Ecological and population genomics of the wild yellow-necked mouse, *Apodemus flavicollis*

Rohan Raval

Supervisor: Dr Jarek Bryk
Department of Biological and Geographical Sciences
School of Applied Sciences
University of Huddersfield

This thesis is submitted to the University of Huddersfield in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

September 2021

For Baa, Bapu and Bobby, whom I miss dearly …

Copyright statement

When in doubt, make a library . . .
Dr Jarek Bryk, Budding Ecologist

Declaration

I, Rohan Raval, declare that this thesis titled "Ecological and population genomics of the wild yellow-necked mouse, *Apodemus flavicollis*" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the University of Huddersfield.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at the University of Huddersfield or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Rohan Raval
September 2021

Acknowledgements

Many PhD students, myself included, often find the experience of a doctoral degree extremely testing both professionally, and personally. Undertaking a PhD is not intended to be easy, and it is, in part, because of the network of support that postgraduates build with the people around them that their efforts bear fruit. These are the people whom I would like to show my appreciation for. You have all made my experience truly memorable.

Firstly, I would like to thank my supervisors, Dr Jarek Bryk and Dr Martin Carr. Jarek, your support, kindness and willingness to provide a free rein during the PhD has been greatly appreciated, and it was a huge motivator for me. Please keep providing your "reality checks"; they were invaluable! I would not have been able to accomplish this without your guiding hand, and yes, bioinformatics rocks! I thank you unreservedly for all you have done, and the faith you have shown in me. Martin, it has been an absolute pleasure to work with you. The spirited debates over a beer were particularly brilliant, and often insightful! I really appreciate the valuable advice you have given me over the last few years. P.s. *Teleopsis thaii* is pretty cool!

Secondly, I would like to thank our collaborators at the Mammalian Research Institute in Białowieża. In particular, to Dr Janek Boratyński and colleagues whom have trapped, sampled and phenotyped the mice used in this thesis. I would also like to extend my gratitude to Dr Marek Kucka and Dr Yingguang Frank Chan. I gained considerable amounts of knowledge from you, and I thank you for your patience and use of your facilities at the Friedrich Miescher Laboratory of the Max Planck Institute for Developmental Biology. Molecular work isn't so bad after all, and Marek, I hope you enjoyed the Fritz-Cola ;) .

To all the fellow PhD students and friends I have made in the Bioinformatics Suite, quirks and all, I loved every moment of my time with you. I will miss adding to the "Great Wall of Comments" and going for that seemingly "quiet" Friday night drink in the pub. Marisa, it was a real pleasure to work with you. Jenni, cheers to your spectacular banana sauce with a side of karaoke! The world is a much better place with dedicated and caring people like you

in it. You will always have your personal cheering section when I'm around.

To Pierre, Manoushka and my goddaughter Albane, my time in Huddersfield was largely defined by you. You are some of the closest to my heart and are truly my family. Pierre, that first whisky in Buttermere was where it all began; we will undoubtedly share many more! And don't forget, we still have that 1963 Strathisla to tick off the list (though I'd much prefer the 1967, no pressure!). Manoushka, you can always depend on me to come knocking on the door for some more of your famous tartiflette. It helped keep me warm through the long, dark Yorkshire winters! You have both been there for me through the best and worst times, to laugh, love and care. I will never forget!

To Dr Shorouk Mombrikotb, look where my time with you has led me! Ice cold hands sampling soil, a few field work fails, burning the midnight oil in the lab, me not listening to you, you not listening to me, a few bruises, twisted ankles (which still hurts on a cold day!), tears, and many many many post-lab breakfasts and dinners later . . . but you know what? . . . WE BOTH MADE IT! Please don't make Saleka go through it all too! Though I must say, if we can go through all of that and still make it out alive, then anything is possible! I am truly grateful for everything.

Lastly, to my parents Ashok and Nayana, I am where I am because of you. You have shown me what I can achieve when I persevere, and taught me to always strive for excellence. You have supported me through university, and continue to support me still. I have had opportunities you could only dream of when growing up, and I will be forever grateful to you for your hard work and sacrifice. I love you both dearly.

Abstract

Species belonging to the genus *Apodemus* are one of the most common and broadly distributed small mammals in the Palearctic, making them ideal for studying ecological and evolutionary processes in natural systems. Although much research has focused on the ecology of *Apodemus sp.*, relatively little is known about the evolutionary processes which govern their natural variation in the wild. This is, in part, due to the limitations of *de novo* genomic research in traditionally non-model organisms. However, recent advances in high-throughput library preparation techniques such as RADseq, have made non-model organisms more accessible for genome wide analyses. This thesis shows how a modified RADseq protocol (quaddRAD) can be used to conduct ecological and evolutionary genomics research in a large population of wild yellow-necked mice, *Apodemus flavicollis*, that is subject to highly seasonal conditions.

I show how a high quality genomic dataset of 21,011 SNPs can be generated using quaddRAD, and discuss the rationale behind the methodology in detail. I then use the genotypes to construct a multi-generational pedigree, and describe the population's demography, fitness and allele frequency dynamics over time. I find significant variation in the genetic contributions of mice to each generation, where by the end of the study, 53.6% of the sampled population was descended from a single individual. Contrary to the expected high levels of inbreeding and low genetic diversity in such a population, I find it is largely panmictic, which suggests a large degree of connectivity to nearby populations allowing genetic rescue through high levels of migration and gene flow.

Finally, I show how heterothermic responses, which reduce an individuals energy budget by up to 65% under harsh conditions, are not only highly variable in the population, but also highly heritable. This suggests heterothermy could be subject to natural selection. I show that heterothermic responses form two distinct thermal strategies in the morphospace, which may be the result of different modes of selection acting on the population to maintain significant natural variation. This thesis shows the viability of quaddRAD for large scale genomics research in wild *Apodemus sp.*, to cement their role as a model organism for ecological and evolutionary genomics research.

Table of contents

List of figures

List of tables

Chapter 1

Apodemus, the under-utilised ecological and evolutionary model

Model systems are not a new concept in ecological, evolutionary and (population) genetic research as they allow researchers to understand the patterns and processes which govern how natural variation is maintained. Their use has been the cornerstone of biological research for centuries, and a wealth of knowledge has been gained through scientific observation and experimentation with these systems. When choosing a model to study ecology and evolutionary biology, researchers must then ask: What makes a good model system?

Historically, choosing an eco-evolutionary model system was determined by its simplicity to observe and manipulate. For example, short generation times and ease of husbandry in genera such as *Mus*, *Caenorhabditis* and *Drosophila*, allow real time observation of evolutionary processes in artificial laboratory environments that forgo the need to wait tens to thousands of years [48, 169, 202]. Similarly, microbial ecologists regularly use microcosm experiments to manipulate and evolve entire ecosystems within a 96-well plate in a concerted and replicable way [181, 185], thus expanding the model to incorporate multi-species dynamics, and better represent natural systems. Laboratory experiments like these, when used appropriately, are undoubtedly an incredibly powerful approach to ecological and evolutionary research. However, their very nature highly simplifies, and tightly controls the

processes they were designed to observe, de-contextualising them from the *natural* systems they are intended to model [163].

Unfortunately, very little is known about the natural history of many of the commonly used model species as they were chosen for specific traits which facilitate adaptation to a laboratory. Thus, they are further differentiated from their wild counterparts [6]. However, exemplified in a recent series of articles published in the journal eLife called The Natural History of Model Organisms, the idea of a model system has begun to shift outside of the laboratory to also encompass more natural/wild organisms and systems. This thesis aims to add to an existing collection of wild models by studying the free-living yellow-necked mouse, *Apodemus flavicollis*, from Białowieża Forest, Poland. I will discuss how such a new model system will supplement our growing body of knowledge on ecological and evolutionary population genomics by providing an excellent new resource for future research.

1.1 Wild ecological and evolutionary model systems

Wild study systems are popular among naturalists and evolutionary biologists to study natural variation, and many exist to better understand how populations adapt, differ and diverge under changing environmental conditions and habitats. Mammalian island populations in particular have received considerable long-term attention, with some examples including, but not limited to: red deer on the Isle of Rum [309], Soay sheep on St Kilda [21], water voles on Coiresa [250] and wild house mice on Skockholm [27], which have been used for quantitative ecological and evolutionary research since MacArthur and Wilson first developed their seminal theory of Island Biogeography [201]. The beauty of these study systems is that they are free running, natural, replicable and isolated *microcosms* that are allowed to evolve and adapt to the changing environment. Furthermore, the distribution and range of an island study species is often defined by the geographical structures and boundaries of the island, allowing reliable estimates of (effective) population sizes. Isolation also means such models systems often exist in allopatry, thus serving as an excellent wild observatory for natural

selection [42], genetic drift and founder effects [250], as well as population divergence and speciation [341]. Other studies include: heritability [257, 277], fitness and fecundity [69, 247], inbreeding [143], parasitology [5], behavioural ecology [278], life history strategy [165], and evolutionary trade-offs [153] to name a few. This is of course an in-exhaustive list, and the range of publications about these model systems numbers in the thousands. It is clear they provide an invaluable resource.

The isolation and equilibrium dynamics of island model systems however, despite the many advantages listed above, have one notable caveat: studies of gene-flow and migration in vertebrates are generally limited to networks of islands which are close enough to allow small amounts of migration, in species which are also easily able to move between them such as the wild house sparrows (*Passer domesticus*) in Helgeland, Norway (but see also Ali and Vences [7] which discuss how long distance dispersal in mammals could occur under rare circumstances). These bird model systems form a dynamic meta-population of source and sink populations which have become popular to study the effects of gene flow and dispersal [161, 292], phenotypic divergence [64], effective population size [17] and conservation biology [193, 276]. Physiological studies in the wild such as on basal metabolic rate and heterothermy however, require mark-recapture experiments and transportation to laboratories to measure the phenotypes, which in bird species require considerable trapping effort.

Rodent models are an alternative system which have long been studied due to their adaptability to novel environments, ease of capture, broad distribution, diversity, and ecological, economical, epidemiological and medigo-genetic importance. For example, many species are commensal with humans and thrive in agricultural landscapes, causing significant damage and financial loss [307]. Many species also form reservoirs for infectious diseases [233, 308], making their proximity to people of particular concern. Studies therefore often focus on the impact of these rodents on humans, even though only 10-15% of species are considered pests [316]. The remaining species have important ecological roles in natural systems [33, 269], and understanding their ecology and natural history has far-reaching implications. Rodents

have therefore formed one of the most important model systems for researchers in many fields of research.

Arguably, the most widely studied rodents belong to the genus *Mus* due to their exceptionally broad distribution. Sub-species of the house mouse, *Mus musculus* for example, form a fascinating model system for the genetic effects of hybrid zones and introgression [147]. *M. musculus domesticus* also forms highly population-specific immunological states, suggesting immune-dependent selection mechanisms could act to determine how the populations respond to infection [2]. However, the species is, rather uniquely among the genus, highly commensal with humans, and very few feral populations exist to study the system in its natural state.

The genus *Apodemus*, which also has a broad range across much of the Palearctic [228], has gained recent attention in ecological and evolutionary research due to an association with natural, and relatively undisturbed old-growth forest habitats [320]. The presence of unique genetic characteristics such as B-chromosomes [356], and their heterothremic life history strategy [31], further makes this an appealing genus to study. Thus, *Apodemus* is now developing into an additional model system to understand what maintains natural variation in the wild.

1.2 Why study *Apodemus*?

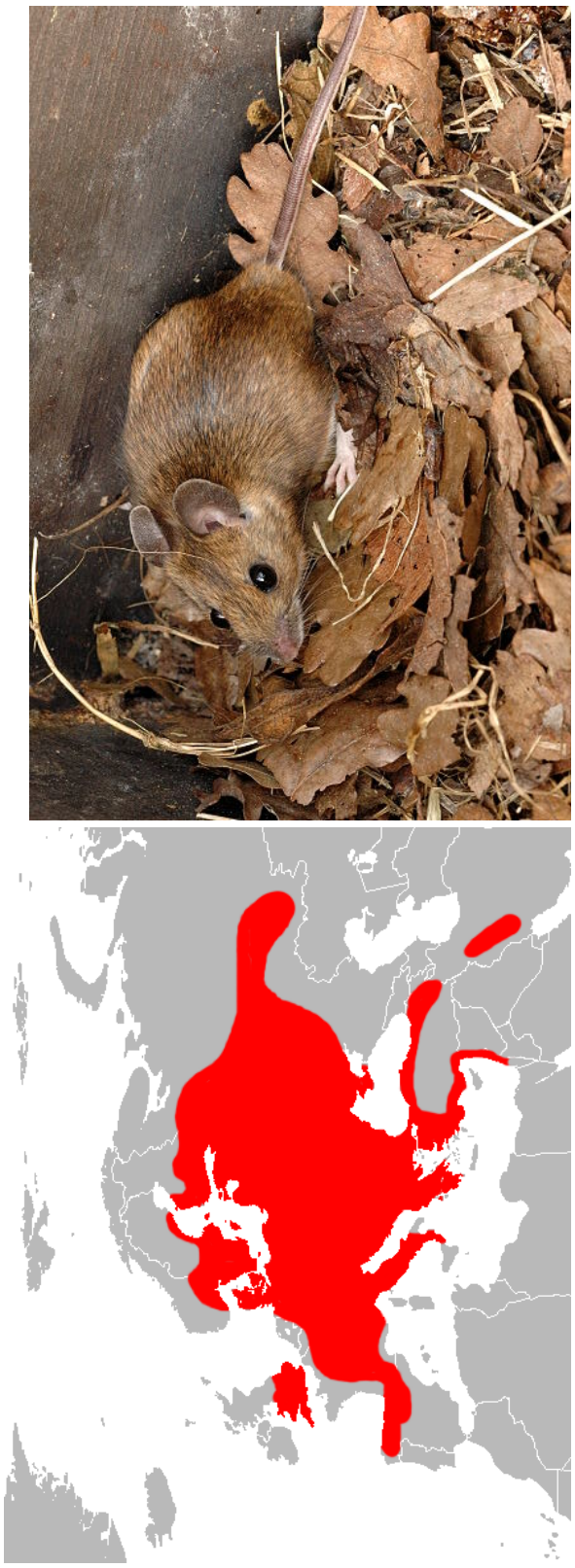
Present across most of the Palearctic [230, 229, 317, 322], *Apodemus* is composed of at least 21 species of mice with a range from Western Europe, through North Africa and the Mediterranean, and as far east as the Japanese Archipelago [53, 219, 229]. Their broad distribution, good fossil record, non-commensal nature and association with old growth forest habitats makes them ideal for ecological and evolutionary studies. Furthermore, some sister species within *Apodemus* are sympatric, and often syntopic, with a large overlap in niches [229, 296]. Despite a large range overlap, similar morphological characteristics and close relatedness, they have not been documented to hybridise in the wild and have distinct genetic characteristics [25, 53, 151, 229]. This is surprising, given viable chimeras have been

created in a laboratory setting with the more distantly related and well studied house mouse, *Mus musculus* (Linnaeus, 1758) [347]. These characteristics therefore make *Apodemus* an appealing genus to research the processes that maintain natural variation in the wild, some of which will be described in this thesis for yellow-necked mice, *Apodemus flavicollis* (Melchior, 1834), from Białowieża Forest in Poland.

1.2.1 *A.flavicollis* in the wild

A. flavicollis is a small (9-12cm), arboreal, nocturnal and forest edge dwelling species common in the Western Palearctic (figure 1.1). They can weigh between 14-45g, though males are typically larger and heavier than females. Females can produce litters of 2-11 pups up to three times per year, and few mice survive more than one year. Although *A. flavicollis* is primarily a forest-dwelling species, it has been known to overwinter in houses, though this is less common [208]. This interaction with humans is of particular epidemiological significance as *Apodemus* is often parasitised by *Borellia* carrying ticks which can cause Lyme disease. *A. flavicollis* is also a known reservoir of tick-borne encephalitis and a number of hantaviruses which can cause haemorrhagic fever [176, 264, 304].

As a largely granivorous species, *A. flavicollis* is ecologically important due to its role in seed dispersal by its hoarding behaviours. The behaviour is known to influence forest composition due to a preference for energy- and nitrogen-rich seeds such as those from beech, oak and hazel, which are cached as far as 13m from the parent tree [150]. In fact, this behavior can account for as much as 100% of seed predation within deciduous forests during masting [74, 149]. It is also strongly linked to an individuals genetics and physiology through differences in boldness, aggression and territoriality, which in turn affects home range sizes (100-2300m^2) and therefore seed distribution patterns [119, 120, 338]. Reliance on seeds as the primary source of food, and seasonal variation in their availability, therefore causes distinct annual cycles of abundance and population density in *A. flavicollis*, with a peak in autumn (September-November) and trough in spring (March-April), with significant variation among years [148, 269].



(a) (b)

Fig. 1.1 a) The extent of *A. flavicollis*' range in the western Palearctic (red). b) *A. flavicollis* in its natural woodland habitat. Images obtained from Wikimedia (available under a public domain) and Wikipedia (available under CC BY-SA 3.0).

To persist in harsh winters and commence breeding under more favourable conditions, *A. flavicollis* has adapted in a number of ways to ensure the greatest individual fitness. A generalist strategy in terms of food choice is one example of this. Limited seed availability during harsh winters leads to increased predation on secondary sources of food such as insects, bird eggs and bats [74]. However, the availability of food during winter is highly stochastic, and the high energy budget of *A. flavicollis* necessitates additional strategies to increase the probability of survival. Like many mammals and birds, *A. flavicollis* has physiologically adapted to difficult environmental conditions through the ability to enter torpor, a controlled hypometabolic and hypothermic state, to lower an individuals energy budget by up to 65% [31, 290]. This allows mice to conserve energy during the winter, so resources can be dedicated to reproduction when food is once again plentiful. Though the expression of torpor is not restricted by seasonality, and can occur at any time when conditions require [290].

We know much about the role of phenotype-environment interactions for the control of torpor. For example, entry into torpor can be entrained by the rhythmicity of an individuals circadian cycle which is influenced by natural light levels, and is therefore affected by seasonality in temperate latitudes across much of *A. flavicollis'* range [170]. We also know that individuals display varying degrees of torpor in response to fasting, which is more likely under harsh winter conditions [31, 32]. While many studies have also researched the molecular processes involved in torpor initiation [45, 227], little is known about the genetic variation responsible for this variation in torpor depth, characterised by individual body temperature when torpid [31, 32, 155], i.e. Is the ability to enter deeper torpor influenced by your genotype, or is it merely a plastic response to environmental conditions? Furthermore, we still do not know how this translates to a population-level response in terms of the evolution of the phenotype, i.e. Are torpor phenotypes heritable and subject to natural selection? These are just two of the questions that will be addressed in this thesis to demonstrate the value of *Apodemus* as an ecological and evolutionary model.

1.2.2 Evolutionary and population genetics in *A. flavicollis*

Much of the genetic research in *Apodemus* till now has focused on highly targeted amplicon sequencing technologies, which amplify variable regions such as the mitochondrial gene *Cytb* [228], 12S rRNA [228, 230, 322], and nuclear genes such as *RAG1* [322]. These variable regions, with mutation rates typically two orders of magnitude greater than the nuclear mutation rate [238], have been very useful in phylogenetic and phylogeographic reconstruction [228–230, 322]. For example, Michaux et al. [230] used sequence variability in the mitochondiral *Cytb* gene to show clear phylogeographic structuring in *A. flavicollis* populations, and this was most likely the result of population isolation and persistence in two refugia during the last glacial maximum. The limitation of these approaches, however, is that they lack the necessary resolution to study selection and gain a better understanding of the processes governing how populations differ and diverge in response. To do this requires a genome-wide approach.

1.2.2.1 Genome-wide approaches for population genetics in *A. flavicollis*

Although the cost of whole genome sequencing significantly reduced after the year 2000 [311], it still remains out of reach for the majority of researchers for population genetic analyses. This is particularly true for eukaryotic species with genome sizes in the order of 10^9 bp, as with *Apodemus*. Alternative approaches, instead utilise polymorphic markers interspersed throughout the genome. These are heritable polymorphic loci that are present in one or more populations, with one example being microsatellites, short sequences of tandemly repeated DNA that are variable in length and sequence, and typically 5-50 base pairs long [80, 333]. In a number of organisms including *Apodemus*, microsatellites have been used in pedigree analysis [119], studies of population structure [204], genetic diversity [172], estimation of mutation rate [203], and the detection of selective sweeps to name a few [103, 280, 297]. These studies typically only require tens of loci, and although it is laborious to develop, microsatellites are relatively inexpensive when such few loci are needed [103]. However, uncovering the genomic basis for a phenotype with a genome wide

association (GWA) for example, requires significantly higher resolution involving typically tens of thousands of independent markers [18, 19, 137, 325].

Unfortunately, significant *a priori* knowledge of the genome to design primers and amplify these variable regions is a prerequisite for analyses using microsatellites, which is why they are labour and time intensive to develop. Thus, they can be costly when more loci are needed, and require significant resources to develop in non-model organisms such as *A. flavicollis*, as little is known about their genome [80, 103]. Although alternatives to microsatellites such as restriction fragment length polymorphism (RFLP), which requires no *a priori* knowledge of the genome, can be used by identifying single nucleotide polymorphisms (SNPs) and form haplotypes for studies of selection [258], they still lack the resolution necessary for GWA, and still require significant development time and cost [80]. Recent evidence also suggests the location of markers such as microsatellites is non-random, and could bias GWA towards certain genomic regions, even if enough are available to conduct the analysis [191]. The advent of high throughput sequencing technologies, alleviates the bottleneck of having too few of markers. SNP arrays for example, typically yield tens to hundreds of thousands of loci [58]. Despite the improved genome wide resolution however, it still does not preclude from the significant resource and time investment required to develop. Furthermore, SNP arrays are specific to a single population, so use in new populations would bias any alleles towards those present in the initial study population [80].

1.2.2.2 Restriction site associated DNA sequencing

Reduced representation sequencing, where only a representative fraction of the genome is sequenced, provides a much more cost effective solution for use in evolutionary and population genetics research compared to whole genome sequencing, while still providing enough resolution for studies of selection and genotype-phenotype associations. As only a fraction of the genome is sequenced and represented by the data per sample, hundreds of individuals can be sequenced simultaneously and genotyped, to yield thousands of markers along the genome at high coverage, and therefore high confidence. One of the most popular methods is restriction site associated DNA sequencing (RAD-seq) [20].

Unlike the targeted amplification and sequencing of microsatellites at known positions along the genome, RAD-seq uses restriction endonucleases to digest high molecular weight genomic DNA at restriction sites interspersed throughout the genome, in both coding and non-coding regions [80], thus resulting in a range of DNA fragment sizes (figure 1.2). Sequencing libraries are prepared by first ligating sequencing adapters, which have been designed to contain complementary ends to the restriction enzyme cut site, to the restricted DNA. The adapters also contain a forward amplification primer and known barcode sequences specific to each sample. The cost effectiveness and efficiency of this protocol becomes evident at this stage, where multiple samples can be pooled together as each sample is identifiable by its barcode. The DNA is then randomly sheared and size selected according to the read length of the sequencing technology ($<$ 700 bp if using Illumina for example). Y-shaped adapters containing the reverse primer sequence are subsequently ligated to the sheared end of the DNA fragments. This forked structure selectively amplifies only DNA fragments that have the forward primer already ligated, and has fully undergone the first-round elongation to complete the complementary reverse strand. Hundreds of samples can be prepared in this way and sequenced together.

In addition to yielding tens of thousands of loci, RAD-seq has the unique advantage of being highly user customisable by the choice of restriction enzymes, which allows researchers to target a specified subset of the genome. For example, many plants have very large genomes due to genome duplications, and have large amounts of non-coding repeat elements prone to higher levels of DNA methylation [225, 305]. Choosing a methylation sensitive restriction enzyme such as MspI, can avoid cutting methylated DNA repeat elements, thus excluding most of them from the sequencing library preparation if desired. In this scenario, it could be argued that exome capture should instead be used as DNA methylation acts to suppress transcription of DNA. However, this would also exclude many regulatory regions present in non-exonic DNA that are not repeat elements, which may be important in the study. Furthermore, exome capture requires a reference genome or transcriptome, which may not be available for non-model organisms [51]. RAD-seq therefore remains a viable and powerful method that enhances a population geneticists toolkit.

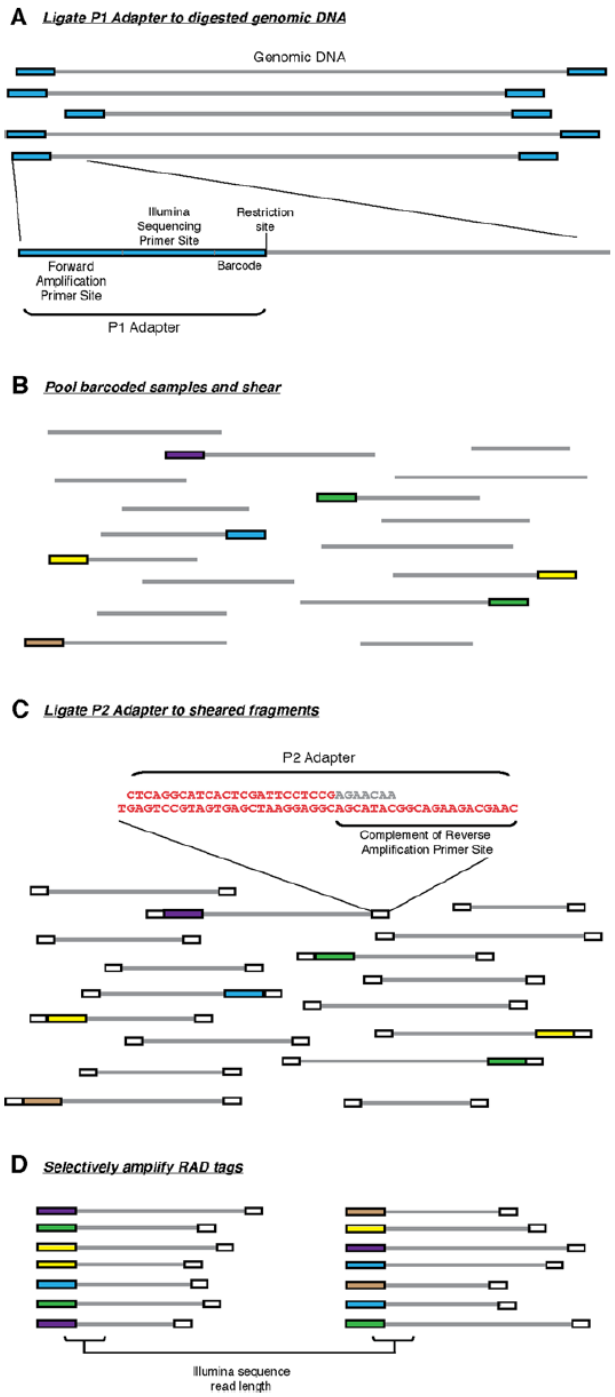


Fig. 1.2 The original RAD-seq protocol by Baird et al. [20]. A) High molecular weight gDNA is digested by a restriction endonuclease and barcoded Illumina P1 adapters are ligated. B) Adapter ligated and barcoded DNA fragments are pooled for multiplexing and sheered randomly. C) Y-shaped Illumina P2 adapters are are ligated. D) Only fragments containing both P1 and P2 adapters are selectively amplified.

Since the original RAD-seq study by Baird et al. [20], many variations have been developed to circumvent the unpredictability of the random shearing step, minimise hands on time for library preparation, and provide more control over the protocol and the target fraction of the genome, including with the use of multiple restriction enzymes [80]. Apart from these benefits, the greatest appeal of these protocols lies in their applicability to non-model organisms such as the genus *Apodemus*. In a recent study by Martin Cerezo et al. [53], RAD-seq was used in *A. flavicollis* and *sylvaticus* to demonstrate how a catalogue of shared markers between two closely related species can be developed to differentiate between them. They also showed the effectiveness of RAD-seq for estimating genetic diversity across their range.

1.2.2.3 Bioinformatics analysis of RAD-seq data

With the decreasing cost in high-throughput, next-generation short read sequencing technologies such as the Illumina platform, and the availability and flexibility of reduced representation library preparation protocols such as RAD-seq, researchers are producing considerably larger genomics data sets in organisms which have been traditionally unstudied at this level, including *Apodemus* [311]. However, drawing meaningful biological conclusions from such vast amounts of data requires efficient and robust pipelines for analysis. Many tools have therefore been developed for each stage of bioinformatic processing of the data, from demultiplexing and adapter trimming [145, 50, 211], sequence quality assessment [13], PCR duplicate removal [49], locus assembly either *de novo* or with a reference genome [111, 221, 271, 285], variant filtering [77, 270, 306] and analysis [198, 351, 355]. These citations given here are of course just a few examples of packages within the broad and expanding field of computational statistical genomics, and as our understanding of the topic improves, the need for more niche analyses increases. The list will therefore undoubtedly grow.

A number of pipelines have been developed specifically for genotyping-by-seqencing approaches such as RAD-seq, due to increasing popularity and various flavours of the protocol. This includes, but is not limited to: pyRAD, designed specifically for phylogenetics and

suited for broad taxonomic scales [91], UNEAK, part of the well known GWA analysis package TASSEL [37, 198], dDOCENT, a bayesian haplotype based variant caller [271], and arguably the most popular due to its versatility, computational efficiency, and relative user friendliness, STACKS [50, 49, 285]. As the different packages have been designed with various algorithms for locus assembly and variant calling, or in the case of pyRAD and UNEAK with different use cases in mind, the choice of pipeline can be critical to ensuring minimal bias from PCR duplicates, genotyping error and the effects of low sequencing coverage [11, 249]. Many studies have therefore compared the efficacy of different programs including STACKS, in multiple use case scenarios with empirical and simulated data [220, 285, 302, 313, 353]. Recent improvements in the STACKS genotype caller have shown it is robust and efficient, and outperforms other software [285]. Furthermore, analysis of RAD-seq data with STACKS in *A. flavicollis* and *sylvaticus*, has recently been successfully demonstrated by Martin Cerezo et al. [53], thus substantiating its use for evolutionary and population genomics research in *Apodemus*.

1.3 The study site - Białowieża forest, Poland

Developing a model for ecological and evolutionary research requires any findings to be placed in the context of the organisms environment. Although *A. flavicollis* has a large range, Białowieża Forest, a $1,250 km^2$ expanse of ancient forest straddling the Polish-Belorussian border was chosen for this study (52^o $30'$-53^o N and 23^o $30'$-24^o $15'E$, figure 1.3). Despite large parts ($\sim 39,000 ha$) of the forest currently being managed, a significant proportion has remained uncleared [174]. It is one of the largest and best preserved forest patches in Europe, and is abundant in *A. flavicollis* making it an ideal location for this study.

1.3.1 The paleoecological history of Białowieża forest

Paleoecological reconstruction of the forest performed with palynological stratigraphy (analysis of pollen trapped in stratified soil samples), has indicated that climate warming, which followed the last glaciation, led to the growth of boreal forests across much of central Europe

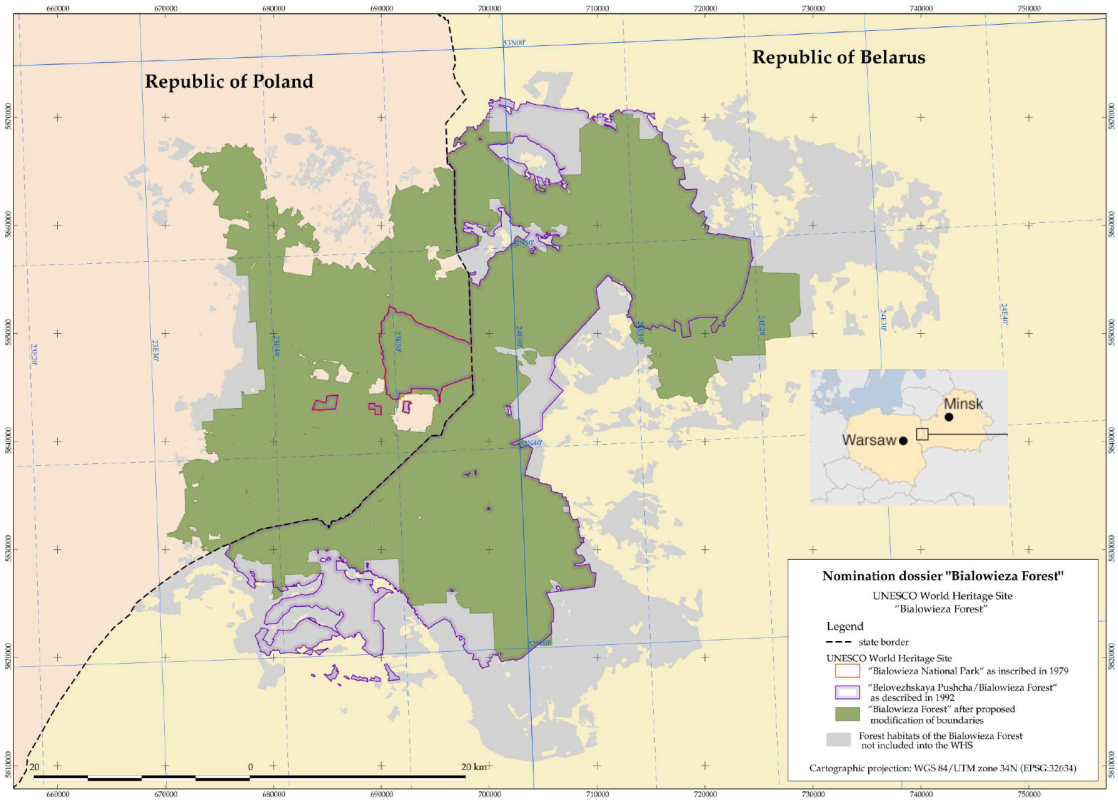


Fig. 1.3 Białowieża forest is located on the border between Poland and Belarus in Central Europe. Białowieża National Park (red outline) is found within the currently recognised boundary of the forest shown in green. These boundaries have changed since their original proposal in 1992 shown with a purple outline. Grey indicates the full extent of the forest habitat. The map was modified from UNESCO, and obtained under CC BY-SA 3.0.

(11,400-10,700 BC). The structure of the sediment cores follow the successional history of the forest closely [180]. At the end of the last glaciation, loose community structure meant large expanses of steppe vegetation were common in Białowieża, and melting of permafrost caused swamps to form in shallow ground. Changes in climate during the Allerød (11,400-10,700 BC) and Younger Dryas (10,700-9,750 BC) periods led to subsequent shifts in forest community structure, which was dominated by birch (*Betula*) and pine (*Pinus*) respectively. Further warming and changes in ground water levels due to increased surface runoff, then led to changes in soil chemistry and water clarity, thus altering the wetland ecosystems drastically. This hydrological instability eventually allowed the growth of more tree species species such as elm (*Ulmus*) and eventually small admixtures of oak (*Quercus*)

and hazel (*Corylus*) between 8,800-7,300 BC, therefore increasing the heterogeneity of the forest ecosystem.

Archaeological and palynological evidence indicates the presence human activity in the forest since the late Neolithic ($\sim$ 3000 BC). This activity, and fluctuations in soil moisture levels caused by prolonged dry periods during 7,300-1,800 BC caused frequent wildfires, further altering the forest ecosystem dramatically [180]. During this time, decreasing ground water levels, as indicated by lithological structuring, caused the drying of marshland and the retreat of pine forest [180]. This allowed mesophillic tree species such as alder (*Alnus*), oak, hazel and hornbeam (*Carpinus betulus*) to proliferate. Oak, hazel and hornbeam now dominate much of the forest due to economic exploitation by humans during the medieval period, where incandescent techniques to alter canopy cover were often used to change the structure of the forest [180, 358]. Highly fertile oak-horbeam habitats also led to an increase in agriculturally associated species around the edge of the forest. These complex events historically caused by variation in the hydrology, temperature and precipitation, and later by anthropogenic activities, have created the non-uniformity of the forest evident today. Large seasonal variations in temperature, precipitation and snow cover also contribute to the uniqueness of the ecology of Białowieża forest [148], where temperatures can vary from -25.0-1.8^oC ($\mu = -4.8^oC$) in winter, and 15.2-22.5^oC ($\mu = 18.4^oC$) in the summer. Snow cover can also vary significantly, from 1-95cm in harsh winters, though wet and mild winters are now becoming increasingly common [269].

1.3.2 The (semi-)natural forest ecosystem for scientific research

Despite the presence of humans in Białowieża Forest for the last $\sim$ 5000 years, pollen records indicate woodland continuity for at least the last $\sim$ 2000 years. This is largely due to the primary activity in the region being hunting and foraging, rather than logging or clearing of the forest [323]. Designated a Royal Forest during the Jagiellonian dynasty (c. 1362-1572 AD), it was largely protected. The continuity of the forest was therefore preserved and the ecological processes that maintain the forest in a mostly natural state remained. The structure and composition of the forest has been drastically altered in the last 250 years however,

mostly through the artificially high abundance of large game for hunting and grazing [234]. Recent control of game densities have now led to an invasion of oak stands which were created by excessive grazing, by pine trees through secondary succession, thus returning the forest to a more natural state [99].

The heterogeneity of the deciduous and riparian forest habitats created by Białowieża Forest's highly varied ecological history (figure 1.4) also supports a plethora of species, including over 2000 macro and lichenised fungi, 1,017 vascular plants, 12,000 insects, over 250 birds, and 59 mammals [26, 30, 174, 234]. Although the Białowieża Forest can no longer be considered *primeval*, the long term continuity and strong regenerative capacity shaped by ecological dynamics, has shown it is a strong subject for ecological research in a (semi-)natural system. The cultural and scientific value of the forest has therefore been recognised internationally, and the forest was placed on the United Nations Educational, Scientific and Cultural Organisation (UNESCO) World Heritage List in 2014. Significant long term biological research has therefore centred around Białowieża Forest, which is home to the Mammalian Research Institute of the Polish Academy of Sciences (MRI-PAS), which manages the *A. flavicollis* trapping site within the Strict Reserve of Białowieża National Park used in this study (52°43'N, 23°52'E).

1.4 Scope of the thesis

The ecology of the genus *Apodemus* has been studied over much of its broad range for almost a century. However, it has not gained prominence among the many other well-studied ecological and evolutionary model organisms, despite its important ecological role in natural systems. This is likely due to the technical limitations of evolutionary genomics research, which until recently has been unachievable in both the wet and dry lab, and further compounded by focus on developing resources for ecologically, epidemiologically and medico-genetically important species such as *Mus musculus*. However, novel protocols that integrate reduced representation library preparation methods like RAD-seq, and the massively-parallel next-generation short read sequencing technologies such as with the

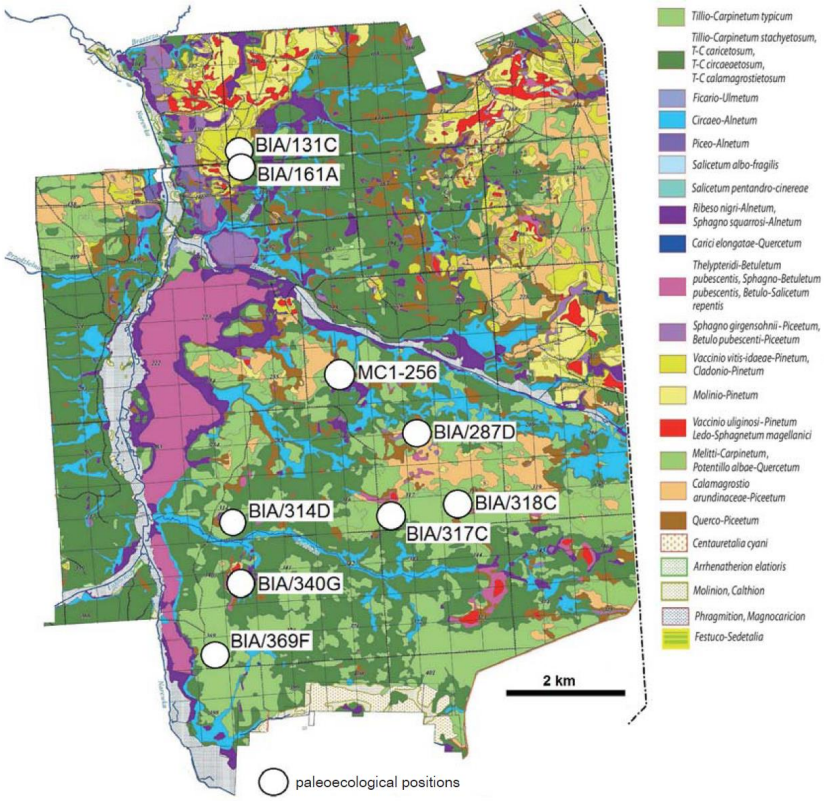


Fig. 1.4 The heterogeneity of Białowieża Forest by species of plants shown in different colours. Figure from Zimny et al. [358]. Circles indicate locations within the forest used for palynological sampling to reconstruct the paleoecological successional dynamics since the last glacial maximum by Zimny et al. [358].

Illumina platform, have revolutionised the way we study the complex evolutionary processes that maintain natural variation in the wild. The efficacy of this approach has now been demonstrated in a number of organisms, including in *Apodemus flavicollis* [53].

The study by Martin Cerezo et al. [53] used RAD-seq to observe the structure of a local population, estimate genetic diversity, and differentiate two species of *Apodemus*. This was a crucial first step in demonstrating the viability of RAD-seq in *Apodemus* for eco-evolutionary research. The aim of this thesis, is to build upon our knowledge of *Apodemus flavicollis* by making individual and population level observations using genome-wide markers to provide insights into some of the processes that maintain natural variation within a single population and species. Using RAD-seq data generated in a large sample of *A. flavicollis* from Białowieża Forest with a custom protocol (n = 672), I will demonstrate how measures

of relatedness within the population can be used to construct wild pedigrees when few field observations are available to support it. I will then describe the population's demography, and show how allele frequencies vary over four generations. I will also show how relatedness estimated from RAD-seq data, can be used to estimate the heritability of the highly variable and ecologically important physiological trait known as torpor. The results presented here will therefore further establish *Apodemus* as an excellent evolutionary and ecological model in natural systems.

The chapters included in this thesis are presented in the format of journal articles, and are briefly summarised below:

Chapter 2: Restriction-site-associated DNA sequencing for population and ecological genomics in the wild yellow-necked mouse, *Apodemus flavicollis*

This chapter shows how a well known variant of RAD-seq (quaddRAD [106]), was customised for use on a study of 672 tissue samples to produce a large-scale genomic dataset for conducting population level genomic analyses. Described in detail are the multiple steps involved in data generation including:

(i) *In silico* simulation of sequencing libraries at the planning stage.

(ii) The experimental protocols involved to produce sequencing libraries.

(iii) The bioinformatics processing and analyses involved to produce a robust dataset for ecological and evolutionary applications following DNA sequencing.

The results are then discussed in terms of the power and pitfalls of the approach, and its potential affects on future analyses.

Chapter 3: Population demography, fitness and allele frequency dynamics in a pedigreed, wild population of *Apodemus flavicollis*

This chapter describes how the genomic data generated in chapter 2 was used to estimate relatedness, and construct a pedigree in the wild population of yellow-necked mice where very limited field observations are available to support the analyses. The pedigree is then used to estimate individual and population level parameters to describe population demography, estimate the fitness of each mouse, and describe the allele frequency dynamics through time in the wild. The results are then discussed in the context of the eco-evolutionary dynamics of *Apodemus flavicollis*.

Chapter 4: The heritability of thermal strategies in wild *Apodemus flavicollis*

Chapter 4 ties together individual and population level data to show how phenotypic variation in the highly plastic heterothermic responses (torpor) can be explained by genetics (heritability) in wild *A. flavicollis*. Two distinct thermal strategies are then described that are constrained by the body mass and sex of an individual. Finally, the results are discussed in an evolutionary context to suggest how such a large degree of phenotypic variation is maintained in the wild.

Chapter 5: Conclusions and future directions

This final chapter summarises the results presented in this thesis, and discusses the future directions of the research using *Apodemus* as an ecological and evolutionary model.

Chapter 2

Restriction-site-associated DNA sequencing for population and ecological genomics in the wild yellow-necked mouse, *Apodemus flavicollis*

2.1 Introduction

Since the advent of next generation sequencing technologies, many sequencing library preparation techniques have been developed to improve accessibility for *de novo* research in non-model organisms. However, none have captured the attention of biologists like restriction-site associated DNA sequencing (RAD-seq) due to its low per-sample cost and hands-on time for library preparation. RAD-seq has now become a powerful tool for use in studies of phylogeography [50, 78, 95], phylogenetics [46, 53], linkage mapping [50], pedigree analysis [103, 189] and genome-wide association [295]. The original RAD-seq protocol by Baird et al. [20] has been modified considerably, and has diversified into many flavours for different use cases [11]. Here I use quaddRAD, a variant of the well known double digest RAD-seq protocol (ddRAD-seq) [106, 261], to generate RAD-seq libraries of a large sample of *A. flavicollis* from a single population, for ecological and population

genomics research. I will describe the protocol in detail from planning to execution, explain the rationale behind the choice of methods, and finally discuss the strengths and limitations of the approach in the context of evolutionary and ecological applications.

2.1.1 Double digest RAD-seq

Despite the low per sample sequencing cost of the original RAD-seq protocol, some limitations slowed its initial uptake among ecological and evolutionary biologists. For example, after digesting gDNA with restriction endonucleases, the protocol relied on the use of an expensive sonicator to further shear the template DNA, and repair fragment ends to allow adapter ligation. This process is completely random as shearing can occur anywhere on the molecule, and produces varying fragment lengths which must be size-selected by gel electrophoresis. Furthermore, sonication was known to fragment with varying degrees of efficiency depending on the length of the restricted template DNA [128]. This leads to some loci being lost if the DNA template is too large or small. ddRAD-seq circumvents this by using two restriction enzymes to more specifically target a desired proportion of the genome (figure 2.1), and produce fragments that are theoretically homologous among samples of the same species. As with sdRAD-seq, size selection can then be used to extract the desired portion of the genome. Hundreds of samples can be multiplexed for high throughput sequencing in this way, using combinations of barcodes within adapter sequences for sample identification.

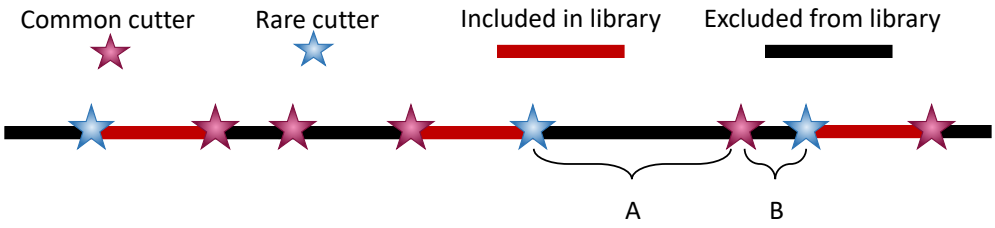


Fig. 2.1 Schematic representation of how genomic representation can be controlled through the use of common and rare cutting restriction enzymes in ddRAD-seq. Precise size selection excludes fragments too large or small, and are represented here by A and B respectively. Only fragments with different cut sites on each end are included in the final sequencing library.

2.1.1.1 quaddRAD - a ddRAD-seq protocol which allows for PCR duplicate removal

Despite its advantages, the ddRAD-seq protocol by Peterson et al. [261] still has two major limitations. Firstly, the unavoidable amplification step to enrich libraries can introduce artefacts such as PCR duplicates. These are clonal reads of template DNA which can skew allele frequencies by increasing genotyping error rates, and inflate estimates of homozygosity [266]. Unlike sdRAD-seq, which allows identification of PCR duplicates due to the random shearing creating different sized fragments, ddRAD-seq produces the same, predictably sized template DNA (figure 2.2). PCR duplicates therefore cannot be identified without further modification of the protocol. One method to identify duplicates is to incorporate a short degenerate base region within adapter sequences, to identify and computationally remove PCR duplicates without greatly increasing the costs for adapter synthesis [300]. This method was adapted for ddRAD-seq by Franchini et al. [106] in the quaddRAD protocol. The *quadd* here, refers to the use of four barcodes which can be incorporated combinatorially into quaddRAD libraries, and increases multiplexing capacity up to 192 samples compared to 96 in standard ddRAD-seq.

Increasing the multiplexing capacity allows quaddRAD to address the second limitation of ddRAD-seq, which is ascertainment bias due to repeated size selection steps when the number of samples to be multiplexed per lane is greater than 96. This is because cutting agarose gels to excise the desired size range can introduce variability in the library size distribution if done imprecisely. Loci can therefore be differentially represented when pooled, often causing some alleles to drop out. Automated size selection has reduced this error significantly (e.g. using the Pippin Prep from Sage Sciences), though can still be as high as 15% per size selection [106]. The higher multiplexing capacity in quaddRAD however, allows more samples to be pooled early in the protocol, so libraries can be amplified simultaneously. Size selection can then take place in a single step at the end of the protocol for up to 192 samples. This also has the added benefit of requiring less gDNA per sample, which is important for precious sample preservation in rare species or historical DNA samples. As with previous versions of RAD-seq, quaddRAD still remains highly customisable by allowing different enzyme combinations to be used based on the users preference.

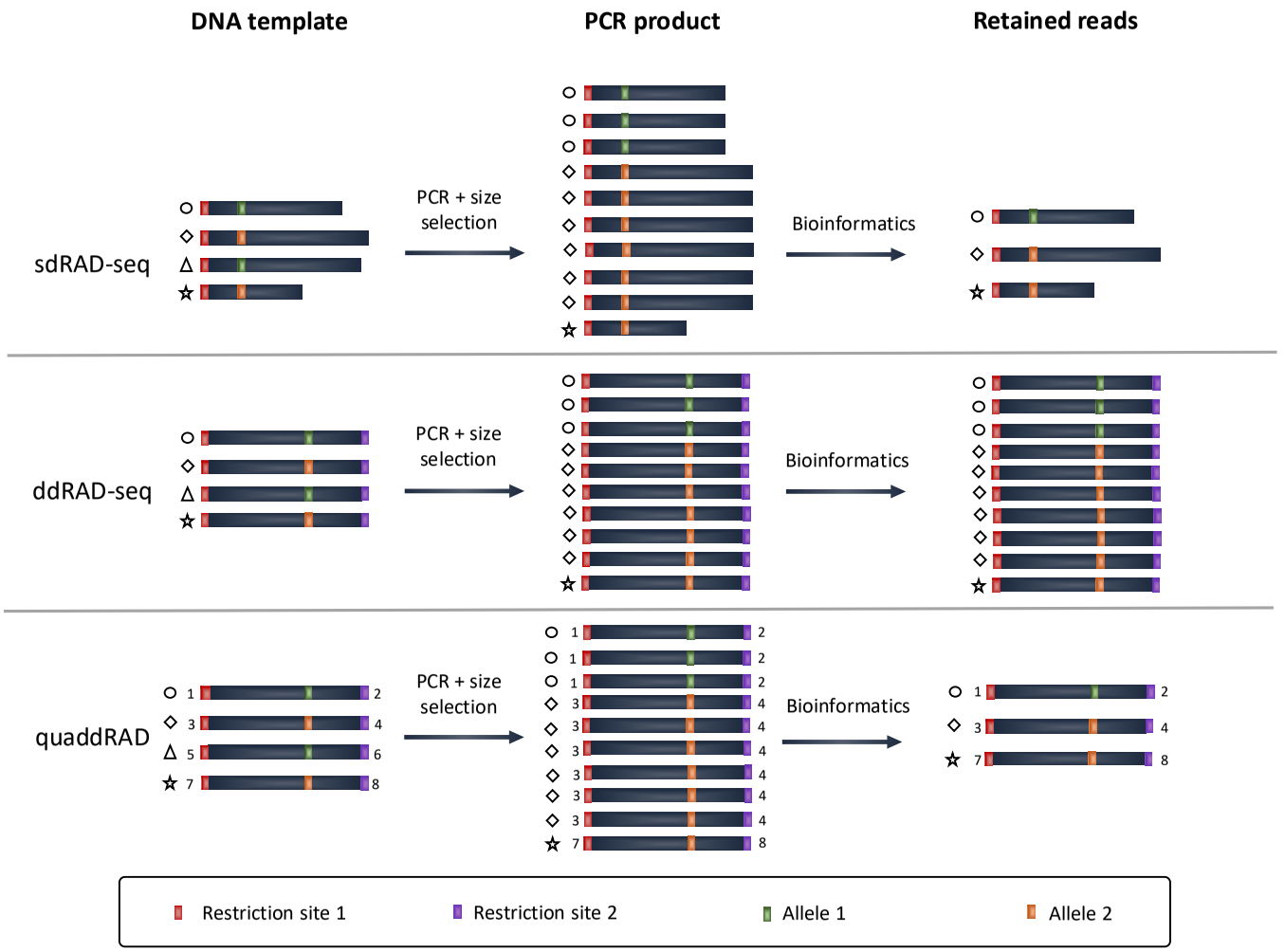


Fig. 2.2 How PCR duplicates for a single heterozygous locus can form in single digest RAD-seq, double digest RAD-seq and quaddRAD protocols. Shapes next to each fragment indicate from which parent fragment the PCR product originates (circle, diamond and star). In sdRAD-seq, duplicates are identified by the length of the fragment. If the same read with the same length is found multiple times, the additional reads are considered duplicates and only one is retained. Similarly, numbers on the quadRAD protocol indicate the unique degenerate base sequence within adapters to identify and remove PCR duplicates. Some alleles (triangles) can *drop out* of the library preparation in any protocol involving size selection. Figure adapted from Andrews et al. [11].

2.1.2 Bioinformatic processing of RAD-seq data using STACKS

Although many different pipelines are available for processing genotyping-by-sequencing data like RAD-seq, few have the versatility of STACKS [49, 50, 285]. High accuracy that is comparable to other software and its relative ease of use [331], have led it to become one of the most widely used RAD-seq analysis pipelines for ecological and evolutionary analyses. The versatility of STACKS lies in its use for both reference-based and *de novo* analyses, using the component pipelines `ref_map.pl` and `denovo_map.pl` respectively (figure 2.3).

Raw sequencing data can be input as either FASTA or FASTQ format into STACKS, and begins with quality control and demultiplexing. As mentioned previously, PCR duplicates are first identified based on the RAD-seq protocol used, and discarded in the program `clone_filter`. The reads can then be demultiplexed based on user provided barcodes, allowing for some mismatches to account for misincorporation of bases during sequencing. As RAD-seq data always starts with a restriction site, cut sites can also be rescued in the same way to ensure as few reads as possible are discarded. This is all conducted in the `process_radtags` program. *De novo* analyses then proceed by building putative alleles by aligning exactly matching reads in the component `ustacks` program, and loci in close sequence space are then merged to form putative loci. A catalogue of shared loci in the population is then built in `cstacks`, and samples are genotyped in `gstacks`. Reference based analyses forgo the steps in `ustacks` and `cstacks`, and instead proceed directly with genotyping in `gstacks`. The final stage is running the `populations` program. This program filters for minor allele frequency to remove low frequency variants if desired, and outputs SNPs shared among a user specified proportion of samples from the population(s) in a variety of formats. These include formats for STRUCTURE analysis [268], phylogenetic analysis using PHYLIP [100], and variant discovery (*.vcf*) [328]. It also computes population level summary statistics such as measures of heterozygosity for population differentiation and inbreeding (F-statistics) [345], and nucleotide diversity (π) [240].

Due to computational constraints, early versions of STACKS utilised a maximum likelihood framework to call SNPs at a given locus for each sample individually [140]. The presence of a SNP was determined by computing the log-likelihood ratio test statistic of the

most likely versus the second most likely SNP, thus providing an estimate of the sequencing error rate. Although this approach can be accurate [331], it requires high sequencing coverage (>15x) to differentiate between true SNPs and sequencing error [312]. Updated versions instead aggregate all samples, and observe biallelic SNPs in the entire population rather than per individual. This information is then fed as a prior into a Bayesian genotype calling model to finally genotype each individual [214, 285]. This is advantageous as it retains the robustness of the model without requiring such high coverage to accurately call SNPs [285].

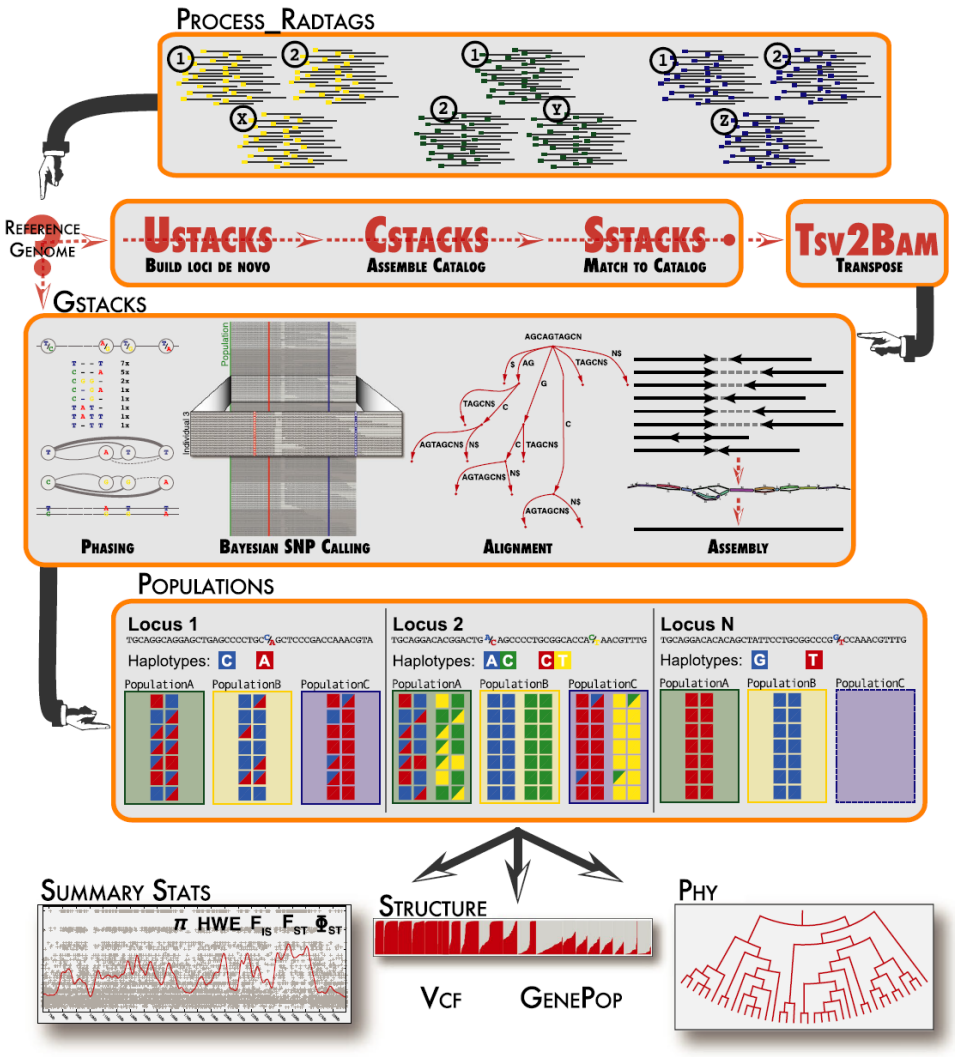


Fig. 2.3 The STACKS V2 pipeline, from demultiplexing of raw sequencing data, to final output of SNPs for use in ecological and evolutionary genomics research. Figure from Rochette et al. [285]

2.2 Materials and methods

2.2.1 *A. flavicollis* trapping and tissue sample collection

672 wild *A. flavicollis* were trapped between 2016-2019 in Białowieża Forest by collaborators in the Mammalian Research Institute at the Polish Academy of Sciences (MRI-PAS, Białowieża, Poland). The plot location within the forest is situated inside the Strict Reserve of Białowieża National Park (52^{o}43'N, 23^{o}52'E - figure 2.4) where 220 wooden traps were baited with oats at 110 trapping locations, set to form a 10x10m grid system in a 0.9ha area of mixed deciduous forest composed mainly of oak, hazel and hornbeam trees. Life history information including sex, body mass and approximate life stage were recorded. Tissue samples were taken as either tail or ear clippings, and stored in 96% ethanol at -20^{o}C until DNA extraction. As the location of trap boxes does not change, the same individuals were expected to be recaptured multiple times in the same trapping season. Passive Integrated Transponder tags (PIT-tags: RF-IDW-1, CBDZOE, Gryfice, Poland) with a unique 16 digit ID that can be read using a scanner were therefore injected subcutaneously on site, so recaptured individuals could be identified and potentially used for repeated phenotypic measures. However, due to the stochastic and natural setting in which the mice live, these PIT-tags can often fall out due to injury or predation, or simply due to improper subcutaneous injection. These recaptured mice cannot be re-identified, and so tissue samples were again taken and new tags injected. Identifying these duplicate samples then requires sequencing and genotyping (see section 2.2.3.3), estimation of genotyping error rates (see section 2.2.3.4) and finally merging of the data into a single sample ID. Due to the invasive nature of the tissue sampling and pit tagging procedures involved in mark-recapture experiments with vertebrates, all experimental procedures outlined in this thesis have been approved by the Local Committee for Ethics in Animal Research and the Ministry of the Environment, Poland (decision numbers 27/2016, 62/2017 and DOP-WPN.287.7.2016.AN.

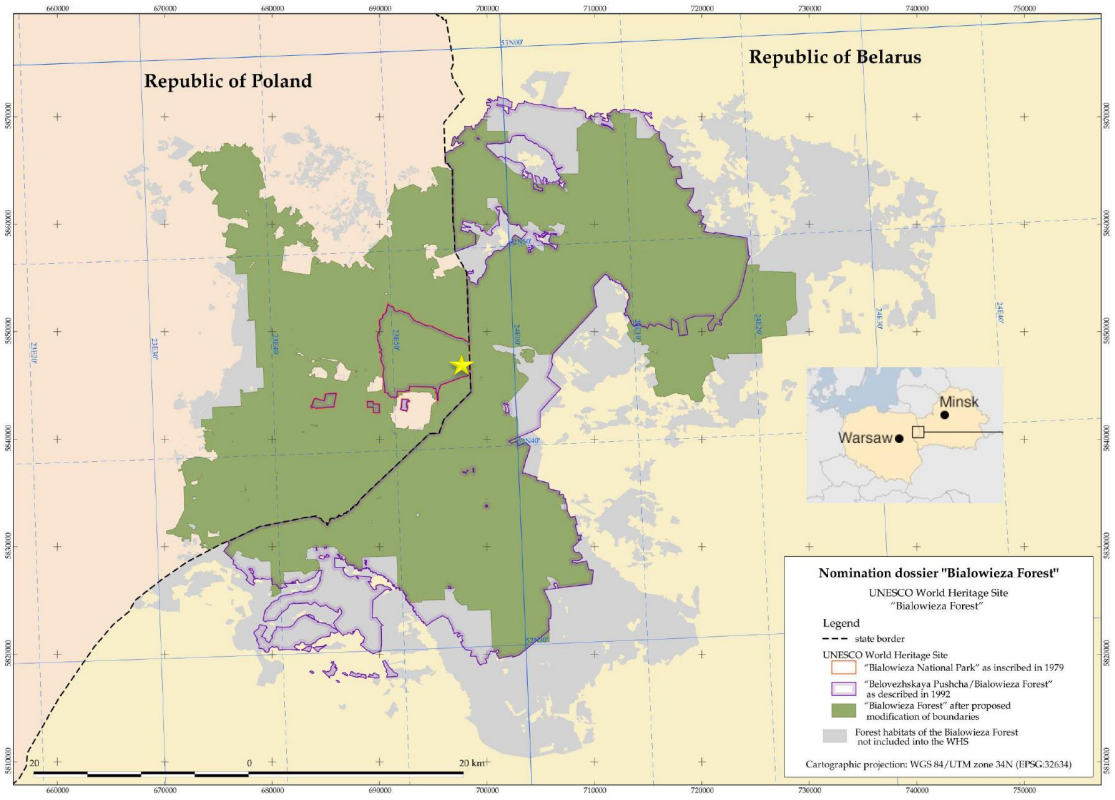


Fig. 2.4 Location of the 0.9 ha trapping site within the Strict Reserve of Białowieża National Park (yellow star inside the purple outline indicating the location of the Strict Reserve, within the forest shown in green). The map was modified from UNESCO, and obtained under CC BY-SA 3.0.

2.2.2 Restriction site Associated DNA sequencing

2.2.2.1 *In silico* restriction digestion and library preparation

Although reduced representation libraries decrease the total number of potential loci captured compared to whole genome sequencing, more individuals can be cost effectively multiplexed while maintaining a relatively high sequencing coverage [11]. This enables higher confidence in the obtained genotypes at a reasonable cost:performance ratio. Different combinations of restriction enzymes can be tested *in silico* without the need for any empirical work. The choice of common or rare cutting restriction enzymes combined with precise size selection results in a reduced representation DNA library customised to produce the desired sequencing coverage. This does however, require the genome sequence to be known beforehand which is not always possible in non-model organisms. Closely related species with a well established

reference can instead be used to ensure an accurate estimate of the final number of loci and sequencing coverage can be obtained. To simulate the library preparation procedure, known restriction sites which are recognised by the chosen enzymes, are searched for along the genome and the sequence is digested. The frequency with which these enzymes cut depends upon the length and representation of each base in the restriction site, and is expected to cut, on average, every 4^n base pairs along the genome, where n is the length of the restriction site. The type of cut (i.e. blunt or staggered), is determined by the enzyme itself. Libraries can then be simulated according to the choice of experimental design using single or double-digest RAD-seq with one or a combination of two restriction enzymes respectively.

Here, to provide finer control over the size and number of expected loci from each individual, ddRAD [261] was chosen to simulate sequencing libraries, and decide the most appropriate enzyme combination to obtain a desired >25X sequencing coverage (J. Catchen, 2018, personal communication). This can be estimated using $C = \frac{R}{NL}$, where C is the predicted coverage, R is the expected number of reads per lane, N the number of samples multiplexed per lane and L is the number of expected loci. ddRAD has proven reliable and effective for studies of evolutionary ecology and genomics in *A. flavicollis* by Martin Cerezo et al. [53], and the protocol has been optimised for this study system, so was preferred over other methods of RAD-seq. Section 2.2.2.4 describes in detail the experimental library preparation protocol used to modify and optimise ddRAD-seq for *A. flavicollis*.

Six different restriction enzyme combinations with varying proportions of bases and restriction site length were used to simulate ddRAD-seq libraries, and represent different cutting frequencies with staggered cuts only (table 2.1). This is to ensure non-specific ligation in the real library preparation protocol is minimised. As no reference genome for *A. flavicollis* is available, the reference for *Mus musculus* v.GRCm38 was used to simulate double digestion of each chromosome *in silico*, using the `insilico.digest` function of the SIMRAD V0.96 package [187] in R V3.5.2 [275]. Simulated library construction involved selecting fragments with different adapters ligated to each end only, to emulate a more realistic library preparation protocol [20] with the `adapt.select` function. Each resulting fragment was then mapped back to each chromosome with `matchPattern`, part

of BIOSTRINGS V2.50.2 [255], to return its index along each chromosome as a list. The distribution of fragment positions was then plotted on a karyoplot of each chromosome using the `kpPlotDensity` function of the KARYOPLOTER V1.8.7 package [118], to check for any bias towards different genomic regions. Analyses were conducted using a custom R script (available on GitHub, see appendix A.1 for details), and parallelised over 8 CPU threads using the `mclapply` function of the PARALLEL V3.4.0 package [275] to improve computational efficiency.

Restriction enzyme	Recognition sequence	Cut type
SbfI	5'...CCTGCA | GG...3' 3'...GG | ACGTCC...5'	staggered
EcoRI	5'...G | AATTC...3' 3'...CTTAA | G...5'	staggered
MspI	5'...C | CGG...3' 3'...GGC | C...5'	staggered
MluCI	5'... | AATT...3' 3'...TTAA | ...5'	staggered
NlaIII	5'... | CATG...3' 3'...GTAC | ...5'	staggered
MseI	5'...T | TAA...3' 3'...AAT | T...5'	staggered

Table 2.1 Restriction enzymes chosen for *in silico* digestion of the *M. musculus* v.GRCm38 reference genome. Only enzymes producing staggered cuts have been chosen to ensure non-specific ligation of adapters does not occur during library preparation. Red vertical bars in the recognition sequence show where the enzymes cut to produce an overhang.

2.2.2.2 DNA extraction and purification

DNA was extracted using a custom protocol in 2mL 96 deep-well plates. 672 tissue samples were cut into small pieces and lysed in 600 μL of lysis buffer (10mM tris-HCl, pH 8.0, 100mM NaCl, 10mM EDTA, 0.5% SDS, filter sterilised at 0.45 μm) and 16 μL proteinase K (20 mg mL^{-1}, Invitrogen) at 58^{o}C overnight. The reaction mixture was agitated throughout the incubation period at 80-100 rpm. 4 μL of RNase A (20 mg mL^{-1}, Invitrogen) was then

added to each lysed sample and incubated at 38°C for one hour, again with agitation at 80-100 rpm. Lysed samples were then incubated on ice for another hour with the addition of 200 μL 5M potassium acetate to precipitate SDS and any SDS-bound proteins, and subsequently centrifuged at 4500g for 30 minutes. 500μL of lysate was then transferred to a new deep well plate, being careful not to disturb the precipitate.

To ensure maximum yield whilst maintaining the integrity of the high molecular weight genomic DNA (gDNA), paramagnetic bead based capture was chosen to isolate and purify DNA from the lysate. This method uses 1 μm magnetite beads with a polystrene core and carboxylate modified polymer coating, to reversibly bind DNA by altering the concentration of NaCl in the solution (Solid Phase Reversible Immobilisation - SPRI) [83, 131]. Maintaining the integrity of high molecular weight gDNA maximises the number of intact restriction sites, and therefore the number of possible polymorphic sites for genotyping [126]. SPRI beads can also be used to size select fragments of DNA depending on the ratio of DNA to binding buffer. Fragments < 100bp do not bind regardless of this ratio. As maximum recovery is important at this stage, a ratio of 1:1 was used to yield > 90% recovery for fragments > 250bp long.

To prepare the SPRI bead mix, 700 μL of stock beads (Sera-Mag Speedbeads, GE Healthcare) were bound on a magnetic rack, washed twice in 10 mM Tris-HCl (pH 8.0) and resuspended in 50 mL of binding buffer (10 mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0, 2.5M NaCl, 20% PEG 8000 and 0.05% tween 20). 500 μL of bead mix was then added to 500 μL lysate that was transferred to a new plate, covered with a tightly-fitting adhesive seal and inverted twice to ensure thorough mixing of the viscous bead mix and lysate while preserving DNA integrity. The mixture was then incubated for 10 minutes at room temperature, and subsequently placed on a magnetic rack till a bead pellet formed. The supernatant was discarded once clear, and DNA bound beads were washed twice with 80% ethanol while still on the magnetic rack. DNA was then finally eluted in 80 μL of 10mM Tris-HCl (pH 8.0). The integrity of the extracted gDNA was qualitatively assessed on a 1% agarose gel with RedSafe nucleic acid staining solution (Chembio), and quantified using the Qubit Broad

Range Assay (Invitrogen) on a fluorometric microplate reader (FLUOstar OPTIMA, BMG). The gDNA for each sample was then normalised to 5 ng μL^{-1} with nuclease free water.

2.2.2.3 Adapter design

quaddRAD, a variant of ddRAD-seq by Franchini et al. [106], was chosen for library preparation and has been modified to reflect the choice of restriction enzymes based on *in silico* DNA digestion (figure 2.5), and the number of samples per lane to be multiplexed (figure 2.6). The advantage of this variant of ddRAD-seq lies in its use of four barcode sequences that allows up to 192 samples to be multiplexed, though only a maximum of 100 samples were multiplexed here to maintain high sequencing coverage.

A total of nine quaddRAD_i5n and nine quaddRAD_i7n adapters were designed with a unique 6bp inner barcode for multiplexing, and each included a 4bp degenerate region for PCR duplicate identification (full adapter and primer sequences available in appendix A.1, tables A.1-A.3). Barcodes were also designed to have a minimum hamming distance of 3, safely allowing 3 mismatches when demultiplexing reads to allow for sequencing error. Adapter overhangs were also modified to match *Mse1* and *Sbf1* restriction enzyme cut sites for ligation to digested gDNA.

2.2.2.4 quaddRAD library preparation

100 ng of gDNA from each sample was digested and adapters ligated in a single 40 μL reaction containing 4 μL 10x CutSmart buffer, 1.5 μL *Mse1* (10 UμL^{-1}), 0.75 μL *Sbf1* (20 U μL^{-1}), 4 μL ATP (10mM), 1 μL T4 DNA ligase (400 U μL^{-1}), 0.75 μL of each quaddRAD_i5n and quaddRAD_i7n adapter (10 μM) and ddH_2O to 40 μL. The reaction mixture was incubated for three hours at 30^{o}C in a thermocycler and stopped with 10 μL of 50 mM EDTA before being pooled according to different inner barcode combinations. Samples were then purified using SPRI beads in a 0.8:1 (beads:sample) ratio to remove fragments < 300bp, washed twice with ethanol (80%), eluted in 30 μL of 10mM tris-HCl and re-quantified using a Qubit 3.0 fluorometer and Qubit high sensitivity assay (Invitrogen).

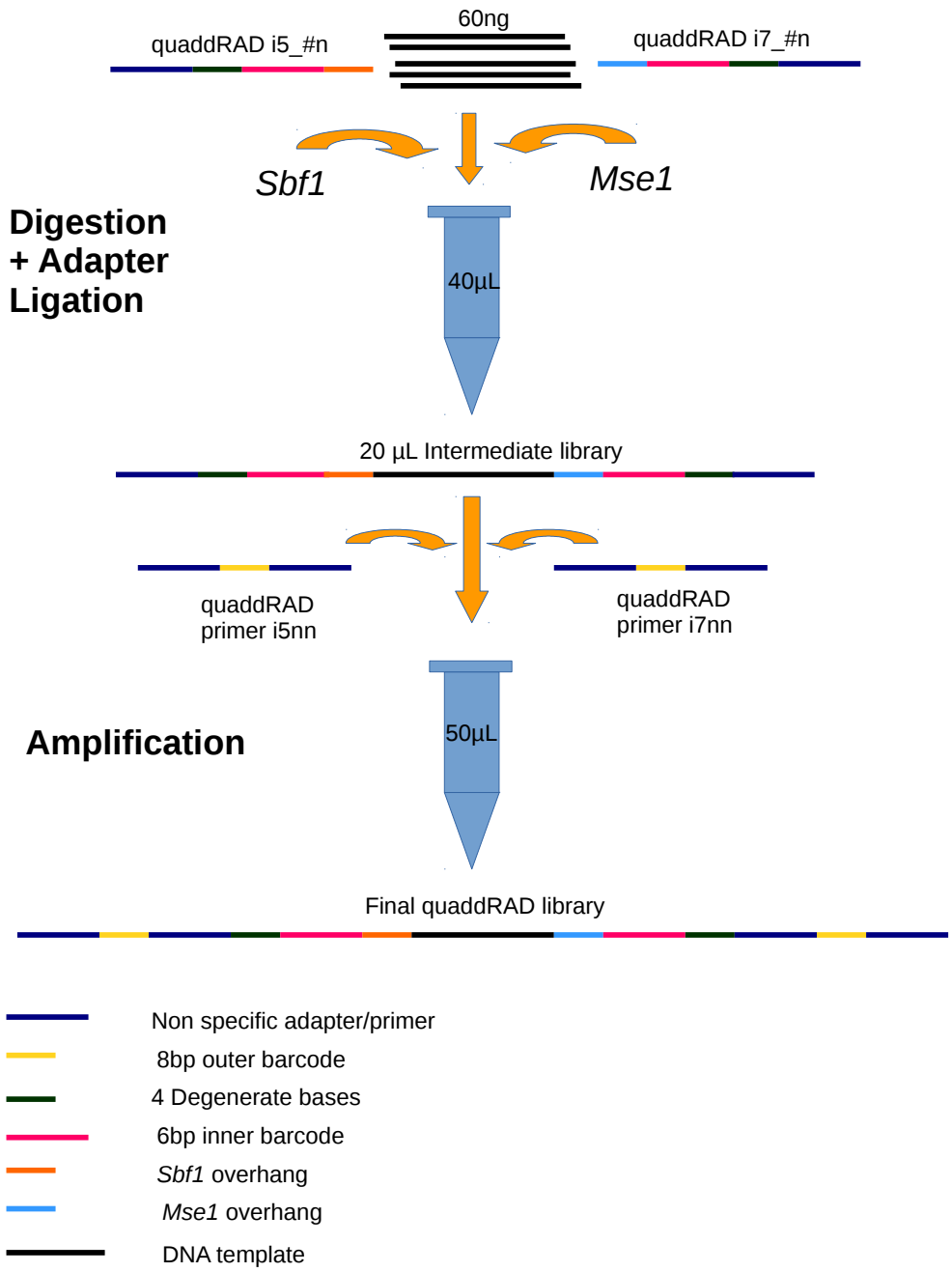


Fig. 2.5 An overview of the quaddRAD library preparation protocol. The adapters here have been modified for compatibility with the restriction enzymes *SbfI* and *MseI* (orange and light blue respectively), and to accommodate the number of samples to be multiplexed by the use of 4 barcode sequences (yellow and pink respectively). Degenerate base sequences allow for PCR duplicate identification (green). Figure adapted from Franchini et al. [106].

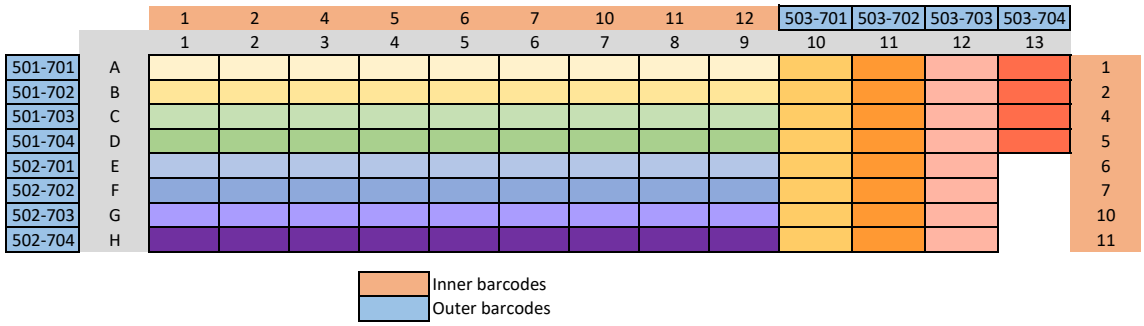


Fig. 2.6 Schema of how samples were multiplexed using combinations of inner and outer barcodes to ensure sequences from all samples were uniquely identifiable. Colours indicate samples belonging to the same outer barcode. A maximum of 100 samples were multiplexed per lane, over seven sequencing lanes.

An indexing PCR was carried out to introduce sequencing primers and a second pair of 8bp Illumina indexes to each pool of digested and inner adapter ligated DNA. The libraries were enriched in a 50 μL reaction mixture containing 4 μL dNTP mix (10 mM), 20 μL 5x Q5 buffer, 4 μL quaddRAD-i50n primer (10 μM), 4 μL quaddRAD-i70n primer (10 μM), 1 μL Q5 high-fidelity DNA polymerase (2 U μL^{-1}), 50 ng of DNA (restricted, ligated and pooled) and ddH_2O to 50 μL. Reaction conditions were as follows: 98^{o}C for 30s, [98^{o}C for 15s, 67^{o}C for 30s, 72^{o}C for 60s] repeated for 14 cycles, followed by a final extension at 72^{o}C for 120 seconds. Each PCR reaction was purified using SPRI beads in a ratio of 0.8:1 (beads:sample), washed twice with ethanol (80%), eluted in 50 μL of tris-HCl (10 mM) and re-quantified using a Qubit 3.0 fluorometer and Qubit broad range assay (Invitrogen). 100 ng of each enriched library was then pooled and size selected to 300-600bp using a Pippin Prep (Sage Science) and sent for paired end sequencing on an Illumina HiSeq 3000 at the Core Genomics Facility in the Max Planck Institute for Developmental Biology, Tübingen, Germany.

2.2.3 Bioinformatic processing

2.2.3.1 Demultiplexing and quality control

Outer barcodes were demultiplexed using BCL2FASTQ2 V2.18 [145] allowing for one mismatch, and sequences were qualitatively assessed using FASTQC [13] to ensure overall

sequencing quality was high (figure 2.8). Demultiplexing of inner barcodes and further quality control were then conducted in STACKS v2.3d [49, 50]. The program `process_radtags` of STACKS demultiplexes sequences based on assigned barcodes, uses a sliding window to remove reads with a PHRED quality score of less than 10, truncates reads to a specified length (in this case to 136bp to remove poor quality sequences at the end of reads, figure 2.8b), and rescues barcodes based on a specified number of mismatches (3 mismatches allowed, see section 2.2.2.3). PCR duplicates were then identified and removed using the `clone_filter` program in STACKS.

2.2.3.2 Locus assembly, parameter selection and genotyping using STACKS

Developing a final dataset of thousands of loci for population genomic analyses can be accomplished via two well documented STACKS pipelines, `denovo_map.pl` and `ref_map.pl` for *de novo* and reference-aligned sequencing data respectively. Regardless of the pipeline used, key parameter selection forms an essential part of the bioinformatic process as it is these parameters which ultimately determine the amount of biologically informative data available for downstream analyses. As no reference sequence is currently available for *Apodemus flavicollis*, `denovo_map.pl` was used here to assemble short read sequences from the RAD-seq data and genotype the samples.

Each component program of `denovo_map.pl` relies on various parameters to control the final assembly of loci within the population. Selection of the value of these parameters therefore depends on a number of factors including: the level of inherent polymorphism within the population of interest, the number of samples being multiplexed during sequencing and therefore sequencing coverage, the length of raw sequencing reads, the sequencing technology being used, and the associated amount of error during library preparation and sequencing. Parameter selection is therefore highly dataset specific, and must be chosen carefully. Three main parameters must be optimised in STACKS: m, the minimum number of reads required to form a putative allele and M, the number of mismatches allowed to merge putative loci, affect how loci are built within each sample in the core program `ustacks`. n, part of the core `cstacks` program, allows for a specified number of mismatches when

forming the catalog of homologous loci in the population. By iterating over these parameters and sequentially changing one whilst fixing the others as described by Paris et al. [256], the user can decide which combination of parameters maximises the biological information within the data. Here, *m*, *M* and *n* were varied between 2-6 (m2-m6), 0-8 (M0-M8) and 0-10 (n0-n10) respectively. While varying each parameter, all other parameters were kept constant at m3, M2 and n0. Data on the number of assembled loci, polymorphic loci and SNPs for *m*, *M* and *n*, contained within the *.log* files of each core component program of `denovo_map.pl`, were extracted using custom *bash* scripts available on GitHub (see appendix A.1 for details) and plotted using the GGPLOT2 V3.2.1 package of the statistical programming environment R V3.5.2.

The final step of `denovo_map.pl` involves running the `populations` module to calculate the number of assembled loci, polymorphic loci and number of SNPs present in at least 80% of the population (*r80* loci) as recommended by Paris *et al.* [256]. As RAD-seq is a reduced representation sequencing approach, it is important that the final subset of loci are representative of the whole genome with minimal bias. Selecting *r80* loci for optimisation ensures they are unlikely to contain a large amount of sequencing error, and serve as a representative proxy for the true genome. Additionally, the presence of multi-allelic SNPs (>2 alleles) in a population is expected to be rare [139], as a particularly large population size is necessary to reach a detectable frequency. Multi-allelic SNPs are therefore considered an artefact of a particularly noisy region of the genome here, and so `populations` was run with the additional `--write_single_snp` argument to ensure only the first of any alternate alleles were written to the final dataset.

2.2.3.3 Identification and merging of duplicate samples

The unpredictability of life in the wild means a number of previously captured mice were missing PIT-tags. As there is no way to verify whether these mice have been captured before, or which ID number they belong to, new PIT-tags were injected with a new ID number under the assumption these were *new* mice. Identification of recaptured individuals with more than

one ID was therefore necessary using the genotypes obtained from STACKS, as duplicated samples can introduce additional population stratification.

In any population, all individuals have portions of the genome which are identical by descent (IBD) due to shared ancestry. A high proportion of the genome found to be identical by descent, which in this case is indicated by a large number of SNPs shared between individuals (coefficient of relatedness, $r \rightarrow 1$), is an indication of either a very high degree of relatedness (e.g. identical twins or clones) or duplicate samples. To ensure duplicates are correctly identified, SNPs obained from STACKS must be stringently filtered to ensure false estimates of relatedness are not calculated.

Often, true low frequency alleles (minor allele frequency, MAF < 0.05) in a population are impossible to distinguish from noise generated by sequencing error, and were therefore filtered using the `--min_maf` argument of the `populations` program. Additionally, SNPs not in Hardy-Weinberg equilibrium (HWE) can indicate a region of the genome under selection. This is rare however, and is more likely an indication of genotyping errors or population stratification, which can lead to false estimates of homozygosity in downstream analyses [59]. SNPs not in HWE were filtered using PLINK V1.9 [270] ($p < 0.05$ with the mid-p adjustment [125]). Similarly, SNPs in linkage disequilibrium can also introduce additional structure, and decrease resolution in downstream analyses [253]. As locus assembly was conducted *de novo*, a true measure of linkage disequilibrium in centiMorgans (cM) is not possible. A statistical approach using PLINK was therefore used to calculate the correlation (r^2) between all pairs of SNPs previously filtered for HWE and MAF, as a proxy for true linkage. If polymorphism within the population is high, it is possible there may be more than one SNP present in some putative loci found by STACKS. Filtering for linkage disequilibrium by estimating r^2 between all pairs of alleles will then remove SNPs found on the same locus, and leave only one for downstream analyses. Individuals with > 70% missing data were also identified using PLINK and removed. Once SNPs were filtered, all individuals in the population were then re-genotyped, excluding those with > 70% missing data, and any SNPs/loci not in linkage equilibrium or HWE using whitelists in `populations`. SNPs were finally output in Variant Call Format (*.vcf*).

Pairwise IBD estimation was conducted in R v4.0.3 using the Bioconductor package SNPRELATE v1.22.0 [354]. Cleaned and filtered SNPs in VCF format and output by `populations`, were first converted to the binary Genome Data Structures (GDS) format using `snpgdsVCF2GDS` in SNPRELATE. This is a prerequisite for SNPRELATE to ensure fast and computationally efficient processing. IBD was then estimated using the functions `snpgdsIBDMoM` (PLINK method of moments [270]), `snpgdsIBDKING` (KING method of moments [205]), `snpgdsGRM` (Genome-Wide Compex Traits Analysis Toolkit - GCTA [351]) and `snpgdsIBDMLE` (Maximum Likelihood Estimation [232]) to ensure congruence among algorithms, and increase confidence in the identification of duplicates. Relatedness estimates from all four methods were then correlated in a pairwise analysis and plotted. Samples which clustered at the highest values of relatedness were then considered duplicated samples, and were subsequently merged by concatenating the raw sequencing data. `denovo_map.pl` was then run one final time as the merger of samples was expected to increase the average sequencing coverage, and alter the MAF of SNPs within the population. To generate the final panel of SNPs, all filtering steps for MAF, HWE and missingness were repeated on the merged data and output in *.vcf* format for further analyses.

2.2.3.4 Calculating the genotyping error rate

As the use of reduced representation sequencing techniques such as RAD-seq have increased, a growing body of evidence suggests specific biases increase genotyping error rates in RAD-seq libraries. Here, per-allele error rates have been estimated based on the sequencing of independently prepared replicate libraries, comparing both within and between sequencing lanes to estimate how much per-allele error rates vary across the entire dataset. The software package TIGER [38] was used to implement an expectation maximisation (EM) algorithm, and estimate the maximum likelihood (ML) estimates of per-allele genotyping error rates. This is based on the probabilities of observing the number of alternative alleles at a locus, given the true genotype and the per-allele genotyping error rate at homozygous and heterozygous sites, ε_0 and ε_1 respectively (table 2.2). This package was chosen as it allows the user to re-calibrate

the genotype likelihood directly within the *.vcf* file based on the estimated genotyping error rates, and take this into account in downstream analyses.

The data was analysed in RV4.0.3 by linear regression according to:

$$\varepsilon_{genotyping} \sim L + C_{max} + \varepsilon, \tag{2.1}$$

where $\varepsilon_{genotyping}$ is the maximum likelihood estimate of the genotyping error rate per sample, L is the sequencing lane, C_{max} is the maximum sequencing coverage per sample, and ε is the residual error of the model. In the event the data was highly skewed, it was Order-Norm transformed using the `orderNorm` function from the R package BESTNORMALIZE V1.6.1, to ensure the error distribution conformed to the assumptions of homoskedasticity. If transformation generated ties in the data, thus preventing it from conforming to a gaussian distribution and meeting the assumptions of linear regression, a negligible amount of random noise was added to the data to ensure each point was unique, thus breaking ties.

	Observed genotype		
True genotype	0	1	2
0	$(1-\varepsilon_0)^2$	$2\varepsilon_0(1-\varepsilon_0)$	ε_0^2
1	$\varepsilon_1(1-\varepsilon_1)$	$(1-\varepsilon_1)^2+\varepsilon_1^2$	$\varepsilon 1(1-\varepsilon_1)$
2	ε_0^2	$2\varepsilon_0(1-\varepsilon_0)$	$(1-\varepsilon_0)^2$

Table 2.2 The probability of observing a true genotype at a given biallelic locus and per allele error rate, at homozygous and heterozygous sites (ε_0 and ε_1 respectively). 0, 1 and 2 denote the number of copies of the alternative allele. Table modified from Bresadola et al. [38].

2.3 Results

2.3.1 Trapping summary

Overall, 287, 274 and 8 mice were caught in 2016, 2017 and 2018 respectively, within the 0.9 ha trapping site inside the Strict Reserve of Białowieża National Park. Of these, 113 were adults and 147 were juveniles. Ages were estimated based on the size and coat colouration of the mice, where younger mice generally have more grey fur and a less pronounced yellow band around the neck compared adults. However, as there is considerable phenotypic variation in *A. flavicollis*, there is often a large amount of uncertainty in estimating their age from field observations. 331 mice (49%) were therefore classed as of an indeterminate age (Maybe adult/juvenile and unknown age, table 2.3).

Sex is much simpler to determine, as males are generally larger/heavier, have longer ano-genital distances, and, particularly during mating, have large descended testes. Females also often have visible nipples, particularly during pregnancy and when postpartum, as they are still suckling their pups. 252 and 325 mice were classed as males and females respectively. 23 samples were found with no associated field information, so were of both undetermined age, sex and trapping year.

Trapping year	Adult	Maybe adult	Juvenile	Maybe juvenile	Unknown age	Males	Females	Unknown sex
2016	54	29	85	33	86	123	164	0
2017	57	28	63	15	111	123	151	0
2018	2	1	0	2	3	4	4	0
Unknown	0	0	0	0	23	2	6	15
Total	113	58	148	50	223	252	325	15

Table 2.3 Breakdown of sampled mice trapped within the 0.9 ha trapping site inside the Strict Reserve of Białowieża National Park.

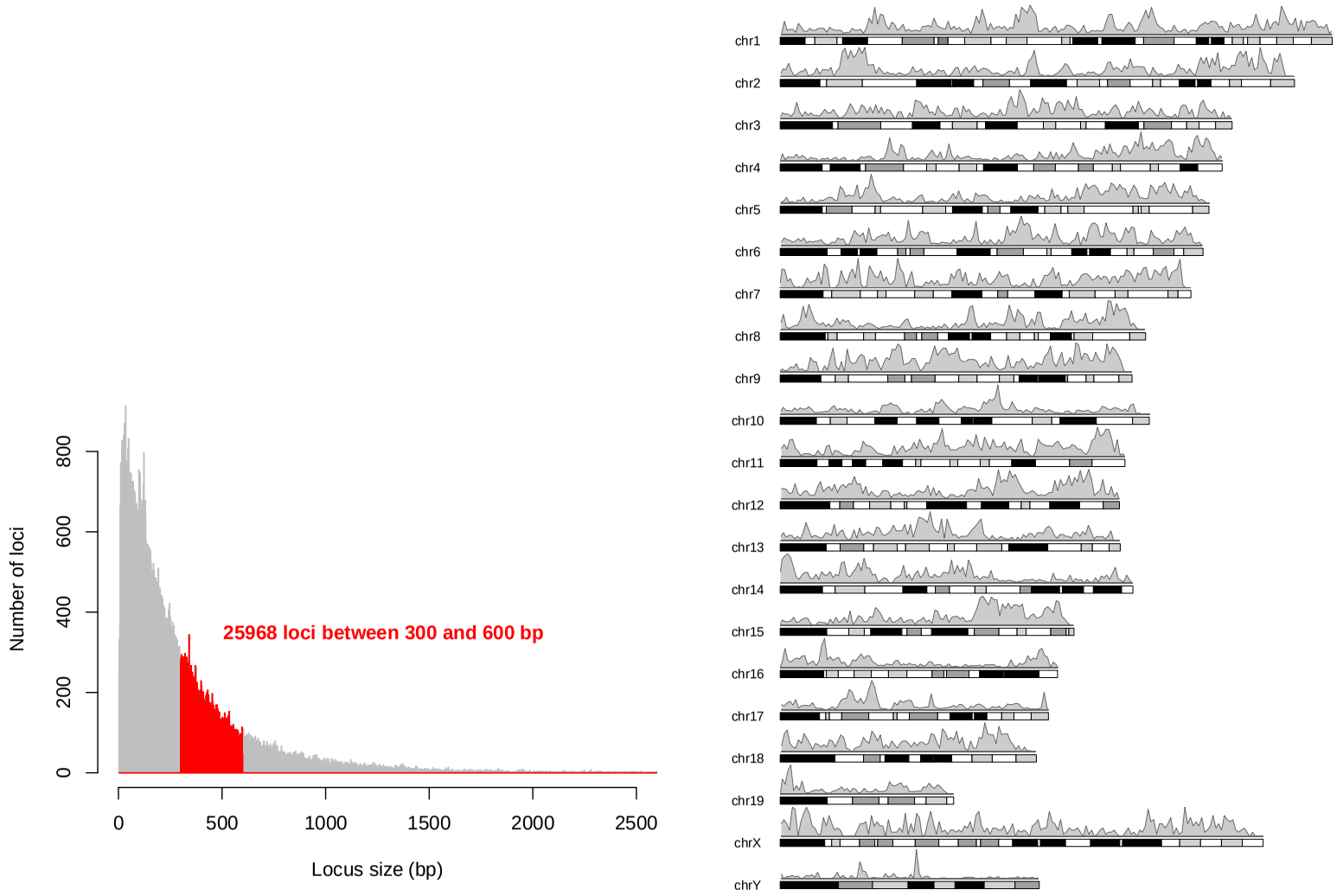
2.3.2 *In silico* restriction digestion and library preparation

Mean fragment lengths resulting from *in silico* restriction digestion of the *Mus musculus* v.GRCm38 reference genome vary between 94-330bp, and yield between 122,166 and 36,273,627 fragments. Expected sequencing coverage from these fragments is between 0.09X and 25.57X (table 2.4). When high quality DNA is extracted, larger mean fragment sizes, as produced by *SbfI* and *MseI*, or *SbfI* and *MluCI*, are desired to ensure the maximum number of variants are sequenced. Although the number of fragments expected from these two enzyme pairs are significantly lower than the other enzyme combinations, and both pairs perform similarly (which is expected as both *MseI and MluCI* have the same length restriction site with the same proportion of adenine and thymine bases), *SbfI* and *MseI* is expected to achieve the highest sequencing coverage whilst still capturing relatively large fragment sizes. Furthermore, Martin Cerezo et al. [53] have also previously used this combination effectively on *A. flavicollis* for phylogeographic analyses. For this study, *SbfI* and *MseI* are thus considered the most appropriate for library preparation and paired end sequencing on an *Illumina HiSeq 3000*, with 151bp read length. All further results will therefore only consider this enzyme combination.

Restriction enzymes	μ fragment size (bp)	N. fragments	Expected μ coverage (x)
EcoRI - MseI	184.45	1,493,745	2.09
MluCI - MseI	93.53	22,618,796	0.14
MspI - MseI	209.37	2,273,207	1.37
NlaIII - MluCI	93.14	36,273,627	0.09
NlaIII - MseI	93.04	22,337,740	0.14
SbfI - MseI	304.52	122,166	25.60
SbfI - MluCI	329.68	128,103	24.4

Table 2.4 Mean fragment lengths and expected sequencing coverage from *in silico* restriction digestion of the *Mus musculus* v.GRCm38 reference genome

Plotting the distribution of the expected number of loci by size indicates 25,968 loci (0.41% of the genome) can be expected from size selecting the prepared libraries between 300-600bp (figure 2.7a). Size selected libraries also indicate good autosomal representation (figure 2.7b), but poor representation of the Y chromosome suggesting studies of sex-linked phenotypes may not be appropriate from libraries generated using *SbfI* and *MseI*.



(a) (b)

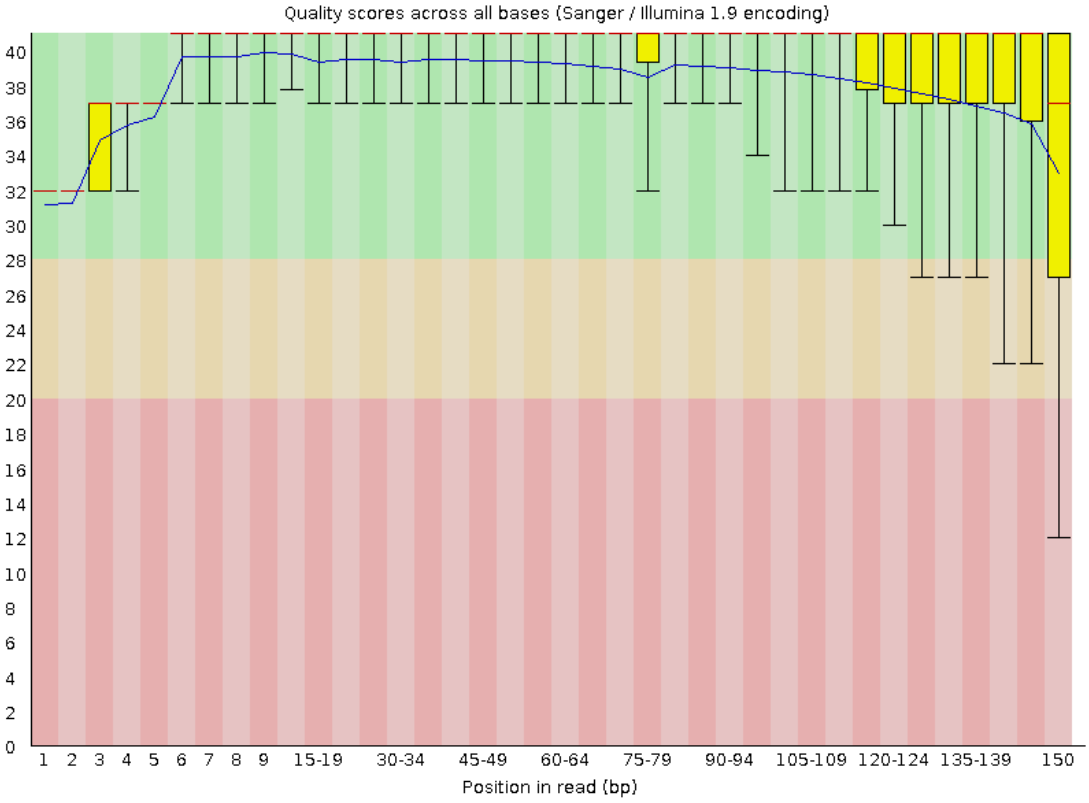
Fig. 2.7 Distribution of sequencing libraries expected from restriction digestion with *SbfI* - *MseI*. a) The resulting fragment size distribution from *in silico* digestion, where red indicates the expected number of fragments once libraries are size selected from 300bp to 600bp. Fragments >2,500bp are not shown (max = 300,118. b) Expected distribution of loci along each chromosome of the *M. musculus* v.GRCm38 reference genome.

2.3.3 Sequencing, demultiplexing and quality control

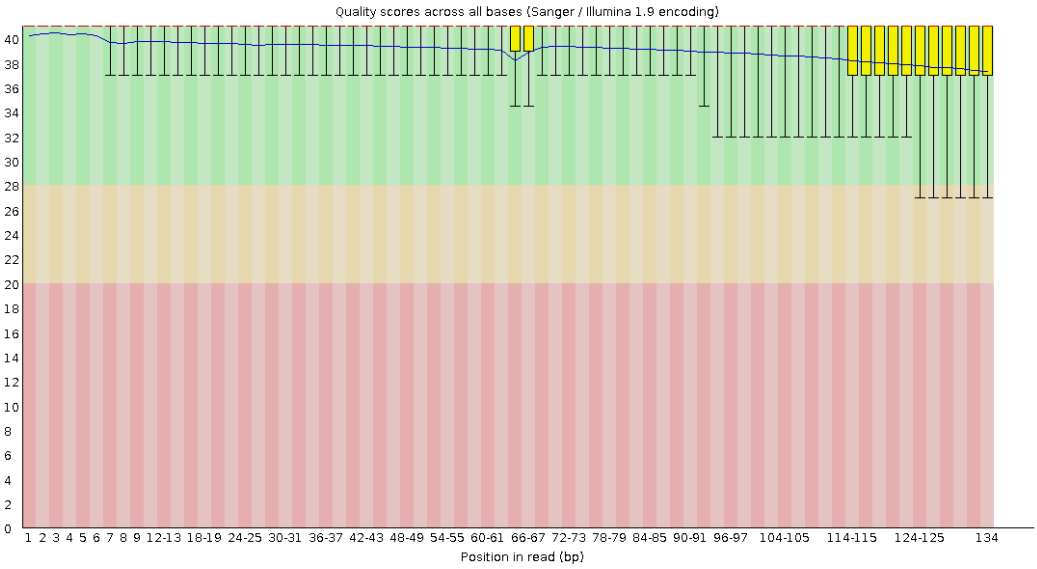
Reads from the Illumina HiSeq 3000 yielded a mean of 7.0×10^6 reads per sample. 95.1% of reads were successfully assigned an outer barcode with one mismatch allowed, and the mean PHRED quality score was 36.9. Raw sequencing data was qualitatively assessed using FASTQC (figure 2.8a), and lower quality at the end of the reads necessitated trimming to 136bp (figure 2.8b). Although the mean PHRED quality score of the raw data was high, the sliding window implemented in `process_radtags` gives a finer resolution to remove sequences where the quality drops below 10. Inner barcodes demultiplexed with more than 3 mismatches are also removed. 5.22% of reads were also identified as PCR duplicates or chimeric sequences respectively using `clone_filter`, and were also discarded. The final mean number of retained reads per sample after demultiplexing and quality control was 6.7×10^6.

2.3.4 Parameter optimisation in STACKS

Increasing the *m* parameter of `ustacks` from two to six increased mean sample coverage from 18.89X to 24.69X (figure 2.9, green) and further increased mean coverage from 25.24X to 31.94X once samples had been collapsed into putative alleles (figure 2.9, purple). As expected, the number of assembled loci decreased from 64,535 to 51,043 for *m2-m6* respectively (figure 2.10). Both mean coverage and the number of assembled loci, stabilised quickly from *m3-m6*, and the consistency in the number of assembled loci present in 80% of the population (*r80* loci), together with high coverage, suggests low sequencing and PCR error. Furthermore, as all other parameters were constant during optimisation, the low variance in the number of polymorphic loci and number of SNPs with each iteration of *m* indicate it has little effect on the levels of polymorphism detected within the population. The highest number of polymorphic loci is detected at *m2* where the number of *r80* loci is 38,579. However, the highest number of SNPs present in 80% of the population was at *m3* (140,421), though this increase is small and very few loci were excluded. Furthermore, slightly higher



(a)



(b)

Fig. 2.8 FASTQC output of sequence quality. Only output from one sequencing file is shown here as an example of quality a) before and b) after truncation to 136bp and removal of adapter sequences. Red signals poor quality reads.

mean coverage at m3 also indicates that the increase in the number of SNPs is likely a real signal, and not due to the promotion of error reads to putative alleles.

Overall, increasing values of *M* decreased the number of assembled loci and increased the number of polymorphic loci and SNPs in the population (figure 2.10). The mean number of assembled loci decreased from 70,239 to 53,539 and increased the mean number of polymorphic loci from 323 to 12,928 for *M0-M8*. Although both appear to stabilise at *M3*, the maximum number of *r80* polymorphic loci detected (37,960) is at *M2* suggesting this is the optimal value of *M*. This is further supported by no new polymorphic *r80* loci being identified after *M1/M2* (figure 2.11), indicating that a higher value of *M* past *M2* will not contribute any further biological information in the population. The mean number of SNPs also increases with each increment in *M*, and by *M2*, the majority of the polymorphism in the population, shown by the number of *r80* SNPs, has already been captured.

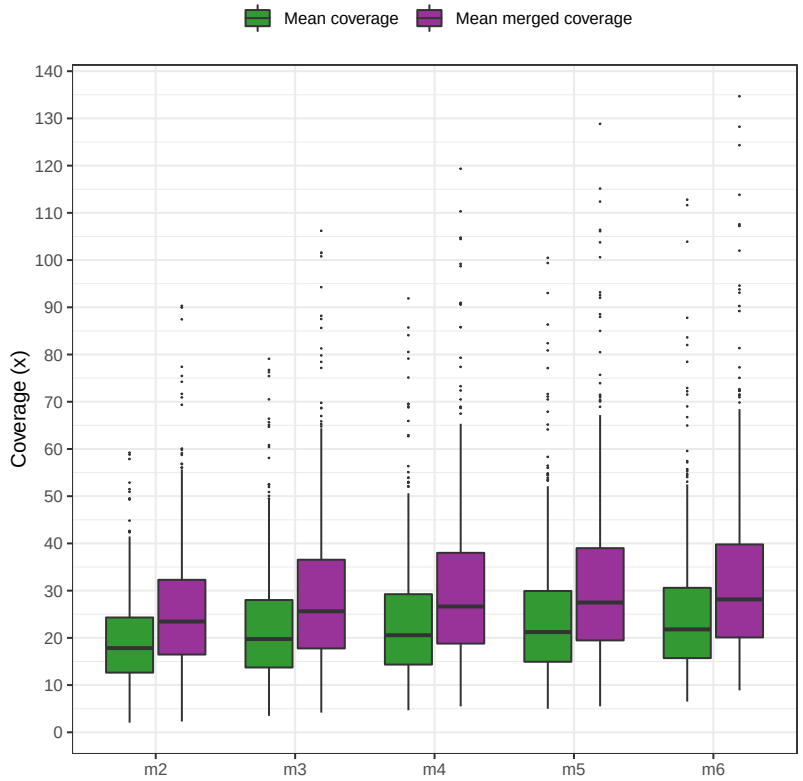


Fig. 2.9 Mean coverage (green) representing the coverage of primary reads, and mean merged coverage (purple) representing the mean coverage after primary reads have been merged into secondary reads.

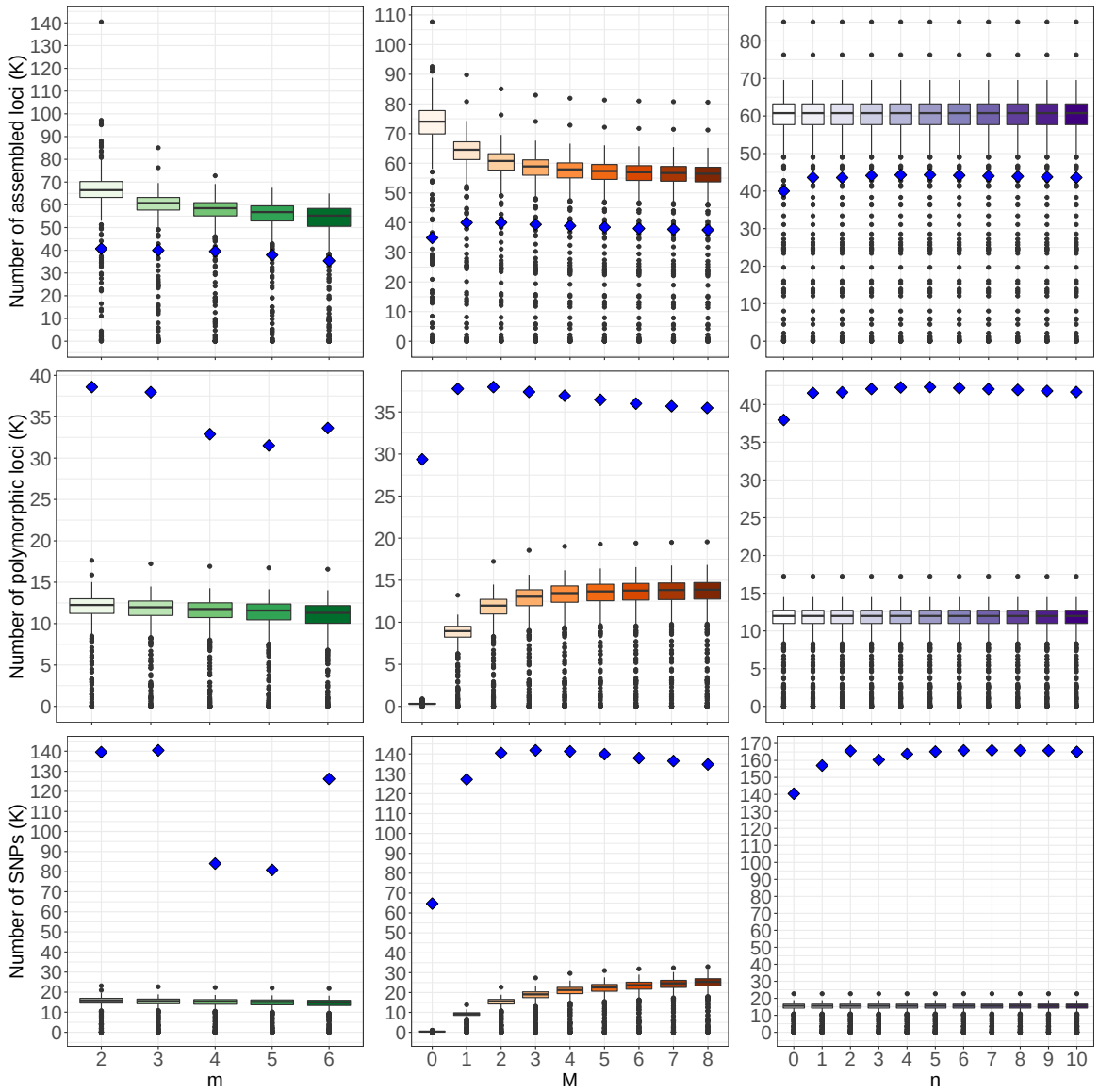


Fig. 2.10 The number of assembled loci, polymorphic loci and SNPs for each increment of *m* (green), *M* (orange) and *n* (purple). The number of *r80* assembled loci, polymorphic loci and SNPs (present in 80% of the population) are indicated by the blue diamonds

After assembly, matching homologous loci across samples in the population to create a catalogue of loci is carried out in `cstacks`, and the number of mismatches is controlled by the parameter, *n*. Here, the number of *r80* SNPs began to asymptote at *n2* with a total of 165,604. Predictably, as building the catalogue only involves matching SNPs from homologous loci across samples, the number of assembled and polymorphic loci does not change once a single fixed difference (*n1*) is introduced.

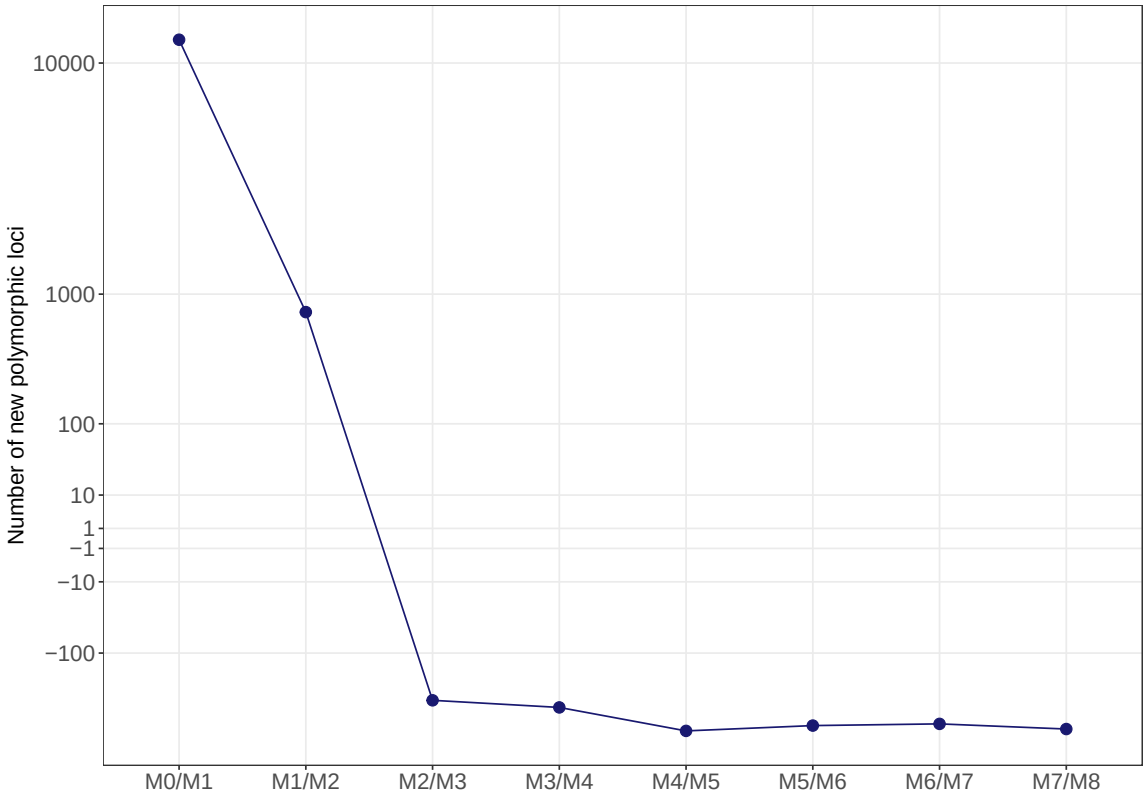


Fig. 2.11 The number of new *r80* polymorphic loci detected in the population between each increment of M.

Based on the above findings, the optimal parameters were decided as *m3*, *M2* and *n2*. Re-running the STACKS `denovo_map.pl` with this optimum parameter combination yielded 408,935 loci across the population, of which 32.37% were polymorphic, with a mean of 1.17 SNPs per locus and 479,593 SNPs overall.

2.3.5 Variant filtering and identification of duplicate samples

Of the total 479,593 SNPs detected following parameter optimisation, 448,533 low frequency variants were removed (MAF < 0.05), 8,624 were removed for not being in HWE ($p < 0.05$). Furthermore, calculating pairwise allele count correlations (r^2) indicated few SNPs were statistically associated (figure 2.12) and considered in linkage disequilibrium. 1,313 SNPs with an $r^2 > 0.035$ were thus removed to conservatively exclude the tail of the distribution.

28 samples were also removed due to having $> 70\%$ missing data, leaving a total of 644 samples and 21,123 remaining SNPs in the population.

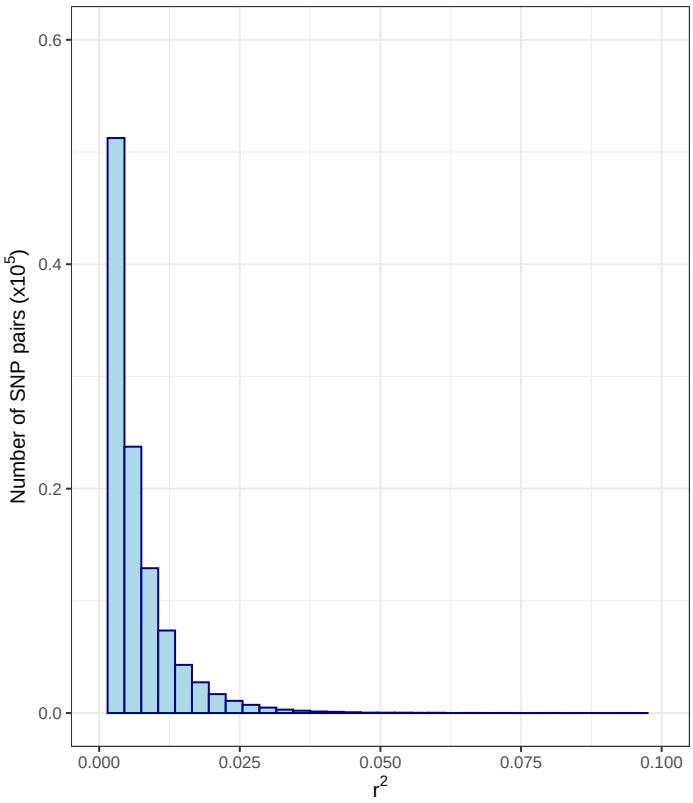


Fig. 2.12 Pairwise squared allele count correlations (r^2) for statistically associated SNPs. Only 13 SNP pairs are correlated with an $r^2 > 0.1$ and are not displayed. The maximum $r^2 = 0.997$ of which there is only one pair.

IBD estimation identified 55 pairs of samples as replicate pairs which clustered tightly in the top right corner of each pairs plot (figure 2.13). All four algorithms also identified the same samples (table 2.5, full table of pairwise relatedness estimates available on GitHub, see appendix A.1 for details). Relatedness in these samples was >0.9 when estimated using MLE, KING MoM and PLINK MoM. The GCTA algorithm was slightly more conservative however, and four pairs of samples were slightly lower ($r = 0.779$-0.898). Regardless, these samples still clustered tightly with all other identified duplicates, and all algorithms are congruent. Overall, estimates of relatedness correlate strongly (Pearson's $\rho_{207,044} = 0.568$-0.964, $p < 0.001$). 12 samples were also found to have more than two replicates, some of which were sequenced across different *Illumina* sequencing lanes. This allowed

for comparisons of genotyping error rates both within and between lanes (see results on Genotyping error rate, section 2.3.5.1).

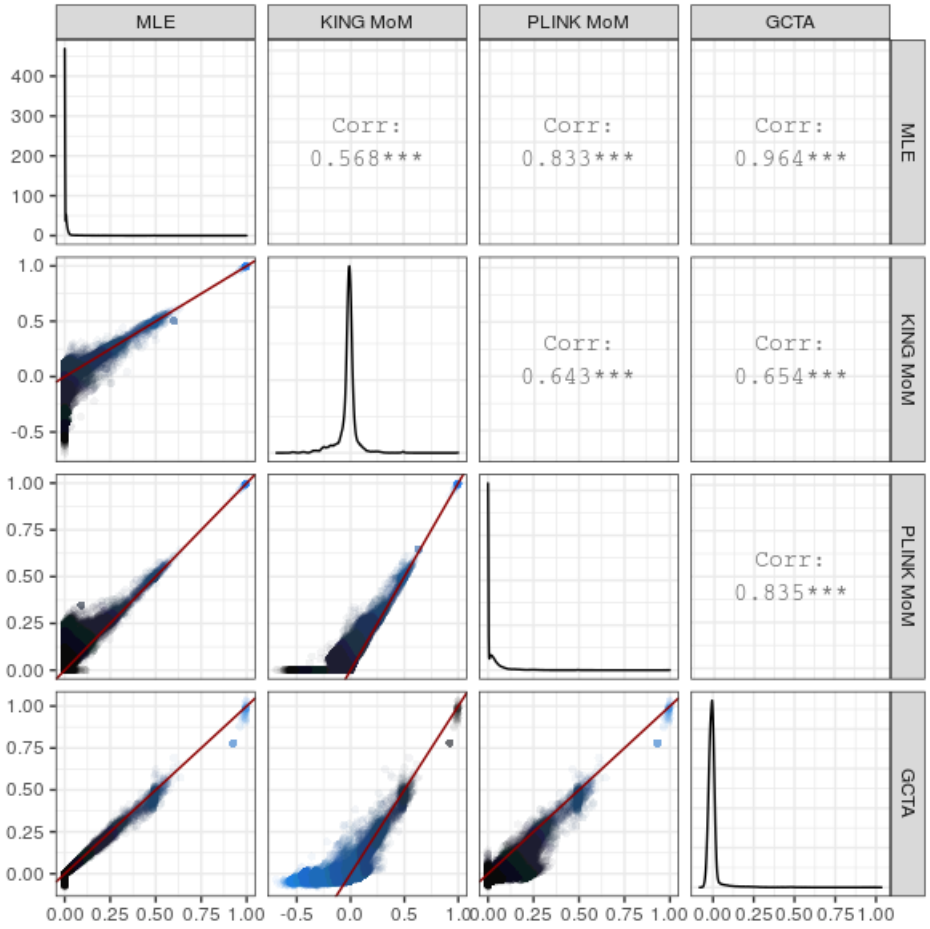


Fig. 2.13 Correlations of relatedness estimated from SNPs that are identical by descent, between Maximum Likelihood Estimation, KING Method of Moments, PLINK Method of Moments and GCTA. Duplicated samples cluster tightly in the top right corner of each plot. Correlation coefficients are calculated using Pearson's product-moment correlation. *** indicates $p < 0.001$. Histograms on the diagonal give an indication of skewness. Shades of blue are purely for visualisation purposes to show where the density of the points is highest (darker is more dense).

Sample 1	Sample 2	MLE	KING MoM	PLINK MoM	GCTA
Af_105	Af_180	0.99035	0.99037	0.99045	0.9614
Af_110	Af_149	0.99557	0.99539	0.99533	0.98982
Af_111	Af_550	0.99729	0.99735	0.99732	0.92776
Af_115	Af_262	0.9965	0.99651	0.99651	0.98535
Af_11	Af_84	0.98463	0.98493	0.98474	0.9669
Af_122	Af_328	0.99365	0.99364	0.99365	1.00175
Af_127	Af_319	0.99756	0.99761	0.99756	0.98355
Af_131	Af_168	0.99811	0.99807	0.99811	1.03318
Af_139	Af_172	0.99889	0.99889	0.9989	0.99526
Af_139	Af_207	0.9961	0.99611	0.99611	0.99093
Af_144	Af_181	0.99463	0.99469	0.99461	0.99535
Af_145	Af_200	0.99046	0.99048	0.99051	0.97839
Af_148	Af_151	0.9965	0.9965	0.99652	0.98935
Af_16	Af_634	0.99643	0.9964	0.99643	1.00091
Af_172	Af_207	0.99642	0.99643	0.99643	0.99039
Af_176	Af_229	0.96284	0.96084	0.96453	0.881
Af_177	Af_615	0.99235	0.99264	0.99242	0.92019
Af_178	Af_28	0.99388	0.99409	0.99396	0.90684
Af_179	Af_79	0.97365	0.97344	0.97422	0.9192
Af_182	Af_499	0.97736	0.97778	0.97763	0.93785
Af_21	Af_243	0.98605	0.98602	0.98622	0.96558
Af_21	Af_62	0.99627	0.99627	0.99628	1.00499
Af_226	Af_57	0.92519	0.9213	0.93115	0.77977
Af_22	Af_660	0.97944	0.97988	0.97973	0.91884
Af_23	Af_604	0.99023	0.99016	0.99015	0.97991
Af_243	Af_62	0.98564	0.98562	0.98581	0.96286
Af_245	Af_61	0.964	0.96368	0.96452	0.93297
Af_259	Af_42	0.99406	0.99398	0.99411	0.96583

Af_264	Af_369	0.99842	0.99841	0.99843	1.02529
Af_269	Af_66	0.99616	0.99621	0.99618	0.96562
Af_269	Af_73	0.98853	0.98876	0.98864	0.93734
Af_276	Af_339	0.99832	0.99853	0.99857	0.99476
Af_309	Af_380	0.99632	0.99632	0.99635	0.98076
Af_309	Af_482	0.99537	0.99537	0.9954	0.98076
Af_314	Af_641	0.99262	0.99264	0.99265	0.97224
Af_330	Af_645	0.99485	0.99489	0.99485	0.99019
Af_342	Af_633	0.99222	0.99239	0.99228	0.94899
Af_347	Af_82	0.99792	0.99805	0.99803	0.92885
Af_348	Af_360	0.99706	0.99713	0.99707	0.95949
Af_373	Af_7	0.99119	0.99159	0.99131	0.89811
Af_380	Af_482	0.99913	0.99913	0.99913	0.99067
Af_413	Af_594	0.99695	0.99693	0.99699	0.952
Af_416	Af_479	0.9989	0.99889	0.9989	0.99917
Af_477	Af_98	0.98195	0.98189	0.98221	0.95007
Af_47	Af_605	0.9946	0.99463	0.99462	0.98182
Af_489	Af_593	0.98915	0.98931	0.98924	0.94933
Af_501	Af_551	0.99882	0.99882	0.99882	0.98075
Af_523	Af_547	0.99812	0.99795	0.99796	1.02968
Af_527	Af_543	0.99168	0.99158	0.99174	0.98065
Af_539	Af_647	0.99709	0.99714	0.99709	0.99008
Af_552	Af_606	0.99437	0.99452	0.99445	0.91806
Af_560	Af_561	0.96329	0.96157	0.96485	0.87386
Af_570	Af_571	0.9974	0.99745	0.99741	0.96163
Af_619	Af_668	0.99875	0.99876	0.99875	1.0117
Af_66	Af_73	0.98689	0.98736	0.98697	0.93241

Table 2.5 Duplicate samples identified by calculating relatedness coefficients with Maximum Likelihood Estimation (MLE), KING and PLINK Method of Moments, and GCTA.

2.3.5.1 Genotyping error rate

As 12 pairs of samples were trapped and re-tagged more than twice (the identity of which was unknown till relatedness was estimated), sequencing these samples involved preparing their libraries independently. They were therefore sequenced across all sequencing lanes allowing comparisons of per allele genotyping error rates across them all. Overall, mean per allele genotyping error rates were 0.020 ($sd = 0.076$). Linear regression indicated that while there was a difference in genotyping error rates between sequencing lanes ($F_{(7,1325)} = 95.58$, $p < 0.0001$), it was only significant in lane three (table 2.6), and was up to an order of magnitude higher than the other lanes. As expected, maximum sequencing coverage per sample was highly inversely proportional with per allele genotyping error rates.

	Estimate	Std. error	*t* value	*p*	
Intercept	0.3540	0.0728	4.859	< 0.0001	****
Lane 2	-0.0521	0.0928	-0.561	0.5747	
Lane 3	1.2886	0.0884	14.566	< 0.0001	****
Lane 4	-0.0554	0.0886	-0.625	0.5319	
Lane 5	-0.0379	0.0920	-0.413	0.6800	
Lane 6	0.1748	0.0929	1.881	0.0602	
Lane 7	0.1121	0.0873	1.284	0.1994	
Maximum coverage	-0.0060	0.0004	-16.040	< 0.0001	****

Table 2.6 Linear regression of genotyping error rates between sequencing lanes, and the effect of sequencing coverage. **** indicates a statistically significant p-value < 0.0001. The intercept represents sequencing Lane 1 with a maximum coverage of 0x.

2.4 Discussion

Highly multiplexed approaches for reduced representation DNA sequencing have now become a cornerstone for maximising the level of biological information from a population of samples, at a reasonable cost:performance ratio. Obtaining reliable genotypes are essential to the robustness of NGS datasets, and careful consideration must be given to each stage of the process from study design and planning, to quality control and analysis, to minimise bias and avoid erroneous biological conclusions from RADseq [11, 47, 75, 80, 106, 239, 312]. Here, I have described the various steps involved in genomic data generation using a customised ddRAD-seq protocol, to sequence and reliably genotype a large sample of *A. flavicollis* from Białowieża National Park.

2.4.1 Choosing the right restriction enzyme combination

One of the advantages of RAD-seq is being able to multiplex many samples and reduce per sample sequencing costs. However, achieving high confidence genotype calls requires high sequencing coverage which can be reduced by a large per-pool sample size [11]. Although 15x coverage can be sufficient to call genotypes in RAD-seq studies [312], higher coverage can increase confidence in genotype calls and should be aimed for when designing a RAD-seq experiment [38, 256]. This becomes particularly important when choosing a protocol. For example, single digest RAD-seq requires an expensive sonicator to further fragment digested gDNA and yield fragments of various sizes, which can then be selectively amplified during library preparation. Although these different fragment lengths can be beneficial by allowing to identify sources of bias such as PCR duplicates [12], sonicators are known to inefficiently shear already short fragments following restriction digestion. This in turn yields unequal sequencing coverage per sample [294], and could result in excess homozygosity in studies of population differentiation. Alternative approaches to RAD-seq study design such as the double digest approach using two restriction enzymes have thus been favoured by many researchers [261]. ddRAD-seq allows greater flexibility and control over sequencing coverage and fragment sizes with the use of two restriction enzymes of different cutting

frequencies. The choice of these enzymes is essential to minimising bias by ensuring the desired portion of the genome is targeted and represented in the final sequencing libraries. Recent variants of ddRAD-seq, including quaddRAD used here, have also incorporated short degenerate base sequences for PCR duplicate identification, forgoing the necessity to shear DNA. When preparing any RAD-seq experiment, it is therefore essential to plan the study based on the expected coverage by choosing the appropriate enzyme(s). *in silico* restriction digestion of a reference genome is one useful method of predicting sequencing coverage and the final number of loci. Although the power of RAD-seq lies in providing access to non-model organisms without *a priori* knowledge of the genome, simulating library preparation using reference genomes of even relatively phylogenetically distant species can still provide useful insights for study design.

Here, SIMRAD was used to conduct *in silico* restriction digestion and library preparation of the *Mus musculus* v.GRCm38 reference genome. Despite the 8-24% divergence between *Apodemus* and *Mus* [301], simulated libraries provided close to the number SNPs from the actual protocol (25,968 predicted and 21,123 actual after filtering), and are predicted to represent $\sim 0.41\%$ of the genome at high coverage. However, high divergence between the two genera but similar numbers of loci suggests that the RAD-seq libraries generated here may over represent conserved genomic regions. Analyses could then struggle to differentiate closely related populations or species in phylogenetic and phylogeographic studies. However, RAD-seq using the restriction enzymes *SbfI* and *MseI* as in this study, has been previously demonstrated to show signatures of differentiation in the two highly morphologically similar sister species *A. sylvaticus* and *A. flavicollis* (1.51% sequence divergence) using 21,000 shared loci [53]. Furthermore, many other studies have shown its utility in phylogenetics, and RAD-seq significantly improves on locus recovery compared to other methods [46, 216]. These studies do require thousands of loci to ensure sufficient power for evolutionary research however, and this must be factored in when designing the study through the choice of restriction enzyme(s).

In a previous study by DaCosta and Sorenson [75], restriction enzymes with a higher proportion of GC bases in the restriction site resulted in significant amplification bias towards

short fragments from GC-rich regions of the genome. The *SbfI* restriction site has a 75% GC content and could also bias the libraries in this way. However, the low CpG content in the *M. musculus* genome [65] suggests amplification bias towards GC rich regions may be of less concern. However, this assumption cannot be made for *A. flavicollis*, as few genomic resources are available for the species. The use of a high fidelity polymerase such as Q5 here, in combination with an AT rich enzyme such as *MseI* may instead overcome these biases [271]. As only fragments with different restriction sites on each end are amplified during library preparation, this should also further minimise bias towards specific genomic regions, which is important in genome wide association studies. This is supported by the larger fragment sizes generated from the use of two frequent cutting enzymes *MspI* and *MseI*, which together have an even representation of all bases (CCGG and TTAA respectively) compared to *EcoRI* and *MseI* (*EcoRI* restriction site: GAATTC). A bias towards AT or GC rich regions should yield shorter DNA fragments, as seen with the use of *EcoRI* despite having a longer restriction site.

No enzyme choice is completely unbiased however, and examination of fragment distributions on a chromosomal level from *SbfI-MseI* indicates low coverage of the Y chromosome. Some sex-linked loci could therefore be excluded from the libraries and potentially limit the data to non-sexually divergent loci. Though it must be noted the significance of the role of sex chromosomes in sexually dimorphic phenotypes is the subject of much discussion [82, 206, 207], and the effect on downstream analyses of excluding loci found on sex chromosomes is currently unknown.

2.4.2 Accounting for genotyping error rate

As restriction sites are broadly shared within species, RAD-seq should theoretically yield the same loci among all samples in a population. However, a growing body of evidence suggests that genotyping error can affect RAD-seq in particular, thus questioning its utility in evolutionary and ecological studies [11, 272]. Genotyping error is known to directly affect downstream analyses in studies of selection [324, 349], inbreeding [122] and population structure and demography [231, 267]. It is therefore imperative that researchers report

genotyping error rates, and account for them in downstream analyses either directly, or by calibrating genotype likelihoods of each SNP [38, 217]. However, this is seldom done in large studies conducted over multiple sequencing lanes. Including replicate samples independently prepared for sequencing has been shown as a simple way to estimate genotyping error rates in RAD-seq studies [38, 217]. Here, these replicates were included not only within, but also between sequencing lanes, and show that although per allele genotyping error rates were low overall, significant variation could be detected among independently prepared libraries. Sequencing lane three in particular appeared to be significantly affected.

Genotyping error can occur due to allele dropout, the failure to amplify one or both alleles at a given locus during a PCR reaction. This can be the result of polymorphisms within the restriction site, and as a consequence restriction endonucleases can no longer recognise it. This allele may not be flanked by another restriction site nearby, and will therefore "drop out" of library preparation, thus causing excess homozygosity at that locus in the population. This type of error affects ddRAD-seq more than its single-digest counterpart, as the protocol requires two restriction enzymes, and the probability of point mutations in restriction sites increases the longer their cumulative length [112]. This also means that restriction enzymes such as *SbfI* with longer recognition sequences, should be disproportionately affected by this phenomenon [197]. However, this should affect all sequencing lanes equally as the same enzymes were used throughout this study, and cannot completely explain why one sequencing lane appeared to have a significantly higher error rate than the others. Furthermore, any error caused by a longer restriction site should be offset by their lower frequencies in the genome.

Amplification bias towards shorter fragments could also cause allele drop out due to a strong positive correlation between read depth and fragment length [79]. This means alleles would drop out not as a result of polymorphism in the restriction site, but rather due to low frequencies in the population resulting in a loss of alleles during quality filtering. More likely, however, is drop out due to non-standard library preparation procedures such as different quantification methods following genomic DNA extraction. Here, all sequencing lanes, apart from lane three, were quantified in a FluoSTAR Optima fluorimetric plate reader allowing 96 samples to be quantified together. Sample DNA concentrations were therefore estimated

based on a standard curve generated from eight DNA standards of known concentrations; these samples were then sequenced on the same lane. Variation in the normalisation of DNA was therefore minimised, and any errors will affect each sample in the same lane similarly as the precise point at which the standard curve loses linearity can be pinpointed, and sample concentrations are more accurately estimated (figure 2.14).

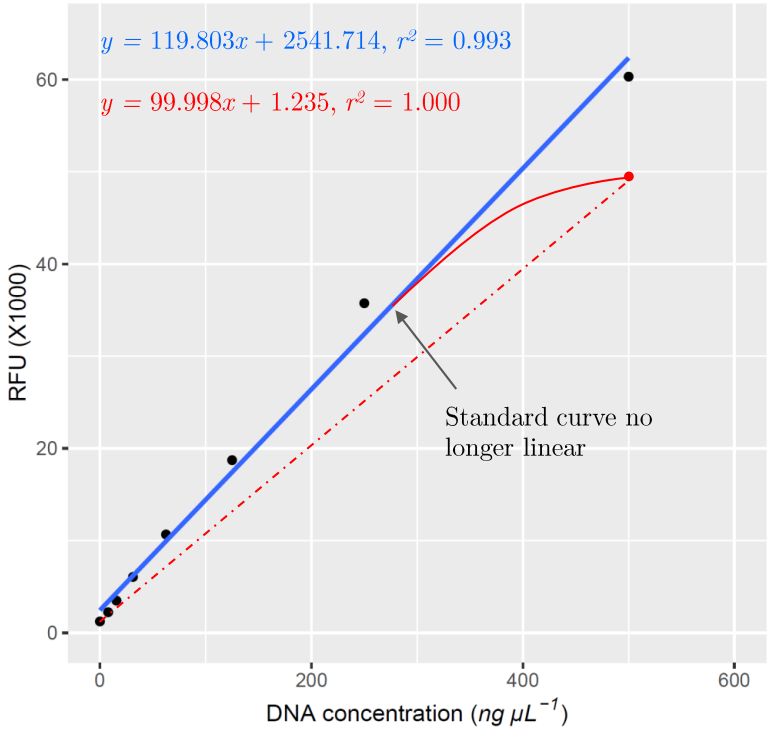


Fig. 2.14 A reliable standard curve is estimated from 8 standards (blue) with an $r^2 > 0.99$. Linearity is lost if the fluorophores in the reagent become saturated with DNA (red, solid) and a linear model can no longer accurately estimate the concentration past the point of inflextion. The highest standard must therefore be removed to improve the fit of the model and any samples estimated higher than this new maximum must be diluted and re-quantified. As the Qubit fluorometer estimates concentrations based on a standard curve generated only from the lowest and highest standards (red, dashed), linearity is assumed along the entire length of the curve and the model no longer accurately predicts the true concentration. The high r^2 is misleading in this case.

Unlike the other lanes, samples sequenced in lane three were quantified individually on a Qubit 3.0 fluorometer due to the unavailability of a plate reader. As the Qubit only requires two DNA standards, each at the limits of detection for the assay to generate the standard curve, linearity in the fluorescence of the samples is therefore assumed up to the theoretical

maximum. This can introduce a significant amount of variation in measurements of DNA quantity, and therefore leads to a higher probability of allele dropout by unequal amplification of DNA from differently represented samples [12, 11, 75]. 30% of samples in lane three have a mean sequencing coverage <10X likely due to this variation, and can be discarded so as not to affect downstream analyses. Though this is not recommended, and should instead be accounted for later by, for example, re-calibrating genotype likelihoods [38]. Furthermore, despite being significantly higher in one lane than the others, the overall mean genotyping error rate reported here is still an order of magnitude lower than reported by Mastretta-Yanes et al. [217] in their study quantifying genotyping error rates in RAD-seq, and are still within the estimates of a previous ddRAD-seq study on *Apodemus* by Martin Cerezo et al. [53].

2.4.3 STACKS parameter optimisation helps separate real biological signals from signals caused by genotyping error

Once sequencing data is obtained in any RAD-seq experiment, stringent bioinformatic processing is essential to ensuring the accuracy of SNP calls and detecting real biological signals. STACKS has become a popular choice for RAD-seq data processing and analysis, due to its versatility and relatively user friendly interface [49, 50]. After raw sequencing reads are demultiplexed and have passed quality control procedures (e.g. identifying and removing PCR duplicates, and reads with low quality or mismatched barcodes), the first stage of processing the data involves interrogating the *m, M* and *n* parameters of `denovo_map.pl`, which is dataset-specific [86, 256]. If sub-optimally specified, incorrect population genomic and evolutionary conclusions can be drawn due to genotyping errors. For example, both Díaz-Arce and Rodríguez-Ezpeleta [86] and Mastretta-Yanes et al. [217] found lower F_{ST} values for higher values of *m*. Here, the optimal parameter set: *m3, M2* and *n2*, were found by iterating over parameter values and plotting the various associated metrics as recommended by Paris et al. [256].

The *m* parameter, which controls the minimum number of reads required to collapse into putative alleles, must be assessed by observing its effect on sequencing coverage, and not

promoting too many error reads to putative alleles. With careful mediation, few true alleles will be excluded [256]. However, as some genotyping error is unavoidable, the promotion of some erroneous reads to putative loci will inevitably occur, particularly if sequencing coverage is low. Erroneous alleles can instead be excluded downstream as they should appear in low frequency if genotyping error is low, and *m* has been assessed properly. Paris et al. [256] therefore suggest that higher values of *m* should be used if coverage is <15x. Although 15X coverage can be sufficient to reliably call genotypes in a population [312], higher coverage would increase confidence in the genotypes and improve accuracy of variant calls. *m3* was selected during optimisation as the maximum number of *r80* SNPs is reached at this point, and the *r80* polymorphic loci decrease sharply at *m4*. More reads are thus collapsed into putative loci, and fewer SNPs are shared among individuals (rare alleles are less likely to be shared among a large proportion of the population) [86].

The *M* parameter dramatically affects the level of polymorphism detected in the population as more mismatches are allowed between alleles to form putative loci. The number of polymorphic loci quickly asymptotes at *M3*, and the majority of *r80* polymorphic loci and SNPs are already captured by *M2* indicating a low level of polymorphism in the population. This could be due to high levels of gene flow into the Białowieża population resulting in many alleles persisting at low frequencies, and is consistent with previous research which found differentiation is generally low among *A. flavicollis* populations across Poland [53, 73, 123]. This is further supported by the low levels of pairwise relatedness among captured mice from Białowieża National Park found here.

Loci also begin to over-merge at higher values of *M* when in close sequence space, and explains why the maximum number of *r80* SNPs occurs at *M3*, whereas the maximum number of *r80* polymorphic loci are detected at *M2*. As no new polymorphic loci are added after *M2*, it was selected as the optimum value. Increasing *M* beyond this point would not yield more biologically useful information, and the additional SNPs at *M3*, if not the result of genotyping error, would otherwise be filtered downstream anyway due to low minor allele frequencies or high statistical association (r^2) with another nearby allele.

2.4.4 Minimising sources of confounding

Confounding factors in genomics datasets such as high relatedness, duplicated samples and SNPs in linkage disequilibrium (LD), must be accounted for in many downstream analyses. As many of these analyses rely on the assumption of independence (of SNPs or samples), high levels of relatedness and LD can mistakenly introduce population stratification which can erroneously indicate selection in evolutionary research [342]. Variant data in this study was therefore stringently quality controlled to minimise possible sources of bias.

2.4.4.1 Filtering SNPs in population genomic analyses

SNP based heritability is often calculated assuming SNPs are independent of one another, i.e. SNPs are not in LD. However, estimating LD requires a reference genome. As assembly was conducted *de novo* here, a statistical proxy for LD was instead used by calculating r^2 between all pairs of variants. Few SNPs already filtered for MAF (< 0.05) were found to be statistically associated, indicating the majority are independent and should not create any false signals of population structure.

Sequencing errors can also introduce additional variation in a population in the form of low frequency SNPs, which can strongly affect population genomics analyses [194]. In small datasets ($\leqslant$ 1000 samples), these errors are often indistinguishable from true minor alleles, and thresholds are therefore often applied to filter these low frequency variants (MAF threshold = 0.01 - 0.1) when estimating heritability or relatedness. True low frequency variants have been attributed to the *missing* heritability of a phenotype, where thousands of minor alleles are expected to additively contribute to the phenotypic variation in a population [335, 352]. Filtering these SNPs is therefore likely to affect estimates of heritability in a downwards bias. However, increasing the power to detect this missing heritability would require dense genotyping data ($\geqslant 10^6$ SNPs) and a large population ($\geqslant$ 1000 samples), which allows for the use of weighted genomic relatedness matrices to account for linkage disequilibrium between alleles which no longer need to be excluded due to a low MAF [105, 183, 314, 335]. As such a large dataset in *A. flavicollis* is unavailable for this study, a more conservative MAF threshold of 0.05 was selected to minimise any bias in downstream

analyses as a result of sequencing error. Furthermore, the genotyping error rate here was calculated as 0.02 ($sd = 0.076$), and this threshold should sufficiently exclude any minor alleles likely to be caused by sequencing error. An even more conservative threshold (> 0.05) is unlikely to reduce noise and increase power in population genomics studies [348], unless the genotyping error rate is estimated as particularly high ($\geqslant 0.05$).

2.4.4.2 Identifying duplicated samples

Duplicated samples were identified by calculating the proportion of SNPs that are IBD, and estimating pairwise relatedness in the population. Four algorithms were used here: MLE, KING MoM, PLINK MoM and GCTA. Relatedness for duplicated samples should theoretically be one (i.e. 100% of SNPs are identical by descent with oneself). As discussed previously however, allele dropout caused by variation in library preparation procedures and sequencing coverage, can result in some variation in the estimates of relatedness. Furthermore, high levels of inbreeding in the population can also result in high relatedness (>50%), making the cutoff for deciding which samples are duplicates somewhat arbitrary. However, when plotting relatedness estimates and correlating the different methods, a small and tight cluster of samples segregated from the rest of the population, the majority of which were related by more than 90%. The exception was four pairs of samples between 0.75-0.9, as estimated by GCTA. These four pairs were estimated as >90% related by MLE, PLINK MoM and KING MoM however, and the lower values appear to be an artefact of the way GCTA estimates relatedness. All other duplicates in all four methods correlate strongly. This congruence among independent estimates of relatedness between maximum likelihood based methods (MLE and GCTA), and method of moments estimators (PLINK and KING), suggest signals of relatedness are real, and first, second and third degree relatives are detectable in the population. This also indicates that the genomic data obtained from quaddRAD can reliably support pedigree-based analyses, estimates of heritability and genome-wide association studies [350, 351]. However, life history information such as age, sex or birth years for each mouse when constructing pedigrees would be necessary to more reliably identify the precise nature of these relationships (e.g. parent-offspring, siblings, grandparent-grandchild) [142].

Although all four sets of estimates correlate strongly, the correlation of KING MoM with the other algorithms is considerably lower. This is because relatedness estimates by KING MoM are not bounded between zero and one, and unrelated individuals are more likely to be assigned a relatedness < 0 (figure 2.13). GCTA also allows relatedness < 0 (and also > 1), and is because the program internally rescales estimates of relatedness. Very low relatedness estimated by GCTA, despite being less than zero, does not exceed -0.081, more similar to MLE and PLINK MoM. Despite these scaling differences, high correlation indicates that this should have little downstream effect [96]. Previous studies however, have shown relatedness estimates from PLINK MoM to perform poorly compared to KING MoM when used with very large (>100,000 SNPs) genome-wide datasets, and led to inflation of *p*-values during association testing [96, 205]. Alternatively, maximum likelihood based methods are considered more robust as they remain unbiased, consistent (the parameter estimate, $\hat{\theta}$, converges to the true value as the sample size $\rightarrow \infty$) and efficient (low mean square error) for large sample sizes [232, 288], though it is far more computationally intensive [354]. As datasets increase in size, computational efficiency should therefore be considered when choosing appropriate methods to estimate relatedness and identify duplicate samples.

Chapter 3

Population demography, fitness and allele frequency dynamics in a pedigreed, wild population of *Apodemus flavicollis*

3.1 Introduction

Winters in Białowieża, Poland can be particularly harsh for small rodents like *Apodemus flavicollis*, as mean daily temperatures are often below freezing and mean daily snow cover can be over $10cm$ deep [269]. Thus, food availability is severely limited during winter, and causes distinctive annual cycles of population growth and decline which closely follow patterns in seedfall [269].

Previous studies estimate the *A. flavicollis* (sub-)population in Białowieża National Park declines by as much as 86% during the winter due to the severity of the climate [269]. This should result in high levels of inbreeding and low genetic diversity, and the (sub-) population should therefore be particularly vulnerable to the effects of drift-driven genetic differentiation, where alleles are more likely to be lost or fixed, and homozygosity to be high relative to the whole Polish population [54]. However, more recent studies indicate very little differentiation between different Polish sub-populations including Białowieża, where pairwise F_{ST}, a measure of genetic differentiation, was measured between 0.007- 0.086 even

with hundreds of kilometres separating sampling sites [53, 73]. Given the annual cycles of growth and decline, it is intriguing how such little population structure is maintained.

Czarnomska et al. [73] have shown that *A. flavicollis* (sub-)populations from northern Poland exhibit high levels of spatial auto-correlation with increased abundance, even as the sampling distance from the point of origin increases. That is, as abundance increases, high levels of genetic similarity can still be detected further away from the origin relative to when abundance is low. This suggests mice are migrating or increasing their range more when conditions are favourable, as local resources become limited by increasing local population density. The probability of gene flow from nearby (sub-)populations to a local population therefore increases, and could be one mechanism that allows the maintenance of high genetic diversity. However, studies of population structure such as by Czarnomska et al., often rely on the use of small numbers of low density markers such as microsatellites. Although such markers perform well for estimating gene flow, population differentiation and relatedness [28, 103, 199, 259, 287], they are limited in their use for real-time, high-resolution, genome-wide studies of genomic variability within a population, and for detecting selection [184, 236, 321, 357]. They are in essence, single point in time studies, rather than truly longitudinal.

Only recently, have more fine scale population level studies of genetic drift, gene-flow and selection become more accessible researchers studying natural populations over multiple generations. For example, with the use of RAD-seq, Fitzpatrick et al.'s recent study on the Trinidadian Guppy (*Poecilia reticulata*) [102], showed using translocation experiments, how gene-flow from a source population can increase individual fitness by heterosis without erosion of local adaptive variation in the sink population (outbreeding depression). Chen et al. [60], also using RAD-seq, showed in a wild population of Florida Scrub-Jays (*Aphelocoma coerulescens*), that the overall genetic contribution of immigrants to a population is large, and the role of selection in short-term allele frequency dynamics in open and natural systems is restricted to a few loci, with allele frequency variation largely driven by variability in survival and reproductive success, with smaller contributions from gene flow over time.

Studies such as these however, often rely on human mediated manipulation of natural systems to conduct experiments, or observe monogamous and/or diurnal species which have long generation times that do not overlap. This makes mark-recapture experiments and estimating relatedness simpler. For more typically r-selected species with short generation times that do overlap, and are also cyclic such as *A. flavicollis*, the applicability of their findings are somewhat limited.

Here I use the set of independent genome-wide SNPs in a wild population of *A. flavicollis* from Białowieża National Park obtained using the quaddRAD protocol described in chapter 2, to estimate relatedness in the population, and construct wild pedigrees to describe the population in terms of its demography and the fitness of each individual. I then track allele frequencies across multiple cohorts to assess the effects of evolutionary forces in a wild population that is vulnerable to significant fluctuations in size, uncover levels of genetic diversity, and show how the population's genomic structure changes over short timescales at a high resolution.

3.1.1 Constructing wild pedigrees

Tracing an individuals genealogy by constructing a pedigree has long been recognised by biologists for its utility to understand patterns of inbreeding, reproductive success, inheritance and describing population demography. Early pedigrees of wild populations however, were often limited to easily observable species where individuals and their offspring could be identified in the field [41]. For wild study systems, in which species may have a long lifespan with late-onset breeding, are nocturnal, elusive, particularly rare, or express extra pair and multiple mating strategies, the traditional method for pedigree construction using field observations alone often results in mis-assigned parentage which can affect downstream analyses such as estimating the heritability of traits [56]. Major developments in DNA sequencing technologies however, have provided a viable alternative to support field observations with molecular evidence, and generate robust pedigrees which accurately reflect the demographic structure of wild populations [103, 259]. Pedigrees supported by genetic data have since been used to uncover levels of inbreeding [90, 159, 279], population demography [90], multiple

mating [279], hybridisation [273], disease transmission [1], and trait heritabilities in the wild [56, 334].

3.1.1.1 Marker selection for pedigree construction

At the core of all methods to assign parentage and construct pedigrees, lie the principles of Mendelian inheritance [103], where in diploid species, offspring inherit one allele from each parent at every locus. Thus, progeny are related to each parent by a mean of 50% (with some notable exceptions such as in populations with high levels of inbreeding [340]). Multiple independent, polymorphic, Mendelian loci in a sample of genotyped individuals, together with even a modest amount of life history information such as birth years, can then be used to identify first, second and third degree familial relationships among a sea of unrelated individuals. This strategy relies on identifying what proportion of the genome is identical by descent [103, 156, 259]. Early approaches to molecular parentage assignment involved polymorphic allozyme markers: large proteins which are polymorphic in their size and charge, but not their function. In principle, they should therefore vary in their mobility through a gel matrix, and can be separated electrophoretically to determine parentage [103]. In practice however, low variability often prevented reliable parentage assignment using these techniques, and the number of studies using pedigrees stagnated until the discovery of hypervariable, codominant markers such as microsatellites [103, 157].

Microsatellites are short stretches of tandemly repetitive DNA occurring throughout the genome [326]. They have high mutation rates compared to other genomic regions [39], and can be used to identify closely related individuals based on the proportion of alleles that are identical by descent. They can be easily amplified from poor quality or low quantities of DNA, and so quickly became the marker of choice for parentage assignment and pedigree construction, particularly in fields such as conservation biology and historical DNA analysis, where genetic material is valued for its rarity [339]. Despite their versatility, microsatellites have some notable drawbacks for pedigree analyses. For example, their use was highly suited to species in which they were abundant and are highly polymorphic, such as in many species of fish [84, 248]. Unfortunately, many other species lack the levels of

polymorphism necessary to discern familial relationships even using these hypervariable loci [259]. Furthermore, developing primers to sequence microsatellite markers in non-model organisms with few available genomic resources, can be laborious and expensive [103, 138].

Alternative markers such as single nucleotide polymorphisms (SNPs) are now gaining favour among biologists for pedigree based analyses [9]. These are single nucleotide differences in the genome present in a significant proportion of the population. Although the power of each individual SNP to determine familial relationships is low compared to a single microsatellite marker, their high abundance throughout the genome mean a few hundred SNPs have enough resolving power to discriminate first, second and third degree relatives, equalling and in some cases outperforming microsatellites in terms of accuracy [10, 103, 330]. Modern high-throughput sequencing technologies have now made SNP discovery a common process [210], and are widely used in ecological and evolutionary studies [182, 235]. Although some methods of developing a SNP panel for parentage analysis such as SNP arrays still require significant investment to develop, reducing genome complexity and multiplexing large numbers of samples, as with RAD-seq for example, has improved the accessibility of genome-wide SNP markers, even in non model organisms, as they require no *a priori* knowledge of the genome [10, 103, 157]. SNPs obtained through reduced representation approaches such as RAD-seq are therefore an appealing choice of marker for pedigree analyses in large and wild populations.

3.1.1.2 Methodological considerations for pedigree construction

Methods to construct pedigrees from genome-wide markers encompass three main categories: exclusion-based methods, kinship-based methods, and likelihood-based methods. As codominant markers such as SNPs and microsatellites follow the Medelian laws of inheritance, it seems intuitive that the list of candidate parents in a population can be refined by simply excluding samples which do not share any loci with a focal individual. This is a fast method of assigning parentage that has the potential for high accuracy, as found by Melo and Hale [226]. However, the authors also acknowledged significant *a priori* life history information is still necessary to improve accuracy further, and the method makes a number of specific

assumptions about the study system. For example, exclusion based methods assume no genotyping error or mutations in the markers. This is less of a problem in a few hypervariable sites such as microsatellites, but the growing use of large numbers of SNPs and the increasing size of modern datasets means this assumption is no longer valid [103]. To account for this, a number of mismatches can instead be specified when excluding potential parents. Though the threshold in the number of mismatches allowed is arbitrary, and implicitly assumes a level of confidence in the genotype accuracy. Furthermore, when used with biallelic SNPs, only homozygous genotypes are used to assign parentage, and a significant number of informative markers for pedigree construction are discarded which could otherwise improve accuracy [103, 209].

Kinship-based methods are an alternative to exclusion-based methods, which work extremely well to discriminate clones, and first and second degree relatives from unrelated individuals (as in chapter 2, used to identify duplicate samples, as well as a number of other studies [303, 315]). However, it is common in wild populations for generations to overlap, and for individuals to be inbred [152]. Further post-hoc analyses, such as calculating Cotterman coefficients, which provide the pairwise probability of zero, one or two alleles at a locus being identical by descent, is then required to distinguish parent-offspring pairs and full siblings, which are both related by 50%. However, this still cannot differentiate between tertiary and avuncular relationships (grandparents, half siblings and aunts or uncles), which are all related by 25%.

Likelihood-based methods provide a significantly more powerful way to construct pedigrees, albeit at the cost of computational efficiency [8, 156, 209, 329]. This method relies on calculating the likelihood of a proposed relationship as a probability, and is based on the observed proportion of genotypes that are identical by descent. Unlike exclusion-based methods, this approach also utilises heterozygous genotypes, which would otherwise have been discarded. The absolute likelihood value provides little information on its own however, and must instead be considered relative to values estimated for alternative relationships. For example, if sample A is proposed as the sire of sample C with a likelihood x, then x must be compared to the probability of the alternative sire D, with likelihood y. By calculating the

likelihood ratio $\Lambda_{AC/DC} = \frac{x}{y}$, a more positive value of $\Lambda_{AC/DC}$ would indicate *A*-*C* is the most likely sire-offspring pair, and can be assigned to the pedigree.

Likelihood-based pedigree construction and parentage assignment has been implemented in several software packages such as COLONY [158], FRANZ [281] and SEQUOIA to name a few (further examples can be found in a recent review by Flanagan and Jones [103]). SEQUOIA offers an advantage over the other software as it gives explicit consideration to a much wider range of possible relationships, including secondary, tertiary and avunculars. This consideration provides the added benefit of allowing to join disconnected clusters of (half-)siblings, which have one or more unsampled parents, by filling in missing pedigree links with *dummy* parents that are unsampled themselves, but are predicted to exist in the population based on these clusters and the presence of their grandparents. Thus, the effect of missing pedigree links on population genomic parameters such as inbreeding coefficients and fitness, are limited [327].

3.1.2 A note on reproductive success as a measure of fitness

The goal of population genetics research is to understand heritable differences in allelic variation over time, at the population level. Ultimately, the forces that drive the evolution of a population include:

i Natural selection, where traits increasing fitness (and the alleles responsible) also increase in frequency (e.g. bright plumage and assortative mating),

ii Genetic drift, where stochasticity in the population causes allele frequencies to vary due to sampling effects, or

iii Gene flow, where migrants in a population introduce new alleles and increase the level of standing genetic variation in the population.

In a theoretical model population where drift and migration are absent, population evolution is ultimately driven by the variation in fitness [251], and pedigrees allow us to estimate the fitness of each individual, and track their genetic contribution to subsequent generations.

However, fitness is a difficult parameter to measure, and is often estimated via proxies such as reproductive success.

Although fitness is, in its most crude definition, a measure of reproductive success, reproductive success is often a poor reflection of fitness [110]. This is because individuals can vary in their *reproductive quality*. For example, consider two focal females, A and B (figure 3.1). In a female-biased population, the probability of successful reproduction is male limited due to their rarity, and if Female A produces only male offspring during her reproductive lifespan, she will likely have more grand-offspring than Female B, which produces the same number of only female offspring. Female A's male offspring will therefore have higher reproductive success than Female B's female offspring, thus giving them a higher reproductive value that should be favoured by natural selection. Taken as a measure of reproductive success however, both Female A and B will have the same fitness, despite the difference in the number of genealogical descendants and reproductive quality. R. A. Fisher first noticed this limitation in his theories on fitness, and proposed an alternative which accounts for not only the quality of the individual, but also of the genealogical descendants [124]. He defined fitness not by reproductive success, but rather the genetic contribution of the focal individual to future generations, thus reflecting differences in reproductive quality well. In this study, three measures of fitness will be declared: the lifetime reproductive success of each individual, and genealogical and genetic contributions of founders to the population. This is because founders caught later in the study are expected to have fewer captured genealogical descendants, and together, these three measures will provide a good overview of fitness within the population.

3.2 Materials and methods

Inferring pedigrees in a natural population with SNPs requires the markers to be rigorously filtered to ensure their independence. In this study, variant data as SNPs in a *.vcf* file were obtained and filtered using the methods described in chapter 2.2, to estimate relatedness and generate pedigrees in a wild population. Briefly, sequencing libraries prepared from

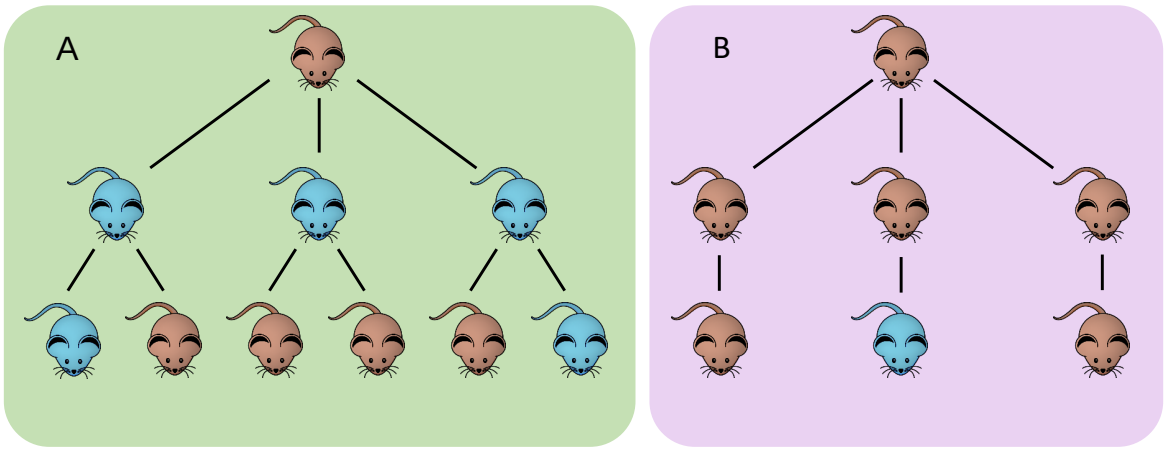


Fig. 3.1 A hypothetical example of how two females with the same reproductive success can have different fitness based on its definition. In a female-biased population, Female A has three male offspring (males shown in blue) and Female B has three female offspring (females shown in red). Both Females A and B have the same fitness when measured as reproductive success, regardless of the quality of the offspring (quality measured as the number of grand-offspring of Female A or B). Measured as the genetic contribution to the population however, Female A has higher fitness as she is likely to have more grand-offspring.

A. flavicollis tail or ear clippings captured in Białowieża National Park, were generated with a modified *quaddRAD* protocol [106]. Sequencing data was demultiplexed using the `process_radtags` module of STACKS V2.3d, and PCR duplicates were removed using `clone_filter`. Samples were then genotyped using the `denovo_map.pl` pipeline using the optimised parameters *m3*, *M2* and *n2*. Unlike many population genomic analyses, thousands of SNPs are not necessary when estimating pedigrees as this can be computationally prohibitive [142]. Between 300-700 independent and high minor allele frequency (MAF) SNPs are recommended for pedigree reconstruction [142], and were obtained by filtering for MAF > 0.45 using the `populations` module of STACKS. SNPs had already been filtered for HWE ($p \leq 0.05$) and linkage disequilibrium (r^2) using PLINK, to ensure SNPs were independent and not an artefact of sequencing error. This ensures no additional structure is introduced into the population. As recommended by Huisman [142], the value for the minimum MAF was decided upon by adjusting the MAF threshold until the number of SNPs passing the filter fall between the recommended range of 300-700 SNPs. This still produced far more than the required number of SNPs however, so 400 of the least statistically correlated SNP

pairs, assumed to come from un-linked loci (lowest r^2), were used to give 701 unique SNPs for pedigree reconstruction.

3.2.1 Pedigree reconstruction

The R package SEQUOIA V2.0.3 [142], which was specifically designed for SNP data, was used to construct pedigrees as it utilises a fast, heuristic hill-climbing algorithm to rule out relationships that are impossible, early in the assignment process. For example, potential offspring must be born after the potential parents. Excluding individuals which do not meet this criterion significantly speeds up computations to maximise the likelihood of each relationship [142]. SEQUOIA also considers multiple generations at once. This is important in wild populations such as *A. flavicollis*, where generations could overlap, and reduces the need to declare candidate parents *a priori*. Though they can be implicitly assumed as part of the exclusion process of SEQUOIA's algorithm, when birth years and sex are provided. This is particularly useful when, as in the case of small, wild and largely nocturnal species like *A. flavicollis*, field observations are limited to support pedigree reconstruction.

To ensure the most reliable pedigrees are constructed, birth years were assigned to each mouse based on their estimated life-stage during field observations. For example, we know the life-span of wild *A. flavicollis* is approximately 12 months [269]. An adult mouse caught in the spring of 2016 could then be estimated as being born in 2015. A juvenile mouse also caught in the spring of 2016 however, could be born in early 2016 or late 2015 (though winter births are less common [269]). In this case, as we know the approximate life stage, we can say the juvenile is born in 2016. Even if incorrect, it ensures the juvenile is placed in the birth cohort after the adult mouse caught in the same season. Thus, the juvenile mouse cannot become a potential parent of the adult mouse, and will be excluded by SEQUOIA. This also allows for the use of samples with some uncertainty in the estimated life stage (maybe adult or juvenile), to be assigned a birth year. As the speed and accuracy of the computations relies on the exclusion of such impossible relationships, as much *a priori* information as possible must be provided to calculate maximum likelihoods, and minimise errors in the

final pedigree [142, 226]. A full breakdown of genotyped mice used to create the pedigree is shown in table 3.1.

Birth year	Adult	Maybe adult	Juvenile	Maybe juvenile	Unknown age	Males	Females	Unknown sex
2015	52	28	2	1	0	54	29	0
2016	59	29	80	32	0	107	93	0
2017	2	1	63	14	0	52	28	0
2018	0	0	0	2	0	1	1	0
Unknown	0	0	2	1	224	111	101	15
Total	113	58	147	50	224	325	252	15

Table 3.1 Breakdown of genotyped mice used to generate the SEQUOIA pedigree, that were trapped within the 0.9 ha trapping site inside the Strict Reserve of Białowieża National Park.

701 filtered SNPs from the *.vcf* file output by the `populations` module of STACKS, was first converted into a *.raw* format in PLINK using the `--recodeA` modifier for input into SEQUOIA. The function `sequoia` was then initially run in R with `MaxSibIter=0` (maximum number of iterations for sibship clustering). This is so sibship clusters are not assigned at first, allowing to visually inspect the resulting parents and ensure no obviously incorrect relationships were found. `Sequoia` was then run again with `MaxSibIter=40`, to instruct the package to give explicit consideration to sibship clusters, and assign secondary and avuncular relationships with a maximum of 40 iterations when maximising the likelihood. Genotyping error rate was specified as `Err=0.020`, as calculated in chapter 2.2.3.4, and all other parameters were kept as default. The distribution of Mendelian errors, which describe alleles present in an individual that cannot be inherited from either assigned parent, were then estimated for each SNP using the `SnpStats` function, and the final pedigree was plotted using the `plot.pedigree` function of the KINSHIP2 V1.8.5 package.

3.2.2 Pedigree accuracy

As a pedigree based on field observations is unavailable to compare with SEQUOIA'S pedigree, two alternative methods have been used to assess its accuracy. Firstly, pairwise genomic relatedness, calculated as the proportion of SNPs that are identical by descent in chapter

2.2.3.3, using all individuals and filtered SNPs (n = 592 after merging duplicates and 21,011 SNPs respectively), was regressed against the pairwise pedigree relatedness which was calculated according to the relationships constructed by SEQUOIA. Unlike pairwise genomic relatedness, pairwise pedigree relatedness is bound to 0.5, 0.25 and 0.125 for primary, secondary and tertiary relationships respectively. High correlation between these two measures should indicate a reliable pedigree has been generated, particularly as the pairwise genomic relatedness has already been used to successfully identify duplicated samples, and also confirmed first, second and third degree relationships are detectable within the population. To generate the pairwise pedigree relatedness matrix, the pedigree output by SEQUOIA was passed onto the `kinship` function of the KINSHIP2 V1.8.5 package in R, and multiplied by two.

Secondly, the `EstConf` function of SEQUOIA was used to calculate the confidence of each assigned dam and sire. 400 SNPs were sampled from the 701 SNPs provided to SEQUOIA using the `nSnp=400` argument, and pedigrees were simulated using this sample to calculate the assignment error rate by comparing each simulated pedigree to the original pedigree generated by the function `sequoia`. Assuming the original pedigree provided to `EstConf` is correct, overall confidence probabilities can then be calculated for dams and sires using: $(\text{number of mismatches between pedigrees} + \text{number of false positives})/2N$, where N is the number of individuals in the pedigree [87, 142]. 60 iterations were used to simulate pedigrees (`nsim=60`), and the default 40% of dams and sires were assumed unsampled (`ParMis=c(0.4,0.4)`). The genotyping error rate (`SnpErr`) was again assumed to be 0.020, and all other parameters remained default. All pedigree analyses were conducted in R V.3.6.3.

3.2.3 Estimating fitness

Fitness in this study was quantified in three ways:

- **Lifetime reproductive success** - the number of individuals caught, genotyped and assigned as an offspring of any focal mouse when constructing a pedigree.

- **Genealogical contribution of founders** - the proportion of pups in the population that are genealogically descended from a focal founder mouse.
- **Genetic contribution of founders** - the expected proportion of alleles in a given generation and locus, that when drawn at random, have a probability equal to the relatedness coefficient, of being identical by descent from the focal individual.

The genetic contribution, G, of a founder was calculated as:

$$G = \frac{1}{n} \sum_{m} \sum_{p} \left(\frac{1}{2} \right)^{g}, \tag{3.1}$$

where n is the number of pups born in a given generation, m is the number of mice in that generation that are related to the focal founder, p is the smallest number of paths in the pedigree connecting the pup to the focal founder, and g is the generational difference between the pup and the focal founder [60]. A focal mouse is any mouse in the population, for which any of the three above mentioned measures of fitness were estimated. A founder was defined as any individual in the pedigreed population who's birth parents are unknown. A founder's parents may be unknown because they exist in the population but were unsampled, or the founder was an immigrant from a nearby population.

3.2.4 Estimating inbreeding

Populations experiencing bottlenecks are expected to present high levels of inbreeding and low genetic diversity due to an excess of homozygotes [241]. Measuring inbreeding is therefore an important way to assess how a population responds to perturbation on a genomic scale, and can be defined as the probability that any two alleles sampled from two different individuals in a population are identical by descent (IBD). That is, that they are inherited from the same ancestor.

Many methods have been developed to measure inbreeding in a population [159]. Here, inbreeding was estimated as F_H, which measures the observed homozygous genotypes in individual i, relative to the mean expected homozygous genotypes in a population conforming to Hardy-Weinberg proportions (randomly mating population) [159]:

$$F_{H_i} = \frac{O(Hom_i) - E(Hom)}{m - E(Hom)}, \tag{3.2}$$

where $O(Hom_i)$ is the number of observed homozygous loci in individual i, and $E(Hom)$ is the mean number of homozygous genotypes expected under HWE from a total of m loci. In a population where individuals are mating randomly and conforms to HWE, $O(Hom_i) \approx E(Hom)$, and F_H will be close to 0. Excess homozygosity and high levels of inbreeding will be indicated by positive values up to 1, as observed homozygosity will be greater than expected, and negative values up to -1 suggest an excess of heterozygotes in the population indicating the population is outbred. F_H was measured using 21,011 SNP genotypes produced as a *.vcf* file by the `populations` module of the STACKS pipeline that have already been quality filtered using the methods described in section 3.2. PLINK was then used to estimate F_H for each mouse using the `--het` modifier. The resulting data was then analysed in R v4.0.3.

3.2.5 Allele frequency dynamics

Of the 592 genotyped samples in the population, only those with a known birth year (n = 366) were analysed for any observable generational trend in allele frequencies. This is to ensure the cohort to which each mouse belongs is known, and mice were selected regardless of whether they were retained in the pedigree or not. First, to see if there was a longitudinal pattern in allele frequencies, a linear model in the form $f_o \sim c + \varepsilon$ was run for each SNP, where f_o is the observed allele frequency, c is the cohort year, and ε the residuals. This model was then repeated for the change in observed allele frequencies, Δf_o, calculated as the difference in allele frequencies between each year, by substituting it for f_o as the response variable. All explanatory variables remained the same. The predicted values for each model were then correlated against the observed values, to see how well the models explain the variation in (the change in) allele frequencies among cohort years.

To test the significance of Δf_o, null allele frequency distributions were simulated using a custom R script (available on GitHub, please see appendix A.2 for details) to compare

the observed and expected (Δf_e) values. Firstly, alleles for every SNP present between 2015-2017 (n = 21,010) for each individual (n = 366), were sampled 1000 times from one of three possibilities: homozygous reference, homozygous alternate and heterozygous, with the probability of sampling given in table 3.2. These null distributions were then simulated 2800 times, and p-values for Δf_o were calculated by counting the number of simulations that Δf_o remained outside the 95^{th} percentile range of Δf_e. All analyses and simulations of allele frequency dynamics were conducted in R v4.0.3.

Zygosity	Sampling probability
Homozygous reference	$(1-f_o)^2$
Heterozygous	$2f_o(1-f_o)$
Homozygous alternate	${f_o}^2$

Table 3.2 The probability of sampling an allele when simulating allele frequency distributions, so p-values can be calculated to assess the significance of allele frequency change between 2015-2017.

3.3 Results

3.3.1 Pedigree reconstruction

Pedigree construction using SEQUOIA assigned 215 maternal and 223 paternal relationships overall, from three generations (figure 3.2, full pedigree available as a tabular format with log-likelihood ratios on GitHub, see appendix A.2 for details). This includes 83, 200, 80 and 2 genotyped mice born in 2015-2018 respectively, of which 151 were female, and 215 were male (a full breakdown of the genotyped samples used for pedigree reconstruction be found in table 3.1). Also shown in figure 3.2, is male Af_237 and female Af_258, which are two founders that made a large genealogical and genetic contribution to the population. These founders are shown here to highlight, relative to the rest of the population, when they were estimated to be born, and the number of offspring they produced. However, please see section 3.3.3 for full results on fitness estimates within the population, and a detailed figure of the respective pedigrees of these two founders showing their full genealogy (figure 3.6b).

During pedigree reconstruction, 191 parent offspring pairs remained unassigned by SEQUOIA due low confidence in the relationship, or due to missing age and birth year data. 65 *dummy* (non-genotyped) individuals were also assigned as individuals that SEQUOIA predicts to exist in the population based on sibship clusters (siblings clustered together but with one or both parents are unsampled, figure 3.3a), but are themselves unsampled. This included 30 *dummy* females and 35 *dummy* males. The number of full siblings per family varied between 1-9 ($mean = 2.08, sd = 1.34$), and the distribution is positively skewed suggesting litter sizes are either generally small, a large number of pups are not surviving long enough to be trapped, or they have migrated out of the local population covered by the trapping site. Half-sibling cluster sizes varied considerably more (1-27), though few clusters of genotyped or *dummy* individuals were composed of greater than 10 mice (figure 3.3b).

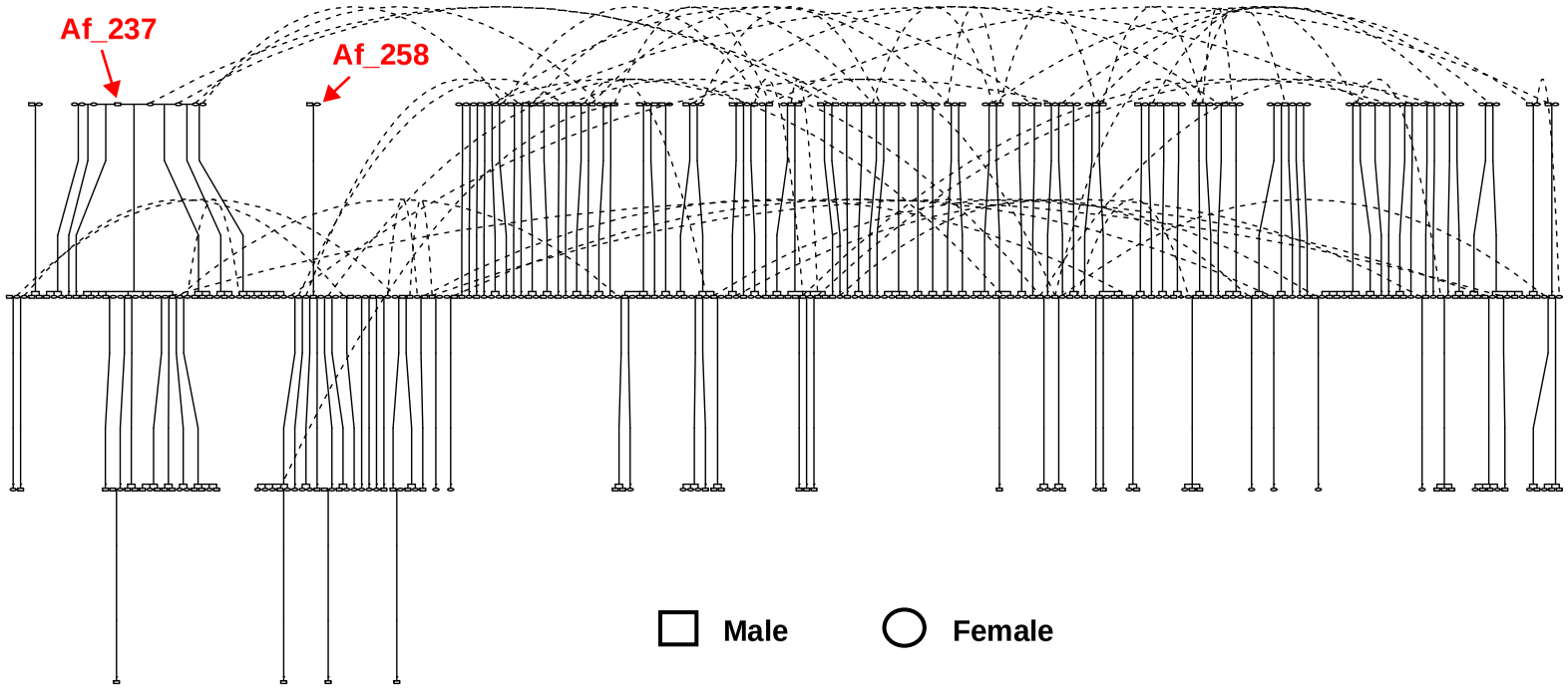


Fig. 3.2 The pedigree of the Białowieża population generated by SEQUOIA. Dashed lines indicate additional matings which resulted in offspring that were also captured. Squares and circles indicate males and females respectively. Only mice with known offspring are plotted for visualisation purposes. Male Af_237 and female Af_258 are shown here to indicate their position in the pedigree, relative to all other pedigreed mice in the population. See section 3.3.3 for more in depth results on the genealogy and fitness of these two founders. A .pdf version of the pedigree can be found on GitHub, which can be zoomed in to show more detail. See appendix A.2 for more information.

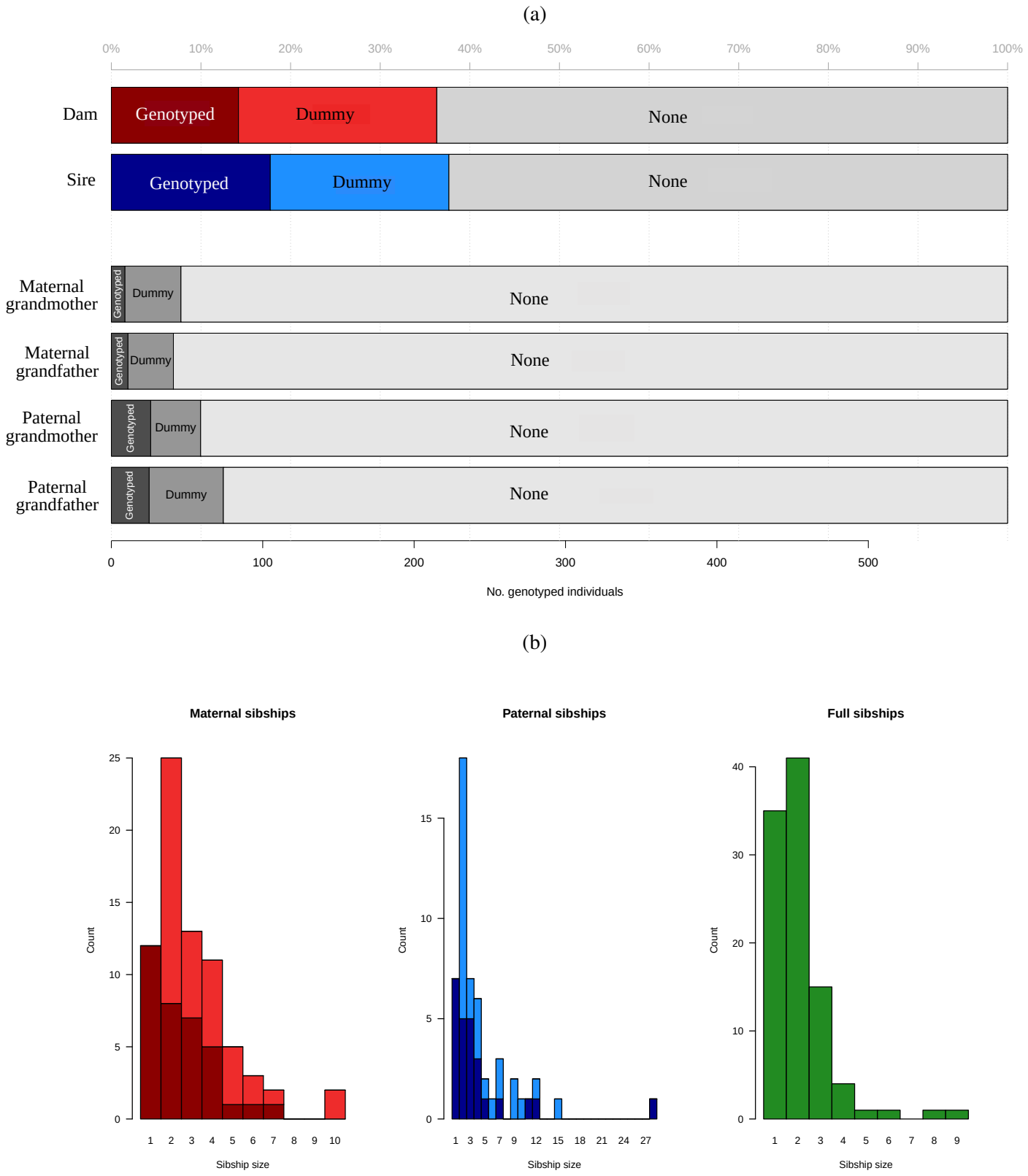


Fig. 3.3 a) The number of parental and grandparental relationships assigned to genotyped individuals. b) The size of full sibling (green), and maternal (red) and paternal (blue) half sibling clusters assigned during pedigree construction. Light red and light blue indicate the number of *dummy* half-siblings assigned.

3.3.2 Pedigree accuracy

Pairwise genomic relatedness (r_{grm}) estimated in section 2.2.3.3 was strongly correlated with pairwise pedigree relatedness (r_{ped}) as estimated by SEQUOIA (Pearson's $\rho = 0.744$, 95% $CI = 0.742$-0.746, $t_{157,120} = 441.24$, $p < 0.001$). This suggests the constructed pedigree serves as a strong proxy for the true pedigree relatedness, and the correlation could be higher still as it appears to be restricted by the large number of unrelated individuals in the population where pedigree relatedness equals zero, but genomic relatedness is much higher (figure 3.4). This is due to low confidence in some relationships which prevent these individuals from being assigned to the pedigree (see GitHub for the full pedigree and full table of possible but unassigned relatives, additional details in appendix A.2). Furthermore, patterns of Mendelian inheritance are random, so r_{grm} is expected to scatter considerably around r_{ped} as observed in figure 3.4. The proportion of $r_{grm} - r_{ped} > 0.2$ can therefore be indicative of pedigree error [142], and was low here (3.8×10^{-5}), thus supporting the findings of the correlation test.

Further to the above estimated pedigree error, the overall mean confidence probability in the assignments from 60 simulations of 400 sampled SNPs is high for both dams and sires (0.998 and 0.990 respectively), and the estimated assignment error rate is low ($\mu = 6.8 \times 10^{-4}$). Though, the error rate is higher for sires than dams ($\mu_{dam} = 7.6 \times 10^{-5}$ and $\mu_{sire} = 6.1 \times 10^{-4}$). However, this method of simulation to estimate assignment error rate and confidence probabilities assumes the pedigree used as a truth set to compare the simulations against, is independently generated, and free from, or has minimal, error (e.g. from reliable field observations or another independently generated genotype dataset). As both the truth set used here and the simulated pedigrees are generated from the same genotype data, this method is somewhat anti-conservative. Thus, assignment error is likely underestimated, and confidence probabilities overestimated. All three methods to assess pedigree accuracy, when taken together, do indicate the final pedigree generated by SEQUOIA is accurate.

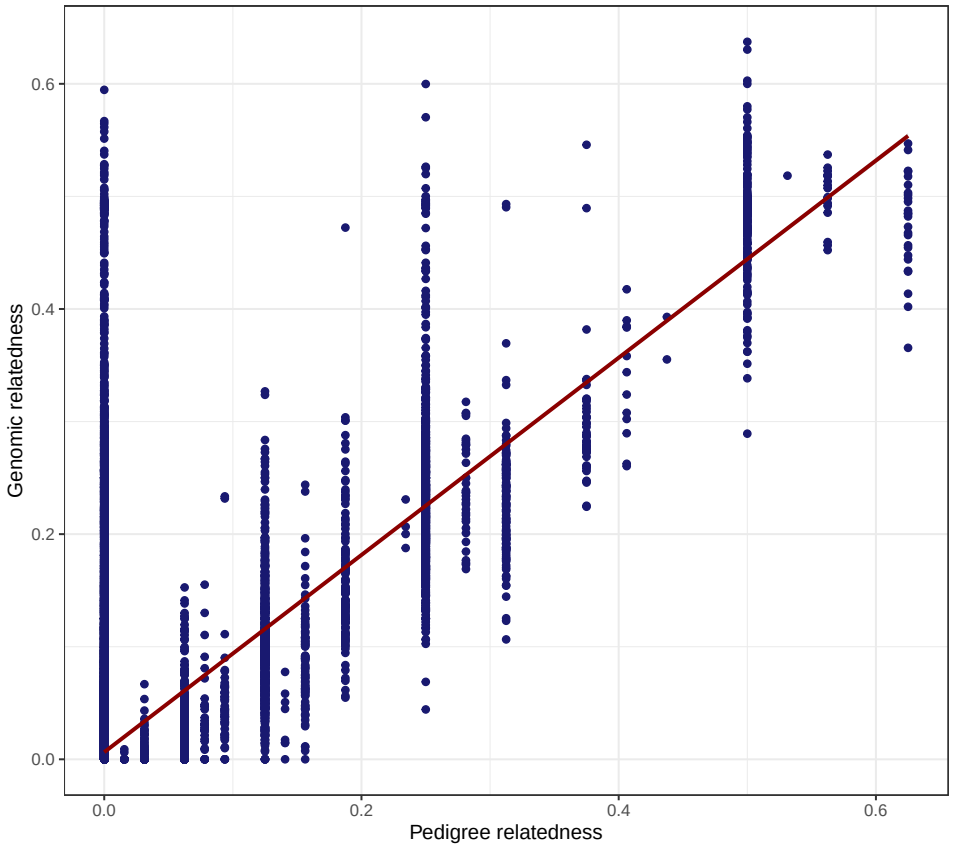


Fig. 3.4 Correlation between pairwise pedigree relatedness as estimated by SEQUOIA using 701 SNPs, and pairwise genomic relatedness estimated from 21,011 SNPs.

3.3.3 Estimates of fitness

Overall, fitness estimated as the lifetime reproductive success for each individual, varied greatly between 0 and 28 (figure 3.5a). Despite this large variation, the mean lifetime reproductive success was very low ($mean = 0.69, sd = 2.07$) due to the 191 possible parent-offspring pairs with low confidence or missing age and birth years (see GitHub for full table of possible but unassigned relatives, additional details in appendix A.2). When considering only the lifetime reproductive success of mice assigned to the pedigree with high confidence, the mean increased to 3.62 offspring per individual ($sd = 3.41$). However, this is still low considering litters can include up to 11 pups, and suggests many pups may not survive to adulthood. Fitness measured as the number of genealogical descendants of founders ($n = 133$), also varied considerably between 1 and 43 ($\mu = 4.20,\ sd = 5.35$, figure 3.5b). As expected, founders caught later in the study have fewer genealogical descendants that were trapped, so the genealogical and genetic contributions of founders to each generation were also calculated. These contributions correlated strongly (Pearson's $\rho = 0.889$, 95% $CI = 0.858$-0.912, $t_{224} = 29.6,\ p < 0.0001$). However, genealogical contributions are approximately twice as large as genetic contributions (paired Wilcoxon test: $V = 11,324,\ p < 0.0001$, median = 0.020 and 0.011 for genealogical and genetic contributions respectively). Figure 3.6a shows the genetic and genealogical contributions to each generation, from two founders to the population, and demonstrates how their genealogical contributions over four generations quickly overtake their genetic contributions. Male Af_237 in particular, is shown to make a particularly large contribution to the population, where 53.6% of individuals in the pedigreed population are descended from this one mouse, which has also contributed 20.1% of the total alleles in the population by the end of the study. In comparison, female Af_258, which despite 19.3% of the population being descended from her, has contributed just 3.8% of the total alleles in the population likely due to a low estimated lifetime reproductive success (one versus 26 for Af_237). A full genealogy of these two founders is shown in figure 3.6b, and a full table of genealogical and genetic contributions of founders to each generation can be found on GitHub (see appendix A.2 for details). For all three measures of fitness, no statistical difference between males or females was found ($p \geq 0.092$, table 3.3).

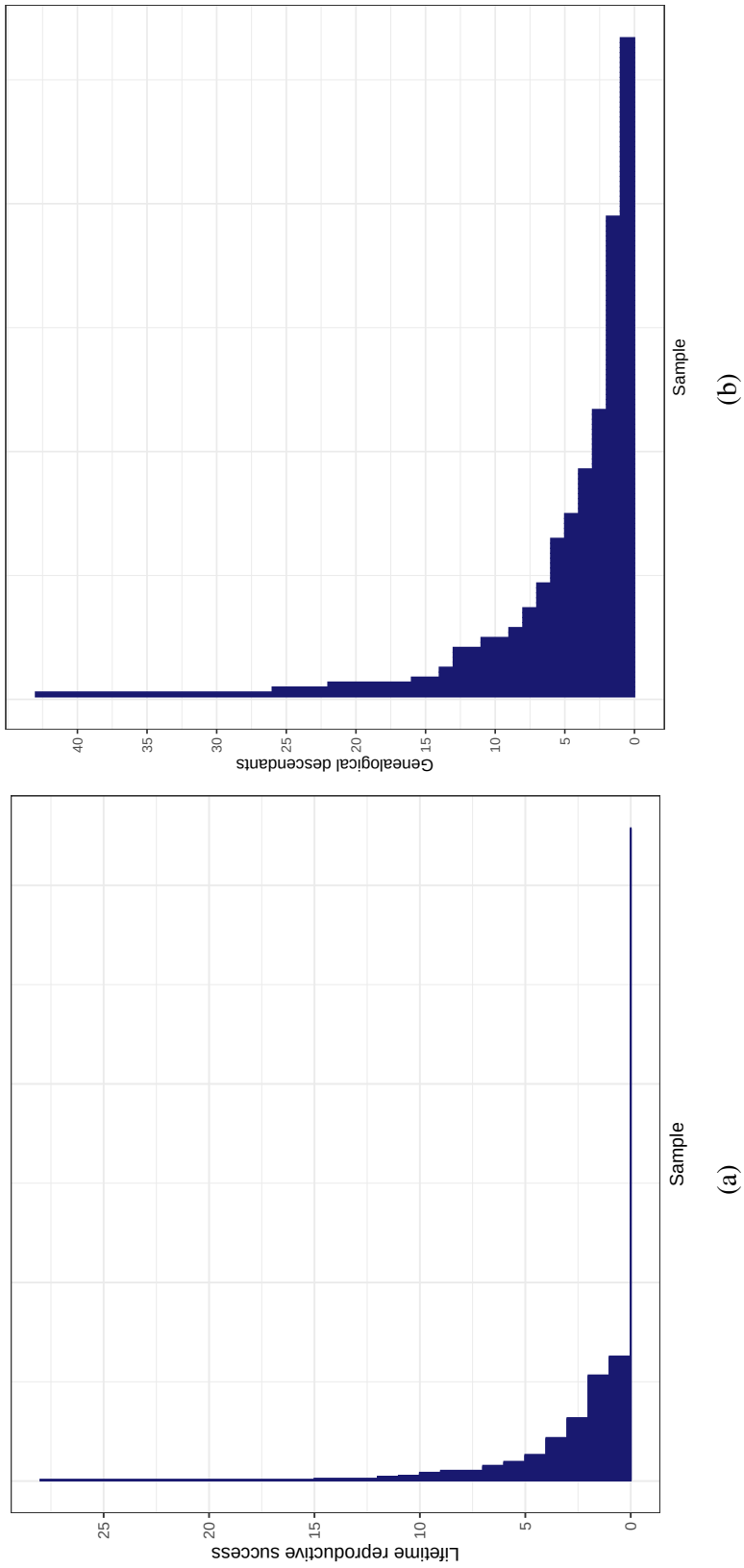
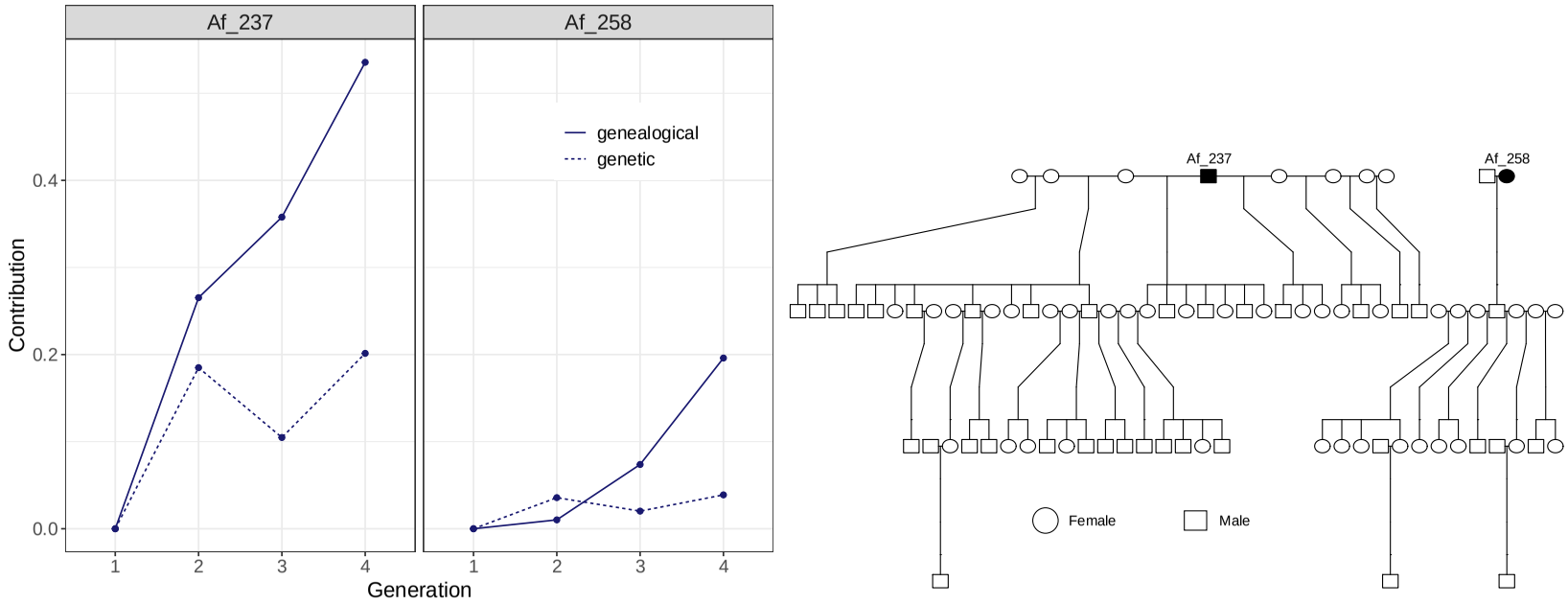


Fig. 3.5 Fitness estimated in the Białowieża population measured as a) lifetime reproductive success of each sample in the population, and b) the number genealogical descendants of each founder in the population.



(a) (b)

Fig. 3.6 a) The genealogical and genetic contributions of two founders (male Af_237 and female Af_258) to the pedigreed population, as a proportion of pups born in a given generation that are descended from the focal founder, and the proportion of alleles in a given generation that are identical by descent from the focal founder respectively. b) The respective pedigrees of these two founders (filled in square and circle) showing their full genealogy. Af_237 had a lifetime reproductive success of 26 pups, whereas Af_258 had one. The total number of genealogical descendants were 43 and 14 respectively.

Fitness measure	W	p
Lifetime reproductive success	3536.5	0.269
Genealogical contribution of founders	14,704	0.348
Genetic contribution of founders	7296	0.092

Table 3.3 Wilcoxon Rank Sum test assessing the statistical difference between male and female lifetime reproductive success, and genealogical and genetic contributions of male and female founders to the population.

3.3.4 Estimates of inbreeding and genetic diversity

Mean F_H in the population was 0.00287 ($sd = 0.0453$) suggesting the population is not inbred overall, and conforms to HWE. This is surprising given the population experiences annual cycles of decline during winter, where high levels of inbreeding and low genetic diversity are expected. However, plotting the distribution of F_H indicates it is highly positively-skewed (coefficient of skewness, $\gamma = 0.598$, figure 3.7), indicating that although the population is not inbred overall, there are more individuals that are inbred than outbred (number of mice where $F_H \geq 0.05 = 47$ and $F_H \leq -0.05 = 12$).

3.3.5 Allele frequency dynamics

84, 200, 80 and 2 genotyped mice were born in 2015, 2016, 2017 and 2018 respectively. As the cohort from 2018 was composed of only two mice, they were discarded from the analysis as they do not provide a representative sample of allele frequencies in the population for that year. Of the total 21,011 filtered and independent SNPs obtained from ddRAD-seq, 21,010 SNPs were present in all cohort years. As only a single SNP was missing in one year of the study, it was excluded from further analyses. The presence of 99.99% of alleles in every year indicates the number of samples in each cohort are representative of the population despite a large number of unassigned individuals in the pedigree, and generational differences in allele frequencies should be reliably captured.

Allele frequencies varied between 0.0233-0.579, 0.0390-0.5410 and 0.0124-0.610 for cohorts from 2015, 2016 and 2017 respectively. Despite SNPs having already been filtered for

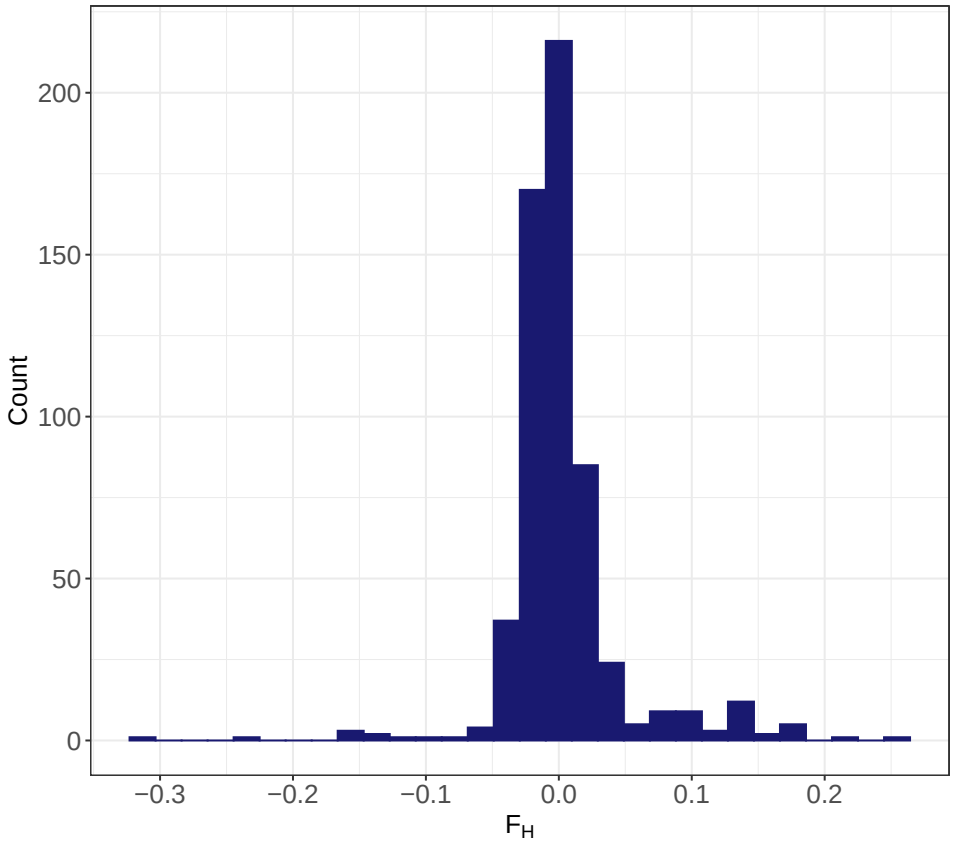


Fig. 3.7 Inbreeding measured as F_H in the Białowieża population. $F_H = 0$ indicates an individual is not inbred, > 0 indicates high levels of inbreeding and excess homozygosity, and < 0 indicates outbreeding with excess heterozygosity. Mean $F_H = 0$ for the whole population indicates the population also conforms to HWE.

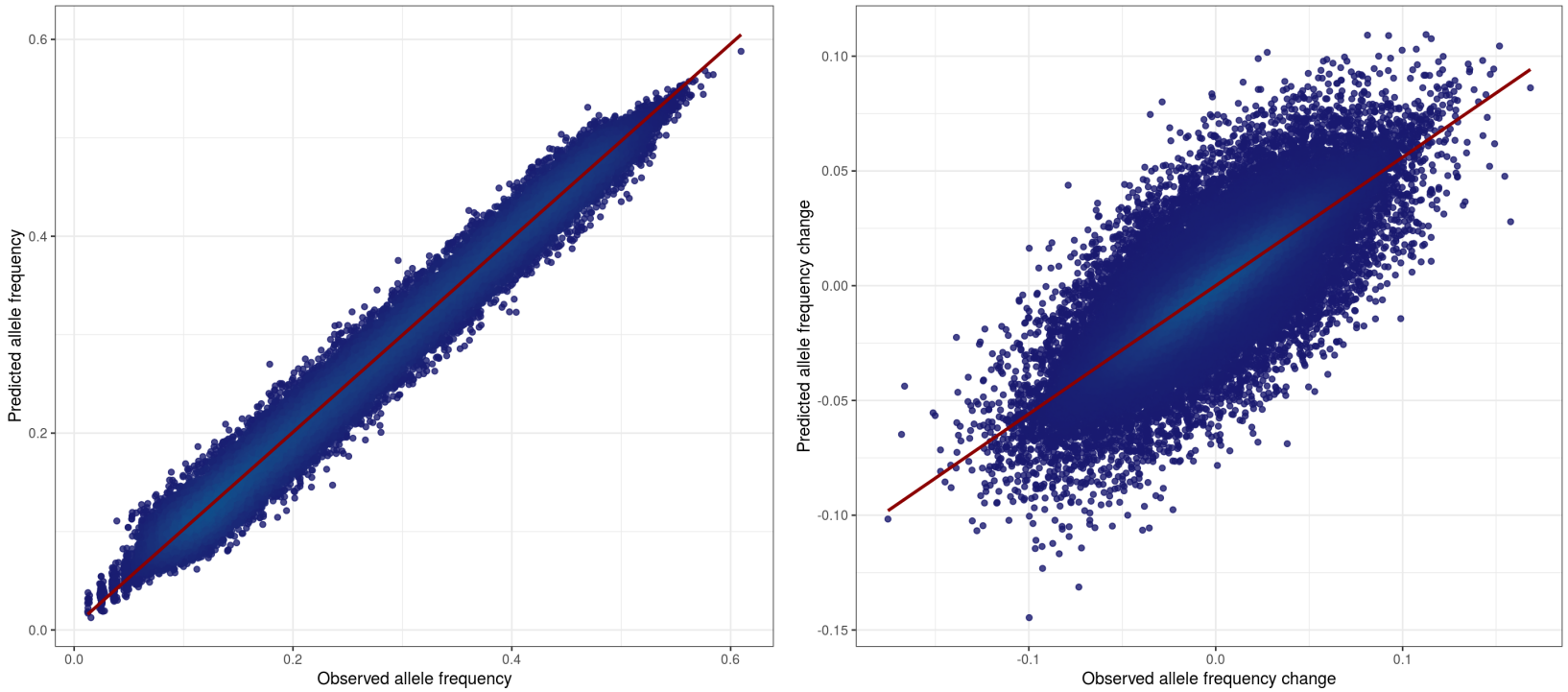
not being in HWE, the observed allele frequency change, Δf_o, was highly variable between -0.217 and +0.164 between 2015-2017 (figure 3.8b) indicating some allele frequencies are changing considerably between years. Model predictions (f_p and Δf_p) from linear regression of f_o and Δf_o with cohort year, also correlate very highly (table 3.4, figure 3.8) indicating allele frequencies are changing linearly and predictably between cohorts. However, the mean Δf_o was very low ($\mu = -0.00068,\ sd = 0.048$).

Surprisingly, 16,534 SNPs (78.7%) increased or decreased in frequency significantly more than expected ($p \leq 0.05$, figure 3.9). Despite the significance of this change, the effect size is generally small, where 11,653 significant SNPs (70.5% of SNPs with p-values ≤ 0.05) showed $\leq 5\%$ change over three years. The remaining SNPs however, are subject to extreme

	T	df	ρ	p
f_o-f_p	2131.5	63,028	0.993	< 0.0001
Δf_o-Δf_p	263.6	42,018	0.790	< 0.0001

Table 3.4 Correlation (Pearson's ρ) between observed and predicted allele frequencies (f_o-f_p), and observed and predicted allele frequency change (Δf_o-Δf_p) between 2015-2017.

changes in frequency. Two examples of this between 2015 and 2017 are shown in figure 3.10, which displays the results from a single simulation of Δf_e, to make clearer the pattern being described. The minor allele in un_19825 appears to significantly decrease by 21.7% between 2015-2017 ($p < 0.05$). Un_16920 on the other hand, appears to increase by 16.1%. Though this is not significantly different from that expected by chance ($p = 0.96$). These results suggest genetic drift or sampling effects alone cannot explain the annual changes in allele frequencies.



(a) (b)

Fig. 3.8 Correlation between a) observed allele frequencies, f_o and b) observed allele frequency change Δ_{f_o}, with the predicted allele frequency, and predicted allele frequency change respectively
, among all cohort years.

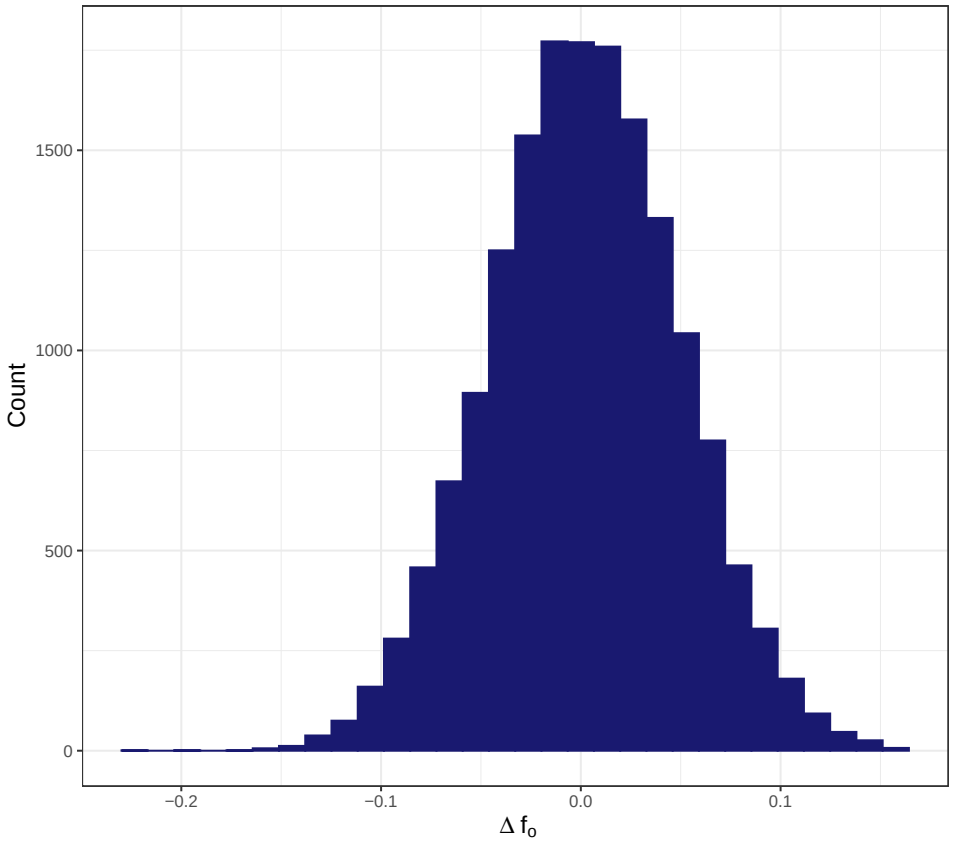


Fig. 3.9 SNPs with a significant ($p \leq 0.05$) observed allele frequency change (Δf_o) between 2015-2017. The majority of SNPs (70.5%) show $< 5\%$ change over three years.

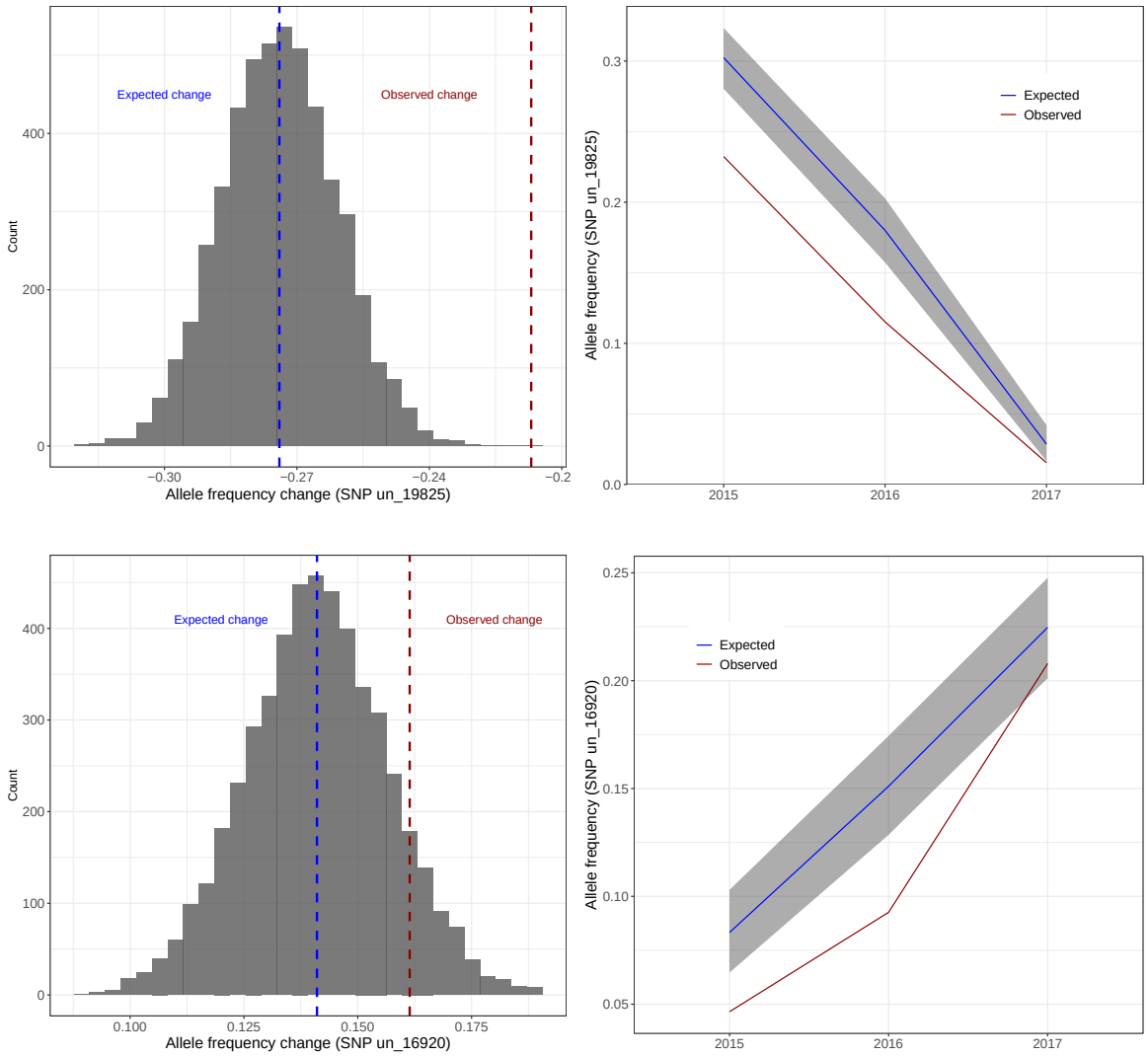


Fig. 3.10 Net expected and observed allele frequency change for a SNP decreasing (top) and increasing (bottom) in the population after a single simulation, sampling alleles 5000 times. Blue and red dashed lines indicate the expected and observed net allele frequency changes between 2015-2017 (left). Grey bars (left) indicate the distribution of net allele frequency change after 5000 samplings. Blue and red solid lines (right) indicate the expected and observed allele frequency for each year between 2015-2017. Grey shading (right) indicates the 95% quantiles of the expected allele frequency distribution.

3.4 Discussion

Apodemus flavicollis in Białowieża National Park are extremely vulnerable to environmental variation, as winter survival is strongly positively correlated with mean winter temperature, and seed crop availability [269]. The population is known to crash by up to 86% of its autumn population size as a result, and exhibits a distinctive cycle of population growth and decline with the seasons (peak in autumn and trough in early spring). As population size and density decrease, localised founder effects should lead to characteristically high levels of inbreeding and low genetic diversity, and genetic drift should increase [29]. Rare alleles are then more likely to be lost, though common alleles can also be subject to its effects [215]. This could in turn reduce fitness, as recessive deleterious alleles accumulate [55, 162]. However, this study has shown that despite the annual population cycles reported previously [269], genetic diversity and inbreeding in *A. flavicollis* from Białowieża National Park are, on average, consistent with a stable population that conforms to Hardy-Weinberg equilibrium.

3.4.1 The density-migration-drift balance

The Białowieża population of *A. flavicollis* has surprisingly shown a remarkable level of genomic resilience to the effects of population crashes in this study. Although 78.7% of alleles show a significant change in allele frequencies between 2015-2017, the effect size is generally small for the vast majority, and the mean change is close to zero. Despite this overall stability in allele frequencies over time, the example of the two SNPs shown in figure 3.10 demonstrates how the remaining 21.3% of allele frequencies can vary considerably in just three years (up to 21.7% frequency change). This could be caused by genetic drift following the inbreeding shown in figure 3.7, which may cause alleles such as SNP un_16920 (figure 3.10, bottom row) to drift considerably, but not more, or less, than expected by random chance. The greater than expected change in alleles such as SNP un_19825 however, suggests drift alone cannot be responsible for this variability. Furthermore, the question remains how the majority of allele frequencies remain relatively stable?

Adams et al. [4] and Robinson et al. [284] show in their studies on a founding population of wolves on Isle Royale, how a single immigrant into a highly inbred population subject to significant genetic drift, can rescue the population from the effects of inbreeding (excess homozygosity and accumulation of deleterious alleles caused by a small effective population size, N_e). As the relative fitness of this immigrant was significantly greater than the individuals within the local population, and the population remained isolated following the migration event, every wolf on the island was soon related to this single male within 2.5 generations. His direct descendants subsequently accounted for 56% of the population, causing a complete selective sweep of the native wolves' genomes. Similarly, some founders in Białowieża National Park also appear to make a disproportionate genealogical and genetic contribution to the population, as is shown by Af_237, which contributed 20.1% of alleles within four generations, and whose descendants account for 53.6% of the pedigreed population.

Unlike the Isle Royale wolves, however, homozygosity here does not appear to be in excess on average, rather, it is close to that expected under HWE. This is more typical of panmictic populations with little genetic structure, and few spatial constraints [61]. Following a harsh winter population crash, a drifting *A. flavicollis* (sub-)population, as indicated here by large changes in allele frequencies, is likely to be buffered against the negative effects of inbreeding and low genetic diversity by immigration from nearby (sub-)populations. This should introduce additional genetic variation into the local gene pool, and would explain why some alleles appear to be significantly drifting, while the majority remain relatively stable. If migration were restricted among (sub-) populations, as with the Isle Royale wolves, *A. flavicollis* from Białowieża would otherwise form a closed system, where heterozygosity (H) would be expected to decline over t generations soon after the first appearance individuals like Af_237, according to:

$$H_t = H_0(1 - \frac{1}{2N_e})^t, \tag{3.3}$$

where N_e is the effective population size, and H_0 the heterozygosity at $t = 0$. This highlights the strong relationship between N_e and the rate at which alleles are lost due to drift, which causes heterozygosity to decline by a factor of $\frac{1}{2N_e}$ at each generation [166] (figure 3.11). This model has been validated previously in wild caught white-footed mice (*Peromyscus*

leucopus)[343], though these mice were captively bred, making the conditions under which heterozygosity declines according to theoretical expectations highly controlled. The study therefore met the assumptions of no migration or mutation, making inferences of natural populations somewhat limited.

With the buffering effects of migration, the effect of drift may not be so pertinent in the short term for *A. flavicollis* in Białowieża, despite the annual population cycling. N_e may therefore be considerably larger than expected, and further agrees with a recent finding by Martin Cerezo and Martin Cerezo et al. [53, 212], which show that genetic differentiation among (sub-)populations of *A. flavicollis* sampled from multiple distant sites ($> 500km$ apart) across Europe, is very low, suggesting there is a large degree of connectivity among sample locations.

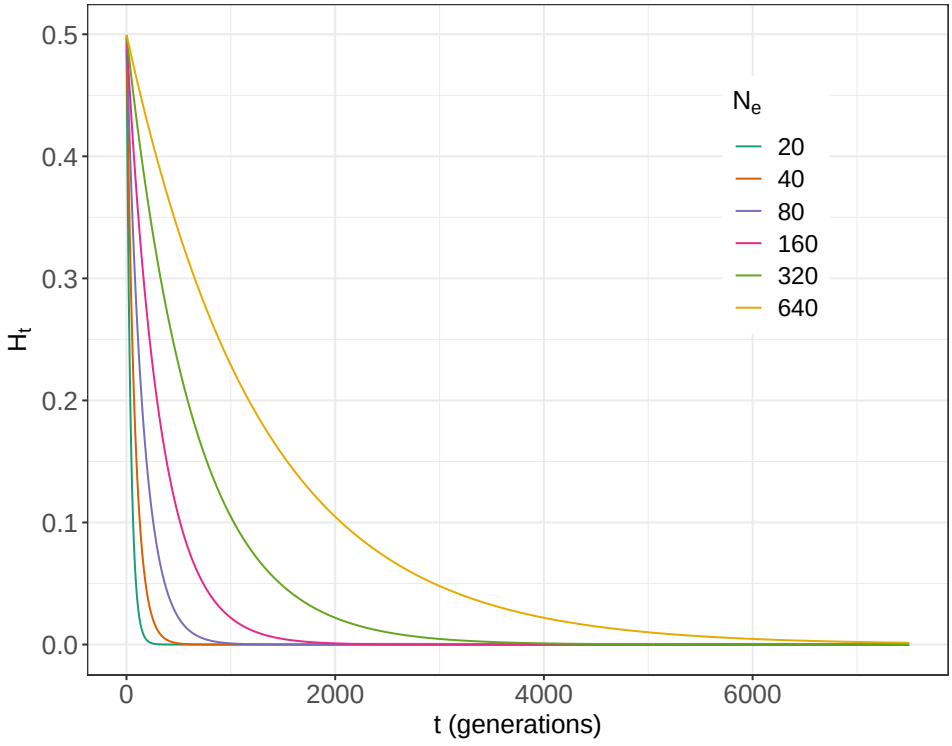


Fig. 3.11 Theoretically decreasing heterozygosity (H_t) over time (t) at varying effective population sizes (N_e). This model assumes a closed population with little or no migration and mutation. Models begin with $H_0 = 0.5$ at $t = 0$, when the population is founded.

Low genetic differentiation in cyclic populations over large distances has also been observed in different species of voles, lemmings and snowshoe hares [28, 92, 282], and

density-dependent dispersal has been suggested as a possible explanation to this pattern (figure 3.12) [192]. This mechanism suggests local demes are partially isolated during low density periods. In the case of *A. flavicollis*, this could be due to deep winter snowfall limiting the capacity for mice to move between sites, and may result in genetic drift at the local scale. Drift could then be compensated for by periods of high population densities, increased migration, and panmixia, which for *A. flavicollis* occur during the summer and autumn when there is low snowfall, and food is plentiful. Also known as the *founder/flush* model [310], it could explain how local demes, or even (sub-)populations that may be declining, are rescued from high levels of inbreeding and genetic drift by migrants moving between demes and interbreeding. This effectively forms a large *meta-population* of source and sink (sub-)populations, where fine-scale spatio-temporal and genetic structuring does not necessarily reflect observations at the meta-population level [263].

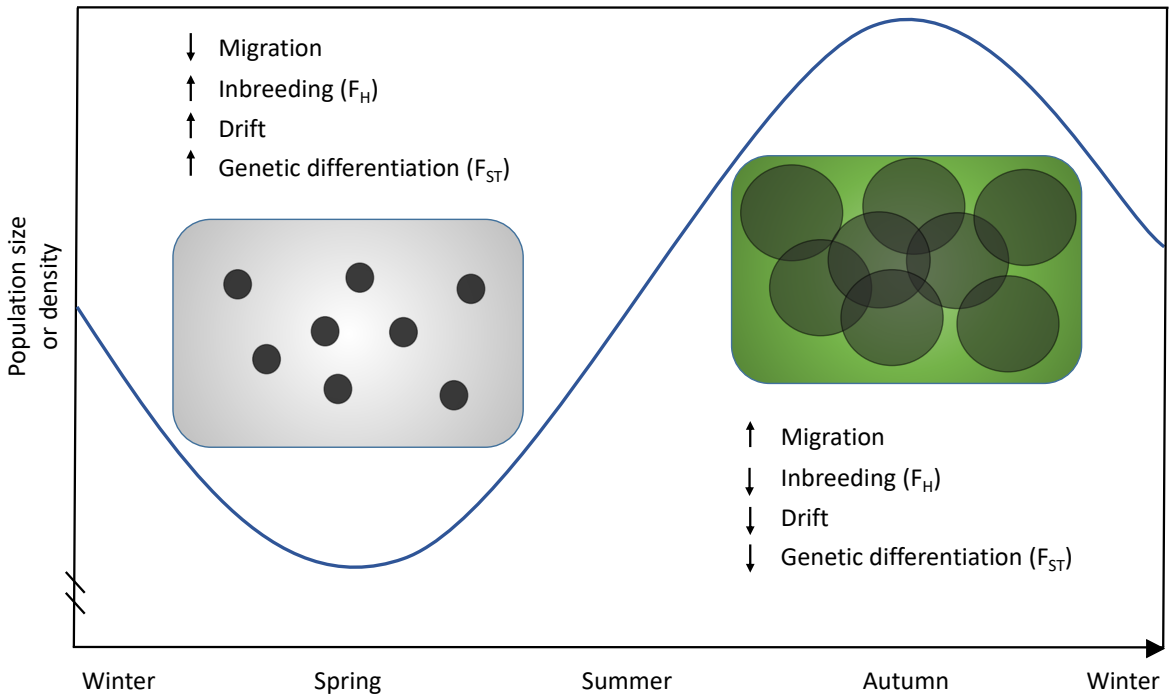


Fig. 3.12 The *founder/flush* model illustrates how a cyclic population exhibits density dependent dispersal, to maintain high levels of genetic diversity despite periodicity in the population size and density. This strikes a balance between population size/density, migration and drift. Only a single period of the cycle is represented here. Black circles represent the relative sizes of sub-populations or local demes at different times of year, and overlapping circles indicate migration is occurring between them.

Density-dependent dispersal was demonstrated by Berthier et al. [28] in a population of fossorial water voles (*Arvicola terrestris*), where genetic differentiation among demes increased when population densities were low, and decreased as the overall population density increased, suggesting periodic migration between demes. Similar to *A. flavicollis*, overall genetic diversity when considering the whole population was also found to be high in *A. terrestris* despite the population cycling. Berthier et al. [28] also demonstrated that genetic diversity reflected the periodicity of population density and size, a result not able to be replicated here due to the lower spatio-temporal resolution of sampling, and the short length of the study. Furthermore, the period of the cycles in *A. terrestris* population size and density are considerably longer than *A. flavicollis* (years versus months), allowing considerably more time for local demes to differentiate. The findings presented here are then even more remarkable considering the lower sampling resolution. The low mean change in allele frequencies in *A. flavicollis* thus indicates density-dependent dispersal and *founder/flush* models could also explain how high levels of genetic diversity are maintained here, albeit at an accelerated rate.

3.4.2 The role of selection in short term allele frequency dynamics

Although *founder/flush* models could explain much of the allele frequency variation in *A. flavicollis* from Białowieża, there of course remains the possibility that these patterns may be the result of natural selection as well as drift and migration. As SNP un_19825 exemplifies, many alleles observed here change significantly more than expected by chance over time. This suggests drift alone is unlikely to be responsible for this variation. Oliver and Piertney [250] also describe on Coiresa, Scotland, a closed island population of *A. terrestris* that experienced long term isolation, where overall, heterozygosity within the *Arte*-DRB locus of the MHC class II region was significantly higher than expected relative to neutrally drifting loci. Following a significant bottleneck where the effective population size reached critically low levels, allele frequencies within the *A. terrestris* population became highly skewed (increased homozygosity), but unlike this study, quickly returned to levels seen prior to the crash. This indicates allelic variation within the locus, which has a functional role T-cell

mediated immunity in mammals, is maintained by a balancing selection equilibrium where heterosis confers a significant fitness advantage, but becomes highly skewed as N_e becomes critically low due to potent drift. There is no indication of a similar trend for alleles observed as under selection in *A. flavicollis* here, though this could be due to the short term of the study.

Different to the *A. terrestris* population in Coiresa, *A. flavicollis* from Białowieża form an open system subject to other evolutionary forces such as migration, which likely limit the effects of positive, negative, balancing and stabilising selection. As many SNPs decreasing in frequency more than expected were already at relatively high frequencies at the start of the study, it is likely SNPs such as un_19825 are weakly deleterious alleles which are not subject to exceptionally strong negative selection. This is because highly deleterious alleles are generally maintained at very low frequencies in a population, if not lost altogether due to their negative effect on fitness [200]. However, due to the short term of this study, the power to reliably detect short-term selection is low [178] due to the compounding effects of drift and migration, and explains why the majority of alleles appear close to HWE, while a small number of significant alleles appear to show strong signatures of selection ($\geqslant \pm 5\%$ change in frequency). Genetic drift and migration therefore probably have a greater effect than selection on short term allele frequency dynamics. Though, this is not to say that natural selection is having no effect at all, rather the pattern being described here is likely a complex interaction between changes in population density, migration and drift acting in the short term, and possibly selection over a much longer time scale.

3.4.3 The limitations of molecular pedigree analyses

Pedigrees in this study were created using data obtained from a modified quaddRAD protocol [106], and, as extensively discussed in chapter 2.4, genotyping errors could be carried forward in downstream analyses. For example, genotyping errors caused by allele drop-out could result in inflated estimates of homozygosity and inbreeding [11]. Though in this study, this is unlikely, as mean homozygosity and levels of inbreeding were found to be low, a result supported by low mean relatedness in the population as found in chapter 2.2.3.3

($\mu_{r\text{-}MLE} = 0.00995,\ sd = 0.0412$), and few pairwise estimates of relatedness were found ≥ 0.55 (as expected for inbred first degree relatives). Furthermore, the mean genotyping error rate found in chapter 2.3.5.1 was also low, despite low sequencing coverage in some samples. Together with the stringent filtering thresholds applied to generate the final panel of SNPs, the effects of genotyping error on downstream analyses should be minimal.

Although the genotyping error rate here was low, there remains a possibility of false familial assignments when constructing the pedigree. Strucken et al. [318] tested the effect of genotyping errors, by varying the stringency of the permitted percentage of genotype mismatches between related individuals when constructing pedigrees, from 0-2%. The study found that setting the threshold for the percentage of allowed mismatches required moderation between allowing for too few mismatches and thus increasing the probability of false negatives, and allowing too many, thus increasing the probability of false positives. However, this study was based on gap-based test statistics to assign familial relationships, and the threshold is arbitrarily set to separate and exclude related and unrelated individuals. Also, gap based methods do not incorporate genotyping error rate directly into the statistical analyses, and assume mutation does not occur. These methods should therefore be avoided [103], particularly as the size of modern genomics datasets using next generation sequencing is increasing. The recommendations by [318] are thus not applicable here.

Unlike gap based statistics, SEQUOIA has integrated genotyping error rates into the maximum likelihood statistics used to construct the pedigree (as provided by `sequoia`'s `err = 0.020` modifier), and should be robust to false assignments. This is supported by the simulations used here to estimate the confidence probability of these assignments, which found the assignment error rate is low ($\mu_{err} = 6.8 \times 10^{-4}$). However, as mentioned in section 3.3.2, this method assumes the pedigree used to compare the simulations to, is true, making this analysis anti-conservative when the pedigree generated by SEQUOIA is used as the truth set, instead of an independently produced pedigree (e.g. from field observations). Assignment error rates may therefore be underestimated here [142]. Pedigree accuracy was therefore also assessed using a SNP-based genomic relatedness matrix, which also found the pedigree to be accurate. Furthermore, Huisman [142] found that pedigree error is $< 10^{-5}$

when $400 \leqslant n_{SNPs} \leqslant 800$ for pedigree construction ($n_{SNPs} = 701$ here), and true assignments were high even with missing life history data. This also lends additional confidence to the final pedigree's accuracy.

One seemingly unavoidable limitation of molecular pedigree construction, is its reliance on as many *a priori* field observations as possible to estimate life history information (e.g. age, birth year, life stage and sex) [142, 226]. For species that can be easily monitored, identified, and are monogamous (e.g. Florida Scub Jays, *Aphelocoma coerulescens* [60, 274]), obtaining this information is made straightforward by marking and recapturing individuals and their offspring. However, for species which are small, largely nocturnal, and exhibit multiple paternities such as *A. flavicollis* [40, 129], mark and recapture studies are considerably more difficult. Unfortunately, this means much of the life history data is missing in this study, as accurately estimating life stages and birth years can be difficult in wild mice. Including such ambiguous data could otherwise introduce large amounts of uncertainty and increased error in the pedigree. The effect of such missing data here was 191 possible parent-offspring pairs which remained unassigned despite the high likelihood of a first degree relationship, and more *dummy* individuals are assigned to the pedigree than actual genotyped mice. Despite these limitations however, the results presented in this study still indicate a strong genomic signature which suggests how allele frequency variation could be maintained over short time scales in *A. flavicollis* from Białowieża National Park.

Chapter 4

The heritability of thermal strategies in wild *Apodemus flavicollis*

4.1 Introduction

4.1.1 The cost of high performance

Life processes are ultimately governed by the flow of energy, and a net gain in energy is essential to enable self maintenance, growth and reproduction [173]. The constant regulation of an organisms internal environment (homeostasis) is essential in complex multicellular organisms. Endothermic homeotherms such as mammals and birds, are able to maintain stable and metabolically favourable body temperatures (T_b) by producing heat internally through elevated metabolic processes, that has provided an evolutionary advantage to allow many species to penetrate different ecological niches, often in the harshest of environments [70, 265]. Maintenance of endothermic homeothermy however, incurs a high metabolic cost (measured as basal metabolic rate - BMR) even when under thermoneutral conditions [3]. In extreme environments, where resources are often scarce and the gradient between internal and ambient temperatures is steep, the energy balance could be tipped into a deficit as the energy required to heat or cool the body to maintain homeothermic conditions (stable internal T_b) is greater than that gained through food. BMR in endotherms, is therefore typically 8-10x higher

than an equivalently sized poikilotherms, which rely on obtaining heat through behavioural means such as basking [94, 224]. Endothermic thermoregulation, despite its high cost, is still considered evolutionarily advantageous as it allows these organisms to maintain high performance in otherwise unsuitable habitats [132, 133]. Many endotherms have therefore evolved different strategies to minimise these costs, such as external morphological features to retain (or expel in hot environments) as much endogenously produced heat as possible (e.g. feathers, fur and fat or large ears in cold or warm environments respectively). Physiological adaptations such as torpor are also a common strategy to minimise energy expenditure in harsh environments, and have been the subject of much research [35, 290, 291].

4.1.2 Torpor as an energy saving strategy

Although most mammals and birds are considered as endothermic homeotherms, many species in fact decrease their energy budget under adverse conditions by entering a torpid state: a period of low responsiveness characterised by a reduction of T_b and metabolic rate to a fraction of the BMR [114, 117]. Metabolic processes are not suspended on entry into torpor, rather they are maintained at a significantly lower rate to reduce the individual's energy budget [115, 246]. This highly regulated hypothermic and hypometabolic strategy described as heterothermic endothermy, can reduce maintenance costs by up to 65% and is common among endotherms [35, 290]. It is generally classified into two main phenotypes: hibernation, and daily torpor.

4.1.2.1 Hibernation versus daily torpor

Hibernation is defined by long torpor bouts lasting multiple days to weeks, is interspersed by brief periods of arousal, and often, but not exclusively, occurs seasonally. Daily torpor however, typically lasts only for a few hours per day [114, 290] and can be easily distinguished from hibernation [290]. For example, seasonal hibernation is significantly longer and deeper than daily torpor [290], but is not continuous. Intermittent periods of rewarming and increased ventilation occur often for a few hours at a time which could be to allow expulsion of waste products accumulated during hibernation [132]. Furthermore, due to the

length of time hibernators are torpid, many species fatten before entry to increase energy reserves, while others hoard food nearby for sustenance during the short periods of arousal [144]. Hibernators also tend to decrease the metabolic rate to 1-5% of BMR, whereas daily heterotherms typically reduce to $\sim 30\%$. However, the metabolic rate in both forms of torpor is strongly correlated with an individual's m_b and is highly variable [117, 31]. Daily heterotherms are on average smaller than hibernators though (18g and 85g respectively), and are typically no heavier than a 9kg, compared to over 200kg in hibernators [97, 117]. Unlike hibernation, daily torpor is not often seasonal, though the length and depth of torpor can increase in seasons where there are severe resource limitations such as during winter in temperate latitudes, or dry seasons in more tropical regions [252, 254]. Furthermore, daily heterotherms do not fatten prior to torpor, and rely on constant foraging to survive [32, 164]. In fact, high proportions of fat can inhibit torpor due to higher leptin levels from brown adipose tissue increasing metabolic activity [116].

4.1.3 The heterothermy *continuum*

Despite the differences between hibernation and daily torpor, it is clear that both are employed as an energy saving mechanism when resources are severely limited. Torpor and hibernation are not mutually exclusive, however, as some species have been shown to exhibit both, dependent on various factors such as body condition [168]. Highly plastic T_b, BMR and torpor duration and length, has led some to suggest the existence of a *continuum* of phenotypes from daily torpor to long term hibernation [35, 36]. For example, Kobbe et al. [168] found in mouse lemurs (*Microcebus griseorufus*) that torpor length varied significantly from short daily events typically lasting only a few hours, to full, long-term hibernation lasting several weeks. Species such as the yellow-necked mouse, *Apodemus flavicollis* however, exhibit daily torpor exclusively [31, 32]. This suggests that heterothermy is subject to phylogenetic constraints [35], and a number of studies have indicated it is highly repeatable [32, 76, 243]. Heterothermic responses could therefore be a heritable trait, and subject to natural selection.

Few of these studies however, investigate the evolutionary processes involved in what maintains such a vast amount of variation at the population level. Here, I estimate the

heritability of heterothermic responses in wild *A. flavicollis*, which undergoes daily torpor in response to external stressors such as starvation [31]. Boratyński et al. [32] indicated heterothermic responses in wild *A. flavicollis* are highly repeatable, and a large amount of the inter-individual phenotypic variation is predicted to be explained by genetic factors. Here, I investigate heterothermic responses in *A. flavicollis* further, and suggest the previously described phenotypic *continuum* [32] may only partially explain the high levels of variation observed here. I then estimate the heritability of heterothermic responses, and suggest how these phenotypes could be maintained in the wild.

4.2 Materials and Methods

4.2.1 Animal handling and experimental procedure

70 mice were trapped by collaborators at the Mammalian Research Institute at the Polish Academy of Sciences (MRI-PAS, Białowieża, Poland) for measurement of heterothermic parameters following Boratyński *et al.* [32]. Brief descriptions of trapping procedures are given here, but for full details please see section 2.2.1.

220 wooden traps baited with oats were laid in a 0.9Ha plot within the Strict Reserve of Białowieża National Park, in two trapping seasons per year (autumn and winter/spring) between 2016 and 2018. Life history information including sex, body mass and approximate life stage were recorded, and tissue samples in the form of tail or ear clippings were taken from each mouse for genotyping. Each sample was subsequently stored in 96% ethanol at -20^oC until DNA extraction. Each mouse was also injected with a Passive Integrated Transponder (PIT-tags: RF-IDW-1, CBDZOE, Grydice, Poland) with unique 16 digit codes for identification purposes.

Once transported to MRI-PAS, mice were kept individually in standard cages at 19 ± 1^oC under a natural light-dark cycle with food and water provided *ad libitum* (carrots, apples and widely available rodent food - Megan, Kraków, Poland). After seven day post-capture acclimatisation in the laboratory, mice were anaesthetised under a 2% mixture of isofluorane and oxygen (Air Products, Warszawa, Poland), and surgically implanted intra-peritoneally

with temperature sensitive, wax coated data loggers, calibrated in a water bath against a high precision (0.1^oC) mercury thermometer at five temperature points between 17^oC and 42^oC. Data loggers were set to measure body temperature (T_b) every 10 minutes with a resolution of 0.062^oC. Mice were injected with a 2% mixture of antibiotics (Enrobioflox 5%, Biowet Sp. z o.o., Gorzów Wielkopolski, Poland) and saline (Baxter Manufacturing Sp. z o.o., Warszawa, Poland), and allowed a minimum of 7 days to recover from the surgery.

4.2.1.1 Fasting experiments to induce torpor

A. flavicollis has a high calorie requirement and must feed continuously to survive [32]. As food restriction can quickly induce torpor, fasting experiments can be used to measure the resulting variation in T_b. However, continued restriction of food (>24h) can result in fatal hypothermia and loss of m_b (body mass), as individuals are unable to return to a normothermic state and wake themselves from torpor without intervention by passively rewarming the mice [32]. Furthermore, as there appears to be no difference in an individuals T_b during torpor between 0-24h and 24-48h [32], it was not necessary, or ethical, to sustain fasting experiments for longer than 24 hours.

After seven days recovery from surgery, m_b was measured ($\pm 0.1g$, ScoutPro 200, Ohaus, Parsippany, NJ, USA) and heterothermy was induced in each mouse by restricting food for a period of 24h, though water was still provided *ad libitum*. After fasting, m_b was measured again and each mouse was allowed to regain any mass lost during the fasting experiment before the measurements were repeated after being allowed a minimum of seven days to recover. Data loggers were then surgically removed, and the mice were released at the place of capture. A number of mice were expected to be recaptured in the following trapping season, and were identified with the unique barcode associated with the sub-cutaneously injected PIT-tag so they were not re-measured. However, a number were found to be missing PIT-tags due to injury, making them unidentifiable. Without any way to confirm these mice had been previously trapped, un-tagged mice were (re)injected with the tag and transported to MRI-PAS for heterothermy measurement. All experimental procedures were approved by

the Local Committee for Ethics in Animal Research and the Ministry of the Environment, Poland (decision numbers 27/2016, 62/2017 and DOP-WPN.287.7.2016.AN).

The Heterothermy Index (H_i) was chosen here as a quantitative measure of the *depth* of an individuals entry into torpor following Boyles et al. [35]:

$$H_i = \sqrt{\frac{\sum(T_{b\text{-}mod} - T_{b\text{-}i})^2}{n-1}} \tag{4.1}$$

where $T_{b\text{-}mod}$ is the modal T_b at the peak of the temperature distribution (i.e. when the individual is in its active phase), $T_{b\text{-}i}$ is the T_b measured at time i, and n is the total number of measurements taken of T_b. The modal T_b when the individual is not torpid was chosen as it serves as a proxy for the theoretical optimum, or *normal*, body temperature experienced by each individual when not torpid. H_i is essentially a measure of the standard deviation from normothermy in degrees Celsius (^oC).

4.2.1.2 Estimating BMR using respirometry

Minimum basal metabolic rate (BMR_{min}) was estimated 1-2 days before fasting experiments, to give an indication of individual maintenance costs in an open flow respirometry system using indirect calorimetry. Air with a known concentration of nitrogen, oxygen and carbon dioxide is pumped into a 0.85l calorimetric chamber, and the relative change in oxygen and carbon dioxide concentration (V_{O_2}, figure 4.1) can be used as a proxy for BMR [101]. Chambers were maintained at an ambient temperature (T_a) of 30°C during daylight hours, and dehumidified air was pumped in to the system (Drierite Co. Ltd, Xenia, OH, USA). Input air streams for each chamber were regulated to $500ml\ min^{-1}$ (calibrated using a soap bubble flow meter - Optiflow 570, Humonic Instruments Inc., USA), and two reference lines were used to measure the input O_2 concentration every 15 minutes. 100ml of output air was measured every second to determine the final O_2 concentration using an FC-10a gas analyser (Sable Systems International), for five minutes. Electronic components of the system were connected to a computer through an analogue-digital interface (U12, Sable

Systems International), and data was recorded using the ExpeData software (Sable Systems International).

Fig. 4.1 simplified representation of an indirect calorimetry set up to measure the BMR of *A. flavicollis*. Arrow thickness indicates the relative change in concentration of gasses between the input and output to measure V_{O_2}.

4.2.2 Statistical analyses and estimating repeatability and heritability

As some individuals were trapped and measured in multiple seasons, degrees of freedom will be artificially inflated in analyses due to pseudo-replication. Mean H_i for each individual was therefore used to compare among grouped variables such as sex, age, trapping year and season. All analyses were conducted in R v4.0.2 unless otherwise stated.

The repeatability (τ), or intra-class correlation coefficient of H_i, was first calculated to ensure within-individual variances in heterothermic responses were low. This can also estimate the upper limit of heritability [98], and is defined mathematically as

$$\tau = \frac{V_I}{V_P}, \tag{4.2}$$

where V_I is the inter-individual variance and V_P the total phenotypic variance. Linear mixed effects regression models to extract the variance components necessary to calculate τ were

fitted in the forms:

$$H_i = \mu + X_1 s + X_2 a + X_3 m_b + X_4 b_{min} + Z_1 Y_{birth} + Z_2 Y_{trapping} + Z_3 I + \varepsilon \tag{4.3}$$

$$BMR_{min} = \mu + X_1 s + X_2 a + X_3 m_b + Z_1 Y_{birth} + Z_2 Y_{trapping} + Z_3 S + Z_4 I + \varepsilon \tag{4.4}$$

using the package LME4 v1.1-23 in R v4.0.2. Here, μ is the global intercept, $X_{1\text{-}4}$ are the design matrices of fixed effects s, a, m_b and b_{min} respectively (sex, age, body mass and BMR_{min}), and $Z_{1\text{-}4}$ are the incidence matrices of random effects Y_{birth}, $Y_{trapping}$, S and I respectively (birth year, trapping year, season and individual ID). ε represents the residuals. If the distribution of residuals did not conform to a gaussian distribution or meet the assumption of homoskedasticity, response variables were Order-Norm transformed using the `orderNorm` function of the BESTNORMALISE package in R v4.0.2 [262]. Variance components were extracted from the models and used to calculate the repeatability of H_i and BMR_{min}.

Calculating SNP-based, narrow-sense heritability (h^2_{SNP}) requires that the pairwise genomic relatedness of each phenotyped individual be known, so the proportion of V_P explained by additive genetic variance (V_A) can be estimated as:

$$h^2_{SNP} = \frac{V_A}{V_P}, \tag{4.5}$$

and allows dominance and epistatic effects to be accounted for. However, this requires a large sample size to reduce standard error (SE) even in a highly heritable phenotype [335]. In a theoretical scenario for example, detecting $h^2_{SNP} > 0$ for a given phenotype when the true $h^2_{SNP} = 0.6$, would still require 1477 samples to reach a power of 0.8 and SE 0.21 (figure 4.2). Thus, heritability was instead estimated according to:

$$h^2_{SNP} = \frac{V_G}{V_G + V_E} = \frac{V_G}{V_F + V_G + V_E} = \frac{V_G}{V_F + V_G + V_{RE} + V_R}, \tag{4.6}$$

where V_G is the overall genetic variance once related individuals have been removed, and V_P is composed of the residual variance due to environmental effects V_E, and V_G. V_E can in turn

be partitioned further into its multiple random effects (V_{RE}) which in this study include Y_{birth}, $Y_{trapping}$ and S, and the residual variance (V_R). Finally, the fixed effects variance (V_F) must also be taken into account to ensure the total phenotypic variance is not underestimated [81], which in this study includes the sex, age, m_b and BMR_{min} of each mouse. This approach to partition variance components into its fixed and random effects was also applied to the calculation of τ.

h^2_{SNP} in unrelated individuals was estimated by fitting a linear mixed-effects model in LME4 V1.1-23 in the forms of formulae 4.3 and 4.4. The variance was partitioned according to formula 4.6. To estimate V_F, the variance was calculated based on the predicted values for models 4.3 and 4.4, which were obtained using `predict` in R V4.0.2. 95% bootstrap confidence intervals were estimated using the `confint` function.

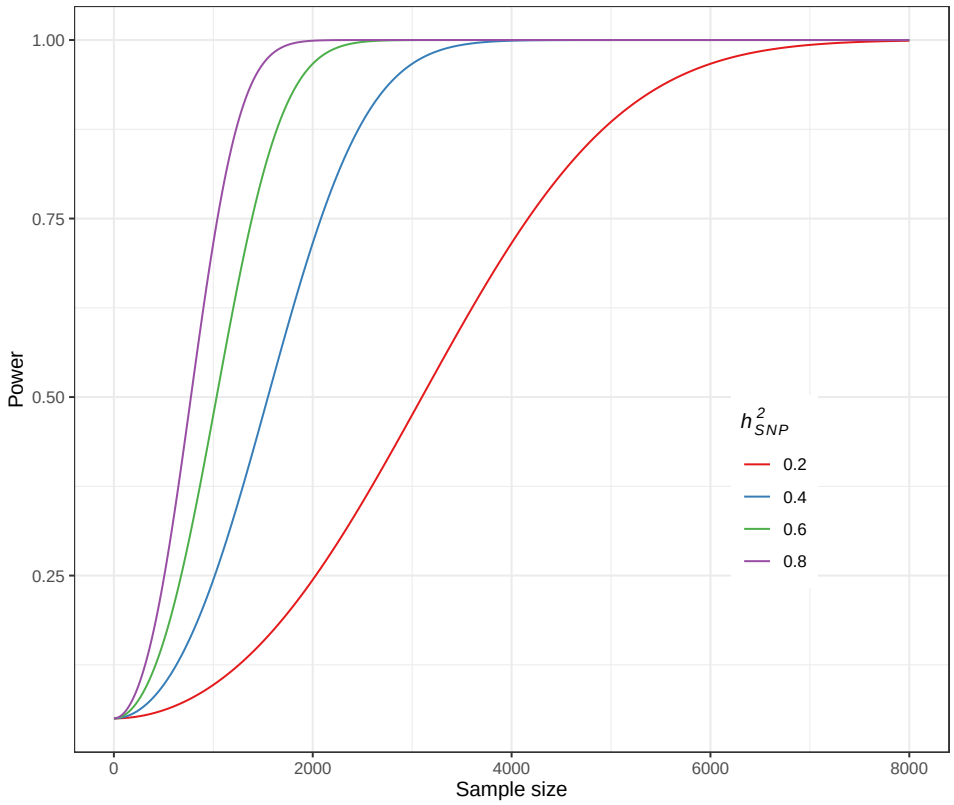


Fig. 4.2 Power curves showing the required sample sizes to detect $h^2_{SNP} > 0$ for a phenotype, given the true $h^2_{SNP} = 0.2$-0.8 represented by colour. Power analysis was performed according to Visscher et al. [335].

4.2.3 Sequencing, bioinformatic processing and estimating relatedness

To estimate V_G and account for any variance in the phenotype that is attributed to shared environmental effects, the relatedness of all phenotyped mice must be known so that highly related individuals can be removed from the analysis. Relatedness was estimated using SNP data obtained from the reduced representation library preparation, sequencing and bioinformatic processing conducted according to the approach described in detail, in section 2.2.2.4 but briefly recounted here.

Sequencing libraries were prepared following a modified quaddRAD protocol [106] and sequenced on an Illumina HiSeq 3000. Outer barcodes were demultiplexed using BCL2FASTQ2 V2.18 [145] and sequences were quality checked using FASTQC [13]. The STACKS V2.3d bioinformatic pipeline was then used to demultiplex sequences according to inner barcodes (`process_radtags`), identify and remove PCR duplicates (`clone_filter`) and assemble loci *de novo* (`denovo_map.pl`). Parameter optimisation was performed by iterating over values of m, M and n, and plotting the associated metrics. SNP data from duplicate samples that were PIT-tagged more than once, were then merged by estimating relatedness calculated as the proportion of the genome that is identical by descent ($r > 0.9$ in duplicates) using SNPRELATE V1.16.0 [354]. Pairwise relatedness among all samples was estimated using 21,011 SNPs filtered for HWE ($p \leq 0.05$), linkage disequilibrium ($r^2 \geqslant 0.035$) and MAF $\geqslant 0.05$. Once duplicates were merged, the pipeline was run again to estimate the relatedness in the phenotyped samples. The final kinship matrix was then multiplied by two and scaled by subtracting the mean relatedness and dividing by the standard deviation to produce the final genomic relatedness matrix (GRM). Using this GRM, cryptically related individuals ($r \geqslant 0.025$) were removed from the analysis to estimate h^2_{SNP}.

4.2.4 Cluster analysis

To better understand if there is any inherent structure in the heterothermy data, a k-means cluster analysis was performed. As clustering algorithms will find structure in even a uniform distribution, the clustering tendency of the data was first assessed to ensure any groups found are not an artefact of the algorithmic clustering process. H_i, m_b and BMR_{min} were first standardised by subtracting the mean and dividing the standard deviation to allow comparisons with a null dataset. A random uniform distribution was then drawn within the range of the true data set for each variable to generate the null. The Hopkins statistic assesses cluster tendency by calculating the probability a data set is drawn from a uniform distribution with no inherent structure [22, 141], and was calculated for both the null and heterothermy data using the `get_clust_tendency` function of the FACTOEXTRA package in R V4.0.2 [160]. A value close to 0.5 indicates no structure is present, and values close to one indicates strong clustering tendency. K-means clustering was then performed for $K = 2$-10 using the `kmeans` function, with the `nstart` argument set to 25 random initial cluster centres. The optimal number of clusters were finally chosen based on a combination of visual assessment and calculation of 30 different indices using the `NbClust` function of the R V 4.0.2 package NBCLUST [57].

4.3 Results

Overall, 34, 31 and five mice were measured for H_i between 2016-2018 respectively (table 4.1). This included 24 females and 46 males. Six adults and 13 juveniles were also measured, however, a further 51 mice were of indeterminate life stage. As there is a possibility that life stage could affect the estimate of H_i, uncertain age classes were treated as missing data in all analyses in this chapter. All mice measured for H_i were also measured for BMR_{min} (see section 4.3.4 for full results on BMR_{min}).

Trapping year	Adult	Maybe adult	Juvenile	Maybe juvenile	Unknown age	Female	Male
2016	4	7	10	3	10	11	23
2017	1	6	3	2	19	10	21
2018	1	1	0	1	2	3	2
Total	6	14	13	6	31	24	46

Table 4.1 Breakdown of mice measured for H_i.

4.3.1 Variation in heterothermic responses

$T_{b\text{-}i}$ varied substantially when mice were torpid during the fasting experiment, and was significantly lower than $T_{b\text{-}mod}$ by 1.21^oC - 14.77^oC (paired Wilcoxon test: $V = 43,071$, $p < 0.0001$, figure 4.3a). As a result, H_i also varied considerably from 0.97^oC - 6.48^oC, and was negatively correlated with m_b ($T_{68} = -3.36,\ p = 0.001,\ \rho = -0.38$, $95\%CI = -0.56$ - -0.16, figure 4.3b). Mass-adjusted BMR_{min} was also found to negatively correlate with mass-adjusted H_i ($T_{291} = -6.22,\ p \leq 0.0001,\ \rho = -0.30$, figure 4.4). As adult *A. flavicollis* can vary in m_b and size markedly, age classes can be very difficult to determine in the field. Only 19 out of 70 mice have been confirmed as adult or juvenile ($n = 6$ and 13 respectively), and mass corrected H_i did not differ between them due to small sample size ($T_{14.5} = 0.58,\ p = 0.57$).

H_i was significantly lower in males than females even when corrected for m_b ($\mu_m = 2.06^oC,\ sd = 0.85^oC, \mu_f = 2.80^oC,\ sd = 1.23^oC$; $T_{64.7} = 8.48,\ p \leq 0.0001$). This indicates heterothermic responses are sexually dimorphic in *A. flavicollis*. Males were also considerably heavier than females ($\mu_m = 42.7g,\ sd = 5.43g,\ \mu_f = 33.6g,\ sd = 3.51g$; $T_{64.7} = -8.48,\ p \leq 0.0001$, figure 4.3b). However, there was no difference in the proportion of m_b lost during torpor between males and females suggesting both sexes bear the same cost (Wilcoxon test $W = 663,\ p = 0.17$). Analysis of variance also shows H_i did not differ between trapping years whether mass corrected or not ($F_{2,67} = 1.76,\ p = 0.18$ and $F_{2,67} = 0.14,\ p = 0.86$ respectively).

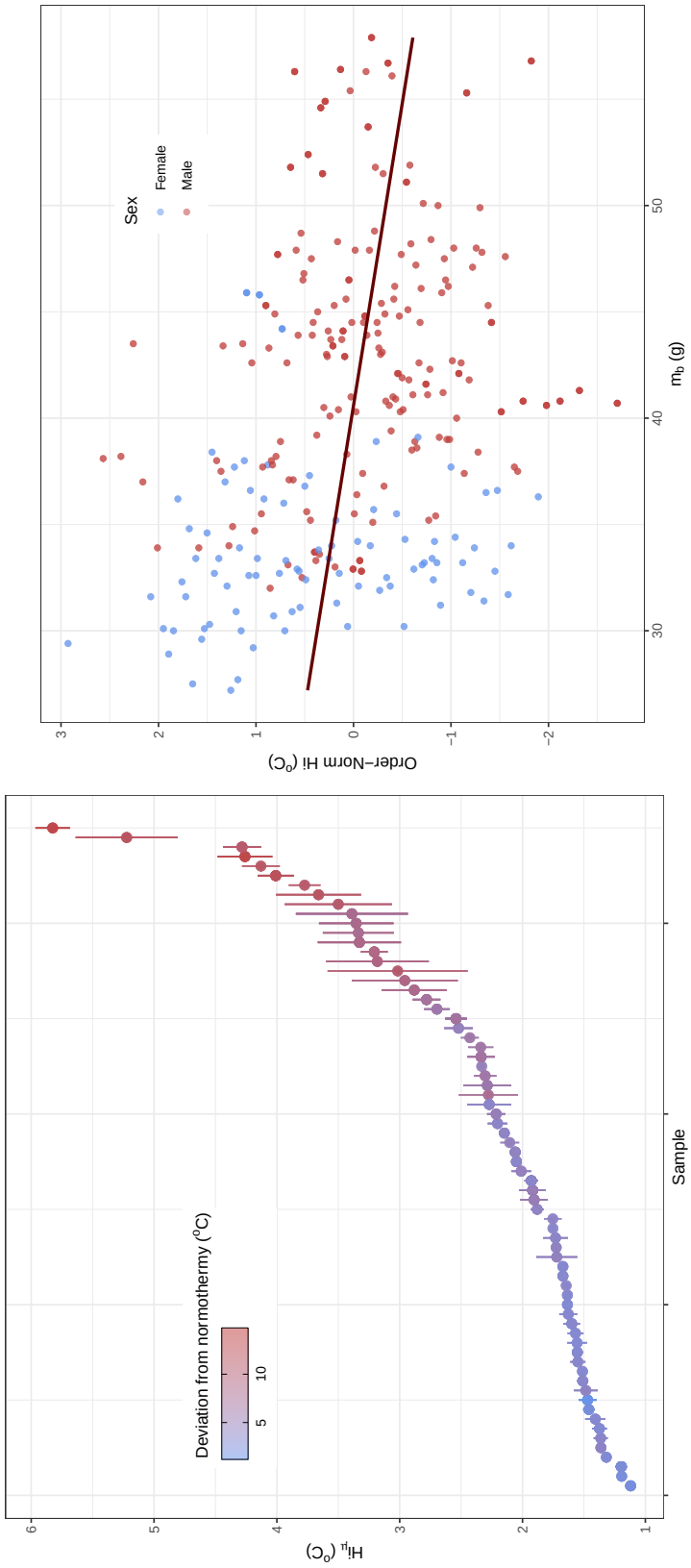


Fig. 4.3 a) Distribution of mean H_i in *A. flavicollis*. Bars indicate standard error, colour indicates $T_{b\text{-}i}$, the deviation from normal T_b. b) The relationship between Order-Norm transformed H_i and m_b. Colour indicates sex.

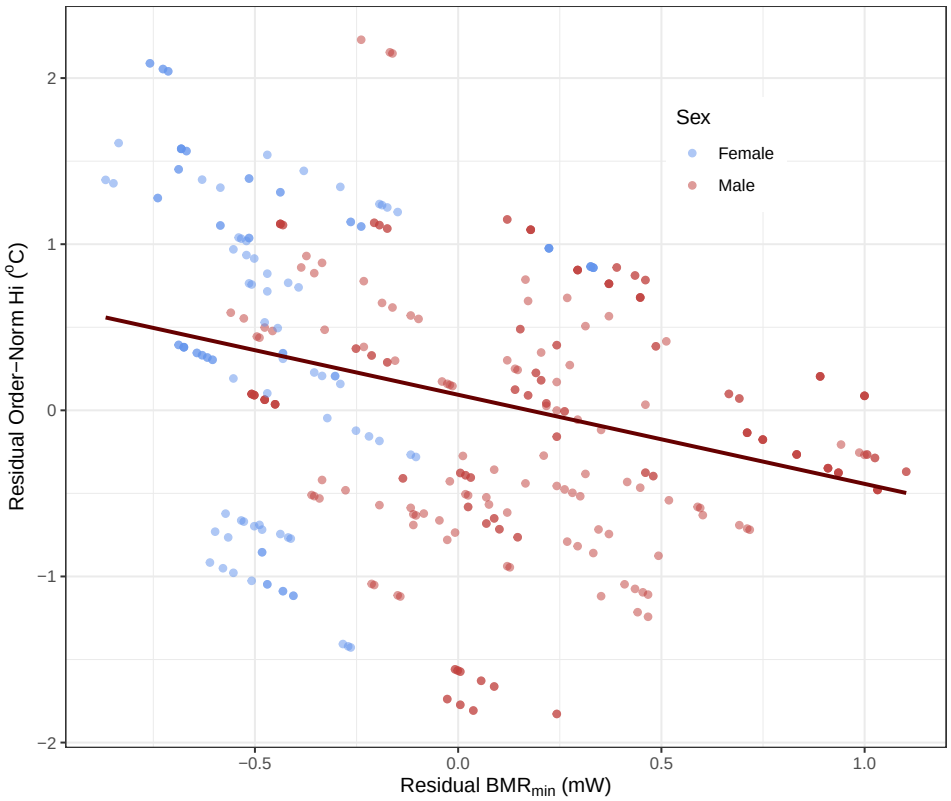


Fig. 4.4 Relationship between mass adjusted Order-Norm transformed H_i and mass adjusted BMR_{min}. Blue and red denote females and males respectively.

4.3.2 Clustering analysis

The Hopkins statistics showed a high cluster tendency for the heterothermy dataset and, as expected, not for the null (0.795 and 0.500 respectively) indicating that real inherent structure exists in the data. K-means cluster analysis revealed two distinct groups in the morphospace indicating the presence of thermal specialists and generalists in the population (figure 4.5). $K = 2$ was chosen for the optimal number of clusters as the indices generated by `NbClust` did not give any clear indication for an optimum (see appendix A.3, table A.4 for details). Visual assessment also indicated a loss of clear structure in the dataset at $K = 3$-10, shown by large amounts of overlap in the clusters (see appendix A.3, figure A.1 for details). Principal components one and two explained 86.3% of the variation in the data overall, and thermal specialists included 192 measures from 36 males, and 15 measures from 3 females. Thermal

generalists on the other hand, included 107 measures from 23 females, and 70 measures from 17 males. This shows overall that males are more often thermal specialists and females are more often thermal generalists, though the sex of the individual is not preclusive to a particular thermal strategy.

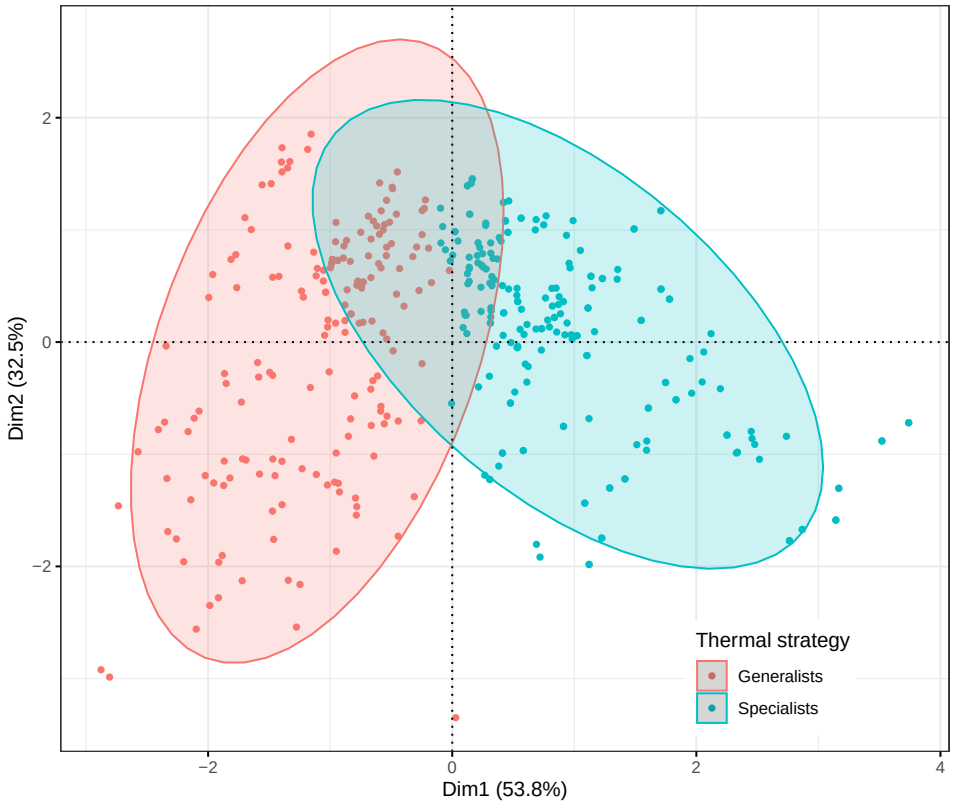


Fig. 4.5 K-means cluster analysis of thermal strategy for $K = 2$. Pink and blue represent thermal generalists and specialists respectively.

4.3.3 Repeatability and heritability of heterothermic responses

To calculate τ, equation 4.3 was simplified based on the significance of the terms. As the age of each mouse was already found to be non-significant, it was removed from the model. To estimate the variance of random effects, the modelling approach explores the parameter space until the best estimate is found. However, if not enough variation exists in the data, these parameters are unable to escape from the boundary of the parameter space and the resulting estimates are unreliable. Y_{birth}, $Y_{trapping}$ and S were therefore also removed as the estimated variance was zero. H_i was found to be highly repeatable in the population as a whole, and for each sex individually ($\tau = 0.68$-0.84, $95\%\ CI = 0.61$-0.86, table 4.2). Pairwise genomic relatedness was low among phenotyped samples overall ($\mu_r = -0.014,\ sd = 0.038$; figure 4.6). However, 73 pairs of individuals were highly related ($r \geqslant 0.025$), 19 pairs of which were first or second degree relations ($r = 0.2$ - 0.55 respectively). Once these highly related individuals were removed, 24 mice were retained for estimating h^2_{SNP}, and heterothermic responses were found to be highly heritable in the population as a whole, and for each sex individually ($h^2_{SNP} = 0.57$-$0.87,\ 95\%\ CI = 0.33$-0.91; table 4.2). h^2_{SNP} was found to be lower in males compared to females, though this difference is not statistically significant as confidence intervals overlap.

Sex	Fixed effects	τ	95% *CI*	h^2_{SNP}	95% *CI*
Overall	$m_b + Sex + BMR_{min}$	0.69	0.61-0.74	0.70	0.58-0.79
	$m_b + Sex$	0.74	0.68-0.78	0.76	0.65-0.82
	m_b	0.75	0.69-0.80	0.78	0.66-0.83
Male	$m_b + BMR_{min}$	0.71	0.64-0.76	0.61	0.39-0.73
	m_b	0.78	0.70-0.83	0.64	0.43-0.75
Female	$m_b + BMR_{min}$	0.76	0.63-0.81	0.80	0.52-0.89
	m_b	0.79	0.68-0.84	0.88	0.68-0.93

Table 4.2 The repeatability and heritability of H_i calculated based on the inclusion of three fixed effects. Models were calculated for both sexes separately and together.

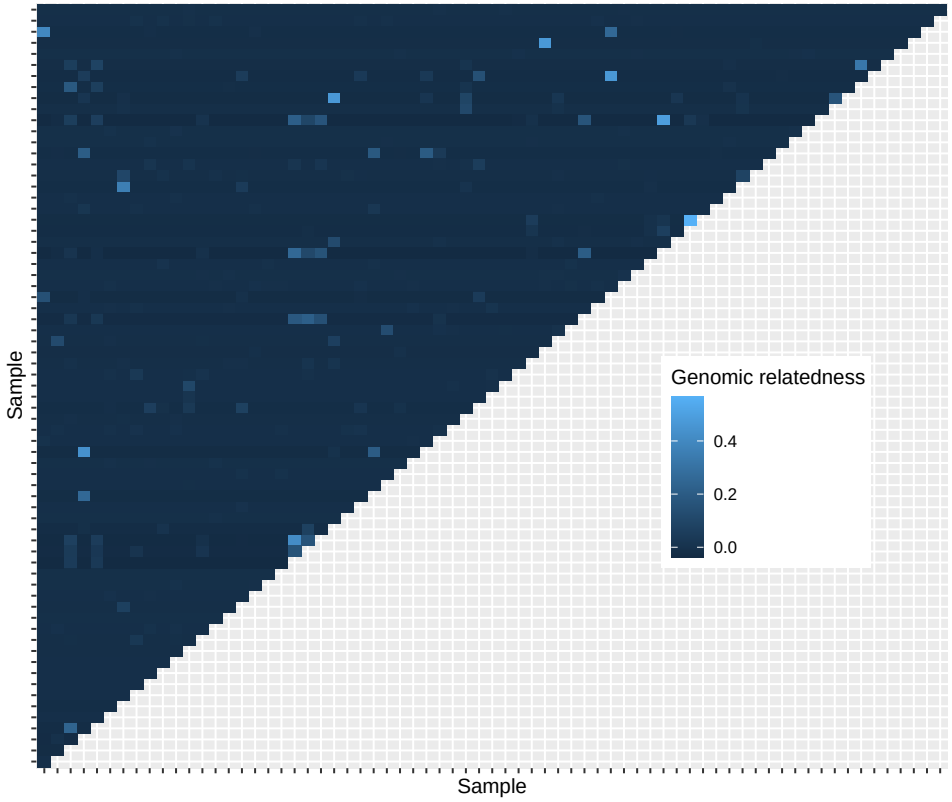


Fig. 4.6 Pairwise genomic relatedness among all phenotyped samples. Lighter blue indicates greater genomic relatedness.

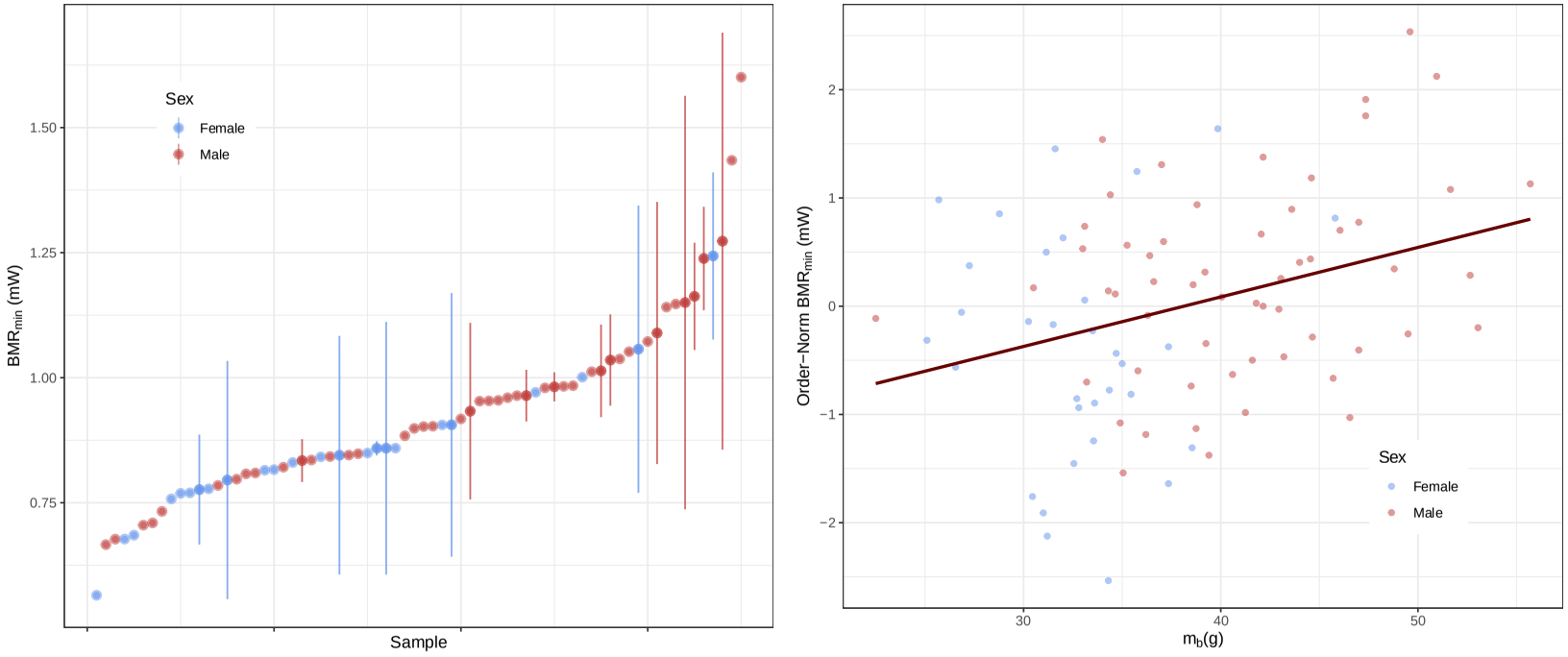
4.3.4 Basal metabolic rate in a wild population

BMR_{min} was measured for all 70 mice phenotyped for heterothermic responses in winter/spring 2016-2018. 19 of these were also trapped and measured in autumn 2016-2017. Overall, BMR_{min} varied between 0.56-1.69mW ($\mu = 0.94mW,\ sd = 0.23mW$, figure 4.7a). As it was significantly correlated with m_b ($\rho = 0.32, T_{87} = 3.16, p < 0.01$; figure 4.7b), BMR_{min} was corrected for m_b in subsequent analyses. There was a significant difference between males and females for mass corrected BMR_{min} ($\mu_{female} = -0.24mW$, $sd_{female} = 0.20mW$, $\mu_{male} = 0.14mW$, $sd_{male} = 0.29mW$; $T_{84.1} = -7.27,\ p \leq 0.0001$), though did not differ between adults and juveniles due to small sample sizes ($T_{14.9} = 1.28,\ p = 0.22$). Linear regression showed m_b corrected BMR_{min} did not differ between 2016-2018, but was significantly higher in winter/spring compared to autumn (table 4.3, figure 4.8). Despite the broad distribution of BMR_{min} in the population overall, and 19 individuals having been measured in

both autumn and winter, repeatability cannot be calculated as the estimated within-individual variance was too low ($V_I \sim 0$) meaning it cannot be partitioned. However, large standard errors due to differences between seasons indicate that repeatability is expected to be very low (figure 4.7a).

	Estimate	Std. error	T	p
Intercept	-0.16	0.08	-2.13	0.04
Trapping year 2017	-0.06	0.07	-0.95	0.34
Trapping year 2018	-0.12	0.15	-0.81	0.42
Season winter/spring	0.25	0.08	3.06	<0.01

Table 4.3 Table of coefficients for linear regression between mass-corrected BMR_{min}, trapping year and trapping season. The baseline intercept represents autumn 2016.



(a) (b)

Fig. 4.7 Variation of BMR_{min} within the Białowieża population. a) Variation in BMR_{min}. Points with lines indicate the mean with standard error which can only be estimated for individuals measured in both autumn and winter/spring ($n = 19$). All other points represent BMR_{min} measured in winter/spring only b) The relationship between m_b and Order-Norm transformed BMR_{min}. Blue indicates females, red indicates males.

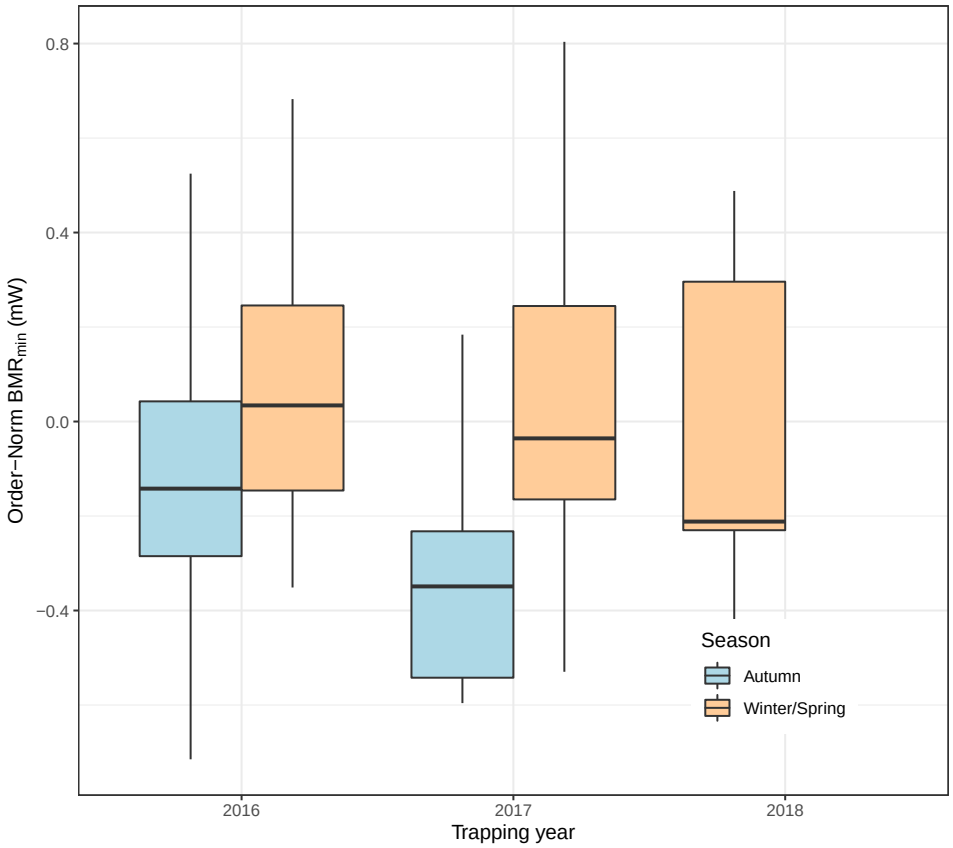


Fig. 4.8 Distribution of BMR_{min} by year and trapping season. BMR_{min} has been Order-Norm transformed and colours represent trapping seasons.

4.4 Discussion

Previous studies have shown that heterothermic responses form a continuum of highly plastic phenotypes [114], and in response to food deprivation is also highly repeatable [32, 76]. This study highlights that heterothermic responses, are not only repeatable, but also highly heritable, variable within the population, and could be subject to natural selection.

4.4.1 Maintenance of variation in heterothermic responses

Phenotypic plasticity has long been suggested as a precursor to natural selection in a variable environment [67]. For example, heterotherms like *A. flavicollis* are able to respond to stressful changes in the environment by lowering their body temperature and metabolic rate, enter into a torpid state and reduce their energy budget by as much as 65%, thus maximising survival probability until conditions are once again favourable [243]. This ensures a balance between energy acquisition and expenditure to avoid a deficit, and allows individuals to maintain a metabolically optimal body temperature when active ($\sim 37^oC$), provided sufficient resources are available [291]. In highly seasonal, temperate forests such as in Białowieża, harsh winter conditions can cause temperatures to drop as low as -20^oC and snow depth can be over 10cm [148]. Food quickly becomes scarce, and winter mortality in *A. flavicollis* can reach 86% [148, 269]. This should impart an extremely large directional selection pressure on the population for any phenotype which can facilitate winter survival. The heterothermy index (H_i) in this case, should be skewed towards higher values if deeper torpor increases survival by reducing individual energy budgets, and the response to selection is strong. The reality is quite different however, and the distribution in heterothermic responses has previously been described as a phenotypic *continuum* within a range of values, indicating the processes that maintain such variation are complex [32, 35]. Unfortunately, the heterothermy *continuum* is often described based on measures such as H_i, which rely on differences in T_b only, and often cannot differentiate between subgroups within the continuum [290, 132]. Inclusion of multiple variables to describe heterothermic responses, as done by Ruf and Geiser [290] and

this study, indicates that heterothermy is a structured phenotype, that is constrained by these multiple interacting and co-variable, secondary phenotypes.

One theory proposes that the large degree of heterothermic variation exists due to constraints by individual body mass (m_b) [130, 168, 188, 337]. For example, Kobbe et al. [168] found that both the duration and frequency of torpor bouts increased in mouse lemurs (*Microcebus griseorufus* (Kollman, 1910)) when an individual's body condition was higher prior to entry. Lemurs were also found to lose a greater proportion of their m_b per day as more time was spent torpid, and the an individual is more likely to enter into deeper torpor when body condition is better, and m_b is higher [337]. This suggests the benefits of deeper, more frequent torpor bouts are conditional on the presence of sufficient fat reserves, and there may otherwise be a large cost for higher H_i [177]. Such phenotypic plasticity leads to the presence of multiple intermediate phenotypes from short bouts of daily torpor to full, long-term hibernation lasting multiple weeks within a single species. In agreement with a study by Vuarin et al. [337] on grey mouse lemurs, it appears heterothermic responses in *A. flavicollis* are also constrained by m_b. However, the trend is reversed where heavier individuals have a lower H_i, and unlike mouse lemurs, they appear to suffer no immediate cost for torpor when measured as the proportion of m_b lost during the fasting experiment. Furthermore, and contrary to the existence of intermediate phenotypes in other species, patterns of two distinct thermal strategies emerge in *A. flavicollis*.

The presence in the population of thermal generalists, which are better able to respond to stressful environmental conditions and enter a deeper torpid state (higher H_i), and thermal specialists, which are more likely to remain at or near normothermic conditions (lower H_i) better explains the patterns found here. Heterothermic responses in *A. flavicollis* cluster well in the morphospace indicating two distinct thermal strategies exist in the population. There is no evidence based on the H_i data presented here to suggest environmental variation drives this variation, and it appears rather, the presence of two strategies is driven by multiple interacting phenotypes (m_b, BMR_{min} and sex in this study). This offers a far more complex picture of heterothermy than can be simply attributed to differences in m_b alone. I propose the description of a heterothermy *continuum* may be incomplete for species which exclusively

enter daily torpor such as *A. flavicollis*, and that heterothermy in the wild may be sexually as well as phylogenetically constrained, suggesting the variation in heterothermic responses may be unique to each species [35, 290]. Few low resolution measurements and high levels of co-variance with continuous variables such as m_b likely hid these patterns in previous studies, and must be accounted for in future.

Schulte-Hostedde et al. [299] showed in three small rodent species that males have proportionally higher muscle mass than females, which may improve mate searching and guarding capacity. However, larger size and a higher proportion of muscle tissue requires more energy to maintain as shown by higher BMR_{min} in *A. flavicollis*, and activity patterns must therefore also remain high to ensure sufficient foraging effort for self maintenance. As daily heterotherms rely primarily on consistent food sources rather than their minimal fat reserves for their primary source of energy, deep torpor may not be beneficial in the heavier mice due to a shift in catabolic processes from lipid to protein metabolism [171]. However, some heavier female *A flavicollis* also appear to be more similar to males and vice versa in terms of their thermal strategy, suggesting the effect of sex is likely secondary to m_b. The interaction of sex and m_b together may therefore ultimately determine the phenotype of each individual within the population.

Androgenisation presents one mechanism to explain how these patterns arise. Greater ano-genital distance, a proxy for higher individual testosterone production in many rodents including *Apodemus sp.*, strongly correlates with body size in both sexes [43, 120]. A study by Cantoni et al. [43] found in California mice (*Peromyscus californicus*) that females born to litters with higher proportions of males were masculinised (larger body sizes and greater ano-genital distances), and close proximity to male foetus' *in utero* is known to result in more masculine physiological, behavioural and morphological phenotypes in adulthood, compared to females surrounded by other females [336]. Increased testosterone exposure in perinatal females is also known to masculinise regulatory centres of the brain such as the hypothalamus, which has direct role in torpor expression by regulating T3 thyroid hormone levels [23, 71, 72, 121, 237, 254]. Furthermore, exogenous testosterone administration results in the complete inhibition of daily torpor in Siberian hamsters [254, 289]. Given the

importance of hypothalmic function and testosterone in growth, development and torpor [72, 62, 134, 135], larger female *A. flavicollis* shown here to have a lower H_i similar to the average male could therefore adopt a more specialist, "male" thermal strategy if they too have been exposed to higher intrauterine testosterone levels. This suggests epigenetic modification in the early stages of development could play a larger role in determining thermal life-history strategies than previously thought.

Maintaining these higher body masses and entering prolonged torpor carries a large amount of risk in *A. flavicollis*. For example, as T_b during torpor ($T_{b\text{-}i}$) is a function of T_a in species with a high surface area:body mass ratio, high thermal inertia is needed following long torpor bouts to passively rewarm when an individual is larger/more massive [113]. Inability to do so can subsequently cause fatal hypothermia as arousal becomes increasingly difficult due a limited capacity for non-shivering thermogenesis [32, 133]. Higher $T_{b\text{-}i}$ and lower H_i in males and large females therefore allows a return to an active phase more easily following torpor, as they are closer to their normothermic condition, thus reducing this risk. The cost of torpor is still considerable, however, and although there appears to be no difference in the immediate costs of daily torpor between the sexes here [337], its effect on long term fitness is unknown. Male and large female *A. flavicollis* may use torpor as a secondary, "emergency-use" strategy which trades-off between conserving energy and its associated costs, when nearing their operational thermal limits, or starvation [76, 136, 177, 245, 298, 344]. The benefits of continued foraging may be greater than the benefits of torpor for these thermal specialists [291], and the observed variation in *A. flavicollis* may then be a result of stabilising selection towards intermediate-low values for thermal specialists that are more likely to be male, and neutral selection in thermal generalists that are more likely to be female. Future, more long term studies observing a wider range of phenotypes could uncover selective differences in heterothermy in a wild population, and elucidate the long term benefits and costs of daily torpor on fitness.

4.4.2 Seasonal variation in BMR

BMR_{min} was found to vary significantly between mice trapped in autumn versus winter/spring in Białowieża Forest, reflecting short term adaptability to environmental variation. This has also been demonstrated in a number of free-living mammals, birds, fish, and invertebrates, as well as humans [146, 175, 186, 244, 283, 346]. For example, Zub et al. [359] also found in a wild population of weasels (*Mustela nivalis*) from Białowieża forest, that variation in resting metabolic rate (RMR) was partially explained by the variation in an individuals permanent environment, which was highly heritable (0.40-0.54). However, permanent environmental effects were small, and did not lower the estimates of additive genetic variance considerably when included in models to estimate the heritability of RMR. As metabolic rate, whether measured as BMR_{min} or RMR, is a complex and highly plastic physiological trait correlated with both environmental variables such as photoperiod, temperature and food availability, as well as secondary phenotypes such m_b, age and sex [223], it is surprising that Zub et al.'s estimates of the heritability of RMR are so high given weasels were not allowed sufficient time to acclimatise to a laboratory environment as in previous studies with similarly high estimates [89, 169, 242, 293]. It could then be argued that high estimates of heritability, when the environment is not strictly controlled for, could then be an artefact of behavioural syndromes and individual stress responses, which may bias estimates of metabolic parameters [14, 44, 179]. Though, it must be noted others have suggested BMR can still evolve independently of physiological and behavioural traits [34, 218]. Unfortunately, low inter-individual variation in BMR_{min} due to low sampling resolution here meant τ and h^2_{SNP} could not be estimated, thus making a direct comparison with the heritability of metabolic parameters from these studies difficult.

In addition to the potential functional link between behaviour and metabolism, one must also consider whether inter-individual variation in metabolic rate is determined by life history strategy. For example, Auer et al. [16] show in brown trout that individual metabolic responses to variation in food availability was markedly different among fish, and consequently affected growth rates; fish that increased metabolic rates the most also grew the fastest in response to greater food availability, but then grew slower than fish whose

metabolic rate was less labile when food was not available *ad libitum*. This suggests the existence of two distinct strategies similar to that found in heterothermic responses in *A. flavicollis* here, and the choice of strategy could affect survival as mice able to mature quicker while food is plentiful in the autumn, may then be advantaged as they have a greater foraging capacity during winter [120]. This could explain why mice caught in autumn were on average smaller/lighter than those caught in winter/spring ($\mu_{aut} = 34.1g,\ sd = 7.23g,\ \mu_{wint/spr} = 39.2h,\ sd = 6.58$; $T_{26.6} = -2.79,\ p = 0.01$), as a greater proportion of mice caught in the autumn may still be growing when food is more plentiful compared to individuals caught in winter/spring which may have matured quickly. This assumes however, that the probability of catching large or small mice in winter/spring is a function of survival probability which may not be the case, and increased trapping effort in future would better elucidate the patterns in the seasonal variation of BMR_{min}.

4.4.3 Heritability of heterthermic responses

Repeatability is often considered the upper limit of heritability in evolutionary studies [98]. By this definition, true narrow-sense heritability (h^2), which accounts for variation resulting from additive genetic effects, should be lower than the repeatability of that trait, τ (formula 4.7), as the proportion of the variation attributed to genetic effects is lower than the overall individual phenotypic variance (V_I in formula 4.2).

$$h^2_{SNP} \leq h^2 \leq H^2_{SNP} \leq \tau \tag{4.7}$$

h^2_{SNP}, which is the phenotypic variance explained by common SNPs, should be lower still as not all SNPs responsible for the observed phenotypic variation are captured during sequencing due to sequencing error and allele dropout resulting from the characteristics of the chosen sequencing protocol [11].

Interestingly, h^2_{SNP} was estimated as higher than τ in this study when estimated for both sexes together, and for females alone. This could be due to some heteroskedasticity and deviation from normality of the residuals in the models (appendix A.3, figures A.2 - A.15,

residuals vs. fitted plots, scale-location plots and Q-Q plots), though, as this represents only a small number of measurements, its effect is likely to be small, and as shown in table 4.2, the difference between τ and h^2_{SNP} is not significant as their confidence intervals broadly overlap.

Under specific conditions, τ can underestimate the upper limit of h^2_{SNP} [88]. For example, calculating τ assumes that measures of fixed effects are independent and free from multi-collinearity. Although multi-colinearity was low between Sex, m_b and BMR_{min} here (appendix A.3, figures A.2 - A.15, plots of variance inflation factors for multi-colinearity), unmeasured secondary phenotypes could also correlate with these fixed effects and H_i. Male mice for example, can be heavier than females due to a higher proportion of muscle mass [24]. However, mice may also be heavier if they have a high proportion of body fat which was not measured in this study. As the metabolic rate can vary between tissue types [93, 109], large variation and multi-colinearity of unmeasured phenotypes such as body fat content/body condition with sex, m_b, BMR_{min} or H_i could non-randomly affect the estimates of the repeatability and heritability of heterothermic responses.

Environmental variation could also affect phenotypic variation [88]. Micro-environmental differences were strictly controlled however, so short term environmental variation on each mouse should not bias the estimates for τ or h^2_{SNP}. This is supported by non-significance in mass-adjusted H_i among trapping years, indicating short-term environmental fluctuations are unlikely to influence heterothermic responses. However, elements of each individuals permanent environment can have lasting effects on an individuals phenotype. One example, already mentioned in section 4.4.1, is the maternal effect of intrauterine testosterone exposure and masculinisation of females. Another example is the effect of nutrition on height in young and adolescent humans, where poor nutrition early in life was found to cause a 20cm difference in height among adults [66]. The latter could be due to a strong genotype-environment interaction and differences in sensitivity to environmental variation. Similar interactions could also occur in *A. flavicollis* and affect H_i.

Such complex interactions between multiple variables when estimating heritability are considered a violation of the assumption of independence for each factor. This assumption is made to allow the estimation of variance components, and differentiate between true additive

genetic variance, dominance effects and environmental variability. The reality however, is that highly complex interactions can occur between single Mendelian alleles, multiple genes and the environment. Estimating variance components accurately would therefore require an infinite number of partitions which is simply impossible. The estimate of h^2_{SNP} given here, is therefore an over-estimate of the true heritability of heterothermic responses, and explains why repeatability may not be its upper limit here. This over-estimation is further compounded by the low power of this study to accurately estimate dominance effects. However, given relatedness was taken into account to remove as much common environmental effects as possible, h^2_{SNP} was still high, and indicates that heterothermic responses could still be subject to natural selection. These estimates of heritability could then be considered a ceiling based on the data presented here, and it is clear that the heterothermic responses of each individual have a strong genetic component, though to what degree still remains unclear. Increased sampling effort in future would allow further variance partitioning to estimate how much variation is due to dominance effects, and thus provide a more accurate estimate of h^2_{SNP}.

Chapter 5

Conclusions and future directions

5.1 The genetic revolution brings us to the wild frontier of model organism research

Since Darwin and Wallace first embarked on their expeditions to study natural variation in the wild, scientists have attempted to uncover the rules which govern how such vast amounts diversity are maintained. It is via model organisms such as *Drosophila melanogaster*, *Mus musculus* and *Arabidopsis thaliana*, that we can test our hypotheses in simplified laboratory studies. Using such models, we have since discovered: how selection can drive phenotypic change [48], how model populations can diverge [85, 167], and even how entire model ecosystems can evolve in response to rapidly changing environmental conditions [107, 108]. Despite the power and many advantages of these studies [169], laboratory model systems are a product of their artificial environment often resulting in very different conclusions from studies of natural/wild study systems [319]; to quote Professor Sarah M. Farris, they "provide an extremely high resolution picture, but a very narrow field of view".

Early population genetics research was limited to studies of phenotypic variation controlled by segregating Mendelian loci in laboratory strains of model organisms (e.g. recessive lethal mutations in *Drosophila* [195]). Although there was a general consensus on the effect these loci have on phenotypic variation, population geneticists struggled to interpret their

observations at the level of the gene due to technological limitations and low resolution. Consequently, attempts at high-resolution studies of population genetics were almost abandoned [190]. As molecular techniques became more sophisticated however, the discovery of highly polymorphic markers such as microsatellites and single nucleotide polymorphisms re-invigorated the field, and biologists began searching for more *natural* models in an attempt to place them in the context of their natural history [6].

Modern techniques such as RAD-seq are now routinely used to produce high-density marker panels for almost any species, without *a priori* knowledge of the genome [20]. This thesis has demonstrated the power of RAD-seq to conduct population genomics research in a natural system.

5.1.1 On the power and pitfalls of RAD-seq. Methodological considerations for a new model system

The versatility of RAD-seq has made it increasingly popular among ecological and evolutionary biologists working in non-model organisms [51, 78, 222]. The scientific community has therefore attempted to extensively validate the protocol [12, 197, 272], and many iterations have since followed to address its limitations [11]. For example, although the original single-digest RAD-seq can provide a cost effective solution for large scale genomics research, it still requires an expensive sonicator to further fragment restricted DNA. Furthermore, it offers little control over the final library sizes, and requires careful refinement to ensure the desired proportion of the genome is represented [128]. Double-digest RAD-seq was developed specifically to improve upon this, though it lacked the ability to identify PCR artefacts such as duplicate and chimeric reads. It also did not have the capacity to multiplex large numbers of samples to reduce costs further [106, 261]. The modified quaddRAD protocol used to prepare libraries here, has now resolved these issues by using a quadruple indexing system to allow up to 192 samples on a single sequencing lane, whilst also allowing to identify PCR artefacts [106, 213].

Despite these clear advantages, particular care must be taken when designing any RAD-seq experiment. For example, although the quadruple indexing system allows for multiplexing 192 samples, such a large number on a single sequencing lane is ill-advised as the mean per-sample coverage decreases, and could result in allele imbalance and dropout [312]. The choice of restriction enzymes is also critical to ensuring minimal bias in the final libraries [75], and *in silico* restriction digestion as conducted in chapter 2, is therefore an invaluable method to predict the number and distribution of loci before commencing the experiment [187, 332]. Even if conducted in relatively phylogenetically distant species as may be required when working *de novo*, simulated RAD-seq experiments can still provide useful insights for planning the protocol.

Although RAD-seq has, to some degree, *democratised* the field of population genomics [51], the design of any experiment must always be considered in the context of the study aims and study system. For example, despite the order of magnitude fewer loci obtained from amplicon-based markers such as microsatellites, they still remain a viable option for the many study systems in which such resources are already available, and in which the analyses do not require high-density loci [103]. Pedigree analyses are one example where microsatellites can perform as equally well as SNPs in large populations, with significantly fewer loci [104, 196]. However, in study systems where heterozygosity is low, as is expected in studies of inbreeding for conservation genetics, microsatellites may not have the resolving power to discern familial relationships with high confidence [184]. Furthermore, genomic resources are often limited in rare and *non-model* species, making microsatellite marker systems expensive and laborious to develop [103, 138].

Although microsatellite markers have previously been used in *Apoedmus* [123, 129, 204, 203], they lack the necessary power for high resolution studies of allele frequency dynamics and selection as conducted in chapter 3. There of course remains the possibility that large numbers of SNPs may exhibit high levels of linkage disequilibrium due to the choice of enzyme combinations, which may then limit the power to detect selection [197]. However, strict quality control procedures ensure that even with some linkage disequilibrium, loci under selection can still be identified [51, 140, 286]. The result in this study, is a clear signal of

selection despite high levels of drift, inbreeding and migration. RAD-seq has therefore proven itself once again, as an incredibly powerful tool for ecological and population genomics research.

5.2 Is another model system needed? A Comparison with other classical models

The famous ungulate model systems of the Scottish islands such as Isle of Rum red deer and Soay sheep on St Kilda, are a classic example of already well established free-living models for ecological and population genomics research. Until its cessation in 1972, the red deer population of the Northern Block on Rum was intensively managed (overwinter feeding by deer managers), and actively culled [260]. The population has since grown to over 900 individuals, though its density appears to be no different from other managed populations still present on the island [260]. Thus, the Isle of Rum red deer form an excellent model system of the ecological and evolutionary dynamics in a stable population which has experienced ecological release in the past. Furthermore, the differences in environmental sensitivities of stags and hinds have made sexual dimorphism and fitness effects in the wild a central theme among the hundreds of published articles about the population [52, 63, 154].

With the exception of the ease of trapping and transportablity to conduct measurements in mice, chapters 3 and 4 demonstrate studies of sexually dimorphic phenotypes and allele frequency dynamics in *A. flavicollis* are comparable to other model systems [127, 153, 60]. For example, the Soay sheep of St Kilda also exhibit population cycles which are determined by the severity of winter conditions [68, 269]. However, due to the considerably shorter generation time of *A. flavicollis*, the genomic signature left by these population maxima and minima is highly accelerated (detectable in just three years here). Longer term studies may then reveal evolutionary trends which would otherwise take considerably longer to observe in models of large herbivores. It is clear the determinants of the dynamics of all these systems including *A. flavicollis* appear to be dictated by multiple factors: winter carrying capacity, age, sex, food availability and stochasticity [68, 127, 269].

Unlike the allopatric ungulate populations however, free-living populations in unbounded habitats such as Białowieża Forest are not restricted by the study site or island. Furthermore, many species in nature are not monogamous like the Florida Scrub Jay [60], as is evident in chapter 3. Wild pedigrees in a small, polygamous species such as *A. flavicollis*, are thus impossible to verify with field observations. The results therefore incorporate considerable amounts of uncertainty and missing data. Though, with advancements in molecular pedigree reconstruction and statistical genetics, this limitation can be accounted for successfully as was done here using SEQUOIA [142]. Consequently, and despite the effects of gene flow from neighbouring (sub-)populations, the signatures of drift and selection are both clearly evident, and likely act in conjunction to stabilise the population. This suggests *A. flavicollis* has an extraordinary resilience to perturbation, and can adapt to persist under harsh conditions. From a conservation perspective, the population from Białowieża Forest may then also offer insights into the effects of inbreeding and genetic rescue in declining populations. The model cannot then be considered atypical of other *natural* populations without further research. The genus *Apodemus* is therefore an excellent new model system to complement those which are already established, and should be considered by researchers in future to conduct ecological and population genomics in the wild.

5.3 Summary of findings from a new model system

5.3.1 The importance of scale in evolutionary studies

I have shown here the importance of scale in evolutionary genomics studies, where processes at the individual and population level can reveal contrasting patterns. For example, as expected in a population which experiences severe population crashes due to harsh winter conditions, many *A. flavicollis* in Białowieża National Park do show signs of inbreeding and reduced genetic diversity. Furthermore, some individuals appear to be making considerably greater genetic contributions to the next generation than others. We can then predict, that at a population scale, allele frequencies should be changing considerably due to genetic drift following population declines. However, on average, inbreeding is low, genetic diversity is

high, and allele frequencies are as expected under Hardy-Weinberg equilibrium. Although some allele frequencies are changing considerably, the majority remain largely stable. This is a pattern which is indicative of a panmictic population experiencing large amounts of migration from surrounding areas, which could act to rescue local (sub-)populations following the winter decline. *A. flavicollis* could therefore have a very large effective population size, and may be considered as a *meta-population*.

Confirming this pattern of decline and rescue in future could involve sampling along a large transect ($> 500km$) during multiple seasons and years, to reveal both local and population level responses to the periodicity in population size observed in Białowieża National Park [269]. This could also uncover a pattern of isolation by distance between (sub-)populations over short time periods, and confirm whether the *founder/flush* model can explain the genetic variation and allele frequency dynamics described here [192, 310].

5.3.2 A phenotypic *continuum* only partly explains the natural variation of heterothermic responses in *A. flavicollis*

The variability in genetic and genealogical contributions to future generations also suggests that phenotypes which increase the probability of survival could be maintained in the population due to natural selection. For example, *A. flavicollis*, as a heterothermic species, is able to minimise its energy budget by entering into a temporarily torpid state when food is restricted. These heterothermic responses were found to be highly heritable here, and previous studies have described a *continuum* of phenotypes that are present in the population [32]. This suggests that a higher heterothermy index may provide some evolutionary advantage for individuals able to more efficiently allocate resources to self maintenance when conditions are harsh. However, a phenotypic continuum only partially explains the variation in heterothermy, which appears to be stratified into two distinct thermal strategies in *A. flavicollis*.

Multiple interacting phenotypes, which account for much of the phenotypic variation of heterothermic responses, could minimise the physiological cost associated with entry

into deep torpor by constraining heterothermic responses [15], and are largely responsible for which thermal strategy is undertaken. The two strategies are: thermal specialists, characterised by larger/heavier mice that are mostly, but not exclusively, male, and thermal generalists, characterised by smaller/lighter mice that are mostly, but not exclusively, female, and appears to best explain the majority of the variation of heterothermy in the population. As heterothermic responses are highly heritable, the presence of these two strategies suggests natural selection may act to stabilise thermal specialists towards intermediate-low values of the heterothermy index, but may be acting neutrally to allow greater phenotypic plasticity in thermal generalists. This should reduce the costs associated with deep torpor and high body mass in thermal specialists, and minimise the energy deficit in lighter thermal generalists.

Future work on thermal strategies and torpor in *A. flavicollis*, should concentrate on phenotyping enough individuals for multiple traits including heterothermic responses in the population, to ensure sufficient power to conduct a genome wide association study [335]. The model should take into account any correlation between these phenotypes [81], and specifically test for sexual dimorphism which appears to be responsible for much of the phenotypic variation. If alleles are found to segregate in sexually divergent phenotypes, validation could uncover the differences in complex molecular mechanisms involving significant genetic, hormonal and epigenetic control of torpor between the sexes. This would then link molecular, individual and population level processes responsible for natural variation in fitness and torpor.

References

[1] Abecasis, G. R., Cookson, W. O., and Cardon, L. R. (2000). Pedigree tests of transmission disequilibrium. *European Journal of Human Genetics*, 8(7):545–551.

[2] Abolins, S., Lazarou, L., Weldon, L., Hughes, L., King, E. C., Drescher, P., Pocock, M. J., Hafalla, J. C., Riley, E. M., and Viney, M. (2018). The ecology of immune state in a wild mammal, mus musculus domesticus. *PLoS biology*, 16(4):e2003538.

[3] Abreu-Vieira, G., Xiao, C., Gavrilova, O., and Reitman, M. L. (2015). Integration of body temperature into the analysis of energy expenditure in the mouse. *Molecular metabolism*, 4(6):461–470.

[4] Adams, J. R., Vucetich, L. M., Hedrick, P. W., Peterson, R. O., and Vucetich, J. A. (2011). Genomic sweep and potential genetic rescue during limiting environmental conditions in an isolated wolf population. *Proceedings of the Royal Society B: Biological Sciences*, 278(1723):3336–3344.

[5] Albery, G. F., Kenyon, F., Morris, A., Morris, S., Nussey, D. H., and Pemberton, J. M. (2018). Seasonality of helminth infection in wild red deer varies between individuals and between parasite taxa. *Parasitology*, 145(11):1410–1420.

[6] Alfred, J. and Baldwin, I. T. (2015). The natural history of model organisms: new opportunities at the wild frontier. *Elife*, 4:e06956.

[7] Ali, J. R. and Vences, M. (2019). Mammals and long-distance over-water colonization: The case for rafting dispersal; the case against phantom causeways. *Journal of Biogeography*, 46(11):2632–2636.

[8] Almudevar, A. (2003). A simulated annealing algorithm for maximum likelihood pedigree reconstruction. *Theoretical population biology*, 63(2):63–75.

[9] Anderson, E. C. and Garza, J. C. (2006). The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, 172(4):2567–2582.

[10] Andrews, K. R., Adams, J. R., Cassirer, E. F., Plowright, R. K., Gardner, C., Dwire, M., Hohenlohe, P. A., and Waits, L. P. (2018). A bioinformatic pipeline for identifying informative snp panels for parentage assignment from rad seq data. *Molecular ecology resources*, 18(6):1263–1281.

[11] Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of radseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2):81.

[12] Andrews, K. R., Hohenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., and Luikart, G. (2014). Trade-offs and utility of alternative radseq methods: Reply to puritz et al. *Molecular ecology*, 23(24):5943–5946.

[13] Andrews, S. (2010). *A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[14] Anestis, A., Lazou, A., Pörtner, H. O., and Michaelidis, B. (2007). Behavioral, metabolic, and molecular stress responses of marine bivalve mytilus galloprovincialis during long-term acclimation at increasing ambient temperature. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 293(2):R911–R921.

[15] Angilletta Jr, M. J., Bennett, A. F., Guderley, H., Navas, C. A., Seebacher, F., and Wilson, R. S. (2006). Coadaptation: a unifying principle in evolutionary thermal biology. *Physiological and Biochemical Zoology*, 79(2):282–294.

[16] Auer, S. K., Salin, K., Rudolf, A. M., Anderson, G. J., and Metcalfe, N. B. (2015). Flexibility in metabolic rate confers a growth advantage under changing food availability. *Journal of Animal Ecology*, 84(5):1405–1411.

[17] Baalsrud, H. T., Sæther, B.-E., Hagen, I. J., Myhre, A. M., Ringsby, T. H., Pärn, H., and Jensen, H. (2014). Effects of population characteristics and structure on estimates of effective population size in a house sparrow metapopulation. *Molecular ecology*, 23(11):2653–2668.

[18] Bahram, S. and Inoko, H. (2007). Microsatellite markers for genome-wide association studies. *Nature Reviews Genetics*, 8(2):164–164.

[19] Bailey-Wilson, J. E., Almasy, L., De Andrade, M., Bailey, J., Bickeböller, H., Cordell, H. J., Daw, E. W., Goldin, L., Goode, E. L., Gray-McGuire, C., et al. (2005). Genetic analysis workshop 14: microsatellite and single-nucleotide polymorphism marker loci for genome-wide scans.

[20] Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A. (2008). Rapid snp discovery and genetic mapping using sequenced rad markers. *PloS one*, 3(10):e3376.

[21] Bamford, D. R., Pemberton, J., Albon, S., Robertson, A., Maccoll, A., Smith, J., Stevenson, I. R., and Clutton-Brock, T. H. (1995). Molecular genetic variation and individual survival during population crashes of an unmanaged ungulate population. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 347(1321):263–273.

[22] Banerjee, A. and Dave, R. N. (2004). Validating clusters using the hopkins statistic. In *2004 IEEE International conference on fuzzy systems (IEEE Cat. No. 04CH37542)*, volume 1, pages 149–153. IEEE.

[23] Bank, J. H., Kemmling, J., Rijntjes, E., Wirth, E. K., and Herwig, A. (2015). Thyroid hormone status affects expression of daily torpor and gene transcription in djungarian hamsters (phodopus sungorus). *Hormones and behavior*, 75:120–129.

[24] Banu, J., Wang, L., and Kalu, D. (2003). Effects of increased muscle mass on bone in male mice overexpressing igf-i in skeletal muscles. *Calcified Tissue International*, 73(2):196–201.

[25] Bartolommei, P., Sozio, G., Bencini, C., Cinque, C., Gasperini, S., Manzo, E., Prete, S., Solano, E., Cozzolino, R., and Mortelliti, A. (2016). Field identification of apodemus flavicollis and apodemus sylvaticus: a quantitative comparison of different biometric measurements. *Mammalia*, 80(5):541–547.

[26] Bernadzki, E., Bolibok, L., Brzeziecki, B., Zajaczkowski, J., and Żybura, H. (1998). Compositional dynamics of natural forests in the bialowieza national park, northeastern poland. *Journal of Vegetation Science*, 9(2):229–238.

[27] Berry, R. and Jakobson, M. (1975). Ecological genetics of an island population of the house mouse (mus musculus). *Journal of Zoology*, 175(4):523–540.

[28] Berthier, K., Charbonnel, N., Galan, M., Chaval, Y., and Cosson, J.-F. (2006). Migration and recovery of the genetic diversity during the increasing density phase in cyclic vole populations. *Molecular Ecology*, 15(9):2665–2676.

[29] Birzu, G., Matin, S., Hallatschek, O., and Korolev, K. S. (2019). Genetic drift in range expansions is very sensitive to density dependence in dispersal and growth. *Ecology Letters*, 22(11):1817–1827.

[30] Bobiec, A. (2002). Living stands and dead wood in the białowieża forest: suggestions for restoration management. *Forest Ecology and Management*, 165(1-3):125–140.

[31] Boratyński, J. S., Iwińska, K., and Bogdanowicz, W. (2018). Body temperature variation in free-living and food-deprived yellow-necked mice sustains an adaptive framework for endothermic thermoregulation. *Mammal Research*, 63(4):493–500.

[32] Boratyński, J. S., Iwińska, K., and Bogdanowicz, W. (2019). An intra-population heterothermy continuum: notable repeatability of body temperature variation in food-deprived yellow-necked mice. *Journal of Experimental Biology*, 222(6).

[33] Boutin, S. (1995). Testing predator–prey theory by studying fluctuating populations of small mammals. *Wildlife Research*, 22(1):89–99.

[34] Bouwhuis, S., Quinn, J. L., Sheldon, B. C., and Verhulst, S. (2014). Personality and basal metabolic rate in a wild bird population. *Oikos*, 123(1):56–62.

[35] Boyles, J. G., Smit, B., and McKechnie, A. E. (2011). A new comparative metric for estimating heterothermy in endotherms. *Physiological and Biochemical Zoology*, 84(1):115–123.

[36] Boyles, J. G., Thompson, A. B., McKechnie, A. E., Malan, E., Humphries, M. M., and Careau, V. (2013). A global heterothermic continuum in mammals.

[37] Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635.

[38] Bresadola, L., Link, V., Buerkle, C. A., Lexer, C., and Wegmann, D. (2020). Estimating and accounting for genotyping errors in rad-seq experiments. *Molecular Ecology Resources*.

[39] Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., and Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *The American Journal of Human Genetics*, 62(6):1408–1415.

[40] Bryja, J., Patzenhauerová, H., Albrecht, T., Mošanskỳ, L., Stanko, M., and Stopka, P. (2008). Varying levels of female promiscuity in four apodemus mice species. *Behavioral Ecology and Sociobiology*, 63(2):251–260.

[41] Bulmer, M. (1973). Inbreeding in the great tit. *Heredity*, 30(3):313–325.

[42] Calsbeek, R. and Irschick, D. J. (2007). The quick and the dead: correlational selection on morphology, performance, and habitat use in island lizards. *Evolution: International Journal of Organic Evolution*, 61(11):2493–2503.

[43] Cantoni, D., Glaizot, O., and Brown, R. E. (1999). Effects of sex composition of the litter on anogenital distance in california mice (peromyscus californicus). *Canadian Journal of Zoology*, 77(1):124–131.

[44] Careau, V., Thomas, D., Humphries, M., and Réale, D. (2008). Energy metabolism and animal personality. *Oikos*, 117(5):641–653.

[45] Carey, H. V., Andrews, M. T., and Martin, S. L. (2003). Mammalian hibernation: cellular and molecular responses to depressed metabolism and low temperature. *Physiological reviews*.

[46] Cariou, M., Duret, L., and Charlat, S. (2013). Is rad-seq suitable for phylogenetic inference? an in silico assessment and optimization. *Ecology and evolution*, 3(4):846–852.

[47] Cariou, M., Duret, L., and Charlat, S. (2016). How and how much does rad-seq bias genetic diversity estimates? *BMC evolutionary biology*, 16(1):1–8.

[48] Castro, J. P., Yancoskie, M. N., Marchini, M., Belohlavy, S., Hiramatsu, L., Kučka, M., Beluch, W. H., Naumann, R., Skuplik, I., Cobb, J., et al. (2019). An integrative genomic analysis of the longshanks selection experiment for longer limbs in mice. *Elife*, 8:e42014.

[49] Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, 22(11):3124–3140.

[50] Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics*, 1(3):171–182.

[51] Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., and Allendorf, F. W. (2017). Unbroken: Radseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular ecology resources*, 17(3):362–365.

[52] Catchpole, E. A., Fan, Y., Morgan, B. J., Clutton-Brock, T., and Coulson, T. (2004). Sexual dimorphism, survival and dispersal in red deer. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(1):1–26.

[53] Cerezo, M. L. M., Kucka, M., Zub, K., Chan, Y. F., and Bryk, J. (2020). Population structure of apodemus flavicollis and comparison to apodemus sylvaticus in northern poland based on rad-seq. *BMC genomics*, 21(1):1–14.

[54] Chakraborty, R. and Nei, M. (1982). Genetic differentiation of quantitative characters between populations or species: I. mutation and random genetic drift. *Genetics Research*, 39(3):303–314.

[55] Charlesworth, D. and Willis, J. H. (2009). The genetics of inbreeding depression. *Nature reviews genetics*, 10(11):783–796.

[56] Charmantier, A. and Reale, D. (2005). How do misassigned paternities affect the estimation of heritability in the wild? *Molecular ecology*, 14(9):2839–2850.

[57] Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61(6):1–36.

[58] Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. (1996). Accessing genetic information with high-density dna arrays. *Science*, 274(5287):610–614.

[59] Chen, B., Cole, J. W., and Grond-Ginsbach, C. (2017). Departure from hardy weinberg equilibrium and genotyping error. *Frontiers in genetics*, 8:167.

[60] Chen, N., Juric, I., Cosgrove, E. J., Bowman, R., Fitzpatrick, J. W., Schoech, S. J., Clark, A. G., and Coop, G. (2019). Allele frequency dynamics in a pedigreed natural population. *Proceedings of the National Academy of Sciences*, 116(6):2158–2164.

[61] Chen, Y.-S., Su, Y.-C., and Pan, W. (2016). Effect of spatial constraints on hardy-weinberg equilibrium. *Scientific reports*, 6(1):1–10.

[62] Clarkson, J. and Herbison, A. E. (2016). Hypothalamic control of the male neonatal testosterone surge. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1688):20150115.

[63] Clutton-Brock, T. H., Guinness, F. E., and Albon, S. D. (1982). *Red deer: behavior and ecology of two sexes*. University of Chicago press.

[64] Cohen, S. B. and Dor, R. (2018). Phenotypic divergence despite low genetic differentiation in house sparrow populations. *Scientific reports*, 8(1):1–12.

[65] Consortium, I. H. G. S. et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409:860–921.

[66] Cooper, C., Dennison, E., Fall, C., and Osmond, C. (2020). Height and body-mass index trajectories of school-aged children and adolescents from 1985 to 2019 in 200 countries: Pooled analysis of 2,182 population-based studies with 65 million participants. *The Lancet*.

[67] Corl, A., Bi, K., Luke, C., Challa, A. S., Stern, A. J., Sinervo, B., and Nielsen, R. (2018). The genetic basis of adaptation following plastic changes in coloration in a novel environment. *Current Biology*, 28(18):2970–2977.

[68] Coulson, T., Catchpole, E. A., Albon, S. D., Morgan, B. J., Pemberton, J., Clutton-Brock, T. H., Crawley, M., and Grenfell, B. T. (2001). Age, sex, density, winter weather, and population crashes in soay sheep. *Science*, 292(5521):1528–1531.

[69] Coulson, T., Milner-Gulland, E., and Clutton-Brock, T. (2000). The relative roles of density and climatic variation on population dynamics and fecundity rates in three contrasting ungulate species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1454):1771–1779.

[70] Crompton, A., Taylor, C. R., and Jagger, J. A. (1978). Evolution of homeothermy in mammals. *Nature*, 272(5651):333–336.

[71] Cubuk, C., Kemmling, J., Fabrizius, A., and Herwig, A. (2017a). Transcriptome analysis of hypothalamic gene expression during daily torpor in djungarian hamsters (phodopus sungorus). *Frontiers in neuroscience*, 11:122.

[72] Cubuk, C., Markowsky, H., and Herwig, A. (2017b). Hypothalamic control systems show differential gene expression during spontaneous daily torpor and fasting-induced torpor in the djungarian hamster (phodopus sungorus). *PloS one*, 12(10):e0186299.

[73] Czarnomska, S. D., Niedziałkowska, M., Borowik, T., and Jędrzejewska, B. (2018). Regional and local patterns of genetic variation and structure in yellow-necked mice-the roles of geographic distance, population abundance, and winter severity. *Ecology and evolution*, 8(16):8171–8186.

[74] Czeszczewik, D., Czortek, P., Jaroszewicz, B., Zub, K., Rowiński, P., and Walankiewicz, W. (2020). Climate change has cascading effects on tree masting and the breeding performance of a forest songbird in a primeval forest. *Science of The Total Environment*, 747:142084.

[75] DaCosta, J. M. and Sorenson, M. D. (2014). Amplification biases and consistent recovery of loci in a double-digest rad-seq protocol. *PloS one*, 9(9):e106713.

[76] Dammhahn, M., Landry-Cuerrier, M., Réale, D., Garant, D., and Humphries, M. M. (2017). Individual variation in energy-saving heterothermy affects survival and reproductive success. *Functional Ecology*, 31(4):866–875.

[77] Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., et al. (2011). The variant call format and vcftools. *Bioinformatics*, 27(15):2156–2158.

[78] Davey, J. W. and Blaxter, M. L. (2010). Radseq: next-generation population genetics. *Briefings in functional genomics*, 9(5-6):416–423.

[79] Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., and Blaxter, M. L. (2013). Special features of rad sequencing data: implications for genotyping. *Molecular ecology*, 22(11):3151–3164.

[80] Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7):499–510.

[81] De Villemereuil, P., Morrissey, M. B., Nakagawa, S., and Schielzeth, H. (2018). Fixed-effect variance and the estimation of repeatabilities and heritabilities: Issues and solutions. *Journal of Evolutionary Biology*, 31(4):621–632.

[82] Dean, R. and Mank, J. E. (2014). The role of sex chromosomes in sexual dimorphism: discordance between molecular and phenotypic data. *Journal of Evolutionary Biology*, 27(7):1443–1453.

[83] DeAngelis, M. M., Wang, D. G., and Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of pcr products. *Nucleic acids research*, 23(22):4742.

[84] DeWoody, J. and Avise, J. (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of fish biology*, 56(3):461–473.

[85] Diamond, S. E., Hawkins, S. D., Nijhout, H. F., and Kingsolver, J. G. (2010). Evolutionary divergence of field and laboratory populations of manduca sexta in response to host-plant quality. *Ecological Entomology*, 35(2):166–174.

[86] Díaz-Arce, N. and Rodríguez-Ezpeleta, N. (2019). Selecting rad-seq data analysis parameters for population genetics: the more the better? *Frontiers in genetics*, 10:533.

[87] Dodge, Y. and Commenges, D. (2006). *The Oxford dictionary of statistical terms*. Oxford University Press on Demand.

[88] Dohm, M. (2002). Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology*, pages 273–280.

[89] Dohm, M. R., Hayes, J. P., and Garland Jr, T. (2001). The quantitative genetics of maximal and basal rates of oxygen consumption in mice. *Genetics*, 159(1):267–277.

[90] Dreger, D. L., Rimbault, M., Davis, B. W., Bhatnagar, A., Parker, H. G., and Ostrander, E. A. (2016). Whole-genome sequence, snp chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. *Disease Models & Mechanisms*, 9(12):1445–1460.

[91] Eaton, D. A. (2014). Pyrad: assembly of de novo radseq loci for phylogenetic analyses. *Bioinformatics*, 30(13):1844–1849.

[92] Ehrich, D. and Jorde, P. E. (2005). High genetic variability despite high-amplitude population cycles in lemmings. *Journal of Mammalogy*, 86(2):380–385.

[93] Elia, M. (1992). Energy expenditure in the whole body. *Energy metabolism: tissue determinants and cellular corollaries*, pages 19–59.

[94] Else, P. and Hulbert, A. (1981). Comparison of the" mammal machine" and the" reptile machine": energy production. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 240(1):R3–R9.

[95] Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., and Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the national academy of sciences*, 107(37):16196–16200.

[96] Eu-Ahsunthornwattana, J., Miller, E. N., Fakiola, M., Jeronimo, S. M., Blackwell, J. M., Cordell, H. J., 2, W. T. C. C. C., et al. (2014). Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genet*, 10(7):e1004445.

[97] Evans, A. L., Singh, N. J., Friebe, A., Arnemo, J. M., Laske, T., Fröbert, O., Swenson, J. E., and Blanc, S. (2016). Drivers of hibernation in the brown bear. *Frontiers in zoology*, 13(1):7.

[98] Falconer, D. S. and Mackay, T. F. (1996). *Introduction to quantitative genetics*. Longman, 4th edition.

[99] Faliński, J. B. (1986). Geobotany. In *Vegetation Dynamics in Temperate Lowland Primeval Forests*, pages 39–111. Springer.

[100] Felsenstein, J. (1993). *PHYLIP (phylogeny inference package), version 3.5 c*. Joseph Felsenstein.

[101] Ferrannini, E. (1988). The theoretical bases of indirect calorimetry: a review. *Metabolism*, 37(3):287–301.

[102] Fitzpatrick, S. W., Bradburd, G. S., Kremer, C. T., Salerno, P. E., Angeloni, L. M., and Funk, W. C. (2020). Genomic and fitness consequences of genetic rescue in wild populations. *Current Biology*, 30(3):517–522.

[103] Flanagan, S. P. and Jones, A. G. (2019). The future of parentage analysis: From microsatellites to snps and beyond. *Molecular Ecology*, 28(3):544–567.

[104] Forstmeier, W., Schielzeth, H., Mueller, J. C., Ellegren, H., and Kempenaers, B. (2012). Heterozygosity–fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Molecular Ecology*, 21(13):3237–3249.

[105] Fournier, T., Abou Saada, O., Hou, J., Peter, J., Caudal, E., and Schacherer, J. (2019). Extensive impact of low-frequency variants on the phenotypic landscape at population-scale. *Elife*, 8:e49258.

[106] Franchini, P., Monné Parera, D., Kautt, A. F., and Meyer, A. (2017). quaddrad: a new high-multiplexing and pcr duplicate removal ddrad protocol produces novel evolutionary insights in a nonradiating cichlid lineage. *Molecular Ecology*, 26(10):2783–2795.

[107] Friman, V.-P., Guzman, L. M., Reuman, D. C., and Bell, T. (2015). Bacterial adaptation to sublethal antibiotic gradients can change the ecological properties of multitrophic microbial communities. *Proceedings of the Royal Society B: Biological Sciences*, 282(1806):20142920.

[108] Friman, V.-P., Jousset, A., and Buckling, A. (2014). Rapid prey evolution can alter the structure of predator–prey communities. *Journal of evolutionary biology*, 27(2):374–380.

[109] Gallagher, D., Belmonte, D., Deurenberg, P., Wang, Z., Krasnow, N., Pi-Sunyer, F. X., and Heymsfield, S. B. (1998). Organ-tissue mass measurement allows modeling of ree and metabolically active tissue mass. *American Journal of Physiology-Endocrinology And Metabolism*, 275(2):E249–E258.

[110] Gardner, A. (2019). Fisher's reproductive value. In *Shackelford T., Weekes-Shackelford V. (eds) Encyclopedia of Evolutionary Psychological Science*. Springer.

[111] Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

[112] Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J.-M., and Estoup, A. (2013). The effect of rad allele dropout on the estimation of genetic variation within and between populations. *Molecular ecology*, 22(11):3165–3178.

[113] Geiser, F. (2001). Hibernation: endotherms. *e LS*.

[114] Geiser, F. (2004). Metabolic rate and body temperature reduction during hibernation and daily torpor. *Annu. Rev. Physiol.*, 66:239–274.

[115] Geiser, F. and Brigham, R. (2012). *Living in a seasonal world*. Springer.

[116] Geiser, F., Körtner, G., and Schmidt, I. (1998). Leptin increases energy expenditure of a marsupial by inhibition of daily torpor. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 275(5):R1627–R1632.

[117] Geiser, F. and Ruf, T. (1995). Hibernation versus daily torpor in mammals and birds: physiological variables and classification of torpor patterns. *Physiological Zoology*, 68(6):935–966.

[118] Gel, B. and Serra, E. (2017). karyoploter : an r bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*, 33(19):3088–3090.

[119] Godsall, B. (2015). *Mechanisms of space use in the wood mouse, Apodemus sylvaticus*. PhD thesis, Imperial College London.

[120] Godsall, B., Coulson, T., and Malo, A. F. (2014). From physiology to space use: energy reserves and androgenization explain home-range size variation in a woodland rodent. *Journal of Animal Ecology*, 83(1):126–135.

[121] Goel, N. and Bale, T. L. (2008). Organizational and activational effects of testosterone on masculinization of female physiological and behavioral stress responses. *Endocrinology*, 149(12):6399–6405.

[122] Gomes, I., Collins, A., Lonjou, C., Thomas, N., Wilkinson, J., Watson, M., and Morton, N. (1999). Hardy–weinberg quality control. *Annals of human genetics*, 63(6):535–538.

[123] Gortat, T., Gryczyńska-Siemiątkowska, A., Rutkowski, R., Kozakiewicz, A., Mikoszewski, A., and Kozakiewicz, M. (2010). Landscape pattern and genetic structure of a yellow-necked mouse apodemus flavicollis population in north-eastern poland. *Acta Theriologica*, 55(2):109–121.

[124] Grafen, A. (2006). A theory of fisher's reproductive value. *Journal of mathematical biology*, 53(1):15–60.

[125] Graffelman, J. and Moreno, V. (2013). The mid p-value in exact tests for hardy-weinberg equilibrium. *Statistical applications in genetics and molecular biology*, 12(4):433–448.

[126] Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R. G., Martino, J. A., Pierson, T., Rogers, S. M., et al. (2015). Impacts of degraded dna on restriction enzyme associated dna sequencing (rads eq). *Molecular Ecology Resources*, 15(6):1304–1315.

[127] Grenfell, B., Price, O., Albon, S., and Glutton-Brock, T. (1992). Overcompensation and population cycles in an ungulate. *Nature*, 355(6363):823–826.

[128] Grokhovsky, S. (2006). Specificity of dna cleavage by ultrasound. *Molecular Biology*, 40(2):276–283.

[129] Gryczyńska-Siemiątkowska, A., Gortat, T., Kozakiewicz, A., Rutkowski, R., Pomorski, J., and Kozakiewicz, M. (2008). Multiple paternity in a wild population of the yellow-necked mouseapodemus flavicollis. *Acta theriologica*, 53(3):251–258.

[130] Hallam, S. L. and Mzilikazi, N. (2011). Heterothermy in the southern african hedgehog, atelerix frontalis. *Journal of Comparative Physiology B*, 181(3):437–445.

[131] Hawkins, T. L., O'Connor-Morin, T., Roy, A., and Santillan, C. (1994). Dna purification and isolation using a solid-phase. *Nucleic acids research*, 22(21):4543.

[132] Heldmaier, G., Ortmann, S., and Elvert, R. (2004). Natural hypometabolism during hibernation and daily torpor in mammals. *Respiratory physiology & neurobiology*, 141(3):317–329.

[133] Heldmaier, G., Steinlechner, S., and Rafael, J. (1982). Nonshivering thermogenesis and cold resistance during seasonal acclimatization in the djungarian hamster. *Journal of comparative physiology*, 149(1):1–9.

[134] Herwig, A., Revel, F., Saboureau, M., Pévet, P., and Steinlechner, S. (2006). Daily torpor alters multiple gene expression in the suprachiasmatic nucleus and pineal gland of the djungarian hamster (phodopus sungorus). *Chronobiology international*, 23(1-2):269–276.

[135] Herwig, A., Saboureau, M., Pévet, P., and Steinlechner, S. (2007). Daily torpor affects the molecular machinery of the circadian clock in djungarian hamsters (phodopus sungorus). *European Journal of Neuroscience*, 26(10):2739–2746.

[136] Hetem, R. S., Maloney, S. K., Fuller, A., and Mitchell, D. (2016). Heterothermy in large mammals: inevitable or implemented? *Biological Reviews*, 91(1):187–205.

[137] Hirschhorn, J. N. and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature reviews genetics*, 6(2):95–108.

[138] Hodel, R. G., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., Gitzendanner, M. A., Douglas, N. A., Germain-Aubrey, C. C., Chen, S., et al. (2016). The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *Applications in plant sciences*, 4(6):1600025.

[139] Hodgkinson, A. and Eyre-Walker, A. (2010). Human triallelic sites: evidence for a new mutational mechanism? *Genetics*, 184(1):233–241.

[140] Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced rad tags. *PLoS genet*, 6(2):e1000862.

[141] Hopkins, B. and Skellam, J. G. (1954). A new method for determining the type of distribution of plant individuals. *Annals of Botany*, 18(2):213–227.

[142] Huisman, J. (2017). Pedigree reconstruction from snp data: parentage assignment, sibship clustering and beyond. *Molecular ecology resources*, 17(5):1009–1024.

[143] Huisman, J., Kruuk, L. E., Ellis, P. A., Clutton-Brock, T., and Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences*, 113(13):3585–3590.

[144] Humphries, M. M., Thomas, D. W., and Kramer, D. L. (2003). The role of energy availability in mammalian hibernation: a cost-benefit approach. *Physiological and Biochemical Zoology*, 76(2):165–179.

[145] Illumina (2016). *BCL2FASTQ conversion software*. https://jp.support.illumina.com/downloads/bcl2fastq-conversion-software-v2-18.html.

[146] Islam, M. J., Kunzmann, A., Bögner, M., Meyer, A., Thiele, R., and Slater, M. J. (2020). Metabolic and molecular stress responses of european seabass, dicentrarchus labrax at low and high temperature extremes. *Ecological Indicators*, 112:106118.

[147] Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskočilová, M. M., Nachman, M. W., Munclinger, P., Macholán, M., Piálek, J., and Tucker, P. K. (2012). Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between mus musculus musculus and m. m. domesticus. *Molecular ecology*, 21(12):3032–3047.

[148] Jedrzejewska, B. and Jedrzejewski, W. (2013). *Predation in vertebrate communities: the Bialowieza Primeval Forest as a case study*, volume 135. Springer Science & Business Media.

[149] Jensen, T. S. (1982). Seed production and outbreaks of non-cyclic rodent populations in deciduous forests. *Oecologia*, 54(2):184–192.

[150] Jensen, T. S. (1985). Seed-seed predator interactions of european beech, fagus silvatica and forest rodents, clethrionomys glareolus and apodemus flavicollis. *Oikos*, pages 149–156.

[151] Jewell, P. and Fullagar, P. (1965). Fertility among races of the field mouse (apodemus sylvaticus) and their failure to form hybrids with the yellow-necked mouse (apodemus flavicollis). *Evolution*, pages 175–181.

[152] Johnson, D. (1977). Inbreeding in populations with overlapping generations. *Genetics*, 87(3):581–591.

[153] Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., and Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, 502(7469):93–95.

[154] Johnston, S. E., Huisman, J., Ellis, P. A., and Pemberton, J. M. (2017). A high-density linkage map reveals sexual dimorphism in recombination landscapes in red deer (cervus elaphus). *G3: Genes, Genomes, Genetics*, 7(8):2859–2870.

[155] Jonasson, K. A. and Willis, C. K. (2012). Hibernation energetics of free-ranging little brown bats. *Journal of Experimental Biology*, 215(12):2141–2149.

[156] Jones, A. G. and Ardren, W. R. (2003). Methods of parentage analysis in natural populations. *Molecular ecology*, 12(10):2511–2523.

[157] Jones, A. G., Small, C. M., Paczolt, K. A., and Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. *Molecular ecology resources*, 10(1):6–30.

[158] Jones, O. R. and Wang, J. (2010). Colony: a program for parentage and sibship inference from multilocus genotype data. *Molecular ecology resources*, 10(3):551–555.

[159] Kardos, M., Luikart, G., and Allendorf, F. W. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity*, 115(1):63–72.

[160] Kassambara, A. and Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.

[161] Kekkonen, J., Seppä, P., Hanski, I., Jensen, H., Väisänen, R., and Brommer, J. (2011). Low genetic differentiation in a sedentary bird: house sparrow population genetics in a contiguous landscape. *Heredity*, 106(1):183–190.

[162] Keller, L. F. and Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in ecology & evolution*, 17(5):230–241.

[163] Kellogg, E. A. and Shaffer, H. B. (1993). Model organisms in evolutionary studies. *Systematic Biology*, 42(4):409–414.

[164] Kennedy, P. and MacFarlane, W. (1971). Oxygen consumption and water turnover of the fat-tailed marsupials dasycercus cristicauda and sminthopsis crassicaudata. *Comparative Biochemistry and Physiology Part A: Physiology*, 40(3):723–732.

[165] Kentie, R., Clegg, S. M., Tuljapurkar, S., Gaillard, J.-M., and Coulson, T. (2020). Life-history strategy varies with the strength of competition in a food-limited ungulate population. *Ecology letters*, 23(5):811–820.

[166] Kimura, M. and Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61(3):763.

[167] Kingsolver, J. G. and Nagle, A. (2007). Evolutionary divergence in thermal sensitivity and diapause of field and laboratory populations of manduca sexta. *Physiological and Biochemical Zoology*, 80(5):473–479.

[168] Kobbe, S., Ganzhorn, J. U., and Dausmann, K. H. (2011). Extreme individual flexibility of heterothermy in free-ranging malagasy mouse lemurs (microcebus griseorufus). *Journal of Comparative Physiology B*, 181(1):165–173.

[169] Konarzewski, M., Książek, A., and Łapo, I. B. (2005). Artificial selection on metabolic rates and related traits in rodents. *Integrative and Comparative Biology*, 45(3):416–425.

[170] Körtner, G. and Geiser, F. (2000). The temporal organization of daily torpor and hibernation: circadian and circannual rhythms. *Chronobiology international*, 17(2):103–128.

[171] Koubi, H., Robin, J., Dewasmes, G., Le Maho, Y., Frutoso, J., and Minaire, Y. (1991). Fasting-induced rise in locomotor activity in rats coincides with increased protein utilization. *Physiology & behavior*, 50(2):337–343.

[172] Kozakiewicz, M., Gryczyńska-Siemiątkowska, A., Panagiotopoulou, H., Kozakiewicz, A., Rutkowski, R., Abramowicz, K., and Gortat, T. (2009). The spatial genetic structure of bank vole (myodes glareolus) and yellow-necked mouse (apodemus flavicollis) populations: the effect of distance and habitat barriers. *Animal Biology*, 59(2):169–187.

[173] Kronfeld-Schor, N. and Dayan, T. (2013). Thermal ecology, environments, communities, and global change: energy intake and expenditure in endotherms. *Annual Review of Ecology, Evolution, and Systematics*, 44:461–480.

[174] Kujawa, A., Orczewska, A., Falkowski, M., Blicharska, M., Bohdan, A., Buchholz, L., Chylarecki, P., Gutowski, J. M., Latałowa, M., Mysłajek, R. W., et al. (2016). The białowieża forest–a unesco natural heritage site–protection priorities. *Forest Research Papers*, 77(4):302–323.

[175] Kvist, A. and Lindström, Å. (2001). Basal metabolic rate in migratory waders: intra-individual, intraspecific, interspecific and seasonal variation. *Functional Ecology*, pages 465–473.

[176] Labuda, M., Kozuch, O., Zuffová, E., Elecková, E., Hails, R. S., and Nuttall, P. A. (1997). Tick-borne encephalitis virus transmission between ticks cofeeding on specific immune natural rodent hosts. *Virology*, 235(1):138–143.

[177] Landes, J., Pavard, S., Henry, P.-Y., and Terrien, J. (2020). Flexibility is costly: Hidden physiological damage from seasonal phenotypic transitions in heterothermic species. *Frontiers in Physiology*, 11:985.

[178] Langmüller, A. M. and Schlötterer, C. (2020). Low concordance of short-term and long-term selection responses in experimental drosophila populations. *Molecular ecology*, 29(18):3466–3475.

[179] Lantová, P., Zub, K., Koskela, E., Šíchová, K., and Borowski, Z. (2011). Is there a linkage between metabolism and personality in small mammals? the root vole (microtus oeconomus) example. *Physiology & behavior*, 104(3):378–383.

[180] Latałowa, M., Zimny, M., Pędziszewska, A., Kupryjanowicz, M., et al. (2016). Postglacial history of białowieża forest-vegetation, climate and human activity. *Parki Narodowe i Rezerwaty Przyrody*, 35(1):3–49.

[181] Lawrence, D., Fiegna, F., Behrends, V., Bundy, J. G., Phillimore, A. B., Bell, T., and Barraclough, T. G. (2012). Species interactions alter evolutionary responses to a novel environment. *PLoS biology*, 10(5):e1001330.

[182] Leaché, A. D. and Oaks, J. R. (2017). The utility of single nucleotide polymorphism (snp) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 48:69–84.

[183] Lee, S. H., Yang, J., Chen, G.-B., Ripke, S., Stahl, E. A., Hultman, C. M., Sklar, P., Visscher, P. M., Sullivan, P. F., Goddard, M. E., et al. (2013). Estimation of snp heritability from dense genotype data. *American journal of human genetics*, 93(6):1151.

[184] Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M.-L., Koskiniemi, J., and Vainikka, A. (2019). Comparing radseq and microsatellites for estimating genetic diversity and relatedness—implications for brown trout conservation. *Ecology and Evolution*, 9(4):2106–2120.

[185] Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME journal*, 11(10):2181–2194.

[186] Leonard, W., Levy, S., Tarskaia, L., Klimova, T., Fedorova, V., Baltakhinova, M., Krivoshapkin, V., and Snodgrass, J. (2014). Seasonal variation in basal metabolic rates among the yakut (sakha) of northeastern siberia. *American Journal of Human Biology*, 26(4):437–445.

[187] Lepais, O. and Weir, J. T. (2014). Sim rad: an r package for simulation-based prediction of the number of loci expected in rad seq and similar genotyping by sequencing approaches. *Molecular ecology resources*, 14(6):1314–1321.

[188] Levesque, D. L. and Tattersall, G. J. (2010). Seasonal torpor and normothermic energy metabolism in the eastern chipmunk (tamias striatus). *Journal of Comparative Physiology B*, 180(2):279–292.

[189] Levine, B. A., Douglas, M. R., Yackel Adams, A. A., Lardner, B., Reed, R. N., Savidge, J. A., and Douglas, M. E. (2019). Genomic pedigree reconstruction identifies predictors of mating and reproductive success in an invasive vertebrate. *Ecology and evolution*, 9(20):11863–11877.

[190] Lewontin, R. C. (1991). Twenty-five years ago in genetics: electrophoresis in the development of evolutionary genetics: milestone or millstone? *Genetics*, 128(4):657.

[191] Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A., and Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular ecology*, 11(12):2453–2465.

[192] Lidicker Jr, W. Z. (2015). Genetic and spatial structuring of the california vole (microtus californicus) through a multiannual density peak and decline. *Journal of Mammalogy*, 96(6):1142–1151.

[193] Liebl, A. L., Schrey, A. W., Andrew, S. C., Sheldon, E. L., and Griffith, S. C. (2015). Invasion genetics: lessons from a ubiquitous bird, the house sparrow passer domesticus. *Current Zoology*, 61(3):465–476.

[194] Linck, E. and Battey, C. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, 19(3):639–647.

[195] Lindsley, D., Edington, C., and Von Halle, E. (1960). Sex-linked recessive lethals in drosophila whose expression is suppressed by the y chromosome. *Genetics*, 45(12):1649.

[196] Liu, T., Li, Q., Kong, L., and Yu, H. (2017). Comparison of microsatellites and snps for pedigree analysis in the pacific oyster crassostrea gigas. *Aquaculture International*, 25(4):1507–1519.

[197] Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., and Storfer, A. (2017). Breaking rad: An evaluation of the utility of restriction site-associated dna sequencing for genome scans of adaptation. *Molecular ecology resources*, 17(2):142–152.

[198] Lu, F., Glaubitz, J., Harriman, J., Casstevens, T., and Elshire, R. (2012). Tassel 3.0–universal network enabled analysis kit (uneak) pipeline documentation. *White Paper*, 2012:1–12.

[199] Luikart, G., England, P. R., Tallmon, D., Jordan, S., and Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature reviews genetics*, 4(12):981–994.

[200] Lynch, M., Ackerman, M. S., Gout, J.-F., Long, H., Sung, W., Thomas, W. K., and Foster, P. L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, 17(11):704.

[201] MacArthur, R. H. and Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, pages 373–387.

[202] Mackay, T. F., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., et al. (2012). The drosophila melanogaster genetic reference panel. *Nature*, 482(7384):173–178.

[203] Makova, K., Patton, J., Krysanov, E. Y., Chesser, R., and Baker, R. (1998). Microsatellite markers in wood mouse and striped field mouse (genus apodemus). *Molecular Ecology*, 7(2):247–248.

[204] Makova, K. D., Nekrutenko, A., and Baker, R. J. (2000). Evolution of microsatellite alleles in four species of mice (genus apodemus). *Journal of Molecular Evolution*, 51(2):166–172.

[205] Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22):2867–2873.

[206] Mank, J. E., Hultin-Rosenberg, L., Webster, M. T., and Ellegren, H. (2008). The unique genomic properties of sex-biased genes: insights from avian microarray data. *BMC genomics*, 9(1):148.

[207] Mank, J. E., Nam, K., Brunström, B., and Ellegren, H. (2010). Ontogenetic complexity of sexual dimorphism and sex-specific selection. *Molecular Biology and Evolution*, 27(7):1570–1578.

[208] Marsh, A. C., Poulton, S., and Harris, S. (2001). The yellow-necked mouse apodemus flavicollis in britain: status and analysis of factors affecting distribution. *Mammal Review*, 31(3-4):203–227.

[209] Marshall, T., Slate, J., Kruuk, L., and Pemberton, J. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular ecology*, 7(5):639–655.

[210] Marth, G. T., Korf, I., Yandell, M. D., Yeh, R. T., Gu, Z., Zakeri, H., Stitziel, N. O., Hillier, L., Kwok, P.-Y., and Gish, W. R. (1999). A general approach to single-nucleotide polymorphism discovery. *Nature genetics*, 23(4):452–456.

[211] Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1):10–12.

[212] Martin Cerezo, M. L. (2019). *European phylogeography and genetic structure of wood and yellow-necked mice Apodemus sylvaticus and Apodemus flavicollis based on whole-genome, high-density genotyping by restriction-site-associated DNA sequencing (RAD-seq)*. PhD thesis, University of Huddersfield.

[213] Martin Cerezo, M. L., Raval, R., Reyes, B. d. H., Kucka, M., Chan, Y. F., and Bryk, J. (2021). Identification and quantification of chimeric sequencing reads in a highly multiplexed rad-seq protocol. *Submitted to Molecular ecology resources*.

[214] Maruki, T. and Lynch, M. (2015). Genotype-frequency estimation from high-throughput sequencing data. *Genetics*, 201(2):473–486.

[215] Masel, J. (2011). Genetic drift. *Current Biology*, 21(20):R837–R838.

[216] Massatti, R., Reznicek, A. A., and Knowles, L. L. (2016). Utilizing radseq data for phylogenetic analysis of challenging taxonomic groups: A case study in carex sect. racemosae. *American Journal of Botany*, 103(2):337–347.

[217] Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., and Emerson, B. C. (2015). Restriction site-associated dna sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular ecology resources*, 15(1):28–41.

[218] Mathot, K. J., Martin, K., Kempenaers, B., and Forstmeier, W. (2013). Basal metabolic rate can evolve independently of morphological and behavioural traits. *Heredity*, 111(3):175–181.

[219] Matsunami, M., Endo, D., Saitou, N., Suzuki, H., and Onuma, M. (2018). Draft genome sequence of japanese wood mouse, apodemus speciosus. *Data in brief*, 16:43–46.

[220] McCartney-Melstad, E., Gidiş, M., and Shaffer, H. B. (2019). An empirical pipeline for choosing the optimal clustering threshold in radseq studies. *Molecular ecology resources*, 19(5):1195–1204.

[221] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome research*, 20(9):1297–1303.

[222] McKinney, G. J., Larson, W. A., Seeb, L. W., and Seeb, J. E. (2017). Rad seq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on breaking rad by lowry et al.(2016). *Molecular ecology resources*, 17(3):356–361.

[223] McNab, B. K. (2002). *The physiological ecology of vertebrates: a view from energetics*. Cornell University Press.

[224] McNab, B. K. (2007). The evolution of energetics in birds and mammals la evolución de la energética en aves y mamíferos. *The quintessential naturalist: honoring the life and legacy of Oliver P. Pearson*, 134:67.

[225] Mehrotra, S. and Goyal, V. (2014). Repetitive sequences in plant nuclear dna: types, distribution, evolution and function. *Genomics, proteomics & bioinformatics*, 12(4):164–171.

[226] Melo, A. T. and Hale, I. (2019). 'apparent': a simple and flexible r package for accurate snp-based parentage analysis in the absence of guiding information. *BMC bioinformatics*, 20(1):1–10.

[227] Melvin, R. G. and Andrews, M. T. (2009). Torpor induction in mammals: recent discoveries fueling new ideas. *Trends in Endocrinology & Metabolism*, 20(10):490–498.

[228] Michaux, J., Chevret, P., Filippucci, M.-G., and Macholan, M. (2002). Phylogeny of the genus apodemus with a special emphasis on the subgenus sylvaemus using the nuclear irbp gene and two mitochondrial markers: cytochrome b and 12s rrna. *Molecular phylogenetics and evolution*, 23(2):123–136.

[229] Michaux, J., Libois, R., and Filippucci, M. (2005). So close and so different: comparative phylogeography of two small mammal species, the yellow-necked fieldmouse (apodemus flavicollis) and the woodmouse (apodemus sylvaticus) in the western palearctic region. *Heredity*, 94(1):52–63.

[230] Michaux, J., Libois, R., Paradis, E., and Filippucci, M.-G. (2004). Phylogeographic history of the yellow-necked fieldmouse (apodemus flavicollis) in europe and in the near and middle east. *Molecular phylogenetics and evolution*, 32(3):788–798.

[231] Miller, C. R., Joyce, P., and Waits, L. P. (2002). Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, 160(1):357–366.

[232] Milligan, B. G. (2003). Maximum-likelihood estimation of relatedness. *Genetics*, 163(3):1153–1167.

[233] Mills, J. (2005). Regulation of rodent-borne viruses in the natural host: implications for human disease. *Infectious diseases from nature: mechanisms of viral emergence and persistence*, pages 45–57.

[234] Mitchell, F. J. and Cole, E. (1998). Reconstruction of long-term successional dynamics of temperate woodland in białowieża forest, poland. *Journal of Ecology*, 86(6):1042–1059.

[235] Morin, P. A., Luikart, G., Wayne, R. K., et al. (2004). Snps in ecology, evolution and conservation. *Trends in ecology & evolution*, 19(4):208–216.

[236] Muñoz, I., Henriques, D., Jara, L., Johnston, J. S., Chávez-Galarza, J., De La Rúa, P., and Pinto, M. A. (2017). Snp s selected by information content outperform randomly selected microsatellite loci for delineating genetic identification and introgression in the endangered dark european honeybee (apis mellifera mellifera). *Molecular ecology resources*, 17(4):783–795.

[237] Murray, E. K., Hien, A., de Vries, G. J., and Forger, N. G. (2009). Epigenetic control of sexual differentiation of the bed nucleus of the stria terminalis. *Endocrinology*, 150(9):4241–4247.

[238] Nabholz, B., Glémin, S., and Galtier, N. (2008). Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Molecular biology and evolution*, 25(1):120–130.

[239] Nazareno, A. G., Bemmels, J. B., Dick, C. W., and Lohmann, L. G. (2017). Minimum sample sizes for population genomics: an empirical study from an amazonian plant species. *Molecular Ecology Resources*, 17(6):1136–1147.

[240] Nei, M. and Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10):5269–5273.

[241] Nei, M., Maruyama, T., and Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, pages 1–10.

[242] Nespolo, R. F., Bacigalupe, L. D., and Bozinovic, F. (2003). Heritability of energetics in a wild mammal, the leaf-eared mouse (phyllotis darwini). *Evolution*, 57(7):1679–1688.

[243] Nespolo, R. F., Verdugo, C., Cortés, P. A., and Bacigalupe, L. D. (2010). Bioenergetics of torpor in the microbiotherid marsupial, monito del monte (dromiciops gliroides): the role of temperature and food availability. *Journal of Comparative Physiology B*, 180(5):767–773.

[244] Norin, T. and Malte, H. (2011). Repeatability of standard metabolic rate, active metabolic rate and aerobic scope in young brown trout during a period of moderate food availability. *Journal of Experimental Biology*, 214(10):1668–1675.

[245] Nowack, J. and Dausmann, K. H. (2015). Can heterothermy facilitate the colonization of new habitats? *Mammal Review*, 45(2):117–127.

[246] Nowack, J., Stawski, C., and Geiser, F. (2017). More functions of torpor and their roles in a changing world. *Journal of Comparative Physiology B*, 187(5):889–897.

[247] Nussey, D. H., Kruuk, L. E., Donald, A., Fowlie, M., and Clutton-Brock, T. H. (2006). The rate of senescence in maternal performance increases with early-life fecundity in red deer. *Ecology Letters*, 9(12):1342–1350.

[248] O'connell, M. and Wright, J. M. (1997). Microsatellite dna in fishes. *Reviews in Fish Biology and Fisheries*, 7(3):331–363.

[249] O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., and Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective snp filtering for molecular ecologists. *Molecular ecology*, 27(16):3193–3206.

[250] Oliver, M. K. and Piertney, S. B. (2012). Selection maintains mhc diversity through a natural population bottleneck. *Molecular Biology and Evolution*, 29(7):1713–1720.

[251] Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, 10(8):531–539.

[252] Ortmann, S., Heldmaier, G., Schmid, J., and Ganzhorn, J. (1997). Spontaneous daily torpor in malagasy mouse lemurs. *Naturwissenschaften*, 84(1):28–32.

[253] Otyama, P. I., Wilkey, A., Kulkarni, R., Assefa, T., Chu, Y., Clevenger, J., O'Connor, D. J., Wright, G. C., Dezern, S. W., MacDonald, G. E., et al. (2019). Evaluation of linkage disequilibrium, population structure, and genetic diversity in the us peanut mini core collection. *BMC genomics*, 20(1):1–17.

[254] Ouarour, A., Kirsch, R., and Pevet, P. (1991). Effects of temperature, steroids and castration on daily torpor in the djungarian hamster (phodopus sungorus). *Journal of Comparative Physiology A*, 168(4):477–481.

[255] Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2019). *Biostrings: Efficient manipulation of biological strings*. R package version 2.50.2.

[256] Paris, J. R., Stevens, J. R., and Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, 8(10):1360–1373.

[257] Pavitt, A. T., Walling, C. A., Pemberton, J. M., and Kruuk, L. E. (2014). Heritability and cross-sex genetic correlations of early-life circulating testosterone levels in a wild mammal. *Biology letters*, 10(11):20140685.

[258] Pegg, E., Doyle, K., Clark, E. L., Jatau, I. D., Tomley, F. M., and Blake, D. P. (2016). Application of a new pcr-rflp panel suggests a restricted population structure for eimeria tenella in uk and irish chickens. *Veterinary parasitology*, 229:60–67.

[259] Pemberton, J. (2008). Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635):613–621.

[260] Pemberton, J. M. and Kruuk, L. E. (2015). *Red deer research on the Isle of Rum NNR: management implications*. Scottish Natural Heritage.

[261] Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest radseq: an inexpensive method for de novo snp discovery and genotyping in model and non-model species. *PloS one*, 7(5):e37135.

[262] Peterson, R. A. and Cavanaugh, J. E. (2019). Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, pages 1–16.

[263] Pierson, J. C., Allendorf, F. W., Drapeau, P., and Schwartz, M. K. (2013). Breed locally, disperse globally: fine-scale genetic structure despite landscape-scale panmixia in a fire-specialist. *PLoS One*, 8(6):e67248.

[264] Plyusnin, A., Vaheri, A., and Lundkvist, Å. (2006). Saaremaa hantavirus should not be confused with its dangerous relative, dobrava virus. *Journal of clinical microbiology*, 44(4):1608–1611.

[265] Polymeropoulos, E. T., Oelkrug, R., and Jastroch, M. (2018). The evolution of endothermy–from patterns to mechanisms. *Frontiers in physiology*, 9:891.

[266] Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, 6(11):847–859.

[267] Pool, J. E., Hellmann, I., Jensen, J. D., and Nielsen, R. (2010). Population genetic inference from genomic sequence variation. *Genome research*, 20(3):291–300.

[268] Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.

[269] Pucek, Z., Jędrzejewski, W., Jędrzejewska, B., and Pucek, M. (1993). Rodent population dynamics in a primeval deciduous forest (białowieża national park) in relation to weather, seed crop, and predation. *Acta Theriologica*, 38(2):199–232.

[270] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3):559–575.

[271] Puritz, J. B., Hollenbeck, C. M., and Gold, J. R. (2014a). ddocent: a radseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2:e431.

[272] Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., and Bird, C. E. (2014b). Demystifying the rad fad. *Molecular ecology*, 23(24):5937–5942.

[273] Quigley, K. M., Bay, L. K., and van Oppen, M. J. (2020). Genome-wide snp analysis reveals an increase in adaptive genetic variation through selective breeding of coral. *Molecular Ecology*, 29(12):2176–2188.

[274] Quinn, J. S., Woolfenden, G. E., Fitzpatrick, J. W., and White, B. N. (1999). Multilocus dna fingerprinting supports genetic monogamy in florida scrub-jays. *Behavioral Ecology and Sociobiology*, 45(1):1–10.

[275] R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[276] Ranke, P. S., Skjelseth, S., Hagen, I. J., Billing, A. M., Pedersen, Å. A. B., Pärn, H., Ringsby, T. H., Sæther, B.-E., and Jensen, H. (2020). Multi-generational genetic consequences of reinforcement in a bird metapopulation. *Conservation Genetics*, 21(3):603–612.

[277] Regan, C., Pilkington, J., Berenos, C., Pemberton, J., Smiseth, P., and Wilson, A. (2017). Accounting for female space sharing in st. kilda soay sheep (ovis aries) results in little change in heritability estimates. *Journal of evolutionary biology*, 30(1):96–111.

[278] Regan, C. E., Pemberton, J. M., Pilkington, J. G., Smiseth, P. T., and Wilson, A. J. (2020). Linking genetic merit to sparse behavioral data: behavior and genetic effects on lamb growth in soay sheep. *Behavioral Ecology*, 31(1):114–122.

[279] Reid, J. M., Keller, L. F., Marr, A. B., Nietlisbach, P., Sardell, R. J., and Arcese, P. (2014). Pedigree error due to extra-pair reproduction substantially biases estimates of inbreeding depression. *Evolution*, 68(3):802–815.

[280] Richard, G.-F., Kerrest, A., and Dujon, B. (2008). Comparative genomics and molecular dynamics of dna repeats in eukaryotes. *Microbiology and molecular biology reviews*, 72(4):686–727.

[281] Riester, M., Stadler, P. F., and Klemm, K. (2009). Franz: reconstruction of wild multi-generation pedigrees. *Bioinformatics*, 25(16):2134–2139.

[282] Rikalainen, K., Aspi, J., Galarza, J. A., Koskela, E., and Mappes, T. (2012). Maintenance of genetic diversity in cyclic populations—a longitudinal analysis in myodes glareolus. *Ecology and Evolution*, 2(7):1491–1502.

[283] Rimbach, R., Pillay, N., and Schradin, C. (2017). Both thyroid hormone levels and resting metabolic rate decrease in african striped mice when food availability decreases. *Journal of Experimental Biology*, 220(5):837–843.

[284] Robinson, J. A., Räikkönen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., and Wayne, R. K. (2019). Genomic signatures of extensive inbreeding in isle royale wolves, a population on the threshold of extinction. *Science advances*, 5(5):eaau0757.

[285] Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve radseq-based population genomics. *Molecular ecology*, 28(21):4737–4754.

[286] Roesti, M., Hendry, A. P., Salzburger, W., and Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. *Molecular ecology*, 21(12):2852–2862.

[287] Roques, S., Chancerel, E., Boury, C., Pierre, M., and Acolas, M.-L. (2019). From microsatellites to single nucleotide polymorphisms for the genetic monitoring of a critically endangered sturgeon. *Ecology and evolution*, 9(12):7017–7029.

[288] Rossi, R. J. (2018). *Mathematical statistics: an introduction to likelihood based inference*. John Wiley & Sons.

[289] Ruby, N. F., Nelson, R. J., Licht, P., and Zucker, I. (1993). Prolactin and testosterone inhibit torpor in siberian hamsters. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 264(1):R123–R128.

[290] Ruf, T. and Geiser, F. (2015). Daily torpor and hibernation in birds and mammals. *Biological Reviews*, 90(3):891–926.

[291] Ruf, T., Klingenspor, M., Preis, H., and Heldmaier, G. (1991). Daily torpor in the djungarian hamster (phodopus sungorus): interactions with food intake, activity, and social behaviour. *Journal of Comparative Physiology B*, 160(6):609–615.

[292] Saatoglu, D., Niskanen, A. K., Kuismin, M., Ranke, P. S., Hagen, I. J., Araya-Ajoy, Y. G., Kvalnes, T., Pärn, H., Rønning, B., Ringsby, T. H., et al. (2021). Dispersal in a house sparrow metapopulation: an integrative case study of genetic assignment calibrated with ecological data and pedigree information. *Molecular Ecology*.

[293] Sadowska, E. T., Labocha, M. K., Baliga, K., Stanisz, A., Wróblewska, A. K., Jagusiak, W., and Koteja, P. (2005). Genetic correlations between basal and maximum metabolic rates in a wild rodent: consequences for evolution of endothermy. *Evolution*, 59(3):672–681.

[294] Sambrook, J. and Russell, D. W. (2006). Fragmentation of dna by sonication. *Cold spring harbor protocols*, 2006(4):pdb–prot4538.

[295] Santure, A. W. and Garant, D. (2018). Wild gwas—association mapping in natural populations. *Molecular ecology resources*, 18(4):729–738.

[296] Sato, J. J., Shimada, T., Kyogoku, D., Komura, T., Uemura, S., Saitoh, T., and Isagi, Y. (2018). Dietary niche partitioning between sympatric wood mouse species (muridae: Apodemus) revealed by dna meta-barcoding analysis. *Journal of Mammalogy*, 99(4):952–964.

[297] Schlötterer, C. (2002). A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, 160(2):753–763.

[298] Schubert, K. A., Boerema, A. S., Vaanholt, L. M., de Boer, S. F., Strijkstra, A. M., and Daan, S. (2010). Daily torpor in mice: high foraging costs trigger energy-saving hypothermia. *Biology letters*, 6(1):132–135.

[299] Schulte-Hostedde, A., Millar, J., and Hickling, G. (2001). Sexual dimorphism in body composition of small mammals. *Canadian Journal of Zoology*, 79(6):1016–1020.

[300] Schweyen, H., Rozenberg, A., and Leese, F. (2014). Detection and removal of pcr duplicates in population genomic ddrad studies by addition of a degenerate base region (dbr) in sequencing adapters. *The Biological Bulletin*, 227(2):146–160.

[301] Serizawa, K., Suzuki, H., and Tsuchiya, K. (2000). A phylogenetic view on species radiation in apodemus inferred from variation of nuclear and mitochondrial genes. *Biochemical Genetics*, 38(1-2):27–40.

[302] Shafer, A. B., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., and Wolf, J. B. (2017). Bioinformatic processing of rad-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8):907–917.

[303] Shcherbina, A., Ricke, D. O., Schwoebel, E., Boettcher, T., Zook, C., Bobrow, J., Petrovick, M., and Wack, E. (2016). Kinlinks: Software toolkit for kinship analysis and pedigree generation from hts datasets. In *2016 IEEE Symposium on Technologies for Homeland Security (HST)*, pages 1–6. IEEE.

[304] Sibold, C., Ulrich, R., Labuda, M., Lundkvist, Å., Martens, H., Schütt, M., Gerke, P., Leitmeyer, K., Meisel, H., and Krüger, D. (2001). Dobrava hantavirus causes hemorrhagic fever with renal syndrome in central europe and is carried by two different apodemus mice species. *Journal of medical virology*, 63(2):158–167.

[305] Simoens, C., Gielen, J., Van Montagu, M., and Inzé, D. (1988). Characterization of highly repetitive sequences of arabidopsis thaliana. *Nucleic acids research*, 16(14):6753–6766.

[306] Sims, D., Ilott, N. E., Sansom, S. N., Sudbery, I. M., Johnson, J. S., Fawcett, K. A., Berlanga-Taylor, A. J., Luna-Valero, S., Ponting, C. P., and Heger, A. (2014). Cgat: computational genomics analysis toolkit. *Bioinformatics*, 30(9):1290–1291.

[307] Singleton, G. (2003). Impacts of rodents on rice production in asia. Technical report.

[308] Singleton, G. and Krebs, C. (2007). The secret world of wild mice. the mouse in biomedical research. 1: 25–51.

[309] Slate, J., Kruuk, L., Marshall, T., Pemberton, J., and Clutton-Brock, T. (2000). Inbreeding depression influences lifetime breeding success in a wild population of red deer (cervus elaphus). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1453):1657–1662.

[310] Slatkin, M. (1996). In defense of founder-flush theories of speciation. *The American Naturalist*, 147(4):493–505.

[311] Snyder, M., Du, J., and Gerstein, M. (2010). Personal genome sequencing: current approaches and challenges. *Genes & development*, 24(5):423–431.

[312] Song, K., Li, L., and Zhang, G. (2016). Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology. *Scientific reports*, 6:35736.

[313] Sovic, M. G., Fries, A. C., and Gibbs, H. L. (2015). Aftr rad: a pipeline for accurate and efficient de novo assembly of rad seq data. *Molecular Ecology Resources*, 15(5):1163–1171.

[314] Speed, D., Hemani, G., Johnson, M. R., and Balding, D. J. (2012). Improved heritability estimation from genome-wide snps. *The American Journal of Human Genetics*, 91(6):1011–1021.

[315] Staples, J., Qiao, D., Cho, M. H., Silverman, E. K., Nickerson, D. A., Below, J. E., of Washington Center for Mendelian Genomics, U., et al. (2014). Primus: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *The American Journal of Human Genetics*, 95(5):553–564.

[316] Stenseth, N. C., Leirs, H., Skonhoft, A., Davis, S. A., Pech, R. P., Andreassen, H. P., Singleton, G. R., Lima, M., Machang'u, R. S., Makundi, R. H., et al. (2003). Mice, rats, and people: the bio-economics of agricultural rodent pests. *Frontiers in Ecology and the Environment*, 1(7):367–375.

[317] Steppan, S. J. and Schenk, J. J. (2017). Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PloS one*, 12(8):e0183070.

[318] Strucken, E., Lee, S., Lee, H., Song, K., Gibson, J., and Gondro, C. (2016). How many markers are enough? factors influencing parentage testing in different livestock populations. *Journal of Animal Breeding and Genetics*, 133(1):13–23.

[319] Stryjek, R. and Pisula, W. (2008). Warsaw wild captive pisula stryjek rats (wwcps)-establishing a breeding colony of norway rat in captivity. *Polish Psychological Bulletin*.

[320] Suchomel, J., Šipoš, J., and Košulič, O. (2020). Management intensity and forest successional stages as significant determinants of small mammal communities in a lowland floodplain forest. *Forests*, 11(12):1320.

[321] Sunde, J., Yıldırım, Y., Tibblin, P., and Forsman, A. (2020). Comparing the performance of microsatellites and radseq in population genetic studies: Analysis of data for pike (esox lucius) and a synthesis of previous studies. *Frontiers in genetics*, 11:218.

[322] Suzuki, H., Filippucci, M. G., Chelomina, G. N., Sato, J. J., Serizawa, K., and Nevo, E. (2008). A biogeographic view of apodemus in asia and europe inferred from nuclear and mitochondrial gene sequences. *Biochemical Genetics*, 46(5-6):329.

[323] Szwagrzyk, J. (2016). Białowieża forest: what it used to be, what it is now and what we want it to be in the future. *Forest Research Papers*, 77(4):291–295.

[324] Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P., and Bouvet, J. (1996). Reliable genotyping of samples with very low dna quantities using pcr. *Nucleic acids research*, 24(16):3189–3194.

[325] Tamiya, G., Shinya, M., Imanishi, T., Ikuta, T., Makino, S., Okamoto, K., Furugaki, K., Matsumoto, T., Mano, S., Ando, S., et al. (2005). Whole genome association study of rheumatoid arthritis using 27 039 microsatellites. *Human molecular genetics*, 14(16):2305–2321.

[326] Tautz, D. and Schlötterer, C. (1994). Simple sequences. *Current opinion in genetics & development*, 4(6):832–837.

[327] Taylor, H. R., Kardos, M. D., Ramstad, K. M., and Allendorf, F. W. (2015). Valid estimates of individual inbreeding coefficients from marker-based pedigrees are not feasible in wild populations with low allelic diversity. *Conservation Genetics*, 16(4):901–913.

[328] Team, F. F. T. (2015). The variant call format (vcf) version 4.2 specification. available at ht tps. *github. com/samtools/hts-specs*.

[329] Thompson, E. A. (1976). Inference of genealogical structure. *Social Science Information*, 15(2-3):477–526.

[330] Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S., and Lovette, I. J. (2018). Double-digest rad sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*, 18(5):953–965.

[331] Torkamaneh, D., Laroche, J., and Belzile, F. (2016). Genome-wide snp calling from genotyping by sequencing (gbs) data: a comparison of seven pipelines and two sequencing technologies. *PloS one*, 11(8):e0161333.

[332] Vendrami, D. L., Forcada, J., and Hoffman, J. I. (2019). Experimental validation of in silico predicted rad locus frequencies using genomic resources and short read data from a model marine mammal. *BMC genomics*, 20(1):1–12.

[333] Vieira, M. L. C., Santini, L., Diniz, A. L., and Munhoz, C. d. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and molecular biology*, 39(3):312–328.

[334] Visscher, P. M. and Goddard, M. E. (2015). A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. *Genetics*, 199(1):223–232.

[335] Visscher, P. M., Hemani, G., Vinkhuyzen, A. A., Chen, G.-B., Lee, S. H., Wray, N. R., Goddard, M. E., and Yang, J. (2014). Statistical power to detect genetic (co) variance of complex traits using snp data in unrelated samples. *PLoS Genet*, 10(4):e1004269.

[336] Vom Saal, F. S. and Bronson, F. (1980). Sexual characteristics of adult female mice are correlated with their blood testosterone levels during prenatal development. *Science*, 208(4444):597–599.

[337] Vuarin, P., Dammhahn, M., and Henry, P.-Y. (2013). Individual flexibility in energy saving: body size and condition constrain torpor use. *Functional Ecology*, 27(3):793–799.

[338] Vukićević-Radić, O., Matic, R., Kataranovski, D., and Stamenković, S. (2006). Spatial organization and home range of apodemus flavicollis and a. agrarius on mt. avala, serbia. *Acta Zoologica Academiae Scientiarum Hungaricae*, 52(1):81–96.

[339] Wandeler, P., Hoeck, P. E., and Keller, L. F. (2007). Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, 22(12):634–642.

[340] Wang, J. (2014). Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of Evolutionary Biology*, 27(3):518–530.

[341] Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., et al. (2015). Islands as model systems in ecology and evolution: prospects fifty years after macarthur-wilson. *Ecology Letters*, 18(2):200–217.

[342] Weir, B. S. and Goudet, J. (2017). A unified characterization of population structure and relatedness. *Genetics*, 206(4):2085–2103.

[343] Willoughby, J. R., Fernandez, N. B., Lamb, M. C., Ivy, J. A., Lacy, R. C., and DeWoody, J. A. (2015). The impacts of inbreeding, drift and selection on genetic diversity in captive breeding populations. *Molecular Ecology*, 24(1):98–110.

[344] Wingfield, J. C. (2003). Control of behavioural strategies for capricious environments. *Animal Behaviour*, 66(5):807–816.

[345] Wright, S. (1950). Genetical structure of populations. *Nature*, 166(4215):247–249.

[346] Wu, M.-X., Zhou, L.-M., Zhao, L.-D., Zhao, Z.-J., Zheng, W.-H., and Liu, J.-S. (2015). Seasonal variation in body mass, body temperature and thermogenesis in the hwamei, garrulax canorus. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 179:113–119.

[347] Xiang, A. P., Mao, F. F., Li, W.-Q., Park, D., Ma, B.-F., Wang, T., Vallender, T. W., Vallender, E. J., Zhang, L., Lee, J., et al. (2008). Extensive contribution of embryonic stem cells to the development of an evolutionarily divergent host. *Human molecular genetics*, 17(1):27–37.

[348] Xiong, S., Hao, Y., Rao, S., Huang, W., Hu, B., Wang, Y., et al. (2009). Effects of cutoff thresholds for minor allele frequencies on hapmap resolution: A real dataset-based evaluation of the chinese han and tibetan populations. *Chinese Science Bulletin*, 54(12):2069–2075.

[349] Xu, J., Turner, A., Little, J., Bleecker, E. R., and Meyers, D. A. (1996). Positive results in association studies are associated with departure from hardy-weinberg equilibrium: hint for genotyping error? *Human Genetics*, 111(6):573–574.

[350] Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., et al. (2010). Common snps explain a large proportion of the heritability for human height. *Nature genetics*, 42(7):565–569.

[351] Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1):76–82.

[352] Young, A. I. (2019). Solving the missing heritability problem. *PLoS genetics*, 15(6):e1008222.

[353] Yu, L.-X., Zheng, P., Bhamidimarri, S., Liu, X.-P., and Main, D. (2017). The impact of genotyping-by-sequencing pipelines on snp discovery and identification of markers associated with verticillium wilt resistance in autotetraploid alfalfa (medicago sativa l.). *Frontiers in Plant Science*, 8:89.

[354] Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of snp data. *Bioinformatics*, 28(24):3326–3328.

[355] Zhou, H., Sinsheimer, J. S., Bates, D. M., Chu, B. B., German, C. A., Ji, S. S., Keys, K. L., Kim, J., Ko, S., Mosher, G. D., et al. (2020). Openmendel: a cooperative programming project for statistical genetics. *Human genetics*, 139(1):61–71.

[356] Zima, J. and Macholán, M. (1995). B chromosomes in the wood mice (genus apodemus). *Acta Theriologica*, 40(Suppl. 3):75–86.

[357] Zimmerman, S. J., Aldridge, C. L., and Oyler-McCance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and snp data for a species of conservation concern. *BMC genomics*, 21:1–16.

[358] Zimny, M., Latałowa, M., and Pędziszewska, A. (2017). *Późnoholoceńska historia lasów Rezerwatu Ścisłego Białowieskiego Parku Narodowego*. Białowieża National Park, Białowieża.

[359] Zub, K., Piertney, S., Szafrańska, P. A., and Konarzewski, M. (2012). Environmental and genetic influences on body mass and resting metabolic rates (rmr) in a natural population of weasel mustela nivalis. *Molecular Ecology*, 21(5):1283–1293.

Appendix A

Due to the length and size of many of the supplementary materials for this thesis, they have been included in a GitHub repository at https://github.com/raval91/Aflav_thesis_supp.git. Below is a summary of the materials contained within it, where all filenames hyperlink directly to the specified file or directory in the digital version of this thesis. For a detailed description of the tables, column names and the information contained with them, please see the `README.md` file, also in the GitHub repository.

A.1 Chapter 2

A.1.1 Code

This directory contains scripts used in chapter 2 to optimise STACKS V2.3D parameters to genotype *A. flavicollis* samples.

- `denovo_map_parameter_optimisation.sh` optimised parameters to genotype samples using the `denovo_map.pl` pipeline of STACKS 2.3D.
- `stacks_data_extraction.sh` extracts the relevant metrics from Stacks 2.3d to optimise parameters
- `denovo_pipeline.sh` runs `denovo_map.pl` on the optimised parameter combination and prepares data for pedigree construction

- `M.musculus_insilico_digestion_parallel_SbfI_MseI.R` runs an *in silico* restriction digest of a reference genome and outputs various plots
- `stacks_metrics.R` plots the stacks metrics to assess which ones are optimal

A.1.2 Tables, figures and data

This directory contains tables and data used in, or generated by, the analyses in chapter 2.

- `A.flavicollis_demographic_data.txt` contains life history and sample ID data on samples from Białowieża National Park
- `adapter_primer_sequences.pdf` contains tables of fixed pairs of adapter sequences containing inner barcodes and Illumina primer sequences containing combinatorially used outer barcodes for quaddRAD.
- `coverage_duplicates_merged.txt` contains sequencing coverage of samples after duplicate samples have been merged
- `parameter_optimisation_metrics.txt` contains the metrics extracted from stacks 2.3d to optimise parameters
- `relatedness_results.txt` contains pairwise relatedness estimated by maximum likelihood, GCTA, KING and PLINK method of moments to identify duplicate samples
- `relatedness_results_duplicatesMerged.txt` contains pairwise relatedness estimated by maximum likelihood estimation after duplicate samples had been merged
- `genotyping_error_rates.txt` contains genotyping error rates as estimated using TIGER
- `sample_groups.txt` contains a list of independently sequenced samples across multiple sequencing lanes

Adapter name	Sequence
i7-top_#01_AGAGTTCG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNAGAGTTCG
i7-top_#02_ACCTGTTG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNACCTGTTG
i7-top_#03_AATCGCCT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNAATCGCCT
i7-top_#04_CTGGTTCA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNCTGGTTCA
i7-top_#05_CGACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNCGACAAGA
i7-top_#06_CAGTCGAA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNCAGTCGAA
i7-top_#07_GTCAGAAC	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNGTCAGAAC
i7-top_#08_GGCAATCT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNGGCAATCT
i7-top_#09_GTGGTCTT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNGTGGTCTT
i7-top_#10_TTGTTCCG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNTTGTTCCG
i7-top_#11_TCGCATTC	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNTCGCATTC
i7-top_#12_TCGAACCA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTVBBNTCGAACCA
i7-bottom_#01_AGAGTTCG	TACGAACTCTNVVBAGATCGGAAGAGCA
i7-bottom_#02_ACCTGTTG	TACAACAGGTNVVBAGATCGGAAGAGCA
i7-bottom_#03_AATCGCCT	TAAGGCGATTNVVBAGATCGGAAGAGCA
i7-bottom_#04_CTGGTTCA	TATGAACCAGNVVBAGATCGGAAGAGCA
i7-bottom_#05_CGACAAGA	TATCTTGTCGNVVBAGATCGGAAGAGCA
i7-bottom_#06_CAGTCGAA	TATTCGACTGNVVBAGATCGGAAGAGCA
i7-bottom_#07_GTCAGAAC	TAGTTCTGACNVVBAGATCGGAAGAGCA
i7-bottom_#08_GGCAATCT	TAAGATTGCCNVVBAGATCGGAAGAGCA
i7-bottom_#09_GTGGTCTT	TAAAGACCACNVVBAGATCGGAAGAGCA
i7-bottom_#10_TTGTTCCG	TACGGAACAANVVBAGATCGGAAGAGCA
i7-bottom_#11_TCGCATTC	TAGAATGCGANVVBAGATCGGAAGAGCA
i7-bottom_#12_TCGAACCA	TATGGTTCGANVVBAGATCGGAAGAGCA

Table A.1 quaddRAD_i7n inner adapter sequences containing 8bp barcodes.

Adapter name	Sequence
i5top_#01_AAGACTGG	CGCTCTTCCGATCTVBBNAAGACTGGTGCA/3Phos/
i5top_#02_ATGTTGGC	CGCTCTTCCGATCTVBBNATGTTGGCTGCA/3Phos/
i5top_#03_ATTGGCTG	CGCTCTTCCGATCTVBBNATTGGCTGTGCA/3Phos/
i5top_#04_CCTCATCT	CGCTCTTCCGATCTVBBNCCTCATCTTGCA/3Phos/
i5top_#05_CGGAATTG	CGCTCTTCCGATCTVBBNCGGAATTGTGCA/3Phos/
i5top_#06_CAAGGTGA	CGCTCTTCCGATCTVBBNCAAGGTGATGCA/3Phos/
i5top_#07_GACTTGAG	CGCTCTTCCGATCTVBBNGACTTGAGTGCA/3Phos/
i5top_#08_GAATCACG	CGCTCTTCCGATCTVBBNGAATCACGTGCA/3Phos/
i5top_#09_GGATTGTC	CGCTCTTCCGATCTVBBNGGATTGTCTGCA/3Phos/
i5top_#10_TCCTTCAC	CGCTCTTCCGATCTVBBNTCCTTCACTGCA/3Phos/
i5top_#11_TGTCAGTG	CGCTCTTCCGATCTVBBNTGTCAGTGTGCA/3Phos/
i5top_#12_TTCTGAGG	CGCTCTTCCGATCTVBBNTTCTGAGGTGCA/3Phos/
i5bottom_#01AAGACTGG	/5Phos/CCAGTCTTNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#02ATGTTGGC	/5Phos/GCCAACATNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#03ATTGGCTG	/5Phos/CAGCCAATNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#04CCTCATCT	/5Phos/AGATGAGGNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#05CGGAATTG	/5Phos/CAATTCCGNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#06CAAGGTGA	/5Phos/TCACCTTGNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#07GACTTGAG	/5Phos/CTCAAGTCNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#08GAATCACG	/5Phos/CGTGATTCNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#09GGATTGTC	/5Phos/GACAATCCNVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#10TCCTTCAC	/5Phos/GTGAAGGANVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#11TGTCAGTG	/5Phos/CACTGACANVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
i5bottom_#12TTCTGAGG	/5Phos/CCTCAGAANVVBAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Table A.2 quaddRAD_17n inner adapter sequences containing 8bp barcodes.

Primer name	Sequence
i501_AGCATGGA	AATGATACGGCGACCACCGAGATCTACAC{AGCATGGA}ACACTCTTTCCCTACACGAC*G
i502_CCTGGAAT	AATGATACGGCGACCACCGAGATCTACAC{CCTGGAAT}ACACTCTTTCCCTACACGAC*G
i503_GCAAGCAA	AATGATACGGCGACCACCGAGATCTACAC{GCAAGCAA}ACACTCTTTCCCTACACGAC*G
i504_TGAGGATG	AATGATACGGCGACCACCGAGATCTACAC{TGAGGATG}ACACTCTTTCCCTACACGAC*G
i701_ACACTCAG	CAAGCAGAAGACGGCATACGAGAT{CTGAGTGT}GTGACTGGAGTTCAGACGTGTGC*T
i702_CAGTCGAA	CAAGCAGAAGACGGCATACGAGAT{TTCGACTG}GTGACTGGAGTTCAGACGTGTGC*T
i703_GGCTCAAT	CAAGCAGAAGACGGCATACGAGAT{ATTGAGCC}GTGACTGGAGTTCAGACGTGTGC*T
i704_TTCCGCTT	CAAGCAGAAGACGGCATACGAGAT{AAGCGGAA}GTGACTGGAGTTCAGACGTGTGC*T

Table A.3 Illumina indexing primers sequences containing combinatorial barcode sequences.

A.2 Chapter 3

A.2.1 Code

This directory contains scripts used in chapter 3 to generate a pedigree using genotype data from *A. flavicollis*, estimate fitness and analyse allele frequency dynamics in the population.

- `allele_freq_sims/cluster` contains scripts to simulate null allele frequency distributions on a high performance cluster:

    1. `MAF_clusterRun_control.sh` instructs the cluster to run the simulations as a batch job
    2. `nullMAFsimulations.sh` parallelises the simulations
    3. `MAF_clusterRun.R` runs the simulations in R
    4. `extract_confint.R` extracts the confidence intervals from the output of the simulations

- `allele_freq_sims/nullMAFsimulations.R` runs the above simulations in parallel on a desktop computer instead of on a cluster
- `allele_frequency_change.R` analyses and plots allele frequency change over time
- `pedigree_construction_R3.6.R` constructs a pedigree using sequoia in R v3.6 and conducts analyses
- `genetic_contributions.R` calculates the genetic and genealogical contributions of founders to a population

A.2.2 Tables, figures and data

This directory contains tables and data used in, or generated by, the analyses in chapter 3.

- `maybe_relatives.txt` contains a list of possible relationships that were excluded from the pedigree
- `pedigree.txt` contains the pedigree generated by SEQUOIA in R v3.6
- `genealogical_genetic_contributions.txt` contains the genealogical and genetic contributions of founders
- `allele_frequencies.txt` observed allele frequencies between 2015-2017
- `allele_frequency_change.txt` contains data on the change in allele frequencies between 2015-2017
- `pvals.txt` contains the p-values of observed vs expected allele frequency change between 2015-2017
- `genetic_diversity.het` contains estimates of genetic diversity (inbreeding coefficients) as calculated by PLINK

A.3 Chapter 4

Index	Optimal clusters	Index value
Hubert	0	0
Dindex	0	0
Frey	1	NA
Duda	2	0.7163
PseudoT2	2	100.1932
Beale	2	0.6704
McClain	2	0.6542
Scott	3	414.4471
TrCovW	3	17577.98
TraceW	3	111.0165
Ratkowsky	3	0.4332
Ball	3	204.8935
KL	4	89.9839
CH	4	274.7578
Hartigan	4	78.64
Marriot	4	8263545
Friedman	4	2.2164
Rubin	4	-0.339
DB	4	1.0062
Silhouette	4	0.3658
PtBiserial	4	0.5732
SDindex	4	2.0272
Dunn	6	0.0432
Cindex	7	0.2406
CCC	10	3.5171
SDbw	10	0.2722

Table A.4 Table of indices generated by NBCLUST to assess the optimal number of clusters for K-means cluster analysis of H_i, BMR_{min} and m_b.

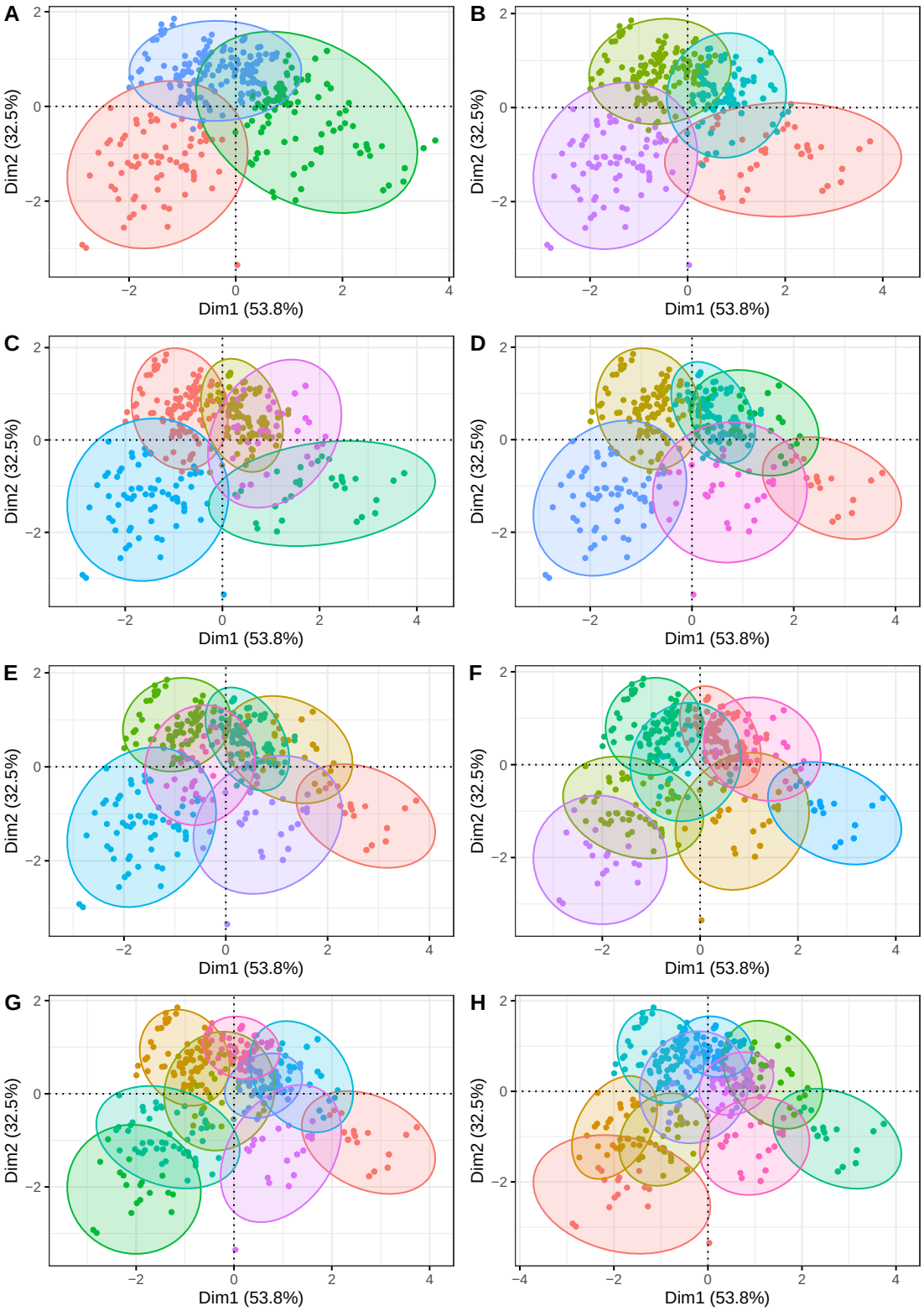


Fig. A.1 K-means cluster analysis performed on H_i, BMR_{min} and m_b for K = 3-10 (A-H respectively).

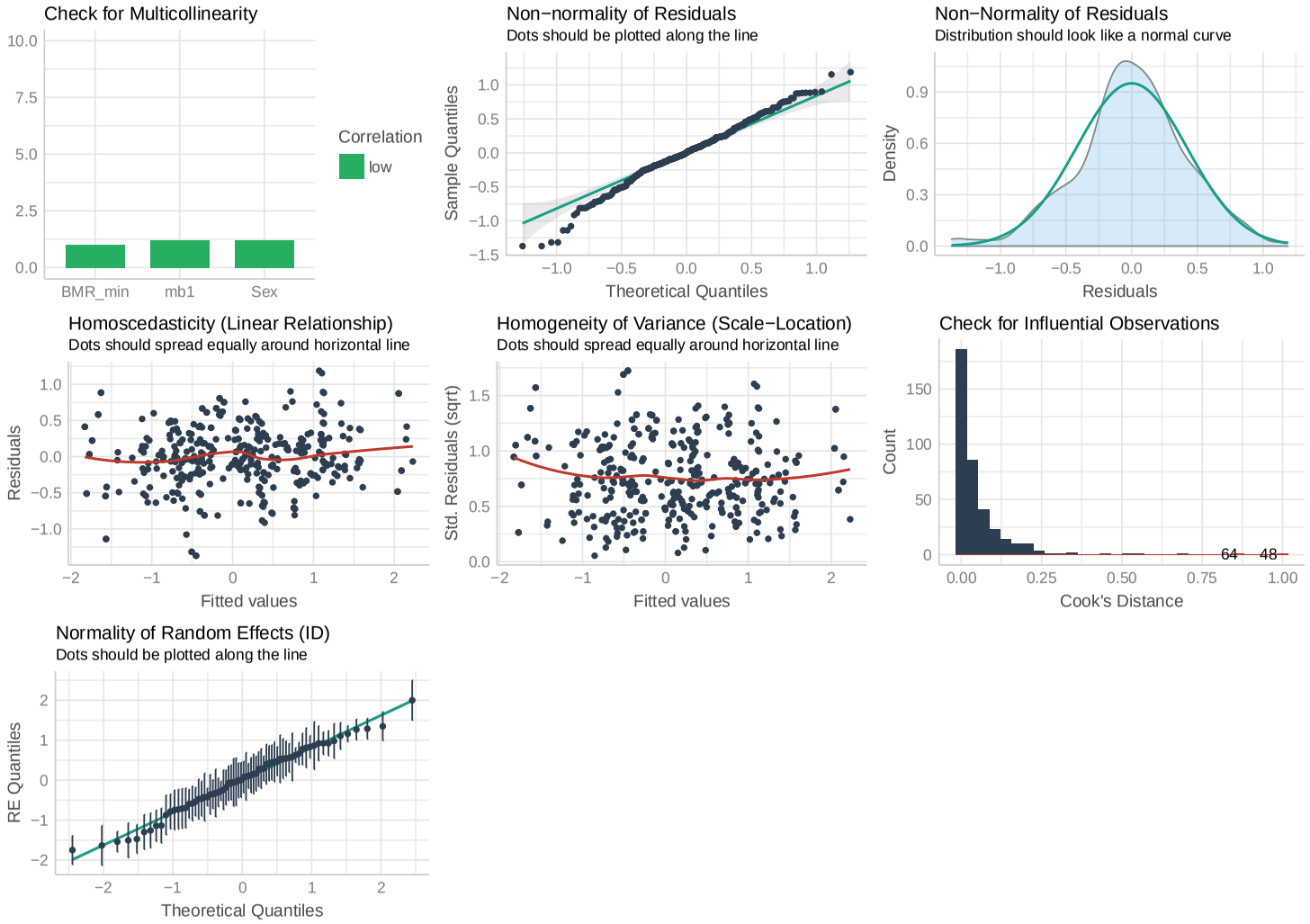


Fig. A.2 Diagnostics plots for the model of repeatability of heterothermic responses. Model includes Sex, $_b$ and BMR_{min} as fixed effects, and sample ID as a random effect.

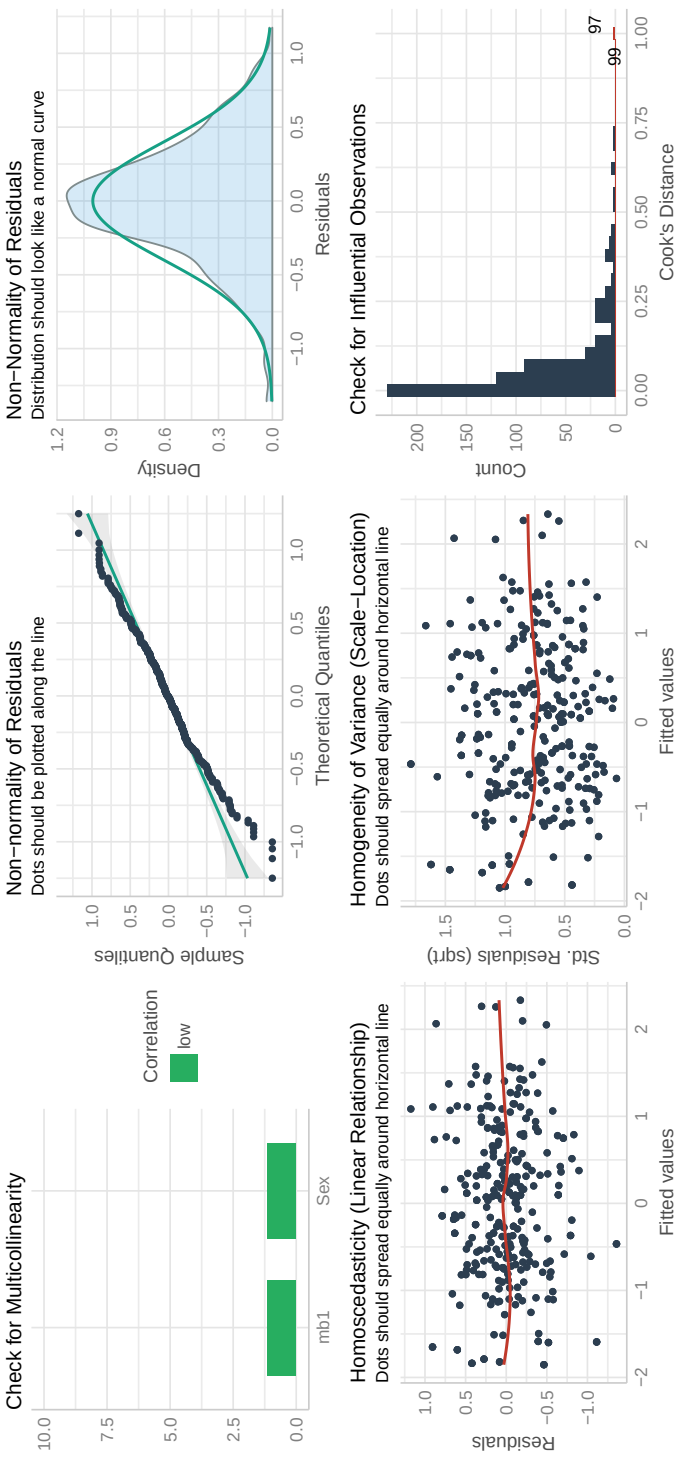


Fig. A.3 Diagnostics plots for the model of repeatability of heterothermic responses. Model includes Sex and m_b as fixed effects, and sample ID as a random effect.

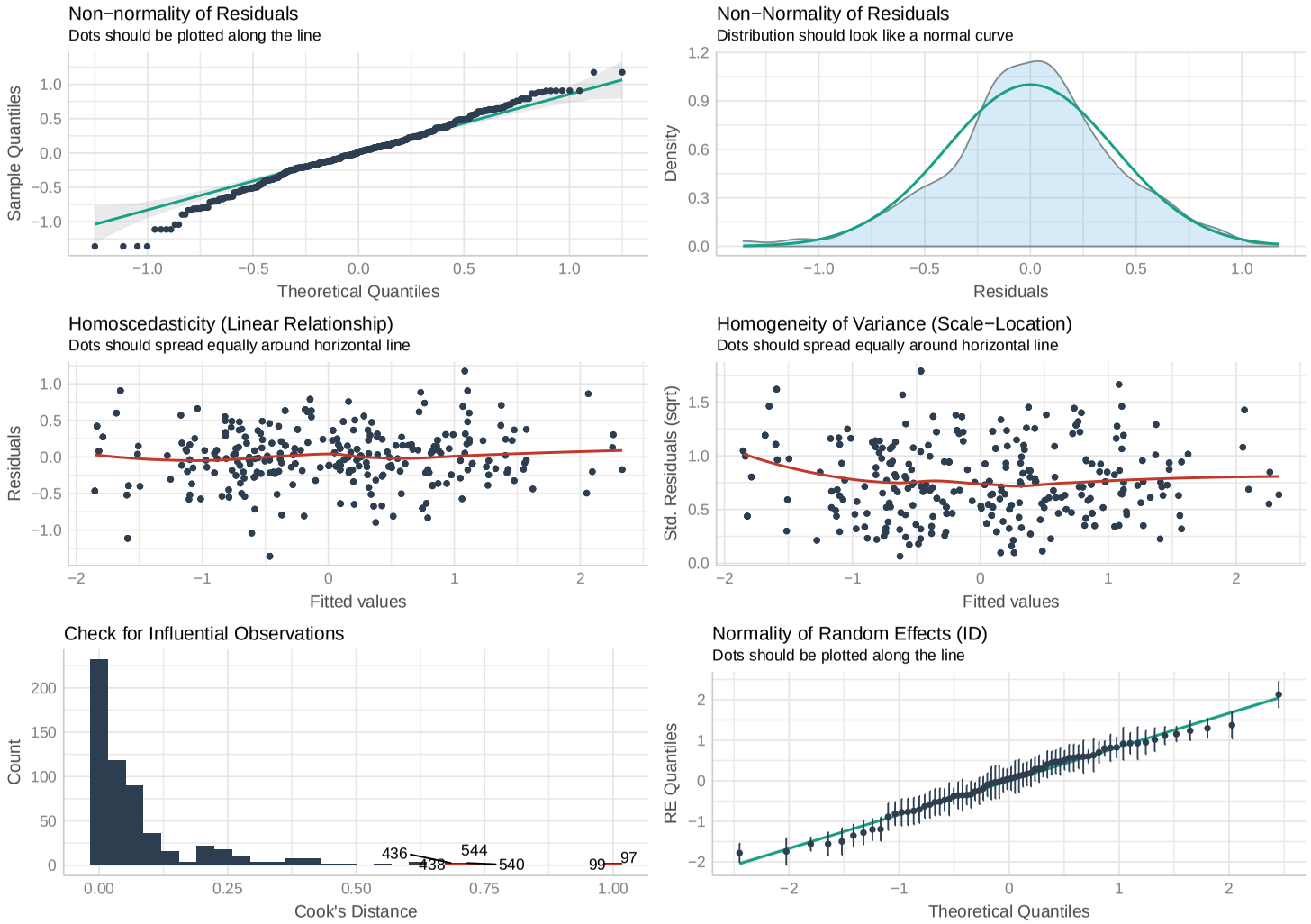


Fig. A.4 Diagnostics plots for the model of repeatability of heterothermic responses. Model includes m_b as a fixed effect, and sample ID as a random effect.

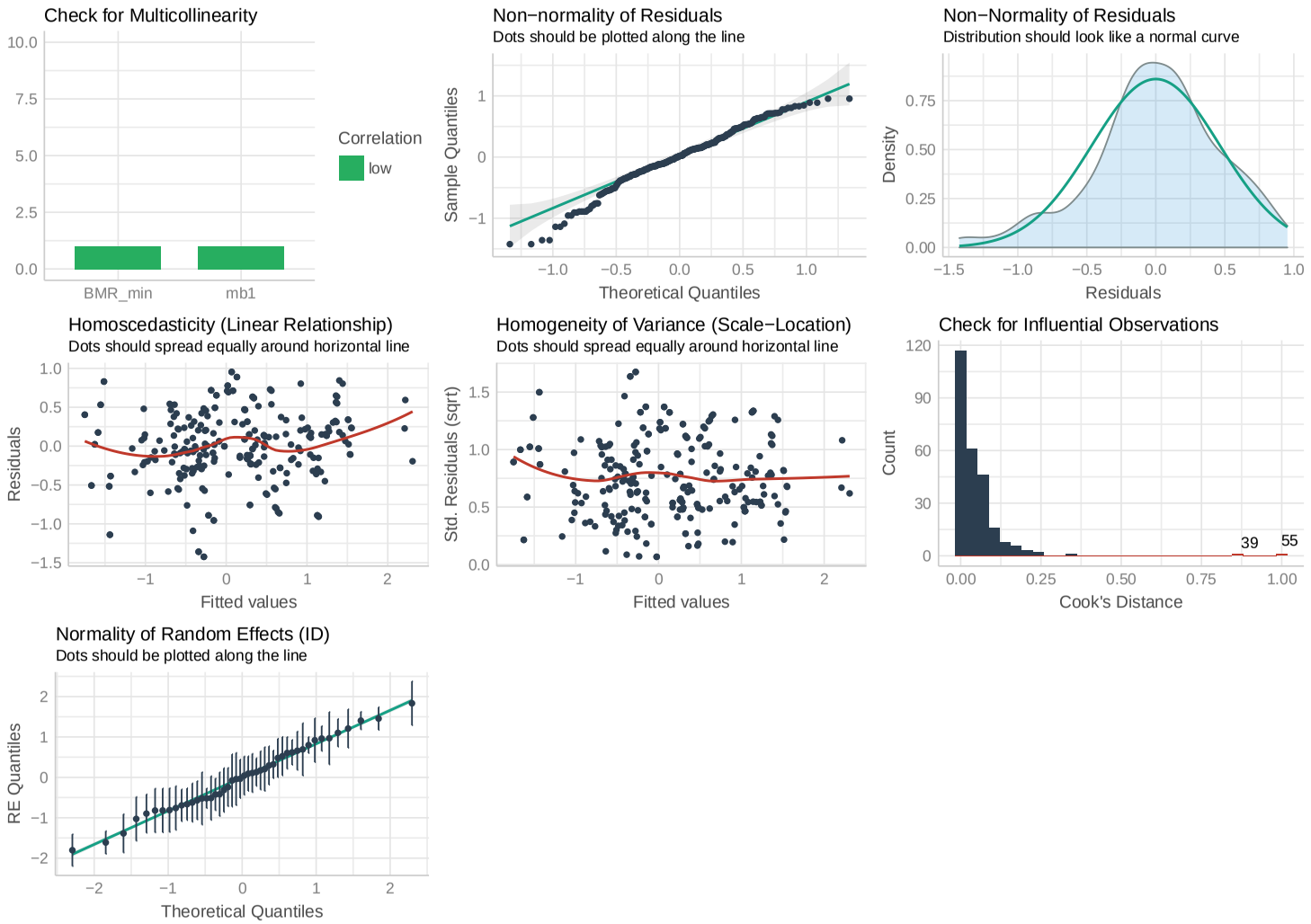


Fig. A.5 Diagnostics plots for the model of repeatability of heterothermic responses in males. Model includes m_b and BMR_{min} as fixed effects, and sample ID as a random effect.

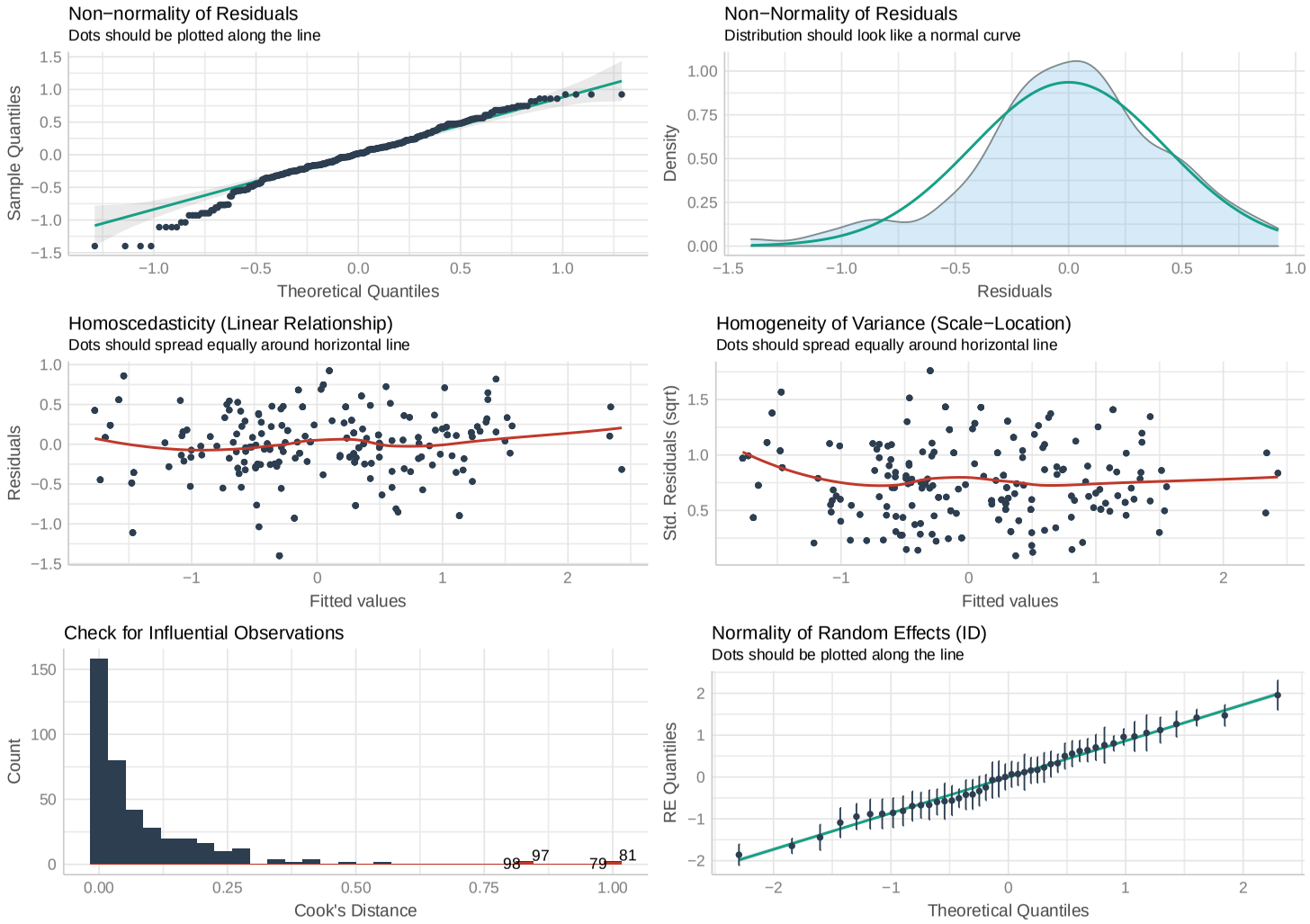


Fig. A.6 Diagnostics plots for the model of repeatability of heterothermic responses in males. Model includes m_b and BMR_{min} as fixed effects, and sample ID as a random effect.

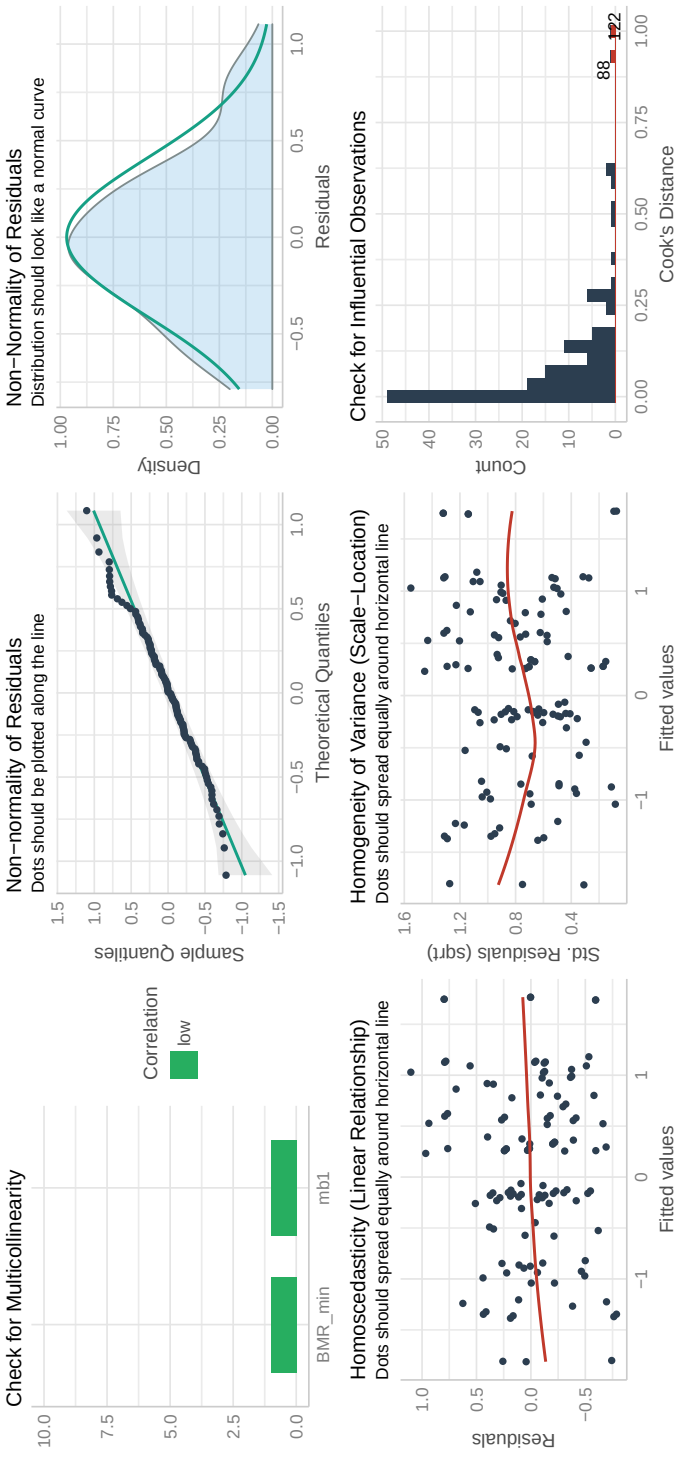


Fig. A.7 Diagnostics plots for the model of repeatability of heterothermic responses in females. Model includes m_b and BMR_{min} as fixed effects, and sample ID as a random effect.

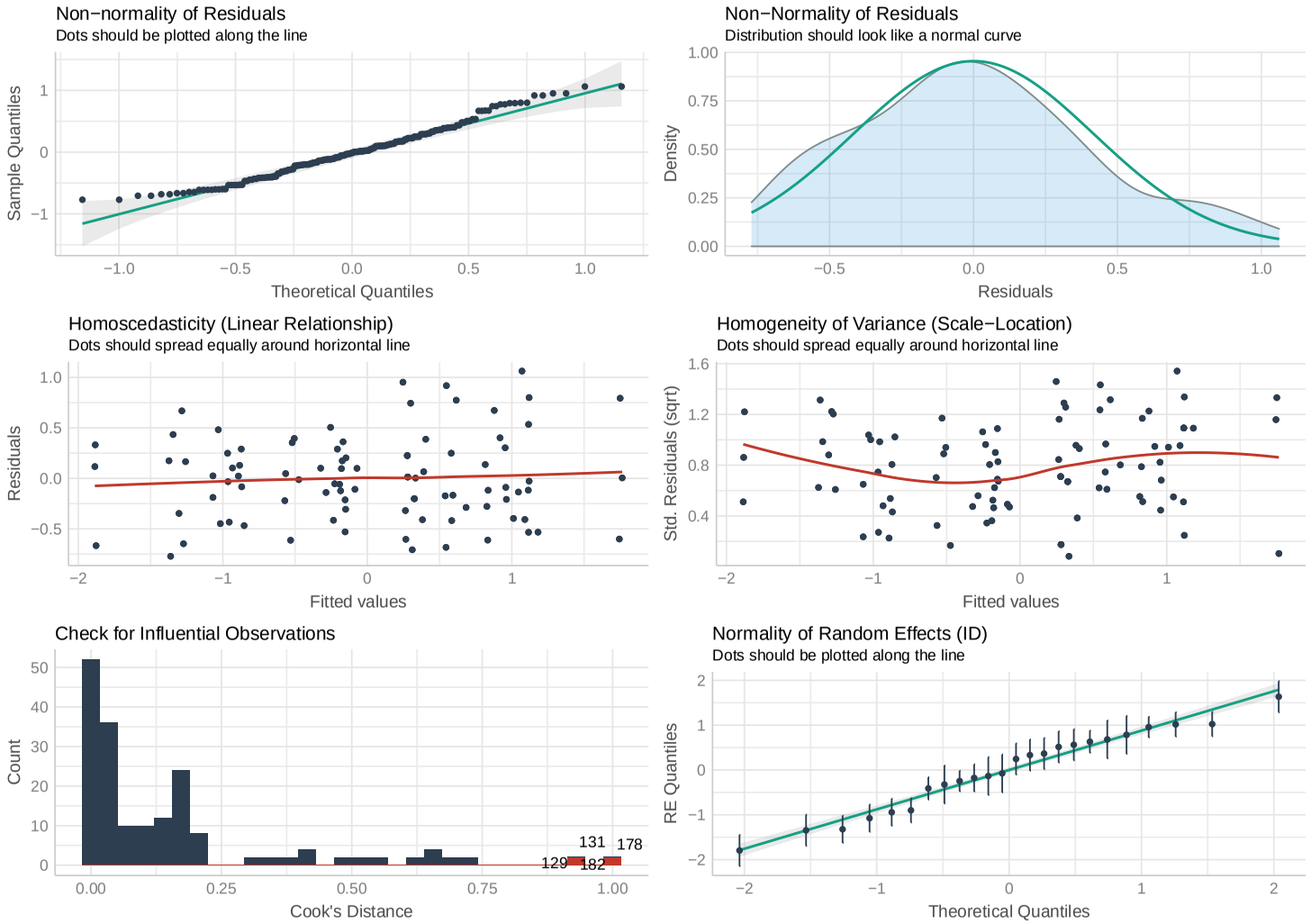


Fig. A.8 Diagnostics plots for the model of repeatability of heterothermic responses in females. Model includes m_b and BMR_{min} as fixed effects, and sample ID as a random effect.

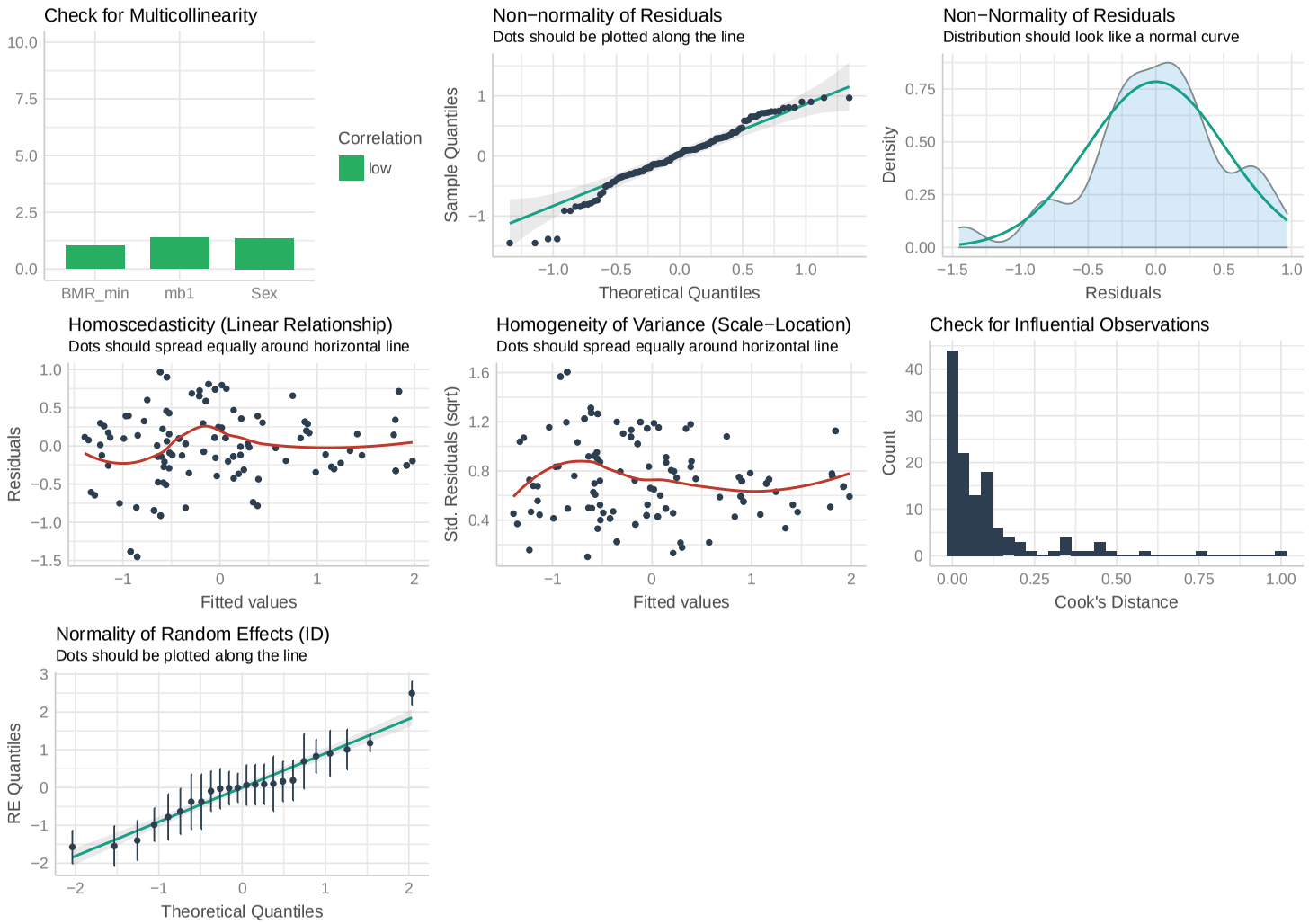


Fig. A.9 Diagnostics plots for the model of heritability of heterothermic responses. Model includes Sex, $_b$ and BMR_{min} as fixed effects, and sample ID as a random effect.

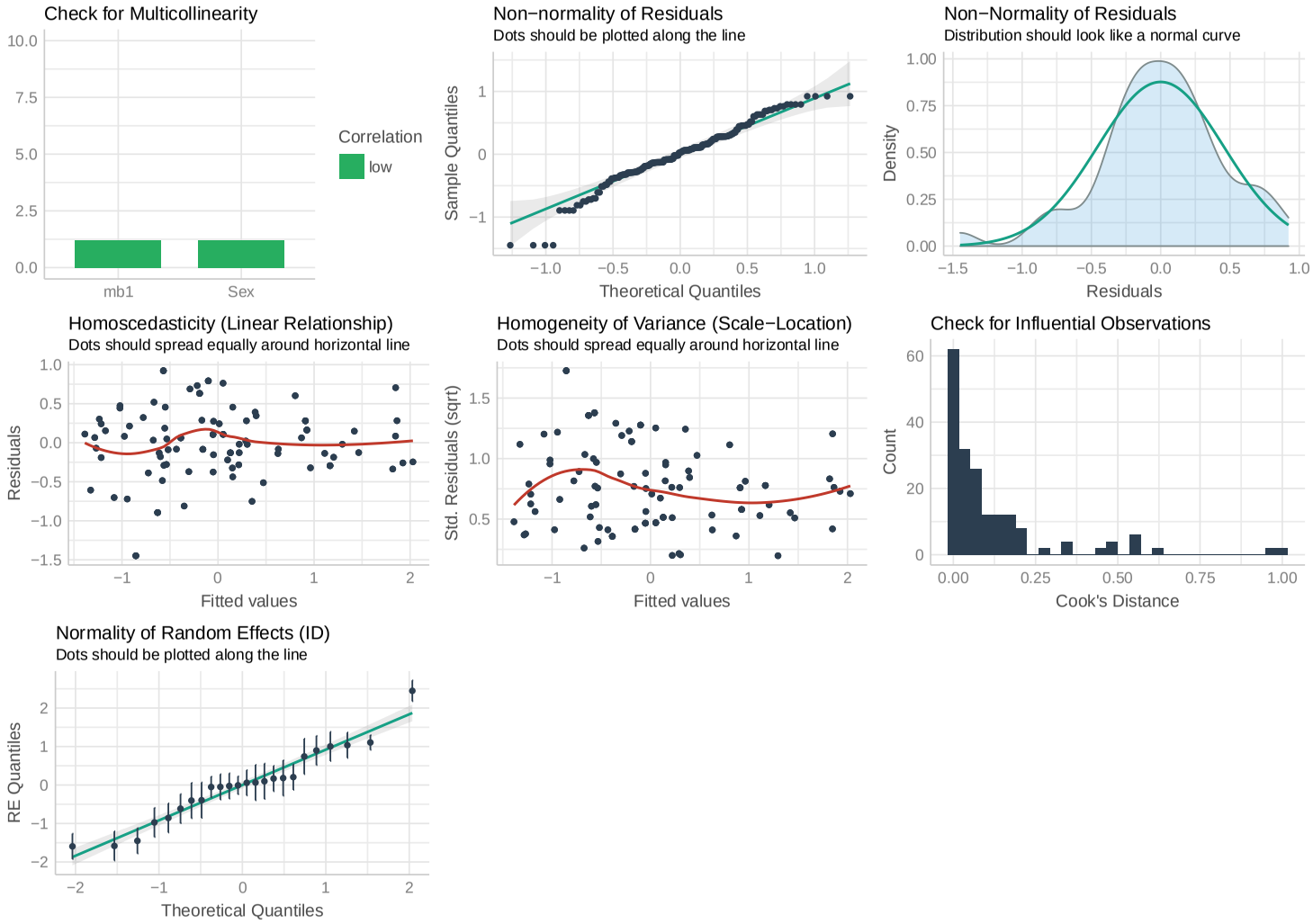


Fig. A.10 Diagnostics plots for the model of heritability of heterothermic responses. Model includes Sex and m_b as fixed effects, and sample ID as a random effect.

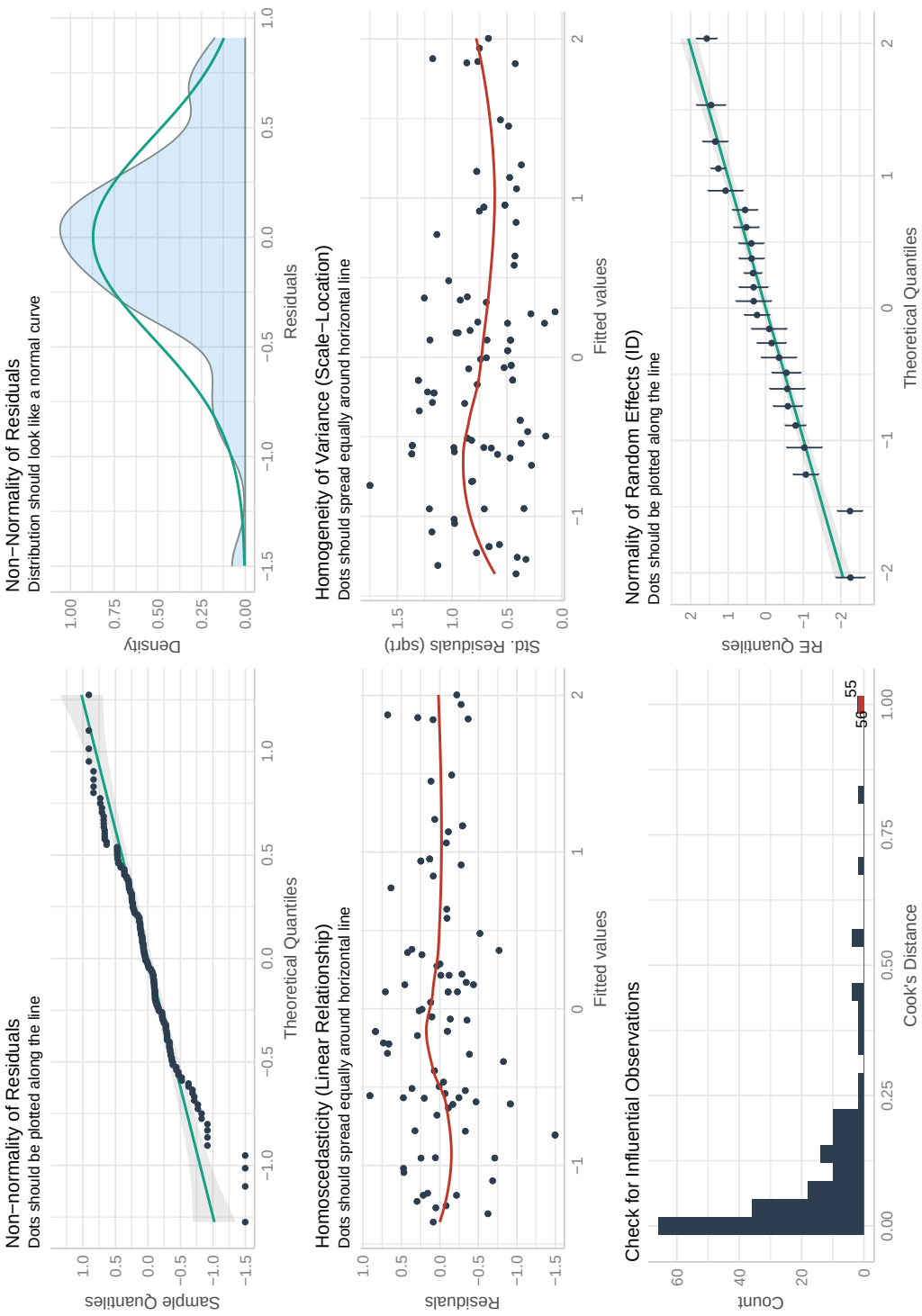


Fig. A.11 Diagnostics plots for the model of heritability of heterothermic responses. Model includes m_b as a fixed effect, and sample ID as a random effect.

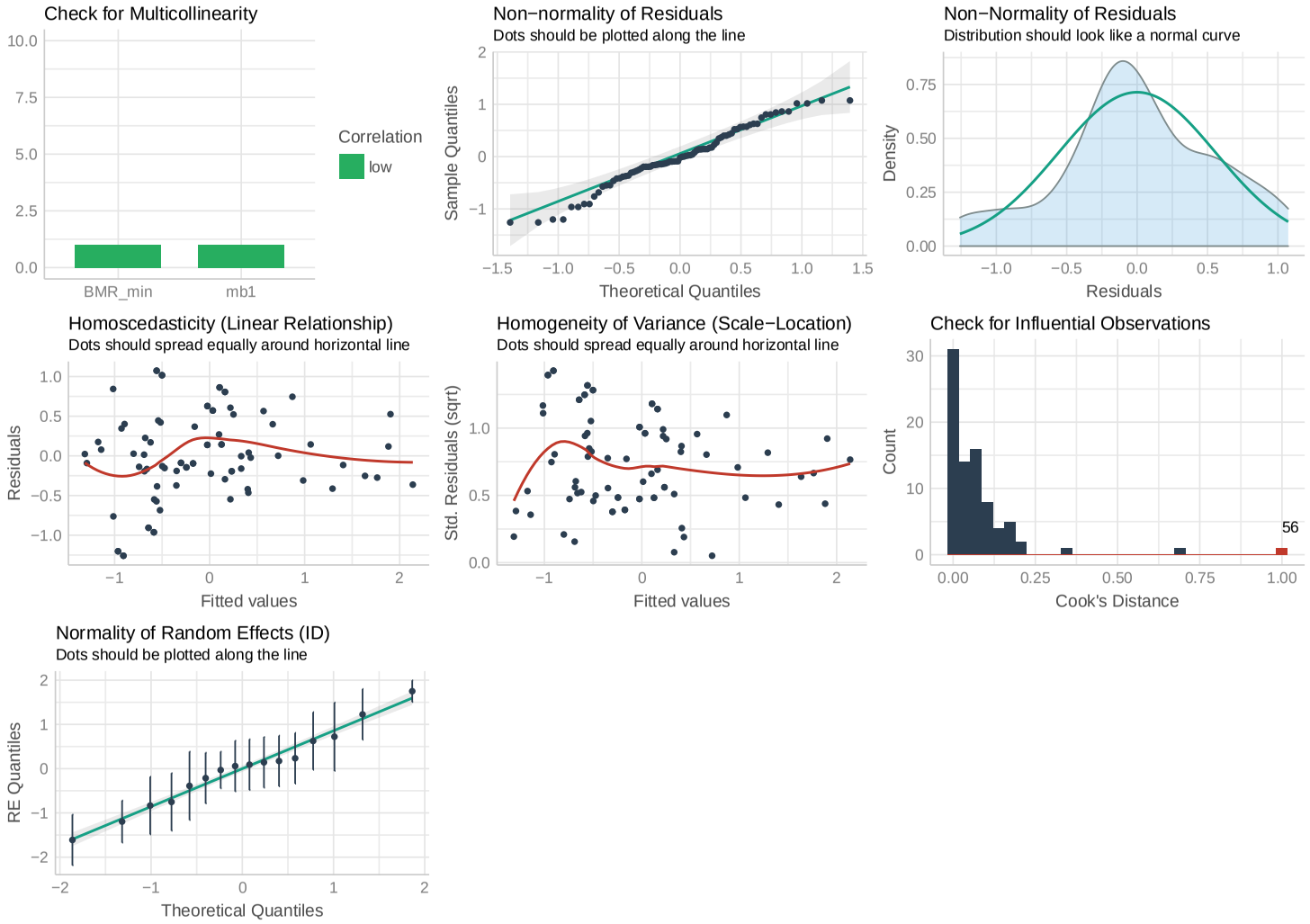


Fig. A.12 Diagnostics plots for the model of heritability of heterothermic responses in males. Model includes m_b and BMR_{min} as fixed effects, and sample ID as a random effect.

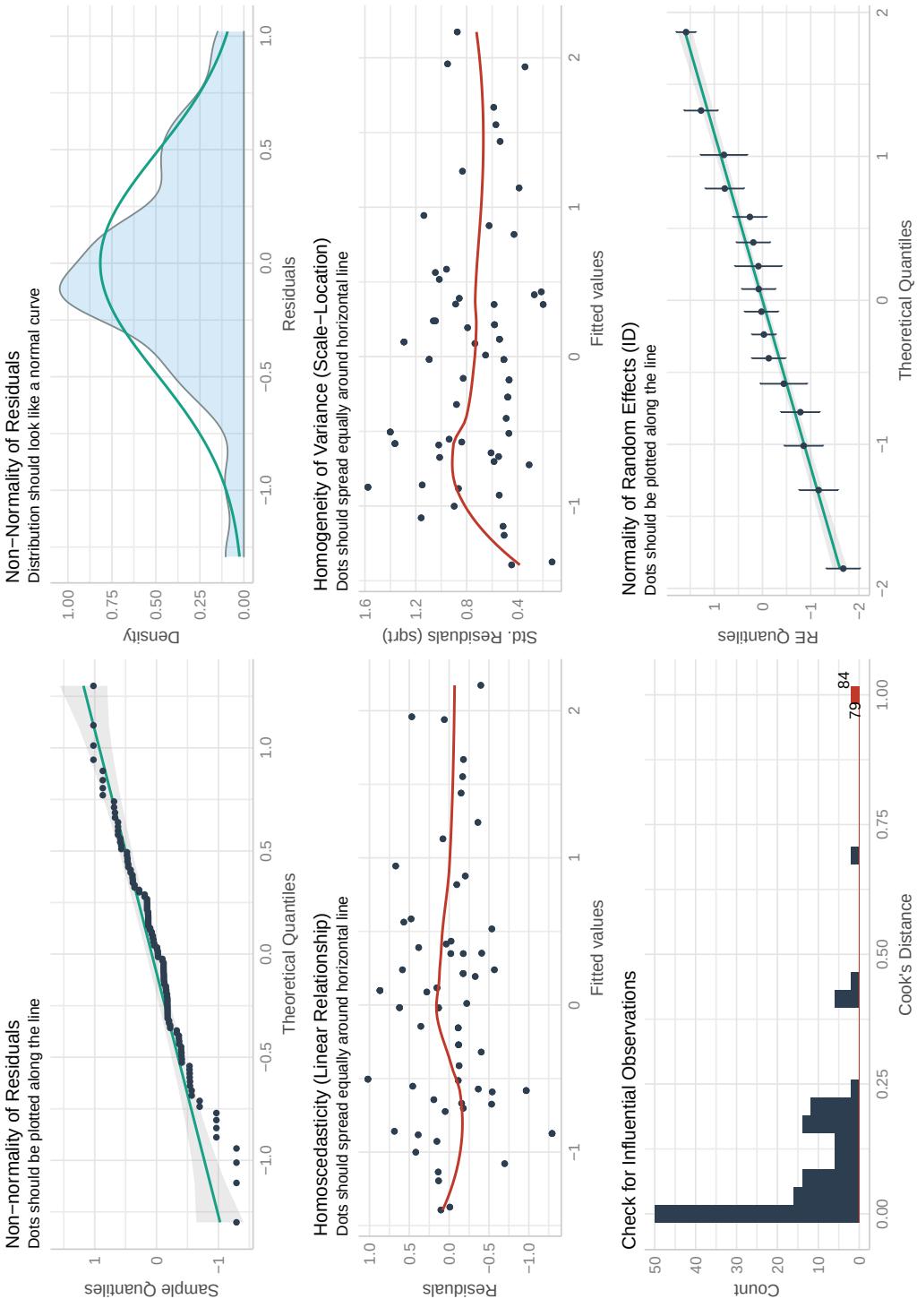


Fig. A.13 Diagnostics plots for the model of heritability of heterothermic responses in males. Model includes m_b and BMR_{min} as fixed effects, and sample ID as a random effect.

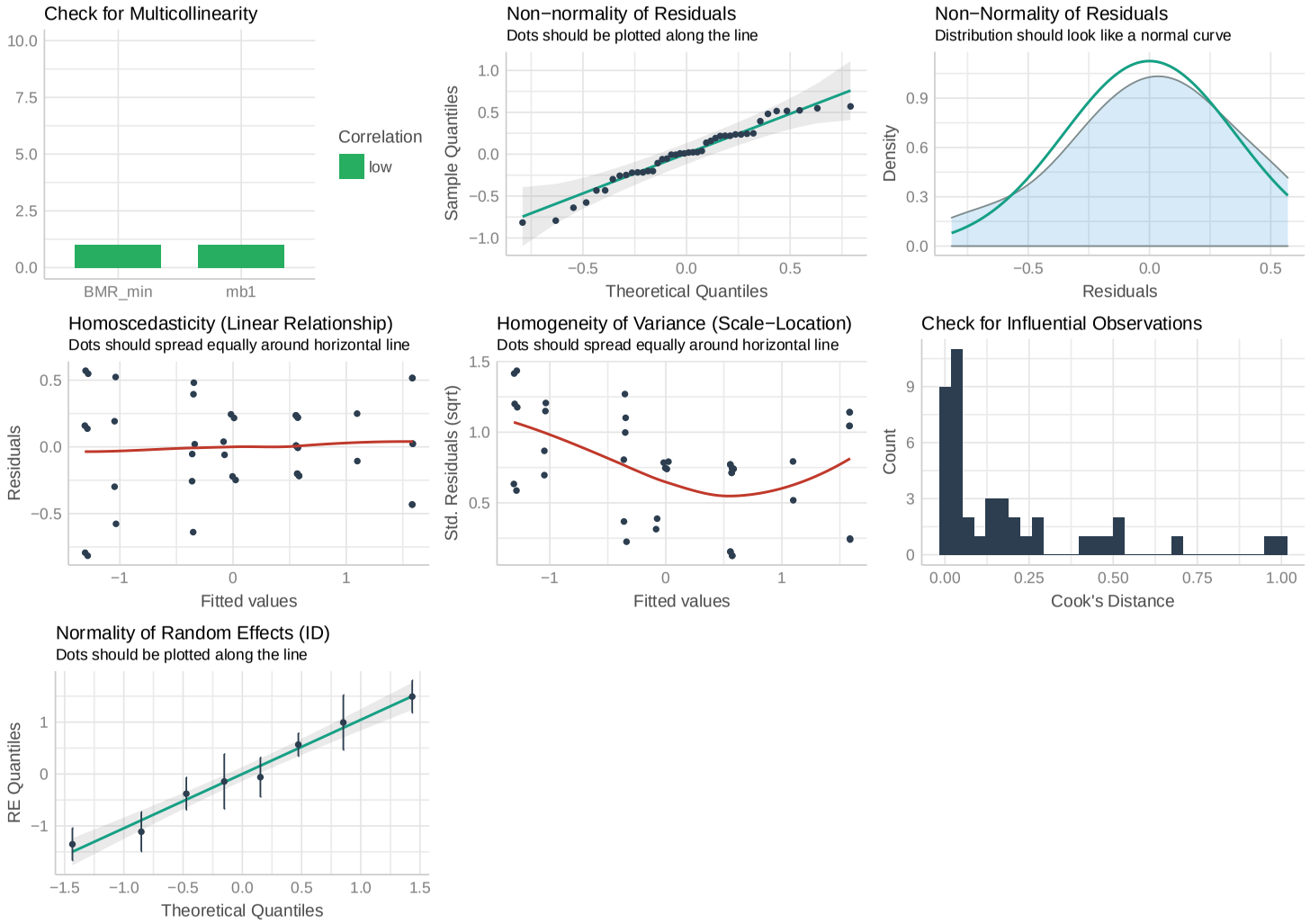


Fig. A.14 Diagnostics plots for the model of heritability of heterothermic responses in females. Model includes m_b and BMR_{min} as fixed effects, and sample ID as a random effect.

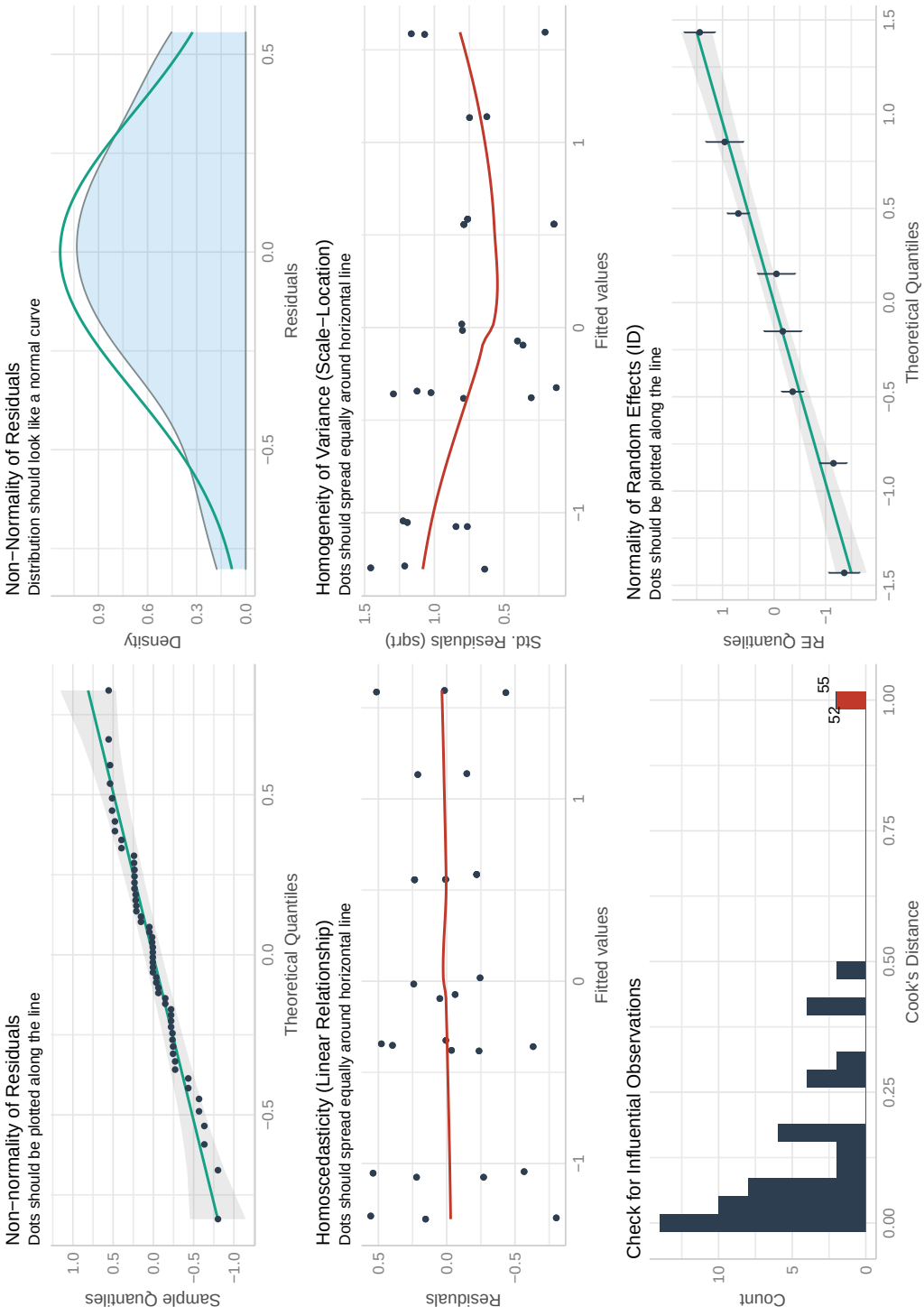


Fig. A.15 Diagnostics plots for the model of heritability of heterothermic responses in females. Model includes m_b and BMR_{min} as fixed effects, and sample ID as a random effect.

A.3.1 Code

This directory contains scripts used to analyse heterothermic responses of *A. flavicollis*.

- `cluster_analysis.R` performs a K-means cluster analysis on the heterothermy data
- `torpor_analysis.R` runs a full analysis if heterothermic responses including estimating repeatability, calculating relatedness and estimating heritability
- `BMR_analysis_REML_final.R` runs a full analysis of BMR_{min} data from *A. flavicollis*

A.3.2 Tables, figures and data

This directory contains tables, data and figures used in, or generated by, the analyses in chapter 4.

- `cluster_analysis/cluster_indices.txt` contains a table of indices to determine optimal number of clusters in K-means cluster analysis
- `cluster_analysis/k3.pdf - cluster_analysis/k10.pdf` are figures of cluster analyses for heterothermy data when K=3-10
- `diagnostic_plots/repeatability/` contains figures of quantile-quantile plots, and residuals vs fitted values plots of repeatability models including sex+mb+bmr, sex+mb and mb as variables in the models. Also included are diagnostic plots for for models of each individual sex
- `diagnostic_plots/heritability/` contains figures of quantile-quantile plots, and residuals vs fitted values plots of heritability models including sex+mb+bmr, sex+mb and mb as variables in the models. Also included are diagnostic plots for for models of each individual sex