Title
An Investigation of the Evolutionary History, Adaptation, and Demography of Canids and Galapagos Rails

Author
Chavez, Daniel

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

An Investigation of the Evolutionary History, Adaptation, and Demography of Canids and Galapagos Rails

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biology

by

Daniel Eduardo Chavez Viteri

2021

ABSTRACT OF THE DISSERTATION

An Investigation of the Evolutionary History, Adaptation, and Demography of Canids and Galapagos Rails

by

Daniel Eduardo Chavez Viteri

Doctor of Philosophy in Biology

University of California, Los Angeles, 2021

Professor Robert Wayne, Chair

One of the fundamental goals in evolution is to understand the process of speciation and adaptation. In this dissertation, we explore this long-lasting question by addressing the genetic basis of species adaptation, and the geological and environmental factors that promote both population differentiation and speciation. To answer these questions, we focused on three model systems. **(1)** The genetic basis of the remarkable adaptations of the African wild dog (*Lycaon pictus*). We found unique mutations associated with cursoriality, hypercanivory, and coat color variation in this species. These mutations evolved ~1.7 million years ago, coinciding with the diversification of large-bodied ungulates. **(2)** The diversification of one of the most rapid speciation in carnivores, the South American canids. We found that this group is derived from a single ancestral population that likely colonized South America three million years ago when the

Panama land bridge was formed. In South America, we found that the Andes promote early diversifications in the eastern region, followed by recent diversification in the west of the Andes. We detected extensive historical gene flow among the youngest lineages, which could have augmented species adaption to different niches. Finally, we found a complex history of adaptive diversification throughout a sequence of past climatic cycles in South America, compounded by recent population declines caused by humans. **(3)** The effect of island isolation and species introduction in the genetic differentiation and diversity of Galapagos rails. We found that the separation of a central landmass in the archipelago around 400 thousand years ago shaped the diversification of rail populations. Our findings show that the eradication of goats was critical to avoiding episodes of severe inbreeding in most populations.

The dissertation of Daniel Eduardo Chavez Viteri is approved.

Michael Edward Alfaro

Kirk Edward Lohmueller

Thomas Bates Smith

Klaus-Peter Koepfli

Robert Wayne, Committee Chair

University of California, Los Angeles

2021

To my wife Mabe,

I couldn't have done it without you.

Table of Contents

List of Figures and Tables

Acknowledgements

I am most grateful to my advisors, Bob Wayne. Thank you for being a great mentor, always showing me the means to develop my full potential, even during the pandemic, and challenging me to accomplish things I thought to be impossible. To my committee members Dr. Michael Alfaro, Dr. Kirk Lohmueller, Dr. Tom Smith, Dr. Klaus-peter Koepfli and Dr. Robert Wayne for your valuable guidance and feedback. Chapter one was possible thanks to the collaboration of Dr. Ilan Gronau, Taylor Hains, Dr. Sergei Kliver, Dr. Klaus-peter Koepfli and Dr. Robert Wayne.

Chapter two was possible thanks to Dr. Ilan Gronau, Taylor Hains, Dr. Rebecca Dikow, Dr. Paul Frandsen, Dr. Eduardo Eizirik, Dr. Klaus-peter Koepfli and Dr. Robert Wayne. I thank Kirk Lohmueller, Jacqueline Robinson and Annabel Beichman for guidance in data analysis. I thank Dr. Gaurav Kandlika, Chris Kyriazis, Audra Huffmeyer for improving the manuscripts. I thank Duyen Tran for the illustrations of the bush dog and maned wolf.

Chapter three was possible thanks to Sebastian Espinoza, Dr. Annabel Beichman, Dr. Jaime Chaves and Dr. Robert K. Wayne. I Also thank Dr. Martin Stervander and Dr. Klaus-peter Koepfli for facilitating the reference genome of the Inaccessible island rail. I thank Dr. Gaurav Kandlika, Dr. Klaus-peter Koepfli, Laila Hualpa and Elana Bachrach for help me to improve the manuscripts. Thanks to the Wayne Lab, for being an inspiration and support through all these years.

Thanks to my family. You have shown me that the only way to success is hard work. To my wife Mabe and child Carlos. You have been the joy and happiness of my life and the fuel that kept me going through years of anxiety and stress.

Curriculum Vitae

Daniel Eduardo Chavez Viteri

EDUCATION

B.S. Catholic Pontifical University of Ecuador. September 10, 2012.
Advisors: M.Sc. Santiago Burneo / Commite: Ph.D. Omar Torres and Ph.D. Santiago Ron
Thesis title: Genetic, morphologic and ecological variation of the genus *Reithrodontomys* (Rodentia: Cricetidae) in Ecuador.

HONORS AND AWARDS

René Fonseca Award, granted by the Ecuadorian Society of Mammalogy, to the 2nd best project of the year 2013. Project: Genetic, morphologic and ecological variation of the genus *Reithrodontomys* (Rodentia: Cricetidae) in Ecuador.

RESEARCH

Publications:

Chavez, D. E. et al. Comparative genomics provides new insights into the remarkable adaptations of the African wild dog (*Lycaon pictus*). Scientific Reports 9, doi:10.1038/s41598-019-44772-5 (2019).

Brito, J., Tinoco, N., **Chavez**, D., Moreno-Cárdenas,P., Batallas,D., and Reed Ojala-Barbour. 2017. New species of arboreal rat of the genus *Rhipidomys* (Cricetidae, Sigmodontinae) from Sangay National Park, Ecuador. Neotropical Biodiversity 3(1).

Camacho, M. A., D. **Chavez** and S. F. Burneo. 2016. A taxonomic revision of the Yasuni Round-eared bat, *Lophostoma yasuni* (Chiroptera: Phyllostomidae). Zootaxa 4114(3): 246–260.

Chavez, D. 2012. Genetic, morphologic and ecological variation of the genus *Reithrodontomys* (Rodentia: Cricetidae) in Ecuador. Quito, Ecuador, Catholic Pontifical University of Ecuador. 127 p

Lee,T., Burneo,S., Cochran,T., **Chavez**, D. 2010. Small Mammals of Santa Rosa, Southwestern Imbabura Province, Ecuador. Outhwestern Imbabura Province, Ecuador. Occasional Papers 290: 1-14.

Publications in preparation:

Chavez, D. E. et al. Splendid diversification: dissecting the evolution of South American canids with whole genome sequences (2021)

Chavez, D. E. et al. Population genomics of the Galapagos rails (*Laterallus spilonota*) (2021)

Presentations:

II CONGRESO LATINOAMERICANO DE MASTOZOOLOGÍA. Chávez, D., Boada, C., Burneo, S. Genetic, morphologic and ecological variation of the genus *Reithrodontomys* (Rodentia: Cricetidae) in Ecuador. Buenos Aires, Argentina. Date: Noviembre 9, 2012 (Presenting author).
II CONGRESO ECUATORIANO DE MASTOZOOLOGÍA. Chávez, D., Boada, C. Genetic variation of the montane fish-eating rat *Neusticomys monticolus* Anthony, 1921 (Rodentia Cricetidae) in Ecuador. Date: Noviembre 9, 2012 (Presenting author).

II CONGRESO ECUATORIANO DE MASTOZOOLOGÍA. Chávez, D., Boada, C., Burneo, S. Genetic, morphologic and ecological variation of the genus *Reithrodontomys* (Rodentia: Cricetidae) in Ecuador. Buenos Aires, Argentina. in Ecuador. Date: Noviembre 9, 2012 (Presenting author).
Primer Simposio de Murciélagos en el Ecuador. Tirira,D., Boada,C., Buitrón-Jurado, G., Chávez, D. Structure and diversity of bat assemblages in the San Francisco River, Esmeraldas Ecuador. Date: May 6, 2009 (Presenting co-author).

Relevant professional & synergistic activities:

Research assistant, Museum of Zoology, Mammalogy. Pontificia Universidad Católica del Ecuador. (Febrary2005 –September,2012).
Researcher for the project: "Genetic characterization of mammals in Yasuní National Park". Pontificia Universidad Católica del Ecuador. (March 2013 -Present).
Researcher: Presence of new species of small-mammals along the Andes of Ecuador. (Oct 2011 -Present).
Laboratory technician, Molecular Biology Laboratory, Pontificia Universidad Católica del Ecuador. (March 2013 -2014).
PhD Candidate. University of California, Los Angeles.

Additional research affiliations:
Research Associate, Museo de Zoología de Vertebrados, Pontificia
Universidad Católica del Ecuador, Quito-Ecuador (September 2012– 2016).
Research Collaborator, Laboratory of Molecular Biology, Pontificia
Universidad Católica del Ecuador, Quito-Ecuador (September 2012– 2016).

FUNDING
Project Research Grant, Pontificia Universidad Católica del Ecuador, 2010. Project: Genetic, morphologic and ecological variation of the genus *Reithrodontomys* and *Thomasomys* (Rodentia: Cricetidae) in Ecuador.
Fellowship Undergraduate Student Training, Pontificia Universidad Católica del Ecuador. 2007-2010.
Conservation Genetics course (Cogen), Pró-Carnivoros, Brazil 2011. The only student granted with a full scholarship.
PhD Scholarship Program in Universities of Excellency. (2014-2017). SENECYT (The Secretariat of Higher Education, Science and Technology of Ecuador).
Summer 2019 fellowship, Ecology and Evolutionary Biology, UCLA.
Winter 2020 fellowship, Ecology and Evolutionary Biology, UCLA.
Lida Scott Brown Fellowship (2020). Ecology and Evolutionary Biology, UCLA.

AREAS OF TECHNICAL EXPERSTISE
Analysis of whole genomes including mapping to a reference genome, filtering, variant calling, population genomics (e.g inference, simulations, ROH, heterozygosity), phylogenomics and positive selection. Also, familiar with *de novo* assembly methods.

Bioinformatics: Proficient in Python, bash (including Computing Cluster), R and git. Familiar with C++.
Molecular techniques for systematic, phylogeographic and population genetics studies (DNA isolation, PCR, sequencing, and cloning).
Extensive field work experience in Ecuador (Specimens identification, sample and data collection, expedition planning and research permits).

TEACHING EXPERIENCE
Teaching assistant. 3 quarters. LS7B (Genetics, Evolution, and Ecology).2018-2019.
Teaching assistant 1 quarter. EEB C177/C234 (Practical Computing for Evolutionary Biologists and Ecologists). 2019
Teaching assistant 1 quarter. 2019. EEB87 (California's DNA: A Field Course).
Teaching assistant 2 quarters. 2020-2021. EEB108 (Biodiversity in the Age of Humans).
Software Carpentry workshop, EEB, UCLA. 2020 (Helper).
Teaching assistant 1 quarter. 2021. EEB116 (Conservation Biology)

Chapter I, Comparative genomics provides new insights into the remarkable adaptations of the African wild dog (*Lycaon pictus*)

Abstract

Within the Canidae, the African wild dog (*Lycaon pictus*) is the most specialized with regards to cursorial adaptations (specialized for running), having only four digits on their forefeet. In addition, this species is one of the few canids considered to be an obligate meat-eater, possessing a robust dentition for taking down large prey, and displays one of the most variable coat colorations amongst mammals. Here, we used comparative genomic analysis to investigate the evolutionary history and genetic basis for adaptations associated with cursoriality, hypercanivory, and coat color variation in African wild dogs. Genome-wide scans revealed unique amino acid deletions that suggest a mode of evolutionary digit loss through expanded apoptosis in the developing first digit. African wild dog-specific signals of positive selection also uncovered a putative mechanism of molar cusp modification through changes in genes associated with the sonic hedgehog (SHH) signaling pathway, required for spatial patterning of teeth, and three genes associated with pigmentation. Divergence time analyses suggest the suite of genomic changes we identified evolved ~1.7 Mya, coinciding with the diversification of large-bodied ungulates. Our results show that comparative genomics is a powerful tool for identifying the genetic basis of evolutionary changes in Canidae.

Introduction

Among the living species of Canidae, the African wild dog (hereafter, AWD) is considered to be the most specialized with regard to adaptations for cursoriality, diet, and coat coloration (Hartstone-Rose et al., 2010). Along with a gracile appendicular skeleton, the most notable characteristic of AWDs is the loss of the first digit on the forefeet. This trait increases their stride length and speed allowing them to pursue prey for long distances in open plain habitats and is unique among living canids (Creel and Creel, 2002). The dentition of the AWD is also exceptional, as the teeth are generally sectorial in shape and the premolars are the largest relative to body size of any living carnivoran except spotted hyenas (Van Valkenburgh, 1991). AWDs also show a transformation of the talonid on the lower first molar (carnassial) from a basin-like crushing depression into a trenchant heel or cutting blade for slicing flesh, which also occurs independently in two other hypercarnivorous canids, the bush dog (*Speothos venaticus*) and the dhole (*Cuon alpinus*). This feature is accompanied by the reduction or loss of post-carnassial molars, also a characteristic of hypercarnivorous canids (Van Valkenburgh and Koepfli, 1993). Fossil evidence suggests that the reduction of the first digit and transformation of the carnassial in AWDs evolved gradually during the Plio-Pleistocene (Martinez-Navarro and Rook, 2003). Finally, AWDs exhibit one of the most variegated coats among mammals, with individuals uniquely differing in pigmentation pattern and color (Shumba et al., 2017) (Figure 1-1a), which suggests the expression of a diversity of genes. The function of this highly individualistic coat pattern is uncertain but may represent an adaptation for concealment, communication or thermoregulation (Walsberg, 1983, Gerald, 2001, Steiner et al., 2007, Candille et al., 2007).

We investigated the genetic origins of adaptations associated with the evolution of cursoriality, hypercarnivory, and coat color variation in the AWD in the context of their evolutionary and demographic history. To accomplish these goals, we sequenced one high coverage AWD genome and utilized three previously reported AWD genomes (Campana et al., 2016, Gopalakrishnan et al., 2018). Additionally, we used three *de novo* AWD reference genomes (Armstrong et al., 2018). Coverage depths are provided in the Supplementary Table 1-S1. These genomes were compared with existing genomes from *Canis* (wolves, coyote and golden jackal) and *Cuon alpinus* (dhole) (Auton et al., 2013, Koepfli et al., 2015, Freedman et al., 2014, Campana et al., 2016, Robinson et al., 2016, vonHoldt et al., 2016, Gopalakrishnan et al., 2018). The genome of the bush dog was not included in this study because it is part of an ongoing research investigation on comparative genomics of South American canids (Armstrong et al., 2018). We hypothesized that genes showing signals of positive selection and other molecular changes in AWDs are associated with digit reduction, tooth morphology, and pigmentation. Furthermore, we aimed to investigate the possibility of convergent evolution at the genetic level, exploring shared signals of selection among the wolf-like canids that have a trenchant heel (AWDs and dholes).

Results and Discussion

Evolutionary history

To provide an accurate evolutionary framework for the comparative genomic analyses of AWDs relative to other wolf-like canids, we first reconstructed the phylogenetic relationships among species of *Canis*, *Cuon*, and *Lycaon*. Species tree analysis using ASTRAL-III (Zhang et al., 2018) produced 102 distinct gene trees from 8,117 25kb alignments sampled from 38

autosomes (~203Mb), with the final species tree being the consensus topology of the 100 replicates from the analysis (Figure 1-1a). The resulting topology shows the Ethiopian wolf basal relative to the rest of *Canis*, with dhole and AWDs as successive sister lineages having diverged earlier (Figure 1-1a and 1-S1). Within *Canis*, the golden jackal and African wolf are independent lineages sister to the clade comprised of the coyote, gray wolf and domestic dog. These patterns of relationship are consistent with previous analyses based on nuclear DNA sequences (Lindblad-Toh et al., 2005, Koepfli et al., 2015, Gopalakrishnan et al., 2018).

We next estimated the age of divergence of the AWD lineage using the inferred tree and two methods. First, we computed average genomic divergence times using MCMCTree (Yang, 2007), in which two fossil priors were used to calibrate nodes of the phylogeny (see Methods for details). Our estimates suggest an average genomic divergence of 3.91 mya (95% HPD = 3.30-4.50 mya) between *Lycaon* and the clade containing *Cuon* and *Canis*, approximately two million years earlier than the earliest fossil evidence recorded for the lineage (Hartstone-Rose et al., 2010). This more ancient divergence is feasible because average genomic divergence captures not only the time since species divergence, but also the time for lineages to coalesce in ancestral populations, which are known to be very large in canids (Freedman et al., 2014). To better address this discrepancy, we thus jointly inferred a complete demographic model using G-PhoCS (Gronau et al., 2011), taking into consideration gene flow between 44 possible pairs of branches in the phylogeny (see Methods for details). Indeed, the species divergence time for *Lycaon* is then estimated at 1.72 mya (95% HPD = 1.70-1.74 mya; Table 1-S2 and Figure 1-1a), which is much closer to estimates from both the fossil record and recent analyses of whole-genome data (Hartstone-Rose et al., 2010, vonHoldt et al., 2016, Gopalakrishnan et al., 2018). Importantly, while our inferred model suggests prevalent gene flow between divergent canid species, *Lycaon*

is inferred to be largely isolated from genetic exchange with other canid lineages. This isolation provided more time for unique genomic adaptations to evolve in isolation.

African wild dogs are uniquely enriched in positively-selected genes related to primary cilia

To identify positive selection events that occurred on protein-coding genes during the evolution of the AWD lineage, the sequencing reads for four AWDs and eight other canid species were mapped to the domestic dog reference assembly (CanFam3.1) to take advantage of the high-quality annotation of the dog reference genome (Table 1-S1). The mapping process was based on the GATK Best Practices pipeline (Methods). For almost all canids, we found that more than 97% of reads successfully mapped to the dog genome. The only exception was a low coverage (12.1x) AWD that had ~93% of the reads mapped to the dog. To avoid potential reference bias from aligning reads to a different species, we further confirmed our results on three recently published *de novo* AWD reference genomes (Armstrong et al., 2018).
After calling genotypes with SAMtools and filtering with GATK 3.7 (McKenna et al., 2010) as well custom python scripts, we identified ~19,000 orthologous protein-coding genes. Among these genes, 18,327 passed our quality filters (no internal stop codon, permissible length, and longest transcript) and were used to identify genes under positive selection using the branch-site model (Yang, 2007). This test was conducted on each multi-species gene alignment generated with PRANK v.150803 (Loytynoja and Goldman, 2005) and using the topology in Figure 1-1a as the guide tree. AWD, dhole, and gray wolf were specified as different foreground branches. A gene was considered positively selected if the value obtained from the likelihood-ratio test comparing a model where the ratio of nonsynonymous substitutions (dN) to synonymous substitutions (dS) was greater than 1 (dN/dS > 1) against a null model where dN/dS = 1.

Significant differences were determinate with a chi-square distribution with 1 degree of freedom (Zhang et al., 2005).

One issue with the branch-site model is that it is highly sensitive to alignment errors. Therefore, we conducted an extensive filtering process on our data, first using SWAMP(Harrison et al., 2014) and then visually inspecting the alignments of genes with $p < 0.05$. After masking regions with unusual enrichment of amino acid changes, we conducted three independent runs for each foreground branch and gene family, and retained the one with the best likelihood-ratio score of each run (Yang and dos Reis, 2011). This guaranteed that large log-likelihood ratios depicted from the branch-site model were not the result of convergence problems of the test (Yang and dos Reis, 2011). Another concern in the exploration for genes under positive selection is the role of multiple nucleotide changes (Venkat et al., 2018). Although these changes may occur simultaneously, the branch-site model assumes that they occur in a successive manner. The result will be unrealistically high likelihood-ratio scores at a codon were nucleotide changes occurred at the same time (Venkat et al., 2018). Among our 12 candidate genes (Figure 1-1a), we identified three genes with multiple nucleotide changes (*CC2d2a*, *TMEM67*, *PAH*) and thus support for the positive selection on these genes should be interpreted with caution. Although it is challenging to elucidate the order of multiple nucleotide changes, our main conclusions are not affected even if we take a conservative approach and do not include such genes in the analysis.

We found 43 genes (Table 1-S3) that were significant at a false discovery rate (FDR) of 20%, after conducting multiple hypothesis testing of 18,327 genes along the three foreground branches (AWD, dhole, and gray wolf). Since only a few genes passed the genome-wide significance threshold, we used all genes with a p-value ≤ 0.01 to test for enrichment of gene functions with G-profiler (Reimand et al., 2007). Ensembl identification of genes with a p-value

≤ 0.01 were input as query lists and the 18,327 total gene set was used as the background list. We allowed a minimum of two genes to overlap between query genes and genes belonging to a gene ontology (GO) term. This resulted in genes with specific signals of selection in AWDs overrepresented in terms related to primary cilia (Figure 1-1a), which are significantly involved in coordinating signaling pathways during mammalian development (Venkatesh, 2017).

The disadvantage of common tests for gene ontology enrichment like G-profiler (Reimand et al., 2007) is that an arbitrary significance cutoff must be specified, and data below that cutoff is expected to be lost. Therefore, instead of just focusing on some outlier genes with high likelihood ratios, we used polysel (Daub et al., 2017) to conduct analyses of polygenic selection across the full set of tested genes. We looked for pathways that were overrepresented with genes having low or moderate likelihood ratios more than would be expected by chance. This model takes likelihood ratio test statistics estimated from the branch-site test and finds weak to moderate polygenic selection within biological pathways. Then, p-values are generated from an empirical null distribution obtained by randomly sampling gene sets in specified pathways. Using this approach, we also found that the "primary cilia" GO category was significantly enriched with a variety of levels of positive selection (Figure 1-1b). To rule out the possibility that this GO category could be enriched just by chance due to the large number of genes, we tested for overrepresentation of significant genes in primary cilia in the gray wolf and dhole and found no evidence of enrichment. Finally, to account for possible errors generated from mapping short reads of the AWD to a different species (domestic dog), we verified every mutation reported in this study (e.g., nucleotide and amino acid deletions and substitutions) with a consensus sequence of the three recently published *de novo* AWD reference genomes (NCBI Bioproject PRJNA488046; Table 1-S1) (Armstrong et al., 2018).

Digit Reduction through apoptosis

Two developmental mechanisms of digit reduction from the ancestral five-digit morphology have been characterized in mammals. One is related to a complete absence of a digit during development through regulation of the transduction of sonic hedgehog (SHH) signaling and the other involves apoptosis of digits during early development (Cooper et al., 2014). The loss of the first digit, as found in AWDs, has been shown to be independent of SHH signaling (Scherz et al., 2007). Therefore, we focused our analyses on genes associated with apoptosis pathways, particularly those related to digit development.

We used the Variant Effect Predictor annotation tool (McLaren et al., 2016) to identify amino acid-changing substitutions unique to the AWD that could have a significant impact on the associated proteins but will be ignored by the branch-site model test. We identified 403 genes with both high and moderate impact. High impact indicates a disruptive substitution that could cause truncation, loss of function, or nonsense-mediated decay of a protein whereas moderate impact indicating a non-disruptive substitution that might change protein functional efficiency. The substitutions we identified were categorized as in-frame indels, frameshift variations, and stop codon gains (see Methods for details).

Strikingly, we found 596 genes with in-frame-deletions, with moderate impact, unique to AWDs. These amino acid deletions were tested for enrichment of GO categories using G-profiler (Reimand et al., 2007). We found that this type of mutation was overrepresented in digit-loss categories with a false discovery rate of 5%. Specifically, the term, “abnormality of the thumb,” was over-represented by the genes *FANCC*, *FANCD2*, and *FANCM*, which are associated with the Fanconi anemia (FA) pathway (Auerbach, 2009). We also observed an overrepresentation of 33 genes with frameshift variation mutations in the olfactory receptor (OR) GO category.

Twenty eight of these genes were also enriched in the olfactory transduction KEGG category, in accordance with the dynamic evolution of OR gene families (Nei et al., 2008).

Our results implicate amino acid deletions in genes associated with the FA pathway in the loss of the first digit in AWDs through an apoptosis pathway that typically directs interdigital cell death (Figure 1-2a and 1-2b). Specifically, in the development of the ancestral five-digit foot, the primary function of apoptosis is to eliminate excessive cells on the interdigital webs which trims the dimension of the digit (Grotewold and Ruther, 2002, Cooper et al., 2014). When this digit individualization occurs during development, apoptosis has only a small effect on digit dimension (Zakeri et al., 1994). Studies have shown that FA proteins form a complex with the CtBP1 protein that results in the repression of the *DKK1* gene. As expression of this gene is restricted to the interdigital area during the early development of digits, its repression prevents apoptosis from extending into the digits (Grotewold and Ruther, 2002, Huard et al., 2014). When the FA-CtBP1 complex fails to properly form, inhibition of *DKK1* expression is removed, thereby permitting apoptosis to extend to the digits. We suggest that the amino acid deletions in *FANCC*, *FANCD2*, and *FANCM* found exclusively in AWDs may reduce the binding affinity of the FA protein complex to CtBP1, thus allowing the loss of the first digit through apoptosis (Figure 1-2b). Our findings are strongly supported by the fact that the mutations we have identified in the FA genes in AWDs are responsible for a condition commonly associated with the absence of the first digit in humans (Webb et al., 2011, Auerbach, 2009) and expression of *DKK1* is related with absence of the first digit in mice (Adamska et al., 2004). Moreover, apoptosis as a mode of evolutionary digit has also has been shown in horses, jerboas, and camels (Cooper et al., 2014).

Hypercarnivory through sonic hedgehog (SHH) signaling

The primary cilium is a hair-like structure that projects from the surface of cells and serves as a sensory organelle, transmitting signals from the extracellular space into the nucleus (Bangs and Anderson, 2017) (Figure 1-3a). This structure is located at the dental epithelium and mesenchyme during the formation of dental cusps (Jernvall and Thesleff, 2000, Hampl et al., 2017). As the primary cilium regulates numerous signaling pathways necessary for odontogenesis, ciliary defects can alter the process of cusp patterning (Thivichon-Prince et al., 2009).

We found AWD-specific amino acid substitutions in components of the primary cilium, specifically the genes *WDR35*, *TMEM67*, *CC2d2a* and *GLI1*. These substitutions suggest a combined regulatory effect, through GLI transcription, on SHH-dependent genes (Figure 1-3a). Specifically, the role of primary cilia depends on the retrograde (to the basal body region) and antegrade (to the tip of the ciliary membrane) intraflagellar transport (IFT) of the transcriptional factor GLI1 (Liem et al., 2012, Hampl et al., 2017) (Figure 1-3a). This transportation occurs through a microtubule structure called the axoneme. The efficacy of GLI mobilization through the axoneme is dependent upon protein-mediated transport by WDR35 (also known as IFT121) as well as the proper docking of the axoneme into the basal body region, which is mediated by TMEM67 and CC2d2a proteins (Veleri et al., 2014, Abdelhamed et al., 2015, Hampl et al., 2017) (Figure 1-3a). Mutations within genes encoding primary cilium components alter mobilization of GLI to the basal body, and hence result in gain or loss of GLI function in this region. An increase in GLI will result in the gain of SHH phenotypes such as growth of molar cusps. In contrast, a decrease in GLI will cause loss of SHH phenotypes (Goetz and Anderson, 2010) such as inhibition of molar cusp development (Jernvall and Thesleff, 2000). We suggest

that the amino acid changes observed in AWDs in *WDR35*, *TMEM67*, *CC2d2a*, and *GLI1* may cause rapid transportation of GLI to the basal body, and consequently overexpression of SHH target genes (Ohazama et al., 2009, Hampl et al., 2017). Variants in the candidate genes reported in this study have been associated with abnormal tooth shape and may thus be related to the exaggerated hypercarnivorous dentition of AWDs, which includes the development of a lower trenchant carnassial (Ohazama et al., 2009, Hoffer et al., 2013, Hampl et al., 2017). Most of the amino acid sites from these genes were unique to AWD even when compared to other placental mammals and marsupials, including two South American canids, *Speothos venaticus* and *Chrysocyon brachyurus* (Chavez et al., unpublished data, Figure 1-S2). Also, specific sites in *CC2d2a* had a relatively high probability of being affected by positive selection as suggested by a Bayes empirical Bayes test (BEB > 0.90). Although nucleotide changes were not recorded for *GLI1* with BEB, mutations in this gene were located within significant windows as suggested by the $\frac{\theta}{D}$ estimate ($p < 0.01$; see Methods section) as estimated from the HKA-like test (Hudson et al., 1987), after verifying that the amount of information within windows of 25kb in length was not driving higher differences between diversity and divergence (Figure 1-S3). Our results suggest that the transduction of SHH through primary cilium may have promoted the modification of a primitive molar with a posterior crushing basin into a trenchant sectorial single cusp in AWDs (Figure 1-3b).

Another gene associated with spatial patterning of the tooth, *CREBBP*, was found to have an amino acid deletion in AWDs. Interestingly, we also observed a two amino acid deletion in the same region in the hypercarnivorous dhole, a canid that also possesses a lower trenchant carnassial heel (Figure 1-3b). *CREBBP* is a strong candidate for the modified carnassial observed in these two hypercarnivorous canids. Notably, this gene is associated with abnormal numbers or

features (talon cusps) of molar cusps in humans (Bloch-Zupan et al., 2007) (Figure 1-3b). Even though *CREBBP* is ubiquitously expressed, the shared amino acid deletions were in the glutamine-rich region of the protein (Figure 1-3c). This region is highly conserved in eukaryotes and serves as the binding site for Smad proteins(Janknecht et al., 1998, Kraus et al., 1999). These proteins are transcriptional factors that regulate the expression of genes located at the dental lamina and mesenchyme, and play important roles in regulating differentiation and proliferation of cells during tooth development (Huang and Chai). Specifically, Smad transcription factors enter the nucleus and bind to the coactivator *CREBBP* and regulate the expression of target genes (Figure 1-3a). We suggest that the observed amino acid deletions observed in *CREBBP* in AWDs and dholes may alter the formation of the Smad-*CREBBP* complex. This ultimately will have a regulatory effect on the expression of TGF-β and BMP dependent genes(Huang and Chai). In Smad knockout mice, dental cusp formation is affected (Huang and Chai). Our results suggest that the formation of the trenchant heel of AWD, initially guided by primary cilium components, may be reinforced by a regulatory effect of Smad transcriptional factors on genes involved in tooth development(Huang and Chai). Also, our findings suggest that this may be the regulatory pathway that also determines the blade-like cusps in the dhole.

The shared amino acid deletions in the *CREBBP* gene observed in AWDs and dholes (Figure 1-3c) could have arisen through different routes. First, the changes could have evolved independently in each species. Alternatively, the changes could have resulted from past adaptive introgression between species, as has been found among species within the *Panthera* lineage (Figueiro et al., 2017). To account for the latter possibility, we conducted tests of admixture in the context of the evolutionary and demographic history of the sampled canids(Gronau et al., 2011). Results from models with and without gene flow among different lineages suggested a

history of extensive admixture among species of *Canis* (Supplementary Discussion; Table 1-S2 and Figure 1-1a), consistent with recent findings (Gopalakrishnan et al., 2018). However, the genomic data for AWDs and dholes suggest little or no gene flow between these lineages and those leading to *Canis* species. Although the lack of post-speciation gene flow between the dhole and AWD could suggest that the amino acid deleting changes in *CREBBP* evolved independently, we do not entirely rule out the possibility of shared post-divergence ancestry. Particularly, we found a low proportion of divergent sites between the AWD and dhole (only 9 divergent sites out of 3,000 flanking sites analyzed) around *CREBBP* that suggest a plausible shared ancestry. Regardless of the mode of evolution of *CREBBP*, the amino acid deletions shared between the AWD and the dhole may reflect similar selective forces favoring hypercarnivory (Van Valkenburgh and Koepfli, 1993).

Positively-selected genes associated with AWD pelage coloration.

Assuming that positive selection could have occurred in genes associated with the unique pelage coloration and patterning seen in AWDs, we tested a set of 151 genes that have been shown to be involved in mammalian pigmentation (Sturm et al., 2001, Raposo and Marks, 2007, Lamoreux et al., 2010, Kaelin and Barsh, 2013, Crawford et al., 2017) using the branch-site test (Yang, 2007). We found six genes with AWD-specific signals of positive selection, three of which are known to have relevant function in coat coloration: *MYO5A*, *HPS6*, and *PAH* (Figure 1-4a). The resulting amino acid changes were confirmed in a consensus sequence ("*Dnv Lycaon pictus*" in Figure 1-4b) of three high-coverage *de novo* AWD genomes. Moreover, most amino acid changes in coat color were unique to AWDs when compared to other species of placental mammals, marsupials, and monotremes (Figure 1-4b).

Considering that positively-selected substitutions related to the pigmentary system might fall outside protein-coding regions and could have also occurred during recent evolutionary history, we used the Hudson-Kreitman-Aguadé (HKA) test to examine regions with high divergence and low diversity for signals of selective sweeps. To conduct this test, we first called genotypes using Haplotype Caller in GATK 3.8 (McKenna et al., 2010). We then calculated polymorphism within 100kb windows among the four AWDs genomes that were mapped to the domestic dog reference genome. At the same time, per-site divergence was calculated between the AWD with the highest coverage (LPI_RKW 4881) and the domestic dog. Considering that demographic effects are expected to have an effect across the entire genome, we used empirical *p*-values to identify loci with extreme values of high divergence and low diversity as evidence for positive selection.

After verifying that the amount of information within 100kb windows was not driving higher differences between diversity and divergence (Figure 1-S3), a total of 159 genes were located within windows with a magnitude of differentiation greater than expected by chance (empirical p-value < 0.01). Among the genes identified in the HKA test, seven were also observed in the branch-site test (Figure 1-1a). From this set of genes, we identified *HPS6* as a candidate gene that may have been recently selected. Among the set of genes with AWD-specific signals of positive selection, we did not observe four genes (*ASIP*, *MITF*, *MLPH*, *PMEL*) that were previously shown to have elevated ratios of non-synonymous/synonymous substitutions in two lower coverage AWD genomes (Campana et al., 2016). For example, these authors reported a stop codon-gain in *PMEL*, but we found this was due to a misorientation in the codon translation frame (Figure 1-S4).

The positively-selected genes associated with pigmentation we detected have notable functions that strongly suggest regulation of the variegated pelage of AWDs (Figure 1-4a). Although phenylalanine hydroxylase (*PAH*) has several pleiotropic effects, phenylalanine levels are closely related to melanin deposition (Nagasaki et al., 1999, Ding et al., 2006, Ding et al., 2008) (Figure 1-4d). Levels of phenylalanine are determined by the *PAH* gene, whose protein catalyzes hydroxylation of phenylalanine to tyrosine (Thony et al., 2000, Ding et al., 2008). Using available information to construct a 3D structure of the PAH protein, we located the AWD-specific mutation in the biopterin H domain (Figure 1-4d), which contains a binding domain for the PAH cofactor (Erlandsen et al., 1997). This finding suggests that the mutation observed in the protein domain of *PAH* could be in part responsible for the AWD coat color pattern, as this gene and its cofactor are known to regulate the proportion of yellow and black fur(Ding et al., 2008) through the conversion of phenylalanine to tyrosine (Thony et al., 2000) (Figure 1-4d). The observed amino acid change in *MYO5A* (myosin 5A) is located at the myosin head region (Figure 1-S5), which is relevant to the motor function of the protein and the transportation of melanosomes (Yao et al., 2016). Mutations in this gene result in patchy color patterns in mice (Meehan et al., 2017) (see Figure 1-4c) and dilute pigment in dogs (Kaelin and Barsh, 2013). The AWD-specific mutations observed in *MYO5A* could be associated with white and black patches by regulating melanosome transport to the bulk of the hair (Wu et al., 1998, Evans et al., 2014) (Figure 1-4c). Similarly, sites under selection in *HPS6* may be responsible for the white and black patches by regulating melanin deposition in melanosomes (Gautam et al., 2004, Raposo and Marks, 2007) (Figure 1-4c). Mutations in this gene are associated with dilution of melanin in mice (Gautam et al., 2004) and Hermansky-Pudlak syndrome in humans (oculocutaneous albinism), a group of autosomal recessive disorders that cause abnormally light

coloring of the hair and skin (Di Pietro et al., 2004). As all AWDs are born black (Creel and Creel, 2002) and develop pigmentation patterns as puppies, longitudinal studies of gene expression will help corroborate the function of these genes in AWD pigmentation.

Conclusions

A commonly observed trend among large mammalian predators that are cursorial is the elongation of limbs and the reduction or loss of digits which allows increased speed and improved pursuit and capture of increasingly faster prey (Van Valkenburgh, 1987). Among canids, the AWD displays the most specialized morphological changes associated with cursoriality, including a unique reduction of the number of digits on the forepaws (Swann, 1904). Similarly. AWDs have specialized carnassial molars that enhance the slicing of flesh. This decreases the consumption time of prey and therefore, the likelihood of encounters with competitors (Van Valkenburgh and Koepfli, 1993). Also, while the function of the conspicuous and individual-specific coat coloration patterns of AWDs is uncertain, it may help confuse both prey and competing predators (Van Valkenburgh, 1996), or it may not be the direct object of natural selection (Anderson et al., 2009, Stahler et al., 2013).

Our comparative genomic analyses suggest that the evolution of cursoriality in AWDs has been driven by a known apoptotic pathway implicated in evolutionary digit loss in other mammals (Cooper et al., 2014) and involved a single major gene pathway. We also found evidence of substitutions and amino acid deletions in genes possibly associated with the hypercarnivorous dentition of AWDs and that changes in one of these genes (*CREBBP*) are also found in the Asiatic dhole. Consequently, our data support the idea that convergent phenotypic evolution can result from genetic changes in the same genes. Our study provides a unique example of genome-

scale adaptive evolution analysis of one of the most successful pack-hunters, the African wild dog, and exemplifies molecular pathways which can iteratively adapt organisms to the challenges of prey capture and consumption.

The unique adaptations observed in African wild dogs was likely facilitated by their unique demographic history. Most large canid linages have experienced gene flow from divergent species, whereas our inferred demographic model suggests that African wild dogs were genetically isolated from other species. Furthermore, divergence dating analyses provide a temporal framework for understanding the general rate of evolution of the molecular changes that underlie the morphological adaptations in AWDs. The earliest fossils of *Lycaon* (*L. sekowei* n. sp.) were described from sites in South Africa and dated to ca. 1.0-1.9 mya and suggest that the development of the hypercarnivorous dentition preceded the morphological changes associated with cursoriality in the modern AWD[1]. Divergence times estimated using a model that considers ancestral population size and post-divergence gene flow suggest that AWDs split from their common ancestor ~1.7 Mya, which is consistent with episodes of faunal turnover and the evolution of faster-moving ungulates during the Pleistocene that likely influence the adaptations of carnivores in sub-Saharan Africa (Turner, 1990; Bobe and Behrensmeyer, 2004). Our study demonstrates that comparative analyses of genomes provide a powerful approach to investigate the genetic basis of unique adaptations in an evolutionary context.

Methods

DNA sample and sequencing

Genomic DNA from a female African wild dog (LPI_RKW 4881) was pair-end sequenced (100 bp) to ~27.9X coverage using an Illumina HiSeq 2000 (Illumina, USA). The

library preparation and genome sequencing was performed by the Vincent J. Coates Genomics Sequencing Laboratory at University of California, Berkeley. The individual that we sequenced originally belonged to the Skukuza pack in Kruger National Park, South Africa, and was originally identified as SF5 (Girman et al., 1997). This sample was selected based on sufficient quantities of high molecular weight DNA using a DNA fluorometer (Qubit 2.0), the NanoDrop spectrophotometer (ThermoFisher, USA) and gel electrophoresis. Genome sequences from 11 other canid species, including three African wild dogs, were obtained from previous studies (Auton et al., 2013, Freedman et al., 2014, Koepfli et al., 2015, Campana et al., 2016, Robinson et al., 2016, vonHoldt et al., 2016) and are detailed in Table 1-S1.

Alignment to the dog reference genome and annotation

An initial quality control of raw reads of LPI_RKW 4881, as well as those from 11 other canids obtained from the literature, was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were then trimmed and filtered for adapters, short reads, and low-quality bases (Q<20) with Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)
before being aligned to the domestic dog genome reference assembly (canFam3.1) using Bowtie2 (Langmead et al., 2009). The percentage of aligned reads to the domestic dog for most species was more than 97 and resulted in different coverage depths per species (Table 1-S1). Variant calling was performed with HaplotypeCaller using the Genome Analysis Toolkit 3.7 (GATK; McKenna et al., 2010) with a series of filtering steps to minimize the presence of false genotypes (Supplementary methods).

Species tree estimation

To reconstruct a phylogenetic tree of the 12 canid genomes, 8,177 sliding-window fragments of 25kb were generated (Supplementary methods) and further aligned with PRANK v.150803 (Loytynoja and Goldman, 2005) using one iteration (-F once option) and the topology shown in Figure 1-1a as the guide tree. Then, windows were trimmed using GBlocks (Castresana, 2000) with default parameters. Trimmed alignments were run with RAxML v8.2.9 (Stamatakis, 2014) under the GTR model for each locus to infer individual maximum likelihood (ML) gene trees with 100 bootstrap replicates. For each locus, the best tree was selected from the RAxML output, while the 100 bootstrap trees were merged into a single file per locus. Additional alignment trimming and tree generation was done using a modified script from the SqCL pipeline (Singhal et al., 2017; phylogeny_align_genetrees_prank.py). The best tree files were concatenated into one file with only 10% missing data tolerated, collapsing branch lengths shorter than 1e-05 substitution per site, and collapsing clades with support less than one using a script from the SqCL pipeline (phylogeny_prep_astrid_astral.py).

The species tree was estimated using ASTRAL-III v.5.5, which models the discordance between gene trees and species trees using a coalescent model (Zhang et al., 2018). We used both the concatenated best-tree and bootstrap tree files as inputs. The analysis was conducted with 100 bootstrap replicates and the best multi-locus tree was selected with ML support values. The best tree was then scored to obtain a posterior probability and quartet support values for each node/clade. The gray fox (*Urocyon cinereoargenteus*) was used as the outgroup to root the tree, based on results from previous molecular systematic investigations of canid relationships (Lindblad-Toh et al., 2005).

Estimation of divergence times

To estimate the ages of divergence among species of *Canis*, *Cuon*, and *Lycaon* (11 species, including the Andean fox and gray fox as outgroups) we first generated alignments for 1,183 single-copy coding orthologues. From the codon alignments, 166,182 four-fold degenerate sites (clock-like) were extracted and concatenated into a data matrix with 6,155 missing sites (3.7%) across the 11 species. Average genomic divergence times were estimated using MCMCTree from the software package PAML4 (Yang, 2007) with the HKY+G model of nucleotide substitution, the topology of the species tree obtained from the ASTRAL-III analyses as input, and 2,200,000 MCMC generations, of which the first 200,000 generations were discarded as burn-in. Other parameter settings used in the analysis are detailed in the Supplementary File. We applied two calibration priors with associated distributions and densities based on the fossil record of the Canidae to calibrate node ages, as previously described[14]. The first prior was set at the root with an age distribution of 9.0-11.9 million years ago (Mya), which provides an approximate age for the split between the tribes Canini and Vulpini, based on the first appearance of *Eucyon*, thought to be an early member of the Canini (Tedford et al., 2009, Slater et al., 2012). The second prior had an age range of 1.1-3.0 Mya ago, based on the earliest fossils of the modern gray wolf, *Canis lupus*, and the earliest known fossils of *Canis*, specifically, *Canis edwardsii* (Tedford et al., 2009).

Phylogenetic analyses suggest *Canis edwardsii* is sister to a clade that includes *Canis aureus* (golden jackal) and *Canis latrans* (coyote) (Tedford et al., 2009). However, since *Canis lupaster* (African wolf) and *Canis simensis* (Ethiopian wolf) are also contained in the genus *Canis* (Lindblad-Toh et al., 2005, Koepfli et al., 2015, Gopalakrishnan et al., 2018), we assumed that the earliest known age of *Canis edwardsii* bracketed all extant species of *Canis*. The MCMC

analysis was repeated twice, as recommended in MCMCTree manual, and no discordance was observed between runs.

Demographic History and Admixture

Twelve canid individuals were used in the demographic analysis, including the domestic dog reference and excluding the gray fox (see Table 1-S1 for the remaining 11 genomes). Sequence alignments were obtained for these 12 genomes at 13,647 putatively neutral noncoding loci computed in previous studies to be short (1 kb long) interspersed (> 30 kb apart) and distant from protein-coding genes (> 10 kb)(Freedman et al., 2014). Of these alignments, 2,535 had more than 10% genotypes missing due to a sequencing depth below four reads or above the 95th coverage percentile. These loci were removed and the remaining 11,112 loci were analyzed using the Generalized Phylogenetic Coalescent Sampler or G-PhoCS (Gronau et al., 2011). We assumed a population phylogeny consistent with the topology of the species tree inferred by ASTRAL-III. After labeling ancestral populations, we modeled gene flow by augmenting this phylogeny with 44 directional migration bands (Supplementary methods). An additional analysis was done assuming a species tree obtained by switching the position of the golden jackal and Ethiopian wolf in the species tree inferred by ASTRAL-III with the same 44 migration bands (see Supplementary Discussion).

We ran the multi-threaded version of G-PhoCS V1.3.2 (https://github.com/gphocs-dev/G-PhoCS) using five threads per run and a standard MCMC setup. Namely, we assumed an exponential distribution with mean of 0.0001 as the prior of all the mutation-scaled population sizes (θ) and divergence times (τ), and a Gamma (α= 0.002, β= 0.00001) distribution for the prior of migration rates (m). Because of the large number of migration bands, the Monte Carlo

Markov chain was let to converge for 200,000 burn-in iterations, after which parameters were sampled every 50 iterations, for the next 400,000 iterations, resulting in a total of 8,000 samples from the approximate posterior distribution. For each parameter, we recorded the mean sampled value and the 95% Bayesian credible interval (CI). Population size estimates (Ne) were obtained from the mutation-scaled samples (θ) by assuming a mutation rate per-generation of $\mu=4.0\times10^{-9}$ (Skoglund et al., 2015), and divergence times (*T*) were calibrated by assuming the same rate and an average generation time of three years. Migration rates were scaled by the duration of time of the migration band, resulting in total rates, which approximate the probability that a lineage experienced migration. Parameter estimates are summarized in Table 1-S2 and Figure 1-1a.

Positive Selection

Coordinates of ~19,000 orthologous genes were obtained using the domestic dog reference genome (canFam3.1) available in the Ensembl database (Zerbino et al., 2018). To exclude paralogous genes, we filtered sites following previous recommendations (McCormack et al., 2013). Specifically, we filtered out sites according to the following criteria: 1) coverage higher than a 95th percentile of distribution; 2) sites that occurred in more than one locus (with fix mate in GATK); and 3) duplicated sites likely generated from PCR libraries (with PCR duplicates in GATK). We also manually checked sequences for signals of duplication events and kept only sites that were bi-alellic.

To reduce the inclusion of false signals of positive selection caused by errors in the alignment process, short regions enriched with unreasonably high rates of nonsynonymous substitutions (dN) sites were masked with the Sliding Window Alignment Masker for PAML (SWAMP) tool(Harrison et al., 2014). Specifically, a two-step masking procedure was

conducted. First, with dN ≥ 10 in a 15-codon window, followed by dN ≥ 3 in a 5-codon. This approach has been proven to effectively remove most of the problematic sequences associated with misalignment(Harrison et al., 2014).

Genes that passed our filters (no internal stop codon, permissible length and longest transcript) were then tested for signals of positive selection, based on the reconstructed species tree (Figure 1-1a), using the branch-site model in PAML 4.8 (Yang, 2007) (Supplementary methods). We run this model on the AWD and the other two pack-hunting species (dhole and gray wolf) were each used as foreground branches. Model A (allowing sites to be under positive selection; fix omega = 0) was compared to the null model A1 (sites may evolve neutrally or under purifying selection; fix omega = 1 and omega = 1). We included only genes with omega values >1, since genes lower than this threshold could be driven by relaxed selection. Statistical significance ($P < 0.05$) was assessed using likelihood ratio tests (LRTs) and chi-square tests. Multiple hypothesis correction for 18,327 protein-coding genes and three foreground branches was conducted with a 20% false discovery rate (FDR) criterion using QVALUE in R(Storey et al., 2017).

To detect recent signals of selection that include non-coding regions such as promoters and enhancers, we aimed to detect selective sweeps through an HKA-like approach (Hudson et al., 1987, Huber et al., 2016, Johnson and Voight, 2018). To determine interspecific variation, genotypic variants were called with Haplotype Caller from the Genome Analysis Toolkit 3.8 (McKenna et al., 2010). Independent gVCF files were created for the four AWD genomes (Table 1-S1) and then joined with the option “CombineGVCFs” from GATK. On the multiple-samples gVCF, per-site polymorphism among the four AWDs was calculated across non-overlapping 100kb windows with 10k steps between windows. At the same time, per-site divergence was

calculated between the high coverage AWD (LPI_RKW 4881) and the domestic dog. These estimates were calculated with the following equation:

$$\frac{\theta}{D} = \frac{\frac{\Sigma_{i=1}^{L} 2p_i(1-p)}{L} * \frac{n}{n-1}}{\frac{\Sigma_{i=1}^{L} X_i + 0.5Y_i}{L}} \quad (1)$$

where *p* is the frequency of one allele in the four AWD genomes, *L* is the total number of callable sites with good quality in the window, *n* is number of sampled chromosomes (n = 8 for 4 diploid individuals), X_i is the number of derived alleles in the AWD (LPI_RKW 4881) with respect to the dog; and Y_i is the number of heterozygous sites.

Windows were required to have at least 10 kb of sequence and sites were filtered for low coverage (less than 3x and no more than a 95th percentile of distribution), low quality variant sites (QUAL<50), missing genotype, non-bi-allelic sites, low quality genotype scores (Q< 20) and regions with high GC content. The windows with the lowest θ/D ratio were candidates for a selective sweep (Huber et al., 2016, Zhai et al., 2009). Empirical *p*-values were obtained for a total of 22,269 windows. To ensure that outliers were not driven by less sequencing data, we plotted the divergence and diversity ratio (θ/D) vs. the number of called bases per window (Figure 1-S3).

To further investigate candidate genes detected to be under positive selection, protein structure information from the relevant gene available in the literature was used. Nucleotide sequences of orthologous genes in FASTA format were translated into amino acids using "vespa.py translate" from the VEPSA tool (Webb et al., 2017). The AWD protein sequences were then aligned to human annotated versions of the orthologous genes within Geneious v11.1.1 (Kearse et al., 2012) to identify amino acid changes that occur within a given protein domain (see Figure 1-S4). When information was available, the effect of amino acid

changes on protein structure and function was evaluated using three-dimensional models depicted from SWISS-MODEL (Biasini et al., 2014). Finally, to evaluate the degree of conservation (constraint) of the AWD-specific amino acid changes and deletions, AWD alignments were compared with available orthologous coding sequences from: 1) 12 other canid species, including corresponding sequences from the genomes *Speothos venaticus* (bush dog) and *Chrysocyon brachyurus* (maned wolf) (unpublished data); and 2) 29 to 41 other species of placental, marsupial and/or monotreme mammals, using the tool OrthoMaM v9 (Douzery et al., 2014). Some amino acid changes in the *PAH* as well as the *WDR35* genes described in the result section were not observed in all samples of AWDs. However, they were confirmed on a consensus sequences of three high-coverage *de novo* ADW reference genomes (*"Dnv Lycaon pictus"* in Figure 1-4b and 1-S2). This suggests that these mutations may be heterozygotes for *PAH* and *WDR35* genes, failing to be detected in some samples due to insufficient coverage.

Enrichment test

Information about the functional impact of amino acid changes was obtained from the Ensembl Variant Effect Predictor tool (McLaren et al., 2016) with the domestic dog (Ensembl's release89) and further used to identify indels (frameshift variation, in-frame insertions and in-frame deletions), loss-of-function mutations (stop codons) and regulatory variants (5' UTR, 3' UTR) that were unique to AWDs. Genes within these categories were then tested for enrichment of Gene Ontology (GO) categories using G-profiler version r1732_e89_eg36 (Reimand et al., 2007), with a Benjamini-Hochberg (BH) false discovery rate (FDR) to correct for multiple testing.

To identify overrepresentation of genes with AWD specific-signals of positive selection depicted from the branch-site model, a GO analysis using G-profiler version r1732_e89_eg36 (Reimand et al., 2007) was performed. Ensembl identification numbers of genes that resulted from the branch-site model ($p<0.05$) were input as query lists, and all the genes tested (18,327 filtered genes) were employed as the background gene list. Multiple hypothesis correction was conducted with the Benjamini-Hochberg FDR method (Benjamini and Hochberg, 1995). Likewise, Ensembl identification numbers of genes within windows that had empirical *p*-values <0.01 from the HKA-like test were input as query lists, and all the genes tested were used as the background gene list.

Polygenic selection

To detect biological pathways overrepresented by weak to moderate signals of selection on the AWD, dhole, and gray wolf, the program polysel (Detection of polygenic selection in gene sets) was employed (Daub et al., 2017). Polysel uses information from all genes depicted by the branch-site model test to find low-level polygenic selection across many genes within a pathway. Significant pathways were identified with an FDR <0.20, after removal of the overlap between pathways with the pruning procedure implemented in polysel (see Supplementary methods for details).

Figures

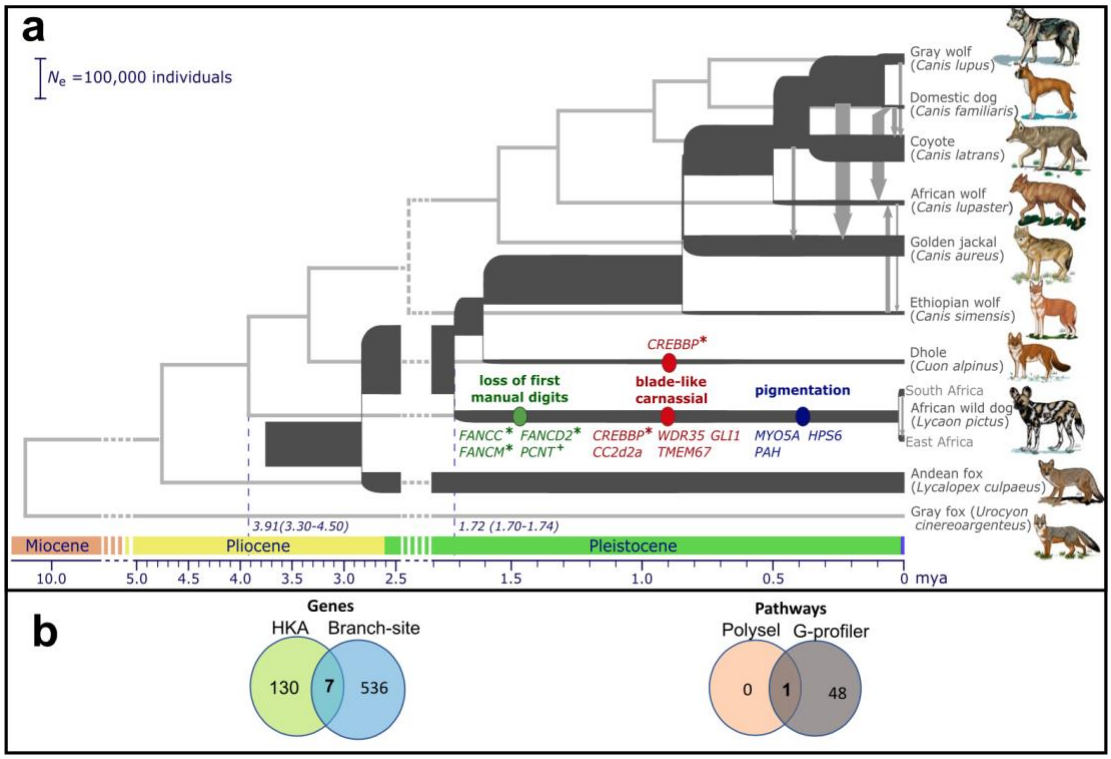

Figure 1-1. Evolutionary history and adaptation in the African wild dog (AWD) and nine other species of canids. **(a)** A species tree was inferred by applying ASTRAL-III to 8,117 25kb-windows (light gray, in background), with internal nodes placed according to average genomic divergence estimated via MCMCTree and calibrated using two fossil priors (see Methods and Figure 1-S1 for details). A demographic model was inferred for the same species (excluding gray fox) by applying G-PhoCS to 11,112 putative neutral 1kb windows (dark gray, in foreground). The same phylogenetic tree topology was assumed, augmented with 44 directed migration bands (see Methods and Table 1-S2 for details). Block arrows depict the eight migration bands inferred with total rates higher than 0.05, with arrow widths scaled proportionally to the estimated total rate. The widths of branches in the demographic model are scaled proportionally to inferred effective population sizes (see scale bar at top-left), and their lengths are scaled proportionally to inferred species divergence times. Both scales assume an average per-generation mutation rate of μ=4.0×10-9 and an average generation time of three years (Skoglund et al., 2015). Species divergence times are much smaller than the average genomic divergence times. Divergence times associated with AWD are indicated at the bottom with 95% Bayesian credible intervals (note the change in time scale between 2.5 – 5 Mya). Genes with signals of positive selection are specified on the branches leading to the AWD and the dhole. Different phenotypic categories are indicated by color; genes marked with an asterisk had in-frame deletions and genes marked with a cross were pseudogenized. Note that CREBBP has undergone parallel adaptation in both lineages. **(b)** Venn diagrams showing shared positively-selected genes (left) and pathways (right) obtained from different analytical approaches. Among the seven genes that resulted in significant scores from both the HKA-like and branch-site tests, only HPS6 was associated with AWD adaptations. Primary cilium was the only pathway that was identified by both G-profiler and polysel.

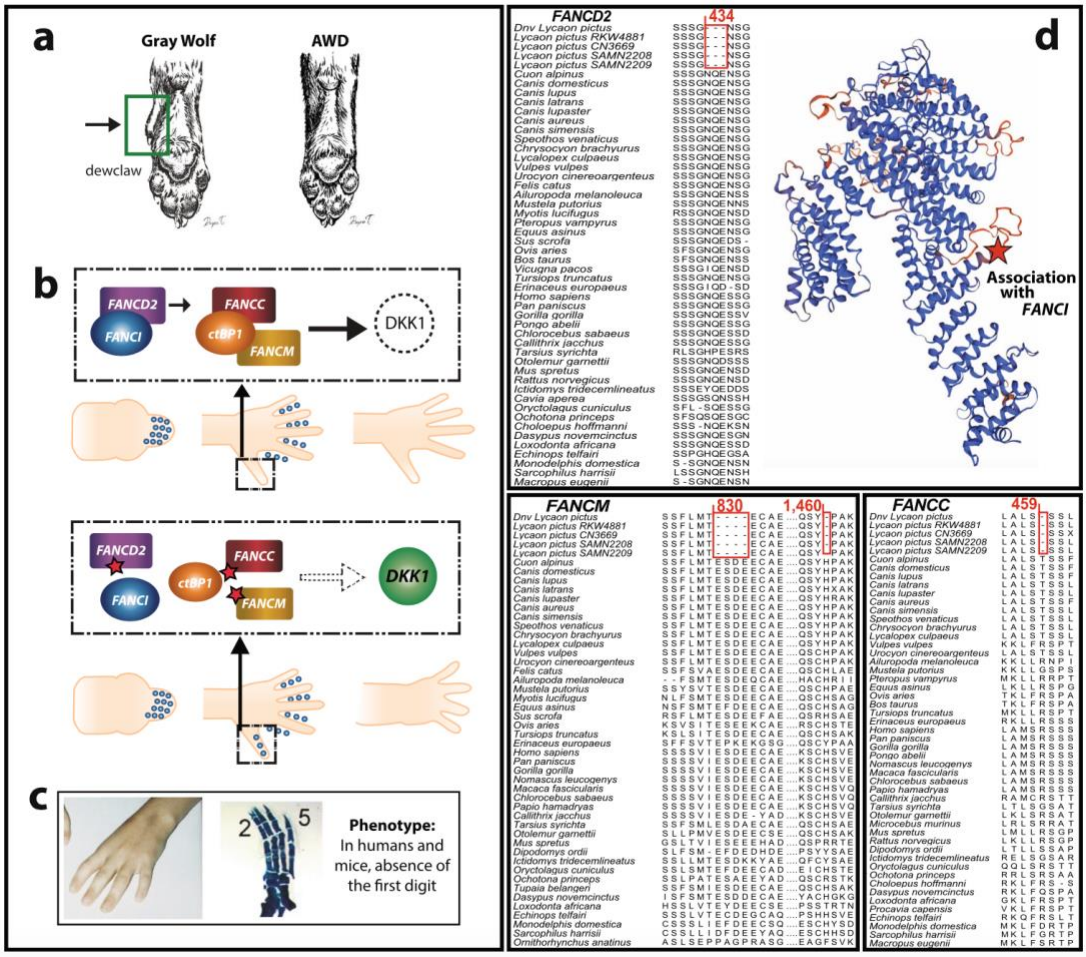

Figure 1-2. Apoptosis of the first digit in the African wild dog. **(a)** Primitive condition of five digits in the Canidae; note the small first digit in gray wolf called the "dewclaw." The absence of the first digit is shown in the African wild dog (AWD). **(b)** Schematic representation of digit reduction and separation of digits. In the normal five-digit pattern scenario shown at the top, apoptosis (blue circles) is restricted to interdigital regions. The first digit, enclosed by a rectangle, is protected from apoptosis by FANCC and FANCM that form a complex with CtBP1 and repress DKK1 (Huard et al., 2014). The stability of this complex is regulated by the FANCD2-FANCI association (Joo et al., 2011). In the scenario shown below, amino acid deletions (red stars) observed in FA genes may reduce both the affinity of FA genes to CtBP1 and the stability of the protein complex. Consequently, the FANCC-FNACM-CtBP1 complex is not formed and DKK1 is not suppressed (indicated by empty arrow). Deficiency of the FA complex activity increases DKK1 expression. As a result, apoptosis expands to the first digit; note blue circles (apoptosis) on the region of the thumb**. (c)** Effect of the mutations in the FA and DKK1 genes in humans (Reprinted from Auerbach © 2009 with permission from Elsevier) and mice (Reprinted from Adamska et al © 2004 with permission from Elsevier). **(d)** Multiple sequence alignments of mammalian FANCD2, FANCM, and FANCC amino acid sequences showing deletions specific to AWDs. Five AWDs are shown; "RWK481" and "SAMN04312208" are individuals from Kruger National Park, South Africa; "CN3669" and "SAMN04312209" are individuals from Kenya and "Dnv Lycaon pictus" is the consensus sequence of three de novo reference AWD genomes (Armstrong et al., 2018). The top panel also shows a 3D protein-model of FANCD2 with the location of the observed amino acid deletion, which is important for the association with FANCI.

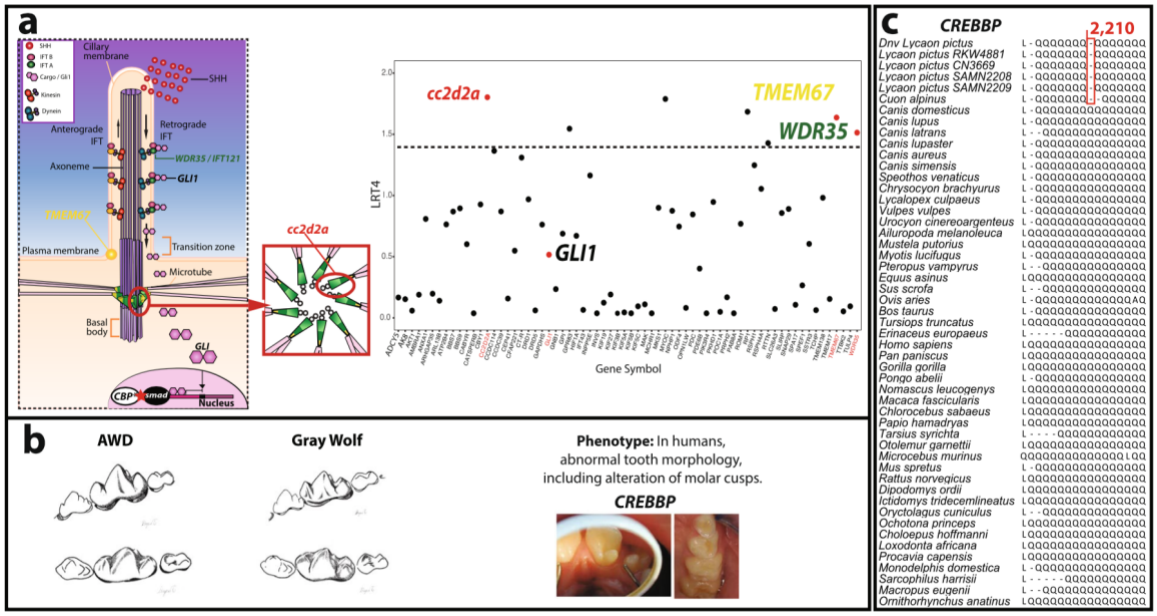

Figure 1-3. Possible mode of evolutionary molar cusp modification in the African wild dog through SHH signaling. **(a)** Figure at the top right shows genes found to be enriched in primary cilium with polysel; genes above dashed line are those found to be enriched with G-profiler as well (see Methods for details). Only gene names with known function in tooth development are shown. Figure at the top left shows a schematic of the primary cilium and the role of candidate genes in SHH transduction. SHH, represented by red circles, reaches the ciliary membrane. Then, transport of transcriptional factors GLIs such as GLI1, through the axoneme, is promoted by WDR35/IFT121 as well as the efficient docking of the axoneme in the plasma membrane, which is conducted by CC2d2a and TMEM67. Together, these components of primary cilium cause rapid accumulation of GLI into the basal body. Ultimately, GLIs will enter the nucleus and promote the expression of SHH-dependent genes. In the case of CREBBP, the observed amino acid deletion (red star) may increase the affinity of this cofactor with smad genes and increase the expression of TGF-β and BMP dependent genes involved in molar cusp development. **(b)** Left-bottom figure showing the single cuspid talonid of the lower first molar (carnassial) in the AWD as opposed to a bi-cusped talonid carnassial in the gray wolf. Right-bottom figures shows the effect that mutations in CBP have on humans (Bloch-Zupan et al., 2007) (reprinted by permission from Wiley-Liss, Inc.; American Journal of Medical Genetics Bloch-Zupan et al © 2007); talon cusp condition is observed; cusps that protrude from the anterior region of incisors on the left and extra cusps on molars on the right. **(c)** Amino acid alignment of the glutamine-rich region of CREBBP for 51 species of mammals showing the deletions specific to AWDs and the dhole. Five AWDs are shown; "RWK481" and "SAMN04312208" are individuals from South Africa; "CN3669" and "SAMN04312209" are individuals from East Africa; "Dnv Lycaon pictus" is the consensus sequence of three de novo reference AWD genomes (Armstrong et al., 2018). The ancestral condition in canids is inferred as 15 glutamine residues.

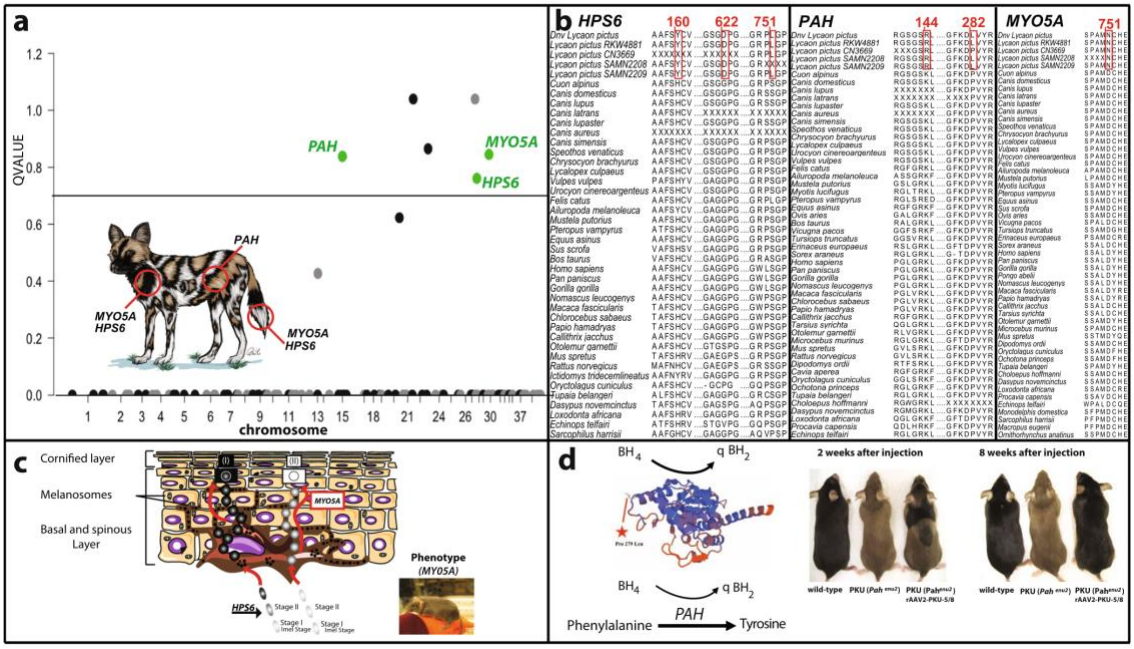

Figure 1-4. Candidate genes and possible mechanisms associated with coat pigmentation and patterning in African wild dogs. **(a)** Plot showing Q-values depicted from the branch-site test after a multiple hypothesis correction of 151 different coat color genes with the AWD as the foreground branch (see methods). Significant genes (Q-value<0.20) are shown above the horizontal blue line; for illustration purposes, Q-values shown on the Y-axis were transformed to -log10. Only names of genes with relevant functions are shown. Illustration at the bottom showing coat color in the AWD; genes are shown in red circles and are placed on the type of color they regulate (e.g., PAH regulates brown and yellow colors). **(b)** Amino acid alignment of coat color candidate genes for an average of 49 species of mammals showing changes specific to AWDs. Five AWDs are shown: "RWK481" and "SAMN04312208" are individuals from Kruger National Park, South Africa; "CN3669" and "SAMN04312209" are individuals from Kenya and "Dnv Lycaon pictus" is the consensus sequence of three de novo reference AWD genomes(Armstrong et al., 2018). **(c)** A schematic showing the role of candidate genes in color pattern. (I) shows that black coat color will need both proper cargo of melanin by HPS6 to melanosomes and its transport by MYO5A to the bulk of the hair. (II) In the case of white blotches or spots, they could be the result of either proper transport of melanosomes but containing no melanin or melanosomes containing melanin that fail to reach the bulk of the hair. Figure on the bottom-right shows the effect of MYO5A mutations on the pelage of a domestic mouse(Meehan et al., 2017). (**d)** 3D model of the PAH protein depicted with SWISS-model (see methods) that shows a mutation on the cofactor of the enzyme (BH_4). The scheme at the bottom illustrates the role of PAH and its cofactor. The right figure shows a gradual recovery of the black color of yellow mice (PKU) with deficiency of PAH (Ding et al., 2006) (Reprinted by permission from Springer Nature: Springer Nature, Gene Therapy Ding et al © 2006).

Appendix 1-I: Supplementary Discussion

Demographic model with alternative topologies

According to our inferred demographic model, gene flow in AWDs is restricted to low rates between the two African populations (7.4%), with other analyzed species experiencing much higher rates of admixture. Most notable of these is the Eurasian golden jackal (*Canis aureus*), for which previous studies have also inferred high rates of ancestral gene flow (Freedman, 2014, Fan et al., 2016).

Since the species tree inferred by ASTRAL-III does not account for gene flow, the location of the golden jackal branch could be falsely inferred as part of the clade containing all wolf species. To ensure that this uncertainty in species tree topology does not influence our estimates of demographic parameters associated with AWDs or dhole (*Cuon alpinus*), we inferred an alternative set of demographic parameters assuming a species tree in which the golden jackal is an outgroup to all wolf populations, including the Ethiopian wolf (obtained by switching the location of the golden jackal and the Ethiopian wolf in the tree).

The estimates of demographic parameters obtained under this alternative model were highly concordant with our original estimates, with overlapping credible intervals for nearly all parameters (see Table 1-S2). The main difference between the two inferred models related to the time period between divergence of the golden jackal and Ethiopian wolf from other populations (between T_{ANC5} and T_{ANC4}). Under the original model (based on the tree inferred by ASTRAL-III), the two species diverged from other lineages during a short period of time roughly 842,500 years ago, with a significant population size decrease during this brief period. However, under the alternative model, the golden jackal diverged 945,600 years ago and the Ethiopian wolf diverged 175,000 years later. This difference may indicate that the golden jackal diverged from

other *Canis* lineages prior to the Ethiopian wolf, and genomic similarity between the golden jackal and several wolf species is now partly explained by post-divergence gene flow. Importantly, this uncertainty in the topology of the species tree does not significantly influence any of the parameters associated with AWDs or the dhole.

Appendix 1-II: Supplementary Methods

Alignment to the dog reference genome and annotation

To conduct a comparison between canid genomes, reads were aligned to the domestic dog with Bowtie2 (Langmead et al., 2009). Then, duplicates that may arise during library construction using PCR were removed with Picard tools 1.80 (https://broadinstitute.github.io/picard/). To correct mapping errors made by genome aligners and to ensure that reads were properly paired, local realignment around indels was conducted with the Genome Analysis Toolkit 3.7 (GATK) IndelRealigner tool (McKenna et al., 2010). To reduce potential errors introduced by sequencing, base quality scores were recalibrated with the GATK BaseRecalibrator tool. Bases at sites with expected variation were masked with "known variants", which was obtained by calling genotypes with three different tools: the GATK HaplotypeCaller, GATK UnifiedGenotyper and SAMtools mpileup (McKenna et al., 2010, Li et al., 2009). Then, genotype calls that were found in two or more of the tools were extracted with BCFtools (Li et al., 2009). To ensure convergence between reported and empirical genotype quality scores, base score quality recalibration was run three times. Only reads that were calibrated, properly mapped and with high quality (Phred scores >= 20) were employed for further analysis.

To minimize the inclusion of erroneous genotypes, a series of filters were applied to called variants. Specifically, sites with a read depth less than 4 or greater than 95th percentiles of depth for each individual genome were filtered out. To reduce bias in downstream analysis due to base composition heterogeneity, CpG islands were removed following previous studies (Marsden et al., 2016). To ensure the inclusion of variants that were both substitutions and bi-allelic, the following sites were removed from the original VCF files: indels, multi-nucleotide polymorphisms, and sites with more than one alternate allele. Finally, to conserve only high-quality reads, only variant sites with high quality scores (GQ≥30) were retained in the final VCF file.

Collection of multi-species sequence alignments

Most of the analyses conducted in this study require multi-species alignments in FASTA format. Therefore, whole-genomes in FASTA format were obtained from BAM files by first creating pileup files using SAMtools 1.2 (mpileup with -uv options) (Li et al., 2009). Pileup files were converted to VCF files with BCFtools 1.1-108-g1844401 (-c option), filtering out low-quality variants (Phred quality scores > 20, min depth coverage > 3x, and maximum depth coverage < 95th percentile of species total coverage). Filtered VCF files were converted to consensus FASTQ files using vcf2fq in vcfutils.pl from SAMtools, and further transformed into FASTA files with seqtk v.1.2 (https://github.com/lh3/seqtk/blob/master/seqtk.c). Once whole genomes in FASTA format were obtained, specific regions (e.g neutral regions, orthologs; see next sections) were extracted with BEDtools v2.26 (-getfasta option), (http://bedtools.readthedocs.io). Finally, the multi-species FASTA sequences were aligned with

PRANK v.150803 (Loytynoja and Goldman, 2005) using one iteration (-F once option) and the topology shown in Figure 1-1a as the guide tree.

Species tree estimation

To generate 25kb genome fragments suitable for a phylogenetic reconstruction, we first called genotypic variants with Haplotype caller with GATK (McKenna et al., 2010). This generated independent gVCF files for 12 canid genomes that were combined with "CombineGVCFs" tool from GATK. Then, in conjunction with SAMtools (Li et al., 2009) and BEDtools (http://bedtools.readthedocs.io), a custom python script was written to generate non-overlapping sliding-window fragments of 25kb that mapped to the 38 canid autosomes. Window size was chosen based on previous studies of relationships among closely related species of carnivores (Kumar et al., 2017). Only windows with more than 700 informative sites were kept. A site was considered informative if at least nine genomes out of twelve had information about their genotype (i.e., either variant or invariant sites). In contrast, a site was considered not informative if more than 4 genomes had masking substitutes of low-quality base calls with 'N's (undetermined bases). This filtered out roughly 5% of the windows genome-wide. FASTA sequences were then obtained for the remaining overlapping windows.

Demographic History and Admixture

The demographic history model was conducting using the topology shown in Fig.1 along with ancestral populations that were label as follows:

ANC1 – ancestral to domestic dog and gray wolf.

ANC2 – ancestral to dog, gray wolf, and coyote.

ANC3 – ancestral to dog, gray wolf, coyote, and African wolf.

ANC4 – ancestral to dog, gray wolf, coyote, African wolf, and golden jackal.

ANC5 – ancestral to the entire *Canis* clade (including Ethiopian wolf).

ANC6 – ancestral to the *Canis* clade and dhole.

AWD – ancestral to the two sampled AWD populations (from East and South Africa)

ANC7 – ancestral to the *Canis* clade, dhole, and *AWD*.

ROOT – ancestral to all sampled individuals, including the Andean fox.

Gene flow was modeled by augmenting the phylogeny of Fig.1 with the 44 directional migration bands:

Thirteen migration bands between the golden jackal and wolf-like canids, as inferred in several previous studies (Freedman, 2014, Fan et al., 2016). Specifically, bands from dog to golden jackal and between the golden jackal and each of gray wolf, African wolf, dhole. Also, bands between the golden jackal and the three populations ancestral to both dog and gray wolf but not to Ethiopian wolf or golden jackal (ANC1, ANC2, and ANC3).

Five migration bands between the coyote and other canids, as inferred by previous studies (vonHoldt et al., 2016, Gopalakrishnan et al., 2018a): Bands from dog to coyote and between the coyote and either the gray wolf or ANC1 (the population directly ancestral to dog and gray wolf).

Eighteen migration bands between AWD and other populations: Bands between the two sampled AWD populations. Also, bands between African wolf or dhole and each of the three AWD populations (including the one directly ancestral to the two sampled populations). Additionally, four bands from gray wolf or dog into each of the two AWD populations.

Eight migration bands involving African wolf, Ethiopian wolf and dhole. Specifically, Bands from the dog and gray wolf to each of dhole, African wolf, and Ethiopian wolf. Also, bands between African wolf and Ethiopian wolf.

Positive Selection

To prepare files for the branch-site test of positive selection (Yang, 2007), a set of python scripts from the tool VESPA (Webb et al., 2017) were used. Only sequences that were confirmed as protein coding (length of the sequence divisible by 3), did not have internal stop codons, and were the longest transcript available, were kept for the positive selection analysis using "vespa.py clean". Sequences were further translated to amino acid alignments with "vespa.py translate" to ensure that no incomplete codons or internal stop codons were included. Amino acid sequences were aligned with PRANK v.150803 (Loytynoja and Goldman, 2005) using one iteration (-F once option) and the topology shown in Figure 1-1a as the guide tree. Then, a database with all amino acid sequences was created with "vespa.py create_database" to assist the next process of backward transformation from amino acid to the nucleotide sequence. The latter step was done with the command "vespa.py map_alignments". A directory structure for PAML analysis was created with a custom bash script. Each individual folder contained a multiple sequence alignment for a specific gene and a phylogenetic tree corresponding to Figure 1-1a with the labeled foreground branch. Also, each file included a control file with information about the null model (fixed Omega) and the model assuming positive selection ("free" Omega).

Polygenic selection

To conduct analyses of polygenic selection across the full set of tested genes in PAML, pathways were extracted from NCBI (https://www.ncbi.nlm.nih.gov/biosystems/), with the options "pathway"[BioSystemType] and "*Canis lupus familiaris*"[Organism]. Then, based on information available in the literature, a subset of pathways relevant to digit development, tooth formation, and pigmentation were chosen as input for polygenic signals associated with species adaptations (Kaelin and Barsh, 2013, Cooper et al., 2014, Jernvall and Thesleff, 2000, Daub et al., 2017). Once the pathways were specified, likelihood ratio test statistics of all genes in each pathway were added in the form of 'SUMSTAT' scores. To estimate the 'SUMSTAT' scores, the ΔlnL4 score was calculated as the fourth root of the log-likelihood ratio previously obtained with the branch-site test for positive selection in PAML4 (Daub et al., 2017). Ensembl gene IDs were transformed into Entrez gene IDs using information within the gene2ensembl file from the NCBI ftp server (ftp://ftp.ncbi.nih.gov/gene/DATA/gene2ensembl.gz). Once a gene set and pathways were specified as input to polysel, a null distribution for each pathway was created by randomly sampling genes from the original gene set to generate pathways of a similar size as the ones provided in the input. Then, p-values were obtained by detecting how often a 'SUMSTAT' score surpasses those from the null distribution. Finally, a multiple testing correction was performed by first removing high-scoring genes from lower level pathways in a process called pruning in polysel. Then, p-values were re-calculated again from the pruned pathways. This process was conducted 300 times, creating an empirical FDR distribution. Significant pathways were identified with an FDR <0.20.

Appendix 1-III: Supplementary Figures

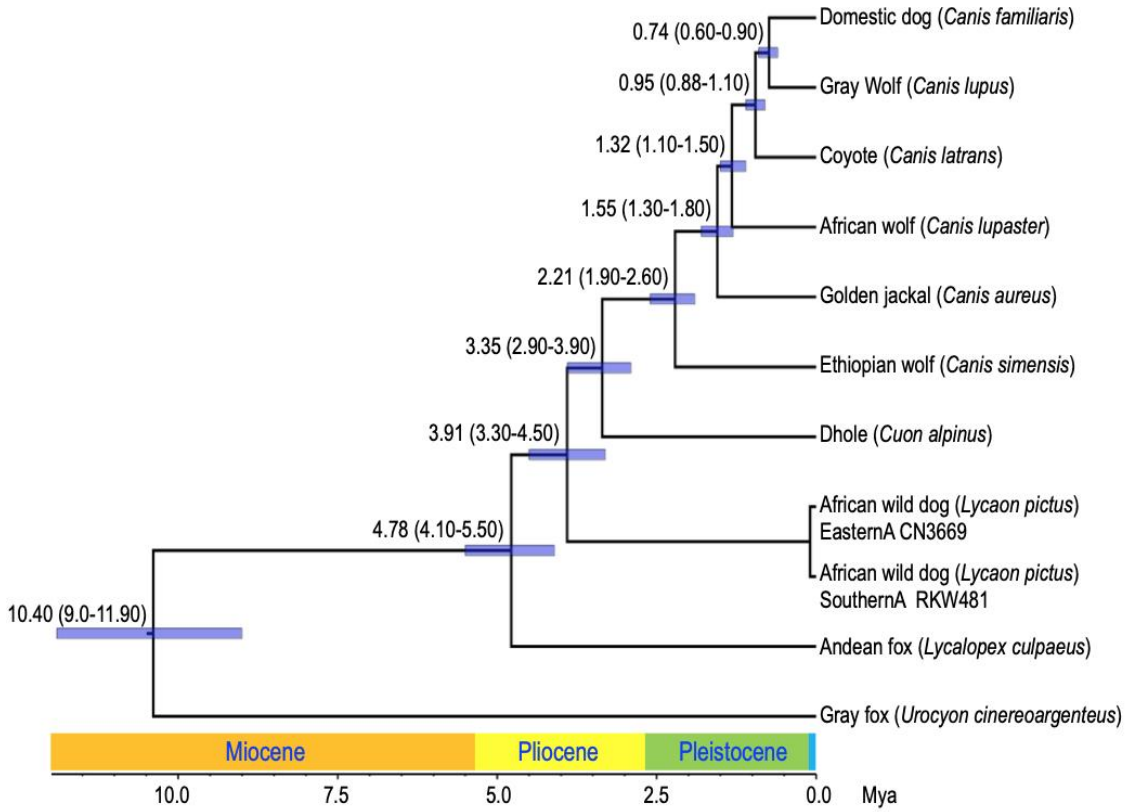

Figure 1-S1. Coalescent species tree inferred by applying ASTRAL-III to 8,117 25kb-windows with internal nodes placed according to average genomic divergence estimated via MCMCTree and calibrated using two fossil priors (see Methods for details).

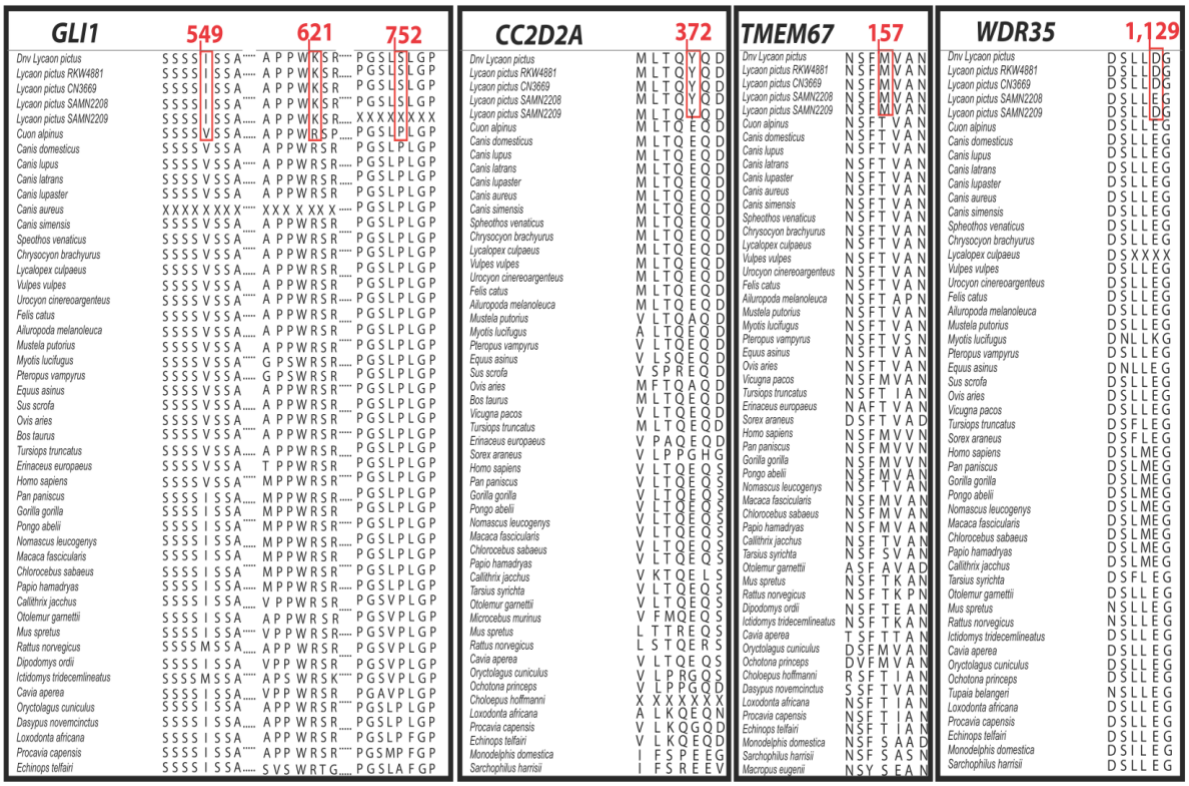

Figure 1-S2. Multiple sequence alignment of mammalian genes that are candidates for blade-like molar cusps in AWDs. Amino acid changes of AWD (Lycaon pictus) are enclosed in red rectangles. Five sequences of AWD are shown; "RWK481" and "SAMN04312208" are individuals from Kruger National Park, South Africa; "CN3669" and "SAMN04312209" are individuals from Kenya and "Dnv_Lycaon pictus" is the consensus of the tree de novo reference AWD genome.

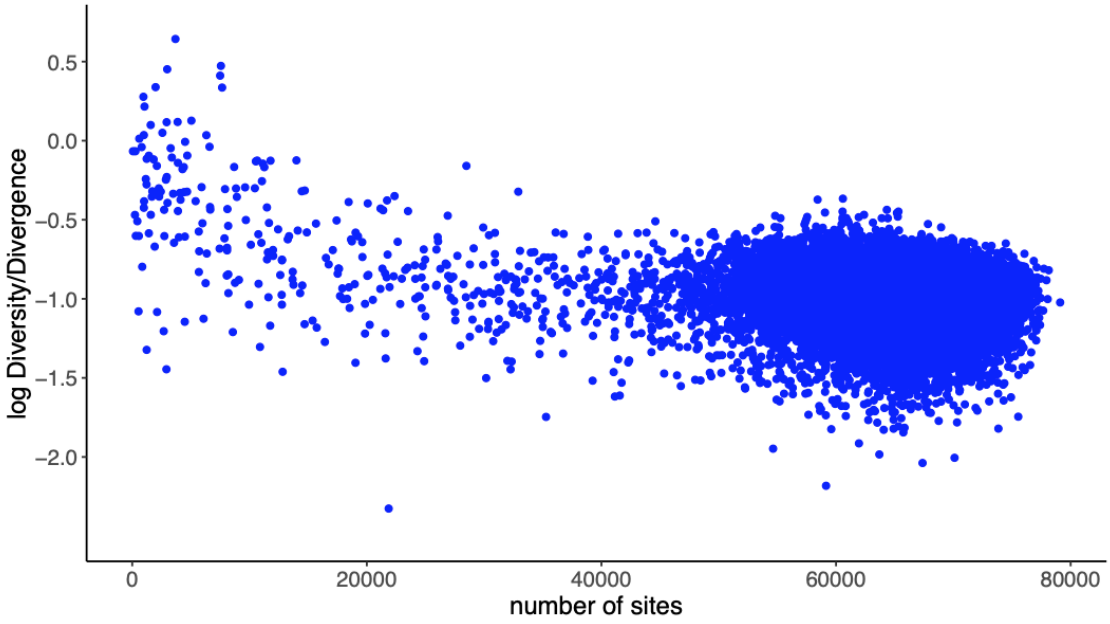

Figure 1-S3. The relationship between the ratio of diversity/divergence and amount of genotype calls (after filtering) within 100-kb windows. Each point represent a window; diversity was calculated among four African wild dog samples, and divergence was calculated by comparison of the AWD (LPI_RKW 4881) with the domestic dog (see Methods).

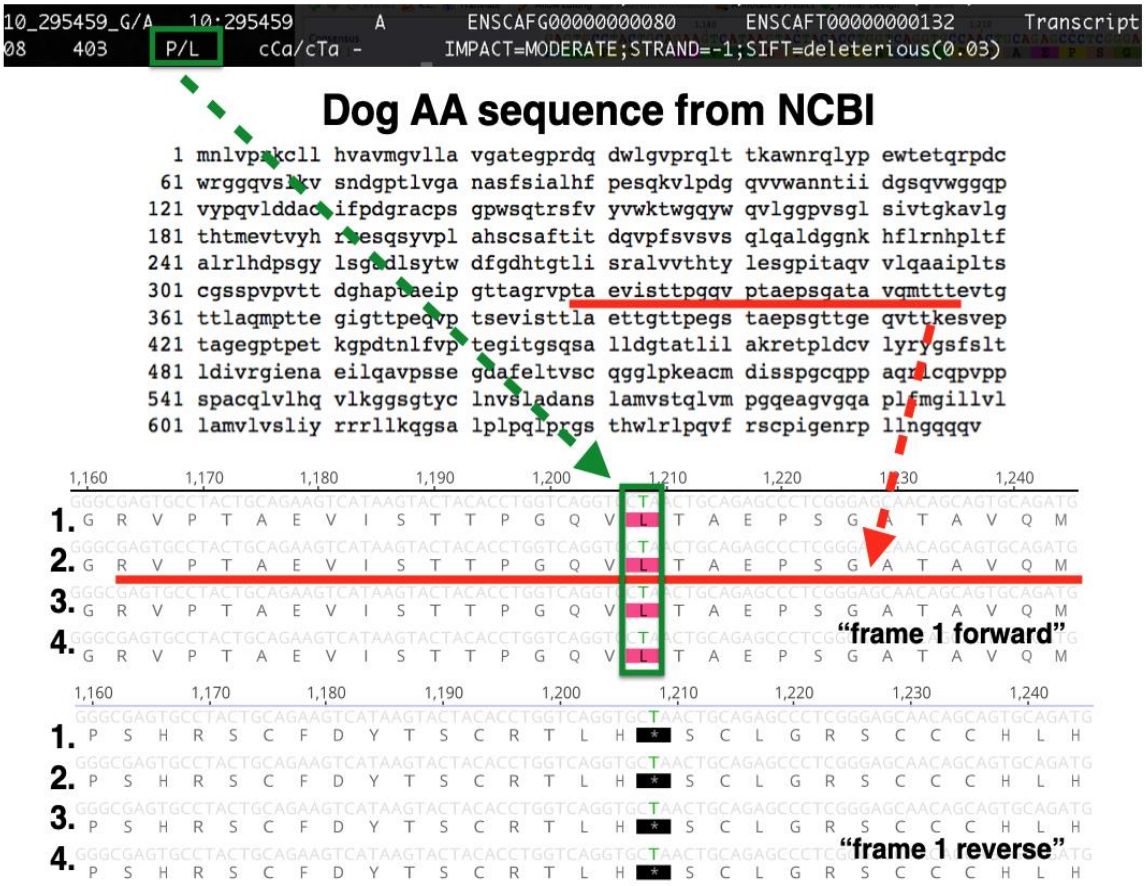

Figure 1-S4. Translation of PMEL CDS sequence with Geneious (Kearse et al. 2012) under two orientations ("forward 1" vs "reverse 1"). Stop codon reported by Campana et al., (2016) is shown by an asterisk in the alignment with the incorrect reading frame ("frame 1 reverse" at the bottom) at position 1023 bp, amino acid (AA) 341. The red arrow and line show the consistency of the correct reading frame ("forward 1" at the top) orientation with reported AA sequences in NCBI. The green arrow and rectangle indicates the concordance of the correct reading orientation ("forward 1") with the AA variation predicted by Variant Effect Predictor tool (black rectangle at the top). Numbers 1-4 indicate sequence identity among the four African wild dogs included in the analysis: 1. SAMN04312209, 2. RWK4881, 3. CN3669, and 4. SAMN04312208. The accession number for the NCBI dog AA sequence is NP_001096686.

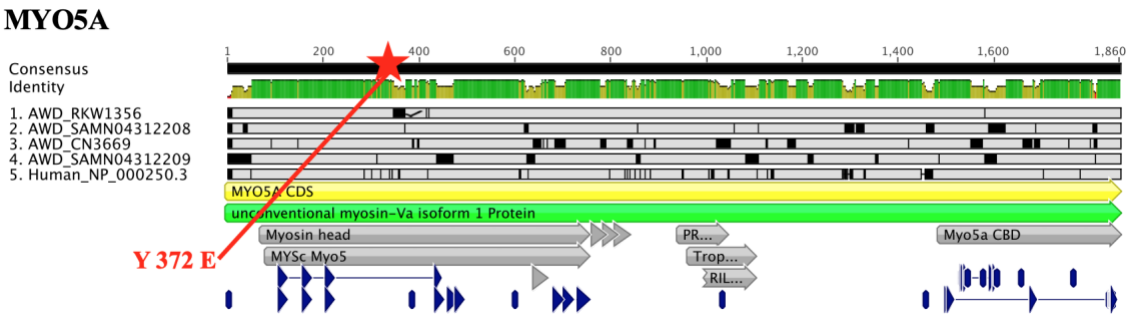

Figure 1-S5. Protein alignments showing that the amino acid mutation unique in AWDs is located at the Myosin 5A head domain. Alignment between the four AWDs (sequences 1-4) and human (sequence 5). RWK481 and SAMN04312208 are individuals from Kruger National Park, South Africa; CN3669 and SAMN04312209 are individuals from Kenya. The red star indicates the amino acid mutation, protein domains are shown in gray.

Appendix 1-IV: Supplementary Tables

Table 1-S1. Basic sequencing statistics related to the genomes analyzed in this study. Statistics were calculated with Qualimap on BAM files after base quality recalibration.

Common name	Scientific name	Sample ID	Reference	Mean depth coverage (X)	Base Pairs (Gbp)	% of reference covered by at least 3 reads	Total number of reads aligned
African wolf	*Canis lupaster*	RKW1356	[15]	27.96	65	98.11	655
Andean fox	*Lycalopex culpaeus*	SRS523207+	[16]	10.98	25.5	97.71	264
AWD East Africa	*Lycaon pictus*	SAMN04312209*	[11]	12.1	28.1	93.94	146
AWD East Africa	*Lycaon pictus*	SAMN10180432* (CN3669)	[106]	17.28	40.2	98.3	521
AWD South Africa	*Lycaon pictus*	SAMN04312208*	[11]	10.67	24.8	96.66	126
AWD South Africa	*Lycaon pictus*	SAMN09924608*	This Study	27.93	65	98.71	674
AWD East Africa	*Lycaon pictus*	SAMN09917439[1]	[13]	69	NA	NA	NA
AWD East Africa	*Lycaon pictus*	SAMN09917479[1]	[13]	46	NA	NA	NA
AWD - captive	*Lycaon pictus*	SAMN09917480[1]	[13]	25	NA	NA	NA
Coyote	*Canis latrans*	RKW13455	[17]	25.67	59.7	98.72	604
Dhole	*Cuon alpinus*	SAMN10180424*	[106]	19.53	45.4	98.35	511
Ethiopian wolf	*Canis simensis*	SAMN10180425*	[106]	9.66	22.5	97.37	247
Golden jackal	*Canis aureus*	SAMN03366713* (RKW1332)	[101]	26.09	60.7	97.46	1102
Gray fox	*Urocyon cinereoargenteus*	SAMN04495241*	[14]	18.47	42.9	97.49	450
Gray wolf	*Canis lupus*	RKW1547	[102]	27.33	63.6	98.76	643

*BioSample IDs as shown in NCBI
+trace.ddb
[1]Genomes employed exclusively for validation of specific mutations.

Table 1-S2. Demographic parameter estimates inferred by G-PhoCS. Analysis was done under a model with population phylogeny with the topology of the species tree inferred by ASTRAL-III and 44 migration bands (middle column) as well as a model in which the golden jackal and the Ethiopian wolf are switched in the population phylogeny (right column). Effective population sizes are calibrated by assuming an average per-site mutation rate of $\mu=4.0\times10^{-9}$ (Skoglund et al., 2015), and divergence times are calibrated assuming the same rate and an average generation time of three years. Total migration rates are obtained by multiplying the mutation-scaled rate with the duration of time of the migration band. For low rates, this approximates the probability that a lineage experienced migration. Total migration rates are shown for the 18 migration bands inferred by G-PhoCS to have the highest total migration rates. The total migration rates of the remaining 26 bands were estimated with mean value 0.0 and 95% Bayesian credible interval below 0.005 in the two runs. The six parameters for which the 95% Bayesian credible intervals from the two analyses do not overlap are shown in bold.

Parameter	Mean estimate (95% Bayesian CI)	
	Tree inferred by ASTRAL-III	Tree with Golden jackal and Ethiopian wolf switched
Population divergence times (years)		
T_{AWD}	26,000 (23,000-29,300)	25,700 (23,100-28,300)
T_{ANC1}	72,200 (66,700-77,800)	70,600 (63,800-81,700)
T_{ANC2}	362,500 (329,300-392,000)	349,800 (317,300-385,400)
$T_{ANC}3$	498,900 (478,100-520,400)	523,000 (492,500-565,000)
$T_{ANC}4$	**842,500 (827,600-859,500)**	**770,700 (748,200-794,100)**
$T_{ANC}5$	**842,700 (827,700-859,600)**	**945,600 (920,800-970,900)**
$T_{ANC}6$	1,610,200 (1,585,600-1,634,900)	1,597,200 (1,569,700-1,624,300)
$T_{ANC}7$	1,722,800 (1,703,900-1,741,600)	1,724,200 (1,705,300-1,742,700)
T_{ROOT}	2,823,400 (2,792,400-2,854,000)	2,823,200 (2,793,400-2,853,300)
Effective population sizes (number of individuals)		
Ne (AWD-SA)	16,400 (14,800-17,900)	15,700 (14,300-17,200)
Ne (AWD-EA)	19,300 (17,600-21,300)	20,100 (18,200-22,100)
Ne (AWD)	29,800 (29,000-30,600)	29,800 (29,000-30,600)
Ne (Gray wolf)	27,100 (25,100-29,300)	26,800 (24,200-30,100)
Ne (Coyote)	65,700 (61,700-69,800)	62,900 (58,600-67,200)
Ne (African wolf)	12,700 (11,700-13,700)	11,800 (10,500-13,000)
Ne (Golden jackal)	51,000 (49,000-53,200)	51,100 (49,000-53,200)
Ne (Ethiopian wolf)	9,100 (8,500-9,700)	9,600 (9,000-10,200)
Ne (Dhole)	11,200 (10,600-11,700)	11,200 (10,600-11,700)
Ne (Andean fox)	50,500 (49,100-51,900)	50,500 (49,100-52,000)
Ne (ANC1)	124,300 (115,500-132,700)	127,800 (117,400-136,600)
Ne (ANC2)	176,400 (140,200-210,300)	190,200 (162,100-220,800)
Ne (ANC3)	122,200 (114,400-130,800)	105,600 (88,300-118,700)
***Ne* (ANC4)**	**12,000 (1,600-35,600)**	**162,200 (138,100-186,100)**
Ne (ANC5)	119,700 (115,400-123,800)	111,400 (105,700-117,000)
Ne (ANC6)	92,200 (69,000-115,800)	105,600 (79,100-130,800)
Ne (ANC7)	166,600 (160,300-173,000)	166,200 (160,000-172,800)
Ne (ROOT)	108,700 (103,400-114,100)	108,700 (103,400-114,000)
Total migration rates		
m(ANC1→Golden jackal)	0.376 (0.306-0.474)	0.405 (0.333-0.469)
m(Dog→African wolf)	0.326 (0.299-0.354)	0.340 (0.307-0.378)
m(Dog→Coyote)	0.155 (0.116-0.192)	0.132 (0.083-0.177)
m(Ethiopian wolf→African wolf)	0.135 (0.124-0.148)	0.148 (0.134-0.163)
m(ANC2→Golden jackal)	0.111 (0.000-0.188)	0.242 (0.159-0.320)
m(Gray wolf→Coyote)	0.092 (0.064-0.129)	0.121 (0.091-0.158)
m(AWD-SA→AWD-EA)	0.074 (0.036-0.135)	0.028 (0.013-0.046)

***m*(African wolf→ Ethiopian wolf)**	**0.060 (0.045-0.073)**	**0.023 (0.011-0.036)**
m(Coyote→Gray wolf)	0.044 (0.028-0.057)	0.041 (0.025-0.058)
m(African wolf→Golden Jackal)	0.042 (0.033-0.052)	0.048 (0.036-0.058)
***m*(Coyote→ANC1)**	**0.041 (0.026-0.063)**	**0.002 (0.000-0.014)**
m(Golden jackal →Dhole)	0.030 (0.025-0.036)	0.033 (0.026-0.040)
m(Dog→ Golden jackal)	0.027 (0.006-0.045)	0.003 (0.000-0.017)
m(Dog→Ethiopian Wolf)	0.010 (0.006-0.014)	0.011 (0.008-0.014)
m(ANC1→Coyote)	0.003 (0.000-0.031)	0.027 (0.000-0.073)
m(Golden jackal→African Wolf)	0.002 (0.000-0.011)	0.001 (0.000-0.010)
m(ANC3→Golden jackal)	0.001 (0.000-0.012)	0.002 (0.000-0.014)
***m*(AWD-EA→WildDog-SA)**	**0.001 (0.000-0.007)**	**0.038 (0.020-0.062)**

Table 1-S3. Genes detected to be under positive selection using the branch-site test in PAML4 and after multiple hypothesis-testing correction on three foreground branches (African wild dogs, dhole, and gray wolf) using 18,327 genes. Only genes significant for the branch of the African wild dog (AWD) and with qvalues < 0.20 are shown. Unavailable gene name is indicated by "N/A.

Ensembl ID	Gene symbol	P value	Q value
ENSCAFG00000014773	HSPG2	0	0
ENSCAFG00000015624	LEO1	8.33E-13	2.29E-08
ENSCAFG00000030852	N/A	8.33E-10	1.53E-05
ENSCAFG00000000070	PHLPP1	7.72E-09	0.000106117
ENSCAFG00000015593	N/A	9.24E-08	0.001016086
ENSCAFG00000004101	BICRA	1.53E-07	0.00136672
ENSCAFG00000006937	N/A	1.74E-07	0.00136672
ENSCAFG00000006714	NLRP14	2.39E-07	0.001642617
ENSCAFG00000005629	N/A	2.95E-07	0.001654988
ENSCAFG00000029964	N/A	3.01E-07	0.001654988
ENSCAFG00000003045	ZNF292	1.31E-06	0.005301932
ENSCAFG00000011482	PABPN1	1.35E-06	0.005301932
ENSCAFG00000016111	FOXK1	1.91E-06	0.007001169
ENSCAFG00000003025	RASSF6	2.13E-06	0.007277162
ENSCAFG00000007121	FEZF2	2.25E-06	0.007277162
ENSCAFG00000012783	NHS	2.50E-06	0.007636528
ENSCAFG00000012447	TRAF3IP1	3.70E-06	0.010707216
ENSCAFG00000024733	N/A	4.52E-06	0.012426158
ENSCAFG00000009552	IMPG2	4.92E-06	0.012881731
ENSCAFG00000023991	N/A	6.50E-06	0.016244977
ENSCAFG00000017716	EEF2K	8.38E-06	0.020032937
ENSCAFG00000007213	CHPT1	1.04E-05	0.023825967
ENSCAFG00000001545	TTC39B	1.09E-05	0.023972588
ENSCAFG00000031056	MANEA	1.25E-05	0.025455093
ENSCAFG00000029006	N/A	1.25E-05	0.025455093
ENSCAFG00000005441	DLL3	1.84E-05	0.036131686
ENSCAFG00000001554	BOP1	3.10E-05	0.058774931
ENSCAFG00000011856	N/A	3.27E-05	0.05993147
ENSCAFG00000003816	SMPD2	5.06E-05	0.089746445
ENSCAFG00000015595	IGSF21	5.71E-05	0.096381965
ENSCAFG00000000938	MCM9	5.94E-05	0.096381965
ENSCAFG00000005330	SH2D4A	5.96E-05	0.096381965
ENSCAFG00000016793	ZNF646	6.47E-05	0.101640003
ENSCAFG00000001178	COL22A1	7.16E-05	0.109355078
ENSCAFG00000004611	GRIN2C	9.37E-05	0.139240732
ENSCAFG00000011155	STARD9	0.000105578	0.151227217
ENSCAFG00000031292	NKX1-2	0.000107267	0.151227217
ENSCAFG00000029563	DIO3	0.000121385	0.166852786
ENSCAFG00000032438	N/A	0.000139159	0.186619007
ENSCAFG00000031975	N/A	0.000144209	0.188786749
ENSCAFG00000025358	KRT72	0.000156208	0.198060984
ENSCAFG00000011046	BCL9	0.000160622	0.198060984
ENSCAFG00000024785	ANKRD40	0.0001621	0.198060984

Reference

ABDELHAMED, Z. A., NATARAJAN, S., WHEWAY, G., INGLEHEARN, C. F., TOOMES, C., JOHNSON, C. A. & JAGGER, D. J. 2015. The Meckel-Gruber syndrome protein TMEM67 controls basal body positioning and epithelial branching morphogenesis in mice via the non-canonical Wnt pathway. *Disease Models & Mechanisms,* 8**,** 527-541.

ADAMSKA, M., MACDONALD, B. T., SARMAST, Z. H., OLIVER, E. R. & MEISLER, M. H. 2004. En1 and Wnt7a interact with Dkk1 during limb development in the mouse. *Developmental Biology,* 272**,** 134-144.

ANDERSON, T. M., VONHOLDT, B. M., CANDILLE, S. I., MUSIANI, M., GRECO, C., STAHLER, D. R., SMITH, D. W., PADHUKASAHASRAM, B., RANDI, E., LEONARD, J. A., BUSTAMANTE, C. D., OSTRANDER, E. A., TANG, H., WAYNE, R. K. & BARSH, G. S. 2009. Molecular and Evolutionary History of Melanism in North American Gray Wolves. *Science,* 323**,** 1339-1343.

ARMSTRONG, E. E., TAYLOR, R. W., PROST, S., BLINSTON, P., VAN-DER-MEER, E., MADZIKANDA, H., MUFUTE, O., MANDISODZA-CHIKEREMA, R., STUELPNAGEL, J., SILLERO-ZUBIRI, C. & PETROV, D. 2018. Cost-effective assembly of the African wild dog (*Lycaon pictus*) genome using linked reads.

AUERBACH, A. D. 2009. Fanconi anemia and its diagnosis. *Mutation Research,* 668**,** 4-10.

AUTON, A., LI, Y. R., KIDD, J., OLIVEIRA, K., NADEL, J., HOLLOWAY, J. K., HAYWARD, J. J., COHEN, P. E., GREALLY, J. M., WANG, J., BUSTAMANTE, C. D. & BOYKO, A. R. 2013. Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. *Plos Genetics,* 9.

BANGS, F. & ANDERSON, K. V. 2017. Primary Cilia and Mammalian Hedgehog Signaling. *Cold Spring Harbor Perspectives in Biology,* 9.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57**,** 289-300.

BIASINI, M., BIENERT, S., WATERHOUSE, A., ARNOLD, K., STUDER, G., SCHMIDT, T., KIEFER, F., CASSARINO, T. G., BERTONI, M., BORDOLI, L. & SCHWEDE, T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research,* 42**,** W252-W258.

BLOCH-ZUPAN, A., STACHTOU, J., EMMANOUIL, D., ARVEILER, B., GRIFFITHS, D. & LACOMBE, D. 2007. Oro-dental features as useful diagnostic tool in Rubinstein-Taybi syndrome. *American Journal of Medical Genetics Part A,* 143A**,** 570-573.

BOBE, R. & BEHRENSMEYER, A. K. 2004. The expansion of grassland ecosystems in Africa in relation to mammalian evolution and the origin of the genus Homo. *Palaeogeography Palaeoclimatology Palaeoecology,* 207**,** 399-420.

CAMPANA, M. G., PARKER, L. D., HAWKINS, M. T. R., YOUNG, H. S., HELGEN, K. M., GUNTHER, M. S., WOODROFFE, R., MALDONADO, J. E. & FLEISCHER, R. C. 2016. Genome sequence, population history, and pelage genetics of the endangered African wild dog (Lycaon pictus). *Bmc Genomics,* 17.

CANDILLE, S. I., KAELIN, C. B., CATTANACH, B. M., YU, B., THOMPSON, D. A., NIX, M. A., KERNS, J. A., SCHMUTZ, S. M., MILLHAUSER, G. L. & BARSH, G. S. 2007. A beta-defensin mutation causes black coat color in domestic dogs. *Science,* 318**,** 1418-1423.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution,* 17**,** 540-552.

COOPER, K. L., SEARS, K. E., UYGUR, A., MAIER, J., BACZKOWSKI, K. S., BROSNAHAN, M., ANTCZAK, D., SKIDMORE, J. A. & TABIN, C. J. 2014. Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature,* 511**,** 41-U537.

CRAWFORD, N. G., KELLY, D. E., HANSEN, M. E. B., BELTRAME, M. H., FAN, S. H., BOWMAN, S. L., JEWETT, E., RANCIARO, A., THOMPSON, S., LO, Y., PFEIFER, S. P., JENSEN, J. D., CAMPBELL, M. C., BEGGS, W., HORMOZDIARI, F., MPOLOKA, S. W., MOKONE, G. G., NYAMBO, T., MESKEL, D. W., BELAY, G., HAUT, J., ROTHSCHILD, H., ZON, L., ZHOU, Y., KOVACS, M. A., XU, M., ZHANG, T. W., BISHOP, K., SINCLAIR, J., RIVAS, C., ELLIOT, E., CHOI, J., LI, S. A., HICKS, B., BURGESS, S., ABNET, C., WATKINS-CHOW, D. E., OCEANA, E., SONG, Y. S., ESKIN, E., BROWN, K. M., MARKS, M. S., LOFTUS, S. K., PAVAN, W. J., YEAGER, M., CHANOCK, S., TISHKOFF, S. A. & PROGRA, N. C. S. 2017. Loci associated with skin pigmentation identified in African populations. *Science,* 358**,** 887-+.

CREEL, S. & CREEL, N. M. 2002. *The African wild dog: behavior, ecology and conservation,* Princeton, New Jersey., Princeton University Press.

DAUB, J. T., MORETTI, S., DAVYDOV, II, EXCOFFIER, L. & ROBINSON-RECHAVI, M. 2017. Detection of Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. *Molecular Biology and Evolution,* 34**,** 1391-1402.

DI PIETRO, S. M., FALCÓN-PÉREZ, J. M. & DELL'ANGELICA, E. C. 2004. Characterization of BLOC-2, a complex containing the Hermansky-Pudlak syndrome proteins HPS3, HPS5 and HPS6., 5.

DING, Z., GEORGIEV, P. & THONY, B. 2006. Administration-route and gender-independent long-term therapeutic correction of phenylketonuria (PKU) in a mouse model by recombinant adeno-associated virus 8 pseudotyped vector-mediated gene transfer. *Gene Therapy,* 13**,** 587-593.

DING, Z., HARDING, C. O., REBUFFAT, A., ELZAOUK, L., WOLFF, J. A. & THONY, B. 2008. Correction of murine PKU following AAV-mediated intramuscular expression of a complete phenylalanine hydroxylating system. *Molecular Therapy,* 16**,** 673-681.

DOUZERY, E. J. P., SCORNAVACCA, C., ROMIGUIER, J., BELKHIR, K., GALTIER, N., DELSUC, F. & RANWEZ, V. 2014. OrthoMaM v8: A Database of Orthologous Exons and Coding Sequences for Comparative Genomics in Mammals. *Molecular Biology and Evolution,* 31**,** 1923-1928.

ERLANDSEN, H., FUSETTI, F., MARTINEZ, A., HOUGH, E., FLATMARK, T. & STEVENS, R. C. 1997. Crystal structure of the catalytic domain of human phenylalanine hydroxylase reveals the structural basis for phenylketonuria. *Nature Structural Biology,* 4**,** 995-1000.

EVANS, R. D., ROBINSON, C., BRIGGS, D. A., TOOTH, D. J., RAMALHO, J. S., CANTERO, M., MONTOLIU, L., PATEL, S., SVIDERSKAYA, E. V. & HUME, A. N. 2014. Myosin-Va and Dynamic Actin Oppose Microtubules to Drive Long-Range Organelle Transport. *Current Biology,* 24**,** 1743-1750.

FAN, Z. X., SILVA, P., GRONAU, I., WANG, S. G., ARMERO, A. S., SCHWEIZER, R. M., RAMIREZ, O., POLLINGER, J., GALAVERNI, M., DEL-VECCHYO, D. O., DU, L. M., ZHANG, W. P., ZHANG, Z. H., XING, J. C., VILA, C., MARQUES-BONET, T., GODINHO, R., YUE, B. S. & WAYNE, R. K. 2016. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Research,* 26**,** 163-173.

FIGUEIRO, H. V., LI, G., TRINDADE, F. J., ASSIS, J., PAIS, F., FERNANDES, G., SANTOS, S. H. D., HUGHES, G. M., KOMISSAROV, A., ANTUNES, A., TRINCA, C. S., RODRIGUES, M. R., LINDEROTH, T., BI, K., SILVEIRA, L., AZEVEDO, F. C. C., KANTEK, D., RAMALHO, E., BRASSALOTI, R. A., VILLELA, P. M. S., NUNES, A. L. V., TEIXEIRA, R. H. F., MORATO, R. G., LOSKA, D., SARAGUETA, P., GABALDON, T., TEELING, E. C., O'BRIEN, S. J., NIELSEN, R., COUTINHO, L. L., OLIVEIRA, G., MURPHY, W. J. & EIZIRIK, E. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Science Advances,* 3.

FREEDMAN, A. H. 2014. Genome Sequencing Highlights the Dynamic Early History of Dogs (vol 10, e1004016, 2014). *Plos Genetics,* 10.

FREEDMAN, A. H., GRONAU, I., SCHWEIZER, R. M., ORTEGA-DEL VECCHYO, D., HAN, E. J., SILVA, P. M., GALAVERNI, M., FAN, Z. X., MARX, P., LORENTE-GALDOS, B., BEALE, H., RAMIREZ, O., HORMOZDIARI, F., ALKAN, C., VILA, C., SQUIRE, K., GEFFEN, E., KUSAK, J., BOYKO, A. R., PARKER, H. G., LEE, C., TADIGOTLA, V., SIEPEL, A., BUSTAMANTE, C. D., HARKINS, T. T., NELSON, S. F., OSTRANDER, E. A., MARQUES-BONET, T., WAYNE, R. K. & NOVEMBRE, J. 2014. Genome Sequencing Highlights the Dynamic Early History of Dogs. *Plos Genetics,* 10.

GAUTAM, R., CHINTALA, S., LI, W., ZHANG, Q., TAN, J., NOVAK, E. K., DI PIETRO, S. M., DELL'ANGELICA, E. C. & SWANK, R. T. 2004. The Hermansky-Pudlak syndrome

3 (cocoa) protein is a component of the biogenesis of lysosome-related organelles complex-2 (BLOC-2). *Journal of Biological Chemistry,* 279**,** 12935-12942.

GERALD, M. S. 2001. Primate colour predicts social status and aggressive outcome. *Animal Behaviour,* 61**,** 559-566.

GIRMAN, D. J., MILLS, M. G. L., GEFFEN, E. & WAYNE, R. K. 1997. A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African wild dog (Lycaon pictus). *Behavioral Ecology and Sociobiology,* 40**,** 187-198.

GOETZ, S. C. & ANDERSON, K. V. 2010. The primary cilium: a signalling centre during vertebrate development. *Nature Reviews Genetics,* 11**,** 331-344.

GOPALAKRISHNAN, S., SINDING, M. H. S., RAMOS-MADRIGAL, J., NIEMANN, J., CASTRUITA, J. A. S., VIEIRA, F. G., CAROE, C., MONTERO, M. D., KUDERNA, L., SERRES, A., GONZALEZ-BASALLOTE, V. M., LIU, Y. H., WANG, G. D., MARQUES-BONET, T., MIRARAB, S., FERNANDES, C., GAUBERT, P., KOEPFLI, K. P., BUDD, J., RUENESS, E. K., HEIDE-JORGENSEN, M. P., PETERSEN, B., SICHERITZ-PONTEN, T., BACHMANN, L., WIIG, O., HANSEN, A. J. & GILBERT, M. T. P. 2018a. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. *Current Biology,* 28**,** 3441-3449.

GOPALAKRISHNAN, S., SINDING, M. H. S., RAMOS-MADRIGAL, J., NIEMANN, J., CASTRUITA, J. A. S., VIEIRA, F. G., CAROE, C., MONTERO, M. D., KUDERNA, L., SERRES, A., GONZALEZ-BASALLOTE, V. M., LIU, Y. H., WANG, G. D., MARQUES-BONET, T., MIRARAB, S., FERNANDES, C., GAUBERT, P., KOEPFLI, K. P., BUDD, J., RUENESS, E. K., HEIDE-JORGENSEN, M. P., PETERSEN, B., SICHERITZ-PONTEN, T., BACHMANN, L., WIIG, O., HANSEN, A. J. & GILBERT, M. T. P. 2018b. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. *Current Biology,* 28**,** 3441-+.

GRONAU, I., HUBISZ, M. J., GULKO, B., DANKO, C. G. & SIEPEL, A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics,* 43**,** 1031-U151.

GROTEWOLD, L. & RUTHER, U. 2002. The Wnt antagonist Dickkopf-1 is regulated by Bmp signaling and c-Jun and modulates programmed cell death. *Embo Journal,* 21**,** 966-975.

HAMPL, M., CELA, P., SZABO-ROGERS, H. L., BOSAKOVA, M. K., DOSEDELOVA, H., KREJCI, P. & BUCHTOVA, M. 2017. Role of Primary Cilia in Odontogenesis. *Journal of Dental Research,* 96**,** 965-974.

HARRISON, P. W., JORDAN, G. E. & MONTGOMERY, S. H. 2014. SWAMP: Sliding Window Alignment Masker for PAML. *Evolutionary Bioinformatics,* 10.

HARTSTONE-ROSE, A., WERDELIN, L., DE RUITER, D. J., BERGER, L. R. & CHURCHILL, S. E. 2010. The Plio-Pleistocene ancestor of wild dogs, *Lycaon sekowein*. sp. *Journal of Paleontology,* 84**,** 299-308.

HOFFER, J. L., FRYSSIRA, H., KONSTANTINIDOU, A. E., ROPERS, H. H. & TZSCHACH, A. 2013. Novel WDR35 mutations in patients with cranioectodermal dysplasia (Sensenbrenner syndrome). *Clinical Genetics,* 83**,** 92-95.

HUANG, X. & CHAI, Y. TGF-β signaling and tooth development. *Chinese Journal of Dental Research,* 13**,** 7-13.

HUARD, C. C., TREMBLAY, C. S., MAGRON, A., LEVESQUE, G. & CARREAU, M. 2014. The Fanconi anemia pathway has a dual function in Dickkopf-1 transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 2152-2157.

HUBER, C. D., DEGIORGIO, M., HELLMANN, I. & NIELSEN, R. 2016. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Molecular Ecology,* 25**,** 142-156.

HUDSON, R. R., KREITMAN, M. & AGUADE, M. 1987. A TEST OF NEUTRAL MOLECULAR EVOLUTION BASED ON NUCLEOTIDE DATA. *Genetics,* 116**,** 153-159.

JANKNECHT, R., WELLS, N. J. & HUNTER, T. 1998. TGF-beta-stimulated cooperation of Smad proteins with the coactivators CBP/p300. *Genes & Development,* 12**,** 2114-2119.

JERNVALL, J. & THESLEFF, I. 2000. Reiterative signaling and patterning during mammalian tooth morphogenesis. *Mechanisms of Development,* 92**,** 19-29.

JOHNSON, K. E. & VOIGHT, B. F. 2018. Patterns of shared signatures of recent positive selection across human populations. *Nature Ecology & Evolution,* 2**,** 713-720.

JOO, W., XU, G. Z., PERSKY, N. S., SMOGORZEWSKA, A., RUDGE, D. G., BUZOVETSKY, O., ELLEDGE, S. J. & PAVLETICH, N. P. 2011. Structure of the FANCI-FANCD2 Complex: Insights into the Fanconi Anemia DNA Repair Pathway. *Science,* 333**,** 312-316.

KAELIN, C. B. & BARSH, G. S. 2013. Genetics of Pigmentation in Dogs and Cats. *Annual Review of Animal Biosciences,* 1**,** 125-156.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

KOEPFLI, K. P., POLLINGER, J., GODINHO, R., ROBINSON, J., LEA, A., HENDRICKS, S., SCHWEIZER, R. M., THALMANN, O., SILVA, P., FAN, Z. X., YURCHENKO, A. A., DOBRYNIN, P., MAKUNIN, A., CAHILL, J. A., SHAPIRO, B., ALVARES, F., BRITO, J. C., GEFFEN, E., LEONARD, J. A., HELGEN, K. M., JOHNSON, W. E., O'BRIEN, S. J., VAN VALKENBURGH, B. & WAYNE, R. K. 2015. Genome-wide

Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Current Biology,* 25**,** 2158-2165.

KRAUS, W. L., MANNING, E. T. & KADONAGA, J. T. 1999. Biochemical analysis of distinct activation functions in p300 that enhance transcription initiation with chromatin templates. *Molecular and Cellular Biology,* 19**,** 8123-8135.

KUMAR, V., LAMMERS, F., BIDON, T., PFENNINGER, M., KOLTER, L., NILSSON, M. A. & JANKE, A. 2017. The evolutionary history of bears is characterized by gene flow across species. *Scientific Reports,* 7.

LAMOREUX, M. L., VÉRONIQUE, D., LIONEL, L. & DOROTHY, B. 2010. *The colors of mice: a model genetic network*, John Wiley & Sons.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,* 10.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA, P. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

LIEM, K. F., ASHE, A., HE, M., SATIR, P., MORAN, J., BEIER, D., WICKING, C. & ANDERSON, K. V. 2012. The IFT-A complex regulates Shh signaling through cilia structure and membrane protein trafficking. *Journal of Cell Biology,* 197**,** 789-800.

LINDBLAD-TOH, K., WADE, C. M., MIKKELSEN, T. S., KARLSSON, E. K., JAFFE, D. B., KAMAL, M., CLAMP, M., CHANG, J. L., KULBOKAS, E. J., ZODY, M. C., MAUCELI, E., XIE, X. H., BREEN, M., WAYNE, R. K., OSTRANDER, E. A., PONTING, C. P., GALIBERT, F., SMITH, D. R., DEJONG, P. J., KIRKNESS, E., ALVAREZ, P., BIAGI, T., BROCKMAN, W., BUTLER, J., CHIN, C. W., COOK, A., CUFF, J., DALY, M. J., DECAPRIO, D., GNERRE, S., GRABHERR, M., KELLIS, M., KLEBER, M., BARDELEBEN, C., GOODSTADT, L., HEGER, A., HITTE, C., KIM, L., KOEPFLI, K. P., PARKER, H. G., POLLINGER, J. P., SEARLE, S. M. J., SUTTER, N. B., THOMAS, R., WEBBER, C., LANDER, E. S. & BROAD INST GENOME SEQUENCING, P. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature,* 438**,** 803-819.

LOYTYNOJA, A. & GOLDMAN, N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 10557-10562.

MARSDEN, C. D., ORTEGA-DEL VECCHYO, D., O'BRIEN, D. P., TAYLOR, J. F., RAMIREZ, O., VILA, C., MARQUES-BONET, T., SCHNABEL, R. D., WAYNE, R. K. & LOHMUELLER, K. E. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** 152-157.

MARTINEZ-NAVARRO, B. & ROOK, L. 2003. Gradual evolution in the African hunting dog lineage - Systematic implications. *Comptes Rendus Palevol,* 2**,** 695-702.

MCCORMACK, J. E., HIRD, S. M., ZELLMER, A. J., CARSTENS, B. C. & BRUMFIELD, R. T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution,* 66**,** 526-538.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20**,** 1297-1303.

MCLAREN, W., GIL, L., HUNT, S. E., RIAT, H. S., RITCHIE, G. R. S., THORMANN, A., FLICEK, P. & CUNNINGHAM, F. 2016. The Ensembl Variant Effect Predictor. *Genome Biology,* 17.

MEEHAN, T. F., CONTE, N., WEST, D. B., JACOBSEN, J. O., MASON, J., WARREN, J., CHEN, C. K., TUDOSE, I., RELAC, M., MATTHEWS, P., KARP, N., SANTOS, L., FIEGEL, T., RING, N., WESTERBERG, H., GREENAWAY, S., SNEDDON, D., MORGAN, H., CODNER, G. F., STEWART, M. E., BROWN, J., HORNER, N., HAENDEL, M., WASHINGTON, N., MUNGALL, C. J., REYNOLDS, C. L., GALLEGOS, J., GAILUS-DURNER, V., SORG, T., PAVLOVIC, G., BOWER, L. R., MOORE, M., MORSE, I., GAO, X., TOCCHINI-VALENTINI, G. P., OBATA, Y., CHO, S. Y., SEONG, J. K., SEAVITT, J., BEAUDET, A. L., DICKINSON, M. E., HERAULT, Y., WURST, W., DE ANGELIS, M. H., LLOYD, K. C. K., FLENNIKEN, A. M., NUTTER, L. M. J., NEWBIGGING, S., MCKERLIE, C., JUSTICE, M. J., MURRAY, S. A., SVENSON, K. L., BRAUN, R. E., WHITE, J. K., BRADLEY, A., FLICEK, P., WELLS, S., SKARNES, W. C., ADAMS, D. J., PARKINSON, H., MALLON, A. M., BROWN, S. D. M., SMEDLEY, D. & INT MOUSE PHENOTYPING, C. 2017. Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nature Genetics,* 49**,** 1231-1238.

NAGASAKI, Y., MATSUBARA, Y., TAKANO, H., FUJII, K., SENOO, M., AKANUMA, J., TAKAHASHI, K., KURE, S., HARA, M., KANEGAE, Y., SAITO, I. & NARISAWA, K. 1999. Reversal of hypopigmentation in phenylketonuria mice by adenovirus-mediated gene transfer. *Pediatric Research,* 45**,** 465-473.

NEI, M., NIIMURA, Y. & NOZAWA, M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature Reviews Genetics,* 9**,** 951-963.

OHAZAMA, A., HAYCRAFT, C. J., SEPPALA, M., BLACKBURN, J., GHAFOOR, S., COBOURNE, M., MARTINELLI, D. C., FAN, C. M., PETERKOVA, R., LESOT, H., YODER, B. K. & SHARPE, P. T. 2009. Primary cilia regulate Shh activity in the control of molar tooth number. *Development,* 136**,** 897-903.

RAPOSO, G. & MARKS, M. S. 2007. Melanosomes-dark organelles enlighten endosomal membrane transport. *Nature Reviews Molecular Cell Biology,* 8**,** 786-797.

REIMAND, J., KULL, M., PETERSON, H., HANSEN, J. & VILO, J. 2007. g : Profiler - a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research,* 35**,** W193-W200.

ROBINSON, J. A., ORTEGA-DEL VECCHYO, D., FAN, Z. X., KIM, B. Y., VONHOLDT, B. M., MARSDEN, C. D., LOHMUELLER, K. E. & WAYNE, R. K. 2016. Genomic Flatlining in the Endangered Island Fox. *Current Biology,* 26**,** 1183-1189.

SCHERZ, P. J., MCGLINN, E., NISSIM, S. & TABIN, C. J. 2007. Extended exposure to Sonic hedgehog is required for patterning the posterior digits of the vertebrate limb. *Developmental Biology,* 308**,** 343-354.

SHUMBA, T., MONTGOMERY, R. A., SILLERO-ZUBIRI, C. & RASMUSSEN, G. S. A. 2017. Morphological variation of wild dogs across Africa. *International Journal of Zoology and Applied Biosciences,* 2**,** 145-154.

SINGHAL, S., GRUNDLER, M., COLLI, G. & RABOSKY, D. L. 2017. Squamate Conserved Loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources,* 17**,** e12-e24.

SKOGLUND, P., ERSMARK, E., PALKOPOULOU, E. & DALEN, L. 2015. Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. *Current Biology,* 25**,** 1515-1519.

SLATER, G. J., HARMON, L. J. & ALFARO, M. E. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution,* 66**,** 3931-3944.

STAHLER, D. R., MACNULTY, D. R., WAYNE, R. K., VONHOLDT, B. & SMITH, D. W. 2013. The adaptive value of morphological, behavioural and life-history traits in reproductive female wolves. *Journal of Animal Ecology,* 82**,** 222-234.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-1313.

STEINER, C. C., WEBER, J. N. & HOEKSTRA, H. E. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. *Plos Biology,* 5**,** 1880-1889.

STOREY, J., BASS, A. J., DABNEY, A. & ROBINSON, D. 2017. qvalue: Q-value estimation for false discovery rate control version 2.10.0. *http://github.com/jdstorey/qvalue*.

STURM, R. A., TEASDALE, R. D. & BOX, N. F. 2001. Human pigmentation genes: identification, structure and consequences of polymorphic variation. *Gene,* 277**,** 49-62.

SWANN, R. L. 1904. IV. Cursorial adaptations. *In:* NATURALIST, T. A. (ed.) *Adaptations to aquatic, arboreal, fossorial and cursorial habits in mammals. - IV Cursorial adaptations.*

TEDFORD, R. H., WANG, X. & TAYLOR, B. E. 2009. Phylogenetic systematics of the North American fossil Caninae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History,* 325**,** 1-218.

THIVICHON-PRINCE, B., COUBLE, M. L., GIAMARCHI, A., DELMAS, P., FRANCO, B., ROMIO, L., STRUYS, T., LAMBRICHTS, I., RESSNIKOFF, D., MAGLOIRE, H. & BLEICHER, F. 2009. Primary Cilia of Odontoblasts: Possible Role in Molar Morphogenesis. *Journal of Dental Research,* 88**,** 910-915.

THONY, B., AUERBACH, G. & BLAU, N. 2000. Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochemical Journal,* 347**,** 1-16.

TURNER, A. 1990. The evolution of the guild of larger terrestrial carnivores during the Plio-Pleistocene in Africa. *Geobios,* 23**,** 349-368.

VAN VALKENBURGH, B. 1987. Skeletal indicators of locomotor behavior in living and extinct carnivores. 7**,** 162-182.

VAN VALKENBURGH, B. 1991. Iterative evolution of hypercarnivory in canids (mammalia, carnivora) - evolutionary interactions among sympatric predators. *Paleobiology,* 17**,** 340-362.

VAN VALKENBURGH, B. 1996. Feeding behavior in free-ranging, large African carnivores. *Journal of Mammalogy,* 77**,** 240-254.

VAN VALKENBURGH, B. & KOEPFLI, K. P. 1993. Cranial and dental adaptations to predation in canids. *Mammals as Predators,* 65**,** 15-37.

VELERI, S., MANJUNATH, S. H., FARISS, R. N., MAY-SIMERA, H., BROOKS, M., FOSKETT, T. A., GAO, C., LONGO, T. A., LIU, P. H., NAGASHIMA, K., RACHEL, R. A., LI, T. S., DONG, L. J. & SWAROOP, A. 2014. Ciliopathy-associated gene Cc2d2a promotes assembly of subdistal appendages on the mother centriole during cilia biogenesis. *Nature Communications,* 5.

VENKAT, A., HAHN, M. W. & THORNTON, J. W. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nature Ecology & Evolution,* 2**,** 1280-1288.

VENKATESH, D. 2017. Primary cilia. *J Oral Maxillofac Pathol,* 21**,** 8-10.

VONHOLDT, B. M., CAHILL, J. A., FAN, Z. X., GRONAU, I., ROBINSON, J., POLLINGER, J. P., SHAPIRO, B., WALL, J. & WAYNE, R. K. 2016. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances,* 2.

WALSBERG, G. E. 1983. Coat color and solar heat gain in animals. *Bioscience,* 33**,** 88-91.

WEBB, A. E., WALSH, T. A. & O'CONNELL, M. J. 2017. VESPA: Very large-scale Evolutionary and Selective Pressure Analyses. *PeerJ Computer Science,* 3**,** 1-16.

WEBB, M. L., ROSEN, H., TAGHINIA, A., MCCARTY, E. R., CERRATO, F., UPTON, J. & LABOW, B. I. 2011. Incidence of Fanconi Anemia in Children with Congenital Thumb Anomalies Referred for Diepoxybutane Testing. *Journal of Hand Surgery-American Volume,* 36A**,** 1052-1057.

WU, X. F., BOWERS, B., RAO, K., WEI, Q. & HAMMER, J. A. 1998. Visualization of melanosome dynamics within wild-type and dilute melanocytes suggests a paradigm for myosin V function in vivo. *Journal of Cell Biology,* 143**,** 1899-1918.

YANG, Z. H. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution,* 24**,** 1586-1591.

YANG, Z. H. & DOS REIS, M. 2011. Statistical Properties of the Branch-Site Test of Positive Selection. *Molecular Biology and Evolution,* 28**,** 1217-1228.

YAO, L. L., SHEN, M., LU, Z. K., IKEBE, M. & LI, X. D. 2016. Identification of the Isoform-specific Interactions between the Tail and the Head of Class V Myosin. *Journal of Biological Chemistry,* 291**,** 8241-8250.

ZAKERI, Z., QUAGLINO, D. & AHUJA, H. S. 1994. Apoptotic cell-death in the mouse limb and its suppression in the hammertoe mutant. *Developmental Biology,* 165**,** 294-297.

ZERBINO, D. R., ACHUTHAN, P., AKANNI, W., AMODE, M. R., BARRELL, D., BHAI, J., BILLIS, K., CUMMINS, C., GALL, A., GIRON, C. G., GIL, L., GORDON, L., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., TO, J. K., LAIRD, M. R., LAVIDAS, I., LIU, Z. C., LOVELAND, J. E., MAUREL, T., MCLAREN, W., MOORE, B., MUDGE, J., MURPHY, D. N., NEWMAN, V., NUHN, M., OGEH, D., ONG, C. K., PARKER, A., PATRICIO, M., RIAT, H. S., SCHUILENBURG, H., SHEPPARD, D., SPARROW, H., TAYLOR, K., THORMANN, A., VULLO, A., WALTS, B., ZADISSA, A., FRANKISH, A., HUNT, S. E., KOSTADIMA, M., LANGRIDGE, N., MARTIN, F. J., MUFFATO, M., PERRY, E., RUFFIER, M., STAINES, D. M., TREVANION, S. J., AKEN, B. L., CUNNINGHAM, F., YATES, A. & FLICEK, P. 2018. Ensembl 2018. *Nucleic Acids Research,* 46**,** D754-D761.

ZHAI, W. W., NIELSEN, R. & SLATKIN, M. 2009. An Investigation of the Statistical Power of Neutrality Tests Based on Comparative and Population Genetic Data. *Molecular Biology and Evolution,* 26**,** 273-283.

ZHANG, C., RABIEE, M., SAYYARI, E. & MIRARAB, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *Bmc Bioinformatics,* 19.

ZHANG, J. Z., NIELSEN, R. & YANG, Z. H. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution,* 22**,** 2472-2479.

Chapter II, Splendid diversification: dissecting the evolution of South American canids with whole genome sequences

Abstract

South American canids represent the most remarkable recent radiation within the dog family (Canidae), comprising 10 species distributed across a wide diversity of habitats and including disparate forms such as the short-legged, hypercarnivorous bush dog and the long-legged, largely-vegetarian maned wolf. Despite the considerable research on these canids, many aspects of their evolutionary history remain uncertain, and the genetic basis of their unique adaptations is unknown. Moreover, human-induced population declines in some species may have led to reduced genetic diversity and adaptive potential, highlighting the need for baseline genomic information to inform conservation planning. We analyzed a total of 31 whole genomes encompassing all extant South American canid species and other related canids to investigate their evolutionary history, demography, current patterns of genetic diversity, and the molecular basis of their adaptations. We found that South American canids are derived from a single ancestral population that likely colonized South America ~3 million years ago. This ancestral lineage first diversified in the eastern region of South America followed by a single colonization event and diversification west of the Andes. We detected extensive historical gene flow among the earliest South American lineages, which could have enhanced adaptation. Our genome-wide scans of selection show that disparate limb proportions in the bush dog and maned wolf may be due to mutations in genes regulating chondrocyte proliferation and enlargement. Further, the maned wolf's ability to eat fruit may be enhanced by genetic variants in genes associated with energy intake from short fatty acids. Similarly, unique genetic variants in the bush dog were consistent with both the development of the interdigital web and dental adaptations for

hypercarnivory. Finally, we found a complex history of adaptive diversification throughout past climatic cycles in South America, compounded by recent population declines caused by humans. Our analysis demonstrates the remarkable potential of genomic analysis to provide new insights into the recent colonization by canids of the South American continent, their species-specific demographic history, degree of admixture and the development of unusual adaptations.

Introduction

In South America, the rapid diversification of canids resulted in ten species ranging in size and conformation from the squat bush dog to the long-legged maned wolf (Figure 2-1). Although this group represents the most rapid recent diversification of canids, its evolutionary history remains uncertain. There are three main aspects of South American (hereafter SA) canid evolution that have yet to be resolved. The first concerns the invasion and dispersion of canids into South America. Although it is widely agreed that the first canids migrated into South America after the formation of the isthmus of Panama ~3 million years ago, there is disagreement about the number of ancestral lineages that entered South America, and therefore, the antiquity of their diverse adaptations (Langguth, 1975, Berta, 1987, Wayne et al., 1997, Webb, 2006, Perini et al., 2010, Eizirik, 2012, Tchaicka et al., 2016, Prevosti and Forasiepi, 2018). When canids arrived in South America, the Andes had already attained their present-day elevation (Mora et al., 2010) and formed a distinct geographic barrier that extended along the entire SA continent serves (Eizirik et al., 1998, Webb, 2006, Koepfli et al., 2007, Favarini, 2011, Patterson et al., 2012, Helgen et al., 2013, Tchaicka et al., 2016). Although canids are found on both sides of the Andes, it is unknown whether the ancestral canid lineage entered the east or west side of the Andes or both sides and subsequently diversified. If the latter is the case, it

establishes the mountain chain as an enabler of diversification and the independent history of canids on either side. Also, it is unclear whether canids migrated across the Andes at one or several points, or if canids circumvented the Andes in the southern range of the continent (Perini et al., 2010, Patterson et al., 2012). Knowing the number of invasions and dispersion of canids in South America is critical to understanding the constraints and mechanisms underlying this burst of speciation.

The second outstanding question pertains to the influence of the Andes on canid demographic history. Climatic changes were distinct on each side of the Andes. In the eastern lowlands, the glacial periods of the Pleistocene led to the expansion of savanna vegetation into areas that now support forest vegetation (Haffer, 1969, Webb, 1991, Haberle and Maslin, 1999, Beerling and Mayle, 2006, Haggi et al., 2017). To the west of the Andes, glacial periods led to the expansion of glaciers in the south and shifts of vegetation zones along the Andean slopes (Clapperton, 1993, Heine, 2000, Kaiser et al., 2007). On balance, these changes may have reduced the favorable habitat of some species but extended the habitat of others (Haffer, 1969, Cossios et al., 2009, Lima-Rezende et al., 2019). Currently, the Andes restrict the geographic range of western species, leaving only a narrow belt between the Andes and the southern Pacific Ocean (Eva et al., 2004). This restricted range affects western species sensitive to direct human persecution and the introduction of domestic species (Jiménez and McMahon, 2015, Miranda et al., 2015, Yahnke et al., 1996). Particularly concerning is the reduction of temperate forest habitat and the introduction of domestic dogs in the Valdivia region of southern Chile, which may have led to the near elimination of Darwin's fox on the mainland, with only 78 individuals estimated as remaining (Jiménez and McMahon, 2015). Due to this recent history of sharp

population decline, the genetic diversity of this remnant population is a pressing conservation concern.

The third poorly resolved question concerns the genetic basis of the highly unique adaptations in SA canids. Two sister species, *Speothos venaticus* (bush dog) and *Chrysocyon brachyurus* (maned wolf), have extraordinary morphological differences that have been subject to extensive evolutionary speculation (Hildebrand, 1952, Walker and Paradiso, 1975, Berta, 1984, Dietz, 1984, Dietz, 1985, Wayne, 1986a, Wayne, 1986b, Van Valkenburgh, 1991). The bush dog is the only obligate meat-eating (hypercarnivorous) canid in the New World that survived the Late Pleistocene extinction of large hypercarnivores, potentially due to its ability to hunt mid-size prey in rain forest habitats (Deutsch, 1983, Berta, 1987, Van Valkenburgh, 1991, Lima et al., 2009). This species possesses a suite of dental and skeletal specializations to process and capture prey in forests, savanna, and wetlands, including short robust legs with webbed feet, long bodies with short tails, a unique extension of the meat cutting blade in the upper P4 and lower M1 teeth (a trenchant heel), and loss of molars which in other canids functions to crush hard plant foods (Van Valkenburgh, 1991). In contrast, the maned wolf is the only large-bodied canid in South America that survived the late Pleistocene extinctions, which may reflect its ability to exploit a wide range of food resources in the savanna-like cerrado environment. The maned wolf is a fruit specialist with as much as 90 percent of its diet being composed of a tomato-like fruit commonly known as “wolf apple” (*Solanum lycocarpum*) (Dietz, 1984, Dietz, 1985, Aragona and Setz, 2001, Bueno and Motta, 2004). The maned wolf also possesses the longest limbs among canids (Figure 2-1). However, it is not a swift runner (Hildebrand, 1952, Dietz, 1984), with its running speed limited by a unique racking gait (lifting both feet on each

side of the body simultaneously) that facilitates movement through tall grassland habitats (Walker and Paradiso, 1975, Dietz, 1985, Harris, 1993). Although these two species have been extensively studied (Hildebrand, 1952, Walker and Paradiso, 1975, Dietz, 1984, Berta, 1984, Dietz, 1985, Wayne, 1986b, Wayne, 1986a, Van Valkenburgh, 1991), the genomic underpinnings of these extraordinary adaptations remain a mystery.

Here, we investigate these three questions concerning SA canid evolution, addressing their phylogeny, recent and historic admixture, demographic history, and patterns of genomic diversity. In addition, we explore the genetic basis of the unique morphological characteristics of the bush dog and maned wolf. To accomplish these goals, we have sequenced 16 new genomes and in total analyzed 31 genomes belonging to 23 different species including the domestic dog (Table 2-S1). Finally, to investigate adaptive evolution and genome variation in the maned wolf and bush dog, we have generated *de novo* genomic assemblies for both species and additionally have sequenced seven wild-caught individuals (four maned wolves and three bush dogs, respectively) to characterize their patterns of genetic diversity in the wild.

Results

Phylogenomics

To provide an accurate evolutionary framework for the comparative genomics analysis of SA canids, we reconstructed a consensus phylogenetic tree of all 10 extant species of SA canids and 12 other species from the genera *Canis, Lupallela, Lycaon,* and *Cuon* (Table 2-S1). We first mapped reads from 30 canid genomes to the domestic dog reference assembly (CanFam3.1). We then extracted 6,716 alignments of 25 kb windows from these mapped genomes. We used

maximum likelihood to construct independent phylogenetic trees from each of these windows. To account for phylogenetic discordance among these trees, we generated a consensus phylogenetic tree with ASTRAL-III (Zhang et al., 2018; Figure 2-1 and 2-S1). Furthermore, we estimated the age of divergences among SA canids using our consensus tree and the MCMC function implemented in PAML4 (Yang, 2007). Our species tree shows that SA canids are monophyletic, with the earliest branching lineage of the bush dog + maned wolf splitting around 3.9 million years ago (mya; 95% HPD = 3.30 – 4.50 mya), followed by a clade containing the crab-eating fox (*Cerdocyon thous*) and enigmatic short-eared dog (*Atelocynus microtis*) that split 2.66 mya (95% HPD = 2.36 – 2.96 mya), and lastly, a clade composed of species of the genus *Lycalopex*. Species within the latter group, which were distributed east of the Andes, started to diversify 2.17 mya (95% HPD = 1.95 - 2.40 mya), whereas those west of the Andes form a monophyletic group that diverged at 1.87 mya (95% HPD = 1.67 - 2.07 mya). These species include the Sechuran fox (*Lycalopex sechurae*), sister to Culpeo fox (*Lycalopex culpaeus*) and Darwin's fox (*Lycalopex fulvipes*), basal to this monophyletic clade (Figure 2-1).

Biogeographic history

Given the potential influence of the Andes on the diversification of SA canids, we reconstructed the ancestral distribution of extant species of SA canids using the R package BioGeoBEARS (Matzke, 2013). We used this tool to estimate the probability that an ancestral lineage was distributed across the west, center, or east of the Andes (see methods for details). Among six different models tested, the DIVALIKE with +J parameter that allows for a jump dispersal event best fit our data (corrected AIC = 81.23%; Figure 2-1 and S2; Table 2-S2). According to this model, there is a 95% probability that lineages splitting earlier than 2.66 mya

originated in the east of the Andes (Figure 2-1). This group includes the bush dog, maned wolf, crab-eating fox, and short-eared dog. Likewise, there is a 67% probability that the hoary fox also originated in the east of the Andes around 2.17 mya (95% HPD = 1.95 - 2.40 mya; see node 21 in Figure 2-1). Thereafter, our model suggests with 78% probability that an ancestral SA fox had a range expansion from east to the west of the Andes around 2 mya (95% HPD = 1.79 – 2.21 mya; see node 20 in Figure 2-1). Following this migration, two ancestral SA fox lineages arose on opposite sides of the Andes. One arose west of the Andes and included the ancestor of Darwin's fox, Sechuran fox, and culpeo fox ~1.87 mya (95% HPD = 1.67 - 2.07 mya) with 100% probability (see node 19 in Figure 2-1). The other lineage arose in the eastern region of the Andes ~1.12 mya (95% HPD = 0.83 – 1.251 mya) with a probability of 85% (see node label 17 in Figure 2-1), and included the ancestor of pampas and SA gray fox.

Interspecific gene flow

We explored the degree of interspecific gene flow among SA canid species using G-PhoCS v1.3.2 (https://github.com/gphocs-dev/G-PhoCS) assuming the topology in Figure 2-1 and a *priori* predicted a higher prevalence of gene flow among more recently diversified species (Figure 2-2a and Table 2-S3). Our results show a minor amount of gene flow (≤ 5.2%) among lineages including the bush dog, maned wolf, crab-eating fox, and short-eared dog that diversified > 2.3 mya (Figure 2-2a and Table 2-S3). Conversely, there is a strong signal of gene flow among species of the genus *Lycalopex* that split more recently, <1.9 mya (Figure 2-2a, b and Table 2-S3). Importantly, there is considerable gene flow (>50%) from pampas fox to SA gray fox, Darwin's fox, and hoary fox (Figure 2-2b). Similarly, there is relatively high gene flow (>10%) from culpeo fox to Darwin's fox (18%) and SA gray fox. Other episodes of admixture

include gene flow from Darwin's fox to SA gray fox (11%) (Figure 2-2b). Overall, our results show a relatively small degree of admixture among the bush dog, maned wolf, crab-eating fox, and short-eared dog (Figure 2-2b and Table 2-S3), indicating that the speciation episodes involving these species are more complete. In contrast, speciation in *Lycalopex* is more recent, with high rates of interspecific historical gene flow that have caused non-independence of these lineages over the course of their radiation (Figure 2-2a and b).

To investigate the effect of admixture on phylogenetic discordance in SA canids, we extracted the quartet frequencies for a subset of internal nodes from the tree inferred by ASTRAL-III (Figure 2-2c). Among different nodes analyzed, two showed unresolved relationships for the pampas fox, SA gray fox, and Darwin's fox (see nodes N20 and N17 in Figure 2-2c). Interestingly, these are the species that showed the highest amount of gene flow (Figure 2-2b). This suggests that the long history of phylogenetic discordance among the youngest clade of SA foxes is likely due to interspecific gene flow (Yahnke et al., 1996, Wayne et al., 1997, Vila et al., 2004, Lindblad-Toh et al., 2005, Perini et al., 2010, Prevosti, 2010, Favarini, 2011, Slater et al., 2009, Tchaicka et al., 2016).

Genetic diversity and ROH

To understand the influence of species demographic history on patterns of genetic variation, we examined per-site heterozygosity in non-overlapping 100 kb windows across the genome of every analyzed individual (Figure 2-3 a-c). We observed three patterns of genome-wide heterozygosity among SA canids. First, five generalist species (crab-eating fox, hoary fox, pampas fox, SA gray fox and to a lesser degree, culpeo fox) had genomes with high

heterozygosity across most autosomes (Figure 2-3a). Second, Darwin's fox had regions of high heterozygosity alternating with stretches of homozygosity that span nearly entire chromosomes (Figure 2-3b). Lastly, four species (short-eared dog, bush dog, maned wolf, and Sechuran fox) had genomes with uniformly low heterozygosity across most autosomes (Figure 2-3c). High levels of heterozygosity in the pampas fox, hoary fox, crab-eating fox, and SA gray fox genomes suggest that these species have sustained large population sizes throughout their recent history (Figure 2-3a). Conversely, the low levels of genome-wide heterozygosity of the short-eared dog, bush dog, maned wolf, and Sechuran fox suggest that persistent small population sizes have caused a historical decline in diversity across all autosomes (Figure 2-3c). The Darwin's fox results suggest that recent population declines and inbreeding (Yahnke et al., 1996, Vila et al., 2004, Farias et al., 2014) may have contributed to the depletion of genetic diversity across long stretches of the genomes that span as much as several megabases in length (Figure 2-3b; see below). Among all the canids in our analysis, the short-eared dog, bush dog, Darwin's fox, and Sechuran fox had the lowest genome-wide heterozygosity (< 0.5 heterozygote sites/kb; Figure 2-3d and 2-4; Table 2-S4), a value that is similar than those observed in endangered canids such as the Ethiopian wolf, African wild dog and dhole (< 0.5 heterozygote sites/kb; Figure 2-4).

Given the observation that several regions of the genomes exhibited depleted genetic variation (Figure 2-3b and 2-c), we quantified the extent of runs of homozygosity (ROH) (Figure 2-3d and 2-3e; Table 2-S4). Long ROH indicate recent population bottlenecks and inbreeding. In contrast, short ROH could indicate ancient bottlenecks as there have been more opportunities for recombination to dissolve long haplotypes. We observed that the individual from Nahuelbuta National Park, Chile, had more sequences in long ROH (> 10 Mb) than any other wild canid

(Figure 2-3d and 2-3e). The total length of these long ROH fragments was 131 Mb and represented 5 % of the genome, as opposed to an average of 18 Mb (<1% of the total genome) in the other nine species of SA canids (excluding the captive bush dog). Likewise, a Darwin's fox from Chiloé Island, located at the west coast of south Chile (Figure 2-3d and 2-3e; Table 2-S4), had high ROH similar to that of a captive bush dog (Figure 2-3d and 2-3e) which was born in captivity and known to be highly inbred. The total length of medium ROH was highest in the Darwin's fox from Chiloé Island and was four times higher than for any other canid, accounting for 935Mb or 37% of the genome (Figure 2-3d and 2-3e; Table 2-S4;). These results are consistent with the relatively low historic population size of 5,300 (95% BI = 5, 100 – 5, 500) estimated by our G-PhoCS model (Figure 2-2a and Table 2-S3) as well as low census estimates for the Darwin's fox (Silva-Rodríguez et al., 2016).

Other species with moderate to low levels of heterozygosity (>1 heterozygotes/kb) including the bush dog, short-eared dog, maned wolf, and Sechuran fox had a dominance of small ROH (Figure 2-3d and 2-3e). On average, small ROH in these canids totaled 512 Mb (Table 2-S4), representing ~20% of the genome. Also, these species were characterized by medium (1-10 Mb) and few long (>10 Mb) ROH that represent 3.4% and 0.96% of the genome, respectively. According to our G-PhoCS model, the population size was relatively small in the short-eared dog, bush dog, maned wolf and Sechuran fox, ranging from 6,500 (95% BI = 6, 200 - 6, 800) in the short-eared dog to 12,000 (95% confidence interval = 11,600 – 12,400) in the maned wolf (Figure 2-2a and Table 2-S3). Species with high levels of heterozygosity (>1 heterozygotes/kb) such as the crab-eating, hoary, pampas, and culpeo foxes (Figure 2-3d and 2-3e; Table 2-S4) had few ROH. In these species, the majority of the ROH were <1 Mb and

represented only 6% of the genome and a total of 106 Mb average length. Also, no long ROH were observed in these species (Figure 2-3d and 2-3e). This is consistent with high population numbers depicted by G-PhoCS, ranging from 13,800 (95% BI = 13,300 – 14,400) in the culpeo fox to 118,400 (95% BI = 114,200 – 122,600) in the pampas fox (Figure 2-2a and Table 2-S3).

Demographic analysis

Given the different population histories suggested by genome diversity and G-PhoCS results, we investigated the historical demography of SA canids using the multiple sequentially Markovian coalescent (MSMC) approach (Schiffels and Durbin, 2014). We scaled the Instantaneous Inverse Coalescence Rate (IICR) from the MSMC model by 2μ for use as a proxy of the effective population size (Ne). We used the IICR with different mutation rates and generation times to estimate plausible time ranges (in thousands of years) of the resulting MSMC trajectories (Figure 2-S3). Some species from the east of the Andes had relatively small Ne values (Figure 2-5a). These species included the maned wolf, which had a stable Ne trajectory around 5,000, as well as the bush dog and short-eared dog, which both demonstrated a decline that began ~20 kya and reached a Ne as low as 2,000. The dip in estimated Ne of the short-eared dog ~10 kya (Figure 2-5a) suggested a sharp recent bottleneck, though this trajectory could also be associated with historic population structure (Mazet et al., 2016, Chikhi et al., 2018). The hoary, pampas, and crab-eating fox, and SA gray fox, all from the east of the Andes, had the highest Ne trajectories of around ~20,000. The latter species had a Ne decline after 20 kya (Figure 2-5a). The Darwin's fox from the west of the Andes had the lowest Ne (~1,000) after a persistent population decline over 20,000 years ago (Figure 2-5b). The remaining two species from the western region of South America, the Sechuran and culpeo foxes, had a long-term

stable Ne trajectory between 5,000 and 8,000 (Figure 2-5b). Overall, our results are consistent with estimates from the G-PhoCS model, showing the lowest Ne among western canids as well as for the bush dog, maned and short-eared dog (Figure 2-2a and Table 2-S3). These estimates ranged from 5,300 (95% BI = 5,100 – 5,500) in the Darwin's fox to 12,000 (95% BI = 11,600 – 12,400) in the maned wolf. Similarly, our results for eastern hoary, pampas, and crab-eating fox and SA gray fox were consistent with G-PhoCS, showing relatively high population size for these species (Figure 2-2a and Table 2-S3) with estimates ranging from 19,200 (95% confidence interval = 18500-19900) in the SA gray fox to 118,400 (95% BI = 114, 200 – 122, 600) in the pampas fox.

Deleterious variation

To further investigate the potential consequences of past population declines, we analyzed the estimated effect on protein-coding variants. We annotated mutations as synonymous or nonsynonymous and further classified nonsynonymous variants as tolerated (no effect on the protein; SIFT score > 0.05) or deleterious (significant impact on protein; SIFT score < 0.05). We grouped synonymous and tolerated mutations as 'benign' and deleterious missense mutations, stop codons, and variants that interrupt splice sites as 'damaging'. The bush dog genomes had on average 227 damaging homozygote derived genotypes, ~31.36% higher than other SA canids that had on average 165 (Figure 2-6 and Table 2-S5). This difference was two times greater than those observed for homozygote benign mutations (Figure 2-6). The bush dog had on average 3,402 benign homozygote derived genotypes, only 12.5% more than the other species, which had on average 3,000 benign homozygote derived genotypes (Table 2-S5). These findings suggest that the accumulation of deleterious variants in the bush dog could be a

consequence of their long-standing small population size. Particularly, selection in small populations is inefficient, resulting in fixation of deleterious variants of moderately and slightly negative effect on fitness, consistent with population genetic theory as well as empirical findings (Marsden et al., 2016, Lohmueller et al., 2008, Robinson et al., 2016, Robinson et al., 2019).

Genetic basis of bush dog and maned wolf adaptations

Among SA canids, bush dogs and maned wolves are the most disparate with regards to diet, ecology, and morphology despite descending from a recent common ancestor ~3.6 mya (Figure 2-1). The most notable difference between the two species is their limb proportions. Changes in gene regulatory regions have been principally associated with limb diversification in mammals (Cretekos et al., 2008, Osterwalder et al., 2018). The two species also differ considerably in their diets, as bush dogs are hypercarnivorous whereas maned wolves are omnivorous, with a heavy reliance on vegetable matter. We used several distinct analyses to identify genes associated with the contrasting limb morphologies and diets of bush dogs and maned wolves. First, we analyzed a set of 17,185 genes to identify amino acid changes that could have undergone selection in the bush dog or maned. These genes were stringently filtered and visually inspected to eliminate false positives for positive selection using the branch-site likelihood ratio test (see methods for details). Briefly, with SWAMP, we masked any region with more than 10 amino acid changes in a 15-codon window and 3 amino acid changes in a 5-codon window. Then, we visually inspected the gene alignments that had $p < 0.05$ and searched for alignment errors. We identified genes with multiple nucleotide changes in a single codon that tended to increase the likelihood-ratio scores (Venkat et al., 2018). To avoid any possible bias

from mapping our samples to the dog reference genome, we confirmed every mutation in each *de novo* assembly of the bush dog and maned wolf.

Following correction for multiple hypothesis testing, we found no significant outlier genes (Table 2-S6 and Table 2-S7). Nevertheless, our results revealed that the Beta-1,4-Galactosyltransferase 7 gene (*B4GALT7*) in the short-legged bush dog was among the top genes showing evidence of positive selection ($p<0.01$). Interestingly, this gene is involved in the synthesis of sulfate proteoglycans (heparan and chondroitin), which are the active molecules responsible for limb elongation through transportation of Indian hedgehog during chondrocyte proliferation (Cortes et al., 2009, Lindahl et al., 2017). Autosomal recessive mutations in *B4GALT7* are directly tied to dwarfism in humans and Friesian horses (Cartault et al., 2015, Leegwater et al., 2016) as well as the short stature phenotype in spondylodysplastic Ehlers-Danlos syndrome (Ritelli et al., 2017).

Morphological and dietary differences between the maned wolf and bush dog may have been driven by many genes. Since these genes would collectively regulate a biological pathway, each could have a moderate to a weak signal of positive selection. Standard tests for gene ontology enrichment such as Gprofiler, which considers only data above an arbitrary significance cutoff, are likely to miss these weaker signals of selection (Daub et al., 2017). Therefore, to explore the possibility of polygenetic signals of selection on the branches leading to the bush dog and maned wolf, we used polysel (Daub et al., 2017). We tested biological pathways related to frugivory in the maned wolf and hypercarnivory in the bush dog, along with limb elongation in both species (Table 2-S8 and Table 2-S9). Our scan of polygenetic signals of selection showed one significant category ($q<0.05$), butanoate metabolism in the maned wolf, which is related to

energy intake from short-chain fatty acids derived from fruit fiber (Figure 2-7 and Table 2-S8). To test if butanoate metabolism is uniquely associated with amino acid changes of the maned wolf, we further evaluated GO categories associated with frugivory in other canid species including coyote, gray wolf, dhole, and African wild dog (Table 2-S10). As expected, butanoate metabolism was not significant in other canids, which further supports its association with the fruit-rich diet of the maned wolf.

Limb elongation in mammals has been found principally at regulatory regions (Cretekos et al., 2008, Osterwalder et al., 2018). Therefore, we examined promoter and enhancer regions conserved in the Canidae but with lineage-specific variation for the bush dog or maned wolf. We analyzed 1 kb windows flanking the start and end sites from 39,704 transcripts belonging to 32,704 genes from the dog genome (canFam3.1). Within each window, we calculated the number of private alleles for the bush dog and maned wolf. In parallel, we calculated the average number of segregating sites across the remaining species (excluding maned wolf and bush dog). This group includes species of the genus, *Lycalopex* (SA foxes), *Canis* (wolves, coyote, and golden jackal), *Cuon* (dhole), *Lycaon* (African wild dog), and *Lupalela* (jackals). With the scores obtained from each window, we calculated empirical *p*-values. In this test, a relatively low p-value indicates a potential promoter or enhancer with a unique feature in either the bush dog or maned wolf. In each window, we searched for the presence of transcription binding sites from the JASPAR database (http://jaspar.genereg.net/). We found 110 genes in proximity to these windows with p-values less than 0.01 (Figure 2-8). Of these, 12 genes were found with known direct or indirect associations with limb development (Figure 2-8; Table 2-S11 and S12). Among these, *IGF1*, *B3GALT5*, and *B4GAL7* are related with chondrocyte proliferation through the synthesis of sulfate proteoglycans and chondrocyte enlargement (Cortes et al., 2009, Cooper et

al., 2013, Lindahl et al., 2017), which are primary genes associated with limb diversification in mammals (Rolian, 2020). To test if bone elongation genes are uniquely associated with bush dog and maned wolf regulatory regions, we further evaluated the proportion of private alleles in the regulatory regions of other canid species including coyote, gray wolf, dhole, and African wild dog. As expected, there were relatively few genes related to bone elongation in other canids (see SI). To further investigate the possibility of coding mutations reinforcing unique variation at regulatory regions, we focused on amino acid changes on genes related with limb development in the bush dog and maned wolf. Among 1,384 genes tested in the bush dog with the branch-site model test in PAML, we observed that *B4GALT7* yielded the most significant results (p=0.005). However, this gene was not significant after correction for multiple hypothesis testing (q=44; Table 2-S13).

Lastly, we identified mutations that may have a considerable impact on protein function and consequently, could underlie morphological differences. Some of these mutations include indels, stop codons, and UTR regions that are ignored by the branch-site test. We used the variant effect predictor tool (VEP) and identified in-frame indels, frameshift variations, and stop codon gains that could have a significant effect on their associated protein. We focused on mutations with a high impact on the protein as they are most likely to have been subject to selection. Also, we chose mutations that were unique to either the bush dog or maned wolf, among 25 species of canids. We found a total of 656 genes in the bush dog with unique mutations causing significant impact on their encoded proteins, including a unique 2-nucleotide insertion in the 3' UTR region of the *MSX1* gene, which affects both tooth and digit development (Weatherbee et al., 2006, Al-Qattan, 2014). Also, we found SULF2, a gene responsible for

sulfation of the receptor heparan sulfate proteoglycan (Ratzka et al., 2008). This result is consistent with our finding above of genetic variants in regulatory regions associated with chondrocyte proliferation through the synthesis of chondroitin sulfate proteoglycans (Cortes et al., 2009, Lindahl et al., 2017). We identified 450 genetic variants unique to the maned wolf with a high impact on their associated protein. Among these, we observed a loss of stop codon in the alcohol dehydrogenase 4 gene (*ADH4*) unique to the maned wolf. This gene facilitates energy transfer in the form of NADH when converting alcohol to ketone. Mutations in this gene have been found among species that have a diet rich in fruits (Carrigan, 2017).

Discussion

The geography of speciation during a rapid radiation

In past studies, divergence time estimates have been used to determine the number of canid invasions of South America (Slater et al., 2009, Perini et al., 2010, Tchaicka et al., 2016). In our analysis, we incorporated population size reconstructions of ancestral lineages to further assess the colonization of canids to South America. Our estimates of ancestral population sizes based on G-PhoCS analyses showed that the ancestral lineage of all SA canids had a relatively small population size of 11,600 (95% BI = 10900-12400) between 2.9 - 2.6 mya (Figure 2-2a). This is consistent with a past migration through the developing Panamanian corridor of Central American ~3mya (Webb, 1978, Webb, 1991, Woodburne, 2010). This corridor was primarily savanna and was narrow, limiting population sizes and acting as a selective filter (Webb, 1978, Webb, 1991, Woodburne, 2010). On reaching South America, canids encountered a vast landscape with few obligate carnivores, and those that were present were functionally distinct, such as marsupials and terror birds (Phorusrhacidae) (Marshall, 1977, Webb, 1991, Prevosti and

Forasiepi, 2018). The habitat diversity and lack of competition in South America likely fostered population growth and speciation of canids (Berta, 1988, Perini et al., 2010, Zurano et al., 2017). At least 13 species diversified in South America, including four hypercarnivorous forms and the recently extinct Falklands wolf (Berta, 1987, Berta, 1988, Prevosti, 2010, Slater et al., 2009, Perini et al., 2010, Austin et al., 2013, Prevosti and Forasiepi, 2018). Our G-PhoCS results are consistent with an increase in diversification around 2.6 mya followed by a 5-fold increase in population size of the ancestor of SA foxes as well as the common ancestor of the bush dog and maned wolf (Figure 2-2a).

Our findings support a single invasion of canids into South America around 3 mya (Woodburne, 2010). This model is challenged by the presence of North American fossils related to the genus of the maned wolf (*Chrysocyon*) and the crab-eating fox (*Cerdocyon*) from the early Pliocene ~5 mya (Torres-Roldán and Ferrusquía-Villafranca, 1981, Berta, 1987, Tedford et al., 2009). This record predates the formation of the Panama land bridge (Bacon et al., 2015, O'Dea et al., 2016). However, these fossil remains are fragmentary, and correct phylogenetic assignment has proven to be difficult. The *Cerdocyon* North American fossil has been related to the Asian raccoon dog, *Nyctereutes procyonoides* (Berta, 1984, Berta, 1987, Tedford et al., 2009), although this species has never been recorded in the Americas (Tedford and Qiu, 1991, Tedford et al., 2009, Wang et al., 2008). The remaining North American fossils, belonging to *Chrysocyon nearticus*, have been related to the maned wolf based on dental features. However, the dentition of these fossils is common to species that are distantly related to SA canids, such as *Canis* and *Vulpes* (Tedford et al., 2009). Despite the uncertainty about the identification of *Ch. nearticus,* these fossils possess the angular process that is characteristic of SA canids (Tedford et

al., 2009). The general dentition of this fossil suggests that it could be a basal lineage related to the ancestor of all SA canids, an inference that is compatible with the single-invasion hypothesis.

Following the colonization of South America via the Panamanian land bridge, our biogeographic model showed that canids first diversified east of the Andes (Figure 2-1), which suggests that they used a natural corridor known as the "savanna route" to disperse from Central America to eastern South America (Webb, 1978). This was the preferred route for most mammals, including felids, procyonids, large horses, and large proboscideans (Berta, 1988, Eizirik et al., 1998, Koepfli et al., 2007, Eizirik et al., 2010, Helgen et al., 2013), which suggests that the Andes may have been a barrier to western dispersal of most mammals during the early stages of the great American interchange. Consequently, canids diversified in a north-south pattern along the east of the Andes. These species include the bush dog, maned wolf, crab-eating fox, short-eared dog, and hoary fox (Figure 2-1). Our results indicate that the ancestral SA fox dispersed to the west of the Andes ~ 2 mya. The dispersal of this ancestral lineage probably occurred during the dry periods of the Pleistocene when savannas expanded in South America (Haffer, 1969, Webb, 1991, Haberle and Maslin, 1999, Beerling and Mayle, 2006). Given that the oldest fossil record of *Lycalopex* has been found in the austral region of the continent (Berta, 1981, Berta, 1987, Berta, 1988), it is possible that this ancestral SA fox circumvented the Andes in southern Patagonia to colonize the west side of the Andes. Multiple species then diversified from this ancestral lineage, including the Pampas fox and SA gray fox in the east and Darwin's fox, culpeo fox, and Sechuran fox to the west of the Andes (Figure 2-1). This burst of speciation is consistent with fossil records suggesting that the center of the *Lycalopex* radiation was

Argentina (Berta, 1987) and establishes the Andes as a critical constraint on early diversification of canids.

Admixture between species and the changing dynamics of habitats

Our biogeographic scenario predicts that lineages resulting from early divergences in eastern South America should have less interspecific admixture than recently diversified species on the west of the Andes. Our results from G-PhoCS partially support this prediction. Particularly, most species west of the Andes had more than 15% admixture (Figure 2-2a and b). In contrast, divergences older than ~2.66 mya east of the Andes have led to lineages with less than 2% of interspecific admixture (Figure 2-2a). These lineages included the maned wolf, bush dog, short-eared dog, and crab-eating fox. The exception to this low level of admixture in eastern lineages were species of the genus *Lycalopex*. Within this group, most of the detected gene flow derives from the eastern pampas fox to other lineages (Figure 2-2b). There was a 90% rate of gene flow from pampas to SA gray fox. These species are morphologically similar and have been suggested to be conspecific (Zunino et al., 1995, Prevosti et al., 2013). Also, their morphological similarities could be the result of extensive hybridization (Wayne et al., 1997, Tchaicka et al., 2016). Additionally, we found 51% of gene flow from pampas to hoary fox (Figure 2-2b), which suggest earlier hybridization between these species since our genomes derive from locations outside any known hybrid zone, and from individuals that have been considered ‘pure’ based on microsatellite markers (Table 2-S1) (Favarini, 2011, Garcez, 2015). The connection of formerly isolated species, during savannas expansion in the Pleistocene, may have provided the conditions for hybridization (Haffer, 1969, Webb, 1991, Haberle and Maslin, 1999, Beerling and Mayle, 2006). It is possible that the habitat of pampas fox extended as far as the west of the Andes

during dry periods, which would explain the observed admixture from pampas to the western Darwin's fox (Figure 2-2b). Overall, our results suggest that interspecific gene flow occurred throughout the radiation of *Lycalopex* foxes, probably augmenting their adaptive potential through the introduction of new alleles and thereby facilitating their rapid ecological diversification (Fuentes and Jaksic, 1979, Jimenez et al., 1995, Zunino et al., 1995, Prevosti et al., 2013, Bubadue et al., 2016, Zurano et al., 2017, Schiaffini et al., 2019), as has been shown in Darwin's finches and *Heliconius* butterflies (Dasmahapatra et al., 2012, Lamichhaney et al., 2015).

Demographic history of SA canids

Different climatic conditions on each side of the Andes may have shaped the demographic history of SA canids. The expansion of ice sheets was likely the most important influence on population size to the west region of the Andes, whereas the expansion of savannas was critical in the eastern region (Haffer, 1969, Webb, 1991, Haberle and Maslin, 1999, Beerling and Mayle, 2006, Haggi et al., 2017). Our demographic results provide new insights into how Pleistocene climatic cycles shaped canid demography. The MSMC model showed a decrease in population size of the short-eared dog and bush dog, both forest-associated eastern species, ~20 kya (Figure 2-5a). Concurrently, there was an increase in population size in species that prefer open habitats, such as the hoary fox, SA gray fox, pampas fox, and crab-eating fox. The population size increase of open-habitat species coupled with a decline of forest specialists could be related to habitat changes that occurred during the last glacial maximum. During this time, there was a 25% reduction of rainfall, which led to a 34-67% expansion of grasslands and a 14-25% reduction of forests (Haberle and Maslin, 1999, Beerling and Mayle, 2006). These habitat

dynamics suggests that the expansion of savannas may have permitted canids from open habitats to extend their ranges and consequently increase their population size. In contrast, the reduction of forests may have decreased the range of forest specialists and consequently led to a reduction of their population size.

For Darwin's, Sechuran, and culpeo foxes found west of the Andes, our MSMC models suggest relatively low population sizes (Figure 2-5b). The Andes restricted the geographic range of western species, leaving a narrow belt between the Andes and the southern Pacific Ocean. The lowlands west of the Andes are only 3% of the width of the eastern lowlands (Eva et al., 2004). The relatively small area of western lowlands could be related to species population size. The Sechuran fox has a minute distribution in northern Peru, which may have restricted its population growth (Cossíos, 2010). In contrast, the culpeo fox has the greatest population size and heterozygosity among the western species, and its geographic range is also the most extensive (Crespo, 1975, González del Solar and Rau, 2004). These results further establish the Andes as an important constraint on genomic variability and adaptive potential.

Our results show that Darwin's fox from Nahuelbuta National Park had low genome-wide heterozygosity, and this species was the only wild canid genome with a considerable proportion of long ROH (Figure 2-3b, 2-3d, and 2-3e), suggesting that this population may have experienced recent inbreeding as a consequence of their severe population decline. Our MSMC model showed a drastic population decline in Darwin's fox over the last 20 kya (Figure 2-5b). This type of trajectory reflects population bottlenecks, since the model tends to smooth them as persistent population declines (Beichman et al., 2017). Population bottlenecks in Darwin's fox could be associated with past climatic changes as well as recent human intervention. This species is distributed in the Valdivia forest, which extends from 35°S to 48°S latitude (Nahuelhual et al.,

2007). During the LGM, ice sheets expanded from the Andes into the pacific coast. The effect of this glacial expansion was to restrict the Valdivia forest's southern limit to 41° S (Kaiser et al., 2007). This habitat reduction may have led to a decrease in population size during the LGM (Villagran, 1988). On a more recent time scale, satellite images have shown that 33% of Darwin's fox native habitat has been lost due to deforestation (Otavo and Echeverria, 2017). This habitat loss has caused a population decline so severe that it almost eliminated Darwin's fox from the mainland (Yahnke et al., 1996, Farias et al., 2014). Current estimates suggest that only 78 individuals remain in two relic populations from Nahuelbuta National Park and the Valdivian costal range (Yahnke et al., 1996, Farias et al., 2014, Jiménez and McMahon, 2015, Silva-Rodríguez et al., 2016). These populations are isolated from each other, which raises concerns about further erosion of genetic diversity. In contrast to mainland populations, Darwin's fox from Chiloé Island showed higher heterozygosity and smaller blocks of ROH, consistent with the larger census size (~500 adults) of this population. Moreover, this population is considered less threatened compared to the mainland populations (Silva-Rodríguez et al., 2016). Nonetheless, our MSMC results suggest that the influence of past population bottlenecks on the genome from the Chiloé Island individual, possibly related to the founding colonization of the island, which likely occurred during the Last Glacial Maximum ~15 kya when the island was connected to the mainland via a land or ice bridge (Moreno et al., 1994). The invasion of Chiloe by land mammals has been confirmed by the presence of mastodons on the island (Moreno et al., 1994). Also, the capability of canids to invade nearby islands has been evident in other SA canids. For instance, the Falkland Islands wolf (*Dusicyon australis*) was able to colonize from the continent nearby islands in eastern South America (Austin et al., 2013). Although the introduction of Darwin foxes to Chiloe by humans may be a possibility, our MCMC analyses suggest the two

populations split ~35 kya (95% confidence interval = 33-48kya), before the earliest presence of humans in southern Chile between 18-14kya (Dillehay et al., 2015).

Genetic basis of bush dog and maned wolf adaptations

Although the bush dog and maned wolf have been subject to extensive morphological study, the genomic mechanisms underlying their unique adaptations have remained a mystery (Langguth, 1975, Dietz, 1984, Dietz, 1985, Wayne, 1986a, Wayne, 1986b, Berta, 1987, Van Valkenburgh, 1987, Van Valkenburgh, 1991, Sheldon, 1992). For a large canid, maned wolves have a novel set of food sources, with a diet consisting primarily of fruits, which presents a challenge to digesting polysaccharides. These complex sugars are indigestible for most carnivores (Clerici et al., 2011, Pereira et al., 2019). In most mammals, including canids, polysaccharides are broken down by bacterial fermentation in the cecum (den Besten et al., 2013, Rahim et al., 2019). Given that maned wolves lack a long intestine to make extensive fermentation possible (Sheldon, 1992), they likely have some physiological adaptations that allow them to process bacterial products more efficiently. Our findings of polygenic signals of selection related to energy intake from short-chain fatty acids provide new insights into the mechanism by which maned wolves digest complex sugars. Specifically, short-chain fatty acids (SCAFs) are processed in the cecum during gut bacterial fermentation for polysaccharides such as fruit fiber (den Besten et al., 2013). Then, unique mutations in the maned wolf may enhance the transformation of SCAFs into an energy precursor, Acetyl-CoA, through the butanoate metabolism process (Figure 2-7). Once Acetyl-CoA is produced, it releases energy in the form of ATP in the citric acid cycle (den Besten et al., 2013, Rahim et al., 2019).

The bush dog and maned wolf have the most divergent body size and limb proportions among canids, which may represent adaptations to flooded forests, and tall grasslands, respectively (Deutsch, 1983, Dietz, 1985). Newborns in both species already have the respective species' characteristic limb proportions (Wayne, 1986a, Wayne, 1986b). Consequently, since differences in limb length at birth are not explained by variation in gestation time, a molecular pathway that decreases the rate of fetal limb growth (while not restricting the growth of the axial skeleton) may exist, such as in various chondrodysplasias (Wayne, 1986a, Wayne, 1986b, Sutton et al., 2005). Our results of unique mutations in regulatory regions and amino acid changes related to chondrocyte proliferation and enlargement provide new insights into this long-standing question. Notably, bone elongation begins when chondrocytes multiply in the proliferation region (Figure 2-9d and 2-9j). Subsequently, these cells grow to become hypertrophic chondrocytes (Rolian, 2020). Finally, there are a series of steps for cell enlargement during which hypertrophic chondrocytes triple their size (Figure 2-9e and 2-9k; Cooper et al., 2013). The genomic windows enriched with lineage-specific alleles next to *B4GAL7* in the maned wolf and *B3GALT5* for the bush dog are associated with limb elongation through the synthesis of sulfate proteoglycans (heparan and chondroitin) (Lindahl et al., 2017). Additionally, the variant in *SULF2* may have a significant impact on the associated protein of bush dogs and is responsible for sulfation of heparan sulfate proteoglycans (Ratzka et al., 2008). Sulfate chondroitin proteoglycans (CSPG) and heparan sulfate proteoglycans (HSPG) transport Indian Hedgehog (Ihh) from the hypertrophic region to the top region of the proliferation zone (Figure 2-9b and 2-h; Cortes et al., 2009). In this region, *Ihh* activates target genes necessary for chondrocyte multiplication (Figure 2-9c and 2-9i). The rate at which new chondrocytes are generated is determined by the proportion of CSPG and HSPG (Cortes et al., 2009). Our findings

suggest that a relatively high rate of sulfation of chondroitin proteoglycans may promote longer limbs in the maned wolf (Figure 2-9a-f) whereas lower sulfation may result in short limbs in the bush dog (Figure 2-9g-l; Cortes et al., 2009). The other gene relevant to limb proportion is *IGF1*, which is significantly enriched with private alleles in the bush dog, and determines the increase of hypertrophic chondrocytes through fluid intake and mass growth (Figure 2-9e and 2-9k; Cooper et al., 2013). Previous studies have shown that *IGF1* is associated with a 40-fold increase of metatarsal cell volume in the jerboa as well as the enlargement of forelimbs on newborn opossums (Sears et al., 2012, Cooper et al., 2013). In dogs, a single mutation at *IGF1* has been associated with size differences among dog breeds (Sutter et al., 2007, Vaysse et al., 2011, Plassais et al., 2019). Given the characteristic short limbs of bush dogs, our results suggest that unique mutations may decrease the expression of *IGF1* and consequently limit chondrocyte elongation (Figure 2-9e). A depletion of IGF1 reduces the duration of the cell enlargement process, thus restricting limb elongation (Figure 2-9k and 2-9l; Cooper et al., 2013). Supporting this view, mice deficient in IGF1 failed to double the volume of their chondrocytes, resulting in short limbs (Wang et al., 1999).

The other unusual characteristics of the bush dog include an interdigital web on the foot that enhances its swimming abilities, as well as molar reduction and the development of a trenchant heel in the carnassial teeth (p^4/M_1) that improves meat cutting. The latter two features are associated with an emphasis on meat rather than plant material in the diet of carnivores (Van Valkenburgh, 1991). Our results from the variant effect predictor tool suggest a candidate gene, *MSX1*, that may be related to molar reduction and interdigital webs. We found an insertion of two bases at the 3' UTR region of the *MSX1* gene that, among canids, is unique to the bush dog and likely affects gene transcription. Previous studies have shown that mutations in this region

can affect the functionality of *MSX1* (Wong et al., 2014, Li et al., 2015), which in turn influences tooth development (Wong et al., 2014). In particular, a decrease in MSX1 reduces BMP4. The latter protein lowers the proliferation of mesenchymal cells during tooth formation (Jernvall and Thesleff, 2000). In the case of digit development, a decrease of *MSX1* expression would increase the effect of its antagonist *FGF8*. The protein expressed from the latter gene prompts the survival of mesodermal cells in the interdigital web (Weatherbee et al., 2006, Al-Qattan, 2014). These findings suggest that the observed insertion at the 3' UTR region may decrease the expression of *MSX1*, with potentially a dual phenotypic effect in the bush dog, depending on the tissue in which the gene is being expressed (loss of molars in tooth development, and interdigital webs formation). If confirmed by experimental data in the future, this dual effect would offer a striking pathway toward morphological innovation and macroevolution evolving genetic changes in a single gene.

Conclusions

Our analysis of whole-genome sequences of all extant species of SA canids supports the hypothesis of a single canid invasion into South America. This finding suggests that the adaptations we have genetically characterized in this study may have evolved exclusively in South America. This result provides new context to the rapid adaptive radiation of these species. For instance, the miniature stature of the bush dog represents an obvious disadvantage for hunting large prey and mitigating competition with large carnivores (Van Valkenburgh, 1991, Van Valkenburgh and Hertel, 1993). However, in South America, the absence of other carnivores and dense Amazon forest of the late Pleistocene may have facilitated the evolution of a diminutive hypercarnivore that can readily locomote in wet rainforests and take advantage of

small and mid-sized prey such as the paca and capybara through social hunting (Berta, 1978, Berta, 1981, Deutsch, 1983, Van Valkenburgh, 1991, Perini et al., 2010, Bubadue et al., 2016, Zurano et al., 2017, Prevosti and Forasiepi, 2018). Similarly, the long limbs of the maned wolf may facilitate locomotion and field of vision through the high grass pampas. Our results suggest that the dramatic evolution of these traits may derive from simple changes in regulation and composition of a few genes that influence limb development, interdigital webbing, and tooth morphology. We have shown that the Andes represented a geographical barrier that prevented canids from colonizing the western region of the Andes early in their evolutionary history. As a consequence of this barrier, canids dispersed southward along the east of the Andes, until reaching the lowest elevation of this mountain range in Patagonia ~2 mya. The expansion of savannas during glacial periods has affected the demographic history and admixture of each South America canid. Notably, there was considerable admixture in *Lycalopex*, suggesting that the rapid adaptation of SA canids to a variety of habitats may have been facilitated by introgression. Admixture also accounts for some of the phylogenetic discordance among pampas fox, SA gray fox, and Darwin's fox. Finally, the expansion of savannas during the Pleistocene cycles has greatly favored population growth of species that prefer open habitats. An important conservation implication of our results is that Darwin's foxes from the mainland are genetically depauperate and potentially suffering from high genetic load and inbreeding depression. Captive breeding and habitat restoration are urgently needed to allow mainland populations to expand to parts of their previous geographic range and retain genetic viability. Our analysis dramatically demonstrates the power of genomic analysis to reveal new evolutionary insights into the causes of speciation and adaptive radiation which have puzzled evolutionary biologists for decades.

Methods

Sequencing Mapping and Genotype calling

We extracted genomic DNA from 19 samples belonging to 10 different species of SA canids using the Qiagen DNEasy kit. Samples were chosen based on sufficient quality and quantity with a DNA fluorometer (Qubit 2.0), a NanoDrop spectrophotometer (ThermoFisher, USA), and gel electrophoresis. We conducted whole-genome sequencing on these extractions using Novaseq with 150 paired end reads from the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley. The samples (Cbr17082018 and Sve16082018) used for the *de novo* assembly in the maned wolf and bush dog were sequenced at the Smithsonian OCIO Data Science Lab (Table 2-S1). Genomes for the North American and Old World canids as well as the culpeo fox were downloaded from the National Center for Biotechnology Information (NCBI) (see Table 2-S1 for further information).

We filtered the raw reads using a modified pipeline from the Genome Analysis Toolkit best practice guide (GATK) (McKenna et al., 2010). Then, we mapped reads with good quality to the domestic dog canFam3.1, using BWA-MEM (Li, 2013). We used GATK HaplotypeCaller to conduct a combined genotype calling from sites that were mapped to the reference genome. We filtered the called genotypes for coverage and quality and kept only genotypes that had a minimum of 5 reads at a given position and a high-quality score (Phred scores >= 20), and no more than the 99th percentile of coverage for each sample. Other variant filtering criteria followed (GATK) Best Practices Guide and Robinson et al. (2019). Briefly, we filtered out CpG islands, indels, multi-nucleotide polymorphisms, and sites with more than one alternate allele.

The command line code used for reads mapping, variant calling, and filtering is available at https://github.com/dechavezv/2nd.paper.v2

To avoid any potential reference bias in our results of positive selection and deleterious variation, we confirm that the sequence of every candidate gene mapped *de novo* reference genomes of SA canids. We first create the *de novo* assemblies of the bush dog and maned wolf using MaSurcA (Zimin et al., 2013) with default parameters. We removed every scaffold less than 500 bp. We used the program assembly stats version 0.1.4 (Trizna, 2020) to obtain the counting and scaffold statistics of the generated *de novo* assemblies for the bush and maned wolf. We used these assemblies to ensure that our candidate mutations were not misalignments. Specifically, we BLASTed every candidate gene to these *de novo* assemblies.

Phylogenetic analysis

We reconstructed a species tree representing the relationships between 31 genomes using ASTRAL-III. First, we extracted 6,716 alignments, each 25 kb in length, from the entire set of 38 autosomal chromosomes. Each independent window was aligned with PRANK v.150803 using iteration (-F once option) and the topology shown in Figure 2-1 as a guide tree. After trimming each multiple species alignments with Gblocks (Castresana, 2000), we calculated the independent phylogeny using RAxML 8.0 (Stamatakis, 2014) under the GTR model. We selected the best phylogeny from each RAXML output. We then created a consensus file with these trees with no more than 10% of missing data. Branch of length < 1e-05 substitution per site were collapsed. Also, we collapsed clades with support lower than one using SqCL pipeline (phylogeny_prep_astrid_astral.py). We chose the best tree from the result depicted from

RAXML and merged the 100 bootstrap trees. We used the best tree and consensus file of bootstrap trees to investigate the discordance between gene trees and the species tree using ASTRAL-III v.5.5. This resulted in maximum likelihood support values that served to choose the best multi-locus tree. We ran 100 bootstrap replicates on this tree and further scored it to calculate the quartet support values and posterior probabilities for each node. We used the gray fox as an outgroup.

To calculated divergence times, we obtained the four-fold degenerate sites from 183 single-copy coding orthologues with a custom python script (https://github.com/mahajrod/MAVR/tree/master/scripts/multiple_alignment/extract_degenerate_sites_from_codon_alignment.py). We concatenated into a supermatrix 166, 182 four-fold degenerate sites across 30 genomes analyzed (including the domestic dog). After keeping a matrix with only 10 % of sites missing, we used the MCMCTree tool from PAML to calculate the divergence times. We used the topology depicted from ASTRAL-III (Figure 2-1). We used a HKY + G model and ran a total of 2,200,000 MCMC generations. We discarded the first 200,000 generations as burn-in. The fossil priors used to calibrate nodes of the phylogeny are shown in Table 2-S14.

BioGeoBEARS

We investigated the geographic origin of extant species of SA canids before possible colonization events across the mountain chain using the R package BioGeoBEARS (Matzke, 2013). This tool estimates the maximum-likelihood distribution of hypothetical ancestors (internal nodes) by modeling shifts between different geographical ranges along the phylogeny

as a function of time. First, we tested three different models: the dispersal-Extinction-Cladogenesis (DEC), dispersal vicariance analysis (DIVALIKE), and the Bayesian analysis of biogeography (BAYAREALIKE) models. Additionally, we test the same three models plus a founder effect parameter on each model named "J": DEC+J, DIVALIKE +J, and BAYAREALIKE+J (Matzke, 2013). The best fitting model was chosen based on the corrected the Akaike information criterion (AIC) scores (Figure 2-S2 and Table 2-S2).

Interspecific gene flow

We ran the multi-threaded version of G-PhoCS V1.3.2 (https://github.com/gphocs-dev/G-PhoCS) based on 11, 636 putatively neutral 1 kb windows following Chavez et al (2019). We used five threads per run and a standard MCMC setup. We assumed an exponential distribution with a mean of 0.0001 as the prior of all the mutation-scaled population sizes (θ) and divergence times (τ) as well as a Gamma (α= 0.002, β= 0.00001) distribution for the prior of migration rates (m). Because of the large number of migration bands, the Monte Carlo Markov chain was let to converge for 200,000 burn-in iterations, after which parameters were sampled every 50 iterations, for the next 400,000 iterations, resulting in a total of 8,000 samples from the approximate posterior distribution. For each parameter, we recorded the mean sampled value and the 95% Bayesian credible interval (CI). Population size estimates (Ne) were obtained from the mutation-scaled samples (θ) by assuming a mutation rate per generation of $\mu=4.0\times10^{-9}$ (Skoglund et al., 2015), and divergence times (*T*) were calibrated by assuming the same rate and an average generation time of three years. Migration rates were scaled by the duration of time of the migration band, resulting in total rates, which approximate the probability that a lineage experienced migration. Parameter estimates are summarized in Table 2-S3.

Genomic Diversity

We examined the site heterozygosity in non-overlapping 100 kb windows across the genome of every SA canids analyzed. We defined heterozygosity as the number of heterozygous genotypes divided by the total number of sites that were called. The total genotypes called within each window included the sum of heterozygous, homozygotes derived, and homozygotes reference genotypes. We kept only windows with no more than 20% of missing data. The script used to calculate 100 kb windows heterozygosity was modified from Robinson et al. (2016 and 2019) and is available at https://github.com/dechavezv/2nd.paper.v2/tree/main/4-Demography/Heterozygosity/WindowHet. We then quantified the extent of runs of homozygosity (ROH) in SA canids using PLINK (Purcell et al., 2007). The parameters chosen to calculate ROH were: SNPs within a window =200, heterozygotes allowed within a window= 3, missing sites within a window=50. We binned these segments into three different size categories using Plink (Purcell et al., 2007; Table S4 and Figure 2-3d). The categories were short ROH as short (1Mb<), medium ROH (>5Mb and <10Mb) long ROH (>10Mb).

MSMC

We used the program MSMC (Schiffels and Durbin, 2014) to calculate the instantaneous inverse coalescence rates (IICR). Then we scaled the IICR from the MSMC model by 2μ to use it as a proxy of the effective population size (Ne). IICR were further scaled to numerous of generation times and mutation rates to estimate plausible time ranges (in thousands of years) of the resulting MSMC trajectories (Figure 2-S3).

Deleterious variation

We assessed the effect of the joint variant calls on the associated proteins using the variant effect predictor tool (McLaren et al., 2016). First, we annotated mutations as synonymous or nonsynonymous. Then we used the SIFT scores to evaluate the effect of coding mutations on the associated protein. The SIFT score was calculated with SIFT (version 5.2.2) by comparing amino acid mutations annotated from the dog reference genome with a protein database from UniRef90 (release 2014_11). We grouped synonymous and tolerated mutations as 'benign' and classified loss-of-function mutations, deleterious missense mutations, and variants that interrupt splice sites as 'damaging' following Robinson et al (2019). Finally, we classified 'benign' and 'damaging' mutations as derived (not matching the reference allele) or reference (matching the reference genome) with respect to the dog genome.

Positive Selection

To obtain the orthologous genes necessary for the branch-site model in PAML (Yang, 2007, we followed Chavez et al (2019). Briefly, we used BioMart in Ensembl (Ensembl Genes 104 release) to get the coordinates of 19,000 genes. These regions were extracted from the annotation of the dog reference genome canFam3.1. To avoid the inclusion of paralogous genes, we used VESPA (Webb et al., 2017)to extract and concatenated different exons from the same transcript. Also, sequences were filtered for quality, coverage ($5< X <95$ percentile). We further confirmed that exons had a permissible length (exons whose length is an exact multiple of three) and no internal stop codon. We kept only the longest transcript for downstream analysis. Once sequences were translated to amino acid sequences we aligned them with prank using the topology obtained by Astral-III as a guide tree. The obtained multiple species alignments were

reverse-translated to nucleotide sequence using Vespa (Webb et al., 2017). The pipeline to extract species 1:1 orthologous genes can be found at https://github.com/dechavezv/2nd.paper.v2/tree/main/2-PositiveSelection

Each ortholog gene from a set of 17,185 genes was tested for signals of positive selection using the branch-site model in PAML 4.8 (Yang, 2007).. We compare a model that allows sites to be under positive selection (dN/dS > 1) along a particular branch in the tree (fix omega = 0) against a model where sites evolve under neutral or purifying selection (dN/dS = 1; fix omega = 1). The likelihood of each model was compared through likelihood ratio tests (LRTs). We determined statistical significance using a chi-square distribution with 1 degree of freedom freedom (Zhang et al., 2005). Two forward branches were tested for selection, the clade of the bush dog and another one for the maned wolf clade. We corrected for multiple hypothesis using a false discovery rate of 0.20 with QVALUE in R. Given the considerable number of genes tested that led to no significant genes (Q-value < 0.2), we further tested the overrepresentation of genes on biological pathways using G-profiler version r1732_e89_eg3629. We provide genes with (p-value < 0.01) as query lists while the full set of tended tested (N= 17,185) was specified as the background gene list. We allowed a minimum of two genes to overlap between the query list and the full set of genes.

Testing for polygenic selection

We used polysel (Daub et al., 2017) to detect biological pathways overrepresented by weak to moderate signals of selection on the bush dog and maned wolf. This tool uses the output from the branch-site model in the bush dog and maned wolf to find polygenic selection across

biological pathways. We extracted these pathways from NCBI (https://www.ncbi.nlm.nih.gov/biosystems/) using the option "pathway"[BioSystemType] and "Canis lupus familiaris"[Organism]. Based on the literature, we chose pathways that are relevant to the unique morphologic features of the bush dog and maned wolf. The characters were diet (carnivorous or frugivorous), limb development, tooth formation, and interdigital membrane development. Polysel uses two inputs. One is the set of biological pathways. The other is the genes set from PAML in the form of 'SUMSTAT' scores. To obtain these scores we took the fourth root of the log-likelihood ratios from PAML. To make the ID of genes match those in the pathway, we converted gene labels into Entrez gene IDs using gene2ensembl from NCBI (ftp://ftp.ncbi.nih.gov/gene/DATA/gene2ensembl.gz). Polysel uses the genes and pathways to reveal a null distribution. This null distribution was created by randomly sampling genes to make new pathways of a similar size. We obtained p-values by comparing the 'SUMSTAT' score between the null distribution and the onset from the original set. After conducting multiple hypothesis correction implemented in polysel, we chose significant pathways with an FDR <0.20.

Enrichment of private alleles at the promoter and regulatory regions

We aimed to detect enrichment of private alleles (alleles unique to one species) in the flanking regions from the bush dog and maned wolf genes. First, we inferred interspecific variation by conducting a join genotyping different canid species with HaplotypeCaller from GATK (McKenna et al., 2010). This group includes species of the genus *Chrysocyon* (maned wolf), *Speothos* (bush dog), *Lycalopex* (SA foxes), *Canis* (wolves, coyote, and golden jackal), *Cuon* (dhole), *Lycaon* (African wild dog), and *Lupalela* (jackals). Then, we combined

independent gVCF files with the "CombineGVCFs" option. We annotated the combined VCF for sites that passed our filter criteria (see annotation of genetic variant section for further details). We used BioMart in Ensembl (Ensembl Genes 104 release) to extract the start and end sites from 39,704 transcripts belonging to 32,704 genes from the dog genome canFam3.1. Then we generated a bed file with 1 kb windows upstream to the transcription start site (promoter region) and downstream to the transcription end site (potential regulatory region). We used this bedfile to calculate the number of private alleles for the bush dog and maned wolf in the combined VCF file containing genotypes calls for the different canid species. Within each 1 kb window, we used a custom python script to calculate the number of private alleles for the bush dog and maned wolf. In parallel, we calculated the average number of segregating sites across the remaining species (excluding maned wolf and bush dog). Considering that unique mutations at flanking regions could be overrepresented in the bush dog and maned wolf due to their relatively long evolutionary history (i.e. long branch lengths), we calculated the difference between private alleles between both species. Mutations overrepresented in one species but not in the other could be candidates of genes under positive selection. These estimates were calculated with the following formula:

$$\frac{P}{S} = \frac{\frac{\Sigma_{i=1}^{L}(A_i + 0.5a_i) - (B_i + 0.5b_i)}{L}}{\frac{\frac{\Sigma_{i=1}^{L} X_i + 0.5Y_i}{n}}{L}}$$

where *A* is the number of private alleles in the bush dog, a is the number of heterozygous sites with private alleles, B is the number of private alleles in the maned wolf genome, b is the number of heterozygous sites with private alleles, *L* is the total number of called sites with good quality

within each widow, *n* is the number of samples analyzed, *Xi* is the number of derived alleles with respect to the dog, and *Yi* is the number of heterozygous sites. After calculating the formula above on 38,542 1 kb window, we calculated empirical *p*-values. To verified that the number of sites was not influencing outlier windows, we plotted P/S vs. the number of sites (Figure 2-S4). Only windows with a minimum of 250 sites were included in the results. The custom pipeline can be found at https://github.com/dechavezv/2nd.paper.v2/tree/main/2-PositiveSelection/07-Selection_Regulatory_Regions/scripts

Figures

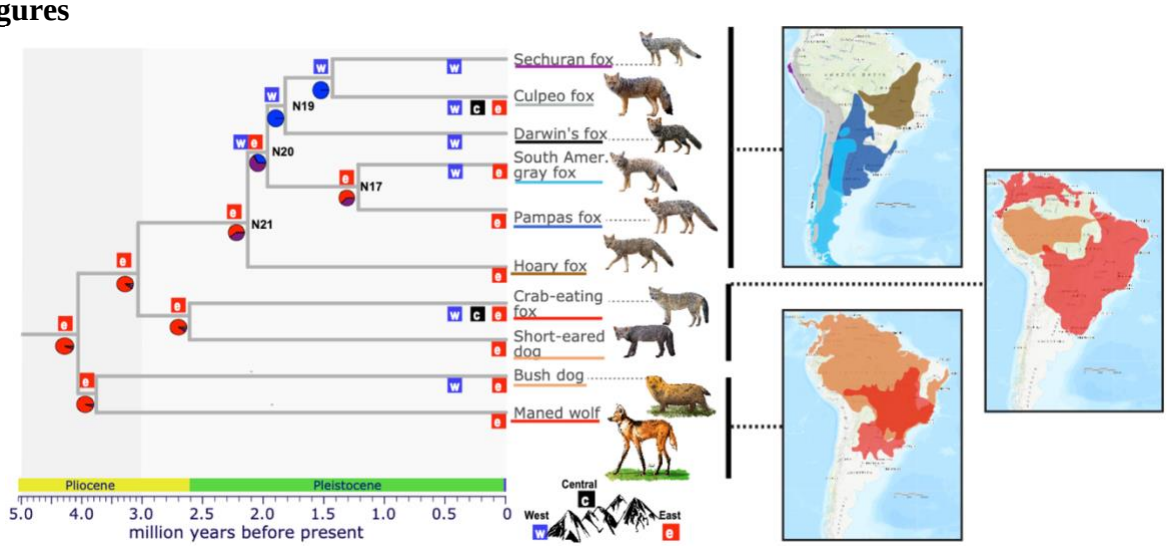

Figure 2-1. BioGeoBEARS (Matzke 2013) ancestral area reconstruction derived from the species tree obtained by ASTRAL-III (Zhang, et al. 2018). This tree was estimated from 6,716 25 kb windows and divergence times using MCMC tool in PAML (Yang 2007). A total of 31 genomes were included (Table 2-S1). The North American and Old World clade are shown in Supplementary Figure 2-1. The best fitting model was DIVALIKE with the parameter "J" that represents a founder event (Figure 2-S2 and Table 2-S2). The estimated ancestral ranges are shown by colored boxes on each node, while boxes at terminal branches indicate the current species distribution; red squares with the letter "e" indicate distributions east of the Andes, black "c" squares indicate the central region of the Andes, and blue "w" squares indicate the west side of the Andes. The probabilities of these ancestral regions are shown in the pie charts below each node. The maps on the right represent species distribution within a major clade. The colored distributions on the map match the colors underlining species names. Canid illustrations were reprinted from (Castelló 2018) © 2018 with permission from Princeton University Press.

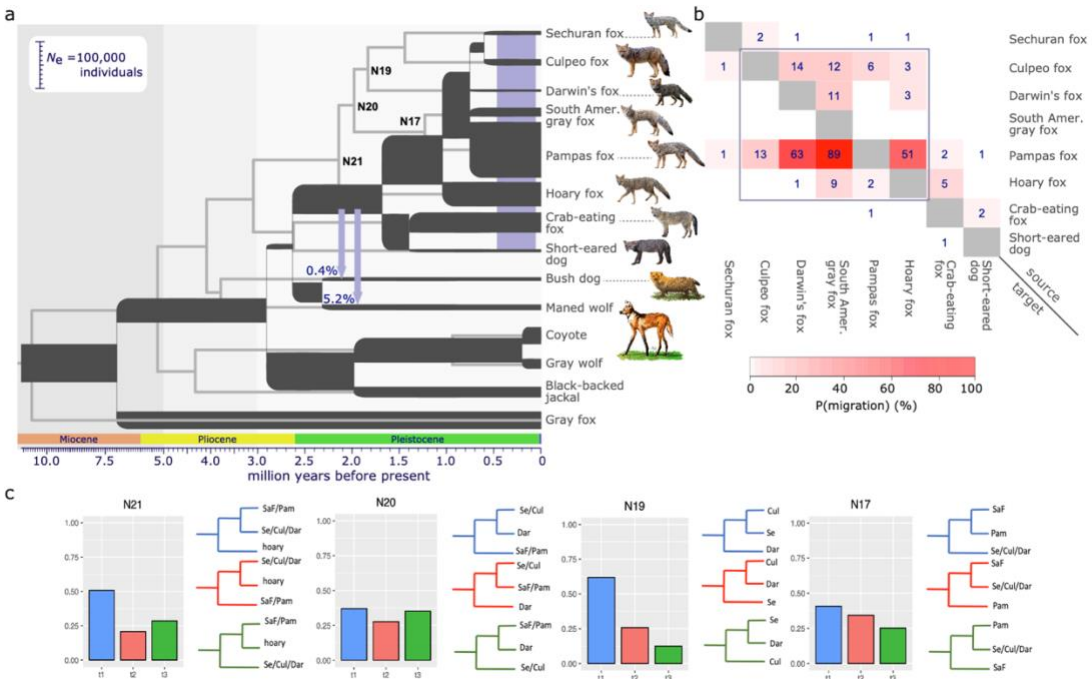

Figure 2-2. (a) Demographic model estimated using G-Phocs (https://github.com/gphocs-dev/G-PhoCS) based on 11, 636 putatively neutral 1 kb windows. The widths of branches in the demographic model are scaled proportionally to inferred effective population sizes (see scale bar at the top left), and their lengths are scaled proportionally to inferred species divergence times. Both scales assume an average per-generation mutation rate of $\mu = 4.0 \times 10{-}9$ and an average generation time of three years. The purple vertical bar along the terminal branches indicates evidence of gene flow after analyzing 37 directed migration bands among all SA canids (see Table 2-S3 for details). The species tree from Figure 2-1 is shown as light gray in the background. Canids illustrations were reprinted from (Castelló 2018) © 2018 with permission from Princeton University Press. **(b)** Probability of admixture among SA foxes. Gene flow probabilities within the genus Lycalopex are shown inside the purple square. **(c)** Quartet frequencies for a subset of internal nodes with three plausible tree topologies. The blue topology represents the tree in the top panel and Figure 2-1. Alternative topologies are shown in red and green colors. The labels on the terminal branches of the trees are Se = Sechuran fox, Dar = Darwin's fox, Cul= culpeo fox, Pam = pampas fox, SaF = South American gray fox. Only nodes with a relative high frequency among different topologies are shown.

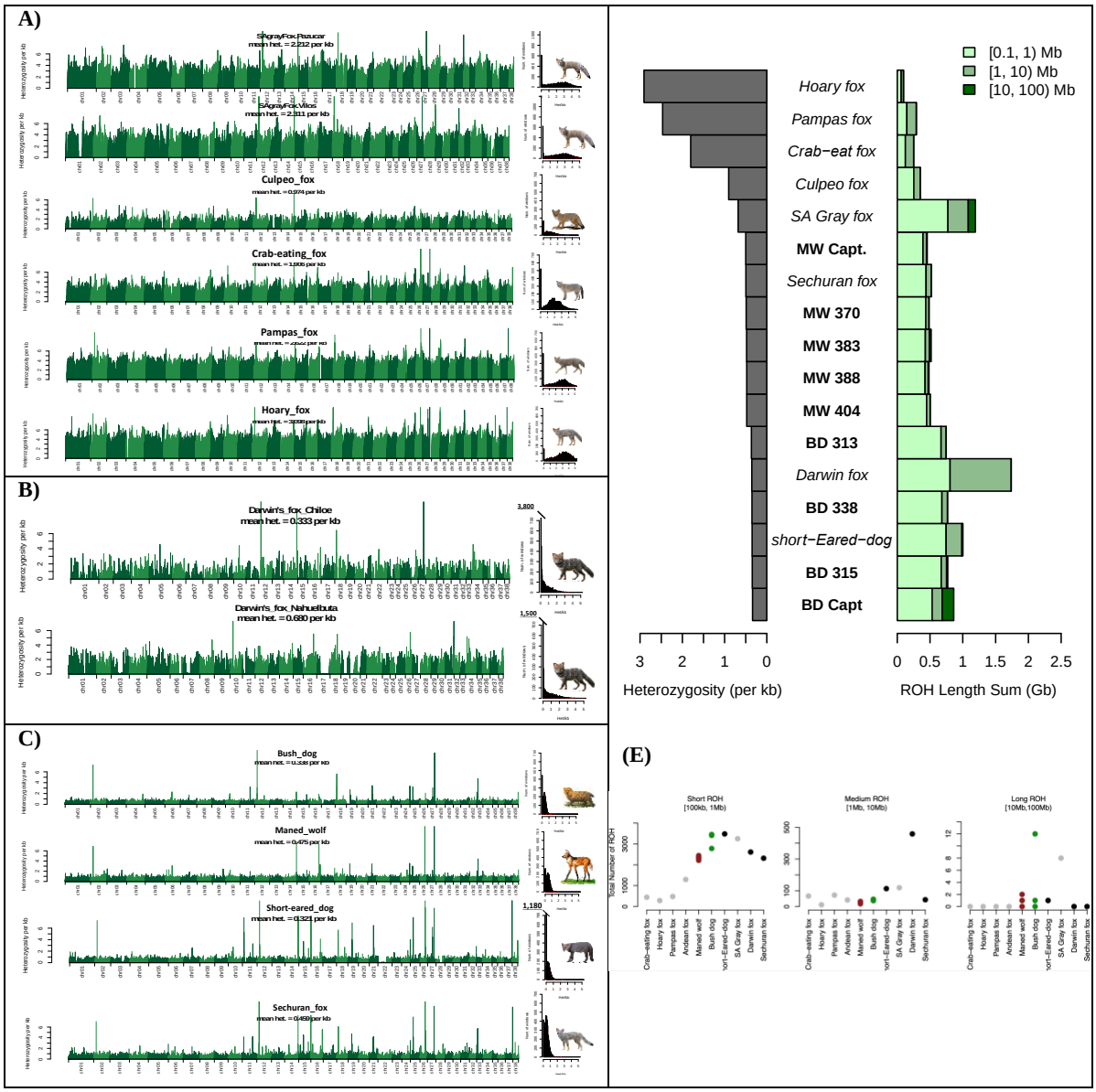

Figure 2-3. Heterozygosity per 100 kb window with a 10 kb step size across the genome of SA canids. Three different patterns were found: **(a)** Species with high heterozygosity thorough the genome. **(b)** Species with high heterozygosity alternating with long stretches of depleted genetic diversity. **(c)** Species with low heterozygosity thorough the genome. Canid illustrations were reprinted from (Castelló 2018) © 2018 with permission from Princeton University Press. **(d)** Histogram with per-site heterozygosity across the autosomal genome (left panel) and summed lengths of ROH of three specific length categories (right panel). On the right panel from top to bottom: Summed lengths of short (< 1 Mb), medium (1 Mb ≤ ROH < 10 Mb), and long (>10 Mb) ROH per individual. **(e)** Panels showing the total number of ROH in the three length categories. From left to right: short ROH indicating ancient inbreeding as in the short-eared dog and bush dog, medium ROH indicating ancient and historic inbreeding as in the Darwin's fox from Chiloé island, and long ROH indicating recent inbreeding as in the case of the Darwin's fox from Nahuelbuta and the inbred captive bush dog.

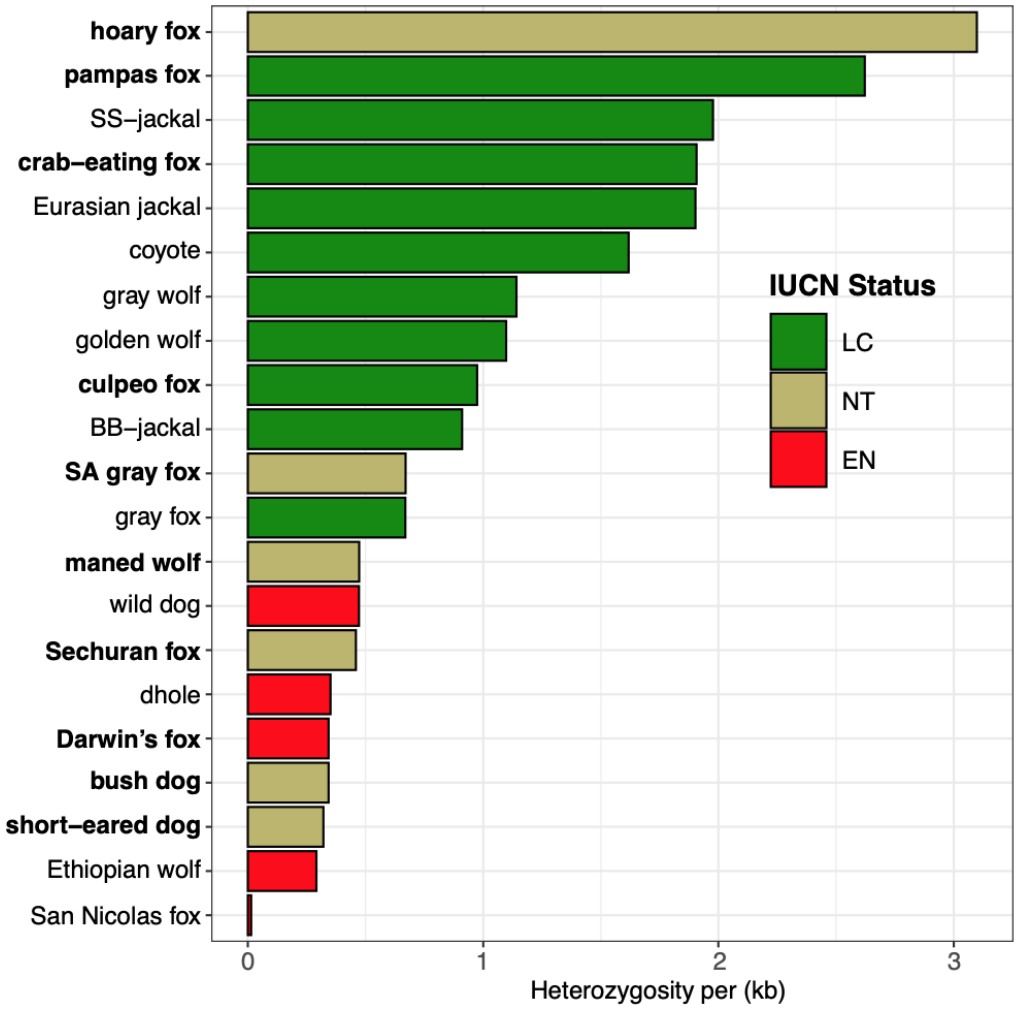

Figure 2-4. Genome-wide heterozygosity of canids. Names in bold indicate South American canids. The IUCN categories are LC = Least Concern, NT = Near Threatened, and EN = Endangered. Endangered canids under the IUCN have less than 0.5 heterozygote sites/kb. This group includes the African wild dog (wild dog), dhole, and Ethiopian wolf. South American canids with a heterozygote below this cutoff are the maned wolf, Sechuran fox, Darwin's fox, bush dog, and short-eared dog. Among these species only the Darwin's fox is listed as endangered.

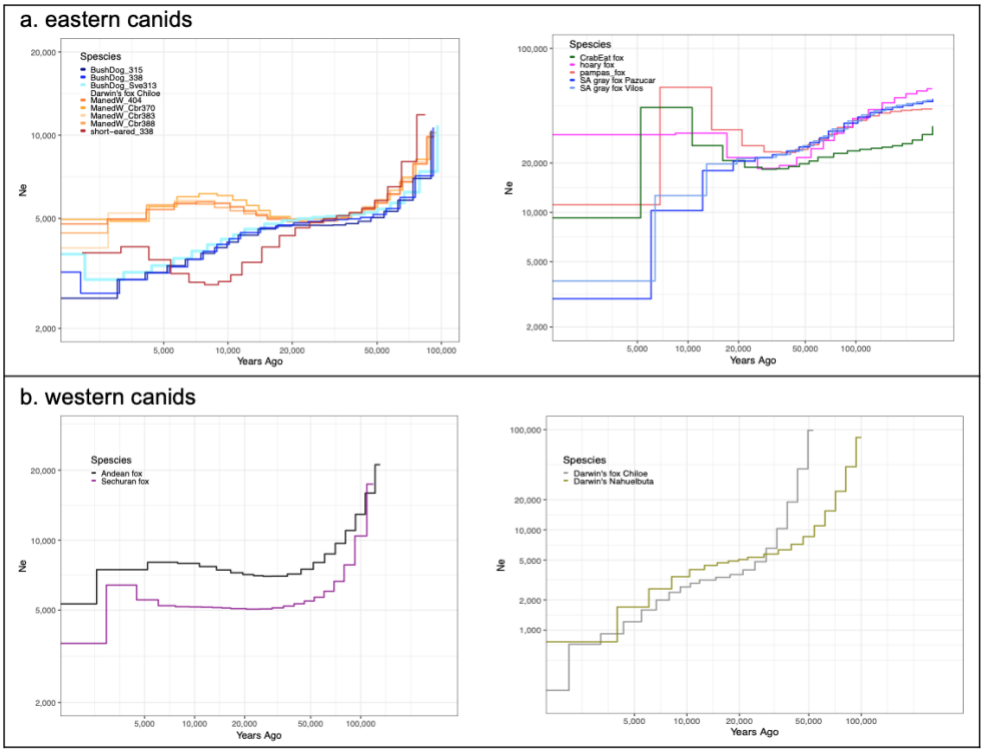

Figure 2-5. Demographic history of South American canids estimated using MSMC (Schiffels and Durbin 2014). The y axis corresponds to the inverse coalescent rate (IICR) scaled by 2μ, which is a proxy of the effective population size (Ne) through time (x-axis in thousands of years). IICR scaled to numerous generation times and mutation rates are shown in Supplementary Figure 2-S3.

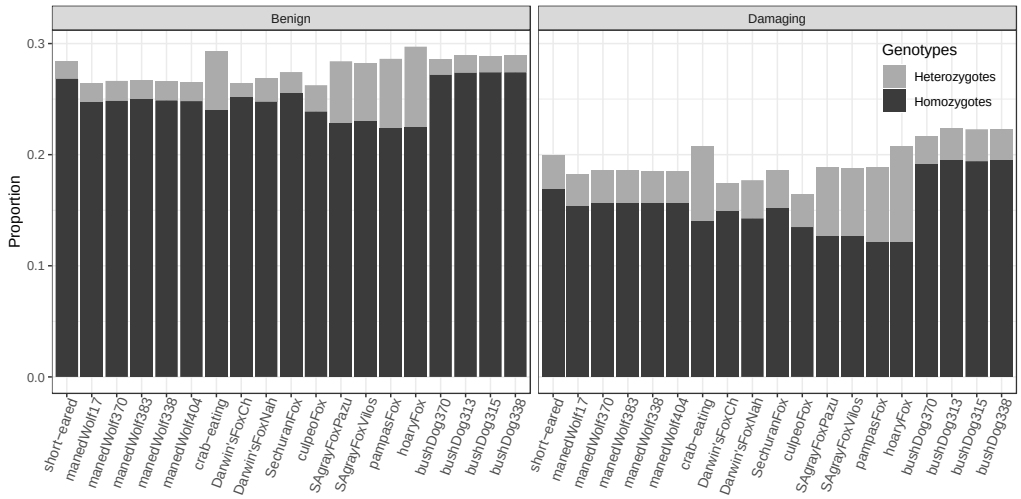

Figure 2-6. Count of derived alleles. Mutations were identified as benign (synonymous and tolerated missense mutations) and damaging (deleterious missense mutations, disruption of splice sites, and gain or loss of a stop codon) using the variant effect predictor tool (McLaren, et al. 2016). Only homozygous derived genotypes and heterozygotes are shown. The full list of alleles including homozygous ancestral genotypes are shown in the Supplementary Table 2-S5. Homozygous derived genotypes in the damaging category are higher in species with smaller populations like the bush dog and short-eared dog. In contrast, damaging mutations are less frequent in species with large populations, like the hoary fox. This difference is less evident under the benign category.

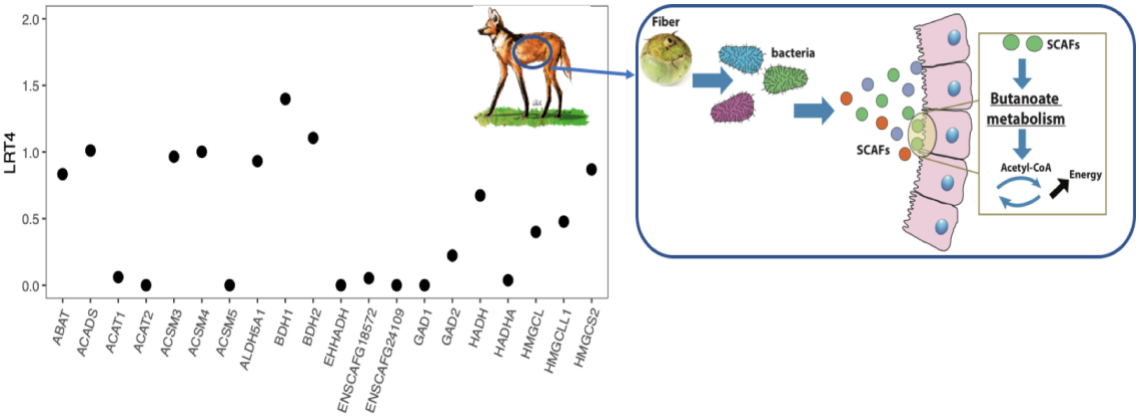

Figure 2-7. Adaptation to digest fruit fiber in the maned wolf. The left panel shows genes found to be enriched in the butanoate metabolism category, as inferred by polysel (Daub, et al. 2017). The right panel shows a schematic of fruit fiber digestion and the role of the candidate metabolic process. First, short fatty acids (SCAFs) represented by colored circles are obtained from fruit fiber fermentation (here represented as the "wolf apple" on the top left). The fermentation process is conducted by bacteria and takes place in the cecum where cells of the intestine, represented in pink, will absorb SCAFS. Within these cells, candidate genes under selection (represented as the word "Butanoate metabolism" in bold) convert SCAFs to energy precursors Acetyl-CoA.

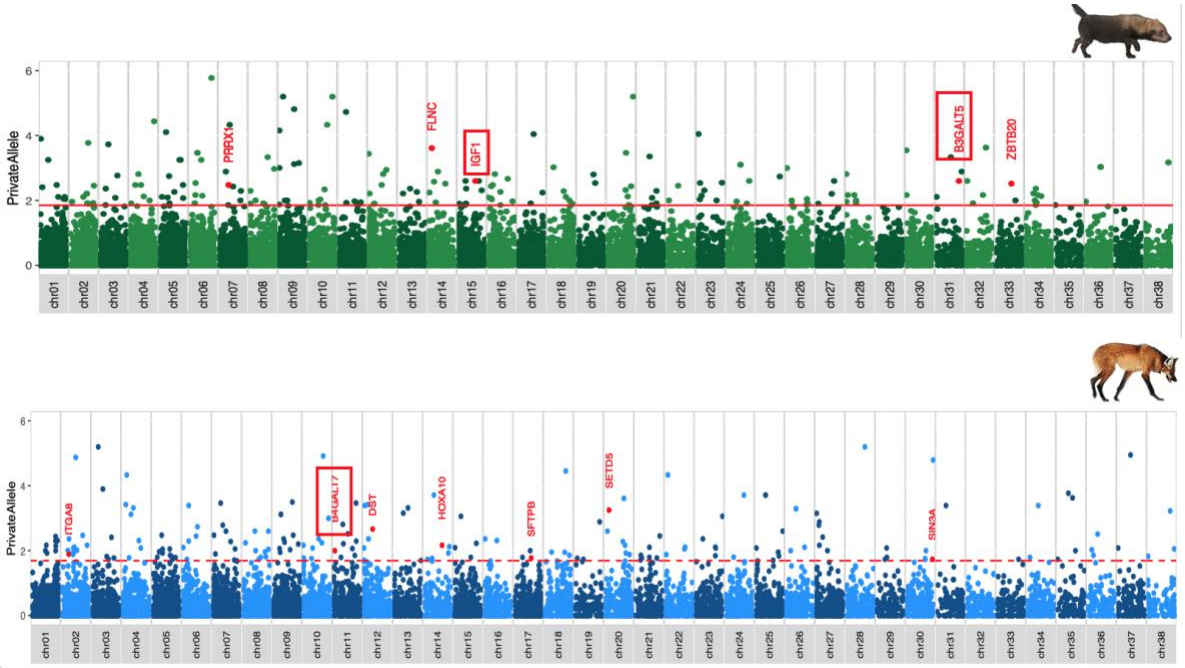

Figure 2-8. Rate of private alleles for the bush dog (top panel) and maned wolf (bottom panel) depicted from 1 kb windows flanking 39,704 transcripts belonging to 32,704 genes. Significant genes (score equivalent to empirical p-value < 0.01) are shown above the horizontal red line. Genes involved in limb development are shown in red. Genes relevant to limb elongation are highlighted by a red rectangle.

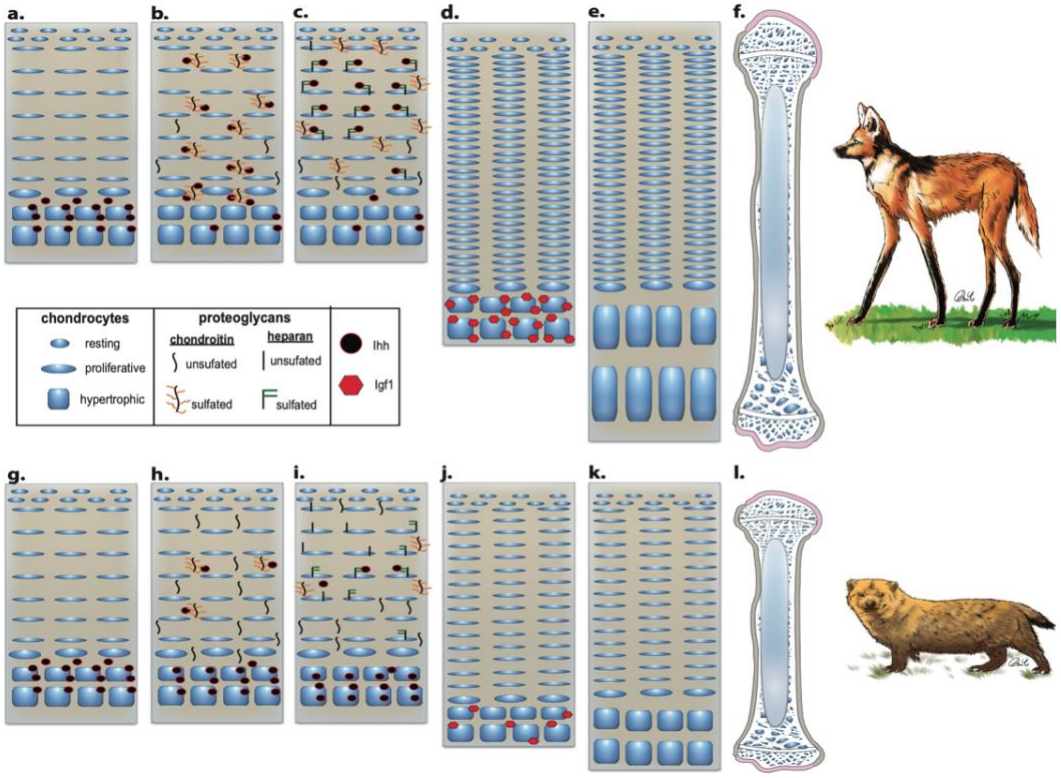

Figure 2-9. Limb elongation in the bush dog and maned wolf. The different panels represent the growth plate and the process of limb elongation in the maned wolf (a-f) and bush dog (g-l). **(a and g)** Small circles at the top show the resting region, the middle-flattened circles represent the proliferative area, and the large squares at the bottom represent the hypertrophic region. The Indian hedgehog protein (Ihh; black circles) is initially produced by hypertrophic chondrocytes. Then as shown in panels **(b and h),** sulfated chondroitin proteoglycans (black oblique lines with orange lines) transport Ihh to the top of the proliferative region. Note that unsulfated proteoglycans (black oblique lines) cannot transport Ihh. Panels **(c and i)** indicate the entrance of Ihh into the target cell facilitated by sulfated heparan proteoglycans (black straight lines with green vertical lines). Note that Ihh will not ligate to unsulfated heparan (black straight lines). Once inside the cell, Ihh promotes the replication of chondrocytes at the proliferative region. Note the multiplication of chondrocytes on panels d and j. Finally, there is an enlargement of hypertrophic chondrocytes (squares at the bottom) regulated by the gene IGF1, represented by red-colored hexagons (**e and k**). The rate of chondrocyte multiplication at the proliferative regions, coupled with hypertrophic enlargement at the bottom (squares), determines the length of the bone **(f and i)**.

Appendix 2-I: Supplementary Figures

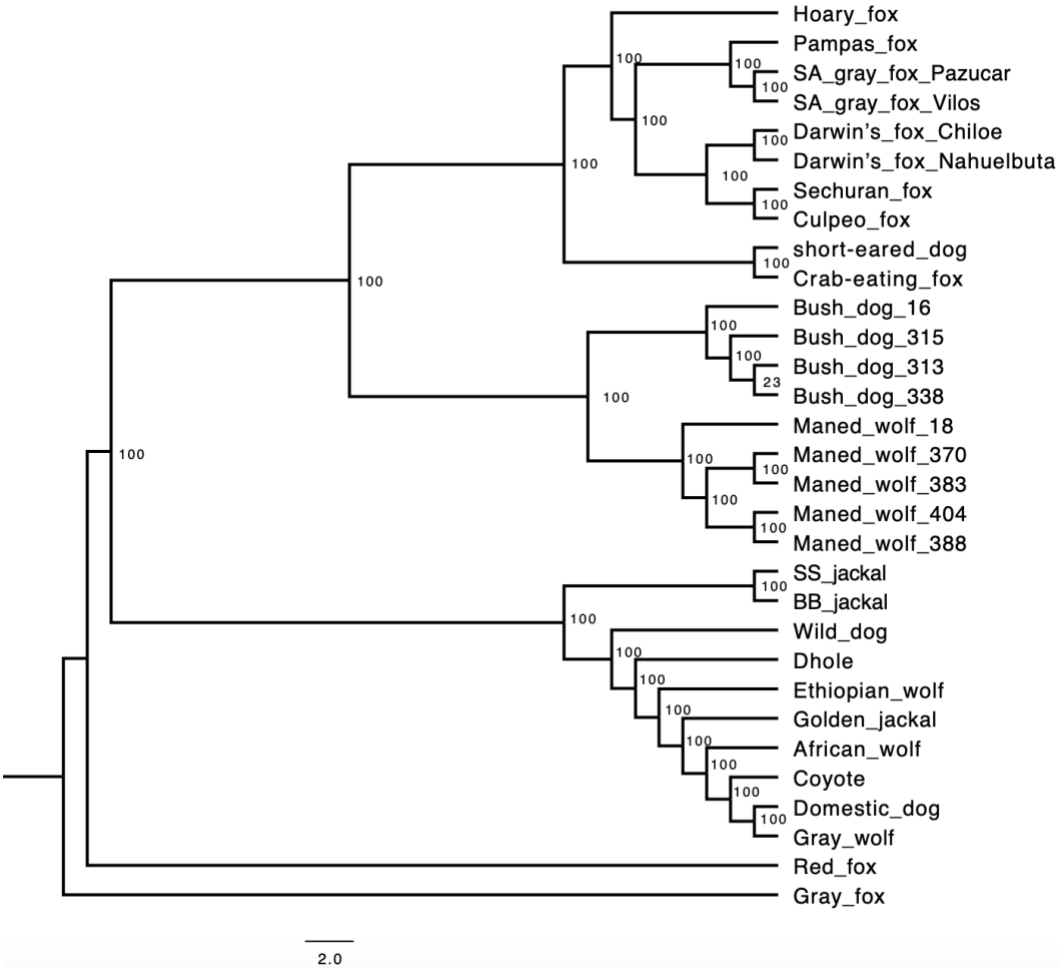

Figure 2-S1. Coalescent species tree based on 6,716 25 kb windows, as inferred by ASTRAL-III (Zhang et al., 2018). A total of 31 genomes corresponding to 23 species were included in the tree (Table 2-S1). Bootstrap supports are shown at the nodes of the tree. The scale bar at the bottom indicates the number of substitutions per site.

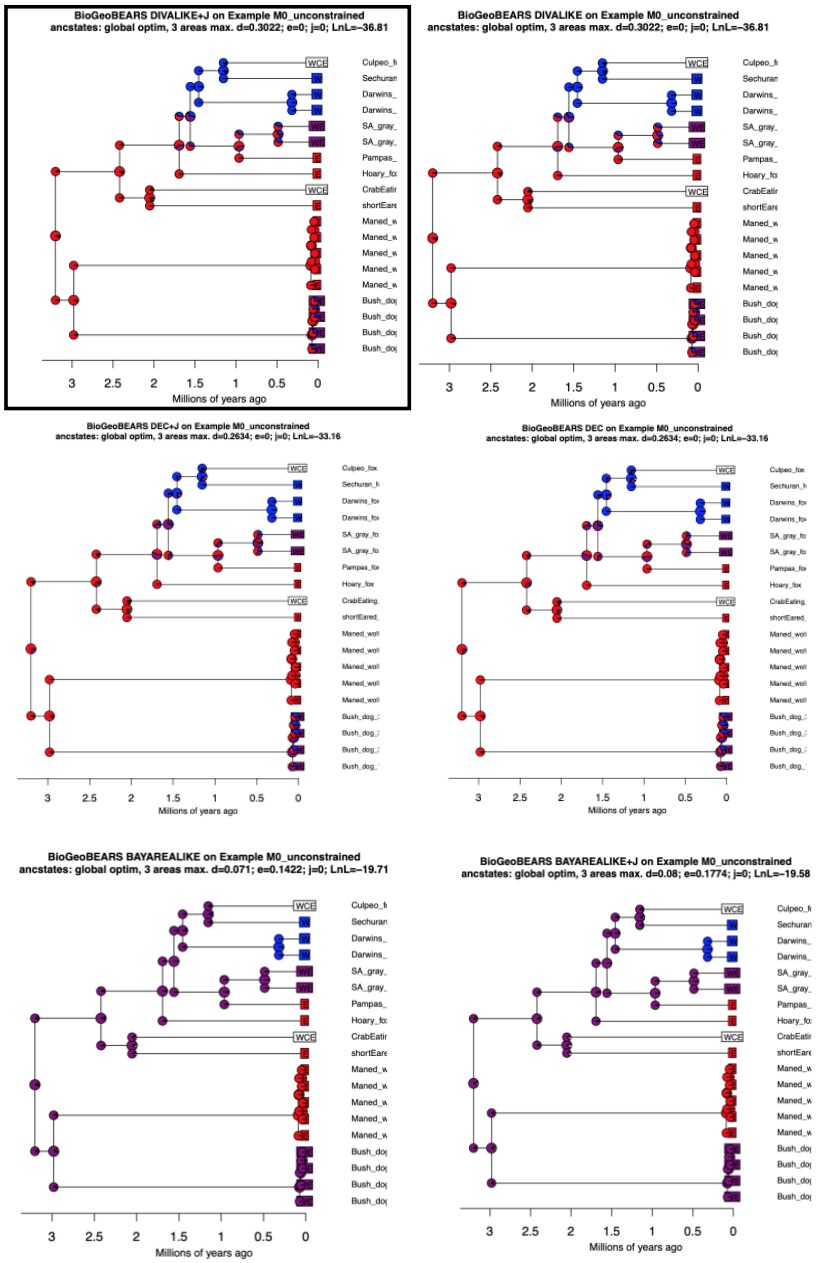

Figure 2-S2. Inference of ancestral distributions depicted by BioGeoBEARS (Matzke, 2013). Six different models were analyzed: DEC, DEC+J, BAYAREALIKE, BAYAREALIKE +J, DIVALIKE, and DIVALIKE+J. The parameter "j" represents a founder event. The probability of different ancestral distributions is indicate by the pie charts located at the nodes of the tree: red = East side of the Andes, blue = West of the Andes, and purple = West and East of the Andes. Letters on terminal branches corresponded to the species' current distribution: W = west of the Andes, C = central region of the Andes, and E = east side of the Andes. The model with the best AIC score was DIVALIKE+J (see Table 2-S2 for a full list of AIC scores). This model suggests that the ancestor of culpeo, Sechuran, Darwin's, pampas, and SA gray fox was the first lineage to colonize the west of the Andes. Species with a current western distribution have an ancestral range at the west of the Andes. Others with an eastern distribution, have an ancestral range at the east of the Andes. Most species with a current distribution across different areas (WE or WCE) have an ancestral distribution at the east of the Andes. This group includes the bush dog, crab-eating fox and SA gray fox. The expectation is the culpeo fox that has an ancestral range at the west of the Andes.

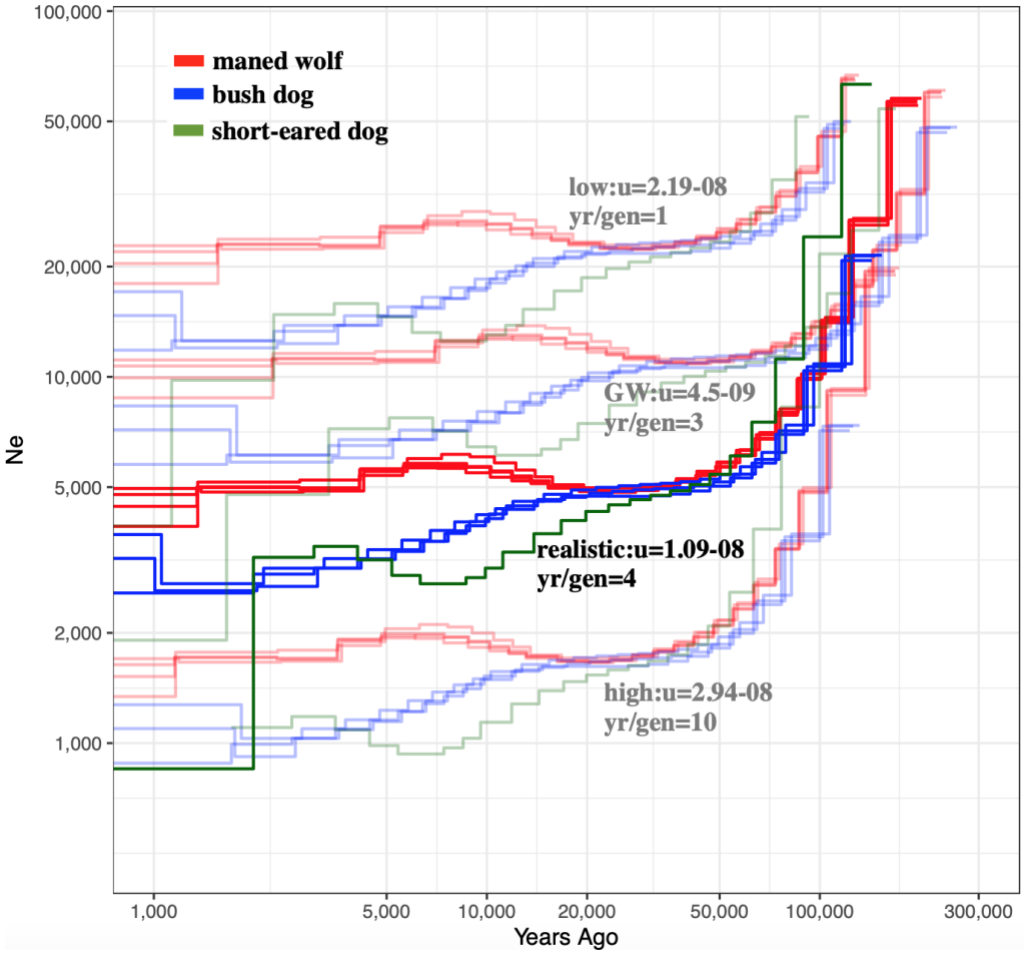

Figure 2-S3. Demographic model of the rain forest specialists (bush dogs and short-eared dogs), and the savanna-like specialist (maned wolf). The figure shows the inference of inverse coalescent rate (IICR) scaled by 2μ, through time (in terms of years), for different paired values of mutation rates indicated by μ, and generation times shown as yer/gen. The "low" mutation rate (faint lines at the top) is the lowest extreme of the values tested, calculated assuming a 1-year generation time and $\mu = 2.19 \times 10{-}8$. The "realistic" mutation rate, which is presented in Figure 2-5, is calculated based on a 4-year generation time and $\mu = 1.19 \times 10{-}8$. The "high" mutation rate (faint lines at the bottom) is calculated based on a 10-year generation time and $\mu = 2.94 \times 10{-}8$. The range of mutation rates and generation times indicate how much the MSMC's trajectories could vary with different generation times and mutation rates. For instance, the decline of the IICR on forest specialists, the bush dog, and short-eared dog, could started anywhere between 15,000 and 25,000 years ago.

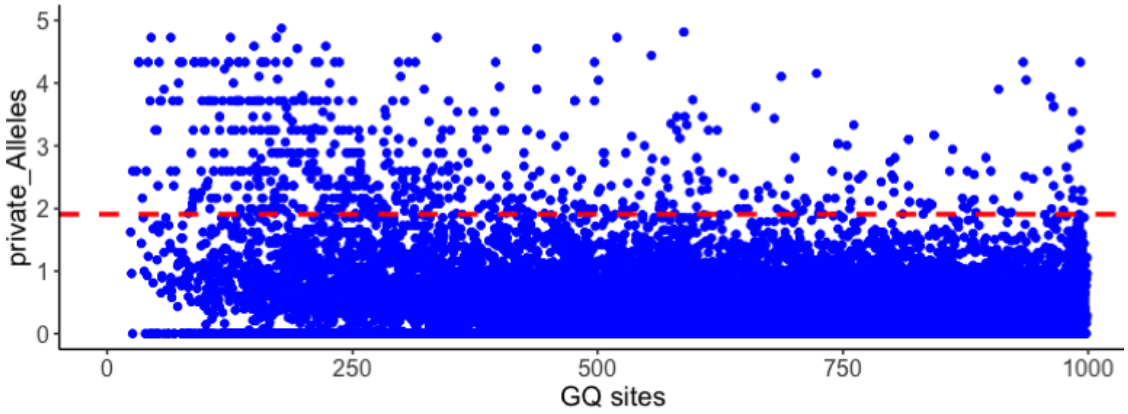

Figure 2-S4. The relationship between private alleles and number of sites within 1 kb windows flanking 39,704 transcripts. The y-axis indicates the proportion of private alleles in the bush dog and maned wolf divided by the average of segregation sites among other canids (see methods for details). Significant genes (score equivalent to empirical p_value < 0.01) are shown above the horizontal red line. The x axis indicates the number of sites within a particular 1kb window that passed our filter criteria (see methods for details). The windows that contained less than 250 good quality sites ("GQ sites") were ignored.

Appendix 2-II: Supplementary Tables

Table 2-S1. Descriptive statistics of the 30 canid genomes analyzed in this study. Statistics were calculated with Qualimap (Okonechnikov et al., 2016) on BAM files that were mapped to the domestic dog canFam3.1 (see methods).

Common name	Scientific name	Sample ID	Reference	Mean depth coverage (X)	Base Pairs (Gbp)	% of reference covered by at least 3 reads
Maned wolf 18	*Chrysocyon brachyurus*	Cbr17082018*	This study	77.7	98.34	98.48
Maned wolf 370	*Chrysocyon brachyurus*	bcbr370	This study	27,01	62.9	98.29
Maned wolf 383	*Chrysocyon brachyurus*	bcbr383	This study	27.96	65	98.37
Maned wolf 388	*Chrysocyon brachyurus*	bcbr388	This study	22.1	51.4	98.23
Maned wolf 404	*Chrysocyon brachyurus*	bcbr404	This study	24.35	56.7	98.26
Bush dog 16	*Speothos venaticus*	Sve16082018*	This study	42.4	98.8	98.5
Bush dog 313	*Speothos venaticus*	bsve313	This study	73.5	171.2	98.55
Bush dog 315	*Speothos venaticus*	bsve215	This study	20.37	47.4	98.15
Bush dog 338	*Speothos venaticus*	bsve338	This study	24.7	57.5	98.22
Short-eared dog	*Atelocynus microtis*	AMI	This study	17.23	40.1	98.23
Crab-eating fox	*Cerdocyon thous*	Cth	This study	23	65.3	98.67
Hoary fox	*Lycalopex vetelus*	Lvet	This study	32.12	74.8	98.72
Pampas fox	*Lycalopex gymnocercus*	Lgy	This study	11.78	27.4	98.03
SA gray fox Pazucar	*Lycalopex griseus*	Lgri.015	This study	16.7	38.6	98.4
SA gray fox Vilos	*Lycalopex griseus*	Lgri.006	This study	18	41.7	98.45
Darwin's fox Chiloe	*Lycalopex fulvipes*	DFU_Ch	This study	16.1	37.4	98.25
Darwin's fox Nahuelbuta	*Lycalopex fulvipes*	DFU_Na	This study	19.3	45	98.4
Sechuran fox	*Lycalopex sechurae*	DSE	This study	16.14	37.6	98.28
Black back jackal	*Lupulella mesomelas*	BBJ	(Perri et al., 2021)	57	132	98.9
Side strip jackal	*Lupulella aduta*	SSJ	Perri, et al. 2021)	37.8	88	97.63
African wolf	*Canis lupaster*	RKW1356	(Koepfli et al., 2015)	27.96	65	98.11
Andean fox	*Lycalopex culpaeus*	SRS523207	(Auton et al., 2013)	10.98	25.5	97.71
AWD South Africa	*Lycaon pictus*	SAMN09924608	This Study	27.93	65	98.71
Coyote	*Canis latrans*	RKW13455	(vonHoldt et al., 2016)	25.67	59.7	98.72

Dhole	*Cuon alpinus*	SAMN10180424	(Gopalakrishnan et al., 2018)	19.53	45.4	98.35
Ethiopian wolf	*Canis simensis*	SAMN10180425	(Gopalakrishnan et al., 2018)	9.66	22.5	97.37
Golden jackal	*Canis aureus*	SAMN03366713 (RKW1332)	(Freedman et al., 2014)	26.09	60.7	97.46
Gray fox	*Urocyon cinereoargenteus*	SAMN04495241	(Robinson et al., 2016)	18.47	42.9	97.49
Gray wolf	*Canis lupus*	RKW1547	(Fan et al., 2016)	27.33	63.6	98.76
Red fox	*Vulpes vulpes*	SAMN06553695	(Kukekova et al., 2018)	70	173.2	98.60

* Samples used for the construction of the de novo assembly

Table 2-S2. Descriptive statistics of the ancestral species distribution models using BioGeoBEARS (Matzke, 2013). The log-likelihoods for different models are shown in column "LnL". Each model contained a specific number of parameters. The best fitting model (shown in bold) was selected based on the corrected Akaike information criterion (AICc) scores. The parameters are d = rate of range-expansion dispersal, e = rate of range-contraction/local extirpation, and j=founder event.

Model	LnL	#Parameters	d	e	j	AICc
DEC	-33.16	2	0.26	2.00E-08	0	71.06
DEC+J	-33.16	3	0.26	1.00E-12	1.00E-05	73.91
DIVALIKE	-36.81	2	0.3	1.00E-12	0	78.38
DIVALIKE+J	**-36.81**	**3**	**0.3**	**1.00E-12**	**1.00E-05**	**81.23**
BAYAREALIKE	-19.71	2	0.071	0.14	0	44.16
BAYAREALIKE+J	-19.58	3	0.08	0.18	1.00E-05	46.77

Table 2-S3. Demographic parameter estimates inferred by G-PhoCS (https://github.com/gphocs-dev/G-PhoCS). Analysis was done under a model with population phylogeny with the topology of the species tree inferred by ASTRAL-III (Figure 2-1) and 37 migration bands (see total migration rates on the left column). Effective population size were calibrated by assuming an average per-site mutation rate of $\mu=4.0\times10^{-9}$ (Skoglund et al., 2015), and divergence times were calibrated assuming the same rate and an average generation time of three years. Total migration rates were obtained by multiplying the mutation-scaled rate with the duration of time of the migration band.

Parameter	Mean estimate (95% Bayesian CI)
	Tree inferred by ASTRAL-III
Population divergence times (years)	
tau_SF_CF	607300(585500-627000)
tau_SA5	757800(737900-778100)
tau_PAMPAS_ANC	759100(738800-779200)
tau_SA4	760000(739700-780000)
tau_SA3	1047900(1020300-1076700)
tau_CEF_SED	1399900(1356000-1442100)
tau_SA2	1684600(1670000-1699800)
tau_BD_MW	2322800(2287300-2359600)
tau_SA1	2636900(2623300-2650300)
tau_GW_CO	202300(192500-212000)
tau_GW_BB	1987100(1958300-2015500)
tau_GW_SA	2912600(2897700-2927400)
tau_ROOT	6994900(6924700-7069400)
Effective population sizes (number of individuals)	
theta_SECHURAN_F	8700(8400-9000)
theta_CULPEO_F	13800(13300-14400)
theta_DARWINS_F	5300(5100-5500)
theta_SA_GRAY_F	19200(18500-19900)
theta_PAMPAS_F	118400(114200-122600)
theta_HOARY_F	52000(49500-54600)
theta_CRAB_EATING_F	43000(41900-44200)
theta_SHORT_EARED_D	6500(6200-6800)
theta_BUSH_D	7900(7600-8200)
theta_MANED_W	12000(11600-12400)
theta_YELLOWSTONE_W	21300(20200-22400)
theta_CALIFORNIA_C	36500(34400-38700)
theta_BB_JACKAL	22800(22200-23400)
theta_GRAY_F	37300(36400-38200)
theta_SF_CF	22200(18100-26500)
theta_SA5	6200(700-15100)
theta_SA4	142000(130500-154500)
theta_SA3	105200(101300-109200)
theta_CEF_SED	77300(65300-89600)
theta_SA2	62800(60900-64800)
theta_BD_MW	42800(35200-50800)
theta_SA1	11600(10900-12400)
theta_GW_CO	52600(51400-53700)
theta_GW_BB	80800(76700-85000)
theta_GW_SA	51400(50200-52600)
theta_ROOT	81100(75700-86400)

theta_SECHURAN_F	8700(8400-9000)
theta_CULPEO_F	13800(13300-14400)
theta_DARWINS_F	5300(5100-5500)
theta_SA_GRAY_F	19200(18500-19900)
theta_PAMPAS_F	118400(114200-122600)
theta_HOARY_F	52000(49500-54600)
theta_CRAB_EATING_F	43000(41900-44200)
theta_SHORT_EARED_D	6500(6200-6800)
theta_BUSH_D	7900(7600-8200)
theta_MANED_W	12000(11600-12400)
theta_YELLOWSTONE_W	21300(20200-22400)
theta_CALIFORNIA_C	36500(34400-38700)
theta_BB_JACKAL	22800(22200-23400)
theta_GRAY_F	37300(36400-38200)
theta_SF_CF	22200(18100-26500)
theta_SA5	6200(700-15100)
theta_SA4	142000(130500-154500)
theta_SA3	105200(101300-109200)
theta_CEF_SED	77300(65300-89600)
theta_SA2	62800(60900-64800)
theta_BD_MW	42800(35200-50800)
theta_SA1	11600(10900-12400)
theta_GW_CO	52600(51400-53700)
theta_GW_BB	80800(76700-85000)
theta_GW_SA	51400(50200-52600)
theta_ROOT	81100(75700-86400)
Total migration rates (Gene flow)	
m_PAMPAS_F->SA_GRAY_F	89.0(87.9-89.9)
m_PAMPAS_F->DARWINS_F	63.2(61.1-64.9)
m_PAMPAS_F->HOARY_F	51.2(48.2-53.9)
m_CULPEO_F->DARWINS_F	13.6(12.2-14.9)
m_PAMPAS_F->CULPEO_F	13.2(11.3-15.1)
m_CULPEO_F->SA_GRAY_F	11.6(10.4-12.8)
m_DARWINS_F->SA_GRAY_F	11.0(9.4-12.5)
m_HOARY_F->SA_GRAY_F	8.7(6.1-11.0)
m_CULPEO_F->PAMPAS_F	5.6(4.7-6.3)
m_HOARY_F->CRAB_EATING_F	5.4(4.4-6.5)
m_SA2->MANED_W	5.2(3.7-6.6)
m_DARWINS_F->HOARY_F	2.9(1.8-3.9)
m_CULPEO_F->HOARY_F	2.8(2.1-3.6)
m_PAMPAS_F->CRAB_EATING_F	2.3(1.5-3.1)
m_HOARY_F->PAMPAS_F	2.2(0.0-4.1)
m_CRAB_EATING_F->SHORT_EARED_D	1.7(0.9-2.6)
m_SECHURAN_F->CULPEO_F	1.7(1.0-2.4)
m_CULPEO_F->SECHURAN_F	1.1(0.5-1.8)
m_PAMPAS_F->SECHURAN_F	1.0(0.2-1.8)
m_CRAB_EATING_F->PAMPAS_F	0.9(0.7-1.1)
m_SECHURAN_F->PAMPAS_F	0.7(0.4-1.1)
m_SECHURAN_F->DARWINS_F	0.7(0.3-1.2)
m_SHORT_EARED_D->CRAB_EATING_F	0.7(0.3-1.1)
m_HOARY_F->DARWINS_F	0.7(0.0-3.3)

m_PAMPAS_F->SHORT_EARED_D	0.7(0.4-1.0)
m_SECHURAN_F->HOARY_F	0.5(0.2-0.9)
m_SA2->BUSH_D	0.4(0.0-2.1)
m_SA3->BUSH_D	0.4(0.0-0.8)
m_DARWINS_F->CULPEO_F	0.4(0.0-1.3)
m_CRAB_EATING_F->SECHURAN_F	0.4(0.2-0.6)
m_SECHURAN_F->SA_GRAY_F	0.3(0.0-0.8)
m_HOARY_F->SECHURAN_F	0.3(0.0-0.8)
m_BUSH_D->SA2	0.3(0.1-0.5)
m_SA2->BD_MW	0.1(0.0-0.6)
m_SA_GRAY_F->DARWINS_F	0.1(0.0-0.5)
m_SA_GRAY_F->HOARY_F	0.0(0.0-0.5)
m_HOARY_F->CULPEO_F	0.0(0.0-0.5)

Table 2-S4. Genome-wide heterozygosity and runs of homozygosity (ROH) of South American canids. ROH were categorized as short (<1Mb), medium (1-10Mb), and long (>10Mb).

Genome	Heterozygosity (kb)	Shot ROH (MB)	Medium ROH (MB)	Long ROH (Mb)
Hoary fox	2.90	57.23	37.38	0
Crab-eat fox	1.78	119.98	128.38	0
Pampas fox	2.47	143.29	144.21	0
SA gray fox los Vilos	2.22	272.99	131.60	11.67
Culpeo fox	0.90	244.43	95.933	0
SA gray fox Pan de azucar	2.11	375.24	95.48	0
Maned wolf Capt.	0.47	368.17	36.82	26.30
Maned wolf 388	0.47	404.38	47.55	11.63
Maned wolf 383	0.48	405.86	64.19	25.71
Sechuran fox	0.49	424.55	75.77	0
Maned wolf 370	0.49	411.28	47.75	0
Maned wolf 404	0.47	428.43	50.79	0
Darwin's fox Ch	0.35	793.13	935.60	0
Bush dog Captive	0.33	515.74	150.12	174.87
Darwin's fox Na	0.68	759.00	293.99	131.03
Bush dog 313	0.36	647.203	73.28	0
Bush dog 315	0.34	652.35	78.75	17.60
Short-Eared-dog	0.34	721.51	235.20	15.09
Bush dog 338	0.35	663.25	81.12	0

Table 2-S5. Genotype counts in South American canids. Homozygote derived genotypes in the damaging category are higher in smaller populations like the bush dog and short-eared dog (bolded) due to increased drift. In contrast, damaging mutations are less frequent in large populations like the hoary fox. This difference is less evident under the benign category.

SampleId	HomAnc	HomDerv	Het	Total	MutType
short-eared dog	8901.2	3333.24	193.87	Total	benign
Maned_wolf_18	9230.8	3098.34	215.66	17647.4	benign
Maned_wolf_370	9148.9	3098.24	222.97	17647.4	benign
Maned_wolf_383	9071.6	3082.87	210.37	17647.4	benign
Maned_wolf_388	9133.7	3096.03	211.82	17647.4	benign
Maned_wolf_404	9156	3093.39	211.61	17647.4	benign
crab-eating fox	8794.3	2986.39	650.55	17647.4	benign
Darwin's_Chiloe	9165.5	3131.95	158.32	17647.4	benign
Darwin's_Nahuelbuta	9092.8	3077.24	258.45	17647.4	benign
Sechuran_fox	9005.8	3164.76	240.68	17647.4	benign
culpeo_fox	8626.6	2793.37	277.95	17647.4	benign
SA_gray_fox_Vilos	8931.2	2842.26	701.42	17647.4	benign
SA_gray_fox_Pazucar	8948.1	2863.45	652.66	17647.4	benign
pampas_fox	8469.9	2653.03	743.95	17647.4	benign
hoary_fox	8776.6	2803.76	908.29	17647.4	benign
bush_dog_16	8915.6	3391.76	174.4	17647.4	benign
bush_dog_313	8856.3	3407.42	192.4	17647.4	benign
bush_dog_315	8838.8	3405.95	188.4	17647.4	benign
bush_dog_338	8833.1	3406.63	187.26	17647.4	benign
short-eared dog	**939.95**	**198.632**	**35.421**	**2838.08**	**damaging**
Maned_wolf_18	974.32	182.921	34	2838.08	damaging
Maned_wolf_370	959.42	183.921	35.263	2838.08	damaging
Maned_wolf_383	948.82	182.368	33.605	2838.08	damaging
Maned_wolf_388	956.29	183.553	33.263	2838.08	damaging
Maned_wolf_404	960.13	183.684	34.868	2838.08	damaging
crab-eating	926.74	163.789	78.368	2838.08	damaging
Darwin's_Chiloe	975.97	176.158	29.368	2838.08	damaging
Darwin's_Nahuelbuta	967.95	167.842	40.421	2838.08	damaging
Sechuran_fox	960.47	178.763	40.579	2838.08	damaging
culpeo_fox	843.4	136.079	29.5	2838.08	damaging
SA_gray_fox_Vilos	964.18	149.974	74.29	2838.08	damaging
SA_gray_fox_Pazucar	963.58	149.579	73.447	2838.08	damaging
pampas_fox	910.16	135.579	76.263	2838.08	damaging
hoary_fox	936.18	142.526	102.08	2838.08	damaging
bush_dog_16	**928.95**	**227.079**	**29.737**	**2838.08**	**damaging**
bush_dog_313	**911.55**	**228.342**	**34.079**	**2838.08**	**damaging**
bush_dog_315	**912.82**	**228**	**33.79**	**2838.08**	**damaging**
bush_dog_338	**909.03**	**227.789**	**33.079**	**2838.08**	**damaging**

Table 2-S6. Candidate genes for positive selection in the bush dog, as identified by the branch-site test in PAML4 (Yang, 2007). Q values were determinate after a multiple hypothesis-testing correction of two foreground branches (bush dog and maned wolf) and 17,185 genes. The top 20 genes are shown. The gene associated with limb elongation is shown in bold.

Ensembel_ID	GeneSymbol	Pvalue	Qvalue
ENSCAFG00000019513	MAPK8IP3	5.71E-07	0.00631248
ENSCAFG00000003587	TRAF1	5.93E-06	0.02536113
ENSCAFG00000000814	HLA-DQB2	1.85E-05	0.0632959
ENSCAFG00000028925	NDUFAF8	3.20E-05	0.09123733
ENSCAFG00000015765	TFE3	5.41E-05	0.13221267
ENSCAFG00000019737	KCNT1	8.02E-05	0.17149768
ENSCAFG00000003181	N/A	0.00010989	0.20887647
ENSCAFG00000017618	DRP2	0.00024554	0.38185934
ENSCAFG00000008863	SCAF4	0.00027261	0.38862827
ENSCAFG00000030681	N/A	0.00036163	0.47587726
ENSCAFG00000003434	VAX2	0.00053129	0.57534564
ENSCAFG00000030605	N/A	0.00053629	0.57534564
ENSCAFG00000000231	**B4GALT7**	**0.00054806**	**0.57534564**
ENSCAFG00000009270	PEBP4	0.00069065	0.62184305
ENSCAFG00000031187	FGF12	0.00074199	0.63465944
ENSCAFG00000005990	C15H12ORF50	0.0007999	0.6516097
ENSCAFG00000029971	OTUD6B	0.0009485	0.70701452
ENSCAFG00000031487	N/A	0.00095772	0.70701452
ENSCAFG00000018673	SMARCA1	0.00102471	0.70701452
ENSCAFG00000001299	ZC3H3	0.00103322	0.70701452

Table 2-S7. Candidate genes for positive selection in the maned wolf, as identified by the branch-site test in PAML4 (Yang, 2007). Q values were determinate after a multiple hypothesis-testing correction of two foreground branches (bush dog and maned wolf) and 17,185 genes. The top 20 genes are shown.

Ensmebl_ID	GeneSymbol	Pvalue	qvalue
ENSCAFG00000004101	BICRA	2.07E-08	0.00035604
ENSCAFG00000010121	CD81	4.43E-05	0.239424
ENSCAFG00000005853	PPP1R27	6.87E-05	0.239424
ENSCAFG00000018235	N/A	6.96E-05	0.239424
ENSCAFG00000019084	NDC1	0.00010119	0.290078
ENSCAFG00000001554	BOP1	0.00014016	0.33940932
ENSCAFG00000006785	N/A	0.00017288	0.33940932
ENSCAFG00000017805	REC114	0.00018369	0.33940932
ENSCAFG00000016111	FOXK1	0.00019733	0.33940932
ENSCAFG00000011857	SLAMF9	0.0002204	0.34462389
ENSCAFG00000001372	SCRIB	0.0004468	0.57128844
ENSCAFG00000009628	GOLT1A	0.00055128	0.57128844
ENSCAFG00000030346	N/A	0.0005755	0.57128844
ENSCAFG00000005759	NPLOC4	0.00061634	0.57128844
ENSCAFG00000031920	N/A	0.00064069	0.57128844
ENSCAFG00000024210	N/A	0.00066799	0.57128844
ENSCAFG00000010729	MRGPRF	0.00067789	0.57128844
ENSCAFG00000001416	GLDC	0.00071138	0.57128844
ENSCAFG00000010839	LYPLAL1	0.00071388	0.57128844
ENSCAFG00000028856	N/A	0.00077641	0.57128844

Table 2-S8. Polygenetic signals of selection in maned wolf's biological pathways. For each pathway and the branch of the maned wolf, the ΔlnL4 values of the genes in different sets were calculated (Methods). The significance of each score was compared against a null distribution of random gene sets of the same size. The most significant pathway in the maned wolf was Butanoate metabolism (bolded), which is related to the energy intake from fruit fiber.

setID	setSize	setScore	setP	setQ	setName
19	**20**	**10.0479332**	**0.000497**	**0.01060714**	**Butanoate metabolism**
27	14	4.6603052	0.09393177	0.71557709	regulation of insulin secretion
73	118	29.3100398	0.0962984	0.71557709	spermatogenesis
39	39	10.9415378	0.09708172	0.71557709	Bile secretion
34	19	5.86964581	0.1050979	0.71557709	Proximal tubule bicarbonate reclamation
42	30	8.25221796	0.14157146	0.76062436	Fat digestion and absorption
63	16	4.64562399	0.16719443	0.76062436	sensory perception of bitter taste
6	19	5.28137777	0.18489384	0.76062436	Inositol phosphate metabolism,
24	22	5.8305167	0.21159295	0.77857962	Galactose metabolism
14	10	2.70179973	0.27463627	0.8902873	Pentose phosphate pathway
36	34	8.0539239	0.30118494	0.8902873	Salivary secretion
17	19	4.46382865	0.34743263	0.8902873	Propanoate metabolism
18	10	2.33599555	0.36848158	0.8902873	Pyruvate metabolism
31	30	6.32340062	0.46627669	0.8902873	Vasopressin-regulated water reabsorption
16	10	2.02903521	0.47887606	0.8902873	Pentose phosphate pathway
11	19	3.95067003	0.48037598	0.8902873	Citrate cycle (TCA cycle)
41	50	10.4343825	0.49042548	0.8902873	Protein digestion and absorption
9	28	5.79945465	0.49242538	0.8902873	Amino sugar and nucleotide sugar metabolism
23	21	4.26092895	0.50514949	0.8902873	Valine, leucine and isoleucine degradation
8	51	10.5694061	0.51124888	0.8902873	Phosphatidylinositol signaling system
37	21	4.09006386	0.55224478	0.90233566	Gastric acid secretion
44	23	4.21779575	0.61633837	0.90233566	Mineral absorption
10	10	1.61389951	0.63333667	0.90233566	Fructose and mannose metabolism
26	53	10.0150271	0.66423358	0.90233566	Vascular smooth muscle contraction
32	15	2.40000638	0.68293171	0.90233566	Aldosterone-regulated sodium reabsorption
25	21	3.22233702	0.77252275	0.90233566	Fatty acid metabolism
3	52	9.03726468	0.77692231	0.90233566	Adipocytokine signaling pathway
20	42	6.94746229	0.7960204	0.90233566	ErbB signaling pathway
38	14	1.86028447	0.79612039	0.90233566	Pancreatic secretion
22	83	14.5817489	0.82381762	0.90233566	mTOR signaling pathway
1	32	4.89266487	0.82881712	0.90233566	Insulin signaling pathway
43	15	1.91973312	0.82891711	0.90233566	Vitamin digestion and absorption

33	20	2.55019202	0.86351365	0.90233566	Endocrine and other factor-regulated calcium reabsorption
2	14	1.56728725	0.86671333	0.90233566	Glycolysis / Gluconeogenesis
35	10	0.90679163	0.87051295	0.90233566	Collecting duct acid secretion
45	16	1.72343607	0.89881012	0.90629679	Melanogenesis

Table 2-S9. Polygenetic signals of selection in the bush dog's biological pathways. For each pathway and the branch of the bush dog, the ΔlnL4 values of the genes in different sets were calculated (Methods). The significance of this score was compared against a null distribution of random gene sets of the same size.

setID	setSize	setScore	setP	setQ	setName
70	15	7.7269093	0.004717	0.15789448	sensory perception of bitter taste
11	20	7.72151245	0.04229965	0.48412229	cholesterol homeostasis
23	11	4.37578306	0.0862819	0.51508114	glutamine metabolic process
4	70	19.8530302	0.13384665	0.51508114	lipid transport
5	24	7.43698281	0.16577086	0.51508114	unsaturated fatty acid metabolic process
52	20	4.91918154	0.43417829	0.84746342	positive regulation of lipid biosynthetic process
56	11	2.65399489	0.4579271	0.84746342	sensory perception of taste
3	41	9.62180346	0.52794721	0.84981589	cellular lipid catabolic process
6	19	2.96092298	0.84091591	0.99339772	regulation of lipid biosynthetic process
32	15	2.03512374	0.87091291	0.99339772	regulation of the fatty acid metabolic process
1	22	2.76604225	0.94650535	0.99339772	regulation of lipid metabolic process
34	19	1.96879644	0.96350365	0.99339772	positive regulation of lipid metabolic process
2	30	2.84398229	0.99440056	0.99339772	fatty acid metabolic process

Table 2-S10. Polygenetic signals of selection in different canids' biological pathways. We calculated the ΔlnL4 values of the genes for each pathway and the branch of the coyote, gray wolf, dhole, and African wild dog (see Methods). The significance of this score was compared against a null distribution of random gene sets of the same size.

Genome	setID	setSize	setScore	setP	setQ	setName
African wild dog	1819	371	131.398984	0.000423	0.31873598	Olfactory transduction
African wild dog	1009	78	31.7263022	0.00783436	0.99706243	Class I MHC mediated antigen processing & presentation
African wild dog	31	17	9.60770042	0.01201426	0.99706243	Butanoate metabolism
African wild dog	1393	13	7.92041778	0.01287945	0.99706243	Pyrimidine metabolism
African wild dog	654	16	9.0217327	0.01419718	0.99706243	Endosomal Sorting Complex Required For Transport (ESCRT)
Dhole	32691	375	87.8833681	0.000493	0.9907233	Olfactory transduction
Dhole	24494	10	6.04757047	0.00109	0.99436894	negative regulation of innate immune response
Dhole	2774	26	9.6684108	0.0052642	0.99436894	Autophagy - other
Dhole	19037	15	6.5676932	0.00614999	0.99436894	aminoglycan catabolic process
Dhole	3443	11	4.896259	0.01316407	0.99436894	Selenocompound metabolism
Gray wolf	1716	75	12.2017602	0.000751	0.45658585	ECM-receptor interaction
Gray wolf	1868	41	8.09818742	0.000984	0.45658585	Type II diabetes mellitus
Gray wolf	666	42	6.90991326	0.00725796	0.89423671	Golgi Associated Vesicle Biogenesis
Gray wolf	1113	18	3.90839539	0.00925635	0.89423671	DNA strand elongation
Gray wolf	1758	28	5.07223123	0.01059789	0.89423671	Mucin type O-glycan biosynthesis

Table 2-S11. Bush dog scores of private alleles (SinglBySeg) at regions flanking 39,704 transcripts. SinglBySeg column represents the proportion of private alleles in bush dogs divided by the average of segregation sites among other canids (see methods for details). The top 20 Candidate genes are shown. The genes related to bone elongation are shown in red.

chromo	Start	End	**SinglBySeg**	Sites Passing	EnsemblID	Gene
chr14	7814773	7815773	3.611111	661	ENSCAFG00000001585	FLNC
chr15	41275699	41276699	2.6	959	ENSCAFG00000007304	**IGF1**
chr31	34355444	34356444	2.6	604	ENSCAFG00000029475	**B3GALT5**
chr33	18571299	18572299	2.51875	755	ENSCAFG00000010823	ZBTB20
chr07	28137504	28138504	2.47619	802	ENSCAFG00000015002	PRRX1
chr03	19850617	19851617	2.363636	965	ENSCAFG00000008302	MEF2C
chr16	27257586	27258586	2.26087	501	ENSCAFG00000006099	DDHD2
chr32	24857235	24858235	2.166667	742	ENSCAFG00000010855	TACR3
chr06	17449879	17450879	2.166667	664	ENSCAFG00000016640	N/A
chr32	24857235	24858235	2.166667	742	ENSCAFG00000010855	TACR3
chr11	21219637	21220637	1.925926	565	ENSCAFG00000000930	AFF4
chr26	30082470	30083470	1.925926	579	ENSCAFG00000014833	DGCR2
chr15	18315887	18316887	1.902439	956	ENSCAFG00000005511	RPGRIP1
chr15	18181960	18182960	1.882759	740	ENSCAFG00000005493	ZNF219
chr11	52518271	52519271	1.833333	546	ENSCAFG00000002281	RECK
chr18	46299919	46300919	1.813953	989	ENSCAFG00000010090	IGF2
chr20	50487161	50488161	1.813953	574	ENSCAFG00000017651	DNM2

Table 2-S12. Maned wolf scores of private alleles (SinglBySeg) at 1kb windows flanking 39,704 transcripts. SinglBySeg column represents the proportion of private alleles in the maned wolf divided by the mean of segregation sites among other canids (see methods for details). The top 20 candidate genes are shown. The genes related to bone elongation are shown in red.

chromo	Start	End	SinglBySeg	Sites Passing	EnsemblID	Gene
chr20	8722345	8723345	3.25	998	ENSCAFG00000005486	SETD5
chr12	23936471	23937471	2.666667	750	ENSCAFG00000002413	DST
chr03	64867819	64868819	2.409756	532	ENSCAFG00000015345	CC2D2A
chr14	40348900	40349900	2.166667	616	ENSCAFG00000002963	HOXA10
chr11	3505713	3506713	2	767	ENSCAFG00000000231	**B4GALT7**
chr02	21120659	21121659	1.890909	579	ENSCAFG00000004656	ITGA8
chr17	39418364	39419364	1.772727	949	ENSCAFG00000007658	SFTPB
chr30	38284938	38285938	1.733333	584	ENSCAFG00000018010	SIN3A
chr12	511945	512945	1.666667	584	ENSCAFG00000000443	N/A
chr23	1562844	1563844	1.659574	597	ENSCAFG00000004389	ABHD12
chr34	33807613	33808613	1.63913	961	ENSCAFG00000014621	MECOM
chr04	71697274	71698274	1.625	761	ENSCAFG00000018695	NIPBL
chr01	116756289	116757289	1.591837	709	ENSCAFG00000006697	N/A
chr10	308502	309502	1.5	986	ENSCAFG00000000080	PMEL
chr26	22725593	22726593	1.493243	654	ENSCAFG00000012243	NEFH
chr32	12313144	12314144	1.474627	950	ENSCAFG00000009840	FAM13A
chr11	21218186	21219186	1.4625	828	ENSCAFG00000000930	AFF4
chr12	2954835	2955835	1.460674	907	ENSCAFG00000001022	BAK1

Table 2-S13. Candidate genes for positive selection in the bush dog, as identified by the branch-site test in PAML4 (Yang, 2007). Q values were determinate after a multiple hypothesis-testing correction of one foreground branch (bush dog) and 1,384 genes associates to limb development. The top 10 genes are shown. The gene associated with limb elongation is shown in bold.

Ensmebl_ID	GeneSymbol	Statistic	Pvalue	qvalue
ENSCAFG00000000231	**B4GALT7**	11.944604	0.00054806	0.44392522
ENSCAFG00000029971	OTUD6B	10.925502	0.0009485	0.44392522
ENSCAFG00000000678	ZFPM2	10.42446	0.00124357	0.44392522
ENSCAFG00000019791	SLC6A17	10.131658	0.00145744	0.44392522
ENSCAFG00000011087	DKK2	9.446112	0.00211597	0.44392522
ENSCAFG00000031248	FAM111A	9.352802	0.00222645	0.44392522
ENSCAFG00000018794	DNAJC21	9.336036	0.00224691	0.44392522
ENSCAFG00000020074	N/A	9.022966	0.00266608	0.46089927
ENSCAFG00000018557	ADAMTS10	8.540634	0.00347305	0.51097231
ENSCAFG00000017717	BBS4	8.428098	0.00369467	0.51097231
ENSCAFG00000008199	FMN1	7.748294	0.00537633	0.61962215
ENSCAFG00000008199	FMN1	7.748294	0.00537633	0.61962215
ENSCAFG00000011318	ANOS1	7.017254	0.00807279	0.85882109
ENSCAFG00000020303	CDH3	4.40933	0.03574288	1
ENSCAFG00000009983	ZNF423	3.339578	0.06763195	1
ENSCAFG00000001179	WNT3A	3.293002	0.06957571	1
ENSCAFG00000015654	PAPSS2	3.07337	0.07958434	1
ENSCAFG00000012877	WDR73	3.04937	0.08076847	1
ENSCAFG00000005190	SEPTIN9	2.910724	0.08799237	1
ENSCAFG00000011801	PHACTR4	2.854708	0.09110678	1

Table 2-S14. Set of fossil calibration priors (minimum and maximum node ages) used for the MCMCTree analyses, as inferred by PAML (Yang, 2007). The divergence time estimates are shown in Figure 2-1 and 2-2a.

Node on tree	Node	Node minimum age(Ma)	Node maximum age(Ma)	Evidence	Taxon Search	Number of studies	Prior model
1	Canini - Vulpini split	7.8	16.9	(Eizirik et al., 2010); TimeTree.org	*Canis* versus *Urocyon+Vulpes*/*Lycaon* versus *Urocyon+Vulpes*	17	Lognormal
2	TMRCA crown Cerdocyonina	4.5	9.6	TimeTree.org	*Lycaon* versus *Chrysocyon*	6	Lognormal
3	TMRCA crown *Lycalopex*	2.51	5.92	TimeTree.org	*Cerdocyon* (+*Atelocynus*) versus *Lycalopex*	6	Lognormal
4	*Lycalopex culpaeus* - *L. sechurae* split	1.25	2.3	TimeTree.org	*Lycalopex culpaeus* versu *L. sechurae*	7	Lognormal

References

AL-QATTAN, M. M. 2014. Formation of normal interdigital web spaces in the hand revisited: implications for the pathogenesis of syndactyly in humans and experimental animals. *Journal of Hand Surgery-European Volume,* 39**,** 491-498.

ARAGONA, M. & SETZ, E. Z. F. 2001. Diet of the maned wolf, Chrysocyon brachyurus (Mammalia : Canidae), during wet and dry seasons at Ibitipoca State Park, Brazil. *Journal of Zoology,* 254**,** 131-136.

AUSTIN, J. J., SOUBRIER, J., PREVOSTI, F. J., PRATES, L., TREJO, V., MENA, F. & COOPER, A. 2013. The origins of the enigmatic Falkland Islands wolf. *Nature Communications,* 4.

AUTON, A., LI, Y. R., KIDD, J., OLIVEIRA, K., NADEL, J., HOLLOWAY, J. K., HAYWARD, J. J., COHEN, P. E., GREALLY, J. M., WANG, J., BUSTAMANTE, C. D. & BOYKO, A. R. 2013. Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. *Plos Genetics,* 9.

BACON, C. D., SILVESTRO, D., JARAMILLO, C., SMITH, B. T., CHAKRABARTY, P. & ANTONELLI, A. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama (vol 112, pg 6110, 2015). *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** E3631-E3631.

BEERLING, D. J. & MAYLE, F. E. 2006. Contrasting effects of climate and CO2 on Amazonian ecosystems since the last glacial maximum. *Global Change Biology,* 12**,** 1977-1984.

BEICHMAN, A. C., PHUNG, T. N. & LOHMUELLER, K. E. 2017. Comparison of Single Genome and Allele Frequency Data Reveals Discordant Demographic Histories. *G3-Genes Genomes Genetics,* 7**,** 3605-3620.

BERTA, A. 1978. South American Carnivora. *In:* MARSHALL, L. G. (ed.). Fossilium Catalogus 1. Animalia pt.

BERTA, A. 1981. Evolution of large canids in South America. Porto Alegre: II Congresso Latino Americano de Paleontología.

BERTA, A. 1984. The Pleistocene Bush Dog *Speothos pacivorus* (Canidae) from the Lagoa Santa Caves, Brazil. *Journal of Mammalogy,* 65**,** 549-559.

BERTA, A. 1987. *Origin, diversification, and zoogeography of the South American Canidae,* Fieldiana. Zoology[FIELDIANA ZOOL.].

BERTA, A. 1988. Quaternary evolution and biogeography of the large South American Canidae (Mammalia, Carnivora) Univ of California Press.

BUBADUE, J. D., CACERES, N., CARVALHO, R. D. & MELORO, C. 2016. Ecogeographical Variation in Skull Shape of South-American Canids: Abiotic or Biotic Processes? *Evolutionary Biology,* 43**,** 145-159.

BUENO, A. D. & MOTTA, J. C. 2004. Food habits of two syntopic canids, the maned wolf (*Chrysocyon brachyurus*) and the crab-eating fox (*Cerdocyon thous*), in southeastern Brazil. *Revista Chilena De Historia Natural,* 77**,** 5-14.

CARRIGAN, M. A. 2017. Hominids adapted to metabolize ethanol long before human-directed fermentation. *American Journal of Physical Anthropology,* 162**,** 138-138.

CARTAULT, F., MUNIER, P., JACQUEMONT, M. L., VELLAYOUDOM, J., DORAY, B., PAYET, C., RANDRIANAIVO, H., LAVILLE, J. M., MUNNICH, A. & CORMIER-DAIRE, V. 2015. Expanding the clinical spectrum of B4GALT7 deficiency: homozygous p.R270C mutation with founder effect causes Larsen of Reunion Island syndrome. *European Journal of Human Genetics,* 23**,** 49-53.

CASTELLÓ, J. R. 2018. Canids of the World: Wolves, Wild Dogs, Foxes, Jackals, Coyotes, and Their Relatives. Princeton University Press.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution,* 17**,** 540-552.

CHIKHI, L., RODRIGUEZ, W., GRUSEA, S., SANTOS, P., BOITARD, S. & MAZET, O. 2018. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. *Heredity,* 120**,** 13-24.

CLAPPERTON, C. M. 1993. Nature of environmental changes in South America at the Last Glacial Maximum. *Palaeogeography Palaeoclimatology Palaeoecology,* 101**,** 189-208.

CLERICI, M., KALLMANN, C., GASPI, F. O. G., MORGANO, M. A., MARTINEZ-BUSTOS, F. & CHANG, Y. K. 2011. Physical, chemical and technological characteristics of *Solanum lycocarpum* A. St. - HILL (Solanaceae) fruit flour and starch. *Food Research International,* 44**,** 2143-2150.

COOPER, K. L., OH, S., SUNG, Y., DASARI, R. R., KIRSCHNER, M. W. & TABIN, C. J. 2013. Multiple phases of chondrocyte enlargement underlie differences in skeletal proportions. *Nature,* 495**,** 375-378.

CORTES, M., BARIA, A. T. & SCHWARTZ, N. B. 2009. Sulfation of chondroitin sulfate proteoglycans is necessary for proper Indian hedgehog signaling in the developing growth plate. *Development,* 136**,** 1697-1706.

COSSIOS, D., LUCHERINI, M., RUIZ-GARCIA, M. & ANGERS, B. 2009. Influence of ancient glacial periods on the Andean fauna: the case of the pampas cat (*Leopardus colocolo*). *Bmc Evolutionary Biology,* 9.

COSSÍOS, E. D. 2010. *Lycalopex sechurae* (Carnivora: Canidae). *Mammalian Species,* 42**,** 1-6.

CRESPO, J. A. 1975. Ecology of the pampas gray fox and the large fox (culpeo). *In:* FOX, M. W. (ed.) *The wild canids.* New York: Van Nostrand Reinhold Company.

CRETEKOS, C. J., WANG, Y., GREEN, E. D., MARTIN, J. F., RASWEILER, J. J., BEHRINGER, R. R. & PROGRA, N. C. S. 2008. Regulatory divergence modifies limb length between mammals. *Genes & Development,* 22**,** 141-151.

DASMAHAPATRA, K. K., WALTERS, J. R., BRISCOE, A. D., DAVEY, J. W., WHIBLEY, A., NADEAU, N. J., ZIMIN, A. V., HUGHES, D. S. T., FERGUSON, L. C., MARTIN, S. H., SALAZAR, C., LEWIS, J. J., ADLER, S., AHN, S. J., BAKER, D. A., BAXTER, S. W., CHAMBERLAIN, N. L., CHAUHAN, R., COUNTERMAN, B. A., DALMAY, T., GILBERT, L. E., GORDON, K., HECKEL, D. G., HINES, H. M., HOFF, K. J., HOLLAND, P. W. H., JACQUIN-JOLY, E., JIGGINS, F. M., JONES, R. T., KAPAN, D. D., KERSEY, P., LAMAS, G., LAWSON, D., MAPLESON, D., MAROJA, L. S., MARTIN, A., MOXON, S., PALMER, W. J., PAPA, R., PAPANICOLAOU, A., PAUCHET, Y., RAY, D. A., ROSSER, N., SALZBERG, S. L., SUPPLE, M. A., SURRIDGE, A., TENGER-TROLANDER, A., VOGEL, H., WILKINSON, P. A., WILSON, D., YORKE, J. A., YUAN, F. R., BALMUTH, A. L., ELAND, C., GHARBI, K., THOMSON, M., GIBBS, R. A., HAN, Y., JAYASEELAN, J. C., KOVAR, C., MATHEW, T., MUZNY, D. M., ONGERI, F., PU, L. L., QU, J. X., THORNTON, R. L., WORLEY, K. C., WU, Y. Q., LINARES, M., BLAXTER, M. L., FFRENCH-CONSTANT, R. H., JORON, M., KRONFORST, M. R., MULLEN, S. P., REED, R. D., SCHERER, S. E., RICHARDS, S., MALLET, J., MCMILLAN, W. O., JIGGINS, C. D. & HELICONIUS GENOME, C. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature,* 487**,** 94-98.

DAUB, J. T., MORETTI, S., DAVYDOV, II, EXCOFFIER, L. & ROBINSON-RECHAVI, M. 2017. Detection of Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. *Molecular Biology and Evolution,* 34**,** 1391-1402.

DEN BESTEN, G., VAN EUNEN, K., GROEN, A. K., VENEMA, K., REIJNGOUD, D. J. & BAKKER, B. M. 2013. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of Lipid Research,* 54**,** 2325-2340.

DEUTSCH, L. A. 1983. An Encounter between bush dog (*Speothos venaticus*) and paca (*Agouti paca*). *Journal of Mammalogy,* 64**,** 532-533.

DIETZ, J. M. 1984. Ecology and social organization of the maned wolf. 392.

DIETZ, J. M. 1985. *Chrysocyon brachyurus*. 234.

DILLEHAY, T. D., OCAMPO, C., SAAVEDRA, J., SAWAKUCHI, A. O., VEGA, R. M., PINO, M., COLLINS, M. B., CUMMINGS, L. S., ARREGUI, I., VILLAGRAN, X. S., HARTMANN, G. A., MELLA, M., GONZALEZ, A. & DIX, G. 2015. New Archaeological Evidence for an Early Human Presence at Monte Verde, Chile. *Plos One,* 10.

EIZIRIK, E. 2012. A molecular view on the evolutionary history and biogeography of Neotropical carnivores (Mammalia, Carnivora). In: Patterson B, Costa L (eds) Bones, clones and Biomes. Chicago: The University of
Chicago Press.

EIZIRIK, E., BONATTO, S. L., JOHNSON, W. E., CRAWSHAW, P. G., VIE, J. C., BROUSSET, D. M., O'BRIEN, S. J. & SALZANO, F. M. 1998. Phylogeographic patterns and evolution of the mitochondrial DNA control region in two neotropical cats (Mammalia, Felidae). *Journal of Molecular Evolution,* 47**,** 613-624.

EIZIRIK, E., MURPHY, W. J., KOEPFLI, K. P., JOHNSON, W. E., DRAGOO, J. W., WAYNE, R. K. & O'BRIEN, S. J. 2010. Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. *Molecular Phylogenetics and Evolution,* 56**,** 49-63.

EVA, H. D., BELWARD, A. S., DE MIRANDA, E. E., DI BELLA, C. M., GOND, V., HUBER, O., JONES, S., SGRENZAROLI, M. & FRITZ, S. 2004. A land cover map of South America. *Global Change Biology,* 10**,** 731-744.

FAN, Z. X., SILVA, P., GRONAU, I., WANG, S. G., ARMERO, A. S., SCHWEIZER, R. M., RAMIREZ, O., POLLINGER, J., GALAVERNI, M., DEL-VECCHYO, D. O., DU, L. M., ZHANG, W. P., ZHANG, Z. H., XING, J. C., VILA, C., MARQUES-BONET, T., GODINHO, R., YUE, B. S. & WAYNE, R. K. 2016. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Research,* 26**,** 163-173.

FARIAS, A. A., SEPULVEDA, M. A., SILVA-RODRIGUEZ, E. A., EGUREN, A., GONZALEZ, D., JORDAN, N. I., OVANDO, E., STOWHAS, P. & SVENSSON, G. L. 2014. A new population of Darwin's fox (*Lycalopex fulvipes*) in the Valdivian Coastal Range. *Revista Chilena De Historia Natural,* 87.

FAVARINI, M. O. 2011. *Relações filogenéticas entre espécies do gênero Lycalopex (Mammalia, Canidae) inferidas com o uso de marcadores do DNA mitocondrial.* Master's thesis, Pontifícia Universidade Católica do Rio Grande do Sul.

FREEDMAN, A. H., GRONAU, I., SCHWEIZER, R. M., ORTEGA-DEL VECCHYO, D., HAN, E. J., SILVA, P. M., GALAVERNI, M., FAN, Z. X., MARX, P., LORENTE-GALDOS, B., BEALE, H., RAMIREZ, O., HORMOZDIARI, F., ALKAN, C., VILA, C., SQUIRE, K., GEFFEN, E., KUSAK, J., BOYKO, A. R., PARKER, H. G., LEE, C., TADIGOTLA, V., SIEPEL, A., BUSTAMANTE, C. D., HARKINS, T. T., NELSON, S. F., OSTRANDER, E. A., MARQUES-BONET, T., WAYNE, R. K. & NOVEMBRE, J. 2014. Genome Sequencing Highlights the Dynamic Early History of Dogs. *Plos Genetics,* 10.

FUENTES, E. R. & JAKSIC, F. M. 1979. Latitudinal size variation of Chilean foxes: tests of alternative hypotheses. *Ecology,* 60**,** 43-47.

GARCEZ, F. S. 2015. *Filogeografia E História Populacional De Lycalopex vetulus (Carnivora, Canidae), incluindo Sua Hibridacão Com L. Gymnocercus.* Master's thesis, Pontifícia Universidade Católica do Rio Grande do Sul.

GONZÁLEZ DEL SOLAR, R. & RAU, J. 2004. Chilla. *Pseudalopex griseus*. *Canids: Foxes, Wolves, Jackals and Dogs. Status Survey and Conservation Action Plan***,** 56-63.

GOPALAKRISHNAN, S., SINDING, M. H. S., RAMOS-MADRIGAL, J., NIEMANN, J., CASTRUITA, J. A. S., VIEIRA, F. G., CAROE, C., MONTERO, M. D., KUDERNA, L., SERRES, A., GONZALEZ-BASALLOTE, V. M., LIU, Y. H., WANG, G. D., MARQUES-BONET, T., MIRARAB, S., FERNANDES, C., GAUBERT, P., KOEPFLI, K. P., BUDD, J., RUENESS, E. K., HEIDE-JORGENSEN, M. P., PETERSEN, B., SICHERITZ-PONTEN, T., BACHMANN, L., WIIG, O., HANSEN, A. J. & GILBERT, M. T. P. 2018. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. *Current Biology,* 28**,** 3441-3449.

HABERLE, S. G. & MASLIN, M. A. 1999. Late Quaternary vegetation and climate change in the Amazon basin based on a 50,000 year pollen record from the Amazon fan, ODP site 932. *Quaternary Research,* 51**,** 27-38.

HAFFER, J. 1969. Speciation in Amazonian forest birds. *Science,* 165**,** 131-&.

HAGGI, C., CHIESSI, C. M., MERKEL, U., MULITZA, S., PRANGE, M., SCHULZ, M. & SCHEFUSS, E. 2017. Response of the Amazon rainforest to late Pleistocene climate variability. *Earth and Planetary Science Letters,* 479**,** 50-59.

HARRIS, S. E. 1993. Horse gaits, balance, and movement.

HEINE, K. 2000. Tropical South America during the Last Glacial Maximum: evidence from glacial, periglacial and fluvial records. *Quaternary International,* 72**,** 7-21.

HELGEN, K. M., PINTO, C. M., KAYS, R., HELGEN, L. E., TSUCHIYA, M. T. N., QUINN, A., WILSON, D. E. & MALDONADO, J. E. 2013. Taxonomic revision of the olingos (*Bassaricyon*), with description of a new species, the Olinguito. *Zookeys***,** 1-83.

HILDEBRAND, M. 1952. An analysis of body proportions in the Canidae. *American Journal of Anatomy,* 90**,** 217-256.

JERNVALL, J. & THESLEFF, I. 2000. Reiterative signaling and patterning during mammalian tooth morphogenesis. *Mechanisms of Development,* 92**,** 19-29.

JIMENEZ, J. E., YANEZ, J. L., TABILO, E. L. & JAKSIC, F. M. 1995. Body size of Chilean foxes: A new pattern in light of new data. *Acta Theriologica,* 40**,** 321-326.

JIMÉNEZ, J. E. & MCMAHON, E. 2015. *Pseudalopex fulvipes*. *In:* SILLERO-ZUBIRI, C., HOFFMANN, M. & MACDONALD, D. W. (eds.) *Canids: Foxes, wolves, jackals and dogs.* Gland, Switzerland and Cambridge, UK: Status survey and conservation action plan, IUCN/SSC Canid Specialist Group.

KAISER, J., LAMY, F., ARZ, H. W. & HEBBELN, D. 2007. Dynamics of the millennial-scale sea surface temperature and Patagonian Ice Sheet fluctuations in southern Chile during the last 70kyr (ODP Site 1233). *Quaternary International,* 161**,** 77-89.

KOEPFLI, K. P., GOMPPER, M. E., EIZIRIK, E., HO, C. C., LINDEN, L., MALDONADO, J. E. & WAYNE, R. K. 2007. Phylogeny of the Procyonidae (Mammalia : Carnivora): Molecules, morphology and the Great American Interchange. *Molecular Phylogenetics and Evolution,* 43**,** 1076-1095.

KOEPFLI, K. P., POLLINGER, J., GODINHO, R., ROBINSON, J., LEA, A., HENDRICKS, S., SCHWEIZER, R. M., THALMANN, O., SILVA, P., FAN, Z. X., YURCHENKO, A. A., DOBRYNIN, P., MAKUNIN, A., CAHILL, J. A., SHAPIRO, B., ALVARES, F., BRITO, J. C., GEFFEN, E., LEONARD, J. A., HELGEN, K. M., JOHNSON, W. E., O'BRIEN, S. J., VAN VALKENBURGH, B. & WAYNE, R. K. 2015. Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Current Biology,* 25**,** 2158-2165.

KUKEKOVA, A. V., JOHNSON, J. L., XIANG, X. Y., SHAOHONG, F. H., LIU, S. P., RANDO, H. M., KHARLAMOVA, A. V., HERBECK, Y., SERDYUKOVA, N. A., XIONG, Z. J., BEKLEMISCHEVA, V., KOEPFLI, K. P., GULEVICH, R. G., VLADIMIROVA, A. V., HEKMAN, J. P., PERELMAN, P. L., GRAPHODATSKY, A. S., O'BRIEN, S. J., WANG, X., CLARK, A. G., ACLAND, G. M., TRUT, L. N. & ZHANG, G. J. 2018. Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nature Ecology & Evolution,* 2**,** 1479-1491.

LAMICHHANEY, S., BERGLUND, J., ALMEN, M. S., MAQBOOL, K., GRABHERR, M., MARTINEZ-BARRIO, A., PROMEROVA, M., RUBIN, C. J., WANG, C., ZAMANI, N., GRANT, B. R., GRANT, P. R., WEBSTER, M. T. & ANDERSSON, L. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature,* 518**,** 371-375.

LANGGUTH, A. 1975. Ecology and evolution in the South American canids. *The Wild Canids.* New York: Litton Educational Publishing.

LEEGWATER, P. A., VOS-LOOHUIS, M., DUCRO, B. J., BOEGHEIM, I. J., VAN STEENBEEK, F. G., NIJMAN, I. J., MONROE, G. R., BASTIAANSEN, J. W. M., DIBBITS, B. W., VAN DE GOOR, L. H., HELLINGA, I., BACK, W. & SCHURINK, A. 2016. Dwarfism with joint laxity in Friesian horses is associated with a splice site mutation in B4GALT7. *Bmc Genomics,* 17.

LI, F. F., HAN, Y., SHI, S. A., LI, X., ZHU, X. D., ZHOU, J., SHAO, Q. L., LI, X. Q. & LIU, S. L. 2015. Characterization of Transcriptional Repressor Gene *MSX1* Variations for Possible Associations with Congenital Heart Diseases. *Plos One,* 10.

LI, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

LIMA, E. D., JORGE, R. S. P. & DALPONTE, J. C. 2009. Habitat use and diet of bush dogs, Speothos venaticus, in the Northern Pantanal, Mato Grosso, Brazil. *Mammalia,* 73**,** 13-19.

LIMA-REZENDE, C. A., ROCHA, A. V., COUTO, A. F., MARTINS, E. D., VASCONCELOS, V. & CAPARROZ, R. 2019. Late Pleistocene climatic changes promoted demographic expansion and population reconnection of a Neotropical savanna-adapted bird, *Neothraupis fasciata* (Aves: Thraupidae). *Plos One,* 14.

LINDAHL, U., COUCHMAN, J., KIMATA, K. & JEFFREY, D. 2017. Proteoglycans and Sulfated Glycosaminoglycans. *In:* VARKI, A., CUMMINGS, R. D. & ESKO J, D. (eds.) *Essentials of Glycobiology* 3rd ed.: Cold Spring Harbor Laboratory Press; 2015-2017.

LINDBLAD-TOH, K., WADE, C. M., MIKKELSEN, T. S., KARLSSON, E. K., JAFFE, D. B., KAMAL, M., CLAMP, M., CHANG, J. L., KULBOKAS, E. J., ZODY, M. C., MAUCELI, E., XIE, X. H., BREEN, M., WAYNE, R. K., OSTRANDER, E. A., PONTING, C. P., GALIBERT, F., SMITH, D. R., DEJONG, P. J., KIRKNESS, E., ALVAREZ, P., BIAGI, T., BROCKMAN, W., BUTLER, J., CHIN, C. W., COOK, A., CUFF, J., DALY, M. J., DECAPRIO, D., GNERRE, S., GRABHERR, M., KELLIS, M., KLEBER, M., BARDELEBEN, C., GOODSTADT, L., HEGER, A., HITTE, C., KIM, L., KOEPFLI, K. P., PARKER, H. G., POLLINGER, J. P., SEARLE, S. M. J., SUTTER, N. B., THOMAS, R., WEBBER, C., LANDER, E. S. & BROAD INST GENOME SEQUENCING, P. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature,* 438**,** 803-819.

LOHMUELLER, K. E., INDAP, A. R., SCHMIDT, S., BOYKO, A. R., HERNANDEZ, R. D., HUBISZ, M. J., SNINSKY, J. J., WHITE, T. J., SUNYAEV, S. R., NIELSEN, R., CLARK, A. G. & BUSTAMANTE, C. D. 2008. Proportionally more deleterious genetic variation in European than in African populations. *Nature,* 451**,** 994-U5.

MARSDEN, C. D., ORTEGA-DEL VECCHYO, D., O'BRIEN, D. P., TAYLOR, J. F., RAMIREZ, O., VILA, C., MARQUES-BONET, T., SCHNABEL, R. D., WAYNE, R. K. & LOHMUELLER, K. E. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** 152-157.

MARSHALL, L. G. 1977. Evolution of the Carnivorous Adaptive Zone in South America. *Major Patterns in Vertebrate Evolution.* Boston, MA: Springer US.

MATZKE, N. J. 2013. BioGeoBEARS: biogeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. Berkeley, CA: University of California, Berkeley.

MAZET, O., RODRIGUEZ, W., GRUSEA, S., BOITARD, S. & CHIKHI, L. 2016. On the importance of being structured: instantaneous coalescence rates and human evolution-lessons for ancestral population size inference? *Heredity,* 116**,** 362-371.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. &

DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20**,** 1297-1303.

MCLAREN, W., GIL, L., HUNT, S. E., RIAT, H. S., RITCHIE, G. R. S., THORMANN, A., FLICEK, P. & CUNNINGHAM, F. 2016. The Ensembl Variant Effect Predictor. *Genome Biology,* 17.

MIRANDA, A., ALTAMIRANO, A., CAYUELA, L., PINCHEIRA, F. & LARA, A. 2015. Different times, same story: Native forest loss and landscape homogenization in three physiographical areas of south-central of Chile. *Applied Geography,* 60**,** 20-28.

MORA, A., BABY, P., RODDAZ, M., PARRA, M., BRUSSET, S., HERMOZA, W. & ESPURT, N. 2010. Tectonic history of the Andes and sub-Andean zones: implications for the development of the Amazon drainage basin. *Amazonia: Landscape and Species Evolution: a Look into the Past***,** 38-60.

MORENO, P. I., VILLAGRAN, C., MARQUET, P. A. & MARSHALL, L. G. 1994. Quaternary Paleobiogeography of Northern and Central Chile. *Revista Chilena De Historia Natural,* 67**,** 487-502.

NAHUELHUAL, L., DONOSO, P., LARA, A., NÚÑEZ, D., OYARZÚN, C. & NEIRA, E. 2007. Valuing ecosystem services of Chilean temperate rainforests. *Environment, Development and Sustainability,* 9**,** 481-499.

O'DEA, A., LESSIOS, H. A., COATES, A. G., EYTAN, R. I., RESTREPO-MORENO, S. A., CIONE, A. L., COLLINS, L. S., DE QUEIROZ, A., FARRIS, D. W., NORRIS, R. D., STALLARD, R. F., WOODBURNE, M. O., AGUILERA, O., AUBRY, M. P., BERGGREN, W. A., BUDD, A. F., COZZUOL, M. A., COPPARD, S. E., DUQUE-CARO, H., FINNEGAN, S., GASPARINI, G. M., GROSSMAN, E. L., JOHNSON, K. G., KEIGWIN, L. D., KNOWLTON, N., LEIGH, E. G., LEONARD-PINGEL, J. S., MARKO, P. B., PYENSON, N. D., RACHELLO-DOLMEN, P. G., SOIBELZON, E., SOIBELZON, L., TODD, J. A., VERMEIJ, G. J. & JACKSON, J. B. C. 2016. Formation of the Isthmus of Panama. *Science Advances,* 2.

OKONECHNIKOV, K., CONESA, A. & GARCIA-ALCALDE, F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics,* 32**,** 292-294.

OSTERWALDER, M., BAROZZI, I., TISSIERES, V., FUKUDA-YUZAWA, Y., MANNION, B. J., AFZAL, S. Y., LEE, E. A., ZHU, Y. W., PLAJZER-FRICK, I., PICKLE, C. S., KATO, M., GARVIN, T. H., PHAM, Q. T., HARRINGTON, A. N., AKIYAMA, J. A., AFZAL, V., LOPEZ-RIOS, J., DICKEL, D. E., VISEL, A. & PENNACCHIO, L. A. 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature,* 554**,** 239-+.

OTAVO, S. & ECHEVERRIA, C. 2017. Fragmentación progresiva y pérdida de hábitat de bosques naturales en uno de los hotspot mundiales de biodiversidad. *Revista Mexicana de Biodiversidad,* 88.

PATTERSON, B., D., SOLARI, S. & VELAZCO, P. L., M. 2012. The Role of the Andes in the Diversification and Biogeography of Neotropical Mammals. *Bones, Clones, and Biomes : The History and Geography of Recent Neotropical Mammals.* University of Chicago Press.

PEREIRA, A. P. A., ANGOLINI, C. F. F., PAULINO, B. N., LAURETTI, L. B. C., ORLANDO, E. A., SILVA, J. G. S., NERI-NUMA, I. A., SOUZA, J., PALLONE, J. A. L., EBERLIN, M. N. & PASTORE, G. M. 2019. A comprehensive characterization of *Solanum lycocarpum* St. Hill and *Solanum oocarpum* Sendtn: Chemical composition and antioxidant properties. *Food Research International,* 124**,** 61-69.

PERINI, F. A., RUSSO, C. A. M. & SCHRAGO, C. G. 2010. The evolution of South American endemic canids: a history of rapid diversification and morphological parallelism. *Journal of Evolutionary Biology,* 23**,** 311-322.

PERRI, A. R., MITCHELL, K. J., MOUTON, A., ALVAREZ-CARRETERO, S., HULME-BEAMAN, A., HAILE, J., JAMIESON, A., MEACHEN, J., LIN, A. T., SCHUBERT, B. W., AMEEN, C., ANTIPINA, E. E., BOVER, P., BRACE, S., CARMAGNINI, A., CAROE, C., CASTRUITA, J. A. S., CHATTERS, J. C., DOBNEY, K., DOS REIS, M., EVIN, A., GAUBERT, P., GOPALAKRISHNAN, S., GOWER, G., HEINIGER, H., HELGEN, K. M., KAPP, J., KOSINTSEV, P. A., LINDERHOLM, A., OZGA, A. T., PRESSLEE, S., SALIS, A. T., SAREMI, N. F., SHEW, C., SKERRY, K., TARANENKO, D. E., THOMPSON, M., SABLIN, M. V., KUZMIN, Y. V., COLLINS, M. J., SINDING, M. H. S., GILBERT, M. T. P., STONE, A. C., SHAPIRO, B., VAN VALKENBURGH, B., WAYNE, R. K., LARSON, G., COOPER, A. & FRANTZ, L. A. F. 2021. Dire wolves were the last of an ancient New World canid lineage. *Nature,* 591**,** 87-+.

PLASSAIS, J., KIM, J., DAVIS, B. W., KARYADI, D. M., HOGAN, A. N., HARRIS, A. C., DECKER, B., PARKER, H. G. & OSTRANDER, E. A. 2019. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications,* 10.

PREVOSTI, F. J. 2010. Phylogeny of the large extinct South American Canids (Mammalia, Carnivora, Canidae) using a "total evidence" approach. *Cladistics,* 26**,** 456-481.

PREVOSTI, F. J. & FORASIEPI, A. M. 2018. South American Fossil Carnivorans (Order Carnivora). *Evolution of South American Mammalian Predators During the Cenozoic: Paleobiogeographic and Paleoenvironmental Contingencies***,** 85-136.

PREVOSTI, F. J., SEGURA, V., CASSINI, G. & MARTIN, G. M. 2013. Revision of the systematic status of patagonian and pampean gray foxes (Canidae: *Lycalopex griseus* and *L. gymnocercus)* using 3D geometric morphometrics. 20.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A. R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. W., DALY, M. J. & SHAM, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics,* 81**,** 559-575.

RAHIM, M., CHILLOUX, J., MARTINEZ-GILI, L., NEVES, A. L., MYRIDAKIS, A., GOODERHAM, N. & DUMAS, M. E. 2019. Diet-induced metabolic changes of the human gut microbiome: importance of short-chain fatty acids, methylamines and indoles. *Acta Diabetologica,* 56**,** 493-500.

RATZKA, A., KALUS, I., MOSER, M., DIERKS, T., MUNDLOS, S. & VORTKAMP, A. 2008. Redundant function of the heparan sulfate 6-O-endosulfatases Sulf1 and Sulf2 during skeletal development. *Developmental Dynamics,* 237**,** 339-353.

RITELLI, M., DORDONI, C., CINQUINA, V., VENTURINI, M., CALZAVARA-PINTON, P. & COLOMBI, M. 2017. Expanding the clinical and mutational spectrum of B4GALT7-spondylodysplastic Ehlers-Danlos syndrome. *Orphanet Journal of Rare Diseases,* 12.

ROBINSON, J. A., ORTEGA-DEL VECCHYO, D., FAN, Z. X., KIM, B. Y., VONHOLDT, B. M., MARSDEN, C. D., LOHMUELLER, K. E. & WAYNE, R. K. 2016. Genomic Flatlining in the Endangered Island Fox. *Current Biology,* 26**,** 1183-1189.

ROBINSON, J. A., RAIKKONEN, J., VUCETICH, L. M., VUCETICH, J. A., PETERSON, R. O., LOHMUELLER, K. E. & WAYNE, R. K. 2019. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances,* 5.

ROLIAN, C. 2020. Endochondral ossification and the evolution of limb proportions. *Wiley Interdisciplinary Reviews-Developmental Biology,* 9.

SCHIAFFINI, M. I., SEGURA, V. & PREVOSTI, F. J. 2019. Geographic variation in skull shape and size of the Pampas fox *Lycalopex gymnocercus* (Carnivora: Canidae) in Argentina. *Mammalian Biology,* 97**,** 50-58.

SCHIFFELS, S. & DURBIN, R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics,* 46**,** 919-925.

SEARS, K. E., PATEL, A., HUBLER, M., CAO, X. Y., VANDEBERG, J. L. & ZHONG, S. 2012. Disparate Igf1 Expression and Growth in the Fore- and Hind Limbs of a Marsupial Mammal (*Monodelphis domestica*). *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution,* 318B**,** 279-293.

SHELDON, J. W. 1992. The natural history of the nondomestic Canidae. San Diego, CA: Academic Press, Inc.

SILVA-RODRÍGUEZ, E., FARIAS, A., MOREIRA-ARCE, D., CABELLO, J., HIDALGO-HERMOSO, E., LUCHERINI, M. & JIMÉNEZ, J. 2016. *Lycalopex fulvipes (errata version published in 2016). The IUCN Red List of Threatened Species 2016: e.T41586A107263066.* [Online]. https://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T41586A85370871.en . . [Accessed 09 February 2021].

SKOGLUND, P., ERSMARK, E., PALKOPOULOU, E. & DALEN, L. 2015. Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. *Current Biology,* 25**,** 1515-1519.

SLATER, G. J., THALMANN, O., LEONARD, J. A., SCHWEIZER, R. M., KOEPFLI, K. P., POLLINGER, J. P., RAWLENCE, N. J., AUSTIN, J. J., COOPER, A. & WAYNE, R. K. 2009. Evolutionary history of the Falklands wolf. *Current Biology,* 19**,** R937-R938.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-1313.

SUTTER, N. B., BUSTAMANTE, C. D., CHASE, K., GRAY, M. M., ZHAO, K. Y., ZHU, L., PADHUKASAHASRAM, B., KARLINS, E., DAVIS, S., JONES, P. G., QUIGNON, P., JOHNSON, G. S., PARKER, H. G., FRETWELL, N., MOSHER, D. S., LAWLER, D. F., SATYARAJ, E., NORDBORG, M., LARK, K. G., WAYNE, R. K. & OSTRANDER, E. A. 2007. A single IGF1 allele is a major determinant of small size in dogs. *Science,* 316**,** 112-115.

SUTTON, V. R., HYLAND, J. C., PHILLIPS, W. A., SCHLESINGER, A. E. & BRILL, P. W. 2005. A dominantly inherited spondylometaphyseal dysplasia with "corner fractures" and congenital scoliosis. *American Journal of Medical Genetics Part A,* 133A**,** 209-212.

TCHAICKA, L., DE FREITAS, T. R. O., BAGER, A., VIDAL, S. L., LUCHERINI, M., IRIARTE, A., NOVARO, A., GEFFEN, E., GARCEZ, F. S., JOHNSON, W. E., WAYNE, R. K. & EIZIRIK, E. 2016. Molecular assessment of the phylogeny and biogeography of a recently diversified endemic group of South American canids (Mammalia: Carnivora: Canidae). *Genetics and Molecular Biology,* 39**,** 442-451.

TEDFORD, R. H. & QIU, Z. 1991. Pliocene *Nyctereutes* (Carnivora: Canidae) from Yushe, Shanxi, with comments on Chinese fossil Racoon-dogs., 29.

TEDFORD, R. H., WANG, X. & TAYLOR, B. E. 2009. Phylogenetic systematics of the North American fossil Caninae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History,* 325**,** 1-218.

TORRES-ROLDÁN, V. & FERRUSQUÍA-VILLAFRANCA, I. 1981. *Cerdocyon sp. nov.* A. (Mammalia, Carnivora) en México y su significacién evolutiva y zoogeográfica en relación a los cánidos sudamericanos. *II Congresso Latino-Americano Paleontologia.* Porto Alegre.

TRIZNA, M. 2020. assembly_stats 0.1.4. Zenodo.

VAN VALKENBURGH, B. 1987. Skeletal indicators of locomotor behavior in living and extinct carnivores. 7**,** 162-182.

VAN VALKENBURGH, B. 1991. Iterative evolution of hypercarnivory in canids (mammalia, carnivora) - evolutionary interactions among sympatric predators. *Paleobiology,* 17**,** 340-362.

VAN VALKENBURGH, B. & HERTEL, F. 1993. Tough Times at La brea: Tooth Breakage in Large Carnivores of the Late Pleistocene. *Science,* 261**,** 456-459.

VAYSSE, A., RATNAKUMAR, A., DERRIEN, T., AXELSSON, E., PIELBERG, G. R., SIGURDSSON, S., FALL, T., SEPPALA, E. H., HANSEN, M. S. T., LAWLEY, C. T., KARLSSON, E. K., BANNASCH, D., VILA, C., LOHI, H., GALIBERT, F., FREDHOLM, M., HAGGSTROM, J., HEDHAMMAR, A., ANDRE, C., LINDBLAD-TOH, K., HITTE, C., WEBSTER, M. T. & CONSORTIUM, L. 2011. Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. *Plos Genetics,* 7.

VENKAT, A., HAHN, M. W. & THORNTON, J. W. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nature Ecology & Evolution,* 2**,** 1280-1288.

VILA, C., LEONARD, J. A., IRIARTE, A., O'BRIEN, S. J., JOHNSON, W. E. & WAYNE, R. K. 2004. Detecting the vanishing populations of the highly endangered Darwin's fox, *Pseudalopex fulvipes*. *Animal Conservation,* 7**,** 147-153.

VILLAGRAN, C. 1988. Late Quaternary Vegetation of Southern Isla Grande de Chiloé, Chile. *Quaternary Research,* 29**,** 294-306.

VONHOLDT, B. M., CAHILL, J. A., FAN, Z. X., GRONAU, I., ROBINSON, J., POLLINGER, J. P., SHAPIRO, B., WALL, J. & WAYNE, R. K. 2016. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances,* 2.

WALKER, E. P. & PARADISO, L. P. 1975. *Mammals of the world*, Johns Hopkins University Press.

WANG, J., ZHOU, J. & BONDY, C. A. 1999. Igf1 promotes longitudinal bone growth by insulin-like actions augmenting chondrocyte hypertrophy. *Faseb Journal,* 13**,** 1985-1990.

WANG, X., TEDFORD, R. H. & ANTÓN, M. 2008. *Dogs: Their Fossil Relatives and Evolutionary History*, Columbia University Press.

WAYNE, R. K. 1986a. Developmental constraints on limb growth in domestic and some wild canids. *Journal of Zoology,* 210**,** 381-399.

WAYNE, R. K. 1986b. Limb morphology of domestic and wild canids: the influence of development on morphological change. *Journal of Morphology,* 187**,** 301-319.

WAYNE, R. K., GEFFEN, E., GIRMAN, D. J., KOEPFLI, K. P., LAU, L. M. & MARSHALL, C. R. 1997. Molecular systematics of the Canidae. *Systematic Biology,* 46**,** 622-653.

WEATHERBEE, S. D., BEHRINGER, R. R., RASWEILER, J. J. & NISWANDER, L. A. 2006. Interdigital webbing retention in bat wings illustrates genetic changes underlying amniote

limb diversification. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 15103-15107.

WEBB, A. E., WALSH, T. A. & O'CONNELL, M. J. 2017. VESPA: Very large-scale Evolutionary and Selective Pressure Analyses. *PeerJ Computer Science,* 3**,** 1-16.

WEBB, D. S. 1991. Ecography and the Great American Interchange. 3.

WEBB, S. D. 2006. The Great American Biotic Interchange: Patterns and processes. *Annals of the Missouri Botanical Garden,* 93**,** 245-257.

WEBB, S. D. C. F. P. D. 1978. A History of Savanna Vertebrates in the New World. Part II: South America and the Great Interchange. *Annual Review of Ecology and Systematics,* 9**,** 393-426.

WONG, S. W., LIU, H. C., HAN, D., CHANG, H. G., ZHAO, H. S., WANG, Y. X. & FENG, H. L. 2014. A novel non-stop mutation in MSX1 causing autosomal dominant non-syndromic oligodontia. *Mutagenesis,* 29**,** 319-323.

WOODBURNE, M. O. 2010. The Great American Biotic Interchange: Dispersals, Tectonics, Climate, Sea Level and Holding Pens. *Journal of Mammalian Evolution,* 17**,** 245-264.

YAHNKE, C. J., JOHNSON, W. E., GEFFEN, E., SMITH, D., HERTEL, F., ROY, M. S., BONACIC, C. F., FULLER, T. K., VANVALKENBURGH, B. & WAYNE, R. K. 1996. Darwin's fox: A distinct endangered species in a vanishing habitat. *Conservation Biology,* 10**,** 366-375.

YANG, Z. H. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution,* 24**,** 1586-1591.

ZHANG, C., RABIEE, M., SAYYARI, E. & MIRARAB, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *Bmc Bioinformatics,* 19.

ZHANG, J. Z., NIELSEN, R. & YANG, Z. H. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution,* 22**,** 2472-2479.

ZIMIN, A. V., MARCAIS, G., PUIU, D., ROBERTS, M., SALZBERG, S. L. & YORKE, J. A. 2013. The MaSuRCA genome assembler. *Bioinformatics,* 29**,** 2669-2677.

ZUNINO, G. E., VACCARO, O. B., CANEVARI, M. & GARDNER, A. L. 1995. Taxonomy of the genus *Lycalopex* (Carnivora: Canidae) in Argentina. *Proceedings of the Biological Society of Washington,* 108**,** 729-747.

ZURANO, J. P., MARTINEZ, P. A., CANTO-HERNANDEZ, J., MONTOYA-BURGOS, J. I. & COSTA, G. C. 2017. Morphological and ecological divergence in South American canids. *Journal of Biogeography,* 44**,** 821-833.

Chapter III, Whole-genome analysis reveals the diversification of rails (Aves: Rallidae) in Galapagos and confirms the success of multiple goat eradication programs

Abstract

Similar to other insular birds around the world, the Galapagos rail (*Laterallus spilonota* Gould, 1841) became flightless following its colonization of the archipelago ~1.2 mya. Despite their short evolutionary history and lack of flight capability, rails have colonized seven different islands spanning the entire width of the archipelago. To date, it is uncertain how the geological history of the Galapagos Islands has shaped the diversification of rails. Galapagos rails were once common on islands with sufficiently high altitudes to support a humid habitat. However, after humans introduced goats to some islands , this habitat was severely reduced due to overgrazing. Habitat loss devastated some rail populations, with less than 50 individuals surviving. Due to this recent history of sharp population decline, the genetic diversity of Galapagos rails is a pressing conservation concern. Additionally, one enigma is the reappearance of rails on the island of Pinta after they were considered extinct. We analyzed a total of 39 whole genomes of Galapagos rails to investigate their evolutionary history and the genome-wide effects of population bottlenecks. We found that the separation of the landmass joining Santiago and Santa Cruz island populations with Isabela island ~400 kya may have led to the isolation of the Isabela population. Subsequently, rails migrated from this landmass to colonize the island of Pinta, where they have been genetically isolated ever since. Finally, our findings show that the eradication of goats was critical to avoiding episodes of severe inbreeding in most populations. One implication of our findings is that the remnant Santa Cruz population, where 40% of captured individuals had a kinship coefficient of 0.25, or sibling-level, may suggest the potential for inbreeding depresssion.

Introduction

The Galapagos archipelago is an excellent system for studying patterns of evolution due to its well-defined geological history and isolation from mainland South America (Garcia-R and Matzke, 2021, Holland and Hadfield, 2004, Poulakakis et al., 2012, Black, 1974, Grant and Grant, 2008). Aside from dispersion among islands, the second-most common mode of diversification in the Galapagos is vicariance, the subdivision of widespread ancestral populations caused by an environmental barrier (Arbogast et al., 2006, Caccone et al., 2002, Parent and Crespi, 2006, Poulakakis et al., 2020, Poulakakis et al., 2012, Steinfartz et al., 2009). Specifically, ~0.4 million years ago, several islands in the central region of the archipelago were connected, constituting a single landmass (Geist et al., 2014, Karnauskas et al., 2017, Schwartz et al., 2018). Rising sea levels over time resulted in the isolation of these islands from one another and led to the diversification of current lineages of giant tortoises, mockingbirds, vermilion flycatchers, and others (Poulakakis et al., 2012, Arbogast et al., 2006, Carmi et al., 2016).

Among the avifauna that inhabit the archipelago, the Galapagos rail (*Laterallus spilonota*) represents an excellent model species for investigating the role of island diversification. The species originated in the Galapagos from a single volant ancestor that arrived from mainland South America ~1.2 mya (Chaves et al., 2020). This suggests that the ancestral population was able to disperse to the majority of the islands (Geist et al., 2014, Karnauskas et al., 2017, Schwartz et al., 2018, Poulakakis et al., 2012). Additionally, the breakup of the previous central landmass of the archipelago might have influenced the pattern of diversification

of the Galapagos rail (Geist et al., 2014, Karnauskas et al., 2017, Schwartz et al., 2018). Critically, similar to other insular birds around the world, rails became flightless subsequent to their arrival (Slikas et al., 2002, Garcia-R and Matzke, 2021, Wright et al., 2016), which may have facilitated further genetic differentiation and diversification due to their limited capacity for dispersion.

A major concern in isolated oceanic islands is the impact of invasive species on native biodiversity. Numerous flightless birds around the world, including rails, have gone extinct due to species introductions by humans (Diamond, 1984, Olson and James, 1982, Steadman, 1995, Duncan and Blackburn, 2004). The Galapagos rail was once common on islands characterized by humid environments and dense ground vegetation at high elevations (Franklin et al., 1979, Rosenberg, 1990, Donlan et al., 2007). When goats were first introduced to the Galapagos islands in 1959 (Hoeck, 1984), their population rapidly grew to at least 20, 000 individuals by 1970 (Weber, 1971). Consequently, goats drastically reduced the highland vegetation through overgrazing (Hamann, 1979, Campbell et al., 2004). As a consequence, rail populations were extirpated from the islands of San Cristobal and Floreana (Rosenberg, 1990, Chaves et al., 2020, Gibbs et al., 2003). Populations from Isabela and Santiago were devastated, with less than 25 individuals estimated to have survived on each island (Donlan et al., 2007, Rosenberg, 1990). On Pinta, rails completely disappeared when goats peaked in population size around 1970 (Franklin et al., 1979). Only after the implementation of goat eradication programs starting in the 1970s did the population size of rails begin to increase (Gibbs et al., 2003, Donlan et al., 2007). The Santiago population increased from 23 to 233 individuals within a span of 18 years (Donlan et al., 2007). Likewise, rails on Pinta became abundant following goat extirpation (Franklin et al.,

1979, Donlan et al., 2007). Despite their recovery, rail populations could still be at risk due to the erosion of genetic diversity and accumulation of harmful genomic mutations during periods of small population size (Lynch et al., 1995, Lynch and Gabriel, 1990, Charlesworth and Willis, 2009). It has been shown that such deleterious variants could have detrimental effects on the fitness of individuals, affecting reproductive success or compromising the immune system in wild populations (Dobrynin et al., 2015, Johnson et al., 2010, Wayne et al., 1991, Charlesworth and Willis, 2009). Therefore, inbreeding remains a conservation concern, as it could lead to population decline, from which the species may never recover (Peterson et al., 2014).

Another major puzzle is the reappearance of rails on the island of Pinta after they were considered extinct (Rosenberg, 1990, Franklin et al., 1979, Chaves et al., 2020). It is uncertain whether the current population represents a recolonization event from a nearby island, or a relic population that survived undetected. Understanding the reappearance of rails on this island is critical for conservation management (Chaves et al., 2020). Specifically, such a relic population could be experiencing inbreeding due to a sharp and extended population bottleneck. Furthermore, surviving populations could possess unique adaptive alleles, which could ultimately increase the probability of survival in changing environments (Smith et al., 2014). The population could also serve as a source for conservation management, i.e. genetic rescue (Johnson et al., 2010, Hogg et al., 2006, Smith et al., 2014). Here, we aimed to investigate: 1) the diversification of rails in the Galapagos Islands; 2) the effect of island isolation and goat introduction on genic diversity and genetic differentiation among rails; and 3) the origins of the reappearance of rails on the island of Pinta. To accomplish these goals, we analyzed

phylogenetic relationships, patterns of genome-wide diversity, and deleterious variation by sequencing 39 whole genomes of endemic Galapagos rails.

Results

Phylogenetic and biogeographic history

To provide an evolutionary framework for the biogeographic history of Galapagos rails, we reconstructed a consensus phylogenetic tree of 39 individuals representing all extant island populations (Table 3-S1). We first mapped reads from Galapagos rail genomes to the Inaccessible Island rail (*Atlantisia rogersi*) reference genome assembly, a species endemic to the Tristan da Cunha archipelago in the southern Atlantic Ocean (Stervander et al., 2019). This reference assembly had a mean coverage of 41x, a total length of 1.168 Gbp and a contig and scaffold N50/L50 of 36,904 bp/8,087 bp (https://www.ncbi.nlm.nih.gov/assembly/GCA_013401215.1). For all Galapagos rail samples, we found that more than 96% of reads successfully mapped to the Inaccessible Island rail (Table 3-S1). We then extracted 6,540 alignments of 25kb windows from these mapped genomes. We used maximum likelihood to construct independent phylogenetic trees from each of these windows. To account for phylogenetic discordance among these trees, we generated a consensus phylogenetic tree with ASTRAL-III (Zhang et al., 2018; Figure 3-1a). Our species tree showed that samples from Isabela and Pinta formed independent monophyletic clades (Figure 3-1a). The latter population was sister to a third clade that included samples from Santiago and Santa Cruz. These island populations were not reciprocally monophyletic (Figure 3-1a).

To elucidate the biogeographical history of rails in the Galapagos Islands, we used the R package BioGeoBEARS (Matzke, 2013). This tool estimates the maximum-likelihood

distribution of hypothetical ancestors (internal nodes) by modeling shifts between different geographical ranges along the phylogeny. First, we tested three different models: dispersal-Extinction-Cladogenesis (DEC); dispersal vicariance analysis (DIVALIKE); and the Bayesian analysis of biogeography (BAYAREALIKE). Additionally, we tested the same three models plus a founder effect parameter on each model named "J": DEC+J, DIVALIKE +J, and BAYAREALIKE+J (Matzke, 2013). Among six different models tested, the BAYAREALIKE was the one that best fit our data, showing a corrected AIC 45% higher than the other five models (Table 3-S2). This model assumes episodes of geographical dispersion, contraction, and range switch of an ancestral population without speciation. According to this model, rails had a relatively large geographical range that contracted into smaller isolated areas. Specifically, there is a 47% probability that Galapagos rails originally occurred across a geographical range that included Isabela, Santa Cruz, and Santiago (see pie charts in Figure 3-1a). This ancestral distribution then contracted and populations split into two different areas. One occurred on Santa Cruz and Santiago and the other occurred on Isabela (Figure 3-1a and 3-1b). Each of this two ancestral ranges were recovered with 100% probability. Following this range split, there was a 100% probability that an ancestral population dispersed to the geographically isolated island of Pinta (Figure 3-1a and 3-1b).

Population structure and admixture

We investigated the amount of genetic differentiation among Galapagos rail populations by calculating the mean pairwise F_{ST} between islands using 10,274 SNPs pruned for linkage disequilibrium from an original set of 651,558 SNPs. We found that rails from Pinta were the most differentiated population relative to other islands, with a pairwise F_{ST} ranging from 0.108

with Santa Cruz to 0.115 with Isabela. Individuals from Isabela were the second-most differentiated population with an F_{ST} ranging from 0.035 with Santa Cruz to 0.115 with Santiago. Rail populations from Santiago and Santa Cruz islands had the lowest F_{ST} value of 0.001 We found some degree of association between F_{ST} and geographical distance between islands. Specifically, values of F_{ST} increased with distance between populations (Figure 3-S1).

To assess the level of population structure among island populations, we used the 10,274 SNPs to conduct a principal component analysis (PCA). To avoid including related individuals in the PCA analysis, we used an identity-by-descent method with PLINK (Purcell et al., 2007) and removed one individual from a pair that had a kinship coefficient > 0.2 from Isabela and four related individuals from Santa Cruz (kinship scores > 0.2; see methods for details). We found that the Pinta population was clearly distinguished from the other three populations (Santiago, Santa Cruz, and Isabela) along PC1. This component represented 7.7 % of the variance, separating samples between the northern and southern islands (Figure 3-2a and 3-2b). PC2 separated the Isabela population, located in the west region of the Galapagos, from islands in the east (Santiago, Santa Cruz, and Pinta) with 5.1% of the variance explained (Figure 3-2a and 3-2b). Populations from Santiago and Santa Cruz cluster together on both PCs (Figure 3-2a), indicating low levels of population structure between these islands. To better elucidate any possible pattern of genetic differentiation between Santiago and Santa Cruz, we excluded from the PCA individuals from Pinta, the most divergent population (Figure 3-S2). Nonetheless, samples from these islands still overlapped on both PC1 and PC2. Additional PCs showed similar results, which confirm a lack of population structure among the Santiago and Santa Cruz populations.

To further investigate the genetic clustering of Galapagos rails, we used the program fastStructure (Raj et al., 2014) using values of k from 1 to 10. The different k values tested had similar marginal likelihoods (Table 3-S3). At k=2, we found individuals from Pinta separating from another group composed of individuals from Isabela, Santiago, and Santa Cruz (Figure 3-2c). These two genetic clusters were observed from k=3 to k=4 (Figure 3-S3a). At k=5 (Figure 3-2c) and k=6 (Figure 3-S3a), individuals from Isabela formed an additional cluster. Within this group, four samples have a relatively high assignment (>20%) to a cluster that includes Santa Cruz and Santiago (see k=5 in Figure 3-2c). The other three individuals belong almost exclusively to the Isabela cluster with a 98% of assignment to this group (see k=5 in Figure 3-2c). These results suggest a signal of mixed ancestry among rails sampled on the island of Isabela (Figure 3-2c).

We investigated admixture in Galapagos rails using TreeMix (Pickrell and Pritchard, 2012), which uses whole genome allele frequencies to discover the best population tree as well as to infer gene flow between different populations. Similar to the ASTRAL-III tree in Figure 3-1a, TreeMix showed that the Santiago and Santa Cruz rail populations are closely related (Figure 3-2d). In contrast, Pinta and Isabela were recovered as independent populations (Figure 3-2d). The former population was the most divergent as evidenced by a relatively long branch in the tree. Notably, this branch reached a drift score of 0.05 which was five times greater than the score of other populations at ~0.01 (see branch length in Figure 3-2d). Importantly, we found evidence of admixture from the Isabela to Santiago population, as suggested by a migration weight value of 0.5 (Figure 3-2d). This score is obtained by comparing the standard error of a model without gene flow (migration edges = 0) with a model that allows for admixture

(migration edges > 0). A positive migration weight suggested that adding gene migration to the model reduced the stand error and improve the fit to the data (see Figure 3-S3b for details).

Genetic diversity

To understand how demographic history has shaped patterns of genetic variation among Galapagos rail populations, we examined heterozygosity in non-overlapping 100 kb windows across the genome of every sequenced individual (Figure 3-3a). We observed that genomes from different islands had high heterozygosity throughout (Figure 3-3a). Also, we observed only a few long stretches which seemed to be depleted of heterozygosity (Figure 3-3a). The level of genome-wide heterozygosity was similar among individuals from different islands and ranged from 0.08 (heterozygosity/kb) on Pinta to 0.10 (heterozygosity/kb) on Isabela and Santiago (Figure 3-3b and Table 3-S4).

To more precisely examine the regions across the genomes depleted of heterozygosity, we quantified the extent of runs of homozygosity (ROH) grouped into three different size categories using PLINK (Purcell et al., 2007; Figure 3-3b). Notably, long ROH (>2 Mb) are a likely consequence of recent close inbreeding due to sharp population declines, whereas ROH of a short (<0.5 Mb) and medium (0.5 - 2 Mb) length could reflect ancient population declines as there would be more time for recombination to break up long segments (Ceballos et al., 2018). ROH were three times higher in all categories in rails from Pinta relative to other islands (Figure 3-3b and 3-3c), consistent with their lower level of heterozygosity (~ 0.08 heterozygosity/kb). The average total length of short ROH in the genome from Pinta was 52.4 Mb, which represents 4% of the genome, whereas the average total length of medium ROH was 68 Mb which represents 5% of the genome (Figure 3-3b and 3-3c; Table 3-S4). Similarly, the total length of

long ROH fragments was ~13 Mb on average (Table 3-S4), which is two times longer than the ROH of rails from other islands and represents less than one percent of the genome (Figure 3-3b and 3-3c). Individuals with relatively higher heterozygosity (~ 0.10 heterozygosity/kb) included samples from Isabela, Santa Cruz, and Santiago (Figure 3-3b). These island populations had fewer small ROH (< 0.5 Mb), that on average ranged from 17.7 Mb in Isabela to 22.3 Mb in Santiago (Figure 3-3b and 3-3c; Table 3-S4) and represent less than 2% of the genome. In these populations, medium ROH (0.5-2 Mb) are also infrequent and represent only 2% of the genomes with values ranging from 16.6 Mb in Isabela genomes to 24.8 Mb in Santiago genomes (Figure 3-3b and 3-3c; Table 3-S4). Also, a few long ROH (>2 Mb) were observed for these individuals, with a total length ranging from 1.5 Mb in genomes from Isabela to 4 Mb in Santa Cruz genomes (Figure 3-3b and 3-3c; Table 3-S4). The exception was one individual from Santiago (LS50), which had more ROH in the three size categories (Figure 3-3b and Table 3-S4). Overall, our results suggest that recent population declines and inbreeding may have been more pronounced in rails on Pinta than those of other islands.

Deleterious variation

To investigate the effects of demography on deleterious variation in rails, we analyzed evidence for the accumulation of deleterious variants as a result of population declines using SnpEff (Cingolani et al., 2012). We annotated mutations in protein-coding regions of the genome as synonymous (as a proxy of neutrality), missense (nonsynonymous amino acid substitution), or loss of function (LOF; e.g., premature stop codon). The latter category is expected to have a higher impact on the phenotype since the function of the protein could be affected, so we regarded LOF mutations as probable deleterious mutations. We found that each population had

similar proportions of homozygous-derived genotypes for each mutation type (synonymous, missense mutation, and LOF; Figure 3-4). For instance, the proportion of homozygous derived genotypes under the loss of function category was ~0.55 for Isabela, Pinta, Santiago and Santa Cruz populations. This suggests that inbreeding in Pinta, as suggested by their relatively long stretches of ROH that total ~13 Mb, have not increased their proportion of homozygous-derived genotypes with respect to the proportion of such genotypes from other populations (Figure 3-4).

Discussion

We reported a large-scale population genomic analysis of Galapagos rails from the islands of Pinta, Isabela, Santiago, and Santa Cruz. Our reconstruction of historic geographic ranges of the different populations included a combination of dispersal and contraction events, reflecting the separation of a central landmass ~400 kya (Geist et al., 2014, Karnauskas et al., 2017, Schwartz et al., 2018). We found that the Isabela population was the first to split from this landmass (Figure 3-1a-c). Importantly, the island of Pinta has been historically isolated (Black, 1974), which suggests that rails migrated from the central landmass to colonize this island (Geist et al., 2014, Karnauskas et al., 2017, Schwartz et al., 2018). The Santa Cruz and Santiago populations were the last populations to diversify, which is consistent with the absence of population structure among these islands (Figure 3-2a and 3-2c). Conversely, the population of Pinta was the most divergent (Figure 3-2d), which confirms that this is a relic population that survived the introduction and overgrazing of their habitat by goats (Kramer and Black, 1970, Franklin et al., 1979, Hamann, 1979, Weber, 1971). This population has more long ROH than the other populations (Figure 3-3b and 3-3c), which suggests recent inbreeding. However, this

episode of inbreeding has not increased their proportion of deleterious mutations with respect to other islands (Figure 3-4), suggesting that-Pinta population may not be experiencing inbreeding depression, which may explain their rapid recovery (Franklin et al., 1979, Donlan et al., 2007). Similarly, the recovery of other island populations following the eradication of goats may be related to the absence of severe inbreeding.

Pattern of island diversification

Similar to other species found on the Galapagos archipelago, ancestors of Galapagos rails likely first colonized San Cristobal, which is the closest island to the South American continent (Arbogast et al., 2006, Caccone et al., 2002, Parent and Crespi, 2006, Poulakakis et al., 2020, Poulakakis et al., 2012, Chaves et al., 2020). However, the subsequent events of diversification of rail populations have been uncertain (Chaves et al., 2020). Their diversification could have been influenced by island dispersion or geographic vicariance as found with numerous other endemic species in the Galapagos (Arbogast et al., 2006, Caccone et al., 2002, Parent and Crespi, 2006, Poulakakis et al., 2020, Poulakakis et al., 2012, Steinfartz et al., 2009). Based on analyses of whole genome sequences, our results provide new insights into how rails diversified across the Galapagos archipelago. According to our biogeographical model, the ancestral geographical distribution of rails includes a single area that included the islands of Santiago, Santa Cruz, and Isabela. This distribution is consistent with current island proximity and geologic history (Black, 1974, Geist et al., 2014, Karnauskas et al., 2017, Schwartz et al., 2018), as the three central islands were once a single landmass that began to separate ~400 kya (Figure 3-1b and 3-1c) with rising sea levels (Schwartz et al., 2018, Poulakakis et al., 2012). The ancestral lineage from this united landmass became the present-day populations of Isabela, Santiago, and Santa Cruz,

probably through vicariance (Figure 3-1b). Interestingly, this model of lineage diversification has been shown for Galapagos tortoises (*Chelonoidis niger* complex) that diverged during the separation of this landmass (Poulakakis et al., 2020, Poulakakis et al., 2012). Tellingly, vermilion flycatchers (*Pyrocephalus nanus*) showed a pattern of genetic clustering remarkably similar to the genetic clusters we found among rails (Figure 3-1a). In both vermilion flycatchers and rails, there is evidence that populations from Isabela in the west split first, followed by Santa Cruz and Santiago populations in the east (Carmi et al., 2016). This splitting pattern suggests that Isabela may have been the first island to separate from the central region (Black, 1974, Schwartz et al., 2018), creating the common diversification pattern observed in numerous lineages including rails (Poulakakis et al., 2020, Poulakakis et al., 2012, Carmi et al., 2016). Our biogeographical model further suggests a past colonization event from the central landmass to the geographically isolated island of Pinta (Schwartz et al., 2018, Poulakakis et al., 2012). Isolated islands are expected to be more genetically differentiated (Sendell-Price et al., 2021), which is consistent with our TreeMix results which places Pinta as the most divergent population (Figure 3-2d). Furthermore, ancient population bottlenecks are expected to be more pronounced in populations that went through a founder event (Sendell-Price et al., 2021, Quesada et al., 2019, Ceballos et al., 2018). Our results support this prediction as they show an overrepresentation of medium ROH in rails from Pinta (Figure 3-3b and 3-3c), which is an indicator of ancient population bottlenecks.

To date, it is uncertain how flightless rails were able to colonize the island of Pinta, which is ~80 km from the central archipelago (Black, 1974). The loss of flight after island colonization is a common pattern in insular birds around the World (Slikas et al., 2002, Garcia-R and Matzke, 2021, Wright et al., 2016). However, the process of flightlessness is not

instantaneous and includes the progressive reduction of flight muscles followed by an enlargement of hind limbs (Olson, 1973, Wright et al., 2016) and includes changes in a variety of genes involved in developmental pathways (Burga et al., 2017). Therefore, it is possible that rails were volant for most of their evolutionary history and originally flew from the central archipelago to the north island of Pinta (White et al., 1993, Schwartz et al., 2018). Previous studies have reported that rails are capable of flying distances up to 30m, which suggests that loss of flight in rails may still be an ongoing process (Franklin et al., 1979). In other birds like Galapagos cormorants, the reduction of wings have make them completely flightless (Burga et al., 2017). Alternatively, the ancestral population or rails may have been flightless and arrived on the island by rafting like tortoises or floating on vegetation like insects (Sequeira et al., 2008, Caccone et al., 2002, Poulakakis et al., 2012, Quesada et al., 2019). These two non-volant modes of dispersion could explain the absence of population structure among the Santiago and Santa Cruz individuals (Figure 3-2a and 3-2c) as well the observed admixture between the Isabela and Santiago genomes (Figure 3-2d).

Eradication of goats from Galapagos Islands: Did it work?

In the Galapagos, rails from San Cristobal and Floreana were decimated by the introduction of goats and have not been seen since 1987 (Rosenberg, 1990). In contrast, rail populations from Pinta and Santiago recovered after the eradication of goats (Rosenberg, 1990, Donlan et al., 2007). Despite the recovery of these populations, they could be at risk due to inbreeding depression following population bottlenecks (Charlesworth and Willis, 2009). Analogously, gray wolves (*Canis lupus*) on Isle Royale in North America originally increased from three to 15 individuals, but due to inbreeding depression, wolf population size declined and has been struggling to recover ever since (Peterson et al., 2014, Robinson et al., 2019). Despite drastic

population declines recorded in the past for the Galapagos rail (Rosenberg, 1990, Donlan et al., 2007), little is known about their genetic diversity (Chaves et al., 2020) or mutational load.

We characterized genome-wide diversity in four different Galapagos rail populations and observed relatively high levels of heterozygosity across the genome (Figure 3-3a). Although we found evidence of long ROH within populations indicative of recent population declines and inbreeding (Figure 3-3b), these constituted less than 1% of the entire genome. Importantly, species with severe inbreeding show a dominance of long blocks of ROH across their genomes, and in some cases, ROH can span entire chromosomes (Kardos et al., 2018, Robinson et al., 2019). Furthermore, rail populations that experienced sharp population declines due to habitat lost (e.g., Pinta) might be expected to have an overrepresentation of deleterious variants as observed in some populations of sea otters and wolves (Beichman et al., 2019, Robinson et al., 2019). However, the proportion of homozygous-derived mutations, including LOF mutations, across the different island populations was similar (Figure 3-4).

We hypothesize that the lack of severe inbreeding in rails on Pinta was due to the relatively short decrease in population size due to the presence of goats which reduced rail habitat. Goats remained at low elevation regions for approximately eight years (Hamann, 1979, Campbell et al., 2004). By the time goats moved to higher elevations and started to reduce ground vegetation around 1970, the goat eradication program had already been established on Pinta (Franklin et al., 1979). Another factor that could have mitigated population decline in rails could be the rapid recovery of their habitat (Hamann, 1979). Although goats completely cleared the ground vegetation in the highlands, the vegetation was reestablished in less than one year due to vegetative reproduction (Hamann, 1979). These plants can rapidly grow asexually from old

stems and rhizomes, as opposed to the plants in the lowlands which recovered relatively slowly due to phases of growing, flowering, and fruiting (Hamann, 1979). Both the rapid eradication of goats and the reestablishment of highland vegetation on Pinta most likely allowed rails to recover without experiencing severe inbreeding (Hamann, 1979, Campbell et al., 2004).

The goat eradication programs on Santiago and Isabela were established 26 years after the program on Pinta. However, the relatively large habitat of these islands coupled with multiple rail subpopulations may have mitigated episodes of inbreeding (Rosenberg, 1990, Gibbs et al., 2003, Donlan et al., 2007). Particularly, Isabela is 15 times larger in area than Pinta (Black, 1974). There are six different volcanoes across Isabela with enough altitude to support suitable habitats for rails. Similar to Galapagos tortoises, rails living on different volcanoes could be genetically differentiated (Beheregaray et al., 2004). Therefore, gene flow from other populations might have restored the genetic variation to small populations in the south that were most affected by goat overgrazing (Rosenberg, 1990, Donlan et al., 2007, Carrion et al., 2011). The existence of different genetic units on Isabela is supported by the presence of rails at the Wolf volcano located in the north region of Isabela island (Rosenberg, 1990). The elimination of approximately 140,000 goats from Isabela may have also mitigated population decline and inbreeding in rails on this island (Carrion et al., 2011).

In the case of Santiago, the elimination of 79,000 goats led to a 16-fold increase in rail population size (Carrion et al., 2011, Donlan et al., 2007). The extirpation of goats from Santiago could explain the lack of severe inbreeding observed among rails from this island. In contrast to Santiago, the eradication of goats in Santa Cruz was minimal, with just 1,700 goats eliminated. Previous surveys have reported a long-standing decline of rail populations on this island due to the presence of another invasive species, the Cinchona tree (*Cinchona pubescens*) that has

further reduced the rails' suitable habitat (Gibbs et al., 2003, Rosenberg, 1990). Among the 11 individuals captured and sequenced from Santa Cruz, 40% had a kinship coefficient of 0.25, a score expected for siblings and which could indicate the initiation of a severe episode of inbreeding. However, high levels of kinship coefficient may also reflect the nature of sampling which could disproportionately sample members of the same pedigree (Chaves et al., 2020).

Evidence of a relic population in the island of Pinta

Around 1970, the ground vegetation of the highland region of Pinta, which rails previously inhabited, was cleared due to overgrazing by goats (Weber, 1971). During two different surveys that year, no rails were detected on Pinta (Franklin et al., 1979, Kramer and Black, 1970). In 1971, a goat eradication program was established, and surprisingly, just one year after this program had started, rails reappeared (Franklin et al., 1979). To date, it is uncertain whether rails recolonized Pinta from a nearby island or if relic populations from Pinta managed to survive (Chaves et al., 2020, Kramer and Black, 1970). Our findings suggest that the post-goat eradication population on Pinta is most likely a relic population. Specifically, our results from the PCA, Structure, and TreeMix analyses suggest that rails from Pinta are genetically divergent and have an evolutionary history largely independent of those on the other islands (Figure 3-1a and 3-2d). Our findings also suggest that rails remained unnoticed during previous surveys conducted by Kramer and Black (1970). Rails are elusive small birds and even big-sized species such as Galapagos tortoises had remained undetected on Pinta (Snow, 1964, Castro, 1969, Franklin et al., 1979). Tortoises were thought to be extinct in the 1960s after multiple attempts to find them (Snow, 1964, Castro, 1969). Ten years later, however, one specimen named "Lonesome George" was found by accident (Vagvolgyi, 1974).

Conclusion

By generating and analyzing whole genome sequence data from 39 individuals, we have provided novel insight into the diversification and population history of Galapagos rails. Our results indicate that the separation of the Galapagos central landmass 400 kya played a critical role in the diversification of endemic species in the Galapagos, including the Galapagos rail (Poulakakis et al., 2012, Carmi et al., 2016, Schwartz et al., 2018, Poulakakis et al., 2020). The rails from Pinta are a long-term isolated population that survived an almost complete loss of habitat due to overgrazing by goats. Conservation efforts should focus on this relic population as it could have unique genomic signatures which could enhance the adaptation to the unique island environment of Pinta (Smith et al., 2014). Moreover, rails from Pinta are the most genetically divergent population. Therefore, the extinction of this small population would represent a significant loss of the genetic diversity for the entire species. Finally, we found that the extirpation of goats mitigated episodes of severe inbreeding among rail populations. One exception could be the population of Santa Cruz, where 45% of individuals analyzed had a kinship coefficient of 0.25, which indicates that these individuals are siblings. Goats have not been significantly eradicated from this island and the invasion of Cinchona trees may further affect rail populations in Santa Cruz. Therefore, we suggest that from a conservation perspective, rails from Santa Cruz island could be experiencing a severe episode of inbreeding. Additional samples are needed to further evaluate the status of this population, but at the moment, we suggest that rails from Santa Cruz could be at an especially high risk of extinction and elimination of goats from the island should be a high priority.

Methods

Sequencing Mapping and Genotype calling

We extracted genomic DNA from 39 samples of Galapagos rails from four different islands (Isabela, Santiago, Santa Cruz, and Pinta) using the Qiagen DNEasy kit (Qiagen, USA). Samples were selected based on sufficient quality and quantity of genomic DNA with a DNA fluorometer (Qubit 2.0), a NanoDrop spectrophotometer (ThermoFisher, USA), and gel electrophoresis. We generated short insert (150 bp) libraries from these extractions using TruSeq Nano PCR-Free kit (Illumina, USA) and then conducted whole-genome sequencing on these libraries using the Novaseq 6000 sequencer (Illumina, USA) with 150 bp paired-end reads at the Vincent J. Coates Genomics Sequencing Laboratory, University of California, Berkeley.

Raw sequencing reads were processed and filtered using a modified pipeline from the Genome Analysis Toolkit (GATK) best practice guide (McKenna et al., 2010). We then mapped reads with good quality to the Inaccessible Island rail (*Atlantisia rogersi*) genome assembly (GCA_013401215.1; Stervander et al., 2019), using BWA-MEM (Li, 2013). We used GATK HaplotypeCaller to conduct joint genotype calling from sites of the Galapagos rail genomes that were mapped to the Inaccessible Island rail reference genome. Then, we filtered the called genotypes for coverage and quality. We kept only genotypes that had a minimum of 5 reads at a given position and a high-quality score (Phred scores >= 20), and no more than the 99th percentile of coverage for each sample. Other variant filtering criteria followed GATK best practices guide (McKenna et al., 2010) and Robinson et al. (2019). We filtered out CpG islands, indels, multi-nucleotide polymorphisms, and sites with more than one alternate allele. The command lines used for read mapping, variant calling, and filtering are available at https://github.com/dechavezv/2nd.paper.v2

Phylogenetic analysis

We reconstructed a species tree representing the relationships among 39 Galapagos rail genomes using ASTRAL-III (Zhang et al., 2018). First, we extracted 6,540 alignments, each 25kb in length, from the entire set of 35 autosomal chromosomes. Each independent window was aligned with PRANK v.150803 using iteration (-F once option) and the topology shown in Figure 3-1a as a guide tree. After trimming each of the multiple alignments with Gblocks (Castresana, 2000), we calculated the maximum-likelihood gene tree phylogeny of the 6,540 alignments using RAxML 8.0 (Stamatakis, 2014) under the GTR model. For each alignment, the best tree was selected from the RAxML output, while the 100 bootstrap trees were merged into a single file per locus. We then created a consensus file with these trees with no more than 10% of missing data. The branch lengths that were shorter than 1e-05 substitution per site were collapsed. Also, we collapsed clades with support lower than one using the SqCL pipeline (phylogeny_prep_astrid_astral.py). We used the best tree and consensus file of bootstrap trees to investigate the discordance between gene trees and the species tree using ASTRAL-III v.5.5 (Zhang et al., 2018). This resulted in maximum likelihood support values that were used to choose the best multi-locus tree. We ran 100 bootstrap replicates on this tree and further scored it to calculate the quartet support values and posterior probabilities for each node. We used the Inaccessible rail as an outgroup to root the tree.

BioGeoBEARS

We investigated the biogeographic history of Galapagos rails using the R package BioGeoBEARS (Matzke, 2013). This tool estimates the maximum-likelihood distribution of hypothetical ancestors (internal nodes) by modeling shifts between different geographical ranges

along the phylogeny as a function of time. First, we tested three different models: dispersal-Extinction-Cladogenesis (DEC), dispersal vicariance analysis (DIVALIKE), and Bayesian analysis of biogeography (BAYAREALIKE). Additionally, we tested the same three models plus a founder effect parameter added to each model named "J": DEC+J, DIVALIKE +J, and BAYAREALIKE+J (Matzke, 2013) . The estimation to determine the best-fitting model was conducted by comparing model likelihoods through the Akaike information criterion (AIC) scores and AIC weights (Table 3-S2). We used default parameters and set the max_range_size =4. The scripts and files used to run these models can be found at https://github.com/dechavezv/2nd.paper.v2/blob/main/3-Phylogenomics/BioGeoBEARS/

Fst, FastStructure and PCA

We calculated F_{ST} for every island-pair population with SNPRelate (Zheng et al., 2012). We used the method *"W&C84"* on .gds files and removed sites that were monomorphic and with no more than 20% missing data. The geographic distances were obtained using the geosphere package in R (https://CRAN.R-project.org/package=geosphere). Then, we created a plot of Fst vs. distance (see Figure 3-S1).

To estimate the level of relatedness among samples, we convert .vcf files to the gds inputs required by PLINK using SNPRelate (Zheng et al., 2012). This tool was used on a subset of ~10k high-quality SNPs pruned for LD, with a r^2 threshold of 0.2, and a maximum missing rate greater than 0.05. Then, we assessed the level of relatedness among rail samples from different islands by estimating identical-by-descent (IBD) using PLINK's method of moments approach (snpgdsIBDMoM; Purcell et al., 2007), with a minor allele frequency cutoff of 0.05.

When the resulting kinship scores between a pair were greater than 0.2, we randomly removed one sample from this pair. As a result, we removed four samples from Santa Cruz and one sample from Isabela.

To evaluate the existence of genetic structure among rail populations, we used SNPrelate to estimate the principal component analysis on 10,274 SNPs pruned for LD (threshold of 0.2) and a minor allele frequency of 0.1. We used windows 500kb in length. Among the different principal components (PC) estimated, we plotted PC1 and PC2 with ggplot2 (Wickham, 2016), as these components had most of the variation explained by the data. We further conducted a Bayesian clustering analysis using fastStructure (Raj et al., 2014). For each model, we tested 10 different partitions from k=2 to k=10 and selected the partition that had the optimal k value based on "chooseK" criteria implemented in fastStructure as well as their marginal likelihoods (Table 3-S3).

Treemix

We used TreeMix (Pickrell and Pritchard, 2012) to further evaluate the relationship among island populations and evaluate levels of admixture among Galapagos rails. Specifically, we used the whole genome allele frequencies of 39 samples to discover the best population tree. This tree was obtained by comparing independent maximum-likelihood scores. TreeMix then compared the covariance of the depicted tree to the observed covariance between populations to identify possible admixture events (Pickrell and Pritchard, 2012). Population pairs that fail to fit the modeled tree could be subject to admixture (Figure 3-3Sb). We tested several models in our whole-genome data for the subset of rails without related individuals as identified by SNPrelate

(Zheng et al., 2012). First, we tested a model with no migration and blocks of 500 SNPs (-k 500). Then, we added eight different migration edges (m=1 to m=10) to the phylogenetic model in TreeMix.-To determinate the best-fitting model, we used the -global option and -se option to calculate the standard errors of the residual covariance. Finally, we considered the residual covariance between different population pairs and divided this difference by the standard error from all different pair populations.

Genomic Diversity

We examined the site heterozygosity in non-overlapping 100kb windows across the genome of all 39 Galapagos rails. We defined heterozygosity as the number of heterozygous genotypes divided by the total number of sites that were called. The total genotypes called within each window included: the sum of heterozygous, homozygous derived, and homozygous reference genotypes. We kept only windows with no more than 20% of missing data. The script used to calculate 100kb windows heterozygosity was modified from Robinson et al. (2019) and is available at https://github.com/dechavezv/2nd.paper.v2/tree/main/4-Demography/Heterozygosity/WindowHet. We then quantified the extent of Runs of Homozygosity (ROH) in rails using PLINK (Purcell et al., 2007). The parameters chosen to calculate ROH were: SNPs within a window =200, heterozygotes allowed within a window= 3, missing sites within a window=50. We binned these segments into three different size categories using PLINK (Purcell et al., 2007) (Table 3-S4 and Figure 3-3b). The size categories were: short (0.5 MB <), medium ROH (0.5- 2 MB), and long ROH (> 2Mb).

Deleterious variation

We used SnpEff (Cingolani et al., 2012) for the functional annotation of each variant of the 39 samples of Galapagos rails. The annotation file corresponded to Inaccessible Island rail (*Atlantisia rogersi*), which is located at https://www.ncbi.nlm.nih.gov/assembly/GCA_013401215.1. SnpEff predicts the effects of genetic variants (e.g. stop-gain variants) and assesses their expected impact. The following categories were retrieved: loss of function (e.g. stop-gain or frameshift variant); missense variant; and synonymous variant. When more than one transcript was available for a single gene, we used the longest transcript for further analysis. The script to annotate and identify deleterious variants can be found at: https://github.com/dechavezv/2nd.paper.v2/tree/main/4-Demography/DeltVariation.

Figures

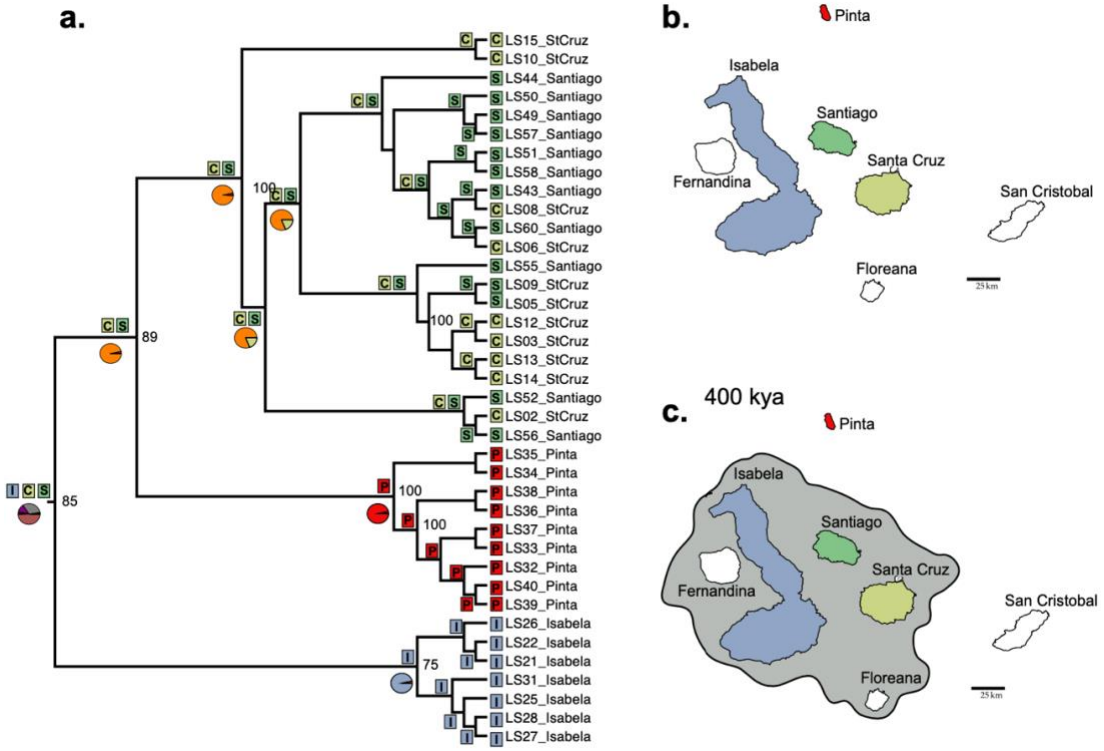

Figure 3-1. (a) Ancestral area reconstruction from BioGeoBEARS, derived from the species tree of 39 Galapagos rails obtained by ASTRAL-III based on 6,540 25kb windows. The best-fitting model was BAYAREALIKE (see Methods). Under this model, the geographical range of an ancestral population could expand, contract or switch to another area, but it does not allow for speciation (Table 3-S2). Numbers on the nodes of the tree indicate bootstrap support (only bootstraps >70 are shown). The estimated ancestral geographical ranges are shown by colored boxes on each node, while colored squares at terminal branches indicate the current distribution of analyzed samples. The color of the squares corresponds to the colors of the islands; the letter "I" indicates Isabela, "P" indicates Pinta, "S" indicates Santiago and "C" indicates Santa Cruz. The ancestral region's probabilities are shown in the pie charts below each node. The pie chart of the most basal node uses purple to indicate the probability of a range that includes "I", "C" and "S". **(b)** Current distribution of the Galapagos rail, showing the island of Pinta as the most geographically isolated island among those where rails occurred. **(c)** The paleogeographic map of the Galapagos Islands 400,000 years ago (Schwartz et al., 2018, Poulakakis et al., 2020, Poulakakis et al., 2012).

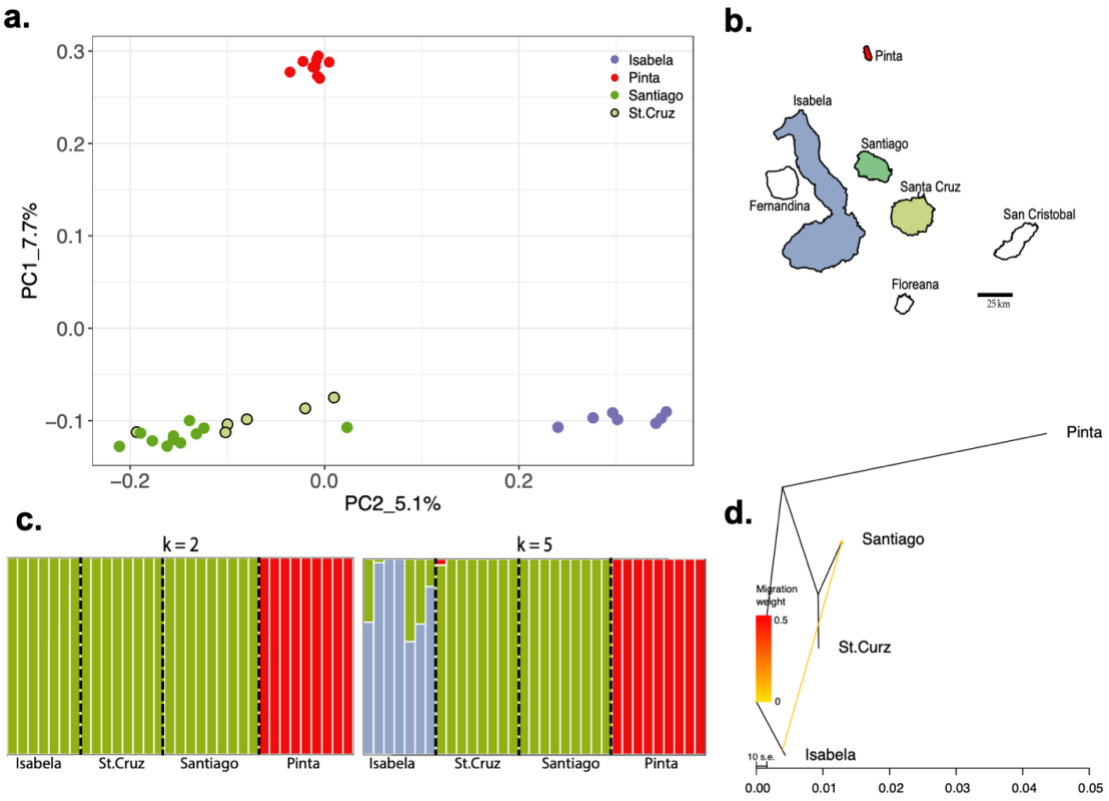

Figure 3-2. (a) Principal component analysis (PCA) of Galapagos rail samples based on 10,274 sites pruned for linkage disequilibrium. The PCA shows that rails from the islands of Pinta and Isabela form distinct clusters. Notably, Pinta is well-differentiated along PC1. This component separated samples into northern versus southern islands (see map on the top right). Rails from Isabela form a distinct cluster along PC2. This component separated samples into west and east island groups (see map on the bottom right). Individuals sampled from the Santiago and Santa Cruz islands overlapped on both PC components, indicating low levels of population structure between these islands (Figure 3-S3a). **(b)** The current distribution of Galapagos rails (colored islands). **(c)** fastSTRUCTURE analysis of the samples with k=2 and k=5. Each column is a single individual, and colors indicate different genetic ancestry groups. Additional values of k=3 and 4 match the k=2 clustering pattern (Figure 3-S3a), whereas k=5 and k=6 are identical (Figure 3-S3a). Dotted vertical lines delimit island samples. **(d)** TreeMix analysis showing drift (x-axis) between Galapagos rail populations with Isabela as the root population. In this tree, rails from Pinta and Isabela form independent lineages. The latter group is the most differentiated population as evidenced by its relatively long branch. One migration event from Isabela to Santiago is shown as an orange line and is related to a positive migration weight score, which favors a model with admixture (see Figure 3-S3b for details).

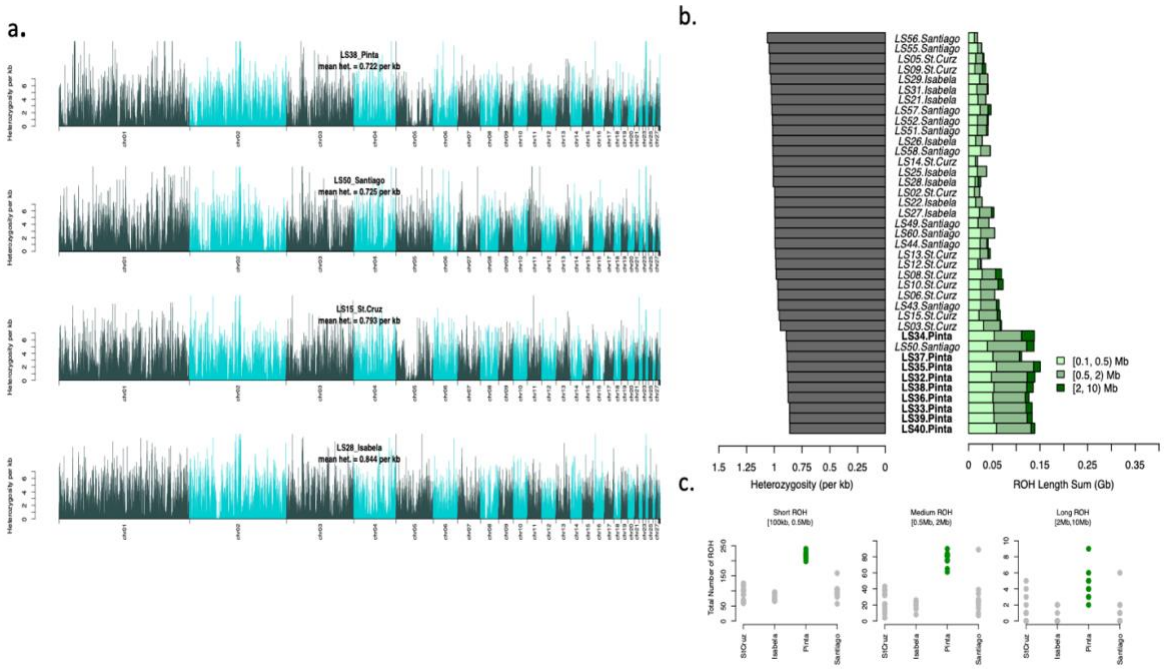

Figure 3-3. Genome-wide diversity and the distribution of ROH in Galapagos rails. **(a)** Heterozygosity per 100 kb window with a 10 kb step size across the genome of Galapagos rails. Only one individual per island is shown (See Figure 3-S4 for all individuals). Rails from different islands showed high heterozygosity. Some stretches of low heterozygosity can be seen across the genome. For instance, these stretches are shown in chromosome 5 of the Pinta and Santa Cruz genomes **(b)** Histogram with per-site heterozygosity across the autosomal genome (left panel) and summed lengths of ROH of three specific length categories (right panel). On the right panel from top to bottom: summed lengths of short (< 0.5 Mb); medium (0.5 Mb ≤ ROH < 2 Mb); and long (>2 Mb) ROH per individual **(c)** Panels showing the total number of ROH in the three length categories. From left to right: short ROH indicate ancient inbreeding as in the Pinta island; medium ROH indicate ancient and historic inbreeding, as in the Pinta island; and long ROH indicate recent inbreeding, as in some individuals from Pinta.

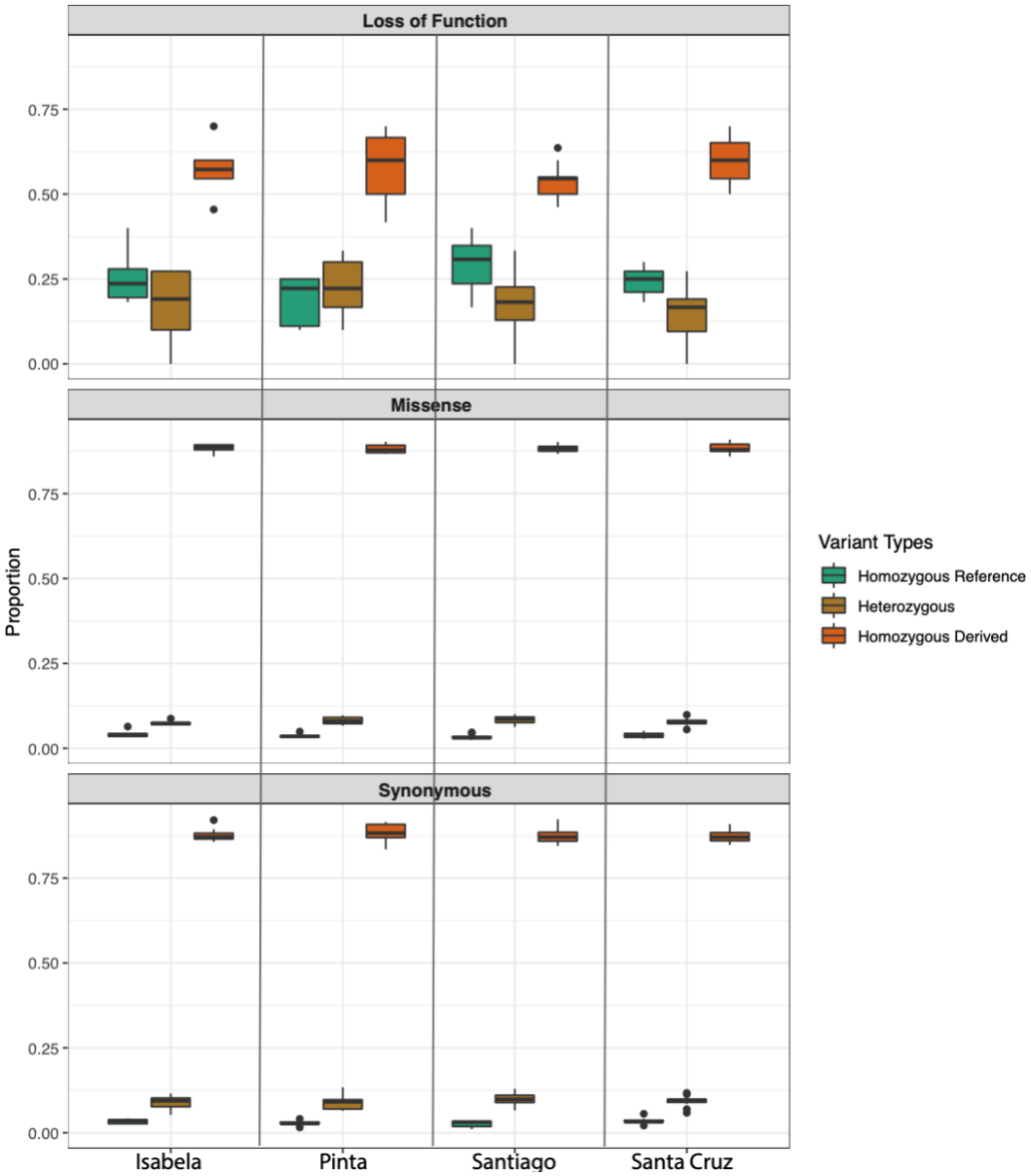

Figure 3-4. The proportion of the three mutation classes across different populations of Galapagos rails. Mutations were identified as either: synonymous (a proxy for neutrality), missense, or loss of function mutation (LOF). The proportion of homozygote-derived and ancestral alleles, as well as heterozygote alleles, are shown. Each population has similar proportions of homozygous-derived genotypes for different mutation types.

Appendix 3-I: Figures

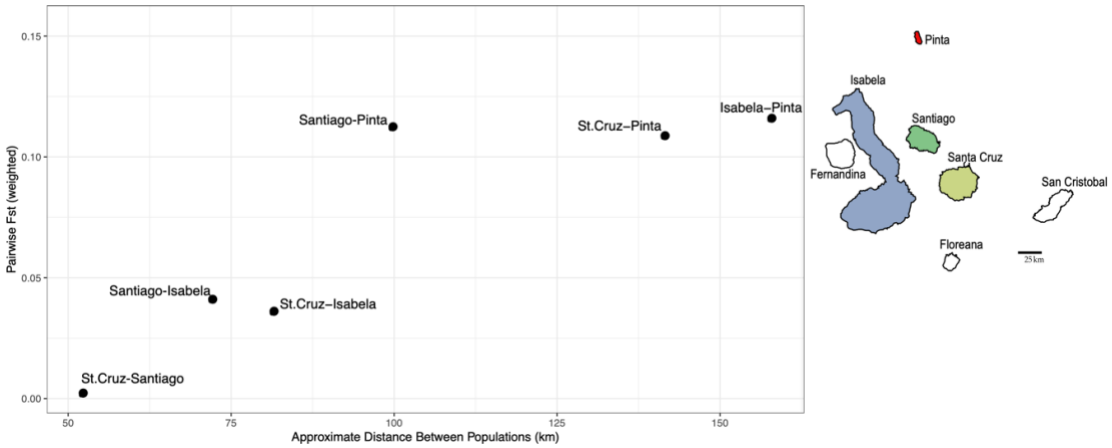

Figure 3-S1. F_{ST} between pairs of populations versus the approximate geographic distance between them. There was some degree of association between F_{ST} and geographical distance. Specifically, values of F_{ST} increased with distance between populations. Rails from Pinta were the most genetically differentiated population relative to other islands (see high F_{ST} values), consistent with Pinta being the most geographically isolated of the four different islands analyzed (Isabela, Santiago, Santa Cruz and Pinta; top right map).

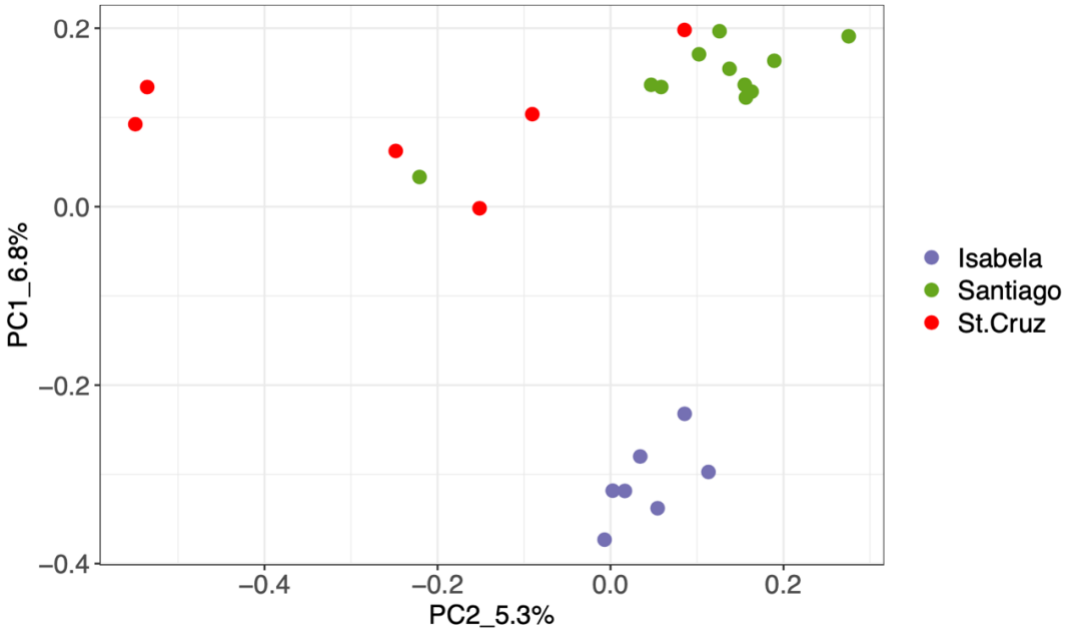

Figure 3-S2. Principal component analysis (PCA) of Galapagos rail samples. The Pinta population was excluded from the original PCA (Figure 3-2a) to further investigate patterns of population structure among samples from Santiago and Santa Cruz. PCA between Santiago and Santa Cruz indicated low levels of population structure between these islands. These islands were also not separated in the fastSTRUCTURE analysis (Figure 3-2c and 3-S3a). The island of Isabela forms a different cluster along PC1.

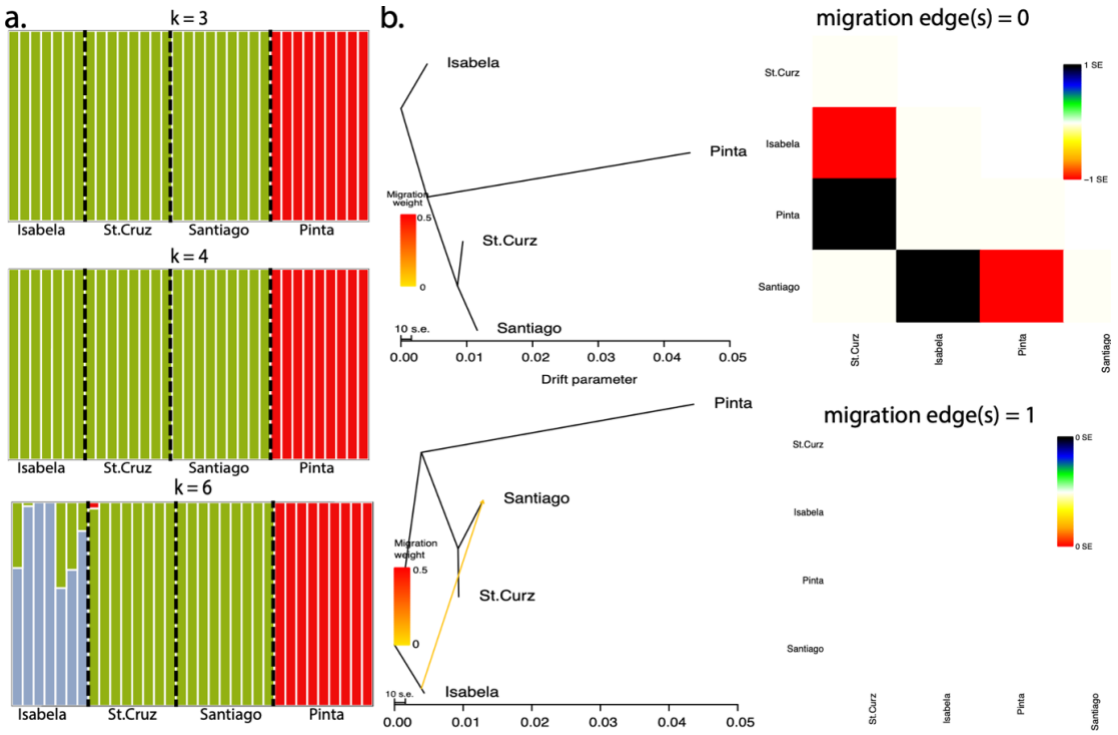

Figure 3-S3

(a) Genetic structure analyses obtained from fastStructure. Different numbers of groups (k = 3, 4, 6) are shown. k=2 and k = 5 are shown in Figure 3-2c. Each column is a single individual, and colors indicate different genetic ancestry groups. Additional groups did not improve the fit to the data. Dotted vertical lines delimit island samples **(b)** TreeMix models and residuals with 0 and 1 migration edges (resulting TreeMix tree with one migration edge also shown in Figure 3-2d). For plots of migration edges on the right, the top right panel suggests that a model with no migration had a positive migration edge standard error of 1. Positive residuals indicate population pairs for which covariance is underestimated by the model, indicating that the model could potentially be improved by adding a migration edge between those populations (Pickrell and Pritchard 2012). When migration was added to the model, the migration edge between those populations reduced the standard error of 0, indicating that adding migration to the model improved the fit of the data. This event of admixture from the Isabela to the Santiago population (Figure 3-2d), is consistent with levels of admixture found at k=5 (Figure 3-2c) and k=6 in the fastStructure analysis. Additional migration edges did not improve the fit to the data.

Appendix 3-II: Tables

Table 3-S1. Descriptive statistics of the 39 rail genomes analyzed in this study. Statistics were calculated with Qualimap (Okonechnikov, et al. 2016) on BAM files that were mapped to the Inaccessible Island rail reference genome (Atlantisia rogersi; see methods for details), downloaded from https://www.ncbi.nlm.nih.gov/assembly/GCA_013401215.1. See methods for further details.from https://www.ncbi.nlm.nih.gov/assembly/GCA_013401215.1. See methods for further details.

Common name	Scientific name	Sample ID	Reference	Mean depth coverage (X)	% of reference covered by at least 3 reads
Galapagos rail	*Laterallus spilonota*	LS02	This study	77.7	96.64
Galapagos rail	*Laterallus spilonota*	LS03	This study	27,01	96.71
Galapagos rail	*Laterallus spilonota*	LS05	This study	41.96	96.7
Galapagos rail	*Laterallus spilonota*	LS06	This study	77.7	96.44
Galapagos rail	*Laterallus spilonota*	LS08	This study	77.7	96.74
Galapagos rail	*Laterallus spilonota*	LS09	This study	40.4	96.71
Galapagos rail	*Laterallus spilonota*	LS10	This study	73.5	96.53
Galapagos rail	*Laterallus spilonota*	LS12	This study	63.6	96.64
Galapagos rail	*Laterallus spilonota*	LS13	This study	77.7	96.57
Galapagos rail	*Laterallus spilonota*	LS14	This study	77.7	96.74
Galapagos rail	*Laterallus spilonota*	LS15	This study	80.4	96.67
Galapagos rail	*Laterallus spilonota*	LS21	This study	31.01	96.41
Galapagos rail	*Laterallus spilonota*	LS22	This study	11.78	96.44
Galapagos rail	*Laterallus spilonota*	LS25	This study	77.7	96.57
Galapagos rail	*Laterallus spilonota*	LS26	This study	18	96.63
Galapagos rail	*Laterallus spilonota*	LS27	This study	77.7	96.67
Galapagos rail	*Laterallus spilonota*	LS28	This study	63.6	96.53
Galapagos rail	*Laterallus spilonota*	LS29	This study	25.49	96.49
Galapagos rail	*Laterallus spilonota*	LS31	This study	77.7	96.65
Galapagos rail	*Laterallus spilonota*	LS32	This study	63.6	96.74
Galapagos rail	*Laterallus spilonota*	LS33	This study	63.6	93.65
Galapagos rail	*Laterallus spilonota*	LS34	This study	40.15	96.73
Galapagos rail	*Laterallus spilonota*	LS35	This study	31.65	96.86
Galapagos rail	*Laterallus spilonota*	LS36	This study	80.4	96.78
Galapagos rail	*Laterallus spilonota*	LS37	This study	77.7	96.84
Galapagos rail	*Laterallus spilonota*	LS38	This study	80.4	96.69
Galapagos rail	*Laterallus spilonota*	LS39	This study	80.4	96.64
Galapagos rail	*Laterallus spilonota*	LS40	This study	80.4	96.96
Galapagos rail	*Laterallus spilonota*	LS43	This study	27.33	96.64
Galapagos rail	*Laterallus spilonota*	LS44	This study	80.4	96.65
Galapagos rail	*Laterallus spilonota*	LS49	This study	31.65	96.66
Galapagos rail	*Laterallus spilonota*	LS50	This study	77.7	96.69
Galapagos rail	*Laterallus spilonota*	LS51	This study	63.6	96.68
Galapagos rail	*Laterallus spilonota*	LS52	This study	63.6	96.64
Galapagos rail	*Laterallus spilonota*	LS55	This study	80.4	96.67
Galapagos rail	*Laterallus spilonota*	LS56	This study	77.7	96.64
Galapagos rail	*Laterallus spilonota*	LS57	This study	40.15	96.73
Galapagos rail	*Laterallus spilonota*	LS58	This study	63.6	96.74
Galapagos rail	*Laterallus spilonota*	LS59	This study	77.7	96.99
Galapagos rail	*Laterallus spilonota*	LS60	This study	80.4	96.84

Table 3-S2. Descriptive statistics of ancestral species distribution models using BioGeoBEARS. The log-likelihood of the different models for Galapagos rails are shown in the column labeled LnL. Each model contained a specific number of parameters. The likelihood of a particular parameter is shown in columns d, e, and j. The best model fit was assessed with the corrected Akaike information criterion (AICc). The best-fitting model is highlighted in bold text. The parameters are: d = rate of range-expansion dispersal, e = rate of range-contraction/local extirpation and j=founder event. Notice that the BAYAREALIKE model has a greater "e" rate, which suggests that ancestral distribution was larger.

model	LnL	numparams	d	e	j	AICc	AICc_wt_vBest
DEC	-35.32	2	0.026	0.032	0	74.99	2.70E-05
DEC+J	-25.34	3	1.00E-12	1.00E-12	0.026	57.39	0.18
DIVALIKE	-34.04	2	0.033	0.047	0	72.42	9.60E-05
DIVALIKE+J	-23.82	3	1.00E-12	1.00E-12	0.026	54.34	0.81
BAYAREALIKE	**-65.97**	**2**	**0.015**	**0.42**	**0**	**136.3**	**1.30E-18**
BAYAREALIKE+J	-28.16	3	1.00E-07	1.00E-07	0.039	63.02	0.011

Table 3-S3. Marginal likelihood from 10 different partitions from fasStructure. Although k =2 indicates the best likelihood, estimates were similar among different partitions. Values of k from 2-6 are shown in Figure 3-2c and 3-S3. Additional k values did not improve the patterns of genetic clustering of Galapagos rails.

Partition(k)	Marginal_Likelihood
1	-1.154131881
2	**-1.133807341**
3	-1.133812405
4	-1.13381443
5	-1.144883074
6	-1.144889645
7	-1.133818013
8	-1.133818261
9	-1.1338189
10	-1.1338201

Table 3-S4. Genome-wide heterozygosity and Runs of homozygosity (ROH) of Galapagos rails. Three different length categories of ROH were used: short (<0.5MB), medium (0.5-2Mb), and long (> 2MB).

Sample	Island	Heterozygosity (kb)	Shot ROH (Mb)	Medium ROH (Mb)	Long ROH (Mb)
LS02	Santa_Cruz	0.10	11.60	10.856856	0
LS03	Santa_Cruz	0.09	30.70	35.13668	2.671282
LS05	Santa_Cruz	0.10	14.67	14.701296	2.280974
LS06	Santa_Cruz	0.10	24.41	29.747538	0
LS08	Santa_Cruz	0.10	27.74	28.392049	12.374174
LS09	Santa_Cruz	0.10	12.62	16.071947	6.886542
LS10	Santa_Cruz	0.10	24.41	37.092603	10.610387
LS12	Santa_Cruz	0.10	19.13	5.253326	2.205742
LS13	Santa_Cruz	0.10	22.88	17.179904	4.553277
LS14	Santa_Cruz	0.10	14.04	4.407426	0
LS15	Santa_Cruz	0.09	21.48	37.867717	6.680367
LS21	Isabela	0.10	18.58	16.432328	0
LS22	Isabela	0.10	14.69	13.210874	0
LS25	Isabela	0.10	18.32	18.984359	0
LS26	Isabela	0.10	14.74	13.576868	0
LS27	Isabela	0.10	22.27	25.487611	4.923703
LS28	Isabela	0.10	12.88	6.656418	5.452212
LS29	Isabela	0.10	23.16	17.07272	0
LS31	Isabela	0.10	17.09	21.749583	2.005268
LS32	Pinta	0.08	47.01	75.433514	15.998257
LS33	Pinta	0.08	51.59	68.448622	12.321384
LS34	Pinta	0.08	52.84	57.851335	26.674368
LS35	Pinta	0.08	57.38	78.066133	14.129899
LS36	Pinta	0.08	50.05	68.175366	7.4409
LS37	Pinta	0.08	50.43	55.199905	4.968104
LS38	Pinta	0.08	52.11371	68.982898	13.415822
LS39	Pinta	0.08	53.143151	69.915789	10.106457
LS40	Pinta	0.08	57.079946	72.798313	7.783014
LS43	Santiago	0.10	24.365117	33.180196	5.775733
LS44	Santiago	0.10	23.535988	14.542941	2.590519
LS49	Santiago	0.10	19.627892	22.87003	0
LS50	Santiago	0.10	39.338068	82.321832	15.448096
LS51	Santiago	0.10	18.18985	18.552248	3.678755
LS52	Santiago	0.10	18.712219	20.127467	2.353979
LS55	Santiago	0.10	17.852014	8.408748	0
LS56	Santiago	0.10	11.916609	6.52217	0
LS57	Santiago	0.10	22.950244	16.794226	6.121631
LS58	Santiago	0.10	24.330224	20.23683	0
LS60	Santiago	0.10	24.403796	29.702631	0

References

ARBOGAST, B. S., DROVETSKI, S. V., CURRY, R. L., BOAG, P. T., SEUTIN, G., GRANT, P. R., GRANT, B. R. & ANDERSON, D. J. 2006. The origin and diversification of Galapagos mockingbirds. *Evolution,* 60**,** 370-382.

BEICHMAN, A. C., KOEPFLI, K. P., LI, G., MURPHY, W., DOBRYNIN, P., KLIVER, S., TINKER, M. T., MURRAY, M. J., JOHNSON, J., LINDBLAD-TOH, K., KARLSSON, E. K., LOHMUELLER, K. E. & WAYNE, R. K. 2019. Aquatic Adaptation and Depleted Diversity: A Deep Dive into the Genomes of the Sea Otter and Giant Otter. *Molecular Biology and Evolution,* 36**,** 2631-2655.

BLACK, J. 1974. Galápagos. Archipelago del Ecuador. Quito: Imprenta Europa.

BURGA, A., WANG, W. G., BEN-DAVID, E., WOLF, P. C., RAMEY, A. M., VERDUGO, C., LYONS, K., PARKER, P. G. & KRUGLYAK, L. 2017. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. *Science,* 356**,** 921-+.

CACCONE, A., GENTILE, G., GIBBS, J. P., FRITTS, T. H., SNELL, H. L., BETTS, J. & POWELL, J. R. 2002. Phylogeography and history of giant Galapagos tortoises. *Evolution,* 56**,** 2052-2066.

CAMPBELL, K., DONLAN, C. J., CRUZ, F. & CARRION, V. 2004. Eradication of feral goats Capra hircus from Pinta Island, Galapagos, Ecuador. *Oryx,* 38**,** 328-333.

CARMI, O., WITT, C. C., JARAMILLO, A. & DUMBACHER, J. P. 2016. Phylogeography of the Vermilion Flycatcher species complex: Multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galapagos-endemic bird species. *Molecular Phylogenetics and Evolution,* 102**,** 152-173.

CARRION, V., DONLAN, C. J., CAMPBELL, K. J., LAVOIE, C. & CRUZ, F. 2011. Archipelago-Wide Island Restoration in the Galapagos Islands: Reducing Costs of Invasive Mammal Eradication Programs and Reinvasion Risk. *Plos One,* 6.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution,* 17**,** 540-552.

CASTRO, M. 1969. Puerto Ayora, Ecuador: Charles Darwin Research Station.

CEBALLOS, F. C., JOSHI, P. K., CLARK, D. W., RAMSAY, M. & WILSON, J. F. 2018. Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics,* 19**,** 220-+.

CHARLESWORTH, D. & WILLIS, J. H. 2009. The genetics of inbreeding depression. *Nature reviews. Genetics,* 10**,** 783-96.

CHAVES, J. A., MARTINEZ-TORRES, P. J., DEPINO, E. A., ESPINOZA-ULLOA, S., GARCIA-LOOR, J., BEICHMAN, A. C. & STERVANDER, M. 2020. Evolutionary

History of the Galapagos Rail Revealed by Ancient Mitogenomes and Modern Samples. *Diversity-Basel,* 12.

CINGOLANI, P., PLATTS, A., WANG, L. L., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. Y. & RUDEN, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly,* 6**,** 80-92.

DIAMOND, J. M. 1984. Historic extinctions: their mechanisms, and les sons for understanding prehistoric extinctions. *In:* MARTIN, P. S. & KLEIN, R. (eds.). Tucson: Quartenary extinctions. University of Arizona Press.

DOBRYNIN, P., LIU, S. P., TAMAZIAN, G., XIONG, Z. J., YURCHENKO, A. A., KRASHENINNIKOVA, K., KLIVER, S., SCHMIDT-KUNTZEL, A., KOEPFLI, K. P., JOHNSON, W., KUDERNA, L. F. K., GARCIA-PEREZ, R., DE MANUEL, M., GODINEZ, R., KOMISSAROV, A., MAKUNIN, A., BRUKHIN, V., QIU, W. L., ZHOU, L., LI, F., YI, J., DRISCOLL, C., ANTUNES, A., OLEKSYK, T. K., EIZIRIK, E., PERELMAN, P., ROELKE, M., WILDT, D., DIEKHANS, M., MARQUES-BONET, T., MARKER, L., JUNWANG, J. B., WANG, J., ZHANG, G. J. & O'BRIEN, S. J. 2015. Genomic legacy of the African cheetah, Acinonyx jubatus. *Genome Biology,* 16.

DONLAN, C. J., CAMPBELL, K., CABRERA, W., LAVOIE, C., CARRION, V. & CRUZ, F. 2007. Recovery of the Galapagos rail (Laterallus spilonotus) following the removal of invasive mammals. *Biological Conservation,* 138**,** 520-524.

DUNCAN, R. P. & BLACKBURN, T. M. 2004. Extinction and endemism in the New Zealand avifauna. *Global Ecology and Biogeography,* 13**,** 509-517.

FRANKLIN, A. B., CLARK, D. A. & CLARK, D. B. 1979. ECOLOGY AND BEHAVIOR OF THE GALAPAGOS RAIL. *Wilson Bulletin,* 91**,** 202-221.

GARCIA-R, J. C. & MATZKE, N. J. 2021. Trait-dependent dispersal in rails (Aves: Rallidae): Historical biogeography of a cosmopolitan bird clade. *Molecular Phylogenetics and Evolution,* 159**,** 107106.

GEIST, D. J., SNELL, H., GODDARD, C. & KURZ, M. D. 2014. A Paleogeographic Model of the Galapagos Islands and Biogeographical and Evolutionary Implications. *Galapagos: a Natural Laboratory for the Earth Sciences,* 204**,** 145-166.

GIBBS, J. P., SHRIVER, W. G. & VARGAS, H. 2003. An assessment of a Galapagos Rail population over thirteen years (1986 to 2000). *Journal of Field Ornithology,* 74**,** 136-140.

GRANT, P. & GRANT, B. 2008. How and Why Species Multiply: The Radiation of Darwin's Finches. Princeton; Oxford: Princeton University Press.

HAMANN, O. 1979. REGENERATION OF VEGETATION ON SANTA-FE AND PINTA ISLANDS, GALAPAGOS, AFTER THE ERADICATION OF GOATS. *Biological Conservation,* 15**,** 215-236.

HOECK, H. N. 1984. Introduced fauna. *In:* PERRY, R. (ed.) *Key environments: Galápagos.* Oxford: Pergamon Press.

HOGG, J. T., FORBES, S. H., STEELE, B. M. & LUIKART, G. 2006. Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B-Biological Sciences,* 273**,** 1491-1499.

HOLLAND, B. S. & HADFIELD, M. G. 2004. Origin and diversification of the endemic Hawaiian tree snails (Achatinellidae : Achatinellinae) based on molecular evidence. *Molecular Phylogenetics and Evolution,* 32**,** 588-600.

JOHNSON, W. E., ONORATO, D. P., ROELKE, M. E., LAND, E. D., CUNNINGHAM, M., BELDEN, R. C., MCBRIDE, R., JANSEN, D., LOTZ, M., SHINDLE, D., HOWARD, J., WILDT, D. E., PENFOLD, L. M., HOSTETLER, J. A., OLI, M. K. & O'BRIEN, S. J. 2010. Genetic Restoration of the Florida Panther. *Science,* 329**,** 1641-1645.

KARDOS, M., AKESSON, M., FOUNTAIN, T., FLAGSTAD, O., LIBERG, O., OLASON, P., SAND, H., WABAKKEN, P., WIKENROS, C. & ELLEGREN, H. 2018. Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution,* 2**,** 124-131.

KARNAUSKAS, K. B., MITTELSTAEDT, E. & MURTUGUDDE, R. 2017. Paleoceanography of the eastern equatorial Pacific over the past 4 million years and the geologic origins of modern Galapagos upwelling. *Earth and Planetary Science Letters,* 460**,** 22-28.

KRAMER, P. & BLACK, J. 1970. Scientific and Conservation Report No. 21. Galapagos, Ecuador: Charles Darwin Research Station.

LI, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

LYNCH, M., CONERY, J. & BURGER, R. 1995. MUTATION ACCUMULATION AND THE EXTINCTION OF SMALL POPULATIONS. *American Naturalist,* 146**,** 489-518.

LYNCH, M. & GABRIEL, W. 1990. MUTATION LOAD AND THE SURVIVAL OF SMALL POPULATIONS. *Evolution,* 44**,** 1725-1737.

MATZKE, N. J. 2013. BioGeoBEARS: biogeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. Berkeley, CA: University of California, Berkeley.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20**,** 1297-1303.

OKONECHNIKOV, K., CONESA, A. & GARCIA-ALCALDE, F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics,* 32**,** 292-294.

OLSON, S. L. 1973. Evolution of the rails of the South Atlantic islands (Aves: Rallidae). Smithsonian Contributions to Zoology.

OLSON, S. L. & JAMES, H. F. 1982. FOSSIL BIRDS FROM THE HAWAIIAN-ISLANDS - EVIDENCE FOR WHOLESALE EXTINCTION BY MAN BEFORE WESTERN CONTACT. *Science,* 217**,** 633-635.

PARENT, C. E. & CRESPI, B. J. 2006. Sequential colonization and diversification of Galapagos endemic land snail genus Bulimulus (Gastropoda, Stylommatophora). *Evolution,* 60**,** 2311-2328.

PETERSON, R. O., VUCETICH, J. A., BUMP, J. M. & SMITH, D. W. 2014. Trophic Cascades in a Multicausal World: Isle Royale and Yellowstone. *Annual Review of Ecology, Evolution, and Systematics, Vol 45,* 45**,** 325-+.

PICKRELL, J. K. & PRITCHARD, J. K. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *Plos Genetics,* 8.

POULAKAKIS, N., MILLER, J. M., JENSEN, E. L., BEHEREGARAY, L. B., RUSSELLO, M. A., GLABERMAN, S., BOORE, J. & CACCONE, A. 2020. Colonization history of Galapagos giant tortoises: Insights from mitogenomes support the progression rule. *Journal of Zoological Systematics and Evolutionary Research,* 58**,** 1262-1275.

POULAKAKIS, N., RUSSELLO, M., GEIST, D. & CACCONE, A. 2012. Unravelling the peculiarities of island life: vicariance, dispersal and the diversification of the extinct and extant giant Galapagos tortoises. *Molecular Ecology,* 21**,** 160-173.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A. R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. W., DALY, M. J. & SHAM, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics,* 81**,** 559-575.

QUESADA, V., FREITAS-RODRIGUEZ, S., MILLER, J., PEREZ-SILVA, J. G., JIANG, Z. F., TAPIA, W., SANTIAGO-FERNANDEZ, O., CAMPOS-IGLESIAS, D., KUDERNA, L. F. K., QUINZIN, M., ALVAREZ, M. G., CARRERO, D., BEHEREGARAY, L. B., GIBBS, J. P., CHIARI, Y., GLABERMAN, S., CIOFI, C., ARAUJO-VOCES, M., MAYORAL, P., ARANGO, J. R., TAMARGO-GOMEZ, I., ROIZ-VALLE, D., PASCUAL-TORNER, M., EVANS, B. R., EDWARDS, D. L., GARRICK, R. C., RUSSELLO, M. A., POULAKAKIS, N., GAUGHRAN, S. J., RUEDA, D. O., BRETONES, G., MARQUES-BONET, T., WHITE, K. P., CACCONE, A. & LOPEZ-OTIN, C. 2019. Giant tortoise genomes provide insights into longevity and age-related disease. *Nature Ecology & Evolution,* 3**,** 87-95.

RAJ, A., STEPHENS, M. & PRITCHARD, J. K. 2014. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics,* 197**,** 573-U207.

ROBINSON, J. A., RAIKKONEN, J., VUCETICH, L. M., VUCETICH, J. A., PETERSON, R. O., LOHMUELLER, K. E. & WAYNE, R. K. 2019. Genomic signatures of extensive

inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances,* 5.

ROSENBERG, D. K. 1990. The impact of introduced herbivores on the Galapagos rail (Laterallus spilonotus). 32.

SCHWARTZ, D. M., SOULE, S. A., WANIESS, V. D. & JONES, M. R. 2018. Identification of Erosional Terraces on Seamounts: Implications for Interisland Connectivity and Subsidence in the Galapagos Archipelago. *Frontiers in Earth Science,* 6.

SENDELL-PRICE, A. T., RUEGG, K. C., ROBERTSON, B. C. & CLEGG, S. M. 2021. An island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology*.

SLIKAS, B., OLSON, S. L. & FLEISCHER, R. C. 2002. Rapid, independent evolution of flightlessness in four species of Pacific Island rails (Rallidae): an analysis based on mitochondrial sequence data. *Journal of Avian Biology,* 33**,** 5-14.

SMITH, T. B., KINNISON, M. T., STRAUSS, S. Y., FULLER, T. L. & CARROLL, S. P. 2014. Prescriptive Evolution to Conserve and Manage Biodiversity. *Annual Review of Ecology, Evolution, and Systematics, Vol 45,* 45**,** 1-+.

SNOW, D. W. 1964. The giant tortoises of the Galapagos Islands: their present status and future chances. 7.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-1313.

STEADMAN, D. W. 1995. PREHISTORIC EXTINCTIONS OF PACIFIC ISLAND BIRDS - BIODIVERSITY MEETS ZOOARCHAEOLOGY. *Science,* 267**,** 1123-1131.

STEINFARTZ, S., GLABERMAN, S., LANTERBECQ, D., RUSSELLO, M. A., ROSA, S., HANLEY, T. C., MARQUEZ, C., SNELL, H. L., SNELL, H. M., GENTILE, G., DELL'OLMO, G., POWELL, A. M. & CACCONE, A. 2009. Progressive colonization and restricted gene flow shape island-dependent population structure in Galapagos marine iguanas (Amblyrhynchus cristatus). *Bmc Evolutionary Biology,* 9.

STERVANDER, M., RYAN, P. G., MELO, M. & HANSSON, B. 2019. The origin of the world's smallest flightless bird, the Inaccessible Island Rail Atlantisia rogersi (Aves: Rallidae). *Molecular Phylogenetics and Evolution,* 130**,** 92-98.

VAGVOLGYI, J. 1974. Pinta tortoise: rediscovered. 27.

WAYNE, R. K., LEHMAN, N., GIRMAN, D., GOGAN, P. J. P., GILBERT, D. A., HANSEN, K., PETERSON, R. O., SEAL, U. S., EISENHAWER, A., MECH, L. D. & KRUMENAKER, R. J. 1991. CONSERVATION GENETICS OF THE ENDANGERED ISLE-ROYALE GRAY WOLF. *Conservation Biology,* 5**,** 41-51.

WEBER, D. 1971. Pinta, Galápagos: Une ile à sauver. *Biological Conservation,* 4**,** 8-12.

WHITE, W. M., MCBIRNEY, A. R. & DUNCAN, R. A. 1993. PETROLOGY AND GEOCHEMISTRY OF THE GALAPAGOS-ISLANDS - PORTRAIT OF A PATHOLOGICAL MANTLE PLUME. *Journal of Geophysical Research-Solid Earth,* 98**,** 19533-19563.

WRIGHT, N. A., STEADMAN, D. W. & WITT, C. C. 2016. Predictable evolution toward flightlessness in volant island birds. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** 4765-4770.

ZHANG, C., RABIEE, M., SAYYARI, E. & MIRARAB, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *Bmc Bioinformatics,* 19.

ZHENG, X. W., LEVINE, D., SHEN, J., GOGARTEN, S. M., LAURIE, C. & WEIR, B. S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics,* 28**,** 3326-3328.