log_{10}(\sigma_{\theta_{\text{anc}}})$	-0.6453	-0.8928	-0.7867	$(-1.7615, -0.2613)$
$\log_{10}(\omega)$	-0.6159	-0.6933	-0.6824	$(-1.33, -0.0294)$

[a] Highest posterior density.

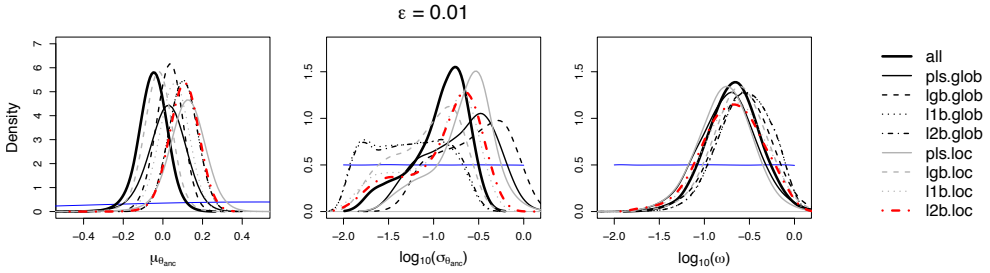

Figure 3.5: Marginal posterior distributions inferred from the Alpine ibex data. Posteriors obtained with tolerance $\epsilon = 0.01$ and various methods for choosing summary statistics are compared. The dotdashed red line corresponds to the method that performed best in the simulation study (`l2b.loc`; Tables 3.2 and 3.3, and Figures 3.3 and 3.4). Thin blue lines give the prior distribution (*cf.* Table 3.1). For pairwise joint posterior distributions, see Figure 3.6. Point estimates and 95% HPD intervals are given in Table 3.4.

The estimates for $\log_{10}(\omega)$ from Table 3.4 imply a proportion of males obtaining access to matings of $\hat{\omega} \approx 0.208$, or about 21%. The 95% HPD interval for ω is (0.047, 0.934). An observational study in a free-ranging ibex deme suggested that roughly 10% of males reproduced (Aeschbacher 1978). More recently, Willisch et al. (2011) conducted a behavioral and genetic study and reported paternity scores for males of different age classes. The weighted mean across age classes from this study is about 14% successful males. Given the many factors that influence such estimates, our result of 21% seems in good agreement with these values, and our 95% HPD interval includes them. Two points are worth noting. First, our 95% HPD interval for ω seems large, which reflects the uncertainty involved in this parameter. Second, when estimating ω, we are essentially estimating the ratio of *recent* effective population size to census population size, $N_e^{(i)}/N$, where $N_e^{(i)}$ is the effective size of a derived deme d_i. This ratio may be smaller than one for many reasons – not just male mating access. Thus, we have strictly speaking estimated the strength of genetic drift due to deviations in reproduction from that in an idealized population. Nevertheless, the good agreement with the independent estimates of male mating access is striking.

In Figure 3.6, we report pairwise joint posterior distributions for `l2b.loc` and $\epsilon = 0.01$. The pairwise joint modes are close to the marginal point estimates in Table 3.4. Moreover, Figure 3.6 suggests no strong correlation among parameters.

3.4.3 Conclusion

We have suggested three variants of boosting for the choice of summary statistics in ABC, and compared them to partial least squares (PLS) regression and to ABC with all candidate summary statistics. Moreover, we proposed to choose summary statistics locally, in the putative neighborhood of the observed data. Overall, the mean of the ancestral mutation rate $\mu_{\theta_{\text{anc}}}$ seemed easier to estimate than its standard deviation $\sigma_{\theta_{\text{anc}}}$ and the male mating access rate ω. In our context, ABC with summary statistics chosen locally via boosting with component-wise linear regression as base procedure and the L_2-loss performed best in terms of accuracy and posterior coverage. However, the difference between the methods was moderate. If the main interest had been in accurate point estimates, but less in good overall posterior properties at the same time, boosting with the negative binomial log-likelihood loss would have been preferable. The performance of the PLS method was intermediate when estimating $\sigma_{\theta_{\text{anc}}}$ and ω, but worst when estimating $\mu_{\theta_{\text{anc}}}$. In general, choosing summary statistics locally slightly improved the accuracy compared to the global choice, but it led to worse posterior coverage for $\mu_{\theta_{\text{anc}}}$. The local version of L_2Boosting with acceptance rate $\epsilon = 0.01$ coped best with this trade-off.

Applying that method to Alpine ibex data, we estimated the mean across loci of the scaled ancestral mutation rate as $\hat{\theta}_{\text{anc}} \approx 1.288$. The estimates for $\sigma_{\theta_{\text{anc}}}$ implied that most of the variation across loci of the mutation rate u was between $7.7 \cdot 10^{-4}$ and $3.5 \cdot 10^{-3}$. The proportion of males obtaining access to matings per breeding season was estimated to $\hat{\omega} \approx 0.21$, which is in good agreement with recent independent estimates. This result suggests that the strong dominance hierarchy in Alpine ibex is reflected in overall genetic diversity, and should therefore be considered an important factor determining the strength of genetic drift.

It should be noted that the results we reported here about the choice of summary statistics are specific to the model and the data. Another method may perform better under a different setting. We think that this is a general aspect of inference with ABC. For the various points where some choice must be made – summary statistics, metric, algorithm, post-rejection adjustment – by nature, no single strategy is best in every case. Rather, the focus should be on

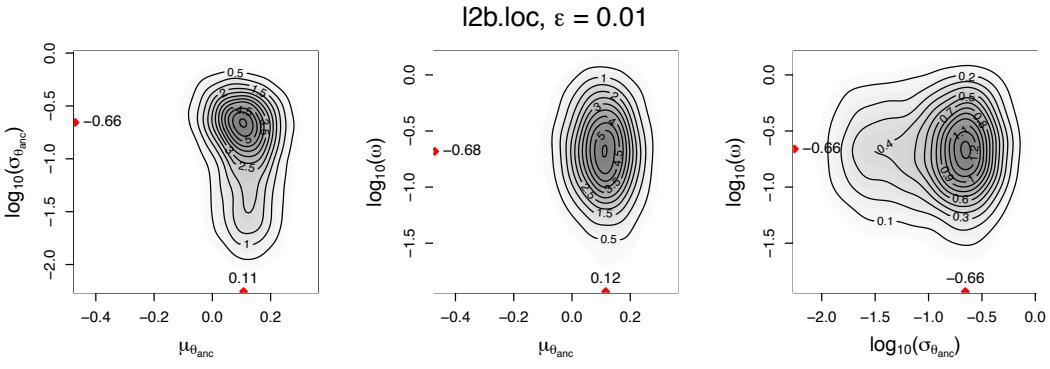

Figure 3.6: Pairwise joint posterior distributions given data observed in Alpine ibex, obtained with tolerance $\epsilon = 0.01$ and summary statistics choosen locally via L_2Boosting (`l2b.loc`). Red triangles denote parameter values corresponding to the pairwise joint modes. Each time, the third parameter has been marginalized over.

choosing the best strategy for a specific problem. In practice, this implies comparing alternatives and assessing performance in a simulation study. Along these lines, there is still scope for new ideas concerning the various choices in ABC (see Beaumont et al. 2010). In particular, the choice of the metric makes ABC a scale-dependent method. This applies both to the ABC algorithm in general, as well as to our suggestion of choosing summary statistics in the putative neighborhood of the truth. Even using the Mahalanobis distance is based on an assumption that is not necessarily appropriate (multivariate normal distribution of variables). In a specific application, a given metric may do better than another one, but it may not be obvious why. Overall, this poses an open problem and motivates future research (Wilkinson 2008).

As more data become available and more complex models are justifiable, it will be necessary that methods of inference keep pace. In principle, ABC is scalable and able to face this challenge. The problems arise in practice, and the combination of approaches devised to tackle them is itself becoming intricate. Researchers may be interested in a single program that implements these approaches and allows for inference with limited effort needed for tuning, simulation and cross-validation. However, such software runs the risk of being treated as a black box. This problem is not unique to ABC, but equally applies to other sophisticated approaches of inference, such as coalescent-based genealogy samplers (Kuhner 2009). In the context of ABC, rather than having a single piece of software, we find it more promising to combine separate pieces of software that each implement a specific step. The appropriate combination must be chosen specifically for any application. It will always be necessary to evaluate the statistical behaviour of any ABC method through simulation-based studies. Such a modular approach has recently been fostered by the developers of `ABCtoolbox` (Wegmann et al. 2010) or the `abc` package for `R` (Csilléry et al. 2011). Here, we contribute to this by providing a flexible simulation program that readily integrates into any ABC procedure.

At the same time as this study was carried out, Fearnhead and Prangle (2011) suggested an interesting related approach for choosing summary statistics in the vicinity of the supposed truth. They proved that, with the L_2-loss, the posterior mean is the optimal summary statistic. Since this is not available in advance, they proposed to first run a pilot ABC study to determine the region of high posterior mass. For this region, they then drew parameters and simulated data to obtain a training data sets. These were then used in a third step to fit a linear regression with the parameters as responses and a vector-valued function of the original summary statistics as explanatory variables. The linear fits were used as summary statistics for the corresponding parameter. A final ABC run was then performed, with a prior restricted to the range established in the first step, and summary statistics as chosen in the third step. Fearnhead and Prangle (2011) claim this to be *semi-automatic* and independent of the choice of statistics, but of course it does depend on the initial choice of candidate statistics and on the choice of the vector-valued function. Moreover, if the (transposed) candidate statistics are uncorrelated, we suspect that their method would be equivalent to using the first component in a univariate PLS regression. In any case, a direct comparison between all the recently proposed methods for the choice of statistics in ABC seems due.

3.5 Appendix

3.5.1 Functional gradient descent boosting algorithm

The general functional gradient descent (FGD) algorithm for boosting, as given by Friedman (2001) and modified by Bühlmann and Hothorn (2007), is:

FGD algorithm:

1. Initialize $\hat{F}^{[0]}(\cdot) \equiv \arg\min_c n^{-1} \sum_{i=1}^{n} L(Y_i, c)$,
   set $m = 0$.
2. Increase m by 1. Compute the negative gradient and evaluate at $\hat{F}^{[m-1]}(\mathbf{X}_i)$:
$$U_i = -\frac{\partial}{\partial F} L(Y_i, F)|_{F=\hat{F}^{[m-1]}(\mathbf{X}_i)}.$$
3. Fit the negative gradient vector $(U_1, \ldots, U_n)$ to $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ by the base procedure:
$$(\mathbf{X}_i, U_i)_{i=1}^{n} \longrightarrow \hat{g}^{[m]}.$$
4. Update $\hat{F}^{[m]}(\cdot) = \hat{F}^{[m-1]}(\cdot) + \nu \hat{g}^{[m]}(\cdot)$, where ν is a step-length factor.
5. Iterate steps 2 to 4 until $m = m_{\text{stop}}$.

Here, ν and m_{stop} are tuning parameters discussed in the main text. The result of this algorithm is a linear combination $\hat{F}(\cdot)$ of base procedure estimates, as shown in equation (3.6) of the main text. In any specific version of boosting, the form of the initial function $\hat{F}^{[0]}(\cdot)$ in step 1, and the negative gradient U_i in step 2 may be expressed explicitly according to the loss function $L(\cdot, \cdot)$ (see SI).

3.5.2 Base procedure: component-wise linear regression

We write the j^{th} component of a vector v as $v^{(j)}$. The following base procedure performs simple component-wise linear regression:

$$\begin{aligned} \hat{g}(\mathbf{X}) &= \hat{\lambda}^{(\hat{\zeta})} \mathbf{X}^{(\hat{\zeta})}, \\ \hat{\lambda}^{(j)} &= \sum_{i=1}^{n} \mathbf{X}^{(j)} U_i \Big/ \sum_{i=1}^{n} \left(\mathbf{X}_i^{(j)}\right)^2, \\ \hat{\zeta} &= \arg\min_{1 \le j \le p} \sum_{i=1}^{n} \left(U_i - \hat{\lambda}^{(j)} \mathbf{X}_i^{(j)}\right)^2, \end{aligned} \tag{3.12}$$

where $\hat{g}(\cdot)$, $\mathbf{X}$ and U_i are as in the FGD algorithm above. This base procedure selects the best variable in a simple linear model in the sense of ordinary least squares fitting (Bühlmann and Hothorn 2007). To see this, note that $\hat{\lambda}^{(j)}$ in (3.12) is the ordinary least squares solution of a linear regression $U_i = \mathbf{X}_i^{(j)} \lambda^{(j)}$, in matrix form $\hat{\lambda}^{(j)} = \left(\mathbf{X}_i^{(j)T} \mathbf{X}_i^{(j)}\right)^{-1} \mathbf{X}_i^{(j)T} U_i$. The choice of the loss functions enters indirectly via U_i (see SI).

3.6 Supporting information: Additional tables

Table 3.5: Deme names, deme numbers and sampling sizes in the Alpine ibex data set

				Genetic sample size[c]		
Deme name	Deme no.[a]	Short name	Internal number[b]	Males	Females	Total
Adula Vial	1	AdulaVial	100	21	16	37
Albris	2	Albris	101	28	33	61
Alpstein	3	Alpstein	102	12	18	30
Bire-Oeschinen	4	BireOesch	103	16	2	18
Brienzer Rothorn	5	BrRothorn	104	21	18	39
Calanda	6	Calanda	105	15	16	31
Churfirsten	7	Churfirsten	106	11	13	24
Crap da Flem	8	CrapFlem	107	16	11	27
Fluebrig	9	Fluebrig	108	17	15	32
Flüela	10	Flüela	109	37	38	75
Foostock	11	Foostock	110	9	18	27
Gastern	12	Gastern	111	5	6	11
Graue Hörner	13	GrHörner	112	21	26	47
Gross Lohner	14	GrLohner	113	15	7	22
Hochwang	15	Hochwang	114	14	14	28
Julier Nord	16	Julier N	115	12	11	23
Julier Süd	17	Julier S	116	12	11	23
Justistal	18	Justistal	117	15	4	19
Macun	19	Macun	118	12	10	22
Oberalp-Frisal	20	Oberalp	134	25	19	44
Oberbauenstock	21	Oberbauen	119	18	12	30
Pilatus	22	Pilatus	120	15	2	17
Mont Pleureur	23	Pleureur	121	22	7	29
Safien-Rheinwald	24	Rheinwald	122	22	13	35
Rothorn-Weissfluh	25	RothWeissfl	123	16	13	29
Schwarzmönch	26	SchwMönch	124	15	17	32
Umbrail	27	Umbrail	125	15	14	29
Val Bever	28	ValBever	126	20	12	32
Wetterhorn	29	Wetterhorn	127	9	10	19
Wittenberg	30	Wittenberg	128	15	6	21
Pierreuse-Gummfluh	31	Pierreuse	133	20	21	41
Wildpark Dählhölzli	32	WPDH	129	0	0	0
Wildpark Interlaken	33	WPIH	130	0	0	0
Wildpark St. Gallen	34	WPPP	131	0	0	0
Wildpark Seiler	35	WPSE	132	0	0	0

[a] As used in main text and Figure 3.1.
[b] As used in scripts and Supporting Files 3.6 and 3.7.
[c] The number of individuals from which genetic samples were taken, both in reality and in the simulations.

3.7 Supporting information: Additional figures

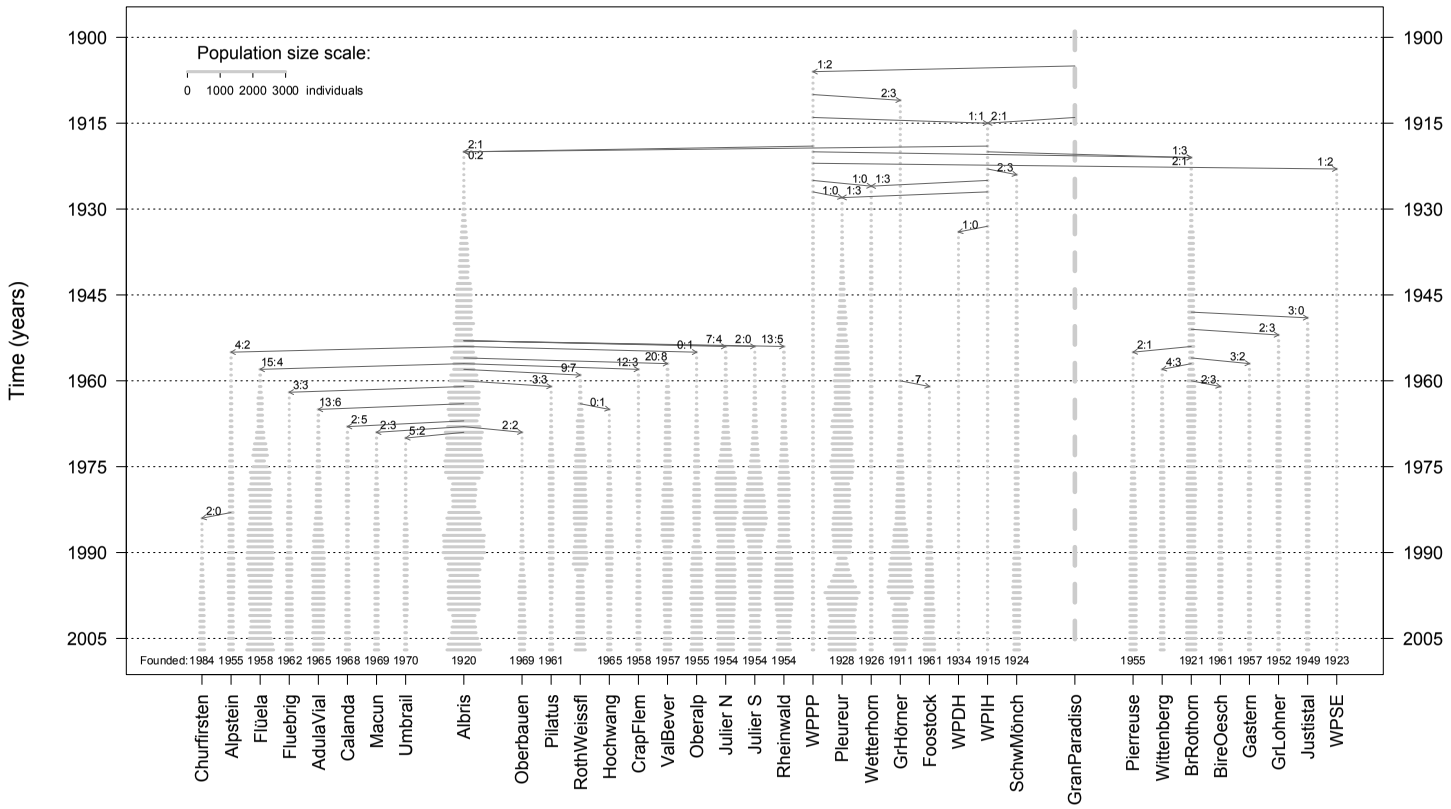

Figure 3.7: Genealogy and demography of Alpine ibex demes analyzed in this study. Time goes from top to bottom, starting in the year 1900 and ending in 2007. Horizontal gray bars represent the known census sizes (Supporting File 3.6 *census sizes*) and arrows show the founder events by which demes were established. The numbers of males and females transferred are given close to the arrow head (males:females; for *Foostock*, the sex of the founders is unknown and only the total number of founders is given). Most demes received further individuals after the initial founder event, but these numbers are not shown here (see Supporting File 3.7 *transfers*). The deme ancestral to all other demes, *GranParadiso*, is shown as a vertical dashed line; its deme size is not known. See also Table 3.5 for the full deme names and Figure 3.1 for the geographical location of demes.

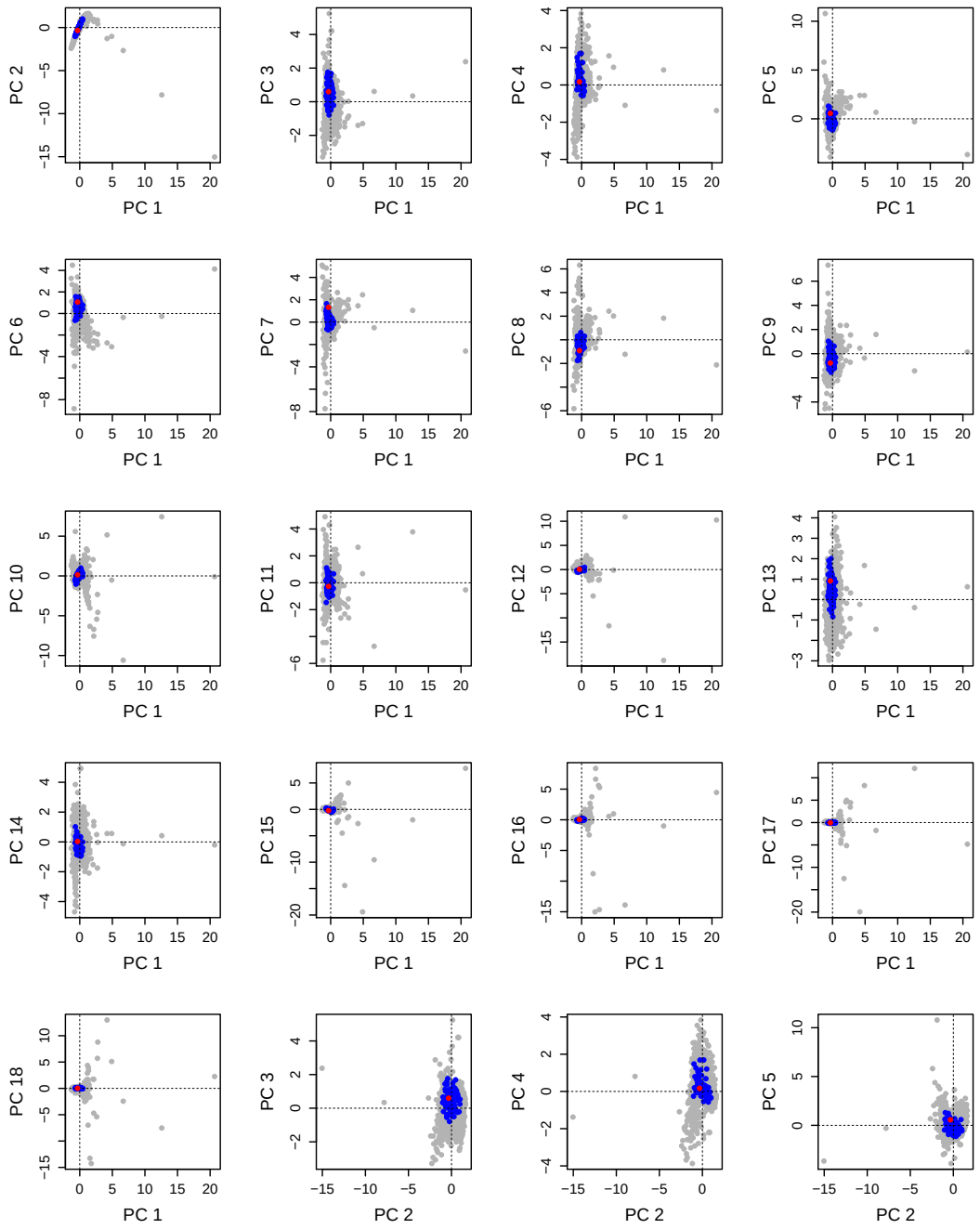

Figure 3.8: *Continued on next page*

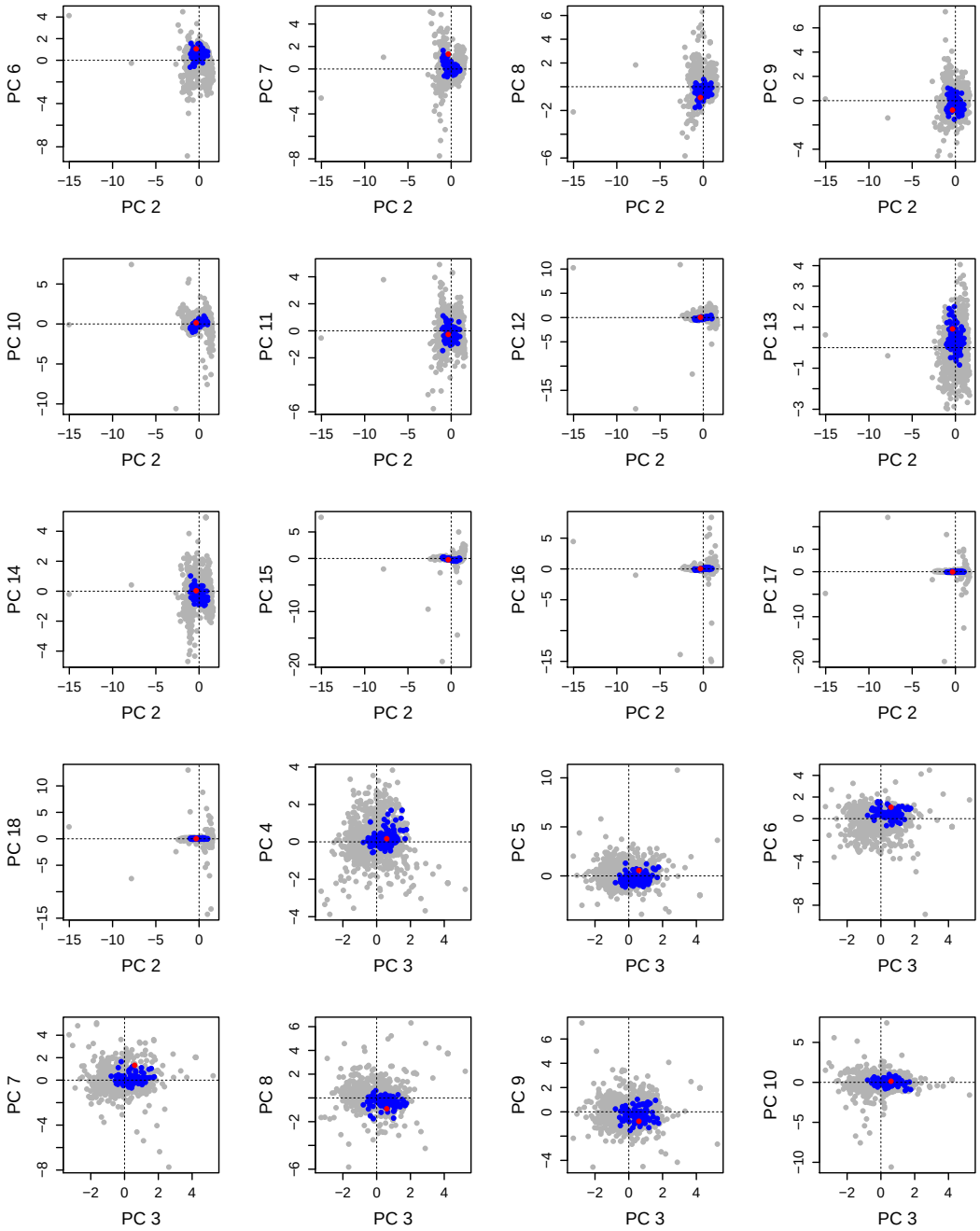

Figure 3.8: *Continued on next page*

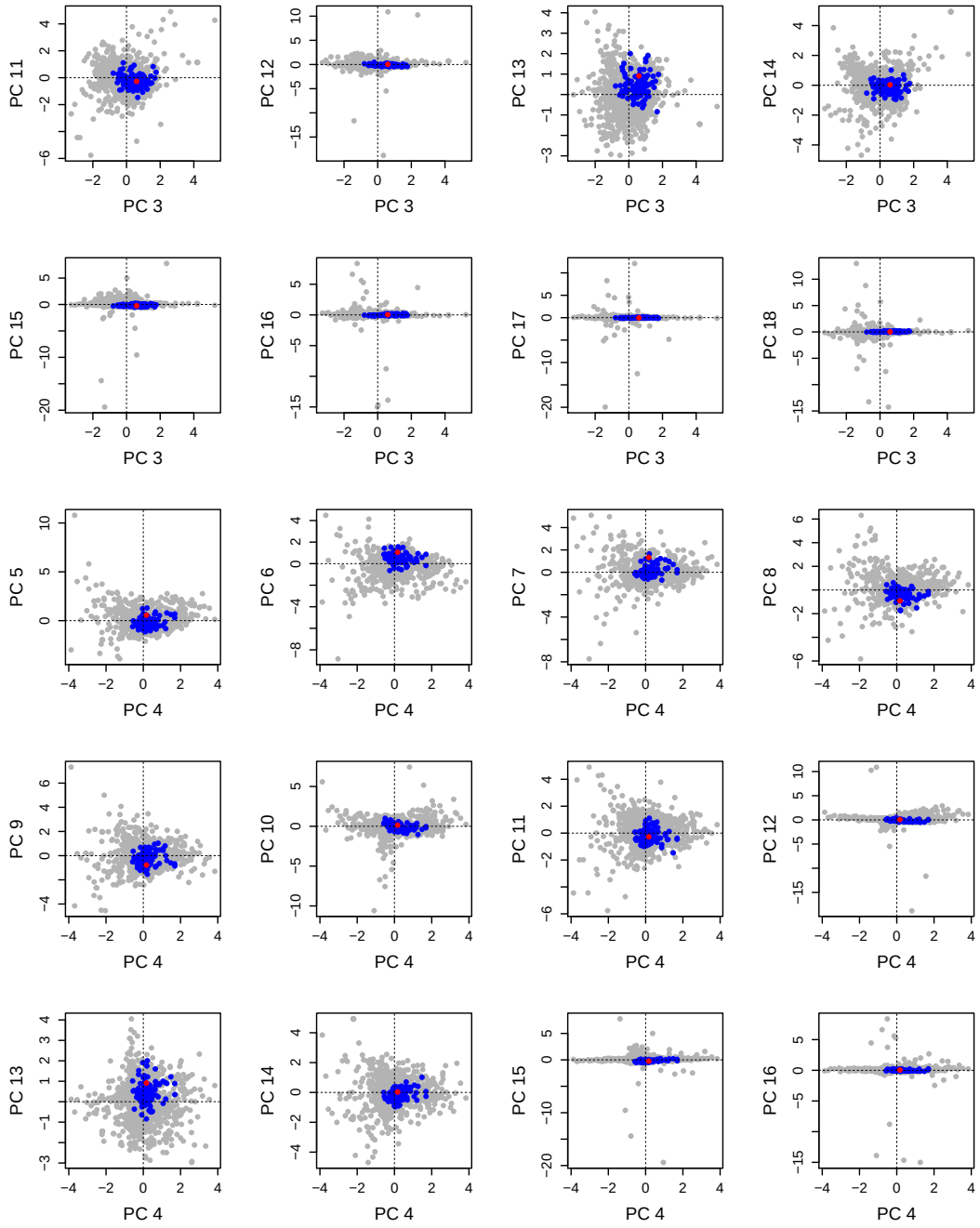

Figure 3.8: *Continued on next page*

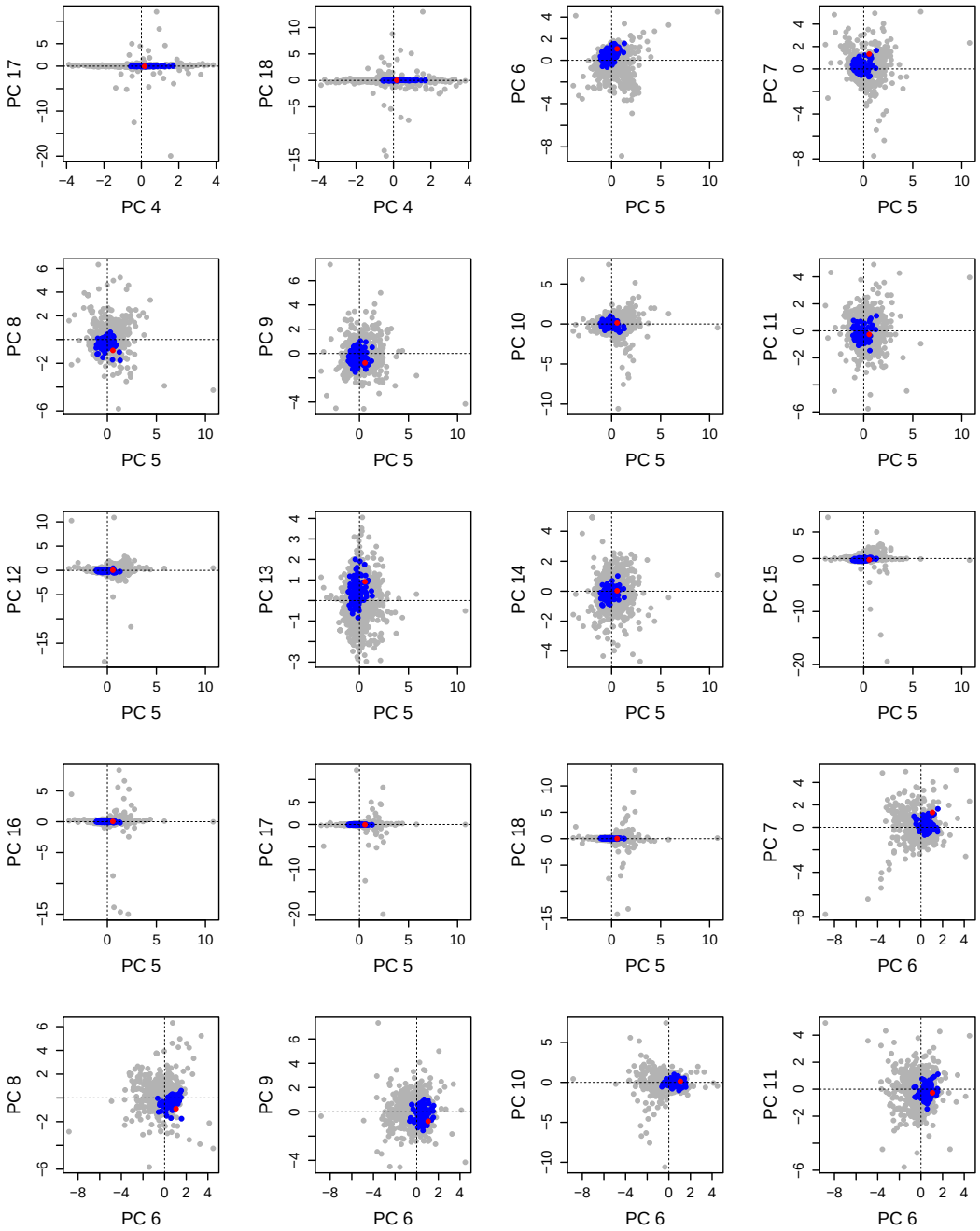

Figure 3.8: *Continued on next page*

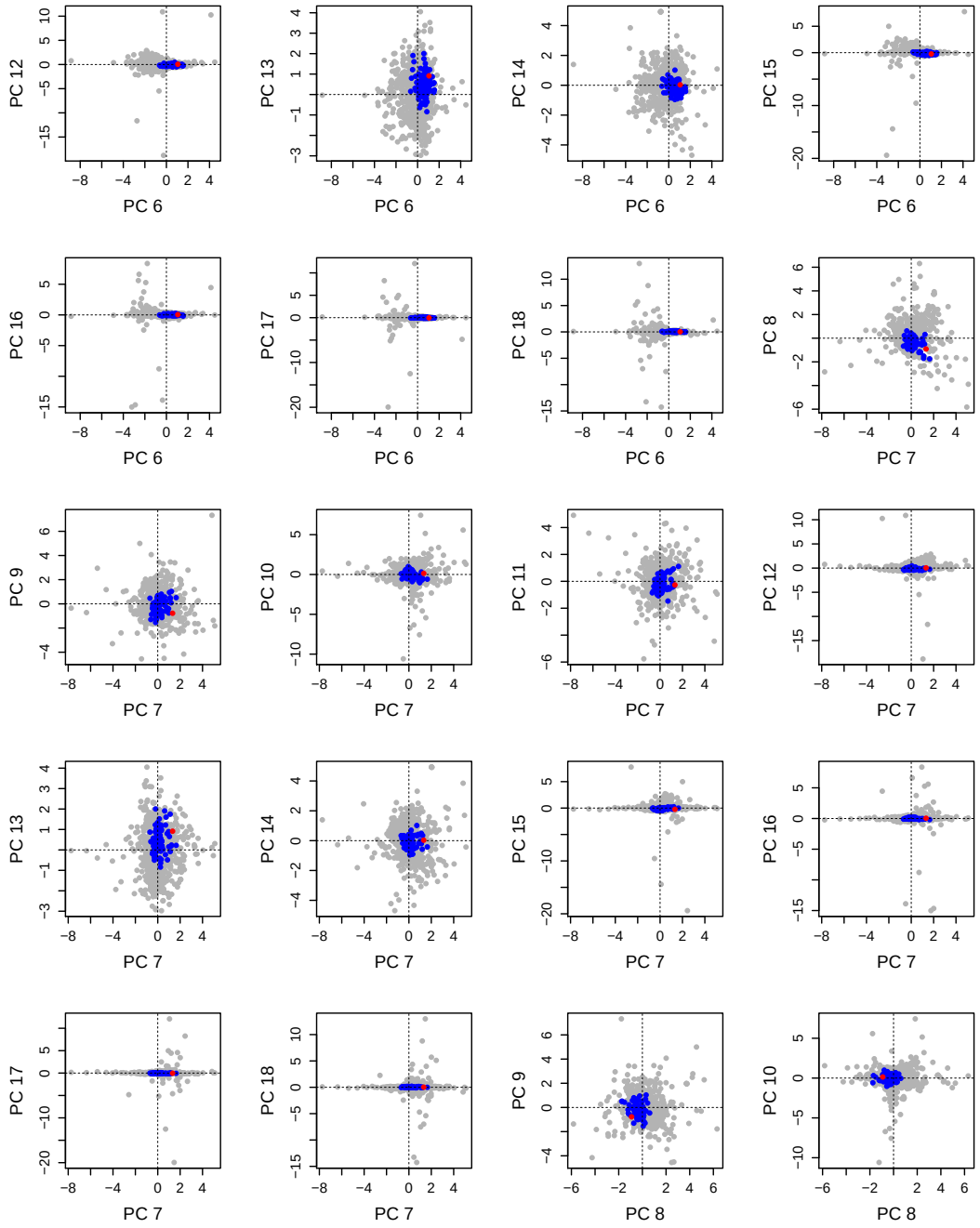

Figure 3.8: *Continued on next page*

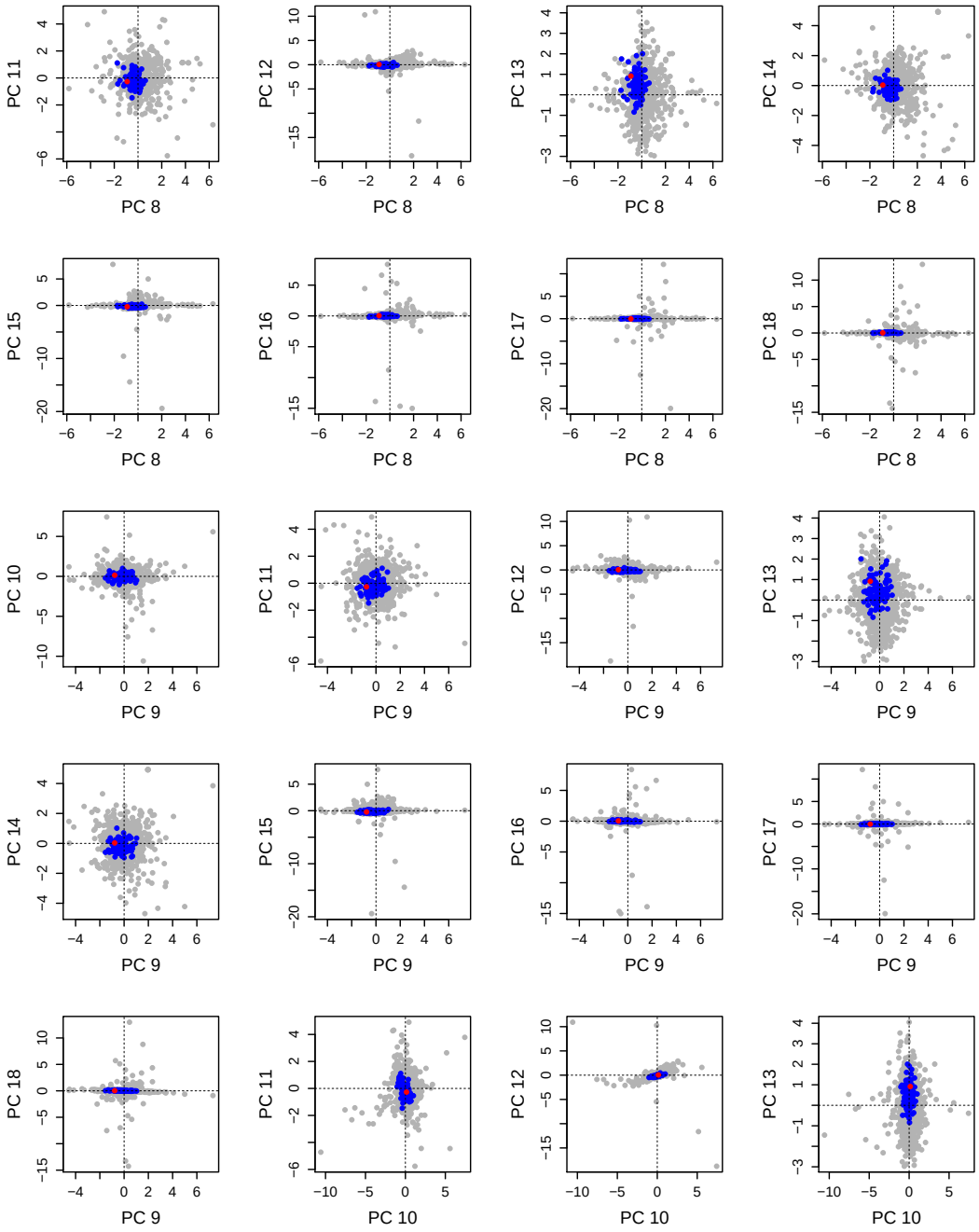

Figure 3.8: *Continued on next page*

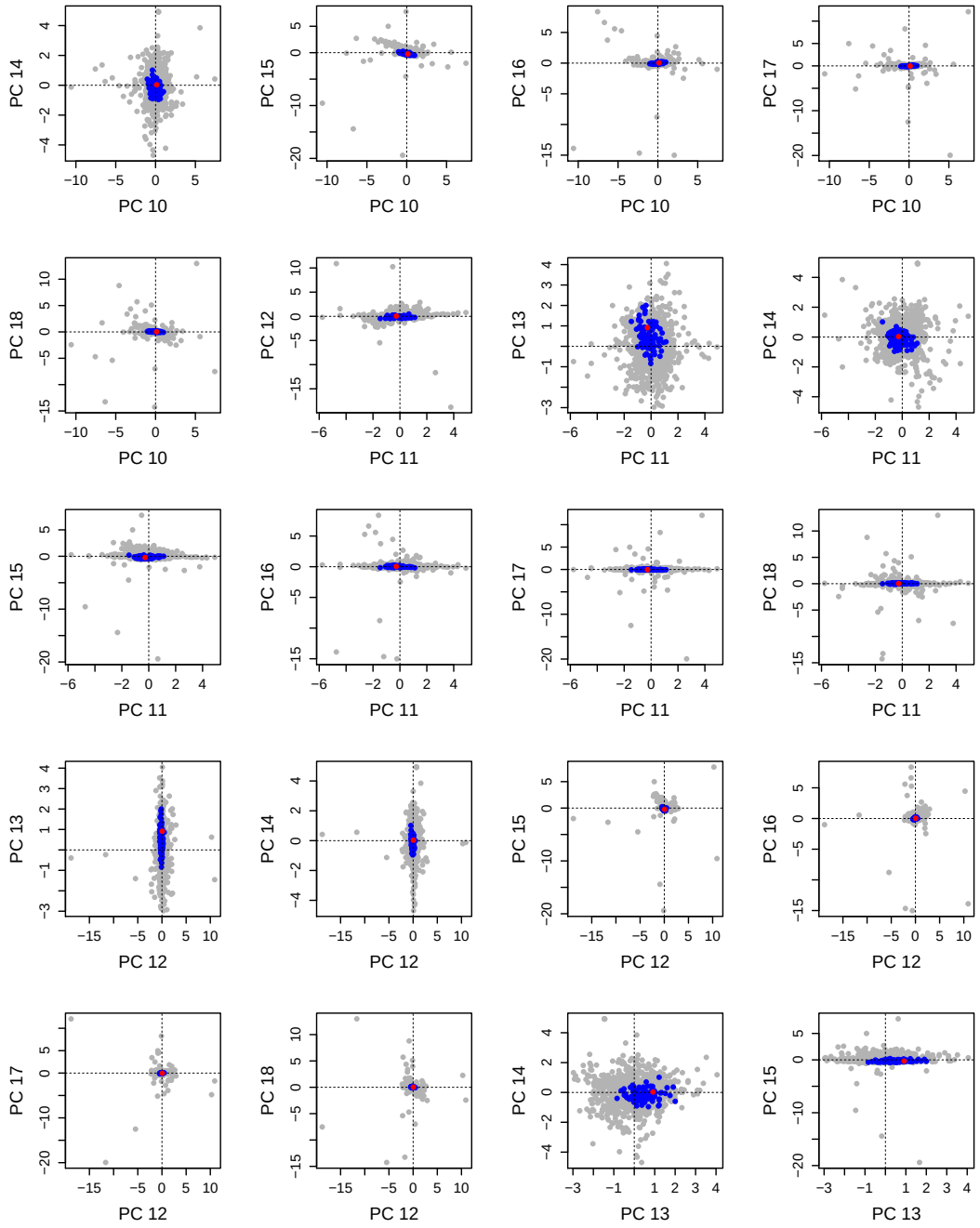

Figure 3.8: *Continued on next page*

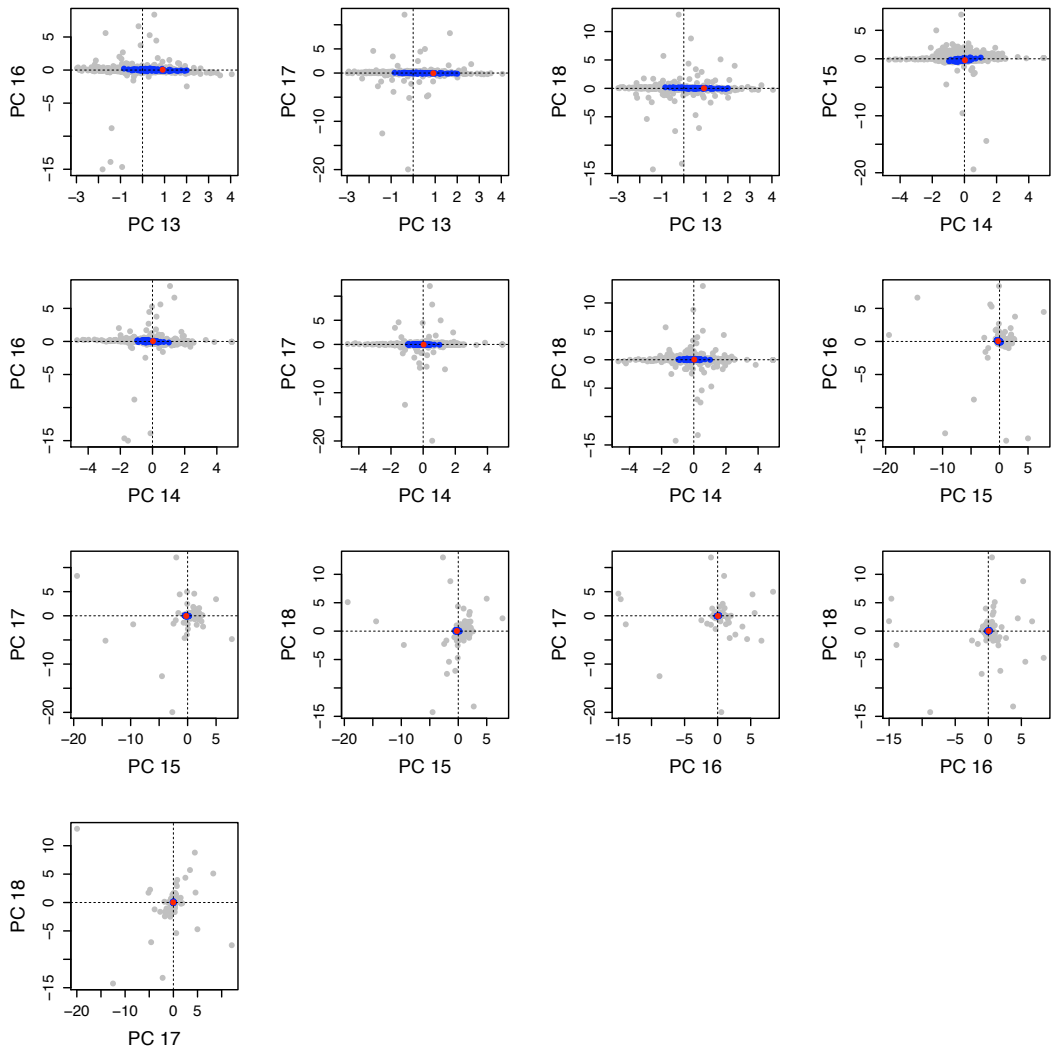

Figure 3.8: *Continued from previous page.* Pairwise prior predictive distribution of PC-rotated summary statistics. Gray points represent $N = 1,000$ simulations with parameter values drawn from the prior. The true value from the ibex data set is shown as a red dot. The fact that it is always embedded in the cloud of gray points means that the model and prior distributions are well specified. The $n' = 100$ points with smallest Euclidean distance from the observation are shown in blue. Those represent simulations used as training data sets for the *local* choice of summary statistics (see main text). In the main study, we used $N = 10^6$ and $n' = 1,000$; smaller numbers are used here for illustration of the principle.

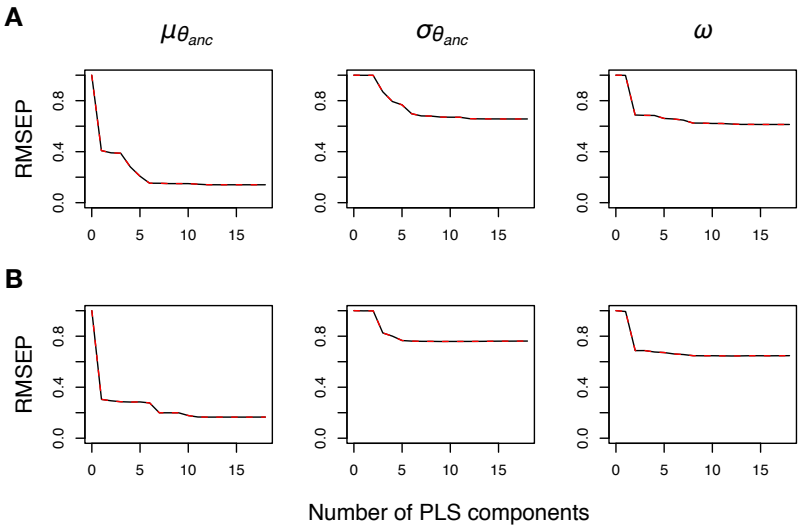

Figure 3.9: Root mean squared error of prediction (RMSEP) for PLS regression as a function of the number of PLS components used. As suggested by (Wegmann et al. 2009a), we chose the number of PLS components to be kept as summary statistics based on these plots. The RMSEP was obtained via leave-one-out cross-validation. (A) Global and (B) local choice of summary statistics via PLS (see main text). In (B), the observation from the ibex data set was used as the center. In both cases, we decided to keep the first ten components as summary statistics.

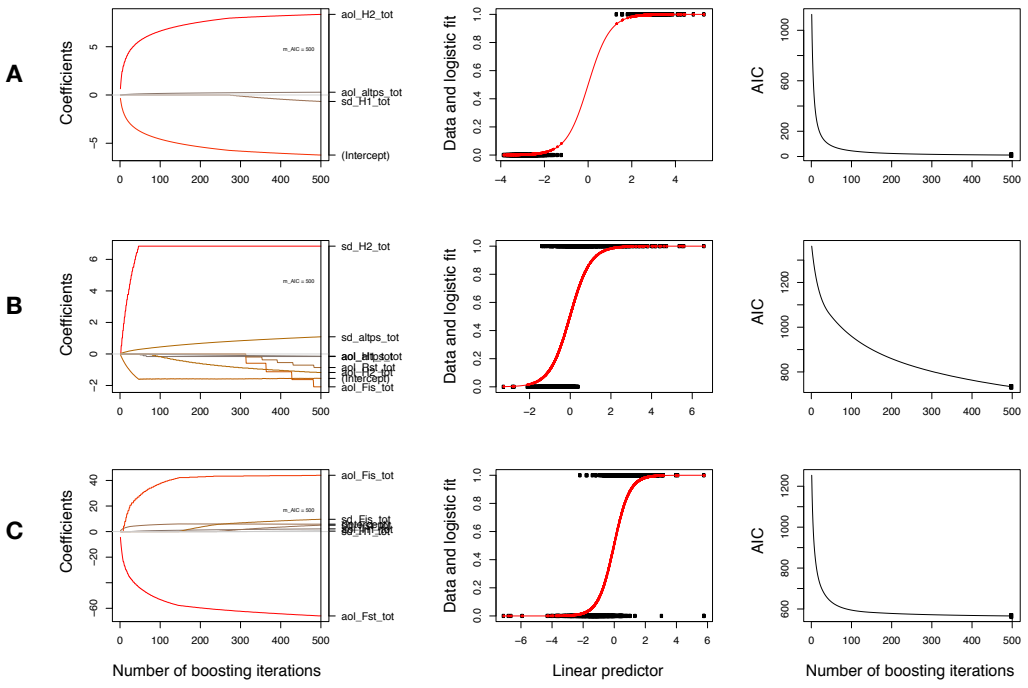

Figure 3.10: Choice of summary statistics via LogitBoost for the three parameters $\mu_{\theta_{\text{anc}}}$ (A), $\sigma_{\theta_{\text{anc}}}$ (B) and ω (C). Left column: Boosted coefficients $\lambda^{[m]}$ as a function of the number of iterations m. Middle column: Binary parameter class variable (Y, black) and logistic fit to the probability $\Pr[Y = 1 \mid \mathbf{X} = \mathbf{x}]$ (red), as a function of the linear predictor. Right column: Quality of fit in terms of AIC as a function of the number of iterations m. The thick black line marks the m_{stop} chosen. In the cases shown here, no minimum AIC was found for $m < 500$.

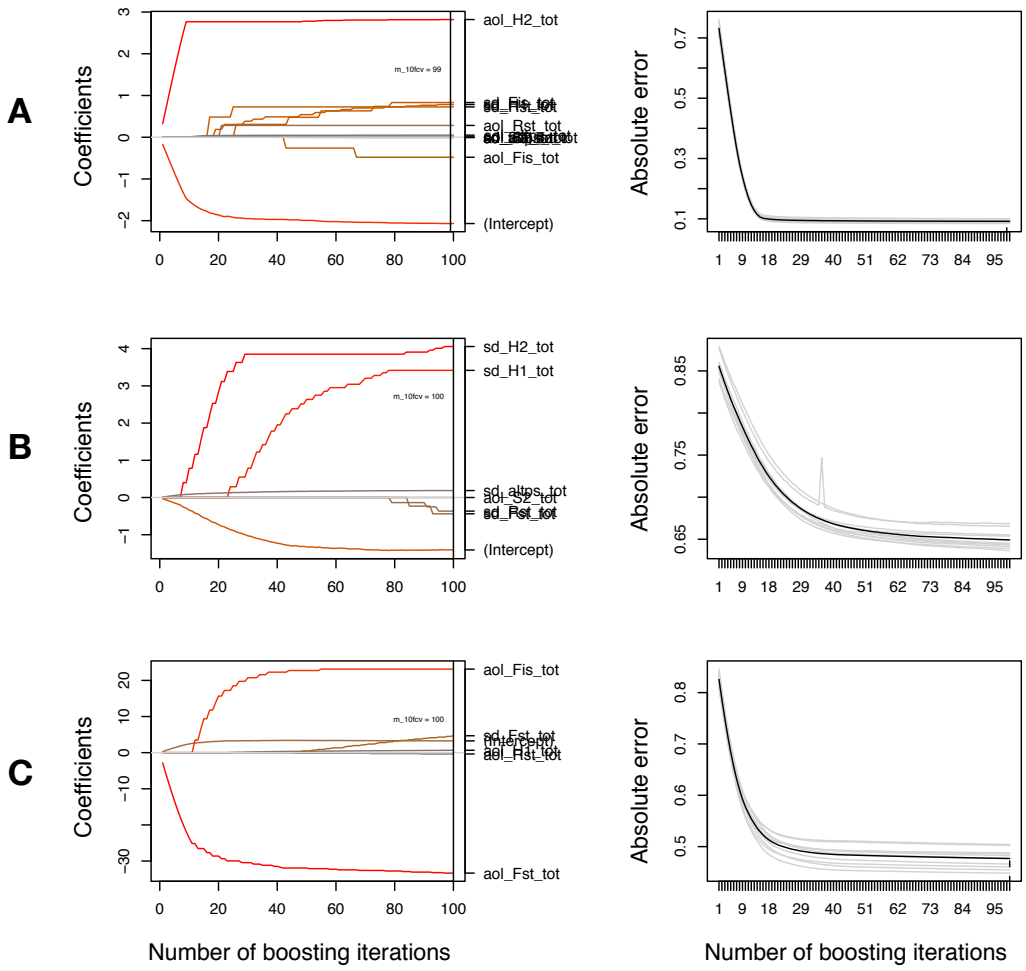

Figure 3.11: Choice of summary statistics via L_1Boosting for the three parameters $\mu_{\theta_{\text{anc}}}$ (A), $\sigma_{\theta_{\text{anc}}}$ (B) and ω (C). Left column: Boosted coefficients $\lambda^{[m]}$ as a function of the number of iterations m. Right column: Quality of fit in terms of the bootstrapping error, as a function of the number of iterations m. The dashed vertical line marks the m_{stop} chosen. In the cases shown here, no minimum absolute error was found for $m < 100$.

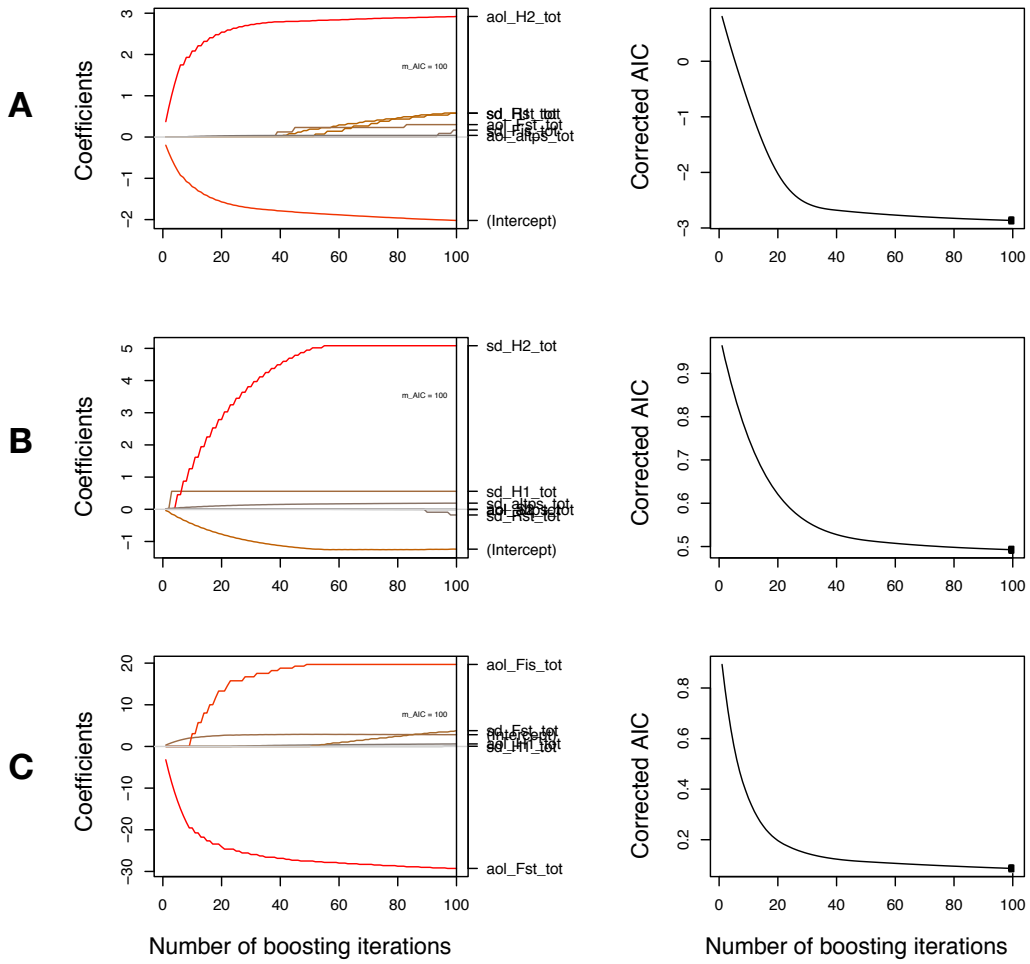

Figure 3.12: Choice of summary statistics via L_2Boosting for the three parameters $\mu_{\theta_{\text{anc}}}$ (A), $\sigma_{\theta_{\text{anc}}}$ (B) and ω (C). Left column: Boosted coefficients $\lambda^{[m]}$ as a function of the number of iterations m. Right column: Quality of fit in terms of the corrected AIC as a function of the number of iterations m. The thick black line marks the m_{stop} chosen. In the cases shown here, no minimum absolute error was found for $m < 100$.

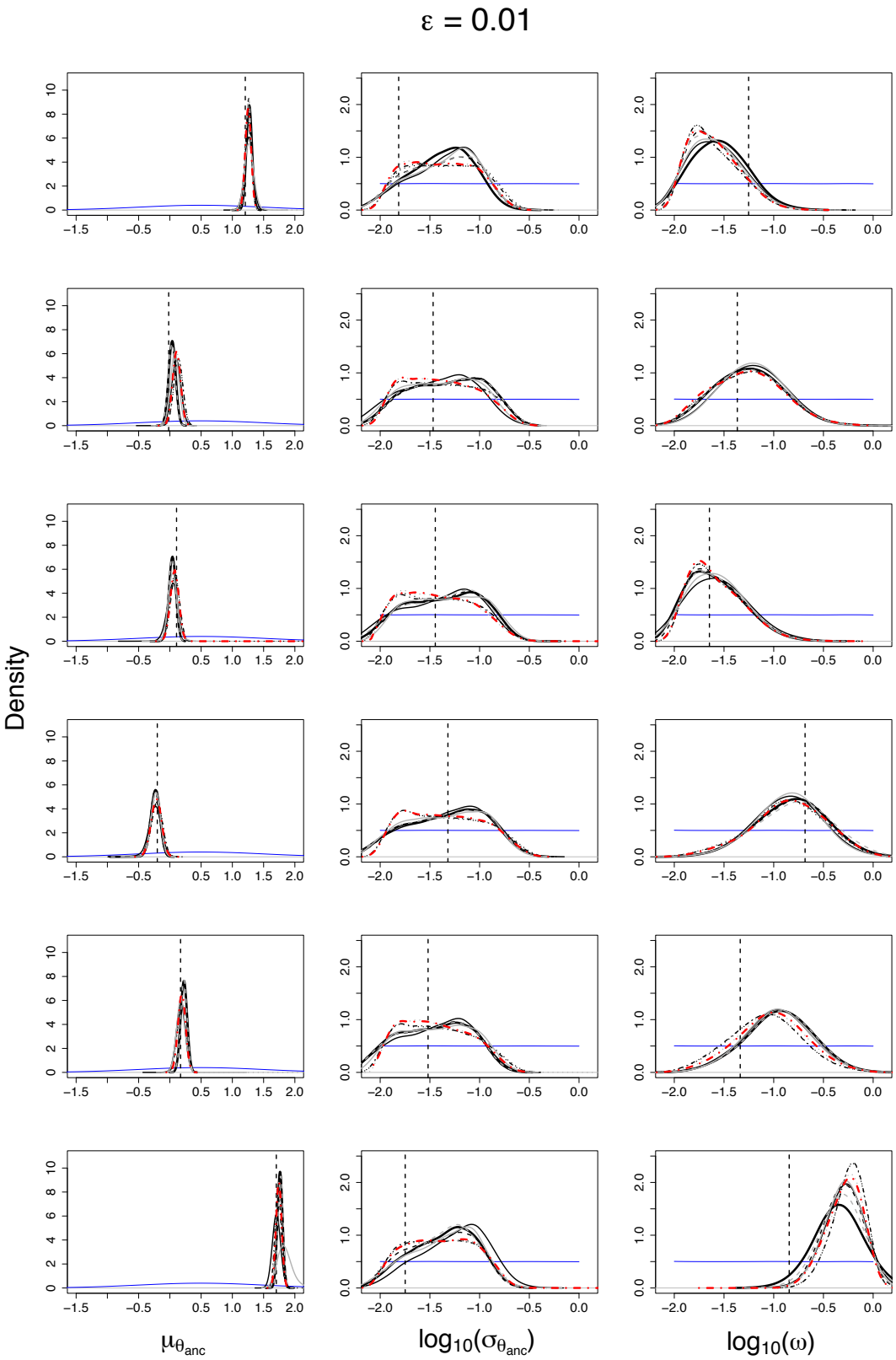

Figure 3.13: Posterior distributions inferred for six random test data sets with acceptance rate $\epsilon = 0.01$. Methods are as described in the main text. True values are given by a dashed vertical line, prior distributions in blue (*cf.* Table 3.1).

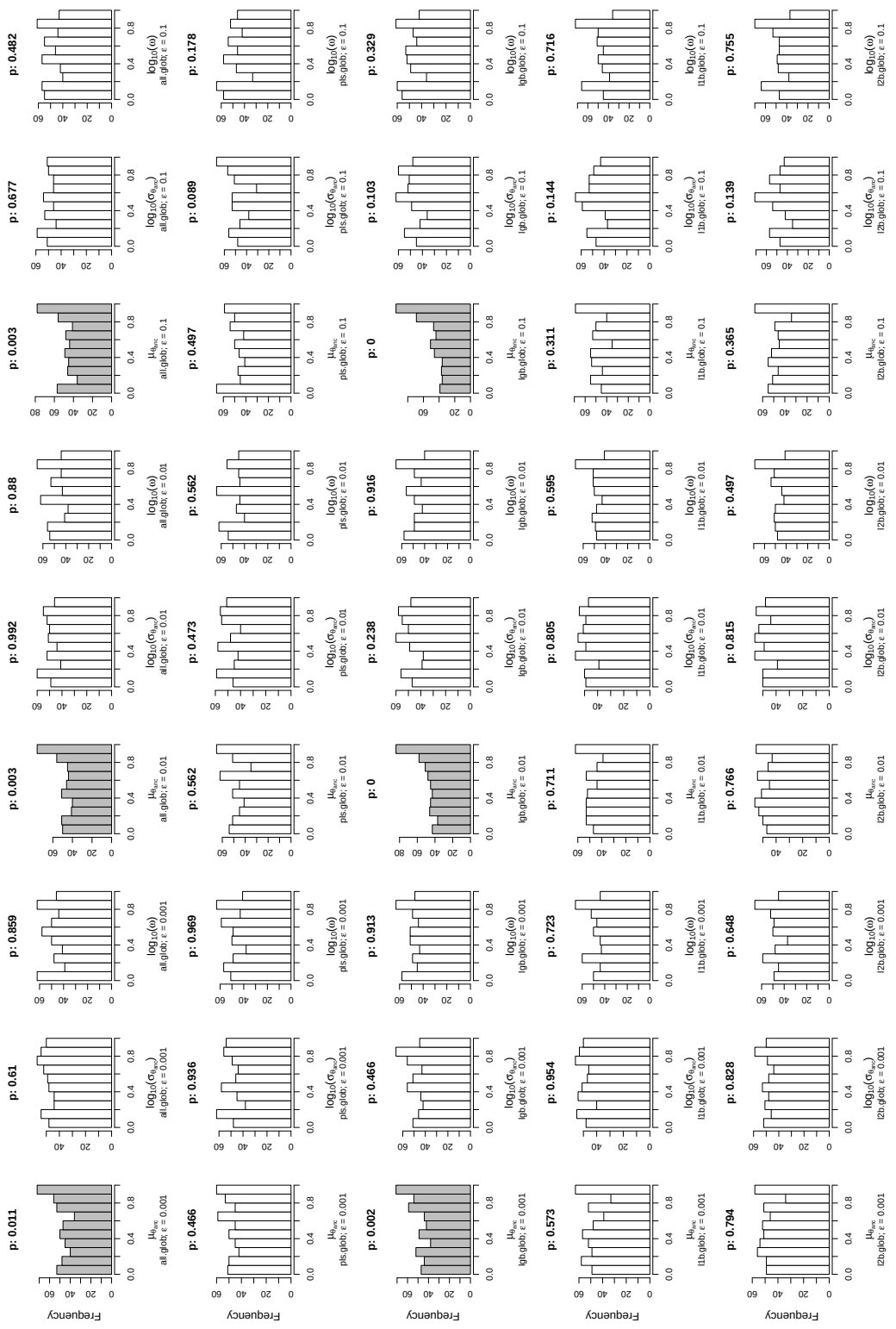

Figure 3.14: Coverage property of posterior distributions inferred with different choices of summary statistics on a global scale. Histograms show the distribution across 500 independent test estimations of the posterior probabilities of the true parameter values. The distribution is expected to be uniform (Wegmann et al. 2009a). Left-skewed or right-skewed distributions indicate that the parameter is on average over- or underestimated, respectively. Peaked or U-shaped distributions result from posterior distributions that are too wide or too narrow, respectively. Non-uniform distributions of posterior probabilities are shaded in gray (p-values from Kolmogorov-Smirnov test as explained in the text).

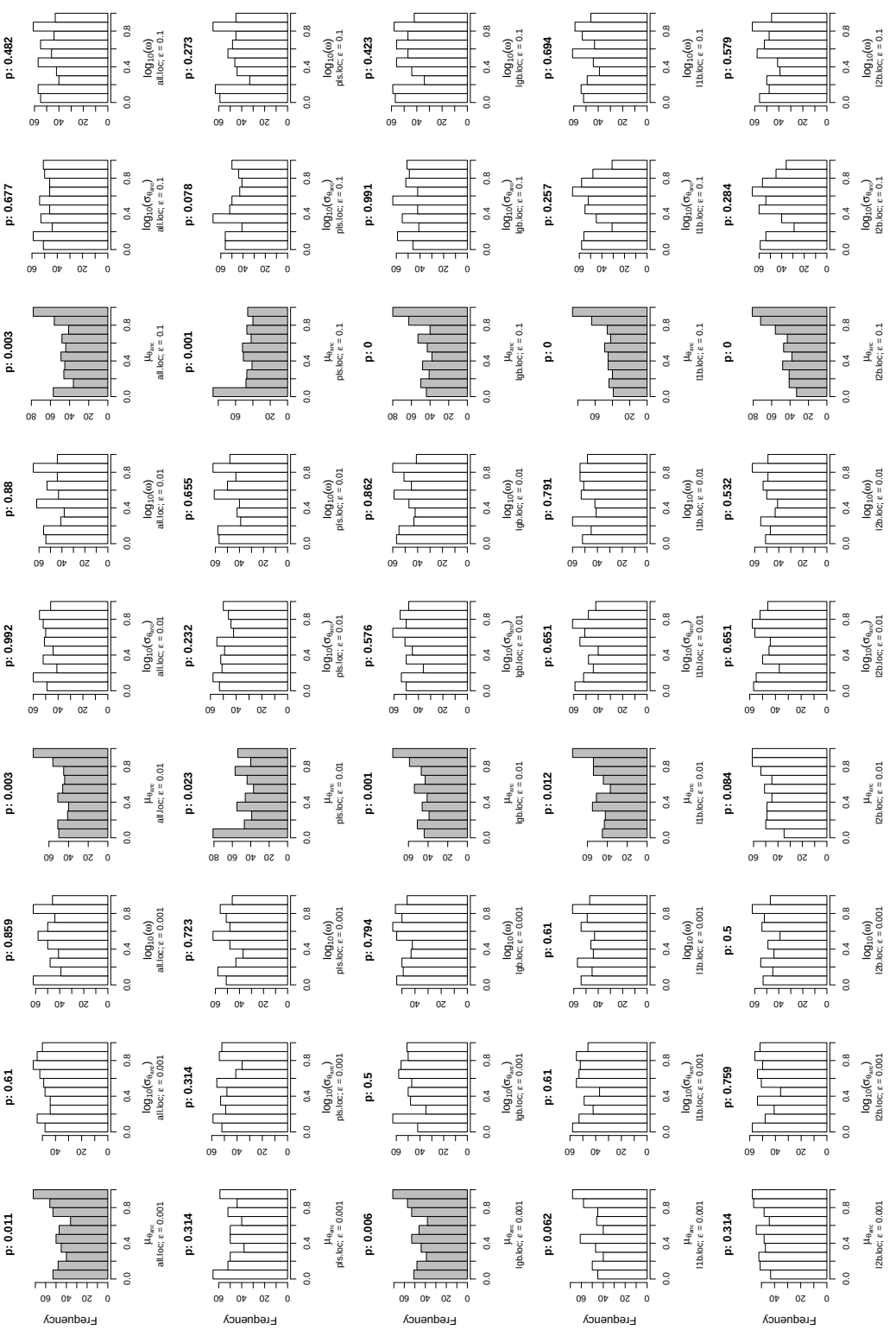

Figure 3.15: Coverage property of posterior distributions inferred with different choices of summary statistics on a local scale. Non-uniform distributions of posterior probabilities of the true parameter are shaded gray (p-values from Kolmogorov-Smirnov test as explained in the text). Note that the first row here corresponds to the first row in Figure 3.14. Further details as in Figure 3.14.

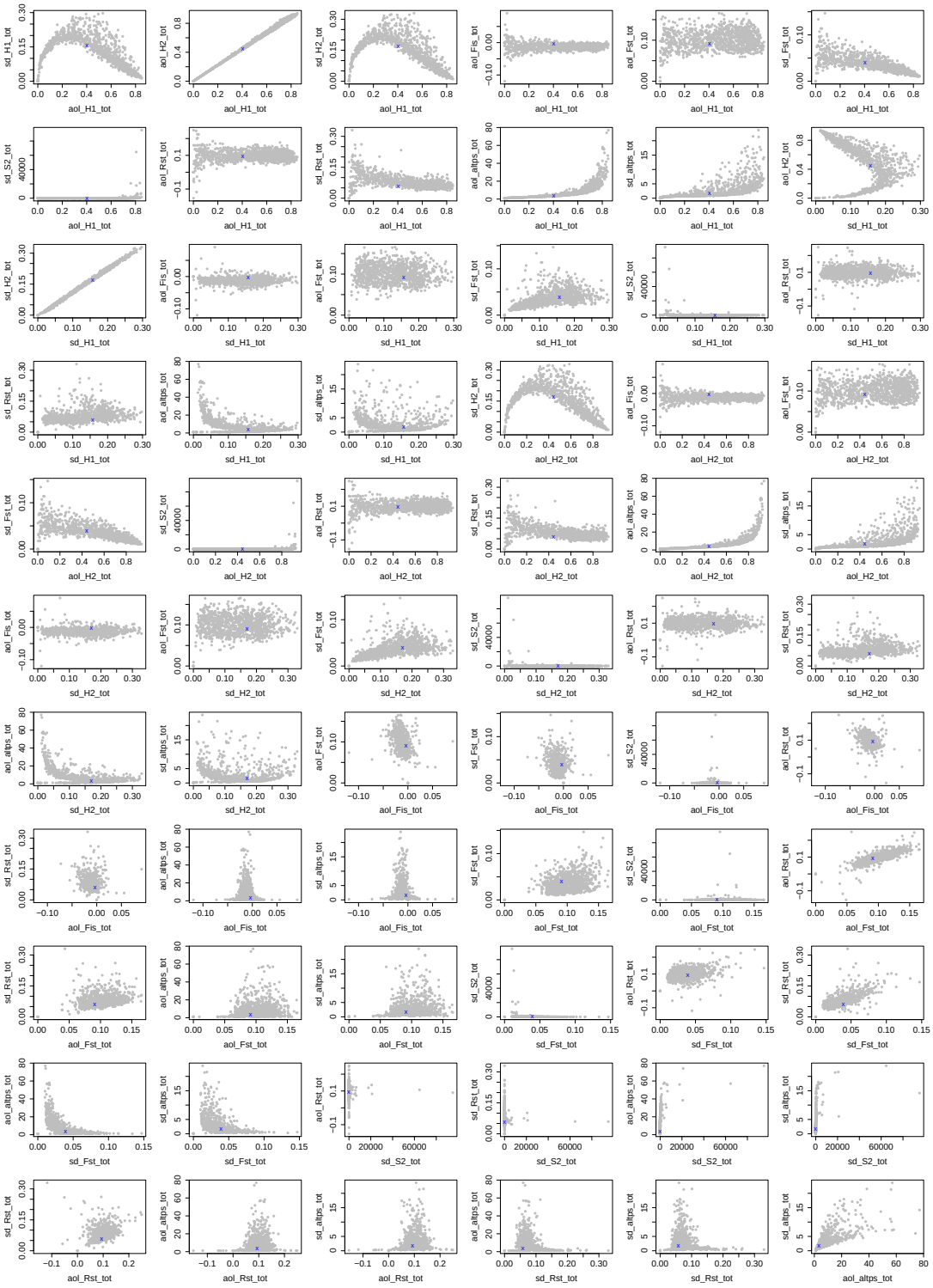

Figure 3.16: Pairwise prior predictive distribution of summary statistics on original scale. Only summary statistics chosen with the `lgb.glob` method are shown. Gray points represent $N = 1,000$ simulations with parameter values drawn from the prior. The true value from the ibex data set is shown as a blue cross; *aol*, average over loci; *sd*, standard deviation over loci.

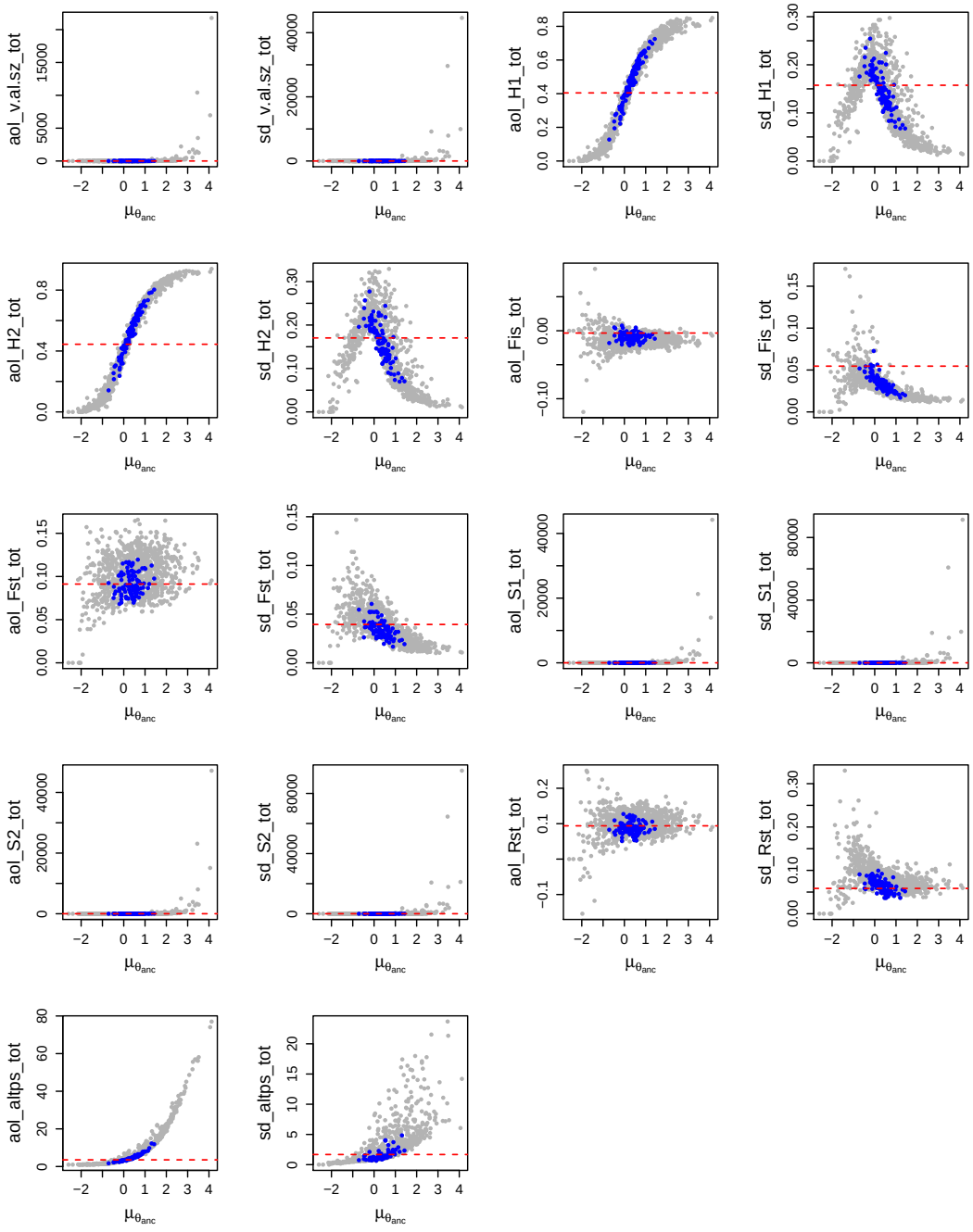

Figure 3.17: Relation between $\mu_{\theta_{\text{anc}}}$ and the candidate summary statistics. The summary statistics are on the y-axis; *aol*, average over loci; *sd*, standard deviation over loci. Gray points represent $N = 1,000$ simulations, the red dashed line corresponds to the observation for Alpine ibex. Blue points represent the $n' = 100$ simulations closest to the observation, where 'closeness' was defined as described in the main text (*cf.* Figure 3.8).

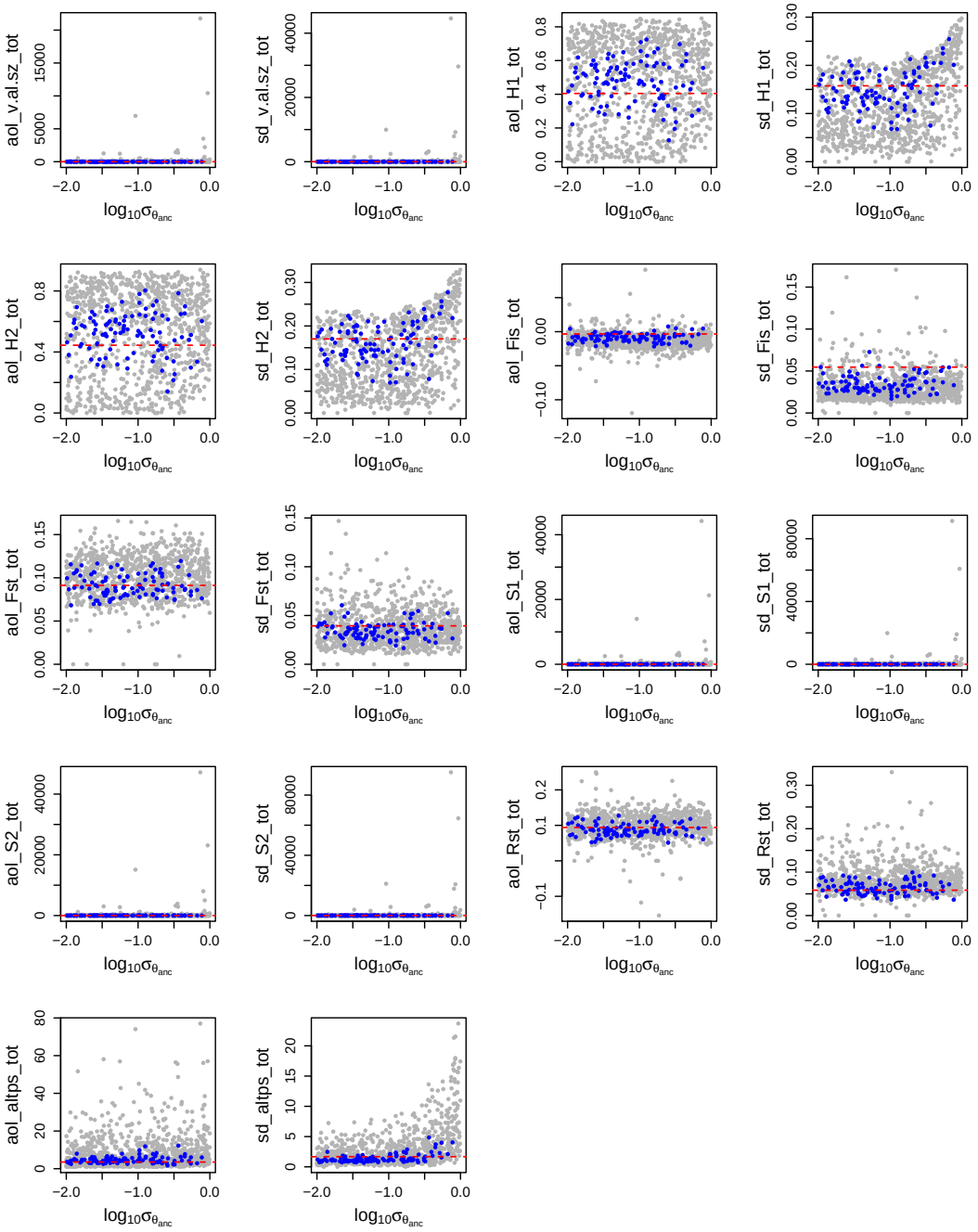

Figure 3.18: Relation between $\sigma_{\theta_{\mathrm{anc}}}$ and the candidate summary statistics. Details as in Figure 3.17.

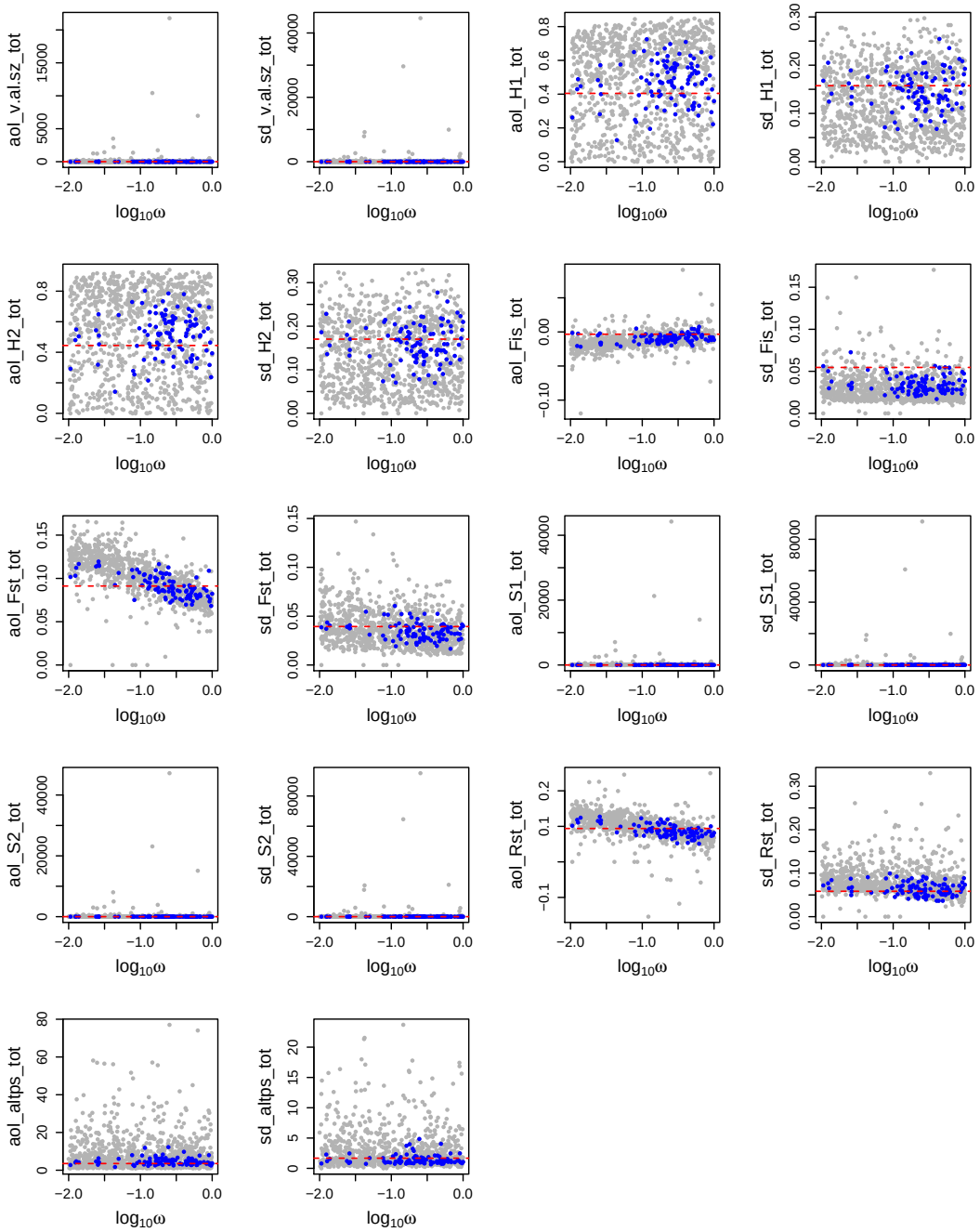

Figure 3.19: Relation between ω and the candidate summary statistics. Details as in Figure 3.17.

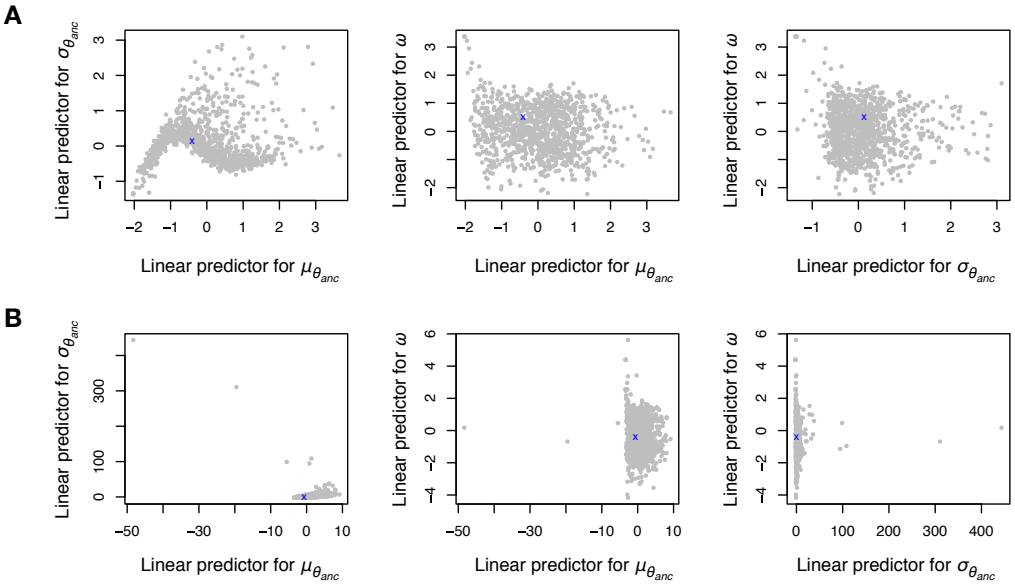

Figure 3.20: Effect of local choice on scale of summary statistics. Summary statistics were chosen with L_2Boosting as explained in the main text. For each parameter, one linear combination of the original statistics is used as the new summary statistic. These linear combinations are plotted against each other. (A) Global choice of summary statistics. (B) Local choice of summary statistics. Gray points represent $N = 1,000$ simulations and the blue cross marks the value observed for Alpine ibex. The local choice of statistics leads to a rescaling compared to the global choice.

3.8 Supporting information: Additional methods

3.8.1 Demography and life cycle in simulations

In the following, we give additional details of the demographic model and the ibex-specific settings used in the simulations. All of this is implemented in the program `SPoCS` (Simulate Populations under Complex Scenarios) written in `Java`^{TM} and available on the website `http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`.

Life cycle

Alpine ibex is a long-lived, middle-sized ungulate species (Toïgo et al. 2002, 2007). We divide the life cycle into years and a year into discrete events, some of which are further described below. We set the maximum age of females and males to 22 and 17 years, respectively (Nievergelt 1966; Toïgo et al. 2007). Females and males reach sexual maturity at an age of 3 years (Nievergelt 1966; Stuwe and Grodinsky 1987; Toïgo et al. 2002), and the expected age of first reproduction for females and males is 4 and 9 years, respectively (Loison et al. 2002; Toïgo et al. 2002). In our simulations, females and males stop reproducing when older than 20 and 15 years, respectively.

Founder/admixture events

A new deme is established by founder individuals taken from previously existing demes. The minimum and maximum age of a founder is 1 and 7 years, respectively, independently of sex. Existing demes may receive further individuals from other demes at later points in time (as specified in Supporting File 3.7 *transfers*). The range of ages allowed for these admixing individuals is the same as for founders. Founder/admixture events take place at the beginning of the year, before the regulating deaths (see below).

Reproduction

Females reproduce according to a baseline fertility parameter f. It gives the probability that, for a given year, a particular female will reproduce. If the female reproduces, she mates with a male randomly chosen from the set of males with access to matings in that year (see below). Given a particular female reproduces, it may have one or two offspring. This is controlled by the twin rate parameter $z := \Pr[\text{twins} \mid \text{female reproduces}]$. We set $f = 0.4$ (Nievergelt 1966; Stuwe and Grodinsky 1987) and $z = 0.08$ (Toïgo et al. 2002). Males can get access to matings if they reached the expected age of first reproduction (9 years) and are then counted as potentially reproducing. If, in a deme, no males older than 9 years are available, all males older than the age of sexual maturity (3 years) are considered potentially reproducing. The proportion of these potentially reproducing males that actually get access to matings is defined as ω (see main text). It is one of the parameters to be estimated in this study.

Deme size control

If the number of offspring required to reach the deme size of the next year cannot be produced by the female baseline fertility f (see above), additional females are allowed to reproduce: Rather than allowing only females to reproduce who reached the expected age of first reproduction (4

years), all females who reached the age of sexual maturity (3 years) may reproduce in this case. If, on the other hand, baseline reproduction results in more individuals than needed to reach the census size of the next year, surplus individuals are removed. These regulating deaths are irrespective of age and sex, and additional to the natural deaths of senescence. In any case, we limit the proportion by which the reproductive need may be overshot per year to 0.2.

Migration

We simulate migration after the regulating deaths, but before reproduction. Females and males must have reached the age of 3 years before they emigrate (they are then 'potential emigrants'). For a given source deme, the total of individuals to be sent to all connected demes (see main text) are put into an emigrant pool. Emigrants are then randomly distributed to the receiver demes in proportions corresponding to the emigration rates.

3.8.2 Explicit forms of minimum expected loss and negative gradient

The FGD algorithm given in the Appendix of the main text is generic. It is instructive to study the explicit form of expressions in step 1 and 2 of this algorithm for the specific loss functions used here. To this purpose, we follow Friedman et al. (2000), Friedman (2001) and Bühlmann and Hothorn (2007).

Population minimizer of expected loss

We first give explicit forms of the population minimizer (3.5) for the three loss functions in equations (3.8), (3.9) and (3.11). These are obtained by minimizing the expectation of the joint distribution of $\mathbf{X}$ and Y, $\mathbb{E}_{\mathbf{X},Y}[L(Y,F)]$, where $L(\cdot,\cdot)$ is the generic loss function and $F = F(\mathbf{X})$. In our context, it is enough to take the expectation conditional on $\mathbf{X} = \mathbf{x}$, $\mathbb{E}_Y[L(Y,F) \mid \mathbf{x}]$.

For the $L1$-loss in (3.8), $F^*(\cdot)$ from (3.5) is obtained as the $F(\cdot)$ that minimizes $\mathbb{E}_Y[|Y-F| \mid \mathbf{x}]$. By the definition of the median, the population minimizer is (Friedman 2001; Bühlmann and Hothorn 2007)

$$F^*(\mathbf{x}) = \text{median}(Y \mid \mathbf{x}). \tag{3.13}$$

For the $L2$-loss in (3.9), the expected loss is $\mathbb{E}_Y[(Y-F)^2/2 \mid \mathbf{x}]$, and $F^*(\cdot)$ is obtained by setting the derivative with respect to F to zero:

$$\begin{aligned} \frac{\partial}{\partial F}\,\mathbb{E}_Y\left[\frac{1}{2}\,(Y-F)^2 \;\middle|\; \mathbf{x}\right] &= \frac{1}{2}\,\frac{\partial \mathbb{E}_Y[Y^2 \mid \mathbf{x}]}{\partial F} - \frac{\partial \mathbb{E}_Y[Y\,F \mid \mathbf{x}]}{\partial F} + \frac{1}{2}\,\frac{\partial \mathbb{E}_Y[F^2 \mid \mathbf{x}]}{\partial F} \\ &= 0 - \mathbb{E}_Y[Y \mid \mathbf{x}] + F(\mathbf{x}) = 0, \end{aligned} \tag{3.14}$$

from which the familiar result

$$F^*(\mathbf{x}) = \mathbb{E}_Y[Y \mid \mathbf{x}] \tag{3.15}$$

follows (Friedman 2001; Bühlmann and Hothorn 2007).

Friedman et al. (2000) show how to derive the population minimizer of the negative binomial log-likelihood in equation (3.11). For notational convenience, we encode the response by $\tilde{Y} = 2Y - 1 \in \{-1, 1\}$. The likelihood in (3.11) can then be written as

$$L(\tilde{Y}, F) = \log\left(1 + e^{-\tilde{Y}F}\right). \tag{3.16}$$

In analogy to our previous definition, we set $p(\mathbf{x}) := \Pr[\tilde{Y} = 1 \mid \mathbf{X} = \mathbf{x}]$, and hence $1 - p(\mathbf{x}) := \Pr[\tilde{Y} = -1 \mid \mathbf{X} = \mathbf{x}]$. Dropping the arguments, we have

$$\begin{aligned} \mathbb{E}_{\tilde{Y}}[L \mid \mathbf{x}] &= \mathbb{E}_{\tilde{Y}}\left[\log\left(1 + e^{\tilde{Y}F}\right) \middle| \mathbf{x}\right] \\ &= p \log\left(1 + e^{-F}\right) + (1 - p) \log\left(1 + e^{F}\right). \end{aligned} \tag{3.17}$$

The partial derivative with respect to F is

$$\mathbb{E}_{\tilde{Y}}\left[\log\left(1 + e^{\tilde{Y}F}\right) \middle| \mathbf{x}\right] = -p \frac{e^{-F}}{1 + e^{-F}} + (1 - p) \frac{e^{F}}{1 + e^{F}}. \tag{3.18}$$

Setting to zero and solving for F, we obtain the population minimizer

$$F^*(\mathbf{x}) = \log\left[\frac{p(\mathbf{x})}{1 - p(\mathbf{x})}\right]. \tag{3.19}$$

Notice that Friedman et al. (2000) and Bühlmann and Hothorn (2007) use a slightly different parameterization, namely setting F equal to *one half* of the logit-transform, such as to have the population minimizer equal to the one for the exponential loss criterion. The population minimizers in (3.13), (3.15) and (3.19) imply that the initial function estimates in step 1 of the FGD algorithm (Appendix) must be set to $F^*(\cdot) \equiv \text{median}(Y)$ for the $L1$-loss, to $F^*(\cdot) \equiv \bar{Y}$ for the $L2$-loss, and to $F^*(\cdot) \equiv \log[\hat{p}/(1 - \hat{p})]$ for the negative binomial log-likelihood loss.

Negative gradient

To calculate the negative gradient vector $(U_1, \ldots, U_n)$ in step 2 of the FGD algorithm (Appendix), we need the partial derivative of the loss function with respect to the target function F. Any element U_i is obtained as this partial derivative evaluated at the previous function estimate $\hat{F}^{[m-1]}(\mathbf{x}_i)$. Formally,

$$U_i = -\frac{\partial}{\partial F} L(Y_i, F)\bigg|_{F = \hat{F}^{[m-1]}(\mathbf{X}_i)}. \tag{3.20}$$

For the $L1$-loss in (3.8), we have

$$-\frac{\partial}{\partial F}\left[|Y_i - F|\right] = \frac{Y_i - F}{|Y_i - F|} = \text{sgn}(Y_i - F), \tag{3.21}$$

which implies the negative gradient component

$$U_i = \text{sgn}\left[Y_i - \hat{F}^{[m-1]}(\mathbf{X}_i)\right] \tag{3.22}$$

in step 2 of the FGD algorithm (*cf.* Friedman 2001).

For the $L2$-loss in (3.9),

$$-\frac{\partial}{\partial F}\left[\frac{1}{2}(Y_i - F)^2\right] = Y_i - F, \tag{3.23}$$

which amounts to

$$U_i = Y_i - \hat{F}^{[m-1]}(\mathbf{X}_i) \tag{3.24}$$

in step 2 of the FGD algorithm (*cf.* Friedman 2001; Bühlmann and Hothorn 2007).

Last, for the negative binomial log-likelihood we again use $\tilde{Y} = 2Y - 1 \in \{-1, 1\}$ and find

$$-\frac{\partial}{\partial F} L(\tilde{Y}_i, F) = -\frac{\partial}{\partial F} \log\left(1 + e^{-\tilde{Y}_i F}\right) = \frac{\tilde{Y}_i \, e^{-\tilde{Y}_i F}}{1 + e^{-\tilde{Y}_i F}}. \tag{3.25}$$

This leads to the negative gradient component

$$U_i = \frac{\tilde{Y}_i \, e^{-\tilde{Y}_i \hat{F}^{[m-1]}(\mathbf{X}_i)}}{1 + e^{-\tilde{Y}_i \hat{F}^{[m-1]}(\mathbf{X}_i)}} \tag{3.26}$$

in step 2 of the FGD algorithm.

Table 3.6: (The table is intended as an online Supporting File (*census sizes*) and therefore not displayed here. It is available on the website `http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`.

Census population sizes of Alpine ibex demes in the Swiss Alps

Table 3.7: (The table is intended as an online Supporting File (*transfers*) and therefore not displayed here. It is available on the website `http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`.

Numbers of Alpine ibex transferred between demes by humans

Chapter 4

Inferring recent migration rates in a complex model with ABC: Joint versus pairwise estimation

The work presented in this chapter was influenced by discussions with Andreas Futschik and Mark Beaumont. They had a strong impact on the design of the study. The chapter is intended for publication in Genetics*, as a companion paper to the one resulting from chapter 3 of this thesis, with Andreas and Mark as co-authors.*

4.1 Introduction

Gene flow via migration or dispersal is of interest for several reasons. First, it is a modulator of speciation and has an impact on species range (Kirkpatrick and Ravigné 2002; Lenormand 2002). Its absence is a requirement for the early phase of allopatric speciation, while some secondary contact is needed to complete speciation via reinforcement (Barton and Hewitt 1985; Servedio and Noor 2003). Theory and recent empirical findings suggest that speciation is possible in the permanent presence of gene flow, although the parameter range may be small (Endler 1977; Gavrilets 2003; Nosil 2008; Barton 2010). Gene flow may swamp locally favoured alleles and therefore limit local adaptation (Morjan and Rieseberg 2004; Nagylaki and Lou 2008). Second, gene flow is one aspect of population history, which is of interest on its own, for instance in the case of human expansion (Rosenberg et al. 2002; Currat and Excoffier 2005). More generally, it is essential for the interpretation of observed patterns of genetic diversity (Charlesworth et al. 2003). Third, gene flow plays a role in the maintenance of genetic diversity, and is therefore of importance in conservation biology. It may reduce the risk of inbreeding depression and fixation of deleterious alleles (Keller and Waller 2002) and has an impact on the definition of management units (Waples and Gaggiotti 2006; Palsboll et al. 2007). Moreover, gene flow is associated with the spread of diseases (Biek and Real 2010), drug resistance (Webster et al. 2008) or genetically modified organisms (Chapman and Burke 2006).

Inferring rates of migration from genetic data has advantages compared to direct observation (Neigel 1997), but it is a formidable challenge under realistic models. We devise an approach for estimating multiple migration rates in an approximate Bayesian framework. It uses summary

statistics and conditions on independent demographic information (*cf.* Estoup et al. 2004). We show that when we split the full estimation problem into sub-problems, the net accuracy increases with the number of parameters to be estimated, relative to the accuracy reached when the full problem is analyzed at once. Our setting is motivated by recent studies of genetic diversity in a re-introduced and spatially subdivided population of Alpine ibex (*Capra ibex*) in Switzerland. Although the species has successfully recovered from near extinction, it remains of conservation concern. Genetic diversity within demes is low and differentiation among demes relatively strong (Stuwe and Scribner 1989; Scribner and Stuwe 1994; Biebach and Keller 2009, 2010). Occasional local outbreaks of diseases such as foot rot (Belloy et al. 2007) or infectious keratoconjunctivitis (Tschopp et al. 2005; Ryser-Degiorgis et al. 2009) raise worries that the respective pathogens (*e.g. Dichelobacter nodosus*, *Mycoplasma conjunctivae*) could spread via migration. Estimating rates and direction of migration are therefore of twofold interest.

In principle, population differentiation is a continuum: the degree of connectivity of demes may vary from panmixia to complete isolation (Figure 1 in Waples and Gaggiotti 2006). In practice, gene flow is often studied from one of two extreme perspectives, either asking about deviations from the null model of panmixia (*e.g.* Bowen et al. 2005; Waples and Gaggiotti 2006), or starting from previously defined demes and asking about their degree of isolation (*e.g.* Lucas et al. 2009). In the case of Alpine ibex, geography and spatial distribution clearly suggest the latter perspective. Nowadays, Alpine ibex live in altitudes of 1,800 to 3,000 meters. Their ranges are restricted to mountain ridges and, usually, deep valleys, bigger roads and rivers are not crossed. Discrete demes can be defined according to the Swiss Federal Office for the Environment (FOEN) and game keepers (Biebach and Keller 2009).

The mathematical treatment of gene flow goes back to Wright (1931) and Haldane (1932). When the first allozyme samples became available (Hubby and Lewontin 1966), the theory was applied to data, essentially using the relationship between migration rate and F_{ST} (Wright 1922; Cockerham and Weir 1993) derived for the island model at equilibrium (Wright 1943). A plethora of studies have since used F_{ST} or modifications of it (Nei 1973; Hudson et al. 1992; Slatkin 1995; Rousset 1996) under this approach. Great effort has been spent on obtaining valid estimates of F_{ST} from genetic data (Nei and Chesser 1983; Weir and Cockerham 1984; Weir and Hill 2002). However, strong – and in most cases unrealistic – assumptions such as symmetric migration rates, constant and equal deme sizes, infinitely many demes and drift-migration equilibrium are made. Violations may result in misleading F_{ST}-based estimates (Whitlock and McCauley 1999; Balloux and Lugon-Moulin 2002, but see Barton and Slatkin 1986). Alternative methods have been proposed, such as maximum-likelihood estimation under the diffusion approximation (Slatkin and Barton 1989), the study of rare alleles (Slatkin 1985; Slatkin and Barton 1989), or cladistic measures of gene flow that compare a gene tree with sampling locations (Slatkin and Maddison 1989; Hey and Machado 2003). Essentially, all these approaches use the island model of migration *and* assume drift-migration equilibrium. In parallel, attempts were made to either relax the assumption of drift-migration equilibrium under the island model (Latter 1973; Takahata and Slatkin 1990; Takahata 1995), or to relax the island-model assumptions of symmetric migration rates (Tufto et al. 1996) and equal deme sizes (Gaggiotti and Excoffier 2000), but keeping the assumption of drift-migration equilibrium.

Coalescent theory (Kingman 1982) and increasing computational power then boosted the development of likelihood-based methods of inference. These use Felsenstein's (1988) equation for the likelihood of parameters that influence the genealogy given observed data, and employ importance sampling (IS; Griffiths and Tavaré 1994a,b; Beerli and Felsenstein 1999) or Markov chain Monte Carlo (MCMC; Kuhner et al. 1995) to explore the large space of potential genealogies. These approaches have later been improved (Stephens and Donnelly 2000; Hey and Nielsen 2007) and implemented in software packages (Bahlo and Griffiths 2000; Beerli and Felsenstein 2001; Kuhner 2006). Wakeley (1996b,a) introduced the isolation with migration (IM) model, in which two demes split at some time in the past and then continue exchanging migrants. He showed that the variance of the number of pairwise differences in DNA sequences from the two demes can be used to distinguish between recent divergence with complete isolation and long-term drift-migration equilibrium (Takahata and Slatkin 1990). Hey and Nielsen (2004) extended the IM model to multiple loci and Hey and Nielsen (2007) devised a more efficient way of integrating Felsenstein's (1988) equation. More recently, Hey (2010) extended inference under the IM model to multiple demes.

Recent development in the context of the IM model marks the state of the art of full-likelihood methods. However, they require that at least parts of the likelihood can be computed analytically. Moreover, correct tuning of MCMC methods may be demanding and time consuming (Kuhner 2009). Incorporation of recombination, more complex population histories or selection pose a challenge to likelihood-based MCMC approaches. In these cases, methods based on summary statistics offer an alternative (Hey and Machado 2003). Summaries of the joint site-frequency spectrum (Wakeley and Hey 1997) have been used in conjunction with MCMC to jointly infer parameters of an IM model accounting for intra-locus recombination (Becquet and Przeworski 2007, 2009; Tellier et al. 2011; Naduvilezhath et al. 2011). A limitation of the IM model is that it assumes constant effective deme sizes, which is not justified in the case we will study here.

A more flexible, but less rigorous framework for inference under complex models without the need of explicitly computing likelihoods is offered by approximate Bayesian computation (ABC; Beaumont 2010). ABC methods i) combine Monte Carlo simulations with a rejection algorithm (Tavaré et al. 1997), ii) allow for some tolerance when rejecting (Fu and Li 1997; Weiss and von Haeseler 1998), and iii) use summary statistics to reduce the number of dimensions (Pritchard et al. 1999, but see Sousa et al. 2009). Various extensions have been proposed to improve the efficiency of the basic ABC algorithm (Marjoram et al. 2003; Wegmann et al. 2009a; Sisson et al. 2007, 2009), to choose summary statistics well (Joyce and Marjoram 2008; Wegmann et al. 2009a; Nunes and Balding 2010; Aeschbacher et al. 2011a, or chapter 3) and to improve posterior density estimation (Beaumont et al. 2002; Blum and François 2010; Leuenberger and Wegmann 2010). One of the main challenges in ABC is the so called curse of dimensionality, which results from the fact that a limited number of simulations is used for rejection in a high-dimensional space (Beaumont 2010). The problem arises when there are many parameters to be estimated jointly, and hence many summary statistics on which to condition. The challenges of inference in problems with high dimensionality are not specific to ABC, however (*e.g.* Hey 2010). ABC has, for instance, been used to compare models of human expansion (Fagundes et al. 2007; Blum and Jakobsson 2011), to infer sex-specific migration in rodents (Hamilton et al.

2005; Wegmann et al. 2010) and to show unidirectional gene flow in chimpanzees (Wegmann and Excoffier 2010).

Models in evolutionary genetics often include sets of parameters for different processes or units of the system, and have a hierarchical structure. If the data reflect this hierarchy, hierarchical Bayes models (HBM) provide an obvious choice (Gelman et al. 2004). In HBM, the distributions of parameters associated with statistical units on a given level are specified in terms of hyperparameters on a higher statistical level. A key feature of HBM is that statistical 'strength' is borrowed across units, which means that estimation of unit-specific parameters is improved by using the same data multiple times. The efficiency of rejection sampling methods can therefore be substantially improved. HBM provide a compromise between either assuming that all statistical units are the same or estimating separate sets of parameters for each unit. In practice, the former may fit the data poorly, while the latter is prone to overfitting.

Recently, Bazin et al. (2010) have proposed an approach for inference with ABC under HBM. They suggested estimating the hyperparameters in a first step, marginal to the parameters on the lower level, and to then infer the remaining parameters conditional on the hyperparameters in a second step. Inference is conditioned on data on the respective levels. The advantage of this two-step procedure is that less memory is needed for storing intermediate values of summary statistics. Yet, as the authors pointed out, it introduces an approximation to the true posterior that goes beyond the usual approximation inherent to ABC. The model we study here is not truly a HBM, because the parameters on the top level (ancestral mutation rate, male mating skew) are not hyperparameters of those on the lower level (migration rates). Nevertheless, the two-step procedure of Bazin et al. (2010) has inspired the approach we are taking here and in chapter 3. In the latter, we have estimated the scaled mutation rate in the ancestral population and the extent of male mating skew as two global parameters. In the current paper, we focus on estimating the strength and direction of migration conditional on the previously inferred global parameters.

An immediate question is what biological entities should be chosen as statistically independent units. One extreme is to consider pairs of demes as basal units (Hoelzel et al. 2007; Lucas et al. 2009). But these pairs are not necessarily independent with respect to migration. On the other extreme, it may be necessary to consider the whole set of demes (and migration rates) jointly. From a statistical perspective, having more, but smaller units, is preferable. From a biological perspective, the optimal choice will often be somewhere between, depending on the connectivity of demes. We compare two alternative choices: a) clusters of demes and b) pairs of demes. The first is justified based on the putative connectivity of the ibex demes, but it suffers from the curse of dimensionality when deme clusters are large. Choice b) reduces the curse of dimensionality, but makes the potentially wrong assumption of pairwise independence. Our results suggest that the error introduced by this assumption is compensated by the reduction of the curse of dimensionality, if the total number of migration rates to be estimated is large. We further confirm that boosting is a valid method for choosing summary statistics and reducing the number of dimensions in ABC, as was proposed in chapter 3.

4.2 Model and parameters

We study a neutral model of a spatially structured population with genetic drift, mutation and migration. The demography includes admixture, subdivision and changes in population size. This model is motivated by the re-introduction of Alpine ibex into the Swiss Alps, which was initiated in 1906. It is described in detail in chapter 3. Here, we only give a brief summary to capture the most relevant aspects. The re-introduction has been documented by game keepers and authorities, such that deme sizes could be reconstructed between 1906 and 2006, and the numbers of founder individuals transferred between demes are known (Couturier 1962; Stuwe and Nievergelt 1991; Scribner and Stuwe 1994; Maudet et al. 2002; Aeschbacher et al. 2011a, or chapter 3).

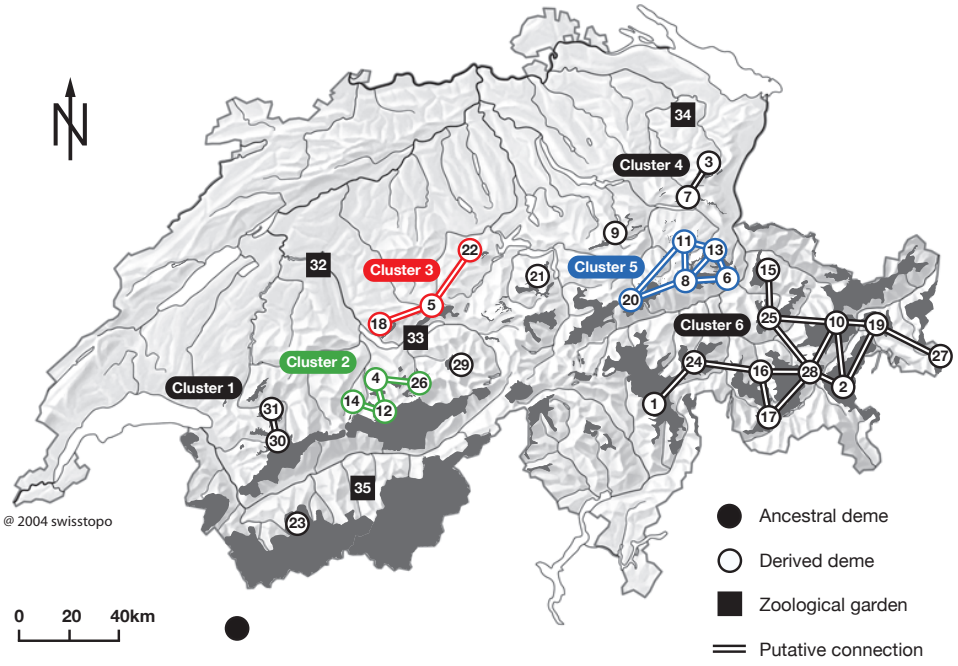

Figure 4.1: Clusters of Alpine ibex demes in the Swiss Alps that are connected by migration. The dark shaded parts represent areas inhabited by ibex. The ancestral deme is located in the *Gran Paradiso* area in Northern Italy, close to the Swiss border. The two demes in the zoological gardens 33 and 34 were first established from the ancestral one. Further demes, including the two in zoological gardens 32 and 35, were derived from demes 33 and 34. Putative connections indicate the pairs of demes for which migration is considered possible. Sets of demes connected by migration are assigned to clusters. Cluster 3 is used for testing alternative methods for choosing summary statistics. Clusters 2, 3 and 5 are used to assess the approximate, pairwise inference scheme described in the text. For deme names see Table 4.6. Map obtained via the Swiss Federal Office for the Environment (FOEN) and modified with permission.

We constructed a model that simulates genetic changes forwards in time, starting with an ancestral deme d_{anc} of unknown effective size, N_e. At times t_i $(i = 1, \ldots, I)$ further demes d_i are derived from d_{anc} or from demes established in the meantime. In our case, $t_1 = 1906$ is the year when the first deme was established in Switzerland with individuals from the ancestral deme d_{anc} in Northern Italy (Figure 4.1). From t_1 onward, the sizes of derived demes follow the

observed census size trajectories. Derived demes may exchange migrants if they are connected. Whether demes are connected depends on information obtained from FOEN, game keepers and on geography (Figure 4.1). For any pair of connected demes d_i and d_j, we define the forward migration rates $\tilde{m}_{i,j}$ and $\tilde{m}_{j,i}$ as the proportion of potential emigrants in deme d_i that migrate to deme d_j per year, and vice versa. We assume that $\tilde{m}_{i,j}$ is constant over time and the same for females and males. Migration in Alpine ibex could be sex specific or density dependent, but so far, we do not know of any evidence for this. Since $\tilde{m}_{i,j}$ is a proportion, the actual number of emigrants from deme d_i may change over time, depending on the size of deme d_i. We use the forward definition because it is straightforward to implement in an individual-based forward simulation with overlapping generations. However, backward migration rates – more commonly found in a theoretical context – can easily be calculated from forward rates assuming that no migrants are lost. We denote the set of all migration rates by $\tilde{\boldsymbol{m}} = \{\tilde{m}_{i,j} : i \neq j,\, i \in \mathcal{J}_m,\, j \in \mathcal{J}_m\}$, where $\mathcal{J}_m$ denotes the set of all demes connected via migration to at least one other deme (Figure 4.1).

A further parameter is the scaled ancestral mutation rate $\theta_{\text{anc}} = 4N_e u$, where N_e is the long-term effective size of d_{anc} up to t_1 and u is the mutation rate per generation and locus. Since we will later consider microsatellite data, we assume the stepwise model of mutation (Ohta and Kimura 1973). In contrast to the time before t_1, we assume no mutation between t_1 and the time of genetic sampling, t_g, because this period represents only 100 years, or about twelve ibex generations. Since u may vary across loci, we employ a hierarchical model, assuming that θ_{anc} is normally distributed across loci with the hyperparameters mean $\mu_{\theta_{\text{anc}}}$ and standard deviation $\sigma_{\theta_{\text{anc}}}$ (see Aeschbacher et al. 2011a, or chapter 3). Last, the proportion of males obtaining access to matings per season is denoted by ω. This parameter is motivated by the high mating skew towards dominant males observed in Alpine ibex (Aeschbacher 1978; Stuwe and Grodinsky 1987; Scribner and Stuwe 1994; Willisch and Neuhaus 2009; Willisch et al. 2011).

As explained later, for ABC we will use summary statistics to infer the migration rates $\tilde{m}_{i,j}$. Population genetic theory suggests that genetic diversity within and differentiation among demes are affected by gene flow (Wright 1931, 1943, 1951; Weir and Cockerham 1984; Cockerham and Weir 1987; Slatkin and Barton 1989; Nath and Griffiths 1996; Neigel 1997). Statistics like gene diversity within demes (expected heterozygosity ${H_1}^{(i)}$ for deme d_i), or the standardized variance of allele frequencies among demes (fixation indices ${F_{\text{ST}}}^{(i)}$ for deme d_i, pairwise ${F_{\text{ST}}}^{(i,j)}$ for demes d_i and d_j) may be used to measure these aspects as functions of the allele frequency distribution within and across demes (but see Whitlock and McCauley (1999) for a word of caution). Figure 4.2 illustrates how these statistics change jointly over time as a function of the strength and direction of gene flow and in the absence of current mutation. Without further information on demography or temporal samples, similar patterns of genetic composition observed at a given point in time may have arisen under rather different scenarios (Nielsen and Wakeley 2001; Hey and Nielsen 2004; Hey 2010; Strasburg and Rieseberg 2010). If, as in our case, historical information such as divergence times and deme genealogies are known, these can be used to condition the inference and discriminate between alternative explanations that would otherwise be hard to distinguish. Moreover, in this setting it is not necessary to assume drift-migration equilibrium.

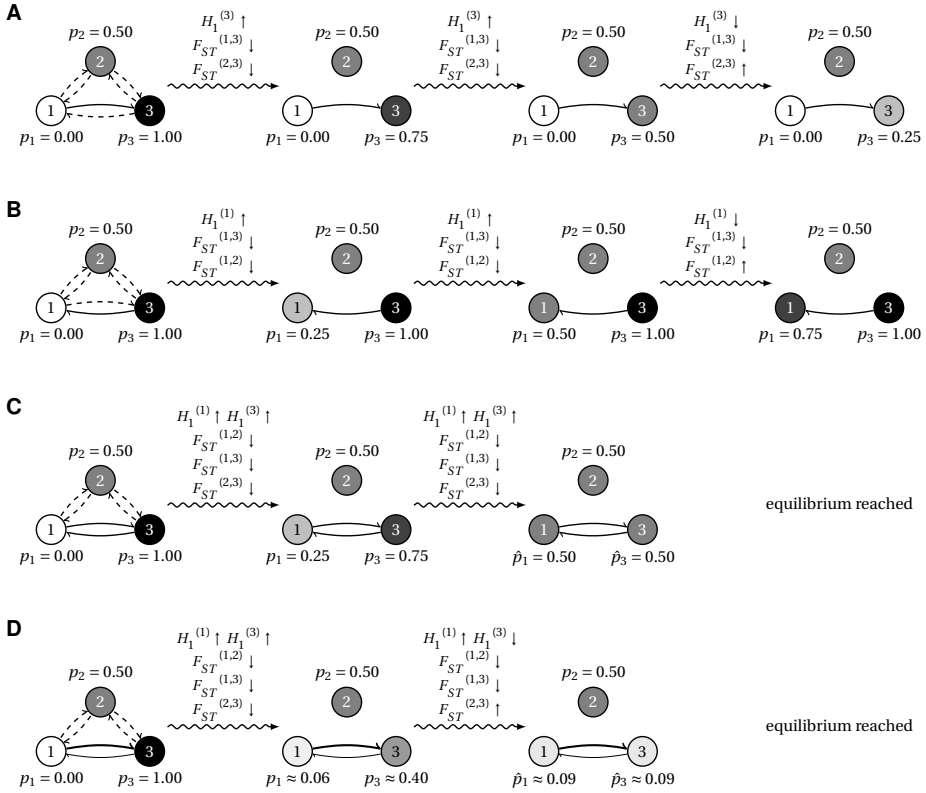

Figure 4.2: Information contained in summary statistics on strength and direction of gene flow. For simplicity, we assume three demes and two alleles, and that demes are large enough for random genetic drift to be ignored. The frequency of the first allele in deme i is p_i, with the degree of shading proportional to p_i. Dashed arrows mark potential paths of migration, solid arrows denote actual migration in each of the examples. The expected change in relevant summary statistics (see text and Table 4.1) as a function of gene flow is shown by arrows pointing up or down. (A) Constant rate of migration from deme 1 to deme 3. (B) Constant rate of migration from deme 3 to deme 1. (C) Constant and symmetric rates of migration between deme 1 and deme 3. (D) Constant, but asymmetric rates of migration between demes 1 and 3. For the numerical example, the rate from deme 1 to deme 3 is ten times the rate in the opposite direction.

4.3 Methods

4.3.1 Reducing the curse of dimensionality

We denote the joint posterior distribution of our model by $\pi(\boldsymbol{\alpha}, \tilde{\boldsymbol{m}} \mid D)$, where D represents the data, and $\boldsymbol{\alpha} = (\mu_{\theta_{\text{anc}}}, \sigma_{\theta_{\text{anc}}}, \omega)$. As pointed out in chapter 3, inferring this distribution is a complex problem due to the large number of parameters that causes a severe curse of dimensionality (Beaumont 2010). Targeting the joint posterior with ABC directly would in principle give a result, but it would be hard to assess its validity. We find it more promising to address intermediate steps and assess them one by one. For this purpose, we realize that the joint posterior may be factorized as

$$\pi(\tilde{\boldsymbol{m}}, \boldsymbol{\alpha} \mid D) = \pi(\tilde{\boldsymbol{m}} \mid \boldsymbol{\alpha}, D)\, \pi(\boldsymbol{\alpha} \mid D). \tag{4.1}$$

In practice, the two factors on the right hand side of (4.1) are individually of interest at least as much as is their product. In chapter 3, we have addressed the direct inference of $\pi(\boldsymbol{\alpha} \mid D)$ via

ABC, marginalizing implicitly over the prior of $\tilde{\boldsymbol{m}}$. For this problem, the curse of dimensionality was moderate, because $\boldsymbol{\alpha}$ comprises only three parameters. There, we have also proposed an approach for obtaining per parameter one linear combination of the original summary statistics, which reduced the number of dimensions to a minimum. These linear combinations were used as new summary statistics when doing ABC. In the current paper, we target the second part of equation (4.1): inferring migration rates conditional on D and previous knowledge of $\boldsymbol{\alpha}$. Here, the curse of dimensionality is still severe: $\tilde{\boldsymbol{m}}$ contains 56 migration rates, requiring at least the same number of summary statistics. For illustration, we may assume one summary statistic per parameter and use the – admittedly stringent – product kernel for rejection (*e.g.* Blum and Tran 2010). Accepting the ten percent closest simulations in each direction, the expected total acceptance rate would be $\epsilon = 0.1^{56}$, which is ridiculously low. In other words, to obtain a reasonable overall acceptance rate – say $\epsilon = 0.1$ – we would need to accept $\sqrt[56]{\epsilon} \approx 96\%$ of simulations in each direction, which comes close to not conditioning on any individual statistic at all. This example is hypothetical: on the one hand, there may well be more than just one summary statistic per parameter; on the other hand, a less stringent rejection kernel would alleviate the problem. Overall, it reveals the need for a strategy to avoid too severe a curse of dimensionality.

From a statistical perspective, a potential solution is to split the full system into (approximately) independent units and analyze them one by one. Such a 'divide and conquer' strategy may indeed also be justified biologically, because the degree to which the genetic composition of individual demes is correlated can vary strongly. In general, such correlations arise from common ancestry of demes, exchange of migrants, and direct or indirect effects of selection. In our case, we ignore selection, but all demes share a common ancestry, and the degree of relatedness and genetic differentiation varies according to the history of re-introduction (Stuwe and Scribner 1989; Scribner and Stuwe 1994; Biebach and Keller 2009, 2010). Defining independent units with respect to the degree of shared ancestry would nevertheless be tricky, because the genealogy of the demes is so intricate (*e.g.* Biebach and Keller 2009; Aeschbacher et al. 2011a, or chapter 3). However, based on geography and knowledge of game keepers, it is relatively straightforward to define subsets of demes that are independent with respect to migration. Since we are interested in migration rates, and since – if present – migration has a more immediate effect on current diversity than shared ancestry, grouping demes according to connectivity seems justified. We call these putatively independent sets of connected demes *deme clusters* and denote them by $\mathcal{C}_\kappa$ ($\kappa = 1, \ldots, K$), where K is the number of deme clusters (see Figure 4.1). Further, we assemble all migration rates associated with cluster $\mathcal{C}_\kappa$ into $\tilde{\boldsymbol{m}}_\kappa = \{\tilde{m}_{i,j} : i \neq j,\, i \in \mathcal{J}_{\mathcal{C}_\kappa},\, j \in \mathcal{J}_{\mathcal{C}_\kappa}\}$, where $\mathcal{J}_{\mathcal{C}_\kappa}$ is the set of demes belonging to deme cluster $\mathcal{C}_\kappa$.

The above assumption of independence of deme clusters implies that the joint likelihood can be factorized accordingly. If we further assume that the priors of $\tilde{\boldsymbol{m}}_\kappa$ conditional on $\boldsymbol{\alpha}$, $\pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha})$, are mutually independent, we can therefore also factorize the posterior distribution (for details, see Bazin et al. 2010). Specifically, the first term on the right hand side of (4.1) may be written as

$$\pi(\tilde{\boldsymbol{m}} \mid \boldsymbol{\alpha},\, D) = \prod_{\kappa=1}^{K} \pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha}, D_\kappa), \tag{4.2}$$

where D_κ is the data relevant for deme cluster $\mathcal{C}_\kappa$. In the context of ABC, this factorization means that the same simulations can be used multiple times for the different deme clusters, which increases overall efficiency. Moreover, the curse of dimensionality is potentially reduced, because fewer parameters need to be estimated jointly in every single step. Inserting (4.2) into (4.1), we have

$$\pi(\tilde{\boldsymbol{m}},\, \boldsymbol{\alpha} \mid D) = \left[\prod_{\kappa=1}^{K} \pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha}, D_\kappa)\right] \pi(\boldsymbol{\alpha} \mid D), \tag{4.3}$$

which focuses our interest on obtaining $\pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha}, D_\kappa)$, with $\boldsymbol{\alpha}$ drawn from $\pi(\boldsymbol{\alpha} \mid D)$. Marginalizing over $\boldsymbol{\alpha}$, we obtain the posterior of the migration rates for any given cluster $\mathcal{C}_\kappa$ as

$$\pi(\tilde{\boldsymbol{m}}_\kappa \mid D) = \int_{\mathcal{A}} \pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha}, D_\kappa)\, \pi(\boldsymbol{\alpha} \mid D)\, \mathrm{d}\boldsymbol{\alpha}, \tag{4.4}$$

with $D_\kappa \subset D$, and $\mathcal{A}$ the domain of $\boldsymbol{\alpha}$ with non-zero prior support. These are the quantities of our principal interest, and the left hand side of (4.4) may be targeted directly using ABC (see below).

If the number of demes in a cluster is large, the curse of dimensionality may still hamper inference of $\pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha}, D_\kappa)$. For instance, cluster 6 (Figure 4.1) consists of eleven demes and comprises 28 migration rates – possibly too many for joint estimation. For this reason, we consider as a further level of hierarchy *pairs of demes*. Clearly, the assumption of pairwise independence of demes with respect to migration is not justified, considering the pattern of connectivity in Figure 4.1. Yet, if the error caused by assuming pairwise independence is compensated by a gain in accuracy due to a reduced curse of dimensionality, such an assumption seems justified in practice. Whether this is the case must be established by a direct comparison of the accuracy achieved with the two approaches.

To formalize this idea, we denote pairs of demes by $\mathcal{P}_\psi$ ($\psi = 1, ..., P$), where P is the number of pairs. In analogy to $\tilde{\boldsymbol{m}}_\kappa$, we introduce $\tilde{m}_\psi = \{\tilde{m}_{i,j} : i \neq j,\, i \in \mathcal{J}_{\mathcal{P}_\psi},\, j \in \mathcal{J}_{\mathcal{P}_\psi}\}$, where $\mathcal{J}_{\mathcal{P}_\psi}$ is the set consisting of the two demes belonging to $\mathcal{P}_\psi$. Therefore, for any ψ, $\tilde{m}_\psi$ comprises just the two rates of migration in the opposite direction along the path connecting a specific deme pair. The marginal posterior for any deme pair – analogous to (4.4) for any deme cluster – is then

$$\pi(\tilde{m}_\psi \mid D) = \int_{\mathcal{A}} \pi(\tilde{m}_\psi \mid \boldsymbol{\alpha}, D_{\kappa:\, \mathcal{P}_\psi \in \mathcal{C}_\kappa})\, \pi(\boldsymbol{\alpha} \mid D)\, \mathrm{d}\boldsymbol{\alpha}, \tag{4.5}$$

where $D_{\kappa:\, \mathcal{P}_\psi \in \mathcal{C}_\kappa}$ denotes the data specific to the deme cluster that contains deme pair $\mathcal{P}_\psi$. The formal equivalent to (4.4) on the level of a given deme cluster $\mathcal{C}_\kappa$, as obtained with the pairwise method, is

$$\pi_{\mathrm{pw}}(\tilde{\boldsymbol{m}}_\kappa \mid D) = \prod_{\psi:\, \mathcal{P}_\psi \in \mathcal{C}_\kappa} \pi(\tilde{m}_\psi \mid D), \tag{4.6}$$

where the product is over all deme pairs $\mathcal{P}_\psi \in \mathcal{C}_\kappa$, and we again assume conditional independence of all priors. The main question of interest is how results obtained from empirical estimates of (4.4) and (4.6) compare. The answer is likely to depend on the number of migration rates in a cluster. For a fixed number of simulations, the more parameters are to be estimated, the better we expect the pairwise method to perform relative to the joint method. To investigate this, we compared joint versus pairwise inference for three clusters with varying number of demes: clusters 2, 3 and 5 with 4, 6 and 14 migration rates, respectively (Figure 4.1).

4.3.2 ABC procedure

In the formal description above, we have conditioned on the data D on the original scale. However, in ABC the data are usually compressed to summary statistics in order to increase the acceptance rate (Pritchard et al. 1999; Marjoram et al. 2003; Sisson et al. 2007). Ideally, summary statistics should be chosen to be Bayes sufficient, meaning that they satisfy

$$\pi(\phi \mid D) = \pi(\phi \mid \mathbf{S}(D)) \tag{4.7}$$

for all values taken by the parameter ϕ and all priors $\pi(\phi)$, where $\mathbf{S}(D)$ denotes a vector of summary statistics computed from the full data (*e.g.* Gelman et al. 2004; Bazin et al. 2010). In many population genetic applications, no sufficient statistics are known and the choice of statistics is therefore a crucial step (Joyce and Marjoram 2008; Wegmann et al. 2009a; Beaumont 2010; Nunes and Balding 2010; Aeschbacher et al. 2011a, or chapter 3). Moreover, as was pointed out by Bazin et al. (2010) in a similar context, equations (4.4) and (4.5) suggest that, given the hierarchical structure of our model, we should use two distinct types of summary statistics: (i) summary statistics that are *symmetric* with respect to the deme clusters (or pairs of demes) and functions of all demes together (*e.g.* means or variances across deme clusters), and (ii) summary statistics that are specific to individual units (deme clusters, or pairs of demes in our case). As a consequence, the requirement for sufficiency stated in equation (4.7) can be relaxed (for details, see Bazin et al. 2010). In the following, we use $\mathbf{s} = \mathbf{S}(D)$ to denote summary statistics that are computed from data D on the level of the whole population, and $\mathbf{u} = \mathbf{U}(D_{\text{unit}})$ for summary statistics computed from data specific to a given unit (deme cluster, or pair of deme). Recall that we have inferred the posterior of $\boldsymbol{\alpha}$ given D before in chapter 3. Here, we focus on inferring the posterior of the $\tilde{\boldsymbol{m}}_\kappa$ marginal to $\boldsymbol{\alpha}$, $\pi(\tilde{\boldsymbol{m}}_\kappa \mid D)$, for each κ. To obtain an ABC approximation to these posteriors with unit-specific candidate statistics $\mathbf{U}(D_{\text{unit}})$, we employed algorithm A below. The algorithm also includes a step for choosing informative summary statistics from the candidate statistics. Throughout, a prime denotes a simulated instance of a parameter or statistic.

Algorithm A:

A.1 For each deme cluster $\mathcal{C}_\kappa$ ($\kappa = 1, ..., K$): Calculate candidate summary statistics $\mathbf{u}_\kappa = \mathbf{U}(D_\kappa)$ from observed data.

A.2 For $t = 1$ to $t = N$:

- i Sample $\boldsymbol{\alpha}'_t$ from $\pi(\boldsymbol{\alpha} \mid D)$ obtained in a previous step.
- ii For $\kappa = 1$ to $\kappa = K$: Sample $\tilde{\boldsymbol{m}}'_{\kappa,t}$ from the conditional prior $\pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha} = \boldsymbol{\alpha}'_t)$.
- iii Simulate data D'_t (for *all* demes, irrespective of deme cluster), conditioning on $\boldsymbol{\alpha}'_t$ and $\tilde{\boldsymbol{m}}'_t = \{\tilde{\boldsymbol{m}}'_{\kappa,t} : \kappa = 1, ..., K\}$.
- iv For $\kappa = 1$ to $\kappa = K$: Calculate candidate summary statistics $\mathbf{u}'_{\kappa,t} = \mathbf{U}(D'_{\kappa,t})$ from data simulated in A.2.iii.

A.3 Sample without replacement $n \leq N$ simulated data points $\langle \mathbf{u}'_{1,t}, ..., \mathbf{u}'_{\kappa,t}, \boldsymbol{\alpha}'_t, \tilde{\boldsymbol{m}}'_{1,t}, ..., \tilde{\boldsymbol{m}}'_{\kappa,t} \rangle$ and use them as a training data set to choose informative sets of summary statistics, $\mathbf{U}_{\tilde{\boldsymbol{m}}_\kappa}$, one set for each κ.

A.4 For $\kappa = 1$ to $\kappa = K$:

- i According to A.3, obtain $\mathbf{u}_{\tilde{\boldsymbol{m}}_\kappa}$ from $\mathbf{u}_\kappa$.
- ii For $t = 1$ to $t = N$: According to A.3, obtain $\mathbf{u}'_{\tilde{\boldsymbol{m}}_{\kappa,t}}$ from $\mathbf{u}'_{\kappa,t}$.
- iii Scale $\mathbf{u}_{\tilde{\boldsymbol{m}}_\kappa}$ and $\mathbf{u}'_{\tilde{\boldsymbol{m}}_\kappa}$ appropriately.
- iv For $t = 1$ to $t = N$: Accept $(\tilde{\boldsymbol{m}}'_{\kappa,t},\, \boldsymbol{\alpha}'_t)$ if $\rho(\mathbf{u}'_{\tilde{\boldsymbol{m}}_\kappa,t}, \mathbf{u}_{\tilde{\boldsymbol{m}}_\kappa}) \leq \delta_\epsilon$, using scaled summary statistics from A.4.iii
- v Estimate the posterior density $\pi(\tilde{\boldsymbol{m}}_\kappa, \boldsymbol{\alpha} \mid D) \approx \pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha}, D)\, \pi(\boldsymbol{\alpha} \mid D)$ from the ϵN accepted points $\langle \mathbf{u}'_{\tilde{\boldsymbol{m}}_\kappa,t},\, \boldsymbol{\alpha}'_t,\, \tilde{\boldsymbol{m}}'_{\kappa,t} \rangle$.

The quantities of interest – the posteriors of $\tilde{\boldsymbol{m}}_\kappa$ marginal to $\boldsymbol{\alpha}$ for any κ as shown in (4.4) – are then approximated by simply discarding $\boldsymbol{\alpha}$ in the results obtained in step A.4.v. Notice the difference between $\mathbf{u}_\kappa = \mathbf{U}(D_\kappa)$ and $\mathbf{u}_{\tilde{\boldsymbol{m}}_\kappa} = \mathbf{U}_{\tilde{\boldsymbol{m}}_\kappa}(D_\kappa)$: the former denotes values of *candidate* summary statistics computed from data of deme cluster $\mathcal{C}_\kappa$; the latter refers to values of summary statistics *chosen to be informative about* $\tilde{\boldsymbol{m}}_\kappa$ – either as a subset of $\mathbf{U}$ or some function of its components – also computed from data of $\mathcal{C}_\kappa$. For step A.2 we performed $N = 10^6$ simulations. In A.2.i, we sampled from the posterior distribution $\pi(\boldsymbol{\alpha} \mid \mathbf{s}_{\boldsymbol{\alpha}}) \approx \pi(\boldsymbol{\alpha} \mid D)$ inferred in chapter 3, where $\mathbf{s}_{\boldsymbol{\alpha}} = \mathbf{S}_{\boldsymbol{\alpha}}(D)$ were chosen to be informative about $\boldsymbol{\alpha}$. These statistics were chosen from a set of candidate statistics $\mathbf{S}$ via L_2-Boosting in the putative vicinity of the observation $\mathbf{s} = \mathbf{S}(D)$, as described in chapter 3. Further, the conditional prior in step A2.ii was assumed to be equal to the unconditional one, that is $\pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha}) = \pi(\tilde{\boldsymbol{m}}_\kappa) = \prod_{i,j} \pi(\tilde{m}_{i,j})$, where the product is over all i, j such that $i \neq j$, $i \in \mathcal{J}_{\mathcal{C}_\kappa}$ and $j \in \mathcal{J}_{\mathcal{C}_\kappa}$. Figure 4.7 suggests that this assumption is justified: It shows that the distribution of $\tilde{m}_{i,j}$ values belonging to simulated data points that were accepted in chapter 3 when inferring $\boldsymbol{\alpha}$ does not deviate from the original $\log_{10}$ uniform prior of the migration rates. This means that the summary statistics $\mathbf{s}_{\boldsymbol{\alpha}}$ used for inferring $\boldsymbol{\alpha}$ in chapter 3 were not informative about the $\tilde{m}_{i,j}$. For the choice of statistics in A.3, we compared a set of methods described in chapter 3 in terms of their accuracy. We restricted this comparison to deme set 3 (Figure 4.1). The methods compared are partial least squares (PLS) regression as suggested by Wegmann et al. (2009a), and three versions of boosting with different loss functions (L_1-, L_2- and logistic loss). For all methods, both a global (focussing on the whole prior range) and a local (focussing on the putative vicinity of the true value only) version was employed. For details and references to alternative approaches, see chapter 3. In A.4.iii, we mean-centered the summary statistics and scaled them to have unit variance, and in A.4.iv, we chose the Euclidean distance as metric $\rho(\cdot)$ (Beaumont et al. 2002). There, δ_ϵ is the threshold chosen such that a proportion of ϵ of the N simulations is accepted. In A.4.v, we performed post-rejection adjustment with a weighted local-linear regression using weights from an Epanechnikov kernel (Beaumont et al. 2002), without additional scaling of parameters. For step A.4 we used the `abc` package (Csilléry et al. 2011) for `R` (R Development Core Team 2011).

Similarly, to obtain an approximation of (4.5) for each deme pair, we used algorithm B (see Appendix), which is essentially obtained from algorithm A by replacing deme cluster $\mathcal{C}_\kappa$ by deme pair $\mathcal{P}_\psi$, κ by ψ, and $\tilde{\boldsymbol{m}}_\kappa$ by $\tilde{m}_\psi$ everywhere except for one subtle difference: In step B.1, even though iteration is over *pairs* of demes, we computed, for each pair, summary statistics from data of the whole respective *cluster* of demes, $\mathbf{u}_\kappa = \mathbf{U}(D_{\kappa:\, \mathcal{P}_\psi \in \mathcal{C}_\kappa})$. The notation is that κ identifies the deme cluster that contains deme pair $\mathcal{P}_\psi$. Accordingly, in B.3 informative statistics

$\mathbf{U}_{\tilde{m}_\psi}$ were chosen from the candidate statistics $\mathbf{U}$. Therefore, while migration rates were estimated independently in pairs, data from the whole corresponding deme cluster were used for each pair. This is important, because demes other than those connected by a given path of migration may also convey information on the rate along that path, while focussing exclusively on data from a pair of demes may be misleading (Figure 4.2; Slatkin 1993; Whitlock and McCauley 1999). Further details mentioned above, following algorithm A, apply analogously to the procedure for pairs of demes.

The results from algorithms A and B represent approximations to $\pi(\tilde{\boldsymbol{m}}_\kappa \mid \boldsymbol{\alpha}, D_\kappa)\,\pi(\boldsymbol{\alpha} \mid D) = \pi(\tilde{\boldsymbol{m}}_\kappa, \boldsymbol{\alpha} \mid D)$ and $\pi(\tilde{m}_\psi \mid \boldsymbol{\alpha}, D_{\kappa:\,\mathcal{P}_\psi \in \mathcal{C}_\kappa})\,\pi(\boldsymbol{\alpha} \mid D) = \pi(\tilde{m}_\psi, \boldsymbol{\alpha} \mid D)$ that go beyond the usual approximation inherent to ABC. The exact explanation is somewhat subtle and we refer to Bazin et al. (2010) for details. The essence is that we are conditioning twice on the data D_κ associated with deme cluster $\mathcal{C}_\kappa$ (or deme pair $\mathcal{P}_\psi$): once when inferring $\boldsymbol{\alpha}$ conditioning on $\mathbf{S}_{\boldsymbol{\alpha}}(D)$ – as done in chapter 3 – and a second time when conditioning on $\mathbf{U}_{\tilde{\boldsymbol{m}}_\kappa}(D_\kappa)$ in algorithm A and $\mathbf{U}_{\tilde{m}_\psi}(D_{\kappa:\,\mathcal{P}_\psi \in \mathcal{C}_\kappa})$ in algorithm B, respectively. The deviation is expected to be small if the number of independent units (deme clusters in the case of algorithm A; pairs of demes in algorithm B) is large, such that the effect of any single unit on the total data is negligible.

It is important to notice that even in the case of joint estimation of all migration rates within a cluster, we focused mainly on *marginal* posterior distributions with respect to the other migration rates. Therefore, when later we report point and interval estimates and coverage properties, these are marginal with respect to the other migration rates.

The set $\mathbf{U}$ of candidate summary statistics we used is given in Table 4.1. For the migration rates, we chose a uniform prior on the $\log_{10}$ scale: $\tilde{m}_{i,j} \sim \log_{10}$-uniform in $\left[10^{-3.5}, 10^{-0.5}\right]$. On the untransformed scale, the limits correspond to about $3.2 \cdot 10^{-4}$ and 0.32, respectively, and therefore range from essentially zero migration to a rate that seems very high for Alpine ibex. This choice of prior imposes a strong belief in low migration rates; a substantial increase in the likelihood is needed to raise an estimate from $10^{-3.5}$ to $10^{-0.5}$. Such a choice nevertheless seems justified, given the potentially high degree of isolation imposed by geographic barriers such as deep valleys, rivers and roads. Notice, however, that $\tilde{m}_{i,j} = 0$ is not included in our prior distribution.

Table 4.1: Candidate summary statistics $\mathbf{U}$

Symbol	Description	Number	Reference
${H_1}^{(i)}$	MAL[a] of within-deme gene diversity in deme d_i	31	NC1983[c]
${F_{IS}}^{(i)}$	MAL of F_{IS} in deme d_i	31	NC1983
${F_{ST}}^{(i)}$	MAL of F_{ST} in deme d_i	31	NC1983
${S_2}^{(i,j)}$	MAL of between-deme MSD[b] in allele length for deme pair (i,j)	465	S1995[d]
${F_{ST}}^{(i,j)}$	MAL of pairwise F_{ST} for deme pair (i,j)	465	NC1983

The column 'Number' refers to the number of times this statistic occurs in the whole data set.
[a] Mean across loci.
[b] Mean squared difference.
[c] Nei and Chesser (1983).
[d] Slatkin (1995).

4.3.3 Simulation study and assessment of performance

To assess different methods for choosing summary statistics and compare the joint with the pairwise estimation procedure, we carried out a simulation study. For each $\epsilon \in \{0.001, 0.01, 0.1\}$, we simulated 500 test data sets with $\tilde{\boldsymbol{m}}$ sampled from the prior distribution and $\boldsymbol{\alpha}$ drawn from $\pi(\boldsymbol{\alpha} \mid D)$ inferred previously in chapter 3. We then estimated marginal posterior distributions for each migration rate and computed measures of accuracy. Similar to Wegmann et al. (2009a), we used the root mean integrated squared error (RMISE), defined as $\text{RMISE}_k = \sqrt{\int_{\Phi^{(k)}} (\phi^{(k)} - \mu_k)^2 \, \pi(\phi^{(k)} \mid \mathbf{s}) \, \mathrm{d}\phi^{(k)}}$, where μ_k is the true value of the k^{th} component of the parameter vector $\boldsymbol{\phi}$ and $\pi(\phi^{(k)} \mid \mathbf{s})$ is the corresponding estimated marginal posterior density. Recall that $\boldsymbol{\phi} = \tilde{\boldsymbol{m}}$ in our case. From this, we obtained the relative absolute RMISE (RARMISE) as $\text{RARMISE}_k = \text{RMISE}_k / |\mu_k|$. We also computed the absolute difference (AE_k) between three marginal posterior point estimates (mode, mean and median) and μ_k. Dividing by $|\mu_k|$, we obtained the relative absolute error (RAE_k). To directly compare the various methods to ABC with all summary statistics, we computed *standardized* variants of the RMISE and AE as follows: If $a_k^{\texttt{all}}$ is the measure of accuracy for ABC with all summary statistics, and a_k^* the one for ABC with the method of interest, the standardized measure was obtained as $a_k^* / a_k^{\texttt{all}}$. As a further criterion, we assessed the coverage property of the inferred posterior distributions. For this, we checked if the posterior probabilities of the true parameter values across the 500 test data sets were uniformly distributed in $[0, 1]$ (*cf.* Wegmann et al. 2009a; Cook et al. 2006). We assessed the uniformity with a Kolmogorov-Smirnov test (Sokal and Rohlf 1981).

4.3.4 Application to Alpine ibex

For the application to Alpine ibex, we used microsatellite allele frequencies and repeat lengths as published in Biebach and Keller (2009) (*see* Figure 4.1 and Table 4.6). ABC simulations and inference were identical to those in the simulation study and implemented in a program called `SPoCS` that we wrote for this purpose. In these simulations, migration occurred between population regulation and reproduction. Females and males must have reached the age of three years before they may emigrate. For a given source deme, the total of individuals to be sent to all connected demes were put into an emigrant pool. Emigrants were then randomly distributed to the receiver demes in proportions corresponding to the emigration rates. Further details are described in the supporting information (SI) of chapter 3. `SPoCS` and a collection of scripts used for inference are available on the website `http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`. We restricted the application to real data to deme clusters 2, 3 and 5 (Figure 4.1), because the main focus of this paper is to compare the joint and pairwise estimation procedure. For a complete analysis of all demes, see chapter 3.

To assess how much our results were influenced by the data as opposed to the prior assumptions, we calculated the Kullback-Leibler divergence D_{KL} (Kullback and Leibler 1951) of the marginal posterior from the corresponding prior, using the `flexmix` package (Leisch 2004; Grün and Leisch 2007, 2008) for `R`. D_{KL} is large if the posterior differs strongly from the prior, which is an indicator for how much the posterior is driven by the data as opposed to the prior. We have also computed the Manhattan and Euclidean distances between prior and posterior

distributions, but the general pattern was very similar to the one obtained for D_{KL} (data not shown).

4.3.5 Comparison to a model without migration

So far, we have considered a model with migration and we have chosen prior distributions for $\tilde{m}_{i,j}$ with no support for $\tilde{m}_{i,j} = 0$. However, it is of interest whether the Alpine ibex data – at least for some deme pairs – are also compatible with a model without migration, and how the two models compare. ABC cannot only be used for estimating parameters of a given model, but has also been employed for model choice (*e.g.* Pritchard et al. 1999; Estoup et al. 2004; Fagundes et al. 2007; Cornuet et al. 2008; Verdu et al. 2009). Bayesian model choice proceeds via the comparison of marginal likelihoods

$$P_k(D) = \int_{\Theta_k} P(D \mid \boldsymbol{\phi}_k)\, \pi_k(\boldsymbol{\phi}_k)\, \mathrm{d}\boldsymbol{\phi}_k, \tag{4.8}$$

where k is a discrete model index and $\boldsymbol{\phi}_k$ and $\pi_k(\cdot)$ are parameters and priors for model k, respectively. This suggests the ratio of the marginal likelihoods as a criterion for model choice. That ratio is called the Bayes factor. For example,

$$B_{12}(D) = \frac{P_1(D)}{P_2(D)} \tag{4.9}$$

is the Bayes factor in favor of model $\mathcal{M} = 1$ compared to the alternative model $\mathcal{M} = 2$ (Robert et al. 2011, and references therein). By Bayes' rule, the posterior probability of a certain model l is given by

$$\pi(\mathcal{M} = l \mid D) = \frac{P_l(D)\, \pi(\mathcal{M} = l)}{\sum_k P_k(D)\, \pi(\mathcal{M} = k)}, \tag{4.10}$$

where $\pi(\mathcal{M} = k)$ is the prior probability assigned to model k. In the case where $\pi(\mathcal{M} = k)$ is the same for all k, it cancels from (4.10), and we see that the ratio of the marginal likelihoods – and hence the Bayes factor – is equal to the ratio of the corresponding posterior probabilities. Here is where ABC comes in naturally, because the average ABC acceptance rate associated with a given model is proportional to the posterior probability corresponding to that model. A necessary condition for this is that identical summary statistics, metric ρ and tolerance ϵ are used across all models. Hence, in practice, an estimate of the Bayes factor is given by the ratio of observed acceptance rates (Robert et al. 2011, and references therein). However, Robert et al. (2011) have shown that the ABC-type Bayes factor does in general *not* converge to the true Bayes factor when the number of simulations goes to infinity, except for some very special cases. In general, they differ by a factor that is equal to the ratio of two quantities that depend on the models compared. The reason is that, even if the summary statistics are sufficient for each of the models separately, in general they are not jointly sufficient with respect to the marginal likelihood *and* the model index. Conclusions drawn from ABC-type Bayes factors and those drawn from the exact Bayes factor will therefore in general not agree, and the correction factor is unknown except for a few special cases (for details, see Robert et al. 2011). Therefore, we consider ABC-type Bayes factors more an explorative tool for model comparison, rather than a robust criterion for model choice. We also performed a simulation study with known

model indices to test the power of the ABC model comparison approach in our setting. For our ABC model comparison procedure, we followed Fagundes et al. (2007). Details are given in the Appendix.

4.4 Results

4.4.1 Comparison of methods for choice of summary statistics

In the following, we summarize results from a comparison of different approaches for choosing summary statistics, obtained in a simulation study with known parameter values. For deme cluster 3 (Figure 4.1) we compared four methods for the choice of summary statistics, each in a global and a local version. Throughout, the point estimators mode, mean and median performed similarly in terms of the standardized absolute error (SAE), but the median was slightly more accurate on average (Table 4.2). The partial least squares (PLS) regression performed significantly worse in terms of the standardized absolute root mean integrated error (SARMISE) than the methods based on boosting. The latter performed similarly amongst each other, with logistic boosting (`lgb`) being most accurate. Interestingly, and in contrast to the results for mutation rate and mating skew in chapter 3, the local versions of the methods resulted in higher SARMISE compared to the global versions (Figure 4.3A).

The same trends applied to the SAE of the median, but to a lesser degree (Figure 4.3B). There was no uniform pattern across methods for the dependence of accuracy on the acceptance rate ϵ. Yet, for the methods based on boosting, the SARMISE tended to be lowest for the intermediate rate $\epsilon = 0.01$ (Figure 4.3). Notice that the differences in accuracy between the methods were small. The median values for SARMISE and SAE in table 4.2 are all very close to one, which means that their performance was similar to ABC with all summary statistics. These values do not reveal anything about absolute accuracy. Nevertheless, the error bars in figure 4.3 suggest that significant differences between methods do exist. Although the global version of logistic boosting (`lgb.glob`) resulted in most accurate point estimates on average, the corresponding coverage properties were unsatisfactory for $\epsilon \geq 0.01$ (rightmost column in Table 4.2). For all methods, the distribution of posterior probabilities of the true value deviated more from a uniform distribution with increasing ϵ. The effect was stronger for the local versions than for the global ones, except for PLS (Table 4.2). Distributions deviating from uniformity were generally left-skewed (data not shown). This means that the true value was found in the lower part of the inferred posterior distribution more often than expected and implies that these methods would overestimate migration rates. The method resulting in the best compromise between accurate point estimation and good posterior coverage was the global version of boosting with the L_2-loss (`l2b.glob`, Table 4.2). For further analyses, we therefore used `l2b.glob` for choosing summary statistics.

4.4.2 Joint versus pairwise estimation of migration rates

We have compared the joint and pairwise estimation procedure for three deme clusters of different size. The accuracy of the pairwise estimation method in terms of the SARMISE and the SAE clearly increased with the size of the deme cluster and the number of migration rates

to be inferred, *relative* to the joint method (Table 4.3). For the smallest cluster – cluster 3 with four migration rates (Figure 4.1) – pairwise estimation resulted in higher SARMISE and SAE than joint estimation. For the intermediate cluster – cluster 2 with six migration parameters – the two methods performed about equally well. For the largest cluster – cluster 5, for which 14 migration rates were to be inferred – the pairwise procedure started outcompeting the joint estimation. In addition, the pairwise estimation method resulted in much better posterior coverage than the joint method. For the latter, posterior probabilities of the true value deviated strongly from uniformity in most cases. In general, posterior coverage became worse with increasing size of the deme cluster (two rightmost columns in Table 4.3).

Table 4.2: Accuracy of different methods for choosing summary statistics, relative to ABC with all candidate summary statistics

Method[a]	ϵ	SARMISE[b]	SAE[c] mode	SAE mean	SAE median	Cov. p[d]
pls.glob	0.001	1.043 (0.074)	1.067 (0.483)	1.037 (0.387)	1.041 (0.365)	0.241
	0.01	1.042 (0.068)	1.075 (0.48)	1.037 (0.298)	1.034 (0.356)	0.093
	0.1	1.039 (0.062)	1.035 (0.443)	1.034 (0.329)	1.037 (0.354)	0.027*
lgb.glob	0.001	0.999 (0.014)	0.998 (0.157)	1.006 (0.086)	1.006 (0.095)	0.062
	0.01	0.998 (0.005)	1.001 (0.083)	1.001 (0.03)	0.999 (0.031)	<0.001*
	0.1	0.999 (0.004)	1.002 (0.039)	0.999 (0.017)	0.998 (0.019)	<0.001*
l1b.glob	0.001	1.005 (0.061)	1.012 (0.493)	1.039 (0.277)	0.998 (0.299)	0.723
	0.01	1.007 (0.056)	1.028 (0.401)	1.010 (0.238)	0.994 (0.293)	0.486
	0.1	1.009 (0.045)	1.012 (0.33)	1.009 (0.258)	0.985 (0.247)	0.102
l2b.glob	0.001	1.007 (0.059)	1.024 (0.479)	1.023 (0.272)	0.989 (0.268)	0.648
	0.01	1.005 (0.052)	1.007 (0.388)	1.006 (0.232)	0.991 (0.27)	0.413
	0.1	1.005 (0.046)	1.005 (0.349)	1.005 (0.236)	0.978 (0.237)	0.137
pls.loc	0.001	1.064 (0.087)	1.091 (0.517)	1.083 (0.394)	1.075 (0.442)	0.4
	0.01	1.059 (0.078)	1.109 (0.49)	1.074 (0.35)	1.082 (0.382)	0.263
	0.1	1.054 (0.071)	1.092 (0.403)	1.083 (0.349)	1.067 (0.413)	0.132
lgb.loc	0.001	1.000 (0.009)	1.000 (0.101)	1.000 (0.058)	1.000 (0.056)	0.043*
	0.01	1.000 (0.005)	1.000 (0.064)	1.000 (0.027)	1.000 (0.028)	<0.001*
	0.1	1.000 (0.003)	1.000 (0.035)	1.000 (0.016)	1.000 (0.017)	<0.001*
l1b.loc	0.001	1.022 (0.068)	1.004 (0.456)	1.044 (0.314)	1.028 (0.341)	0.078
	0.01	1.020 (0.053)	1.067 (0.445)	1.037 (0.269)	1.011 (0.263)	0.031*
	0.1	1.019 (0.046)	1.054 (0.367)	1.021 (0.252)	1.027 (0.264)	0.002*
l2b.loc	0.001	1.017 (0.062)	1.037 (0.463)	1.046 (0.306)	1.035 (0.317)	0.078
	0.01	1.015 (0.051)	1.047 (0.396)	1.030 (0.259)	1.021 (0.259)	0.087
	0.1	1.018 (0.044)	1.043 (0.362)	1.027 (0.24)	1.012 (0.258)	0.005*

The table shows results for rates of migration between demes of cluster 3 (Figure 4.1). SARMISE and SAE (see below) are given as the median across 500 independent estimations with true values drawn from the prior (median absolute deviation in parentheses). For each test set, we computed the geometric mean of the measures of accuracy across parameters before averaging across test sets (*cf.* Tables 4.7 and 4.8). Migration rates were estimated on the $\log_{10}$ scale.

[a] `pls`, partial least squares regression (PLS) with the first five comonents used as statistics; `lgb`, logistic boosting; `l1b`, boosting with L_1 loss; `l2b`, boosting with L_2 loss; `glob`, global version; `loc`, local version (for details, see text and Aeschbacher et al. 2011a).

[b] Standardized absolute root mean integrated squared error.

[c] Standardized absolute error with respect to the true value.

[d] P-value from a Kolmogorov-Smirnov test for the uniformity of the posterior probabilities of the true values ($*$: $p < 0.05$).

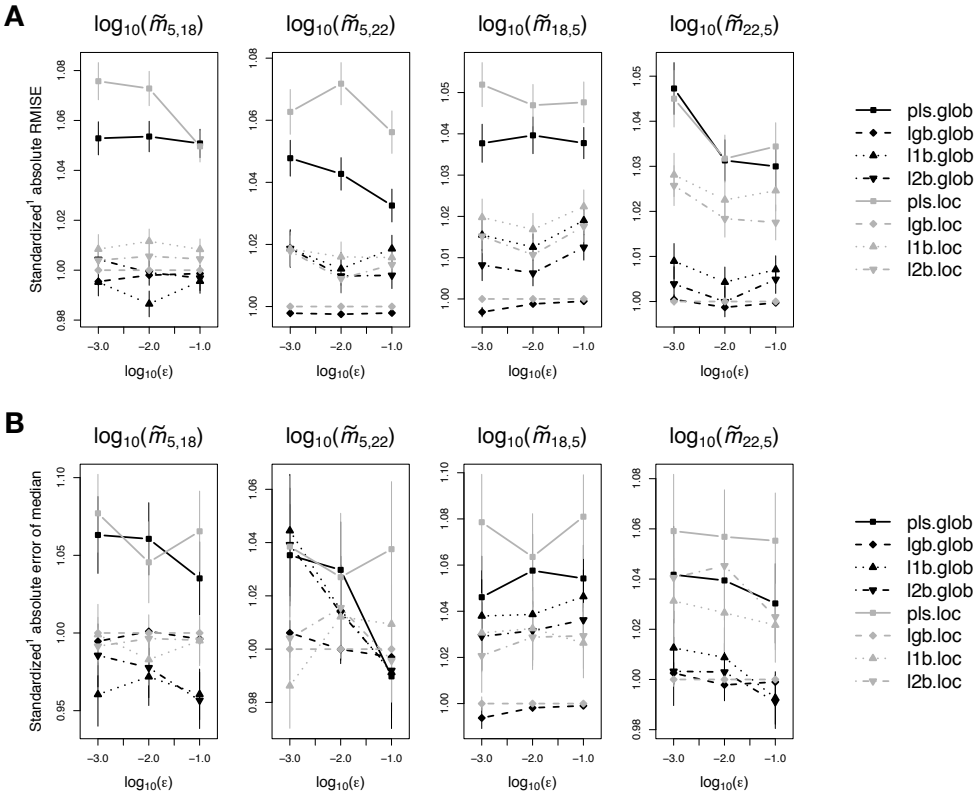

Figure 4.3: Standardized accuracy of different methods for choosing summary statistics as a function of the acceptance rate (ϵ) for deme cluster 3. [1]Standaridized means that, before averaging across test sets, we divided the measures of accuracy for the respective method by the measure of accuracy obtained with all candidate summary statistics. (A) Root mean integrated squared error (RMISE), relative to the RMISE obtained with all summary statistics. (B) Absolute error of the posterior median, relative to the one obtained with all summary statistics. Plotted are the medians across $n = 500$ independent test estimations with true values drawn from the prior (error bars denote the $\mathrm{median} \pm \mathrm{MAD}/\sqrt{n}$, where MAD is the median absolute deviation). For typical values (geometric means) across parameters, see Table 4.2.

The above results originate from a relative comparison of two approaches, but they do not reveal in an intuitive way how accurate the inferred migration rates really were. For this, we plotted the distribution of the ratio of point estimates (posterior median) to true values (Figure 4.4 for deme cluster 3 and Figures 4.8 and 4.9 for culsters 2 and 5). The rates were brought to the raw scale before the ratio was computed. As expected, the distribution is centered around a ratio of 1:1 in all cases, implying that on average the estimates were unbiased. However, the tails of the distribution reach to hundredfold under- or overestimation, with a slight skew towards overestimation (*e.g.* Figure 4.4). Comparing the joint and the pairwise estimation method, we found a slight tendency for the ratio to be closer to 1 for the pairwise method with increasing size of the deme cluster (compare Figures 4.4, 4.8 and 4.9).

The accuracy of point estimates also depended on the true value. The relation between estimated and true value was approximately linear with an expected slope of 1 only in the center of the prior distribution. True values in the lower range of the prior distribution were often overestimated, while true values in the upper range were underestimated (Figure 4.5 for deme cluster 3 and Figures 4.10 and 4.11 for culsters 2 and 5). This may be an effect of the

Table 4.3: Accuracy of pairwise estimation of migration rates relative to joint estimation per cluster, for deme clusters of different

Cluster[a]	ϵ	SARMISE[b]	SAE[c] mode	SAE mean	SAE median	Cov. p_j[d]	Cov. p_{pw}[e]
3 [4]	0.001	1.193 (0.195)	1.271 (0.749)	1.230 (0.612)	1.239 (0.619)	0.2	0.759
3 [4]	0.01	1.186 (0.173)	1.253 (0.77)	1.228 (0.591)	1.222 (0.609)	0.134	0.64
3 [4]	0.1	1.158 (0.152)	1.286 (0.755)	1.172 (0.597)	1.206 (0.619)	0.002*	0.199
2 [6]	0.001	1.088 (0.19)	1.090 (0.747)	1.041 (0.641)	1.028 (0.64)	0.018*	0.453
2 [6]	0.01	1.074 (0.171)	1.032 (0.697)	1.039 (0.599)	1.021 (0.646)	0.034*	0.418
2 [6]	0.1	1.048 (0.158)	0.983 (0.673)	1.024 (0.606)	0.987 (0.609)	<0.001*	0.159
5 [14]	0.001	0.934 (0.175)	0.804 (0.513)	0.780 (0.441)	0.784 (0.442)	0.021*	0.122
5 [14]	0.01	0.928 (0.161)	0.794 (0.58)	0.826 (0.452)	0.797 (0.427)	<0.001*	0.08
5 [14]	0.1	0.926 (0.154)	0.786 (0.569)	0.826 (0.459)	0.806 (0.444)	<0.001*	0.004*

SARMISE and SAE (see below) are given as the median across 500 independent estimations with true values drawn from the prior (median absolute deviation in parentheses). For each test set, we computed the geometric mean of the measures of accuracy accros parameters before averaging across test sets. Migration rates were estimated on the $\log_{10}$ scale, summary statistics chosen with the `l2b.glob` method, and $\epsilon = 0.01$. For parameter-specific values see Tables 4.9 to 4.11.

[a] ID of deme cluster as shown in Figure 4.1; the corresponding number of migration rates given is in brackets.

[b] Standardized absolute root mean integrated squared error with respect to the joint estimate.

[c] Standardized absolute error of the pairwise estimate with respect to the joint estimate.

[d] P-value from a Kolmogorov-Smirnov test for the uniformity of posterior probabilities of the true values ($*$: $p < 0.05$), for the joint estimation procedure.

[e] As in [d], but for the pairwise estimation procedure.

log-uniform prior in combination with the non-zero rejection tolerance of the ABC algorithm, causing biased point estimates to pile up at the sharp boundaries of the prior. Moreover, even in cases where the observed slope was close to 1 over the whole prior range (*e.g.* for $\tilde{m}_{5,18}$ in Figure 4.5), only about a fraction of $R^2 \approx 0.7^2 \approx 0.5$ of the total variance was explained (R is the Pearson product-moment correlation coefficient). There was no obvious difference in these patterns between the joint and pairwise estimation method.

4.4.3 Estimates for Alpine ibex and comparison to model without migration

Results from the simulation study above suggest that – for the model and set of methods considered here – summary statistics are best chosen via boosting with the L_2-loss, the preferable

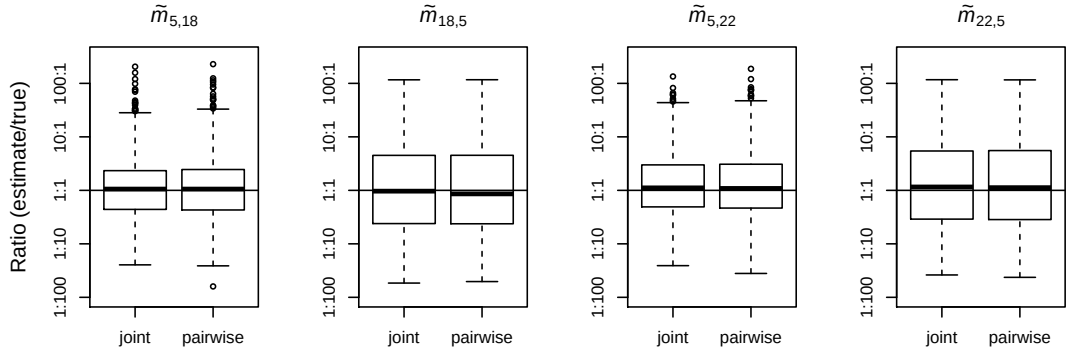

Figure 4.4: Ratio of posterior point estimate (median) to true value for the joint and pairwise estimation method. Box plots summarize data from 500 test data sets with true values sampled from the prior. Boxes show the interquartile range and whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. Note the logarithmic scale. As an example, the four parameters belonging to cluster 3 are shown (see Figures 4.8 and 4.9 for clusters 2 and 5).

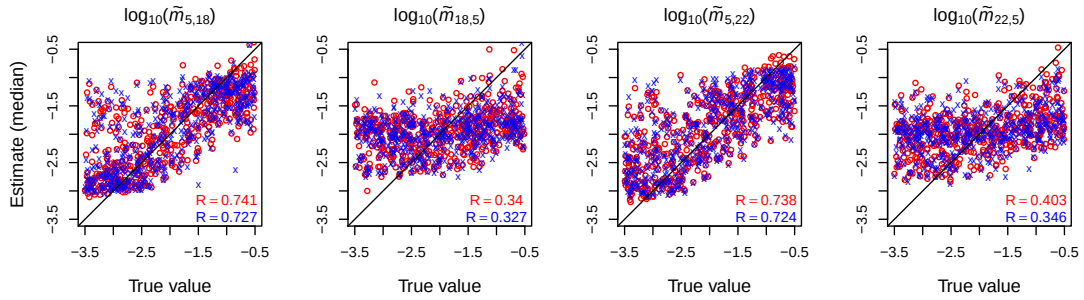

Figure 4.5: Correlation of posterior point estimate and true value for the joint (red circles) and pairwise (blue crosses) estimation method across 500 test data sets. The black line shows the expected ratio of 1:1 and R is the Pearson product-moment correlation coefficient. Plots are shown for the four parameters belonging to cluster 3 (see Figures 4.10 and 4.11 for clusters 2 and 5).

acceptance rate is $\epsilon = 0.01$, and pairwise estimation of migration rates outcompetes joint estimation as the total number of migration rates to be estimated increases. Applying this to the Alpine ibex data from deme clusters 2, 3 and 5 (Figure 4.1), we obtained posterior distributions of migration rates that are summarized in Table 4.4. For comparison, we also give the results from the joint estimation procedure. Before interpreting these estimates, we first wanted to know for which pairs of demes the model with migration (mig) has decisively more support than a model without migration (nomig). The probability that mig is the true model given the data and given the power of the ABC model comparison procedure, $P[\text{mig} \mid p_{\text{mig}}]$, was high ($> 0.95$) for three pairs of demes, (12, 14), (8, 11) and (8, 13), and marginally higher than 0.5 in the case of deme pair (6, 13). In addition, the ABC-type Bayes factor suggested weak support for mig for deme pairs (4, 12) and (11, 13), but given the conceptual difficulties that come with the ABC-type Bayes factor (see Methods), we give more importance to $P[\text{mig} \mid p_{\text{mig}}]$. A more detailed record of the model comparison procedure is provided in Figures 4.12 and 4.13.

Referring to Table 4.6 for the deme names, we conclude that there is evidence for migration for deme pairs (Gastern, Gross Lohner), (Crap da Flem, Foostock) and (Crap da Flem, Graue Hörner). For these deme pairs, the parameter estimates obtained with the pairwise estimation approach and given on the $\log_{10}$ scale in Table 4.4 translate into the following values on the untransformed scale: $\hat{\tilde{m}}_{12,14} \approx 0.086$ with a highest posterior density (HPD) interval of (0.015, 0.489); $\hat{\tilde{m}}_{14,12} \approx 0.005$ (< 0.001, 0.124); $\hat{\tilde{m}}_{8,11} \approx 0.003$ (<0.001, 0.040); $\hat{\tilde{m}}_{11,8} \approx 0.009$ (< 0.001, 0.229); $\hat{\tilde{m}}_{8,13} \approx 0.004$ (< 0.001, 0.053); and $\hat{\tilde{m}}_{13,8} \approx 0.005$ (< 0.001, 0.142). Recall that $\tilde{m}_{i,j}$ is the *annual* emigration rate from deme i to deme j. The point estimates above seem very low, except for $\tilde{m}_{12,14}$, and the HPD intervals are large for $\tilde{m}_{12,14}$ and very large for the other rates. This impression is confirmed by the Kullback-Leibler divergences D_{KL} of the posterior distributions from their corresponding priors (Table 4.4). Among migration rates that belong to deme pairs for which the migration model is justified, D_{KL} associated with $\tilde{m}_{12,14}$ is clearly the largest. Figure 4.6 illustrates this and also gives the joint posterior distributions for the migration rates connecting the two demes in a pair. Overall, there is support for a model with migration for three deme pairs, but only one of the migration rates ($\tilde{m}_{12,14}$) is associated with a posterior distribution that clearly differs from the prior distribution. There is evidence for an annual rate of migration from deme Gastern to deme Gross Lohner of about 0.09, but little

support by the data for migration in the opposite direction, as suggested by a posterior of $\tilde{m}_{14,12}$ that is very close to its prior.

4.5 Discussion

In the context of multiple populations, a central question is whether the full data are needed for accurate estimation of parameters or if subsets of the data, even pairs of demes, provide enough information (Hoelzel et al. 2007; Lucas et al. 2009; Hey 2010). The main goal of this study was to assess if estimating migration rates independently for subsets of demes is a valid strategy when multiple migration rates are to be estimated, and when discrete demes can be defined *a priori*. We have compared joint estimation per deme cluster to separate estimation for each pair of demes, using an approximate Bayesian computation (ABC) approach. The intuition was that pairwise estimation would reduce the curse of dimensionality and therefore increase the efficiency of ABC. At the same time, it was not obvious to what extent the assumption of pairwise independence would counteract this potential gain by decreasing the accuracy. We had speculated that the trade-off would depend on the number of migration rates to be estimated jointly, and hence on the size of a deme cluster.

Table 4.4: Point and interval estimates of migration rates for Alpine ibex data

		Joint estimation			Pairwise estimation		
Cluster	Param.	Median[a]	95% HPDI[b]	D_{KL}[c]	Median	95% HPDI	D_{KL}
3	$\tilde{m}_{5,18}$	−2.670	(−3.755, −1.519)	0.805	−2.933	(−3.517, −2.201)	0.843
	$\tilde{m}_{18,5}$	−1.804	(−3.379, −0.641)	0.220	−2.075	(−3.507, −0.798)	0.137
	$\tilde{m}_{5,22}$	−3.197	(−3.609, −2.648)	1.526	−3.100	(−3.513, −2.556)	1.414
	$\tilde{m}_{22,5}$	−2.499	(−3.478, −1.449)	0.349	−2.526	(−3.474, −1.473)	0.369
2	$\tilde{m}_{4,12}$	−1.727	(−3.095, −0.739)	0.322	−1.853	(−3.302, −0.68)	0.261
	$\tilde{m}_{12,4}$	−2.361	(−3.659, −1.168)	0.497	−2.545	(−3.525, −1.405)	0.330
	$\tilde{m}_{4,26}$	−2.782	(−3.504, −1.947)	0.674	−2.852	(−3.503, −2.074)	0.762
	$\tilde{m}_{26,4}$	−2.949	(−3.534, −2.278)	0.980	−2.882	(−3.518, −2.067)	0.703
	$\tilde{m}_{12,14}$	*−1.008*	(−2.358, −0.137)	1.198	*−1.063*	(−1.82, −0.311)	0.996
	$\tilde{m}_{14,12}$	*−1.963*	(−3.44, −0.729)	0.190	*−2.265*	(−3.555, −0.907)	0.194
5	$\tilde{m}_{6,8}$	−2.083	(−3.589, −0.833)	0.312	−2.321	(−3.484, −0.836)	0.160
	$\tilde{m}_{8,6}$	−2.546	(−3.831, −1.352)	0.857	−2.781	(−3.522, −1.992)	0.674
	$\tilde{m}_{6,13}$	−1.541	(−3.233, −0.419)	0.331	−2.125	(−3.481, −0.916)	0.149
	$\tilde{m}_{13,6}$	−2.394	(−3.496, −1.262)	0.528	−2.859	(−3.52, −2.181)	0.796
	$\tilde{m}_{8,11}$	*−2.190*	(−3.609, −0.949)	0.384	*−2.505*	(−3.541, −1.402)	0.330
	$\tilde{m}_{11,8}$	*−1.995*	(−3.571, −0.762)	0.311	*−2.031*	(−3.445, −0.641)	0.120
	$\tilde{m}_{8,13}$	*−2.000*	(−3.636, −0.791)	0.334	*−2.436*	(−3.542, −1.273)	0.276
	$\tilde{m}_{13,8}$	*−2.408*	(−3.969, −1.28)	0.774	*−2.328*	(−3.489, −0.848)	0.184
	$\tilde{m}_{8,20}$	−2.589	(−3.561, −1.557)	0.465	−2.496	(−3.541, −1.261)	0.300
	$\tilde{m}_{20,8}$	−2.936	(−3.911, −1.818)	1.418	−2.923	(−3.566, −1.915)	0.740
	$\tilde{m}_{11,13}$	−1.888	(−3.54, −0.718)	0.299	−2.210	(−3.541, −1.064)	0.180
	$\tilde{m}_{13,11}$	−2.423	(−3.677, −1.277)	0.598	−2.563	(−3.539, −1.518)	0.398
	$\tilde{m}_{11,20}$	−2.482	(−3.417, −1.456)	0.360	−2.107	(−3.422, −0.743)	0.215
	$\tilde{m}_{20,11}$	−2.795	(−3.615, −1.953)	0.730	−2.990	(−3.573, −2.273)	0.863

[a] Posterior median on $\log_{10}$ scale; *italic* if the migration model had strong support for the respective pair of demes (*cf.* Table 4.5).
[b] Highest posterior density interval.
[c] Kullback-Leibler divergence of posterior distribution from prior distribution.

Table 4.5: Comparison between the model with migration, $\mathcal{M} = \text{mig}$, and the one without, $\mathcal{M} = \text{nomig}$

Cluster	Parameter pair	p_{mig} [a]	$\hat{B}_{\text{ABC}}$ [b]		β_{mig} [c]	β_{nomig} [d]	P[mig | p_{mig}] [e]	
3	$\tilde{m}_{5,18}, \tilde{m}_{18,5}$	0.216	0.275		0.797	0.942	0.212	
	$\tilde{m}_{5,22}, \tilde{m}_{22,5}$	0.072	0.078		0.927	0.981	0.026	
2	$\tilde{m}_{4,12}, \tilde{m}_{12,4}$	0.549	1.217	·	0.700	0.901	0.458	
	$\tilde{m}_{4,26}, \tilde{m}_{26,4}$	0.155	0.183		0.816	0.956	0.068	
	$\tilde{m}_{12,14}, \tilde{m}_{14,12}$	0.999	1053.136	****	0.724	0.921	0.999	***
5	$\tilde{m}_{6,8}, \tilde{m}_{8,6}$	0.408	0.690		0.886	0.959	0.333	
	$\tilde{m}_{6,13}, \tilde{m}_{13,6}$	0.704	2.379	·	0.883	0.963	0.687	·
	$\tilde{m}_{8,11}, \tilde{m}_{11,8}$	1.000	∞	****	0.983	0.984	1.000	***
	$\tilde{m}_{8,13}, \tilde{m}_{13,8}$	0.996	266.391	****	0.939	0.974	0.999	***
	$\tilde{m}_{8,20}, \tilde{m}_{20,8}$	0.038	0.040		0.884	0.960	0.069	
	$\tilde{m}_{11,13}, \tilde{m}_{13,11}$	0.905	9.511	*	0.869	0.967	0.357	
	$\tilde{m}_{11,20}, \tilde{m}_{20,11}$	0.375	0.599		0.927	0.975	0.352	

[a] ABC approximation to posterior probability of $\mathcal{M} = \text{mig}$ given the data, $p_{\text{mig}} = \pi_{\text{ABC}}(\mathcal{M} = \text{mig} \mid \mathbf{u}_{\tilde{m}_\psi})$

[b] ABC-type Bayes factor in favor of $\mathcal{M} = \text{mig}$; classification code for $\hat{B}_{\text{ABC}}$ according to Jeffreys (1961): ' ' ($\mathcal{M} = \text{nomig}$ supported) 1 '·' (barely worth mentioning) $10^{1/2}$ '*' (substantial) 10 '**' (strong) $10^{3/2}$ '***' (very strong) 100 '****' (decisive support for $\mathcal{M} = \text{mig}$).

[c] Fraction of 1000 simulations performed under the mig model for which $\mathcal{M} = \text{mig}$ was correctly inferred as the true model

[d] Fraction of 1000 simulations performed under the nomig model for which $\mathcal{M} = \text{nomig}$ was correctly inferred as the true model

[e] Probability that $\mathcal{M} = \text{mig}$ is the true model, given the observed value of p_{mig} and given the power of the ABC model comparison procedure to correctly distinguish the two models (see text for details); classification code for P[mig | p_{mig}]: ' ' 0.5 '·' 0.8 '*' 0.9 '**' 0.95 '***'.

4.5.1 Pairwise estimation of migration rates more accurate for many parameters

Our main result is that the accuracy of the pairwise estimation method increased relative to that of joint estimation as the number of parameters increased. This supports our previous intuition. We found evidence that the curse of dimensionality annihilated the initial advantage of the joint estimation method for small numbers of parameters. Applying the method to data from Alpine ibex demes, we have inferred posterior distributions for migration rates between pairs of potentially connected demes that belong to deme clusters 2, 3 and 5 (Figure 4.1). However, when comparing the model with migration to one without migration, we found strong support for the migration model for only three deme pairs. The posterior distributions of the corresponding migration rates had a very wide highest posterior density interval and were close to their prior, except for $\tilde{m}_{12,14}$, the annual rate of emigration from deme Gastern to deme Gross Lohner. This is the only case for which we conclude strong support for gene flow via migration. In this study, we have only analyzed a subset of all possible demes, namely those from clusters of a moderate size (Figure 4.1). At least for the model studied here, our results about better performance of the pairwise method compared to the joint method encourage the extension of the analysis to the remaining migration rates, some of which belong to a large cluster of connected demes, including 28 migration rates (Figure 4.1). Such a study is currently in progress.

Moreover, we found that boosting of a linear regression with the L_2 loss function performed best for the choice of summary statistics. This confirms the conclusion from chapter 3. However, in the current paper, we found that choosing summary statistics on the global scale – over the whole range of prior support – yielded slightly more accurate results than focussing the choice on the putative neighborhood of the true parameter value. The opposite was found in chapter 3,

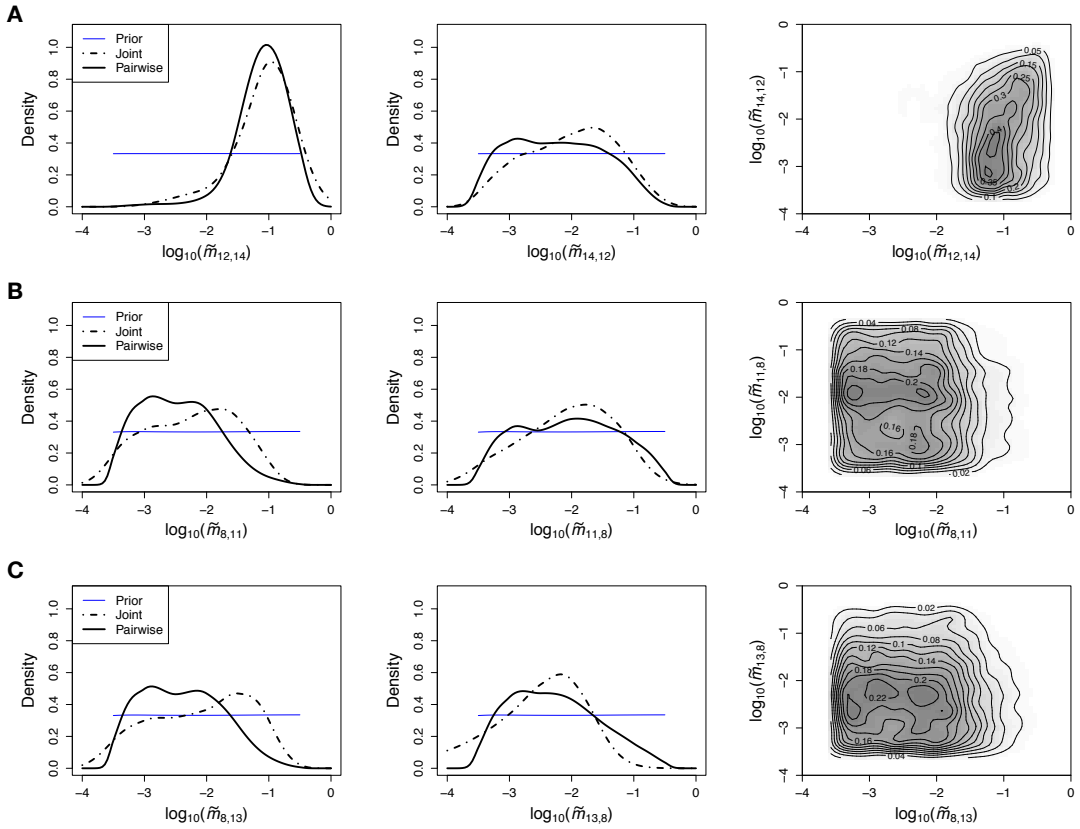

Figure 4.6: Posterior distributions of migration rates for three pairs of Alpine ibex demes, inferred with the joint (dotdashed line) and pairwise (solid line) estimation method. Each row belongs to one of the three deme pairs for which the migration model had very strong support (*cf.* Table 4.4). The first and second plot in a row give the marginal posterior for the annual emigration rates $\tilde{m}_{i,j}$ from deme i to deme j and vice versa, respectively. The third plot in a row shows the joint posterior distribution of $\tilde{m}_{i,j}$ and $\tilde{m}_{j,i}$ obtained with the pairwise method. (A) Deme pair (12, 14). (B) Deme pair (8, 11). (C) Deme pair (8, 13). See Figure 4.1 for the geographic location of the demes.

where the mean and standard deviation of the scaled mutation rate in the ancestral population and the extent of male mating skew were the parameters of interest. It is possible that whether the global or local choice of statistics is preferable depends on the parameter, its scale and the prior distribution. This needs to be explored further. In the current study, the local choice of statistics did not only result in less accurate point estimates, but also led to unsatisfactory posterior coverage, with migration rates being overestimated in general. The global methods suffered less from this.

4.5.2 Stepwise analysis of a hierarchical model

The essence of our approach was to divide the problem into (approximately) independent units, and analyse each of them separately, conditioning on previously estimated global parameters. The units in our case were sets of demes – either clusters or pairs. The global parameters on which we conditioned were the scaled ancestral mutation rate and the extent of male mating skew. This setting is reminiscent of a recent study by Bazin et al. (2010), where the units were

the different loci, and the interest was in inferring locus-specific mutation rates and selection coefficients. An important difference to Bazin et al. (2010) is that the global parameters in their case were the hyperparameters of the distribution of the locus-specific parameters. In our case, the global parameters referred to other processes (mutation, drift) than the unit-specific ones do (migration). Hence, while the setting of Bazin et al. (2010) was truly hierarchically Bayesian, ours was not. In principle, we could have introduced a hyperprior for the unit-specific migration rates. We have chosen not to do so, because it was not obvious in advance what would be the appropriate statistical units. Nevertheless, subject to some modifications, the idea of the two-step procedure proposed by Bazin et al. (2010) was relevant in our context. Specifically, in chapter 3 we had estimated the global parameters – the scaled mutation rate in the ancestral deme and the extent of male mating skew – conditioning on summary statistics computed from data across all demes, and marginal to all other parameters. For the second step, we focussed in this paper on the unit-specific parameters – the migration rates.

A crucial assumption was that the statistical units (deme clusters or pairs of demes) were independent with respect to migration. This was justified for the deme clusters simply by our definition of a cluster. Taking deme clusters as local units, for large clusters we ran into the problem of having to perform ABC with many migration rates at the same time, which was prone to the curse of dimensionality. We therefore zoomed in on a lower level, pairs of demes connected by migration. While statistical independence was now also violated with respect to migration, the number of parameters for which ABC rejection had to bee performed jointly was reduced to two. This reduced the number of summary statistics and hence dimensions.

4.5.3 Advantages and limitations

We have used uniform priors on the $\log_{10}$ scale, which has the advantage of equal probability for all values on the corresponding scale. The sharp boundaries allowed limiting the range at some threshold of choice. However, setting this range is somewhat arbitrary. Moreover, the discrete boundaries may amplify undesired effects, such as piling up of posterior mass close to the boundary, or projection of posterior density out of the prior support. With a log-uniform prior, the value $\tilde{m}_{i,j} = 0$ was not included. We therefore employed an ABC-type model comparison procedure to compare the migration model to an alternative model without migration. Although straightforward to implement, ABC-type model comparison comes with a conceptual hitch, because the summary statistics used in the compared models are in general not sufficient for model comparison. This issue remains open for further research (Robert et al. 2011).

Some highest posterior density (HPD) intervals in Table 4.4 reach beyond the prior limits. This could be an effect of the local linear regression projecting simulated parameter values out of the prior range, a problem observed and discussed before (*e.g.* Beaumont et al. 2002; Estoup et al. 2004; Leuenberger and Wegmann 2010; Beaumont 2010). One *ad hoc* solution would be to scale the parameter values prior to regression; an alternative is to discard points outside the bounded prior (Beaumont 2010). The effect might further be attenuated by using priors without sharp boundaries.

The temporal scale on which gene flow affects the genetic composition of populations is co-determined by the rates of mutation and genetic drift (Felsenstein 1982; Neigel 1997). As a

consequence, migration rates are usually scaled by either of the two (Wakeley and Hey 1997; Beerli and Felsenstein 2001; Nielsen and Wakeley 2001; Hey and Nielsen 2004). In our context, we studied migration on a very short time scale for which mutation could be ignored, and for which deme genealogy and deme sizes were known. Since we conditioned the inference of migration rates on known population history and on the previously estimated *ancestral* mutation rate (see chapter 3), there was no further need for scaling.

We have used individual-based simulations to accurately fit population history, in particular the founder events, to detailed historical records. Moreover, biological details regarding mating and reproduction, as well as overlapping generations, could be incorporated in a straightforward manner. A conceptual issue, however, arose when implementing migration. We had defined $\tilde{m}_{i,j}$ as the proportion of potential emigrants in deme d_i that migrate to deme d_j per year. If d_i is connected to more than one – say K – demes, the sum of emigrant proportions attracted by each of the receiving demes may exceed 1. Since the number of emigrants in d_i is a finite number, we had to normalize the original proportions such that they summed to 1. Therefore, a given value of $\tilde{m}_{i,j}$ may result in a varying number of emigrants, depending on the rates $\tilde{m}_{i,k}$, where $k \neq j$ and $k, j \in \{1, \ldots, K\}$. An alternative to normalizing the emigration rates would have been to define joint prior distributions that account for this constraint, leading to conditional dependence of individual priors. A related issue is that interpretation of our emigration rates is not straightforward. A given value of $\tilde{m}_{i,j}$ does not immediately translate into a number of emigrants, unless the total number of emigrants in d_i and the emigration rates into other connected demes are known. In practice, we therefore suggest running a set of simulations *a posteriori*, with migration rates equal to previously obtained point estimates or drawn from posterior distributions, and keeping track of migrant numbers. Moreover, we have ignored the effect of non-sampled demes. These exist, but to our knowledge none is likely to be connected via migration to any of the sampled demes.

The model for which our findings apply is rather specific, tailored to fit the ibex scenario. The flipside is that generalizations for other models cannot be made from our results without further investigation. However, the procedure by which we obtained our results is not restricted to this particular setting. This flexibility reflects an advantage of ABC over alternative methods. In any case, simulations and analyses need to be carried out to validate any particular application of ABC.

4.5.4 General perspective

The precision and accuracy with which parameters can be estimated depend on the information that the data contain about these, on the approach used to extract that information, and on the uncertainty about the underlying model(s). In our case, the demographic model was known. No degrees of freedom had to be sacrificed to compare alternative models. This is not usually the case (*e.g.* Takahata 1995; Fagundes et al. 2007; Blum and Jakobsson 2011). In a similar situation, Estoup et al. (2004) used demographic and geographic information on the spread of cane toad (*Bufo marinus*) to condition inference from genetic data. However, demography was not fully known, so that a set of compatible demographic models had to be compared. The indirect estimates of demographic parameters such as effective deme sizes and effective founder

sizes were reliable, while precise estimation of migration rates was problematic (Estoup et al. 2004). A similar pattern applies to recent results by Hey (2010) for the IM model with multiple demes.

In general, demography and migration have potentially confounding effects on the genetic composition of populations. Although theory by Wakeley (1996b,a) shows that for certain models (*e.g.* the IM model), DNA sequence data reveal information for differentiation, it is not obvious how this scales to more complex models. We suspect that in our case, conditioning on a known demographic model was essential. Even then, the shape of some posterior distributions implied considerable uncertainty. We do not know to what degree this was due to the lack of information in the data as opposed to the insufficiency of our ABC approach to extract it. This point raises the general question as to what extent it is possible to infer gene flow via migration at all in realistic settings. This opens perspectives for future theoretical work that will hopefully soon be verifiable with DNA sequence data from large samples in a spatial context.

4.6 Appendix

4.6.1 ABC algorithm with choice of summary statistics for pairwise method

The following algorithm was used for the pairwise estimation procedure described in the main text.

Algorithm B:

B.1 For each deme pair $\mathcal{P}_\psi$ ($\psi = 1, ..., P$): Calculate candidate summary statistics $\mathbf{u}_\kappa = \mathbf{U}(D_{\kappa:\,\mathcal{P}_\psi \in \mathcal{C}_\kappa})$ from observed data, where κ identifies the deme cluster that contains deme pair $\mathcal{P}_\psi$.

B.2 For $t = 1$ to $t = N$:

- i Sample $\boldsymbol{\alpha}'_t$ from $\pi(\boldsymbol{\alpha} \mid D)$ obtained in a previous step.
- ii For $\psi = 1$ to $\psi = P$: Sample $\tilde{m}'_{\psi,t}$ from the conditional prior $\pi(\tilde{m}_\psi \mid \boldsymbol{\alpha} = \boldsymbol{\alpha}'_t)$.
- iii Simulate data D'_t (for *all* demes, irrespective of deme cluster or deme pair), conditioning on $\boldsymbol{\alpha}'_t$ and $\tilde{\boldsymbol{m}}'_t = \{\tilde{m}'_{\psi,t} : \psi = 1, ..., P\}$.
- iv For $\psi = 1$ to $\psi = P$: Calculate candidate summary statistics $\mathbf{u}'_{\kappa,t} = \mathbf{U}(D'_{\kappa:\,\mathcal{P}_\psi \in \mathcal{C}_\kappa,\,t})$ from data simulated in B.2.iii.

B.3 Sample without replacement $n \leq N$ simulated data points $\langle \mathbf{u}'_{1,t}, ..., \mathbf{u}'_{\kappa,t}, \boldsymbol{\alpha}'_t, \tilde{\boldsymbol{m}}'_{1,t}, ..., \tilde{m}'_{\psi,t} \rangle$ and use them as a training data set to choose informative sets of summary statistics, $\mathbf{U}_{\tilde{m}_\psi}$, one set for each ψ.

B.4 For $\psi = 1$ to $\psi = P$:

- i According to B.3, obtain $\mathbf{u}_{\tilde{m}_\psi}$ from $\mathbf{u}_\kappa$.
- ii For $t = 1$ to $t = N$: According to B.3, obtain $\mathbf{u}'_{\tilde{m}_{\psi,t}}$ from $\mathbf{u}'_{\kappa,t}$.
- iii Scale $\mathbf{u}_{\tilde{m}_\psi}$ and $\mathbf{u}'_{\tilde{m}_\psi}$ appropriately.
- iv For $t = 1$ to $t = N$: Accept $(\tilde{m}'_{\psi,t}, \boldsymbol{\alpha}'_t)$ if $\rho(\mathbf{u}'_{\tilde{m}_\psi,t}, \mathbf{u}_{\tilde{m}_\psi}) \leq \delta_\epsilon$, using scaled summary statistics from B.4.iii
- v Estimate the posterior density $\pi(\tilde{m}_\psi, \boldsymbol{\alpha} \mid D) \approx \pi(\tilde{m}_\psi \mid \boldsymbol{\alpha}, D)\,\pi(\boldsymbol{\alpha} \mid D)$ from the ϵN accepted points $\langle \mathbf{u}'_{\tilde{\boldsymbol{m}}_\psi,t}, \boldsymbol{\alpha}'_t, \tilde{m}'_{\psi,t} \rangle$.

Further details are as given in the main text after algorithm A, with ψ and $\mathcal{P}_\psi$ replaced by κ and $\mathcal{C}_\kappa$, respectively. In particular, we again assume that the conditional prior in step B.2.ii is equal to the unconditional one, *i.e.* that $\pi(\tilde{m}_\psi \mid \boldsymbol{\alpha} = \boldsymbol{\alpha}'_t) = \pi(\tilde{m}_\psi)$. Figure 4.7 shows that this assumption is justified.

4.6.2 Details of ABC model comparison procedure

Proceeding essentially as proposed by Fagundes et al. (2007), we performed 10^6 simulations under both the migration (mig) and the no-migration (nomig) model, and then calculated the ABC-type Bayes factor, $\hat{B}_{\mathrm{ABC}}$, and the posterior probability of the mig model, $p_{\mathrm{mig}} = \pi_{\mathrm{ABC}}(\mathcal{M} = \mathrm{mig} \mid \mathbf{u}_{\tilde{m}_\psi})$, from the acceptance rates. We did so for each deme pair independently, using the same summary statistics $\mathbf{u}_{\tilde{m}_\psi}$ as for parameter estimation under the mig

model (see Methods). To assess the power of this procedure in recovering the true model, we simulated 1000 test data sets under each model and performed ABC model comparison for each test data set. We then calculated the proportion of times the correct model was chosen, $\beta_{\text{mig}} = P[\mathcal{M} = \text{mig inferred} \mid \mathcal{M} = \text{mig true}]$, and $\beta_{\text{nomig}} = P[\mathcal{M} = \text{nomig inferred} \mid \mathcal{M} = \text{nomig true}]$, where we considered a model as 'inferred' when its posterior probability was $p_k > 0.5$, where $k \in \{\text{mig}, \text{nomig}\}$. Recall that $p_{\text{mig}} = \pi_{\text{ABC}}(\mathcal{M} = \text{mig} \mid \mathbf{u}_{\tilde{m}_\psi})$ and, accordingly, $p_{\text{nomig}} = 1 - p_{\text{mig}}$. These considerations allowed us to compute the probability that $\mathcal{M} = \text{mig}$ is the true model, given our estimate of p_{mig} from the real data and given the power of the model comparison procedure

$$\begin{aligned} P[\mathcal{M} = \text{mig true} \mid p_{\text{mig}}] & \\ &= \frac{P[p_{\text{mig}} \mid \mathcal{M} = \text{mig true}]\ P[\mathcal{M} = \text{mig true}]}{\sum_{k \in \{\text{mig,nomig}\}} P[p_{\text{mig}} \mid \mathcal{M} = k \text{ true}]\ P[\mathcal{M} = k \text{ true}]} \\ &= \frac{P[p_{\text{mig}} \mid \mathcal{M} = \text{mig true}]}{\sum_{k \in \{\text{mig,nomig}\}} P[p_{\text{mig}} \mid \mathcal{M} = k \text{ true}]}, \end{aligned} \tag{4.11}$$

where the last equality holds if $P[\mathcal{M} = \text{mig true}] = P[\mathcal{M} = \text{nomig true}]$, which is the case if we perform the same number of test simulations under each model, as we do. The probabilities $P[p_{\text{mig}} \mid \mathcal{M} = k \text{ true}]$ are obtained from the empirical distribution of posterior model probabilities p_{mig} resulting from 1000 test simulations under each model (see Methods and Figure 4.13). We used a rejection tolerance of $\epsilon = 0.05$ and applied a logistic regression correction step to estimate the posterior model probabilities from the empirical acceptance rates (*e.g.* Fagundes et al. 2007). Throughout, we used the `abc` package (Csilléry et al. 2011) for `R`.

4.7 Supporting information: Additional tables

Table 4.6: Deme names, deme numbers and sampling sizes in the Alpine ibex data set

				Genetic sample size[c]		
Deme name	Deme no.[a]	Short name	Internal number[b]	Males	Females	Total
Adula Vial	1	AdulaVial	100	21	16	37
Albris	2	Albris	101	28	33	61
Alpstein	3	Alpstein	102	12	18	30
Bire-Oeschinen	4	BireOesch	103	16	2	18
Brienzer Rothorn	5	BrRothorn	104	21	18	39
Calanda	6	Calanda	105	15	16	31
Churfirsten	7	Churfirsten	106	11	13	24
Crap da Flem	8	CrapFlem	107	16	11	27
Fluebrig	9	Fluebrig	108	17	15	32
Flüela	10	Flüela	109	37	38	75
Foostock	11	Foostock	110	9	18	27
Gastern	12	Gastern	111	5	6	11
Graue Hörner	13	GrHörner	112	21	26	47
Gross Lohner	14	GrLohner	113	15	7	22
Hochwang	15	Hochwang	114	14	14	28
Julier Nord	16	Julier N	115	12	11	23
Julier Süd	17	Julier S	116	12	11	23
Justistal	18	Justistal	117	15	4	19
Macun	19	Macun	118	12	10	22
Oberalp-Frisal	20	Oberalp	134	25	19	44
Oberbauenstock	21	Oberbauen	119	18	12	30
Pilatus	22	Pilatus	120	15	2	17
Mont Pleureur	23	Pleureur	121	22	7	29
Safien-Rheinwald	24	Rheinwald	122	22	13	35
Rothorn-Weissfluh	25	RothWeissfl	123	16	13	29
Schwarzmönch	26	SchwMönch	124	15	17	32
Umbrail	27	Umbrail	125	15	14	29
Val Bever	28	ValBever	126	20	12	32
Wetterhorn	29	Wetterhorn	127	9	10	19
Wittenberg	30	Wittenberg	128	15	6	21
Pierreuse-Gummfluh	31	Pierreuse	133	20	21	41
Wildpark Dählhölzli	32	WPDH	129	0	0	0
Wildpark Interlaken	33	WPIH	130	0	0	0
Wildpark St. Gallen	34	WPPP	131	0	0	0
Wildpark Seiler	35	WPSE	132	0	0	0

[a] As used in main text and Figure 4.1.
[b] As used in scripts.
[c] The number of individuals from which genetic samples were taken, both in reality and in the simulations.

Table 4.7: Accuracy of different methods for choosing summary statistics on a global scale

Method	ϵ	Param.	RARMISE[a]	RAE[b] mode	RAE mean	RAE median	Cov. p[c]
all	0.001	$\tilde{m}_{5,18}$	0.426 (0.262)	0.208 (0.207)	0.185 (0.178)	0.187 (0.186)	0.61
		$\tilde{m}_{5,22}$	0.415 (0.213)	0.225 (0.191)	0.216 (0.195)	0.214 (0.192)	0.017*
		$\tilde{m}_{18,5}$	0.496 (0.216)	0.336 (0.265)	0.302 (0.245)	0.298 (0.232)	0.685
		$\tilde{m}_{22,5}$	0.491 (0.183)	0.355 (0.284)	0.298 (0.255)	0.305 (0.258)	0.121
	0.01	$\tilde{m}_{5,18}$	0.414 (0.238)	0.210 (0.198)	0.185 (0.173)	0.198 (0.186)	0.02*
		$\tilde{m}_{5,22}$	0.409 (0.207)	0.239 (0.193)	0.212 (0.202)	0.230 (0.199)	<0.001*
		$\tilde{m}_{18,5}$	0.484 (0.189)	0.368 (0.268)	0.310 (0.238)	0.302 (0.228)	0.58
		$\tilde{m}_{22,5}$	0.475 (0.168)	0.380 (0.274)	0.299 (0.257)	0.301 (0.244)	0.091
	0.1	$\tilde{m}_{5,18}$	0.414 (0.219)	0.216 (0.191)	0.193 (0.176)	0.201 (0.184)	<0.001*
		$\tilde{m}_{5,22}$	0.412 (0.193)	0.241 (0.193)	0.226 (0.206)	0.230 (0.192)	<0.001*
		$\tilde{m}_{18,5}$	0.483 (0.186)	0.386 (0.251)	0.314 (0.247)	0.307 (0.234)	0.381
		$\tilde{m}_{22,5}$	0.475 (0.161)	0.404 (0.252)	0.300 (0.249)	0.314 (0.254)	0.205
pls.glob	0.001	$\tilde{m}_{5,18}$	0.447 (0.261)	0.213 (0.207)	0.199 (0.205)	0.195 (0.197)	0.61
		$\tilde{m}_{5,22}$	0.437 (0.228)	0.231 (0.219)	0.217 (0.202)	0.212 (0.193)	0.078
		$\tilde{m}_{18,5}$	0.502 (0.193)	0.362 (0.318)	0.303 (0.243)	0.321 (0.241)	0.573
		$\tilde{m}_{22,5}$	0.490 (0.148)	0.376 (0.309)	0.320 (0.25)	0.323 (0.244)	0.241
	0.01	$\tilde{m}_{5,18}$	0.440 (0.254)	0.209 (0.203)	0.199 (0.203)	0.197 (0.202)	0.319
		$\tilde{m}_{5,22}$	0.420 (0.222)	0.240 (0.2)	0.214 (0.203)	0.219 (0.197)	0.03*
		$\tilde{m}_{18,5}$	0.494 (0.178)	0.387 (0.327)	0.307 (0.24)	0.313 (0.243)	0.617
		$\tilde{m}_{22,5}$	0.479 (0.143)	0.408 (0.315)	0.320 (0.258)	0.320 (0.255)	0.404
	0.1	$\tilde{m}_{5,18}$	0.433 (0.238)	0.217 (0.207)	0.204 (0.21)	0.214 (0.208)	0.117
		$\tilde{m}_{5,22}$	0.416 (0.204)	0.245 (0.199)	0.223 (0.206)	0.228 (0.197)	0.005*
		$\tilde{m}_{18,5}$	0.491 (0.172)	0.397 (0.308)	0.306 (0.242)	0.308 (0.232)	0.351
		$\tilde{m}_{22,5}$	0.474 (0.136)	0.419 (0.319)	0.319 (0.253)	0.319 (0.253)	0.564
lgb.glob	0.001	$\tilde{m}_{5,18}$	0.423 (0.256)	0.206 (0.202)	0.182 (0.179)	0.194 (0.188)	0.61
		$\tilde{m}_{5,22}$	0.417 (0.219)	0.228 (0.194)	0.211 (0.195)	0.215 (0.187)	0.015*
		$\tilde{m}_{18,5}$	0.493 (0.212)	0.331 (0.284)	0.308 (0.256)	0.304 (0.235)	0.648
		$\tilde{m}_{22,5}$	0.495 (0.192)	0.375 (0.289)	0.302 (0.26)	0.312 (0.256)	0.164
	0.01	$\tilde{m}_{5,18}$	0.414 (0.24)	0.206 (0.193)	0.184 (0.174)	0.196 (0.184)	0.032*
		$\tilde{m}_{5,22}$	0.406 (0.21)	0.238 (0.187)	0.213 (0.2)	0.228 (0.195)	<0.001*
		$\tilde{m}_{18,5}$	0.483 (0.192)	0.364 (0.263)	0.312 (0.244)	0.303 (0.227)	0.629
		$\tilde{m}_{22,5}$	0.478 (0.171)	0.381 (0.281)	0.299 (0.255)	0.301 (0.247)	0.121
	0.1	$\tilde{m}_{5,18}$	0.416 (0.222)	0.215 (0.184)	0.192 (0.175)	0.199 (0.184)	0.001*
		$\tilde{m}_{5,22}$	0.411 (0.195)	0.241 (0.192)	0.222 (0.207)	0.228 (0.189)	<0.001*
		$\tilde{m}_{18,5}$	0.483 (0.19)	0.382 (0.252)	0.312 (0.245)	0.307 (0.231)	0.357
		$\tilde{m}_{22,5}$	0.475 (0.16)	0.409 (0.258)	0.298 (0.251)	0.313 (0.252)	0.192
l1b.glob	0.001	$\tilde{m}_{5,18}$	0.440 (0.296)	0.192 (0.176)	0.179 (0.19)	0.175 (0.178)	0.5
		$\tilde{m}_{5,22}$	0.439 (0.247)	0.233 (0.207)	0.215 (0.201)	0.206 (0.192)	0.263
		$\tilde{m}_{18,5}$	0.506 (0.214)	0.366 (0.306)	0.305 (0.256)	0.320 (0.263)	0.685
		$\tilde{m}_{22,5}$	0.500 (0.178)	0.363 (0.315)	0.311 (0.255)	0.314 (0.254)	0.466
	0.01	$\tilde{m}_{5,18}$	0.435 (0.289)	0.206 (0.185)	0.176 (0.186)	0.174 (0.179)	0.554
		$\tilde{m}_{5,22}$	0.426 (0.233)	0.235 (0.22)	0.212 (0.199)	0.208 (0.195)	0.158
		$\tilde{m}_{18,5}$	0.496 (0.206)	0.386 (0.309)	0.309 (0.253)	0.314 (0.265)	0.828
		$\tilde{m}_{22,5}$	0.489 (0.171)	0.382 (0.322)	0.306 (0.259)	0.310 (0.252)	0.293
	0.1	$\tilde{m}_{5,18}$	0.423 (0.255)	0.208 (0.171)	0.185 (0.19)	0.184 (0.176)	0.353
		$\tilde{m}_{5,22}$	0.421 (0.213)	0.236 (0.207)	0.217 (0.2)	0.217 (0.195)	0.084
		$\tilde{m}_{18,5}$	0.495 (0.193)	0.392 (0.301)	0.309 (0.254)	0.315 (0.257)	0.42
		$\tilde{m}_{22,5}$	0.488 (0.166)	0.394 (0.288)	0.302 (0.253)	0.309 (0.245)	0.233

continued on next page

Table 4.7: Continued from previous page

Method	ϵ	Param.	RARMISE[a]	RAE[b] mode	RAE mean	RAE median	Cov. p[c]
l2b.glob	0.001	$\tilde{m}_{5,18}$	0.446 (0.301)	0.188 (0.173)	0.176 (0.188)	0.174 (0.178)	0.4
		$\tilde{m}_{5,22}$	0.434 (0.245)	0.225 (0.204)	0.207 (0.201)	0.209 (0.194)	0.314
		$\tilde{m}_{18,5}$	0.505 (0.218)	0.371 (0.297)	0.309 (0.263)	0.308 (0.256)	0.859
		$\tilde{m}_{22,5}$	0.496 (0.184)	0.336 (0.316)	0.302 (0.244)	0.300 (0.242)	0.536
	0.01	$\tilde{m}_{5,18}$	0.438 (0.293)	0.200 (0.186)	0.184 (0.19)	0.176 (0.178)	0.35
		$\tilde{m}_{5,22}$	0.425 (0.232)	0.236 (0.213)	0.213 (0.2)	0.215 (0.189)	0.245
		$\tilde{m}_{18,5}$	0.496 (0.21)	0.380 (0.314)	0.307 (0.26)	0.312 (0.26)	0.666
		$\tilde{m}_{22,5}$	0.488 (0.177)	0.368 (0.313)	0.298 (0.249)	0.291 (0.244)	0.525
	0.1	$\tilde{m}_{5,18}$	0.430 (0.265)	0.211 (0.183)	0.184 (0.196)	0.186 (0.175)	0.632
		$\tilde{m}_{5,22}$	0.417 (0.219)	0.237 (0.207)	0.215 (0.202)	0.218 (0.195)	0.075
		$\tilde{m}_{18,5}$	0.493 (0.2)	0.392 (0.304)	0.309 (0.25)	0.317 (0.256)	0.468
		$\tilde{m}_{22,5}$	0.485 (0.172)	0.388 (0.295)	0.299 (0.248)	0.303 (0.24)	0.487

The table shows results for rates of migration between demes of cluster 3 (Figure 4.1). RARMISE and RAE (see below) are given as the median across 500 independent estimations with true values drawn from the prior (median absolute deviation in parentheses). The parameters were estimated on the $\log_{10}$ scale. The indices (i, j) to the migration rates ($\tilde{m}_{i,j}$) refer to deme numbers given in Figure 4.1 and Table 4.6.

[a] Relative absolute root mean integrated squared error (see text) with respect to the true value.

[b] Relative absolute error with respect to the true value.

[c] P-value from a Kolmogorov-Smirnov test for the unifiormity of the posterior probabilities of the true values ($*$: $p < 0.05$).

Table 4.8: Accuracy of different methods for choosing summary statistics on a local scale

Method	ϵ	Param	RARMISE	RAE mode	RAE mean	RAE median	Cov. p
pls.loc	0.001	$\tilde{m}_{5,18}$	0.430 (0.244)	0.208 (0.195)	0.199 (0.197)	0.203 (0.195)	0.828
		$\tilde{m}_{5,22}$	0.445 (0.24)	0.237 (0.206)	0.221 (0.209)	0.229 (0.205)	0.148
		$\tilde{m}_{18,5}$	0.503 (0.159)	0.384 (0.312)	0.321 (0.257)	0.326 (0.256)	0.828
		$\tilde{m}_{22,5}$	0.488 (0.134)	0.410 (0.318)	0.330 (0.264)	0.334 (0.259)	0.241
	0.01	$\tilde{m}_{5,18}$	0.424 (0.225)	0.214 (0.21)	0.203 (0.199)	0.203 (0.193)	0.723
		$\tilde{m}_{5,22}$	0.431 (0.225)	0.250 (0.211)	0.227 (0.215)	0.234 (0.206)	0.09
		$\tilde{m}_{18,5}$	0.490 (0.156)	0.407 (0.314)	0.323 (0.256)	0.323 (0.266)	0.573
		$\tilde{m}_{22,5}$	0.476 (0.128)	0.438 (0.309)	0.327 (0.263)	0.334 (0.265)	0.3
	0.1	$\tilde{m}_{5,18}$	0.414 (0.199)	0.214 (0.195)	0.212 (0.202)	0.219 (0.204)	0.567
		$\tilde{m}_{5,22}$	0.422 (0.199)	0.252 (0.199)	0.236 (0.218)	0.241 (0.209)	0.041*
		$\tilde{m}_{18,5}$	0.483 (0.147)	0.442 (0.294)	0.324 (0.255)	0.330 (0.261)	0.509
		$\tilde{m}_{22,5}$	0.470 (0.126)	0.471 (0.267)	0.325 (0.263)	0.329 (0.263)	0.393
lgb.loc	0.001	$\tilde{m}_{5,18}$	0.424 (0.26)	0.205 (0.198)	0.188 (0.182)	0.187 (0.182)	0.859
		$\tilde{m}_{5,22}$	0.417 (0.217)	0.226 (0.193)	0.215 (0.195)	0.213 (0.184)	0.013*
		$\tilde{m}_{18,5}$	0.499 (0.222)	0.324 (0.27)	0.305 (0.25)	0.305 (0.234)	0.61
		$\tilde{m}_{22,5}$	0.495 (0.189)	0.373 (0.285)	0.303 (0.26)	0.303 (0.263)	0.164
	0.01	$\tilde{m}_{5,18}$	0.417 (0.242)	0.214 (0.201)	0.187 (0.179)	0.195 (0.182)	0.059
		$\tilde{m}_{5,22}$	0.408 (0.207)	0.239 (0.187)	0.212 (0.198)	0.228 (0.194)	<0.001*
		$\tilde{m}_{18,5}$	0.488 (0.195)	0.368 (0.274)	0.313 (0.242)	0.302 (0.23)	0.518
		$\tilde{m}_{22,5}$	0.478 (0.168)	0.387 (0.287)	0.303 (0.255)	0.296 (0.249)	0.164
	0.1	$\tilde{m}_{5,18}$	0.417 (0.222)	0.215 (0.191)	0.193 (0.179)	0.198 (0.187)	0.002*
		$\tilde{m}_{5,22}$	0.411 (0.194)	0.246 (0.197)	0.222 (0.205)	0.227 (0.191)	<0.001*
		$\tilde{m}_{18,5}$	0.485 (0.188)	0.380 (0.26)	0.316 (0.251)	0.311 (0.233)	0.178
		$\tilde{m}_{22,5}$	0.478 (0.162)	0.405 (0.261)	0.300 (0.25)	0.314 (0.254)	0.363
l1b.loc	0.001	$\tilde{m}_{5,18}$	0.443 (0.272)	0.189 (0.176)	0.183 (0.177)	0.188 (0.184)	0.121
		$\tilde{m}_{5,22}$	0.435 (0.217)	0.234 (0.197)	0.215 (0.188)	0.213 (0.183)	0.001*
		$\tilde{m}_{18,5}$	0.507 (0.196)	0.369 (0.314)	0.318 (0.263)	0.320 (0.266)	0.37
		$\tilde{m}_{22,5}$	0.489 (0.163)	0.372 (0.308)	0.307 (0.242)	0.307 (0.257)	0.288
	0.01	$\tilde{m}_{5,18}$	0.437 (0.257)	0.201 (0.178)	0.190 (0.184)	0.184 (0.179)	0.124
		$\tilde{m}_{5,22}$	0.431 (0.215)	0.241 (0.2)	0.208 (0.184)	0.213 (0.181)	0.001*
		$\tilde{m}_{18,5}$	0.497 (0.195)	0.397 (0.318)	0.318 (0.255)	0.318 (0.253)	0.452
		$\tilde{m}_{22,5}$	0.476 (0.158)	0.417 (0.313)	0.312 (0.246)	0.302 (0.245)	0.674
	0.1	$\tilde{m}_{5,18}$	0.432 (0.244)	0.212 (0.191)	0.190 (0.175)	0.195 (0.174)	0.01*
		$\tilde{m}_{5,22}$	0.424 (0.198)	0.255 (0.204)	0.215 (0.19)	0.218 (0.178)	<0.001*
		$\tilde{m}_{18,5}$	0.496 (0.189)	0.406 (0.281)	0.317 (0.252)	0.319 (0.252)	0.317
		$\tilde{m}_{22,5}$	0.472 (0.145)	0.424 (0.301)	0.312 (0.247)	0.308 (0.244)	0.767
l2b.loc	0.001	$\tilde{m}_{5,18}$	0.440 (0.269)	0.197 (0.167)	0.184 (0.179)	0.189 (0.181)	0.219
		$\tilde{m}_{5,22}$	0.428 (0.212)	0.234 (0.202)	0.218 (0.188)	0.216 (0.181)	0.003*
		$\tilde{m}_{18,5}$	0.508 (0.216)	0.356 (0.301)	0.321 (0.258)	0.315 (0.256)	0.241
		$\tilde{m}_{22,5}$	0.487 (0.169)	0.367 (0.305)	0.311 (0.252)	0.313 (0.245)	0.536
	0.01	$\tilde{m}_{5,18}$	0.436 (0.275)	0.202 (0.178)	0.187 (0.172)	0.191 (0.179)	0.126
		$\tilde{m}_{5,22}$	0.428 (0.212)	0.237 (0.194)	0.216 (0.184)	0.222 (0.186)	0.004*
		$\tilde{m}_{18,5}$	0.499 (0.201)	0.373 (0.324)	0.324 (0.262)	0.323 (0.256)	0.429
		$\tilde{m}_{22,5}$	0.477 (0.164)	0.392 (0.312)	0.305 (0.25)	0.310 (0.254)	0.719
	0.1	$\tilde{m}_{5,18}$	0.428 (0.243)	0.215 (0.18)	0.190 (0.175)	0.196 (0.179)	0.028*
		$\tilde{m}_{5,22}$	0.424 (0.198)	0.249 (0.198)	0.223 (0.202)	0.227 (0.189)	<0.001*
		$\tilde{m}_{18,5}$	0.493 (0.191)	0.392 (0.293)	0.327 (0.273)	0.324 (0.25)	0.398
		$\tilde{m}_{22,5}$	0.473 (0.15)	0.427 (0.291)	0.309 (0.254)	0.311 (0.25)	0.608

Details as in Table 4.7.

Table 4.9: Accuracy of pairwise estimation of migration rate compared to fully joint estimation for deme cluster 3

ϵ	Param.	SARMISE[a]	SAE[b] mode	SAE mean	SAE median	Cov. p_j[c]	Cov. p_{pw}[d]
0.001	$\tilde{m}_{5,18}$	1.013 (0.093)	1.037 (0.477)	1.016 (0.289)	1.018 (0.282)	0.4	0.573
	$\tilde{m}_{18,5}$	1.398 (0.577)	1.596 (1.897)	1.524 (1.617)	1.575 (1.694)	0.314	0.723
	$\tilde{m}_{5,22}$	1.042 (0.445)	1.068 (1.199)	1.129 (1.198)	1.085 (1.217)	0.4	0.181
	$\tilde{m}_{22,5}$	1.402 (0.532)	1.778 (2.082)	1.517 (1.516)	1.632 (1.653)	0.314	0.5
0.01	$\tilde{m}_{5,18}$	1.003 (0.079)	1.029 (0.33)	1.007 (0.259)	1.010 (0.264)	0.35	0.61
	$\tilde{m}_{18,5}$	1.374 (0.565)	1.700 (1.989)	1.541 (1.594)	1.572 (1.673)	0.245	0.587
	$\tilde{m}_{5,22}$	1.017 (0.406)	1.023 (1.179)	1.082 (1.128)	1.087 (1.189)	0.35	0.207
	$\tilde{m}_{22,5}$	1.399 (0.521)	1.795 (2.223)	1.533 (1.553)	1.596 (1.632)	0.245	0.565
0.1	$\tilde{m}_{5,18}$	0.987 (0.08)	1.022 (0.329)	0.997 (0.302)	0.996 (0.28)	0.632	0.759
	$\tilde{m}_{18,5}$	1.333 (0.533)	1.660 (1.945)	1.486 (1.536)	1.554 (1.615)	0.075	0.576
	$\tilde{m}_{5,22}$	1.011 (0.368)	1.042 (1.164)	1.064 (1.146)	1.049 (1.173)	0.632	0.132
	$\tilde{m}_{22,5}$	1.380 (0.459)	1.799 (2.302)	1.527 (1.539)	1.630 (1.686)	0.075	0.295

The table shows results for rates of migration between demes of cluster 3 (Figure 4.1). SARMISE and SAE (see below) are given as the median across 500 independent estimations with true values drawn from the prior (median absolute deviation in parentheses). Migration rates were estimated on the $\log_{10}$ scale.

[a] Standardized absolute root mean integrated squared error. Standardized means that before averaging across test sets, we divided the measure of accuracy obtained with the pairwise estimation approach by the one obtained with the joint estimation approach (see text for details).

[b] Standardized absolute error of the pairwise estimate with respect to the joint estimate.

[c] P-value from a Kolmogorov-Smirnov test for the uniformity of the posterior probabilities of the true values ($*$: $p < 0.05$), for the joint estimation procedure.

[d] As in [c], but for the pairwise estimation procedure.

Table 4.10: Accuracy of pairwise estimation of migration rate compared to fully joint estimation for deme cluster 2

ϵ	Param.	SARMISE[a]	SAE[b] mode	SAE mean	SAE median	Cov. p_j[c]	Cov. p_{pw}[d]
0.001	$\tilde{m}_{4,12}$	1.087 (0.176)	1.078 (0.68)	1.066 (0.469)	1.064 (0.507)	0.888	0.219
	$\tilde{m}_{12,4}$	1.173 (0.46)	1.377 (1.596)	1.228 (1.264)	1.193 (1.281)	0.134	0.148
	$\tilde{m}_{4,26}$	1.185 (0.707)	1.287 (1.564)	1.163 (1.279)	1.146 (1.285)	0.888	0.219
	$\tilde{m}_{26,4}$	0.854 (0.365)	0.949 (1.195)	0.899 (0.987)	0.931 (1.08)	0.134	0.954
	$\tilde{m}_{12,14}$	1.111 (0.453)	0.990 (1.192)	1.161 (1.207)	1.112 (1.16)	0.888	0.888
	$\tilde{m}_{14,12}$	1.151 (0.469)	1.341 (1.636)	1.246 (1.284)	1.289 (1.414)	0.134	0.466
0.01	$\tilde{m}_{4,12}$	1.079 (0.157)	1.131 (0.698)	1.071 (0.444)	1.072 (0.495)	0.871	0.261
	$\tilde{m}_{12,4}$	1.156 (0.416)	1.352 (1.58)	1.214 (1.298)	1.146 (1.215)	0.139	0.107
	$\tilde{m}_{4,26}$	1.165 (0.678)	1.325 (1.65)	1.169 (1.291)	1.171 (1.263)	0.871	0.241
	$\tilde{m}_{26,4}$	0.833 (0.337)	0.982 (1.227)	0.904 (0.981)	0.913 (1.037)	0.139	0.984
	$\tilde{m}_{12,14}$	1.077 (0.438)	1.022 (1.192)	1.150 (1.186)	1.071 (1.129)	0.871	0.704
	$\tilde{m}_{14,12}$	1.103 (0.42)	1.439 (1.702)	1.234 (1.255)	1.242 (1.368)	0.139	0.327
0.1	$\tilde{m}_{4,12}$	1.052 (0.147)	1.049 (0.674)	1.072 (0.449)	1.034 (0.472)	0.146	0.531
	$\tilde{m}_{12,4}$	1.133 (0.387)	1.229 (1.36)	1.168 (1.233)	1.176 (1.231)	0.173	0.172
	$\tilde{m}_{4,26}$	1.156 (0.588)	1.169 (1.48)	1.145 (1.234)	1.180 (1.29)	0.146	0.46
	$\tilde{m}_{26,4}$	0.827 (0.296)	0.873 (1.059)	0.899 (0.986)	0.905 (1.002)	0.173	0.771
	$\tilde{m}_{12,14}$	1.040 (0.379)	0.909 (1.051)	1.108 (1.128)	1.041 (1.073)	0.146	0.399
	$\tilde{m}_{14,12}$	1.101 (0.406)	1.336 (1.521)	1.207 (1.211)	1.232 (1.331)	0.173	0.523

The table shows results for rates of migration between demes of cluster 2 (Figure 4.1). Further details as in Table 4.9.

Table 4.11: Accuracy of pairwise estimation of migration rate compared to fully joint estimation for deme cluster 5

ϵ	Param.	SARMISE	SAE mode	SAE mean	SAE median	Cov. p_j	Cov. p_{pw}
0.001	$\tilde{m}_{6,8}$	1.019 (0.099)	1.059 (0.786)	1.034 (0.358)	1.053 (0.405)	0.4	0.573
	$\tilde{m}_{8,6}$	0.930 (0.298)	0.793 (0.857)	0.917 (0.831)	0.853 (0.86)	0.648	0.954
	$\tilde{m}_{6,13}$	1.033 (0.356)	1.119 (1.251)	1.068 (1.01)	1.101 (1.126)	0.4	0.888
	$\tilde{m}_{13,6}$	0.723 (0.263)	0.701 (0.808)	0.638 (0.674)	0.609 (0.652)	0.648	0.988
	$\tilde{m}_{8,11}$	1.061 (0.373)	1.229 (1.543)	1.099 (1.067)	1.165 (1.09)	0.4	0.5
	$\tilde{m}_{11,8}$	0.976 (0.365)	0.898 (1.064)	0.979 (0.945)	0.953 (0.983)	0.648	0.98
	$\tilde{m}_{8,13}$	1.055 (0.369)	1.105 (1.306)	1.052 (0.989)	1.096 (1.04)	0.4	0.432
	$\tilde{m}_{13,8}$	0.890 (0.308)	0.705 (0.805)	0.856 (0.817)	0.754 (0.768)	0.648	0.241
	$\tilde{m}_{8,20}$	0.995 (0.358)	1.036 (1.261)	0.982 (0.929)	0.946 (0.963)	0.4	0.013*
	$\tilde{m}_{20,8}$	0.876 (0.334)	0.824 (0.917)	0.818 (0.832)	0.780 (0.822)	0.648	0.288
	$\tilde{m}_{11,13}$	0.992 (0.355)	1.132 (1.361)	0.975 (0.918)	1.037 (1.021)	0.4	0.759
	$\tilde{m}_{13,11}$	0.862 (0.334)	0.824 (0.916)	0.812 (0.852)	0.807 (0.865)	0.648	0.432
	$\tilde{m}_{11,20}$	0.905 (0.354)	0.910 (1.105)	0.839 (0.82)	0.834 (0.826)	0.4	0.062
	$\tilde{m}_{20,11}$	0.854 (0.331)	0.796 (0.875)	0.752 (0.783)	0.746 (0.767)	0.648	0.5
0.01	$\tilde{m}_{6,8}$	1.017 (0.087)	1.043 (0.524)	1.039 (0.272)	1.044 (0.351)	0.543	0.49
	$\tilde{m}_{8,6}$	0.917 (0.303)	0.827 (0.922)	0.902 (0.812)	0.850 (0.844)	0.696	0.931
	$\tilde{m}_{6,13}$	1.030 (0.344)	1.171 (1.402)	1.056 (0.984)	1.069 (1.084)	0.543	0.763
	$\tilde{m}_{13,6}$	0.718 (0.273)	0.672 (0.81)	0.646 (0.667)	0.627 (0.676)	0.696	0.975
	$\tilde{m}_{8,11}$	1.048 (0.356)	1.264 (1.489)	1.073 (1.038)	1.156 (1.172)	0.543	0.573
	$\tilde{m}_{11,8}$	0.974 (0.368)	0.892 (1.06)	0.982 (0.938)	0.938 (0.919)	0.696	0.993
	$\tilde{m}_{8,13}$	1.049 (0.363)	1.117 (1.308)	1.030 (0.993)	1.105 (1.066)	0.543	0.295
	$\tilde{m}_{13,8}$	0.885 (0.307)	0.694 (0.842)	0.841 (0.779)	0.796 (0.811)	0.696	0.252
	$\tilde{m}_{8,20}$	0.992 (0.351)	0.973 (1.198)	1.026 (0.962)	0.938 (0.905)	0.543	0.013*
	$\tilde{m}_{20,8}$	0.868 (0.323)	0.737 (0.834)	0.815 (0.838)	0.796 (0.855)	0.696	0.137
	$\tilde{m}_{11,13}$	1.005 (0.346)	1.084 (1.267)	0.984 (0.899)	1.018 (1.017)	0.543	0.479
	$\tilde{m}_{13,11}$	0.844 (0.327)	0.779 (0.923)	0.810 (0.846)	0.804 (0.857)	0.696	0.436
	$\tilde{m}_{11,20}$	0.902 (0.352)	0.843 (1.056)	0.847 (0.864)	0.867 (0.848)	0.543	0.031*
	$\tilde{m}_{20,11}$	0.859 (0.326)	0.783 (0.861)	0.735 (0.773)	0.713 (0.766)	0.696	0.285
0.1	$\tilde{m}_{6,8}$	1.016 (0.079)	1.047 (0.477)	1.043 (0.256)	1.042 (0.309)	0.475	0.515
	$\tilde{m}_{8,6}$	0.905 (0.289)	0.755 (0.858)	0.881 (0.812)	0.838 (0.813)	0.547	0.701
	$\tilde{m}_{6,13}$	1.025 (0.322)	1.157 (1.411)	1.036 (0.934)	1.066 (1.016)	0.475	0.762
	$\tilde{m}_{13,6}$	0.719 (0.255)	0.674 (0.772)	0.655 (0.697)	0.639 (0.705)	0.547	0.481
	$\tilde{m}_{8,11}$	1.029 (0.333)	1.221 (1.425)	1.058 (1.024)	1.103 (1.089)	0.475	0.543
	$\tilde{m}_{11,8}$	0.964 (0.337)	0.910 (1.092)	0.985 (0.938)	0.958 (0.927)	0.547	0.909
	$\tilde{m}_{8,13}$	1.031 (0.352)	1.028 (1.229)	1.016 (0.986)	1.059 (1.03)	0.475	0.354
	$\tilde{m}_{13,8}$	0.877 (0.291)	0.634 (0.759)	0.840 (0.775)	0.796 (0.805)	0.547	0.261
	$\tilde{m}_{8,20}$	0.976 (0.345)	1.011 (1.226)	1.006 (0.94)	0.940 (0.899)	0.475	0.01*
	$\tilde{m}_{20,8}$	0.866 (0.298)	0.705 (0.794)	0.836 (0.868)	0.798 (0.856)	0.547	0.056
	$\tilde{m}_{11,13}$	0.985 (0.321)	1.012 (1.165)	0.982 (0.889)	1.010 (0.971)	0.475	0.209
	$\tilde{m}_{13,11}$	0.840 (0.313)	0.683 (0.809)	0.791 (0.823)	0.806 (0.862)	0.547	0.234
	$\tilde{m}_{11,20}$	0.899 (0.326)	0.893 (1.108)	0.872 (0.864)	0.866 (0.885)	0.475	0.01*
	$\tilde{m}_{20,11}$	0.836 (0.301)	0.683 (0.759)	0.725 (0.8)	0.731 (0.779)	0.547	0.309

The table shows results for rates of migration between demes of cluster 5 (Figure 4.1). Further details as in Table 4.9.

4.8 Supporting information: Additional figures

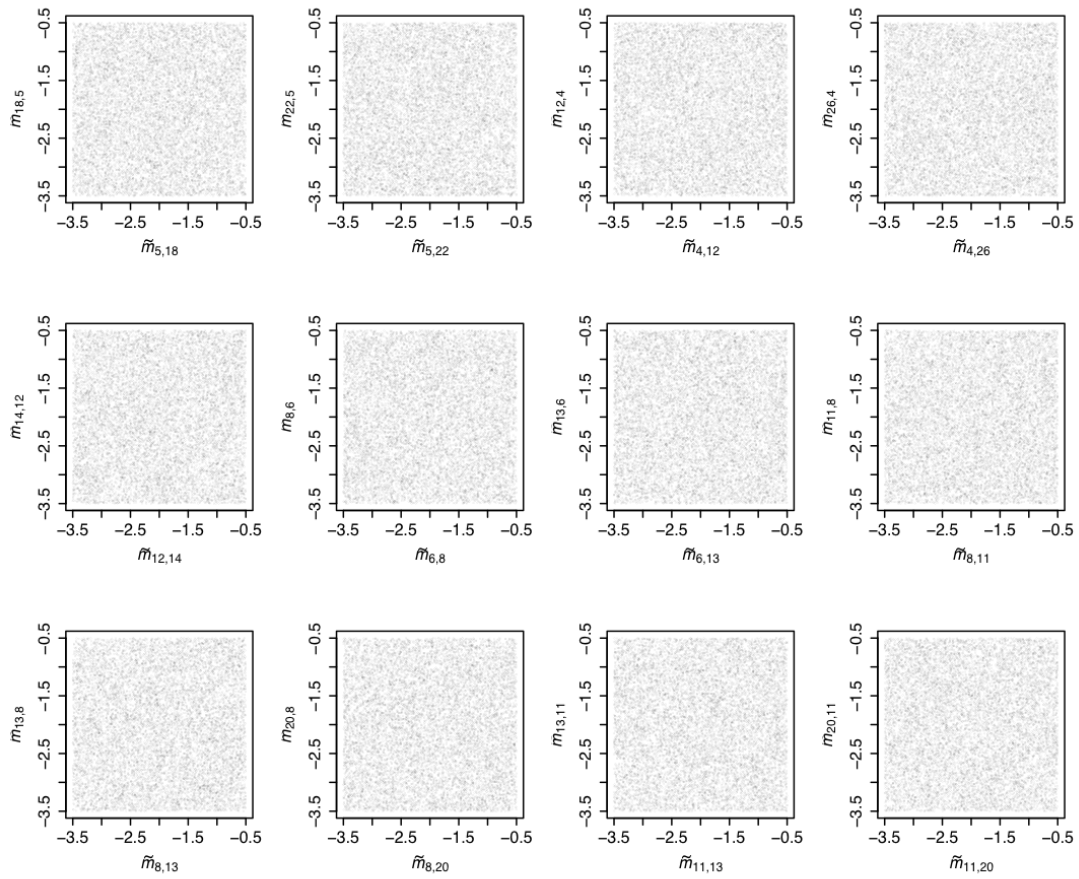

Figure 4.7: Values of $\tilde{m}_{j,i}$ that belong to data points accepted when inferring $\boldsymbol{\alpha}$ with ABC in chapter 3, plotted against corresponding values of $\tilde{m}_{i,j}$. There is no deviation from the log_{10} uniform prior distribution, and no obvious correlation between $\tilde{m}_{j,i}$ and $\tilde{m}_{i,j}$. This suggests that the summary statistics $\mathbf{s}_{\boldsymbol{\alpha}}$ used to infer $\boldsymbol{\alpha}$ in chapter 3 were not informative about the migration rates, and that the assumption that the prior of the migration rates conditional on $\boldsymbol{\alpha}$ is equal to the unconditional prior is justified. Notice that both axes are on the log_{10} scale.

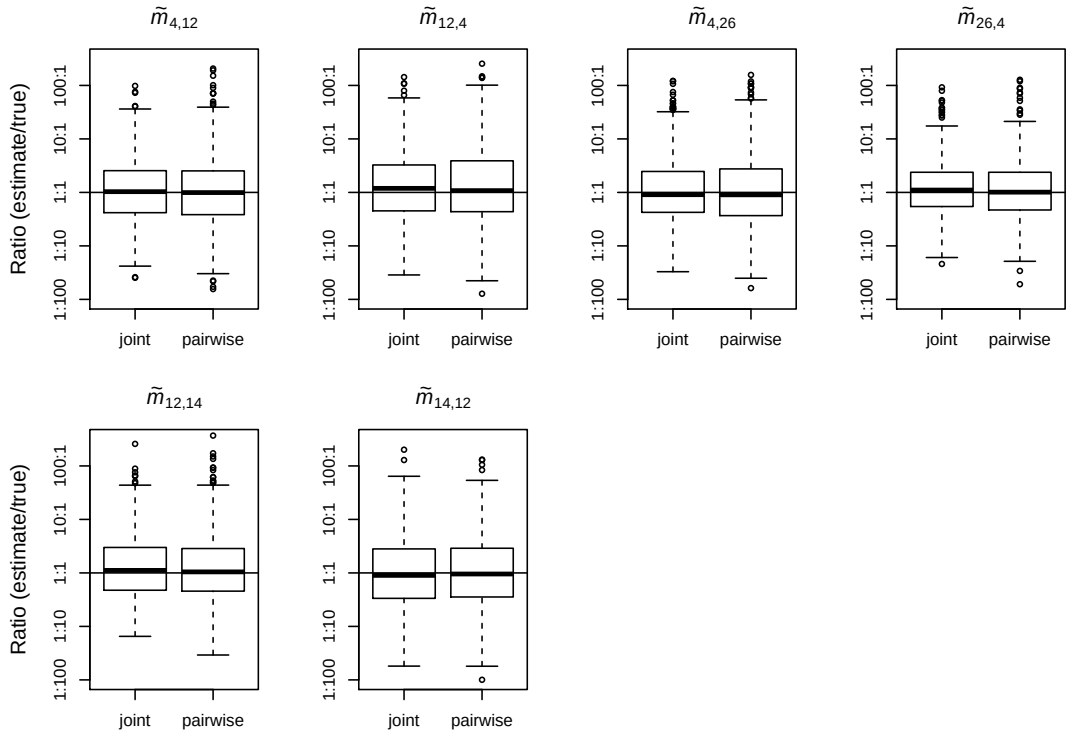

Figure 4.8: Ratio of posterior point estimate (median) to true value for the joint and pairwise estimation method. Box plots summarize data from 500 test data sets with true values sampled from the prior. Boxes show the interquartile range and whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. Note the logarithmic scale. The six parameters belonging to cluster 2 are shown (see Figure 4.4 for cluster 3 and Figure 4.9 for cluster 5).

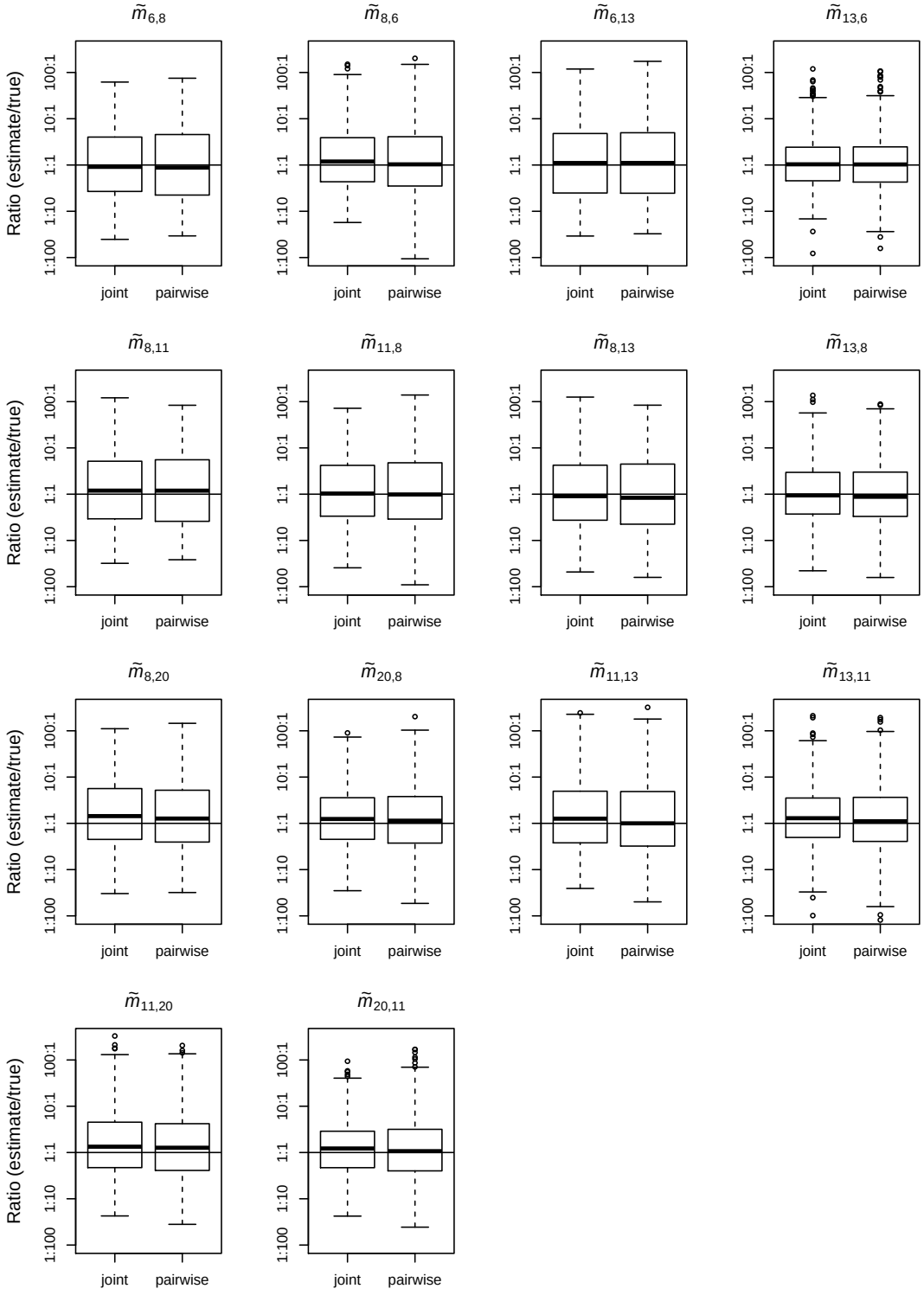

Figure 4.9: Ratio of posterior point estimate (median) to true value for the joint and pairwise estimation method for cluster 5. Further details as in Figure 4.8.

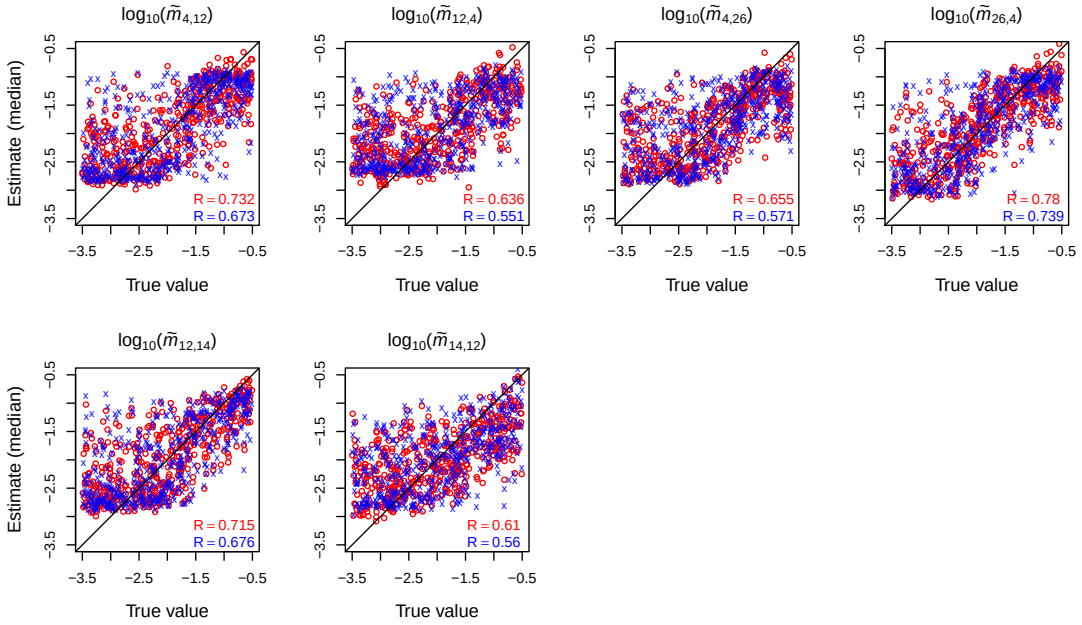

Figure 4.10: Correlation of posterior point estimate and true value for the joint (red circles) and pairwise (blue crosses) estimation method across 500 test data sets. The black line shows the expected ratio of 1:1 and R is the Pearson product-moment correlation coefficient. Plots are shown for the six parameters belonging to cluster 2 (see Figure 4.5 for cluster 3 and Figure 4.11 for cluster 5).

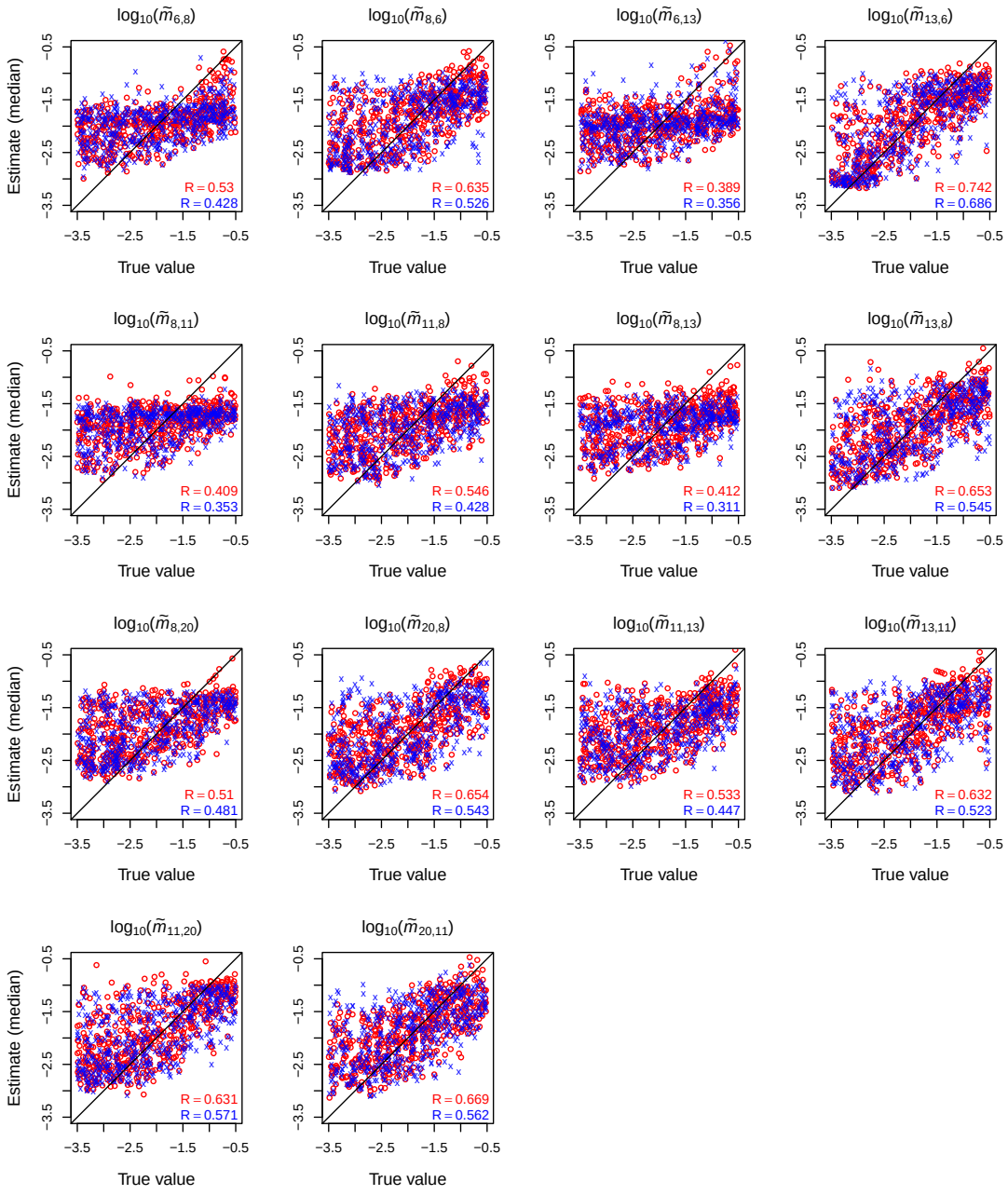

Figure 4.11: Correlation of posterior point estimate and true value for the joint (red circles) and pairwise (blue crosses) estimation for cluster 5. Further details as in Figure 4.10.

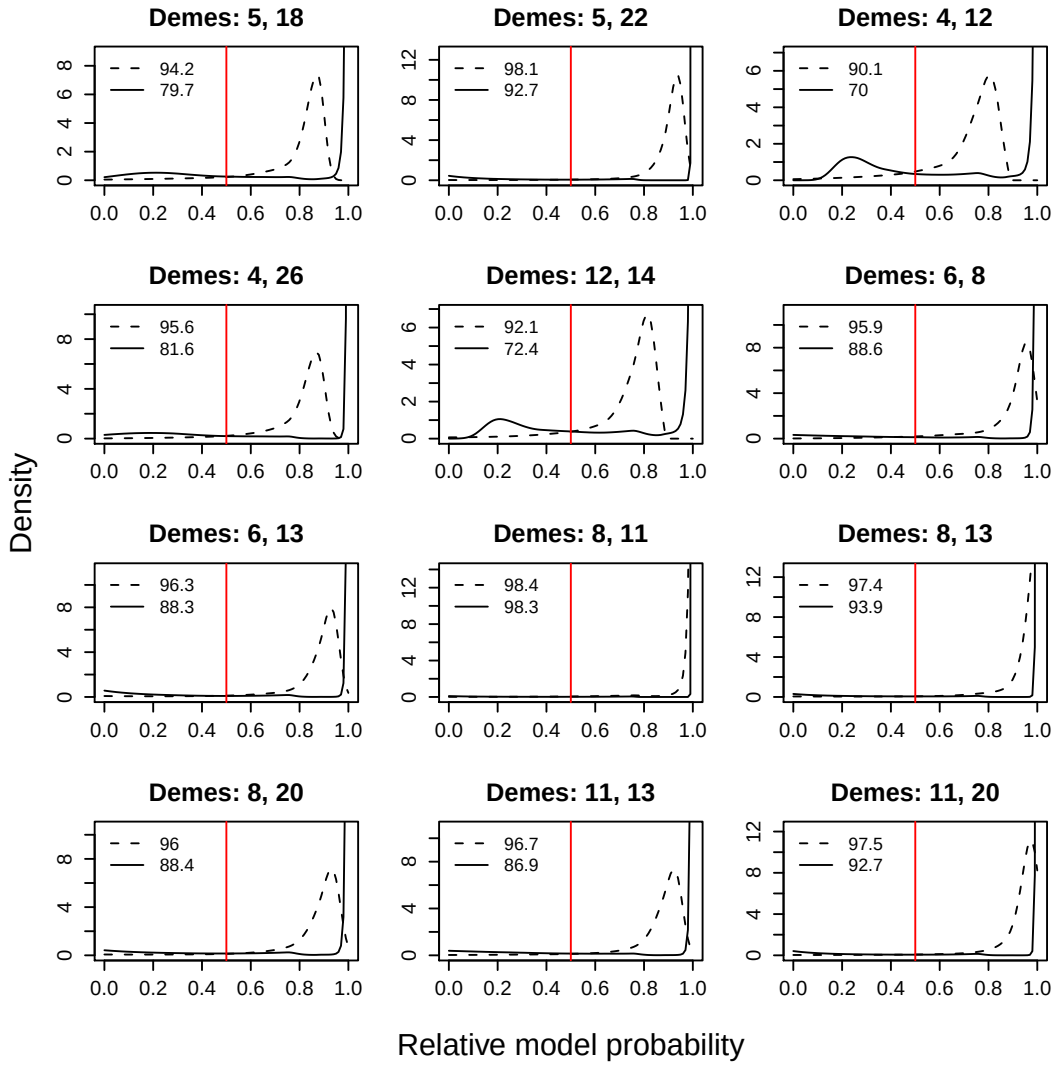

Figure 4.12: Probability density fitted to the empirical distribution of relative probabilities of the mig (solid line) and the nomig (dashed line) model when they are the true model. Empirical distributions were obtained by simulating 1000 data sets under the mig (with migration rates drawn from the prior) and nomig model (migration rates set to zero). The relative model probabilities were then estimated according to the ABC model comparison procedure explained in the text. The area under the curve to the right of the vertical line gives the proportion of times the true model was correctly recovered ($p_{\mathrm{mig}} > 0.5$). These proportions are given as percentages by the two numbers in the plot, or as β_{mig} and β_{nomig} in Table 4.5. The rejection tolerance was $\epsilon = 0.05$.

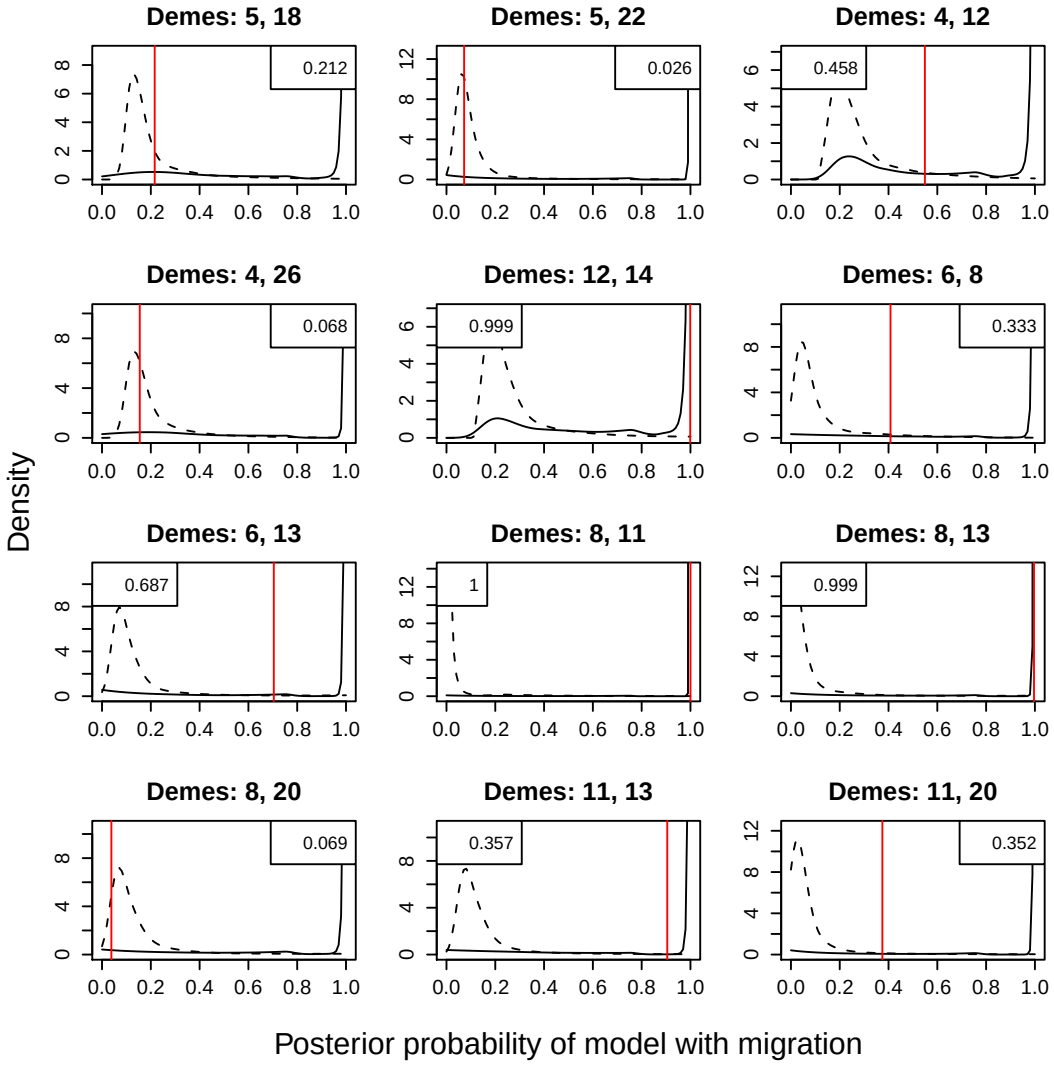

Figure 4.13: Probability density fitted to the empirical distribution of relative probabilities of the mig model when the mig (solid line) and the nomig (dasheded line) model are the true models. Empirical distributions were obtained by simulating 1000 data sets under the mig (with migration rates drawn from the prior) and nomig model (migration rates set to zero). The density estimates of the two models at the posterior probability of the mig model, p_{mig}, were used to compute the probability that mig is the correct model given p_{mig} (vertical line; *cf.* Table 4.5). This probability is given by the number in the corner of the plots.

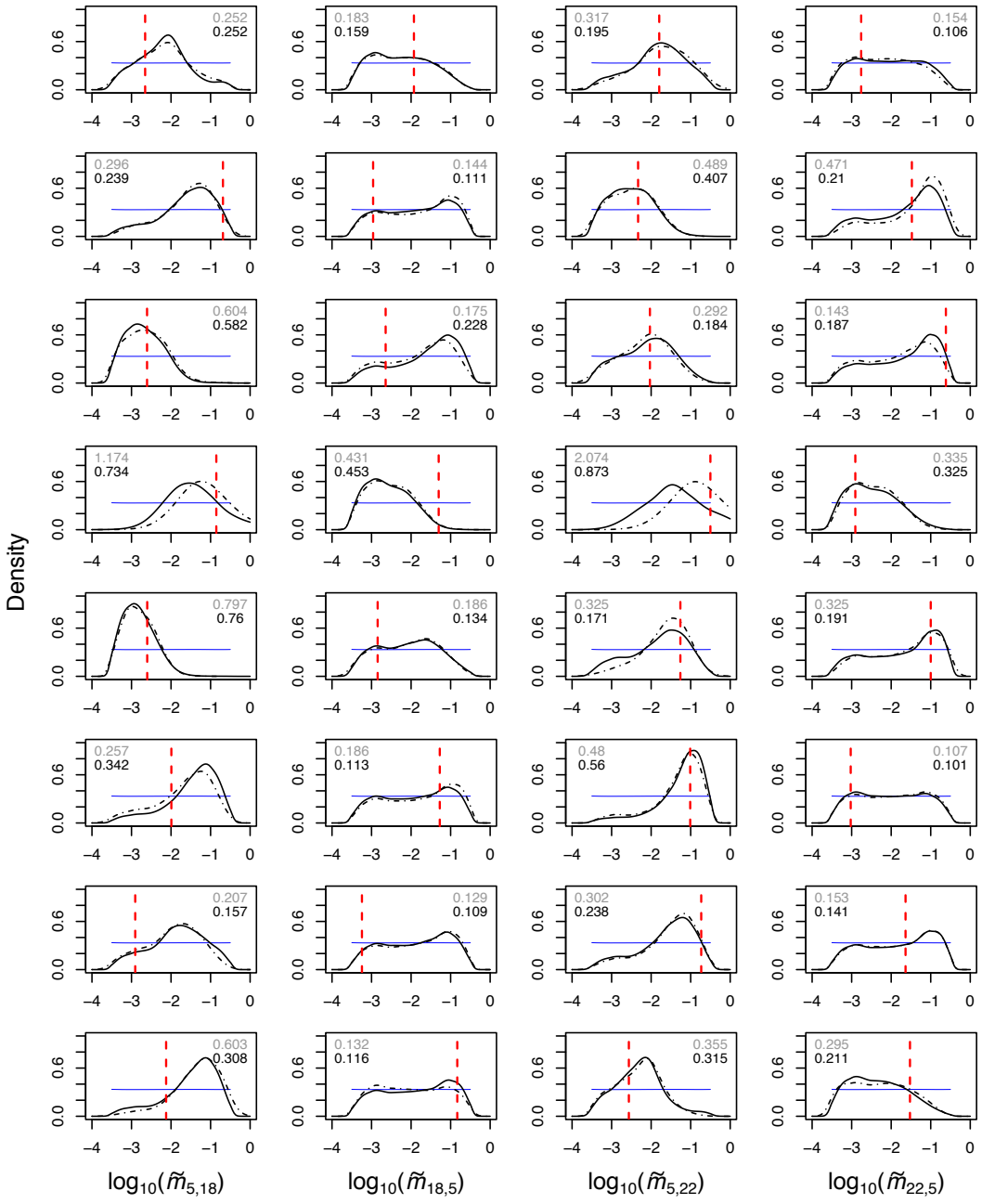

Figure 4.14: Posterior distributions of migration rates for deme cluster 3 obtained in eight independent test runs with true values drawn from the prior. Thin blue line for the prior; dotdashed black line for the posterior inferred with the joint estimation method; solid black line for the posterior inferred with the pairwise estimation method; red vertical line for the true value. One row corresponds to one test data set, and the Kullback-Leibler divergence of the posterior from the prior is given in gray and black print for the joint and pairwise method, respectively.

Chapter 5

The fate in the wild of an MHC allele shared with a domesticated species: Combining short- and long-term evidence for selection

This chapter is the result of a collaboration with Lukas Keller, Christine Grossen, Iris Biebach and Nick Barton. Lukas, Christine and Iris provided genetic data and helped designing the study. Nick suggested the matrix iteration approach, helped with the parameterization of fitness and provided tips for efficient implementation. A version of this chapter – with shortened introduction and discussion – is intended for publication in Evolution*, with Lukas, Christine, Iris and Nick as co-authors.*

5.1 Introduction

The Major Histocompatibility Complex (MHC) is a family of genes involved in immune response in vertebrates. Out of three classes, MHC class II genes code for proteins on the surface of antigen-presenting cells (macrophages, B cells and dendritic cells). The surface proteins bind extracellular pathogen-derived peptides and present them to T-helper cells, thus triggering the adaptive immune response. MHC has also been shown to play a role in mate choice (*e.g.* Radwan et al. 2008), kin-recognition and pre-natal survival (Edwards and Hedrick 1998). In most vertebrates, MHC genes are highly diverse, especially in regions that code for the peptide-binding parts of the molecule (Garrigan and Hedrick 2003; Radwan et al. 2010). Balancing selection, negative-assortative mating and maternal-fetal interactions have been proposed as evolutionary explanations for the maintenance of this diversity (Hedrick 1994). Although effects on mate choice (*e.g.* Thoß et al. 2011) and pre-natal survival (*e.g.* Knapp et al. 1996; Ober et al. 1998) have been reported, they do not seem to be the rule. There is increasing evidence for parasite pressure to be the main source of selection on MHC (Bernatchez and Landry 2003; Garrigan and Hedrick 2003; Piertney and Oliver 2006; Radwan et al. 2010). Overdominance or selection varying in time or space are the most likely underlying mechanisms (Hedrick et al. 1976; Garrigan and Hedrick 2003, but see van Oosterhout (2009) for a complementary hypothesis). The overdominance hypothesis states that heterozygotes have an advantage because

they can recognize a broader range of pathogens. This is based on the fact that both proteins are expressed on the cell surface. This assumes that MHC alleles are expressed codominantly. Selection varying in time may be due to negative-frequency dependence (*e.g.* Borghans et al. 2004), where rare MHC mutations are favored because parasites have not yet developed resistance against them. As the frequency of these alleles increases, parasites coevolve and the advantage disappears. Local adaptation to spatially heterogeneous parasite communities is an example of spatially varying selection (Bernatchez and Landry 2003). Evidence for both overdominance and selection varying in time or space has been reported in free-living vertebrates and under laboratory conditions (*e.g.* Paterson et al. 1998; Miller et al. 2001; Charbonnel and Pemberton 2005; Meyer-Lucht and Sommer 2005; Piertney and Oliver 2006; Mona et al. 2008; Fraser et al. 2010, see Sommer (2005) for an extensive review). In addition, there is growing evidence for the association of specific MHC genotypes or individual alleles with susceptibility to infection (Arkush et al. 2002; Sommer 2005; Radwan et al. 2010).

MHC diversity may be of conservation concern. Bottlenecks or spatial subdivision reduce genetic diversity and increase the chance of matings among relatives. This may lead to inbreeding depression (*e.g.* Keller and Waller 2002) and reduced immune response (Reid et al. 2003, 2007). It has been suggested that reduced diversity at MHC causes higher susceptibility to infectious disease and population decline (O'Brien and Evermann 1988; Coltman et al. 1999a; Arkush et al. 2002). This has been challenged by Gutierrez-Espeleta et al. (2001) who found high MHC diversity in bighorn sheep (*Ovis canadensis*) in spite of a strong population decline (see also Aguilar et al. 2004). Moreover, it is difficult to separate MHC-specific effects on fitness from effects of inbreeding in general (Sommer 2005; Hansson and Westerberg 2008; Radwan et al. 2010), and from selection acting elsewhere on the genome (Santucci et al. 2007; Thoß et al. 2011). Comparing MHC diversity to neutral diversity may give further insight. For natural populations such comparisons do not yield consistent results across studies. Some have reported higher spatial differentiation at MHC compared to neutral loci (*e.g.* Miller et al. (2001) in sockeye salmon (*Oncorhynchus nerka*)), others found a more uniform spatial distribution at MHC compared to neutral markers (*e.g.* Mona et al. (2008) in Alpine chamois (*Rupicapra rupicapra*), Santucci et al. (2007) in mouflon (*Ovis orientalis musimon*)), and in Soay sheep (*Ovis aries*) the result depended on the period considered (Charbonnel and Pemberton 2005). Lukas et al. (2004) observed high MHC diversity in two gorilla populations of very different effective size, whereas Mona et al. (2008) found patterns of MHC diversity in Alpine chamois populations that could be explained by demography alone. In humans, it seems that selection has not fully removed the traces of demography at MHC (Currat et al. 2010, and references therein). Thus, while differences in patterns of diversity at MHC versus neutral variation may be a signature of selection, does their absence mean that there is no selection? Not necessarily, since demography may overwhelm selection in shaping diversity at MHC, and tests for selection might mislead. It is therefore important to disentangle the effects of demography and selection in empirical studies of MHC diversity. This may be challenging, because demography is complex, including changes in population size (bottlenecks), population genealogy (geographic origin, founder events, admixture) and migration.

Empirical studies in which the impacts on MHC variation of genetic drift, migration and selection can be well separated are still rare. Overall, there are several lines of evidence for

selection on MHC. However, not much is known about its strength. Here, we study genetic variation at exon 2 of the MHC class IIa gene DRB in a spatially structured population of Alpine ibex (*Capra ibex ibex*) in the Swiss Alps. We use the distribution of observed allele frequencies across ibex colonies for inference on the strength of selection and its mode of dominance. Thereby, we take into account demography and migration. This study profits from the fact that population history has been well documented over the time scale of interest (1906 till 2006).

Alpine ibex were almost extinct by the beginning of the 18^{th} century (Stuwe and Scribner 1989). One population survived in the Gran Paradiso area, Northern Italy. At the end of the 19^{th} century, first attempts were made to re-establish former populations in the Swiss Alps. Since pure bred Alpine ibex were hard to obtain, hybrids between ibex and domestic goat were used. None of these hybrid releases was successful. From 1906 on, pure Alpine ibex were brought from the Gran Paradiso population to two zoos in St. Gall and Interlaken, Switzerland, and bred there in captivity. Starting from 1911, a first set of former colonies were re-established with pure ibex bred in captivity. These colonies were used as a reservoir for further transfers. The re-introduction has been documented in great detail (Couturier 1962; Nievergelt 1966; Stuwe and Nievergelt 1991; Scribner and Stuwe 1994; Maudet et al. 2002; Biebach and Keller 2009). This provides the opportunity to condition genetic inference on information that would otherwise have to be inferred from genetic data first (*e.g.* Mona et al. 2008). In the following, we use the term 'deme' (Gilmour and Gregor 1939) for the spatially separated colonies (subpopulations).

Alpine ibex is a protected species, but in some demes in the Swiss Alps annual culls have been carried out since 1978 to prevent damage to forests. Relatively low genetic diversity within, and moderate to strong differentiation between demes at 37 neutral microsatellites has been reported by Biebach and Keller (2009). Overall, the re-introduction history is well reflected in today's genetic composition. Grossen (2005) found that there are only two haplotypes of exon 2 at DRB (Figure 5.1) present in the whole population. One is specific to Alpine ibex, the other one is shared with domestic goat (*C. aegagrus hircus*). The two haplotypes differ in 29 out of 198 base pairs (14.6%). There are three hypothetical explanations for this trans-species polymorphism (TSP): i) a shared ancestral polymorphism (SAP), ii) introgression via hybridisation, and iii) convergent evolution (homoplasy). The phylogenetics of the genus *Capra* has not been fully resolved (but see Mannen et al. 2001; Kazanskaya et al. 2007). It has been estimated that the lines of Alpine ibex and domestic goat split from their most recent common ancestor about 6 million years ago (L. Keller, personal communication). Although an observed sequence divergence of 14.6% between the haplotypes seems large, we currently do not have enough information on polymorphism in domestic goat to argue for or against homoplasy. On the other hand, TSPs at MHC genes have been reported in various vertebrate taxa (Rodentia: Cutrera and Lacey (2007); Felidae: Wei et al. (2010); Mustelidae: Becker et al. (2009); Ursidae: Goda et al. (2010); *Spenicus* penguins: Kikkawa et al. (2009); Primates: Garrigan and Hedrick (2003)), and for ruminants in particular (Gutierrez-Espeleta et al. 2001; Worley et al. 2006; Ballingall et al. 2010). These TSPs are usually interpreted as being SAPs maintained by balancing selection (Takahata and Nei 1990; Takahata 1990; Garrigan and Hedrick 2003). Introgression via hybridisation could also lead to TSPs and create genealogies similar to bal-

ancing selection. In case of Alpine ibex, this alternative hypothesis is justified. Hybridisation with domestic goat has repeatedly been observed in nature, and hybrid offspring are viable and fertile (Giacometti et al. 2004). Backcrosses of hybrids with pure ibex could potentially establish and integrate in natural colonies if environmental conditions are not too harsh. One case has been documented in the Swiss Alps, but those hybrids were culled (Giacometti et al. 2004). For the analyses and result in this study, it is not directly relevant if the TSP is a SAP or due to introgression, because we focus on the most recent ten generations, for which the data suggest that the goat haplotype has already been present in ibex. However, under some assumptions, our results may render one or the other explanation for the TSP more likely.

Grossen (2005) reported two potential signals of selection at the ibex MHC: i) higher spatial differentiation (F_{ST}) for MHC-linked markers compared to neutral ones, ii) a (non-significant) trend for increasing number of nematode (Strongylida) egg counts in feces with increasing frequency of the 'goat' haplotype (Grossen 2005, p. 29). Recently, Ch. Grossen also pointed out a correlation between heterozygosity and age at sampling (personal communication).

Here, we study a larger and different set of demes compared to Grossen (2005). We assess whether the observed genetic variation within and between demes at exon 2 of DRB has been shaped by demography only, or if signals of selection are present. We combine short- and medium-term evidence (see below) and set up a drift-selection-migration model to estimate the strength of selection (s). Thereby, we account for the history of re-introduction, explore different modes of dominance and investigate the influence of gene flow via migration. We also aim at confirming the suggested correlation between age at sampling and heterozygosity. In particular, we use this information to learn about dominance and to condition the estimation of s.

Selection may be classified according to the time scale on which it occurs (Black and Hedrick 1997; Garrigan and Hedrick 2003), and tests have been developed for different time scales and types of data (Nielsen 2001; Garrigan and Hedrick 2003; Nielsen 2005). Following Garrigan and Hedrick (2003), we distinguish between selection in the current generation (short-term), selection over several generations in the same species (microevolutionary time scale, medium-term),

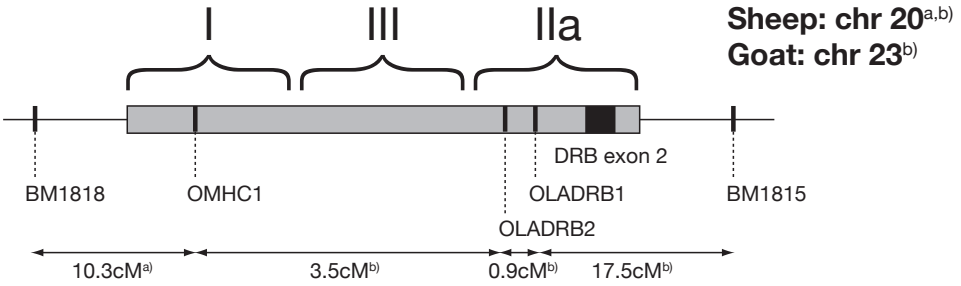

Figure 5.1: Part of the sheep (*Ovis aries*) chromosome 20 with the MHC class I, III and IIa regions, including linked microsatellites. The corresponding chromosome in goat (*Capra*) is chromosome 23 (Vaiman et al. 1996). DRB exon 2 is closely linked to the microsatellites OLADRB1 and OLADRB2. The OLADRB2 allele with repeat length 277 (A_1 in the main text) is diagnostic for the DRB exon 2 allele that is shared between Alpine ibex and domestic goat. Details and genetic distances from [a)]Paterson et al. (1998) and [b)]Maddox et al. (2001).

and selection over the history of species (macroevolutionary time scale, long-term). Examples of signals of selection in the current generation are deviation from Hardy-Weinberg proportions (HWE) or correlations between genotype and traits influencing fitness. Examples of signals (and tests) over several generations are deviations of population genetic diversity from neutral expectation (*e.g.* Ewens-Watterson test, Watterson 1978) or a geographical distribution of genetic variation that is incongruent with neutrality (*e.g.* Lewontin-Krakauer test and related approaches, Lewontin and Krakauer 1973; Beaumont and Nichols 1996). Finally, signals of selection over the history of species may be found in ratios of nonsynonymous to synonymous diversity within and between species (*e.g.* McDonald Kreitman test, McDonald and Kreitman 1991) or by comparison of observed site frequency spectra to the neutral expectation (*e.g.* Tajima's D test, Tajima 1989). Trans-species polymorphisms (TSPs) are also a signal of selection on that scale. Clear separation between the medium- and long-term signals is not always possible. For example, tests that are based on the site frequency spectrum may apply to both time scales. Garrigan and Hedrick (2003) also pointed out an important problem for inference: long-term signals generally take a long time to establish. Once present, however, it also takes a long time for them to disappear, even in the absence of selection. Therefore, if the TSP (trans-species polymorphism) at MHC between Alpine ibex and domestic goat were a SAP (shared ancestral polymorphism), this cannot be taken as evidence for selection still acting on the microevolutionary time scale that is of interest in our study. Another issue is that most tests for selection – or 'neutrality tests' – can at most reject the null model of neutral evolution for a given model. In general, they do not provide a direct estimate of the strength of selection, or even the mode of dominance. For this, a specific model and assumptions about selection (and dominance) are needed.

To overcome some of these problems, we use a combination of short- and medium-term analyses. In particular, we first investigate the correlation between genotype and age at sampling, both for a diagnostic microsatellite linked to DRB (OLADRB2) and 37 neutral markers. This short-term signal may reveal information about dominance. For the medium-term analysis, we use a modification of a method by Beaumont and Nichols (1996) to infer the spatial configuration of selection. This method uses the variation of observed allele frequencies across demes as information. We then set up a drift-selection-migration model and develop a matrix iteration approach that allows for likelihood-based inference on the strength of selection, the dominance coefficient and the initial allele frequency. This method is also part of the medium-term analysis and makes use of the full observed distribution of allele frequencies. Similar approaches have been used before by Keightley and Eyre-Walker (2007) and Zeng and Charlesworth (2009). Last, we assess the influence of gene flow via migration on the estimated selection coefficient.

5.2 Model and parameters

We model the change in time of allele frequencies in a spatial context and under the influence of demography (drift), selection and migration. We consider diploid monoecious individuals and one locus with two alleles, where allele A_1 is the allele shared with goat ('goat' allele) and A_2 is the allele found only in ibex. We let time be discrete in units of one generation and assume, for the moment, that generations are non-overlapping. We will later account for the presence

of two sexes, overlapping generations and other complications that apply to Alpine ibex by calculating an appropriate effective deme size.

5.2.1 Demography and spatial structure

We denote by t the time in generations. Forward in time, we start at time t_0 with one ancestral deme d_0, and an initial allele frequency of p_{init}. After some generations, further demes are derived from the ancestral deme (Figure 5.2). We denote these derived demes by d_α, where $\alpha \in \{1, 2, \ldots, \Gamma\}$, and the corresponding times when they were founded by t_α. In our case, $\Gamma = 13$. All derived demes do not need to be founded from the ancestral deme at the same point in time, but the time at which the first one is founded is given by $t_f := t_1$. Notice that the indices i to the times t_i do not reflect a temporal order, except that t_1 refers to the deme established first. We assume that the allele frequency in the ancestral deme (p_0) is constant between t_f and the time when the last derived deme is founded, t_Γ.

The demography of a derived deme d_α is determined by a sequence of values of its deme size $N^{(\alpha)}$,

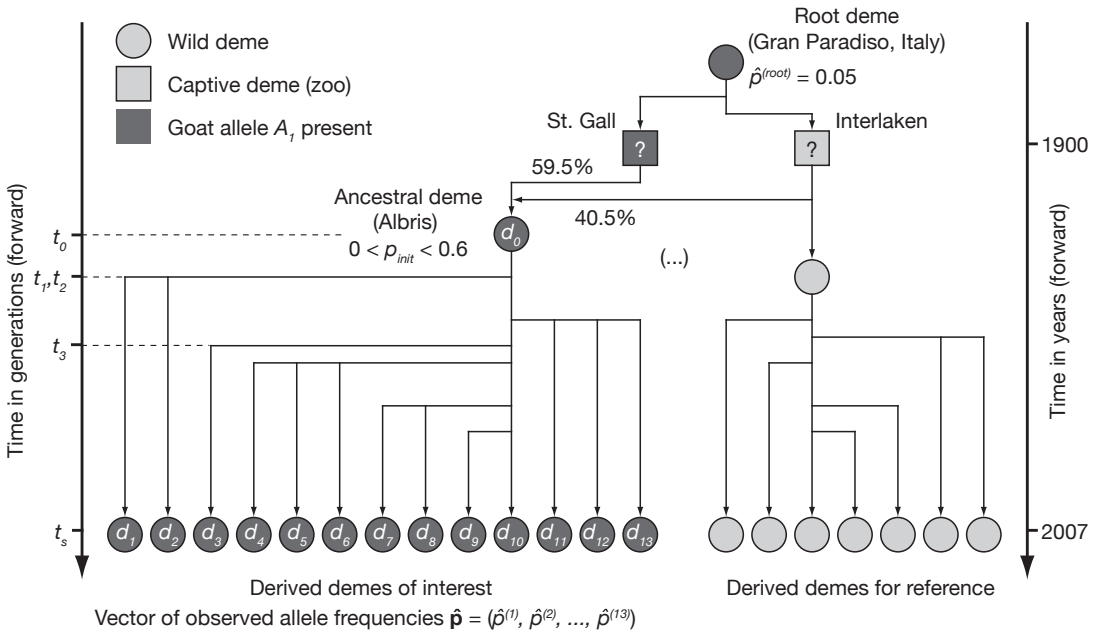

Figure 5.2: Schematic representation of the founding history and genealogy of Alpine ibex demes in the Swiss Alps. Only a relevant subset of demes is shown. This includes the two zoo populations in St. Gall and Interlaken, where ibex were bred in captivity. The question marks mean that we do not know via direct observation if the 'goat' allele, A_1, was present in these two demes. We indirectly inferred its presence or absence via observations in the respective set of derived demes. The estimate of the frequency of A_1 in the root deme, $\hat{p}^{(root)}$, is from year 2007, not from before 1900. Time in years is given on the right, time in generations on the left, where t_α is the time when deme α was founded and t_s the time of genetic sampling. Percentages along the arrows leading to d_0 refer to the proportions of founders originating from the two zoos. These numbers limit the range of p_{init}, the initial allele frequency of A_1 in d_0. The topology of founder events for the derived demes is only schematic (*cf.* Table 5.5, and Figure 5.12 in SI for details).

$$\mathcal{N}^{(\alpha)} := \left(N_{t_\alpha}^{(\alpha)}, N_{t_\alpha+1}^{(\alpha)}, \ldots, N_{t_s-1}^{(\alpha)}, N_{t_s}^{(\alpha)} \right). \tag{5.1}$$

We call $\mathcal{N}^{(\alpha)}$ the deme size trajectory of deme d_α. The first value of a trajectory, $N_{t_\alpha}^{(\alpha)}$, corresponds to the number of founders that is drawn from the ancestral deme d_0 to establish deme d_α. The last value of the trajectory, $N_{t_s}^{(\alpha)}$, corresponds to the deme size of deme d_α at the time of genetic sampling t_s (Figure 5.2). For the deme size trajectory of the ancestral deme, replace α by 0 in (5.1). Derived demes experience immigration of individuals from a common migrant pool at a rate m per generation. We assume that the immigrant is large enough so that its genetic composition changes deterministically over time.

Genetic sampling of derived demes takes place at time t_s, before migration and selection. In reality, not all individuals belonging to a deme were sampled. We model this by drawing without replacement the corresponding number of alleles from the deme. This number follows a hypergeometric distribution.

5.2.2 Migration, selection and genetic drift

Our model assumes the following life cycle. We start with zygotes in generation t and assume that they reach the adult stage immediately. Young adults experience viability selection before they become older adults. We assume soft selection, which means that selection does not change the relative deme sizes. After selection, a proportion m of adults is replaced by immigrants from a common immigrant pool according to the continent-island model (Wright 1931). Reproduction and deme size regulation follow migration and lead to zygotes of generation $t+1$.

We denote the fitnesses of the two homozygote genotypes A_1A_1 and A_2A_2 by w_{11} and w_{22}, and we assume that there is no position effect, such that the heterozygotes A_1A_2 and A_2A_1 have the same fitness w_{12}. We further assume that the fitnesses are the same in all demes. The marginal fitnesses of alleles A_1 and A_2 in deme d_α are

$$w_{1,\alpha}(t) := w_{11}p_\alpha(t) + w_{12}(1-p_\alpha(t)), \quad w_{2,\alpha}(t) := w_{12}p_\alpha(t) + w_{22}(1-p_\alpha(t)), \tag{5.2}$$

respectively, where $p_\alpha(t)$ is the frequency of allele A_1 in deme d_α at time t. The mean fitness in deme d_α at time t is then given by

$$\bar{w}_\alpha = w_{1,\alpha}p_\alpha + w_{2,\alpha}(1-p_\alpha), \tag{5.3}$$

where we have omitted the explicit notation of time for simplicity. The allele frequency p_α^* after selection is

$$p_\alpha^* = p_\alpha \frac{w_{1,\alpha}}{\bar{w}_\alpha}. \tag{5.4}$$

Notice that this holds for all demes, including the ancestral deme for which α must be replaced by 0 in (5.4). For the derived demes ($\alpha \in \{1, 2 \ldots, \Gamma\}$), the initial frequency is $p_\alpha(t_\alpha)$ and determined by the sampling of founders from the ancestral deme d_0. For the ancestral deme (indicated by $\alpha = 0$), the initial frequency $p_0(t_0)$ is equal to p_{init}. Equation (5.4) also holds analogously for the immigrant pool, for which we assumed that the allele frequency changes deterministically. In generation t the frequency p_I^* of allele A_1 after selection in the immigrant pool is given by

$$p_I^* = p_I \frac{w_{1,I}}{\bar{w}_I}, \tag{5.5}$$

where $w_{1,I}$ and $\bar{w}_I$ are defined analogously to (5.2) and (5.3). The initial value, $p_I(t_f)$, is assumed to be equal to the frequency of A_1 in the ancestral deme at the time when the first derived deme was founded, *i.e.* $p_I(t_f) = p_0(t_f)$.
We consider two alternative fitness parameterizations, one for over- and underdominance, and one for intermediate dominance (directional selection). In the former, the fitnesses are defined as

$$w_{11} := 1 - s(1 - \phi), \quad w_{12} := 1, \quad w_{22} := 1 - s\phi, \tag{5.6}$$

where s is the selection coefficient and ϕ ($0 \leq \phi \leq 1$) is the frequency that allele A_1 reaches at the internal equilibrium (stable for overdominance, unstable for underdominance). Alternatively, ϕ may be interpreted as a dominance coefficient. There is overdominance whenever $s > 0$ and $0 < \phi < 1$, and underdominance whenever $s < 0$ and $0 < \phi < 1$, whereby ϕ specifies the degree of asymmetry 'in favor' of the A_1A_1 genotype. The closer ϕ is to 1, the closer is the fitness of A_1A_2 to the fitness of A_1A_1. In the second parameterization, we define

$$w_{11} := 1 - s, \quad w_{12} := 1 - hs, \quad w_{22} := 1, \tag{5.7}$$

where s is again the selection coefficient and h is the dominance coefficient. Notice that the special cases of $h = 0$ (full recessivity of A_1) and $h = 1$ (full dominance of A_1) are covered by the ϕ-notation, if $\phi = 0$ for the former case, and if $\phi = 1$ and s is re-defined as $-s/(1 - s)$ for the latter case. Therefore, to omit redundancy, we constrain h such that $0 < h < 1$. For parametrization (5.7), if $s > 0$ there is directional selection against A_1, if $s < 0$ there is directional selection in favour of A_1. Overall, we require that $w_{ij} \geq 0 \; \forall i, j$.

In the derived demes, migration follows after selection. The allele frequency in deme d_α in generation t after migration is given by

$$p_\alpha^{**} = p_\alpha^* + m(p_I^* - p_\alpha^*), \tag{5.8}$$

where p_α^* and p_I^* are given by (5.4) and (5.5), respectively.

To model the effect of genetic drift we assume that reproduction and deme size regulation in each deme follow the Wright-Fisher model (Fisher 1930; Wright 1931). This implies random mating with selfing and random union of gametes. With these assumptions, the number j of copies of allele A_1 in a derived deme d_α in generation $t + 1$ follows a binomial distribution with $2N_{t+1}^{(\alpha)}$ trials and probability of success equal to the allele frequency after selection and migration in the previous generation, $p_\alpha^{**}(t)$.

5.2.3 Parameters

We briefly recall the four parameters of the model: p_{init} is the initial frequency of the 'goat' allele (A_1) in the ancestral deme d_0 at time t_0; ϕ is the allele frequency that would be reached at the deterministic internal equilibrium of the selection dynamics in the case of overdominance, *i.e.* for $w_{12} > w_{11}$ and $w_{12} > w_{22}$ (otherwise it is unstable); alternatively, h is the dominance coefficient for intermediate dominance schemes; s is the selection coefficient; and m is the proportion of genes contributed by immigrants from the migrant pool each generation. Recall further that ϕ may be interpreted as a dominance parameter, with symmetric over- or underdominance if $\phi = 0.5$. For an overview of parameters and symbols used, see Table 5.1.

Table 5.1: List of parameters and symbols used in the main text.

Symbol	Description
A_1, A_2	Two alleles, A_1 shared with domestic goat
d_0	Ancestral deme
d_α, $\alpha \in \{1, 2, \ldots, \Gamma\}$	Derived deme
Γ	Number of derived demes (13 in this case)
t	Time in generations
t_0, t_α	Time of foundation of d_0 and d_α, respectively
t_f, t_Γ	Time of first and last founder event
$\mathcal{N}^{(\alpha)} = (N_{t_\alpha}^{(\alpha)}, \ldots, N_{t_s}^{(\alpha)})$	Deme size trajectory of d_α ($\alpha = 0$ for d_0)
p_{init}	Initial frequency of A_1 in d_0
p_0	Constant frequency of A_1 in d_0 between t_f and t_Γ
$p_{t_s}^{(\alpha)}$	Frequency of A_1 in d_α at the time of sampling
p_α, p_α^*, p_α^{**}	Frequency of A_1 in d_α before and after selection, and after migration
p_I, p_I^*, p_I^{**}	Frequency of A_1 in the immigrant pool before and after selection, and after migration
w_{ij}	Relative fitness of genotype A_iA_j
$w_{i,\alpha}$, $\bar{w}_\alpha$	Marginal fitness of A_i in d_α, mean fitness in d_α
s	Selection coefficient
ϕ	Dominance coefficient for over- or underdominance
h	Dominance coefficient for directional selection
m	Immigration rate from migrant pool to derived demes
$\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_1, \ldots, \hat{p}_\Gamma)^T$	Vector of frequencies of A_1 observed at time t_s
$\mathbf{Q}_{t \to t'}^{(\alpha)}$	Matrix of transition probabilities $q_{ij}^{(\alpha)}(t)$ between times t and t' ($\alpha = 0$ for ancestral deme d_0)
$\mathbf{F}^{(\alpha)}$	Vector of transition probabilities $f_{kl}^{(\alpha)}$ for the founder event of d_α

5.3 Data and methods

5.3.1 Data

Demographic data

Census sizes of ibex populations in the Swiss Alps have been recorded since 1911, and the numbers of males and females transferred between populations in the process of re-introduction have been documented. These data were available from the literature (Couturier 1962; Nievergelt 1966) or provided by the Swiss Federal Office for the Environment (FOEN), the cantons, and the Swiss National Park. For some periods and populations, no census data were available. We interpolated missing values linearly, if the gap of missing data was only one year, or exponentially, if values for two or more successive years were missing.

Phenotypic and genetic data

We obtained tissue samples of Alpine ibex culled between 2005 and 2007, and blood or tissue samples collected during the same period from a small number of additional individuals (Biebach and Keller 2009). The age of individuals at the time of sampling and the sex were determined. A total of 421 individuals were genotyped at three microsatellites linked to the MHC complex on chromosome 23: OLADRB1, OLADRB2 and OMHC1 (Figure 5.1; Vaiman et al. 1996; Paterson et al. 1998; Maddox et al. 2001; Grossen 2005; Biebach and Keller 2009). Individuals

were also genotyped at 37 putatively neutral microsatellites as described in Biebach and Keller (2009).

Data sets and standard tests for neutrality and linkage

The model described in the previous section applies to a subset of 14 demes that have in common a relatively simple genealogy: one deme is ancestral to 13 derived demes (Figure 5.2). We built a first data set, data_mhc_14, with all samples from these 14 demes for which age, deme, sex and the OLADRB2 genotype are known. A total of 307 samples (138 females, 169 males) fulfilled these criteria (Table 5.5). We used this data set for both the matrix iteration approach (medium-term signals) and the analysis of genotype versus age at sampling (short-term). While the matrix iteration approach is a very general approach, our specific implementation is justified only for data set data_mhc_14, because the demographic model applies only to these 14 demes. The requirements with respect to demography are less stringent for the remaining analyses, and samples from nine additional demes could be used for those. This extended data set, data_mhc_23, contains a total of 421 samples (189 females, 232 males). Genotypes, deme names and further details of the samples in both data sets are provided in Tables 5.6 and 5.9 in the Supporting Information (SI).

Marker BM1225 was monomorphic for data_mhc_14 and therefore excluded from this data set. The frequency of the 'goat' allele (A_1) did not differ between females and males in both data sets (Mantel-Haenszel chi-squared test: $\chi^2_1 = 1.607$, $p = 0.205$ for data_mhc_14, and $\chi^2_1 = 0.724$, $p = 0.395$ for data_mhc_23). We tested for deviations from HWE and for linkage disequilibrium (LD) using GENEPOP vesion 4.0.9 (Raymond and Rousset 1995; Rousset 2008). For data_mhc_14, none of the neutral markers showed significant deviation from HWE. We found a marginally significant heterozygote deficit at OLADRB1 ($p = 0.049$) and OMHC1 ($p = 0.037$) in deme Julier Süd for data_mhc_14. For data_mhc_23, we additionally observed significant heterozygote excess for OLADRB1 in Cape Moine ($p = 0.018$) and marginally significant excess in Flüela ($p = 0.054$). We observed no deviation from HWE at OLADRB2 (but see below for differences between age classes). No correction for multiple testing was applied for results on HWE. Not surprisingly, we found highly significant pairwise LD between the three MHC-linked markers OLADRB1, OLADRB2 and OMHC1 ($p < 10^{-4}$; significant after correction by Holm (1979)) both for data_mhc_14 and data_mhc_23 when samples were pooled across demes. When samples were not pooled, we observed significant LD in some but not all demes. We found no significant LD between any pair of neutral markers, although some share a chromosome. In the following, we denote the vector of allele frequency estimates by $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_\Gamma)^T$, where $\hat{p}_\alpha$ is the observed frequency of A_1 in deme d_α.

5.3.2 Detecting medium-term signals of selection

Matrix iteration approach based on the full allele frequency distribution

For parameter estimation we set up a framework that represents evolution in each deme as a Markov chain and allows inference via matrix iteration (Ewens 1979; Keightley and Eyre-Walker 2007; Zeng and Charlesworth 2009). For a given derived deme d_α, there is a transition matrix $\mathbf{Q}^{(\alpha)}$ that contains the probabilities $q_{ij}^{(\alpha)}$ of going from a state with i copies of allele

A_1 in generation t (with deme size $N_t^{(\alpha)}$) to a state with j copies in generation $t+1$ (with deme size $N_{t+1}^{(\alpha)}$): $\mathbf{Q}_{t\to t+1}^{(\alpha)} = \left\{q_{ij}^{(\alpha)}(t)\right\}$, $i \in 0,1,\ldots,2N_t^{(\alpha)}$ and $j \in 0,1,\ldots,2N_{t+1}^{(\alpha)}$. Since deme sizes may change from generation to generation, the resulting Markov chain is time-inhomogeneous. For the ancestral deme, d_0, the corresponding transition matrix is $\mathbf{Q}_{t\to t+1}^{(0)} = \left\{q_{ij}^{(0)}(t)\right\}$. The transition probabilities q_{ij} can be obtained by combining equations (5.8), (5.5) and (5.4). For the derived demes, they account for viability selection, migration and drift, while for the ancestral deme we do not need to consider migration (details in the Appendix). Multiplying the transition matrices $\mathbf{Q}_{t\to t+1}$ over the desired number of generations yields the transition probabilities over the total time span and therefore the joint probability distribution of going from state i at time t' to state j at time t'': $\mathbf{Q}_{t'\to t''} = \mathbf{Q}_{t'\to t'+1} \cdot \mathbf{Q}_{t'+1\to t'+2} \cdot \ldots \cdot \mathbf{Q}_{t''-1\to t''}$. Derived demes evolve from time t_α to time t_s, so that the transition matrix of interest is $\mathbf{Q}_{t_\alpha\to t_s}^{(\alpha)}$. Analogously, $\mathbf{Q}_{t_0\to t_f}^{(0)}$ is the transition matrix for the ancestral deme from t_0 to t_f.

When a derived deme d_α is founded by sampling from the ancestral deme, we express this in a similar way with a vector of transition probabilities $\mathbf{F}^{(\alpha)} = \left\{f_{kl}^{(\alpha)}\right\}$, where k is the number of A_1 alleles in the ancestral deme at time t_f and l is the number of copies among the founders of the derived deme at the time of founding. The $f_{kl}^{(\alpha)}$ are calculated from the binomial distribution (see Appendix). To speed up computation, we truncated the binomial distribution at the mean $\pm$ four times the standard deviation. We normalized the truncated distribution such that the total probability mass was equal to 1. This way, we did not spend time computing very low probabilities and were thus able to reduce the computation time by more than 50%. The error introduced by the truncation accumulated as we iterated the matrices, but it was negligible ($<$ 0.005) after twelve generations, which is more than the maximum number of iterations needed for this study (data not shown).

For the matrix iteration approach we simplified the full founding history of the derived demes. In reality, most derived demes received founder individuals in several years. These founding years may, but need not have taken place in consecutive years; there may be gaps when no individuals were released. Reflecting such details in the Markov implementation would have been tedious. Therefore, we determined one single point in time of establishment (t_α) per derived deme. We did so by defining as the year of establishment the year by which at least 50% of the total number of founders of a particular derived deme had been released. We choose the deme size in the year of establishment as the number of founders. For the founder individuals, we also dropped the distinction between males and females. Notice that we did not make these simplifications for the simulations resulting in the distribution of F_{ST} versus diversity, described in the following subsection. See Table 5.5 and Figure 5.2 for details of the founder events and demography.

We obtained the likelihood of the parameters given the observed allele frequencies, $\hat{\mathbf{p}}$, and the deme size trajectories, $\mathcal{N}$, as the probability of $\hat{\mathbf{p}}$ given the parameters and $\mathcal{N}$, *i.e.* $L(s,m,\phi,p_{\mathrm{init}};\hat{\mathbf{p}},\mathcal{N}) = \mathrm{P}[\hat{\mathbf{p}} \mid s,m,\phi,p_{\mathrm{init}},\mathcal{N}]$. Here and in the following, ϕ may be replaced by h depending on the dominance scheme (*cf.* equations (5.6) and (5.7)). Computing the joint likelihood surface on a dense four-dimensional parameter grid would be prohibitive. Therefore, we first limited the range of m to three distinct values $\{0.0,\ 0.1,\ 0.2\}$. Second, we set the dominance coefficient ϕ to values from 0.00 to 1.00 in steps of 0.125, and computed the joint

likelihood of s and p_{init} for each. We did so for values of s ranging from -1.0 to 1.0 in steps of 0.1 for over- and underdominance, and from 0.00 to 0.95 in steps of 0.05 for directional selection. For given values of m and ϕ, we computed $L(s, p_{\text{init}}; \hat{\mathbf{p}}, \mathcal{N}, m, \phi)$. The likelihood function and its derivation are given in the Appendix.

Spatial partition of variance in allele frequency

Selection in a spatial context may reduce or enhance genetic differentiation among demes at the gene of interest, depending on whether it is spatially homogeneous or heterogeneous. The impact of selection may be confounded by demography and gene flow. Beaumont and Nichols (1996) suggested plotting the distribution of the standardized variance across demes in allele frequency (F_{ST}) versus gene diversity between demes (heterozygosity) for a large set of neutral loci simulated under the island model (Wright 1931). Comparing this neutral distribution to the gene of interest may then reveal evidence for selection. Although results by Beaumont and Nichols (1996) were robust to a range of demographic deviations from the island model, we used a more realistic scenario to obtain the neutral distribution. The scenario reflects in detail the history of re-introduction of Alpine ibex into the Swiss Alps. We used our software SPoCS (`http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`) to simulate 10^5 replications of neutral evolution given this demography. For each replicate, we simulated one bialellic neutral marker. We drew the initial allele frequency from a uniform distribution between 0 and 0.5 and assumed no mutation and no migration. We followed Beaumont and Nichols (1996) in computing and plotting the distribution and the 95% and 50% quantiles for F_{ST}, and we used a sliding window comprising 10^4 points to estimate the quantiles.

5.3.3 Detecting short-term signals of viability selection

We performed standard statistical analyses to assess the correlation of age at sampling with genetic composition as a potential short-term (within-generation) signal of selection (Garrigan and Hedrick 2003). Since most of our samples are from harvested individuals, age at sampling in those cases coincides with age at culling. Age at sampling is not a direct measure of natural survival (Hadfield 2008), and it would be misleading to treat it as a response to the genetic composition in statistical analyses. Rather, we used it as an index of viability in the following sense: In response to viability selection, we expect the genetic constitution of individuals sampled at different ages to change as a function of the latter. It then makes sense to consider the genetic constitution as a response – in the statistical sense – to age at sampling. Throughout, we assume that culling (and hence sampling) and the genotype of an individual are uncorrelated. We further assume that the correlation between viability and genotype at OLADRB2 is constant across cohorts.

Correlation of age at sampling with zygosity and genotype

First, we investigated heterozygosity as a function of age at sampling, deme and sex by fitting Generalized Linear Models (GLM) with binomial error distribution (logistic regression). We compared models with various combinations of predictors and performed model selection based on the Akaike Information Criterion (AIC; Akaike 1974; Burnham and Anderson 2002) and on

the area under the Receiver Operating Characteristic curve (AUC; Fawcett 2006). We did this for OLADRB2, for the two other MHC-linked markers OLADRB1 and OMHC1, and for the putatively neutral markers. The latter is to make sure that the pattern observed at OLADRB2 is not due to a genome-wide heterozygosity effect. If more than two alleles were observed per locus, we did the analysis for all possible heterozygotes. For the neutral loci, we also computed a standardized version of multilocus-heterozygosity (Coltman et al. 1999a; Slate et al. 2004) and regressed it against the predictors. This is an alternative way of testing for a genome-wide heterozygosity effect. The multilocus-heterozygosity of individual i is obtained as

$$H_i = \frac{\sum_{l=1}^{L} h_{l,i}}{\sum_{l=1}^{L} \bar{h}_l}, \tag{5.9}$$

where $h_{l,i}$ is the heterozygosity (0 if homozygote and 1 if heterozygote) of individual i at locus l ($l = 1 \dots L$), and $\bar{h}_l = \frac{1}{k} \sum_{i=1}^{k} h_{l,i}$ is the mean heterozygosity across all individuals k typed at locus l.

Second, for OLADRB2 we extended the analyses from zygosity (heterozygous versus homozygous) to the full genotype, which may reveal more information about the dominance scheme. We fitted pairwise logistic regressions explaining the difference between any two of the three potential OLADRB2 genotypes in response to age at sampling, deme, sex and first-order interaction terms. Alternatively, we treated the genotype as a three-level response in a multinomial logistic regression, with age at sampling, deme and sex as predictors (again allowing for first-order interactions). We did all statistical analyses in `R` version 2.11 (R Development Core Team 2011), using the `aod` package (Lesnoff and Lancelot 2009) for Wald tests of significance in logistic regressions (Agresti 1990) and the `mlogit` package (Croissant 2008) for multinomial logistic regression (see SI for details).

Change in deviation from HWE as a function of age at sampling

Viability selection with under- or overdominance is expected to change the ratio of heterozygotes to homozygotes. To assess this, we pooled all individuals from different demes and then grouped them into two age classes, using the global median (5.25 years) as boundary. A total of 307 samples (138 females, 169 males) were included, of which 161 (74 females, 87 males) in age class 1 and 146 (64 females, 82 males) in age class 2. We computed the deviation from HWE of the proportion of heterozygotes, F_{IS}, for both age classes with `GENEPOP` version 4.0.9 (Raymond and Rousset 1995; Rousset 2008). We then computed the change in F_{IS} as a function of age, $\Delta F_{\mathrm{IS}} = F_{\mathrm{IS}}^{(\mathrm{age2})} - F_{\mathrm{IS}}^{(\mathrm{age1})}$, and compared ΔF_{IS} for OLADRB2 to the distribution of ΔF_{IS} for the 36 putatively neutral markers (BM1225 was monomorphic in data_mhc_14). Since we did not account for population structure, our estimates of F_{IS} are subject to the Wahlund effect (Wahlund 1928; Wright 1931). Because we are interested in the relative change in F_{IS} within one generation, not in the absolute values, this should not be a problem.

5.3.4 Estimating effective deme size from demographic data

In the matrix iteration approach, we have modeled genetic drift according to the idealized conditions of a Wright-Fisher population. This implies assumptions that cannot be justified

for Alpine ibex. Generations are overlapping, and the mating system does not result in equal contribution of parents to the gamete gene pool and random union of gametes. To account for these deviations, we calculated effective deme sizes N_e (Wright 1931) from demographic data, following Nunney (1993), and then used those in the matrix iteration approach (see SI for details).

5.4 Results

5.4.1 Evidence for viability selection, and its mode of dominance

Negative correlation between heterozygosity and age at sampling specific to OLADRB2

We found that the probability of an individual being heterozygous at OLADRB2 decreased with increasing age at sampling. This was independent of the exact structure of the GLM, as long as age at sampling was included as a predictor. Both criteria for model selection, AIC and AUC, yielded similar results (Tables 5.11, 5.12 and 5.13). For data_mhc_23, the best-compromise model between AIC and AUC was the one that includes all predictors, but no interaction terms. Age at sampling had a significant negative effect, -0.0625 (95% confidence interval: $[-0.1199, -0.0080]$) on the logit of the probability of an individual being heterozygous ($p \sim 0.0281$), deme had a marginally significant joint effect ($p \sim 0.063$) and sex had no significant effect. The overall effect of deme was caused by significantly positive effects of the levels Calanda, Macun, Safien-Rheinwald, Rothorn-Weissfluh, Wittenberg (all $p < 0.05$), and Cape au Moine ($p < 0.001$; see SI for details). These deme-specific effects are likely due to varying degrees of genetic drift to which the demes were exposed during re-introduction (Biebach and Keller 2009, 2010). For data_mhc_14, the best model was the one with age as the only predictor (Table 5.13). Age at sampling had a negative effect on the probability of being heterozygous (-0.0595 $[-0.1242, 0.0007]$ on the logit scale), but the effect was only marginally significant ($p \sim 0.0608$).

We also observed a negative effect of age at sampling on heterozygosity for allele 184 of the OLADRB1 locus, both for data_mhc_23 and data_mhc_14. The effect was significant for data_mhc_23 (-0.0557 on the logit-scale $[-0.1090, -0.0051]$, $p \sim 0.0351$), but not significant for data_mhc_14. The result is not surprising given that allele 184 is in strong linkage disequilibrium with allele 277 (the 'goat' alllele) of OLADRB2. None of the other alleles (174, 178, 170) of OLADRB1 showed a significant relationship between heterozygosity and age. We also observed no significant relation for the biallelic MHC-linked marker OMHC1.

For the majority of alleles at the putatively neutral markers (37 in case of data_mhc_23, 36 for data_mhc_14), we observed no statistical correlation between heterozygosity and age at sampling. However, a small number of alleles showed a significant correlation. We observed both positive and negative correlations, and it remains to be shown if these just correspond to the proportion of false positives to be expected under neutrality, or if they reflect effects of selection (see SI for details). None of these alleles is in linkage disequilibrium with the 'goat' allele of OLADRB2. Overall, there is therefore no evidence for a genome-wide correlation of heterozygosity with age at sampling. This was confirmed by regression of multilocus-heterozygosity against age at sampling, deme and sex: There was no significant effect of age at sampling on multilocus-heterozygosity in both data sets (see SI for details).

Reduced age at sampling of heterozygotes compared to ibex homozygotes

Since results from the multinomial logistic regression were essentially the same for data_mhc_23 and data_mhc_14, we only state those for the latter here and present results for data_mhc_23 in the SI. Five demes (Calanda, Macun, Oberalp-Frisal, Safien-Rheinwald and Rothorn-Weissfluh) had a positive single-level effect on the odds of genotype A_2A_2 relative to A_1A_2 (data not shown). We therefore modified the original factor deme with 14 levels to a factor with only two levels, one containing the five demes just mentioned (deme set 1), and the second containing all other demes (deme set 2). Within deme set 2, the probability of an individual having genotype A_2A_2 relative to A_1A_2 increased significantly as a function of age at sampling (0.1460 [0.0138, 0.2782] on the logit-scale, $p \sim 0.0304$; Table 5.2 and Figure 5.3). There was no effect of age at sampling within deme set 1. The contrasts between the other pairs of genotypes (A_1A_1 versus A_1A_2, and A_1A_1 versus A_2A_2) were not significantly influenced by age at sampling. The results from the pairwise logistic regressions supported this finding: The best model was one that distinguished between two sets of demes (see Tables 5.26 to 5.28 in SI for details on model selection). For the subset consisting of demes Macun, Oberalp-Frisal and Rothorn-Weissfluh, the probability of an individual being A_2A_2-homozygous compared to heterozygous was significantly lower compared to the other demes (-2.0652 [-3.2153, -0.9923] on the logit-scale, $p \sim 0.0002$; Table 5.28). *Within* this subset of demes, however, the probability that an individual is A_2A_2-homozygous compared to heterozygous increased significantly as a function of age at sampling (0.2142 [0.0912, 0.0508], $p \sim 0.0188$; Table 5.28). In summary, we confirmed the negative correlation between age at sampling and heterozygosity at OLADRB2, at least for subsets of demes for which there was enough statistical power. However, we obtained no further insight into the effect of age at sampling on the ratios of A_1A_2 to A_1A_1 and of A_2A_2 to A_1A_1. Hence, we remain uncertain about the mode of dominance. Both, underdominance or directional selection against the 'goat' allele (intermediate dominance including full recessivity), are compatible with the short-term signals of viability selection (Figure 5.3). Remember, however, that viability is only one aspect of fitness, and that we have ignored sexual selection.

Increase in F_{IS} at OLADRB2 as a function of age at sampling

Pooling samples from different demes, we found that the change in deviation from HWE as a function of age at sampling, $\Delta F_{\mathrm{IS}} = F_{\mathrm{IS}}^{(\mathrm{age2})} - F_{\mathrm{IS}}^{(\mathrm{age1})}$, was strongly positive for OLADRB2. ΔF_{IS} for OLADRB2 was slightly beyond the upper limit of the distribution of ΔF_{IS} obtained for 36 neutral microsatellites (Figure 5.4). Hence, the proportion of OLADRB2 heterozygotes dropped significantly as a function of age at sampling. The mean ΔF_{IS} for the neutral loci was not different from 0 (one sample t-test, $t = 0.346$, df $= 35$, $p > 0.7$). This applies analogously to data_mhc_23 (Figure 5.18). Overall, these results confirm the previous findings of a decrease in heterozygosity at OLADRB2 as a function of age at sampling.

Medium-term evidence for spatially homogeneous selection

The short-term signals of viability selection reported above build up within one generation, but are wiped out by mating and reproduction. In the following, we turn to medium-term signals, which accumulate across generations and are potentially reflected in the observed allele frequencies. For data_mhc_14 we investigated the genetic differentiation across demes at OLADRB2

measured in terms of F_{ST} and compared it to the distribution expected under neutrality, as well as to the other markers. Accounting for the dependence of F_{ST} on total diversity, we found that the observed spatial differentiation at OLADRB2 is low compared to a large number of neutral loci simulated under the demographic scenario of Alpine ibex (Figure 5.5). F_{ST} for OLADRB2 was within, but close to the lower bound of, the interval between the empirical 5%

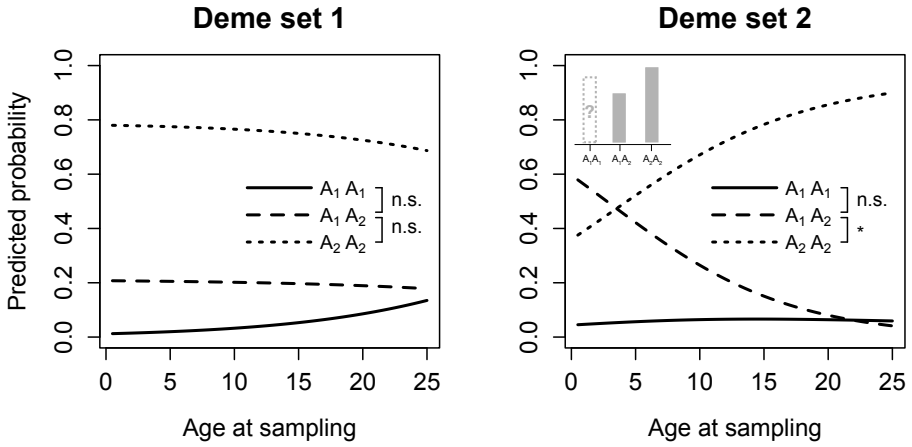

Figure 5.3: Predicted probability of the three OLADRB2 genotypes as a function of age at sampling. The predictions were obtained from the multinomial logistic regression model given in Table 5.2. The right figure is for demes in set 2 = {Calanda, Macun, Oberalp-Frisal, Safien-Rheinwald, Rothorn-Weissfluh}, the left for all other demes in data_mhc_14. A_1 denotes the 'goat' allele, A_2 the ibex allele. For demes in set 2, the odds of A_2A_2 versus A_1A_2 increase significantly as a function of age at sampling ($p \sim 0.03$; Table 5.2). The small plot in the right figure qualitatively illustrates the range of compatible dominance schemes, namely underdominance or selection against A_1 with intermediate dominance.

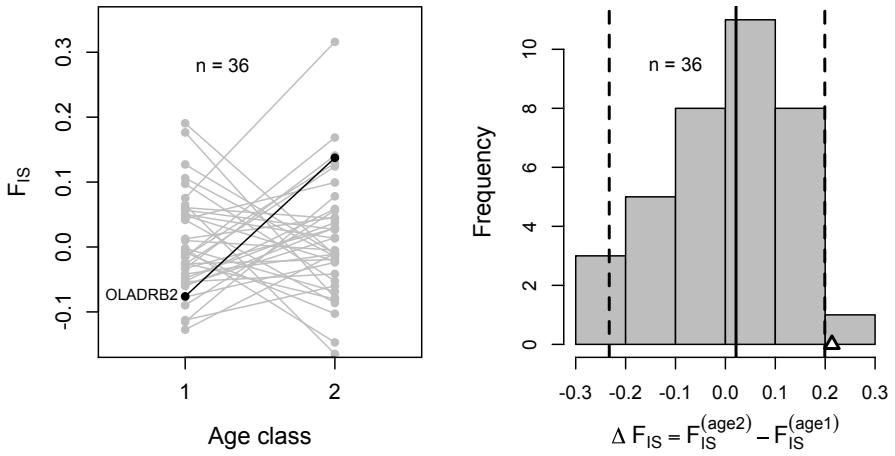

Figure 5.4: Change in deviation from HWE measured by F_{IS} as a function of age. Age class 1 comprises individuals of age up to and including 5.25 years, and age class 2 those of age older than 5.25 years. (A) Gray symbols belong to 36 neutral microsatellites, and the black symbols represents the MHC-linked marker OLADRB2. (B) The change in F_{IS} as a function of age (ΔF_{IS}) for OLADRB2 (triangle) is compared to the distribution of ΔF_{IS} obtained for the 36 neutral markers. Vertical lines represent the median (solid) and the 5% and 95% quantiles (dashed) of the neutral distribution. Plots are shown for the data set data_mhc_14 (see text for details and Figure 5.18 for analogous plots for data_mhc_23).

Table 5.2: Estimates of effects on the contrast between all three OLADRB2 genotypes from a multinomial logistic regression for data_mhc_14.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p
gtp.ola2 $\sim$ age + I(deme $\in \mathcal{D}$) + age:I(deme $\in \mathcal{D}$)						
	a1a1	−2.8502	0.8741	−4.5635	−1.1370	0.0011 **
	a2a2	1.3511	0.3022	0.7588	1.9434	<0.0001 ***
	a1a1:age	0.1025	0.0888	−0.0715	0.2766	0.2481
	a2a2:age	−0.0068	0.0404	−0.0860	0.0725	0.8673
	a1a1:I(deme $\in \mathcal{D}$)	0.2485	1.1657	−2.0362	2.5332	0.8312
	a2a2:I(deme $\in \mathcal{D}$)	−1.8438	0.4697	−2.7645	−0.9231	<0.0001 ***
	a1a1:age:I(deme $\in \mathcal{D}$)	0.0158	0.1349	−0.2487	0.2802	0.9069
	a2a2:age:I(deme $\in \mathcal{D}$)	0.1460	0.0675	0.0138	0.2782	0.0304 *

Model, the multinomial logistic regression model fitted to explain the odds of the three OLADRB2 genotypes (gtp.ola2 $\in \{a1a1, a1a2$ and $a2a2\}$) as a function of the predictors; Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code: *** for $0 < p \leq 0.001$, ** for $0.001 < p \leq 0.01$, * for $0.01 < p \leq 0.05$, · for $0.05 < p \leq 0.1$ and ' ' for $0.1 < p \leq 1$. I(deme $\in \mathcal{D}$) = 1 if deme $\in \mathcal{D}$, and I(deme $\in \mathcal{D}$) = 0 if deme $\notin \mathcal{D}$. The set $\mathcal{D}$ contains the demes with a significant single-level effect on the genotype in a more extended model: Calanda, Macun, Oberalp-Frisal, Safien-Rheinwald and Rothorn-Weissfluh. The estimates for I(deme $\in \mathcal{D}$) are given for the effect of I(deme $\in \mathcal{D}$) = 1 compared to the default I(deme $\in \mathcal{D}$) = 0. All estimates are relative to the heterozygous genotype A_1A_2 (gtp.ola2 = a1a2).

and 95% quantiles. It was also lower than the one for the majority of the other markers (Figure 5.5A). The result suggests that, if OLADRB2 is under selection, the selection pressure is more likely to be spatially homogeneous than heterogeneous (Lewontin and Krakauer 1973, 1975; Beaumont and Nichols 1996). In addition, the observed F_{ST} was very high compared to the neutral expectation for the putatively neutral markers BM1415 and SR-CRSP08, and very low for INRABERN172, TGLA73, BM302, TGLA10, OARHH35 and OARFCB20 (Figure 5.5A). These markers might in fact not be neutral, but under spatially heterogeneous or homogeneous selection, respectively. If loci are not artificially made biallelic, the effect vanishes for BM302 (4 alleles) and TGLA10 (3), however.

5.4.2 Likelihood-based estimates of strength of selection

Recall that for the matrix iteration approach we needed estimates of effective deme sizes that account for demography, life cycle and mating system in Alpine ibex. Table 5.5 summarizes in the outer right column the trajectories of per-generation effective sizes for each deme, estimated from demographic data. The matrix iteration approach allowed us in principle to obtain the joint likelihood of the selection (s) and dominance (ϕ or h) coefficient, the migration rate (m) and the initial frequency (p_{init}) of the 'goat' allele A_1. In practice, there is a high computational cost for evaluating the likelihood on a dense enough grid, and it is not obvious how to present four-dimensional joint likelihood surfaces. Therefore, we start without migration. Further, we only consider a small set of values for the dominance coefficients. For each dominance coefficient, we obtained the joint likelihood surface of s and p_{init}. Since s is of most interest, whereas p_{init} is a confounding parameter, we were mainly interested in the marginal likelihood of s with respect to p_{init}. At the end, we will assess the effect of gene flow.

Under- and overdominance without migration

In the case of under- and overdominance (where ϕ is the dominance coefficient; *cf.* equation (5.6)) there was most support for ϕ around 0.125 (Tabe 5.3). Given $\phi = 0.125$, the marginal

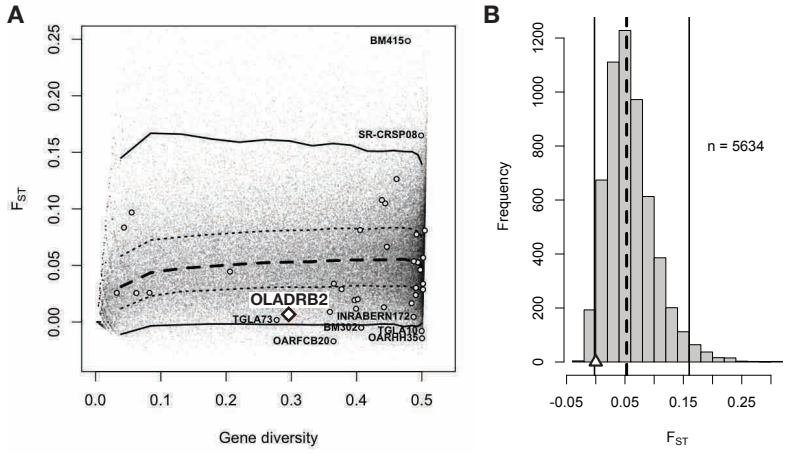

Figure 5.5: The distribution of the standardized variance in allele frequency across 14 derived demes (F_{ST}) as a function of gene diversity (expected between-deme heterozygosity). (A) The point cloud represents the distribution of values simulated under neutrality for the demographic scenario of Alpine ibex ($n \approx 9 \cdot 10^4$ biallelic loci). The median, 25% and 75%, and 75% and 95% quantiles are given by the bold dashed line, the thin dashed lines and the solid line, respectively. F_{ST} and gene diversity observed for real loci are given by the diamond for OLADRB2 and the circles for the other MHC-linked markers (OLADRB1, OMHC1) and the 36 neutral markers. Names of markers with extreme values are shown. For loci with more than two alleles, we pooled all alleles except the one with the major allele frequency. (B) The F_{ST} distribution conditioning on the gene diversity being equal to the one observed for OLADRB2 (0.296) $\pm$ 0.02 ($n = 5634$ simulations). This corresponds to taking a vertical slice through the plot in (A) at a gene diversity of about 0.3. The triangle represents OLADRB2 and the vertical lines have the same meaning as the respective lines in (A).

maximum likelihood estimate (MLE) of s was 0.974, which suggests overdominance ($s > 0$). Table 5.3 summarizes the results for various values of ϕ and gives 95% credible intervals for s. Figure 5.6 shows the respective marginal likelihood curves of s. Interestingly, the observed allele frequency spectrum contained quite some information about p_{init} (Figure 5.19). Therefore, in Table 5.3 we also provide joint maximum-likelihood estimates of s and p_{init}. Comparing the marginal and joint estimates of s, there is considerable difference for $\phi = 0$ and $\phi = 0.125$, but for $\phi \geq 0.25$, the two estimates become very similar. Although some point estimates of s were relatively high, $s = 0$ was included in all 95% credible intervals. Therefore, a drift-only scenario cannot be excluded for any ϕ considered here. Figure 5.7 further illustrates the relation between s and p_{init}. First, it is obvious that if we had a precise point estimate of p_{init}, we would be able to make more precise inference on s and, in some cases, to exclude a drift-only scenario. Second, it illustrates how crucial the effect of p_{init} is on the point estimate of s. Given ϕ, the conditional MLE of s may be negative or positive, depending on p_{init} (*e.g.* Figure 5.7A).

Conditioning on underdominance without migration

Taken on its own, the matrix iteration approach yielded most support for asymmetric overdominance, with an equilibrium frequency of the 'goat' allele A_1 of about 0.125. But Figure 5.6 and Table 5.3 show that the observed allele frequency spectrum may also be explained by other scenarios. Without knowing the true dominance coefficient in advance, choosing the one with highest marginal likelihood seems justified (Table 5.3). In our case, however, the short-term

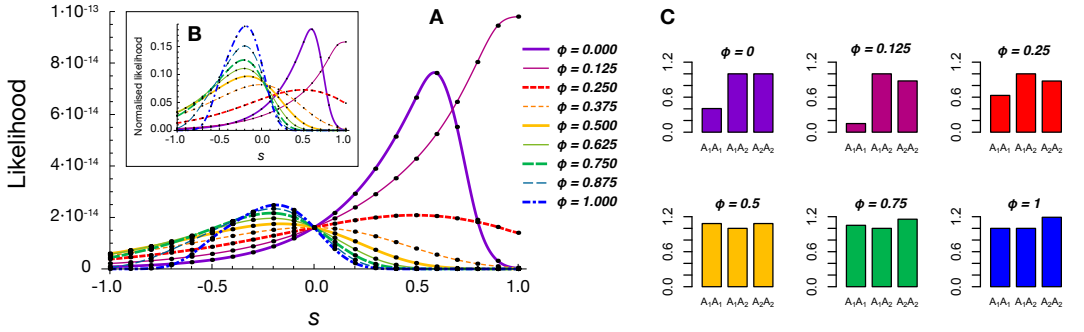

Figure 5.6: Likelihood of the selection coefficient s for various dominance coefficients ϕ without migration. The likelihood curves are marginal with respect to the initial frequency p_{init} of the 'goat' allele A_1. Fitnesses are parameterized as in equation (5.6) in the text. A_1 is fully recessive if $\phi = 0$ and fully dominant if $\phi = 1$. For $\phi \notin \{0, 1\}$, there is overdominance if $s > 0$ and underdominance if $s < 0$. (A) The likelihoods are not normalized. Therefore, the areas under the curves indicate the relative support for the respective values of ϕ (*cf.* Table 5.3). (B) As in (A) but with likelihoods normalized such that the area under the curve is 1. In a Bayesian view, these curves correspond to the posterior distribution of s given a uniform prior on the normal scale. The curves in (A) and (B) were obtained by third-order interpolation of points computed for values of s on a grid from –1.0 to 1.0 with step size 0.1 (black dots). (C) Relative fitnesses of the three genotypes for some values of ϕ and the respective MLE of s. The plot in the middle of the top row corresponds to the most likely scenario.

signals of selection provide an indicator for the mode of dominance. They suggest a disadvantage of heterozygotes compared to the ibex-homozygotes, which rules out overdominance. In the following, we therefore conditioned on underdominance, including the two marginal cases of full recessivity and full dominance of A_1. Figure 5.10 and Table 5.7 show that, in this case, the relative support for different values of ϕ was similar, with $\phi = 0.75$ being slightly preferred. For $\phi \in \{0, 0.125, 0.25, 0.375\}$, the MLE of s was 0, suggesting that drift-only is most likely for

Table 5.3: Likelihood-based estimates of selection (s) and dominance (ϕ) coefficient with under- or overdominance, without migration.

Dominance scheme	ϕ	L_ϕ [a]	B.F.	$\hat{s}_\phi$	HPD	$\{\widehat{s, p_{\text{init}}}\}_\phi$
A_1 fully recessive	0.000	4.210	0.680	0.595	$(-0.306, 0.869)$	$\{0.50, 0.36\}$
Overdom. if $s > 0$, underdom. if $s < 0$	0.125	6.193	1.000	0.974	$(-0.352, 1.000)$	$\{0.79, 0.30\}$
·	0.250	2.877	0.465	0.493	$(-0.718, 1.000)$	$\{0.50, 0.14\}$
·	0.375	1.970	0.318	-0.017	$(-0.990, 0.608)$	$\{0.01, 0.18\}$
·	0.500	1.811	0.292	-0.165	$(-0.995, 0.344)$	$\{-0.10, 0.21\}$
·	0.625	1.776	0.287	-0.205	$(-0.971, 0.239)$	$\{-0.20, 0.54\}$
·	0.750	1.721	0.278	-0.209	$(-0.930, 0.196)$	$\{-0.20, 0.29\}$
·	0.875	1.550	0.250	-0.200	$(-0.790, 0.179)$	$\{-0.20, 0.31\}$
A_1 fully dominant	1.000	1.332	0.215	-0.190	$(-0.613, 0.142)$	$\{-0.20, 0.35\}$

$L_\phi = \sum_{s \in \mathcal{S}} L(\phi, s; D) = \sum_{s \in \mathcal{S}} P(D|\phi, s)$ is an approximation to the marginal likelihood of ϕ, $L(\phi; D) = P(D|\phi) = \int_{\mathcal{S}} P(D|\phi, s)P(s|\phi)\mathrm{d}s = \int_{\mathcal{S}} P(D|\phi, s)P(s)\mathrm{d}s$, where $\mathcal{S}$ is the set of possible values for s, and the last equality holds because ϕ and s are independent. The Bayes Factor (B.F.) is here defined as $L_\phi/\max(L_\phi)$, and therefore denotes the support for any model compared to the one with the maximum marginal likelihood ($\phi = 0.125$ in this case). The maximum-likelihood estimate (MLE) of s given ϕ is provided by $\hat{s}_\phi$. In a Bayesian perspective, this is equal to the posterior mode, since the prior was uniform on the normal scale. HPD, 95% highest posterior density interval of s. Point and interval estimates correspond to likelihood curves displayed in Figure 5.6. These were obtained after marginalizing out the initial allele frequency p_{init}. The last column gives the joint MLE of s and p_{init}, which is obtained if p_{init} is not marginalized out (*cf.* Figure 5.19 for the full likelihood surface).

[a] In multiples of 10^{-13}

these dominance coefficients. For $\phi \in \{0.5, 0.625, 0.75, 0.875, 1\}$, the MLE of s was negative, which implies underdominance. Moreover, A_1 is preferred in these cases whenever its frequency is larger than the respective value of ϕ, and disfavored if its frequency is below ϕ. However, the 95% credible interval included $s = 0$ in all cases, such that drift-only cannot be excluded.

Intermediate dominance without migration

The short-term signals of viability selection were either compatible with underdominance, or with intermediate dominance and directional selection against the 'goat' allele A_1 (*cf.* Figure 5.3B). In the previous paragraph, we have dealt with medium-term evidence in the case of underdominance. We will now turn to the case of intermediate dominance. Recall that for this case we parameterized the relative fitnesses as in equation (5.7), with dominance coefficient h, where $0 < h < 1$ and $0 \leq s \leq 1$. Hence, A_1 is partially recessive if $0 < h < 0.5$ and partially dominant if $0.5 < h < 1$. There is no dominance if $h = 0.5$. For the values of h assessed, we found most support in terms of the marginal likelihood for $h = 0.125$, and decreasing support for increasing values of h (Figure 5.8A and Table 5.4). Marginal likelihood curves and posterior distributions of s are shown for various h in Figures 5.8A and 5.8B, respectively. The MLE of s given $h = 0.125$ was 0.5, and the 95% credible interval did not include $s = 0$ (Table 5.4). Hence, we found support for medium-term negative selection against A_1, with A_1 being partially recessive. Although the relative support for higher values of h decreased, as long as $h < 0.5$, the 95% credible interval for s did not incude $s = 0$, and for no dominance ($h = 0.5$),

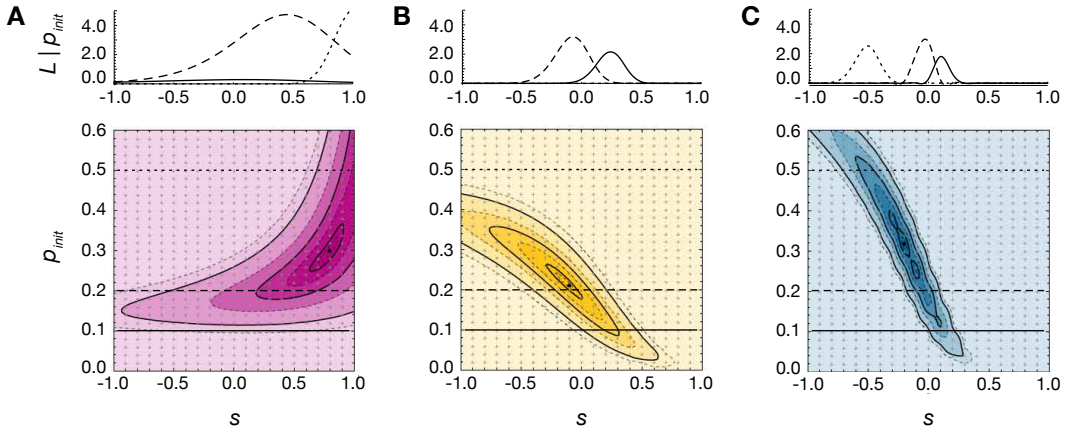

Figure 5.7: The effect of the initial allele frequency p_{init} on the likelihood of the selection coefficient s, for under- and overdominance. The joint likelihood surface of p_{init} and s is shown in the bottom row (third-order interpolation was applied). The top row shows the conditional likelihoods $L|p_{\text{init}}$ of s given three specific values of p_{init} (solid line: 0.1, dashed line: 0.2, dotted line: 0.5). This corresponds to taking horizontal slices from the surface plots at respective positions. Fitnesses are parameterized as in equation (5.6) in the text. The dotted and solid black lines denote regions of highest posterior density for levels of support of 99%, 95%, 75%, 50%, 25%, 5% and 1%. Crosses denote parameter combinations for which exact values were computed, and the surface was obtained by third-order interpolation. (A) $\phi = 0.125$ (B) $\phi = 0.5$ and c) $\phi = 0.875$. The corresponding marginal likelihood curves in Figure 5.6 are obtained by summing over the range of possible values of p_{init}, $0.0 < p_{\text{init}} < 0.6$. $L|p_{\text{init}}$ is shown in units of 10^{-15}. In (B) the dotted line coincides with the abscissa and is therefore hardly visible. See Figure 5.19 for more values of ϕ.

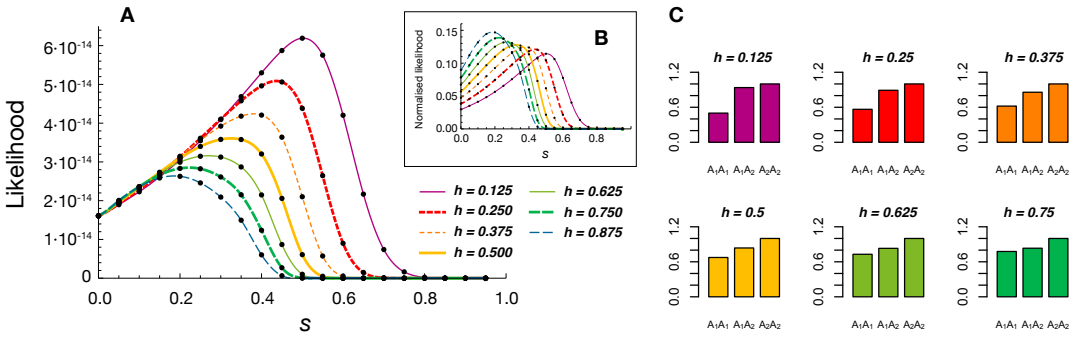

Figure 5.8: Likelihood of the selection coefficient s for various dominance coefficients h without migration. The likelihood curves are marginal with respect to the initial frequency p_{init} of the 'goat' allele A_1. Fitnesses are parameterized as in equation (5.7). For $0 < h < 1$ (and $0 \leq s \leq 1$, as is the case here), dominance is intermediate. A_1 is partially recessive if $0 < h < 0.5$ and partially dominant if $0.5 < h < 1$; there is no dominance if $h = 0.5$. The limiting case of full recessivity of A_1 ($h = 0$) is equivalent to the case of $\phi = 0$ in Figure 5.6 and therefore not plotted again. (A) The likelihoods are not normalized and the areas under the curves indicate the relative support for the various values of h (*cf.* Table 5.4). (B) As in (A) but with likelihoods normalized such that the area under the curve is 1. In a Bayesian view, these curves correspond to the posterior distribution of s given a uniform prior on the normal scale. The curves were obtained by third-order interpolation of points computed for values of s on a grid from 0.0 to 0.95 with step size 0.05 (black dots). (C) Relative fitnesses of the three genotypes for some values of h and the respective MLE of s. The top left plot corresponds to the most likely scenario. Notice the similarity with the most likely scenarios in Figure 5.6C ($\phi = 0$ and 0.125).

the lower bound of the interval was at 0.008. However, for partial dominance, the drift-only case ($s = 0$) could not be excluded (Table 5.4). Table 5.4 also shows the joint maximum-likelihood estimates of s and p_{init}. Both, the joint and marginal MLE of s are similar for all values of h. While the estimate of s decreased monotonously with increasing h, the estimate of p_{init} first increased, with a maximum of 0.42 for $h = 0.25$, and then decreased h increased further. The effect on the estimate of s caused by marginalizing out p_{init} is illustrated in Figure 5.9. As for under- and overdominance (Figure 5.7), knowing p_{init} would allow for preciser and more accurate inference on s.

The effect of gene flow via migration

The main effect of gene flow via migration was to reduce the marginal likelihood of s compared to the case with $m = 0$. Second, migration tended to smooth out the likelihood curves (Figure 5.11A for intermediate dominance and Figure 5.25A for under- and overdominance). The effect on the mode of the posterior was minor (Figures 5.11B and 5.25B). These results are expected, since if selection is spatially homogeneous, and migration happens after selection, gene flow from the common migrant pool supports the effect of selection and balances differences among demes. As a consequence, weaker selection is needed to achieve the same net change in allele frequency as in the case without migration. The highest migration rate we considered is $m = 0.2$. Some of our previous analyses not reported here suggest that migration rates between Alpine ibex demes in the Swiss Alps do not exceed this value (see chapter 3).

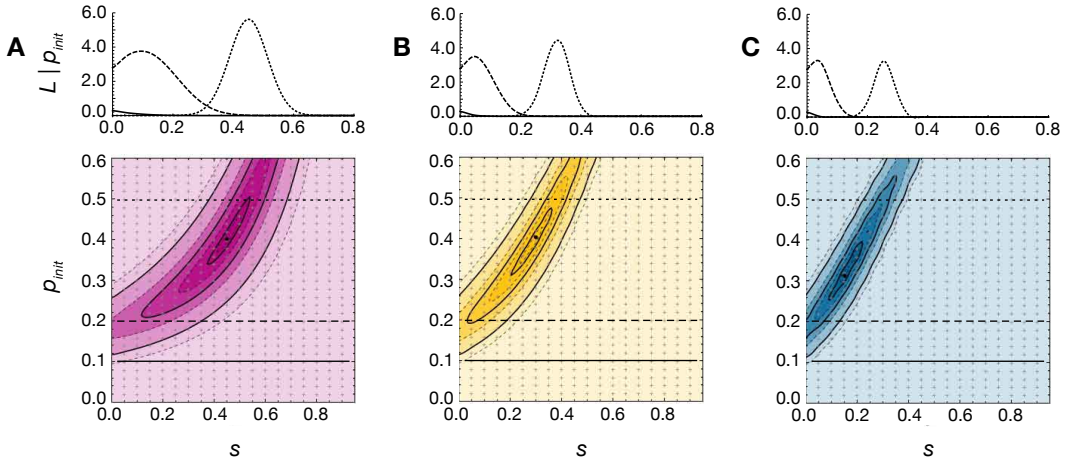

Figure 5.9: The effect of the initial allele frequency p_{init} on the likelihood of the selection coefficient s, for intermediate dominance. Fitnesses are parameterized as in equation (5.6) in the text. (A) $\phi = 0.125$ (B) $\phi = 0.5$ and c) $\phi = 0.875$. The corresponding marginal likelihood curves in Figure 5.8 are obtained by summing over the range of possible values of p_{init}, $0.0 < p_{\text{init}} < 0.6$. $L|p_{\text{init}}$ is shown in units of 10^{-15}. Further details as in Figure 5.9. See Figure 5.22 for intermediate values of h.

Summary

The results on the medium-term signals of selection can be summarized as follows. We distinguished between i) under- and overdominance, and ii) intermediate dominance. For i), we further distinguished between i.a) not conditioning on short-term evidence, and i.b) conditioning on the short-term evidence. For ii), no further distinction was needed, because this case was in agreement with short-term evidence. In case i.a), the vector of observed allele frequencies was best explained by overdominant selection, with equilibrium frequency around 0.125, and $\hat{s}_\phi \approx 0.97$. For i.b) – excluding overdominance as suggested by the short-term analyses – we found weak support for underdominance and an unstable internal equilibrium ϕ at about 0.75. For $\phi < 0.5$, there was most support for a drift-only scenario ($\hat{s}_\phi = 0$), although these cases

Table 5.4: Likelihood-based estimates of selection (s) and dominance (h) coefficient with intermediate dominance, without migration.

Dominance scheme	h	L_h[a]	B.F.	$\hat{s}_h$	HPD	$\{\widehat{s, p_{\text{init}}}\}_h$
A_1 partially recessive	0.125	5.415	1.000	0.500	(0.035, 0.661)	{0.45, 0.40}
·	0.250	4.202	0.776	0.435	(0.025, 0.574)	{0.40, 0.42}
·	0.375	3.397	0.627	0.380	(0.016, 0.512)	{0.35, 0.41}
No dominance	0.500	2.822	0.521	0.326	(0.008, 0.463)	{0.30, 0.40}
A_1 partially dominant	0.625	2.391	0.442	0.271	(0.000, 0.421)	{0.25, 0.37}
·	0.750	2.057	0.380	0.223	(0.000, 0.394)	{0.20, 0.35}
·	0.875	1.790	0.331	0.186	(0.000, 0.368)	{0.15, 0.31}

L_h, the Bayes Factor (B.F.), $\hat{s}_h$, HPD and $\{\widehat{s, p_{\text{init}}}\}_h$ are as in Table 5.3, with ϕ replaced by h. Point and interval estimates correspond to likelihood curves displayed in Figure 5.8. These are obtained after marginalizing out the initial allele frequency p_{init}. The last column gives the joint MLE of s and p_{init}, which is obtained if p_{init} is not marginalized out. For full recessivity and full dominance of the 'goat' allele, see Table 5.3.
[a] In multiples of 10^{-13}

had slightly less overall support than cases with $\phi \geq 0.50$ and $\hat{s}_\phi < 0$. For directional selection (ii), we found relatively strong support for partial recessivity of the 'goat' allele, with $\hat{s}_h \approx 0.5$. Relative support decreased with increasing degree of dominance h. A drift-only scenario could be excluded for $h < 0.5$, but not for $h \geq 0.5$. Importantly, the relative fitnesses implied by the most likely scenarios of i.a) and ii), respectively, are not so different (compare Figures 5.6C and 5.8C). Both clearly suggest lowest fitness for the 'goat' homozygote (A_1A_1) and heterozygote fitness close to that of 'ibex' homozygotes A_2A_2. In case i.a), the heterozygotes are as fit ($\phi = 0$) or slightly fitter ($\phi = 0.125$), in case ii) they are slightly less fit than A_2A_2 ($h = 0.125$). The truth might be somewhere in between, since we have only considered discrete values of ϕ and h. Further, it turned out that integrating out the initial allele frequency p_{init} could also be avoided; the vector of observed allele frequencies allowed for a joint estimation of s and p_{init}. The marginal and joint point estimates of s were similar in most cases. The effect of migration on inference of s was weak, as long as m was not too high.

5.5 Discussion

In the following, we first discuss biological implications and address the apparent contradiction between short- and medium-term evidence. We then revisit some of our assumptions and discuss advantages and limitations of our approach.

5.5.1 Biological implications

Linking results to the biological function of MHC

Using the matrix iteration approach exclusively we found strongest support for asymmetric overdominance. In contrast, the short-term results ruled out overdominance, but suggested underdominance or intermediate dominance. Intermediate dominance in the form of directional selection against the 'goat' haplotype was also well supported by the matrix iteration approach. This is compatible with the hypothesis of selection varying in space or time. Since we found lower than expected variance in allele frequency across demes, selection seems to be uniform in space. Selection varying in time remains as a potential mechanism for explaining the observed pattern of genetic diversity at the MHC locus. However, to confirm this, we would need samples from more than one point in time. Such samples are currently available only to a limited extent for the demes studied here (Biebach and Keller 2010). Therefore, while we found evidence for selection acting on DRB, our results do not provide conclusive evidence for the underlying mechanism. Combining short- and medium-term evidence, it seems that the DRB allele shared with domestic goat has been selected against during about the last ten generations (1920 until 2006). Whether this is in the form of balancing selection with very low equilibrium frequency or strictly directional selection remains open. Similarly, whether or not the disadvantage of the 'goat' allele is permanent or transient is not evident. Selection in favour of or against particular MHC alleles has been reported before in free-living populations (*e.g.* Paterson et al. 1998; Meyer-Lucht and Sommer 2005) or humans (*e.g.* Thursz et al. 1997), and is a plausible scenario given the role of MHC in pathogen defense (Sommer 2005; Radwan et al. 2010, but see Black and Hedrick (1997)).

A negative correlation between age at sampling and heterozygosity at DRB has been suggested before by Ch. Grossen (personal communication). Here, we have confirmed this partly by statistical analyses. Heterozygotes were on average younger at culling compared to 'ibex' homozygotes (A_2A_2). But we found no significant difference between heterozygotes and 'goat' homozygotes (A_1A_1). We have interpreted this as a signal of viability selection on a short time scale, either in the form of heterozygote disadvantage or directional selection against the 'goat' allele (A_1). This interpretation is indirectly supported by the trend for an increase in nematode fecal egg counts with increasing population frequency of the 'goat' allele reported by Grossen (2005). If heterozygotes carry more parasites, they may accumulate higher mortality and therefore be underrepresented among individuals of high age at sampling.

We found low spatial differentiation for OLADRB2 among demes (Figure 5.5) and concluded that selection is homogeneous in space. This might be a consequence of a spatially homogeneous parasite community. In contrast, Grossen (2005) reported higher F_{ST} for MHC-linked markers compared to neutral ones. This would argue for spatially heterogeneous selection, possibly due to adaptation to genetically different parasite populations. The contradictory results may be explained by the fact that not exactly the same sets of demes were analyzed in the two studies. While the 14 focal demes of our study are all geographically close, Grossen (2005) studied fewer (six) demes, and these were on average further appart from each other. The difference in spatial scale between the two studies might correlate with changes in the selection scheme.

How do our maximum likelihood estimates of s (0.97 for $\phi = 0.125$; 0.5 for $h = 0.125$) compare to those in previous studies? Aguilar et al. (2004) reported similar values (0.5–0.95) for DRB for the San Nicolas Island fox population. Studies on various MHC genes in humans found estimates that vary over several orders of magnitude: 0.05–0.605 (Black and Hedrick 1997), 0.0007–0.042 (Satta et al. 1994), and 0.022 (Currat et al. 2010). Overall, our estimates seem high, but are plausible. For example, the observed change in the proportion of heterozygotes between age classes 1 and 2 (Figure 5.4) is about 1.3 times the one expected with $s = 0.5$ and $h = 0.125$ under a deterministic model. van Oosterhout (2009) recently suggested a model that complements the hypothesis of overdominance or spatio-termporally varying selection. It incorporates purifying selection on recessive deleterious mutations linked to the MHC. Lower selection coefficients are needed at the MHC to explain the same data that, under the traditional balancing selection hypothesis, would lead to very high estimates of s. On the other hand, the diffusion approximation suggests that, for selection to leave a signal in populations of low effective size, s must be high relative to $1/N_e$, since it is $2N_es$ that determines the effective strength of selection (Wright 1945; Charlesworth and Charlesworth 2010). In our case, deme-specific values of N_e are low, namely of order 10^2 (Table 5.5). It is therefore not surprising that we could reliably detect only moderate to strong selection (s of order 10^{-2} and higher) in our medium-term analysis.

Explanations for the disparity between short- and medium-term evidence

Disparity between short- and medium-term evidence of selection is common in empirical studies and there are various explanations (Arnold 1992; Coltman et al. 2001; Merilä et al. 2001; Kruuk et al. 2002; Wilson et al. 2005b). First, as mentioned earlier, the distribution of allele

frequencies may not be sufficient for inference on the model parameters. Several evolutionary scenarios might lead to similar distributions. Although our matrix iteration approach uses the whole distribution for inference, the differences might be hard to detect with data from only thirteen derived demes and incomplete sampling within demes. Ideally, one would like to combine the likelihoods of the short- and medium-term analyses to obtain a single measure. This is compromised in our case by a fact that, on its own, provides a second explanation: While the short-term analysis focussed on viability – which is only one aspect of fitness – the matrix iteration also captures other aspects of fitness. Although heterozygotes seemed less viable, they might be more fertile. In the medium-term, this might compensate lower viability and cause the apparent contradiction. As a third explanation is related to this: selection on correlated traits may impose a constraint on the locus of interest (Merilä et al. 2001). For instance, if exon 2 of DRB were linked to another gene or a QTL with antagonistic effects on lifetime fitness, short- and medium-term signals do not need to be consistent. A striking example of such a constraint in a free-living species was recently documented by Gratten et al. (2008) for coat color in Soay sheep on St. Kilda. Dark color is genetically associated with larger body size (Clutton-Brock et al. 1997; Gratten et al. 2008), which in turn is heritable (Wilson et al. 2005a, 2007) and positively correlated with survival (Wilson et al. 2005b) and reproductive success (Coltman et al. 1999b). Yet, the population frequency of dark color decreased over a period of 20 years. Gratten et al. (2008) showed that this is due to linkage between the color locus and another QTL with negative effect on lifetime fitness. This imposes a direct fitness cost on the dark allele that outweighs the expected benefit of being larger. Fourth, a change over time in the selection regime – both in the strength or the optimum phenotype – may explain disagreement between short- and medium-term evidence (Merilä et al. 2001). For example, selection may depend on population density (*e.g* Coltman et al. 1999b; Cutrera and Lacey 2006) or alter with changes in the environment (*e.g.* Haldane 1924; Steward 1977; Merilä et al. 2001; Cook 2003).

A combination of these explanations might apply in our case. We have merely relied on genetic data sampled at one point in time, and on the age at sampling as an index of survival. In the absence of direct measures of selection – which would require phenotypic data relevant to fitness and estimates of heritability of those traits – there is little scope for specific conclusions regarding the mechanisms underlying the observed signals of selection.

5.5.2 Approach and assumptions

Inference of selection and model complexity

When inferring the strength of selection, there is a choice between general models that make strong assumptions and more specific models that account for various sources of uncertainty. The first strategy is usually associated with higher statistical power and analytical solutions, but biologically important details might be missed. This may lead to wrong conclusions, which motivates the second strategy (Zeng and Charlesworth 2009). A number of neutrality tests assume equilibrium of evolutionary forces (Schierup et al. 2000; Garrigan and Hedrick 2003; Charlesworth and Charlesworth 2010), and the common principle is to compare candidates of genes or sites under selection to putatively neutral ones (*e.g.* Lewontin and Krakauer 1975; Watterson 1978; McDonald and Kreitman 1991). Rejection of the null model (neutrality) is

the strongest result that may be obtained. This is fine if the goal is to obtain a set of neutral markers for inference on demography (Vitalis et al. 2001) – as was often the case in original applications. What if selection is of interest? Showing that a gene or site is under selection is one task. Estimating its strength and mode of dominance another, more difficult one. One must revert to models that explicitly parameterize the aspects of selection that are of interest. These parameters can then be estimated via maximum likelihood or in a Bayesian context. An important requirement for the validity of such inference, however, is that the demography is known and incorporated. One cannot infer details about the demography and test for the homogeneity of loci at the same time. This would lead to Felsenstein's (1982) "infinitely many parameters" problem (Vitalis et al. 2001). Recent studies do account for demography (*e.g.* Aguilar et al. 2004; Mona et al. 2008; Currat et al. 2010), and robust tests are being developed (Li 2011), but often this is intricate on its own (*e.g.* Barton and Etheridge 2004; Novembre et al. 2005; Keightley and Eyre-Walker 2007; Nielsen et al. 2007). The effect of alternative dominance schemes, however, seems to be ignored or only marginally addressed in most studies (*e.g.* Black and Hedrick 1997; Aguilar et al. 2004; Zeng and Charlesworth 2009, but see Lynd et al. (2010)). This may be justified when dominance does not affect the equilibrium state, which is the case for directional selection, but not for overdominance. Moreover, whether or not equilibrium has been reached is not obvious in general. We have tried to avoid some of these problems by i) explicitly conditioning on known demography, ii) allowing for a wide range of dominance schemes, and iii) not assuming equilibrium. We have used a modification of Beaumont and Nichols (1996) to show that selection is uniform in space. For parameter estimation, we then employed an explicit model of selection, migration and drift.

What information do these approaches use? The one similar to Beaumont and Nichols (1996) compares the F_{ST} of candidate loci to the distribution expected under the null model of neutrality. F_{ST} is essentially the observed between-deme variance, divided by the expected (maximum) value of this variance (Wright 1931). The variance (second moment) of the observed allele frequency distribution is affected by the spatial configuration of selection. For a one-locus model with two alleles, spatially uniform selection reduces the variance across demes, whereas heterogeneous selection maintains or increases it. The matrix iteration approach, on the other hand, uses the full information of the observed allele frequency distribution (Figures 5.13 and 5.14). Importantly, that distribution is not only sensitive to the strength of selection (s). Assuming spatially uniform selection, the variance in allele frequencies across derived demes also carries information about the initial allele frequency p_{init} (think of the binomial variance as a function of its parameter).

The advantages of our matrix iteration approach are the following. It provided a full-likelihood framework for joint parameter estimation. We did not assume equilibrium of evolutionary processes, and could incorporate demography via effective deme sizes and an explicit model of migration. It accounted for the uncertainty about the nuisance parameter p_{init}. Last, it is straightforward to include prior information if such is available. The drawback of the approach is its high computational cost due to the construction of large stochastic matrices. The size of those depends on N_e. The problem may be overcome using a scaling argument based on the diffusion approximation (Hill and Robertson 1966).

Assumptions revisited

Throughout, we assumed viability selection and ignored potential effects on fertility. Also, we assumed that mating is random with respect to the MHC. If the MHC were involved in mate choice or other aspects relevant to reproduction (*e.g.* Thoß et al. 2011), the matrix iteration approach could still be used. However, the model would have to be adjusted to account for those aspects. This would require additional parameters and might limit the statistical power. Selection acting at the stage of reproduction or early development would not have been detected by our short-term analysis. Further, we have used a one-locus model. The signatures of selection detected at the DRB locus might be a joint effect of selection acting at DRB, at other MHC genes and elsewhere in the genome (van Oosterhout 2009; Thoß et al. 2011).

For the short-term analyses, we used the age at sampling (equivalent to age at culling in most cases) as an indicator of survival. This way, we only caught one aspect of viability (Hedrick et al. 1976). We were also exposed to the 'invisible fraction' problem (Hadfield 2008), because individuals that died before sampling were not observed and their genotypes are thus unknown. A further note of caution seems appropriate given a meta-analysis by Chapman et al. (2009) of studies on heterozygosity-fitness correlations. The authors found that generally the observed patterns are in disagreement with population genetic predictions, and the effects only explain a small ($< 1\%$) proportion of the variance in phenotypic characters. We further assumed that culling and genotype were uncorrelated. We do not know of any evidence for the opposite in Alpine ibex, although there might be visual or behavioral differences between carriers of the 'goat' and the ibex allele at OLADRB2, with consequences on the probability of being culled (see Giacometti et al. 2004, for body weight, horn growth and morphology of F1 and F2 hybrids). For an example of such biases in bighorn sheep rams, see Coltman et al. (2003). Compared to this, however, the Swiss hunting scheme for Alpine ibex is very restrictive and less prone to such effects. Last, we assumed that, conditional on the genotype, individuals are exchangeable across time with respect to viability selection. This would be hampered, if selection pressure changes on a shorter time scale than that of generations.

In the model used for the medium-term analyses, we have ignored mutation for the $\sim$90 years ($\sim$10 generations) between reintroduction and sampling. This is justified for two reasons. First, even in the most extreme case of a star-shaped genealogy, the per-site mutation rate would have to be extremely high ($\geq 10^{-6}$) for at least one mutation to be expected. Second, the two DRB haplotypes observed in the derived demes are identical to those still present in the original deme in Italy (Figure 5.2), and allele 277 of OLADRB2 is fully diagnostic for the 'goat' haplotype at exon 2 of DRB. Second, we assumed that the allele frequency in the ancestral deme (p_0) was constant during the 16 years (1.7 generations) between the establishment of the first and last derived deme. This may have introduced a bias, but it simplified the likelihood function considerably. Third, we approximated migration by a continent-island model, which implies a common gene pool of infinitely many immigrants and a global immigration rate. Although the demes are geographically close, this is certainly an oversimplification. A detailed study on pairwise migration rates is underway. Fourth, we have assumed that selection coefficient, dominance coefficient and migration rate are constant over time. In the absence of time-series data, this is the most parsimonious assumption.

Uncertainty about N_e

In the matrix iteration approach we relied on demographic estimates of effective deme sizes N_e. We obtained these estimates from various parameters that were more or less accurately estimated in previous studies (see Appendix and SI), but are in general difficult to estimate. Our estimates of N_e are therefore error-prone. To assess the potential effect on the marginal estimate of the selection coefficient s, we repeated the inference with values of N_e increased or decreased by 20% of the actual estimate. As shown in Table 5.8, the effect on the MLE of s was moderate, with a maximum relative bias of 13% in case of $\phi = 0.50$. The width of the 95% credible interval was only marginally affected. For intermediate dominance, the MLE of s was negatively correlated with N_e (Table 5.8B). This is reminiscent of the fact that in the diffusion approximation, only the compound parameter $N_e s$ is relevant. However, the relation we observed between s and N_e was not perfectly proportional. The reason is that time is constrained by the demographic scenario (Figure 5.2). This hampers the scaling argument in our case.

5.5.3 Conclusion and outlook

Using a combination of approaches, we found signatures of spatially uniform selection at exon 2 of DRB (MHC class II) in a structured population of Alpine ibex in the Swiss Alps. Scenarios with either asymmetric overdominance or directional selection against the haplotype shared with domestic goat were most likely. Other, less likely dominance schemes were most compatible with a drift-only scenario, however. Assuming a constant selection pressure over the last 10 generations, and that short-term signatures of selection can be used to condition the analysis over the microevolutionary time scale, it seems that the 'goat' haplotype is selected against in Alpine ibex. Extrapolating further back in time, this would imply that the trans-species polymorphism is more likely a consequence of relatively recent introgression, rather than a shared ancestral polymorphism (SAP). Otherwise, we would expect it to have been lost (in case of directional selection, at least). However, the question of SAP versus recent introgression should be addressed independently in the future, using sequence data of the flanking region of the MHC. Differences at synonymous sites between goat and ibex haplotype would support a shared ancestral polymorphism (or ancient introgression), rather than recent introgression. Similarly, the pattern of LD along the chromosome may be informative.

Although current literature reflects a consensus that selection has been acting on MHC in many taxa, the pattern and conclusions about mechansisms are surprisingly heterogeneous. No general rule has emerged from studies on natural or human populations. Our work adds to this complexity, in particular providing support for selection under two distinct, contradicting modes of dominance. There is clearly a need for further empirical studies. In particular, insight may be obtained by i) considering different time scales on which selection may act and over which its signatures persist; ii) taking into account the effects of demography and alternative modes of dominance; iii) using time series data and/or samples from multiple populations; and iv) showing causal relationships between MHC variation and fitness. Those are challenging requirements. Recent studies satisfy some, but more effort seems needed to complete our understanding of MHC evolution.

5.6 Appendix

5.6.1 Transition probabilities

For derived demes ($\alpha \in \{1, 2, \ldots, \Gamma\}$), recalling equations (5.8), (5.5) and (5.4), the transition probabilities are given by

$$q_{ij}^{(\alpha)} = \binom{2N_{t+1}^{(\alpha)}}{j} \left(p_\alpha^{**}\right)^j \left(1 - p_\alpha^{**}\right)^{2N_{t+1}^{(\alpha)} - j}, \tag{A1}$$

where

$$\begin{aligned} p_\alpha^{**} &\overset{(5.8)}{=} p_\alpha^* + m(p_I^* - p_\alpha^*), \\ p_I^* &\overset{(5.5)}{=} p_I(t)\frac{w_{1,I}}{\bar{w}_I}, \\ p_\alpha^* &\overset{(5.4)}{=} p_\alpha(t)\frac{w_{1,\alpha}}{\bar{w}_\alpha}, \end{aligned} \tag{A2}$$

and

$$p_\alpha(t) = \frac{i}{2N_t^{(\alpha)}}. \tag{A3}$$

For the ancestral deme, there is no migration, and we obtain

$$q_{ij}^{(0)} = \binom{2N_{t+1}^{(0)}}{j} \left(p_0^*\right)^j \left(1 - p_0^*\right)^{2N_{t+1}^{(0)} - j}, \tag{A4}$$

where

$$p_0^* \overset{(5.4)}{=} p_0(t)\frac{w_{1,0}}{\bar{w}_0}, \tag{A5}$$

and

$$p_0(t) = \frac{i}{2N_t^{(0)}}. \tag{A6}$$

Recall that in equations (A1) and (A4) the marginal and mean fitnesses are functions of $p(t)$, which we have omitted for simplicity.

The transition probabilities for the founder events are obtained according to the binomial distribution as

$$f_{kl}^{(\alpha)} = \binom{2N_{t_\alpha}^{(\alpha)}}{l} p_0(t_f)^l \big(1 - p_0(t_f)\big)^{2N_{t_\alpha}^{(\alpha)} - l}, \quad \text{where} \quad p_0(t_f) = \frac{k}{2N_{t_f}^{(0)}}. \tag{A7}$$

5.6.2 Derivation of likelihood function

Here, we describe the derivation of the likelihood of the parameters s and p_{init} given the data $(\hat{\mathbf{p}}, \mathcal{N})$ and some fixed values of m and ϕ (or h): $L(s, p_{\text{init}}; \hat{\mathbf{p}}, \mathcal{N}, m, \phi)$. For simplicity, we omit m and ϕ in the notation from now on. We define X as the number of copies of allele

A_1 in the ancestral deme at time t_f; Y_α as the number of copies in deme d_α at the time of founding, t_α; and Z_α as the number of copies in deme α at the time of sampling, t_s. We do not have information about the Y_α, so that we have to sum over all possible outcomes of this intermediate state. We know Z_α directly from the observation $\hat{\mathbf{p}}$. X is determined via p_{init}, the initial allele frequency in the ancestral deme d_0. The likelihood of interest is then equal to the probability of the observed allele frequencies $\hat{\mathbf{p}}$, given the parameters s and p_{init} and the deme size trajectories $\mathcal{N}$:

$$L\left(s, p_{\text{init}}; \hat{\mathbf{p}}, \mathcal{N}\right) = P\left[\mathbf{Z} = (Z_1, Z_2, \ldots, Z_\Gamma) \mid s, p_{\text{init}}, \mathcal{N}\right] \tag{A8}$$

$$= \sum_{x \in \mathcal{X}} P\left[X = x \mid s, p_{\text{init}}, \mathcal{N}^{(0)}\right] \prod_{\alpha=1}^{\Gamma} P\left[Z_\alpha = z_\alpha \mid X = x, s, \mathcal{N}^{(\alpha)}\right], \tag{A9}$$

where $\mathcal{X}$ is the set of possible values that X can take. Recall that $\mathcal{N}^{(0)}$ is the deme size trajectory of the ancestral deme, whereas $\mathcal{N}^{(\alpha)}$ is the one for the derived deme d_α, with $\alpha \in \{1, 2, 3, \ldots, \Gamma\}$. The probabilities $P\left[Z_\alpha = z_\alpha \mid X = x, s, \mathcal{N}^{(\alpha)}\right]$ are given by

$$P\left[Z_\alpha = z_\alpha \mid X = x\right] = \sum_{y_\alpha \in \mathcal{Y}_\alpha} P\left[Y_\alpha = y_\alpha \mid X = x\right] \cdot P\left[Z_\alpha = z_\alpha \mid Y_\alpha = y_\alpha\right], \tag{A10}$$

where $\mathcal{Y}_\alpha$ is the set of values that Y_α can take, and we have dropped the conditioning on s, p_{init} and $\mathcal{N}$ for simplicity. Equation (A10) makes explicit the summation over all unobserved outcomes of the variable Y_α mentioned in the main text, and introduces the founder event explicitly via $P\left[Y_\alpha = y_\alpha \mid X = x\right]$.

It is straightforward to relate the probabilities $P\left[X \mid s, p_{\text{init}}, \mathcal{N}^{(0)}\right]$, $P\left[Y_\alpha \mid X, \mathcal{N}^{(0)}\right]$, and $P\left[Z_\alpha \mid Y_\alpha, s, \mathcal{N}^{(\alpha)}\right]$ to the transition matrices introduced in the main text (Methods). First, $P\left[X \mid s, p_{\text{init}}, \mathcal{N}^{(0)}\right]$ is obtained from the transition matrix $\mathbf{Q}^{(0)}_{t_0 \to t_f}$. It corresponds to the probability distribution given by the row of $\mathbf{Q}^{(0)}_{t_0 \to t_f}$ that reflects the transition from an initial number of $i = \left\lfloor p_{\text{init}} \cdot 2N^{(0)}_{t_0} \right\rceil$ copies of A_1 to any possible number x of copies at time t_f, where we use $\lfloor r \rceil$ to denote the nearest integer from r. Second, $P\left[Y_\alpha \mid X, \mathcal{N}^{(0)}\right]$ is the probability distribution given by the vector that contains the transition probabilities of going from $X = x$ copies of A_1 in the ancestral deme at time t_f to any possible number y_α of copies in deme d_α at time t_α. So, the desired probability distribution is given by the row vector $\mathbf{F}^{(\alpha)}$. Third, $P\left[Z_\alpha \mid Y_\alpha, s, \mathcal{N}^{(\alpha)}\right]$ is obtained from the transition matrix $\mathbf{Q}^{(\alpha)}_{t_\alpha \to t_s}$. It is the probability distribution given by the row of $\mathbf{Q}^{(\alpha)}_{t_\alpha \to t_s}$ that corresponds to going from a number of $Y_\alpha = y_\alpha$ copies of A_1 in deme d_α at time t_α to any possible number z_α of copies in deme d_α at the time of genetic sampling, t_s.

Recall from the description of the model (Model and parameters) that p_{init} is a nuisance parameter and expected to be correlated to s. Since we do not have data to estimate p_{init}, we are also interested in the likelihood of s marginal to p_{init}. This is obtained by summing the joint likelihood over the range of values that p_{init} can take:

$$L\left(s; \hat{\mathbf{p}}, \mathcal{N}\right) = \sum_{p_{\text{init}} \in \mathcal{P}} L\left(s, p_{\text{init}}; \hat{\mathbf{p}}, \mathcal{N}\right), \tag{A11}$$

where $\mathcal{P}$ is the set of values in (0.0, 0.6] that p_{init} can take (*cf.* Figure 5.2). We provide an illustration of the approach in the SI. MATHEMATICA notebooks implementing the matrix approach are available from the corresponding author or may be downloaded from the website `http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`.

5.7 Supporting information: Additional tables

Table 5.5: Founder events, demography and trajectories of effective deme sizes.

Deme	Name	τ_{first}	τ_{last}	F_{tot}	F_m	F_f	τ_{50}	F_e	g_α	$\mathcal{N}^{(\alpha)}$
d_0	Albris[a]	1920	1934	42	16	26	1927	25	9	{25, 137, 241, 340, 478, 463, 479, 498, 410}
d_1	Adula-Vial	1965	1965	19	13	6	1965	15	5	{15, 45, 98, 176, 162}
d_2	Calanda	1968	1970	36	15	21	1970	26	4	{26, 29, 41, 56}
d_3	Crap da Flem	1958	1963	27	16	11	1958	7	5	{7, 32, 36, 42, 52}
d_4	Flüela	1958	1987	42	30	12	1959	40	5	{40, 238, 342, 357, 326}
d_5	Hochwang	1965	1973	40	21	19	1971	32	4	{32, 67, 74, 49}
d_6	Julier Nord	1954	1970	109	76	33	1965	74	5	{74, 225, 276, 280, 248}
d_7	Julier Süd	1954	1970	41	30	11	1957	16	6	{16, 44, 160, 271, 177, 155}
d_8	Macun	1969	1980	53	36	17	1974	22	4	{22, 45, 52, 56}
d_9	Safien-Rheinw.	1954	1965	29	17	12	1954	17	6	{17, 49, 128, 156, 245, 198}
d_{10}	Rothorn-Weissfl.	1959	1971	77	47	30	1962	52	5	{52, 155, 173, 121, 114}
d_{11}	Umbrail	1970	1979	59	38	21	1976	17	3	{17, 49, 35}
d_{12}	Val Bever	1957	1971	137	91	46	1961	56	5	{56, 99, 146, 116, 95}
d_{13}	Oberalp-Frisal	1955	1970	65	42	23	1966	32	5	{32, 67, 87, 154, 157}

τ_{first}, τ_{last}, year in which first/last founders were released, respectively; F_{tot}, F_m, F_f, number of individuals, males and females, respectively, released as founders into the deme (all originating from d_0); τ_{50}, year by which at least 50% of founders had been released (corresponds to t_α in Figure 5.2); F_e, effective number of founders used in the matrix iteration approach (equal to $\mathcal{N}^{(\alpha)}_{t_\alpha}$ in main text); g_a, age of deme in generations; $\mathcal{N}^{(\alpha)}$, trajectory of effective deme sizes for deme α, estimated from demographic data (see text for details).
[a] The Albris deme is ancestral to all other derived demes. It was established with 25 individuals (59.5%; 11 males, 14 females) from the St. Gall zoo and with 17 (40.5%; 5, 12) from the Interlaken zoo (*cf.* Figure 5.2, and Figure 5.12).

Table 5.6: Genotypes, age and sex of sampled Alpine ibex (large table provided as a spreadsheet on `http://pub.ist.ac.at/~saeschbacher/phd_e-sources/`).

Table 5.7: Likelihood-based estimates of selection (s) and dominance (ϕ) coefficient conditioning on underdominance, without migration.

Dominance scheme	ϕ	$L_\phi{}^a$	B.F.	$\hat{s}_\phi$	HPD
Full recessivity of the 'goat' allele A_1	0.000	0.594	0.386	0.000	(−0.781, 0.000)
Underdominance ($-1 \leq s \leq 0$)	0.125	0.784	0.509	0.000	(−0.862, 0.000)
·	0.250	1.009	0.656	0.000	(−0.892, 0.000)
·	0.375	1.232	0.800	−0.018	(−0.903, 0.000)
·	0.500	1.410	0.916	−0.165	(−0.902, 0.000)
·	0.625	1.517	0.985	−0.205	(−0.892, 0.000)
·	0.750	1.539	1.000	−0.209	(−0.869, 0.000)
·	0.875	1.415	0.919	−0.200	(−0.748, 0.000)
Full dominance of the 'goat' allele A_1	1.000	1.229	0.799	−0.190	(−0.574, 0.000)

Details are as in Table 5.3, with the difference that, here, we conditioned the inference on underdominance, *i.e.* we required $-1 \leq s \leq 0$. Point and interval estimates correspond to likelihood curves displayed in Figure 5.10. The marginal cases of $\phi = 0.00$ and $\phi = 1.00$ are included for comparison.
[a] In multiples of 10^{-13}.

Table 5.8: The effect on the estimation of the selection coefficient (s) of uncertainty about effective deme size (N_e).

A) Over- and underdominance

	$N_e{}^{(0)}$		$0.8 \times N_e{}^{(0)}$				$1.2 \times N_e{}^{(0)}$			
ϕ	$\hat{s}_\phi{}^{(0)}$	HPD$^{(0)}$	$\hat{s}_\phi$	HPD	s_{rel}	l_{rel}	$\hat{s}_\phi$	HPD	s_{rel}	l_{rel}
0.00	0.595	(−0.306, 0.869)	0.595	(−0.304, 0.873)	1.00	1.00	0.592	(−0.396, 0.861)	0.99	1.07
0.50	−0.165	(−0.995, 0.344)	−0.144	(−0.972, 0.364)	0.87	1.00	−0.184	(−1.000, 0.336)	1.12	1.00
1.00	−0.190	(−0.613, 0.142)	−0.200	(−0.625, 0.133)	1.05	1.00	−0.181	(−0.603, 0.148)	0.95	0.99

B) Intermediate dominance

	$N_e{}^{(0)}$		$0.8 \times N_e{}^{(0)}$				$1.2 \times N_e{}^{(0)}$			
ϕ	$\hat{s}_\phi{}^{(0)}$	HPD$^{(0)}$	$\hat{s}_\phi$	HPD	s_{rel}	l_{rel}	$\hat{s}_\phi$	HPD	s_{rel}	l_{rel}
0.25	0.435	(0.025, 0.574)	0.444	(0.035, 0.588)	1.02	1.01	0.429	(0.018, 0.563)	0.99	0.99
0.50	0.326	(0.008, 0.463)	0.342	(0.014, 0.476)	1.05	1.02	0.309	(0.003, 0.453)	0.95	0.93
0.75	0.223	(0.000, 0.394)	0.240	(0.000, 0.400)	1.08	1.02	0.210	(0.000, 0.389)	0.94	0.99

The marginal MLE of s given ϕ ($\hat{s}_\phi$, in A) or h ($\hat{s}_h$, in B), and the highest posterior density interval (HPD) are shown for two scenarios in which the original estimates of effective deme size, $N_e{}^{(0)}$, was decreased or increased by 20%, respectively. The ratio of the respective point estimate, $\hat{s}$, relative to $\hat{s}^{(0)}$, is given by s_{rel}, and the corresponding ratio for the length of the HPD is given by l_{rel}.

5.8 Supporting information: Additional figures

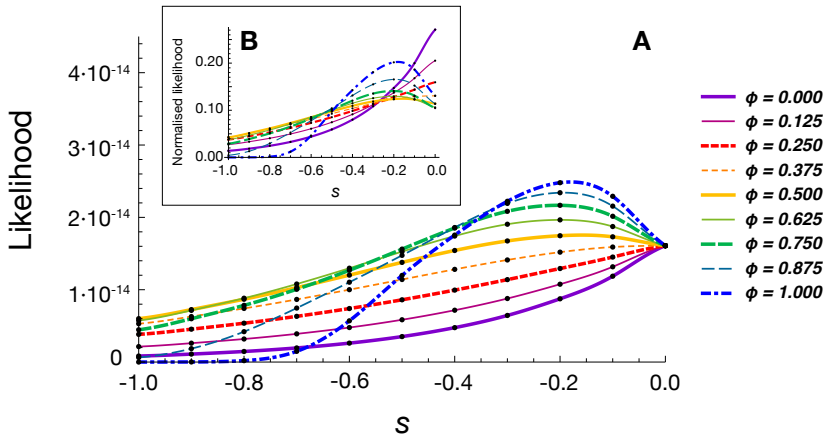

Figure 5.10: Likelihood of the selection coefficient s for various degrees of underdominance without migration. Details are as in Figure 5.6 in the main text, with the difference that, here, we conditioned the inference on underdominance, *i.e.* we required $-1 \leq s \leq 0$. (A) The likelihoods are not normalized. Therefore, the areas under the curves indicate the relative support for the respective values of ϕ (*cf.* Table 5.7). (B) As in (A) but with likelihoods normalized such that the area under the curve is 1. Further details as in Figure 5.6 in the main text. For comparison, the marginal cases of full recessivity and full dominance of A_1, $\phi = 0.00$ and $\phi = 1.00$, are included.

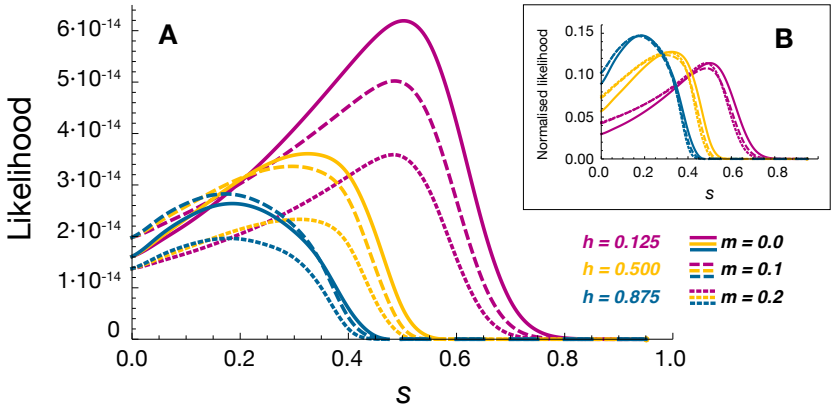

Figure 5.11: The effect of gene flow via migration (at rate m) on the marginal likelihood of the selection coefficient s intermediate dominance. (A) The likelihoods are not normalized and the areas under the curves indicate the relative support for the different migration rates m, given h. (B) Likelihoods normalized such that the area under the curve is 1. In a Bayesian view, these curves correspond to the posterior distribution of s given a uniform prior on the normal scale. Other details as in Figure 5.8 in the main text. For under- and overdominance, see Figure 5.25.

5.9 Supporting information: Additional data and methods

5.9.1 Demography and effective deme size

Figure 5.12 gives the trajectories of effective deme sizes for each deme.

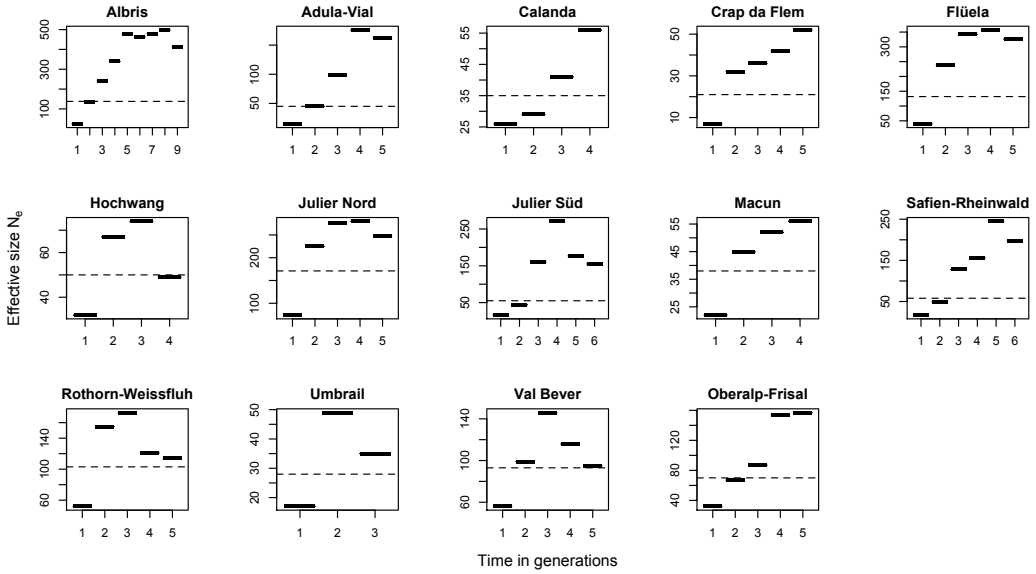

Figure 5.12: Trajectories of effective deme sizes $\mathcal{N}^{(\alpha)}$ over time in generations. Black bars show effective deme size N_e computed according to Nunney (1991, 1993) and described in the Appendix of the main text. Dashed lines show the harmonic mean over time of N_e, which was used to illustrate the principle of the matrix iteration approach in Figures 5.13 and 5.14. Generations are discrete, and are aligned on the real time scale such that their last generation coincides with the time of sampling t_s (*cf.* main text, Figure 5.2 and Table 5.5).

5.9.2 Genotypic raw data

Table 5.6 (in electronic format only) provides genotypes, age at sampling and sex of sampled Alpine ibex. Table 5.9 gives the abbreviations of deme names used, along with sample sizes and observed frequencies of the 'goat' allele A_1.

5.9.3 Heterozygosity versus age at sampling

To quantify the relationship between heterozygosity, age at sampling, deme and sex, we fitted Generalized Linear Models (GLM) with various combinations of predictors, assuming a binomial distribution of the error terms (logistic regression). We performed model selection based on the Akaike Information Criterion (AIC; Akaike 1974; Burnham and Anderson 2002). We regressed the response variable heterozygosity against combinations of the explanatory variables age, deme and sex. In the stepwise fitting process, we allowed for all first-order interaction terms to be explored. Recall that we are mainly interested in the effect of age, but would like to account for the potentially confounding effects of the two covariates deme and sex. We kept as the best models the one with minimum AIC (AIC_{min}) and those with substantial support,

Table 5.9: Deme name abbreviations, sample sizes and estimated frequencies $\hat{p}$ of the 'goat' allele at OLADRB2.

Deme Name	Abbreviation	data_mhc_14	Males	Females	Total	$\hat{p}$
Albris	Albris	TRUE	17	21	38	0.1447
Adula-Vial	AdulaVial	TRUE	20	16	36	0.0972
Calanda	Calanda	TRUE	9	10	19	0.2368
Cape au Moine	CapeMoine	FALSE	10	11	21	0.3095
Crap da Flem	CrapFlem	TRUE	3	3	6	0.0833
Fergen-Seetal	FergenSeetal	FALSE	3	1	4	0.1250
Flüela	Fluuela	TRUE	23	11	34	0.0735
Gran Paradiso (Rhemes)	GPRhemes	FALSE	4	0	4	0.1667
Graue Hörner	GrHörner	FALSE	15	14	29	0.1034
Hochwang	Hochwang	TRUE	12	13	25	0.2200
Julier Nord	Julier N	TRUE	2	2	4	0.2500
Julier Süd	Julier S	TRUE	8	6	14	0.1071
Macun	Macun	TRUE	12	10	22	0.2727
Oberalp-Frisal	Oberalp	TRUE	5	5	10	0.3000
Pierreuse-Gummfluh	Pierreuse	FALSE	9	9	18	0.0833
Safien-Rheinwald	Rheinwald	TRUE	17	13	30	0.2500
Rothorn-Weissfluh	RothWeiss	TRUE	14	11	25	0.2400
Umbrail	Umbrail	TRUE	13	11	24	0.1667
Val Bever	ValBever	TRUE	14	6	20	0.1000
Vals	Vals	FALSE	1	1	2	0.2500
Weisshorn	Weisshorn	FALSE	6	5	11	0.0455
Wittenberg	Wittenberg	FALSE	14	6	20	0.2000
Wildpark Goldau	WPGA	FALSE	2	4	6	0.2500

All demes are shown in which both the goat (A_1) and ibex (A_2) allele were found and which therefore make up data set data_mhc_23 (see main text for details). The field 'data_mhc_14' indicates if samples from a deme are present in data_mhc_14, the data set used for the matrix iteration approach.

i.e. with $\Delta_i < 2$, where $\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$ is the difference between the AIC of model i and $\mathrm{AIC}_{\min}$. We treated all explanatory variables as fixed effects and obtained p-values for the effects from a Wald test (Agresti 1990), as available in the `glm` function of the `stats` package in `R` (R Development Core Team 2011). We judged significance based on the threshold of 0.05 and computed 95% confidence intervals of the estimates using the `confint` function of the `stats` package in `R` (R Development Core Team 2011). For the explanatory variable deme, which is a multilevel factor, we assessed its overall significance by a Wald test as provided by the `aod` package in `R` (Lesnoff and Lancelot 2009).

The AIC for model selection aims at a trade-off between explanation of the data and prediction. A model with many parameters (effects) will potentially fit the data better, but be less general. AIC punishes for this by adding twice the number of parameters to the negative log-likelihood. Hence, for two models with equal likelihoods, the one with fewer parameters is preferred. Alternatively, logistic regression models may be compared according to their discriminating power, *i.e.* by how well the model predicts the true response given some explanatory value. Discriminating power can be expressed as the ratio of the true positive rate (TPR, sensitivity) to the false positive rate (FPR, 1 − specificity), TPR/FPR, as the FPR is changed. The curve obtained by plotting the empirical TPR versus FPR is called Receiver Operating Characteristic (ROC) curve. The ROC hence illustrates the relative trade-off between the benefits (true positives) and costs (false positives; Fawcett 2006). The ROC of a model that makes random decisions (a random classifier in the machine learning chargon) would correspond to

a straight line with slope 1. The more concave the ROC is, the better is the model at discriminating. The point (0, 1) represents perfect classification, and the area under the ROC (called AUC) is a measure of overall discriminating performance of the model. However, a very concave ROC may arise merely due to overfitting, and there is no obvious threshold for when to accept or reject a model (but see Forman 2002). We use the AUC as an alternative criterion for model selection besides AIC.

A potential heterozygosity-age relationship for OLADRB2 does not need to be specific to exon 2 of the MHC class II complex (or the surrounding area), but might reflect a genome-wide heterozygosity effect. In the latter case, we would expect to see a significant heterozygosity-age relationship for many other (neutral) loci. If the effect is specific to the MHC class II genes, however, we expect to see no systematic negative relationship between age and heterozygosity at other markers than those in tight linkage to exon 2 of the MHC class II complex (OLADRB1, OLADRB2, OMHC1). When regressing heterozygosity at other markers than OLADRB2 against the covariates, we started off with the same samples as used for OLADRB2 (data sets data_mhc_14 and data_mhc_23), but excluded those with missing values for the genotype of the respective marker. We performed stepwise fitting of a Generalized Linear Model (GLM) with binomial error distribution (logistic regression) for each locus and, within each locus, for each allele. Model selection was done via AIC as described above for OLADRB2. For a given focal allele, we determined heterozygosity after assigning all other alleles to a separate class.

5.9.4 Genotype versus age at sampling

Pairwise logistic regression

OLADRB2 is a biallelic locus with alleles A_1 (the 'goat' allele) and A_2. Hence, there are three potential genotypes, A_1A_1, A_1A_2 and A_2A_2, and three pairwise comparisons among them. Let the binary response variable a1a1.a1a2 contrast the genotype A_1A_1 with A_1A_2, where a1a1.a1a2 = 0 for an A_1A_1 individual and a1a1.a1a2 = 1 for an A_1A_2 individual. Analogously, let a1a1.a2a2 and a1a2.a2a2 be the contrasting factors for the remaining two comparisons. For each pairwise comparison, we fitted a GLM with binomial error distribution (logistic regression) with age at sampling, deme and sex as potential predictors. We performed stepwise model selection based on the Akaike Information Criterion (AIC; Akaike 1974; Burnham and Anderson 2002) and we allowed for first-order interaction terms to be included. We kept as the best models the one with minimum AIC ($\mathrm{AIC}_{\min}$) and those with substantial support, *i.e.* with $\Delta_i < 2$, where $\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$ is the difference between the AIC of model i and $\mathrm{AIC}_{\min}$. We treated all explanatory variables as fixed effects and obtained p-values for the effects from a Wald test, as available in the `glm` function of the `stats` package in `R` (R Development Core Team 2011). We judged significance based on the threshold of 0.05 and computed 95% confidence intervals of the estimates using the `confint` function of the `stats` package in `R` (R Development Core Team 2011). For the explanatory factors with multiple levels (deme, interactions), we assessed the overall significance with a Wald test as provided by the `aod` package in `R` (Lesnoff and Lancelot 2009).

Multinomial logistic regression

Alternatively, we defined ola2.gtp as the three-level response in a multinomial regression. It takes the values a1a1, a1a2 and a2a2 for the three potential OLADRB2 genotypes A_1A_1, A_1A_2 and A_2A_2, respectively. We fitted multinomial logistic regressions to explain ola2.gtp in terms of the predictors age at sampling, deme, sex and their interactions. We treated all predictors as individual-specific variables and used the `mlogit` package (Croissant 2008) for `R` (R Development Core Team 2011).

5.9.5 Estimating effective deme size from demographic data

Here, we give formal details of how we estimated effective deme sizes from demographic data. Specifically, we account for overlapping generations, the variance in reproductive success, differences between the two sexes and population growth. Nunney (1991, 1993) approximated earlier results on this by Hill (1972, 1979) and formulated them in terms of quantities that may be estimated more easily from wild populations. Hill and Nunney derived their results assuming that deme sizes are constant over time and that the age structure is stable. We will present a heuristic way of adjusting this to changing deme sizes. We start with equation (A3) in the Appendix of Nunney (1993) which gives the effective size as

$$N_e = \frac{N_A(T/A)}{1 + \left\{I_{b_m}/r + I_{b_f}/(1-r) - 2/\left[(1-r)b_f\right]\right\}/(4A) + (I_{A_m} + I_{A_f})/2}, \tag{A12}$$

where N_A is the number of adults and T is the mean generation time of males and females, defined as the average age of a parent. A is the average adult life span, defined as $(A_m + A_f)/2$, where A_m and A_f are the average life span of male and female adults, respectively. A_m and A_f are given by the average age of death of males and females, respectively, minus $(M-1)$, where M is the age at which juveniles of either sex start reproducing. The N_A adults have a sex ratio (proportion of males) of r. The standardized variances (variance/mean^2) in life span and in seasonal fecundity are defined as I_{A_m} and I_{b_m} for males, and as I_{A_f} and I_{b_f} for females, respectively. Seasonal mean fecundity of a female (the average number of offspring reared to independence) is b_f (*cf.* Nunney 1993).

To estimate the standardized variance in female seasonal fecundity, I_{b_f}, we proceeded as follows. The proportion of adult females that reproduce in a given mating season, ρ_f, and the expected number of offspring reared to independence by a female, b_f, are related as

$$b_f = \rho_f(1+z), \tag{A13}$$

where z is the proportion of twin births. One can therefore estimate ρ_f by $\hat{b}_f/(1+\hat{z})$. If K is the number of offspring per female per season, b_f is the expectation of K, *i.e.* $b_f = E(K) = \rho_f(1+z)$. The variance of K is given by $V(K) = E(K^2) - E(K)^2 = \rho_f(1-3z) - \rho_f^2(1+z)^2$. Hence, the standardized variance in seasonal female fecundity is

$$I_{b_f} = \frac{V(K)}{E(K)^2} = \frac{V(K)}{b_f} = \frac{\rho_f(1-3z) - \rho_f^2(1+z)^2}{\left(\rho_f(1+z)\right)^2} = \frac{1+3z-\rho_f(1+z)^2}{\rho_f(1+z)^2}. \tag{A14}$$

The standardized variance in male seasonal fecundity depends on the mating system. For dominance hierarchy, the system that applies to Alpine ibex, it is given by equation (19) in

Nunney (1993) as

$$I_{b_m} = \frac{r}{\rho_f(1-r)} + \frac{1-\rho_m}{\rho_m}, \tag{A15}$$

where ρ_m is the proportion of dominant males (*i.e.* the proportion of males getting access to matings per season). In the Online SI we show how we estimated the ingredients to formulae (A12) to (A15).

We applied formula (A12) to obtain, for each deme, a series of local in time estimates of the effective size. We then substituted these series for the deme size trajectories $\mathcal{N}^{(\alpha)}$ and $\mathcal{N}^{(0)}$ as introduced in (5.1). By 'local in time' we mean that we obtained one estimate of N_e per time segment. We chose the length of such a segment to be equal to the average generation time $T = 9$ years (generation time for Alpine ibex). We started dividing time into segments (*i.e.* generations) at the time of sampling, τ_s (in units of one year), and then went backwards in time year by year, closing a segment (generation) every $T = 9$ years. The time in years of existence of a deme is not necessarily a multiple of the generation time. In such cases, if less than five years were remaining after the last complete generation, we assigned them to the last complete generation. If five or more years remained, we lumped them into a new generation. Therefore, we obtained the number of generations over which a derived deme d_α existed as $g_\alpha := \lfloor (\tau_s - \tau_\alpha)/T \rceil$, where τ_α is the year in which the first founder individual was released to deme d_α, and $\lfloor x \rceil$ means rounded to the next integer. Analogously, we set $g_0 := \lfloor (\tau_s - \tau_0)/T \rceil$ for the ancestral deme d_0. We further obtained the time of sampling in units of generations as

$$t_s = \left\lfloor \frac{\tau_s - \tau_0}{T} \right\rceil = g_0, \tag{A16}$$

and the time of foundation in units of generations as

$$t_\alpha = t_s - g_\alpha \quad \text{and} \quad t_0 = t_s - g_0 \tag{A17}$$

for derived demes and the ancestral deme, respectively. Combining (A16) and (A17) then asserts that $t_0 = t_s - g_0 = 0$. From the recorded census size estimates $N_{c,\tau}^{(\alpha)}$, we obtained the corresponding numbers of adults $N_{A,\tau}^{(\alpha)} = \hat{a} N_{c,\tau}^{(\alpha)}$ in derived demes, and analogously for the ancestral deme, replacing α by 0. We then set the per generation number of adults, $N_{A,t}^{(\alpha)}$, equal to the harmonic mean of the corresponding annual numbers of adults:

$$N_{A,t}^{(\alpha)} = T \left(\sum_{\tau \in \mathcal{T}} \frac{1}{N_{A,\tau}^{(\alpha)}} \right)^{-1}, \tag{A18}$$

where $\mathcal{T}$ is the set of all years τ that are assigned to generation t, *i.e.* $\mathcal{T} = \{\tau : t \leq \lfloor \tau/T \rfloor < t+1\}$. Substituting (A18) for N_A in (A12), we obtained the local in time estimate $N_{e,t}^{(\alpha)}$ for a derived deme d_α in generation t. For the ancestral deme, replace α by 0 in (A18).

5.9.6 Parameter values used for the estimation of effective deme sizes

In the following, we describe how we estimated the parameters that were needed to compute effective deme sizes from demographic data. The formal aspects of this are given in the previous paragraph. A list of symbols used is provided in Table 5.10.

Table 5.10: List of symbols used in the estimation of effective deme sizes.

Symbol	Description
N_c	Total census size of a constant population
N_A	Number of adults ($\geq$ 3 years) in a constant population
a	N_A/N_c
T	Mean generation time (average age of a parent)
A_m, A_f, A	Average life span of male and female adults
A	Average adult live span
M	Age at which juveniles start reproducing
r	Adult sex ratio (proportion of males)
I_{A_m}, I_{A_f}	Standardized variance in male and female life span
I_{b_m}, I_{b_f}	Standardized variance in male and female seasonal fecundity
b_f	Average number of offspring reared to independence by a female
ρ_f	Proportion of adult females that reproduce in a given season
ρ_m	Proportion of dominant males (access to matings) in a season
z	Proportion of twin births
τ	Time in years (*cf.* t in Table 5.1)
τ_s, τ_α	Year of sampling, year of first founder event in deme d_α
g_α	Number of generations for which an ancestral deme d_α existed
g_0	Number of generations for which the ancestral deme d_0 existed
$N_{c,\tau}^{(\alpha)}$	Census size estimate for deme d_α in year τ
$N_{A,\tau}^{(\alpha)}$	Estimated number of adults for deme d_α in year τ
$N_{A,t}^{(\alpha)}$	Estimated number of adults for deme d_α in generation t
$N_{e,t}^{(\alpha)}$	Local in time estimate of effective size of deme d_α in generation t

We assumed that Alpine ibex start reproducing at an age of $M = 3$ years (Nievergelt 1966; Stuwe and Grodinsky 1987; Toïgo et al. 2002). Nievergelt (1966) argues that it may be higher ($M \geq 4$). The generation time T has been estimated previously to about 9 years (Stuwe and Grodinsky 1987; Scribner and Stuwe 1994, and see Jacobson et al. (2004) for potential environmental effects on T). In order to estimate N_A, r and b_f, we used time series of detailed census data that included the counts for different age classes of either sex. Such counts were available for a limited period of time for the demes considered here, but also for further demes in the Swiss Alps. From these data, we estimated the ratio a of N_A to the total census size N_c as the number of adults ($\geq$ 3 years of age) of either sex divided by the total number of individuals. Similarly, we estimated r as the number of adult males divided by the total number of adults. To obtain an estimate of b_f, we divided the number of kids ($<$ 1 year of age) of either sex by the number of adult females (*cf.* Nievergelt 1966). Recall that we had defined b_f with respect to offspring *reared to independence*. Ibex kids are independent at an age of about six to twelve months. Strictly speaking, we might thus have overestimated b_f, since some of the juveniles younger than one year might still have died until independence. However, different studies concluded that, once ibex kids survived the first six weeks, their mortality is very low until the age of one year (Nievergelt 1966). Our estimate should therefore be reliable. For each of a, r and b_f, the estimates were more or less constant over time and remarkably similar among the 28 (24 for b_f) demes for which these estimates were available, although the demes varied substantially in their demographic history (Table 5.5). We therefore considered it justified to use the same estimate per parameter for each deme. Specifically, we obtained $\hat{r} \approx 0.5$ for the proportion of males, $\hat{a} \approx 0.7$ for the proportion of adults, and $\hat{b}_f \approx 0.4$ kids per female per year. Our estimate of r agrees well with previous studies on Alpine ibex: Jacobson et al.

(2004) found that r varied between 0.43 and 0.53 in the Gran Paradiso deme over a 45-year time series, with a mean below 0.5. The authors also reported a slight correlation between r and the total population density in response to environmental conditions. Scribner and Stuwe (1994) assumed $r \approx 0.5$, and Nievergelt (1966) also reported an estimate of about 0.5. For b_f, the number of offspring reared to independence per female, Stuwe and Grodinsky (1987) observed values between 0.78 and 0.99 in a captive ibex population. The authors argued that b_f should be lower in wild populations. Nievergelt (1966) presented estimates of b_f between 0.44 and 0.74 for six wild demes with differing dynamics. He discussed potential reasons for varying fecundity. One observation was that demes in the colonizing phase showed higher values compared to those close to the carrying capacity. Overall, our estimate seems low compared to those from the previous studies. A potential explanation is that the 28 demes considered here are most likely close to their carrying capacity.

To quantify the standardized variance in male and female adult life span (I_{A_m} and I_{A_f}) we needed an estimate of the mean and the variance in adult life span. Toïgo et al. (2007) presented results on sex and age-specific survival in a wild ibex deme in the Belledonne-Sept-Laux Reserve in France. In their 25-year capture-mark-recapture study, the authors found that both females and males show a highly conservative life-history tactic. Prime-aged (2–8 years) and old adults (8–13) enjoyed very high survival, and mortality increased only afterwards, at senescence (> 13 years). This pattern is rather exceptional among ungulates, especially for males. It may be explained by a conservative male reproductive tactic: mainly dominant males get access to matings, and dominance is correlated with size, body weight and horn size. So, by surviving to an advanced age, males may reach high reproductive success (Willisch 2009; Willisch and Neuhaus 2009, 2010; Willisch et al. 2011). The data by Toïgo et al. (2007) suggest a slight difference between males and females, though. While for male survival was very high up to about ten years and then dropped clearly, for females, survival decreased much less strongly and more linearly after about eight years. The pattern for male ibex reported by Toïgo et al. (2007) is in agreement with previous results by Nievergelt (1966). In order to capture the high survival of prime-aged and old adults, but also the difference in survival between senescent males and females, we modelled adult life span as a random variable that follows a negative binomial distribution. For females, we got a good fit to the data by Toïgo et al. (2007) with a mean adult life span, A_f, of six years and a dispersion parameter ν (the shape parameter of the gamma mixing distribution) of 1. For males, we obtained good agreement with Toïgo et al. (2007) and Nievergelt (1966) with $A_m = 6$ years and $\nu = 4$. The advantage of parameterizing via the negative binomial distribution is that the variance can be expressed as a function of the two parameters: $\mathrm{var}(A_\gamma) = A_\gamma + A_\gamma^2/\nu$, where γ is m for males and f for females. From this, we obtained estimates of the standardized variances in adult lifespan, I_{A_m} and I_{A_f}, of 0.417 and 1.167, respectively.

Toïgo et al. (2002) reported data from a wild population that imply an estimate of the proportion of twin births (z) of ~ 0.08. In contrast, results by Stuwe and Grodinsky (1987) imply a twin birth rate of ~ 0.17, but for a captive population. We used $\hat{z} \approx 0.08$, which (combining with $\hat{b}_f \approx 0.4$ from above, and using equation (A13) in the Appendix resulted in an estimate of $\hat{\rho}_f \approx 0.370$. Scribner and Stuwe (1994) have previously estimated that about 50% of all females reproduce. Toïgo et al. (2002) found that reproductive success differed between

colonizing and well-established demes. For colonizing demes, they reported proportions of 0.4 to 0.8; for well-established ones, they found values between 0.3 and 0.5 (*cf.* Stuwe and Grodinsky 1987, who obtained higher estimates for a zoo population). Our estimate $\hat{\rho}_f$ is therefore in the range of previously reported values. Plugging into equation (A14) in the Appendix our estimates $\hat{z} \approx 0.08$ and $\hat{\rho}_f \approx 0.37$, we obtained for the standardized variance in female seasonal fecundity $\hat{I}_{b_f} \approx 1.870$. Our previous investigations suggested an estimate for the proportion of dominant males of $\hat{\rho}_m \approx 0.2$ (details not shown). Combining with the estimates of $\hat{r} \approx 0.5$ and $\hat{\rho}_f \approx 0.37$ from above, and using equation (A15) in the Appendix, we obtained an estimate of the standardized variance in male seasonal fecundity of $\hat{I}_{b_m} \approx 6.7$. This is in good agreement with a recent study by Willisch et al. (2011) who presented point estimates ranging from 4.8 to 8.0 for different models using paternities inferred from genetic data.

5.9.7 Illustration of the matrix iteration approach

Figures 5.13 and 5.14 illustrate the approach we used to make joint inference about the parameters, given the vector of observed allele frequencies. In essence, transition matrices that describe the evolutionary processes are iterated, and the likelihood of the parameters is then computed from elements of these matrices. For details, see the main text.

In Figures 5.13 and 5.14 the likelihood-based inference is motivated in the following way. The vertical bars – one for each deme – give the density of the focal allele frequency at the time of sampling, p_{sampling}. The darker the shading, the higher the density for the respective value of p_{sampling}. These densities were obtained by iterating the deme-specific transition matrices, as explained in the main text (to save computing time, a slightly simplified demographic scenario with constant deme sizes was used here; *cf.* Figure 5.12). The black circles denote the observed focal allele frequency in each deme. The likelihood-based inference may then be understood in the following way: The darker the vertical bar at the position of a circle, the more support there is for the parameter combination used to generate the density of p_{sampling}. Combining over all demes, one obtains the likelihood of the parameter combination given the observed data – which is equivalent to the probability of the data given the parameters. This approach makes use of the full information contained in the observed focal allele frequencies. The demes are arranged according to their effective size, such that the effect of genetic drift decreases from left to right. The number of individuals sampled per deme (numbers in parentheses after deme names) also has an influence on the shape of the density of p_{sampling}: The higher the sampling size, the smoother the transition between different shades of gray in the vertical bars. The approach thus has the advantage of accounting both for genetic drift and sampling. Different evolutionary scenarios for under- and overdominance are compared.

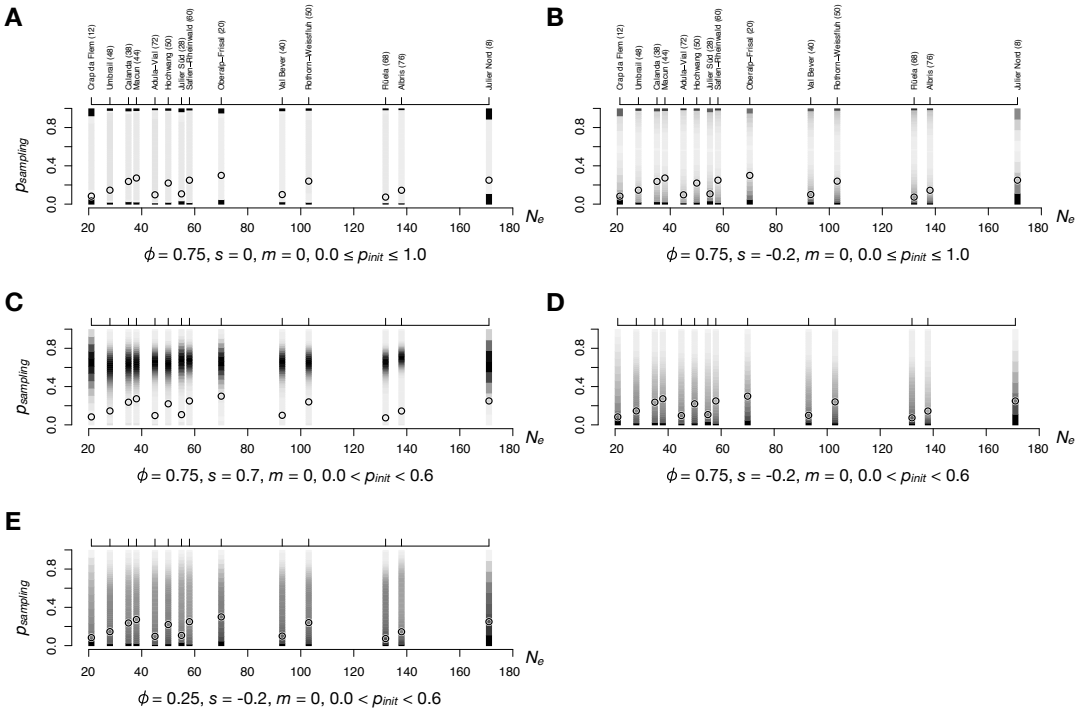

Figure 5.13: Illustration of likelihood-based inference with the matrix iteration approach for under- and overdominance. Details and symbols are explained in the corresponding text of the SI. Each panel shows a different evolutionary scenario, with parameter combinations given below each panel. Specifically, (A) versus (B) shows the contrast between drift-only and selection. (B) versus (D) illustrates the effect of limiting the range of the initial allele frequency $p_{\rm init}$ (see main text). (C) versus (D) shows the effect of changing the sign of the selection coefficient s: In (C), $s > 0$ and there is overdominance (balancing selection) with a stable polymorphic equilibrium at $\phi = 0.75$. In (D), $s < 0$ and there is underdominance with an unstable internal equilibrium at $\phi = 0.75$; the dynamics depend on $p_{\rm init}$. Finally, (D) versus (E) compares two values of ϕ and hence gives an intuition for how inference on the degree of dominance is possible. For intermediate dominance, see Figure 5.14.

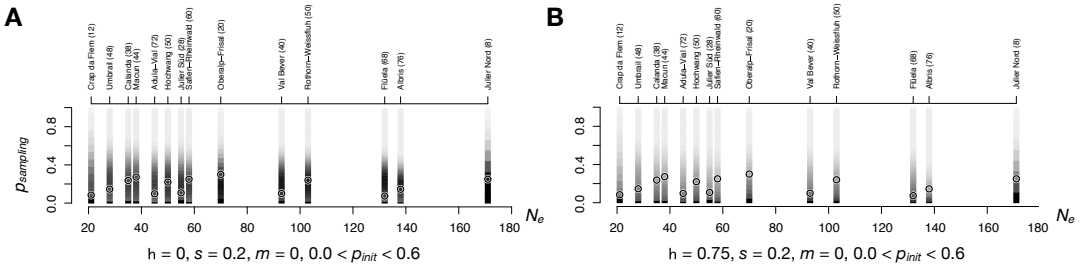

Figure 5.14: Illustration of likelihood-based inference with the matrix iteration approach for intermediate dominance. Details are as in Figure 5.14. The difference is that there is intermediate dominance here, with directional selection against the focal allele (see equation (5.7) in the main text for the definition of fitnesses). (A) The focal allele is fully recessive. (B) The focal allele is partially dominant. In both cases, the allele frequency will approach the equilibrium value of 0. However, before the equilibrium is reached, there is information in the data about the degree of dominance. The approach is therefore appropriate for cases in which the sampled populations have not necessarily reached evolutionary stasis.

5.10 Supporting information: Additional results

5.10.1 Statistical correlation between age at sampling and genetic composition

Heterozygosity at OLADRB2

The distribution of age at sampling in the different demes is shown in Figure 5.15 and the probability of being heterozygous given a certain age at sampling is given in Figure 5.16. In the following, we present detailed results on the relationship between the response (heterozygosity) and the predictors (age at sampling, deme and sex) inferred via logistic regression.

Table 5.11 summarises the GLM fitting process and its result for data_mhc_23. The model with most support (minimum AIC_i) included age as the only predictor. The model with age and sex as predictors also had substantial support ($\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min} = 1.56$). In the model with age as the only predictor, age had a significant negative effect, -0.0584 (95% confidence interval: $[-0.1118, -0.0078]$), on the logit of the probability of an individual being heterozygous at OLADRB2 ($p \sim 0.0273$). For the model with age and sex as predictors, the effect of age was -0.0587 $[-0.1118, -0.0083]$ on the logit scale ($p \sim 0.0257$), and males showed a non-significantly lower chance of being heterozygous compared to females (-0.1433 $[-0.5698$, $0.2841]$ on the logit scale, $p \sim 0.5099$).

Table 5.11: Stepwise model selection via AIC for the explanation of heterozygosity at OLADRB2, using data_mhc_23.

Step	Model i	Add./rem.	Df	Dev	AIC_i	Δ_i	w_i	e_i
1	het ~ age + deme + sex							
	.	− deme	418	499.46	505.46	*1.56*	0.23	0.46
	.	− sex	397	461.81	509.81	5.91	0.03	0.05
	.	none	396	461.68	511.68	7.78	0.01	0.02
	.	+ age:sex	395	461.65	513.65	9.75	0.00	0.01
	.	+ age:deme	374	420.43	514.43	10.53	0.00	0.01
	.	− age	397	466.76	514.76	10.86	0.00	0.00
	.	+ deme:sex	375	434.53	526.53	22.63	0.00	0.00
2	het ~ age + sex							
	.	− sex	419	499.90	503.90	*0.00*	0.49	1.00
	.	+ age:sex	417	499.22	507.22	3.32	0.09	0.19
	.	− age	419	504.71	508.71	4.81	0.04	0.09
3	het ~ age							
	.	− age	420	505.05	507.05	3.15	0.1	0.21

Add./rem., the term that was added (+) or removed (−) relative to the current model; Df, residual degrees of freedom; Dev, residual deviance; $\mathrm{AIC}_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \mathrm{AIC}_i - \min(\mathrm{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) \,/\, \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; het, heterozygosity at the OLADRB2 locus (binary response); age, age at sampling (continuous); deme, factor with 23 levels; sex, factor with 2 levels. Only those models are shown which were visited during the stepwise fitting procedure, and duplicate lines were removed. Interaction terms were added only if the corresponding main effects were also present.

Table 5.12 lists the area under the ROC (AUC) for some models of interest (first two columns). As expected, the complete model (with all interaction terms included) resulted in

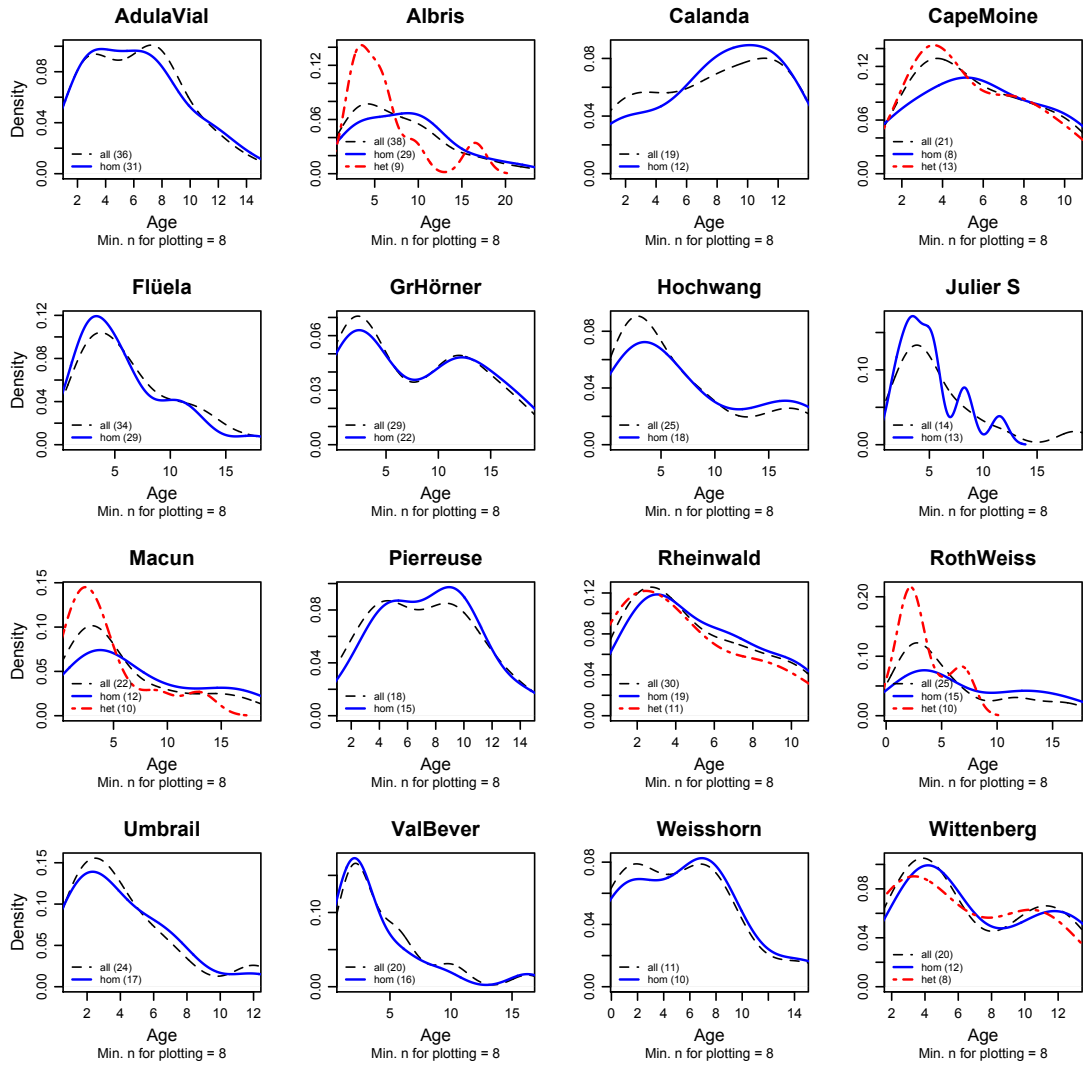

Figure 5.15: Distribution of age at sampling depending on the genotype (zygosity) at the OLADRB2 microsatellite. One plot is shown for each deme in which both alleles (277 and 293) occur and for which at least eight samples ($n = 8$) were available. The dashed (black) line applies when all individuals are pooled, the solid (blue) line applies to homozygous individuals, and the dot-dashed (red) line to heterozygotes. For five demes all three lines could be plotted. For four of them (Albris, Cape au Moine, Macun, Rothorn-Weissfluh), there seems to be a relative excess of heterozygotes at lower ages and, correspondingly, a heterozygote deficiency at higher ages. For deme Rheinwald this pattern is also true, but much weaker, and for Wittenberg it does not apply. Numbers in parentheses give the numbers of data points from which the densities were estimated. Abbreviated deme names are used (see Table 5.9).

the largest AUC, but was prone to overfitting. Comparing to Table 5.11 suggests that a good compromise between the two approaches for model choice via AIC and AUC is provided by the model that includes all predictors, but no interactions (het $\sim$ age + deme + sex). Figure 5.17 compares the ROC of this model to the curve of the best model chosen with AIC (het $\sim$ age) and the curve for the model that maximises the AUC (het $\sim$ age $*$ deme $*$ sex). For the 'best compromise' model (het $\sim$ age + deme + sex), age had a significant negative effect (-0.0625 $[-0.1199, -0.0080]$, $p \sim 0.0281$), deme had a marginally significant joint effect ($p \sim 0.063$) and sex had no significant effect. The effect of deme was mainly caused by significantly positive

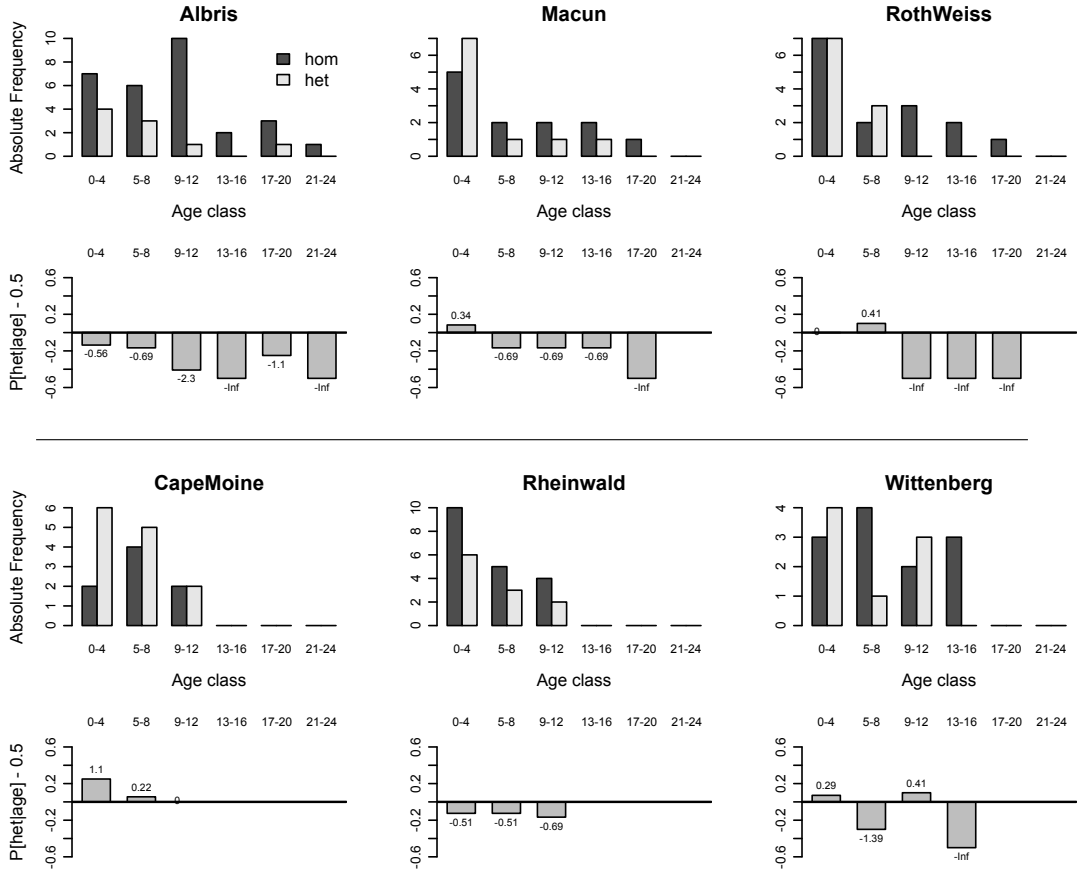

Figure 5.16: Absolute and relative frequencies of OLADRB2 genotypes (zygosity) as a function of age at sampling. Plots are shown for demes in which both OLADRB2 alleles were present and in which the sample size of either homozygotes and heterozygotes was at least eight. The first and the third row show the absolute frequencies of homozygotes and heterozygotes as a function of age class. The second and fourth row illustrate the probability that an individual is heterozygous given it was in a certain age class when culled, Pr[het|age], as a function of age class. The numbers on top of and below the bars give the log odds of this probability. For all demes, there is a tendency for Pr[het|age] to dicrease with increasing age at sampling. The observation is affected by the size of the age class, (4 years as shown here) but the general trend applies for other sizes, too (not shown). In general, the ratio of heterozygotes to homozygotes decreases as a function of age at sampling. Abbreviated deme names are used (see Table 5.9).

effects of the levels Calanda, Macun, Safien-Rheinwald, Rothorn-Weissfluh, Wittenberg (all $p < 0.05$), and Cape au Moine ($p < 0.001$). These deme-specific effects are likely due to differences in ancestral genetic composition and demography, *i.e.* genetic drift (Biebach and Keller 2009, 2010). We conclude that both strategies for model selection gave preference to a model in which age at sampling is included as a predictor. Independently of the exact model chosen, age at sampling had a significant negative effect on the probability of an individual being heterozygous at OLADRB2. These results apply to the set data_mhc_23.

For the set data_mhc_14, model selection based on AIC suggested as the best model the one with age as the only predictor, as for data_mhc_23 (Table 5.13). For this model, age at sampling had a negative effect on the probability of being heterozygous (-0.0595 [-0.1242, 0.0007] on

the logit scale), the effect being marginally significant ($p \sim 0.0608$). The models (het $\sim$ age + sex) and (het $\sim$.) also had substantial support ($\Delta_i < 2$, Table 5.13). Importantly, the latter is the one with no predictors, which means that there is not much scope for explaining heterozygosity in data set data_mhc_14. For (het $\sim$ age + sex), age again had a marginally significant negative effect (-0.0594 [-0.1233, 0.0002], $p \sim 0.0584$), while males had a non-significantly lower probability of being heterozygous than females (-0.2554 [-0.7649, 0.2538] on the logit scale, $p \sim 0.3247$). The performance of the competing models in terms of the AUC is shown in Table 5.12 (right column).

Overall, we found that the probability of an individual being heterozygous at OLADRB2 decreased with increasing age at sampling. This was true independently of the exact model structure, as long as age at sampling was included as a predictor. The negative effect was significant for the larger data set (data_mhc_23), but only marginally significant for the smaller one (data_mhc_14). Recall that data_mhc_14 contains those samples that we used for the matrix iteration approach (long-term analysis). In contrast, data_mhc_23 contains samples from additional demes, and statistical power might be higher there.

Table 5.12: Discriminating power of different GLMs explaining heterozygosity at OLADRB2.

Model	AUC_{23}	AUC_{14}
het $\sim$ age	0.560	0.556
het $\sim$ deme	0.683	0.661
het $\sim$ sex	0.516	0.530
het $\sim$ age + deme	0.679	0.678
het $\sim$ age + sex	0.568	0.579
het $\sim$ deme + sex	0.683	0.663
het $\sim$ age + deme + sex	0.703	0.684
het $\sim$ age + deme + age:deme	0.755	0.750
het $\sim$ age * deme * sex	0.834	0.812

AUC, the area under the ROC, the subscripts 23 and 14 referring to the data sets data_mhc_23 and data_mhc_14, respectively (see text for details). The larger the AUC, the higher the discriminating power of a model, and for random classification AUC = 0.5.

Heterozygosity at other MHC-linked markers

Recall that we found a negative correlation between age at sampling and heterozygosity at the OLADRB2 marker linked to the MHC class II gene DRB: increasing age showed a negative effect on the probability of an individual being heterozygous. This effect was significant when we used samples from all demes with both the ibex and the 'goat' allele present (data_mhc_23), and marginally significant when we considered only samples from those demes used in the matrix iteration approach (data_mhc_14).

For set data_mhc_23 we observed a significant negative relationship (-0.0557 [-0.1090, -0.0051], $p \sim 0.0351$) also for allele 184 of the OLADRB1 locus, a marker physically linked to both OLADRB2 and MHC class II genes. The result is not surprising given that allele 184 is in strong linkage disequilibrium with allele 277 of OLADRB2. None of the other alleles (174, 178, 170) of OLADRB1 showed a significant relationship between heterozygosity and age. For

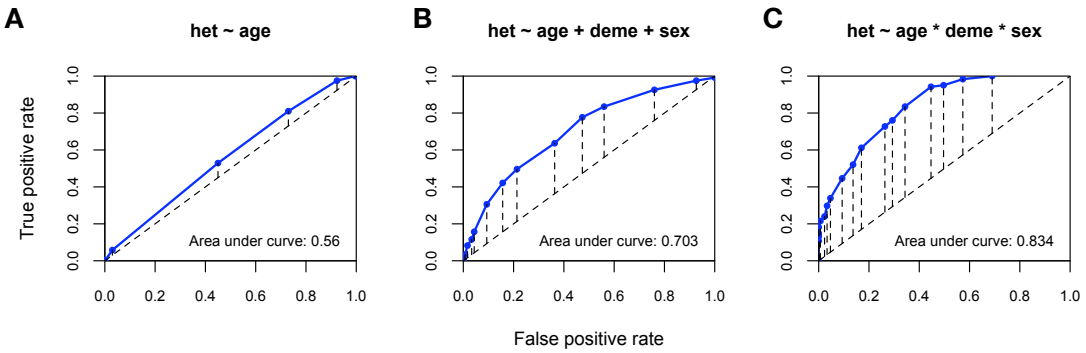

Figure 5.17: Discriminating power of different GLMs explaining heterozygosity at OLADRB2 as a function of age at sampling. (A) The simplest model with age as the only predictor. (B) An intermediate model with age, deme and sex as predictors (best compromise between model selection via AIC versus AUC). (C) The model with highest discriminating power, including age, deme, sex and all interaction terms as predictors. The graphs apply to data set data_mhc_23 (see text for details).

Table 5.13: Stepwise model selection via AIC for the explanation of heterozygosity at OLADRB2, using data_mhc_14.

Step	Model i	Add./rem.	Df	Dev	AIC_i	Δ_i	w_i	e_i
1	het ~ age + deme + sex							
	.	− deme	304	353.62	359.62	*1.03*	0.21	0.6
	.	+ age:deme	278	303.13	361.13	2.54	0.1	0.28
	.	− sex	292	334.00	364.00	5.41	0.02	0.07
	.	none	291	333.51	365.51	6.92	0.01	0.03
	.	− age	292	337.10	367.10	8.51	0.00	0.01
	.	+ age:sex	278	333.10	367.10	8.51	0.00	0.01
	.	+ deme:sex	278	324.43	382.43	23.84	0.00	0.00
2	het ~ age + sex							
	.	− sex	305	354.59	358.59	*0.00*	0.35	1.00
	.	− age	305	357.43	361.43	2.84	0.08	0.24
	.	+ age:sex	303	353.54	361.54	2.95	0.08	0.23
3	het ~ age							
	.	− age	306	358.34	360.34	*1.75*	0.14	0.42

Add./rem., the term that was added (+) or removed (−) relative to the current model; Df, residual degrees of freedom; Dev, residual deviance; $\text{AIC}_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \text{AIC}_i - \min(\text{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) \,/\, \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; het, heterozygosity at the OLADRB2 locus (binary response); age, age at sampling (continuous); deme, factor with 14 levels; sex, factor with 2 levels. Further details as in Table 5.11.

OMHC1 (biallelic) we did not observe a significant relationship between heterozygosity and age at sampling.

For set data_mhc_14, no other MHC-linked allele showed an effect on heterozygosity with a p-value equal to or smaller than the one of allele 277 of OLADRB2. Allele 184 of OLADBR1 showed a nonsignificant negative effect of age (-0.0294 [-0.3462, 0.2486] on the logit scale; $p \sim 0.8407$).

Putatively neutral markers

From the total of 44 other available markers we excluded seven, because, in previous analyses (Biebach and Keller 2009), they have either shown signatures of spatially uniform (BM4208) or spatially heterogeneous selection (OARHH62), are linked to a quantitative trait locus (ETH10), to MHC class II genes (BM1258, BM1818, on chromosome 23), to INFG (OARKP6), or are not in Hardy-Weinberg equilibrium (HWE) (SR-CRSP07).

For the remaining 37 markers, a total of 130 stepwise searches for model selection were performed (one search for each allele). Among these, 22 (16.9%) resulted in a 'best' model that included age as main effect for the data set data_mhc_23. However, this effect was significant at a threshold of 0.05 for only three (2.3%) alleles. The first was allele 134 of marker MILSTS076 (seven alleles, chromosome 9 (Vaiman et al. 1996, citing Kemp et al. (1995)), for which age had a positive effect (0.1009) on heterozygosity ($p \sim 0.0049$). The second was allele 151 of marker OARFCB48 (2 alleles, chromosome 17 (Vaiman et al. 1996, citing Bishop et al. (1994)), for which age had a negative effect (-0.1333, $p \sim 0.0126$). The third was allele 123 of marker MCM73 (four alleles in total, chromosome 4 (Vaiman et al. 1996, citing Crawford et al. (1995)), for which age had a negative effect (-0.0466) on heterozygosity ($p \sim 0.0447$).

For data_mhc_14, 24 (21.2%) out of 113 best models (one for each allele) contained age at sampling as a main effect. This effect was significant at a level equal to or lower than that of allele 277 of OLADRB2 for six alleles (5.3%). For the following four of them the effect of age was positive: allele 134 of MILSTS076 (see above), allele 115 of MCM152 (3 alleles, chromosome 13, (Mainguy et al. 2005, citing Crawford et al. (1995)), and allleles 100 and 116 of SR-CRSP25 (3 alleles, chromosome unknown, (Maddox et al. 2001; Maudet et al. 2002)). For the remaining two alleles – allele 272 of INRABERN185 (4 alleles, chromosome 18, (Luikart et al. 1999)) and allele 151 of OARFCB48 (see above) – the effect was negative (-0.0540 and -0.2012 on the logit scale, with $p \sim 0.0466$ and $p \sim 0.0222$ respectively).

Overall, this suggests that there was no genome-wide negative effect of age at sampling on heterozygosity. However, the significant relationship found for a small number of alleles at neutral markers wait for an explanation. They might just correspond to the proportion expected under the null hypothesis. Neither of those alleles are in linkage disequilibrium with allele 277 of OLADRB2 (data not shown).

Multilocus-heterozygosity at putatively neutral markers

For data_mhc_23, the best model was the one with deme as the only predictor, where deme had a highly significant overall effect ($p < 0.0001$). Two further models with substantial support ($\Delta_i < 2$) were the one with deme and sex as predictors and the one with age and deme as predictors. However, in the former, the effect of sex (males versus females) was not significant (-0.0085 [-0.0518, 0.0348] on the normal scale, $p > 0.7$), and in the latter the effect of age was not significant (0.0009 [-0.0041, 0.0059], $p > 0.7$). For data_mhc_14, the model without any predictors had most support. The two models with either sex or age as the only predictor also had substantial support ($\Delta_i < 2$), but these predictors had non-significant effects (-0.0196 [-0.0695, 0.0304], $p > 0.40$ for males versus females, and 0.0003 [-0.0054, 0.0059] $p > 0.90$ for age). Deme had no effect anymore compared to data_mhc_23 because two demes with strong

effects (Pierreuse-Gummfluh, Wittenberg) are not included in data_mhc_14. Taken together, these results suggest that there is no genome-wide relationship between heterozygosity and age at sampling, which confirms the previous result obtained by fitting models for individual alleles.

Pairwise logistic regression of OLADRB2 genotypes

We start by presenting the results for data set data_mhc_23 and will come to data_mhc_14 later. Four models explaining the contrast between heterozygotes (A_1A_2) and goat homozygotes (A_1A_1) have substantial support. Two of them include age as a predictor (Table 5.14). In those, the probability of A_1A_2 versus A_1A_1 tends to decrease with increasing age at sampling. However, none of the effects is significant (Table 5.15). For the contrast between goat homozygotes (A_1A_1) and ibex homozygotes (A_2A_2), there are also four models with substantial support (Table 5.16). Again, none of the respective predictors has a significant effect, but the probability of A_2A_2 relative to A_1A_1 tends to decrease with increasing age (Table 5.17). There are two models with substantial support in explaining the difference between heterozygotes and ibex homozygotes (A_2A_2), both including age as a predictor (Table 5.18): Both suggest that, the older individuals were at sampling, the higher the probability that they had the A_2A_2 genotype compared to the A_1A_2 genotype. The effect of age is significant for both models (0.057 on the logit scale, $p < 0.035$; Table 5.19). These results agree with our previous finding that the probability of being heterozygous decreases with age at sampling. Additionally, they suggest that this effect must be due to lower survival of heterozygotes compared to ibex homozygotes. On the other hand, the contrasts between heterozygotes and goat homozygotes, as well as between the two homozygote genotypes, can only be weakly explained by age at sampling. They tend to suggest heterozygote disadvantage, but directional selection against the 'goat' allele (*i.e.* intermediate dominance) cannot be excluded with certainty. We assume that the lack of significance for the two comparisons under question (A_1A_2 versus A_1A_1, and A_2A_2 versus A_1A_1) is due to the small number of A_1A_1 individuals in the sample.

We now turn to data set data_mhc_14. For the contrasts in probability between genotypes A_1A_2 and A_1A_1, and between A_2A_2 and A_1A_1, the results were completely analogous to those obtained above with data set data_mhc_23 (Tables 5.20 and 5.21, and Tables 5.22 and 5.23, respectively). The outcome for the probability of being homozygous for the ibex allele A_2A_2 compared to heterozygous (A_1A_2) was more intricate. Three models had substantial support, all of them including age as a predictor. Among those, two additionally included deme and the interaction of deme with age as predictors (Table 5.24). The best model was the one with age, deme and the interaction between age and deme as predictors (a1a2.a2a2 $\sim$ age + deme + age:deme). In there, age at sampling had no significant effect (0.0112 on the logit scale, $p \sim 0.9404$), deme had a marginally significant overall effect ($p \sim 0.0620$), but the interaction between age and deme had no significant overall effect ($p \sim 0.4000$; Table 5.26). The overall effect of deme was due to significant negative single-level effects of demes Macun (-2.5139, $p \sim 0.0435$) and Rothorn-Weissfluh (-2.8550, $p \sim 0.0265$), and a marginally significant negative single-level effect of deme Oberalp-Frisal (-3.6980, $p \sim 0.0672$; Table 5.26). The second best model additionally included sex as a predictor with non-significant effect (Table 5.27). The third best model was (a1a2.a2a2 $\sim$ age + sex), where age had a marginally significant positive effect on a1a2.a2a2 (0.0561; $p \sim 0.751$) and sex had no effect ($p \sim 0.2639$; Table 5.27). To better

Table 5.14: Stepwise model selection for the contrast between the OLADRB2 genotypes A_1A_1 and A_1A_2, using data_mhc_23.

Step	Model i	Add./rem.	Df	Dev	AIC_i	Δ_i	w_i	e_i
1	a1a1.a1a2 ∼ age + deme + sex							
	.	− deme	129	72.425	78.425	*1.02*	0.18	0.60
	.	− sex	108	57.675	105.675	28.27	0.00	0.00
	.	none	107	57.404	107.404	30.00	0.00	0.00
	.	− age	108	59.548	107.548	30.15	0.00	0.00
	.	+ age:sex	106	56.709	108.709	31.31	0.00	0.00
	.	+ deme:sex	91	38.224	120.224	42.82	0.00	0.00
	.	+ age:deme	90	42.151	126.151	48.75	0.00	0.00
2	a1a1.a1a2 ∼ age + sex							
	.	− sex	130	73.401	77.401	*0.00*	0.30	1.00
	.	− age	130	74.612	78.612	*1.21*	0.16	0.55
	.	+ age:sex	128	71.585	79.585	2.18	0.10	0.34
3	a1a1.a1a2 ∼ age							
	.	− age	131	75.725	77.725	*0.32*	0.26	0.85

Add./rem., the term that was added (+) or removed (−) relative to the current model; Df, residual degrees of freedom; Dev, residual deviance; $\text{AIC}_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \text{AIC}_i - \min(\text{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) \,/\, \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; *a1a1.a1a2*, binary response contrasting the OLADRB2 genotypes A_1A_1 versus A_1A_2 (0 for A_1A_1 and 1 for A_1A_2); age, age at sampling (continuous); deme, factor with 23 levels; sex, factor with 2 levels. Only those models are shown which were visited during the stepwise fitting procedure, and duplicate lines were removed. Interaction terms were added only if the corresponding main effects were also present.

account for the effect of deme, but at the same time avoid overfitting, we simplified the model (a1a2.a2a2 ∼ age + deme + age:deme) by pooling samples from demes other than Macun, Oberalp-Frisal and Rothorn-Weissfluh. We explored four models of that kind, among which one had clearly most support. It explains the probability of A_2A_2 relative to A_1A_2 by age at sampling, by whether or not the deme is in the set {Macun, Oberalp-Frisal, Rothorn-Weissfluh}, and by the interaction of these two predictors (Table 5.25). This model is instructive: it suggests that the probability of being homozygous for the ibex allele compared to heterozygous is significantly lower in demes Macun, Oberalp-Frisal and Rothhorn-Weissfluh relative to the other demes (-2.0652 on the logit scale, $p < 0.001$). However, within these three demes, increasing age at sampling is significantly positively correlated with the probability of being homozygous for the ibex allele compared to being heterozygous (0.2142, $p < 0.02$; Table 5.28). Overall, the results for data_mhc_14 support those we obtained with data_mhc_23: the older an individual at sampling, the lower the probability that it was heterozygous (A_1A_2) compared to homozygous for the ibex allele (A_2A_2). In addition, the results for data_mhc_14 reveal an effect of deme – namely demes Macun, Oberalp-Frisal and Rothorn-Weissfluh – which most likely reflects the founder effect.

Table 5.15: Estimates of effects on the contrast between OLADRB2 genotypes A_1A_1 and A_1A_2 for data_mhc_23.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p	
a1a1.a1a2 ~ age							
	(Intercept)	3.1180	0.6080	2.0253	4.4403	<0.0001	***
	age	−0.1113	0.0712	−0.2515	0.0335	0.1180	
a1a1.a1a2 ~ 1							
	(Intercept)	2.3979	0.3149	1.8281	3.0746	<0.0001	***
a1a1.a1a2 ~ age + sex							
	(Intercept)	2.7930	0.6616	1.6133	4.2437	<0.0001	***
	age	−0.1057	0.0700	−0.2444	0.0361	0.1311	
	sexmale	0.6407	0.6595	−0.6254	2.0326	0.3313	
a1a1.a1a2 ~ sex							
	(Intercept)	2.0971	0.4005	1.3815	2.9757	<0.0001	***
	sexmale	0.6754	0.6527	−0.5738	2.0574	0.3007	

Model, the models with substantial support in explaining a1a1.a1a2 as a function of the predictors (*cf.* Table 5.14 for a summary on model choice); Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code: *** for $0 < p \leq 0.001$, ** for $0.001 < p \leq 0.01$, * for $0.01 < p \leq 0.05$, · for $0.05 < p \leq 0.1$ and ' ' for $0.1 < p \leq 1$.

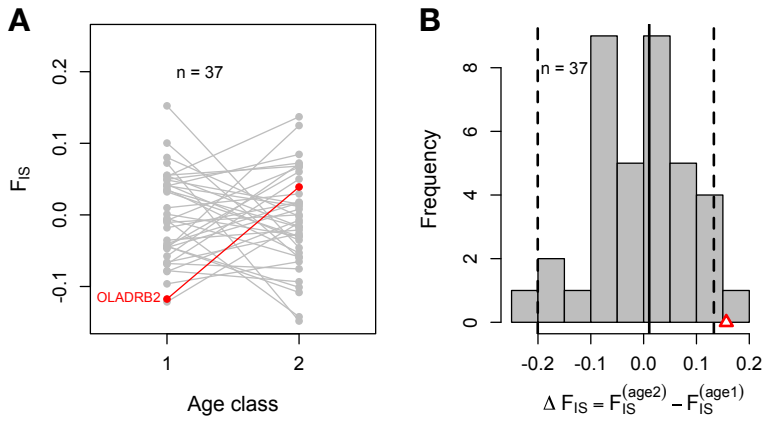

Figure 5.18: Change in deviation from HWE measured by F_{IS} as a function of age. Age class 1 comprises individuals of age up to and including 5.25 years, and age class 2 those of age older than 5.25 years. (A) Gray symbols belong to 37 neutral microsatellites, and the symbols represents the MHC-linked marker OLADRB2. (B) The change in F_{IS} as a function of age (ΔF_{IS}) for OLADRB2 (red triangle) is compared to the distribution of ΔF_{IS} obtained for the 37 neutral markers. Vertical lines represent the median (solid) and the 95% credibility interval (dashed) of the neutral distribution. Plots are shown for the data set data_mhc_23 (see text for details).

Table 5.16: Stepwise model selection for the contrast between the OLADRB2 genotypes A_1A_1 and A_2A_2, using data_mhc_23.

Step	Model i	Add./rem.	Df	Dev	AIC$_i$	Δ_i	w_i	e_i
1	a1a1.a2a2 ∼ age + deme + sex							
	.	− deme	297	92.243	98.243	*1.92*	0.13	0.38
	.	− age	276	74.750	122.750	26.43	0.00	0.00
	.	− sex	276	76.314	124.314	27.99	0.00	0.00
	.	none	275	74.701	124.701	28.38	0.00	0.00
	.	+ age:sex	274	74.268	126.268	29.95	0.00	0.00
	.	+ age:deme	255	56.707	146.707	50.39	0.00	0.00
	.	+ deme:sex	256	59.728	147.728	51.41	0.00	0.00
2	a1a1.a2a2 ∼ age + sex							
	.	− age	298	92.544	96.544	*0.22*	0.31	0.89
	.	− sex	298	93.847	97.847	*1.53*	0.16	0.47
	.	+ age:sex	296	91.767	99.767	3.45	0.06	0.18
3	a1a1.a2a2 ∼ sex							
	.	− sex	299	94.321	96.321	*0.00*	0.34	1.00
4	a1a1.a2a2 ∼ 1							
	.	+ deme	277	76.454	122.454	26.13	0.00	0.00

Add./rem., the term that was added (+) or removed (−) relative to the current model; Df, residual degrees of freedom; Dev, residual deviance; $\text{AIC}_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \text{AIC}_i - \min(\text{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) / \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; *a1a1.a2a2*, binary response contrasting the OLADRB2 genotypes A_1A_1 versus A_2A_2 (0 for A_1A_1 and 1 for A_2A_2); age, age at sampling (continuous); deme, factor with 23 levels; sex, factor with 2 levels. Only those models are shown which were visited during the stepwise fitting procedure, and duplicate lines were removed. Interaction terms were added only if the corresponding main effects were also present.

Table 5.17: Estimates of effects on the contrast between OLADRB2 genotypes A_1A_1 and A_2A_2 for data_mhc_23.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p	
a1a1.a2a2 ∼ 1							
	(Intercept)	3.2685	0.3072	2.7171	3.9330	<0.0001	***
a1a1.a2a2 ∼ sex							
	(Intercept)	2.8824	0.3884	2.1971	−0.3885	<0.0001	***
	sexmale	0.8312	0.6379	3.7428	2.1890	0.1926	
a1a1.a2a2 ∼ age							
	(Intercept)	3.5796	0.5625	2.5622	4.7912	<0.0001	***
	age	−0.0443	0.0630	−0.1629	0.0885	0.4815	
a1a1.a2a2 ∼ age + sex							
	(Intercept)	3.1347	0.6188	2.0355	4.4943	<0.0001	***
	age	−0.0338	0.0607	−0.1495	0.0937	0.5778	
	sexmale	0.7971	0.6426	−0.4344	2.1616	0.2148	

Model, the models with substantial support in explaining a1a1.a2a2 as a function of the predictors (*cf.* Table 5.16 for a summary on model choice); Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code as in Table 5.15.

Table 5.18: Stepwise model selection for the contrast between the OLADRB2 genotypes A_1A_2 and A_2A_2, using data_mhc_23.

Step	Model i	Add./rem.	Df	Dev	AIC_i	Δ_i	w_i	e_i
1	a1a2.a2a2 ∼ age + deme + sex							
	.	− deme	407	492.11	498.11	*1.41*	0.23	0.49
	.	− sex	386	453.42	501.42	4.72	0.04	0.09
	.	none	385	453.15	503.15	6.45	0.02	0.04
	.	+ age:sex	384	453.15	505.15	8.45	0.01	0.01
	.	+ age:deme	363	411.84	505.84	9.14	0.00	0.01
	.	− age	384	457.96	505.96	9.26	0.00	0.01
	.	+ deme:sex	364	426.63	518.63	21.93	0.00	0.00
2	a1a2.a2a2 ∼ age + sex							
	.	− sex	408	492.70	496.70	*0.00*	0.47	1.00
	.	+ age:sex	406	491.94	499.94	3.24	0.09	0.20
3	a1a2.a2a2 ∼ age							
	.	− age	409	497.47	499.47	2.77	0.12	0.25

Add./rem., the term that was added (+) or removed (−) relative to the current model; Df, residual degrees of freedom; Dev, residual deviance; $\mathrm{AIC}_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \mathrm{AIC}_i - \min(\mathrm{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) \,/\, \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; *a1a2.a2a2*, binary response contrasting the OLADRB2 genotypes A_1A_2 versus A_2A_2 (0 for A_1A_2 and 1 for A_2A_2); age, age at sampling (continuous); deme, factor with 23 levels; sex, factor with 2 levels. Only those models are shown which were visited during the stepwise fitting procedure, and duplicate lines were removed. Interaction terms were added only if the corresponding main effects were also present.

Table 5.19: Estimates of effects on the contrast between OLADRB2 genotypes A_1A_2 and A_2A_2 for data_mhc_23.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p	
a1a2.a2a2 ∼ age							
	(Intercept)	0.5303	0.1884	0.1626	0.9024	0.0049	∗∗
	age	0.0566	0.0266	0.0057	0.1103	0.0334	∗
a1a2.a2a2 ∼ age + sex							
	(Intercept)	0.4361	0.2234	0.0015	0.8789	0.0509	·
	age	0.0568	0.0264	0.0061	0.1101	0.0318	∗
	sexmale	0.1691	0.2188	-0.2607	0.5981	0.4395	*n.s.*

Model, the models with substantial support in explaining a1a2.a2a2 as a function of the predictors (*cf.* Table 5.18 for a summary on model choice); Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code as in Table 5.15.

Table 5.20: Stepwise model selection for the contrast between the OLADRB2 genotypes A_1A_1 and A_1A_2, using data_mhc_14.

Step	Model i	Add./rem.	Df	Dev	AIC_i	Δ_i	w_i	e_i
1	a1a1.a1a2 ∼ age + deme + sex							
	.	− deme	91	64.892	70.892	*1.48*	0.16	0.48
	.	− sex	79	57.675	87.675	18.26	0.00	0.00
	.	none	78	57.404	89.404	19.99	0.00	0.00
	.	− age	79	59.548	89.548	20.14	0.00	0.00
	.	+ age:sex	77	56.709	90.709	21.30	0.00	0.00
	.	+ deme:sex	66	38.224	94.224	24.81	0.00	0.00
	.	+ age:deme	66	42.151	98.151	28.74	0.00	0.00
2	a1a1.a1a2 ∼ age + sex							
	.	− sex	92	65.413	69.413	*0.00*	0.34	1.00
	.	− age	92	67.060	71.060	*1.65*	0.15	0.44
	.	+ age:sex	90	64.564	72.564	3.15	0.07	0.21
3	a1a1.a1a2 ∼ age							
	.	− age	93	67.858	69.858	*0.45*	0.27	0.80

Add./rem., the term that was added (+) or removed (−) relative to the current model; Df, residual degrees of freedom; Dev, residual deviance; $\mathrm{AIC}_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \mathrm{AIC}_i - \min(\mathrm{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) \,/\, \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; $a1a1.a1a2$, binary response contrasting the OLADRB2 genotypes A_1A_1 versus A_1A_2 (0 for A_1A_1 and 1 for A_1A_2); age, age at sampling (continuous); deme, factor with 14 levels; sex, factor with 2 levels. Only those models are shown which were visited during the stepwise fitting procedure, and duplicate lines were removed. Interaction terms were added only if the corresponding main effects were also present.

Table 5.21: Estimates of effects on the contrast between OLADRB2 genotypes A_1A_1 and A_1A_2 for data_mhc_14.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p	
a1a1.a1a2 ∼ age							
	(Intercept)	2.7409	0.6005	1.6583	4.0454	<0.0001	***
	age	−0.1129	0.0706	−0.2532	0.0300	0.1100	
a1a1.a1a2 ∼ 1							
	(Intercept)	2.0209	0.3209	1.4369	2.7071	<0.0001	***
a1a1.a1a2 ∼ age + sex							
	(Intercept)	2.4733	0.6785	1.2684	3.9732	0.0003	***
	age	−0.1046	0.0701	−0.2452	0.0360	0.1360	
	sexmale	0.4845	0.6784	−0.8273	1.9047	0.4751	
a1a1.a1a2 ∼ sex							
	(Intercept)	1.7677	0.4090	1.0306	2.6589	<0.0001	***
	sexmale	0.5837	0.6641	−0.6885	1.9845	0.3794	

Model, the models with substantial support in explaining a1a1.a1a2 as a function of the predictors (*cf.* Table 5.20 for a summary on model choice); Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code: *** for $0 < p \leq 0.001$, ** for $0.001 < p \leq 0.01$, * for $0.01 < p \leq 0.05$, · for $0.05 < p \leq 0.1$ and ' ' for $0.1 < p \leq 1$.

Table 5.22: Stepwise model selection for the contrast between the OLADRB2 genotypes A_1A_1 and A_2A_2, using data_mhc_14.

Step	Model i	Add./rem.	Df	Dev	AIC_i	Δ_i	w_i	e_i
1	a1a1.a2a2 ~ age + deme + sex							
	.	− deme	221	85.369	91.369	*1.62*	0.14	0.45
	.	− age	209	74.750	104.750	15.00	0.00	0.00
	.	− sex	209	76.314	106.314	16.56	0.00	0.00
	.	none	208	74.701	106.701	16.95	0.00	0.00
	.	+ age:sex	207	74.268	108.268	18.52	0.00	0.00
	.	+ age:deme	196	56.707	112.707	22.95	0.00	0.00
	.	+ deme:sex	195	59.728	117.728	27.97	0.00	0.00
2	a1a1.a2a2 ~ age + sex							
	.	− age	222	85.821	89.821	*0.07*	0.31	0.97
	.	− sex	222	87.144	91.144	*1.39*	0.16	0.50
	.	+ age:sex	220	84.802	92.802	3.05	0.07	0.22
3	a1a1.a2a2 ~ sex							
	.	− sex	221	87.753	89.753	*0.00*	0.32	1.00
4	a1a1.a2a2 ~ 1							
	.	+ deme	210	76.454	104.454	14.7	0.00	0.00

Add./rem., the term that was added (+) or removed (−) relative to the current model; Df, residual degrees of freedom; Dev, residual deviance; $\text{AIC}_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \text{AIC}_i - \min(\text{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) / \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; *a1a1.a2a2*, binary response contrasting the OLADRB2 genotypes A_1A_1 versus A_2A_2 (0 for A_1A_1 and 1 for A_2A_2); age, age at sampling (continuous); deme, factor with 14 levels; sex, factor with 2 levels. Only those models are shown which were visited during the stepwise fitting procedure, and duplicate lines were removed. Interaction terms were added only if the corresponding main effects were also present.

Table 5.23: Estimates of effects on the contrast between OLADRB2 genotypes A_1A_1 and A_2A_2 for data_mhc_14.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p	
a1a1.a2a2 ~ 1							
	(Intercept)	2.9634	0.3092	2.4072	3.6311	<0.0001	***
a1a1.a2a2 ~ sex							
	(Intercept)	2.5539	0.3924	1.8586	3.4203	<0.0001	***
	sexmale	0.8720	0.6419	−0.3557	2.2363	0.1744	
a1a1.a2a2 ~ age							
	(Intercept)	3.3092	0.5572	2.2978	4.5067	<0.0001	***
	age	−0.0499	0.0622	−0.1668	0.0818	0.4224	
a1a1.a2a2 ~ age + sex							
	(Intercept)	2.8513	0.6098	1.7632	4.1870	<0.0001	***
	age	−0.0410	0.0597	−0.1546	0.0850	0.4923	
	sexmale	0.8417	0.6454	−0.3948	2.2108	0.1922	

Model, the models with substantial support in explaining a1a1.a2a2 as a function of the predictors (*cf.* Table 5.22 for a summary on model choice); Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code as in Table 5.21.

Table 5.24: Stepwise model selection for the contrast between the OLADRB2 genotypes A_1A_2 and A_2A_2, using data_mhc_14.

Step	Model i	Add./rem.	Df	Dev	AIC$_i$	Δ_i	w_i	e_i
1	a1a2.a2a2 ~ age + deme + sex							
	.	+ age:deme	267	294.48	352.48	*1.63*	0.20	0.44
	.	− deme	281	346.66	352.66	*1.81*	0.18	0.40
	.	− sex	281	325.60	355.60	4.75	0.04	0.09
	.	none	280	324.79	356.79	5.94	0.02	0.05
	.	− age	281	328.07	358.07	7.22	0.01	0.03
	.	+ age:sex	279	324.15	358.15	7.30	0.01	0.03
	.	+ deme:sex	267	316.54	374.54	23.69	0.00	0.00
2	a1a2.a2a2 ~ age + deme + sex + age:deme							
	.	− sex	268	294.85	350.85	*0.00*	0.45	1.00
	.	+ age:sex	266	294.06	354.06	3.21	0.09	0.20
	.	+ deme:sex	254	286.78	370.78	19.93	0.00	0.00

Add./rem., the term that was added (+) or removed (−) relative to the current model; Df, residual degrees of freedom; Dev, residual deviance; AIC$_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \text{AIC}_i - \min(\text{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) \,/\, \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; *a1a2.a2a2*, binary response contrasting the OLADRB2 genotypes A_1A_2 versus A_2A_2 (0 for A_1A_2 and 1 for A_2A_2); age, age at sampling (continuous); deme, factor with 14 levels; sex, factor with 2 levels. Only those models are shown which were visited during the stepwise fitting procedure, and duplicate lines were removed. Interaction terms were added only if the corresponding main effects were also present.

Table 5.25: Performance of models with reduced levels of demes in explaining the contrast between the OLADRB2 genotypes A_1A_2 and A_2A_2, using data_mhc_14.

Model i	Df	Dev	AIC$_i$	Δ_i	w_i	e_i
a1a2.a2a2 ~ age + deme + age:deme	268	294.845	350.845	10.02	0.01	0.01
a1a2.a2a2 ~ age + I(deme = Macun) + I(deme = Oberalp) + I(deme = RothWeiss)	291	339.557	349.557	8.73	0.01	0.01
a1a2.a2a2 ~ age + I(deme = Macun) + I(deme = Oberalp) + I(deme = RothWeiss) + sex)	290	338.317	350.317	9.49	0.01	0.01
a1a2.a2a2 ~ age + I(deme ∈ {Macun, Oberalp, RothWeiss})	293	339.736	345.736	4.91	0.08	0.09
a1a2.a2a2 ~ age + I(deme ∈ {Macun, Oberalp, RothWeiss}) + age:I(deme ∈ {Macun, Oberalp, RothWeiss})	292	332.829	340.829	*0.00*	0.90	1.00

Df, residual degrees of freedom; Dev, residual deviance; AIC$_i = -2\log(L_i) + 2k$, Akaike Information Criterion for model i, where L_i is the likelihood and k_i the number of parameters estimated under model i; $\Delta_i = \text{AIC}_i - \min(\text{AIC}_i)$, italic for the best model and models with substantial support ($\Delta_i < 2$); $w_i = \exp(-\Delta_i/2) \,/\, \sum_{r=1}^{R} \exp(-\Delta_r/2)$, Akaike weight; $e_i = w_i/w_{\max}$, evidence ratio relative to best model; $a1a2.a2a2$, binary response contrasting the OLADRB2 genotypes A_1A_2 versus A_2A_2 (0 for A_1A_2 and 1 for A_2A_2); age, age at sampling (continuous); deme, factor with 14 levels; sex, factor with 2 levels. I(a) denotes an indicator, taking the value 1 if a is true and 0 else. Such indicators are used to reduce the levels of the factor deme to those with a significant effect (*cf.* Tables 5.26 and 5.27).

Table 5.26: Estimates of effects on the contrast between OLADRB2 genotypes A_1A_2 and A_2A_2 for data_mhc_14, part 1.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p
a1a2.a2a2 ∼ age + deme + age:deme						
	(Intercept)	1.7247	1.0130	−0.1281	3.9566	0.0887 ·
	age	0.0112	0.1505	−0.2775	0.3348	0.9404
	demeAlbris	−1.3404	1.2326	−3.9281	0.9922	0.2768
	demeCalanda	−1.2323	1.4625	−4.1998	1.6780	0.3995
	demeCrapFlem	−4.1930	4.1753	−16.7399	3.2346	0.3153
	demeFlueela	1.5101	1.4983	−1.4088	4.6808	0.3135
	demeHochwang	−1.7084	1.2329	−4.3028	0.6204	0.1658
	demeJulier N	104.1874	11434.6510	2931.7533	8710.2284	0.9927
	demeJulier S	62.7910	3469.2199	1020.3946	1146.1607	0.9856
	demeMacun	−2.5139	1.2453	−5.1466	−0.1844	0.0435 ∗
	demeOberalp	−3.6980	2.0203	−8.5578	−0.1176	0.0672 ·
	demeRheinwald	−1.9414	1.2600	−4.5825	0.4415	0.1234
	demeRothWeiss	−2.8550	1.2866	−5.5925	−0.4674	0.0265 ∗
	demeUmbrail	−0.7388	1.2632	−3.3629	1.6902	0.5586
	demeValBever	0.2882	1.3759	−2.4794	3.0748	0.8341
	age:demeAlbris	0.0919	0.1745	−0.2677	0.4357	0.5985
	age:demeCalanda	−0.0164	0.1924	−0.4134	0.3559	0.9319
	age:demeCrapFlem	0.7025	0.8119	−0.4119	3.6197	0.3869
	age:demeFlueela	−0.2087	0.1889	−0.6045	0.1519	0.2692
	age:demeHochwang	0.1310	0.1861	−0.2423	0.5184	0.4816
	age:demeJulier N	−27.2207	2856.1854	NA	300.1088	0.9924
	age:demeJulier S	−4.3348	243.4479	−182.8066	52.5800	0.9858
	age:demeMacun	0.1405	0.1828	−0.2309	0.5081	0.4419
	age:demeOberalp	0.4058	0.3716	−0.1922	1.4697	0.2748
	age:demeRheinwald	0.1207	0.2009	−0.2847	0.5194	0.5482
	age:demeRothWeiss	0.2805	0.2162	−0.1304	0.7593	0.1944
	age:demeUmbrail	−0.0349	0.2082	−0.4553	0.3836	0.8669
	age:demeValBever	−0.1353	0.2016	−0.5547	0.2611	0.5022
	deme (overall, Wald test)					0.0620 ·
	age:deme (overall, Wald test)					0.4000

Model, the models with substantial support in explaining a1a2.a2a2 as a function of the predictors (only one model shown here; *cf.* Table 5.24 for a summary on model choice and Table 5.27 for the remaining models); Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code as in Table 5.21. The single-level effects of demes are relative to the one of deme AdulaVial.

Table 5.27: Estimates of effects on the contrast between OLADRB2 genotypes A_1A_2 and A_2A_2 for data_mhc_14, part 2.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p
a1a2.a2a2 ∼ age + deme + sex + age:deme						
	(Intercept)	1.6387	1.0235	−0.2371	3.8869	0.1094
	age	0.0091	0.1508	−0.2801	0.3329	0.9521
	demeAlbris	−1.3390	1.2319	−3.9266	0.9912	0.2771
	demeCalanda	−1.2239	1.4638	−4.1943	1.6884	0.4031
	demeCrapFlem	−4.4190	4.2359	−16.9898	3.0904	0.2968
	demeFlueela	1.4365	1.4983	−1.4841	4.6057	0.3377
	demeHochwang	−1.7177	1.2328	−4.3125	0.6101	0.1635
	demeJulier N	104.0413	11507.8822	2949.7743	8765.1282	0.9928
	demeJulier S	62.8751	3467.9970	705.4432	831.8066	0.9855
	demeMacun	−2.5231	1.2468	−5.1597	−0.1918	0.0430 ∗
	demeOberalp	−3.6666	2.0124	−8.4971	−0.0906	0.0685 ·
	demeRheinwald	−1.9322	1.2620	−4.5774	0.4546	0.1257
	demeRothWeiss	−2.8228	1.2874	−5.5623	−0.4339	0.0283 ∗
	demeUmbrail	−0.7193	1.2622	−3.3427	1.7062	0.5688
	demeValBever	0.2672	1.3750	−2.5005	3.0500	0.8459
	sexmale	0.1784	0.2946	−0.4015	0.7563	0.5447
	age:demeAlbris	0.0944	0.1743	−0.2649	0.4376	0.5883
	age:demeCalanda	−0.0156	0.1926	−0.4129	0.3571	0.9354
	age:demeCrapFlem	0.7409	0.8244	−0.3911	3.6572	0.3688
	age:demeFlueela	−0.2010	0.1892	−0.5973	0.1599	0.2879
	age:demeHochwang	0.1350	0.1854	−0.2376	0.5196	0.4665
	age:demeJulier N	−27.1689	2873.3136	NA	302.1227	0.9925
	age:demeJulier S	−4.3323	243.2135	−182.6344	52.5176	0.9858
	age:demeMacun	0.1408	0.1824	−0.2301	0.5072	0.4402
	age:demeOberalp	0.3984	0.3671	−0.1967	1.4513	0.2778
	age:demeRheinwald	0.1169	0.2016	−0.2894	0.5170	0.5618
	age:demeRothWeiss	0.2748	0.2160	−0.1358	0.7531	0.2033
	age:demeUmbrail	−0.0397	0.2078	−0.4597	0.3772	0.8486
	age:demeValBever	−0.1368	0.2014	−0.5556	0.2590	0.4969
	deme (overall, Wald test)					0.0700 ·
	age:deme (overall, Wald test)					0.4100
a1a2.a2a2 ∼ age + sex						
	(Intercept)	0.4551	0.2591	−0.0486	0.9702	0.0791 ·
	age	0.0561	0.0315	−0.0038	0.1204	0.0751 ·
	sexmale	0.2918	0.2612	−0.2210	0.8052	0.2639

Model, the models with substantial support in explaining a1a2.a2a2 as a function of the predictors (*cf.* Table 5.24 for a summary on model choice); Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code as in Table 5.21. The single-level effects of demes are relative to the one of deme AdulaVial.

Table 5.28: Estimates of effects on the contrast between OLADRB2 genotypes A_1A_2 and A_2A_2 for data_mhc_14, with two subsets of demes.

Model	Coefficient	Estimate	SE	2.5%	97.5%	p	
a1a2.a2a2 ∼ age + I(deme = Macun) + I(deme = Oberalp) + I(deme = RothWeiss)							
	(Intercept)	0.8006	0.2339	0.3464	1.2654	0.0006	***
	age	0.0566	0.0321	−0.0044	0.1218	0.0775	·
	I(deme = Macun)	−1.0490	0.4665	−1.9708	−0.1197	0.0245	*
	I(deme = Oberalp)	−0.9051	0.6904	−2.2717	0.5238	0.1898	
	I(deme = RothWeiss)	−0.7926	0.4438	−1.6568	0.1020	0.0741	·
a1a2.a2a2 ∼ age + I(deme = Macun) + I(deme = Oberalp) + I(deme = RothWeiss) + sex							
	(Intercept)	0.6446	0.2702	0.1205	1.1830	0.0171	*
	age	0.0557	0.0317	−0.0046	0.1202	0.0786	·
	I(deme = Macun)	−1.0576	0.4679	−1.9821	−0.1257	0.0238	*
	I(deme = Oberalp)	−0.9067	0.6930	−2.2784	0.5269	0.1908	
	I(deme = RothWeiss)	−0.7879	0.4459	−1.6562	0.1106	0.0772	·
	sexmale	0.2959	0.2655	−0.2254	0.8179	0.2652	
a1a2.a2a2 ∼ age + I(deme ∈ {Macun, Oberalp, RothWeiss})							
	(Intercept)	0.8015	0.2339	0.3474	1.2663	0.0006	***
	age	0.0564	0.0321	−0.0045	0.1216	0.0783	·
	I(deme ∈ {Macun, Oberalp, RothWeiss})	−0.9115	0.3145	−1.5282	−0.2905	0.0038	**
a1a2.a2a2 ∼ age + I(deme ∈ {Macun, Oberalp, RothWeiss}) + age:I(deme ∈ {Macun, Oberalp, RothWeiss})							
	(Intercept)	1.0718	0.2595	0.5704	1.5905	<0.0001	***
	age	0.0099	0.0351	−0.0573	0.0813	0.7781	
	I(deme ∈ {Macun, Oberalp, RothWeiss})	−2.0652	0.5627	−3.2153	−0.9923	0.0002	***
	age:I(deme ∈ {Macun, Oberalp, RothWeiss})	0.2142	0.0912	0.0508	0.4161	0.0188	*

Model, the models with substantial support in explaining a1a2.a2a2 as a function of the predictors (*cf.* tables 5.24 and 5.25 for a summary on model choice); Coefficient, name of predictor; Estimate, estimated effect of predictor (on the logit scale); SE, standard error; 2.5% and 97.5%, the limits of the 95% confidence interval; p, p-value (Wald test), significance code as in Table 5.21. I(a) denotes an indicator, taking the value 1 if a is true and 0 else. The estimates for such indicators are given for the effect of I(a) = 1 compared to the default I(a) = 0.

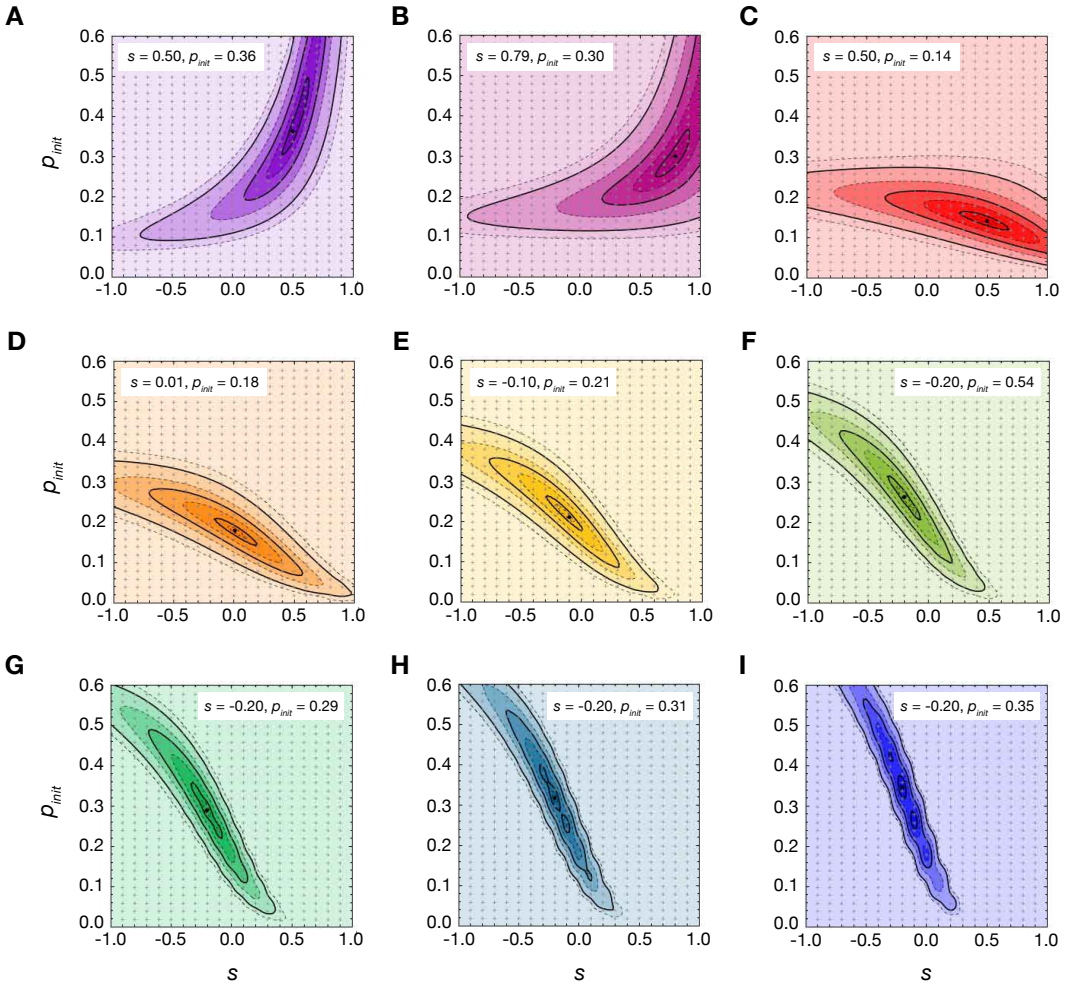

Figure 5.19: Joint likelihood surface of selection coefficient s and initial frequency p_{init} of the 'goat' allele A_1, for under- and overdominance and without migration. Joint maximum-likelihood estimates are given in the boxes. The dotted and solid black lines denote regions of highest posterior density for levels of support of 99%, 95%, 75%, 50%, 25%, 5% and 1%. Crosses denote parameter combinations for which exact values were computed, and the surface was obtained by third-order interpolation. Fitnesses are parameterized as in equation (5.6) in the main text. (A) Dominance coefficient $\phi = 0.00$, (B) $\phi = 0.125$, (C) $\phi = 0.25$, (D) $\phi = 0.375$, (E) $\phi = 0.50$, (F) $\phi = 0.625$, (G) $\phi = 0.75$, (H) $\phi = 0.875$, (I) $\phi = 1.00$. For marginal likelihoods of s with respect to p_{init} see Figure 5.6.

5.10.2 Additional results from the matrix iteration approach

Figure 5.19 shows the joint likelihood surface of the selection coefficient (s) and the initial frequency of the focal allele (p_{init}), for under- and overdominance. Figure 5.20 and Table 5.29, and Figure 5.21 and Table 5.30 provide likelihood curves and parameter estimates for migration at rate $m = 0.1$ and $m = 0.2$. These results should be compared to Figure 5.6 and Table 5.3 without migration in the main text.

Figure 5.22 gives the joint likelihood surface of the selection coefficient, s, and the initial frequency of the focal allele, p_{init}, for intermediate dominance. Figure 5.23 and Table 5.31, and

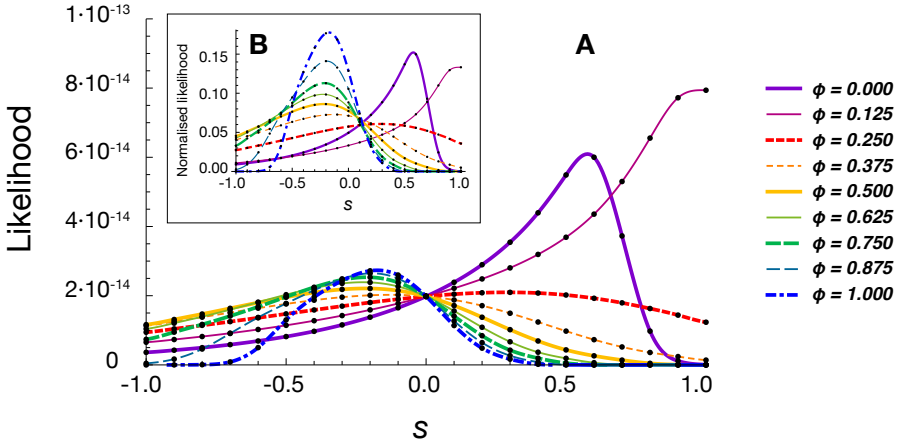

Figure 5.20: Likelihood of the selection coefficient s for various dominance coefficients ϕ with migration rate $m = 0.1$. The likelihood curves are marginal with respect to the initial frequency p_{init} of the 'goat' allele A_1. Fitnesses are parameterized as in equation (5.6) in the main text. A_1 is fully recessive if $\phi = 0$ and fully dominant if $\phi = 1$. For $\phi \notin \{0, 1\}$, there is overdominance if $s > 0$ and underdominance if $s < 0$. (A) The likelihoods are not normalized. Therefore, the areas under the curves indicate the relative support for the various values of ϕ (*cf.* Table 5.29). (B) As in (A) but with likelihoods normalized such that the area under the curve is 1. In a Bayesian view, these curves correspond to the posterior distribution of s given a uniform prior on the normal scale. The curves in (A) and (B) were obtained by third-order interpolation of points computed for values of s on a grid from –1.0 to 1.0 with step size 0.1 (black dots).

Figure 5.24 and Table 5.32 provide likelihood curves and parameter estimates for migration at rate $m = 0.1$ and $m = 0.2$. These results should be compared to those in Figure 5.8 and Table 5.4 for intermediate dominance without migration in the main text.

Table 5.29: Likelihood-based estimates of selection (s) and dominance (ϕ) coefficient with under- or overdominance and migration rate $m = 0.1$.

Dominance scheme	ϕ	$L_\phi{}^a$	B.F.	$\hat{s}_\phi$	HPD
A_1 fully recessive	0.000	4.006	0.672	0.580	(−0.651, 0.824)
Overdom. if $s > 0$, underdom. if $s < 0$	0.125	5.959	1.000	0.967	(−0.655, 1.000)
·	0.250	3.453	0.579	0.290	(−0.837, 1.000)
·	0.375	2.781	0.467	−0.127	(−1.000, 0.627)
·	0.500	2.564	0.430	−0.217	(−1.000, 0.394)
·	0.625	2.423	0.407	−0.226	(−1.000, 0.266)
·	0.750	2.244	0.377	−0.211	(−0.980, 0.214)
·	0.875	1.876	0.315	−0.197	(−0.786, 0.228)
A_1 fully dominant	1.000	1.539	0.258	−0.177	(−0.594, 0.189)

$L_\phi = \sum_{s\in\mathcal{S}} L(\phi, s; D) = \sum_{s\in\mathcal{S}} P(D|\phi, s)$ is an approximation to the marginal likelihood of ϕ, $L(\phi; D) = P(D|\phi) = \int_{\mathcal{S}} P(D|\phi, s)P(s|\phi)\mathrm{d}s = \int_{\mathcal{S}} P(D|\phi, s)P(s)\mathrm{d}s$, where $\mathcal{S}$ is the set of possible values for s, and the last equality holds because ϕ and s are independent. The Bayes Factor (B.F.) is here defined as $L_\phi/\max(L_\phi)$, and therefore denotes the support for any model compared to the one with the maximum marginal likelihood (*i.e.* to $\phi = 0.125$). The maximum-likelihood estimate of s given ϕ is provided by $\hat{s}_\phi$. In a Bayesian perspective, this is equal to the posterior mode, since the prior was uniform on the normal scale. HPD, highest posterior density interval of s. Point and interval estimates correspond to likelihood curves displayed in Figure 5.20.

[a] In multiples of 10^{-13}.

Table 5.30: Likelihood-based estimates of selection (s) and dominance (ϕ) coefficient with under- or overdominance and migration rate $m = 0.2$.

Dominance scheme	ϕ	$L_\phi{}^a$	B.F.	$\hat{s}_\phi$	HPD
A_1 fully recessive	0.000	2.873	0.650	0.576	(−0.754, 0.796)
Overdom. if $s > 0$, underdom. if $s < 0$	0.125	4.421	1.000	0.966	(−0.731, 1.000)
·	0.250	2.483	0.562	0.218	(−0.965, 0.900)
·	0.375	2.047	0.463	−0.203	(−1.000, 0.598)
·	0.500	1.904	0.431	−0.261	(−1.000, 0.377)
·	0.625	1.802	0.408	−0.251	(−1.000, 0.258)
·	0.750	1.659	0.375	−0.227	(−1.000, 0.191)
·	0.875	1.334	0.302	−0.202	(−0.773, 0.228)
A_1 fully dominant	1.000	1.077	0.244	−0.186	(−0.579, 0.190)

$L_\phi = \sum_{s\in\mathcal{S}} L(\phi, s; D) = \sum_{s\in\mathcal{S}} P(D|\phi, s)$ is an approximation to the marginal likelihood of ϕ, $L(\phi; D) = P(D|\phi) = \int_{\mathcal{S}} P(D|\phi, s)P(s|\phi)\mathrm{d}s = \int_{\mathcal{S}} P(D|\phi, s)P(s)\mathrm{d}s$, where $\mathcal{S}$ is the set of possible values for s, and the last equality holds because ϕ and s are independent. The Bayes Factor (B.F.) is here defined as $L_\phi/\max(L_\phi)$, and therefore denotes the support for any model compared to the one with the maximum marginal likelihood (*i.e.* to $\phi = 0.125$). The maximum-likelihood estimate of s given ϕ is provided by $\hat{s}_\phi$. In a Bayesian perspective, this is equal to the posterior mode, since the prior was uniform on the normal scale. HPD, highest posterior density interval of s. Point and interval estimates correspond to likelihood curves displayed in Figure 5.21.
[a] In multiples of 10^{-13}.

Figure 5.25 illustrates the effect of gene flow via migration on the marginal likelihood of s, when there is under- or overdominance. This is similar to Figure 5.11 for intermediate dominance.

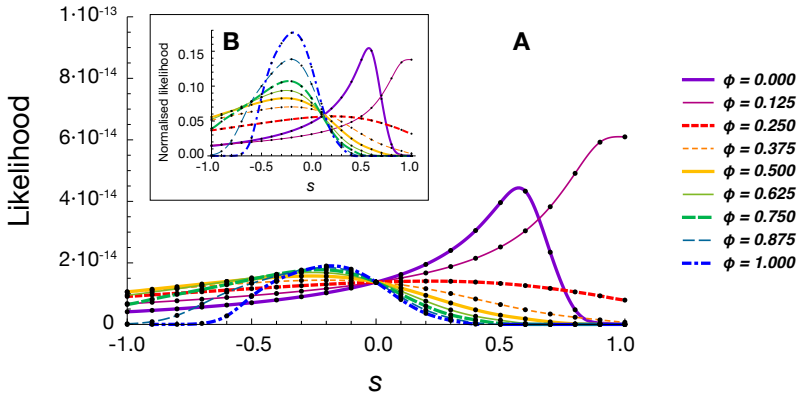

Figure 5.21: Likelihood of the selection coefficient s for various dominance coefficients ϕ with migration rate $m = 0.2$. The likelihood curves are marginal with respect to the initial frequency p_{init} of the 'goat' allele A_1. Fitnesses are parameterized as in equation (5.6) in the main text. A_1 is fully recessive if $\phi = 0$ and fully dominant if $\phi = 1$. For $\phi \notin \{0, 1\}$, there is overdominance if $s > 0$ and underdominance if $s < 0$. (A) The likelihoods are not normalized. Therefore, the areas under the curves indicate the relative support for the various values of ϕ (*cf.* Table 5.30). (B) As in (A) but with likelihoods normalized such that the area under the curve is 1. In a Bayesian view, these curves correspond to the posterior distribution of s given a uniform prior on the normal scale. The curves in (A) and (B) were obtained by third-order interpolation of points computed for values of s on a grid from –1.0 to 1.0 with step size 0.1 (black dots).

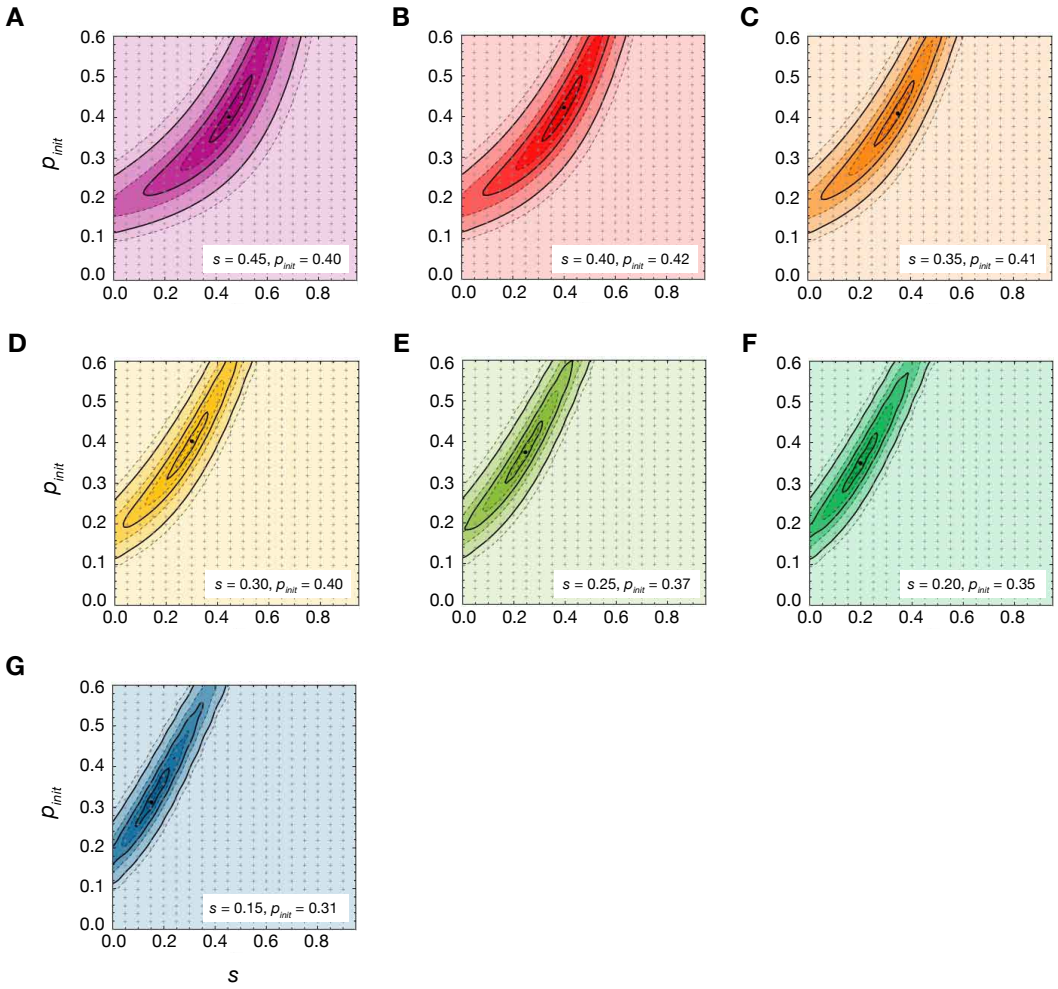

Figure 5.22: Joint likelihood surface of selection coefficient s and initial frequency p_{init} of the 'goat' allele A_1, for intermediate dominance and without migration. Joint maximum-likelihood estimates are given in the boxes. The dotted and solid black lines denote regions of highest posterior density for levels of support of 99%, 95%, 75%, 50%, 25%, 5% and 1%. Crosses denote parameter combinations for which exact values were computed, and the surface was obtained by third-order interpolation. Fitnesses are parameterized as in equation (5.7) in the main text. (A) Dominance coefficient $h = 0.125$, (B) $h = 0.325$, (C) $h = 0.375$, (D) $h = 0.50$, (E) $h = 0.625$, (F) $h = 0.75$, (G) $h = 0.875$, (H) $h = 1.00$. For marginal likelihoods of s with respect to p_{init} see Figure 5.8 in the main text.

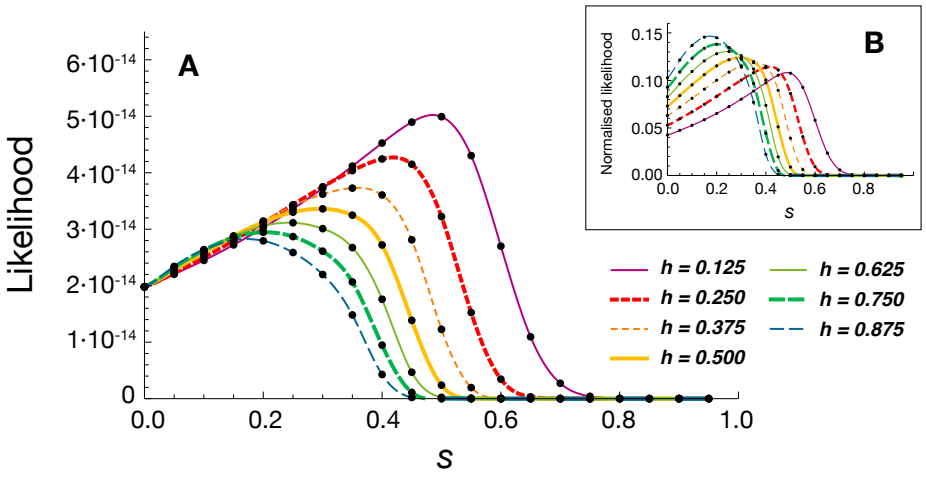

Figure 5.23: Likelihood of the selection coefficient s for various dominance coefficients h migration rate $m = 0.1$. The likelihood curves are marginal with respect to the initial frequency p_{init} of the 'goat' allele A_1. Fitnesses are parameterized as in equation (5.7) in the main text. For $0 < h < 1$ (and $0 \leq s \leq 1$, as is the case here), dominance is intermediate. A_1 is partially recessive if $0 < h < 0.5$ and partially dominant if $0.5 < h < 1$; there is no dominance if $h = 0.5$. The limiting case of full recessivity of A_1 ($h = 0$) is equivalent to the case of $\phi = 0$ in Figure 5.20 and therefore not plotted again. (A) The likelihoods are not normalized and the areas under the curves indicate the relative support for the various values of h (*cf.* Table 5.31). (B) As in (A) but with likelihoods normalized such that the area under the curve is 1. In a Bayesian view, these curves correspond to the posterior distribution of s given a uniform prior on the normal scale. The curves were obtained by third-order interpolation of points computed for values of s on a grid from 0.0 to 0.95 with step size 0.05 (black dots).

Table 5.31: Likelihood-based estimates of selection (s) and dominance (h) coefficient with intermediate dominance and migration rate $m = 0.1$.

Dominance scheme	h	$L_h{}^a$	B.F.	$\hat{s}_h$	HPD
Partial recessivity of the 'goat' allele A_1	0.125	4.641	1.000	0.485	(0.014, 0.620)
.	0.250	3.737	0.805	0.418	(0.007, 0.536)
.	0.375	3.139	0.676	0.362	(0.001, 0.476)
No dominance	0.500	2.712	0.584	0.295	(0.000, 0.435)
Partial dominance of the 'goat' allele A_1	0.625	2.389	0.515	0.244	(0.000, 0.404)
.	0.750	2.134	0.460	0.203	(0.000, 0.379)
.	0.875	1.928	0.415	0.172	(0.000, 0.358)

$L_h = \sum_{s\in\mathcal{S}} L(h, s; D) = \sum_{s\in\mathcal{S}} P(D|h, s)$ is an approximation to the marginal likelihood of h, $L(h; D) = P(D|h) = \int_{\mathcal{S}} P(D|h, s)P(s|h)\mathrm{d}s = \int_{\mathcal{S}} P(D|h, s)P(s)\mathrm{d}s$, where $\mathcal{S}$ is the set of possible values for s, and the last equality holds because h and s are independent. The Bayes Factor (B.F.) is here defined as $L_h/\max(L_h)$, and therefore denotes the support for any model compared to the one with the maximum marginal likelihood (*i.e.* to $h = 0.125$). The maximum-likelihood estimate of s given h is provided by $\hat{s}_h$. In a Bayesian perspective, this is equal to the posterior mode, since the prior was uniform on the normal scale. HPD, highest posterior density interval of s. For full recessivity and full dominance of the 'goat' allele, see Table 5.29. Point and interval estimates correspond to likelihood curves displayed in Figure 5.23.

[a] In multiples of 10^{-13}.

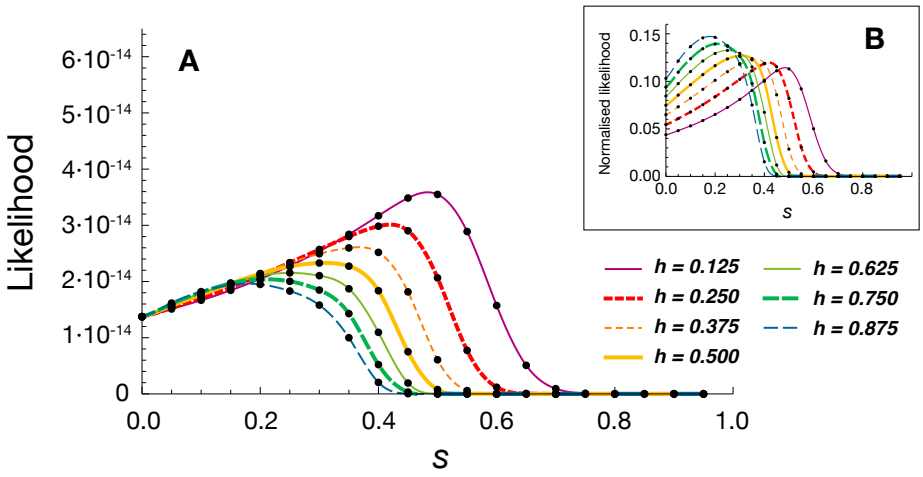

Figure 5.24: Likelihood of the selection coefficient s for various dominance coefficients h migration rate $m = 0.2$. The likelihood curves are marginal with respect to the initial frequency p_{init} of the 'goat' allele A_1. Fitnesses are parameterized as in equation (5.7) in the main text. For $0 < h < 1$ (and $0 \leq s \leq 1$, as is the case here), dominance is intermediate. A_1 is partially recessive if $0 < h < 0.5$ and partially dominant if $0.5 < h < 1$; there is no dominance if $h = 0.5$. The limiting case of full recessivity of A_1 ($h = 0$) is equivalent to the case of $\phi = 0$ in Figure 5.21 and therefore not plotted again. (A) The likelihoods are not normalized and the areas under the curves indicate the relative support for the various values of h (*cf.* Table 5.32). (B) As in (A) but with likelihoods normalized such that the area under the curve is 1. In a Bayesian view, these curves correspond to the posterior distribution of s given a uniform prior on the normal scale. The curves were obtained by third-order interpolation of points computed for values of s on a grid from 0.0 to 0.95 with step size 0.05 (black dots).

Table 5.32: Likelihood-based estimates of selection (s) and dominance (h) coefficient with intermediate dominance and migration rate $m = 0.2$.

Dominance scheme	h	L_h[a]	B.F.	$\hat{s}_h$	HPD
Partial recessivity of the 'goat' allele A_1	0.125	3.138	1.000	0.483	(0.015, 0.606)
·	0.250	2.518	0.802	0.419	(0.009, 0.525)
·	0.375	2.117	0.675	0.367	(0.003, 0.468)
No dominance	0.500	1.835	0.585	0.315	(0.000, 0.425)
Partial dominance of the 'goat' allele A_1	0.625	1.625	0.518	0.257	(0.000, 0.396)
·	0.750	1.462	0.466	0.210	(0.000, 0.371)
·	0.875	1.332	0.425	0.178	(0.000, 0.351)

$L_h = \sum_{s \in \mathcal{S}} L(h, s; D) = \sum_{s \in \mathcal{S}} P(D|h, s)$ is an approximation to the marginal likelihood of h, $L(h; D) = P(D|h) = \int_{\mathcal{S}} P(D|h, s)P(s|h)\mathrm{d}s = \int_{\mathcal{S}} P(D|h, s)P(s)\mathrm{d}s$, where $\mathcal{S}$ is the set of possible values for s, and the last equality holds because h and s are independent. The Bayes Factor (B.F.) is here defined as $L_h/\max(L_h)$, and therefore denotes the support for any model compared to the one with the maximum marginal likelihood (*i.e.* to $h = 0.125$). The maximum-likelihood estimate of s given h is provided by $\hat{s}_h$. In a Bayesian perspective, this is equal to the posterior mode, since the prior was uniform on the normal scale. HPD, highest posterior density interval of s. For full recessivity and full dominance of the 'goat' allele, see Table 5.30. Point and interval estimates correspond to likelihood curves displayed in Figure 5.24.

[a] In multiples of 10^{-13}.

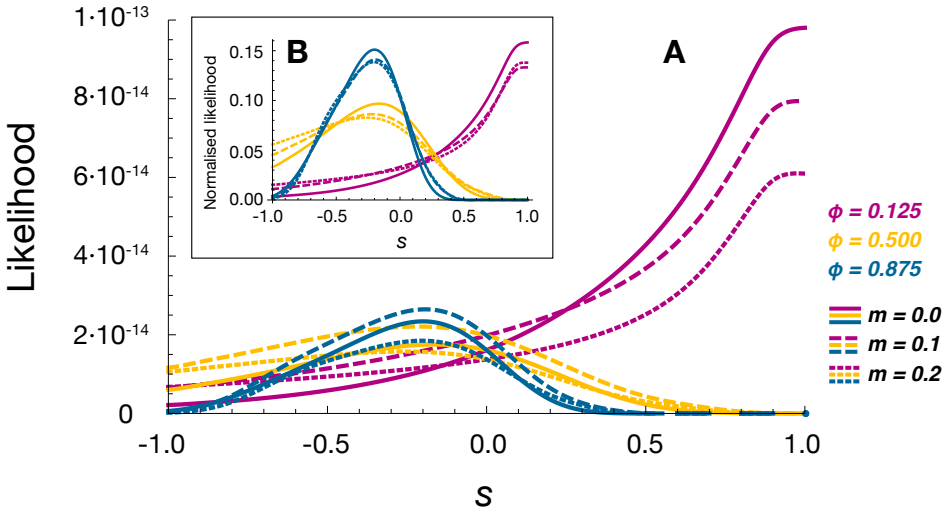

Figure 5.25: The effect of gene flow via migration (at rate m) on the marginal likelihood of the selection coefficient s with under- or overdominance. (A) The likelihoods are not normalized and the areas under the curves indicate the relative support for the different migration rates m, given ϕ. (B) Likelihoods normalized such that the area under the curve is 1. In a Bayesian view, these curves correspond to the posterior distribution of s given a uniform prior on the normal scale. Other details as in Figure 6 in the main text. For intermediate dominance (directional selection), see Figure 5.11.

Bibliography

Aeschbacher, A., 1978. Das Brunftverhalten des Alpensteinbocks. Eugen Rentsch Verlag, Erlenbach-Zürich.

Aeschbacher, S., 2007. Contrasting observed and simulated genetic structure of bottlenecked Alpine ibex populations reveals evidence for gene flow. Master's thesis, Zoological Museum, University of Zurich, Switzerland.

Aeschbacher, S., A. Futschik, and M. A. Beaumont, 2011a. Choice of summary statistics in ABC via boosting and application to the estimation of mutation rates and mating skew in Alpine ibex (*Capra ibex*). In preparation.

———, 2011b. Inferring recent migration rates in a complex model with ABC: joint versus pairwise estimation. In preparation.

Agresti, A., 1990. Categorial Data Analysis. John Wiley & Sons, Inc.

Aguilar, A., G. Roemer, S. Debenham, M. Binns, D. Garcelon, and R. K. Wayne, 2004. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc. Natl. Acad. Sci. U.S.A. 101:3490–3494.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19:716–723.

Arkush, K. D., A. R. Giese, H. L. Mendonca, A. M. McBride, G. D. Marty, and P. W. Hedrick, 2002. Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. Canad. J. Fish. Aquat. Sci. 59:966–975.

Arnold, S. J., 1992. Constraints on phenotypic evolution. Am. Nat. 140:S85–S107.

Bahlo, M. and R. C. Griffiths, 2000. Inference from gene trees in a subdivided population. Theor. Popul. Biol. 57:79–95.

Ballingall, K. T., M. S. Rocchi, D. J. McKeever, and F. Wright, 2010. Trans-species polymorphism and selection in the MHC class II DRA genes of domestic sheep. PLoS One 5.

Balloux, F. and N. Lugon-Moulin, 2002. The estimation of population differentiation with microsatellite markers. Mol. Ecol. 11:155–165.

Barton, N. H., 2000. Genetic hitchhiking. Philosophical Transactions of the Royal Society B: Biological Sciences 355:1553–1562.

———, 2010. What role does natural selection play in speciation? Phil. Trans. R. Soc. B. Biol. Sci. 365:1825–1840.

Barton, N. H., D. E. G. Briggs, J. A. Eisen, D. B. Goldstein, and N. H. Patel, 2007. Evolution. Cold Spring Harbor Laboratory Press, New York.

Barton, N. H. and A. M. Etheridge, 2004. The effect of selection on genealogies. Genetics 166:1115–1131.

Barton, N. H. and G. M. Hewitt, 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113 – 148.

Barton, N. H. and M. Slatkin, 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity 56:409–415.

Bazin, E., K. J. Dawson, and M. A. Beaumont, 2010. Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. Genetics 185:587–602.

Beaumont, M. A., 2010. Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41:379–406.

Beaumont, M. A., J. M. Cornuet, J. M. Marin, and C. P. Robert, 2009. Adaptive approximate Bayesian computation. Biometrika 96:983–990.

Beaumont, M. A. and R. A. Nichols, 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. B. 263:1619–1626.

Beaumont, M. A., R. Nielsen, C. Robert, J. Hey, O. Gaggiotti, L. Knowles, A. Estoup, M. Panchal, J. Corander, M. Hickerson, S. A. Sisson, N. Fagundes, L. Chikhi, P. Beerli, R. Vitalis, J. M. Cornuet, J. Huelsenbeck, M. Foll, Z. H. Yang, F. Rousset, D. Balding, and L. Excoffier, 2010. In defence of model-based inference in phylogeography – Reply. Mol. Ecol. 19:436–446.

Beaumont, M. A. and B. Rannala, 2004. The Bayesian revolution in genetics. Nat. Rev. Genet. 5:251–261.

Beaumont, M. A., W. Zhang, and D. J. Balding, 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Becker, L., C. Nieberg, K. Jahreis, and E. Peters, 2009. MHC class II variation in the endangered European mink *Mustela lutreola* (l. 1761) – consequences for species conservation. Immunogenetics 61:281–288.

Becquet, C. and M. Przeworski, 2007. A new approach to estimate parameters of speciation models with application to apes. Genome Res. 17:1505–1519.

———, 2009. Learning about modes of speciation by computational approaches. Evolution 63:2547–2562.

Beerli, P. and J. Felsenstein, 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152:763–773.

———, 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc Natl Acad Sci USA 98:4563–4568.

Belloy, L., M. Giacometti, P. Boujon, and A. Waldvogel, 2007. Detection of *Dichelobacter nodosus* in wild ungulates (*Capra ibex ibex* and *Ovis aries musimon*) and domestic sheep suffering from foot rot using a two-step polymerase chain reaction. J. Wildl. Dis. .

Bernatchez, L. and C. Landry, 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J. Evol. Biol. 16:363–377.

Bertorelle, G., A. Benazzo, and S. Mona, 2010. Abc as a flexible framework to estimate demography over space and time: some cons, many pros. Mol. Ecol. 19:2609–2625.

Biebach, I. and L. F. Keller, 2009. A strong genetic footprint of the re-introduction history of Alpine ibex (*Capra ibex ibex*). Mol. Ecol. 18:5046–5058.

———, 2010. Inbreeding in reintroduced populations: the effects of early reintroduction history and contemporary processes. Conserv. Genet. 11:527–538.

Biek, R. and L. A. Real, 2010. The landscape genetics of infectious disease emergence and spread. Mol. Ecol. 19:3515–3531.

Bishop, M. D., S. M. Kappes, J. W. Keele, R. T. Stone, S. Sunden, G. A. Hawkins, S. S. Toldo, R. Fries, M. D. Grosz, J. Yoo, and C. W. Beattie, 1994. A genetic linkage map for cattle. Genetics 136:619–639.

Black, F. L. and P. W. Hedrick, 1997. Strong balancing selection at HLA loci: Evidence from segregation in South Amerindian families. Proc. Natl. Acad. Sci. U.S.A. 94:12452–12456.

Blum, M. and O. François, 2010. Non-linear regression models for approximate Bayesian computation. Stat. Comp. 20:63–73.

Blum, M. G. B. and M. Jakobsson, 2011. Deep divergences of human gene trees and models of human origins. Mol. Biol. Evol. 28:889–898.

Blum, M. G. B. and V. C. Tran, 2010. HIV with contact tracing: a case study in approximate Bayesian computation. Biostat. 11:644–660.

Borghans, J. A. M., J. B. Beltman, and R. J. De Boer, 2004. MHC polymorphism under host-pathogen coevolution. Immunogenetics 55:732–739.

Bowen, B. W., A. L. Bass, L. Soares, and R. J. Toonen, 2005. Conservation implications of complex population structure: lessons from the loggerhead turtle (*Caretta caretta*). Mol. Ecol. 14:2389–2402.

Box, G. E. P. and D. R. Cox, 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26:211–252.

Bühlmann, P. and T. Hothorn, 2007. Boosting algorithms: Regularization, prediction and model fitting. Stat. Sci. 22:477–505.

Burnham, K. P. and D. R. Anderson, 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd ed. Springer.

Chapman, J. R., S. Nakagawa, D. W. Coltman, J. Slate, and B. C. Sheldon, 2009. A quantitative review of heterozygosity-fitness correlations in animal populations. Mol. Ecol. 18:2746–2765.

Chapman, M. A. and J. M. Burke, 2006. Letting the gene out of the bottle: the population genetics of genetically modified crops. New Phytol. 170:429–443.

Charbonnel, N. and J. Pemberton, 2005. A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. Heredity 95:377–388.

Charlesworth, B. and D. Charlesworth, 2010. Elements of Evolutionary Genetics. Roberts & Company Publishers.

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.

Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Clutton-Brock, T. H., K. Wilson, and I. R. Stevenson, 1997. Density-dependent selection on horn phenotype in Soay sheep. Phil. Trans. R. Soc. B. Biol. Sci. 352:839–850.

Cockerham, C. C. and B. S. Weir, 1987. Correlations, descent measures: drift with migration and mutation. Proc Natl Acad Sci USA 84:8512–8514.

———, 1993. Estimation of gene flow from F-statistics. Evolution 47:855–863.

Coltman, D. W., P. O'Donoghue, J. T. Jorgenson, J. T. Hogg, C. Strobeck, and M. Festa-Bianchet, 2003. Undesirable evolutionary consequences of trophy hunting. Nature 426:655–658.

Coltman, D. W., J. Pilkington, L. E. B. Kruuk, K. Wilson, and J. M. Pemberton, 2001. Positive genetic correlation between parasite persistance and body size in a free-living ungulate population. Evolution 55:2116–2125.

Coltman, D. W., J. G. Pilkington, J. A. Smith, and J. M. Pemberton, 1999a. Parasite-mediated selection against inbred soay sheep in a free-living, island population. Evolution 53:1259–1267.

Coltman, D. W., J. A. Smith, D. R. Bancroft, J. Pilkington, A. D. C. MacColl, T. H. Clutton-Brock, and J. M. Pemberton, 1999b. Density-dependent variation in lifetime breeding success and natural and sexual selection in Soay rams. Am. Nat. 154:730–746.

Cook, L. M., 2003. The rise and fall of the Carbonaria form of the peppered moth. Quart. Rev. Biol. 78:399–417.

Cook, S. R., A. Gelman, and D. B. Rubin, 2006. Validation of software for Bayesian models using posterior quantiles. J. Comp. Graph. Stat. 15:675–692.

Cornuet, J.-M., F. Santos, M. A. Beaumont, C. P. Robert, J.-M. Marin, D. J. Balding, T. Guillemaud, and A. Estoup, 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24:2713–2719.

Couturier, M. A. J., 1962. Le bouquetin des Alpes – *Capra aegagrus ibex ibex L.* Chez l'auteur, 54, Rue Thiers, Allier.

Crawford, A. M., K. G. Dodds, A. J. Ede, C. A. Pierson, G. W. Montgomery, H. G. Garmonsway, A. E. Beattie, K. Davies, J. F. Maddox, S. W. Kappes, R. T. Stone, T. C. Nguyen, J. M. Penty, E. A. Lord, J. E. Broom, J. Buitkamp, W. Schwaiger, J. T. Epplen, P. Matthew, M. E. Matthews, D. J. Hulme, K. J. Beh, R. A. McGraw, and C. W. Beattie, 1995. An autosomal genetic linkage map of the sheep genome. Genetics 140:703–724.

Croissant, Y., 2008. mlogit: multinomial logit model. R package version 0.1-2.

Csilléry, K., M. Blum, and O. François, 2011. Tools for Approximate Bayesian Computation (ABC). The Comprehensive R Archive Network.

Csilléry, K., M. G. B. Blum, O. E. Gaggiotti, and O. François, 2010. Approximate Bayesian computation (ABC) in practice. Trends Ecol. Evol. 25:410–418.

Currat, M. and L. Excoffier, 2005. The effect of the neolithic expansion on european molecular diversity. Proc. R. Soc. B. 272:679–688.

Currat, M., E. S. Poloni, and A. Sanchez-Mazas, 2010. Human genetic differentiation across the Strait of Gibraltar. BMC Evol. Biol. 10.

Cutrera, A. P. and E. A. Lacey, 2006. Major histocompatibility complex variation in talas tuco-tucos: The influence of demography on selection. J. Mammal. 87:706–716.

———, 2007. Trans-species polymorphism and evidence of selection on class II MHC loci in tuco-tucos (rodentia: Ctenomyidae). Immunogenetics 59:937–948.

Dahm, R., 2008. Discovering dna: Friedrich miescher and the early years of nucleic acid research. Human Genet. 122:565–581. 10.1007/s00439-007-0433-0.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray.

Di Rienzo, A., P. Donnelly, C. Toomajian, B. Sisk, A. Hill, M. L. Petzl-Erler, G. K. Haines, and D. H. Barch, 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics 148:1269–1284.

Diggle, P. J., 1979. On parameter estimation and goodness-of-fit testing for spatial point patterns. Biometrics 35:87–101.

Diggle, P. J. and R. J. Gratton, 1984. Monte carlo methods of inference for implicit statistical models. J. R. Stat. Soc. Ser. B 46:193–227.

Edwards, S. V. and P. W. Hedrick, 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol. Evol. 13:305–311.

Endler, J. A., 1977. Monographs in Population Biology No 10 – Geographic Varioation, Speciation, and Clines. Princeton University Press.

Estoup, A. and B. Angers, 1998. Microsatellites and minisatellites for molecular ecology: Theoretical and empirical considerations. Pp. 55–86, *in* G. R. Carvalho, ed. Advances in Molecular Ecology, vol. 306. IOS Press.

Estoup, A., M. Beaumont, F. Sennedot, C. Moritz, and J.-M. Cornuet, 2004. Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, *Bufo marinus*. Evolution 58:2021–2036.

Estoup, A. and J.-M. Cornuet, 1999. Microsatellite evolution: inference from population data. Pp. 49–65, *in* D. B. Goldstein and C. Schloetterer, eds. Microsatellites – Evolution and Application. Oxford University Press Inc.

Ewens, W. J., 1972. The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3:87–112.

———, 1979. Mathematical Population Genetics. 2 (2004) ed. Springer.

Fagundes, N. J. R., N. Ray, M. Beaumont, S. Neuenschwander, F. M. Salzano, S. L. Bonatto, and L. Excoffier, 2007. Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. USA 104:17614–17619.

Fan, J. and I. Gijbels, 1996. Local polynomial modelling and its applications. Chapman & Hall/CRC.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Let. 27:861–874.

Fearnhead, P. and D. Prangle, 2011. Constructing summary statistics for approximate Bayesian computation: Semi-automatic ABC. In Review .

Felsenstein, J., 1982. How can we infer geography and history from gene frequencies? J. Theor. Biol. 96:9–20.

———, 1988. Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. 22:521–565.

Fisher, R. A., 1922a. On the dominance ratio. Proc. R. Soc. Edin. 42:321–341.

———, 1922b. On the mathematical foundations of theoretical statistics. Phil. Trans. R. Soc. A 222:309–368.

———, 1930. The Genetical Theory of Natural Selection. Oxford University Press.

Forman, G., 2002. A method for discovering the insignificance of one's best classifier and the unlearnability of a classification task. *in* Data Mining Lessons Learned Workshop, the 19th International Conference on Machine Learning (ICML).

Fraser, B. A., I. W. Ramnarine, and B. D. Neff, 2010. Temporal variation at the MHC class IIb in wild populations of the guppy (*Poecilia reticulata*). Evolution 64:2086–2096.

Frazer, K. A., D. G. Ballinger, D. R. Cox, D. A. Hinds, and L. L. e. Stuve, 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861.

Freund, Y., 1995. Boosting a weak learning algorithm by majority. Inform. Comput. 121:256–285.

Freund, Y. and R. E. Schapire, 1996. Experiments with a new boosting algorithm. Machine Learning: Proceedings of the Thirteenth International Conference Pp. 148–156.

———, 1999. A short introduction to boosting. J. Japan. Soc. Artif. Intell. 14:771–780.

Friedman, J., T. Hastie, and R. Tibshirani, 2000. Special invited paper. additive logistic regression: A statistical view of boosting. Ann. Stat. 28:337–374.

Friedman, J. H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29:1189–1232.

Fu, Y. X. and W. H. Li, 1997. Estimating the age of the common ancestor of a sample of DNA sequences. Mol. Biol. Evol. 14:195–199.

Gaggiotti, O. E. and L. Excoffier, 2000. A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances. Proc. R. Soc. B. 267:81–87.

Garrigan, D. and P. W. Hedrick, 2003. Perspective: Detecting adaptive molecular polymorphism: Lessons from the MHC. Evolution 57:1707–1722.

Gavrilets, S., 2003. Perspective: Models of speciation: What have we learned in 40 years? Evolution 57:2197–2215.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004. Bayesian Data Analysis. 2 ed. Chapman & Hall/CRC.

Giacometti, M., R. Roganti, D. De Tann, N. Stahlberger-Saitbekova, and G. Obexer-Ruff, 2004. Alpine ibex *Capra ibex ibex* x domestic goat *C. aegagrus domestica* hybrids in a restricted area of southern Switzerland. Wildlife Biol. 10:137–143.

Gilmour, J. S. L. and J. W. Gregor, 1939. Demes - a suggested new terminology. Nature 144:333–333.

Goda, N., T. Mano, P. Kosintsev, A. Vorobiev, and R. Masuda, 2010. Allelic diversity of the MHC class II DRB genes in brown bears (*Ursus arctos*) and a comparison of DRB sequences within the family Ursidae. Tissue Antigens 76:404–410.

Gratten, J., A. J. Wilson, A. F. McRae, D. Beraldi, P. M. Visscher, J. M. Pemberton, and J. Slate, 2008. A localized negative genetic correlation constrains microevolution of coat color in wild sheep. Science 319:318–320.

Griffiths, R. C. and S. Tavaré, 1994a. Ancestral inference in population genetics. Stat. Sci. 9:307–319.

———, 1994b. Simulating probability distributions in the coalescent. Theor. Popul. Biol. 46:131–159.

Grossen, C., 2005. MHC variability in Alpine ibex (*Capra ibex ibex*) populations. Masters' thesis, Swiss Federal Institute of Technology Zurich (ETH), Switzerland.

Grøtan, V., B.-E. Sæther, F. Filli, and S. Engen, 2008. Effects of climate on population fluctuations of ibex. Glob. Change Biol. 14:218–228.

Grün, B. and F. Leisch, 2007. Fitting finite mixtures of generalized linear regressions in R. Comp. Stat. Data Analysis 51:5247–5252.

———, 2008. FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters. J. Stat. Softw. 28:1–35.

Gutierrez-Espeleta, G. A., P. W. Hedrick, S. T. Kalinowski, D. Garrigan, and W. M. Boyce, 2001. Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? Heredity 86:439–450.

Hadfield, J. D., 2008. Estimating evolutionary parameters when viability selection is operating. Proc. R. Soc. B. 275:723–734.

Haldane, J. B. S., 1924. A mathematical theory of natural and artificial selection – Part II – The influence of partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of mendelian populations, and on natural selection. Proc. Cambr. Phil. Soc. Biol. Sci. 1:158–163.

———, 1932. The Causes of Evolution. 2 (1993) ed. Princeton University Press.

Halliburton, R., 2004. Introduction to Population Genetics. Pearson Prentice Hall.

Hamilton, G., M. Currat, N. Ray, G. Heckel, M. Beaumont, and L. Excoffier, 2005. Bayesian estimation of recent migration rates after a spatial expansion. Genetics 170:409–417.

Hansson, B. and L. Westerberg, 2008. Heterozygosity-fitness correlations within inbreeding classes: local or genome-wide effects? Conserv. Genet. 9:73–83.

Hastie, T., R. Tibshirani, and J. Friedman, 2011. The Elements of Statistical Learning – Data Mining, Inference, and Prediction. 2 (2011) ed. Springer Verlag.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109.

Hedrick, P. W., 1994. Evolutionary genetics of the Major Histocompatibility Complex. Am. Nat. 143:945–964.

Hedrick, P. W., M. E. Ginevan, and E. P. Ewing, 1976. Genetic polymorphism in heterogeneous environments. Annu. Rev. Ecol. Syst. 7:1–32.

Hein, J., M. H. Schierup, and C. Wiuf, 2005. Gene Genealogies, Variation and Evolution – A Primer in Coalescent Theory. Oxford University Press.

Hershey, A. D. and M. Chase, 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. 36:39–56.

Hey, J., 2010. Isolation with migration models for more than two populations. Mol. Biol. Evol. 27:905–920.

Hey, J. and C. A. Machado, 2003. The study of structured populations – New hope for a difficult and divided science. Nat. Rev. Genet. 4:535–543.

Hey, J. and R. Nielsen, 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760.

———, 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. USA 104:2785–2790.

Hill, W. G., 1972. Effective size of populations with overlapping generations. Theor. Popul. Biol. 3:278–289.

———, 1979. A note on effective population size with overlapping generations. Genetics 92:317–322.

Hill, W. G. and A. Robertson, 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294. M3 - 10.1017/S0016672300010156.

Hoelzel, A. R., J. Hey, M. E. Dahlheim, C. Nicholson, V. Burkanov, and N. Black, 2007. Evolution of population structure in a highly social top predator, the killer whale. Mol. Biol. Evol. 24:1407–1415.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6:65–70.

Hothorn, T., P. Buhlmann, T. Kneib, S. M., and H. B., 2011. mboost: Model-based boosting, R package version 2.0-11.

Hubby, J. L. and R. C. Lewontin, 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics 54:577–594.

Hudson, R. R., D. D. Boos, and N. L. Kaplan, 1992. A statistical test for detecting geographic subdivision. Mol. Biol. Evol. 9:138–151.

Hudson, R. R. and N. L. Kaplan, 1988. The coalescent process in models with selection and recombination. Genetics 120:831–840.

Jacobson, A. R., A. Provenzale, A. von Hardenberg, B. Bassano, and M. Festa-Bianchet, 2004. Climate forcing and density dependence in a mountain ungulate population. Ecology 85:1598–1610.

Jeffreys, H., 1961. Theory of Probability. Oxford Classic Texts in the Physical Sciences, Oxford, 3rd ed. Oxford University Press.

Joyce, P. and P. Marjoram, 2008. Approximately sufficient statistics and Bayesian computation. Stat. Appl. Genet. Mol. Biol. 7.

Kaplan, N. L., T. Darden, and R. R. Hudson, 1988. The coalescent process in models with selection. Genetics 120:819–829.

Kazanskaya, E., M. Kuznetsova, and A. Danilkin, 2007. Phylogenetic reconstructions in the genus *Capra* (Bovidae, Artiodactyla) based on the mitochondrial DNA analysis. Russ. J. Genet. 43:181–189.

Keightley, P. D. and A. Eyre-Walker, 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics 177:2251–2261.

Keller, L. F. and D. M. Waller, 2002. Inbreeding effects in wild populations. Trends Ecol. Evol. 17:230–241.

Kemp, S. J., O. Hishida, J. Wambugu, A. Rink, A. J. Teale, M. L. Longeri, R. Z. Ma, Y. Da, H. A. Lewin, and W. Barendse, 1995. A panel of polymorphic bovine, ovine and caprine microsatellite markers. Anim. Genet. 26:299–306.

Kikkawa, E. F., T. T. Tsuda, D. Sumiyama, T. K. Naruse, M. Fukuda, M. Kurita, R. P. Wilson, Y. LeMaho, G. D. Miller, M. Tsuda, K. Murata, J. K. Kulski, and H. Inoko, 2009. Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus *Spheniscus*). Immunogenetics 61:341–352.

Kimura, M., 1984. The neutral theory of molecular evolution, chap. 2, Pp. 15–33. Cambridge University Press.

Kimura, M. and J. F. Crow, 1964. The number of alleles that can be maintained in a finite population. Genetics 49:725–738.

Kimura, M. and T. Ohta, 1969. Average number of generations until fixation of a mutant gene in a finite population. Genetics 61:763–771.

———, 1974. Probability of gene fixation in an expanding finite population. Proc. Natl. Acad. Sci. USA 71:3377–3379.

———, 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc. Natl. Acad. Sci. USA 75:2868–2872.

Kingman, J. F. C., 1982. On the genealogy of large populations. J. Appl. Probab. 19:27–43.

Kirkpatrick, M. and V. Ravigné, 2002. Speciation by natural and sexual selection: Models and experiments. Am. Nat. 159:22–35.

Knapp, L., J. Ha, and G. Sackett, 1996. Parental MHC antigen sharing and pregnancy wastage in captive pigtailed macaques. J. Reprod. Immunol. 32:73–88.

Kruuk, L. E. B., J. Slate, J. M. Pemberton, S. Brotherstone, F. Guinness, and T. Clutton-Brock, 2002. Antler size in red deer: Heritability and selection but no evolution. Evolution 56:1683–1695.

Kuhner, M. K., 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22:768–770.

———, 2009. Coalescent genealogy samplers: windows into population history. Trends Ecol. Evol. 24:86–93.

Kuhner, M. K., J. Yamato, and J. Felsenstein, 1995. Estimating effective population size and mutation rate from sequence data using metropolis-hastings sampling. Genetics 140:1421–1430.

Kullback, S. and R. A. Leibler, 1951. On information and sufficiency. Ann. Math. Stat. 22:79–86.

Latter, B. D. H., 1973. Island model of population differentiation – general solution. Genetics 73:147–157.

Leisch, F., 2004. FlexMix: A general framework for finite mixture models and latent class regression in R. J. Stat. Softw. 11:1–18.

Lenormand, T., 2002. Gene flow and the limits to natural selection. Trends Ecol. Evol. 17:183–189.

Lesnoff, M. and R. Lancelot, 2009. aod: Analysis of Overdispersed Data. R package version 1.1-31.

Leuenberger, C. and D. Wegmann, 2010. Bayesian computation and model selection without likelihoods. Genetics 184:243–252.

Lewontin, R. C. and J. Krakauer, 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175–195.

———, 1975. Testing the heterogeneity of F values. Genetics 80:397–398.

Li, H. P., 2011. A new test for detecting recent positive selection that is free from the confounding impacts of demography. Mol. Ecol. Evol. 28:365–375.

Lin, K., H. Li, C. Schlötterer, and A. Futschik, 2011. Distinguishing positive selection from neutral evolution: Boosting the performance of summary statistics. Genetics 187:229–244.

Loader, C. R., 1996. Local likelihood density estimation. Ann. Stat. 24:1602–1618.

Loison, A., C. Toïgo, J. Appolinaire, and J. Michallet, 2002. Demographic processes in colonizing populations of isard (*Rupicapra pyrenaica*) and ibex (*Capra ibex*). J. Zool. 256:199–205.

Lucas, L., Z. Gompert, J. Ott, and C. Nice, 2009. Geographic and genetic isolation in spring-associated *Eurycea*; salamanders endemic to the Edwards Plateau region of Texas. Conserv. Genet. 10:1309–1319.

Luikart, G., M.-P. Biju-Duval, O. Ertugrul, Y. Zagdsuren, C. Maudet, and P. Taberlet, 1999. Power of 22 microsatellite markers in fluorescent multiplexes for parentage testing in goats (*Capra hircus*). Anim. Genet. 30:431–438.

Lukas, D., B. J. Bradley, A. M. Nsubuga, D. Doran-Sheehy, M. M. Robbins, and L. Vigilant, 2004. Major histocompatibility complex and microsatellite variation in two populations of wild gorillas. Mol. Ecol. 13:3389–3402.

Lynd, A., D. Weetman, S. Barbosa, A. E. Yawson, S. Mitchell, J. Pinto, I. Hastings, and M. J. Donnelly, 2010. Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae s.s.* Mol. Biol. Evol. 27:1117–1125.

MacKay, D., 2003. Information theory, inference, and learning algorithms. Cambridge University Press.

Maddox, J. F., K. P. Davies, A. M. Crawford, D. J. Hulme, D. Vaiman, E. P. Cribiu, B. A. Freking, K. J. Beh, N. E. Cockett, N. Kang, C. D. Riffkin, R. Drinkwater, S. S. Moore, K. G. Dodds, J. M. Lumsden, T. C. van Stijn, S. H. Phua, D. L. Adelson, H. R. Burkin, J. E. Broom, J. Buitkamp, L. Cambridge, W. T. Cushwa, E. Gerard, S. M. Galloway, B. Harrison, R. J. Hawken, S. Hiendleder, H. M. Henry, J. F. Medrano, K. A. Paterson, L. Schibler, R. T. Stone, and B. van Hest, 2001. An enhanced linkage map of the sheep genome comprising more than 1000 loci. Genome Res. 11:1275–1289.

Mahalanobis, P. C., 1936. On the generalized distance in statistics. Proc. Nat. Ins. Sci. India 2:49–55.

Mainguy, J., A. S. Llewellyn, K. Worley, S. D. Côté, and D. W. Coltman, 2005. Characterization of 29 polymorphic artiodactyl microsatellite markers for the mountain goat (*Oreamnos americanus*). Mol. Ecol. Notes 5:809–811.

Malécot, G., 1969. The Mathematics of Heredity. W. H. Freeman and Company, San Francisco.

Mannen, H., Y. Nagata, and S. Tsuji, 2001. Mitochondrial DNA reveal that domestic goat (*Capra hircus*) are genetically affected by two subspecies of bezoar (*Capra aegagurus*). Biochem. Genet. 39:145–154.

Marjoram, P., J. Molitor, V. Plagnol, and S. Tavaré, 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. USA 100:15324–15328.

Marjoram, P. and S. Tavaré, 2006. Modern computational approaches for analysing molecular genetic variation data. Nat. Rev. Genet. 7:759–770.

Maudet, C., C. Miller, B. Bassano, C. Breitenmoser-Wursten, D. Gauthier, G. Obexer-Ruff, J. Michallet, P. Taberlet, and G. Luikart, 2002. Microsatellite DNA and recent statistical methods in wildlife conservation management: applications in Alpine ibex [*Capra ibex (ibex)*]. Mol. Ecol. 11:421–436.

Maynard Smith, J. and J. Haigh, 1974. Hich-hiking effect of a favorable gene. Genet. Res. 23:23–35.

McDonald, J. H. and M. Kreitman, 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.

Merilä, J., B. C. Sheldon, and L. E. B. Kruuk, 2001. Explaining stasis: microevolutionary studies in natural populations. Genetica 112:199–222.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092.

Mevik, B.-H. and R. Wehrens, 2007. The pls package: Principal component and partial least squares regression in R. J. Stat. Softw. 18:1–24.

Meyer-Lucht, Y. and S. Sommer, 2005. MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). Mol. Ecol. 14:2233–2243.

Miller, K. M., K. H. Kaukinen, T. D. Beacham, and R. E. Withler, 2001. Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. Genetica 111:237–257.

Mona, S., B. Crestanello, S. Bankhead-Dronnet, E. Pecchioli, S. Ingrosso, S. D'Amelio, L. Rossi, P. G. Meneguz, and G. Bertorelle, 2008. Disentangling the effects of recombination, selection, and demography on the genetic variation at a major histocompatibility complex class II gene in the alpine chamois. Mol. Ecol. 17:4053–4067.

Morjan, C. L. and L. H. Rieseberg, 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. 13:1341–1356.

Naduvilezhath, L., L. E. Rose, and D. Metzler, 2011. Jaatha: a fast composite-likelihood approach to estimate demographic parameters. Mol. Ecol. 20:2709–2723.

Nagylaki, T., 1989. Gustave Malécot and the transition from classical to modern population genetics. Genetics 122:253–268.

Nagylaki, T. and Y. Lou, 2008. The Dynamics of Migration-Selection Models, vol. 1922, chap. 4, Pp. 117–170. Springer Berlin / Heidelberg.

Nath, H. B. and R. C. Griffiths, 1996. Estimation in an island model using simulation. Theor. Popul. Biol. 50:227–253.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323.

Nei, M. and R. K. Chesser, 1983. Estimation of fixation indexes and gene diversities. Ann. Hum. Genet. 47:253–259.

Neigel, J. E., 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. Annu. Rev. Ecol. Syst. 28:105–128.

Nielsen, R., 2001. Statistical tests of selective neutrality in the age of genomics. Heredity 86:641–647.

———, 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39:197–218.

Nielsen, R., I. Hellmann, M. Hubisz, C. Bustamante, and A. G. Clark, 2007. Recent and ongoing selection in the human genome. Nat. Rev. Genet. 8:857–868.

Nielsen, R. and J. Wakeley, 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. Genetics 158:885–896.

Nievergelt, B., 1966. Der Alpensteinbock (*Capra ibex L.*) in seinem Lebensraum. Ein ökologischer Vergleich. Mammalia depicta. Verlag Paul Parey, Hamburg, Berlin.

Nosil, P., 2008. Speciation with gene flow could be common. Mol. Ecol. 17:2103–2106.

Novembre, J., A. P. Galvani, and M. Slatkin, 2005. The geographic spread of the CCR5 Δ32 HIV-resistance allele. PLoS Biol. 3:e339.

Nunes, M. A. and D. J. Balding, 2010. On optimal selection of summary statistics for approximate Bayesian computation. Stat. Appl. Genet. Mol. Biol. 9.

Nunney, L., 1991. The influence of age structure and fecundity on effective population size. Proc. R. Soc. B. 246:71–76.

———, 1993. The influence of mating system and overlapping generations on effective population-size. Evolution 47:1329–1341.

Ober, C., T. Hyslop, S. Elias, L. Weitkamp, and W. Hauck, 1998. Human leukocyte antigen matching and fetal Loss: results of a 10 year prospective study. Human Reprod. 13:33–38.

O'Brien, S. J. and J. F. Evermann, 1988. Interactive influence of infectious disease and genetic diversity in natural populations. Trends Ecol. Evol. 3:254–259.

Ohta, T. and M. Kimura, 1972. Fixation time of overdominant alleles influenced by random fluctuation of selection intensity. Genet. Res. 20:1–7.

———, 1973. Model of mutation appropriate to estimate number of electrophoretically detectable alleles in a finite population. Genet. Res. 22:201–204.

van Oosterhout, C., 2009. A new theory of MHC evolution: beyond selection on the immune genes. Proc. R. Soc. B. 276:657–665.

Palsboll, P. J., M. Berube, and F. W. Allendorf, 2007. Identification of management units using population genetic data. Trends Ecol. Evol. 22:11–16.

Paterson, S., K. Wilson, and J. M. Pemberton, 1998. Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* l.). Proc. Natl. Acad. Sci. U.S.A. 95:3714–3719.

Pearson, H., 2006. What is a gene? Nature 441:398–401.

Piertney, S. B. and M. K. Oliver, 2006. The evolutionary ecology of the major histocompatibility complex. Heredity 96:7–21.

Pritchard, J., M. Seielstad, A. Perez-Lezaun, and M. Feldman, 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. 16:1791–1798.

Provine, W. B., 1971. The Origins of Theoretical Population Genetics. The University of Chicago Press.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Radwan, J., A. Biedrzycka, and W. Babik, 2010. Does reduced MHC diversity decrease viability of vertebrate populations? Biol. Conserv. 143:537–544.

Radwan, J., A. Tkacz, and A. Kloch, 2008. MHC and preferences for male odour in the bank vole. Ethology 114:827–833.

Raiffa, H. and R. Schlaifer, 1968. Applied Statistical Decision Theory. John Wiley and Sons, Ltd.

Ratmann, O., O. Jorgensen, T. Hinkley, M. Stumpf, S. Richardson, and C. Wiuf, 2007. Using likelihood-free inference to compare evolutionary dynamics of the protein networks of h. pylori and p. falciparum. PLoS Comput. Biol. 3:2266–2278.

Raymond, M. and F. Rousset, 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. J. Hered. 86:248–249.

Reid, J. M., P. Arcese, and L. F. Keller, 2003. Inbreeding depresses immune response in song sparrows (*Melospiza melodia*): direct and inter-generational effects. Proc. R. Soc. B. 270:2151–2157.

Reid, J. M., P. Arcese, L. F. Keller, K. H. Elliott, L. Sampson, and D. Hasselquist, 2007. Inbreeding effects on immune response in free-living song sparrows (*Melospiza melodia*). Proc. R. Soc. B. 274:697–706.

Robert, C. P., J.-M. Cornuet, J.-M. Marin, and N. S. Pillai, 2011. Lack of confidence in approximate bayesian computation model choice. Proc. Natl. Acad. Sci. USA .

Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. W. Feldman, 2002. Genetic structure of human populations. Science 298:2381–2385.

Rousset, F., 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics 142:1357–1362.

———, 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Res. 8:103–106.

Rubin, D. B., 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. Ann. Stat. 12:1151–1172.

Ryser-Degiorgis, M.-P., D. F. Bischof, N. Marreros, C. Willisch, C. Signer, F. Filli, G. Brosi, J. Frey, and E. M. Vilei, 2009. Detection of *Mycoplasma conjunctivae* in the eyes of healthy, free-ranging Alpine ibex: Possible involvement of Alpine ibex as carriers for the main causing agent of infectious keratoconjunctivitis in wild Caprinae. Vet. Microbiol. 134:368 – 374.

Sæther, B.-E., S. Engen, F. Filli, R. Aanes, W. Schröder, and R. Andersen, 2002. Stochastic population dynamics of an introduced Swiss population of the ibex. Ecology 83:3457–3465.

Santucci, F., K. M. Ibrahim, A. Bruzzone, and G. M. Hewit, 2007. Selection on MHC-linked microsatellite loci in sheep populations. Heredity 99:340–348.

Satta, Y., C. Ohuigin, N. Takahata, and J. Klein, 1994. Intensity of natural selection at the Major Histocompatibility Complex loci. Proc. Natl. Acad. Sci. U.S.A. 91:7184–7188.

Schapire, R. E., 1990. The strength of weak learnability. Mach. Learn. 5:197–227.

Schierup, M. H., X. Vekemans, and D. Charlesworth, 2000. The effect of subdivision on variation at multi-allelic loci under balancing selection. Genet. Res. 76:51–62.

Scribner, K. T. and M. Stuwe, 1994. Genetic relationships among Alpine ibex *Capra ibex* populations reestablished from a common ancestral source. Biol. Conserv. 69:137–143.

Servedio, M. R. and M. A. F. Noor, 2003. The role of reinforcement in speciation: Theory and data. Annu. Rev. Ecol. Evol. Syst. 34:339–364.

Sisson, S. A. and Y. Fan, 2010. Likelihood-free Markov chain Monte Carlo. ArXiv:1001.2058v1.

Sisson, S. A., Y. Fan, and M. M. Tanaka, 2007. Sequential Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. USA 104:1760–1765.

———, 2009. Correction for Sisson et al., Sequential Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. USA 106:16889–16889.

Slate, J., P. David, K. G. Dodds, B. A. Veenvliet, B. C. Glass, T. E. Broad, and J. C. McEwan, 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity 93:255–265.

Slatkin, M., 1985. Rare alleles as indicators of gene flow. Evolution 39:53–65.

———, 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.

———, 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.

Slatkin, M. and N. H. Barton, 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43:1349–1368.

Slatkin, M. and W. P. Maddison, 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123:603–613.

Sokal, R. R. and J. F. Rohlf, 1981. Biometry – The Principles and Practice of Statistics in Biological Research. 2 ed. W. H. Freeman and Company, New York.

Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. 2:1742–9994.

Sousa, V. C., M. Fritz, M. A. Beaumont, and L. Chikhi, 2009. Approximate Bayesian computation without summary statistics: The case of admixture. Genetics 181:1507–1519.

Stephens, M. and P. Donnelly, 2000. Inference in molecular population genetics. J. Roy. Stat. Soc. B 62:605–635.

Steward, R. C., 1977. Industrial and non-industrial melansim in peppered moth, *Biston betularia* (L). Ecol. Entomol. 2:231–243.

Strasburg, J. L. and L. H. Rieseberg, 2010. How robust are "isolation with migration" analyses to violations of the IM model? a simulation study. Mol. Biol. Evol. 27:297–310.

Stuwe, M. and C. Grodinsky, 1987. Reproductive biology of captive Alpine ibex (*Capra i. ibex*). Zoo Biol. 6:331–339.

Stuwe, M. and B. Nievergelt, 1991. Recovery of Alpine ibex from near extinction – the result of effective protection, captive breeding, and reintroduction. Appl. Anim. Behav. Sci. 29:379–387.

Stuwe, M. and K. T. Scribner, 1989. Low genetic variablility in reintroduced Alpine ibex (*Capra ibex ibex*) populations. J. Mammal. 70:370–373.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA plymorphism. Genetics 123:585–595.

Takahata, N., 1990. A simple genealogical structure of strongly balanced allelic lines and transspecies evolution of polymorphism. Proc. Natl. Acad. Sci. U.S.A. 87:2419–2423.

———, 1995. A genetic perspective on the origin and history of humans. Annu. Rev. Ecol. Syst. 26:343–372.

Takahata, N. and M. Nei, 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of Major Histocompatibility Complex loci. Genetics 124:967–978.

Takahata, N. and M. Slatkin, 1990. Genealogy of neutral genes in two partially isolated populations. Theor. Popul. Biol. 38:331 – 350.

Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly, 1997. Inferring coalescence times from DNA sequence data. Genetics 145:505–518.

Tellier, A., P. Pfaffelhuber, B. Haubold, L. Naduvilezhath, L. E. Rose, T. Städler, W. Stephan, and D. Metzler, 2011. Estimating parameters of speciation models based on refined summaries of the joint site-frequency spectrum. PLoS ONE 6:e18155–.

Thoß, M., P. Ilmonen, K. Musolf, and D. J. Penn, 2011. Major histocompatibility complex heterozygosity enhances reproductive success. Mol. Ecol. 20:1546–1557.

Thursz, M. R., H. C. Thomas, B. M. Greenwood, and A. V. S. Hill, 1997. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. Nat. Genet. 17:11–12.

Toïgo, C., J. M. Gaillard, M. Festa-Bianchet, E. Largo, J. Michallet, and D. Maillard, 2007. Sex- and age-specific survival of the highly dimorphic Alpine ibex: evidence for a conservative life-history tactic. J. Anim. Ecol. 76:679–686.

Toïgo, C., J. M. Gaillard, D. Gauthier, I. Girard, J. P. Martinot, and J. Michallet, 2002. Female reproductive success and costs in an alpine capital breeder under contrasting environments. Ecoscience 9:427–433.

Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interf. 6:187–202.

Tschopp, R., J. Frey, L. Zimmermann, and M. Giacometti, 2005. Outbreaks of infectious keratoconjunctivitis in alpine chamois and ibex in Switzerland between 2001 and 2003. Vet. Record 157:13–18.

Tufto, J., S. Engen, and K. Hindar, 1996. Inferring patterns of migration from gene frequencies under equilibrium conditions. Genetics 144:1911–1921.

Vaiman, D., L. Schibler, F. Bourgeois, A. Oustry, Y. Amigues, and E. P. Cribiu, 1996. A genetic linkage map of the male goat genome. Genetics 144:279–305.

Verdu, P., F. Austerlitz, A. Estoup, R. Vitalis, M. Georges, S. ThÈry, A. Froment, S. Le Bomin, A. Gessain, J.-M. Hombert, L. Van der Veen, L. Quintana-Murci, S. Bahuchet, and E. Heyer, 2009. Origins and genetic diversity of pygmy hunter-gatherers from western central africa. Current Biology 19:312–318.

Vitalis, R., K. Dawson, and P. Boursot, 2001. Interpretation of variation across marker loci as evidence of selection. Genetics 158:1811–1823.

Wahlund, S., 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betracthet. Hereditas 11:65–106.

Wakeley, J., 1996a. Distinguishing migration from isolation using the variance of pairwise differences. Theor. Popul. Biol. 49:369 – 386.

———, 1996b. Pairwise differences under a general model of population subdivision. J. Genet. 75:81–89.

———, 2004. Recent trends in population genetics: More data! More math! Simple models? J. Hered. 95:397–405.

———, 2009. Coalescent theory – an introduction. 1 ed. Roberts & Company Publishers.

Wakeley, J. and J. Hey, 1997. Estimating ancestral population parameters. Genetics 145:847–855.

Waples, R. S. and O. Gaggiotti, 2006. 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.

Watson, J. D. and F. H. C. Crick, 1953. Molecular structure of nucleic acids: A structure for Deoxyribose Nucleic Acid. Nature 171:737–738.

Watterson, G. A., 1978. Homozygosity test of neutrality. Genetics 88:405–417.

Webster, L. M., P. C. Johnson, A. Adam, B. K. Mable, and L. F. Keller, 2008. Absence of three known benzimidazole resistance mutations in *Trichostrongylus tenuis*, a nematode parasite of avian hosts. Veterinary Parasitology 158:302 – 310.

Wegmann, D. and L. Excoffier, 2010. Bayesian inference of the demographic history of chimpanzees. Mol. Biol. Evol. 27:1425–1435.

Wegmann, D., C. Leuenberger, and L. Excoffier, 2009a. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics 182:1207–1218.

———, 2009b. Using ABCtoolbox. University of Bern, Computational and Molecular Population Genetics Laboratory, 3012 Bern, Switzerland. `http://www.cmpg.iee.unibe.ch/content/softwares__services/computer_programs/abctoolbox/index_eng.html`.

Wegmann, D., C. Leuenberger, S. Neuenschwander, and L. Excoffier, 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. BMC Bioinformatics 11.

Wei, K., Z. Zhang, X. Wang, W. Zhang, X. Xu, F. Shen, and B. Yue, 2010. Lineage pattern, trans-species polymorphism, and selection pressure among the major lineages of feline Mhc-DRB peptide-binding region. Immunogenetics 62:307–317.

Weir, B. S. and C. C. Cockerham, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Weir, B. S. and W. G. Hill, 2002. Estimating F-Statistics. Annu. Rev. Genet. 36:721–750.

Weiss, G. and A. von Haeseler, 1998. Inference of population history using a likelihood approach. Genetics 149:1539–1546.

Whitlock, M. C. and D. E. McCauley, 1999. Indirect measures of gene flow and migration: F_{ST} not equal $1/(4Nm + 1)$. Heredity 82:117–125.

Wilkinson, R. D., 2008. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. ArXiv:0811.3355v1.

Williamson, S. H., R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen, and C. D. Bustamante, 2005. Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proc. Natl. Acad. Sci. USA 102:7882–7887.

Willisch, C., I. Biebach, U. Koller, T. Bucher, N. Marreros, M.-P. Ryser-Degiorgis, L. Keller, and P. Neuhaus, 2011. Male reproductive pattern in a polygynous ungulate with a slow life-history: the role of age, social status and alternative mating tactics. Evol. Ecol. Online First (May 2011).

Willisch, C. S., 2009. The ecology of reproduction in long-lived male Alpine ibex (*Capra ibex*): the role of age, dominance and alternative mating tactics. Ph.D. thesis, Université de Neuchâtel, Switzerland.

Willisch, C. S. and P. Neuhaus, 2009. Alternative mating tactics and their impact on survival in adult male Alpine ibex (*Capra ibex ibex*). J. Mammal. 90:1421–1430.

———, 2010. Social dominance and conflict reduction in rutting male Alpine ibex, *Capra ibex*. Behav. Ecol. 21:372–380.

Wilson, A. J., D. W. Coltman, J. M. Pemberton, A. D. J. Overall, K. A. Byrne, and L. E. B. Kruuk, 2005a. Maternal genetic effects set the potential for evolution in a free-living vertebrate population. J. Evol. Biol. 18:405–414.

Wilson, A. J., J. M. Pemberton, J. G. Pilkington, T. H. Clutton-Brock, D. W. Coltman, and L. E. B. Kruuk, 2007. Quantitative genetics of growth and cryptic evolution of body size in an island population. Evol. Ecol. 21:337–356.

Wilson, A. J., J. G. Pilkington, J. M. Pemberton, D. W. Coltman, A. D. J. Overall, K. A. Byrne, and L. E. B. Kruuk, 2005b. Selection on mothers and offspring: Whose phenotype is it and does it matter? Evolution 59:451–463.

Worley, K., J. Carey, A. Veitch, and D. W. Coltman, 2006. Detecting the signature of selection on immune genes in highly structured populations of wild sheep (*Ovis dalli*). Mol. Ecol. 15:623–637.

Wright, S., 1922. Coefficients of inbreeding and relationship. Am. Nat. 56:330–338.

———, 1931. Evolution in mendelian populations. Genetics 16:97–159.

———, 1937. The distribution of gene frequencies in populations. Proc. Natl. Acad. Sci. USA 23:307–320.

———, 1943. Isolation by distance. Genetics 28:114–138.

———, 1945. The differential equation of the distribution of gene frequencies. Proc. Natl. Acad. Sci. USA 31:382–389.

———, 1951. The genetical structure of populations. Ann. Eugenics 15:323–354.

Yule, G. U., 1902. Mendel's laws and their probable relations to intra-racial heredity. New Phytologist 1:193–207.

Zeng, K. and B. Charlesworth, 2009. Estimating selection intensity on synonymous codon usage in a nonequilibrium population. Genetics 183:651–662.