# On the balance of drift and selection: the evolution of the Orkney vole

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

**Xuejing Wang**

von China

Leiter der Arbeit:
Prof. Dr. Gerald Heckel
Institut für Ökologie und Evolution, Universität Bern

# On the balance of drift and selection: the evolution of the Orkney vole

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

**Xuejing Wang**

von China

Leiter der Arbeit:
Prof. Dr. Gerald Heckel
Institut für Ökologie und Evolution, Universität Bern

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, 6.6.2023

Der Dekan
Prof. Dr. Marco Herwegh

# Public Defense

Institute of Ecology and Evolution,

University of Bern,

6, June 2023

**Composition of the jury:**

| | |
|---|---|
| Chair: | Prof. Dr. Catherine Peichel |
| Supervisor: | Prof. Dr. Gerald Heckel |
| Examiner: | Prof. Dr. Lukas Keller, University of Zurich |

# Table of Contents

# Summary

Demography and selection have been under the spotlight for a long time in evolutionary biology. As human activities lead to increased risk of population fragmentation and biological invasion, both of which involve drastic demographic changes, the importance of advanced knowledge about the impacts of genetic drift in the short and long term is increasing. Populations on islands are ideal models to study adaptive and non-adaptive evolutionary processes simultaneously. In isolated island populations, the efficacy of purifying selection is reduced by genetic drift, leading to accumulation of deleterious variants in homozygous state, hence reduced fitness (mutation load). On the other hand, island populations often show phenotypic differences when compared to the continental populations (island syndrome), which are considered to be related to divergent selection posed by the contrast in environmental factors.

In this thesis, I used the Orkney vole (*Microtus arvalis orcadensis*) as my model to investigate the genomic consequences of bottlenecks and long-term isolation. The Orkney vole was introduced by Neolithic farmers from the European continent, and isolated since introduction for over 5,000 years, providing a unique opportunity to study the long-term effects of isolation in nature. In **Chapter 1**, I reconstructed the detailed demographic history of Orkney populations and found that Orkney voles have been through a strong bottleneck related to the introduction. I further investigated the mutation load in Orkney populations and found high fixation of potential deleterious alleles. In **Chapter 2**, I looked at the genomic landscape of Orkney voles and found genome-wide relaxation of purifying selection. I performed genomic tests for divergent selective sweeps, and detected signatures indicating the reduction of positive selection in Orkney voles related to their increase of body size.

The research of mutation load and selection in most species has been mainly constrained to autosomes so far. Inherited along with the autosomes, the sex chromosomes undergo disparate evolutionary paths not only because of their functions but also differences in ploidy. **In Chapter 3**, I first assembled the sex chromosomes of the common vole. With population genomic data, I found that the autosomes, X and Y chromosomes had different levels of genetic diversity, accumulation of deleterious alleles, and genetic responses to severe bottlenecks. Such differences are likely correlated to the ploidy of the chromosomes and sex-biased mutation rates.

# General Introduction

Demography and selection are of great weight on the balance of evolution. They jointly shape the genomes of the species and neither can be ignored (Glinka, et al. 2003). Isolated populations are of major concern in conservation biology, ecology and evolutionary biology (Mayr 1963; Kruess and Tscharntke 1994). Human activities have affected and will continue to affect ecosystems worldwide. On one hand, fragmented habitats create isolated populations and leads to increased risks of extinction in many species (Schipper, et al. 2008; Crooks, et al. 2017). On the other hand, human transportation of organisms, whether consciously or not, may result in biological invasions given suitable conditions (Luque, et al. 2014; Pringle, et al. 2019). Even though both scenarios typically involve a certain degree of population isolation, their demographic changes are opposed. To further understand the evolutionary history and future performance of isolated populations, which are influenced by genetic diversity and the quality of genetic variation (Colautti, et al. 2017), island populations are ideal systems in nature to study the short and long-term effects of isolation (Mayr 1963; Latter 1973; Whittaker and Fernández-Palacios 2007). There are several aspects of the evolution of island populations along with demographic changes I focused on in this thesis: mutation load, the shifts of selection on the islands, and population genetics of sex chromosomes. With my study of Orkney voles, I intend to provide new insights into long-term evolutionary effects of genetic drift on isolated population.

**Mutation load and demographic changes**

Out of new mutations under selection, most have a negative effect on individual fitness (Eyre-Walker and Keightley 2007). Deleterious mutations are thus common in the genome (Lohmueller 2014). They are usually effectively purged by purifying selection and maintained at low frequency and heterozygous state in large populations (Kirkpatrick and Jarne 2000; Bertorelle, et al. 2022). However, in nature, demographic changes such as bottlenecks and population isolation are common, which lead to genetic drift out-weighing selection in the genome in certain cases (Mayr 1963; Lynch, et al. 1995; Simons, et al. 2014; Gravel 2016). Indeed, in small populations, random genetic drift is the major evolutionary force driving the loss of genetic diversity, and may leave long-lasting signatures even if the

population size increases (Nei, et al. 1975; Simons, et al. 2014). In such conditions, deleterious variants might persist and reach high frequency. Consequently, they are more likely to be homozygous, resulting in a loss of fitness which is called “mutation load” (Kimura, et al. 1963; Lynch, et al. 1995; González-Martínez, et al. 2017; Perrier, et al. 2020; Bertorelle, et al. 2022). Mildly deleterious mutations are more likely to persist, and strongly deleterious mutations are more likely to be lost over time by being purged by purifying selection (Kirkpatrick and Jarne 2000; Crnokrak and Barrett 2002; van der Valk, et al. 2021).

Beside population size changes and fragmentation, expanding populations can also face an increase of homozygous state of deleterious alleles correlated with loss of fitness (expansion load). Theory has shown that the dynamics of deleterious alleles at the edge of spatially expanding populations is initially the same as that after a series of bottlenecks (Excoffier and Ray 2008). Due to drift, the proportion of standing deleterious alleles increases while their total number remains unchanged (Peischl and Excoffier 2015; Aris-Brosou 2019). The expansion load can last for thousands of generations, especially in the absence of gene flow from core populations (Peischl, et al. 2016).

It is important to understand the distribution and dynamics of deleterious mutations and their relationship to fitness not only for conservation genetics, but also for further understanding of evolutionary genomics (DeWoody, et al. 2021; Kardos, et al. 2021). However, empirical studies on the mutation load in isolated populations have mainly focused on a limited range of species, for example model species (Mallet and Chippindale 2011; Peischl, et al. 2018; Aris-Brosou 2019) or endangered wild populations (Huisman, et al. 2016; Robinson, et al. 2018; Grossen, et al. 2020; Stoffel, et al. 2021). It is uncommon to have populations in nature with a clear founding history and being isolated beyond a few hundred generations (Huisman, et al. 2016; Stoffel, et al. 2021) to study the long-term effects of mutation load driven by genetic drift.

**Disentangling selection from demography**

It remains challenging to detect signals of selection in the presence of demographic changes, as both demography and selection shape the genome jointly (Glinka, et al. 2003; Li, et al. 2012). Strong genetic drift can decrease the efficacy of purifying selection and leads to extensive relaxed selection overall in the genome (Gravel 2016; Kutschera, et al. 2019). Such background noise in the genome brought by genetic drift poses challenges to selection tests

as the signals from random drift can be similar to those of selection (Hahn 2008; Li, et al. 2012; Moinet, et al. 2022). Unlike genetic drift that affects the whole genome, selection usually works at specific genes and leaves local signatures in contrast with the genomic background (Nosil, et al. 2009; Ellegren, et al. 2012; Ahrens, et al. 2018). Typical signals include local shifts ("genomic islands") of decreased genetic diversity, extreme Tajima's D, or increased $F_{ST}$ (Poelstra, et al. 2014; Wolf and Ellegren 2016). The genes in or close to the genomic islands are commonly considered as under selection, and potentially correlate with the process of adaptation and divergence of the species (Ellegren, et al. 2012; Malinsky, et al. 2015; Marques, et al. 2018). However, under certain circumstances, the genomic patterns arising from demographic changes are similar to those of selection, and obscure the typical "genomic island" feature (Stephan 2016). For example, inbreeding brings long regions with low diversity similar to the action of selective sweeps (Harris, et al. 2018), and segregating functional alleles in the population can be found both in the condition of population structure and balancing selection (Johri, et al. 2022).

In the presence of demographic changes, methods based on extended haplotype homozygosity (EHH, Sabeti, et al. 2002), such as iHES (Voight, et al. 2006), XP-EHH (Sabeti, et al. 2007) and REHH (Qanbari, et al. 2010), are suitable for detecting selective sweeps. The EHH methods are designed for detecting long homozygous regions around beneficial mutations and use the genomic background for correction, which increases the robustness to population size changes (Ferrer-Admetlla, et al. 2014; Weigand and Leese 2018). Multiple genetic measurements are usually used and combined for selection tests to obtain robust results, for example with Fisher's combined probability test and hidden Markov model (HMM) clustering (Ayub, et al. 2013; Marques, et al. 2018). The density-based spatial clustering of applications with noise (DBSCAN) algorithm (Ester, et al. 1996), which is easy in use to combine multiple genetic measurements, also has good potential for selection tests, yet has not been fully explored. In short, the DBSCAN algorithm clusters data points with multiple dimensions (genetic measurements in my study) based on their distance between each other (Schubert, et al. 2017). DBSCAN can thus detect outlier regions under the simple assumption that genomic regions affected by strong selection should be rare and statistically distinct from most of the genome (Johri, et al. 2022).

**Mutation load on the sex chromosomes**

Assemblies of sex chromosomes are still scarce, limited by technical difficulties caused by low depth of X or W chromosome, heterogeny and similarity between the two sex chromosomes (Tomaszkiewicz, et al. 2017). The formation of sex chromosomes starts from accumulation of sexually antagonistic mutations close to the sex-determining gene, leading to loss of recombination in the region (Vicoso 2019; Rafati, et al. 2020). This region accumulates deleterious mutations on the non-recombining chromosome (e.g. Y in XY systems), tends to degenerate, and gradually differentiates into two sex chromosomes (Vicoso 2019). Highly heteromorphic sex chromosomes with short homologous regions are common in mammals, for example in mice (Soh, et al. 2014) and humans (Lahn and Page 1999).

Recent theoretical and empirical studies on the deleterious mutations in the genome (Peischl and Excoffier 2015; Robinson, et al. 2018; Aris-Brosou 2019; Grossen, et al. 2020; Bertorelle, et al. 2022) usually focus on diploid systems, or to be specific, the autosomes of diploid species. As selection functions on the individual level, it is important to further understand the fitness effects of deleterious mutations on sex chromosomes to have a whole picture of mutation load related to deleterious mutation both on autosomes and sex chromosomes.

There are three types of ploidies in diploid species with heteromorphic sex chromosomes (Laporte and Charlesworth 2002; Immler 2019). Here I use the XY system as an example and assume a sex ratio of 1 for simplification. First, autosomes and the pseudoautosomal regions of sex chromosomes are diploid going through recombination each generation. Second, non-recombining regions of the Y chromosome and mitochondrial DNA are haploid have an $N_e$ equal to 1/4 of the autosomes. Last, the regions on the X chromosome outside of pseudoautosomal regions are diploid in females and haploid in males. The X chromosome is diploid in the female and haploid in the male, going through recombination in 2/3 of its history, and with $N_e$ equal to 3/4 of the autosomes (Bachtrog 2006). In the haploid system, every mutation is dominant, and selection against deleterious alleles is stronger than in the diploid system (Valero, et al. 1992; Szövényi, et al. 2014). Thus, the sex chromosomes should experience stronger purging of deleterious alleles due to a certain degree of haploidy. Furthermore, the mutation rate is higher in males in many species (Hurst and Ellegren 1998), which adds an extra layer of complexity to the population genetics of sex chromosomes. The

genetic diversity and dynamics of deleterious alleles on sex chromosomes and autosomes may thus not be the same under demographic changes.

**Island populations on the balance of drift and selection**

Island populations are ideal for studying the effect of both genetic drift and selection in nature with often simpler environments and the clear geographic barrier of water (Foster 1964; Losos and Ricklefs 2009; Grant and Grant 2014). On one hand, island populations usually experience bottlenecks and genetic isolation in the absence of gene flow which can lead to strong genetic drift (Foster 1964; Whittaker and Fernández-Palacios 2007; Rogers and Slatkin 2017). For example, the effects of genetic drift together with inbreeding led to mutation load and inbreeding depression in the population of wolves on Isle Royale (Robinson, et al. 2019) and in Soay sheep (Stoffel, et al. 2021). This is a phenomenon often observed in endangered species with similar demographic changes (Xue, et al. 2015; Bozzuto, et al. 2019). On the other hand, the island environment usually differs from the continental origin of the newly arrived population in abiotic and biotic aspects (Whittaker and Fernández-Palacios 2007). Such environmental change may cause divergent selection on the island populations (Losos and Ricklefs 2009), which can result in morphological and genetical divergence in a short period (Foster 1964; Hofman, et al. 2015; Renaud, et al. 2015; Benítez-López, et al. 2021; Chevret, et al. 2021).

The changes of the island populations in morphology, ecology and behavior compared to their continental relatives are referred to as “island syndrome” (Foster 1964; Adler and Levins 1994; Losos and Ricklefs 2009). A shift of in body size, known as “island rule” or Foster's rule (Foster 1964), is often observed: small species become bigger on islands (Lomolino 1985) while species with large body size tend to become smaller (Benítez-López, et al. 2021). It is commonly observed in rodents on islands whose weight can increase to the double of continental individuals (Angerbjörn 1986; Adler and Levins 1994; Payseur and Jing 2021; Renom, et al. 2021). Genetic studies of the island syndrome have mainly focused on the genes under selection on the island which are related to the phenotypic changes of island populations (Raia, et al. 2010; Choi, et al. 2021; Payseur and Jing 2021). Meanwhile, evidence has been found that the island syndrome may also be caused by relaxion of selection. Island populations may be released from certain selective pressures because of the different environments. For example, lack of predators on the island may lead to change on body size

(Adler and Levins 1994) or coloration (Bliard, et al. 2020). Furthermore, in the face of genetic drift, the efficacy of selection especially on mildly deleterious mutations can be reduced over the genome (Gravel 2016; Peischl, et al. 2018). For example, globally relaxed purifying selection may have affected immune functions of bird populations on islands (Barthe, et al. 2022).

**The Orkney vole**

In this thesis, I mainly focused on the common vole (*Microtus arvalis*) on the Orkney archipelago in the north of Scotland, specifically referred to as the Orkney vole *(M. a. orcadensis*, Millais 1904), as a model to study the genomic consequences of long isolation. Common voles are small herbivorous rodents widely distributed across Europe (Zimmermann 1959; Heckel, et al. 2005; Lischer, et al. 2014). They have very high capability of reproduction and extensive population size fluctuations in many areas (Lambin, et al. 2006). There are four major evolutionary lineages in Europe, which are genetically and geographically distinct, but morphologically cryptic (Heckel, et al. 2005; Lischer, et al. 2014). The origin of the Orkney vole is the Western lineage covering Belgium, France and northern Spain (**Chapter 1**; Heckel, et al. 2005; Martínková, et al. 2013).

The history of Neolithic humans and associated organisms (Stanton, et al. 2016; Chevret, et al. 2021) on Orkney islands has been well studied thanks to many Neolithic settlements in Orkney (Richards 1996). The demographic and genetic history of the Orkney vole is tightly associated with human presence. Orkney voles were introduced by Neolithic farmers over 5,000 years ago likely as a food resource (Romaniuk, et al. 2016). It is one of the oldest known human introductions of a wild species into an island system (Haynes, et al. 2003; Martínková, et al. 2013). Martínková, et al. (2013) found that the source of introduction was most likely the coast of Belgium and northern France based on mitochondrial and microsatellite markers. The Orkney vole is widespread on seven islands of Orkney, and the census population size was estimated to be one million (Reynolds 1992). The populations likely experienced early divergence between the islands isolated by the sea (Phillips 2004; Smith, et al. 2019). The bones of the Orkney vole found in archaeological sites on multiple islands dated to a short period after the introduction (Martínková, et al. 2013; Cucchi, et al. 2014; Fraser 2015; Romaniuk, et al. 2016) support this hypothesis. The Eday population is exceptional compared to the populations with long history, that 15 voles were recently

introduced in 1987. And Eday island was fully covered by fast-expanding voles in three years (Mike Cockram, personal communication), showing Orkney voles' strong ability of reproduction and dispersal even at modern time. The body size of the Orkney vole changed rapidly in less than a century after its introduction (Cucchi, et al. 2014), and it has currently roughly twice the weight compared to its continental conspecifics in Belgium and northern France. Despite the morphological difference, there is no reproductive isolation between the Orkney vole and the common vole on the European continent (Zimmermann 1959).

The genus *Microtus* (Cricetidae) comprises high variation of sex chromosomes, including the typical short Y chromosome (e.g. *M. arvalis*, Lemskaya, et al. 2010), extremely long X and Y chromosomes (e.g. *M. levis*, Lemskaya, et al. 2010), multiple karyotypes in one species (e.g. *M. thomasi*, Rovatsos, et al. 2017), and an X0/XY sex determination system (M. oregoni, Charlesworth and Dempsey 2001), in which the female has only one X chromosome while the male has both X and Y chromosomes. The karyotype of the common vole (2n=46) is formed by 22 pairs of autosomes, an intermediate sized X chromosome and a short Y chromosome. However, the X and Y chromosomes have no synapsis (Ashley, et al. 1990), which is rare among mammals.

**Outline of this thesis**

Overall, the Orkney vole is an exceptional system for the study of the genetics and evolution of bottlenecked and isolated populations in the long term. With its clear history of introduction and analyses with full genomes, we were able to address important evolutionary questions, including the dynamics of mutation load in long isolation, the shifts of selection and its link to the island syndrome, and the effects of demographic changes on the genetic variation on sex chromosomes.

In **Chapter 1**, I reconstructed a detailed demographic history of Orkney voles with genomic resequencing data. Based on this, I then assessed the genomic consequences arising from their ancient colonization events. The strong genetic bottlenecks and exceptional long isolation potentially led to mutation load which reduced the fitness of Orkney voles. With 48 resequenced genomes, I was able to test some of the theoretical predictions on the dynamics of deleterious variation in natural populations after long isolation.

In **Chapter 2**, I examined the exceptional genomic background of Orkney voles shaped by genetic drift and tested for relaxed selection throughout the genome. I also tested for

divergent positive selection to understand the evolutionary changes of Orkney voles, including extreme body size increase compared to the continental conspecifics. To detect hard selective sweeps in the genome, I applied a machine learning algorithm to identify outlier regions potentially under selection, and compared the results with other existing methods of for outlier identification.

In **Chapter 3**, I provided the first draft assembly of the sex chromosomes of the common vole. I then compared the genetic diversity and the dynamics of deleterious mutations between the autosomes and sex chromosomes. Using the genomes of Orkney voles and the continental voles from the Western lineage, I also investigated the different impacts of a population bottleneck on the genetic variation on autosomes and sex chromosomes.

# References


Adler GH, Levins R. 1994. The Island Syndrome in Rodent Populations. The Quarterly Review of Biology 69:473-490.

Ahrens CW, Rymer PD, Stow A, Bragg J, Dillon S, Umbers KDL, Dudaniec RY. 2018. The search for loci under selection: trends, biases and progress. Molecular Ecology 27.

Angerbjörn A. 1986. Gigantism in island populations of wood mice (*Apodemus*) in Europe. Oikos:47-56.

Aris-Brosou S. 2019. Direct evidence of an increasing mutational load in humans. Molecular Biology and Evolution 36:2823-2829.

Ashley T, Jaarola M, Fredga K. 1990. Absence of synapsis during pachynema of the normal sized sex chromosomes of Microtus arvalis. Hereditas 111:295-304.

Ayub Q, Yngvadottir B, Chen Y, Xue Y, Hu M, Vernes SC, Fisher SE, Tyler-Smith C. 2013. FOXP2 targets show evidence of positive selection in European populations. The American Journal of Human Genetics 92:696-706.

Bachtrog D. 2006. A dynamic view of sex chromosome evolution. Current opinion in genetics & development 16:578-585.

Barthe M, Doutrelant C, Covas R, Melo M, Illera JC, Tilak M-K, Colombier C, Leroy T, Loiseau C, Nabholz B. 2022. Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome. BioRxiv:2021.2011.2021.469450.

Benítez-López A, Santini L, Gallego-Zamorano J, Milá B, Walkden P, Huijbregts MAJ, Tobias JA. 2021. The island rule explains consistent patterns of body size evolution in terrestrial vertebrates. Nature Ecology & Evolution 5:768-786.

Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, Morales HE, Van Oosterhout C. 2022. Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics 23:1-12.

Bliard L, Paquet M, Robert A, Dufour P, Renoult JP, Grégoire A, Crochet P-A, Covas R, Doutrelant C. 2020. Examining the link between relaxed predation and bird coloration on islands. Biology Letters 16:20200002.

Bozzuto C, Biebach I, Muff S, Ives AR, Keller LF. 2019. Inbreeding reduces long-term growth of Alpine ibex populations. Nature Ecology and Evolution 3:1359-1364.

Charlesworth B, Dempsey ND. 2001. A model of the evolution of the unusual sex chromosome system of Microtus oregoni. Heredity 86:387-394.

Chevret P, Hautier L, Ganem G, Herman J, Agret S, Auffray J-C, Renaud S. 2021. Genetic structure in Orkney island mice: isolation promotes morphological diversification. Heredity 126:266-278.

Choi JY, Dai X, Alam O, Peng JZ, Rughani P, Hickey S, Harrington E, Juul S, Ayroles JF, Purugganan MD, et al. 2021. Ancestral polymorphisms shape the adaptive radiation of *Metrosideros* across the Hawaiian Islands. Proceedings of the National Academy of Sciences 118:e2023801118.

Colautti RI, Alexander JM, Dlugosch KM, Keller SR, Sultan SE. 2017. Invasions and extinctions through the looking glass of evolutionary ecology. Philosophical Transactions of the Royal Society B: Biological Sciences 372.

Crnokrak P, Barrett SC. 2002. Perspective: purging the genetic load: a review of the experimental evidence. Evolution 56:2347-2358.

Crooks KR, Burdett CL, Theobald DM, King SRB, Di Marco M, Rondinini C, Boitani L. 2017. Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. Proceedings of the National Academy of Sciences of the United States of America 114:7635-7640.

Cucchi T, Barnett R, Martínková N, Renaud S, Renvoisé E, Evin A, Sheridan A, Mainland I, Wickham-Jones C, Tougard C, et al. 2014. The changing pace of insular life: 5000 years of microevolution in the orkney vole (*microtus arvalis orcadensis*). Evolution 68:2804-2820.

DeWoody JA, Harder AM, Mathur S, Willoughby JR. 2021. The long-standing significance of genetic diversity in conservation. Molecular Ecology 30:4147-4154.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756-760.

Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends in Ecology & Evolution 23:347-351.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:610-618.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. Molecular Biology and Evolution 31:1275-1291.

Foster JB. 1964. Evolution of mammals on islands. Nature 202:234-235.

Fraser SM. 2015. Mammals in Late Neolithic Orkney (with reference to mammal bone recovered from Links of Noltland, Westray). The University of Edinburgh.

Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics 165:1269-1278.

González-Martínez SC, Ridout K, Pannell JR. 2017. Range expansion compromises adaptive evolution in an outcrossing plant. Current Biology 27:2544-2551.e2544.

Grant PR, Grant BR. 2014. 40 Years of evolution: Darwin's finches on Daphne Major Island. Princeton University Press.

Gravel S. 2016. When is selection effective? Genetics 203:451-462.

Grossen C, Guillaume F, Keller LF, Croll D. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nature communications 11:1001.

Hahn MW. 2008. Toward a selection theory of molecular evolution. Evolution 62:255-265.

Harris RB, Sackman A, Jensen JD. 2018. On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. PLOS Genetics 14:e1007859.

Haynes S, Jaarola M, Searle JB. 2003. Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. Molecular Ecology 12:951-956.

Heckel G, Burri R, Fink S, Desmet J-F, Excoffier L. 2005. Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. Evolution 59:2231-2242.

Hofman CA, Rick TC, Hawkins MTR, Funk WC, Ralls K, Boser CL, Collins PW, Coonan T, King JL, Morrison SA, et al. 2015. Mitochondrial genomes suggest rapid evolution of dwarf California Channel Islands Foxes (*Urocyon littoralis*). PloS one 10:1-15.

Huisman J, Kruuk LEB, Ellis PA, Clutton-Brock T, Pemberton JM. 2016. Inbreeding depression across the lifespan in a wild mammal population. Proceedings of the National Academy of Sciences 113:3585-3590.

Hurst LD, Ellegren H. 1998. Sex biases in the mutation rate. Trends in Genetics 14:446-452.

Immler S. 2019. Haploid selection in" diploid" organisms. Annual Review of Ecology, Evolution, and Systematics 50:219-236.

Johri P, Aquadro CF, Beaumont M, Charlesworth B, Excoffier L, Eyre-Walker A, Keightley PD, Lynch M, McVean G, Payseur BA, et al. 2022. Recommendations for improving statistical inference in population genomics. PLoS Biology 20:e3001669.

Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM, Tallmon DA, Funk WC. 2021. The crucial role of genome-wide genetic variation in conservation. Proceedings of the National Academy of Sciences 118:e2104642118.

Kimura M, Maruyama T, Crow JF. 1963. The mutation load in small populations. Genetics 48:1303.

Kirkpatrick M, Jarne P. 2000. The effects of a bottleneck on inbreeding depression and the genetic load. The American Naturalist 155:154-167.

Kruess A, Tscharntke T. 1994. Habitat fragmentation, species loss, and biological control. Science 264:1581-1584.

Kutschera VE, Poelstra JW, Botero-Castro F, Dussex N, Gemmell NJ, Hunt GR, Ritchie MG, Rutz C, Wiberg RAW, Wolf JBW. 2019. Purifying selection in corvids is less efficient on islands. Molecular Biology and Evolution 37:469-474.

Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. Science 286:964-967.

Lambin X, Bretagnolle V, Yoccoz NG. 2006. Vole population cycles in northern and southern Europe: Is there a need for different explanations for single pattern? Journal of Animal Ecology 75:340-349.

Laporte V, Charlesworth B. 2002. Effective population size and population subdivision in demographically structured populations. Genetics 162:501-519.

Latter B. 1973. The island model of population differentiation: a general solution. Genetics 73:147-157.

Lemskaya NA, Romanenko SA, Golenishchev FN, Rubtsova NV, Sablina OV, Serdukova NA, O'Brien PC, Fu B, Yiğit N, Ferguson-Smith MA. 2010. Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). III. Karyotype relationships of ten *Microtus* species. Chromosome research 18:459-471.

Li J, Li H, Jakobsson M, Li S, SjÖDin P, Lascoux M. 2012. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Molecular Ecology 21:28-44.

Lischer HEL, Excoffier L, Heckel G. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: Implications for the evolutionary history of *Microtus* voles. Molecular Biology and Evolution 31:817-831.

Lohmueller KE. 2014. The distribution of deleterious genetic variation in human populations. Current opinion in genetics & development 29:139-146.

Lomolino MV. 1985. Body size of mammals on islands - the island rule reexamined. American Naturalist 125:310-316.

Losos JB, Ricklefs RE. 2009. Adaptation and diversification on islands. Nature 457:830-836.

Luque GM, Bellard C, Bertelsmeier C, Bonnaud E, Genovesi P, Simberloff D, Courchamp F. 2014. The 100th of the world's worst invasive alien species. Biological Invasions 16:981-985.

Lynch M, Conery J, Burger R. 1995. Mutation accumulation and the extinction of small populations. The American Naturalist 146:489-518.

Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350:1493-1498.

Mallet M, Chippindale A. 2011. Inbreeding reveals stronger net selection on *Drosophila melanogaster* males: implications for mutation load and the fitness of sexual females. Heredity 106:994-1002.

Marques DA, Jones FC, Di Palma F, Kingsley DM, Reimchen TE. 2018. Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. Nature Ecology and Evolution 2.

Martínková N, Barnett R, Cucchi T, Struchen R, Pascal M, Pascal M, Fischer MC, Higham T, Brace S, Ho SYW, et al. 2013. Divergent evolutionary processes associated with colonization of offshore islands. Molecular Ecology 22:5205-5220.

Mayr E. 1963. Animal species and evolution. Belknap Press of Harvard University Press.

Millais J. 1904. On a new British vole from the Orkney islands. Zoologist 8:241-246.

Moinet A, Schlichta F, Peischl S, Excoffier L. 2022. Strong neutral sweeps occurring during a population contraction. Genetics 220.

Nei M, Maruyama T, Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. Evolution 29:1-10.

Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18:375-402.

Payseur BA, Jing P. 2021. Genomic targets of positive selection in giant mice from Gough Island. Molecular Biology and Evolution 38:911-926.

Peischl S, Dupanloup I, Bosshard L, Excoffier L. 2016. Genetic surfing in human populations: from genes to genomes. Current Opinion in Genetics and Development 41:53-61.

Peischl S, Dupanloup I, Foucal A, Jomphe M, Bruat V, Grenier JC, Gouy A, Gilbert KJ, Gbeha E, Bosshard L, et al. 2018. Relaxed selection during a recent human expansion. Genetics 208:763-777.

Peischl S, Excoffier L. 2015. Expansion load: recessive mutations and the role of standing genetic variation. Molecular Ecology 24:2084-2094.

Perrier A, Sánchez-Castro D, Willi Y. 2020. Expressed mutational load increases toward the edge of a species' geographic range. Evolution:evo.14042.

Phillips T. 2004. Seascapes and landscapes in Orkney and northern Scotland. World archaeology 35:371-384.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410-1414.

Pringle RM, Kartzinel TR, Palmer TM, Thurman TJ, Fox-Dobbs K, Xu CCY, Hutchinson MC, Coverdale TC, Daskin JH, Evangelista DA, et al. 2019. Predator-induced collapse of niche structure and species coexistence. Nature 570:58-64.

Qanbari S, Pimentel E, Tetens J, Thaller G, Lichtner P, Sharifi A, Simianer H. 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. Animal Genetics 41:377-389.

Rafati N, Chen J, Herpin A, Pettersson ME, Han F, Feng C, Wallerman O, Rubin C-J, Péron S, Cocco A, et al. 2020. Reconstruction of the birth of a male sex chromosome present in Atlantic herring. Proceedings of the National Academy of Sciences:202009925.

Raia P, Guarino FM, Turano M, Polese G, Rippa D, Carotenuto F, Monti DM, Cardi M, Fulgione D. 2010. The blue lizard spandrel and the island syndrome. BMC evolutionary biology 10:1-16.

Renaud S, Gomes Rodrigues H, Ledevin R, Pisanu B, Chapuis J-L, Hardouin EA. 2015. Fast evolutionary response of house mice to anthropogenic disturbance on a Sub-Antarctic island. Biological Journal of the Linnean Society 114:513-526.

Renom P, de-Dios T, Civit S, Llovera L, Sánchez-Gracia A, Lizano E, Rando JC, Marquès-Bonet T, Kergoat GJ, Casanovas-Vilar I. 2021. Genetic data from the extinct giant rat from Tenerife (Canary Islands) points to a recent divergence from mainland relatives. Biology Letters 17:20210533.

Reynolds P. 1992. The impact of changes in land-use in Orkney, on the vole *Microtus arvalis orcadensis* and its avian predators. University of Aberdeen.

Richards C. 1996. Monuments as landscape: creating the centre of the world in late Neolithic Orkney. World archaeology 28:190-208.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK. 2018. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Current Biology 28:3487-3494.e3484.

Robinson JA, Räikkönen J, Vucetich LM, Vucetich JA, Peterson RO, Lohmueller KE, Wayne1 RK. 2019. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Science Advances 5:eaau0757.

Rogers RL, Slatkin M. 2017. Excess of genomic defects in a woolly mammoth on Wrangel island. PLOS Genetics 13:e1006601.

Romaniuk AA, Shepherd AN, Clarke DV, Sheridan AJ, Fraser S, Bartosiewicz L, Herman JS. 2016. Rodents: food or pests in Neolithic Orkney. Royal Society open science 3:160514.

Rovatsos MT, Marchal JA, Romero-Fernández I, Arroyo M, Athanasopoulou EB, Sánchez A. 2017. Extensive Sex Chromosome Polymorphism of *Microtus thomasi/Microtus atticus* Species Complex Associated with Cryptic Chromosomal Rearrangements and Independent Accumulation of Heterochromatin. Cytogenetic and Genome Research 151:198-207.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature

449:913-918.

Schipper J, Chanson JS, Chiozza F, Cox NA, Hoffmann M, Katariya V, Lamoreux J, Rodrigues AS, Stuart SN, Temple HJ. 2008. The status of the world's land and marine mammals: diversity, threat, and knowledge. Science 322:225-230.

Schubert E, Sander J, Ester M, Kriegel HP, Xu XW. 2017. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. Acm Transactions on Database Systems 42:1-21.

Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is insensitive to recent population history. Nature Genetics 46:220-224.

Smith DE, Barlow NL, Bradley SL, Firth CR, Hall AM, Jordon J, Long D. 2019. Quaternary sea level change in Scotland. Earth and Environmental Science Transactions of the Royal Society of Edinburgh 110:219-256.

Soh YS, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, Fulton RS, Kremitzki C, Koutseva N, Mueller JL. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159:800-813.

Stanton DWG, Mulville JA, Bruford MW. 2016. Colonization of the Scottish Islands via long-distance Neolithic transport of red deer (*Cervus elaphus*). Proceedings of the Royal Society B: Biological Sciences 283:20160095.

Stephan W. 2016. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Molecular Ecology 25:79-88.

Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM. 2021. Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. Nature communications 12:2972.

Szövényi P, Devos N, Weston DJ, Yang X, Hock Z, Shaw JA, Shimizu KK, McDaniel SF, Wagner A. 2014. Efficient purging of deleterious mutations in plants with haploid selfing. Genome biology and evolution 6:1238-1252.

Tomaszkiewicz M, Medvedev P, Makova KD. 2017. Y and W chromosome assemblies: approaches and discoveries. Trends in Genetics 33:266-282.

Valero M, Richerd S, Perrot V, Destombe C. 1992. Evolution of alternation of haploid and diploid phases in life cycles. Trends in Ecology & Evolution 7:25-29.

van der Valk T, de Manuel M, Marques-Bonet T, Guschanski K. 2021. Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. BioRxiv:696831.

Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nature Ecology & Evolution 3:1632-1641.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biology 4:e72.

Weigand H, Leese F. 2018. Detecting signatures of positive selection in non-model species using genomic data. Zoological Journal of the Linnean Society 184:528-583.

Whittaker RJ, Fernández-Palacios JM. 2007. Island biogeography: ecology, evolution, and conservation: Oxford University Press.

Wolf JBW, Ellegren H. 2016. Making sense of genomic islands of differentiation in light of speciation. Nature 18:87-100.

Xue Y, Prado-martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, et al. 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science 348:242-245.

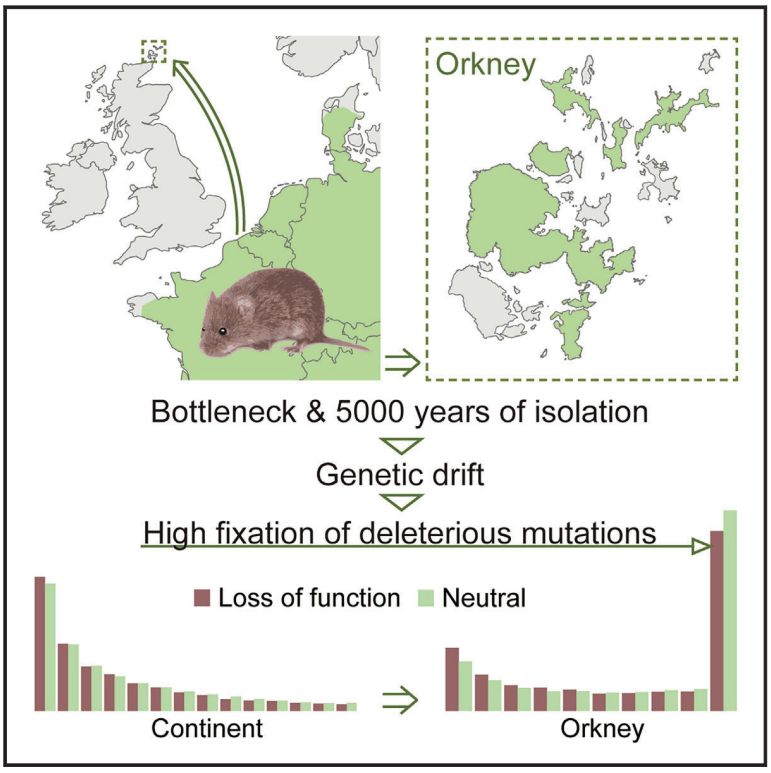
Zimmermann K. 1959. Über eine Kreuzung von Unterarten der Feldmaus *Microtus arvalis*. Zool. Jb. Syst 87:1-12.

Article

Current Biology

# Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal

## Graphical abstract




## Highlights

- Common voles on Orkney remained completely isolated for more than 5,000 years
- Genetic drift led to a strong reduction of genetic diversity and population divergence
- Orkney voles have high levels of detrimental mutations, especially on small islands
- Simulations suggest purging of highly deleterious alleles, while mild ones persisted

## Authors

Xuejing Wang, Stephan Peischl, Gerald Heckel

## Correspondence

gerald.heckel@unibe.ch

## In brief

Wang et al. show that common voles on the Orkney archipelago have remained genetically isolated for more than 5,000 years after human introduction. Orkney voles lost most genetic diversity and harbor high levels of inferred strongly deleterious mutations, yet simulations and large current population sizes suggest rather mild effects on fitness.

Wang et al., 2023, Current Biology *33*, 1–16
May 22, 2023 
https://doi.org/10.1016/j.cub.2023.04.042

CellPress

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042

Current Biology

CellPress
OPEN ACCESS

Article

# Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal

Xuejing Wang,[1] Stephan Peischl,[2,3] and Gerald Heckel[1,3,4,*]
[1]Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
[2]Interfaculty Bioinformatics Unit, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
[3]Swiss Institute of Bioinformatics, Amphipôle, Quartier UNIL-Sorge, 1015 Lausanne, Switzerland
[4]Lead contact
*Correspondence: gerald.heckel@unibe.ch
https://doi.org/10.1016/j.cub.2023.04.042

## SUMMARY

Increased human activities caused the isolation of populations in many species—often associated with genetic depletion and negative fitness effects. The effects of isolation are predicted by theory, but long-term data from natural populations are scarce. We show, with full genome sequences, that common voles (*Microtus arvalis*) in the Orkney archipelago have remained genetically isolated from conspecifics in continental Europe since their introduction by humans over 5,000 years ago. Modern Orkney vole populations are genetically highly differentiated from continental conspecifics as a result of genetic drift processes. Colonization likely started on the biggest Orkney island and vole populations on smaller islands were gradually split off, without signs of secondary admixture. Despite having large modern population sizes, Orkney voles are genetically depauperate and successive introductions to smaller islands resulted in further reduction of genetic diversity. We detected high levels of fixation of predicted deleterious variation compared with continental populations, particularly on smaller islands, yet the fitness effects realized in nature are unknown. Simulations showed that predominantly mildly deleterious mutations were fixed in populations, while highly deleterious mutations were purged early in the history of the Orkney population. Relaxation of selection overall due to benign environmental conditions on the islands and the effects of soft selection may have contributed to the repeated, successful establishment of Orkney voles despite potential fitness loss. Furthermore, the specific life history of these small mammals, resulting in relatively large population sizes, has probably been important for their long-term persistence in full isolation.

## INTRODUCTION

Isolated populations are of major concern in conservation biology, human and animal health, and evolutionary biology.[1,2] As human activities have affected and will continue to affect ecosystems worldwide, fragmentation of many species into isolated populations is ongoing and may lead to increased risks of local or global extinction.[3,4] Human transfer of organisms to previously unoccupied areas can result in populations that may either vanish or thrive and turn into biological invasions given suitable conditions.[5,6] Even though both scenarios typically involve isolated populations, their demographic trajectories are diametrically opposed and are thought to be influenced by the amount and quality of genetic variation.[7]

Isolation of populations leads to loss of genetic diversity over time, and the rate of the loss is inversely related to the effective population size ($N_e$).[2] In small populations, the loss of genetic diversity is mostly governed by random genetic drift, which affects neutral and non-neutral genetic variation equally, and might leave long-lasting signatures even if a population expanded afterward.[8,9] Deleterious mutations may reduce the fitness of populations, a phenomenon for which the term "mutation load" was coined.[10–13] Deleterious mutations are usually present at low frequency in large populations and mostly express their negative fitness effects when inbreeding occurs (inbreeding load).[14] The loss of diversity in small populations can cause recessive deleterious variants to drift to high frequency and be more likely present in homozygous form. This can lead to an increased mutation load (drift load)[15] of mildly and moderately deleterious mutations but also to efficient purging of strongly deleterious mutations over time by purifying selection.[16–18]

Theory has shown that, at the front of spatially expanding populations initially going through a series of bottlenecks, the proportion of deleterious alleles in a homozygous state may increase while the total number of deleterious alleles remains unchanged.[19,20] Such an increase correlates with a decrease in fitness of derived populations and has thus been called "expansion load." The expansion load can last for thousands of generations, especially in the absence of gene flow from core populations.[21] Empirical studies on the genetic effects of deleterious variation in isolated populations have mainly focused either on model species[19,22,23] in experiments[24,25] or on

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042



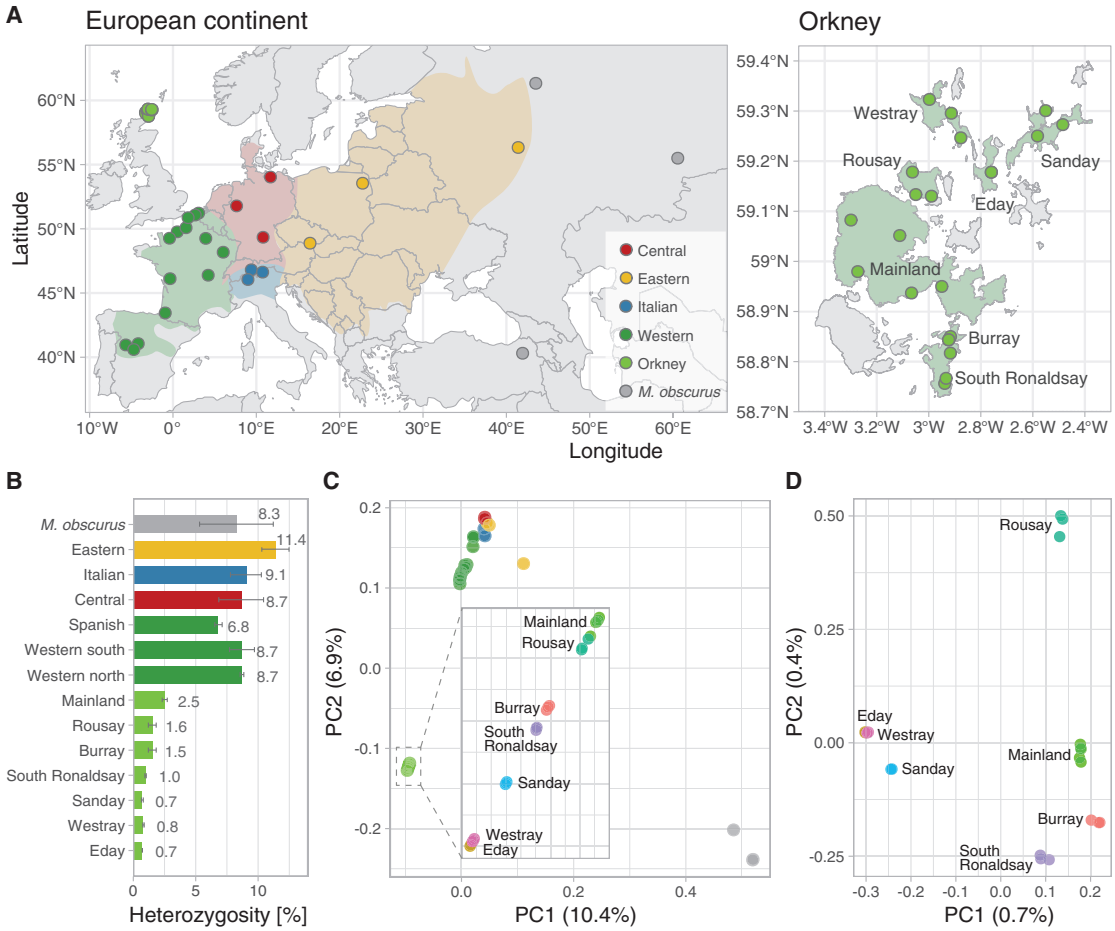



**Figure 1. Distribution, genetic diversity and divergence of common voles**

(A) Sampling locations and distribution range of the common vole *Microtus arvalis* modified after IUCN Red List.[42,43] Major evolutionary lineages in *Microtus arvalis* are coded as: Eastern lineage, yellow; Italian lineage, blue; Central lineage, red; Western lineage, dark green; Orkney, light green. Sampling locations of *M. obscurus* are marked with gray dots. One genome was sequenced for each location except for Eday (n = 2) and one location of Burray (n = 2).

(B) The average percentage of heterozygous sites and the standard deviation marked as the error bar in each group.

(C) Principal component analysis (PCA) of genetic variation based on 57.5 million SNPs without missing data (54.9% of all) with a zoom of the Orkney voles in the insert (x axis: −0.097 to −0.092; y axis: −0.128 to −0.118). Patterns in Orkney populations correlate with heterozygosity in (B) and indicate the varied strength of genetic drift during the history.

(D) PCA of Orkney vole genomes only.

See also Tables S1–S3.

endangered wild populations,[26–30] but knowledge of the effects in the wild exceeding a period of a few hundred generations[29,30] remains scarce.

Islands are natural systems to study the short- and long-term effects of isolation on populations.[2,31,32] The Orkney archipelago in the north of Scotland has served as an iconic model for studying the history of Neolithic human settlements and culture[33] and associated organisms.[34,35] The Orkney vole (*Microtus arvalis orcadensis*)[36] results from one of the oldest-known human introductions of a wild species to an island system[37,38] and provides a unique system to study the long-term consequences of genetic isolation. Common voles on Orkney are geographically separated from their conspecifics in continental Europe (Figure 1) and were once believed to be an independent species based on morphology. Common voles are small herbivorous rodents with a very high capability of reproduction and extensive population size fluctuations in many areas.[39] In Europe, four major evolutionary lineages are found that are genetically and geographically distinct but morphologically cryptic (Figure 1A).[40,41]

Limited mitochondrial and nuclear DNA information has shown that Orkney voles belong to the Western evolutionary lineage within the species and were most likely introduced by Neolithic farmers from coastal Belgium or France more than 5,000 years ago,[37,38] probably as a food item.[44] Orkney voles are present and abundant on seven of the islands, with their census population size estimated to be over one million.[45] Their bones have been found on multiple Orkney islands in archeological sites dating back to the Neolithic, which supports early splits between vole populations[38,44,46,47] at a time when most islands of Orkney were already isolated by the sea.[48,49] However, the last intentional introduction of voles to an unoccupied Orkney island,

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042





**Table 1. Genomic diversity of common voles (*Microtus arvalis*) and the outgroup *M. obscurus***

| Lineage | Group | Number of genomes | Heterozygosity | Deleterious homozygous | Deleterious heterozygous | Deleterious sites total | Deleterious alleles |
|---|---|---|---|---|---|---|---|
| *M. obscurus* | – | 3 | 8.3% | – | – | – | – |
| Eastern | – | 3 | 11.4% | – | – | – | – |
| Italian | – | 3 | 9.1% | – | – | – | – |
| Central | – | 3 | 8.7% | – | – | – | – |
| Western | north | 7 | 8.7% | 1,371 | 2,661 | 4,032 | 5,468 |
| | south | 3 | 8.7% | 1,598 | 2,849 | 4,447 | 6,045 |
| | Spanish | 3 | 6.8% | 1,995 | 2,456 | 4,451 | 6,446 |
| | overall | 13 | 8.2% | 1,585 | 2,657 | 4,242 | 5,827 |
| Orkney | Mainland | 5 | 2.5% | 2,419 | 911 | 3,330 | 5,750 |
| | Rousay | 3 | 1.6% | 2,580 | 597 | 3,177 | 5,757 |
| | Burray | 3 | 1.5% | 2,602 | 597 | 3,199 | 5,801 |
| | South Ronaldsay | 3 | 1.0% | 2,690 | 431 | 3,121 | 5,813 |
| | Sanday | 3 | 0.7% | 2,662 | 334 | 2,996 | 5,659 |
| | Westray | 3 | 0.8% | 2,683 | 383 | 3,066 | 5,749 |
| | Eday | 2 | 0.7% | 2,705 | 331 | 3,036 | 5,741 |
| | overall | 22 | 1.4% | 2,598 | 556 | 3,155 | 5,753 |

Groups in the Western lineage are defined based on geographical origin. Heterozygosity is shown as the average percentage of heterozygous sites. Deleterious genetic variation in Orkney voles and the Western evolutionary lineage is reported as the average numbers of derived deleterious homozygous and heterozygous sites, deleterious sites, and deleterious alleles, rounded up to the whole number.

Eday, occurred only in 1987 (M. Cockram, personal communication), which indicates the potential of human interference with the demographic and genetic history of seemingly isolated island populations.

Here, we use full genomic data to first reconstruct the history of Orkney voles throughout the Neolithic until the modern period relative to their continental conspecifics. On this base, we then assess the genomic consequences of ancient colonization events in Orkney potentially leading to strong genetic bottlenecks, thousands of generations of isolation, and mutation load. This allows us to empirically test some of the theoretical predictions regarding changes in the number, frequency, and heterozygosity of deleterious variation in natural mammal populations after extremely long isolation.

## RESULTS

### Genomic diversity and diversification

We sequenced 45 common vole and three additional *Microtus obscurus* genomes at a mean read depth of 26× (Tables 1 and S1). This included 22 individuals from all seven Orkney islands occupied by *M. arvalis*, 14 individuals chosen to cover the continental distribution range of the Western evolutionary lineage in the species, and three samples from each of the other major evolutionary lineages (Figure 1A). The genetic diversity of Orkney voles, reported as the average percentage of heterozygous sites among all called single-nucleotide polymorphism (SNP, 100.8 million in total) sites of each population, was extremely low compared with continental populations (Figure 1B; Table 1). On the European continent, genetic diversity in the Eastern evolutionary lineage was the highest (11.4%). Within the Western evolutionary lineage, Spanish populations had the lowest heterozygosity (6.8%), while Western-south and Western-north populations had equal levels (8.7%) despite the divergence within the lineage. In Orkney, voles on Mainland Orkney had the highest heterozygosity (2.5%), which is 1.5–3.4 times more than in the other islands (Figure 1B). Orkney voles on islands closer to Mainland Orkney had higher heterozygosity than voles on the islands—Sanday, Westray, and Eday—farther to the north (Figures 1A and 1B). Though the Eday vole population started from only 15 individuals from Westray released in 1987, the population did not show a strong decrease in heterozygosity compared with Westray.

A principal component analysis (PCA) of all genomes without missing data showed Orkney voles as highly divergent from the continental populations (Figure 1C). Most of the variance was explained by differences between *M. arvalis* and *M. obscurus*, but Orkney voles formed a cluster separate from all other individuals spanning the distribution range of the species between Spain and Russia. On PC1, PC2, and PC3, Orkney voles were closest to individuals from the northern part of the Western evolutionary lineage (Table S2). In the detailed view, populations are ordered almost perfectly linear to levels of heterozygosity (Figure 1B) probably due to the effects of genetic drift. Orkney Mainland was closest to the continental populations, which gives support to this largest island as the original founding site of Orkney voles. The pattern in the PCA showed Rousay and Burray as genetically closest to Mainland Orkney, consistent with the geography. South Ronaldsay was beyond Burray, and populations from the most distant northern islands, Sanday and Westray, were farther out (Figure 1A; Table S3). As expected, the Eday population was close to Westray voles that were the source of the introduction. Focusing only on Orkney voles confirmed the genetic distinctness of populations on separate islands and little variation within populations. The three northern islands separated from the rest on PC1, and the

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042



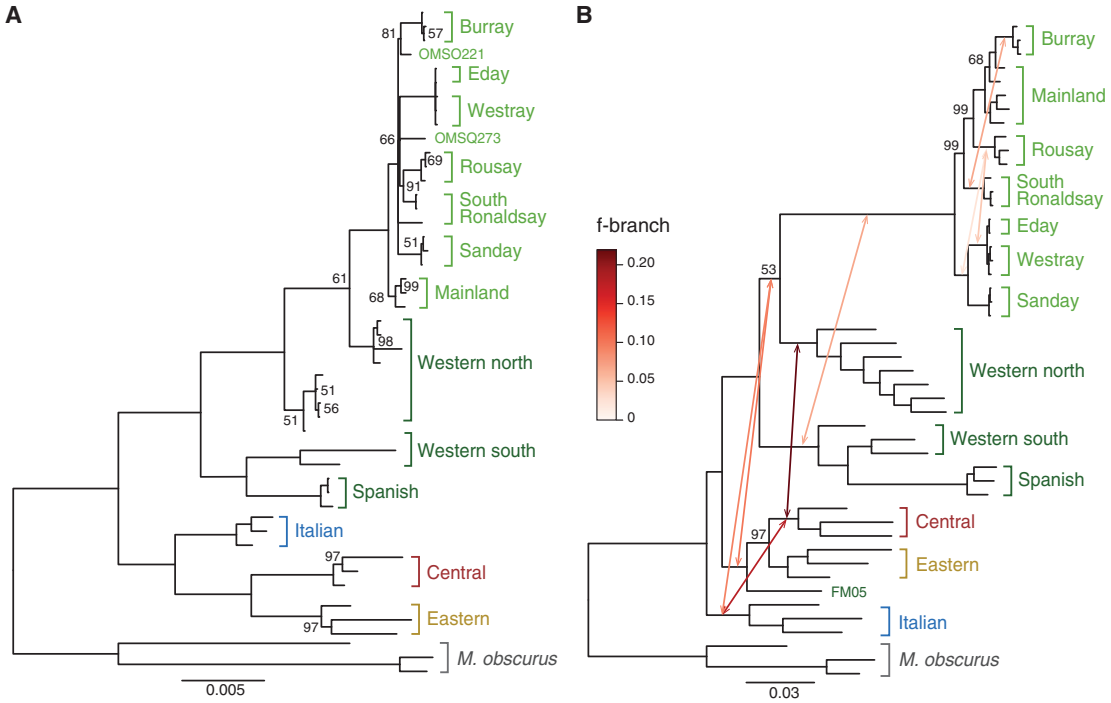



**Figure 2. Phylogenetic relationships and gene flow inference**

(A) Bayesian phylogeny based on complete mitochondrial vole genomes.

(B) ML tree conducted with IQ-TREE based on 5 million random nuclear single-nucleotide polymorphism sites without missing data. Arrows show the *f*-statistics with $p < 0.01$ and the colors representing the values of *f*-statistics. Note that *f*-statistics do not represent the direction of gene flow. All posterior probabilities in (A) and (B) were equal to 1 unless shown as percentage on the tree. Information on evolutionary lineage and grouping of the vole individuals is given in Table 1. Individuals that did not cluster with the geographical groups are marked with the sample name.

See also Figures S1 and S2.

southern islands separated from each other on PC2 (Figure 1D; Table S4).

## Phylogeny and admixture

Phylogenetic analyses of both nuclear and mitochondrial genomes showed Orkney voles as distinct, with the closest affinity to individuals from the northern area of the Western evolutionary lineage (Figure 2). Mainland individuals were either at basal positions or grouped with voles from other islands on the mitochondrial tree, potentially due to preserved ancestral diversity in this biggest Orkney population.[38]

Nuclear genomes strongly supported the notion that Orkney voles have remained isolated from continental populations for a very long period, as no evidence of post-introduction admixture between Orkney populations and the continent was found (Figures 2B, S1A, and S2). Signals of genomic admixture were mostly detected between basal branches on the continent, which is consistent with the sharing of ancestral variation and limited gene flow due to incomplete reproductive isolation between evolutionary lineages.[42,50–52] An individual from north-eastern France (FM05) showed strong admixture (Figure S1A) between the Western-north and the Central lineage and was thus not included in further, more specific analyses. *f*-branch statistics suggested also limited ancestral gene flow between continental lineages and Orkney voles (Figures 2B and S2). The signal of admixture between the Western-south branch and ancestral Orkney is likely a consequence of shared ancestral variation that was passed over to the Orkney population at the time of founding (for details, see Malinsky et al.[53]). The results of ADMIXTURE analyses support that Orkney vole genomes are distinct from current continental populations, with only very minor sharing of ancestries (Figure S1).

Within the monophyletic cluster of Orkney samples, individuals from each island clustered separately except for Orkney Mainland (Figures 2 and S1B), suggesting isolation and divergence between the islands. Populations from the northern Orkney islands (Sanday, Westray, and Eday) were diverged from the rest in the phylogeny and in ADMIXTURE analyses (Figure S1B). *f*-branch statistics suggested multiple admixture events between the Orkney islands, mainly at the early stage of divergence (Figures 2B and S2). Treemix or ADMIXTURE detected no signals of admixture between islands, except for the South Ronaldsay and Burray populations (Figure S1B), where recent gene flow between these nearby islands cannot be excluded.

## Demographic history of Orkney voles

We used the Orkney Mainland population to represent Orkney voles in demographic analyses because of its preserved diversity and absence of potential admixture after introduction (Figures 2B and S2). Analyses with SMC++[54] showed a strong drop in the effective population size of Orkney Mainland voles

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042





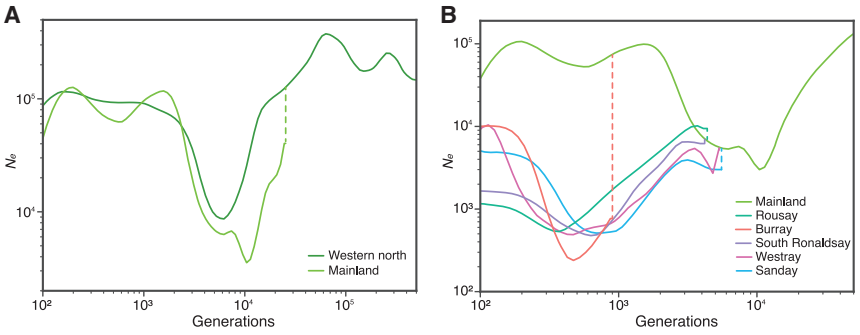



**Figure 3. Demographic history of Orkney voles**

(A) Demographic histories of common voles on Orkney Mainland relative to the genetically and geographically closest Western-north populations on the continent inferred by SMC++. Solid curves show variation in the effective population size $N_e$ over time and the dashed vertical line represents the inferred time of divergence. The generation time is 0.5 year.

(B) Demographic histories of voles on smaller Orkney islands relative to Orkney Mainland. Multiple analyses were combined into one graph for comparison. The Eday population is not shown because of its very recent introduction from Westray in 1987. See also Table S4.

ca. 5,300 years ago (bootstrap median 4,875, 5$^{th}$ percentile 3,014, 95$^{th}$ percentile 5,410), with a minimum $N_e$ of 2,781 (Figure 3B; bootstrap median 3,153, 5$^{th}$ percentile 2,275, 95$^{th}$ percentile 3,771). The time of this bottleneck closely fits the dates from radiocarbon analyses on vole bones (oldest 5,100 years old) and the previously estimated time (about 5,000 years) of the introduction to Orkney.[38] The split between Orkney Mainland and the continental population was estimated to be ~8,400 years ago (Figure 3A; Table S4). This is much older than the bottleneck on Orkney Mainland and probably a consequence of the Western-north individuals not being the descendants of the precise source populations for the founding of Orkney (see discussion).[38] After a relatively slow initial population size increase in Orkney, the vole population grew particularly fast in the last millennium and remained, with some fluctuation, at an $N_e$ of roughly 100,000. Surprisingly, this is the same order of magnitude that the population size in the Western-north group reached after a less severe bottleneck about 2,000–3,000 years ago.

The comparison between Orkney islands showed that the demographic history of Orkney Mainland voles differed from populations on the smaller islands, but there were similar trends among the latter (Figure 3B). With the Mainland population as the potential source, the splits of most island populations (Sanday, Westray, Rousay, and South Ronaldsay) were estimated at between 2,000 and 3,000 years ago (Table S4). The population on Burray split only about 450 years ago, at a time when all Orkney populations, including the Mainland, experienced a phase with the lowest effective population sizes in their history (Figure 3B). Populations on the smaller islands showed partially signs of very strong growth only in the last 200 years (e.g., on Burray from ~250 to 10,000). We attempted also to get insights into the history of the recently introduced voles on Eday through analyses with SMC++. The estimated population size changes are relatively similar to those on Westray, the source for the introduction. The split time of about 3 years ago (Table S4) is clearly an underestimate, but this is not surprising because of the limits of SMC++ to estimate very recent demographic changes. Events occurring less than ~160 generations since introduction are too recent to be estimated reliably.[55] Orkney voles had runs of homozygosity (ROH) longer than 1 Mb, covering a total of 2.1%–5.4% of the genome (mean = 3.6%), which is 8–20 times longer compared with the continental voles (mean = 0.25%; range 0%–0.35%). However, the ROH regions in Orkney voles were short (mean 1.4 Mb, longest 3.4 Mb; mean continental voles 1.6 Mb, longest 4.1 Mb), indicating that the ROHs resulted from the deeper demographic history rather than inbreeding due to recent common ancestors.

## Accumulation of deleterious mutations

Assessing the potential functional relevance of genetic variation showed that Orkney voles accumulated high amounts of homozygous, putatively deleterious mutations in their specific history compared with continental populations (Figures 4A, 4B, and S3A; Table 1). We identified in total 11,562 SNPs and 22,006 indels with defined derived states in Orkney that were classified as highly deleterious (mainly loss-of-function mutations, e.g., frameshift mutations; see STAR Methods), with 9.7 million neutral SNPs and 1.8 million neutral indels as comparison. On Orkney Mainland, over 40% of the deleterious variants were fixed in the population. On smaller islands, 63.6%–89.9% of the deleterious variants were fixed (Figures 4A and S3A). The majority of the fixed variants were shared between Mainland Orkney and the other islands, probably inherited from the ancestral population of Orkney (Figure S4). Gene enrichment analysis suggested that the affected genes were distributed widely in many fundamental pathways without specific enrichment, and we did not detect under-represented gene ontology terms (results not shown).

We used individual-based simulations to investigate whether this large proportion of fixations is a plausible outcome of the estimated demographic history. Using a distribution of fitness effects (gamma distributed) with mean $s = -0.01$ (Ns = 50; N being the effective population size during the bottleneck), we found very good agreement between observed and simulated site frequency spectra (SFS) for both neutral and deleterious alleles (Figure 4A). The mean strength of selection and the shape parameter of the distribution of fitness effects (DFE, alpha = 1) were chosen such that a broad range of selection coefficients are contained in the DFE. Using refined DFEs (Figure S5) that incorporate a negative relationship between the strength of selection and the dominance coefficient yielded very similar SFS,[56] in particular whether the DFE was wide enough to include a substantial proportion of nearly neutral mutations (Ns < 1 during the bottleneck, Figure S5). Re-tracing the fixation events in the simulations showed that they mostly occurred during the ancestral bottleneck more than 5,000 years ago (Figure S5C). Furthermore, virtually all mutations that became fixed fell below

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042



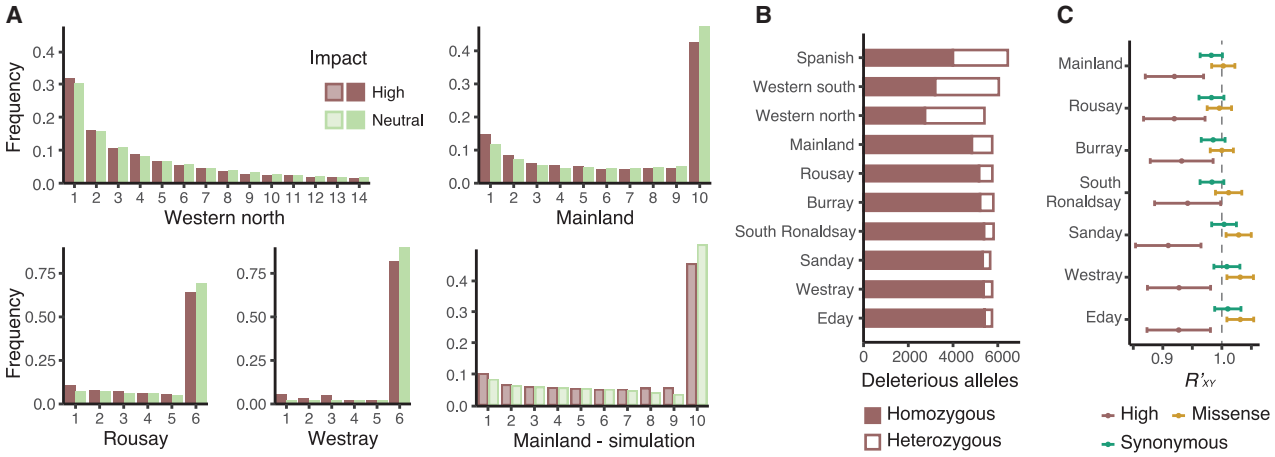



**Figure 4. Accumulation of deleterious genetic variation in Orkney voles**

(A) Unfolded site frequency spectra of derived high-impact (dark red) and neutral sites (green) for real data, including nucleotide polymorphism sites (SNP) and indels, or for simulated data. A very high proportion of deleterious mutations were fixed in Orkney vole populations.

(B) Number of high-impact deleterious alleles per individual for Orkney and continental voles from the Western evolutionary lineage. Deleterious alleles were more often found in homozygous state in Orkney voles.

(C) $R'_{XY}$ of highly deleterious, missense, and synonymous SNPs of Orkney populations. The error bars stand for ± 2 standard errors.

See also Figures S3–S5.

the threshold for nearly neutral mutations during the bottleneck (Figure S5). This shows that only mildly deleterious mutations could fix, whereas more strongly deleterious mutations were successfully purged from the population in the simulations. Including a proportion of lethal and sub-lethal mutations (3% of mutations with s = −0.75 or s = −1) did not affect our simulation results (Figure S3), in line with the observation that genomic patterns of deleterious diversity are composed almost exclusively of small effect mutations across all simulations.

We investigated the accumulation of deleterious alleles in empirical data with the $R_{XY}$ method.[57,58] The ratio $R'_{XY}$ of highly deleterious mutations in Orkney populations was close but significantly lower than 1 when calculated against the Western-north group (Figure 4C). This showed that a number of highly deleterious alleles were removed from Orkney populations due to purging,[58] especially in the northern islands (Sanday, Westray, and Eday). However, $R'_{XY}$ values of missense SNPs exceeding 1 suggest that the removal of weakly deleterious alleles was less efficient in Orkney populations (Figure 4C). We further tested for evidence of pseudo-overdominance (POD) in Orkney voles that could potentially reduce the expression of fitness effects. POD is a form of balancing selection in regions of low recombination that leads to the masking of recessive deleterious variants by maintaining complementary haplotypes at high frequencies.[59,60] However, we did not find signatures of POD, such as an excess of deleterious mutations in low recombination rate regions nor an excess of intermediate frequency variants in the SFS of deleterious variants (Figures 4A and S4). Also, the SFS from our simulations showed no evidence of potential POD affecting Orkney voles (Figure 4A and S5).

The genomic patterns of homozygosity of Orkney populations were consistent with the theoretical predictions for range expansions. The number of sites with derived deleterious alleles in a homozygous state observed per genome was strongly correlated with the number of sites with derived neutral alleles in homozygous states ($R^2$ = 0.992, p < 2.2e−16). The same pattern was found for the number of heterozygous sites ($R^2$ = 0.992, p < 2.2e−16). This correlation is mainly due to genetic drift because the alleles with lower initial frequency are more likely to be lost during the bottleneck.[61] Despite much lower heterozygosity overall, the total number of deleterious alleles per individual in Orkney voles was comparable to their conspecifics on the continent (t test; Orkney versus Western lineage p = 0.96; Figure 4B). Very similar patterns were observed when classifying the mutations according to different functional consequences (frameshift mutations, premature stop codons, start codon lost and missense mutations; Figure S3B).

## DISCUSSION

Our analyses show that the Orkney vole is likely one of the oldest cases of human introduced wild species on islands with complete genetic isolation for more than 5,000 years (Figures 2 and 3A). The early separation between Orkney islands makes these voles an ancient ''experiment'' with repeated trials on the evolution of isolated populations. The outcome of this exceptionally long-lasting experiment supports population genetics theory on the consequences of bottlenecks and founder events[8,10,62,63] beyond most time frames investigated in natural systems so far.[64,65] Remarkably, the demographic trajectories of the populations are largely decoupled from the depauperate genomic background of Orkney voles. Orkney voles have high modern $N_e$ as indicated by demographic reconstruction (Figure 3; Table S4) and they have experienced ecological success,[45] despite the ubiquitous genomic signs associated with potential negative fitness consequences that still persist after thousands of generations in isolation.

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042





## Multiple replicates of long-lasting isolation induced by humans

Genetic isolation of Orkney voles for over 10,000 generations after human introduction is exceptional among the studied systems of vertebrates. The most similar example is the island fox (*Urocyon littoralis*), which started colonizing the California Channel Islands about 9,000 years ago (9,000 generations for fox), and for which the later close association with humans contributed to their spread and recent gene flow among populations.[64] Other completely isolated vertebrate populations with relatively clear history have been separated from conspecific populations for hundreds of generations (e.g., Italian brown bears[66]) and, rarely, a few thousand generations (e.g., Soay sheep[29]). Except for domestic animals like sheep,[29] there was either no human transfer involved in establishing these completely isolated populations or it remains unknown. Overall, human introductions of vertebrates apart from domestic animals are very numerous but they concern mostly comparatively very recent cases that led to biological invasions.[6,67,68]

The strongest bottleneck in the history of Orkney voles ca. 5,300 years ago probably represents the establishment of founding populations[46] with increasing human occupation at a time when most Orkney islands were already separated by the North Atlantic Ocean.[48,49] This time point is consistent with the oldest vole remains on Mainland dated 4,800 years old, while the earliest archeological evidence of voles on Orkney dated to 5,600 years ago stems from the island Papa Westray in the north of Westray.[38] The estimated split between Orkney Mainland and the continental population ~8,400 years ago coincides with the earliest evidence for Mesolithic humans in Orkney,[69] but this time estimate may have been pushed earlier by population structure within the Western lineage (e.g., shown in Heckel et al.[40]). A genetic replacement event in the Western-north region[38] occurred probably due to land-use changes on the continent within the last two millenia[38]—much after the separation of Orkney populations. The signal of introgression from the Western-south group to the ancestor of the Orkney voles (Figures 2B and S2) is thus likely caused by shared ancestral variation in the Western lineage that is still partially preserved in Orkney populations.

The demographic history of Orkney voles was tightly linked to human activities. Early archeological records from multiple islands[46] show that voles have spread fast across the Orkney archipelago given suitable ecological conditions and likely with the help of Neolithic humans. The signals of gene flow between islands (Figures 2B and S2) probably relate mostly to this relatively early phase of colonization. Our divergence times estimates for extant Orkney populations will thus not necessarily reflect the first waves of vole introduction, but rather the order in which island populations were split off. Human transport of voles around the Orkney system may have continued for centuries after the initial colonization and our divergence time estimates are likely to also reflect such gene flow.[70]

The absence of voles from Papa Westray in modern times demonstrates that some of the initial island populations have vanished. Human actions have affected other Orkney islands like Shapinsay[71] and Eday (M. Cockram, personal communication) where voles went extinct in the early 20th century.[46] It is unclear what the causes of these extinctions were and whether mostly ecological change or also genetically caused fitness decline ("mutational meltdown")[72] was involved. The phase of low $N_e$ on all Orkney islands around 500 to 200 years ago is compatible with negative effects of agricultural practices that reduced vole habitats by using burned kelp and hay as fertilizer.[73] The Orkney vole system provides a unique opportunity to distinguish between the scenarios of repeated ancient introductions versus continued gene flow. In the future, the concrete timing of events can be provided by combining more genomic information from modern vole populations with time series analyses of ancient DNA based on the rich vole bone material recovered during archeological excavations.[38,46,47]

## Extreme accumulation of deleterious mutations

The bottleneck related to the founding event of Orkney voles has left the shared genomic legacy of a large proportion of fixed derived sites that include many potentially highly deleterious alleles. Homozygosity of loss-of-function mutations (Figure 4) has reached an extremely high level in Orkney voles compared with populations of other mammals that have gone through much shorter periods of isolation.[27,28,57] Our individual-based simulations suggest that the fixation of many deleterious mutations in Orkney voles likely happened shortly after the bottleneck started (Figure S5C). Fixation at the founding stage of Orkney populations can also explain the sharing of most fixed deleterious mutations among successively colonized islands (Figure S4). It is noteworthy that these genomic patterns resulting from relatively few consecutive founding events are largely consistent with the patterns predicted by theoretical models of population contraction[74] and of extensive range expansions in continuous or discrete space,[20,21] and also with empirical patterns in human populations.[22] This similarity may stem from the fact that evolutionary forces at the front of expanding populations can have effects that are very similar to that of a single bottleneck.[75] The amount of drift during a spatial expansion comprising serial founder events is determined by the harmonic mean of population sizes at the expansion front.[76] Thus, a single bottleneck and a spatial expansion with the same harmonic mean of population size and the same duration will yield very similar genomic signatures as the spatial expansion. Relatively large population sizes per island may have contributed to maintaining the high frequency of derived deleterious alleles over the extended period in isolation.[77]

The relatively large effective population size of Orkney voles even during the initial bottleneck may also explain why we found no strong signal of reduction in the number of deleterious alleles in contrast to e.g., Mountain gorilla,[57] Bengal tiger,[78] or reintroduced Alpine ibex,[27] large mammals typically with population sizes of dozens to hundreds. Genetic purging in the narrow sense as a consequence of inbreeding is expected to decrease the number of deleterious alleles[79] and a few dozens of generations can lead to detectable signatures in genomes.[27] This process may be subtle in the Orkney vole system given even large effective population sizes per island compared with the mentioned big mammals (Figure 3). Indeed, the amount of genetic drift and inbreeding is determined by the harmonic mean of population sizes over time,[80] and the harmonic mean gives most weight to the lowest population sizes. Populations of voles with high growth rates can rapidly regain large sizes after

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042



introduction given suitable environments, and thus experience far less drift and fixation through bottlenecks than those of large mammals with much smaller growth rates. Furthermore, distant bottlenecks have a smaller impact on current genomic diversity, inbreeding and mutation load as compared with more recent ones (see Table 1 in Robinson et al.[81]). For these demographic reasons, we would not expect strong signals of purging in Orkney vole populations in contrast to the big mammal species that experienced prolonged and recent bottlenecks at much lower population sizes.

However, even the Eday population, which started from only 15 individuals in 1987 and whose ancestors have gone through multiple bottlenecks in the last 5,000 years, did not show a noticeable reduction of the total number of deleterious alleles (Figure 4; Table 1). Systematically collected data on its expansion are not available but the 27 $km^2$ island was already largely occupied by voles 4 years after introduction (M. Cockram, personal communication). This suggests that the current mutation load effectively leads to only minor reduction in the absolute fitness of Orkney voles and that selection has at least in part been soft in the Orkney archipelago. Deleterious variation with the strongest effects on survival or reproduction may have been purged already thousands of generations ago at the initial stages of Orkney colonization[13] as shown in our simulations (Figure S5). However, extinction of a few island populations mentioned above circumstantially suggests that hard selection from environmental or intrinsic genetic pressure may have had a large effect on certain populations.

Several potential explanations exist for the demographic success of Orkney voles despite the relatively large amount of inferred deleterious mutations. Reduced inter-specific competition could contribute to predominantly soft selection on Orkney, such that fitness is mostly affected by intra-specific competition.[82] Furthermore, selection could have been relaxed overall due to generally benign environmental conditions on Orkney (e.g., stable climatic conditions, lower predation pressure, lower pathogen burden) compared with continental Europe,[46,83] resulting in a DFE that is shifted toward more neutral variants. Selective sweeps of beneficial mutations that compensate effects of deleterious mutations might also play a role, but these are very difficult to identify after long bottlenecks.[74,75]

The realized fitness effects of the deleterious alleles in the present-day Orkney populations remain unknown. There is no information on current demographic trajectories for any of the Orkney islands, and published data on survival rates or other fitness relevant parameters in natural populations are absent. Our own field data suggest a lower reproductive rate in Orkney populations (mean number of embryos: 4.4; N = 80 litters) than in the Western lineage (mean: 5.5[84]), but Orkney voles have probably a higher early survival rate and longer life expectancy based on data from captivity.[85,86] Given the large recent effective population size, it is possible that some of the Orkney populations carry a high inbreeding load.[87] In order to improve our knowledge about the realized or masked effects of deleterious variation, it would be interesting to expose and quantify it at the individual level, for example, through crossing experiments with voles with high or low genetic load. Experimental comparisons between Orkney and continental voles could clarify the extent of differences in absolute and relative fitness between these populations. Field experiments in different environments might elucidate particularly components of hard and soft selection affecting populations and individuals differently (e.g., Barrett et al.[88]). Similar experimental approaches with Orkney voles could also shed light on how demographic events, such as prolonged periods at reduced population sizes in isolated populations, can set the stage for the occurrence of heterosis and POD after secondary contact.[60]

## The success of an ancient off-site introduction

The Orkney vole system can be regarded as an ancient off-site introduction and may thus enable a comparative glimpse into the potential future of isolated vertebrate populations of concern. In conservation biology, the need is rising to better understand the relationship between fitness and genome-wide genetic variation.[87,89] Genetic diversity is the hallmark of evolutionary potential of the populations, and reduced diversity and increased homozygosity of deleterious mutations for extended periods of time are expected to lead to increased risk of extinction of isolated populations.[90] However, many Orkney vole populations have prevailed for thousands of generations. One of the beneficial factors may be that the initial effective sizes were not extremely small compared with other systems mentioned above and may have left sufficient functionally relevant diversity to cope with the environment.

An important reason for the persistence of Orkney voles may consist in the life history of the common vole. Unlike the big mammals that have been mostly studied,[26,27,57,91] the common vole is a typical *r*-strategist with short life expectancy and high reproductivity, and extensive fluctuations as part of its normal demography of populations.[92,93] Recent work suggested that species with shorter life spans tend to bear more deleterious mutations,[17] possibly driven by insufficient selection on the genes involved in the late stage of life, or due to genetic drift from repeated bottlenecks.[94] Moreover, organisms with high reproductive rates are less likely affected by severe fitness loss due to inbreeding depression than slowly reproducing species because of more potential for variation among offspring.[95] Yet, more theoretical and empirical studies are necessary to disentangle the evolutionary connection between life history traits and mutation load. Thus, it would be important to consider for off-site introductions, e.g., of species of conservation concern, not only general life history traits but also information about historical levels of deleterious variation in genomes.[96,97] Although being of no conservation concern itself, the Orkney vole offers the opportunity to examine in more detail the relative importance of adaptive versus chance events for the persistence of isolated populations in the future.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042







## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.04.042.

## ACKNOWLEDGMENTS

We thank Julian Branscombe, John Crossley, Keith Dobney, and Sydney Gauld for logistical support, and the following landowners and institutions for access for fieldwork: C. and N. Bichan, M. and D. Cockram, S. and A. Gauld, A. and D. Thomson, S. and L. Hagan, J. and S. Crossley, B. and L. Hamill, B. Turnbull and the Westray Golf Club, E. Spence, R. and A. Manson, and the Orkney Islands council. We thank Susanne Tellenbach, Melanie Hiltbrunner, Joana Rechsteiner, and Guy Schnidrig for help with fieldwork in Orkney, and S. Braaker, M. Fischer, E. Kindler, B. Gauffre, A. Labutin, N. Bulatova, L. Yalkovskaya, M. Ratkiewicz, and C. Polat for assistance with sample collection. We thank the Next Generation Sequencing Platform of the University of Bern and Pamela Nicholson for excellent sequencing services. Alexandre Thiéry and Alexandre Gouy provided support for bioinformatics. We thank Nina Marchi, Manuel Schweizer, and David Marques for their comments on the manuscript. We also thank Donald Waller, Marty Kardos, and one anonymous reviewer for helpful comments and suggestions. Computation was performed in part on UBELIX (http://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern. X.W. received a scholarship for PhD study from the China Scholarship Council (no. 201706380049). This study was supported by Swiss National Science Foundation grant 31003A_176209 to G.H.

## AUTHOR CONTRIBUTIONS

G.H. conceptualized the study; G.H. designed research; X.W. and G.H. performed research; X.W. analyzed genomic data; S.P. performed evolutionary simulations; X.W., S.P., and G.H. wrote the paper.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 28, 2022
Revised: December 20, 2022
Accepted: April 17, 2023
Published: May 12, 2023

## REFERENCES

1. Kruess, A., and Tscharntke, T. (1994). Habitat fragmentation, species loss, and biological control. Science *264*, 1581–1584.
2. Mayr, E. (1963). Animal Species and Evolution*797* (Belknap Press of Harvard University Press).
3. Schipper, J., Chanson, J.S., Chiozza, F., Cox, N.A., Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A.S., Stuart, S.N., and Temple, H.J. (2008). The status of the world's land and marine mammals: diversity, threat, and knowledge. Science *322*, 225–230.
4. Crooks, K.R., Burdett, C.L., Theobald, D.M., King, S.R.B., Di Marco, M., Rondinini, C., and Boitani, L. (2017). Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. Proc. Natl. Acad. Sci. USA *114*, 7635–7640.
5. Pringle, R.M., Kartzinel, T.R., Palmer, T.M., Thurman, T.J., Fox-Dobbs, K., Xu, C.C.Y., Hutchinson, M.C., Coverdale, T.C., Daskin, J.H., Evangelista, D.A., et al. (2019). Predator-induced collapse of niche structure and species coexistence. Nature *570*, 58–64.
6. Luque, G.M., Bellard, C., Bertelsmeier, C., Bonnaud, E., Genovesi, P., Simberloff, D., and Courchamp, F. (2014). The 100th of the world's worst invasive alien species. Biol. Invas. *16*, 981–985.
7. Colautti, R.I., Alexander, J.M., Dlugosch, K.M., Keller, S.R., and Sultan, S.E. (2017). Invasions and extinctions through the looking glass of evolutionary ecology. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20160031.
8. Nei, M., Maruyama, T., and Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. Evolution *29*, 1–10.
9. Simons, Y.B., Turchin, M.C., Pritchard, J.K., and Sella, G. (2014). The deleterious mutation load is insensitive to recent population history. Nat. Genet. *46*, 220–224.
10. Kimura, M., Maruyama, T., and Crow, J.F. (1963). The mutation load in small populations. Genetics *48*, 1303–1312.
11. González-Martínez, S.C., Ridout, K., and Pannell, J.R. (2017). Range expansion compromises adaptive evolution in an outcrossing plant. Curr. Biol. *27*, 2544–2551.e4.
12. Perrier, A., Sánchez-Castro, D., and Willi, Y. (2020). Expressed mutational load increases toward the edge of a species' geographic range. Evolution *74*, 1711–1723.
13. Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H.E., and Van Oosterhout, C. (2022). Genetic load: genomic estimates and applications in non-model animals. Nat. Rev. Genet. *23*, 1.
14. Charlesworth, D., and Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. *18*, 237–268.
15. Lynch, M., Conery, J., and Burger, R. (1995). Mutation accumulation and the extinction of small populations. Am. Nat. *146*, 489–518.
16. Crnokrak, P., and Barrett, S.C. (2002). Perspective: purging the genetic load: a review of the experimental evidence. Evolution *56*, 2347–2358.
17. . Preprint at van der Valk, T., de Manuel, M., Marques-Bonet, T., and Guschanski, K. (2021). Estimates of genetic load suggest frequent purging of deleterious alleles in small populations bioRxiv. https://doi.org/10.1101/696831.
18. Kirkpatrick, M., and Jarne, P. (2000). The effects of a bottleneck on inbreeding depression and the genetic load. Am. Nat. *155*, 154–167.
19. Aris-Brosou, S. (2019). Direct evidence of an increasing mutational load in humans. Mol. Biol. Evol. *36*, 2823–2829.
20. Peischl, S., and Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. Mol. Ecol. *24*, 2084–2094.
21. Peischl, S., Dupanloup, I., Bosshard, L., and Excoffier, L. (2016). Genetic surfing in human populations: from genes to genomes. Curr. Opin. Genet. Dev. *41*, 53–61.
22. Peischl, S., Dupanloup, I., Foucal, A., Jomphe, M., Bruat, V., Grenier, J.C., Gouy, A., Gilbert, K.J., Gbeha, E., Bosshard, L., et al. (2018). Relaxed selection during a recent human expansion. Genetics *208*, 763–777.
23. Mallet, M.A., and Chippindale, A.K. (2011). Inbreeding reveals stronger net selection on *Drosophila melanogaster* males: implications for mutation load and the fitness of sexual females. Heredity *106*, 994–1002.
24. Jiménez, J.A., Hughes, K.A., Alaks, G., Graham, L., and Lacy, R.C. (1994). An experimental study of inbreeding depression in a natural habitat. Science *266*, 271–273.
25. Meagher, S., Penn, D.J., and Potts, W.K. (2000). Male–male competition magnifies inbreeding depression in wild house mice. Proc. Natl. Acad. Sci. USA *97*, 3324–3329.

26. 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. Sci. Adv. *5*, eaau0757.

27. Grossen, C., Guillaume, F., Keller, L.F., and Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat. Commun. *11*, 1001.

28. Robinson, J.A., Brown, C., Kim, B.Y., Lohmueller, K.E., and Wayne, R.K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Curr. Biol. *28*, 3487–3494.e4.

29. Stoffel, M.A., Johnston, S.E., Pilkington, J.G., and Pemberton, J.M. (2021). Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. Nat. Commun. *12*, 2972.

30. Huisman, J., Kruuk, L.E.B., Ellis, P.A., Clutton-Brock, T., and Pemberton, J.M. (2016). Inbreeding depression across the lifespan in a wild mammal population. Proc. Natl. Acad. Sci. USA *113*, 3585–3590.

31. Whittaker, R.J., and Fernández-Palacios, J.M. (2007). Island Biogeography: Ecology, Evolution, and Conservation (Oxford University Press).

32. Latter, B.D. (1973). The island model of population differentiation: a general solution. Genetics *73*, 147–157.

33. Richards, C. (1996). Monuments as landscape: creating the centre of the world in late Neolithic Orkney. World Archaeol. *28*, 190–208.

34. Stanton, D.W.G., Mulville, J.A., and Bruford, M.W. (2016). Colonization of the Scottish Islands via long-distance Neolithic transport of red deer (*Cervus elaphus*). Proc. Biol. Sci. *283*, 20160095.

35. Chevret, P., Hautier, L., Ganem, G., Herman, J., Agret, S., Auffray, J.C., and Renaud, S. (2021). Genetic structure in Orkney island mice: isolation promotes morphological diversification. Heredity *126*, 266–278.

36. Millais, J. (1904). On a new British vole from the Orkney Islands. Zoologist *8*, 241–246.

37. Haynes, S., Jaarola, M., and Searle, J.B. (2003). Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. Mol. Ecol. *12*, 951–956.

38. Martínková, N., Barnett, R., Cucchi, T., Struchen, R., Pascal, M., Pascal, M., Fischer, M.C., Higham, T., Brace, S., Ho, S.Y.W., et al. (2013). Divergent evolutionary processes associated with colonization of offshore islands. Mol. Ecol. *22*, 5205–5220.

39. Lambin, X., Bretagnolle, V., and Yoccoz, N.G. (2006). Vole population cycles in northern and southern Europe: is there a need for different explanations for single pattern? J. Anim. Ecol. *75*, 340–349.

40. Heckel, G., Burri, R., Fink, S., Desmet, J.-F., and Excoffier, L. (2005). Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. Evolution *59*, 2231–2242.

41. Lischer, H.E.L., Excoffier, L., and Heckel, G. (2014). Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. Mol. Biol. Evol. *31*, 817–831.

42. Beysard, M., and Heckel, G. (2014). Structure and dynamics of hybrid zones at different stages of speciation in the common vole (*Microtus arvalis*). Mol. Ecol. *23*, 673–687.

43. Yigit, N., Hutterer, R., Kryštufek, B., and Amori, G.; The IUCN red list threatened species (2016). *Microtus arvalis*. e.T13488A22351133. https://doi.org/10.2305/IUCN.UK.2016-2.RLTS.T13488A22351133.en.

44. Romaniuk, A.A., Shepherd, A.N., Clarke, D.V., Sheridan, A.J., Fraser, S., Bartosiewicz, L., and Herman, J.S. (2016). Rodents: food or pests in Neolithic Orkney. R. Soc. Open Sci. *3*, 160514.

45. Reynolds, P. (1992). The Impact of Changes in Land-Use in Orkney, on the Vole *Microtus arvalis orcadensis* and Its Avian Predators.. PhD thesis (University of Aberdeen).

46. Cucchi, T., Barnett, R., Martínková, N., Renaud, S., Renvoisé, E., Evin, A., Sheridan, A., Mainland, I., Wickham-Jones, C., Tougard, C., et al. (2014). The changing pace of insular life: 5000 years of microevolution in the Orkney vole (*Microtus arvalis orcadensis*). Evolution *68*, 2804–2820.

47. Fraser, S.M. (2015). Mammals in Late Neolithic Orkney (with reference to mammal bone recovered from Links of Noltland, Westray). PhD thesis (The University of Edinburgh).

48. Phillips, T. (2004). Seascapes and landscapes in Orkney and northern Scotland. World Archaeol. *35*, 371–384.

49. Smith, D.E., Barlow, N.L.M., Bradley, S.L., Firth, C.R., Hall, A.M., Jordan, J.T., and Long, D. (2019). Quaternary sea level change in Scotland. Earth Environ. Sci. Trans. R. Soc. Edinb. *110*, 219–256.

50. Braaker, S., and Heckel, G. (2009). Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole (*Microtus arvalis*). Mol. Ecol. *18*, 2518–2531.

51. Saxenhofer, M., Schmidt, S., Ulrich, R.G., and Heckel, G. (2019). Secondary contact between diverged host lineages entails ecological speciation in a European hantavirus. PLoS Biol. *17*, e3000142.

52. Saxenhofer, M., Labutin, A., White, T.A., and Heckel, G. (2022). Host genetic factors associated with the range limit of a European hantavirus. Mol. Ecol. *31*, 252–265.

53. Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., and Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. *2*, 1940–1955.

54. Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. *49*, 303–309.

55. Patton, A.H., Margres, M.J., Stahlke, A.R., Hendricks, S., Lewallen, K., Hamede, R.K., Ruiz-Aravena, M., Ryder, O., McCallum, H.I., Jones, M.E., et al. (2019). Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in Tasmanian devils. Mol. Biol. Evol. *36*, 2906–2921.

56. Agrawal, A.F., and Whitlock, M.C. (2011). Inferences about the distribution of dominance drawn from yeast gene knockout data. Genetics *187*, 553–566.

57. Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N., et al. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science *348*, 242–245.

58. Do, R., Balick, D., Li, H., Adzhubei, I., Sunyaev, S., and Reich, D. (2015). No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. Nat. Genet. *47*, 126–131.

59. Gilbert, K.J., Pouyet, F., Excoffier, L., and Peischl, S. (2020). Transition from background selection to associative overdominance promotes diversity in regions of low recombination. Curr. Biol. *30*, 101–107.e3.

60. Abu-Awad, D., and Waller, D. (2023). Conditions for maintaining and eroding pseudo-overdominance and its contribution to inbreeding depression. Peer Community J. *3*, e8.

61. Kimura, M. (1962). On the probability of fixation of mutant genes in a population. Genetics *47*, 713–719.

62. Slatkin, M., and Excoffier, L. (2012). Serial founder effects during range expansion: A spatial analog of genetic drift. Genetics *191*, 171–181.

63. Deshpande, O., Batzoglou, S., Feldman, M.W., and Cavalli-Sforza, L.L. (2009). A serial founder effect model for human settlement out of Africa. Proc. Biol. Sci. *276*, 291–300.

64. Funk, W.C., Lovich, R.E., Hohenlohe, P.A., Hofman, C.A., Morrison, S.A., Sillett, T.S., Ghalambor, C.K., Maldonado, J.E., Rick, T.C., Day, M.D., et al. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Mol. Ecol. *25*, 2176–2194.

65. Martin, C.H., and Höhna, S. (2018). New evidence for the recent divergence of Devil's Hole pupfish and the plausibility of elevated mutation rates in endangered taxa. Mol. Ecol. *27*, 831–838.

66. Benazzo, A., Trucchi, E., Cahill, J.A., Maisano Delser, P.M., Mona, S., Fumagalli, M., Bunnefeld, L., Cornetti, L., Ghirotto, S., and Girardi, M.

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042





(2017). Survival and divergence in a small group: the extraordinary genomic history of the endangered Apennine brown bear stragglers. Proc. Natl. Acad. Sci. USA *114*, E9589–E9597.

67. Iannella, A., Peacock, D., Cassey, P., and Schwensow, N. (2019). Genetic perspectives on the historical introduction of the European rabbit (*Oryctolagus cuniculus*) to Australia. Biol. Invas. *21*, 603–614.
68. Puckett, E.E., Magnussen, E., Khlyap, L.A., Strand, T.M., Lundkvist, Å., and Munshi-South, J. (2020). Genomic analyses reveal three independent introductions of the invasive brown rat (*Rattus norvegicus*) to the Faroe Islands. Heredity *124*, 15–27.
69. Farrell, M., Bunting, M.J., Lee, D.H.J., and Thomas, A. (2014). Neolithic settlement at the woodland's edge: palynological data and timber architecture in Orkney, Scotland. J. Archaeol. Sci. *51*, 225–236.
70. Leaché, A.D., Harris, R.B., Rannala, B., and Yang, Z. (2014). The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. *63*, 17–30.
71. Miller, G.S. (1912). Catalogue of the Mammals of Western Europe (Europe Exclusive of Russia): in the Collection of the British Museum (Trustees of the British Museum).
72. Lynch, M., Bürger, R., Butcher, D., and Gabriel, W. (1993). The mutational meltdown in asexual populations. J. Hered. *84*, 339–344.
73. Fenton, A. (1997). The Northern Isles Orkney and Shetland (Dundurn).
74. Moinet, A., Schlichta, F., Peischl, S., and Excoffier, L. (2022). Strong neutral sweeps occurring during a population contraction. Genetics *220*, iyac021.
75. Schlichta, F., Peischl, S., and Excoffier, L. (2022). The impact of genetic surfing on neutral genomic diversity. Mol. Biol. Evol. *39*, msac249.
76. Wakeley, J. (2009). Coalescent Theory (Roberts & Company).
77. Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R.D., Wayne, R.K., and Lohmueller, K.E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc. Natl. Acad. Sci. USA *113*, 152–157.
78. Khan, A., Patel, K., Shukla, H., Viswanathan, A., van der Valk, T., Borthakur, U., Nigam, P., Zachariah, A., Jhala, Y.V., Kardos, M., et al. (2021). Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. Proc. Natl. Acad. Sci. USA *118*, e2023018118.
79. García-Dorado, A. (2012). Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. Genetics *190*, 1461–1476.
80. Crow, J.F., and Kimura, M. (1970). An Introduction to Population Genetics Theory (Burgess Publishing Company).
81. Robinson, J., Kyriazis, C.C., Yuan, S.C., and Lohmueller, K.E. (2023). Deleterious variation in natural populations and implications for conservation genetics. Annu. Rev. Anim. Biosci. *11*, 93–114.
82. Agrawal, A.F., and Whitlock, M.C. (2012). Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. Annu. Rev. Ecol. Evol. Syst. *43*, 115–135.
83. Adler, G.H., and Levins, R. (1994). The island syndrome in rodent populations. Q. Rev. Biol. *69*, 473–490.
84. Spitz, F. (1974). Démographie du campagnol des champs *Microtus arvalis* en Vendée. Ann. Zool. Écologie Anim. *6*, 259–312.
85. Leslie, P.H., Tener, J.S., Vizoso, M., and Chitty, H. (1955). The longevity and fertility of the Orkney vole, *Microtus orcadensis*, as observed in the laboratory. Proc. Zool. Soc. Lond. *125*, 115–125.
86. Daketse, M.-J., and Martinet, L. (1977). Effect of temperature on the growth and fertility of the field-vole, *Microtus arvalis*, raised in different daylength and feeding conditions. Ann. Biol. anim. Bioch. Biophys. *17*, 713–721.
87. Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Hauser, S., Hedrick, P.W., Miller, J.M., Tallmon, D.A., and Funk, W.C. (2021). The crucial role of genome-wide genetic variation in conservation. Proc. Natl. Acad. Sci. USA *118*, e2104642118.
88. Barrett, R.D.H., Laurent, S., Mallarino, R., Pfeifer, S.P., Xu, C.C.Y., Foll, M., Wakamatsu, K., Duke-Cohan, J.S., Jensen, J.D., and Hoekstra, H.E. (2019). Linking a mutation to survival in wild mice. Science *363*, 499–504.
89. DeWoody, J.A., Harder, A.M., Mathur, S., and Willoughby, J.R. (2021). The long-standing significance of genetic diversity in conservation. Mol. Ecol. *30*, 4147–4154.
90. Allendorf, F.W., Hohenlohe, P.A., and Luikart, G. (2010). Genomics and the future of conservation genetics. Nat. Rev. Genet. *11*, 697–709.
91. Kleinman-Ruiz, D., Lucena-Perez, M., Villanueva, B., Fernández, J., Saveljev, A.P., Ratkiewicz, M., Schmidt, K., Galtier, N., García-Dorado, A., and Godoy, J.A. (2022). Purging of deleterious burden in the endangered Iberian lynx. Proc. Natl. Acad. Sci. USA *119*, e2110614119.
92. Boyce, C.C.K., and Boyce, J.L., III. (1988). Population biology of *Microtus arvalis*. I. Lifetime reproductive success of solitary and grouped breeding females. J. Anim. Ecol. *57*, 711–722.
93. Tkadlec, E., and Zejda, J. (1995). Precocious breeding in female common voles and its relevance to rodent fluctuations. Oikos *73*, 231–236.
94. Cui, R., Medeiros, T., Willemsen, D., Iasi, L.N.M., Collier, G.E., Graef, M., Reichard, M., and Valenzano, D.R. (2019). Relaxed selection limits lifespan by increasing mutation load. Cell *178*, 385–399.e20.
95. Hedrick, P.W., Hellsten, U., and Grattapaglia, D. (2016). Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of *Eucalyptus grandis*. New Phytol. *209*, 600–611.
96. Hohenlohe, P.A., Funk, W.C., and Rajora, O.P. (2021). Population genomics for wildlife conservation and management. Mol. Ecol. *30*, 62–82.
97. Robinson, J.A., Kyriazis, C.C., Nigenda-Morales, S.F., Beichman, A.C., Rojas-Bracho, L., Robertson, K.M., Fontaine, M.C., Wayne, R.K., Lohmueller, K.E., Taylor, B.L., et al. (2022). The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science *376*, 635–639.
98. Lunter, G., and Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. *21*, 936–939.
99. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.
100. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.
101. Danecek, P., and McCarthy, S.A. (2017). BCFtools/csq: haplotype-aware variant consequences. Bioinformatics *33*, 2037–2039.
102. Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. *61*, 539–542.
103. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). JModelTest 2: more models, new heuristics and parallel computing. Nat. Methods *9*, 772.
104. Martin, M., Patterson, M., Garg, S., Fischer, O.S., Pisanti, N., Klau, G.W., Schöenhuth, A., and Marschall, T. (2016). WhatsHap: fast and accurate read-based phasing. Preprint at bioRxiv. https://doi.org/10.1101/085050.
105. Delaneau, O., Zagury, J.-F., Robinson, M.R., Marchini, J.L., and Dermitzakis, E.T. (2019). Accurate, scalable and integrative haplotype estimation. Nat. Commun. *10*, 5436.
106. Hermann, P., Heissl, A., Tiemann-Boege, I., and Futschik, A. (2019). *LDJump*: estimating variable recombination rates from population genetic data. Mol. Ecol. Resour. *19*, 623–638.

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042



107. Zhou, Y., Browning, S.R., and Browning, B.L. (2020). A fast and simple method for detecting identity-by-descent segments in large-scale data. Am. J. Hum. Genet. *106*, 426–437.

108. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

109. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534.

110. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.

111. Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite – fast *D*-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. *21*, 584–595.

112. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly *6*, 80–92.

113. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. *47*, W191–W198.

114. Haller, B.C., and Messer, P.W. (2019). SLiM 3: forward genetic simulations beyond the Wright–Fisher model. Mol. Biol. Evol. *36*, 632–637.

115. Fink, S., Excoffier, L., and Heckel, G. (2004). Mitochondrial gene diversity in the common vole *Microtus arvalis* shaped by historical divergence and local adaptations. Mol. Ecol. *13*, 3501–3514.

116. Baca, M., Popović, D., Lemanik, A., Bañuls-Cardona, S., Conard, N.J., Cuenca-Bescós, G., Desclaux, E., Fewlass, H., Garcia, J.T., Hadravova, T., et al. (2023). Ancient DNA reveals interstadials as a driver of common vole population dynamics during the last glacial period. J. Biogeogr. *50*, 183–196.

117. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., and Roazen, D. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at bioRxiv. https://doi.org/10.1101/201178.

118. Stumpf, M.P.H., and McVean, G.A.T. (2003). Estimating recombination rates from population-genetic data. Nat. Rev. Genet. *4*, 959–968.

119. Chapman, N.H., and Thompson, E.A. (2001). Linkage disequilibrium mapping: the role of population history, size, and structure. Adv. Genet. *42*, 413–437.

120. Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.F., Thomas, M.A., Haussler, D., and Jacob, H.J. (2004). Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. *14*, 528–538.

121. Palamara, P.F., Francioli, L.C., Wilton, P.R., Genovese, G., Gusev, A., Finucane, H.K., Sankararaman, S., Genome of the Netherlands Consortium, Sunyaev, S.R., de Bakker, P.I., et al. (2015). Leveraging distant relatedness to quantify human mutation and gene-conversion rates. Am. J. Hum. Genet. *97*, 775–789.

122. Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., Pan, X., Gong, M., Zhang, Y., Guo, Y., et al. (2020). The origin of domestication genes in goats. Sci. Adv. *6*, eaaz5216.

123. Boyko, A.R., Williamson, S.H., Indap, A.R., Degenhardt, J.D., Hernandez, R.D., Lohmueller, K.E., Adams, M.D., Schmidt, S., Sninsky, J.J., and Sunyaev, S.R. (2008). Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. *4*, e1000083.

124. Kim, B.Y., Huber, C.D., and Lohmueller, K.E. (2017). Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics *206*, 345–361.

125. Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nat. Rev. Genet. *8*, 610–618.

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042





## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited data | | |
| Reference genome *Microtus arvalis* | NCBI website (https://www.ncbi.nlm.nih.gov/bioproject/737461) | PRJNA737461 |
| Mitochondrial genome *Microtus arvalis* | NCBI website (https://www.ncbi.nlm.nih.gov/nuccore/MG948434.1) | MG948434.1 |
| Gene annotation *Microtus arvalis* | A. Gouy, G.H., X.W. et al., unpublished data | N/A |
| Raw resequencing data | This study | PRJNA963218 (https://www.ncbi.nlm.nih.gov/sra/PRJNA963218) |
| Software and algorithms | | |
| Stampy 1.0.32 | Lunter and Goodson[98] | https://github.com/uwb-linux/stampy |
| GATK 4.1.6.0 | DePristo et al.[99] | https://gatk.broadinstitute.org |
| BWA 0.7.17 | Li et al.[100] | https://github.com/lh3/bwa |
| BCFtools 1.9 | Danecek and McCarthy[101] | https://samtools.github.io/bcftools |
| MrBayes 3.2.7 | Ronquist et al.[102] | https://nbisweden.github.io/MrBayes |
| JModelTest 2.7 | Darriba et al.[103] | https://github.com/ddarriba/jmodeltest2 |
| WhatsHap 1.0 | Martin et al.[104] | https://whatshap.readthedocs.io/en/latest |
| SHAPEIT4 4.1.2 | Delaneau et al.[105] | https://github.com/odelaneau/shapeit4 |
| R package LDJump | Hermann et al.[106] | https://github.com/PhHermann/LDJump |
| hap-IBD | Zhou et al.[107] | https://github.com/browning-lab/hap-ibd |
| PLINK 1.9 | Purcell et al.[108] | https://www.cog-genomics.org/plink |
| IQ-TREE 2.1.4 | Minh et al.[109] | http://www.iqtree.org |
| Treemix 1.13 | Pickrell et al.[110] | https://web.stanford.edu/group/pritchardlab/software.html |
| R package "optM" | https://cran.r-project.org/web/packages/OptM/index.html | |
| Dsuite 0.5 | Malinsky et al.[111] | https://github.com/millanek/Dsuite |
| SMC++ 1.15.2 | Terhorst et al.[54] | https://github.com/popgenmethods/smcpp |
| SnpEff 5.0 | Cingolani et al.[112] | https://pcingola.github.io/SnpEff/ |
| g:Profiler | Raudvere et al.[113] | https://biit.cs.ut.ee/gprofiler/ |
| SLiM 3.2 | Haller et al.[114] | https://www.slimframework.com/ |

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gerald Heckel (gerald.heckel@unibe.ch).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

Resequencing short-read Fastq files generated in this study have been deposited to SRA (BioProject ID: PRJNA963218). Code for data processing and analysis, and simulation used in this study have been deposited to Github (https://github.com/s-peischl/Wang_et_al_2023).

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042



## EXPERIMENTAL MODEL AND SUBJECT DETAILS

Tissue samples were obtained from individuals captured with snap traps and stored in absolute ethanol or as material stored at -20 °C. The samples were collected either for earlier studies Martínková et al.,[38] Beysard and Heckel,[42] Heckel et al.,[40] Fink et al.,[115] Baca et al.,[116] or in 2019.

## METHOD DETAILS

### Sampling and Sequencing

DNA was extracted using the phenol-chloroform method. DNA quality and concentration were checked with 1% agarose gels, Qubit fluorometer (Life Technologies) and NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific). Sequencing libraries were produced with Illumina TruSeq DNA PCR-Free Library Prep Kit and sequenced on Illumina Hiseq 2000 or Novaseq 6000 by the NGS platform of the University of Bern (2 x 150 cycles).

### Variant calling and filtering

In total, the genomes of 45 common voles covering most of the species' distribution range and three additional *Microtus obscurus* were sequenced (Figure 1A; Tables 1 and S1; see supplemental information for the methods). Raw reads were mapped to the reference genome of *M. arvalis* assembled at chromosome scale (BioProject ID: PRJNA737461, A. Gouy, G.H., X.W., et al., unpublished data) using Stampy 1.0.32[98] with default parameters except "—substitutionrate" set to 0.05. After mapping, duplicated reads were marked and removed using MarkDuplicates tool from GATK 4.1.6.0.[99] Variants were called individually with GATK HaplotypeCaller[117] using the GVCF pipeline.

After calling, SNPs and indels were filtered separately. For each individual, variants with read depth lower than 5 or higher than 150, or genotype quality (GQ) lower than 20 were marked as missing sites. For each SNP, the overall filter parameters were: QD<10.0, SOR>3.0, FS>60.0, MQ<40.0, MQRankSum<-12.5 and ReadPosRankSum<-8.0. The overall filter parameters for indels were: QD<10.0, FS>60.0, and ReadPosRankSum<-8.0. Indel data was only used in the analysis of mutation load. Given the broad distribution of samples from different evolutionary lineages with multiple levels of divergence in the common vole,[40,41] setting a filter of allele frequency would cause a bias towards less heterozygosity (for example up to 39% less when set to 0.05) in samples of *M. obscurus*, from the Italian lineage and some from the Western lineage, so variants with low allele frequency were preserved. With overall high sequencing depth and quality, the accuracy of called singletons was expected to be high. Keeping singletons was also essential for SFS analysis. Indel data was only used in the analysis of mutation load. Only variants from 22 chromosome-scale scaffolds corresponding to the autosomes were used for analyses.

### Mitochondrial genome analyses

Raw reads were mapped to the complete mitochondrial genome of *M. arvalis* (GenBank: MG948434.1) using BWA 0.7.17[100] with default parameters. Duplicated reads were marked and removed with GATK MarkDuplicates. Joint SNP calling was done with BCFtools[101] mpileup (version 1.9) and only sites with QUAL higher than 50 were kept. For each individual, the full mitochondrial sequence was converted from VCF with reference (GenBank: MG948434.1) using BCFtools consensus function, and sites with read depth lower than 3 or GQ lower than 99 were marked as missing sites. Phylogenetic reconstructions based on full mitochondrial sequences were achieved using MrBayes 3.2.7[102] with the GTR+I+G substitution model which was chosen with JModelTest 2.7.[103] Two runs were completed with 4 chains, 2,000,000 iterations and 25% burn-in to make sure the final standard deviation of split frequencies was lower than 0.01. The mitochondrial sequences of *M. obscurus* were used as outgroup for phylogenetic reconstruction.

### Recombination map and mutation rate

To estimate the mutation rate, SNP phasing and a recombination map were needed. Two steps were taken to obtain phased genome sequences. First, for each individual, SNPs were phased into small blocks based on reads covering two SNPs or more using WhatsHap 1.0.[104] Then the pre-phased data of individuals from Orkney and the Western lineage were pooled together and phased with SHAPEIT4 4.1.2[105] taking into account read-based blocks with an error rate of 0.0001. The specific default parameter set designed for sequencing data (–sequencing) was used. Though WhatsHap was originally designed for long-read sequencing, using short-read data with high depth can still effectively reduce the runtime and increase the accuracy for grouped phasing with SHAPEIT4.[105]

Phased SNP data was converted to fasta files using the BCFtools consensus function and missing sites were masked as "N". Considering that population structure could bias the estimation of linkage disequilibrium (LD),[118,119] only five individuals from the northern coastal area of the Western lineage were used for the estimation. The R package LDJump,[106] designed for small sample size, was used to produce the recombination map. Default parameters were used except for the segment length set to 2kb. The population recombination parameter $\rho$ was then converted to cM using the autosomal average recombination rate 0.63 cM/Mb for rat.[120]

The genome wide mutation rate was estimated using the TMRCA (time to the most recent common ancestor) regression method from Palamara et al.[121] which is based on the genetic differences of identical-by-descent (IBD) segments from an inbred population. In short, the average mismatch rate of IBD segments longer than a certain length linearly regresses to the posterior mean age of these segments. Five individuals from Mainland Orkney were used for mutation rate estimation. IBD segments were detected using phased

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042





SNP data and the estimated recombination map with hap-IBD[107] and default parameters. The mismatch rate of each segment was calculated by directly counting mismatch sites. IBD segments longer than 0.5 cM and only MRCA timepoints with more than 100 segments were used for regression.

### Nuclear genomic diversity and divergence

The numbers of SNPs and heterozygous sites were counted using BCFtools 1.9. PCA was done using PLINK 1.9[108] on either SNPs of all individuals or only Orkney vole individuals without any missing data. We used IQ-TREE 2.1.4[109] and Treemix 1.13[110] to obtain a nuclear phylogeny and estimate the number and extent of historical migration events. To avoid excessive run times of the Maximum Likelihood reconstruction of the phylogeny, we subsampled our dataset to 5 million random SNPs without any missing data. The three *M. obscurus* individuals were used as outgroup. Implemented substitution model test, 1,000 ultrafast bootstraps and 1,000 SH-like approximate likelihood ratio tests were performed with IQ-TREE. Independent runs of Treemix were performed with 0 to 8 migration events added to the tree with 10 iterations for each number of migration events. The optimal number of migration events was estimated with the R package "optM" using the "Evanno" method (https://cran.r-project.org/web/packages/OptM/). For the Western evolutionary lineage, three geographical groups of genomes were defined according to Treemix results.

To further examine our dataset for genetic admixture between Orkney and continental voles and between different Orkney islands, *f*-branch statistics were calculated with Dsuite[111] based on the phylogeny obtained with IQ-TREE. This method is based on the $f_4$-ratio and can assign admixture events to specific branches given a phylogenetic tree.[53] Additionally, we performed population clustering with ADMIXTURE for either all *M. arvalis* individuals in the dataset or only for Orkney individuals to help interpret the admixture and gene flow events.

### Demographic history

To investigate the demographic history and the sequence of colonization and divergence of voles on different Orkney islands, we used SMC++[54] to estimate the change of population sizes and divergence times. Population size changes were first estimated for each population from 100 to 500,000 generations ago, with 40 iterations using pchip spline, 30 spline knots, and regularization penalty at 8. We used the mutation rate estimate of $8.7\times10^{-9}$ per generation (SI) and a generation time of 0.5 years.[41] The divergence time was then estimated between Orkney Mainland voles and the Western-north group which is closest to the ancestral populations. We obtained further estimates for the split between Orkney Mainland and the other islands where Orkney voles occur, except for Eday island. Given the known introduction and release of 15 voles from Westray in Eday in 1987 by M. Cockram (personal communication), we attempted to estimate the split time between Eday and Westray.

### Mutation load

In order to estimate the frequency of deleterious mutations in Orkney voles and Western lineage populations, SnpEff 5.0[112] was used to annotate the functional effects of the variants. Gene annotation of *M. arvalis* (BioProject ID: PRJNA737461, A. Gouy, G.H., X.W., et al., unpublished data) was used to locate SNPs and indels in genes. Variants marked with the impact category "high" by SnpEff were considered deleterious mutations. This impact category includes mutations heavily affecting the function of the protein, for example mutations which eliminate start or stop codons, or frameshifting insertions and deletions. The number of variants with different functional consequences was counted per individual. Putatively neutral variants were defined as variants assumed to have no effect on any gene (category "modifier" defined by SnpEff) and at least 10 Kb away from any gene. The genomes of the three *M. obscurus* individuals were used to define derived states of the genotypes. Only those sites were kept where the allele frequencies in *M. obscurus* were zero, and for each site the alternative genotype was considered as the derived allele. Note that the derived states were inferred from a single outgroup species and thus may not be as reliable as when based on multiple species. To avoid the uncertainty of ancestral state, non-biallelic variants were filtered. Variants that were heterozygous in more than 50% of the individuals or had missing data were also filtered. The unfolded allele site frequency spectrum (SFS) of deleterious mutations and neutral mutations was then calculated for each group of the Western evolutionary lineage and the Orkney islands. Statistical enrichment and underrepresentation analyses of Gene Ontology terms were performed on g:Profiler web server[113] using genes with deleterious variants fixed in Orkney populations and the gene list of mouse as reference.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Demographic history

For each SMC++ analysis, we produced 50 bootstrap replicates for each population with the script from Zheng et al.[122] and performed a SMC++ run for each replicate with the same parameters as above. Due to the limitation of computational time, the size of each bootstrap replicate was reduced to 5 chromosomes of 100 Mb formed by 5 Mb blocks randomly chosen from the original genomes. For the reported population sizes and split times, we then estimated the median, 5$^{th}$ percentile and 95$^{th}$ percentile for each in R.

### Mutation load

To estimate the change of selection efficacy in Orkney voles compared to their continental conspecifics, we calculated the standardized rate of private deleterious allele $R'_{XY}$[58] of Orkney populations relative to the Western-north group. For each Orkney population,

Please cite this article in press as: Wang et al., Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.04.042



$R_{XY}$ of highly deleterious, missense (SnpEff impact category "moderate" in coding regions), and synonymous (SnpEff impact category "low" in coding regions) SNPs were calculated and divided by $R_{XY}$ of neutral SNPs to acquire the standardized $R'_{XY}$ to eliminate the effect of unequal distance to the outgroup. The standard error was estimated by jackknifing of 100 blocks of equal number of SNPs.

### Individual based simulations

We conducted forward-time, individual-based simulations using SLiM v3.2.[114] We modeled the Mainland population as a single panmictic population with a demographic history that mimics the one inferred using SMC++. We simplified the demographic history in a 3-epoch model with instantaneous changes in population size. We first simulated an ancestral population of 150,000 diploid individuals for a burn-in period of 450,000 generations. After 12,000 generations the population size changed to 5,000 individuals for 8,000 generations. Then the population increased to the current effective population size of 75,000 individuals for another 4,000 generations (see Figure S5C for an illustration of the demographic model). We then sampled 10 individuals from the population to estimate the expected SFS corresponding to the number of genomes from Orkney Mainland. First, we simulated 100,000 sites with a recombination rate of $10^{-6}$ between consecutive sites and the estimated mutation rate (see supplemental information) mentioned above. We simulated both neutral and deleterious SNPs. The distribution of fitness effects for deleterious mutations was modelled as a Gamma distribution with mean -0.01 and shape parameter 0.1. The dominance coefficient of deleterious mutations was set to h = 0.25. To speed up computation times all parameters were rescaled by a factor of 50 (see chapter 5.5 in the SLiM manual). We chose the parameters of the DFE to span the range of plausible values for selection coefficients of deleterious mutations estimated in natural populations.[123–125] Second, to test the robustness of our results and to further investigate the role of the shape of the DFE, we performed an additional set of simulations with a refined DFE combining different dominance coefficients, a lower recombination rate of $10^{-8}$ between consecutive sites, and smaller scaling factors to allow for broader DFEs including more large-effect mutations including sub-lethal and lethal mutations (Figure S5). We further examined the simulation results for evidence of pseudo-overdominance (POD) that could potentially reduce the expression of fitness effects, including low recombination rate regions with an excess of deleterious mutations or an excess of intermediate frequency deleterious variants in the SFS.

## Supplementary Information

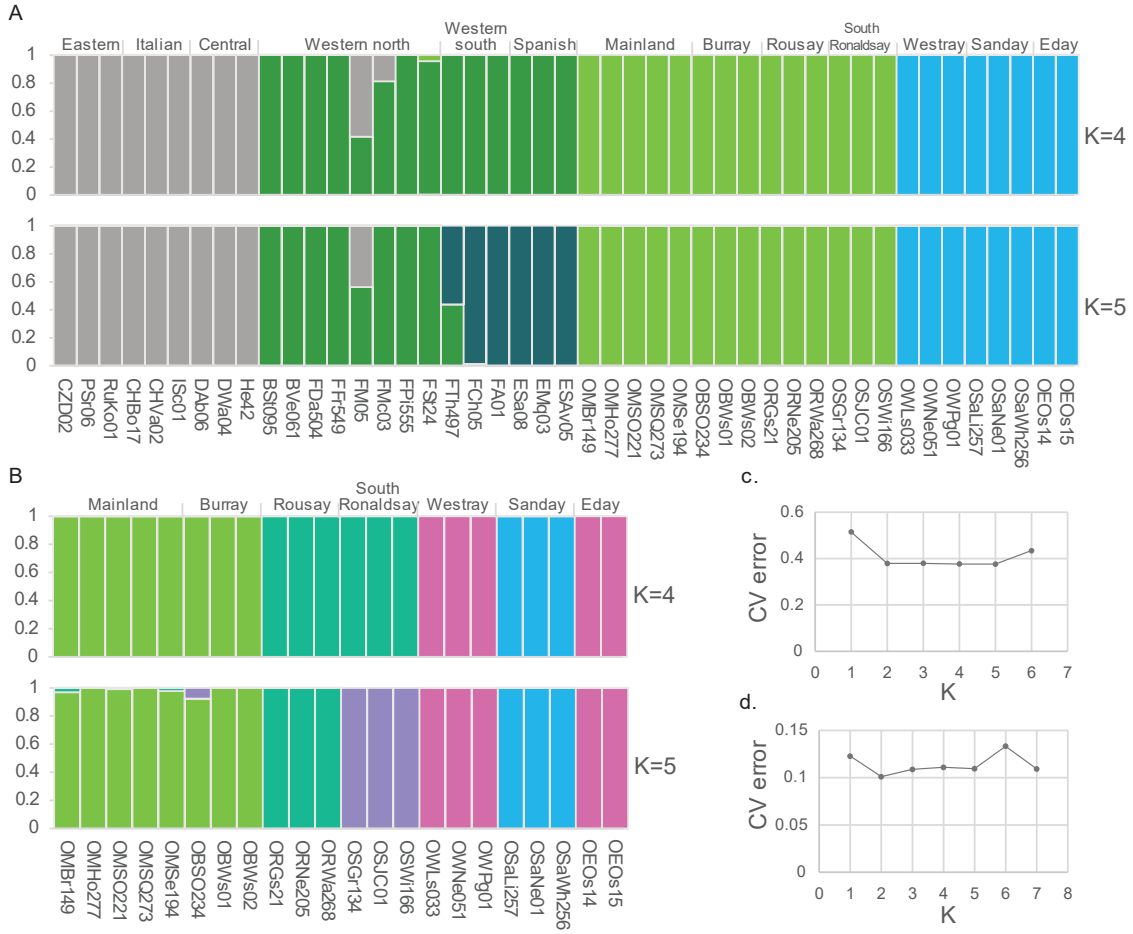



**Figure S1. ADMIXTURE analyses of *M. arvalis* individuals; related to Figure 2B**

**A.** Admixture plot with K=4 and K=5 showing the separation of southern and northern Orkney populations from continental voles. No sign of gene flow from the continent to Orkney was found while some continental individuals showed evidence for admixture, for example FM05. **B.** ADMIXTURE plots of Orkney individuals (K=4 and K=5) which separated voles on most islands from each other analogous to Figure 1. The populations of Mainland and Burray had close ancestry, as well as the Eday and Westray populations (Eday voles were introduced from Westray; see text). **C.** CV errors of ADMIXTURE of all *M. arvalis* individuals of K from 1 to 6. The lowest CV error was 0.377 when K=4. CV errors of ADMIXTURE of Orkney individuals of K from 1 to 7. The lowest CV error was 0.101 when K=2 which reflected the separation of south and north Orkney island populations as in **A**.

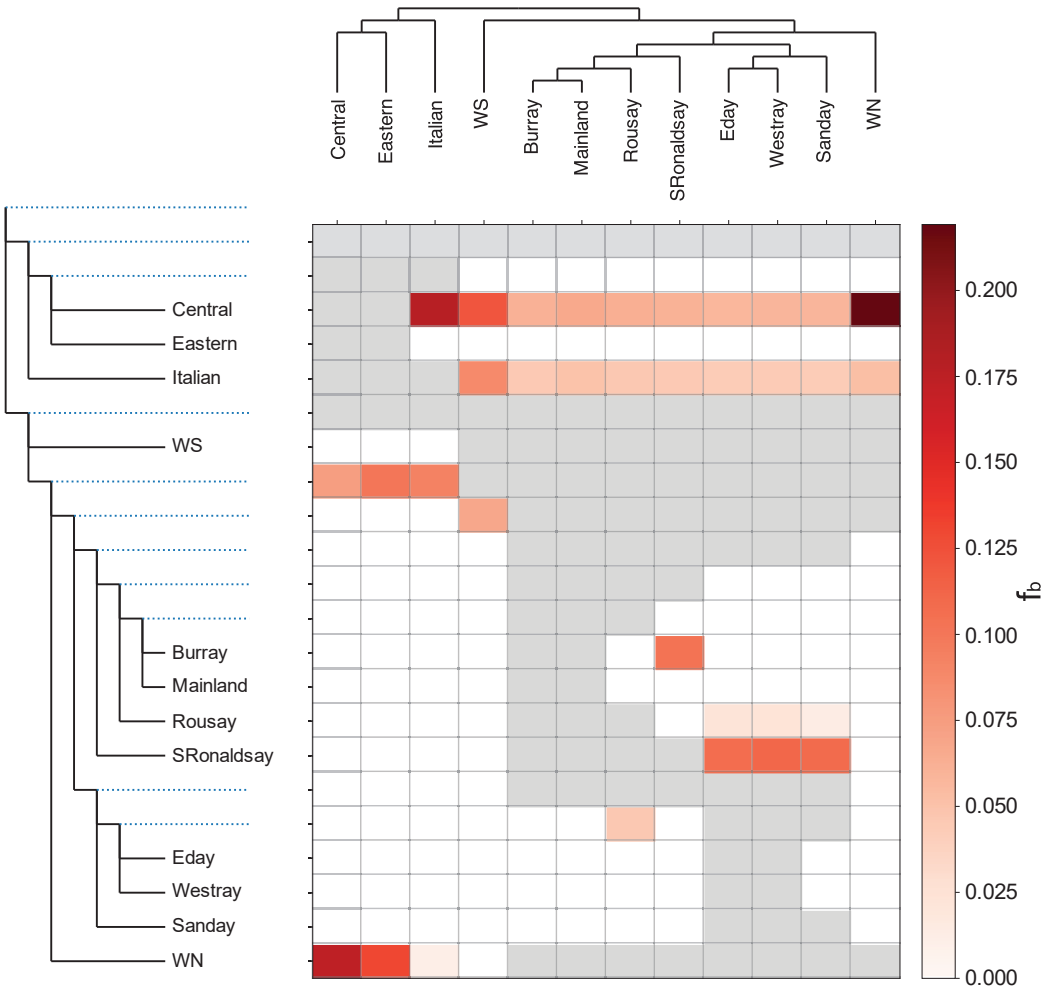



**Figure S2. Excess allele sharing indicated by *f*-branch statistics; related to Figure 2B**

The admixed individual FM05 (see Results) was not included in this analysis. In the matrix, the extended tree is on the y axis, with dashed lines representing internal branches, and the P3 population is on the x axis. Grey blocks refer to the pairs that cannot be tested. Spanish samples were merged into the Western south (WS) population. The significant *f*b values ($p$<0.01) between Orkney populations and Central lineage likely represent shared ancestry with Western north (WN), rather than real admixture. Similarly, the *f*b values between Westray, Sanday and Eday to South Ronaldsay (SRonaldsay) indicate probably admixture between ancestral South Ronaldsay and the common ancestor of these three northern islands.

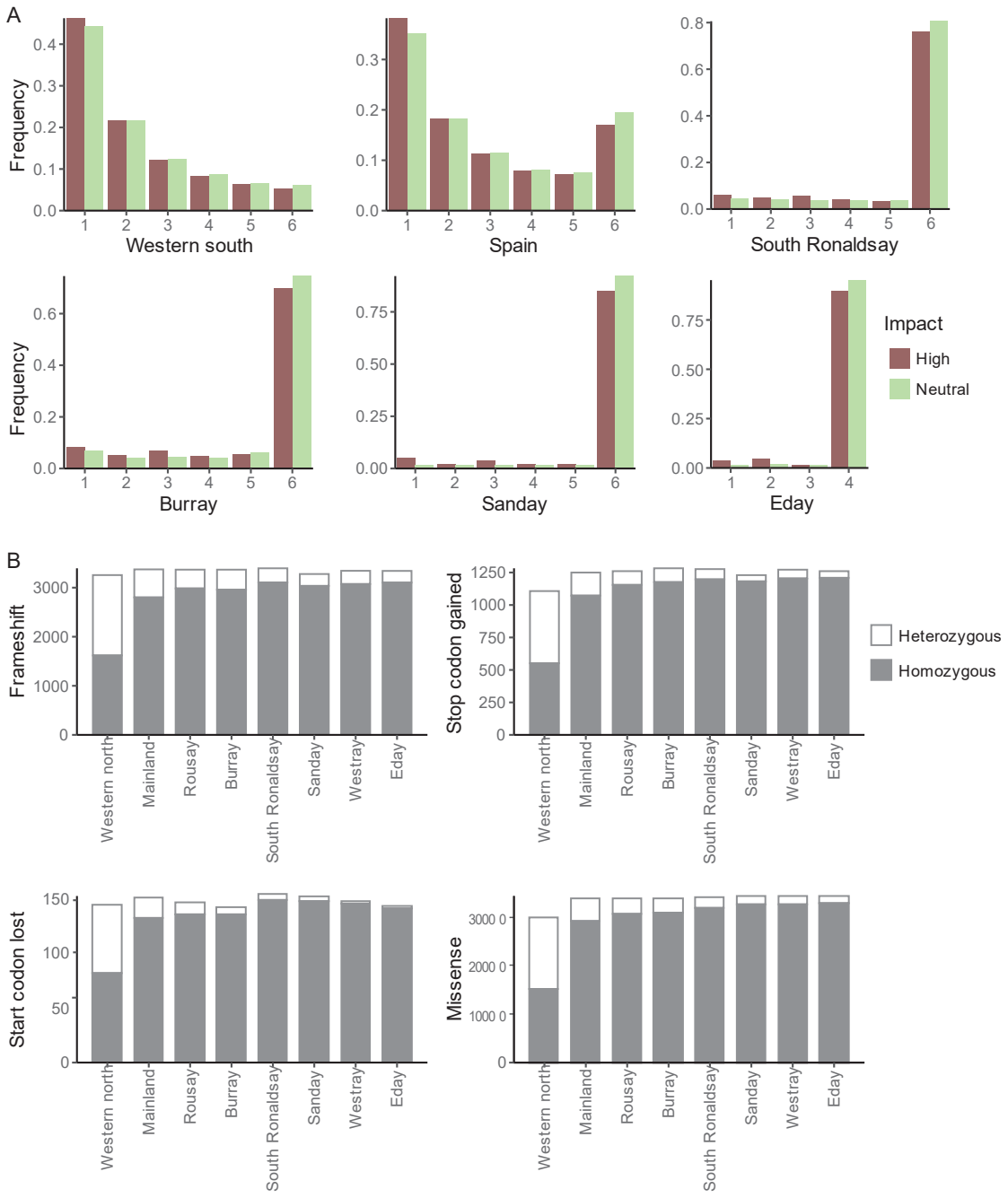



**Figure S3. Extended data on the deleterious mutations; related to Figure 4**

**A.** Unfolded site frequency spectra of derived high-impact (dark red) and neutral sites (green) including nucleotide polymorphism sites and indels for Orkney vole populations and groups of genomes in the Western evolutionary lineage that are not shown in Figure 4. **B.** The average number of alleles with different functional consequences per individual for Orkney populations and continental voles from the Western north group.

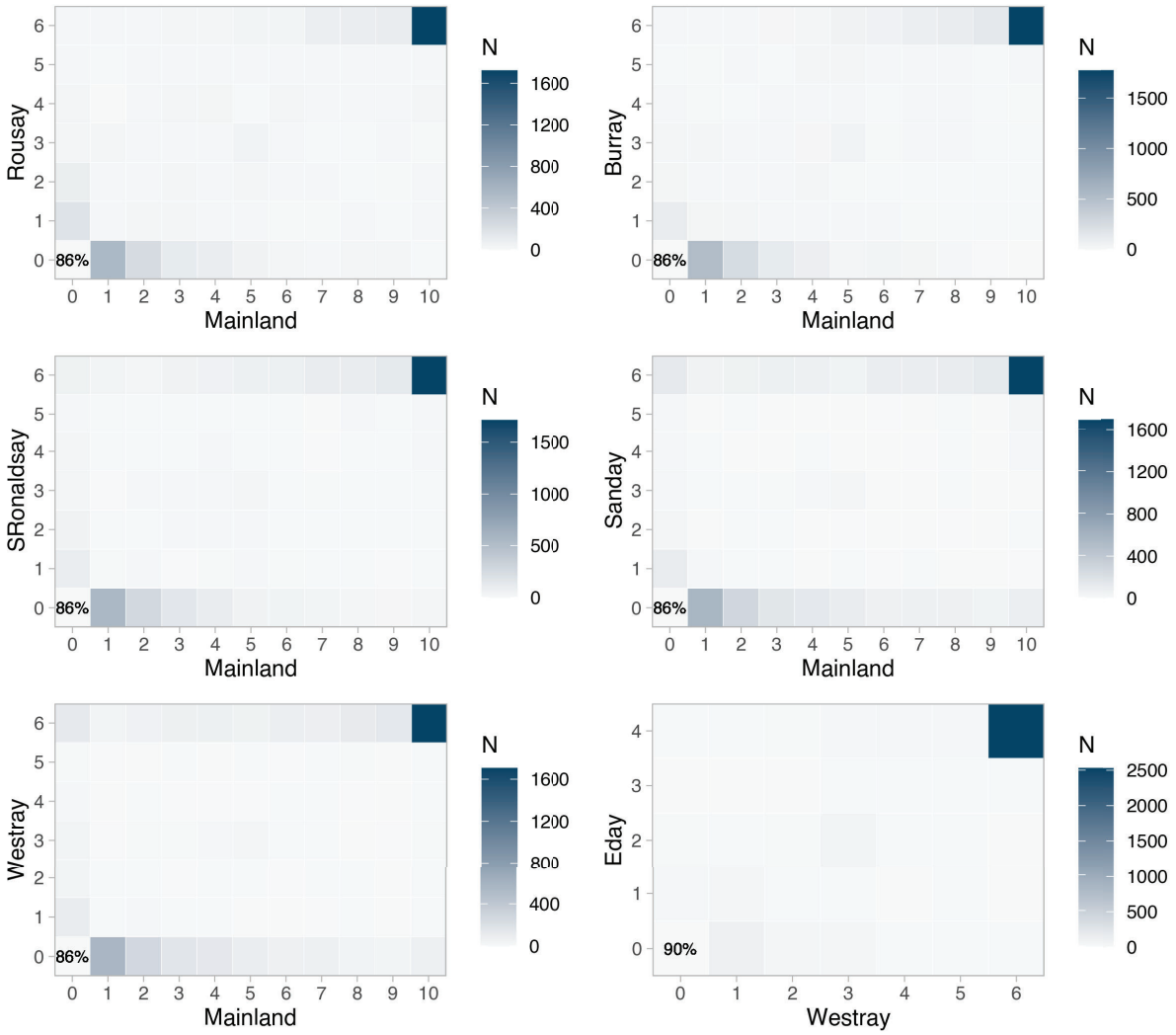



**Figure S4. two-dimensional site allele frequency spectra (2D-SFS) of deleterious sites of Orkney vole populations; related to Figure 4A**

The number shown in the bottom-left is the percentage of sites without any deleterious alleles in the pair of populations.

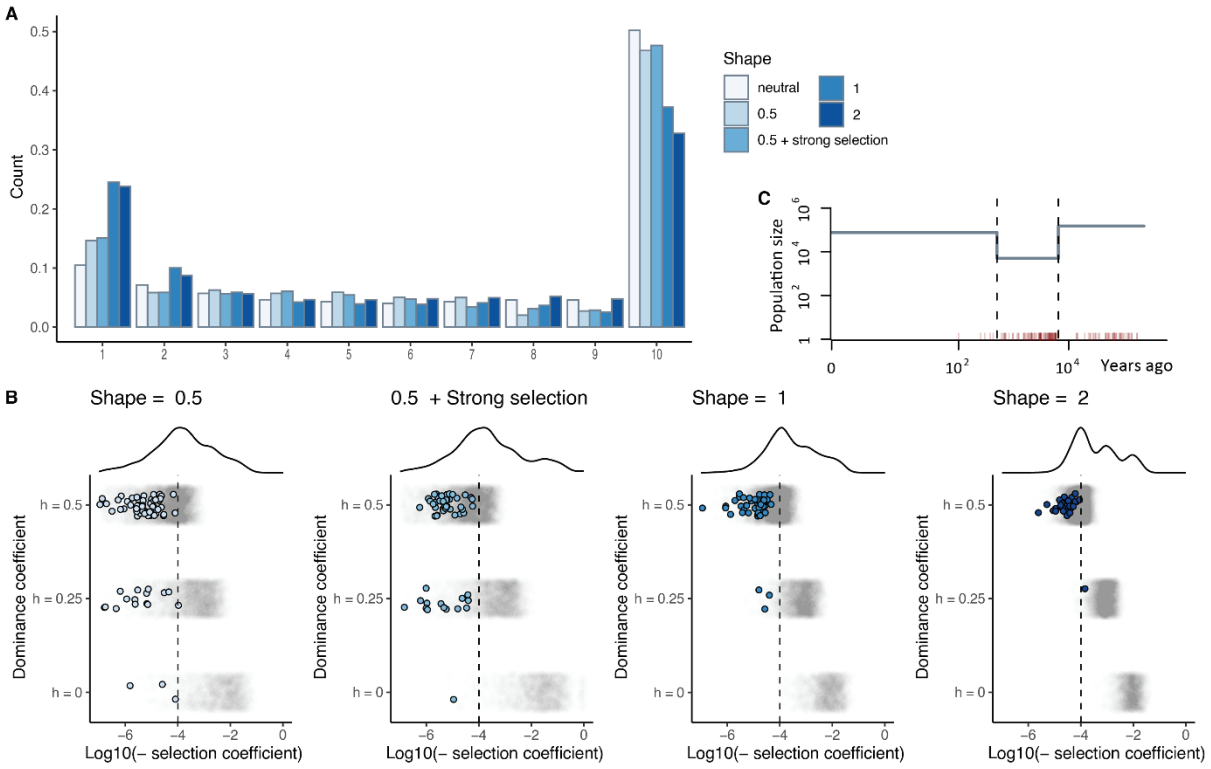



**Figure S5. Extended simulations with varied Distributions of Fitness Effects; related to Figure 4A**

Site Frequency Spectra (SFS) obtained from individual based simulations of the demographic history of Orkney voles. Distributions of Fitness Effects (DFE) were generated as mixtures of three different gamma distributions with mean s = -0.01 (fully recessive, h = 0), mean s = -0.001 (partially recessive, h = 0.25), and mean s = -0.0001 (co-dominant, h = 0.5). The proportions of these three classes of mutations were 0.2, 0.3, and 0.5, respectively. We varied the shape parameter of the gamma distributions to consider narrower and wider DFEs. **A**. the SFS for a sample of size 10, averaged over 100 replicates, for three DFEs as well as neutral simulations. We included additionally a case with strong selection (only shown for the DFE with shape = 0.5), where the mean effect of fully recessive mutations was s = -0.05 instead of s = -0.01, and 3% of mutations were recessive lethal or sub-lethal (s = -1 or s = -0.75). **B**. examples of the DFEs. The x-axis shows the log10 of the absolute value of the selection coefficient of deleterious mutations. The semi-transparent gray dots represent a sample of 1000 mutations from each of the DFEs and the blue dots show fixed mutations found in a single simulation replicate. The density estimate on top of the panels displays the DFE irrespective of the dominance coefficients. The vertical dashed line shows the critical s value such that *Ns = 1* during the bottleneck. Each simulation consisted of 10 kb and the recombination rate was set to $10^{-8}$ between consecutive sites. To speed up the run time of these simulations (except for the strong selection scenario), we rescaled all parameters by a factor of 10 (see SLiM manual chapter 5.5). **C.** Model of the demographic history used in individual-based simulations of the vole population on Orkney Mainland. The simplified 3-epoch model mimicking the inferred demographic history is shown with the solid line while the dashed lines represent the start and end points of the bottleneck with instantaneous changes in population size. The ancestral population was 150,000 diploid individuals. The population size changed after 12,000 generations (0.5 years) to 5,000 individuals for 8,000 generations. Then the population increased to the current effective population size of 75,000 individuals for another 4,000 generations. The red lines show the timepoints of fixation of deleterious alleles in the population.

| Individual | Lineage | Subgroup | Latitude | Longitude | Depth | %SNP-missing | %indel-missing | Heterozygous sites |
|---|---|---|---|---|---|---|---|---|
| **BSt095** | Western | Western north | 51.2232 | 3.07275 | 22.7 | 2.80% | 2.64% | 8556802 |
| **BVe061** | Western | Western north | 51.07315 | 2.66803 | 40.8 | 1.39% | 1.55% | 8643462 |
| **FDa504** | Western | Western north | 49.76666 | 0.51666 | 23.8 | 2.13% | 2.29% | 8849443 |
| **FFr549** | Western | Western north | 50.08333 | 1.56666 | 23.3 | 2.40% | 2.38% | 8471666 |
| **FMc03** | Western | Western north | 49.25437 | 3.93216 | 19.0 | 3.47% | 3.16% | 9173598 |
| **FPi555** | Western | Western north | 50.86666 | 1.81666 | 22.6 | 2.35% | 2.56% | 8422519 |
| **FSt24** | Western | Western north | 46.38548 | 4.22641 | 29.7 | 1.76% | 2.15% | 8586945 |
| **FA01** | Western | Western south | 43.30067 | 1.17211 | 28.9 | 1.97% | 2.46% | 7829303 |
| **FCh05** | Western | Western south | 46.11654 | -0.35008 | 18.4 | 4.65% | 3.66% | 8262807 |
| **FTh497** | Western | Western south | 49.26508 | -0.41399 | 32.9 | 1.55% | 1.85% | 9853799 |
| **EMq03** | Western | Spain | 41.05869 | -4.16692 | 26.0 | 2.30% | 2.83% | 6775268 |
| **ESa08** | Western | Spain | 40.95909 | -5.66137 | 21.3 | 3.39% | 3.40% | 6880383 |
| **ESAv05** | Western | Spain | 40.59966 | -4.70519 | 20.3 | 3.23% | 3.58% | 6597593 |
| **FM05** | Western | *admixed | 48.17736 | 6.02823 | 22.5 | 2.36% | 2.52% | 10119600 |
| **OBSO234** | Orkney | Burray | 58.85002 | -2.91659 | 23.4 | 2.57% | 2.92% | 1855239 |
| **OBWs01** | Orkney | Burray | 58.84349 | -2.92363 | 33.3 | 1.98% | 2.59% | 1255967 |
| **OBWs02** | Orkney | Burray | 58.84349 | -2.92363 | 22.7 | 2.74% | 3.32% | 1348561 |
| **OEOs14** | Orkney | Eday | 59.17797 | -2.75946 | 26.0 | 2.48% | 3.23% | 679365 |
| **OEOs15** | Orkney | Eday | 59.17797 | -2.75946 | 27.2 | 2.33% | 3.02% | 669188 |
| **OMBr149** | Orkney | Mainland | 59.08242 | -3.29908 | 22.4 | 3.26% | 3.21% | 2599235 |
| **OMHo277** | Orkney | Mainland | 58.93692 | -3.06632 | 25.8 | 2.22% | 2.61% | 2607042 |
| **OMSe194** | Orkney | Mainland | 59.05154 | -3.11165 | 48.0 | 2.33% | 2.72% | 2684887 |
| **OMSO221** | Orkney | Mainland | 58.95 | -2.95 | 25.0 | 2.64% | 2.93% | 2214464 |
| **OMSQ273** | Orkney | Mainland | 58.9796 | -3.27287 | 22.5 | 1.48% | 1.79% | 2360249 |
| **ORGs21** | Orkney | Rousay | 59.13011 | -2.98977 | 34.3 | 1.90% | 2.52% | 1213829 |
| **ORNe205** | Orkney | Rousay | 59.13333 | -3.04999 | 27.9 | 2.07% | 2.59% | 1626903 |
| **ORWa268** | Orkney | Rousay | 59.1782 | -3.0628 | 30.5 | 1.89% | 2.38% | 1818755 |
| **OSaLi257** | Orkney | Sanday | 59.2502 | -2.58204 | 30.3 | 2.37% | 2.55% | 789992 |
| **OSaNe01** | Orkney | Sanday | 59.27294 | -2.48481 | 28.4 | 2.29% | 2.78% | 650409 |
| **OSaWh256** | Orkney | Sanday | 59.30079 | -2.55056 | 16.0 | 9.23% | 5.75% | 685883 |
| **OSGr134** | Orkney | South Ronaldsay | 58.81695 | -2.91688 | 30.4 | 2.09% | 2.48% | 1099542 |
| **OSJC01** | Orkney | South Ronaldsay | 58.75641 | -2.93866 | 30.3 | 2.08% | 2.73% | 1058722 |
| **OSWi166** | Orkney | South Ronaldsay | 58.76666 | -2.93333 | 31.6 | 2.00% | 2.41% | 984461 |
| **OWLs033** | Orkney | Westray | 59.2956 | -2.91336 | 18.8 | 5.04% | 4.24% | 856422 |
| **OWNe051** | Orkney | Westray | 59.24703 | -2.873534 | 14.2 | 13.73% | 7.46% | 698533 |
| **OWPg01** | Orkney | Westray | 59.32317 | -2.9977 | 28.5 | 2.20% | 2.92% | 810811 |
| **CHBo17** | Italian | / | 46.80605 | 9.39923 | 22.1 | 4.86% | 4.36% | 9684498 |
| **CHVa02** | Italian | / | 46.05165 | 8.99273 | 29.4 | 1.97% | 2.43% | 7702578 |
| **ISc01** | Italian | / | 46.62605 | 10.75749 | 25.4 | 2.00% | 2.31% | 9700295 |
| **CZD02** | Eastern | / | 48.87907 | 16.44188 | 27.8 | 2.14% | 3.09% | 12595886 |
| **PSr06** | Eastern | / | 53.54811 | 22.75329 | 23.0 | 2.12% | 2.26% | 10435148 |
| **RuKo01** | Eastern | / | 56.33333 | 41.41666 | 35.2 | 1.44% | 1.69% | 11294029 |
| **DAb06** | Central | / | 51.78452 | 7.63334 | 19.5 | 3.18% | 2.71% | 7583091 |
| **DWa04** | Central | / | 54.03135 | 11.70194 | 24.4 | 2.25% | 2.39% | 7750945 |
| **He42** | Central | / | 49.34092 | 10.81468 | 37.4 | 1.12% | 1.15% | 10931971 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **RuCh01** | *M. obscurus* | / | 61.33066 | 43.52866 | 16.8 | 6.11% | 5.43% | 5379010 |
| **RuSu03** | *M. obscurus* | / | 55.5 | 60.55999 | 21.9 | 3.27% | 3.89% | 7870579 |
| **TuEr01** | *M. obscurus* | / | 40.30937 | 41.94739 | 18.2 | 6.75% | 5.27% | 10886235 |

**Table S1. Individual data for *Microtus* vole genomes analyzed in this study, related to Figure 1A**
Group information, sampling site, sequencing depth, the percentages of missing sites of SNPs (%SNP-missing) and indels (%indel-missing) and the number of heterozygous SNP sites (#Het) are given for each individual.

| Individual | Lineage | Subgroup | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|---|---|
| **BSt095** | Western | Western north | -0.00336063 | 0.109603 | -0.00303446 | 0.0468423 | -0.18777 |
| **BVe061** | Western | Western north | -0.00324143 | 0.110266 | -0.00297539 | 0.0467155 | -0.188467 |
| **FDa504** | Western | Western north | 6.36E-05 | 0.119078 | -0.00316085 | 0.0450569 | -0.189521 |
| **FFr549** | Western | Western north | -0.00185932 | 0.114894 | -0.00337375 | 0.0524377 | -0.195195 |
| **FMc03** | Western | Western north | 0.00618018 | 0.128589 | -0.00439819 | 0.0732666 | -0.17779 |
| **FPi555** | Western | Western north | -0.00250651 | 0.113285 | -0.00328137 | 0.0507322 | -0.193972 |
| **FSt24** | Western | Western north | -0.00279577 | 0.105227 | -0.00243398 | 0.0387778 | -0.160881 |
| **FA01** | Western | Western south | 0.00972864 | 0.129245 | 0.00779879 | -0.21807 | -0.0314594 |
| **FCh05** | Western | Western south | 0.00774023 | 0.12573 | 0.00656136 | -0.189312 | -0.0507309 |
| **FTh497** | Western | Western south | 0.00435178 | 0.123384 | 0.00282178 | -0.0998147 | -0.108374 |
| **EMq03** | Western | Spain | 0.0214537 | 0.163874 | 0.015462 | -0.46858 | 0.0495989 |
| **ESa08** | Western | Spain | 0.0214917 | 0.163512 | 0.0152934 | -0.46652 | 0.0496905 |
| **ESAv05** | Western | Spain | 0.0213519 | 0.163342 | 0.0154187 | -0.467534 | 0.0495426 |
| **FM05** | Western | *admixed | 0.020451 | 0.15148 | -0.00605415 | 0.109621 | -0.137883 |
| **OBSO234** | Orkney | Burray | -0.0941354 | -0.122073 | 0.000465358 | 0.00564962 | 0.0136569 |
| **OBWs01** | Orkney | Burray | -0.0940807 | -0.121948 | 0.000451477 | 0.00557276 | 0.013744 |
| **OBWs02** | Orkney | Burray | -0.0940345 | -0.121892 | 0.00043578 | 0.00563678 | 0.0136653 |
| **OEOs14** | Orkney | Eday | -0.0966906 | -0.127437 | 0.0004861 | 0.0057878 | 0.0155439 |
| **OEOs15** | Orkney | Eday | -0.0967129 | -0.127495 | 0.000480173 | 0.0058356 | 0.0154982 |
| **OMBr149** | Orkney | Mainland | -0.0923516 | -0.118377 | 0.000457288 | 0.00513878 | 0.012899 |
| **OMHo277** | Orkney | Mainland | -0.0924707 | -0.118579 | 0.000479395 | 0.00525201 | 0.0128278 |
| **OMSe194** | Orkney | Mainland | -0.0926512 | -0.119124 | 0.000435271 | 0.00551318 | 0.0132438 |
| **OMSO221** | Orkney | Mainland | -0.0924392 | -0.118555 | 0.000462864 | 0.00508199 | 0.0128579 |
| **OMSQ273** | Orkney | Mainland | -0.0923912 | -0.118483 | 0.000445557 | 0.00543793 | 0.0129853 |
| **ORGs21** | Orkney | Rousay | -0.0929521 | -0.119667 | 0.000472001 | 0.00553429 | 0.013354 |
| **ORNe205** | Orkney | Rousay | -0.0927273 | -0.119239 | 0.000465245 | 0.00550219 | 0.0132363 |
| **ORWa268** | Orkney | Rousay | -0.0929333 | -0.119607 | 0.000460548 | 0.00542785 | 0.013246 |
| **OSaLi257** | Orkney | Sanday | -0.0944869 | -0.122876 | 0.000472978 | 0.00559587 | 0.0140949 |
| **OSaNe01** | Orkney | Sanday | -0.0944749 | -0.122792 | 0.000499214 | 0.00545119 | 0.014118 |
| **OSaWh256** | Orkney | Sanday | -0.0944579 | -0.122758 | 0.000491238 | 0.00546692 | 0.0140875 |
| **OSGr134** | Orkney | South Ronaldsay | -0.0965998 | -0.127273 | 0.000470277 | 0.00591264 | 0.0155011 |
| **OSJC01** | Orkney | South Ronaldsay | -0.0966327 | -0.127307 | 0.000478878 | 0.00584066 | 0.0154462 |
| **OSWi166** | Orkney | South Ronaldsay | -0.096572 | -0.127192 | 0.000470966 | 0.00585167 | 0.0154598 |
| **OWLs033** | Orkney | Westray | 0.0495132 | 0.178327 | -0.0134125 | 0.169452 | 0.00163409 |
| **OWNe051** | Orkney | Westray | 0.51914 | -0.239088 | -0.387861 | -0.0362379 | -0.0113412 |
| **OWPg01** | Orkney | Westray | 0.110683 | 0.130391 | -0.0686751 | 0.157916 | 0.00628911 |
| **CHBo17** | Italian | / | 0.0401214 | 0.173819 | -0.00279671 | 0.134385 | 0.311544 |
| **CHVa02** | Italian | / | 0.0410811 | 0.16531 | 0.00162428 | 0.105061 | 0.56778 |
| **ISc01** | Italian | / | 0.0420177 | 0.165021 | 0.0012826 | 0.104576 | 0.529707 |
| **CZD02** | Eastern | / | 0.0460093 | 0.181036 | -0.00958728 | 0.168201 | -0.0131323 |
| **PSr06** | Eastern | / | 0.519759 | -0.238385 | -0.375777 | -0.0351368 | -0.0107132 |
| **RuKo01** | Eastern | / | -0.0955013 | -0.124986 | 0.000478164 | 0.00583341 | 0.0147834 |
| **DAb06** | Central | / | 0.0409011 | 0.186653 | -0.00938127 | 0.177285 | -0.0701322 |
| **DWa04** | Central | / | 0.0415438 | 0.188188 | -0.00943071 | 0.181078 | -0.0664695 |
| **He42** | Central | / | 0.0414022 | 0.18649 | -0.00889122 | 0.174006 | -0.0612767 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **RuCh01** | *M. obscurus* | / | 0.48507 | -0.201652 | 0.83795 | 0.0227496 | -0.0205761 |
| **RuSu03** | *M. obscurus* | / | -0.0955206 | -0.125033 | 0.000477938 | 0.00587576 | 0.0148683 |
| **TuEr01** | *M. obscurus* | / | -0.0954747 | -0.124926 | 0.000475571 | 0.00584776 | 0.0147811 |

**Table S2. Grouping information for *Microtus* genomes and values of PC1 to PC5 from PCA using all individuals; related to Figure 1C**

| Individual | Island | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|---|
| **OBSO234** | Burray | 0.201523 | -0.17043 | -0.170216 | 0.00858302 | -0.335765 |
| **OBWs01** | Burray | 0.220612 | -0.175631 | -0.254436 | -0.013317 | -0.40665 |
| **OBWs02** | Burray | 0.218619 | -0.175969 | -0.248818 | -0.00674784 | -0.394307 |
| **OEOs14** | Eday | -0.302199 | 0.0239649 | -0.10107 | 0.238947 | -0.00013107 |
| **OEOs15** | Eday | -0.302009 | 0.0223684 | -0.101594 | 0.236685 | -0.00229697 |
| **OMBr149** | Mainland | 0.177204 | -0.00333955 | -0.147216 | -0.0410357 | 0.393832 |
| **OMHo277** | Mainland | 0.175421 | -0.0331951 | -0.167016 | -0.0251309 | 0.226731 |
| **OMSe194** | Mainland | 0.17803 | -0.0124248 | -0.157712 | -0.031345 | 0.358753 |
| **OMSO221** | Mainland | 0.178725 | -0.0425566 | -0.173797 | -0.0198431 | 0.0856802 |
| **OMSQ273** | Mainland | 0.178869 | -0.0134086 | -0.154171 | -0.0412999 | 0.424471 |
| **ORGs21** | Rousay | 0.133658 | 0.500469 | 0.159909 | 9.84E-05 | -0.121682 |
| **ORNe205** | Rousay | 0.137619 | 0.493026 | 0.149986 | -0.00668511 | -0.10424 |
| **ORWa268** | Rousay | 0.131155 | 0.454407 | 0.148906 | -0.0069068 | -0.113274 |
| **OSaLi257** | Sanday | -0.242413 | -0.0580044 | 0.0675154 | -0.470708 | -0.0257492 |
| **OSaNe01** | Sanday | -0.244413 | -0.0584164 | 0.0677365 | -0.470792 | -0.0266314 |
| **OSaWh256** | Sanday | -0.24302 | -0.0577259 | 0.0657356 | -0.466384 | -0.02494 |
| **OSGr134** | South Ronaldsay | 0.107702 | -0.257734 | 0.382405 | 0.132375 | -0.0132262 |
| **OSJC01** | South Ronaldsay | 0.0880207 | -0.247781 | 0.463516 | 0.14541 | 0.039019 |
| **OSWi166** | South Ronaldsay | 0.0893187 | -0.25556 | 0.469587 | 0.143862 | 0.0385544 |
| **OWLs033** | Westray | -0.294943 | 0.0235996 | -0.0994296 | 0.23433 | 0.00108739 |
| **OWNe051** | Westray | -0.297686 | 0.021723 | -0.100262 | 0.230885 | 0.00111762 |
| **OWPg01** | Westray | -0.296796 | 0.0225521 | -0.100941 | 0.229764 | -0.000102658 |

**Table S3. Grouping information for *Microtus* genomes and the values of PC1 to PC5 from PCA with Orkney individuals only; related to Figure 1D**

| | Split time in years | | | | Recent $N_e$ | | | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | Median | $P_5$ | $P_{95}$ | Estimation | Median | $P_5$ | $P_{95}$ |
| **Mainland Orkney** | 8447 | 10665 | 8089 | 16326 | 74209 | 49297 | 3105 | 87771 |
| **Rousay** | 2185 | 2239 | 2143 | 2360 | 1023 | 695 | 476 | 1244 |
| **Burray** | 451 | 393 | 326 | 536 | 8741 | 856 | 332 | 146142 |
| **South Ronaldsay** | 2094 | 2054 | 510 | 2231 | 1466 | 1146 | 489 | 2394 |
| **Sanday** | 2772 | 2822 | 2591 | 2862 | 4690 | 2611 | 332 | 6040 |
| **Westray** | 2655 | 2656 | 2399 | 2661 | 2495 | 1825 | 370 | 18089 |
| **Eday** | 3 | 2 | 1 | 14 | 1325 | 761 | 232 | 2254 |

**Table S4. Estimates of population history of Orkney populations; related to Figure 3B**

Time of split of Orkney vole populations from ancestral population and average $N_e$ 100 years ago (200 generations) estimated by SMC++. The median, 5$^{th}$ percentile ($P_5$) and 95$^{th}$ percentile ($P_{95}$) are reported for bootstrapping repeats. For the split time, Mainland Orkney was analyzed relative to the ancestral Western north group on the continent. Eday was analyzed compared to the ancestral population on Westray. Other islands were analyzed relative to Orkney Mainland.

Chapter 2

# Islands of adaptation in an ocean of genetic drift: genome-wide relaxation of selection in an insular rodent

Xuejing Wang[1], Gerald Heckel[1,2]

[1]Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland; [2]Swiss Institute of Bioinformatics, Quartier Sorge - Batiment Amphipole, 1015 Lausanne, Switzerland

*Gerald Heckel

Email: gerald.heckel@unibe.ch

## Abstract

Island populations experience often different ecological and demographic conditions than their counterparts on the continent, resulting in divergent evolutionary forces affecting their genomes. Random genetic drift and selection may both leave their imprints although the relative strength depends strongly on the specific conditions. Here we address their roles in contributing to the island syndrome in a rodent with unusually clear history of isolated populations. Common voles (*Microtus arvalis*) were introduced by humans on the Orkney archipelago north of Scotland more than 5000 years ago and rapidly evolved to exceptional big size. Our analyses show that the genomes of Orkney voles were dominated by genetic drift, with strikingly low diversity, extremely variable Tajima's D and very high divergence from continental conspecifics. Increased dN/dS ratios over a wide range of genes in Orkney voles indicated genome-wide relaxation of purifying selection. We found evidence of hard sweeps on key genes of the lipid metabolism pathway only in continental voles. The marked increase of body size in Orkney - a typical phenomenon of the island syndrome - may thus be associated to the relaxation of positive selection on genes related to lipid metabolism. However, a hard sweep on immune genes of Orkney voles may reflect the divergent ecological conditions and the history of human introduction. The isolated Orkney voles show that adaptive changes may still impact the evolutionary trajectories of isolated populations despite the pervasive consequences of genetic drift at the genome level.

**Keywords**: genomic landscape; genetic drift; relaxed selection; island syndrome

## Introduction

Islands are often considered to be ideal systems to observe adaptation in nature (Foster 1964; Losos and Ricklefs 2009; Grant and Grant 2014). A population newly arrived on an island faces a novel environment with at least some abiotic and biotic differences compared to its origin (Whittaker and Fernández-Palacios 2007). The shift of environment can lead to a drastic shift in the selective regime (Losos and Ricklefs 2009), which can sculpt island populations drastically different from their continental relatives both in morphology and genetics (Foster 1964; Hofman, et al. 2015; Lamichhaney, et al. 2015; Renaud, et al. 2015; Benítez-López, et al. 2021; Chevret, et al. 2021).

Compared to their continental relatives, island populations often experience a variety of changes of morphological, ecological and behavioral traits, which are referred to as “island syndrome” (Foster 1964; Adler and Levins 1994; Losos and Ricklefs 2009). One of the classical patterns of change is that species with small body size tend to become bigger on islands, while species with large body size tend to become smaller (“island rule” or Foster's rule; Foster 1964; Benítez-López, et al. 2021). The strength of evolutionary patterns following the island rule differs between taxa, but increase of body mass up to 100% has been observed repeatedly in island populations (Lomolino 1985; Angerbjörn 1986; Adler and Levins 1994; Payseur and Jing 2021; Renom, et al. 2021). The shift of body size can be fast, for example within two hundred years for the house mouse on Gough Island (Rowe-Rowe and Crafford 1992). Studies of the genetics underlying the island syndrome have mainly focused on detecting signatures of directional selection on the genes related to shifted phenotypes (Raia, et al. 2010; Choi, et al. 2021; Payseur and Jing 2021). However, the traits of island populations can also be changed by relaxation of purifying or positive selection, in two ways. First, the strength of selection can be reduced compared to the ancestral situation due to external changes; for example lower predation pressure, may lead to change of body size (Adler and Levins 1994) or coloration (Bliard, et al. 2020). Second, the efficacy of selection, mainly on mildly deleterious variation can be strongly reduced under genetic drift (Gravel 2016; Peischl, et al. 2018); for example with effects on immune functions of island birds (Barthe, et al. 2022).

The demographic history of island systems poses particular challenges to characterizing the contribution of selection to evolutionary change (Glinka, et al. 2003; Li, et al. 2012). Island populations usually experience bottlenecks at the founding stage and at least

partial genetic isolation, which may result in strong genetic drift and long-lasting genome-wide consequences (Ellegren, Smeds, Burri, Olason, Backström, Kawakami, Künstner, Mäkinen, Nadachowska-Brzyska, Qvarnström, et al. 2012; Consortium 2015; Friis, et al. 2018). The genomic background patterns caused by genetic drift make it particularly challenging to identify the genomic signatures of genes under selection like locally decreased genetic diversity and extreme Tajima's D, or increased $F_{ST}$ (Hahn 2008; Li, et al. 2012; Moinet, et al. 2022). Genomic patterns resulting from strong genetic drift may conceal the typical "genomic island" features of selection (Stephan 2016). For example, a long region with low diversity resulting from genetic drift could be interpreted as caused by a selective sweep (Harris, et al. 2018) or segregating nonsynonymous alleles as a consequence of balancing selection (Johri, et al. 2022).

Linkage disequilibrium (LD) based methods measuring extended haplotype homozygosity (EHH, Sabeti, et al. 2002) are suitable for detecting recent positive selection in the face of genetic drift from population bottlenecks, as LD changes more quickly than site frequency patterns (Ferrer-Admetlla, et al. 2014; Weigand and Leese 2018). Methods like iHES (Voight, et al. 2006), XP-EHH (Sabeti, et al. 2007) and REHH (Qanbari, et al. 2010) look for long regions of homozygosity caused by selective sweep and use the statistics of full genomes as the background for significance test. EHH-based methods. Nowadays multiple genetic measurements in sliding windows are usually combined, for example with Fisher's combined probability test or hidden Markov model (HMM) clustering (Ayub, et al. 2013; Marques, et al. 2018), to obtain robust results. By combined test for multiple signatures in different categories, for example LD based XP-EHH and derived allele frequency, which are nearly uncorrelated in the way being affected by variations of recombination rate and genetic drift in neutral genomic regions, the accuracy of detection of positive selection can be boosted (Grossman, et al. 2010).

In this study, we investigate the genomic interplay between selection and genetic drift leading to the island syndrome in a system with exceptionally long and clear history of isolation. Common voles (*Microtus arvalis*) were introduced to the Orkney archipelago in the north of Scotland by Neolithic farmers over 5,000 years ago likely as a food resource (Haynes, et al. 2003; Martínková, et al. 2013; Romaniuk, et al. 2016). Common voles are very widespread in continental Europe with multiple distinct evolutionary lineages (Heckel, et al. 2005; Beysard and Heckel 2014) but Orkney vole ancestry can be traced back only to

populations in Belgium or northern France (Wang et al. in press; Martínková, et al. 2013). The species is absent from the rest of the British Isles but abundant on seven Orkney islands. Archaeological material showed that the body size of voles increased rapidly in the first centuries after introduction (Cucchi, et al. 2014) leading to the current Orkney vole (*M. a. orcadensis*) phenotype with twice the body mass of its continental relatives (Millais 1904). The Orkney vole experienced a severe bottleneck through the introduction with a 20 to 30-fold decrease in the effective population size ($N_e$), but the $N_e$ remained relatively large at several thousand even during the bottleneck (Wang, et al. Accepted). However, genetic drift in complete isolation for approximately 10 000 generations has left strong impacts in the genome of Orkney voles, including low genetic diversity and the accumulation of homozygous deleterious alleles (Martínková, et al. 2013; Wang, et al. in submission).

Our analyses utilize the exceptionally clear history of the Orkney vole for assessing the long-term effects of genetic drift and selection on completely isolated natural vertebrate populations relative to populations that remained unaffected by isolation. Importantly, common vole populations on the continent underwent demographic fluctuations similar in timing and magnitude to Orkney voles ($N_e$ ~50,000- 80,000 in the last 2,000 generations; Wang et al. in press). We thus have the potential to identify effects of the long-term history of isolation and of the interactions with the specific environment for these mammal genomes. We hypothesized that generally benign environmental conditions for common voles on the Orkney islands and stronger bottlenecks in the colonization history (Wang et al. in press) may have resulted in a relaxation of selection compared to continental populations (Gravel 2016; Kutschera, et al. 2020). Relaxation of purifying selection could lead to overall elevated levels of non-synonymous variation in Orkney populations (Kryazhimskiy and Plotkin 2008). However, differences in the environment for thousands of generations at relatively large population sizes may also enable divergent selection and thus an adaptive cause for the rapid and drastic body size change, and other evolutionary changes of Orkney voles. Our comparative analyses of Orkney and continental populations reveal the differential interactions between neutral and selective forces that resulted in the evolution of the island syndrome of Orkney voles.

## Results

### Extreme genomic background of Orkney voles

We obtained full genome data from the resequencing of 14 Orkney voles and from 11 voles from Belgium and northern France as the region of origin for the ancient human introduction to Orkney (Figure 1a; Wang et al. in press). We refer to the latter samples in the following as continental voles. We further used three *M. obscurus* as outgroup in this study (Wang et al. in press). The average mapping depth was 26.0. After filtering, 23.7 million SNPs were kept, including 411,340 SNPs located in coding regions.

We characterized genomic differences between Orkney and continental voles in 39,784 genomic windows of 50 Kb. Genetic diversity of Orkney voles was almost one order of magnitude lower than for continental voles (median $\pi$: Orkney = $2.5\times10^{-4}$ vs. continent = $2.2\times10^{-3}$; Z-test, $p < 2.2\times10^{-16}$, Figure 2) with nearly complete diversity loss ($\pi < 1\times10^{-4}$) in 37% windows of the genome. In continental voles, only 0.7% of the windows had $\pi < 1\times10^{-4}$. Tajima's D of Orkney voles was overall slightly negative but with a very wide range throughout the genome (median -0.50, SE = 3.00, Figure 2) suggestive of signatures of population contraction (Tajima's D > 0) or expansion (Tajima's D < 0) depending on the genomic region. In contrast, Tajima's D of continental voles was slightly positive overall with a much narrower range of individual estimates (median 0.32, SE = 1.02; K-S test, $p < 2.2\times10^{-16}$; Figure 2). Genetic differentiation between Orkney and continental voles reached a very high level with median $F_{ST}$ = 0.47 (Figure 2). The distribution of differentiation was very homogeneous along the whole genome with only 1% of all genomic windows showing $F_{ST}$ estimates smaller than 0.05, and 43% higher than 0.5 (Figure 2, Figure S1).

### Overall increased dN/dS in Orkney

Among the SNPs in coding regions, 2.0% had loss-of-function (LOF) alleles, 47.4% had nonsynonymous alleles and 50.9% had synonymous alleles. 0.4% SNPs were multiallelic and had alleles in more than one functional category. Compared to continental voles, dN/dS ratios were increased for a larger number of genes in Orkney voles (Figure 1). The dN/dS ratios of roughly half of all genes (18,112 out of 36,074) were higher than 0 in at least one group. For these genes, we further analyzed those 8,213 genes which have a mouse homolog and dN/dS > 0. The average dN/dS ratio of Orkney (0.199, SE = 0.195) was 21% higher than in continental

voles (0.165, SE = 0.161; K-S test, $p < 2.2\times10^{-16}$; Figure 1b,). The dN/dS ratio was increased for 4,427 genes in Orkney voles by 0.121 (SE = 0.145) on average and decreased for 3,719 genes by 0.069 (SE = 0.089) on average compared to continental voles (Figure 1c).

**Divergent positive selection shaping the genome**

We used the density-based spatial clustering of applications with noise (DBSCAN) algorithm (Ester, et al. 1996; Hahsler, et al. 2019) to identify outlier regions potentially affected by positive selection against the genomic background. The DBSCAN algorithm clusters data points in multidimensional space based on their distance between each other, and data points with long distances from the majority are outliers (Schubert, et al. 2017). For each 50 Kb window, we included 6 parameters for clustering: Δπ ($\pi_{continent} - \pi_{Orkney}$), ΔTajima's D (Tajima's $D_{continent}$ – Tajima's $D_{Orkney}$), ΔAF (deviation of major allele frequency, $AF_{continent} - AF_{Orkney}$), $F_{ST}$, recombination rate and median XP-EHH. DBSCAN identified three clusters of genomic windows and 925 outlier windows (Fig 3). The major cluster containing 95.7% of all windows represented most regions of the genome and was considered as the genomic background not strongly affected by positive selection. Two further clusters of 11 and 9 genomic windows and the 925 outlier windows were then considered in the next step as potentially positively selected regions.

The outliers defined by DBSCAN include any region with extreme parameters, for example with highest recombination rates, so we used a hard filter of 1$^{st}$ percentile on XP-EHH to determine regions that potentially went through hard selective sweeps (see Methods). There were 52 regions ranging from 50 Kb to 200 Kb in length within the highest 1$^{st}$ percentile of median XP-EHH and thus considered as positively selected regions only in Orkney voles. Within the lowest 1$^{st}$ percentile of median XP-EHH suggesting positive selection only in continental voles, there were 115 regions ranging from 50 Kb to 700 Kb (Figure 3). Additional HMM clustering primarily identified 1332 outliers which included 321 (35%) of the DBSCAN outlier windows. A total of 43 primary HMM outlier windows were within the highest 1st percentile of median XP-EHH, overlapping with 19 windows from the DBSCAN result (26%). There were 214 HMM outlier windows within the lowest 1st percentile of median XP-EHH and a high overlap with the DBSCAN result (134 windows, 75%).

In total, 164 genes orthologous to mouse were contained in the regions identified as under hard sweep with DBSCAN in continental voles (example in Figure 4a), 113 of which were

in the regions also detected by HMM clustering. The average dN/dS ratio of the genes detected with DBSCAN in continental voles was 0.068. Only one gene (2900026A02Rik) with unknown function had dN/dS higher than 1. The genes showed significant enrichment mainly in two types of Gene Ontology Biological Process terms (31 terms in total, Table S2): histone monoubiquitination and lipid metabolic process (Table 1). The genes belonging to the term histone H2A-K119 monoubiquitination include Bmi1 and several Pcgf genes and are regulated by insulin-like growth factors in mammals (Nacerddine, et al. 2012). A series of cytochrome P450 genes in the second most significant term, epoxygenase P450, which were also all detected by HMM, play an important role in lipid metabolisim and affect body size in mammals by regulating and oxidizing steroid hormones (e.g., sex hormones, glucocorticoids), fatty acids, and xenobiotics (Medhora, et al. 2007; Knockaert, et al. 2011).

In Orkney voles, there were 46 genes orthologous to mouse identified with DBSCAN as under hard sweeps, and 17 of them belong to the GO term immune response (GO:0006955) (example in Figure 4b). Among 40 significant GO terms, 33 were related to immune response (Table 2, Table S3), including response to protozoans, MHC class II protein and T cell activation. The average dN/dS ratio of the detected genes in Orkney voles was 0.051, and none was higher than 1. With HMM clustering, 29 genes were identified. Nineteen of these overlapped with DBSCAN results. Sixteen of which belong to GO term immune response, suggesting a strong signature of positive selection on these immune genes captured by both methods.

## Discussion

In this study, we showed with an island system that both genetic drift and shift of selection shaped the genome drastically. Demographic history of prolonged bottleneck and long isolation could lead to unusual genomic landscape and relaxation of purifying selection in the population. We also showed the potential role of relaxed selection on the evolution of island syndrome that the increase of body size of Orkney voles was likely due to reduced positive selection on small body size related to reduced competition on the islands.

### Genome wide relaxation of purifying selection

The overall increased dN/dS ratios in Orkney voles compared to continental voles suggests a lower effectiveness of selection in the population that experienced long term isolation (Figure 1). The increase of dN/dS in over half of the inferred genes compared to continental voles was likely the consequence of relaxed purifying selection (Gravel 2016; Settepani, et al. 2016; Kutschera, et al. 2020), consistent with the fixation rate of loss-of-function mutations in Orkney (Wang et al. in press). Both lower selective pressure from the environment (e.g. less competition on food resources, see discussion below about body size) and reduced efficacy of selection due to strong genetic drift (Gravel 2016) are expected to play a role in the relaxation of selection. It is possible that positive selection on particular genes contributed to the global increase of dN/dS ratios, but we do not consider it as the major driver because of the high number of affected genes. The dN/dS ratio exceeding one is an indication of positive selection (Nielsen and Yang 2003), and positive selection can leave widespread traces over the genome (Carneiro, et al. 2012; Enard, et al. 2014). However, the distribution of dN/dS ratios was centered close to zero in both populations (Figure 1b), showing the dominance of purifying selection across the genome (Kimura 1977; Kryazhimskiy and Plotkin 2008). Furthermore, only 3% of the genes with increased dN/dS ratios in Orkney voles had dN/dS reaching 1, none of which were detected in our test of hard sweep, indicating limited genes affected by positive selection.

### Extreme demography led to extreme genomic landscape

Our results showed that the rising level of genetic drift may flood the genomic landscape globally and cover the genomic islands of selection, which may bring problems to

the study of adaptation of island populations. The reduction of genetic diversity could be preserved in the population through an ancient bottleneck (Nei, et al. 1975), and the decrease affected the genome globally (Fig 3). Regions of exceptionally high differentiation have been used to identify parts of the genome under divergent selection (Beaumont 2005; Ellegren, Smeds, Burri, Olason, Backström, Kawakami, Künstner, Mäkinen, Nadachowska-Brzyska and Qvarnström 2012). In our study, mean $F_{ST}$ was already very high, and the $F_{ST}$ values of candidate regions for hard sweeps were lower than the genomic background on average (Z-test, p = 0.0001, Fig 3, Fig S1). This can be explained by the reduction of genetic diversity of Orkney populations, likely together with private mutations emerged during the isolation (Holsinger and Weir 2009; Meirmans and Hedrick 2011). The landscape of Tajima's D strongly varied in the Orkney population (SD = 3.00) compared to continental voles (Figure 2, Figure S1), and directly using Tajima's D as a measurement of selection is impossible in this case. Our results showed that a prolonged bottleneck and long isolation can leave global marks of contraction (leads to positive Tajima's D) and expansion (leads to negative Tajima's D) in the genome (Tajima 1989) even after thousands of generations' recovery to large population.

The candidate regions for hard sweeps in continental voles had typical patterns of positive selection, including decreased $\pi$, negative Tajima's D and shifted allele frequencies (Nielsen, et al. 2005). The high overlap between hard sweep regions in continental voles detected by HMM and DBSCAN methods also showed the robustness of DBSCAN method. However, in Orkney voles, the genetic parameters of candidate regions of hard sweep overlapped with the background, especially $\pi_{Orkney}$, Tajima's D and $F_{ST}$, and did not show typical patterns of "genomic islands" (Fig 4b, Fig S1). Two factors may play a role to the fact that fewer hard sweep regions detected for Orkney voles: a true difference in selection compared to continental conspecifics (see the discussion below), and reduced statistical power for detection caused by genetic drift (Grossman, et al. 2010; Ferrer-Admetlla, et al. 2014).

**Island syndrome potentially caused by relaxed selection**

Relaxation from selection against large body size may have contributed to the island syndrome phenotype of Orkney voles. Our test found signatures of hard sweeps related to body size only in continental voles, contrary to studies of giant island mice in which selection acted directly on growth-related genes in island populations (Payseur and Jing 2021; Renom, et al. 2021). Genes related to lipid metabolism in multiple windows were detected to be under

positive selection and may play a role in keeping the continental voles at a small size. In details, these genes include cytochrome P450 genes (*CYP*) which are regulated by growth factors (Michaelis, et al. 2003), and are related to vessel growth (Michaelis, et al. 2003; Medhora, et al. 2007) and obesity (Takahashi, et al. 1999; Knockaert, et al. 2011). Further genes may be related to early sexual maturity and stop of body growth, including growth differentiation factor 1 (*GDF1*), which is essential for neural development (Lee 1990) and connected with obesity in mice (Onishi, et al. 2016); sterol carrier protein 2 (*SCP2*) and adrenodoxin (*ADX*), both involved in steroid hormone, especially sex hormone synthesis (Mendis-Handagama, et al. 1992; Huang, et al. 2022). Genes related to histone H2A-K119 monoubiquitination () were also detected under selection. Genetic changes that decrease the insulin-like growth factor signaling were found to increase lifespan in mammals (Brown-Borg 2015). However, histone H2A as part of the chromatin is evolutionarily conserved (Molaro, et al. 2018) and plays a fundamental role in transcriptional repression during organism development (Tamburri, et al. 2020). Thus, we could not draw a conclusion whether and how the hard sweep signatures on the genes involved in histone monoubiquitination were related to the biological changes of Orkney voles.

It is likely that Orkney voles were released from the selective pressure of small size. The size of molars of Orkney voles increased rapidly after introduction for over 20% (Cucchi, et al. 2014), indicating fast evolution of the island syndrome of the Orkney vole. The molar size of ancient voles (800 – 22,000 years ago), except in Spain, was 5% bigger than for modern samples (Cucchi, et al. 2014). The body sizes decreased both on the continent and in Orkney in the past centuries (Cucchi, et al. 2014), indicating the selection of body size was related to competition and population density. In a broad range of animal species, body size is negatively correlated with population density, both at population levels e.g. freshwater invertebrates (Schmid, et al. 2000) and Damaraland mole-rat (Finn, et al. 2018), and species levels e.g. birds (Juanes 1986) and herbivorous mammals (Damuth 1981). Snow voles (*Chionomys nivalis*) in the Alps are under selection for small body size and early maturity, which reduces food requirements and increases survival through winter (Bonnet, et al. 2017). It was likely that the Orkney voles experienced less competition during the phase of fast expansion after the introduction until reaching the carrying capacity. The decrease of body size of Orkney voles in the past centuries may thus be explained by increased competition after reaching the carrying capacity (Rowe-Rowe and Crafford 1992; Schmid, et al. 2000).

**Human influenced evolution of immunity**

The Orkney population experienced strong sweeps on genes related to antigen processing and defense to protozoans which pointed to the selective pressure of pathogens. It seems possible that these signals relate to the introduction history of Orkney voles that involved transport and storage of the voles by the Neolithic farmers (Romaniuk, et al. 2016), as infectious diseases are in general easy to transmit among the hosts in high density (Hochberg 1991; Laakkonen, et al. 1999; Altizer, et al. 2006). Data on differences in the pathogen communities between Orkney and European continent are very limited, but pathogen diversity (amoebas in rodents and Cryptosporidium in cattles, Fulton and Joyner 1948; Morrison, et al. 2008) or infection rate (John Cunningham virus in humans, Paratuberculosis in cattles, Poskanzer, et al. 1980; Beasley, et al. 2011) was found to be lower in Orkney. It is not likely that the selective pressure of pathogens was stronger on Orkney islands, which are an isolated archipelago with low biodiversity (Essl, et al. 2013), than on the European continent (Keesing, et al. 2010). On the other hand, the window with the highest median XP-EHH in Chr15 (Fig 4b) contained a series of MHC class II genes (e.g. H2-Aa and H2-Eb) which are known to be under balancing selection, though not found in the common vole, in many other species (Hedrick 1998; Aguilar, et al. 2004; Tollenaere, et al. 2008; Barthe, et al. 2022). The genetic diversity of this window ($\pi$=0.0076) is 2.5-fold higher than the average in the continental group, which is also an indication of balancing selection (Charlesworth 2006). Therefore, we cannot completely exclude the possibility that reduced efficacy of balancing selection on MHC II genes, either due to drift or the environment in Orkney, contributed to local reduction of genetic diversity and enhanced the signature of hard sweep (Sabeti, et al. 2007).

## Conclusion

Orkney voles showed genome-wide relaxation of purifying selection and an unusual genomic landscape. These patterns reflected strong and long-lasting impact of genetic drift on the genome of populations experienced prolonged population bottleneck and long isolation dominating the whole landscape of the genome. We found evidence that the size increase of the Orkney vole might be caused by relaxed positive selection of small body size,

indicating the potential role of relaxed selection for island syndrome. Thus, to investigate the evolutionary processes involved in the island syndrome, it is necessary to consider not only selection directly affecting the island population, but also selection ceased on the island.

## Material and Methods

### Calling and annotation of the genetic variants

In this study, we included 14 samples of *M. arvalis* from the Orkney islands (2 from each island) and 11 from Belgium and northern France (referred to as continental voles). Three *M. obscurus* samples were used as outgroup. All samples used in this study were collected for previous studies (Wang et al. in press; Braaker and Heckel 2009; Martínková, et al. 2013; Lischer, et al. 2014). DNA extraction was performed with the phenol-chloroform method. DNA quality and concentration were checked with 1% agarose gels, Qubit fluorometer (Life Technologies) and NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific). Sequencing libraries were produced with Illumina TruSeq DNA PCR-Free Library Prep Kit. The sequencing was done with Illumina Hiseq 2000 or Novaseq 6000 by the NGS platform of the University of Bern. Raw reads of all individuals (Figure 1a, Table S1) were mapped to the reference genome of *M. arvalis* (BioProject ID: PRJNA737461, Gouy et al. submitted) using Stampy 1.0.32. Duplicated reads were marked and filtered using GATK 4.0.10 (Poplin, et al. 2017). SNP calling was performed with GATK Haplotypecaller following the GVCF pipeline (Poplin, et al. 2017). Individual genotypes with read depth lower than 5 or higher than 100 were marked as missing sites. For each SNP, the overall filter parameters were: QD<15.0, SOR>3.0, FS>60.0, MQ<40.0, MQRankSum<-12.5, ReadPosRankSum<-8.0, heterozygosity=100% and mean-depth over all individuals>50. SNP filtering was performed with GATK VariantFiltration and VCFtools 0.1.16 (Danecek, et al. 2011). Only SNPs without any missing data were kept for further analyses. The functional effect of the SNPs was annotated with SnpEff 5.0 using the gene annotation from Gouy *et al*. (submitted). We only used the 13,543 genes which have a orthologue in mouse for further analyses.

### Calculation of dN/dS

The dN/dS ratio was calculated for each annotated gene based on its longest transcript. The exon sequences were produced for each individual by replacing the corresponding sites

of the reference sequence with SNP information of each individual using GATK FastaAlternateReferenceMaker. Heterozygous sites were replaced with IUPAC codes. For each gene, pairwise dN/dS ratios were calculated between *M. obscurus* and the other samples, and then averaged for Orkney or continental voles. The dN/dS calculation was performed with KaKs_Calculator 2.0 (Wang, et al. 2010) using the "NG" method (Nei and Gojobori 1986).

**Detection of hard sweeps**

Several genetic parameters were estimated along the genome in non-overlapping 50 Kb windows. With VCFtools, we calculated π and Tajima's D in 50 Kb windows for both Orkney and continental voles, and $F_{ST}$ between the groups. The major allele frequency (AF) was calculated per SNP and averaged for each window with BEDTools 2.28 (Quinlan and Hall 2010). The cross-population extended haplotype homozygosity test (XP-EHH, Simonson, et al. 2010) was performed for each SNP using R package "rehh" (Gautier and Vitalis 2012) with default parameters. XP-EHH tests specifically for hard sweep by comparing the haplotype homozygosity between two populations. An XP-EHH value strongly deviating from zero indicates a hard selective sweep in one group but not in the other (Simonson, et al. 2010). No ancestral state was defined for the SNPs, so the dataset was regarded as non-polarized. The median XP-EHH value for the SNPs in each 50 kb window was used to represent the state of the window.

A density-based clustering algorithm DBSCAN (Schubert, et al. 2017) was used to determine outlier regions potentially under positive selection. DBSCAN was used based on the assumption that genomic regions affected by strong positive selection should be rare and statistically distinct from most of the genome (Johri, et al. 2022), which was generally affected by genetic drift (Wolf and Ellegren 2017). Parameters used for clustering were: $\Delta\pi$ ($\pi_{continent} - \pi_{Orkney}$), ΔTajima's D (Tajima's $D_{continent}$ – Tajima's $D_{Orkney}$), ΔAF ($AF_{continent} - AF_{Orkney}$), $F_{ST}$, recombination rate and median XP-EHH value. The clustering was performed with R package "dbscan" (Hahsler, et al. 2019). The radius of the epsilon neighborhood (eps) was set to the "knee" value (the value when a curve changes suddenly, calculated following Satopaa, et al. 2011) of the distribution of the kNN distances of all the data points (k=6). The number of minimum points required for a cluster was set to 7 (number of dimensions plus 1) as recommended by the package. A Hidden Markov model (HMM) clustering method (Hofer, et al. 2012) was used to identify outlier regions independently as comparison. Fisher's combined

probability test was applied to five out of six parameters except recombination rate with R package “poolr” to obtain one combined *p* value for each window. The quantile of $F_{ST}$ was transformed to a one-tailed *p* value, and the quantile of the remaining four parameters were transformed to two-tailed *p* values. Each *p* value was then corrected for multiple testing with the Benjamini & Hochberg correction implemented in R function “p.adjust”. The corrected *p* values were converted to *q* values (1-*p*) and Z-transformed, and used as the input to assign the windows to three states using the Viterbi algorithm with R package “HiddenMarkov” following Marques, et al. (2016). The regions containing the windows with combined *p* value < 0.01 were considered as outliers.

To focus on the potential regions that experienced hard selective sweeps, a hard filter of median XP-EHH value was applied to the outliers detected by DBSCAN and HMM clustering. Outlier regions with the median XP-EHH value within the highest 1st percentile among all regions were considered potential positively selected regions only in Orkney voles. Outlier regions with the median XP-EHH value within the lowest 1st percentile were regarded as potential positively selected regions only in continental voles. For both categories of regions, functional enrichment analyses of Gene Ontology terms were performed on g:Profiler web server (Raudvere, et al. 2019) with a p-value threshold at 0.01 for the genes falling at least 50% in the regions.

## Acknowledgements

We thank the Next Generation Sequencing Platform of the University of Bern for sequencing services. This study was supported by Swiss National Science Foundation grant 31003A_176209 to GH. XW received a scholarship for PhD study from the China Scholarship Council (No. 201706380049). Part of the computations was performed on UBELIX, the HPC cluster at the University of Bern (http://www.id.unibe.ch/hpc).

## References

Adler GH, Levins R. 1994. The Island Syndrome in Rodent Populations. The Quarterly Review of Biology 69:473-490.

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK. 2004. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proceedings of the National Academy of Sciences 101:3490-3494.

Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P. 2006. Seasonality and the dynamics of infectious diseases. Ecology Letters 9:467-484.

Angerbjörn A. 1986. Gigantism in island populations of wood mice (*Apodemus*) in Europe. Oikos:47-56.

Ayub Q, Yngvadottir B, Chen Y, Xue Y, Hu M, Vernes SC, Fisher SE, Tyler-Smith C. 2013. FOXP2 targets show evidence of positive selection in European populations. The American Journal of Human Genetics 92:696-706.

Barthe M, Doutrelant C, Covas R, Melo M, Illera JC, Tilak M-K, Colombier C, Leroy T, Loiseau C, Nabholz B. 2022. Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome. BioRxiv:2021.2011.2021.469450.

Beasley L, Truyers I, Mellor D, Norquay R, Duthie S, Ellis K. 2011. Prevalence of Johne's disease among cattle in Orkney. Veterinary Record-English Edition 169:50.

Beaumont MA. 2005. Adaptation and speciation: what can Fst tell us? Trends in Ecology & Evolution 20:435-440.

Benítez-López A, Santini L, Gallego-Zamorano J, Milá B, Walkden P, Huijbregts MAJ, Tobias JA. 2021. The island rule explains consistent patterns of body size evolution in terrestrial vertebrates. Nature Ecology & Evolution 5:768-786.

Beysard M, Heckel G. 2014. Structure and dynamics of hybrid zones at different stages of speciation in the common vole (*Microtus arvalis*). Molecular Ecology 23:673-687.

Bliard L, Paquet M, Robert A, Dufour P, Renoult JP, Grégoire A, Crochet P-A, Covas R, Doutrelant C. 2020. Examining the link between relaxed predation and bird coloration on islands. Biology Letters 16:20200002.

Bonnet T, Wandeler P, Camenisch G, Postma E. 2017. Bigger is fitter? Quantitative genetic decomposition of selection reveals an adaptive evolutionary decline of body mass in a wild rodent population. PLoS Biology 15:e1002592.

Braaker S, Heckel G. 2009. Transalpine colonisation and partial phylogeographic erosion by dispersal in the common vole (*Microtus arvalis*). Molecular Ecology 18:2518-2531.

Brown-Borg HM. 2015. The somatotropic axis and longevity in mice. American Journal of Physiology-Endocrinology and Metabolism 309:E503-E510.

Carneiro M, Albert FW, Melo-Ferreira J, Galtier N, Gayral P, Blanco-Aguiar JA, Villafuerte R, Nachman MW, Ferrand N. 2012. Evidence for widespread positive and purifying selection across the European rabbit (*Oryctolagus cuniculus*) genome. Molecular Biology and Evolution 29:1837-1849.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLOS Genetics 2:e64.

Chevret P, Hautier L, Ganem G, Herman J, Agret S, Auffray J-C, Renaud S. 2021. Genetic structure in Orkney island mice: isolation promotes morphological diversification. Heredity 126:266-278.

Choi JY, Dai X, Alam O, Peng JZ, Rughani P, Hickey S, Harrington E, Juul S, Ayroles JF, Purugganan MD, et al. 2021. Ancestral polymorphisms shape the adaptive radiation of *Metrosideros* across the Hawaiian Islands. Proceedings of the National Academy of Sciences 118:e2023801118.

Consortium GP. 2015. A global reference for human genetic variation. Nature 526:68-74.

Cucchi T, Barnett R, Martínková N, Renaud S, Renvoisé E, Evin A, Sheridan A, Mainland I, Wickham-Jones C, Tougard C, et al. 2014. The changing pace of insular life: 5000 years of microevolution in the orkney vole (*Microtus arvalis orcadensis*). Evolution 68:2804-2820.

Damuth J. 1981. Population density and body size in mammals. Nature 290:699-700.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756-760.

Enard D, Messer PW, Petrov DA. 2014. Genome-wide signals of positive selection in human evolution. Genome research 24:885-895.

Essl F, Moser D, Dirnböck T, Dullinger S, Milasowszky N, Winter M, Rabitsch W. 2013. Native, alien, endemic, threatened, and extinct species diversity in European countries. Biological Conservation 164:90-

97.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. Molecular Biology and Evolution 31:1275-1291.

Finn K, Parker DM, Bennett NC, Zöttl M. 2018. Contrasts in body size and growth suggest that high population density results in faster pace of life in Damaraland mole-rats (*Fukomys damarensis*). Canadian Journal of Zoology 96:920-927.

Foster JB. 1964. Evolution of mammals on islands. Nature 202:234-235.

Friis G, Fandos G, Zellmer AJ, McCormack JE, Faircloth BC, Milá B. 2018. Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. Molecular Ecology 27:5137-5153.

Fulton J, Joyner L. 1948. Natural Amoebic Infections in Laboratory Rodents. Nature 161:66-68.

Gautier M, Vitalis R. 2012. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28:1176-1177.

Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics 165:1269-1278.

Grant PR, Grant BR. 2014. 40 Years of evolution: Darwin's finches on Daphne Major Island. Princeton University Press.

Gravel S. 2016. When is selection effective? Genetics 203:451-462.

Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, et al. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327:883-886.

Hahn MW. 2008. Toward a selection theory of molecular evolution. Evolution 62:255-265.

Hahsler M, Piekenbrock M, Doran D. 2019. dbscan: Fast Density-Based Clustering with R. Journal of statistical software 91:1-30.

Harris RB, Sackman A, Jensen JD. 2018. On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. PLOS Genetics 14:e1007859.

Haynes S, Jaarola M, Searle JB. 2003. Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. Molecular Ecology 12:951-956.

Heckel G, Burri R, Fink S, Desmet J-F, Excoffier L. 2005. Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. Evolution 59:2231-2242.

Hedrick PW. 1998. Balancing selection and MHC. Genetica 104:207-214.

Hochberg ME. 1991. Non-linear transmission rates and the dynamics of infectious disease. Journal of Theoretical Biology 153:301-321.

Hofer T, Foll M, Excoffier L. 2012. Evolutionary forces shaping genomic islands of population differentiation in humans. BMC genomics 13:107.

Hofman CA, Rick TC, Hawkins MTR, Funk WC, Ralls K, Boser CL, Collins PW, Coonan T, King JL, Morrison SA, et al. 2015. Mitochondrial genomes suggest rapid evolution of dwarf California Channel Islands Foxes (*Urocyon littoralis*). PloS one 10:1-15.

Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting $F_{ST}$. Nature Reviews Genetics 10:639-650.

Huang Q, Luo L, Han X, Li F, Zhang X, Tian M. 2022. Low-dose perfluorooctanoic acid stimulates steroid hormone synthesis in Leydig cells: Integrated proteomics and metabolomics evidence. Journal of Hazardous Materials 424:127656.

Johri P, Aquadro CF, Beaumont M, Charlesworth B, Excoffier L, Eyre-Walker A, Keightley PD, Lynch M, McVean G, Payseur BA, et al. 2022. Recommendations for improving statistical inference in population genomics. PLoS Biology 20:e3001669.

Juanes F. 1986. Population density and body size in birds. The American Naturalist 128:921-929.

Keesing F, Belden LK, Daszak P, Dobson A, Harvell CD, Holt RD, Hudson P, Jolles A, Jones KE, Mitchell CE, et al. 2010. Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature

468:647-652.

Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275-276.

Knockaert L, Fromenty B, Robin M-A. 2011. Mechanisms of mitochondrial targeting of cytochrome P450 2E1: physiopathological role in liver injury and obesity. The FEBS Journal 278:4252-4260.

Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. PLOS Genetics 4:e1000304.

Kutschera VE, Poelstra JW, Botero-Castro F, Dussex N, Gemmell NJ, Hunt GR, Ritchie MG, Rutz C, Wiberg RAW, Wolf JBW. 2020. Purifying selection in corvids is less efficient on islands. Molecular Biology and Evolution 37:469-474.

Laakkonen J, Henttonen H, Niemimaa J, Soveri T. 1999. Seasonal dynamics of Pneumocystis carinii in the field vole, *Microtus agrestis*, and in the common shrew, *Sorex araneus*, in Finland. Parasitology 118:1-5.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin CJ, Wang C, Zamani N, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371-375.

Lee S-J. 1990. Identification of a novel member (GDF-1) of the transforming growth factor-β superfamily. Molecular Endocrinology 4:1034-1040.

Li J, Li H, Jakobsson M, Li S, SjÖDin P, Lascoux M. 2012. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Molecular Ecology 21:28-44.

Lischer HEL, Excoffier L, Heckel G. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: Implications for the evolutionary history of *Microtus* voles. Molecular Biology and Evolution 31:817-831.

Lomolino MV. 1985. Body size of mammals on islands - the island rule reexamined. American Naturalist 125:310-316.

Losos JB, Ricklefs RE. 2009. Adaptation and diversification on islands. Nature 457:830-836.

Marques DA, Jones FC, Di Palma F, Kingsley DM, Reimchen TE. 2018. Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. Nature Ecology and Evolution 2.

Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. PLOS Genetics 12:e1005887.

Martínková N, Barnett R, Cucchi T, Struchen R, Pascal M, Pascal M, Fischer MC, Higham T, Brace S, Ho SYW, et al. 2013. Divergent evolutionary processes associated with colonization of offshore islands. Molecular Ecology 22:5205-5220.

Medhora M, Dhanasekaran A, Gruenloh SK, Dunn LK, Gabrilovich M, Falck JR, Harder DR, Jacobs ER, Pratt PF. 2007. Emerging mechanisms for growth and protection of the vasculature by cytochrome P450-derived products of arachidonic acid and other eicosanoids. Prostaglandins & Other Lipid Mediators 82:19-29.

Meirmans PG, Hedrick PW. 2011. Assessing population structure: $F_{ST}$ and related measures. Molecular ecology resources 11:5-18.

Mendis-Handagama S, Watkins PA, Gelber SdJ, Scallen TJ. 1992. Leydig cell peroxisomes and sterol carrier protein-2 in luteinizing hormone-deprived rats. Endocrinology 131:2839-2845.

Michaelis UR, Fisslthaler B, Medhora M, Harder D, Fleming I, Busse R. 2003. Cytochrome P450 2C9-derived epoxyeicosatrienoic acids induce angiogenesis via cross-talk with the epidermal growth factor receptor. The FASEB Journal 17:770-772.

Millais J. 1904. On a new British vole from the Orkney islands. Zoologist 8:241-246.

Moinet A, Schlichta F, Peischl S, Excoffier L. 2022. Strong neutral sweeps occurring during a population contraction. Genetics 220:iyac021.

Molaro A, Young JM, Malik HS. 2018. Evolutionary origins and diversification of testis-specific short histone H2A variants in mammals. Genome research 28:460-473.

Morrison LJ, Mallon ME, Smith HV, MacLeod A, Xiao L, Tait A. 2008. The population structure of the *Cryptosporidium parvum* population in Scotland: A complex picture. Infection, Genetics and Evolution 8:121-129.

Nacerddine K, Beaudry J-B, Ginjala V, Westerman B, Mattiroli F, Song J-Y, Van Der Poel H, Ponz OB, Pritchard C, Cornelissen-Steijger P. 2012. Akt-mediated phosphorylation of Bmi1 modulates its oncogenic potential, E3 ligase activity, and DNA damage repair activity in mouse prostate cancer. The Journal of clinical investigation 122:1920-1932.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3:418-426.

Nei M, Maruyama T, Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. Evolution 29:1-10.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. Genome research 15:1566-1575.

Nielsen R, Yang Z. 2003. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. Molecular Biology and Evolution 20:1231-1239.

Onishi Y, Fukasawa K, Ozaki K, Iezaki T, Yoneda Y, Hinoi E. 2016. GDF1 is a novel mediator of macrophage infiltration in brown adipose tissue of obese mice. Biochemistry and Biophysics Reports 5:216-223.

Payseur BA, Jing P. 2021. Genomic targets of positive selection in giant mice from Gough Island. Molecular Biology and Evolution 38:911-926.

Peischl S, Dupanloup I, Foucal A, Jomphe M, Bruat V, Grenier JC, Gouy A, Gilbert KJ, Gbeha E, Bosshard L, et al. 2018. Relaxed selection during a recent human expansion. Genetics 208:763-777.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv:201178.

Poskanzer DC, Sever JL, Sheridan JL, Prenney LB. 1980. Multiple sclerosis in the Orkney and Shetland Islands. IV: Viral antibody titres and viral infections. Journal of Epidemiology and Community Health 34:258.

Qanbari S, Pimentel E, Tetens J, Thaller G, Lichtner P, Sharifi A, Simianer H. 2010. A genome‐wide scan for signatures of recent selection in Holstein cattle. Animal Genetics 41:377-389.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

Raia P, Guarino FM, Turano M, Polese G, Rippa D, Carotenuto F, Monti DM, Cardi M, Fulgione D. 2010. The blue lizard spandrel and the island syndrome. BMC evolutionary biology 10:1-16.

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Research 47:W191-W198.

Renaud S, Gomes Rodrigues H, Ledevin R, Pisanu B, Chapuis J-L, Hardouin EA. 2015. Fast evolutionary response of house mice to anthropogenic disturbance on a Sub-Antarctic island. Biological Journal of the Linnean Society 114:513-526.

Renom P, de-Dios T, Civit S, Llovera L, Sánchez-Gracia A, Lizano E, Rando JC, Marquès-Bonet T, Kergoat GJ, Casanovas-Vilar I. 2021. Genetic data from the extinct giant rat from Tenerife (Canary Islands) points to a recent divergence from mainland relatives. Biology Letters 17:20210533.

Romaniuk AA, Shepherd AN, Clarke DV, Sheridan AJ, Fraser S, Bartosiewicz L, Herman JS. 2016. Rodents: food or pests in Neolithic Orkney. Royal Society open science 3:160514.

Rowe-Rowe D, Crafford J. 1992. Density, body size, and reproduction of feral house mice on Gough Island. South African Journal of Zoology 27:1-5.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449:913-918.

Schmid P, Tokeshi M, Schmid-Araya J. 2000. Relation between population density and body size in stream

communities. Science 289:1557-1560.

Schubert E, Sander J, Ester M, Kriegel HP, Xu XW. 2017. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. Acm Transactions on Database Systems 42:1-21.

Settepani V, Bechsgaard J, Bilde T. 2016. Phylogenetic analysis suggests that sociality is associated with reduced effectiveness of selection. Ecology and Evolution 6:469-477.

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, et al. 2010. Genetic evidence for high-altitude adaptation in Tibet. Science 329:72-75.

Stephan W. 2016. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Molecular Ecology 25:79-88.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Takahashi J, Furuhata Y, Ikeda A, Takahashi M, Iwata H, Kazusaka A, Fujita S. 1999. Characterization of hepatic cytochrome P450 isozyme composition in the transgenic rat expressing low level human growth hormone. Xenobiotica 29:1203-1212.

Tamburri S, Lavarone E, Fernández-Pérez D, Conway E, Zanotti M, Manganaro D, Pasini D. 2020. Histone H2AK119 mono-ubiquitination is essential for polycomb-mediated transcriptional repression. Molecular Cell 77:840-856.e845.

Tollenaere C, Bryja J, Galan M, Cadet P, Deter J, Chaval Y, Berthier K, Ribas Salvador A, Voutilainen L, Laakkonen J. 2008. Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. Journal of Evolutionary Biology 21:1307-1320.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biology 4:e72.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. Genomics, Proteomics & Bioinformatics 8:77-80.

Weigand H, Leese F. 2018. Detecting signatures of positive selection in non-model species using genomic data. Zoological Journal of the Linnean Society 184:528-583.

Whittaker RJ, Fernández-Palacios JM. 2007. Island biogeography: ecology, evolution, and conservation. Oxford University Press.

Wolf JB, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics 18:87.

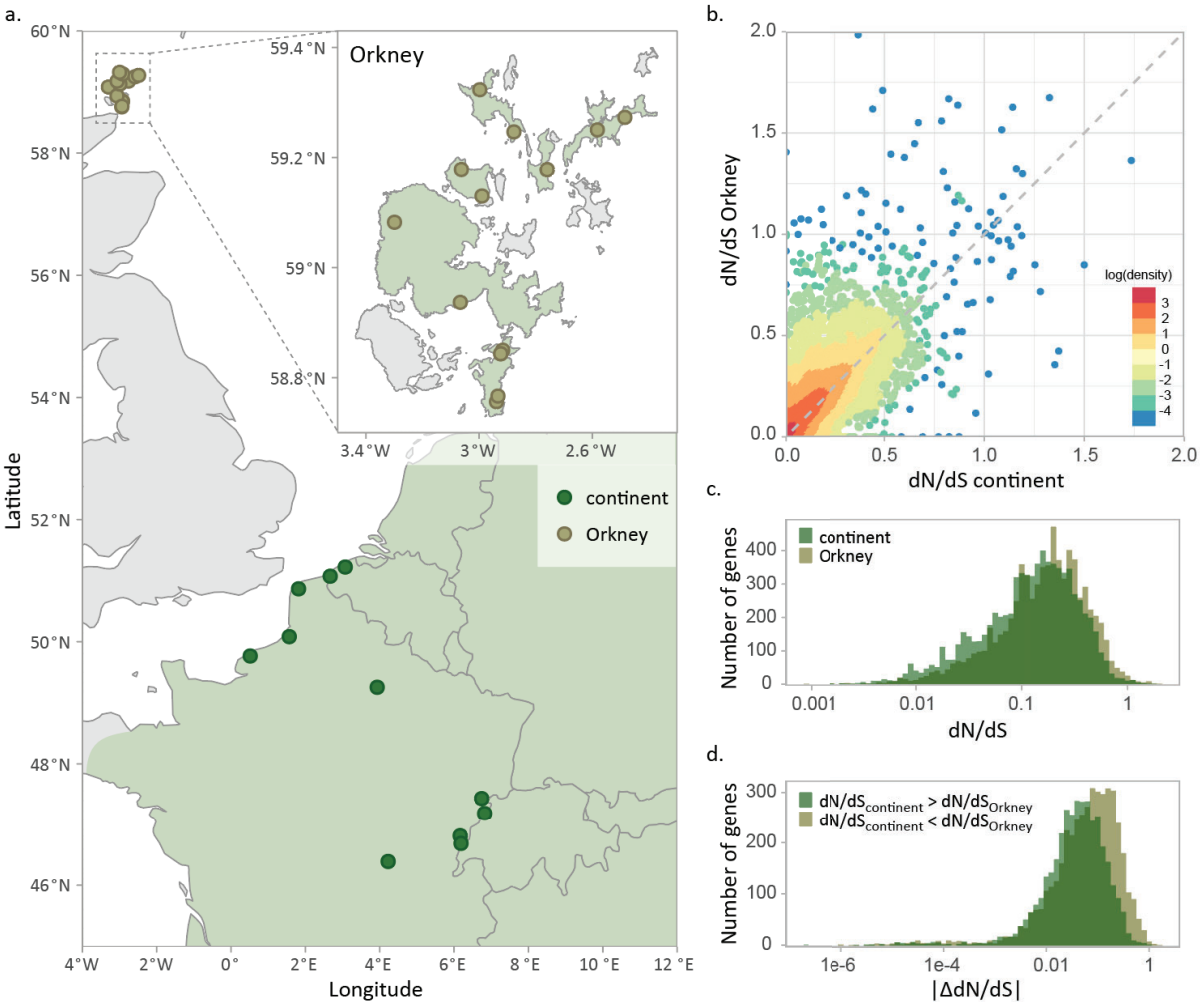



**Figure 1**. **a.** Sample locations of common voles *M. arvalis* in Western Europe and Scotland. Continental samples marked with green and Orkney samples with brown. The distribution range of *M. arvalis* was marked with light green. Notr that the species is absent from the British Isles except for Orkney. **b.** For each gene, the $d_N/d_S$ ratio of continental voles was plotted against the ratio of Orkney voles. There were more genes with higher ratio in Orkney than in the continent. The dashed line marks equal $d_N/d_S$ in both groups. The density of the points is marked with colors. **c.** Distribution of $d_N/d_S$ ratios of Orkney voles vs continental voles. **d.** For the genes with $|\Delta d_N/d_S| > 0$ between Orkney and continental voles, values were overall biased towards higher $d_N/d_S$ ratio in Orkney voles.

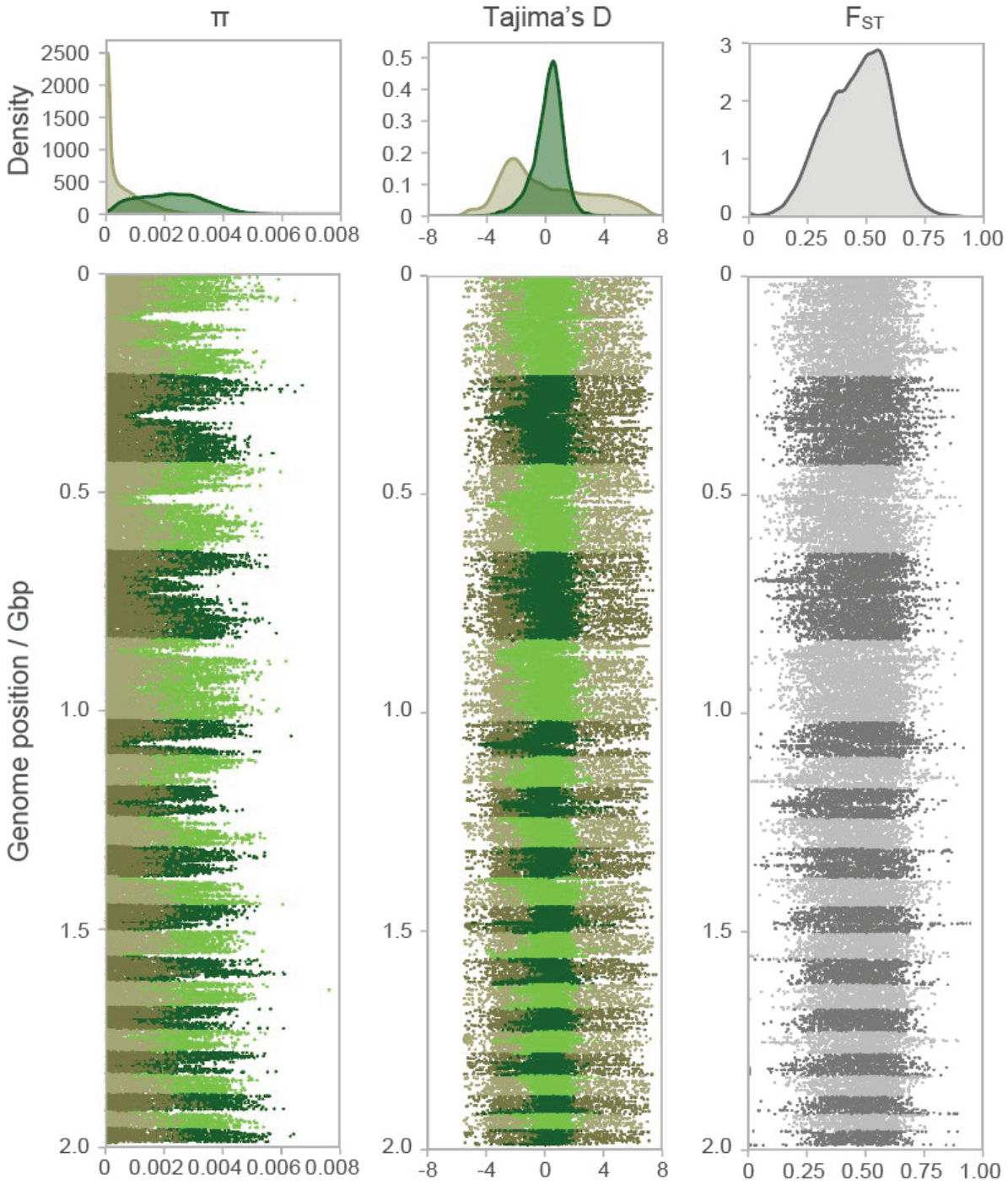



**Figure 2**. Genome-wide diversity and differentiation for common voles *M. arvalis* from the Orkney archipelago (brown) versus continental individuals (green). Density distributions (upper panel) and Manhattan plots (lower panel) for π, Tajima's D and $F_{ST}$ in 50 Kb windows along the genome. Different chromosomes are marked with lighter and darker colors.

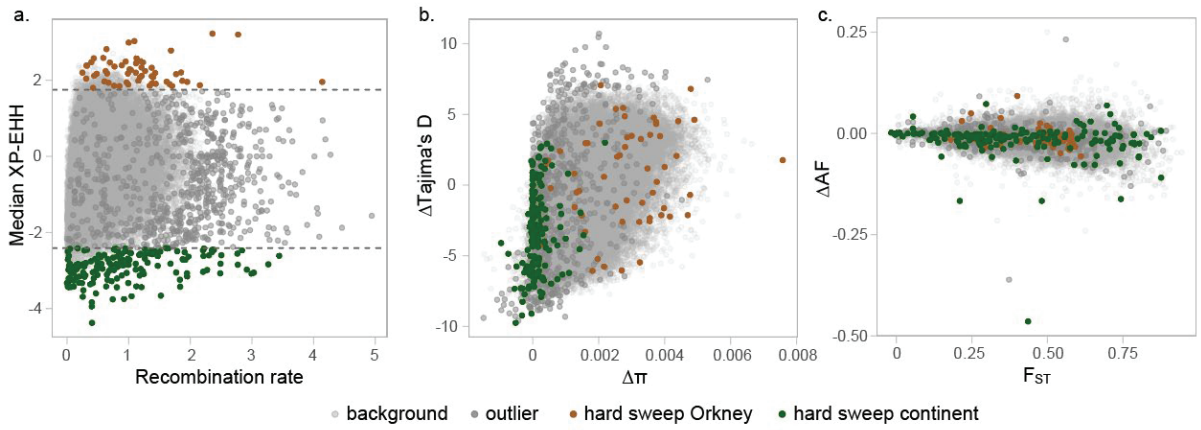



**Figure 3**. The distributions of genomic parameters used for tests of positive selection with Dbscan. Each dot refers to one 50 Kb window. **a.** The absolute median XP-EHH value of the outlier windows, which measures relative haplotype homozygosity, was negatively correlated ($R^2$=0.16) with recombination rate. The dashed lines show the threshold of highest and lowest 1% of XP-EHH. **b.** Regions positively selected only in continental voles had lower Δπ and ΔTajima's D, while regions positively selected only in Orkney showed large variation. **c.** The major allele frequency difference of the regions under hard selective sweep showed overall no obvious deviation from the genomic background (K-S test, $p$ = 0.32, Z-test $p$ = 0.19) and mean $F_{ST}$ values lower than the average value over the genome (Z-test $p$ = 0.0001).

**Table 1**. The 10 most significantly enriched GO Biological Process terms of genes that experienced hard sweeps in continental voles with the lowest $p_{adj}$ ($p$ value of enrichment adjusted with g:SCS method for multiple testing correction). Term size refers to the number of genes in the GO term. Intersection size refers to the number of genes under hard sweep. The GO terms listed are either related to histone monoubiquitination (Histone) or lipid metabolism (Lipid).

| Term ID | Term name | $p_{adj}$ | Term size | Intersection size | Relate to |
|---|---|---|---|---|---|
| **GO:0036353** | histone H2A-K119 monoubiquitination | 4.94E-08 | 9 | 5 | Histone |
| **GO:0019373** | epoxygenase P450 pathway | 1.43E-06 | 34 | 6 | Lipid |
| **GO:0035518** | histone H2A monoubiquitination | 3.28E-06 | 18 | 5 | Histone |
| **GO:0006805** | xenobiotic metabolic process | 6.64E-06 | 122 | 8 | Lipid |
| **GO:0071466** | cellular response to xenobiotic stimulus | 7.13E-06 | 179 | 9 | Lipid |
| **GO:0006629** | lipid metabolic process | 9.96E-06 | 1371 | 20 | Lipid |
| **GO:0033522** | histone H2A ubiquitination | 3.68E-05 | 28 | 5 | Histone |
| **GO:0010390** | histone monoubiquitination | 5.31E-05 | 30 | 5 | Histone |
| **GO:0009410** | response to xenobiotic stimulus | 7.89E-05 | 316 | 10 | Lipid |
| **GO:0051276** | chromosome organization | 8.2E-05 | 981 | 16 | Histone |

**Table 2**. The 10 most significantly enriched GO Biological Process terms of genes that experienced hard sweeps only in the Orkney vole population. $p_{adj}$ refers to $p$ value of enrichment adjusted with g:SCS method for multiple testing correction. Term size refers to the number of genes in the GO term. Intersection size refers to the number of genes under hard sweep. All the GO terms listed are immune related.

| Term ID | Term name | $p_{adj}$ | Term size | Intersection size |
|---|---|---|---|---|
| **GO:0006955** | immune response | 5.07E-11 | 1794 | 17 |
| **GO:0048002** | antigen processing and presentation of peptide antigen | 5.17E-08 | 71 | 6 |
| **GO:0002478** | antigen processing and presentation of exogenous peptide antigen | 7.13E-08 | 30 | 5 |
| **GO:0042832** | defense response to protozoan | 1.18E-07 | 33 | 5 |
| **GO:0001562** | response to protozoan | 1.88E-07 | 36 | 5 |
| **GO:0019884** | antigen processing and presentation of exogenous antigen | 2.17E-07 | 37 | 5 |
| **GO:0019882** | antigen processing and presentation | 9.88E-07 | 115 | 6 |
| **GO:0019886** | antigen processing and presentation of exogenous peptide antigen via MHC class II | 2.03E-06 | 18 | 4 |
| **GO:0002495** | antigen processing and presentation of peptide antigen via MHC class II | 4.85E-06 | 22 | 4 |
| **GO:0002504** | antigen processing and presentation of peptide or polysaccharide antigen via MHC class II | 7.03E-06 | 24 | 4 |

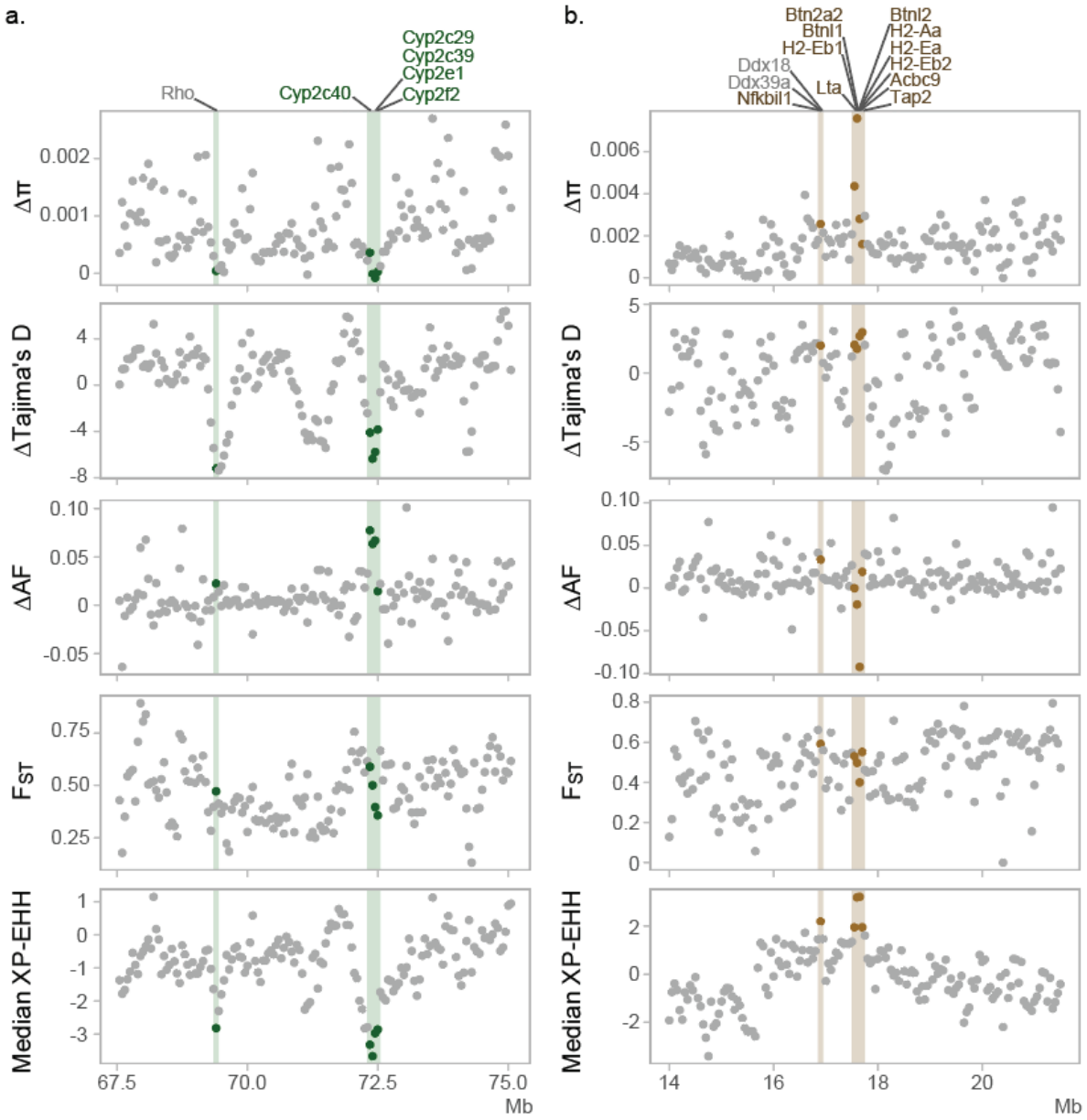



**Figure 4**. Examples of genomic landscapes close to the regions for which hard selective sweeps were detected in common voles from Orkney or the continent. Δπ ($\pi_{Western} - \pi_{Orkney}$), ΔTajima's D (Tajima's $D_{Western}$ – Tajima's $D_{Orkney}$), ΔAF ($AF_{Western} - AF_{Orkney}$), $F_{ST}$ and median XP-EHH are shown. Each point represents one 50 Kb window. Positively selected windows were marked with vertical lines. Genes overlapped with positively selected regions (outlier windows determined with DBSCAN with the highest or lowest one percentile of median XP-EHH value) were given on top. **A.** Positively selected regions of continental voles had low Δπ, ΔTajima's D and ΔAF. Genes belonging to the epoxygenase P450 pathway on *M. arvalis* chromosome 4 were marked with green. **b.** In Orkney voles, Δπ, ΔTajima's D and ΔAF of positively selected regions did not form clusters outstanding from the genomic background. Genes related to antigen processing and presentation on chromosome 15 were marked with brown.

**Supplementary Information for**

# Islands of selection in an ocean of genetic drift: genome-wide relaxation of selection of the Orkney vole

Xuejing Wang[a], Gerald Heckel[a,b]

[a]Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland; [b]Swiss Institute of Bioinformatics, Quartier Sorge - Batiment Amphipole, 1015 Lausanne, Switzerland

*Gerald Heckel

Email: gerald.heckel@unibe.ch

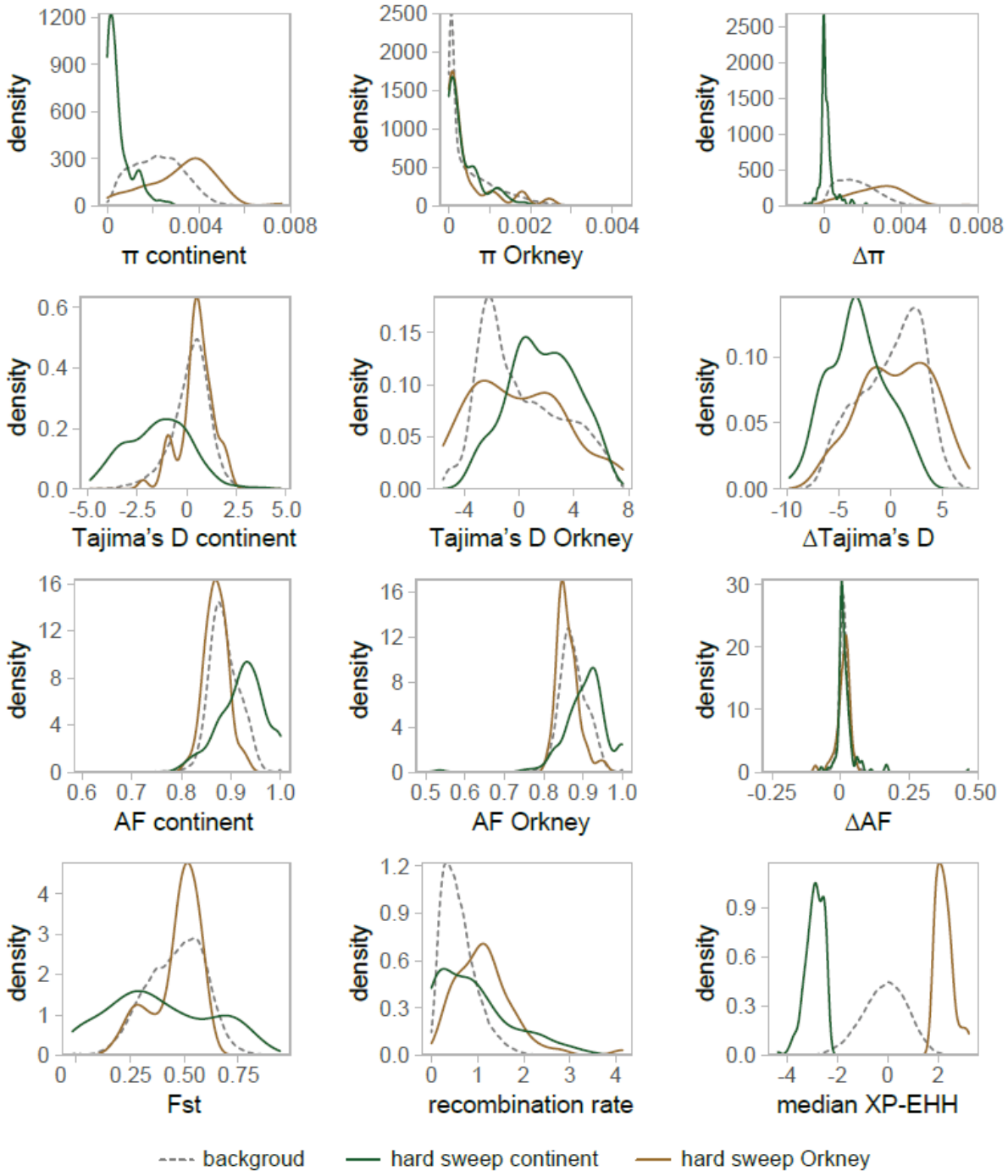



**Figure S1.** The density distributions of genomic parameters including π, Tajima's D and allele frequency (AF) in both groups; between groups: Δπ ($\pi_{continent} - \pi_{Orkney}$), ΔTajima's D (Tajima's $D_{continent}$ – Tajima's $D_{Orkney}$), ΔAF ($AF_{continent} - AF_{Orkney}$), $F_{ST}$ and median XP-EHH value; and recombination rate. The distribution of background windows (the majority cluster from DBSCAN, N=38,859) are shown as dashed line in grey. The distribution of windows of hard sweeps in continental voles (N=178) are shown as solid line in green and in Orkney voles (N=57) in brown.

**Table S1.** The samples used in this study with their sampling location and mapping depth.

| Individual | Group | Latitude | Longitude | Depth |
|---|---|---|---|---|
| **BSt095** | continent | 51.2232 | 3.07275 | 22.7 |
| **BVe061** | continent | 51.07315 | 2.66803 | 40.8 |
| **FBv02** | continent | 46.80561 | 6.164746 | 17.4 |
| **FCc02** | continent | 47.18444 | 6.822475 | 27.4 |
| **FDa504** | continent | 49.76666 | 0.51666 | 23.8 |
| **FFr549** | continent | 50.08333 | 1.56666 | 23.3 |
| **FGr14** | continent | 46.68058 | 6.18723 | 37.5 |
| **FMc03** | continent | 49.25437 | 3.93216 | 19.0 |
| **FOg02** | continent | 47.43055 | 6.742983 | 16.9 |
| **FPi555** | continent | 50.86666 | 1.81666 | 22.6 |
| **FSt24** | continent | 46.38548 | 4.22641 | 29.7 |
| **OBSO234** | Orkney | 58.85002 | -2.91659 | 23.4 |
| **OBWs01** | Orkney | 58.84349 | -2.92363 | 33.3 |
| **OEOs14** | Orkney | 59.17797 | -2.75946 | 26.0 |
| **OEOs15** | Orkney | 59.17797 | -2.75946 | 27.2 |
| **OMBr149** | Orkney | 59.08242 | -3.29908 | 22.4 |
| **OMHo277** | Orkney | 58.93692 | -3.06632 | 25.8 |
| **ORGs21** | Orkney | 59.13011 | -2.98977 | 34.3 |
| **ORWa268** | Orkney | 59.1782 | -3.0628 | 30.5 |
| **OSJC01** | Orkney | 58.75641 | -2.93866 | 30.3 |
| **OSWi166** | Orkney | 58.76666 | -2.93333 | 31.6 |
| **OWLs033** | Orkney | 59.2956 | -2.91336 | 18.8 |
| **OWPg01** | Orkney | 59.32317 | -2.9977 | 28.5 |
| **OSaLi257** | Orkney | 59.2502 | -2.58204 | 30.3 |
| **OSaNe01** | Orkney | 59.27294 | -2.48481 | 28.4 |
| **RuCh01** | *M. obscurus* | 61.33066 | 43.52866 | 16.8 |
| **RuSu03** | *M. obscurus* | 55.5 | 60.55999 | 21.9 |
| **TuEr01** | *M. obscurus* | 40.30937 | 41.94739 | 18.2 |

**Table S2.** Significant GO term enrichments of the candidate genes experienced hard sweep in continental voles. The threshold of p value was 0.05.

| Term name | Term ID | Adjusted *p*-value | Term size | Intersection size |
|---|---|---|---|---|
| **histone H2A-K119 monoubiquitination** | GO:0036353 | 4.82E-08 | 9 | 5 |
| **epoxygenase P450 pathway** | GO:0019373 | 1.37447E-06 | 34 | 6 |
| **histone H2A monoubiquitination** | GO:0035518 | 3.20649E-06 | 18 | 5 |
| **xenobiotic metabolic process** | GO:0006805 | 6.25664E-06 | 122 | 8 |
| **cellular response to xenobiotic stimulus** | GO:0071466 | 6.64774E-06 | 179 | 9 |
| **lipid metabolic process** | GO:0006629 | 8.34618E-06 | 1371 | 20 |
| **histone H2A ubiquitination** | GO:0033522 | 3.59218E-05 | 28 | 5 |
| **histone monoubiquitination** | GO:0010390 | 5.18412E-05 | 30 | 5 |
| **chromosome organization** | GO:0051276 | 7.15874E-05 | 981 | 16 |
| **response to xenobiotic stimulus** | GO:0009410 | 7.29914E-05 | 316 | 10 |
| **xenobiotic catabolic process** | GO:0042178 | 0.000103051 | 68 | 6 |
| **olefinic compound metabolic process** | GO:0120254 | 0.000115447 | 177 | 8 |
| **cellular lipid metabolic process** | GO:0044255 | 0.000126687 | 1023 | 16 |
| **positive regulation of RNA metabolic process** | GO:0051254 | 0.000134035 | 1786 | 21 |
| **positive regulation of nucleobase-containing compound metabolic process** | GO:0045935 | 0.000138588 | 1957 | 22 |
| **arachidonic acid metabolic process** | GO:0019369 | 0.000218435 | 77 | 6 |
| **positive regulation of transcription by RNA polymerase II** | GO:0045944 | 0.000395745 | 1261 | 17 |
| **histone ubiquitination** | GO:0016574 | 0.000597001 | 48 | 5 |
| **positive regulation of transcription, DNA-templated** | GO:0045893 | 0.000840321 | 1654 | 19 |
| **positive regulation of nucleic acid-templated transcription** | GO:1903508 | 0.000840321 | 1654 | 19 |
| **positive regulation of RNA biosynthetic process** | GO:1902680 | 0.000855668 | 1656 | 19 |
| **dosage compensation by inactivation of X chromosome** | GO:0009048 | 0.001373953 | 24 | 4 |
| **chromatin organization** | GO:0006325 | 0.001435657 | 548 | 11 |
| **positive regulation of macromolecule biosynthetic process** | GO:0010557 | 0.001502981 | 1893 | 20 |
| **dosage compensation** | GO:0007549 | 0.001924257 | 26 | 4 |
| **regulation of gene expression, epigenetic** | GO:0040029 | 0.003673809 | 124 | 6 |
| **long-chain fatty acid metabolic process** | GO:0001676 | 0.004032546 | 126 | 6 |

| | | | | |
|---|---|---|---|---|
| **protein monoubiquitination** | GO:0006513 | 0.004609895 | 72 | 5 |
| **monocarboxylic acid metabolic process** | GO:0032787 | 0.006574609 | 642 | 11 |
| **icosanoid metabolic process** | GO:0006690 | 0.009459448 | 146 | 6 |
| **unsaturated fatty acid metabolic process** | GO:0033559 | 0.009459448 | 146 | 6 |

**Table S3.** Significant GO term enrichments of the candidate genes experienced hard sweep in Orkney voles. The threshold of p value was 0.05.

| Term name | Term ID | Adjusted *p*-value | Term size | Intersection size |
|---|---|---|---|---|
| immune response | GO:0006955 | 5.07E-11 | 1794 | 17 |
| antigen processing and presentation of peptide antigen | GO:0048002 | 5.17E-08 | 71 | 6 |
| antigen processing and presentation of exogenous peptide antigen | GO:0002478 | 7.13E-08 | 30 | 5 |
| defense response to protozoan | GO:0042832 | 1.18E-07 | 33 | 5 |
| response to protozoan | GO:0001562 | 1.88E-07 | 36 | 5 |
| antigen processing and presentation of exogenous antigen | GO:0019884 | 2.17E-07 | 37 | 5 |
| antigen processing and presentation | GO:0019882 | 9.88E-07 | 115 | 6 |
| antigen processing and presentation of exogenous peptide antigen via MHC class II | GO:0019886 | 2.03E-06 | 18 | 4 |
| antigen processing and presentation of peptide antigen via MHC class II | GO:0002495 | 4.85E-06 | 22 | 4 |
| antigen processing and presentation of peptide or polysaccharide antigen via MHC class II | GO:0002504 | 7.03E-06 | 24 | 4 |
| response to other organism | GO:0051707 | 1.65E-05 | 1705 | 12 |
| response to external biotic stimulus | GO:0043207 | 1.69E-05 | 1709 | 12 |
| response to biotic stimulus | GO:0009607 | 2.17E-05 | 1748 | 12 |
| biological process involved in interspecies interaction between organisms | GO:0044419 | 4.07E-05 | 1851 | 12 |
| defense response to other organism | GO:0098542 | 5.24E-05 | 1162 | 10 |
| MHC class II protein complex assembly | GO:0002399 | 7.23E-05 | 9 | 3 |
| peptide antigen assembly with MHC class II protein complex | GO:0002503 | 7.23E-05 | 9 | 3 |
| defense response to Gram-positive bacterium | GO:0050830 | 1.15E-04 | 126 | 5 |
| peptide antigen assembly with MHC protein complex | GO:0002501 | 1.42E-04 | 11 | 3 |
| positive regulation of immune system process | GO:0002684 | 1.59E-04 | 978 | 9 |
| MHC protein complex assembly | GO:0002396 | 1.89E-04 | 12 | 3 |
| leukocyte activation | GO:0045321 | 2.53E-04 | 1034 | 9 |

| | | | | |
|---|---|---|---|---|
| response to interferon-gamma | GO:0034341 | 2.75E-04 | 150 | 5 |
| T cell activation | GO:0042110 | 3.77E-04 | 518 | 7 |
| regulation of immune system process | GO:0002682 | 5.00E-04 | 1483 | 10 |
| adaptive immune response | GO:0002250 | 5.37E-04 | 546 | 7 |
| adhesion of symbiont to host | GO:0044406 | 5.82E-04 | 17 | 3 |
| cell activation | GO:0001775 | 6.00E-04 | 1146 | 9 |
| lymphocyte activation | GO:0046649 | 8.29E-04 | 862 | 8 |
| regulation of immune response | GO:0050776 | 1.25E-03 | 911 | 8 |
| positive regulation of T cell activation | GO:0050870 | 1.45E-03 | 210 | 5 |
| defense response | GO:0006952 | 1.83E-03 | 1711 | 10 |
| positive regulation of leukocyte cell-cell adhesion | GO:1903039 | 2.32E-03 | 231 | 5 |
| response to bacterium | GO:0009617 | 3.10E-03 | 1029 | 8 |
| biological adhesion | GO:0022610 | 3.58E-03 | 1423 | 9 |
| cellular response to interferon-gamma | GO:0071346 | 5.48E-03 | 122 | 4 |
| positive regulation of cell-cell adhesion | GO:0022409 | 5.53E-03 | 276 | 5 |
| T cell differentiation | GO:0030217 | 6.03E-03 | 281 | 5 |
| antigen processing and presentation of endogenous peptide antigen | GO:0002483 | 7.14E-03 | 38 | 3 |
| antigen processing and presentation of endogenous antigen | GO:0019883 | 7.73E-03 | 39 | 3 |

Chapter 3

# Sex-specific patterns of genetic diversity and mutation load in a small mammal

Xuejing Wang[1], Gerald Heckel[1,2]

[1]Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland; [2]Swiss Institute of Bioinformatics, Quartier Sorge - Batiment Amphipole, 1015 Lausanne, Switzerland

*Gerald Heckel

Email: gerald.heckel@unibe.ch

## Abstract

The sex chromosomes have evolutionary paths and population genetic features different from autosomes but empirical data from non-model species are still rare. Here we compared patterns of genetic diversity between sex chromosomes and autosomes based on the first draft assembly of the sex chromosomes of the common vole (*Microtus arvalis*). The X chromosome of *M. arvalis* is highly conserved compared to the X of mouse, while the Y chromosome shows much more divergence. We assessed the extent of potential sex-specific differences in patterns of genetic variation in natural populations without or with severe historical bottleneck followed by long-lasting absence of gene flow. We based this on genome-wide resequencing data of common voles from the European continent and their conspecifics on the Orkney islands in northern Scotland which were isolated there for over 5,000 years. Genetic diversity and heterozygosity of the X chromosome were low compared to autosomes even in the continental individuals. This indicates that partial haploidy and recombination together promote efficient background selection. The X chromosome also showed stronger evidence of purging of deleterious alleles compared to the autosomes. In contrast, potential deleterious alleles on the Y chromosome remained at low frequency while many neutral mutations drifted to fixation. The genetic patterns of the Y chromosome indicated strong selection on haploid mutations in Orkney voles, and fast accumulation of divergence due to a male-biased mutation rate. Taken together this shows that a prolonged historical bottleneck and long-term isolation can be followed by distinct genetic dynamics of

autosomes and sex chromosomes.

**Keywords**: genome assembly; sex chromosome; mutation load; *Microtus arvalis*; rodent

## Introduction

The evolution of sex chromosomes is on a disparate path comparing to the autosomes. In the classic model of (Charlesworth 1991), sex chromosomes start as autosomes, by accumulating sexually antagonistic mutations close to the sex-determining gene, which leads to local loss of recombination. The non-recombining region loaded with deleterious mutations tends to degenerate, and reduce the length of e.g. the Y chromosome, which can result in highly differentiated sex chromosomes with very short homologous regions left (Vicoso 2019), for example in mice (Soh, et al. 2014) and humans (Lahn and Page 1999). In diploid species with heteromorphic XY chromosomes, three ploidy systems exist, which are expected to differentiate the population genetic features of the chromosomes (Laporte and Charlesworth 2002; Immler 2019). First, autosomes and the pseudoautosomal regions of sex chromosomes going through recombination each generation behave as typical diploid alleles. Second, the non-recombining region of the Y chromosome behaves as a haploid allele and has $N_e$ equal to 1/4 of the autosomes when assuming a sex ratio of 1. Last, the X chromosome is diploid in the females and haploid in the males, with $N_e$ equal to 3/4 of the autosomes. In many species the mutation rate is higher in males (Hurst and Ellegren 1998; Bergeron, et al. 2023), which may lead to stronger differentiation on the population genetic dynamics of sex chromosomes and autosomes.

Deleterious mutations are common in the genome (Lohmueller 2014) as most new mutations which change the amino acid have a negative fitness effect (Eyre-Walker and Keightley 2007). In large populations, recessive deleterious mutations are usually in haploid form, and stay at low frequency due to effective selection (Kirkpatrick and Jarne 2000; Bertorelle, et al. 2022). In small populations, deleterious variants can drift to high frequency and are more likely to be homozygous, resulting in a loss of fitness which is called “mutation load” (Lynch, et al. 1995). It is thus important to estimate the accumulation state of deleterious mutations and their relationship to fitness in conservation genetics (DeWoody, et al. 2021; Kardos, et al. 2021). Recent studies showed that genetic drift and purging of deleterious mutations together shape the distribution of deleterious mutations in the genome (Robinson, et al. 2018; Aris-Brosou 2019; Grossen, et al. 2020; Chapter 1). However, most studies focused on the autosomes likely due to the difficulty and unavailability of assemblies of the sex chromosomes (Tomaszkiewicz, et al. 2017). In haploid individuals, every

deleterious mutation is at dominant state, which leads to stronger purging of deleterious alleles than in diploid systems (Valero, et al. 1992; Szövényi, et al. 2014). Deleterious mutations on sex chromosomes should thus be exposed to stronger selection because of certain degrees of haploidy. As selection acts at individual, rather than chromosomal level, it is necessary to include sex chromosomes when studying the adaptive process of the species.

The vole genus *Microtus* (Cricetidae) is a group of small rodents found in Europe, northern Asia, and North America. The sex chromosomes of *Microtus* have high variation, including typical short Y chromosomes (e.g. *M. arvalis*, Lemskaya, et al. 2010), X and Y chromosomes longer than any autosome (e.g. *M. levis*, Lemskaya, et al. 2010), intraspecific karyotypic diversity (e.g. *M. thomasi*, Rovatsos, et al. 2017), and an X0/XY sex determination system (*M. oregoni*, Charlesworth and Dempsey 2001). Such diversity is exceptional in mammals, yet so far there is only one chromosomal-level assembly of sex chromosomes of *Microtus* species, *M. oregoni* (Couger, et al. 2021).

To gain more insights into the evolution of genetic variation on sex chromosomes of *Microtus*, we focused on the common vole (*M. arvalis*). It is widespread in Europe with distinct genetic lineages associated with the geographic distribution (Heckel, et al. 2005). The karyotype of *M. arvalis* (2n=46) is formed by 22 pairs of autosomes, an intermediate sized X chromosome and a short Y chromosome, with no synapsis observed between the X and Y chromosomes (Ashley, et al. 1990). The existing genome assembly of *M. arvalis*, produced with Hi-C scaffolding (Belton, et al. 2012), contains 22 chromosome-level scaffolds representing the autosomes, yet no sex chromosome was assembled (Gouy et al. in submission). For the *de novo* sequencing, a depth filter higher than half of the average depth was used for the male sample to acquire a high-quality assembly, and the sex chromosomes were thus discarded.

The common vole is also widely distributed on the Orkney archipelago in the north of Scotland with a high census population estimated to be one million (Reynolds 1992). Previous genetic and archeological studies showed that it was likely introduced by Neolithic farmers as a food resource (Haynes, et al. 2003; Romaniuk, et al. 2016) over 5,000 years ago (Martínková, et al. 2013). The common voles were most likely introduced from the coast of Belgium and northern France (Martínková, et al. 2013) and went through a severe bottleneck (Chapter 1). The decrease of effective population size ($N_e$) was estimated to be ~30 fold, from over 100 thousand to several thousands, and recovered back to 50-100 thousand in the recent ~1,000

years (Wang, et al. in press). The prolonged severe bottleneck, though with high estimated $N_e$, led to strong genetic drift, which has left remarkable impacts in the autosomal genome of Orkney voles, including low genetic diversity and accumulation of derived homozygous deleterious alleles (Chapter 1; Chapter 2). Thus, the Orkney vole can serve as a system to observe the change of genetic variation on sex chromosomes through a bottleneck given its relatively recent divergence time and demographic changes compared to the time scales of speciation in many vertebrates of usually millions of years (Sepkoski 1998; Rabosky, et al. 2013).

In this study we provide a draft assembly of the sex chromosomes of the common vole. With the new reference sequences, we compared the genetic diversity and distribution of potential deleterious mutations on the autosomes and sex chromosomes in populations without or with an extended bottleneck and long absence of gene flow. We hypothesized that the genetic diversity on sex chromosomes was not at the expected value, 3/4 or 1/4 of that of autosomes, because of sex-biased mutation rate and stronger purifying selection during haploid phase. We also expected that purging of deleterious mutations on sex chromosomes was stronger than on autosomes in the Orkney population because of the bottleneck.

## Results

### The assembly of sex chromosomes

We assembled both sex chromosomes from a heterochromatic (male) sample sequenced with the regular Hi-C method (Belton, et al. 2012) using a pipeline of two rounds of assembling (see Methods) simplified from Peichel, et al. (2020). The final assembly of sex chromosomes is 96.7 Mb (Details in Supplementary Information). The X chromosome was assembled to one scaffold of 88.4 Mb in the first round with an N-rate of 8.7%. It showed strong synteny only with the X chromosome of mouse in two sequential blocks (Figure 1, S1), which confirmed that it was X chromosomal sequence of *M. arvalis*. In the second round, three scaffolds, 8.3 Mb in total, were assembled and verified as the draft assembly of the Y chromosome. The synteny maps showed short syntenic regions with Chr7, Chr11, X and Y chromosomes of mouse (Figure 1, S1). The shortest 1.0 Mb scaffold had regions of synteny to the Y chromosome of mouse, and most sequences assembled with male-specific k-mers were

mapped to it (29 sequences, 29,167 bp in total).

On the scaffold of the X chromosome, 1190 genes were predicted, 995 of which were homologues of mouse genes, and 707 were homologous with genes on the mouse X chromosome. On the scaffolds of the Y chromosome, 339 genes were predicted, 285 were homologous with mouse genes, and 5 were homologous with genes on mouse Y chromosome (Uba1y, Ddx3y, Zfy1, Kdm5d and Eif2s3y).

**Distinct levels of genetic variation between autosomes and sex chromosomes**

The genetic diversity of the X chromosome, measured as nucleotide diversity π, was much lower than for autosomes (Figure 2, Table S3). The mean values of autosomal π in 50 kb windows of both continental (samples from Belgium and northern France which are genetically closest to Orkney voles, $1.83 \times 10^{-3}$) and Orkney voles ($4.80 \times 10^{-4}$) were roughly 3 times higher than π of the X chromosome (continent: $6.29 \times 10^{-4}$, Orkney: $1.76 \times 10^{-4}$). The heterozygosity, measured in the proportion of heterozygous sites, was also lower on X chromosome than autosomes (continent: 33% of autosomes, Orkney: 41% of autosomes; Figure 3). On the other hand, the Y chromosome which is expected to have 1/4 $N_e$ of the autosomes, maintained relatively high genetic diversity (mean π continent: $1.42 \times 10^{-3}$, Orkney: $3.38 \times 10^{-4}$) equivalent to over 70% of the autosomes (Figure 2). The median π of the Y chromosome of Orkney voles ($2.42 \times 10^{-4}$) was even higher than of the autosomes ($2.31 \times 10^{-4}$), indicating that genetic variation was maintained in the haploid Y chromosome.

The X chromosome carried significantly less derived alleles of all functional categories in both continental and Orkney voles (Figure 3; t-test, all $p < 10^{-10}$). The deviation was highest for the number of loss-of-function (LOF) alleles, with the X chromosome carrying only half as many LOF alleles per 1 Mb of coding regions compared to autosomes. For the Y chromosome, the proportion of derived neutral and missense alleles was equivalent to autosomes, while the proportion of derived LOF and synonymous alleles were lower than for autosomes (t-test; all $p < 0.05$).

**Impacts of the extensive bottleneck on sex chromosomes**

Genetic diversity was lower over the genome in bottlenecked Orkney voles compared to their continental conspecifics. The density distribution of π in Orkney voles is shifted towards the right, with the deviation between median and mean values increased (Figure 2,

Table S3). On autosomes and the X chromosome, the heterozygosity of Orkney voles was lower compared to continental voles in all functional categories (Figure 3). The total number of derived alleles was slightly higher in Orkney voles ($p < 5 \times 10^{-8}$), while the number of derived alleles on sex chromosomes remained at the same level, except for a significantly higher proportion of intergenic (neutral) derived alleles on the Y chromosome (Figure 3).

The sex chromosomes showed distinct patterns in the fixation of derived LOF mutations in Orkney voles compared to the autosomes. While the LOF mutations fixed at a similarly high level (39%) as neutral mutations (41%) on the autosomes, the site frequency spectrum (SFS) of LOF mutations on the Y chromosome strongly deviated from neutral mutations towards lower site frequency (Figure 4). The X chromosome showed a deviation at moderate level, characterized by an increased proportion of singletons of LOF mutations (26%) compared to neutral mutations (13%) and a decreased fixation ratio (53% to 47%).

The level of accumulation of potential deleterious mutations in the coding region (LOF, missense and synonymous alleles) on the sex chromosomes were lower compared to the autosomes in Orkney voles with $R^{a}_{XY}$ lower than 1 (Figure 5a). When compared to the neutral mutations, the X chromosome accumulated missense and synonymous mutations, while Y chromosome showed evidence of strong purging against all the mutations within the coding region (Figure 5b).

## Discussion

### The assembly of the sex chromosomes of M. *arvalis*

In this study, we provided a draft assembly of the sex chromosomes of the common vole, with the X at chromosomal level and Y in long scaffolds. However, the quality of our assembly of sex chromosomes is lower than the previous assembly of the autosomes. The N-rate of the autosomal assembly was 3.3%, much lower than for both X (8.7%) and Y (12.6%) chromosomes. This is likely because certain reads from the sex chromosomes mapped to the autosomes and were not used in the assembly, which led to longer gaps when scaffolding.

The homologous pseudoautosomal regions cause difficulties during assembling sex chromosomes in many species (Tomaszkiewicz, et al. 2017). This effect is subtle in our case as the sex chromosomes of M. *arvalis* share likely, if any, very short homologous regions. They are highly divergent (Mayorov, et al. 1996) with distinct size difference, which indicates ancient divergence. They have also no observable physical contact at all during meiosis which is rare, especially for typical short Y and long X chromosomes without accumulation of heterochromatin (Ashley, et al. 1990). Thus, the error rate of mismatches in the homologous regions should be low during scaffolding. However, our methods can only assemble the nonrecombining regions of the Y chromosome, as the read depth of pseudoautosomal regions is expected to be at the same level of autosomes and they are assembled at the same stage with autosomes. Thus, these methods are not suitable for species with largely homogeneous sex chromosomes, for example many fish and frogs (Ross, et al. 2009; Jeffries, et al. 2018).

The X chromosome of *M. arvalis* is strongly homologous to mouse X chromosome (Figure 1), which was expected as the X chromosome is highly conserved in most mammals (Rodríguez Delgado, et al. 2009; Livernois, et al. 2012). The Y scaffolds, on the other hand, share only one short homologous region harboring five genes, four of which (Uba1y, Zfy1, Kdm5d and Eif2s3y) are among the most conserved MSY (male-specific region of the Y chromosome) genes (Li, et al. 2013). This is likely partially due to the fragmentation of the assembly, but also due to the fast evolution of sex chromosomes in *Microtus*, indicated by highly divergent karyotypes in related species (Mayorov, et al. 1996; Lemskaya, et al. 2010). Genomic resources of *Microtus* are scarce and limited to few species at the moment (Rovatsos, et al. 2017; Lamelas, et al. 2018; Couger, et al. 2021). The diversity of sex chromosomes in *Microtus* (Lemskaya, et al. 2010), and lack of recombination between the sex

chromosomes in several species in Cricetidae (Ashley and Fredga 1994) also reveals great potential to study the evolution of the Y chromosome in this highly diverse rodent group. For example, comparative genomics between species could help to understand whether fast speciation is related to chromosomal rearrangements in this clade (Rieseberg 2001; Faria and Navarro 2010).

**Potential effects of male biased mutation rate**

The difference on the genetic diversity between sex chromosomes and autosomes (Figure 2, 3, Table S3) likely resulted from a higher mutation rate in males than in females that is frequently observed in mammals (Wilson Sayres and Makova 2011). The male mutation bias α (male mutation rate divided by female mutation rate) of most mammals ranges from 1 to 4, and the α of mouse and rat is estimated around 2 (Sayres, et al. 2011). The α of *M. arvalis* or closely-related species is unavailable. Taking the α of mouse at 2, and under the assumption of a sex ratio of 1, the mutation rate of Y is 1.5 times of X, and 1.3 times of the autosomes, and the expected genetic diversity of Y is 1/3 of the autosomes. This does not match the pattern observed in the continental population, where the genetic diversity of Y is 2/3 of the autosomes (Figure 2, Table S3). In the case of an equal sex ratio and neutral evolution, the genetic diversity of Y chromosome should not reach half of autosomes arithmetically. The causes for unexpectedly high genetic diversity on Y chromosome are unknown but possible factors include high α of *M. arvalis*, sex ratio biased towards male (Bryja, et al. 2005), and lack of purifying selection in the pseudogenes and repeated regions on the Y chromosome (Bachtrog 2013; Vicoso 2019). However, the exceptionally low genetic diversity of X (1/3 of the autosomes) and low heterozygosity (Figure 3) cannot be explained by a bias in mutation rates.

**Selection on haploid and diploid chromosomes**

Our results showed that the sex chromosomes likely experience stronger purifying and positive selection. In haploid systems, every mutation is dominant, while recessive mutations are sheltered in diploid systems (Valero, et al. 1992; Immler 2019). Deleterious or beneficial mutations on the Y chromosome are expected to be exposed to stronger selection than on the X, and on X chromosome stronger than on autosomes (Singh, et al. 2008). The U-shaped SFS of X and Y of the continental voles (Figure 4) is a typical feature of genomic regions under

a selective or neutral sweep (Huber, et al. 2016; Moinet, et al. 2022), indicating stronger linked selection than on the autosomes.

On the other hand, recombination, as an efficient way to remove deleterious alleles (Barton 1995; Comeron, et al. 1999), occurs only on X chromosomes in females, but ceased completely on the Y chromosome for *M. arvalis*. Thus, the low genetic diversity on the X chromosome (Figure 2, 3, Table S3) is likely caused by strong background selection empowered by both the exposure of deleterious mutations in the haploid phase and recombination in the diploid phase. The low burden of derived LOF mutations on X (Figure 3) also supports the hypothesis of stronger purifying selection on the X chromosome.

The ploidy may also play a role in the change of genetic variance in the face of strong drift caused by a severe bottleneck of ~30 fold population decrease for several thousand generations. The extant derived LOF mutations of Orkney voles drifted to fixation together with the neutral mutations (Figure 4). The fixation rate of LOF mutations on the Y chromosome was low, highly different from the rate of neutral mutations (Figure 4), showing that the efficacy of selection was still high on the haploid parts of the Y chromosome even with strong genetic drift (Otto, et al. 2015). This is supported by the $R'_{XY}$ ratios of LOF, missense and synonymous mutations lower than 1, indicating global purifying selection on deleterious alleles (Grossen, et al. 2020). The X chromosome, as a haploid-diploid chromosome, showed the intermediate shift of fixation rate between the autosomes and the Y chromosome (Figure 4). The $R^a_{XY}$ of X is lower than 1 for all mutations, likely due to the low genetic diversity compared to the autosomes leading to smaller genetic distance. Contrary to the Y chromosome, the $R'_{XY}$ ratios of missense and synonymous mutations of the X chromosome were higher than 1, showing potential accumulation of mildly deleterious mutations on X during the bottleneck (Do, et al. 2015). However, our result of stronger purifying selection on the Y chromosome is not in conflict with the classical theory of genetic degeneration of the Y chromosome resulting from the accumulation of loss-of-function mutations and repetitive DNA (Bachtrog 2013). First, mammalian Y chromosomes are at the late stage of evolution beyond massive degeneration, so the functional genes on the Y chromosomes are conserved and stable (Vicoso 2019). Second, our results reflect the genetic changes at population level within the species, while the degeneration of Y chromosome is usually only detectable at the interspecific level.

In summary, we showed that the autosomes, X and Y have distinct features of genetic

variation and reactions to the population bottleneck, likely affected by both male biased mutation rate and the difference in ploidy. So far, the focus of research of mutation load on sex chromosomes has been mainly on sex chromosome turnover (Blaser, et al. 2013; El Taher, et al. 2021). Only a limited number of studies has looked at individual and populational effects of deleterious mutations on sex chromosomes (Hollis and Houle 2011; Mallet, et al. 2011). As selection acts on individuals rather than individual chromosomes, it is necessary to investigate further how the mutation load on sex chromosomes influences the evolutionary process at population level.

## Material and Methods

### De novo assembly of sex chromosomes on the basis of assembled autosomes

Reads from the same sequencing libraries used for the reference genome of *M. arvalis* (only 22 autosomes assembled, BioProject ID: PRJNA737461, Gouy et al. in submission) were used for the assembly of sex chromosomes, including four libraries of 150 bp paired-end Illumina reads, long-range sequencing libraries produced with the “Chicago” (Putnam, et al. 2016) and Hi-C (Belton, et al. 2012) methods by Dovetail Genomics (Santa Cruz, CA). Reads from Illumina libraries were trimmed using Trimmomatic (Bolger, et al. 2014). Trimmed reads were mapped back to the reference genome. Since the depth of pseudo-autosomal and repeat regions of the Y chromosome can be as high as autosomes, some regions of sex chromosomes may already have been assembled. Read pairs not mapped to the 22 chromosome-sized scaffolds were extracted for the first step of *de novo* assembly. ABySS 2.1 (Simpson, et al. 2009) was used to assemble short reads to contigs with K-mer at 55 (estimated by Dovetail for the reference assembly). Only unmapped reads and their mates from Chicago and Hi-C libraries were used for scaffolding. Read pairs not mapped to the autosomal assembly were first mapped and filtered with HiCUP 0.7.4 (Wingett, et al. 2015). Next, 3D-DNA (Dudchenko, et al. 2017) was used for scaffolding using *de novo* assembled contigs which were longer than 1 kb, and filtered reads from Chicago and Hi-C libraries. Hi-C assembly was visualized, manually reviewed and corrected in Juicebox Assembly Tools 1.11 (Durand, et al. 2016).

The first round of scaffolding resulted in one long scaffold representing the X

chromosome (see the next section of Methods for the validation of sex chromosome assembly). According to the karyotype of *M. arvalis*, the length of the Y chromosome should be roughly 1/8 to 1/4 of the X chromosome (Lemskaya, et al. 2010), but the remaining scaffolds except the X chromosome were only 3.5 Mb, about 4% of X chromosome assembly. Therefor to acquire a more complete assembly of the Y chromosome, a second round of assembling was performed with the pipeline from Peichel, et al. (2020). The principle is to repeat the scaffolding of the Y chromosome with higher tolerance of repeated sequences and excluding contigs from the X chromosome from the first round.

For the second round of assembly, read pairs from Chicago and Hi-C libraries not mapped to either autosomes or X chromosome were used for scaffolding with the same parameters with HiCUP. Next, 3D-DNA was used for scaffolding using non-X chromosome contigs and filtered reads from Chicago and Hi-C libraries. The minimum length of the contigs used for scaffolding was 1 kb. Multiple trials were done using different thresholds of tandem repeat coverage (--editor-repeat-coverage = 2, 4, 6… 16) to test for the optimal repeat level allowed in scaffolding. The best assembly of the Y chromosome was chosen based on the length of the longest scaffold and the total length of the long scaffolds. The five longest scaffolds (>500 Mb) from the best assembly were kept for validation.

**Validation and annotation of the sex chromosome assembly**

Each assembly was compared with the mouse genome (GRCm39, GenBank accession: GCA_000001635.9) for synteny using Minimap2 (Li 2018) on the website of D-Genies (Cabanettes and Klopp 2018). The divergence of the Y chromosome between different mammal species is usually high because of its fast evolution. Thus, an extra k-mer based method was used to validate whether the assembly from the second round was part of the Y chromosome. Raw reads of 150 bp paired-end sequencing of 10 individuals, 5 males and 5 females, from the Central lineage of *M. arvalis* (Table S1, the same evolutionary lineage as the sample used for *de novo* assembly) were used to identify male specific k-mers with KmerGO 1.5.0 (Wang, et al. 2020). Note that these samples were not used for the following mutation load analyses because Central lineage is genetically distinct from Orkney voles and thus does not serve as a good comparison (Martínková, et al. 2013; Chapter 1). The k-mer length was set to 55 and the minimal K-mer occurring times was set to 4. The male specific k-mers were then used to assemble into sequences using Cap3 (Huang and Madan 1999), with

segment pair score cutoff (-i) of 30, chain score cutoff (-j) of 31, overlap length cutoff (-o) of 18, overlap similarity score cutoff (-s) of 300 and max number of word matches (-t) of 100. Assembled sequences > 1 kb were mapped to the potential Y chromosome assembly using Minimap2 on the website of D-Genies. Only the scaffolds having male-specific sequence mapped to them were considered as Y chromosome scaffolds.

To perform gene annotation on the new assembly of sex chromosomes, repeat regions were first identified with RepeatModeler 2.0.3 (Flynn, et al. 2020) and RepeatMasker 4.1.4 (Chen 2004). For annotation, the training of Augustus (Stanke and Waack 2003) was performed using the annotation of autosomes. Gene prediction was performed for the assembly of sex chromosomes with transposable elements (TE) hard-masked and simple repeats soft-masked using Augustus. The synteny map between the final assembly of *M. arvalis* sex chromosomes and mouse sex chromosomes were produced with the MCscan pipeline from JCVI 2.1.7 (Tang, et al. 2008) using coding sequence (CDS) regions.

**Mutation load on different chromosomes**

Samples from the Western evolutionary lineage (northern France and Belgium) and Orkney were used to compare genetic variation between the autosomes and sex chromosomes, and to investigate the change of genetic variation after a severe historical bottleneck. Tissue samples preserved in absolute ethanol were used for DNA extraction with the phenol-chloroform method. The quality and concentration of the DNA extraction were checked with Qubit fluorometer (Life Technologies) and NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific), and on 1% agarose gels. The library preparation with Illumina TruSeq DNA PCR-Free Library Prep Kit, and sequencing on Illumina Hiseq 2000 or Novaseq 6000 were performed by the NGS platform of the University of Bern.

The resequencing data of 11 samples from Western lineage (referred as continent, 4 males and 7 females) and 16 Orkney voles (8 males and 8 females) were mapped to the reference genome of autosomes and newly assembled sex chromosomes. The mapping was performed with BWA 0.7.17 MEM algorithm (Li and Durbin 2009). The duplicated reads were filtered with GATK 4.0.10 (Poplin, et al. 2017). Then we performed SNP calling with GATK Haplotypecaller following the GVCF pipeline (Poplin, et al. 2017). For male samples, the ploidy of sex chromosomes was set to 1 so that only haploid SNPs were called. To acquire reliable SNP data, we filtered the SNP data with GATK VariantFiltration and VCFtools 0.1.16 (Danecek,

et al. 2011) with following parameters: QD<10.0, SOR>3.0, FS>60.0, MQ<40.0, MQRankSum<-12.5, ReadPosRankSum<-8.0, GQ<20, heterozygosity=100% (not used for male sex chromosomes), mean-depth over all individuals>50 (>25 for male sex chromosomes), individual genotypes <5 (<3 for male sex chromosomes) or >150 (>75 for male sex chromosomes). Only SNPs with no missing data were used in further analyses. The site frequency of each SNP for either continental or Orkney samples was calculated with VCFtools. The genetic diversity π for each site was calculated from site frequencies in R. The average π in 50 kb windows was calculated with BEDTools 2.28 (Quinlan and Hall 2010). The functional effect of each SNP was annotated with SnpEff 5.0 (Cingolani, et al. 2012). We compared the number and distribution of loss-of-function (LOF), missense, synonymous and intergenic variants (5 kb away from any gene, considered as neutral variants) as a gradient from highly deleterious to neutral mutations. The number of SNPs in heterozygous or homozygous state was counted per individual, averaged within each group, and controlled both by ploidy and by the length of CDS or intergenic region. The unfolded site frequency spectra (SFS) of LOF and neutral SNPs were also reported. We used LOF mutations as a proximate of deleterious mutations, however the real distribution of fitness effect can be wide, from nearly neutral to lethal (Eyre-Walker and Keightley 2007).

We calculated the rate of private alleles $R_{XY}$ (Do, et al. 2015) of Orkney voles relative to the continental group. For autosomes, X chromosome or Y chromosome, $R_{XY}$ of highly deleterious, missense, and synonymous SNPs were calculated and divided by $R_{XY}$ of neutral SNPs of autosomes to acquire the standardized $R^{a}_{XY}$ in order to reveal the state of accumulation or purging related to the genomic background. We also reported standardized $R'_{XY}$ controlled with the $R_{XY}$ of neutral SNPs from each chromosome(s), to show the state of functional mutations compared to the chromosomal background. The standard error was estimated by jackknifing of 100 blocks of equal number of SNPs for autosomes, or 20 blocks for sex chromosomes.

## Acknowledgements

We thank Zuyao Liu for advice on sex chromosome assembly and annotation. We thank Stephan Peischl and Stefan Strütt for comments on the results. We thank the Next Generation Sequencing Platform of the University of Bern for sequencing services. This study was supported by the Swiss National Science Foundation grant 31003A_176209. XW received a scholarship for PhD study from the China Scholarship Council (No. 201706380049).

## References

Aris-Brosou S. 2019. Direct evidence of an increasing mutational load in humans. Molecular Biology and Evolution 36:2823-2829.

Ashley T, Fredga K. 1994. The curious normality of the synaptic association between the sex chromosomes of two arvicoline rodents: *Microtus oeconomus* and *Clethrionomys glareolus*. Hereditas 120:105-111.

Ashley T, Jaarola M, Fredga K. 1990. Absence of synapsis during pachynema of the normal sized sex chromosomes of *Microtus arvalis*. Hereditas 111:295-304.

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nature Reviews Genetics 14:113-124.

Barton N. 1995. A general model for the evolution of recombination. Genetics Research 65:123-144.

Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods 58:268-276.

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St. Leger J, Shao C, et al. 2023. Evolution of the germline mutation rate across vertebrates. Nature 615:285-291.

Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, Morales HE, Van Oosterhout C. 2022. Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics 23:1-12.

Blaser O, Grossen C, Neuenschwander S, Perrin N. 2013. Sex-chromosome turnovers induced by deleterious mutation load. Evolution 67:635-645.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30:2114-2120.

Bryja J, Nesvadbová J, Heroldová M, Jánová E, Losík J, Trebatická L, Tkadlec E. 2005. Common vole (*Microtus arvalis*) population sex ratio: biases and process variation. Canadian Journal of Zoology 83:1391-1399.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6:e4958.

Charlesworth B. 1991. The evolution of sex chromosomes. Science 251:1030-1033.

Charlesworth B, Dempsey ND. 2001. A model of the evolution of the unusual sex chromosome system of *Microtus oregoni*. Heredity 86:387-394.

Chen N. 2004. Using Repeat Masker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics 5:4.10. 11-14.10. 14.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6:80-92.

Comeron JM, Kreitman M, Aguadé M. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics 151:239-249.

Couger MB, Roy SW, Anderson N, Gozashti L, Pirro S, Millward LS, Kim M, Kilburn D, Liu KJ, Wilson TM. 2021. Sex chromosome transformation and the origin of a male-specific X chromosome in the creeping

vole. Science 372:592-600.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

DeWoody JA, Harder AM, Mathur S, Willoughby JR. 2021. The long-standing significance of genetic diversity in conservation. Molecular Ecology 30:4147-4154.

Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D. 2015. No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. Nature Genetics 47:126-131.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356:92-95.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell systems 3:99-101.

El Taher A, Ronco F, Matschiner M, Salzburger W, Böhne A. 2021. Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. Science Advances 7:eabe8215.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:610-618.

Faria R, Navarro A. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends in Ecology & Evolution 25:660-669.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences 117:9451-9457.

Grossen C, Guillaume F, Keller LF, Croll D. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nature communications 11:1001.

Haynes S, Jaarola M, Searle JB. 2003. Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. Molecular Ecology 12:951-956.

Heckel G, Burri R, Fink S, Desmet J-F, Excoffier L. 2005. Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. Evolution 59:2231-2242.

Hollis B, Houle D. 2011. Populations with elevated mutation load do not benefit from the operation of sexual selection. Journal of evolutionary biology 24:1918-1926.

Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome research 9:868-877.

Huber CD, DeGiorgio M, Hellmann I, Nielsen R. 2016. Detecting recent selective sweeps while controlling for mutation rate and background selection. Molecular Ecology 25:142-156.

Hurst LD, Ellegren H. 1998. Sex biases in the mutation rate. Trends in Genetics 14:446-452.

Immler S. 2019. Haploid selection in "diploid" organisms. Annual Review of Ecology, Evolution, and Systematics 50:219-236.

Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, Borzée A, Barrow LN, Canestrelli D, Crochet P-A, Dufresnes C. 2018. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nature communications 9:1-11.

Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM, Tallmon DA, Funk WC. 2021. The crucial role of genome-wide genetic variation in conservation. Proceedings of the National Academy of Sciences 118:e2104642118.

Kirkpatrick M, Jarne P. 2000. The effects of a bottleneck on inbreeding depression and the genetic load. The American Naturalist 155:154-167.

Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. Science 286:964-967.

Lamelas L, Arroyo M, Fernández FJ, Marchal JA, Sánchez A. 2018. Structural and evolutionary relationships in the giant sex chromosomes of three *Microtus* species. Genes 9:27.

Laporte V, Charlesworth B. 2002. Effective population size and population subdivision in demographically structured populations. Genetics 162:501-519.

Lemskaya NA, Romanenko SA, Golenishchev FN, Rubtsova NV, Sablina OV, Serdukova NA, O'Brien PC, Fu B, Yiğit N, Ferguson-Smith MA. 2010. Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). III. Karyotype relationships of ten *Microtus* species. Chromosome research 18:459-471.

Li G, Davis BW, Raudsepp T, Wilkerson AJP, Mason VC, Ferguson-Smith M, O'Brien PC, Waters PD, Murphy WJ. 2013. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome research 23:1486-1495.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760.

Livernois AM, Graves JAM, Waters PD. 2012. The origin and evolution of vertebrate sex chromosomes and dosage compensation. Heredity 108:50-58.

Lohmueller KE. 2014. The distribution of deleterious genetic variation in human populations. Current opinion in genetics & development 29:139-146.

Lynch M, Conery J, Burger R. 1995. Mutation accumulation and the extinction of small populations. The American Naturalist 146:489-518.

Mallet MA, Bouchard JM, Kimber CM, Chippindale AK. 2011. Experimental mutation-accumulation on the X chromosome of *Drosophila melanogaster* reveals stronger selection on males than females. BMC Evolutionary Biology 11:1-12.

Martínková N, Barnett R, Cucchi T, Struchen R, Pascal M, Pascal M, Fischer MC, Higham T, Brace S, Ho SYW, et al. 2013. Divergent evolutionary processes associated with colonization of offshore islands. Molecular Ecology 22:5205-5220.

Mayorov V, Adkison L, Vorobyeva N, Khrapov E, Kholodhov N, Rogozin I, Nesterova T, Protopopov A, Sablina O, Graphodatsky A. 1996. Organization and chromosomal localization of a B1-like containing repeat of *Microtus subarvalis*. Mammalian Genome 7:593-597.

Moinet A, Schlichta F, Peischl S, Excoffier L. 2022. Strong neutral sweeps occurring during a population contraction. Genetics 220.

Otto SP, Scott MF, Immler S. 2015. Evolution of haploid selection in predominantly diploid organisms. Proceedings of the National Academy of Sciences 112:15952-15957.

Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM, Kingsley DM, et al. 2020. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome biology 21:177.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv:201178.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome research 26:342-350.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, Alfaro ME. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature communications 4:1958.

Reynolds P. 1992. The impact of changes in land-use in Orkney, on the vole *Microtus arvalis orcadensis* and its avian predators. University of Aberdeen.

Rieseberg LH. 2001. Chromosomal rearrangements and speciation. Trends in Ecology & Evolution 16:351-358.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK. 2018. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Current Biology 28:3487-3494.e3484.

Rodríguez Delgado CL, Waters PD, Gilbert C, Robinson TJ, Graves JAM. 2009. Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. Chromosome research 17:917-926.

Romaniuk AA, Shepherd AN, Clarke DV, Sheridan AJ, Fraser S, Bartosiewicz L, Herman JS. 2016. Rodents: food or pests in Neolithic Orkney. Royal Society open science 3:160514.

Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). PLOS Genetics 5:e1000391.

Rovatsos MT, Marchal JA, Romero-Fernández I, Arroyo M, Athanasopoulou EB, Sánchez A. 2017. Extensive Sex Chromosome Polymorphism of *Microtus thomasi/Microtus atticus* Species Complex Associated with Cryptic Chromosomal Rearrangements and Independent Accumulation of Heterochromatin. Cytogenetic and Genome Research 151:198-207.

Sayres MAW, Venditti C, Pagel M, Makova KD. 2011. Do variations in substitution rates and male mutation bias correlate with life-history traits? a study of 32 mammalian genomes. Evolution 65:2800-2815.

Sepkoski JJ. 1998. Rates of speciation in the fossil record. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 353:315-326.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. Genome Research 19:1117-1123.

Singh ND, Larracuente AM, Clark AG. 2008. Contrasting the efficacy of selection on the X and autosomes in *Drosophila*. Molecular Biology and Evolution 25:454-467.

Soh YS, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, Fulton RS, Kremitzki C, Koutseva N, Mueller JL. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159:800-813.

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19:ii215-225.

Szövényi P, Devos N, Weston DJ, Yang X, Hock Z, Shaw JA, Shimizu KK, McDaniel SF, Wagner A. 2014. Efficient purging of deleterious mutations in plants with haploid selfing. Genome biology and evolution 6:1238-1252.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and Collinearity in Plant Genomes. Science 320:486-488.

Tomaszkiewicz M, Medvedev P, Makova KD. 2017. Y and W chromosome assemblies: approaches and discoveries. Trends in Genetics 33:266-282.

Valero M, Richerd S, Perrot V, Destombe C. 1992. Evolution of alternation of haploid and diploid phases in life cycles. Trends in Ecology & Evolution 7:25-29.

Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nature Ecology & Evolution 3:1632-1641.

Wang Y, Chen Q, Deng C, Zheng Y, Sun F. 2020. KmerGO: a tool to identify group-specific sequences with k-mers. Frontiers in microbiology 11:2067.

Wilson Sayres MA, Makova KD. 2011. Genome analyses substantiate male mutation bias in many species. Bioessays 33:938-945.

Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. 2015. HiCUP: pipeline for mapping and processing Hi-C data. F1000Research 4:1310.

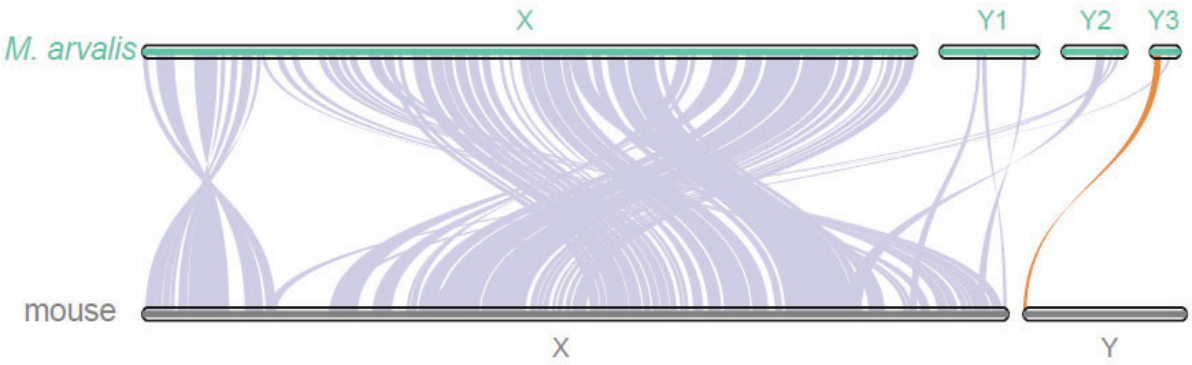



**Figure 1**. Synteny map between the draft assemblies of the sex chromosomes of *M. arvalis* and mouse based on homologous coding sequence (CDS) regions. The X chromosomes of *M. arvalis* and of mouse had clear and strong synteny while the Y chromosomes had little. The orange line represents a homologous region between one of the Y scaffolds (Y3) of *M. arvalis* and mouse Y chromosome.

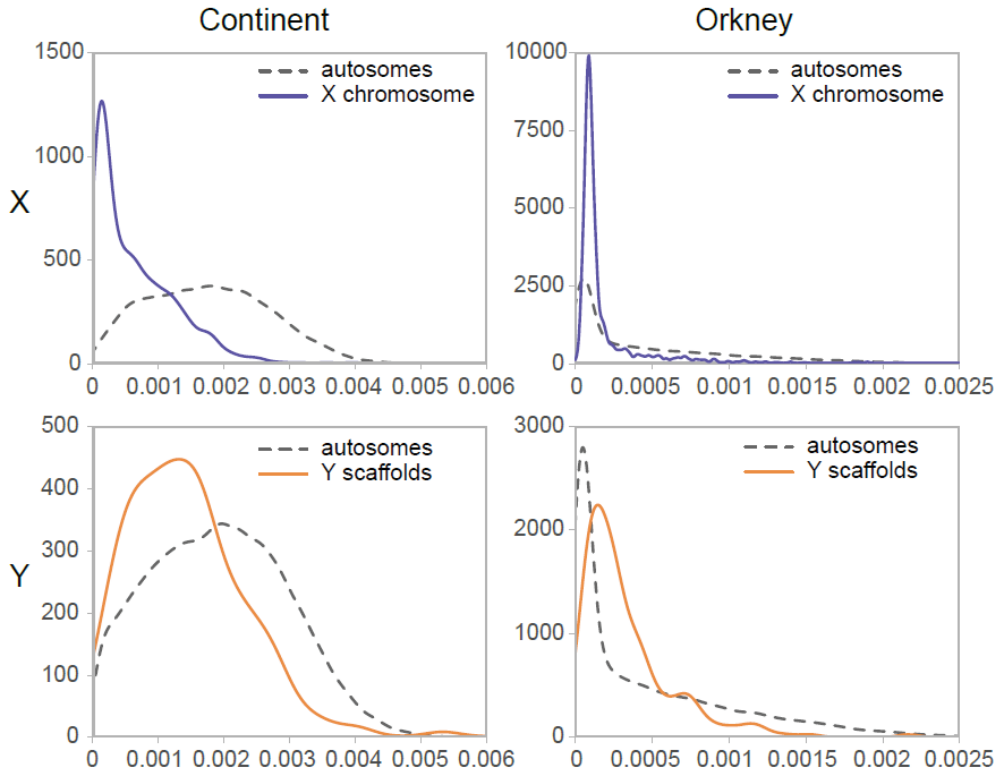



**Figure 2.** The density distributions of π measured in 50-Kb windows for autosomes and sex chromosomes with π on the x-axis and density on the y-axis. Only female samples were used for the X panels (both autosomes and X chromosome) and only males for the Y panels. The genetic diversity of sex chromosomes of continental voles was lower than for autosomes. Y scaffolds had higher diversity than the X chromosome despite lower $N_e$. The Orkney voles experienced genome-wide decrease of π, while the genetic diversity of Y scaffolds remained at a level similar to the autosomes.

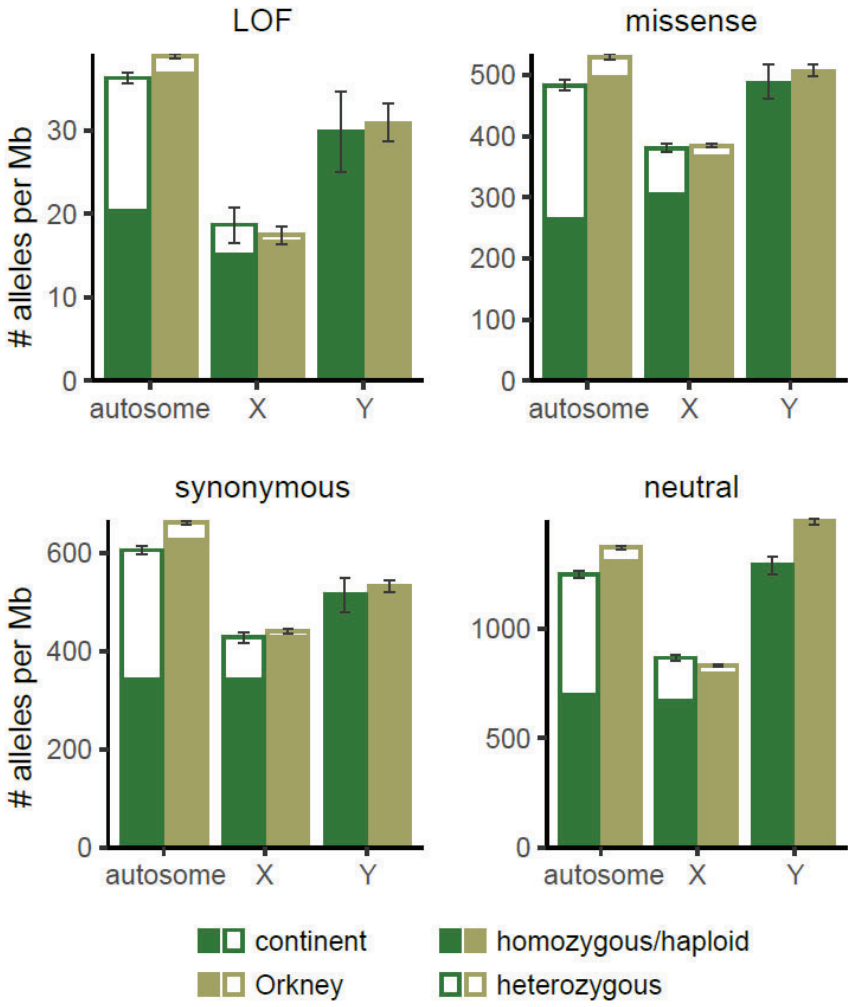



**Figure 3.** Numbers of derived alleles of different functional impacts per Mb per copy on different chromosomes. Numbers of loss-of-function (LOF), missense and synonymous mutations were standardized by division with the length of CDS, and the number of neutral mutations with the length of intergenic regions (5 kb away from any gene). Only female samples were used for X and only males for Y chromosome analyses. The error bar shows the standard error of the number of alleles per individual.

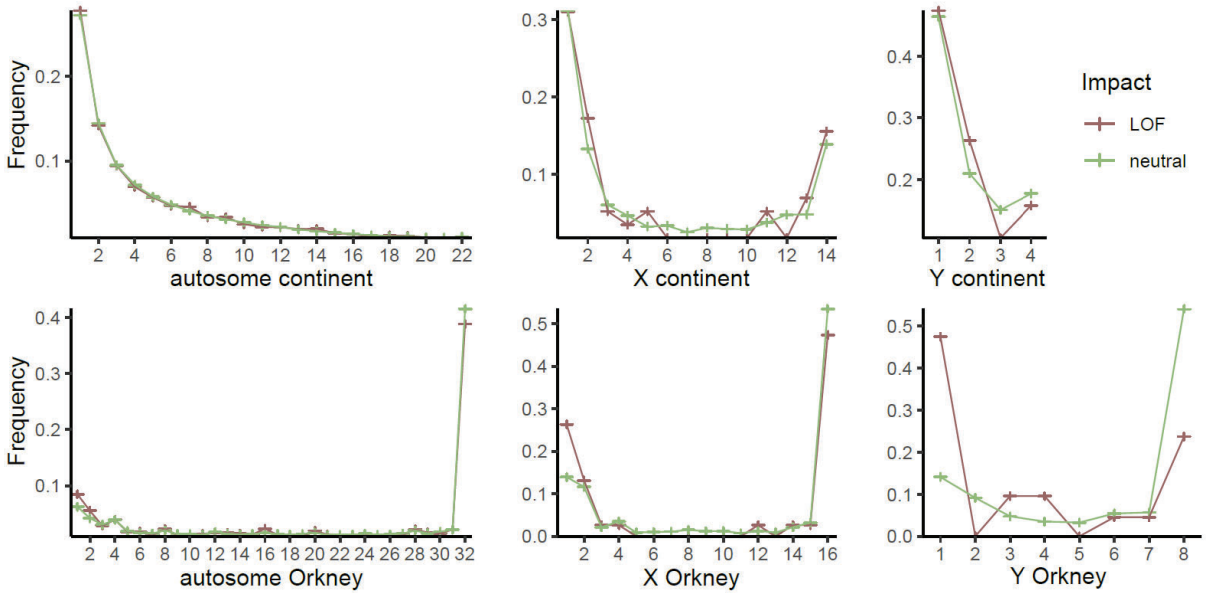



**Figure 4.** Unfolded site frequency spectra (SFS) of loss-of-function (LOF) and neutral SNPs on autosomes and sex chromosomes. Compared to autosomes, the SFS of LOF SNPs on sex chromosomes of Orkney voles were shifted towards lower frequency. The deviation between LOF and neutral spectra was stronger on the Y chromosome than on the X chromosome.

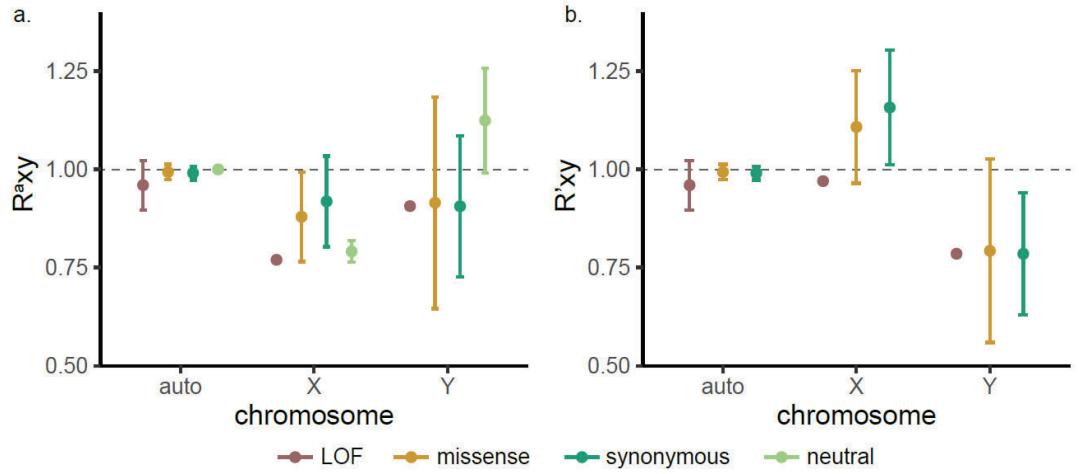



**Figure 5.** $R_{XY}$ relative to $R_{XY}$ of neutral SNPs on autosomes (**a**, $R^{a}_{XY}$) and $R_{XY}$ controlled by neutral $R_{XY}$ of each chromosome (**b**, $R'_{XY}$) of different chromosomes. The error bars stand for ± 2 standard errors (SE). **a.** The SE of $R_{XY}$ of loss-of-function (LOF) SNPs on sex chromosomes were high (0.2 to 0.6; not shown on the figures). Sex chromosomes of Orkney voles accumulated less mutations in coding regions compared to autosomes, and the X chromosome accumulated less than the Y chromosome. **b.** Within chromosomes, the accumulation of derived mutations in coding regions, except LOF mutations, was higher than in intergenic regions on the X chromosome, and the accumulation in coding regions was lower than in intergenic regions on the Y chromosome.

**Supplementary Information for**

# Sex-specific patterns of genetic diversity and mutation load in a small mammal

Xuejing Wang[a], Gerald Heckel[a,b]

[a]Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland; [b]Swiss Institute of Bioinformatics, Quartier Sorge - Batiment Amphipole, 1015 Lausanne, Switzerland

*Gerald Heckel

Email: gerald.heckel@unibe.ch

# Supplementary Information

**Details about the assembly of sex chromosomes of *Microtus arvalis***

The first step of assembly from short paired-end libraries produced very fragmented results of 1,249,460 contigs with N50 of 6702 bp, L50 at 4543 bp, and the longest contig with a length of 78714 bp. A total number of 22,721 contigs longer than 1 kb were used for the first round of scaffolding. There were two scaffolds longer than 1 Mb, 88.4 Mb and 2.5 Mb, potentially representing the X and Y chromosome. No long match was found between the new assembly of the X chromosome and the 22 scaffolds of autosomes from the reference genome (Figure S2), showing that the new assembly did not overlap with the autosomal assembly.

In the second round of assembly, 7388 contigs not assembled to the scaffold of the likely X chromosome were used. Multiple trials of the threshold of tandem repeat coverage were performed. The length of the longest scaffolds in the assembly was roughly correlated with the threshold (Table S2). The assembly with repeat coverage at 12 was the longest yet had much noise in the Hi-C map (Figure S3a), thus the second-longest assembly with repeat coverage at 8 was used as the draft for the final assembly of the Y chromosome (Figure S3b). After manual editing in Juicebox, the longest five scaffolds were kept for further validation. The total length of the five scaffolds was 9.9 Mb with the longest scaffold at 5.4 Mb and N-rate of 12.6%.

Using male specific k-mers, 40 sequences longer than 1 kb were assembled. The longest sequence was 3,174 bp and the total length was 39,942 bp. Among the male specific sequences, 30 were mapped to three scaffolds from the second round (5.4, 3.9 and 1.0 Mb) and none was mapped to the assembly of the X chromosome. Thus, these three scaffolds with the total length of 8.3 Mb, were considered the draft assembly of Y chromosome.

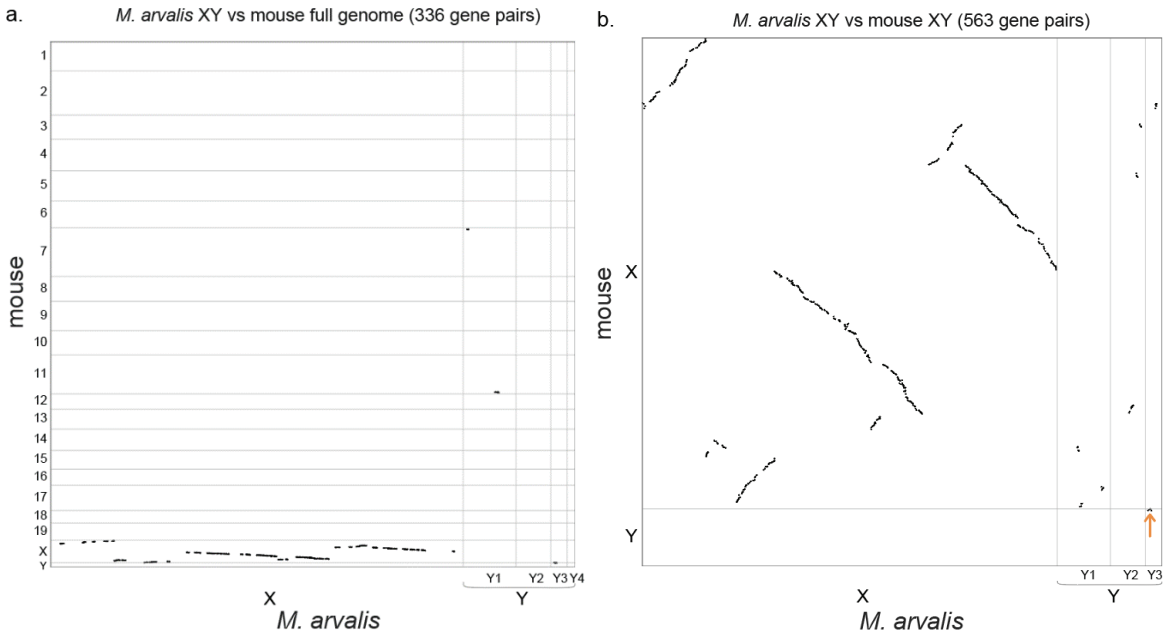



**Figure S1.** Synteny map between the sex chromosomes of *M. arvalis* and the mouse genome (**a**) or sex chromosomes (**b**) based on coding regions. **a.** The X chromosome of *M. arvalis* was homologous only with the X chromosome of mouse, while the Y scaffolds of *M. arvalis* had several homologous regions with mouse autosomes 7 and 11. **b.** The X chromosomes of *M. arvalis* and mouse maintained strong synteny in long blocks throughout the chromosome. The Y scaffolds of *M. arvalis* only shared short homologous regions with the sex chromosomes of mouse.

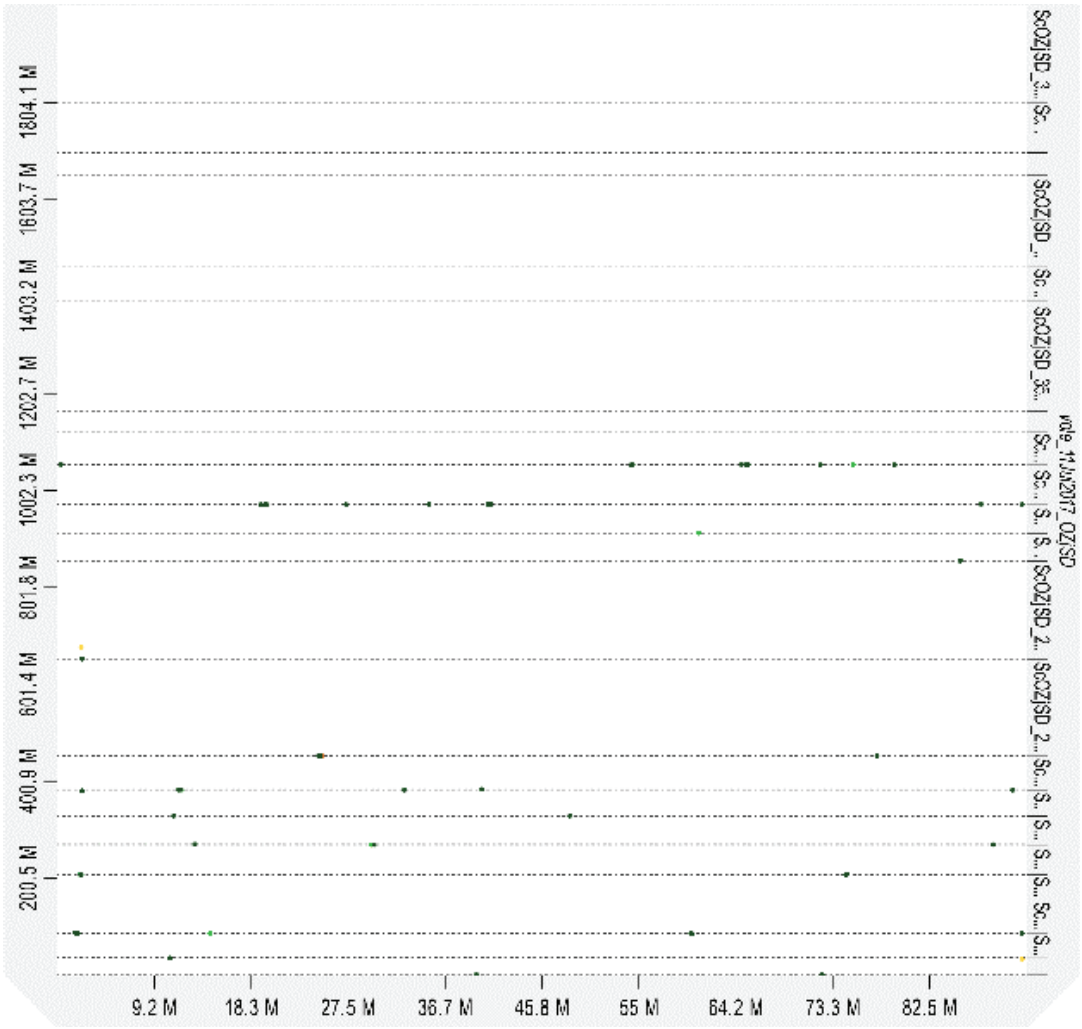
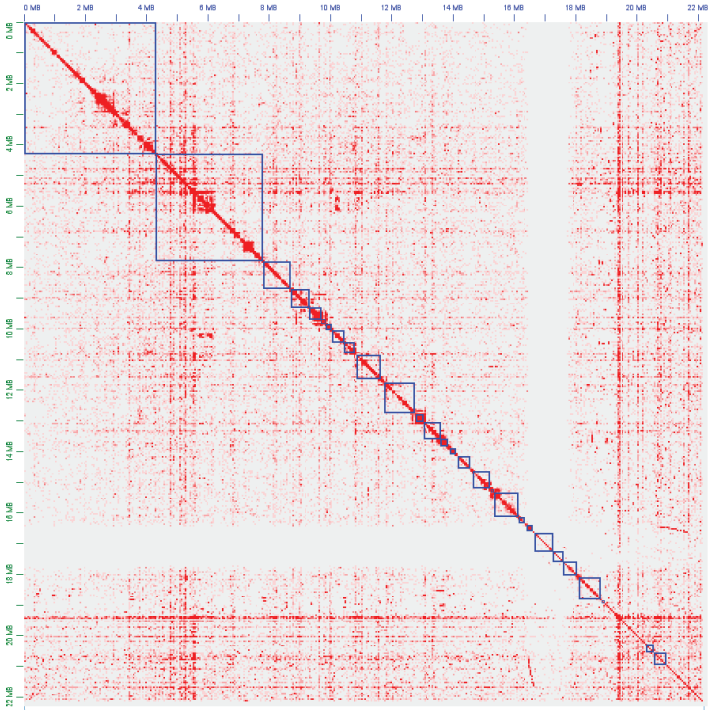



**Figure S2.** Synteny map between the first round of assembly of sex chromosomes and the reference genome of autosomes. The X-axis refers to the assembly of sex chromosomes and the Y-axis refers to the reference genome of autosomes. Dark green dots showed the synteny regions between the assembly of sex chromosomes and small scaffolds from reference genome.

**a.**



**b.**

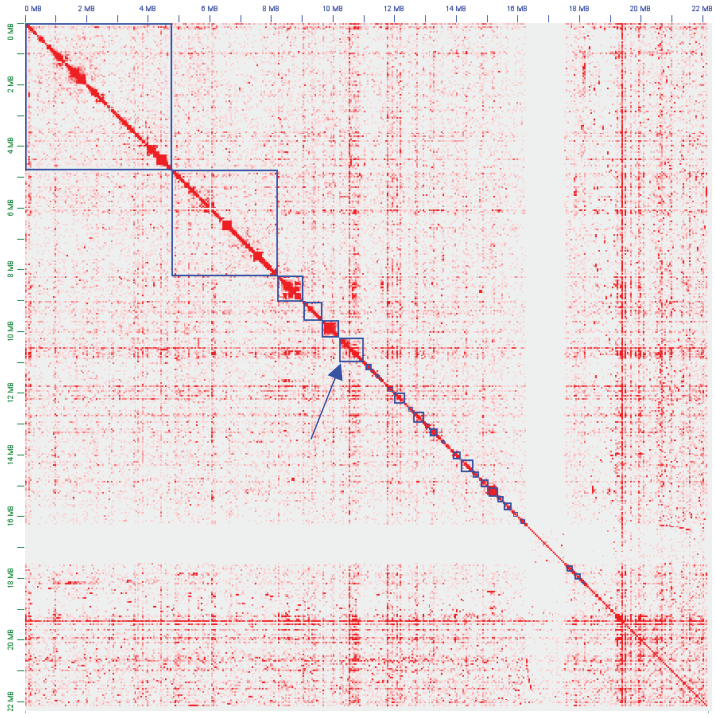


**Fig S3. a.** The Hi-C map of the assembly with repeat coverage at 12 showed much noise especially in the second scaffold. **b**. Manually reviewed and edited assembly with repeat coverage at 8. The sixth scaffold which is the fourth longest (marked with blue arrow), has obvious noise that causes a cross in the map, so it was excluded from the result.

**Table S1.** The samples used in this study with their sex, sampling location and mapping depth. The samples from the Central lineage were used for the validation of the Y assembly. The samples from the Western lineage (referred to as continental voles) and Orkney were used in the analyses of genetic variation. The *M. levis* samples were used as outgroup.

| Individual | Group | Sex | Latitude | Longitude | Depth |
|---|---|---|---|---|---|
| Eugene | Central lineage | M | 46.85 | 7.74 | 22.9 |
| CHIn04 | Central lineage | M | 47.01 | 7.11 | 22.8 |
| CHDa02 | Central lineage | M | 46.62 | 6.55 | 22.1 |
| DBu02 | Central lineage | M | 49.08 | 12.37 | 36.6 |
| DFl01 | Central lineage | M | 50.13 | 11.75 | 25.5 |
| He42 | Central lineage | F | 49.34 | 10.81 | 37.4 |
| DWa04 | Central lineage | F | 54.03 | 11.70 | 24.4 |
| DAb06 | Central lineage | F | 51.78 | 7.63 | 19.5 |
| DLu05 | Central lineage | F | 51.09 | 12.33 | 16.9 |
| NLHe04 | Central lineage | F | 52.94 | 5.96 | 21.6 |
| BSt095 | Western lineage | F | 51.22 | 3.07 | 22.7 |
| BVe061 | Western lineage | F | 51.07 | 2.67 | 40.8 |
| FBv02 | Western lineage | M | 46.81 | 6.16 | 17.4 |
| FCc02 | Western lineage | M | 47.18 | 6.82 | 27.4 |
| FDa504 | Western lineage | F | 49.77 | 0.52 | 23.8 |
| FFr549 | Western lineage | F | 50.08 | 1.57 | 23.3 |
| FGr14 | Western lineage | F | 46.68 | 6.19 | 37.5 |
| FMc03 | Western lineage | F | 49.25 | 3.93 | 19.0 |
| FOg02 | Western lineage | M | 47.43 | 6.74 | 16.9 |
| FPi555 | Western lineage | F | 50.87 | 1.82 | 22.6 |
| FSt24 | Western lineage | M | 46.39 | 4.23 | 29.7 |
| OBSO234 | Orkney | F | 58.85 | -2.92 | 23.4 |
| OBWs01 | Orkney | M | 58.84 | -2.92 | 33.3 |
| OEOs14 | Orkney | M | 59.18 | -2.76 | 26.0 |
| OEOs15 | Orkney | F | 59.18 | -2.76 | 27.2 |
| OMBr149 | Orkney | F | 59.08 | -3.30 | 22.4 |
| OMDo01 | Orkney | M | 59.08 | -3.21 | 32.1 |
| OMWr01 | Orkney | M | 58.99 | -2.93 | 35.3 |
| OMHo277 | Orkney | F | 58.94 | -3.07 | 25.8 |
| ORGs21 | Orkney | M | 59.13 | -2.99 | 34.3 |
| ORWa268 | Orkney | F | 59.18 | -3.06 | 30.5 |
| OSJC01 | Orkney | M | 58.76 | -2.94 | 30.3 |
| OSWi166 | Orkney | F | 58.77 | -2.93 | 31.6 |
| OWLs033 | Orkney | F | 59.30 | -2.91 | 18.8 |
| OWPg01 | Orkney | M | 59.32 | -3.00 | 28.5 |
| OSaLi257 | Orkney | F | 59.25 | -2.58 | 30.3 |
| OSaNe01 | Orkney | M | 59.27 | -2.48 | 28.4 |
| MroNLo01 | *M. levis* | M | 78.22 | 15.62 | 21.1 |
| MroNLo02 | *M. levis* | F | 78.22 | 15.62 | 21.6 |

**Table S2**. Lengths of big scaffolds in different assemblies. Lengths are shown in Mb. The two longest assemblies were marked in bold.

| **Repeat coverage** | 2 | 4 | 6 | **8** | 10 | **12** | 14 | 16 |
|---|---|---|---|---|---|---|---|---|
| **Longest scaffold (Mb)** | 3.4 | 3.9 | 4.9 | **3.5** | 4.3 | **4.3** | 3.6 | 3.6 |
| **Top 3 scaffolds (Mb)** | 5.2 | 6.1 | 6.9 | **8.3** | 6.6 | **8.8** | 6.8 | 6.8 |
| **Top 5 scaffolds (Mb)** | 6.6 | 7.7 | 8.1 | **9.9** | 8.5 | **10.6** | 8.6 | 8.6 |

**Table S3.** The median and mean values of π in 50 kb windows on different chromosomes. The values of autosomes (A) were calculated for males (M) and females (F) separately.

| **Group** | **Chromosome** | **Median** | **Mean** | **Median/Mean** |
|---|---|---|---|---|
| **Orkney** | A (F) | 0.00024 | 0.00048 | 0.50 |
| | A (M) | 0.00023 | 0.00048 | 0.48 |
| | X | 0.00010 | 0.00018 | 0.58 |
| | Y | 0.00024 | 0.00034 | 0.72 |
| **Continent** | A (F) | 0.00174 | 0.00177 | 0.98 |
| | A (M) | 0.00191 | 0.00193 | 0.99 |
| | X | 0.00045 | 0.00063 | 0.72 |
| | Y | 0.00131 | 0.00142 | 0.92 |

# General Discussion

In the genome of the Orkney vole, genetic drift has outweighed natural selection, while both forces brought it to the present form as a unique evolutionary case. Intense genetic drift emerged from the severe bottleneck and long isolation related to the introduction of Orkney voles by humans (**Chapter1**; Haynes, et al. 2003; Martínková, et al. 2013). The genomic consequences of genetic drift, including loss of genetic diversity (Mayr 1963), accumulation of homozygous deleterious alleles (Kimura, et al. 1963) and relaxation of selection (Gravel 2016) showed long-lasting impacts even after recovery from a bottleneck. On the other hand, selection has played a role regarding immunity and the change of body size of Orkney voles.

The extreme genomic landscape of the Orkney vole demonstrated genetic characteristics which are similar to many endangered species (Xue, et al. 2015; Grossen, et al. 2020; Bertorelle, et al. 2022; Robinson, et al. 2022). However, the ecological success of Orkney voles indicates no strong loss of absolute fitness, showing that the relationship between observed genetic patterns and realized fitness is complex. Furthermore, the autosomes and sex chromosomes showed different genetic patterns suggesting potential populational impacts of the mutation load on sex chromosomes.

**History of the Orkney vole is closely associated with humans**

The demographic inferences in **Chapter 1** show that the Orkney vole is likely one of the oldest cases of human introduction of island populations, which has remained isolated for more than 10,000 generations, consistent with previous studies (Martínková, et al. 2013; Cucchi, et al. 2014). The demographic inference showed that the introduction happened around 5,000 years ago, consistent with previous dating of the bones found in Neolithic settlements in Orkney (Martínková, et al. 2013; Cucchi, et al. 2014), and the founding population experienced a severe bottleneck related to the introduction. The fast spread of Orkney voles into multiple islands was likely related to the activities of Neolithic Orkney residents, supported by archeological records expanding for a few hundred years after introduction (Cucchi, et al. 2014), and signals of early gene flow between islands. The introduction may have brought selective pressure to Orkney voles regarding immunity as

shown in **Chapter 2**. The high density of the voles during the transport and management by the Neolithic farmers (Romaniuk, et al. 2016) likely increased the transmission of pathogens (Hochberg 1991; Laakkonen, et al. 1999; Altizer, et al. 2006).

The most comparable system is the Soay sheep which was also introduced by Neolithic farmers around 6,000 years ago to Orkney and the St Kilda archipelago (Clutton-Brock and Pemberton 2004). Despite the similar time of introduction, the Orkney vole had longer isolation measured in the number of generations because of the shorter generation time (0.5 compared to 3 years for the sheep), and much higher modern effective and census population sizes (Stoffel, et al. 2021). Together with the locally extinct red deer which was also introduced to Orkney in Neolithic time (Stanton, et al. 2016), both cases showed the ability of human transportation of animals in ancient times. The effects of such ancient movements can last a long time and influence the local ecosystem permanently as both Orkney vole and Soay sheep have maintained a stable modern population and integrated into the ecosystem of the islands (Reynolds 1992; Clutton-Brock and Pemberton 2004).

The Orkney vole can be considered as a "Neolithic invasive species" in three ways: human-related introduction and related demographic changes, a degree of isolation from its original population (in our case complete isolation), and the capability of fast reproduction and expansion (Colautti, et al. 2017; Courchamp, et al. 2017). The case of the Orkney vole indicates that in the long term, some invasive species may be incorporated into the local ecosystem and may have reached a new balance (Walther, et al. 2009; Shackleton, et al. 2019). For example, two invasive plant species have become the most visited plants by local birds on Galápagos islands (Traveset, et al. 2015). Of course, reaching the new balance has an extremely high price, since the alien species pose threats to local species, or even to the whole ecosystem, which is the major concern of invasion biology (Luque, et al. 2014; Courchamp, et al. 2017). The establishment of the Eday population in 1987 may serve as an example of a (seemingly) harmless introduction (Mike Cockram, personal communication) to understand how the disturbed ecosystem returned to a stable state, but further ecological investigations are required.

**Advances to expand the understanding of mutation load**

The populations on multiple Orkney islands can be considered as an exceptionally long "experiment" with repeated trials on the evolution of bottlenecked and isolated populations.

In **Chapter 1** and **Chapter 3**, the result of this "experiment" showed typical consequences of bottlenecks and founder events predicted by population genetics theory. The genetic consequences include loss of genetic diversity and accumulation of deleterious mutations in the homozygous state, which is expected to associate with loss of fitness (Xue, et al. 2015; Robinson, et al. 2018; Grossen, et al. 2020).

Both genetic diversity and homozygosity of deleterious mutations serve as estimates of genetic quality in conservation biology, and are expected to be associated with each other (Allendorf, et al. 2010; Díez-del-Molino, et al. 2018; Hohenlohe, et al. 2021). Reduced genetic diversity and expressed deleterious mutations in homozygous form are expected to increase the risk of extinction of affected populations (Lynch, et al. 1995; Allendorf, et al. 2010; Kardos, et al. 2021). Surprisingly, the flourishing Orkney vole populations showed these two genetic characteristics at extreme levels without suffering obvious loss of fitness. This shows that the method I used for estimating mutation load, by approximating the fitness effects of variants to their functional effect, cannot represent the broad difference of fitness effects between the variants and reveals the need of further studies (Kardos, et al. 2021). The distribution of fitness effects (DFE) of functional variants is only available in a few species (e.g. Wloch, et al. 2001; Eyre-Walker, et al. 2006; Eyre-Walker and Keightley 2007; Castellano, et al. 2019; Gilbert, et al. 2022), and remains unknown in the Orkney populations, which is the case in most non-model species. Thus, in this thesis and similar studies (Palkopoulou, et al. 2015; Grossen, et al. 2020; Robinson, et al. 2022), the functional annotation of the alleles was used to approximate the fitness effects of the alleles. For example, loss-of-function (LOF) mutations were treated as highly deleterious variants, synonymous or intergenic mutations as neutral variants, etc. However, our forward simulations in **Chapter 1** showed that the homozygous LOF mutations we observed likely had only mildly deleterious fitness effects while the highly deleterious variants were probably purged much earlier during the bottleneck. This shows the limit of equating inferred functional effects to realized fitness effects when estimating mutation load in the population.

The diversity and dynamics of the DFE may bring additional difficulties to the comparison between populations and species. Certain life history traits, such as life span and reproductive rate, affect the ability to tolerate deleterious mutations of the population. Species with shorter life span have higher numbers of deleterious variants (van der Valk, et al. 2021), likely related to relaxed selection on the genes involved in the late stage of life and

genetic drift from repeated bottlenecks (Cui, et al. 2019). Species with high reproductive rate have a higher genetic potential in the offspring (Hedrick, et al. 2016) and are less likely affected by inbreeding depression compared to species with fewer offspring. The common vole is a typical *r*-strategist with short life expectancy and high reproductivity (Boyce and Boyce III 1988; Tkadlec and Zejda 1995). This may explain the difference between the extreme genomic patterns of the Orkney vole and the observations in many other large mammals, e.g. ibex, tigers or foxes (Grossen, et al. 2020; Khan, et al. 2021; Robinson, et al. 2022). The fitness effects of certain variants are also related to the environment posing selection on the population, and thus the DFE evolves following shifts in the environment (Kardos, et al. 2021). Furthermore, the DFE may vary between the chromosomes within the genome, as the dominance coefficient of mutations differs between the autosomes and the sex chromosomes (Immler 2019). To better understand how mutation load arises during demographic changes, and affects the ecological performance of the population, it is necessary to face the challenge and fill the gap between the genetic variation and its fitness effects, for example, with field experiments estimating detailed life history traits.

**Linking genomic signals to the island syndrome**

Current selection tests or genome-wide methods usually target genes with large effects linked to phenotypic changes and detect the signals standing out from the neutral background (Kimura 1977; Kryazhimskiy and Plotkin 2008). However, the application of these methods in genetic studies of the island syndrome have the following limits. First, many traits involved in the island syndrome are complex traits controlled by a high number of genes, for example behavior changes (Chabris, et al. 2015) and body size (Gray, et al. 2015). In the case of multiple genes under moderate selective pressure, the signatures may not be captured. Second, depending on the methods used, shift of selective pressure can be revealed not only by genomic signatures detected related to adaptation in the island population, but instead in the continental population. In **Chapter 2**, I found signals of positive selection related to body size (likely against large size) only in continental voles. This result indicates that the Orkney voles were released from the selective pressure of being small, rather direct selection on the growth-related genes in island populations found by studies of giant island rodents (Payseur and Jing 2021; Renom, et al. 2021). This explanation is backed up in three aspects. The body size of the common voles on the continent has decreased in the past centuries, shown by 5%

decrease of molar size compared to ancient samples dated to 800 – 22,000 years ago (Cucchi, et al. 2014). The other hint is that snow voles (*Chionomys nivalis*) in the Alps were found to be under selection for small body size and fast maturity to survive winter with less energy cost (Bonnet, et al. 2017). Similarly, the continental voles may have experienced selective pressure related to food resources and temperature, which the Orkney voles may be released from. Third, the genome-wide reduction of efficacy of selection because of genetic drift, as shown in **Chapter 1** (high fixation rate of LOF mutations) and **Chapter 2** (overall increase of $d_N/d_S$ ratios), may have impacts on the function of a broad range of genes and thus affect certain traits of island populations, yet related studies are still very limited (Grueber, et al. 2013; Barthe, et al. 2022).

**Prospects of sequencing and computational tools for genetic studies of wild species**

This PhD work is the first comprehensive genomic study of the Orkney vole which expanded our knowledge of the evolution of rodents. So far, genomic studies have been focused on a limited number of species with particular concern for health, the economy or conservation. Taking NCBI Genbank as an example (Sayers, et al. 2023), there are reference genomes of 532 mammal species comprising about 8% of all extant mammal species (Burgin, et al. 2018). Primates are the most sequenced order of mammal, with more than 14% of the species sequenced, while only 5% of Rodentia is covered. This coverage is very high compared to other taxa. For example the myriapods, which play an essential ecological role in breaking down decayed plants (Minelli 2011), have only nine genomes from over 13,000 species sequenced. Studies on previously neglected organisms can bring new ideas to both theoretical and applied evolutionary biology. For example, a majority of modern population genetics theory is based on haploid or diploid system (Crow and Kimura 1970; Kimura and Ohta 1971), while studies of species with multiple ploidy levels (Pruvost, et al. 2013; Stift, et al. 2019; Gerstein and Sharp 2021) may help us to better understand the evolutionary cause and consequences of ploidy changes.

With the fast development of sequencing techniques, it is time to apply advanced genomic tools to a wider range of branches on the tree of life (Hohenlohe, et al. 2021). We are going through an explosion of genomic data as the cost of sequencing drops drastically and the efficiency increases with the latest sequencing platforms. The biggest obstacle to genomic studies of wild species, especially at the population level, will be not the cost of

sequencing, but rather the accessibility of genomic resources and practical analysis methods. The development of third generation sequencing e.g., PacBio (Rhoads and Au 2015) and Nanopore Technologies (Deamer, et al. 2016), together with Hi-C sequencing (Belton, et al. 2012), has made it possible and affordable to produce *de novo* genome assemblies at the chromosome level with high quality (Athanasopoulou, et al. 2022). It was difficult to assemble especially the Y or W chromosome only with NGS technology due to the heterogeny of the sex chromosomes. In **Chapter 3**, I pushed our data of NGS and Hi-C sequencing to the limit and was unable to assemble the complete Y chromosome. Nowadays, long read sequencing, or combined with short read NGS sequencing provides budget and high-quality solutions to sex chromosome assembling (Peichel, et al. 2020; Couger, et al. 2021).

The genomic resources include not only the reference genomes, but also additional descriptive features of the genome, such as gene annotation, mutation rate and recombination rate, which are essential for phylogenetic, genomic and population genetic analyses (Stein 2001; Nachman 2002; Hodgkinson and Eyre-Walker 2011; Lynch, et al. 2016). Bioinformatical methods have been developed for estimation of the mutation rate (Palamara, et al. 2015) and recombination rate (Hermann, et al. 2019) using population genomics data from a small number of samples, which I have applied in **Chapter 1**. Such new methods are usually developed based on and for model species, for example humans, and have limits when applied to a broader range of species with diverse life history and genetic characters. A similar issue of applying adequate methods when studying non-model species occurs also in the downstream analyses. Taking demographic inference as an example, SMC methods (PSMC, MSMC, SMC++, etc.) have been widely used (Li and Durbin 2011; Schiffels and Durbin 2014; Terhorst, et al. 2017; Mather, et al. 2020). In **Chapter 1**, I used SMC++ (Terhorst, et al. 2017) to infer the population sizes and split times of the Orkney populations. We did not use the more widespread method MSMC (Schiffels and Durbin 2014), because of its limit of sample size and massive requirement of computational resources. Each method has its advantages and restrictions, and with the appearance of new methods, it has become one of the biggest difficulties of any genomic study to choose the best tool to answer the question with a reasonable cost of resource, time, and learning effort. International collaborative networks, such as European Reference Genome Atlas (Formenti, et al. 2022), may be a solution to such information barriers of methodology.

**Concluding remarks**

With the comprehensive genomic analyses of the extreme example of the Orkney vole, I have expanded the understanding of the genetics of isolated populations. The results have revealed the importance of looking into details in future studies of mutation load, for example the distribution of fitness effects of the alleles, both in the Orkney vole system and in other organisms. It has also shown the necessity of expanding research of species adaptation combining demography and selection together to a wider range of species with diverse life histories and ecological traits, and to the whole genome, including the sex chromosomes. The fast development of genomic technologies and explosively emerging genomic data bring us great power to tackle long-standing questions, such as the impacts of demographic changes on local adaptation, or to fill the gap between genetic variation and its fitness effects in real world. Challenge is opportunity, and I believe that in the near future, we will arrive at a new stage of understanding the evolution on the balance of drift and selection.

# References


Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 11:697-709.

Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P. 2006. Seasonality and the dynamics of infectious diseases. Ecology Letters 9:467-484.

Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A. 2022. Third-generation sequencing: the spearhead towards the radical transformation of modern genomics. Life 12:30.

Barthe M, Doutrelant C, Covas R, Melo M, Illera JC, Tilak M-K, Colombier C, Leroy T, Loiseau C, Nabholz B. 2022. Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome. BioRxiv:2021.2011.2021.469450.

Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods 58:268-276.

Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, Morales HE, Van Oosterhout C. 2022. Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics 23:1-12.

Bonnet T, Wandeler P, Camenisch G, Postma E. 2017. Bigger is fitter? Quantitative genetic decomposition of selection reveals an adaptive evolutionary decline of body mass in a wild rodent population. PLoS Biology 15:e1002592.

Boyce CC, Boyce III JL. 1988. Population biology of *Microtus arvalis*. I. Lifetime reproductive success of solitary and grouped breeding females. The Journal of Animal Ecology:711-722.

Burgin CJ, Colella JP, Kahn PL, Upham NS. 2018. How many species of mammals are there? Journal of Mammalogy 99:1-14.

Castellano D, Macià MC, Tataru P, Bataillon T, Munch K. 2019. Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. Genetics 213:953-966.

Chabris CF, Lee JJ, Cesarini D, Benjamin DJ, Laibson DI. 2015. The fourth law of behavior genetics. Current Directions in Psychological Science 24:304-312.

Clutton-Brock TH, Pemberton JM. 2004. Soay sheep: dynamics and selection in an island population. Cambridge University Press.

Colautti RI, Alexander JM, Dlugosch KM, Keller SR, Sultan SE. 2017. Invasions and extinctions through the looking glass of evolutionary ecology. Philosophical Transactions of the Royal Society B: Biological Sciences 372.

Couger MB, Roy SW, Anderson N, Gozashti L, Pirro S, Millward LS, Kim M, Kilburn D, Liu KJ, Wilson TM. 2021. Sex chromosome transformation and the origin of a male-specific X chromosome in the creeping vole. Science 372:592-600.

Courchamp F, Fournier A, Bellard C, Bertelsmeier C, Bonnaud E, Jeschke JM, Russell JC. 2017. Invasion biology: specific problems and possible solutions. Trends in Ecology & Evolution 32:13-22.

Crow JF, Kimura M. 1970. An introduction to population genetics theory. Burgess Publishing Company.

Cucchi T, Barnett R, Martínková N, Renaud S, Renvoisé E, Evin A, Sheridan A, Mainland I, Wickham-Jones C, Tougard C, et al. 2014. The changing pace of insular life: 5000 years of microevolution in the orkney vole (*microtus arvalis orcadensis*). Evolution 68:2804-2820.

Cui R, Medeiros T, Willemsen D, Iasi LNM, Collier GE, Graef M, Reichard M, Valenzano DR. 2019. Relaxed Selection Limits Lifespan by Increasing Mutation Load. Cell 178:385-399.e320.

Deamer D, Akeson M, Branton D. 2016. Three decades of nanopore sequencing. Nature Biotechnology 34:518-524.

Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L. 2018. Quantifying Temporal Genomic Erosion in Endangered Species. Trends in Ecology and Evolution 33:176-185.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:610-618.

Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173:891-900.

Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J. 2022. The era of reference genomes in conservation genomics. Trends in Ecology & Evolution 37:197-202.

Gerstein AC, Sharp NP. 2021. The population genetics of ploidy change in unicellular fungi. FEMS Microbiology Reviews 45:fuab006.

Gilbert KJ, Zdraljevic S, Cook DE, Cutter AD, Andersen EC, Baer CF. 2022. The distribution of mutational effects on fitness in Caenorhabditis elegans inferred from standing genetic variation. Genetics 220:iyab166.

Gravel S. 2016. When is selection effective? Genetics 203:451-462.

Gray MM, Parmenter MD, Hogan CA, Ford I, Cuthbert RJ, Ryan PG, Broman KW, Payseur BA. 2015. Genetics of rapid and extreme size evolution in island mice. Genetics 201:213-228.

Grossen C, Guillaume F, Keller LF, Croll D. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nature communications 11:1001.

Grueber CE, Wallis GP, Jamieson IG. 2013. Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. Mol Ecol 22:4470-4482.

Haynes S, Jaarola M, Searle JB. 2003. Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. Molecular Ecology 12:951-956.

Hedrick PW, Hellsten U, Grattapaglia D. 2016. Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of Eucalyptus grandis. New Phytologist 209:600-611.

Hermann P, Heissl A, Tiemann-Boege I, Futschik A. 2019. *LDJump* : Estimating variable recombination rates from population genetic data. Molecular ecology resources 19:623-638.

Hochberg ME. 1991. Non-linear transmission rates and the dynamics of infectious disease. Journal of Theoretical Biology 153:301-321.

Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. Nature Reviews Genetics 12:756-766.

Hohenlohe PA, Funk WC, Rajora OP. 2021. Population genomics for wildlife conservation and management. Molecular Ecology 30:62-82.

Immler S. 2019. Haploid selection in "diploid" organisms. Annual Review of Ecology, Evolution, and Systematics 50:219-236.

Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM, Tallmon DA, Funk WC. 2021. The crucial role of genome-wide genetic variation in conservation. Proceedings of the National Academy of Sciences 118:e2104642118.

Khan A, Patel K, Shukla H, Viswanathan A, van der Valk T, Borthakur U, Nigam P, Zachariah A, Jhala YV, Kardos M, et al. 2021. Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. Proceedings of the National Academy of Sciences 118:e2023018118.

Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275-276.

Kimura M, Maruyama T, Crow JF. 1963. The mutation load in small populations. Genetics 48:1303.

Kimura M, Ohta T. 1971. Theoretical aspects of population genetics. Princeton University Press.

Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. PLOS Genetics 4:e1000304.

Laakkonen J, Henttonen H, Niemimaa J, Soveri T. 1999. Seasonal dynamics of Pneumocystis carinii in the field vole, *Microtus agrestis*, and in the common shrew, *Sorex araneus*, in Finland. Parasitology 118:1-5.

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

Luque GM, Bellard C, Bertelsmeier C, Bonnaud E, Genovesi P, Simberloff D, Courchamp F. 2014. The 100th of the world's worst invasive alien species. Biological Invasions 16:981-985.

Lynch M, Ackerman MS, Gout J-f, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. Nature Reviews Genetics 17:704-714.

Lynch M, Conery J, Burger R. 1995. Mutation accumulation and the extinction of small populations. The American Naturalist 146:489-518.

Martínková N, Barnett R, Cucchi T, Struchen R, Pascal M, Pascal M, Fischer MC, Higham T, Brace S, Ho SYW, et al. 2013. Divergent evolutionary processes associated with colonization of offshore islands. Molecular Ecology 22:5205-5220.

Mather N, Traves SM, Ho SYW. 2020. A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. Ecology and Evolution 10:579-589.

Mayr E. 1963. Animal species and evolution. Belknap Press of Harvard University Press.

Minelli A. 2011. The Myriapoda. Brill Leiden.

Nachman MW. 2002. Variation in recombination rate across the genome: evidence and implications. Current opinion in genetics & development 12:657-663.

Palamara PF, Francioli LC, Wilton PR, Genovese G, Gusev A, Finucane HK, Sankararaman S, Sunyaev SR, De Bakker PIW, Wakeley J, et al. 2015. Leveraging distant relatedness to quantify human mutation and gene-conversion rates. American Journal of Human Genetics 97:775-789.

Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A, et al. 2015. Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. Current Biology 25:1395-1400.

Payseur BA, Jing P. 2021. Genomic targets of positive selection in giant mice from Gough Island. Molecular Biology and Evolution 38:911-926.

Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM, Kingsley DM, et al. 2020. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome biology 21:177.

Pruvost NB, Hoffmann A, Reyer HU. 2013. Gamete production patterns, ploidy, and population genetics reveal evolutionary significant units in hybrid water frogs (*Pelophylax esculentus*). Ecology and Evolution 3:2933-2946.

Renom P, de-Dios T, Civit S, Llovera L, Sánchez-Gracia A, Lizano E, Rando JC, Marquès-Bonet T, Kergoat GJ, Casanovas-Vilar I. 2021. Genetic data from the extinct giant rat from Tenerife (Canary Islands) points to a recent divergence from mainland relatives. Biology Letters 17:20210533.

Reynolds P. 1992. The impact of changes in land-use in Orkney, on the vole *Microtus arvalis orcadensis* and its avian predators. University of Aberdeen.

Rhoads A, Au KF. 2015. PacBio sequencing and its applications. Genomics, proteomics & bioinformatics 13:278-289.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK. 2018. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Current Biology 28:3487-3494.e3484.

Robinson JA, Kyriazis CC, Nigenda-Morales SF, Beichman AC, Rojas-Bracho L, Robertson KM, Fontaine MC, Wayne RK, Lohmueller KE, Taylor BL, et al. 2022. The critically endangered vaquita is

not doomed to extinction by inbreeding depression. Science 376:635-639.

Romaniuk AA, Shepherd AN, Clarke DV, Sheridan AJ, Fraser S, Bartosiewicz L, Herman JS. 2016. Rodents: food or pests in Neolithic Orkney. Royal Society open science 3:160514.

Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Sherry ST, Yankie L, Karsch-Mizrachi I. 2023. GenBank 2023 update. Nucleic Acids Research 51:D141-D144.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. Nature genetics 46:919-925.

Shackleton RT, Shackleton CM, Kull CA. 2019. The role of invasive alien species in shaping local livelihoods and human well-being: A review. Journal of Environmental Management 229:145-157.

Stanton DWG, Mulville JA, Bruford MW. 2016. Colonization of the Scottish Islands via long-distance Neolithic transport of red deer (*Cervus elaphus*). Proceedings of the Royal Society B: Biological Sciences 283:20160095.

Stein L. 2001. Genome annotation: from sequence to biology. Nature Reviews Genetics 2:493-503.

Stift M, Kolář F, Meirmans PG. 2019. Structure is more robust than other clustering methods in simulated mixed-ploidy populations. Heredity 123:429-441.

Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM. 2021. Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. Nature communications 12:2972.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nature Genetics 49:303-309.

Tkadlec E, Zejda J. 1995. Precocious breeding in female common voles and its relevance to rodent fluctuations. Oikos 73:231-236.

Traveset A, Olesen JM, Nogales M, Vargas P, Jaramillo P, Antolín E, Trigo MM, Heleno R. 2015. Bird–flower visitation networks in the Galápagos unveil a widespread interaction release. Nature communications 6:1-6.

van der Valk T, de Manuel M, Marques-Bonet T, Guschanski K. 2021. Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. BioRxiv:696831.

Walther G-R, Roques A, Hulme PE, Sykes MT, Pyšek P, Kühn I, Zobel M, Bacher S, Botta-Dukát Z, Bugmann H. 2009. Alien species in a warmer world: risks and opportunities. Trends in Ecology & Evolution 24:686-693.

Wloch DM, Szafraniec K, Borts RH, Korona R. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. Genetics 159:441-452.

Xue Y, Prado-martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, et al. 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science 348:242-245.

# Acknowledgements

My PhD started in an unexpected way and have been through an unexpected way. It would not have been possible for me to finish it without the help of many people in the last five years.

I especially would like to thank Gerald Heckel for giving me the opportunity to work with him when I was close to drop out, for devoting so many efforts to supervision and education, and for listening all the time. I would also like to thank Laurent Excoffier for running the lab and leading all the members. I also thank Lukas Keller for the evaluation of my thesis and Katie Peichel for accepting to be the chair at my defense.

Additionally, I would like to thank Susanne Tellenbach for the support in the lab and in my daily life; Stephan Peischl for his beautiful simulations and great help; Kim Gilbert and Nina Marchi, for answering my questions at any time; and Adamantia Kapopoulou for being the best officemate. I thank Zuyao Liu for countless help and discussions in science and life, and Xiangyi Li Richter for career guidance and sharing a hobby of balcony gardening. Both of them are like my family especially when my real family was too far to reach. I also would like to thank all the members of the CMPG Lab and many other colleagues in IEE. The pandemic was nothing close to pleasing, but it was my pleasure to go through it with my colleagues.

Nature is my object of study, my motivation and my inspiration. I would like to thank the voles sacrificed for my project. I also thank the great nature of Switzerland, the mountains, the lakes, the plants and the animals, for bringing much color to my life.

Finally, thanks to my parents who provided the best support: always believing in me, and all my friends around the world who supported me beyond the differences of time.

# Declaration of consent

on the basis of Article 18 of the PromR Phil.-nat. 19

Name/First Name:

Registration Number:

Study program:

Bachelor ☐ Master ☐ Dissertation ☐

Title of the thesis:

Supervisor:

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of September 5th, 1996 and Article 69 of the University Statute of June 7th, 2011 is authorized to revoke the doctoral degree awarded on the basis of this thesis.
For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with theses submitted by others.

Place/Date

Signature