THE TAMING OF THE SHREW(‘S GENOME): AN OMICS APPROACH TO STUDYING THE WORLD’S SMALLEST MAMMALS

A Dissertation Submitted to the Committee on Graduate Studies

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Environmental and Life Sciences Ph.D. Graduate Program

January 2026

Abstract

The Taming of the Shrew('s Genome): An Omics Approach to Studying the World's Smallest Mammals - Marie-Laurence Cossette

The diversification of mammals has been shaped by climatic fluctuations and geological changes over millions of years. Among them, shrews (Soricidae) stand out as one of the most diverse mammalian lineages. Shrews are found on most continents and have evolved remarkable adaptations at the species and population levels. Although evolutionary studies of shrews have been limited by a lack of genomic resources, this work aimed to address this gap by developing such resources and using various omics approaches to explore adaptation and divergence in shrews, with a particular focus on an isolated population on Bon Portage Island (BPI), Nova Scotia, Canada. BPI shrews exhibit distinctive foraging habits and possess a unique allele for a dietary enzyme, which may represent an adaptation to their specialized diet. My research involved assembling and annotating *de novo* genomes from three shrew species. In Chapter 2 I conducted a comparative genomic analysis of 20 mammals (including four shrew species) to identify lineage-specific adaptations including accelerated regions, gene family expansions, and positively selected genes. I found shrew-specific variants in genes associated with the nervous, metabolic, and auditory systems, which may underlie some of their unique traits. In Chapter 3, I examined morphological and epigenetic divergence between mainland and island populations of masked shrews (*Sorex cinereus*), including BPI. Island shrews exhibited smaller body sizes and signs of accelerated biological aging, marked by DNA methylation differences enriched in developmental and digestive pathways. Chapter 4 focused on analyzing

genome-wide SNP data to identify regions of differentiation, alongside RNA-seq data to perform a differential gene expression analysis between BPI shrews and other populations. The results from both analyses revealed patterns of differentiation in genes associated with fatty acid metabolism and metabolic regulation that are likely linked to their specialized, largely marine-based diet. Additionally, I reconstructed the demographic history of Nova Scotia masked shrew populations, revealing that the divergence of the BPI population appears to coincide with rising sea levels following the last glacial maximum. These findings shed light on mechanisms of adaptation and divergence, illustrating how ecological pressures, geographic isolation, and dietary specialization shape genomic, epigenomic, and transcriptomic landscapes.

Keywords

genome assembly, comparative genomics, population genomics, island syndrome, epigenetics, transcriptomics, shrews

Preface

I have written my dissertation following a manuscript format as all data chapters are either published or being prepared for submission in a peer-reviewed journal. Specifically, Chapter 2 was published in *Genome Biology and Evolution* in 2024, Chapter 3 was published in *Molecular* Ecology in 2022, and Chapter 4 is in preparation for submission in *Molecular Biology and Evolution.* The style of each data chapter is specific to their respective journal formatting requirements, however, in the interest of space, I used a single citation format throughout with a single reference section at the end. Where collaborative work was involved, I have used the collective "we". In addition to my own research, I had the opportunity to collaborate with the Mammalian Methylation Consortium and with Dr. Jeong's group at Seoul National University during a semester abroad. These experiences, along with others, have led to my collaboration on the following:

1. Kim, D., Choi, J., Kim, D., **Cossette, M-L.**, Kim, S., Tsai, C-L., Baek, M.J., Park, S-J., Borzée, A., Shin, S., Jeong, C. A deep population stratification and ongoing local adaptations of the range-expanding lovebug *Plecia longiforceps*. *Commun Biol* (in review).

2. Lu, A. T., [and 185 others, including **Cossette M-L**], Horvath, S. (2023). Universal DNA methylation age across mammalian tissues. *Nature Aging*, 3(9), 1144–1166. https://doi.org/10.1038/s43587-023-00462-6

3. Haghani, A., [and 160 others, including **Cossette M-L**], Horvath, S. (2023). DNA methylation networks underlying mammalian traits. *Science*, 381(6658). https://doi.org/10.1126/science.abq5693

4. Wolf, J. F., MacKay, L., Haworth, S. E., **Cossette, M-L.,** Dedato, M. N., Young, K. B., Elliott, C. I., & Oomen, R. A. (2021). Preprinting is positively associated with early career researcher status in ecology and evolution. *Ecology and Evolution*, 11(20), 13624–13632. https://doi.org/10.1002/ece3.8106

Acknowledgments

I begin by acknowledging the land on which this work was carried out and by expressing my gratitude to the First Peoples who have long cared for it. This research was conducted on the traditional territories of the Mississauga (Michi Saagiig) Anishinaabeg and the Mi'kmaq. I recognize my privilege and remain committed to continually educating myself and others. I am deeply grateful for the opportunity to work on this land and for the enduring care, stewardship, and teachings of its First Peoples.

Many thanks to all my family and friends. To all the women in my life, I cherish our friendships. I have been privileged to witness some of your life's biggest milestones during this PhD, and I am so fortunate to have you all in my life. To my Korean friends, thank you for welcoming me even though you all thought I was a bit too outgoing and talked too much. I value the time you've taken to share your culture with me. Maman, Papa, Charlot, Loulou, et Papi je vous aime énormément. Nala, my soul cat, should earn an honorary degree for her attendance at nearly every lab meeting (arguably more consistent than some students) from her preferred seat… my lap. Throughout the countless hours spent on this dissertation, she has matched my dedication with equal commitment to napping by my side.

Don, thank you for welcoming me to Nova Scotia and for letting me continue the BPI shrew legacy. I've enjoyed every email update from you and hearing about your family and travels. Even from afar, I have felt your support and am grateful for everything. To Joana and everyone else who has been involved in my project along the way, thank you.

Aaron, I started in the lab as a 20-year-old third-year undergrad, and now at 27 (and almost a Dr.!), I'm closing this chapter of my life. These past few years have flown by, and I am truly grateful for all the guidance and support you've given me along the way. I've enjoyed our Slack messages, though I admit half the time I'm either being ignored, not understanding what you're saying, or receiving seemingly passive-aggressive replies (I know they aren't, but FYI, replying with only "okay." makes my generation think you're upset). You are a researcher I admire and a person whose commitment to justice I deeply respect. Thank you for being exactly the mentor I needed every step of the way.

Finalement, Brian, mon autre moitié, je pourrais probablement écrire un essai aussi long que cette thèse pour te dire à quel point je t'apprécie et à quel point je suis reconnaissante de t'avoir dans ma vie, mais je vais essayer de faire court (je sais, moi et la concision). Merci pour nos innombrables rendez-vous café-travail et tous les matcha lattes pendant notre mission d'essayer chaque café en ville. Merci pour chaque fois que tu as conduit jusqu'à Picton, ou honnêtement n'importe où, même après tes shifts de nuit, juste pour que je puisse travailler dans la voiture. Merci de ne pas m'en vouloir trop quand la lumière de mon écran te réveillait à 2 h du matin, de tolérer le chaos total qu'est mon bureau et de t'occuper de notre Bibitte d'amour chaque fois que j'étais partie. Je suis tellement reconnaissante pour tous les merveilleux souvenirs que nous avons créés ensemble, et j'ai hâte de voir où la vie nous mènera. Je t'aime d'amour.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

AR	accelerated region
BP	base pair
BPI	Bon Portage Island
CGI	CpG islands
CpG	cytosine-phosphate-guanine
DEG	differentially expressed gene
DGE	differential gene expression
DNAm	DNA methylation
d_{XY}	absolute nucleotide divergence
F_{ST}	fixation index
HBD	homozygous-by-descent
Hi-C	high-throughput chromosome conformation capture
HiFi	high fidelity
Iso-Seq	isoform sequencing
kya	thousand years ago
LGM	Last Glacial Maximum
LI	Long Island
LIS	Laurentide Ice Sheet
LR	linked read
MSMC	multiple sequentially Markovian coalescent
mya	million years ago
N_e	effective population size
NS	Nova Scotia
ON	Ontario
PCA	principal component analysis
SC	Sandy Cove
SFS	site-frequency spectrums
SMC	sequentially Markovian coalescent
SNP	single nucleotide polymorphism
SR	short read
SV	structural variants
WGS	whole genome sequencing
π	nucleotide diversity

CHAPTER 1: General Introduction

Diversification of mammals

Following the extinction of non-avian dinosaurs 66 million years ago, mammals rapidly diversified to occupy the newly available ecological niches (Hunter, 2020). Since then, their evolutionary trajectories have been shaped by a combination of global climatic and geological changes, such as shifting continents, climate oscillations, habitat transformations, and region-specific pressures (Steinthorsdottir et al., 2021). More recently, the glaciation cycles of the Quaternary Period (2.85 mya to present) influenced the demographic and adaptive histories of species (Fu & Wen, 2023; Hewitt, 2000; Shafer et al., 2010). Advancing and retreating ice sheets repeatedly fragmented habitats, created refugia, and temporarily exposed land bridges (Hewitt, 2000). These dynamic geological processes shaped patterns of movement and migration across and between continents. In many species, glacial cycles promoted population isolation leading to reduced gene flow, enhanced genetic drift, and local adaptation (Avila-Cervantes et al., 2021; Hansen et al., 2023; Howell et al., 2025; Sherlock et al., 2025). Over time, these processes have resulted in genomic and phenotypic divergence, playing a key role in shaping global biodiversity.

Island populations tend to exhibit behavioural, morphological, and ecological differences compared to their mainland counterparts (Baeckens & Van Damme, 2020). Such differences can be attributed to, but not limited to, changes in predation, competition, and resources (Benítez-López et al., 2021; Lomolino et al., 2013). This phenomenon, known as the island syndrome, has been observed across a wide range of taxa (Jezierski et al., 2024; Matthews, 2023). For instance, island foxes (*Urocyon*

littoralis) on the Channel Islands are significantly smaller than their mainland conspecifics (Wayne et al., 1991), and conversely, the Flores giant rat (*Papagomys armandvillei*) in Indonesia exemplify island gigantism (Locatelli et al., 2012). These examples highlight the role of islands in shaping species traits and driving evolutionary divergence, where isolation and selective pressures foster distinct genetic and phenotypic outcomes.

Omic tools to study evolution

Investigating how species and populations diverge and evolve requires integrating information from multiple molecular layers, including the genome, epigenome, and transcriptome. Divergence at the species, population, and individual levels is shaped by a range of mechanisms operating across these layers (Snead & Clark, 2022). Understanding how these molecular components contribute to evolution begins with the central dogma of molecular biology, which outlines the directional flow of information from DNA to RNA to protein (Crick, 1970). This framework links heritable genetic variation (genome) to regulatory modifications (epigenome) and gene expression patterns (transcriptome), ultimately shaping phenotypic traits and influencing diversity and adaptation.

The genome, which is the complete set of an organism's hereditary information encoded in DNA, forms the instructions for biological function and is the fuel of evolutionary change. In recent years, large-scale initiatives have prioritized the assembly of high-quality genomes across diverse species (e.g. Lewin et al., 2018; Rhie et al., 2021; Teeling et al., 2018). These assembled genomes are the foundation for investigating evolutionary relationships through comparative genomics and serve as

references to study genetic variation within and between populations (Lewin et al., 2018). Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation and serve as valuable markers for detecting adaptive changes in a genome (Syvänen, 2001). In some cases, genomic regions that are typically highly conserved across related species exhibit an unusually high density of SNPs in a specific lineage (Ferris et al., 2018; Pollard, Salama, King, et al., 2006). These segments are known as accelerated regions (ARs) and often reflect signatures of positive selection or lineage-specific adaptation (Pollard, Salama, King, et al., 2006). For example, human ARs are the fastest-evolving regions of the human genome and have been associated with important processes such as neurodevelopment and cognitive function (Girskis et al., 2021; Pollard, Salama, King, et al., 2006). In addition to SNPs, structural variants (SVs), which include deletions, insertions, duplications, and inversions of 50 base pairs or more, are also a source of genomic variation (Mahmoud et al., 2019). Though less common, SVs often can have important functional impacts, influencing gene structure, dosage, or regulation (Alonge et al., 2020; Hurles et al., 2008). In comparative genomics, SVs complement SNP analysis by revealing large-scale genomic changes that drive species-specific traits and adaptations, offering additional insight into the structural and evolutionary dynamics of genomes.

Genomic data are also fundamental to population-level studies, where they are used to infer phylogenetic relationships, reconstruct demographic histories, and investigate how evolutionary forces, such as natural selection, genetic drift, and gene flow, drive population divergence over time (e.g. Bourgeois & Warren, 2021; Kessler & Shafer, 2024; Martchenko & Shafer, 2023). Metrics such as F_{ST} quantify relative genetic differentiation between populations, while absolute divergence can be

measured by statistics such as d_{XY} (Irwin et al., 2018). Within populations, diversity is captured by metrics such as nucleotide diversity (π) and runs of homozygosity (ROH), which reflect levels of genetic variation and inbreeding (Irwin et al., 2018; Shafer & Kardos, 2025). Genetic health can be further assessed through GERP scores and mutation load, which provide insights into the burden of deleterious mutations (Davydov et al., 2010; Henn et al., 2015). Additionally, Tajima's D help detect deviations from neutral evolution, indicating potential selection or demographic shifts (Korneliussen et al., 2013). Collectively, these approaches and many more provide insights into the evolutionary dynamics shaping populations through time.

The epigenome consists of chemical modifications to DNA that regulate gene expression without altering the underlying DNA sequence (Al Aboud et al., 2021), and has been associated with divergence between species and populations (Meröndun et al., 2019; Vernaz, Malinsky, et al., 2021). These modifications include non-coding RNAs, histone modifications, and DNA methylation (Moore et al., 2013). Non-coding RNAs, while not encoding proteins, play essential regulatory roles by influencing chromatin structure and guiding processes like DNA methylation and histone modifications (Rinn & Chang, 2012). Histone modifications affect chromatin structure by changing how DNA interacts with histone proteins, which controls the accessibility of DNA for transcription (Al Aboud et al., 2021). DNA methylation, which primarily occurs through the addition of methyl groups to cytosine bases at CpG sites (Moore et al., 2013), can either suppress or activate gene expression depending on where these changes take place in the genome (Dhar et al., 2021; Wan et al., 2015). Together, these epigenetic mechanisms dynamically shape gene activity and cellular function throughout an organism's lifetime. Furthermore, methylation status follows predictable

patterns as organisms age which can be used to age individuals (Horvath, 2013; Prado et al., 2021; Wilkinson et al., 2021). These epigenetic clocks can predict chronological age in different mammalian species (Jasinska et al., 2021; Prado et al., 2021; Raj et al., 2021; Schachtschneider et al., 2021). While methylation data can be used to estimate age, instances where epigenetic age outpaces chronological age reflect a phenomenon known as epigenetic age acceleration (Horvath, 2013). This can be influenced by genetics, environmental exposures, lifestyle factors, or disease state (Oblak et al., 2021). These patterns offer insight into biological aging and population-level differences in health and longevity.

The transcriptome is the complete set of RNA molecules produced by a cell or tissue at a given time (Dong & Chen, 2013). It represents the temporal and spatial dynamics of gene activity, revealing the extent to which genes are being expressed in specific tissues under varying conditions (Dong & Chen, 2013). The transcriptome is highly dynamic and plastic in nature, constantly changing in response to developmental cues, environmental stimuli, and regulatory influences from both genomic and epigenetic factors, shaping phenotypes (Bogan & Yi, 2024). One of the primary tools for studying the transcriptome is RNA sequencing (RNA-seq), which enables genome-wide quantification of gene expression (Z. Wang et al., 2009). This allows researchers to identify differentially expressed genes (DEGs) between tissues, individuals, populations, and species (Fair et al., 2020; Lin et al., 2016). For example, RNA-seq has been used to uncover gene expression differences underlying altitude adaptation in highland versus lowland populations of deer mice (*Peromyscus* spp.), providing insights into the molecular basis of physiological adaptation (Cheviron et al., 2012)

Together, genomic, epigenomic, and transcriptomic data provide complementary insights on the mechanisms driving divergence and adaptation. Integrating these layers connects genetic variation with regulatory and expression changes that can drive phenotypic diversity and shape evolution across species and populations (Cai et al., 2021; M. S. Clark et al., 2023; Talavera et al., 2025). However, disentangling these complex relationships remains challenging. While integrative analyses have been performed at multiple biological levels, identifying consistent, overlapping trends across datasets is often complex (e.g. L. U. Gleason et al., 2023; Luo et al., 2025; T. Zhao et al., 2024). Yet, when signals do overlap, they reveal associations with functionally significant pathways under selection, offering insights into the molecular basis of adaptation (Lai et al., 2025; C. Li et al., 2024).

Sequencing technologies

Rapid technological advancements and decreasing costs have led to a variety of new library preparation and sequencing methods, each designed to overcome the limitations of earlier approaches (Mardis, 2017). Conventional (Illumina) short-read sequencing produces reads between 50 to 300 bp in length and offers high accuracy; while it is the standard for population resequencing studies, short-reads on their own fail to produce highly contiguous assemblies of complex genomes (Alkan et al., 2011). Short read lengths have limited accuracy in assembly of repetitive regions leading to fragmented contigs and incomplete genome reconstructions (Ji et al., 2011). To address this, technologies like mate-pair and 10x Genomics linked-read sequencing were developed. Both methods provided long-range genomic information by grouping of reads from the same DNA molecule, facilitating scaffolding, haplotype phasing,

structural variant detection, and ultimately *novo* genome assembly (Spies et al., 2017). However, mate-pair sequencing is no longer the preferred method in many modern genome assembly pipelines and the discontinuation of 10x Genomics' linked-read technology has created a gap that is increasingly being filled by high-accuracy long-read approaches.

Long-read sequencing such as PacBio high-fidelity (HiFi) have a 99.9% accuracy for long (10 - 25 kb) single-molecule reads (Wang (王博) et al., 2025). Circular DNA molecules called SMRTbell templates are repeatedly read by a DNA polymerase, allowing the same DNA fragment to be sequenced multiple times (Rhoads & Au, 2015). These multiple passes are then used to build a consensus sequence, correcting random sequencing errors and producing what are known as HiFi reads (Wang (王博) et al., 2025). The length and accuracy of HiFi reads make them well-suited for *de novo* genome assembly, detection of structural variants, and comprehensive characterization of complex genomic regions (Wang (王博) et al., 2025). In fact, long-read data form the backbone (i.e. primary sequence) of most modern assembly pipelines (e.g. Rhie et al., 2021).

Long-read sequencing data still require scaffolding, ideally to the chromosome level, to achieve complete and accurate genome assemblies. Hi-C sequencing can be used capture the three-dimensional (3D) organization of the genome within the nucleus. It works by capturing and sequencing pairs of DNA fragments that are physically close to each other in 3D space (Korbel & Lee, 2013). This is done by crosslinking chromatin, digesting the DNA, ligating nearby fragments, and then sequencing the resulting chimeric fragments (Belton et al., 2012). The data is used to

generate a contact map that reveals how different regions of the genome interact with each other (Yaffe & Tanay, 2011). Hi-C can be paired with other sequencing methods to assemble genomes as it helps with chromosome level scaffolding and haplotype resolution (Zeng et al., 2024). For example, the Vertebrate Genomes Project uses Hi-C scaffolding of HiFi contigs to produce chromosome-scale assemblies, resulting in some of the most complete vertebrate genomes to date (Rhie et al., 2021).

Quantifying epigenetic variation, specifically DNA methylation, can be conducted using a suite of techniques. While PacBio sequencing can directly detect methylation, most assays require converting unmethylated cytosines into uracils (typically via bisulfite or enzymatic reaction), while methylated cytosines remain unchanged (S. J. Clark et al., 2006). Methylation status of sites can then can be assessed using methylation arrays or sequencing, which allow for high-throughput quantification of DNA methylation at specific CpG sites across the genome (Arneson et al., 2022; Morris et al., 2014). The result is a high-resolution methylation profile for each sample, typically reported as β-values ranging from 0 (unmethylated) to 1 (fully methylated) (Bibikova et al., 2009). While direct sequencing is an option, challenges with library preparation and SNP calling have made high-density arrays a popular alternative (Arneson et al., 2022).

RNA sequencing (RNA-seq) is a widely used technique to analyze the transcriptome and historically has used short-read Illumina sequence. The process begins with RNA extraction, followed by conversion into complementary DNA, which is then fragmented and sequenced (Z. Wang et al., 2009). The resulting short reads are mapped to a reference genome or assembled *de novo* to quantify gene expression levels, identify differentially expressed genes, or detect novel transcripts (Z. Wang et

al., 2009). However, due to the limited length of these reads, accurately reconstructing full-length transcripts and distinguishing between closely related isoforms can be challenging (Steijger et al., 2013). In contrast, Isoform Sequencing (Iso-Seq) uses long-read sequencing technology to capture full-length RNA transcripts without the need for assembly (Minio et al., 2019). This approach enables identification of alternative splicing, transcript boundaries, and novel isoforms by capturing full-length transcripts, making it powerful for analyzing complex transcriptomes and improving genome annotations (L. Zhao et al., 2019).

Study system

Shrews (Soricidae), within the order Eulipotyphla, first appeared over 50 million years ago (Yuan et al., 2024) and have become one of the most diverse mammalian taxa with more than 400 extant species (Burgin et al., 2018). They are classified into three living sub-families, with the most speciose being Crocidurinae (white-toothed shrews), followed by Soricinae (red-toothed shrews), and Myosoricinae (African shrews) (George, 1986, 1988; Hutterer, 2005). Shrews exhibit remarkable ecological versatility, having colonized a wide range of habitats across the globe (Churchfield, 1990), which has led to diverse physiological and behavioral adaptations between species and populations. For example, some species have developed echolocation to navigate their environment (Chai et al., 2020; Forsman & Malmquist, 1988; Tomasi, 1979), while others possess venomous saliva used to subdue prey (Kita et al., 2004). Aquatic shrews dive underwater to hunt (Mendes-Soares & Rychlik, 2009), and many species show seasonal metabolic changes (Thomas, Baldoni, et al., 2025) and reversible reductions in skull morphology, body mass and organ size during winter to

conserve energy, followed by regrowth in the spring and summer (Lázaro & Dechmann, 2021). Shrews are among the smallest, shortest-lived mammals (Churchfield, 1990) and have an extremely high mass-specific metabolic rate (Ochocińska & Taylor, 2005) requiring them adapt physiologically to seasonal variations in resource availability and environmental conditions.

Bon Portage Island (BPI) lies approximately 3 km off the southwest coast of Nova Scotia, Canada. The island is about 3 km in length and less than 1 km in width, encompassing roughly 340 acres of undeveloped land. It is believed to have been separated from the mainland around 20 kya (Roland, 1982; Stewart & Baker, 1992). The island's shores are rich in marine algae, which wash up and create ideal conditions for dense populations of invertebrates such as kelp flies (*Coelopidae*) and sand fleas (*Platorchestia platensis*). The masked shrew (*Sorex cinereus*) population on the island travel up to 60 meters from the island's interior to the shoreline to feed on these invertebrates, thus transferring significant amounts of energy back into the island's terrestrial ecosystem (Stewart et al., 1989). They are the only *Sorex* shrews known to feed on sand fleas (MacPherson & Stewart, 2003; Stewart et al., 1989). Also, previous work found that BPI shrews have a unique allele for the cytosolic non-specific dipeptidase 2 (*CNDP2*) gene (MacPherson & Stewart, 2003), involved in intracellular amino acid and dipeptide metabolism (Pfeffer et al., 2024). It is hypothesized that this unique allele reflects an adaptive response, enhancing their ability to digest sand fleas more efficiently (MacPherson & Stewart, 2003; McAlpine, 2009).

Thesis objectives and data chapters

When I began my PhD in 2020, there were few genomic resources available for shrews, limiting opportunities to explore the mechanisms driving their evolution. Since then, the field has advanced significantly, with the addition of several new genome assemblies and an increasing number of studies focused on shrews. I have contributed to this by generating and annotating four *de novo* shrew genomes. Alongside this, I have developed and applied genomic, epigenomic, and transcriptomic tools to investigate patterns of evolution and adaptation in shrews at a species and population level, with a focus on BPI shrews, enabling deeper insights into their diversification, adaptation, and unique evolutionary trajectories.

In **Chapter 2** I conducted a comparative genomic analysis of four shrew species against 16 other mammals to identify lineage-specific variants in shrews. I used two publicly available shrew genomes and generated two additional *de novo* assemblies for the maritime (*Sorex maritimensis*) and smoky (*Sorex fumeus*) shrews. I identified accelerated regions, gene families undergoing significant size change, and positively selected genes to provide insights into the evolutionary mechanism driving shrew adaptation.

Chapter 3 aimed to examine the relationship between geographical isolation and both morphological and epigenetic divergence between island masked shrews and their mainland counterparts. I analyzed morphological trends in body size and weight across populations. Then, using methylation data, I developed an epigenetic clock for the masked shrew to assess epigenetic aging rates and quantified differentially methylated patterns between populations.

Expanding on the previous chapter, in **Chapter 4** I assembled and annotated a high-quality *de novo* genome for the masked shrew using HiFi and Hi-C data to investigate the evolutionary dynamics of the insular population on BPI. I reconstructed and modeled the demographic history of BPI shrews, examined how genome-wide SNP variation and differential gene expression contribute to their divergence from the mainland, and evaluated the extent to which these molecular mechanisms overlap.

Overall, my PhD research aimed to develop genomic resources and apply a multi-omics approach, integrating genomic, epigenomic, and transcriptomic data, as well as morphological data to investigate adaptation and divergence in shrews, with a focus on the masked shrew population from BPI.

CHAPTER 2: Comparative Genomics of the World's Smallest Mammals Reveals Links to Echolocation, Metabolism, and Body Size Plasticity

Authors:

Marie-Laurence Cossette, Donald T Stewart, Aaron B A Shafer

A version of this chapter was published in *Genome Biology and Evolution*

Contributions:

M.-L.C., A.B.A.S., and D.T.S. conceived the study and collected the samples. M.-L.C. and D.T.S. performed the molecular laboratory work. M.-L.C. performed the bioinformatic analyses with contribution from A.B.A.S. for the smoky shrew genome assembly. M.-L.C. and A.B.A.S. wrote the manuscript, and D.T.S. reviewed it.

Complete citation:

Cossette, M.-L., Stewart, D. T., & Shafer, A. B. A. (2024). Comparative genomics of the world's smallest mammals reveals links to echolocation, metabolism, and body size plasticity. Genome Biology and Evolution, 16(11).

https://doi.org/10.1093/gbe/evae225

Abstract

Originating over 50 million years ago, shrews (Soricidae) have diversified into around 400 species worldwide. Shrews display a wide array of adaptations, with some species having developed distinctive traits such as echolocation, underwater diving, and venomous saliva. Accordingly, these tiny insectivores are ideal to study the genomic mechanisms of evolution and adaptation. We conducted a comparative genomic analysis of four shrew species and 16 other mammals to identify genomic variations unique to shrews. Using two existing shrew genomes and two de novo assemblies for the maritime (*Sorex maritimensis*) and smoky (*Sorex fumeus*) shrews, we identified mutations in conserved regions of the genomes, also known as accelerated regions, gene families that underwent significant expansion, and positively selected genes. Our analyses unveiled shrew-specific genomic variants in genes associated with the nervous, metabolic, and auditory systems, which can be linked to unique traits in shrews. Notably, genes suggested to be under convergent evolution in echolocating mammals exhibited accelerated regions in shrews, and pathways linked to putative body size plasticity were detected. These findings provide insight into the evolutionary mechanisms shaping shrew species, shedding light on their adaptation and divergence over time.

Keywords

comparative genomics, genome assembly, accelerated regions, positive selection, gene duplication.

Introduction

Mammals have undergone successive episodes of rapid diversification resulting in the emergence of over 6,000 extant species worldwide (Burgin et al., 2018). Shrews (Soricidae) appeared over 50 million years ago (Yuan et al., 2024), and have become one of the most diverse mammalian taxa, having undergone numerous colonization, speciation, and extinction events (Reumer, 1989). Shrews consist of more than 400 extant species (Burgin et al., 2018) classified into three sub-families: Soricinae (red-toothed shrews), Crocidurinae (white-toothed shrews), and Myosoricinae (African shrews) (Hutterer, 2005). This diversification is also reflected in the diverse karyotypes both between (Schlitter et al., 1999) and within species (White et al., 2019; Wójcik et al., 2002), including sex chromosomes (Sharman, 1956).

Shrews have successfully colonized a diverse array of habitats across the globe (George, 1986, 1988) ranging from tropical forests to grasslands and arid regions (Churchfield, 1990). This has led to a range of adaptations, such as echolocation to navigate their surroundings (Chai et al., 2020; Forsman & Malmquist, 1988; Tomasi, 1979), aquatic diving capabilities to hunt prey (Mendes-Soares & Rychlik, 2009), venomous saliva for predation (Kita et al., 2004), and metabolic shifts (Thomas, Baldoni, et al., 2023) and reversible body size changes to survive winter (Lázaro & Dechmann, 2021). Shrews are among the smallest, shortest-lived mammals (Churchfield, 1990) and have an extremely high metabolic rate (Ochocińska & Taylor, 2005) which requires them to consume up to 125% of their body weight in food each day (Churchfield, 1990). These adaptations and overall diversity of shrews make for a unique system to investigate genome evolution in the context of mammalian diversity and evolution.

Approximately 10% of the human genome appears evolutionary constrained across mammals (Christmas et al., 2023). These highly conserved regions are predicted to be functionally important (Bi et al., 2023; Ponting, 2017), but may occasionally exhibit lineage-specific increases in nucleotide substitutions, which are referred to as accelerated regions (ARs) (Ferris et al., 2018; Pollard, Salama, King, et al., 2006). ARs are thought to contribute to species-specific traits (Ferris et al., 2018; Levchenko et al., 2018) and can result from evolutionary forces such as positive selection and GC-biased gene conversion (Bi et al., 2023; Ferris et al., 2018; Hubisz & Pollard, 2014; Pollard, Salama, King, et al., 2006). Ferris et al. (2018) and Tollis et al. (2021) identified ARs enriched near immune system and DNA damage response genes in elephants (*Loxodonta*), which are known to show resistance to cancer. Similarly, human ARs have been associated with proteins hypothesized to be important in neurodevelopment (Pollard, Salama, Lambert, et al., 2006).

On a larger scale, gene duplication provides new genetic material that can generate novel phenotypes that selection can act upon (Magadum et al., 2013). In bats (*Myotis*), repeated duplications of the protein kinase R (*PKR*) gene have been linked to immunity to viruses (Jacquet et al., 2022). Similarly, gene loss also has the potential to lead to phenotypic evolution and diversity (Helsen et al., 2020; Sharma et al., 2018). The loss of AMP deaminase 3 (*AMPD3*) in sperm whales (*Physeter macrocephalus*) has likely improved O_2 transport (Sharma et al., 2018). Other variants, such as larger structural rearrangements, also contribute to genomic diversity and evolution (Damas et al., 2021, 2022), highlighting the complex nature of the genomic architecture underlying phenotypic evolution.

Comparative genomics can be used to identify unique and shared genomic features, providing insight on the genetic basis of diversity and patterns involved in species-specific traits. Understanding gene family evolution, including adaptive expansion (i.e. duplication of genes) and adaptive contraction (loss of genes), has increasingly become a focus of molecular evolutionary genetics as the number of fully sequenced genomes has been increasing (Rhie et al., 2021). Here, we assembled two *de novo* shrew genomes for the maritime shrew (*Sorex maritimensis*), which is endemic to Canada (Stewart et al., 2002), and the smoky shrew (*Sorex fumeus*) (Figure S2.1a). Although restricted to Canada, the maritime shrew is part of the *Sorex* subgenus of *Sorex* shrews that are predominantly found in Eurasia (Fumagalli et al., 1999). The smoky shrew is a member of the subgenus *Otisorex*, members of which are predominantly found in the Nearctic region (George, 1988). We included two previously assembled shrew genomes for the Etruscan shrew (*Suncus etruscus*) and the Eurasian common shrew (*Sorex araneus*) and compared them against 16 other mammal genomes (Figure S2.1b). We aimed to uncover shrew-specific genomic changes associated with their distinctive phenotypes, with the prediction that shared genomic variants among shrew species will be associated with unique phenotypes and traits shared among these species. We focused on characterizing ARs, gene family duplications, expansion, and contraction events, and positively selected genes in shrew to provide insights into the evolutionary mechanisms driving shrew diversity and adaptation.

Results

Genome assembly and annotations

For the smoky shrew assembly (GenBank assembly accession: GCA_029834395.2), the five HiFi cells produced 115,806,042,303 bp of HiFi data, or ~43× coverage from the estimated 2.7 GB genome (Cossette et al., 2023). The mean length of the HiFi reads was 14,652 bp. We generated 368,278,690 paired Hi-C reads. The final smoky shrew assembly was 2.87 GB with an N50 of 42.40 Mb and over 99% of the genome being in scaffolds >50 kb long (Table 2.1). Approximately 51.9% of the genome was made of repetitive sequences (Table 2.2). The BUSCO assessment was 93.4% complete, of which 91.8% were single-copy. Distribution plots showed a high percentage of single-copy k-mers in the final assembly. We generated 890,398,361 raw 150 bp paired-end reads (~110× coverage) and 96,242,024,700 bp of 10× reads. The maritime shrew genome was 2.43 GB (GenBank assembly accession: GCA_030324115.1) and comprised of 91,327 scaffolds with an N50 of 84.5 kb. A total of 69.8% of the assembly was in scaffolds over 50 kb long, and 41.7% of the genome was classified as repetitive sequences. Out of the BUSCO orthologs from the mammalian data set, 72.9% were complete with 70.6% being single-copy. The potential impact of this lower quality genome assembly on subsequent analyses is discussed below (see Discussion).

A total of 20,380 protein-coding genes were identified in the smoky shrew genome and 21,191 in the maritime shrew genome. The assembled maritime and smoky shrew mitochondrial genomes were 16,979 bp and 17,051 bp long, respectively; all 37 mitochondrial genes were present in both assemblies.

Accelerated regions

We identified 598,180 conserved regions (~1% of the genome) (Table S2.1) in our alignments, comparable to other mammalian studies (Ferris et al., 2018; Tollis et al., 2021). From these regions, 307,989, ~51% of conserved regions, were in exons. We found 2,643 common shrew ARs in 1,627 genes, 21,581 maritime shrew ARs in 4,528 genes, 4,272 smoky shrew ARs in 2,257 genes, and 21,354 Etruscan shrew ARs in 6,787 genes (Table 2.3, Table S2.1). Soricidae shared 35 ARs, and the *Sorex* species shared 404 ARs (Figure 2.1a). Notable ARs were found in genes related to the nervous system such as the growth associated protein 43 (*GAP43*), fibroblast growth factor receptor 1 (*FGFR1*), and class III β-tubulin (*TUBB3*) (Table 2.4). Also, genes involved in metabolic pathways such as the adiponectin receptor 1 (*ADIPOR1*) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) had shared ARs in *Sorex* and Soricidae, respectively (Table 2.4). The common shrew and maritime shrew genomes shared ARs in the cadherin related 23 (*CDH23*) and otoferlin (*OTOF*) genes, which are associated with echolocation capabilities in mammals (Shen et al., 2012; Table 2.4). The pathway analyses revealed genes associated with nervous system development, olfactory receptors, insulin and energy metabolism pathways, muscle contraction, and cardiac conduction as well as keratinization (Figure 2.1, Figure S2.2).

Gene family size evolution

Orthofinder assigned 476,077 (97.0% of total) genes to 20,489 orthogroups. Of these, 7,954 orthogroups included at least one ortholog from each of the 20 species and 1,810 orthogroups were species-specific. A total of 533 genes were identified as

single-copy genes across all 20 mammals. The maritime shrew had 15,755 (74% of genes) genes assigned to orthogroups (Figure 2.2a) and had the most species-specific orthogroups (Figure 2.2b), which may be due to inaccurate gene predictions resulting from the fragmented assembly (Table 2.1). The three other shrew species' assemblies had on average 37 unique gene families each (Figure 2.2b). Orthofinder further identified 2,722 gene duplication events in the common shrew, 2,658 in the maritime shrew, 2,328 in the smoky shrew, and 2,306 in the Etruscan shrew (Figure 2.2c, Table 2.3). There were 42 gene duplication events shared between all shrew species and 62 between the *Sorex* genus (Figure 2.2c, Table 2.3). The number of gene families undergoing expansion ranged between 904 and 1,476 in shrews with up to 407 significantly expanding gene families in the common shrew (Figure 2.2d). The maritime shrew appeared to have undergone more contractions (n = 7,285) than any other shrew species which ranged between 629 and 1,598 contractions. This discrepancy may be due to incomplete gene predictions resulting from the fragmented assembly (Table 2.1). The orthogroup associated with the serine racemase (*SRR*) gene, which is involved in neurogenesis (Roychaudhuri et al., 2023), appeared to have undergone significant expansion in the *Sorex* shrews (Table 2.4). Other significantly expanded gene families in all Soricidae were associated with nervous system development genes and cellular response to stress and starvation, olfactory receptors, and the immune system (Figure 2.2e). The gene families that underwent significant contractions shared between species were related to olfactory receptors and blood group system pathways (Figure 2.2e).

Episodic diversifying selection in shrews

We calculated the ratio of non-synonymous to synonymous substitutions (dN/dS) using aBSREL to identify positive selection. Out of the 533 single-copy genes tested, nine were detected as being under positive selection in the common shrew, 97 in the maritime shrew, 31 in the smoky shrew, and 116 in the Etruscan shrew when the Soricidae common ancestor and all descendent branches were selected as the foreground branches (Figure 2.3a, Table 2.3). A total of 62 genes were detected as being under positive selection in the Soricidae node, and 47 were detected as being selected for in the *Sorex* (Figure 2.3a, Table 2.3). Ninety-four percent of these genes had trees supporting shrews as a monophyletic group. In all shrew nodes and species, except for the common shrew, the ankyrin 2 (*ANK2*) gene, involved in making proteins found in the brain and heart, was under positive selection (Figure 2.3b, Table 2.4). Other genes under selection included anoctamin 10 (*ANO10*), myosin IXB (*MYO9B*), and reticulon 4 receptor (*RTN4R*) which are all associated with the nervous system (Table 2.4). Top enriched pathways for positively selected genes were related to protein metabolism and various GTPases cycles (Figure 2.3c).

Discussion

Over the course of mammalian evolution, species have adapted to diverse ecological niches, evolving an array of specialized traits and behaviors. Shrews, in particular, have evolved unique phenotypes and abilities tailored to their various environments, including venomous saliva (Kita et al., 2004), echolocation (Chai et al., 2020; Forsman & Malmquist, 1988; Tomasi, 1979), hunting underwater (Mendes-

Soares & Rychlik, 2009), and reversible body size changes (Lázaro & Dechmann, 2021). Soricidae are also known to have one of the shortest lifespans and highest mass-specific metabolic rate of all mammals (Ochocińska & Taylor, 2005). Here, we assembled two *de novo* shrew genomes and conducted comparative genomic analyses between four shrews and 16 other mammal species to test our prediction that shared genomic variants among shrew species are likely associated with unique phenotypes and abilities specific to shrews. We identified ARs, gene family expansion and contractions, as well as genes undergoing positive selection in Soricidae.

Genome evolution underling phenotypic changes in shrews

Shrews have poor vision which is mostly used to detect light intensity; in contrast, their olfactory, tactile, and acoustic senses are well developed (Churchfield, 1990). Certain Nearctic shrew species, and the common shrew, have even developed echolocating abilities to navigate their surroundings (Churchfield, 1990; Forsman & Malmquist, 1988; Gould, 1969). We found shared ARs in the coding region of the *CDH23* gene and 3′UTR region of the *OTOF* gene in the common and maritime shrews (Table 2.4). Both genes are involved in the auditory system and more specifically echolocation in mammals (Y.-Y. Shen et al., 2012) . Chai et al. (2020) found evidence of convergent evolution between the common shrew and other echolocating mammals for both these genes. The common shrew is the only species included in our analyses confirmed to echolocate, but our findings indicate the potential for such ability in the maritime shrew, as it is the most closely related species and shares ARs with the common shrew for the *CDH23* and *OTOF* genes. Echolocation is likely more widespread among shrews (i.e. *Blarina*) (M. E. Gleason et al., 2023), including other *Sorex* (Buchler,

1976), thus detection in the maritime shrew is not unexpected and this might be a *Sorex*-wide trait.

Survival in cold climates involves species adopting strategies such as migration, hibernation, or entering a state of torpor. Shrews do not hibernate nor migrate long distances (Churchfield, 1990). Their high metabolic rate prevents them from building up sufficient fat storage (Churchfield, 1990); therefore, to survive through the winter, some shrew species have evolved a way to lower their energetical demand by undergoing reversible seasonal changes in body size and mass, known as Dehnel's phenomenon (Lázaro & Dechmann, 2021). During the winter, brain mass can shrink around 20% in the common shrew and regrow up to 17% for the summer (Lázaro, Hertel, et al., 2017). This has also been shown in the Etruscan shrew (Ray et al., 2020), which enters torpor (Fons et al., 1997), and other Sorex shrews (e.g. *Sorex minutus*) (Bartkowska et al., 2008; Lázaro & Dechmann, 2021). It remains uncertain if the maritime shrew and smoky shrew exhibit Dehnel's phenomenon, however both species belong to the *Sorex* genus and inhabit Northern climates, akin to other shrew species known to undergo these size changes.

We observed hundreds of ARs in shrews associated with nervous system development and axon guidance pathways (Figure 2.1b), including specific genes such as *TUBB3*, *GAP43*, and *FGFR1* (Table 2.4). In all *Sorex* species, ARs were found in the 5′UTR regions of the *TUBB3* and *GAP43* genes. *TUBB3* plays a role in the regeneration of axons (Latremoliere et al., 2018), while *GAP43* is involved in neurite outgrowth and synaptic plasticity in the hippocampus (Aigner et al., 1995; Y. J. Lee et al., 2023), a region of the brain known to shrink during the winter and regrow in the spring in the common shrew (Lázaro et al., 2018). Furthermore, ARs were identified in the coding

region of the *FGFR1* gene, which is also involved in hippocampal growth (Ohkubo et al., 2004). These findings were further supported by the gene family expansions. Gene families associated with the *SRR*, *PURA*, and *UGT8* genes, which are all involved with the brain and the nervous system (Table 2.4), had undergone significant expansion in all *Sorex* species. And in all Soricidae species, significantly expanded gene families were associated with SLITs and ROBOs pathways (Figure 2.2e) which are involved in neocortical development, a region responsible for sight and hearing (Gonda et al., 2020). The pathway analysis further revealed that positively selected genes in shrew were related to Rho GTPase pathways (Figure 2.3c). Rho GTPases, such as RhoA, Cdc42, and Rac1, are involved in neuronal development, neurodegeneration, and synaptic plasticity (Stankiewicz & Linseman, 2014; H. Zhang et al., 2021). Finding such genes and pathways associated with the brain and nervous system consistently in all analyses and shared between all species, we hypothesize, is indicative of the genetic mechanisms involved in seasonal brain size plasticity.

Shrews might also undergo a metabolic shift from lipid to glucose as a fuel source to survive the winter. Work by Thomas et al. (2023) found evidence that lipid metabolite concentration decreased throughout the winter in the common shrew, possibly promoting carbohydrate metabolism during the harsh winter months. Thomas et al. (2023) found differentially expressed genes and pathways associated with insulin and cholesterol between shrew brain and liver samples across seasons. When consumed, carbohydrates breakdown into glucose, and insulin then regulates if it is used as a source of energy or stored (Dube et al., 2013). We identified ARs and positively selected genes in shrews associated with Rho GTPase pathways (Figure 2.3, Figure S2.2). Rho GTPase are also involved in metabolic homeostasis more specifically

glucose metabolism (Møller et al., 2019). Furthermore, an AR was identified in the 3′UTR region of the *ADIPOR1* gene in *Sorex* (Table 2.4). *ADIPOR1* plays a crucial role in the regulation of glucose and lipid metabolism (Yamauchi et al., 2007). All shrew species shared an AR in the *GAPDH* coding region, which is involved in glycolysis, the process that breaks down glucose into energy (Table 2.4). Furthermore, gene families that underwent significant expansions were associated with cellular response to stress and starvation (Figure 2.2e). These findings highlight the relationship between genomic variants unique to shrews and their distinctively fast metabolism and wintering strategy and offers candidate genes for further hypothesis testing (Table 2.4).

Addressing bias from assembly and annotation differences

The divergent assembly strategies and genome quality statistics (Table 2.1) presented us with a natural study to compare the impact of reference genome that warrants comment. Assembly inconsistencies between genomes can lead to overestimation of genomic differences between taxa when conducting comparative genomic analyses (Denton et al., 2014). Our shrew assemblies consisted of three highly contiguous genomes (Etruscan, smoky, and common shrew), and one fragmented genome (maritime shrew). In the AR analysis, the maritime shrew and Etruscan shrew assemblies revealed a high number of ARs; however, we found no clear pattern between number of ARs and genome assembly quality or contiguity (Figure S2.3). For good measure, we repeated the analysis to identify ARs in the African elephant (*Loxodonta africana*) and big brown bat (*Eptesicus fuscus*) and compared these values to the study done by Ferris et al. (2018). Our results were consistent with

their work, with a significant lower number of ARs in the African elephant (n = 1,904) compared to the bat (n = 20,939; Table S2.1). Thus, the AR analysis appears to not be impacted by highly fragmented assemblies like the maritime shrew, likely due to the bioinformatic approach of mapping short fragments.

The maritime shrew had fewer genes assigned to orthogroups and more species-specific orthogroups compared to all other shrews. Here, this is likely due to the maritime shrew's lower genome contiguity and not having a standardized RefSeq annotation like all other species. Depending on the sequencing and assembly method, *de novo* genome assemblies can be fragmented resulting in inaccurate structure and number of predicted genes (Denton et al., 2014). However, despite detecting considerably more orthogroups in the maritime shrew, the number of duplicated regions and genes under positive selection did not deviate significantly from the Etruscan shrew (Table 2.3), suggesting that none of our metrics were drastically impacted by fragmentation. Still, to avoid biases due to assembly and annotation differences between species, we focused on overall trends found in Soricidae and *Sorex* instead of species-specific trends. This minimizes overestimation of ARs, gene duplication event, or gene family expansion/contraction and positively selected genes identified in shrews. Overall, our study found shrew-specific genomic variants in genes associated with the nervous, auditory, metabolic, and olfactory systems, most of which can be linked to unique phenotypes or traits in shrews.

Materials and Methods

Sampling and sequencing

A single smoky shrew (*S. fumeus*) was live captured from Peterborough, Ontario (Animal Care Certificate no. 26234). Heart, liver, brain, and tail tissue were collected immediately after euthanasia and frozen. DNA was extracted from the heart and liver using the *MagAttract HMW DNA Kit* from QIAGEN. Pooled DNA extracts were sent to The Centre for Applied Genomics (TCAG) in Toronto, Ontario, Canada, to sequence five HiFi cells on the PacBio Sequel II instrument. Tissue samples were sent to Phase Genomics (Seattle, Washington) to generate a Hi-C library using the Phase Genomics Proximo Animal kit version 4.0.

Genomic sequencing of the maritime shrew was undertaken as part of the CanSeq150 program (https://www.cgen.ca/canseq150). A single maritime shrew (*S. maritimensis*) was live captured and euthanized following the guidelines of the Canadian Council on Animal Care see Dawe et al. (2009). Liver tissue was collected immediately and frozen. Genomic DNA was extracted from frozen tissues using a standard proteinase K-phenol-chloroform protocol (Sambrook et al., 1989). DNA and tissues were stored at −80°, and DNA extracts from this maritime shrew were sent to TCAG for sequencing. Short-read (SR) libraries were generated using PCR-free preparation and sequenced on two lanes on the Illumina HiSeqX instrument and with 150 bp paired-end reads. Linked reads (LR) were prepared using the 10× genome library after selecting fragments > 15 kb using BluePippin. The LR were sequenced on one Illumina HiSeq X Lane with 150 bp paired-end reads.

Genome assemblies and annotation

We assembled the smoky shrew reference genome following the updated Vertebrate Genome Pipeline (Rhie et al., 2021). The HiFi bam files were filtered using bamtools version 2.5.1 (Barnett et al., 2011) (rq ≥ 0.99) and converted to fastq files using samtools version 1.12 (Danecek et al., 2021) via the fastq flag. We trimmed residual HiFi adapter sequences using cutadapt version 3.4 (M. Martin, 2011). The primary assembly was constructed using hifiasm version 0.16.1-r375 (Cheng et al., 2021, 2022) allowing for the integration of the Hi-C reads. We used the default level of purge duplication of three for non-trio assembly. We ran the hifiasm python module on 32 threads with 228 GB of RAM. We used Merqury version 1.3 (Rhie et al., 2020) with meryl from Canu version 2.2 (Koren et al., 2017) to assess genome assembly quality via k-mer copy number analysis.

For the maritime shrew genome, we used BBMap version 35.8 (Bushnell, 2022) to remove adapters and trim low-quality data from the fastq files. Kmergenie version 1.7048 (Chikhi & Medvedev, 2014) was used to find the k-mer size for downstream analysis. The genome was assembled using a tiered approach: (i) using only the SR data with w2rap-contigger with -K set to 144 (Clavijo, 2021) and (ii) with only the 10X LR data using Supernova version 2.1.1 (Weisenfeld et al., 2018) using the --maxreads filter set to "all". The pseudohap2 style was selected for the assembly output. Following these assemblies, both SR and LR assembly versions were merged via quickmerge version 0.3 (Chakraborty et al., 2016) with the LR as the backbone and the default parameters. Scaff10x version 5.0 (Ning et al., n.d.) was used to further polish the maritime shrew genome.

Nuclear genome completeness was assessed for both assemblies using Benchmarking universal single-copy orthologues (BUSCO) version 3.0.2 (Simão et al., 2015) by comparing the genomes to highly conserved genes in mammals. The smoky shrew mitochondrial genome was assembled using the PacBio subreads and the mitochondrial genome of a previous smoky shrew assembly (NCBI accession: GCA_026122425.1) (Cossette et al., 2023) as a backbone with MitoHifi version 3.2 (Uliano-Silva et al., 2023). The mitochondrial genome for the maritime shrew was assembled using MitoZ (Meng et al., 2019a) with the function "mitoz all" on the fastq reads and parameters --genetic_code 2 --clade Chordata --kmers_megahit 59 63 79 99 119 141 --requiring_taxa Chordata and was manually curated.

The smoky shrew genome annotation was generated by NCBI using their eukaryotic genome annotation pipeline that integrated RNAseq data (SRA RNA-Seq accession: SRX20204431, SRX20204430) from Cossette et al. (2023). We generated the maritime shrew annotation in-house via GenSAS version 6.0 (Humann et al., 2019) as it was too fragmented for the NCBI pipeline. Repeat regions were identified using RepeatModeler version 2.0.1 (Smit & Hubley, 2023b) and RepeatMasker version 4.1.1 (Smit & Hubley, 2023a). Here, we used the smoky shrew liver and heart RNA reads from Cossette et al. (2023), as we did not have RNA data for the maritime shrew, and mapped them to the maritime shrew genome using HISAT2 version 2.2.1 (Kim et al., 2019). The resulting BAM files were used by Augustus version S3.4.0 (Stanke et al., 2006) for gene prediction. The NCBI refseq vertebrate-mammalian protein database available on GenSAS was aligned to the genome using DIAMOND version 2.0.11 (Buchfink et al., 2015). Augustus and DIAMOND were also run using the common shrew, mSorAra2.pri (GenBank accession: GCF_027595985.1), protein fasta file.

EVidenceModeler version 1.1.1 (Haas et al., 2008) was used to generate a consensus gene set using the Augustus (1× weight) and DIAMOND (5× weight) outputs. Gene function was assigned to our gene consensus model with DIAMOND and InterProScan version 5.53-87.0 (P. Jones et al., 2014).

Detection of ARs in shrew genomes

We generated a multiple alignment file (MAF) consisting of 20 mammalian genome assemblies to identify ARs in the four shrew species. We downloaded the pairwise syntenic net alignment files for the 16 mammals, including the common shrew, against the human (hg38) genome (Table S2.2) from the UCSC database (Kent et al., 2002). Using LASTZ version 1.04.03 (Harris, 2007), we generated pairwise alignments against the hg38 genome for the smoky, maritime, and Etruscan shrew genomes. The LASTZ run parameters were set to K = 3,000, L = 3,000, Y = 9,400, E = 30, H = 2,000, and O = 400 based on UCSC's hg38 100-way conservation parameters for their common shrew alignment. The alignments for each species were chained with the chainMinScore = 3000 and linearGap = medium options and subsequently netted using kentutils tools version 401 (UCSC Genome Browser, 2020). Our resulting pairwise alignments for each shrew species were then aligned to the other mammal pairwise alignments to create one MAF using the roast function from MULTIZ version 11.2 (Blanchette et al., 2004) and the tree topology in Figure S2.1b, supported by Murphy et al. (2001) and Meredith et al. (2011).

The MAF was filtered to only keep alignments to the 22 main autosomal chromosomes of the hg38 genome. We used the msa_view function from the PHAST package version 1.5 (Hubisz et al., 2011) to extract 4-fold degenerate (4D) sites from

the non-shrew species in our alignments and based on the human annotations (NCBI accession: GCA_000001405.15) from which we only kept coding sequences. The output was used to estimate a non-conserved phylogenetic model using phyloFit with the substitution model REV and the EM algorithm option. The PhastCons function from RPHAST version 1.6.11 (Hubisz et al., 2011) was used to identify conserved regions in the non-shrew species with the parameters set to expected.length = 45, target.coverage = 0.3, rho = 0.31, and viterbi = TRUE. We filtered the predicted conserved regions to sites that aligned to all the shrew species and at least 18 species in total. We regularized the length of the conserved regions to 50 bp regions to simplify the likelihood ratio tests. PhyloP was run using the acceleration (ACC) mode with the --features option to identify ARs within the predicted conserved regions for each of our shrew species. This was done by running phyloP sequentially on each shrew terminal branch using the full species alignment. Nonparametric simulations were run to calculate empirical P-values by generating 100,000 synthetic alignments and running phyloP on these alignments to obtain a null distribution of log likelihood ratios. Statistically significant ARs were defined with a false discovery rate threshold of 5%. Genes overlapping these ARs were identified using bedtools version 2.30.0 (Quinlan & Hall, 2010) and the hg38 genome annotations.

Gene family size evolution

We identified orthogroups among protein sequences from all 20 mammal genomes (Figure S2.1b, Table S2.2) to identify gene duplication events. Here, an orthogroup is a set of genes that have descended from a single ancestral gene in the last common ancestor to our species of interest and all other mammalian genomes in

the set (Emms & Kelly, 2019). We downloaded the protein files from NCBI RefSeq for each species (Table S2.2), except the maritime shrew for which we used our GenSAS annotation. We further filtered the data to keep the longest transcript variant per gene in each species' file to minimize run time and increase accuracy. OrthoFinder version 2.5.4 (Emms & Kelly, 2019) with the parameters -M msa, -S diamond, -A mafft, -z, -T fasttree was used to obtain 20,489 orthogroups and identify duplicated genes. We then used CAFE version 4.2 (De Bie et al., 2006) with a multi-lambda model to analyze the evolution of orthogroup sizes and used an error estimation model to account for genome assembly errors. A Monte Carlo resampling procedure was applied to each branch and node to compute family-wide P-values. For $P \leq 0.01$, the Viterbi method was used to calculate exact P-values and identify gene families that have experienced significant expansion ($P \leq 0.01$) (De Bie et al., 2006).

Non-synonymous to synonymous rate ratio

We used an adaptive branch site random effects likelihood (aBSREL) model (M. D. Smith et al., 2015) with the HyPhy package version 2.5.49 (Pond et al., 2005) to identify possible episodic diversifying selection in shrews by calculating the ratio of non-synonymous to synonymous substitutions (dN/dS). The analysis was performed on the 533 single-copy genes identified by OrthoFinder. Multiple sequence alignment for each protein was converted to sequence codon alignments using PAL2NAL version 14.1 (Suyama et al., 2006) to input in aBSREL using the coding sequence available on RefSeq (Table S2.2). We examined the individual gene trees for each single-copy gene alignment and found that 80% of these trees supported shrews as a monophyletic group. We used the consensus species tree output from the OrthoFinder analysis,

which is based on approximately 20,000 orthogroups and also includes shrews as a monophyletic group, as the input tree for all aBSREL runs. We ran the aBSREL analysis twice to estimate dN/dS, first (i) with the branch leading to the common ancestor of the shrews and all its descendent branches selected as the foreground; and (ii) with no foreground branches specification and all branches tested for positive selection. An R script (Jarva, 2023) was used to parse and extract data from the nested json file outputs.

Gene pathway analysis

Gene and protein summaries were manually retrieved from UniProt (Bateman et al., 2023) and NCBI (National Center for Biotechnology Information (NCBI), 1988). Genes that were associated with ARs, duplications events, statistically significant family size changes, and positive selection in shrews were linked to the corresponding human ortholog gene ID when possible, to input in DAVID, the Database for Annotation, Visualization and Integrated Discovery (D. W. Huang et al., 2009; Sherman et al., 2022), to identify gene pathways for each analysis.

Acknowledgments

We would like to acknowledge that the work for this study was carried out on the traditional territory of the Mississauga (Michi Saagiig) Anishinaabeg and the Mi'kmaq People. We are grateful to have had the opportunity to work on this land and thank the First Peoples for their care, stewardship, and teaching. We would like to thank The Centre for Applied Genomics in Toronto, Ontario, Canada and Phase

Genomics Seattle for their expertise and insight with sequencing as well as the reviewers and editor for their valuable feedback on this manuscript.

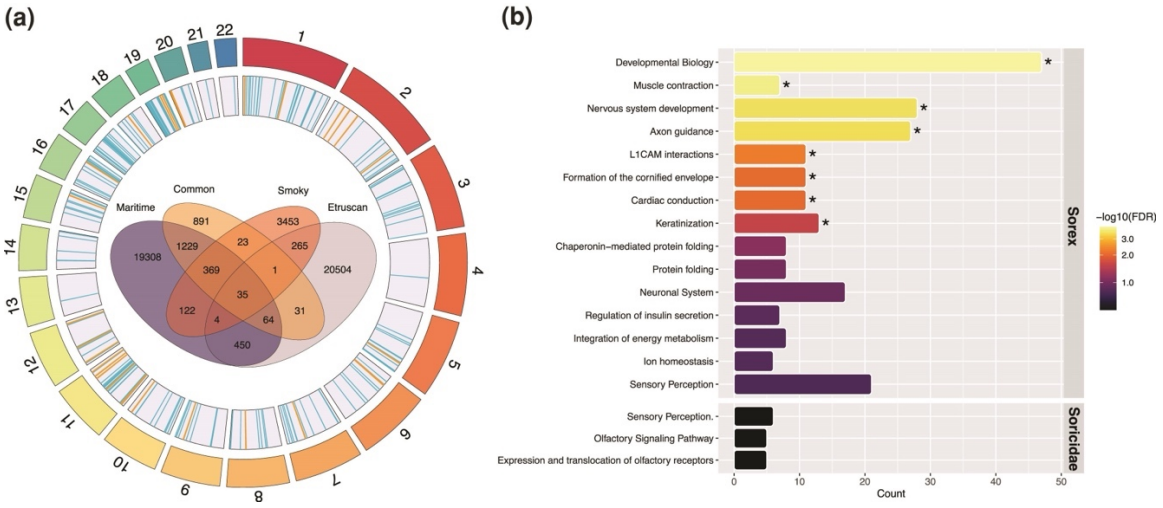

Figure 2.1. Accelerated region (AR) analysis results. (a) Location of ARs shared between Sorex species (n = 404) in blue and Soricidae (n = 35) in yellow mapped to the human (hg38) genome. Venn diagram for the number of overlapping ARs between shrew species. (b) Top Reactome pathways for overlapping ARs in Sorex and Soricidae. Asterisk represents significant pathways (5% FDR). Top pathways for individual shrew species can be found in the Supplemental (Figure S2.2).

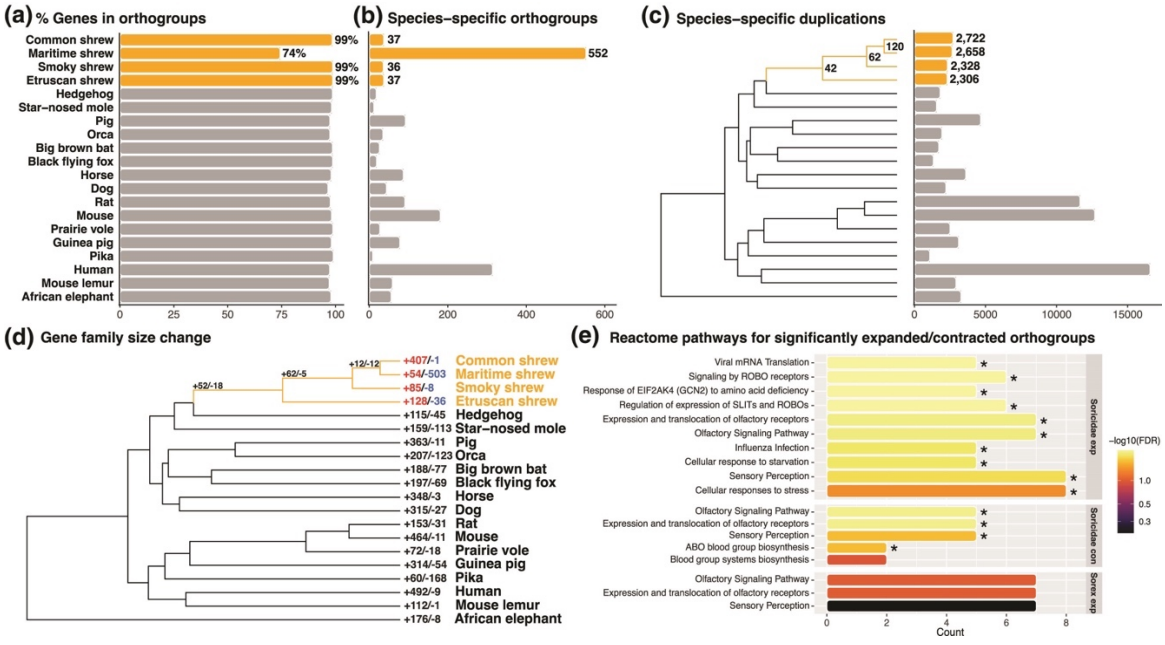

Figure 2.2. OrthoFinder and CAFE analyses results. (a) Percentage of genes from each species assigned to orthogroups. (b) Number of species-specific orthogroups. (c) Number of species-specific duplication events. Number of duplication events shared by all descendants of the shrew nodes are also represented. (d) Number of significantly (P-value < 0.01) expanded/contracted orthogroups across 20 mammalian species. Shrew nodes and branches are highlighted. (e) Top Reactome pathways for the significantly expanded/contracted orthogroups shared between Sorex and Soricidae species. Asterisk represents significant pathways (5% FDR).

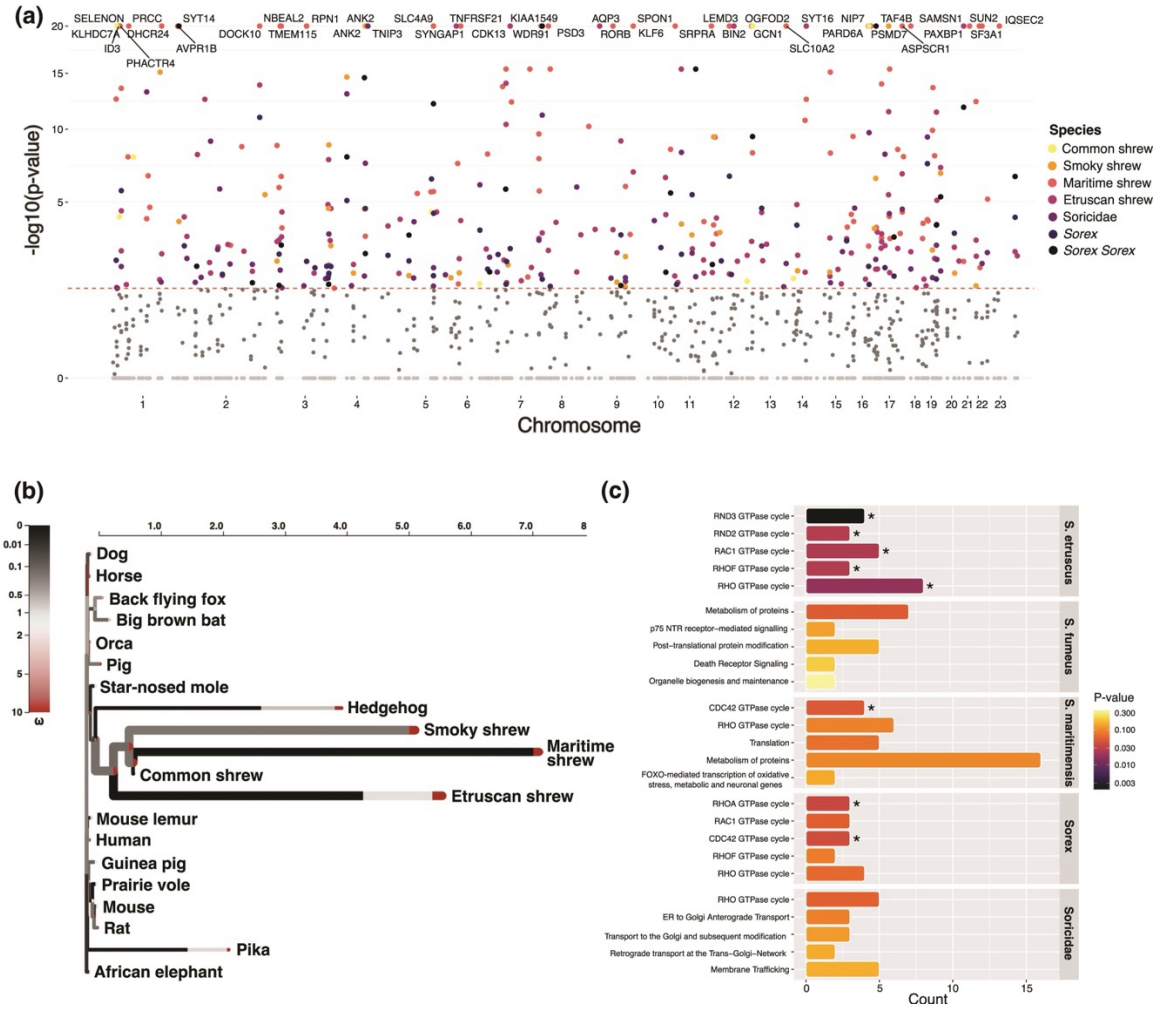

Figure 2.3. aBSREL (dN/dS) analysis results. (a) Manhattan plot of Bonferroni-Holm corrected P-values used as evidence for selection for all 533 single-copy genes from the foreground run mapped to the human (hg38) genome. Dark grey points represent genes with non-significant P-values for shrew nodes and species. Light grey points represent genes from all other mammal nodes and species. The red dotted line represents the 0.05 significance threshold. Top genes for shrew nodes and species are indicated by the respective name. (b) HyPhy Vision plot for the different omega (ω) rate distributions in each branch for the ANK2 gene. The color of the segments indicates the rate (ω), while the length of each segment represents the proportion of sites exhibiting that specific ω rate. Thicker branches highlight those identified as having undergone diversifying positive selection (corrected P-value < 0.05). (c) Top Reactome pathways for the positively selected genes from the dN/dS analysis in shrew nodes and species. Asterisk represents significant pathways (P-value < 0.05).

Table 2.1. Genome assembly statistics for the smoky shrew genome generated from PacBio HiFi long-reads and Hi-C reads, the maritime shrew genome generated from 10X long-reads and short-reads, and the common shrew and Etruscan shrew genomes downloaded from NCBI.

	Smoky shrew	Maritime shrew	Common shrew	Etruscan shrew
Main genome assembly size	2.87 GB	2.43 GB	2.42 GB	2.47 GB
Main genome number of scaffolds	499	91,327	12,478	147
Main genome scaffold N/L50	42.4 Mb/23	84.5 kb/7,135	22.8 Mb/36	131.9 Mb/8
Main genome scaffold N/L90	17.4 Mb/49	15.3 kb/29,605	4.2 Mb/129	87.5 Mb/17
Max scaffold length	91.7 Mb	2.5 Mb	60.2 Mb	208.2 Mb
Number of scaffolds > 50 kb	365	14,697	473	70
% main genome in scaffolds > 50 kb	99.9%	69.8%	98.4%	99.9%
Genome coverage	43.0×	76.5×	120.0×	142.6×
Number of protein-coding genes	20,380	21,191	19,080	19,819

Table 2.2. Summary of repeats in the smoky and maritime shrew genomes in length occupied (bp) and percentage of the genome.

	Smoky shrew	Maritime shrew
SINE	182,692,788 (6.36%)	141,858,027 (6.90%)
LINE	614,478,406 (21.38%)	366,060,918 (17.81%)
LTR elements	41,000,372 (1.43%)	30,015,478 (1.46%)
DNA transposons	10,884,988 (0.38%)	6,592,309 (0.32%)
Unclassified	500,269,535 (17.41%)	198,021,546 (9.63%)
Rolling-circles	686,527,20 (2.39%)	58,004,316 (2.82%)
Small RNA	20,490,319 (0.71%)	18,114,174 (0.88%)
Satellites	2,682,795 (0.09%)	2,057,146 (0.10%)
Simple repeats	39,822,407 (1.39%)	29,813,949 (1.45%)
Low complexity	9,066,276 (0.32%)	5,801,032 (0.28%)
Total	1,421,387,886 (51.86%)	856,338,895 (41.65%)

Table 2.3. Summary statistics from each analysis per shrew species/node.

	Smoky shrew	Maritime shrew	Common shrew	Etruscan shrew	*Sorex*	Soricidae
# accelerated regions	4,272	21,581	2,643	21,354	404	35
# gene duplications	2,328	2,658	2,722	2,306	62	42
# significant+/−orthogroups	+85/−8	+54/−503	+407/−1	+128/−36	+62/−5	+52/−18
# genes under selection	31	97	9	116	62	47

Table 2.4. Genes associated to the nervous system, auditory system, and metabolism in shrew species/nodes identified in each analysis based on the human hg38 annotations (NCBI accession: GCA_000001405.15).

	Node/species	Gene	Function
Accelerated regions	*Sorex sorex*	CDH23	Critical component of hair bundle formation (Palma et al., 2001)
	Sorex sorex	OTOF	Involved in vesicle release at the synapse between inner hair cells and auditory nerve fibers (Roux et al., 2006)
	Sorex	GAP43	Associated to neurite outgrowth and synaptic plasticity (Aigner et al., 1995; Y. J. Lee et al., 2023)
	Sorex	TUBB3	Involved in axon regeneration (Latremoliere et al., 2018)
	Sorex	CPNE6	Regulates dendritic spine structural plasticity, learning and memory (Reinhard et al., 2016)
	Sorex	FGFR1	Necessary for hippocampal growth (Ohkubo et al., 2004)
	Sorex	ADIPOR1	Plays a role in glucose and lipid metabolism (Yamauchi et al., 2007)
	Soricidae	GAPDH	Involved in glycolysis pathway (Seidler, 2013)
	Soricidae	CSNK2A1	Mutations associated with neurodevelopmental abnormalities (Okur et al., 2016)
Duplication/CAFE	*Sorex*	SRR	Involved in synaptic plasticity (Wong et al., 2020) and affects adult neurogenesis (Roychaudhuri et al., 2023)
	Sorex	PURA	Important for postnatal brain development (Khalili et al., 2003)
	Sorex	UGT8	Involved in the synthesis of glycosphingolipid of myelin (Dzięgiel et al., 2010)
dN/dS	Soricidae *Sorex* *Sorex sorex* *S. fumeus* *S. maritimensis* *S. etruscus*	ANK2	Variants associated to cardiac and neurological disorders (York et al., 2022)

	Sorex *Sorex sorex* *S. maritimensis* *S. etruscus*	ANO10	Variants are known to cause spinocerebellar ataxia, a neurological disorder (Chrysanthou et al., 2022)
	Soricidae *Sorex* *S. maritimensis* *S. etruscus*	MYO9B	Modulates dendritic morphogenesis in the brain (Long et al., 2013)
	S. fumeus *S. etruscus*	RTN4R	Plays an essential role in regulating axonal regeneration and plasticity in the central nervous system (Kimura et al., 2017)
	Soricidae *S. maritimensis*	ACLY	Directs glucose metabolic fluxes to de novo lipogenesis (R. Li et al., 2023)
	Sorex	CH25H	Modulates cholesterol homeostasis (J. Zhao et al., 2020)

CHAPTER 3: Epigenetics and Island-Mainland Divergence in an Insectivorous Small Mammal

Authors: Marie–Laurence Cossette, Donald T. Stewart, Amin Haghani, Joseph A. Zoller, Aaron B. A. Shafer, Steve Horvath

A version of this chapter was published in *Molecular Ecology*

Contributions:

M.-L.C., A.B.A.S., D.T.S. and S.H. conceived the study. M.-L.C., A.B.A.S. and D.T.S. collected the samples. M.-L.C. and D.T.S. performed the molecular laboratory work. M.-L.C. performed the bioinformatic analyses with contribution from A.H. and J.Z. for the epigenetic clock scripts. M.-L.C. and A.B.A.S. wrote the manuscript, and D.T.S., A.H., J.Z. and S.H. reviewed it.

Complete citation:

Cossette, M., Stewart, D. T., Haghani, A., Zoller, J. A., Shafer, A. B. A., & Horvath, S. (2023). Epigenetics and island-mainland divergence in an insectivorous small mammal. Molecular Ecology, 32(1), 152–166.

https://doi.org/10.1111/mec.16735

Abstract

Geographically isolated populations, specifically island-mainland counterparts, tend to exhibit phenotypic variation in many species. The so-called island syndrome occurs when different environmental pressures lead to insular divergence from mainland populations. This phenomenon can be seen in an island population of Nova Scotia masked shrews (*Sorex cinereus*), which have developed a specialized feeding habit and digestive enzyme compared to their mainland counterparts. Epigenetic modifications, such as DNA methylation (DNAm), can impact phenotypes by altering gene expression without changing the DNA sequence. Here, we used a *de novo* shrew genome assembly and a mammalian methylation array profiling 37 thousand conserved CpGs to investigate morphological and DNA methylation patterns between island and mainland populations. Island shrews were morphologically and epigenetically different than their mainland counterparts, exhibiting a smaller body size. A gene ontology enrichment analyses of differentially methylated CpGs implicated developmental and digestive system related pathways. Based on our shrew epigenetic clock, island shrews might also be aging faster than their mainland counterparts. This study provides novel insight on phenotypic and epigenetic divergence in island-mainland mammal populations and suggests an underlying role of methylation in island-mainland divergence.

Keywords

DNA methylation, epigenetic clock, genome assembly, island divergence, island syndrome, masked shrew

Introduction

Insular populations tend to exhibit divergent traits compared to their continental counterparts (Adler & Levins, 1994; Foster, 1964; Goltsman et al., 2005; Meröndun et al., 2019; Novosolov et al., 2013). This phenomenon is known as the island syndrome and refers to ecological and morphological differences between island and mainland populations (Baeckens & Van Damme, 2020). Altered predation, competition, environment, and geographic isolation are thought to lead to phenotypic differences between mainland and insular populations (Baeckens & Van Damme, 2020; Benítez-López et al., 2021; Lomolino et al., 2013).

Relatively little is known about the specific mechanisms causing phenotypic changes in island–mainland scenarios, whether it be genetic drift, natural selection, or plasticity (Baeckens & Van Damme, 2020; Benítez-López et al., 2021). Miller et al. (2021) found heritable variations for deer mice for both mass and length associated with outlier loci between island–mainland populations and suggested that natural selection underlies such phenotypic divergence. Island–mainland populations, however, do not always exhibit clear signals of nuclear genetic divergence with divergent phenotypes (e.g. Meröndun et al., 2019). In some cases, different epigenetic signatures, which include chemical modifications that regulate gene expression (Al Aboud et al., 2021), have been associated with levels of divergence between species and populations exhibiting different phenotypes (e.g. Meröndun et al., 2019; Vernaz et al., 2021) and unlike DNA, epigenetic modifications can respond quickly to environmental pressures even within an individual's lifetime (Anastasiadi et al., 2021). For example, differential epigenetic patterns associated with phenotypic traits have been noted for body size in sheep (*Ovis sp.*) (Cao et al., 2015) and lynx (*Lynx*

canadensis) (Meröndun et al., 2019), for pigmentation levels in rabbits (*Oryctolagus cuniculus*) (Y. Chen et al., 2021) and for breeding success in barn swallows (*Hirundo rustica*) (Saino et al., 2017).

Epigenetic modifications may take a variety of forms, these include DNA methylation (DNAm), histone modifications and non-coding RNA activity, all of which can regulate gene expression without requiring changes in the DNA sequence (Moore et al., 2013). Differential patterns of epigenetic modification have been associated with early stages of speciation (Vernaz, Hudson, et al., 2021; Vernaz, Malinsky, et al., 2021; Vogt, 2017). DNAm involves the transfer of a methyl group to a cytosine base (Moore et al., 2013) and can occur throughout the genome but mostly occurs on CpG dinucleotide sites (Moore et al., 2013). Depending on where in the genome these modifications occur, they can either supress or activate gene expression (Dhar et al., 2021; Wan et al., 2015) and in some cases, methylation of only one CpG can inhibit the binding of transcription factors (Iguchi-Ariga & Schaffner, 1989; Watt & Molloy, 1988).

CpG islands (CGI) are regions of the genome that contain a high frequency of CpGs relative to other regions of the mammalian genome (Illingworth & Bird, 2009). CGIs are also known to often overlap promoter regions (Deaton & Bird, 2011; P. A. Jones, 1999; Saxonov et al., 2006). These sites might be evolutionarily important in mammals since many CGIs are conserved between mice and humans (Illingworth et al., 2010; Moore et al., 2013). It is theorized that over time differential methylation can lead to permanent heritable (cis) nuclear changes, referred to as genetic assimilation (Anastasiadi et al., 2021). This can happen when environmental conditions persist through generations favouring a specific phenotype. Over time, this can result in

selection for a permanent genetic basis (e.g. point mutations) of the plastic epigenetically influenced phenotype (Anastasiadi et al., 2021).

In addition, DNAm levels, particularly hypermethylation, on specific CpG sites changes throughout an individual's lifetime, making it possible to age individuals based on their methylation levels (Horvath, 2013; Lu et al., 2021; Prado et al., 2021; Wilkinson et al., 2021). These epigenetic clocks can predict chronological age in different mammalian species (Jasinska et al., 2021; Lu et al., 2021; Prado et al., 2021; Raj et al., 2021; Schachtschneider et al., 2021), and can be used to validate current aging techniques based on field observations (Robeck et al., 2021). Epigenetic clocks can also be used to predict demographic trends for conservation efforts (Prado et al., 2021) and have been shown to be elevated in inbred individuals, suggesting that inbreeding can lead to aging at a faster rate (Larison et al., 2021a). Likewise, metabolic syndrome in humans (i.e. obesity) also results in increased epigenetic aging (H. S. Lee & Park, 2020; Nannini et al., 2019). It stands to reason that the differential selection pressures underlying island--mainland patterns might also alter the DNAm aging rate.

Shrews of the family Soricidae appeared over 50 million years ago (Yuan et al., 2024). Hundreds of extant shrew species are now found on most continents (Churchfield, 1990; George, 1986) and have evolved unique traits such as echolocation (Chai et al., 2020; Forsman & Malmquist, 1988; Tomasi, 1979), venomous saliva (Kita et al., 2004) and the ability to dive underwater (Catania et al., 2008; Mendes-Soares & Rychlik, 2009). Shrews have the highest mass--specific metabolic rate of all mammals (Ochocińska & Taylor, 2005). They are also one of the smallest and shortest--lived mammals, with many species having a maximum lifespan of ~17 months (Churchfield, 1990), making them important to study from an aging

perspective. Shrews of the genus *Sorex* show differing trends in body size variation across their range; populations of certain species at higher latitudes tend to have smaller body sizes, which is counter to the predictions of Bergmann's rule (Ochocinska & Taylor, 2003; Yom-Tov & Yom-Tov, 2005), whereas some populations of other species are larger at more northerly latitudes (Huggins & Kennedy, 1989). *Sorex* shrews also exhibit morphological plasticity; the Eurasian common shrew (*Sorex araneus*), for example, manifests seasonal and reversible changes in skull morphology, body mass and organ size (Lázaro, Dechmann, et al., 2017; Schaeffer et al., 2020; Taylor et al., 2013). Whether plastic or genetic in nature, these phenotypic variations throughout winter and latitude are generally explained as a response to limited resource availability (Lázaro, Dechmann, et al., 2017; Ochocinska & Taylor, 2003; Schaeffer et al., 2020; Taylor et al., 2013; Yom-Tov & Yom-Tov, 2005).

Masked or cinereus shrews (*Sorex cinereus*) are the most abundant shrew in North America (Nagorsen, 1996). The range of this species spans much of the northern United States and most of Canada (Nagorsen, 1996) with a varied diet consisting primarily of invertebrates (Bellocq et al., 1994; Churchfield, 1990; Whitaker Jr. & French, 1984). One population of island dwelling masked shrews off the southern coast of Nova Scotia became isolated from the mainland during the late Wisconsin glaciation around 20,000 years ago (Roland, 1982; Stewart & Baker, 1992) and has developed a specialized feeding habit in response to their littoral environment (Stewart et al., 1989). Masked shrews on Bon Portage Island (BPI) are the only population of *Sorex* known to mainly feed on sand fleas (*Platorchestia platensis*) (MacPherson & Stewart, 2003; Stewart et al., 1989). *Platorchestia platensis* is a type of amphipod typically not consumed by insectivores, although Churchfield (1990) observed that the

population of the lesser white–toothed shrew (*Crocidura suaveolens*) on St. Martin's, Isles of Scilly, in the United Kingdom, also feed on amphipods. In the case of *S. cinereus*, electrophoresis assays by Stewart and Baker (1992) found that the BPI masked shrew population expressed a unique (slow) allele for Peptidase A and showed evidence of drift. It was later hypothesized that due to the importance of sand fleas on BPI, the shrew's digestive enzyme, Peptidase A, had undergone an adaptive response (MacPherson & Stewart, 2003). Here, using a *de novo* assembled and annotated genome, we built a masked shrew epigenetic clock and quantified differential methylation patterns between populations to examine the association between geographical isolation and epigenetic divergence. We hypothesized that island shrews will exhibit different phenotypes than mainland shrews based on the island syndrome and that divergent methylation patterns could underly these differences.

Materials and Methods

Sampling, sexing, and aging shrews

Masked shrew samples were collected from four locations: Sandy Cove (44°49′N, 66°09′W), North Mountain (45°18′N, 64°46′W), Long Island (44°38′N, 66°23′W) and Bon Portage Island (43°46′N, 65°75′W) in Nova Scotia, Canada (Figure S3.1). All shrews were trapped between mid-July and late August to limit seasonal morphological differences, such as skull and body size (Lázaro, Dechmann, et al., 2017; Lázaro et al., 2019, 2021). Shrews were euthanized using cervical dislocation or thoracic compressions (Animal Care Certificate No. 26234). An Ohaus CS200 scale was used to weight each shrew in grams with an accuracy of one decimal place. A vinyl

Almedic ruler was used to measure full body length with and without tail in millimetres (Figure S3.2). Body length values without tail were used for the analyses. A Preciva electronic digital caliper was used to measure the full skull length from the most protruding anterior to the most protruding backward point of the skull (Figure S3.2). DNA was extracted from liver as this is where digestive enzymes are found; tail and fetal tissue were selected to help build the epigenetic clock (see Lu et al., 2021). Liver and tail samples produce high DNA yields and are easy to locate in shrews. We used a DNeasy Blood & Tissue Kit from QIAGEN following the *Purification of Total DNA from Animal Tissues (Spin-Column Protocol)* to extract DNA from whole fetuses, ~ 20-30 mm tail snips and ~ 20 mg of liver samples from each individual. Shrews were left in dermestid beetle tanks for a week to remove all remaining tissue from the skulls. A Leica EZ4 microscope was used to assess age class for each shrew based on teeth wear according to the method of Pruitt (Pruitt, 1954; Rudd, 1955). Age in months was estimated based on the trapping date, the age class, and the known reproductive season of masked shrews assuming a maximum lifespan of ~17 months (Churchfield, 1990). Shrews were sexed by gel electrophoresis of PCR amplicons of the Y chromosome-linked SRY gene (Cervantes et al., 2013; Matsubara et al., 2001) using custom primers (Table S3.1); sex assignments were subsequently validated with the methylation assay.

Genome assembly and annotation

A single smoky shrew (*Sorex fumeus*) from Peterborough County (44°53′N, 78°05′W), Ontario, Canada had DNA and RNA extracted from the heart and liver immediately after euthanasia. We used the MagAttract HMW DNA Kit from QIAGEN

following the *Disruption/Lysis of Tissue* and *Manual Purification of High-Molecular-Weight Genomic DNA from Fresh or Frozen Tissue* protocols to extract DNA from ~ 20 mg of heart and liver samples. Both extracts were then pooled together. We used the Monarch Total RNA Miniprep Kit from New England BioLabs to extract RNA from ~ 20 mg of heart and liver following the kit's protocol for tissue samples. Extracts were sent to The Centre for Applied Genomics, Toronto, Ontario, Canada, for sequencing. Short-read (SR) libraries were generated using Truseq PCR-free preparation and sequenced on two lanes on an Illumina HiSeq X platform with 150 bp paired-end reads. Linked reads (LR) were prepared using the 10X genome library after selecting fragments >20Kb using BluePippin. The linked reads were sequenced on one Illumina HiSeq X Lane. Twenty-five million RNA reads were sequenced on a HiSeq2500 with 126 bp paired-end reads.

We used BBMap v.35.8 (Bushnell, 2022) to remove adapters and trim low-quality data from the fastq files. Contaminant screening was performed using Kraken2 v.2.0.9-beta (Wood & Salzberg, 2014) and all unclassified sequences were kept. Kmergenie v.1.7048 (Chikhi & Medvedev, 2014) was used to find the kmer size for downstream analysis. The genome was assembled using a tiered approach: 1) using only the SR data with Meraculous v.2.2.4 (Chapman et al., 2011) and 2) with only the LR data using Supernova v.2.1.1 (Weisenfeld et al., 2018) using the --maxreads filter set to "all". The pseudohap2 style was selected for the assembly output. Following these assemblies, both SR and LR assembly versions were merged via quickmerge v.0.3 (Chakraborty et al., 2016) with the LR as the backbone. Scaff10x v.5.0 (Ning et al., n.d.) was used to further polish the *S. fumeus* genome. Genome completeness was assessed using Benchmarking Universal Single-copy Orthologs (BUSCO) v.3.0.2 (Simão

et al., 2015) by comparing the hybrid genome to highly conserved genes in mammals based on the *mammalia_odb9* dataset. The mitochondrial genome was assembled using MitoZ v.2.4-alpha (Meng et al., 2019b). Genome annotation was conducted via GenSAS v.6.0 (Humann et al., 2019) using the assembled genome and RNAseq data. Here, repeat regions of the genome were identified using RepeatModeler v.1.0.1.1 (Smit & Hubley, 2023b) and RepeatMasker v.4.0.7 (Smit & Hubley, 2023a). Smoky shrew liver and heart RNA reads were mapped to the genome using HISAT2 v.2.1.0 (Kim et al., 2019). The resulting BAM file was used by BRAKER v.2.1.1 (Hoff et al., 2019) and EVidenceModeler v1.1.1 (Haas et al., 2008) to predict features in the genome. Functional annotation to find common protein sequences between the smoky shrew and the common shrew, *Sorex araneus* (GenBank assembly accession: GCA_000181275.2), and other vertebrate mammals was done using BLAST v.2.11.0 (Altschul et al., 1990), SwissProt v.2.11.0 (Bairoch & Apweiler, 1996) and InterProScan v.5.44-79.0 (Jones et al., 2014) databases. The assembled genome and shared annotations were used only to identify *Sorex* specific probes.

Epigenetic clocks

DNA methylation profiles for each sample (n = 39, Table S3.2) were assayed on the HorvathMammalMethylChip40, a novel Infinium array that profiles up to 37,000 highly conserved CpGs across the genome (Arneson et al., 2022). The SeSaMe normalization method was used to define beta values for each probe (Zhou et al., 2018). Unsupervised hierarchical clustering based on inter-array correlation coefficients revealed that samples cluster by tissue type. The methylation array measurements can assess the sex of each sample based on the X chromosomal CpGs.

One fetal sample was removed due to inaccurate sexing. Only probes that aligned to the smoky shrew genome using QuasR v.1.12.0 (Gaidatzis et al., 2015) were retained for downstream analysis (n = 29,609). Elastic net regression models (alpha = 0.5, 10-fold internal cross-validation) were applied to the DNAm data (covariates) and the chronological age estimate (dependent variable). The resulting multivariate regression model will be referred to informally as the masked shrew epigenetic clock. The alpha value of the glmnet R software package v.4.1-3 (J. H. Friedman et al., 2010) was set to 0.5 to be midpoint between a ridge and lasso type regression. Elastic net regressions help prevent overfitting the model to the data by selecting variables (CpGs) and shrinking them simultaneously (Zou & Hastie, 2005). Three different models were used as per Lu et al. (2021); 1) no transformation to the chronological age, 2) log-transformed chronological age (log(x+1)), and 3) square root transformed chronological age (sqrt(x+1)) with offsets of 1 year to avoid negative values for our fetal samples. To assess the accuracy of our epigenetic clock models we used a leave-one-out (LOO) cross-validation method. This method consists of training the epigenetic clock models with a subset of the data and testing how well it can predict other observed data. We report the LOO estimates of the Pearson correlation between age and its DNAm based estimates as well as the median absolute error (MAE) between DNAm age and observed age (estimated chronological age in years).

Statistical analyses and functional gene enrichment

We ran a series of models in R v.1.3.959 (R Core Team, 2021). Using linear models, we separately regressed body size, weight, and skull length against location, age, and sex (i.e., phenotype ~ location + age + sex) of 27 masked shrews (Table S3.2).

Here, location was a categorical variable of island or mainland (see Table 3.1). Additionally, we extracted the residuals from the epigenetic clock for each sample and ran a Welch Two Sample T-test using location, sex and age group to test if different groupings epigenetically aged at comparable rates. Positive residuals were indicative of an epigenetically older individual than the linear model would predict.

Epigenome-wide association studies (EWAS) were conducted using the limma package v.3.44 (Ritchie et al., 2015) with 36 methylomes (Table S3.2); note we removed the fetal samples from this analysis to avoid biased results associated with developmental genes and pathways. Using the β values of aligned probes, we ran the following model: DNAm ~ location + sex + age + tissue including animal ID as a random effect. This model was run two times; once with BPI and Long Island grouped together as 'Island' in the model with all other locations assigned as 'Mainland' and once with BPI versus all other populations. The latter grouping was meant to isolate possible unique dietary adaptations of BPI shrews. The island versus mainland model was also run with only Long Island samples to test for a potential bias of having more BPI samples in the island grouping.

A *p* value < 2e-6 was set according to the Bonferroni-corrected threshold as the cutoff for significance to be conservative (Johnson et al., 2010). The location of probes in relation to the gene annotations were identified using the ChIPseeker v.1.24.0 (Yu et al., 2015) and GenomicFeatures v.1.40.1 (Lawrence et al., 2013) packages. Probe locations were identified as intron, exon, downstream up to 300 bp from the transcriptional start site (TSS), promoter up to 10 kb upstream and 1 kb downstream from the TSS, and distal intergenic for probes that were not part of any other category. Enriched pathways associated with the significant CpGs for the EWAS models were

identified with rGREAT v.1.20 (McLean et al., 2010) using the human Hg19 genome annotation to allow for comparisons (Jasinska et al., 2021; Prado et al., 2021; Schachtschneider et al., 2021). We accounted for bias resulting from the design of the mammalian array by using the 14,290 CpGs that mapped to our shrew gene annotations as background. Following Schachtschneider et al. (2021), from these probes we filtered and kept the top 500 hypermethylated and top 500 hypomethylated sites based on *p* values to use for the enrichment analysis.

Results

Genome assembly and annotations

We assessed the quality and completeness of our genome assembly to ensure confidence in our downstream analyses. The SR and LR assembly statistics are provided in Table S3.3; the final hybrid genome was 2.66 Gb with an N50 of 1.99 Mb and L50 of 291 scaffolds. The GC content was 42.88% with over 86% of the genome being in scaffolds over 50 Kb and the largest scaffold being around 22 million base pairs. Close to 95% of BUSCOs were identified in this assembly (Table S3.3): 3504 of 4101 were complete single-copies, 74 were complete duplicates and 301 were fragmented. Repeats represented 38.60% of the genome (Table S3.4) with long interspersed nuclear elements (LINEs) and unclassified repeats representing 20.17% and 15.39% of the genome respectively. The functional annotation pipeline identified 20,067 protein coding genes (Table S3.3). The metrics assessed were comparable to other genome assemblies using similar sequencing approaches (e.g., Etherington et al., 2020; Wolf et al., 2022). The raw sequence data have been deposited in the Short

Read Archive (SRA) under accession number PRJNA826195. The genome assembly has been deposited in GenBank/NCBI under accession number JAOANZ000000000.

Epigenetic clocks

We built a shrew methylation clock to assess epigenetic aging rates between different shrew populations (n = 24 from islands versus n = 6 from mainland). The methylation data consisted of 27 tail, 9 liver, and 3 fetal tissue DNA samples selected from 30 different individuals (Table S3.2). One fetal sample was dropped due to quality control metrics (wrong sex). The average age of our samples was seven and a half months; the youngest shrews being fetuses and the oldest one being an estimated 16 months (Table 3.1). Out of the 37,000 probes, 29,609 were aligned to our shrew genome. We constructed and cross-validated epigenetic clocks for different masked shrew tissue types, which included liver, tail, fetus, and multi-tissue clock. The final clocks were based on the methylation profiles of 21 CpGs, similar to clocks from other species such as elephants or cats (e.g., Prado et al., 2021; Raj et al., 2021), our clocks were highly accurate with a LOO cross validation estimate of the Pearson correlation between age and its predicted value r >= 0.94 and a median absolute error of MAE = 0.12 years for the multi-tissue clock, representing approximately a 6-week difference between the estimated DNAm age and the teeth-based chronological age. The liver DNAm clock had the lowest MAE of 0.07 years (~3.5 weeks), followed by the tail clock with MAE around 0.14 (~7 weeks) (Figure 3.1).

Island-mainland differences

Morphological and epigenetic age differences between 21 island and 6 mainland masked shrew populations (Table 3.2) were examined using linear regressions. Mass, body length and skull length differed between the mainland and island samples (Table 3.2). Our models including age, sex, and island-mainland covariates explained a high percentage of variance: skull length (Adj. R^2 = 0.70), body length (Adj. R^2 = 0.33) and weight (Adj. R^2 = 0.27; Table 3.2). In all models, location was the variable that correlated most to the change in body size and weight (*p* value < 0.01; Table 3.2). All size metrics were negatively correlated to being on an island (Table 3.2, Figure 3.2). Age was correlated to body length whereas sex did not have a significant effect in all models (Table 3.2).

Epigenetic clock residuals were slightly elevated on islands indicative of individuals epigenetically older than predicted (t = - 1.77, df = 18.83, *p* value = 0.09); residual variation was higher on the island (mean of 0.01) compared to the mainland (mean of -0.02). There was also a difference in residuals among the sexes (t = -1.78, df = 33.93, *p* value = 0.08) but not for age groups (yearlings versus older) (t = -0.04, df = 26.59, *p* value = 0.97).

EWAS Island effects

EWAS were conducted using the methylation data consisting of the 27 tail, 9 liver samples (Table S3.2) to identify differentially methylated genes and pathways between island and mainland masked shrews. A total of 469 differentially methylated CpG sites (218 hypermethylated and 251 hypomethylated relative to mainland) were identified (*p* value < 2e-6) between our island and mainland populations (Figure 3.3a).

The most divergent CpG was on the *EWS* exon (*p* value = 5.76e-18) a gene that encodes a multifunctional protein involved in various cellular processes such as gene expression, cell signaling, and RNA processing and transport. Top differentially methylated CpGs were located on or near genes associated to developmental and adipose-related processes: *MEST* exon (*p* value = 1.01e-09), *PHIP* intergene (*p* value = 3.72e-08) *SIM1* intergene (*p* value = 1.28e-08) and metabolism: *ATPA* exon (*p* value = 8.74e-11) and *ZBTB7C* promotor (*p* value = 3.53e-09) (Table 3.3, Figure 3.3a, b). The top hypermethylated CpGs were associated to three different fibroblast growth factor receptors pathways (*p* value < 5.17e-10) (Figure 3.3c). Other significantly enriched pathways included synaptic and adipose-related processes (Figure 3.3c).

Isolating BPI identified 112 significant CpGs (*p* value < 2e-6), 59 were hypermethylated and 53 were hypomethylated compared to other populations (Figure 3.4a). The top significant hypomethylated CpG in BPI samples was on the *SMO* exon (*p* value = 1.06e-24; Table 3.3, Figure 3.4a, b). Top CpGs were most often hypomethylated in BPI island samples (Table 3.3, Figure 3.4a, b). Pathway analysis highlighted that differentially methylated sites in BPI shrews compared to other populations seem to affect digestive and intestinal phenotypes and functions (Figure 3.4c). Additionally, other phenotypes related to morphology, more specifically the skeleton, tail, and cartilage showed pathway enrichment in BPI (Figure 3.4c). When BPI was removed from the island-mainland model (i.e. only Long Island), digestive enzymes were not identified as being a significantly enriched pathway (Figure S3.3).

Discussion

The island syndrome describes the distinctive morphological differences between island-mainland populations due to different ecological and environmental pressures (Baeckens & Van Damme, 2020). Here, we examined phenotypic divergence and epigenetic modifications, specifically DNAm, which has the potential to increase phenotypic plasticity within an individual (Anastasiadi et al., 2021). Our analysis showed differential methylation patterns and distinct phenotypes between island and mainland masked shrew populations. Over 29,000 CpG methylation probes aligned to our genome, consistent with other mammalian studies (Larison et al., 2021b; Pinho et al., 2022; Sugrue et al., 2021), and a robust epigenetic clock was generated (Figure 3.1). The differential methylation between island-mainland populations showed a remarkable overlap between enriched pathways and measured phenotypic differences. While some association assessments suffer from false positives and "story-telling" (Pavlidis et al., 2012), there was an unambiguous connection to the measured body phenotypes (Figure 3.2), dietary differences (MacPherson & Stewart, 2003; McAlpine, 2009; Stewart et al., 1989), and the observed enriched pathways (Figure 3.3c, Figure 3.4c). Thus, the clear morphological and dietary island-mainland differences appear to be, in part, under epigenetic influence.

Highly diversified DNAm methylation among populations can be a strategy to take advantage of unpredictable environments (Anastasiadi et al., 2021); moreover, epigenetic modifications have been shown to allow isogenic lines to adjust their phenotype to maximize fitness (Van Egeren et al., 2018). Bon Portage Island masked shrew populations have lower genetic diversity than their mainland counterparts (Stewart & Baker, 1992), but likely have a broader ecological niche due to reduced

interspecific competition; masked shrews on mainland Nova Scotia are broadly sympatric with six other species of shrews, and diet studies demonstrate that these soricids differentially exploit available invertebrate food resources (Whitaker & French, 1984). Without standing genetic variation, differential methylation among offspring could be a bet-hedging strategy to allow for exploration and exploitation of a range of food resources on the island (Anastasiadi et al., 2021). Further, DNAm can theoretically lead to genetic modifications through the process of genetic assimilation where methylated cytosines have a higher frequency of mutagenesis resulting in novel alleles (Anastasiadi et al., 2021). Differential methylation is an effective strategy that individuals in genetically impoverished populations may use to generate phenotypic diversity. Thus, we hypothesize that BPI shrews are in the process of methylation-induced genetic assimilation, with the high frequency of the unique peptidase A allele in this population being one outcome of this process.

Phenotypic and epigenetic divergence

Here, insular masked shrews exhibited different phenotypes than their mainland counterparts; morphologically, this presented as decreased body length, mass, and skull length on islands. Body size appears to be a heritable polygenic trait in mammals (Bonnet et al., 2017; Posbergh & Huson, 2021; Postma, 2014) with many underlying mechanisms, often making it difficult to clearly understand the role of each gene or pathway involved (e.g., Anderson et al., 2022). Cytosine methylation levels measured in highly conserved stretches of DNA have been linked to traits such as maximum lifespan in mammals (Haghani et al., 2021) and has been suggested as a universal mechanism to regulates body mass and morphological development in

animals in other association studies (Cao et al., 2015; Haghani et al., 2021). Genetic variation in quantitative trait loci (QTL) almost certainly play a role in phenotypic variations between insular and mainland populations – and might also influence methylation (J. T. Bell et al., 2011; A. K. Smith et al., 2014). We suggest that pooled genome sequencing approaches (e.g., island versus mainland samples) are a feasible option to test for nuclear genomic associations of this nature.

Differential methylated pathways between island and mainland shrew populations for all tissue types included fibroblast growth factors (FGF), which are crucial for animal body development (Oulion et al., 2012), and adipose-related genes and processes. Haghani *et al.* (2021) also found that across mammals, species weight was correlated to development and adipose-related pathways, possibly due to DNAm affecting the expression of adipose tissue and regulation of lipid storage (Ma & Kang, 2019). In masked shrews, CpGs on the homolog to mesoderm-specific transcript protein (*MEST*) exon and single-minded homolog 1 (*SIM1*) intergenic region were hypo and hypermethylated, respectively, in island shrews compared to mainland shrews. *MEST* expression is known to be correlated to adipocytes in humans (Karbiener et al., 2015) and mice (Takahashi et al., 2005), while *SIM1* expression is associated with food intake regulation and obesity (Holder et al., 2004; Yang et al., 2006), both possibly affecting body mass in shrews. Differential methylation of genes such as *SIM1* might affect shrews' metabolism and ultimately body size by regulating the expression of hormones involved in signaling for food intake. Further, island shrews had a hypomethylated CpG in the pleckstrin homology domain-interacting protein (*PHIP*) intergenic region, a probable regulator of insulin-like growth factor signaling pathways (Farhang-Fallah et al., 2002), which reinforces the idea that insulin growth factors are

involved in mammalian life-history variation (Swanson & Dantzer, 2014). The overall associations between DNAm and genes related to body composition possibly indicates that methylation inhibits or activates various genes during developmental stages in shrews, but future studies are needed to better understand such mechanisms.

Masked shrew insular populations appear to have non-significantly decreased body mass compared to their mainland counterparts; BPI masked shrews occur in high densities (Downie, 1986; Stewart et al., 1989; Telfer, 1984), which should lead to small animals having larger body size to better compete for resources (Juette et al., 2020). However, the discrepancy with this expected phenotype for island shrews might be due to their diet. Previous studies have shown that food scarcity contributes to the selection of smaller body size in Palearctic shrews, which might also be the case for Nearctic populations (Ochocinska & Taylor, 2003; Yom-Tov & Yom-Tov, 2005). Churchfield (2002) found that in locations where higher biomass resources are scarce, shrews tend be smaller in size and survive by consuming less nutritional but more abundant sources of food. This might be the case for BPI shrews feeding on sand fleas. Supporting this idea, top EWAS genes involved in energy homeostasis and metabolism were differentially methylated between mainland-island populations. For example, ATP Synthase F1 Subunit Alpha (*ATPA*) catalyzes the synthesis of ATP, the main molecule used to store and transfer energy in cells (Erecińska & Silver, 1989). Similarly, the zinc finger and BTB domain–containing 7c (*ZBTB7C)* gene promotor was also differentially methylated in island shrews. *ZBTB7C* is an important metabolic regulating gene for blood-glucose homeostasis during fasting (Choi et al., 2019).

BPI shrews also showed differentially methylated sites associated with various processes related to the digestive system, including peptidase activity. This is of

significance as previous studies noted that masked shrews on BPI have developed a specialized littoral feeding habit (MacPherson & Stewart, 2003; Stewart et al., 1989). Previous work showed that BPI shrews possess a unique polymorphism for the digestive enzyme Peptidase A (Stewart & Baker, 1992), which was hypothesized to potentially facilitate digestion of sand flea exoskeletons by preferentially cleaving proteins unique to amphipods (McAlpine, 2009). Digestive system related pathways being in the top enrichment category indicate that these dietary habits seem to be one of the most significant molecular differences between this insular population and others and support earlier allozyme work (MacPherson & Stewart, 2003; McAlpine, 2009; Stewart & Baker, 1992). Our work remains correlative, but future genomic assays can be used to test our genetic assimilation hypothesis, while mouse models or common-garden experiments could be used to assess the localization of these CpGs and phenotypic associations to provide information on possible mechanism of enriched regions on masked shrew methylomes.

Epigenetic age

DNA methylation and chronological age can be used to build epigenetic clocks to identify genes associated with aging and predict how methylation patterns change throughout an individual's lifetime (Larison et al., 2021b; Lu et al., 2021; Prado et al., 2021; Wilkinson et al., 2021). Epigenetic clocks have been constructed for a wide range of species, showing consistent trends with CpGs associated to developmental genes and pathways (Lu et al., 2021; Pinho et al., 2022; Raj et al., 2021; Schachtschneider et al., 2021). Second generation clocks, such as the DNAm GrimAge, account for variables such as chronological age, health factors and mortality, to predict

lifespan and healthspan (C. G. Bell et al., 2019; Lu et al., 2019). In humans, DNAm GrimAge has been studied in relation to various comorbidities and social factors (Lu et al., 2019; McCrory et al., 2021). Inbreeding has been associated with decreasing longevity and an increase in health concerns in various species (Keller & Waller, 2002; Skotarczak et al., 2020; Yordy et al., 2020) and recently was linked to accelerated DNAm aging in zebras (*Equus quagga*) (Larison et al., 2021b). We used these patterns to explore accelerated aging as a component of island-mainland divergence.

Masked shrews tend to live on average 14 months; accurately aging small mammals such as shrews can be difficult since they are such small creatures and current methods rely on somewhat subjective classification based on teeth wear, body length measurements and weight (Pruitt, 1954; Rudd, 1955). Our epigenetic clock validated the current aging techniques used in the field, and overall, the clock was highly predictive (R = 0.95; MAE = 0.12) (Figure 3.1). Further, when we examined the clock model residuals, island shrews appeared to age slightly faster than mainland shrews. Parasites can affect hosts' life-history traits through the energetic investment of maintaining an immune system response (Cooper et al., 2012; Cowan et al., 2009; Morand & Harvey, 2000), possibly resulting in faster biological aging. Group size has been shown to be correlated with prevalence and intensity of parasitism (Cote & Poulinb, 1995). The population density of shrews on BPI is extremely high, at least in some years (Downie, 1986; Stewart et al., 1989; Telfer, 1984) possibly resulting in an increased incidence of various parasites and pathogens. Overall, this study reports divergent phenotypic trends between mainland-island masked shrew populations, and more broadly support the possible role of epigenetics in shaping phenotypic divergence. Future research should investigate epigenetic aging in insular populations,

especially when inbreeding (Larison et al., 2021) or phenotypic divergence (i.e., this study) is present.

Acknowledgments

We want to thank the reviewers for the valuable insight and comments provided. This work was supported by CanSeq150 (CGEn) (ABAS, DTS and MLC); Ontario Graduate Scholarship (MLC), Natural Sciences and Engineering Research Council of Canada (NSERC) Alexander Graham Bell Canada Graduate Scholarship-Master's (MLC) and NSERC Discovery grants (ABAS and DTS); Compute Canada Resources for Research Groups (ABAS); Canadian Foundation for Innovation: John R. Evans Leaders Fund (ABAS); Ontario Early Researcher Awards (ABAS). SH was supported by the Paul G. Allen Frontiers Group. We wish to acknowledge that the work for this study was carried out on the traditional territory of the Mississauga (Michi Saagiig) Anishinaabeg and the Mi'kmaq People. We are grateful to have had the opportunity to work on this land and thank the First Peoples for their care, stewardship, and teaching.

Figures and Tables

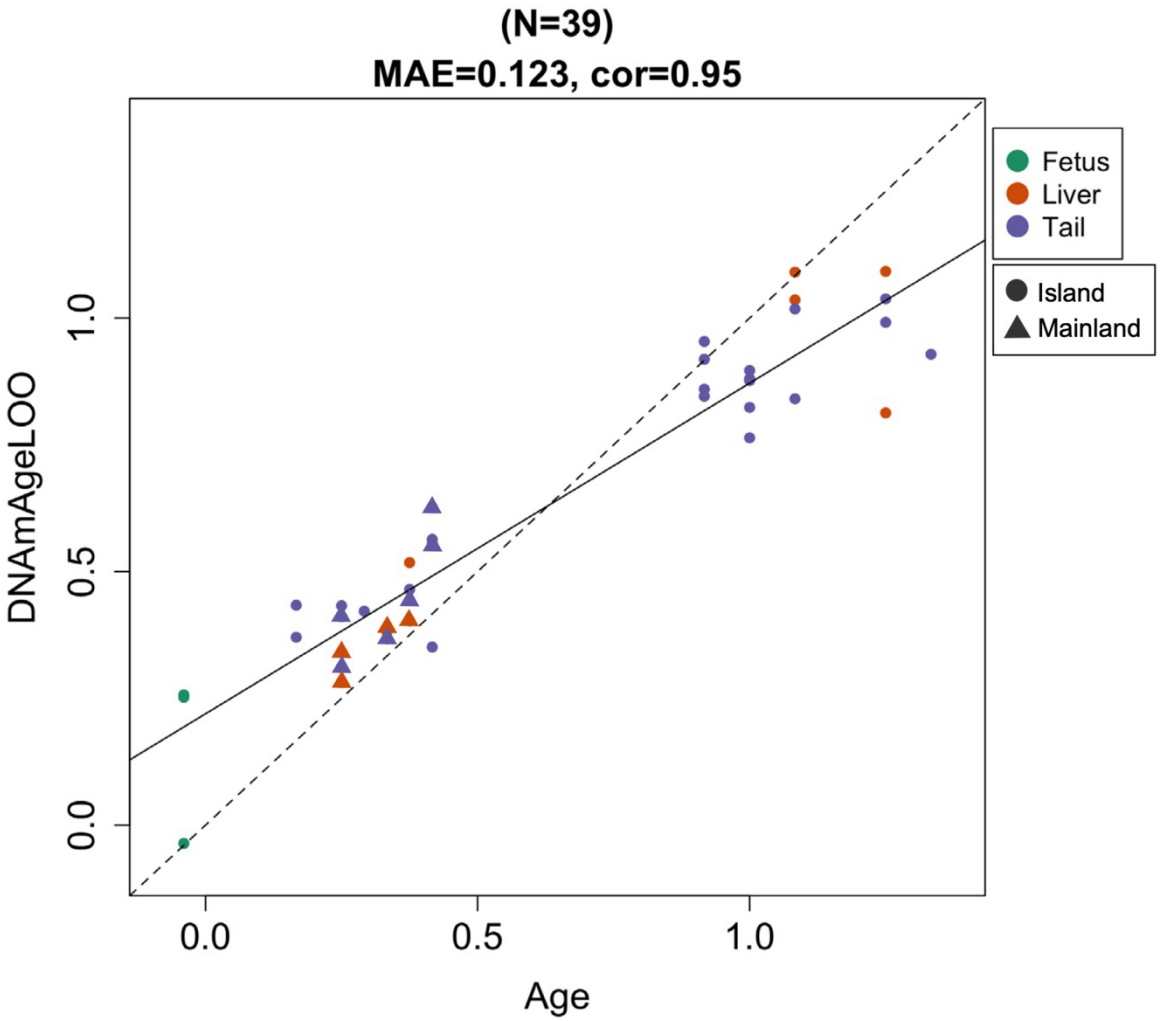

Figure 3.1. Cross-validation study of multi-tissue epigenetic clock for masked shrews. Leave-one-sample-out (LOO) estimate (y-axis, in units of years) versus chronological age (x-axis, in unit of years). The linear regression of epigenetic age is indicated by a solid line while the diagonal line (y = x) is depicted by a dashed line. "MAE" represents the mean absolute error and "cor" represents the correlation coefficient.

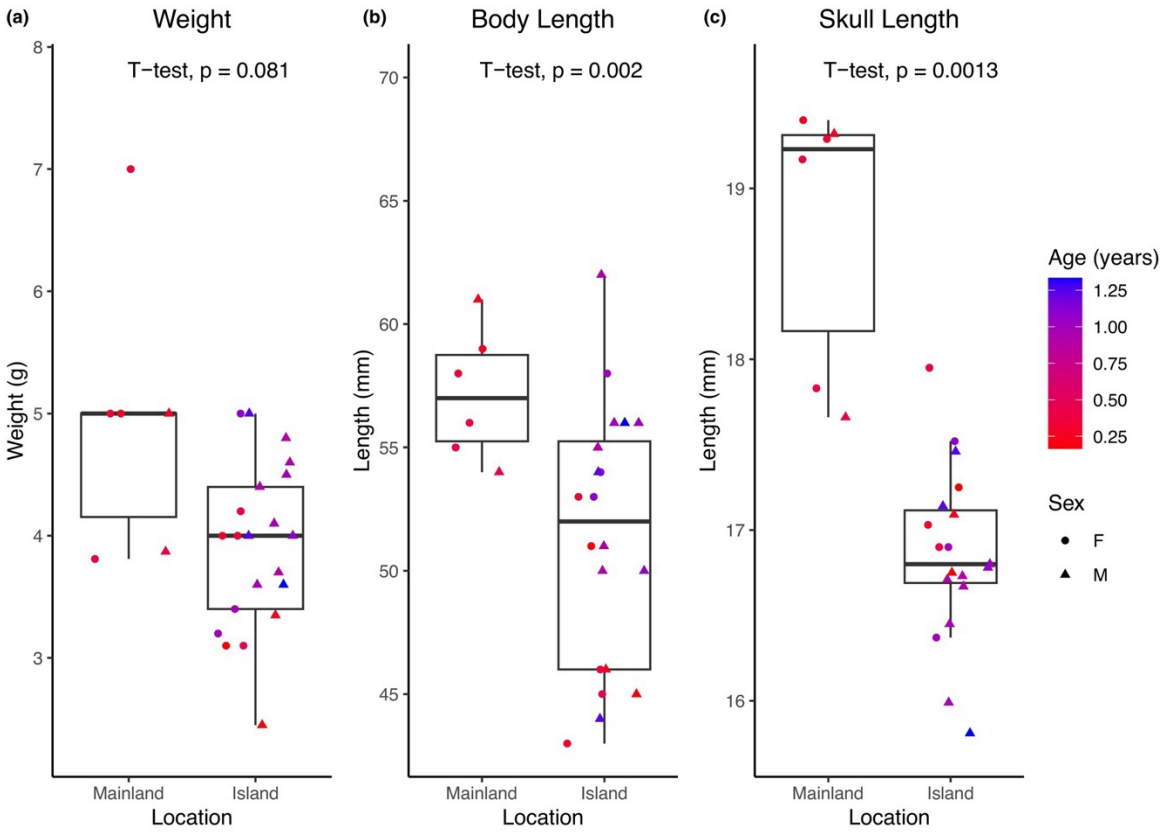

Figure 3.2. Boxplot of (a) weight (g), (b) body length (mm) and (c) skull length (mm) between mainland and island masked shrew populations.

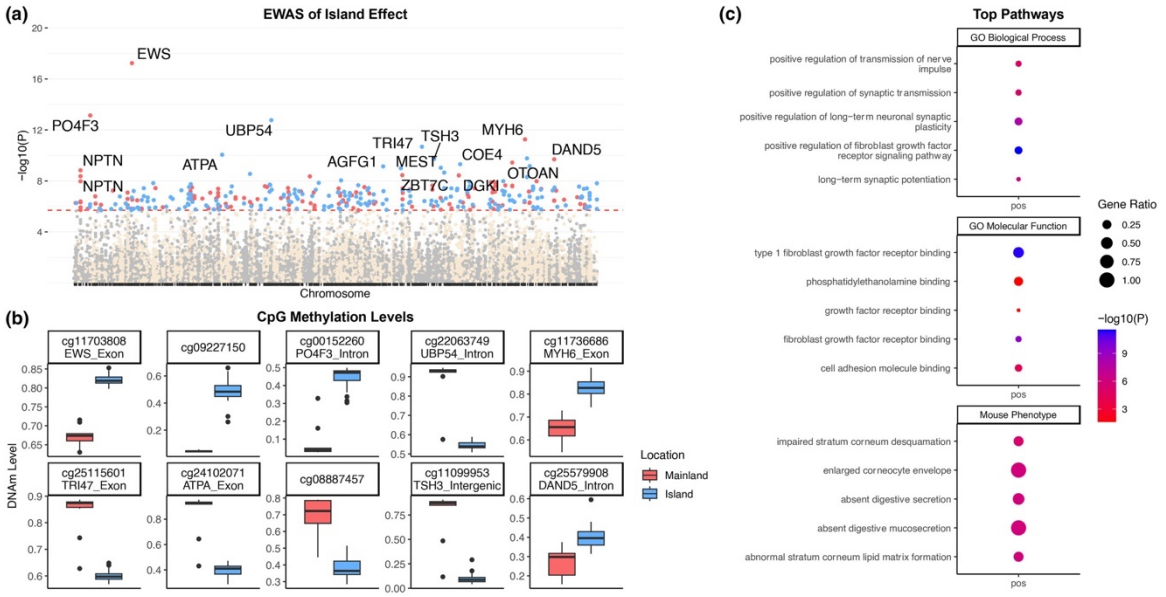

Figure 3.3. Epigenome-wide association (EWAS) of the multivariate regression model between island and mainland masked shrew populations (location, age, sex and tissue). (a) Manhattan plots of the EWAS for location (island versus mainland). The coordinates are estimated based on the alignment of 29,609 Mammalian array probes to our smoky shrew genome assembly. The direction of associations with < 2e-6 (red dotted line) is highlighted by red (hypermethylated) and blue (hypomethylated) colors relative to mainland methylation levels. Top 15 CpGs are indicated by their neighboring genes. (b) DNAm levels of island samples versus mainland for the top 10 significant CpGs (p value) associated to location. (c) Enrichment analysis of the top CpGs with positive (hypermethylated) and negative (hypomethylated) correlations to island populations. The gene-level enrichment analysis was carried out using the GREAT software. Background probes were limited to 14,290 probes that had shrew gene annotations. The top ontologies with most significance pathways for island samples were selected based on Bonferroni corrected p values.

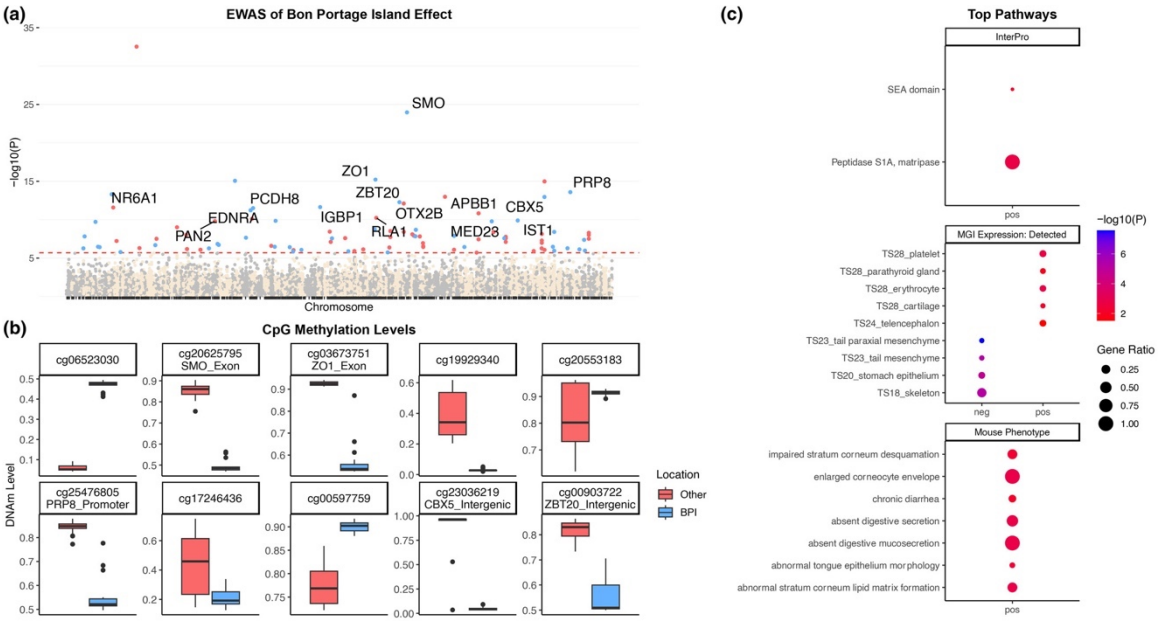

Figure 3.4. Epigenome-wide association (EWAS) of the multivariate regression model between Bon Portage Island and other masked shrew populations (location, age, sex and tissue). (a) Manhattan plots of the EWAS for location (Bon Portage Island (BPI) versus others). The coordinates are estimated based on the alignment of 29,609 Mammalian array probes to our smoky shrew genome assembly. The direction of associations with < 2e-6 (red dotted line) is highlighted by red (hypermethylated) and blue (hypomethylated) colors relative to non-BPI methylation levels. Top 15 CpGs are indicated by their neighboring genes. (b) DNAm levels of BPI samples versus others for the top 10 significant CpGs (p value) associated to location. (c) Enrichment analysis of the top CpGs with positive (hypermethylated) and negative (hypomethylated) correlations to BPI. The gene-level enrichment analysis was carried out using the GREAT software. Background probes were limited to 14,290 probes that had shrew gene annotations. The top 3 ontologies with most significance pathways for BPI were selected based on Bonferroni corrected p values.

Table 3.1. Trapping location of samples, N = total number of samples, number of females, minimum and maximum age. Age is represented as the estimated age in months divided by 12. Note, age of 0.00 represents fetus.

Location	Island	N	No. of females	Age (min/max)
Bon Portage Island, NS	Yes	16	6	0.167/1.33
Long Island, NS	Yes	9	4	0.00/1.25
Sandy Cove, NS	No	4	3	0.250/0.375
North Mountain, NS	No	2	1	0.417/0.417

Table 3.2. Summary statistics for the linear regression models for weight, body length and skull length of mainland versus island masked shrews. Sex values are relative to females and location relative to mainland populations. The β-coefficient (estimate), 95% confidence interval (CI), p value and adjusted R2 are provided.

	Estimate	**95% CI**	***P* value**	**Adj. R^2**
Model 1: Weight ~ Age + Sex + Location				
Age	0.96	−0.03 to 1.95	0.056	0.27
Sex	−0.21	−0.89 to 0.47	0.535	
Location	−1.42	−2.26 to −0.57	**0.002**	
Model 2: Body length ~ Age + Sex + Location				
Age	6.95	1.27 to 12.63	**0.019**	0.33
Sex	−0.30	−4.22 to 3.63	0.877	
Location	−8.80	−13.64 to −3.96	**0.001**	
Model 3: Skull length ~ Age + Sex + Location				
Age	−0.38	−1.11 to 0.36	0.298	0.70
Sex	−0.35	−0.86 to 0.16	0.170	
Location	−1.64	−2.27 to −1.01	**<0.001**	

Table 3.3. Top 5 genes near outlier CpGs for island versus mainland EWAS and BPI versus all other populations. Functions of each gene are abbreviated from Uniprot.

	Gene name	CpG	Direction	Location (p)	Putative function
Island-mainland					
	EWS exon	cg11703808	pos	5.76e-18	Transcriptional repressor and might play a role in the tumorigenic process.
	PO4F3 intron	cg00152260	pos	7.32e-14	Transcriptional activator. Involved in the auditory system development.
	UBP54 intron	cg22063749	neg	1.71e-13	Has no peptidase activity.
	MYH6 exon	cg11736686	pos	5.48e-12	Actin-based motor molecules with ATPase activity.
	TRI47 exon	cg25115601	neg	2.07e-11	Mediates the ubiquitination and proteasomal degradation of CYLD.
BPI					
	SMO exon	cg20625795	neg	1.06e-24	Associates with the patched protein (PTCH) to transduce the hedgehog's proteins signal.
	ZO1 exon	cg03673751	neg	6.04e-16	Related to movement of substances through the paracellular space. Plays a role in the regulation of cell migration.
	PRP8 promoter	cg25476805	neg	2.57e-14	Plays role in pre-mRNA splicing.
	CBX5 intergenic	cg23036219	neg	1.10e-13	Associated with epigenetic repression. Involved in the formation of functional kinetochore.
	ZBT20 intergenic	cg00903722	neg	5.24e-13	Transcription factor that may be involved in hematopoiesis, oncogenesis, and immune responses. Plays a role in postnatal myogenesis.

CHAPTER 4: Genomic and Transcriptomic Signatures of Post-Glacial Divergence in an Insular Shrew Population

Authors: Marie–Laurence Cossette, Donald T. Stewart, Aaron B. A. Shafer

This chapter is in preparation for submission in *Molecular Biology and Evolution*.

Contributions:

M.-L.C., A.B.A.S., and D.T.S. conceived the study and collected the samples. M.-L.C. and D.T.S. performed the molecular laboratory work. M.-L.C. performed the bioinformatic analyses. M.-L.C. wrote the manuscript, and A.B.A.S and D.T.S. reviewed it.

Abstract

Geographic isolation and local selection drive population divergence, leaving distinct genomic signatures. In North America, repeated glacial cycles fragmented habitats, restricted gene flow, and reshaped migration routes, influencing the evolutionary trajectories of many species and populations. One such case is the masked shrew (*Sorex cinereus*) population on Bon Portage Island (BPI), off the coast of Nova Scotia (Canada), which exhibits notable differences in morphology, behavior, and diet compared to their mainland conspecifics. To investigate the evolutionary and adaptive mechanisms underlying this divergence, we generated a high-quality *de novo* genome assembly and annotation and analyzed whole genome resequencing and transcriptome data from four distinct masked shrew populations. Demographic modeling suggests that the BPI population became isolated from the mainland approximately 16-17 kya, coinciding with the post-glacial sea-level rise following the Last Glacial Maximum (LGM), leading to a decline in effective population size. Analyses of genomic islands of differentiation revealed evidence of divergent selection in the BPI population, particularly in genes involved in fatty acid metabolism. Transcriptomic analysis also identified differentially expressed genes (DEGs) also associated with fatty acid metabolism and metabolic regulation. However, overlap between genomic islands of differentiation and DEGs was limited to one region, suggesting a complex architecture and pathways underlying adaptive divergence. Together, our findings highlight the role of historical isolation and divergent selection shaping the genomic and transcriptomic landscape of an insular masked shrew population.

Keywords

genome scans, differential gene expression, shrews, demographic history, island population, fatty acid metabolism, divergent selection, genomic islands of differentiation

Introduction

Divergence among populations is often driven by historical isolation and shaped by local adaptation, leaving signatures across the genome (Brown et al., 2016; Byer et al., 2021). The repeated glaciation events of the Quaternary Period played a major role in shaping the demographic and adaptive history of North American species (Hewitt, 2000; Shafer et al., 2010). As glaciers advanced and retreated, they repeatedly fragmented habitats, established refugia, and intermittently exposed land bridges (Hewitt, 2000). These geological events shaped patterns of movement and migration for many species. In some cases, these events led to population isolation, particularly among island groups formerly connected to the mainland, resulting in reduced gene flow, increased genetic drift, and local adaptation, ultimately creating genomic and phenotypic divergence (e.g. Avila-Cervantes et al., 2021; Hansen et al., 2023; Howell et al., 2025; Sherlock et al., 2025).

Whole genome sequencing (WGS) has become an essential tool to infer phylogenies, reconstruct demographic histories, and understand how evolutionary forces drive population divergence over time (e.g. Bourgeois & Warren, 2021; Kessler & Shafer, 2024; Martchenko & Shafer, 2023). Whole genome sequencing enables the detection of genomic islands of differentiation, regions of the genome that show

elevated relative genetic differentiation in allele frequencies (F_{ST}), between populations or species (Irwin et al., 2018; Wright, 1949). These islands can arise under different scenarios, each of which will exhibit different patterns for nucleotide diversity within populations (π), absolute nucleotide divergence between populations (d_{XY}), and Tajima's D (Irwin et al., 2018). Genomic islands of differentiation, characterized by elevated F_{ST} and d_{XY} values, are strong candidates for barrier loci which contribute to reproductive isolation between populations experiencing gene flow (Irwin et al., 2018; Ravinet et al., 2017). In allopatric populations, divergent selection pressures acting on unevenly inherited ancestral polymorphisms can lead to the fixation of different alleles, also resulting in signals of elevated F_{ST} and d_{XY} (Guerrero & Hahn, 2017; Han et al., 2017; B. Wang et al., 2019). When combined with reduced π and negative Tajima's D, these patterns likely indicate loci under selection since the population split (Montejo-Kovacevich et al., 2022; Zhai et al., 2009). While WGS effectively detects such genomic patterns, it does not directly reveal the functional impacts of these variants on phenotypes.

Transcriptomic data capture gene activity and can help reveal molecular mechanisms through which genetic variants influence traits. RNA sequencing (RNA-seq) quantifies genome-wide patterns of gene expression, allowing the identification of differentially expressed genes (DEGs) in population-level studies (e.g. Cheviron et al., 2012; Pravosudov et al., 2013). Differential gene expression (DGE) directly influences phenotypes by altering the levels and timing of protein synthesis (Zuckerkandl & Pauling, 1965), which can affect an organism's development, physiology, and response to environmental challenges, thereby shaping observable traits and adaptive potential (Romero et al., 2012). While WGS can be used to identify genomic regions and

candidate loci under selection, RNA-seq provides insights into phenotypic variation by quantifying transcriptomic abundance and differences across populations. Thus, integrating WGS and RNA-seq data offers a more comprehensive understanding of the processes driving population divergence (Y. Huang et al., 2019; Wooldridge et al., 2022).

Bon Portage Island (BPI), located 3 km off the southwest coast of Nova Scotia, Canada, is home to a population of masked shrews (*Sorex cinereus*). The island spans about 3 km in length and less than 1 km in width, covering 340 acres of undeveloped land. It is believed to have been separated from the mainland around 20 kya (Roland, 1982; Stewart & Baker, 1992). The shrews on BPI exhibit considerable phenotypic variation from the mainland shrews, such as a smaller body size, accelerated epigenetic aging, and a unique feeding habit and peptidase digestive enzyme – all consistent with insular evolution (Cossette et al., 2023; MacPherson & Stewart, 2003; Stewart et al., 1989; Stewart & Baker, 1992). Notably, BPI shrews travel up to 60 meters to forage in intertidal zones (Stewart et al., 1989) where they consume kelp flies (*Coelopidae*) and they are the only *Sorex* shrews known to feed on sand fleas (*Platorchestia platensis*), a behavior likely driven by the island's distinct coastal environment (MacPherson & Stewart, 2003; Stewart et al., 1989). Previous work found that BPI shrews have a unique allele for the cytosolic non-specific dipeptidase 2 (*CNDP2*) gene (MacPherson & Stewart, 2003; Nandi, 2025), involved in intracellular amino acid and dipeptide metabolism (Pfeffer et al., 2024). It is hypothesized that *CNDP2* has undergone adaptive evolution in shrews to enhance their ability to digest sand fleas more efficiently (MacPherson & Stewart, 2003; McAlpine, 2009). Furthermore, Cossette *et al.* (2023) found differentially methylated CpGs between BPI

and mainland shrews linked to genes associated with digestive system related pathways, providing additional support for the hypothesis that BPI shrews are undergoing local adaptation related to their specialized diet.

Here, we assembled and annotated a high quality *de novo* genome, and sequenced whole genomes and transcriptomes from four distinct masked shrew populations to investigate the evolutionary history and evidence of divergent selection in the insular population on BPI. Our objectives were to (i) reconstruct and model the demographic history of the BPI shrew population, (ii) detect candidate genomic islands of differentiation potentially shaped by divergent selection, and (iii) identify DEGs between BPI shrews and mainland populations, as well as (iv) assess the extent of the overlap between these genomic islands of differentiation and DEGs. We hypothesized that the distinctive phenotypes observed in BPI shrews, such as their specialized feeding behavior, are the result of local adaptation to the island's coastal environment. We predict that these phenotypic traits are associated with genomic regions shaped by divergent selection, and that genes involved in digestion (i.e. *CNDP2*) and metabolism will show both genetic differentiation and differential expression between BPI and other populations. Furthermore, we expect some overlap between genomic islands of differentiation and DEGs, as regions near these islands may harbor genes that contribute to phenotypic divergence through regulatory changes in gene expression (McGirr & Martin, 2020; Verta & Jones, 2019).

Results

Genome assembly and annotations

The three HiFi cells from one individual produced 86,791,021,970 bp (~30x coverage) of data with an average length of 18,225 bp. We generated 538,468,674 paired Hi-C reads. The final primary assembly was 2.82 Gb in 343 scaffolds and N50 and L50 values of 80,745,324 bp and 13 respectively (Table 4.1). The primary assembly had a Mercury QV of 57.99 and completeness of 84.81. Approximately 49.14% of the genome was made of repetitive sequences and the BUSCO assessment revealed that 93.79% of orthologs were present and single-copy (Figure S4.1). A total of 15,123 unique protein-coding genes with 50% match were identified (Table 4.1). The assembled mitochondrial genome was 17,082 bp long and all 37 mitochondrial genes were present (Figure S4.2).

Genome Summary Statistics and Population Divergence

ANGSD called 80,926,708 SNPs across the 24 masked shrew samples with coverage varying between (6X and 18X) (Table S4.1). The PCA from allele frequencies and the NGSadmix analysis clustered the samples in 4 different populations with no admixed individuals (Figure 4.1b, c). The inbreeding coefficient values were between -0.0265 and 0.0393, and nucleotide diversity values from 2.80 x 10^{-3} to 7.43 x 10^{-3} (Table 4.2). Tajima's D values were negative in the Ontario (ON) population (-0.123), while the highest positive values were observed in the island populations: Long Island (LI) (0.517) and BPI (0.727). Pairwise averaged F_{ST} values ranged from 0.0827 to 0.236, with the lowest differentiation observed between the two geographically closest

populations, Sandy Cove (SC) and LI. The highest level of genetic differentiation was found between the island populations, BPI and LI. Pairwise averaged d_{XY} values were similar among all Nova Scotia populations, ranging from 5.72 x 10^{-3} to 5.82 x 10^{-3}. d_{XY} values between BPI and ON was 8.36 x 10^{-3}. All populations had low F_{HBD} values (<0.3%).

The highest f_2 estimate was observed between BPI and LI (0.431), and the lowest was observed between LI and SC (0.189) (Table S4.2). All f_3 values were positive, indicating no evidence of admixture in any of the tested population triplets. Furthermore, all f_4 tests of the form f_4(ON, B; C, D) yielded estimates near zero, ranging from -5.59×10^{-4} to 3.57×10^{-4}, consistent with either minimal gene flow or closely spaced divergence times among BPI, LI and SC.

Demographic modeling

We performed SMC analyses on all individual genomes and at the population level for higher-coverage samples. The three Nova Scotia populations showed a peak in N_e around 100 kya, followed by a steady decline with a broadly similar trajectory across all three (Figure 4.2a). However, the SC population displayed a rapid expansion approximately 5 kya, increasing from an estimated 20,000 to 70,000 individuals. In contrast, LI and BPI populations had flat N_e estimates of approximately 13,000 individuals each at this time. The ON population showed a distinct trajectory, diverging from other groups around 100 kya, with a wider range of N_e values across individuals and generally higher N_e than all other populations.

Using Momi2 we estimated the ON population split from the Nova Scotia population around 175 kya (Figure 4.2b). Within the Nova Scotia groups, BPI diverged

approximately 16–17 kya, followed by LI separating from the mainland population around 12–13 kya. Consistent with the MSMC results, the current N_e of BPI is estimated at approximately 12,000 individuals, while LI is estimated at around 14,000 individuals. Current N_e estimates based on Momi2 suggest slightly higher values compared to MSMC, particularly for SC (~218,000) and ON (~672,000).

Differential gene expression

To investigate gene expression differences among populations, we compared gene expression profiles of BPI shrews to those from SC and LI. A total of 244 DEGs were identified, with 177 significantly underexpressed genes and 67 significantly overexpressed genes in BPI shrews compared to SC and LI (Figure 4.3a, Table S4.3). The gene with the most significantly reduced expression in BPI was *FCN2* (LFC = -4.00, adjusted p-value = 1.32×10^{-08}), which encodes a protein involved in the innate immune system (Messias-Reason et al., 2009). In contrast, *PRODH* (LFC = 3.89, adjusted p-value = 3.41×10^{-14}) was the gene with the most significantly increased expression and is involved in proline catabolism (Tanner, 2019). Genes associated with fatty acid metabolism (e.g. *ACOT4*, *ACOT11*, *HADH,* and *PECR*), and to metabolic regulation (*LEPR, PDK4*), were significantly differentially expressed between BPI and the other populations (Table 4.3). Overall, differentially expressed genes were associated with various pathways including immune system, structural and development processes, as well as signaling pathways (Table S4.4).

Genomic islands of differentiation and DEGs overlap

We identified 128 candidate genomic islands of differentiation (top 5% elevated F_{ST} and d_{XY}) shared between BPI-SC and BPI-LI comparisons but absent in SC-LI. These islands harbored 52 annotated genes that were associated with various pathways, including fatty acid elongation and lipid metabolism (*ELOVL2/5, FABP5/P3*, and *CEBPZ*), bone development and growth factor (*BMP5/7,* and *NODAL*), and chromatin regulation (*H2BC14, H1-4,* and *Histone H2B type 1*) (Figure 4.3b, c, Table S4.5, Table S4.6). Of the 52 genes, only one, *SNAI2*, was differentially expressed between BPI and the other populations. *SNAI2* encodes a transcription factor that represses gene expression (Zhou et al., 2019).

Discussion

Geographic isolation and local adaptation lead to population divergence over time (Hope et al., 2013). In North America, repeated glacial cycles fragmented habitats, which shaped the evolutionary paths of many species and populations (Shafer et al., 2010). One clear example is the masked shrew population on BPI, which exhibit distinct phenotypic differences from their mainland counterparts. To understand the evolutionary processes behind this divergence, we assembled and annotated a high-quality *de novo* genome and analyzed whole-genome and transcriptome data from four masked shrew populations. This allowed us to reconstruct the demographic history of the BPI population and to investigate how genomic and transcriptomic patterns differ between island and mainland shrews.

Impact of the Last Glacial Maximum on shrews in Nova Scotia

During the last glaciation cycle, the Laurentide Ice Sheet (LIS) began forming over the eastern Canadian Arctic around 116–110 kya and underwent several episodes of expansion and retreat before reaching its maximum extent around 26–25 kya (Stokes, 2017). At its peak, the LIS completely covered present day Nova Scotia (Dalton et al., 2020, 2022). As the ice advanced and retreated, it reshaped Nova Scotia's landscape through erosion and deposition, carving valleys and forming geographic features across the region. When the ice began to melt ~20 kya, global sea levels rose significantly; this post-glacial sea-level rise inundated coastal areas and fragmented the land, creating over 3,800 islands around Nova Scotia (Parker, 2012). The change in sea level played a key role in the isolation of island populations, including masked shrews on BPI, which are thought to have been cut off from the mainland during this post-Wisconsin glacial period less than 20 kya (Stewart & Baker, 1992).

Our demographic analyses support this scenario (Figure 4.2), where shrew populations in Nova Scotia appear to have diverged from the ON population before the last glaciation cycle, around 175 kya, suggesting long-term separation shaped by earlier glacial cycles. Historical patterns show a distinct demographic trajectory for ON, including earlier divergence and consistently higher N_e than the Nova Scotia populations, consistent with a presumed larger *Sorex cinereus* population. Following this, estimates place the divergence of BPI from the Nova Scotia mainland population at approximately 16-17 kya, consistent with the timing of sea-level rise and island isolation (Figure 4.2c). Shortly after, the LI population diverged from the mainland around 12-13 kya, likely reflecting continued fragmentation due to rising waters. Furthermore, the lack of admixture and clear population structure between BPI, SC,

and LI populations support limited gene flow, driven by geographic distance and physical isolation (Figure 4.1).

All Nova Scotia populations appear to have undergone population decline during the Wisconsin glaciation, consistent with broader patterns observed in other mammals during this period (Blois et al., 2010; Mann et al., 2015). Post-glaciation, the SC population, which is part of mainland Nova Scotia, appears to have rebounded whereas both island populations continued declining. Smaller isolated populations are generally expected to exhibit reduced genetic diversity, increased inbreeding, and a higher genetic load (Frankham et al., 2010). In line with these expectations, nucleotide diversity was lower in the island populations compared to the mainland, but F values did not reflect elevated levels of inbreeding (Table 4.2). Consistent with this, there were few long HBD segments detected, with only one segment over 1 Mb found in an individual from BPI (Table S4.7). Based on recombination rates from other *Sorex* species (Borodin et al., 2008), an HBD segment of this length likely traces back to a common ancestor more than 100 generations ago, and thus recent inbreeding is unlikely to be a concern on the island, contradictory to other shrew species where inbreeding has been documented (Duarte et al., 2003; Stockley et al., 1993). Here, the low levels of diversity observed are likely the result of historical bottlenecks and prolonged, relatively small effective population size. Furthermore, the effects of any inbreeding in the island populations may be partially mitigated by fast purging of deleterious alleles (Duarte et al., 2003), particularly since *Sorex* shrews experience rapid (yearly) generational turnover and relatively high reproductive output (Churchfield, 1990). Another factor that may reduce this effect is a polyandrous mating system, in which females mate with several males, resulting in multiple paternities

within a single litter (Batova et al., 2021). Although it is unknown whether this population exhibits such behavior, other shrew species within the genus *Sorex* have been documented to do so (Tegelström et al., 1991). These mechanisms would facilitate the maintenance of small but stable populations despite reduced genetic diversity, allowing the long-term persistence of island populations.

Evidence of divergent selection and differential expression of fatty acid related genes

Regions of high F_{ST} and d_{XY} are often interpreted as signatures of divergence with gene flow, potentially highlighting loci involved in reproductive isolation (Irwin et al., 2018). Similar patterns can also arise without gene flow, especially when geographically varying selection acts on ancestral balanced polymorphisms that have been unequally sorted among descendant lineages (B. Wang et al., 2019). If the descendant populations inherit ancestral polymorphisms that are unequally sorted, and experience divergent environmental pressures (e.g., differences in diet), this may drive the fixation of alternate alleles in each lineage over time, resulting in genomic islands of differentiation with both elevated F_{ST} and d_{XY} (Guerrero & Hahn, 2017; Han et al., 2017; B. Wang et al., 2019). These patterns are consistent with loci under selection since the population split, particularly when reduced nucleotide diversity (π) and negative Tajima's D are also observed (Montejo-Kovacevich et al., 2022; Zhai et al., 2009).

BPI shrews have developed a specialized diet adapted to their littoral habitat, feeding primarily of sand fleas (*Platorchestia platensis*) and kelp flies (*Coelopidae*) (Stewart et al., 1989). Both are fatty prey species, notably rich in polyunsaturated fatty acids (PUFAs) (Albalat et al., 2025; Jiménez-Prada et al., 2018; Nandi, 2025; Warwas

et al., 2024), which are produced by algae and accumulate through the food web (Harwood, 2019). As mammals are unable to synthesize PUFAs *de novo*, these essential nutrients must be acquired through their diet (Burdge, 2018). Once ingested, PUFAs are converted into acyl-CoA derivatives, enabling their function in membrane synthesis, gene regulation, inflammatory responses, neural development, and energy production via β-oxidation (Benatti et al., 2004; Calder, 2015; Mititelu et al., 2024). Candidate genomic islands of differentiation between BPI and SC and LI harbored genes involved in fatty acids metabolism and adipocytes (Figure 4.3b, c, Table 4.3). Notably, *ELOVL2* and *ELOVL5* catalyze the first, rate-limiting step of the elongation of long-chain fatty acids (LCFAs) and PUFAs (Monroig et al., 2016). *FABP5* and *FABP5P3* regulate intracellular levels of fatty acids (Xu et al., 2022) and *CEBPZ* modulated the differentiation and proliferation of preadipocytes (Y. Chen et al., 2022). These genes were in windows with elevated F_{ST} and d_{XY}, reduced π, and negative Tajima's D in BPI populations, suggesting potential local adaptation to their PUFA-rich diet.

Several lipid and fatty acid metabolism genes were also significantly differentially expressed in BPI shrews compared to the other populations (Figure 4.3a). *ACOT4* and *ACOT11* are acyl-CoA thioesterase enzymes that hydrolyze fatty acyl-CoA, playing key roles in regulating fatty acid metabolism (Tillander et al., 2017). *HADH* encodes the mitochondrial enzyme 3-hydroxyacyl-CoA dehydrogenase, which catalyzes a crucial step in the β-oxidation of fatty acids to generate ATP (X. Wang et al., 2022), and *PECR* is a peroxisomal enzyme that catalyzes the last step of the fatty acid elongation cycle (B. Shen et al., 2019). Furthermore, gene expression for the leptin receptor (*LEPR*) was significantly lower in the livers of BPI shrews compared to other populations. Leptin is a hormone secreted by adipose tissue that regulates appetite

and energy expenditure by signaling satiety and modulating metabolic processes (J. M. Friedman & Halaas, 1998; Klok et al., 2007). Previous studies have suggested that leptin plays a critical role in modulating seasonal feeding behavior in shrews (Nieminen & Hyvärinen, 2000; Thomas et al., 2025), suggesting that altered leptin signaling in this species may be part of a more comprehensive adaptive strategy to maintain their high energetical demands by sustaining foraging and metabolic flexibility. *PDK4*, which regulates energy metabolism by minimizing carbohydrate oxidation (Buck et al., 2002), was also differentially expressed in BPI shrews. Given that masked shrews exhibit exceptionally high metabolic rates and possess minimal fat reserves, they are highly vulnerable to starvation during periods of food scarcity (Churchfield, 1990) and are likely under strong selective pressure to optimize fuel use based on the most accessible energy sources. Consequently, differential expression of genes associated with metabolism and energy regulation in BPI shrews may reflect adaptive molecular mechanisms that help meet their extreme energetic demands.

Overlaps between islands of differentiations and DEGs

Both the islands of differentiation and DGE analyses identified outliers associated with fatty acid metabolism, with some specifically linked to PUFAs. However, we observed limited direct overlap between the genomic islands of differentiation and DEGs, consistent with results from other studies (Szukala et al., 2023; Wegner et al., 2020). This discrepancy may reflect the influence of *trans*-regulatory mechanisms, where expression is modulated by regulatory elements located elsewhere in the genome rather than *cis*-acting variation at the expressed gene itself (Signor & Nuzhdin, 2018). Additionally, transcriptional changes are highly dynamic and can be rapidly

induced (e.g., within minutes) through regulatory responses to environmental stimuli, rather than being solely determined by fixed genetic differences (DeBiasse & Kelly, 2016; López-Maury et al., 2008), Accordingly, BPI shrews may display DEGs as a consequence of their current diet and environment rather than fixed evolutionary differences. Although this study focused on the liver because of its role in digestion and metabolism and the presence of *CNDP2*, gene expression is tissue-specific, meaning that genomic islands of differentiation could affect expression in other tissues that were not examined.

Furthermore, genomic windows containing the *CNDP2* gene were not among the top genomic islands of differentiation (Table S4.8). Although this pattern suggests the surrounding region may not be undergoing linked selection, it remains possible that selection is acting specifically on the allele rather than on the broader genomic window. All sampled individuals in the BPI population carried the unique allele, with 70% of individuals being homozygous, while the allele was completely absent in individuals from the SC and LI populations (Nandi, 2025). Additionally, there were no significant differences in overall *CNDP2* gene expression between the BPI population and the other populations, indicating that the unique allele does not appear to affect total transcript abundance. However, this does not exclude the possibility that the allele influences isoform-level variation or protein function. Notably, the allele results in a nonsynonymous substitution from glutamic acid to lysine, which computational predictions suggest may slightly reduce protein stability (Nandi, 2025). Overall, our findings highlight the complex molecular basis of local adaptation in BPI shrews.

Materials and Methods

Sampling

Masked shrew specimens were collected in July and August 2022 from four locations: Peterborough County, in Ontario (ON), Canada, and Sandy Cove (SC), Long Island (LI), and Bon Portage Island (BPI) in Nova Scotia, Canada (Figure 4.1a). Pitfall cone traps were used and monitored at a maximum of 12-hour intervals. Shrews were euthanized using cervical dislocation or thoracic compressions (Animal Care Certificate no. 26234). We collected liver, heart, tail and hindfoot samples and stored half of the tissues in RNAlater and the other half in empty Eppendorf tubes and flash froze them with liquid nitrogen. A single smoky shrew (*S. fumeus*) was collected as an outgroup.

Genome sequencing, assembly, and annotation

Masked shrew tail tissue from one individual was submitted to the DNA Sequencing & Genotyping Center at the University of Delaware, Newark, USA, for high molecular weight DNA extraction. PacBio HiFi library preparation and sequencing on three SMRT cells using the PacBio Sequel II platform was conducted. Arima High Coverage Hi-C library preparation and sequencing on an Illumina NextSeq 2000 for 300 cycles on a P2 flow cell to generate 150 bp paired-end reads was performed. We assembled the masked shrew genome following the Vertebrate Genomes Project (VGP) pipeline v2.0 (Larivière et al., 2024). Prior to running the pipeline, we used cutadapt V4.8 (M. Martin, 2011) to remove reads that contain adapters. We ran the pipeline with default parameters and include the purge_dups steps. For the YaHS scaffolding step we used the "-e 'GATC,GANTC,CTNAG,TTAA' -l 5000 --no-contig-ec" flags.

Subsequently, we ran the Assembly-decontamination-VGP9 pipeline v0.1.4 on the primary and alternate assemblies to identify any contaminated and/or mitochondrial scaffolds. WindowMasker v1.0 (Morgulis et al., 2006) was used to mask highly repetitive and low complexity regions from the primary assembly. Genome completeness was assessed using Benchmarking universal single-copy orthologues (BUSCO) version 5.8.0 (Simão et al., 2015) with the mammalia_odb10 data set.

Liver and heart tissue samples from the same individual as mentioned above were submitted for Iso-Seq RNA extractions at the DNA Sequencing & Genotyping Center. PacBio Iso-Seq library preparation for each tissue was done before sequencing on one cell using the PacBio Sequel IIe platform. We received demultiplexed full-length non-concatemer (FLNC) polyA removed data. To annotate the genome, we assembled transcripts with the Iso-Seq and population RNA-Seq data (see RNA sequencing strategy below) separately and then merged the resulting transcript annotations to create a comprehensive gene model. First, we identified isoforms from the Iso-Seq data using the nf-core/isoseq pipeline v2.0.0 (P. A. Ewels et al., 2020; Guizard et al., 2023) starting at the mapping entry point for the Iso-Seq data. We used the default parameters and selected minimap2 (H. Li, 2018) as aligner for mapping. We then ran TAMA Merge (Kuo et al., 2020) to combine the output BED files from both heart and liver samples. Secondly, we assessed the RNA-seq data using FastQC v0.11.9 (Andrews, 2010), then removed adapters and filtered out short (< 20bp) and low quality (q < 20) reads using cutadapt v5.0 (Martin, 2011). We used the masked shrew genome and GTF file resulting from the Iso-Seq annotation step above to align the RNA-seq reads using STAR v2.7.11b (Dobin et al., 2013). We only kept unique reads and used MarkDuplicates v3.1.0 (Broad Institute, 2019) to identify and remove duplicates.

The reads were then assembled into transcripts and expression levels were estimated for each sample with StringTie v2.2.3 (Pertea et al., 2015). We retained novel transcripts identified by StringTie that were not present in the Iso-Seq annotation file and had an FPKM value of at least 0.5 in all samples (Y. Zhao et al., 2024).

To predict open reading frames (ORF) and identify potential targets for nonsense-mediated decay (NMD) from both Iso-Seq and RNA-seq transcript annotations we used the TAMA GO: ORF and NMD predictions pipeline (Kuo et al., 2020). The output amino acid sequences were aligned to the Swiss-Prot mammal database (Bairoch & Apweiler, 1997) using blast+ v2.14.0 (Camacho et al., 2009) with an evalue cutoff of 1×10^{-10}. Results from the Iso-Seq and RNA-seq transcript annotations were parsed and merged to create the final annotation file. The picard v3.1.0 CollectAlignmentSummaryMetrics command (Broad Institute, 2019), samtools v1.20 flagstat (Danecek et al., 2021) and MultiQC v1.13 (P. Ewels et al., 2016) were used to obtain alignment summaries both here and in several other analyses described later in the manuscript.

Whole genome sequencing and variant calling

We extracted DNA from the shrew tissue samples mentioned above and from samples previously collected by Cossette et al. (2023) (Table S4.1) using the Qiagen DNeasy Blood and Tissue Kit following manufacturer's instructions and checked concentrations using Invitrogen Qubit assays. DNA was submitted to The Centre for Applied Genomics (TCAG) in Toronto, Canada, for PCR-free library preparation. WGS was performed on an Illumina NovaSeq X 25B flow cell (1 lane) for some samples, and on a NovaSeq S4 flow cell (0.5 lanes) for others, generating 150 bp paired-end reads.

We used FastQC v0.11.9 (Andrews, 2010) to assess the quality of the raw reads. Trimmomatic v0.36 (Bolger et al., 2014) and fastp v0.23.4 (S. Chen et al., 2018) were used to trim adapters and remove polyG tails respectively. We used bwa-mem2 v2.2.1 (H. Li & Durbin, 2009) with default parameters and included read group information to align the reads to the masked shrew genome assembly and samtools v1.20 (Danecek et al., 2021) to sort the output. We used the picard MarkDuplicate v3.1.0 (Broad Institute, 2019) command to identify duplicates and Sambamba view v0.8.0 (Tarasov et al., 2015) to remove them. GATK RealignerTargetCreator and IndelRealigner v3.8 commands were used for local realignment around indels. Sambamba flagstat was used for quality checks and the final coverage was calculated using mosdepth v0.3.8 (Pedersen & Quinlan, 2018). The data from two smoky shrews, one sourced from NCBI GenBank (GCA_029834395.2) were included in the data processing steps to facilitate single nucleotide polymorphism (SNP) polarization for subsequent analyses.

We ran ANGSD v0.940 (Korneliussen et al., 2014) to call variants for all samples (including smoky shrews) with the following parameters: -SNP_pval 1e-6 -minMapQ 30 -minQ 30 -skipTriallelic 1 -doMajorMinor 1 -gl 2 -doGlf 2 -doMaf 1 -doPost 1 -doGeno 4 -doCounts 1 -doBcf 1. Using the output VCF, we polarized SNPs to identify the “ancestral” and “derived” states of each variant, using the smoky shrew as an outgroup (Esteva et al., 2010). We used est-sfs v.2.04 (Keightley & Jackson, 2018) with the Kimura2 parameter model to infer the ancestral state probabilities for each SNP. A custom script from Battilani et al. (2025) was then used to extract the most probable ancestral state for each variant and align it with its corresponding genomic

position. These ancestral states were subsequently used to overwrite the matching positions in the reference genome using the base_changer2.py script (Lipworth, 2017).

Population structure, summary statistics

We used PCAngsd v1.2 (Meisner & Albrechtsen, 2018) and NGSadmix v33 (Skotte et al., 2013) to investigate population clustering. The inbreeding coefficient (F) for each individual was calculated with PLINK v2.00-20231024-avx2 (Purcell et al., 2007). All subsequent genome-wide analyses were conducted using sliding windows of 50 Kb with a 10 Kb step size. For each population, nucleotide diversity (π) was estimated using vcftools v0.1.16 (Danecek et al., 2011), and Tajima's D were computed using an unfolded site-frequency spectrums (SFS) generated by ANGSD v0.940 (Korneliussen et al., 2014). F_{ST} and d_{XY} values, were calculated using popgenWindows.py v0.5 (S. Martin, 2025).

To assess inbreeding levels for each individual, we identified homozygous-by-descent (HBD) segments on the scaffold making up the N90 of the genome (n=47). We used RZooRoH v0.3.2 (Bertrand et al., 2019) and the genotype probabilities generated by ANGSD. The analyses were conducted using the default 10K model, which assumes multiple HBD classes. The proportion of the genome covered by HBD (F_{HBD}) was then calculated by dividing the total sum of the HBD that were over 500 kb per individual by the genome's N90 length.

We used ADMIXTOOLS v2.0 (Maier et al., 2023) in r v.4.4.0 (R Core Team, 2021) to investigate genetic relationships, population divergence, and potential admixture among the studied populations. Pairwise f_2-statistics were calculated across all population pairs to quantify genetic drift. To test for admixture, we computed f_3-

statistics in the form f_3(A; B, C), evaluating all possible combinations among BPI, SC, and LI populations. Additionally, we calculated f_4-statistics in the form f_4(ON, B; C, D) for all possible population combinations to test whether the populations fit a simple tree-like relationship and to identify signals of gene flow or shared ancestry.

Demographic modeling

We ran the multiple sequentially Markovian coalescent (MSMC) model v2.1.4 (Schiffels & Wang, 2020) to estimate the historical effective population size (N_e) of the masked shrew populations using every WGS sample. Following the MSMC2 workflow (Rick, 2024) we created a 35-mer mappability mask file for scaffolds making up the N90 of the genome using Heng Li's SNPable program (2009). We called variants using bcftools v1.21 mpileup (Danecek et al., 2021) for sites with at least 10X coverage to generate the input for the MSMC2 runs. First, (i) we ran MSMC2 on each sample separately, and then (ii) with samples that had an overall coverage ~15X (BPI n=5, SC n=3) from the same population using the unphased data. We ran 20 iterations for each run with the 1*2+25*1+1*2+1*3 segmentation pattern, used a generation time of 1 year and the default MSMC2 mutation rate 1.25e-8.

We used Momi v2.1.20 (Kamm et al., 2020) to estimate current N_e and divergence times among the shrew populations. Raw allele counts were computed using the polarized SNPs, restricted to the N90 scaffolds. The counts were then used to generate an unfolded site frequency spectrum (SFS). Demographic models were fit across a range of N_e and divergence time parameters, informed in part by estimates from the MSMC results. Including migration parameters did not improve model fit. The best-fitting model was selected based on the highest log-likelihood score and was re-

run using 100 bootstrap replicates of the SFS to assess uncertainty in parameter estimates.

Differential gene expression analysis

We extracted RNA from liver and heart tissue samples stored in RNAlater using the QIAGEN QIAwave DNA/RNA Mini Kit following manufacturer's instructions and checked concentrations using Invitrogen Qubit assays. RNA extracts were submitted to TCAG for NEB Ultra II Directional polyA RNA preparation. Sequencing was done on an Illumina NovaSeq S4 to generate 150 bp paired-end reads with 15-25 million reads/sample. All data were used for genome annotation, as described above; however, only liver RNA-seq samples were used for the DGE analysis, as our focus was on population-level differences in gene expression associated with the digestive system, and *CNDP2* is expressed in the liver. We used the final merged (Iso-Seq + RNA-seq) GTF file, keeping only genes that had a minimum 50% match, to align the RNA-seq reads using STAR v2.7.11b (Dobin et al., 2013) to generate gene counts. Prefiltering was done to remove genes with low (< 5) read counts across all samples to avoid biases in fold change due to weakly expressed genes (Love et al., 2014; Singh et al., 2017). DGE was analyzed using DESeq2 v1.48.1 (Love et al., 2014) in r v.4.4.2 (R Core Team, 2021) with the experimental design accounting for location. The DESeq function was then applied to normalize counts, estimate dispersion, and fit a negative binomial model to test for differential expression between BPI and non-BPI shrews. Genes with an adjusted p-value (FDR) < 0.05 were considered significantly differentially expressed.

Genomic islands of differentiation and overlaps

Regions of genomic differentiation between Nova Scotia shrew populations were identified through pairwise comparisons based on elevated F_{ST} and d_{XY} values (see above). Candidate genomic islands of differentiation were defined as regions falling within the top 5% of each metric (Gu et al., 2023), shared between the BPI-SC and BPI-LI comparisons, but absent in the SC-LI comparison. We identified direct overlaps, without any flanking regions, between the candidate genomic islands of differentiation and DEGs using the GenomicRanges package v.1.58.0 (Lawrence et al., 2013) in r v.4.4.2 (R Core Team, 2021). DAVID, the Database for Annotation, Visualization and Integrated Discovery (D. W. Huang et al., 2009; Sherman et al., 2022), was used to identify enriched gene pathways for each analysis.

Acknowledgments

We would like to acknowledge that the work for this study was carried out on the traditional territory of the Mississauga (Michi Saagiig) Anishinaabeg and the Mi'kmaq. We are grateful to have had the opportunity to work on this land and thank the First Peoples for their care, stewardship, and teaching. We would like to thank the Jeong lab at Seoul National University, South Korea, for their support and guidance. We would also like to acknowledge The Centre for Applied Genomics in Toronto, Ontario, Canada, and the DNA Sequencing & Genotyping Center at the University of Delaware, Newark, US, for their expertise and insight with sequencing.

Figures and Tables

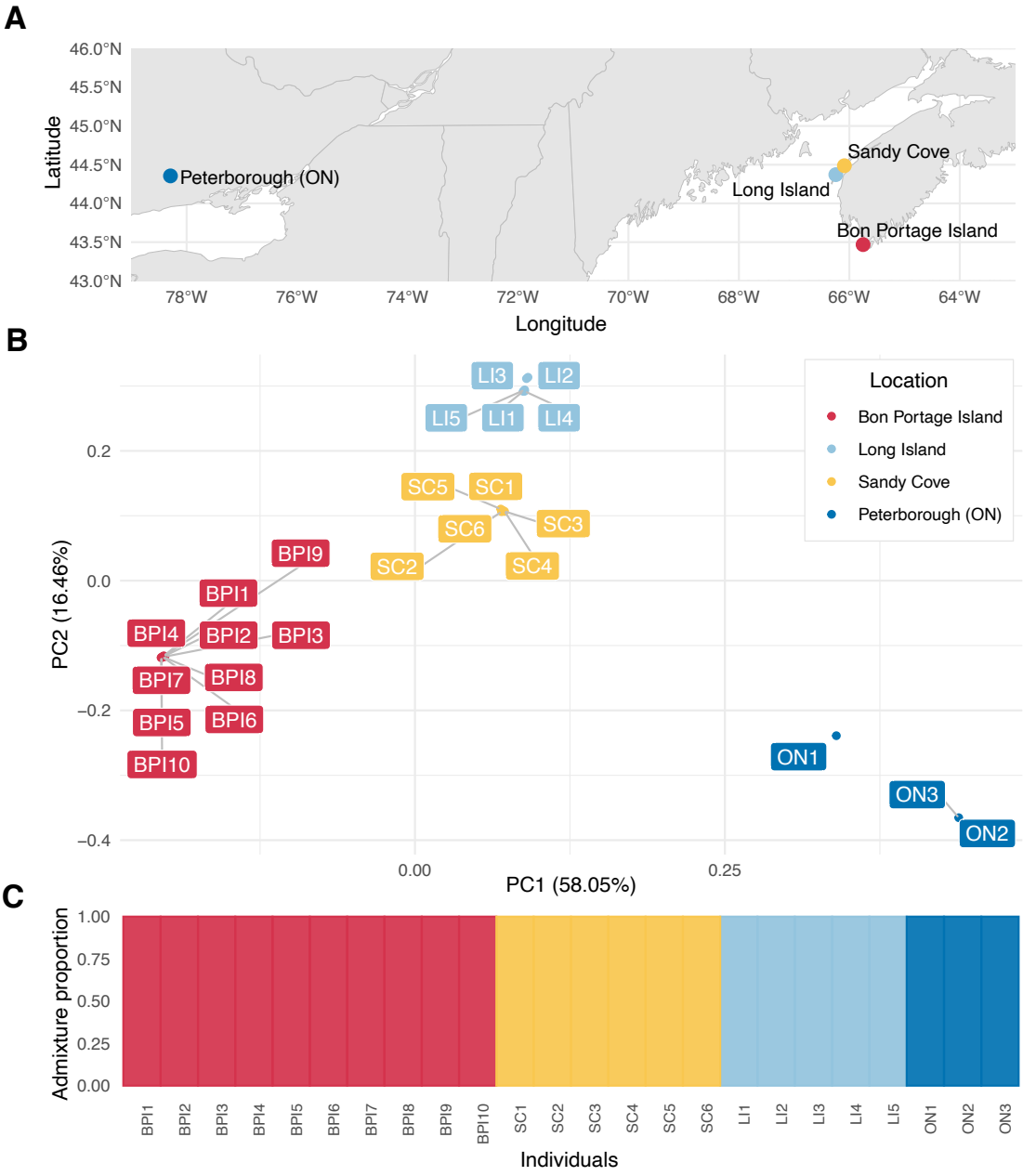

Figure 4.1. (a) Map of sampling locations. Population structure inferred from (b) PCA and (c) NGSadmix based on SNPs from WGS data.

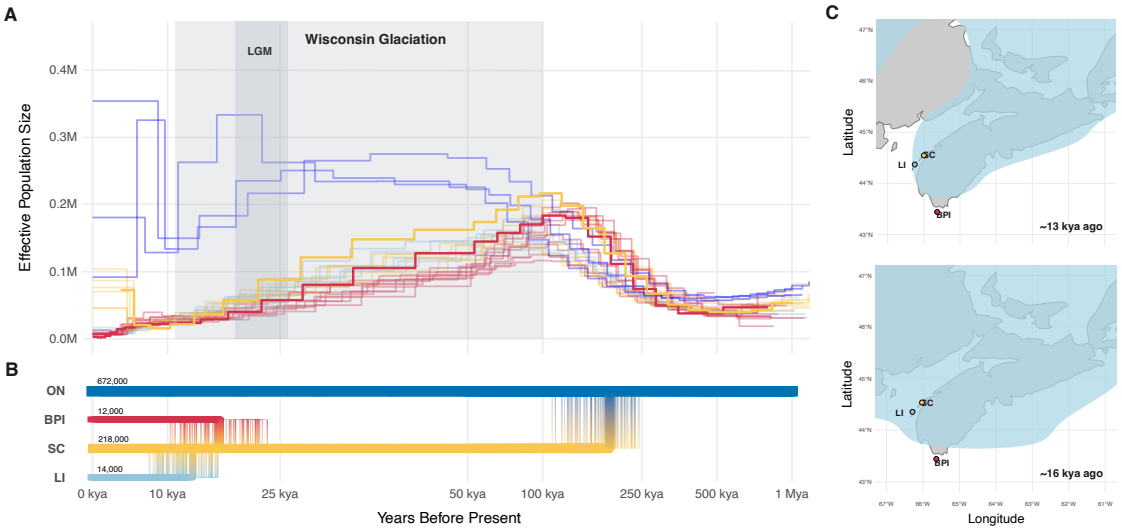

Figure 4.2. Demographic history and modeling. (a) SMC plot estimating historical effective population size of each shrew population. (b) Estimated split times and present effective population sizes from the Momi analysis with 100 bootstrap replicates. (c) Estimated ice coverage over Nova Scotia at the estimates split times of the BPI and LI populations from the mainland.

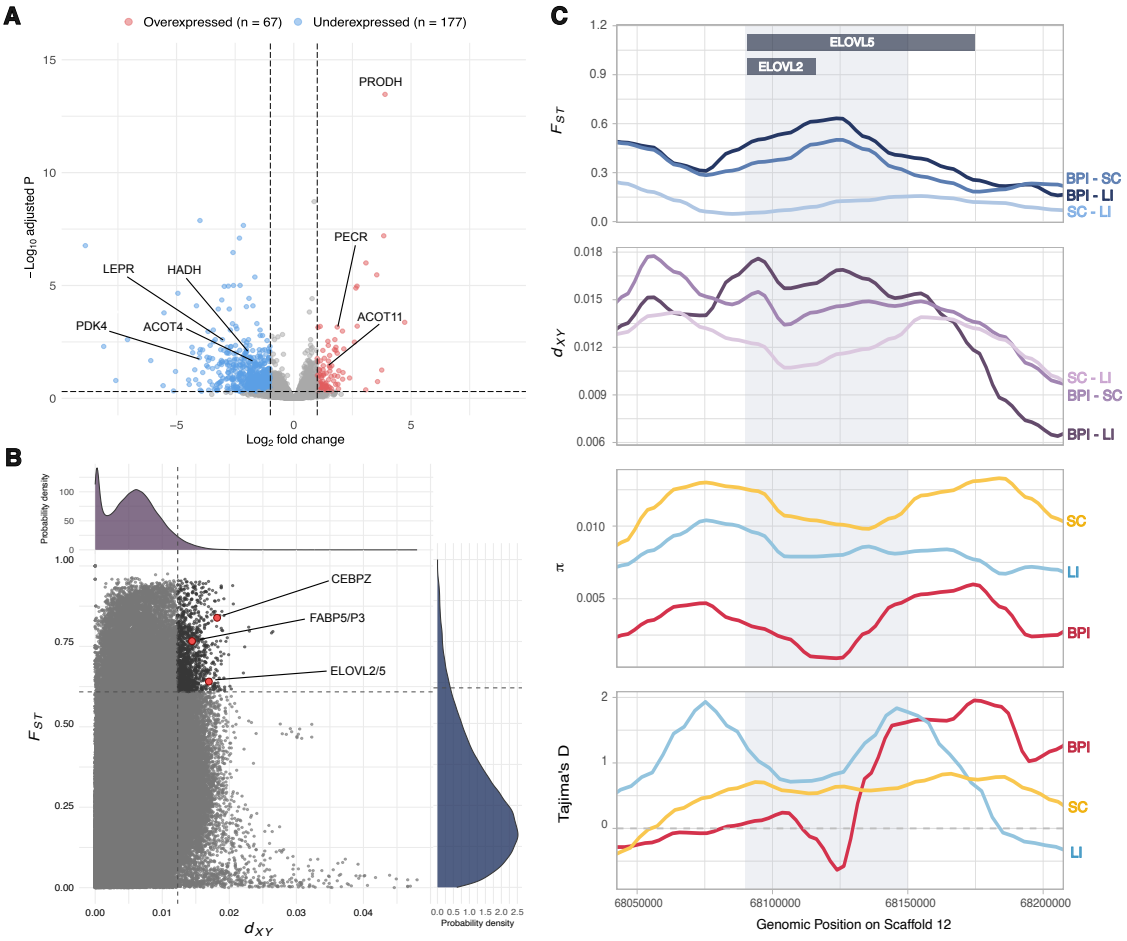

Figure 4.3. (a) Volcano plot of DEGs between BPI and SCI + LI from liver tissues. (b) Top 5% outliers for F_{ST} and d_{XY} between BPI and LI populations. Annotated genes in (a) and (b) are genes associated with lipid metabolism and regulation. (c) Genome scans (50 kb windows with 10 kb steps) for candidate island of differentiation in BPI shrews compared to SC and LI populations, overlapping ELOVL2 and ELOVL5 genes.

Table 4.1. Genome assembly metrics for the masked shrew genome generated from PacBio HiFi long-reads and Hi-C reads.

	Primary assembly
Assembly size	2.82 Gb
# of scaffolds	343
Scaffold N/L50	80,745,324 bp/13
Scaffold N/L90	6,969,911 bp/47
Average scaffold length	8,223,106 bp
Longest scaffold length	159,956,999 bp
Shortest scaffold length	19,156 bp
# of contigs	1,187
Contigs N/L50	5,462,505 bp/144
Contigs N/L90	1,310,465 bp/542
Average contig length	2,376,037 bp
Longest contig length	31,895,011 bp
Shortest contig length	19,156 bp
Number of gaps	844
GC content	43.02%
Merqury QV	57.99
Merqury completeness	84.81
% genome masked	49.14%
BUSCO results	S:8653, D:97, F:126, M:350, n:9226
# proteins-coding genes >50% match	15,123
Mitochondrial genome length	17,082 bp

Table 4.2. Summary statistics for within and between populations analyses. Values for π, Tajima's D, F_{ST}, and d_{XY} are averages from the sliding windows.

	BPI	**SC**	**LI**	**ON**
n	10	6	5	3
#snps	30,948,385	46,978,069	32,253,653	55,385,074
F	-0.0265	0.0393	-0.0211	0.0178
π	0.00280	0.00561	0.00370	0.00743
Tajima's D	0.727	0.201	0.517	-0.123
F_{HBD} > 500 kb (%)	0.281	0.163	0.186	0
	BPI - LI	**BPI - SC**	**SC - LI**	**BPI - ON**
F_{ST}	0.236	0.151	0.0827	0.223
d_{XY}	0.00572	0.00582	0.00571	0.00836

Table 4.3. Top DEGs and genes overlapping islands of differentiation between BPI and other NS populations, associated with lipid metabolism and regulation.

DEGs			
Name	**LFC**	**Adjusted p-value**	**Function**
ACOT4	-1.90	2.63×10^{-02}	Hydrolyzes fatty acyl-CoAs (Tillander et al., 2017)
ACOT11	1.51	3.64×10^{-02}	Hydrolyzes fatty acyl-CoAs (Tillander et al., 2017)
HADH	-2.04	9.80×10^{-03}	Mitochondrial 3-hydroxyacyl-CoA dehydrogenase, catalyzes fatty acid β-oxidation (X. Wang et al., 2022)
PECR	1.85	6.94×10^{-04}	Catalyzes the last step of the fatty acid elongation cycle (B. Shen et al., 2019)
LEPR	-3.04	2.50×10^{-03}	Regulates appetite and energy expenditure (J. M. Friedman & Halaas, 1998; Klok et al., 2007)
PDK4	-4.02	1.59×10^{-02}	Regulates energy metabolism by minimizing carbohydrate oxidation (Buck et al., 2002)

Islands of differentiation							
Name	F_{ST}		d_{XY}		**π BPI**	**Tajima's D BPI**	**Function**
	BPI - SC	**BPI - LI**	**BPI - SC**	**BPI - LI**			
ELOV2/ ELOV5	0.474 – 0.502	0.614 – 0.634	0.0142 – 0.0146	0.0160 – 0.0169	9.00×10^{-04} – 1.00×10^{-03}	-0.139 – -0.652	Catalyze elongation of long-chain fatty acids (Monroig et al., 2016)
FABP5/ FABP5P3	0.458 – 0.483	0.654 – 0.769	0.0129 – 0.0130	0.0129 – 0.0146	$5 .00 \times 10^{-04}$	-0.620 – -0.480	Regulate intracellular levels of fatty acids (Xu et al., 2022)
CEBPZ	0.262 – 0.422	0.705 – 0.849	0.0132 – 0.0150	0.0146 – 0.0188	8.00×10^{-04} – 3.10×10^{-03}	-2.505 – -2.243	Modulated the differentiation and proliferation of preadipocytes (Y. Chen et al., 2022)

CHAPTER 5: General Discussion

Mammals have undergone repeated episodes of diversification driven by major historical climatic shifts and geological events, which have shaped evolution at both the species and population level (Burgin et al., 2018; Hewitt, 2000). These processes have led to a wide range of specializations that allowed individuals to survive in diverse and often changing environments. Among mammals, shrews (family Soricidae) stand out as one of the most diverse groups, with over 400 extant species inhabiting various habitats worldwide (Burgin et al., 2018). This has led to a range of specialized traits such as echolocation (Forsman & Malmquist, 1988; Tomasi, 1979), venomous saliva (Kita et al., 2004), extreme seasonal physiological changes like Dehnel's phenomenon (Lázaro et al., 2017, 2019), and unique foraging behaviors and diets (Mendes-Soares & Rychlik, 2009; Stewart et al., 1989). Understanding how such traits arose and are maintained requires examining both macroevolutionary and microevolutionary trends. The aim of my Ph.D. was to develop genomic resources to explore the mechanisms driving these phenotypic adaptations in shrews. This involved using comparative genomic analyses at the species level and a multi-omics approach, integrating genomic, epigenomic, and transcriptomic data, at the population level.

Comparative genomics of shrews

I assembled high-quality genomes for three shrew species: the masked shrew (*Sorex cinereus*), maritime shrew (*S. maritimensis*), and two smoky shrew genomes (*S. fumeus*). In Chapter 2, I used these resources to identify genomic variants unique to shrews and to explore whether these variants are associated with some of their distinct

traits. Analyses of accelerated regions (ARs), gene family expansions, and positively selected genes revealed outliers associated with the nervous system development, such as genes linked to axon regeneration, neurite outgrowth, synaptic plasticity, and hippocampal growth. Notably, the hippocampus is a region of the brain known to shrink during the winter and regrow in the spring in common shrews (*S. araneus*), a pattern associated with Dehnel's phenomenon (Lázaro et al., 2018). Furthermore, I also found outliers in all analyses that were either related to lipid and glucose metabolism, or cellular responses to stress and starvation. Shrews exhibit the highest mass-specific metabolic rate of any mammal and do not hibernate or rely on large fat stores to meet their energetic demands (Churchfield, 1990). Thus, adaptations related to metabolism and energy regulation are likely critical for sustaining the extreme metabolic demands of shrews, particularly during cold and resource-scarce seasons. My findings suggest that shrew-specific genomic variants have likely contributed to the physiological and neurological adaptations that enable their survival under extreme metabolic pressure and environmental stress, particularly during winter months for northern species.

Evolutionary history of Nova Scotia masked shrew populations

Geographic isolation and local adaptation are drivers of population divergence over time. In North America, repeated glacial cycles fragmented habitats and reshaped the evolutionary paths of many populations (Shafer et al., 2010). An example of this is the masked shrew population on Bon Portage Island (BPI), Nova Scotia, which exhibits distinct phenotypic differences from their mainland counterparts (MacPherson & Stewart, 2003; Stewart et al., 1989). Prolonged isolation on BPI has likely contributed

to both phenotypic and behavioral divergence, with island shrews exhibiting traits such as reduced body size and signs of accelerated aging relative to mainland conspecifics (Chapter 3). These shrews also have a specialized diet and carry a unique allele for the cytosolic non-specific dipeptidase 2 (*CNDP2*) gene, a digestive enzyme believed to have undergone adaptive changes in response to dietary pressures (MacPherson & Stewart, 2003; Stewart et al., 1989).

In Chapter 4, I explored the demographic history of Nova Scotia masked shrew populations, including BPI, tracing their divergence from a common ancestor shared with the Ontario lineage to approximately 175 kya. Since then, the data revealed a prolonged and steady decline in effective population (N_e) size across all Nova Scotia populations, with the most marked reductions observed in island populations, each now estimated to have an N_e close to 13,000 individuals. Demographic modeling suggests that BPI and Long Island (LI) became isolated approximately 16-17 kya and 12-13 kya, respectively, a timing that coincides with major climatic and geological shifts during the Last Glacial Maximum (Stokes, 2017). These shifts, particularly post-glacial sea-level rise, played a key role in fragmenting habitats and isolating populations on islands, specifically in Nova Scotia (Parker, 2012). Furthermore, the absence of detectable gene flow since their divergence provides strong evidence of long-term genetic isolation, which likely facilitated genetic drift and adaptive divergence in these island populations.

Building on this, I employed a multi-omics approach, integrating genomic, epigenomic, and transcriptomic data, to uncover varying levels of differentiation across genes and pathways between BPI masked shrews and other populations. Notably, Chapter 3 revealed epigenetic modifications, specifically differential DNA methylation

pattens near genes linked to digestive function and peptidase activity. In Chapter 4, genomic and transcriptomic analyses pointed to divergence in metabolic related genes and pathways, particularly those associated with fatty acid metabolism and metabolic regulation. This divergence across multiple biological layers, suggest that BPI shrews may have undergone functional shifts in physiology and downstream energy metabolism relating to their largely marine-based diet. Such changes are likely reflective of selection pressures imposed by their insular environment. However, direct overlap between these different omic layers was limited, suggesting that divergence in BPI shrews is a process driven by a complex genomic architecture functioning at multiple layers. This complexity shows how environmental pressures shape phenotypes through multiple mechanisms, providing a deeper understanding of local adaptation beyond genetics alone.

Collectively, my dissertation advances our understanding of the mechanisms underlying adaptation and divergence in shrews. It demonstrates that geographic isolation and distinct selection pressures leave signatures across their genomic, transcriptomic, and epigenomic landscapes. By integrating comparative genomics, population genetics, and multi-omic approaches, this work uncovers key molecular pathways that mediate responses to these evolutionary forces and contribute to the emergence of divergent and unique phenotypes.

Literature cited

Adler, G. H., & Levins, R. (1994). The Island Syndrome in Rodent Populations. *The Quarterly Review of Biology*, *69*(4), 473–490. https://doi.org/10.1086/418744

Aigner, L., Arber, S., Kapfhammer, J. P., Laux, T., Schneider, C., Botteri, F., Brenner, H.-R., & Caroni, P. (1995). Overexpression of the neural growth-associated protein GAP-43 induces nerve sprouting in the adult nervous system of transgenic mice. *Cell*, *83*(2), 269–278. https://doi.org/10.1016/0092-8674(95)90168-X

Al Aboud, N., Tupper, C., & Jialal, I. (2021). *Genetics, Epigenetic Mechanism*. StatPearls Publishing.

Albalat, A., Spence, R., Sprague, M., & Powell, A. (2025). Composition of sandhopper, Talitrus saltator and potential as a novel aquaculture feed. *Crustacean Research*, *54*(0), 9–17. https://doi.org/10.18353/crustacea.54.0_9

Alkan, C., Sajjadian, S., & Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nature Methods*, *8*(1), 61–65. https://doi.org/10.1038/nmeth.1527

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., Levy, Y., Harel, T. H., Shalev-Schlosser, G., Amsellem, Z., Razifard, H., Caicedo, A. L., Tieman, D. M., Klee, H., Kirsche, M., … Lippman, Z. B. (2020). Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell*, *182*(1), 145-161.e23. https://doi.org/10.1016/j.cell.2020.05.021

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anastasiadi, D., Venney, C. J., Bernatchez, L., & Wellenreuther, M. (2021). Epigenetic inheritance and reproductive mode in plants and animals. *Trends in Ecology and Evolution*, *36*(12), 1124–1140. https://doi.org/10.1016/j.tree.2021.08.006

Anderson, S. J., Côté, S. D., Richard, J. H., & Shafer, A. B. A. (2022). Genomic architecture of phenotypic extremes in a wild cervid. *BMC Genomics*, *23*(1). https://doi.org/10.1186/s12864-022-08333-x

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc.

Arneson, A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. Bin, Vu, H., Maciejewski, E., Yao, M., Li, C. Z., Lu, A. T., Morselli, M., Rubbi, L., Barnes, B., Hansen, K. D., Zhou, W., Breeze, C. E., Ernst, J., & Horvath, S. (2022). A mammalian methylation array for profiling methylation levels at conserved sequences. *Nature Communications*, *13*(1), 783. https://doi.org/10.1038/s41467-022-28355-z

Avila-Cervantes, J., Arias, C., Venegas-Anaya, M., Vargas, M., Larsson, H. C. E., & McMillan, W. O. (2021). Effect of the Central American Isthmus on gene flow and divergence of the American crocodile (Crocodylus acutus). *Evolution*, *75*(2), 245–259. https://doi.org/10.1111/evo.14139

Baeckens, S., & Van Damme, R. (2020). The island syndrome. *Current Biology*, *30*(8), R338–R339. https://doi.org/10.1016/j.cub.2020.03.029

Bairoch, A., & Apweiler, R. (1996). The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Research*, *24*(1), 21–25. https://doi.org/10.1093/nar/24.1.21

Bairoch, A., & Apweiler, R. (1997). The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Research*, *25*(1), 31–36. https://doi.org/10.1093/nar/25.1.31

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, *27*(12), 1691–1692. https://doi.org/10.1093/bioinformatics/btr174

Bartkowska, K., Djavadian, R. L., Taylor, J. R. E., & Turlejski, K. (2008). Generation recruitment and death of brain cells throughout the life cycle of Sorex shrews (Lipotyphla). *European Journal of Neuroscience*, *27*(7), 1710–1721. https://doi.org/10.1111/j.1460-9568.2008.06133.x

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., … Zhang, J. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Batova, O. N., Vasilieva, N. A., Titov, S. V., Savinetskaya, L. E., & Tchabovsky, A. V. (2021). Female polyandry dilutes inbreeding in a solitary fast-living hibernator. *Behavioral Ecology and Sociobiology*, *75*(10), 145. https://doi.org/10.1007/s00265-021-03086-1

Battilani, D., Gargiulo, R., Caniglia, R., Fabbri, E., Madrigal, J. R.-, Fontsere, C., Ciucani, M. M., Gopalakrishnan, S., Girardi, M., Fracasso, I., Mastroiaco, M., Ciucci, P., & Vernesi, C. (2025). Beyond population size: Whole-genome data reveal bottleneck legacies in the peninsular Italian wolf. *Journal of Heredity*, *116*(1), 10–23. https://doi.org/10.1093/jhered/esae041

Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., Christensen, B. C., Gladyshev, V. N., Heijmans, B. T., Horvath, S., Ideker, T., Issa, J. P. J., Kelsey, K. T., Marioni, R. E., Reik, W., Relton, C. L., Schalkwyk, L. C., Teschendorff, A. E., Wagner, W., … Rakyan, V. K. (2019). DNA methylation aging clocks: Challenges and recommendations. *Genome Biology*, *20*(1), 1–24. https://doi.org/10.1186/s13059-019-1824-y

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., Gilad, Y., & Pritchard, J. K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology*, *12*(1). https://doi.org/10.1186/gb-2011-12-1-r10

Bellocq, M. I., Bendell, J. F., & Innes, D. G. L. (1994). Diet of Sorex cinereus, the Masked Shrew, in Relation to the Abundance of Lepidoptera Larvae in Northern Ontario. *American Midland Naturalist*, *132*(1), 68. https://doi.org/10.2307/2426201

Belton, J.-M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi–C: A comprehensive technique to capture the conformation of genomes. *Methods*, *58*(3), 268–276. https://doi.org/10.1016/j.ymeth.2012.05.001

Benatti, P., Peluso, G., Nicolai, R., & Calvani, M. (2004). Polyunsaturated Fatty Acids: Biochemical, Nutritional and Epigenetic Properties. *Journal of the American College of Nutrition*, *23*(4), 281–302. https://doi.org/10.1080/07315724.2004.10719371

Benítez-López, A., Santini, L., Gallego-Zamorano, J., Milá, B., Walkden, P., Huijbregts, M. A. J., & Tobias, J. A. (2021). The island rule explains consistent patterns of body size evolution in terrestrial vertebrates. *Nature Ecology and Evolution*, *5*(6), 768–786. https://doi.org/10.1038/s41559-021-01426-y

Bertrand, A. R., Kadri, N. K., Flori, L., Gautier, M., & Druet, T. (2019). RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods in Ecology and Evolution*, *10*(6), 860–866. https://doi.org/10.1111/2041-210X.13167

Bi, X., Zhou, L., Zhang, J.-J., Feng, S., Hu, M., Cooper, D. N., Lin, J., Li, J., Wu, D.-D., & Zhang, G. (2023). Lineage-specific accelerated sequences underlying primate evolution. *Science Advances*, *9*(22). https://doi.org/10.1126/sciadv.adc9507

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., & Gunderson, K. L. (2009). Genome-Wide Dna Methylation Profiling Using Infinium® Assay. *Epigenomics*, *1*(1), 177–200. https://doi.org/10.2217/epi.09.14

Blanchette, M., Kent, W. J., Riemer, C., Elnitski, L., Smit, A. F. A., Roskin, K. M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E. D., Haussler, D., & Miller, W. (2004). Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. *Genome Research*, *14*(4), 708–715. https://doi.org/10.1101/gr.1933104

Blois, J. L., McGuire, J. L., & Hadly, E. A. (2010). Small mammal diversity loss in response to late-Pleistocene climatic change. *Nature*, *465*(7299), 771–774. https://doi.org/10.1038/nature09077

Bogan, S. N., & Yi, S. V. (2024). Potential Role of DNA Methylation as a Driver of Plastic Responses to the Environment Across Cells, Organisms, and Populations. *Genome Biology and Evolution*, *16*(2). https://doi.org/10.1093/gbe/evae022

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

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*(1), e1002592. https://doi.org/10.1371/journal.pbio.1002592

Borodin, P. M., Karamysheva, T. V, Belonogova, N. M., Torgasheva, A. A., Rubtsov, N. B., & Searle, J. B. (2008). Recombination Map of the Common Shrew, Sorex araneus (Eulipotyphla, Mammalia). *Genetics*, *178*(2), 621–632. https://doi.org/10.1534/genetics.107.079665

Bourgeois, Y. X. C., & Warren, B. H. (2021). An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Molecular Ecology*, *30*(23), 6036–6071. https://doi.org/10.1111/mec.15989

Broad Institute, G. R. (2019). *Picard toolkit*. Http://Broadinstitute.Github.Io/Picard/.

Brown, R. P., Paterson, S., & Risse, J. (2016). Genomic Signatures of Historical Allopatry and Ecological Divergence in an Island Lizard. *Genome Biology and Evolution*, *8*(11), 3618–3626. https://doi.org/10.1093/gbe/evw268

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Buchler, E. R. (1976). The use of echolocation by the wandering shrew (Sorex vagrans). *Animal Behaviour*, *24*(4), 858–873. https://doi.org/10.1016/S0003-3472(76)80016-4

Burdge, G. C. (2018). Polyunsaturated Fatty Acid Biosynthesis and Metabolism in Adult Mammals. In *Polyunsaturated Fatty Acid Metabolism* (pp. 15–30). Elsevier. https://doi.org/10.1016/B978-0-12-811230-4.00002-8

Burgin, C. J., Colella, J. P., Kahn, P. L., & Upham, N. S. (2018). How many species of mammals are there? *Journal of Mammalogy*, *99*(1), 1–14. https://doi.org/10.1093/jmammal/gyx147

Bushnell, B. (2022). *BBTools*. https://sourceforge.net/projects/bbmap/

Byer, N. W., Holding, M. L., Crowell, M. M., Pierson, T. W., Dilts, T. E., Larrucea, E. S., Shoemaker, K. T., & Matocq, M. D. (2021). Adaptive divergence despite low effective population size in a peripherally isolated population of the pygmy rabbit, Brachylagus idahoensis. *Molecular Ecology*, *30*(17), 4173–4188. https://doi.org/10.1111/mec.16040

Cai, S., Shen, Q., Huang, Y., Han, Z., Wu, D., Chen, Z., Nevo, E., & Zhang, G. (2021). Multi-Omics Analysis Reveals the Mechanism Underlying the Edaphic Adaptation in Wild Barley at Evolution Slope (Tabigha). *Advanced Science*, *8*(20). https://doi.org/10.1002/advs.202101374

Calder, P. C. (2015). Functional Roles of Fatty Acids and Their Effects on Human Health. *Journal of Parenteral and Enteral Nutrition*, *39*(1S). https://doi.org/10.1177/0148607115595980

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Cao, J., Wei, C., Liu, D., Wang, H., Wu, M., Xie, Z., Capellini, T. D., Zhang, L., Zhao, F., Li, L., Zhong, T., Wang, L., Lu, J., Liu, R., Zhang, S., Du, Y., Zhang, H., & Du, L. (2015). DNA methylation Landscape of body size variation in sheep. *Scientific Reports*, *5*. https://doi.org/10.1038/srep13950

Catania, K. C., Hare, J. F., & Campbell, K. L. (2008). Water shrews detect movement, shape, and smell to find prey underwater. *Proceedings of the National Academy of Sciences*, *105*(2), 571–576. https://doi.org/10.1073/pnas.0709534104

Cervantes, F. A., Arcangeli, J., & Alonso Morales, R. A. (2013). The SRY gene of the y chromosome successfully identifies gender of endemic Mexican shrews (Mammalia: Soricomorpha). *Mammalia*, *77*(3), 335–339. https://doi.org/10.1515/mammalia-2011-0123

Chai, S., Tian, R., Rong, X., Li, G., Chen, B., Ren, W., Xu, S., & Yang, G. (2020). Evidence of echolocation in the common shrew from molecular convergence with other echolocating mammals. *Zoological Studies*, *59*, 1–11. https://doi.org/10.6620/ZS.2020.59-4

Chakraborty, M., Baldwin-Brown, J. G., Long, A. D., & Emerson, J. J. (2016). Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Research*, *44*(19), 1–12. https://doi.org/10.1093/nar/gkw654

Chapman, J. A., Ho, I., Sunkara, S., Luo, S., Schroth, G. P., & Rokhsar, D. S. (2011). Meraculous: De novo genome assembly with short paired-end reads. *PLoS ONE*, *6*(8). https://doi.org/10.1371/journal.pone.0023501

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, Y., Gao, L., Lin, T., Pei, X., Gao, Q., Chen, J., Zhang, Y., Wu, X., Li, Z., & Zhang, Z. (2022). C/EBPZ modulates the differentiation and proliferation of preadipocytes. *International Journal of Obesity*, *46*(3), 523–534. https://doi.org/10.1038/s41366-021-01020-z

Chen, Y., Hu, S., Liu, M., Zhao, B., Yang, N., Li, J., Chen, Q., Zhou, J., Bao, G., & Wu, X. (2021). Analysis of Genome DNA Methylation at Inherited Coat Color Dilutions of Rex Rabbits. *Frontiers in Genetics*, *11*(January), 1–9. https://doi.org/10.3389/fgene.2020.603528

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170–175. https://doi.org/10.1038/s41592-020-01056-5

Cheng, H., Jarvis, E. D., Fedrigo, O., Koepfli, K.-P., Urban, L., Gemmell, N. J., & Li, H. (2022). Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology*, *40*(9), 1332–1335. https://doi.org/10.1038/s41587-022-01261-x

Cheviron, Z. A., Bachman, G. C., Connaty, A. D., McClelland, G. B., & Storz, J. F. (2012). Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice. *Proceedings of the National Academy of Sciences*, *109*(22), 8635–8640. https://doi.org/10.1073/pnas.1120523109

Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, *30*(1), 31–37. https://doi.org/10.1093/bioinformatics/btt310

Choi, W. il, Yoon, J. H., Song, J. Y., Jeon, B. N., Park, J. M., Koh, D. I., Ahn, Y. ho, Kim, K. S., Lee, I. K., & Hur, M. W. (2019). Zbtb7c is a critical gluconeogenic transcription factor that induces glucose-6-phosphatase and phosphoenylpyruvate carboxykinase 1 genes expression during mice fasting. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1862*(6), 643–656. https://doi.org/10.1016/j.bbagrm.2019.04.001

Christmas, M. J., Kaplow, I. M., Genereux, D. P., Dong, M. X., Hughes, G. M., Li, X., Sullivan, P. F., Hindle, A. G., Andrews, G., Armstrong, J. C., Bianchi, M., Breit, A. M., Diekhans, M., Fanter, C., Foley, N. M., Goodman, D. B., Goodman, L., Keough, K. C., Kirilenko, B., … Zhang, X. (2023). Evolutionary constraint and innovation across hundreds of placental mammals. *Science*, *380*(6643). https://doi.org/10.1126/science.abn3943

Chrysanthou, A., Ververis, A., & Christodoulou, K. (2022). ANO10 Function in Health and Disease. *The Cerebellum*, *22*(3), 447–467. https://doi.org/10.1007/s12311-022-01395-3
Churchfield, S. (1990). *The Natural History of Shrews*. Cornell University Press.
Churchfield, S. (2002). Why are shrews so small? The costs and benefits of small size in northern temperate Sorex species in the context of foraging habits and prey supply. *Acta Theriologica*, *47*(S1), 169–184. https://doi.org/10.1007/BF03192486
Clark, M. S., Hoffman, J. I., Peck, L. S., Bargelloni, L., Gande, D., Havermans, C., Meyer, B., Patarnello, T., Phillips, T., Stoof-Leichsenring, K. R., Vendrami, D. L. J., Beck, A., Collins, G., Friedrich, M. W., Halanych, K. M., Masello, J. F., Nagel, R., Norén, K., Printzen, C., … Mock, T. (2023). Multi-omics for studying and understanding polar life. *Nature Communications*, *14*(1), 7451. https://doi.org/10.1038/s41467-023-43209-y
Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L., & Frommer, M. (2006). DNA methylation: Bisulphite modification and analysis. *Nature Protocols*, *1*(5), 2353–2364. https://doi.org/10.1038/nprot.2006.324
Clavijo, B. J. (2021). *w2rap-contigger*. https://github.com/bioinfologics/w2rap-contigger
Cooper, N., Kamilar, J. M., & Nunn, C. L. (2012). Host longevity and parasite species richness in mammals. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0042190
Cossette, M., Stewart, D. T., Haghani, A., Zoller, J. A., Shafer, A. B. A., & Horvath, S. (2023). Epigenetics and island-mainland divergence in an insectivorous small mammal. *Molecular Ecology*, *32*(1), 152–166. https://doi.org/10.1111/mec.16735
Cote, I. M., & Poulinb, R. (1995). Parasitism and group size in social animals: a meta-analysis. *Behavioral Ecology*, *6*(2), 159–165. https://doi.org/10.1093/beheco/6.2.159
Cowan, K. M., Shutler, D., Herman, T. B., & Stewart, D. T. (2009). Splenic mass of masked shrews, sorex cinereus, in relation to body mass, sex, age, day of the year, and bladder nematode, liniscus (=capillaria) maseri, infection. *Journal of Parasitology*, *95*(1), 228–230. https://doi.org/10.1645/GE-1566.1
Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0
Dalton, A. S., Margold, M., Stokes, C. R., Tarasov, L., Dyke, A. S., Adams, R. S., Allard, S., Arends, H. E., Atkinson, N., Attig, J. W., Barnett, P. J., Barnett, R. L., Batterson, M., Bernatchez, P., Borns, H. W., Breckenridge, A., Briner, J. P., Brouard, E., Campbell, J. E., … Wright, H. E. (2020). An updated radiocarbon-based ice margin chronology for the last deglaciation of the North American Ice Sheet Complex. *Quaternary Science Reviews*, *234*, 106223. https://doi.org/10.1016/j.quascirev.2020.106223
Dalton, A. S., Stokes, C. R., & Batchelor, C. L. (2022). Evolution of the Laurentide and Innuitian ice sheets prior to the Last Glacial Maximum (115 ka to 25 ka). *Earth-Science Reviews*, *224*, 103875. https://doi.org/10.1016/j.earscirev.2021.103875

Damas, J., Corbo, M., Kim, J., Turner-Maier, J., Farré, M., Larkin, D. M., Ryder, O. A., Steiner, C., Houck, M. L., Hall, S., Shiue, L., Thomas, S., Swale, T., Daly, M., Korlach, J., Uliano-Silva, M., Mazzoni, C. J., Birren, B. W., Genereux, D. P., … Zhang, X. (2022). Evolution of the ancestral mammalian karyotype and syntenic regions. *Proceedings of the National Academy of Sciences*, *119*(40). https://doi.org/10.1073/pnas.2209139119

Damas, J., Corbo, M., & Lewin, H. A. (2021). Vertebrate Chromosome Evolution. *Annual Review of Animal Biosciences*, *9*(1), 1–27. https://doi.org/10.1146/annurev-animal-020518-114924

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLoS Computational Biology*, *6*(12), e1001025. https://doi.org/10.1371/journal.pcbi.1001025

Dawe, K. L., Shafer, A. B. A., Herman, T. B., & Stewart, D. T. (2009). Diffusion of nuclear and mitochondrial genes across a zone of secondary contact in the maritime shrew, Sorex maritimensis: implications for the conservation of a Canadian endemic mammal. *Conservation Genetics*, *10*(4), 851–857. https://doi.org/10.1007/s10592-008-9645-7

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, *22*(10), 1269–1271. https://doi.org/10.1093/bioinformatics/btl097

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes and Development*, *25*(10), 1010–1022. https://doi.org/10.1101/gad.2037511

DeBiasse, M. B., & Kelly, M. W. (2016). Plastic and Evolved Responses to Global Change: What Can We Learn from Comparative Transcriptomics?: Table 1. *Journal of Heredity*, *107*(1), 71–81. https://doi.org/10.1093/jhered/esv073

Denton, J. F., Lugo-Martinez, J., Tucker, A. E., Schrider, D. R., Warren, W. C., & Hahn, M. W. (2014). Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLoS Computational Biology*, *10*(12), e1003998. https://doi.org/10.1371/journal.pcbi.1003998

Dhar, G. A., Saha, S., Mitra, P., & Nag Chaudhuri, R. (2021). DNA methylation and regulation of gene expression: Guardian of our health. In *Nucleus (India)* (Vol. 64, Issue 3, pp. 259–270). Springer. https://doi.org/10.1007/s13237-021-00367-y

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dong, Z., & Chen, Y. (2013). Transcriptomics: Advances and approaches. *Science China Life Sciences*, *56*(10), 960–967. https://doi.org/10.1007/s11427-013-4557-2

Downie, A. B. (1986). *Spatial distribution, movements and age structure in island and mainland populations of the masked shrew (Sorex cinereus Kerr)* [B.Sc. Honours thesis]. Acadia University, Wolfville, N.S.

Duarte, L. C., Bouteiller, C., Fontanillas, P., Petit, E., & Perrin, N. (2003). Inbreeding in the greater white-toothed shrew, Crocidura russula. *Evolution*, *57*(3), 638–645. https://doi.org/10.1111/j.0014-3820.2003.tb01555.x

Dube, S., Errazuriz, I., Cobelli, C., Basu, R., & Basu, A. (2013). Assessment of insulin action on carbohydrate metabolism: physiological and non-physiological methods. *Diabetic Medicine*, *30*(6), 664–670. https://doi.org/10.1111/dme.12189

Dzięgiel, P., Owczarek, T., Plażuk, E., Gomułkiewicz, A., Majchrzak, M., Podhorska-Okołów, M., Driouch, K., Lidereau, R., & Ugorski, M. (2010). Ceramide galactosyltransferase (UGT8) is a molecular marker of breast cancer malignancy and lung metastases. *British Journal of Cancer*, *103*(4), 524–531. https://doi.org/10.1038/sj.bjc.6605750

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Erecińska, M., & Silver, I. A. (1989). ATP and Brain Function. *Journal of Cerebral Blood Flow & Metabolism*, *9*(1), 2–19. https://doi.org/10.1038/jcbfm.1989.2

Esteva, M., Cervantes, F. A., Brant, S. V., & Cook, J. A. (2010). Molecular phylogeny of long-tailed shrews (genus Sorex) from México and Guatemala. *Zootaxa*, *2615*(1). https://doi.org/10.11646/zootaxa.2615.1.3

Etherington, G. J., Heavens, D., Baker, D., Lister, A., McNelly, R., Garcia, G., Clavijo, B., Macaulay, I., Haerty, W., & di Palma, F. (2020). Sequencing smart: De novo sequencing and assembly approaches for a non-model mammal. *GigaScience*, *9*(5). https://doi.org/10.1093/gigascience/giaa045

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*(3), 276–278. https://doi.org/10.1038/s41587-020-0439-x

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fair, B. J., Blake, L. E., Sarkar, A., Pavlovic, B. J., Cuevas, C., & Gilad, Y. (2020). Gene expression variability in human and chimpanzee populations share common determinants. *ELife*, *9*. https://doi.org/10.7554/eLife.59929

Farhang-Fallah, J., Randhawa, V. K., Nimnual, A., Klip, A., Bar-Sagi, D., & Rozakis-Adcock, M. (2002). The Pleckstrin Homology (PH) Domain-Interacting Protein Couples the Insulin Receptor Substrate 1 PH Domain to Insulin Signaling Pathways Leading to Mitogenesis and GLUT4 Translocation. *Molecular and Cellular Biology*, *22*(20), 7325–7336. https://doi.org/10.1128/mcb.22.20.7325-7336.2002

Ferris, E., Abegglen, L. M., Schiffman, J. D., & Gregg, C. (2018). Accelerated Evolution in Distinctive Species Reveals Candidate Elements for Clinically Relevant Traits, Including Mutation and Cancer Resistance. *Cell Reports*, *22*(10), 2742–2755. https://doi.org/10.1016/j.celrep.2018.02.008

Forsman, K. A., & Malmquist, M. G. (1988). Evidence for echolocation in the common shrew, Sorex araneus. *Journal of Zoology*, *216*(4), 655–662. https://doi.org/10.1111/j.1469-7998.1988.tb02463.x

Foster, J. B. (1964). Evolution of Mammals on Islands. *Nature*, *202*, 234–235. https://doi.org/10.1038/202234a0

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2010). Genetic consequences of small population sizes. In *Introduction to Conservation Genetics* (pp. 161–181). Cambridge University Press.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Friedman, J. M., & Halaas, J. L. (1998). Leptin and the regulation of body weight in mammals. *Nature*, *395*(6704), 763–770. https://doi.org/10.1038/27376

Fu, J., & Wen, L. (2023). Impacts of Quaternary glaciation, geological history and geography on animal species history in continental East Asia: A phylogeographic review. *Molecular Ecology*, *32*(16), 4497–4514. https://doi.org/10.1111/mec.17053

Fumagalli, L., Taberlet, P., Stewart, D. T., Gielly, L., Hausser, J., & Vogel, P. (1999). Molecular phylogeny and evolution of Sorex shrews (Soricidae: insectivora) inferred from mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, *11*(2), 222–235. https://doi.org/10.1006/mpev.1998.0568

Gaidatzis, D., Lerch, A., Hahne, F., & Stadler, M. B. (2015). QuasR: quantification and annotation of short reads in R. *Bioinformatics*, *31*(7), 1130–1132. https://doi.org/10.1093/bioinformatics/btu781

George, S. B. (1986). Evolution and Historical Biogeography of Soricine Shrews. *Systematic Zoology*, *35*(2), 153–162. https://doi.org/10.2307/2413427

George, S. B. (1988). Systematics, Historical Biogeography, and Evolution of the Genus Sorex. *Journal of Mammalogy*, *69*(3), 443–461. https://doi.org/10.2307/1381337

Girskis, K. M., Stergachis, A. B., DeGennaro, E. M., Doan, R. N., Qian, X., Johnson, M. B., Wang, P. P., Sejourne, G. M., Nagy, M. A., Pollina, E. A., Sousa, A. M. M., Shin, T., Kenny, C. J., Scotellaro, J. L., Debo, B. M., Gonzalez, D. M., Rento, L. M., Yeh, R. C., Song, J. H. T., … Walsh, C. A. (2021). Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron*, *109*(20), 3239-3251.e7. https://doi.org/10.1016/j.neuron.2021.08.005

Gleason, L. U., Fekete, F. J., Tanner, R. L., & Dowd, W. W. (2023). Multi-omics reveals largely distinct transcript- and protein-level responses to the environment in an intertidal mussel. *Journal of Experimental Biology*, *226*(22). https://doi.org/10.1242/jeb.245962

Gleason, M. E., Eddington, V. M., & Kloepper, L. N. (2023). Acoustic behavior in the northern short-tailed shrew (Blarina brevicauda): Ultrasonic click production in a

novel environment. *The Journal of the Acoustical Society of America*, *154*(1), 411–417. https://doi.org/10.1121/10.0020071

Goltsman, M., Kruchenkova, E. P., Sergeev, S., Volodin, I., & Macdonald, D. W. (2005). “Island syndrome” in a population of Arctic foxes (Alopex lagopus) from Mednyi Island. *Journal of Zoology*, *267*(4), 405–418. https://doi.org/10.1017/S0952836905007557

Gonda, Y., Namba, T., & Hanashima, C. (2020). Beyond Axon Guidance: Roles of Slit-Robo Signaling in Neocortical Formation. *Frontiers in Cell and Developmental Biology*, *8*. https://doi.org/10.3389/fcell.2020.607415

Gould, E. (1969). Communication in three genera of shrews (Soricidae): Suncus, Blarina, and Cryptotis . *Communications in Behavioral Biology*, *3*, 11–31.

Gu, T.-T., Wu, H., Yang, F., Gaubert, P., Heighton, S. P., Fu, Y., Liu, K., Luo, S.-J., Zhang, H.-R., Hu, J.-Y., & Yu, L. (2023). Genomic analysis reveals a cryptic pangolin species. *Proceedings of the National Academy of Sciences*, *120*(40). https://doi.org/10.1073/pnas.2304096120

Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, *26*(20), 5362–5368. https://doi.org/10.1111/mec.14290

Guizard, S., Miedzinska, K., Smith, J., Smith, J., Kuo, R. I., Davey, M., Archibald, A., & Watson, M. (2023). nf-core/isoseq: simple gene and isoform annotation with PacBio Iso-Seq long-read sequencing. *Bioinformatics*, *39*(5). https://doi.org/10.1093/bioinformatics/btad150

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Haghani, A., Lu, A., Li, C., Robeck, T., Belov, K., Breeze, C., Brooke, R., Clarke, S., Faulkes, C., Fei, Z., Ferguson, S., Finno, C., Gladyshev, V., Gorbunova, V., Goya, R., Hogan, A., Hogg, C., Hore, T., Kiaris, H., … Horvath, S. (2021). DNA Methylation Networks Underlying Mammalian Traits. *BioRxiv*. https://doi.org/10.1101/2021.03.16.435708

Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches. *Genome Research*, *27*(6), 1004–1015. https://doi.org/10.1101/gr.212522.116

Hansen, C. C. R., Láruson, Á. J., Rasmussen, J. A., Ballesteros, J. A. C., Sinding, M. S., Hallgrimsson, G. T., von Schmalensee, M., Stefansson, R. A., Skarphédinsson, K. H., Labansen, A. L., Leivits, M., Sonne, C., Dietz, R., Skelmose, K., Boertmann, D., Eulaers, I., Martin, M. D., Helgason, A. S., Gilbert, M. T. P., & Pálsson, S. (2023). Genomic diversity and differentiation between island and mainland populations of white-tailed eagles (Haliaeetus albicilla). *Molecular Ecology*, *32*(8), 1925–1942. https://doi.org/10.1111/mec.16858

Harris, R. S. (2007). *Improved pairwise alignment of genomic DNA*. The Pennsylvania State University.

Harwood, J. (2019). Algae: Critical Sources of Very Long-Chain Polyunsaturated Fatty Acids. *Biomolecules*, *9*(11), 708. https://doi.org/10.3390/biom9110708

Helsen, J., Voordeckers, K., Vanderwaeren, L., Santermans, T., Tsontaki, M., Verstrepen, K. J., & Jelier, R. (2020). Gene Loss Predictably Drives Evolutionary Adaptation. *Molecular Biology and Evolution*, *37*(10), 2989–3002. https://doi.org/10.1093/molbev/msaa172

Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G., & Gravel, S. (2015). Estimating the mutation load in human genomes. *Nature Reviews Genetics*, *16*(6), 333–343. https://doi.org/10.1038/nrg3931

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). *Whole-Genome Annotation with BRAKER* (pp. 65–95). https://doi.org/10.1007/978-1-4939-9173-0_5

Holder, J. L., Zhang, L., Kublaoui, B. M., Dileone, R. J., Oz, O. K., Bair, C. H., Lee, Y.-H., Zinn, A. R., Lloyd, J., & Zinn, R. (2004). Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice. *Am J Physiol Endocrinol Metab*, *287*, 105–113. https://doi.org/10.1152/ajpendo.00446.2003.-Haploinsufficiency

Hope, A. G., Takebayashi, N., Galbreath, K. E., Talbot, S. L., & Cook, J. A. (2013). Temporal, spatial and ecological dynamics of speciation among amphi-Beringian small mammals. *Journal of Biogeography*, *40*(3), 415–429. https://doi.org/10.1111/jbi.12056

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Howell, E. K., Nolfo-Clements, L. E., & Payseur, B. A. (2025). Population History Across Timescales in an Urban Archipelago. *Genome Biology and Evolution*, *17*(4). https://doi.org/10.1093/gbe/evaf048

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, Y., Feulner, P. G. D., Eizaguirre, C., Lenz, T. L., Bornberg-Bauer, E., Milinski, M., Reusch, T. B. H., & Chain, F. J. J. (2019). Genome-Wide Genotype-Expression Relationships Reveal Both Copy Number and Single Nucleotide Differentiation Contribute to Differential Gene Expression between Stickleback Ecotypes. *Genome Biology and Evolution*, *11*(8), 2344–2359. https://doi.org/10.1093/gbe/evz148

Hubisz, M. J., & Pollard, K. S. (2014). Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Current Opinion in Genetics & Development*, *29*, 15–21. https://doi.org/10.1016/j.gde.2014.07.005

Hubisz, M. J., Pollard, K. S., & Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings in Bioinformatics*, *12*(1), 41–51. https://doi.org/10.1093/bib/bbq072

Huggins, J. A., & Kennedy, M. L. (1989). Morphologic Variation in the Masked Shrew (Sorex cinereus) and the Smoky Shrew (S. fumeus). *American Midland Naturalist*, *122*(1), 11–25. https://doi.org/10.2307/2425678

Humann, J. L., Lee, T., Ficklin, S., & Main, D. (2019). *Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. In: Kollmar M. (eds) Gene*

Prediction. Methods in Molecular Biology. *1962*, 29–51. https://doi.org/10.1007/978-1-4939-9173-0
Hunter, P. (2020). The rise of the mammals. *EMBO Reports*, *21*(11). https://doi.org/10.15252/embr.202051617
Hurles, M. E., Dermitzakis, E. T., & Tyler-Smith, C. (2008). The functional impact of structural variation in humans. *Trends in Genetics*, *24*(5), 238–245. https://doi.org/10.1016/j.tig.2008.03.001
Hutterer, R. (2005). Order Soricomorpha. In Wilson D. E. & Reeder D. M. (Eds.), *Mammal species of the world: a taxonomic and geographic reference* (Third Edition, pp. 220–311). The Johns Hopkins University Press.
Iguchi-Ariga, S. M., & Schaffner, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes & Development*, *3*(5), 612–619. https://doi.org/10.1101/gad.3.5.612
Illingworth, R. S., & Bird, A. P. (2009). CpG islands - "A rough guide." *FEBS Letters*, *583*(11), 1713–1720. https://doi.org/10.1016/j.febslet.2009.04.012
Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome. *PLoS Genetics*, *6*(9). https://doi.org/10.1371/journal.pgen.1001134
Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Grossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, *27*(23), 4839–4855. https://doi.org/10.1111/mec.14858
Jacquet, S., Culbertson, M., Zhang, C., El Filali, A., De La Myre Mory, C., Pons, J.-B., Filippi-Codaccioni, O., Lauterbur, M. E., Ngoubangoye, B., Duhayer, J., Verez, C., Park, C., Dahoui, C., Carey, C. M., Brennan, G., Enard, D., Cimarelli, A., Rothenburg, S., Elde, N. C., … Etienne, L. (2022). Adaptive duplication and genetic diversification of protein kinase R contribute to the specificity of bat-virus interactions. *Science Advances*, *8*(47). https://doi.org/10.1126/sciadv.add7540
Jarva, T. (2023). *parse_all_absREL_json*. https://github.com/stupornova33/parse_absREL_json/blob/main/parse_all_absREL_json
Jasinska, A. J., Haghani, A., Zoller, J. A., Li, C. Z., Arneson, A., Ernst, J., Kavanagh, K., Jorgensen, M. J., Mattison, J. A., Wojta, K., Choi, O.-W., DeYoung, J., Li, X., Rao, A. W., Coppola, G., Freimer, N. B., Woods, R. P., & Horvath, S. (2021). Epigenetic clock and methylation studies in vervet monkeys. *GeroScience*, *0123456789*. https://doi.org/10.1007/s11357-021-00466-3
Jezierski, M. T., Smith, W. J., & Clegg, S. M. (2024). The island syndrome in birds. *Journal of Biogeography*, *51*(9), 1607–1622. https://doi.org/10.1111/jbi.14720
Ji, Y., Shi, Y., Ding, G., & Li, Y. (2011). A new strategy for better genome assembly from very short reads. *BMC Bioinformatics*, *12*(1), 493. https://doi.org/10.1186/1471-2105-12-493
Jiménez-Prada, P., Hachero-Cruzado, I., Giráldez, I., Fernández-Diaz, C., Vilas, C., Cañavate, J. P., & Guerra-García, J. M. (2018). Crustacean amphipods from

marsh ponds: a nutritious feed resource with potential for application in Integrated Multi-Trophic Aquaculture. *PeerJ*, *6*, e4194. https://doi.org/10.7717/peerj.4194

Johnson, R. C., Nelson, G. W., Troyer, J. L., Lautenberger, J. A., Kessing, B. D., Winkler, C. A., & O'Brien, S. J. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, *11*(1). https://doi.org/10.1186/1471-2164-11-724

Jones, P. A. (1999). The DNA methylation paradox. *Trends in Genetics*, *15*(1), 34–37. https://doi.org/10.1016/S0168-9525(98)01636-9

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Juette, T., Garant, D., Jameson, J. W., & Réale, D. (2020). The island syndrome hypothesis is only partially validated in two rodent species in an inland–island system. *Oikos*, *129*(11), 1739–1751. https://doi.org/10.1111/oik.07249

Kamm, J., Terhorst, J., Durbin, R., & Song, Y. S. (2020). Efficiently Inferring the Demographic History of Many Populations With Allele Count Data. *Journal of the American Statistical Association*, *115*(531), 1472–1487. https://doi.org/10.1080/01621459.2019.1635482

Karbiener, M., Glantschnig, C., Pisani, D. F., Laurencikiene, J., Dahlman, I., Herzig, S., Amri, E. Z., & Scheideler, M. (2015). Mesoderm-specific transcript (MEST) is a negative regulator of human adipocyte differentiation. *International Journal of Obesity*, *39*(12), 1733–1741. https://doi.org/10.1038/ijo.2015.121

Keightley, P. D., & Jackson, B. C. (2018). Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site. *Genetics*, *209*(3), 897–906. https://doi.org/10.1534/genetics.118.301120

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology and Evolution*, *17*(5), 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, and D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Kessler, C., & Shafer, A. B. A. (2024). Genomic Analyses Capture the Human-Induced Demographic Collapse and Recovery in a Wide-Ranging Cervid. *Molecular Biology and Evolution*, *41*(3). https://doi.org/10.1093/molbev/msae038

Khalili, K., Del Valle, L., Muralidharan, V., Gault, W. J., Darbinian, N., Otte, J., Meier, E., Johnson, E. M., Daniel, D. C., Kinoshita, Y., Amini, S., & Gordon, J. (2003). Purα Is Essential for Postnatal Brain Development and Developmentally Coupled Cellular Proliferation As Revealed by Genetic Inactivation in the Mouse. *Molecular and Cellular Biology*, *23*(19), 6857–6875. https://doi.org/10.1128/MCB.23.19.6857-6875.2003

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kimura, H., Fujita, Y., Kawabata, T., Ishizuka, K., Wang, C., Iwayama, Y., Okahisa, Y., Kushima, I., Morikawa, M., Uno, Y., Okada, T., Ikeda, M., Inada, T., Branko, A., Mori, D., Yoshikawa, T., Iwata, N., Nakamura, H., Yamashita, T., & Ozaki, N. (2017). A novel rare variant R292H in RTN4R affects growth cone formation and possibly contributes to schizophrenia susceptibility. *Translational Psychiatry*, *7*(8), e1214–e1214. https://doi.org/10.1038/tp.2017.170

Kita, M., Nakamura, Y., Okumura, Y., Ohdachi, S. D., Oba, Y., Yoshikuni, M., Kido, H., & Uemura, D. (2004). Blarina toxin, a mammalian lethal venom from the short-tailed shrew Blarina brevicauda: Isolation and characterization. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(20), 7542–7547. https://doi.org/10.1073/pnas.0402517101

Klok, M. D., Jakobsdottir, S., & Drent, M. L. (2007). The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. *Obesity Reviews*, *8*(1), 21–34. https://doi.org/10.1111/j.1467-789X.2006.00270.x

Korbel, J. O., & Lee, C. (2013). Genome assembly and haplotyping with Hi-C. *Nature Biotechnology*, *31*(12), 1099–1101. https://doi.org/10.1038/nbt.2764

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Korneliussen, T. S., Moltke, I., Albrechtsen, A., & Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics*, *14*(1), 289. https://doi.org/10.1186/1471-2105-14-289

Kuo, R. I., Cheng, Y., Zhang, R., Brown, J. W. S., Smith, J., Archibald, A. L., & Burt, D. W. (2020). Illuminating the dark side of the human transcriptome with long read transcript sequencing. *BMC Genomics*, *21*(1), 751. https://doi.org/10.1186/s12864-020-07123-7

Lai, D., Qu, X., Wang, Z., He, S., Li, X., Bao, Q., Sun, G., Zhang, J., Zhu, Y., & Yi, G. (2025). Integrative Genomic, Transcriptomic and Epigenomic Analysis Reveals cis-regulatory Contributions to High-altitude Adaptation in Tibetan Pigs. *Molecular Biology and Evolution*, *42*(7). https://doi.org/10.1093/molbev/msaf169

Larison, B., Pinho, G. M., Haghani, A., Zoller, J. A., Li, C. Z., Finno, C. J., Farrell, C., Kaelin, C. B., Barsh, G. S., Wooding, B., Robeck, T. R., Maddox, D., Pellegrini, M., & Horvath, S. (2021a). Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids. *Communications Biology*, *4*(1), 1412. https://doi.org/10.1038/s42003-021-02935-z

Larison, B., Pinho, G. M., Haghani, A., Zoller, J. A., Li, C. Z., Finno, C. J., Farrell, C., Kaelin, C. B., Barsh, G. S., Wooding, B., Robeck, T. R., Maddox, D., Pellegrini, M., & Horvath, S. (2021b). Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids. *Communications Biology*, *4*(1). https://doi.org/10.1038/s42003-021-02935-z

Larivière, D., Abueg, L., Brajuka, N., Gallardo-Alba, C., Grüning, B., Ko, B. J., Ostrovsky, A., Palmada-Flores, M., Pickett, B. D., Rabbani, K., Antunes, A., Balacco, J. R., Chaisson, M. J. P., Cheng, H., Collins, J., Couture, M., Denisova, A., Fedrigo, O., Gallo, G. R., … Formenti, G. (2024). Scalable, accessible and reproducible reference genome assembly and evaluation in Galaxy. *Nature Biotechnology*, *42*(3), 367–370. https://doi.org/10.1038/s41587-023-02100-3

Latremoliere, A., Cheng, L., DeLisle, M., Wu, C., Chew, S., Hutchinson, E. B., Sheridan, A., Alexandre, C., Latremoliere, F., Sheu, S.-H., Golidy, S., Omura, T., Huebner, E. A., Fan, Y., Whitman, M. C., Nguyen, E., Hermawan, C., Pierpaoli, C., Tischfield, M. A., … Engle, E. C. (2018). Neuronal-Specific TUBB3 Is Not Required for Normal Neuronal Function but Is Essential for Timely Axon Regeneration. *Cell Reports*, *24*(7), 1865-1879.e9. https://doi.org/10.1016/j.celrep.2018.07.029

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, *9*(8). https://doi.org/10.1371/journal.pcbi.1003118

Lázaro, J., & Dechmann, D. K. N. (2021). Dehnel's phenomenon. *Current Biology*, *31*(10), R463–R465. https://doi.org/10.1016/j.cub.2021.04.006

Lázaro, J., Dechmann, D. K. N., LaPoint, S., Wikelski, M., & Hertel, M. (2017). Profound reversible seasonal changes of individual skull size in a mammal. *Current Biology*, *27*(20), R1106–R1107. https://doi.org/10.1016/j.cub.2017.08.055

Lázaro, J., Hertel, M., LaPoint, S., Wikelski, M., Stiehler, M., & Dechmann, D. K. N. (2017). Cognitive skills of common shrews (Sorex araneus) vary with seasonal changes in skull size and brain mass. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.166595

Lázaro, J., Hertel, M., Muturi, M., & Dechmann, D. K. N. (2019). Seasonal reversible size changes in the braincase and mass of common shrews are flexibly modified by environmental conditions. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-38884-1

Lázaro, J., Hertel, M., Sherwood, C. C., Muturi, M., & Dechmann, D. K. N. (2018). Profound seasonal changes in brain size and architecture in the common shrew. *Brain Structure and Function*, *223*(6), 2823–2840. https://doi.org/10.1007/s00429-018-1666-5

Lázaro, J., Nováková, L., Hertel, M., Taylor, J. R. E., Muturi, M., Zub, K., & Dechmann, D. K. N. (2021). Geographic patterns in seasonal changes of body mass, skull, and brain size of common shrews. In *Ecology and Evolution* (Vol. 11, Issue 6, pp. 2431–2448). John Wiley and Sons Ltd. https://doi.org/10.1002/ece3.7238

Lee, H. S., & Park, T. (2020). The influences of DNA methylation and epigenetic clocks, on metabolic disease, in middle-aged Koreans. *Clinical Epigenetics*, *12*(1). https://doi.org/10.1186/s13148-020-00936-z

Lee, Y. J., Jeong, Y. J., Kang, E. J., Kang, B. S., Lee, S. H., Kim, Y. J., Kang, S. S., Suh, S. W., & Ahn, E. H. (2023). GAP-43 closely interacts with BDNF in hippocampal neurons and is associated with Alzheimer's disease progression. *Frontiers in Molecular Neuroscience*, *16*. https://doi.org/10.3389/fnmol.2023.1150399

Levchenko, A., Kanapin, A., Samsonova, A., & Gainetdinov, R. R. (2018). Human Accelerated Regions and Other Human-Specific Sequence Variations in the

Context of Evolution and Their Relevance for Brain Development. *Genome Biology and Evolution*, *10*(1), 166–188. https://doi.org/10.1093/gbe/evx240
Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., … Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences*, *115*(17), 4325–4333. https://doi.org/10.1073/pnas.1720115115
Li, C., Chen, B., Langda, S., Pu, P., Zhu, X., Zhou, S., Kalds, P., Zhang, K., Bhati, M., Leonard, A., Huang, S., Li, R., Cuoji, A., Wang, X., Zhu, H., Wu, Y., Cuomu, R., Gui, B., Li, M., … Wang, X. (2024). Multi-omic Analyses Shed Light on The Genetic Control of High-altitude Adaptation in Sheep. *Genomics, Proteomics & Bioinformatics*, *22*(2). https://doi.org/10.1093/gpbjnl/qzae030
Li, H. (2009, November 10). *SNPable Regions*. Https://Lh3lh3.Users.Sourceforge.Net/Snpable.Shtml.
Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, R., Meng, M., Chen, Y., Pan, T., Li, Y., Deng, Y., Zhang, R., Tian, R., Xu, W., Zheng, X., Gong, F., Liu, J., Tang, H., Ding, X., Tang, Y., Annane, D., Chen, E., Qu, H., & Li, L. (2023). ATP-citrate lyase controls endothelial gluco-lipogenic metabolism and vascular inflammation in sepsis-associated organ injury. *Cell Death & Disease*, *14*(7), 401. https://doi.org/10.1038/s41419-023-05932-8
Lin, Y., Golovnina, K., Chen, Z.-X., Lee, H. N., Negron, Y. L. S., Sultana, H., Oliver, B., & Harbison, S. T. (2016). Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual Drosophila melanogaster. *BMC Genomics*, *17*(1), 28. https://doi.org/10.1186/s12864-015-2353-z
Lipworth, S. (2017, August 2). *base_changer*. Https://Github.Com/Samlipworth/Base_changer/Blob/Master/Base_changer2.Py.
Locatelli, E., Due, R. A., van den Bergh, G. D., & van den Hoek Ostende, L. W. (2012). Pleistocene survivors and Holocene extinctions: The giant rats from Liang Bua (Flores, Indonesia). *Quaternary International*, *281*, 47–57. https://doi.org/10.1016/j.quaint.2012.04.005
Lomolino, M. V., Van der Geer, A. A., Lyras, G. A., Palombo, M. R., Sax, D. F., & Rozzi, R. (2013). Of mice and mammoths: Generality and antiquity of the island rule. *Journal of Biogeography*, *40*(8), 1427–1439. https://doi.org/10.1111/jbi.12096
Long, H., Zhu, X., Yang, P., Gao, Q., Chen, Y., & Ma, L. (2013). Myo9b and RICS Modulate Dendritic Morphology of Cortical Neurons. *Cerebral Cortex*, *23*(1), 71–79. https://doi.org/10.1093/cercor/bhr378
López-Maury, L., Marguerat, S., & Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics*, *9*(8), 583–593. https://doi.org/10.1038/nrg2398

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, A. T., Fei, Z., Haghani, A., Robeck, T. R., Zoller, J. A., Li, C. Z., Zhang, J., Ablaeva, J., Adams, D. M., Almunia, J., Ardehali, R., Arneson, A., Baker, C. S., Belov, K., Blac, P., & Horvath, S. (2021). Universal DNA methylation age across mammalian tissues. *BioRxiv*, 1–28. https://doi.org/https://doi.org/10.1101/2021.01.18.426733

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., Assimes, T. L., Ferrucci, L., & Horvath, S. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*, *11*(2), 303–327. https://doi.org/10.18632/aging.101684

Luo, M., Zhao, J., Merilä, J., Barrett, R. D. H., Guo, B., & Hu, J. (2025). The interplay between epigenomic and transcriptomic variation during ecotype divergence in stickleback. *BMC Biology*, *23*(1), 70. https://doi.org/10.1186/s12915-025-02176-0

Ma, X., & Kang, S. (2019). Functional implications of DNA methylation in adipose biology. *Diabetes*, *68*(5), 871–878. https://doi.org/10.2337/dbi18-0057

MacPherson, I. S., & Stewart, D. T. (2003). Characterisation of a unique dipeptidase allele in an insular population of masked shrews, Sorex cinereus (Insectivora: Soricidae). *Biochemical Systematics and Ecology*, *31*(6), 573–580. https://doi.org/10.1016/S0305-1978(02)00218-1

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics*, *92*(1), 155–161. https://doi.org/10.1007/s12041-013-0212-8

Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biology*, *20*(1), 1–14. https://doi.org/10.1186/s13059-019-1828-7

Maier, R., Flegontov, P., Flegontova, O., Işıldak, U., Changmai, P., & Reich, D. (2023). On the limits of fitting complex models of population history to f-statistics. *ELife*, *12*. https://doi.org/10.7554/eLife.85492

Mann, D. H., Groves, P., Reanier, R. E., Gaglioti, B. V., Kunz, M. L., & Shapiro, B. (2015). Life and extinction of megafauna in the ice-age Arctic. *Proceedings of the National Academy of Sciences*, *112*(46), 14301–14306. https://doi.org/10.1073/pnas.1516573112

Mardis, E. R. (2017). DNA sequencing technologies: 2006–2016. *Nature Protocols*, *12*(2), 213–218. https://doi.org/10.1038/nprot.2016.182

Martchenko, D., & Shafer, A. B. A. (2023). Contrasting whole-genome and reduced representation sequencing for population demographic and adaptive inference: an alpine mammal case study. *Heredity*, *131*(4), 273–281. https://doi.org/10.1038/s41437-023-00643-4

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Martin, S. (2025, February 28). *genomics_general*. Https://Github.Com/Simonhmartin/Genomics_general.

Matsubara, K., Ishibashi, Y., Ohdachi, S., & Matsuda, Y. (2001). A new primer set for sex identification in the genus Sorex (Soricidae, Insectivora). *Molecular Ecology Notes*, *1*(4), 241–242. https://doi.org/10.1046/j.1471-8278.2001.00089.x

Matthews, T. J. (2023). Evolution: The rise and fall of island dwarfs and giants. *Current Biology*, *33*(12), R684–R686. https://doi.org/10.1016/j.cub.2023.04.062

McAlpine, C. S. (2009). *Molecular characterization of the unique Peptidase A Allele within a population of Sorex cinereus on Bon Portage Island, Nova Scotia* [B.Sc. Honours thesis, Acadia University]. https://scholar.acadiau.ca/islandora/object/theses:571

McCrory, C., Fiorito, G., Hernandez, B., Polidoro, S., O'Halloran, A. M., Hever, A., Ni Cheallaigh, C., Lu, A. T., Horvath, S., Vineis, P., & Kenny, R. A. (2021). GrimAge Outperforms Other Epigenetic Clocks in the Prediction of Age-Related Clinical Phenotypes and All-Cause Mortality. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *76*(5), 741–749. https://doi.org/10.1093/gerona/glaa286

McGirr, J. A., & Martin, C. H. (2020). Ecological divergence in sympatry causes gene misexpression in hybrids. *Molecular Ecology*, *29*(14), 2707–2721. https://doi.org/10.1111/mec.15512

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., & Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, *28*(5), 495–501. https://doi.org/10.1038/nbt.1630

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731. https://doi.org/10.1534/genetics.118.301336

Mendes-Soares, H., & Rychlik, L. (2009). Differences in swimming and diving abilities between two sympatric species of water shrews: Neomys anomalus and Neomys fodiens (Soricidae). *Journal of Ethology*, *27*(3), 317–325. https://doi.org/10.1007/s10164-008-0122-z

Meng, G., Li, Y., Yang, C., & Liu, S. (2019a). MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, *47*(11), e63–e63. https://doi.org/10.1093/nar/gkz173

Meng, G., Li, Y., Yang, C., & Liu, S. (2019b). MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, *47*(11). https://doi.org/10.1093/nar/gkz173

Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., Goodbla, A., Eizirik, E., Simão, T. L. L., Stadler, T., Rabosky, D. L., Honeycutt, R. L., Flynn, J. J., Ingram, C. M., Steiner, C., Williams, T. L., Robinson, T. J., Burk-Herrick, A., Westerman, M., … Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science*, *334*(6055), 521–524. https://doi.org/10.1126/science.1211028

Meröndun, J., Murray, D. L., & Shafer, A. B. A. (2019). Genome-scale sampling suggests cryptic epigenetic structuring and insular divergence in Canada lynx. *Molecular Ecology*, *28*(13), 3186–3196. https://doi.org/10.1111/mec.15131

Messias-Reason, I. J., Schafranski, M. D., Kremsner, P. G., & Kun, J. F. J. (2009). Ficolin 2 (FCN2) functional polymorphisms and the risk of rheumatic fever and

rheumatic heart disease. *Clinical and Experimental Immunology*, *157*(3), 395–399. https://doi.org/10.1111/j.1365-2249.2009.03975.x

Miller, J. M., Garant, D., Perrier, C., Juette, T., Jameson, J. W., Normandeau, E., Bernatchez, L., & Réale, D. (2021). Linking genetic, morphological, and behavioural divergence between inland island and mainland deer mice. *Heredity*. https://doi.org/10.1038/s41437-021-00492-z

Minio, A., Massonnet, M., Figueroa-Balderas, R., Vondras, A. M., Blanco-Ulate, B., & Cantu, D. (2019). Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development. *G3 Genes|Genomes|Genetics*, *9*(3), 755–767. https://doi.org/10.1534/g3.118.201008

Mititelu, M., Lupuliasa, D., Neacșu, S. M., Olteanu, G., Busnatu, Ștefan S., Mihai, A., Popovici, V., Măru, N., Boroghină, S. C., Mihai, S., Ioniță-Mîndrican, C.-B., & Scafa-Udriște, A. (2024). Polyunsaturated Fatty Acids and Human Health: A Key to Modern Nutritional Balance in Association with Polyphenolic Compounds from Food Sources. *Foods*, *14*(1), 46. https://doi.org/10.3390/foods14010046

Møller, L. L. V., Klip, A., & Sylow, L. (2019). Rho GTPases—Emerging Regulators of Glucose Homeostasis and Metabolic Health. *Cells*, *8*(5), 434. https://doi.org/10.3390/cells8050434

Monroig, Ó., Lopes-Marques, M., Navarro, J. C., Hontoria, F., Ruivo, R., Santos, M. M., Venkatesh, B., Tocher, D. R., & C. Castro, L. F. (2016). Evolutionary functional elaboration of the Elovl2/5 gene family in chordates. *Scientific Reports*, *6*(1), 20510. https://doi.org/10.1038/srep20510

Montejo-Kovacevich, G., Meier, J. I., Bacquet, C. N., Warren, I. A., Chan, Y. F., Kucka, M., Salazar, C., Rueda-M, N., Montgomery, S. H., McMillan, W. O., Kozak, K. M., Nadeau, N. J., Martin, S. H., & Jiggins, C. D. (2022). Repeated genetic adaptation to altitude in two tropical butterflies. *Nature Communications*, *13*(1), 4676. https://doi.org/10.1038/s41467-022-32316-x

Moore, L. D., Le, T., & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, *38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Morand, S., & Harvey, P. H. (2000). Mammalian metabolism, longevity and parasite species richness. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1456), 1999–2003. https://doi.org/10.1098/rspb.2000.1241

Morgulis, A., Gertz, E. M., Schäffer, A. A., & Agarwala, R. (2006). WindowMasker: window-based masker for sequenced genomes. *Bioinformatics*, *22*(2), 134–141. https://doi.org/10.1093/bioinformatics/bti774

Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., & Beck, S. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics*, *30*(3), 428–430. https://doi.org/10.1093/bioinformatics/btt684

Murphy, W. J., Eizirik, E., Johnson, W. E., Zhang, Y. P., Ryder, O. A., & O'Brien, S. J. (2001). Molecular phylogenetics and the origins of placental mammals. *Nature*, *409*(6820), 614–618. https://doi.org/10.1038/35054550

Nagorsen, D. (1996). *Opossums, Shrews and Moles of British Columbia*. UBC Press.

Nandi, G. (2025). *Molecular characterization of the PEP A (CNDP2) gene in masked shrews (Sorex cinereus) on Bon Portage Island* [Acadia University]. https://doi.org/https://doi.org/10.26108/2E94-J684

Nannini, D. R., Joyce, B. T., Zheng, Y., Gao, T., Liu, L., Yoon, G., Huan, T., Ma, J., Jacobs, D. R., Wilkins, J. T., Ren, J., Zhang, K., Khan, S. S., Allen, N. B., Horvath, S., Lloyd-Jones, D. M., Greenland, P., & Hou, L. (2019). Epigenetic age acceleration and metabolic syndrome in the coronary artery risk development in young adults study. *Clinical Epigenetics*, *11*(1). https://doi.org/10.1186/s13148-019-0767-1

National Center for Biotechnology Information (NCBI). (1988). *National Library of Medicine (US), National Center for Biotechnology Information*. from: https://www.ncbi.nlm.nih.gov/

Ning, Z., Giordano, F., & Harry, E. (n.d.). *Scaff10X*. https://github.com/wtsi-hpag/Scaff10X

Novosolov, M., Raia, P., & Meiri, S. (2013). The island syndrome in lizards. *Global Ecology and Biogeography*, *22*(2), 184–191. https://doi.org/10.1111/j.1466-8238.2012.00791.x

Oblak, L., van der Zaag, J., Higgins-Chen, A. T., Levine, M. E., & Boks, M. P. (2021). A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. *Ageing Research Reviews*, *69*, 101348. https://doi.org/10.1016/j.arr.2021.101348

Ochocinska, D., & Taylor, J. R. E. (2003). Bergmann's rule in shrews: geographical variation of body size in Palearctic Sorex species. *Biological Journal of the Linnean Society*, *78*(3), 365–381. https://doi.org/10.1046/j.1095-8312.2003.00150.x

Ochocińska, D., & Taylor, J. R. E. (2005). Living at the physiological limits: Field and maximum metabolic rates of the common shrew (Sorex araneus). *Physiological and Biochemical Zoology*, *78*(5), 808–818. https://doi.org/10.1086/431190

Ohkubo, Y., Uchida, A. O., Shin, D., Partanen, J., & Vaccarino, F. M. (2004). Fibroblast Growth Factor Receptor 1 Is Required for the Proliferation of Hippocampal Progenitor Cells and for Hippocampal Growth in Mouse. *The Journal of Neuroscience*, *24*(27), 6057–6069. https://doi.org/10.1523/JNEUROSCI.1140-04.2004

Okur, V., Cho, M. T., Henderson, L., Retterer, K., Schneider, M., Sattler, S., Niyazov, D., Azage, M., Smith, S., Picker, J., Lincoln, S., Tarnopolsky, M., Brady, L., Bjornsson, H. T., Applegate, C., Dameron, A., Willaert, R., Baskin, B., Juusola, J., & Chung, W. K. (2016). De novo mutations in CSNK2A1 are associated with neurodevelopmental abnormalities and dysmorphic features. *Human Genetics*, *135*(7), 699–705. https://doi.org/10.1007/s00439-016-1661-y

Oulion, S., Bertrand, S., & Escriva, H. (2012). Evolution of the FGF Gene Family. *International Journal of Evolutionary Biology*, *2012*, 1–12. https://doi.org/10.1155/2012/298147

Palma, F. Di, Holme, R. H., Bryda, E. C., Belyantseva, I. A., Pellegrino, R., Kachar, B., Steel, K. P., & Noben-Trauth, K. (2001). Mutations in Cdh23, encoding a new type of cadherin, cause stereocilia disorganization in waltzer, the mouse model for Usher syndrome type 1D. *Nature Genetics*, *27*(1), 103–107. https://doi.org/10.1038/83660

Parker, M. (2012). *Ghost Islands of Nova Scotia*. Pottersfield Press.

Pavlidis, P., Jensen, J. D., Stephan, W., & Stamatakis, A. (2012). A critical assessment of storytelling: Gene ontology categories and the importance of validating

genomic scans. *Molecular Biology and Evolution*, *29*(10), 3237–3248. https://doi.org/10.1093/molbev/mss136

Pedersen, B. S., & Quinlan, A. R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics (Oxford, England)*, *34*(5), 867–868. https://doi.org/10.1093/bioinformatics/btx699

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Pfeffer, T., Krug, S. M., Kracke, T., Schürfeld, R., Colbatzky, F., Kirschner, P., Medert, R., Freichel, M., Schumacher, D., Bartosova, M., Zarogiannis, S. G., Muckenthaler, M. U., Altamura, S., Pezer, S., Volk, N., Schwab, C., Duensing, S., Fleming, T., Heidenreich, E., … Peters, V. (2024). Knock-out of dipeptidase CN2 in human proximal tubular cells disrupts dipeptide and amino acid homeostasis and para- and transcellular solute transport. *Acta Physiologica*, *240*(4). https://doi.org/10.1111/apha.14126

Pinho, G. M., Martin, J. G. A., Farrell, C., Haghani, A., Zoller, J. A., Zhang, J., Snir, S., Pellegrini, M., Wayne, R. K., Blumstein, D. T., & Horvath, S. (2022). Hibernation slows epigenetic ageing in yellow-bellied marmots. *Nature Ecology & Evolution*, *6*(4), 418–426. https://doi.org/10.1038/s41559-022-01679-1

Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J. S., Bejerano, G., Baertsch, R., Rosenbloom, K. R., Kent, J., & Haussler, D. (2006). Forces Shaping the Fastest Evolving Regions in the Human Genome. *PLoS Genetics*, *2*(10), e168. https://doi.org/10.1371/journal.pgen.0020168

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J. S., Katzman, S., King, B., Onodera, C., Siepel, A., Kern, A. D., Dehay, C., Igel, H., Ares, M., Vanderhaeghen, P., & Haussler, D. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, *443*(7108), 167–172. https://doi.org/10.1038/nature05113

Pond, S. L. K., Frost, S. D. W., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, *21*(5), 676–679. https://doi.org/10.1093/bioinformatics/bti079

Ponting, C. P. (2017). Biological function in the twilight zone of sequence conservation. *BMC Biology*, *15*(1), 71. https://doi.org/10.1186/s12915-017-0411-5

Posbergh, C. J., & Huson, H. J. (2021). All sheeps and sizes: a genetic investigation of mature body size across sheep breeds reveals a polygenic nature. *Animal Genetics*, *52*(1), 99–107. https://doi.org/10.1111/age.13016

Postma, E. (2014). Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data, better estimates? In Charmantier Anne, Garant Dany, & Kruuk Loeske E. B. (Eds.), *Quantitative Genetics in the Wild* (pp. 16–33). Oxford University Press.

Prado, N. A., Brown, J. L., Zoller, J. A., Haghani, A., Yao, M., Bagryanova, L. R., Campana, M. G., E. Maldonado, J., Raj, K., Schmitt, D., Robeck, T. R., & Horvath, S. (2021). Epigenetic clock and methylation studies in elephants. *Aging Cell*, *20*(7), 1–11. https://doi.org/10.1111/acel.13414

Pravosudov, V. V., Roth, T. C., Forister, M. L., LaDage, L. D., Kramer, R., Schilkey, F., & van der Linden, A. M. (2013). Differential hippocampal gene expression is associated with climate-related natural variation in memory and the hippocampus in food-caching chickadees. *Molecular Ecology*, *22*(2), 397–408. https://doi.org/10.1111/mec.12146

Pruitt, W. O. (1954). Aging in the Masked Shrew, Sorex cinereus cinereus Kerr. *Journal of Mammalogy*, *35*(1), 35–39. https://doi.org/10.2307/1376070

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raj, K., Szladovits, B., Haghani, A., Zoller, J. A., Li, C. Z., Black, P., Maddox, D., Robeck, T. R., & Horvath, S. (2021). Epigenetic clock and methylation studies in cats. *GeroScience*, *43*(5), 2363–2378. https://doi.org/10.1007/s11357-021-00445-8

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Ray, S., Li, M., Koch, S. P., Mueller, S., Boehm-Sturm, P., Wang, H., Brecht, M., & Naumann, R. K. (2020). Seasonal plasticity in the adult somatosensory cortex. *Proceedings of the National Academy of Sciences*, *117*(50), 32136–32144. https://doi.org/10.1073/pnas.1922888117

Reinhard, J. R., Kriz, A., Galic, M., Angliker, N., Rajalu, M., Vogt, K. E., & Ruegg, M. A. (2016). The calcium sensor Copine-6 regulates spine structural plasticity and learning and memory. *Nature Communications*, *7*(1), 11613. https://doi.org/10.1038/ncomms11613

Reumer, J. W. F. (1989). Speciation and evolution in the Soricidae (Mammalia: Insectivora) in relation with the paleoclimate. *Revue Suisse de Zoologie.*, *96*, 81–90. https://doi.org/10.5962/bhl.part.117758

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., … Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1), 245. https://doi.org/10.1186/s13059-020-02134-9

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and its Applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Rick, J. (2024, March 7). *MSMC Tutorial and Scripts for running MSMC2*. Https://Github.Com/Jessicarick/Msmc2_scripts.

Rinn, J. L., & Chang, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, *81*(1), 145–166. https://doi.org/10.1146/annurev-biochem-051410-092902

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Robeck, T. R., Fei, Z., Lu, A. T., Haghani, A., Jourdain, E., Zoller, J. A., Li, C. Z., Steinman, K. J., DiRocco, S., Schmitt, T., Osborn, S., Van Bonn, B., Katsumata, E., Mergl, J., Almunia, J., Rodriguez, M., Haulena, M., Dold, C., & Horvath, S. (2021). Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Communications Biology*, *4*(1). https://doi.org/10.1038/s42003-021-02179-x

Roland, A. E. (1982). *Geological background and physiography of Nova Scotia*. Nova Scotian Institute of Science.

Romero, I. G., Ruvinsky, I., & Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, *13*(7), 505–516. https://doi.org/10.1038/nrg3229

Roux, I., Safieddine, S., Nouvian, R., Grati, M., Simmler, M.-C., Bahloul, A., Perfettini, I., Le Gall, M., Rostaing, P., Hamard, G., Triller, A., Avan, P., Moser, T., & Petit, C. (2006). Otoferlin, Defective in a Human Deafness Form, Is Essential for Exocytosis at the Auditory Ribbon Synapse. *Cell*, *127*(2), 277–289. https://doi.org/10.1016/j.cell.2006.08.040

Roychaudhuri, R., Atashi, H., & Snyder, S. H. (2023). Serine Racemase mediates subventricular zone neurogenesis via fatty acid metabolism. *Stem Cell Reports*, *18*(7), 1482–1499. https://doi.org/10.1016/j.stemcr.2023.05.015

Rudd, R. L. (1955). Age, Sex, and Weight Comparisons in Three Species of Shrews. *Journal of Mammalogy*, *36*(3), 323–339. https://doi.org/10.2307/1375674

Saino, N., Ambrosini, R., Albetti, B., Caprioli, M., De Giorgio, B., Gatti, E., Liechti, F., Parolini, M., Romano, A., Romano, M., Scandolara, C., Gianfranceschi, L., Bollati, V., & Rubolini, D. (2017). Migration phenology and breeding success are predicted by methylation of a photoperiodic gene in the barn swallow. *Scientific Reports*, *7*(January), 1–10. https://doi.org/10.1038/srep45412

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual* (2nd ed.). Cold Spring Harbor Laboratory Press.

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences*, *103*(5), 1412–1417. https://doi.org/10.1073/pnas.0510310103

Schachtschneider, K. M., Schook, L. B., Meudt, J. J., Shanmuganayagam, D., Zoller, J. A., Haghani, A., Li, C. Z., Zhang, J., Yang, A., Raj, K., & Horvath, S. (2021). Epigenetic clock and DNA methylation analysis of porcine models of aging and

obesity. *GeroScience*, *43*(5), 2467–2483. https://doi.org/10.1007/s11357-021-00439-6
Schaeffer, P. J., O'Mara, M. T., Breiholz, J., Keicher, L., Lázaro, J., Muturi, M., & Dechmann, D. K. N. (2020). Metabolic rate in common shrews is unaffected by temperature, leading to lower energetic costs through seasonal size reduction. *Royal Society Open Science*, *7*(4). https://doi.org/10.1098/rsos.191989
Schiffels, S., & Wang, K. (2020). *MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent* (pp. 147–166). https://doi.org/10.1007/978-1-0716-0199-0_7
Schlitter, D. A., Hutterer, R., Maddalena, T., & Robbins, L. W. (1999). New karyotypes of shrews (Mammalia: Soricidae) from Cameroon and Somalia. *Annals of the Carnegie Museum*, *68*(1), 1–14. https://doi.org/10.5962/p.226615
Seidler, N. W. (2013). *GAPDH and Intermediary Metabolism* (pp. 37–59). https://doi.org/10.1007/978-94-007-4716-6_2
Shafer, A. B. A., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*(21), 4589–4621. https://doi.org/10.1111/j.1365-294X.2010.04828.x
Shafer, A. B. A., & Kardos, M. (2025). Runs of Homozygosity and Inferences in Wild Populations. *Molecular Ecology*, *34*(3). https://doi.org/10.1111/mec.17641
Sharma, V., Hecker, N., Roscito, J. G., Foerster, L., Langer, B. E., & Hiller, M. (2018). A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications*, *9*(1), 1215. https://doi.org/10.1038/s41467-018-03667-1
Sharman, G. B. (1956). Chromosomes of the Common Shrew. *Nature*, *177*(4516), 941–942. https://doi.org/10.1038/177941a0
Shen, B., Yang, Z., Han, S., Zou, Z., Liu, J., Nie, L., Dong, W., Li, E., Liu, S., Zhao, Z., & Wu, R. (2019). Bta-miR-124a Affects Lipid Metabolism by Regulating PECR Gene. *BioMed Research International*, *2019*, 1–9. https://doi.org/10.1155/2019/2596914
Shen, Y.-Y., Liang, L., Li, G.-S., Murphy, R. W., & Zhang, Y.-P. (2012). Parallel Evolution of Auditory Genes for Echolocation in Bats and Toothed Whales. *PLoS Genetics*, *8*(6), e1002788. https://doi.org/10.1371/journal.pgen.1002788
Sherlock, M. B., Wilkinson, M., Maddock, S. T., Nussbaum, R. A., Day, J. J., & Streicher, J. W. (2025). Submerged Corridors of Ancient Gene Flow in an Island Amphibian. *Molecular Ecology*, *34*(9). https://doi.org/10.1111/mec.17742
Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, *50*(W1), W216–W221. https://doi.org/10.1093/nar/gkac194
Signor, S. A., & Nuzhdin, S. V. (2018). The Evolution of Gene Expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007
Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Singh, P., Börger, C., More, H., & Sturmbauer, C. (2017). The Role of Alternative Splicing and Differential Gene Expression in Cichlid Adaptive Radiation. *Genome Biology and Evolution*, *9*(10), 2764–2781. https://doi.org/10.1093/gbe/evx204

Skotarczak, E., Szwaczkowski, T., & Ćwiertnia, P. (2020). Effects of inbreeding, sex and geographical region on survival in an American bison (Bison bison) population under a captive breeding program. *European Zoological Journal*, *87*(1), 402–411. https://doi.org/10.1080/24750263.2020.1797194

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics*, *195*(3), 693–702. https://doi.org/10.1534/genetics.113.154138

Smit, A., & Hubley, R. (2023a). *RepeatMasker Open-4.0*. http://www.repeatmasker.org

Smit, A., & Hubley, R. (2023b). *RepeatModeler Open-1.0*. http://www.repeatmasker.org

Smith, A. K., Kilaru, V., Kocak, M., Almli, L. M., Mercer, K. B., Ressler, K. J., Tylavsky, F. A., & Conneely, K. N. (2014). Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics*, *15*(1). https://doi.org/10.1186/1471-2164-15-145

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution*, *32*(5), 1342–1353. https://doi.org/10.1093/molbev/msv022

Snead, A. A., & Clark, R. D. (2022). The Biological Hierarchy, Time, and Temporal 'Omics in Evolutionary Biology: A Perspective. *Integrative And Comparative Biology*, *62*(6), 1872–1886. https://doi.org/10.1093/icb/icac138

Spies, N., Weng, Z., Bishara, A., McDaniel, J., Catoe, D., Zook, J. M., Salit, M., West, R. B., Batzoglou, S., & Sidow, A. (2017). Genome-wide reconstruction of complex structural variants using read clouds. *Nature Methods*, *14*(9), 915–920. https://doi.org/10.1038/nmeth.4366

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(Web Server), W435–W439. https://doi.org/10.1093/nar/gkl200

Stankiewicz, T. R., & Linseman, D. A. (2014). Rho family GTPases: key players in neuronal development, neuronal survival, and neurodegeneration. *Frontiers in Cellular Neuroscience*, *8*. https://doi.org/10.3389/fncel.2014.00314

Steijger, T., Abril, J. F., Engström, P. G., Kokocinski, F., Hubbard, T. J., Guigó, R., Harrow, J., & Bertone, P. (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods*, *10*(12), 1177–1184. https://doi.org/10.1038/nmeth.2714

Steinthorsdottir, M., Coxall, H. K., de Boer, A. M., Huber, M., Barbolini, N., Bradshaw, C. D., Burls, N. J., Feakins, S. J., Gasson, E., Henderiks, J., Holbourn, A. E., Kiel, S., Kohn, M. J., Knorr, G., Kürschner, W. M., Lear, C. H., Liebrand, D., Lunt, D. J., Mörs, T., … Strömberg, C. A. E. (2021). The Miocene: The Future of the Past. *Paleoceanography and Paleoclimatology*, *36*(4). https://doi.org/10.1029/2020PA004037

Stewart, D. T., & Baker, A. J. (1992). Genetic differentiation and biogeography of the masked shrew in Atlantic Canada. *Canadian Journal of Zoology*, *70*(1), 106–114. https://doi.org/10.1139/z92-016

Stewart, D. T., Herman, T. B., & Teferi, T. (1989). Littoral feeding in a high-density insular population of Sorex cinereus. *Canadian Journal of Zoology*, *67*(8), 2074–2077. https://doi.org/10.1139/z89-295

Stewart, D. T., Perry, N. D., & Fumagalli, L. (2002). The maritime shrew, Sorex maritimensis (Insectivora: Soricidae): A newly recognized Canadian endemic. *Canadian Journal of Zoology*, *80*(1), 94–99. https://doi.org/10.1139/z01-207

Stockley, P., Searle, J. B., MacDonald, D. W., & Jones, C. S. (1993). Female multiple mating behaviour in the common shrew as a strategy to reduce inbreeding. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *254*(1341), 173–179. https://doi.org/10.1098/rspb.1993.0143

Stokes, C. R. (2017). Deglaciation of the Laurentide Ice Sheet from the Last Glacial Maximum. *Cuadernos de Investigación Geográfica: Geographical Research Letters*, *43*(2), 377–428.

Sugrue, V. J., Zoller, J. A., Narayan, P., Lu, A. T., Ortega-Recalde, O. J., Grant, M. J., Bawden, C. S., Rudiger, S. R., Haghani, A., Bond, D. M., Hore, R. R., Garratt, M., Sears, K. E., Wang, N., Yang, X. W., Snell, R. G., Hore, T. A., & Horvath, S. (2021). Castration delays epigenetic aging and feminizes dna methylation at androgen-regulated loci. *ELife*, *10*, 1–20. https://doi.org/10.7554/eLife.64932

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*(Web Server), W609–W612. https://doi.org/10.1093/nar/gkl315

Swanson, E. M., & Dantzer, B. (2014). Insulin-like growth factor-1 is associated with life-history variation across Mammalia. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1782). https://doi.org/10.1098/rspb.2013.2458

Syvänen, A.-C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, *2*(12), 930–942. https://doi.org/10.1038/35103535

Szukala, A., Lovegrove-Walsh, J., Luqman, H., Fior, S., Wolfe, T. M., Frajman, B., Schönswetter, P., & Paun, O. (2023). Polygenic routes lead to parallel altitudinal adaptation in Heliosperma pusillum (Caryophyllaceae). *Molecular Ecology*, *32*(8), 1832–1847. https://doi.org/10.1111/mec.16393

Takahashi, M., Kamei, Y., Ezaki, O., Takahashi, Y., & Kamei, O. E. (2005). Mest/Peg1 imprinted gene enlarges adipocytes and is a marker of adipocyte size. *Am J Physiol Endocrinol Metab*, *288*, 117–124. https://doi.org/10.1152/ajpendo

Talavera, A., Palmada-Flores, M., Martínez-Freiría, F., Mochales-Riaño, G., Burriel-Carranza, B., Estarellas, M., Fernández-Guiberteau, D., Camina, Á., Ursenbacher, S., Vörös, J., Halpern, B., Pla, D., Calvete, J. J., Mikheyev, A. S., Marquès-Bonet, T., & Carranza, S. (2025). Unveiling the Evolutionary History of European Vipers and Their Venoms From a Multi-Omic Approach. *Molecular Ecology*. https://doi.org/10.1111/mec.70019

Tanner, J. J. (2019). Structural Biology of Proline Catabolic Enzymes. *Antioxidants & Redox Signaling*, *30*(4), 650–673. https://doi.org/10.1089/ars.2017.7374

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics*, *31*(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Taylor, J. R. E., Rychlik, L., & Churchfield, S. (2013). Winter reduction in body mass in a very small, nonhibernating mammal: Consequences for heat loss and metabolic rates. *Physiological and Biochemical Zoology*, *86*(1), 9–18. https://doi.org/10.1086/668484

Teeling, E. C., Vernes, S. C., Dávalos, L. M., Ray, D. A., Gilbert, M. T. P., & Myers, E. (2018). Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. *Annual Review of Animal Biosciences*, *6*(1), 23–46. https://doi.org/10.1146/annurev-animal-022516-022811

Tegelström, H., Searle, J., Brookfield, J., & Mercer, S. (1991). Multiple paternity in wild common shrews (Sorex araneus) is confirmed by DNA-fingerprinting. *Heredity*, *66*(3), 373–379. https://doi.org/10.1038/hdy.1991.47

Telfer, K. A. (1984). *Microhabitat associations, morphopology, and demography of shrews (Soricidae) in Nova Scotia* [B.Sc. Honours thesis]. Acadia University, Wolfville, N.S.

Thomas, W. R., Baldoni, C., Zeng, Y., Carlson, D., Holm-Jacobsen, J., Muturi, M., von Elverfeldt, D., Nieland, J., Dechmann, D. K. N., Corthals, A., & Dávalos, L. M. (2023). *Dynamic metabolic and molecular changes during seasonal shrinking in Sorex araneus*. https://doi.org/10.1101/2023.10.02.560485

Thomas, W. R., Dechmann, D. K. N., Nieland, J., Baldoni, C., Carlson, D., Elverfeldt, D. von, Holm-Jacobsen, J., Muturi, M., Corthals, A., & Dávalos, L. M. (2023). Molecular mechanisms of seasonal brain shrinkage and regrowth in Sorex araneus. *BioRxiv*, 2023.10.02.560485. https://doi.org/10.1101/2023.10.02.560485

Thomas, W. R., Richter, T., O'Neil, E. T., Baldoni, C., Corthals, A., von Elverfeldt, D., Nieland, J. D., Dechmann, D., Hunter, R., & Davalos, L. M. (2025). Seasonal and comparative evidence of adaptive gene expression in mammalian brain size plasticity. *ELife*, *13*. https://doi.org/10.7554/eLife.100788

Tillander, V., Alexson, S. E. H., & Cohen, D. E. (2017). Deactivating Fatty Acids: Acyl-CoA Thioesterase-Mediated Control of Lipid Metabolism. *Trends in Endocrinology & Metabolism*, *28*(7), 473–484. https://doi.org/10.1016/j.tem.2017.03.001

Tollis, M., Ferris, E., Campbell, M. S., Harris, V. K., Rupp, S. M., Harrison, T. M., Kiso, W. K., Schmitt, D. L., Garner, M. M., Aktipis, C. A., Maley, C. C., Boddy, A. M., Yandell, M., Gregg, C., Schiffman, J. D., & Abegglen, L. M. (2021). Elephant Genomes Reveal Accelerated Evolution in Mechanisms Underlying Disease Defenses. *Molecular Biology and Evolution*, *38*(9), 3606–3620. https://doi.org/10.1093/molbev/msab127

Tomasi, T. E. (1979). Echolocation by the Short-Tailed Shrew Blarina brevicauda. *Journal of Mammalogy*, *60*(4), 751–759. https://doi.org/10.2307/1380190

UCSC Genome Browser. (2020). *kentutils* (401). https://github.com/ucscGenomeBrowser/kent

Uliano-Silva, M., Ferreira, J. G. R. N., Krasheninnikova, K., Darwin Tree of Life Consortium, Formenti, G., Abueg, L., Torrance, J., Myers, E. W., Durbin, R., Blaxter, M., & McCarthy, S. A. (2023). MitoHiFi: a python pipeline for

mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*, *24*(288), 1–13. https://link.springer.com/content/pdf/10.1186/s12859-023-05385-y.pdf

van Egeren, D., Madsen, T., & Michor, F. (2018). Fitness variation in isogenic populations leads to a novel evolutionary mechanism for crossing fitness valleys. *Communications Biology*, *1*(1). https://doi.org/10.1038/s42003-018-0160-1

Vernaz, G., Hudson, A. G., Santos, M. E., Fischer, B., Carruthers, M., Shechonge, A. H., Gabagambi, N. P., Tyers, A. M., Ngatunga, B. P., Malinsky, M., Durbin, R., Turner, G. F., Genner, M. J., & Miska, E. A. (2021). Epigenetic Divergence during Early Stages of Speciation in an African Crater Lake Cichlid Fish. *BioRxiv*, *doi:10.1101/2021.07.30.435319*, 1–12.

Vernaz, G., Malinsky, M., Svardal, H., Du, M., Tyers, A. M., Santos, M. E., Durbin, R., Genner, M. J., Turner, G. F., & Miska, E. A. (2021). Mapping epigenetic divergence in the massive radiation of Lake Malawi cichlid fishes. *Nature Communications*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-26166-2

Verta, J.-P., & Jones, F. C. (2019). Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *ELife*, *8*. https://doi.org/10.7554/eLife.43785

Vogt, G. (2017). Facilitation of environmental adaptation and evolution by epigenetic phenotype variation: Insights from clonal, invasive, polyploid, and domesticated animals. *Environmental Epigenetics*, *3*(1), 1–17. https://doi.org/10.1093/eep/dvx002

Wan, J., Oliver, V. F., Wang, G., Zhu, H., Zack, D. J., Merbs, S. L., & Qian, J. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-1271-4

Wang, B., Mojica, J. P., Perera, N., Lee, C.-R., Lovell, J. T., Sharma, A., Adam, C., Lipzen, A., Barry, K., Rokhsar, D. S., Schmutz, J., & Mitchell-Olds, T. (2019). Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in Boechera stricta. *Genome Biology*, *20*(1), 126. https://doi.org/10.1186/s13059-019-1729-9

Wang, X., Song, H., Liang, J., Jia, Y., & Zhang, Y. (2022). Abnormal expression of HADH, an enzyme of fatty acid oxidation, affects tumor development and prognosis (Review). *Molecular Medicine Reports*, *26*(6), 355. https://doi.org/10.3892/mmr.2022.12871

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Wang (王博), B., Jia (贾鹏), P., Gao (高胜寒), S., Zhao (赵焕焕), H., Zheng (郑高洋), G., Xu (许林峰), L., & Ye (叶凯), K. (2025). Long and Accurate: How HiFi Sequencing is Transforming Genomics. *Genomics, Proteomics & Bioinformatics*, *23*(1). https://doi.org/10.1093/gpbjnl/qzaf003

Warwas, N., Berdan, E. L., Xie, X., Jönsson, E., Roques, J. A. C., Doyle, D., Langeland, M., Hinchcliffe, J., Pavia, H., & Sundell, K. (2024). Seaweed Fly Larvae Cultivated on Macroalgae Side Streams: A Novel Marine Protein and Omega-3 Source for Rainbow Trout. *Aquaculture Nutrition*, *2024*(1). https://doi.org/10.1155/2024/4221883

Watt, F., & Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & Development*, *2*(9), 1136–1143. https://doi.org/10.1101/gad.2.9.1136

Wayne, R. K., George, S. B., Gilbert, D., Collins, P. W., Kovach, S. D., Girman, D., & Lehman, N. (1991). A morphologic and genetic study of the island fox, Urocyon littoralis. *Evolution*, *45*(8), 1849–1868. https://doi.org/10.1111/j.1558-5646.1991.tb02692.x

Wegner, K. M., Lokmer, A., & John, U. (2020). Genomic and Transcriptomic Differentiation of Independent Invasions of the Pacific Oyster Crassostrea gigas. *Frontiers in Ecology and Evolution*, *8*. https://doi.org/10.3389/fevo.2020.567049

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2018). Corrigendum: Direct determination of diploid genome sequences. *Genome Research*, *28*(4), 757–767. https://doi.org/10.1101/gr.235812.118

Whitaker, J. O. Jr., & French, T. W. (1984). Foods of six species of sympatric shrews from New Brunswick. *Canadian Journal of Zoology*, *62*, 622–626.

Whitaker Jr., J. O., & French, T. W. (1984). Foods of six species of sympatric shrews from New Brunswick. *Canadian Journal of Zoology*, *62*(4), 622–626. https://doi.org/10.1139/z84-091

White, T. A., Wójcik, J. M., & Searle, J. B. (2019). Shrews, Chromosomes and Speciation. In *Shrews, Chromosomes and Speciation*. Cambridge University Press. https://doi.org/10.1017/9780511895531.007

Wilkinson, G. S., Adams, D. M., Haghani, A., Lu, A. T., Zoller, J., Breeze, C. E., Arnold, B. D., Ball, H. C., Carter, G. G., Cooper, L. N., Dechmann, D. K. N., Devanna, P., Fasel, N. J., Galazyuk, A. V., Günther, L., Hurme, E., Jones, G., Knörnschild, M., Lattenkamp, E. Z., … Horvath, S. (2021). DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-21900-2

Wójcik, J. M., Ratkiewicz, M., & Searle, J. B. (2002). Evolution of the common shrew, Sorex araneus: chromosomal and molecular aspects. *Acta Theriologica*, *47*(S1), 139–167. https://doi.org/10.1007/BF03192485

Wolf, J. F., Bowman, J., Keobouasone, S., Taylor, R. S., & Wilson, P. J. (2022). A de novo genome assembly and annotation of the southern flying squirrel (Glaucomys volans). *G3: Genes, Genomes, Genetics*, *12*(1). https://doi.org/10.1093/G3JOURNAL/JKAB373

Wong, J. M., Folorunso, O. O., Barragan, E. V., Berciu, C., Harvey, T. L., Coyle, J. T., Balu, D. T., & Gray, J. A. (2020). Postsynaptic Serine Racemase Regulates NMDA Receptor Function. *The Journal of Neuroscience*, *40*(50), 9564–9575. https://doi.org/10.1523/JNEUROSCI.1525-20.2020

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3). https://doi.org/10.1186/gb-2014-15-3-r46

Wooldridge, T. B., Kautt, A. F., Lassance, J.-M., McFadden, S., Domingues, V. S., Mallarino, R., & Hoekstra, H. E. (2022). An enhancer of Agouti contributes to parallel evolution of cryptically colored beach mice. *Proceedings of the National Academy of Sciences*, *119*(27). https://doi.org/10.1073/pnas.2202862119

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, *15*(1), 323–354.

Xu, B., Chen, L., Zhan, Y., Marquez, K. N. S., Zhuo, L., Qi, S., Zhu, J., He, Y., Chen, X., Zhang, H., Shen, Y., Chen, G., Gu, J., Guo, Y., Liu, S., & Xie, T. (2022). The Biological Functions and Regulatory Mechanisms of Fatty Acid Binding Protein 5 in Various Diseases. *Frontiers in Cell and Developmental Biology*, *10*. https://doi.org/10.3389/fcell.2022.857919

Yaffe, E., & Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature Genetics*, *43*(11), 1059–1065. https://doi.org/10.1038/ng.947

Yamauchi, T., Nio, Y., Maki, T., Kobayashi, M., Takazawa, T., Iwabu, M., Okada-Iwabu, M., Kawamoto, S., Kubota, N., Kubota, T., Ito, Y., Kamon, J., Tsuchida, A., Kumagai, K., Kozono, H., Hada, Y., Ogata, H., Tokuyama, K., Tsunoda, M., … Kadowaki, T. (2007). Targeted disruption of AdipoR1 and AdipoR2 causes abrogation of adiponectin binding and metabolic actions. *Nature Medicine*, *13*(3), 332–339. https://doi.org/10.1038/nm1557

Yang, C., Gagnon, D., Vachon, P., Tremblay, A., Levy, E., Massie, B., & Michaud, J. L. (2006). Adenoviral-Mediated Modulation of Sim1 Expression in the Paraventricular Nucleus Affects Food Intake. *Journal of Neuroscience*, *26*(26), 7116–7120. https://doi.org/10.1523/JNEUROSCI.0672-06.2006

Yom-Tov, Y., & Yom-Tov, J. (2005). Global warming, Bergmann's rule and body size in the masked shrew Sorex cinereus Kerr in Alaska. *Journal of Animal Ecology*, *74*(5), 803–808. https://doi.org/10.1111/j.1365-2656.2005.00976.x

Yordy, J., Kraus, C., Hayward, J. J., White, M. E., Shannon, L. M., Creevy, K. E., Promislow, D. E. L., & Boyko, A. R. (2020). Body size, inbreeding, and lifespan in domestic dogs. *Conservation Genetics*, *21*(1), 137–148. https://doi.org/10.1007/s10592-019-01240-x

York, N. S., Sanchez-Arias, J. C., McAdam, A. C. H., Rivera, J. E., Arbour, L. T., & Swayne, L. A. (2022). Mechanisms underlying the role of ankyrin-B in cardiac and neurological health and disease. *Frontiers in Cardiovascular Medicine*, *9*. https://doi.org/10.3389/fcvm.2022.964675

Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, *31*(14), 2382–2383. https://doi.org/10.1093/bioinformatics/btv145

Zeng, X., Yi, Z., Zhang, X., Du, Y., Li, Y., Zhou, Z., Chen, S., Zhao, H., Yang, S., Wang, Y., & Chen, G. (2024). Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nature Plants*, *10*(8), 1184–1200. https://doi.org/10.1038/s41477-024-01755-3

Zhai, W., Nielsen, R., & Slatkin, M. (2009). An Investigation of the Statistical Power of Neutrality Tests Based on Comparative and Population Genetic Data. *Molecular Biology and Evolution*, *26*(2), 273–283. https://doi.org/10.1093/molbev/msn231

Zhang, H., Ben Zablah, Y., Zhang, H., & Jia, Z. (2021). Rho Signaling in Synaptic Plasticity, Memory, and Brain Disorders. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.729076

Zhang, S., Hulver, M. W., McMillan, R. P., Cline, M. A., & Gilbert, E. R. (2014). The pivotal role of pyruvate dehydrogenase kinases in metabolic flexibility. *Nutrition & Metabolism*, *11*(1), 10. https://doi.org/10.1186/1743-7075-11-10

Zhao, J., Chen, J., Li, M., Chen, M., & Sun, C. (2020). Multifaceted Functions of CH25H and 25HC to Modulate the Lipid Metabolism, Immune Responses, and Broadly Antiviral Activities. *Viruses*, *12*(7), 727. https://doi.org/10.3390/v12070727

Zhao, L., Zhang, H., Kohnen, M. V., Prasad, K. V. S. K., Gu, L., & Reddy, A. S. N. (2019). Analysis of Transcriptome and Epitranscriptome in Plants Using PacBio Iso-Seq and Nanopore-Based Direct RNA Sequencing. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00253

Zhao, T., Guan, X., Hu, Y., Zhang, Z., Yang, H., Shi, X., Han, J., Mei, H., Wang, L., Shao, L., Wu, H., Chen, Q., Zhao, Y., Pan, J., Hao, Y., Dong, Z., Long, X., Deng, Q., Zhao, S., … Fang, L. (2024). Population-wide DNA methylation polymorphisms at single-nucleotide resolution in 207 cotton accessions reveal epigenomic contributions to complex traits. *Cell Research*, *34*(12), 859–872. https://doi.org/10.1038/s41422-024-01027-x

Zhao, Y., Chen, Z., Hu, M., Liu, H., Zhao, H., Huang, Y., Jiang, M., Li, S., Li, G., Zhu, C., Hu, W., & Luo, D. (2024). Integrating Iso-seq and RNA-seq data for the reannotation of the greater amberjack genome. *Scientific Data*, *11*(1), 675. https://doi.org/10.1038/s41597-024-03495-7

Zhou, W., Gross, K. M., & Kuperwasser, C. (2019). Molecular regulation of Snai2 in development and disease. *Journal of Cell Science*, *132*(23). https://doi.org/10.1242/jcs.235127

Zhou, W., Triche, T. J., Laird, P. W., & Shen, H. (2018). SeSAMe: Reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Research*, *46*(20), 1–15. https://doi.org/10.1093/nar/gky691

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Zuckerkandl, E., & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. In *Evolving Genes and Proteins* (pp. 97–166). Elsevier. https://doi.org/10.1016/B978-1-4832-2734-4.50017-6

Appendix I: Supplementary Material for Chapter 2

Figures

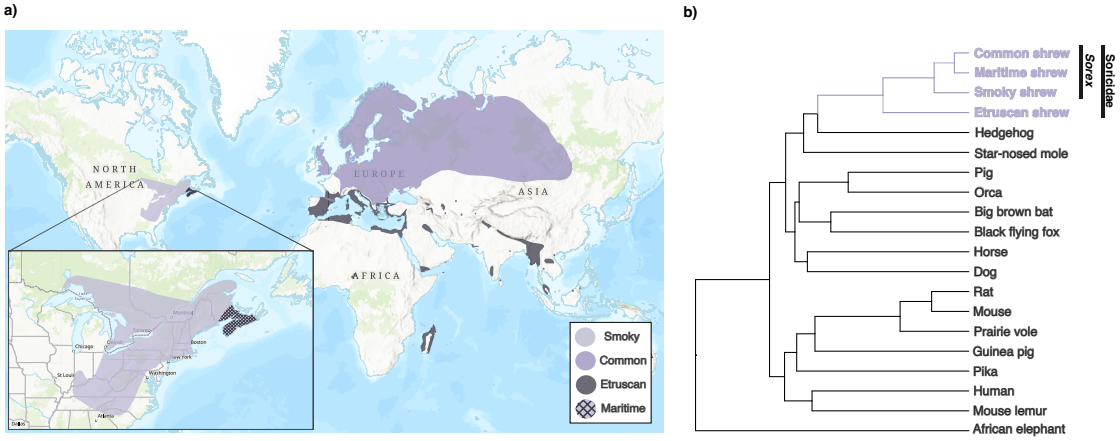

Figure S2.1. (a) Shrew species' range according to the IUCN. (b) Tree topology used for the accelerated region analysis.

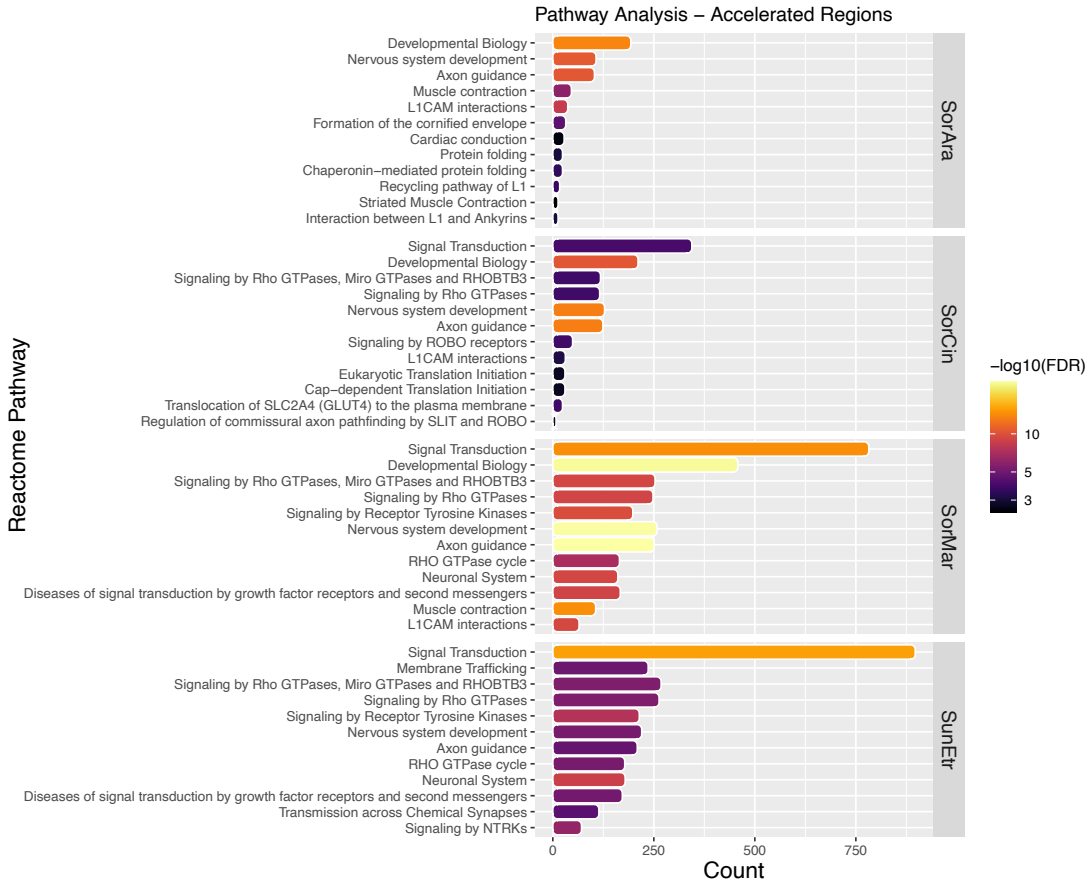

Figure S2.2. Top Reactome pathways for ARs found in each shrew species. All pathways are significant pathways (5% FDR).

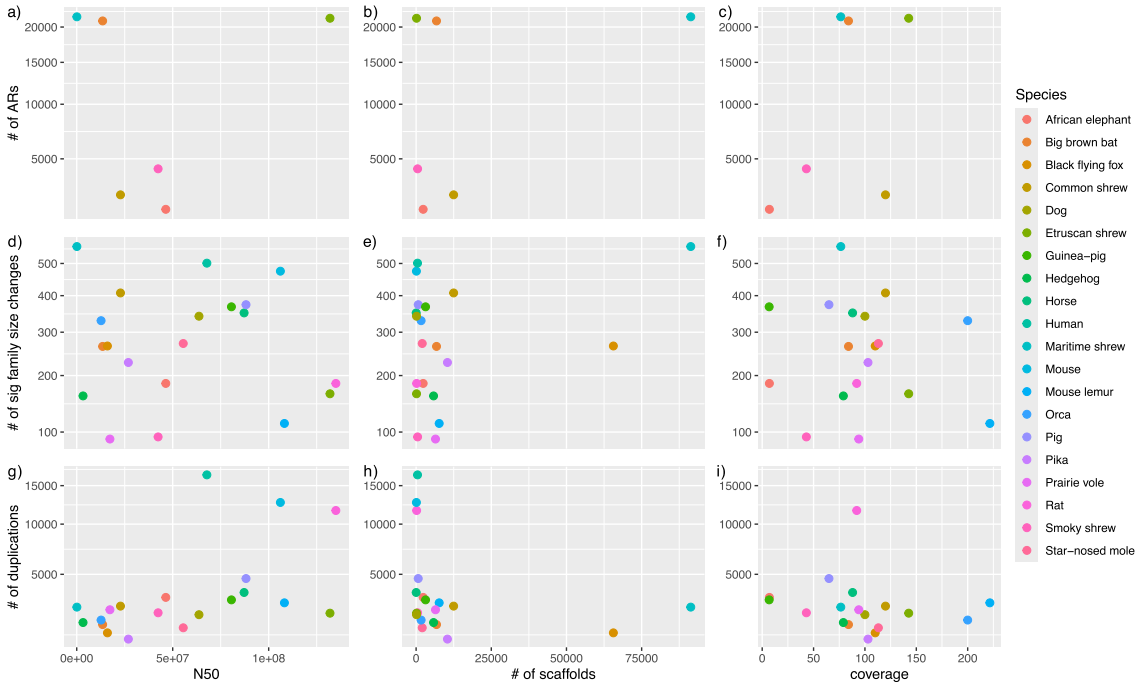

Figure S2.3. Analysis results versus genome assembly statistics for each species. Number of ARs (a) versus N50, (b) number of scaffolds, and (c) coverage. Number of gene families that underwent significant size change (d) versus N50, (e) number of scaffolds, and (f) coverage. Number of gene duplications (g) versus N50, (h) number of scaffolds, and (i) coverage.

Tables

Table S2.1. Conserved and accelerated region statistics aligned to the hg38 chromosomes (NCBI accession: GCA_000001405.15) for all shrew species, and the African elephant and big brown bat.

Chromosome	# 50 bp conserved regions	# ARs S. etruscus	# ARs S. fumeus	# ARs S. maritimensis	# ARs S. araneus	#ARs L. africana	#ARs E. fuscus
1	53,160	1,494	548	2,183	259	67	2,870
2	62,618	2,638	315	1,262	145	362	1,942
3	45,064	1,131	365	1,317	139	177	1,253
4	28,793	962	156	615	104	73	1,266
5	38,505	983	219	1,393	212	38	1,475
6	32,985	1,358	325	871	144	93	422
7	31,500	988	124	939	124	158	915
8	26,899	986	233	512	67	86	1,166
9	29,383	778	154	735	73	178	436
10	26,248	2,080	153	841	126	77	928
11	30,470	1,056	244	1,667	260	65	1,144
12	27,793	1,237	344	1,371	197	28	175
13	18,667	536	121	395	39	190	324
14	24,584	805	80	806	73	10	343
15	20,174	1,143	104	816	111	54	1,146
16	21,988	651	187	1,138	117	70	1,890
17	26,095	978	166	1,600	181	15	228
18	16,292	355	77	348	30	54	1,685
19	13,331	457	142	1,575	126	41	768
20	13,448	421	120	609	70	35	51
21	4,008	103	19	120	14	24	467
22	6,175	214	76	468	32	9	45
TOTAL	598,180	21,354	4,272	21,581	2,643	1,904	20,939

*Table S2.2. Species name, genome version, and NCBI RefSeq accession number of the protein.faa and cds_from_genomic.fna files downloaded and used for the OrthoFinder and aBSREL analyses respectively. The synNet.maf files used for each species for the accelerated region analysis matched the genome versions listed here. *There is no RefSeq annotation file for the maritime shrew. We used the maritime shrew protein.faa file from our GenSAS annotation pipeline (available on gitlab).*

Species	Genome version	NCBI RefSeq accession
Black flying fox (*Pteropus alecto*)	ASM32557v1	GCF_000325575.1
Dog (*Canis lupus familiaris*)	CanFam6	GCF_000002285.5
Domestic guinea pig (*Cavia porcellus*)	Cavpor3.0	GCF_000151735.1
Star-nosed mole (*Condylura cristata*)	ConCri1.0	GCF_000260355.1
Big brown bat (*Eptesicus fuscus*)	EptFus1.0	GCF_000308155.1
Horse (*Equus caballus*)	EquCab3.0	GCF_002863925.1
Western European hedgehog (*Erinaceus europaeus*)	EriEur2.0	GCF_000296755.1
Human (*Homo sapiens*)	GRCh38.p14	GCF_000001405.40
African savanna elephant (*Loxodonta africana*)	Loxafr3.0	GCF_000001905.1
Gray mouse lemur (*Microcebus murinus*)	Mmur_3.0	GCF_000165445.2
Prairie vole (*Microtus ochrogaster*)	MicOch1.0	GCF_000317375.1
House mouse (*Mus musculus*)	GRCm39	GCF_000001635.27
Etruscan shrew (*Suncus etruscus*)	mSunEtr1.pri.cur	GCF_024139225.1
American pika (*Ochotona princeps*)	OchPri3.0	GCF_000292845.1
Killer whale (*Orcinus orca*)	Oorc_1.1	GCF_000331955.2
Norway rat (*Rattus norvegicus*)	mRatBN7.2	GCF_015227675.2
Common shrew (*Sorex araneus*)	SorAra2.0	GCF_000181275.1
Smoky shrew (*Sorex fumeus*)	SorFum_2.1	GCF_029834395.1
Maritime shrew (*Sorex maritimensis*)	SorMar_1.0*	GCA_030324115.1*
Pig (*Sus scrofa*)	Sscrofa11.1	GCF_000003025.6

Appendix II: Supplementary Material for Chapter 3

Figures

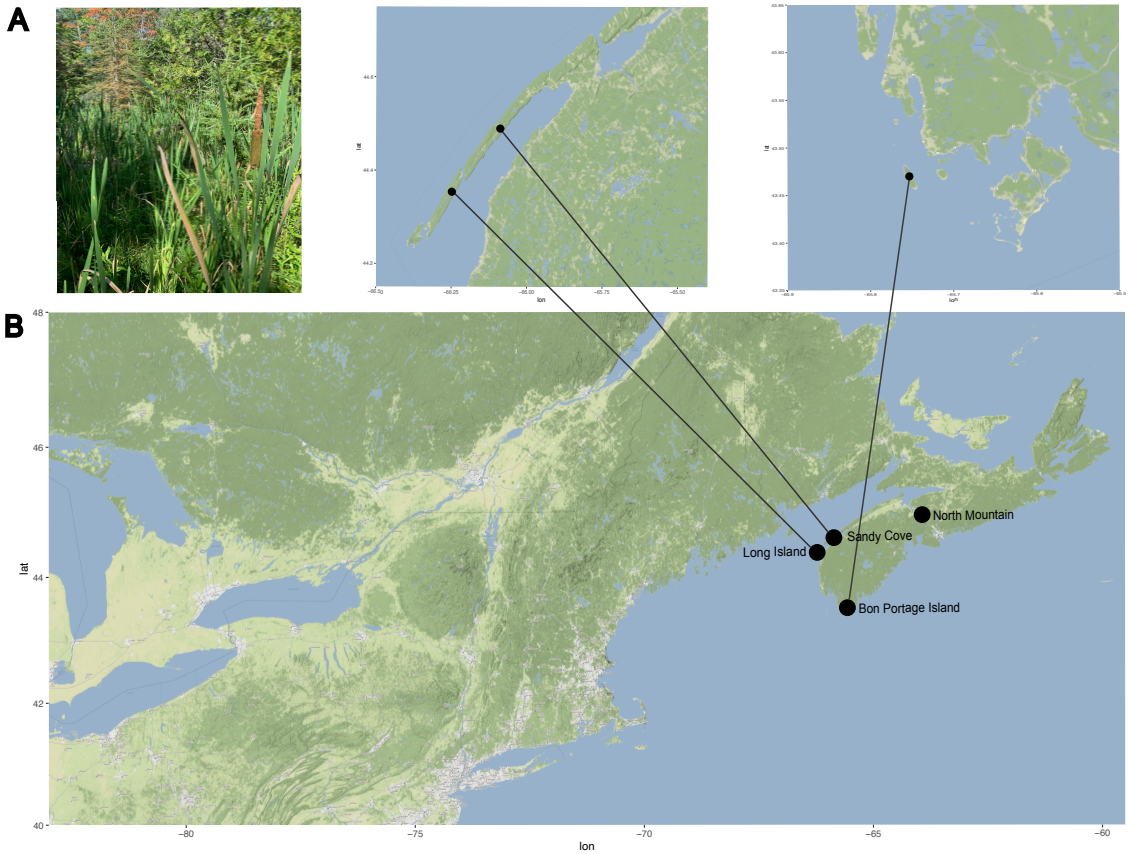

Figure S3.1. Trapping information for our masked shrew samples. (a) Example picture of the type of ecosystem where traps were set. (b) Map of trapping locations in Nova Scotia. Traps in Sandy Cove and North Mountain were set on the mainland. Bon Portage is located ~ 2.5 km from the mainland and Long Island is separated by ~ 750 m of water from the mainland. Both islands require a ferry or boat to reach.

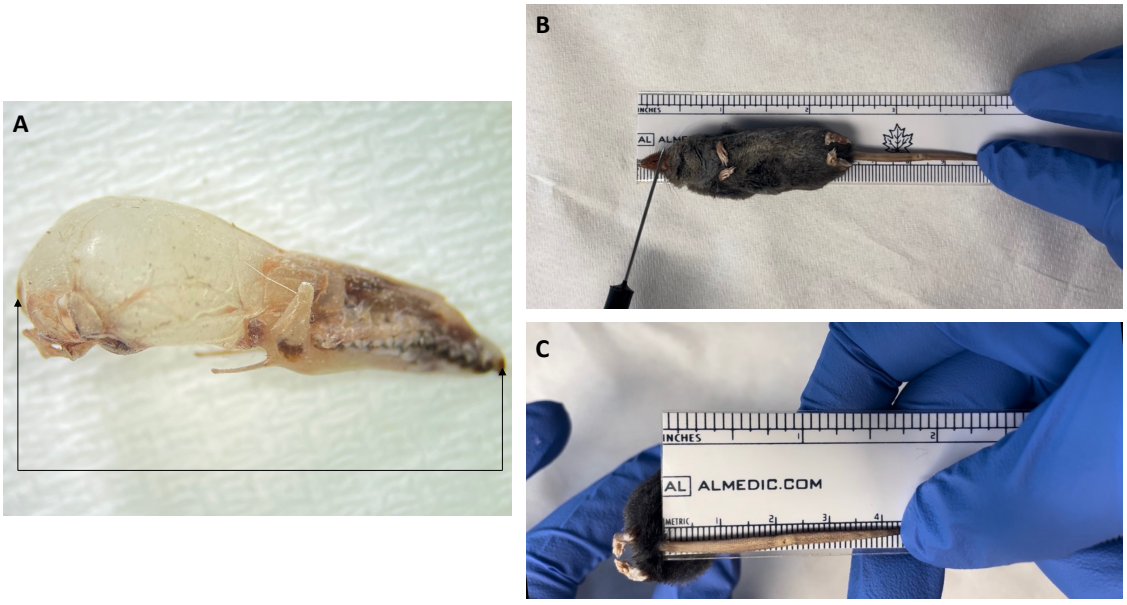

Figure S3.2. Measurements taken for each individual. (a) Full skull length from the most protruding anterior to most protruding backward point of the skull, (b) full body length and (c) tail length.

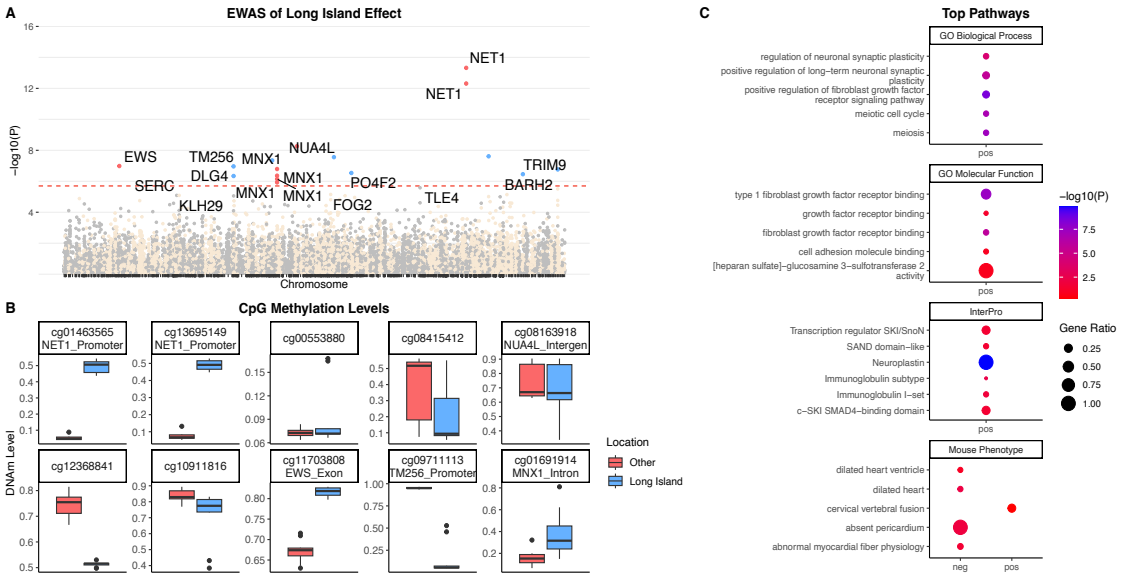

Figure S3.3. Epigenome-wide association (EWAS) of the multivariate regression model between Long Island and other masked shrew populations (location, age, sex and tissue). (a) Manhattan plots of the EWAS for location (Long Island vs others). The coordinates are estimated based on the alignment of Mammalian array probes to our masked shrew genome assembly. The direction of associations with < 2e-6 (red dotted line) is highlighted by red (hypermethylated) and blue (hypomethylated) colors relative to non-Long Island methylation levels. Top 15 CpGs are indicated by their neighboring genes. (b) DNAm levels of Long Island samples versus others for the top 10 significant CpGs (p value) associated to location. (c) Enrichment analysis of the top CpGs with positive (hypermethylated) and negative (hypomethylated) correlations to Long Island. The gene-level enrichment analysis was carried out using the GREAT software. Background probes were limited to 14290 probes that had shrew gene annotations. The top 4 ontologies with most significance pathways for Long Island were selected based on Bonferroni corrected p values.

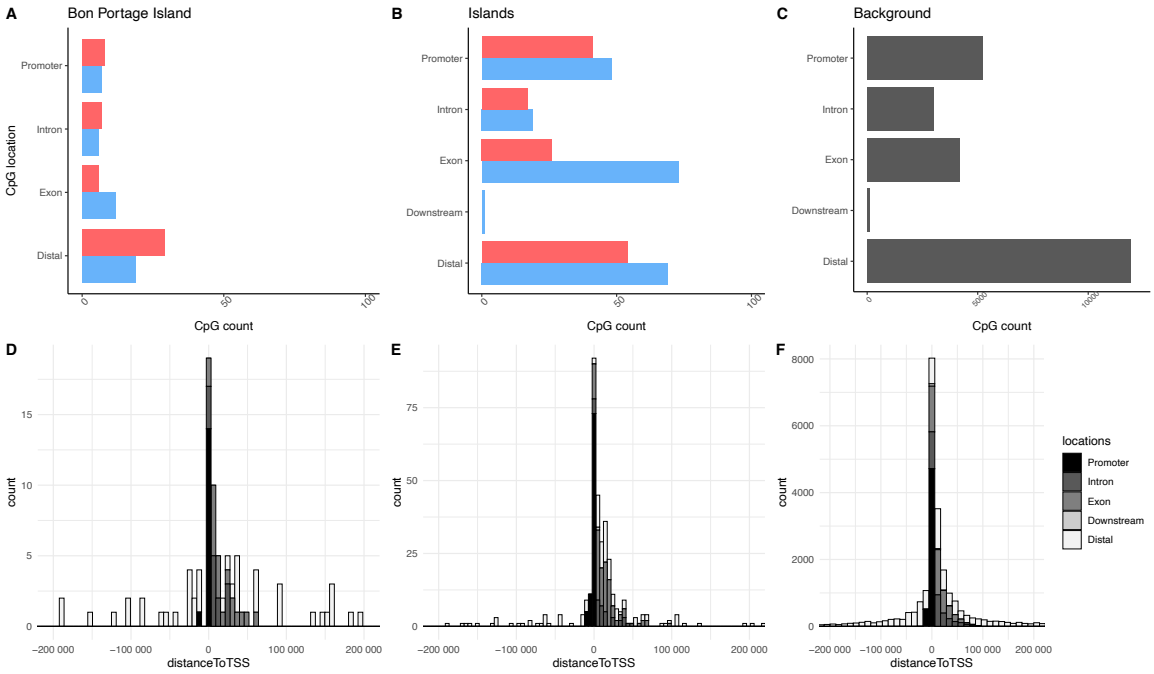

Figure S3.4. Location and distance to transcriptional start site (TSS) of significant (p value < 2e-6) probes aligned to the smoky shrew genome for (a) & (d) Bon Portage Island shrews relative other populations, (b) & (e) Island shrews relative to mainland shrews and (c) & (f) all probes. Blue bars represent hypomethylated CpGs and red bars represent hypermethylated CpGs. Probe locations were identified as intron, exon, downstream up to 300 bp from the transcriptional start site (TSS), promoter up to 10 kb upstream and 1 kb downstream from the TSS, and distal intergenic for probes that were not part of any other category.

Tables

Table S3.1. Master mix formula. Custom primer: 5′-CATGGTGTGGGCTCGCAATC-3 and 5′ CTGCCTGTAGTCTCTGTGCC-3′. PCR cycles as follow: denature at 93 °C for 2 minutes, run 30 cycles of 15 seconds at 93 °C, 20 seconds at 60 °C and 20 seconds at 72 °C.

Reagents	Conc. of stock solution	Volume per individual reaction
Buffer	5X	4 µL
dNTPs	2mM	2 µL
$MgCl_2$	50mM	1 µL
BSA	3mg/mL	1 µL
Primer	10µM	0.8 µL
Taq polymerase	5U/µL	0.2 µL
DNA		2 µL
Water		11 µL
Total volume (before DNA)		20 µL

*Table S3.2. Breakdown of sample information. n = represents the total number of samples used for each analysis. *represents second sample from a same individual.*

SampleID	Tissue	Sex	Teeth class	Age (months)	Location	Used for epigenetic clock (n = 39)	Used for body size analysis (n = 27)	Used for EWAS (n = 36)
1213	Tail	F	2	3.5	Bon Portage Island, NS	yes	yes	yes
1215	Tail	M	1	2	Bon Portage Island, NS	yes	yes	yes
1216	Tail	M	2	3	Bon Portage Island, NS	yes	yes	yes
1217	Tail	F	2	5	Bon Portage Island, NS	yes	yes	yes
1222	Tail	M	3	11	Bon Portage Island, NS	yes	yes	yes
1223	Tail	F	1	5	Bon Portage Island, NS	yes	yes	yes
1225	Tail	F	1	2	Bon Portage Island, NS	yes	yes	yes
1226	Tail	M	3	12	Bon Portage Island, NS	yes	yes	yes
1227	Tail	M	3	12	Bon Portage Island, NS	yes	yes	yes
1228	Tail	M	3	12	Bon Portage Island, NS	yes	yes	yes
1229	Tail	F	3	12	Bon Portage Island, NS	yes	yes	yes
1231	Tail	M	3	11	Bon Portage Island, NS	yes	yes	yes
1232	Tail	F	3	12	Bon Portage Island, NS	yes	yes	yes
1233	Tail	M	3	11	Bon Portage Island, NS	yes	yes	yes
1234	Tail	M	4	16	Bon Portage Island, NS	yes	yes	yes
1235	Tail	M	3	11	Bon Portage Island, NS	yes	yes	yes
1335	Tail	M	2	5	North Mountain, NS	yes	yes	yes
1336	Tail	F	2	5	North Mountain, NS	yes	yes	yes
SC001	Tail	F	2	4.5	Sandy Cove, NS	yes	no	yes
SC002	Tail	F	2	3	Sandy Cove, NS	yes	no	yes
SC003	Tail	F	2	4	Sandy Cove, NS	yes	no	yes
SC004	Tail	M	2	3	Sandy Cove, NS	yes	no	yes
LI001	Tail	M	4	15	Long Island, NS	yes	no	yes
LI002	Tail	F	4	13	Long Island, NS	yes	no	yes
LI003	Tail	F	2	4.5	Long Island, NS	yes	no	yes
LI004	Tail	M	4	13	Long Island, NS	yes	no	yes
LI005	Tail	M	4	15	Long Island, NS	yes	no	yes
SC001	Liver*	F	2	4.5	Sandy Cove, NS	yes	yes	yes
SC002	Liver*	F	2	3	Sandy Cove, NS	yes	yes	yes
SC003	Liver*	F	2	4	Sandy Cove, NS	yes	yes	yes
SC004	Liver*	M	2	3	Sandy Cove, NS	yes	yes	yes
LI001	Liver*	M	4	15	Long Island, NS	yes	yes	yes
LI002	Liver*	F	4	13	Long Island, NS	yes	yes	yes
LI003	Liver*	F	2	4.5	Long Island, NS	yes	yes	yes
LI004	Liver*	M	4	13	Long Island, NS	yes	yes	yes
LI005	Liver*	M	4	15	Long Island, NS	yes	yes	yes
LI002_FET1	Fetus	F	0	0	Long Island, NS	yes	no	no
LI002_FET2	Fetus	M	0	0	Long Island, NS	yes	no	no
LI002_FET3	Fetus	F	0	0	Long Island, NS	no	no	no
LI002_FET4	Fetus	M	0	0	Long Island, NS	yes	no	no

Table S3.3. Assembly statistics for smoky shrew genome assemblies.

	Short-read assembly	Linked-read assembly	Hybrid assembly
Assembly size	1.92 Gb	2.66 Gb	2.66 Gb
N50/L50	2.77 Kb/175,854 scaffolds	1.99 Mb/291 scaffolds	1.99 Mb/291 scaffolds
N90/L90	1.23 Kb/607,012 scaffolds	0.16 Mb/1,755 scaffolds	0.16 Mb/1,753 scaffolds
Number of scaffolds	779,274	84,561	84,558
Largest scaffold	91,699	22,233,027	22,233,027
% Genome in scaffolds > 50 Kb	0.03	86.15	86.14
GC content (%)	42.63	42.88	42.88
BUSCOs identified			3,879 (95%)
Number of protein coding genes			20,067

Table S3.4 Summary of repeats in the smoky shrew genome.

	Length (bp)	Percentage of genome (%)
SINE	53,159,155	2.10
LINE	509,850,416	20.17
LTR elements	2,632,939	0.10
DNA elements	665,221	0.03
Unclassified	389,117,414	15.39
Small RNA	0	0
Satellites	0	0
Simple repeats	20,505,599	0.81
Low complexity	0	0
Total	975,930,744	38.60

Appendix III: Supplementary Material for Chapter 4

Figures

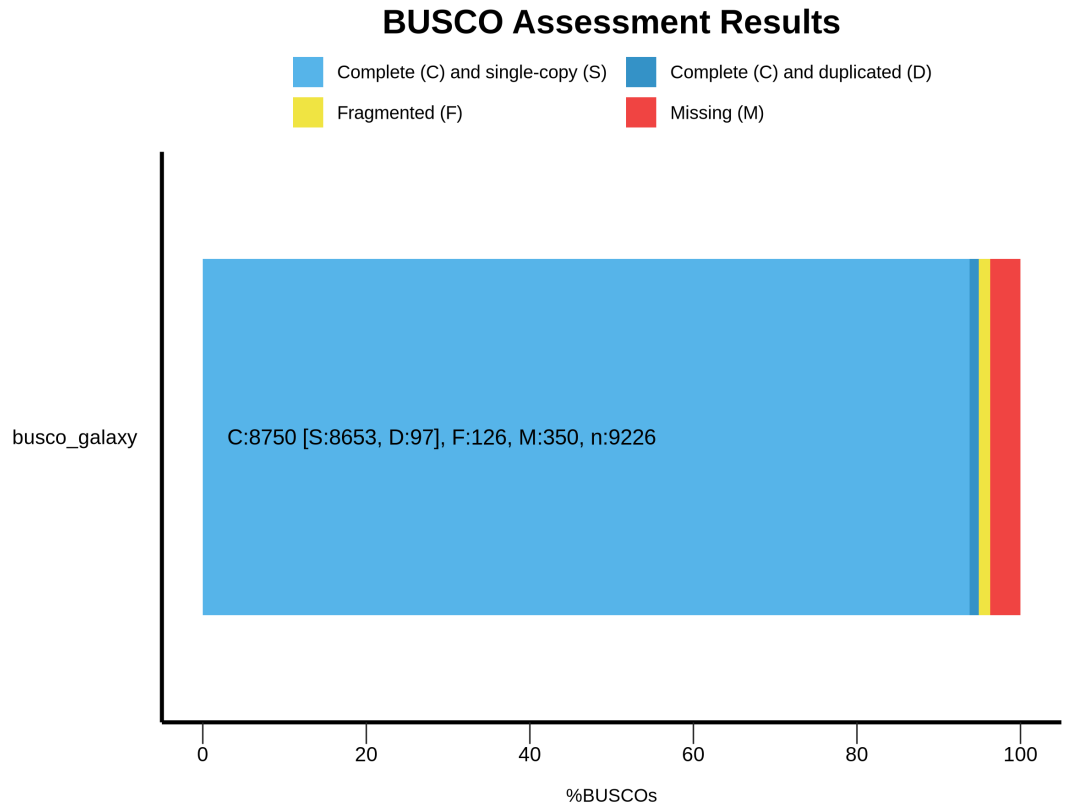

Figure S4.1. Primary assembly BUSCO assessment results for the mammalia_odb10 data set.

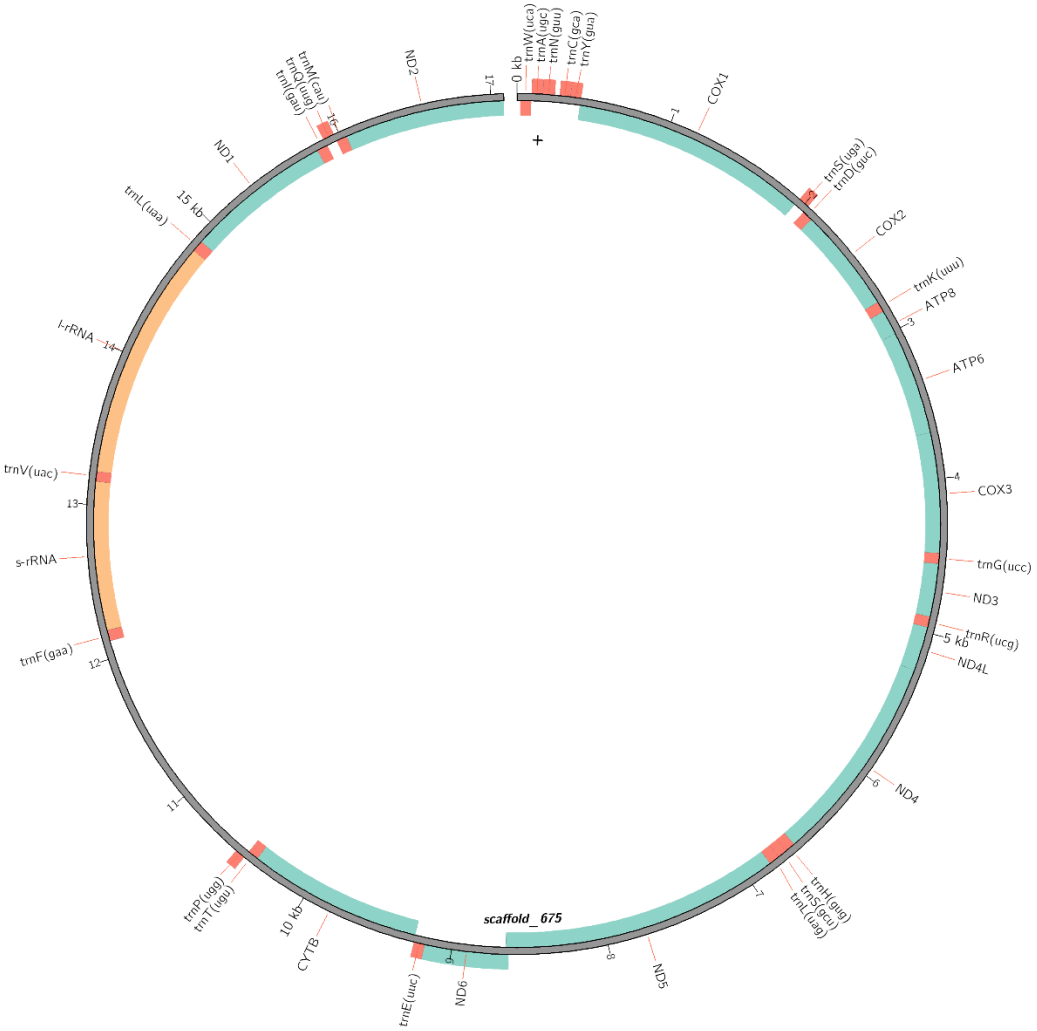

Figure S4.2. Masked shrew mitochondrial genome annotations Circos plot.

Tables

*Table S4.1. Sample information. * Denotes smoky shrew samples.*

Sample ID	PCA ID	*Species*	Sex	Location	WGS Mean Coverage	*RNA-seq tissue*	From Cossette et al. 2023
1223	BPI1	*S. cinereus*	F	BPI	6.7X	*NA*	yes
1228	BPI2	*S. cinereus*	M	BPI	8.1X	*NA*	yes
1229	BPI3	*S. cinereus*	F	BPI	6.6X	*NA*	yes
LI002	LI1	*S. cinereus*	F	LI	6.7X	*NA*	yes
LI003	LI2	*S. cinereus*	F	LI	6.9X	*NA*	yes
LI004	LI3	*S. cinereus*	M	LI	7.7X	*NA*	yes
LI005	LI4	*S. cinereus*	M	LI	8.0X	*NA*	yes
NS1	BPI4	*S. cinereus*	M	BPI	9.2X	Liver	no
NS2	BPI5	*S. cinereus*	M	BPI	**14.7X**	Heart, Liver	no
NS3	BPI6	*S. cinereus*	M	BPI	**17.7X**	Liver	no
NS4	BPI7	*S. cinereus*	M	BPI	**14.5X**	Liver	no
NS5	BPI8	*S. cinereus*	F	BPI	**14.1X**	Liver	no
NS6	BPI9	*S. cinereus*	M	BPI	6.2X	Liver	no
NS7	BPI10	*S. cinereus*	M	BPI	**14.8X**	Liver	no
NS8	SC6	*S. cinereus*	F	SC	7.5X	Liver	no
NS10	SC1	*S. cinereus*	M	SC	**13.1X**	Liver	no
NS11	SC2	*S. cinereus*	M	SC	**15.8X**	Liver	no
NS12	LI5	*S. cinereus*	M	LI	6.9X	Heart, Liver	no
NS13	SC3	*S. cinereus*	M	SC	6.5X	Heart, Liver	no
NS14	SC4	*S. cinereus*	M	SC	7.3X	Liver	no
NS15	SC5	*S. cinereus*	M	SC	**15.7X**	Liver	no
ON4	ON1	*S. cinereus*	*NA*	ON	6.5X	Heart, Liver	no
ON5	ON2	*S. cinereus*	*NA*	ON	8.6X	Heart, Liver	no
ON6	ON3	*S. cinereus*	*NA*	ON	**13.7X**	Liver	no
NS9*	*NA*	*S. fumeus*	*NA*	SC	**14.6X**	*NA*	no
SC2*	*NA*	*S. fumeus*	M	ON	6.3X	*NA*	yes

Table S4.2. F-statistics done with genome N90 (72,716,233 SNPs).

f			
pop1	**pop2**	**est**	**se**
BPI	LI	0.43055543	0.00240651
BPI	ON	0.38777294	0.00190862
LI	ON	0.32865797	0.001784
BPI	SC	0.28038394	0.00180212
SC	ON	0.22242853	0.00174191
LI	SC	0.18898498	0.00124119

f_3 **(A; B, C)**							
pop1 (A)	**pop2 (B)**	**pop3 (C)**	**est**	**se**	**z**	**p**	**n**
BPI	SC	LI	0.59218144	0.0177494	33.3634629	4.65E-244	10384226
SC	BPI	LI	0.02152528	0.00260015	8.27847868	1.25E-01	10384226
LI	BPI	SC	0.28684796	0.00723655	39.6387631	0	10384226

f_4 **(ON, B; C, D)**							
pop1	**pop2 (B)**	**pop3 (C)**	**pop4 (D)**	**est**	**se**	**z**	**p**
ON	SC	LI	BPI	-5.59E-04	1.96E-04	-2.8502261	0.00436882
ON	LI	SC	BPI	-2.02E-04	2.17E-04	-0.930139	0.3522991
ON	BPI	SC	LI	3.57E-04	2.15E-04	1.66114207	0.09668492

Table S4.3. Significant DEGs between BPI vs SC & LI.

baseMean	**log2FoldChange**	**lfcSE**	**stat**	**pvalue**	**padj**	**gene name**	**uniprot accession**
120.4144107	1.798886667	0.44432824	4.048553537	5.15E-05	0.005422961	8ODP	P36639
41.44717994	-2.020878393	0.393075766	-5.141193044	2.73E-07	8.66E-05	A1AG1	P02763
146.6957401	-3.424308341	0.81031377	-4.225904171	2.38E-05	0.003122915	ABCA9	Q8K449
457.2754087	0.711316162	0.218667858	3.252952534	0.001142126	0.039156392	ABCG5	Q99PE7
95.40289497	1.513206457	0.460814967	3.283761521	0.001024316	0.036423981	ACO11	Q8VHQ9
978.2074655	-1.901143603	0.554298217	-3.429820886	0.00060398	0.026267945	ACOT4	Q8BWN8
363.5283868	-2.006625001	0.476984263	-4.206899804	2.59E-05	0.003284119	ACTA	P62738
122.0230527	-2.599971454	0.806667647	-3.22310119	0.001268107	0.042607747	ADAM8	Q05910
150.8362839	-2.590801457	0.413281845	-6.268848937	3.64E-10	3.46E-07	ADH6	P28332
58.36291782	-2.092445614	0.56760679	-3.686435135	0.000227417	0.014423947	ADRB2	O70431
154.2630261	-1.0674719	0.33798496	-3.15834142	0.001586696	0.049493213	AFAP1	Q80YS6
452.8479846	0.975721648	0.288812034	3.378396786	0.000729098	0.029834218	AIG1	Q9NVV5
5397.020754	-3.166600974	0.935920313	-3.383408746	0.00071592	0.029453334	AK1CA	D3ZF77
119.1145539	-2.028779093	0.509133759	-3.984766396	6.75E-05	0.006590984	AL1A3	P47895
2940.373276	-2.017214192	0.521006037	-3.871767404	0.000108049	0.008477951	ANXA2	A2SW69
683.060662	-1.646566638	0.487728488	-3.375990288	0.000735505	0.029935453	ANXA4	P08132

52.07466365	-2.5199011	0.702492826	-3.587084463	0.000334396	0.018058984	AP1M2	Q3SYW1
17.00739363	-3.078842631	0.91966329	-3.347793332	0.000814577	0.032054351	APCD1	Q8J025
421.9802635	0.68513605	0.184063376	3.72228341	0.000197429	0.012843025	APLP2	P15943
38.72344369	-1.397743245	0.427052875	-3.272998094	0.001064132	0.03732308	ARSK	Q6UWY0
116.4463356	0.728360922	0.219586349	3.316968129	0.00091	0.034118275	ASAP3	Q8TDY4
28.61898379	-2.443323778	0.680994727	-3.587874738	0.000333384	0.018058984	ASGL1	Q4R7U8
1180.643391	0.989465476	0.285281424	3.468383826	0.000523599	0.024063061	ATP7A	Q04656
225.47261	2.70563131	0.576571549	4.692620219	2.70E-06	0.000632395	ATRAP	Q6RW13
20.13694281	-3.566228459	0.866737818	-4.114541196	3.88E-05	0.004513839	ATS5	Q9TT92
20.76659179	-1.586511444	0.475267431	-3.338144677	0.000843398	0.032584284	AURKB	O70126
9.105099293	-2.573629379	0.777036334	-3.312109441	0.000925953	0.034546223	B3GL1	Q864U6
54.49358724	-1.682448079	0.481052417	-3.497431919	0.000469761	0.022628784	B3GN5	Q9BYG0
29.90270027	-2.181178013	0.677406803	-3.219893872	0.001282381	0.042607747	BMP3	Q8BHE5
60.69121731	-2.316753482	0.354395858	-6.537191199	6.27E-11	7.95E-08	CAD17	P55281
14.17780895	-3.08882313	0.871432105	-3.54453676	0.000393304	0.020790277	CAHM5	Q8N5C1
33.92775619	-3.658828943	1.130646382	-3.236050637	0.001211959	0.040996527	CAN6	Q9Y6Q1
225.6092528	0.71611974	0.173020717	4.138924816	3.49E-05	0.004353707	CC28B	Q9BUN5
111.9711997	-1.348138618	0.410451912	-3.284522686	0.001021553	0.036423981	CCBE1	Q3MI99
159.3834339	-2.148382462	0.316544296	-6.786988392	1.14E-11	2.18E-08	CCL11	Q9TTQ4
187.1732652	-2.448301654	0.744892443	-3.286785462	0.00101338	0.0363813	CD44	P14745
20.64045551	-2.821068917	0.848957646	-3.322979573	0.000890614	0.033556762	CD80	P33681
49.44964473	-2.010161491	0.516626963	-3.890934145	9.99E-05	0.008085395	CDKA1	O14519
100.0737132	1.035033532	0.222822565	4.645101959	3.40E-06	0.000718623	CEP63	Q4KLY0
28.70779308	-4.942501753	0.908016634	-5.443184151	5.23E-08	2.21E-05	CES1P	Q9UKY3
672.8716223	-1.021755046	0.263290366	-3.880715664	0.00010415	0.008257106	CISD3	P0C7P0
472.4640366	-0.676753299	0.154365406	-4.384099508	1.16E-05	0.001886014	CIZ1	Q9ULV3
507.212004	-1.547607599	0.418978522	-3.693763562	0.000220959	0.014132109	CML2	O97664
2.445403488	-3.270659061	1.005916603	-3.251421689	0.001148294	0.039191327	CMTD1	Q86VU5
40.14313032	-4.308865735	1.121803205	-3.841017492	0.000122525	0.009515721	CO3A1	P02461
7.080449592	-2.535161382	0.716625383	-3.537638271	0.000403723	0.021046122	CP2BJ	O55071
17.17540866	-3.350089517	0.77837928	-4.303929464	1.68E-05	0.002494998	CP2D7	A0A087X1C5
136.2436185	-0.822192115	0.191285248	-4.298251565	1.72E-05	0.002494998	CPSF1	Q10569
476.5106875	0.580383075	0.14864375	3.904523913	9.44E-05	0.007726479	CS025	Q1LZF3
447.3588086	-1.063391849	0.291688985	-3.645635954	0.000266731	0.015860096	CSRP1	P97315
26.57328727	-8.106370918	1.986251177	-4.081241593	4.48E-05	0.005024614	CTHR1	Q96CG8
515.9099839	1.016912909	0.293655021	3.462950868	0.000534286	0.024205062	CTNS	A7MB63
41.61974933	-1.529425095	0.435477729	-3.512062715	0.000444643	0.021975184	CYBR1	Q925G2
968.5891668	-1.240010544	0.313405291	-3.956571828	7.60E-05	0.00698095	CYL1B	A0A8V8TMC4
144.1872845	1.240928418	0.316621001	3.919286514	8.88E-05	0.007521562	CYREN	Q09HN1
68.63383866	-1.485527072	0.454549883	-3.268127717	0.001082615	0.037797169	DAAM2	Q80U19

105.7269586	0.784326836	0.194368963	4.035247312	5.45E-05	0.005610008	DC4L1	Q3SXM0
163.6437804	-2.705488707	0.592284076	-4.567890339	4.93E-06	0.000937404	DDR1	Q08345
27.48899581	1.752367709	0.407282385	4.30258654	1.69E-05	0.002494998	DDT4L	Q8VD50
141.6701572	-1.241717733	0.365317891	-3.399006083	0.000676312	0.028282481	DEN2B	P78524
235.3677405	0.711815971	0.211192449	3.370461272	0.000750425	0.030349805	DI3L1	Q5R5N8
786.3020842	0.666925186	0.13298963	5.014866106	5.31E-07	0.0001496	DJB14	Q0IIE8
1559.730323	0.780790404	0.146579542	5.326735181	1.00E-07	3.81E-05	DMP34	Q6PK57
79.26722489	3.538985312	0.60322559	5.866769204	4.44E-09	3.38E-06	DNSL3	Q13609
71.80603183	-2.29053732	0.405074718	-5.654604496	1.56E-08	9.91E-06	DOC10	Q8BZN6
380.752252	0.851944717	0.265885516	3.204178736	0.001354484	0.044427262	DOK4	Q8TEW6
124.1338249	-2.676780156	0.65595422	-4.080742337	4.49E-05	0.005024614	DPYL3	Q14195
830.8777582	-1.437099435	0.428945962	-3.350304145	0.000807229	0.031999051	DUS14	O95147
9.376352262	-3.047795247	0.876707884	-3.476409079	0.000508177	0.023583728	EFCB6	Q4R8T1
138.0039357	-2.469703968	0.52667251	-4.689259309	2.74E-06	0.000632395	EMP1	P54850
450.1701284	-0.931871931	0.211646173	-4.402970849	1.07E-05	0.00176672	ERO1A	Q8R4A1
10.50217554	1.854096731	0.487715807	3.801592453	0.000143769	0.010074666	FANCF	E9Q5Z5
21.77407149	-2.476277278	0.774303113	-3.198072223	0.001383496	0.044999103	FBLN5	Q5EA62
57.96541333	0.995500139	0.313036298	3.180142833	0.001472025	0.046876906	FBW12	Q6X9E4
475.8769714	-4.0014153	0.57995229	-6.899559446	5.22E-12	1.32E-08	FCN2	P57756
678.9694095	0.736870965	0.184351562	3.997096407	6.41E-05	0.006424506	FCSD2	O94868
9.27558237	-2.137105232	0.654761887	-3.263942626	0.001098734	0.038011207	FEZ1	P97577
43.52462789	-3.069237516	0.579942815	-5.292310613	1.21E-07	4.38E-05	FHL2	O35115
124.1144492	-1.076117756	0.338174121	-3.182141058	0.001461906	0.046750279	FRMD6	Q96NE9
134.0657464	-2.12402774	0.498463222	-4.261152374	2.03E-05	0.002764091	FXYD5	P59647
108.6296969	-1.743370724	0.498282039	-3.498762922	0.000467422	0.022628784	GBP6	Q5R9T9
5.985675213	-3.58701365	0.986272895	-3.636938285	0.000275898	0.015979469	GFPT2	Q08DQ2
17.84011875	-1.282894696	0.365504174	-3.509931721	0.000448222	0.021981309	GLCTK	Q8IVS8
19.66329683	-2.231432926	0.620087693	-3.598576379	0.000319964	0.017770117	GLT16	Q9JJ61
637.7111218	-1.573939001	0.330492185	-4.762409136	1.91E-06	0.000485316	GNB5	Q6PNB6
222.3952196	-1.660864249	0.511852524	-3.244810119	0.00117529	0.039933614	GPC3	P51654
186.8673825	-0.729529556	0.206597476	-3.531163935	0.000413735	0.021219542	GPR39	O43194
14.689465	4.73030789	0.986512822	4.79497862	1.63E-06	0.000426983	GUC2D	Q02846
8.784240836	-2.141915547	0.574798629	-3.726375535	0.000194253	0.012745336	H31	Q6LED0
69.68456508	-2.080335213	0.545359742	-3.814610893	0.000136398	0.009800783	HA19	P14431
274.0187994	0.447019109	0.137151002	3.25932077	0.001116793	0.038461145	HAUS2	Q9NVX0
71.24464793	-2.044325223	0.536271248	-3.812110441	0.000137785	0.009800783	HCDH	P00348
2863.047576	0.537511139	0.166956236	3.219473267	0.001284263	0.042607747	HECD3	Q5T447
314.6138479	-1.077140026	0.269922098	-3.99055888	6.59E-05	0.006515587	HID1	Q8IV36
151.2756978	-4.149878629	0.803424275	-5.165239289	2.40E-07	7.95E-05	HLAF	P33617
219.6936953	-1.182523893	0.313857546	-3.76770898	0.000164753	0.011096744	HMGN1	P02316

4.203400236	-3.699264002	1.006417112	-3.675676771	0.00023722	0.014866517	HOGA1	Q86XE5
31.78285869	-1.932951618	0.448407884	-4.310699445	1.63E-05	0.002494998	HSPB1	P04792
35.35460568	-1.323000535	0.336065112	-3.936738711	8.26E-05	0.007337254	HSPB6	Q148F8
17.7336575	-2.456142693	0.642264553	-3.824191575	0.000131202	0.009800783	HVCN1	Q96D96
252.1911204	-2.427038818	0.52655217	-4.609303614	4.04E-06	0.00083108	IEX1	Q7YR42
195.0276273	-1.704694449	0.46639225	-3.655065981	0.000257116	0.015655269	IFI44	Q8BV66
92.03298525	1.528042634	0.409312578	3.73319247	0.000189068	0.012513014	IFIT2	P09913
45.81345503	-2.257133456	0.678924286	-3.324573154	0.00088554	0.033531566	IRF6	P97431
10.96289395	-1.928909989	0.587785817	-3.28165453	0.001032	0.036423981	ITA2	P53710
168.3766447	-1.242471378	0.364680572	-3.407012804	0.00065678	0.027666341	ITA2B	P53711
201.3178161	1.454398068	0.34283891	4.242219962	2.21E-05	0.002955199	ITB4	P16144
27.25368224	-2.61293802	0.549676888	-4.753589023	2.00E-06	0.000490632	KCJ15	Q99712
62.6838154	-1.463456733	0.448285893	-3.264561201	0.001096338	0.038011207	KCNN4	O15554
6733.296323	0.865764695	0.261663808	3.30869103	0.000937332	0.034800171	LACB2	P21664
39.61450958	-3.040520545	0.708405232	-4.292063928	1.77E-05	0.002494998	LEPR	P48356
49.31384733	-2.038605196	0.544630185	-3.743099909	0.000181764	0.012135133	LHPL6	Q8BM86
20.40463872	-2.769375551	0.703630441	-3.935838174	8.29E-05	0.007337254	LOX5	P09917
89.97524915	-2.792718059	0.497523392	-5.613239701	1.99E-08	1.06E-05	LRCC1	Q69ZB0
323.6022259	-2.962897759	0.530260197	-5.58762995	2.30E-08	1.09E-05	LYOX	P45845
54.17824706	-2.553360012	0.67361372	-3.790540391	0.00015032	0.010400769	M4A6A	Q9H2W1
14.64430522	1.546777661	0.480554762	3.218733398	0.001287581	0.042607747	MALL	Q91X49
195.0222564	2.07676146	0.457461595	4.539750407	5.63E-06	0.001045507	MATN1	P51942
77.55815037	-3.88778446	0.984660789	-3.948349019	7.87E-05	0.007130059	MEDAG	Q5VYS4
363.8654597	-1.814103261	0.504371061	-3.59676318	0.000322202	0.017770117	MFAP4	P55083
13.25882129	-3.431917262	0.750363743	-4.573671493	4.79E-06	0.000935277	MIME	Q62000
88.43848375	0.652883344	0.185132994	3.526563957	0.000420989	0.021219542	MLEC	Q14165
9.469055572	-2.172469254	0.640939545	-3.389507282	0.000700183	0.028962481	MLRT	P24732
123.418015	-2.000339678	0.630042684	-3.174927237	0.001498741	0.047528817	MMP14	Q5RES1
52.7990586	1.621173511	0.492363655	3.292634409	0.000992534	0.035801799	MOC2B	A4FUY7
29.14304196	-2.743282955	0.770944978	-3.558338189	0.000373209	0.020003455	MS4A3	Q920C4
393.582049	-3.988390381	1.003845769	-3.973110714	7.09E-05	0.006834491	MSTN1	Q32KU9
161.1732773	-1.249224691	0.353889085	-3.529989323	0.000415576	0.021219542	MYL9	P29269
104.1610346	3.078563621	0.506154671	6.082258635	1.19E-09	1.00E-06	MYRIP	Q8NFW9
40.7534011	-1.612090982	0.472391217	-3.412618446	0.00064342	0.02751161	N4BP3	O15049
27.98635293	1.477530906	0.421121999	3.508557874	0.000450543	0.021981309	NDUS7	P0CB83
29.05564121	-1.357923041	0.395318673	-3.435008603	0.000592535	0.026068116	NDUV2	P04394
43.21095332	-0.98437671	0.269672051	-3.650273383	0.000261961	0.015823713	NEDD1	P33215
31.99977035	-4.018644071	1.052309128	-3.8188817	0.000134058	0.009800783	NELL2	Q5R3Z7
442.561466	-0.955583031	0.273443358	-3.494628789	0.000474722	0.022723934	NOS2	O19114
760.7808539	-1.431919941	0.323723754	-4.423277324	9.72E-06	0.001644227	NT2NA	Q7Z3S9

9.050476683	-2.320332713	0.696368642	-3.332046524	0.000862098	0.03287206	OSR1	Q08DS3
46.9561796	-1.435116637	0.387388607	-3.704591752	0.000211731	0.013656679	OX2G	Q5RAL8
22.02091523	-2.767740944	0.72821715	-3.800708268	0.000144283	0.010074666	PA2G5	P97391
1438.468236	-3.641240728	1.032453659	-3.526783692	0.00042064	0.021219542	PA2GA	P14555
50.85701262	1.258328859	0.382035733	3.293746499	0.000988616	0.035801799	PACRG	A5PK71
41.49806582	-3.463391765	0.959606558	-3.609178926	0.000307168	0.017317432	PAI1	P13909
26.56140635	-2.87342556	0.83752907	-3.430836806	0.000601722	0.026267945	PAMR1	Q5RDI1
174.5072831	-1.8741588	0.455104073	-4.118088391	3.82E-05	0.004513839	PARPT	Q7Z3E1
175.0962037	-0.87262829	0.231081708	-3.776275927	0.000159191	0.010915315	PCGF2	P35227
498.8703338	-0.500365219	0.15835692	-3.159730677	0.00157915	0.04946055	PCNA	P61258
93.71379456	-7.091066697	1.651796854	-4.292941157	1.76E-05	0.002494998	PCOC2	Q8R4W6
361.0648122	-1.910165138	0.357711725	-5.339956748	9.30E-08	3.72E-05	PD1L1	Q9NZQ7
52.37750479	-1.42032294	0.389350323	-3.647930554	0.000264361	0.01584293	PDE5A	O54735
4.974568106	-4.023177002	1.104915267	-3.641163374	0.000271409	0.015889943	PDK4	O54937
68.03310207	1.853461833	0.397911731	4.657972337	3.19E-06	0.000694426	PECR	Q99MZ7
5651.730736	-1.634540512	0.498155239	-3.281187035	0.001033712	0.036423981	PGDH	Q3T0C2
207.4992211	-1.089069219	0.303619123	-3.586958576	0.000334557	0.018058984	PGH1	P05979
557.7149405	0.801594482	0.235101331	3.409570157	0.000650653	0.027665492	PHLB3	Q6NSJ2
363.4838546	-1.332506935	0.301240612	-4.423397385	9.72E-06	0.001644227	PIGR	O70570
863.5055738	1.125531198	0.240598182	4.678053621	2.90E-06	0.000648302	PLIN3	Q5RAV8
22.30352179	-2.693033649	0.662144847	-4.067136761	4.76E-05	0.005174862	PLM	O00168
100.5842635	-1.74926014	0.447128855	-3.912205894	9.15E-05	0.007566069	PLMN	P00747
1092.876818	1.001053493	0.276946586	3.614608534	0.000300802	0.017085092	PON1	P27169
435.7493392	-1.058431207	0.272242982	-3.887818158	0.000101149	0.008103661	PON2	Q15165
8.005384655	-3.654611366	0.807592865	-4.525314082	6.03E-06	0.001092829	POPD1	B8Q0B2
258.3898476	0.486642161	0.127311651	3.82244796	0.000132133	0.009800783	PP1R8	Q12972
6.017053981	-2.458979167	0.735648367	-3.342601271	0.000829971	0.032348001	PP4RL	Q9P1A2
287.004424	3.836686234	0.581405531	6.598984758	4.14E-11	6.30E-08	PPIP1	O43586
38.12598202	3.892177404	0.449132651	8.665986306	4.48E-18	3.41E-14	PROD	O43272
47.93017644	-1.60772061	0.479598946	-3.352218814	0.000801666	0.031944933	PRS23	Q1LZE9
1083.71542	0.775798527	0.235123102	3.299541893	0.000968428	0.035436074	PRUN1	Q6AYG3
8.101517755	-1.839711796	0.574372905	-3.202991959	0.001360078	0.044427262	PSMD5	Q0P5A6
97.63627922	1.302878619	0.359611652	3.623015584	0.000291188	0.016663413	PTAFR	P25105
24.89269187	-3.717013871	1.124578149	-3.305251729	0.000948911	0.034889681	PVR	P32506
10.37214258	1.794840996	0.458126566	3.917784141	8.94E-05	0.007521562	PXMP2	Q2KIY1
4597.957248	-1.077293492	0.295676139	-3.643491477	0.000268965	0.015868909	QSOX2	Q3TMX7
300.7541906	1.472026363	0.35794094	4.112483926	3.91E-05	0.004513839	RAB10	Q5R5U1
57.5308862	-1.028666429	0.322639916	-3.188280116	0.001431218	0.046156786	RALYL	Q8BTF8
81.65412278	-1.886462764	0.514272437	-3.668216746	0.000244248	0.014991711	RASF2	P50749
13.02142896	-1.828824604	0.527773765	-3.465167702	0.000529901	0.024150156	RBMS3	Q6XE24

34.36321866	-2.573998448	0.456365995	-5.64020649	1.70E-08	9.94E-06	RCAN2	Q5RE25
93.52775287	-2.00591007	0.602104817	-3.331496469	0.000863804	0.03287206	RCN3	Q96D15
384.6268184	2.135479964	0.534271634	3.996992969	6.42E-05	0.006424506	RDHE2	Q8N3Y7
75.72559011	-1.435497108	0.426055145	-3.369275377	0.000753661	0.030349805	RECK	O95980
15.80585317	-6.099090934	1.725885653	-3.533890511	0.000409491	0.021201588	REG4	Q9BYZ8
176.6499448	-2.756014808	0.833507603	-3.306526296	0.000944605	0.034889681	RGS2	O08849
859.9552106	-1.349567666	0.3675818	-3.671475751	0.000241154	0.014922137	RIPK2	O43353
497.2556107	0.708470953	0.207974383	3.406529889	0.000657943	0.027666341	RM13	Q3SYS1
40.02090968	-1.437313545	0.314115697	-4.575745678	4.75E-06	0.000935277	RM47	Q08DT6
82.1474207	-1.459471487	0.382828336	-3.812339244	0.000137658	0.009800783	RN150	Q9ULK6
12.97216347	-1.888933937	0.586256811	-3.22202472	0.001272882	0.042607747	RNS1A	Q8VD93
510.235929	0.64163242	0.182282549	3.519988191	0.000431566	0.021609532	RT30	P82924
20.1538306	-2.841578879	0.70073844	-4.055120596	5.01E-05	0.005371484	RTL4	Q6ZR62
431.7846497	-1.138560787	0.266876382	-4.266247831	1.99E-05	0.002750864	S14L3	Q9UDX4
1224.303678	0.710105528	0.2078742	3.416034934	0.000635401	0.027322241	S19A2	O60779
142.8385857	0.941987424	0.271641972	3.467753592	0.000524828	0.024063061	S53A1	Q9TU72
1235.415372	2.649281891	0.477627868	5.546748978	2.91E-08	1.30E-05	S5A1	A5PJS2
75.99820868	1.733362965	0.43813061	3.956269946	7.61E-05	0.00698095	SALL2	Q9Y467
46.48421855	-1.741362072	0.337048605	-5.166501347	2.39E-07	7.95E-05	SDHA	Q0QF17
82.73195036	-1.693086703	0.513469706	-3.297344873	0.000976036	0.035543579	SEM3D	O95025
21.33632859	-3.090308766	0.899397181	-3.43597782	0.000590419	0.026068116	SEPR	A5D7B7
2284.560166	0.508536891	0.159780621	3.182719449	0.001458989	0.046750279	SFT2A	Q5SSN7
87.49055869	-1.015074813	0.249014413	-4.076369716	4.57E-05	0.005045791	SFXN5	Q8CFD0
3675.79523	0.588890027	0.166141704	3.544504557	0.000393352	0.020790277	SGPL1	O95470
16.43485512	-2.861541686	0.749760561	-3.816607375	0.000135299	0.009800783	SNAI2	O08954
14.22507953	-2.16708117	0.649553182	-3.336264418	0.000849124	0.032639797	SNURF	A5A6K1
54.46609665	0.957973601	0.244613965	3.916267011	8.99E-05	0.007521562	SNX4	A1A4L0
50.86306226	-1.042323469	0.299052884	-3.4854152	0.000491374	0.023228864	SOX7	Q9BT81
88.27940478	-2.038112085	0.493857611	-4.126922496	3.68E-05	0.004495491	SPB8	Q5BIR5
4.485252248	-3.466921724	1.093126478	-3.171565042	0.001516199	0.047882952	SPON2	Q9WV75
18.63688535	-2.885534094	0.828454033	-3.48303464	0.000495764	0.023291737	SRCRL	Q8BV57
179.2783523	-3.221147077	0.633389043	-5.085574358	3.67E-07	0.000110297	STAR	P49675
14.21578616	-2.985273047	0.760392097	-3.925965381	8.64E-05	0.007471112	STC1	P52823
9.377578242	-2.577502675	0.767600446	-3.357870215	0.000785455	0.031463667	STEA3	Q658P3
16.68293591	-2.275247322	0.718911285	-3.164851311	0.001551622	0.04879916	SUCA	O19069
2.651953732	-4.340769888	1.072751406	-4.046389374	5.20E-05	0.005422961	SV2A	Q29397
167.2389102	0.456873998	0.143025567	3.194351946	0.001401451	0.045389137	SYHM	A5D7V9
171.0221654	-1.470950407	0.371361265	-3.960968866	7.46E-05	0.00698095	T150B	A7MBB3
739.2455149	-1.842141535	0.501329189	-3.674514817	0.000238302	0.014866517	TAGL	Q9TS87
762.0047456	0.718240015	0.166399216	4.316366589	1.59E-05	0.002494998	TCTP	P61288

105.9002169	-2.484554791	0.602441143	-4.124145272	3.72E-05	0.004495491	TGM5	O43548
1096.328669	0.512934889	0.151168262	3.393138763	0.000690966	0.0287374	TIDC1	Q9NPL8
1272.465407	-0.787255931	0.235215623	-3.346954257	0.000817047	0.032054351	TLE5	P63003
13.17148431	-1.922559002	0.55652427	-3.454582498	0.000551146	0.024675111	TM154	Q6P9G4
141.8511736	0.742549309	0.216684746	3.426864712	0.000610593	0.026404686	TM177	Q53S58
14.2133784	2.588464364	0.614006507	4.21569533	2.49E-05	0.003212226	TM215	A7MB05
2056.596644	0.868271955	0.120186904	7.224347478	5.04E-13	1.92E-09	TM7S3	Q5FVF4
1228.748226	-2.751296326	0.541558938	-5.080326691	3.77E-07	0.000110297	TPH1	P17532
53.94619614	-1.889876704	0.519799023	-3.635783485	0.000277137	0.015979469	TR11B	A5D7R1
259.6623156	-0.862012094	0.249440326	-3.455784822	0.000548693	0.024675111	TRAM2	Q15035
10.10009162	-2.113415274	0.587415614	-3.597819367	0.000320896	0.017770117	TRFL	O77811
6.287065084	-2.112651301	0.596724781	-3.540411544	0.000399504	0.020969803	TRIB3	Q0VCE3
626.5708631	-0.440717677	0.116932962	-3.768977268	0.000163918	0.011096744	TRRAP	Q80YV3
462.8782319	-0.88510207	0.253776418	-3.487723863	0.000487151	0.02317316	TSN14	Q8QZY6
2914.070739	0.560922067	0.174943934	3.206296175	0.001344555	0.044300475	TSN9	O75954
39.3900117	-3.115086741	0.786395351	-3.961222223	7.46E-05	0.00698095	TSP1	Q28178
40.7032269	-2.722977141	0.792712057	-3.435014162	0.000592523	0.026068116	TSP3	P49746
120.1473612	0.761536113	0.199014174	3.82654209	0.000129956	0.009800783	TTI1	O43156
22.10516724	-2.828113602	0.737226815	-3.836151295	0.000124977	0.009608114	TULP3	O88413
1351.366315	-0.482481356	0.108290789	-4.455423791	8.37E-06	0.001481979	UBAP1	Q9NZ09
4.45417097	-5.533995361	1.108735603	-4.991266938	6.00E-07	0.000163051	VISL1	P62762
45.73514666	-8.889069233	1.389354628	-6.397984399	1.57E-10	1.71E-07	VPK25	P63125
26.38706006	1.13743528	0.323524498	3.515762444	0.000438493	0.021812884	VTNC	P22458
192.4434156	2.696763271	0.481244119	5.603732417	2.10E-08	1.06E-05	VWF	Q62935
26.73149034	-2.290839494	0.658358559	-3.479622862	0.00050212	0.023445619	WSB1	Q9Y6I7
25.10594254	-2.301538261	0.58535343	-3.931877979	8.43E-05	0.007373471	WSCD1	Q658N2
91.76563926	-1.658642753	0.285269215	-5.81430685	6.09E-09	4.21E-06	ZBED5	A4Z944
10.66275255	-1.680986729	0.503051577	-3.341579283	0.000833032	0.032348001	ZN691	Q3TDE8

Table S4.4. Top DEGs pathways between BPI vs SC & LI.

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_CC_DIRECT	GO:0043235~receptor complex	6	2.44	0.06	A5D7R1, P16144, Q99PE7, O70570, P48356, Q08345	236	93	10395	2.84	1.00	1.00	0.99
GOTERM_MF_DIRECT	GO:0001664~G protein-coupled receptor binding	4	1.63	0.03	A5PK71, P16144, O88413, P57756	231	30	9943	5.74	1.00	1.00	1.00
GOTERM_CC_DIRECT	GO:0044297~cell body	6	2.44	0.01	A5PK71, P62738, Q14195	236	55	10395	4.81	0.91	0.26	0.26
KEGG_PATHWAY	rno04820:Cytoskeleton in muscle cells	4	1.63	0.01	O35115, P62738	116	18	4969	9.52	0.96	1.00	1.00
GOTERM_BP_DIRECT	GO:0019470~4-hydroxyproline catabolic process	2	0.81	0.05	O43272, Q86XE5	228	2	9968	43.72	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0006954~inflammatory response	9	3.66	0.08	O43353, O43586, Q62935, Q05910, Q9TTQ4, P14555, O97664, P02763, P25105	228	196	9968	2.01	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0045627~positive regulation of T-helper 1 cell differentiation	2	0.81	0.09	O43353, P33681	228	4	9968	21.86	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0070374~positive regulation of ERK1 and ERK2 cascade	8	3.25	0.01	O43353, P78524, P62738, P14555, P97391	228	106	9968	3.30	1.00	0.74	0.74
UP_KW_MOLECULAR_FUNCTION	KW-0675~Receptor	19	7.72	0.04	O43353, Q920C4, O70431, Q9TU72, O70570, Q08345, O97664, P53710, Q9NVV5, Q9H2W1, Q8BV57, A5D7R1, P16144, Q6RW13, P33681, Q9NZQ7, P48356, O43194, P25105	155	476	6374	1.64	0.92	0.63	0.63
GOTERM_BP_DIRECT	GO:0032760~positive regulation of tumor necrosis factor production	6	2.44	0.01	O43353, Q9WV75, Q28178, P04792, P02763, P25105	228	58	9968	4.52	1.00	0.74	0.74
GOTERM_BP_DIRECT	GO:0042742~defense response to bacterium	6	2.44	0.01	O43353, Q9WV75, Q9TT92, O19114, Q5R9T9, P14555	228	62	9968	4.23	1.00	0.89	0.88
GOTERM_CC_DIRECT	GO:0009986~cell surface	15	6.10	0.01	O77811, Q28178, A5D7B7, P51654, P53710, P14431, P16144, P00747, P33681, Q5RAL8, Q05910, P08132, P55281, P97391, Q8QZY6	236	294	10395	2.25	0.88	0.26	0.26
UP_SEQ_FEATURE	DOMAIN:Ig-like C2-type	4	1.63	0.10	O95025, P33681, Q5RAL8, Q9NZQ7	232	50	10393	3.58	1.00	1.00	1.00
GOTERM_MF_DIRECT	GO:0002020~protease binding	6	2.44	0.02	P02461, P00747, Q3MI99, Q62935, Q28178, A5D7B7	231	69	9943	3.74	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:0007160~cell-matrix adhesion	6	2.44	0.01	P02461, P16144, Q05910, P22458, P53710, P49746	228	51	9968	5.14	1.00	0.52	0.52
GOTERM_BP_DIRECT	GO:0007229~integrin-mediated signaling pathway	5	2.03	0.05	P02461, P16144, Q05910, P55281, P53710	228	61	9968	3.58	1.00	1.00	0.99
GOTERM_MF_DIRECT	GO:0005178~integrin binding	10	4.07	0.00	P02461, P16144, Q62935, Q05910, Q28178, A5D7B7, P55281, P22458, P53710, Q5EA62	231	102	9943	4.22	0.25	0.10	0.10
GOTERM_BP_DIRECT	GO:0043588~skin development	3	1.22	0.10	P02461, P97431, Q04656	228	23	9968	5.70	1.00	1.00	0.99
UP_SEQ_FEATURE	DOMAIN:VWFC	4	1.63	0.01	P02461, Q62935, Q28178, Q5R3Z7	232	17	10393	10.54	0.99	1.00	1.00
INTERPRO	IPR001007:VWF_dom	4	1.63	0.01	P02461, Q62935, Q28178, Q5R3Z7	246	17	10695	10.23	1.00	1.00	1.00
SMART	SM00214:VWC	4	1.63	0.01	P02461, Q62935, Q28178, Q5R3Z7	129	17	5400	9.85	0.61	0.23	0.23
INTERPRO	IPR008160:Collagen	4	1.63	0.05	P02461, Q96CG8, Q3MI99, P57756	246	37	10695	4.70	1.00	1.00	1.00

GOTERM_MF_DIRECT	GO:0005201~extracellular matrix structural constituent	5	2.03	0.01	P02461, Q96CG8, Q8R4W6, P55083, P49746	231	40	9943	5.38	1.00	1.00	1.00
GOTERM_CC_DIRECT	GO:0062023~collagen-containing extracellular matrix	18	7.32	0.00	P02461, Q9TT92, Q62935, Q28178, Q62000, P51654, P57756, P51942, P45845, Q8BV57, Q96CG8, P00747, Q8R4W6, P55083, P22458, P02763, P49746, Q5EA62	236	149	10395	5.32	0.00	0.00	0.00
UP_KW_BIOLOGICAL_PROCESS	KW-0391~Immunity	15	6.10	0.01	P09913, O43353, Q9WV75, Q96D96, O77811, O43586, P33617, P57756, Q9UDX4, Q8BV57, P14431, O15049, Q5R9T9, Q9NZQ7, O15554	129	342	6212	2.11	0.60	0.46	0.46
GOTERM_BP_DIRECT	GO:0009615~response to virus	7	2.85	0.01	P09913, P62738, Q9TU72, P04792	228	73	9968	4.19	1.00	0.52	0.52
GOTERM_BP_DIRECT	GO:0050796~regulation of insulin secretion	3	1.22	0.07	P09917, O19114, P00348	228	20	9968	6.56	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0061044~negative regulation of vascular wound healing	2	0.81	0.05	P09917, P13909	228	2	9968	43.72	1.00	1.00	0.99
GOTERM_CC_DIRECT	GO:0009897~external side of plasma membrane	9	3.66	0.05	P14431, P00747, P33681, P33617, Q62935, Q28178, Q9NZQ7, P48356, P53710	236	177	10395	2.24	1.00	1.00	0.99
GOTERM_CC_DIRECT	GO:0032398~MHC class Ib protein complex	2	0.81	0.09	P14431, P33617	236	4	10395	22.02	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0006955~immune response	10	4.07	0.00	P14431, P33681, P33617, Q8BV66, Q9TTQ4, Q28178, Q5R9T9, Q9NZQ7, P22458, P25105	228	140	9968	3.12	1.00	0.45	0.44
UP_SEQ_FEATURE	DOMAIN:Phospholipase A2	2	0.81	0.07	P14555, P97391	232	3	10393	29.86	1.00	1.00	1.00
INTERPRO	IPR033113:PLipase_A2_His_AS	2	0.81	0.07	P14555, P97391	246	3	10695	28.98	1.00	1.00	1.00
INTERPRO	IPR016090:PLipase_A2_dom	2	0.81	0.07	P14555, P97391	246	3	10695	28.98	1.00	1.00	1.00
INTERPRO	IPR001211:PLipase_A2	2	0.81	0.07	P14555, P97391	246	3	10695	28.98	1.00	1.00	1.00
INTERPRO	IPR033112:PLipase_A2_Asp_AS	2	0.81	0.07	P14555, P97391	246	3	10695	28.98	1.00	1.00	1.00
SMART	SM00085:PA2c	2	0.81	0.07	P14555, P97391	129	3	5400	27.91	1.00	1.00	1.00
INTERPRO	IPR036444:PLipase_A2_dom_sf	2	0.81	0.09	P14555, P97391	246	4	10695	21.74	1.00	1.00	1.00
UP_KW_MOLECULAR_FUNCTION	KW-0646~Protease inhibitor	5	2.03	0.09	P15943, Q5BIR5, P51654, O95980, P13909	155	71	6374	2.90	1.00	0.91	0.91
GOTERM_BP_DIRECT	GO:0033627~cell adhesion mediated by integrin	3	1.22	0.10	P16144, P22458, P53710	228	23	9968	5.70	1.00	1.00	0.99
GOTERM_CC_DIRECT	GO:0005604~basement membrane	6	2.44	0.00	P16144, P62738, A2SW69	236	45	10395	5.87	0.63	0.14	0.14
UP_KW_MOLECULAR_FUNCTION	KW-0401~Integrin	3	1.22	0.10	P16144, Q05910, P53710	155	22	6374	5.61	1.00	0.91	0.91
UP_SEQ_FEATURE	DOMAIN:EGF-like	11	4.47	0.00	P16144, Q3MI99, P05979, Q05910, Q5RDI1, Q28178, Q5R3Z7, Q7Z3S9, P51942, P49746, Q5EA62	232	92	10393	5.36	0.03	0.01	0.01
INTERPRO	IPR000742:EGF-like_dom	11	4.47	0.00	P16144, Q3MI99, P05979, Q05910, Q5RDI1, Q28178, Q5R3Z7, Q7Z3S9, P51942, P49746, Q5EA62	246	99	10695	4.83	0.08	0.08	0.08
UP_KW_DOMAIN	KW-0245~EGF-like domain	11	4.47	0.00	P16144, Q3MI99, P05979, Q05910, Q5RDI1, Q28178, Q5R3Z7, Q7Z3S9, P51942, P49746, Q5EA62	180	105	7076	4.12	0.01	0.00	0.00
GOTERM_CC_DIRECT	GO:0043005~neuron projection	8	3.25	0.05	P17532, A5PK71, Q96CG8, P05979, Q04656, Q29397, O15554, O97664	236	145	10395	2.43	1.00	1.00	0.99

GOTERM_BP_DIRECT	GO:0046470~phosphatidylcholine metabolic process	3	1.22	0.03	P27169, P14555, P97391	228	13	9968	10.09	1.00	1.00	0.99
OMIM_DISEASE	Coronary artery disease, susceptibility to~Coronary artery disease, susceptibility to	2	0.81	0.05	P27169, Q15165	33	2	1235	37.42	0.92	1.00	1.00
UP_SEQ_FEATURE	DISULFID:In form B	2	0.81	0.07	P27169, Q15165	232	3	10393	29.86	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:1901335~lactone catabolic process	2	0.81	0.07	P27169, Q15165	228	3	9968	29.15	1.00	1.00	0.99
INTERPRO	IPR051288:Serum_paraoxonase/arylesterase	2	0.81	0.07	P27169, Q15165	246	3	10695	28.98	1.00	1.00	1.00
INTERPRO	IPR002640:Arylesterase	2	0.81	0.07	P27169, Q15165	246	3	10695	28.98	1.00	1.00	1.00
INTERPRO	IPR008364:Paraoxonase2	2	0.81	0.07	P27169, Q15165	246	3	10695	28.98	1.00	1.00	1.00
GOTERM_MF_DIRECT	GO:0102007~acyl-L-homoserine-lactone lactonohydrolase activity	2	0.81	0.07	P27169, Q15165	231	3	9943	28.70	1.00	1.00	1.00
GOTERM_MF_DIRECT	GO:0016491~oxidoreductase activity	9	3.66	0.03	P28332, Q3T0C2, P09917, Q925G2, D3ZF77, P04394, Q8R4A1, P47895, Q99MZ7	231	161	9943	2.41	1.00	1.00	1.00
KEGG_PATHWAY	hsa00830:Retinol metabolism	3	1.22	0.08	P28332, Q8N3Y7, P47895	116	21	4969	6.12	1.00	1.00	1.00
GOTERM_MF_DIRECT	GO:0005509~calcium ion binding	22	8.94	0.00	P29269, P27169, Q28178, A2SW69, Q5R3Z7, Q7Z3S9, P14555, P51942, Q13609, Q96D15, P62762, P61288, Q3MI99, Q05910, Q5RDI1, P24732, P08132, P55281, P97391, Q4R8T1, P49746, Q5EA62	231	333	9943	2.84	0.01	0.01	0.01
GOTERM_CC_DIRECT	GO:0001725~stress fiber	6	2.44	0.03	P29269, P62738, Q80YS6	236	77	10395	3.43	1.00	0.75	0.74
UP_KW_MOLECULAR_FUNCTION	KW-0514~Muscle protein	7	2.85	0.00	P29269, P62738, Q9TS87, P24732	155	49	6374	5.87	0.07	0.07	0.07
UP_SEQ_FEATURE	DOMAIN:EF-hand 3	5	2.03	0.04	P29269, P62762, P24732, Q4R8T1, Q96D15	232	56	10393	4.00	1.00	1.00	1.00
COG_ONTOLOGY	Signal transduction mechanisms / Cytoskeleton / Cell division and chromosome partitioning / General function prediction only	4	1.63	0.00	P29269, P62762, P24732, Q96D15	22	20	1167	10.61	0.07	0.08	0.08
GOTERM_BP_DIRECT	GO:0042130~negative regulation of T cell proliferation	5	2.03	0.00	P33681, O54735, Q9NZQ7, P14555, P97391	228	22	9968	9.94	0.84	0.17	0.17
INTERPRO	IPR051713:T-cell_Activation_Regulation	2	0.81	0.05	P33681, Q9NZQ7	246	2	10695	43.48	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:0048251~elastic fiber assembly	5	2.03	0.00	P45845, P02461, Q04656, P55083, Q5EA62	228	9	9968	24.29	0.04	0.01	0.01
GOTERM_BP_DIRECT	GO:0030199~collagen fibril organization	4	1.63	0.04	P45845, P02461, Q04656, Q8R4A1	228	33	9968	5.30	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0030324~lung development	5	2.03	0.03	P45845, P02461, Q3MI99, Q5RES1, P51654	228	51	9968	4.29	1.00	1.00	0.99
GOTERM_CC_DIRECT	GO:0005581~collagen trimer	5	2.03	0.02	P45845, P02461, Q96CG8, Q3MI99, P57756	236	47	10395	4.69	1.00	0.64	0.63

GOTERM_MF_DIRECT	GO:0005518~collagen binding	5	2.03	0.01	P45845, Q3MI99, Q62935, Q8R4W6, Q08345	231	37	9943	5.82	0.99	1.00	1.00
UP_KW_LIGAND	KW-0186~Copper	4	1.63	0.07	P45845, Q658P3, Q04656, P15943	111	37	4168	4.06	0.92	0.76	0.76
GOTERM_BP_DIRECT	GO:0001649~osteoblast differentiation	5	2.03	0.06	P45845, Q8BHE5, O35115, Q96CG8, O08954	228	63	9968	3.47	1.00	1.00	0.99
KEGG_PATHWAY	ocu04510:Focal adhesion	3	1.22	0.01	P54850, P24732, P22458	116	6	4969	21.42	0.96	1.00	1.00
INTERPRO	IPR050373:Fibrinogen_C-term_domain	2	0.81	0.09	P55083, P57756	246	4	10695	21.74	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:0043269~regulation of monoatomic ion transport	2	0.81	0.05	P59647, O00168	228	2	9968	43.72	1.00	1.00	0.99
INTERPRO	IPR000272:Ion-transport_regulator_FXYD	2	0.81	0.05	P59647, O00168	246	2	10695	43.48	1.00	1.00	1.00
INTERPRO	IPR047297:FXYD_motif	2	0.81	0.05	P59647, O00168	246	2	10695	43.48	1.00	1.00	1.00
GOTERM_MF_DIRECT	GO:0099106~ion channel regulator activity	2	0.81	0.05	P59647, O00168	231	2	9943	43.04	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:1903278~positive regulation of sodium ion export across plasma membrane	2	0.81	0.09	P59647, O00168	228	4	9968	21.86	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0006936~muscle contraction	6	2.44	0.00	P62738, P24732, O00168	228	32	9968	8.20	0.62	0.10	0.10
GOTERM_BP_DIRECT	GO:0008217~regulation of blood pressure	6	2.44	0.00	P62738, Q6RW13, P05979	228	43	9968	6.10	0.98	0.29	0.29
GOTERM_BP_DIRECT	GO:0090090~negative regulation of canonical Wnt signaling pathway	5	2.03	0.08	P63003, Q96CG8, O08954, P51654, Q6XE24	228	72	9968	3.04	1.00	1.00	0.99
GOTERM_CC_DIRECT	GO:0032010~phagolysosome	2	0.81	0.07	Q05910, P97391	236	3	10395	29.36	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0061302~smooth muscle cell-matrix adhesion	2	0.81	0.05	Q08345, P22458	228	2	9968	43.72	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0072180~mesonephric duct morphogenesis	2	0.81	0.05	Q08DS3, P51654	228	2	9968	43.72	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0072111~cell proliferation involved in kidney development	2	0.81	0.07	Q08DS3, P51654	228	3	9968	29.15	1.00	1.00	0.99
UP_SEQ_FEATURE	MOTIF:Cell attachment site	5	2.03	0.01	Q28178, P22458, P55083, P53710, Q5EA62	232	43	10393	5.21	1.00	1.00	1.00
UP_SEQ_FEATURE	REPEAT:TSP type-32	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	REPEAT:TSP type-35	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	REPEAT:TSP type-34	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	REPEAT:TSP type-31	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	REPEAT:TSP type-33	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	REPEAT:TSP type-37	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00

UP_SEQ_FEATURE	REPEAT:TSP type-3 8	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	REPEAT:TSP type-3	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	REPEAT:TSP type-3 6	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	DOMAIN:TSP C-terminal	2	0.81	0.09	Q28178, P49746	232	4	10393	22.40	1.00	1.00	1.00
INTERPRO	IPR003367:Thrombospondin_3-like_rpt	2	0.81	0.09	Q28178, P49746	246	4	10695	21.74	1.00	1.00	1.00
INTERPRO	IPR008859:Thrombospondin_C	2	0.81	0.09	Q28178, P49746	246	4	10695	21.74	1.00	1.00	1.00
INTERPRO	IPR028974:TSP_type-3_rpt	2	0.81	0.09	Q28178, P49746	246	4	10695	21.74	1.00	1.00	1.00
INTERPRO	IPR017897:Thrombospondin_3_rpt	2	0.81	0.09	Q28178, P49746	246	4	10695	21.74	1.00	1.00	1.00
KEGG_PATHWAY	bta04512:ECM-receptor interaction	3	1.22	0.01	Q28178, Q29397, P53710	116	7	4969	18.36	0.99	1.00	1.00
UP_SEQ_FEATURE	DOMAIN:Laminin G-like	3	1.22	0.01	Q28178, Q5R3Z7, P49746	232	7	10393	19.20	1.00	1.00	1.00
INTERPRO	IPR048287:TSPN-like_N	3	1.22	0.01	Q28178, Q5R3Z7, P49746	246	8	10695	16.30	1.00	1.00	1.00
SMART	SM00210:TSPN	3	1.22	0.01	Q28178, Q5R3Z7, P49746	129	8	5400	15.70	0.85	0.38	0.38
GOTERM_BP_DIRECT	GO:0002062~chondrocyte differentiation	4	1.63	0.04	Q32KU9, P02461, Q08DS3, P51942	228	32	9968	5.46	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0042593~glucose homeostasis	5	2.03	0.08	Q32KU9, P09917, P48356, 097664, 054937	228	73	9968	2.99	1.00	1.00	0.99
GOTERM_BP_DIRECT	GO:0035988~chondrocyte proliferation	3	1.22	0.02	Q32KU9, P52823, Q5RES1	228	11	9968	11.92	1.00	1.00	0.99
GOTERM_CC_DIRECT	GO:0005576~extracellular region	40	16.26	0.00	Q32KU9, Q9BYZ8, Q6UWY0, Q9TTQ4, Q28178, O70570, Q5R3Z7, Q62000, Q7Z3S9, P14555, P51942, Q8BV57, Q96CG8, P00747, Q15165, O88413, Q9UKY3, Q8R4W6, P02763, P97391, P21664, Q5EA62, P27169, P02461, Q9TT92, P09917, O77811, Q62935, A2SW69, Q1LZE9, O95980, P57756, Q13609, P45845, Q8BHE5, Q148F8, Q5RDI1, P48356, P55083, P49746	236	700	10395	2.52	0.00	0.00	0.00
UP_KW_CELLULAR_COMPONENT	KW-0964~Secreted	47	19.11	0.00	Q32KU9, Q9WV75, Q9BYZ8, Q6UWY0, Q9TTQ4, Q28178, A5D7B7, O70570, Q8R4A1, Q5R3Z7, Q62000, Q7Z3S9, P14555, Q08345, P51942, Q8BV57, Q96CG8, A5D7R1, P00747, Q3MI99, O88413, Q9NZQ7, Q9UKY3, Q8R4W6, P22458, P02763, P97391, P13909, P21664, Q5EA62, P27169, P02461, O95025, Q9TT92, O77811, Q62935, A2SW69, P52823, Q1LZE9, P57756, Q13609, P45845, Q8BHE5, Q148F8, Q5RDI1, P48356, P55083	226	759	9590	2.63	0.00	0.00	0.00
GOTERM_BP_DIRECT	GO:0002040~sprouting angiogenesis	3	1.22	0.09	Q3MI99, Q28178, O95980	228	22	9968	5.96	1.00	1.00	0.99
INTERPRO	IPR018097:EGF_Ca-bd_CS	5	2.03	0.02	Q3MI99, Q5R3Z7, Q7Z3S9, P49746, Q5EA62	246	43	10695	5.06	1.00	1.00	1.00
INTERPRO	IPR000152:EGF-type_Asp/Asn_hydroxyl_site	4	1.63	0.07	Q3MI99, Q5R3Z7, Q7Z3S9, Q5EA62	246	41	10695	4.24	1.00	1.00	1.00
INTERPRO	IPR001881:EGF-like_Ca-bd_dom	8	3.25	0.00	Q3MI99, Q5RDI1, Q28178, Q5R3Z7, Q7Z3S9, P51942, P49746, Q5EA62	246	56	10695	6.21	0.23	0.13	0.13
SMART	SM00179:EGF_CA	8	3.25	0.00	Q3MI99, Q5RDI1, Q28178, Q5R3Z7, Q7Z3S9, P51942, P49746, Q5EA62	129	56	5400	5.98	0.04	0.04	0.04

SMART	SM00181:EGF	8	3.25	0.00	Q3MI99, Q5RDI1, Q28178, Q5R3Z7, Q7Z3S9, P51942, P49746, Q5EA62	129	81	5400	4.13	0.32	0.19	0.19
GOTERM_MF_DIRECT	GO:0070403~NAD + binding	3	1.22	0.05	Q3T0C2, P00348, P47895	231	16	9943	8.07	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:0007567~parturition	2	0.81	0.05	Q3T0C2, P25105	228	2	9968	43.72	1.00	1.00	0.99
UP_KW_BIOLOGICAL_PROCESS	KW-0276~Fatty acid metabolism	7	2.85	0.06	Q3T0C2, Q8VHQ9, P05979, P00348, Q8BWN8, P97391, Q99MZ7	129	137	6212	2.46	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:0006508~proteolysis	12	4.88	0.07	Q4R7U8, Q9Y6Q1, Q9TT92, P63125, P00747, O77811, Q05910, Q5RDI1, A5D7B7, Q1LZE9, P57756	228	294	9968	1.78	1.00	1.00	0.99
UP_SEQ_FEATURE	DOMAIN:EF-hand 6	2	0.81	0.09	Q4R8T1, Q96D15	232	4	10393	22.40	1.00	1.00	1.00
UP_SEQ_FEATURE	DOMAIN:EGF-like 2; calcium-binding	3	1.22	0.10	Q5R3Z7, P49746, Q5EA62	232	24	10393	5.60	1.00	1.00	1.00
INTERPRO	IPR049883:NOTCH1_EGF-like	4	1.63	0.03	Q5R3Z7, Q7Z3S9, P49746, Q5EA62	246	31	10695	5.61	1.00	1.00	1.00
UP_SEQ_FEATURE	DOMAIN:EGF-like 5; calcium-binding	3	1.22	0.05	Q5R3Z7, Q7Z3S9, Q5EA62	232	16	10393	8.40	1.00	1.00	1.00
SMART	SM00215:VWC_out	2	0.81	0.07	Q62935, Q5R3Z7	129	3	5400	27.91	1.00	1.00	1.00
INTERPRO	IPR036291:NAD(P)-bd_dom_sf	7	2.85	0.08	Q658P3, P28332, Q3T0C2, O19069, P00348, Q8N3Y7, Q99MZ7	246	130	10695	2.34	1.00	1.00	1.00
UP_KW_MOLECULAR_FUNCTION	KW-0560~Oxidoreductase	23	9.35	0.00	Q658P3, P28332, Q3T0C2, P09917, Q925G2, O19114, P05979, O55071, Q8R4A1, Q99MZ7, P45845, P17532, Q3TMX7, A5PJS2, Q0QF17, P00348, D3ZF77, A0A087X1C5, P04394, Q8N3Y7, P47895, O43272, P0CB83	155	496	6374	1.91	0.22	0.13	0.13
UP_KW_LIGAND	KW-0274~FAD	6	2.44	0.09	Q658P3, Q3TMX7, O19114, Q0QF17, Q8R4A1, O43272	111	91	4168	2.48	0.96	0.76	0.76
KEGG_PATHWAY	ocu04371:Apelin signaling pathway	2	0.81	0.07	Q6PNB6, O19114	116	3	4969	28.56	1.00	1.00	1.00
KEGG_PATHWAY	ocu04926:Relaxin signaling pathway	2	0.81	0.09	Q6PNB6, O19114	116	4	4969	21.42	1.00	1.00	1.00
GOTERM_MF_DIRECT	GO:0051087~protein-folding chaperone binding	5	2.03	0.10	Q6PNB6, P00747, Q148F8, Q04656, Q62935	231	76	9943	2.83	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:0060021~roof of mouth development	5	2.03	0.01	Q7Z3E1, P97431, O95470, O08954, Q08DS3	228	36	9968	6.07	1.00	0.69	0.68
GOTERM_BP_DIRECT	GO:0060828~regulation of canonical Wnt signaling pathway	5	2.03	0.00	Q80U19, A0A8V8TMC4, Q9BT81, P51654, O95980	228	23	9968	9.50	0.89	0.18	0.18
GOTERM_BP_DIRECT	GO:2000050~regulation of non-canonical Wnt signaling pathway	2	0.81	0.05	Q80U19, P51654	228	2	9968	43.72	1.00	1.00	0.99
UP_SEQ_FEATURE	DISULFID:Interchain	7	2.85	0.01	Q8BHE5, P02461, O14519, Q02846, Q28178, P52823, P49746	232	90	10393	3.48	1.00	1.00	1.00
UP_KW_PTM	KW-0165~Cleavage on pair of basic residues	7	2.85	0.04	Q8BHE5, Q9TT92, P63125, P00747, A5D7B7, Q5RES1	192	104	8050	2.82	0.62	0.25	0.25
GOTERM_MF_DIRECT	GO:0050840~extracellular matrix binding	3	1.22	0.07	Q8BV57, Q9TT92, P22458	231	19	9943	6.80	1.00	1.00	1.00
GOTERM_CC_DIRECT	GO:0005886~plasma membrane	72	29.27	0.00	Q8J025, O43353, Q96NE9, Q658P3, P54850, Q96D96, A5D7B7, O70570, A7MBB3, P14555, A0A8V8TMC4, P63125, P16144, P00747, P33681, Q5RAL8, O88413, O94868, O15554, Q8N3Y7, P97391, Q8QZY6, Q925G2, Q9TU72, Q5R5U1, O97664, O60779, Q99712, Q02846,	236	2268	10395	1.40	0.36	0.07	0.07

UP_SEQ_FEATURE	TOPO_DOM:Extracellular	47	19.11	0.00	P50749, Q05910, P48356, P08132, P55281, P78524, O70431, Q5FVF4, O08849, Q08345, P59647, Q8N5C1, O00168, Q9H2W1, Q04656, Q15165, Q5R5N8, Q9NZQ7, Q9NZ09, P25105, Q6PNB6, O95025, Q920C4, O75954, O43586, O77811, P97577, Q99PE7, Q91X49, O43548, A2SW69, P51654, O95980, P57756, Q9NVV5, Q6RW13, Q5RES1, A1A4L0, O43194, Q14165, Q9ULV3, B8Q0B2	232	1029	10393	2.05	0.00	0.00	0.00
UP_SEQ_FEATURE	CARBOHYD:N-linked (GlcNAc...) asparagine	77	31.30	0.00	Q8J025, Q658N2, P33617, Q96D96, O70431, A5D7B7, A7MBB3, O70570, P15943, Q08345, P59647, Q8N5C1, O00168, Q9H2W1, P63125, P16144, P33681, Q5RAL8, Q04656, A0A087X1C5, Q9NZQ7, P25105, Q8QZY6, Q920C4, Q925G2, O75954, Q99PE7, Q9TU72, Q9ULK6, Q29397, O97664, Q6P9G4, O60779, Q9ULV3, P50749, Q9NVV5, P14431, Q99712, Q02846, Q7YR42, Q6RW13, Q05910, P48356, Q5RES1, P55281, O43194, Q15035, B8Q0B2	232	1734	10393	1.99	0.00	0.00	0.00
UP_SEQ_FEATURE	TOPO_DOM:Cytoplasmic	57	23.17	0.00	Q8J025, Q658P3, P54850, Q28178, A5D7B7, O70570, A7MBB3, P51942, Q96D15, Q8BV57, P63125, Q9BYG0, A5D7R1, Q96CG8, P16144, P00747, Q3MI99, P33681, Q5RAL8, Q9UKY3, Q8R4W6, P22458, P02763, P13909, Q8QZY6, Q5EA62, P27169, Q864U6, Q9TT92, P05979, Q62935, Q9ULK6, Q29397, O97664, Q1LZE9, O60779, Q8BHE5, Q02846, Q3TMX7, Q7YR42, Q05910, P48356, P55083, P55281, A7MB63, Q15035, Q658N2, Q9BYZ8, P33617, Q5FVF4, Q6UWY0, Q8R4A1, Q5R3Z7, Q08345, Q7Z3S9, Q04656, Q15165, A0A087X1C5, Q9NZQ7, P25105, O95025, Q8K449, O75954, O77811, Q99PE7, P52823, P51654, O95980, P53710, P57756, P14431, Q5RDI1, O43194, P49746, Q14165, B8Q0B2	232	1668	10393	1.53	0.49	0.17	0.17
UP_KW_DOMAIN	KW-0812~Transmembrane	81	32.93	0.07	Q8J025, Q86VU5, Q658P3, P54850, Q96D96, O95470, A5D7B7, O70570, A7MBB3, P63125, Q9BYG0, P16144, Q3MI99, P33681, Q5RAL8, O15554, Q8N3Y7, Q53S58, Q86XE5, Q8QZY6, Q864U6, Q925G2, P05979, Q9TU72, Q9ULK6, Q29397, O97664, Q1LZE9, Q6P9G4, O60779, Q9UDX4, Q99712, Q02846, Q3TMX7, Q7YR42, A5PJS2, Q9NPL8, Q05910, P48356, P55281, A7MB63, Q15035, Q658N2, P33617, O70431, Q5FVF4, Q6UWY0, Q5SSN7, P15943, Q2KIY1, Q08345, P59647, Q8N5C1, O00168, Q0IIE8, Q9H2W1, Q04656, Q15165, A0A087X1C5, Q9NZQ7, P25105, O95025, Q920C4, Q8K449, O75954, A7MB05, Q99PE7, Q91X49, P53710, Q6X9E4, Q9NVV5, P14431, Q6RW13, Q8CFD0, Q5RES1, Q8BM86, O43194, Q9JJ61, Q14165, B8Q0B2	180	2769	7076	1.15	0.82	0.47	0.47
UP_KW_DOMAIN	KW-1133~Transmembrane helix	80	32.52	0.08	Q8J025, Q86VU5, Q658P3, P54850, Q96D96, O95470, A5D7B7, O70570, A7MBB3, P63125, Q9BYG0, P16144, Q3MI99, P33681, Q5RAL8, O15554, Q8N3Y7, Q53S58, Q86XE5, Q8QZY6,	180	2747	7076	1.14	0.86	0.47	0.47

UP_SEQ_FEATURE	TRANSMEM:Helical	74	30.08	0.01	Q864U6, Q925G2, P05979, Q9TU72, Q9ULK6, Q29397, O97664, Q1LZE9, Q6P9G4, O60779, Q9UDX4, Q99712, Q02846, Q3TMX7, Q7YR42, A5PJS2, Q9NPL8, Q05910, P48356, P55281, A7MB63, Q15035, Q658N2, P33617, O70431, Q5FVF4, Q6UWY0, Q5SSN7, P15943, Q2KIY1, Q08345, P59647, Q8N5C1, O00168, Q0IIE8, Q9H2W1, Q04656, Q15165, A0A087X1C5, Q9NZQ7, P25105, O95025, Q920C4, Q8K449, O75954, A7MB05, Q99PE7, Q91X49, P53710, Q6X9E4, Q9NVV5, Q6RW13, Q8CFD0, Q5RES1, Q8BM86, O43194, Q9JJ61, Q14165, B8Q0B2	232	2549	10393	1.30	1.00	1.00	1.00
GOTERM_MF_DIRECT	GO:0017147~Wnt-protein binding	3	1.22	0.04	Q8J025, Q96CG8, O95980	231	14	9943	9.22	1.00	1.00	1.00
UP_KW_DOMAIN	KW-0732~Signal	81	32.93	0.00	Q8J025, Q9WV75, Q3SXM0, Q28178, Q9TTQ4, O70570, Q62000, P14555, P51942, Q96D15, Q8BV57, P63125, A5D7R1, Q96CG8, P16144, P00747, Q3MI99, P33681, Q5RAL8, Q9UKY3, Q8R4W6, P22458, P02763, P97391, P13909, A5D7V9, Q5EA62, P27169, P28332, Q9TT92, Q4KLY0, P05979, Q62935, Q9ULK6, Q1LZE9, Q6P9G4, Q13609, P45845, Q8BHE5, Q02846, Q3TMX7, Q9NPL8, Q05910, P48356, P55083, P55281, P0CB83, A7MB63, Q9BYZ8, P33617, Q5FVF4, Q6UWY0, O19069, Q8R4A1, P15943, Q5R3Z7, Q08345, P59647, Q7Z3S9, O00168, Q9Y6I7, Q15165, Q9NZQ7, P02461, O95025, O75954, O77811, P52823, P51654, O95980, P53710, P57756, Q9NVV5, P14431, Q6RW13, Q5RDI1, Q5RES1, Q8BM86, P49746, Q14165	180	1755	7076	1.81	0.00	0.00	0.00
UP_KW_PTM	KW-0325~Glycoprotein	82	33.33	0.00	Q8J025, Q9WV75, Q658P3, P54850, Q28178, Q9TTQ4, A5D7B7, O70570, A7MBB3, Q62000, P51942, Q96D15, Q8BV57, P63125, Q9BYG0, A5D7R1, Q96CG8, P16144, P00747, Q3MI99, P33681, Q5RAL8, Q9UKY3, Q8R4W6, P22458, P02763, P13909, Q8QZY6, Q5EA62, P27169, Q864U6, Q9TT92, P05979, Q62935, Q9ULK6, Q29397, O97664, Q1LZE9, O60779, Q8BHE5, Q02846, Q3TMX7, Q7YR42, Q05910, P48356, P55083, P55281, A7MB63, Q15035, Q658N2, Q9BYZ8, P33617, Q5FVF4, Q6UWY0, Q8R4A1, P15943, Q5R3Z7, Q08345, Q7Z3S9, Q04656, Q15165, A0A087X1C5, Q9NZQ7, P25105, P02461, O95025, Q8K449, O75954, O77811, Q99PE7, P52823, P51654, O95980, P53710, P57756, P14431, Q5RDI1, O43194, P49746, Q14165, B8Q0B2	192	1895	8050	1.81	0.00	0.00	0.00
GOTERM_MF_DIRECT	GO:0004666~prostaglandin-endoperoxide synthase activity	2	0.81	0.05	Q96CG8, P05979	231	2	9943	43.04	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:0006813~potassium ion transport	4	1.63	0.03	Q99712, O15554, P59647, O00168	228	30	9968	5.83	1.00	1.00	0.99

INTERPRO	IPR030417:MS4A	2	0.81	0.09	Q9H2W1, Q920C4	246	4	10695	21.74	1.00	1.00	1.00
GOTERM_MF_DIRECT	GO:0008201~heparin binding	10	4.07	0.00	Q9TT92, O77811, Q9BYZ8, Q28178, P15943, Q5R3Z7, Q8R4W6, P22458, P97391, P49746	231	94	9943	4.58	0.14	0.08	0.08
GOTERM_BP_DIRECT	GO:0030198~extracellular matrix organization	11	4.47	0.00	Q9TT92, P02461, A5D7R1, Q04656, Q62935, Q8R4A1, P15943, Q5RES1, P22458, O95980, P51942	228	66	9968	7.29	0.00	0.00	0.00
GOTERM_BP_DIRECT	GO:0001916~positive regulation of T cell mediated cytotoxicity	3	1.22	0.07	Q9UDX4, P14431, P33617	228	19	9968	6.90	1.00	1.00	0.99
GOTERM_MF_DIRECT	GO:0042605~peptide antigen binding	3	1.22	0.10	Q9UDX4, P14431, P33617	231	23	9943	5.61	1.00	1.00	1.00
GOTERM_CC_DIRECT	GO:0030670~phagocytic vesicle membrane	5	2.03	0.03	Q9UDX4, P14431, Q96D96, Q04656, Q5R5U1	236	52	10395	4.24	1.00	0.75	0.74
UP_KW_MOLECULAR_FUNCTION	KW-1278~Translocase	7	2.85	0.01	Q9UDX4, Q925G2, Q8K449, Q99PE7, Q04656, P04394, P0CB83	155	68	6374	4.23	0.32	0.13	0.13
UP_KW_BIOLOGICAL_PROCESS	KW-0130~Cell adhesion	13	5.28	0.00	Q9WV75, 043586, Q62935, Q28178, A5D7B7, P53710, P16144, P55083, P55281, P22458, B8Q0B2, P49746, Q5EA62	129	219	6212	2.86	0.14	0.16	0.16
GOTERM_MF_DIRECT	GO:0001530~lipopolysaccharide binding	3	1.22	0.08	Q9WV75, O77811, P25105	231	20	9943	6.46	1.00	1.00	1.00
GOTERM_BP_DIRECT	GO:0007155~cell adhesion	12	4.88	0.02	Q9WV75, P16144, 043586, Q62935, Q28178, A5D7B7, Q08345, P55083, P53710, P49746, Q5EA62, B8Q0B2	228	232	9968	2.26	1.00	1.00	0.99
GOTERM_CC_DIRECT	GO:0005615~extracellular space	41	16.67	0.00	Q9WV75, P33617, Q9TTQ4, A5D7B7, O70570, Q8R4A1, Q5R3Z7, Q62000, P14555, Q08345, Q8BV57, Q96CG8, A5D7R1, P00747, P61288, Q3MI99, Q14195, P22458, P02763, P13909, Q5EA62, P27169, P02461, O95025, Q9TT92, P36639, P09917, O77811, Q62935, P52823, Q1LZE9, Q5BIR5, P57756, P14431, P45845, Q8BHE5, Q3TMX7, P48356, Q5RES1, P55083, P04792	236	667	10395	2.71	0.00	0.00	0.00
GOTERM_CC_DIRECT	GO:0031012~extracellular matrix	10	4.07	0.00	Q9WV75, P45845, Q8BV57, Q9TT92, A5D7R1, Q3MI99, Q28178, Q62000, Q5RES1, P51942	236	93	10395	4.74	0.07	0.01	0.01
UP_KW_PTM	KW-1015~Disulfide bond	65	26.42	0.00	Q9WV75, Q28178, Q9TTQ4, A5D7B7, O70570, Q62000, P14555, P51942, Q8BV57, P63125, A5D7R1, P16144, P00747, Q3MI99, P33681, Q5RAL8, Q8R4W6, P22458, P02763, P97391, Q8QZY6, Q5EA62, P27169, Q9TT92, P05979, Q62935, O97664, Q1LZE9, Q13609, P45845, Q8BHE5, O14519, Q02846, Q3TMX7, Q05910, P48356, Q9BYZ8, P33617, O70431, Q8R4A1, P15943, Q5R3Z7, Q08345, Q7Z3S9, Q8N5C1, Q15165, Q9NZQ7, P25105, P21664, P02461, O95025, O75954, O77811, P52823, P51654, O95980, P53710, P57756, P14431, Q5RDI1, Q5RES1, O43194, Q9JJ61, P49746	192	1407	8050	1.94	0.00	0.00	0.00
UP_KW_LIGAND	KW-0106~Calcium	32	13.01	0.00	Q9WV75, Q6UWY0, Q28178, Q5R3Z7, Q7Z3S9, P14555, Q08345, Q96D15, P16144, P61288, Q3MI99, Q15165, P24732, P97391, Q5EA62, P29269, P27169, P02461, Q9TT92, P09917, O43548, A2SW69, P53710, P57756, Q13609, P62762, P08132, Q5RES1, P55083, P55281, Q4R8T1, P49746	111	458	4168	2.62	0.00	0.00	0.00
UP_KW_CELLULAR_COMPONENT	KW-0272~Extracellular matrix	11	4.47	0.00	Q9WV75, Q9TT92, P02461, Q96CG8, Q62935, Q28178, A2SW69, Q62000, P55083, P51942, Q5EA62	226	109	9590	4.28	0.01	0.00	0.00

Table S4.5. Genes overlapping islands of differentiation in BPI-SC, BPI-LI, but not in SC-LI.

ID	Gene Name
P97776	ADAM metallopeptidase domain 18(Adam18)
Q28660	ADAM metallopeptidase domain 2(ADAM2)
Q61072	ADAM metallopeptidase domain 9(Adam9)
Q9UMR7	C-type lectin domain family 4 member A(CLEC4A)
P53569	CCAAT/enhancer binding protein zeta(Cebpz)
A8MTT3	CEBPZ opposite strand(CEBPZOS)
Q7Z3D6	D-glutamate cyclase(DGLUCY)
Q8BH86	D-glutamate cyclase(Dglucy)
D4A612	ELOVL fatty acid elongase 2(Elovl2)
Q920L7	ELOVL fatty acid elongase 5(Elovl5)
Q5R6J0	Ecm29 proteasome adaptor and scaffold(ECPAS)
Q5VYK3	Ecm29 proteasome adaptor and scaffold(ECPAS)
Q6PDI5	Ecm29 proteasome adaptor and scaffold(Ecpas)
Q8BJL1	F-box protein 30(Fbxo30)
P62932	F-box protein 40(Fbxo40)
P16591	FER tyrosine kinase(FER)
P07332	FES proto-oncogene, tyrosine kinase(FES)
P10412	H1.4 linker histone, cluster member(H1-4)
P10854	H2B clustered histone 14(H2bc14)
Q96PU5	NEDD4 like E3 ubiquitin protein ligase(NEDD4L)
Q15042	RAB3 GTPase activating protein catalytic subunit 1(RAB3GAP1)
Q9UBZ9	REV1 DNA directed polymerase(REV1)
O14802	RNA polymerase III subunit A(POLR3A)
Q6NW29	RWD domain containing 4(RWDD4)
D3ZID8	UBX domain protein 2A(Ubxn2a)
Q9H6R7	WD repeat and coiled coil containing(WDCP)
Q6NV72	WD repeat and coiled coil containing(Wdcp)
Q6PJI9	WD repeat domain 59(WDR59)
Q8C0M0	WD repeat domain 59(Wdr59)
Q96I59	asparaginyl-tRNA synthetase 2, mitochondrial(NARS2)
Q5R9I5	aspartyl-tRNA synthetase 1(DARS1)
Q3SYZ4	aspartyl-tRNA synthetase 1(DARS1)
P14868	aspartyl-tRNA synthetase 1(DARS1)
Q922B2	aspartyl-tRNA synthetase 1(Dars1)
P22003	bone morphogenetic protein 5(BMP5)
P34819	bone morphogenetic protein 7(BMP7)
P34819	bone morphogenetic protein 7(BMP7)
Q76LD0	chordin-like 1(Chrdl1)

P59266	fat storage-inducing transmembrane protein 2(Fitm2)
A8MUU1	fatty acid binding protein 5 pseudogene 3(FABP5P3)
Q01469	fatty acid binding protein 5(FABP5)
Q9H2C0	gigaxonin(GAN)
Q5R893	histone H2B type 1-C/E/F/G/I(LOC100441051)
Q9NQS7	inner centromere protein(INCENP)
Q0VC92	mago homolog B, exon junction complex subunit(MAGOHB)
Q2KIZ8	minichromosome maintenance complex component 6(MCM6)
Q2KID9	mitochondrial ribosomal protein S5(MRPS5)
Q15080	neutrophil cytosolic factor 4(NCF4)
Q96S42	nodal growth differentiation factor(NODAL)
Q7M6Y3	phosphatidylinositol binding clathrin assembly protein(Picalm)
O94806	protein kinase D3(PRKD3)
Q8VE37	regulator of chromosome condensation 1(Rcc1)
Q5RAQ8	ribosomal protein S24(RPS24)
O08954	snail family transcriptional repressor 2(Snai2)
Q9XT74	solute carrier family 11 member 1(SLC11A1)
Q9XT74	solute carrier family 11 member 1(SLC11A1)
Q2T9X8	spindle and centriole associated protein 1(SPICE1)
Q5RKG1	spindle and centriole associated protein 1(Spice1)
A6QLC7	trafficking protein particle complex subunit 11(TRAPPC11)
Q7Z392	trafficking protein particle complex subunit 11(TRAPPC11)
B1AZA5	transmembrane protein 245(Tmem245)
D3ZXD8	transmembrane protein 245(Tmem245)
Q5H8C4	vacuolar protein sorting 13A(Vps13a)
Q5RBY9	zinc finger protein 2(ZNF2)
Q96CX3	zinc finger protein 501(ZNF501)

Table S4.6. Genes overlapping islands of differentiation in BPI-SC, BPI-LI, but not in SC-LI top pathways.

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0007030~Golgi organization	4	6.06	0.02	D3ZID8, Q7Z392, Q96CX3, A6QLC7	61	127	14354	7.41	1.00	0.28	0.27
INTERPRO	IPR002549:AI-2E-like	2	3.03	0.01	D3ZXD8, B1AZA5	66	2	15277	231.47	0.85	0.06	0.05
GOTERM_MF_DIRECT	GO:0031492~nucleosomal DNA binding	2	3.03	0.08	P10412, Q8VE37	60	21	14318	22.73	1.00	0.69	0.66
GOTERM_MF_DIRECT	GO:0030527~structural constituent of chromatin	3	4.55	0.01	P10854, P10412, Q5R893	60	42	14318	17.05	0.82	0.21	0.21
GOTERM_CC_DIRECT	GO:0000786~nucleosome	3	4.55	0.04	P10854, P10412, Q5R893	64	78	14900	8.95	1.00	0.70	0.70
UP_KW_PTM	KW-0013~ADP-ribosylation	3	4.55	0.09	P10854, P10412, Q5R893	56	105	11732	5.99	0.63	0.35	0.35
SMART	SM00427:H2B	2	3.03	0.02	P10854, Q5R893	30	5	7685	102.47	0.44	0.12	0.11
INTERPRO	IPR000558:Histone_H2B	2	3.03	0.02	P10854, Q5R893	66	5	15277	92.59	0.99	0.11	0.10
INTERPRO	IPR055333:HISTONE_H2B_site	2	3.03	0.02	P10854, Q5R893	66	5	15277	92.59	0.99	0.11	0.10
UP_SEQ_FEATURE	REGION:Binding site for the 3'-end of tRNA	3	4.55	0.00	P14868, Q3SYZ4, Q5R9I5	64	3	14888	232.63	0.01	0.00	0.00
INTERPRO	IPR004523:Asp-tRNA_synthase_2	4	6.06	0.00	P14868, Q922B2, Q3SYZ4, Q5R9I5	66	4	15277	231.47	0.00	0.00	0.00
GOTERM_MF_DIRECT	GO:0004815~aspartate-tRNA ligase activity	4	6.06	0.00	P14868, Q922B2, Q3SYZ4, Q5R9I5	60	5	14318	190.91	0.00	0.00	0.00
GOTERM_BP_DIRECT	GO:0006422~aspartyl-tRNA aminoacylation	4	6.06	0.00	P14868, Q922B2, Q3SYZ4, Q5R9I5	61	5	14354	188.25	0.00	0.00	0.00
UP_SEQ_FEATURE	REGION:Aspartate	4	6.06	0.00	P14868, Q922B2, Q3SYZ4, Q5R9I5	64	5	14888	186.10	0.00	0.00	0.00
GOTERM_CC_DIRECT	GO:0017101~aminoacyl-tRNA synthetase multienzyme complex	4	6.06	0.00	P14868, Q922B2, Q3SYZ4, Q5R9I5	64	24	14900	38.80	0.02	0.02	0.02
KEGG_PATHWAY	hsa00970:Aminoacyl-tRNA biosynthesis	2	3.03	0.09	P14868, Q96I59	34	20	7412	21.80	1.00	1.00	1.00
UP_SEQ_FEATURE	DOMAIN:OB	3	4.55	0.00	P14868, Q96I59, Q5R9I5	64	5	14888	139.58	0.04	0.01	0.01
GOTERM_MF_DIRECT	GO:0003676~nucleic acid binding	4	6.06	0.10	P14868, Q96I59, Q922B2, Q3SYZ4	60	266	14318	3.59	1.00	0.76	0.73
INTERPRO	IPR002312:Asp/Asn-tRNA-synth_IIb	5	7.58	0.00	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	66	7	15277	165.34	0.00	0.00	0.00
INTERPRO	IPR004364:Aa-tRNA-synt_II	5	7.58	0.00	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	66	11	15277	105.21	0.00	0.00	0.00
INTERPRO	IPR004365:NA-bd_OB_tRNA	5	7.58	0.00	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	66	13	15277	89.03	0.00	0.00	0.00
UP_SEQ_FEATURE	DOMAIN:Aminoacyl-transfer RNA synthetases class-II family profile	5	7.58	0.00	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	64	25	14888	46.53	0.00	0.00	0.00
INTERPRO	IPR006195:aa-tRNA-synth_II	5	7.58	0.00	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	66	30	15277	38.58	0.00	0.00	0.00

INTERPRO	IPR045864:aa-tRNA-synth_II/BPL/LPL	5	7.58	0.00	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	66	38	15277	30.46	0.00	0.00	0.00
UP_KW_MOLECULAR_FUNCTION	KW-0030~Aminoacyl-tRNA synthetase	5	7.58	0.00	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	39	68	9248	17.44	0.00	0.00	0.00
UP_KW_BIOLOGICAL_PROCESS	KW-0648~Protein biosynthesis	5	7.58	0.02	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	40	224	9049	5.05	0.44	0.19	0.19
UP_KW_MOLECULAR_FUNCTION	KW-0436~Ligase	5	7.58	0.02	P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	39	244	9248	4.86	0.38	0.08	0.07
PIR_SUPERFAMILY	PIRSF000632:TyrPK_fps	2	3.03	0.01	P16591, P07332	8	4	2074	129.63	0.09	0.05	0.05
UP_SEQ_FEATURE	REGION:Important for interaction with membranes containing phosphoinositides	2	3.03	0.02	P16591, P07332	64	4	14888	116.31	0.99	0.12	0.11
INTERPRO	IPR035849:Fes/Fps/Fer_SH2	2	3.03	0.02	P16591, P07332	66	4	15277	115.73	0.98	0.10	0.09
INTERPRO	IPR016250:Tyr-prot_kinase_Fes/Fps	2	3.03	0.02	P16591, P07332	66	4	15277	115.73	0.98	0.10	0.09
INTERPRO	IPR050198:Non-receptor_tyrosine_kinases	2	3.03	0.07	P16591, P07332	66	18	15277	25.72	1.00	0.31	0.29
SMART	SM00055:FCH	2	3.03	0.08	P16591, P07332	30	22	7685	23.29	0.92	0.41	0.39
INTERPRO	IPR001060:FCH_dom	2	3.03	0.09	P16591, P07332	66	23	15277	20.13	1.00	0.38	0.35
GOTERM_BP_DIRECT	GO:0021502~neural fold elevation formation	3	4.55	0.00	P34819, P22003	61	6	14354	117.66	0.12	0.04	0.04
GOTERM_BP_DIRECT	GO:1905069~allantois development	3	4.55	0.00	P34819, P22003	61	7	14354	100.85	0.17	0.04	0.04
GOTERM_BP_DIRECT	GO:0003344~pericardium morphogenesis	3	4.55	0.00	P34819, P22003	61	9	14354	78.44	0.27	0.04	0.04
UP_KW_BIOLOGICAL_PROCESS	KW-0891~Chondrogenesis	3	4.55	0.00	P34819, P22003	40	9	9049	75.41	0.02	0.01	0.01
GOTERM_MF_DIRECT	GO:0070700~BMP receptor binding	3	4.55	0.00	P34819, P22003	60	10	14318	71.59	0.09	0.05	0.05
GOTERM_BP_DIRECT	GO:0061384~heart trabecula morphogenesis	3	4.55	0.00	P34819, P22003	61	10	14354	70.59	0.33	0.04	0.04
GOTERM_BP_DIRECT	GO:0060710~chorio-allantoic fusion	3	4.55	0.00	P34819, P22003	61	10	14354	70.59	0.33	0.04	0.04
GOTERM_BP_DIRECT	GO:0060037~pharyngeal system development	3	4.55	0.00	P34819, P22003	61	10	14354	70.59	0.33	0.04	0.04
INTERPRO	IPR001111:TGF-b_propeptide	3	4.55	0.00	P34819, P22003	66	10	15277	69.44	0.16	0.01	0.01
GOTERM_BP_DIRECT	GO:0003272~endocardial cushion formation	3	4.55	0.00	P34819, P22003	61	11	14354	64.18	0.38	0.04	0.04
GOTERM_BP_DIRECT	GO:1900006~positive regulation of dendrite development	3	4.55	0.00	P34819, P22003	61	12	14354	58.83	0.44	0.05	0.05
GOTERM_BP_DIRECT	GO:0030902~hindbrain development	3	4.55	0.00	P34819, P22003	61	15	14354	47.06	0.60	0.06	0.06
GOTERM_BP_DIRECT	GO:0060411~cardiac septum morphogenesis	3	4.55	0.00	P34819, P22003	61	15	14354	47.06	0.60	0.06	0.06
GOTERM_BP_DIRECT	GO:0048738~cardiac muscle tissue development	3	4.55	0.00	P34819, P22003	61	25	14354	28.24	0.92	0.13	0.13
GOTERM_BP_DIRECT	GO:0051216~cartilage development	3	4.55	0.02	P34819, P22003	61	49	14354	14.41	1.00	0.28	0.27

GOTERM_BP_DIRECT	GO:0009880~embryonic pattern specification	3	4.55	0.00	P34819, Q96S42	61	15	14354	47.06	0.60	0.06	0.06
UP_SEQ_FEATURE	DOMAIN:TGF-beta family profile	4	6.06	0.00	P34819, Q96S42, P22003	64	15	14888	62.03	0.01	0.00	0.00
SMART	SM00204:TGFB	4	6.06	0.00	P34819, Q96S42, P22003	30	19	7685	53.93	0.00	0.00	0.00
INTERPRO	IPR015615:TGF-beta-rel	4	6.06	0.00	P34819, Q96S42, P22003	66	19	15277	48.73	0.01	0.00	0.00
INTERPRO	IPR017948:TGFb_CS	4	6.06	0.00	P34819, Q96S42, P22003	66	19	15277	48.73	0.01	0.00	0.00
INTERPRO	IPR001839:TGF-b_C	4	6.06	0.00	P34819, Q96S42, P22003	66	19	15277	48.73	0.01	0.00	0.00
GOTERM_BP_DIRECT	GO:0060391~positive regulation of SMAD protein signal transduction	4	6.06	0.00	P34819, Q96S42, P22003	61	33	14354	28.52	0.17	0.04	0.04
INTERPRO	IPR029034:Cystine-knot_cytokine	4	6.06	0.00	P34819, Q96S42, P22003	66	38	15277	24.37	0.12	0.01	0.01
UP_KW_MOLECULAR_FUNCTION	KW-0202~Cytokine	4	6.06	0.00	P34819, Q96S42, P22003	39	57	9248	16.64	0.04	0.02	0.02
GOTERM_MF_DIRECT	GO:0005125~cytokine activity	4	6.06	0.00	P34819, Q96S42, P22003	60	58	14318	16.46	0.21	0.08	0.07
UP_KW_MOLECULAR_FUNCTION	KW-0339~Growth factor	4	6.06	0.00	P34819, Q96S42, P22003	39	75	9248	12.65	0.09	0.03	0.03
GOTERM_MF_DIRECT	GO:0008083~growth factor activity	4	6.06	0.01	P34819, Q96S42, P22003	60	94	14318	10.15	0.60	0.15	0.15
UP_KW_PTM	KW-0165~Cleavage on pair of basic residues	4	6.06	0.04	P34819, Q96S42, P22003	56	161	11732	5.20	0.36	0.34	0.34
GOTERM_BP_DIRECT	GO:0034389~lipid droplet organization	2	3.03	0.09	P59266, Q15042	61	22	14354	21.39	1.00	0.53	0.52
INTERPRO	IPR000463:Fatty_acid-bd	2	3.03	0.04	Q01469, A8MUU1	66	9	15277	51.44	1.00	0.17	0.16
INTERPRO	IPR031259:ILBP	2	3.03	0.04	Q01469, A8MUU1	66	9	15277	51.44	1.00	0.17	0.16
INTERPRO	IPR000566:Lipocln_cytosolic_FA-bd_dom	2	3.03	0.09	Q01469, A8MUU1	66	23	15277	20.13	1.00	0.38	0.35
UP_KW_PTM	KW-0007~Acetylation	27	40.91	0.06	Q0VC92, Q920L7, Q6NW29, P10854, Q5VYK3, P10412, Q96I59, Q922B2, Q5R893, B1AZA5, A6QLC7, Q7Z392, O14802, Q01469, Q96PU5, Q2KIZ8, D3ZXD8, Q9H6R7, Q8BH86, P14868, Q28660, Q7M6Y3, Q3SYZ4, Q5RAQ8, Q5R9I5, Q6NV72	56	4224	11732	1.34	0.47	0.34	0.34
UP_KW_LIGAND	KW-0446~Lipid-binding	5	7.58	0.02	Q15080, Q01469, P16591, P07332, A8MUU1	25	283	6178	4.37	0.18	0.20	0.20
UP_KW_CELLULAR_COMPONENT	KW-0967~Endosome	8	12.12	0.04	Q15080, Q5H8C4, Q6PDI5, Q5VYK3, Q9XT74, Q5R6J0, Q96PU5	61	729	13800	2.48	0.66	0.56	0.56
GOTERM_BP_DIRECT	GO:0045730~respiratory burst	3	4.55	0.00	Q15080, Q9XT74	61	10	14354	70.59	0.33	0.04	0.04

GOTERM_BP_DIRECT	GO:0006909~phagocytosis	3	4.55	0.04	Q15080, Q9XT74	61	77	14354	9.17	1.00	0.37	0.36
UP_KW_CELLULAR_COMPONENT	KW-0158~Chromosome	6	9.09	0.06	Q2KIZ8, P10854, P10412, Q5R893, Q9NQS7, Q8VE37	61	487	13800	2.79	0.82	0.56	0.56
UP_KW_MOLECULAR_FUNCTION	KW-0238~DNA-binding	9	13.64	0.10	Q2KIZ8, P10854, Q96CX3, Q9UBZ9, P10412, O08954, Q5R893, Q5RBY9, Q8VE37	39	1163	9248	1.84	0.94	0.37	0.33
INTERPRO	IPR012340:NA-bd_OB-fold	6	9.09	0.00	Q2KIZ8, P14868, Q96I59, Q922B2, Q3SYZ4, Q5R9I5	66	102	15277	13.62	0.02	0.00	0.00
INTERPRO	IPR031387:SPICE1	2	3.03	0.01	Q2T9X8, Q5RKG1	66	2	15277	231.47	0.85	0.06	0.05
GOTERM_BP_DIRECT	GO:0046599~regulation of centriole replication	2	3.03	0.02	Q2T9X8, Q5RKG1	61	6	14354	78.44	1.00	0.29	0.29
GOTERM_BP_DIRECT	GO:0051310~metaphase chromosome alignment	3	4.55	0.00	Q2T9X8, Q5RKG1, Q9NQS7	61	9	14354	78.44	0.27	0.04	0.04
GOTERM_BP_DIRECT	GO:0090307~mitotic spindle assembly	3	4.55	0.02	Q2T9X8, Q5RKG1, Q9NQS7	61	45	14354	15.69	1.00	0.28	0.27
UP_SEQ_FEATURE	REGION:Disordered	47	71.21	0.07	Q5H8C4, Q6NW29, P10854, Q5VYK3, Q9UMR7, Q5R6J0, Q2KID9, P10412, Q96I59, P53569, Q5R893, B1AZA5, Q5RKG1, Q9NQS7, Q15080, Q76LD0, Q6PDI5, Q96PU5, P62932, Q8BJL1, Q6PJI9, P07332, Q15042, O94806, Q2KIZ8, D3ZXD8, Q9XT74, P34819, Q2T9X8, Q9H6R7, Q8BH86, Q28660, Q7M6Y3, O08954, Q8VE37, Q5RAQ8, P16591, Q61072, Q9UBZ9, P22003, Q7Z3D6, Q8C0M0, Q5RBY9, Q6NV72	64	9406	14888	1.16	1.00	0.37	0.33
GOTERM_BP_DIRECT	GO:0010467~gene expression	4	6.06	0.03	Q5H8C4, Q8BJL1, Q28660	61	153	14354	6.15	1.00	0.30	0.29
GOTERM_CC_DIRECT	GO:0005765~lysosomal membrane	5	7.58	0.09	Q5H8C4, Q9XT74, Q6PJI9, Q8C0M0	64	396	14900	2.94	1.00	0.90	0.90
GOTERM_BP_DIRECT	GO:0006412~translation	4	6.06	0.10	Q5RAQ8, Q2KID9, P14868, Q96I59	61	261	14354	3.61	1.00	0.57	0.55
INTERPRO	IPR018358:Disintegrin_CS	4	6.06	0.00	Q61072, P97776, Q28660	66	8	15277	115.73	0.00	0.00	0.00
INTERPRO	IPR034027:Reprolysin_adamalysin	4	6.06	0.00	Q61072, P97776, Q28660	66	10	15277	92.59	0.00	0.00	0.00
SMART	SM00050:DISIN	4	6.06	0.00	Q61072, P97776, Q28660	30	13	7685	78.82	0.00	0.00	0.00
INTERPRO	IPR002870:Peptidase_M12B_N	4	6.06	0.00	Q61072, P97776, Q28660	66	12	15277	77.16	0.00	0.00	0.00

SMART	SM00608:ACR	4	6.06	0.00	Q61072, P97776, Q28660	30	14	7685	73.19	0.00	0.00	0.00
INTERPRO	IPR036436:Disintegrin_dom_sf	4	6.06	0.00	Q61072, P97776, Q28660	66	13	15277	71.22	0.00	0.00	0.00
INTERPRO	IPR001762:Disintegrin_dom	4	6.06	0.00	Q61072, P97776, Q28660	66	13	15277	71.22	0.00	0.00	0.00
INTERPRO	IPR006586:ADAM_Cys-rich	4	6.06	0.00	Q61072, P97776, Q28660	66	14	15277	66.13	0.01	0.00	0.00
UP_SEQ_FEATURE	DOMAIN:Disintegrin	4	6.06	0.00	Q61072, P97776, Q28660	64	17	14888	54.74	0.01	0.00	0.00
UP_SEQ_FEATURE	DOMAIN:Peptidase M12B	4	6.06	0.00	Q61072, P97776, Q28660	64	17	14888	54.74	0.01	0.00	0.00
INTERPRO	IPR001590:Peptidase_M12B	4	6.06	0.00	Q61072, P97776, Q28660	66	17	15277	54.46	0.01	0.00	0.00
INTERPRO	IPR024079:MetalloPept_cat_dom_sf	4	6.06	0.00	Q61072, P97776, Q28660	66	40	15277	23.15	0.13	0.01	0.01
GOTERM_MF_DIRECT	GO:0004222~metalloendopeptidase activity	4	6.06	0.00	Q61072, P97776, Q28660	60	69	14318	13.83	0.32	0.10	0.09
UP_SEQ_FEATURE	DOMAIN:EGF-like	4	6.06	0.03	Q61072, P97776, Q28660	64	150	14888	6.20	1.00	0.17	0.15
INTERPRO	IPR000742:EGF-like_dom	4	6.06	0.03	Q61072, P97776, Q28660	66	157	15277	5.90	1.00	0.14	0.13
UP_KW_DOMAIN	KW-0245~EGF-like domain	4	6.06	0.03	Q61072, P97776, Q28660	45	164	10178	5.52	0.38	0.47	0.47
SMART	SM00591:RWD	3	4.55	0.00	Q6NW29, Q6PJI9, Q8C0M0	30	11	7685	69.86	0.02	0.01	0.01
UP_SEQ_FEATURE	DOMAIN:RWD	3	4.55	0.00	Q6NW29, Q6PJI9, Q8C0M0	64	11	14888	63.44	0.22	0.02	0.01
INTERPRO	IPR006575:RWD_dom	3	4.55	0.00	Q6NW29, Q6PJI9, Q8C0M0	66	12	15277	57.87	0.22	0.01	0.01
INTERPRO	IPR016135:UBQ-conjugating_enzyme/RWD	3	4.55	0.02	Q6NW29, Q6PJI9, Q8C0M0	66	56	15277	12.40	0.99	0.12	0.11
UP_SEQ_FEATURE	DOMAIN:Proteasome component Ecm29 N-terminal	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	3	14888	232.63	0.01	0.00	0.00
INTERPRO	IPR024372:Ecm29	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	66	3	15277	231.47	0.01	0.00	0.00
UP_SEQ_FEATURE	REPEAT:HEAT 27	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	6	14888	116.31	0.07	0.01	0.01
UP_SEQ_FEATURE	REPEAT:HEAT 25	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	9	14888	77.54	0.15	0.01	0.01
UP_SEQ_FEATURE	REPEAT:HEAT 26	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	9	14888	77.54	0.15	0.01	0.01
UP_SEQ_FEATURE	REPEAT:HEAT 22	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	10	14888	69.79	0.18	0.01	0.01
UP_SEQ_FEATURE	REPEAT:HEAT 21	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	10	14888	69.79	0.18	0.01	0.01
UP_SEQ_FEATURE	REPEAT:HEAT 24	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	10	14888	69.79	0.18	0.01	0.01
UP_SEQ_FEATURE	REPEAT:HEAT 23	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	10	14888	69.79	0.18	0.01	0.01
GOTERM_BP_DIRECT	GO:0043248~proteasome assembly	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	61	15	14354	47.06	0.60	0.06	0.06
UP_SEQ_FEATURE	REPEAT:HEAT 20	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	15	14888	46.53	0.37	0.03	0.02
UP_SEQ_FEATURE	REPEAT:HEAT 19	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	17	14888	41.05	0.45	0.03	0.03
UP_SEQ_FEATURE	REPEAT:HEAT 17	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	17	14888	41.05	0.45	0.03	0.03

UP_SEQ_FEATURE	REPEAT:HEAT 18	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	17	14888	41.05	0.45	0.03	0.03
UP_SEQ_FEATURE	REPEAT:HEAT 16	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	19	14888	36.73	0.53	0.04	0.03
UP_SEQ_FEATURE	REPEAT:HEAT 15	3	4.55	0.00	Q6PDI5, Q5VYK3, Q5R6J0	64	24	14888	29.08	0.70	0.05	0.05
UP_SEQ_FEATURE	REPEAT:HEAT 14	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	64	26	14888	26.84	0.75	0.06	0.05
GOTERM_MF_DIRECT	GO:0070628~proteasome binding	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	60	27	14318	26.51	0.52	0.15	0.14
UP_SEQ_FEATURE	REPEAT:HEAT 13	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	64	27	14888	25.85	0.78	0.06	0.06
UP_SEQ_FEATURE	REPEAT:HEAT 12	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	64	28	14888	24.92	0.80	0.06	0.06
UP_SEQ_FEATURE	REPEAT:HEAT 11	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	64	30	14888	23.26	0.84	0.07	0.06
UP_SEQ_FEATURE	REPEAT:HEAT 10	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	64	33	14888	21.15	0.89	0.07	0.06
UP_SEQ_FEATURE	REPEAT:HEAT 9	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	64	35	14888	19.94	0.92	0.08	0.07
UP_SEQ_FEATURE	REPEAT:HEAT 8	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	64	40	14888	17.45	0.96	0.10	0.09
UP_SEQ_FEATURE	REPEAT:HEAT 7	3	4.55	0.01	Q6PDI5, Q5VYK3, Q5R6J0	64	43	14888	16.23	0.98	0.11	0.10
UP_SEQ_FEATURE	REPEAT:HEAT 6	3	4.55	0.02	Q6PDI5, Q5VYK3, Q5R6J0	64	53	14888	13.17	1.00	0.15	0.13
UP_SEQ_FEATURE	REPEAT:HEAT 5	3	4.55	0.03	Q6PDI5, Q5VYK3, Q5R6J0	64	58	14888	12.03	1.00	0.17	0.15
GOTERM_CC_DIRECT	GO:0030134~COPII-coated ER to Golgi transport vesicle	3	4.55	0.03	Q6PDI5, Q5VYK3, Q5R6J0	64	59	14900	11.84	0.97	0.70	0.70
UP_SEQ_FEATURE	REPEAT:HEAT 4	3	4.55	0.03	Q6PDI5, Q5VYK3, Q5R6J0	64	67	14888	10.42	1.00	0.21	0.19
UP_SEQ_FEATURE	REPEAT:HEAT 3	3	4.55	0.04	Q6PDI5, Q5VYK3, Q5R6J0	64	72	14888	9.69	1.00	0.24	0.21
GOTERM_CC_DIRECT	GO:0030139~endocytic vesicle	3	4.55	0.04	Q6PDI5, Q5VYK3, Q5R6J0	64	77	14900	9.07	1.00	0.70	0.70
GOTERM_CC_DIRECT	GO:0000502~proteasome complex	3	4.55	0.04	Q6PDI5, Q5VYK3, Q5R6J0	64	78	14900	8.95	1.00	0.70	0.70
UP_SEQ_FEATURE	CROSSLNK:Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in SUMO1)	3	4.55	0.04	Q6PDI5, Q5VYK3, Q5R6J0	64	79	14888	8.83	1.00	0.27	0.24
UP_SEQ_FEATURE	REPEAT:HEAT 1	3	4.55	0.05	Q6PDI5, Q5VYK3, Q5R6J0	64	82	14888	8.51	1.00	0.28	0.25
UP_SEQ_FEATURE	REPEAT:HEAT 2	3	4.55	0.05	Q6PDI5, Q5VYK3, Q5R6J0	64	82	14888	8.51	1.00	0.28	0.25
UP_KW_CELLULAR_COMPONENT	KW-0647~Proteasome	3	4.55	0.05	Q6PDI5, Q5VYK3, Q5R6J0	61	80	13800	8.48	0.74	0.56	0.56
GOTERM_CC_DIRECT	GO:0005793~endoplasmic reticulum-Golgi intermediate compartment	3	4.55	0.07	Q6PDI5, Q5VYK3, Q5R6J0	64	105	14900	6.65	1.00	0.87	0.87
GOTERM_BP_DIRECT	GO:0036503~ERAD pathway	3	4.55	0.08	Q6PDI5, Q5VYK3, Q5R6J0	61	110	14354	6.42	1.00	0.52	0.50
GOTERM_CC_DIRECT	GO:0005771~multivesicular body	4	6.06	0.00	Q6PDI5, Q5VYK3, Q5R6J0, Q96PU5	64	50	14900	18.63	0.15	0.08	0.08
INTERPRO	IPR049567:WDR59-like	2	3.03	0.01	Q6PJI9, Q8C0M0	66	2	15277	231.47	0.85	0.06	0.05
INTERPRO	IPR049566:WDR59_RTC1-like_RING_Znf	2	3.03	0.01	Q6PJI9, Q8C0M0	66	2	15277	231.47	0.85	0.06	0.05

INTERPRO	IPR039456:WDR59_mRING-H2-C3H3C2	2	3.03	0.01	Q6PJI9, Q8C0M0	66	2	15277	231.47	0.85	0.06	0.05
GOTERM_CC_DIRECT	GO:0035859~Seh1-associated complex	2	3.03	0.02	Q6PJI9, Q8C0M0	64	5	14900	93.13	0.94	0.70	0.70
GOTERM_CC_DIRECT	GO:0061700~GATOR2 complex	2	3.03	0.06	Q6PJI9, Q8C0M0	64	15	14900	31.04	1.00	0.87	0.87
GOTERM_CC_DIRECT	GO:0005774~vacuolar membrane	2	3.03	0.09	Q6PJI9, Q8C0M0	64	22	14900	21.16	1.00	0.90	0.90
UP_KW_BIOLOGICAL_PROCESS	KW-0892~Osteogenesis	4	6.06	0.00	Q76LD0, P34819, P22003	40	34	9049	26.61	0.02	0.01	0.01
GOTERM_BP_DIRECT	GO:0030509~BMP signaling pathway	4	6.06	0.00	Q76LD0, P34819, P22003	61	61	14354	15.43	0.67	0.06	0.06
GOTERM_BP_DIRECT	GO:0001503~ossification	4	6.06	0.00	Q76LD0, P34819, P22003	61	78	14354	12.07	0.89	0.12	0.12
UP_KW_BIOLOGICAL_PROCESS	KW-0221~Differentiation	6	9.09	0.05	Q76LD0, P34819, Q96PU5, P97776, P22003	40	463	9049	2.93	0.84	0.45	0.44
UP_KW_MOLECULAR_FUNCTION	KW-9996~Developmental protein	8	12.12	0.01	Q76LD0, P34819, Q96S42, P97776, P22003, Q7M6Y3, O08954	39	555	9248	3.42	0.17	0.04	0.03
UP_KW_MOLECULAR_FUNCTION	KW-0217~Developmental protein	8	12.12	0.01	Q76LD0, P34819, Q96S42, P97776, P22003, Q7M6Y3, O08954	39	555	9248	3.42	0.17	0.04	0.03
GOTERM_CC_DIRECT	GO:0045202~synapse	6	9.09	0.03	Q76LD0, Q01469, P14868, Q922B2, Q3SYZ4, Q5R9I5	64	420	14900	3.33	0.99	0.70	0.70
UP_SEQ_FEATURE	DOMAIN:Trafficking protein particle complex subunit 11	2	3.03	0.01	Q7Z392, A6QLC7	64	2	14888	232.62	0.89	0.07	0.06
UP_SEQ_FEATURE	DOMAIN:Trafficking protein particle complex subunit 11 C-terminal	2	3.03	0.01	Q7Z392, A6QLC7	64	2	14888	232.62	0.89	0.07	0.06
INTERPRO	IPR021773:TPC11	2	3.03	0.01	Q7Z392, A6QLC7	66	2	15277	231.47	0.85	0.06	0.05
INTERPRO	IPR025876:TRAPPC11_C	2	3.03	0.01	Q7Z392, A6QLC7	66	2	15277	231.47	0.85	0.06	0.05
GOTERM_BP_DIRECT	GO:0045054~constitutive secretory pathway	2	3.03	0.02	Q7Z392, A6QLC7	61	5	14354	94.12	1.00	0.28	0.27
GOTERM_CC_DIRECT	GO:0030008~TRAPP complex	2	3.03	0.07	Q7Z392, A6QLC7	64	17	14900	27.39	1.00	0.87	0.87
GOTERM_BP_DIRECT	GO:0061635~regulation of protein complex stability	2	3.03	0.09	Q7Z392, A6QLC7	61	22	14354	21.39	1.00	0.53	0.52
PIR_SUPERFAMILY	PIRSF037204:UCP037204	2	3.03	0.01	Q8BH86, Q7Z3D6	8	2	2074	259.25	0.05	0.05	0.05
GOTERM_MF_DIRECT	GO:0047820~D-glutamate cyclase activity	2	3.03	0.01	Q8BH86, Q7Z3D6	60	2	14318	238.63	0.67	0.16	0.15
UP_SEQ_FEATURE	DOMAIN:D-glutamate cyclase-like C-terminal	2	3.03	0.01	Q8BH86, Q7Z3D6	64	2	14888	232.62	0.89	0.07	0.06
INTERPRO	IPR017135:D-Glu_cyclase_mito	2	3.03	0.01	Q8BH86, Q7Z3D6	66	2	15277	231.47	0.85	0.06	0.05
INTERPRO	IPR025504:GLUCM_C	2	3.03	0.01	Q8BH86, Q7Z3D6	66	2	15277	231.47	0.85	0.06	0.05
INTERPRO	IPR009906:D-Glu_cyclase	2	3.03	0.01	Q8BH86, Q7Z3D6	66	2	15277	231.47	0.85	0.06	0.05
INTERPRO	IPR038021:Putative_hydro-lyase	2	3.03	0.01	Q8BH86, Q7Z3D6	66	2	15277	231.47	0.85	0.06	0.05

GOTERM_BP_DIRECT	GO:0006536~glutamate metabolic process	2	3.03	0.09	Q8BH86, Q7Z3D6	61	22	14354	21.39	1.00	0.53	0.52
GOTERM_BP_DIRECT	GO:0030261~chromosome condensation	2	3.03	0.05	Q8BJL1, P10412	61	12	14354	39.22	1.00	0.42	0.41
INTERPRO	IPR031890:Fbxo30/Fbxo40	2	3.03	0.01	Q8BJL1, P62932	66	3	15277	154.31	0.94	0.08	0.07
INTERPRO	IPR043013:Znf_TRAF_N	2	3.03	0.01	Q8BJL1, P62932	66	3	15277	154.31	0.94	0.08	0.07
INTERPRO	IPR001293:Znf_TRAF	2	3.03	0.04	Q8BJL1, P62932	66	10	15277	46.29	1.00	0.19	0.17
UP_SEQ_FEATURE	ZN_FING:TRAF-type	2	3.03	0.06	Q8BJL1, P62932	64	14	14888	33.23	1.00	0.33	0.29
GOTERM_BP_DIRECT	GO:0007059~chromosome segregation	3	4.55	0.04	Q8BJL1, Q9NQS7, Q8VE37	61	76	14354	9.29	1.00	0.37	0.36
KEGG_PATHWAY	rno00062:Fatty acid elongation	2	3.03	0.02	Q920L7, D4A612	34	4	7412	109.00	0.73	1.00	1.00
GOTERM_MF_DIRECT	GO:0009922~fatty acid elongase activity	2	3.03	0.02	Q920L7, D4A612	60	6	14318	79.54	0.96	0.33	0.31
GOTERM_BP_DIRECT	GO:0034626~fatty acid elongation, polyunsaturated fatty acid	2	3.03	0.02	Q920L7, D4A612	61	6	14354	78.44	1.00	0.29	0.29
GOTERM_BP_DIRECT	GO:0034625~fatty acid elongation, monounsaturated fatty acid	2	3.03	0.02	Q920L7, D4A612	61	6	14354	78.44	1.00	0.29	0.29
INTERPRO	IPR002076:ELO_fam	2	3.03	0.03	Q920L7, D4A612	66	6	15277	77.16	1.00	0.13	0.12
GOTERM_BP_DIRECT	GO:0019367~fatty acid elongation, saturated fatty acid	2	3.03	0.03	Q920L7, D4A612	61	7	14354	67.23	1.00	0.31	0.30
KEGG_PATHWAY	rno01040:Biosynthesis of unsaturated fatty acids	2	3.03	0.03	Q920L7, D4A612	34	7	7412	62.29	0.90	1.00	1.00
KEGG_PATHWAY	rno01212:Fatty acid metabolism	2	3.03	0.04	Q920L7, D4A612	34	10	7412	43.60	0.96	1.00	1.00
GOTERM_BP_DIRECT	GO:0042761~very long-chain fatty acid biosynthetic process	2	3.03	0.06	Q920L7, D4A612	61	14	14354	33.62	1.00	0.46	0.44
GOTERM_BP_DIRECT	GO:0006636~unsaturated fatty acid biosynthetic process	2	3.03	0.06	Q920L7, D4A612	61	16	14354	29.41	1.00	0.46	0.45
GOTERM_BP_DIRECT	GO:0030148~sphingolipid biosynthetic process	3	4.55	0.01	Q920L7, D4A612, 094806	61	28	14354	25.21	0.96	0.16	0.15
GOTERM_BP_DIRECT	GO:2000009~negative regulation of protein localization to cell surface	2	3.03	0.06	Q96PU5, Q7M6Y3	61	14	14354	33.62	1.00	0.46	0.44
GOTERM_BP_DIRECT	GO:1903861~positive regulation of dendrite extension	2	3.03	0.06	Q96PU5, Q7M6Y3	61	15	14354	31.37	1.00	0.46	0.45
GOTERM_BP_DIRECT	GO:0007281~germ cell development	2	3.03	0.10	Q96S42, P16591	61	25	14354	18.82	1.00	0.58	0.56
UP_SEQ_FEATURE	REGION:Interaction with HCK	2	3.03	0.01	Q9H6R7, Q6NV72	64	2	14888	232.62	0.89	0.07	0.06
INTERPRO	IPR028041:WDCP	2	3.03	0.01	Q9H6R7, Q6NV72	66	2	15277	231.47	0.85	0.06	0.05
GOTERM_BP_DIRECT	GO:0007155~cell adhesion	6	9.09	0.01	Q9UMR7, P16591, Q61072, P07332, Q28660	61	338	14354	4.18	1.00	0.27	0.26
GOTERM_BP_DIRECT	GO:0032496~response to lipopolysaccharide	3	4.55	0.09	Q9XT74, P16591	61	118	14354	5.98	1.00	0.53	0.52
GOTERM_BP_DIRECT	GO:0060586~multicellular organismal-level iron ion homeostasis	3	4.55	0.01	Q9XT74, Q7M6Y3	61	37	14354	19.08	1.00	0.25	0.24

GOTERM_BP_DIRECT	GO:0006879~intracellular iron ion homeostasis	3	4.55	0.06	Q9XT74, Q7M6Y3	61	95	14354	7.43	1.00	0.46	0.45
GOTERM_BP_DIRECT	GO:0001818~negative regulation of cytokine production	3	4.55	0.01	Q9XT74, Q9UMR7	61	31	14354	22.77	0.98	0.18	0.18

Table S4.7. HBD segments > 500 kb.

id	start_snp	end_snp	start_pos	end_pos	# snps	length	HBD class	population	scaffold
8	2096539	2132904	75947985	77535207	36366	1587223	9	SC	11
19	491799	505669	25449597	26580397	13871	1130801	9	BPI	1
9	570614	607775	15379906	16438767	37162	1058862	9	SC	19
2	1739028	1765655	58544990	59539951	26628	994962	9	BPI	15
1	1849208	1870579	142409111	143397282	21372	988172	9	BPI	1
19	1873413	1904221	50287995	51237204	30809	949210	9	BPI	5
19	284066	295577	11595855	12496932	11512	901078	9	BPI	1
11	1070936	1084527	87692957	88555392	13592	862436	9	SC	1
19	1506126	1515971	121655310	122511448	9846	856139	9	BPI	1
4	453491	484202	18671172	19524812	30712	853641	9	LI	10
13	1512756	1521456	122228261	122998619	8701	770359	9	SC	1
5	817880	824276	62811122	63579424	6397	768303	9	LI	1
2	60663	72009	4539365	5305338	11347	765974	9	BPI	10
9	475371	485025	24179644	24943658	9655	764015	9	SC	1
15	470724	479954	23795619	24552574	9231	756956	9	BPI	1
2	1439605	1466698	44628307	45385093	27094	756787	9	BPI	20
17	1023039	1032501	84276490	85031597	9463	755108	9	BPI	1
10	1503241	1511595	121375279	122128091	8355	752813	9	LI	1
19	600807	621329	30666489	31413421	20523	746933	9	BPI	12
1	487422	496517	25131853	25849290	9096	717438	9	BPI	1
6	230292	252637	7028294	7742799	22346	714506	9	LI	41
20	3026030	3045791	89910916	90624193	19762	713278	9	BPI	9
19	471454	479954	23842457	24552574	8501	710118	9	BPI	1
17	471475	479954	23843842	24552574	8480	708733	9	BPI	1
16	471497	479954	23844400	24552574	8458	708175	9	BPI	1
20	471497	479954	23844400	24552574	8458	708175	9	BPI	1
6	471497	479954	23844400	24552574	8458	708175	9	LI	1
4	1240210	1248827	100507901	101212317	8618	704417	9	LI	1
6	1071465	1082865	87725257	88429294	11401	704038	9	LI	1
5	989695	1006521	38903058	39601991	16827	698934	9	LI	19
19	1372782	1382804	111771524	112461235	10023	689712	9	BPI	1
19	1659475	1671938	131859190	132548535	12464	689346	9	BPI	1
1	2107517	2126740	73382347	74066517	19224	684171	9	BPI	3
1	1147239	1158747	44550774	45224929	11509	674156	9	BPI	16
9	2529256	2551439	91483846	92155487	22184	671642	9	SC	2
20	95724	104771	3717574	4388973	9048	671400	9	BPI	17
1	1471573	1479969	118590396	119261394	8397	670999	9	BPI	1

4	909587	915353	73486691	74154603	5767	667913	9	LI	1
17	1851410	1866558	142514320	143181979	15149	667660	9	BPI	1
19	421633	430905	20719768	21386267	9273	666500	9	BPI	1
9	231777	252637	7080080	7742799	20861	662720	9	SC	41
19	203068	214068	6823004	7482475	11001	659472	9	BPI	1
9	2576886	2597846	92940510	93599946	20961	659437	9	SC	2
19	1749147	1771793	46470942	47127381	22647	656440	9	BPI	5
20	287054	295585	11845949	12497347	8532	651399	9	BPI	1
1	2010775	2028553	70050961	70700770	17779	649810	9	BPI	3
1	789248	793564	59256765	59898373	4317	641609	9	BPI	1
20	119253	125446	8796791	9435619	6194	638829	9	BPI	27
1	784628	793239	27360093	27998595	8612	638503	9	BPI	26
11	1132853	1155959	44159824	44794344	23107	634521	9	SC	16
1	1439080	1455753	41112768	41744550	16674	631783	9	BPI	7
14	2847958	2864963	103167343	103794779	17006	627437	9	BPI	3
3	76937	84154	5640924	6265202	7218	624279	9	BPI	10
1	1770222	1779705	137893381	138514853	9484	621473	9	BPI	1
13	1384418	1393371	112571947	113190865	8954	618919	9	SC	1
4	1659112	1670379	131845391	132462951	11268	617561	9	LI	1
4	819921	824998	63052493	63668307	5078	615815	9	LI	1
19	1782072	1799413	60094005	60708832	17342	614828	9	BPI	15
1	2862360	2880495	95072756	95686727	18136	613972	9	BPI	6
19	1553781	1566548	125049182	125663099	12768	613918	9	BPI	1
1	1360198	1368476	110788894	111401489	8279	612596	9	BPI	1
13	703199	729856	16615072	17225500	26658	610429	9	SC	22
19	2054625	2072296	61489781	62098229	17672	608449	9	BPI	8
10	487917	495256	25155653	25763103	7340	607451	9	LI	1
19	2072357	2092382	62099585	62706245	20026	606661	9	BPI	8
20	630494	634745	37642869	38246470	4252	603602	9	BPI	1
19	2030974	2038756	151137370	151733900	7783	596531	9	BPI	1
1	908070	918904	31646215	32240135	10835	593921	9	BPI	29
10	213865	227827	6546786	7138752	13963	591967	9	LI	13
6	1073724	1088445	44408732	44997950	14722	589219	9	LI	2
1	1383721	1392328	112529790	113118419	8608	588630	9	BPI	1
1	1505219	1511964	121568712	122155819	6746	587108	9	BPI	1
11	819356	824282	62993363	63580283	4927	586921	9	SC	1
6	72727	83331	5013426	5599910	10605	586485	9	LI	3
19	1383818	1392323	112534328	113118124	8506	583797	9	BPI	1
7	1855300	1867817	142663348	143247137	12518	583790	9	LI	1
5	1275469	1287606	42827134	43409860	12138	582727	9	LI	10

18	2570193	2588361	75747829	76327705	18169	579877	9	BPI	9
2	2034715	2042753	151421306	151998299	8039	576994	9	BPI	1
2	1851432	1864872	142514921	143090826	13441	575906	9	BPI	1
1	2834047	2854904	108102164	108677057	20858	574894	9	BPI	4
5	645140	657579	29178935	29753311	12440	574377	9	LI	18
16	1982171	2000732	59556423	60129319	18562	572897	9	BPI	13
14	2037674	2055532	73447053	74019829	17859	572777	9	BPI	2
13	2038339	2047245	151695055	152267481	8907	572427	9	SC	1
9	1946700	1958637	146991490	147563341	11938	571852	9	SC	1
14	1745901	1765486	46377687	46946271	19586	568585	9	BPI	5
10	1792584	1802781	139313798	139882250	10198	568453	9	LI	1
1	1438564	1457863	44603344	45169312	19300	565969	9	BPI	20
14	630833	634762	37684330	38249478	3930	565149	9	BPI	1
5	1514282	1520380	122349489	122913819	6099	564331	9	LI	1
8	2069619	2089103	75194292	75756824	19485	562533	9	SC	11
1	1541537	1556179	47739130	48301056	14643	561927	9	BPI	6
1	231633	248447	7073965	7635397	16815	561433	9	BPI	41
18	2110483	2128015	63201308	63761638	17533	560331	9	BPI	8
19	1360414	1375245	47033584	47593746	14832	560163	9	BPI	10
19	738576	742159	52536890	53097036	3584	560147	9	BPI	1
19	1098828	1102775	41824714	42384349	3948	559636	9	BPI	26
13	425399	432846	20943885	21502213	7448	558329	9	SC	1
6	3323325	3323707	105029053	105587279	383	558227	9	LI	7
1	1516795	1523812	122606060	123163659	7018	557600	9	BPI	1
14	1792654	1802618	139316348	139871122	9965	554775	9	BPI	1
20	1852136	1864929	142540874	143094064	12794	553191	9	BPI	1
2	1875823	1889029	143611309	144164470	13207	553162	9	BPI	1
8	1075930	1084527	88003482	88555392	8598	551911	9	SC	1
10	799011	803692	60621492	61172723	4682	551232	9	LI	1
7	2604811	2619908	93831168	94381420	15098	550253	9	LI	2
15	1360198	1367575	110788894	111338471	7378	549578	9	BPI	1
10	1298959	1304020	105623291	106172010	5062	548720	9	LI	1
2	491404	497831	25407893	25955347	6428	547455	9	BPI	1
13	486742	493802	25081367	25627236	7061	545870	9	SC	1
19	590283	607544	16378387	16923601	17262	545215	9	BPI	6
8	3637132	3658714	126679037	127223648	21583	544612	9	SC	4
9	1361384	1368537	110864982	111409048	7154	544067	9	SC	1
3	2494896	2510359	82067665	82611042	15464	543378	9	BPI	6
19	593976	598875	33567972	34110430	4900	542459	9	BPI	1
14	1186003	1192487	96150704	96692885	6485	542182	9	BPI	1

17	1271129	1282928	42638893	43180676	11800	541784	9	BPI	10
1	74482	92606	2391565	2932545	18125	540981	9	BPI	46
10	3006577	3006797	104083929	104624816	221	540888	9	LI	6
1	1105516	1119514	46995295	47535869	13999	540575	9	BPI	23
2	1146197	1157356	44524850	45063278	11160	538429	9	BPI	16
4	1176941	1188605	35923276	36461250	11665	537975	9	LI	15
2	944962	959388	28851045	29387407	14427	536363	9	BPI	14
6	1285592	1291107	104440679	104976637	5516	535959	9	LI	1
14	1510838	1516734	122068992	122604434	5897	535443	9	BPI	1
9	1492886	1497174	120491331	121026709	4289	535379	9	SC	1
18	3026030	3041271	89910916	90446231	15242	535316	9	BPI	9
19	1686712	1703116	59442227	59977252	16405	535026	9	BPI	10
19	1690235	1700409	56687464	57221251	10175	533788	9	BPI	15
14	2030974	2038056	151137370	151671043	7083	533674	9	BPI	1
9	1586699	1595137	126784146	127317156	8439	533011	9	SC	1
2	1048583	1067606	35209353	35741641	19024	532289	9	BPI	10
1	1818838	1827868	140855695	141386896	9031	531202	9	BPI	1
6	799173	803692	60642307	61172723	4520	530417	9	LI	1
19	1474736	1495708	40458000	40988356	20973	530357	9	BPI	9
10	1686843	1703109	59447391	59977151	16267	529761	9	LI	10
18	1805338	1813603	140075072	140604420	8266	529349	9	BPI	1
1	2132064	2146887	63935609	64463763	14824	528155	9	BPI	8
10	1328266	1333443	108069798	108597629	5178	527832	9	LI	1
13	2218270	2230376	77968348	78495986	12107	527639	9	SC	3
2	2495340	2510272	82080721	82608086	14933	527366	9	BPI	6
2	729920	744078	31607550	32134475	14159	526926	9	BPI	11
6	289905	296406	12040007	12566102	6502	526096	9	LI	1
2	1442044	1457117	35497047	36022815	15074	525769	9	BPI	24
5	1071465	1080075	87725257	88250700	8611	525444	9	LI	1
9	99001	106972	4008054	4533389	7972	525336	9	SC	17
19	479970	486669	24553686	25078762	6700	525077	9	BPI	1
14	1360635	1367575	110813463	111338471	6941	525009	9	BPI	1
18	820718	824998	63143755	63668307	4281	524553	9	BPI	1
1	934885	940763	76074082	76598237	5879	524156	9	BPI	1
11	849531	854545	66855638	67379624	5015	523987	9	SC	1
1	1708939	1723767	57523787	58046804	14829	523018	9	BPI	15
18	1880708	1898031	50513129	51034821	17324	521693	9	BPI	5
14	1656026	1666006	46417213	46938783	9981	521571	9	BPI	8
17	2433426	2446663	79665237	80185653	13238	520417	9	BPI	6
9	1642287	1648947	130714117	131234507	6661	520391	9	SC	1

11	1125160	1132348	91536327	92055573	7189	519247	9	SC	1
6	533132	534337	17678396	18197524	1206	519129	9	LI	30
9	204226	212106	6893077	7410602	7881	517526	9	SC	1
19	162480	176449	10622576	11138736	13970	516161	9	BPI	11
1	2457307	2471822	71832726	72347774	14516	515049	9	BPI	9
3	1576456	1588998	40384031	40898601	12543	514571	9	BPI	5
13	66132	77568	2922841	3437273	11437	514433	9	SC	1
9	2491527	2508440	90296486	90808560	16914	512075	9	SC	2
16	1978777	1990088	71891248	72401856	11312	510609	9	BPI	11
13	1787289	1796646	139031768	139541844	9358	510077	9	SC	1
6	1252689	1258368	36205429	36715419	5680	509991	9	LI	28
10	309173	317401	13623401	14132696	8229	509296	9	LI	1
20	1375487	1382804	111951972	112461235	7318	509264	9	BPI	1
20	491799	497896	25449597	25958844	6098	509248	9	BPI	1
2	827202	842689	19856134	20365284	15488	509151	9	BPI	22
1	1734796	1748377	49497764	50006248	13582	508485	9	BPI	8
19	128366	137978	10265691	10774119	9613	508429	9	BPI	16
20	1808030	1816233	140244844	140752976	8204	508133	9	BPI	1
4	1067158	1080122	44135869	44643771	12965	507903	9	LI	2
12	1076617	1084527	88048136	88555392	7911	507257	9	SC	1
19	2391494	2400071	72188195	72694788	8578	506594	9	BPI	13
2	252578	267746	5960667	6467038	15169	506372	9	BPI	36
10	818585	822806	62900611	63406802	4222	506192	9	LI	1
2	2056396	2070870	61544164	62050018	14475	505855	9	BPI	8
19	503974	523039	16911150	17416309	19066	505160	9	BPI	17
1	2224775	2236227	63234278	63739134	11453	504857	9	BPI	9
2	2682747	2699454	89043190	89547236	16708	504047	9	BPI	6
14	426308	437449	23246213	23750023	11142	503811	9	BPI	23
6	1186837	1192770	96211619	96715312	5934	503694	9	LI	1
19	61750	68775	4631024	5134630	7026	503607	9	BPI	10
7	1254199	1260641	101658006	102160844	6443	502839	9	LI	1
14	2239052	2250316	80780532	81283311	11265	502780	9	BPI	2
17	2030974	2037612	151137370	151639288	6639	501919	9	BPI	1
13	1372782	1380123	111771524	112272665	7342	501142	9	SC	1
5	1323985	1329461	107676376	108176471	5477	500096	9	LI	1

Table S4.8. Summary metrics for windows overlapping the CNDP2 gene and DGE output.

scaffold	Window start	Window end	Fst BPI_SC	Fst BPI_LI	Fst SC_LI	dxy BPI_SC	dxy BPI_LI	dxy SC_LI	pi_BPI	pi_SC	pi_LI	Tajima's D BPI	Tajima's D SC	Tajima's D LI
6	23220001	23270000	0.2766	0.4619	0.1445	0.0119	0.0114	0.0105	0.0033	0.0107	0.005	-0.102308	0.454067	-0.528947
6	23230001	23280000	0.2775	0.4315	0.1453	0.0127	0.0123	0.0122	0.0034	0.0117	0.0065	-0.130701	0.439819	0.05491
6	23240001	23290000	0.2831	0.4271	0.1437	0.0117	0.0112	0.0116	0.0028	0.0109	0.0065	-0.054796	0.400517	0.228972
6	23250001	23300000	0.2991	0.4331	0.1407	0.0104	0.0102	0.0107	0.0022	0.0098	0.0064	-0.006311	0.447854	0.496787
6	23260001	23310000	0.3048	0.3931	0.1466	0.0092	0.0081	0.0096	0.0018	0.0086	0.0058	0.037284	0.524358	0.399288
6	23270001	23320000	0.3168	0.364	0.1616	0.0077	0.0057	0.0083	0.0012	0.0075	0.0045	0.139627	0.513771	0.529929
6	23280001	23330000	0.315	0.3643	0.1452	0.0054	0.0029	0.0055	0.0005	0.0057	0.0025	0.492193	0.454994	0.207027
6	23290001	23340000	0.2824	0.3056	0.1352	0.0045	0.0021	0.0045	0.0005	0.0051	0.0019	0.190844	0.414947	-0.079746

baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
1645.958	0.2649	0.1732	1.5296	0.1261	0.5405