Integrating evolution and genomics to investigate social development in wolf-dog hybrids

INTEGRATING EVOLUTION AND GENOMICS TO INVESTIGATE SOCIAL DEVELOPMENT IN WOLF-DOG HYBRIDS

A Dissertation Presented

By

Xue(Shirley) Li

Submitted to the Faculty of the

University of Massachusetts Morningside Graduate School of Biomedical Sciences

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 5th, 2023

Bioinformatics & Computational Biology

INTEGRATING EVOLUTION AND GENOMICS TO INVESTIGATE SOCIAL DEVELOPMENT IN WOLF-DOG HYBRIDS

A Dissertation Presented

By

Xue Li

This work was undertaken in the Morningside Graduate School of Biomedical Sciences

Program in Bioinformatics and Integrative Biology

Under the mentorship of

Elinor Karlsson, Ph.D., Thesis Advisor

Zhiping Weng, Ph.D., Member of Committee

Elaine Lim, Ph.D., Member of Committee

Michael Lodato, Ph.D., Member of Committee

Emma Teeling, Ph.D., External Member of Committee

Christopher Sassetti, Ph.D., Chair of Committee

Mary Ellen Lane, Ph.D., Dean of the Graduate School of Biomedical Sciences

May 5th, 2023

DEDICATION

To my parents, who loved me endlessly and always supported me,

my husband, Xiaoyi, who has always believed in my potential and been a guiding light in my journey,

and

Elsa and Hulk, my beloved furry companions, who bring me immense joy and happiness each and every day.

ACKNOWLEDGEMENTS

It has been a wonderful experience studying at UMASS Chan Medical School and I am deeply grateful to the individuals who have played a vital role in helping me with my graduate study and research.

First and foremost, I would like to express my deep appreciation to my mentor, Dr. Elinor Karlsson, whose constant support, guidance, and encouragement have been indispensable to my success throughout my graduate school journey. When I first joined her research group, I struggled with self-doubt and a lack of confidence in my abilities. However, Elinor's unwavering belief in me and her willingness to invest her time and energy in my development has been invaluable in improving my self-esteem and becoming a confident scientist. I really enjoy working with her while structuring scientific questions, analyzing data, interpreting and visualizing results, and writing papers and grants. I am truly grateful for her generosity with her time and expertise. She has inspired me to enjoy science and to approach research with enthusiasm and curiosity, and I will always treasure the wonderful experiences and valuable memories I gained under her mentorship.

I want to acknowledge the members of my thesis research advisory and dissertation examination committees, including Christopher Sassetti, Zhiping Weng, Manuel Garber, Athma Pai, Elain Lim, and Michael Ladato, for their invaluable feedback and support throughout my graduate studies. I am especially grateful to Dr. Emma Teeling for joining my thesis defense all the way from Ireland.

I want to convey my heartfelt appreciation to my colleagues for their unwavering support and encouragement throughout my academic journey. Their readiness to share their expertise, give constructive feedback, and provide motivation has been invaluable. I am fortunate to have such incredible colleagues and friends, and I am grateful for their ongoing support and camaraderie. In particular, I would like to express my gratitude to Diane Genereux for her steadfast assistance, guidance, and encouragement in both research and personal matters since joining UMASS eight years ago. Additionally, I would like to thank Kathryn Lord for her invaluable knowledge of animal behaviors, which has been essential to the success of the wolfdog hybrid project.

I owe an immeasurable debt of gratitude to my parents, Li Yanfeng (李岩峰) and Zhang Weixia (张维侠). They have supported me in every possible way throughout my life, from helping me with my schoolwork as a child to cheering me on as I pursued my PhD. Despite the distance between us, they never once questioned my decision to study abroad. My parents never had the opportunity to attend college themselves, but they instilled in me a deep appreciation for learning. They worked tirelessly to provide for our family, and their sacrifices have made it possible for me to pursue my dreams. From funding my education to providing emotional support during difficult times, their selflessness has been a constant reminder of the power of love and determination. As the first PhD in my family, I am keenly aware of the role that my parents' support has played in my success. I am deeply grateful for everything they have done for me, and I will continue to work hard to make them proud. Thank you, Mom and Dad, for your steadfast belief in me. I love you both more than words can express.

Throughout my academic journey, I have been fortunate to have the support of my husband, Xiaoyi Wang (王潇毅). His understanding, encouragement, belief in my abilities, and commitment to my success have been crucial in helping me work persistently and make progress along the way. He has been there for me every step of the way and has never once doubted my potential. I am grateful for the amazing adventures we've shared and the countless memories we've made while exploring the world together. Thank you, Xiaoyi, for being my partner in life and being a guiding light of my self-cultivation. I could not have done this without you.

Lastly, I would like to express my gratitude to my two furry companions, Elsa and Hulk. They have helped me become a better person, more organized and able to balance life and research. Their wagging tails, affectionate licks, and boundless energy have been a source of joy and happiness, reminding me of the importance of taking breaks and enjoying life's simple pleasures.

ABSTRACT

Domesticated dogs separated from wolves around ~5000-7000 generations ago, with major differences in early social development that have enabled them to survive, and thrive, in close proximity to humans. Due to their unique evolutionary history and accessibility, canines serve as a natural model system to study the genetic factors underlying behavior adaptation within and between subspecies. The wolf/dog system can not only advance the understanding of evolutionary processes, but also help to better understand the neurogenetic pathways involved in human psychiatric disorders. Although wolves and dogs split relatively recently in evolutionary time, they are genetically distinct populations with numerous differences across their genome. This population structure makes it impossible to confidently associate particular genomic variants with domestication-related traits by simply comparing dogs and wolves.

In this dissertation, I identified genes and pathways associated with domestication-related behavior using an unusual admixed population of wolf-dog hybrids housed in sanctuaries across the United States. I developed methods and approaches to map behavioral phenotypes in wolf-dog hybrids, and explored the overlap with dog social behaviors, and human psychiatric conditions. I first characterized the population history of wolf-dog hybrids using techniques including exploratory principal component analysis, ancestry calling, and population differentiation test. I defined the behavioral phenotypes by dimensional reduction analysis of coded video data, and identified associations

between genes and regulatory elements with those phenotypes using admixture mapping and association test. Finally, I investigated the functional and biological mechanisms underlying the associated regions by gene-set analysis. I discovered that regions associated with domestication-related behavioral differences are enriched for brain expressed genes, especially those enriched in early infancy.

To further investigate the candidate regions associated with canine domestication, I leveraged a powerful new data resource comparing the genomes of 240 mammalian species. Using data from massively parallel reporter assay experiments in human cells, I confirmed that this resource can distinguish which bases have regulatory function. Overall, variants in highly constrained positions are more likely to alter cellular function. In addition, I showed that dogs with ancestry from a single breed, which have shorter lifespans than outbred dogs, are also more likely to carry variants in constrained positions, suggesting they impact fitness. In the wolf-dog hybrids, I cataloged candidate causal variants that differed between dogs and wolves and were highly constrained across mammals.

Overall, this thesis demonstrates how new genomic tools and data resources can be leveraged to investigate exceptional evolutionary adaptations in other species that may offer insight into human diseases. By utilizing the wolf-dog hybrid population, we can re-trace the ancient genetic changes of domestication that led to divergence of canine social and developmental

behaviors, and potentially uncover genetic pathways that contribute to social behavioral disorders such as autism spectrum disorders.

PREFACE

Parts of this dissertation appear in the publication listed below:

Xue Li, K. Morril, B.Logan, D. Genereux, G. Chauhan, H. J. Noh, K. Megquier, R. Swofford, J. Turner-Maier, M. Koltookian, … …,K. A. Lord, E. K. Karlsson, Investigating genetics of social behaviors in wolfdog hybrids. (In submission)

K. L. Moon, H. J. Huson, K. Morrill, M. Wang, **Xue Li**, K. Srikanth, Zoonomia Consortium, G. J. Svenson, E. K. Karlsson, B. Shapiro, Comparative genomics of Balto, a famous historic dog, captures lost diversity of 1920s sled dogs. *Science*. **380**, eabn5887 (2023).

P. F. Sullivan, J. R. S. Meadows, S. Gazal, B. N. Phan, **Xue Li**, D. P. Genereux, M. X. Dong, M. Bianchi, G. Andrews, S. Sakthikumar, J. Nordin, A. Roy, M. J. Christmas, V. D. Marinescu, C. Wang, O. Wallerman, J. Xue, S. Yao, Q. Sun, J. Szatkiewicz, J. Wen, L. M. Huckins, A. Lawler, K. C. Keough, Z. Zheng, J. Zeng, N. R. Wray, Y. Li, J. Johnson, J. Chen, Zoonomia Consortium, B. Paten, S. K. Reilly, G. M. Hughes, Z. Weng, K. S. Pollard, A. R. Pfenning, K. Forsberg-Nilsson, E. K. Karlsson, K. Lindblad-Toh, Leveraging Base Pair Mammalian Constraint to Understand Genetic Variation and Human Disease. *Science*. **380**, eabn2937 (2023).

Other publications from my graduate study that are not included in the thesis:

K. Morrill, J. Hekman, **Xue Li,** J. McClure, B. Logan, L. Goodman, M. Gao, Y. Dong, M. Alonso, E. Carmichael, N. Snyder-Mackler, J. Alonso, H. J. Noh, J. Johnson, M. Koltookian, C. Lieu, K. Megquier, R. Swofford, J. Turner-Maier, M. E. White, Z. Weng, A. Colubri, D. P. Genereux, K. A. Lord, E. K. Karlsson, Ancestry-inclusive dog genomics challenges popular breed stereotypes. *Science*. **376**, eabk0639 (2022).

K. Megquier, J. Turner-Maier, K. Morrill, **Xue Li**, J. Johnson, E. K. Karlsson, C. A. London, H. L. Gardner, The genomic landscape of canine osteosarcoma cell lines reveals conserved structural complexity and pathway alterations. *PLoS One*. **17**, e0274383 (2022).

K. Megquier, D. P. Genereux, J. Hekman, R. Swofford, J. Turner-Maier, J. Johnson, J. Alonso, **Xue Li**, K. Morrill, L. J. Anguish, M. Koltookian, B. Logan, C. R. Sharp, L. Ferrer, K. Lindblad-Toh, V. N. Meyers-Wallen, A. Hoffman, E. K. Karlsson, BarkBase: Epigenomic Annotation of Canine Genomes. *Genes* . **10** (2019), doi:10.3390/genes10060433.

J. R. Xue, A. Mackay-Smith, K. Mouri, M. F. Garcia, M. X. Dong, J. F. Akers, M. Noble, **Xue Li**, Zoonomia Consortium, K. Lindblad-Toh, E. K. Karlsson, J. P. Noonan, T. D. Capellini, K. J Brennand, R. Tewhey, P. C. Sabeti, S. K. Reilly. The functional and evolutionary impacts of human-specific deletions in conserved elements. *Science*. **380**, eabn2253 (2023).

Table of contents

List of Figures

List of Tables

Table 4.11 Contingency table for Fisher's exact test for gene overlap

List of Symbols, Abbreviations, or Nomenclature

Basic Genetics	
genetic variant	a variation in the DNA sequence of an individual, which may be small, such as a single nucleotide polymorphism (SNP), or larger, such as a deletion or insertion of multiple nucleotides
allele	one of multiple versions of a genetic variant
genotype	maternal and paternal alleles carried by an individual
SNP	single-nucleotide polymorphism, a type of genetic variant
indel	insertion or deletion, a type of genetic variant
phenotype	an observable expression of a trait or characteristic
neutral variant	a genetic variant having no effect on a phenotype of interest
causal variant	a genetic variant having direct effects on a phenotype of interest
Population Genetics	

linkage disequilibrium	non-random association of alleles at different but nearby genomic locations
genetic selection	non-random changes in allele frequency in a population over generations
genetic differentiation	differences in allele frequency between populations that may be the result of genetic drift or genetic selection
population structure	non-random patterns of genetic variation within and between populations

Selection, Evolution, and Comparative Genomics

artificial selection	a process of natural selection mediated by human decisions to breed for or against particular phenotypes in other species
negative selection	a process of natural selection that results in removal of disadvantageous genetic variants (a.k.a. purifying selection)
evolutionary constraint	observed resistance to mutation of DNA sequences aligned across species suggesting negative (purifying) selection
evolutionary conservation	observed similarity of DNA sequences aligned across species
homologs	within or across species: genes or genetic elements generated from a common ancestral source DNA sequence, and which may or may not perform the same function

Admixture, Dog breeding	
lineage	ancestral relationships traced and supported across generations; in dogs, lineage refers to common descent attributed to a type of dog (e.g. terriers)
dog breed	any subpopulation of dogs defined by genetic similarity and shared lineage
modern breed	a subpopulation of dogs intentionally bred and genetically isolated according to a stringent registration, lineage, closed studbook, or standard
global ancestry	genetic composition of an individual attributed to its ancestors, which may comprise admixture from several lineages (e.g. a dog may derive half its ancestry from a specific spaniel line, a quarter from bull-and-terrier dogs, and a quarter from village dogs in Costa Rica)
breed ancestry	ancestry content derived from specific breed populations (e.g. a dog who is 50% golden retriever and 50% poodle)
wolf ancestry	ancestry content derived from wolves (e.g. dog have 0% wolf ancestry, wolfdog hybrids can have wolf ancestry anywhere between 0% and 100%)

local ancestry	the genetic ancestry of an individual at a particular chromosomal location, where an individual can have 0, 1 or 2 copies of an allele derived from each ancestral population
admixture	the process by which individuals from different populations mate and produce offspring, resulting in exchange of genetic material from those populations via random assortment and crossover
introgression	the transfer of genetic material between species following hybridization and backcrossing to the parental species (we also use this term to describe the transfer of genetic material between wolves and dogs despite they belong to the same species)

Canine population

purebred dog	a dog having full-bred ancestry from a single breed, often intentionally bred though not necessarily registered
mixed-breed dog	a dog having mixed breed ancestry, intentional or unintentional
non-breed dog	a dog lacking in recent or any breed ancestry, e.g. village dogs
indigenous dog	a non-breed dog with genetic signatures derived from ancient dogs of a given geographic region
village dog	a non-breed dog free-living and free-breeding near people

grey wolf	a large canine native to Eurasia and North America
Eurasian wolf	subspecies of grey wolf native to Europe and Asia
North American wolf	subspecies of grey wolf native to North American
Coyote	a species of canine native to North America
Wolfdog hybrid	an animal that is part wolf and part domestic dog
Wolfdog breed	Czechoslovakian wolfdog or Saarloos wolddog
Captive wolf	grey wolves that live in captivity
Captive wolfdog	wolfdogs that live in captivity
Coyote-dog	an animal that is part coyote and part domestic dog

Genomic Analysis

WGS	whole genome sequencing
MAF	minor allele frequency
GWAS	genome-wide association study, a series of statistical tests for association across independent genome-wide markers from a set of unrelated individuals stratified by a phenotype of interest

Manhattan plot	a plot of test statistics or *p*-values across genomic coordinates
Admixture mapping	a genetic research method used to identify genetic variants associated with diseases or traits in populations that have a history of recent genetic admixture
QADM	a chi-squared with 1 degree of freedom generated by MIXSCORE to perform local ancestry admixture association
QSNP1	a chi-squared with 1 degree of freedom generated by MIXSCORE to SNP association condition on local ancestry status
QSUM	the sum of QADM and QSNP1, a chi-squared score with 2 degree of freedom
Population differentiation test	a statistical method used in population genetics to assess the degree of genetic differentiation between two or more populations
ANOVA	a statistical method used to analyze the differences between the means of three or more groups
PCA	a dimensionality reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set

CHAPTER I. INTRODUCTION

I.1 Wolves and dogs

Evolution of dogs from wolves

Dogs (*Canis familiaris*) were domesticated from wolves (*Canis lupus*) around 15,000 years ago (Vilà et al. 1997; Lindblad-Toh et al. 2005). However, the exact location and process of this evolution from wolves is not fully understood. Fossil records indicate that domestication may have occurred independently in various regions of the world, including East Asia and the Middle East (Vilà et al. 1997; Savolainen et al. 2002; Larson et al. 2012; Ding et al. 2012). Domestication likely occurred when a small subpopulation of ancient wolves began to scavenge on human garbage and follow human groups, which increased their chances of survival due to food availability. Additionally, humans began to select “tamer” wolves that exhibit more docile or mild behavior and a shorter flight distance and train these wolves for various tasks such as hunting, guarding, and tracking. These two actions by humans and wolves together gradually shifted the small population of wolves from wild to domesticated animals (Lindblad-Toh et al. 2005).

Morphological change

Since domestication, many morphological changes have occurred. Wolves have long and lean bodies, tall and strong legs for running far distances, sharp strong teeth to grab, hold, and dissect prey, large brains for processing complex tasks

and, when they live in cold climates, dense fur. As all wolves experience very similar selective pressures, they are visually more homogeneous than dogs. Despite being descended from wolves, domesticated dog breeds exhibit considerable variability in physical characteristics. Some characteristics, like floppy ears, a small jaw, short and abruptly-stopped muzzle, curled tails, and white patches on the body, have evolved since domestication (Lord et al. 2020). Other characteristics, like smaller body size, extra-long hair, and curly fur are the result of very recent artificial selection.

Domestication-related social behaviors

To survive in the anthropogenic niche, domesticated dogs have undergone many behavioral changes. In the wild, wolves have essential behaviors such as collaborative hunting, and a high threshold of novelty avoidance to protect themselves from danger (MacNulty et al. 2007). Domesticated dogs, which have evolved to live with humans, show reduced flight distance and the ability to quickly form interspecies social bonds (Moretti et al. 2015; Zapata et al. 2016). A study comparing wolves and dogs has shown that dogs exhibit reduced fear of novelty characterized by reduced latency of approach and reduced time spent exploring novel objects (Moretti et al. 2015).

Existing approaches to study social behavior differences

To study the social and behavioral differences between wolves and dogs, researchers have adopted both phenotypic and genetic approaches. Specific behaviors such as interspecies social bonding, problem-solving abilities, and response to novelty have been evaluated using various techniques (Persson et al. 2016; Burkhard et al. 2022; Hansen Wheat et al. 2022; Pörtl and Jung 2017). To characterize the complex behavior phenotypes, researchers utilize ethological coding, a method in the biological study of animal behavior, involving the systematic measurement of observable behaviors of animals (Hansen Wheat et al. 2022). Typically, an ethogram is generated by cataloging discrete, objective behaviors from videos recorded over a limited time period. Behaviors are then converted to quantitative phenotypes by recording the frequency, latency, and/or duration of a specific behavior. To minimize the possibility of bias assessment, multiple individuals are required to establish inter-rater reliability in the coding process. In addition to behavior characterization, recent advances in sequencing techniques have enabled researchers to pair the behavior data with genomic data of wolves and dogs at the population level. Selection scans on dogs have uncovered genes that appear to be under selection in dogs, including genes preferentially expressed in the brain (Axelsson et al. 2013). A comparison of the genomes of village dogs and wolves has highlighted the role of the neural crest in dog domestication (Pendleton et al. 2018), and research on gene expression

patterns in dogs, wolves, and coyotes has identified rapid changes in brain gene expression involved in adaptation (Saetre et al. 2004).

I.2 Primary critical period of socialization in canines

The primary critical period of socialization is crucial in understanding the behavioral differences between wolves and dogs. Both genetic and environmental factors contribute to the distinct social behaviors observed in adult wolves and dogs. The critical period, however, represents a key developmental window early in life when rapidly developing brains are highly sensitive to both environmental and genetic influences, and therefore drastically influence adult social behavior later in life (Scott and Marston 1950). Our previous studies have shown that wolves and dogs have different timing of this period (Figure 1.1), resulting in differences in processing early sensory experiences, even when raised in the same environment (Lord 2013).

Both wolves and dogs have a four week window of critical period which initiates when they start walking. Since wolves begin walking two weeks earlier than dogs, their critical period starts two weeks ahead of dogs. For dogs, the critical period begins at four weeks, and they do not hesitate to explore new things. Over the following four weeks, they become increasingly fearful of new things until, at eight weeks old, they will no longer explore novelty, and the critical period ends. However, observational data suggest that in wolves, they begin to explore the world at two weeks of age, and this period closes at six weeks of age with

avoidance of novelty (Lord 2013). This alteration in timing completely changes their early experiences. They share the same developmental timing for their sensory systems. While dogs' critical period starts once all senses are functional, wolves are still functionally blind and deaf at the beginning of their critical period. Thus, dogs and wolves have very different experiences during this period, even when raised in identical environments. Furthermore, while dogs can collect diverse sensory information during this period, wolves primarily receive olfactory information. Wolves do not hear the world around them until one week into their critical period, and they do not see the world until halfway through the critical period.

This difference in sensory input during the critical period of socialization is interpreted to explain why dogs more easily form interspecies social bonds, are better at integrating stimuli from multiple sensory channels, and are less sensitive to novelty (Lord 2013).

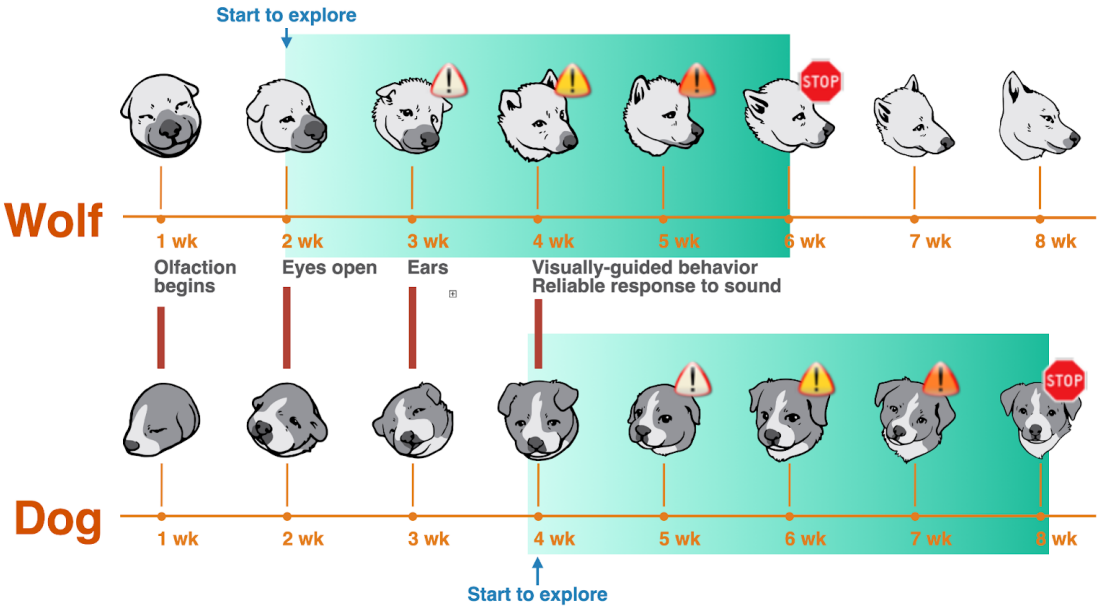


Figure 1.1 Primary critical period of socialization. The four-week period is highlighted in green for both wolves and dogs. The developing timing of sensory systems are the same for wolves and dogs. Wolves start to walk two weeks earlier than dogs, which leads to the critical period starting two weeks earlier than dogs.

I.3 Canines as a natural model system for neurodevelopmental disorders

Domesticated dogs share their environment and food with humans and sometimes experience similar diseases such as cancer, aging-related issues, and psychiatric disorders (Megquier et al. 2022; Karlsson and Lindblad-Toh 2008; Kaeberlein et al. 2016). Many medications have proven effective in both dogs and humans – including many psychotropic drugs applied by veterinary behaviorists for canine behavioral disorders – making dogs a natural model

system for various human neuropsychiatric conditions (Starkey et al. 2005; Pageat 2005).

For instance, dogs are an excellent model for studying obsessive-compulsive disorder (OCD) (Tang et al. 2014; Noh et al. 2017). Research by Tang et al. identified potential genetic risk factors and candidate genes and pathways associated with OCD using GWAS and targeted sequencing in high risk dog breeds (Tang et al. 2014). Noh et al. employed a multispecies approach, integrating evolutionary and regulatory information across human, mouse, and dog genomes to identify candidate genes and pathways associated with OCD (Noh et al. 2017). This research demonstrates the utility of the dog model while also highlighting the potential of comparative genomics across species to better comprehend complex psychiatric disorders such as OCD.

Companion dogs have also been proposed as a translational model for autism spectrum disorder (ASD), a neurological and developmental disorder that affects how people interact with others, communicate, learn, and behave (Topál et al. 2019). When dog puppies were isolated in a confined area from 2 to 14 weeks of age with no socialization, extremely altered social behavioral phenotypes were observed, including the tendency to withdraw reactions from humans and the inability to establish interspecies bonds (Freedman et al. 1961). However, generating these model phenotypes through intentional or unintentional social deprivation of pups during critical periods raises serious ethical concerns, and

therefore this points to a need for comparative studies in which differences in early experiences affecting development of social behavior occur naturally and without welfare-reducing interventions.

In my dissertation, I propose utilizing wolfdog hybrids as a natural model to study the neurogenetic pathways of psychiatric conditions, particularly those with early onset, where early brain development and life experiences play a significant role, such as ASD (Topál et al. 2019). ASD is believed to result from dysregulated early brain development during a period when the brain is highly plastic and responsive to environmental input ((Klin et al. 2020; Shultz et al. 2018). The critical period hypothesis posits that an imbalance between synaptic excitation and inhibition during early brain development leads to altered neuronal connectivity and abnormal behaviors in individuals with ASD ((LeBlanc and Fagiolini 2011). Social experiences and interactions shape the developing brain during this period, and disruptions can lead to social and communication deficits characteristic of ASD. Despite numerous genes being found to be associated with ASD, each accounts for only a small amount of variance (Robinson et al. 2016; Baranova et al. 2021; Quesnel-Vallières et al. 2019). Many studies use mice as model systems but mice have different brain structures and social behaviors compared to humans, and the range of behaviors in human ASD, from mild to severe, cannot be reflected in a mouse model ((Kazdoba et al. 2016).

Wolfdog hybrids provide the benefits to circumvent those problems. This model is ideal due to the similarity in social behaviors between the humans and canines, as well as the ability to conduct non-invasive research. Furthermore, wolves and dogs have distinct primary critical periods of socialization, offering naturally different early experiences for both groups. Hybrids, exhibiting social behaviors distinct but derived from both dogs and wolves, are speculated to possess a varying critical period that commences anytime after 2 weeks and concludes prior to 8 weeks. This results in a range of early experiences among hybrids, which could serve as a means to connect not just genotype to behavioral characteristics, but also to establish a connection from genotype, via early development, to modifications in adult behavior.

I.4 Admixture between wolf and dogs

In nature, there is evidence of admixture between wolves and domestic dogs in several different populations, such as Italian wolves, Mexican wolves, and Ethiopian wolves (Vila and Wayne 1999; Schweizer et al. 2018; Randi and Lucchini 2002; Pilot et al. 2019). A recent study investigated the genetic relationship between wolves, dogs, and other wolf-like canids from around the world and found evidence of such admixture (Gopalakrishnan et al. 2018). Intentional breeding of wolves and dogs dates back to the 1950s when breeders started experimenting with crossing wolves and domestic dogs. Initially, the goal

was to create a breed that would combine the physical traits of wolves with the docile temperament of dogs. This resulted in breeds such as the Czechoslovakian Wolfdog (83-85% German Shepherd Dog and 15-17% wolf) and Saarloos wolfdog (66-77% German Shepherd Dog and 23-34% wolf) (Pilot et al. 2019; Moravčíková et al. 2021; Sommese et al. 2021).

Captive wolfdog hybrid population in this study

Wolfdog hybrids we use in this study are distinct from the natural hybrids rarely produced in the wild and from the wolfdog breeds intentionally produced in the 1950s (Figure 1.2). The majority of these domestically bred wolfdogs can trace their lineage back to 1950s fur farms, around the same time as the Czechoslovakian Wolfdog and Saarloos Wolfdog emerged in Europe (Muller 2021). The wolfdog hybrids used in this study reside in North American sanctuaries. The wolfdog hybrids found in sanctuaries are speculated to have a different ancestry compared to the two widely recognized wolfdog breeds that have German Shepherd Dogs as their sole dog breed ancestors. Although no documentation is available regarding the breeding process, these sanctuary hybrids are usually a mixture of German Shepherd, Alaskan Malamute, Siberian Husky, and occasionally Chow Chows and Akitas in their lineage. Their wolf heritage may originate from a population that has been bred in captivity for an extended period.

Wolfdog hybrids, inheriting a blend of two distinct behavior patterns from their ancestors, may have shown any degree of social behavior from wolf end of the spectrum to the dog end. Due to this unique variability, they serve as an exceptional population for mapping genomic regions associated with behavioral differences. Conventional association tests like GWAS are not effective when comparing wolves to dogs due to ubiquitous genomic differences. Wolfdog hybrids escape these limitations. The hybrids possess mosaic genomes and a documented admixture history, enabling more precise mapping of genomic regions associated with behavioral differences through admixture mapping combined with GWAS. Moreover, since these hybrids reside in sanctuaries throughout the US, controlled environments facilitate accurate phenotyping of all subjects through practical behavioral tests. We propose captive wolfdog hybrids as an ideal model for detecting the genomic changes responsible for the differences in early development and social behavior between wolves and dogs.

Figure 1.2 Captive wolfdog hybrids. Two wolfdog hybrids looking at each other during a typical day at sanctuary. Photo courtesy of Kathryn Lord and Brittney Kenney.

I.5 Evolution and comparative genomics

Comparative genomics examines unusual patterns of variation, elements that are changing, usually slow or quick, across species' genomes to identify potential functional significance. This concept predates the sequencing of the completion of the Human Genome Project in 2003. Early comparative genomic studies focused on organisms with small genome size. A study of yeast, worms, and flies revealed common proteins and similar non redundant protein sets (Rubin et al. 2000). Further research led to the significant revision of the yeast gene catalog

and the discovery of new regulatory motifs (Kellis et al. 2003). As sequencing and genome assembly became more cost-effective, the focus shifted from microorganisms and invertebrates to more complex organisms with larger genomes and intricate biological functions. In 2011, the 29 Mammals Project sequenced 29 mammals, assessed sequence conservation among them, and identified constrained elements at 12 base pair resolution (Lindblad-Toh et al. 2011). These regions proved to be more enriched for the heritability of complex diseases than any other functional annotation. More recently, the Zoonomia project, an international collaboration between many research groups including the Karlsson Lab, generated the most extensive mammalian comparative genomics resource to date by aligning the whole genomes of 240 species and the protein-coding regions of 428 vertebrates (Zoonomia Consortium 2020). Our latest research, published in Science (Apr 26), uses comparative genomics to investigate genetic variants underlying species-level phenotypes and leverages evolutionary constraints to expedite the search for genetic changes associated with complex human diseases.

CHAPTER II. LEVERAGE EVOLUTION AND COMPARATIVE GENOMICS TO CONNECT VARIATION TO FUNCTION

II.1 Preface

The findings discussed in this chapter have been adapted from two significant studies:

K. L. Moon, H. J. Huson, K. Morrill, M. Wang, **Xue Li**, K. Srikanth, Zoonomia Consortium, G. J. Svenson, E. K. Karlsson, B. Shapiro, Comparative genomics of Balto, a famous historic dog, captures lost diversity of 1920s sled dogs. *Science*. **380**, eabn5887 (2023).

P. F. Sullivan, J. R. S. Meadows, S. Gazal, B. N. Phan, **Xue Li**, D. P. Genereux, M. X. Dong, M. Bianchi, G. Andrews, S. Sakthikumar, J. Nordin, A. Roy, M. J. Christmas, V. D. Marinescu, C. Wang, O. Wallerman, J. Xue, S. Yao, Q. Sun, J. Szatkiewicz, J. Wen, L. M. Huckins, A. Lawler, K. C. Keough, Z. Zheng, J. Zeng, N. R. Wray, Y. Li, J. Johnson, J. Chen, Zoonomia Consortium, B. Paten, S. K. Reilly, G. M. Hughes, Z. Weng, K. S. Pollard, A. R. Pfenning, K. Forsberg-Nilsson, E. K. Karlsson, K. Lindblad-Toh, Leveraging Base Pair Mammalian Constraint to Understand Genetic Variation and Human Disease. *Science*. **380**, eabn2937 (2023).

The Zoonomia Consortium is credited with the creation of the evolutionary constraint score, phyloP. The sequencing of Balto's genome was carried out by Heather Huson group from Cornell University and Beth Shapiro Group from the University of California, Santa Cruz. Following these significant contributions, I carried out variant calling using the raw reads generated from high-coverage DNA sequencing and performed the analysis described in this chapter.

II.2 Abstract

In Chapter II, we leveraged the power of comparative genomics through two case studies, emphasizing its potential in identifying disease-related genetic variations and understanding their biological mechanisms.

We utilized human-centric phyloP score to anticipate the regulatory functional significance of human genomic variants, employing data from massively parallel reporter assays (MPRA). Utilizing MPRA data, we discerned that phyloP scores can differentiate between sequence backgrounds with and without regulatory activity. Notably, phyloP scores highlighted variants demonstrating allele-specific effects. We further explored the predictive power of the phyloP score on variant effect, comparing it to variant effects evaluated by saturation mutagenesis MPRA. The results revealed the phyloP score as a potent predictor for regulatory effects of mutations in regulatory elements.

Our investigation into Balto, the lone sequenced individual from its population, showcases how comparative genomics can amplify our ability to study genotype-phenotype associations using limited cross-population datasets. Merging genomic data from various sources, we use dog-centric phyloP scores and SNP effect annotations to devise an analytical strategy. This enables us to infer genetic and phenotypic traits of Balto and its population, demonstrating the essential role of comparative genomics, particularly when dealing with limited information about the studied population.

II.3 Background

II.3.1 Efforts made to characterize genomic variation in complex traits

The genome-wide association study (GWAS) is a widely adopted bioinformatics approach to study complex traits and identify genomic variants that are statistically linked to disease risk or particular traits. More than 5,700 GWAS have now been conducted for more than 3,300 traits (Uffelmann et al. 2021). However, due to linkage disequilibrium, these associations often span regions with multiple variants, making it difficult to pinpoint the specific causal variant among the linked variants. To tackle this problem, considerable efforts have been dedicated to annotating human genetic variations at both coding and non-coding regions. These efforts include ENCODE and GTEx which generate large amounts of epigenomic data at both tissue and single-cell level and gnomAD and TOPMed which infer deleterious effects from allele frequencies and locations in coding sequences. Although functional validation of identified loci is essential to verify the functional relevance of the variants, this has significant limitations in terms of cost and time. Fine-mapping methods such as PAINTOR (Gong et al. 2019) and PolyFUN (Weissbrod et al. 2020) have been developed to address this issue, but they all require specific inputs. These inputs are often disease-dependent or cell type-dependent, and thus are sometimes inaccessible, which poses challenges for their use in research. The incorporation of evolutionary constraints through

phyloP scores, which are independent of cell type and disease type, offer strong predictive power to help with fine-mapping (Fong and Capra 2021; Pankratov et al. 2022).

II.3.2 Zoonomia project: whole-genome alignment of 240 species and resolve constraint to single nucleotides.

The Zoonomia Comparative project offers unprecedented insight into evolutionary constraint, and has been instrumental in studying mammalian evolution and identifying functional variants involved in human diseases. It is the most comprehensive resource of its kind with whole genomes aligned for 240 species, a significant improvement over the previous alignment of 100 vertebrates, with protein-coding sequences aligned for 428 species (Zoonomia Consortium 2020). The project used a reference-free, graph-based approach (Armstrong et al. 2020) to align sequences missing from the human genome.

The project has yielded valuable insights into the genetic basis of human diseases by using evolutionary constraint as a tool to identify functional variants. Furthermore, it has addressed various questions about mammalian evolution, including the annotation of regions of exceptional constraint and acceleration in the human genome, as well as the investigation of the origins of mammalian traits.

These findings have been critical in advancing our understanding of mammalian evolution and the genetic underpinnings of human diseases. In summary, the

Zoonomia Comparative project has made significant contributions to the field of evolutionary biology and has provided a valuable resource for researchers interested in understanding the genetic basis of human diseases.

II.3.3 Conservation constraints as a tool to study past individuals without data from their contemporaries.

In recent years, advancements in the extraction of ancient DNA have made it possible to reconstruct high-coverage nuclear genomes from fossils and historical specimens (Perri et al. 2021; Thomson et al. 2014). However, interpreting genetic information from individuals who lived in the past can be difficult without access to data from their contemporaries. By combining current-day population-level genomic data with trait association catalogs, researchers can deduce the genetic and phenotypic characteristics of deceased individuals and the populations to which they belonged. The Zoonomia resource has emerged as a possible comparative tool to aid in the interpretation of paleogenomic data, encompassing both coding and rapidly evolving regulatory elements.

II.4 Results

II.4.1 Conservation constraints predict variant regulatory effect at single base resolution

Massively Parallel Reporter Assays (MPRA) (Tewhey et al. 2016) have been employed to rapidly examine thousands of genomic variants for their possible regulatory impacts on gene expression. While the functional results from these high-throughput techniques are beneficial for pinpointing potential causal alleles, incorporating constraint scores can provide additional insights into the functional variants. To explore this, we integrated our Zoonomia phyloP scores with three MPRA studies: (1) a 3'UTR MPRA (Griesemer et al. 2021) assaying 12,173 3' untranslated region (3'UTR) variants with evidence of associating with genome-wide association studies (GWAS) and human evolutionary adaptation, (2) an eQTL MPRA (Tewhey et al. 2016) assaying 32,373 variants from 3,642 cis-expression quantitative trait loci (eQTL) and control regions, and (3) a saturation mutagenesis MPRA that examines 20 disease-associated gene promoters and enhancers (Kircher et al. 2019).

Conservation pattern for regulatory elements at single base pair resolution

We generated three metrics to measure the conservation level of the testing elements or variants in each group: (1) Percentage of constraint sites: the percentage of testing elements or variants in each group having phyloP score

above FDR 0.05 cutoff: 2.27 (2) The maximum phyloP score in the testing oligo; (3) The phyloP score at the variant position centered in the oligo.

Utilizing the 3'UTR MPRA data and the metric `percentage of constraint sites`, we discovered that phyloP scores were capable of distinguishing between sequence backgrounds with and without regulatory activity (Figure 2.1A; Neutral vs. Active: P=3.3E-5). Moreover, phyloP scores emphasized the variants exhibiting allele-specific effects (Figure 2.1A; Neutral vs. Skew: 3.9E-5). The same pattern is observed in the eQTL MPRA data (figure 2.1D). Both the metrics 'maximum phyloP score in oligo' and 'phyloP score' at the variant position can distinguish between neutral and active oligos, and between neutral and skewed variants, except for 'phyloP score' in the case of eQTL (Figure 2.1B,C,E,F). A potential reason for this is that eQTLs, which are genomic regions that influence the expression levels of one or more genes, may not be as conserved as other functional elements like promoters or 3'UTRs. For eQTLs, 'maximum phyloP score in oligo' serves as a more effective metric for predicting functional importance, as it can determine whether the eQTL is situated within a more conserved element or not (Figure 2.1E). For 3'UTR MPRA, we also investigated this pattern using data from individual cell types and a similar pattern was observed (Figure 2.2).

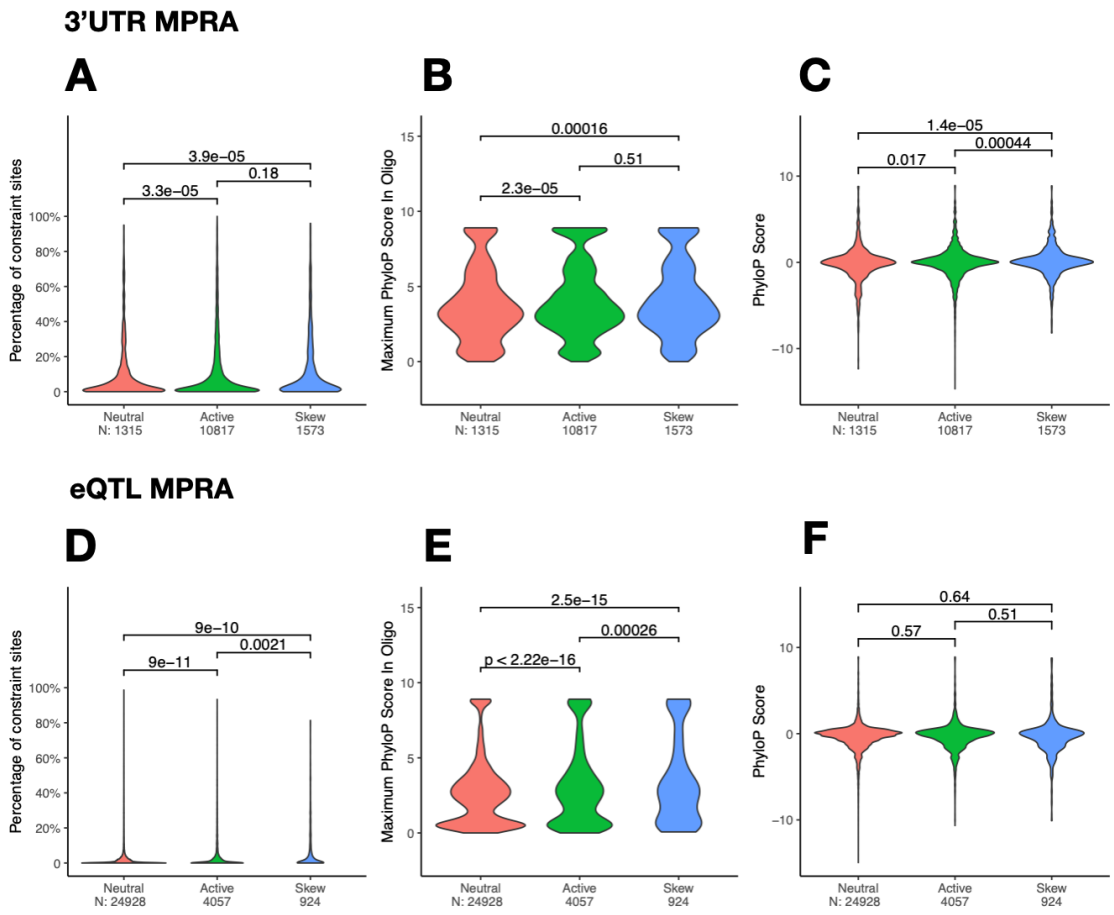
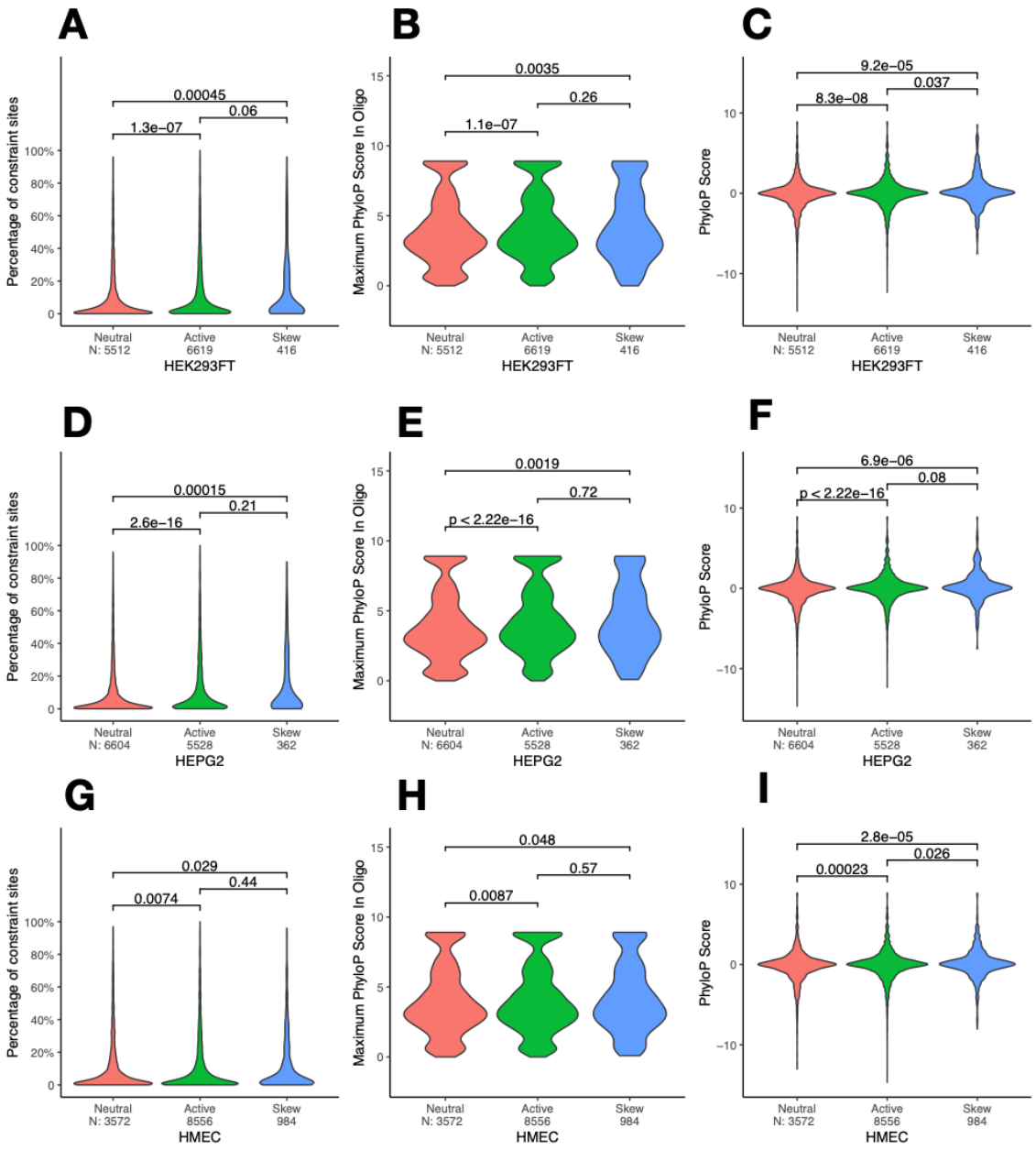


Figure 2.1 Conservation status for variants with different regulatory potential (Neutral, Active, Skew). Metrics to measure conservation: `percentage of constraint sites`, `maximum phyloP score within an oligo` and `phyloP score at the selectively tested base`. (A) 3'UTR variants (B) eQTL variants.


A
Percentage of constraint sites
0.00045
0.06
1.3e−07
100%
80%
60%
40%
20%
0
Neutral
N: 5512
Active
6619
Skew
416
HEK293FT
B
Maximum PhyloP Score In Oligo
0.0035
0.26
1.1e−07
15
10
5
0
Neutral
N: 5512
Active
6619
Skew
416
HEK293FT
C
PhyloP Score
9.2e−05
0.037
8.3e−08
10
0
−10
Neutral
N: 5512
Active
6619
Skew
416
HEK293FT
D
Percentage of constraint sites
0.00015
0.21
2.6e−16
100%
80%
60%
40%
20%
0
Neutral
N: 6604
Active
5528
Skew
362
HEPG2
E
Maximum PhyloP Score In Oligo
0.0019
0.72
p < 2.22e−16
15
10
5
0
Neutral
N: 6604
Active
5528
Skew
362
HEPG2
F
PhyloP Score
6.9e−06
0.08
p < 2.22e−16
10
0
−10
Neutral
N: 6604
Active
5528
Skew
362
HEPG2
G
Percentage of constraint sites
0.029
0.44
0.0074
100%
80%
60%
40%
20%
0
Neutral
N: 3572
Active
8556
Skew
984
HMEC
H
Maximum PhyloP Score In Oligo
0.048
0.57
0.0087
15
10
5
0
Neutral
N: 3572
Active
8556
Skew
984
HMEC
I
PhyloP Score
2.8e−05
0.026
0.00023
10
0
−10
Neutral
N: 3572
Active
8556
Skew
984
HMEC

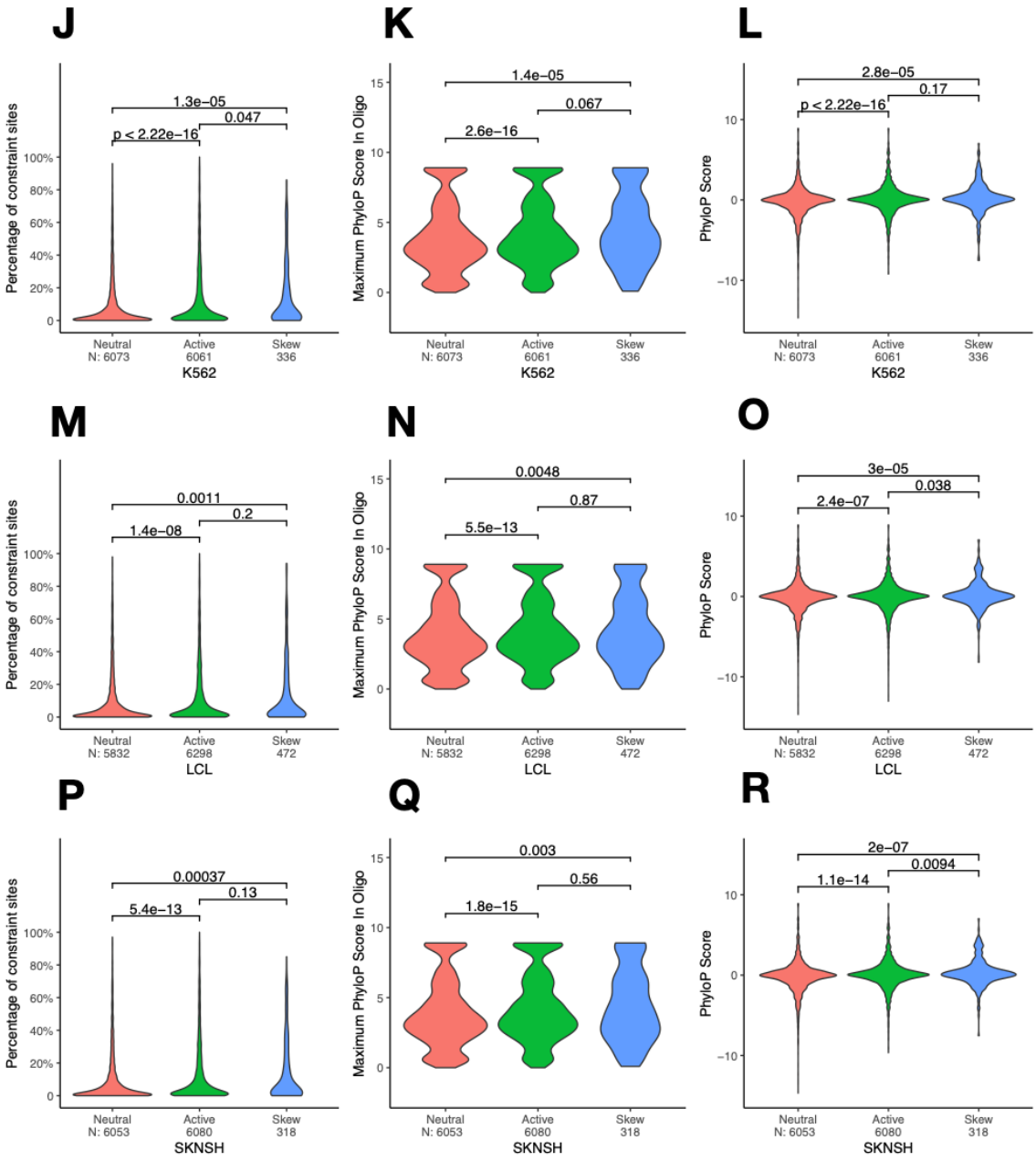


Figure 2.2 Conservation status for 3'UTR variants with different regulatory potential (Neutral, Active, Skew) in six cell types. Metrics to measure conservation: `percentage of constraint sites`, `maximum phyloP score within an oligo` and `phyloP score at the selectively tested base`. (A-C) HEK293FT, (D-F) HEPG2, (G-I) HMEC, (J-L) K562, (M-O) LCL, (P-R) SKNSH.

Predicated variant effect validated at single base resolution

To investigate the predictive ability of phyloP score on variant effect, we compared phyloP score to variant effect assessed by a saturation mutagenesis MPRA. We observed that phyloP score was a strong predictor for variant effect within the LDLR promoter (Spearman correlation coefficient=0.51), with five of the the most constrained sites providing the strongest regulatory effects and also tagging pathogenic ClinVar positions (Figure 2.3). We investigated this pattern in the other 19 promoters or enhancers and found that high constraint is not indicative of correlated effect across the oligo and that cell line specific effects will alter the correlation (correlation ranges from 0.04 to 0.10 for the TERT locus in HEK293T and SF7996 cell lines) (Table 2.1).

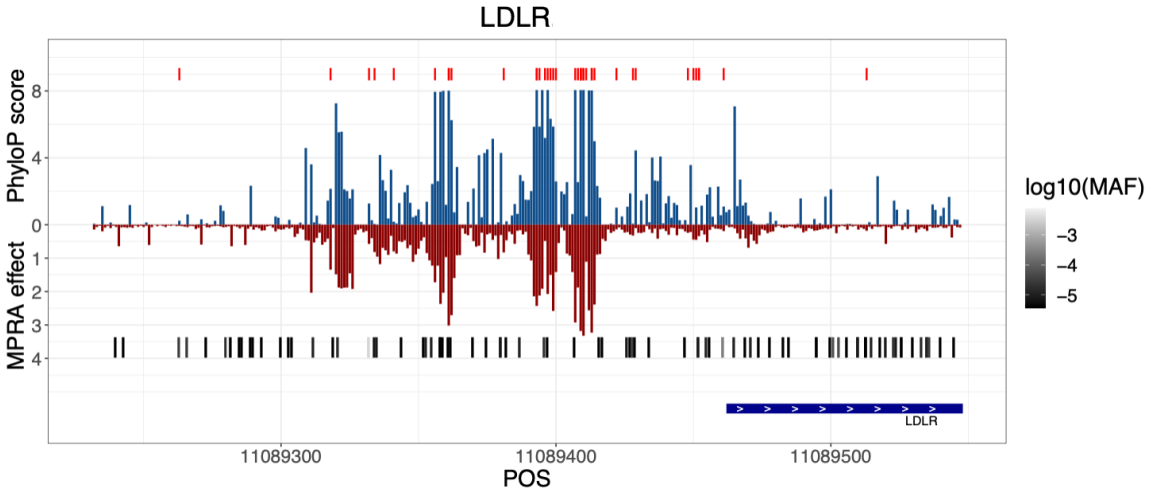


Figure 2.3. PhyloP score plotted in comparison to MPRA effect. Red bar represents MPRA effect, blue bar represents phyloP score. ClinVar pathogenic variants are plotted on the top to highlight known deleterious positions where available.

Table 2.1 Correlation between phyloP score and MPRA effect on each element

Regulatory Space	Name	Cell line	Spearman correlation coefficient
Enhancer	SORT1	HepG2	-0.09
	UCE UC88	Neuro-2a	-0.02
	MYC	LNCaP+100nM DHT	-0.04
	MYC	HEK293T	-0.01
	BCL11A	HEL 92.1.7	0.06
	TCF7L2	MIN6	0.1
	ZRS	NIH/3T3	0.13
	RET	Neuro-2a	0.15
	ZFAND3	MIN6	0.31
	IRF6	HaCaT	0.23
	IRF4	SK-MEL-28	0.35
Promoter	FOXE1	HeLa	0
	Factor IX (F9)	HepG2	0.02
	TERT	HEK293T	0.04
	MSMB	K562	0.04
	HNF4A (P2)	HEK293T	0.13
	TERT	SF7996	0.1
	GP1BB	HEL 92.1.7	0.05
	HBB	HEL 92.1.7	0.19
	HBG1	HEL 92.1.7	0.19

PKLR	K562	0.21
LDLR	HepG2	0.51

II.4.2 Conservation constraints contrast the burden of deleterious alleles in historical and contemporary populations

To illustrate how comparative genomic analysis can provide information for analyzing genomes from the past. We investigated the genomic variants in Balto, the famous sled dog who delivered diphtheria serum to the children of Nome, Alaska, during a 1925 outbreak. We generated a 40.4 fold-coverage nuclear genome from Balto's underbelly skin using protocols for degraded samples and successfully called variants against a panel described by Morrill et al. ((Morrill et al. 2022). We compared Balto to working sled dogs, sled dog breeds, other breeds, village dogs (free-breeding dogs without known breed ancestry), and other canids. We identified evolutionary constrained bases using phyloP evolutionary constraint scores from the dog-referenced version of the 240 species Zoonomia alignment. Ancestry analysis places Balto in a clade of sled dog breeds and working sled dogs and closest to the Alaskan sled dogs (Figure 2.4).

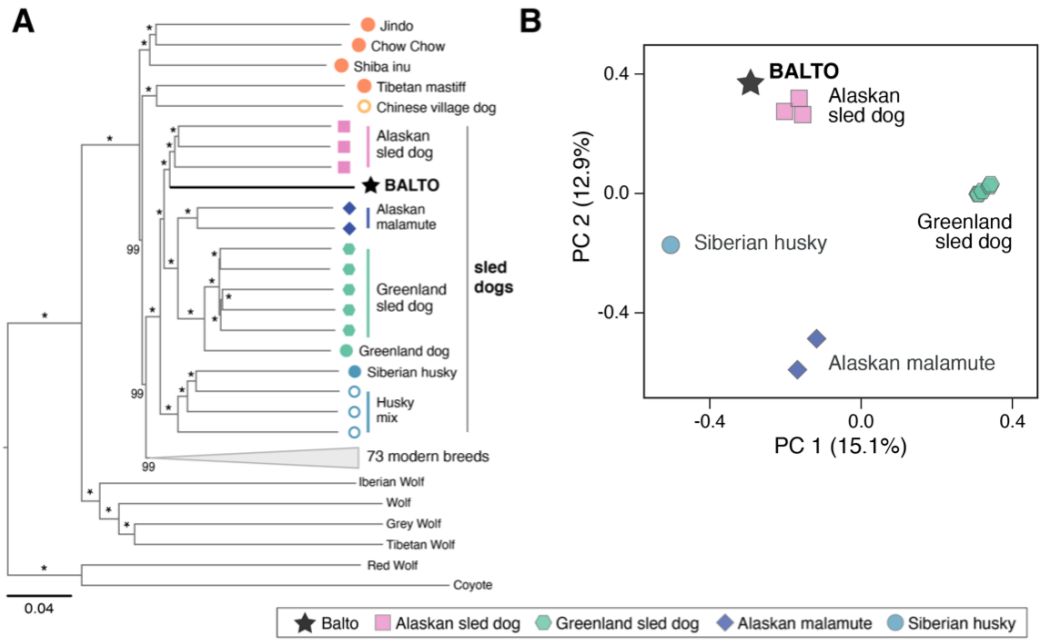


Figure 2.4. Balto clusters most closely with Alaskan sled dogs. (A) Neighbor-joining tree clusters Balto (★) most closely with the outbred, working population of Alaskan sled dogs, and a part of a clade of sled dog populations. (B) Similarly, principal component analysis puts Balto near, but not in, a cluster of Alaskan sled dogs.

Table 2.2 Variant sets used in the analysis of Balto genetics.

Variant set name	Description	Frequency status	Gene set enrichment analysis
Unique, rare variants	Observed in representative dog from a population (once or twice) and not observed in representatives from any of the other populations.	“rare” variants	
Unique, rare variants and evolutionary constrained	**Unique, rare variants** and with a phyloP score above FDR 0.05 cutoff of 2.56	“rare” variants	
Unique, rare variants and missense	**Unique, rare variants** and a missense mutation	“rare” variants	
Derived, common variants	Observed twice in the representative dog from a population and not observed in wolves, for each of the populations	“common” variants	
Derived, common variants, and evolutionary constrained & missense	**Derived, common variants** and with phyloP score above the FDR 0.01cutoff of 3.52 and a missense mutation	“common” variants	Yes

To evaluate the genetic health of Balto’s population of origin, we developed an analytical approach that leveraged the Zoonomia 240 species constraint scores and required only a single individual from each population (necessary since Batlo is the only available representative of his population. Briefly, we selected one

individual at random from each breed or population (57 dogs total) and scored variant positions as either evolutionarily constrained (and more likely to be damaging) or not using the Zoonomia phyloP scores. We also identified variants likely to be “rare” (low frequency) in each dog’s breed or population (Table 2.2, Figure 2.5). Because we couldn’t directly measure population allele frequencies with only a single representative dog, we defined “rare” variants as heterozygous or homozygous variants unique to that dog among all 57 representative dogs. This metric effectively identifies variants occurring at unusually low frequencies (Figure 2.5).

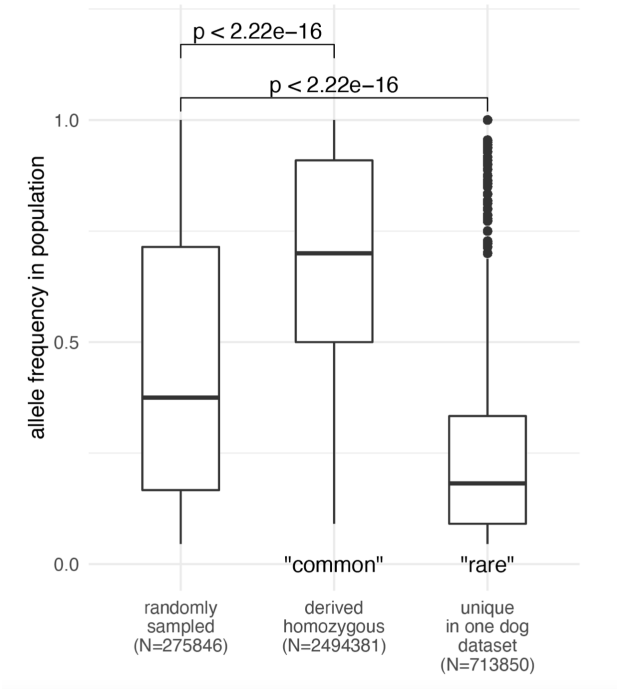


Figure 2.5. Metrics for defining rare and common variants using data from a single representative individual. The allele frequencies, across breeds and populations, of derived homozygous variants (proxy for “common” variation) and variants unique to a single dog (proxy for “rare” variants) show that these definitions do effectively distinguish common and rare variants from randomly sampled variants.

Balto and modern working sled dogs had a lower burden of `rare, potentially damaging variation`, indicating they represent genetically healthier populations (Shindyapina et al. 2020) than breed dogs. Balto and the working sled dogs had significantly fewer potentially damaging variants (missense or constrained) than any breed dog, including the sled dog breeds (Figure 2.6). The pattern persists even in the less genetically diverse Greenland sled dog. Selection for fitness in working sled dog populations appears more effective in removing damaging genetic variation than selection to meet a breed standard.

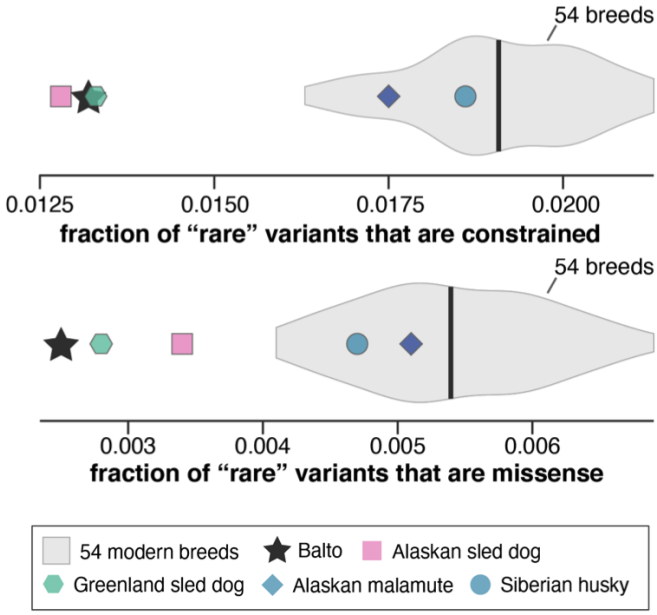


Figure 2.6. The distribution of unique, rare variants annotated by either evolutionary constrained (FDR<0.05) or a missense mutation in single dogs representing modern breeds, working sled dogs, and in Balto.

We found an enrichment for unusual functional variation in Balto's population consistent with adaptation to the extreme environments in which early 20th century sled dogs worked. We identified variants in Balto's genome that were new (not seen in wolves) and likely to be common in his population (homozygous in Balto; Table 2.2). We further filtered for variants that were both protein-altering (missense) and evolutionarily constrained (FDR<0.01), and thus likely to be functional (Table 2.2). Balto was no more likely to carry such variants than dogs from 54 other populations (Figure 2.7), but in Balto these variants tended to disrupt tissue development genes (GO:0009888; 24 genes; 3.02-fold enrichment; FDR=0.013) (table 2.3). This enrichment was unique to Balto, and most of the variants were rare or missing in other dog populations. Even when all GO biological process gene sets are tested in all 57 dogs, Balto's enrichment in tissue development genes is highly unusual. It ranks 4th out of 888,573 dog/set pairs tested (Table 2.3). Phenotype associations from human disease studies suggest that these variants could have influenced skeletal and epithelial development including joint formation, body weight, coordination, and skin thickness (Table 2.4) (Robinson et al. 2008).

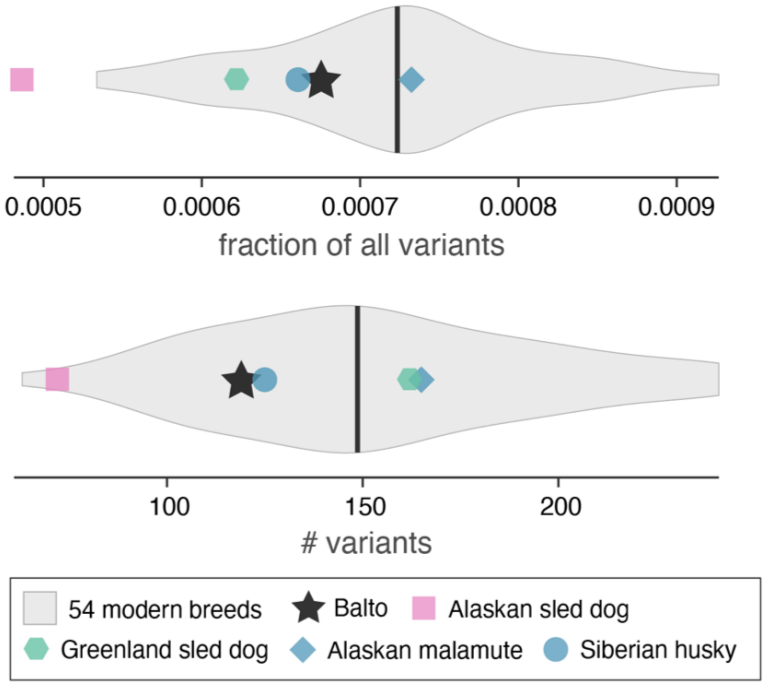


Figure 2.7 The distribution of missense and highly evolutionarily constrained (FDR<0.01) single nucleotide polymorphisms in single dogs representing modern breeds, working sled dogs, and in Balto.

Table 2.3 Gene ontology for homozygous, missense variants from Balto and other breeds.

term label	term	number in reference	expected	dog	FDR
plasma membrane bounded cell projection organization	GO:0120036	1082	6.41	IrishWolfhound01	0.0017
cell projection organization	GO:0030030	1131	6.70	IrishWolfhound01	0.0019
neurogenesis	GO:0022008	1223	6.00	Elo01	0.0077
tissue development	**GO:0009888**	**1687**	**7.95**	**Balto**	**0.0131**
flavonoid glucuronidation	GO:0052696	9	0.07	Keeshond01	0.0335
neuron differentiation	GO:0030182	1018	4.99	Elo01	0.0192
neuron development	GO:0048666	819	4.02	Elo01	0.0145
plasma membrane bounded cell projection organization	GO:0120036	1082	6.10	Pomeranian01	0.0471
cilium assembly	GO:0060271	316	1.87	IrishWolfhound01	0.0149
extracellular structure organization	GO:0043062	295	1.88	FrenchBulldog01	0.0314
xenobiotic glucuronidation	GO:0052697	11	0.08	Keeshond01	0.0316
external encapsulating structure organization	GO:0045229	297	1.89	FrenchBulldog01	0.0223
neuron projection development	GO:0031175	656	3.70	Pomeranian01	0.0407
cell development	GO:0048468	1668	8.18	Elo01	0.0205
generation of neurons	GO:0048699	1081	5.30	Elo01	0.0175
plasma membrane bounded cell projection organization	GO:0120036	1082	5.31	Elo01	0.0148

cell projection organization	GO:0030030	1131	6.37	Pomeranian01	0.0302
plasma membrane bounded cell projection assembly	GO:0120031	405	2.40	IrishWolfhound01	0.0210
cilium organization	GO:0044782	346	2.05	IrishWolfhound01	0.0228
cell projection assembly	GO:0030031	421	2.49	IrishWolfhound01	0.0219
cell projection organization	GO:0030030	1131	5.55	Elo01	0.0230
neurogenesis	GO:0022008	1223	6.89	Pomeranian01	0.0465
neuron projection development	GO:0031175	656	3.22	Elo01	0.0440

Table 2.4 Gene set enrichment analysis of Balto's homozygous, missense variants for Human Phenotype Ontology terms.

Gene Set Name	Genes in Gene Set	Description	Genes in Overlap	FDR.q.value
HP_HIGH_PALATE	635	High palate	13	1.44E-05
HP_ABNORMALITY_OF_THE_KNEE	324	Abnormality of the knee	10	1.44E-05
HP_DECREASED_BODY_WEIGHT	1301	Decreased body weight	17	1.80E-05
HP_ABNORMALITY_OF_PRENATAL_DEVELOPMENT_OR_BIRTH	850	Abnormality of prenatal development or birth	14	1.80E-05

HP_APLASIA_HYPOPLASIA_AFFECTING_BONES_OF_THE_AXIAL_SKELETON	799	Aplasia/hypoplasia affecting bones of the axial skeleton	13	5.67E-05
HP_ABNORMALITY_OF_LOWER_LIMB_JOINT	595	Abnormality of lower limb joint	11	0.000129
HP_ABNORMALITY_OF_THE_OUTER_EAR	1062	Abnormality of the outer ear	14	0.000138
HP_ABNORMAL_LIP_MORPHOLOGY	919	Abnormal lip morphology	13	0.000149
HP_ABNORMAL_CONJUGATE_EYE_MOVEMENT	938	Abnormal conjugate eye movement	13	0.000149
HP_ABNORMAL_JAW_MORPHOLOGY	943	Abnormal jaw morphology	13	0.000149
HP_ABNORMALITY_OF_FACIAL_SKELETON	1119	Abnormality of facial skeleton	14	0.000149
HP_BOWING_OF_THE_LEGS	193	Bowing of the legs	7	0.000149
HP_ABNORMAL_EYELID_MORPHOLOGY	1120	Abnormal eyelid morphology	14	0.000149
HP_ABNORMAL_PALATE_MORPHOLOGY	963	Abnormal palate morphology	13	0.000154
HP_ABNORMALITY_OF_THE_NECK	543	Abnormality of the neck	10	0.000186
HP_ABNORMAL_JOINT_MORPHOLOGY	1001	Abnormal joint morphology	13	0.000186
HP_ABNORMALITY_OF_THE_DENTITION	839	Abnormality of the dentition	12	0.000186
HP_ABNORMAL_TENDON_MORPHOLOGY	696	Abnormal tendon morphology	11	0.000201
HP_ABNORMALITY_OF_THE_PALPEBRAL_FISSURES	710	Abnormality of the palpebral fissures	11	0.000223

HP_FUNCTIONAL_MOTOR_DEFICIT	448	Functional motor deficit	9	0.000257
HP_APLASIA_HYPOPLASIA_INVOLVING_THE_SKELETON	1053	Aplasia/hypoplasia involving the skeleton	13	0.000257
HP_ABNORMAL_RESPIRATORY_SYSTEM_PHYSIOLOGY	1239	Abnormal respiratory system physiology	14	0.000264
HP_ABNORMALITY_OF_THE_CALF	243	Abnormality of the calf	7	0.000344
HP_ABNORMAL_HAIR_MORPHOLOGY	1100	Abnormal hair morphology	13	0.000358
HP_FEEDING_DIFFICULTIES	945	Feeding difficulties	12	0.000424
HP_ABNORMALITY_OF_COORDINATION	1002	Abnormality of coordination	14	9.10E-05
HP_PRENATAL_MOVEMENT_ABNORMALITY	198	Prenatal movement abnormality	7	0.000154
HP_THICKENED_SKIN	406	Thickened skin	9	0.000165
HP_LIMB_JOINT_CONTRACTURE	347	Limb joint contracture	8	0.00034
HP_HAMARTOMA	75	Hamartoma	5	0.000223

II.5 Discussion

In chapter II, I presented two case studies that demonstrate the power of comparative genomics in identifying genetic variations that contribute to the risk of developing diseases and in comprehending their biological mechanisms. The phyloP score, which measures evolutionary constraint, is particularly advantageous because it is mechanistically agnostic.

Using human-centered phyloP score, we tested the power of this metric to predict functional importance of individual human genomic variants regardless of the disease type and cell type. The identification of functional variants, particularly those that are rare and affect only a small number of individuals, is a challenging task that, when using human data only, requires sequencing tens of thousands of samples. Here, with the genome sequence of 240 mammalian species, we are able to tackle this problem. This work illustrates the potential for relatively small cross-species datasets to inform the search for the genomic basis of human disease.

In turn, our Balto study illustrates how comparative genomics enhances the ability to investigate genotype-phenotype associations using small cross-population datasets. Our study investigated the genetics of Balto, the only sequenced individual available from its population, combined with genomic data from pre-existing datasets including 125 dogs from 56 breeds or populations, using population-level phenotypes. By integrating dog-centered phyloP scores

and SNP effect annotations, we developed an analytical approach to deduce the genetic and phenotypic traits of Balto and its population. This analysis would not have been possible without comparative genomics. The dog-centered phyloP score allowed us to measure functional significance, even though we had only limited knowledge of Balto and its population.

II.6 Materials and Methods

II.6.1 Massively parallel reporter assay datasets

MPRA datasets

To investigate the ability of phyloP score to predict functional importance of variation, we assembled MPRA datasets from three sources (Table 2.5). (1) An eQTL set in an MPRA that evaluates 32,373 variants from 3,642 cis-expression quantitative trait loci and control regions in lymphoblastoid cell lines (Tewhey et al. 2016). (2) A 3'UTR set in an MPRA that assesses 12,173 3'UTR variants across six human cell lines, HEK293FT, HEPG2, HMEC, K562, LCL, and SKNSH (Griesemer et al. 2021). (3) A promoter-enhancer set in a saturation mutagenesis MPRA that examines 20 disease-associated gene promoters and enhancers, generating functional measurements for over 30,000 single nucleotide substitutions and deletions (Kircher et al. 2019).

Table 2.5. Massively parallel reporter assay (MPRA) datasets.

Name	Type of element	Cell type	Publication
3'UTR MPRA	3' UTR variants	HEK293T, HEPG2, HMEC, K562, LCL, SKNSH	Tewhey et al. (Tewhey et al. 2016)
eQTL MPRA	eQTL variants	LCL	Griesemer et al. (Griesemer et al. 2021)
Promoter-enhancer MPRA	Promoter and enhancer	Promoter or enhancer-specific cell type	Kircher et al. (Kircher et al. 2019)

We annotated positions under evolutionary constraint by examining sites overlapping phyloP evolutionary constraint scores from the human-referenced version of the 240 species Cactus alignment (Armstrong et al. 2020; Zoonomia Consortium 2020). We calculated the constraint score cutoffs at various false discovery rates (FDR) (Table 2.6).

Table 2.6 Human-centered phyloP score cutoffs at various false discovery rates (FDR)

Total % constrained	phyloP cutoff	percentage of genome (total: 3.09Gb)	bases (Mb)
< 1% FDR	3.27	1.87%	58
< 3% FDR	2.59	2.67%	82
< 5% FDR *	2.27	3.26%	101

< 10% FDR	1.78	4.69%	145
< 20% FDR	1.25	7.56%	233

Statistical analysis for 3'UTR and eQTL MPRA data

For the 3'UTR MPRA, the construct contained a maximum of 101bp of genomic sequence, with 12,134 variants representing the signals of recent human evolutionary adaptation, or drawn from the results of GWAS analyses, some with linked eQTL effect. For the eQTL MPRA set, the coordinates for the 150 bp or genomic sequence for the constructs and the 29,108 unique variant positions were downloaded. These represent a combination of GWAS and eQTL hits drawn from lymphoblastoid cell lines (LCLs), with results drawn from populations of European and West African ancestry. Genomic positions were lifted from hg19 to hg38 and annotated with mammalian phyloP scores.

3'UTR or eQTL MPRA oligos were categorized into three groups based on activity: (1) Neutral, oligo with no regulatory activity for either allele; (2) Active, oligo with either allele showing regulatory activity; and (3) Skew, differential expression is observed between the two target alleles within an oligo. For the 3'UTR analysis, activity was extended to include any of the six cell lines. A t-test was performed to assess differences between the groups.

Statistical analysis for promoter-enhancer MPRA data

For the promoter-enhancer MPRA analysis, we annotated every position in the promoter or enhancer element with a phyloP score. For the final regulatory space, the MPRA data from the error-prone PCR saturation mutagenesis of 9 enhancer and 11 promoter disease-associated regions, size ranging from 186 to 600 bp, was downloaded (Kircher et al. 2019). That data set contained the functional measurements for over 30,000 single nucleotide substitutions and deletions, as each position from the region was mutated to the other three nucleotides or a deletion, and an MPRA effect read out reported. We calculated the average of the MPRA effects for each position to represent the relative ability of this position. These positions were subsequently annotated with mammalian phyloP scores, and the Spearman correlation between phyloP score and MPRA effect for each promoter region computed. The correlation ranged from negligible in the *FOXE1* locus (0.00) to highest in the *LDLR* locus (0.51), and so we also annotated the regions using ClinVar pathogenic variants and plotted the results for visual inspection.

II.6.2 An analytical approach to study Balto genetics

Population representative sampling

As the only representative of his population, Balto was included alongside one randomly chosen sample from each of the 57 other populations to identify unique, population-specific genetic variants among 67,085,518 biallelic single nucleotide polymorphisms. This group comprised Balto, one Alaskan sled dog, one Greenland sled dog, and 54 modern purebred dogs, including a Siberian husky and an Alaskan malamute (Table 2.7 samples selected in the `Population Variants Analysis`). Similarly, we selected an additional 5 to 11 random samples from 10 modern breeds where possible, as well as all remaining Greenland sled dog samples, to evaluate the population-wide allele frequency of these variants (Table 2.7 samples selected in the `Population Frequency Analysis`).

Table 2.7. Whole genome data with sample metadata and membership in analysis datasets.

SRA BioSample	Group	Representative in `Population Variants Analysis`	Representative in `Population Frequency Analysis`
SAMN23691391	working sled dog	Balto, Alaskan Sled Dog (1920s)	Alaskan Sled Dog (1920s)
SAMN23691392	working sled dog		Alaskan Sled Dog (modern)
SAMN23691393	working sled dog	Alaskan Sled Dog (modern)	Alaskan Sled Dog (modern)
SAMN23691394	working sled dog		Alaskan Sled Dog (modern)

SAMN14210383	working sled dog		Greenland Sled Dog
SAMN14210374	working sled dog		Greenland Sled Dog
SAMN14210381	working sled dog		Greenland Sled Dog
SAMN14210379	working sled dog	Greenland Sled Dog	Greenland Sled Dog
SAMN14210377	working sled dog		Greenland Sled Dog
SAMN03801690	sled dog breed	Siberian Husky	Siberian Husky
SAMEA3539249	sled dog breed		Siberian Husky
SAMN03168389	sled dog breed		Siberian Husky
SAMN08872810	sled dog breed		Alaskan Malamute
SAMN03168378	sled dog breed		Alaskan Malamute
SAMN02194722	breed dog	Afghan Hound	
SAMN03580390	breed dog	Airedale Terrier	
SAMEA4346711	breed dog	Alpine Dachsbracke	
SAMN08872960	breed dog		American Cocker Spaniel
SAMN08872961	breed dog		American Cocker Spaniel
SAMN08872962	breed dog		American Cocker Spaniel
SAMN08872963	breed dog		American Cocker Spaniel
SAMN08872964	breed dog		American Cocker Spaniel
SAMN03323677	breed dog	Australian Cattle Dog	
SAMN08872826	breed dog	Beagle	
SAMEA3928140	breed dog		Bearded Collie
SAMEA4505492	breed dog	Bearded Collie	Bearded Collie

SAMEA4504824	breed dog		Bearded Collie
SAMEA4505493	breed dog		Bearded Collie
SAMEA4504827	breed dog		Bearded Collie
SAMN03801646	breed dog	Belgian Sheepdog	Belgian Sheepdog
SAMN08872843	breed dog		Belgian Sheepdog
SAMN08872844	breed dog		Belgian Sheepdog
SAMN08872845	breed dog		Belgian Sheepdog
SAMN08872846	breed dog		Belgian Sheepdog
SAMN08872849	breed dog		Belgian Tervuren
SAMN08872850	breed dog		Belgian Tervuren
SAMN08872851	breed dog		Belgian Tervuren
SAMN08872852	breed dog		Belgian Tervuren
SAMN08872853	breed dog		Belgian Tervuren
SAMN08872859	breed dog	Belgian Tervuren	
SAMN03801648	breed dog		Bernese Mountain Dog
SAMN08872866	breed dog	Bernese Mountain Dog	Bernese Mountain Dog
SAMN08872868	breed dog		Bernese Mountain Dog
SAMN08872875	breed dog		Bernese Mountain Dog
SAMN08872876	breed dog		Bernese Mountain Dog
SAMN04196853	breed dog	Black and Tan Coonhound	
SAMN03323668	breed dog	Black Russian Terrier	
SAMN03801649	breed dog	Bloodhound	
SAMN01908107	breed dog		Border Collie

SAMN03801650	breed dog		Border Collie
SAMN03801651	breed dog		Border Collie
SAMEA4504821	breed dog		Border Collie
SAMEA3724571	breed dog	Border Collie	Border Collie
SAMN08872905	breed dog	Borzoi	
SAMN08872906	breed dog	Boston Terrier	
SAMN08872908	breed dog	Bouvier des Flandres	
SAMN02921305	breed dog	Boxer	
SAMN08872913	breed dog		Bull Terrier
SAMN08872914	breed dog		Bull Terrier
SAMN08872915	breed dog		Bull Terrier
SAMN08872916	breed dog		Bull Terrier
SAMN08872917	breed dog		Bull Terrier
SAMN08872918	breed dog	Bull Terrier	
SAMN08872926	breed dog	Carolina Dog	
SAMN03801656	breed dog	Chihuahua	
SAMN03075611	breed dog	Chinese Crested	
SAMN03801658	breed dog	Chow Chow	
SAMEA4506890	breed dog	Elo	
SAMN03801654	breed dog	English Bulldog	
SAMN03580391	breed dog	English Springer Spaniel	
SAMEA4505501	breed dog	Entlebucher Mountain Dog	
SAMEA3928146	breed dog	French Bulldog	

SAMN02585201	breed dog		German Shepherd Dog
SAMN02585202	breed dog		German Shepherd Dog
SAMN02585203	breed dog		German Shepherd Dog
SAMN02585204	breed dog		German Shepherd Dog
SAMN02585205	breed dog		German Shepherd Dog
SAMN03580392	breed dog	German Shepherd Dog	
SAMEA3928144	breed dog	German Wirehaired Pointer	
SAMN03801669	breed dog	Great Dane	
SAMN08873062	breed dog	Greater Swiss Mountain Dog	
SAMN03801672	breed dog	Irish Wolfhound	
SAMN03801674	breed dog	Italian Greyhound	
SAMN03580400	breed dog	Jack Russell Terrier	
SAMN03168383	breed dog	Jamthund	
SAMN08873128	breed dog	Keeshond	
SAMN08873130	breed dog	Komondor	
SAMEA2417015	breed dog	Labrador Retriever	Labrador Retriever
SAMEA2417016	breed dog		Labrador Retriever
SAMN03801675	breed dog		Labrador Retriever
SAMN04196848	breed dog		Labrador Retriever
SAMN08873155	breed dog		Labrador Retriever
SAMN04196863	breed dog	Lowchen	
SAMN03801664	breed dog	Mastiff (English)	

SAMN03323675	breed dog	Scotia Duck Tolling Retriever	
SAMN03801676	breed dog	Pekingese	
SAMEA4506892	breed dog	Pomeranian	
SAMN08873219	breed dog		Portuguese Water Dog
SAMN08873220	breed dog		Portuguese Water Dog
SAMN08873221	breed dog		Portuguese Water Dog
SAMN08873222	breed dog		Portuguese Water Dog
SAMN08873223	breed dog		Portuguese Water Dog
SAMN03801684	breed dog	Rottweiler	
SAMN03801685	breed dog	Saint Bernard	
SAMN03801686	breed dog	Saluki	
SAMN08873258	breed dog	Samoyed	
SAMN03580401	breed dog	Scottish Deerhound	
SAMN08873266	breed dog	Shetland Sheepdog	
SAMN01974493	breed dog		Tibetan Mastiff
SAMN02570458	breed dog		Tibetan Mastiff
SAMN02585152	breed dog		Tibetan Mastiff
SAMN02585153	breed dog		Tibetan Mastiff
SAMN02585154	breed dog		Tibetan Mastiff
SAMN02585160	breed dog	Tibetan Mastiff	
SAMN03580403	breed dog	Tibetan Terrier	
SAMN03801693	breed dog		West Highland White Terrier
SAMN08873423	breed dog		West Highland White Terrier

SAMN03580395	breed dog		West Highland White Terrier
SAMN03580396	breed dog		West Highland White Terrier
SAMN03580397	breed dog		West Highland White Terrier
SAMN04195509	breed dog	Yorkshire Terrier	
SAMN03168352	village dog		Indigenous - China
SAMN03168353	village dog		Indigenous - China
SAMN03168354	village dog		Indigenous - China
SAMN01974490	village dog		Village Dog - China
SAMN01974491	village dog		Village Dog - China

Dog-referenced mammalian evolutionary constraint

We selected biallelic SNPs under evolutionary constraint by examining sites overlapping phyloP evolutionary constraint scores from the dog-referenced version of the 240 species Cactus alignment (Armstrong et al. 2020; Zoonomia Consortium 2020). We calculated the constraint score cutoffs at various false discovery rates (FDR) (Table 2.8).

Table 2.8. Dog-centered phyloP score cutoffs at various false discovery rates (FDR).

Total % constrained	phyloP cutoff	percentage of genome (total: 2.33Gb)	bases (Mb)
< 1% FDR	3.52	2.11%	49.2
< 3% FDR	2.88	2.84%	66.0

< 5% FDR *	2.56	3.43%	79.8
< 10% FDR	2.10	4.69%	109.2
< 20% FDR	1.59	7.02%	163.4

Unique, rare, and potentially deleterious variants

We first identified all “population-unique” variants, defined as those observed in the representative dog from a population (either once or twice) and not observed in representatives from any of the other populations (Table 2.2, Table 2.7). With this method, we identified 206,164 population-unique variants for Balto, 120,279 for the Alaskan sled dog, 119,482 variants for the Greenland sled dog, 120,780 unique to the Alaskan malamute, and 133,200 unique to the Siberian husky.

We hypothesized that Balto, and dogs from modern working sled dog populations, should carry fewer rare, potentially damaging genetic variants than modern, pedigreed sled dog breeds selected primarily for aesthetics (Siberian husky and Alaskan malamute). We confirmed that population-unique variants tend to be uncommon by calculating the allele frequencies in its population using ten breeds and one working dog group: bearded collie, Belgian sheepdog, Belgian Tervuren, Bernese mountain dog, Border Collie, Bull terrier, Entlebucher Sennenhund, German shepherd dog, Greenland sled dog, Labrador retriever, and Tibetan mastiff, each with 5 to 11 dogs.

We used Zoonomia PhyloP scores and SNPeff annotations to identify which population-unique variants were either “evolutionarily constrained” (phyloP score above the FDR 0.05 cutoff of 2.56) or a missense mutation (determined using SnpEff (Cingolani et al. 2012), and thus more likely to have functional consequences (Table 2.2). We grouped the dogs into working dog groups including Balto, Alaskan Sled dog, and Greenland Sled dog, and modern breeds including all the other 54 dogs. We then applied Student’s t-test on the percentage of “evolutionarily constrained” or missense mutation for the two groups.

Derived, common, and potentially beneficial variants

We identified “homozygous derived” variants, defined as those observed twice in the representative dog from a population and not observed in wolves, for each of the populations (Table 2.2, Table 2.7). With this method, we identified 176,135 homozygous derived variants for Balto, 148,036 variants for Alaskan sled dog, 260,457 variants for Greenland sled dog, 225,270 variants for Alaskan Malamute, and 189,188 variants for Siberian Husky. We confirmed that homozygous variants in each representative dog tend to be “common” in their population by calculating the allele frequency of the homozygous derived variants in its own breed using the same set of dogs described above for calculating allele frequency for population-unique variants (Figure 2.5).

Gene ontology for homozygous, missense variants from Balto and other breeds.

We further defined variants likely to be functional as those that were both “highly evolutionarily constrained” (defined by phyloP score above the FDR>0.01 cutoff of 3.52) and a missense mutation. We annotated the variant by genes, and performed gene set enrichment against all Gene Ontology Biological Process gene sets (http://geneontology.org/) using the R package rbioapi v. 0.7.4 (Rezwani et al. 2022; Mi et al. 2021). We set the reference organism to human (‘organism = 9606’), as its annotation is more comprehensive, and the annotation dataset to GO Biological Processes (‘annot_dataset = "GO:0008150"’). To assess how unusual the observed enrichment was, we extended the enrichment analysis to test all GO Biological process gene sets with fewer than 2500 genes (15,589 sets), and all 57 dogs that had been selected to represent their populations (including Balto), totalling 888,573 tests. If a variant mapped to multiple genes, all genes were included.

We also tested for overlap between Balto’s variant genes and genes implicated in particular phenotypes in human studies using the Human Phenotype Ontology (Köhler et al. 2021) and the “Investigate gene sets" feature provided by GSEA (http://www.gsea-msigdb.org/).

CHAPTER III. POPULATION HISTORY OF WOLFDOG HYBRIDS

III.1 Preface

Kathryn Lord and Brittney Logan collected DNA samples from all studied hybrids. The process of DNA extraction from the fur samples was expertly managed by Gaurav Chauhan. Concurrently, Gencove was responsible for extracting DNA from saliva samples and further contributed by conducting low-pass sequencing and imputation. I performed all the analysis described in this chapter.

III.2 Abstract

This chapter provides an in-depth investigation into the population history of wolf-dog hybrids. Utilizing principal component analysis, we examined the relationships among wolf, dog, and wolfdog hybrid genomes. We devised a Hidden Markov Model-based approach to infer local ancestry within this admixed population. Through both global and local ancestral calling, we identified regions predominantly characterized by wolf ancestry. We discussed the potential selection within captive wolfdog hybrid populations. Furthermore, we conducted a thorough investigation into inbreeding and kinship patterns, as well as the extent of linkage disequilibrium in wolves, dogs, and wolf-dog hybrids. Our findings suggest that the hybrids have undergone sufficient generations of admixture, making them suitable for trait mapping, which will be explored in chapter IV.

III.3. Background

The emergence of wolf-dog hybrids is the result of multiple generations of interbreeding between North American wolves and domesticated dogs. Although wolves and dogs have significantly different genomes, they are members of the same species and are capable of producing fertile offspring. Their genome sizes are similar to other mammals, including humans, containing around 2.5 billion DNA base pairs (Sinding et al. 2021). Wolves and dogs have an average of one single nucleotide polymorphism (SNP) difference per 580 base pairs in their genomes (Ostrander and Wayne 2005).

Wolf-dog hybrids are rare in the wild despite the fact that they can mate and reproduce healthy offspring. The main reason is that their behavior is distinct from both wolves and dogs. They are not friendly enough to be around humans due to the wolf component, and, due to the dog component, they lack abilities such as hunting and traveling long distances, which would allow them to survive in the wild. In the wild, introgression occurs in both directions; over time, only beneficial alleles survive in the genome of admixed individuals. For example, there are instances where black wolves have acquired the gene responsible for black coat color through interbreeding with domestic dogs, these black-colored wolves have higher fitness than gray-colored wolves (Caniglia et al. 2013; Schweizer et al. 2018). A similar example of transfer in the other direction is the transfer of an adaptive variant of EPAS1 and linked variants that potentially

function in hypoxia response at high elevation, from highland wolves to highland dogs. (vonHoldt et al. 2017). In both cases, only the beneficial allele and the linked variants persist in the genome of mixed individuals.

The upcoming two chapters, chapter III and IV, will discuss how we used the captive hybrid population, which displays a diverse spectrum of behaviors ranging from very dog-like to very wolf-like, to identify genetic alterations that transpired during domestication. The principal aim of this chapter is to delve into the population history of hybrids and their relationship with their ancestral populations.

III.4. Results

III.4.1 Wolf, dog, and wolfdog hybrid genomes

We enrolled 90 putative wolf-dog hybrids (wolfdogs) from five sanctuaries across North America, and collected DNA from saliva or fur samples. For each animal, we recorded sex, age, and living conditions (number of animals in the enclosure and enclosure size) (Table 3.1). We densely genotyped each animal with a DNA sample using low-pass sequencing [coverage: 1.0× ± 0.6× (±SD)] and imputed against a panel of 435 deeply sequenced canids that included 393 dogs, 36 wolves and 6 other wild canids (Morrill et al. 2022). For the 76 successfully

sequenced hybrids, we called on average of 31,790,793±282,460 (±SD) SNPs and 13,375,106 ±121,185 (±SD) indels per dog.

Table 3.1 Data collection for 90 hybrids.

ID	Population	Morpho-logy	Genoty-ping	Behavior	Enclosure size (CM^2)	Number of animals in habitat
WD001	Wolfdog	1	1	0	856	2
WD002	Wolfdog	1	1	1	279	2
WD003	Wolfdog	1	1	1	404	2
WD004	Wolfdog	1	1	1	419	2
WD005	Wolfdog	1	1	1	932	2
WD006	Wolfdog	1	1	1	1447	3
WD007	Wolfdog	1	1	1	1728	2
WD008	Wolfdog	1	1	0	171	2
WD009	Wolfdog	1	1	1	454	2
WD010	Wolfdog	1	1	1	815	2
WD011	Wolfdog	1	1	1	29	2
WD012	Wolfdog	1	1	1	549	2
WD013	Wolfdog	1	1	1	1447	3
WD014	Wolfdog	1	1	1	880	2
WD015	Wolfdog	1	1	1	419	2
WD016	Wolfdog	1	1	1	1095	2
WD017	Wolfdog	1	1	1	241	2

WD018	Wolfdog	1	1	1	1728	2
WD019	Wolfdog	1	1	1	2885	2
WD020	Wolfdog	1	1	1	1045	2
WD021	Wolfdog	1	1	1	1572	2
WD022	Wolfdog	1	1	0	NA	NA
WD023	Wolfdog	1	1	1	404	2
WD024	Wolfdog	1	1	1	260	2
WD025	Wolfdog	1	1	1	29	2
WD026	Wolfdog	1	1	1	1045	2
WD027	Wolfdog	1	1	1	815	2
WD028	Wolfdog	1	1	1	691	1
WD029	Wolfdog	1	1	1	691	1
WD030	Wolfdog	1	1	1	880	2
WD031	Wolfdog	1	1	0	164	2
WD032	Wolfdog	1	1	1	1045	2
WD033	Wolfdog	1	1	1	691	1
WD034	Wolfdog	1	1	1	549	2
WD035	Wolfdog	1	1	1	1095	2
WD036	Wolfdog	1	1	1	1095	2
WD037	Wolfdog	1	1	1	2885	2
WD038	Wolfdog	1	1	1	639	2
WD039	Wolfdog	1	1	1	856	1
WD040	Wolfdog	1	1	1	691	1
WD041	Wolfdog	1	1	1	1572	2

WD042	Wolfdog	1	1	1	691	1
WD043	Wolfdog	1	1	1	691	1
WD044	Wolfdog	1	1	1	1095	2
WD045	Wolfdog	1	1	1	932	2
WD046	Wolfdog	1	1	1	267	2
WD047	Wolfdog	1	1	1	267	2
WD048	Wolfdog	1	1	1	631	2
WD049	Wolfdog	1	1	1	1984	3
WD050	Wolfdog	1	1	1	2758	1
WD051	Wolfdog	1	1	1	101	1
WD052	Wolfdog	1	1	1	112	2
WD053	Wolfdog	1	1	1	631	2
WD054	Wolfdog	1	1	1	1984	3
WD055	Wolfdog	1	1	1	292	2
WD056	Wolfdog	1	1	1	292	2
WD057	Wolfdog	1	1	1	261	1
WD058	Wolfdog	1	1	1	218	1
WD059	Wolfdog	1	1	1	292	3
WD060	Wolfdog	1	1	1	332	2
WD061	Wolfdog	1	1	1	346	1
WD062	Wolfdog	1	1	1	1984	3
WD063	Wolfdog	1	1	1	218	1
WD064	Wolfdog	1	1	1	499	2
WD065	Wolfdog	1	1	1	499	2

WD066	Wolfdog	1	1	1	2885	4
WD067	Wolfdog	1	1	1	2885	4
WD068	Wolfdog	1	1	1	2885	4
WD069	Wolfdog	1	1	1	29	2
WD070	Wolfdog	1	1	1	1728	2
WD071	Wolfdog	1	1	1	29	2
WD072	Wolfdog	1	1	1	224	1
WD073	Wolfdog	1	1	1	1728	2
WD074	Wolfdog	1	1	0	NA	NA
WD075	Wolfdog	1	1	1	241	2
WD076	Wolfdog	1	1	1	171	2
WD077	Wolfdog	1	0	1	275	2
WD078	Wolfdog	1	0	1	275	2
WD079	Wolfdog	1	0	1	184	1
WD080	Wolfdog	1	0	1	639	2
WD081	Wolfdog	1	0	1	1045	2
WD082	Wolfdog	1	0	1	1447	3
WD083	Wolfdog	1	0	1	260	2
WD084	Wolfdog	1	0	1	2885	3
WD085	Wolfdog	1	0	1	112	2
WD086	Wolfdog	1	0	0	NA	NA
WD087	Wolfdog	1	0	0	419	2
WD088	Wolfdog	1	0	0	164	2
WD089	Wolfdog	1	0	0	151	1

WD090	Wolfdog	1	0	0	NA	NA

Wolfdog hybrids are admixed of dogs and North American wolves. To explore the population structure of hybrids and the relationship between them and other canids, we constructed a `diverse canine reference panel` including 110 dogs, 33 North American and 26 European wolves, and 3 coyotes (Table 3.9, Table 3.10). We conducted a principal component analysis. We first explored the relationship between our hybrid population and wolves, dogs, and coyotes using principal component analysis (Figure 3.1). The first component, which explains 44% of the genetic variability, shows all but two of the putative wolf-dog hybrids falling along a continuum between modern dogs and North American wolves. The second component, which explains 8.31% of the genetic variability, separates coyotes from wolves, and the intermediate position of two wolfdogs on this component suggests they may have substantial coyote ancestry. Within the wolf cluster, regional geographic groupings are evident, North American wolves and Eurasian wolves form two sub clusters. Notably, despite large variability in dog morphology, dogs are much more genetically close to each other than are wolves and form a tight cluster distinct from any other canine species.

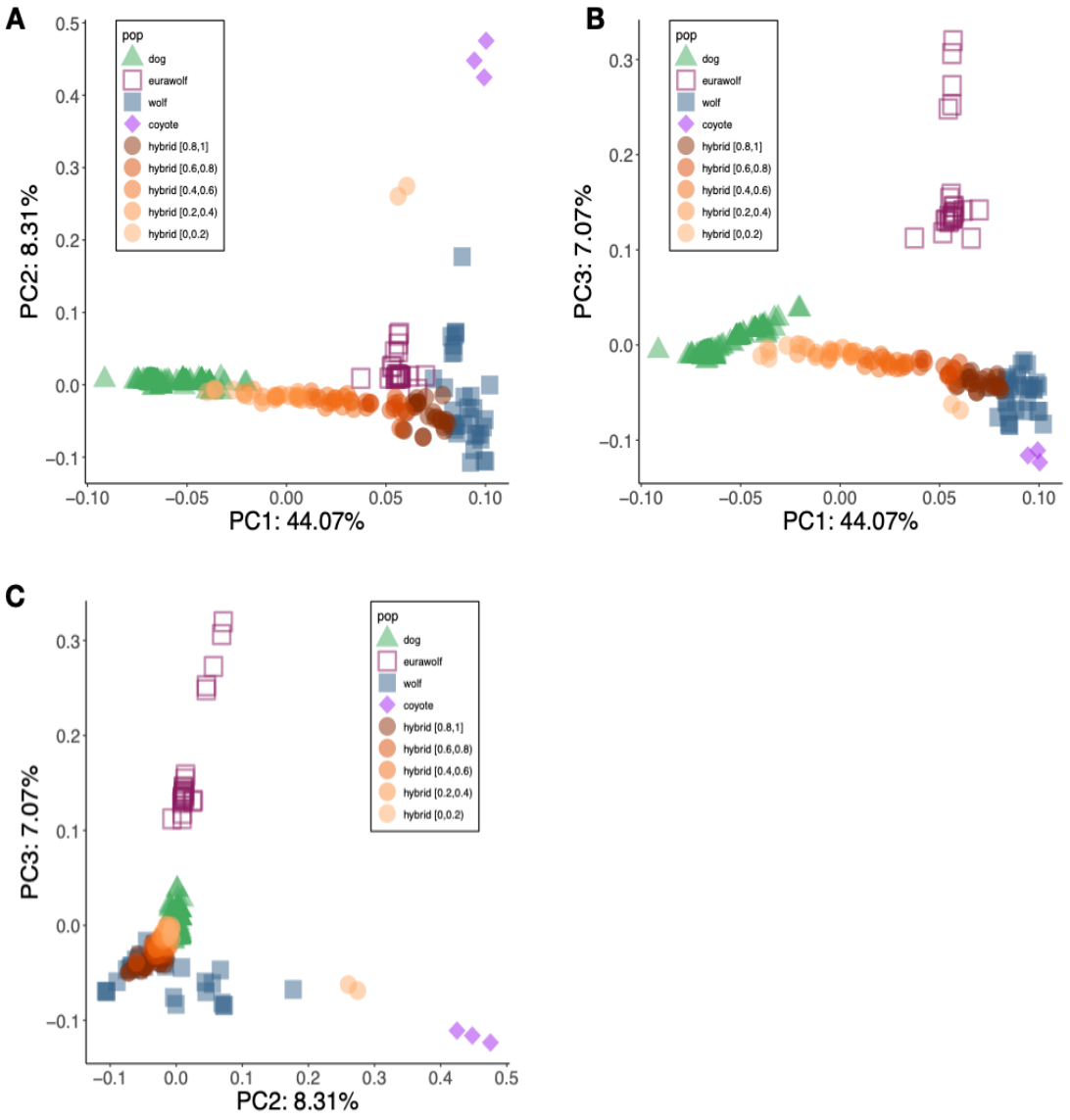


Figure 3.1 Principal component analysis of 214 reference canids and 76 wolf-dog and wolf-dog-coyote hybrids on 2,412,489 markers. Color represents different canine populations, with light orange represents low wolf content hybrids and dark orange represents high wolf content hybrids. (A) PC1 vs. PC2 (B) PC1 vs. PC3 (C) PC2 vs. PC3.

To determine genetic distances between populations, we computed a matrix of pairwise Fst estimates from the 2,412,489 autosomal SNPs. The dogs were more divergent from North American wolves (Fst: 0.19) than from Eurasian wolves (Fst: 0.16), which is consistent with introgression between free-ranging dogs and wild wolves in Europe. The hybrids, which were descendants of dogs and North American wolves, showed greater genetic similarity to North American wolves (Fst: 0.05) than to Eurasian wolves (Fst: 0.06). Furthermore, the hybrids showed a greater genetic distance from dogs (Fst: 0.10) than to wolves (Fst: 0.05 to North American wolves), suggesting a higher proportion of wolf ancestry in this hybrid population (Table 3.2). This finding is further supported by the global ancestry analysis, which is discussed in subsequent sections.

Table 3.2 Fst among dog, wolf, and hybrid populations

	Hybrid	Dog	North American Wolf
Dog	0.0971		
North American Wolf	0.0459	0.1946	
Eurasian Wolf	0.0556	0.1587	0.0663

As a population formed by the mixing of two different ancestral groups, hybrids are expected to have shorter linkage blocks. To confirm whether our data align with this prediction, we calculated the r^2 values for both the hybrid and ancestral populations (Figure 3.2). The results showed that, on average, the r^2 value in

hybrids decreases to less than 0.2 at a length of 4.2kb. This value is intermediate between that of wolves (2.3kb) and dogs (10kb for village dogs and >50kb for purebred dogs).

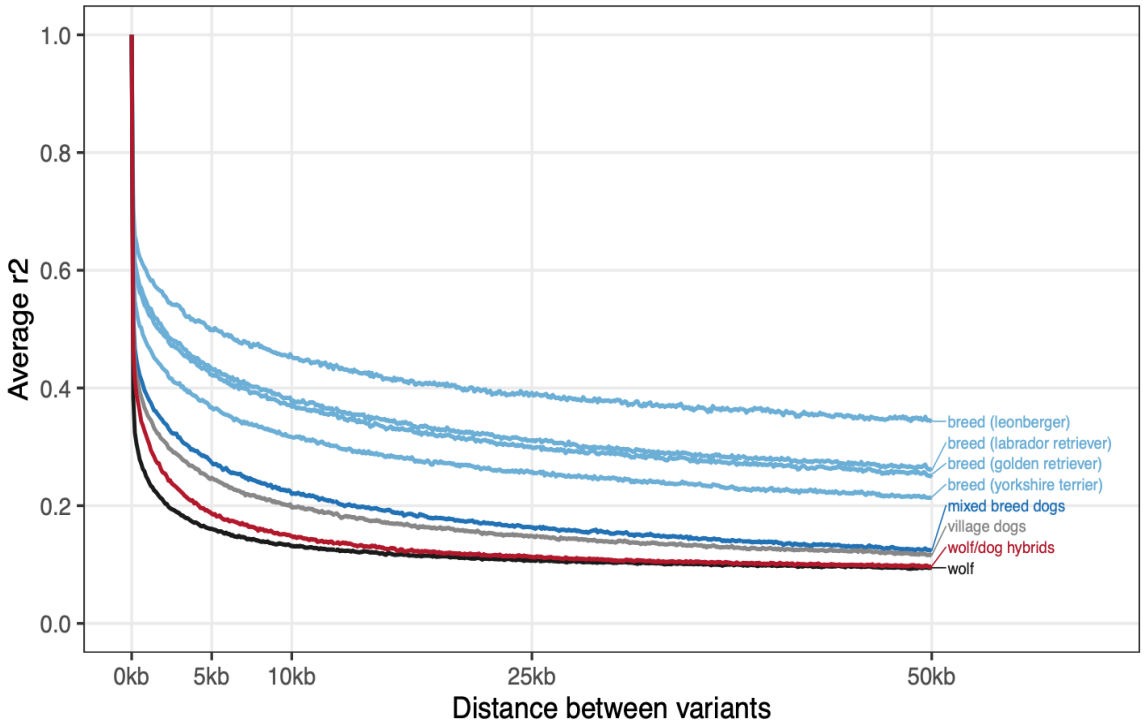


Figure 3.2 Decay of linkage disequilibrium in hybrid population and ancestral populations including wolves and dogs. Linkage blocks in hybrids (red) were observed to be slightly longer than those in wolves but much shorter than village dogs and purebred dogs. Linkage disequilibrium was calculated using 25 individuals, randomly selected from all dogs, wolves, and hybrids within each population. The analysis measured r^2 between each of 20,000 randomly chosen SNPs and all variants with a minor allele frequency >0.025 within a 100kb range of the index SNP.

III.4.2 Inbreeding and kinship

We evaluated the inbreeding status of the hybrids and the parental populations, wolves and dogs (Table 3.3). On average, hybrid individuals exhibited lower levels of inbreeding (F_{RoH} 0.019 ± SD: 0.018, n= 76) compared to both domesticated dogs (breed dogs: F_{RoH} 0.145 ± SD: 0.057, n=96; non-breed dogs: F_{RoH} 0.066 ± SD: 0.065, n=16) and the reference wolf population (F_{RoH} 0.062 ± SD: 0.080, n=59). This observation aligns with the fact that recombination breaks long genomic blocks into smaller pieces in admixed populations. The average length of ROH is also shorter in hybrid individuals compared to dog populations (377kb ± SD: 41kb). As expected, purebred dogs have the highest levels of inbreeding and longer ROH (491kb ± SD: 39kb). The length of ROH varies considerably in wolf population (377kb ± SD: 170kb). In theory, wolves living in the wild should have relatively lower levels of inbreeding and shorter ROH, as they have not undergone artificial selection by humans. However, studies have revealed that certain wolf populations that face harsh living conditions maintain a small effective population size and experience high levels of inbreeding (Robinson et al. 2019). (Table 3.3, Figure 3.3)

Table 3.3. Inbreeding coefficient and runs of homozygosity for wolf, dog, and hybrid populations.

	INBREEDING COEFFICIENT (FROH)				ROH LENGTH (KB)			
POP	Mean	SD	Max	Min	Mean	SD	Max	Min
Purebred Dog	0.1454	0.0571	0.2599	0.0153	491	39	553	331
Coyote	0.0073	0.0116	0.0206	0.0000	216	188	339	0
Hybrid	0.0194	0.0188	0.0844	0.0004	377	41	516	305
Village dog	0.0665	0.0647	0.2210	0.0143	402	64	530	347
Wolf	0.0617	0.0806	0.3560	0.0000	377	170	572	0

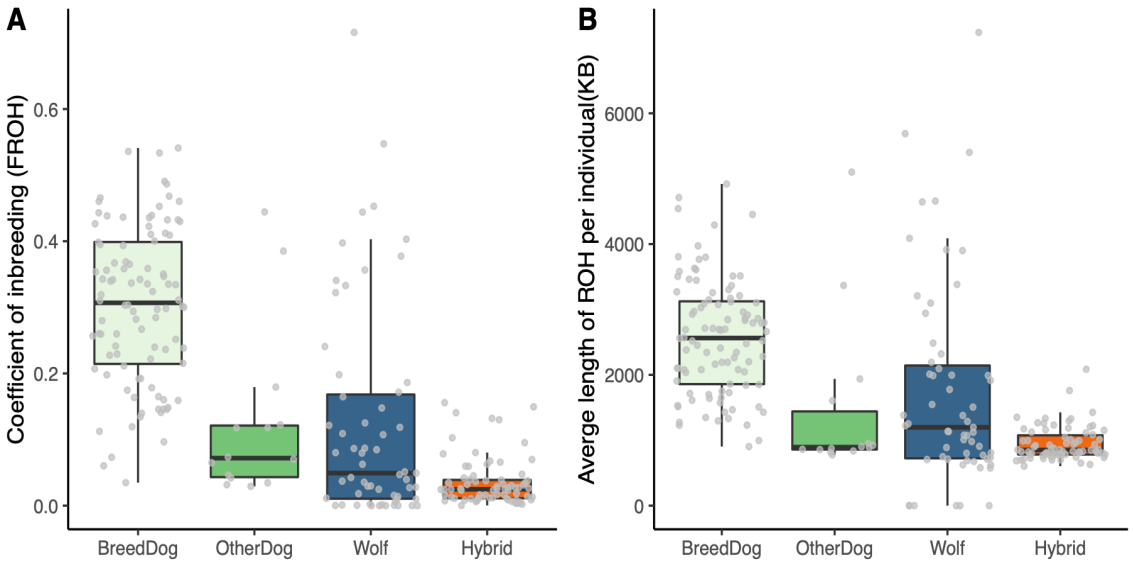


Figure 3.3 Runs of homozygosity and inbreeding coefficients for hybrids and ancestral populations. Hybrid population includes 74 wolfdogs and 2 coydogs;, dog population includes 96 breed dogs and 14 non-breed dogs (village dog and mixed-breed dogs);, and wolf population includes 33 North American wolves and 26 Eurasian wolves.

Kinship among hybrids doesn't indicate high relatedness (Table 3.4). Because the hybrids in our analysis reside at a small number of sanctuaries, it is possible that they are more related to one another than are hybrids overall. If so, this could influence our association analysis (discussed in chapter IV). To examine the kinship among hybrids, we estimated the kinship coefficient using 12,004,470 biallelic autosomal SNPs for a total of 74 wolfdog hybrids. Based on the coefficient values (k), the relationships were classified as unrelated (k<=0), second-degree (0.125<=k<0.25), or first-degree (k>=0.25). Out of the 2701 pairs, the majority (85.4%) were unrelated. The average kinship scores among the 393 related pairs were 0.03 (SD: 0.04). Among the related pairs, 21 unique wolfdog hybrids in 15 pairs exceeded the second-degree kinship, and 2 unique wolfdog hybrids in one pair exceeded the first-degree kinship.

Table 3.4. Pairs of hybrids that exceed second-degree kinship.

ID1	ID2	Kinship Coefficient	Relationship
WD069	WD025	0.367	First-degree
WD047	WD037	0.230	Second-degree
WD070	WD045	0.224	Second-degree
WD036	WD012	0.215	Second-degree
WD039	WD012	0.207	Second-degree
WD039	WD036	0.207	Second-degree
WD045	WD018	0.204	Second-degree

WD043	WD042	0.202	Second-degree
WD062	WD049	0.201	Second-degree
WD070	WD018	0.199	Second-degree
WD023	WD014	0.193	Second-degree
WD068	WD066	0.148	Second-degree
WD036	WD001	0.131	Second-degree
WD039	WD001	0.127	Second-degree
WD038	WD032	0.127	Second-degree

III.4.3 Global ancestral calling and breed composition in wolf-dog hybrids

Supervised admixture analysis confirms the wolfdogs are predominantly a mix of widely varying proportions of dog and North American wolf ancestry. We measured wolf, coyote, and dog breed ancestry using supervised admixture analysis and a `global ancestry calling panel` (Table 3.10) of 2163 reference canines from 115 populations: 108 modern breeds with at least 4 dogs per breed, 11 Greenland sled dogs, 3 regional village dog populations (4 Nigerian village dogs, 5 Vietnamese village dogs, 55 Chinese village dogs), 2 wolf populations (19 North American wolves and 25 Eurasian wolves), and 3 coyotes. This approach reports the ancestries and the proportion for ancestries for each hybrid, in addition to the ancestries in the reference panel, this approach also reports

another category “not detected” in the situation that it cannot assign an ancestry to the hybrid with confidence. On average, this approach reported an average of 5.99% “not detected” for 76 hybrids (Figure 3.4C).

The two wolfdogs with apparent coyote ancestry in the principle component analysis are predominantly coyotes (94% and 100%) (Figure 3.4A). Prior to genetic analysis, both animals were suspected to have wolf and dog ancestry, highlighting the difficulty of discerning ancestry without genetic data. After excluding the two high-content coyotes, 64% of detected wolfdog hybrid ancestry is wolf (97% of which is North American wolf), 35.5% is dog and <0.05% is coyote (Figure 3.4C; Table 3.5). The percentage of wolf ancestry varies widely, with individual wolfdogs carrying between 9% and 98% wolf ancestry (N=74, median=61%, interquartile range=44%) (Table 3.5). Most wolfdogs (57 out of 74) have no coyote ancestry and none has more than 5.3% coyote (mean=0.5%±1.0% (±SD)).

Dog breed ancestry in wolfdogs matches predominantly to sled dogs and other breeds with some physical characteristics popularly considered wolf-like, such as upright ears, dense, fluffy coats, and large size. Six breeds contribute more than 5% ancestry to individual wolfdogs: Alaskan malamute (54 wolfdogs), Siberian husky (39 wolfdogs), German shepherd (36 wolfdogs), Greenland sled dog (5 dogs), Great Pyrenees (4 wolfdogs) and Samoyed (2 wolfdogs) (Table 3.5).

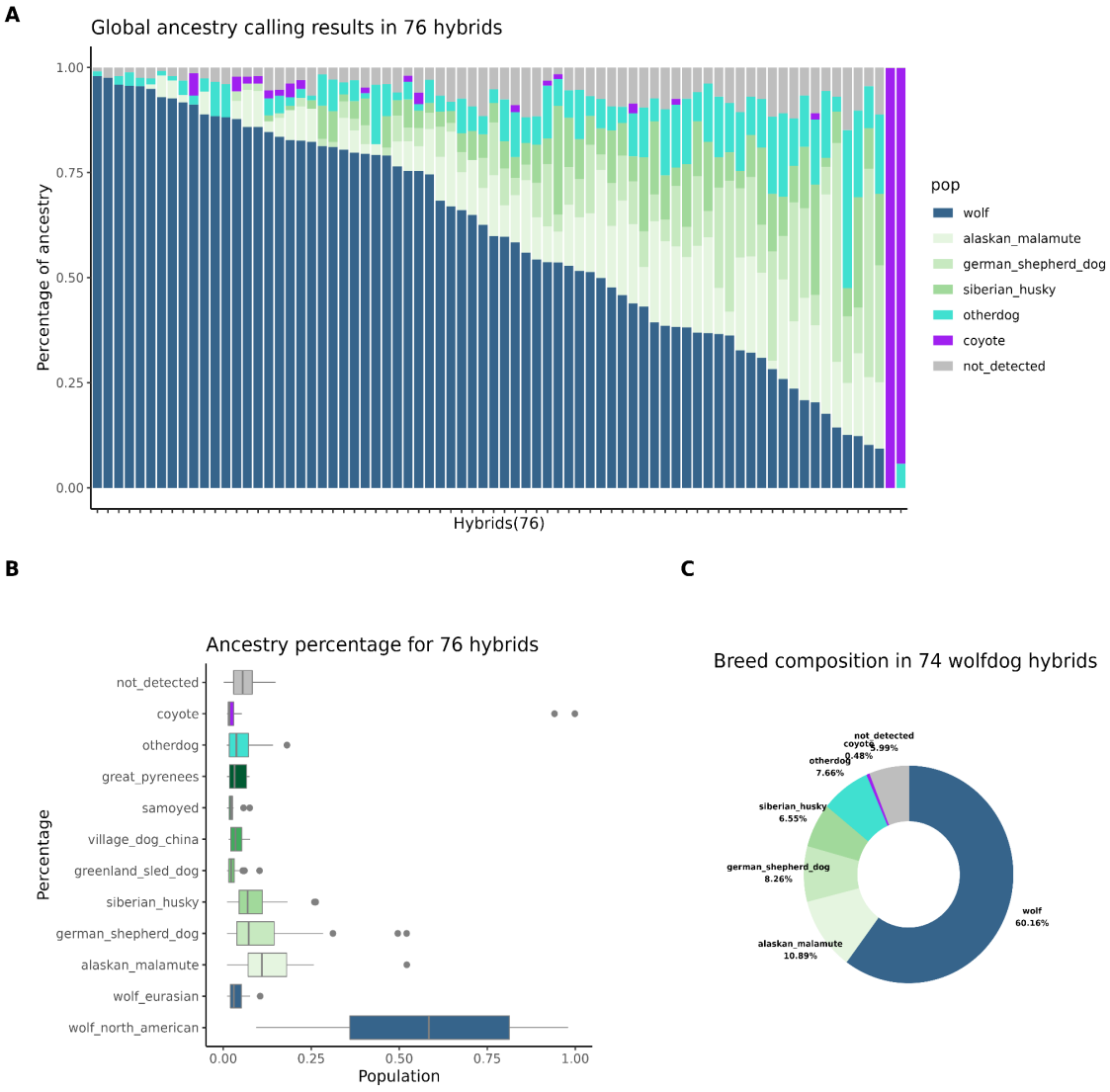


Figure 3.4. Global ancestry calling for 74 wolfdogs and 2 coydogs. (A) Ancestry composition per hybrid, two coyote-dogs are included and represented by purple. (B) Ancestry percentage distribution per ancestral population, 76 hybrids are included. (C) Breed composition for 74 wolf-dog hybrids.

Table 3.5. Global ancestry calling results for 74 wolfdogs and 2 coydogs. (NA: North American; EA: Eurasian; AM: Alaskan Malamute; GSheD: German Shepherd Dog; SH: Siberian Husky; GP: Great Pyrenees; GSledD: Greenland Sleddog; SA: Samoyed; Coy: Coyote)

ID	Pop	Wolf (NA)	Wolf (EA)	AM	GSheD	SH	GP	GSledD	SA	Coy
WD001	Wolfdog	83%	0%	8%	2%	0%	0%	0%	0%	2%
WD002	CoyDog	0%	0%	0%	0%	0%	0%	0%	0%	94%
WD003	Wolfdog	56%	3%	8%	9%	6%	0%	1%	0%	0%
WD004	Wolfdog	52%	2%	7%	4%	26%	0%	2%	0%	1%
WD005	Wolfdog	95%	0%	1%	0%	0%	0%	0%	0%	0%
WD006	Wolfdog	57%	3%	14%	13%	5%	0%	1%	0%	0%
WD007	Wolfdog	69%	5%	11%	4%	2%	0%	2%	0%	0%
WD008	Wolfdog	43%	0%	9%	8%	18%	0%	2%	0%	0%
WD009	Wolfdog	32%	1%	22%	16%	4%	0%	3%	3%	0%
WD010	Wolfdog	28%	4%	23%	19%	10%	2%	1%	0%	0%
WD011	Wolfdog	79%	0%	0%	2%	7%	4%	0%	0%	0%
WD012	Wolfdog	88%	0%	4%	2%	0%	0%	0%	0%	3%
WD013	Wolfdog	17%	4%	24%	21%	15%	0%	2%	0%	0%
WD014	Wolfdog	14%	0%	16%	52%	7%	0%	0%	0%	0%
WD015	Wolfdog	36%	1%	23%	6%	16%	0%	5%	2%	0%
WD016	Wolfdog	34%	2%	6%	31%	7%	0%	0%	0%	0%
WD017	Wolfdog	26%	2%	2%	26%	14%	0%	0%	2%	0%
WD018	Wolfdog	89%	0%	5%	0%	0%	2%	0%	0%	0%

WD019	Wolfdog	13%	0%	12%	16%	7%	0%	6%	6%	0%
WD020	Wolfdog	60%	3%	9%	7%	6%	0%	1%	2%	0%
WD021	Wolfdog	56%	0%	17%	4%	5%	0%	2%	0%	0%
WD022	Wolfdog	21%	3%	16%	28%	9%	0%	0%	0%	0%
WD023	Wolfdog	10%	0%	16%	50%	10%	0%	3%	0%	0%
WD024	Wolfdog	20%	0%	18%	7%	26%	0%	2%	3%	1%
WD025	Wolfdog	88%	0%	0%	0%	0%	7%	0%	0%	0%
WD026	Wolfdog	82%	0%	8%	0%	2%	0%	0%	0%	0%
WD027	Wolfdog	57%	2%	11%	4%	5%	0%	0%	8%	2%
WD028	Wolfdog	69%	10%	3%	0%	0%	0%	0%	0%	0%
WD029	Wolfdog	81%	1%	7%	0%	1%	0%	2%	0%	3%
WD030	Wolfdog	26%	0%	24%	10%	9%	1%	5%	0%	0%
WD031	Wolfdog	37%	2%	24%	6%	17%	0%	1%	0%	0%
WD032	Wolfdog	81%	0%	0%	1%	7%	8%	0%	0%	0%
WD033	Wolfdog	73%	6%	2%	4%	6%	0%	0%	0%	1%
WD034	Wolfdog	93%	3%	0%	0%	0%	0%	0%	0%	0%
WD035	Wolfdog	78%	7%	2%	0%	1%	0%	0%	0%	2%
WD036	Wolfdog	86%	0%	9%	2%	0%	0%	0%	0%	2%
WD037	Wolfdog	44%	3%	15%	10%	13%	0%	1%	0%	0%
WD038	Wolfdog	81%	0%	0%	2%	8%	6%	0%	0%	0%
WD039	Wolfdog	86%	0%	9%	2%	0%	0%	0%	0%	2%
WD040	Wolfdog	96%	0%	0%	0%	0%	0%	0%	0%	0%
WD041	Wolfdog	77%	3%	5%	3%	4%	0%	0%	0%	0%
WD042	Wolfdog	98%	0%	0%	0%	0%	0%	0%	0%	0%

WD043	Wolfdog	98%	0%	0%	0%	0%	0%	0%	0%	0%
WD044	Wolfdog	60%	5%	13%	0%	7%	0%	2%	0%	0%
WD045	Wolfdog	93%	0%	5%	0%	0%	0%	0%	0%	0%
WD046	Wolfdog	9%	0%	16%	28%	17%	0%	6%	0%	0%
WD047	Wolfdog	48%	4%	13%	9%	9%	0%	2%	0%	0%
WD048	Wolfdog	77%	6%	4%	1%	0%	0%	0%	0%	1%
WD049	Wolfdog	61%	6%	18%	2%	3%	0%	1%	0%	0%
WD050	Wolfdog	37%	1%	17%	7%	15%	0%	5%	0%	0%
WD051	Wolfdog	42%	4%	23%	11%	8%	0%	3%	0%	0%
WD052	Wolfdog	35%	4%	19%	14%	4%	0%	2%	2%	1%
WD053	Wolfdog	51%	3%	8%	15%	6%	2%	0%	1%	1%
WD054	Wolfdog	51%	3%	10%	8%	9%	0%	2%	3%	0%
WD055	Wolfdog	44%	0%	19%	10%	6%	0%	0%	3%	2%
WD056	Wolfdog	18%	0%	52%	7%	2%	0%	10%	0%	0%
WD057	Wolfdog	73%	2%	6%	4%	4%	0%	0%	0%	3%
WD058	Wolfdog	12%	0%	20%	11%	26%	0%	6%	0%	0%
WD059	Wolfdog	29%	1%	20%	11%	18%	0%	4%	0%	0%
WD060	Wolfdog	31%	6%	11%	24%	7%	0%	2%	2%	0%
WD061	Wolfdog	51%	0%	14%	16%	7%	0%	0%	0%	0%
WD062	Wolfdog	44%	6%	26%	4%	6%	0%	3%	1%	0%
WD063	Wolfdog	51%	1%	15%	9%	12%	0%	0%	1%	0%
WD064	Wolfdog	86%	5%	0%	0%	0%	0%	0%	0%	5%
WD065	Wolfdog	37%	2%	20%	5%	11%	0%	3%	2%	0%
WD066	Wolfdog	76%	0%	6%	6%	4%	0%	0%	0%	0%

WD067	Wolfdog	63%	3%	9%	5%	4%	0%	0%	0%	0%
WD068	Wolfdog	80%	0%	8%	4%	2%	0%	0%	0%	0%
WD069	Wolfdog	88%	0%	0%	0%	0%	7%	0%	0%	0%
WD070	Wolfdog	93%	0%	4%	0%	0%	0%	0%	0%	0%
WD071	Wolfdog	96%	0%	0%	0%	0%	2%	0%	0%	0%
WD072	Wolfdog	75%	0%	7%	3%	7%	0%	0%	0%	1%
WD073	Wolfdog	89%	3%	2%	0%	0%	0%	0%	0%	0%
WD074	Wolfdog	61%	8%	10%	7%	4%	0%	0%	3%	0%
WD075	Wolfdog	31%	6%	21%	17%	13%	0%	3%	2%	0%
WD076	CoyDog	0%	0%	0%	0%	0%	0%	0%	0%	100%

III.4.4 Local ancestral calling in wolf-dog hybrids

We modified a Hidden Markov Model-based approach to assign ancestral state to genome blocks as either homozygous dog, homozygous wolf, or heterozygous wolf and dog. This approach was evaluated using simulated hybrids and can achieve an accuracy of 0.95 (SD:0.05). Among 74 wolfdog hybrids, the average wolf content estimated using the local ancestry inference approach was 63.3%, which is highly consistent with the estimate from the global ancestry calling approach. The latter method detected 60% wolf ancestry in the entire genome and with a 5% genome undetected in the hybrid population. If we only consider the genomes that are successfully called for ancestry, 63.8% of the successfully

called genome is wolf ancestry. It has a correlation coefficient of 0.997 with the wolf ancestry called from local ancestry inference approach (P-value < 2.2e-16) between the wolf content called by these two approaches (Figure 3.5A).

For each hybrid, we calculated ancestral switches, median, and maximum ancestral block lengths. The longest ancestral block, which derived from dog ancestry, was 102Mb and covered the entire chromosome 1. To control for false ancestral state switches, we smoothed short blocks less than 1Mb and assigned them the same ancestry status as their neighboring blocks. Thus the shortest ancestral block was at least 1Mb.

The number of switches was negatively correlated with the amount of wolf content, while the median ancestral block size was positively correlated , suggesting that hybrids with more dog content underwent more generations of admixture (Figure 3.5B,C). This finding was further supported by the relationship between ancestral states and individual blocks of all hybrids (Figure 3.5D). The size of ancestral blocks was highly dependent on ancestral status, with the "Wolf" state displaying longer blocks (median: 8.5Mb) than the "Dog" state (3.8Mb) or the "Heterozygous of wolf and dog" state (4.3Mb).

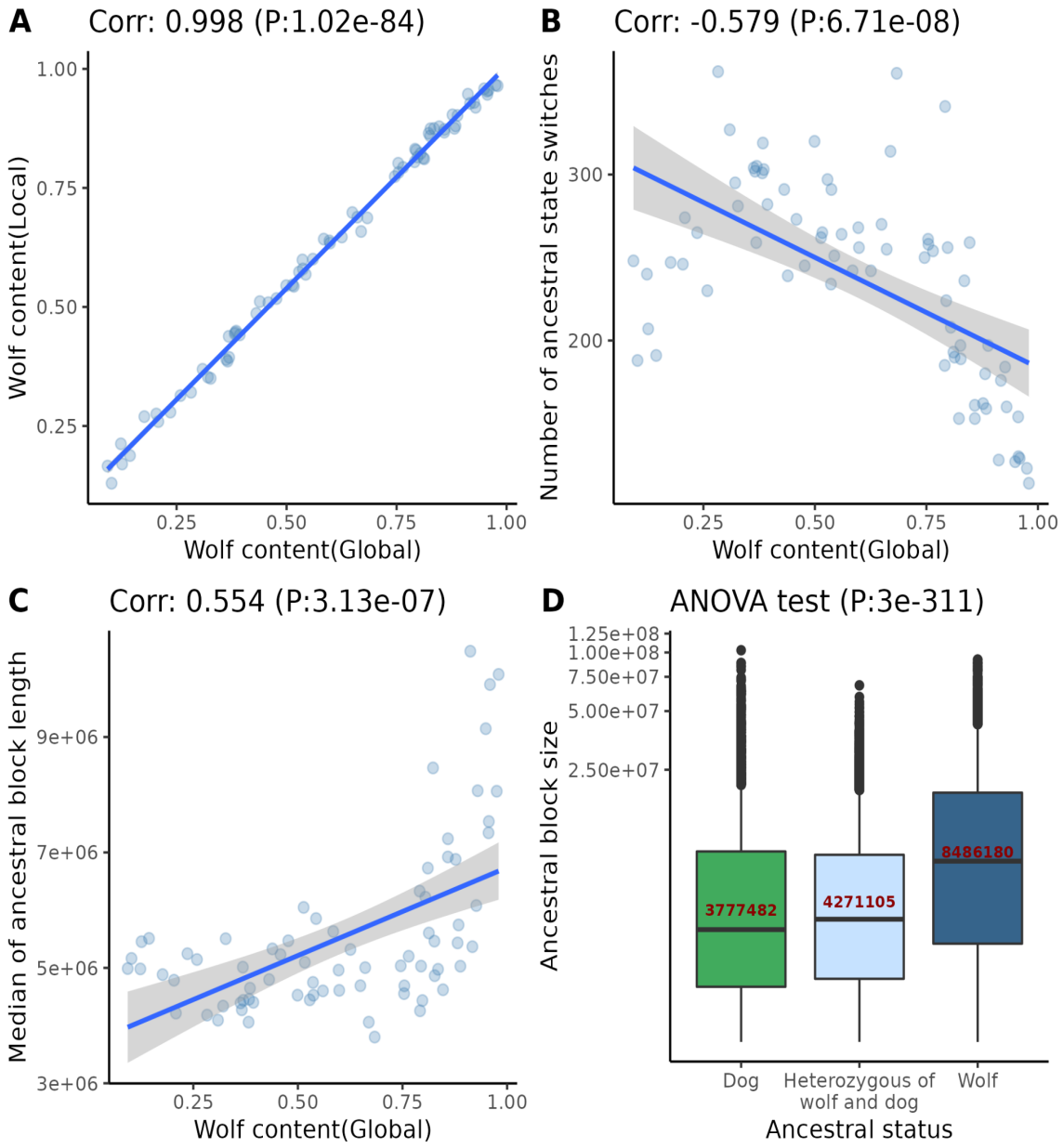


Figure 3.5. Number of ancestral state switches and ancestral block size highly depend on wolf content. (A) Comparing wolf contents estimated from global and local approaches. (B) Number of ancestral state switches versus wolf content. (C) Median of ancestral block length versus wolf content. (D) Individual block size distributions among three ancestral states: dog, wolf, and heterozygous of wolf and dog.

Furthermore, the observation of a negative correlation between wolf content and inbreeding coefficient or average length of runs of homozygosity supports the notion that the dog population is more inbred than the wolf population (Figure 3.6).

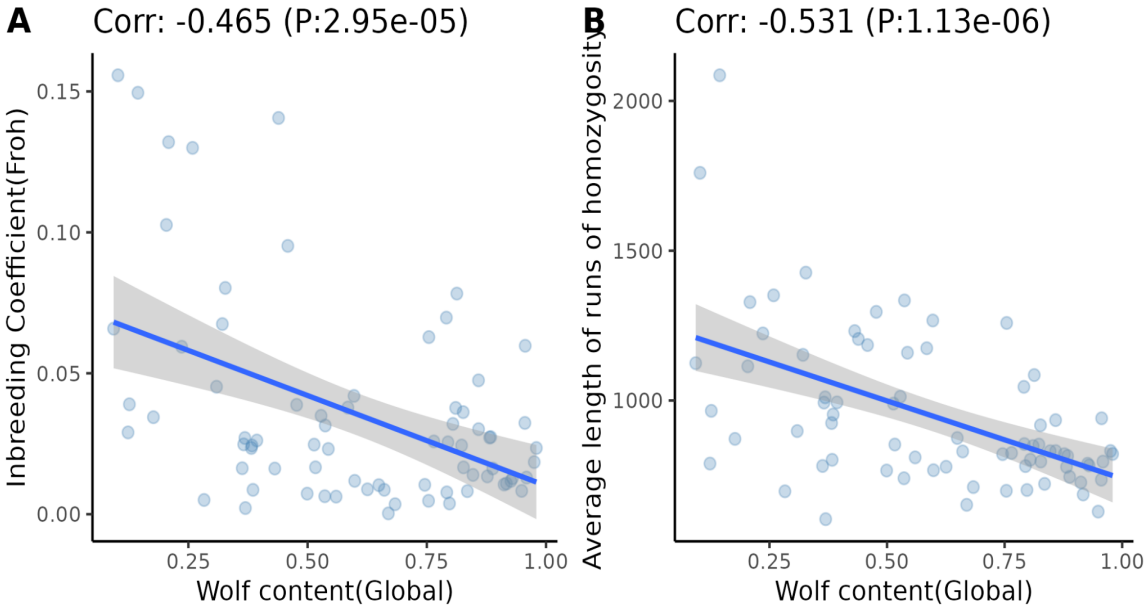


Figure 3.6. High wolf-content hybrids are more inbred. (A) Inbreeding coefficient negatively correlated with wolf content. (B) Average length of runs of homozygosity is negatively correlated with wolf content.

III.4.5 Identification and investigation of regions over-represented by wolf ancestry

We computed the per-site wolf ancestry for 74 wolfdog hybrids and identified "wolf over-represented" regions as those with per-site wolf ancestry greater than the 99% quantile (77.6%) and "dog over-represented" regions as those with per-site wolf ancestry lower than the 1% quantile (50.7%) (Figure 3.7). We identified 9 "wolf over-represented" regions comprising 645 genes with an average size of 4.6Mb (SD: 8.9Mb). There were 12 "dog over-represented" regions including 218 genes with an average size of 2.8Mb (SD: 4.3Mb) (Table 3.6).

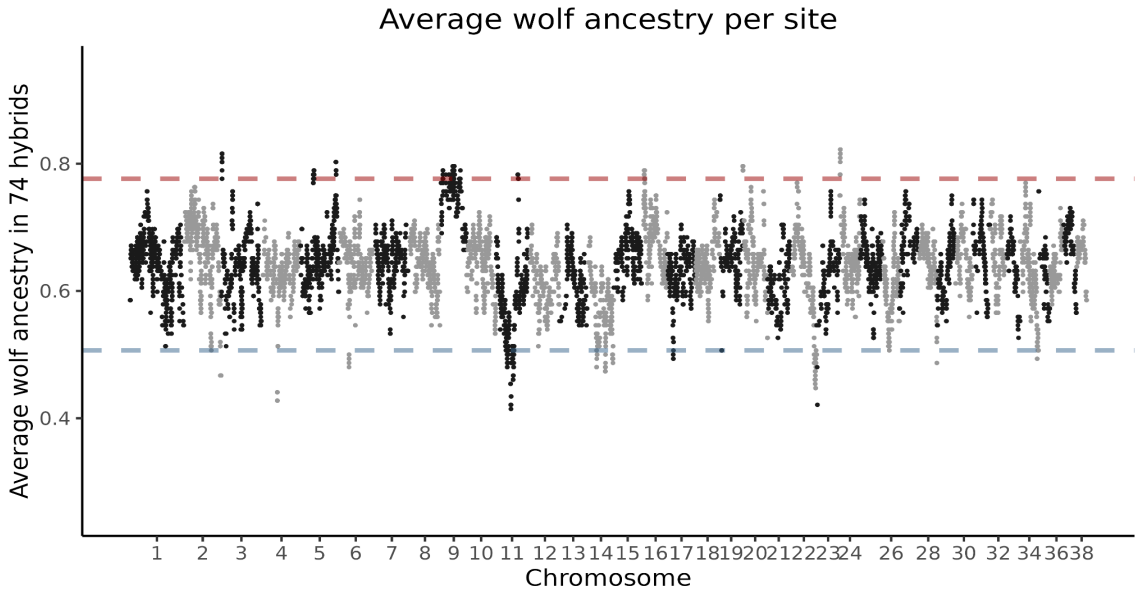


Figure 3.7 Manhattan plot for per-site wolf ancestry on 25,640 ancestry informative markers (AIMs). 99% quantile is labeled by red dashed line and 1% quantile is labeled by blue dashed line. Points above red line represents “wolf overrepresented” and points below blue line represents “dog overrepresented”

Table 3.6. Wolf and dog over-represented regions in 74 wolfdog hybrids.

Chr	Start	End	Num of genes	Wolf ancestry	Group
3	2166769	4241529	4	80.3%	Wolf-represented
5	30366673	31882942	49	78.6%	Wolf-represented
5	81644680	82810046	49	79.3%	Wolf-represented
9	6937488	35168932	425	78.3%	Wolf-represented
9	46838726	47686283	23	78.9%	Wolf-represented
11	50447326	51339913	23	78.3%	Wolf-represented
16	5882497	6964479	24	78.6%	Wolf-represented
20	1	3747832	39	79.3%	Wolf-represented
24	1278095	3183784	9	79.9%	Wolf-represented
2	83273638	85000867	35	46.7%	Dog-represented
4	36385220	37641327	22	43.4%	Dog-represented
6	23066031	24270941	17	49.0%	Dog-represented
11	26217529	41709532	52	50.0%	Dog-represented
14	21712276	23534860	12	49.0%	Dog-represented
14	41565117	42087340	3	48.7%	Dog-represented
14	57660895	59236066	4	49.7%	Dog-represented
17	13085486	13479035	5	49.7%	Dog-represented
22	55600429	61424290	39	50.0%	Dog-represented
23	1035130	4319886	29	45.1%	Dog-represented
28	39978530	39983423	0	48.7%	Dog-represented

34	38827333	39010618	0	49.3%	Dog-represented

Gene ontology analysis showed that regions over-represented in wolves are significantly enriched in genes responsible for muscle building, such as intermediate filament organization (GO:0045109) (FDR: 1.9e-10), as well as in genes related to skeleton and tissue development, such as the pathway for skeletal system morphogenesis (GO:0048705) (FDR: 2.8e-2) and tissue development (GO:0009888) (FDR: 2.9e-2) (Table 3.7).

Regions overrepresented in wolf and dogs show gene enrichment seems possible given the history of wolfdog. Breeders have intentionally bred hybrids to possess physical characteristics resembling wolves and behaviors resembling dogs (Muller 2021). Therefore, it is not surprising that the hybrid cohort inherits genes responsible for muscle, skeleton, and tissue development predominantly from wolves. The gene ontology analysis did not detect any significantly enriched pathways in regions over-represented in dogs; however, the top 20 identified pathways were enriched in immune-related genes (Table 3.8). One plausible explanation for this is that hybrids live in captive environments where the population density is higher than for their wild wolf ancestors. The acquisition of dog immune genes may benefit individuals living in captivity.

Table 3.7. GO ontology for genes in the wolf-represented regions.

GO biological process	Number of genes in term	Observed	Expected	Fold Enrichment	Raw P-value	FDR
intermediate filament organization (GO:0045109)	70	22	2.07	10.61	1.21E-14	1.90E-10
intermediate filament-based process (GO:0045103)	93	22	2.76	7.98	1.59E-12	8.32E-09
intermediate filament cytoskeleton organization (GO:0045104)	92	22	2.73	8.07	1.32E-12	1.04E-08
epithelium development (GO:0060429)	1072	64	31.76	2.02	3.37E-07	7.56E-04
epithelial cell differentiation (GO:0030855)	620	43	18.37	2.34	7.35E-07	1.44E-03
supramolecular fiber organization (GO:0097435)	580	41	17.18	2.39	9.85E-07	1.72E-03
positive regulation of macromolecule biosynthetic process (GO:0010557)	1943	96	57.57	1.67	1.27E-06	2.00E-03
positive regulation of DNA-templated transcription (GO:0045893)	1715	84	50.81	1.65	8.70E-06	1.14E-02

positive regulation of nucleic acid-templated transcription (GO:1903508)	1715	84	50.81	1.65	8.70E-06	1.24E-02
positive regulation of RNA biosynthetic process (GO:1902680)	1721	84	50.99	1.65	1.20E-05	1.45E-02
positive regulation of biosynthetic process (GO:0009891)	2087	97	61.83	1.57	1.37E-05	1.54E-02
positive regulation of RNA metabolic process (GO:0051254)	1848	88	54.75	1.61	1.65E-05	1.73E-02
embryonic skeletal system development (GO:0048706)	129	15	3.82	3.92	1.80E-05	1.77E-02
positive regulation of cellular biosynthetic process (GO:0031328)	2048	95	60.68	1.57	2.12E-05	1.96E-02
skeletal system morphogenesis (GO:0048705)	228	20	6.76	2.96	3.59E-05	2.82E-02
tissue development (GO:0009888)	1726	82	51.14	1.6	3.46E-05	2.85E-02
embryo development ending in birth or egg hatching (GO:0009792)	677	41	20.06	2.04	3.28E-05	2.86E-02
chordate embryonic development (GO:0043009)	655	40	19.41	2.06	3.91E-05	2.92E-02

chromosome segregation (GO:0007059)	291	23	8.62	2.67	4.40E-05	3.13E-02
embryonic skeletal system morphogenesis (GO:0048704)	94	12	2.78	4.31	5.42E-05	3.69E-02

Table 3.8 GO ontology for genes in the dog-represented regions.

GO biological process	Number of genes in term	Observed	Expected	Fold Enrichment	Raw P-value	FDR
RNA destabilization (GO:0050779)	78	6	0.78	7.73	1.83E-04	3.59E-01
natural killer cell activation involved in immune response (GO:0002323)	25	4	0.25	16.07	1.80E-04	4.02E-01
immune system development (GO:0002520)	757	19	7.54	2.52	2.57E-04	4.03E-01
B cell proliferation (GO:0042100)	52	5	0.52	9.66	2.45E-04	4.27E-01
cell activation involved in immune response (GO:0002263)	192	9	1.91	4.71	1.78E-04	4.65E-01
SCF-dependent proteasomal ubiquitin-dependent protein catabolic process (GO:0031146)	48	5	0.48	10.46	1.73E-04	5.43E-01
lymphocyte activation involved in immune response (GO:0002285)	128	7	1.27	5.49	3.90E-04	5.56E-01

leukocyte activation involved in immune response (GO:0002366)	188	9	1.87	4.81	1.53E-04	6.00E-01
regulation of peptidyl-serine phosphorylation of STAT protein (GO:0033139)	23	4	0.23	17.47	1.35E-04	7.04E-01
regulation of mitotic cell cycle phase transition (GO:1901990)	332	11	3.31	3.33	6.27E-04	8.20E-01
interneuron migration (GO:1904936)	15	3	0.15	20.09	6.91E-04	8.33E-01
negative regulation of systemic arterial blood pressure (GO:0003085)	22	3	0.22	13.7	1.85E-03	8.54E-01
response to nitrogen compound (GO:1901698)	1062	22	10.57	2.08	1.28E-03	8.74E-01
aspartate family amino acid metabolic process (GO:0009066)	49	4	0.49	8.2	1.84E-03	8.75E-01
brush border assembly (GO:1904970)	5	2	0.05	40.17	1.97E-03	8.81E-01
leukocyte proliferation (GO:0070661)	145	7	1.44	4.85	7.92E-04	8.87E-01
aspartate family amino acid biosynthetic process (GO:0009067)	21	3	0.21	14.35	1.64E-03	8.87E-01
lymphocyte proliferation (GO:0046651)	123	6	1.22	4.9	1.78E-03	8.99E-01
response to exogenous dsRNA (GO:0043330)	49	4	0.49	8.2	1.84E-03	9.02E-01
mononuclear cell proliferation (GO:0032943)	127	6	1.26	4.74	2.07E-03	9.04E-01

III.5 Discussion

III.5.1 Potential selection on captive wolfdog hybrids population

Wolfdog hybrids emerged through intentional breeding in the 1950s. The goal to create a breed that exhibits wolf-like physical features alongside a dog-like temperament has remained consistent. Over time, various combinations of wolves, dogs, and wolfdogs have been bred. Certain traits reach high frequency in these hybrid populations due to breeder preferences or captive living conditions. We propose that wolfdog hybrids in captivity favors wolf-like appearances, dense fur, long and lean legs, erect ears, and a gradual muzzle stop. Our research later in chapter IV reveals that, on average, these hybrids have appearances more similar to wolves, possessing an average of 14 out of 18 wolf-version morphological traits (Figure 4.1). The genes prevalent in wolf-overrepresented regions are enriched in tissue, skeletal, and muscle development, suggesting that these wolf-derived traits are favored in wolfdog hybrid populations (Table 3.7).

Considering the captive environment, higher population density and frequent human interaction might introduce pathogens that cause infectious diseases in the animals living there. We hypothesized that wolfdog hybrids are likely to carry immune-related genes from their dog ancestry since dogs share environments with humans and are continually exposed to various pathogens. My findings are consistent with this prediction. While not statistically significant, the top 20

identified pathways in regions over-represented in dogs are enriched with immune-related genes, including natural killer cell activation and immune system development pathways (Table 3.8). However, additional tests should be conducted to further explore the selection of immune genes in captive populations. Wolfdog hybrids are rarely found in the wild, but we can compare captive wolves to their wild counterparts to determine if the same regions are present. Additionally, we can examine other species, such as foxes, which have both domesticated and wild populations, for further insights.

III.5.2 Benefits of adding more North American wolves in imputation panel

Wolfdog hybrids have been identified as a combination of domesticated dogs and North American wolves. Our imputation panel includes 29 wolves, but only 6 of them are North American wolves. From the exploratory principal component analysis, we found that North American wolves and Eurasian wolves are genetically close to each other with Fst 0.066 (Figure 3.1). However, population differentiation may exist between the two populations due to geographical distance. The set of six North American wolves does not adequately represent the North American wolf population, and as a result, the genomic variants specific to these wolves cannot be captured or imputed in the wolfdog hybrids. To enhance our imputation panel, I propose to incorporate more North American

wolves to the imputation panel which will allow us to identify those unique variants within the population that may hold significant functional importance.

III.6 **Materials and Methods**

III.6.1 Ethical note

This research was approved by the Broad Institute of Harvard MIT and the University of Massachusetts Medical School. Written consent was provided by the owners of each location before research was conducted. All subjects included in this study were up to date on vaccinations and housed legally with all state required permits or licenses.

III.6.2 Data collection and genotyping

Genomic DNA extraction from saliva and fur samples was carried out through either in-house procedures or by Gencove (Gencove, Inc., New York, NY) or Neogen (Neogen Corporation - GeneSeek Operations, Lincoln, NE) as previously described (Morrill et al. 2022). A minimum concentration of 10ng/ul was required for low coverage sequencing (1X). 76 hybrid individuals (Table 3.1) were sequenced with this procedure at depths ranging from 0.5x to 1.1x. Sequencing reads were processed into imputed autosomal variant calls through Gencove's *loimpute* software (Wasik et al. 2021) using the copying model

described in Li and Stephens (Li and Stephens 2003). It utilizes an `imputation haplotype reference panel` (Table 3.10) that compares the data to 435 whole genomes (Morrill et al. 2022) from 287 pedigreed dogs of various breeds, 6 dogs of unknown ancestry, 100 indigenous village dogs worldwide, 36 wolves, and 6 other wild canids. This panel includes 46,349,043 autosomal variants, including 32,438,672 SNPs and 13,910,371 indels.

III.6.3 Construction of a diverse canine reference panel

We constructed a `diverse canine reference panel` (Table 3.9, Table 3.10) representing three species: domestic dogs, wolves, and coyotes. This dataset came from two sources: (1) 139 dogs and wolves are from the `imputation haplotype reference panel` with 435 whole genomes (Morrill et al. 2022). The 139 samples include 97 purebred dogs from 95 breeds, 10 village dogs, 1 unknown breed ancestry, 2 greenland sled dogs, 1 Alaskan sled dog, 3 coyotes, and 29 wolves (6 North American wolves and 23 Eurasian wolves). (2) 33 published whole-genome sequences including 27 North American wolves, 3 Eurasian wolves, and 3 coyotes, reads were aligned to *Canis lupus familiaris* (dog) reference genome (CanFam3.1) using BWA-MEM v0.7.17 (Li and Durbin 2009). We used GATK (v4.0.4) HaplotypeCaller to call variants for the 33 canids against the `imputation haplotype reference panel`. After calling, we merged the 33 samples with the 139 samples to construct a `diverse canine reference panel`,

filtered out sites with a missing rate greater than 5%. The final dataset includes 172 canids with 4,306,664 SNPs and 446,129 INDELs, with an average 1.97 variants per kb, with coverage ranges from 4x to 73x.

Table 3.9. Samples in diverse reference panel

SRA BioSample	Group	Breed or Population	Source
SAMEA1521941	Breed dog	English Cocker Spaniel	In-house VCF panel
SAMN03323675	Breed dog	Nova Scotia Duck Tolling Retriever	In-house VCF panel
SAMEA3723573	Breed dog	Petit Basset Griffon Vendeen	In-house VCF panel
SAMN02194722	Breed dog	Afghan Hound	In-house VCF panel
SAMN03580381	Breed dog	Airedale Terrier	In-house VCF panel
SAMN04196850	Breed dog	Airedale Terrier	In-house VCF panel
SAMEA3449656	Breed dog	Alaskan Husky	In-house VCF panel
SAMN08872810	Breed dog	Alaskan Malamute	In-house VCF panel
SAMN08872960	Breed dog	American Cocker Spaniel	In-house VCF panel
SAMN08872813	Breed dog	American Hairless Terrier	In-house VCF panel
SAMN03323677	Breed dog	Australian Cattle Dog	In-house VCF panel
SAMN04196845	Breed dog	Basenji	In-house VCF panel
SAMN08872826	Breed dog	Beagle	In-house VCF panel
SAMEA3928140	Breed dog	Bearded Collie	In-house VCF panel
SAMEA4504823	Breed dog	Belgian Malinois	In-house VCF panel
SAMN03801646	Breed dog	Belgian Sheepdog	In-house VCF panel
SAMN08872849	Breed dog	Belgian Tervuren	In-house VCF panel

SAMEA4506887	Breed dog	Berger Blanc Suisse	In-house VCF panel
SAMN06159678	Breed dog	Berger Picard	In-house VCF panel
SAMN03801648	Breed dog	Bernese Mountain Dog	In-house VCF panel
SAMN04196853	Breed dog	Black and Tan Coonhound	In-house VCF panel
SAMN03323668	Breed dog	Black Russian Terrier	In-house VCF panel
SAMN03801649	Breed dog	Bloodhound	In-house VCF panel
SAMN01908107	Breed dog	Border Collie	In-house VCF panel
SAMN03580407	Breed dog	Border Terrier	In-house VCF panel
SAMN08872905	Breed dog	Borzoi	In-house VCF panel
SAMN08872906	Breed dog	Boston Terrier	In-house VCF panel
SAMN08872908	Breed dog	Bouvier des Flandres	In-house VCF panel
SAMN02921305	Breed dog	Boxer	In-house VCF panel
SAMN08872913	Breed dog	Bull Terrier	In-house VCF panel
SAMN03801656	Breed dog	Chihuahua	In-house VCF panel
SAMN03075611	Breed dog	Chinese Crested	In-house VCF panel
SAMN03580382	Breed dog	Chinook	In-house VCF panel
SAMN03801658	Breed dog	Chow Chow	In-house VCF panel
SAMN03801659	Breed dog	Clumber Spaniel	In-house VCF panel
SAMEA3121338	Breed dog	Dachshund	In-house VCF panel
SAMN08872976	Breed dog	Dalmatian	In-house VCF panel
SAMN03801662	Breed dog	Doberman Pinscher	In-house VCF panel
SAMN03801654	Breed dog	English Bulldog	In-house VCF panel
SAMN03580391	Breed dog	English Springer Spaniel	In-house VCF panel
SAMEA4504828	Breed dog	Entlebucher Mountain Dog	In-house VCF panel

SAMN03168391	Breed dog	Finnish Lapphund	In-house VCF panel
SAMN03801666	Breed dog	Flat-Coated Retriever	In-house VCF panel
SAMEA3928146	Breed dog	French Bulldog	In-house VCF panel
SAMN02585201	Breed dog	German Shepherd Dog	In-house VCF panel
SAMN03801668	Breed dog	Golden Retriever	In-house VCF panel
SAMN03801669	Breed dog	Great Dane	In-house VCF panel
SAMN03801670	Breed dog	Great Pyrenees	In-house VCF panel
SAMN08873062	Breed dog	Greater Swiss Mountain Dog	In-house VCF panel
SAMN03168381	Breed dog	Greenland Dog	In-house VCF panel
SAMN03801671	Breed dog	Greyhound	In-house VCF panel
SAMN08873112	Breed dog	Irish Setter	In-house VCF panel
SAMN08873115	Breed dog	Irish Water Spaniel	In-house VCF panel
SAMN03801673	Breed dog	Italian Greyhound	In-house VCF panel
SAMN03580384	Breed dog	Jack Russell Terrier	In-house VCF panel
SAMN03580387	Breed dog	Kerry Blue Terrier	In-house VCF panel
SAMEA4505491	Breed dog	Lagotto Romagnolo	In-house VCF panel
SAMEA3121328	Breed dog	Landseer	In-house VCF panel
SAMN03168384	Breed dog	Lapponian Herder	In-house VCF panel
SAMN03801664	Breed dog	Mastiff (English)	In-house VCF panel
SAMN08873175	Breed dog	Miniature Poodle	In-house VCF panel
SAMN03168382	Breed dog	Norwegian Elkhound	In-house VCF panel
SAMN03801676	Breed dog	Pekingese	In-house VCF panel
SAMN03145702	Breed dog	Pembroke Welsh Corgi	In-house VCF panel
SAMN03168386	Breed dog	Peruvian Inca Orchid	In-house VCF panel

SAMN03801665	Breed dog	Pointer (English)	In-house VCF panel
SAMEA4506892	Breed dog	Pomeranian	In-house VCF panel
SAMN03580388	Breed dog	Portuguese Podengo	In-house VCF panel
SAMN08873219	Breed dog	Portuguese Water Dog	In-house VCF panel
SAMN03801682	Breed dog	Pug	In-house VCF panel
SAMN03801683	Breed dog	Rhodesian Ridgeback	In-house VCF panel
SAMN03801684	Breed dog	Rottweiler	In-house VCF panel
SAMN03801685	Breed dog	Saint Bernard	In-house VCF panel
SAMN03801686	Breed dog	Saluki	In-house VCF panel
SAMN08873258	Breed dog	Samoyed	In-house VCF panel
SAMN03580401	Breed dog	Scottish Deerhound	In-house VCF panel
SAMN03801688	Breed dog	Scottish Terrier	In-house VCF panel
SAMN08873266	Breed dog	Shetland Sheepdog	In-house VCF panel
SAMN03801690	Breed dog	Siberian Husky	In-house VCF panel
SAMEA3928143	Breed dog	Sloughi	In-house VCF panel
SAMN03323670	Breed dog	Soft Coated Wheaten Terrier	In-house VCF panel
SAMN03168380	Breed dog	Spanish Galgo	In-house VCF panel
SAMEA3164479	Breed dog	Spanish Water Dog	In-house VCF panel
SAMN03801691	Breed dog	Standard Poodle	In-house VCF panel
SAMN03323676	Breed dog	Standard Schnauzer	In-house VCF panel
SAMN03168387	Breed dog	Swedish Lapphund	In-house VCF panel
SAMN01974493	Breed dog	Tibetan Mastiff	In-house VCF panel
SAMN03801692	Breed dog	Toy Poodle	In-house VCF panel
SAMN03801693	Breed dog	West Highland White Terrier	In-house VCF panel

SAMEA3928145	Breed dog	Yorkshire Terrier	In-house VCF panel
SAMEA4346711	Breed dog	Alpine Dachsbracke	In-house VCF panel
SAMN04908310	Breed dog	Border Collie	In-house VCF panel
SAMN08872926	Breed dog	Carolina Dog	In-house VCF panel
SAMN08873012	Breed dog	Fonni's Dog	In-house VCF panel
SAMN08873128	Breed dog	Keeshond	In-house VCF panel
SAMN08873139	Breed dog	Labrador Retriever	In-house VCF panel
SAMN04196846	Other Dog	mixed Breed	In-house VCF panel
SAMN14210383	Sled dog	Greenland sled dog	In-house VCF panel
SAMN14210375	Sled dog	Greenland sled dog	In-house VCF panel
TBD	Sled dog	Alaskan sled dog	In-house VCF panel
SAMN02585198	Village dog	unknown	In-house VCF panel
SAMN02585191	Village dog	Indigenous - China, Kunming	In-house VCF panel
SAMN03168352	Village dog	Indigenous - China	In-house VCF panel
SAMN03168373	Village dog	Indigenous - Nigeria	In-house VCF panel
SAMN03168368	Village dog	Indigenous - Vietnam	In-house VCF panel
SAMEA1521940	Village dog	Village Dog - Aboriginal Camp	In-house VCF panel
SAMN01974490	Village dog	Village Dog - China	In-house VCF panel
SAMN02485586	Village dog	Village Dog - India	In-house VCF panel
SAMN02485599	Village dog	Village Dog - Namibia	In-house VCF panel
SAMN02485604	Village dog	Village Dog - Portugal	In-house VCF panel
SAMN01974486	Eurasian wolf	Wolf (Altai, Russia)	In-house VCF panel
SAMN01974488	Eurasian wolf	Wolf (Bryansk, Russia)	In-house VCF panel
SAMN03366711	Eurasian wolf	Wolf (China)	In-house VCF panel

SAMN01974487	Eurasian wolf	Wolf (Chukotka, Russia)	In-house VCF panel
SAMN03366712	Eurasian wolf	Wolf (Croatia)	In-house VCF panel
SAMN02921311	Eurasian wolf	Wolf (India)	In-house VCF panel
SAMN03652997	Eurasian wolf	Wolf (Inner Mongolia, China)	In-house VCF panel
SAMN03653003	Eurasian wolf	Wolf (Inner Mongolia, China)	In-house VCF panel
SAMN02921312	Eurasian wolf	Wolf (Iran)	In-house VCF panel
SAMN03168394	Eurasian wolf	Wolf (LiaoNing, China)	In-house VCF panel
SAMN02921316	Eurasian wolf	Wolf (Portugal)	In-house VCF panel
SAMN03652998	Eurasian wolf	Wolf (Qinghai, China)	In-house VCF panel
SAMN03653001	Eurasian wolf	Wolf (Qinghai, China)	In-house VCF panel
SAMN03168396	Eurasian wolf	Wolf (Shanxi, China)	In-house VCF panel
SAMN02921319	Eurasian wolf	Wolf (Spain)	In-house VCF panel
SAMN03653004	Eurasian wolf	Wolf (Tibet, China)	In-house VCF panel
SAMN03652999	Eurasian wolf	Wolf (Tibet, China)	In-house VCF panel
SAMN03653002	Eurasian wolf	Wolf (Xinjiang, China)	In-house VCF panel
SAMN03653000	Eurasian wolf	Wolf (Xinjiang, China)	In-house VCF panel
SAMN03168397	Eurasian wolf	Wolf (Xinjiang, China)	In-house VCF panel
SAMN03168399	Eurasian wolf	Wolf (Xinjiang, China)	In-house VCF panel
SAMN03168398	Eurasian wolf	Wolf (Xinjiang, China)	In-house VCF panel
SAMN03168400	Eurasian wolf	Wolf	In-house VCF panel
SAMN04851099	Eurasian wolf	Wolf	Public Dataset
SAMN06219538	Eurasian wolf	Wolf	Public Dataset
SAMN06219539	Eurasian wolf	Wolf	Public Dataset

SAMN02921315	North American wolf	Wolf (Mexico, North America)	In-house VCF panel
SAMN02921310	North American wolf	Wolf (Great Lakes)	In-house VCF panel
SAMN02921314	North American wolf	Wolf (Mexico, North America)	In-house VCF panel
SAMN03801694	North American wolf	Wolf (Unknown Locale)	In-house VCF panel
SAMN02921320	North American wolf	Wolf (Yellowstone)	In-house VCF panel
SAMN02921321	North American wolf	Wolf (Yellowstone)	Public Dataset
SAMN02921322	North American wolf	Wolf (Yellowstone,)	Public Dataset
SAMN04209256	North American wolf	Wolf	In-house VCF panel
SAMN10662574	North American wolf	Wolf ArcticBaffin	Public Dataset
SAMN10662575	North American wolf	Wolf ArcticEllesmere	Public Dataset
SAMN10662576	North American wolf	Wolf ArcticNunavut	Public Dataset
SAMN10662573	North American wolf	WolfArcticVictoria	Public Dataset
SAMN10662590	North American wolf	WolfIsleRoyale1	Public Dataset
SAMN10662579	North American wolf	WolfIsleRoyale2	Public Dataset
SAMN10662591	North American wolf	Wolf minnesota	Public Dataset
SAMN10662585	North American wolf	Wolf minnesota	Public Dataset
SAMN10662577	North American wolf	Wolf Quebec	Public Dataset
SAMN10246085	North American wolf	Wolf Alaska	Public Dataset

SAMN10246087	North American wolf	Wolf BaffinNorth	Public Dataset
SAMN10246091	North American wolf	Wolf Pacific Coast	Public Dataset
SAMN10246093	North American wolf	Wolf saskatchewan	Public Dataset
SAMN10246088	North American wolf	Wolf BaffinSouth	Public Dataset
SAMN10246094	North American wolf	Wolf St Lawrence Island	Public Dataset
SAMN10246095	North American wolf	Wolf Toronto	Public Dataset
SAMN10246096	North American wolf	Wolf Victoria Island	Public Dataset
SAMN10246086	North American wolf	Wolf Atlantic Coast	Public Dataset
SAMN10246089	North American wolf	Wolf BanksIsland	Public Dataset
SAMN10246090	North American wolf	Wolf Ellesmere	Public Dataset
SAMN10246092	North American wolf	Wolf Qamanirjuaq	Public Dataset
SAMN10246098	North American wolf	Wolf AlgonquinPark	Public Dataset
SAMN10180430	North American wolf	Wolf greenland	Public Dataset
SAMN10180431	North American wolf	Wolf Ellesmere	Public Dataset
SAMN08005342	North American wolf	Wolf	Public Dataset
SAMN10180421	Coyote	Coyote Alaska	Public Dataset
SAMN10180422	Coyote	Coyote Mexico	Public Dataset
SAMN10180423	Coyote	Coyote Missouri	Public Dataset

III.6.3 Principal component analysis to explore relationships among populations

Principal component analysis (PCA) was used to explore the dominant relationships of the reference canids. PCA was implemented in Plink (`--pca`) on 2,412,489 common markers between the 172 reference samples and hybrids. We reported the top 10 principal components of the variance-standardized relationship matrix and calculated the variance of each principal component.

III.6.4 Fst estimation

Fst was estimated using vcftools v0.1.16 (`--fst`) between populations and species suggested by PCA, hybrids, domesticated dogs, Eurasian wolves, and North American wolves. The average Fst between any two populations is calculated as the mean of Fst for all 2,412,489 variants.

III.6.5 Homozygosity and inbreeding

We scanned runs of homozygosity across 2,412,489 variants in the hybrids and individuals from ancestral wolf and dog populations .Using PLINK 1.9 with settings: a minimum run length of 500kb (`--homozyg-kb 500`) at a density of 100kb per SNP (`--homozyg-density 100`) with no two SNPs more than 10kb apart (`--homozyg-gap 10`) and only 1 heterozygous genotypes tolerated per window (`--homozyg-window-het 1`) ((Ceballos et al. 2018). We then calculated

the autosomal ROH-estimated coefficient of inbreeding (FROH) from the total ROH segment length divided by the total SNP-covered length across autosomes (2,203,765 kilobases) where ROH detection was possible ((McQuillan et al. 2008; Sams and Boyko 2019).

III.6.6 Kinship and relatedness

Kinship among the 74 wolfdog hybrids were estimated using PLINK 1.9 (`--king`). The relationships were classified into three groups based on the coefficient values (k): unrelated (k<=0), second-degree (0.125<=k<0.25), or first-degree (k>=0.25).

III.6.7 Decay of linkage disequilibrium

Decay of linkage disequilibrium was estimated by calculating r2 between each of 20,000 randomly chosen SNPs and all variants with a minor allele frequency >0.025 within a 100kb range of the index SNPs (`−−ld−window−kb 105`, `−−ld−window−r2 0`, `−−maf 0.025`, `−−ld−window 5000`)

III.6.8 Global ancestry calling

We inferred unsupervised and supervised global ancestry proportions for 76 putative hybrids, including 2 hybrids with apparent coyote ancestry, 6 wolves sampled from two litters born in Canada, and 5 pet dogs from the Darwin's Ark project (darwinsark.org) with inferred wolf ancestry (Table 3.10).

First, we inferred optimal clusters using 5-fold cross validation of K= {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25} clusters on 403,074 biallelic SNPs from chromosome 38. We ran unsupervised admixture analysis on the query samples using 6,070,900 SNPs selected by LD-pruning pairs of r2>0.8 in 50 kb windows across all 32,438,672 autosomal biallelic SNPs for K= 2 and 3 clusters using ADMIXTURE (Alexander and Lange 2011) in unsupervised mode (random seed: 43) and 20 bootstrap replicates to estimate standard errors. As all Canadian wolf samples were inferred to have full assignment to clusters # (K=2) and # (K=3), we refer to these clusters as the “wolf cluster”, clusters # (K=2) and # (K=3) as the “dog cluster”, and cluster # (K=3) as the “coyote cluster”.

We then selected publicly available genotype data from 2163 reference canines among 115 populations: 108 modern breeds with at least 4 dogs per breed, 11 Greenland sled dogs, 3 regional village dog populations (4 Nigerian village dogs, 5 Vietnamese village dogs, 55 Chinese village dogs), 2 wolf populations (19 North American wolves and 25 Eurasian wolves), and 3 coyotes. In PLINK (v2.00a3LM) (Purcell et al. 2007), we identified 4,267,732 biallelic single nucleotide polymorphisms with <1% missing genotypes, and calculated the Wright’s F-statistics using Hudson method ((Bhatia et al. 2013; Weir and Cockerham 1984) for (1) each dog breed versus all other purebred dogs; (2) all village dogs versus all other purebred dogs; (3) each regional village dog population; (4) all wolves versus all other dogs; (5) North American wolves versus Eurasian wolves; and (6) coyotes versus dogs and wolves. We selected

1,858,634 SNPs with FST>0.5 across all comparisons, and performed LD-based pruning in 50kb windows for r2>0.2 to extract 363,718 markers for global ancestry inference. We merged genotype data for these SNPs from query samples with genotype data from reference samples, then performed global ancestry inference using ADMIXTURE (Alexander and Lange 2011) in supervised mode (random seed: 43) and 20 bootstrap replicates to estimate standard errors. Population weights under 1% were discarded from both unsupervised and supervised individual global ancestry proportions.

III.6.10 Local ancestry inference

Hidden Markov Model

We modified a previously published HMM-based approach (Patterson et al. 2004) (GitHub repo: TBD) to screen along the genome of individuals with recently mixed ancestry and identify which segments have been inherited from either of the ancestral populations; we tailed this to the canine population. This method assumes that our hybrid population has been recently derived by the mixing of wolf and dog populations.

We defined the “ancestry state” as whether an individual has 0, 1, or 2 alleles from the wolf population; this is a “hidden state” which is not observed. We defined the “observed state” as whether an individual carries 0, 1, or 2 alternative alleles at each locus. The sequence of ancestry states along chromosomes can

be represented as a Markov chain on three hidden states in which the transition probabilities vary according to the genetic distance (probability of historical recombination) between markers. We composed a set of ancestry informative markers with a fixation index (Fst) > 0.8 between North American wolves and domesticated dogs. The emission probability matrix is marker-dependent and is inferred for a given marker from the allele frequencies in each population. The HMM moves through those markers and at each marker, uses the observed genotypes and the correlation between nearby markers imposed by the model to produce a probability map of ancestry. Finally, we exploited the *Viterbi* algorithm to select the sequence of hidden states which reaches the maximum joint probability.

Local ancestry inference

Using the above approach we inferred the local ancestry of hybrids. This analysis included 25,640 autosomal ancestry informative markers (AIMs). We first calculated Fst for 4,874,271 variants in the `diverse canid reference panel` between the hybrid ancestral populations, North American wolves and domesticated dogs, then filtered variants by Fst 0.8 which gave us 30,482 variants. These markers have large frequency differences between the two ancestral populations. We then mapped the variants to hybrid genomic variants arriving at a final set of 25,640 variants as our ancestry informative markers, with an average of 1 marker per 81.2 kb (min:1bp; max:4.6Mb) (Table 3.10).

Simulation of wolf-dog hybrids

We used a Monte Carlo approach (GitHub repo: TBD) to generate simulated admixed individuals with known ancestry per haplotype. We split the reference dataset into two folds, the first fold includes 55 dogs and 16 North American wolves, the second fold includes 55 dogs and 17 North American wolves. In order to simulate admixed individuals with different levels of admixture, we performed admixture from generation 1 to 50 as follows: at each simulation run, a minimum of 5 and a maximum of 14 random individuals were selected randomly as parent haplotypes contributing to the admixture, this includes 10-28 haplotypes in total. In the first generation, two haplotypes from different individuals were randomly drawn from the parent population. Recombinations were treated as a Poisson event occurring on average once every Morgan. After each generation, the simulated child haplotypes were added to the parent haplotype pool and the same procedure is repeated until reaching the desired number of generations of admixture. We used the simulated hybrids to assess the performance of the local ancestry inference method described above.

III.6.11 Identification of wolf-overrepresented regions and gene ontology

We inferred local ancestry for 74 hybrids and calculated the percentage with wolf ancestry at each SNP. (Ex: 100% means all hybrids inherit wolf ancestry at this locus and 0% means all hybrids inherit dog ancestry at this locus.) We then

ranked the sites by wolf ancestry percentage, with the highest being 82% and lowest 41%. We then defined the top 1% (wolf ancestry > 77.6%) as "wolf-overrepresented regions" and the bottom 1% (wolf ancestry <50.7%) as “dog-overrepresented regions”. We performed gene set enrichment analysis using the PANTHER (http://www.pantherdb.org/) online tool.

III.6.12 Dataset clarification

We have three reference panels with different sets of individuals for different analysis. `Imputation haplotype reference panel` and `global ancestry calling panel` were constructed by colleagues (Morrill et al. 2022) and `Diverse canine reference` and `Local ancestry inference panel` were constructed in this study.

Table 3.10. Reference panels clarification.

Name	Samples	Usage
Imputation haplotype reference panel (Morrill et al. 2022)	435 individuals including 287 pedigreed dogs of various breeds, 6 dogs of unknown ancestry, 100 indigenous village dogs worldwide, 36 wolves, and 6 other wild canids	Variant calling and imputation

Global ancestry calling panel (Morrill et al. 2022)	2163 reference canines from 115 populations: 108 modern breeds with at least 4 dogs per breed, 11 Greenland sled dogs, 3 regional village dog populations (4 Nigerian village dogs, 5 Vietnamese village dogs, 55 Chinese village dogs), 2 wolf populations (19 North American wolves and 25 Eurasian wolves), and 3 coyotes.	Global ancestry calling (Figure 3.3)
Diverse canine reference panel	172 individuals including 110 dogs, 33 North American wolves, 26 Eurasian wolves, and 3 coyotes	Exploratory PCA (Figure 3.1)
Local ancestry inference panel	110 dogs and 33 North American wolves (subset of `diverse canine reference panel`)	Local ancestry inference

Table 3.11. Hybrids set in data analyses.

Analysis	Samples
Morphology collection	88 wolfdog hybrids and 2 coyote dog hybrids
Behavior collection	78 wolfdog hybrids and 2 coyote dog hybrids
Genotyping	74 wolfdog hybrids and 2 coyote dog hybrids
Global ancestry calling	74 genotyped wolfdog hybrids and 2 coyote dog hybrids
Local ancestry calling	74 genotyped wolfdog hybrids
Association study on morphology	74 genotyped wolfdog hybrids with morphology data collected
Association study on behaviors	69 genotyped wolfdog hybrids with behavior data collected

CHAPTER IV. MAPPING AESTHETICS AND SOCIAL BEHAVIORS IN WOLF-DOG HYBRIDS

IV.1 Preface

Kathryn Lord and Brittney Logan undertook the effort of gathering morphological data, designing and conducting behavioral novelty tests, and transforming novelty test videos into quantifiable behavioral measurements. Following this, I took charge of delineating the behavioral traits for the association test and carried out all subsequent analyses.

IV.2 Abstract

In this chapter, we employed an integrative approach combining admixture mapping and genome-wide association studies to identify genes and pathways connected to domestication-related behavior. Our study utilized an exceptional admixed population of wolf-dog hybrids housed in sanctuaries across the United States. We delineated behavioral phenotypes through dimensional reduction analysis of coded video data and detected associations between these phenotypes and genes and regulatory elements using admixture mapping and association tests. Additionally, we delved into the functional and biological mechanisms underpinning the associated regions through gene-set analysis. Our findings highlight that regions associated with domestication-related behavioral differences are predominantly enriched for genes expressed in the brain, particularly those active during early infancy. In addition, we examined the gene expression patterns across various developmental periods for candidate genes,

focusing on those with high expression during early infancy. This led to the identification and prioritization of a set of genes for subsequent functional exploration.

IV.3. Background

The existence of major behavioral distinctions between dog and wolf along with their relatively recent divergence, has brought these subspecies to the forefront of numerous population-level analyses. These studies have discovered genes that appear to be under selection in dogs, including many predominantly expressed in the brain (Axelsson et al. 2013). However, genomic comparisons between dogs and wolves are of limited impact, as association studies exploring differences between these two populations cannot determine which genes correspond to which differences. This is because, as for any study that compares members of one group to members of another, all dogs differ from all wolves in terms of various behavioral and morphological traits, resulting in every genetic difference being linked to all phenotypic differences. To address this issue, I conducted genome-wide association study (GWAS) combined with admixture mapping on wolfdog hybrids with a range of wolf:dog ancestry ratios and various behavioral phenotypes.

We combined GWAS with admixture mapping to increase the statistical power of mapping social behaviors in wolfdog hybrids. Genome-wide association study is

a widely used approach in genetics research to identify associations between genetic variants and specific traits or diseases (Uffelmann et al. 2021). This has proven to be an effective approach for identifying causal variants in domesticated dogs, a population with homogeneous dog ancestry, as demonstrated in our previous study (Morrill et al. 2022). Admixture mapping is a genetic technique used to identify genes and genetic variants associated with specific traits or diseases in populations with mixed ancestry. It was initially developed to map diseases in human admixed populations, such as Latino and African American populations with recent ancestry from two or more ancestral groups (Pasaniuc et al. 2011; Pino-Yanes et al. 2015), and was later adapted to non-human species, including cattle where many breeds or populations are hybrids of two divergent ancestral genomes (Kassahun et al. 2015; Cai et al. 2018). The key assumption for admixture mapping is that causal allele frequencies vary between ancestral populations and the trait is differentially distributed in the ancestral populations. Under this assumption, individual wolfdog hybrids exhibiting more wolf-like behavior in the admixed population would be expected to have a higher proportion of wolf ancestry at the causal locus. Conversely, individuals displaying more dog-like behavior would be expected to have a lower proportion of wolf ancestry at the causal locus. Therefore, we applied admixture mapping to identify regions that have local ancestry differently distributed between wolf-like and dog-like hybrids and used GWAS to further investigate the candidate regions and prioritize casual genes.

In Chapter IV, I present my research on mapping aesthetic and behavioral traits in wolfdog hybrids using GWAS combined with admixture mapping. I discuss the collection of phenotypic data including morphology and behaviors, as well as the use of GWAS and admixture mapping to link genotype to phenotype. Additionally, I interpret the identified candidate loci with data from population differentiation, and mammalian constraints. Finally, I explore the potential of utilizing hybrids as a natural model for studying human psychiatric disorders, such as autism spectrum disorder (ASD).

IV.4. Results

IV.4.1 Aesthetic traits that differ between wolves and dogs

In our study, we evaluated 18 aesthetic traits that are typically used to differentiate wolves from most dogs (Table 4.1). While wolves tend to have a more consistent physical appearance, dogs can exhibit a range of appearances that may either resemble or differ from that of wolves. Although there are no specific physical traits that are exclusive to wolves and absent in dogs, we focused on traits that are commonly used to distinguish wolfdogs from dogs. More than 90% of the wolfdog hybrids exhibited the wolf-version in 8 of these traits, including mesocephalic face shape, double coat, dense ear fur, erect ears,

median fur length, no coat pattern, front dew claw, and tail held straight and below the level of the back. In contrast, only 9 wolfdog hybrids carry small ears, the wolf-version for ear size, with the rest exhibiting median or large ears (Figure 4.1). Overall, the hybrids more often had the appearances in wolf-version than non wolf-version, with an average of 14.0 (SD:2.1) traits in wolf-version among all 18 traits (Figure 4.2B). Percent wolf ancestry was positively associated with the number of wolf-version traits exhibited in individual hybrids, indicating that ancestry has a big effect on the appearance of hybrids (Figure 4.2A).

Table 4.1 Aesthetic traits mapped in wolfdog hybrids

Morphological Categories	**Type**	**Typical Wolf**	**Dog**
Fur Coat	Single, Double	Double	All Options
Fur Length	Short, Medium, Long	Medium	All Options
Fur Base Color	Agouti, Black, White, BlackTan, Tan	Agouti	All Options
Fur Pattern	WhiteSpots, ColoredSpots, BlockedColor, Grizzled, BlackMuzzle, EyeBrows, Dilute, Foreheadstripe, WidowPeak	None	All Options
Face Mask Type	None, Diffuse, Defined	None or Diffuse	All Options
Nail Color	Dark, Light, Mix	Dark	All Options
Dew Claw	Front, Front&Back	Front	All Options
Paw Size (Compared to body)	Other, Large	Large	All Options

Ear Type	Erect, Drop, Erect+Drop	Erect	All Options
Ear Size (Compared to head)	Small, Medium, Large	Small	All Options
Ear Fur	Dense, Sparse	Dense	All Options
Nose Color	Black, Pink, MixBoth	Black	All Options
Face shape	Brac, Mes, Doli	Mes	All Options
Muzzle Stop	Abrupt, Gradual	Gradual	All Options
Chest Shape	Compressed, Barrel, Average	Compressed	All Options
Leg length	Short, Medium, Long	Long	All Options
Eye color	Yellow, Brown, Black, Blue	Yellow or Brown	All Options
Tail Carriage	Straight, Curled, Other	Down Straight	All Options

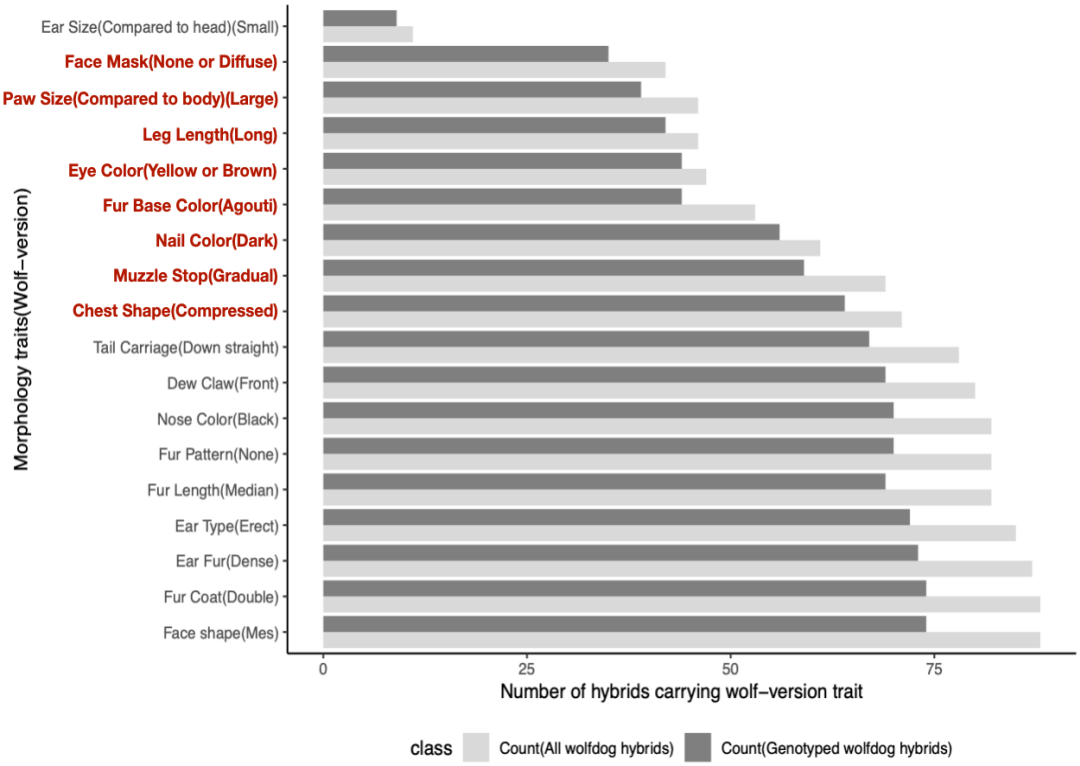


Figure 4.1. Hybrids exhibit more wolf-version morphological traits. Hybrids are counted within two groups, 74 genotyped wolfdog hybrids (dark gray) and all 88 wolfdog hybrids (light gray). Traits that are used for association tests are marked in red.

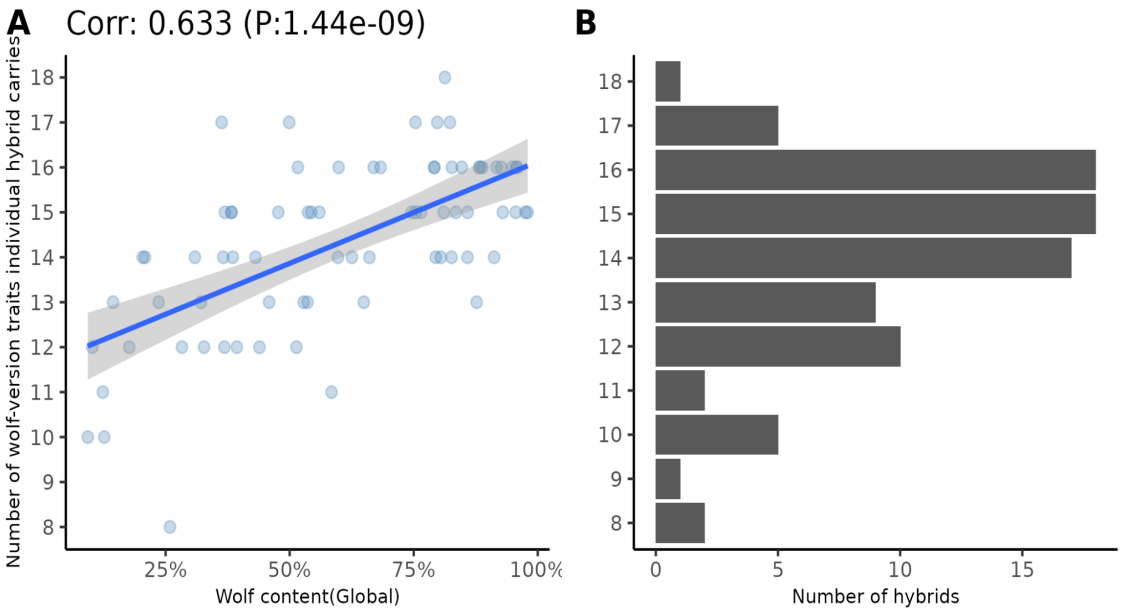


Figure 4.2. High wolf-content wolfdog hybrids carry more wolf-version morphological traits. (A) Relationship between wolf content and the number of wolf-version morphological traits an individual hybrid carries. (B) Distribution of hybrids in categories. Categorized are defined by the number of wolf-version morphological traits an individual hybrid carries.

IV.4.2 Mapping aesthetic traits in wolf-dog hybrids

We conducted SNP association mapping and admixture mapping on 8 morphological traits for which more than 15% but less than 85% of genotyped hybrids carried the wolf-version (Figure 4.1). We performed this using a cohort of 74 wolfdog hybrids with both genotype and morphology data collected (Table 3.11). One additional morphological trait of black coat color is included as well, with 19 hybrids exhibiting black color and 55 hybrids exhibiting non-black color. Our data shows that the genotyped hybrids are a good representation subset of

hybrids in our study population. All phenotypes are used as quantitative traits to enable the same significance threshold as the behavioral traits described in the following sections.

Using MIXSCORE (Pasaniuc et al. 2011), we estimated SNP association scores conditioned on local ancestry (QSNP1, chi-squared score with 1 dof), as well as admixture scores that associates the local ancestry to the continuous phenotype with genome-wide ancestry as a covariate (QADM, chi-squared score with 1 dof). We then summed the QSNP1 and QADM score to produce a QSUM score (chi-squared score with 2 dof) to represent the association strength at each testing variant (Pasaniuc et al. 2011).

Our study successfully replicated published associations for morphological trait `black coat` color, including a gene CBD103 (P=1.70E-08) which codes for beta-defensin, with a mutation that can bind to Mc1r and shift the melanocyte from producing pheomelanin (yellow pigment) to eumelanin (black pigment) (Candille et al. 2007) (Figure 4.3A). Additionally, we identified new associations for the morphological trait of `muzzle stop` with gene FOXP1 (1.71E-09), codes for the forkhead box (FOX) protein which may regulate bone development in jaw (Cesario et al. 2016) and gene GALNTL6 (P=1.35E-08), which codes for a protein involved in protein O-linked glycosylation via threonine. This gene is

expressed in the facial bone primordium and is associated with cleft lip (Curtis et al. 2021) (Figure 4.3B).

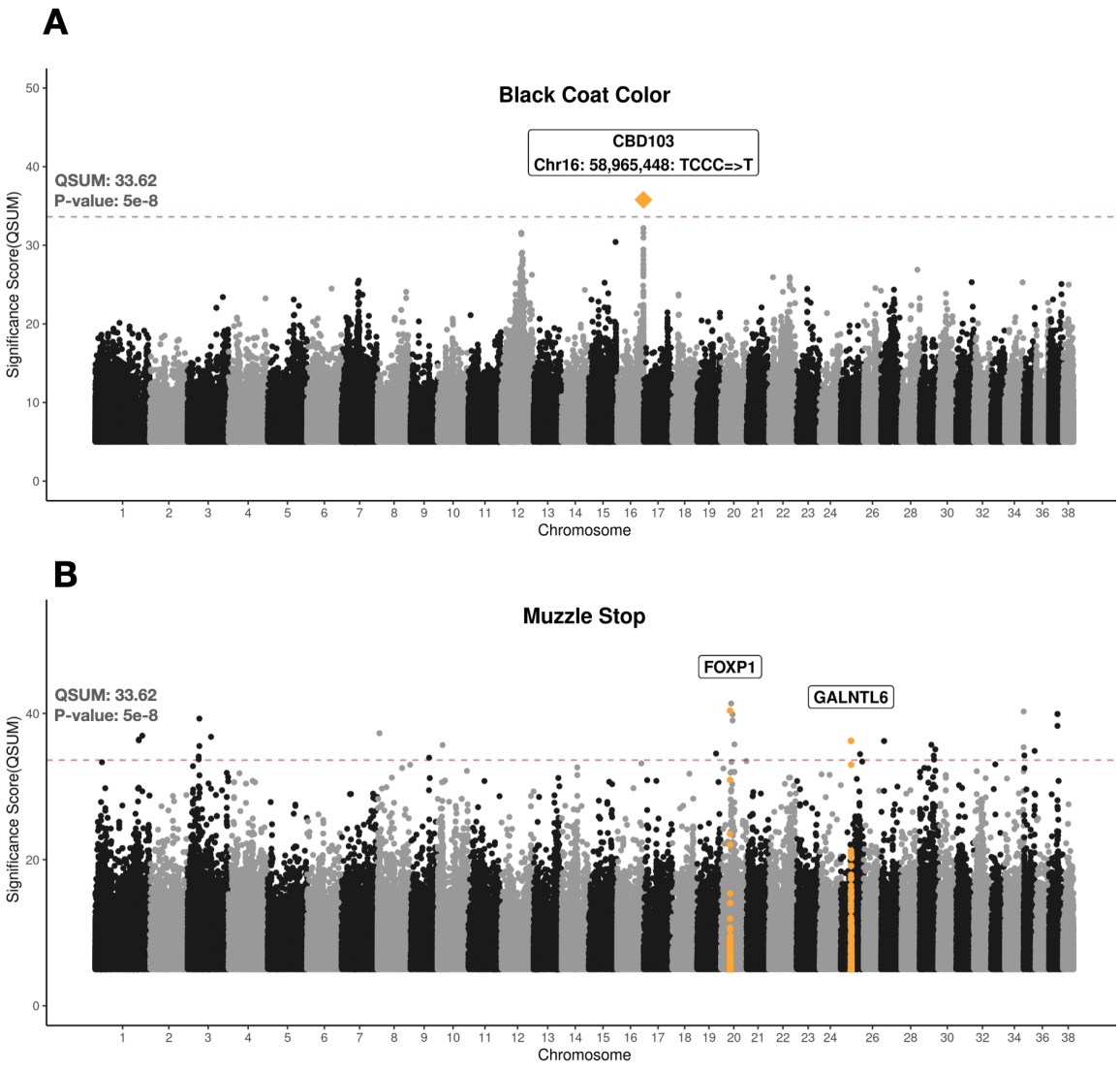


Figure 4.3 Genome-wide associations for two aesthetic traits. Manhattan plot of QSUM score for (A) black coat color (B) muzzle stop. Red line represents genome-wide significance at P-value 5e-8 (QSUM score: 33.62). Candidate causal variants and genes are highlighted in orange.

IV.4.3 Measuring domestication-related behavior in hybrids

Novelty test to measure behaviors

Compared to wolves, dogs typically demonstrate less fear of novelty and greater ability to generalize from one new experience to another (Moretti et al. 2015). To assess wolfdogs' response to novelty, we first introduced a familiar object into their enclosure and video recorded their behavior for 10 minutes (Figure 4.4). We then repeated the test with a novel object and finally with both objects together. Each 10-minute video was scored by one or two observers using an ethogram Modified from a previous study (Moretti et al. 2015) (Table 4.2). Videos scored by two observers were used to calculate Inter-observer reliability with Cohen's Kappa (Sun 2011) and resulted in Cohen's Kappa reliability of average 0.83 (SD: 0.12).

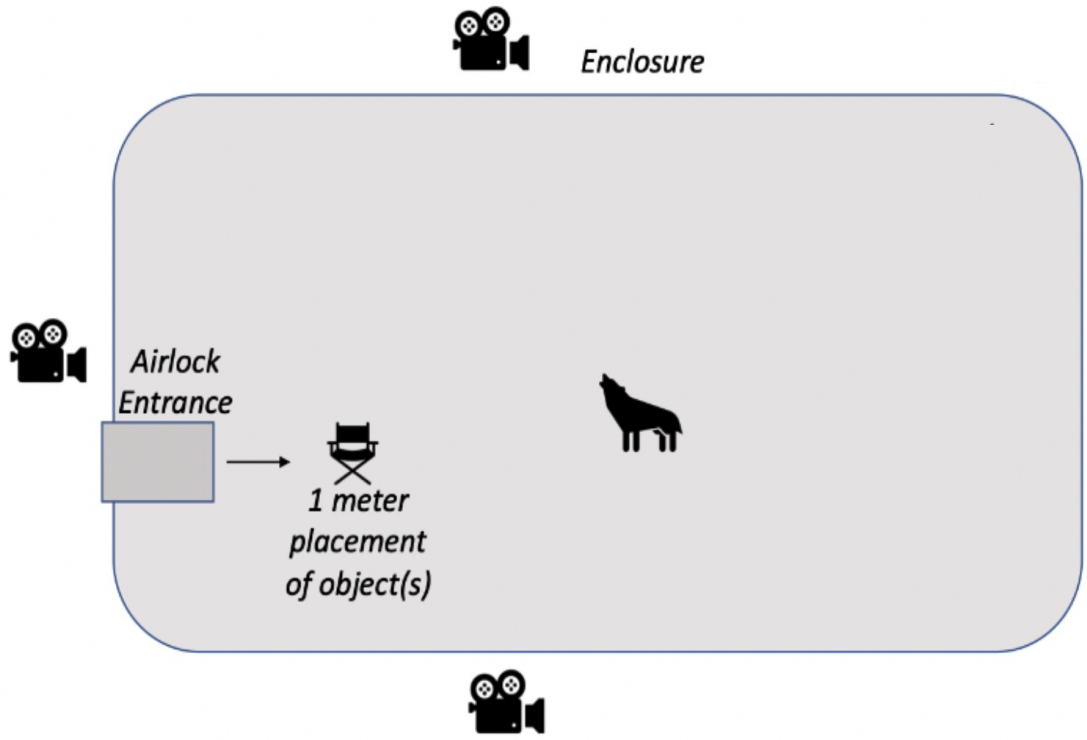


Figure 4.4. Enclosure setup. At least two cameras were placed on different sides of the enclosure, testing object (s) were placed one meter away from the entrance.

Variation in wolfdog responses to familiar, novel, or paired objects

We measured response to novelty in 80 wolfdogs by scoring 14 measurements for each of the three tests (Table 4.2). For each test, we recorded frequency, latency and, when applicable, duration for five distinct actions: whether the animal is (1) out of sight, (2) shows fear, or (3) investigates, (4) contacts or (5) manipulates the object. Nearly all the wolfdog hybrids investigated the object in all three tests, although fewer contacted or manipulated it (Table 4.3). For both

dogs and wolves, novel objects attract more attention than familiar objects. This is reflected by our observation that more hybrids performed investigate, contact, manipulate, and fear behaviors to a novel object than a familiar object (32% increase of activity). The setup of pairing the novel object together with a familiar object aims to test the ability of wolfdogs to generalize a familiar object in the presence of a novel object. When considering all hybrids at once, we see slightly higher attention to paired objects than a familiar object (14% increase of activity) and slightly lower attention to paired objects than a novel object (14% decrease of activity) (Table 4.3).

Table 4.2. Novelty Test Ethogram. Behavioral categories coded modified from Moretti et al. (Moretti et al. 2015). All behaviors except fear were scored as latency, frequency and duration. Fear behavior was scored as latency and frequency only.

Behavior	Description
Investigate	Sniffing or looking toward object without touching it
Contact	Touching object or close enough to be touching object
Manipulation	Moving all or some of the object with mouth, paw and/or nose
Fear	Jumping, running away, flinching, crouching, and/or tucking tail
Out of sight	The animal is not visible enough to determine if it is engaged in any of the other behaviors in the ethogram

Table 4.3. Hybrid's activity when tested with different objects. Activity was measured by the total number of hybrids that performed a behavior at least once during the test (total count: 80).

Behavior	Familiar Object	Novel Object	Paired Object
Investigate	56	67	67
Contact	53	57	55
Manipulation	14	29	15
Fear	16	30	21
Out of sight	67	65	63
Total activity (excluding out of sight)	139	183	158
Fold-change relative to "familiar object"		**1.32 (32% increase)**	**1.14 (14% increase)**
Fold-change relative to "novel object"			**0.86 (14% decrease)**

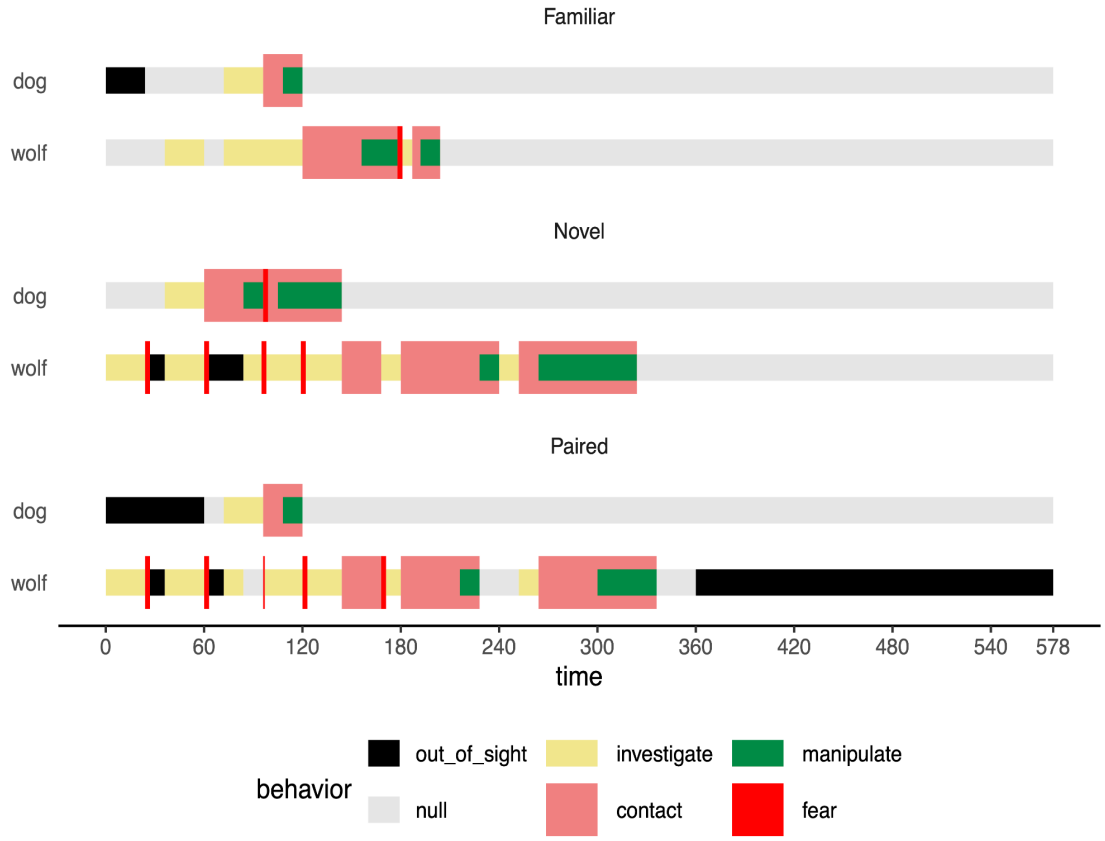


Figure 4.5. Schematic of wolf or dog's response to familiar, novel, or paired objects. Colors represent behavior categories across the 578 second test. Based on previous findings by Morettil et al. (Moretti et al. 2015), wolves exhibited more interest (investigation, contact, manipulation, and fear) than dogs in response to both novel and familiar objects, with this difference being particularly noticeable in the case of novel objects. With regard to paired objects, dogs showed better generalization ability than wolves, as they found paired objects to be less interesting than familiar objects, as evidenced by their lower level of activity. Conversely, wolves demonstrated poorer generalization ability than dogs, responding with greater fear to paired objects than to novel object alone, leading to higher activity levels and sometimes even causing them to hide far away from the objects.

Behavioral differences among hybrids with different levels of wolf ancestry

Animals with a wolf-like response to all three objects are expected to have a higher frequency and longer latency for the investigate, contact, manipulate, fear, and out of sight behaviors, as well as a longer duration for all behaviors except fear (for which the duration is not measurable) ((Moretti et al. 2015) (Figure 4.5).

We used the novelty test data to ask about the correlation of wolf-like behavior to genomic wolf-content for the 69 hybrids with both genotype and behavior data collected. Five among the 42 behavioral measurements have positive correlations with wolf content with heritability ranging from 0.38 to 0.99 (Table 4.4, Figure 4.6). Hybrids with high wolf ancestry show more fear in all three testing objects. For paired objects, hybrids with high wolf ancestry exhibited fear behavior much earlier than those with low wolf ancestry, and stayed out of sight more frequently.

Table 4.4 Five behavioral measurements that are significantly correlated with wolf content.

Object	Measurement	Behavior	Correlation Coefficient	P-value	Heritability	Heritability P-value
Paired	Frequency	Fear	0.38	1.43E-03	0.973997	1.49E-04
Paired	Latency	Fear	-0.37	1.94E-03	0.697853	2.79E-03
Paired	Frequency	Out of sight	0.30	1.37E-02	0.999999	3.08E-03
Novel	Frequency	Fear	0.29	1.57E-02	0.825666	2.79E-03
Familiar	Frequency	Fear	0.25	3.96E-02	0.337654	4.83E-02

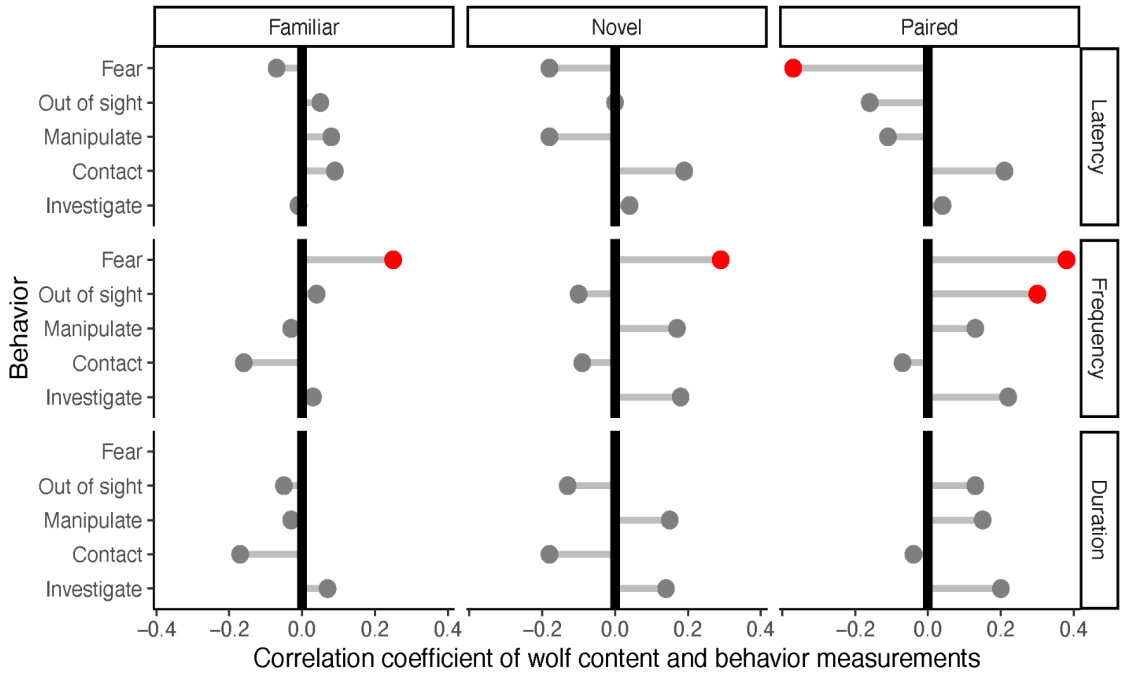


Figure 4.6 Correlation coefficient of wolf content and 42 behavioral measurements. Significant correlations are represented by red dots.

IV.4.4 Defining behavioral phenotypes for association test

To reduce the dimensionality of the novelty test results, we defined behavioral phenotypes for association testing using principal component analysis (Figure 4.7). The first three components each explained more than 10% of the variance (26%, 14%, and 10% respectively). We find that the first component (`PC1`)

captures an animal's interest in the object in any of the three tests. Higher-scoring animals investigated, contacted and manipulated the object more quickly, frequently, and for longer periods than lower-scoring animals. The second component (`PC2`) reflects fear response or caution in all three tests. Higher-scoring animals reacted fearfully sooner and more often in all three tests, and tended to investigate the novel object for longer while also being less willing to contact or manipulate it. On the third component (`PC3`), higher-scoring animals were less likely to engage with the familiar object (exhibiting longer latency and lower frequency and duration) and more likely to engage with, and react fearfully, to both the novel and paired objects.

We generated a new score `PC1+PC2+PC3` for each animal by summing all three principal components. Higher values for this composite score represent more wolf-like behaviors, indicating more interest as well as more fear for objects, especially for novel objects either alone or paired with familiar. This new score `PC1+PC2+PC3` accounts for 60% of variance, and heritability is much higher compared to individual principal components (Table 4.5). We assessed the relationship between wolf-like behaviors and wolf ancestry by calculating the correlation coefficients between the principal components and wolf ancestry for each principal component and the sum of the three. All of them are positively correlated with percent wolf ancestry, with `PC3` and `PC1+PC2+PC3` significantly correlated (Figure 4.8).

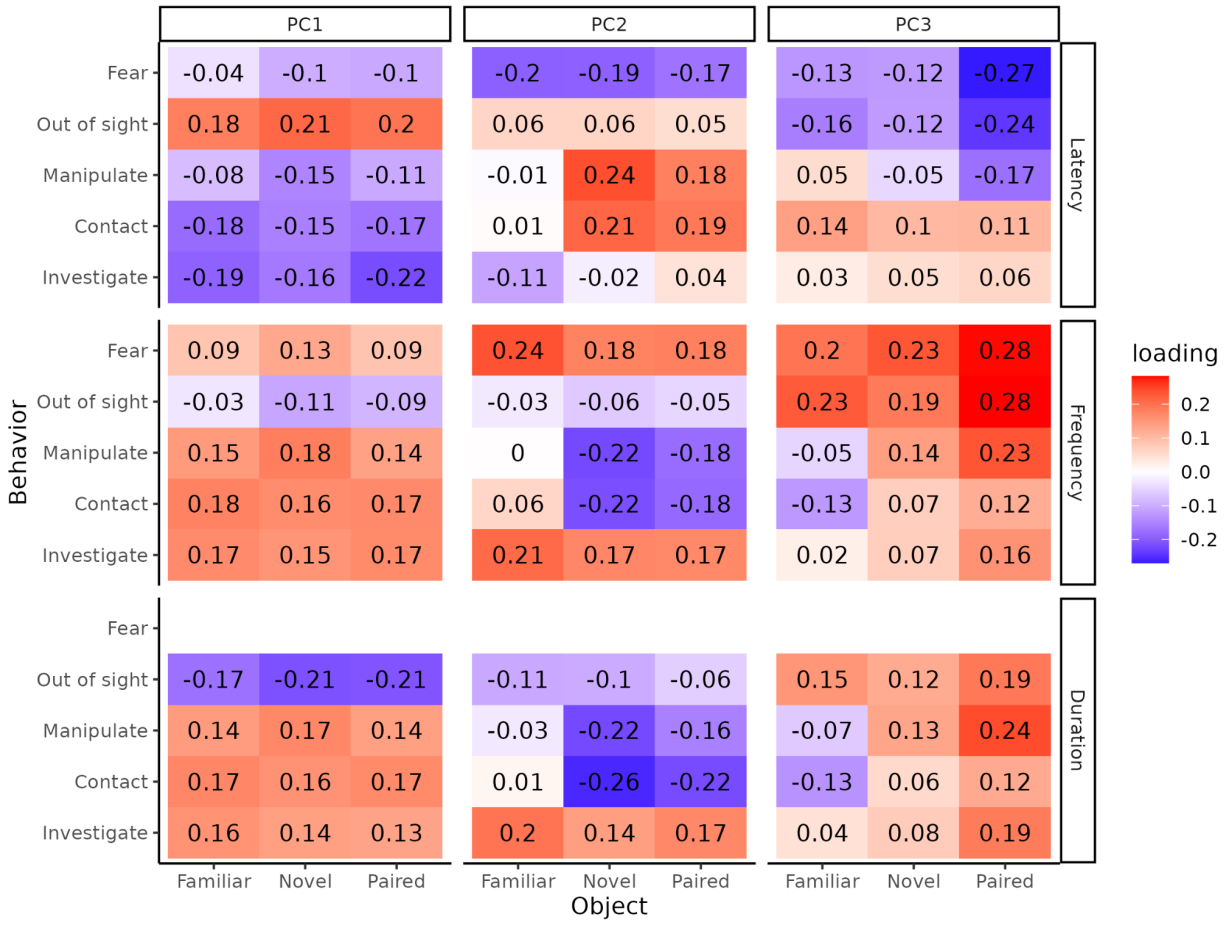


Figure 4.7. Loading of measurements on behavior PCA. We performed principal analysis of 42 measurements across 80 hybrids and extracted the top three principal components. Measurements captured by each component and loadings are reported.

Table 4.5 Heritability of the top three principal components and the sum of them.

Trait	Heritability	SE	Heritability P-value	Correlation coefficient with wolf ancestry	Corr P-value
PC1	0.84	0.48	7.34E-02	0.05	7.04E-01
PC2	1.00	0.40	1.06E-03	0.16	1.88E-01
PC3	1.00	0.35	1.76E-03	0.37	1.55E-03
PC1+PC2+PC3	1.00	0.37	1.06E-04	0.29	1.44E-02

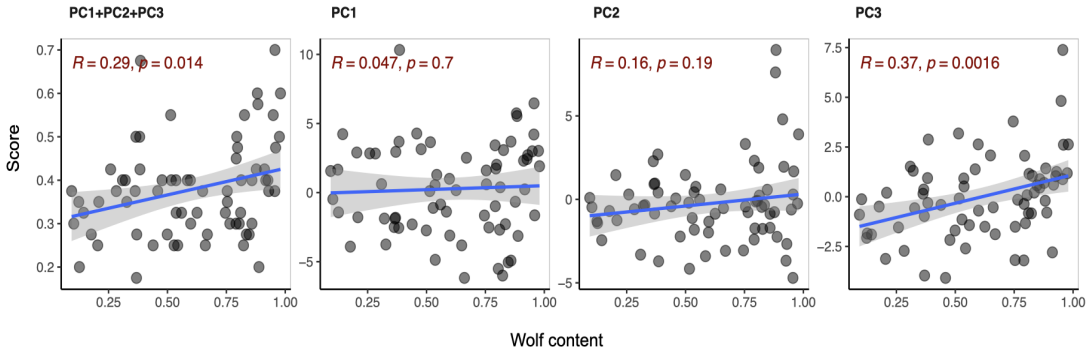


Figure 4.8. Correlation between wolf ancestry and behavioral principal components (PC) and the sum of the PCs. A linear regression line was fit based on the points in the scatter plot and the correlation coefficient is highlighted in red.

IV.4.5 Mapping social behaviors in wolf-dog hybrids

To identify the loci associated with scores for each of the first three principal components of behavior (PC1, PC2, and PC3), and the combination of the three (PC1+PC2+PC3), we utilized the same methodology as for the mapping of aesthetic traits. We used statistics (QADM) from admixture mapping to report associated candidate regions and used statistics from SNP association conditioned on local ancestry (QSNP1) as additional evidence to prioritize causal genes for further analysis.

Our admixture mapping method employed a linear-mixed effect model with local ancestry as independent variable, phenotype as a dependent variable, and genome-wide wolf ancestry as a covariate. At each testing position, we calculated a QADM score that reflected the likelihood of exhibiting wolf-like behaviors when one or two wolf alleles were present. The recent history of admixture gives rise to long-range correlation in QADM scores across the genome, thus the threshold for genome-wide significance levels of admixture mapping is substantially lower than that for the genotype test. On the basis of a permutation approach, a QADM score as 16.45 yielded a genome-wide FDR of 0.05 was used for genome-wide significance and a QADM score as 13.96

yielded a genome-wide FDR of 0.2 was used for suggestive significance threshold.

Admixture mapping identified regions associated with social behaviors

Using this approach, we identified a significant locus (QADM: 19.38, FDR<1%) on chromosome 19 that covered 1,565,112 base pairs and contained the genesHNMT, SPOPL, and NXPH2 for the phenotype `PC1+PC2+PC3` (Figure 4.9A, Figure 4.10). Additionally, there was another associated locus (QADM: 15.58, FDR: 8%) situated 1.6 Mb downstream of this region that spanned 1,413,994 base pairs and encompassed three genes, LRP1B, KYNU, and ARHGAP15 (Figure 4.9A). The human homolog of these two loci is situated on 2q22, and its deletion in humans has been associated with severe intellectual disability (Mulatinho et al. 2012). We also detected another region (QADM: 15.01; FDR: 11%) on chromosome 34 spanning 435,342 bases (Figure 4.9A), which contained the genes ST6GAL1, RTP1, MASP1, and RTP4, with MASP1 playing a vital role in the complement pathway and migration of neurons in the developing cortex (Fagan et al. 2017). Furthermore, we identified a region (QADM:16.54; FDR:4%) significantly associated with phenotype `PC3` on chromosome 9 that spanned 541,964 bases and contained 21 genes (Figure 4.9D). However, there is no notable evidence linking these genes to the behavioral phenotype, and further studies are necessary to investigate the possible role of regulatory elements in this region.

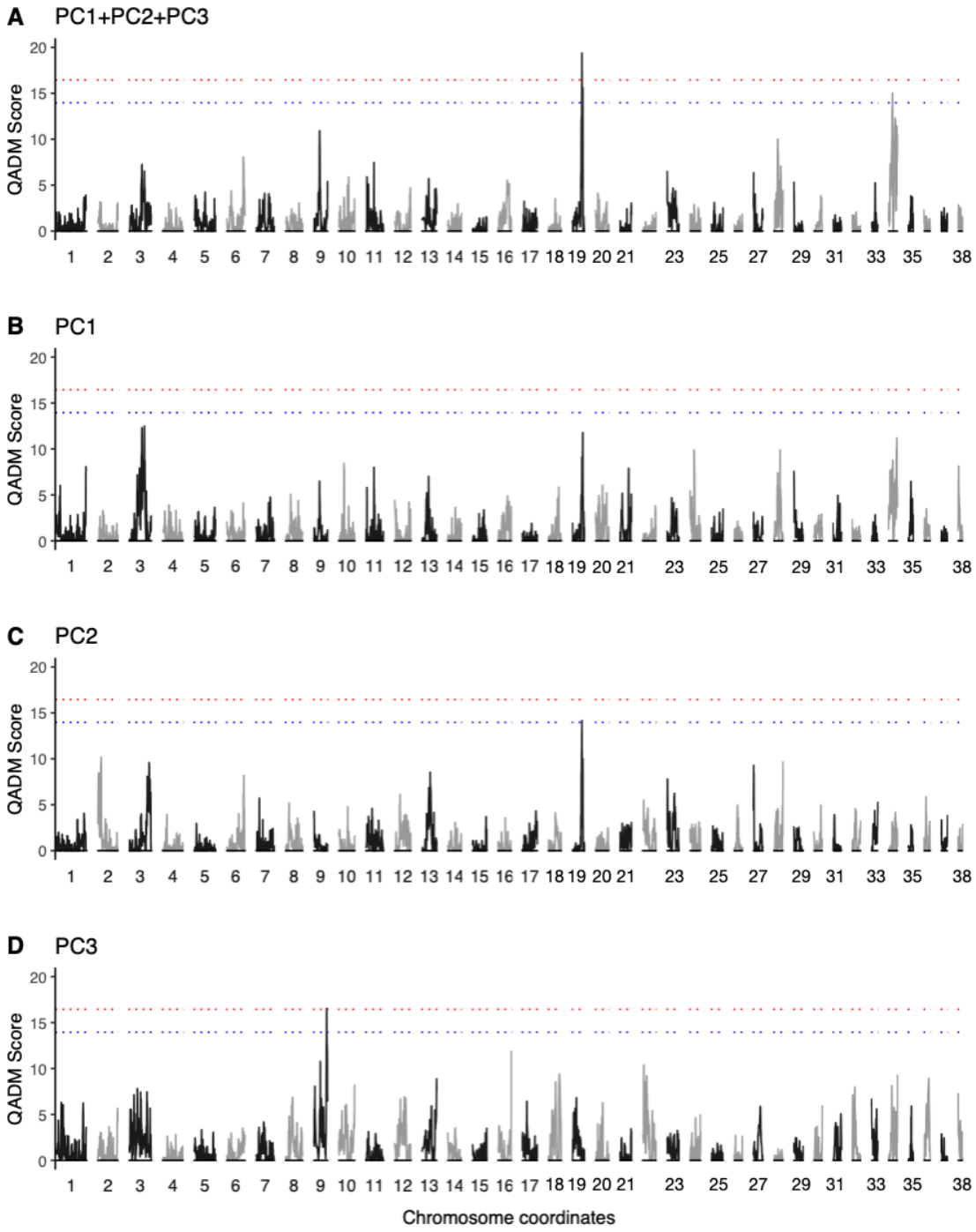


Figure 4.9. Admixture mapping scan results for trait PC1, PC2, PC3, and PC1+PC2+PC3. Neighboring variants with the same QADM score are merged into one region. Genome-wide significance as FDR 0.05 is highlighted in red, suggestive significance as FDR 0.2 is highlighted in blue.

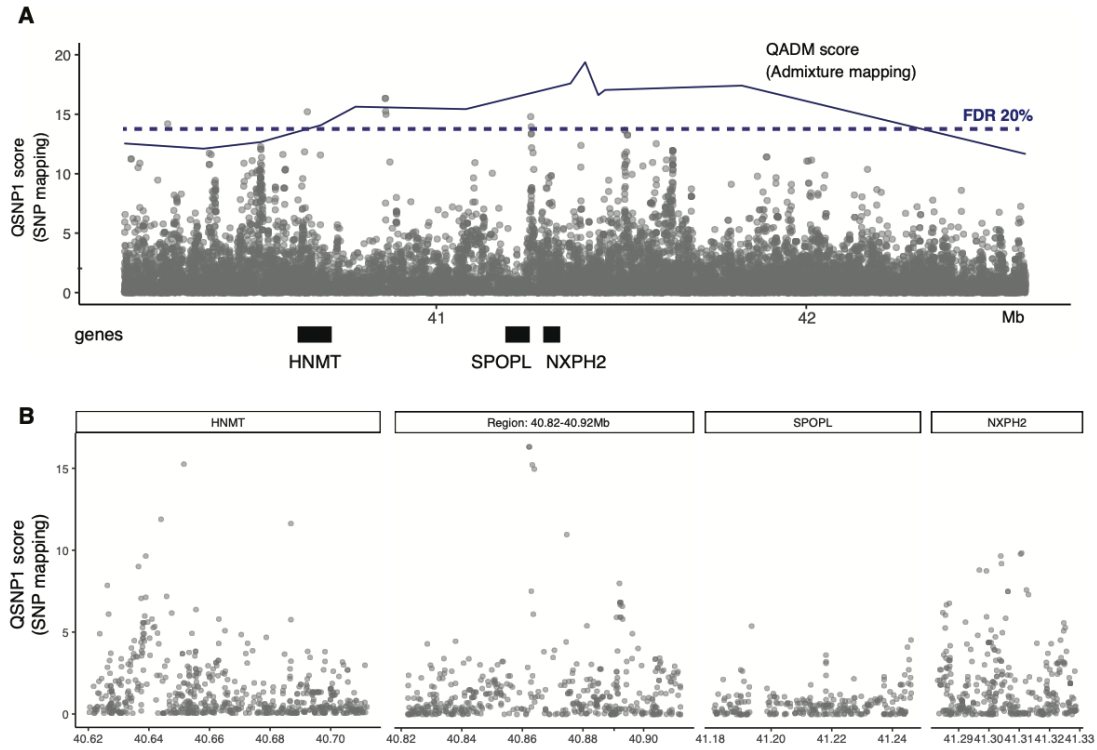


Figure 4.10. QADM, QSNP1 score of candidate region on 19 associated with trait `PC1+PC2+PC3`. (A) The QADM scores are shown by a continuous dark blue line, while the QSNP scores for every variant tested are denoted by dots. (B) The QSNP1 scores for potential genes and regulatory areas within the chromosome 19 candidate region are displayed. The scale of chromosome coordinates is given in megabases (Mb).

Investigate the population differentiation and conservation in candidate regions

We assessed the candidate loci using Fst and phyloP scores and calculated the regional average for each locus (Table 4.6). For the locus on chr19 spanning from 40.6Mb to 42.2Mb and on chr34 spanning from 55.4Mb to 56.0Mb, associated with `PC1+PC2+PC3`, the average Fst and phyloP score are both higher than the genome-wide average. For the locus on chr9 associated with `PC3`, the Fst score is higher than genome-wide average while the phyloP score is lower than genome-wide average. For the locus on chr19 spanning from 43.8Mb to 45.2Mb, both Fst and phyloP are lower than genome-wide average.

In principle, regions associated with the behavioral difference between wolves and dogs should exhibit greater differentiation, resulting in a high Fst. We observed this pattern in three out of the four regions identified. Conservation constraint, reflected by phyloP score, has a resolution at one base pair and is better suited for fine-mapping rather than evaluating the overall pattern. We recognize that region-wise phyloP score is not an ideal metric to measure conservation of the candidate loci. Instead, it should be employed to prioritize regions or genes for further investigation.

Table 4.6. Fst and phyloP score in candidate regions.

Trait	Chromosome coordinate	QADM score	Average Fst	Average phyloP score
PC1+PC2+PC3	19: 40,622,108-42,187,220	19.3767	0.240	-0.211
PC1+PC2+PC3	19: 43,807,826-45,221,820	15.5768	0.165	-0.269
PC1+PC2+PC3	34: 19,463,002-19,898,344	15.0147	0.222	-0.202
PC3	9: 55,415,820-55,957,784	16.5435	0.212	-0.279
Genome-wide average	NA	NA	0.194	-0.243

IV.4.6 Assessment of environmental factors on behavior

The positive correlation between wolf ancestry and behavioral principal components (PC) and the combined components showed that ancestry is a key factor in explaining variation in behaviors. Heritability of the four phenotypes ranges from 0.84 to 1.0 indicates that genetics can explain much of the variation as well (Table 4.5). However, the heritability calculation is largely influenced by the sample size and population structure in admixed populations. Further analysis with an approach suitable for small size and admixture population is needed. Here, we view the results as a guiding result and focus on the relative difference between components (Yang et al. 2011; Luo et al. 2021). Behaviors are complex traits and can be influenced by other factors, such as age, sex, and living conditions. To assess that, we performed an Analysis of Variance (ANOVA)

(Faraway 2002; Sthle and Wold 1989) test to explore how those factors, in addition to ancestry and genetics, influence the behavioral phenotype (Figure 4.11).

Our analysis revealed that the effect of age, sex, and living conditions on behavioral phenotypes could not be disregarded (Figure 4.11). For `PC1+PC2+PC3`, although wolf ancestry explained the largest variation (>10%) in phenotype, age, enclosure size, and total numbers of animals in the same enclosure also have significant effects, and each explains more than 5% of variance. For `PC1`, enclosure size has the largest effect on phenotype while ancestry has very minimal effect. This is consistent with the correlation analysis that only a minor positive correlation is observed between `PC1` and wolf ancestry. For `PC2`, sex has the largest effect on phenotype and ancestry has a relatively lower effect. For `PC3`, only ancestry has a significant effect which explains more than 12% of the variation.

We mapped behaviors in hybrids using all three principal components and combined components and later used them in the evaluation of gene set enrichment and genetic overlap with dog behaviors and human psychiatric disorders. Despite the fact that ancestry has minimal effect on trait `PC1` and `PC2`, they effectively capture the variation in hybrid behaviors. Those heritable behavioral phenotypes, although influenced by environmental factors, are still

good candidates for trait mapping as long as we carefully control for those factors. Ancestry doesn't have a perfect correlation with phenotype, even in morphological traits. In practice, it is possible that a hybrid with high wolf ancestry may display dog-like behaviors simply due to the acquisition of a few "dog" genes with large effect sizes.

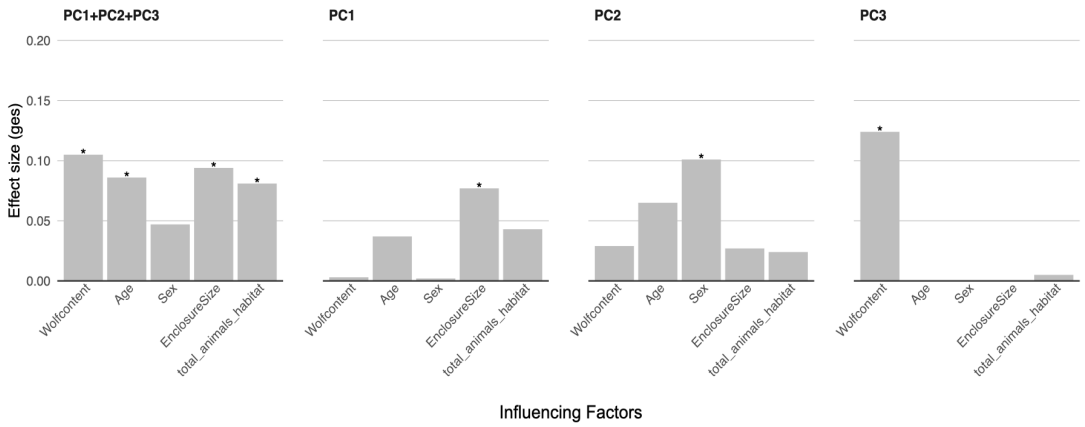


Figure 4.11. Analysis of variance in behavioral principal components and combined components among 69 wolfdog hybrids. Analysis of variance showed the effect size of ancestry exceeds 10% for principal component 3 and combined score.

IV.4.7 Test for enrichment of brain-expressed genes

We considered the QSUM score which combines SNP association and admixture association for association strength at single variants for gene-set enrichment

test. QSUM score is a chi-squared score with two degrees of freedom. We converted it to P-values using the `pchisq` function in R. We tested all four behavioral phenotypes for enrichment in gene sets defined from (1) human tissue-level gene expression from the GTEx portal (GTEx Consortium et al. 2017) and (2) human brain subregion-level gene expression from different developmental periods from Brainspan (Miller et al. 2014) (Table 4.7). The regions associated with the behavioral principal components of wolfdog hybrids showed enrichment for brain-expressed genes, but no gene set passed multiple test corrections (Figure 4.12). The highest enrichment was observed for `PC1` in genes expressed in brain clusters V1C and PFC-MS during the prenatal and infancy developmental periods (Figure 4.13). For PC1, we also observed enrichment in genes expressed in the cerebellar hemisphere from GTEx data independent of developmental periods (Figure 4.12). The regions associated with the phenotype `PC1+PC2+PC3` showed enrichment in genes expressed in all four brain clusters during brain development and were enriched in genes expressed in the cerebellum and basal ganglia independent of developmental period (Figure 4.13).

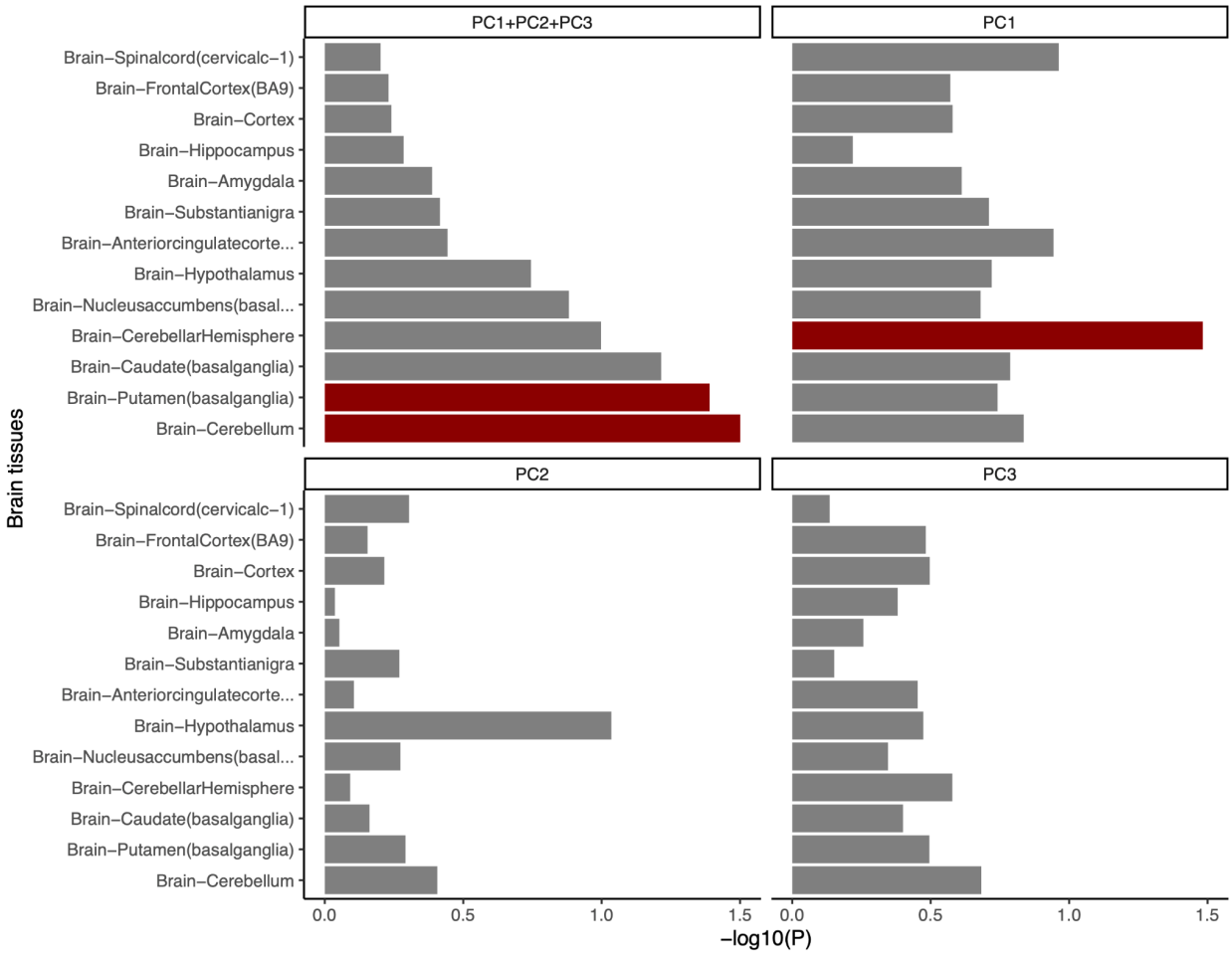


Figure 4.12. Brain-expressed gene set enrichment. Gene set enrichment of behavior associated regions to human brain tissue-expressed genes sourced from the GTEx Portal. None of the enrichments survived p-value adjustment using Benjamini-Hochberg procedure. Gene sets with empirical P-value less than 0.05 are highlighted in red.

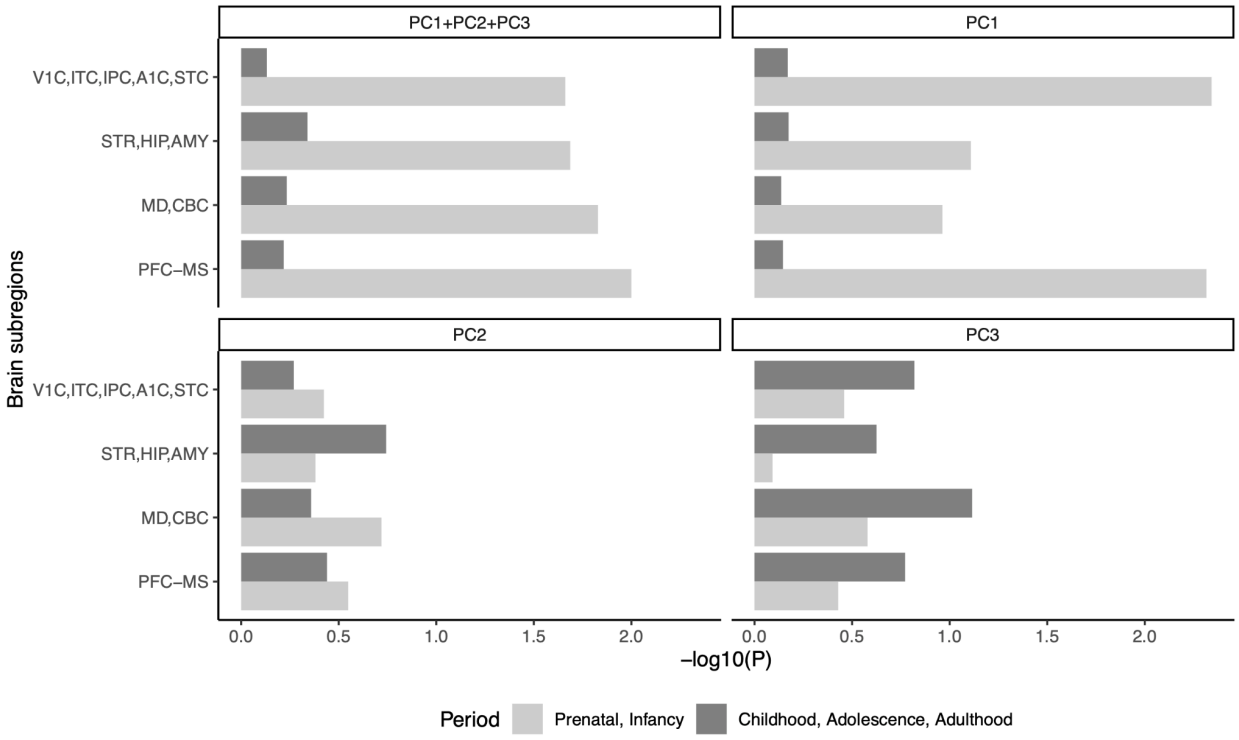


Figure 4.13. Brain-expressed gene set enrichment at different developing periods. Gene set enrichment of behavior associated regions to human brain tissue-expressed genes sourced from the brainspan database. None of the enrichments survived p-value adjustment using Benjamini-Hochberg procedure. Different developing periods are grouped into two categories, ages at 1 year old or younger are categorized as `prenatal, infancy`, ages older than 1 year old are categorized as `childhood, adolescence, and adulthood`.

Table 4.7 Hierarchical clustering of brain regions based on transcriptional similarity during fetal development divides the brain regions into four groups (Willsey et al. 2013).

Cluster	Synonyms	Brain region
Cluster1	V1C	primary visual cortex
	ITC	inferolateral temporal cortex
	IPC	posteroventral (inferior) parietal cortex
	A1C	primary auditory cortex

	STC	posterior (caudal) superior temporal cortex
Cluster2	STR	striatum
	HIP	hippocampus
	AMY	amygdaloid complex
Cluster3	MD	mediodorsal nucleus of thalamus
	CBC	cerebellar cortex
Cluster4	M1C	primary motor cortex
	S1C	primary somatosensory cortex
	VFC	ventrolateral prefrontal cortex
	MFC	anterior (rostral) cingulate (medial prefrontal) cortex
	DFC	dorsolateral prefrontal cortex
	OFC	orbital frontal cortex

IV.4.8 Overlap with dog phenotypes, and human psychiatric disorders

We conducted Fisher's exact tests (Nagel et al. 2018; Romero et al. 2022) to examine genetic overlap for top associated genes for each pair comparing all hybrid behavioral components to all 149 dog phenotypes surveyed in Darwin's dog for different P-value thresholds (P=1e-4, 1e-3, 1e-2, 5e-2, and 1e-1). To calculate the strength of gene association with the phenotype, we utilized MAGMA and obtained a P-value for each gene, which was then used as input for Fisher's exact test. At a P-value threshold of 0.05, we reported the top five dog

survey questions that overlap with hybrid behavior components (Table 4.8). Our findings revealed significant overlap between PC1+PC2+PC3 and the survey question “Dog is easily startled by unexpected contact with objects” (2.03E-6). Furthermore, we observed overlap of genomic associations to hybrid behavioral components and dog survey questions related to eating behaviors. Further analysis is needed to investigate the overlapping genes and variants.

Table 4.8. Gene enrichment overlap with dog phenotypes. We corrected the empirical P-value using Bonferroni Correction with a significant P-value falling below 5E-8.

Hybrid behavioral components	**P**	**Dog phenotype (Survey questions in Darwin’s dog project)**
PC1+PC2+PC3	**4.38E-07**	**DOG is as active as HE has been**
PC1+PC2+PC3	**2.03E-06**	**DOG is easily startled by unexpected contact with objects (e.g., tripping, brushing against a door frame).**
PC1+PC2+PC3	1.57E-04	DOG is gentle with small children
PC1+PC2+PC3	6.82E-04	What percent of DOG's diet is human food?
PC1+PC2+PC3	1.63E-03	DOG often gets human food
PC1	**4.52E-08**	**DOG damages doors, gates, or walls**
PC1	**1.56E-05**	**I think DOG could do with losing some weight**
PC1	2.81E-04	DOG is boisterous
PC1	4.51E-04	DOG tucks tail between legs and/or flattens ears when approached by an unfamiliar person
PC1	5.36E-04	DOG is easily startled by unexpected contact with objects (e.g., tripping, brushing against a door frame).

PC2	2.52E-04	DOG is relaxed when greeting people
PC2	5.34E-04	DOG inspects unfamiliar foods before deciding whether to eat them
PC2	1.55E-03	DOG seems to be hungry all the time
PC2	1.73E-03	DOG refuses to eat certain human foods that other dogs enjoy
PC2	3.23E-03	DOG is choosy about what HE eats
PC3	**3.29E-06**	**DOG enjoys playing with toys**
PC3	**2.87E-05**	**DOG gets bits of human food when we are eating**
PC3	7.43E-04	I am happy with DOG's weight
PC3	2.50E-03	DOG would eat anything
PC3	3.36E-03	DOG shows barrier aggression

We also examined genetic overlap for top associated genes for each pair comparing all hybrid behavioral components to three human psychiatric disorders including ASD (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium 2017), OCD (International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS) 2018), and ADHD (Demontis et al. 2019). We didn't observe significant overlap between any pair. We recognized that fisher's exact test is not the best approach to examine overlap across different species. We propose to use cross-trait cross-species LD score regression (Nagel et al. 2018; Bulik-Sullivan et al. 2015a, 2015b) to evaluate the overlap between canine behaviors and human behavioral disorders.

IV.5 Discussion

IV.5.1 Hybrids serve as a great natural model to study domestication

In this chapter, I applied admixture mapping and genome-wide association study (GWAS) to analyze social behaviors in wolf-dog hybrids, successfully identifying candidate regions on chromosomes 19 and 34. These regions are enriched with brain-expressed genes, particularly during early infancy. This research highlights the potential of using the captive wolf-dog hybrid population as a model to study behavioral changes following dog domestication and to potentially investigate neuropsychiatric pathways involved in early onset psychiatric disorders, such as autism spectrum disorder (ASD).

Dogs serve as a significant model species due to their phenotypic diversity and importance in studying the origins of human genetic diseases. The wolf-dog hybrid model can provide insights into psychiatric conditions that may not be adequately represented in dogs alone. Wolves have adapted to different environmental niches and developed characteristics distinct from dogs with genetic similarity. Investigating the pathways involved in this process could offer valuable information on human psychiatric disorders resulting from early development dysregulation.

Our study demonstrates that behavioral phenotyping of natural populations can be achieved with controlled environmental parameters. Despite environmental influences, we can extract phenotypic variation from raw behavioral

measurements, which can be explained by genetic variation and used in association studies. Utilizing advanced computational methods and integrating more data types, such as behavioral data on wolf, dogs, and wolfdog hybrid pups during critical periods, we can further explore the genetic alterations of animal domestication and evolution

IV.5.2 Phenotyping accuracy influence the statistical power to map behaviors

In order to obtain an accurate assessment of animals' behavior, we developed a "novelty test" that carefully defines various behavioral types. The test involves recording videos during a 10-minute period, which are then coded by two independent researchers to ensure greater accuracy. Despite the meticulous design and execution of the study, our ANOVA results indicate that certain measurements may still be influenced by environmental factors. It is possible that such factors are sometimes correlated with wolf ancestry, making it difficult to distinguish between the effects of environmental factors and those of ancestry (Table 4.9). Moreover, controlling the size of the enclosure and the total number of animals in the habitat is challenging, and even if possible, moving animals into the same enclosure or separating companion animals for testing will likely alter their response to the testing object (Moretti et al. 2015). Thus, I propose to apply a more practical approach, incorporating environmental factors as covariates in

the statistical test, which would help to eliminate their influence on the phenotype.

Table 4.9. Correlation among ancestry and environmental factors

	Age	Sex	Enclosure Size	Total animals in habitat
Wolf content	-0.15	-0.10	**0.25**	**-0.15**
Age		0.09	0.11	0.02
Sex			0.04	**0.26**
Enclosure Size				**0.38**

IV.5.3 Limited accessibility of canine functional data

We examined the candidate regions using human-derived functional data, such as gene expression data from GTEx and BrainSpan. However, we lose some information when we lift overed the genome from one species' genome to another, despite the similarities between humans and canines. Although we successfully lifted over 16,329 dog genes to human genes, we lost nearly one quarter of all genes in the process. To overcome this limitation, we need to generate more functional data for canines and canine brains so that we can investigate our candidate regions using functional data from the same species and eliminate the noise introduced by human data.

We matched the critical period of socialization, which can start at 2 weeks and end at 8 weeks, to human developmental stages from 6 months to 1 year old and investigated the temporal gene expression within each of the human developmental periods. However, we cannot verify this comparison since not enough studies have directly compared the brains of the two species at different developmental periods. To circumvent this problem, we can use gene expression data from dogs and directly explore the data at the same age as the critical period.

IV.6. Materials and methods

IV.6.1 Morphology collection

We collected 18 morphological traits by observing the animals in the enclosure. Two researchers were present and reached 100% agreement with each other on the trait a hybrid carries. When the animal was not visible to researchers during their visit, we sent out surveys for the owner to answer the questions regarding the specific morphological traits. We also asked owners for photographs of animals (full body image of the front and side) for two researchers to confirm the owners responses. The category and description of the eighteen morphological traits are summarized in Table 4.1.

IV.6.2 Novelty test to collect behavior data

Behaviors of individual subjects were recorded in response to three different conditions where objects were placed into subjects' home enclosures. During the first condition a single familiar object is placed 1-3 meters into the enclosure and then removed. In the second condition, a single novel object is placed in the same location and removed. Lastly, in the third condition the same familiar and novel objects are placed together in contact with each other in the sample location.

Enclosure setup

For each of the conditions, at least two Sony FDR-AX33 Digital 4K Video Cameras were placed around the outside perimeter, to capture animals' response. The animals were lured away from the entrance by a researcher or caretaker from outside the enclosure. After placement, subjects were given 10 minutes to interact with the object (s) without humans present. Once the condition had concluded the subjects were again lured away to the other side of the enclosure, and the object (s) were removed. All novel objects were immediately cleaned with Nature's Miracle Stain and Odor Remover spray, and allowed to dry. Between the conditions, animals were given a minimum of five minutes or until they returned to a relaxed state (e.g. behaviors expressed before presentations such as resting, exploring, drinking and the absence of distress

behaviors such as tail tucking, crouching, shaking, or attempts to hide) (Figure 4.4).

Behavior measurement

Each test lasted for 10 minutes, however, unavoidable interruption can occur in the field, thus all videos were clipped to match the shortest uninterrupted testing period of 578 seconds. The same test was performed with three different object arrangements, familiar object, novel object, or paired objects (familiar and novel object together) in the same enclosure. There are 42 measurements in total which can be categorized into five groups: investigation, contact, manipulation, fear, and out of sight. Some animals didn't start a specific behavior during the test and in this case we filled in the latency as 578 seconds and assumed that a behavior happens after that time (Table 4.10).

Table 4.10. Measurements on behavioral phenotype, "response to novelty".

Behavior phenotype			**Median (s)**	**SD (s)**	**Number of animals conduct this action**	**Correlation coefficient between measurement and wolf ancestry**
Object	Measure-ment	Action				
Paired	Latency	Fear	578.00	213.14	21	-0.41
Paired	Frequency	Fear	0.00	3.24	21	0.29
Novel	Frequency	Fear	0.00	3.61	30	0.29
Novel	Duration	Investigate	11.70	37.26	67	0.29

Novel	Frequency	Investigate	4.00	15.77	67	0.27
Paired	Frequency	Investigate	3.00	8.73	67	0.27
Familiar	Latency	Fear	578.00	196.35	16	-0.26
Familiar	Frequency	Fear	0.00	1.32	16	0.25
Paired	Duration	Investigate	6.65	28.75	67	0.23
Novel	Latency	Fear	578.00	244.49	30	-0.23
Novel	Latency	Contact	73.10	246.61	57	0.23
Familiar	Duration	Investigate	4.75	11.46	56	0.22
Paired	Latency	Out of Sight	10.80	230.01	63	-0.22
Paired	Frequency	Out of Sight	4.00	10.06	63	0.21
Paired	Latency	Contact	80.40	250.49	55	0.16
Novel	Latency	Manipulate	578.00	233.63	29	-0.15
Familiar	Frequency	Investigate	2.00	4.06	56	0.15
Familiar	Duration	Contact	9.10	34.77	53	-0.14
Familiar	Latency	Contact	61.30	256.26	53	0.13
Paired	Latency	Manipulate	578.00	180.68	15	-0.12
Familiar	Frequency	Contact	2.00	2.43	53	-0.11
Novel	Latency	Out of Sight	63.10	226.16	65	0.10
Novel	Duration	Out of Sight	86.00	200.06	65	-0.09
Novel	Frequency	Contact	3.00	4.92	57	0.08
Novel	Frequency	Manipulate	0.00	6.19	29	0.08
Paired	Latency	Investigate	30.60	207.12	67	-0.07
Paired	Frequency	Contact	2.00	3.73	55	0.07
Familiar	Duration	Manipulate	0.00	23.31	14	-0.07

Familiar	Latency	Investigate	56.85	252.27	56	0.07
Novel	Duration	Manipulate	0.00	50.24	29	0.06
Novel	Frequency	Out of Sight	3.00	5.74	65	-0.05
Familiar	Frequency	Manipulate	0.00	1.76	14	-0.05
Familiar	Frequency	Out of Sight	2.00	9.30	67	0.05
Novel	Latency	Investigate	41.50	205.38	67	0.04
Familiar	Duration	Out of Sight	51.90	218.45	67	-0.03
Paired	Duration	Out of Sight	75.55	201.80	63	-0.02
Paired	Duration	Contact	10.55	62.45	55	-0.02
Novel	Duration	Contact	33.75	73.29	57	-0.02
Familiar	Latency	Manipulate	578.00	186.69	14	-0.02
Paired	Duration	Manipulate	0.00	47.18	15	0.00
Familiar	Latency	Out of Sight	71.80	215.01	67	0.00
Paired	Frequency	Manipulate	0.00	3.16	15	0.00

Ethological coding

We used ethological coding (Hansen Wheat et al. 2022) to code behavioral data from video (Table 4.2). The program allows for multiple categories of behaviors; in our case two. Categories are made up of mutually exclusive behaviors within each category. The output provides a column for each behavioral category listing the specific behaviors selected for each second of the video. We used Cohen's Kappa to evaluate inter-observer reliability (Sun 2011) on 52% of the tests. This

statistic allows us to calculate reliability for each individual behavior in both their frequency and duration (Jansen et al. 2003; Hallgren 2012).

IV.6.3 Principal component analysis on behavioral traits

We collected behavior data on 42 measurements (Table 4.10). To reduce the dimension of the phenotype data and at the same time eliminate multicollinearity between measurements, we performed principal component analysis (PCA). There are 42 measurements in total which can be categorized into six groups: investigate, contact, manipulation, fear, and out of sight. The test ends at 578 seconds, and some animals didn't start a specific behavior during the test and in this case we filled in the latency as 578 seconds and assumed that a behavior happens after that time. Prior PCA, we performed log transformation on the raw measurement to shift the distribution closer to a normal distribution. We performed PCA on all 80 animals with behavioral phenotypes, then 69 wolf-dog hybrids with genotype information were kept for admixture mapping and SNP mapping. We selected the top three PCs that each carries at least 10% of the variance. We used the R function `princomp` to do the analysis.

IV.6.4 Genetic relationship matrix and narrow-sense heritability

We utilized the Genome-wide Complex Trait Analysis (GCTA) software tool (Yang et al. 2011) to assess genetic relationship matrix and linkage disequilibrium (LD)

scores in 250 kb regions. We set block size as 10,000kb with an overlap of 5,000kb between blocks and generate a genome-wide autosomal genetic relationship matrix (GRM) as well as multiple GRMs calculated from SNPs stratified into LD score quartiles ((Evans et al. 2018) with a relatedness cutoff of 0.75. For the estimation of heritability (h^2_{SNP}), we employed restricted maximum likelihood (REML) analysis using the GRM for all SNPs (`no-lds`). We noticed that the estimation of heritability can only be used as a guiding purpose due to the low accuracy of heritability estimation in sample size.

IV.6.5 Admixture mapping using MIXSCORE

We performed genome-wide association studies and admixture mappings for the morphological traits and behavioral phenotypes in genotyped hybrids using MIXSCORE (Pasaniuc et al. 2011), combining the signals from both SNP association and admixture association tests to increase statistical power for identifying causal SNPs. We incorporated wolf content estimated from the global ancestry calling approach into the method for admixture association using the `QADM` statistics, while for SNP association, we used the `QSNP1` statistics, which incorporate the local ancestry information of each SNP. Our study involved more than 12 million autosomal, biallelic SNPs with a minor allele frequency

greater than 0.05, and a `QSUM` score was calculated for each SNP to represent the association strength combining signals from both tests.

IV.6.6 Gene-set enrichment analysis

To assess the enrichment of sets of functionally related genes, we applied MAGMA v 1.09 (de Leeuw et al. 2015) on the admixture mapping QSUM statistics for 4 behavioral phenotypes, including three principal components, PC1, PC2, PC3 and the sum of three components `PC1+PC2+PC3`. MAGMA calculated gene-wide p-values by combining the p-values of all SNPs inside genes while accounting for gene size, number of SNPs in a gene, and LD. Using gene-based p-values, it tested for enrichment of association signals in genes belonging to the same set.

Gene sets were compiled from three main sources. First, top genes expressed in GTEx tissues (GTEx Consortium et al. 2017). We compiled a GTEx gene set by choosing the top 100 expressed genes in 13 brain subregions. Second, top genes expressed in brainspan brain regions at different developing periods (Miller et al. 2014). Brainspan data covers 13 developing stages from 8 post-conception weeks (PCW) to 40 years old and includes samples on 16 anatomical structures. A previous study utilizing brainspan data has performed hierarchical clustering of brain regions based on transcriptional similarity during fetal development and divided the brain regions into four groups which also reflect actual topographical proximity and functional segregation (Table 4.7). We

divided the developing period into two groups, the `prenatal, infancy` group includes samples 1 year old or younger, and the `childhood, adolescence, adulthood` group includes samples older than 1 year old. We merged samples from brain regions in each cluster at the same developing stage into one by taking the average as the expression level for each gene.

Third, curated neuropsychiatric genes. The autism spectrum disorder (ASD) gene set included 820 genes from the SFARI database which centers on genes implicated in ASD susceptibility (Abrahams et al. 2013). The obsessive-compulsive disorder (OCD) gene set comprised 62 manually curated genes from OCDB (Privitera et al. 2015), a database collecting genes, miRNAs and drugs for OCD. The schizophrenia (SCZ) gene set combined genes from two studies, the PGC2 GWAS in 2014 (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014) and the UK CLOZUK GWAS in 2018 (Pardiñas et al. 2018), which gave us a total of 371 genes.

IV.6.7 Genetic overlap tests for top associated genes

Fisher's exact tests were used to examine the overlap in the top associated genes across all hybrids phenotype to all dog behavioral phenotypes, and three human psychiatric disorders including ASD (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium 2017), ADHD (Demontis et al.

2019), and OCD (International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS) 2018).

Prior to this analysis, we performed gene-based analysis modified from gene-based genome-wide association analyses (Nagel et al. 2018) on 4 hybrid behavioral phenotypes, 149 dog behavioral phenotypes, and three human psychiatric disorders.

For hybrid behavioral phenotypes, we converted the QSUM scores (a chi-squared score with 2 dof) into p-value using `pchisq` R function, from the admixture association and SNP association test as input. For dog behavioral phenotypes, we used the P values from the genome-wide association studies as input. For human psychiatric disorders, we used P values from the genome-wide association studies as input. Because only summary statistics are publicly available and all the three studies are performed in the European population. We utilized the European genomes from the 1000 genome project (1000 Genomes Project Consortium et al. 2015) as the genotype input for MAGMA.

This gene-based analysis tests the joint signal of all SNPs in a gene with the phenotype, while accounting for LD between those SNPs, thus uncovering gene-level signals that may go unnoticed in SNP-based analysis. We performed this analysis on 16,329 genes homologous between humans and canines. We tested the overlap at different P-value thresholds, 0.1, 0.05, 0.01, 0.005, 0.0001. At each threshold, we counted the number of significant genes associated with

each of the two traits and got a table depicted below (Table 4.11). The categories in the table below are exclusive, meaning that a gene in cell A is not included in the count of cells B and C. We then performed one-side Fisher's exact tests to test the increase of overlapping genes associated with both traits. For more specific information on how this test was conducted, we refer to the supplementary information of Nagel et al. (Nagel et al. 2018).

Table 4.11 Contingency table for Fisher's exact test for gene overlap

A Number of genes < $P_{threshold}$ for both traits	B Number of genes < $P_{threshold}$ for trait 1 & > $P_{threshold}$ for trait 2
B Number of genes > $P_{threshold}$ for trait 1 & < $P_{threshold}$ for trait 2	D Number of genes > $P_{threshold}$ for both traits

CHAPTER V. CONCLUSION

V.1 Conclusion

My dissertation demonstrates the effectiveness of comparative genomics methodologies to investigate the genetic basis and evolutionary history of social development in wolf-dog hybrids. Chapter II emphasizes the value of evolutionary constraints in identifying genetic variations that contribute to the phenotypic differences and in comprehending their biological mechanisms. Chapter III delves into the population history of wolf-dog hybrids and examines potential selective stress in this captive population. In Chapter IV, I mapped social behaviors in wolfdog hybrids using SNP association and admixture mapping.

My research incorporating comparative genomics into analysis of wolfdog hybrids has two significant implications.

Firstly, it highlights the importance of species and population diversity in untangling complex phenotypes. With 240 genomes from 240 species sequenced and aligned, we generated a data source, phyloP score, which can predict functional importance at single base pair resolution. In turn, using one dog per breed or population, together with phyloP score, we deduced the genetics and phenotypes of Balto and his population. With 69 wolfdog hybrids which have diverse genetic backgrounds and cover a wide range of behavior phenotypes, we are able to map social behaviors at statistical significance in a remarkably small cohort of individuals. Having genomic data from distinct

populations, or species, together with a breadth of phenotypes can significantly decrease the samples we need to reach statistical power and can enhance our efforts to understand the genetic basis of various phenotypes in various populations, subspecies, or species.

As an extension of the wolfdog hybrid project, we have collected genomic and behavioral data in dog pups and propose to collect more data in wolf pups this summer. We will combine the critical period behavior data from the two subspecies, with behavior data from adult wolfdog hybrids, and population-level data for adult wolves and dogs to further investigate the genetic mechanisms of social behaviors. This diverse dataset, encompassing individuals from different populations and developmental stages, will significantly increase the power of trait mapping and improve our understanding of how genetics and environment interact in shaping social behavior.

Secondly, we demonstrated that wolfdog hybrids present a unique opportunity for trait mapping. Data from just 69 individuals enabled us to map intricate social behaviors within an admixed population for the first time. Besides our research on hybrids, the same approach has been employed in other species to examine phenotypes such as disease resistance (Kim and Rothschild 2014) and morphology (Brelsford et al. 2017) in human (Pino-Yanes et al. 2015) and non-human species (Kassahun et al. 2015).

Hybridized populations (comprising two species) or admixed populations (consisting of two populations from the same species, e.g., wolfdog hybrids) offer

exceptional opportunities in the field of evolutionary biology. Hybridized or admixed populations arise naturally in both ancient and recent times, under environmental or breeding stress, leaving genetic signatures that can be utilized for a wide array of studies beyond trait mapping (Brelsford et al. 2017), including evolutionary adaptation (Brauer et al. 2023), biodiversity conservation (Gese et al. 2015), and speciation (Payseur and Rieseberg 2016).

In conclusion, by comparing genomes across an extensive range of evolutionary time, we can more effectively detect unusual biological patterns and gain a deeper comprehension of evolution and the mechanisms underlying phenotypes ranging from diseases to social behavior.

V.2 Future directions

To delve deeper into the potential of utilizing wolfdog hybrids as a natural model for ASD, I suggest two specific future directions to advance the research on their social behaviors.

V.2.1 Mapping “response to stranger” behaviors in wolfdog hybrids

For behavior mapping, we utilized a "novelty test" to observe the animals' response to new stimuli. This is inspired by the fact that wolves have a longer flight distance than dogs to avoid danger in the wild, conversely, dogs have been domesticated for thousands of years and are often more tolerant of human presence, as they have been bred to work closely with humans in an environment with constant new stimuli (Zimen 1987; Karlsson et al. 2007).

Interspecies social bonding is another significant difference between wolves and dogs (Lenkei et al. 2020; Hansen Wheat et al. 2022; Morey 2010). While both can form social bonds, wolves may view humans as potential threats or competitors, and thus it takes more time and effort to establish social bonds with them. On the other hand, dogs often see humans as social companions and partners (Nagasawa et al. 2009).

To capture such behaviors, in addition to the "novelty test", we conducted a "stranger test", modified from the previous approach (Hansen Wheat et al. 2022; Foyer et al. 2014), by having a person stand still on the other side of the fence facing the animal for two minutes. We recorded videos and planned to code them into quantitative measurements using the same procedure as for the "novelty test". Our next step is to map the behaviors observed during the "stranger test” in wolfdogs. We will first evaluate the behavior spectrum in wolfdog hybrids categories into different levels of wolf ancestry load, then reduce the dimensions of measurements using principal component analysis, defining the phenotype for

top selected PCs, and then repeating the SNP association and admixture association mapping.

V.2.2 LD score regression to evaluate the genetic overlap between hybrid behaviors and human psychiatric conditions

The methods employed in this thesis to investigate the genetic overlap between hybrid behaviors and human psychiatric disorders have several limitations. Firstly, the small sample size of the hybrid study has hindered our ability to discover many causal genetic variants (Ball 2013). Additionally, cross-species comparisons are challenging due to the information loss that occurs when comparing one species to another. We attempted to utilize MAGMA to test for the enrichment of human psychiatric disorder-related genes but did not find any significant enrichment. However, this method requires a predefined set of genes to be tested, and previous GWAS work has shown that even with super large sample sizes, only part of heritability is enriched in the top candidate regions, with the missing heritability enriched in common variants that do not reach genome-wide significance (Eichler et al. 2010). Thus a set of predefined genes are not sufficient to test the overlap between the two. To compare whole-genome to whole-genome, we applied Fisher's exact test method (Nagel et al. 2018) and tested gene enrichment overlap at a high P-value threshold (less significant). However, due to the differences in population structure between admixed and

homogeneous populations (Bryc et al. 2010), as well as differences across species (Luu et al. 2020), we were unable to obtain a clear pattern of enrichment status.

To address this issue in future studies, we propose using LD score regression (Bulik-Sullivan et al. 2015b; Finucane et al. 2018, 2015) in combination with evolutionary constraints (Zoonomia Consortium 2020). Our approach involves lifting the candidate regions to humans and applying LD score regression to obtain the heritability of those human homologous regions. We will also filter our candidate loci by phyloP score and regress on those conserved candidate regions. We plan to expand our analysis to include hybrid-to-human, dog-to-human, and hybrid-to-dog comparisons. Additionally, we will lift human regions to dogs to complete our analysis. This new approach will allow us to overcome the limitations of the previous methods and improve our understanding of genome-to-genome comparisons across different species.

REFERENCES

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. *Nature* **526**: 68–74.

Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-Basu S, Packer A. 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **4**: 36.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**: 246.

Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, et al. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**: 246–251.

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. 2017. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* **8**: 21.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*. http://www.nature.com/nature/journal/v495/n7441/abs/nature11837.html.

Ball RD. 2013. Designing a GWAS: power, sample size, and data structure. *Methods Mol Biol* **1019**: 37–98.

Baranova J, Dragunas G, Botellho MCS, Ayub ALP, Bueno-Alves R, Alencar RR, Papaiz DD, Sogayar MC, Ulrich H, Correa RG. 2021. Autism Spectrum Disorder: Signaling Pathways and Prospective Therapeutic Targets. *Cell Mol Neurobiol* **41**: 619–649.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the impact of rare variants. *Genome Res* **23**: 1514–1521.

Brauer CJ, Sandoval-Castillo J, Gates K, Hammer MP, Unmack PJ, Bernatchez L, Beheregaray LB. 2023. Natural hybridization reduces vulnerability to climate change. *Nat Clim Chang* **13**: 282–289.

Brelsford A, Toews DPL, Irwin DE. 2017. Admixture mapping in a hybrid zone reveals

loci associated with avian feather coloration. *Proc Biol Sci* **284**. http://dx.doi.org/10.1098/rspb.2017.1106.

Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H. 2010. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proceedings of the National Academy of Sciences* **107**: 8954–8961.

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan L, et al. 2015a. An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**: 1236–1241.

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. 2015b. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**: 291–295.

Burkhard ME, Range F, Ward SJ, Robinson LM. 2022. Bonded by nature: Humans form equally strong and reciprocated bonds with similar raised dogs and wolves. *Front Psychol* **13**: 1044940.

Cai Z, Guldbrandtsen B, Lund MS, Sahana G. 2018. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. *BMC Genomics* **19**: 656.

Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS. 2007. A beta-defensin mutation causes black coat color in domestic dogs. *Science* **318**: 1418–1423.

Caniglia R, Fabbri E, Greco C, Galaverni M, Manghi L, Boitani L, Sforzi A, Randi E. 2013. Black coats in an admixed wolf× dog pack is melanism an indicator of hybridization in wolves? *Eur J Wildl Res* **59**: 543–555.

Ceballos FC, Hazelhurst S, Ramsay M. 2018. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics* **19**: 106.

Cesario JM, Almaidhan AA, Jeong J. 2016. Expression of forkhead box transcription factor genes Foxp1 and Foxp2 during jaw development. *Gene Expr Patterns* **20**: 111–119.

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

Curtis SW, Chang D, Lee MK, Shaffer JR, Indencleef K, Epstein MP, Cutler DJ, Murray JC, Feingold E, Beaty TH, et al. 2021. The PAX1 locus at 20p11 is a potential

genetic modifier for bilateral cleft lip. *HGG Adv* **2**. http://dx.doi.org/10.1016/j.xhgg.2021.100025.

de Leeuw CA, Mooij JM, Heskes T, Posthuma D. 2015. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**: e1004219.

Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, Baldursson G, Belliveau R, Bybjerg-Grauholm J, Bækvad-Hansen M, et al. 2019. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**: 63–75.

Ding Z-L, Oskarsson M, Ardalan A, Angleby H, Dahlgren L-G, Tepeli C, Kirkness E, Savolainen P, Zhang Y-P. 2012. Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity* **108**: 507–514.

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* **11**: 446–450.

Evans TM, Bira L, Gastelum JB, Weiss LT, Vanderford NL. 2018. Evidence for a mental health crisis in graduate education. *Nat Biotechnol* **36**: 282–284.

Fagan K, Crider A, Ahmed AO, Pillai A. 2017. Complement C3 Expression Is Decreased in Autism Spectrum Disorder Subjects and Contributes to Behavioral Deficits in Rodents. *Mol Neuropsychiatry* **3**: 19–27.

Faraway JJ. 2002. Practical Regression and Anova using R. http://csyue.nccu.edu.tw/Practical%20Regression%20and%20Anova%20using%20R.pdf (Accessed April 17, 2023).

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, et al. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**: 1228–1235.

Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh P-R, Lareau C, Shoresh N, et al. 2018. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**: 621–629.

Fong SL, Capra JA. 2021. Modeling the evolutionary architectures of transcribed human enhancer sequences reveals distinct origins, functions, and associations with human-trait variation. *Mol Biol Evol*. http://dx.doi.org/10.1093/molbev/msab138.

Foyer P, Bjällerhag N, Wilsson E, Jensen P. 2014. Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test. *Appl Anim Behav Sci* **155**: 93–100.

Freedman DG, King JA, Elliot O. 1961. Critical period in the social development of dogs. *Science* **133**: 1016–1017.

Gese EM, Knowlton FF, Adams JR, Beck K, Fuller TK, Murray DL, Steury TD, Stoskopf MK, Waddell WT, Waits LP. 2015. Managing hybridization of a recovering endangered species: The red wolf Canis rufus as a case study. *Curr Zool* **61**: 191–205.

Gong Y, Greenbaum J, Deng H-W. 2019. A statistical approach to fine-mapping for the identification of potential causal variants related to human intelligence. *J Hum Genet* **64**: 781–787.

Gopalakrishnan S, Sinding M-HS, Ramos-Madrigal J, Niemann J, Samaniego Castruita JA, Vieira FG, Carøe C, Montero M de M, Kuderna L, Serres A, et al. 2018. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. *Curr Biol* **28**: 3441–3449.e5.

Griesemer D, Xue JR, Reilly SK, Ulirsch JC, Kukreja K, Davis JR, Kanai M, Yang DK, Butts JC, Guney MH, et al. 2021. Genome-wide functional screen of 3'UTR variants uncovers causal variants for human disease and evolution. *Cell* **184**: 5247–5260.e19.

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. 2017. Genetic effects on gene expression across human tissues. *Nature* **550**: 204–213.

Hallgren KA. 2012. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol* **8**: 23–34.

Hansen Wheat C, Larsson L, Berner P, Temrin H. 2022. Human-directed attachment behavior in wolves suggests standing ancestral variation for human-dog attachment bonds. *Ecol Evol* **12**: e9299.

International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). 2018. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry* **23**: 1181–1188.

Jansen RG, Wiertz LF, Meyer ES, Noldus LPJJ. 2003. Reliability analysis of observational data: problems, solutions, and software implementation. *Behav Res Methods Instrum Comput* **35**: 391–399.

Kaeberlein M, Creevy KE, Promislow DEL. 2016. The dog aging project: translational geroscience in companion animals. *Mamm Genome* **27**: 279–288.

Karlsson EK, Lindblad-Toh K. 2008. Leader of the pack: gene mapping in dogs and other model organisms. *Nat Rev Genet* **9**: 713–725.

Karlsson J, Eriksson M, Liberg O. 2007. At what distance do wolves move away from an approaching human? *Can J Zool* **85**: 1193–1197.

Kassahun Y, Mattiangeli V, Ameni G, Hailu E, Aseffa A, Young DB, Hewinson RG, Vordermeier HM, Bradley DG. 2015. Admixture mapping of tuberculosis and pigmentation-related traits in an African-European hybrid cattle population. *Front Genet* **6**: 210.

Kazdoba TM, Leach PT, Yang M, Silverman JL, Solomon M, Crawley JN. 2016. Translational Mouse Models of Autism: Advancing Toward Pharmacological Therapeutics. *Curr Top Behav Neurosci* **28**: 1–52.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254.

Kim E-S, Rothschild MF. 2014. Genomic adaptation of admixed dairy cattle in East Africa. *Front Genet* **5**: 443.

Kircher M, Xiong C, Martin B, Schubach M, Inoue F, Bell RJA, Costello JF, Shendure J, Ahituv N. 2019. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat Commun* **10**: 3583.

Klin A, Micheletti M, Klaiman C, Shultz S, Constantino JN, Jones W. 2020. Affording autism an early brain development re-definition. *Dev Psychopathol* **32**: 1175–1189.

Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, et al. 2021. The Human Phenotype Ontology in 2021. *Nucleic Acids Res* **49**: D1207–D1217.

Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, Stahl PW, Piper PJ, Lingaas F, Fredholm M, et al. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci U S A* **109**: 8878–8883.

LeBlanc JJ, Fagiolini M. 2011. Autism: A “Critical Period” Disorder? *Neural Plast* **2011**. https://www.hindawi.com/journals/np/2011/921680/ (Accessed August 26, 2020).

Lenkei R, Újváry D, Bakos V, Faragó T. 2020. Adult, intensively socialized wolves show features of attachment behaviour to their handler. *Sci Rep* **10**: 17296.

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

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**: 476–482.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.

Li N, Stephens M. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**: 2213–2233.

Lord K. 2013. A Comparison of the Sensory Development of Wolves (Canis lupus lupus) and Dogs (Canis lupus familiaris). *Ethology* **119**: 110–120.

Lord KA, Larson G, Coppinger RP, Karlsson EK. 2020. The History of Farm Foxes Undermines the Animal Domestication Syndrome. *Trends Ecol Evol* **35**: 125–136.

Luo Y, Li X, Wang X, Gazal S, Mercader JM, 23 and Me Research Team, SIGMA Type 2 Diabetes Consortium, Neale BM, Florez JC, Auton A, et al. 2021. Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. *Hum Mol Genet* **30**: 1521–1534.

Luu P-L, Ong P-T, Dinh T-P, Clark SJ. 2020. Benchmark study comparing liftover tools for genome conversion of epigenome sequencing data. *NAR Genom Bioinform* **2**: lqaa054.

MacNulty DR, Mech LD, Smith DW. 2007. A Proposed Ethogram of Large-Carnivore Predatory Behavior, Exemplified by the Wolf. *J Mammal* **88**: 595–605.

McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, et al. 2008. Runs of homozygosity in European populations. *Am J Hum Genet* **83**: 359–372.

Megquier K, Turner-Maier J, Morrill K, Li X, Johnson J, Karlsson EK, London CA, Gardner HL. 2022. The genomic landscape of canine osteosarcoma cell lines reveals conserved structural complexity and pathway alterations. *PLoS One* **17**: e0274383.

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res* **49**: D394–D403.

Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, et al. 2014. Transcriptional landscape of the prenatal human brain. *Nature* **508**: 199–206.

Moravčíková N, Kasarda R, Židek R, Vostrý L, Vostrá-Vydrová H, Vašek J, Čílová D. 2021. Czechoslovakian Wolfdog Genomic Divergence from Its Ancestors Canis lupus, German Shepherd Dog, and Different Sheepdogs of European Origin. *Genes* **12**. http://dx.doi.org/10.3390/genes12060832.

Moretti L, Hentrup M, Kotrschal K, Range F. 2015. The influence of relationships on neophobia and exploration in wolves and dogs. *Anim Behav* **107**: 159–173.

Morey D. 2010. *Dogs: Domestication and the Development of a Social Bond*. Cambridge University Press.

Morrill K, Hekman J, Li X, McClure J, Logan B, Goodman L, Gao M, Dong Y, Alonso M, Carmichael E, et al. 2022. Ancestry-inclusive dog genomics challenges popular breed stereotypes. *Science* **376**: eabk0639.

Mulatinho MV, de Carvalho Serao CL, Scalco F, Hardekopf D, Pekova S, Mrasek K, Liehr T, Weise A, Rao N, Llerena JC Jr. 2012. Severe intellectual disability, omphalocele, hypospadia and high blood pressure associated to a deletion at 2q22.1q22.3: case report. *Mol Cytogenet* **5**: 30.

Muller SM. 2021. Companion Cyborgs: Untethering Discourses About Wolf-Hybrids. *Environmental Communication* **15**: 99–114.

Nagasawa M, Kikusui T, Onaka T, Ohta M. 2009. Dog's gaze at its owner increases owner's urinary oxytocin during social interaction. *Horm Behav* **55**: 434–441.

Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S. 2018. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat Commun* **9**: 905.

Noh HJ, Tang R, Flannick J, Dushlaine CO, Swofford R, Howrigan D, Genereux DP, Johnson J, van Grootheest G, Grunblatt E, et al. 2017. Integrating evolutionary and regulatory information with multispecies approach implicates genes and pathways in obsessive-compulsive disorder. *Nat Commun* 107193.

Ostrander EA, Wayne RK. 2005. The canine genome. *Genome Res* **15**: 1706–1716.

Pageat P. 2005. Assessing prolactinemia in anxious dogs (Canis familiaris): interest in diagnostic value and use in the selection of the most appropriate psychotropic drug. *Current Issues and Research in Veterinary Behavioral Medicine* 155.

Pankratov V, Yunusbaeva M, Ryakhovsky S, Zarodniuk M, Estonian Biobank Research Team, Yunusbayev B. 2022. Prioritizing autoimmunity risk variants for functional analyses by fine-mapping mutations under natural selection. *Nat Commun* **13**: 7069.

Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, et al. 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**: 381–389.

Pasaniuc B, Zaitlen N, Lettre G, Chen GK, Tandon A, Kao WHL, Ruczinski I, Fornage M, Siscovick DS, Zhu X, et al. 2011. Enhanced statistical tests for GWAS in admixed populations: assessment using African Americans from CARe and a Breast Cancer Consortium. *PLoS Genet* **7**: e1001371.

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, et al. 2004. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* **74**: 979–1000.

Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Mol Ecol* **25**: 2337–2360.

Pendleton AL, Shen F, Taravella AM, Emery S, Veeramah KR, Boyko AR, Kidd JM. 2018. Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. *BMC Biol* **16**: 64.

Perri AR, Mitchell KJ, Mouton A, Álvarez-Carretero S, Hulme-Beaman A, Haile J, Jamieson A, Meachen J, Lin AT, Schubert BW, et al. 2021. Dire wolves were the last of an ancient New World canid lineage. *Nature* **591**: 87–91.

Persson ME, Wright D, Roth LSV, Batakis P, Jensen P. 2016. Genomic Regions Associated With Interspecies Communication in Dogs Contain Genes Related to Human Social Disorders. *Sci Rep* **6**: 33439.

Pilot M, Moura AE, Okhlopkov IM, Mamaev NV, Alagaili AN, Mohammed OB, Yavruyan EG, Manaseryan NH, Hayrapetyan V, Kopaliani N, et al. 2019. Global Phylogeographic and Admixture Patterns in Grey Wolves and Genetic Legacy of An Ancient Siberian Lineage. *Sci Rep* **9**: 17328.

Pino-Yanes M, Gignoux CR, Galanter JM, Levin AM, Campbell CD, Eng C, Huntsman S, Nishimura KK, Gourraud P-A, Mohajeri K, et al. 2015. Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. *J Allergy Clin Immunol* **135**: 1502–1510.

Pörtl D, Jung C. 2017. Is dog domestication due to epigenetic modulation in brain? *Dog behavior* **3**: 21–32.

Privitera AP, Distefano R, Wefer HA, Ferro A, Pulvirenti A, Giugno R. 2015. OCDB: a database collecting genes, miRNAs and drugs for obsessive-compulsive disorder. *Database* **2015**: bav069.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575.

Quesnel-Vallières M, Weatheritt RJ, Cordes SP, Blencowe BJ. 2019. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. *Nat Rev Genet* **20**: 51–63.

Randi E, Lucchini V. 2002. Detecting rare introgression of domestic dog genes into wild wolf (Canis lupus) populations by Bayesian admixture analyses of microsatellite variation. *Conserv Genet* **3**: 29–43.

Rezwani M, Pourfathollah AA, Noorbakhsh F. 2022. rbioapi: user-friendly R interface to biologic web services' API. *Bioinformatics* **38**: 2952–2953.

Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J, Maller J, Samocha KE, Sanders SJ, Ripke S, et al. 2016. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet* **48**: 552–555.

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

Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. 2008. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet* **83**: 610–615.

Romero C, Werme J, Jansen PR, Gelernter J, Stein MB, Levey D, Polimanti R, de Leeuw C, Posthuma D, Nagel M, et al. 2022. Exploring the genetic overlap between twelve psychiatric disorders. *Nat Genet* **54**: 1795–1802.

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, et al. 2000. Comparative genomics of the eukaryotes. *Science* **287**: 2204–2215.

Saetre P, Lindberg J, Leonard JA, Olsson K, Pettersson U, Ellegren H, Bergström TF, Vilà C, Jazin E. 2004. From wild wolf to domestic dog: gene expression changes in the brain. *Molecular Brain Research* **126**: 198–206.

Sams AJ, Boyko AR. 2019. Fine-Scale Resolution of Runs of Homozygosity Reveal Patterns of Inbreeding and Substantial Overlap with Recessive Disease Genotypes in Domestic Dogs. *G3* **9**: 117–123.

Savolainen P, Zhang Y-P, Luo J, Lundeberg J, Leitner T. 2002. Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**: 1610–1613.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**: 421–427.

Schweizer RM, Durvasula A, Smith J, Vohr SH, Stahler DR, Galaverni M, Thalmann O, Smith DW, Randi E, Ostrander EA, et al. 2018. Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. *Mol Biol Evol* **35**: 1190–1209.

Scott JP, Marston MV. 1950. Critical periods affecting the development of normal and mal-adjustive social behavior of puppies. *J Genet Psychol* **77**: 25–60.

Shindyapina AV, Zenin AA, Tarkhov AE, Santesmasses D, Fedichev PO, Gladyshev VN. 2020. Germline burden of rare damaging variants negatively affects human healthspan and lifespan. *Elife* **9**. http://dx.doi.org/10.7554/eLife.53449.

Shultz S, Klin A, Jones W. 2018. Neonatal Transitions in Social Behavior and Their Implications for Autism. *Trends Cogn Sci* **22**: 452–469.

Sinding M-HS, Gopalakrishnan S, Raundrup K, Dalén L, Threlfall J, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium, et al. 2021. The genome sequence of the grey wolf, Canis lupus Linnaeus 1758. *Wellcome Open Res* **6**: 310.

Sommese A, Valsecchi P, Pelosi A, Prato-Previde E. 2021. Comparing behavioural characteristics of Czechoslovakian Wolfdogs, German shepherds and Labrador retrievers in Italy and the Czech Republic. *Appl Anim Behav Sci* **237**: 105300.

Starkey MP, Scase TJ, Mellersh CS, Murphy S. 2005. Dogs really are man's best friend—Canine genomics has applications in veterinary and human medicine! *Brief Funct Genomics* **4**: 112–128.

Sthle L, Wold S. 1989. Analysis of variance (ANOVA). *Chemometrics Intellig Lab Syst* **6**: 259–272.

Sun S. 2011. Meta-analysis of Cohen's kappa. *Health Serv Outcomes Res Methodol* **11**: 145–163.

Tang R, Noh HJ, Wang D, Sigurdsson S, Swofford R, Perloski M, Duxbury M, Patterson EE, Albright J, Castelhano M, et al. 2014. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biol* **15**: R25.

Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, Andersen KG, Mikkelsen TS, Lander ES, Schaffner SF, et al. 2016. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell* **165**: 1519–1529.

Thomson VA, Lebrasseur O, Austin JJ, Hunt TL, Burney DA, Denham T, Rawlence NJ, Wood JR, Gongora J, Girdland Flink L, et al. 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. *Proc Natl Acad Sci U S A* **111**: 4826–4831.

Topál J, Román V, Turcsán B. 2019. The dog (Canis familiaris) as a translational model of autism: It is high time we move from promise to reality. *WIREs Cogn Sci* **10**: 509.

Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T. 2021. Genome-wide association studies. *Nature Reviews Methods Primers* **1**: 1–21.

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK. 1997. Multiple and ancient origins of the domestic dog. *Science* **276**: 1687–1689.

Vila C, Wayne R. 1999. Hybridization between wolves and dog. *Conserv Biol* **13**: 195–198.

vonHoldt B, Fan Z, Ortega-Del Vecchyo D, Wayne RK. 2017. EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. *PeerJ* **5**: e3522.

Wasik K, Berisa T, Pickrell JK, Li JH, Fraser DJ, King K, Cox C. 2021. Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genomics* **22**: 197.

Weir BS, Cockerham CC. 1984. ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. *Evolution* **38**: 1358–1370.

Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, Schoech AP, van de Geijn B, Reshef Y, Márquez-Luna C, et al. 2020. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat Genet* **52**: 1355–1363.

Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, et al. 2013. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**: 997–1007.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**: 76–82.

Zapata I, Serpell JA, Alvarez CE. 2016. Genetic mapping of canine fear and aggression. *BMC Genomics* **17**: 572.

Zimen E. 1987. Ontogeny of approach and flight behavior towards humans in wolves, poodles and wolf-poodle hybrids. *Man and wolf* 275–292.

Zoonomia Consortium. 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**: 240–245.

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. Nature 526: 68–74.

Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-Basu S, Packer A. 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism 4: 36.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12: 246.

Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, et al. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature 587: 246–251.

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. 2017. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism 8: 21.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. http://www.nature.com/nature/journal/v495/n7441/abs/nature11837.html.

Ball RD. 2013. Designing a GWAS: power, sample size, and data structure. Methods Mol Biol 1019: 37–98.

Baranova J, Dragunas G, Botellho MCS, Ayub ALP, Bueno-Alves R, Alencar RR, Papaiz DD, Sogayar MC, Ulrich H, Correa RG. 2021. Autism Spectrum Disorder: Signaling

Pathways and Prospective Therapeutic Targets. Cell Mol Neurobiol 41: 619–649.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the impact of rare variants. Genome Res 23: 1514–1521.

Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H. 2010. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proceedings of the National Academy of Sciences 107: 8954–8961.

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan L, et al. 2015a. An atlas of genetic correlations across human diseases and traits. Nat Genet 47: 1236–1241.

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. 2015b. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47: 291–295.

Burkhard ME, Range F, Ward SJ, Robinson LM. 2022. Bonded by nature: Humans form equally strong and reciprocated bonds with similar raised dogs and wolves. Front Psychol 13: 1044940.

Cai Z, Guldbrandtsen B, Lund MS, Sahana G. 2018. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. BMC Genomics 19: 656.

Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS. 2007. A beta-defensin mutation causes black coat color in domestic dogs. Science 318: 1418–1423.

Caniglia R, Fabbri E, Greco C, Galaverni M, Manghi L, Boitani L, Sforzi A, Randi E. 2013. Black coats in an admixed wolf× dog pack is melanism an indicator of hybridization in wolves? Eur J Wildl Res 59: 543–555.

Ceballos FC, Hazelhurst S, Ramsay M. 2018. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. BMC Genomics 19: 106.

Cesario JM, Almaidhan AA, Jeong J. 2016. Expression of forkhead box transcription factor genes Foxp1 and Foxp2 during jaw development. Gene Expr Patterns 20: 111–119.

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

Curtis SW, Chang D, Lee MK, Shaffer JR, Indencleef K, Epstein MP, Cutler DJ, Murray JC, Feingold E, Beaty TH, et al. 2021. The PAX1 locus at 20p11 is a potential genetic modifier for bilateral cleft lip. HGG Adv 2. http://dx.doi.org/10.1016/j.xhgg.2021.100025.

de Leeuw CA, Mooij JM, Heskes T, Posthuma D. 2015. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 11: e1004219.

Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, Baldursson G, Belliveau R, Bybjerg-Grauholm J, Bækvad-Hansen M, et al. 2019. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet 51: 63–75.

Ding Z-L, Oskarsson M, Ardalan A, Angleby H, Dahlgren L-G, Tepeli C, Kirkness E, Savolainen P, Zhang Y-P. 2012. Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. Heredity 108: 507–514.

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 11: 446–450.

Evans TM, Bira L, Gastelum JB, Weiss LT, Vanderford NL. 2018. Evidence for a mental health crisis in graduate education. Nat Biotechnol 36: 282–284.

Fagan K, Crider A, Ahmed AO, Pillai A. 2017. Complement C3 Expression Is Decreased in Autism Spectrum Disorder Subjects and Contributes to Behavioral Deficits in Rodents. Mol Neuropsychiatry 3: 19–27.

Faraway JJ. 2002. Practical Regression and Anova using R. http://csyue.nccu.edu.tw/Practical%20Regression%20and%20Anova%20using%20R.pdf (Accessed April 17, 2023).

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, et al. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47: 1228–1235.

Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh P-R, Lareau C, Shoresh N, et al. 2018. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet 50: 621–629.

Fong SL, Capra JA. 2021. Modeling the evolutionary architectures of transcribed human enhancer sequences reveals distinct origins, functions, and associations with human-trait variation. Mol Biol Evol. http://dx.doi.org/10.1093/molbev/msab138.

Foyer P, Bjällerhag N, Wilsson E, Jensen P. 2014. Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test. Appl Anim Behav Sci 155: 93–100.

Freedman DG, King JA, Elliot O. 1961. Critical period in the social development of dogs. Science 133: 1016–1017.

Gong Y, Greenbaum J, Deng H-W. 2019. A statistical approach to fine-mapping for the identification of potential causal variants related to human intelligence. J Hum Genet 64: 781–787.

Gopalakrishnan S, Sinding M-HS, Ramos-Madrigal J, Niemann J, Samaniego Castruita JA, Vieira FG, Carøe C, Montero M de M, Kuderna L, Serres A, et al. 2018. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. Curr Biol 28: 3441–3449.e5.

Griesemer D, Xue JR, Reilly SK, Ulirsch JC, Kukreja K, Davis JR, Kanai M, Yang DK, Butts JC, Guney MH, et al. 2021. Genome-wide functional screen of 3'UTR variants uncovers causal variants for human disease and evolution. Cell 184: 5247–5260.e19.

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. 2017. Genetic effects on gene expression across human tissues. Nature 550: 204–213.

Hallgren KA. 2012. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol 8: 23–34.

Hansen Wheat C, Larsson L, Berner P, Temrin H. 2022. Human-directed attachment behavior in wolves suggests standing ancestral variation for human-dog attachment bonds. Ecol Evol 12: e9299.

International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). 2018. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. Mol Psychiatry 23: 1181–1188.

Jansen RG, Wiertz LF, Meyer ES, Noldus LPJJ. 2003. Reliability analysis of observational data: problems, solutions, and software implementation. Behav Res Methods Instrum Comput 35: 391–399.

Kaeberlein M, Creevy KE, Promislow DEL. 2016. The dog aging project: translational geroscience in companion animals. Mamm Genome 27: 279–288.

Karlsson EK, Lindblad-Toh K. 2008. Leader of the pack: gene mapping in dogs and other model organisms. Nat Rev Genet 9: 713–725.

Karlsson J, Eriksson M, Liberg O. 2007. At what distance do wolves move away from an approaching human? Can J Zool 85: 1193–1197.

Kassahun Y, Mattiangeli V, Ameni G, Hailu E, Aseffa A, Young DB, Hewinson RG,

Vordermeier HM, Bradley DG. 2015. Admixture mapping of tuberculosis and pigmentation-related traits in an African-European hybrid cattle population. Front Genet 6: 210.

Kazdoba TM, Leach PT, Yang M, Silverman JL, Solomon M, Crawley JN. 2016. Translational Mouse Models of Autism: Advancing Toward Pharmacological Therapeutics. Curr Top Behav Neurosci 28: 1–52.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423: 241–254.

Kircher M, Xiong C, Martin B, Schubach M, Inoue F, Bell RJA, Costello JF, Shendure J, Ahituv N. 2019. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. Nat Commun 10: 3583.

Klin A, Micheletti M, Klaiman C, Shultz S, Constantino JN, Jones W. 2020. Affording autism an early brain development re-definition. Dev Psychopathol 32: 1175–1189.

Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, et al. 2021. The Human Phenotype Ontology in 2021. Nucleic Acids Res 49: D1207–D1217.

Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, Stahl PW, Piper PJ, Lingaas F, Fredholm M, et al. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proc Natl Acad Sci U S A 109: 8878–8883.

LeBlanc JJ, Fagiolini M. 2011. Autism: A “Critical Period” Disorder? Neural Plast 2011. https://www.hindawi.com/journals/np/2011/921680/ (Accessed August 26, 2020).

Lenkei R, Újváry D, Bakos V, Faragó T. 2020. Adult, intensively socialized wolves show features of attachment behaviour to their handler. Sci Rep 10: 17296.

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

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478: 476–482.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803–819.

Li N, Stephens M. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165: 2213–2233.

Lord K. 2013. A Comparison of the Sensory Development of Wolves (Canis lupus lupus) and Dogs (Canis lupus familiaris). Ethology 119: 110–120.

Lord KA, Larson G, Coppinger RP, Karlsson EK. 2020. The History of Farm Foxes Undermines the Animal Domestication Syndrome. Trends Ecol Evol 35: 125–136.

Luu P-L, Ong P-T, Dinh T-P, Clark SJ. 2020. Benchmark study comparing liftover tools for genome conversion of epigenome sequencing data. NAR Genom Bioinform 2: lqaa054.

MacNulty DR, Mech LD, Smith DW. 2007. A Proposed Ethogram of Large-Carnivore Predatory Behavior, Exemplified by the Wolf. J Mammal 88: 595–605.

McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, et al. 2008. Runs of homozygosity in European populations. Am J Hum Genet 83: 359–372.

Megquier K, Turner-Maier J, Morrill K, Li X, Johnson J, Karlsson EK, London CA, Gardner HL. 2022. The genomic landscape of canine osteosarcoma cell lines reveals conserved structural complexity and pathway alterations. PLoS One 17: e0274383.

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res 49: D394–D403.

Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, et al. 2014. Transcriptional landscape of the prenatal human brain. Nature 508: 199–206.

Moravčíková N, Kasarda R, Židek R, Vostrý L, Vostrá-Vydrová H, Vašek J, Čílová D. 2021. Czechoslovakian Wolfdog Genomic Divergence from Its Ancestors Canis lupus, German Shepherd Dog, and Different Sheepdogs of European Origin. Genes 12. http://dx.doi.org/10.3390/genes12060832.

Moretti L, Hentrup M, Kotrschal K, Range F. 2015. The influence of relationships on neophobia and exploration in wolves and dogs. Anim Behav 107: 159–173.

Morey D. 2010. Dogs: Domestication and the Development of a Social Bond. Cambridge University Press.

Morrill K, Hekman J, Li X, McClure J, Logan B, Goodman L, Gao M, Dong Y, Alonso M, Carmichael E, et al. 2022. Ancestry-inclusive dog genomics challenges popular breed stereotypes. Science 376: eabk0639.

Mulatinho MV, de Carvalho Serao CL, Scalco F, Hardekopf D, Pekova S, Mrasek K, Liehr T, Weise A, Rao N, Llerena JC Jr. 2012. Severe intellectual disability, omphalocele, hypospadia and high blood pressure associated to a deletion at 2q22.1q22.3: case report. Mol Cytogenet 5: 30.

Muller SM. 2021. Companion Cyborgs: Untethering Discourses About Wolf-Hybrids. Environmental Communication 15: 99–114.

Nagasawa M, Kikusui T, Onaka T, Ohta M. 2009. Dog's gaze at its owner increases owner's urinary oxytocin during social interaction. Horm Behav 55: 434–441.

Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S. 2018. Item-level analyses reveal genetic heterogeneity in neuroticism. Nat Commun 9: 905.

Noh HJ, Tang R, Flannick J, Dushlaine CO, Swofford R, Howrigan D, Genereux DP, Johnson J, van Grootheest G, Grunblatt E, et al. 2017. Integrating evolutionary and regulatory information with multispecies approach implicates genes and pathways in obsessive-compulsive disorder. Nat Commun 107193.

Ostrander EA, Wayne RK. 2005. The canine genome. Genome Res 15: 1706–1716.

Pageat P. 2005. Assessing prolactinemia in anxious dogs (Canis familiaris): interest in diagnostic value and use in the selection of the most appropriate psychotropic drug. Current Issues and Research in Veterinary Behavioral Medicine 155.

Pankratov V, Yunusbaeva M, Ryakhovsky S, Zarodniuk M, Estonian Biobank Research Team, Yunusbayev B. 2022. Prioritizing autoimmunity risk variants for functional analyses by fine-mapping mutations under natural selection. Nat Commun 13: 7069.

Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, et al. 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50: 381–389.

Pasaniuc B, Zaitlen N, Lettre G, Chen GK, Tandon A, Kao WHL, Ruczinski I, Fornage M, Siscovick DS, Zhu X, et al. 2011. Enhanced statistical tests for GWAS in admixed populations: assessment using African Americans from CARe and a Breast Cancer Consortium. PLoS Genet 7: e1001371.

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, et al. 2004. Methods for high-density admixture mapping of disease genes. Am J Hum Genet 74: 979–1000.

Pendleton AL, Shen F, Taravella AM, Emery S, Veeramah KR, Boyko AR, Kidd JM. 2018. Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. BMC Biol 16: 64.

Perri AR, Mitchell KJ, Mouton A, Álvarez-Carretero S, Hulme-Beaman A, Haile J, Jamieson A, Meachen J, Lin AT, Schubert BW, et al. 2021. Dire wolves were the last of an ancient New World canid lineage. Nature 591: 87–91.

Persson ME, Wright D, Roth LSV, Batakis P, Jensen P. 2016. Genomic Regions Associated With Interspecies Communication in Dogs Contain Genes Related to Human Social Disorders. Sci Rep 6: 33439.

Pilot M, Moura AE, Okhlopkov IM, Mamaev NV, Alagaili AN, Mohammed OB, Yavruyan EG, Manaseryan NH, Hayrapetyan V, Kopaliani N, et al. 2019. Global

Phylogeographic and Admixture Patterns in Grey Wolves and Genetic Legacy of An Ancient Siberian Lineage. Sci Rep 9: 17328.

Pino-Yanes M, Gignoux CR, Galanter JM, Levin AM, Campbell CD, Eng C, Huntsman S, Nishimura KK, Gourraud P-A, Mohajeri K, et al. 2015. Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. J Allergy Clin Immunol 135: 1502–1510.

Pörtl D, Jung C. 2017. Is dog domestication due to epigenetic modulation in brain? Dog behavior 3: 21–32.

Privitera AP, Distefano R, Wefer HA, Ferro A, Pulvirenti A, Giugno R. 2015. OCDB: a database collecting genes, miRNAs and drugs for obsessive-compulsive disorder. Database 2015: bav069.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.

Quesnel-Vallières M, Weatheritt RJ, Cordes SP, Blencowe BJ. 2019. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. Nat Rev Genet 20: 51–63.

Randi E, Lucchini V. 2002. Detecting rare introgression of domestic dog genes into wild wolf (Canis lupus) populations by Bayesian admixture analyses of microsatellite variation. Conserv Genet 3: 29–43.

Rezwani M, Pourfathollah AA, Noorbakhsh F. 2022. rbioapi: user-friendly R interface to biologic web services' API. Bioinformatics 38: 2952–2953.

Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J, Maller J, Samocha KE, Sanders SJ, Ripke S, et al. 2016. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nat Genet 48: 552–555.

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

Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. 2008. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 83: 610–615.

Romero C, Werme J, Jansen PR, Gelernter J, Stein MB, Levey D, Polimanti R, de Leeuw C, Posthuma D, Nagel M, et al. 2022. Exploring the genetic overlap between twelve psychiatric disorders. Nat Genet 54: 1795–1802.

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, et al. 2000. Comparative genomics of the

eukaryotes. Science 287: 2204–2215.

Saetre P, Lindberg J, Leonard JA, Olsson K, Pettersson U, Ellegren H, Bergström TF, Vilà C, Jazin E. 2004. From wild wolf to domestic dog: gene expression changes in the brain. Molecular Brain Research 126: 198–206.

Sams AJ, Boyko AR. 2019. Fine-Scale Resolution of Runs of Homozygosity Reveal Patterns of Inbreeding and Substantial Overlap with Recessive Disease Genotypes in Domestic Dogs. G3 9: 117–123.

Savolainen P, Zhang Y-P, Luo J, Lundeberg J, Leitner T. 2002. Genetic evidence for an East Asian origin of domestic dogs. Science 298: 1610–1613.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511: 421–427.

Schweizer RM, Durvasula A, Smith J, Vohr SH, Stahler DR, Galaverni M, Thalmann O, Smith DW, Randi E, Ostrander EA, et al. 2018. Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. Mol Biol Evol 35: 1190–1209.

Scott JP, Marston MV. 1950. Critical periods affecting the development of normal and mal-adjustive social behavior of puppies. J Genet Psychol 77: 25–60.

Shindyapina AV, Zenin AA, Tarkhov AE, Santesmasses D, Fedichev PO, Gladyshev VN. 2020. Germline burden of rare damaging variants negatively affects human healthspan and lifespan. Elife 9. http://dx.doi.org/10.7554/eLife.53449.

Shultz S, Klin A, Jones W. 2018. Neonatal Transitions in Social Behavior and Their Implications for Autism. Trends Cogn Sci 22: 452–469.

Sinding M-HS, Gopalakrishnan S, Raundrup K, Dalén L, Threlfall J, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium, et al. 2021. The genome sequence of the grey wolf, Canis lupus Linnaeus 1758. Wellcome Open Res 6: 310.

Sommese A, Valsecchi P, Pelosi A, Prato-Previde E. 2021. Comparing behavioural characteristics of Czechoslovakian Wolfdogs, German shepherds and Labrador retrievers in Italy and the Czech Republic. Appl Anim Behav Sci 237: 105300.

Starkey MP, Scase TJ, Mellersh CS, Murphy S. 2005. Dogs really are man's best friend—Canine genomics has applications in veterinary and human medicine! Brief Funct Genomics 4: 112–128.

Sthle L, Wold S. 1989. Analysis of variance (ANOVA). Chemometrics Intellig Lab Syst 6: 259–272.

Sun S. 2011. Meta-analysis of Cohen's kappa. Health Serv Outcomes Res Methodol 11: 145–163.

Tang R, Noh HJ, Wang D, Sigurdsson S, Swofford R, Perloski M, Duxbury M, Patterson EE, Albright J, Castelhano M, et al. 2014. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. Genome Biol 15: R25.

Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, Andersen KG, Mikkelsen TS, Lander ES, Schaffner SF, et al. 2016. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell 165: 1519–1529.

Thomson VA, Lebrasseur O, Austin JJ, Hunt TL, Burney DA, Denham T, Rawlence NJ, Wood JR, Gongora J, Girdland Flink L, et al. 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. Proc Natl Acad Sci U S A 111: 4826–4831.

Topál J, Román V, Turcsán B. 2019. The dog (Canis familiaris) as a translational model of autism: It is high time we move from promise to reality. WIREs Cogn Sci 10: 509.

Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T. 2021. Genome-wide association studies. Nature Reviews Methods Primers 1: 1–21.

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK. 1997. Multiple and ancient origins of the domestic dog. Science 276: 1687–1689.

Vila C, Wayne R. 1999. Hybridization between wolves and dog. Conserv Biol 13: 195–198.

vonHoldt B, Fan Z, Ortega-Del Vecchyo D, Wayne RK. 2017. EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. PeerJ 5: e3522.

Wasik K, Berisa T, Pickrell JK, Li JH, Fraser DJ, King K, Cox C. 2021. Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. BMC Genomics 22: 197.

Weir BS, Cockerham CC. 1984. ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. Evolution 38: 1358–1370.

Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, Schoech AP, van de Geijn B, Reshef Y, Márquez-Luna C, et al. 2020. Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nat Genet 52: 1355–1363.

Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, et al. 2013. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155: 997–1007.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88: 76–82.

Zapata I, Serpell JA, Alvarez CE. 2016. Genetic mapping of canine fear and aggression. BMC Genomics 17: 572.

Zimen E. 1987. Ontogeny of approach and flight behavior towards humans in wolves, poodles and wolf-poodle hybrids. Man and wolf 275–292.

Zoonomia Consortium. 2020. A comparative genomics multitool for scientific discovery and conservation. Nature 587: 240–245.

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. Nature 526: 68–74.

Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-Basu S, Packer A. 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism 4: 36.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12: 246.

Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, et al. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature 587: 246–251.

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. 2017. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism 8: 21.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. http://www.nature.com/nature/journal/v495/n7441/abs/nature11837.html.

Ball RD. 2013. Designing a GWAS: power, sample size, and data structure. Methods Mol Biol 1019: 37–98.

Baranova J, Dragunas G, Botellho MCS, Ayub ALP, Bueno-Alves R, Alencar RR, Papaiz DD, Sogayar MC, Ulrich H, Correa RG. 2021. Autism Spectrum Disorder: Signaling Pathways and Prospective Therapeutic Targets. Cell Mol Neurobiol 41: 619–649.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the impact of rare variants. Genome Res 23: 1514–1521.

Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H. 2010. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proceedings of the National Academy of Sciences 107: 8954–8961.

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan L, et al. 2015a. An atlas of genetic correlations across human diseases and traits. Nat Genet 47: 1236–1241.

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. 2015b. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47: 291–295.

Burkhard ME, Range F, Ward SJ, Robinson LM. 2022. Bonded by nature: Humans form equally strong and reciprocated bonds with similar raised dogs and wolves. Front Psychol 13: 1044940.

Cai Z, Guldbrandtsen B, Lund MS, Sahana G. 2018. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. BMC Genomics 19: 656.

Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS. 2007. A beta-defensin mutation causes black coat color in domestic dogs. Science 318: 1418–1423.

Caniglia R, Fabbri E, Greco C, Galaverni M, Manghi L, Boitani L, Sforzi A, Randi E. 2013. Black coats in an admixed wolf× dog pack is melanism an indicator of hybridization in wolves? Eur J Wildl Res 59: 543–555.

Ceballos FC, Hazelhurst S, Ramsay M. 2018. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. BMC Genomics 19: 106.

Cesario JM, Almaidhan AA, Jeong J. 2016. Expression of forkhead box transcription factor genes Foxp1 and Foxp2 during jaw development. Gene Expr Patterns 20: 111–119.

Curtis SW, Chang D, Lee MK, Shaffer JR, Indencleef K, Epstein MP, Cutler DJ, Murray JC, Feingold E, Beaty TH, et al. 2021. The PAX1 locus at 20p11 is a potential genetic modifier for bilateral cleft lip. HGG Adv 2. http://dx.doi.org/10.1016/j.xhgg.2021.100025.

de Leeuw CA, Mooij JM, Heskes T, Posthuma D. 2015. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 11: e1004219.

Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, Baldursson G, Belliveau R, Bybjerg-Grauholm J, Bækvad-Hansen M, et al. 2019. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet 51: 63–75.

Ding Z-L, Oskarsson M, Ardalan A, Angleby H, Dahlgren L-G, Tepeli C, Kirkness E,

Savolainen P, Zhang Y-P. 2012. Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. Heredity 108: 507–514.

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 11: 446–450.

Evans TM, Bira L, Gastelum JB, Weiss LT, Vanderford NL. 2018. Evidence for a mental health crisis in graduate education. Nat Biotechnol 36: 282–284.

Fagan K, Crider A, Ahmed AO, Pillai A. 2017. Complement C3 Expression Is Decreased in Autism Spectrum Disorder Subjects and Contributes to Behavioral Deficits in Rodents. Mol Neuropsychiatry 3: 19–27.

Faraway JJ. 2002. Practical Regression and Anova using R. http://csyue.nccu.edu.tw/Practical%20Regression%20and%20Anova%20using%20R.pdf (Accessed April 17, 2023).

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, et al. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47: 1228–1235.

Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh P-R, Lareau C, Shoresh N, et al. 2018. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet 50: 621–629.

Fong SL, Capra JA. 2021. Modeling the evolutionary architectures of transcribed human enhancer sequences reveals distinct origins, functions, and associations with human-trait variation. Mol Biol Evol. http://dx.doi.org/10.1093/molbev/msab138.

Foyer P, Bjällerhag N, Wilsson E, Jensen P. 2014. Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test. Appl Anim Behav Sci 155: 93–100.

Freedman DG, King JA, Elliot O. 1961. Critical period in the social development of dogs. Science 133: 1016–1017.

Gong Y, Greenbaum J, Deng H-W. 2019. A statistical approach to fine-mapping for the identification of potential causal variants related to human intelligence. J Hum Genet 64: 781–787.

Gopalakrishnan S, Sinding M-HS, Ramos-Madrigal J, Niemann J, Samaniego Castruita JA, Vieira FG, Carøe C, Montero M de M, Kuderna L, Serres A, et al. 2018. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. Curr Biol 28: 3441–3449.e5.

Griesemer D, Xue JR, Reilly SK, Ulirsch JC, Kukreja K, Davis JR, Kanai M, Yang DK, Butts JC, Guney MH, et al. 2021. Genome-wide functional screen of 3’UTR variants

uncovers causal variants for human disease and evolution. Cell 184: 5247–5260.e19.

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. 2017. Genetic effects on gene expression across human tissues. Nature 550: 204–213.

Hansen Wheat C, Larsson L, Berner P, Temrin H. 2022. Human-directed attachment behavior in wolves suggests standing ancestral variation for human-dog attachment bonds. Ecol Evol 12: e9299.

Herbeck YE, Eliava M, Grinevich V, MacLean EL. 2022. Fear, love, and the origins of canid domestication: An oxytocin hypothesis. Comprehensive Psychoneuroendocrinology 9: 100100.

International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). 2018. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. Mol Psychiatry 23: 1181–1188.

Kaeberlein M, Creevy KE, Promislow DEL. 2016. The dog aging project: translational geroscience in companion animals. Mamm Genome 27: 279–288.

Karlsson EK, Lindblad-Toh K. 2008. Leader of the pack: gene mapping in dogs and other model organisms. Nat Rev Genet 9: 713–725.

Karlsson J, Eriksson M, Liberg O. 2007. At what distance do wolves move away from an approaching human? Can J Zool 85: 1193–1197.

Kassahun Y, Mattiangeli V, Ameni G, Hailu E, Aseffa A, Young DB, Hewinson RG, Vordermeier HM, Bradley DG. 2015. Admixture mapping of tuberculosis and pigmentation-related traits in an African-European hybrid cattle population. Front Genet 6: 210.

Kazdoba TM, Leach PT, Yang M, Silverman JL, Solomon M, Crawley JN. 2016. Translational Mouse Models of Autism: Advancing Toward Pharmacological Therapeutics. Curr Top Behav Neurosci 28: 1–52.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423: 241–254.

Kircher M, Xiong C, Martin B, Schubach M, Inoue F, Bell RJA, Costello JF, Shendure J, Ahituv N. 2019. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. Nat Commun 10: 3583.

Klin A, Micheletti M, Klaiman C, Shultz S, Constantino JN, Jones W. 2020. Affording

autism an early brain development re-definition. Dev Psychopathol 32: 1175–1189.

Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, et al. 2021. The Human Phenotype Ontology in 2021. Nucleic Acids Res 49: D1207–D1217.

Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, Stahl PW, Piper PJ, Lingaas F, Fredholm M, et al. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proc Natl Acad Sci U S A 109: 8878–8883.

LeBlanc JJ, Fagiolini M. 2011. Autism: A “Critical Period” Disorder? Neural Plast 2011. https://www.hindawi.com/journals/np/2011/921680/ (Accessed August 26, 2020).

Lenkei R, Újváry D, Bakos V, Faragó T. 2020. Adult, intensively socialized wolves show features of attachment behaviour to their handler. Sci Rep 10: 17296.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478: 476–482.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803–819.

Li N, Stephens M. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165: 2213–2233.

Li Y, Wang G-D, Wang M-S, Irwin DM, Wu D-D, Zhang Y-P. 2014. Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. Genome Biol Evol 6: 3115–3121.

Lord K. 2013. A Comparison of the Sensory Development of Wolves (Canis lupus lupus) and Dogs (Canis lupus familiaris). Ethology 119: 110–120.

Lord KA, Larson G, Coppinger RP, Karlsson EK. 2020. The History of Farm Foxes Undermines the Animal Domestication Syndrome. Trends Ecol Evol 35: 125–136.

Luu P-L, Ong P-T, Dinh T-P, Clark SJ. 2020. Benchmark study comparing liftover tools for genome conversion of epigenome sequencing data. NAR Genom Bioinform 2: lqaa054.

MacNulty DR, Mech LD, Smith DW. 2007. A Proposed Ethogram of Large-Carnivore Predatory Behavior, Exemplified by the Wolf. J Mammal 88: 595–605.

McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, et al. 2008. Runs of homozygosity in European populations. Am J Hum Genet 83: 359–372.

Megquier K, Turner-Maier J, Morrill K, Li X, Johnson J, Karlsson EK, London CA, Gardner HL. 2022. The genomic landscape of canine osteosarcoma cell lines

reveals conserved structural complexity and pathway alterations. PLoS One 17: e0274383.

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res 49: D394–D403.

Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, et al. 2014. Transcriptional landscape of the prenatal human brain. Nature 508: 199–206.

Moravčíková N, Kasarda R, Židek R, Vostrý L, Vostrá-Vydrová H, Vašek J, Čílová D. 2021. Czechoslovakian Wolfdog Genomic Divergence from Its Ancestors Canis lupus, German Shepherd Dog, and Different Sheepdogs of European Origin. Genes 12. http://dx.doi.org/10.3390/genes12060832.

Moretti L, Hentrup M, Kotrschal K, Range F. 2015. The influence of relationships on neophobia and exploration in wolves and dogs. Anim Behav 107: 159–173.

Morey D. 2010. Dogs: Domestication and the Development of a Social Bond. Cambridge University Press.

Morrill K, Hekman J, Li X, McClure J, Logan B, Goodman L, Gao M, Dong Y, Alonso M, Carmichael E, et al. 2022. Ancestry-inclusive dog genomics challenges popular breed stereotypes. Science 376: eabk0639.

Mulatinho MV, de Carvalho Serao CL, Scalco F, Hardekopf D, Pekova S, Mrasek K, Liehr T, Weise A, Rao N, Llerena JC Jr. 2012. Severe intellectual disability, omphalocele, hypospadia and high blood pressure associated to a deletion at 2q22.1q22.3: case report. Mol Cytogenet 5: 30.

Muller SM. 2021. Companion Cyborgs: Untethering Discourses About Wolf-Hybrids. Environmental Communication 15: 99–114.

Nagasawa M, Kikusui T, Onaka T, Ohta M. 2009. Dog’s gaze at its owner increases owner's urinary oxytocin during social interaction. Horm Behav 55: 434–441.

Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S. 2018. Item-level analyses reveal genetic heterogeneity in neuroticism. Nat Commun 9: 905.

Noh HJ, Tang R, Flannick J, Dushlaine CO, Swofford R, Howrigan D, Genereux DP, Johnson J, van Grootheest G, Grunblatt E, et al. 2017. Integrating evolutionary and regulatory information with multispecies approach implicates genes and pathways in obsessive-compulsive disorder. Nat Commun 107193.

Ostrander EA, Wayne RK. 2005. The canine genome. Genome Res 15: 1706–1716.

Pageat P. 2005. Assessing prolactinemia in anxious dogs (Canis familiaris): interest in diagnostic value and use in the selection of the most appropriate psychotropic drug. Current Issues and Research in Veterinary Behavioral Medicine 155.

Pankratov V, Yunusbaeva M, Ryakhovsky S, Zarodniuk M, Estonian Biobank Research Team, Yunusbayev B. 2022. Prioritizing autoimmunity risk variants for functional analyses by fine-mapping mutations under natural selection. Nat Commun 13: 7069.

Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, et al. 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50: 381–389.

Pasaniuc B, Zaitlen N, Lettre G, Chen GK, Tandon A, Kao WHL, Ruczinski I, Fornage M, Siscovick DS, Zhu X, et al. 2011. Enhanced statistical tests for GWAS in admixed populations: assessment using African Americans from CARe and a Breast Cancer Consortium. PLoS Genet 7: e1001371.

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, et al. 2004. Methods for high-density admixture mapping of disease genes. Am J Hum Genet 74: 979–1000.

Pendleton AL, Shen F, Taravella AM, Emery S, Veeramah KR, Boyko AR, Kidd JM. 2018. Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. BMC Biol 16: 64.

Perri AR, Mitchell KJ, Mouton A, Álvarez-Carretero S, Hulme-Beaman A, Haile J, Jamieson A, Meachen J, Lin AT, Schubert BW, et al. 2021. Dire wolves were the last of an ancient New World canid lineage. Nature 591: 87–91.

Persson ME, Wright D, Roth LSV, Batakis P, Jensen P. 2016. Genomic Regions Associated With Interspecies Communication in Dogs Contain Genes Related to Human Social Disorders. Sci Rep 6: 33439.

Pilot M, Moura AE, Okhlopkov IM, Mamaev NV, Alagaili AN, Mohammed OB, Yavruyan EG, Manaseryan NH, Hayrapetyan V, Kopaliani N, et al. 2019. Global Phylogeographic and Admixture Patterns in Grey Wolves and Genetic Legacy of An Ancient Siberian Lineage. Sci Rep 9: 17328.

Pino-Yanes M, Gignoux CR, Galanter JM, Levin AM, Campbell CD, Eng C, Huntsman S, Nishimura KK, Gourraud P-A, Mohajeri K, et al. 2015. Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. J Allergy Clin Immunol 135: 1502–1510.

Pörtl D, Jung C. 2017. Is dog domestication due to epigenetic modulation in brain? Dog behavior 3: 21–32.

Privitera AP, Distefano R, Wefer HA, Ferro A, Pulvirenti A, Giugno R. 2015. OCDB: a database collecting genes, miRNAs and drugs for obsessive-compulsive disorder. Database 2015: bav069.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome

association and population-based linkage analyses. Am J Hum Genet 81: 559–575.

Quesnel-Vallières M, Weatheritt RJ, Cordes SP, Blencowe BJ. 2019. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. Nat Rev Genet 20: 51–63.

Randi E, Lucchini V. 2002. Detecting rare introgression of domestic dog genes into wild wolf (Canis lupus) populations by Bayesian admixture analyses of microsatellite variation. Conserv Genet 3: 29–43.

Rezwani M, Pourfathollah AA, Noorbakhsh F. 2022. rbioapi: user-friendly R interface to biologic web services' API. Bioinformatics 38: 2952–2953.

Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J, Maller J, Samocha KE, Sanders SJ, Ripke S, et al. 2016. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nat Genet 48: 552–555.

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

Romero C, Werme J, Jansen PR, Gelernter J, Stein MB, Levey D, Polimanti R, de Leeuw C, Posthuma D, Nagel M, et al. 2022. Exploring the genetic overlap between twelve psychiatric disorders. Nat Genet 54: 1795–1802.

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, et al. 2000. Comparative genomics of the eukaryotes. Science 287: 2204–2215.

Saetre P, Lindberg J, Leonard JA, Olsson K, Pettersson U, Ellegren H, Bergström TF, Vilà C, Jazin E. 2004. From wild wolf to domestic dog: gene expression changes in the brain. Molecular Brain Research 126: 198–206.

Sams AJ, Boyko AR. 2019. Fine-Scale Resolution of Runs of Homozygosity Reveal Patterns of Inbreeding and Substantial Overlap with Recessive Disease Genotypes in Domestic Dogs. G3 9: 117–123.

Savolainen P, Zhang Y-P, Luo J, Lundeberg J, Leitner T. 2002. Genetic evidence for an East Asian origin of domestic dogs. Science 298: 1610–1613.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511: 421–427.

Schweizer RM, Durvasula A, Smith J, Vohr SH, Stahler DR, Galaverni M, Thalmann O, Smith DW, Randi E, Ostrander EA, et al. 2018. Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. Mol Biol Evol 35: 1190–1209.

Scott JP, Marston MV. 1950. Critical periods affecting the development of normal and mal-adjustive social behavior of puppies. J Genet Psychol 77: 25–60.

Shultz S, Klin A, Jones W. 2018. Neonatal Transitions in Social Behavior and Their Implications for Autism. Trends Cogn Sci 22: 452–469.

Sinding M-HS, Gopalakrishnan S, Raundrup K, Dalén L, Threlfall J, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium, et al. 2021. The genome sequence of the grey wolf, Canis lupus Linnaeus 1758. Wellcome Open Res 6: 310.

Sommese A, Valsecchi P, Pelosi A, Prato-Previde E. 2021. Comparing behavioural characteristics of Czechoslovakian Wolfdogs, German shepherds and Labrador retrievers in Italy and the Czech Republic. Appl Anim Behav Sci 237: 105300.

Starkey MP, Scase TJ, Mellersh CS, Murphy S. 2005. Dogs really are man's best friend—Canine genomics has applications in veterinary and human medicine! Brief Funct Genomics 4: 112–128.

Sthle L, Wold S. 1989. Analysis of variance (ANOVA). Chemometrics Intellig Lab Syst 6: 259–272.

Tang R, Noh HJ, Wang D, Sigurdsson S, Swofford R, Perloski M, Duxbury M, Patterson EE, Albright J, Castelhano M, et al. 2014. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. Genome Biol 15: R25.

Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, Andersen KG, Mikkelsen TS, Lander ES, Schaffner SF, et al. 2016. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell 165: 1519–1529.

Thomson VA, Lebrasseur O, Austin JJ, Hunt TL, Burney DA, Denham T, Rawlence NJ, Wood JR, Gongora J, Girdland Flink L, et al. 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. Proc Natl Acad Sci U S A 111: 4826–4831.

Topál J, Román V, Turcsán B. 2019. The dog (Canis familiaris) as a translational model of autism: It is high time we move from promise to reality. WIREs Cogn Sci 10: 509.

Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T. 2021. Genome-wide association studies. Nature Reviews Methods Primers 1: 1–21.

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK. 1997. Multiple and ancient origins of the domestic dog. Science 276: 1687–1689.

Vila C, Wayne R. 1999. Hybridization between wolves and dog. Conserv Biol 13: 195–198.

vonHoldt B, Fan Z, Ortega-Del Vecchyo D, Wayne RK. 2017a. EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. PeerJ 5: e3522.

vonHoldt BM, Shuldiner E, Koch IJ, Kartzinel RY, Hogan A, Brubaker L, Wanser S, Stahler D, Wynne CDL, Ostrander EA, et al. 2017b. Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. Sci Adv 3: e1700398.

Wasik K, Berisa T, Pickrell JK, Li JH, Fraser DJ, King K, Cox C. 2021. Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. BMC Genomics 22: 197.

Weir BS, Cockerham CC. 1984. ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. Evolution 38: 1358–1370.

Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, Schoech AP, van de Geijn B, Reshef Y, Márquez-Luna C, et al. 2020. Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nat Genet 52: 1355–1363.

Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, et al. 2013. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155: 997–1007.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88: 76–82.

Zapata I, Serpell JA, Alvarez CE. 2016. Genetic mapping of canine fear and aggression. BMC Genomics 17: 572.

Zimen E. 1987. Ontogeny of approach and flight behavior towards humans in wolves, poodles and wolf-poodle hybrids. Man and wolf 275–292.

Zoonomia Consortium. 2020. A comparative genomics multitool for scientific discovery and conservation. Nature 587: 240–245.