PUCRS

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

FERNANDO RICARDO VIEIRA LOPES

FILOGENÔMICA DE OTARIIDAE E HISTÓRIA EVOLUTIVA DE LOBOS-MARINHOS DA AMÉRICA DO SUL

Porto Alegre
2019

PÓS-GRADUAÇÃO - *STRICTO SENSU*

Pontifícia Universidade Católica
do Rio Grande do Sul

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA DE CIÊNCIAS DA SAÚDE E DA VIDA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E EVOLUÇÃO DA BIODIVERSIDADE

FILOGENÔMICA DE OTARIIDAE E HISTÓRIA EVOLUTIVA DE LOBOS-MARINHOS DA AMÉRICA DO SUL

Fernando Ricardo Vieira Lopes
Orientador: Dr. Sandro L. Bonatto
Coorientadora: Dra. Larissa Rosa de Oliveira

TESE DE DOUTORADO

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
Porto Alegre - RS - Brasil
2019

Lobo-marinho-peruano-chileno
Foto: Juan Jeri

SUMÁRIO

Peruvian/Chilean fur seals pups in Punta San Juan, Peru
Foto: Juan Jeri

"It always seems impossible until it's done."

- Nelson Mandela

À minha família.

Primeiramente, obrigado aos melhores orientadores que um doutorando poderia ter. Essa tese é fruto de nosso trabalho conjunto.

Ao meu orientador, Dr. Sandro Bonatto, pela oportunidade de desenvolver um projeto incrível. Obrigado pela liberdade intelectual, pela presença diária, pelas discussões a cada novo resultado gerado, pelo incentivo constante e por alavancar meu conhecimento e pensamento crítico em Biologia Evolutiva.

À minha coorientadora de Doutorado e orientadora de vida acadêmica, Dra. Larissa Oliveira. Nesse agosto de 2019 completamos 10 anos de colaboração. Obrigado, minha mãe científica, por ter me posto nos trilhos da genética de lobos-marinhos, por viabilizar o estudo com os lobinhos peruanos (e todos os outros, mas em especial este), por ter ajudado a construir minha carreira profissional e ter acreditado no meu potencial desde o primeiro minuto até hoje. Obrigado por cada contato profissional (nacional e internacional) e por me ensinar os passos da vida acadêmica.

À CAPES pela bolsa de Doutorado e pela bolsa de Doutorado Sanduíche. Ao CNPq, à FAPERGS e a Navy-NICOP pelo financiamento do projeto.

À PUCRS pela infraestrutura disponibilizada e pelo maravilhoso período de convivência no campus da universidade.

À minha família. À minha mãe Adeline Lopes, irmã Fernanda Lopes, afilhada e sobrinha mais linda do mundo, Maya Lopes, meu pai Eurico Lopes (*in memorian*), "namorida" Nathaly Miranda e sogra Iona Lucas. Obrigado por compreenderem as dificuldades de ascensão na vida acadêmica, por cada minuto de apoio, por cada mãe, pai, irmã, sogra, amor-trocínio! Sem dúvida, as mulheres mais fortes e batalhadoras que conheço. Nenhum dia começaria e terminaria com sucesso se não fosse por cada uma de vocês!

Ao amor da minha vida, Nathaly Miranda. Obrigado por compartilhar cada segundo comigo. Por estar presente em cada café da manhã, almoço, janta e "lanchos". Por cada ônibus lotado, vazio, escaldante e congelante que vivenciamos nas manhãs, tardes e noites. Obrigado por atravessar o mundo comigo e por viver os momentos mais incríveis de nossas vidas. Obrigado, também, por compartilhar os piores momentos. Crescemos juntos nas dificuldades. Obrigado pelas séries "pra esvaziar a cabeça" e pelo apoio incondicional. Obrigado por cada taça de vinho e copo de cerveja que não me deixaram surtar. Obrigado por ser meu sustento, meu início, meio e fim. Obrigado por me

compreender, principalmente nos momentos mais difíceis. Obrigado por ser e estar na minha vida. Te amo!

Aos amigos, em especial, Pedro Silva e Angelo Werner pelos momentos de descontração, pelas cervejas, jantas, churrascos e gargalhadas que tornaram os anos de doutorado mais fáceis de serem vivenciados.

À amiga, colega, e ex-bolsista de iniciação científica Amanda Kessler. Muito obrigado por ter me ajudado a decifrar como executar inúmeras análises, pelas sugestões textuais e de cores, por todas as revisões e extrações de DNA. Obrigado também pelos momentos de desabafo coletivo, pelos conselhos e almoços.

Ao bolsista de iniciação científica Yago Beux pelas análises, pelos infinitos cafés e por ter me salvado diversas vezes em muitas análises, principalmente nas "emergenciais". Te coorientatar também foi um aprendizado!

À Jéssica Pereira pelas centenas de extrações de DNA na sua iniciação científica.

Aos genômicos pelo cotidiano, pelas sugestões e discussões na sala do café. Em especial Gustavo Lorenzana, Fabrício Garcez, Manuel Escalona, Fernanda Trindade e Sarah Santos pelos códigos salvadores e discussões aprofundadas. Meus especiais agradecimentos às técnicas Giovanna Oliveira e Vanessa Engers por todo auxílio e empenho na elaboração das bibliotecas genômicas e sequenciamentos de DNA.

Aos professores do Programa de Pós-graduação em Ecologia e Evolução da Biodiversidade. Em especial às bancas de acompanhamento e qualificação compostas pelos Doutores Eduardo Eizirik, Nelson Fontoura e Paulo Henrique Ott (UERGS).

Aos colaboradores e suas instituições que permitiram a completude dos manuscritos. Em especial Dr. Jochen B. W. Wolf, Dra. Claire Peart e Ludwig-Maximilians-Universität München que me receberam no Doutorado Sanduíche.

A Juan Jeri pelas fotos que ilustraram a tese.

Muito obrigado a todas as pessoas e instituições não mencionadas anteriormente e que auxiliaram no desenvolvimento da tese.

A sistemática, a filogenia e a história evolutiva de Otariidae (lobos e leões-marinhos) são debatidas desde as descrições das primeiras espécies por exploradores naturalistas europeus. Atualmente, diversas relações, em particular dentro de *Arctocephalus*, permanecem confusas. Estudos moleculares recentes envolvendo espécies da família basearam-se somente em poucos fragmentos de DNA, em sua maioria, regiões mitocondriais concatenadas, com o objetivo de decifrar a biogeografia dos lobos e leões-marinhos. Assim como para a Otariidae, as relações entre *Arctocephalus australis* (lobo-marinho-peruano-chileno e lobo-marinho-sul-americano), *A. galapagoensis* (lobo-marinho-de-Galápagos) e *A. forsteri* (lobo-marinho-da-Nova Zelândia) também são bastante discutidas. Embora atualmente reconhecidas como espécies plenas, todos os estudos com um enfoque populacional, aplicando mais de um indivíduo por táxon, demonstraram a parafilia deste grupo. Em dois manuscritos aplicamos uma abordagem genômica para esclarecer as relações filogenéticas de Otariidae e o processo evolutivo de *A. australis* (e suas duas subunidades populacionais), *A. galapagoensis* e *A. forsteri*. Para isso, avaliamos 15 genomas completos, sendo 12 sequenciados pelo nosso grupo de pesquisa e três obtidos do GenBank. Além disso, também sequenciamos uma biblioteca de representação reduzida (método ddRAD-seq) composta por indivíduos de *A. australis* (n=15/13, subpopulações do sul do Chile/Oceano Atlântico e Peru/Norte do Chile, respectivamente), *A. galapagoensis* (n=10) e *A. forsteri* (n=9). Em nosso trabalho, encontramos um elevado grau de discordância genealógica ao longo dos genomas, o qual atribuímos, principalmente, à separação incompleta de linhagens gênicas. Identificamos que o elevado número de topologias únicas se deu à um explosivo período de especiação conduzido pelo resfriamento das águas superficiais do Oceano Pacífico, na transição Plio-Pleistoceno há ~3 Ma. Além disso, métodos variados de reconstruções filogenéticas recuperaram, de forma consistente, a monofilia do gênero *Arctocephalus*, sugerindo a manutenção da atual nomenclatura taxonômica. Análises com genomas completos e bibliotecas de representação reduzida também mostraram a clara delimitação entre *A. australis*, *A. galapagoensis* e *A. forsteri*. Encontramos, também, evidências de que a subpopulação peruana/chilena de *A. australis* compartilha componentes genômicos entre *A. galapagoensis* e a subpopulação do sul do Chile/Oceano Atlântico de *A. australis*. Análises posteriores suportaram que esta mistura reflete uma antiga origem híbrida da subunidade populacional do lobo-marinho-peruano/chileno. Também detectamos a presença de indivíduos puros de *A. galapagoensis* (primeiramente identificados como *A. australis*) em Isla Foca, uma colônia reprodutiva isolada e localizada há ~1000 km ao norte da distribuição principal de *A. australis* (subunidade peruano/chilena) e à leste da distribuição de *A. galapagoensis*. Dois retrocruzamentos entre o táxon híbrido e *A. galapagoensis* também foram detectados em Isla Foca e em Punta San Juan, Peru, a maior colônia reprodutiva dos lobos-marinhos-peruanos-chilenos. Esses achados demonstram a necessidade de futuros estudos em Isla Foca para compreender o seu papel nas dinâmicas populacionais, na distribuição e na conservação dos táxons mencionados. Nos artigos apresentados delimitamos espécies, esclarecemos a complexa história evolutiva de lobos e leões-marinhos e resolvemos questionamentos fundamentais sobre as relações internas de Otariidae. Também demonstramos a importância de considerar as discordâncias genealógicas na interpretação da história evolutiva e de considerar o fluxo gênico interespecífico como força evolutiva na geração de biodiversidade.

Palavras-chave: filogenia; filogeografia; evolução; pinípedes; genômica

The systematics and phylogeny of Otariidae have been extensively studied for over two centuries since the first species descriptions by the European explorers. Yet, several relationships, particularly the monophyly within *Arctocephalus*, remain unclear. Recent molecular phylogenies only used few concatenated mitochondrial or nuclear genes. Similarly, the relationships within the clade that encompasses the Galapagos (*A. galapagoensis*), New Zealand (*A. forsteri*), South American and Peruvian-Chilean fur seals (*A. australis* subpopulations) are also debated. Although currently recognized as full species, the studies that assessed more than one individual per taxon showed the group paraphyly. In two manuscripts presented in this thesis, we applied a genomic approach to clarify conflicting Otariidae relationships and to shed light on the evolutionary processes of South American, Peruvian-Chilean, Galapagos and New Zealand fur seals. Here, we assessed 15 genomes (12 sequenced by our group and three retrieved from GenBank) and sequenced a reduced representation library composed by 47 individuals of the South American and Peruvian-Chilean fur seals (n= 15/13), Galapagos fur seals (n=10) and New Zealand fur seals (n=9). Our phylogenies showed a high-level of genealogical discordances that we assigned mainly to the incomplete lineage sorting promoted by an explosive adaptive-radiation at ~3Mya in the Plio-Pleistocene transition, potentially induced by the eastern Pacific sea surface cooling. We also found a strong support for the *Arctocephalus* monophyly, suggesting the maintenance of the current taxonomic nomenclature. Whole-genome and reduced representation libraries analyses showed well-delimited species for the clade encompassing South American, Peruvian-Chilean, Galapagos and New Zealand fur seal. Additionally, Peruvian-Chilean fur seals showing shared genomic components with Galapagos and South American fur seals. Posterior analyses supported that this admixture reflected the hybrid origin of the Peruvian-Chilean fur seals and the ancient evolutionary history of this taxa. We also detected two pure individuals of Galapagos fur seals in Isla Foca, a remote rookery located ~1,000 km from the main distributions of the Peruvian-Chilean and Galapagos fur seals. Besides that, we found a secondary contact of the Peruvian-Chilean with Galapagos fur seals in this locality and in Punta San Juan, southern coast of Peru. These findings highlight that new studies are needed to understand the role of Isla Foca in the population dynamics, distribution and conservation of the Peruvian-Chilean and Galapagos fur seals. The manuscripts presented in this thesis untangled the complex evolutionary history of the fur seals and sea lions by illuminating internal Otariidae relationships and showing the importance of the interspecific gene flow in the biodiversity promotion.

Keywords: phylogeny; phylogeography; evolution; pinnipeds; genomics

Esta tese é composta por quatro capítulos, dos quais dois estão formatados como artigos científicos e serão submetidos a dois periódicos de alto impacto. O primeiro artigo, que trata da filogenômica de Otariidae, será submetido ao periódico *Systematic Biology*. O segundo manuscrito, que trata da história evolutiva do grupo *Arctocephalus australis*, *Arctocephalus galapagoensis* e *Arctocephalus forsteri*, será submetido ao periódico *Molecular Biology and Evolution*.

1. Introdução

Na sequência serão abordadas as aplicações do sequenciamento de alto desempenho no estudo da sistemática filogenética, bem como será discutido o estado atual da filogenômica na reconstrução de árvores de espécies. Também será contextualizado o estado da arte das relações filogenéticas de lobos e leões-marinhos (Otariidae).

1.1 Sequenciamento de alto desempenho: aplicações na filogenômica

A popularização na geração de genomas tornou possível a construção de genomas de alta qualidade e redundância, também conhecidos como genomas de referência, para uma ampla diversidade de organismos, incluindo os pinípedes (morsas, focas, lobos e leões-marinhos) (Foote et al. 2015; Humble et al. 2016). Até o momento, estão publicados os genomas completos de três espécies de Otariidae: o lobo-marinho-do-norte, *Callorhinus ursinus* (GeneBank Bioproject PRJNA475116), o leão-marinho-de-Galápagos, *Zalophus wollebaeki* (Shafer et al. 2016), uma das espécies mais basais da família; e o lobo-marinho-antártico, *Arctocephalus gazella* (Humble et al. 2016).

Como exemplo dos avanços analíticos propiciados pela popularização na geração dos genomas podemos citar as análises filogenômicas. A filogenômica é campo da biologia molecular onde ocorre a intersecção entre evolução e filogenética no contexto genômico. O termo filogenômica também se refere às análises e estudos que reconstroem as histórias evolutivas com base no sequenciamento de genomas completos ou de grandes porções dos genomas de um ou diversos organismos. Quatro alicerces sustentam a filogenômica (Philippe e Blanchette 2007): (1) Predição da função gênica; (2) estudo e esclarecimento das relações evolutivas entre táxons; (3) evolução das famílias gênicas e; (4) predição e detecção de transferências laterais de genes.

Como exemplos de estudos filogenômicos podemos citar a publicação de Misof et al. (2014) e Foote et al (2015). No primeiro, os autores buscaram identificar, através da análise de sequências de nucleotídeos e aminoácidos dos insetos atuais, os tempos de divergência, as relações filogenéticas e os padrões de evolução do grupo. No segundo, os autores sequenciaram o genoma de três espécies de mamíferos marinhos pertencentes à diferentes ordens: o manatí (*Trichechus manatus*), a morsa (*Odobenus rosmarus*) e a orca (*Orcinus orca*); com o objetivo de investigar a evolução convergente

de fenótipos destas três espécies. Os autores encontraram, através de análises genômicas comparativas, substituições nucleotídicas convergentes que estariam sob seleção natural positiva associadas às adaptações ao ambiente marinho (Foote et al. 2015).

O aumento expressivo no número de regiões genômicas, quando comparado aos marcadores tradicionais de DNA, como as sequências mitocondriais curtas e os microssatélites, por exemplo, possibilitam maior precisão e refinamento em estudos de genética de populações, de filogeografia e de demografia histórica (Duran et al. 2009, Hoglund 2009, Ekblom e Galindo 2010). Entretanto, apesar dos preços de sequenciamento de alto desempenho terem reduzido abruptamente nos últimos 10 anos e de suas técnicas analíticas terem acompanhado este avanço, a aplicação de genomas completos em estudos filogeográficos pode ser um fator limitante para a execução de determinadas análises (Ekblom e Wolf 2014), pois estes estudos requerem a amostragem da distribuição completa das espécies e um número representativo de indivíduos amostrados (Avise 2009).

Para estudos de filogeografia é recomendado que a amostragem cubra toda a distribuição das espécies-alvo com diversos indivíduos em cada localidade amostrada, para que se possa compreender, inferir e testar como os processos ambientais influenciaram na distribuição das linhagens evolutivas ao longo do tempo (Avise 2009, Kumar e Kumar 2018). A ampla distribuição de algumas espécies, como a do lobo-marinho-sul-americano, *Arctocephalus australis*, que se distribui ao longo de toda a América do Sul, poderia aumentar a complexidade analítica devido ao alto volume de dados que poderiam ser gerados (Ekblom e Wolf 2014) para analisar os genomas completos de vários indivíduos das diferentes localidades amostradas. Para reduzir esta complexidade foram criadas estratégias de subamostragens genômicas, onde regiões-alvo são selecionadas para otimizar o número de indivíduos, o custo relativo e a quantidade de loci gênicos com boa redundância para responder as perguntas levantadas (Baird et al. 2008; Peterson et al. 2012).

Dentre as estratégias existentes, um conjunto de protocolos de bibliotecas de representação reduzida de DNA conhecidos como RADseq (*Restriction-site-Associated DNA sequencing*) vem sendo empregado de maneira eficiente em uma grande quantidade de trabalhos (Rowe et al. 2011; Andrews et al. 2016), incluindo aqueles que necessitam a utilização de amostras de DNA não-invasivo e/ou com um certo grau de degradação do DNA (Graham et al. 2015). O protocolo original foi proposto por Baird et al. (2008) e diversas adaptações surgiram, dentre elas o método ddRADseq (*Double Digest Restriction-site-Associated DNA sequencing* - Peterson et al. 2012) e suas adaptações. No método ddRADseq, duas enzimas de restrição são usadas para fragmentar o genoma, possibilitando uma maior precisão na seleção dos tamanhos, quantidade dos fragmentos e cobertura almejados, permitindo um melhor controle das regiões representadas na biblioteca (Peterson et al. 2012). Três características tornaram o protocolo de RADseq popular no estudo de organismos não-

modelos (da Fonseca et al. 2016): (1) custo-benefício em relação ao número de indivíduos e marcadores moleculares; (2) a não obrigatoriedade de um genoma de referência a ser incluído nas análises; e (3) o *trade-off* entre preço relativo (custo vs. volume de dados) e exigência computacional quando comparados ao sequenciamento de genomas completos (Peterson et al. 2012). Adicionalmente, estudos recentes têm mostrado que a utilização de genomas de referência e diferentes *pipelines* de análise para o mesmo grupo de dados sejam testados para dar confiabilidade aos resultados encontrados (Shafer et al. 2016).

Dentre as variações dos métodos de ddRADseq podemos citar, como exemplo, os métodos *Low-cost* ddRADseq e *hy*RAD (Kess et al. 2016; Suchan et al. 2016). Tais variações permitem ao usuário uma aplicação adequada do método às necessidades de sua pesquisa. O primeiro, por exemplo, apresenta uma preparação otimizada do conjunto de protocolos, reduzindo o número de passos totais e o custo final da preparação das bibliotecas (Kess et al. 2016); enquanto o segundo permite a aplicação em amostras de coleção ou "DNA difícil". Nesta abordagem, os fragmentos biotinilados de RAD, que cobrem uma fracção aleatória do genoma, são usados como iscas para capturar fragmentos homólogos nas bibliotecas de sequenciamento (Suchan et al. 2016).

Além disso, o método de sequenciamento de DNA via ddRADseq tem sido amplamente utilizado em estudos de genômica evolutiva, filogeografia, filogenômica e genômica da conservação que visam a compreender melhor a delimitação de espécies e unidades populacionais (Andrews et al. 2016). Dentre as aplicações do método se pode citar a busca por estrutura populacional, investigação de casos hibridação e interpretação da história demográfica, principalmente em um nível intraespecífico ou interespecífico de grupos recentemente divergidos (Andrews et al. 2016). Outras aplicações também têm sido feitas (Lowry et al. 2016) como, por exemplo, *genome scans* para detectar loci envolvidos em convergências adaptativas, detecção de loci de características quantitativas (QTL), entre outros (Baxter et al 2011; Chutimanitsakun et al. 2011).

1.2 O estado atual da filogenômica: a reconstrução de árvores de espécies a partir de múltiplas regiões gênicas e as discordâncias genealógicas dentro de genomas

Durante muitos anos sistematas moleculares têm debatido as discordâncias entre os diferentes métodos de reconstruções filogenéticas e suas potenciais e reais discordâncias genealógicas para estudos filogenéticos e filogeográficos (Pamilo e Nei 1988; Maddison 1997; Edwards 2009). Entre as mais recentes discussões dentro do campo da filogenômica estão o paradigma da escolha do método de reconstrução filogenética de sequências *multiloci* (McVay e Carstens 2013; Liu et al. 2015)

e a aplicação de múltiplas regiões gênicas com o objetivo de reconstruir a história evolutiva de um determinado grupo taxonômico (Desluc et al. 2005).

Os principais modelos de reconstruções *multiloci* atualmente discutidos na literatura são as concatenações de sequências e os coalescentes multiespécies, ou *species tree* (Liu et al. 2009; Liu et al. 2015; Springer e Gatesy 2016; Edwards et al. 2016) (fig. 1). Assim, com o advento dos dados de sequenciamento em larga escala para estudos de sistemática filogenética molecular e filogeografia têm intensificado os debates sobre quais dos métodos são os mais apropriados para análises de múltiplas regiões gênicas (Springer e Gatesy 2016; Edwards et al. 2016). Embora alguns estudos em filogenômica tenham sugerido que a maioria das relações filogenéticas obtidas por meio de concatenação de sequências e *species tree* (fig. 1) são consistentes entre elas ou diferem umas das outras sem suporte estatístico (Tonini et al. 2015), alguns exemplos de relações conflitantes e com suporte estatístico têm destacado que as diferenças significativas entre os métodos aplicados estão nas bases analíticas dos dois modelos. Ou seja, divergem na essência de seus funcionamentos, neste caso, o tratamento individual da árvore de genes (McVay e Carstens, 2013; Liu et al. 2015; Edwards et al. 2016).

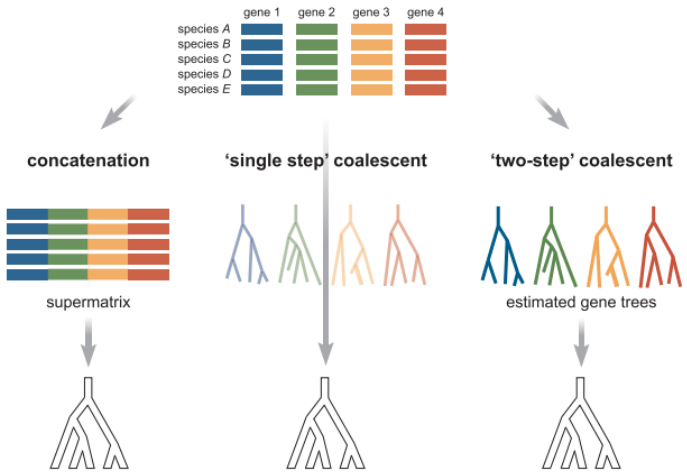


Figura 1. Diferenças no tratamento das regiões multigênicas para as estimativas filogenéticas utilizando dados concatenados (à esquerda) e os métodos de árvores de espécies (centro e direita). Figura retirada em sua plenitude de Liu et al. (2015).

O método de concatenação estima a filogenia unindo os dados de múltiplos genes de uma amostra. Nessa abordagem, mais tradicional, os dados são tratados como um único *locus* gênico e, essencialmente, a estimativa da genealogia de cada *locus* é a média através de genes (e.g., Manthey et al. 2016). Nessa abordagem se esperaria que a acurácia da filogenia melhorasse com o aumento do número de sítios variáveis no alinhamento (Hills et al. 1994), além de uma congruência, *a priori*, das topologias das árvores de genes. Entretanto, se a congruência é inexistente, então, se espera erros na estimativa filogenética (Carstens e Knowles 2007; Liu et al. 2009). Na figura 1 (à esquerda), se pode

ver uma representação esquemática de uma supermatriz de dados concatenados como um "supergene", uma abordagem que pode ser aplicada em análises filogenéticas com marcadores tradicionais ou de sequenciamento de DNA de larga escala em programas como o RAxML (Stamatakis 2006), que utiliza como base a Máxima Verossimilhança.

Nos métodos filogenômicos multiespécies, se infere a filogenia assumindo a independência e a livre recombinação dos genes, permitindo lidar com sinais evolutivos heterogêneos (variação na composição de bases nucleotídicas entre as regiões gênicas, diferentes taxas evolutivas, de recombinação gênica e incongruência topológica entre as árvores de genes, por exemplo) provindos das populações e seus indivíduos amostrados (Edwards 2009; Szöllosi et al. 2015). Quando cópias gênicas são amostradas de diferentes populações/espécies, as árvores gênicas podem apresentar diferentes histórias evolutivas e discordar genealogicamente. Ao menos sete processos têm sido descritos como fonte de discordância entre as árvores de genes: transferência horizontal de genes, hibridação, divergência incompleta das linhagens gênicas, estrutura genômica, demografia, seleção natural e estruturação filogeográfica (Maddison 1997; Davidson et al. 2015; Nacir e Linder 2015). Neste contexto, este método tem se tornado padrão na delimitação de espécies em estudos filogeográficos que utilizam uma abordagem *multiloci* (Fujita et al. 2012; Edwards e Knowles 2014). No centro da figura 1, se observa a classe de *species tree* denominada *one-step,* utilizada no programa BEAST e StarBEAST (Bouckeart et al. 2014). Nesta classe, as árvores de genes são, concomitantemente, estimadas pelo método Bayesiano, dadas as sequências *multiloci* e seus respectivos *priors* em função da verossimilhança. À direita, na figura 1, se encontra o modelo coalescente *two-steps,* utilizada no programa ASTRAL-III (Zhang et al. 2018). Nessa abordagem, as árvores de genes são geradas individualmente em programas como o RAxML e, posteriormente, utilizadas de entrada para uma estimativa de *species tree*.

Neste sentido, a evidência mais importante para aqueles que precisam decidir quanto ao uso de concatenação ou *specie tree* seja o grau de incongruência entre as regiões gênicas. Ou seja, se as árvores de genes forem em sua maioria congruentes e seus ramos suficientemente longos (antigos) para permitir a segregação completa das linhagens (McVay e Carstens 2013), então a concatenação dos dados é justificada.

Springer e Gatesy (2016) publicaram duras críticas a respeito da implementação de *species tree* como modelo para análises em filogenômica. As críticas se direcionaram, principalmente, ao relacionar erros nos sets de dados de algumas das mais conhecidas e citadas publicações em filogenômica (McCormack et al. 2012; Song et al. 2012; Xi et al. 2014). Além disso, estes autores declararam que as publicações mencionadas violaram aspectos básicos dos modelos de coalescência

de espécies, aplicando, assim, o teste de maneira errônea. Isso teria resultado em filogenias não confiáveis dos respectivos organismos abordados e em conclusões equivocadas nos manuscritos.

No entanto, Edwards et al. (2016), além de criticar o tom combativo e sarcástico de Springer e Gatesy (2016), rebateram e reanalisaram todos os pontos criticados pelos autores. Edwards et al. (2016) mostraram que os erros nas análises (principalmente em Xi et al., 2014) tiveram baixo impacto nas conclusões finais dos artigos. Além disso, Edwards et al. (2016) chamaram a atenção para diversos pontos de má interpretação do modelo de árvore de espécies destacados na publicação de Springer e Gatesy (2016). Também provaram, através de simulações, que o modelo de concatenação revela um comportamento errático nas análises e que o mesmo tende a produzir resultados com uma confiança exacerbada e espúria. Na réplica à crítica, Edwards et al. (2016) também defenderam que os pontos ressaltados por Springer e Gatesy (2016) são invalidados a partir do momento que a concatenação é considerada um tipo especial de *species tree* onde todas as árvores de genes são congruentes entre si.

Até recentemente, grande parte dos métodos e dos estudos que utilizaram estimativas de *species tree* como base foram conduzidas dentro da premissa da não existência de fluxo gênico pós divergência e da heterogeneidade nas árvores de genes fruto apenas de ILS (Fujita et al., 2012). No entanto, ignorar a presença de migração pode causar vieses nas inferências coalescentes multiespécies, bem como na estimativa dos parâmetros biológicos (Holder et al. 2001, Leaché et al. 2014). Leaché et al. (2014) detectaram que o fluxo gênico entre espécies não-irmãs pode resultar em erros topológicos, desvios significativos nas estimativas de tamanhos populacionais e compressão dos tempos de divergência, principalmente em ramos mais profundos da filogenia (Wen e Nakhleh 2018).

Diversos estudos já estão documentados e mostram que, apesar de difícil detecção, a diferenciação entre ILS e hibridação como fontes das discordâncias genealógicas podem e devem ser testadas (e.g. Rheindt et al., 2014, Figueiró et al. 2017), pois, em muitos casos, a especiação pode ocorrer de forma gradativa, permitindo fluxo gênico em grupos em processo de especiação ou em espécies já bem delimitadas (Leaché et al. 2014).

Como as evidências para introgressão genômica entre espécies diferentes tem crescido rapidamente, há uma pressão constante para a criação e utilização de métodos que infiram filogenias *species tree* e seus parâmetros associados na presença de hibridação (Pease e Hahn 2015; Wen e Nakhleh 2018; Jones 2019). Os principais métodos que consideram a migração na reconstrução de árvores de espécies diferem, principalmente, em como eles quantificam as relações entre os táxons e o desbalanço de filogenias discordantes (Pease e Hahn 2018). Alguns métodos, como PhyloNet (Than

et al. 2008) e o SpeciesNetwork (Zhang et al. 2017), implementado no pacote BEAST (Bouckeart et al. 2014), usam os dados empíricos de sequências nucleotídicas para reconstruir árvores de genes e, posteriormente, conciliar o resultado das topologias em uma filogenia reticulada (Sang e Zhong 2000; Holder et al. 2001; Paese e Hahn 2018). Outros métodos, como a estatística D (e.g. ABBA-BABBA e D_{FOIL}), distinguem ILS de introgressão ao calcular o desbalanço nas trocas de alelos entre dois táxons dada uma determinada topologia e baseando-se na média das diferenças nucleotídicas ou nas distâncias par-a-par (Durand et al. 2011; Paese e Hahn 2018). Além disso, métodos baseados em filogenias reticuladas também permitem estimar a proporção de introgressão entre as reticulações. Programas como o G-PhoCS (*Generalized Phylogenetic Coalescent Sampler*) (Gronau et al. 2011) tornam possível o refinamento das análises mencionadas. Este programa utiliza como entrada alinhamentos múltiplos de *loci* neutros e independentes extraídos de genomas completos. Assim, é possível inferir os tamanhos populacionais ancestrais, os tempos de divergência e taxas de migração entre bandas definidas pelo usuário (entre genomas e entre genomas seus ancestrais - Gronau et al. 2011).

1.3 Incertezas sobre as relações filogenéticas em Otariidae

Embora Otariidae (lobos e leões-marinhos) seja composta por espécies carismáticas e popularmente conhecidas, em nível de sistemática filogenética permanece confusa, sendo fruto de extenso debate científico há pelo menos dois séculos (e.g., Sclater 1897; Wynen et al 2001; Demeré e Berta 2003; Yonezawa et al 2009; Berta e Churchill 2012; Churchill et al. 2014; Nyakatura e Bininda-Emonds 2014; Berta et al. 2018).

Otariidae foi tradicionalmente classificada em duas grandes subfamílias: Otariinae (leões-marinhos) e Arctocephalinae (lobos-marinhos). Otariinae estaria composta por cinco gêneros: *Eumetopias*, *Neophoca*, *Otaria*, *Zalophus*, *Phocarctos* e *Neophoca*. Enquanto Arctocephalinae estaria composta por dois gêneros: *Arctocephalus*, incluindo oito espécies, e *Callorhinus*, composta por uma única espécie (Riedman 1990; Reynolds 1999).

Esta classificação, no entanto, foi refutada por se basear em apenas dois caracteres diagnósticos: a presença/ausência de subcamada de pelo e presença de cinco ou seis caninos superiores. Diversos estudos posteriores, entretanto, baseados em caracteres morfológicos, moleculares ou em análises integradas, descreveram as espécies do gênero *Arctocephalus* como mais semelhantes à Otariinae do que ao gênero *Callorhinus*, indicando Arctocephalinae como grupo parafiléticos e *Callorhinus* como espécie basal à família (Wynen et al. 2001; Árnason et al. 2006; Yonezawa et al. 2009; Churchill et al. 2014).

Estudos mais recentes, apontaram que origem de Otariidae teria ocorrido no hemisfério norte, com um único evento de especiação em direção ao hemisfério sul. Estes estudos identificaram que o grupo basal estaria composto pelos gêneros *Callorhinus*, *Zalophus* e *Eumetopias* (fig. 2 - Yonezawa et al. 2009; Nyakatura e Bininda-Emonds 2012; Berta e Churchill 2012; Churchill et al. 2014). Além disso, apontaram para a posição basal de *C. ursinus* em relação à Otariidae e para a parafilia do gênero *Arctocephalus* (e.g., Yonezawa et al. 2009; Nyakatura e Bininda-Emonds 2012; Churchill et al. 2014).

Embora estudos moleculares tenham mostrado que as relações filogenéticas permanecem estáveis nos ramos mais basais da filogenia de Otariidae (fig. 2), os ramos que, provavelmente, surgiram no hemisfério sul (gêneros *Arctocephalus*, *Otaria*, *Phocarctos* e *Neophoca*) permanecem sem uma delimitação clara e com relações não bem suportadas (Wynen et al. 2001; Árnason et al. 2006; Yonezawa et al. 2009; Nyakatura e Bininda-Emonds 2012; Berta e Churchill 2012; Churchill et al. 2014; Berta et al. 2018). O histórico confuso e a não resolução das posições filogenéticas da maior parte das espécies de Otariidae trazem à luz a problemática e a falta eficiência dos métodos tradicionais para resolver as questões internas do grupo.

Mesmo com o histórico problemático na sistemática de Otariidae, diversas inferências biogeográficas e até mesmo reclassificações taxonômicas foram feitas apenas em poucos caracteres morfológicos, pequenos fragmentos mitocondriais ou em análises combinadas envolvendo estes marcadores (Wynen et al. 2001; Árnason et al. 2006; Higdon et al. 2007; Yonezawa et al. 2009; Nyakatura e Bininda-Emonds 2012; Berta e Churchill 2012; Churchill et al. 2014). Berta e Churchill (2012) tentaram reconciliar as incongruências até então conhecidas e, baseados em filogenia combinada de dados publicados, realizaram uma revisão taxonômica de todos os pinípedes (morsas, focas, lobos e leões-marinhos) (fig. 2A). Baseados em apenas uma topologia, estes autores propuseram a manutenção do nome *Arctocephalus* apenas para a *Arctocephalus pusillus*, espécie tipo do gênero. Para as espécies remanescentes de *Arctocephalus* (*A. australis, A. galapagoensis, A. tropicalis, A. gazella* e *A. philippii*) sugeriram a transferência para o nome *Arctophoca*, nomenclatura prioritária proposta por Gardner e Robbins (1999), embora *A. tropicalis* e *O. flavescens* tenham sido reconstruídos junto ao clado de *A. pusillus*, porém sem resolução. No mesmo estudo, Berta e Churchill (2012) sugeriram quatro subespécies para *Arctocephalus australis* (*Arctophoca australis*), sendo que *Arctocephalus forsteri* estaria condicionada à subespécie deste táxon: *A. a. australis* (Ilhas Falklands), *A. a. gracilis* (América do Sul continental), *A. a. forsteri* (Nova Zelândia) e uma subespécie para a população peruana (Oliveira et al. 2008).

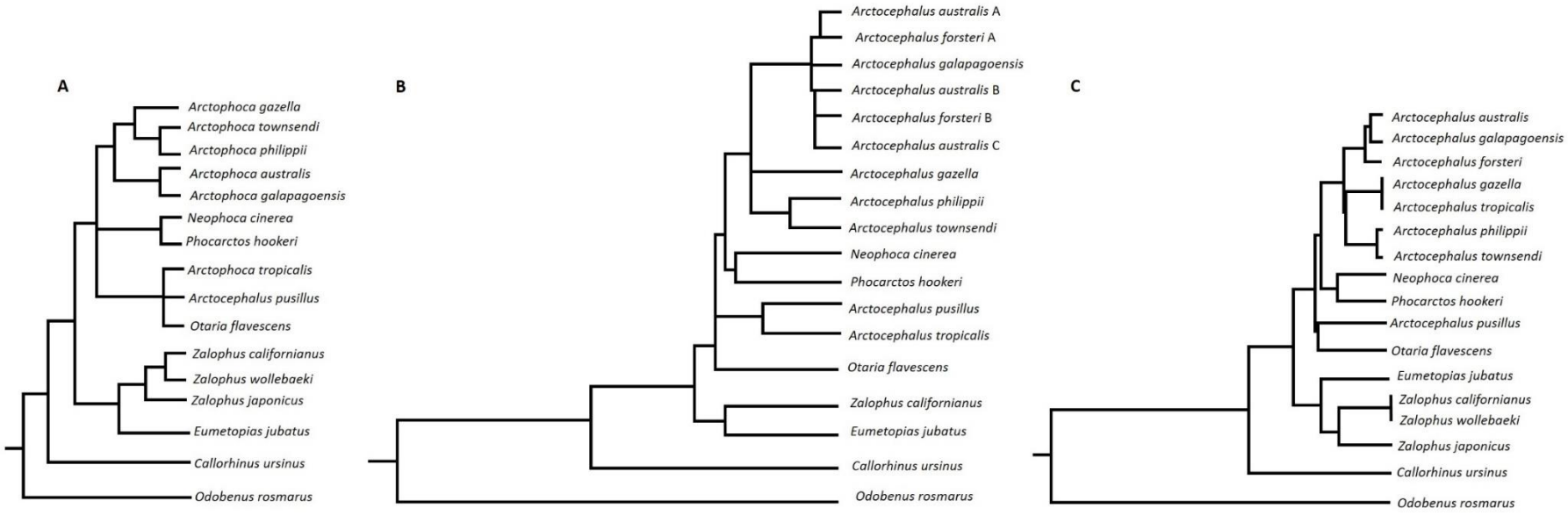


Figura 2. Filogenias mostrando as incertezas nas relações de Otariidae, bem como para o clado de *A. australis*, *A. galapagoensis* e *A. forsteri*. (A) Filogenia de Berta e Churchill (2012), baseada em topologias já publicadas (molecular e morfologia) e propondo reclassificação taxonômica para o gênero *Arctocephalus*; (B) Filogenia de mitogenoma de Yonezawa et al. (2009) e (C) Filogenia *supertree* de Nyakatura e Bininda-Emmonds (2012).

Em uma das publicações mais recente a respeito do tema, Churchill et al. (2014) conduziram, pela primeira vez, uma série de análises cladísticas combinando evidências morfológicas, moleculares e biogeográficas (tolerância das espécies às diferentes temperaturas superficiais dos oceanos). Este estudo mostrou ramos internos pobremente suportados para o grupo de lobos e leões-leões marinhos do hemisfério sul, principalmente quando as filogenias foram recuperadas apenas com caracteres morfológicos. Entre os principais fatores atuando na diminuição da variabilidade morfológica interespecífica estariam que os otarídeos, quando comparados aos outros pinípedes, apresentam uma menor variação na forma do crânio; que as recorrentes hibridações poderiam influenciar na morfologia; que a rápida divergência e que a perda de especialização ecológica dentro das linhagens poderiam diminuir as diferenças morfológicas entre as espécies (Yonezawa et al. 2009; Berta e Churchill 2012; Churchill et al. 2014). Churchill et al. (2014) também declaram que, apesar da filogenia molecular ter resultado em uma árvore com maior suporte estatístico, a quantidade total de regiões gênicas foi limitada e predominada por genes mitocondriais. Assim, as topologias recuperadas poderiam estar refletindo apenas a história deste gene (Maddison 1997; McVay e Carstens 2013).

Seguindo o histórico acima, a história evolutiva de *A. australis* e espécies relacionadas, *Arctocephalus galapagoensis* e *Arctocephalus forsteri*, também tem sido extremamente discutida. Com base no tamanho dos crânios, King (1954) sugeriu a existência de três subespécies de *A. australis*: *A. a. australis*, com localidade tipo as Ilhas Falkland (Malvinas); *A. a. gracilis,* que se distribuiria ao longo de todo a região continental da América do Sul; e a subespécie *A. a. galapagoensis*, com localidade tipo o arquipélago de Galápagos, no Equador. Revendo sua reclassificação taxonômica, King (1969; 1983) elevou *A. forsteri* ao *status* de espécie plena. Em estudo de morfometria craniana, Repenning et al. (1971), acordou ao *status* de espécie plena não somente para *A. forsteri* como também para *A. galapagoensis* e *A. australis*.

Wynen et al. (2001) e Yonezawa et al. (2009), ao recuperar a filogenia com base na região citocromo b (mtDNA) e do DNA mitocondrial completo, marcador não recombinante e de origem matrilinear, demonstraram a parafilia para o clado destas espécies (fig. 2B). Os autores apresentaram dois grupos principais: um composto por indivíduos de *A. australis* e *A. galapagoensis* com *A. forsteri* sendo basal; e o segundo composto apenas por *A. australis* e *A. forsteri*, mostrando-os como clados irmãos (Wynen et al. 2001; Yonezawa et al. 2009).

Em relação à *A. australis*, Oliveira et al. (2008), ao avaliar colônias do Peru e do Uruguai, reconheceram através da análise de morfometria craniana e de diferenças significativas nas frequências alélicas de sete *loci* de microssatélites de DNA a existência de, ao menos, duas unidades evolutivamente significativas (do inglês *Evolutionarily Significant Unit* - ESU) para cada uma das

localidades avaliadas (Uruguai e Peru). Ao comparar o DNA mitocondrial de amostras de lobos-marinhos-sul-americanos do Peru, Chile, Argentina e Uruguai, Rodrigues et al. (2018) reforçaram os achados de Oliveira et al. (2008), mostrando o isolamento genético da ESU peruana. Além disso, Rodrigues et al. (2018) também mostraram a conectividade genética entre as populações do sul do Chile e do Oceano Atlântico.

Embora a história evolutiva de *A. australis*, *A. galapagoensis* e *A. forsteri* tenha sido bastante discutida, até o presente não há estudo publicado que abranja a total distribuição, nem que tenha buscado delimitar estas espécies. Desta forma, nosso estudo se baseou na abordagem genômica para tentar esclarecer os conflitos existentes.

1.4 Referências Bibliográficas

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Genetics, 17: 81-92

Avise J. 2009. Phylogeography: retrospect and prospect. Journal of Biogeography, 36: 3-15

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE, 3: e3376

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, Blaxter ML. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PLoS ONE, 6: e19315

Berta A, Churchill M. 2012. Pinniped taxonomy: Review of currently recognized species and subspecies, and evidence used for their description. Mamm Rev. 42:207-234

Berta A, Churchill M, Boessenecker RW. 2018. The origin and evolutionary biology of pinnipeds: seals, sea lions and walruses. Annu Rev Earth Planet Sci. 46:203-228

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software plataform for Bayesian evolutionary analysis. PLoS Comput Biol. 10:e1003537

Carstens BC, Knowles LL. 2007. Estimating phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Syst Biol. 56:400-411

Churchill M, Boessenecker RW, Clementz MT. 2014. Colonization of the southern hemisphere by fur seals and sea lions (Carnivora: Otariidae) revealed by combined evidence phylogenetic and Bayesian biogeographical analysis. Zool J Linn Soc. 172:200-225

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistué L, Corey A, Filichkina T, Johnson EA, Hayes PM. 2011. Construction and application for QTL of a Restriction Site Associated DNA (RAD) linkage map in barley. BMC Genomics, 12: 1-13

Committee on Taxonomy of Marine Mammals. 2018. List of marine mammal species and subspecies. Available from https://www.marinemammalscience.org/species-information/list-marine-mammal-species-subspecies

da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, Zepeda-Mendoza ML, Campos PF. 2016. Next-generation biology: sequencing and data analysis approaches for non-model organisms. Marine Genomics, 30: 3-13

Davidson R, Vachaspati P, Mirarab S, Warnow T. 2015. Phylogenomic species tree estimation in the presence of incomplete lineage sorting and horizontal gene transfer. BMC Genomics doi: 10.1186/1471-2164-16-S10-S1

Desluc F, Brinkmann H, Philippe H. 2005. Phylogenetics and the reconstruction of the three of life. Nat Rev Genet. 6:361-375

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Test for ancient admixture between closely related populations. Mol Biol Evol. 28:2239-2252

Duran C, Appleby N, Edwards D, Batley J. 2009. Molecular genetic markers: discovery, applications, data storage and visualisation. Curr Bionformatics, 61: 16–27

Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1-19

Edwards DL, Knowles LL. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? Proc R Soc Biol Sci. 281:20132765

Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM, Lemmon AR et al. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol Phylogenet Evol. 94:447-462

Ekblom R, Galindo J. 2010. Applications of next-generation sequencing in molecular ecology of non-model organisms. Heredity, 107: 1-15

Ekblom R, Wolf JBW. 2014. A field guide to whole-genome sequencing, assembly and annotation. Evolutionary Applications, 7: 1026-1042

Figueiró HV, Li G, Trindade FJ, Assis J, Pais F, Fernandes G, Santos SHD, Huges GM, Linderoth T, Bi K et al. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci Adv. 3:e1700299

Foote AD, Liu Y, Thomas GWC, Vinar T, Alföldi J et al. 2015. Convergent Evolution of the genomes of marine mammals. Nature Genetics, 47: 272-275

Fujita T, Barraclough TG. 2013. Delimiting species using single locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. Syst Biol. 62:707-724

Gardner AL, Robbins CB. 1999. *Arctocephalus* F. Cuvier, 1826 and *Callorhinus* Gray, 1859 (Mammalia, Pinnipedia): proposed conservation by the designation of *Phoca pusilla* Schreber, [1775] as the type species of *Arctocephalus*; and *Otaria* Péron, 1816 and *Eumetopias* Gill, 1866: proposed conservation by the designation of *Phoca leonina* Molina, 1782 as the type species of Otaria. Bull Zool Nomen. 56:136-141

Graham CF, Glen TC, McArthur AG, Boreham DR, Kieran T, Lance S, Manzon RG, Martino JA, Pierson T, Rogers SM, Wilson JY, Somers, CM. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Molecular Ecology Resources, 15: 1304-1315

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. Nat Genet. 43:1031-1034

Higdon JW, Bininda-Emonds ORP, Beck RMD, Ferguson SH. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evol Biol. 7:216

Hills DM, Huelsenbeck JP, Cunningham CW. 1994. Application and accuracy of molecular phylogenies. Science 264:671-677

Hoglund J. 2009. Evolutionary Conservation Genetics. Oxford University Press, New York

Holder MT, Anderson JA, Holloway AK. 2001. Difficulties in detecting hybridization. Syst Biol. 50:978-982

Humble E, Dasmahapatra KK, Martinez-Barrio A, Gregório I, Forcada J, Polikeit A, Goldsworthy SD, Goebel ME, Kalinowski J, Wolf JBW, Hoffman JI. 2018. RAD sequencing and a hybrid fur seal genome assembly reveal rapidly decating linkage disequilibrium, global population structure and evidence for inbreeding. G3 (Bethesda): 2709-2722

Jones G. 2019. Divergence estimation in the presence of incomplete lineage sorting. Syst Biol. 68:19-31

Kess T, Gross J, Harper F, Boulding EG. 2016. Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. Journal of Moluscan Studies, 82: 104-109

King J. 1954. The otariid seals of the Pacific coast of America. Bull Brit Mus Zool. 2:311-337

King J. 1969. The identity of the fur seals of Australia. Aust J Zool. 17:841-853

King, J. 1983. Seals of the world. 2nd edition. Cornell University Press, Ithaca, NY

Kumar R, Kumar V. 2018. A review of phylogeography: biotic and abiotic factors. Geology, Ecology, and Landscapes, 4: 268-274

Leaché AD, Harris RB, Rannala B, Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst Biol. 63:17-30.

Liu L, Yu L, Kubatko L, Pearls DK, Edwards SV. 2009. Coalescent methods for estimating phylogenetic trees. Mol Phylogenet Evol. 53:320-328.

Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species tree under coalescent model. BMC Evol Biol. 10:302.

Liu L, Xi Z, Wu S, Davis CC, Edwards SV. 2015. Estimating phylogenetic trees from genome-scale data. Annals of the New York Academy of Sciences, doi 10.1111/nyas.12747

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. 2016. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Res. 17: 142-152

Maddison WP. 1997. Gene trees in species trees. Syst Biol. 46:523-536

Manthey JD, Campillo LC, Burns KJ, Moyle RG. 2016. Comparison of Target-capture and Restriction Site Associated DNA sequencing for phylogenomics: a test in Cardinalid tanagers (Aves: Genus *Piranga*). Syst Biol. doi: 10.1093/sysbio/syw005

McVay JD, Carstens BC. 2013. Phylogenetic model choice: justifying a species tree or concatenation analysis. J Phylolgenet Evol Biol. 1:114

Misof B, Liu S, Meusemann K, Donath A, Mayer C et al. 2014. Phylogenomics resolves the timing and pattern of insect Evolution. Science, 7: 763-767

Nacir Y, Linder HP. 2015. Species delimitation and relationships: the dance of seven veils. Taxon 64:3-16

Nyakatura K, Bininda-Emonds ORP. 2012. Updating the evolutionary history of Carnivora (Mammalia): A new species-level supertree complete with divergence time estimates. BMC Biol. 10:12. doi:10.1186/1741-7007-10-12

Oliveira L, Hoffman JI, Hingst-Zaher E, Majluf P, Muelbert MMC, Morgante JS. 2008. Morphological and genetic evidence for two evolutionarily significant units (ESUs) in the South American fur seal, *Arctocephalus australis*. Conserv Genet. 9:1451-1466

Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. Mol Biol Evol. 5:568-83

Pease JB, Hahn MW. 2015. Detection and Polarization of introgression in a five-taxon-phylogeny. Syst Biol. 64:651-662.

Peterson BK, Webber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model-species. PLoS ONE, 7:e37135

Philippe H, Blanchette M. 2007. Overview of the first phylogenomics conference. BMC Evolutionary Biology, 7(suppl. 1): 1-4

Repenning CA, Peterson RS, Hubbs CL. 1971. Contributions to the systematics of the southern fur seals, with particular reference to the Juan Fernández and Guadalupe species. Antar Res S. 18:1-34

Reynolds JE, Odell DK, Rommel SA. 1999. Marine mammals of the world. Melbourne University Press, Melbourne

Rheindt FE, Fujita MK, Wilton PR, Edwards SV. 2014. Introgression and phenotypic assimilation *Zimmerius* flycatchers (tyranidae): population genetic and phylogenetic inferences from genome-wide SNPs. Syst Biol. 63:134-152

Riedman M. 1990. The Pinnipeds: Seals, Sea Lions, and Walruses. University of California Press, Berkley. 439 p.

Rodrigues P, Seguel M, Gutiérrez J, Pavés HJ, Verdugo C. 2018. Genetic connectivity of the South American fur seal (*Arctocephalus australis*) across Atlantic and Pacific oceans revealed by mitochondrial genes. Aquat Conserv Mar Fresh Ecosyst. 28:315-323

Rowe HC, Renaut S, Guggisberg A. 2011. A RAD in the realm of next-generation sequencing Technologies. Molecular Ecology, 20: 3499-3502

Sang T, Zhong Y. 2000. Testing hybridization hypothesis based on incongruent gene trees. Syst. Biol. 49:422-434

Shafer AB, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW. 2016. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution, 8: 907-917

Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc Nat Acad Sci USA. 109:14942-14947

Springer MS, Gatesy J. 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94:1-33

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analysis with thousands of taxa and mixed models. Bioinformatics 22:2688-2690

Suchan T, Pitteloud C, Gerasimova NS, Kostikova A, Schmid S, Arrigo N, Pajkovic M, Ronikier M, Alvarez N. Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens. PLoS One, 11: e0151651

Szöllosi G, Tannier E, Daubin V, Boussau B. 2015. The inference of gene trees with species trees. Syst Biol. 64:042-062

Than C, Ruths D, Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Informatics 9:322

Tonini J, Moore A, Stern D, Shcheglovitova M, Ortí G. 2015. Concatenation and Species Tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Currents Tree of Life. doi: 10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be

Wen D, Nakhleh L. 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. Syst Biol. 67:439-457

Wynen LP, Goldsworthy SD, Insley SJ, Adams M, Bickham JW, Francis J, Gallo JP, Hoelzel AR, Majluf P, White RWG et al. 2001. Phylogenetic relationships within eared seals (Otariidae: Carnivora): implications for the historical biogeography of the family. Mol Phylogenet Evol. 21:270-284

Xi Z, Liu L, Davis CC. 2014. Coalescent versus concatenation methods and the placement as *Amborella* as sister to water lillies. Syst Biol. 63:919-932

Yonezawa T, Kohno N, Hasegawa M. 2009. The monophyletic origin of sea lion and fur seals (Carnivora: Otariidae) in the southern hemisphere. Gene 441:89-99

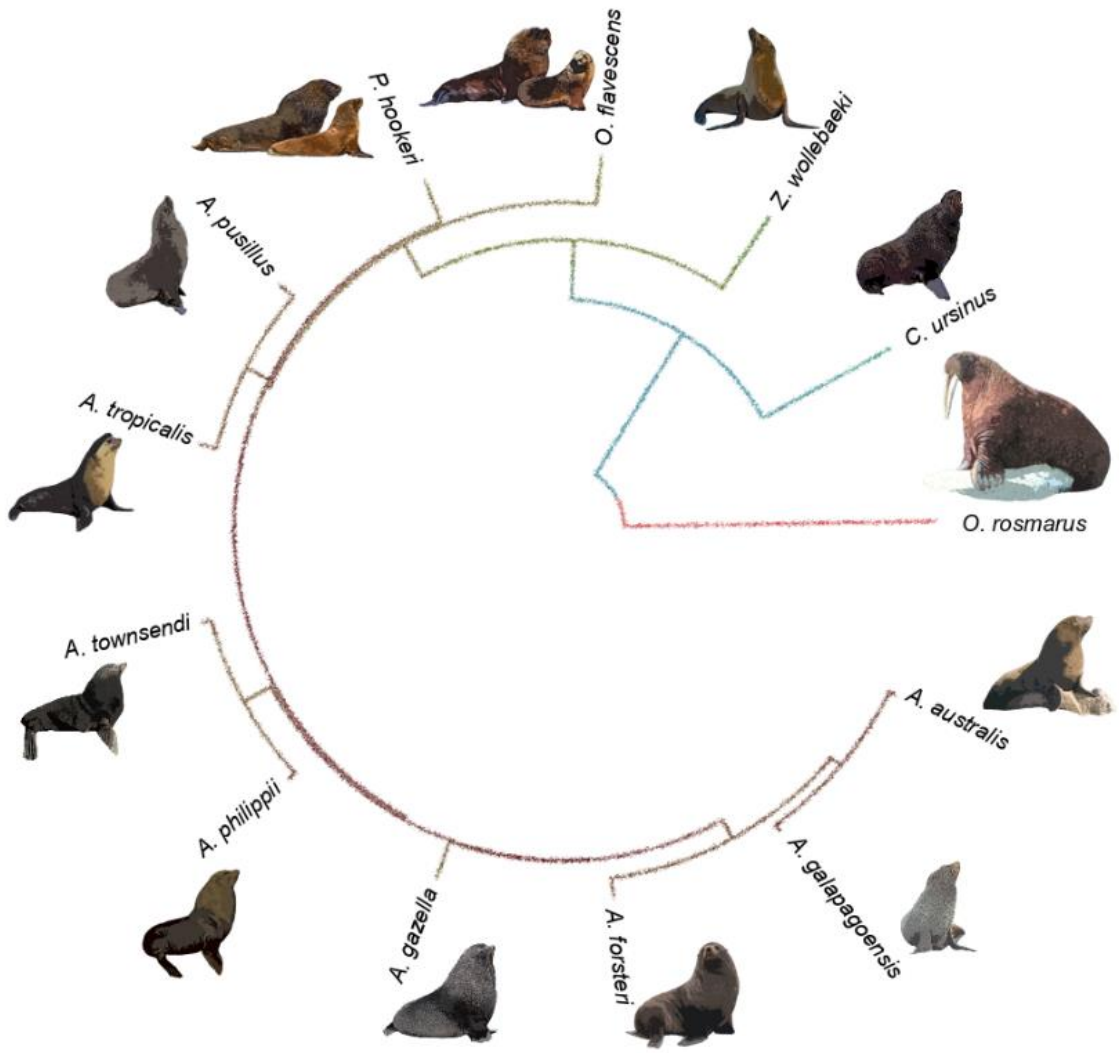
Zhang C, Ogilvie HA, Drummond AJ, Stadler T. 2017. Bayesian inference of species networks from multilocus sequence data. Mol Biol Evol. 35:504-517

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19 (suppl 6):153

Título:

Phylogenomics of Otariidae reveal an explosive radiation at the colonization of the Southern Hemisphere and massive Incomplete Lineage Sorting

Artigo Científico abordando a filogenia de Otariidae com base em genomas completos e formatado conforme as normas do periódico *Systematic Biology*.



Running head

Phylogenomics of Fur Seals and Sea Lions

Phylogenomics of Otariidae reveal an explosive radiation at the colonization of the Southern Hemisphere and massive Incomplete Lineage Sorting

Authors

Fernando Lopes[1*], Larissa Oliveira[2,3], Amanda Kessler[1], Enrique Crespo[4], Susana Cárdenas-Alayza[5], Patricia Majluf[5], Maritza Sepúlveda[6], Robert L. Brownell Jr.[7], Valentina Franco-Trecu[8], Diego Páez-Rosas[9], Jaime Chaves[9], Carolina Loch[10], Bruce Robertson[11], Karina Acevedo-Whitehouse[12], Fernando Elorriaga-Verplancken[13], Stephen P. Kirkman[14], Claire Peart[15], Jochen B. W. Wolf[15], Sandro L. Bonatto[1*]

Affiliations

[1] Escola de Ciências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brasil
[2] Programa de Pós-graduação em Biologia, Universidade do Vale do Rio dos Sinos, São Leopoldo, Brasil
[3] GEMARS, Grupo de Estudos de Mamíferos Aquáticos do Rio Grande do Sul
[4] CONICET, Centro Nacional Patagónico - CENPAT, Puerto Madryn, Argentina
[5] Centro para la Sostenibilidad Ambiental, Universidad Peruana Cayetano Heredia, Lima, Peru
[6] Departamento de Biología y Ciencias Ambientales, Universidad de Valparaiso, Valparaiso, Chile
[7] NOAA, National Oceanic and Atmospheric Administration, La Jolla, United States of America
[8] Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay
[9] Colegio de Ciencias Biológicas y Ambientales, COCIBA, Universidad San Francisco de Quito, Quito, Ecuador
[10] Department of Oral Sciences, University of Otago, Dunedin, New Zealand
[11] Department of Zoology, University of Otago, Dunedin, New Zealand
[12] Facultad de Ciencias Naturales, Universidad Autónoma de Querétaro, Querétaro, Mexico
[13] Instituto Politecnico Nacional (CICIMAR-IPN), Centro Interdisciplinario de Ciencias Marinas, La Paz, Mexico
[14] Department of Environmental Affairs, Oceans and Coasts, Cape Town, South Africa
[15] Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, Münich, Germany

Corresponding authors (*): *slbonatto@pucrs.br; fernando.lopes@edu.pucrs.br*

Phone number: +55(51)33534727

Ipiranga 6681, 12C 134. 90619-900. Porto Alegre, RS, Brazil

Keywords: Phylogenomics, ILS, hybridization, Pliocene, Pleistocene, monophyly

ABSTRACT

The phylogeny and systematics of fur seals and sea lions (Otariidae) have been studied for a long time with the use of diverse types of data. However, despite the increasing size of the molecular data, only a few relationships have reached acceptance, such as the existence of a clade with mostly southern hemisphere species. The phylogenetic relationship between genera and several species of the latter clade is highly discordant between studies and the monophyletism of *Arctocephalus* remains unclear. Divergence times in the group vary largely between studies. This lack of resolution also hindered the understanding of the biogeographical history of this taxon, such as when and how the trans-equatorial dispersal occurred. Here we used high-coverage whole genomes for 13 of the 15 species of Otariidae to unravel its phylogeny and its bearing on the taxonomy and biogeographical history. Despite extreme topological discordance among gene trees, we found a fully supported species tree that agrees with the well-accepted relationships and resolved all uncertainties still prevalent, such as the monophyly of *Arctocephalus*. We found a relatively recent trans-hemispheric dispersal that originated the southern clade and diversified in rapid succession between 3 to 2.5 Mya. The first emerged species was *Otaria* followed by *Phocarctos* and then by four lineages of *Arctocephalus*. We found that this quasi-simultaneous speciation within the southern clade led to an extensive incomplete lineage sorting throughout the genomes and resulted in a high level of genealogical discordance, explaining the incongruence among and within prior phylogenetic studies of the family.

INTRODUCTION

For decades it was assumed that simply adding more molecular data (more genes) we would, at some point, find perfect phylogenies even with simple phylogenetic methods (Rokas et al. 2003; Faircloth et al. 2013; Hoban et al. 2013; McCormack and Faircloth 2013). Studies using whole genome data, however, found that inference of the true species tree, if such a thing exists, may be extremely challenging for some parts of the tree of life (Nakhleh 2013).

For example, the studies that partitioned genomic data into smaller regions (e.g., genes or independent genomic fragments) found a high degree of genealogical discordance among gene trees (Peter 2016; Harris and DeGiorgio 2016; Elworth et al. 2018; Jones 2019). Theoretical and empirical studies have shown that these incongruences may have three main causes. First, incorrect estimation of the gene trees (e.g., caused by reduced phylogenetic information). Secondly, by incomplete lineage sorting (ILS), found when ancestral polymorphism is persistent between successive speciation events (see Maddison and Knowles 2006; Oliver 2013). Third, by introgression between lineages (hybridization) (e.g., Rheindt et al. 2014; Figueiró et al. 2017; Zhang et al. 2017).

Almost all phylogenetic methods to date assume only ILS, ignoring the consequences of hybridization in the phylogenetic reconstruction (Stamatakis 2014; Drummond and Bouckaert 2015). Despite this, simulated data have shown problems for recovering the correct topologies, node heights and population sizes under multispecies coalescent models when migration between non-sister species is present but ignored (Leaché et al. 2014; Wen and Nakhleh 2018). Consequently, resolving relationships among fast-radiated taxonomic groups that putatively underwent both ILS and hybridization has proven challenging (Chakrabarty et al. 2017; Esselstyn et al. 2017; Reddy et al. 2017).

Otariidae (fur seals and sea lions) is composed by 15 species, according to the Committee on Taxonomy of the Society for Marine Mammalogy (Committee on Taxonomy 2018), although there are uncertainties for the specific status for some of them, such as *A. philippii* and *A. townsendi*. This

family occurs in the Northern Pacific Ocean and Southern Hemisphere, where it is widely distributed from tropical to polar regions (Churchill et al. 2014; Berta et al. 2018). Although the systematics and phylogeny of the group have been extensively discussed for over two centuries (Sclater 1897; Wynen et al. 2001; Deméré et al. 2003; Árnason et al. 2006; Yonezawa et al. 2009; Berta and Churchill 2012; Nyakatura and Bininda-Emonds 2012; Churchill et al. 2014; Berta et al. 2018), several relationships, in particular within *Arctocephalus*, the most diverse Otariidae genus, remain unclear (Yonezawa et al. 2009; Berta and Churchill 2012). For example, older morphology-based studies usually suggested the grouping of *C. ursinus* and *Arctocephalus* in Arctocephalinae, encompassing fur seals and characterized by small body size and thick pelage, and all other species in Otariinae, encompassing sea lions and characterized by larger body sizes and reliance on blubber rather than fur for thermal insulation (Berta and Demeré 1986; see review in Berta et al. 2018). On the other hand, more recent published phylogenies, that mostly applied few concatenated mitochondrial or nuclear genes in supermatrices (e.g., Yonezawa et al. 2009), combined data source (mtDNA, nDNA and morphology e.g., Churchill et al. 2014) or combined phylogenies (e.g., Berta and Churchill et al. 2012; Nyakatura and Bininda-Emonds 2014; Berta et al. 2018) usually do not support the subfamilies. On the contrary, the Southern Hemisphere otariids (i.e., *Otaria*, *Neophoca, Phocarctos*, and *Arctocephalus*) were grouped in the so-called southern clade, which is considered the sister clade of the sea lions of the Northern Hemisphere (i.e., *Zalophus* and *Eumetopias*) (Yonezawa et al. 2009, Churchill et al. 2014).

Another important discordance between studies is related to the monophyly of *Arctocephalus*. A combined phylogeny based on published morphological and molecular data found *Arctocephalus sensu lato* as paraphyletic (Berta and Churchill 2012). They proposed *Arctocephalus* to be limited to *Arctocephalus pusillus*, the type species, and the transference of the remaining species to *Arctophoca*. Otherwise, a Carnivora supertree study suggested that the use of *Arctophoca* could be premature due to the uncertainty about phylogenetic relationships (Nyakatura and Bininda-Emonds 2014). The Committee on Taxonomy of Marine Mammals that firstly supported the separation of *Arctocephalus*

and *Arctophoca* presently suggest the conservative use of *Arctocephalus sensu lato* for all southern fur seals pending further studies (Committee on Taxonomy 2018).

The divergence times and the biogeography of the group also present similarities and some differences between studies, the latter particularly caused by the differences in the phylogenetic trees. The most recent biogeographical studies usually agree on a North Pacific origin of Otariidae and support one main trans-equatorial dispersal towards the Southern Pacific Ocean, where a southern clade emerged. This dispersal and the diversification of the southern clade have been estimated between ~7 and 6 Mya (Yonezawa et al. 2009, Churchill et al. 2014, Berta et al. 2018). The main period of diversification within *Arctocephalus* has been proposed to have occurred around 4-3 Mya (Nyakatura and Bininda-Emonds 2012) or recently as <1 Mya (Berta et al. 2018).

Although there is a better understanding of the relationships of basal Otariidae and how they reached South Hemisphere, it is still unclear the events of speciation in this portion of the globe (Berta et al. 2018). In this study, we inferred the relationships and estimated the divergence times of 13 species of fur seals and sea lions by using whole genome sequence data and several different phylogenetic approaches aiming to clarify the Otariidae evolutionary history. Here, we also considered topological incongruences caused by incomplete lineage sorting or introgression and their impacts on our results.

MATERIAL AND METHODS

Data collection and Genome Sequencing

A total of nine tissue samples were collected from skin biopsies from pups, or fresh carcasses found ashore. Piglet ear notch pliers were used to collect ~0.5 cm^3 from skin biopsies. The samples were stored in ethanol 70% and cryo-preserved in -20 ºC freezer. Genomic DNA extractions were carried out with DNeasy Tissue Kit (Qiagen) following the standard protocol.

We sequenced the whole genome of one individual from each of the seven species of *Arctocephalus* and two other monospecific genera (*Phocarctos* and *Otaria*) (Table 1). Genomic

libraries were prepared with Illumina DNA PCR-free or TruSeq Nano kits with an insert size of 350 bp, and two libraries were sequenced (PE150) per lane on the Illumina HiSeq X platform. Furthermore, genome raw reads from *Arctocephalus gazella* (SRR2658542-61 - Humble et al. 2016), *Zalophus wollebaeki* (SRR4431580-81 - Shafer et al. 2016), and the outgroup *Odobenus rosmarus* (ANOP00000000 - Foote et al. 2015) were retrieved from NCBI Short Reads Archive (https://www.ncbi.nlm.nih.gov/sra). Since we had already started several analyses when the *Callorhinus ursinus* genome raw reads (SRR7278673) were made available, we did not include it in some less critical but time-consuming analyses.

Our study then contains 12 of the 15 extant Otariidae species (all *Arctocephalus*, *Phocarctos*, *Otaria*, *Zalophus*, and *Callorhinus*). The three species left presented well-resolved positions in Otariidae phylogenies published in the last decade (see Yonezawa et al. 2009; Berta and Churchill 2012; Nyakatura and Bininda-Emonds 2012; Berta et al. 2018): *Z. californianus*, the sister species of Galapagos sea lion (*Z. wollebaeki*), considered subspecies until very recently, *Neophoca cinerea*, the sister species of *P. hookeri*, and *Eumetopias jubatus* that is basal to *Zalophus* spp.

Table 1. Species sampling information.

Species	Common name	Localities	Geographical coordinates	GeneBank (SRR)	Source
Arctocephalus australis	South American fur seal	Provincia de Chubut (Argentina)	45°06'S, 65°24'W	-	Current study
Arctocephalus galapagoensis	Galapagos fur seal	Isla Santiago (Ecuador)	0°14'S, 90°51'W	-	Current study
Arctocephalus forsteri	New Zealand fur seal	Otago Harbour (New Zealand)	45°49'S, 170°39'E	-	Current study
Arctocephalus philippii	Juan Fernández fur seal	Juan Fernández Islands (Chile)	33°46'S, 80°46'W	-	Current study
Arctocephalus townsendi	Guadalupe fur seal	Corral Canyon Beach (USA)	34°01'N, 118°01W	-	Current study
Arctocephalus pusillus	Cape fur seal	Kleinsee (South Africa)	29°33'S, 17°00'E	-	Current study
Arctocephalus tropicalis	Subantarctic fur seal	Coast of Rio Grande do Sul (Brazil)	29°59'S, 50°08'W	-	Current study
Phocarctos hookeri	New Zealand sea lion	Otago Peninsula (New Zealand)	45°50'S, 170°43'E	-	Current study
Otaria flavescens	South American sea lion	Punta San Juan (Peru)	15°22'S, 75°11'W	-	Current study
Arctocephalus gazela	Antarctic fur seal	N.A (Antarctic continent)	N.A.	2658542-61	Humble et al. 2018
Zalophus wollebaeki	Galapagos sea lion	Galapagos Archipelago (Ecuador)	0°14'S, 90°51'W	4431580-81	Shafer et al. 2016
Callorhinus ursinus	Northern fur seal	N.A.	N.A.	7278673	Árnason et al. 2006

Sequencing quality control was performed using FastQC (Andrews 2010). Reads were trimmed for vestigial adapters, mapped against *O. rosmarus* genome and locally realigned using the bam_pipeline implemented on PALEOMIX 1.2.13.2 (Schubert et al. 2014) following the parameters: reads with length-size < 100 bp and Phred-score < 30 were filtered out by AdapterRemoval v2 (Schubert et al. 2016); the remaining paired-end reads were mapped using BWA 0.7.17 (Li and Durbin 2009) and -mem algorithm; paired-end reads with mapping quality Phred-score < 20, unmapped reads and single-reads were discarded from the downstream pipeline; reads that were sequenced more than two or less than one standard deviation from the average of coverage of each genome (Supplementary Table S1) were not considered and were removed (Arnold et al. 2013; Gautier et al. 2013). PCR duplicates were detected and removed by Picard Tools 2.18.5 (broadinstitute.github.io/picard/) and miscalling indels were locally realigned by GATK 3.8 (McKenna et al. 2010).

Consensus, Alignments and SNP Calling

Consensus sequences of all genomes were generated and aligned per scaffold with ANGSD 0.921 (Korneliussen et al. 2014) using the parameters *doFasta 2*, *doCounts1,* and *explode 1*. Single-nucleotide polymorphisms (SNPs) were called following the filters: *uniqueOnly 1*, *remove_bads 1*, *only_proper_pairs 1*, *C 50*, *baq 1*, *setMinDepth* 120, *setMaxDepth 1200*, *setMinDepthInd 5*, *setMaxDepthInd 100*, *doCounts 1*, *GL 1*, *doMajorMinor 1*, *SNP_pval 1e-3*, *doGeno 32*, *doPost 1*, *doPlink2*. After the SNP calling, the generated PLINK variant panel was converted to VCF format with Plink 1.9 (Chang et al. 2015). We removed repetitive, coding, and transposons using BEDTools 2.27.0 maskfasta option (Quinlan and Hall 2010) and the General Feature Format File of the *O. rosmarus* genome.

Phylogenetic information, phylogenomic analyses, and species trees estimation

For some phylogenetic analyses, per-scaffold alignments were partitioned into sets of nonoverlapping genomic fragments (GFs) of 10, 20, 50, 80, 100, and 200 kilobases (kb). To reduce the effect of linkage disequilibrium between GFs, they were separated by 100 kb, following Humble

et al. (2018). Scaffolds smaller than the GF partition size were excluded, and sites with more than 20% of missing data (i.e., not sequenced in more than three species in the alignment) were removed with trimAl v1.4 (Capella-Gutierrez et al. 2009). After these filters, alignments smaller than half of the original alignment size were discarded. To reduce the effect of intra-fragment genetic recombination on the recovered topologies, we used the software 3Seq on *full run mode* (Lam et al. 2017) to detect likely recombinant sequences. We removed the alignments with *p-value* < 0.01 after Bonferroni correction. To assess the amount of genetic information content on GFs, we randomly sampled 10,000 GFs of 50 kb and used the AMAS tool (Borowiec 2016) to count the number of parsimony informative sites in the alignments and to estimate the genetic distances between two closely related fur seals (*A. australis* and *A. galapagoensis*).

First, we estimated species phylogenies using large amounts of sequence data. We used the VCF2Dis script (github.com/BGI-shenzhen) to estimate the p-distance matrix for the whole genome. With the matrix, we estimated a neighbor-joining tree in the software PHYLIP 3.697 (Felsestein 1989). We also estimated maximum-likelihood (ML) trees for the ten largest scaffolds with RAxML-HPC-PTHREADS 8.2 (Stamatakis 2014) using GTR+G as best fit substitution model, estimated by JModelTest2 (Darriba et al. 2012), and 100 bootstrap replicates.

Next, we estimated ML trees with RAxML as above for each GF in all GF partitions (10 to 200 kb). For each set of GF trees, we count the frequency of each topology with Newick Utilities 1.1 (Junier and Zdobnov 2010), a suite of Unix shell tools for processing phylogenetic trees. The GFs ML trees were also used to estimate a maximum quartet support species tree (exact search) with the multispecies coalescent model (MSC) of ASTRAL-III (Zhang et al. 2018).

We also estimated the species tree and the divergence times with the Bayesian Inference method StarBEAST2 implemented in BEAST 2.5.2 package (BEAUti, BEAST, TreeAnotator - Rambaut and Drummond 2010; Bouckaert et al. 2014; Ogilvie et al. 2017). Since this Bayesian analysis is very time-consuming and that the ASTRAL species trees of all GF datasets, except the 10 kb GF, were identical (see Results), we used 300 randomly selected GFs of the 50 kb dataset only. The main priors

used were: linked clock models, constant population sizes, HKY substitution model with empirical base frequencies, an estimated six gamma categories site model and Yule tree model. To estimate divergence times, we used as prior a strict molecular clock with a log-normal distribution and standard mammalian genomic mutation rate of 1×10^{-8} bp^{-1} gen^{-1} (Kumar and Subramanian 2001) with large standard deviation of 0.4 (5% and 95% quantiles of 4×10^{-9} and 4×10^{-8} bp^{-1} gen^{-1}, respectively) to account for other values found in the literature. We also added two calibration points in the phylogeny. One at the origin of *Arctocephalus* spp. clade + *P. hookeri*, based on the age of the oldest *Arctocephalus* fossil record (*Arctocephalus* sp. nov. - Varswater Formation of South Africa) and that constrained the origin of this group to a lower bound of 2.7 Mya (Avery and Klein 2011). We also set the date of the root as a normal prior with a mean of 20 Mya (± 3.0) in the divergence between Otariidae and Odobenidae (Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012). We ran a Bayesian Markov Chain Monte Carlo (MCMC) of 500,000,000 steps sampled each 20,000 with a burn-in of 10%. To reduce the effect of underestimation of the internal branches induced by undetected hybridizations in our dataset (Leaché et al. 2014, Elworth et al. 2019), we also estimated a StarBEAST2 species tree using only the GF of 50 kb whose ML tree topology was identical to our species tree (as estimated by ASTRAL-III and the previous StarBEAST analyses, see results), with the same parameters as above. We checked the MCMC runs with Tracer 1.7 (Rambaut and Drummond 2007).

As an additional estimation of divergence times, the species tree recovered by both ASTRAL-III and StarBEAST2 was used as input in the Bayesian species tree estimation of the BP&P program (Ziheng 2015; Flouri et al. 2018) using the same 300 GFs of 50 kb applied in the initial StarBEAST analysis. We use the following the parameters: an MCMC chain of 2,000,000 replicates with burn-in of 200,000, with theta prior of 0.01 and tau prior of 0.02. The theta prior specifies the inverse-gamma prior, the number of differences per kb, and the tau specifies the divergence time parameter for the root. For this analysis, the divergence times were calibrated based on the age of the root as above (Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012).

All trees were visualized and edited for clarity on FigTree 1.4.4 (Rambaut 2017), DensiTree 2 (Bouckaert and Heled 2014) or Dendroscope 3 (Huson and Scornavacca 2012).

Gene trees simulation based on the species tree

To test if the high topological discordances between the GFs could be explained by ILS alone, we simulated gene trees under a multispecies coalescent framework using as input our species tree estimated by StarBEAST2, that included branch lengths and effective population sizes. We simulated 10,000 gene trees with the function sim.coaltree.sp in the R phylogenetic package Phybase (Liu and Yu 2010) and calculated the linear Pearson's correlation between their frequency distribution with our empirical topology frequency (similar to Wang et al. 2018).

Mitochondrial Genome Phylogeny

We obtained the mitochondrial genomes of the fur seals and sea lions by mapping the reads with PALEOMIX 1.2.13.2 against the mitochondrial genome of *A. townsendi* retrieved from GenBank (NC008420). The Bayesian phylogenetic tree was reconstructed with BEAST 2.5.2 package following the parameters: HKY substitution model, with six gamma categories, estimated base frequencies and Yule Tree Model; Uncorrelated Lognormal Clock Model with a mean substitution rate of 2% site^{-1} million year^{-1}, a log-normal distribution and 0.9 of standard deviation; 10,000,000 MCMC steps sampled each 1,000; and the calibration point in the origin of *P. hookeri* and *Arctocephalus* spp., the same used in the previous StarBEAST2 estimation.

Introgression between species

For these analyses, we did not use the two earliest diverged species (*C. ursinus* and *Z. wollebaeki*). To estimate possible introgression between the species, we calculated f_3-statistics and f_4-statistics based on the panel of SNPs in threepop and fourpop modules, respectively, of *TreeMix* package (Pickrell and Pritchard 2012; Harris and DeGiorgio 2016). Threepop explicitly tests whether a taxon of interest *A* is the result of admixture between two other *B* and *C*: positive values indicate no evidence of admixture among tested species, while negative values indicate admixture. Fourpop

estimates unrooted four-population phylogenies to visualize shared genetic drift among taxa. For a f_4 ((A,B),(C,D)) topology, the defined paths are *A* to *B* and *C* to *D*, this results in $f_4 = 0$. If f_4 is a positive value, this indicates introgression from *A* to *C* and/or from *B* to *D*. Otherwise, a negative value indicates introgression from *A* to *D* and/or *B* to *C*. Significant f_4 values may also be interpreted as a rejection of the given phylogenetic relationship between the four taxa (Peter 2016; Zhenge and Janke 2018)). The significance of f_3 and f_4-statistics is based on Z-score and was calculated over 872 jackknife blocks of 50,000 SNPs. Significantly positive ($Z > 3$) and significantly negative ($Z < -3$) values reject the null hypothesis. We plotted the distribution of f_4-values with the function f4stats of the admixturegraph (Leppälä et al. 2017), an R package.

RESULTS

We successfully sequenced the whole genome of nine fur seals and sea lions, and including the three previously published genomes, we recovered an average of 83% percent of the walrus reference genome with an average coverage of 24X (± 7X) (see Supplementary Table S1). The largest scaffold is 126.48 million bases (Mb) and the ten largest summed around 845 Mb, ~35% of the whole genome (2.4 Gb - Fig. 1). BEDtools maskfasta identified ~40% of the reference genome as repetitive regions that were masked, remaining a high-quality alignment of ~1.1 Gb for further analyses. After all filtering (removing the masked regions, missing data, GFs with less than 50% of the original information, and those with signal of intra-locus recombination), we obtained between 14,012 (with 10 kb) and 5,701 (with 200 kb) GFs 100 kb apart each other (Table 2) for the gene trees analyses.

Whole Genome Phylogenetic Inferences

The phylogenetic trees were highly concordant across different methods. The Bayesian species tree (estimated with StarBEAST2 using 300 GF of 50 kb) (Fig. 1), the ASTRAL-III species trees (from thousands of ML trees from GFs from 20kb to 200 kb) (Supplementary Fig. S1), the ML trees of eight of the ten largest scaffolds (Supplementary Fig. S2) and the NJ tree (estimated using the genetic distances among the whole genomes) (~1 Gb, Supplementary Fig. S3) resulted in the same

tree topology with high support (most or all branches presented full support), hereafter named as the Otariidae species tree.

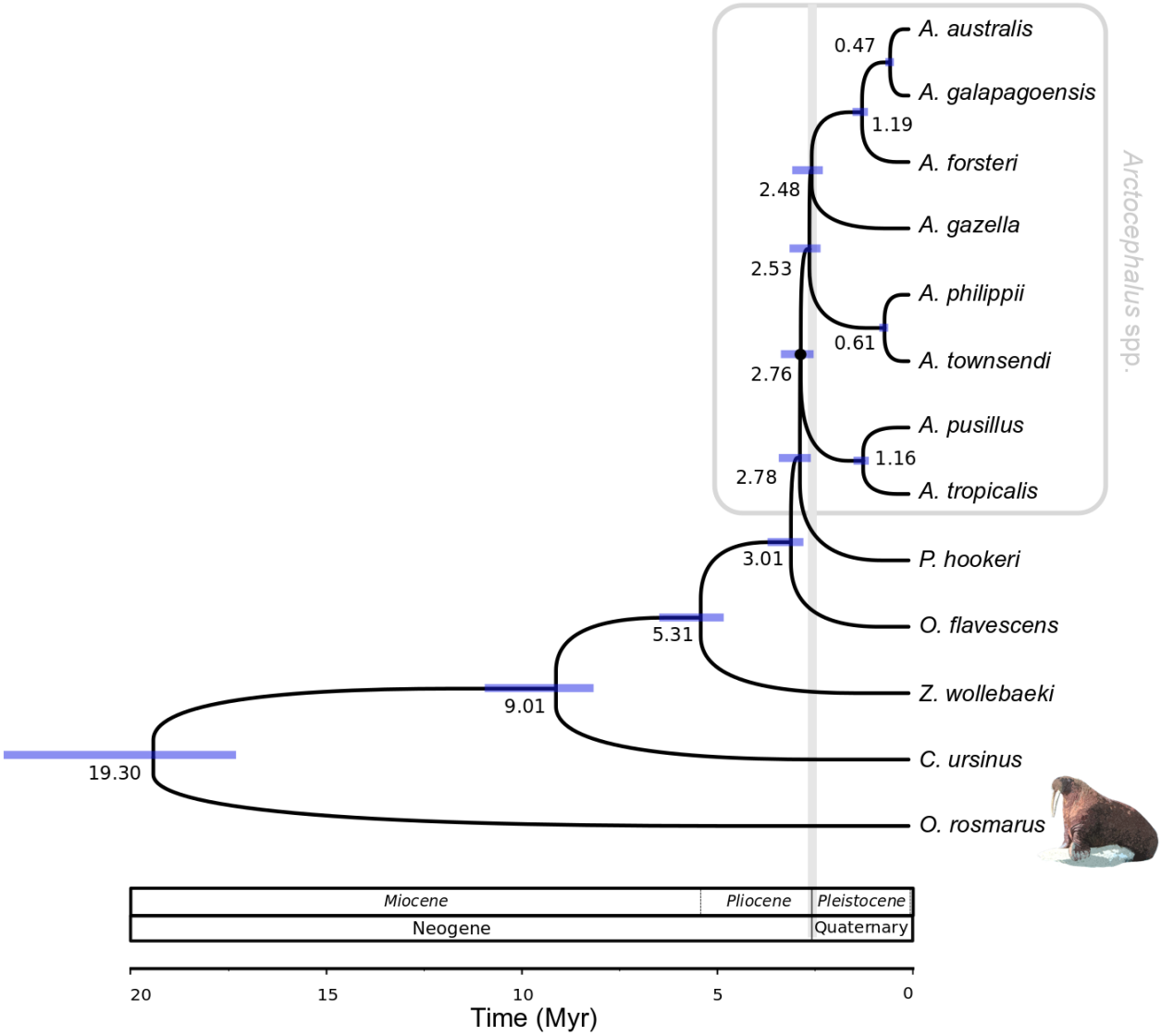


Figure 1. Calibrated species tree and the divergence times using the Bayesian inference of StarBEAST2 and 300 GF of 50 kb. The transition between Pliocene and Pleistocene is highlighted with a gray bar. Blue bars represent the divergence time confidence interval of 95%. Almost all nodes with HPD = 1, the exception is the *Arctocephalus* node (HPD = 0.74), shown as a black dot in the phylogeny.

This species tree supported the monophyly of *Arctocephalus*, that is very close to *P. hookeri* and *O. flavescens*, with *Z. wollebaeki* more distantly related and *C. ursinus* in a basal position in Otariidae. Within *Arctocephalus*, there are four main lineages, following this sequence: *A. pusilius+A. tropicalis*, *A. phillippii+A. townsendi*, *A. gazella*, and the clade *A. forsteri+A. galapagoensis+A. australis*. Three alternative topologies were found in these analyses, one in which *P. hookeri* and the *A. tropicalis+A. pusillus* clade switched position (ASTRAL-III with GFs 10 kb,

Supplementary Fig. S1) and two in which *A. gazella* was found in two different positions within *Arctocephalus* (found in two ML scaffold trees) (Supplementary Fig.S2).

The mitochondrial genome phylogeny (Fig. 2 and Supplementary Fig. S4) is also very similar to the nuclear genome species tree, presenting high posterior probabilities for most nodes, with two differences: (1) the switched position between *P. hookeri* and the *A. tropicalis+A. pusillus* clade, as in the ASTRAL species tree of 10 kb GFs; 2) the sister relationship of *A. australis* with *A. forsteri*, instead of with *A. galapagoensis*.

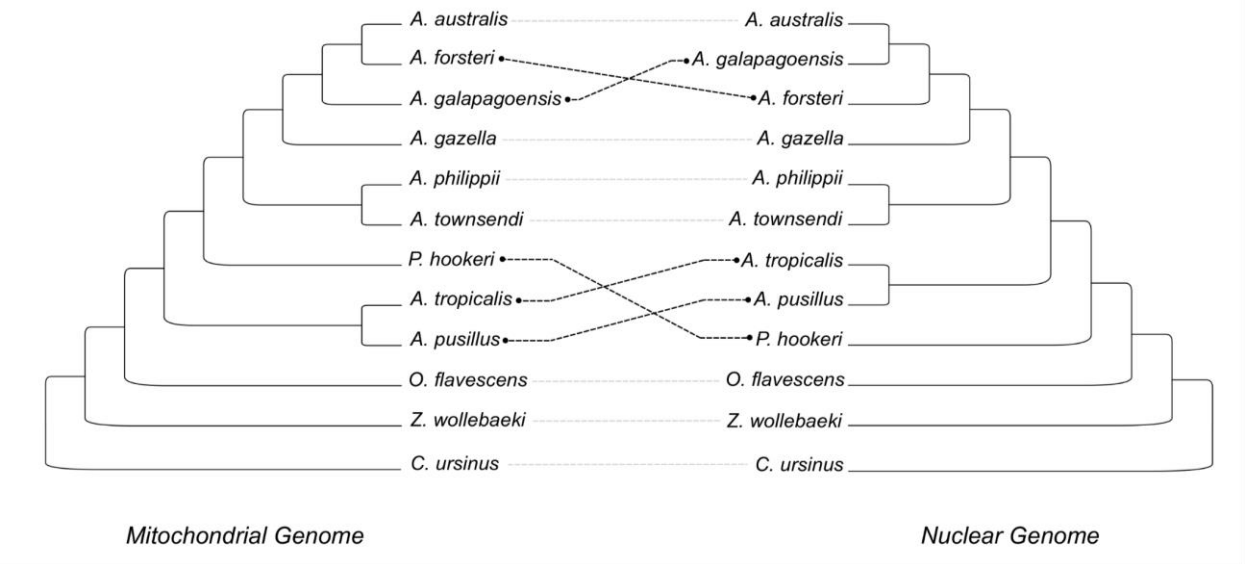


Figure 2. Comparison of mitochondrial and nuclear genomes topology.

The species tree divergence times estimated with StarBEAST2 and BP&P were very similar (Fig. 1 and Supplementary Fig. S5). The divergence between the Walrus and the Otariidae is 19.3 Mya (C.I. 95% = 17.2 - 23.1), and within the latter, the *C. ursinus* diverged ~9 Mya Mya (C.I. 95% = 8.0 – 10.8) followed by the *Z. wollebaeki* 5.3 Mya (C.I. 95% = 4.7 - 6.3). After that, around the Pliocene to Pleistocene transition, six lineages diverged almost simultaneously (between ~3 Mya and 2.5 Mya), originating in sequence: *O. flavescens*, *P. hookeri*, and the four main *Arctocephalus* lineages described above. Specifically, *Arctocephalus* diversification began ~2.8 Mya (C.I. 95% = 2.4 - 3.3), the divergence times between *A. pusillus* and *A. tropicalis* and between *A. forsteri* and *A. austalis+A. galapagoensis* were very similar, ~1.2 Mya, and the two groups that diverged more recently were *A. phillippii+A. townsendi*, and *A. australis+A. galapagoensis*, at ~0.5 Mya. The main

difference between the StarBEAST2 and the BP&P was that in the latter the divergence of *A. gazelle* was almost simultaneous with *A. forsteri* (~1.2 Mya, Fig. 1 and Supplementary Figure S5). Finally, we assessed if the almost simultaneous divergence time for the six lineages estimated in our species tree could be an artefact caused by the underestimation of divergence times in methods that do not account for introgression. The StarBEAST2 inference using only the 113 50 kb GFs whose ML trees were identical to the species tree and estimated the same topology. The divergence times were almost identical to the general species trees, in special the six nodes related to the explosive radiation (Supplementary Fig. S5), supporting the latter is not an artefact of unaccounted hybridizations.

Genome fragment information content and phylogenetic discordance

When the ML phylogeny of each GF was estimated separately, we found hundreds to thousands different trees in each GF dataset (Supplementary Fig. S6), most of them occurred just once or a few times (that is, were estimated from one or a few GFs). The most frequent topology in the 10 kb dataset occurred in only 55 of the GFs (i.e., in ~0.4% of the GF) and although the frequency of the most common topology in each dataset increased with the increase in size of the GFs, even in the 200 kb data set the most frequent topology was found only ~3.7% (Table 2, Supplementary Fig. S6). Except for the 10 kb dataset, the other five datasets presented the same tree as the most frequent one. Considering the 50 kb GFs as an example, the mean variable sites between the two closest species (*A. australis* and *A. galapagoensis*) is ~40 and the mean parsimony informative sites in the alignments is ~200 (Supplementary Fig S7), suggesting that there is enough variation in all GF sizes to allow the inference of GF gene trees.

Table 2. The ten most frequent Maximum Likelihood GF trees recovered in RaxML and the absolute frequencies of occurrence in the different set of windows sizes.

Order	Topology	10kb	20kb	50kb	80kb	100kb	200kb	SGT
1	(((((((Aaus,Agal),Afor),Agaz),(Aphi,Atow)),(Apus,Atro)),Phoo),Ofla),Zwol),Oros)	45	**79**	**169**	**167**	**170**	**215**	**367**
2	(((((((Aaus,Agal),Afor),Agaz),(Aphi,Atow)),Phoo),((Apus,Atro),Ofla)),Zwol),Oros)	33	52	96	124	137	153	126
3	(((((((Aaus,Agal),Afor),Agaz),(Aphi,Atow)),(Apus,Atro)),(Ofla,Phoo)),Zwol),Oros)	36	43	78	106	95	118	96
4	(((((((Aaus,Agal),Afor),(Aphi,Atow)),(Agaz,(Apus,Atro))),Phoo),Ofla),Zwol),Oros)	41	57	67	82	86	109	42
5	(((((((Aaus,Agal),Afor),(Agaz,(Aphi,Atow))),(Apus,Atro)),Phoo),Ofla),Zwol),Oros)	34	70	97	100	93	100	209
6	(((((((Aaus,Agal),Afor),(Agaz,(Aphi,Atow))),Phoo),(Apus,Atro)),Ofla),Zwol),Oros)	**55**	71	95	80	110	87	343
7	(((((((Aaus,Agal),Afor),Agaz),(Aphi,Atow)),(Apus,Atro)),Ofla),Phoo),Zwol),Oros)	28	46	112	94	108	86	95
8	(((((((Aaus,Agal),Afor),(Aphi,Atow)),Agaz),Phoo),(Apus,Atro)),Ofla),Zwol),Oros)	34	62	69	79	71	82	331
9	(((((((Aaus,Agal),Afor),Agaz),(Aphi,Atow)),((Apus,Atro),Ofla)),Phoo),Zwol),Oros)	26	51	71	81	81	75	3
10	(((((((Aaus,Agal),Afor),(Aphi,Atow)),Agaz),(Apus,Atro)),Phoo),Ofla),Zwol),Oros)	18	67	94	98	93	75	214
	N° of GFs/Trees	14,012	13,029	10,806	9,175	8,310	5,701	10,000

Hybridization vs. incomplete lineage sorting

First, we used *f3* and *f4*-statistics to test if the high level of topological discordance we found could be explained, at least, partially by past events of hybridization (introgression) between the species.

We found no evidence of introgression between any pair of species studied here since all f_3 values resulted in significantly positive values. On the other hand, f_4-statistics identified several significant sets of shared drift pathways between fur seals and sea lions (Fig. 3 and Supplementary Fig. S8) that could be interpreted as signals of introgression or as supporting an alternative phylogenetic relationship between the species considered (Peter 2016; Zhenge and Janke 2018). The strongest signals ($f_4 > 0.01$) were found in those phylogenies where *A. australis* and *A. forsteri* are sharing drift pathway. The other significant *f4* values were very close to zero ($f_4 < 0.001$, Fig. 3 and Supplementary Fig. S8). Note that with the exception of the tests that support introgression between *A. australis* and *A. forsteri* (Fig. 3a), all the other significant results implied on an alternative phylogenetic relationship between the six clades that diverged almost simultaneously (see above and Figs. 1, 3 and Supplementary Figs. S5 and S8), that is, where the internal branches were extremely

small. Therefore, we next tested if the gene trees discordances (see above) and the *f4*-statistics results could be explained by ILS alone.

The simulations that we used allowed ILS but not introgression between species. The gene trees were simulated from our species tree, including the divergence times (branch lengths) and effective population sizes (Fig. 1). Simulated gene trees were similar to the observed distributions, in particular the 200 kb GFs partition (Supplementary Fig. S6). The species tree was also the most frequent topology (Table 2). The coefficient of correlation for the observed data from the 50 and 200 kb and simulated topologies was 0.73 (Supplementary Fig. S9), which is high considering the simulated data are true gene trees not subjected by the uncertainties of the estimation of the empirical dataset.

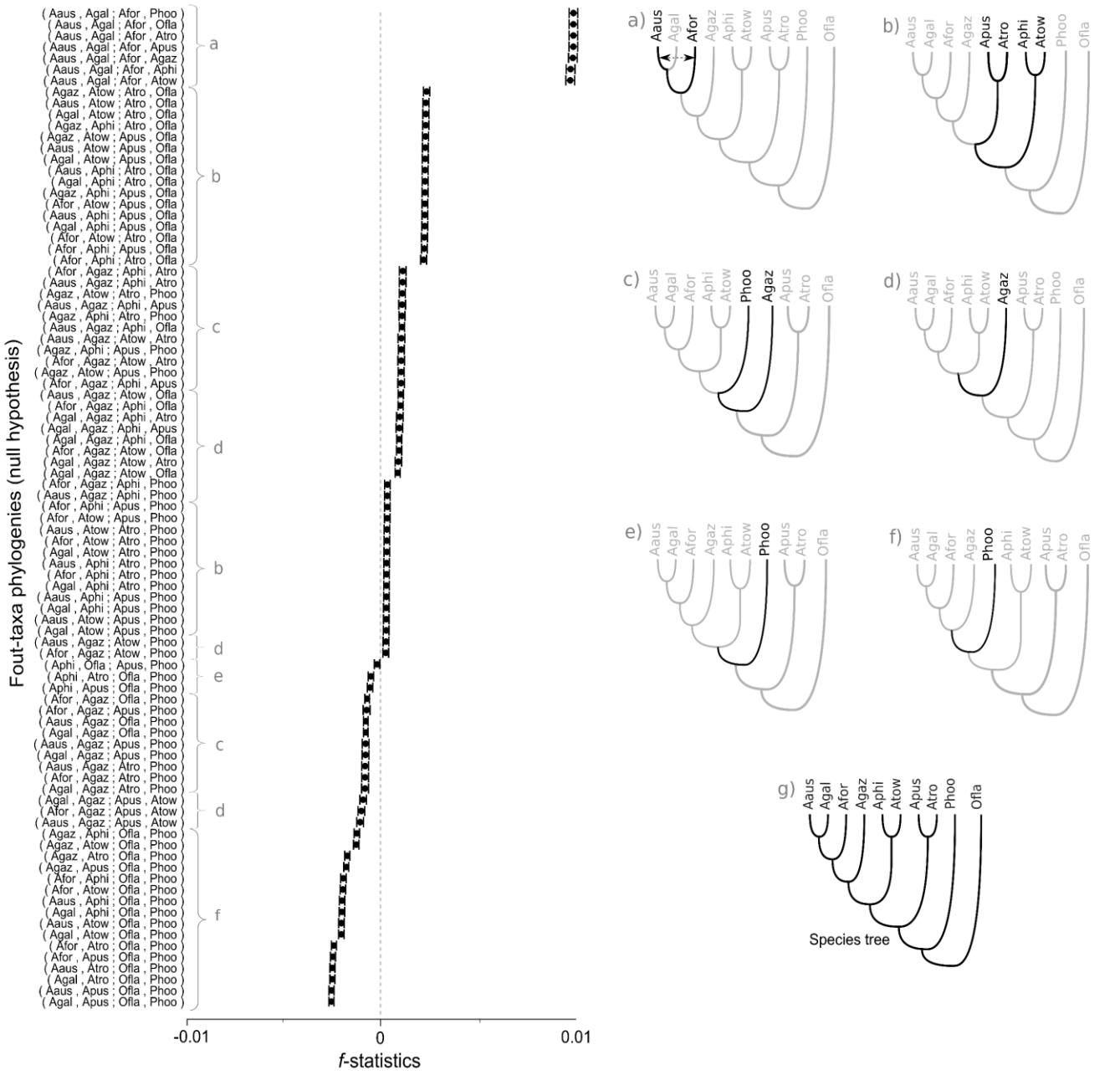


Figure 3. *F*$_4$-statistics distribution (x-axis) for the four-taxa phylogenies combinations (y-axis) with significant values (|Z-score| > 3). From *a* to *f*, the alternative branch position implied by the result. The letter *g* depicts the species tree (null model).

DISCUSSION

Otariidae phylogenomics

Here we presented the first whole-genome species tree of Otariidae that was consistently recovered with high support by different approaches. Our phylogeny agrees with the well-accepted relationships and resolved all uncertainties still prevalent to date in the group, such as the monophyly of *Arctocephalus*. We did not obtain the sequences for the monotypic *Neophoca* and *Eumetopias*.

However, the phylogenetic position of *E. jubatus* in a clade with *Zalophus* is widely accepted as it was found in all recent molecular and combined studies. Perhaps the most important omission in our study is *N. cinerea*, since, although some recent molecular studies support that it forms a clade with *P. hookeri* (e.g., Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012), other results suggested that it may be positioned in different places (Deméré and Berta 2003), the most extreme as a sister group to all Otariidae except *C. ursinus* (Churchill et al. 2014). To obtain and analyze *Neophoca* genome with the data presented here may be the most important next step to fully resolve Otariidae phylogeny.

The positions of *C. ursinus* as a sister species to all other Otariidae and the *Zalophus+Eumetopias* clade as sister to the rest of the family are also supported by the most recent studies (Wynen et al. 2001; Árnason et al. 2006; Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012; Berta and Churchill 2012; Churchill et al. 2014), therefore, fully refuting the validity of Arctocephalinae and Otariinae. These consistent results are expected since we showed *C. ursinus* as very divergent from all *Arctocephalus* and a non-monophyly of the remaining sea lions (Fig. 2).

However, the relationships between *Arctocephalus* spp., *P. hookeri*, *O. flavescens*, and *N. cinerea*, which was called the southern clade (Churchill et al. 2014) had no consensus so far, except for a few subgroups within *Arctocephalus* (Berta et al. 2018). Our dated species tree shows that most of the speciations within the southern clade were almost simultaneous (see below) and this could explain the high number of different phylogenetic hypotheses proposed for this group to date. Besides, with our whole genome data, we were able to find a highly supported phylogeny for this clade, the Southern American sea lion (*O. flavescens*) being the first to diverge around 3 Mya, followed by the New Zealand sea lion (*P. hookeri*) and a monophyletic *Arctocephalus*, both ~2.8 Mya. The previously controversial monophyly of *Arctocephalus* is here highly supported by many different phylogenetic approaches, which in consequence reject the use of *Arctophoca* for some of its species (*sensu* Berta and Churchill 2012).

Within *Arctocephalus*, four main lineages originated in a fast succession between ~2.8 Mya and 2.5 Mya. The first to diverge was *A. pusillus+A. tropicalis*, and their basal position within *Arctocephalus* was also found in several of the most recent studies (e.g., Berta et al. 2018), although some studies found them to be even more basal than *Phocarctos* (e.g., Yonezawa et al. 2009). The divergence time between them is ~1.2 Mya. Next originated the *A. phillippii* and *A. townsendi* sister pair, that in turn, diverged recently, at 0.6 Mya, which is expected since they are so similar that some authors considered *A. townsendi* a subspecies of *A. phillippii* (e.g., Committee on Taxonomy, 2018) and they grouped in all previous molecular phylogenies (Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012). Considering their morphological differences, the divergence time, that is similar or older than between *A. australis* and *A. galapagoensis* and their geographic isolation, we agree with most of the recent literature on their taxonomic status of full species (Repenning et al. 1971; Higdon et al. 2007; Yonezawa et al. 2009; Nyakatura et al. 2012; Berta and Churchill 2012; Aurioles-Gamboa 2015; Berta et al. 2018). The grouping of *A. gazella* with *A. forsteri+A. australis+A. galapagoensis* clade was found in some recent studies (e.g., Yonezawa et al. 2009; Churchill et al. 2014) although in most of the cases *A. gazella* was found in a polytomy with other *Arctocephalus* (e.g., Yonezawa et al. 2009; Berta et al. 2018) or in other positions. Finally, the *A. forsteri+A. galapagoensis+A. australis* clade was also highly expected, since for a long time they were considered as subspecies of *A. australis*, and only recently their specific status have been widely accepted as they are closely related in all previous studies. Nevertheless, there are some questioning on the species status of the New Zealand fur seals (*A. forsteri*) (see Berta and Churchill 2012) and indeed we found evidence of past low introgression between them and the South American fur seals (see below). However, we support *A. forsteri* as full species based on the same arguments, we metioned above for *A. phillippii* and *A. townsendi* also considering that they diverged from *A. australis+A. galapagoensis* more than 1 Mya.

The switched positions between *P. hookeri* and *A. tropicalis+A. pusillus* in the phylogeny of the mitochondrial genomes and 10 kb GF (Fig. 2, Supplementary Fig. S1 and S4) helps to explain

why most of the previous molecular studies recovered a non-monophyletic *Arctocephalus*, since mtDNA constitutes most or the total data used in these studies (e.g., Árnason et al. 2006; Higdon et al. 2007; Yonezawa et al. 2009; Churchill et al. 2014; Berta et al. 2018). The position of *A. forsteri* as the sister species of *A. australis* in our mtDNA tree instead of *A. galapagoensis*, as in our nuclear genome species tree, is also a common relationship in other mtDNA phylogenies (e.g., Wynen et al 2001; Yonezawa et al. 2009). However, studies that sequenced more individuals of *A. forsteri* and *A. australis* found multiple lineages in each species that are intermixed (e.g., Yonezawa et al. 2009). The relationship we found here, and the intermixing of lineages may have been caused by ILS, since the grouping of *A. australis* and *A. forsteri* occurred ~0.8% of the simulated trees, that considered ILS only. However, introgression may also have played a role in the history of this group, and this is supported by the f_4-statistics that indicated admixture between *A. australis* and *A. forsteri* (see below).

Divergence Times and Biogeographical Inferences

Our results agree with most divergence times estimates and the fossil record that suggest a northern Pacific origin for Otariidae and the split from Odobenidae at ~19 Mya, in the lower Miocene (Fig. 2 and Supplementary Fig. S5 - Demeré et al. 2003; Árnason et al, 2006; Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012; Churchill et al. 2014; Berta et al. 2018), when early Odobenidae and Otariidae lived in sympatry (Boessenecker and Churchill 2015). The divergence of *C. ursinus* at ~9 Mya is also similar to most estimates (e.g., Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012) although Berta et al. (2018) suggest a much older divergence ~16 Mya. Although a comparison of our results with previous divergence times (Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012; Berta et al. 2018) is not straightforward, given the differences in the topologies, some major points can be discussed. First, no previous study detected the explosive radiation at the beginning of the diversification of the southern clade around the Pliocene-Pleistocene boundary. Second, the divergences between *Zalophus* and the southern clade (~5.3 Mya) and between the main lineages in the latter (3-2.5 Mya) is younger than most previous estimates. As an extreme example, Berta et al. (2018) estimated the divergence between *Otaria*, *Phocarctos* and *Arctocephalus*

> 6 Mya. On the other hand, Berta et al. (2018) time estimates for the diversification within *Arctocephalus* (except for the *A. pusillus*+*A. tropicalis* clade) are very recent, < 1 Mya.

A relatively recent trans-hemispheric dispersal towards the Southern Hemisphere, likely through the west coast of South America achieved great support lately (see Yonezawa et al. 2009; Churchill et al. 2014, Berta et al. 2018), and our phylogenomic results strongly agree with this scenario. To estimate the minimum number of such dispersal events it is necessary to have additional data, such as the age and phylogenetic position of fossils. In the Southern Hemisphere, most of the otariid fossils date back to the Pleistocene, with *Hydrarctos* and *Arctocephalus* sp. nov. known from sediments of th Pliocene. The *Hydrarctos* fossil (Muizon 1978; Muizon and DeVries 1985), the oldest known to South America, belongs to the Pisco Formation of Peru, however, its phylogenetic position is uncertain. It was more consistently recovered within the southern clade (as sister to *Arctocephalus*) but was also found outside the southern clade, as the sister taxon to all extant otariids. Our results do not support its position inside the southern clade, since this group diversification started ~3 Mya, and the youngest date of the fossil is ~3.9 Mya, although the fossil may be as old as ~6.6 Mya (see Churchill et al. 2014). Therefore, *Hydrarctos* is likely the sister clade to the southern clade or, assuming the oldest dates, may represent an independent (and eventually extinct) trans-equatorial dispersal that preceded the one that originated the extant southern clade. *Arctocephalus* sp. nov., a fossil that belongs to the Varswater formation of South Africa, was dated between ~2.7-5 Mya and we used its most recent date as the minimum age for *Arctocephalus* in our StarBEAST2 calibrated species tree. Our point estimate for the origin of *Arctocephalus* was ~2.8 (C.I. 95% ranging from 2.4 to 3.3 Mya, Fig. 2), close to the minimum limit. The divergence times of the species tree reestimated using only the calibrated point at the root (20 Mya) with BP&P (Fig. S5) showed values very similar to those of StarBEAST trees, therefore supporting our estimates of diversification of the southern clade dates between ~3 and 2.5 Mya.

Based on these results, dispersal to the Southern Hemisphere seem to have occurred anytime between ~5 Mya, in the split of the southern clade from the northern *Zalophus* group, and ~3 Mya,

the beginning of the explosive radiation within the southern clade. However, if *Hydrarctos* is considered a member of the southern clade and its minimum age of ~4 Mya, the southern dispersal could have occurred more than 1 My before the burst of diversification of the extant species. It seems rather useless to speculate at which specific moment the trans-hemispheric dispersal would have occurred in this large interval (1 to 2 My), since it could have been a single, perhaps fortuitous event, only mildly facilitated by conditions such as sea temperature (Churchill et al. 2014). Nevertheless, the early Pliocene (~4 Mya) and the mid Pliocene (~3 Mya) were characterized by warm temperature (Fedorov et al. 2013; Rousselle et al., 2013) that may have isolated the incipient taxa in the South Hemisphere and limited new events of speciation in the Southern Pacific Ocean. On the other hand, the diversification of the southern clade in six lineages distributed over the Southern Hemisphere, at a relatively short time interval (~3.0-2.5 Mya) was likely promoted by important climatic events. At 3 Mya a sharp global cooling started, associated with the beginning of the Northern Hemisphere Glaciation (Marlow et al. 2000). The strong environmental changes caused by the concomitant global cooling of the Plio-Pleistocene transition and the total closure of Panama Isthmus would have provided suitable niche in all Southern Pacific Ocean (O'Dea et al. 2012; Churchill et al. 2014). Thus, it may have opened the way for long dispersal events, the establishment of new colonies, the consequently local adaptation to new niches and the rapid speciation from a potential large ancestral population could have occured.

Our biogeographical scenario is similar in broad terms to those presented earlier (mainly by Yonezawa et al. 2009 and Churchill et al. 2014), but the different divergence dates, and details of the phylogeny resulted in alternative hypothesis. For example, a simultaneous southern dispersal and the diversification of the southern clade around 3 Mya (see above) may explain the absence of otariids in the North Atlantic waters, since the total closure of the Central American Seaway finished ~3 Mya (Churchill et al. 2014).

Genealogical Discordances, ILS and Introgression

We found an extremely high degree of topological discordances throughout genomes (including the mtDNA), with many topologies appearing only once, even in the GFs of 200 kb (Table 2, Supplementary Fig. S6). This gene trees discordance could not be assigned to lack of information in the GFs since the average of internal nodes support (83.7%) is very high in gene trees from the 200 kb GFs. These results explain the high degree of discordances between the phylogenies estimated by previous studies and why it has been so difficult to find a robust classification for the Otariidae based on few genes. The explosive radiation that originated the southern clade lineages generated very short internal branches, which increased the occurrence of ILS and did not allow the accumulation of numbers of substitutions on these nodes, the latter caused poorly supported gene trees based on standard-size gene regions (McCormack et al. 2013; Suh et al. 2015). Therefore, to estimate with confidence the evolutionary history of groups that originated by explosive radiation and with extremely high levels of gene tree discordances, whole genome datasets in a multispecies coalescence model framework seem to be the best choice.

Topological discordance between genomic regions is not unusual and is being observed with increasing frequency in recent phylogenomic studies (e.g., Martin et al. 2013; Li et al. 2016; Árnason et al., 2018). The sources of topological discordances are mainly attributed to ILS, as suggested above or to hybridization (Bravo et al. 2019). Recent genomic studies have shown that introgression has been very frequent in the history of several groups (e.g., Figueiró et al. 2017; Masello et al. 2019). We do not support any relevant role for introgression in most of the evolutionary history of the Otariidae, since there is no signal of hybridization between species with the f_3-statistics and the pattern of gene tree discordances could be explained by ILS alone. Except for the signal between *A. australis* and *A. forsteri,* which presented a strong signal of introgression on the f_4-statistics. We hypothesize that all other *f4* significant and weak values actually support the ILS model and the almost polytomous relationship between the six main clades of the southern group (see results).

Considering the above results, in our species tree inferences, we used methods that implemented MSC models that consider only the occurrence of ILS. However, there has been recent implementations that included hybridization in the MSC models, with the purpose to estimate species networks (Zhang et al. 2017; Wen et al. 2018; Wen and Nakhleh 2018; Jones 2019). To assess if we could obtain a more complete picture of the relationships within the family we tried to recover a species network using two of these methods: the SpeciesNetwork (Zhang et al. 2018) and DENIM (Jones 2019), both implemented in StarBEAST2. For these analyses we used 100 GFs of 1 and 5 kb selected among the most variables (using the initial bases) from the 300 GFs of 50 kb used in the StarBEAST2 analyzes (Fig. 2 and Supplementary Fig. S5). Unfortunately, the Bayesian estimations did not stabilize even after long runs (>1 billion steps, DENIM) or we recovered several topologies that are very different from all other main topologies we recovered with all other methods as presented above (with SpeciesNetwork). We hypothesized that the source of these inconsistencies could be related to the elevated number of taxa (since these methods are recommended for few, <6 taxa). The virtual polytomy between six lineages and the consequent high levels of ILS in our dataset may also have contributed to the non-stabilization of the analyses.

The relationships within the clade with *A. australis, A. galapagoensis* and *A. forsteri* seem to reflect a complex scenario since we found pieces of evidence for both introgression and ILS between them. Besides that, previous studies have found the absence of reciprocal monophyly on the mtDNA between the species and the possible existence of at least one cryptic species (King 1954; Repenning et al. 1971; Wynen et al. 2001; Yonezawa et al. 2009; Oliveira and Brownell 2014). Therefore, we defer the discussion about the evolutionary history of this group for a further study.

FUNDING

This study was financed by CNPq (Brazil) and FAPERGS (PRONEX 12/2014) (Brazil) grants to SLB, US Navy NICOP 2015 granted to Eduardo Eizirik and SLB, Coordenação de Aperfeiçoamento

de Pessoal de Nivel Superior (CAPES) (Brazil**)** – Finance Code 001 PhD scholarship to FL, and Society for Marine Mammalogy Small Grants-in-aid of Research 2016 granted to FL.

ACKNOWLEDGEMENTS

We are very thankful for the support of the following people and institutions: Laboratório de Biologia Genômica e Molecular-PUCRS staff for support, general suggestions and discussions; Jochen B. W. Wolf Group and Ludwig-Maximilians-Universität München which kindly hosted FL in his sandwich PhD; GEMARS, NOAA-SWFSC and Kelly Robertson that contributed with samples; Nathaly Miranda who reviewed many versions of this manuscript; and all people and institutions that directly or indirectly contributed with the construction of this manuscript.

REFERENCES

Árnason U., Lammers F., Kumar V., Nilsson M.A., Janke A. 2018. Whole-genome sequencing of the blue whale and other rorqual finds signatures for introgressive gene flow. Sci. Adv. 4:eaap9873

Árnason U., Gullberg A., Janke A., Kullberg M., Lehman N., Petrov E.A., Väinölä R. 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. Mol. Phylogenet. Evol. 41:345-354

Arnold B., Corbett-Detig R.B., Hartl D., Bomblies K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol. Ecol. 22:3179-3190.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Aurioles-Gamboa D. 2015. *Arctocephalus philippii*. The IUCN Red List of Threatened Species 2015: e.T2059A61953525

Avery G., Klein R.G. 2011. Review of fossil phocid and otariid seals from the southern and western coasts of South Africa. T. Roy. Soc. S. Afr. 66:14-24

Berta A., Churchill M. 2012. Pinniped taxonomy: Review of currently recognized species and subspecies, and evidence used for their description. Mamm. Rev. 42:207-234

Berta A., Churchill M., Boessenecker R.W. 2018. The origin and evolutionary biology of pinnipeds: seals, sea lions and walruses. Annu. Rev. Earth. Planet. Sci. 46:203-228

Berta A., Deméré T.A. 1986. *Callorhinus gilmorei* n. sp., (Carnivora: Otariidae) from the San Diego Formation (Blancan) and its implications for otariid phylogeny. Trans. San. Diego. Soc. Nat. Hist. 21:111-116

Boessenbecker R.W., Churchill M. 2015. The oldest known fur seal. Biol. Lett. 11:20140835

Borowiec M.L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4:e1660

Bouckaert R., Heled J. 2014. DensiTree2: seeing trees through the forest. Doi: https://doi.org110/012401

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C. H., Xie D., Suchard M. A., Rambaut A., Drummond A. J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol., 10:e1003537

Bravo G.A., Antonelli A., Bacon C.D., Bartoszek K., Blom M.P.K., Huynh S., Jones G., Knowles L.L., Lamichhaney S., Marcussen T., Morlon H., Nakhleh L.K., Oxelman B., Pfeil B., Schliep A., Wahlberg N., Werneck F.P., Wiedenhoeft J., Willows-Munro S., Edwards S.V. 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ 7:e6399

Capella-Gutierrez S., Silla-Martinez J.M., Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyzes. Bioinformatics 25:1972-1973

Chakrabarty P., Faircloth B.C., Alda F., Ludt W.B., McMahan C.D., Near T.J., Dornburg A., Albert J.S., Arroyave J., Stiassny M.L.J., Sorenson L., Alfaro M.E. 2017. Phylogenomic systematics of Ostariophysan fishes: ultraconserved elements support the surprising non-monophyly of Characiformes. Syst. Biol. 66:881-895

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., Lee J.J. 2015. Second-generation PLINK: rising to the challenge of larger and richer data sets. GigaScience 4

Churchill M., Boessenecker R.W., Clementz M.T. 2014. Colonization of the southern hemisphere by fur seals and sea lions (Carnivora: Otariidae) revealed by combined evidence phylogenetic and Bayesian biogeographical analysis. Zool. J. Linn. So.c 172:200-225

Committee on Taxonomy of Marine Mammals. 2018. List of marine mammals species and subspecies. Available fromhttps://www.marinemammalscience.org/species-information/list-marine-mammal-species-subspecies

Darriba D., Taboada G.L., Doallo R., Posada D. 2012. jModelTest 2: More models, new heuristics and parallel computing. Nat. Methods 9:722

Deméré T.A., Berta A., Adam P.J. 2003. Pinnipedimorph evolutionary biogeography. B. Am. Mus. Nat. His. 279:32-76

Drummond A.J., Bouckaert R.R. 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Press. Cambridge 244 pp.

Elworth R.A.L., Ogilvie H.A., Zhu J., Nakhleh L. 2019. Advances in computational methods for phylogenetic networks in the presence of hybridization. In Warnow T (ed.) Bioinformatics and Phylogenetics: Seminal Contributions of Bernard Moret. Springer International Publishing

Esselstyn J.A., Oliveros C.H., Swanson M.T., Faircloth B.C. 2017. Investigating difficult nodes in the placental mammal tree with expanded taxon sampling and thousands of ultraconserved elements. Genome Biol. Evol. 9:2308-2321

Faircloth B.C., Sorenson L., Santini F., Alfaro M.E. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon target sequencing of ultraconserved elements (UCEs). PLoS One 8:e65923

Fedorov A.V., Brierly C.M., Lawrence K.T., Liu Z., Dekens P.S., Ravelo A.C. 2013. Patterns and mechanisms of early Pliocene warmth. Nature 496:43-49

Felsestein J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5:164-166

Figueiró H.V., Li G., Trindade F., Assis J., Pais F., Fernandes G., Santos S.H., Hughes G.M., Komissarov A., Antunes A., Trinca C.S., Rodrigues M.R., Linderoth T., Bi K., Silveira L., Azevedo F.C.C., Kantek D., Ramalho E., Brassaloti R.A., Villela P.M.S., Nunes A.L.V., Teixeira

R.H.F., Morato R.G., Loska D., Saragüeta P., Gabaldón T., Teeling E.C., O'Brien S.J.O., Nielsen R., Coutinho L.L., Oliveira G., Murphy W.J., Eizirik E. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci. Adv. 7:e1700299

Flouri T., Jiao X., Rannala B., Yang Z. 2018. Species Tree Inference with BPP using Genomic Sequences and the Multispecies Coalescent. Mol. Biol. Evol. 35:2585-2593

Foote A.D., Liu Y., Thomas G.W., Vinař T., Alföldi J., Deng J., Dugan S., van Elk C.E., Hynter M.E., Joshi V., Khan Z., Kovar C., Lee S.L., Linbald-Toh K., Mancia A., Nielsen R., Qin X., Qu J., Raney B.J., Vijay N., Wolf J.B.W., Hahn M.W., Muzny D.M., Worley K.C., Gilbert M.T., Gibbs R.A. 2015. Convergent Evolution of the genomes of marine mammals. Nat. Genet. 47:272-275

Gautier M., Gharbi K., Cezard T., Foucaud J., Kerdelhué C., Pudlo P., Cornuet J.M., Estoup A. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol. Ecol. 22:3165–3178

Harris A.M., DeGiorgio M. 2012. Admixture and ancestry inference from ancient and modern samples through measures of population genetic drift. Hum. Biol.114

Higdon J.W., Bininda-Emonds O.R.P., Beck R.M.D., Ferguson S.H. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evol. Biol. 7:216

Hoban S.M., Gaggiotti O.E., Bertorelle G. 2013. The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulation-based study. Mol. Ecol. 22:3444-3450

Humble E., Martinez-Barrio A., Forcada J., Trathan P.N., Thorne M.A.S., Hoffmann M., Wolf J.B.W., Hoffman J.I. 2016. A draft fur seal genome provides insights into factors affecting SNP validation and how to mitigate them. Mol. Ecol. Res. 16(4):909-921

Humble E., Dasmahapatra K.K., Martinez-Barrio A., Gregório I., Forcada J., Polikeit A.C., Goldsworthy S.D., Goebel M.E., Kalinowski J., Wolf J.B.W., Hoffman J.I. 2018. RAD Sequencing and a Hybrid Antarctic Fur Seal Genome Assembly Reveal Rapidly Decaying Linkage Disequilibrium, Global Population Structure and Evidence for Inbreeding. G3 (Bethesda) 31:2709-2722

Huson D.H., Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol 6:1061-1067

Jones G.R. 2019. Divergence estimation in the presence of incomplete lineage sorting and migration. Syst. Biol. 68:19-31

Junier T., Zdobnov E.M. 2010. The newick utilities: high-throughput phylogenetic tree processing using unix shell. Bioinformatics 26:1669-1670

King J. 1954. The otariid seals of the Pacific coast of America. Bull. Brit. Mus. Zool. 2:311-337

Korneliussen T.S., Albrechtsen A., Nielsen R. 2014. ANGSD: analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356

Kumar S., Subramanian S. 2001. Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. U. S. A. 22:803-808

Lam H.M., Ratmann O., Boni M.F. 2018. Improved algorithmic complexity for 3seq recombination detection algorithm. Mol. Biol. Evol. 35:247-251

Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17-30

Leppälä K., Nielsen S., Mailund T. 2017. admixturegraph: an R package for admixture graph manipulation and fitting. Bioinformatics 33:1738-1740

Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754-60

Li G., Davis B.W., Eizirik E., Murphy W.J. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Res. 26:1-11

Liu L., Yu L. 2010. Phybase: an R package for species tree analysis. Bioinformatics 26:962-963

Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21-30

Marlow J.R., Lange C.B., Wefer G., Rosell-Melé A. 2000. Upwelling intensification as part of the Pliocene-Pleistocene climate transition. Science 22:2288-2291

Martin S.H., Dasmahapatra K.K., Nadeau N.J., Salazar C., Walters J.R., Fraser S., Blaxter M., Manica A., Mallet J., Jiggins C.D. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23:1817-1828

Masello J.F., Quillfeldt P., Sandoval-Castellanos E., Alderman R., Calderón L., Cherel Y., Cole T.L., Cuthbert R.J., Marin M., Massaro M., Navarro J., Phillips R.A., Shepherd L.D., Suazo C.g., Weimerskirch H, Moodley Y. Additive traits lead to feeding advantage and reproductive isolation, promoting homoploid hybrid speciation. Mol. Biol. Evol. 36:1671-1685

McCormack J.E., Faircloth B.C. 2013. Next-generation phylogenetics takes root. Mol. Ecol. 22:19-21

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., dePristo M.A. 2010. The Genome Analysis Toolkit: a MapReference framework for analyzing next-generation DNA sequencing data. Genome Res. 20:297-303

Muizon C. 1978. *Arctocephalus* (*Hydrarctos*) lomasiensis, subgen. nov. et. nov sp., un unouvel Otariidae du Mio-Pliocene de Sacaco. B. I'Inst. Franc. d'Etu. And. 7:169189

Muizon C., DeVries T.J. 1985. Geology and paleontology of late cenozoic marine deposits in the Sacaco area (Peru). Geol. Rundsch. 74:547-563

Nakhleh L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. Trends. Ecol. Evol. 28:719-728

Nyakatura K., Bininda-Emonds O.R.P. 2012. Updating the evolutionary history of Carnivora (Mammalia): A new species-level supertree complete with divergence time estimates. BMC Biol. 10:12. doi:10.1186/1741-7007-10-12

O'Dea A., Hoyos N., Rodriguez F., Degracia B., De Gracia C. 2012. History of upwelling in the Tropical Eastern Pacific and the palaeogeography of the Isthmus of Panama. Palaeograph. Palaeoclimat. Palaeoecol. 348-349:59-66

Ogilvie H.A., Bouckaert R.R., Drummond A.J. 2017. StarBEAST2 Brings faster species tree inference and accurate estimates of substitution rates. Mol. Biol. Evol. 34:2101-2114

Oliveira L., Brownell Jr. R.L. 2014. Taxonomic status of two subspecies of South American fur seals: *Arctocephalus australis australis* vs. *A. a. gracilis*. Mar. Mamm. Sci. 30:1258-1263

Oliver J.C. 2013. Microevolutionary process generates phylogenomic discordance at ancient divergence. Evolution 5:568-583

Peter B.M. 2016. Admixture, population structure, and *F*-statistics. Genetics 202:1485-1501

Pickrell J.K., Pritchard J.K. 2012 Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8:e1002967

Quinlan A.R., Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 6:841-842

Rambaut A., Drummond A.J. 2007. Tracer v1.7 Available from https://github.com/beast-dev/tracer/releases/latest

Rambaut A., Drummond A.J. 2010. TreeAnnotator version 1.6.1 Available from http://beast.bio.ed.ac.uk

Rambaut A. 2017. FigTree v1.4.4. Available from: https://github.com/rambaut/figtree/

Reddy S., Kimball R.T., Pandey A., Hosner P.A., Braun M.J., Hackett S.J., Han K.L., Harshmann J., Huddleston C.J., Kingston S., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Witt C.C., Yuri T.B., Braun E.L. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian Tree of Life more than taxon sampling. Syst. Biol. 66:857-879

Repenning C.A., Peterson R.S., Hubbs C.L. 1971. Contributions to the systematics of the southern fur seals, with particular reference to the Juan Fernández and Guadalupe species. Antarct. Res. S. 18:1-34

Rheindt F.E., Fujita M.K., Wilton P.R., Edwards S.V. 2014. Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. Syst. Biol. 63:134-152

Rokas A., Williams B.L., King N., Carrol S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798-804

Rouselle G., Beltran C., Sicre M.A., Raffi I., De Rafélis M. 2013. Changes in sea-surface conditions in the Equatorial Pacific during the middle Miocene–Pliocene as inferred from coccolith geochemistry. Earth Plant Sci Lett. 361:412-421

Schubert M., Ermini L., Sarkissian C.D., 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. Nat. Protoc. 9:1056-1082

Schubert M., Lindgreen S., Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9:88

Sclater P.L. 1897. On the distribution of marine mammals. Proc. Zool. Soc. London, 349–359

Shafer A.B.A., Peart C.R., Tusso S., Maayan I., Brelsford A., Wheat C.W., Wolf J.B.W. 2016. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Meth. Ecol. Evol. 8:907-017

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

Suh A., Smeds L., Ellegreen H. 2015. The Dynamics of Incomplete Lineage Sorting across the Ancient Adaptive Radiation of Neoavian Birds. PLoS Biol. 13:e1002224

Wang K., Lenstra J.A., Liu L., Hu Q., Ma T., Qiu Q., Liu J. 2018. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. Comm. Biol. 1:169

Wen D., Nakhleh L. 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. Syst. Biol. 67:439-457

Wen D., Yu Y., Nakhleh L. 2018. Inferring phylogenetic networks using Phylonet. Syst. Biol. 67:735-740

Wynen L.P., Goldsworthy S.D., Insley S.J., Adams M., Bickham J.W., Francis J., Gallo J.P., Hoelzel A.R., Majluf P., White R.W.G., Slade R. 2001. Phylogenetic relationships within eared seals

(Otariidae: Carnivora): implications for the historical biogeography of the family. Mol. Phylogenet. Evol. 21:270-284

Yonezawa T., Kohno N., Hasegawa M. 2009. The monophyletic origin of sea lion and fur seals (Carnivora: Otariidae) in the southern hemisphere. Gene 441:89-99

Zhang C., Ogilvie H.A., Drummond A.J., Stadler T. 2017. Bayesian inference of species networks from multilocus sequence data. Mol. Biol. Evol. 35:504-517

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19 (suppl 6):153

Zheng Y., Janke A. 2018. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. BMC Bioinformatics. 19:10

Ziheng Y. 2015. The BPP program for species tree estimation and species delimitation. Curr. Zool. 61:854-865

Table S1. PALEOMIX statistics for each species mapped against *O. rosmarus* genome (2.4 Gb genome).

	Library	Coverage	Proportion Ref. Genome Mapped	Fraction total hits vs. n° of reads retained	Average of hit length	Fraction of PCR duplicates
A. australis	TruSeq Nano	21.34	0.85	0.91	149.81	0.06
A. galapagoensis	TruSeq Nano	15.42	0.79	0.90	150.14	0.12
A. forsteri	TruSeq PCR Free	26.05	0.86	0.91	151.40	0.05
A. gazella	GenBank	32.44	0.83	0.89	191.46	0.07
A. philippii	TruSeq PCR Free	17.44	0.85	0.90	148.90	0.06
A. townsendi	TruSeq PCR Free	28.20	0.84	0.90	148.82	0.07
A. pusillus	TruSeq PCR Free	25.28	0.84	0.91	150.59	0.08
A. tropicalis	TruSeq PCR Free	28.63	0.82	0.91	150.50	0.10
P. hookeri	TruSeq PCR Free	17.49	0.85	0.90	149.11	0.05
O. flavescens	TruSeq PCR Free	22.95	0.86	0.91	151.76	0.05
Z. wollebaeki	GenBank	6.04	0.89	0.91	147.05	0.01
C. ursinus	GenBank	27.67	0.70	0.84	133.63	0.17

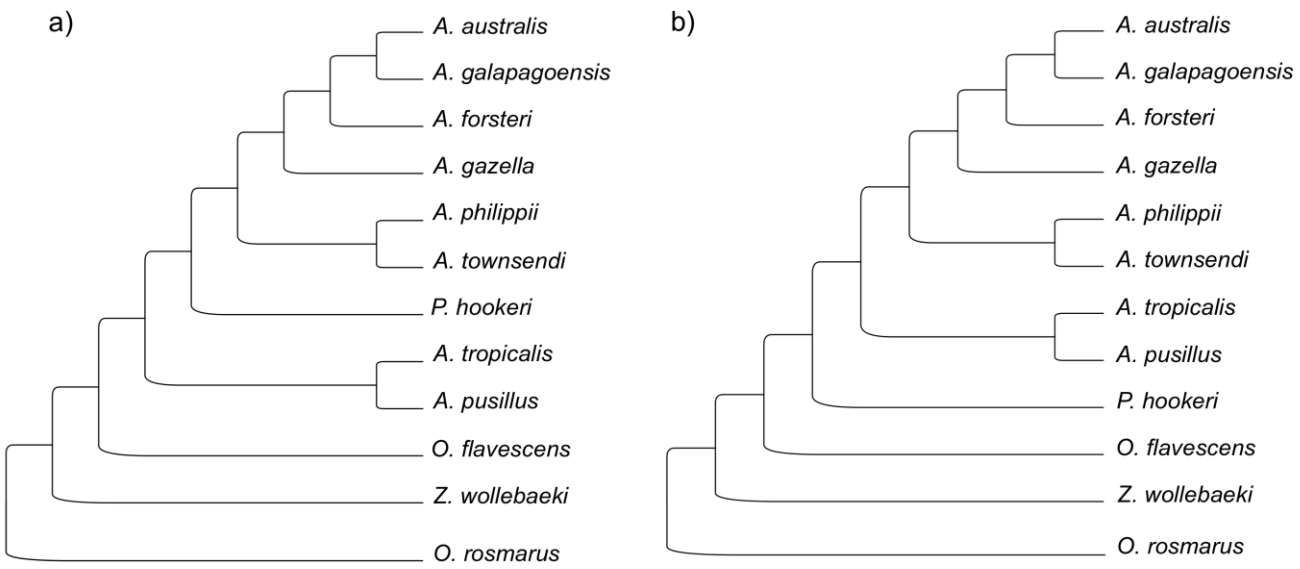


Figure S1. Species trees recovered with ASTRAL-III for the different genomic fragment sizes. All branches with support of 1. (A) Species tree recovered for the 10 kb GF. (B) Species tree recovered for the 20, 50, 80, 100 and 200 kb GFs.

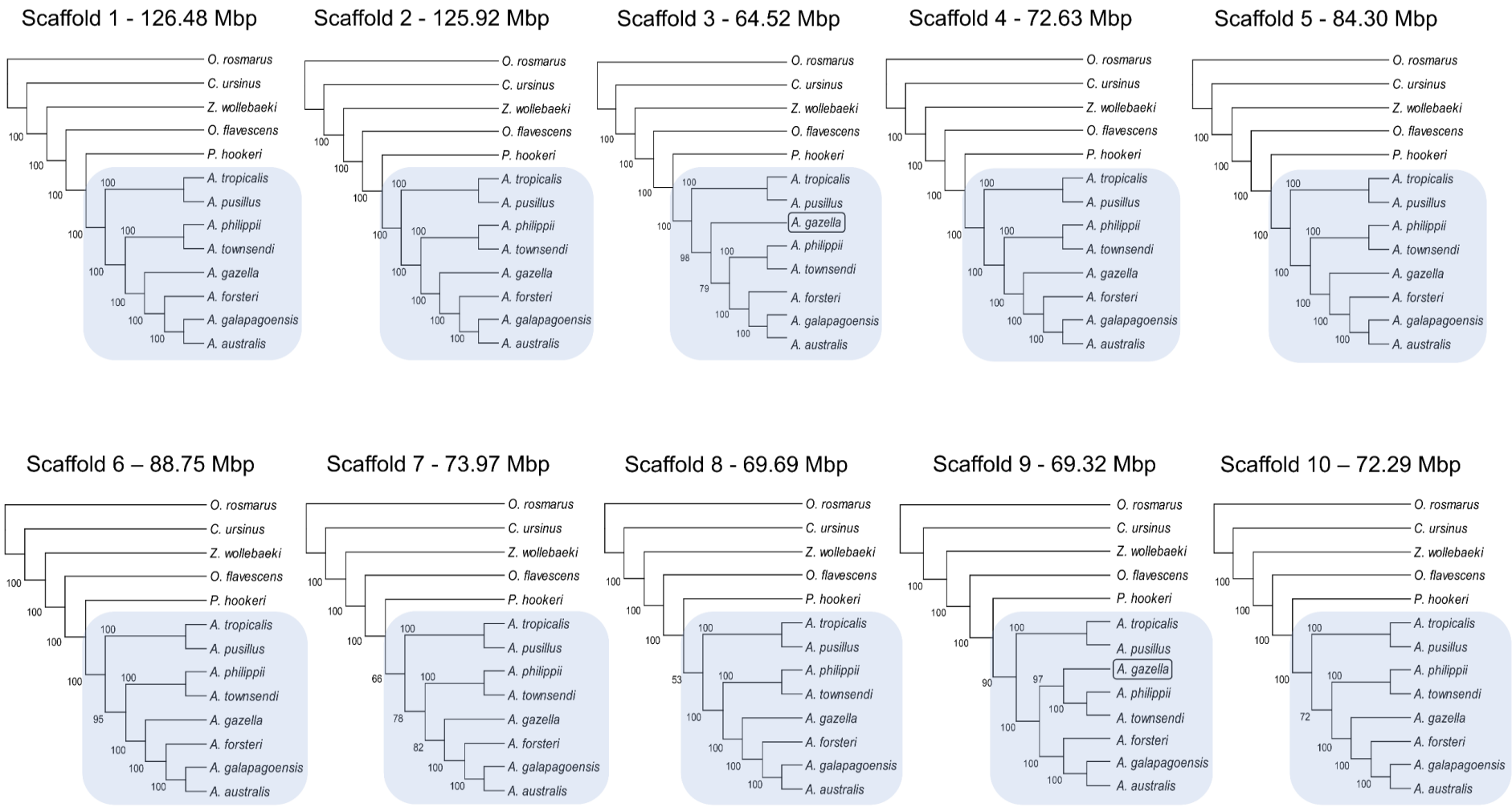


Figure S2. Scaffold sizes and their respective ML bootstrapped topology showing the monophyly of *Arctocephalus*.

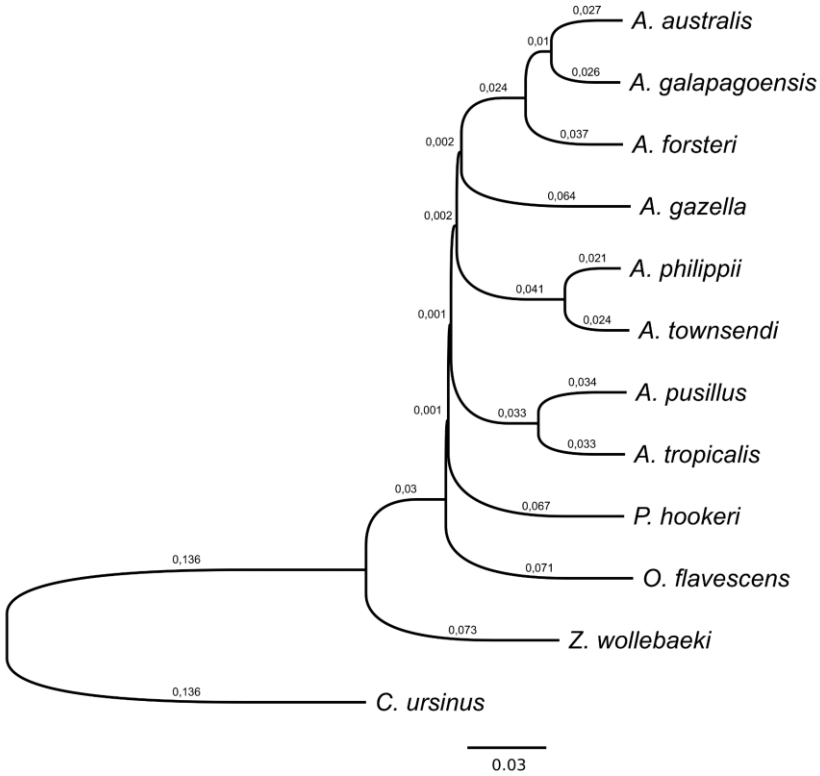


Figure S3. Whole-genome genetic distance tree showing the same topology as the species trees recovered in the GFs of 20-200 kb.

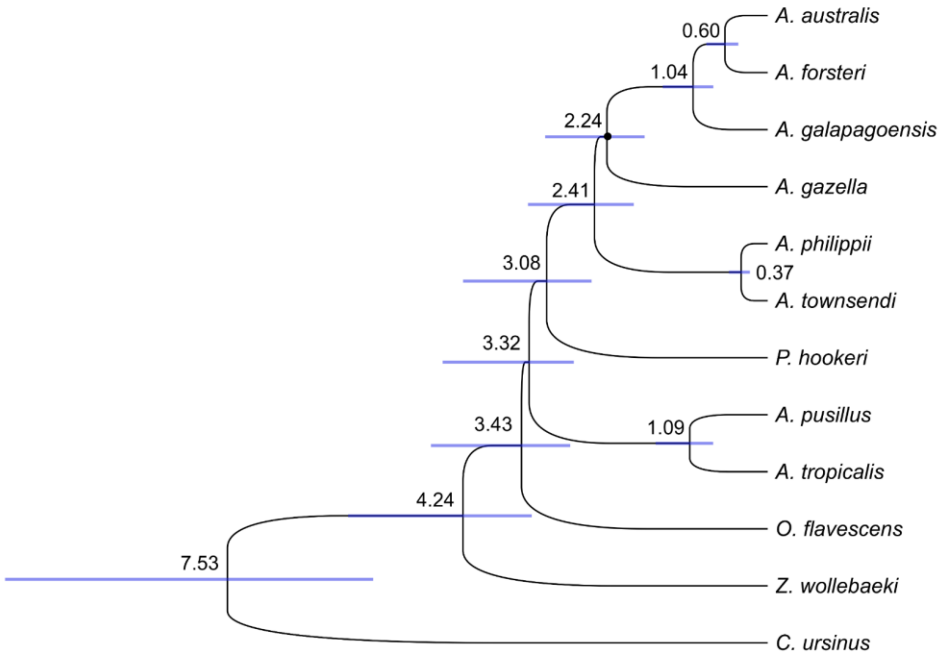


Figure S4. Bayesian mitogenome phylogeny. Almost all nodes with HPD = 1, the exception is the highlighted with a black dot (HPD = 0.90).

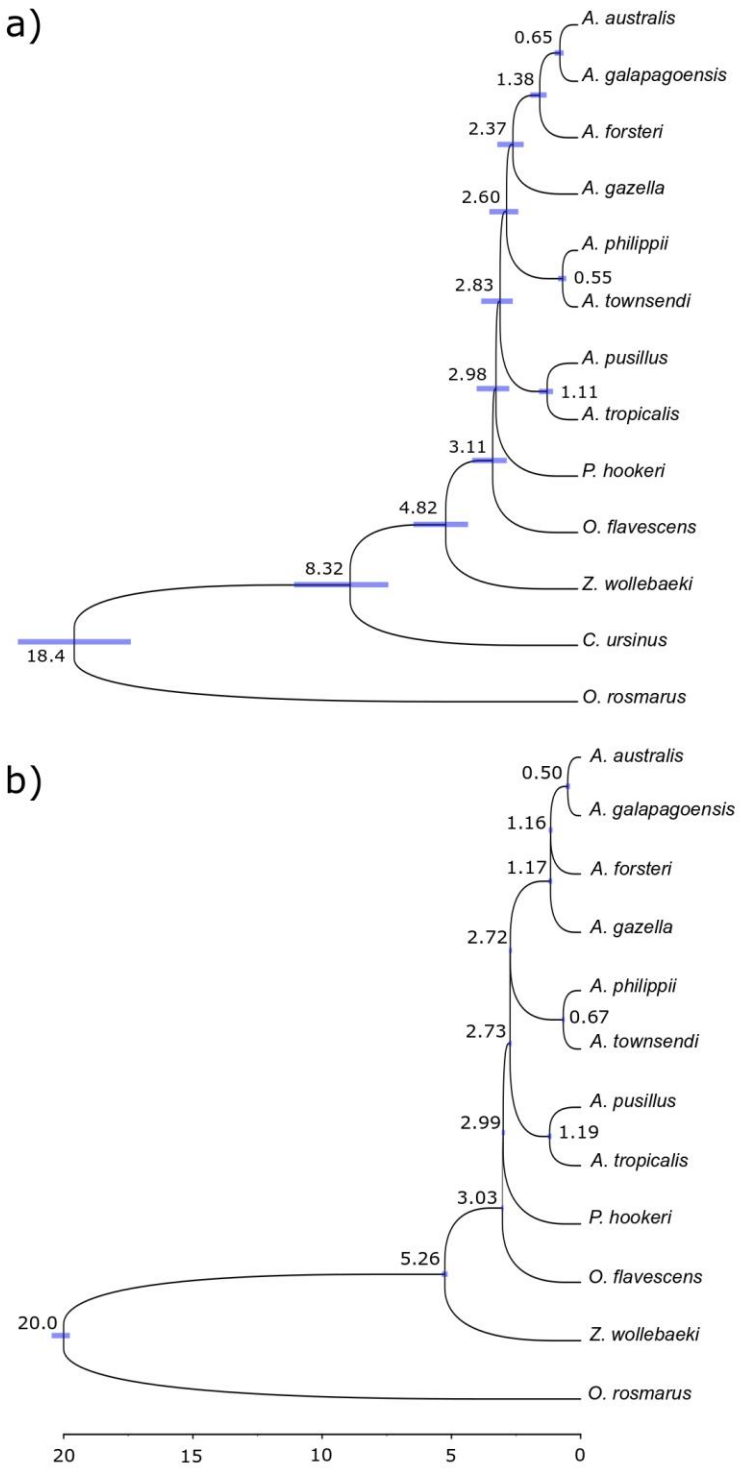


Figure S5. Multispecies coalescent trees and the divergence times recovered with 300 GF of 50 kb with two different approaches: (a) Calibrated species tree recovered with StarBEAST2. Most nodes with HPD = 1. (b) Divergence times (Mya) estimated with BP&P. Blue bars represent the confidence interval of 95%.

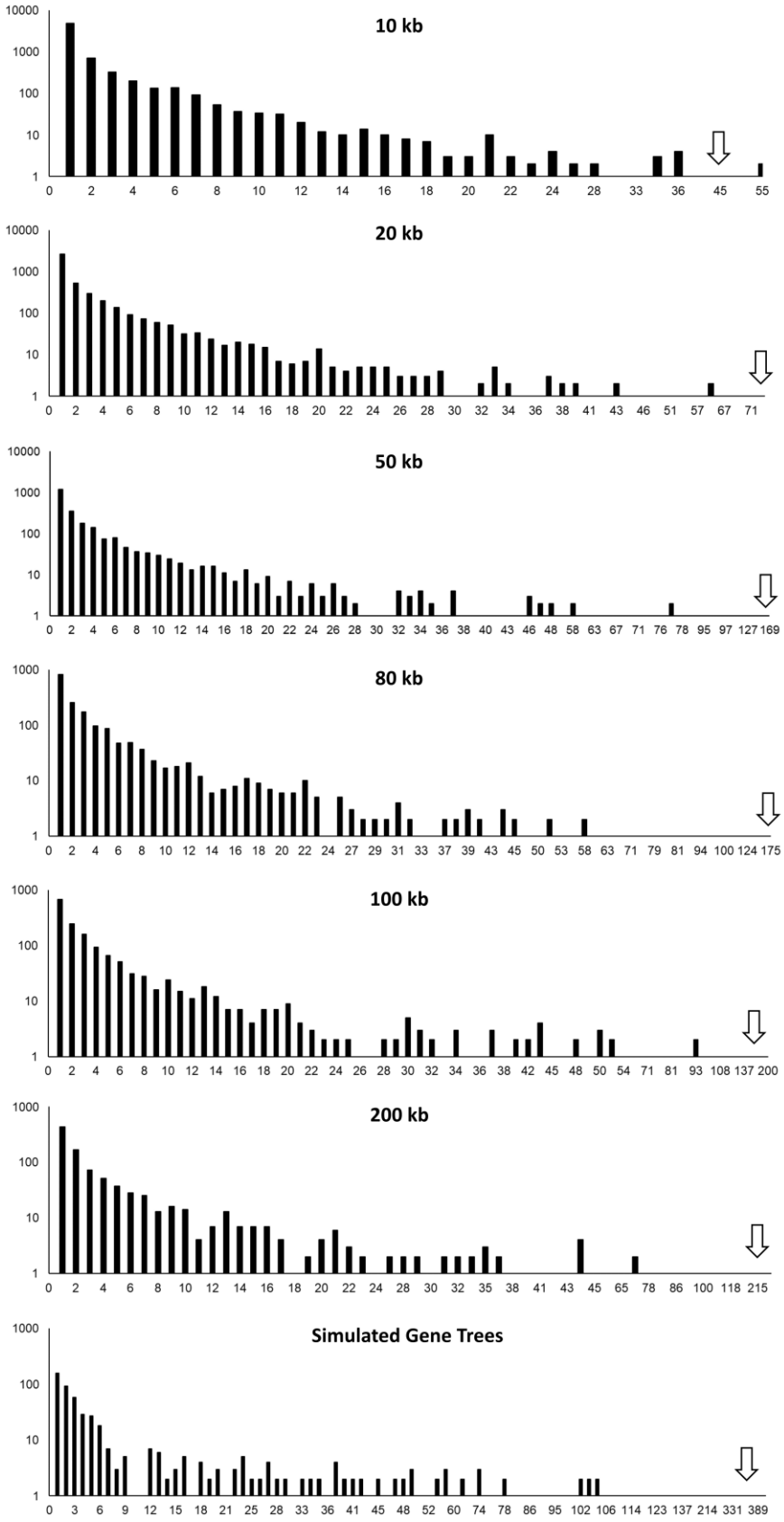


Figure S6. The extensive incongruence of topologies in different partitions. The x-axis presents the absolute topology frequency and the y-axis, cumulative frequencies (log-scale). The hollow arrows indicate the frequency of the species tree.

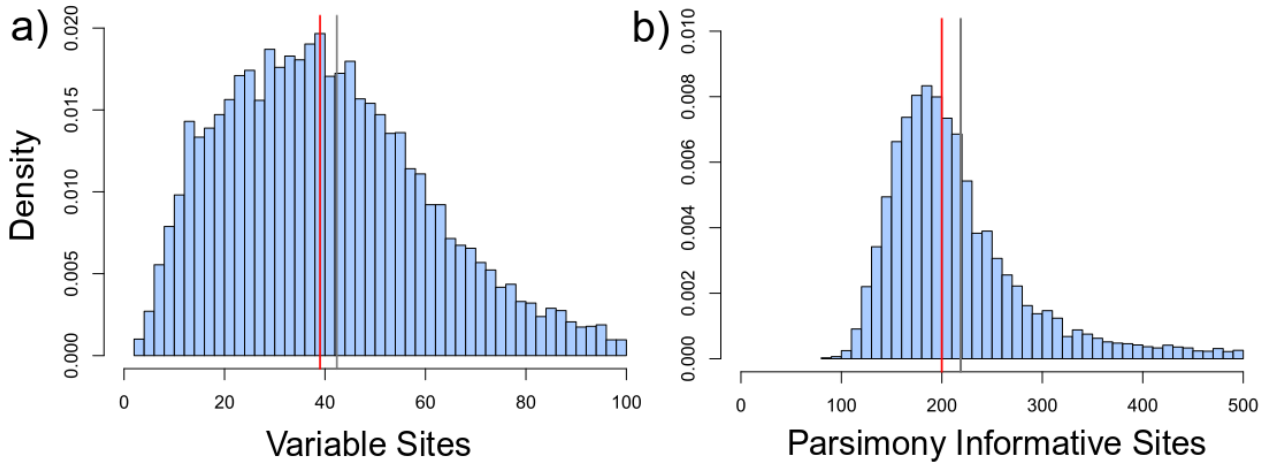


Figure S7. Average (red line) and median (grey line) of genetic variation in the GFs of 50 kb. (a) Variable sites between two closest species (*A. australis* and *A. galapagoensis*). (b) Parsimony informative sites in the multispecies GFs.

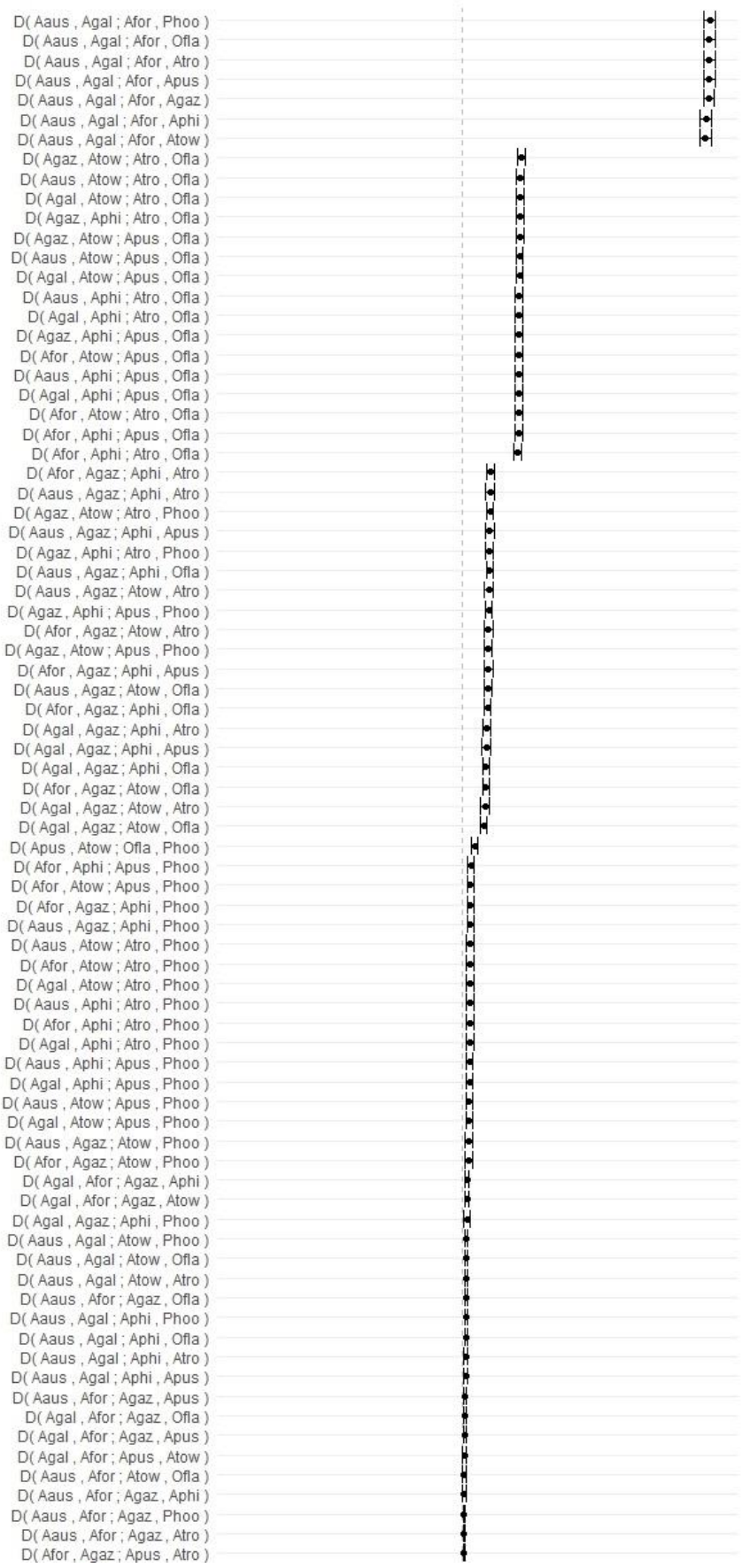
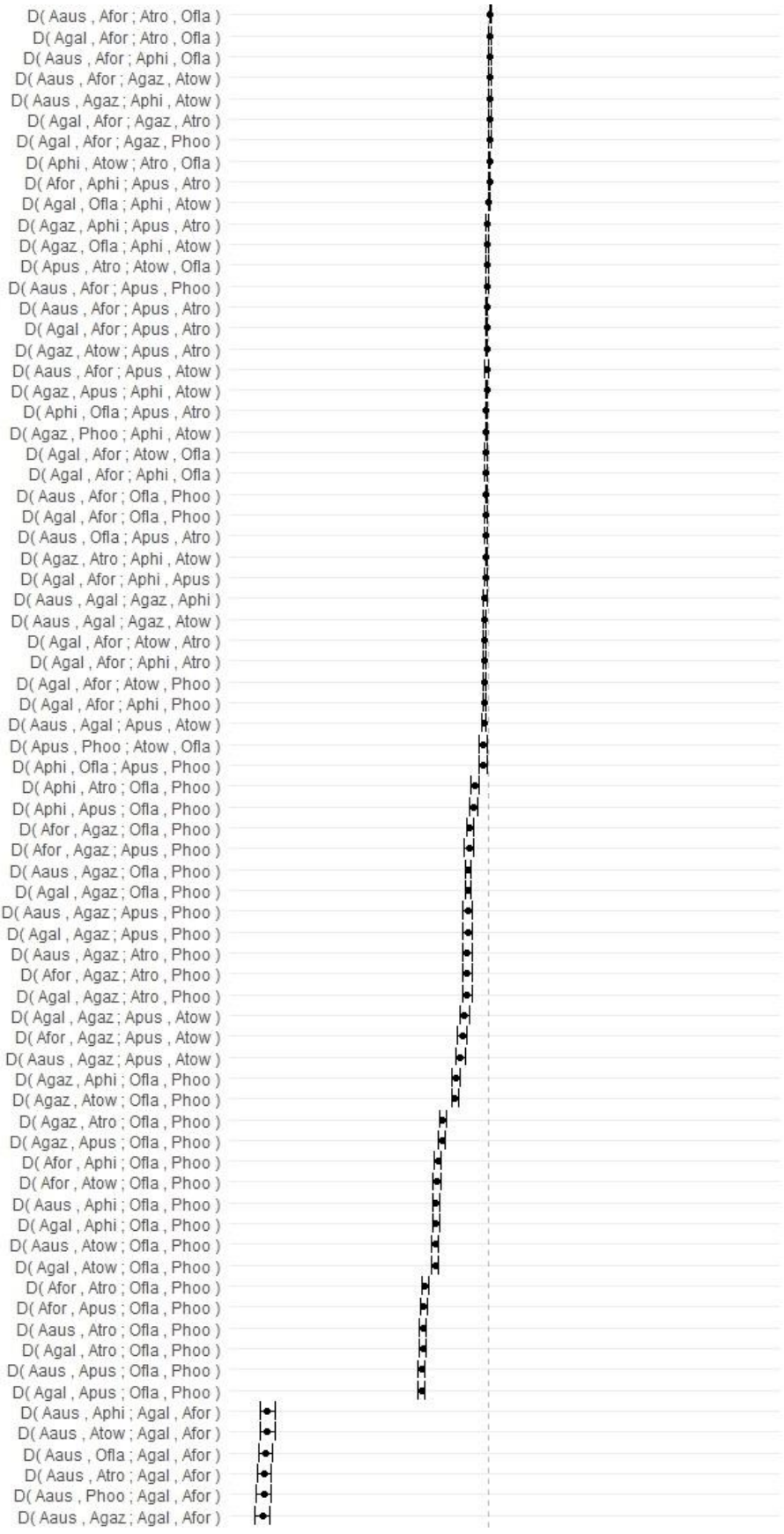


Figure S8. F_4-statistics distribution (x-axis) for all significant combination of four-taxa phylogenies (y-axis) with significant values (|Z-score| > 3).

Figure S8. Cont.



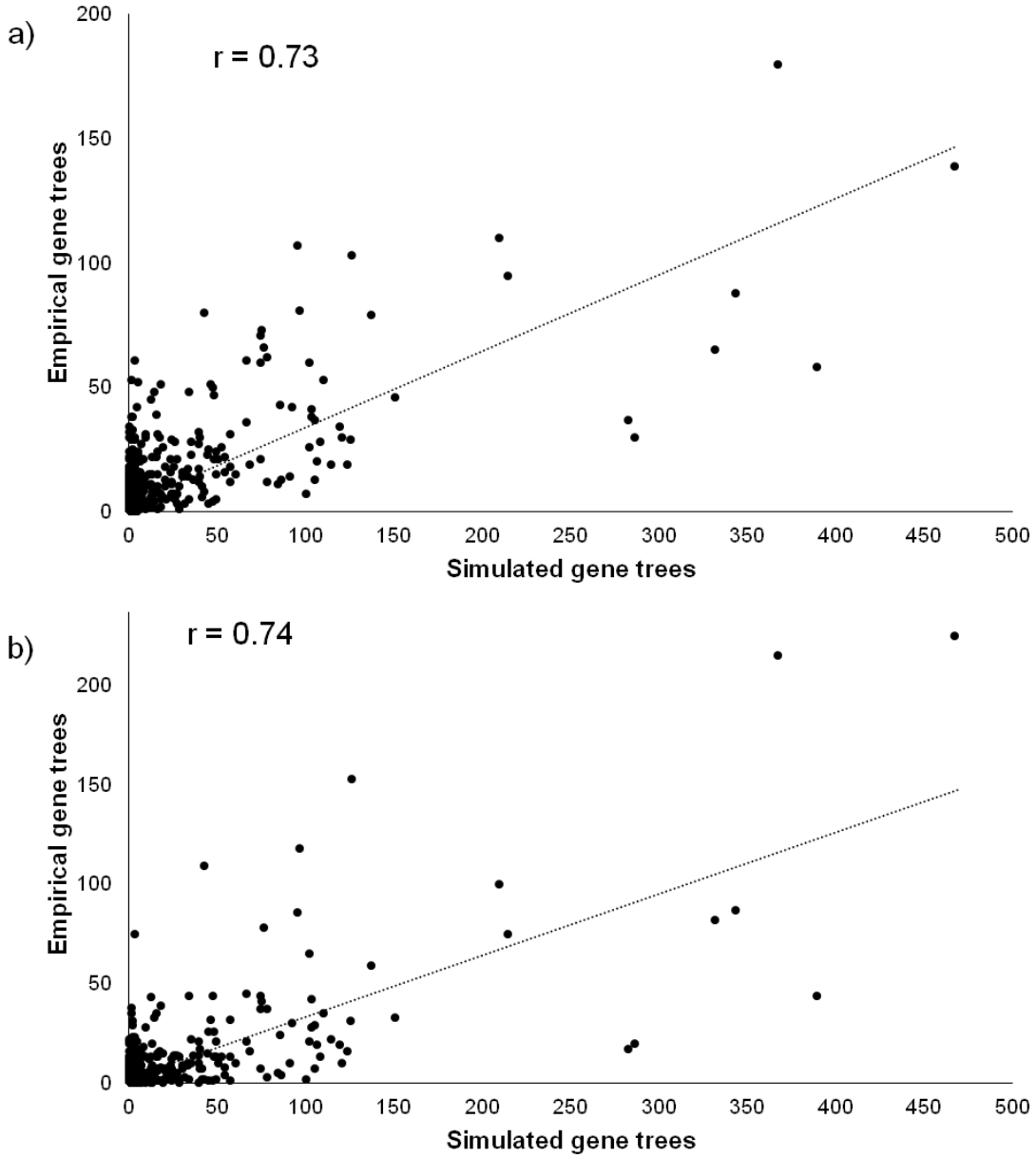
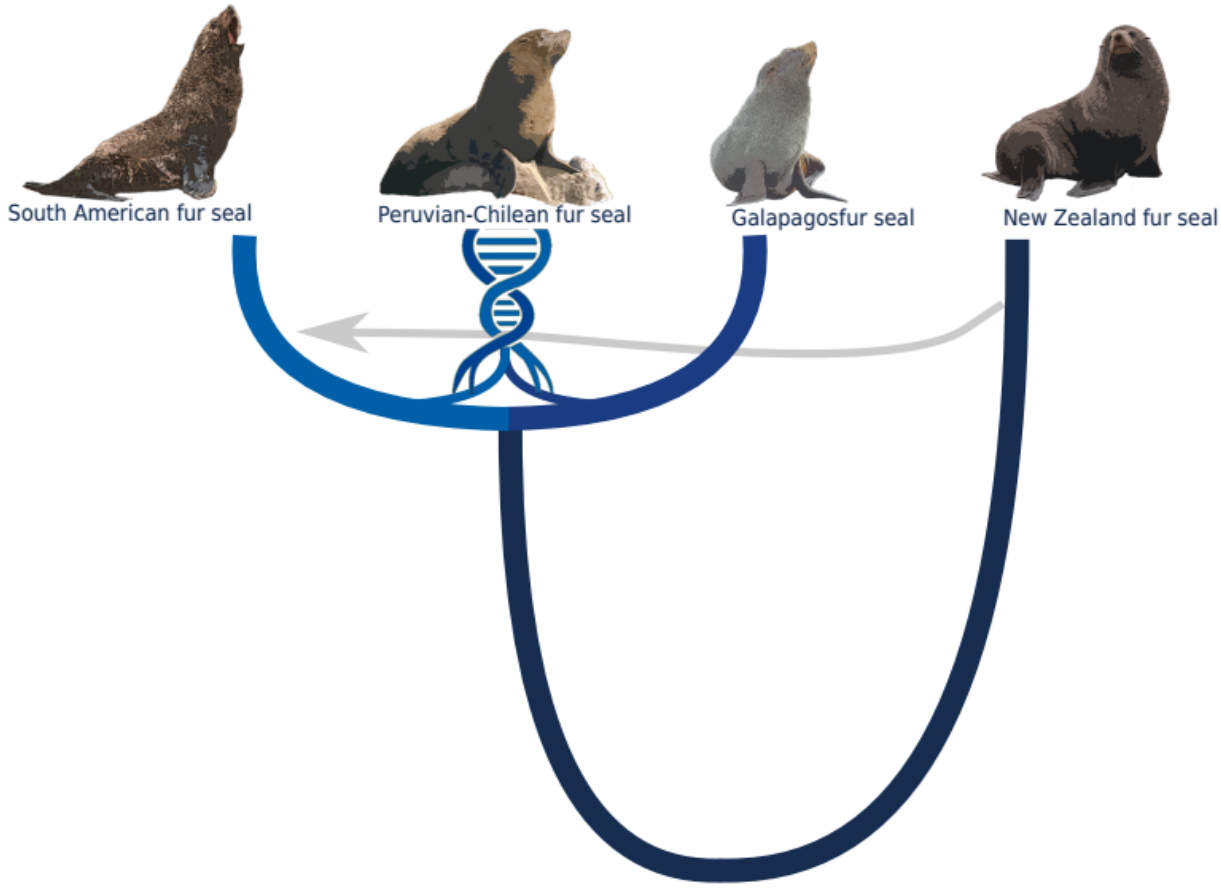


Figure S9. Topologies correlation. (a) 50 kb GFs and (b) 200 kb GFs.

Título:

There and there again: Genomic data indicate the Peruvian-Chilean fur seal is an ancient hybrid between the South American and the Galapagos species and that the latter is expanding and reproducing in the continent again

Artigo Científico abordando a história evolutiva das espécies do clado *Arctocephalus australis*, *A. galapagoensis* e *A. forsteri* com base em genomas completos e bibliotecas de representação reduzida ddRAD-seq. Formatado conforme as normas do periódico *Molecular Biology and Evolution*.



Article

Discoveries

There and there again: Genomic data indicate the Peruvian-Chilean fur seal is an ancient hybrid between the South American and the Galapagos species and that the latter is expanding and reproducing in the continent again

Fernando Lopes[1], Larissa Oliveira[2,3], Yago Beux[1], Amanda Kessler[1], Susana Cárdenas-Alayza[4], Patricia Majluf[4], Diego Páez-Rosas[5], Jaime Chaves[5], Enrique Crespo[6], Robert Brownell Jr.[7], Maritza Sepúlveda[8], Valentina Franco-Trecu[9], Carolina Loch[10], Bruce Robertson[11], Claire Peart[12], Jochen B. W. Wolf[12], Sandro L. Bonatto[1]

[1] Escola de Ciências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brasil
[2] Programa de Pós-graduação em Biologia, Universidade do Vale do Rio dos Sinos, São Leopoldo, Brasil
[3] GEMARS, Grupo de Estudos de Mamíferos Aquáticos do Rio Grande do Sul
[4] Centro para la Sostenibilidad Ambiental, Universidad Peruana Cayetano Heredia, Lima, Peru
[5] Colegio de Ciencias Biológicas y Ambientales, COCIBA, Universidad San Francisco de Quito, Quito, Ecuador
[6] CONICET, Centro Nacional Patagónico - CENPAT, Puerto Madryn, Argentina
[7] NOAA, National Oceanic and Atmospheric Administration, La Jolla, United States of America
[8] Departamento de Biología y Ciencias Ambientales, Universidad de Valparaiso, Valparaiso, Chile
[9] Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay
[10] Department of Oral Sciences, University of Otago, Dunedin, New Zealand
[11] Department of Zoology, University of Otago, Dunedin, New Zealand
[12] Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, München, Germany

Authors for correspondence: E-mail: slbonatto@pucrs.br; fernando.lopes@edu.pucrs.br
Phone number: +55(51)33534727
Ipiranga 6681, 12C 134. 90619-900. Porto Alegre, RS, Brazil

Keywords: Genomics, pinnipeds, RAD-seq, whole-genome, hybridizarion

Short running head: *The hybrid origin of the Peruvian-Chilean fur seals*

Abstract

Hybridization has been reported as an alternative evolutionary force contributing to adaptation and biological diversification, mainly with the increasing use of genomic data. Here, we used whole genomes and RADseq data to elucidate the evolutionary history of three related fur seals species: *Arctocephalus galapagoensis* (Galapagos fur seal), *A. forsteri* (New Zealand fur seal), and *A. australis* (South American fur seal), from which the Peruvian-Chilean population has been suggested to be a separated Evolutionary Significant Unit. Our phylogenomic results support the basal position of the New Zealand fur seals, a close relationship between the Galapagos and South American fur seals, and that the Peruvian-Chilean population is differentiated from both. Surprisingly, admixture analyses showed that the Peruvian-Chilean fur seal is a taxon of hybrid origin, still retaining in its genome around 50% of both parental species, the South American and Galapagos fur seals. The ABC analysis strongly supported that this hybrid speciation occurred thousands of generations ago. We also detected one admixed individual with the Galapagos fur seal in Punta San Juan, southern coast of Peru and another in Isla Foca, an isolated rookery located in the northern Peruvian coast. In the latter, we also found two pure Galapagos fur seals. Our study elucidates the importance of Peruvian-Chilean fur seals as an independent hybrid taxon and the enigmatic role of Isla Foca in the population dynamics, distribution and conservation of the Peruvian-Chilean and Galapagos fur seals.

Introduction

Hybridization occurs due to incomplete reproductive isolation when two populations or distinct species mate (Harisson 1990). The evolutionary consequences of this phenomenon can vary depending on the viability and fertility of the offspring. In some situations, the effects of interspecific hybridization extend further than the contact zone, causing the influx of foreign alleles from either or both parental species into the genome of other species, resulting in genomic introgression (Bullini 1994; Downling and Secor 1997). Postdivergence gene flow is common in vertebrates and has been reported for many groups such birds (Lamichhaney et al. 2018), fishes (Meier et al. 2018) and mammals (Cahill et al. 2018; Teng et al. 2017) being prevalent in fast-radiating groups (Mallet 2005, 2007).

One of the most extreme cases of hybridization is the hybrid speciation, that is quite common in plants but extremely rare in mammals. The hybrid speciation occurs when hybridization played a principal role in the origin of a new species or lineage (Mallet 2007). Among mammals, the hybrid speciation is mainly homoploid, meaning a formation of the new species via hybridization with no whole-genome duplication and no increase in ploidy (Schumer et al. 2018). The homoploid hybrid speciation is less frequent in nature because the genes and alleles are frequently overcome by gene flow from the parental species, difficulting its stabilization as a new lineage (Schumer et al. 2014; 2018). Despite this, new genetic recombinations, that emerge from parental alleles in the introgressed groups can lead to new phenotypes, which may be well-adapted to the environment, improving the fitness and giving an evolutionary advantage to the hybrids (Tobler and Carson 2010; Abbott and Rieseberg 2012; Abbott et al. 2013).

The spatial divergence is another important mechanism that can play an important role in the establishment of homoploid hybrids. Hybrids frequently have expanded ecological tolerance and are ecologically divergent from the parental species. Taking this into to account, the geographical

isolation provided by founder events might allow admixed genomes to become stabilized before they are prevailed by gene flow from parental species (Abbott and Rieseberg 2012).

One of the latest divergent clade of Otariidae encompasses three species: *Arctocephalus galapagoensis* (Galapagos fur seal), endemic to the Galapagos Archipelago; *A. australis* (South American and Peruvian-Chilean fur seals), widespread along the coast of South America; and *A. forsteri* (New Zealand fur seal), distributed in New Zealand and Southern Australia, (fig. 1A) (Chilvers and Goldsworthy 2015; Trillmich et al. 2015; Cárdenas-Alayza et al. 2016). According to the current literature, *A. australis* is distributed from Isla Foca (5°12'S; 81°12'W), in Northern Peru, to Mejillones, in the Northern coast of Chile (23°00'S; 70°20'W); and from Guafo Island (43°35'S; 74°42'W), Southern coast of Chile, to Ilha dos Lobos (29°20'S; 49°42'W), Southern coast of Brazil (Oliveira et al. 2008; Cárdenas-Alayza et al. 2016). This characterizes a gap of 2,000 km in the species distribution from Guafo Island to Mejillones (Rodrigues et al. 2018).

The fur seals and sea lions are charismatic species and well-known by the general public. However, from the point of view of the taxonomy and systematics, the internal relationships of *Arctocephalus* are controversial and have been extensively discussed (Sclater 1987; King 1954, 1969, 1983; Repenning et al. 1971; Yonezawa et al. 2009; Berta and Churchill 2012; Berta et al. 2018). Based on skull morphology, King (1954; 1969) suggested *A. forsteri* as a valid species and at least three subspecies of *A. australis*: *A. a. australis* (from the Falkland Islands); *A. a. gracilis* (from the mainland of South America); and *A. a. galapagoensis* (from Galapagos Islands). Repenning et al. (1971), however, analysed the skull morphometry and proposed New Zealand, Galapagos and South American fur seals, including the Falkland Islands, as full species. King (1983), revisited her taxonomic review and also recognized *A. galapagoensis* as a separate species, agreeing with Repenning et al. (1971). Based on published studies of morphology and genetics and the current distribution, Oliveira and Brownell (2014) suggested *A. a. gracilis* as a junior synonym of *A. a. australis*.

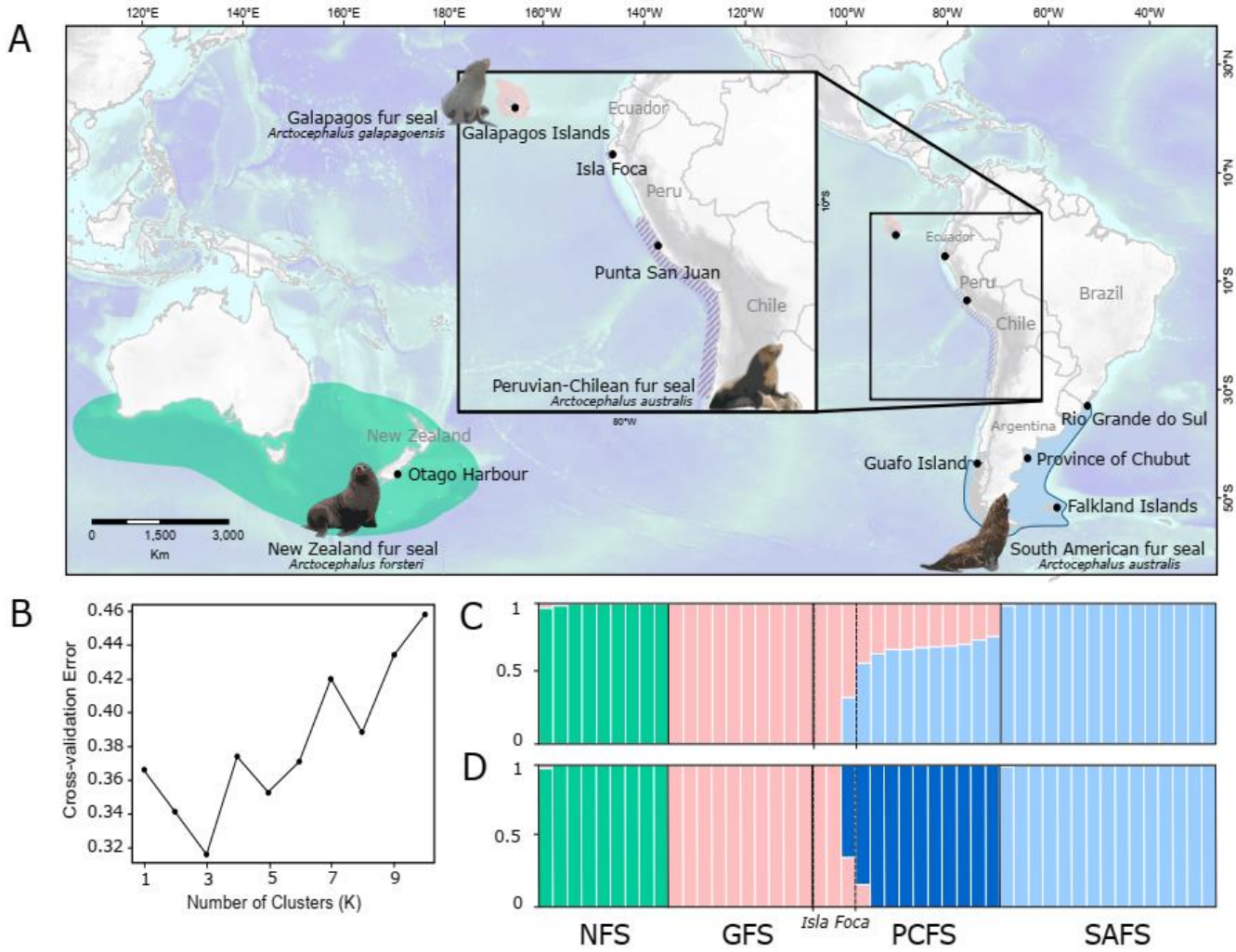


Figure 1. (A) Distribution and localities assessed in this study for each fur seal species. New Zealand fur seal (NZFS), green (n = 9); Galapagos fur seal (GFS), pink (n = 10), Peruvian-Chilean fur seal (PCFS), hatched blue and pink lines (n = 13); and South American fur seal (SAFS), light blue (n = 15). (B) Cross-validation chart showing the number of clusters and K=3 with lowest error rate. (C) Admixture graphic based on RADseq SNPs representing K=3. (D) Admixture graphic showing K=4 and Peruvian fur seal having its unique genomic component.

Mitochondrial phylogenies showed the paraphyly for this clade (Lento et al. 1997; Wynen et al. 2001; Yonezawa et al. 2009). By analyzing mtDNA cytochrome-b and estimating the divergence times using the mitogenome of all Otariidae, Yonezawa et al. (2009) found three mitochondrial lineages of *A. australis* and two lineages of *A. forsteri*, one of them grouping with *A. australis* from Peru and another one being the sister taxon of *A. australis* from Atlantic coast. The split time from the most common ancestor of this group with *A. gazella* was estimated ~1.4 Mya.

Berta and Churchill (2012), using previously published topologies for all pinnipeds, proposed a new taxonomic organization based on a combined phylogeny. To avoid issues of paraphyly, the authors suggested the genus *Arctocephalus* be limited to *A. pusillus*, the type species. Similarly, the

transference of remaining *Arctocephalus* to *Arctophoca* (Berta and Churchill 2012; Committee on Taxonomy 2018). In the same manuscript, these authors recommended the recognition of four subspecies of *Arctocephalus australis* (*Arctophoca australis*): (1) *A. a. australis* from the Falkland Islands, (2) an unnamed subspecies from Peru, (3) *A. a. gracilis* from Brazil and Argentina, and (4) *A. a. forsteri* from New Zealand. This classification is currently used by the International Union for Conservation of Nature (IUCN) in the red list of threatened species (Chillvers and Goldsworthy 2015; Cárdenas-Alayza et al. 2016). However, Nyakatura and Bininda-Emonds (2014), in an updating of Carnivora supertree, suggested that the use of *Arctophoca* could be premature due to the uncertainty about the phylogenetic relationships of many Otariidae species. The Committee on Taxonomy of Marine Mammals, that firstly supported Berta and Churchill (2012), returned to use *Arctocephalus* temporarily, but with no definition so far.

By analyzing 318 skulls with standard and geometric morphometrics, Oliveira et al. (2005, 2008) revealed pronounced geographic variation in size and shape of the skulls and a remarkable variation in sexual dimorphism of *A. australis* from Peru and Uruguay/Brazil. In the same study, the analyses of seven highly polymorphic microsatellite loci revealed highly significant differences in allele frequency between these two localities. The morphological and genetic variations associated with different breeding systems and significant variation on sexual dimorphism in the Peruvian-Chilean and Brazilian/Uruguayan population led these authors to suggest different Evolutionary Significant Units (ESUs) or even a different species (Oliveira et al. 2008; Oliveira and Brownell 2014). Túnez et al. (2007, 2013) and Rodrigues et al. (2018) also found significant genetic differences in partial mitochondrial DNA sequences between Pacific and Atlantic populations, corroborating with the findings of Oliveira et al. (2008). Significant differences in the reproduction times, resulting in an asynchrony of the breeding season, were also found between the two localities (Pavés et al. 2016). These differences were not assigned to demographic, oceanographic or photoperiod conditions in the breeding areas, neither to latitudinal effects of the breeding phenology, reinforcing the independent evolutionary histories of the two ESUs (Pavés et al. 2016).

Despite the complex taxonomical history, *A. forsteri*, *A. galapagoensis* and *A. australis* are currently recognized as full species, the last one with two subpopulations: the Peruvian-Chilean and the South American fur seal subpopulations (Cárdenas-Alayza et al. 2016; Committee on Taxonomy 2018).

Here, we generated and analysed six whole-genomes and reduced representation libraries of three species and its subpopulations, hereafter called by their common names: *A. forsteri* (New Zealand fur seals - NZFS), *A. galapagoensis* (Galapagos fur seal - GFS) and *A. australis* from two subpopulations (Peruvian-Chilean fur seal - PCFS and South American fur seal - SAFS). The analyses performed allowed us to delimitate species, compare different scenarios of speciation, assess signals of hybridization among different groups of fur seals, and detect the hybrid origin of the Peruvian-Chilean subpopulation.

Results

Sequencing and SNP Filtering Information

We successfully sequenced six whole-genomes of fur seals (n=1 per locality): for the South American fur seals from Argentina, Chile and Falklands (SAFS), and one from each the Peruvian-Chilean (PCFS), New Zealand (NZFS) and the Galapagos fur seals (GFS) (fig. 1A). The total average coverage, including the genome of Antarctic fur seal (AFS) raw reads retrieved from GenBank (n = 1), was 23.6X (± 6.2) (table S1). The filtered whole-genome SNP panel recovered 13,316,418 polymorphisms. The mean depth of the RADseq sequenced libraries was 49.5X (fig. S1). After the SNP calling and hard filtering, were kept 47 of 72 high-depth sequenced samples (fig. S1, S2 and table S2) with 3,198 polymorphisms (fig. S3).

Population Clustering and Structure

The pairwise F_{ST} with RADseq data indicated a low to moderate population differentiation (0.03-0.09) between the four SAFS localities. The differentiation between PCFS and the other

populations are much higher than between the SAFS localities (0.14-0.18). Between species divergences (GFS vs NZFS vs SAFS populations) are higher than the previous ones (0.20-0.32) (table S3).

Admixture analysis with RADseq data suggests three as the most appropriate number of genomic clusters for these populations (K=3; fig. 1B). The Admixture plot for K=3 showed samples of the GFS, NZFS and SAFS as well-delimited groups, while Peruvian-Chilean fur seals (sampled at Punta San Juan - PSJ) appeared as an admixed population between GFS and SAFS (fig. 1C), with a higher genomic component from the SAFS. However, for the three individuals sampled in the isolated Isla Foca (in northern Peru), while one presented a mixed ancestry similar to the PCFS samples (although with a higher proportion from GFS), the other two presented a pure GFS genomic component. When we set K=4 genomic clusters, all but one of the Peruvian individuals from PSJ are pure for their own genetic component. Again, two individuals from Isla Foca were assigned as pure GFS and the other one as mixed between the Peruvian and Galapagos fur seals (fig. 1D).

The PCA analysis of the whole genomes presented the three species well separated, with the three SAFS individuals from Argentina, Falklands and Chile closely together (Fig 2B). The individual from the Peruvian (PSJ) population was found in an almost intermediate position between the SASF

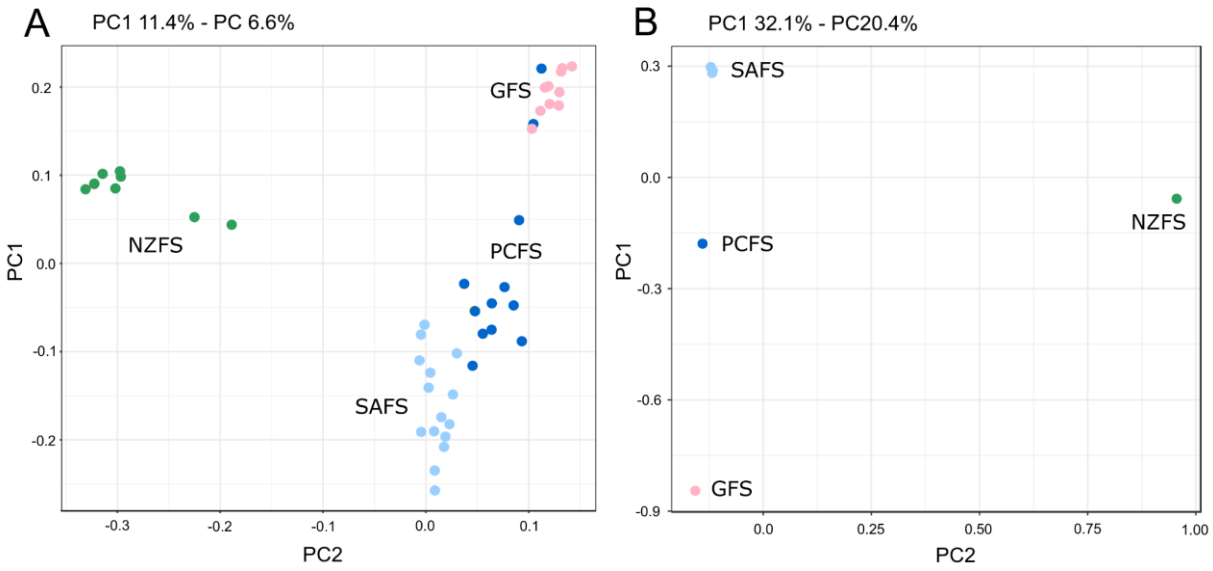


Figure 2. (A) PCA for the ddRAD dataset and (B) whole-genomes.

and GSF. Results of the PCA analysis from RADseq SNPs with all samples were very similar to the admixture results (Fig. 2A). The species were well separated in clusters, with samples from the Peruvian population slightly close to the SAFS. Again, two samples from the Peruvian Isla Foca grouped inside the GFS cluster and a third was in an intermediary position between this taxon and PCFS of PSJ. A more detailed PCA from SAFS *sensu lato*, the Peruvian cluster and distinguished samples from Isla Foca is shown in the supplementary figure S4.

Genomic Introgression Events

Since the results of admixture and PCA analyses with RADseq clearly showed the Peruvian population as admixed, we applied the f_3 and f_4-statistics to the whole genomes to investigate further signals of past hybridization. The f_3 analysis showed strong and significant evidence ($f_3 < 0$) that PCFS is the result of admixture between SAFS and GFS, with all three SAFS localities presenting virtual identical results. The combination with NZFS and GFS also showed evidence for admixture, however with a weaker signal (fig. S5 and table S4).

The f_4 analysis, that estimates unrooted four-population phylogenies to visualize shared genetic drift among taxa, also indicated PCFS as an admixed taxon (fig. S5 and table S5). The f_4 results strongly rejected the proposed null hypothesis when PCFS population is the sister taxon of SAFS in a four-taxa phylogeny, indicating a shared drift pathway with GFS. It also rejects the null hypothesis when PCFS is the sister taxon of GFS, pointing to a shared drift pathway with SAFS. We also found weak signals of introgression between NZFS and SAFS ($f_4 < 0.003$) and between AFS and SAFS ($f_4 > -0.001$) (fig. S6).

We used TreeMix to estimate with whole genome data the phylogenetic relationship of populations and species, including admixture events. In a first analysis in which we maintained each SAFS individual as a separate terminal, PCFS grouped with the Chilean SAFS, but very close to the individuals from AR and FK, which are indistinguishable (fig. 3A). The analysis also found that 0.49 of the PCFS genome came from introgression from GFS. Because of the undifferentiation between

AR and FK, we ran a second analysis joining these two individuals in an Atlantic population of SAFS (fig. 3B). Interestingly, now the PCFS grouped with GFS, but with 0.48 of the genome introgressed from the Chilean SASF. These results suggested that PCFS was essentially a hybrid between SAFS and GFS, with half of its genome received from each ancestral. The result of the Admixturegraph (fig. S7) found the same results, estimating the admixture proportion PCFS with GFS at 0.46 and with Chilean SAFS of 0.54. TreeMix results found an introgression from NZFS into SAFS of about 4-10% and a large branch length for PCFS (fig. 3), also a signal of an admixed population. Finally, in the analysis with AR and FK separated, there was a signal of a large admixture (0.31) between FK and CH individuals, which we interpreted as indicating the close relationship between the three SAFS.

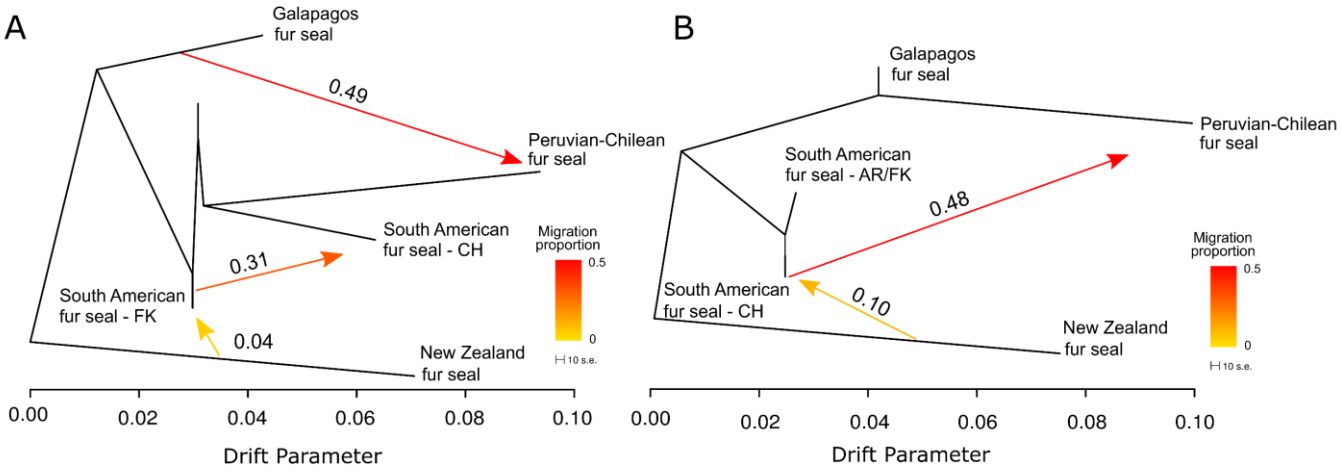


Figure 3. Maximum Likelihood tree of TreeMix with two migration events. (A) Showing the Peruvian-Chilean fur seal grouping with South American of Chile (CH) and AR (Argentina) and FK (Falklands) as independent branches. (B) Showing the Peruvian-Chilean fur seal grouping with Galapagos fur seal when AR and FK are considered from the same population. Both analyses show a migration of ~50% from the genome of the Galapagos fur seal or South American fur seal to Peruvian-Chilean fur seal.

Test of scenarios

We used an ABC approach to test different scenarios for the origin of the Peruvian-Chilean taxon using the RADseq data. The results supported the hybrid origin of PCFS from GFS and SAFS with a posterior probability of 1 against the two alternative pure divergence scenario (fig. S8A and 8C). We also estimated with this approach the age of the hybridization and the contribution of the two parental genomes to the PCFS. The mode value of the ratio of mixture event ('ra') in the PCFS was 61% for SAFS and 39% for GFS. The posterior distribution of the $\tau 1$ (the time when PCFS arose

by hybridization) showed that the event of admixture is not recent but have occurred likely thousands of generations ago. However, the absolute estimation for divergence times and effective population sizes should not be considered here given the absence of mutation model parameterization in the SNPs coalescence model in DIYABC (Cornuet et al. 2014).

Phylogenomic Inferences in the Presence of Gene Flow

Given the scenarios with different events of genomic introgression, we used G-PHoCS to estimate divergence times and rates of migration in the presence of gene flow. This analysis estimated almost simultaneous speciation of SAFS, PCFS and GFS that occurred between 352 and 403 Kya, with small and overlapped intervals, after the split from NZFS at 1.5 Mya (fig. 4). Although estimates of the migration parameters did not stabilize well, resulting in large confidence intervals, the median value that represents the proportion of the genome of the PCFS received from the GFS was 0.48, the same values estimated above with TreeMix and Admixturegraph. A large and similar *Ne* for each extant species and the ancestor population were also estimated (fig. 4 and table S6). The PCFS showed the smallest *Ne*, however, with the largest HPD interval, and NZFS showed the largest *Ne* (table S6).

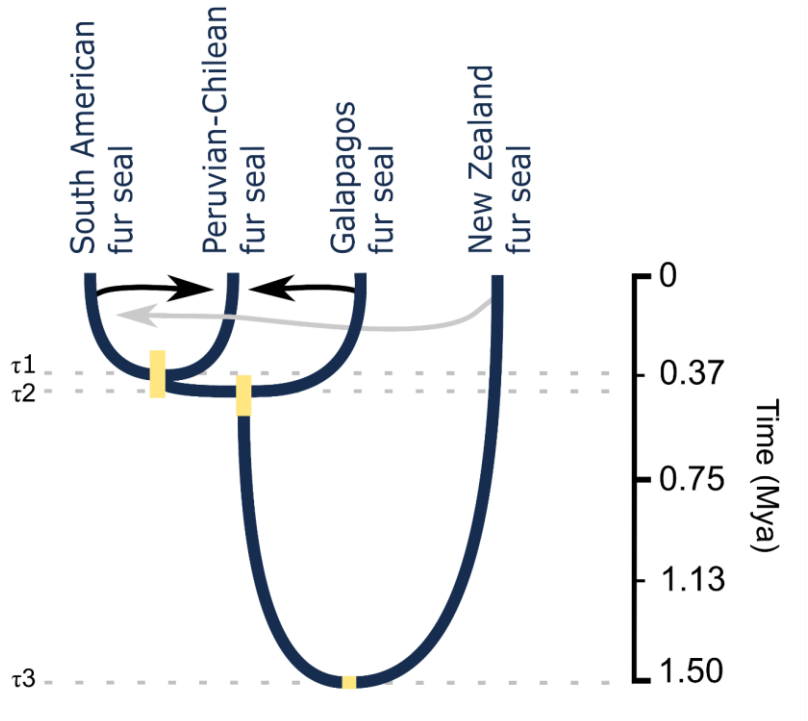


Figure 4. Divergence times recovered with G-PHoCS showing the fast period of speciation of the SAFS, PCFS and GFS.

Population size changes through time

The demographic histories of five fur seals genomes with the PSMC analysis were presented in figure 5 and S9. The SAFS of the Atlantic Ocean (Argentina and Falklands) presented almost identical trajectories, while the SAFS from Chile presented some differences. GFS showed a long-term distinct evolutionary trajectory, frequently presenting smaller *Ne* than the other fur seals. PCFS presented a history like the other SAFS differing more clearly only more recently indicating the beginning of an independent evolutionary history from the other South American fur seals.

Mitochondrial DNA phylogeny

To compare our data with the previous phylogenies published in the literature, we also

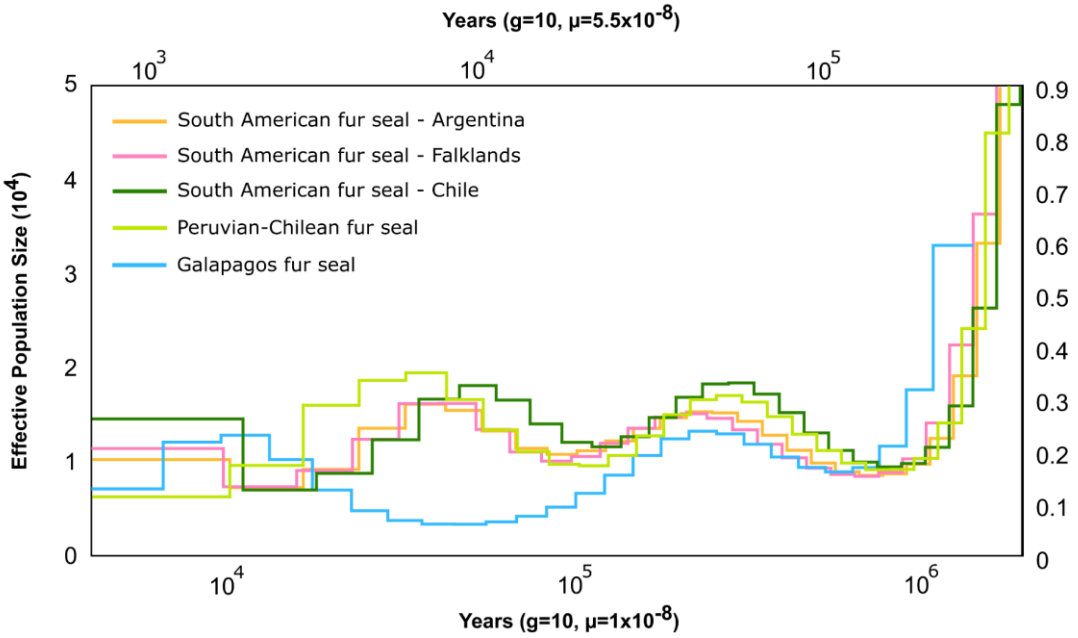


Figure 5. Historical *Ne* using the PSMC analyses for five fur seal genomes. The x-axis shows the time, and the y-axis shows *Ne*. The left and bottom axis show the plot scaled using a conservative mammalian mutation rate (μ) of 1 × 10^{-8} bp^{-1} gen^{-1} for a generation time (g) of 10 years. The above and right axis show the plot scaled using a mutation rate estimated previously from 300 genomic fragments as 5.5 × 10^{-8} bp^{-1} gen^{-1} (Lopes et al. in prep).

sequenced the control region of the mtDNA from the samples used in the RADseq analysis. Our phylogeny showed a very similar topology to those already reported (fig. S10, table S7 and S8). However, samples from PCFS appeared grouped in two separated clades. The major lineage was derived from a SAFS clade (PCFS Clade A), and the minor (clade B) as a sister group of the largest lineage of the NZFS (clade A). The NZFS sequences also appeared separated in two clades, being

basal to the group of PCFS (Clade A) and SAFS. Two of three individuals sampled in Isla Foca (Aaus11 and Aaus53), that were assigned to GFS in the Admixture analyses of RADseq, also clustered with the Galapagos species in the mtDNA phylogeny. The third specimen, with the nuclear DNA identified as a mixed individual between SAFS and GFS, clustered with PCFS (clade A) (fig. S10 and table S8).

Discussion

We presented a detailed analysis of the evolutionary history of the Peruvian-Chilean, South American, Galapagos and New Zealand fur seals using both whole genome and RADseq data. We detected at least three events of interspecific gene flow, including the hybrid speciation of the Peruvian-Chilean fur seals, shedding light on the group diversification process and species delimitation.

Interspecific gene flow and the hybrid origin of the Peruvian-Chilean fur seals

Besides the hybridization that originated the Peruvian-Chilean fur seals, we also found evidence for small but significant past introgression between the New Zealand and South American fur seals and between the Antarctic and the South American fur seals (figs. 3, S5-S7) and a secondary contact of the Galapagos with the Peruvian-Chilean fur seals (fig. 1C-D). Signals for past introgressive hybridization have been identified with genomic data in many mammal groups, such as humans, bears, big cats, whales, stoats (Figueiró et al. 2017; Kumar et al. 2017; Árnason et al. 2018; Cahill et al. 2018; Colella et al. 2018). In pinnipeds (fur seals, sea lions, phocids and walruses) there are more than 1,100 cases of hybrid individuals registered, of which 19.2 % involved New Zealand fur seals (Schaurich et al. 2012). No cases involving South American, Peruvian-Chilean or Galapagos fur seals were registered in the literature until now. We have previously found genomic evidence of past hybridization in Otariidae only between New Zealand and South American fur seals (Lopes et al. in prep.).

Gene flow between species of fur seals that are distributed far away from each other is plausible. The species of the group are highly dispersive and influenced by environmental conditions, such as marine currents and oceanographic phenomena, they can swim long distances, sometimes more than 5,000 km from their reproductive sites (Páez-Rosas et al. 2017; Millman et al. 2019). For example, vagrants Galapagos fur seals were already recorded during *El Niño* events in the coasts of Mexico, Costa Rica, Guatemala, Colombia and Ecuador, however, with no register of offspring (Capella et al. 2002; Aurioles-Gamboa et al. 2004; Félix et al. 2007; Montero-Cordero et al. 2010; Páez-Rosas et al. 2017; Quintana-Rizzo et al. 2017). In the Atlantic Ocean, the northernmost record of an *Arctocephalus* sp. occurred in the São Pedro and São Paulo archipelago (00°56'N, 29°22'W), which is about 1000 km away from the Brazilian and Africa mainland. This record could correspond to three potential species: *Arctocephalus tropicalis*, the Subantarctic fur seal, *Arctocephalus pusillus pusillus*, South African fur seal and *A. australis*, the South American fur seal (Millman et al. 2019). Thus, the great dispersal capability, agonistic behavior, great dispersal capability and weak mechanisms of reproductive isolation can be promoting interspecific gene flow in fur seals (Carey 1992; Amos 2007; Lancaster et al. 2007; Schaurich et al. 2012; Millman et al. 2019).

The hybrid speciation of the Peruvian-Chilean fur seals was strongly supported by the test of scenarios of ABC (fig. S8) and evidenced by multiple analyses of genomic admixture (figs. 1-3, S5-S8). The proportions of introgression of each parental species in the hybrid taxon genome in most of the results are closer to 50%-50%. The variation in these estimates between some analyses, that ranged from around 40 to 60%, most likely reflects the different phylogenetic relationships between the three species underlying the analyses. For example, in the TreeMix analyses (figs. 3 and S7), we found the Peruvian-Chilean fur seals grouping either with Galapagos or South American fur seals. This is expected since these analyses first estimated a dichotomous tree and afterwards estimate migration between them. Being a hybrid taxon with almost identical contribution from both parental species, parental taxa will group first by minor differences in the algorithms or the data set. The hybrid taxon also has its genetic diversity increased due to the influx of alleles from the different parental

species (Zalapa et al. 2010), which supports the longest branch lengths of the Peruvian-Chilean fur seals in the species tree phylogenetic reconstruction (fig. S11 and supplementary material S1) and drift parameter of TreeMix (figs. 3 and S7). The test of scenarios with ABC (fig. S8) also rejected scenarios of divergence without admixture between these three taxa, including the hypothesis that the Peruvian-Chilean fur seal is the sister taxon of South American fur seals (Wynen et al. 2001; Yonezawa et al. 2009).

We also showed that there seems to have been a recent secondary contact between Galapagos and Peruvian-Chilean fur seals. We found one individual in Punta San Juan, in the Southern coast of Peru, with mixed ancestry with Galapagos fur seals (fig. 1D), suggesting at least one event of successful offspring between Galapagos and Peruvian-Chilean fur seals. More interestingly, we found two pure Galapagos fur seals and another mixed (mostly with Galapagos ancestry) individual in Isla Foca, an islet very close to the continent (<1 km), located ~1,000 km to the north of Peruvian-Chilean fur seals distribution and to the east of the closest Galapagos island. The presence of these individuals in the coast of Peru indicates a new and complex scenario for the Galapagos species until now considered endemic to the Galapagos Archipelago (Trillmich 2015). Isla Foca was recognized as an extension of the Peruvian-Chilean fur seals distribution only in the last update of the IUCN red list of threatened species (Cárdenas-Alayza et al. 2016). Although our sampling is very limited, our results are suggestive that this colony may represent an established Galapagos fur seal colony at the South American coast, showing a putative zone of contact between the two taxa.

In the last census there were ~50 fur seals living on the islet, but this number vary from year to year (Cárdenas-Alayza and Oliveira, personal communication), mainly during *El Niño* events, the driving force that leads to strong demographic oscillations in the Galapagos and Peruvian-Chilean fur seals populations due to the lack of primary productivity affecting the food chain in the Equatorial Pacific ocean (Oliveira et al. 2008, 2009; Trillmich 2015). Unfortunately, there is not much biological information available about the islet (Cárdenas-Alayza and Oliveira, personal communication) and further studies are needed for understanding the role of Isla Foca and the potential effect of

environmental anomalies, such as climate change and *El Niño* events (see Oliveira et al. 2006; Oliveira et al. 2009), in this locality and the consequences for the Galapagos and Peruvian-Chilean fur seals population dynamics, distribution and conservation.

Species delimitation

Until now, phylogenetic relationships, biogeographical histories and even taxonomic reclassification within Otariidae and *Arctocephalus* were inferred based on small mtDNA sequences (Lento et al. 1997; Wynen et al. 2001; Árnason et al. 2006; Higdon et al. 2007; Yonezawa et al. 2009; Berta and Churchill 2012; Rodrigues et al. 2018). Among these studies, those that applied more than one individual per species supported the paraphyly for the crown group assessed in our study (e.g., Wynen et al. 2001; Yonezawa et al. 2009).

The species trees recovered with genomic fragments showed the same topology with the Antarctic fur seal being basal to the group, followed by the New Zealand fur seal and Galapagos and Peruvian-Chilean fur seal as the sister clade of the South American fur seal (fig. S11). However, the phylogenetic position and divergence time of the Peruvian-Chilean taxon in these analyses should be considered with caution, since these methods do not deal with the effect of introgression (Wen and Nakhleh 2018). Simulated data have shown biased estimates for node heights and population sizes under multispecies coalescent models when migration between non-sister species is present but ignored (Leaché et al. 2014; Wen and Nakhleh 2018). We, therefore, used the G-PHoCS approach, a Bayesian method that incorporates interspecific gene flow in the estimation of time, population sizes and migration. This analysis indicated more recent divergence times (fig. 4) than the those published in the literature (Higdon et al. 2007; Yonezawa et al. 2009; Nyakatura and Bininda-Emonds 2012) and recovered in the MSC models of StarBEAST and ASTRAL (see fig. S11 and supplementary material S1), showing an almost simultaneous speciation for the Galapagos, Peruvian-Chilean and South American fur seals around 400 Kya (fig. 4). Similar divergence times were found in the split

of South American and Galapagos fur seals in our previous whole genome study of Otariidae (Lopes et al. in prep.).

The topology estimated with mtDNA (fig. S10) agreed with the paraphyly of the crown group found in previous studies (Lento et al. 1997; Wynen et al. 2001; Yonezawa et al. 2009; Rodrigues et al. 2018). Our RADseq results, however, showed that the three species are well-delimited based on the nuclear genome. Given the mode of inheritance of the mitochondrial marker, the paraphyly reported can also reveal an ancient and complex event of origin and dispersal of the crown group (Yonezawa et al. 2009; Salis et al. 2016; Dussex et al. 2016; Rodrigues et al. 2018). Therefore, previous biogeographical inferences based only on the mitochondrial gene tree history should be revisited. To untangle this complex evolutionary history at a finer scale would be necessary additional genomes.

Previous studies on morphology, genetics, and reproduction biology (Oliveira et al. 2005, 2008; Rodrigues et al. 2018; Pavés et al. 2016) found the Peruvian-Chilean fur seal to be a very divergent taxon within *A. australis*, with at least, a status of ESU. Besides, at least three additional pieces of evidence supported a long-term evolutionary history of the Peruvian-Chilean fur seals as an independent population. First, in the admixture graph with K=4 genomic clusters, this population presented its unique genomic component (fig. 1D). Second, the mitochondrial DNA tree showed two very distinct lineages (fig. S10) with long branches for this population, one derived from South American fur seals (clade A) and another clustering with the major lineage of the New Zealand fur seals. These two clades were also found by Rodrigues et al. (2018), evidencing a complex scenario in the crown group origin and dispersal and ancient evolutionary history. Third, the test of scenarios of ABC also supported an ancient history with the hybridization estimated to have occurred thousands of generations ago (see results and fig. S8). Therefore, our discovery that the Peruvian-Chilean fur seal originated from an ancient hybridization and that it has its own genetic component but that it still preserves its mixed origin, implies that, in evolutionary terms, it could not be considered as being a subgroup (e.g., an ESU or subspecies) of only one of its parental species. Consequently, the correct

biological and taxonomic solution would be to grant this taxon with a full species status within *Arctocephalus* as previously suggested by Oliveira and Brownell (2014).

Our results also showed a lack of population structure among *A. australis* main populations (Chilean, Argentinean and from Falklands) or even between Southern Pacific and Atlantic oceans (Oliveira et al., unpublished data; Rodrigues et al. 2018), not supporting the three mitochondrial clusters within the Uruguayan/Brazilian ESUs found by Túnez et al. (2007; 2013).

Materials and Methods

Sampling

A total of 72 tissue samples were collected representing the species distribution of SAFS, PCFS, GFS and NZFS. The effective number of samples (after filters) and localities for each species are described in figure 1 and table S1. Skin biopsies (~0.5 cm^3) were obtained with piglet notch pliers from pups or from recently dead animals found ashore, stored in ethanol 70% and cryo-preserved in -20 ºC. Genomic DNA extractions were carried out with DNeasy Tissue Kit (Qiagen) following the manufacturer's protocol. DNA yield was quantified using a spectrophotometer NanoDrop Lite, and quality assessed on a 2% agarose gel.

Whole-genomes libraries and Sequencing

We sequenced six whole-genomes of fur seals (n=1 per locality): for the South American fur seals from Argentina, Chile and Falklands (SAFS), and one from each the Peruvian-Chilean (PCFS), New Zealand (NZFS) and the Galapagos fur seals (GFS) (fig. 1A). Genomic libraries were prepared with Illumina DNA PCR-free or TruSeq Nano kits with an insert size of 350 bp and 150 bp of paired-end reads. The libraries (two libraries/lane) were sequenced by Macrogen Inc. (Seoul, South Korea) on an Illumina HiSeq X platform.

ddRAD-seq Library and Sequencing

ddRAD-seq libraries were prepared in-house using the version of Kess et al. (2016), a modified protocol of Peterson et al. (2012). Before the library preparation, we performed an *in-silico* digest based on the reference genome of Antarctic fur seal (AFS), *A. gazella* (Humble et al. 2016) and using the python script designed by DaCosta and Sorenson (2014) to ensure a sufficient number of RAD loci with satisfactory coverage to properly address the research question being investigated. The enzymes *SphI* and *Eco-RI* were selected, aiming ~15,000 genomic loci and 50x of coverage.

For each sample, 20 ng of genomic DNA was digested with restriction enzymes. P1 and P2 barcode combinations were ligated to the generated fragments for individual identification before samples were pooled. DNA fragments between 355 and 555 base pairs (bp) were selected using 1.5% agarose gel electrophoresis. These fragments were purified with Invitrogen PureLink Quick Gel Extraction Kit and magnetic bead purification to remove remained fragments below 250 bp. Genomic library was sequenced in a platform Illumina HiSeq 2500 using single-end 2x100 bp reads module in The Centre of Applied Genomics of The Hospital for Sick Children (Toronto, Canada). Stacks 1.46 (Catchen et al. 2011) was used for demultiplexing.

Partial mitochondrial DNA Sequencing

For identifying the matrilineal inheritance of the PCFS (see Results), we sequenced the mitochondrial DNA (mtDNA) control region (CR) using the following primers: L15926 5´- TCA AAG CTT ACA CCA GTC TTG TAA ACC - 3´ (Kocher et al., 1989); H16498 5´- CCT GAA GTA GGA ACC AGA TG- 3´ (Meyer et al., 1990). Amplifications for were carried out in 20 µl with the following conditions: 1.5 mM MgCl2, 200 µM of each dNTP, 0.1 µM of each primer, 1 U of Platinum Taq DNA polymerase (Invitrogen), 1X PCR buffer (Invitrogen), 0.2% - 0.4% Triton and 2 µl of DNA (approximately 50 ng). Thermocycling conditions were: 1 cycle of 5 min at 94 °C, 35 cycles of 30 s at 94 °C, 60 s at 65 °C and 60 s at 72 °C; and 1 final cycle of 7 min at 72 °C. Amplification products were purified with shrimp alkaline phosphatase and exonuclease I (Amersham Biosciences). The

purified were sequenced in both directions using and run in an ABI 3730XLs at Macrogen Inc. Partial mtDNA CR of the PCFS were checked in CHROMASPRO 1.7.4 (http://technelysium.com.au).

Data Quality Control and Mapping

Data quality was assessed with FastQC software (Andrews 2010). Reads were trimmed for vestigial adapters, mapped against masked genomes of Antarctic fur seal, *A. gazella*, (Humble et al. 2018) for RADseq and the annotated genome of the Walrus, *O. rosmarus*, for whole-genomes (Foote et al. 2015), and locally realigned using the bam_pipeline module implemented on PALEOMIX 1.2.13.2 (Schubert et al. 2014). The detailed PALEOMIX bam_pipeline and settings: for RADseq and whole-genome sequenced libraries we used the parameters, respectively: reads with length-size < 60/100 bp and Phred-score < 20/30 were filtered out by AdapterRemoval v2 (Schubert et al. 2016); the remained reads were mapped using BWA 0.7.17 (Li and Durbin 2009) and -mem algorithm. Single/Paired-end reads (RADseq and whole-genomes, respectively) with mapping quality Phred-score < 20, unmapped reads and single-reads (only for whole-genomes) were discarded from the downstream pipeline; reads that were sequenced more than double or less than one standard deviation from the coverage average were not considered and were removed (Arnold et al. 2013; Gautier et al. 2013). PCR duplicates were detected and removed by Picard Tools 2.18.5 (broadinstitute.github.io/picard/). Then, miscalling indels were locally realigned by GATK 3.8 (McKenna et al. 2010).

To assemble loci, call and filter SNPs, we used the program ANGSD 0.929 (Korneliussen et al. 2014). Detailed parameters used for filtering and mapping RADseq and whole-genomes reads are described in supplementary material S2. For whole-genomes, we used only sites with 100% of information; for RADseq, we removed all individuals and sites with less than 70% of genomic information (samples and sites). The distribution of missing data within the SNP panel was graphically represented with Matrix Condenser (de Medeiros and Farell 2018). For admixture

analyzes, to minimize the genetic linkage and consider all SNP as independent loci, we thinned the matrix of SNPs to a single SNP per contig with VCFtools (Danecek et al. 2011).

Population Clustering and Structure

We performed exploratory analyses to assess how individuals and genomes cluster together with a Principal Component Analyzes (PCA) using PCAngsd for RADseq and whole-genome datasets (Meisner and Albrechtsen 2018). For the RADseq dataset and based on PCA results, we used ANGSD Site Frequency Spectrum to assess the weighted pairwise F_{ST} population differentiation.

With the Bayesian clustering implemented in the NGSadmix ANGSD (Meisner and Albrechtsen 2018) and Admixture 1.3 (Alexander et al. 2009) (supplementary material S3), we evaluated the admixture proportions among individuals and populations/species. The best-fitted number of the genetic cluster was identified with cross-validation and ten independent replicates calculated in Admixture (from 1 to 10 clusters). The results were graphically visualized with Pophelper 2.2.9 (Francis 2017).

Admixture analyses

To estimate introgression between taxa based on whole genomes we used f_3 *and* f_4-statistics with threepop and fourpop modules implemented on TreeMix suite (Pickrell and Pritchard 2012; Harris and DeGiorgio 2016). F_3-statistics explicitly tests whether a taxon of interest *A* is the result of admixture between two other *B* and *C.* Positive values indicate that there is no evidence of admixture among tested species, while negative values indicate admixture; f_4-statistics estimates unrooted four-population phylogenies to visualize shared genetic drift among taxa (Harris and DeGiorgio 2016). For a f_4 ((A,B),(C,D)) topology, the defined paths are *A* to *B* and *C* to *D*, this results in $f_4 = 0$. If f_4 is a positive value, this indicates drift paths *A* to *C* and *B* to *D*. Otherwise, a negative value indicates *A* to *D* and *B* to *C*. TreeMix infers a maximum likelihood tree of populations with a given number of events of introgression.

The analyzes of threepop and fourpop were performed over 811 blocks of 5,000 SNPs. The significance of f_3 and f_4-statistics, respectively, were based on Z-score, calculated by the jackknife method. Significantly positive (Z > 3) and significantly negative (Z < -3) values reject the null hypothesis (Harris and DeGiorgio 2016). We plotted the distribution of f_3 and f_4 and values with admixturegraph (Leppälä et al. 2017), an R package.

Test of hybridization scenarios

We used approximate Bayesian computation (ABC) method implemented on DIYABC 2.1.0 (Cornuet et al. 2014) to test the most likely demographic scenario for the origin of PCFS with the RADseq SNP panel. In this analysis, the NZFS was excluded to maintain simply the demographic scenarios and improve the efficiency of the estimation. We tested three scenarios (Fig. S10) that represent some of our results and phylogenetic relationships previously reported in the literature: (1) a scenario where PCFS originated as a hybrid between SAFS and GFS, as suggested by our results; (2) PCFS as the sister group of SAFS, as the literature (Wynen et al. 2001; Oliveira et al. 2008) and some of our results; (3) PCFS as the sister group of GFS, as our ASTRAL-III and StarBEAST2 species trees suggested (supplementary material S1). Scenarios 2 and 3 are pure divergence scenarios, without admixture, to avoid that the three scenarios converge to a single scenario. Distribution of the priors for the parameters was presented in figure S8. All available summary statistics parameters of genetic diversity, F_{ST} distances and Nei's distances were used. We ran several preliminary analyses of 500,000 simulations for each scenario to adjusting the distributions of the parameters. After the draft analyses, we performed a final run of 5,000,000 simulations for each evolutionary model. Scenarios pre-evaluation and model checking was performed with PCA within DIYABC. To compute the posterior probability of the best scenario, we used the logistic regression with 1% closest similarity to the observed data and number of intermediate values = 10.

Phylogenomic Inferences in the Presence of Gene Flow with G-PHoCS

To infer migration rates, divergence times and population sizes (*Ne*) with the whole genomes, we used the Generalized Phylogenetic Coalescent Sampler method (G-PHoCS - Gronau et al. 2011). To minimize the effect of genetic recombination, the analysis was done over 20,000 intergenic and random neutral loci of 5 kb, 900,000 MCMC iterations, and NZFS as root (S4 and fig. S11). Based on the previous analysis we applied the following migration bands: GFS -> PCFS; SAFS->PCFS and NZFS->SAFS. We ran the analysis twice to ensure the convergence of results.

The pipeline used for selecting neutral loci and information of parameters estimation is based on Freedman et al. (2014) and Gronau et al. (2011). For detailed information, see supplementary material S4. The divergence times were calibrated using the time of divergence between NZFS and SAFS, PCFS and GFS estimated with StarBEAST2 (Ogilvie et al. 2017 - supplementary material S1and fig. S11).

Population size changes through time

Changes in N_e through time for fur seals were inferred from genomes using the Partial Sequentially Markovian Coalescent model of PSMC 0.6.5-r67 (Li and Durbin 2012). First, we generated the inputs using Samtools mpileup 1.9, applying a minimum mapping and base quality Phred-score > 30 and the option -C 50. After that, we used vcfutils vcf2fq perl script (Danecek et al. 2011) to apply a minimum individual read depth of 10 and a maximum of 100. We set the PSMC analysis using -N25, for the number of cycles of the algorithm, -t15, as the upper limit for the most recent common ancestor (TMRCA), -r5, for the initial θ/ρ, and -p 4 + 25*2 + 4 + 6 atomic intervals. The reconstructed population history was plotted using psmc_plot.pl script applying two substitution rates: a conservative mammalian substitution rate of 1×10^{-8} bp^{-1} gen^{-1}; and a substitution rate of 5.5×10^{-8} bp^{-1} gen^{-1} obtained with MSC model of StarBEAST2 (supplementary material S1) for 300 genomic fragments of 50 kb and 10 years per generation time (Nachman and Crowell 2000; Trilmmich 2015; Lopes et al. in prep).

Mitochondrial phylogeny

We captured the mtDNA CR from the raw reads of whole-genomes of the NZFS, GFS, PCFS and SAFS by mapping these reads with PALEOMIX against the complete mitochondrial genome of *A. townsendi* (GenBank NC008420). Sequenced mtDNA CR and captured sequences were automatically aligned with MUSCLE and edited on MEGA X (Kumar et al. 2018) with CR sequences retrieved from GenBank (Table S2): NZFS (Salis et al. 2016), GFS (Lopes et al. 2015) and SAFS (Crespo et al. 2015; Oliveira et al., unpublished data; Rodrigues et al. 2018). In order to identify clusters of PCFS mitochondrial lineages, a Neighbor-joining tree was recovered over 256 bp of CR with Mega X and 100 bootstrap replicates.

Acknowledgements

We are very thankful for the support of the following people and institutions: Cordenação de Aperfeiçoamento de Pessoal de Nível Superior-CAPES and Pontifícia Universidade Católica do Rio Grande do Sul-PUCRS by the institutional scholarship of FL; Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul-FAPERGS PRONEX 12/2014 and US Navy NICOP for the funding; All members of the Laboratório de Biologia Genômica e Molecular-PUCRS by the support, general suggestions and discussions, mainly the undergraduate student Jessica Pereira and the technicians Vanessa Endres and Giovanna Oliveira that helped in the libraries preparation; Jochen B. W. Wolf Group and Ludwig-Maximilians-Universität München which kindly hosted FL in his sandwich PhD; Nathaly Miranda who reviewed carefully many versions of this manuscript; all people and institutions that directly or indirectly contributed with the study.

Author contributions

FL, SLB and LO designed the study with contributions of JBWW and CP; FL and YB performed the analysis with contributions of SLB and AK; FL and SLB wrote the manuscript with

the support of the other authors; FL, AK, LO, SCA, PM, DPR, JC, EC, RBJ, MS, VFT, CL, BR, JBWW and CP prepared the samples, provide and processed data.

Competing interests

The authors declare no competing interests.

References

Abbott RJ, Rieseberg LH. 2012. Hybrid speciation. Encyclopaedia of Life Sciences. In: eLS. John Wiley & Sons, Ltd: Chichester.

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R. et al. 2013. Hybridization and Speciation. J Evol Biol. 26:229-246.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655-1664.

Amos W. 2007. Mix and match - hybridization reveals hidden complexity in seal breeding behaviour. Mol Ecol. 16:3066-3068.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Árnason U, Gullberg A, Janke A, Kullberg M, Lehman N, Petrov EA, Väinölä R. 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. Mol Phylogenet Evol. 41:345-354.

Árnason U, Lammers F, Kumar V, Nilsson MA, Janke A. 2018. Whole-genome sequencing of the blue whale and other rorqual finds signatures for introgressive gene flow. Sci Adv. 4:eaap9873.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol Ecol. 22:3179-3190.

Aurioles-Gamboa D, Schramm Y, Mesnick S. 2004. Galapagos fur seal, *Arctocephalus galapagoensis*, in Mexico. Latin J Aquat Mammal. 3:77-80.

Berta A, Churchill M. 2012. Pinniped taxonomy: Review of currently recognized species and subspecies, and evidence used for their description. Mamm Rev. 42:207-234.

Berta A, Churchill M, Boessenecker RW. 2018. The origin and evolutionary biology of pinnipeds: seals, sea lions and walruses. Annu Rev Earth Planet Sci. 46:203-228.

Bullini, L. 1994. Origin and evolution of animal hybrid species. Trends Ecol Evol. 9:422-426.

Cahill JA, Heintzman PD, Harris K, Teasdale MD, Kapp J, Soares AERS, Stirling I, Bradley D, Edwards CJ, Graim K et al. 2018. Genomic Evidence of Widespread Admixture from Polar Bears into Brown Bears during the Last Ice Age. Mol Biol Evol. 35:1120-1129.

Capella JJ, Flores-Gonzalez L, Falk-Fernandez P, Palacios DM. 2002. Regular appearance of otariid pinnipeds along the Colombian Pacific coast. Aquat Mammal. 28:67-72.

Cárdenas-Alayza S, Oliveira L, Crespo E. 2016. *Arctocephalus australis*. The IUCN Red List of Threatened Species: e.T2055A45223529.

Carey PW. 1992. Agonistic Behaviour in Female New Zealand Fur Seals, *Arctocephalus forsteri*. Ethology 92:70-80.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. G3 (Bethesda) 1:171-182.

Chilvers BL, Goldsworthy SD. 2015. *Arctocephalus forsteri*. The IUCN Red List of Threatened Species 2015:e.T41664A45230026.

Colella JP, Lan T, Schuster SC, Talbot SL, Cook JA, Lindqvist C. 2018. Whole-genome analysis of Mustela erminea finds that pulsed hybridization impacts evolution at high latitudes. Nature Comm Biol. 1:51.

Committee on Taxonomy of Marine Mammals. 2018. List of marine mammal species and subspecies. Available from https://www.marinemammalscience.org/species-information/list-marine-mammal-species-subspecies.

Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin JM, Estoup A. 2014. DIYABC 2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 15:1187-1189.

Crespo EA, Schiavini A, Garcia N, Franco-Trecu V, Goodal RNP, Rodriguez D, Morgante JS, Oliveira L. 2015. Status, population trend and genetic structure of South American fur seals, *Arctocephalus australis*, in southwestern Atlantic waters. Mar Mamm Sci. 31:866-890.

DaCosta JM, Sorenson MD. 2014. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS One. 4:e106713.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. 2011. The variant call format and VCFtools. Bioinformatics 1:2156-2158.

de Medeiros BAS, Farrell BD. 2018 Whole-genome amplification in double-digest RADseq results in adequate libraries but fewer sequenced loci. PeerJ. 6:e5089.

Dowling TE, Secor CL. 1997. The role of hybridization and introgression in the diversification of animals. Ann Rev Ecol Syst. 28: 93-619.

Dussex N, Robertson BC, Salis AT, Kalinin A, Best H, Gemmel NJ. 2016. Low Spatial Genetic Differentiation Associated with Rapid Recolonization in the New Zealand Fur Seal *Arctocephalus forsteri*. J Hered. 107:581-592.

Félix F, Jiménez P, Falconí J, Echeverry O. 2007. New records and first births of the Galapagos fur seal, Arctocephalus galapagoensis (Heller, 1904), from the mainland coast of Ecuador. Rev Biol Mar Oceanogr. 42:77-82.

Figueiró HV, Li G, Trindade F, Assis J, Pais F, Fernandes G, Santos SH, Hughes GM, Komissarov A, Antunes A et al. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci. Adv. 7:e1700299.

Foote AD, Liu Y, Thomas GW, Vinař T, Aföldi J, Deng J, Dugan S, van Elk CE, Hynter ME, Joshi V et al. 2015. Convergent Evolution of the genomes of marine mammals. Nat Genet. 47:272-275.

Francis RM. 2017. Pophelper: an R package and web app to analyse and visualize population structure. Mol Ecol Resour. 17:27-32.

Freedman AH, Gronau I, Schweitzer RM, Ortega-Del Vecchyo DO, Han E, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B et al. 2014. Genome sequencing highlights the dynamic early history of dogs. PLoS Genet. e:1004016.

Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM, Estoup A. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol Ecol. 22:3165–3178.

Gronau I, Hubisz M, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. Nat Genet. 43:1031-1034.

Harris AM, DeGiorgio M. 2012. Admixture and ancestry inference from ancient and modern samples through measures of population genetic drift. Hum Biol. 114.

Harisson RG. 1990. Hybrid zones: windows on the evolutionary process. Oxford surveys in evolutionary biology. 7:69-128.

Higdon JW, Bininda-Emonds ORP, Beck RMD, Ferguson SH. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evol Biol. 7:216.

Humble E, Martinez-Barrio A, Forcada J, Trathan PN, Thorne MA, Hoffman M, Wolf JBW, Hoffman JI. 2016. A draft fur seal genome provides insights into factors affecting SNP validation and how to mitigate them. Mol Ecol Resour. 16:909-921.

Humble E, Dasmahapatra KK, Martinez-Barrio A, Gregório I, Forcada J, Polikeit AC, Goldsworthy SD, Goebel ME, Kalinowski J, Wolf JBW et al. 2018. RAD Sequencing and a Hybrid Antarctic Fur Seal Genome Assembly Reveal Rapidly Decaying Linkage Disequilibrium, Global Population Structure and Evidence for Inbreeding. G3 (Bethesda). 31:2709-2722.

Kess T, Gross J, Harper F, Boulding EG. 2016. Low-cost ddRAD-seq method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. J Mollus Stud. 82:104-109.

King J. 1954. The otariid seals of the Pacific coast of America. Bull Brit Mus Zool. 2:311-337.

King J. 1969. The identity of the fur seals of Australia. Aust J Zool. 17:841-853.

King, J. E. 1983. Seals of the world. 2nd edition. Cornell University Press, Ithaca, NY.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A. 86:6196-6200.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356.

Kumar V, Lammers F, Bidon T, Pfenninger M, Kolter L, Nilsson MA, Janke A. 2017. The evolutionary history of bears is characterized by gene flow across species. Sci Rep. 7:46487.

Kumar S, Stecher G, Li M, Knyaz C, Kumar K. 2018. Mega X: Molecular Evolutionary Genetic Analysis across computing platforms. Mol Biol Evol. 35:1547-1549.

Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018. Rapid hybrid speciation in Darwin's finches. Science 359:224-228.

Lancaster M, Goldsworthy SD, Sunnucks P. 2007. Multiple mating strategies explain unexpected genetic mixing of New Zealand fur seals with two congenerics in a recently recolonizes population. Mol Ecol. 16:5267-5276.

Leaché AD, Harris RB, Rannala B, Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst Biol. 63:17-30.

Lento G, Haddon M, Chambers GK, Baker CS. 1997. Genetic variation of Southern Hemisphere fur seals (*Arctocephalus* spp.): investigation of population structure and species identity. J Heredity 88:202-206.

Leppälä K, Nielsen S, Mailund T. 2017. admixturegraph: an R package for admixture graph manipulation and fitting. Bioinformatics 33:1738-1740.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754-60.

Li H, Durbin R. 2012. Inferrence of Human population history from whole-genome sequence of a single individual. Nature 475:493-496.

Lopes F. Hoffman JI, Valiati VH, Bonatto SL, Wolf JBW, Trillmich F, Oliveira L. 2015. Fine-scale matrilineal population structure of the Galapagos fur seal and its implications for conservation management. Conserv Genet. 4:1099-1113.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol Evol. 20:229-237.

Mallet J. 2007. Hybrid speciation. Nature 446:279.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, dePristo MA. 2010. The Genome Analysis Toolkit: a MapReference framework for analyzing next-generation DNA sequencing data. Genome Res. 20:297-303.

Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. 2018. Genomics of parallel ecological speciation in Lake Victoria cichlids. Mol Biol Evol. 35:1489-1506.

Meisner J, Albretchtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. Genetics 210:719-731.

Meyer A, Kocher TD, Basasiwaki P, Wilson AC. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347:550-553.

Milmann L, Machado R, Oliveira L, Ott PH. 2019. Far away from home: the presence of fur seal (*Arctocephalus* sp.) in the equatorial Atlantic Ocean. Polar Biol. 42:817-822.

Montero-Cordero A, Fernández DM, Hernández-Mora G. 2010. Mammalia, Carnivora, Otariidae, Arctocephalus galapagoensis Heller, 1904: first continental record for Costa Rica. Check List 6:630-632.

Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics 156:297-304.

Nyakatura K, Bininda-Emonds ORP. 2012. Updating the evolutionary history of Carnivora (Mammalia): A new species-level supertree complete with divergence time estimates. BMC Biol. 10:12. doi:10.1186/1741-7007-10-12.

Ogilvie HA, Bouckaert RR, Drummond AJ. 2017. StarBEAST2 Brings faster species tree inference and accurate estimates of substitution rates. Mol Biol Evol. 34:2101-2114.

Oliveira L, Arias-Schreiber M, Meyer D, Morgante JS. 2006. Effective population size in a bottlenecked fur seal population. Biol Conserv. 131:505-509.

Oliveira L, Brownell RL. 2014. Taxonomic status of two subspecies of South American fur seals: *Arctocephalus australis australis* vs. *A. a. gracilis*. Mar Mammal Sci. 30:1258-1263.

Oliveira L, Hingst-Zaher E, Morgante JS. 2005. Size and shape sexual dimorphism in the skull of the South American fur seal, *Arctocephalus australis* (Zimmermann, 1783) (Carnivora: Otariidae). Latin J Aquat Mammal. 4:27-40.

Oliveira L, Hoffman JI, Hingst-Zaher E, Majluf P, Muelbert MMC, Morgante JS. 2008. Morphological and genetic evidence for two evolutionarily significant units (ESUs) in the South American fur seal, *Arctocephalus australis*. Conserv Genet. 9:1451-1466.

Oliveira L, Meyer D, Hoffman J, Majluf P, Morgante JS. 2009. Evidence of a genetic bottleneck in an El Niño affected population of South American fur seals, *Arctocephalus australis*. J Mar Biol Assoc UK. 89:1717-1725.

Páez-Rosas D, Valdovinos LA, Elorriaga-Verplancken FR. 2017. Northernmost Record of the Galapagos Fur Seal (Arctocephalus galapagoensis): A Consequence of Anomalous Warm Conditions Around the Galapagos Archipelago. Aquat Mammal. 43:629-634.

Pavés H, Schlatter RP, Franco-Trecu V, Crespo E, Siefeld W, Araos V, Giesecke R, Batallés LM, Capozzo HL. 2016. Breeding season of the South American fur seal (*Arctocephalus australis*, Otariidae:Carnivora): New data for establishing independent evolutionary histories? Rev Biol Mar Oceano. 51:241-253.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 7:e37135.

Pickrell JK. Pritchard JK. 2012 Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8:e1002967.

Quintanna-Rizzo E, Garcia L, López RJ, Tobar-Hurtado S, López-Roulet A. 2017. First record of a Galapagos fur seal (*Arctocephalus galapagoensis*) in Guatemala; Mar Biodivers Record. 10:24.

Repenning CA, Peterson RS, Hubbs CL. 1971. Contributions to the systematics of the southern fur seals, with particular reference to the Juan Fernández and Guadalupe species. Antar Res S. 18:1-34.

Rodrigues P, Seguel M, Gutiérrez J, Pavés HJ, Verdugo C. 2018. Genetic connectivity of the South American fur seal (*Arctocephalus australis*) across Atlantic and Pacific oceans revealed by mitochondrial genes. Aquat Conserv Mar Fresh Ecosyst. 28:315-323.

Salis AT, Easton LJ, Robertson BC, Gemmel N, Smith IWG, Weisler MI, Waters JM Rawlence NJ. 2016. Mith or relict? Does ancient DNA detect the enigmatic Upland seal? Mol Phylogenet Evol. 97:101-106.

Schaurich M, Lopes F, Oliveira L. 2012. Revisão sobre o fenômeno da hibridização em cetáceos e pinípedes. Neotrop Biol Conserv. 7:199-209.

Schubert M, Ermini L, Sarkissian CD, Jónsson H, Ginolhac A, Schaefer R, Martin MD, Fernández R, Kircher M, McCue M et al. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat Protoc. 9:1056-1082.

Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88.

Schumer M, Rosenthal GG, Andolfatto P. 2014. How common is homoploid hybrid speciation? Evolution 68:1553-1560.

Schumer M, Rosenthal GG, Andolfatto P. 2018. What do we mean when we talk about hybrid speciation? Heredity 120:379.

Sclater P.L. 1897. On the distribution of marine mammals. Proc Zool Soc London p.349–359.

Teng H, Zhang Y, Shi C, Mao F, Cai W, Lu L, Zhao F, Sun Z, Zhang J. 2017. Population genomics reveals speciation and introgression between brown Norway rats and their sibling species. Mol Biol Evol. 34:2214-2228.

Tobler M, Carson EW. 2010. Environmental variation, hybridization, and phenotypic diversification in Cuatro Ciénegas pupfishes. J Evol Biol. 23:1475-1489.

Trillmich F. 2015. *Arctocephalus galapagoensis*. The IUCN Red List of Threatened Species 2015:e.T2057A45223722.

Túnez JI, Centrón D, Capozzo HL, Cassini MH. 2007. Geographic distribution and diversity of mitochondrial DNA haplotypes in South American sea lions (*Otaria flavescens*) and fur seals (*Arctocephalus australis*). Mammal Biol. 4:193-203.

Túnez, JI, Cappozzo HL, Paves H, Albareda DA, Cassini MH. 2013. The role of Pleistocene glaciations in shaping the genetic structure of South American fur seals (*Arctocephalus australis*). New Zeal J Mar Fresh. 47:139-152.

Wen D, Nakhleh L. 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. Syst Biol. 67:439-457.

Wynen LP, Goldsworthy SD, Insley SJ, Adams M, Bickham JW, Francis J, Gallo JP, Hoelzel AR, Majluf P, White RWG et al. 2001. Phylogenetic relationships within eared seals (Otariidae: Carnivora): implications for the historical biogeography of the family. Mol Phylogenet Evol. 21:270-284.

Yonezawa T, Kohno N, Hasegawa M. 2009. The monophyletic origin of sea lion and fur seals (Carnivora: Otariidae) in the southern hemisphere. Gene 441:89-99.

Zalapa JE, Brunet J, Guries RP. 2010. The extent of hybridization and its impact on the genetic diversity and population structure of an invasive tree, *Ulmus pumila* (Ulmaceae). Evol Appl. 3:157-168.

There and there again: Genomic data indicate the Peruvian-Chilean fur seal is an ancient hybrid between the South American and the Galapagos species and that the latter is expanding and reproducing in the continent again

___

Supplementary material

Fernando Lopes[1], Larissa Oliveira[2,3], Yago Beux[1], Amanda Kessler[1], Susana Cárdenas-Alayza[4], Patricia Majluf[4], Diego Páez-Rosas[5], Jaime Chaves[5], Enrique Crespo[6], Robert Brownell Jr.[7], Maritza Sepúlveda[8], Valentina Franco-Trecu[9], Carolina Loch[10], Bruce Robertson[11], Claire Peart[12], Jochen B. W. Wolf[12], Sandro L. Bonatto[1]

[1] Escola de Ciências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brasil
[2] Programa de Pós-graduação em Biologia, Universidade do Vale do Rio dos Sinos, São Leopoldo, Brasil
[3] GEMARS, Grupo de Estudos de Mamíferos Aquáticos do Rio Grande do Sul
[4] Centro para la Sostenibilidad Ambiental, Universidad Peruana Cayetano Heredia, Lima, Peru
[5] Colegio de Ciencias Biológicas y Ambientales, COCIBA, Universidad San Francisco de Quito, Quito, Ecuador
[6] CONICET, Centro Nacional Patagónico - CENPAT, Puerto Madryn, Argentina
[7] NOAA, National Oceanic and Atmospheric Administration, La Jolla, United States of America
[8] Departamento de Biología y Ciencias Ambientales, Universidad de Valparaiso, Valparaiso, Chile
[9] Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay
[10] Department of Oral Sciences, University of Otago, Dunedin, New Zealand
[11] Department of Zoology, University of Otago, Dunedin, New Zealand
[12] Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, Münich, Germany

August 29, 2019

Contents

S1. Phylogenomic Inference in the Absence of Gene Flow

Per-scaffold alignments obtained from the whole genomes dataset were partitioned into sets of nonoverlapping genomic fragments of 50 kilobases (kb), with a step size of 100 kb between each alignment to avoid the effect of Linkage Disequilibrium (see Humble et al. 2018). Sites with missing data and masked regions were removed with trimAl v1.4 (Capella-Gutierrez et al. 2009). Scaffolds smaller than 50 kb were excluded and only sites with 100% of the data were kept in the downstream pipeline. Alignments smaller than half of the original alignment size and/or non-informative alignments were discarded.

For each genomic fragment, phylogenetic Maximum Likelihood (ML) trees were computed with RAxML-HPC-PTHREADS 8.2 (Stamatakis 2014) using GTR+G as best fit substitution model estimated by JModelTest2 (Darriba et al. 2012) and AFS as outgroup. The outputs of RAxML of each dataset were used to produce the maximum quartet support species tree (exact search) under the Multispecies Coalescent Codel (MSC) of ASTRAL-III (Zhang et al. 2018).

To estimate the root age, we also recovered the species tree with StarBEAST2 template implemented in BEAST 2.5.2 package (BEAUti, BEAST, TreeAnotator - Rambaut and Drummond 2010; Drummond and Bouckaert et al. 2014; Ogilvie et al. 2017). This analysis, as the ASTRAL-III approach, assumes that the source of genealogical discordances is due to Incomplete Lineage Sorting (ILS), not considering events of migration in the model and then over or underestimating the times of speciation (Wen and Nakhleh 2018). In this analyzes, we selected the 300 most variable genomic fragments estimated with AMAS tool (Borowiec 2016). The priors used were: a strict molecular clock with a relative mutation rate = 1, linked clock models, constant population sizes, HKY substitution model with empirical base frequencies, an estimated gamma site model and Yule tree model. We run a Bayesian Markov Chain Monte Carlo (MCMC) of 500,000,000 steps sampled each 20,000 with a burn-in of 10%. The divergence times were calibrated based on the split time between AFS and the remaining species at 2.4 Mya. This node age was based on the calibrated whole-genome phylogenies for almost all Otariidae species (Lopes et al. in prep.).

The tree recovered under Multispecies Species Coalescent Model of ASTRAL-III does not consider gene flow in the phylogenetic reconstruction (fig. S11A). This tree showed AFS and NZFS as the basal species as expected. Surprisingly, Peruvian-Chilean fur seal (PCFS) grouped with Galapagos fur seal (GFS), both being the sister group of the South American fur seal (SAFS). Here, we also highlight that the branch length of PCFS is longer than that of the other species. The StarBEAST MSC tree presented the same topology. The divergence time in the split of PCFS and GFS should be interpreted with caution since the method do not consider gene flow.

S2. SNP calling and filtering parameters

ddRAD-seq parameters:

SNPs were called with ANGSD 0.921 (Korneliussen et al. 2014) using the parameters:

doPlink 2, uniqueOnly 1, remove_bads 1, only_proper_pairs 1, C 50, baq 1, setMinDepth 235, setMaxDepth 4700, setMinDepthInd 5, setMaxDepthInd 100, doCounts 1, GL 2, doMajorMinor 1, doGlf 2, SNP_pval 1e-3, minInd 38, doGeno 4, doPost 1, doMaf, postCutoff 0.99, geno_minDepth 4. After the SNP calling, the variant panel was converted to VCF format with Plink 1.9 (Chang et al. 2015).

Whole-genome parameters:

From the mapped reads, SNPs were called using ANGSD 0.921 (Korneliussen et al. 2014) with the parameters: *doFasta 2, doCounts1* and *explode 1*. Single-nucleotide polymorphisms (SNPs) were called with settings: *uniqueOnly 1, remove_bads 1, only_proper_pairs 1, C 50, baq 1, setMinDepth* 120, *setMaxDepth 1200, setMinDepthInd 5, setMaxDepthInd 100, doCounts 1, GL 2, doMajorMinor 1, SNP_pval 1e-3, doGeno 32, doPost 1, doPlink2*. After the SNP calling, the variant panel was converted to VCF format with Plink 1.9 (Chang et al. 2015).

Parameters meaning:

Quality filtering

- only_proper_pairs =1: include only proper pairs (pairs of read with both mates mapped correctly). 1: include only proper (default), 0: use all reads. Only relevant for paired end data;
- -C 50: adjust mapQ for excessive mismatches (as SAMtools). The coefficient for downgrading mapping quality for reads containing excessive mismatches. Given a read with a phred-scaled probability q of being generated from the mapped position, the new mapping quality is about sqrt((INT-q)/INT)*INT. A zero value disables this functionality; if enabled, the recommended value for BWA is 50;
- Baq=1: BAQ is the Phred-scaled probability of a read base being misaligned. It greatly helps to reduce false SNPs caused by misalignments. BAQ is calculated using the probabilistic realignment method described in the paper "Improving SNP discovery by base alignment quality", Heng Li, Bioinformatics, Volume 27, Issue 8 ;
- uniqueOnly=1: remove reads that have multiple best hits. 0 no (default), 1 remove;
- remove_bads=1: same as the samtools flags -x which removes read with a flag above 255 (not primary, failure and duplicate reads). 0 no 1 remove (default).

Haplotype calling

- doGeno=4: print the called genotype as AA, AC, AG;
  =32: write the posterior probabilities of the 3 genotypes as binary;

- doPost=1: estimate the posterior genotype probability based on the allele frequency as a prior;
- GL=2: Genotype likelihoods obtained by using GATK algorithm
- postCutoff=0.99: call only a genotype with a posterior above this threshold;
- minInd=28: only keep sites with at least minIndDepth (default is 1) from at least [int] individuals

- -doMaf=2: Here the major allele is assumed to be known (inferred or given) however the minor allele is not determined. Instead, we sum over the 3 possible minor alleles weighted by their probabilities;
- SNP_pval= 1e-3: only work with sites with a p-value less than 1e-3;
- -doGlf=2: beagle haplotype imputation and be performed directly on genotype likelihoods. To generate beagle input file use;
- doMajorMinor=1: from the input for either sequencing data like bam files or from genotype likelihood data like. The major and minor allele can be inferred directly from likelihoods;
- doCounts=1: use -doCounts 1 in order to count the bases at each site after filters.

S3. Individual clustering and population structure parameters used in ANGSD

Principal Component Analysis (PCA) for the whole genomes

uniqueOnly 1, remove_bads 1, only_proper_pairs 1, C 50, baq 1, setMinDepth 30, setMaxDepth 600, setMinDepthInd 5, setMaxDepthInd 100, doCounts 1, GL 2, doMajorMinor 1, doMaf 1, SNP_pval 1e-3, doGeno 32, doPost 1, nThreads 10, doPlink 2

Principal Component Analysis (PCA) for the ddRAD-seq

uniqueOnly 1, remove_bads 1, only_proper_pairs 1, C 50, baq 1, setMinDepth 5, setMaxDepth 235, setMinDepthInd 5, setMaxDepthInd 4700, doCounts 1, GL 2, doMajorMinor 1, doMaf 1, SNP_pval 1e-3, doGeno 32, doPost 1, -nThreads 10, doPlink 2

Admixture and FST for ddRAD-seq

doPlink 2, uniqueOnly 1, remove_bads 1, only_proper_pairs 1, C 50, baq 1, setMinDepth 235, setMaxDepth 4700, setMinDepthInd 5, setMaxDepthInd 100, doCounts 1, GL 2, doMajorMinor 1, doGlf 2, SNP_pval 1e-3, minInd 28, doGeno 4, doPost 1, doMaf 2, postCutoff 0.99, geno_minDepth 4

The option *-doSaf 1* was applied with the parameters above described to generate the Site Frequency Spectrum for FST population differentiation.

S4. G-PHoCS pipeline for selecting neutral loci

#use bedtools slop to remove CDS and adjacent regions
bedtools slop -b 10000 -i ODros.CDS.bed -g ODros.GENOME > ODros.annotated.cds.20kbSlop.bed
#get from the ".GENOME" file the sacaffold sizes. Add a column of zero between the name and scaffold sizes
cat ODros.GENOME | awk '{print $1,0,$2}' > ref_ODros.scaff.bed

#ensure that it is tab delimited; remove b from a and save only the regions equal or greater than 5 kb
bedtools subtract -a ref_ODros.scaff.bed -b ODros.annotated.cds.20kbSlop.bed | awk '($3-$2) >= 5000' > ref_ODros.nonCoding.5kbWin.bed

#making windows of 5 kb
bedtools makewindows -b ref_ODros.nonCoding.5kbWin.bed -w 5000 > ref_ODros.5kbwindows.bed

#randomly selecting fragments
shuf > ref_ODros.5kbwindows.bed | head -20000 | sort -k 1,1 -k2,2n > ref_ODros.shuf20k.bed

#converting zipped vcf into bcf zipped vcf
bcftools view -Oz GPhoCS.vcf.gz > GPhoCS.Oz.vcf.gz

#indexing the vcf file
tabix -p vcf GPhoCS.Oz.vcf.gz

#selecting bed regions
bcftools view -R ref_ODros.shuf20k.bed GPhoCS.Oz.vcf.gz > GPhoCS_targetregions20k.vcf.gz

#converting to bcf gzipped vcf again
bcftools view -Oz GPhoCS_targetregions20k.vcf.gz > GPhoCS_targetregions20kOz.vcf.gz

#indexing the new vcf
tabix -p vcf GPhoCS_targetregions20kOz.vcf.gz

#making consensus from vcf target regions
for samp in $(bcftools query -l GPhoCS_targetregions20kOz.vcf.gz); do bcftools consensus -s $samp -I -f ODros.fasta GPhoCS_targetregions20kOz.vcf.gz > $samp.consensus.fasta; done

#creating genomic windows
for f in *; do sample=${f%.consensus*}; echo $sample ; bedtools getfasta -fi $f -bed ref_ODros.shuf20k.bed > consensus/windows/$sample.windows.fasta; done

#split the genomic regions into individual alignments

Supplementary references

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. Peer J. 4:e1660.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyzes. Bioinformatics. 25:1972-1973.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer data sets. GigaScience 4.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: More models, new heuristics and parallel computing. Nat Methods. 9:722.

Drummond AJ, Bouckaert RR. 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Press. Cambridge 244 pp.

Humble E, Dasmahapatra KK, Martinez-Barrio A, Gregório I, Forcada J, Polikeit AC, Goldsworthy SD, Goebel ME, Kalinowski J, Wolf JBW et al. 2018. RAD Sequencing and a Hybrid Antarctic Fur Seal Genome Assembly Reveal Rapidly Decaying Linkage Disequilibrium, Global Population Structure and Evidence for Inbreeding. G3 (Bethesda) 31:2709-2722.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356.

Ogilvie HA, Bouckaert RR, Drummond AJ. 2017. StarBEAST2 Brings faster species tree inference and accurate estimates of substitution rates. Mol Biol Evol. 34:2101-2114.

Rambaut A, Drummond AJ. 2010. TreeAnnotator version 1.6.1 Available from http://beast.bio.ed.ac.uk.

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

Wen D, Nakhleh L. 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. Syst Biol. 67:439-457.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19 (suppl 6):153.

Table S1. PALEOMIX statistics for each species mapped against *O. rosmarus* genome (2.4 Gb genome).

	Library	Coverage of unique hits	%Mapped of the ref. genome	Fraction total hits vs. n° of reads retained	Average of hit length	Fraction of PCR duplicates
SAFS - Argentina	TruSeq Nano	21.34	0.85	0.91	149.81	0.06
SAFS - Falklands	TruSeq Nano	17.39	0.85	0.89	148.09	0.04
SAFS - Chile	TruSeq Nano	32.04	0.81	0.90	150.73	0.10
PCFS	TruSeq PCR Free	27.91	0.82	0.92	149.11	0.11
GFS	TruSeq Nano	15.42	0.79	0.90	150.14	0.12
NZFS	TruSeq PCR Free	26.05	0.86	0.91	151.40	0.05
AFS	GenBank	32.44	0.83	0.89	191.46	0.07

Table S2. Sampling information for fur seals samples used in RADseq analyses.

Species	Common name	ID code	Locality	Country	Geographical coordinates	Filtered out
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus01	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus02	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus03	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus04	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus05	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus06	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus07	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus08	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus09	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus12	Punta San Juan	Peru	15°22'S, 75°11'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus10	Isla Foca	Peru	05°12'S, 81°12'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus11	Isla Foca	Peru	05°12'S, 81°12'W	No
Arctocephalus australis	Peruvian-Chilean fur seal	Aaus53	Isla Foca	Peru	05°12'S, 81°12'W	No
Arctocephalus australis	South American fur seal	Aaus13	Isla Guafo	Chile	43°34'S, 74°45'W	Yes
Arctocephalus australis	South American fur seal	Aaus14	Isla Guafo	Chile	43°34'S, 74°45'W	No
Arctocephalus australis	South American fur seal	Aaus15	Isla Guafo	Chile	43°34'S, 74°45'W	Yes
Arctocephalus australis	South American fur seal	Aaus16	Isla Guafo	Chile	43°34'S, 74°45'W	No
Arctocephalus australis	South American fur seal	Aaus17	Isla Guafo	Chile	43°34'S, 74°45'W	No
Arctocephalus australis	South American fur seal	Aaus18	Isla Guafo	Chile	43°34'S, 74°45'W	No
Arctocephalus australis	South American fur seal	Aaus19	Isla Guafo	Chile	43°34'S, 74°45'W	No
Arctocephalus australis	South American fur seal	Aaus20	Isla Guafo	Chile	43°34'S, 74°45'W	No
Arctocephalus australis	South American fur seal	Aaus21	Isla Guafo	Chile	43°34'S, 74°45'W	Yes
Arctocephalus australis	South American fur seal	Aaus22	Isla Guafo	Chile	43°34'S, 74°45'W	No
Arctocephalus australis	South American fur seal	Aaus23	Province of Chubut	Argentina	45°06'S, 65°24'W	No
Arctocephalus australis	South American fur seal	Aaus24	Province of Chubut	Argentina	45°06'S, 65°24'W	No
Arctocephalus australis	South American fur seal	Aaus25	Province of Chubut	Argentina	45°06'S, 65°24'W	Yes
Arctocephalus australis	South American fur seal	Aaus26	Province of Chubut	Argentina	45°06'S, 65°24'W	No
Arctocephalus australis	South American fur seal	Aaus27	Province of Chubut	Argentina	45°06'S, 65°24'W	No
Arctocephalus australis	South American fur seal	Aaus28	Province of Chubut	Argentina	45°06'S, 65°24'W	Yes
Arctocephalus australis	South American fur seal	Aaus29	Province of Chubut	Argentina	45°06'S, 65°24'W	Yes
Arctocephalus australis	South American fur seal	Aaus30	Province of Chubut	Argentina	45°06'S, 65°24'W	No
Arctocephalus australis	South American fur seal	Aaus31	Province of Chubut	Argentina	45°06'S, 65°24'W	Yes
Arctocephalus australis	South American fur seal	Aaus32	Province of Chubut	Argentina	45°06'S, 65°24'W	Yes
Arctocephalus australis	South American fur seal	Aaus33	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus34	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus35	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus36	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus37	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus38	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus39	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus40	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	No
Arctocephalus australis	South American fur seal	Aaus41	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus42	Rio Grande do Sul	Brazil	29°20'S, 49°'42'W	Yes
Arctocephalus australis	South American fur seal	Aaus43	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	No
Arctocephalus australis	South American fur seal	Aaus44	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	Yes
Arctocephalus australis	South American fur seal	Aaus45	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	Yes
Arctocephalus australis	South American fur seal	Aaus46	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	Yes
Arctocephalus australis	South American fur seal	Aaus47	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	No
Arctocephalus australis	South American fur seal	Aaus48	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	Yes
Arctocephalus australis	South American fur seal	Aaus49	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	Yes
Arctocephalus australis	South American fur seal	Aaus50	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	Yes
Arctocephalus australis	South American fur seal	Aaus51	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	Yes
Arctocephalus australis	South American fur seal	Aaus52	Falkland Islands	British Overseas Territory	52°10'S, 60°55'W	Yes
Arctocephalus galapagoensis	Galapagos fur seal	Agal01	Cape Ibbetson	Ecuador	0°32'S, 90°43'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal02	Cape Ibbetson	Ecuador	0°32'S, 90°43'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal03	Punta Flores	Ecuador	0°05'S, 91°29'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal04	Punta Flores	Ecuador	0°05'S, 91°29'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal05	Punta Flores	Ecuador	0°05'S, 91°29'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal06	Cape Marshall	Ecuador	0°01'N, 91°12'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal07	Cape Marshall	Ecuador	0°01'N, 91°12'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal08	Cape Marshall	Ecuador	0°01'N, 91°12'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal09	Cape Marshall	Ecuador	0°01'N, 91°12'W	No
Arctocephalus galapagoensis	Galapagos fur seal	Agal10	Cape Marshall	Ecuador	0°01'N, 91°12'W	No
Arctocephalus forsteri	New Zealand fur seal	Afor01	Otago Harbour	New Zealand	45°49'S, 170°38E	No
Arctocephalus forsteri	New Zealand fur seal	Afor02	Otago Harbour	New Zealand	45°49'S, 170°38E	No
Arctocephalus forsteri	New Zealand fur seal	Afor03	Otago Harbour	New Zealand	45°49'S, 170°38E	No
Arctocephalus forsteri	New Zealand fur seal	Afor04	Otago Harbour	New Zealand	45°49'S, 170°38E	No
Arctocephalus forsteri	New Zealand fur seal	Afor05	Otago Harbour	New Zealand	45°49'S, 170°38E	No
Arctocephalus forsteri	New Zealand fur seal	Afor06	Otago Harbour	New Zealand	45°49'S, 170°38E	No
Arctocephalus forsteri	New Zealand fur seal	Afor07	Otago Harbour	New Zealand	45°49'S, 170°38E	No
Arctocephalus forsteri	New Zealand fur seal	Afor07	Otago Harbour	New Zealand	45°49'S, 170°38E	No
Arctocephalus forsteri	New Zealand fur seal	Afor09	Otago Harbour	New Zealand	45°49'S, 170°38E	No

Table S3. F_{ST} pairwise comparisons. SAFS - South American fur seals (localities: Brazil - BR (n=1), Argentina - AR (n=5), Falkland Islands - FK (n=2), Chile (n=3); PCFS - Peruvian-Chilean fur seals (n=10); GFS - Galapagos fur seals (n=10); and NZFS - New Zealand fur seals (n=9).

	SAFS BR	**SAFS FK**	**SAFS CH**	**PCFS**	**GFS**	**NZFS**
SAFS Argentina	0.03	0.08	0.03	0.14	0.31	0.25
SAFS Brazil	-	0.09	0.05	0.14	0.32	0.24
SAFS Falklands	-	-	0.04	0.18	0.30	0.26
SAFS Chile	-	-	-	0.13	0.25	0.20
PCFS	-	-	-	-	0.22	0.29
GFS	-	-	-	-	-	0.40

Table S4. Whole genomes f_3-statistics for the Peruvian-Chilean fur seal (PCFS) population and its specific population combination. Z-score < 0 is evidence for genomic introgression. GFS (Galapagos fur seal), NZFS = (New Zealand fur seal) and the different localities of the South American fur seal (SAFS AR - Argentina, FK - Falkland Islands, CH – Chile). Bold lines show the significant signals of admixture.

Admixed Population	**Population Combination**	***f*-statistics**	**Standard Error**	**Z-score**
PCFS	AR ; FK	0.013	0.0004	35.08
PCFS	AR ; CH	0.013	0.0004	34.52
PCFS	**AR ; GFS**	**-0.015**	**0.0003**	**-55.27**
PCFS	AR ; NZFS	0.004	0.0003	13.43
PCFS	FK ; SAFS CH	0.013	0.0004	35.76
PCFS	**FK ; GFS**	**-0.015**	**0.0003**	**-54.09**
PCFS	FK ; NZFS	0.005	0.0003	17.17
PCFS	**CH ; GFS**	**-0.015**	**0.0003**	**-55.42**
PCFS	CH ; NZFS	0.004	0.0003	13.77
PCFS	**GFS ; NZFS**	**-0.007**	**0.0003**	**-27.28**

Table S5. f_4-statistics for all four-taxa phylogeny indicating genomic introgression from SAFS (SAFS AR - Argentina, FK - Falkland Islands, CH – Chile) and GFS (Galapagos fur seal) in the PCFS (Peruvian-Chilean fur seal) genome. All values are significant.

W	X	Y	Z	*f*-statistics	Z-score
SAFS AR	PCFS	SAFS FK	GFS	0.028	77.41
SAFS AR	GFS	SAFS FK	PCFS	0.028	75.79
SAFS AR	PCFS	SAFS FK	NZFS	0.009	32.13
SAFS AR	NZFS	SAFS FK	PCFS	0.008	28.23
SAFS AR	NZFS	SAFS FK	GFS	0.015	51.21
SAFS AR	GFS	SAFS FK	NZFS	0.016	52.96
SAFS AR	PCFS	SAFS CH	GFS	0.028	76.30
SAFS AR	GFS	SAFS CH	PCFS	0.028	77.12
SAFS AR	PCFS	SAFS CH	NZFS	0.009	31.39
SAFS AR	NZFS	SAFS CH	PCFS	0.009	31.43
SAFS AR	NZFS	SAFS CH	GFS	0.016	53.81
SAFS AR	GFS	SAFS CH	NZFS	0.017	54.05
SAFS AR	PCFS	GFS	NZFS	-0.019	-63.83
SAFS AR	NZFS	PCFS	GFS	0.007	34.12
SAFS AR	GFS	PCFS	NZFS	-0.011	-34.27
SAFS FK	PCFS	SAFS CH	GFS	0.028	75.49
SAFS FK	GFS	SAFS CH	PCFS	0.029	77.58
SAFS FK	PCFS	SAFS CH	NZFS	0.008	29.59
SAFS FK	NZFS	SAFS CH	PCFS	0.009	33.19
SAFS FK	NZFS	SAFS CH	GFS	0.016	54.24
SAFS FK	GFS	SAFS CH	NZFS	0.016	53.03
SAFS FK	PCFS	GFS	NZFS	-0.020	-65.76
SAFS FK	NZFS	PCFS	GFS	0.007	32.10
SAFS FK	GFS	PCFS	NZFS	-0.013	-37.47
SAFS CH	PCFS	GFS	NZFS	-0.019	-65.16
SAFS CH	NZFS	PCFS	GFS	0.008	34.14
SAFS CH	GFS	PCFS	NZFS	-0.012	-34.61
PCFS	AFS	GFS	NZFS	0.047	160.90
PCFS	GFS	NZFS	AFS	0.001	13.16
PCFS	NZFS	GFS	AFS	0.048	167.44
SAFS CH	GFS	PCFS	AFS	-0.007	-25.97
SAFS CH	AFS	PCFS	GFS	0.007	38.56
SAFS CH	PCFS	GFS	AFS	-0.014	-58.71
SAFS FK	PCFS	GFS	AFS	-0.015	-58.26
SAFS FK	GFS	PCFS	AFS	-0.008	-28.12
SAFS AR	PCFS	GFS	AFS	-0.014	-57.30
SAFS AR	GFS	PCFS	AFS	-0.007	-26.38
SAFS AR	AFS	PCFS	GFS	0.007	39.57
SAFS FK	AFS	PCFS	GFS	0.007	39.13

Table S6. G-PHoCS parameters estimation and conversion into absolute values. SAFS - South American fur seals; PCFS - Peruvian-Chilean fur seals; GFS - Galapagos fur seals; and NZFS - New Zealand fur seals.

For G-PHoCS analysis we assumed a generation time of 10 years (Trillmich et al. 2015). Node ages were calibrated with root age (split time from NZFS) estimated with StarBEAST2 (see supplementary material S1).

	Median Θ G-PHoCS	Θ G-PHoCS HPD 95%	Median Θ	Θ HPD 95%	Median Ne	Ne HPD 95%	Generation time
			= Θ G-PHoCS/1E04	= Θ G-PHoCS HPD 95%/1E04	= Θ/4 x g x μ		(g)
SAFS	16.46	15.74-17.18	1.65E-03	1.57E-03-1.72E-03	194,716	186,199-203,233	10
PCFS	9.39	3.47-16.36	9.39E-04	3.47E-04-1.64E-03	111,080	41,049-193,533	
GFS	19.80	18.46-21.15	1.98E-03	1.85E-03-2.21E-03	224,227	218,375-250,197	
NZFS	30.00	28.68-31.38	3.00E-03	2.86E-03-3.19E-03	354,890	339,274-371,215	
theta root	10.95	10.79-11.09	1.10E-03	1.08E-03-1.10E-03	129,535	127,642-131,191	

	Median τ G-PHoCS	τ G-PHoCS HPD 95%	Median τ	τ HPD 95%	Median Time	Time HPD 95%	μ from τ Root
			= τ G-PHoCS/1E4		= Real τ / Time		u=Tau root/root age
Tau 1 (SAFS+PCFS)	0.787	0.744-0.827	7.87E-05	7.44E5-8.27E5	372,390	352,050-391,320	2,11E-10
Tau 2 (SAFS+PCFS+GFS)	0.827	0.800-0.852	8.27E-05	8.00E5-8.52E5	391,320	378,549-403,155	
Tau root (SAFS+PCFS+GFS+NZFS)	0.317	3.13-3.21	3.17E-04	3.31E4-3.21E4	1,500,000	1,481,073-1,518,927	

	Median migration (m) G-PHoCS	m G-PHoCS HPD 95%	Median m	m HPD 95 %	Median Migration Rate (M)	Migration Rate (M) HPD 95%
	(m)		= m/1E-03		M = m x τ	
GFS->PCFS	610.95	97.17-1428.37	610,950	91,170-1428,370	48,08	7.65-112.41
SAFS->PCFS	388.64	66.32-896.75	388,640	66,320-89,670	30,53	5.22-70.57
NZFS->SAFS	2.24	2.08-2.40	2,240	2,080-2,400	0,71	0.67-0.77

Table S7 Mitochondrial DNA Control Region information used in the current study and obtained from GenBank.

Species	Common name	Localities	Country	N° Samples	GeneBank	Source
Arctocephalus australis	Peruvian-Chileanfur seal	Punta San Juan	Peru	10	-	Current study
Arctocephalus australis	Peruvian-Chilean-Chilean/Galapagos fur seal	Isla Foca	Peru	3	-	Current study
Arctocephalus australis	South American fur seal	Isla Guafo	Chile	14	MG383525-38	Rodrigues et al. 2018
Arctocephalus australis	South American fur seal	Cabo Polonio	Uruguay	3	KM593065-67	Crespo et al. 2015
Arctocephalus australis	South American fur seal	RS Coast	Brazil	14	KM593068-81	Crespo et al. 2015
Arctocephalus australis	South American fur seal	Mar del Plata	Argentina	3	KM593082-83	Crespo et al. 2015
Arctocephalus australis	South American fur seal	Province of Chubut	Argentina	10	KM593084-93	Crespo et al. 2015
Arctocephalus australis	South American fur seal	Falkland Islands	British Overseas Territory	21	-	Oliveira et al., unpublished data
Arctocephalus galapagoensis	Galapagos fur seal	Galapagos Islands	Ecuador	30	KM030335-44-44;65-74;96-404	Lopes et al. 2015
Arctocephalus forsteri	New Zealand fur seal	South Island	New Zealand	97	KU510625-722	Salis et al. 2016
Arctocephalus gazella	Antarctic fur seal	Antarctic continent	-	2	AF384378-9	Wynen et al. 2001

Table S8. Peruvian-Chilean fur seals mitochondrial and nuclear DNA identification. Hybrid=Peruvian-Chilean + Galapagos fur seal; mtDNA clade=clade of the Neighbor-joining tree (table S7). Identification of the mtDNA captured from whole-genomes raw reads. PCFS=Peruvian-Chileanfur seal, GFS=Galapagos fur seal, SAFS=South American fur seal, NZFS=New Zealand fur seal.

Peruvian-Chilean sample	mtDNA	mtDNA clade	nDNA
Aaus01	Peruvian-Chilean	PCFS A	Hybrid
Aaus02	Peruvian-Chilean	PCFS B	Hybrid
Aaus03	Peruvian-Chilean	PCFS A	Hybrid
Aaus04	Peruvian-Chilean	PCFS A	Hybrid
Aaus05	Peruvian-Chilean	PCFS B	Hybrid
Aaus06	Peruvian-Chilean	PCFS A	Hybrid
Aaus07	Peruvian-Chilean	PCFS A	Hybrid
Aaus08	Peruvian-Chilean	PCFS B	Hybrid
Aaus09 (whole-genome)	Peruvian-Chilean	PCFS A	Hybrid
Aaus12	Peruvian-Chilean	PCFS A	Hybrid
Aaus10	Peruvian-Chilean	PCFS A	Hybrid
Aaus11	Galapagos	GFS	Galapagos
Aaus53	Galapagos	GFS	Galapagos
Aaus19 (whole-genome)	South American	SAFS	South American
Aaus31 (whole-genome)	South American	SAFS	South American
Aaus45 (whole-genome)	South American	SAFS	South American
Afor* (whole-genome)	New Zealand	NZFS	New Zealand
Agal* (whole-genome)	Galapagos	GFS	GFS

*Not sequenced in the ddRAD-seq libraries.

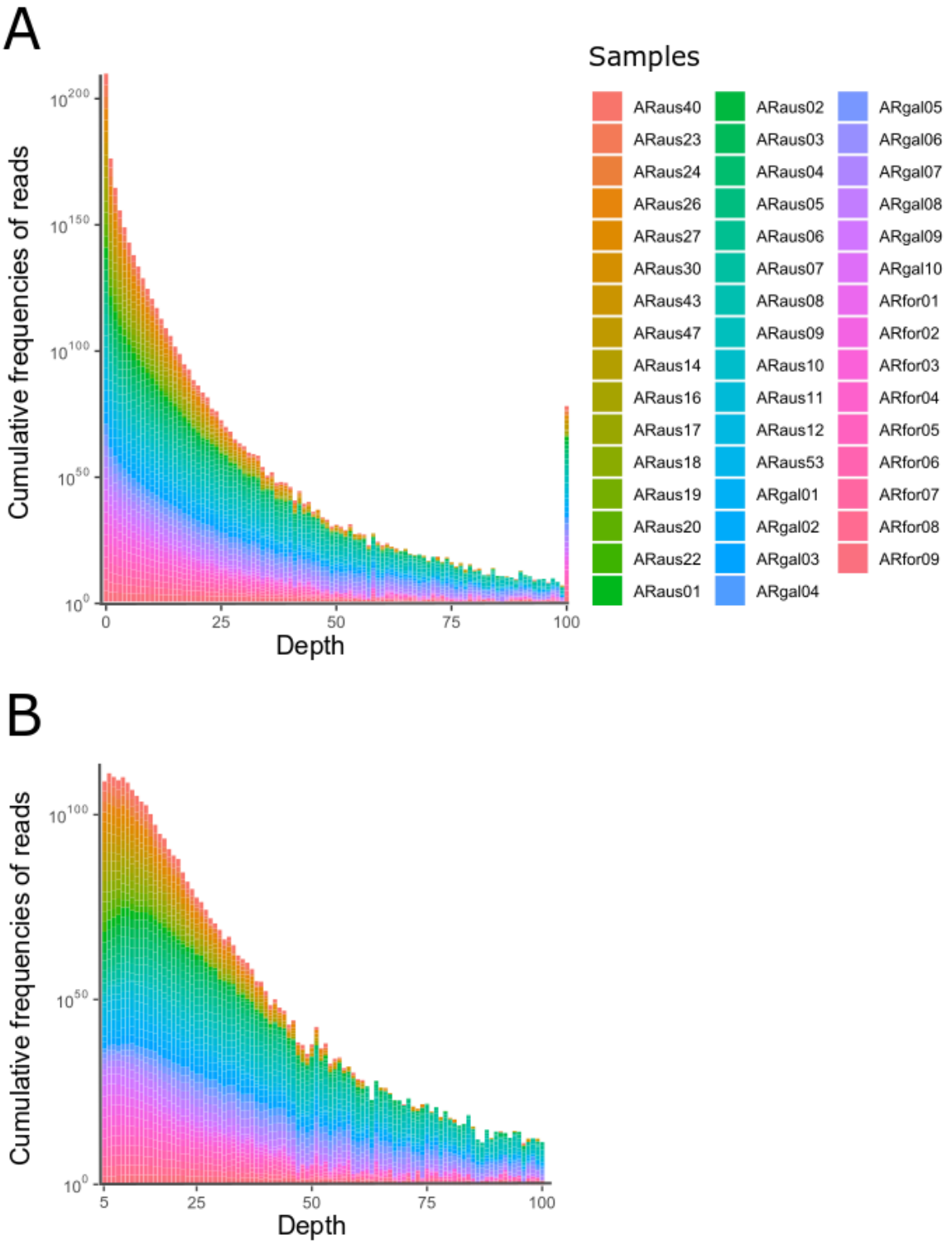


Figure S1. ddRAD-seq cumulative frequencies of reads depths per sample (A) before and (B) after SNP filtering.

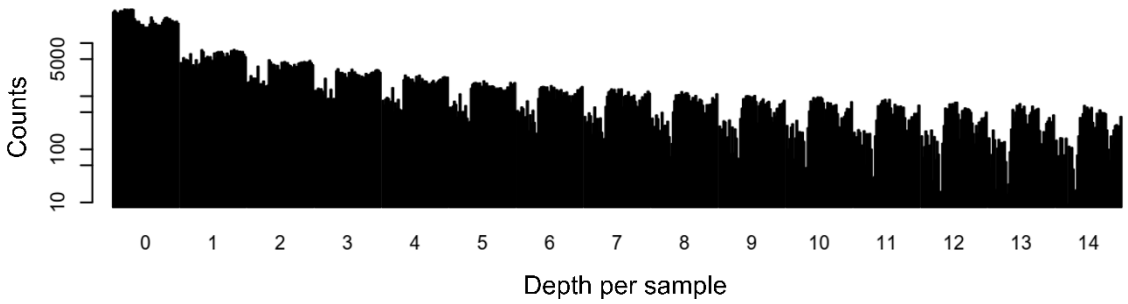


Figure S2. ddRAD-seq pre-filtering depth per sample. The x-axis shows the total depth until 14X; the y-axis shows the counts of reads to each sample (black bar) and its respective depth. The chart was pruned to show the reads that were removed after the minimum individual depth per sample (<5).

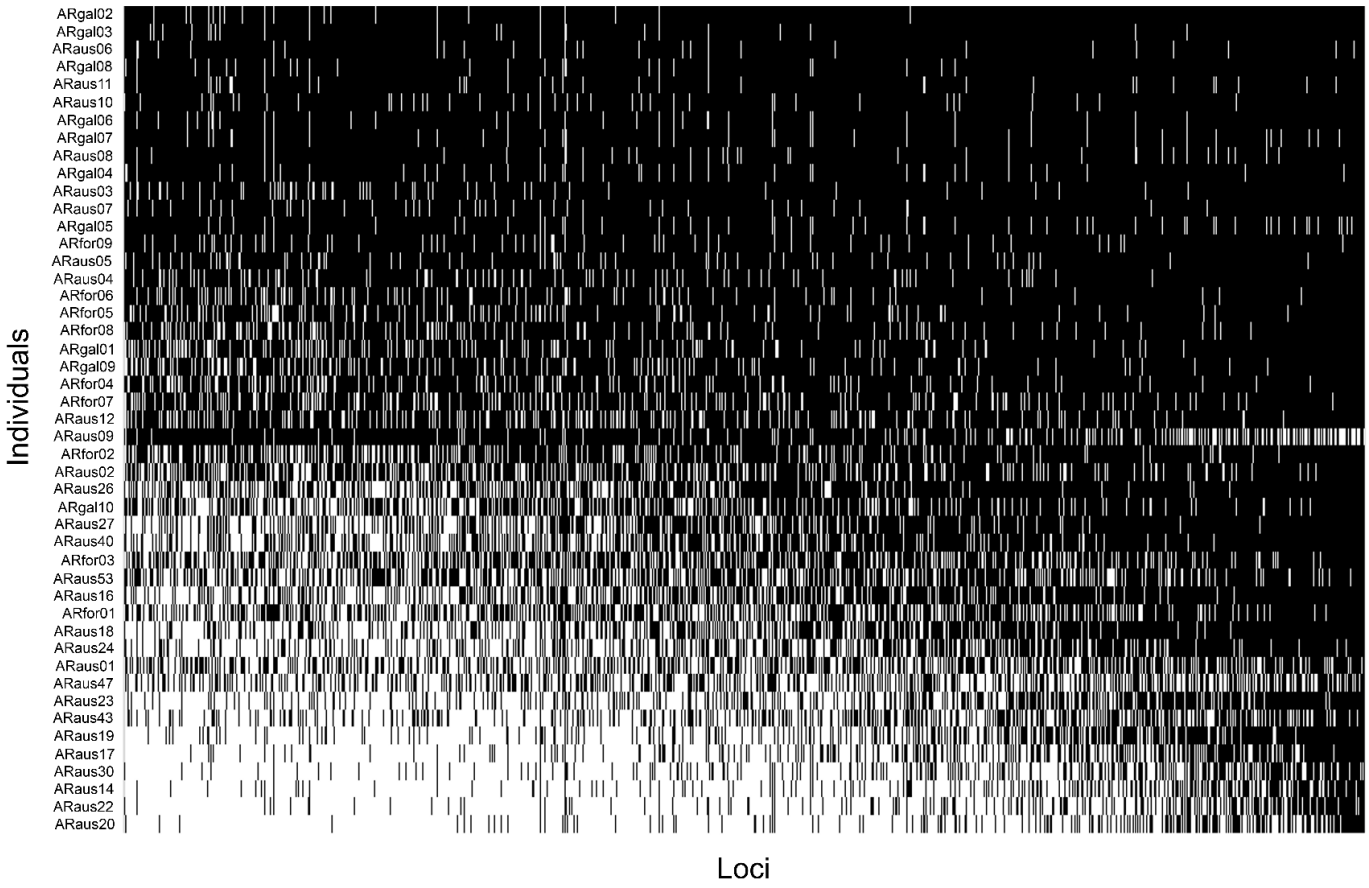


Figure S3. Matrix of the ddRAD-seq SNP dataset representing the 3,198 SNPs and showing the presence (black) and absence (white) of genomic information by loci and sample. The individuals were sorted by the amount of data (higher to lower).

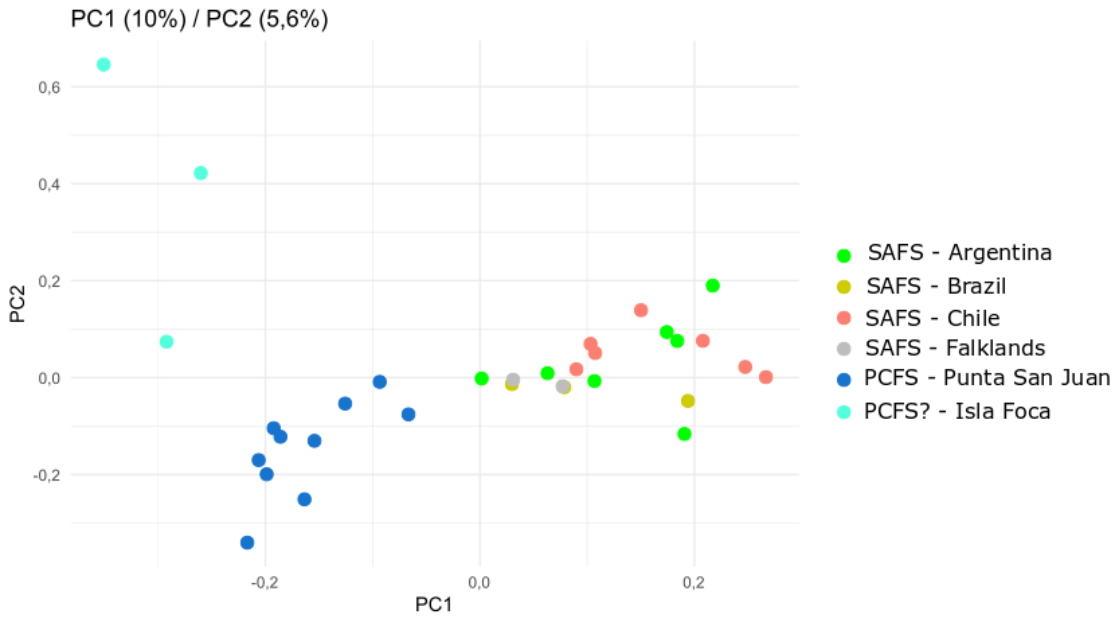


Figure S4. PCA of the South American fur seals (SAFS) and Peruvian-Chilean fur seals (PCFS) samples.

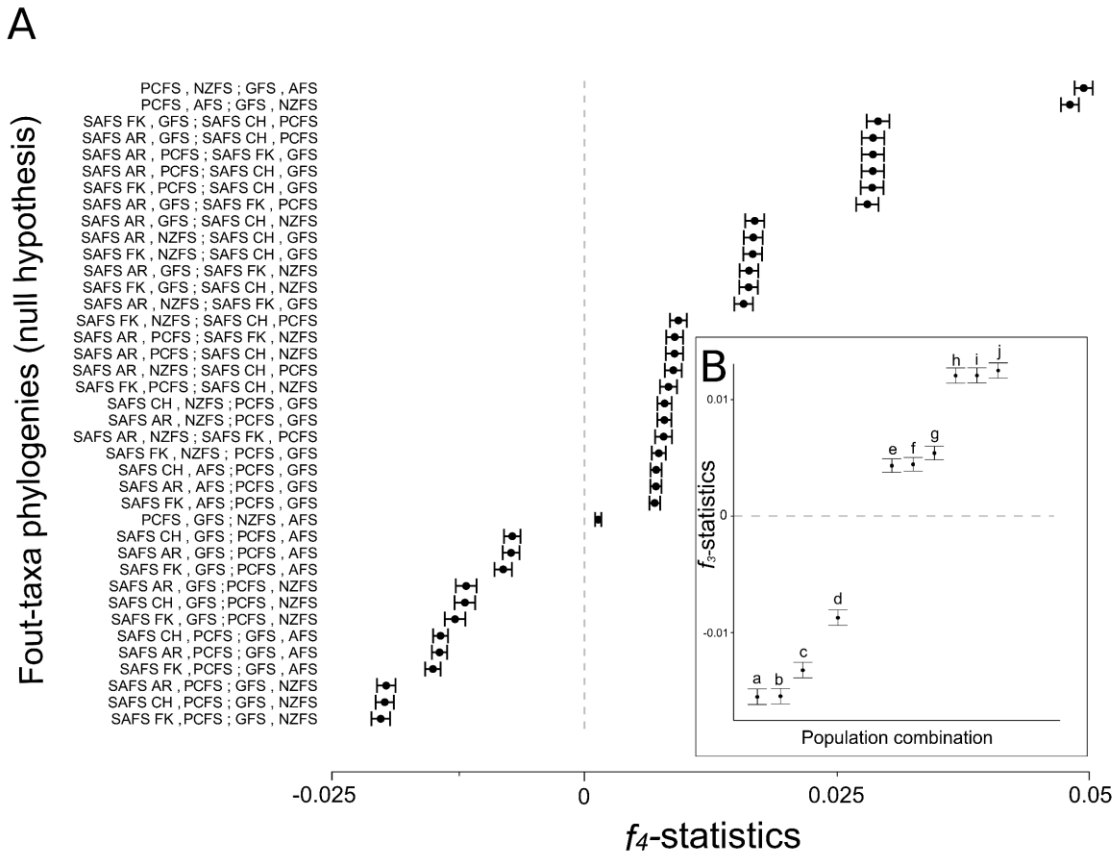


Figure S5. (A) Four-taxa phylogenies and f_4-statistics distribution showing the evidence for genomic introgression in the Peruvian-Chilean (PCFS) genome. (B) f_3-statistics showing PCFS as the result of admixture between Galapagos fur seal (GFS), New Zealand fur seal (NZFS) and the different localities of the South American fur seal (SAFS AR - Argentina, FK - Falkland Islands, CH - Chile). Admixture ($f_3 < 0$): (a) SAFS CH+GFS, (b) SAFS AR+GFS, (c) SAFS FK+GFS, (d) NZFS+GFS; no admixture ($f_3 > 0$): (e) SAFS CH+NZFS, (f) SAFS AR+NZFS, (g) SAFS FK+NZFS, (h) SAFS CH+AR, (i) SAFS AR+FK and (j) SAFS CH+FK. All admixture results are significant |Z-score| > 3.

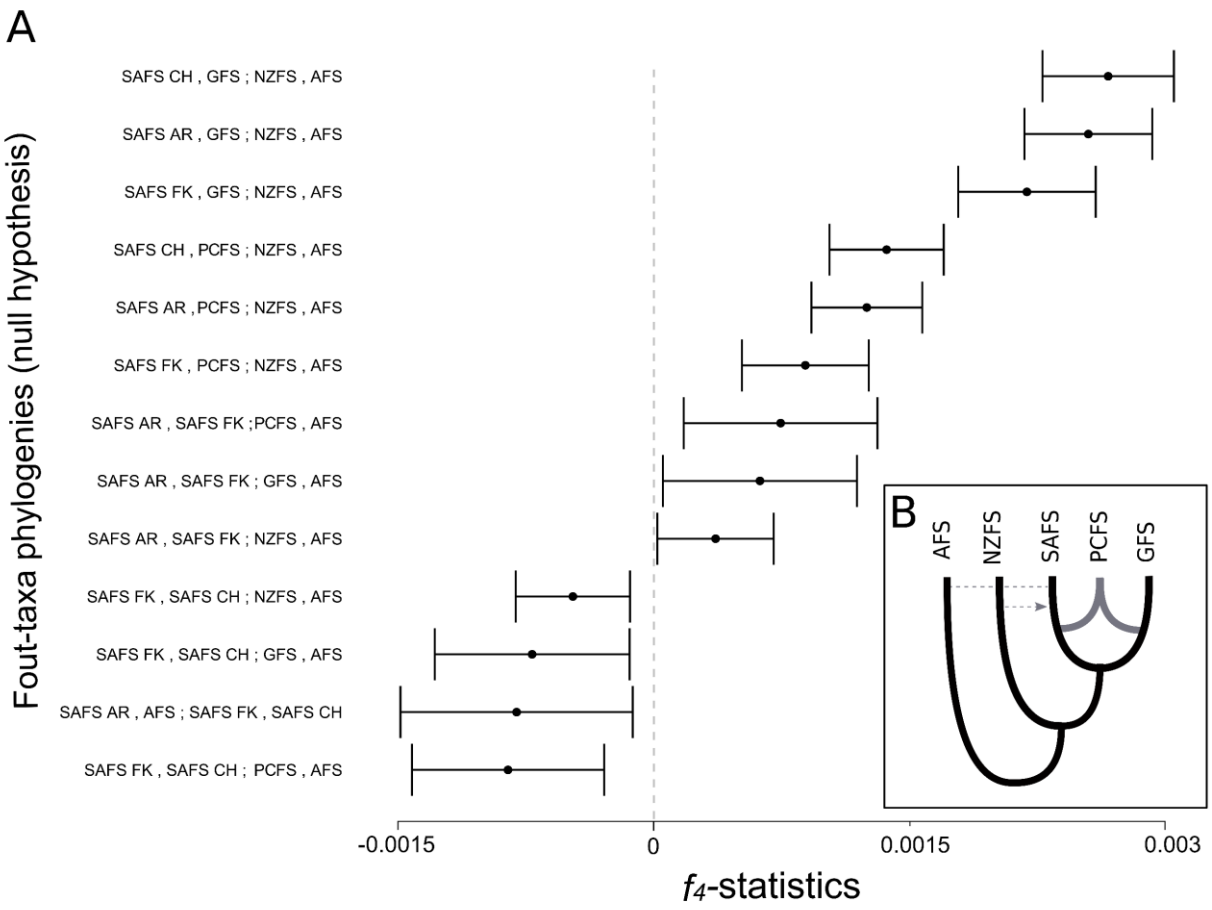


Figure S6. (A) Four-taxa phylogenies and f_4-statistics distribution showing the evidence for genomic introgression from New Zealand fur seal (NZFS) and Antarctic fur seal (AFS) in the South American fur seal genome (SAFS AR - Argentina, FK - Falkland Islands and CH – Chile). All admixture results are significant |Z-score| > 3. (B) Migration events NZFS to all SAFS, and from AFS to SAFS of Falkland Island.

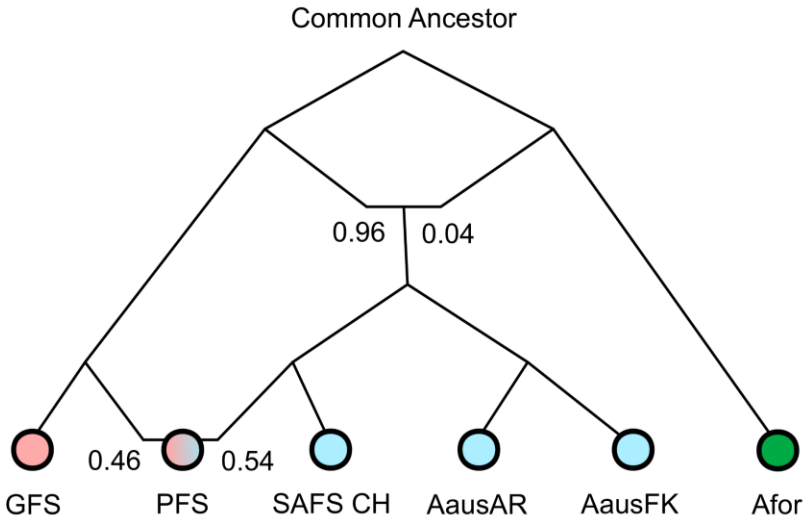
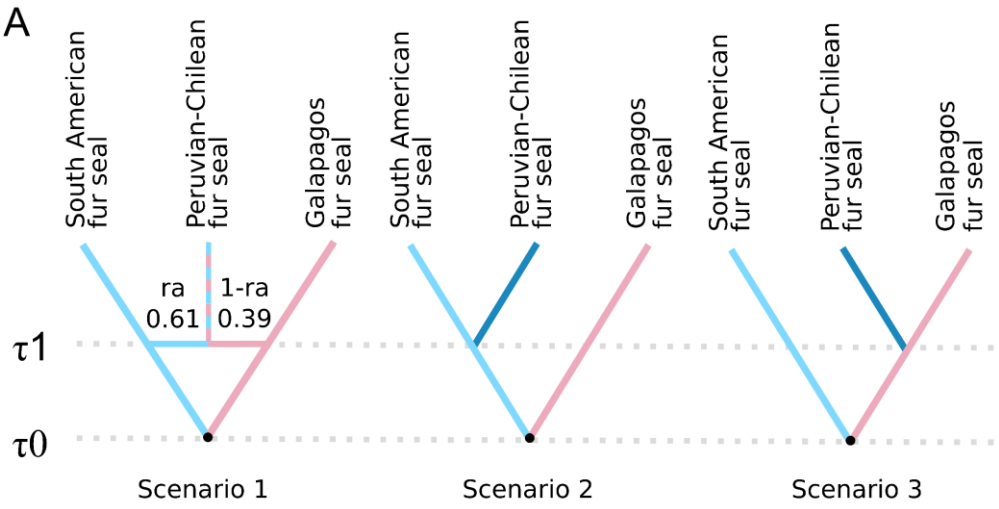
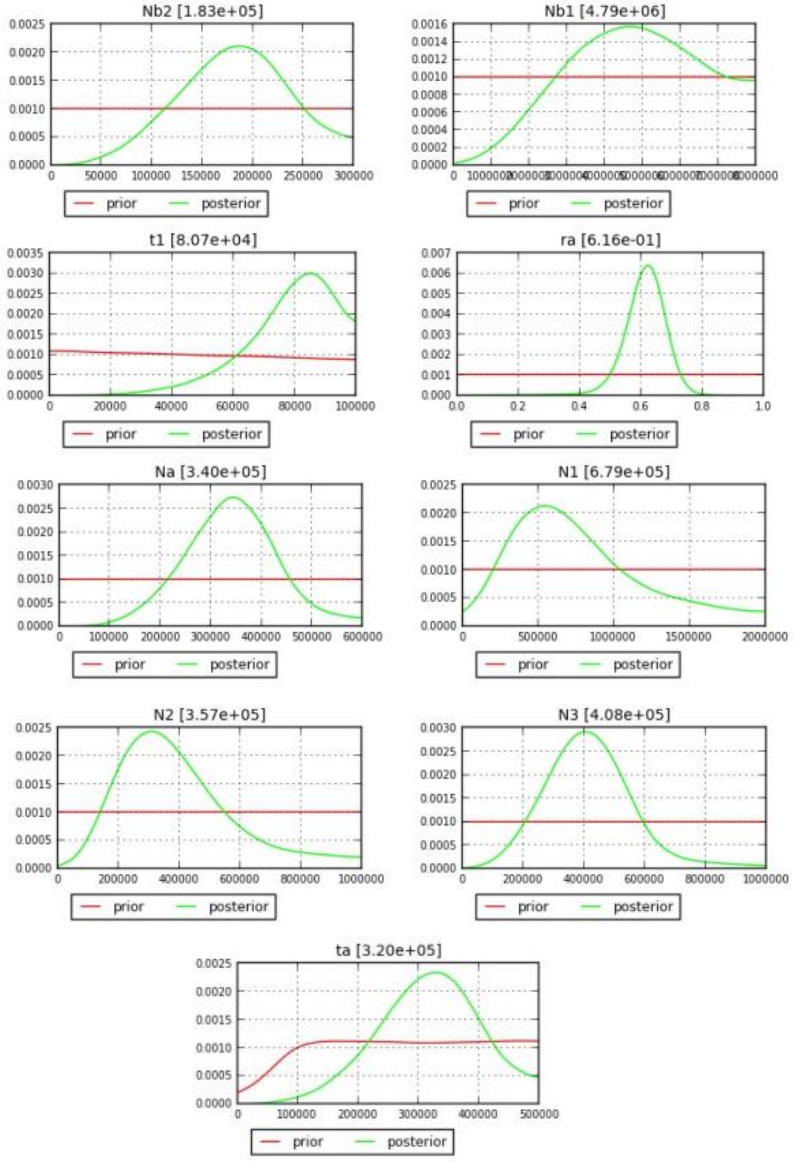


Figure S7. Admixturegraph showing migration events and proportion of admixture.

A


South American fur seal
Peruvian-Chilean fur seal
Galapagos fur seal
South American fur seal
Peruvian-Chilean fur seal
Galapagos fur seal
South American fur seal
Peruvian-Chilean fur seal
Galapagos fur seal
ra
0.61
1-ra
0.39
τ1
τ0
Scenario 1
Scenario 2
Scenario 3

B


Nb2 [1.83e+05]
Nb1 [4.79e+06]
t1 [8.07e+04]
ra [6.16e-01]
Na [3.40e+05]
N1 [6.79e+05]
N2 [3.57e+05]
N3 [4.08e+05]
ta [3.20e+05]
prior
posterior

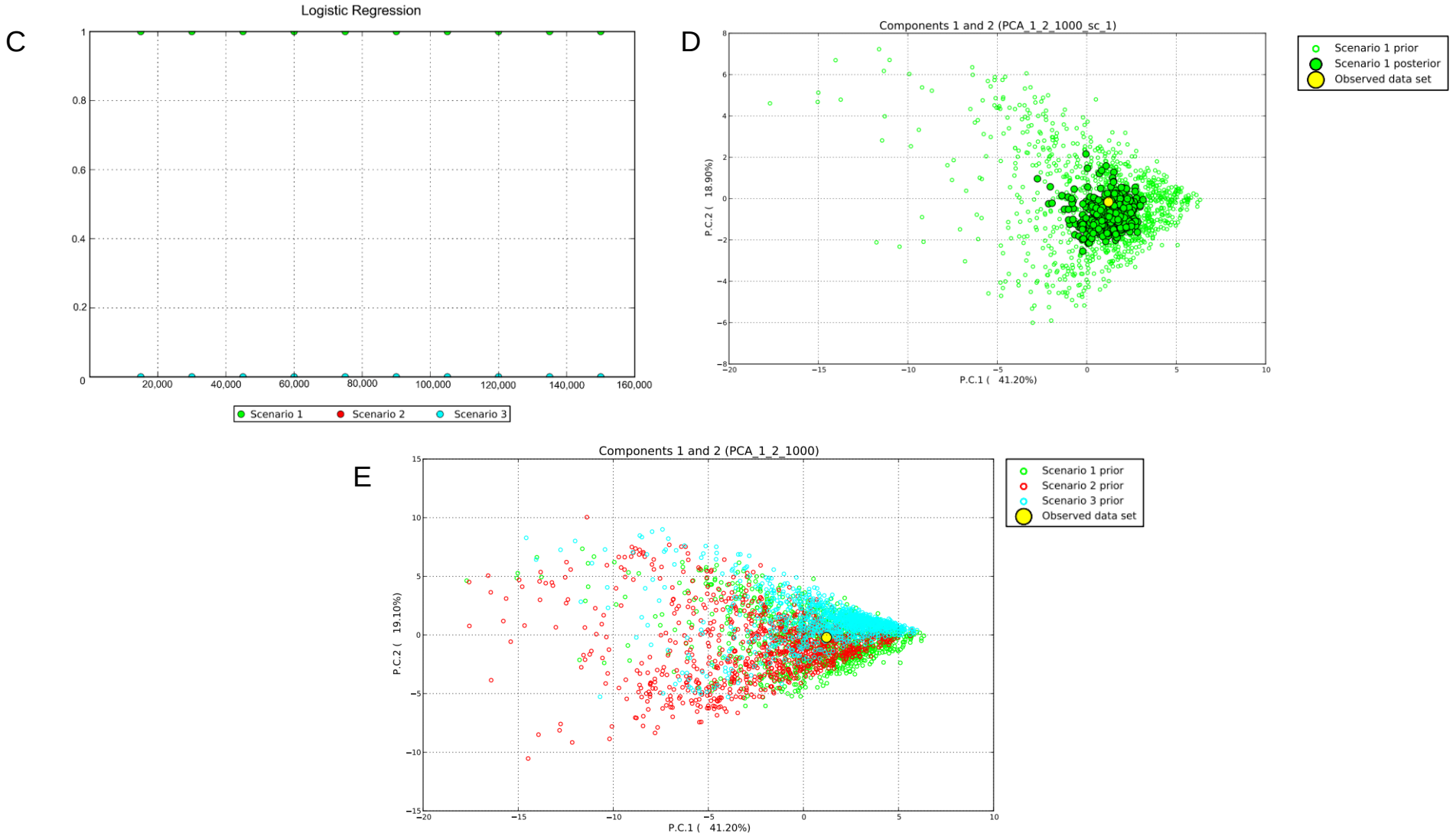


Figure S8. ABC analysis. (A) Comparison of scenatios; (B) parameter distributions; (C) scenario choice with logistic regression; (D) model checking; (E) PCA pre-evaluated scenario prior combination.

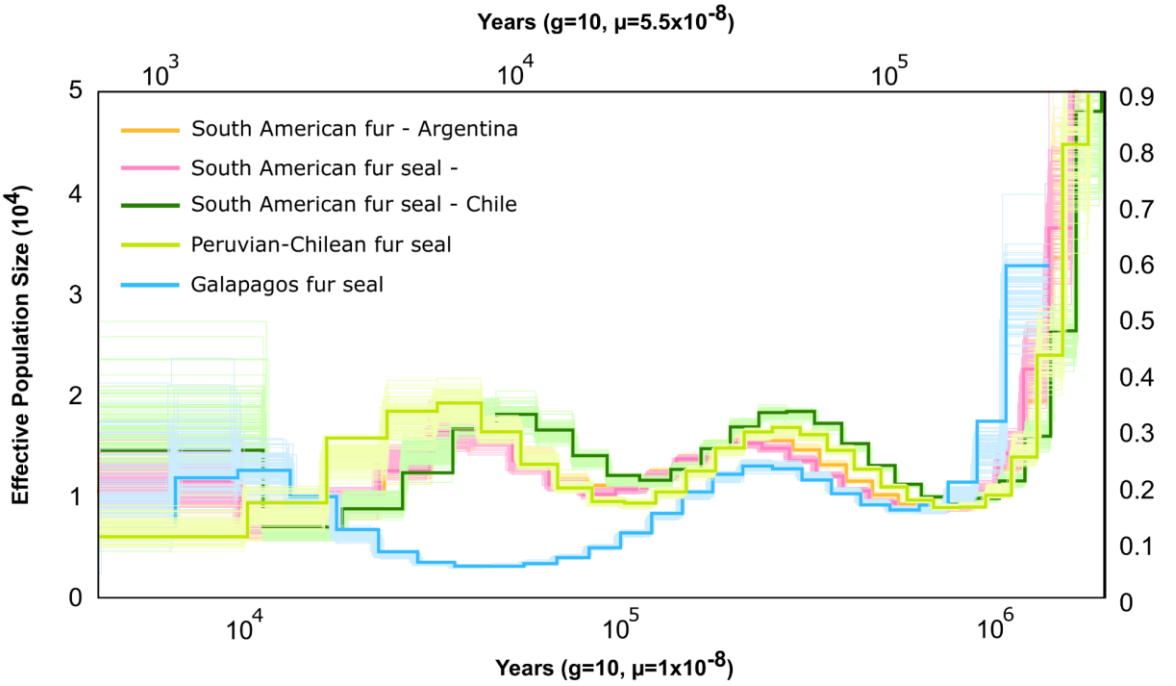


Figure S9. Historical *Ne* using PSMC analyses and their replicates for five fur seal genomes. The x-axis shows time, and the y-axis shows *Ne*. The left and bottom axis show the plot scaled using a conservative mammalian mutation rate (μ) of 1×10^{-8} substitutions nucleotide^{-1} generation^{-1} for a generation time (g) of 10 years. The above and right axis show the plot scaled using a mutation rate estimated from 300 genomic fragments of 50 kb (15 Mb) in a phylogeny for Otariidae (Lopes et al. in prep) and a mutation rate of 5.5×10^{-8} substitutions nucleotide^{-1} generation^{-1}.

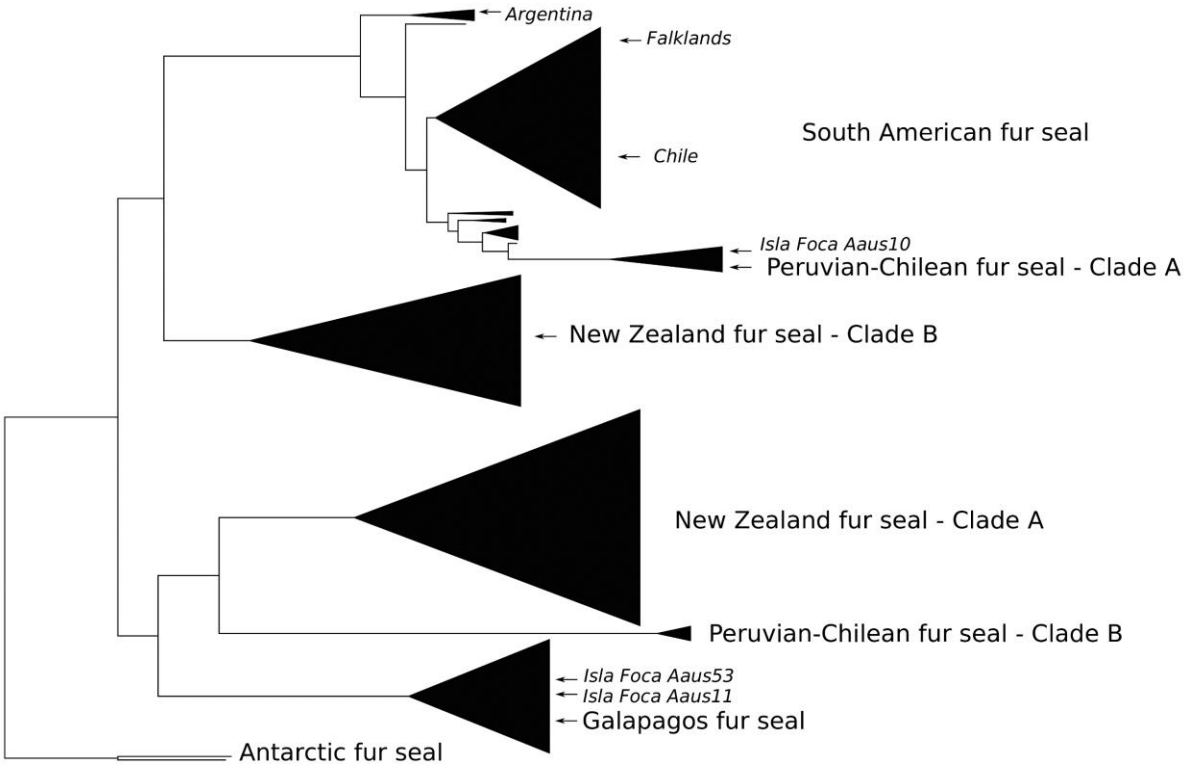


Figure S10. Mitochondrial control region Neighbor-joining. The arrows indicate the position of mtDNA captured from whole-genomes raw reads or the position of the samples of Isla Foca (see table S1 and S8)

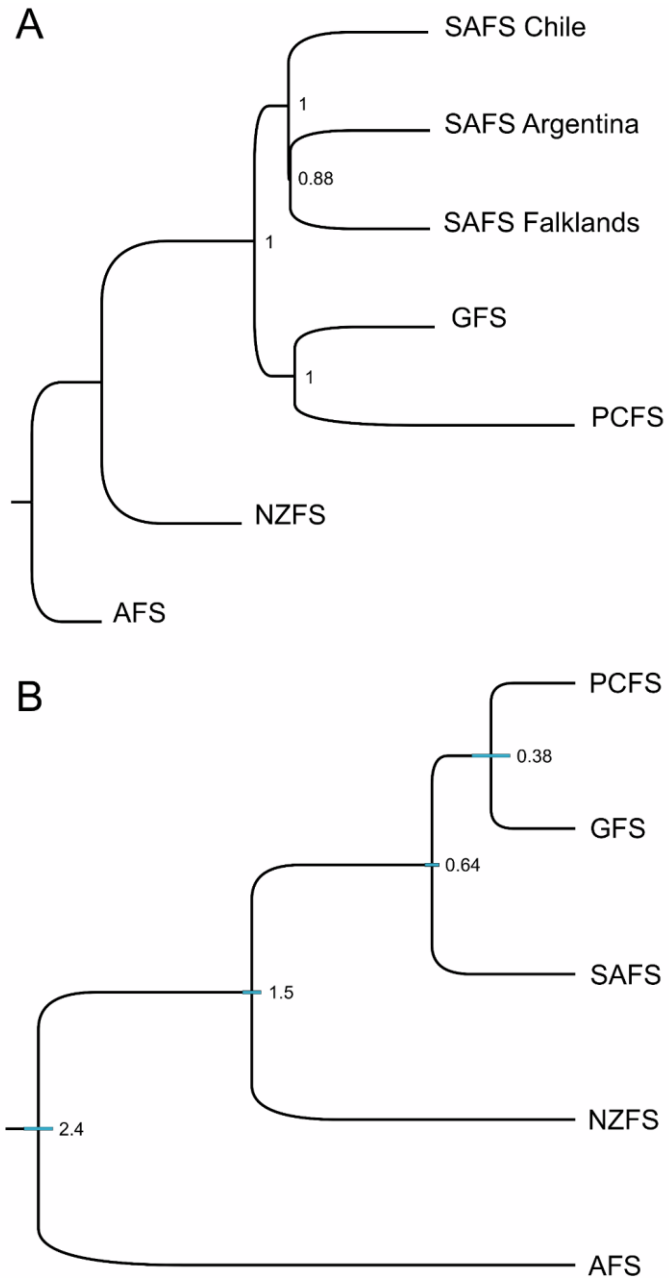


Figure S11. (A) Species tree recovered with ASTRAL-III. The node values are the support. (B) StarBEAST2. All nodes with HPD = 1. The values in the nodes are the point estimates of divergence times. Blue bars represent the HPD interval of 95%.

2. Conclusões gerais

Até o presente, todos os estudos que buscaram compreender a história evolutiva de Otariidae basearam suas análises filogenéticas, inferências biogeográficas (Lento et al. 1997; Wynen et al. 2001; Árnason e al. 2006; Yonezawa et al. 2009; Berta e Churchill 2012; Rodrigues et al. 2018) e até reclassificações taxonômicas (Berta e Churchill 2012; Nyakatura e Bininda-Emonds 2014; Berta et al. 2018) a partir de poucos fragmentos de DNA, com uma única exceção reportada na literatura científica (Higdon et al. 2007). Adicionalmente, todos os trabalhos que aplicaram mais de uma região gênica concatenaram as sequências de DNA (Higdon et al. 2007; Yonezawa et al. 2009; Nyakatura e Bininda-Emonds 2012) e não consideraram as discordâncias genealógicas ao longo dos genomas. Como resultado, diversas topologias muito diferentes foram propostas, resultando em uma interpretação confusa sobre a biogeografia de Otariidae (Yonezawa et al. 2009; Berta e Churchill 2012; Churchill et al. 2014).

Para estimar uma filogenia robusta e entender a história evolutiva de Otariide, nós analisamos genomas completos de 12 lobos e leões-marinhos utilizando métodos filogenéticos variados como neighbor-joining, máxima verossimilhança e bayesiano. Também buscamos compreender as fontes do conflito vigente nas relações de Otariidae. De uma maneira geral, as nossas diferentes análises convergiram para uma filogenia única e robusta, mostrando consistência nos resultados encontrados. Além disso, nossa filogenia concorda com os relacionamentos mais robustos publicados anteriormente (Yonezawa et al. 2009; Churchill et al. 2014), tais como um único evento de dispersão e colonização do hemisfério sul a partir de um grupo basal originado do hemisfério norte. Porém, nossas estimativas de tempo de divergência são diferentes das anteriores. Em nosso estudo, estimamos que o isolamento do clado do hemisfério sul ocorreu há aproximadamente 5 Mya, na transição entre o Mioceno e o Plioceno, período conhecido como "Plioceno quente" (5,3 – 2,6 Mya) (O'Dea et al. 2012). Nossas reconstruções filogenéticas também mostraram um rápido e intenso período de especiação ocorrido na transição entre o Plioceno e o Pleistoceno, há ~3 Mya, faixa de tempo quando emergiram ao menos 5 clados. Estes múltiplos eventos de especiação, que ocorreram quase simultaneamente, resultaram em um alto grau de discordância genealógica promovida pelo fenômeno da separação incompleta de linhagens gênicas (*ILS – Incomplete Lineage Sorting*). Neste segmento da árvore não detectamos a presença significativa de introgressão entre as linhagens, embora tenhamos identificado fluxo gênico interespecífico entre algumas espécies, porém de forma

menos preponderante. Dado que os eventos de especiação de lobos e leões-marinhos estão fortemente vinculados à temperatura superficial dos oceanos, os eventos de especiação provavelmente foram conduzidos pelo resfriamento das águas superficiais do Oceano Pacífico que, à época, também haveria sido influenciado pelo fechamento total do istmo do Panamá há ~3 Ma (O'Dea et al. 2012).

Mesmo com o surgimento quase que simultâneo de distintos clados e o alto nível de incongruência topológica ao longo dos genomas, as principais análises executadas, entre elas as árvores de espécies e máxima verossimilhança dos 10 maiores *scaffolds* (representam ~35% do genoma), apontaram para a monofilia de *Arctocephalus*, um resultado incomum em estudos moleculares publicados até então (Wynen et al. 2001; Árnason et al. 2006; Higdon et al. 2009; Yonezawa et al 2009; Berta e Churchill 2012; Nyakatura e Bininda Emonds 2012; Churchil et al. 2014; Berta et al. 2018). O forte suporte para a monofilia do gênero, recuperada com dados de genomas completos, também sugere a manutenção da nomenclatura taxonômica estabelecida na literatura e rejeita a reclassificação proposta por Berta e Churchill (2012).

A delimitação das espécies do clado que engloba o lobo-marinho-sul-americano/peruano-chileno (*A. australis*), lobo-marinho-de-Galápagos (*A. galapagoensis*) e o lobo-marinho-da-Nova Zelândia (*A. forsteri*) tem sido bastante discutida desde as primeiras descrições das espécies do grupo (Sclater 1987; Oliveira e Brownell 2014). Os últimos estudos de sistemática molecular, que avaliaram mais de um indivíduo por espécie, apontaram a parafilia mitocondrial das três espécies (Wynen et al. 2001; Yonezawa et al. 2009). Apesar disso, o Comitê de Taxonomia de Mamíferos Marinhos (Committee on Taxonomy 2018) considera estes três grupos como espécies plenas, também reconhecendo o lobo-marinho-peruano/chileno e o lobo-marinho-sul-americano como uma subespécie de *A. australis* (Oliveira et al. 2008; Berta e Churchill 2012; Committee on Taxonomy 2018).

As filogenias de árvores de espécies baseadas nos genomas completos, as PCAs (de genomas e bibliotecas de representação reduzida) e a análise de mistura para os 47 indivíduos sequenciados, mostraram que os lobos-marinhos-sul-americanos, de Galápagos e da Nova Zelândia formam grupos claramente delimitados, com a espécie da Oceania sendo o táxon basal. A análise de mistura indicou três clusters populacionais (K) como o mais provável, mostrando que o genoma de todos os indivíduos de Punta San Juan, a maior colônia do lobo-marinho-peruano/chileno, são compostos por componentes do lobo-marinho-de-Galápagos (entre 30 e 40%) e do lobo-marinho-sul-americano (entre 60 e 70%). Análises de mistura de genomas completos também indicaram que o genoma do lobo-marinho-peruano/chileno é composto por 40 a 60% de cada uma das espécies parentais.

Dado que em publicações anteriores o táxon peruano/chileno era considerado grupo irmão e derivado do lobo-marinho-sul-americano e que em nossas análises hora agrupava com este táxon, hora com o lobo-marinho-de-Galápagos, testamos, com análises ABC, o cenário mais provável para a origem do lobo-marinho-peruano/chileno. Dentre os três cenários avaliados, obtivemos com probabilidade posterior de 1.0 a origem híbrida deste táxon. Essas análises de ABC corroboraram as estimativas acima das proporções genômicas herdadas das espécies parentais, além de apontarem que a mistura que deu origem ao lobo-marinho-peruano/chileno ocorreu há possivelmente milhares de gerações no passado. Assim, diversas evidências, em conjunto, suportaram a origem híbrida e a história evolutiva independente do lobo-marinho-peruano/chileno: as análises de introgressão genômica, teste de cenários ABC, a longa história evolutiva recuperada nas filogenias mitocondriais, as análises de admixture K=4 clusters populacionais (que mostrou um componente populacional único para o lobo-marinho-peruano/chileno) e a trajetória demográfica do grupo. Outras características morfológicas (Oliveira et al. 2005, 2008), genéticas (Oliveira et al. 2008; Rodrigues et al. 2018), comportamentais e ecológicas (Pavés et al. 2016) também corroboram com a história independente e sustentam a origem híbrida (Masello et al. 2019). Além da sua história evolutiva independente, a origem híbrida do táxon peruano-chileno, que ainda mantem aproximadamente metade do seu genoma derivado de cada espécie parental, implica que não seria evolutiva e taxonomicamente correto considerar o táxon como um subgrupo infra específico de nenhuma das espécies parentais. Sendo assim, a solução mais adequada é considerar este táxon como uma espécie plena (cf. Oliveira e Brownell 2014).

Nosso estudo também mostrou que dois de três indivíduos coletados em Isla Foca, localizada a ~1000 km ao norte da distribuição do lobo-marinho-peruano-chileno e à leste da ilha mais próxima do arquipélago de Galápagos, contrariaram a identificação de coleta, que indicava serem o lobo-marinho-peruano/chileno, e foram atribuídos ao lobo-marinho-de-Galápagos (nos genomas mitocondriais e nucleares). O terceiro indivíduo, cujo genoma mitocondrial foi atribuído ao lobo-marinho-sul-americano, apresentou componentes nucleares do lobo-marinho-sul-americano e de Galápagos. Esses achados evidenciam a importância de Isla Foca nas dinâmicas populacionais, nas distribuições e na conservação do lobo-marinho-peruano/chileno e do lobo-marinho-de-Galápagos, até então considerado endêmico ao arquipélago de Galápagos.

Toda a costa peruana, incluindo Isla Foca, e o arquipélago de Galápagos estão no centro de eventos de *El Niño*, que ao afetar a produtividade primária na região do Pacífico Equatorial, afetam também toda a cadeia trófica e são responsáveis por drásticos declínios populacionais de lobos-marinhos da região (Oliveira et al. 2006, 2009; Trillmich 2015). Além disso, devido à indisponibilidade de alimento para forrageio, os eventos de *El Niño* induzem os lobos-marinhos-de-

Galápagos a dispersões erráticas, onde se perdem nas correntes marinhas e acabam encalhando muito distante de sua distribuição original, podendo ser encontrados nas regiões costeiras de países como México, Costa Rica, Guatemala, Equador e Colômbia (Capella et al. 2002; Aurioles-Gamboa et al. 2004; Félix et al. 2007; Montero-Cordero et al. 2010; Páez-Rosas et al. 2017; Quintana-Rizzo et al. 2017). Apesar disso, nosso estudo foi o primeiro a demonstrar que esta espécie é capaz de se reproduzir e gerar prole longe de suas principais colônias reprodutivas. Isso corrobora necessidade de futuros estudos em Isla Foca para que se possa compreender o seu papel enigmático na genética populacional dos lobos-marinhos-peruanos/chilenos e de Galápagos.

Embora tenhamos coberto a maioria das espécies de Otariidae no primeiro artigo, a filogenia da família não está completa, devido à ausência de *Eumetopias jubatus*, *Zalophus californianus* e especialmente *Neophoca cinerea.* Será importante sequenciar o genoma destas espécies e, desta forma, testar a posição basal de *E. jubatus* e *Z. californianus* em relação à *Arctocephalus* spp., *Otaria flavescens* e *P. hookeri*, bem como a posição da última como espécie irmã de *N. cinerea*, relação reportada na maioria dos estudos moleculares. Pretendemos, também, sequenciar mais genomas de lobos-marinhos da América do Sul (lobos-marinhos-peruanos/chilenos, sul-americano e de Galápagos) e da Oceania (lobo-marinho-da-Nova Zelândia) para decifrar o processo complexo de especiação até os grupos atuais, dadas a história contada pelas filogenias mitocondriais. Com mais genomas, também pretendemos definir, em fina escala, os processos de introgressão do lobo-marinho-da-Nova Zelândia para o lobo-marinho-sul-americano, definindo as proporções do aporte genômico da mistura em em cada uma das principais colônias reprodutivas avaliadas neste estudo (Uruguai/Brasil, Argentina, Falklands e Chile). Além disso, com mais genomas sequenciados, também pretendemos dar continuidade ao estudo e identificar e quantificar as composições das regiões genômicas a partir dos eventos de introgressão, bem como estimar precisamente o tempo das misturas e origem do táxon peruano/chileno.

Os artigos apresentados nessa tese ajudaram no entendimento da complexa história evolutiva de lobos e leões-marinhos, contribuindo e esclarecendo questionamentos fundamentais sobre as relações internas de Otariidae, fruto de dois séculos de debates científicos. Os artigos apresentados mostraram a importância de utilizar a abordagem genômica na interpretação da história evolutiva de grupos complexos. Aqui também evidenciamos a extensiva discordância genealógica existente ao longo dos genomas de lobos e leões-marinhos. Estes tiveram como fonte principal a separação incompleta das linhagens gênicas promovida por explosivas radiações adaptativas, eventos que se apresentaram como uma característica evolutiva de Otariidae. Nesta tese também demonstramos o importante papel do fluxo gênico interespecífico nas inconsistências topológicas e nas estimativas de

parâmetros biológicos, mas, principalmente, o de atuar como força evolutiva na geração de biodiversidade.

2.1 Referências bibliográficas

Árnason U, Gullberg A, Janke A, Kullberg M, Lehman N, Petrov EA, Väinölä R. 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. Mol Phylogenet Evol. 41:345-354.

Aurioles-Gamboa D, Schramm Y, Mesnick S. 2004. Galapagos fur seal, *Arctocephalus galapagoensis*, in Mexico. Latin J Aquat Mammal. 3:77-80.

Berta A, Churchill M. 2012. Pinniped taxonomy: Review of currently recognized species and subspecies, and evidence used for their description. Mamm Rev. 42:207-234.

Berta A, Churchill M, Boessenecker RW. 2018. The origin and evolutionary biology of pinnipeds: seals, sea lions and walruses. Annu Rev Earth Planet Sci. 46:203-228.

Capella JJ, Flores-Gonzalez L, Falk-Fernandez P, Palacios DM. 2002. Regular appearance of otariid pinnipeds along the Colombian Pacific coast. Aquat Mammal. 28:67-72.

Cárdenas-Alayza S, Oliveira L, Crespo E. 2016. *Arctocephalus australis*. The IUCN Red List of Threatened Species: e.T2055A45223529.

Churchill M, Boessenecker RW, Clementz MT. 2014. Colonization of the southern hemisphere by fur seals and sea lions (Carnivora: Otariidae) revealed by combined evidence phylogenetic and Bayesian biogeographical analysis. Zool J Linn Soc. 172:200-225.

Committee on Taxonomy of Marine Mammals. 2018. List of marine mammal species and subspecies. Available from https://www.marinemammalscience.org/species-information/list-marine-mammal-species-subspecies.

Félix F, Jiménez P, Falconí J, Echeverry O. 2007. New records and first births of the Galapagos fur seal, Arctocephalus galapagoensis (Heller, 1904), from the mainland coast of Ecuador. Rev Biol Mar Oceanogr. 42:77-82.

Higdon JW, Bininda-Emonds ORP, Beck RMD, Ferguson SH. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evol Biol. 7:216.

Lento G, Haddon M, Chambers GK, Baker CS. 1997. Genetic variation of Southern Hemisphere fur seals (*Arctocephalus* spp.): investigation of population structure and species identity. J Heredity 88:202-206.

Masello JF, Quillfeldt P, Sandoval-Castellanos E, Alderman R, Calderón L, Cherel Y, Cole TL, Cuthbert RJ, Marin M, Massaro M et al. 2019. Additive traits lead to feedinf advantage and reproductive isolation, promoting homoploid hybrid speciation. Mol Biol Evol. 36:1671-1685.

Montero-Cordero A, Fernández DM, Hernández-Mora G. 2010. Mammalia, Carnivora, Otariidae, Arctocephalus galapagoensis Heller, 1904: first continental record for Costa Rica. Check List 6:630-632.

Nyakatura K, Bininda-Emonds ORP. 2012. Updating the evolutionary history of Carnivora (Mammalia): A new species-level supertree complete with divergence time estimates. BMC Biol. 10:12. doi:10.1186/1741-7007-10-12.

O'Dea A, Hoyos N, Rodriguez F, Degracia B, De Gracia C. 2012. History of upwelling in the Tropical Eastern Pacific and the palaeogeography of the Isthmus of Panama. Palaeograph. Palaeoclimat Palaeoecol. 348-349:59-66.

Oliveira L, Arias-Schreiber M, Meyer D, Morgante JS. 2006. Effective population size in a bottlenecked fur seal population. Biol Conserv. 131:505-509.

Oliveira L, Brownell RL. 2014. Taxonomic status of two subspecies of South American fur seals: *Arctocephalus australis australis* vs. *A. a. gracilis*. Mar Mammal Sci. 30:1258-1263.

Oliveira L, Hingst-Zaher E, Morgante JS. 2005. Size and shape sexual dimorphism in the skull of the South American fur seal, *Arctocephalus australis* (Zimmermann, 1783) (Carnivora: Otariidae). Latin J Aquat Mammal. 4:27-40.

Oliveira L, Hoffman JI, Hingst-Zaher E, Majluf P, Muelbert MMC, Morgante JS. 2008. Morphological and genetic evidence for two evolutionarily significant units (ESUs) in the South American fur seal, *Arctocephalus australis*. Conserv Genet. 9:1451-1466.

Oliveira L, Meyer D, Hoffman J, Majluf P, Morgante JS. 2009. Evidence of a genetic bottleneck in an El Niño affected population of South American fur seals, *Arctocephalus australis*. J Mar Biol Assoc UK. 89:1717-1725.

Páez-Rosas D, Valdovinos LA, Elorriaga-Verplancken FR. 2017. Northernmost Record of the Galapagos Fur Seal (Arctocephalus galapagoensis): A Consequence of Anomalous Warm Conditions Around the Galapagos Archipelago. Aquat Mammal. 43:629-634.

Pavés H, Schlatter RP, Franco-Trecu V, Crespo E, Siefeld W, Araos V, Giesecke R, Batallés LM, Capozzo HL. 2016. Breeding season of the South American fur seal (*Arctocephalus australis*, Otariidae:Carnivora): New data for establishing independent evolutionary histories? Rev Biol Mar Oceano. 51:241-253.

Quintanna-Rizzo E, Garcia L, López RJ, Tobar-Hurtado S, López-Roulet A. 2017. First record of a Galapagos fur seal (Arctocephalus galapagoensis) in Guatemala; Mar Biodivers Record. 10:24.

Rodrigues P, Seguel M, Gutiérrez J, Pavés HJ, Verdugo C. 2018. Genetic connectivity of the South American fur seal (*Arctocephalus australis*) across Atlantic and Pacific oceans revealed by mitochondrial genes. Aquat Conserv Mar Fresh Ecosyst. 28:315-323.

Sclater P.L. 1897. On the distribution of marine mammals. Proc Zool Soc London p.349–359

Wynen LP, Goldsworthy SD, Insley SJ, Adams M, Bickham JW, Francis J, Gallo JP, Hoelzel AR, Majluf P, White RWG et al. 2001. Phylogenetic relationships within eared seals (Otariidae: Carnivora): implications for the historical biogeography of the family. Mol Phylogenet Evol. 21:270-284.

Yonezawa T, Kohno N, Hasegawa M. 2009. The monophyletic origin of sea lion and fur seals (Carnivora: Otariidae) in the southern hemisphere. Gene 441:89-99.

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