PUCRS

ESCOLA DE CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E EVOLUÇÃO DA BIODIVERSIDADE
DOUTORADO EM ECOLOGIA E EVOLUÇÃO DA BIODIVERSIDADE

GUSTAVO PABLO LORENZANA PIÑA

POPULATION GENOMICS OF JAGUARS (PANTHERA ONCA): COMPARATIVE ASSESSMENT OF DIVERSITY ON DIFFERENT GENOMIC AND SPATIO-TEMPORAL SCALES

Porto Alegre

2019

PÓS-GRADUAÇÃO - *STRICTO SENSU*

Pontifícia Universidade Católica
do Rio Grande do Sul

Pontifícia Universidade Católica do Rio Grande do Sul

Programa de Pós-Graduação em Ecologia e Evolução da Biodiversidade

Population genomics of jaguars (*Panthera onca*): comparative assessment of diversity on different genomic and spatio-temporal scales

Gustavo Pablo Lorenzana Piña

Tese de Doutorado

Porto Alegre

2019

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL

ESCOLA DE CIENCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E EVOLUÇÃO DA BIODIVERSIDADE

Population genomics of jaguars (*Panthera onca*): comparative assessment of diversity on different genomic and spatio-temporal scales

Gustavo Pablo Lorenzana Piña

Orientador: Dr. Eduardo Eizirik

TESE DE DOUTORADO

PORTO ALEGRE – BRASIL

SUMÁRIO

“There are as many atoms in a single molecule of your DNA
as there are stars in the typical galaxy.
We are, each of us, a little universe.”

Neil deGrasse Tyson

“The visions we offer our children shape the future.
It matters what those visions are.
Often they become self-fulfilling prophecies.
Dreams are maps.”

Carl Sagan

AGRADECIMENTOS/ACKNOWLEDGMENTS

I am greatly indebted to several people. First, I want to thank Eduardo Eizirik for giving me the opportunity to start in the field of wildlife genetics. The learning curve was very steep but most enjoyable due to his great support and encouragement. Thanks a lot Duda!

Thanks to all the colleagues and friends at Lab Genoma, PUCRS, for the their assistance and camaraderie, especially to Henrique Figueiró, Fernando Lopes, Fabricio Garcez, Isadora Franca, Fernanda Trindade, Sarah Santos and Manuel Riveros. From the core area of the *Green Corridor*, Marcia Kaim helped me a lot with logistical support and a close friendship throughout my studies.

Professors Sandro Bonatto, Nelson Fontoura and Júlio César Bicca-Marques from PUCRS, contributed with useful comments and suggestions that helped to greatly improve this dissertation. Pavel Dobrynin and Marty Kardos provided valuable insights to perform genomic analyses.

Special thanks to all the researchers that were involved in fieldwork, sample collection and lab procedures, Taiana Haag, Laura Heidtmann Emiliano Ramalho, Guilherme Dias, Tomas Hrbek, Izeni Farias, Priscilla Villela, Luiz Coutinho, and Oliver Ryder.

Ronaldo Morato, Nelson Fagundes, and Klaus-Peter Koepfli kindly accepted to be part of the examination committee and their comments and suggestions fostered a fruitful discussion that helped to improve the quality of final drafts for submission.

I thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for granting me the scholarship number 141123/2015-0 to develop my studies.

Last but not least, I want to thank my family for all the love and support, at a merely distance of 9,000 km. Gracias Pá, Má, Fridita & Leo.

ABSTRACT

The jaguar is the largest predator in the Neotropics and an iconic species in many Native American cultures. It is the only extant representative of genus *Panthera* in the western hemisphere, and is the target of considerable attention from the scientific and conservation communities, given its threatened status across its continental range. Due to habitat loss and direct human persecution, it has already lost over half of its historical range, and some of its remaining populations are isolated and critically endangered. The jaguar has been the focus of several genetic studies and was the first large Neotropical mammal to have its genome sequenced. However, many outstanding questions remain regarding its genetic diversity, population structure and evolutionary history. For example, no genetic study has investigated diversity or structure of jaguar populations across the Amazon region, a major stronghold for the species and an important baseline against which other biomes can be compared. In addition, as jaguar genetic studies transition to genome-wide approaches, an important issue is to assess the performance of different methods, such as alternative strategies to generate and sequence reduced-representation libraries. Such comparisons are still rare in the literature, and jaguar datasets offer a useful opportunity for such an assessment. Finally, as the Jaguar Genome Project moves forward and begins to include population genomic studies, it is relevant to assess the potential of whole genome sequences generated from multiple individuals to investigate the historical demography of different populations, and their power to inform conservation efforts on behalf of this species. This dissertation addresses these three topics, each of which constitutes the focus of a scientific manuscript. In the first study, I employed 11 microsatellite loci to characterize the genetic variability and population structure of Amazonian jaguars, and then performed integrated analyses incorporating previously published data for the same markers collected in the Atlantic Forest and Pantanal biomes. All indices of genetic diversity were consistently higher for the Amazonian population. No genetic subdivision was detected in the Amazon, indicating large-scale connectivity across a sampled area spanning more than three thousand kilometers. We observed that the Atlantic Forest as a whole still retains considerable levels of genetic diversity, but this is currently partitioned among fragments which are increasingly isolated and subjected to heavy anthropic disturbance. The second study reports the collection of genotyping-by-sequencing (GBS) data from 20 wild jaguars representing five different biomes, and for which whole-exome sequencing (WES) data had already been collected by our group. We performed multiple analyses of both genome-wide datasets, estimating genetic diversity and population differentiation indices, and assessing the impact of different parameter settings on these comparisons. We observed that changes in parametrization led to measurable differences in summary statistics for each jaguar population, both between approaches and among distinct analytical batches within each approach, especially for GBS. Diversity was consistently higher for the Amazonian and Pantanal populations, with the Caatinga exhibiting the

lowest diversity and highest differentiation from other regions. Our results show that some parameters do influence estimates of diversity and differentiation in ways that may not be fully predictable, highlighting the importance of careful fine-tuning of parameters for obtaining robust and unbiased genomic diversity estimates. The third manuscript describes the generation of eight novel complete jaguar genomes, and their analyses jointly with three other genomes representing different geographic regions. These 11 genomes were analyzed using the pairwise sequentially Markovian (PSMC) method, and also characterized in terms of their runs of homozygosity (ROH) content to investigate more recent phases of their demographic history. Our PSMC results were very consistent among individuals, and indicated that jaguar populations have undergone pronounced cycles of demographic fluctuations in the last 1-2 million years. In addition, the Arizona individual stood out in showing a steeper decline in the last 30,000 years, likely as a result of a recent history of founder events at the edge of the species' range. As for the ROH analyses, we found a relatively modest burden of homozygosity across most jaguar populations. However, representatives from the Arizona and Atlantic Forest populations showed signals of recent bottlenecks and, in the latter case, inbreeding. These results demonstrate the potential of genome-wide datasets to investigate jaguar demographic history in unprecedented detail, and open up new avenues for conservation genetic efforts targeting this species. Overall, the three studies contained in this dissertation illustrate the use of different types of markers (from traditional microsatellites to whole-genome sequences) and analyses targeting different spatial and temporal scales, to characterize the evolutionary history of a flagship carnivore. Hopefully these studies will contribute to enhance our understanding of jaguar biology and evolution, and provide useful information to be incorporated into conservation efforts on its behalf.

Keywords: Carnivora, Felidae, endangered species, genetic diversity, population structure, demographic history, microsatellite loci, Single Nucleotide Polymorphism (SNP).

RESUMO

A onça-pintada é o maior predador dos Neotrópicos e uma espécie icônica em muitas culturas nativas americanas. É o único representante existente do gênero *Panthera* no hemisfério ocidental, e é alvo de considerável atenção por parte das comunidades científica e de conservação, devido ao seu status ameaçado em toda a sua extensão continental. Devido à perda de habitat e à perseguição humana direta, ela já perdeu mais da metade de sua área histórica, e algumas de suas populações remanescentes estão isoladas e criticamente ameaçadas de extinção. A onça-pintada tem sido foco de vários estudos genéticos e foi o primeiro grande mamífero Neotropical a ter seu genoma sequenciado. No entanto, muitas questões permanecem pendentes quanto à sua diversidade genética, estrutura populacional e história evolutiva. Por exemplo, nenhum estudo genético investigou a diversidade ou estrutura das populações de onça-pintada em toda a região amazônica, um importante reduto para as espécies e uma linha de base importante na qual outros biomas podem ser comparados. Além disso, como os estudos genéticos da onça-pintada transitam para abordagens genômicas, uma questão importante é avaliar o desempenho de diferentes métodos, como estratégias alternativas para gerar e sequenciar bibliotecas de representação reduzida. Tais comparações ainda são raras na literatura, e os conjuntos de dados de onça-pintada oferecem uma oportunidade útil para tal avaliação. Finalmente, à medida que o Projeto Genoma Jaguar avança e começa a incluir estudos genômicos populacionais, é relevante avaliar o potencial de sequências de genomas completos geradas para múltiplos indivíduos para investigar a demografia histórica de diferentes populações e seu poder de informar os esforços de conservação em favor da população desta espécie. Esta dissertação aborda esses três tópicos, cada um dos quais constitui o foco de um manuscrito científico. No primeiro estudo, empreguei 11 loci microssatélites para caracterizar a variabilidade genética e a estrutura populacional de onças-pintadas na Amazônia, e então realizei análises integradas incorporando dados previamente publicados para os mesmos marcadores coletados nos biomas Mata Atlântica e Pantanal. Todos os índices de diversidade genética foram consistentemente maiores para a população amazônica. Nenhuma subdivisão genética foi detectada na Amazônia, indicando conectividade em grande escala em uma área amostrada que abrange mais de três mil quilômetros. Observamos que a Mata Atlântica como um todo ainda mantém níveis consideráveis de diversidade genética, mas isso é atualmente particionado entre fragmentos cada vez mais isolados e sujeitos a fortes perturbações antrópicas. O segundo estudo relata a coleta de dados de genotipagem-por-seqüenciamento (GBS) de 20 onças-pintadas representando cinco diferentes biomas, e para os quais dados do seqüenciamento de exoma completo (WES) já haviam sido coletados pelo nosso grupo. Realizamos análises múltiplas de ambos os conjuntos de dados genômicos, estimando a diversidade genética e os índices de diferenciação populacional, e avaliando o impacto de diferentes configurações de parâmetros nessas comparações. Observamos que mudanças na parametrização levaram a diferenças mensuráveis nas estatísticas

sumarias de cada população de onça-pintada, tanto entre abordagens quanto entre lotes analíticos distintos dentro de cada abordagem, especialmente para GBS. A diversidade foi consistentemente maior para as populações da Amazônia e do Pantanal, com a Caatinga exibindo a menor diversidade e maior diferenciação de outras regiões. Nossos resultados mostram que alguns parâmetros influenciam as estimativas de diversidade e diferenciação de maneiras que podem não ser totalmente previsíveis, destacando a importância de um cuidadoso ajuste fino dos parâmetros para a obtenção de estimativas robustas e imparciais da diversidade genômica. O terceiro manuscrito descreve a geração de oito novos genomas completos de onça pintada e suas análises em conjunto com outros três genomas representando diferentes regiões geográficas. Esses 11 genomas foram analisados pelo método pareado seqüencial Markoviano (PSMC) e também caracterizados em termos de segmentos de homozigosidade (ROH) para investigar fases mais recentes de sua história demográfica. Nossos resultados do PSMC foram muito consistentes entre os indivíduos e indicaram que as populações de onça-pintada sofreram pronunciados ciclos de flutuações demográficas nos últimos 1-2 milhões de anos. Além disso, o indivíduo do Arizona destacou-se em mostrar um declínio mais acentuado nos últimos 30.000 anos, provavelmente como resultado de uma história recente de eventos de fundadores no limite da área de distribuição da espécie. Quanto às análises de ROH, encontramos uma carga relativamente modesta de homozigosidade na maioria das populações de onça-pintada. No entanto, representantes das populações do Arizona e da Mata Atlântica mostraram sinais de gargalos de garrafa recentes e, no último caso, endogamia. Esses resultados demonstram o potencial de conjuntos de dados genômicos para investigar a história demográfica da onça-pintada em detalhes sem precedentes, e abrem novos caminhos para esforços genéticos de conservação visando essa espécie. No geral, os três estudos contidos nesta dissertação ilustram o uso de diferentes tipos de marcadores (a partir de microssatélites até sequências de genomas completos) e análises dirigidas a diferentes escalas espaciais e temporais, para caracterizar a história evolutiva de um carnívoro emblemática. Esperamos que esses estudos contribuam para melhorar nossa compreensão da biologia e evolução da onça-pintada e forneçam informações úteis para serem incorporadas aos esforços de conservação em seu nome.

Palavras chave: Carnivora, Felidae, espécie ameaçada, diversidade genética, estrutura populacional, história demográfica, loci microssatélite, polimorfismo de nucleotídeo simples (SNP).

APRESENTAÇÃO

A presente tese doutoral está estruturada na forma de artigos científicos (Capítulos II, III e IV), precedidos por uma Introdução geral (Capítulo I). As implicações e limitações dos principais achados são sumarizados na Discussão Geral (Capítulo V). Os artigos científicos estão em fase final de preparação na língua inglesa e serão submetidos aos jornais *Diversity and Distributions* (Cap. I), *Methods in Ecology and Evolution* (Cap II), e *Current Biology* (Cap. III), seguindo as normas editoriais de cada periódico, após a incorporação dos comentários e sugestões emitidas pela banca examinadora.

CHAPTER I – GENERAL INTRODUCTION

Natural phenomena occur at different scales, from the infinitesimal to the infinitely large, from subatomic particles to galaxies. In every order of magnitude, patterns disclose processes, awaiting to be discovered and described. The genome is not an exception. It is a *miniverse* on its own, enclosing the vast biological legacy shared by all organisms. Organisms such as mammals, despite being greatly outnumbered by other taxonomic groups, exert a disproportionate influence on the ecosystems where they occur (Brown & Maurer, 1986), and are extremely complex organisms. The three billion base pairs of a typical mammalian genome (Gregory et al. 2007) encode a vast amount of information than can be mined at different scales, providing insights for specific questions. Biological questions that, as in other scientific endeavors, need to be tackled using ingenuity and perseverance. The ongoing technological revolution is allowing us to explore problems in increasing detail. This is the framework employed throughout this dissertation. From a few traditional, hypervariable genetic markers such as microsatellite loci, to thousands of single nucleotide polymorphisms (SNP), to whole-genome sequences. An armory of genetic markers to parse and unveil the way genetic diversity is structured across space and time, within populations and species. Species such as the jaguar (*Panthera onca*), lord of the night and symbol of might in Pre-Columbian cosmologies (Benson, 1998).

The jaguar is the largest felid species in the Americas, and it is a keystone top-predator whose presence is linked to healthy, productive ecosystems (Miller et al., 2001; Thornton et al., 2015; Morato et al., 2018). It has a muscular body, with strong limbs and neck, and a yellowish coat with dark rosettes (Figure 1), attaining over 100 kg in open habitats such as the Llanos and Pantanal, while smaller individuals are found in forested environments such as the Amazon basin and Central America (Seymour, 1989). The jaguar is the only extant species of big cat belonging to the subfamily Pantherinae occurring in the Western Hemisphere. It diverged from the ancestor of lions and leopards ca. 3.65 million years ago (Mya) in Eurasia (Figure 2), crossing Beringia and colonizing North America by the middle Pleistocene (Kurtén & Anderson, 1980), reaching South America as part of a recent dispersal pulse within the Great American Biotic Interchange (GABI), *ca.* 1.8 Mya (Woodburne et al. 2010).

Figure 1. Male jaguar in the Pantanal. Photo courtesy: Daniel Kantek

Until the first half of the 20th century, jaguars were widely distributed in the western hemisphere, from southwestern United States to northern Argentina, occupying several tropical and subtropical biomes (Sanderson et al. 2002; Zeller, 2007), each of them harboring different abiotic conditions as well as prey, competitor and pathogen ensembles. Since then, threats on the species have steadily worsened due to illicit and legal human activities. Back in the 1960s, the beauty of the jaguar's pelage unleashed a boom in illegal trade of pelts, until their listing on Appendix I of the CITES agreement in 1973 (Smith, 1976; Swank & Teer, 1989). Currently, extensive clearing of primary forests for agriculture and cattle ranching is increasingly destroying large tracts of habitat (Zeilhofer et al., 2014; Paviolo et al., 2016), and the events of predation on livestock generally lead to retaliation poaching or broader persecution (Michalski et al., 2006). These factors have caused global population declines, including extirpation from Uruguay and El Salvador, eradicating the species from 40-55% of its historic range (Sanderson et al. 2002; Zeller, 2007, de la Torre et al., 2018). Most of the remaining populations are becoming increasingly small and isolated, making them prone to further decline, not only by demographic stochastic factors, but also from the effects of genetic drift and inbreeding (Eizirik et al. 2008).

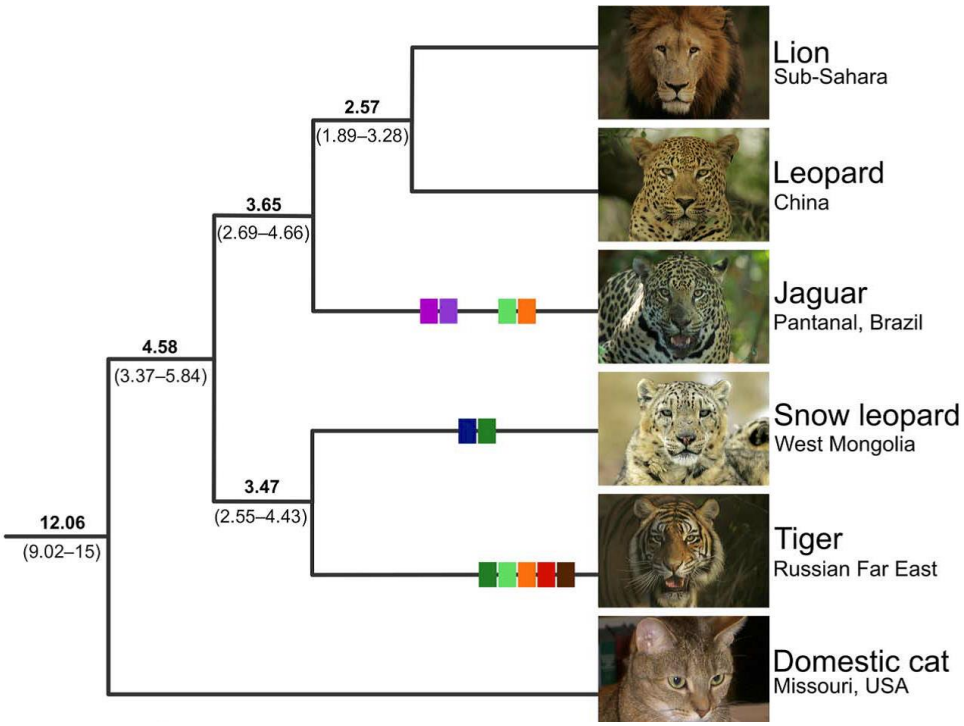


Figure 2. Phylogenetic relationships of the five extant species belonging to the genus *Panthera*. The ancestor of jaguars diverged from the ancestral species that gave rise to the lion and the leopard, about 3.65 Mya. Figure from Figueiró et al. (2017).

As an apex predator, the jaguar has extensive spatial requirements, which turns this animal into a "landscape species", suitable to be regarded as an umbrella taxon in the realm of Conservation Biology (Copolillo et al., 2004; de Barros et al., 2014; Thornton et al., 2014). Studying and protecting this cat would thus help to preserve ecological processes and functions (Ripple et al., 2014). Conversely, local extirpation of this felid is one of the first warning signals of defaunation (Dirzo et al., 2014), leading to empty forest syndromes and trophic cascade effects (Redford, 1992; Wilkie et al., 2011; Jorge et al., 2013). This is especially relevant in tropical ecosystems, which sustain a disproportionate share of global terrestrial biodiversity, including thousands of endemic species (Mittermeier et al., 1998; Jenkins et al., 2013). In this context, development of cutting-edge biological research in those regions is urgently needed. For example, the Amazon basin, despite being the largest, most important global expanse of primary habitat for jaguars (Sanderson et al., 2002; Jędrzejewski et

al. 2018), has been the subject of relatively few ecological and genetic studies (e.g. Tobler et al., 2013; Foster et al., 2013; Roques et al., 2016).

The first studies that applied molecular techniques to investigate jaguar phylogeography and population/conservation genetics (e.g. Eizirik et al. 2001, 2008; Ruiz-Garcia et al. 2006) relied on mitochondrial DNA (mtDNA) and/or microsatellites, and this traditional suite of markers has continued to be employed in almost all studies performed to date. Eizirik et al. (2001) found moderate to high levels of gene diversity and considerably low nucleotide diversity among jaguars sampled across the species' range, without strong population structure, which is possibly due to a recent demographic expansion and high connectivity on broad spatial scales, sustained by long-range dispersal, likely driven by males (Eizirik et al. 2008). Major geographic barriers such as the Amazon River and the Darien straits between northern South America and Central America, were suggested as elements that restricted historical gene flow, producing measurable genetic differentiation among four incompletely isolated phylogeographic groups (South America south of the Amazon; northern South America; Central America; and Guatemala + Mexico).

More recently, Roques et al. (2015), analyzing 11 microsatellite loci typed in individuals from Brazil and Mexico, found a pronounced genetic structure, identifying four genetically differentiated areas. Genetic differentiation was related to geographic distance, but they also found evidence of the effects of habitat degradation on genetic patterns. Within the sampled portion of the Amazon rainforest, jaguars showed high levels of genetic diversity and panmixia across large distances, while the genetic diversity was reduced near the limits of the species' range, likely due to population contractions. Mexican jaguars were highly differentiated from Brazilian jaguars and genetically depauperated. In addition, an isolated population from the Caatinga biome in northeastern Brazil showed the genetic effects of a recent demographic reduction, occurred within the last 20-30 years, which may reflect the region's contemporary habitat deterioration.

Other assessments have added evidence on the jaguar's high sensitivity to habitat loss and fragmentation. In the Brazilian Atlantic Forest, environmental degradation has been severe enough to

promote significant differentiation induced by genetic drift among population remnants (Haag et al. 2010; Valdez et al. 2015, Srbek-Araújo et al., 2018), and a similar, though less severe pattern was detected in Belize (Wultsch et al. 2016). All of these studies stressed the need for further sampling that could reveal finer genetic patterns.

Therefore, this dissertation had the goal of filling gaps of knowledge on jaguar genetics, using a comparative and integrative approach that took advantage of Next-generation-sequencing (NGS), or more properly, High-throughput sequencing (HTS) techniques. I developed the research under a population genetics framework with well-defined ecoregions sustaining jaguar demes. The whole dissertation focused on some of the major and more diverse tropical biomes of South America, the Amazon and Atlantic rainforests, the Pantanal wetland, the Cerrado savanna, and the Caatinga dry forest. The ample, continental scale of this region allowed the creation of a meaningful geographic context for the sets of analyzed samples. In addition, same of the same samples were subjected to more than one genetic technique, which allowed obtaining sets of genetic and genome-wide markers, and even whole-genome sequences, that can be a valuable resource for comparative analyses. The results are presented and discussed in the three subsequent chapters, with a fifth and final chapter summarizing the main insights and limitations, presenting closing remarks, along with future directions given current technological and biological challenges and opportunities.

This dissertation follows up on the genetic and genomic research efforts that have been developed in the last several years at the PUCRS Laboratory of Genomics and Molecular Biology, which culminated in the first complete genome sequenced for a Neotropical mammal (Figueiró et al., 2017). That comprehensive paper characterized the genome of the jaguar through a *de novo* assembly and annotation, and included the detection of genes with signatures of positive selection, as well as the identification of historical introgression with other conspecific members in the genus *Panthera*. I extensively employed that genome assembly as a reference for aligning and mapping the genomic resources that I obtained and assessed as part of population-level analyses described in Chapters 3 and 4.

In Chapter 2, novel data for microsatellite loci, in conjunction with genotypes for the same markers published in three previous studies, were used to perform a comparative assessment of jaguar neutral genetic variability and population structure across the focal biomes. This analysis encompassed the broadest geographic coverage of jaguar genetic samples within the Amazon rainforest up to now, as well as the largest microsatellite dataset so far analyzed for this species.

Chapter 3 constitutes a *sui generis* comparison between two major approaches currently used to reduce genome complexity in large-scale sequencing studies: restriction enzyme associated DNA sequencing (RADseq) and exome target capture. Standard diversity and differentiation metrics were estimated from genome-wide single nucleotide polymorphisms (SNPs) derived from each type of approach using the same jaguar individuals, with emphasis on the Amazon and the Pantanal regions. As far as I could ascertain, this is the first study in which genomic data generated using these two methods were simultaneously analyzed for the same set of individuals.

Chapter 4 harnessed the power of whole-genome resequencing to analyze jaguar demographic history, contemporary bottlenecks and inbreeding signatures. Using complete genomes from 11 different jaguar individuals (eight of which were sequenced specifically for this study), historical demography in different biomes was inferred using the Pairwise Sequentially Markovian Coalescent (PSMC) method, while recent inbreeding was evaluated through the identification of long runs of homozygosity (ROH) along the autosomes. Jointly, these analyses and their underlying dataset have opened up new avenues to investigate jaguar evolution in unprecedented detail, as well to empower the development and refinement of conservation strategies on behalf of this increasingly threatened species.

CHAPTER II – High genetic diversity and large-scale connectivity of jaguars (*Panthera* onca) in their main global stronghold, the Amazon rainforest

High genetic diversity and large-scale connectivity of jaguars (*Panthera onca*) in their largest global stronghold, the Amazon rainforest.

Jaguar population genetics

Gustavo Lorenzana[1], Laura Heidtmann[1], Taiana Haag[1], Emiliano Ramalho[2], Guilherme Dias[3], Tomas Hrbek[3], Izeni Farias[3], Eduardo Eizirik[1,4]

[1] Laboratório de Biologia Genômica e Molecular, Escola de Ciências, PUCRS, Porto Alegre, Brazil

[2] Instituto de Desenvolvimento Sustentável Mamirauá, Tefé, Brazil

[3] Departamento de Genética, UFAM, Manaus, Brazil

[4] Instituto Pró-Carnívoros, Atibaia, Brazil

Acknowledgments

GL, LH and EE acknowledge CNPq/Brazil for financial support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior – Brasil (CAPES) – Finance 18 Code 001.

Abstract

Aim

Jaguar population genetics has so far not been investigated on a broad spatial scale in the Amazon rainforest, which constitutes the largest remaining block of continuous habitat for the species. Given its continuity, it serves not only as a stronghold but also as a reference for jaguar population genetics against which fragmented landscapes can be compared. We thus assessed genetic diversity and structure of Amazonian jaguars and then compared them with data collected in two other major South American biomes in which the species has faced different amounts of habitat loss and fragmentation.

Location

South America: Amazon, Atlantic Forest, Pantanal.

Methods

Using 11 microsatellite loci, we characterized the genetic variability and population structure of Amazonian jaguars, and performed integrated analyses incorporating previously published data for the same markers collected in the Atlantic Forest and Pantanal biomes.

Results

All indices of genetic diversity were consistently higher for the Amazonian population. Allelic richness was 4-fold higher than that of the Pantanal. No genetic subdivision was detected in the Amazon, indicating large-scale connectivity across a sampled area spanning more than three thousand kilometers. No signals of recent population bottlenecks were detected for this deme. We corroborated the inference of anthropic-driven structure and bottlenecks for two Atlantic Forest subpopulations.

Main conclusions

Our results support the view that the Amazon rainforest is a critically important stronghold for jaguars, comprising a relatively large, highly diverse, panmictic population, allowing a glimpse into the patterns of genetic connectivity that characterized this species prior to human intervention. In contrast, the Atlantic Forest populations as a whole still retain considerable levels of genetic diversity, but it is currently partitioned among fragments which are increasingly isolated and subjected to heavy anthropic disturbance. These results have important implications for jaguar conservation planning, as we corroborate the inference that Atlantic Forest populations are in critical condition, and provide a genetic baseline to which they can be compared.

Keywords: fragmentation, population structure, Felidae, tropical rainforest, Pantanal, Atlantic Forest

Introduction

Tropical ecosystems harbour a large proportion of global biological diversity, reaching more than 50% of the world's terrestrial biodiversity (Gardner, Barlow, Sodhi, & Peres, 2010). Increasing human activities on those regions are exerting pressure on the biota, reducing local abundance and causing defaunation, driving thousands of species to extinction even before they are discovered (Dirzo et al., 2014). Habitat loss and fragmentation are two of the main threats to species survival, especially for large carnivores (Crooks, 2002; Costa, Leite, Mendes, & Ditchfield, 2005), such as the jaguar (*Panthera onca*). This felid is the top predator of the Neotropics, and given its keystone role, constitutes an umbrella and flagship species for biodiversity conservation (Thornton et al., 2016). Globally, it is considered 'Near Threatened' by the IUCN (Caso et al., 2008), but it is categorized as Endangered or Vulnerable in most national red lists across its distribution (e.g. ESA, 1973; Rodríguez-Mahecha et al., 2006; SEMARNAT 2010; Aprile et al., 2012).

In Brazil, jaguars currently occur in five out of six major biomes, and their populations are subjected to different threats on a regional basis, making them more vulnerable in some areas than others (Sollmann, Tôrres, & Silveira, 2008; Nijhawan, 2012). It is 'Critically Endangered' in the Atlantic Forest due to a drastic population reduction during the last three decades (Beisiegel, Sana, & Moraes, 2012), as this biome is severely imperilled by habitat loss and fragmentation (Tabarelli, Pinto, Silva, Hirota, & Bedê, 2005). In contrast, the Amazon and the Pantanal, given their extent, habitat suitability and comparatively lower levels of fragmentation, are regarded as the two main strongholds for the jaguar, both nationally and globally, although its status is 'Vulnerable' in both of these biomes (Cavalcanti, Azevedo, Tomás, Boulhosa, & Crawshaw Jr, 2012; de Oliveira, Ramalho, & de Paula, 2012).

The Brazilian portion of the Amazon covers nearly 3.5 million km^2, and it is assumed that jaguars occupy most of this area (de Oliveira et al., 2012). For this reason, this biome is regarded as the most important block of continuous habitat for jaguars, harbouring one of the largest populations of the species, with good perspectives for long-term persistence. Nevertheless, jaguars in this biome are threatened by illegal hunting, and the so-called "arc of deforestation" is advancing on the eastern

79 and southern portions of the region, already representing a loss of 18% of the originally forested area
80 (de Oliveira et al., 2012). Likewise, the Pantanal is one of the largest wetlands in the world,
81 encompassing 140,000 km2 (85% of which remain conserved), with jaguars occupying between 88,000
82 and 125,000 km2 (Cavalcanti et al., 2012). In the Pantanal, retaliatory hunting of jaguars that prey on
83 cattle is the main threat to the species' survival. Interestingly, ecotourism focused on jaguars in this
84 region is currently fifty times more profitable than cattle ranching (Tortato, Izzo, Hoogesteijn, & Peres,
85 2017), which has helped to alleviate the hunting pressure. In sharp contrast, the Atlantic Forest is a
86 biodiversity hotspot with a high degree of endemism (Myers, Mittermeier, Mittermeier, de Fonseca,
87 & Kent, 2000), whose primary cover has been decimated in the last four decades, declining from 1.3
88 million to 150,000 km^2 (Ribeiro, Metzger, Martensen, Ponzoni, & Hirota, 2009). Currently, jaguars
89 occupy less than 50% of this area, persisting in small, isolated fragments in which jaguars also suffer
90 from prey depletion and illegal hunting (Beisiegel et al., 2012; Paviolo et al., 2016).
91 As a large mammalian carnivore, jaguars have high mobility and, as a result, could potentially
92 attain high levels of dispersal and gene flow across the landscape (Tammeleht et al., 2010; Row et al.,
93 2012). However, relatively few molecular studies with jaguars have been published to date. Jaguars
94 have shown moderate to high levels of genetic diversity (Eizirik et al., 2001; Ruiz-Garcia, Payán, Murillo,
95 & Alvarez, 2006), without evidence of strong population structure across their range, possibly due to
96 a recent population expansion and high connectivity on broad spatial scales. Major geographical
97 barriers such as the Amazon River and perhaps the Darien strait were suggested as having restricted
98 historical gene flow among four incompletely isolated phylogeographic groups: southern South
99 America, northern South America, Central America and Mexico-Guatemala (Eizirik et al., 2001).
100 However, the authors of that study stressed the need for further sampling that could reveal a finer
101 pattern of subdivision or isolation by distance on a regional level.
102 In-depth analyses of regional jaguar populations in Brazil initially revealed that a recently
103 fragmented area of the inner Atlantic Forest showed evidence of drift-induced population
104 differentiation and loss of allelic richness, driven by anthropogenic habitat loss and isolation (Haag et

al., 2010). The problem is so severe that one of the sampled populations ("Porto Primavera") was extirpated due to the flooding of a hydroelectric dam before that study was published. Valdez et al. (2015) further analysed these subpopulations in conjunction with jaguars sampled at four sites within the southern Pantanal, and found that the latter region forms a single genetic cluster with higher genetic diversity than each of the Atlantic forest demes. Subsequently, Srbek-Araujo, Haag, Chiarello, Salzano, & Eizirik (2018) analysed an isolated population from the coastal Atlantic Forest, and demonstrated that it also bears signs of anthropogenic loss of diversity, at a rate that may be even higher than that of the inland fragments.

Any genetic study is sensitive to the geographic scale considered in the analysis, potential gaps in sampling, and numbers of makers and their information content (Radespiel & Bruford, 2014). Furthermore, ancient demographic process left genetic imprints in edge-populations (vs. core-populations) that are analogous to signals detected in shrinking populations subject to contemporaneous anthropic-driven drift (Slatkin & Excoffier, 2012), potentially hindering the disentanglement of the underlying process. For instance, jaguars have shown a marked population structure altogether but a weak signal of isolation by distance across Central America, which increased when Mexican (edge-) populations were included in the analysis (Wultsch et al., 2016a; Wultsch, Waits, & Kelly, 2016b). Similarly, comparing 11 microsatellite loci typed in jaguars from Brazil and Mexico, Roques et al. (2016) found a marked genetic structure, with samples from Brazil forming three genetic clusters, corresponding to the Amazon/Cerrado, the Pantanal and the Caatinga. Genetic differentiation was not only related to geographic distance, but also to the intensity of drift, as the isolated population from the Caatinga showed low allelic richness and reduced gene flow relative to the other areas within Brazil. This is a likely consequence of a recent (within the last 20 to 30 years) demographic reduction, which may reflect the Caatinga region contemporary habitat deterioration. Jaguars sampled in the Amazon rainforest showed high levels of genetic diversity and panmixia across considerable distances, while the genetic diversity was lower towards the limits of the species' range (Mexico, Caatinga and Pantanal). However, Roques et al. (2016) did not survey the Amazon as a whole, as their geographic

sampling of this vast region was restricted to a north-south transect covering only the central portion of the biome, leaving large sampling gaps in the eastern and western Amazon. In addition, that study did not include comparisons with Atlantic Forest populations, which have been found to be severely impacted by recent fragmentation (Haag et al., 2010; Srbek-Araujo et al., 2018).

In this context, the aim of this study was to survey the jaguar's genetic variability and population structure across the Amazon, and to perform comparative analyses of this data set jointly with those reported previously for Atlantic Forest (Haag et al., 2010; Srbek-Araujo et al., 2018) and southern Pantanal (Valdez et al., 2015) populations. In particular, we aimed to employ standardized molecular markers to assess the hypothesis that jaguars in the large, continuous Amazon rainforest show greater levels of genetic diversity and population size and connectivity than in the highly fragmented Atlantic Forest. We included the Pantanal biome as a control for high-quality habitat availability, as this later region currently harbours roughly the same extension as the sum of Atlantic Forest remnant fragments. This result would further corroborate our previous inference that the population structure observed in the Atlantic Forest is anthropogenic (Haag et al., 2010; Srbek-Araujo et al., 2018), and stress the importance of generating baseline data for jaguar genetics and ecology in a habitat that still retains large-scale continuity.

The specific aims of this study were as follows:

1. To contribute data on jaguar population structure and genetic diversity in the Amazon region, which currently represents its main stronghold for global conservation, but is still understudied due to its vastness and inaccessibility.

2. To compare these results with those previously published for two different biomes, the Pantanal and the Atlantic Forest, which are subjected to different intensities of anthropogenic disturbance.

3. To summarize the amounts of genetic diversity and population structure in these populations, characterizing their spatial distribution within and among biomes.

4. To provide baseline data for assessment of jaguar vulnerability to genetic erosion in its core

157 range, as well as in other areas, given current and projected scenarios of habitat degradation.

158

159 Methods

160 2.1 Sampling protocol

161 We obtained samples of biological material from 73 Amazonian jaguars, including blood
162 samples from animals captured for field ecology studies or kept in captivity, and pelt/hair samples from
163 material confiscated by local environmental authorities or from specimens kept in museum collections
164 (Supporting Information Appendix S1). Field-captured animals were covered by capture permit 11095-
165 8, issued by SISBIO/ICMBio, Brazil. The overall Amazonian sample encompassed three sub-regions:
166 upper Amazon (n=46), northeastern Amazon (n=18) and southeastern Amazon (n=9) (Figure 1). Blood
167 samples were preserved with EDTA, followed by mixing with an equal volume of the buffer TES (100
168 uM Tris, 100 uM EDTA, 2% SDS). Pelts, tissues and hairs were preserved in 96% ethanol. Faecal samples
169 were stored in sterile vials containing silica gel at a ratio of 4g silica: 1g stool (Wasser, Houston, Koehler,
170 Cadd, & Fain, 1997). All samples were stored at -20°C prior to DNA extraction.

171

172 2.2 Data collection and dataset construction

173 We performed DNA extractions from Amazonian samples using the commercial kits Puregene
174 DNA Purification Kit (GENTRA), ChargeSwitch Forensic DNA Purification Kit (Invitrogen), or QIAamp
175 DNA Stool Mini Kit (QIAGEN), following the manufacturers' instructions. We used the DNA extracts to
176 perform genetic analyses using 13 microsatellite loci, one with a dinucleotide repeat (FCA742), two
177 with trinucleotide repeats (F146 and F98), and ten with tetranucleotide repeats (FCA741, FCA740,
178 FCA723, FCA453, FCA441, FCA391, F124, F85, F53 and F42). We scored microsatellite alleles using a
179 MegaBACE 1000 automated sequencer and the ET-ROX 550 size standard, and then analysed them
180 with the accompanying Genetic Profiler software v.2.2, as described by Haag et al. (2010).

181 To allow comparisons on a broader scale, we jointly analysed these Amazonian data with
182 genotypic matrices generated by Haag et al. (2010) and Srbek-Araujo et al. (2018) for the Atlantic

Forest (n=59, and n=11, respectively), as well as Valdez et al. (2015) for the Pantanal (n=52). These studies used the same loci, and their data are available on the Dryad digital repository (https://doi.org/10.5061/dryad.1884/1; https://doi.org/10.5061/dryad.371c6). Genotyping for all these previous studies and for the present one was performed with the same protocols and equipment, including replicated control samples to allow identical binning of microsatellite alleles.

2.3 Genotyping Quality Control

All datasets were screened for genotyping errors and missing data using the 'strataG' package v.2.0.2 (Archer, Adams, & Schneiders, 2016). First, we identified samples with missing loci using a threshold equal to 0.69, i.e. only individuals genotyped for at least nine out of 13 (69%) loci were included in the analysis. We then assessed the percentage of missing samples per locus, using a cut-off value of 0.20. We removed loci below this threshold from the analysis. For the novel Amazon dataset, we also checked for duplicate genotypes, using an identity threshold of 1.0. We assessed departures from Hardy-Weinberg Equilibrium (HWE) proportions, using the exact test of Guo & Thompson (1992) for heterozygote deficit, as well as linkage disequilibrium between loci in 'genepop' v.1.0.5. For both tests, we estimated *P*-values by the Markov chain method with 10,000 dememorization steps, 200 batches and 5,000 iterations per batch. For some downstream analysis (i.e. effective population size estimation, see below), we previously tested for the presence of closely related individuals (parent-offspring, and full-siblings) using the software ML-Relate v.1 (Kalinowski, Wagner, & Taper, 2006).

2.4 Genetic diversity and Population structure

We calculated standard diversity and differentiation indices with 'adegenet' v.2.1.1 (Jombart, 2008) and 'diveRsity' v.1.9.90 (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013) packages in R, and GenAlex v.6.503 in Excel (Peakall & Smouse, 2012). We also calculated allelic richness by rarefaction using HP-Rare v.1 (Kalinowski, 2005). We assessed population structure with *F*-statistics computed in

GenAlex, including pairwise standardized measures (G_{st}), which are better suited for hypervariable markers, such as microsatellite loci, than F_{st} indices (Hedrick, 2005), using 1,000 permutations to estimate *P*-values.

In addition, we used Bayesian clustering in Structure v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000) parallelized with StrAuto v1.0 (Chhatre & Emerson, 2017) to reduce running time. The optimal value of *k* was defined using the Puechmaille method (Puechmaille, 2016) calculated on the Structure Selector web server (Li & Liu, 2018), based on 20 replicates per *k*, with 2^6 burn-in steps and 2^6 additional Markov Chain Monte Carlo (MCMC) sampled generations per run. Many studies have based the choice of the optimal *k* using the Evanno approach (Janes et al., 2016). However, it has been shown (Gilbert et al., 2012) that this works well only for datasets that harbour at least two genetic clusters; therefore, it does not perform well when the population shows no structure (i.e. *k*=1). Moreover, the Puechmaille method has shown a better performance than Evanno's technique in cases of uneven sampling (Puechmaille et al., 2016), as is the case in the present study. The genetic clusters for the best value of *k* were visualized in geographic space through the interpolation of the admixture coefficients onto a South America raster map, using the R script provided by Jay et al., (2012), as a companion to the spatially explicit Bayesian clustering approach Tess v.2.3 (Chen, Durand, Forbes, & François, 2007). For this, the 20 replicate runs of Structure generated with the optimal *k* value were merged with CLUMMP v.1.1.2 (Jakobsson & Rosenberg, 2007) using the greedy algorithm and 10,000 repeat configurations, in order to generate a single admixture matrix (*Q*-matrix) as an input for Tess. Finally, we also ran Structure with the LOCPRIOR option, using the putative population origin of each sample as a prior (Supporting Information Appendix S1). Isolation by distance (IBD) patterns were assessed within and among biomes using individual-based pairwise Mantel tests (Mantel, 1967), comparing genotypic (proportion of shared alleles) and geographic matrices with the distance-based module and the correlogram module in GenAlex.

2.5 Effective population size and contemporary bottlenecks

We estimated the contemporary effective population size (N_e) for each of the inferred populations using the programs Speed-Ne v.2.3 (Hamilton, Tartakovsky, & Battocletti, 2018), NeEstimator v2.1 (Do et al., 2014), and LDNE v.1.31 (Waples & Do, 2008), incorporating two values for the minor allele frequency (MAF, 0 and 0.01), and discarding seven closely related individuals detected by the relatedness analysis (Supporting Information Appendix S1). Finally, we searched for signals of drastic contemporary population reductions with the software Bottleneck v.1.2.02 (Piry, Luikart, & Cornuet, 1999).

Results

3.1. Dataset features

For the joint data set, using the 0.69 threshold of genotyped loci, we discarded three individuals that did not meet this criterion. After checking for exact duplicate genotypes, one additional individual was removed from the Amazon dataset (likely deriving from tube mislabelling during sample collection or processing), as well as another one showing an excess of homozygous genotypes. Two out the 13 loci showed more than 20% of missing genotypes: F124 (n=48; 24.7% missing) and FCA741 (n=41.5; 21.4% missing), and we removed them from further analyses, for a final dataset of 190 individuals reliably genotyped at 11 loci. Before estimating effective population size, we removed seven individuals from the Amazon dataset that potentially could downwardly bias the estimates, which were part of two parent-offspring pairs, three full-sibling pairs, and one full-sibling triplet (Supporting Information Appendix S1).

The Amazon population showed no significant deviations from HWE (P>0.05), except for the loci FCA740, FCA391 and F98, which presented a heterozygote deficit. For the Atlantic Forest dataset, two loci (FCA723, FCA441) showed signs of heterozygote deficit. The linkage disequilibrium test did not detected any significant non-random associations between pairwise locus comparisons. Since there was no consistent trend of the same loci showing departures from equilibrium, and to maximize information content, we kept the full dataset for all the analyses described below.

261 3.2 Genetic diversity

262 Overall, Amazon jaguars showed considerably high levels of genetic variability across most of
263 the loci (Table 1), with most of the estimates being higher than those of the Atlantic Forest and the
264 Pantanal (Table 2). Confidence intervals for the estimates of Allelic richness (*Ar*) per locus did not
265 overlap among the three biomes, indicating significantly higher diversity in the Amazon than in the
266 Atlantic Forest, which was significantly more diverse than the Pantanal. Expected heterozygosity
267 followed the same pattern, but observed heterozygosity showed the opposite trend, with lower values
268 in the Amazon (Table 2). Total and private alleles ranged from 10.2 and 2.7 for the Amazon, to 6.5 and
269 0.40 for the Pantanal (Table 2).

270

271 3.3 Population structure

272 *F*-statistics among major populations were quite low, with F_{st} values ranging from 0.037–
273 0.052), although their confidence intervals did not overlap zero, indicating modest but significant
274 differentiation among biomes (Table 3). G_{st} values were higher, and followed the same trend,
275 indicating that the highest levels of differentiation were observed between the Atlantic Forest and the
276 Pantanal, and the lowest ones between the Pantanal and the Amazon.

277 Changes in allelic frequencies identified four major clusters of population subdivision, one
278 corresponding to the Amazon, the second one representing the Pantanal, and the third and fourth
279 dividing the Atlantic Forest into two spatial domains (Figure 2). One of them grouped the Green
280 Corridor (the southern block of the Upper Parana Atlantic Forest [UPAF]) with the coastal Vale
281 population, on opposite sides of the surveyed region, while the other group assembled individuals
282 from a central area, comprising the small fragments of the northern block on the UPAF (Porto
283 Primavera, Ivinhema and Morro do Diabo).

284 Extensive admixture was observed among the three biomes, and the Amazon cluster included
285 a few individuals with a large proportion of ancestry coming from the Pantanal and the Atlantic Forest
286 (Figure 2a). In the next hierarchical level of structure, neither the Amazon nor the Pantanal showed

287 further subdivision (*k*=1 each), whereas the Atlantic forest showed a marked structure into five genetic
288 clusters (Supporting Information Appendix S1).
289
290 3.4 Isolation by distance
291 We did not find significant patterns of isolation by distance (IBD) within and among the biomes
292 (Figure 3). However, Mantel tests of the proportion of shared alleles vs. geographic distance indicated
293 a slight inverse relationship for all the biomes except the Amazon. This pattern was clearer for the
294 Atlantic Forest (Spearman *R*= -0.475; Figure 3c), followed by the three biomes assessed jointly (*R*= -
295 0.222; Figure 3d) and the Pantanal by itself (*R*= -0.178, Figure 3b). The *R*-value for the Amazon was
296 nearly null (*R*= 0.034, Figure 3a), although the spatial correlogram indicated that this small signal of
297 IBD derives from a negative correlation between genetic similarity and geographic distance observed
298 up to a distance of 400 km (Figure 4). Within this range, the negative correlation is significantly
299 different from the null expectation (of no correlation) up to a distance of 150 km between sampling
300 points.
301
302 3.5 Effective population size and bottlenecks
303 Estimates of contemporary effective population size based on linkage disequilibrium were
304 lowest for the Atlantic Forest and highest for the Amazon, ranging from 20 to 887 individuals,
305 respectively (Table 4). Using these figures and assuming that N_e represents on average one tenth of
306 the census size (N_c) for a given population (Frankham 1995), we estimate that N_c point estimates range
307 from 1,152 to 8,877 individuals in the Amazon; 499 to 812 in the southern Pantanal, and 169 to 262 in
308 the Atlantic Forest (Table 4). We did not detect signals of recent bottlenecks for the Amazon and
309 Pantanal populations. However, when we performed the analysis on the four separate clusters of the
310 Atlantic Forest, the Morro do Diabo and Ivinhema demes appeared bottlenecked.
311
312 4 Discussion

313 4.1. General patterns

314 Genetic diversity studies constitute a pillar in the field of conservation biology, although their
315 practical application has not been fully achieved in so far (Hoban et al., 2013; Rivers, Brummitt,
316 Lughadha, & Meagher, 2014; de la Torre, González-Maya, Zarza, Ceballos, & Medellín, 2018). As a
317 contribution to fill this gap, we analysed the most broadly distributed set of genetic samples for
318 Amazonian jaguars sampled to date, and directly compared it with two other biomes, serving as a
319 baseline for the assessment of jaguar population genetics across species' range. As a result, we
320 highlight the following features. The Amazonian jaguar population showed (1) moderate to high levels
321 of microsatellite diversity, for example as assessed by allelic richness; (2) large-scale connectivity with
322 signals of panmixia across thousands of kilometres, both south and north of the Amazon River; (3)
323 relatively large effective size population with no signals of recent bottlenecks. The Pantanal population
324 displayed (4) lower genetic diversity but a relatively large effective population size derived from just a
325 small surveyed portion of that area; while for the Atlantic Forest population we corroborated (5)
326 intermediate levels of diversity, with a marked structure due to strong signals of anthropic-driven drift,
327 and even recent bottlenecks in two of their demes.

328 The high diversity and long-distance connectivity observed in the Amazon highlight the
329 importance of this region as the most extensive stronghold for this species. It is noteworthy that
330 genetic variability comparisons with the Pantanal population are constrained by the relatively
331 restricted geographic area surveyed by Valdez et al. (2015), but in the case of the Atlantic Forest, our
332 comparison was useful to confirm the effects of habitat loss and fragmentation in that biome (Haag et
333 al., 2010; Srbek-Araujo et al., 2018). Currently, the lack of genetic studies on jaguars using historical
334 samples from museums and collections, such as those performed in other big cat species (e.g. Dures
335 et al., 2019), justifies our use of the Amazon population as a baseline, assuming that it retains most of
336 the variability that has been lost from some other areas due to large-scale habitat degradation.

337

338 Accordingly, the level of genetic differentiation among jaguar populations sampled in these

three biomes supports the view that this species has historically attained high levels of gene flow on a broad geographic scale (Eizirik et al. 2001). This pattern can be explained by the high dispersal potential of jaguars, favoured by high quality, continuous habitat, which in turn allowed gene flow across the Neotropics. These inferred high levels of connectivity contrast with observations of stronger population differentiation based on mtDNA markers (e.g. Eizirik et al., 2001), and support the view that this species exhibits a male-biased dispersal pattern, as has been described for other big cats (e.g. Smith, 1993; Gour et al, 2013; Fattebert, Balme, Dickerson, Slotow, & Hunter, 2015). Similar instances of higher variability and less structured populations towards the centre of the species range had been documented elsewhere for jaguars (Roques et al., 2016) and other large mammalian carnivores such as tigers in Nepal (Thapa et al., 2018), leopards in South Africa (McManus et al., 2015), black bears in Florida (Dixon et al., 2007), and wolverines in Montana (Cegelski, Waits, & Anderson, 2003). In all of these instances, habitat fragmentation was the underlying factor causing differentiation at peripheral populations.

4.2 High diversity in the Amazon and genetic drift in the Atlantic Forest

With the sole exceptions of observed heterozygosity (H_o) and inbreeding coefficient (F_{is}), summary statistics indicated that the Amazon rainforest sustains one of the most diverse jaguar populations in South America, as inferred from its comparison to the Atlantic Forest and the Pantanal populations (Table 2). It is expected that this patterns holds range-wide, since previous studies have shown lower variability levels in other peripheral biomes not assessed in this study, such as the Caatinga in Brazil, Mesoamerican forests, and subtropical Mexico (Roques et al., 2016; Wultsch et al. 2016a, b). This assertion is supported by the levels of diversity reported by Roques et al. (2016) for the Amazon [H_e (0.805) and H_o (0.848)], which were higher than those of other populations, except for the central-range Cerrado biome. Gene diversity level was similar to the value documented in this study [H_e(0.76)], although direct comparisons are hampered by the fact that different loci were employed in each assessment. Likewise, our diversity estimates are higher than those reported for the tropical

365 rainforest in Belize (H_e=0.57; H_o=0.57) by Wultsch et al. (2016b), but again the set of loci is different,
366 precluding a more direct comparison. The higher F_{is} and lower H_o values are the result of several closely
367 related individuals, consistent in two parent-offspring dyads, three full-sibling dyads, and one full-
368 sibling triad) detected in the Amazon population (Supporting Information Appendix S1).

369 An interesting observation was that the Atlantic forest as a whole still retains genetic diversity
370 levels similar to those in the Pantanal, but the most isolated subpopulation (Morro do Diabo) showed
371 even lower values (H_o=0.55; H_e=0.50; Haag et al., 2010) than those documented for Belize. It is
372 remarkable that the heavily fragmented Atlantic Forest demes retain, altogether, rather high levels of
373 diversity, likely representing a large portion of their historic variability share. However, jaguars in this
374 highly fragmented region are under a metapopulational dynamic, where each remaining population is
375 subject to genetic stochastic effects (Dixon et al., 2007), losing its variability by drift and even being at
376 risk of local extirpation (Jędrzejewski et al., 2017; Thatte, Joshi, Vaidyanathan, Landguth, &
377 Ramakrishnan, 2018).

378

379 4.3 High connectivity in the Amazon

380 All the metrics were consistent in showing large-scale demographic connectivity encompassing
381 thousands of kilometres across the Amazon basin. As a result, we infer that the lack of population
382 subdivision in this vast region implies far-reaching amounts of gene flow throughout the landscape. A
383 significant signal of IBD was detected from 0 to 150 km, and this pattern is expected as the individuals
384 are more closely related in shorter distances, with genetic relatedness gradually fading away (Zanin et
385 al., 2016). The extent of the genetic neighbourhood, where genetic correlation is negatively associated
386 with distance, was estimated to lie between 300-400 km (Figure 4). This seems biologically reasonable
387 in terms of the high vagility and social organization of jaguars (i.e. one male overlapping the home
388 range of three or more females), especially in a continuous, productive habitat such as the Amazon.
389 Similar results were reported for tigers in the Sundarbans (Aziz et al., 2018).

390 Our results could represent one of the few possible snapshots of large-scale jaguar population

connectivity before severe human intervention, illustrating the occurrence of historical panmixia throughout the tropical forested biomes across the species' range, from the Atlantic Forest in southeastern South America to the Mayan forest in Mesoamerica. Local discontinuities may occur in areas such as the Pantanal, perhaps driven by adaptive differentiation in ecological and/or behavioural traits (Figueiró et al., unpublished), but much of the interruption of long-range gene flow observed in recent studies is likely to have been exacerbated by human-driven drift. In this sense, Wultsch et al. (2016a) found signals of interruption of panmixia in northern Central America, between the Mayan forest, which is the largest tract of Neotropical rainforest outside of the Amazon, and the Honduran population, probably due to a drastic habitat loss between these two regions.

In this context, it is extremely important to maintain the connectivity in the Amazon, as large-scale deforestation" is advancing in the southern limits of the biome. Projections indicate that by 2050, the Amazon will lose 40% of its area, and the protected areas network will not be sufficient to fully protect its biodiversity (Soares-Filho et al., 2006), as deforestation, poaching and illegal fishing and mining continue as the main threats (Kauano, Silva, & Michalski, 2017). Reversing this trend and maintaining large-scale connectivity across this biome will be critical not only for jaguars (Silveira, Sollmann, Jácomo, Diniz Filho, & Tôrres, 2014), but also for many other components of Amazonian biodiversity (Lees & Peres, 2008).

In spite of their high vagility, jaguars may be more vulnerable than other species to human-induced fragmentation. For example, Figueiredo et al. (2015) identified no genetic structure between ocelots (*Leopardus pardalis*) sampled at Morro do Diabo and the Green Corridor, contrasting with pattern observed in jaguars, suggesting that the latter are more sensitive to genetic erosion driven by anthropic disturbance. A likely explanation is that ocelots possess larger effective population sizes (due to smaller body size and higher density) in the same area, thus taking longer to show the effects of genetic drift. An additional possibility is that ocelots are more capable of navigating through the human-dominated matrix (Zimbres, Peres, Penido, & Machado, 2018), maintaining gene flow across fragments in a way that jaguars no longer can. A similar trend was reported for ocelots, pumas and

417 jaguars in Belize (Wultsch et al., 2016b), as ecological and behavioural differences among these species
418 could determine the potential and effective amounts of gene flow among populations. However, this
419 pattern also seems to be dependent on the time elapsed since habitat perturbation and its intensity,
420 as well as habitat productivity. Ocelots occurring in southern Texas, on the northern limit of their
421 range, where two subpopulations occur in small blocks of semiarid habitat isolated from each other by
422 approximately 30 km of cropland matrix, show small N_e (<14) and high differentiation (F_{st}=0.163;
423 Janečka et al., 2011). In general, top predators are very sensitive to habitat perturbation (Dutta et al.,
424 2012), but this sensitivity can be attenuated by differences in ecosystem productivity (Jędrzejewski et
425 al. 2017) and their natural recolonization capability (Malaney, Lackey, Beckmann, & Matocq, 2018).
426 This situation also raises the question about restoring connectivity of landscapes subjected to
427 heavy anthropic perturbation, such as the Atlantic Forest (Ribeiro et al., 2009; Silveira et al. 2014), and
428 constitutes a warning about the negative effects of fragmentation that could occur in less disturbed
429 regions such as the Pantanal, whose extent is much smaller than that of the Amazon. For example, the
430 coastal Vale population still retains some of the shared diversity present in the interior of the Atlantic
431 Forest (i.e. Green Corridor), likely represented by ancestral alleles. However, its size and degree of
432 isolation make it very difficult to maintain gene flow with other coastal subpopulations persisting in
433 that biome, such as those described by Souza et al. (2017), or even in nearby biomes such as the
434 Cerrado and Caatinga. This strengthens the previous evidence that the marked change in allele
435 frequencies in the central populations of the biome has caused a genetic differentiation that reflects
436 its contemporary, fast degradation (Haag et al., 2010; Valdez et al., 2015). The magnitude of the
437 deforestation rates in the Atlantic Forest has already propitiated defaunation and cascade effects
438 across this biome (Jorge, Galetti, Ribeiro, & Ferraz, 2013), and management actions such as restoring
439 connectivity through riparian and mountainous corridors are urgently needed (Castilho, Hackbart,
440 Pivello, & dos Santos, 2015), using spatially-explicit approaches on gene flow (Reddy, Cushman,
441 Srivastava, Sarkar, & Shivaji, 2017).
442

443 4.4 Effective population size

444 Given the vast extension of the Amazon rainforest, it is probable that we obtained an
445 underestimate of effective population size in this study. This issue has been identified when calculating
446 N_e using linkage disequilibrium estimators (Wang, 2005, Waples & Do, 2010), as this method has a
447 better performance when population size is small, as is the case for the Atlantic Forest in our
448 assessment. On the other hand, our inference that the Amazon population approaches panmixia
449 indicates that our sample may be sufficiently representative of the biome as a whole to allow an
450 inference of its overall effective size. In any event, our estimate was sufficient to demonstrate that the
451 Amazon appears to sustain a much larger breeding population than the other two assessed biomes, at
452 least twice the size of the southern Pantanal's and almost eight times larger than the Atlantic Forest's.

453 Despite the massive expanse of Amazon basin, it is currently losing primary cover, which can
454 lead to jaguar demographic reductions and local extirpations as the agrarian frontier continues
455 encroaching on the rainforest. As for the Pantanal and the Atlantic Forest biomes, primary habitat
456 currently extends over roughly equivalent areas (~100,000 km^2), although in the former case it forms
457 a single, continuous block, while in the latter the remaining area is fragmented into thousands of small
458 patches. Up to 80% of those patches are smaller than half a square kilometre (Ribeiro et al., 2009),
459 which is too small to sustain even a single jaguar individual, partially explaining the very low N_e (17-26)
460 estimates for that biome. Indeed, those figures put Atlantic Forest jaguars below the N_e=50 threshold
461 proposed by Franklin & Frankham (1998) to avoid short-term risks due to inbreeding (Rutledge et al.,
462 2017).

463

464 4.5 Concluding remarks

465 On the basis of the results presented here, and the need to further refine these inferences, we
466 recommend that continuous molecular surveys (including genome-wide approaches) be performed
467 throughout the jaguar range, addressing demographic as well as adaptive questions, and providing
468 updated information on the genetic health of natural populations. This would help to rapidly detect

changes that can further compromise the persistence of jaguars throughout their distribution, enabling improved management actions in the context of long-term conservation strategies that integrate multiple spatial scales.

We conclude by stressing the importance of maintaining connectivity regionally and across the species' range to ensure that gene flow persists within and across biomes, including those that still represent large strongholds for the species, such as the Amazon rainforest. To achieve this goal, it is critical to monitor the loss of genetic diversity driven by human-induced fragmentation and population isolation, and to actively restore gene flow it in some cases. In this context, it is noteworthy that rampant habitat loss is currently taking place in the southeastern Amazon across the "arc of deforestation", which represents an imminent threat (or perhaps already a reality) of gene flow interruption with adjacent biomes such as the Cerrado, Caatinga and Pantanal. Recent trends of increased deforestation and weakened enforcement of environmental protection in this region are alarming, and have been the focus of extensive concern by the scientific community (e.g. Abessa, Famá, & Buruaem, 2019; Kehoe et al., 2019). The situation in the Atlantic Forest is even more worrisome, as evidence has accumulated demonstrating that drastic habitat fragmentation takes only a few decades to induce severely negative effects (both genetic and demographic) on wildlife species such as jaguars. Urgent action is needed to avoid that the Amazon rainforest follows a similar trajectory in the next few decades, which could lead to disastrous effects on a global scale.

References

Abessa, D., Famá, A., & Buruaem, L. (2019). The systematic dismantling of Brazilian environmental laws risks losses on all fronts. *Nature Ecology & Evolution*, 3, 510-511.

Aprile, G., Cuyckens, E., De Angelo, C., Di Bitetti, M., Lucherini, M., Muzzachiodi, N., ... Soler, L. (2012). Familia Felidae. In: R. A. Ojeda, V. Chillo & G. B. Diaz Isenrath (Eds.), Libro rojo de mamíferos amenazados de la Argentina (pp 92-101). Mendoza; Sociedad Argentina para el Estudio de los Mamíferos (SAREM).

Archer, F.I., Adams, P.E., & Schneiders, B.B. (2016) strataG: an R package for manipulating, summarizing and analyzing population genetic data. *Molecular Ecology Resources*, 17, 5-11.

Aziz, M. A., Smith, O., Barlow, A., Tollington, S., Islam, M. A., & Groombridge, J. J. (2018). Do rivers influence fine-scale population genetic structure of tigers in the Sundarbans? *Conservation Genetics*, 19, 1137-1151.

Beisiegel, B., Sana, D., & Moraes, E. (2012). The jaguar in the Atlantic Forest. *Cat News Spec*, 7, 14-18.

Caso, A., Lopez-Gonzalez, C, Payan, E., Eizirik, E., de Oliveira, T., Leite-Pitman, R., ... Valderrama, C. (2008). *Panthera onca*. The IUCN Red List of threatened species 2008: e.T15953A5327466. Downloaded on 15 October 2017.

Castilho, C. S., Hackbart, V. C., Pivello, V. R., & dos Santos, R. F. (2015). Evaluating landscape connectivity for *Puma concolor* and *Panthera onca* among Atlantic forest protected areas. *Environmental Management*, 55, 1377-1389.

Cavalcanti, S. M. C., Azevedo, F. C. C., Tomás, W. M., Boulhosa, R. L. P., Crawshaw, Jr. P. G. (2012). The status of the jaguar in the Pantanal. *Cat News Spec* 7, 29-34.

Cegelski, C. C., Waits, L. P., & Anderson, N. J. (2003). Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Molecular Ecology*, 12, 2907-2918.

Chen, C., Durand, E., Forbes, F., & François, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, 7, 747- 756.

Chhatre, V. E., & Emerson K. J. (2017). StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, 18, 192.

Costa, L. P., Leite, Y. L. R., Mendes, S. L., & Ditchfield, A. D. (2005). Mammal conservation in Brazil. *Conservation Biology*, 19, 672-679.

Crooks, K. R. (2002). Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology* 16, 488-502.

de la Torre, J. A., González-Maya, J. F., Zarza, H., Ceballos, G., & Medellín, R. A. (2018). The jaguar's spots are darker than they appear: assessing the global conservation status of the jaguar *Panthera onca*. *Oryx*, 52, 300-315.

de Oliveira, T.G., Ramalho, E.E., de Paula, R.C. (2012) Red List assessment of the jaguar in Brazilian Amazonia. *Cat News Spec*, 7, 8-13.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. *Science* 345, 401-406.

Dixon, J. D., Oli, M. K., Wooten, M. C., Eason, T. H., McCown, J. W., & Cunningham, M. W. (2007). Genetic consequences of habitat fragmentation and loss: the case of the Florida black bear (*Ursus americanus floridanus*). *Conservation Genetics*, 8, 455-464.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14, 209-214.

Dures, S. G., Carbone, C., Loveridge, A. J., Maude, G., Midlane, N., Aschenborn, O., & Gottelli, D. (2019). A century of decline: loss of genetic diversity in a southern African lion-conservation stronghold. *Diversity and Distributions*. First published online: 11 March 2019 https://doi.org/10.1111/ddi.12905.

Dutta, T., Sharma, S., Maldonado, J. E., Wood, T.C., Panwar, H.S., & Seidensticker, J. (2012). Fine-scale population genetic structure in a wide-ranging carnivore, the leopard (*Panthera pardus fusca*) in central India. *Diversity and Distributions*, 19, 760–771.

Endangered Species Act (ESA). (1973) 16 U.S.C. Sections 1531-1544. December 28, 1973.

Eizirik, E., Kim, J. H., Menotti-Raymond, M., Crawshaw Jr, P. G., O'Brien, S. J., & Johnson, W. E. (2001). Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, Mammalia, Felidae). *Molecular Ecology*, 10, 65-79.

Fattebert, J., Balme, G., Dickerson, T., Slotow, R., & Hunter, L. (2015). Density-dependent natal dispersal patterns in a leopard population recovering from over-harvest. *PloS One*, 10,

547 e0122355.
548 Figueiredo, M., Cervini, M., Rodrigues, F., Eizirik, E., Azevedo, F., Cullen, L., ... Galetti, P. (2015). Lack of
549 population genetic structuring in ocelots (*Leopardus pardalis*) in a fragmented landscape.
550 *Diversity*, 7, 295-306.
551 Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: A review.
552 *Genetical Research*, 66, 95-107.
553 Franklin, J.R. & Frankham, R. (1998) How large must populations be to retain evolutionary potential?
554 *Animal Conservation*, 1, 69-73.
555 Gardner, T.A., Barlow, J., Sodhi, N.S. & Peres, C.A. (2010). A multi-region assessment of tropical forest
556 biodiversity in a human-modified world. *Biological Conservation*, 143, 2293-2300.
557 Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J. S., ... Vines, T. H. (2012).
558 Recommendations for utilizing and reporting population genetic analyses: the reproducibility
559 of genetic clustering using the program STRUCTURE. *Molecular Ecology*, 21, 4925-4930.
560 Gour, D. S., Bhagavatula, J., Bhavanishankar, M., Reddy, P. A., Gupta, J. A., Sarkar, M. S., ... Shivaji, S.
561 (2013). Philopatry and dispersal patterns in tiger (*Panthera tigris*). *PloS One*, 8, e66956.
562 Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for
563 multiple alleles. *Biometrics*, 48, 361-372.
564 Haag, T., Santos, A. S., Sana, D. A., Morato, R. G., Cullen Jr, L., Crawshaw Jr, P. G., ... Eizirik, E. (2010).
565 The effect of habitat fragmentation on the genetic structure of a top predator: loss of diversity
566 and high differentiation among remnant populations of Atlantic Forest jaguars (*Panthera*
567 *onca*). *Molecular Ecology*, 19, 4906-4921.
568 Hamilton, M. B., Tartakovsky, M., & Battocletti, A. (2018). SPEED-NE: software to simulate and estimate
569 genetic effective population size (N_e) from linkage disequilibrium observed in single samples.
570 *Molecular Ecology Resources*, 18, 714-728.
571 Hedrick, P. (2005). A standardized genetic differentiation measure. *Evolution*, 59, 1633-1638.
572 Hoban, S. H., Hauffe, C., Perez-Espona, S., Arntzen, J. W., Bertorelle, G., Bryja, J., … Bruford, M. W.

(2013). Bringing genetic diversity to the forefront of conservation policy and management. *Conservation Genetics Resources*, 5, 593-598.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801-1806.

Janečka, J. E., Tewes, M. E., Laack, L. L., Caso, A., Grassman Jr, L. I., Haines, A. M., ... Honeycutt, R. L. (2011). Reduced genetic diversity and isolation of remnant ocelot populations occupying a severely fragmented landscape in southern Texas. *Animal Conservation*, 14, 608-619.

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K= 2 conundrum. *Molecular Ecology*, 26, 3594-3602.

Jay, F., Manel, S., Alvarez, N., Durand, E. Y., Thuiller, W., Holderegger, R., ... François, O. (2012). Forecasting changes in population genetic structure of alpine plants in response to global warming. *Molecular Ecology*, 21, 2354-2368.

Jędrzejewski, W., Boede, E. O., Abarca, M., Sánchez-Mercado, A., Ferrer-Paris, J. R., Lampo, M., ... Robinson, H. S. (2017). Predicting carnivore distribution and extirpation rate based on human impacts and productivity factors; assessment of the state of jaguar (*Panthera onca*) in Venezuela. *Biological Conservation*, 206, 132-142.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405.

Jorge, M. L. S., Galetti, M., Ribeiro, M. C., & Ferraz, K. M. P. (2013). Mammal defaunation as surrogate of trophic cascades in a biodiversity hotspot. *Biological Conservation*, 163, 49-57.

Kalinowski, S. T. (2005). HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology Notes*, 5, 187-189.

Kalinowski, S. T., Wagner, A. P., & Taper, M. L. (2006). ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, 6, 576-579.

Kauano, É. E., Silva, J. M., & Michalski, F. (2017). Illegal use of natural resources in federal protected

areas of the Brazilian Amazon. *PeerJ*, 5, e3902.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4, 782-788.

Kehoe, L., Reis, T., Virah-Sawmy, M., Balmford, A., Kuemmerle, T., & 604 signatories. Make EU trade with Brazil sustainable. *Science*, 364, 341.

Lees, A. C., & Peres, C. A. (2008). Conservation value of remnant riparian forest corridors of varying quality for Amazonian birds and mammals. *Conservation Biology*, 22, 439-449.

Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, 18, 176-177.

Malaney, J. L., Lackey, C. W., Beckmann, J. P., & Matocq, M. D. (2018). Natural rewilding of the Great Basin: Genetic consequences of recolonization by black bears (*Ursus americanus*). *Diversity and Distributions*, 24, 168-178.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209-220.

Myers, N., Mittermeier, R. A, Mittermeier, C. G., de Fonseca, G. AB. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853-858.

McManus, J. S., Dalton, D. L., Kotze, A., Smuts, B., Dickman, A., Marshall, J. P. & Keith, M. (2015). Gene flow and population structure of a solitary top carnivore in a human-dominated landscape. *Ecology and Evolution*, 5, 335-344.

Nijhawan, S. (2012) Conservation units, priority areas and dispersal corridors for jaguars in Brazil. *Cat News Spec*, 7, 43-47.

Paviolo, A., De Angelo, C., Ferraz, K. M., Morato, R. G., Pardo, J. M., Srbek-Araujo, A. C., ... Azevedo, F. (2016). A biodiversity hotspot losing its top predator: The challenge of jaguar conservation in the Atlantic Forest of South America. *Scientific Reports*, 6, 37147.

Peakall, P. E., & Smouse, R. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software

for teaching and research—an update. *Bioinformatics*, 28, 2537.

Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, 90, 502-503.

Pritchard, J. K., Stephens, M., Donnelly, P. J. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16, 608-627.

Radespiel, U. & Bruford M. W. (2014). Fragmentation genetics of rainforest animals: insights from recent studies. *Conservation Genetics*, 15, 245-260.

Reddy, P. A., Cushman, S. A., Srivastava, A., Sarkar, M. S., & Shivaji, S. (2017). Tiger abundance and gene flow in Central India are driven by disparate combinations of topography and land cover. *Diversity and Distributions*, 23, 863-874.

Ribeiro, M. C., Metzger, J. P., Martensen, A. C., Ponzoni, F. J., & Hirota, M. M. (2009). The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, 142, 1141-1153.

Rivers, M. C., Brummitt, N. A., Lughadha, E. N., & Meagher, T. R. (2014). Do species conservation assessments capture genetic diversity?. *Global Ecology and Conservation*, 2, 81-87.

Rodríguez-Mahecha, J. V., Alberico, M., Trujillo, F., Jorgenson, J. (2006) Libro rojo de los mamíferos de Colombia. Bogotá: Conservación Internacional Colombia.

Roques, S., Sollman, R., Jácomo, A., Tôrres, N., Silveira, L., Chávez, C., ... Palomares, F. (2016). Effects of habitat deterioration on the population genetics and conservation of the jaguar. *Conservation Genetics*, 17, 125-139.

Row, J. R., Gomez, C., Koen, E. L., Bowman, J., Murray, D. L., & Wilson, P. J. (2012). Dispersal promotes high gene flow among Canada lynx populations across mainland North America. *Conservation*

651 *Genetics*, 13, 1259-1268.
652 Ruiz-Garcia, M., Payán, E., Murillo, A., & Alvarez, D. (2006). DNA microsatellite characterization of the
653 jaguar (*Panthera onca*) in Colombia. *Genes & Genetic Systems*, 81, 115-127.
654 Rutledge, L. Y., Desy, G., Fryxell, J. M., Middel, K., White, B. N., & Patterson, B. R. (2017). Patchy
655 distribution and low effective population size raise concern for an at-risk top predator.
656 *Diversity and Distributions,* 23, 79-89.
657 SEMARNAT (2010) Norma Oficial Mexicana NOM-059-SEMARNAT-2010. Diario Oficial de la Federación
658 (DOF). México.
659 Silveira, L., Sollmann, R., Jácomo, A. T., Diniz Filho, J. A., & Tôrres, N. M. (2014). The potential for large-
660 scale wildlife corridors between protected areas in Brazil using the jaguar as a model species.
661 *Landscape Ecology*, 29, 1213-1223.
662 Slatkin, M., & Excoffier, L. (2012). Serial founder effects during range expansion: a spatial analog of
663 genetic drift. *Genetics*, 191, 171-181.
664 Smith, J. L. D. (1993). The role of dispersal in structuring the Chitwan tiger population. *Behaviour* 124,
665 165-195
666 Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., ... &
667 Schlesinger, P. (2006). Modelling conservation in the Amazon basin. *Nature*, 440, 520.
668 Sollmann R., Tôrres, N. M., & Silveira, L. (2008). Jaguar conservation in Brazil: the role of protected
669 areas. *Cat News Spec* 4, 15-20.
670 Souza, A. S. M. D. C., Saranholi, B. H., Crawshaw Jr, P. G., Paviolo, A. J., Rampim, L. E., Sartorello, L., &
671 Galetti Jr, P. M. (2017). Re-discovering jaguar in remaining coastal Atlantic Forest in
672 southeastern Brazil by non-invasive DNA analysis. *Biota Neotropica*, 17, e20170358.
673 Srbek-Araujo, A. C., Haag, T., Chiarello, A. G., Salzano, F. M., & Eizirik, E. (2018). Worrisome isolation:
674 noninvasive genetic analyses shed light on the critical status of a remnant jaguar population.
675 *Journal of Mammalogy*, 99, 397-407.
676 Tabarelli, M., Pinto, L. P., Silva, J. M., Hirota, M., & Bedê, L. (2005). Challenges and opportunities for

biodiversity conservation in the Brazilian Atlantic Forest. *Conservation Biology*, 19, 695-700.

Tammeleht, E., Remm, J., Korsten, M., Davison, J., Tumanov, I., Saveljev, A., ... & Saarma, U. (2010). Genetic structure in large, continuous mammal populations: the example of brown bears in northwestern Eurasia. *Molecular Ecology*, 19, 5359-5370.

Thapa, K., Manandhar, S., Bista, M., Shakya, J., Sah, G., Dhakal, M., ... Karmacharya, D. (2018). Assessment of genetic diversity, population structure, and gene flow of tigers (*Panthera tigris tigris*) across Nepal's Terai Arc Landscape. *PloS One*, 13, e0193495.

Thatte, P., Joshi, A., Vaidyanathan, S., Landguth, E., & Ramakrishnan, U. (2018). Maintaining tiger connectivity and minimizing extinction into the next century: Insights from landscape genetics and spatially-explicit simulations. *Biological Conservation*, 218, 181-191.

Thornton, D., Zeller, K., Rondinini, C., Boitani, L., Crooks, K., Burdett, C., ... Quigley, H. (2016). Assessing the umbrella value of a range-wide conservation network for jaguars (*Panthera onca*). *Ecological Applications*, 26, 1112-1124.

Tortato, F. R., Izzo, T. J., Hoogesteijn, R., & Peres, C. A. (2017). The numbers of the beast: valuation of jaguar (*Panthera onca*) tourism and cattle depredation in the Brazilian Pantanal. *Global Ecology and Conservation*, 11, 106-114.

Valdez, F. P., Haag, T., Azevedo, F. C., Silveira, L., Cavalcanti, S. M., Salzano, F. M., & Eizirik, E. (2015). Population genetics of jaguars (*Panthera onca*) in the Brazilian Pantanal: molecular evidence for demographic connectivity on a regional scale. *Journal of Heredity*, 106, 503-511.

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1395-1409.

Waples, R. S., & Do, C. (2008). LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, 8, 753-756.

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, 3, 244-262.

Wasser, S. K., Houston, C. S., Koehler, G. M., Cadd, G. G., & Fain, S. R. (1997). Techniques for application of faecal DNA methods to field studies of Ursids. *Molecular Ecology*, 6, 1091-1097.

Wultsch, C., Caragiulo, A., Dias-Freedman, I., Quigley, H., Rabinowitz, S., & Amato, G. (2016a). Genetic diversity and population structure of Mesoamerican jaguars (*Panthera onca*): implications for conservation and management. *PloS One*, 11, e0162377.

Wultsch, C., Waits, L. P., & Kelly, M. J. (2016b). A comparative analysis of genetic diversity and structure in jaguars (*Panthera onca*), pumas (*Puma concolor*), and ocelots (*Leopardus pardalis*) in fragmented landscapes of a critical Mesoamerican linkage zone. *PloS One*, 11, e0151043.

Zanin, M., Adrados, B., González, N., Roques, S., Brito, D., Chávez, C., ... Palomares, F. (2016). Gene flow and genetic structure of the puma and jaguar in Mexico. *European Journal of Wildlife Research*, 62, 461-469.

Zimbres, B., Peres, C. A., Penido, G., & Machado, R. B. (2018). Thresholds of riparian forest use by terrestrial mammals in a fragmented Amazonian deforestation frontier. *Biodiversity and Conservation*, 27, 2815-2836.

Data Accessibility/Availability Statement:

Part of the data used in this paper have been published previously (Haag et al., 2010; Valdez et al., 2015, Srbek-Araujo et al., 2018). Two of these previous data sets have been deposited in the Dryad digital repository (https://doi.org/10.5061/dryad.1884/1; https://doi.org/10.5061/dryad.371c6). The third previous data set (Srbek-Araujo et al., 2018) and the novel data reported here have also been deposited in Dryad (accession XXX).

729 Tables

730

Table 1. Genetic diversity at 11 loci microsatellite in three populations of jaguars in South America. *N* number of individuals genotyped, *A* number of alleles, *AR* allelic richness, *PA* private alleles, H_e expected heterozygosity, H_o observed heterozygosity

	Amazon (n=71) (this study)						Atlantic Forest (n=68) (Haag et al. 2010; Srbek-Araujo et al. 2018)						Pantanal (n=51) (Valdez et al. 2015)					
Locus	*N*	*A*	*AR*	*PA*	H_e	H_o	*N*	*A*	*AR*	*PA*	H_e	H_o	*N*	*A*	*AR*	*PA*	H_e	H_o
FCA742	67	26	23.5	8.6	0.93	0.78	65	14	14.3	0.5	0.88	0.83	50	13	12.8	0.1	0.86	0.90
FCA723	66	9	7.8	2.1	0.63	0.52	66	7	6.7	1.0	0.67	0.48	48	7	6.9	1.9	0.63	0.67
FCA740	68	6	5.7	0.7	0.77	0.72	64	5	4.7	0	0.71	0.72	50	5	5.0	0	0.69	0.58
FCA441	69	8	7.9	1.9	0.80	0.75	68	6	6.0	0	0.74	0.59	48	7	6.9	1.0	0.62	0.52
FCA391	66	8	7.9	0	0.85	0.79	66	8	7.9	0.1	0.76	0.80	51	6	6.0	0	0.76	0.86
F98	71	5	4.6	1.6	0.80	0.69	67	4	4.0	1.0	0.55	0.60	50	3	3.0	0	0.66	0.66
F53	64	16	14.6	4.4	0.86	0.70	66	12	11.4	1.1	0.85	0.86	48	6	6.0	0	0.77	0.79
F146	62	8	7.2	2.7	0.42	0.35	68	5	4.7	0	0.59	0.54	48	3	3.0	0	0.30	0.23
F85	65	13	12.3	2.8	0.78	0.63	62	12	11.8	2.4	0.80	0.77	48	9	8.9	1.0	0.78	0.82
F42	65	13	12.4	2.5	0.88	0.72	58	9	8.9	0	0.77	0.69	48	7	7.9	0	0.84	0.85
FCA453	62	9	8.5	2.7	0.73	0.65	62	6	5.7	0.7	0.71	0.63	46	6	6.0	0	0.76	0.80

731

732

733

Table 2. Summary of population genetic parameters for Amazon, Atlantic Forest and Pantanal jaguar populations, based on 11 autosomal microsatellite loci. Number of genotyped individuals (*N*), mean number of observed alleles per loci (*Na*), mean number of effective alleles per locus (*Nf*), mean number of private alleles per loci (Np), allelic richness (*Ar*), rarefied allelic richness (*Af*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), inbreeding coefficient (*Fis*), confidence interval 95% (*CI*), standard error (*SE*)

Population	*N*	*Na*	*Nf*	*Np*	*Ar*	*(CI)*	*Af **	*Ho (SE)*	*He (SE)*	*Fis*	*(CI)*
Amazon	71	11	5.64	3.27	9.38	(8.55–10.18)	10.22	0.674 (0.041)	0.759 (0.043)	0.11	(0.062–0.142)
Atlantic Forest	68	8	4.26	0.55	7.75	(7.27–8.18)	7.82	0.684 (0.038)	0.730 (0.030)	0.06	(0.002–0.103)
Pantanal	51	6.5	3.98	0.36	6.36	(5.91–6.73)	6.58	0.698 (0.060)	0.698 (0.047)	0.02	(-0.044–0.054)

* Allelic richness rarefied to 92 gene copies (*N*=46)

Table 3. Fixation indices reflecting jaguar population differentiation in three South American biomes. Values above the diagonal are F_{st} *Nei*, with G_{st} *Hed* in parentheses; values below the diagonal are 95% confidence intervals for F_{st}

Population	Amazon	Pantanal	Atlantic Forest
Amazon	-	0.037 (0.124)	0.041 (0.149)
Pantanal	0.025-0.055	-	0.052 (0.164)
Atlantic Forest	0.032-0.064	0.045-0.097	-

745

746

747

748

749

750

751

752

753

754

755

756

757

758

Table 4. Contemporary effective population size (N_e) of jaguars estimated for three South American biomes (AM: Amazon; PA: Pantanal; AF: Atlantic Forest) based on linkage disequilibrium at 11 microsatellite loci, and extrapolated census size (N_c) based on the N_e point estimates

Population (sample size)	Estimates		NeEstimator[a]		LDNe[b]			spEED-Ne[c]		N_c [d] (range)	
			MAF 0+	MAF 0.01	MAF 0+	MAF 0.01		MAF 0+	MAF 0.01		
AM (n=64)	Point estimate	N_e	**115.2**	**724.4**	**119.2**	**887.7**		**278.9**	**278**	**1,152**	**8,877**
	95% CIs	(1)	87.3-164.1	241.4-inf	89.6-172.2	257.4-inf	(3)	116.4-inf	116.3-inf		
		(2)	43.2-inf	133.9-inf	62.2-499.3	183.6-inf	(2)	192.9-503.8	192.4-500.8		
PA (n=51)	Point estimate	N_e	**79.8**	**79.8**	**81.2**	**81.2**		**49.9**	**50.5**	**499**	**812**
	95% CIs	(1)	53.3-141.4	53.3-141.4	54.0-145.6	54.0-145.6	(3)	43.3-58.9	43.8-59.6		
		(2)	41.8-291.4	41.8-291.4	48.4-189.0	48.4-189.0	(2)	44.1-57.5	44.5-58.3		
AF (n=68)	Point estimate	N_e	**26.1**	**20.4**	**26.2**	**20.5**		**16.9**	**16.9**	**169**	**262**
	95% CIs	(1)	22.6-30.3	17.8-23.5	22.7-30.4	17.8-23.6	(3)	14.3-20.73	14.3-20.73		
		(2)	17.2-41.6	13.8-30.9	22.7-30.4	17.9-23.5	(2)	15.8-18.2	15.8-18.2		

(1) Parametric

(2) Jackknife on samples (individuals)

(3) Jackknife on loci

a NeEstimator v2.1 (Do et al., 2014)

b LDNe v1.31 (Waples & Do, 2008)

c spEED-Ne v.2.3 (Hamilton et al., 2018)

d Extrapolated census size N_c, where N_e represents one tenth of N_c

MAF Minor allele frequency

Inf Infinity

765

766

Figures

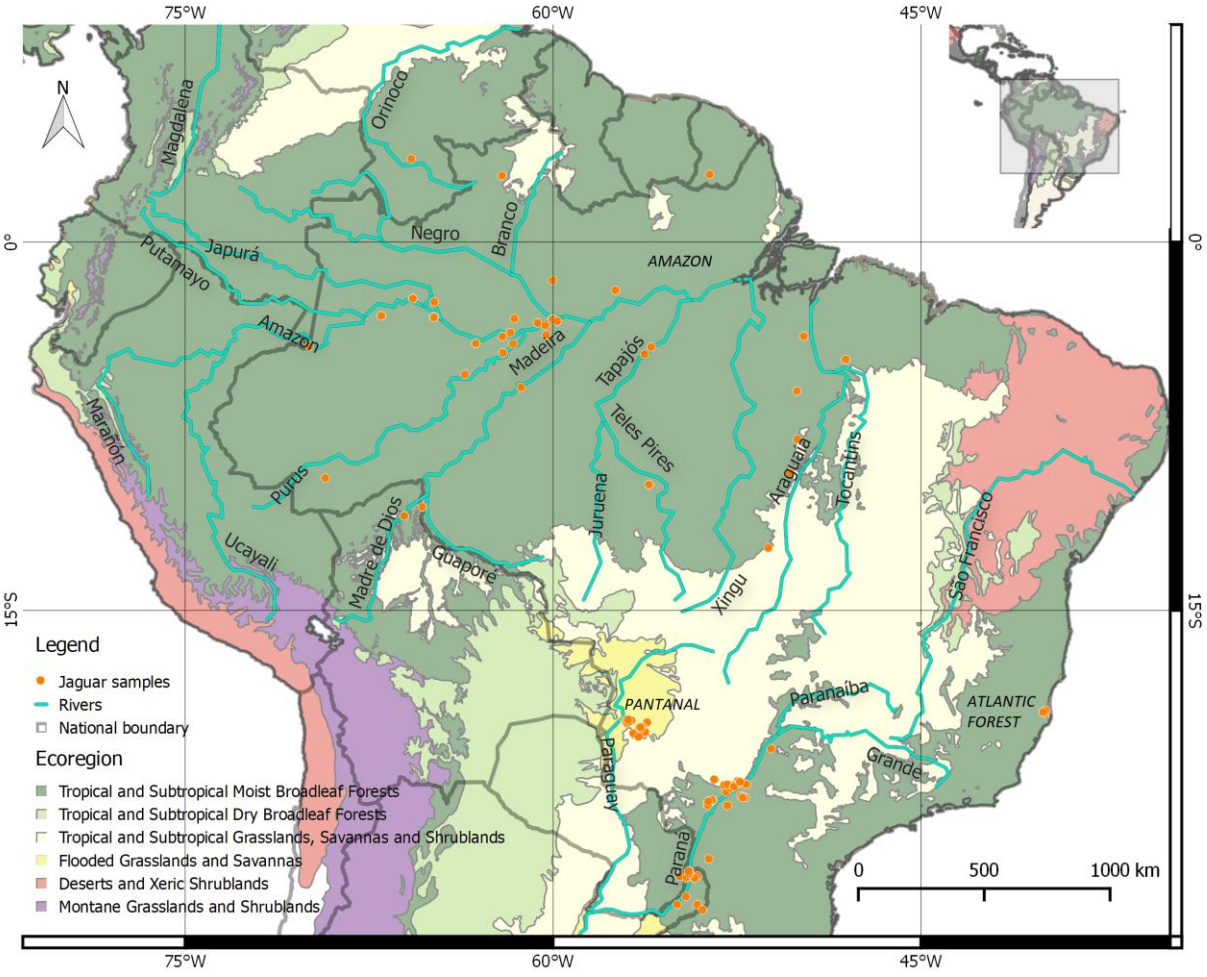
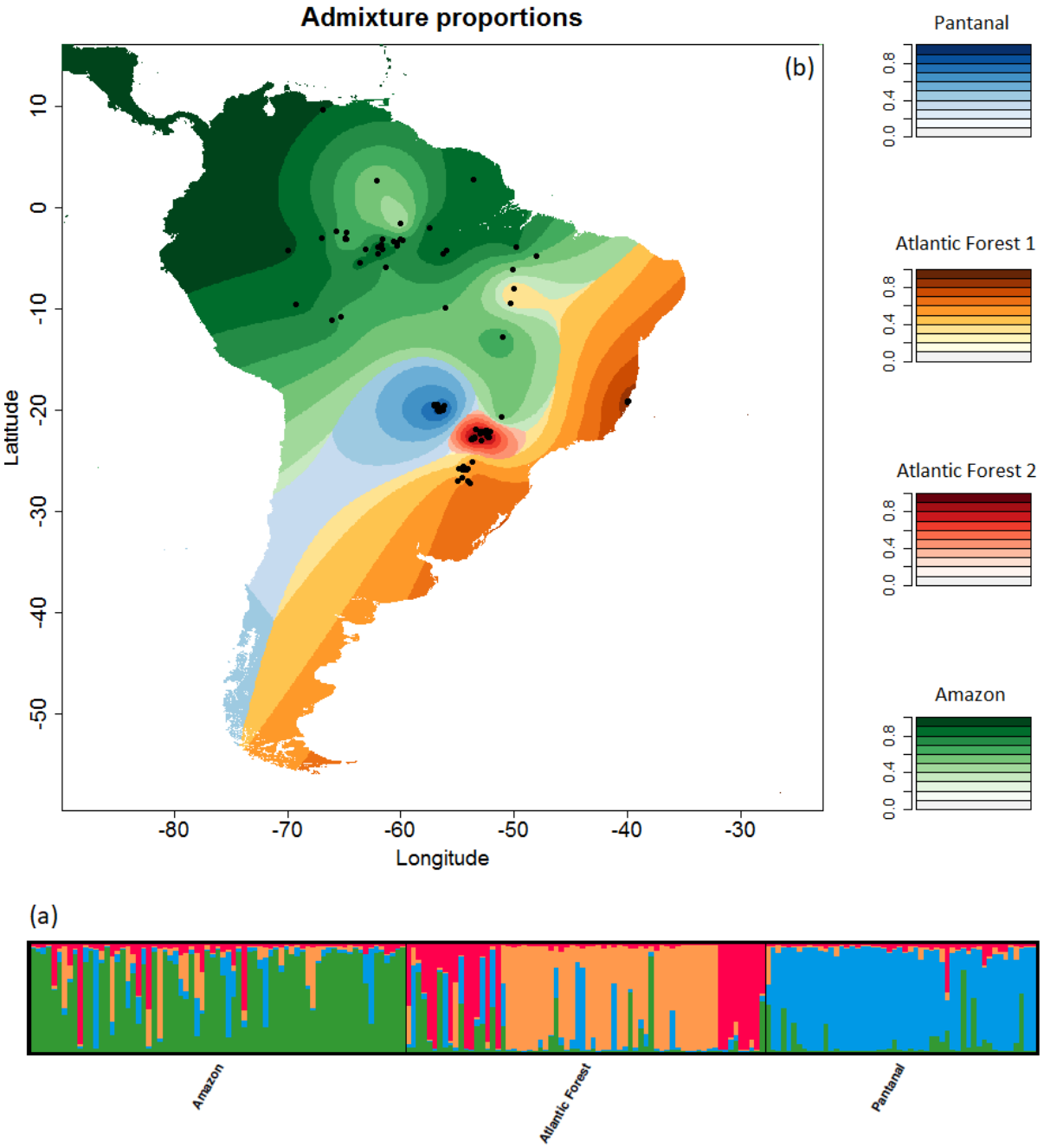


Figure 1

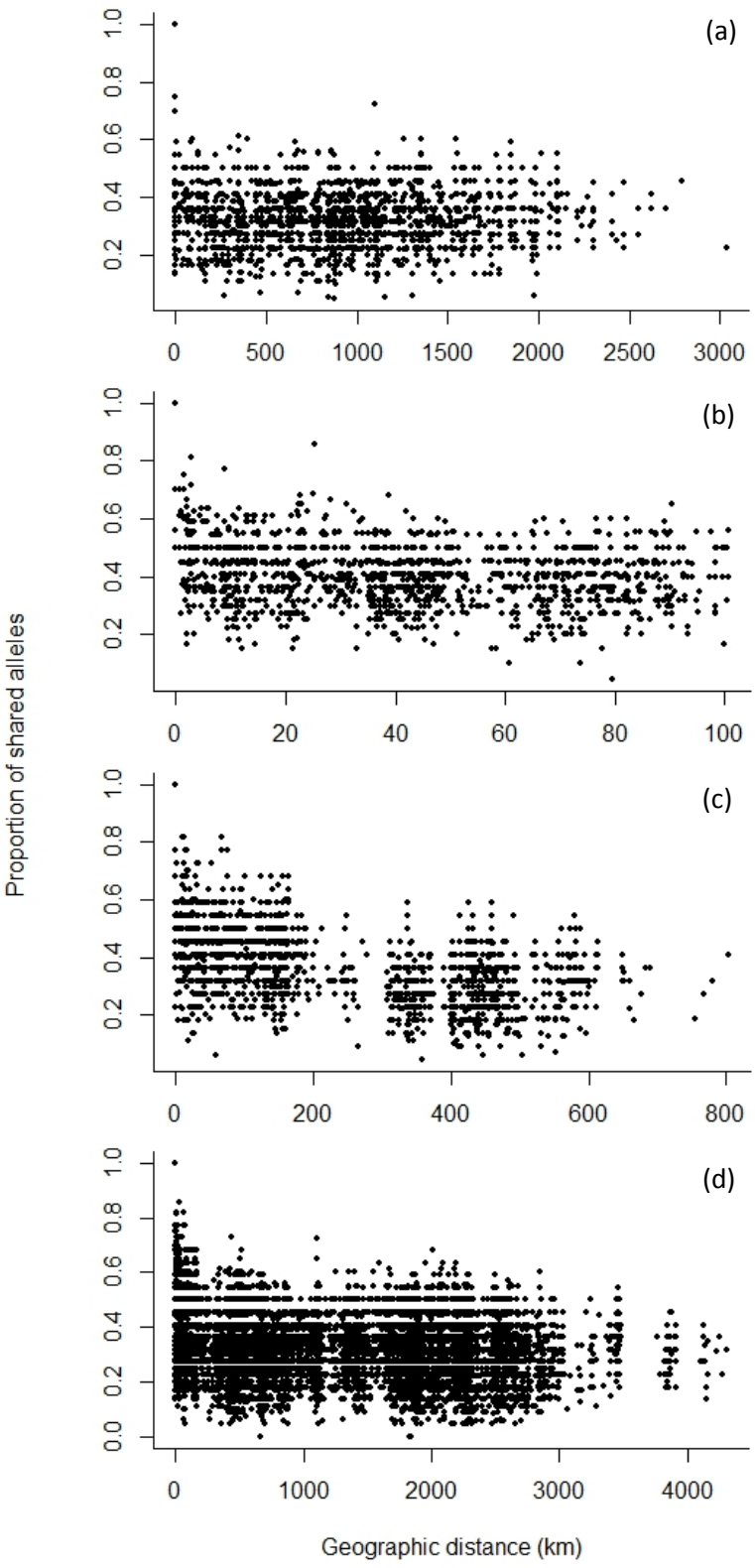
Study area in South America. Points represent the sampling location for genotyped jaguars.



779

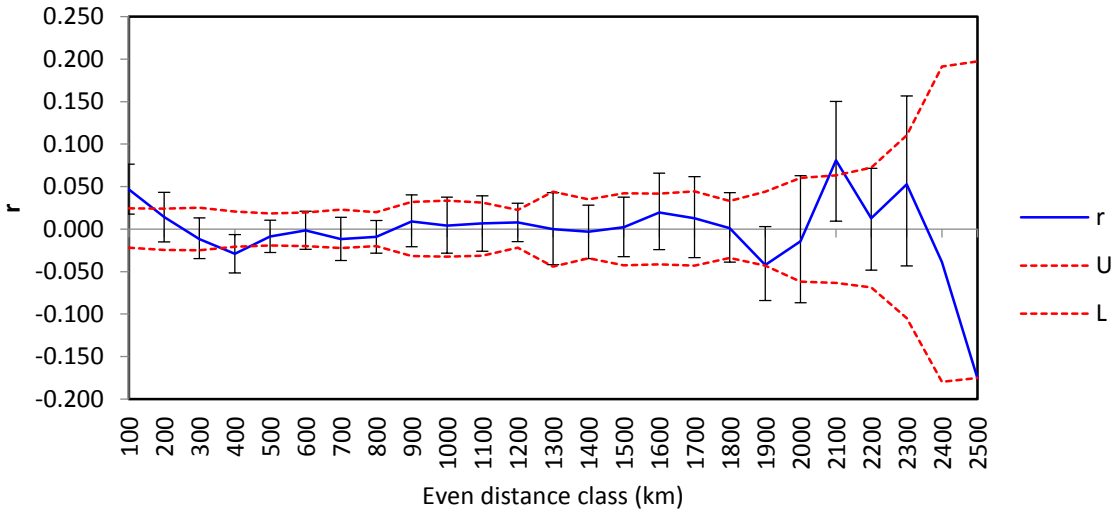
780 **Figure 2**

781 Jaguar population structure in three South American biomes. Inference of genetic clusters (*K*) was
782 based on the Puechmaille method using correlated allele frequencies. (a) Vertical bars represents each
783 individual jaguar, and the colour of the bar shows the percentage of membership (*Q*) to the distinct
784 clusters. (b) Spatially-explicit interpolation of admixture coefficients.



785 **Figure 3**

786 Regional isolation-by-distance patterns in South American jaguars. Isolation-by-distance was assessed
787 by plotting pairwise proportion of shared alleles calculated in GenAlEx, versus pairwise Euclidean
788 distances (km) across the (a) Amazon, (b) Pantanal, (c) Atlantic Forest and (d) the three populations
789 altogether



790

791 **Figure 4**

792 Correlogram showing spatial autocorrelation for Amazon jaguars. The genetic correlation coefficient
793 (*r*) is plotted as a function of geographic distance across defined spatial distance classes (100 km).
794 Dashed red lines represent upper (*U*) and lower (*L*) bounds of the null hypothesis of no spatial structure
795 based on 10,000 random permutations. Error bars represent 95% confidence intervals for *r* estimates,
796 based on 1,000 bootstrap replications.

797

CHAPTER III Comparative assessment of GBS (RADseq) and whole-exome sequencing for estimating genetic diversity and geographic structure in natural jaguar populations

Comparative assessment of GBS (RADseq) and whole-exome sequencing for estimating genetic diversity and geographic structure in natural jaguar populations

Running title: Jaguar genomic diversity

Gustavo P. Lorenzana[1], Henrique V. Figueiró[1], Priscilla Villela[2], Luiz L. Coutinho[2], Eduardo Eizirik[1,3]

[1] Laboratório de Biologia Genômica e Molecular, PUCRS, Porto Alegre, Brazil

[2] Centro de Genômica Funcional, ESALQ-USP, Piracicaba, Brazil

[3] Instituto Pró-Carnívoros, Atibaia, Brazil

Acknowledgments

GL, HVF and EE acknowledge CNPq/Brazil and CAPES/Brazil for financial support.

Abstract

Aim

Biologists currently have an assortment of high-throughput sequencing techniques that allow them to study population dynamics at an increasing level of detail. The utility of diversity estimates, especially for threatened species such a as the jaguar (*Panther onca*) depends on their robustness to recover meaningful approximations while filtering out noise produced by artifacts. Employing two genome-wide reduced representation approaches, we obtained population-level

summary statistics for five jaguar demes occurring in major South American biomes that differ in habitat expanse and productivity, as well as in the intensity of anthropic threats.

Location

Amazon, Atlantic Forest, Cerrado, Caatinga and Pantanal biomes, South America.

Methods

Using one restriction-enzyme-based (or RADseq) approach (genotyping-by-sequencing, GBS) and one targeted capture approach (whole-exome sequencing, WES), we obtained thousands of SNPs for the same set of jaguar individuals (n=20). We estimated expected heterozygosity (H_{exp}) and nucleotide diversity (π) for the focal populations, and estimated F_{st} values among them, using the *populations* program implemented in the package *Stacks*. We then repeated the process allocating the full dataset into two subsets that equalized sampling size. For each dataset, we varied the magnitude of three filtering parameters (*r*, *p* and MAF), which control the number of joint SNPs parsed within and among populations. We then compared GBS and WES estimates through the joint estimation of H_{exp} and π, using each jaguar deme as a replicate for each batch.

Results

Changes in parametrization had measurable differences in summary statistics for each jaguar deme, both between approaches and among distinct analytical batches within each approach, especially for GBS. H_{exp} and π were consistently higher for the Amazonian and Pantanal populations, with Caatinga exhibiting the lowest diversity and higher pairwise F_{st} values. For the *joint* estimates, we found no statistically significant differences between approaches and batches, and we observed that the application of intermediate stringency filtering for population characterization provided the most precise and consistent results across the analyses.

Main conclusions

To our knowledge, this is the first instance of simultaneous use of GBS and WES for estimating population genetic parameters from the same set of individuals. Our results show that some parameters do influence estimates of diversity and differentiation in ways that may not be fully predictable, highlighting the importance of careful fine-tuning of parameters for obtaining robust and unbiased genomic diversity estimates. As expected, the Amazon and Pantanal biomes sustain more diverse jaguar populations in comparison with adjacent biomes, reaffirming their status as major global strongholds for jaguars.

Keywords: Genotyping-by-sequence, whole exome sequencing, SNP, *Panthera onca*, tropical biomes.

INTRODUCTION

Genetic diversity is a key feature of species and populations, representing a major surrogate for setting priorities and guidelines in conservation biology, and thus requiring robust and unbiased estimates (Moritz 2002; Coates et al., 2018). Until recently, variability metrics were calculated using a few traditional genetic markers such as mitochondrial DNA and microsatellite loci (Allendorf et al., 2010), but currently novel genotyping methods that rely on a reduced representation of whole genomes represent a valuable choice to investigate non-model species (Andrews & Luikart, 2014). Collectively known as “genomic enrichment” or “reduced representation sequencing” (RRS), they include several techniques that vary on aspects such as initial quantity and quality of extracted DNA, library construction, means of locus identification and reconstruction, SNP calling, and downstream utilities (Harvey et al., 2016).

Two major RRS approaches are targeted capture (Ng et al., 2009; Mamanova et al., 2010; Jones & Good, 2015), and restriction-site associated DNA sequencing (RADseq, Hohenloe et al., 2010; Andrews et al., 2016). In the former, specific portions of the genome, such as ultraconserved elements (Smith et al., 2013) or coding regions (Hodges et al., 2007), are the intended targets to which the sequencing efforts are directed. In the latter, sequencing is focused on genomic regions located next to cut sites recognized by restriction enzymes (Campbell et al., 2018). A variety of protocols exists within the restriction enzyme subdivision, and sometimes the terms are used interchangeably. For our purposes, we recognize genotyping-by sequencing (GBS), as proposed by Elshire et al. (2011), as nested within RADseq, as this latter term is more intuitively informative.

Among target capture techniques, whole-exome sequencing (WES) has been primarily used in biomedical research (Pabinger et al., 2013), with some recent studies applying it for wildlife species (e.g. Förster et al., 2018). On the other hand, RADseq has been mostly applied to characterize genomic diversity in fisheries, crop varieties, and livestock (Gorjanc et al., 2015; Kim et al., 2016; Li & Wang, 2017), although its use in wildlife species has steadily increased in recent years (Andrews et al., 2016). The main potential of WES lies in its ability to detect variants in functional regions, a feature that until very recently was out of reach for molecular studies using neutral markers. However, the need of a reference genome or a transcriptome, used as a template for the design of probes, represents a constraint for studying species with scarce genetic resources. The main advantage of RADseq lies in the abundance of cut sites along the genome and the fact that a reference genome of the focal species is not strictly necessary to generate large panels of genomic markers at low cost (Angeloni et al., 2011).

Despite their growing popularity in genetic research, there are virtually no direct assessments comparing the performance of different RRS approaches for population genomic applications in non-model species. For example, most WES studies have a biomedical scope, aiming

to identify variants that either predispose to particular diseases in humans (e.g. schizophrenia or autism, Gilissen et al., 2012), confer resistance to physiological risk factors (e.g. ability to thrive on a fat-rich animal diet, Hsieh et al., 2017), or prompt local adaptation to harsh environments (e.g. high-altitude genes, Yi et al., 2010). Conversely, the fact that RADseq markers are generally anonymous (i.e. their genomic position is not known) limits their capability to explore these issues directly (but see Catchen et al., 2017). As a consequence, there is a scarcity of controlled datasets (e.g. comprising the same individuals) with which these two approaches can be rigorously compared.

Such a comparison would be useful in the context of assessing the accuracy of population genetic estimates derived from these two genome-wide approaches. For example, some researchers have pointed out issues when dealing with RADseq data (Davey et al., 2011), arguing that they can either underestimate (Arnold et al, 2013; Cariou et al., 2016) or overestimate (Gautier et al., 2013) genetic diversity, potentially leading to biased inferences. In spite of the relevance of further testing these possibilities, such an assessment is usually hampered by the lack of comparable datasets for RADseq and an independent genome-wide approach such as WES.

To address this issue, we compared the performance of two RRS approaches (RADseq and WES) using samples from the same set of wild-caught jaguar (*Panthera onca*) individuals. We estimated genetic diversity and population structure metrics, focusing primarily on the Amazon and the Pantanal biomes, two major global strongholds for the species, for which we had a larger sample size, complemented with additional samples coming from the Atlantic Forest, Cerrado and Caatinga biomes. Specifically, we tested several combinations of parameters for assembling loci and calling and filtering SNPs, aiming to assess congruency among estimated metrics, quantify approach-specific biases, and provide recommendations for future studies. In spite of the conceptual and

119 methodological differences between approaches, we were interested in assessing their ability to
120 recover consistent genetic signals for this elusive carnivore.
121
122 METHODS
123 We comparatively assessed the performance of exome-capture and GBS approaches,
124 selecting 20 individuals for which the former approach had been performed in a previous study
125 (Figueiró et al., in prep.), and collected GBS data for them. The samples cover the five Brazilian
126 biomes where jaguars currently occur: Amazon and Pantanal, with *n*=7 each, as well as Atlantic
127 Forest, Cerrado and Caatinga, with *n*=2 each.
128 Exome probe design was performed from genomic data available for the five *Panthera*
129 species (i.e. tiger, lion, jaguar, leopard, and snow leopard). For the annotation and selection of genic
130 regions, Figueiró et al. (in prep.) followed the pipeline proposed by Bi et al. (2012), including in the
131 probes 500bp on each flank of the coding region of each gene. Genomic libraries were constructed
132 according to the protocol proposed by Meyer & Kirchner (2010). Exome capture was performed
133 using a custom-designed Nimblegen Capture Kit (Roche), and sequencing was conducted on two
134 lanes of the Illumina HiSeq 2500 platform, with 100bp reads and *ca.* 300bp insert size.
135 The GBS experiment was carried out following Elshire et al.'s (2011) protocol with minor
136 modifications. DNA extraction was performed with a *DNeasy Blood & Tissue* kit (Qiagen), and DNA
137 concentration and quality were assessed with Qubit (Invitrogen) and NanoDrop 2000 (Thermo
138 Scientific), respectively. Libraries were prepared using the *Pst*I restriction enzyme for digestion in a
139 35µL total volume containing 2µL DNA (50ng), 3.5µL NEB Buffer3, 0.8µL *Pst*I enzime (10U/µL) and
140 28.7µL water. A reaction for adaptor ligation was performed in 30µL: 6µL adaptors (0.06pmol); 5µL
141 T4 DNA Ligase buffer (10X); 1µL T4 DNA enzyme ligase (400U/µL; New England Biolabs) and 18µL
142 water. After adaptor ligation, samples were pooled by adding 10µL per sample, and purification was

143 performed using a *QIAquick PCR Purification* Kit (Qiagen). Restriction fragments were amplified in a
144 total volume of 50µL with 15µL of pooled DNA, 1X Taq Master Mix (New England Biolabs) and
145 20pmol of complementary primers. After PCR, the products were purified using magnetic beads
146 (Agencourt AMPure XP), with a fragment size of 200-450 bp, and libraries were quantified on real-
147 time PCR. Libraries were diluted to 2nMol, denatured and eluted to attain a clustering concentration
148 of 16pM. Sequencing was performed on one lane of the Illumina HiSeq2500 platform.

149 GBS raw reads were demultiplexed using a score *S*=30, and barcodes were removed in the
150 *process radtags* module of Stacks v2.0. Quality scores were checked with FastQC and cleaned reads
151 were processed with the *ref_map* and *de_novo* wrappers in Stacks v2.0. In the former case, raw
152 reads were aligned against the jaguar reference genome (Figueiró et al., 2017), using Bowtie2
153 (Langmead et al., 2012) with default parameters. For the *de novo* approach, we tested several
154 parameters to assemble loci from RAD *tags*, but for the purpose of this study, we focused on two of
155 these batches: *dnv-def*, with default parameters and *dnv-cov,* which increases coverage per locus.
156 The *m* parameter controls the number of reads required to assemble a *stack,* or putative locus, and
157 was set to *m*=3, and *m*=6 for each of those batches, respectively. The latter is the highest value that
158 *m* can attain in order to increase coverage before facing the potential problem of assembling paralog
159 loci (Paris et al., 2017; Rochette & Catchen 2017). The catalog loci resulting from these two batches
160 were then aligned against the reference genome, as advised by Paris et al. (2017), to generate two
161 additional batches: *dnv-defa* (de novo default aligned) and *dnv-cova* (de novo increased coverage
162 aligned).

163 SNP calling and population-based filtering are crucial given their effects on downstream
164 analysis and thus were the main aspects under scrutiny during this study. For the former, after
165 quality filtering and assessment of coverage per individual and per site, SNP calling was conducted

with Samtools v1.9 (Li et al., 2009) using up to 70% of individuals with site depth between 2x and 30x.

Exome-derived SNPs (Figueiró et al., in prep.) were split into three subsets: exon-derived (more likely to contain adaptive segments), flanking regions (more likely to contain neutral segments), and whole exome (exons + flanking regions). We used a *bed* file with genomic positions to include exons or otherwise exclude them (to obtain flanking regions), and output the respective VCF files in VCFtools. Then we ran a clustering analysis using the flanking region SNPs (enriched for neutral markers) to generate a PCA plot in R, to verify if we were indeed dealing with discrete populations (Fig. S2).

As for the GBS batches, two parameters determined the stringency of population-based filters applied to the SNP-calling process: i) the minimum proportion of individuals within a population required to process a locus (*r*), and ii) the minimum number of populations required to process a locus (*p*). We ran *populations* varying *p* from 1 to 5, in the cases of full and rarefied datasets, which included Amazon, Pantanal, Cerrado, Atlantic Forest, and Caatinga populations, and 1-2 for the Amazon-Pantanal dataset. Full datasets (*n*=20) contained all the individual samples belonging to the five populations. The rarefied dataset comprised the same samples for the latter three biomes, as well as two randomly selected individuals each from the Amazon and Pantanal. Finally, the Amazon-Pantanal (Ama-Pan) dataset (*n*=14), comprised only the individuals coming from those biomes. In all cases, *r* was set to 0.7. We applied different filtering strategies on these datasets in order to generate three groups of datasets (i.e. lax, intermediate, and stringent), each encompassing both WES and GBS batches (Table 1).

We comparatively assessed genetic diversity and population differentiation using standard summary statistics obtained from WES- and GBS-derived SNP subsets. We focused on two metrics, expected heterozygosity (or gene diversity) and nucleotide diversity (π), which were obtained

running the *populations* module versions 2.0 and 2.1 for GBS and WES, respectively. The newer version was use for the WES data because v2.0 exhibited a bug that impeded parsing external (i.e. not generated in Stacks) VCF files. To test for statistically significant differences between reduced-representation approaches, we applied a Wilcoxon test. To see if the populations were significantly different we used a Kruskal-Wallis test. Population structure was assessed using the fixation index F_{st}, and smoothed pairwise F_{st} between Pantanal and Amazon, using a custom script provided by Rochette & Catchen (2017). Final comparisons were carried out among r*efmap*, d*nv-cova* and *flank* batches, as these harbored most of the global variability among batches and approaches.

We were interested in testing the reduced-representation approaches at three levels: 1) internal consistency within approaches. We wanted to see if different GBS batches recovered similar signals of diversity and differentiation among populations, and the same reasoning was applied to WES batches; 2) consistency between approaches. To look for differences between the metrics estimated by the two approaches, and to assess the probable causes for the observed patterns; and 3) biological comparison among biomes, to determine if some populations are more diverse than others.

RESULTS

The exome-capture experiment covered 25,441 genes (Figueiró et al. in prep). From this, we identified 115,649 variant sites distributed in 2050 scaffolds. Of these variant sites, 103,376 were located within exons and 12,273 were located in their flanking regions. SNP calling and filtering resulted in 0% missing data across the 20 samples, and thus there was no need for further filtering using the *r* and *p* parameters applied to the GBS dataset.

As for the GBS experiment, the sequencing lane comprised 53 individuals (jaguars and other Neotropical felid species [not shown]), yielding 190×10^6 raw reads. Of these, 139×10^6 were retained

after demultiplexing. The mean number of reads per individual was 2.5×10^6 (ranging from 58 to 9.8×10^6). For the 20 jaguar individuals analyzed in the study, approximately 68×10^6 raw reads were retrieved through the *process radtags* module. Of these, 8.7×10^5 were removed due to ambiguous *RADtag* cut sites (no mismatch allowed), and 14.5×10^6 did not meet our quality criteria (score limit *s*=30), thus retaining 52.5×10^6 reads for downstream analyses. The mean number of retained reads per individual sample in this subset was 2.62×10^6 (ranging from 4×10^5 to 9.8×10^6; Figure S1). The mean number (and range) of retained reads per putative population were as follows: Amazon: 3.7×10^6 (4×10^5 to 9.8×10^6); Pantanal: 1.7×10^6 (4.3×10^5 to 3.2×10^6); Cerrado: 1.8×10^6 (1×10^6 to 2.7×10^6); Atlantic Forest: 2.3×10^6 (9.5×10^5 to 3.7×10^6); and Caatinga: 2.7×10^6 (2.2×10^6 to 3.1×10^6). Considering all batches across the three datasets and filter stringency, the mean number of retained loci and SNPs after population filtering ranged from 4,857 to 348,870 (Tables 1, S1).

Genetic diversity metrics

Expected heterozygosity

Gene diversity ranged from 0.08 to 0.22 and from 0.12 to 0.28 for full and rarefied datasets, respectively (Fig S3). In both datasets, lax and intermediate filtering yielded higher estimates for GBS batches than for whole-exome-sequencing (WES) batches, while strict filtering led to the opposite results (Fig. S3 top row). Rarefied datasets showed the same results as full datasets, with a higher variability in estimated values across datasets, although they yielded estimates that are more consistent across internal batches (Fig S3, middle row). The Amazon-Pantanal (AMPA) dataset yielded closer estimates among batches, with a large effect of strict filtering based on MAF, probably because of the exclusion of rare alleles, which inflated the estimates of the remaining alleles up to near 0.35 (Fig. S3 bottom row).

Nucleotide diversity

In general, nucleotide diversity (π) values showed great variability, from 0.2 to 0.35. Again, for full and rarefied datasets, lax and intermediate filtering yielded higher estimates for GBS than for WES, while strict filtering caused the opposite pattern, and intermediate filtering was more stable. In the Amazon-Pantanal dataset, GBS batches provided higher values than exome batches, especially when using strict filtering, attaining π=0.35 in that case. This is an expected outcome as the exome region is more conserved (due to purifying selection) than the more random regions covered by GBS data.

Population structure

Across all batches, the highest F_{st} values where consistently observed between Atlantic Forest and Caatinga, both of which are subjected to severe anthropic disturbance, while the lowest values were found between the larger, contiguous biomes of the Amazon and Cerrado (Fig. 3). In those cases, differentiation was moderate, surpassing F_{st} values of 0.20. In general, WES estimates were lower than those obtained with GBS data.

DISCUSSION

Approach consistency and population genomic insights

First level: Internal consistency within approaches.

The GBS runs showed the following results: a) the *ref map* batch recovered lower diversity (expected heterozygosity and nucleotide diversity) values than *de novo* batches. b) Increasing the number of reads (*stacks*) required to assemble a *de novo* locus (and thus achieving higher mean coverage) as well as c) aligning the catalog loci to a reference genome, further increased the diversity estimates. On the other hand, exome-derived runs showed very similar metrics, with exons showing slightly lower diversity than flanking regions. As the vast majority of SNPs were located in

exons, the Whole exome estimates showed closer values to the exonic results than to those of the flanking regions.

Second level: Consistency between approaches.

The comparison between the two approaches revealed several trends: a) null population filtering (p=1) retrieved higher values of diversity and population differentiation for the GBS runs in comparison with WES batches, especially for the Amazon and Pantanal populations; b) intermediate filtering (p=3) retrieved more similar estimates between approaches; and c) when population filtering was set to maximum (p=5), GBS metrics decreased sharply even below WES values.

Third level: Biological comparison among populations.

From a population perspective, the data showed two trends: a) diversity metrics for the Amazon population were consistently higher than those of the other assessed populations; b) a fairly regular difference among biomes was observed in most batches, with the Amazon being the most diverse biome, and the Caatinga the least diverse, with the Pantanal, Cerrado and Atlantic Forest lying in between.

We further compared the performance of the GBS and WES batches, through the *joint* estimation of diversity metrics using the H_e and π values for each jaguar populations as a replicate for each batch. Accordingly, we recalculated those metrics with five replicas per batch (Figs 1-2). This analysis showed that the differences observed between GBS and WES batches were not statistically significant, with a few exceptions for the strictly filtered datasets. Thus, we favored the intermediate filtering approach as the more adequate in most instances, providing more precise and consistent estimates across batches and approaches even at low sample sizes, such as the ones used in this study (Figs. 1-2 middle panels).

Even though the two discussed approaches yielded equivalent inferences for our purposes, there are logistical and cost considerations that could lead one to decide to use one method over

285 the other. We encourage the use of RADseq approaches to tackle population genomics research in
286 species with few available genomic resources, as has been stated elsewhere (Andrews et al., 2016).
287 However, we should stress some precautions, especially for wet lab practices. One identified issue
288 with RADseq data is the unequal proportion of sequenced reads among samples, which constitutes
289 an experimental error (Gautier et al., 2013). In our case, we observed a 24-fold difference between
290 the two extreme samples with less (bPon167) and more (bPon377) retained reads (Figure S1).
291 In this study, we compared WES- and GBS-derived SNP data. The results showed a scenario
292 where the approaches are congruent and the degree of coherence and precision largely depends
293 on the magnitude of population filters applied. Ultimately, we found no statistical differences
294 between batches. However, it has been shown that RADseq may systematically underestimate
295 variant sites due to the presence of polymorphisms at restriction sites, which makes it impossible
296 to observe the associated allele (Arnold et al., 2013; Gautier et al., 2013). This allele dropout
297 phenomenon at restriction sites (Cariou et al., 2016; Nunziata & Weisrock 2018) leads to an
298 underestimation of genetic diversity. One way that has been proposed to circumvent this issue is
299 increasing coverage. Our results supported this assertion, as increasing coverage in the *De Novo*
300 batches was accompanied by a concomitant increase in diversity metrics, approaching the values
301 observed with the WES batches.
302 As described by Paris et al. (2017), we found supporting evidence that the *ref map* wrapper
303 acts as a filter in itself, because it aligns each raw read to the reference genome, while the *de novo*
304 module first identifies putative alleles from every individual in a sampled population and then
305 merges them into putative loci, thus recovering more loci. For our GBS datasets, *ref map* consistently
306 yielded lower estimates than *dnv-cov* batches.
307 Given budgetary and bioinformatics constraints, for population genomic applications we
308 recommend the following. In cases where a reference genome is not available, but a large number

of samples across several populations are, it is better to employ RADseq. Conversely, when a reference genome is obtainable, but only a few populations have been sampled, and the inter-population environmental heterogeneity is high, it could be advantageous to choose WES approaches, especially when one is interested in assessing adaptive variation. Currently, both of these approaches are still more practical for population genomic applications than whole-genome sequencing. Overall, our results demonstrate the usefulness of performing controlled comparisons between different approaches to generate genome-wide markers for population-level studies, and highlights the analytical factors that may lead to consistent *versus* discrepant results emerging from these methods.

REFERENCES

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11, 697–709.

Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. *Molecular Ecology*, 23, 1661–1667.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G. & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17, 81.

Angeloni, F., Wagemaker, N., Vergeer, P., & Ouborg, J. (2012). Genomic toolboxes for conservation biologists. *Evolutionary Applications*, 5, 130–143.

Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, 22, 3179–3190.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., & Good, J. M. (2012). Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, 13, 403.

Buerkle, A. C., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: how low should we go?. *Molecular Ecology*, 22, 3028–3035.

Campbell, E. O., Brunet, B. M. T., Dupuis, J. R. & Sperling, F. A. H. (2018). Would an RRS by any other name sound as RAD? *Methods in Ecology and Evolution*, 23, 1920–1927.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics*, 1, 171–182.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140.

Catchen, J. M, Hohenlohe, P. A., Bernatchez, L., Andrews, K. R. & F. W. Allendorf. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources* 17, 362–365.

Cariou, M., Duret, L., & Charlat, S. (2013). Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution*, 3, 846–852.

Cariou, M., Duret, L., & Charlat, S. (2016). How and how much does RAD-seq bias genetic diversity estimates?. BMC evolutionary biology, 16(1), 240.

Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, 6, 165.

354 Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special
355 features of RAD Sequencing data: implications for genotyping. *Molecular Ecology*, 22, 3151–
356 3164.
357 Dussex, N., Taylor, H. R., Stovall, W. R., Rutherford, K., Dodds, K. G., Clarke, S. M., & Gemmell, N. J.
358 (2018). Reduced representation sequencing detects only subtle regional structure in a
359 heavily exploited and rapidly recolonizing marine mammal species. *Ecology and Evolution,*
360 8, 8736–8749.
361 Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011).
362 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS*
363 *One*, 6, e19379.
364 Förster, D. W., Bull, J. K., Lenz, D., Autenrieth, M., Paijmans, J. L., Kraus, R. H., Nowak, C., Bayerl, H.,
365 Kuehn, R., Saveljev, A. P., Sindičić, M., Hofreiter, M., Schmidt, K., & Fickel, J. (2018). Targeted
366 resequencing of coding DNA sequences for SNP discovery in nonmodel species. *Molecular*
367 *Ecology Resources*, 18, 1356–1373.
368 Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J.-M. & Estoup, A.
369 (2013). The effect of RAD allele dropout on the estimation of genetic variation within and
370 between populations. *Molecular Ecology*, 22, 3165–3178.
371 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. (2012). Disease gene identification
372 strategies for exome sequencing. *European Journal of Human Genetics* 20, 490–497.
373 Gorjanc, G., Cleveland, M. A., Houston, R. D., & Hickey, J. M. (2015). Potential of genotyping-by-
374 sequencing for genomic selection in livestock populations. *Genetics Selection Evolution*, 47,
375 12.

376 Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture
377 versus restriction site associated DNA sequencing for shallow systematics. *Systematic*
378 *Biology*, 65, 910–924.
379 Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., Middle, C. W., Rodesch, M. J.,
380 Albert, T. J., Hannon, G. J. & McCombie, W. R. (2007). Genome-wide in situ exon capture for
381 selective resequencing. *Nature Genetics*, 39, 1522.
382 Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010).
383 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD
384 tags. *PLoS Genetics*, 6(2), e1000862.
385 Heppenheimer E. et al., 2018 High genomic diversity and candidate genes under selection
386 associated with range expansion in eastern coyote (*Canis latrans*) populations. *Ecology and*
387 *Evolution*, DOI: 10.1002/ece3.4688.
388 Hsieh, P., Hallmark, B., Watkins, J., Karafet, T. M., Osipova, L. P., Gutenkunst, R. N., & Hammer, M.
389 F. (2017). Exome sequencing provides evidence of polygenic adaptation to a fat-rich animal
390 diet in indigenous Siberian populations. Molecular Biology and Evolution, 34, 2913–2926.
391 Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics.
392 *Molecular Ecology*, 25, 185–202.
393 Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L. S., & Paterson, A. H. (2016). Application of
394 genotyping by sequencing technology to a variety of crop breeding programs. *Plant Science*,
395 242, 14-22.
396 Langmead, B., & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*,
397 9:357-359.

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., & Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, 7, 355–369.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Li, Y. H., & Wang, H. P. (2017). Advances of genotyping-by-sequencing in fisheries and aquaculture. *Reviews in Fish Biology and Fisheries*, 27, 535–559.

Luca, F., Hudson, R. R., Witonsky, D. B., & Di Rienzo, A. (2011). A reduced representation approach to population genetic analyses and applications to human evolution. *Genome Research*, 21, 1087–1098.

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J. & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, 7, 111.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15, 28–41.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 6, pdb-prot5448.

Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, 51, 238–254.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R. & Hohenlohe, P. A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22**,** 2841–2847.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A., & Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461, 272–276.

Pabinger, S., A. Dander, M. Fischer, R. Snajder, M. Sperk, M. Efremova, B. Krabichler, M. R. Speicher, J. Zschocke, and Z. Trajanoski. 2014. A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*, 15, 256–78.

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for STACKS. *Methods in Ecology and Evolution*, 8, 1360–1373.

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12, 2640.

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2013). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic biology*, 63, 83–95.

Smitz N, Jouvenet O, Ambwene Ligate F, Crosmary WG, Ikanda D, et al., 2018. A genome-wide data assessment of the African lion (*Panthera leo*) population genetic structure and diversity in Tanzania. *PLOS ONE*, 13: e0205395.

Suchan, Tomasz, Camille Pitteloud, Nadezhda S Gerasimova, and Anna Kostikova. 2016. Hybridization Capture Using RAD Probes (HyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. *Plos One* 1–22. doi:10.1371/journal.pone.0151651.

Szulkin, M., P. A. Gagnaire, N. Bierne, and A. Charmantier. 2016. Population Genomic Footprints of Fine-Scale Differentiation between Habitats in Mediterranean Blue Tits. *Molecular Ecology*, 25: 542–558.

Vega, R., Vázquez-Domínguez, E., White, T. A., Valenzuela-Galván, D., & Searle, J. B. (2017). Population genomics applications for conservation: the case of the tropical dry forest dweller *Peromyscus melanophrys*. *Conservation Genetics*, 18, 313–326.

Wang, J. Raskin, L., Samuels, D.C., Shyr Y, Guo Y. 2015. Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics* 31: 318–323.

Willing E-M, Dreyer C, van Oosterhout C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE*, 7(8): e42649.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X., Pool, J. E., Xu, X., Jiang, H., Vinckenbosch, N. Korneliussen, T. S. et al., (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 329, 75–78.

468

469

470 TABLES

471

Table 1. Dataset partition according to sample size and filtering parameters. *p* represents the minimum number of populations in which a locus must be present for it to be processed, while *r** is the minimum percentage of individuals in a population required to process a locus for that population. MAF, minor allele frequency.

Dataset	Sample size	Filters		
		Lax	Intermediate	Stringent
Full	20	*p*=1	*p=3*	*p=5*
Rarefied	10	*p*=1	*p=3*	*p=5*
Ama-Pan	14	*p*=1	*p=2*	*p=2;* MAF=0.1

472 *In all cases, *r* was set to 0.7.

Table 2. Summary of number of loci and SNPs per approach and stringency of population filters

			GBS			WES		
dataset	filter	n	$\bar{x}$	min	max	$\bar{x}$	min	max
FULL	lax	Loci	218446	138581	348870	77099.3	12273	115649
		SNP	37309	20590	55544	61844.0	9855	92766
	int	Loci	59549.8	27221	103950	77099.3	12273	115649
		SNP	24998.6	11797	40645	61844.0	9855	92766
	str	Loci	20843.8	8266	47297	77099.3	12273	115649
		SNP	8802.2	3874	17255	61844.0	9855	92766
RAREFIED	lax	Loci	218327.4	144736	338108	77099.3	12273	115649
		SNP	37079.4	26673	49208	49926.0	7996	74889
	int	Loci	69900	41994	100469	77099.3	12273	115649
		SNP	23708.8	13615	35328	49926.0	7996	74889
	str	Loci	17567.6	8425	27317	77099.3	12273	115649
		SNP	7347.8	3224	13980	49926.0	7996	74889
AMPA	lax	Loci	213732.4	135930	342090	77099.3	12273	115649
		SNP	30715.2	16952	45707	56619.3	9041	84929
	int	Loci	41736.4	21440	65333	77099.3	12273	115649
		SNP	17481.8	8650	28663	56619.3	9041	84929
	str	Loci	34553.4	17982	53042	30150.0	4857	45225
		SNP	7680.8	3949	11901	30150.0	4857	45225

FIGURES

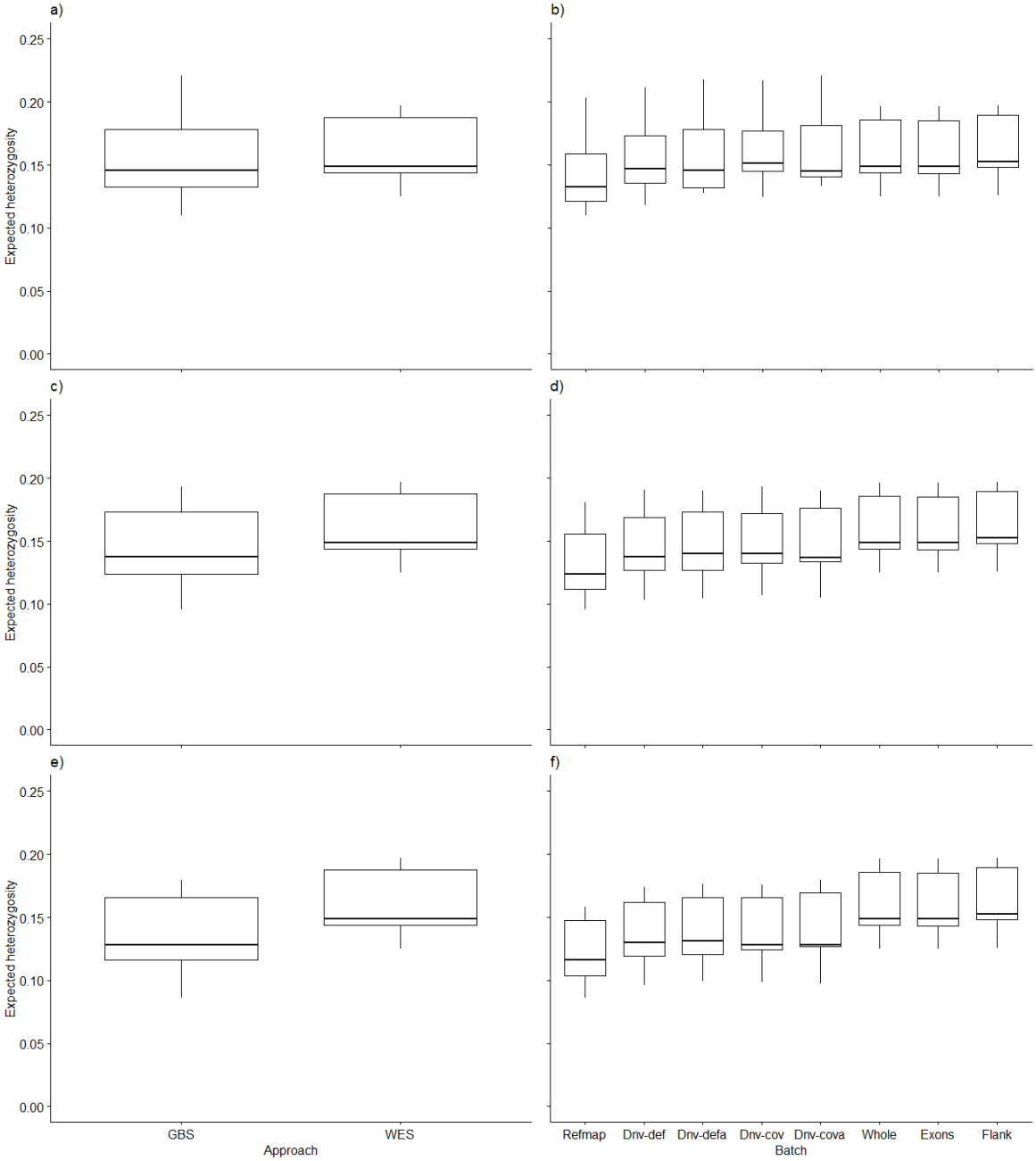


Fig 1a. Mean *joint* expected heterozygosity (H_{exp}) calculated per batch and reduced representation approach, using H_{exp} estimates of five jaguar populations as replicates for each batch. Full dataset (n=20): a-b) lax filter; c-d) intermediate filter; e-f) strict filter. No significant differences were found except for e) [Kruskal-Wallis chi-squared = 5.8398, df = 1, p-value = 0.01567].

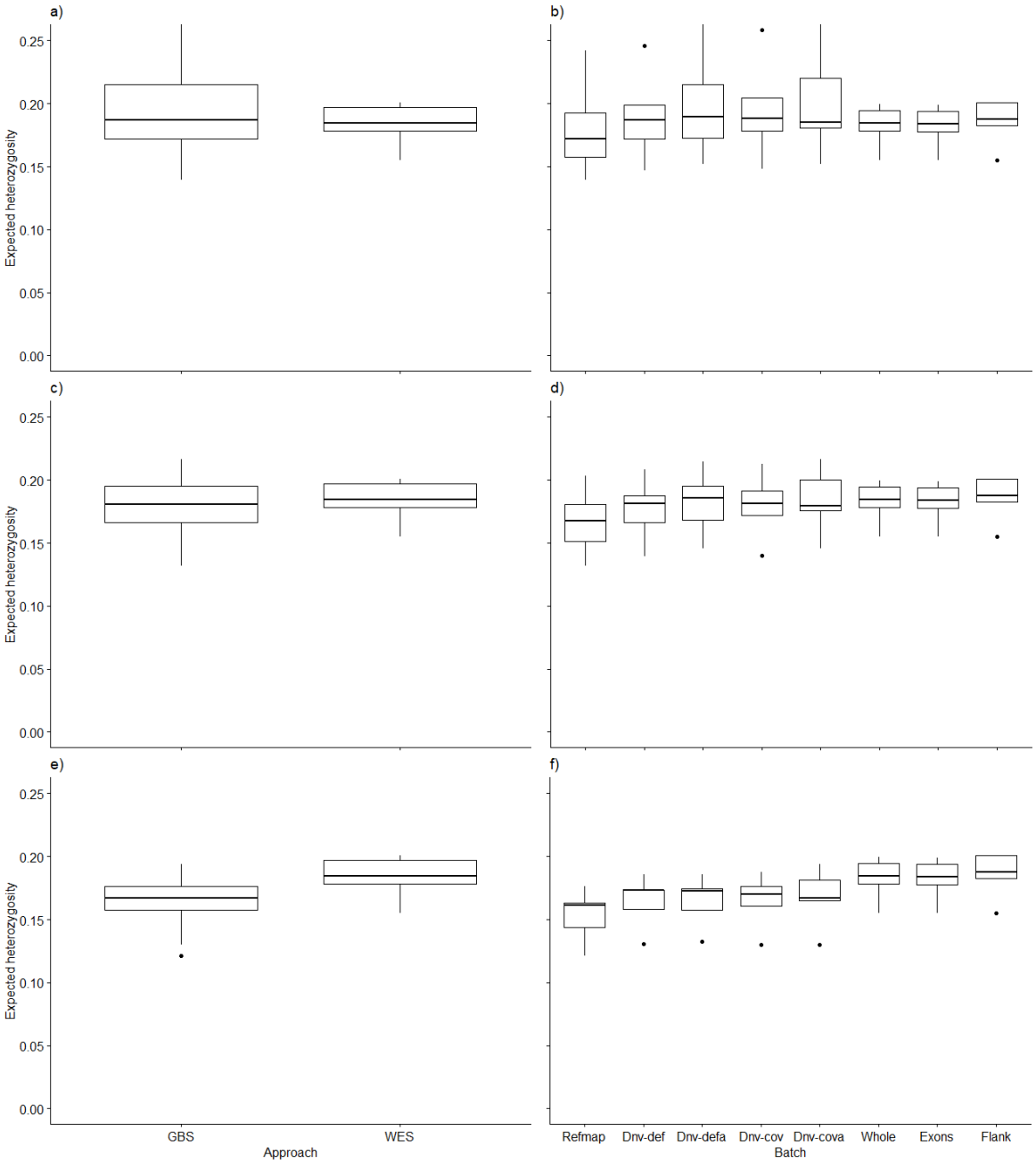


Fig 1b. Mean *joint* expected heterozygosity (H_{exp}) calculated per batch and reduced representation approach*,* using H_{exp} estimates of five jaguar populations as replicates for each batch. Rarefied dataset (n=10): a-b) lax filter; c-d) intermediate filter; e-f) strict filter. No significant differences were found except for e) [Kruskal-Wallis chi-squared = 8.8533, df = 1, p-value = 0.002926].

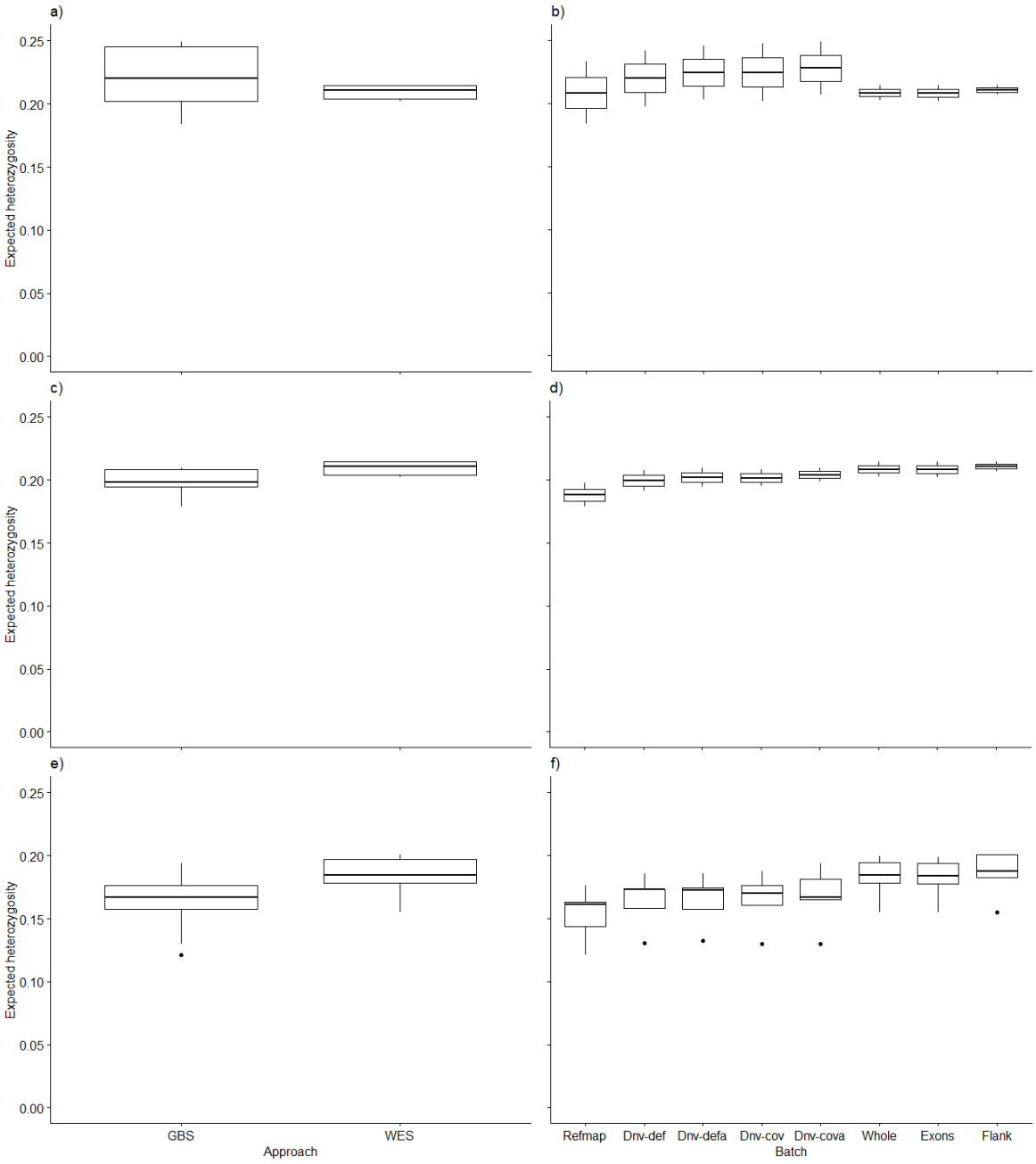


Fig 1c. Mean *joint* expected heterozygosity (H_{exp}) calculated per batch and reduced representation approach*,* using H_{exp} estimates of two jaguar populations as replicates for each batch. Amazon-Pantanal dataset (n=14): a-b) lax filter; c-d) intermediate filter; e-f) strict filter. No significant differences were found except for e) [Kruskal-Wallis chi-squared = 8.8533, df = 1, p-value = 0.002926].

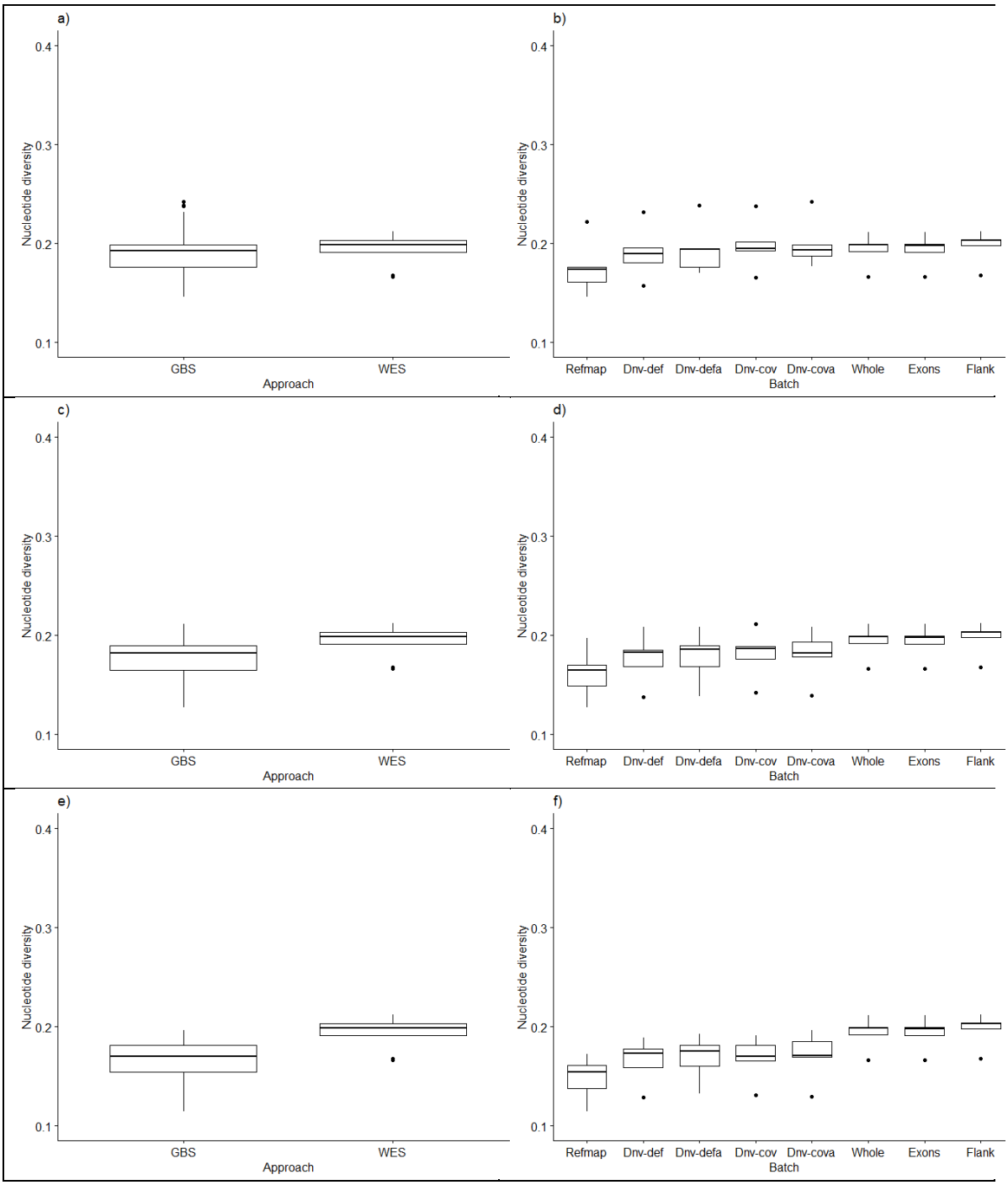


Fig 2a. Mean *joint* nucleotide diversity (*Pi*) calculated per batch and reduced representation approach*,* using *Pi* estimates of five jaguar populations as replicates for each batch. Full dataset (n=20): a-b) lax filter; c-d) intermediate filter; e-f) strict filter.

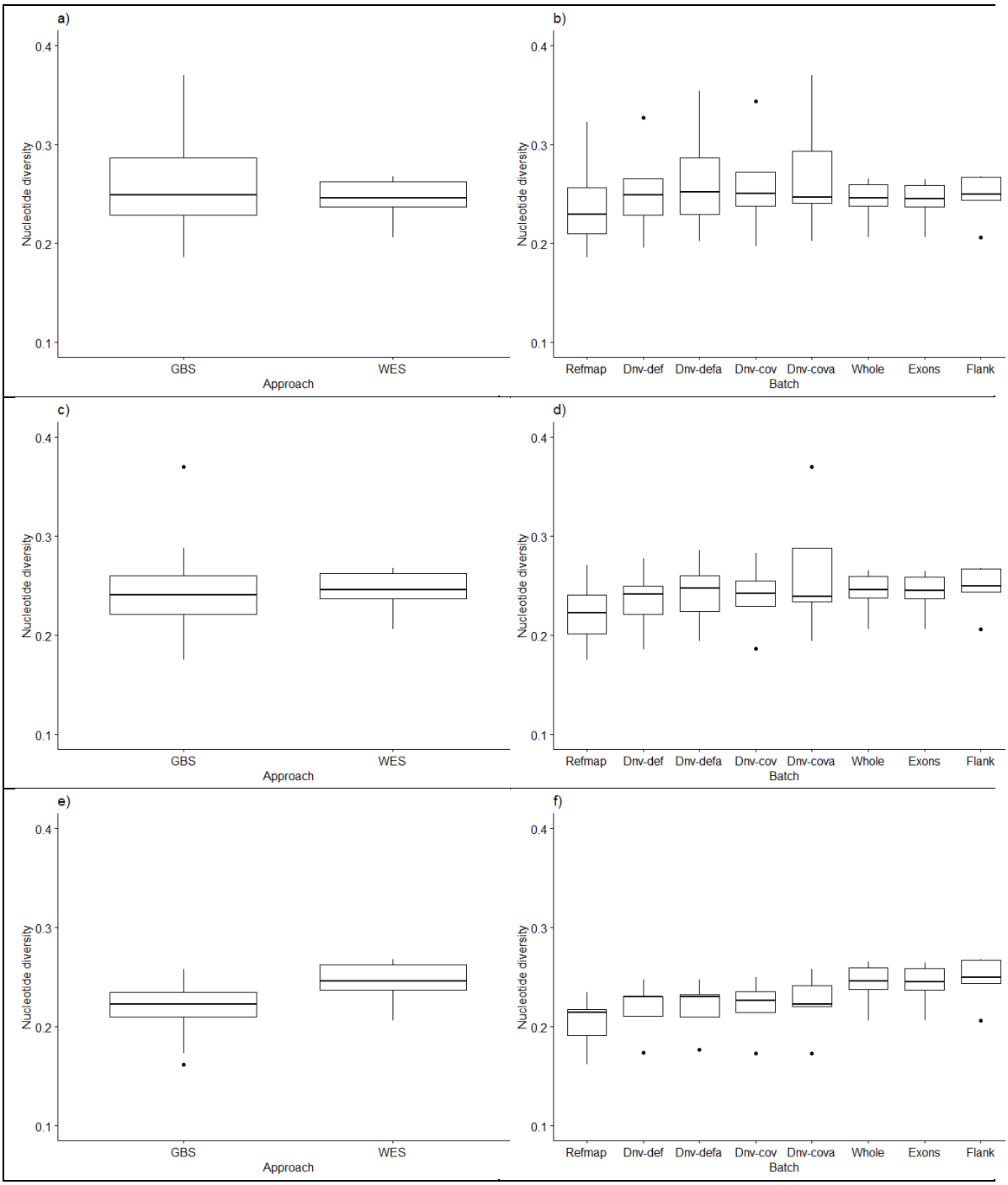


Fig 2b. Mean *joint* nucleotide diversity (*Pi*) calculated per batch and reduced representation approach, using *Pi* estimates of five jaguar populations as replicates for each batch. Rarified dataset (n=14): a-b) lax filter; c-d) intermediate filter; e-f) strict filter.

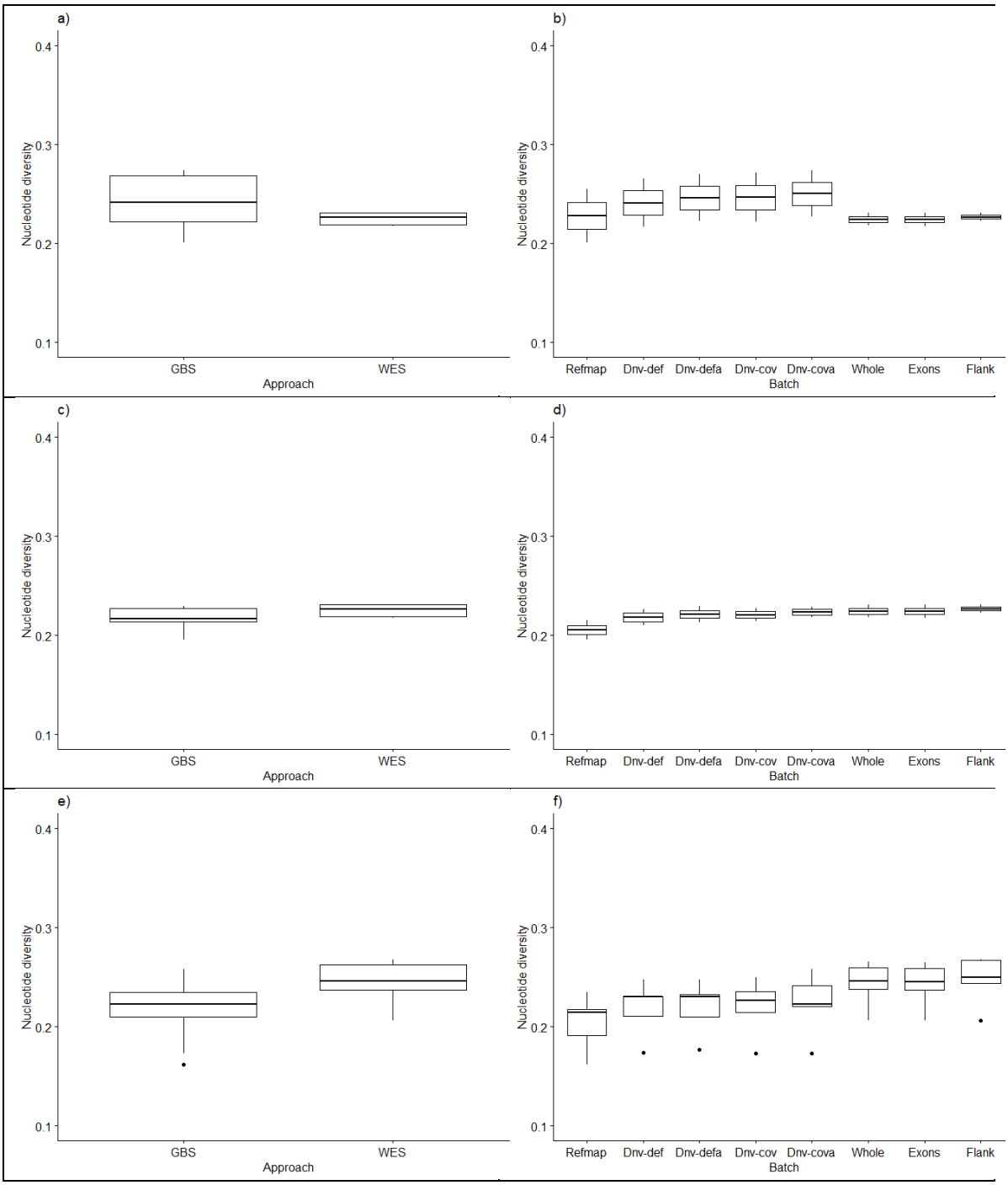


Fig 2c. Mean *joint* nucleotide diversity (*Pi*) calculated per batch and reduced representation approach*,* using *Pi* estimates of five jaguar populations as replicates for each batch. Amazon-Pantanal dataset (n=14): a-b) lax filter; c-d) intermediate filter; e-f) strict filter.

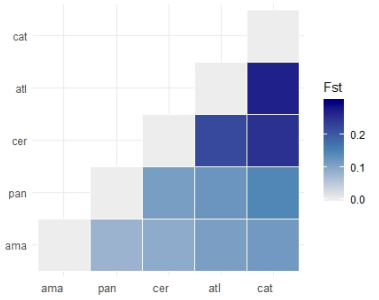
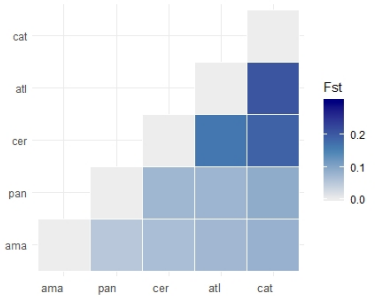
[insert: Dnv-Cov batch – lax filtering]	[insert: Flanking batch– lax filtering]
[insert: Dnv-Cov batch – intermediate filtering]	[insert: Flanking batch – intermediate filtering]
	

Figure 3. F_{st} heatmaps calculated from Full dataset. Top row: Middle. Bottom: full dataset, strict filtering. Left column: Dnv-Cov; right column: Flanking

SUPPLEMENTARY

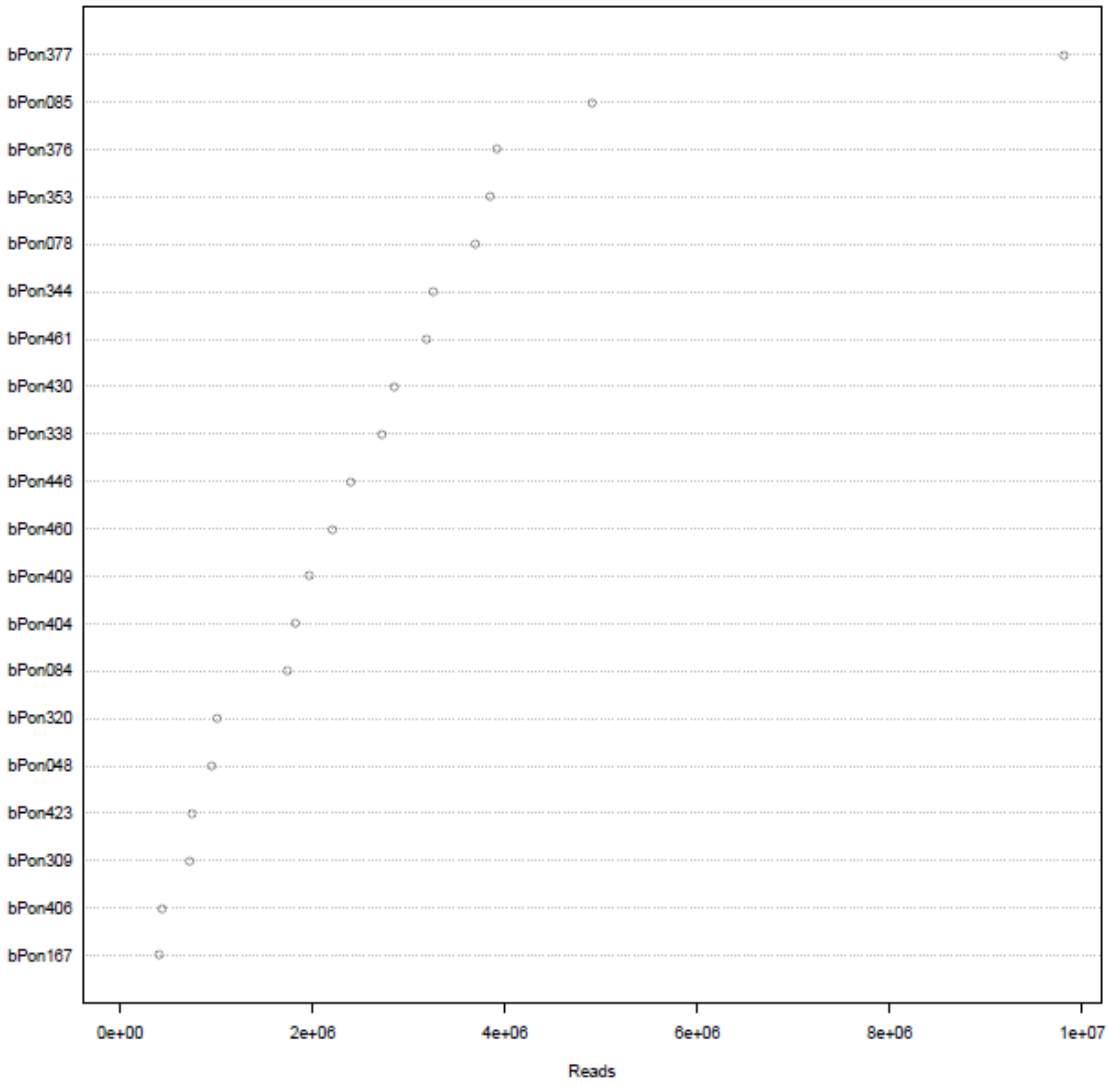


Figure S1. Retained reads per sample – GBS data.

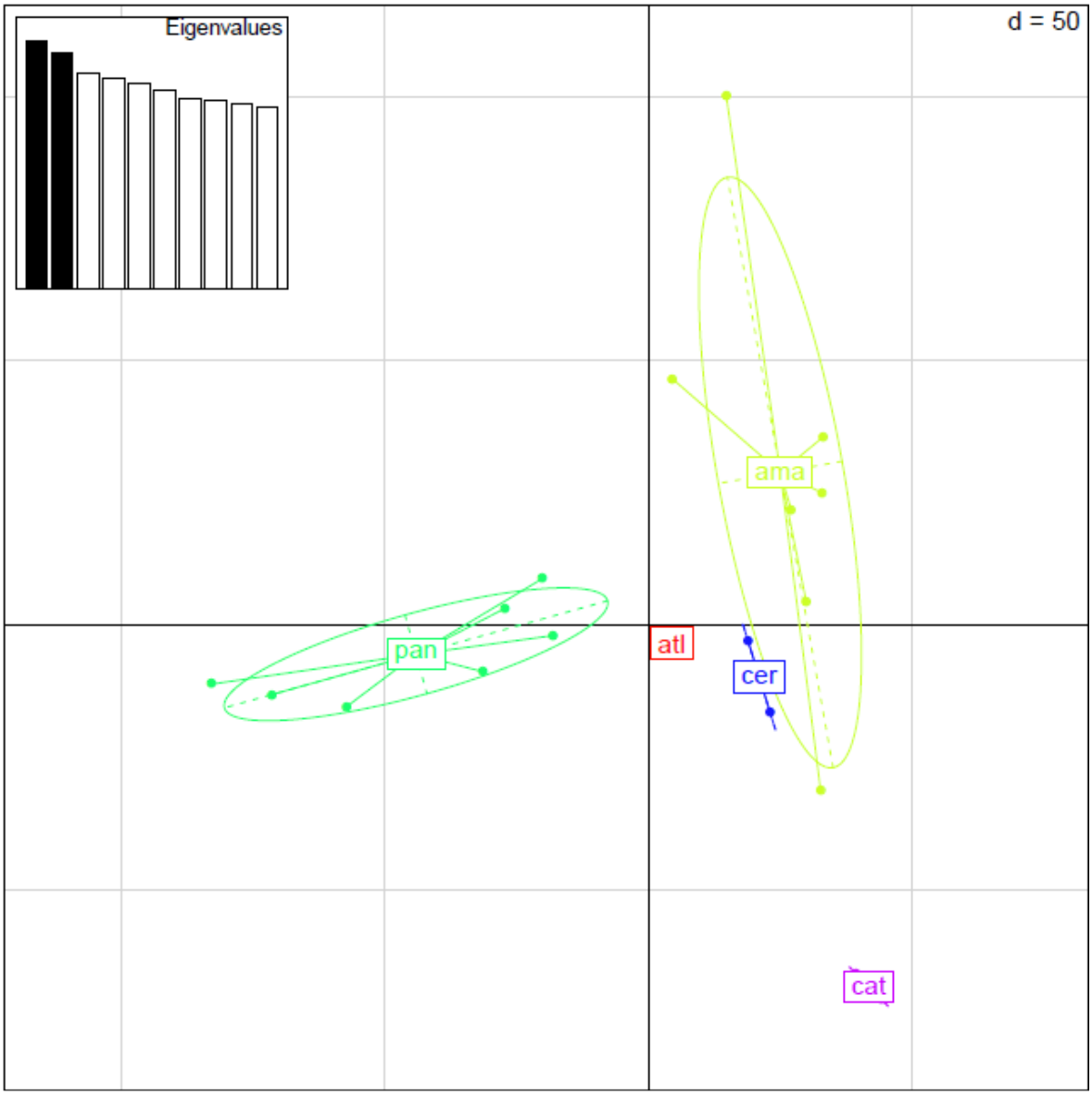


Figure S2. PCA 1x2 Flanking regions of the exome-derived data.


Ama
Pan
Cer
Atl
Cat

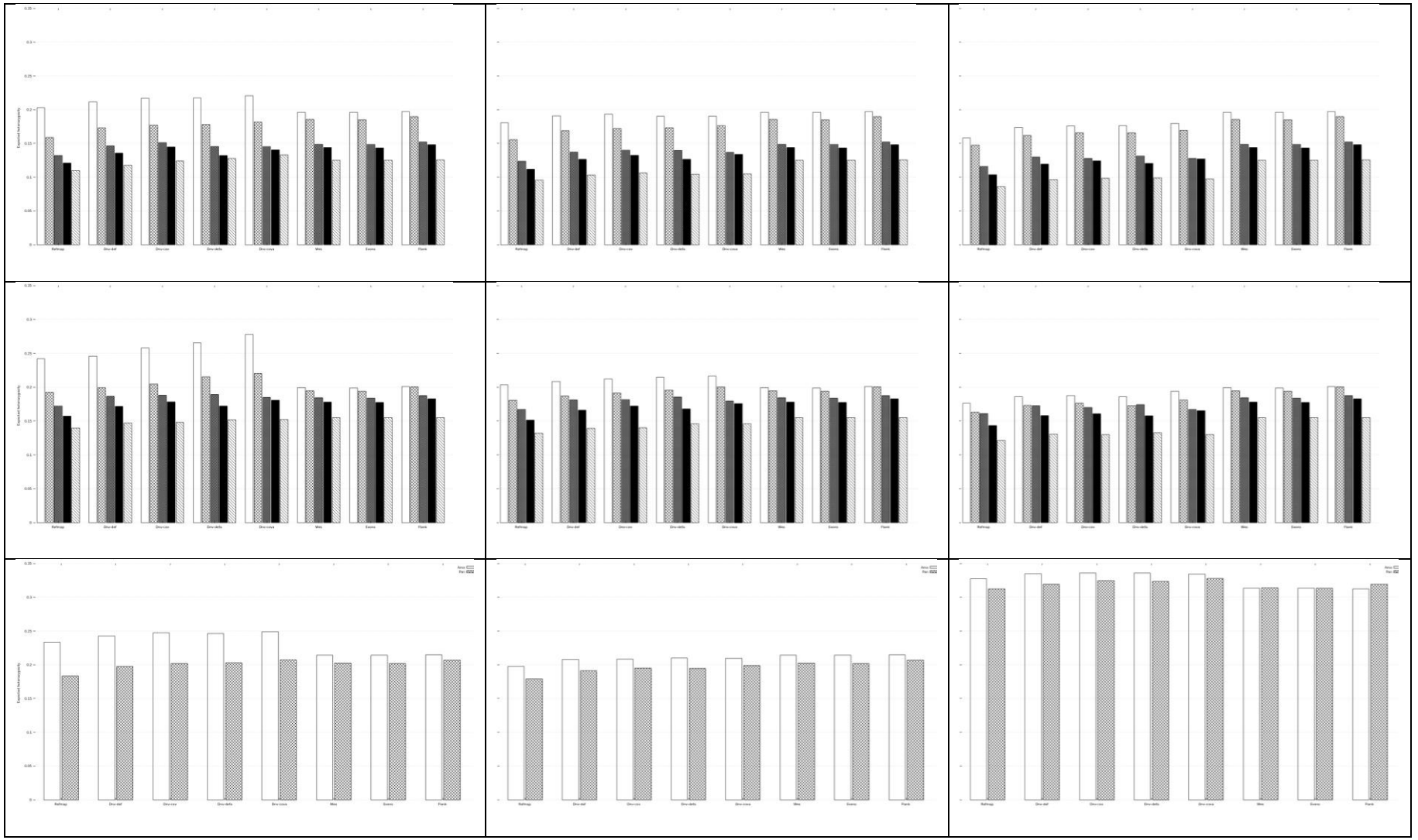


Figure S3. Gene diversity (expected heterozygosity). Top row: full dataset. Middle row: rarefied dataset. Bottom row: Amazon-Pantanal dataset. In all panels, from left to right: lax filtering, intermediate filtering, and strict filtering.

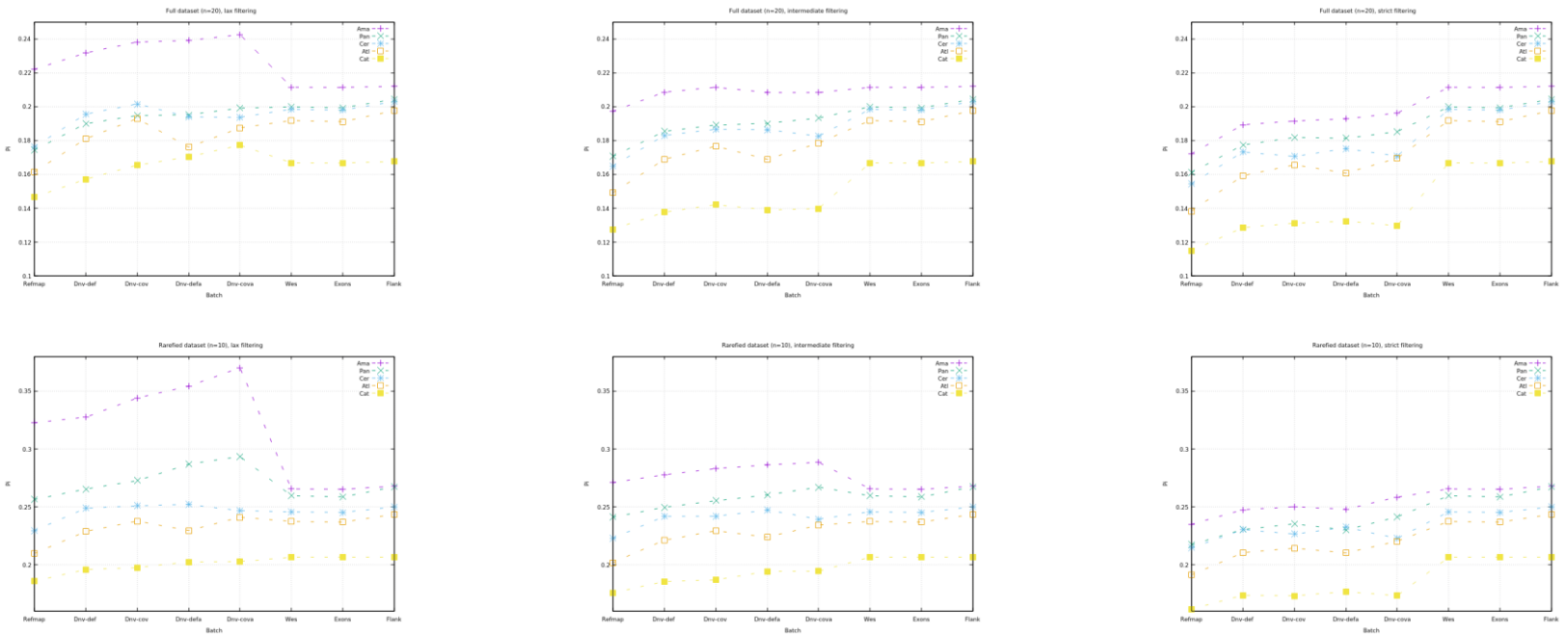


Figure S4. Nucleotide diversity (π) across five jaguar populations in Brazil estimated with five GBS and three exome sequencing SNP batches. Top row: full datasets (n=20). Bottom row: rarefied datasets (n=10). In all cases r=0.7 (minimum percentage of individuals in a population required to process a locus for that population). The parameter p, minimum number of populations a locus must be present in to process a locus, varied from p=1 in lax filtering (left column), to p=3 for intermediate filtering (center column) and, p=5 for strict filtering (right column).

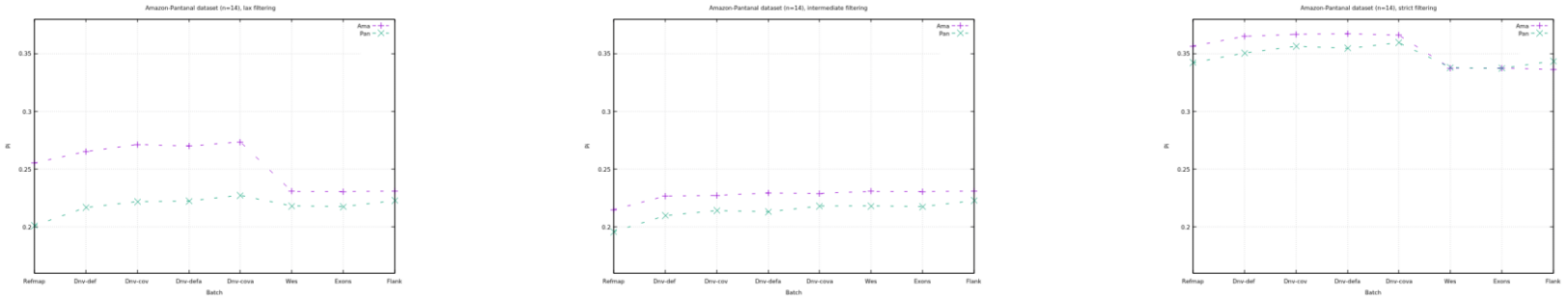


Figure S5. Nucleotide diversity (π) in Amazon and Pantanal jaguar populations (n=14), estimated with five GBS and three exome sequencing SNP batches. The r parameter (minimum percentage of individuals in a population required to process a locus for that population) was set to r=0.7. The parameter p, minimum number of populations a locus must be present in to process a locus, varied from p=1 in lax filtering (left column) to p=2 for both intermediate (center column) and strict filtering (right column). For strict filtering, we also required a minor allele frequency (MAF) of 0.1.

Table S1. Summary of genotyping statistics for eight batches in three datatasets

	Ref-map	Dnv-def	Dnv-cov	Dnv-defa	Dnv-cova	Wes	Exons	Flanking
				Full dataset (n=20), lax filtering				
Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273
Mean genotyped sites per locus	47.8	47.39	47.62	44.89	39.4	1	1	1
Kept loci after filtering	348870	248589	184340	171850	138581	115649	103376	12273
Total sites	32127413	22870188	16959280	15810200	12749452	115649	103376	12273
Genomic sites	15043419	--	--	--	--	115649	103376	12273
Variant sites	55544	46874	33267	30270	20590	92766	82911	9855
				Full dataset (n=20), intermediate filtering				
Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273
Mean genotyped sites per locus	92.2	92	92	92	92	1	1	1
Kept loci after filtering	103950	74195	50156	42227	27221	115649	103376	12273
Total sites	9584659	6825940	4614352	3884884	2504332	115649	103376	12273
Genomic sites	8469054	--	--	--	--	115649	103376	12273
Variant sites	40645	32453	21309	18789	11797	92766	82911	9855
				Full dataset (n=20), strict filtering				
Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273

Mean genotyped sites per locus	92.2	92	92	92	92	1	1	1
Kept loci after filtering	47297	22436	13034	13186	8266	115649	103376	12273
Total sites	4363397	2064112	1199128	1213112	760472	115649	103376	12273
Genomic sites	3890028	--	--	--	--	115649	103376	12273
Variant sites	17255	10601	6087	6194	3874	92766	82911	9855
				Rarefied dataset (n=10), lax filtering				
Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273
Mean genotyped sites per locus	47.9	57.61	61.49	68.97	67.78	1	1	1
Kept loci after filtering	338108	246675	185223	176895	144736	115649	103376	12273
Total sites	31135203	22694100	17040516	16274340	13315712	115649	103376	12273
Genomic sites	14808984	--	--	--	--	115649	103376	12273
Variant sites	49208	42310	32765	34441	26673	74889	66893	7996
				Rarefied dataset (n=10), intermediate filtering				
Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273
Mean genotyped sites per locus	92.2	92	92	92	92	1	1	1
Kept loci after filtering	100469	85095	62142	59800	41994	115649	103376	12273
Total sites	9263139	7828740	5717064	5501600	3863448	115649	103376	12273

Genomic sites	8317916	--	--	--	--	115649	103376	12273
Variant sites	35328	29154	20640	19807	13615	74889	66893	7996
				Rarefied dataset (n=10), strict filtering				
Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273
Mean genotyped sites per locus	92.2	92	92	92	92	1	1	1
Kept loci after filtering	27317	24312	15047	12737	8425	115649	103376	12273
Total sites	2520501	2236704	1384324	1171804	775100	115649	103376	12273
Genomic sites	2335488	--	--	--	--	115649	103376	12273
Variant sites	13980	9107	5649	4779	3224	74889	66893	7996
				Amazon-Pantanal dataset (n=14), lax filtering				
Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273
Mean genotyped sites per locus	34.3	40.43	39.98	38.49	32.99	1	1	1
Kept loci after filtering	342090	242586	179861	168195	135930	115649	103376	12273
Total sites	31502468	22317912	16547212	15473940	12505560	115649	103376	12273
Genomic sites	10498269	--	--	--	--	115649	103376	12273
Variant sites	45707	38543	27103	25271	16952	84929	75888	9041
				Amazon-Pantanal dataset (n=14), intermediate filtering				

Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273
Mean genotyped sites per locus	92.24	92	92	92	92	1	1	1
Kept loci after filtering	65333	53792	34488	33629	21440	115649	103376	12273
Total sites	6026086	4948864	3172896	3093868	1972480	115649	103376	12273
Genomic sites	5337966	--	--	--	--	115649	103376	12273
Variant sites	28663	22197	13915	13984	8650	84929	75888	9041
				Amazon-Pantanal dataset (n=14), strict filtering				
Genotyped loci	358215	268682	202249	192692	157234	115649	103376	12273
Mean genotyped sites per locus	92.23	92	92	92	92	1	1	1
Kept loci after filtering	53042	44811	28923	28009	17982	45225	40368	4857
Total sites	4892392	4122612	2660916	2576828	1654344	45225	40368	4857
Genomic sites	4414839	--	--	--	--			
Variant sites	11901	9965	6248	6341	3949	45225	40368	4857

1

CHAPTER IV Whole-genome sequences shed light onto demographic history and contemporaneous genetic erosion of natural jaguar populations

Whole-genome sequences shed light onto demographic history and contemporaneous genetic erosion of natural jaguar populations

Running title: Jaguar demographic history

Gustavo P. Lorenzana[1], Henrique V. Figueiró[1], Oliver A. Ryder[2], Eduardo Eizirik[1,3] *

[1] Laboratório de Biologia Genômica e Molecular, PUCRS, Porto Alegre, Brazil

[2] San Diego Zoo Institute for Conservation Research

[3] Instituto Pró-Carnívoros, Atibaia, Brazil

* Provisional author list (dissertation version).

Acknowledgments

GL, HVF and EE acknowledge CNPq/Brazil and CAPES/Brazil for financial support.

Abstract

Aim

Whole genome sequencing (WGS) constitutes the current state-of-the-art technique in many fields within the areas of genetics, evolution and molecular biology. The vast amount of data contained in a single genome represents an incredibly detailed record of past events in the lineage of its bearer, and may forecast the evolutionary potential of that particular taxon into the future. The jaguar (*Panthera onca*) is the apex predator of the Neotropics and its study and conservation are important to preserve vital ecosystem functions in the regions where it occurs. Here we employed jaguar WGS

data to infer its demographic history and assess signals of recent inbreeding in different portions of its geographic range.

Location

Amazon, Atlantic Forest, Cerrado, Caatinga and Pantanal biomes, South America; Southwestern USA/Northwestern Mexico, North America.

Methods

We sequenced eight novel genomes of jaguars collected in five biomes in South America, along with an additional genome from the species' northernmost population in Mexico/USA. We also incorporated recently sequenced genomes from two additional jaguars, creating a comparative dataset that represented multiple populations across the species' present-day range. Using these data, we modeled demographic history using the pairwise sequentially Markovian (PSMC) method, and assessed signals of inbreeding through the identification and characterization of genomic runs of homozygosity (ROH).

Results

PSMC plots indicated that jaguar lineages leading up to the analyzed populations had an effective population size (N_e) ranging from $4x10^4$ to $8.5x10^4$ *ca.* 1-2 Mya. Then they sharply declined to *ca*. $1.5x10^4$ around 0.5 Mya. A gradual expansion ensued, peaking at $2x10^4$ near 30,000 years ago, followed by a new reduction until 10,000 years ago, which was steeper for the Arizona individual. As for the contemporaneous analyses, we found a relatively small ROH burden across most jaguar populations, compared to other conservation-concern large carnivores. However, representatives from the Arizona and Atlantic Forest populations showed signals of recent bottlenecks and, in the latter case, inbreeding.

Main conclusions

Demographic history reconstruction is a key feature in the assessment of effective population size fluctuations over time. Our PSMC results were very consistent among individuals, and indicated that jaguar populations have undergone pronounced cycles of demographic fluctuations in the last 1-2 million years. In addition, the Arizona individual stood out in showing a steeper decline in the last 30,000 years, likely as a result of a recent history of founder events at the edge of the species' range. In their turn, ROHs constitute a component of genomic architecture with a great utility as an indicator of recent demographic history and adaptive potential of populations. Among South American jaguar populations, the Amazon appears as the largest population, with Pantanal, Cerrado and Caatinga sustaining smaller populations, while some individuals sampled in Atlantic Forest fragments showed signals of genomic erosion. In the northernmost extreme in the species' range, the Arizona individual genome bore the signature of a recently bottlenecked population, highlighting the need to urgently implement conservation measures to improve the chances of long-term survival of this felid in that area.

Keywords: WGS, PSMC, ROH, inbreeding

INTRODUCTION

Quaternary climatic oscillations shaped the patterns of divergence, distribution, abundance and population structure of many contemporary species. Jaguars diverged from other Eurasian *Panthera* species in the early Pliocene, approximately 3.7 million years ago (Mya), colonizing North America though the Bering strait by the end of that epoch, about 3 Mya (refs). They expanded their range southwards, across Central and South America, and by the end of the Pleistocene (13-11 thousand years ago [kya]), along with many other megafauna species, jaguars went extinct in the northern Hemisphere. The remaining populations thrived in strongholds located in the more stable equatorial ecosystems of South America (O'Brien & Johnson, 2007). Afterwards, a late range

77 expansion, advancing from the south, allowed jaguars to resettle Central America and subtropical
78 North America (including what is now northern Mexico and southwestern US), but the details of these
79 demographic processes and their genetic consequences have yet to be clarified.

80 Presently, the jaguar's core range is located in the Amazon basin. Sheer size and historical
81 climatic and geomorphological stability helped to preserve continuous lowland tropical forests
82 throughout glacial cycles (Colinvaux et al., 2000; Hoorn et al., 2010). Thus, it may be hypothesized that
83 this region constitutes the oldest and/or most genetically diverse population for several Neotropical
84 cat species, including modern jaguars. At the same time, the Amazon basin is surrounded by other
85 major biomes, such as the Pantanal, Cerrado, and Atlantic Forest, which may also have played
86 important roles in the maintenance of genetic diversity over time, since previous studies on jaguar
87 phylogeography using traditional markers have shown low population structure and substantial
88 amounts of gene flow on a broad spatial scale (Eizirik et al., 2001). Given that patterns of genetic
89 diversity are mainly determined by evolutionary history (Nadachowska-Brzyska et al., 2016),
90 estimating demographic trajectories of distinct demes, and their respective effective population sizes,
91 taking advantage of high-throughput sequencing and genome-wide markers, is now feasible and
92 advisable.

93 Whole-genome sequencing (WGS) allows researchers to tackle a myriad of questions with
94 unprecedented statistical power. For example, inference of demographic history using a single genome
95 per population is now possible using the pairwise sequentially Markovian coalescent (PSMC) approach
96 (Li & Durbin, 2011). This allows the estimate of effective population sizes through time, splitting times,
97 and periods of growth and decline for different demes (or taxonomic unit) in a straightforward routine.
98 In addition to its relevance in terms of reconstructing evolutionary history, such an assessment is also
99 relevant from a conservation perspective. Although the jaguar is globally considered to be less
100 threatened than other big cats, some of its populations are critically endangered, and may lose genetic
101 diversity very quickly and likely go extinct in the near future if the ongoing processes of habitat loss
102 and fragmentation remain unrestrained (De La Torre et al., 2018; Paviolo et al., 2016).

In addition, present-day estimates of genetic diversity are not always correlated with the conservation status of threatened species (Brüniche-Olsen et al., 2018; Díez del Molino, 2018), and some signals of genomic erosion may help to identify populations facing cryptic threats due to historical bottlenecks or persistently small population size. These warning signs are loss of genome-wide diversity, accumulation of runs of homozygosity (ROH), and alterations in genomic structural variants such as deletions or insertions (Díez del Molino, 2018). For the purpose of this study, we are mainly interested in ROH, as their identification represents a good trade-off between analytical requirements and information content, amid the relatively simple estimation of genome-wide heterozygosity and the complex identification of structural variants. Long ROH are caused when haplotypes inherited from each parent are identical (i.e. they are identical by descent, IBD), forming continuous chromosomal segments of homozygous genotypes (Ceballos et al, 2018). The number and length of ROH reflect individual demographic history, while the full homozygosity burden is useful to investigate recent episodes of drastic population reduction potentially leading to inbreeding depression.

In this context, our goal was to investigate the demographic history and to estimate both historical and contemporary effective population sizes of several jaguar subpopulations across the species' range (from southeastern Brazil to Arizona, USA) by employing PSMC and ROH-based methods. We aimed to identify long ROH that could entail an increased vulnerability of particular demes to reduction in fitness and ultimately to local extirpation. As a prediction, we expected the Amazon population to show less marked fluctuations in N_e over time, as well to as to harbor individuals with fewer and shorter ROH compared to demes located at the species' range edge or in areas subjected to heavy anthropic disturbance.

METHODS

DNA sample preparation and genome resequencing

Our data set included eight novel jaguar genomes sequenced specifically for this study, using DNA extracts that had been previously employed in a whole-exome sequencing project (Figueiró et al., in prep.). The newly sequenced genomes (Table 1) originated from all five major Brazilian biomes in which jaguars still occur: Atlantic Forest (n=3), Amazon (*n*=1), Pantanal (n=2), Cerrado (n=1) and Caatinga (n=1). In addition, our dataset also included two recently sequenced jaguar genomes (Santos et al., in prep.) that originated from different individuals sampled in the Amazon and Atlantic Forest. Finally, we also included a recently sequenced genome from a jaguar sampled in Arizona, USA, at the northernmost edge of the species' range. As a whole, our dataset comprised 11 jaguar genomes that represent much of the extant geographic distribution of this felid (Figure 1).

For the genomes sequenced in this study, libraries were constructed using the *TruSeq DNA PCR Free* or *TruSeq Nano DNA* kits (Illumina), with ~350 insert size. Sequencing was performed in an Illumina HiSeqX platform, with 2 or 3 samples multiplexed on the same lane, achieving ~20x coverage per genome. The genome from the Arizona individual was sequenced using the Discovar approach, as part of a large mammalian sequencing effort carried out by the Broad Institute, and attained 37x coverage.

Short read alignment and mapping; SNP calling and variant filtering

The jaguar assembly reference (Figueiró et al., 2017) was used to align the resequenced genomes by employing the Paleomix *bam pipeline* (Schubert et al., 2014). SNP calling was performed using Samtools *mpileup* according to the following parameters for base ("Q 20") and mapping quality ("C 50"). Unfiltered VCF files were parsed with the *SNPcleaner* Perl script from the ngsTools package (Fumagalli et al., 2014), using the following parameters: minimum and maximum site depth ("d 10", "D 100"), minimum number of 'covered' individuals ("k 1"), minimum read depth for an individual to be considered 'covered' ("u 10"), and process nonvariants ("v").

Downstream analyses

Inference of demographic history

We initially applied the Pairwise Sequentially Markovian Coalescent (PSMC; Li & Durbin 2011) method, which relies on the identification of hetero- and homozygous sites along the genome. Genomic regions with a high proportion of homozygous sites are indicative of recent coalescent times, while predominance of heterozygous sites reveal older coalescent events (Beichman et al., 2018). PSMC uses this information to recover trajectories in effective population size over broad temporal scales, displaying distinctive fluctuations of population growth and decline, that can be generalized into demographic syndromes such as expansion, contraction, and constant size. We used a realigned alignment file as input, and assumed a generation time of 5 years, and a mutation rate of 0.1e-8.

Runs of homozygosity (ROH)

For this analysis, chromosomal positions on resequenced genomes were reconstructed using the same reference jaguar assembly, which in this case was transposed to the chromosome-scale domestic cat assembly (*Felis_catus_9.0*, www.ncbi.nlm.nih.gov/genome/78?genome_assembly_id=356698) using Chromosomer (Tamazian et al., 2016). Each individual's VCF file was sorted and TAB-delimited indexed by genomic position (*tabix*) with VCFtools (Danecek et al., 2011), and then merged into a single VCF file containing the full set of variant sites per individual. Finally, that VCF file was converted into a genotypic matrix, recoded in *012* format, in which genotypes are represented as 0, 1 and 2, with the number representing the sum of non-reference alleles, thus specifying homozygous for alternative allele, heterozygous, and homozygous for reference allele, respectively.

Individual inbreeding was quantified with the statistic F_{ROH}, which is based on a likelihood ratio method that describes the logarithm of the odds (LOD), following the parameterization proposed by Kardos et al., (2018) and their custom R script. On each chromosome, a sliding window comprising 100 adjacent SNPs, a step size of 10 SNPs, and a minimum number of genotyped SNPs of 50, was used to calculate the IBD (identical by descent) LOD score. In order to avoid potential unfiltered genotyping

errors that could break long ROHs, we allowed the occurrence of 2% heterozygous SNPs in each IBD segment.

From this dataset, we characterized ROHs of different lengths as belonging to three categories: long ROH (>10Mb) derived from recent, closely related inbreeding; short ROH (<1Mb), indicative of more ancient inbreeding; and intermediate ROH (1-10Mb). Homozygous tracts were plotted using a custom script in R (Kardos et al., 2018). This helped us to calculate metrics used as a proxy to assess inbreeding: a) F_{ROH}, that describes the proportion of the genome that is IBD (completely homozygous; McQuillan et al., 2008); b) ROH burden, number and length of ROH (Ceballos et al., 2018); and c) ROH overlap, to identify those chromosomal regions showing long ROH in two or more individuals (Saremi et al., 2018).

RESULTS

We analyzed whole genome sequences of 11 jaguars sampled in six ecoregions across the species' range, and aligned them against the reference assembly reported by Figueiró et al. (2017), either directly (for PSMC) or after transposition to domestic cat chromosomal scale coordinates (for ROH analyses). PSMC analyses were run using diploid consensus sequences for each sample, encoded in a *fasta-like* format (*psmcfa*), containing at least one heterozygous site per bin. Runs of homozygosity were detected in each individual, after parsing 8.5×10^6 biallelic SNPs identified on 18 autosomes across the 11 genomes.

Ancient demographic history – PSMC

Estimates of past fluctuations in effective population size (N_e) obtained with the pairwise sequentially Markovian coalescent method (PSMC) are shown in Fig. 1. Inferred plots for all individuals begin 1 – 2 Mya, with initial N_e values ranging from of 4×10^4 to 8.5×10^4. Then all plots show a sharp decline in N_e, plummeting to *ca.* 1.5×10^4 *ca.* 0.5 Mya. Subsequently a gradual population expansion

ensued, peaking at *ca.* $2x10^4$ around 30,000 years ago, followed by a new round of that affected all individuals. Interestingly, the trajectories of all South American individuals remain similar after this point, reaching a N_e of *ca.* $1x10^4$ at the end of the plot, 10,000 years ago, while the Arizona individual shows a distinct path, with a steeper decline in this final phase, ending with a N_e of $3x10^3$.

Recent demographic history – ROH

We identified 8383 runs of homozygosity (ROH) longer than 10 Kb, adding up to 3.17 Gb across the 11 individual genomes. The Arizona jaguar showed the heaviest burden, with 722 Mb distributed in 2219 ROH, indicating that 30% of its genome is homozygous (Table S1). On the other hand, AF017 showed the least burden, with less than 2% of its genome in IBD tracts.

As for long (>1 Mb) ROH, we detected 537 segments totaling 793.9 Mb. Again, the Arizona jaguar showed the heaviest ROH burden, with 173.1 Mb in 111 homozygous segments, including the single ROH detected in the 'very long' size class (with 5.65 Mb in length), and another segment spanning 4.99 Mb. Individuals AF017 and PA462 exhibited the lowest sum total length of ROH (SROH), with 6.8 and 12 Mb distributed in 4 and 8 segments, respectively (Table S1, Fig. 2). On average, 3% of the 11 genomes appeared to comprise long ROH, with a range between 0.3% and 7% (Table S2).

Fig. 2 summarizes the mean distributions of the three different ROH size classes in each sampled biome. Long ROH spanned typically 1-2 Mb for all populations, with no significant differences among populations (Mann-Whitney rank sum test: χ^2=7.6, df=5, p=0.18). Considering intermediate and long ROH jointly, mean ROH length decreased to around 0.2-0.5 Mb, resulting in statistically significant differences among populations (χ^2=170.5, df=5, p=2.2e-16). Then we applied pairwise comparisons using the Wilcoxon rank sum test, which showed that Arizona was different from the rest of the biomes (Table S3). Lastly, considering all three ROH size classes jointly yielded a slight reduction in means relative to the previous batch, denoting a very small contribution of short ROH to SROH, with significant differences among populations (χ^2=142.6, df=5, p=2.2e-16). Again, Arizona was different from the rest,

except from the Amazon population. The Atlantic Forest showed significant differences from Amazon and Caatinga, while the Pantnal was different from the Amazon (Table S3).

On an individual basis, the mean for most samples was 0.2-0.5 and 1.0-1-5 Mb for short and long ROH, respectively (Fig. 3). Thus, most of the IBD tracts were located on the lower end of the expected spectrum of their respective categories. The pattern for short ROH is consistent among individuals, indicating similar signals of ancient inbreeding. Regarding long ROH, we detected four major cases depicted in Fig. 3: (a) Higher enrichment around the most frequent mean size across individuals (AM378). b) Spread density around a higher mean (AM017 and PA462). c) Higher density around the most frequent mean size with some enrichment for long ROH (AF048, AM404, CE411). d) Concentration around the mean with moderate enrichment for long ROH (AF052, AF395, PA342, CA460, jagAriz).

DISCUSSION

Ancient demographic history

The PSMC analysis led to interesting insights regarding jaguar demographic history. An initial observation was that patterns were remarkably consistent among individuals for most of the reconstructed period, supporting the robustness of the estimates and indicating that all sampled individuals share the same population history throughout most of the retrieved time. This shared history corroborates inferences from previous studies (e.g. Eizirik et al. 2001) that jaguar populations have remained highly connected across the species' range until recent times. An interesting departure from this shared pattern was observed in the Arizona individual in the most recent phase of the assessed history (< 30,000 years ago), when this lineage exhibited a steeper decline in effective size relative to the South American individuals, reaching a 3-fold lower N_e than the others by 10,000 years ago. This pattern may result from a more recent process of colonization of this region at the northern end of the species' current range, driven by an expansion of South American lineages, as suggested previously (Eizirik et al. 2001). Such a scenario implies extinction of pre-existing North American jaguar

lineages and recolonization from South America, in a process that may have included a succession of founder events. This may have been exacerbated by the lower productivity of habitats in this area, relative to moister environments farther south, which may have contributed to keep jaguar densities and effective sizes at lower levels in this area. Such hypotheses can be further tested by expanding the geographic sampling of assessed genomes, and especially by including representatives of Central American and Mexican jaguar populations.

When assessing the PSMC trajectories of the full suite of analyzed jaguars, another interesting point emerges. As we are dealing with a lineage occurring predominantly in South America, where climatic oscillations during the Quaternary (Pliocene and Pleistocene) were not as drastic as in the northern hemisphere (Hoorn et al., 2010), one could expect more stable effective populations sizes (N_e) on a broad temporal scale. However, that is not the case as jaguar PSMC trajectories keep a good resemblance with patterns inferred for boreal species such as American black bears (Miller et al., 2012; Kumar et al., 2017) and Eurasian lynx (Abascal et al., 2016). One possible explanation for this match is that in the southern hemisphere, global climatic oscillations, especially changes in precipitation patterns due to sea level fluctuation, caused cycles of increased aridity (Werneck, 2011), that could have affected austral species in an analogous fashion to glacial cycles. Currently, the lack of PSMC models for other Neotropical mammal species hinders comparison, but the lowest N_e is attained around 10-12 Kya, coinciding with the Last Maximum Glacial. The PSMC method has poor resolution for periods near the present, due to the reduced number of coalescent events that have occurred at more recent times (Ellegren, 2014). Thus, to assess contemporary demographic processes, other methods can be employed, such as the characterization of individual patterns of genome-wide variability, as assessed the presence of runs of homozygosity (ROHs) (Díez del Molino et al., 2018).

Recent demographic history

Our results suggest that currently, most of the South American jaguar populations are relatively large in comparison with other large mammalian carnivores, but even so they are very

283 sensitive to rapid genomic erosion triggered by anthropic-driven drift and inbreeding. The longest ROH
284 we identified across the sampled jaguars reached 5.65 Mb in length, while a single ROH of a severely
285 inbred island wolf showed a single ROH measuring 95.8 Mb in length (Kardos et al., 2018). This ROH
286 size represents more than half of the total long-ROH burden in the most heavily ROH-burdened
287 jaguars, such as AF395, AF052 and jagAriz (Table S2). Similarly, Mcquillan et al., (2008) detected a few
288 long ROHs, adding up to 100 Mb, in inbred European human populations.

289 Very long ROH have been documented as well in the cheetah (Dobrynin et al. 2015), Iberian
290 lynx (Abascal et al., 2016), and Florida panthers (Saremi et al., 2018). We did not detect in jaguars any
291 very long ROH (>10 Mb), such as those observed in such highly inbred species. Actually, up to 93% of
292 cheetahs' genomes, 58% in Florida panthers' and 40% in Iberian lynx resulted to be homozygous. The
293 closest instance in our dataset is given by the Arizona individual, which displayed a 30% genome-wide
294 homozygosity.

295 Our intra-population sampling is small, but the ample geographic coverage of the samples
296 allowed us to describe recent demographic patterns among populations. The sum total length of ROHs
297 plotted against the total number of ROHs (Fig. 4), allowed us to identify some demographic processes
298 that can be further tested by increasing the sample size per population. On the lower side of the
299 spectrum lie outbred populations, with fewer and shorter ROHs (i.e. low NROH and low SROH). At the
300 other extreme are the inbred and/or bottlenecked populations, with more numerous and longer ROHs
301 (i.e. higher NROH and SROH). Accordingly, the Amazon, North Pantanal and Cerrado biomes sustain
302 relatively large jaguar populations, and they showed low NROH and SROH. The individual from the
303 southern Pantanal (PA342), showed signals of inbreeding that need to be further tested by
304 resequencing additional individuals from that area. The individual AF017, sampled in a transition area
305 between the Atlantic Forest and the Cerrado and Pantanal biomes (and whose jaguar population has
306 subsequently been extirpated – see Haag et al. [2010]) seems to be an admixed individual that
307 harbored signatures of connectivity with adjacent habitats, prior to anthropic-driven population
308 bottlenecks that seem to characterize remaining Atlantic Forest fragments. The Arizona jaguar

exhibited strong signatures of a bottlenecked population, likely influenced by its location at the northern edge of the species' range, at suboptimal conditions due to habitat productivity and prey biomass, and perhaps also affected by pressure from human sources of disturbance.

The major source of concern arising from our results pertains to jaguars AF395 and AF052, which bear signs of a bottlenecked and inbred population. The former individual, coming from the Atlantic Forest Green Corridor, showed worrysome signals of inbreeding, even more severe than those observed for other Atlantic Forest individuals coming from populations that previously had been identified as having comparatively stronger signals of genetic drift due to habitat loss and fragmentation (Haag et al. 2010). This can be attributed to the fact of that individual sample was the most recently collected among the analyzed individuals for this biome. This suggests that increased inbreeding has occurred very fast, after recent massive deforestation in that biome, leaving forest fragments with only a few related individuals. This result matches the scenario provided by Zanin et al., (2015), under which this population is highly vulnerable to extinction, with an estimated mean time to extinction of less than 60 years, which is mainly attributed to the low population density in that region. The genome-wide results presented here add to the sense of urgency of restoring connectivity among remaining jaguar populations in Atlantic Forest fragments (as highlighted by Srbek-Araújo., 2018), so as to rescue their genetic diversity and evolutionary potential, likely already impacted by loss of adaptive alleles and inbreeding.

Further resequencing of more individuals across the jaguar's range, considering a temporal sampling, i.e. sub-setting samples by collection date, at least for the last few decades, would increase the cloud of points along the regression curve of ROH/SROH, which could improve the resolution on the temporal dynamics of the ROH burden, and further clarify the recent demographic history of jaguars. Overall, these results illustrate the power of genome-wide analyses to perform in-depth assessments of demographic history, illuminating not only evolutionary trajectories leading to present-day populations, but also providing critical information that can be incorporated into conservation strategies on their behalf.

335 REFERENCES

336 Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J. L., Vlasova, A., Marcet-Houben, M., Martínez-
337 Cruz, B., Cheng, J. C.,… & Godoy, J. A. (2016). Extreme genomic erosion after recurrent
338 demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, 1, 1–19.
339 Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using genomic data to infer historic
340 population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and*
341 *Systematics*, 49, 433–456.
342 Brüniche-Olsen, A., Kellner, K. F., Anderson, C. J., & DeWoody, J. A. (2018). Runs of homozygosity have
343 utility in mammalian conservation and evolutionary studies. *Conservation Genetics*, 19, 1295–
344 1307.
345 Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity:
346 windows into population history and trait architecture. *Nature Reviews Genetics*, 19, 220–234.
347 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter,
348 G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group
349 (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.
350 De la Torre, J. A., González-Maya, J. F., Zarza, H., Ceballos, G., & Medellín, R. A. (2018). The jaguar's
351 spots are darker than they appear: assessing the global conservation status of the jaguar
352 *Panthera onca*. *Oryx*, 52, 300–315.
353 Dobrynin, P., Liu, S., Tamazian, G., Xiong, Z., Yurchenko, A. A., Krasheninnikova, K., ... & Kuderna, L. F.
354 (2015). Genomic legacy of the African cheetah, Acinonyx jubatus. *Genome Biology*, 16, 277.
355 Díez del Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2018). Quantifying
356 temporal genomic erosion in endangered species. *Trends in Ecology & Evolution*, 33, 176–185.
357 Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in*
358 *Ecology & Evolution*, 29, 51–63.
359 Fumagalli, M., Vieira, F. G., Linderoth, T., & Nielsen, R. (2014). ngsTools: methods for population
360 genetics analyses from next-generation sequencing data. *Bioinformatics*, 30, 1486–1487.

Hoorn, C., Wesselingh, F. P., Ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C. L., Figuereido, J. P., Jaramillo, C., Riff, D., Negri, F. R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T. & Antonelli, A. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, 330, 927–931.

Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, 9, 1205–1218.

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., 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.

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, 46487.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475, 493–496.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F., Campbell, H., & Wilson, J. F. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, 83, 359–372.

Miller, W., Schuster, S. C., Welch, A. J., Ratan, A., Bedoya-Reina, O. C., Zhao, F., ... & Lindqvistc, C. (2012). Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences*, 109, E2382–E2390.

Paviolo, A., De Angelo, C., Ferraz, K. M. P. M. B., Morato, R. G., Martinez Pardo, J., Srbek-Araujo, A. C., Beisiegel, B. de M., Lima, F., Sana, D., Xavier da Silva, M., Velázquez, M. C., Cullen, L., Crawshaw Jr., P., Jorge, M. L. S. P., Galetti, P. M., Di Bitetti, M. S., de Paula, R. C., Eizirik, E., Aide, T. M.,

Cruz, P., Perilli, M. L. L., Souza, A. S. M. C, Quiroga, V., Nakano, E., Ramírez Pinto, F., Fernández, S., Costa, S., Moraes Jr., E. A., & Azevedo, F. (2016). A biodiversity hotspot losing its top predator: The challenge of jaguar conservation in the Atlantic Forest of South America. *Scientific Reports*, 6, 37147.

Saremi, N. F., Supple, M. A., Byrne, A., Cahill, J. A., Coutinho, L. L., Dalen, L., ... & Shapiro, B. (2018). Mountain lion genomes provide insights into genetic rescue of inbred populations. *bioRxiv*, 482315. doi: http://dx.doi.org/10.1101/482315.

Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E. & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, 9, 1056–1082.

Tamazian, G., Dobrynin, P., Krasheninnikova, K., Komissarov, A., Koepfli, K. P., & O'Brien, S. J. (2016). Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences. *GigaScience*, 5, 38.

Werneck, F. P. (2011). The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. *Quaternary Science Reviews*, 30, 1630–1648.

Zanin, M., Palomares, F., & Brito, D. (2015). The jaguar's patches: viability of jaguar populations in fragmented landscapes. *Journal for Nature Conservation*, *23*, 90–97.

TABLES

413

414 FIGURES

415

416

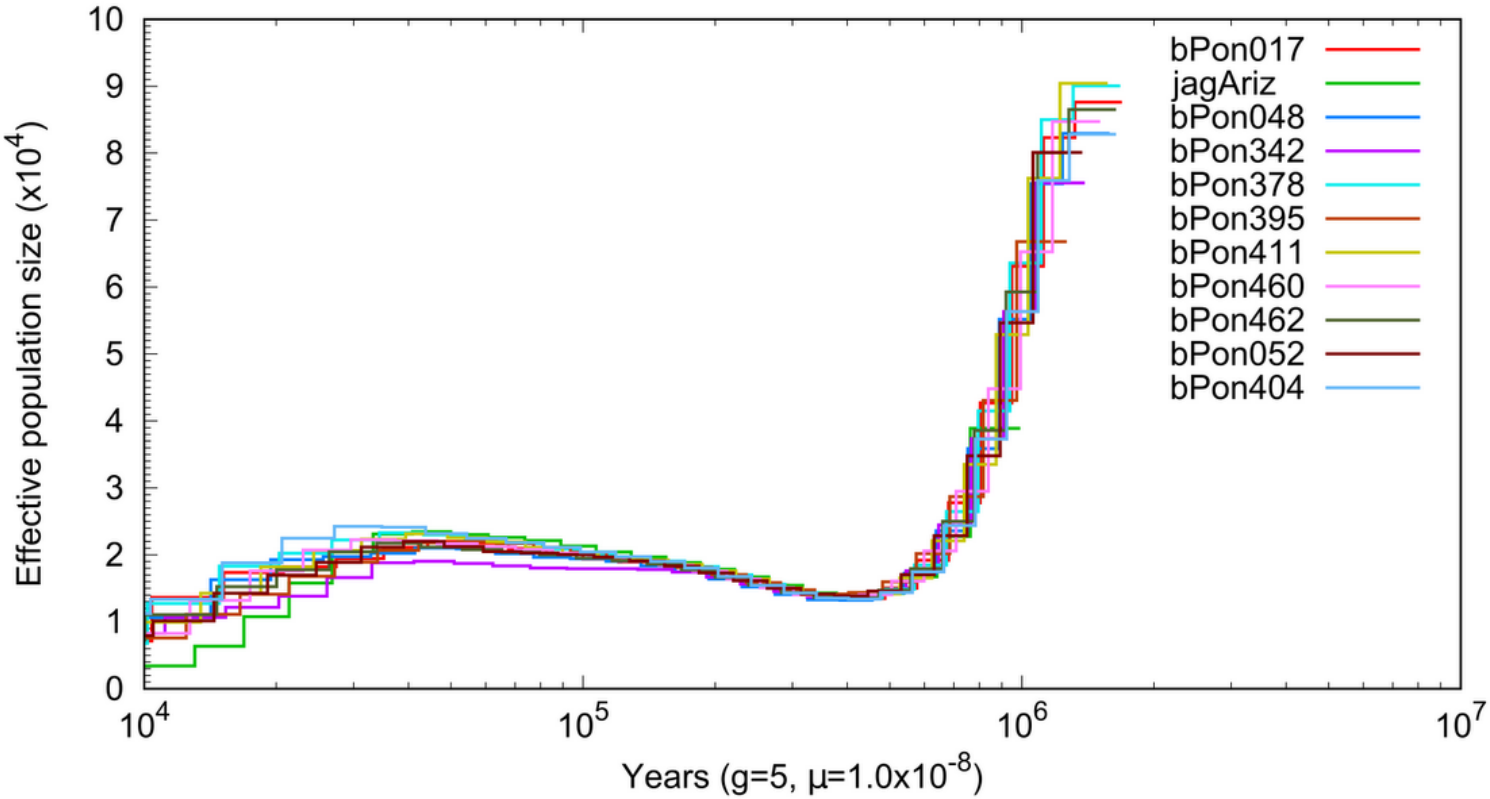
417

418

419

420

421



422

423 Figure 1. Reconstruction of historical fluctuations in effective population size (N_e) based on the genome-wide distribution of heterozygous sites, using the
424 Pairwise Sequential Markovian coalescent (PSMC) applied to newly sequenced jaguar genomes.
425

426

427

428

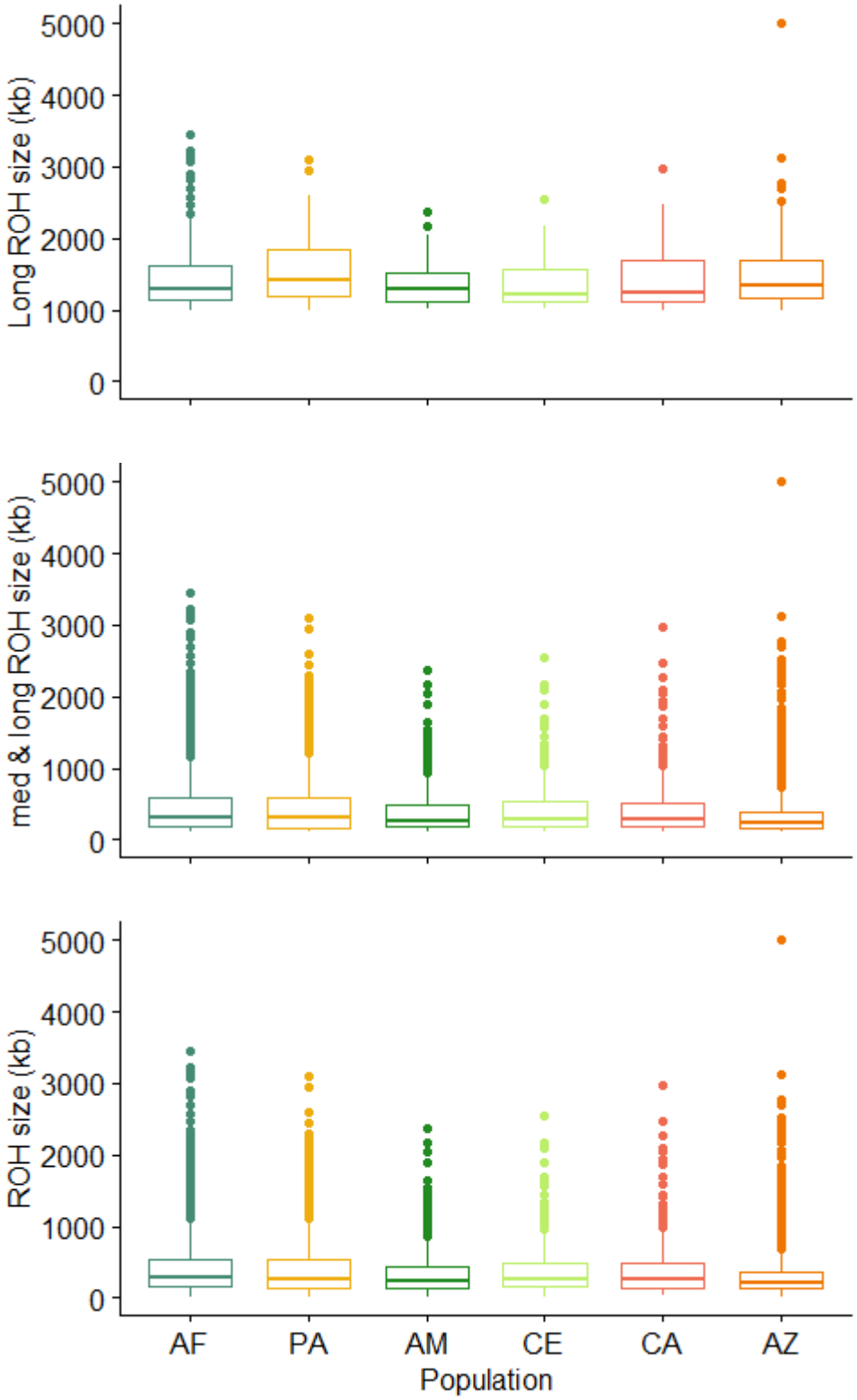


Figure 2. Box plots of the mean size for ROHs in three different size classes: long (top panel), intermediate plus long (middle panel), and all segments (bottom panel). ROHs were searched for in the 18 autosomes of each individual, considering homozygous tracts longer than 10 kb, and then grouped per population. For graphical purposes, the point corresponding to the single very long ROH (>5 *Mb*), detected in the AZ jaguar genome, is not shown. AF: Atlantic Forest; PA: Pantanal; AM: Amazon; CE: Cerrado; CA: Caatinga; AZ: Arizona.

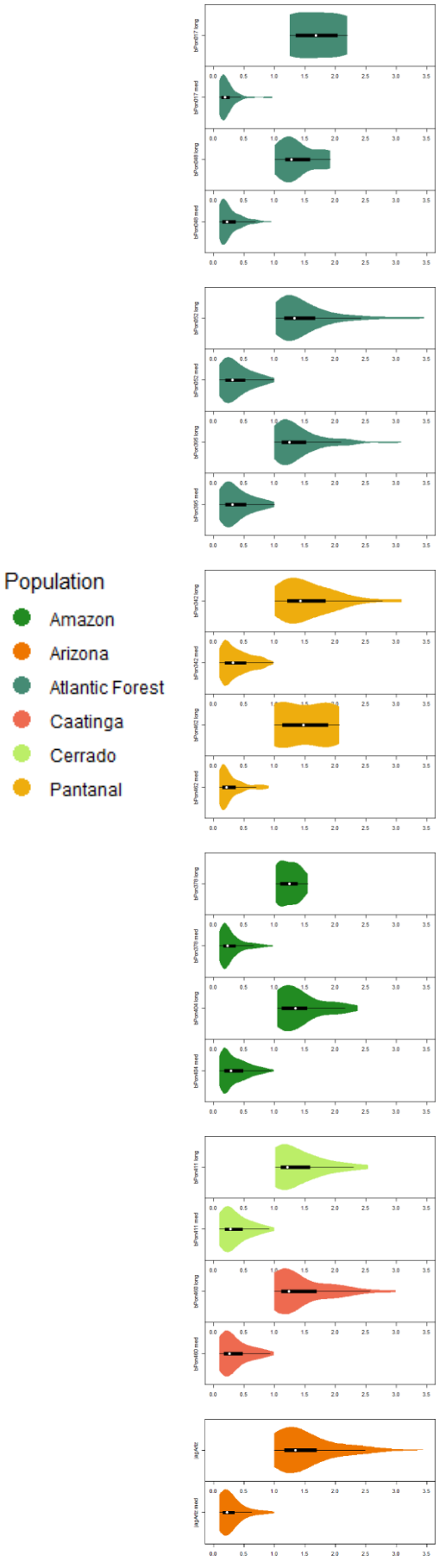

Population
Amazon
Arizona
Atlantic Forest
Caatinga
Cerrado
Pantanal

Figure 3. Violin plots of the mean size for long and intermediate ROHs, calculated for the 18 autosomes, considering homozygous tracts longer than 10 kb. Individuals are grouped by their population of origin in South America (from top to bottom: Atlantic Forest, Pantanal, Amazon, Cerrado and Caatinga), and North America (Sonora-Arizona).

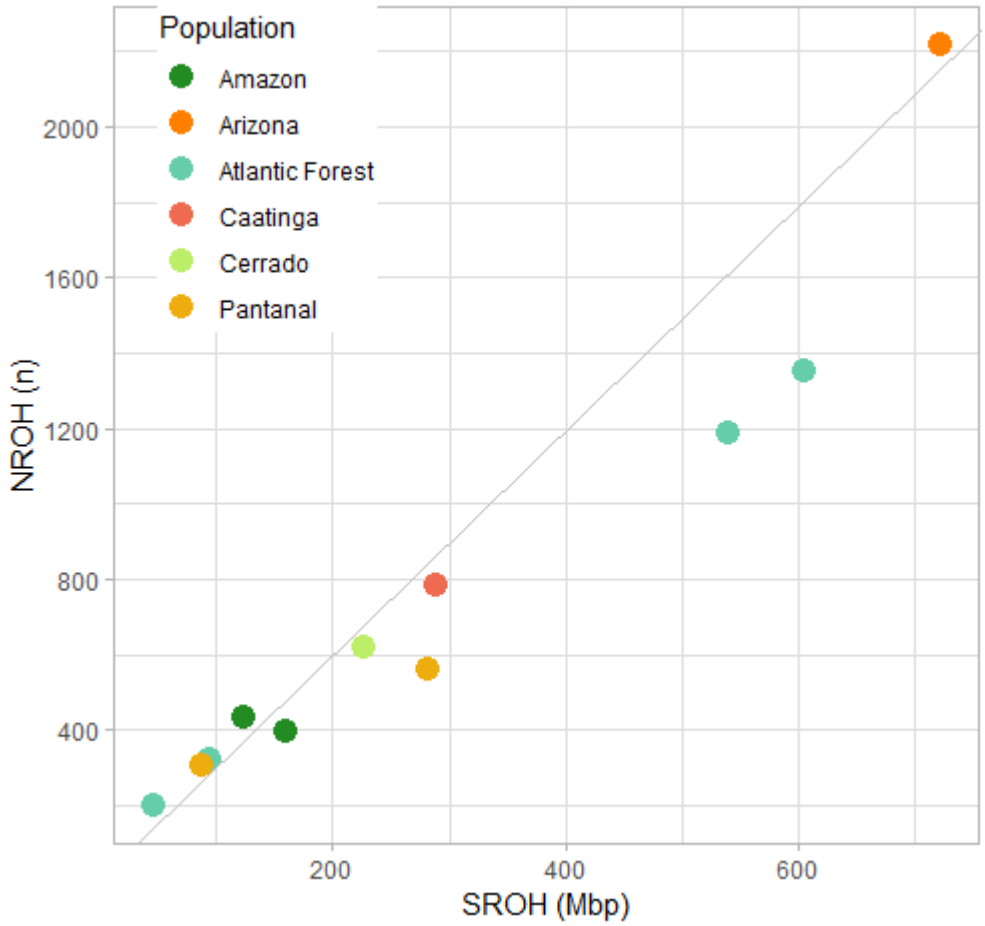


Figure 4. Total number of ROH (*NROH*) compared to the sum total length of ROH (*SROH*) in 11 jaguar genomes. The recent demographic history of each population can be inferred by the relative position of each individual along the plot (*cf*. Fig. 1 in Ceballos et al., 2018). The burden of ROH is related to population size, with smaller populations harboring more and longer ROH than larger populations. The Arizona jaguar position in the upper right corner is characteristic of a bottlenecked population. Atlantic Forest AF395 and AF052 individuals fit the pattern expected for a bottlenecked and inbred population. North Pantanal (PA462) and Amazon sustain larger populations than the Cerrado, Caatinga and South Pantanal (PA342), with the latter individual showing signals of inbreeding. AF017, in the lower left corner of the plot, seems to be an admixed individual.

SUPPLEMENTARY

Table S1. Summary of ROHs per individual sample (generated with Plink).

Population	Individual	Short ROH		Medium ROH		Long ROH		Total		
		0.01 - 0.1 Mb		0.1 - 1 Mb		> 1 Mb				
		n	length	n	length	n	length	n	length	% *
Atlantic Forest	bPon017	48	3.49	148	35.01	4	6.81	200	45.31	1.89
Atlantic Forest	bPon048	56	4.01	249	69.00	15	20.82	320	93.83	3.91
Atlantic Forest	bPon052	119	8.91	955	358.97	115	172.31	1189	540.18	22.51
Atlantic Forest	bPon395	77	5.37	1163	443.79	112	156.66	1352	605.82	25.24
Pantanal	bPon342	39	2.71	452	173.46	68	105.99	559	282.15	11.76
Pantanal	bPon462	48	3.24	252	71.60	8	11.97	308	86.82	3.62
Amazon	bPon378	67	4.90	356	105.58	11	13.69	434	124.17	5.17
Amazon	bPon404	45	2.99	322	113.28	31	44.10	398	160.37	6.68
Cerrado	bPon411	76	5.74	520	185.47	25	35.14	621	226.35	9.43
Caatinga	bPon460	84	6.23	662	228.70	37	53.26	783	288.18	12.01
Arizona	jagAriz	248	19.07	1860	529.82	111	173.12	2219	722.00	30.08
		907	66.64	6939	2314.67	537	793.86	8383	3175.18	

*Sum total length of ROH/Genome size.

Short: 10-100 kb; Medium: 100-1000 kb; Long: >1000 kb

Table S2. Summary of ROHs >1Mb per individual (generated with Plink).

Population	Individual	NROH	SROH		ROH (Mb)		
		n	Mb	% *	$\bar{x}$	min	max
Atlantic Forest	bPon017	4	6.8	0.28	1.70	1.25	2.19
Atlantic Forest	bPon048	15	20.8	0.87	1.39	1.01	1.92
Atlantic Forest	bPon052	115	172.3	7.18	1.50	1.02	3.45
Atlantic Forest	bPon395	112	156.7	6.53	1.40	1.00	3.08
Pantanal	bPon342	68	106.0	4.42	1.56	1.01	3.09
Pantanal	bPon462	8	12.0	0.50	1.50	1.00	2.06
Amazon	bPon378	11	13.7	0.57	1.24	1.03	1.55
Amazon	bPon404	31	44.1	1.84	1.42	1.05	2.37
Cerrado	bPon411	25	35.1	1.46	1.41	1.02	2.54
Caatinga	bPon460	37	53.3	2.22	1.44	1.00	2.99
Arizona	jagAriz	111	173.1	7.21	1.56	1.00	5.65
Total		537	793.9				

*SROH/Genome size

Table S3. Pairwise Wilcoxon test results for the three ROH size classes

	Pop	AF	PA	AM	CE	CA
	PA	1	-	-	-	-
Wilcoxon	AM	2.90E-07	0.0087	-	-	-
adj/Bonferroni	CE	0.254	1	0.3167	-	-
All ROH	CA	0.0327	1	0.6006	1	-
	AZ	2.00E-16	2.00E-09	0.2931	1.30E-05	2.30E-05

		AF	PA	AM	CE	CA
	PA	1	-	-	-	-
Wilcoxon	AM	3.10E-05	0.10327	-	-	-
adj/Bonferroni	CE	1	1	0.35655	-	-
Med & Long	CA	0.02436	1	1	1	-
	AZ	2.00E-16	1.40E-11	0.00048	1.70E-09	4.00E-07

		AF	PA	AM	CE	CA
	PA	0.55	-	-	-	-
Wilcoxon	AM	1	0.77	-	-	-
adj/Bonferroni	CE	1	1	1	-	-
Long ROH	CA	1	1	1	1	-
	AZ	1	1	1	1	1

CHAPTER V – GENERAL DISCUSSION

Global biodiversity is increasingly threatened by human activities. The severity of the ongoing crisis is of such magnitude that it is now considered the sixth episode of mass extinction in the planet's known history (Ceballos et al., 2015), even warranting a change in geological time nomenclature, as were leaving the Holocene, entering into the Anthropocene (Lewis & Maslim, 2015).

We have only a poor understanding about the global consequences that can unfold in the coming decades, but the signals are menacing. We are losing hundreds, perhaps thousands of species in very short spans of time, many of them vanishing even before being discovered and described. Certain taxonomic groups, such as amphibians, appear more vulnerable than others (Wake & Vredenburg, 2008). Mammals are far from safe (Ceballos et al, 2017). Their position as tertiary consumers in most of the ecosystems where they occur means that they can only attain relatively low population densities. The situation worsens for habitat and/or dietary specialists, such as forest hypercanivores. Losing one species is sad, losing many species unleashes ecological havoc, but from a purely esthetical perspective, few losses should be comparable to the disappearance of majestic beasts such as tigers, lions and jaguars.

In recent decades, big cat species have been used as flagship icons for global conservation and their role as surrogate taxa for defining and setting priorities can only increase in the years to come. Unfortunately, the efforts of the academic community have not been as fully operational as required to respond to the fast-paced threats, through the translation of research insights into effective conservation policies. Genomics represents the most recent addition to the repertoire of Conservation Biology approaches (Schafer et al., 2015), and as such it conveys a great potential to help fill in the gaps in a more expedited way.

Our understanding of biological phenomena is been revolutionized by high-throughput sequencing (HTS). The advent of these technologies and a steady reduction in their costs now allows sequencing the whole genome of wildlife (non-model) species, which holds a huge potential for the development of high-resolution molecular studies. Now we are able to search and characterize

patterns at an increasing level of detail, with a concomitant improvement in inferential power on the underlying processes. These technological advances include a variety of techniques to perform genome-wide assessments. Depending on the questions and applications, researchers choose the focal regions under scrutiny, from specific loci to long haplotypes and chromosomes, up to whole genomes scanned through sliding-windows approaches.

The results of this dissertation derive from the complementary application of four distinct molecular approaches: short-tandem repeat (STR) genotyping, restriction-site associated DNA sequencing (RADseq), whole-exome sequencing (WES), and whole-genome sequencing (WGS). Each of these methods varies in the way loci are recovered and makers are called and filtered, which are key steps for downstream analyses, ultimately determining the robustness and predictive power of resulting inferences. The project emerged as a follow-up on the international collaborative efforts led by the PUCRS Laboratory of Genomics and Molecular Biology, which constitute the ongoing Jaguar Genome Project.

In the first study (paper 1), we reported a genetic study including the most comprehensive geographic coverage of Amazon jaguar samples to date, which allowed us to demonstrate that this region sustains one of the largest and most genetically diverse jaguar populations, and highlight its use as a baseline against which fragmented demes can be assessed and monitored. A limitation of this study is that, in spite of its usefulness for using Amazon jaguar diversity estimates as a global yardstick, the markers it employed (microsatellites) are being progressively replaced by HTS, which will require that this assessment be repeated in the near future with novel, genome-wide markers.

In the second study (paper 2), we compared the performance of GBS and WES to characterize genomic diversity in five different jaguar populations We elaborated on the null hypothesis that both approaches should recover a similar biological signal despite differences in data collection strategy and analytical approaches. The alternative hypothesis is that estimates are sensitive to these differences, and may be affected by changes in the way SNP calling is performed and population filter parameters are applied. We showed that both under- and over-parametrization have measurable effects on the

inferences drawn. Thus, we stress the importance of fine-tuning the parametrization process in order to reduce bias-inducing artifacts, so as to increase precision in the recovery of meaningful biological information for the populations of interest.

As for the third study (paper 3), we used whole genome sequence data to generate models of demographic history, both ancient as contemporaneous. Despite low sampling size, we were able to recover interesting signals about historical demography and recent genomic erosion driven by anthropic disturbance. Future directions in this study, which will be further expanded prior to publication, are to perform additional analyses of demographic history, as well as to increase the dataset by inclusion of more samples.

Genomic data allow the recovery of fine patterns in evolutionary history. When we sample many individuals from a single population, we can adequately infer the patterns within that population, but we remain agnostic if that pattern holds for other populations. Conversely, when we sample a few individuals, or even one individual per population, we can infer some aspects of demographic history as well as some inter-population patterns, thanks to the massive amount of data provided by WGS. As the field develops, both approaches will be gradually integrated, with large amounts of genomic data collected simultaneously for population-level samples, thus allowing an even better inference on that taxon's evolutionary history.

Our major findings indicate that the Amazon basin represents the main global stronghold for jaguars, while the Atlantic Forest jaguar's condition is critical, as small-population adverse effects are becoming more evident for this deme. As next steps, we recommend continuing molecular research, especially using HTS. In the next fear years, it should be advisable to increase geographic coverage to include at least one whole genome sequenced per biome across the jaguar's range. This would help to clarify local adaptation patterns and relative susceptibilities to different threats among demes. For the demes showing precarious outlooks, such as the Atlantic Forest, genomic-empowered guidelines for crisis management plans can be applied, such as restoring habitat connectivity and deciding about the best options for genetic rescue, such as translocating individuals among demes or even artificial

insemination of females using sperm from other isolated areas. While the challenges for jaguar conservation in human-dominated landscapes remain large, genomic approaches now offer additional and refined tools in the armory of resources that can be employed in this inter-disciplinary effort.

REFERENCES

Benson, E. P. (1998). The lord, the ruler: Jaguar symbolism in the Americas. In N. J. Saunders (Ed.), *Icons of power: feline symbolism in the Americas* (pp. 53–76). London: Routledge.

Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using genomic data to infer historic population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics*, 49, 433–456.

Brown, J. H., & Maurer, B. A. (1986). Body size, ecological dominance and Cope's rule. *Nature*, 324, 248.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, 1(5), e1400253.

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, 114, E6089–E6096.

Coppolillo, P., Gomez, H., Maisels, F., & Wallace, R. (2004). Selection criteria for suites of landscape species as a basis for site-based conservation. *Biological Conservation*, 115, 419–430.

De Barros, A. E., Macdonald, E. A., Matsumoto, M. H., Paula, R. C., Nijhawan, S., Malhi, Y., & Macdonald, D. W. (2014). Identification of areas in Brazil that optimize conservation of forest carbon, jaguars, and biodiversity. *Conservation Biology*, 28, 580–593.

De la Torre, J. A., González-Maya, J. F., Zarza, H., Ceballos, G., & Medellín, R. A. (2018). The jaguar's spots are darker than they appear: assessing the global conservation status of the jaguar Panthera onca. *Oryx*, *52*, 300–315.

Foster, V. C., Sarmento, P., Sollmann, R., Tôrres, N., Jácomo, A. T., Negrões, N., Fonseca, C. & Silveira, L. (2013). Jaguar and puma activity patterns and predator-prey interactions in four Brazilian biomes. *Biotropica*, 45, 373–379.

Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., Leitch, I. J., Murray, B. G., Kapraun, D. F., Greilhuber, J. & Bennett, M. D. (2006). Eukaryotic genome size databases. *Nucleic Acids Research*, 35(suppl_1), D332–D338.

Haag, T., Santos, A. S., Sana, D. A., Morato, R. G., Cullen Jr, L., Crawshaw Jr, P. G., De Angelo, C., Di Bitetti, M. S., Salzano, F. M. & Eizirik, E. (2010). The effect of habitat fragmentation on the genetic structure of a top predator: loss of diversity and high differentiation among remnant populations of Atlantic Forest jaguars (*Panthera onca*). *Molecular Ecology*, 19, 4906–4921.

Jędrzejewski, W., Robinson, H. S., Abarca, M., Zeller, K. A., Velasquez, G., Paemelaere, E. A., … & Quigley, H. (2018). Estimating large carnivore populations at global scale based on spatial predictions of density and distribution–Application to the jaguar (*Panthera onca*). *PloS One*, *13,* e0194719.

Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences*, 110, E2602–E2610.

Jorge, M. L. S., Galetti, M., Ribeiro, M. C., & Ferraz, K. M. P. (2013). Mammal defaunation as surrogate of trophic cascades in a biodiversity hotspot. *Biological Conservation*, 163, 49–57.

Kurtén, B., & Anderson, E. (1980). *Pleistocene mammals of North America*. New York: Columbia University Press.

Lewis, S. L., & Maslin, M. A. (2015). Defining the anthropocene. *Nature*, 519, 171.

Michalski, F., Boulhosa, R. L. P., Faria, A., & Peres, C. A. (2006). Human–wildlife conflicts in a fragmented Amazonian forest landscape: determinants of large felid depredation on livestock. *Animal Conservation*, *9*, 179–188.

Miller, B., Dugelby, B., Foreman, D., Del Río, C. M., Noss, R., Phillips, M., Reading, R., Soulé, M. E., Terborgh, J., & Willcox, L. (2001). The importance of large carnivores to healthy ecosystems. *Endangered Species Update*, 18, 202–210.

Mittermeier, R. A., Myers, N., Thomsen, J. B., Da Fonseca, G. A., & Olivieri, S. (1998). Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. *Conservation Biology*, 12, 516–520.

Morato, R. G., Connette, G. M., Stabach, J. A., De Paula, R. C., Ferraz, K. M. P. M., Kantek, D. L. Z., Miyazaki, S. S., Pereira, T. D. C., Silva, L. C., Paviolo, A., De Angelo, C., Di Bitetti, M. S., Cruz, P., Lima, F., Cullen, L., Sana, D. A., Ramalho, E. E., Carvalho, M. M., da Silva, M. M., Moraes, M. D. F., Vogliotti, A., May Jr., J. A., Haberfeld, M., Rampim, L., Sartorello, L., Araujo, G. R., Wittenmeyer, G., Ribeiro, M. C., & Leimgruber, P. (2018). Resource selection in an apex predator and variation in response to local landscape characteristics. *Biological Conservation*, 228, 233–240.

Paviolo, A., De Angelo, C., Ferraz, K. M. P. M. B., Morato, R. G., Martinez Pardo, J., Srbek-Araujo, A. C., Beisiegel, B. de M., Lima, F., Sana, D., Xavier da Silva, M., Velázquez, M. C., Cullen, L., Crawshaw Jr., P., Jorge, M. L. S. P., Galetti, P. M., Di Bitetti, M. S., de Paula, R. C., Eizirik, E., Aide, T. M., Cruz, P., Perilli, M. L. L., Souza, A. S. M. C, Quiroga, V., Nakano, E., Ramírez Pinto, F., Fernández, S., Costa, S., Moraes Jr., E. A., & Azevedo, F. (2016). A biodiversity hotspot losing its top predator: The challenge of jaguar conservation in the Atlantic Forest of South America. *Scientific Reports*, 6, 37147.

Redford, K. H. (1992). The empty forest. *BioScience*, 42, 412–422.

Ripple, W. J., Estes, J. A., Beschta, R. L., Wilmers, C. C., Ritchie, E. G., Hebblewhite, M., Berger, J., Elmhagen, B., Letnic, M., Nelson, M. P., Schmitz, O. J., Smith, D. W., Wallach, A. D., & Wirsing, A. J. (2014). Status and ecological effects of the world's largest carnivores. *Science*, 343, 1241484.

Roques, S., Sollman, R., Jácomo, A., Tôrres, N., Silveira, L., Chávez, C., da Luz, X. B. G., Magnusson, W. W. E., Godoy, J. A., & Palomares, F. (2016). Effects of habitat deterioration on the population genetics and conservation of the jaguar. Conservation Genetics, 17, 125–139.

Ruíz-García, M., Payán, E., Murillo, A., & Álvarez, D. (2006). DNA microsatellite characterization of the jaguar (*Panthera onca*) in Colombia. *Genes & Genetic Systems,* 81, 115–127.

Sanderson, E. W., Redford, K. H., Chetkiewicz, C. L. B., Medellin, R. A., Rabinowitz, A. R., Robinson, J. G., & Taber, A. B. (2002). Planning to save a species: the jaguar as a model. *Conservation Biology*, 16, 58–72.

Seymour, K. L. (1989). *Panthera onca*. *Mammalian Species*, 340, 1–9.

Shafer, A. B., Wolf, J. B., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., ... & Fawcett, K. D. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, *30*, 78–87.

Smith, N. J. (1976). Spotted cats and the Amazon skin trade. *Oryx*, 13, 362–371.

Srbek-Araujo, A. C., Haag, T., Chiarello, A. G., Salzano, F. M., & Eizirik, E. (2018). Worrisome isolation: noninvasive genetic analyses shed light on the critical status of a remnant jaguar population. *Journal of Mammalogy*, 99, 397–407.

Swank, W. G., & Teer, J. (1989). Status of the jaguar–1987. *Oryx*, 23, 14–21.

Valdez, F. P., Haag, T., Azevedo, F. C., Silveira, L., Cavalcanti, S. M., Salzano, F. M., & Eizirik, E. (2015). Population genetics of jaguars (*Panthera onca*) in the Brazilian Pantanal: molecular evidence for demographic connectivity on a regional scale. *Journal of Heredity*, 106, 503–511.

Wake, D. B., & Vredenburg, V. T. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences*, *105*(Supplement 1), 11466–11473.

Wilkie, D. S., Bennett, E. L., Peres, C. A., & Cunningham, A. A. (2011). The empty forest revisited. *Annals of the New York Academy of Sciences*, 1223, 120–128.

Woodburne, M. O. (2010). The Great American Biotic Interchange: dispersals, tectonics, climate, sea level and holding pens. *Journal of Mammalian Evolution*, 17, 245–264.

Wultsch, C., Waits, L. P., & Kelly, M. J. (2016). A comparative analysis of genetic diversity and structure in jaguars (*Panthera onca*), pumas (*Puma concolor*), and ocelots (*Leopardus pardalis*) in fragmented landscapes of a critical Mesoamerican linkage zone. *PloS One*, 11, e0151043.

Zeilhofer, P., Cezar, A., Torres, N. M., de Almeida Jacomo, A. T., & Silveira, L. (2014). Jaguar *Panthera onca* habitat modeling in landscapes facing high land-use transformation pressure. Findings from Mato Grosso, Brazil. *Biotropica*, 46, 98–105.

Zeller, K. (2007). Jaguars in the new millennium data set update: the state of the jaguar in 2006. New York: Wildlife Conservation Society.

Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br