

COMPARATIVE BIOINFORMATIC AND MOLECULAR EVOLUTIONARY ANALYSIS OF
CHORDATE GENES AND GENOMES

A Thesis
Submitted to
the Temple University Graduate Board

In Partial Fulfillment of
the Requirements for the
Degree
MASTER OF SCIENCE

by
David Northover
December 2020

Thesis Advisor:

Dr. David A Liberles, Department of Biology

ABSTRACT

As knowledge of evolutionary processes has expanded over the years, we have deepened our understanding about how they drive organismal, cellular, and molecular biology and the factors beyond natural selection that are involved. Nevertheless, selection maintains a role in fixing and maintaining successful adaptations to new niches, whether from environmental change or organismal migration. Adaptation should not be considered solely on the level of individual genes and point substitutions as selection occurs on multiple levels. Examination on these multiple levels can further aid in understanding the constraints on evolution and how organisms can attain a phenotype.

Here we present two packages of tools for the examination of selection on the levels of protein structure and genetic pathways as well as on the individual gene and sequence levels., followed by examples of potential applications. First, we present a package of Application Programming Interface (API) tools that simplifies use of The Adaptive Evolutionary Database. Second, we present a package of tools implemented in the Rust programming language for fast and reliable analysis of phylogenetic data.

Then we describe the phenotypic data and methodology for use of these tools to analyze evolution on multiple levels, where genomic data is available. A broad scale analysis of the protein structural properties of evolutionary genetic changes in proteins is developed and described. We also present an organization of phenotypic data for mammals in the arctic biome, an ancestral reconstruction of the evolution of the phenotypic traits under study, and demonstrate a methodology to apply the tool packages to this cohort when sufficient genomic data is available.

ACKNOWLEDGMENTS

Thanks to David Liberles, Russell Hermansen, Tristan Stark, other members of the Liberles Research Group, and Kevin Northover for their advice, suggestions, and aid.

TABLE OF CONTENTS

ABSTRACT.....	i
ACKNOWLEDGMENTS.....	ii
LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
1. INTRODUCTION.....	1
1.1. Discussion of Genetics	1
1.2. Discussion of Protein Structure	4
1.3. Discussion of Evolution	4
1.4. Convergent Evolution	7
1.5. Evolutionary Constraints.....	9
1.6. Evolutionary Determinism	10
1.7. Thesis roadmap.....	11
2. CONTEXT OF WORK.....	13
2.1. Existing Methods of Convergent Evolution Analysis.....	13
2.2. Environment-specific Convergent Evolution	14
2.3. Unique Features of this Approach	15
2.4. Reasoning for Embarking on this Analysis	16
3. TAED (THE ADAPTIVE EVOLUTIONARY DATABASE)	17
3.1. TAED Background.....	17
3.2. Extending Usefulness with an API.....	18
3.3. The TAED API.....	19
4. PHYLO, A RUST LIBRARY OF PHYLOGENETIC TOOLS	22
4.1. Needs for Rapid, Integrated Tooling.....	22
4.2. Features and advantages of Phylo.....	23
4.3. Use of Phylo in Addressing Broader Evolutionary Problems	24
4.4. Library Handling and Publishing	25
5. PROTEIN EVOLUTION.....	26
5.1. Protein Evolution	26
5.2. Methodology.....	26
5.3. Results.....	28
5.4. Distribution within Solvent Accessibility Categories.....	32
6. PHENOTYPIC DATA.....	35
6.1. How is a Phenotype Constructed?.....	35

6.2.	Compilation of Phenotypic Data	36
6.3.	Phylogenetic Display of Phenotypic Data	37
6.4.	Convergent Phenotype Presence and Frequency.....	39
7.	ANCESTRAL SOURCES OF PHENOTYPIC TRAITS	41
7.1.	Summary	41
7.2.	Concealing Coloration.....	42
7.3.	Seasonal Pelage.....	44
7.4.	Body-Mass Variation	46
7.5.	Hoarding Behavior	48
8.	NOVEL METHODOLOGY OF CONVERGENT EVOLUTION ANALYSIS.....	51
8.1.	Background	51
8.2.	Necessary Data Sources	51
8.3.	Pathway Analysis.....	52
8.4.	Selection Analysis.....	52
8.5.	Sequence Analysis.....	53
8.6.	Use in Examining Evolutionary Constraint.....	54
	BIBLIOGRAPHY/REFERENCES CITED.....	55
	APPENDIX 1.....	70
	Arctic Categorical Phenotypic Data	70
	APPENDIX 2.....	73
	Arctic Additional Phenotypic Data	73
	APPENDIX 3.....	77
	Phenotypic Data For Ancestral Reconstruction	77

LIST OF TABLES

Table 1 - Site Substitution Rate by Secondary Structure and Solvent Accessibility	28
Table 2 - Site Distribution in Buried Regions	32
Table 3 - Site Distribution in Exposed Regions.....	33
Table 4 - Convergence Summary	41
Table 5 - Types of Hoarding Behavior	48

LIST OF FIGURES

Figure 1 - Paths through TAED API.....	20
Figure 2 - BLAST Call Sequence in Python.....	21
Figure 3 - Example of Using Tree Handling to Find Negatively Selected Lineages	23
Figure 4 - Retrieving Selection Data For a Record From Python.....	24
Figure 5 - Excerpt of Integration of Protein Structure and Phylogenetic Data for Analysis	25
Figure 6 - Phylogenetic Map of Convergent Phenotypes	38
Figure 7 - Ancestral Reconstruction of Concealing Coloration	43
Figure 8 - Seasonal Pelage Variation.....	45
Figure 9 - Ancestral Reconstruction of Seasonal Body Mass.....	47
Figure 10 - Pattern of Hoarding Behavior in Mammals	50

1. INTRODUCTION

1.1. Discussion of Genetics

Starting in the 19th Century scientists have been slowly unraveling the biological basis for the diversity of life, from the birth of Mendelian Genetics to the more complicated interactions understood to drive it today. Initial understanding of inherited traits considered them driven by a single gene, from two copies (alleles), one inherited from each of two parents in sexually reproducing species (1). Each gene was considered to act independently from all other genes, and in a simple dominance relationship where if the alleles differed one would control the expression of the trait and the other would have no role (1).

This view rapidly became outdated and we now understand that genetic interactions are much more complex. Alleles do not only interact in simple dominance, but in co-dominance (where both alleles contribute to the trait separately), incomplete dominance (where the trait is intermediate), or more complex dominance relationships where 1 dominant allele is not necessarily sufficient to express the trait correctly. Moreover, genetic interactions were identified as more complex than one gene- one trait. Pleiotropy, first defined in 1910, is any allele that affects multiple organismal traits. Epistasis, initially identified and defined as the masking of one gene by another (2) has been expanded and understood more broadly to explore any way the expression of the function of a gene is affected by other genes. Much of this is now understood at a biochemical level.

These changes, as well as the discovery of the structure of DNA (3) have led to the broader study of an organism's complete set of DNA in the field of genomics. DNA

sequencing has enabled this study by leading to the large-scale production of large datasets of sequencing data. This data both extensively covers certain key organisms to more readily identify the role of specific genes and sweeps across a large space of organisms to understand how genomes have changed over time.

Simply assembling genetic sequences is not sufficient to understand them, given the complexity of genetic interactions, however. Most of the DNA in an organism does not code for specific proteins, though the proportion of functional non-coding DNA is unclear. The ENCODE project, a systematic attempt to annotate the human genome, suggested that 80.4% of the genome as a whole is transcribed at some point and as such has a potential role in trait determination (4). However, in addition to the ENCODE's project vulnerability to false positives, the transcription of a section of DNA does not indicate that the transcribed section is functional and there are known pseudogenes that do not have any functional activity despite active transcription. Further, only 10% of the genome is evolutionary conserved (5), and it would be difficult for much of the further 70% to have extensive roles in conserved trait determination when they are not reliably conserved. Further, as most mutations are deleterious, an increase in functional areas of the genome creates an increase in the mutational load; that is, the amount of deleterious mutations that selection must be able to reliably eliminate in order for a population to survive (6). Such a high rate of functionality is also inconsistent with the known diversity of genomes across life. The C-value, a measure of genome size, varies dramatically across species and does not correlate with organismal complexity, as seen in the massive genomes of some species of plants (7).

While broader-scale genomics is one answer to the question of complex trait determination, it has still shown strong limitations in current approaches. Even with large

datasets, it can be difficult to have sufficient statistical power to be able to identify complex interactions. A common target for analysis is Single Nucleotide Polymorphisms (SNPs), alleles identified by the substitution of a single nucleotide in the gene. Genome-wide association studies (GWAS) take the approach of attempting to identify which SNPs are most associated with variation in a trait; however, they are often limited to known SNPs rather than arbitrary substitutions or full genetic sequences, and do not consider more complex interactions. It is also not always clear whether the associations identified are of functional genes or just those linked to functional genes by linkage disequilibrium or other, non-functional relations. Such matters can be compensated for but some commonly used methods are insufficiently conservative (8). Whether due to this or other factors, they do not generally find SNPs with an individually significant role in driving a trait, with the explosion of GWAS studies in 2007 usually finding alleles with <1.5% contribution when considering disease-associated risk (9), a trend that has continued across studies and fields.

A study of the mechanisms of gene interaction can be helpful to start to track down what we are missing here. While Francis Crick identified the central dogma of molecular biology as “the detailed residue-by-residue transfer of sequential information ... cannot be transferred back from protein to either protein or nucleic acid” (10) proteins, alone and in tandem with each other, are selected upon by evolution and drive change in the genome. These interactions are complex and not fully understood, but a high-level starting point is the wide-spectrum interaction of proteins in biochemical pathways.

The Kyoto Encyclopedia of Genes and Genomes (11) has systematically characterized work linking together genomic and functional data in a way that allows genes to be placed within the biochemical pathways they act in. While many of these pathways

are used across almost all cells, this is not always the case and pathway activity can vary strongly depending on the role of the cell.

1.2. Discussion of Protein Structure

As proteins act in 3D physical space, at the individual level of a gene formed into a protein, it is the structure of the protein and its physical properties that are ultimately responsible for the driving of phenotype. Proteins are driven by multiple levels of structure: The sequence of amino acids (primary structure); distinct local structural patterns formed by subsets of the full sequence (secondary structure) (12); the overall 3D structure created by the polypeptide coded by a gene (tertiary structure); and in some cases the interactions of separate polypeptide chains with each other (quaternary structure).

These constructions accommodate the separate parts of the protein that act in and respond to evolutionary pressures separately. The portion of a protein most variable under selection are generally the binding sites, where interaction with other proteins and molecules occur. Much of the rest of the protein is dedicated to maintaining its structure, including both the hydrophobic regions interior to the protein and the hydrophilic regions facing the exterior. These face their own selective pressure, as the shape of the protein is still important to its functional role, even beyond matters such as maintaining stability; issues such as the orientation of key functional residues and the prevention of aggregations and non-specific binding sites are amongst additional issues posed by protein structure (13)

1.3. Discussion of Evolution

Much genetic change is driven by genetic drift as alleles and their associated traits in a population change according to sometimes random processes (14). This leads to the

neutral theory of molecular evolution (14), where most genetic change is neutral to the fitness of an organism.

The most common genetic change driven by natural selection is negative, purifying selection, wherein novel genetic changes are selected against and removed from the population over time. This is maintained even in areas under frequent, novel selective pressure such as the immune system (15). This maintenance is vastly more frequent than other methods of selection as genetic mutation is usually deleterious to an organism if not neutral (16) and so organisms must regularly purge these negative changes to maintain fitness. Negative selection is also the primary mechanism in stabilizing selection, where the maintenance goes beyond clearly deleterious alleles into cases where a stable phenotype is maintained in order to remain competitive in a specific niche (17). Novel mutations can still be selected for in stabilizing selection, however, where they improve the ability to maintain the given phenotype.

However, stabilizing selection does not allow for adaptation to new niches. Constructive neutral evolution has been hypothesized to be able to drive complex traits that on their own contribute to organismal fitness (18); this would suggest that selection acts at least in part as a stochastic filter of generated traits (19). However, on its own this does not drive the mutations underlying the traits to fixation. That requires either directional or disruptive selection, driven by positive selection of new alleles from the less frequent beneficial genetic mutations. In directional selection, organisms are driven towards an extreme to adopt a new phenotype. This is essential in the adaptation to a new niche and may be the primary cause of phenotypic diversification (20). In disruptive (diversifying) selection, on the other hand, organisms are driven to not one, but multiple new phenotypes

that strongly contrast with one another and the original phenotype that they are diverging from (21). As natural selection cannot select on variation it does not have, “The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time” (22). The polymorphism resulting from diversifying selection provides an increase in genetic variation in fitness and allows for adoption of multiple more optimal phenotypes, either to slice the existing organism’s niche into multiple, smaller niches, or to allow for subsets of an organism’s population to react differently to changes in external environment.

All these selective processes can be complicated by linkage disequilibrium (LD), where loci and the traits they determine are associated at a different rate than what would be expected by random chance. Background selection and selective sweeps of a deleterious or beneficial allele cause genetic draft, a change in frequency of alleles associated via linkage disequilibrium, depending on the recombination frequency and range of fitness in a population (23). When strong directional selection occurs, this can strongly reduce localized genetic variation around the selective locus (24), possibly fixing deleterious mutations and throwing off stabilizing selection.

The measures of fitness driving natural selection are a result of the phenotype’s interaction with the natural environment (25) leading to the frequent independent adoption of the same feature both at the broader level of homoplasy and the more narrow molecular level (26). When the organisms are unrelated and evolve independently, this can be subdivided into parallel and convergent evolution. Parallel evolution occurs when organisms sharing one phenotype move to another shared phenotype; under natural selection this occurs twice as often versus neutrality, even down to the level of identical

mutations (27). Convergent evolution occurs when organisms with distinct phenotypes converge on a new shared phenotype and will be discussed in more detail.

1.4. Convergent Evolution

As multiple species can share overlapping niches, certain adaptations represent fitness peaks in the evolutionary landscape and are adopted by multiple organisms (28). At one extreme, this can occur extremely rapidly from distinct mutations (29). As such, while all selection requires mutational opportunity, convergent evolution is not substantially constrained by mutational opportunity, particularly as it can occur at distinct levels of the organism, such as the molecular or the morphological. This potential variation in genetic structure for the same trait can inform us about how traits are generated. Convergent evolution, alongside parallel evolution and reversion to ancestral traits, is one cause of homoplasy (30), the independent gain or loss of a certain trait in unrelated lineages. Multiple methods for detecting convergent evolution use homoplasy at the genetic level as part of their analysis (31). Homoplasy can be calculated based on extant phylogenetic data. The Consistency Index (32) measures the minimum amount of homoplasy implied by a tree. More recently, the Homoplasy Excess Ratio (HER) (33) measures the observed excess homoplasy (HE) vs the maximum possible excess homoplasy (MHE), using the formula

$$HER = 1.0 - \frac{HE}{MHE}$$

Recent research indicates convergent evolution is underdetected (34). Current methods are primarily based on similar sequence changes across matching genes, which can produce misleading results as these methods do not consider convergence derived from distinct mutations, or that selected upon at a different level of organization. Even with limitations, genome wide analysis has suggested that convergent evolution is significantly more pervasive than previously calculated (34). New approaches focused on identifying the roots of known morphological convergent evolution may suggest new paths that can correct this error elsewhere.

This does require known morphological similarities that can be associated with convergent evolution absent specific genetic analysis. Natural Selection fundamentally depends on the notion that the individuals better adapted to their environment survive and thrive (35). To identify convergently evolved phenotypes at the morphological level in a way that allows for examining their causes with minimal noise, the most straightforward approach is to focus on a specific biome, which provides many species with a shared environment. These species will have distinct genetic backgrounds and genetic ancestors that do not live in the biome. As such, phenotypic adaptations that provide high fitness in the shared biome but lower fitness outside of it can be inferred as convergent evolution. Here we are looking at the arctic biome, as food scarcity issues, temperatures, and varied snow cover have driven convergent adaptations that can be identified in the phenotypic data.

The evolution may not be clearly detectable at the gene level alone. Traits can arise by changes in interactions between genes that alone cannot be distinguished from the background of random changes. When genes that participate in a shared molecular

pathway evolve, the pathway can show highly significant indications of positive selection despite a lack of detectable enrichment at the genetic level (36). As clearly observable phenotypes are usually the result of a complex mix of molecular interactions rather than individual genetic changes (37), accurate accounting for convergent evolution must consider action at the higher, pathway level.

Pathways are not the only level above individual genes that can be examined, however. Genes act via the proteins that they transcribe, and the properties of the resulting structure drive the resulting phenotype. First, a protein must be stable, and the structural properties of a 3 dimensional protein are complex; many factors affect protein stability across multiple avenues of selection (38). Second, most protein interactions with other molecules occur in the narrowly defined area of the binding site, whose sequences are not optimized for stability but are predictable by focusing on ligand binding efficiency (39) even when it can reduce protein stability (40). This surface optimization also manifests in suppression of alternate active sites (13). Third, proteins can contain internal subunits called domains that can have independently identifiable structure, function and evolution (41). These domains face their own, distinct selective pressure and evolve at different rates (42). Further, this combination of factors leads to different levels of selection on other levels of protein structure, such as secondary structure (42).

1.5. Evolutionary Constraints

These forces do not act with complete freedom. Evolution occurs under multiple constraints. Mutational pressure is the most direct, as evolution cannot select on variation that does not occur. It is not the most parsimonious explanation in all situations for the lack of organismal adaptability to nearby, unclaimed niches or environmental changes that

similarly situated organisms adapt to (43). In the former, sufficient time has occurred that strict mutational pressure is no longer a factor, yet populations do not expand their range fully; possible reasons include range fragmentation and migration from central populations (44). In the latter, some species may adapt to an environmental change while others don't, despite identical exposure, time, and room (43).

Further, despite the capability of rapid evolvability, organisms possess remarkable morphological stability. This is particularly clear in the fossil record (45), providing clear examples of “evolutionary stasis”, minimal morphological change along a lineage for up to millions of years (45). While stabilizing selection undoubtedly plays a role in this, geographical constraints, phenotypic constraints, and epistatic constraints based on genetic interactions have also been proposed (46).

Such matters have spawned a wide study of the constraints of evolution as a modification to neo-Darwinian synthesis. The consideration of further factors as evolutionary constraints has included many other matters, from the ecological to the mechanical, resulting in a profusion of terminology (47).

1.6. Evolutionary Determinism

These matters lay bare that we do not currently understand how things evolve, and to what extent the genotype and phenotype of modern organism are dictated by deterministic versus random forces. This elaborates on the classic question of what would occur if we were able to “replay the tape” of life (48) alongside the questions of historical contingency; to seek answers to the question of how deterministic evolution is, even when the environment outside the organism is kept the same. Can certain phenotypes be obtained only in specific fashion from matching genotypes? How free is genetic change to

alter the proteins and other building blocks of an organism? How do interactions between proteins in genetic pathways mediate this? Answering these questions will improve understanding of how evolution occurs and how constraints manifest themselves in the underlying genetic code.

1.7. Thesis roadmap

The current state of convergent evolution analysis and previous results are described in Chapter 2 to set the context for the approach of this work. The value of the Arctic biome as a laboratory for this type of investigation and the new features of our approach are also identified in this chapter. Further, we identify the use of convergent evolution as a testbed for examining the path and constraints on evolutionary development.

In Chapter 3 we present data relating to The Adaptive Evolutionary Database (49) and the extension of its usability by the creation and continued development of an API, for the ability to integrate multiple different sources of genomic and biological information alongside the genetic sequences themselves.

Chapter 4 describes Phylo, a library of high-performance tools for Phylogenetic analysis implemented in the Rust programming language, to handle the large and integrated datasets used in the analysis.

Chapter 5 provides an overview of the approach used to combine the protein structural data and gene family databases and model protein structural evolution, as described in (42).

In Chapter 6 we provide a short summary of the data sources for phenotypes and gene families used in this investigation and present an analysis of phenotypic data for Arctic

and Antarctic mammals to display the frequency and type of convergent evolution displayed.

Chapter 7 presents an ancestral reconstruction of the phenotypic traits under study to understand the role of convergent evolution and alternate explanations (such as gain and loss) for the emergence of the traits.

A methodology to carry out analyses of pathway, gene and protein region enrichment, as well as sequence-level changes in convergently evolving lineages and the interrelationship between changes at each level is reported in Chapter 8.

2. CONTEXT OF WORK

2.1. Existing Methods of Convergent Evolution Analysis

Multiple methods have been used to identify convergent evolution in genetic data. Phenotypic similarities cause deformations in phylogenetic trees due to the increased similarities of the underlying genetics (31). One approach to identify convergent evolution has been to locate the inconsistencies created here, with varying methods to address either continuous or categorical data. These methods are limited, however; they only capture instances where the convergence is dependent on the same underlying genetic changes. This limits such methods to studying at the gene level, rather than the regulatory, protein or pathway level. The method is also of low power; the convergent signal must be strong enough to shift the tree topology despite the more numerous genetic changes indicating the ancestral signal. Further, they are naïve towards known phenotypic data. Another approach is to contrast the phylogenetic and phenetic tree. This has the advantage of not being naïve to existing knowledge of phenomic data but is restricted in the kinds of phenomic data that can be used due to needing some level of continuous data (31). It retains the other issues associated with tree inconsistency analyses.

More recent work has introduced newer methods. Functional Enrichment Analysis can be used with genetic sequences and molecular-level phenotypic data to identify groups of genes affected by substitutions in phenotypically convergent lineages (50). However, this requires a certain fine-grained level of phenotypic data that is currently only present for model organisms, and the phenotypes used will not be a clean match to a convergent phenotype in a different species. Marcovitz uses the Mammalian Phenotype Ontology, which is currently largely restricted to mice.

These methods are particularly limited in that they identify only sequence-level changes in specific genes and so do not capture the full breadth of the areas of the genome that can be selected upon. Regulatory sequences play their own role in evolution, and as gene-level analyses do not sequence or compare them, nor run a comparative transcriptomic analysis of the expressed RNA, study of this role is entirely absent from these analyses. Protein structure and pathways are entirely absent from these analyses, despite having roles in evolutionary change that can be identified when gene and sequence level changes cannot be distinguished from the genetic background.

2.2. Environment-specific Convergent Evolution

Fitness of an organism's phenotype cannot be separated from the organism's environment, as it is the interaction between an organism with that phenotype and the environment that determines an individual's fitness (25). When the selective pressure is strong enough, even completely isolated organisms may adopt the same mutation (51). The Arctic and Antarctic are two similar shared environments that present similar environmental challenges across a wide range of species in large areas, particularly regarding temperature regulation, seasonal variation, and the demands of cryptic coloration. Moreover, organisms in the arctic niche have close genetic relationships to non-arctic species from which they differentiated by migration and isolation. This combination of distantly related species with a shared environment, all of which are more closely related to species that do not share the same environment, is an excellent situation to examine the nature of convergent evolution.

The situation is not perfect, however. Gene flow, the transfer of genetic information between populations, remains a factor as much of the arctic is contiguous with non-arctic areas. Gene flow constrains speciation when it does not lead to the founding of

new populations (52), though it is not necessarily entirely absent during speciation (53). The level of gene flow in arctic species varies based on both historical and current population dispersal and behavior, such as long distance migration (54). Even within a species geographically isolated populations can genetically diverge from related, non-isolated populations despite similar environment and life history (55). The net effect of these genetic factors is to complicate the consideration of interspecies vs intraspecies genetic difference for widely distributed arctic species versus their closely related non-arctic cousins. These complicating factors are increasing in importance with the effects of climate change as current arctic habitats are changing rapidly. Endemic arctic species have faced habitat loss and some non-arctic species have widened their distribution (56). Examination of ancestral DNA during previous climate change found evidence of gene flow in populations where it does not currently occur (57). This means that closely related species to endemic arctic species are often found in the modern arctic as well.

2.3. Unique Features of this Approach

Rather than running an analysis naïve to current knowledge of phenotypic data, we have chosen the approach of identifying existing convergent phenotypes and doing a thorough analysis of the underlying genetic data at more levels than the genetic tree. Enrichment analyses have been used at the gene level as part of methods before (50) but less so with higher level, known convergent phenotypes. Use of selection data allows for less dependence on identical changes to identify convergence events. Pathway analyses is a novel addition to convergent evolution analysis; as genetic changes can be reflected in pathway shifts this provides an opportunity to identify them where they had not previously been located. Protein structure similarly provides insight into a previously understudied path to analyzing directional selection resulting in convergent evolution. The increased

robustness multiple analytical approaches provide aids in accounting for the ongoing environmental shifts mentioned above, though no approach is immune.

Looking at individual genetic sequences provides a different angle, as it tracks the consistency of genetic changes driving convergence across separate organisms, and how frequently these genetic changes are shared. As multiple existing convergent evolution analysis depend on genetic sequence similarity (31) this can provide insight into the reliability of these analysis as well as providing novel data about the mechanism of convergent evolution at the gene level.

As described further below, the phenotypic data was collected from extant references and categorized for analysis.

2.4. Reasoning for Embarking on this Analysis

Convergent evolution is an excellent testbed for understanding the scope of and limitations of evolutionary change. Understanding which organisms undergo convergent evolution to a trait and which do not can identify situations where mechanical, geographic, and phenotypic constraints prevent adaptation. Other constraints can be examined on the genetic level by developing the understanding of the extent to which convergent evolutionary changes towards the same phenotype share a similar path along the genetic lineage of an organism, and the distinctness of the final resulting genotypes. Examining convergent phenotypes from multiple perspectives will shed light on the genomic constraints on the development of phenotypes at the organismal level. Such insights can be extended to examine the path of evolution beyond convergent evolution.

3. TAED (THE ADAPTIVE EVOLUTIONARY DATABASE)

3.1. TAED Background

To analyze evolutionary information on multiple levels we must first collect the necessary data into a coherent framework. For examining integrated phylogenetic, pathway, and protein data we turn to The Adaptive Evolutionary Database (TAED). TAED was constructed by Russell Hermansen following the pipeline outlined in (49). The pipeline includes generation of gene families from single-linkage clustering of BLAST results from chordate genes found in GenBank. A point accepted mutation (PAM) distance threshold of 120 was used for gene family construction. Gene families were refined for quality using an iterative method controlling for pairwise percent identity and the fraction of pairwise aligned gaps. Gene families were then aligned using MAFFT (58) and phylogenetic trees were constructed using PhyML (59). Gene tree – species tree reconciliation against the NCBI chordate taxonomy was implemented to determine putative duplication events and gene tree roots using SoftParsMap. Putative rates of evolution were then calculated using the branches model from PAML and dN/dS rates were computed. BLAST was then performed on TAED gene families against the KEGG database (11) to determine KEGG pathway relatedness and against PDB (60) to determine protein structure for each gene in TAED. All branches, including specifically those found to have a dN/dS > 1 (putatively evolving under positive selection) were mapped to the corresponding chordate species tree to determine along what lineage the elevated rates of evolution occurred and which proteins evolved rapidly on the same species tree lineage. Roots of all gene families were additionally mapped to the chordate species tree. To determine the approximate family root age for each gene family, information from TimeTree (61) was collected and root ages determined

in MYA (millions of years ago). Domain classification information was gathered from the CATH database (62). Putative functional annotations were assigned to each gene family based on NCBI nomenclature and KEGG pathway annotations when available.

3.2. Extending Usefulness with an API

As initially built, The Adaptive Evolutionary Database required access either through website usage or via downloading the entire database and flat files, then querying the database directly for information. This required a heavy local footprint for efficient programmable access, or for access to large portions of the dataset, as downloading large numbers of files from the existing website is cumbersome, and the full dataset takes up a large amount of space. The buildout of a new Application Programming Interface (API) allows us to address this issue, by providing access to the TAED information in a way more convenient to users who need to access to the larger scale datasets necessary to work extensively with the database. It offers this both via a generic API callable by multiple methods, and small group of python tools allowing users to construct queries without knowledge of either SQL or calling the website via POST or JSON, assembling the calls itself.

The presence of an API allows for more effective use of the TAED application and favors analyses in several ways. First, building analytical methods on top of the API rather than direct access to the files allows distribution of the methods without distribution of the database and/or flat files depending on what is utilized. Second, it allows for updates of the database over time to be used without redownloading. Third, it allows simpler and more coherent access from programming languages, as direct API calls are clearer to end users and do not require knowledge of the internal structure of either the database or the flat file

compilations. These matters are necessary for robustness of the TAED database as an evolutionary resource.

3.3. The TAED API

Portions of the TAED API were originally published in (49).

First, the API provides an easy way to access to the alignment and phylogenetic tree data files that make up the genomic records in the site. They can be delivered either in a standard flat file format or parsed and sent in the widely used JSON file format. Further, the links to these files can be fetched via multiple possible queries, allowing the systematic fetching of genomic data without downloading extraneous datasets.

For example, to fetch all alignment and phylogenetic data associated with agouti on genes on lineages with at least 5 total taxa, you can call the following url:

https://liberles.cst.temple.edu/TAED/json/search?min_taxa=5&gene=agouti

Such as with:

```
wget
https://liberles.cst.temple.edu/TAED/json/search?min_taxa=5&gene=agouti
> agouti_data.json
```

to retrieve a JSON object usable by multiple programming languages.

Alternately, the API can provide the information residing in the database in formats more immediately accessible. This is particularly true for certain data (such dNdS and KEGG

data) that need additional parsing when directly accessed from the database. API documentation is ongoing at (63).

The API also provides a method to perform a BLAST search against the records extant in the database. BLAST (64) is a series of tools used to identify sequences in a database or file similar to a provided sequence. The TAED API takes the website's existing ability to run a BLAST search against the TAED database (and thus find related sequences with associated phylogenetic information) and provides it in a simpler and programmatically accessible fashion, necessary when significant numbers of sequences are being checked.

Figure 1 shows an overview of paths through the API.

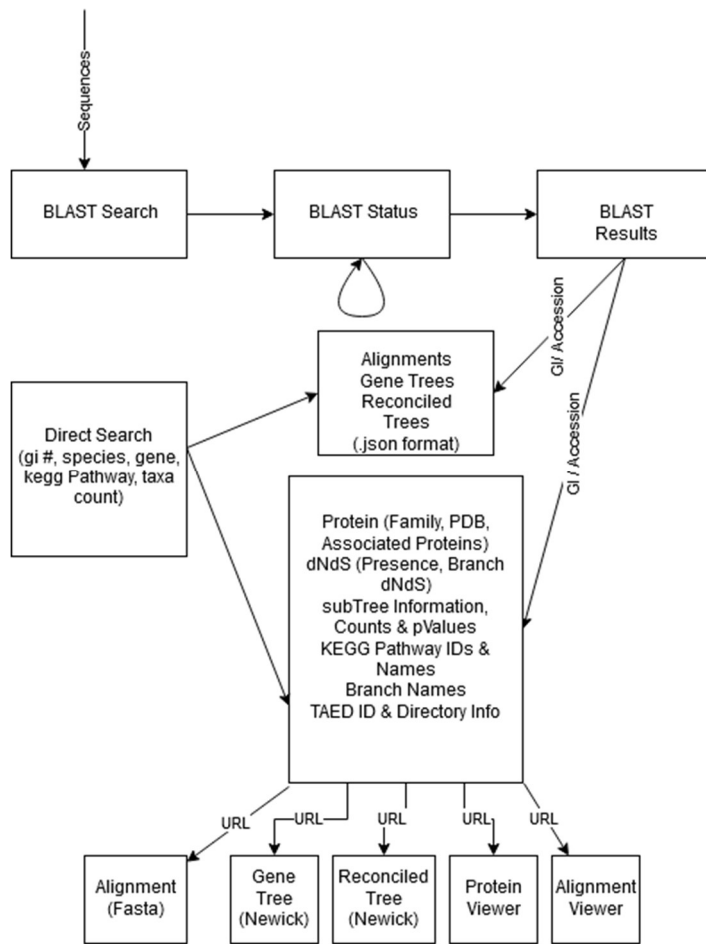


Figure 1 - Paths through TAED API

All of this is wrapped in python integration, allowing cleaner scripting than bash /
wget calls, as well as integration with BioPython (65) and other tools.

```
# Getting blast data for a sequence.
b_search = BLASTSearch(search={"job_name": "example", "e_value": 10, "max_hits": 50,
                              "sequence":
"MRPGIDSTDNAGRKGAAINANEAMLTAALLSCALLLALPATQGAQMGLAP"})
print(b_search.__dict__)
if not b_search.status["run_status"] == BLASTStatus.ERROR:
    result = b_search.run_web_query(in_use_url)    # Start the BLAST Run.

    if result.status["run_status"] == BLASTStatus.ERROR: # Early error.
        print("\nSomething went wrong in the initial query.")
        print(result.status["error_message"])

    while result.get_remote_status(in_use_url + "Status") == BLASTStatus.IN_PROGRESS:
        sleep(5) # Wait to finish

    # If we're not complete, there's an issue somewhere.
    if result.get_remote_status(in_use_url + "Status") != BLASTStatus.COMPLETE:
        # Error Handling
    else:
        print("\nWe're complete and can fetch the BLAST result.")
        final_data = result.get_remote_data(in_use_url + "Result")
```

Figure 2 - BLAST Call Sequence in Python

4. PHYLO, A RUST LIBRARY OF PHYLOGENETIC TOOLS

4.1. Needs for Rapid, Integrated Tooling

Bioinformatics is heavily dependent on the tools used to read, manage and manipulate the data. Most analysis pipelines are a mix of batch scripts, scripting languages, and command line tools, with some integration with faster, compiled tooling. Input and output generally consist of human-readable but verbose output formats, varying from the simple and easy to parse to the extremely complex and not designed for machine readability.

As the number of bioinformatics datasets grows, ever larger amounts of data must be parsed and handled. This is particularly true when multiple lines of genomic data are combined, as protein, phylogenetic, and alignment data all have different requirements for their storage and recovery. Efficient and reliable handling of this data is necessary in large-scale genomic analysis, particularly given the need for efficiency in both memory and speed for navigating datasets that are often far larger than can be stored in memory at a single time.

There are multiple extant data formats in use for handling trees for phylogenetic analysis alone, including Nexus (66), Newick (67), and NHX (68) amongst others. Complex protein data is usually distributed in the common Protein Data Bank Format (69), while the fasta/fastp formats (70) are the most common for simple sequences of residues or nucleotides, although others such as PHYLIP (71) are used in conjunction with phylogenetic analysis. Existing tools for analysis also create their own formats that must be handled separately, including the Phylogenetic Analysis by Maximum Likelihood (PAML) toolset (72),

as well as tools related to Protein Secondary Structure, using the Dictionary of Secondary Structure of Proteins (DSSP) format (73).

4.2. Features and advantages of Phylo

To address these issues, a unified tooling in a rapid, stable language will help enable analyses that are difficult or impractical using long toolchains and unclear data formats. The Phylo library is our approach. Phylo library parsing and navigation of these sizable datasets provides high performance and safe access to data stored in these formats, including handling malformed input files or those that do not strictly follow all standards, while outputting alongside the strict format definitions.

```
/// Finds negative Selection across a lineage in a tree.
fn find_neg_selection<'a>(node: &'a Clade, parents: &'a mut Vec<&'a
Clade>) -> Vec<(String, String)> {
    let mut name_pairs = Vec::new();
    if is_not_bootstrap_node(&node.name) && node.is_neg_select().unwrap() {
        for x in (0..(parents.len() - 1)).rev() {
            if is_not_bootstrap_node(&parents[x].name) {
                name_pairs.push((&node.name.clone(), &parents[x].name.clone()));
                break;
            }
        }
    }

    parents.push(node);
    for child in &node.children {
        name_pairs.append(&mut find_neg_selection(&child, &mut parents.clone()));
    }
    parents.pop();
    name_pairs
}
```

Figure 3 - Example of Using Tree Handling to Find Negatively Selected Lineages

Additional support was also built to support analysis of this data. The computationally heavy portions of the following methods of analysis (Protein Evolution and Convergent Evolution analyses) was implemented as part of the Phylo library and is to be

released as reusable portions of code easily portable to allow application to different sets of lineages than they have been initially applied to.

Further, foreign function interfaces allow the use of the library from other languages while maintaining the advantages of the library.

```
def get_PAML_Selection(paml):  
    print(f"Getting PAML Data for {paml}")  
    paml_dimensions = phylo_lib.array_meta()  
    paml_data = phylo_lib.get_paml_selection(paml, paml_dimensions)  
    return paml_data, paml_dimensions
```

Figure 4 - Retrieving Selection Data For a Record From Python

4.3. Use of Phylo in Addressing Broader Evolutionary Problems

Support for a wide variety of data handling sources allows phylo to integrate analysis across multiple types of data more easily than alternate methods. Integration of multiple datasets is important for modern genetic analyses, whether examining evolution as in this case, or for extending other types of analyses such as GWAS studies to integrate pathway (74,75), phylogenetic (76) or other methods of data. Being able to handle these sets of data with the same toolset simplifies the necessary interactions. While other libraries such as BioPython (65,77) or Bio++ (78) exist and handle significant amounts of data, each comes with tradeoffs in speed or reliability, and they are not necessarily built to handle all the data at hand in the analysis for this paper.

```

let buf_reader = BufReader::new(file);
prot = match serde_json::from_reader(buf_reader) {
    Ok(protein) => ExtendedProteinStructure::extend(protein,
pdb_number),
.....

let mut paml_file = match
PAML::read(&format!("{}/{}/{}.subTree_tree_1.paml_rooted",config.data["pam
l"],directory,taedfilenumber)) {
    Ok(res) => (res),
....

match
paml_file.append(&format!("{}/{}/{}_1.RST",config.data["paml"],directory,t
aedfilenumber)) {
    Some(_) => (),
....

paml_file.regenerate_records();
paml_file.append_prot_structure(&prot);

match paml_file.get_ancestral_alignment() {
    Some(extended_alignment) => prot.add_family(paml_file.records,
extended_alignment, taedfilenumber),
    None => continue
};

```

Figure 5 - Excerpt of Integration of Protein Structure and Phylogenetic Data for Analysis

4.4. Library Handling and Publishing

Rust packages are publicly available in the crates.io repository, which allows easy package management within the rust ecosystem. Phylo will be publicly published once the final versions of its external-facing API are established and tested, with the source code available separately on github.

5. PROTEIN EVOLUTION

Sections 5.2 and 5.3 are from “Characterizing Lineage-Specific Evolution and the Processes Driving Genomic Diversification in Chordates” (79) which has been published in BMC Evolutionary Biology.

5.1. Protein Evolution

When looking at evolution in coding sequences, the primary consideration is the amino acid sequence, dependent on the subset of non-synonymous nucleotide changes that alter the residue at a specific position. A variety of models (80,81) currently exist to analyze evolution based on substitutions along these lines, each attempting to strike a balance between accuracy and simplicity. This evolution has further occurred under constraints induced by protein structure; this has been examined in the context of protein kinetics and structurally disordered proteins (82) but not at the level of secondary structure or solvent accessibility. The latter issues have their own influence on protein stability (83) and have internal stability based on their particular physical properties (84). While this stability has complicating factors such as peptide length (85) as the stability is a result of the physics acting on the residues in the structure, examining the evolution of proteins at the secondary structure and solvent accessibility level will help understand the constraints that evolution operates on, and from there the potential ways phenotype can be generated by genotype.

5.2. Methodology

Protein information was determined from stored PDB information associated with each gene family. To show that sites in different locations and belonging to different structures evolve at different rates, DSSP (73) values were used to ascertain the relative

solvent accessibility (RSA) and secondary structure of individual sites within the protein was obtained. Membrane proteins and multimers were removed from the dataset based on identifying information in the PDB data. Sites were binned based on RSA using maximum surface areas from Tien et. al (86); sites with a ratio greater than 0.20 were marked as exposed and buried otherwise, and then further categorized according to secondary structure. PAML analysis was used to determine the maximum likelihood ancestral sequence for each gene associated with a protein and the results controlled for lineages with $dN/dS > 1$ and lineages with a $dN/dS < 0.5$. dN/dS values of 0 or between 0.5 and 1 were ignored, as were any sites that did not align with the PDB sequence or were not one of the most common 20 amino acids. To determine the significance of the calculated values, two-tailed non-parametric bootstrapping was performed. For each lineage, simulated datasets of size matching the total substituted residue count were generated, using the distribution of all sites on the respective lineages as a baseline.

5.3. Results

Table 1 - Site Substitution Rate by Secondary Structure and Solvent Accessibility

	Positively Selected Lineages (dN/dS > 1)			Negatively Selected Lineages (dN/dS < 0.5)		
	Substituted Sites	P	All Sites	Substituted Sites	p	All Sites
<i>Helix</i>	30.2826%	<0.0001	34.0597%	35.2377%	<0.0001	36.6327%
<i>Exposed</i>	17.1142%	0.0002	16.4580%	20.2511%	<0.0001	17.3397%
<i>Buried</i>	13.1684%	<0.0001	17.6017%	14.9866%	<0.0001	19.2929%
<i>α-Helix</i>	26.1346%	<0.0001	30.1108%	31.1229%	<0.0001	32.7617%
<i>Exposed</i>	14.3764%	0.1659	14.1740%	17.5250%	<0.0001	15.1730%
<i>Buried</i>	11.7582%	<0.0001	15.9368%	13.5979%	<0.0001	17.5887%
<i>3₁₀ Helix</i>	3.4956%	0.0098	3.3039%	3.5115%	<0.0001	3.2213%
<i>Exposed</i>	2.3597%	<0.0001	2.0045%	2.4163%	<0.0001	1.8907%
<i>Buried</i>	1.1359%	0.0004	1.2994%	1.0952%	<0.0001	1.3306%
<i>π-Helix</i>	0.6524%	0.8047	0.6449%	0.6033%	<0.0001	0.6497%
<i>Exposed</i>	0.3780%	<0.0001	0.2794%	0.3098%	<0.0001	0.2761%
<i>Buried</i>	0.2743%	0.0005	0.3655%	0.2935%	<0.0001	0.3736%
<i>β-Sheet</i>	23.2104%	<0.0001	21.7820%	18.2981%	<0.0001	19.8385%
<i>Exposed</i>	8.9360%	<0.0001	7.1998%	7.3255%	<0.0001	6.2661%
<i>Buried</i>	14.2744%	0.0361	14.5822%	10.9726%	<0.0001	13.5724%
<i>β-Bridge</i>	1.1095%	0.7913	1.0984%	0.9888%	<0.0001	1.0382%
<i>Exposed</i>	0.5644%	0.0004	0.4641%	0.4876%	<0.0001	0.4194%
<i>Buried</i>	0.5451%	0.0081	0.6343%	0.5012%	<0.0001	0.6188%

Table 1, Continued - Site Substitution Rate by Secondary Structure and Solvent Accessibility

	Positively Selected Lineages (dN/dS > 1)			Negatively Selected Lineages (dN/dS < 0.5)		
	Substituted Sites	P	All Sites	Substituted Sites	p	All Sites
Turn	12.0729%	<0.0001	11.0540%	12.3561%	<0.0001	11.0859%
Exposed	9.7554%	<0.0001	8.3283%	9.9588%	<0.0001	8.1517%
Buried	2.3175%	<0.0001	2.7257%	2.3973%	<0.0001	2.9342%
Bend	10.4763%	<0.0001	9.7416%	10.2628%	<0.0001	9.6989%
Exposed	7.9179%	<0.0001	6.8552%	7.8004%	<0.0001	6.6547%
Buried	2.5584%	<0.0001	2.8864%	2.4624%	<0.0001	3.0443%
Coil	22.8482%	0.0006	22.2643%	22.8565%	<0.0001	21.7058%
Exposed	15.2151%	<0.0001	13.2976%	15.3522%	<0.0001	12.7858%
Buried	7.6331%	<0.0001	8.9667%	7.5044%	<0.0001	8.9201%
Buried (All Sites)	40.4969%	<0.0001	47.3970%	38.8245%	<0.0001	48.3826%
Exposed (All Sites)	59.5031%	<0.0001	52.6030%	61.1755%	<0.0001	51.6174%

P-Value based on parametric bootstrapping, n=20000, and adjusted for multiple comparisons

The combination of gene families and information from the Protein Databank allows examination of how selection acts on a protein structural level. Gene families with associated protein structures were collated and aligned to the PDB alongside maximum likelihood ancestral sequences calculated by PAML.

The resulting profile is significantly different than the profile of non-substituted sites in

the background on those lineages. For both positively and negatively selected lineages, fewer substituted sites are buried relative to all sites on the protein; this is true both looking at all sites, and sites of any specific secondary structure, except for β -Sheet ($p = 0.0361$) and β -Bridge ($p = 0.0081$) sites on positively selected lineages, which was not significant after a multiple testing correction. The result in β -Bridge sites may simply be a matter of lower power due to the relatively small number of β -Bridge residues compared to most other secondary structures. β -Sheet sites are the most commonly substituted buried site on positive lineages (14.2744% vs 13.1684% for all helices), though α -Helix sites, as well as helices in general, are more common amongst all sites (15.9368% and 17.6017% vs 14.5822% for β -Sheet).

Negatively selected lineages consistently show an increase in the prevalence of exposed residues across all secondary structures, but this is not universal for positively selected lineages. α -Helix sites are the most frequent in the dataset and show no change in prevalence of exposed sites compared to non-substituted sites under positive selection. This suggests α -Helices have a significant selective constraint lacking in even other helix structures. 3_{10} Helix sites show an overall increase in substitution rates in negatively selected lineages, unlike other helices but consistent with bends, turns and coil sites. This is likely linked to their lower stability and higher proportion of exposed vs buried sites.

In terms of secondary structure when both exposed and buried regions are considered together, substitutions are more likely to occur across less structured regions (Turns, Bends, and Coil areas) that are more likely to be exposed than buried on both positively and negatively selected lineages, but also β -Sheet sites on positively selected lineages and 3_{10} Helix sites on negatively selected lineages. The changes in prevalence for each secondary structure is strongly related to the buried/exposed ratio of their own residues (particularly in negatively selected sites), so solvent exposure, while a significant factor, is not the only one. This corresponds with

observations seen in other studied (87).

The lack of significant change in *β-Sheet* buried sites on positively selected lineages suggests that positive selection is freer to act on it than comparable *α-Helix* sites, which have a considerable and significant drop in frequency amongst substituted (13.1684%) rather than all (17.6017%) sites. The *β-Sheet* site changes also point at differences between positive and negative selection. Unlike in positively selected lineages, in negatively selected lineages a smaller proportion of substituted sites are buried *β-Sheet* sites compared to all sites. This suggests the difference on positively selected lineages is not simply due to lower fragility in *β-Sheet* structure, but an active role for *β-Sheet* internal structure in driving evolution of new functionality. It should also be considered that, in general, positively selected lineages have fewer *α-Helix* (30.1108% vs 32.7617%) and more *β-Sheet* (21.7820% vs 19.8385%) sites compared to negatively selected lineages. Since, as discussed earlier, certain gene families and pathways are under more frequent positive selection than others, the lower selective constraint on *β-Sheet* sites has a long-term impact on protein structure.

β-Bridge sites did not show a reduction in prevalence for substitutions on positively selected lineages. As these sites are used to hydrogen bond, particularly between *β-sheets*, a likely explanation for these substitutions is to allow for protein restructuring. Compensatory driven changes are a less likely explanation, as negatively selected lineages where they are more likely than positively selected ones show a significant reduction in *β-Bridge* prevalence amongst substituted sites (0.9295% vs 0.9587%, $p < 0.0001$).

5.4. Distribution within Solvent Accessibility Categories

The data can also be considered within each solvent accessibility category, buried and exposed separately.

Table 2 - Site Distribution in Buried Regions

	Positively Selected Lineages (dN/dS > 1)			Negatively Selected Lineages (dN/dS < 0.5)		
	Substituted Sites	P	All Sites	Substituted Sites	p	All Sites
<i>Helix</i>	32.5170%	<0.0001	37.1367%	38.6009%	<0.0001	39.8757%
<i>α-Helix</i>	29.0348%	<0.0001	33.6241%	35.0239%	<0.0001	36.3534%
<i>3₁₀ Helix</i>	2.8049%	0.5194	2.7415%	2.8209%	0.0071	2.7501%
<i>π-Helix</i>	0.6773%	0.07955	0.7712%	0.7560%	0.2512	0.7723%
<i>β-Sheet</i>	35.2481%	<0.0001	30.7660%	28.2621%	0.00315	28.0522%
<i>β-Bridge</i>	1.3460%	0.9145	1.3383%	1.2909%	0.5158	1.2790%
<i>Turn</i>	5.7227%	0.8389	5.7508%	6.1748%	0.00445	6.0645%
<i>Bend</i>	6.3176%	0.1176	6.0898%	6.3424%	0.2025	6.2921%
<i>Coil</i>	18.8485%	0.7664	18.9183%	19.3290%	<0.0001	18.4365%

P-Value based on parametric bootstrapping, n=20000, and adjusted for multiple comparisons

Table 3 - Site Distribution in Exposed Regions

	Positively Selected Lineages (dN/dS < 0.5)		Negatively Selected Lineages (dN/dS < 0.5)			
	Substituted Sites	P	All Sites	Substituted Sites	p	All Sites
Helix	28.7618%	<0.0001	31.2871%	33.1033%	<0.0001	33.5928%
<i>α</i>-Helix	24.1608%	<0.0001	26.9453%	28.6471%	<0.0001	29.3951%
<i>3₁₀</i> Helix	3.9657%	0.1578	3.8107%	3.9498%	<0.0001	3.6629%
<i>π</i>-Helix	0.6353%	0.0138	0.5312%	0.5064%	0.0134	0.5348%
<i>β</i>-Sheet	15.0177%	<0.0001	13.6871%	11.9745%	0.0015	12.1395%
<i>β</i>-Bridge	0.9486%	0.222	0.8822%	0.7970%	0.2726	0.8125%
Turn	16.3948%	0.0086	15.8323%	16.2790%	<0.0001	15.7926%
Bend	13.3067%	0.1576	13.0320%	12.7509%	0.00715	12.8923%
Coil	25.5703%	0.2444	25.2792%	25.0953%	<0.0001	24.7703%

P-Value based on parametric bootstrapping, n=20000, and adjusted for multiple comparisons

When examined in this fashion, the data reinforces the change in Helix and β -Sheet helixes dependent on selection and solvent accessibility. Under positive selection, β -Sheet represent a considerably higher % of buried substituted sites compared to all buried sites (35.2481% v 30.7660%) while Helixes, driven by the change in α -Helixes (29.0348% v 33.6241%) represent fewer. This is also true for exposed regions, but less dramatically in both cases. Other structures do not significantly differ in the distribution between buried and exposed sites under positive selection, suggesting solvent accessibility is the most important factor in the change in frequency in these sites, but as some of these sites are rare some difference may be found in a

broader dataset.

Under negative selection, the matter is different. While α -Helixes do show a slight drop, it is smaller than in positive selection, but β -Sheet sites do not increase in either buried or exposed locations. Less ordered sites do increase in relative prevalence, however; Coil in both buried and exposed sites, and Turn in exposed sites. Less expected are 3_{10} Helixes, which show an increase in exposed site prevalence on negatively selected lineages. This may be related to their role in coordinating α -Helixes in some proteins (88), providing a role of compensatory mutations.

6. PHENOTYPIC DATA

6.1. How is a Phenotype Constructed?

In contrast the genotype, clearly definable from the sequence of nucleotides that defines the organism's genetic code, phenotype has a more complicated definition. An observable trait of an organism, phenotypes can include molecular phenotypes (89) at the protein/RNA level, cellular phenotypes (90), cognitive (89) phenotypes, behavioral (91) phenotypes, extended phenotypes incorporating an organisms impact on the environment (92) amongst others in addition to classical morphological phenotypes (93). The precise definition within phenotypes can also be unclear. For example, coat coloration in the wild can often be clearly categorized within or across organisms; e.g. fox coloration falling into clear categories that can be linked to MC1R and agouti activity (94). Nevertheless, while this categorization is a phenotype, coat coloration is not purely identical for individuals within each category, reflecting another, more precisely defined phenotype that has additional environmental and biological causes. Not all traits can be cleanly categorized however; others such as height require quantitative analysis or arbitrary categorization.

Such categorization and data collection raise challenges, as categories and data collection methods must be consistent across studies to be synchronized. Existing phenotypic data is not available in the organized fashion that genotypic data is in such areas as the NCBI library. There are several phenotypic databases, but published ontologies must be maintained and not all are; e.g. PhenomicDB (95) which is currently defunct. Those that are stable and comprehensive are often focused on specific model organisms, such as the Mammalian Phenotype Ontology (MP) (96), which focuses almost exclusive on mice despite

a theoretically broader classification system. It is also more narrowly focused on molecular and disease phenotypes. Attempts have been made to merge this data across organisms (97) but though useful such work is still limited to the species covered on the datasets, and works primarily on identifying orthologous genotypes and phenotypes that can be matched cleanly (98), currently useful primarily for disease phenotypes.

6.2. Compilation of Phenotypic Data

As such we are approaching the method based primarily categorical analysis while maintaining exposure of the underlying data, collected in a coherent fashion. This allows for consistent analysis methods among traits while allowing for re-use of the underlying data under different categorical requirements, such as when exposed to different data sets. The phenotypic data for this study is therefore drawn from multiple published sources. References on Arctic (99) and Antarctic (100) wildlife, academic websites (101,102) and mammal hoarding behavior (103) have been consulted, with some reference paper supplementation(104,105).

Pelt coloration was separated by winter and summer coloration, with detailed coloration data recorded, then further organized by creating separate categorization for those displaying strong seasonal coloration or white cryptic coloration. Body mass was taken from the middle of specified ranges, averaged between male and female equally, then categorized into tiers, as well as consideration of the seasonality. Hoarding behavior was distributed among scatter hoarding, larder hoarding, limited caching of food, and a combination of the above. Fuller descriptions of these can be seen in table 5. Body fat % was not consistently covered in the literature (106–109) and was dropped from the analysis due to incomplete coverage and inability to reconcile disparate data in a coherent fashion.

6.3. Phylogenetic Display of Phenotypic Data

The categorized phenotypic data can be displayed alongside the phylogeny of the species under study, and we do so in Figure 6. The background color indicates the nature of the pelage; Blue indicates seasonal change, Yellow indicates concealment coloration matching snow; Green indicates both phenotypes are present. The branch coloration and pattern match hoarding behavior: Red indicates larder hoarding; Orange scatter hoarding; Violet the presence of both; and Green the presence of limited caching behavior. Dotted lines indicate scatter hoarding behavior is additionally present but only rarely. Spheres at leaf nodes indicate that species has body mass that varies strongly with season. Branch length 1.0 on leaf nodes indicates an unknown branch length due to lack of phylogenetic data.

As an additional note, the species *Clethrionomys Rufocanus*, *Clethrionomys Rutilus*, *Halichoerus Grypus*, *Martes Zibellina* and *Physeter Macrocephalus* were omitted due to a lack of precise phylogenetic data.

6.4. Convergent Phenotype Presence and Frequency

Hoarding behavior is found across the animal kingdom (103) but is far from universal, and more broadly varies in its role in the species' diet when present. Primary larder hoarding is widespread (31% of non-marine mammal species) but regular scatter hoarding (independently or shared with larder hoarding) is rarer (13% of non-marine mammal species), while bear hoarding behavior is primarily limited caching of food in place. Across multiple species, scatter hoarding behavior is associated with complicated memory structures in the hippocampus (110,111), suggesting a more complex path necessary to converge on this behavior compared to changes in coloration or simpler larder hoarding. However, when present it is a larger portion of the species diet than the larder hoarding in arctic animals, such as *Sorex* and *Microtus* (99,103). This suggests a reason for the lower prevalence via less mutational opportunity and higher constraints of purifying selection.

The arctic phenotypic data (see appendices) showed distinct distributions of the convergent adaptations. Gray and Brown are the most common colorations in mammals (112) and this holds true in arctic mammals, partly due to the presence of animals with wide ranges including non-arctic areas. Despite this, concealing coloration adapted to snow (36% of non-marine mammal species) and coloration changing seasonally (33% of non-marine mammal species) are common singular convergent phenotypes. This should not be surprising, as cryptic coloration is the primary driver of mammal pelage coloration (112) and the mutational opportunity for pigmentation is high, due to the presence of over 150 involved genes and the existence of multiple distinct alleles displaying the same phenotype (113).

Least common of the studied phenotypes was the indication of large swings in seasonal body mass (15% of non-marine mammal species), primarily bears and lemmings. The lower % here suggests that the use of stored body fat to handle periods of low food availability is disfavored versus alternate techniques, except in certain cases such as hibernation.

7. ANCESTRAL SOURCES OF PHENOTYPIC TRAITS

7.1. Summary

Ancestral reconstruction of the origin of traits amongst a subset of the animals under study was performed, examining the origin of these traits and whether convergence is clearly present. A subset was used because the number of related organisms made categorical analyses of all species in question impractical. The arctic and non-arctic species used are in Appendix 2. Recorded phenotypes and full trees with species names are in the supplementary material; images with no names are used here for clarity as the size of the trees prevents names from being clearly distinguishable.

Table 4 - Convergence Summary

Clades showing gain/loss of function within the clade may have convergence at the basal clade level.

	WHITE CONCEALING COLORATION	SEASONAL PELAGE	SEASONAL BODY MASS VARIATION	HOARDING BEHAVIOR
CLADES / SPECIES SHOWING CONVERGENCE	<i>Vulpes lagopus, Uncia, Ursus maritimus, Mustela erminea, Lynx (lynx, canadensis), Canis lupus (arctos, variaibilis, labradorius)</i>	<i>Lynx (lynx, canadensis); Mustela (putorius, nivalis, erminea), Martes (pennant, zibelina)*, Vulpes lagopus</i>	<i>Tremarctos ornatus, Pusa sibirica, Mephitis mephitis</i>	<i>Canis, Lyncalopex, Martes, Vomela peregusna, Ictonyx striatus, Ursus, Lutra, Lontra canadensis, Panthera, Spilogale, Hyaenidae, Meles, Neofelis, Gulo gulo, Prionallurus viverrinus, Eira Barbara, Urocyon cinerecargenteus, Taxidae taxus, Poecilogale albinucha, Mustela, Genetta, Felis (nigripes, silvestris), Lynx</i>

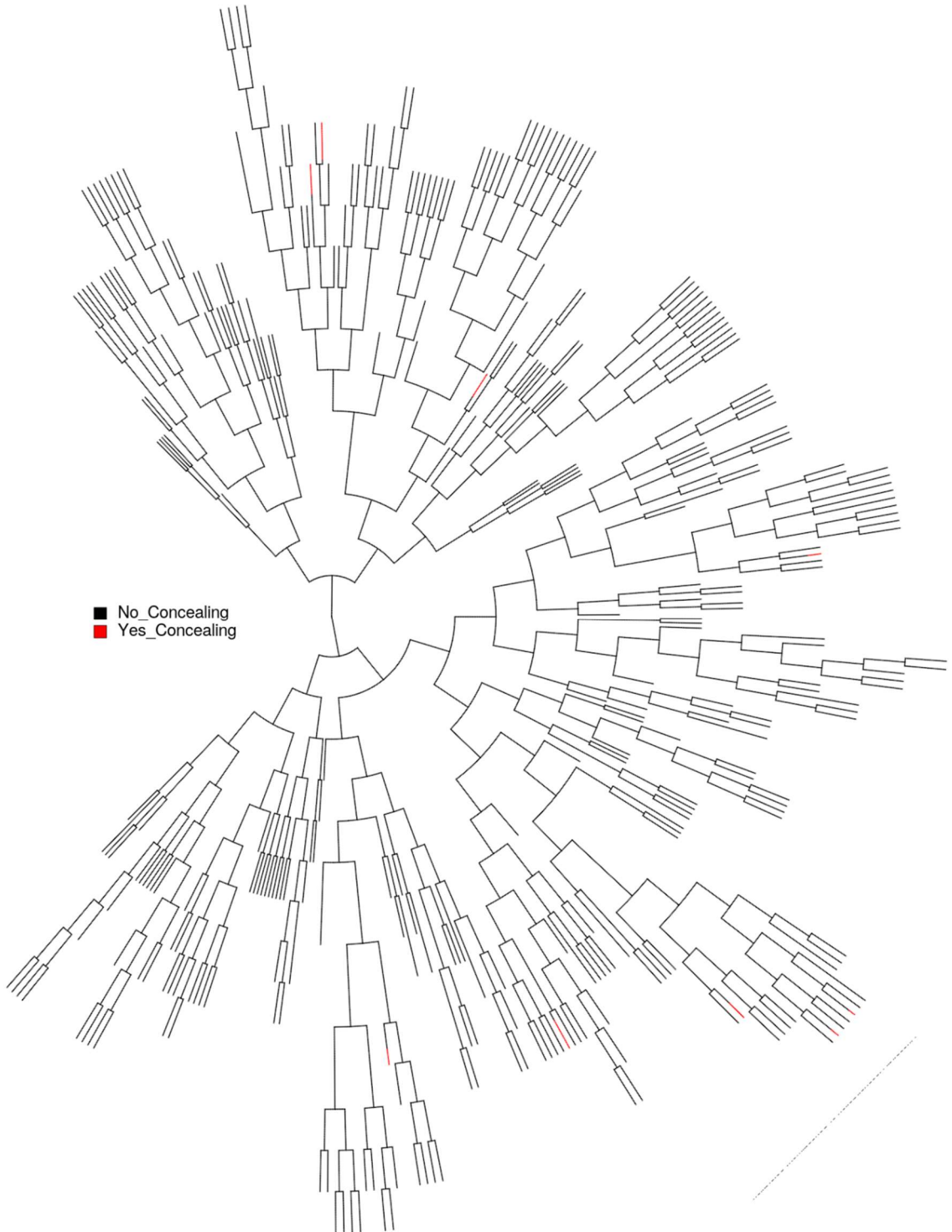
Table 4, Continued - Convergence Summary

	WHITE CONCEALING COLORATION	SEASONAL PELAGE	SEASONAL BODY MASS VARIATION	HOARDING BEHAVIOR
CLADES SHOWING GAIN/LOSS OF FUNCTION	<i>Lynx</i>	<i>Martes, Lynx</i>	<i>Ursus, Procyon, Meles</i>	<i>Vulpes, Mustela, Lynx, Canis</i>

7.2. Concealing Coloration

White concealing coloration arises independently in several distinct lineages, each separate from each other. This also occurs at the subspecies level; in *Canis lupus*, 3 subspecies have distinct coloration patterns compared to other subspecies; given that they are widely spaced amongst a significant number of subspecies, gain and loss of function is less likely as a driver than convergence. As white concealing coloration is a type of cryptic coloration, an extremely widespread adaptation(114), this is unsurprising. The presence of the trait in both *Lynx canadensis* and *Lynx lynx* suggests it may have evolved in the *Lynx* lineage and then lost by the non-arctic members of the lineage. However, as *Lynx lynx* is basal, the origin of the trait in the clade can be inferred to be convergent with the more distantly related species and subspecies.

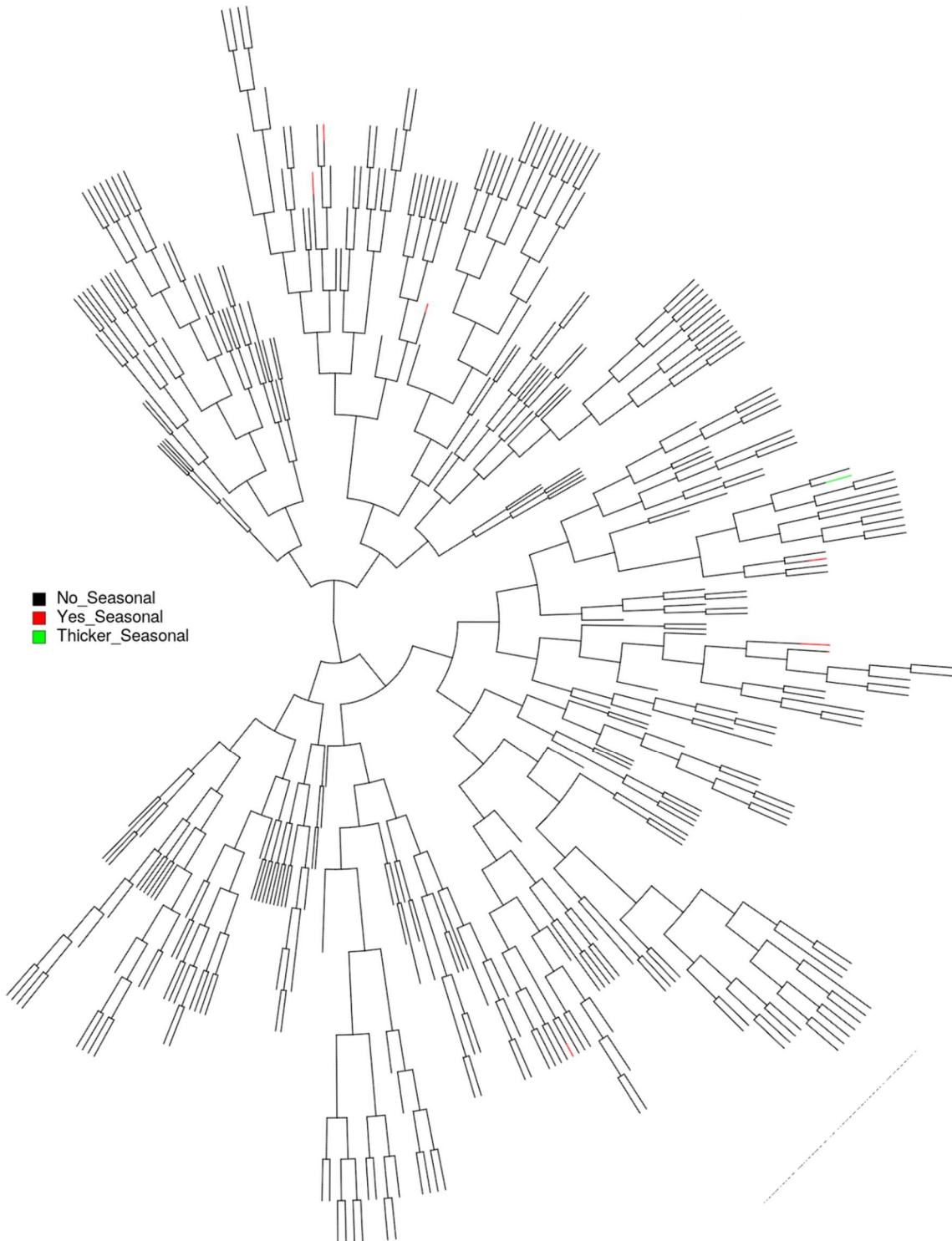
Figure 7 - Ancestral Reconstruction of Concealing Coloration



7.3. Seasonal Pelage

Multiple clades showed distinct seasonal pelage variation. *Lynx* and *Mustela* had several species with significant seasonal pelage variation not shared by their close phylogenetic relations, showing a strong indication of convergent evolution towards the trait. The *Martes* lineage also has species with the trait; however, while *Pennant* and *ibeline* were identified as obtaining the trait independently, they are closely related, near the bottom of the clade and there may have been a gain of function at the base of the clade, followed by a loss of function in the lineage leading to the rest of the genus. *Vulpes lagopus* (previously *Alopex lagopus*), which is endemic to the arctic unlike its cousin *Vulpes Vulpes*, retains strong seasonal pelage lacking in its non-endemic cousin.

Figure 8 - Seasonal Pelage Variation



7.4. Body-Mass Variation

Body Mass Variation is a trait that finds uses in non-arctic animals as well, and indeed the arctic but non-hibernating *Ursus maritimus* lacks the trait, while its cousins in the *Ursus arctos* lineage do not; here the trait was gained in an ancestor and lost by the arctic species. The species *Tremarctos ornatus* (which is non-arctic, but mountainous), *Pusa sibirica*, and *Mephitis mephitis* gain the trait independently of their close phylogenetic relationships, suggesting convergent adaptation drives the trait in some cases, but not all. In *Meles*, the arctic species and non-arctic species share the adaptation; the most basal species (*Meles leucrus* and *Meles canescens*) have ranges with extensive arctic or mountainous (which have similar selective pressure) regions, indicating a possibility that the phenotype evolved in response to that pressure and was retained as the ranges expanded. That the trait is adaptive to prey scarcity in non-arctic regions (115) and can be shared across a clade including non-arctic species indicates that the trait is not presumptively convergent due to the pressure of the arctic environment but must be considered in the light of broader phylogenetic data.

Figure 9 - Ancestral Reconstruction of Seasonal Body Mass



7.5. Hoarding Behavior

Hoarding behavior is common across species and is not limited to arctic regions; however, the trait is highly variable, as cache duration, location, method of storage, and role in the organism's food diet varies across species. For our investigation, the types of hoarding behavior have been classified into 8 groups (See Table 5). Data on hoarding behavior relies more extensively on extended observation and field work than the other phenotypic traits under study, so this classification is a best effort given available data. Sources included reference texts(99,100,103), websites(101,102), and specific papers(105,116–119) when data was not otherwise available.

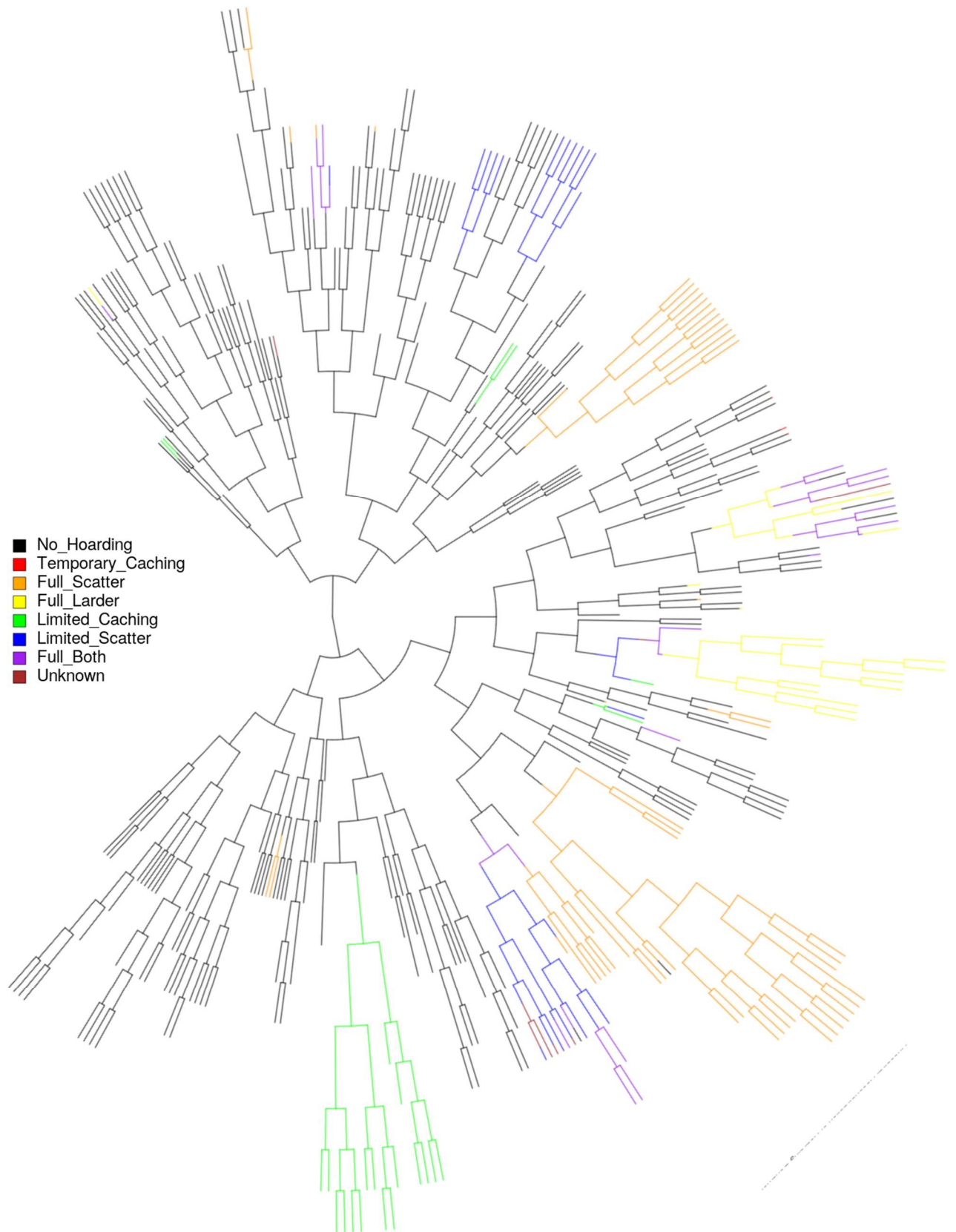
Table 5 - Types of Hoarding Behavior

No Hoarding	Full Both	Full Scatter	Full Larder	Temporary Caching	Limited Caching	Limited Scatter	Unknown
Organisms where no hoarding behavior has been identified.	Organisms that cache extensively both in a central larder and elsewhere.	Organisms that make extensive use of caching in scattered locations.	Organisms that cache food in a larder.	Organisms that cache food with no concealment temporarily.	Organisms that conceal food in place.	Organisms that scatter caches of a small subset of their diet.	Organisms sufficiently unknown that recorded data is not present.

Hoarding behavior is clearly convergent, as similar hoarding behavior arises across widely different clades. The link to the arctic biome is not consistent. In both arctic (*Gulo gulo*, *Vulpes*

lagopus) and non-arctic (*Eira Barbara*, *Urocyon cinereoargentus*) lineages there are individual species with divergent hoarding behavior to their neighbors, while similar to more distantly related species, although gain and loss occurs along certain lineages as well (*Mustela*). But in many cases this convergence applies to larger lineages than can cross arctic boundaries, including *Canis* and much of *Vulpes*. The complexities of the phenotype contribute to this; as the literature is qualitative observational studies, the data is not fully reliable and the variation in the trait is not fully captured. Arctic mammals (and other arctic organisms, such as birds) may share the categorical trait with their non-arctic cousins but use hoarding and caching to provide a much greater portion of their diet. But this possibility cannot be investigated given the current state of the literature, and would need a systematic observational analysis, and likely at significant expense. In the meantime, like body mass, evaluating the effect of the arctic environment on the trait needs strong consideration of the relevant phylogeny.

Figure 10 - Pattern of Hoarding Behavior in Mammals



8. NOVEL METHODOLOGY OF CONVERGENT EVOLUTION ANALYSIS

8.1. Background

Here we present a methodology of approaching Convergent Evolution analysis that will address the issues identified earlier. This can be applied when phenotypic data is categorized as described in the previous section. It additionally requires extensive genomic data, fully phylogenetically indexed. The protein analysis also requires extant protein structures to be associated with the genes.

Initial analysis was performed for the phenotypes described in section (6) but because of the lack of systematic genomic data, the analysis should be viewed as a pilot study, gearing up for a future where there are less sparse genomic resources for these species. Many of the convergent phenotypes lack enough genome sequences in extant datasets. The aquatic mammals are well-covered but lack most of the studied adaptations due to distinct evolutionary pressures. Land-bound arctic mammals possessing the majority of the phenotypic diversity under study do not have the same genomic coverage as the aquatic mammals yet.

We present the analysis as follows in order to describe the novel approach and make it available for future use.

8.2. Necessary Data Sources

Data is sourced from all genes associated with each species. Selection data is required for Pathway and Gene enrichment analysis; it is sourced via comparison with the ancestor determined via PAML phylogenetic analysis. PAML data is also used to source alignments to

identify sites that change relative to ancestral Protein structural and regional data is sourced via the PDB protein structures associated with each considered gene by TAED.

8.3. Pathway Analysis

Pathways are measured for enrichment by identifying both the number of genes associated with that pathway that are under selection and the total number of genes associated with the pathway. These counts are summed across all species sharing a phenotype for each of the phenotypes under investigation, and the proportion of genes in a pathway under selection is evaluated based on this summed count on a per-phenotype basis.

The same enrichment analysis is then performed separately for genes belonging to organisms in a phylogenetic relationship with each species under study; they are categorized according to the phenotypes of the species they are related to.

Both sets of analyses are evaluated for significance via bootstrapping, testing the null hypothesis of no effect from convergence, randomly distributing positive selection amongst a number of genes equal to the number identified by the analysis.

8.4. Selection Analysis

Genes are measured for enrichment by identifying the number of genes for each species that are under selection and the total number of genes for each species. These counts are summed across all species sharing a phenotype for each investigation, and the proportion of sites under selection is calculated on a per-phenotype basis. The same enrichment analysis is then performed separately for genes associated with organisms in a

phylogenetic relationship with each species under study; they are categorized according to the phenotypes of the species they are related to.

Both sets of analyses are evaluated for significance via bootstrapping, testing the null hypothesis of no effect from convergence, randomly distributing a number of selected genes equal to the number identified by the analysis. Protein Region Analysis

PDB records are associated with each gene for the species under selection via TAED, which was originally done via BLAST (49). DSSP data is also generated for each PDB for easier access for certain kinds of analysis, via the DSSP tool (73). For each protein, the location of substitutions is identified, classifying by domain, binding site, and secondary structure, calculating total substitutions per category, as well as total site counts. Site and substitution count are summed per phenotype for all proteins from organisms sharing that phenotype, and substitution rate and distribution is calculated based on those values. The results are analyzed for significance via bootstrapping against a null hypothesis of no effect of convergence, based on the random distribution of substitutions.

8.5. Sequence Analysis

Sequence data for each gene in the studied species and their ancestor is compiled. The ancestor and descendent sequences are compared, and the number of matched and unmatched substitutions are calculated. These totals are compiled for each phenotype according to the phenotypes each species belong to and the ratio of matched substitutions to overall substitutions is recorded. Random substitutions are then simulated via bootstrapping based on the total number of substitutions to determine significance.

8.6. Use in Examining Evolutionary Constraint

The analysis above, when applied to a large dataset, will help provide answers to several of the posed questions about the nature of evolution. The variation (or lack thereof) at the genome level between convergent phenotypes identified in the sequence will provide information on to what genetic constraint limits the genotype behind a phenotype. The gene and selection analyses provide information on how the evolution occurs and any constraints operating on a less comprehensive level. Pathway analysis will provide information on how pathway interactions mediate the other sets of analyses.

BIBLIOGRAPHY/REFERENCES CITED

1. Mendel G. Versuche liber Pflanzen- Hybriden. Verhandlungen des Naturforschenden Vereines, Abhandlungern. 1866;4:3–47.
2. Seward AC. Darwin and modern science. Z Indukt Abstamm VererbungsI [Internet]. 1910 Dec [cited 2019 Jun 27];3(1):109–10. Available from:
<http://link.springer.com/10.1007/BF02047720>
3. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature [Internet]. 1953 Apr [cited 2019 May 29];171(4356):737–8. Available from: <http://www.nature.com/articles/171737a0>
4. Ecker JR, Bickmore WA, Barroso I, Pritchard JK, Gilad Y, Segal E. ENCODE explained. Nature [Internet]. 2012 Sep 5 [cited 2019 May 29];489(7414):52–4. Available from: <http://www.nature.com/articles/489052a>
5. Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E. On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE. Genome Biol Evol [Internet]. 2013 [cited 2019 May 29];5(3):578–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23431001>
6. Graur D. An Upper Limit on the Functional Fraction of the Human Genome. Genome Biol Evol [Internet]. 2017 Jul 1 [cited 2019 Jun 28];9(7):1880–5. Available from: <https://academic.oup.com/gbe/article/9/7/1880/3952726>
7. Pellicer J, Fay MF, Leitch IJ. The largest eukaryotic genome of them all? Bot J Linn Soc [Internet]. 2010 Sep 15 [cited 2019 Jun 28];164(1):10–5. Available from: <https://academic.oup.com/botlinnean/article-lookup/doi/10.1111/j.1095-8339.2010.01072.x>

8. Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics* [Internet]. 2010 Dec 22 [cited 2019 Jun 30];11(1):724. Available from: <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-724>
9. Ku CS, Loy EY, Pawitan Y, Chia KS. The pursuit of genome-wide association studies: where are we now? *J Hum Genet* [Internet]. 2010 Apr 19 [cited 2019 May 29];55(4):195–206. Available from: <http://www.nature.com/articles/jhg201019>
10. Crick F. Central Dogma of Molecular Biology. *Nature* [Internet]. 1970 Aug [cited 2019 May 29];227(5258):561–3. Available from: <http://www.nature.com/articles/227561a0>
11. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* [Internet]. 2016 Jan 4 [cited 2019 Feb 7];44(D1):D457-62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26476454>
12. PAULING L, COREY RB, BRANSON HR. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* [Internet]. 1951 Apr [cited 2019 Jun 27];37(4):205–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14816373>
13. Chi PB, Liberles DA. Selection on protein structure, interaction, and sequence. *Protein Sci* [Internet]. 2016 Jul [cited 2019 Feb 7];25(7):1168–78. Available from: <http://doi.wiley.com/10.1002/pro.2886>
14. KIMURA M. Evolutionary Rate at the Molecular Level. *Nat* 1968 2175129 [Internet]. 1968 [cited 2019 Jun 25];217(5129):624. Available from: <https://www-nature-com.libproxy.temple.edu/articles/217624a0>
15. Noh S, Geist KS, Tian X, Strassmann JE, Queller DC. Genetic signatures of microbial

- altruism and cheating in social amoebas in the wild. *Proc Natl Acad Sci* [Internet]. 2018;115(12):201720324. Available from: <http://www.pnas.org/lookup/doi/10.1073/pnas.1720324115>
16. Hietpas RT, Jensen JD, Bolon DNA. Experimental Illumination of a fitness landscape. *Proc Natl Acad Sci U S A* [Internet]. 2011 [cited 2019 Jun 25];108(19):7896. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3093508/>
 17. Schmalhausen II. Factors of Evolution: The Theory of Stabilizing Selection. *J Am Med Assoc* [Internet]. 2011 Mar 25 [cited 2019 Jun 27];142(12):949. Available from: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1950.02910300087029>
 18. Stoltzfus A. On the possibility of constructive neutral evolution. *J Mol Evol*. 1999;
 19. Stoltzfus A. Constructive neutral evolution: Exploring evolutionary theory's curious disconnect. *Biol Direct*. 2012;7:1–13.
 20. Rieseberg LH, Widmer A, Arntz AM, Burke JM. Directional selection is the primary cause of phenotypic diversification. *Proc Natl Acad Sci U S A* [Internet]. 2002 Sep 17 [cited 2019 Jun 27];99(19):12242–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12221290>
 21. Mather K. Polymorphism as an Outcome of Disruptive Selection. *Evolution* (N Y) [Internet]. 1955 Mar 1 [cited 2019 Jun 27];9(1):52–61. Available from: <http://doi.wiley.com/10.1111/j.1558-5646.1955.tb01513.x>
 22. Fisher RA. The Genetical Theory of Natural Selection. *Proc Roy Soc Edinburgh*. 1930;42:321–41.
 23. Neher RA, Shraiman BI. Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics* [Internet]. 2011 Aug [cited 2019 Jun 30];188(4):975–96. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21625002>

24. Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* [Internet]. 2002;160(2):765–77. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11861577><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1461968>
25. Wade MJ, Kalisz S. The Causes of Natural Selection. *Evolution* (N Y). 2006;44(8):1947.
26. Conway Morris S. Evolution and convergence: some wider considerations. In: *The deep structure of biology: is convergence sufficiently ubiquitous to give a directional signal?* West Conshohocken , PA: Templeton Foundation Press; 2009. p. 46–67.
27. Orr HA. The probability of parallel evolution. *Evolution* (N Y) [Internet]. 2005 Jan 1 [cited 2019 Jun 27];59(1):216–20. Available from: <http://doi.wiley.com/10.1111/j.0014-3820.2005.tb00907.x>
28. Winemiller KO, Fitzgerald DB, Bower LM, Pianka ER. Functional traits, convergent evolution, and periodic tables of niches. *Ecol Lett*. 2015;18(8):737–51.
29. Pascoal S, Cezard T, Eik-Nes A, Gharbi K, Majewska J, Payne E, et al. Rapid convergent evolution in wild crickets. *Curr Biol*. 2014;24(12):1369–74.
30. Wake DB. Homoplasy: The Result of Natural Selection , or Evidence of Design Limitations? *Am Nat*. 1991;138(3):543–67.
31. Speed MP, Arbuckle K. Quantification provides a conceptual basis for convergent evolution. *Biological Reviews* [Internet]. 2016 May 1 [cited 2017 Nov 12];815–29. Available from: <http://doi.wiley.com/10.1111/brv.12257>
32. Kluge AG, Farris JS. Quantitative Phyletics and the Evolution of Anurans. *Syst Biol* [Internet]. 1969 Mar 1 [cited 2019 Jun 25];18(1):1–32. Available from: <https://academic.oup.com/sysbio/article-lookup/doi/10.1093/sysbio/18.1.1>
33. Archie JW. Homoplasy Excess Ratios: New Indices for Measuring Levels of Homoplasy in

- Phylogenetic Systematics and a Critique of the Consistency Index. *Syst Zool* [Internet]. 1989 Sep 1 [cited 2019 Jun 25];38(3):253. Available from:
<https://academic.oup.com/sysbio/article-lookup/doi/10.2307/2992286>
34. Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, et al. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* [Internet]. 2013 Oct 4 [cited 2018 Jun 7];502(7470):228–31. Available from:
<http://www.nature.com/articles/nature12511>
35. Darwin C. *On the Origin of Species by Means of Natural Selection*. 1859;502.
36. Daub JT, Moretti S, Davydov II, Excoffier L, Robinson-Rechavi M. Detection of Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. *Mol Biol Evol* [Internet]. 2017 [cited 2019 Jun 25];34(6):1391–402. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/28333345>
37. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, et al. Genotype to Phenotype: A Complex Problem. *Science* (80-) [Internet]. 2010 Apr 23 [cited 2019 Jul 23];328(5977):469–469. Available from:
<https://science.sciencemag.org/content/328/5977/469.full>
38. Van den Burg B, Eijsink VG. Selection of mutations for increased protein stability. *Current Opinion in Biotechnology*. 2002.
39. Chakrabarti R, Klibanov AM, Friesner RA. Computational prediction of native protein ligand-binding and enzyme active site sequences. *Proc Natl Acad Sci U S A* [Internet]. 2005 Jul 19 [cited 2019 Jun 30];102(29):10153–8. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/15998733>
40. O’Hara BP, Wilson SA, Lee AWL, Roe SM, Siligardi G, Drew RE, et al. Structural adaptation to selective pressure for altered ligand specificity in the *Pseudomonas aeruginosa* amide

- receptor, AmiC. *Protein Eng Des Sel* [Internet]. 2000 Feb 1 [cited 2019 Jun 30];13(2):129–32. Available from: <https://academic.oup.com/peds/article-lookup/doi/10.1093/protein/13.2.129>
41. Wetlaufer DB. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* [Internet]. 1973 Mar 1 [cited 2019 Jun 26];70(3):697–701. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/4351801>
 42. Hermansen RA, Northover DE, Shank SD, Liberles DA. Characterizing Lineage-Specific Evolution and the Processes Driving Genomic Diversification in Chordates. Unpublished. 2019;
 43. Bradshaw AD. The Croonian Lecture, 1991: Genostasis and the Limits to Evolution. *Philos Trans Biol Sci* [Internet]. 1991;333(1267):289–305. Available from: <http://www.jstor.org/stable/55615>
 44. Bridle JR, Vines TH. Limits to evolution at range margins: when and why does adaptation fail? *Trends Ecol Evol* [Internet]. 2007;22(3):140–7. Available from: <http://www.sciencedirect.com/science/article/pii/S0169534706003636>
 45. Eldredge N, Thompson JN, Brakefield PM, Gavrillets S, Jablonski D, Jackson JBC, et al. The dynamics of evolutionary stasis. *Paleobiology*. 2005;31(sp5):133–45.
 46. Hansen TF, Houle D. Evolvability, Stabilizing Selection, and the problem of stasis. *Phenotypic Integr Stud Ecol Evol complex phenotypes*. 2004;130–54.
 47. Antonovics J, van Tienderen PH. Ontoecogenophyloconstraints? The chaos of constraint terminology. *Trends Ecol Evol* [Internet]. 1991;6(5):166–8. Available from: <http://www.sciencedirect.com/science/article/pii/0169534791900597>
 48. Gould SJ. *Wonderful life: the Burgess Shale and the nature of history*. WW Norton & Company; 1990.

49. Hermansen RA, Oswald BP, Knight S, Shank SD, Northover D, Korunes KL, et al. The Adaptive Evolution Database (TAED): A New Release of a Database of Phylogenetically Indexed Gene Families from Chordates. *J Mol Evol* [Internet]. 2017 Aug 9 [cited 2019 Feb 7];85(1–2):46–56. Available from: <http://link.springer.com/10.1007/s00239-017-9806-8>
50. Marcovitz A, Turakhia Y, Gloude-mans M, Braun BA, Chen HI, Bejerano G. A novel unbiased test for molecular convergent evolution and discoveries in echolocating, aquatic and high-altitude mammals. *doi.org* [Internet]. 2017 Jul 31 [cited 2017 Nov 12];170985. Available from: <https://www.biorxiv.org/content/early/2017/07/31/170985>
51. Unckless RL, Orr HA. Dobzhansky–Muller incompatibilities and adaptation to a shared environment. *Hered* 2009 1023 [Internet]. 2009 Jan 14 [cited 2019 Jun 26];102(3):214. Available from: <https://www.nature.com/articles/hdy2008129>
52. Slatkin M. Gene Flow and the Geographic Structure of Natural Populations. *Am J Trop Med Hyg* [Internet]. 2010;82(2):235–42. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2813164&tool=pmcentrez&rendertype=abstract>
53. Hey J. Recent advances in assessing gene flow between diverging populations and species. *Curr Opin Genet Dev*. 2006;16(6):592–6.
54. Carmichael LE, Krizan J, Nagy JA, Fuglei E, Dumond M, Johnson D, et al. Historical and ecological determinants of genetic structure in arctic canids. *Mol Ecol*. 2007;16(16):3466–83.
55. Al ET, Dalén L, Fuglei EVA, Hersteinsson P, Kapel CMO, Roth JD, et al. Population history and genetic structure of a circumpolar species : the arctic fox. *Biol J Linn Soc*. 2005;79–89.
56. Scheffers BR, De Meester L, Bridge TCL, Hoffmann AA, Pandolfi JM, Corlett RT, et al. The

- broad footprint of climate change from genes to biomes to people. *Science* [Internet]. 2016 Nov 11 [cited 2019 Jun 26];354(6313):aaf7671. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27846577>
57. Prost S, Smirnov N, Fedorov VB, Sommer RS, Stiller M, Nagel D, et al. Influence of climate warming on arctic mammals? new insights from ancient DNA studies of the collared lemming *Dicrostonyx torquatus*. *PLoS One*. 2010;5(5).
58. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* [Internet]. 2013 Apr [cited 2019 Feb 7];30(4):772–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23329690>
59. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* [Internet]. 2010 Mar 29 [cited 2019 Feb 7];59(3):307–21. Available from: <https://academic.oup.com/sysbio/article/59/3/307/1702850>
60. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* [Internet]. 2000 Jan 1 [cited 2019 Feb 7];28(1):235–42. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.1.235>
61. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* [Internet]. 2017 Jul 1 [cited 2019 Feb 7];34(7):1812–9. Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msx116>
62. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, et al. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res* [Internet]. 2015 Jan [cited 2019 Feb 7];43(Database issue):D376-81. Available from:

- <http://www.ncbi.nlm.nih.gov/pubmed/25348408>
63. Northover D. TAED API Page [Internet]. Available from:
https://liberles.cst.temple.edu/TAED/TAED_API.php
 64. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* [Internet]. 1997 Sep 1 [cited 2019 Jul 22];25(17):3389–402. Available from:
<https://academic.oup.com/nar/article/25/17/3389/1061651>
 65. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* [Internet]. 2009 Jun 1 [cited 2019 Jul 19];25(11):1422–3. Available from:
<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp163>
 66. Maddison DR, Swofford DL, Maddison WP. Nexus: An Extensible File Format for Systematic Information. Cannatella D, editor. *Syst Biol* [Internet]. 1997 Dec 1 [cited 2019 Jul 21];46(4):590–621. Available from:
<http://academic.oup.com/sysbio/article/46/4/590/1629695>
 67. Olsen G. The Newick's 8:45 Tree Format Standard [Internet]. 1990. p. 1990. Available from: http://evolution.genetics.washington.edu/phylip/newick_doc.html
 68. Christian M. Zmasek. NHX - New Hampshire eXtended [version 2.0] [Internet]. Available from: <http://www.phylosoft.org/NHX/>
 69. Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, et al. Protein Data Bank (PDB): Database of Three-Dimensional Structural Information of Biological Macromolecules. *Acta Crystallogr Sect D Biol Crystallogr* [Internet]. 1998 Nov 1 [cited 2019 Jul 21];54(6):1078–84. Available from: <http://scripts.iucr.org/cgi->

bin/paper?S0907444998009378

70. Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* [Internet]. 1990 Jan 1 [cited 2019 Jul 21];183(C):63–98. Available from: <https://www.sciencedirect.com/science/article/pii/007668799083007V>
71. Retief JD. Phylogenetic Analysis Using PHYLIP. In: Misener S, Krawetz SA, editors. *Bioinformatics Methods and Protocols* [Internet]. Totowa, NJ: Humana Press; 1999. p. 243–58. Available from: <https://doi.org/10.1385/1-59259-192-2:243>
72. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* [Internet]. 2007 Apr 18 [cited 2019 Jul 21];24(8):1586–91. Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msm088>
73. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* [Internet]. 1983 Dec [cited 2019 Feb 7];22(12):2577–637. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6667333>
74. Mukherjee S, Kim S, Ramanan VK, Gibbons LE, Nho K, Glymour MM, et al. Gene-based GWAS and biological pathway analysis of the resilience of executive functioning. *Brain Imaging Behav* [Internet]. 2014 Mar;8(1):110–8. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24072271>
75. Jia P, Wang L, Meltzer HY, Zhao Z. Pathway-based analysis of GWAS datasets: effective but caution required. *Int J Neuropsychopharmacol*. 2011 May;14(4):567–72.
76. Collins C, Didelot X. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLOS Comput Biol* [Internet]. 2018 Feb 5;14(2):e1005958. Available from: <https://doi.org/10.1371/journal.pcbi.1005958>

77. Talevich E, Invergo BM, Cock PJ, Chapman BA. Bio.Phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinformatics* [Internet]. 2012 Dec 21 [cited 2019 Jul 19];13(1):209. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-209>
78. Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, et al. Bio++: Efficient Extensible Libraries and Tools for Computational Molecular Evolution. *Mol Biol Evol* [Internet]. 2013 Aug 1 [cited 2019 Jul 19];30(8):1745–50. Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst097>
79. Northover DE, Shank SD, Liberles DA. Characterizing lineage-specific evolution and the processes driving genomic diversification in chordates. *BMC Evol Biol*. 2020 Feb;20(1):24.
80. Thorne JL. Models of protein sequence evolution and their applications. *Curr Opin Genet Dev*. 2000;10(6):602–5.
81. Perron U, Moal IH, Thorne JL, Goldman N. Probabilistic Models for the Study of Protein Evolution [Internet]. *Handbook of Statistical Genomics*. 2019. p. 330–47. (Wiley Online Books). Available from: <https://doi.org/10.1002/9781119487845.ch12>
82. Siltberg-Liberles J, Grahnen JA, Liberles DA. The evolution of protein structures and structural ensembles under functional constraint. *Genes (Basel)* [Internet]. 2011 Oct 28;2(4):748–62. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24710290>
83. Parthiban V, Gromiha MM, Hoppe C, Schomburg D. Structural analysis and prediction of protein mutant stability using distance and torsion potentials: Role of secondary structure and solvent accessibility. *Proteins Struct Funct Bioinforma* [Internet]. 2007 Jan 1;66(1):41–52. Available from: <https://doi.org/10.1002/prot.21115>
84. Podtelezhnikov AA, Wild DL. Reconstruction and stability of secondary structure

- elements in the context of protein structure prediction. *Biophys J* [Internet]. 2009 Jun 3;96(11):4399–408. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/19486664>
85. Finkelstein A V, Badretdinov AY, Ptitsyn OB. Physical reasons for secondary structure stability: α -Helices in short peptides. *Proteins Struct Funct Bioinforma* [Internet]. 1991 Apr 1;10(4):287–99. Available from: <https://doi.org/10.1002/prot.340100403>
86. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum allowed solvent accessibilities of residues in proteins. *PLoS One* [Internet]. 2013 [cited 2019 Feb 7];8(11):e80635. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24278298>
87. Chi PB, Kim D, Lai JK, Bykova N, Weber CC, Kubelka J, et al. A new parameter-rich structure-aware mechanistic model for amino acid substitution during evolution. *Proteins Struct Funct Bioinforma* [Internet]. 2018 Feb 1 [cited 2019 Jul 16];86(2):218–28. Available from: <http://doi.wiley.com/10.1002/prot.25429>
88. KENDREW JC, BODO G, DINTZIS HM, PARRISH RG, WYCKOFF H, PHILLIPS DC. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* [Internet]. 1958;181(4610):662–6. Available from: <https://doi.org/10.1038/181662a0>
89. Kaufmann WE, Abrams MT, Chen W, Reiss AL. Genotype, molecular phenotype, and cognitive phenotype: Correlations in fragile X syndrome. *Am J Med Genet*. 1999;83(4):286–95.
90. Wheelock MJ, Johnson KR. Cadherins as Modulators of Cellular Phenotype. *Annu Rev Cell Dev Biol* [Internet]. 2003 Nov 1;19(1):207–35. Available from: <https://doi.org/10.1146/annurev.cellbio.19.011102.111135>
91. Kukekova A V., Temnykh S V., Johnson JL, Trut LN, Acland GM. Genetics of behavior in the silver fox. *Mamm Genome*. 2012;23(1–2):164–77.
92. Dawkins R. *The extended phenotype*. Vol. 8. Oxford University Press Oxford; 1982.

93. Kukekova A V., Trut LN, Oskina IN, Johnson JL, Temnykh S V., Kharlamova A V., et al. A meiotic linkage map of the silver fox, aligned and compared to the canine genome. *Genome Res.* 2007;17(3):387–99.
94. Våge DI, Lu D, Klungland H, Lien S, Adalsteinsson S, Cone RD. A non-epistatic interaction of agouti and extension in the fox, *Vulpes vulpes*. *Nat Genet.* 1997;15(3):311.
95. Groth P, Pavlova N, Kalev I, Tonov S, Georgiev G, Pohlenz H-D, et al. PhenomicDB: a new cross-species genotype/phenotype resource. *Nucleic Acids Res [Internet]*. 2007 Jan [cited 2019 Jul 12];35(Database issue):D696-9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16982638>
96. Smith CL, Eppig JT. The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mamm Genome [Internet]*. 2012 Oct 9 [cited 2017 Sep 29];23(9–10):653–68. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22961259>
97. Köhler S, Doelken SC, Ruef BJ, Bauer S, Washington N, Westerfield M, et al. Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. *F1000Research [Internet]*. 2014;2(30). Available from: <http://openr.es/2td>
98. Robinson PN, Webber C. Phenotype Ontologies and Cross-Species Analysis for Translational Research. *PLoS Genet.* 2014;10(4).
99. Sale R, Michelsen P. A complete guide to Arctic wildlife. Firefly Books; 2012.
100. Shirihai H. A complete guide to Antarctic wildlife.
101. International Union for Conservation of Nature. Red List for Threatened Species [Internet]. 2020. Available from: <https://www.iucnredlist.org>
102. Myers P, Espinosa R, Parr CS, Jones T, Hammond GS, Dewey TA. Animal Diversity Web

[Internet]. Available from: <http://animaldiversity.org/about/>

103. Vander Wall SB. Food hoarding in animals. University of Chicago Press; 1990.
104. Braun JK, Eaton TS, Mares MA. *Marmota Caligata* (Rodentia: Sciuridae). *Mamm Species*. 2011;43(884):155–71.
105. Allen ML, Wittmer HU, Ingaki A, Yamazaki K, Koike S. Food caching by bears: a literature review and new observations for Asiatic and American black. *Zoology*. 2020;21:267–72.
106. Wang JM, Zhang YM, Wang DH. Photoperiodic regulation in energy intake, thermogenesis and body mass in root voles (*Microtus oeconomus*). *Comp Biochem Physiol - A Mol Integr Physiol*. 2006;145(4):546–53.
107. Stirling I, Thiemann GW, Richardson E. Quantitative Support for a Subjective Fatness Index for Immobilized Polar Bears. *J Wildl Manage*. 2008;72(2):568–74.
108. Nieminen P, Mustonen AM. Uniform fatty acid mobilization from anatomically distinct fat depots in the sable (*Martes zibellina*). *Lipids*. 2007;42(7):659–69.
109. Aguilar A, Borrell A. Patterns of Lipid Content and Stratification in the Blubber of Fin Whales (*Balaenoptera physalus*). *J Mammal*. 1990;71(4):544–54.
110. Lucas JR, Brodin A, De Kort SR, Clayton NS. Does hippocampal size correlate with the degree of caching specialization? *Proc R Soc B Biol Sci* [Internet]. 2004 [cited 2017 Nov 11];271(1556):2423–9. Available from: <http://www.jstor.org.libproxy.temple.edu/stable/4142998>
111. Pravosudov V V., Clayton NS. A test of the adaptive specialization hypothesis: Population differences in caching, memory, and the hippocampus in black-capped chickadees (*Poecile atricapilla*). *Behav Neurosci* [Internet]. 2002 [cited 2017 Nov 11];116(4):515–22. Available from: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0735-7044.116.4.515>
112. Caro T. The colours of extant mammals. *Semin Cell Dev Biol* [Internet]. 2013 Jun 1 [cited

2019 Jul 22];24(6–7):542–52. Available from:

<https://www.sciencedirect.com/science/article/pii/S1084952113000529>

113. Hofreiter M, Schöneberg T. The genetic and evolutionary basis of colour variation in vertebrates. *Cell Mol Life Sci* [Internet]. 2010 Aug 14 [cited 2019 Jul 22];67(15):2591–603. Available from: <http://link.springer.com/10.1007/s00018-010-0333-7>
114. Endler JA. Disruptive and cryptic coloration. *Proceedings Biol Sci* [Internet]. 2006 Oct 7;273(1600):2425–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/16959630>
115. Poulle M-L, Crête M, Huot J. Seasonal variation in body mass and composition of eastern coyotes. *Can J Zool* [Internet]. 1995 Sep 1;73(9):1625–33. Available from: <https://doi.org/10.1139/z95-193>
116. Careau V, Giroux J-F, Berteaux D. Cache and carry: hoarding behavior of arctic fox. *Behav Ecol Sociobiol* [Internet]. 2007 Oct 11 [cited 2017 Nov 1];62(1):87–96. Available from: <http://link.springer.com/10.1007/s00265-007-0441-z>
117. Rasmussen JL, Tilson RL. Food provisioning by adult maned wolves (*Chrysocyon brachyurus*). *Z Tierpsychol*. 1984;65(4):346–52.
118. Sovada MA, Roy CC, Telesco DJ. Seasonal food habits of swift fox (*Vulpes velox*) in cropland and rangeland landscapes in western Kansas. *Am Midl Nat*. 2001;145(1):101–11.
119. Geffen E, Hefner R, Macdonald DW, Ucko M. Diet and foraging behavior of Blanford's foxes, *Vulpes cana*, in Israel. *J Mammal*. 1992;73(2):395–402.

APPENDIX 1

ARCTIC CATEGORICAL PHENOTYPIC DATA

Species	<i>Snow Concealment?</i>	<i>Seasonal Pel- age?</i>	<i>Hoarding Behavior</i>	<i>Seasonal Bo Mass Stability</i>
alces_alces	No	No	No Hoarding	Stable
vulpes_lagopus	Yes	Yes	Scatter & Larder	Stable
balaena_mysticetus	No	No	No Hoarding	Stable
balaenoptera_acutorostrata	No	No	No Hoarding	Stable
balaenoptera_borealis	No	No	No Hoarding	Stable
balaenoptera_musculus	No	No	No Hoarding	Stable
balaenoptera_physalus	No	No	No Hoarding	Stable
berardius_bairdii	No	No	No Hoarding	Stable
callorhinus_ursinus	No	No	No Hoarding	Stable
canis_lupus	Yes	No	No Hoarding	Stable
castor_canadensis	No	No	Larder	Stable
urocitellus_parryii	No	No	Larder	Stable
clethrionomys_rufocanus	No	No	No Hoarding	Stable
clethrionomys_rutilus	No	No	No Hoarding	Stable
cystophora_cristata	No	No	No Hoarding	Seasonal
delphinapterus_leucas	No	No	No Hoarding	Seasonal
dicrostonyx_groenlandicus	Yes	Yes	No Hoarding	Seasonal
dicrostonyx_hudsonius	Yes	Yes	No Hoarding	Seasonal
dicrostonyx_richardsoni	Yes	Yes	No Hoarding	Seasonal
dicrostonyx_torquatus	Yes	Yes	No Hoarding	Stable
dicrostonyx_vinogradovi	Yes	Yes	No Hoarding	Stable
enhydra_lutris	No	No	No Hoarding	Stable
erignathus_barbatus	No	No	No Hoarding	Stable
eschrichtius_robustus	No	No	No Hoarding	Stable
eubalaena_glacialis	No	No	No Hoarding	Stable
eumetopias_jubatus	No	No	No Hoarding	Stable
gulo_gulo	No	No	Scatter & Larder	Stable
halichoerus_grypus	No	No	No Hoarding	Stable
hydrurga_leptonyx	No	No	No Hoarding	Stable
hyperoodon_ampullatus	No	No	No Hoarding	Stable
lagenorhynchus_albirostris	No	No	No Hoarding	Stable
lemmus_sibiricus	No	No	No Hoarding	Stable
lemmus_trimucronatus	No	No	No Hoarding	Stable
leptonychotes_weddellii	No	No	No Hoarding	Stable
lepus_americanus	Yes	Yes	No Hoarding	Stable
lepus_arcticus	Yes	Yes	No Hoarding	Stable
lepus_othus	Yes	Yes	No Hoarding	Stable
lepus_timidus	Yes	Yes	No Hoarding	Stable

lutra_canadensis	No	No	No Hoarding	Stable
lutra_lutra	No	No	No Hoarding	Stable
lynx_canadensis	Yes	Yes	Scatter & Larder	Stable
lynx_lynx	Yes	Yes	Scatter & Larder	Stable
marmota_broweri	No	No	No Hoarding	Stable
marmota_caligata	No	No	No Hoarding	Stable
marmota_camtschatica	No	No	No Hoarding	Stable
martes_zibellina	No	No	No Hoarding	Stable
megaptera_novaeangliae	No	No	No Hoarding	Stable
mesoplodon_stejnegeri	No	No	No Hoarding	Stable
microtus_abbreviatus	No	No	Larder	Stable
microtus_gregalis	No	No	Larder	Stable
microtus_hyperboreus	No	No	Larder	Stable
microtus_miurus	No	No	Larder	Stable
microtus_oecconomus	No	No	Larder	Stable
microtus_pennsylvanicus	No	No	Larder	Stable
microtus_xanthognathus	No	No	Larder	Stable
monodon_monoceros	No	No	Larder	Stable
mustela_erminea	Yes	Yes	Scatter & Larder	Stable
mustela_nivalis	Yes	Yes	Scatter & Larder	Stable
mustela_vison	No	No	No Hoarding	Stable
myopus_schisticolor	No	No	No Hoarding	Stable
ochotona_hyperborea	No	Yes	Scatter	Stable
odobenus_rosmarus	No	No	No Hoarding	Stable
ondatra_zibethicus	No	No	Larder	Stable
orcinus_orca	No	No	No Hoarding	Stable
ovibos_moschatus	No	No	No Hoarding	Stable
ovis_dalli	Yes	No	No Hoarding	Stable
ovis_nivicola	No	No	No Hoarding	Stable
phoca_fasciata	No	No	No Hoarding	Stable
phoca_groenlandica	No	No	No Hoarding	Stable
phoca_largha	No	No	No Hoarding	Stable
phoca_vitulina	No	No	No Hoarding	Stable
phocoena_phocoena	No	No	No Hoarding	Stable
physeter_macrocephalus	No	No	No Hoarding	Stable
phocoenoides_dalli	No	No	No Hoarding	Stable
pusa_hispida	No	No	No Hoarding	Stable
rangifer_tarandus	Yes	Yes	No Hoarding	Stable
sorex_caecutiens	No	No	Larder	Stable
sorex_cinereus	No	No	Larder	Stable
sorex_hydrodomus	No	No	Larder	Stable
sorex_jacksoni	No	No	Larder	Stable
sorex_monticolus	Yes	Yes	Larder	Stable
sorex_portenkoi	Yes	Yes	Larder	Stable
sorex_tundrensis	Yes	Yes	Larder	Stable
sorex_ugyunak	Yes	Yes	Larder	Stable

synaptomys_borealis	No	No	No Hoarding	Stable
ursus_americanus	No	No	Limited Caching	Stable
ursus_arctos	No	No	Limited Caching	Seasonal
ursus_maritimus	Yes	No	Limited Caching	Stable
vulpes_vulpes	No	No	Scatter & Larder	Seasonal

APPENDIX 2

ARCTIC ADDITIONAL PHENOTYPIC DATA

<i>Species</i>	<i>Pelage (Winter)</i>	<i>Pelage (Summer)</i>	<i>Weight (kg)</i>
<i>alces alces</i>	Dark Brown; Underparts: Paler	Dark Brown; Underparts: Paler	550
<i>alopex lagopus</i>	White	Brown	6
<i>balaena mysticetus</i>	Black Underparts: White	Black Underparts: White	77111
<i>balaenoptera acutorostrata</i>	Skin (Dorsal: Dark Gray; Ventral: White)	Skin (Dorsal: Dark Gray; Ventral: White)	9979
<i>balaenoptera borealis</i>	Skin (Dorsal: Gray; Ventral: Paler)	Skin (Dorsal: Gray ; Ventral: Paler)	13381
<i>balaenoptera musculus</i>	Skin (Blue Gray)	Skin (Blue Gray)	110223
<i>balaenoptera physalus</i>	Skin (Slate Gray Dorsal; Ventral: Paler)	Skin (Slate Gray Dorsal; Paler Ventral)	63503
<i>berardius bairdii</i>	Skin (Dark Gray-Brown)	Skin (Dark Gray-Brown)	9525
<i>callorhinus ursinus</i>	M: Dark Brown F: Dorsal: Gray-Brown; Ventral: Paler	M: Dark Brown F: Dorsal: Gray-Brown; Ventral: Paler	120
<i>canis lupus</i>	White	White	55
<i>castor canadensis</i>	Dark Brown	Dark Brown	24.5
<i>urocitellus parryii</i>	Tawny Upperparts: Flecked	Tawny Upperparts: Flecked	0.69
<i>clethrionomys rufocanus</i>	Chestnut; Gray; Underparts: Pale Gray	Chestnut; Gray; Underparts: Pale Gray	0.0325
<i>clethrionomys rutilus</i>	Red-Brown; Golden-Brown; Underparts: Cream Pale Buff	Red-Brown; Golden-Brown; Underparts: Cream Pale Buff	0.0275
<i>cystophora cristata</i>	Silver-Gray Mottled: Black	Silver-Gray Mottled: Black	247.5
<i>delphinapterus leucas</i>	Skin (Creamy White)	Skin (Creamy White)	1000
<i>dicrostonyx groenlandicus</i>	White	Gray-Brown; Mixed; Underparts: Paler	0.045
<i>dicrostonyx hudsonius</i>	White	Gray-Brown Dorsal: Black	0.0625
<i>dicrostonyx richardsoni</i>	White	Dark Red-Brown	0.0625
<i>dicrostonyx torquatus</i>	White	Red-Brown; Mixed; Underparts: Paler	0.0625
<i>dicrostonyx vinogradovi</i>	White	Pale-Gray Patches: Gray-Cream	0.075
<i>enhydra lutris</i>	Variable-Brown	Variable-Brown	30
<i>erignathus barbatus</i>	Gray-Brown	Gray-Brown	250
<i>eschrichtius robustus</i>	Skin (Dark Gray; Light Gray; White)	Skin (Dark Gray; Light Gray; White)	25401
<i>eubalaena glacialis</i>	Skin (Black Patches: White)	Skin (Black Patches: White)	76657

<i>eumetopias jubatus</i>	Buff Red-Brown	Buff Red-Brown	600
<i>gulo gulo</i>	Dark Brown Stripes: Yellow-Brown	Dark Brown Stripes: Yellow-Brown	24
<i>halichoerus grypus</i>	Dark Gray Gray Brown Patches: Light Gray	Dark Gray Gray Brown Patches: Light Gray	212.5
<i>hydrurga leptonyx</i>	Blue-Gray Ventral: Paler	Blue-Gray Ventral: Paler	400
<i>hyperoodon ampullatus</i>	Skin (Dorsal: Brown Ventral: Paler)	Skin (Dorsal: Brown Ventral: Paler)	7484
<i>lagenorhynchus albirostris</i>	Skin (Dorsal: Black Patches: Gray-White Ventral: White)	Skin (Dorsal: Black Patches: Gray-White Ventral: White)	200
<i>lemmus sibiricus</i>	Brown-Yellow Stripes: Black	Brown-Yellow Stripes: Black	0.09
<i>lemmus trimucronatus</i>	Tawny Spots: Brown Buff Underparts: Paler	Tawny Spots: Brown Buff Underparts: Paler	0.0925
<i>leptonychotes weddellii</i>	Blue-Gray Spots: Grey-White	Blue-Gray Spots: Grey-White	500
<i>lepus americanus</i>	White	Red-Brown Underparts: Pale Gray	1.55
<i>lepus arcticus</i>	White	Gray Gray-Brown	4.5
<i>lepus othus</i>	White	Red-Brown Underparts: Paler	5.45
<i>lepus timidus</i>	White	Dark Brown Grey Brown; Underparts: Paler	3.9
<i>lutra canadensis</i>	Dark Brown Underparts: Cream	Dark Brown Underparts: Cream	9.5
<i>lutra lutra</i>	Dark Brown Underparts: Cream	Dark Brown Underparts: Cream	12
<i>lynx canadensis</i>	Mod: Whiter; Mod: Paler	Yellow/Brown; Pale Cream Below	12.5
<i>lynx lynx</i>	Mod: Grayer; Mod: Thicker	Gray Gray Brown; Spots: Dark	21.5
<i>marmota broweri</i>	Gray-Silver	Gray-Silver	3.25
<i>marmota caligata</i>	Silver-Gray, Reddish-Brown, Black	Silver-Gray, Reddish-Brown, Black	5.375
<i>marmota camtschatica</i>	Brown Underparts: Paler	Brown Underparts: Paler	4
<i>martes zibellina</i>	Dark Brown Patches: Paler	Dark Brown Patches: Paler	1.45
<i>megaptera novaeangliae</i>	Skin (Dark-Gray Black)	Skin (Dark-Gray Black)	3855
<i>mesoplodon stejnegeri</i>	Skin (Dark Gray Ventral: Paler)	Skin (Dark Gray Ventral: Paler)	
<i>microtus abbreviatus</i>	Brown; Yellow Brown; Underparts: Paler	Brown; Yellow Brown; Underparts: Paler	0.06
<i>microtus gregalis</i>	Light-Ochre Dark-Brown Spots: Darker Underparts: Greyish_Brown	Light-Ochre Dark-Brown Spots: Darker Underparts: Greyish_Brown	0.065

<i>microtus hyperboreus</i>	Gray-Brown Underparts: Silver-Gray	Gray-Brown Underparts: Silver-Gray	0.07
<i>microtus miurus</i>	Dark-Brown; Tawny; Un- derparts: Gray-Brown	Dark-Brown; Tawny; Un- derparts: Gray-Brown	0.04
<i>microtus oeconomus</i>	Dark-Brown Spots: Buff; Underparts: Pale-Buff	Dark-Brown Spots: Buff; Underparts: Pale Buff	0.045
<i>microtus pennsylvanicus</i>	Red-Brown Dark Brown; Underparts: Pale- Gray-Brown	Red-Brown Dark Brown; Underparts: Pale- Gray-Brown	0.05
<i>microtus xanthognathus</i>	Dark-Brown Gray- Brown Patch: Yellow- Brown; Underparts: Paler	Dark-Brown Gray- Brown Patch: Yellow- Brown; Underparts: Paler	0.14
<i>monodon monoceros</i>	Skin (Blue Dark Gray Mottled: White)	Skin (Blue Dark Gray Mottled: White)	1200
<i>mustela erminea</i>	White	Brown Underparts: White	0.165
<i>mustela nivalis</i>	White	Brown Underparts: White	0.0425
<i>mustela vison</i>	Dark Brown Patches: White	Dark Brown Patches: White	0.9
<i>myopus schisticolor</i>	Dark-Gray Underparts: Paler	Dark-Gray Underparts: Paler	0.0325
<i>ochotona hyperborea</i>	Gray-Brown Reddish- Brown	Brown Dark Brown	127.5
<i>odobenus rosmarus</i>	Pink to Brown (Blood Flow Variation)	Pink to Brown (Blood Flow Variation)	1350
<i>ondatra zibethicus</i>	Red-Brown Underparts: Paler	Red-Brown Underparts: Paler	1.2
<i>orcinus orca</i>	Dorsal: Black Ventral: White	Dorsal: Black Ventral: White	4082
<i>ovibos moschatus</i>	Dark-Brown Gray- Brown	Dark-Brown Gray- Brown	310
<i>ovis dalli</i>	White	White	80
<i>ovis nivicola</i>	Brown Underparts: Paler	Brown Underparts: Paler	85
<i>phoca fasciata</i>	Dark Brown Bands: Pale Gray Cream	Dark Brown Bands: Pale Gray Cream	77.5
<i>phoca groenlandica</i>	Silver-Gray Highlights: Black	Silver-Gray Highlights: Black	227.5
<i>phoca largha</i>	Dorsal: Pale Gray Gray Brown Mid-Brown Ven- tral: Paler	Pale Gray Gray Brown Mid-Brown	95
<i>phoca vitulina</i>	(Silvery-Gray Buff w/ Dark Gray) Dorsal: Dark Grey-Brown Dark Brown	(Silvery-Gray Buff w/ Dark Gray) Dorsal: Dark Grey-Brown Dark Brown	90
<i>phocoena phocoena</i>	Dark gray Pale gray	Dark gray Pale gray	57.5

<i>physeter macrocephalus</i>	Skin (Dark Gray Patches: White)	Skin (Dark Gray Patches: White)	29483.5
<i>phocoenoides dalli</i>	Skin (Black Patches: White Ventral: White)	Skin (Black Patches: White Ventral: White)	190
<i>pusa hispida</i>	Dorsal: Dark-Gray Gray-Brown Rings: Pale-Gray; Ventral: Silver Silver-Brown Gray-Brown	Dorsal: Dark-Gray Gray-Brown Rings: Pale-Gray; Ventral: Silver Silver-Brown Gray-Brown	77.5
<i>rangifer tarandus</i>	Mod: Paler	Dark Brown Gray Brown; Underparts: Paler	105
<i>sorex caecutiens</i>	Dark-Brown Underparts: Silver-White	Dark-Brown Underparts: Silver-White	0.0055
<i>sorex cinereus</i>	Brown Grey-Brown; Underparts: Paler	Brown Grey-Brown; Underparts: Paler	0.004
<i>sorex hydrodromus</i>	Brown Underparts: Paler	Brown Underparts: Paler	0.0035
<i>sorex jacksoni</i>	Brown; Buff Buff-Gray	Brown; Buff Buff-Gray	0.0045
<i>sorex monticolus</i>	Mod: Darker	Red-Brown Underparts: Paler	0.0075
<i>sorex portenkoi</i>	Dark-Brown; Pale-Gray	Brown; Yellow Brown; Underparts: Paler	0.00375
<i>sorex tundrensis</i>	Dark-Brown; Pale-Gray	Dark-Brown; Pale-Gray-Brown; Underparts: Pale Gray	0.0075
<i>sorex ugyunak</i>	Mod: Darker	Brown; Pale-Brown	0.00425
<i>synaptomys borealis</i>	Gray-Brown Patches: Gray; Red-Brown Patches: Gray	Gray-Brown Patches: Gray; Red-Brown Patches: Gray	0.03
<i>ursus americanus</i>	Black	Black	82.5
<i>ursus arctos</i>	Dark-Brown	Dark-Brown	435
<i>ursus maritimus</i>	Pale-Yellow	Pale-Yellow	362.5
<i>vulpes vulpes</i>	Red-Brown Underparts: White	Red-Brown Underparts: White	7

APPENDIX 3

PHENOTYPIC DATA FOR ANCESTRAL RECONSTRUCTION

<i>Species</i>	<i>Snow Concealment?</i>	<i>Seasonal Pelage?</i>	<i>Hoarding Behavior</i>	<i>Seasonal Bo Mass Stability</i>
<i>Alces alces</i>	No	No	No Hoarding	Stable
<i>Alces alces alces</i>	No	No	No Hoarding	Stable
<i>Alopex lagopus</i>	Yes	Yes	Full Both	Stable
<i>Balaenoptera acutorostrata</i>	No	No	No Hoarding	Stable
<i>Balaenoptera acutorostrata scam- moni</i>	No	No	No Hoarding	Stable
<i>Balaenoptera borealis</i>	No	No	No Hoarding	Stable
<i>Balaenoptera musculus</i>	No	No	No Hoarding	Stable
<i>Balaenoptera physalus</i>	No	No	No Hoarding	Stable
<i>Callorhinus ursinus</i>	No	No	No Hoarding	Stable
<i>Canis lupus arctos</i>	Yes	No	Full Scatter	Stable
<i>Canis lupus variabilis</i>	Yes	No	Full Scatter	Stable
<i>Canis lupus labradorius</i>	Yes	No	Full Scatter	Stable
<i>Cystophora cristata</i>	No	No	No Hoarding	Stable
<i>Delphinapterus leucas</i>	No	No	No Hoarding	Stable
<i>Dicrostonyx groenlandicus</i>	Yes	Yes	No Hoarding	Seasonal
<i>Dicrostonyx hudsonius</i>	Yes	Yes	No Hoarding	Seasonal
<i>Dicrostonyx richardsoni</i>	Yes	Yes	No Hoarding	Seasonal
<i>Dicrostonyx torquatus</i>	Yes	Yes	No Hoarding	Seasonal
<i>Dicrostonyx vinogradovi</i>	Yes	Yes	No Hoarding	Seasonal
<i>Erignathus barbatus</i>	No	No	No Hoarding	Stable
<i>Eschrichtius robustus</i>	No	No	No Hoarding	Stable
<i>Gulo gulo</i>	No	No	Full Both	Stable
<i>Hyperoodon ampullatus</i>	No	No	No Hoarding	Stable
<i>Lagenorhynchus albirostris</i>	No	No	No Hoarding	Stable
<i>Lemmus sibiricus</i>	No	Thicker Sea- sonal	No Hoarding	Stable
<i>Lemmus trimucronatus</i>	No	No	No Hoarding	Stable
<i>Lepus americanus</i>	Yes	Yes	No Hoarding	Stable
<i>Lepus arcticus</i>	Yes	Yes	No Hoarding	Stable
<i>Lepus othus</i>	Yes	Yes	No Hoarding	Stable
<i>Lepus timidus</i>	Yes	Yes	No Hoarding	Stable
<i>Lepus timidus varronis</i>	Yes	Yes	No Hoarding	Stable
<i>Lynx canadensis</i>	Yes	Yes	Full Both	Stable

<i>Lynx lynx</i>	Yes	Yes Thicker Seasonal	Full Both	Stable
<i>Marmota breweri</i>	No	No	No Hoarding	Stable
<i>Marmota camtschatica</i>	No	No	No Hoarding	Stable
<i>Marmota camtschatica doppelmayri</i>	No	No	No Hoarding	Stable
<i>Megaptera novaeangliae</i>	No	No	No Hoarding	Stable
<i>Microtus abbreviatus</i>	No	No	Full Larder	Stable
<i>Lasiopodomys gregalis</i>	No	No	Full Larder	Stable
<i>Microtus hyerboreus</i>	No	No	Full Larder	Stable
<i>Microtus miurus</i>	No	No	Full Larder	Stable
<i>Microtus oeconomus</i>	No	No	Full Larder	Stable
<i>Microtus pennsylvanicus</i>	No	No	Full Larder	Stable
<i>Microtus pennsylvanicus terraenovae</i>	No	No	Full Larder	Stable
<i>Microtus xanthognathus</i>	No	No	Full Larder	Stable
<i>Monodon monoceros</i>	No	No	No Hoarding	Stable
<i>Mustela erminea</i>	Yes	Yes	Full Both	Stable
<i>Mustela nivalis</i>	Yes	Yes	Full Both	Stable
<i>Neovison vison</i>	No	No	No Hoarding	Stable
<i>Myopus schisticolor</i>	No	No	No Hoarding	Stable
<i>Odobenus rosmarus</i>	No	No	No Hoarding	Stable
<i>Odobenus rosmarus laptevi</i>	No	No	No Hoarding	Stable
<i>Odobenus rosmarus divergens</i>	No	No	No Hoarding	Stable
<i>Odobenus rosmarus rosmarus</i>	No	No	No Hoarding	Stable
<i>Ondatra zibethicus</i>	No	No	Full Larder	Stable
<i>Ovibos moschatus</i>	No	No	No Hoarding	Stable
<i>Ovibos moschatus moschatus</i>	No	No	No Hoarding	Stable
<i>Ovibos moschatus wardi</i>	No	No	No Hoarding	Stable
<i>Phoca fasciata</i>	No	No	No Hoarding	Stable
<i>Phoca groenlandica</i>	No	No	No Hoarding	Stable
<i>Phoca larga</i>	No	No	No Hoarding	Stable
<i>Phoca vitulina</i>	No	No	No Hoarding	Stable
<i>Phoca vitulina vitulina</i>	No	No	No Hoarding	Stable
<i>Phoca vitulina concolor</i>	No	No	No Hoarding	Stable
<i>Phoca vitulina stejnegeri</i>	No	No	No Hoarding	Stable
<i>Phoca vitulina richardii</i>	No	No	No Hoarding	Stable
<i>Phocoena phocoena</i>	No	No	No Hoarding	Stable
<i>Phocoena phocoena vomerina</i>	No	No	No Hoarding	Stable
<i>Physeter macrocephalus</i>	No	No	No Hoarding	Stable
<i>Pusa hispida</i>	No	No	No Hoarding	Stable
<i>Pusa hispida hispida</i>	No	No	No Hoarding	Stable
<i>Pusa hispida saimensis</i>	No	No	No Hoarding	Stable

<i>Pusa hispida ladogensis</i>	No	No	No Hoarding	Stable
<i>Pusa hispida botnica</i>	No	No	No Hoarding	Stable
<i>Rangifer tarandus</i>	Yes	Yes	No Hoarding	Stable
<i>Rangifer tarandus granti</i>	Yes	Yes	No Hoarding	Stable
<i>Rangifer tarandus groenlandicus</i>	Yes	Yes	No Hoarding	Stable
<i>Rangifer tarandus pearyi</i>	Yes	Yes	No Hoarding	Stable
<i>Rangifer tarandus fennicus</i>	Yes	Yes	No Hoarding	Stable
<i>Rangifer tarandus tarandus</i>	Yes	Yes	No Hoarding	Stable
<i>Rangifer tarandus caribou</i>	Yes	Yes	No Hoarding	Stable
<i>Rangifer tarandus platyrhincus</i>	Yes	Yes	No Hoarding	Stable
<i>Sorex arcticus</i>	Yes	Yes	Full Larder	Stable
<i>Sorex cinereus</i>	Yes	Yes	Full Larder	Stable
<i>Sorex tundrensis</i>	Yes	Yes	Full Larder	Stable
<i>Sorex ugyunak</i>	Yes	Yes	Full Larder	Stable
<i>Ursus arctos</i>	No	No	Limited Caching	Seasonal
<i>Ursus arctos middendorffi</i>	No	No	Limited Caching	Seasonal
<i>Ursus arctos horribilis</i>	No	No	Limited Caching	Seasonal
<i>Ursus maritimus</i>	Yes	No	Full Both	Stable
<i>Vulpes vulpes</i>	No	No	Full Both	Stable
<i>Vulpes lagopus beringensis</i>	Yes	Yes	Full Both	Stable
<i>Acinonyx jubatus</i>	No	No	No Hoarding	Stable
<i>Ailuropoda melanoleuca qinlingensis</i>	No	No	No Hoarding	Stable
<i>Ailurus fulgens styani</i>	No	No	No Hoarding	Stable
<i>Aonyx capensis</i>	No	No	No Hoarding	Stable
<i>Aonyx cinerea</i>	No	No	No Hoarding	Stable
<i>Arctictis binturong</i>	No	No	No Hoarding	Stable
<i>Arctocephalus australis</i>	No	No	No Hoarding	Stable
<i>Arctocephalus forsteri</i>	No	No	No Hoarding	Stable
<i>Arctocephalus galapagoensis</i>	No	No	No Hoarding	Stable
<i>Arctocephalus gazella</i>	No	No	No Hoarding	Stable
<i>Arctocephalus philippii</i>	No	No	No Hoarding	Stable
<i>Arctocephalus pusillus doriferus</i>	No	No	No Hoarding	Stable
<i>Arctocephalus pusillus pusillus</i>	No	No	No Hoarding	Stable
<i>Arctocephalus townsendi</i>	No	No	No Hoarding	Stable
<i>Arctocephalus tropicalis</i>	No	No	No Hoarding	Stable
<i>Arctogalidia trivirgata</i>	No	No	No Hoarding	Stable
<i>Arctonyx collaris</i>	No	No	No Hoarding	Stable
<i>Atelocynus microtis</i>	No	No	No Hoarding	Stable
<i>Atilax paludinosus</i>	No	No	No Hoarding	Stable
<i>Bassaricyon gabbii gabbii</i>	No	No	No Hoarding	Stable
<i>Bassaricyon medius medius</i>	No	No	No Hoarding	Stable
<i>Bassaricyon medius orinomus</i>	No	No	No Hoarding	Stable

<i>Bassaricyon neblina neblina</i>	No	No	No Hoarding	Stable
<i>Bassariscus astutus</i>	No	No	No Hoarding	Stable
<i>Bassariscus sumichrasti</i>	No	No	No Hoarding	Stable
<i>Bdeogale crassicauda</i>	No	No	No Hoarding	Stable
<i>Bdeogale nigripes</i>	No	No	No Hoarding	Stable
<i>Canis adustus</i>	No	No	Full Scatter	Stable
<i>Canis anthus</i>	No	No	Full Scatter	Stable
<i>Canis aureus</i>	No	No	Full Scatter	Stable
<i>Canis indica</i>	No	No	Full Scatter	Stable
<i>Canis latrans</i>	No	No	Full Scatter	Stable
<i>Canis mesomelas elongae</i>	No	No	Full Scatter	Stable
<i>Canis rufus</i>	No	No	Full Scatter	Stable
<i>Canis simensis</i>	No	No	No Hoarding	Stable
<i>Caracal caracal</i>	No	No	No Hoarding	Stable
<i>Catopuma badia</i>	No	No	No Hoarding	Stable
<i>Catopuma temminckii</i>	No	No	No Hoarding	Stable
<i>Cerdocyon thous</i>	No	No	No Hoarding	Stable
<i>Chrotogale owstoni</i>	No	No	No Hoarding	Stable
<i>Chrysocyon brachyurus</i>	No	No	No Hoarding	Stable
<i>Civettictis civetta</i>	No	No	No Hoarding	Stable
<i>Conepatus chinga</i>	No	No	No Hoarding	Stable
<i>Conepatus leuconotus figginsi</i>	No	No	No Hoarding	Stable
<i>Conepatus leuconotus leuconotus</i>	No	No	No Hoarding	Stable
<i>Conepatus mesoleucus</i>	No	No	No Hoarding	Stable
<i>Conepatus semistriatus</i>	No	No	No Hoarding	Stable
<i>Crocuta crocuta</i>	No	No	Limited Caching	Stable
<i>Crossarchus alexandri</i>	No	No	No Hoarding	Stable
<i>Crossarchus ansorgei nigricolor</i>	No	No	No Hoarding	Stable
<i>Crossarchus obscurus</i>	No	No	No Hoarding	Stable
<i>Crossarchus platycephalus</i>	No	No	No Hoarding	Stable
<i>Cryptoprocta ferox</i>	No	No	No Hoarding	Stable
<i>Cuon alpinus lepturus</i>	No	No	No Hoarding	Stable
<i>Cynictis penicillata</i>	No	No	No Hoarding	Stable
<i>Cynogale bennettii</i>	No	No	No Hoarding	Stable
<i>Diplogale hosei</i>	No	No	No Hoarding	Stable
<i>Eira barbara</i>	No	No	Limited Caching	Stable
<i>Enhydra lutris kenyoni</i>	No	No	No Hoarding	Stable
<i>Enhydra lutris nereis</i>	No	No	No Hoarding	Stable
<i>Eumetopias jubatus</i>	No	No	No Hoarding	Stable
<i>Eupleres goudotii</i>	No	No	Unknown	Stable
<i>Felis catus</i>	No	No	No Hoarding	Stable
<i>Felis chaus</i>	No	No	No Hoarding	Stable

<i>Felis margarita</i>	No	No	No Hoarding	Stable
<i>Felis nigripes</i>	No	No	Full Scatter	Stable
<i>Felis silvestris bieti</i>	No	No	No Hoarding	Stable
<i>Felis silvestris cafra</i>	No	No	No Hoarding	Stable
<i>Felis silvestris lybica</i>	No	No	No Hoarding	Stable
<i>Felis silvestris ornata</i>	No	No	No Hoarding	Stable
<i>Felis silvestris silvestris</i>	No	No	Full Scatter	Stable
<i>Fossa fossana</i>	No	No	No Hoarding	Stable
<i>Galerella nigrata</i>	No	No	No Hoarding	Stable
<i>Galerella pulverulenta</i>	No	No	No Hoarding	Stable
<i>Galerella sanguinea</i>	No	No	Full Larder	Stable
<i>Galictis cuja</i>	No	No	No Hoarding	Stable
<i>Galictis vittata</i>	No	No	No Hoarding	Stable
<i>Galidia elegans dambrensis</i>	No	No	No Hoarding	Stable
<i>Galidia elegans elegans</i>	No	No	No Hoarding	Stable
<i>Galidia elegans occidentalis</i>	No	No	No Hoarding	Stable
<i>Galidictis fasciata</i>	No	No	No Hoarding	Stable
<i>Genetta angolensis</i>	No	No	Full Scatter	Stable
<i>Genetta bourloni</i>	No	No	Full Scatter	Stable
<i>Genetta cristata</i>	No	No	Full Scatter	Stable
<i>Genetta felina</i>	No	No	Full Scatter	Stable
<i>Genetta genetta</i>	No	No	Full Scatter	Stable
<i>Genetta johnstoni</i>	No	No	Full Scatter	Stable
<i>Genetta maculata</i>	No	No	Full Scatter	Stable
<i>Genetta pardina</i>	No	No	Full Scatter	Stable
<i>Genetta poensis</i>	No	No	Full Scatter	Stable
<i>Genetta servalina</i>	No	No	Full Scatter	Stable
<i>Genetta thierryi</i>	No	No	Full Scatter	Stable
<i>Genetta tigrina</i>	No	No	Full Scatter	Stable
<i>Genetta victoriae</i>	No	No	Full Scatter	Stable
<i>Halichoerus grypus grypus</i>	No	No	No Hoarding	Stable
<i>Helarctos malayanus</i>	No	No	No Hoarding	Stable
<i>Helogale hirtula</i>	No	No	No Hoarding	Stable
<i>Helogale parvula</i>	No	No	No Hoarding	Stable
<i>Hemigalus derbyanus</i>	No	No	No Hoarding	Stable
<i>Herpestes auropunctatus</i>	No	No	No Hoarding	Stable
<i>Herpestes brachyurus</i>	No	No	No Hoarding	Stable
<i>Herpestes edwardsii</i>	No	No	No Hoarding	Stable
<i>Herpestes fuscus</i>	No	No	No Hoarding	Stable
<i>Herpestes ichneumon</i>	No	No	No Hoarding	Stable
<i>Herpestes javanicus</i>	No	No	No Hoarding	Stable
<i>Herpestes naso</i>	No	No	No Hoarding	Stable

<i>Herpestes semitorquatus</i>	No	No	No Hoarding	Stable
<i>Herpestes smithii</i>	No	No	No Hoarding	Stable
<i>Herpestes urva</i>	No	No	No Hoarding	Stable
<i>Herpestes vitticollis</i>	No	No	No Hoarding	Stable
<i>Hyaena hyaena</i>	No	No	No Hoarding	Stable
<i>Hydrictis maculicollis</i>	No	No	No Hoarding	Stable
<i>Hydrurga leptonyx</i>	No	No	No Hoarding	Stable
<i>Ichneumia albicauda</i>	No	No	No Hoarding	Stable
<i>Ictonyx libyca</i>	No	No	No Hoarding	Stable
<i>Ictonyx striatus</i>	No	No	Full Larder	Stable
<i>Leopardus colocolo</i>	No	No	No Hoarding	Stable
<i>Leopardus geoffroyi</i>	No	No	No Hoarding	Stable
<i>Leopardus guigna</i>	No	No	No Hoarding	Stable
<i>Leopardus guttulus</i>	No	No	No Hoarding	Stable
<i>Leopardus jacobita</i>	No	No	No Hoarding	Stable
<i>Leopardus pardalis albescens</i>	No	No	No Hoarding	Stable
<i>Leopardus tigrinus</i>	No	No	No Hoarding	Stable
<i>Leopardus wiedii</i>	No	No	No Hoarding	Stable
<i>Leptailurus serval</i>	No	No	No Hoarding	Stable
<i>Leptonychotes weddellii</i>	No	No	No Hoarding	Stable
<i>Liberiictis kuhni</i>	No	No	No Hoarding	Stable
<i>Lobodon carcinophaga</i>	No	No	No Hoarding	Stable
<i>Lontra canadensis</i>	No	No	Temporary Caching	Stable
<i>Lontra felina</i>	No	No	No Hoarding	Stable
<i>Lontra longicaudis</i>	No	No	No Hoarding	Stable
<i>Lontra provocax</i>	No	No	No Hoarding	Stable
<i>Lutra lutra</i>	No	No	Temporary Caching	Stable
<i>Lutra nippon</i>	No	No	Temporary Caching	Stable
<i>Lutra sumatrana</i>	No	No	No Hoarding	Stable
<i>Lutrogale perspicillata</i>	No	No	No Hoarding	Stable
<i>Lycalopex culpaeus</i>	No	No	Full Scatter	Stable
<i>Lycalopex fulvipes</i>	No	No	Full Scatter	Stable
<i>Lycalopex griseus</i>	No	No	Full Scatter	Stable
<i>Lycalopex gymnocercus</i>	No	No	Full Scatter	Stable
<i>Lycalopex sechurae</i>	No	No	Full Scatter	Stable
<i>Lycalopex vetulus</i>	No	No	Full Scatter	Stable
<i>Lycaon pictus</i>	No	No	Limited Scatter	Stable
<i>Lyncodon patagonicus</i>	No	No	No Hoarding	Stable
<i>Lynx pardinus</i>	No	No	Limited Scatter	Stable
<i>Lynx rufus</i>	No	No	Full Scatter	Stable
<i>Macrogalidia musschenbroekii</i>	No	No	No Hoarding	Stable
<i>Martes americana atrata</i>	No	No	Full Larder	Stable

<i>Martes americana caurina</i>	No	No	Full Larder	Stable
<i>Martes americana humboldtensis</i>	No	No	Full Larder	Stable
<i>Martes americana nesophila</i>	No	No	Full Larder	Stable
<i>Martes americana sierrae</i>	No	No	Full Larder	Stable
<i>Martes flavigula</i>	No	No	Full Larder	Stable
<i>Martes foina</i>	No	No	Full Larder	Stable
<i>Martes martes</i>	No	No	Full Larder	Stable
<i>Martes melampus melampus</i>	No	No	Full Larder	Stable
<i>Martes melampus tsuensis</i>	No	No	Full Larder	Stable
<i>Martes pennanti</i>	No	Yes	Full Larder	Stable
<i>Martes zibellina linkouensis</i>	No	Yes	Full Larder	Stable
<i>Meles anakuma</i>	No	No	No Hoarding	Seasonal
<i>Meles canescens</i>	No	No	No Hoarding	Seasonal
<i>Meles leucurus amurensis</i>	No	No	No Hoarding	Seasonal
<i>Meles meles arcalus</i>	No	No	Full Scatter	Seasonal
<i>Meles meles meles</i>	No	No	Full Scatter	Seasonal
<i>Mellivora capensis</i>	No	No	No Hoarding	Stable
<i>Melogale moschata subaurantiaca</i>	No	No	No Hoarding	Stable
<i>Melogale personata</i>	No	No	No Hoarding	Stable
<i>Melursus ursinus</i>	No	No	No Hoarding	Stable
<i>Mephitis macroura</i>	No	No	No Hoarding	Stable
<i>Mephitis mephitis</i>	No	No	No Hoarding	Seasonal
<i>Mirounga angustirostris</i>	No	No	No Hoarding	Stable
<i>Mirounga leonina</i>	No	No	No Hoarding	Stable
<i>Monachus monachus</i>	No	No	No Hoarding	Stable
<i>Mungos mungo</i>	No	No	No Hoarding	Stable
<i>Mungotictis decemlineata</i>	No	No	No Hoarding	Stable
<i>Mustela africana</i>	No	No	Full Both	Stable
<i>Mustela altaica</i>	No	No	No Hoarding	Stable
<i>Mustela eversmannii</i>	No	No	Full Larder	Stable
<i>Mustela felipei</i>	No	No	No Hoarding	Stable
<i>Mustela frenata</i>	No	No	Full Larder	Stable
<i>Mustela itatsi</i>	No	No	Full Both	Stable
<i>Mustela kathiah</i>	No	No	No Hoarding	Stable
<i>Mustela lutreola</i>	No	No	Full Both	Stable
<i>Mustela nigripes</i>	No	No	No Hoarding	Stable
<i>Mustela nudipes</i>	No	No	Unknown	Stable
<i>Mustela putorius furo</i>	No	Thicker Seasonal	No Hoarding	Stable
<i>Mustela sibirica</i>	No	No	Full Both	Stable
<i>Mustela strigidorsa</i>	No	No	No Hoarding	Stable
<i>Mydaus javanensis</i>	No	No	No Hoarding	Stable
<i>Mydaus marchei</i>	No	No	No Hoarding	Stable

<i>Nandinia binotata</i>	No	No	No Hoarding	Stable
<i>Nasua narica</i>	No	No	No Hoarding	Stable
<i>Nasua nasua rufa</i>	No	No	No Hoarding	Stable
<i>Nasuella olivacea</i>	No	No	No Hoarding	Stable
<i>Neofelis diardi</i>	No	No	Limited Caching	Stable
<i>Neofelis nebulosa nebulosa</i>	No	No	Limited Caching	Stable
<i>Neomonachus schauinslandi</i>	No	No	No Hoarding	Stable
<i>Neophoca cinerea</i>	No	No	No Hoarding	Stable
<i>Nyctereutes procyonoides albus</i>	No	No	No Hoarding	Stable
<i>Nyctereutes procyonoides koreensis</i>	No	No	No Hoarding	Stable
<i>Nyctereutes procyonoides procyonoides</i>	No	No	No Hoarding	Stable
<i>Nyctereutes procyonoides viverrinus</i>	No	No	No Hoarding	Stable
<i>Ommatophoca rossii</i>	No	No	No Hoarding	Stable
<i>Osbornictis piscivora</i>	No	No	No Hoarding	Stable
<i>Otaria byronia</i>	No	No	No Hoarding	Stable
<i>Otocolobus manul</i>	No	Yes	No Hoarding	Stable
<i>Otocyon megalotis</i>	No	No	No Hoarding	Stable
<i>Paguma larvata</i>	No	No	No Hoarding	Stable
<i>Panthera leo atrox</i>	No	No	No Hoarding	Stable
<i>Panthera leo bleyenberghi</i>	No	No	No Hoarding	Stable
<i>Panthera leo krugeri</i>	No	No	No Hoarding	Stable
<i>Panthera leo leo</i>	No	No	No Hoarding	Stable
<i>Panthera leo persica</i>	No	No	No Hoarding	Stable
<i>Panthera leo spelaea</i>	No	No	No Hoarding	Stable
<i>Panthera onca mesembrina</i>	No	No	No Hoarding	Stable
<i>Panthera pardus fusca</i>	No	No	Limited Scatter	Stable
<i>Panthera pardus japonensis</i>	No	No	Limited Scatter	Stable
<i>Panthera pardus orientalis</i>	No	No	Limited Scatter	Stable
<i>Panthera pardus saxicolor</i>	No	No	Limited Scatter	Stable
<i>Panthera tigris altaica</i>	No	No	Limited Scatter	Stable
<i>Panthera tigris amoyensis</i>	No	No	Limited Scatter	Stable
<i>Panthera tigris corbetti</i>	No	No	Limited Scatter	Stable
<i>Panthera tigris jacksoni</i>	No	No	Limited Scatter	Stable
<i>Panthera tigris sondaica</i>	No	No	Limited Scatter	Stable
<i>Panthera tigris sumatrae</i>	No	No	Limited Scatter	Stable
<i>Panthera tigris tigris</i>	No	No	Limited Scatter	Stable
<i>Panthera uncia</i>	No	No	No Hoarding	Stable
<i>Paracynictis selousi</i>	No	No	No Hoarding	Stable
<i>Paradoxurus hermaphroditus</i>	No	No	No Hoarding	Stable
<i>Paradoxurus jerdoni</i>	No	No	No Hoarding	Stable

<i>Paradoxurus zeylonensis</i>	No	No	No Hoarding	Stable
<i>Parahyaena brunnea</i>	No	No	Limited Caching	Stable
<i>Pardofelis marmorata</i>	No	No	No Hoarding	Stable
<i>Phoca largha</i>	No	No	No Hoarding	Stable
<i>Phocarcos hookeri</i>	No	No	No Hoarding	Stable
<i>Poecilogale albinucha</i>	No	No	Full Scatter	Stable
<i>Poiana richardsonii</i>	No	No	No Hoarding	Stable
<i>Potos flavus</i>	No	No	No Hoarding	Stable
<i>Prionailurus bengalensis bengalensis</i>	No	No	No Hoarding	Stable
<i>Prionailurus bengalensis chinensis</i>	No	No	No Hoarding	Stable
<i>Prionailurus bengalensis euptilurus</i>	No	No	No Hoarding	Stable
<i>Prionailurus iriomotensis</i>	No	No	No Hoarding	Stable
<i>Prionailurus planiceps</i>	No	No	No Hoarding	Stable
<i>Prionailurus rubiginosus</i>	No	No	No Hoarding	Stable
<i>Prionailurus viverrinus</i>	No	No	Full Scatter	Stable
<i>Prionodon linsang</i>	No	No	No Hoarding	Stable
<i>Prionodon pardicolor</i>	No	No	No Hoarding	Stable
<i>Procyon cancrivorus</i>	No	No	No Hoarding	Stable
<i>Procyon lotor hirtus</i>	No	No	No Hoarding	Seasonal
<i>Procyon lotor lotor</i>	No	No	No Hoarding	Seasonal
<i>Procyon lotor pallidus</i>	No	No	No Hoarding	Seasonal
<i>Procyon minor</i>	No	No	No Hoarding	Seasonal
<i>Profelis aurata</i>	No	No	No Hoarding	Stable
<i>Proteles cristatus</i>	No	No	No Hoarding	Stable
<i>Pteronura brasiliensis</i>	No	No	No Hoarding	Stable
<i>Puma concolor coryi</i>	No	No	No Hoarding	Stable
<i>Puma concolor cougar</i>	No	No	No Hoarding	Stable
<i>Puma yagouaroundi</i>	No	No	No Hoarding	Stable
<i>Pusa caspica</i>	No	No	No Hoarding	Stable
<i>Pusa sibirica</i>	No	No	No Hoarding	Seasonal
<i>Rhynchogale melleri</i>	No	No	No Hoarding	Stable
<i>Salanoia concolor</i>	No	No	No Hoarding	Stable
<i>Speothos venaticus</i>	No	No	Limited Caching	Stable
<i>Spilogale gracilis latifrons</i>	No	No	Full Scatter	Stable
<i>Spilogale putorius</i>	No	No	Full Scatter	Stable
<i>Suricata suricatta</i>	No	No	No Hoarding	Stable
<i>Taxidea taxus jacksoni</i>	No	No	Full Scatter	Stable
<i>Tremarctos ornatus</i>	No	No	No Hoarding	Seasonal
<i>Uncia Uncia</i>	Yes	No	No Hoarding	Stable
<i>Urocyon cinereoargenteus</i>	No	No	Full Both	Stable
<i>Urocyon littoralis catalinae</i>	No	No	No Hoarding	Stable
<i>Urocyon littoralis clementae</i>	No	No	No Hoarding	Stable

<i>Urocyon littoralis dickeyi</i>	No	No	No Hoarding	Stable
<i>Urocyon littoralis littoralis</i>	No	No	No Hoarding	Stable
<i>Urocyon littoralis santacruzae</i>	No	No	No Hoarding	Stable
<i>Urocyon littoralis santarosae</i>	No	No	No Hoarding	Stable
<i>Ursus americanus altifrontalis</i>	No	No	Limited Caching	Seasonal
<i>Ursus americanus americanus</i>	No	No	Limited Caching	Seasonal
<i>Ursus americanus carlottae</i>	No	No	Limited Caching	Seasonal
<i>Ursus americanus cinnamomum</i>	No	No	Limited Caching	Seasonal
<i>Ursus americanus kermodei</i>	No	No	Limited Caching	Seasonal
<i>Ursus americanus vancouveri</i>	No	No	Limited Caching	Seasonal
<i>Ursus thibetanus formosanus</i>	No	No	Limited Caching	Seasonal
<i>Ursus thibetanus japonicus</i>	No	No	Limited Caching	Seasonal
<i>Ursus thibetanus mupinensis</i>	No	No	Limited Caching	Seasonal
<i>Ursus thibetanus thibetanus</i>	No	No	Limited Caching	Seasonal
<i>Ursus thibetanus ussuricus</i>	No	No	Limited Caching	Seasonal
<i>Viverra megaspila</i>	No	No	No Hoarding	Stable
<i>Viverra zibetha</i>	No	No	No Hoarding	Stable
<i>Viverricula indica</i>	No	No	No Hoarding	Stable
<i>Vormela peregusna</i>	No	No	Full Larder	Stable
<i>Vulpes cana</i>	No	No	No Hoarding	Stable
<i>Vulpes chama</i>	No	No	Limited Scatter	Stable
<i>Vulpes corsac</i>	No	No	Limited Scatter	Stable
<i>Vulpes ferrilata</i>	No	No	Unknown	Stable
<i>Vulpes macrotis zinseri</i>	No	No	Limited Scatter	Stable
<i>Vulpes pallida</i>	No	No	Unknown	Stable
<i>Vulpes rueppellii</i>	No	No	Unknown	Stable
<i>Vulpes velox</i>	No	No	Limited Scatter	Stable
<i>Vulpes zerda</i>	No	No	Full Scatter	Stable
<i>Zalophus californianus</i>	No	No	No Hoarding	Stable
<i>Zalophus japonicus</i>	No	No	No Hoarding	Stable
<i>Zalophus wollebaeki</i>	No	No	No Hoarding	Stable
<i>Canis lupus</i>	No	No	Full Scatter	Stable
<i>Microtus pennsylvanicus chihuahuensis</i>	No	No	No Hoarding	Stable
<i>Mustela nivalis heptneri</i>	No	No	Full Both	Stable
<i>Phocoena phocoena relicta</i>	No	No	No Hoarding	Stable
<i>Vulpes vulpes schrenckii</i>	No	No	Full Both	Stable
<i>Vulpes vulpes japonica</i>	No	No	Full Both	Stable
<i>Vulpes vulpes montana</i>	No	No	Full Both	Stable
<i>Balaenoptera acutorostrata dwarf minke whale</i>	No	No	No Hoarding	Stable
<i>Canis lupus laniger</i>	No	No	Full Scatter	Stable

<i>Ursus arctos syriacus</i>	No	No	Limited Caching	Seasonal
<i>Ursus arctos arctos</i>	No	No	Limited Caching	Seasonal
<i>Ursus arctos isabellinus</i>	No	No	Limited Caching	Seasonal