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

AMANDA KESSLER

GENÔMICA DA ORIGEM HÍBRIDA DO GOLFINHO *STENELLA CLYMENE*

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

PROGRAMA DE PÓS GRADUAÇÃO EM ECOLOGIA E EVOLUÇÃO DA BIODIVERSIDADE

GENÔMICA DA ORIGEM HÍBRIDA DO GOLFINHO *STENELLA CLYMENE*

Amanda Kessler

DISSERTAÇÃO DE MESTRADO

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL

Av. Ipiranga 6681 - Caixa Postal 1429

Fone: (051) 320-3500

CEP 90619-900 Porto Alegre - RS

Brasil

2019

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL

ESCOLA DE CIÊNCIAS

PROGRAMA DE PÓS GRADUAÇÃO EM ECOLOGIA E EVOLUÇÃO DA BIODIVERSIDADE

GENÔMICA DA ORIGEM HÍBRIDA DO GOLFINHO *STENELLA CLYMENE*

Amanda Kessler

Orientador: Dr. Sandro L. Bonatto

DISSERTAÇÃO DE MESTRADO

Porto Alegre - RS - Brasil

2019

Sumário

Agradecimentos

Primeiramente, gostaria de agradecer ao meu orientador Sandro Bonatto, que esteve ativamente presente em todas as etapas do projeto, que sempre esteve disposto a discutir os métodos e os resultados e que fez com que esse trabalho acontecesse da melhor maneira;

Ao meu colega de laboratório, co-autor do artigo e amigo, Fernando Lopes, que esteve presente em todas as etapas do projeto, foi um grande co-orientador no início da minha vida acadêmica na iniciação científica e que sempre esteve disposto a me ajudar seja no trabalho ou na vida pessoal;

Aos meus colegas de laboratório Manuel Escalona, Henrique Figueiró, Fernanda Trindade e Sarah Santos que me ajudaram em diversos momentos com auxílios de bioinformática e em discussões de artigos;

A Giovanna Oliveira, colega e técnica de laboratório, por toda a ajuda de bancada, pela manutenção e cuidado das amostras no laboratório

A minha amiga e colega de laboratório Vera de Ferran, que esteve presente nesses dois anos me apoiando tanto nas questões pessoais quanto no auxílio na edição de algumas imagens deste trabalho;

A todos os meus colegas do laboratório de Biologia Genômica e Molecular por todo o apoio diário;

As minhas amigas Paula Neves e Giovanna Carboni, que sempre acreditaram em mim e sempre estiveram dispostas e me ouvir e aconselhar;

Aos meus pais, que foram a minha maior fonte de inspiração e meu porto seguro, que não mediram esforços para que eu pudesse realizar todos os meus sonhos;

A Capes que me concedeu a bolsa de estudos;

Resumo

A hibridação tem se mostrado muito mais prevalente em animais do que se pensava anteriormente, em especial com o uso crescente de dados genômicos. Foi sugerido recentemente, com base em dados morfológicos e moleculares (mtDNA e poucos loci nucleares), que o golfinho Clymene, *Stenella clymene*, originou-se da hibridação entre *Stenella longirostris* e *Stenella coeruleoalba*. Aqui nós usamos os genomas nucleares e mitocondriais destas espécies e outras três espécies de golfinhos para testar esta hipótese e reconstruir sua história evolutiva. A árvore de espécies baseada nos genomas nucleares, altamente suportada, mostrou que *S. clymene* e *S. longirostris* são espécies irmãs, assim como *S. coeruleoalba* e *T. truncatus*, posicionadas em dois clados diferentes. Em contraste, na filogenia do genoma mitocondrial, *S. longirostris* e *S. coeruleoalba* trocaram de posição, na qual *S. coeruleoalba* é irmã de *S. clymene* e *S. longirostris* está distantemente relacionada a elas, próxima de *T. truncatus*. Análises de mistura e introgressão não mostraram evidência de *S. clymene* como espécie híbrida, mas encontramos fortes evidências de um evento de introgressão no qual cerca de 40% do genoma do ancestral de *S. clymene* e *S. longirostris* veio de *S. coeruleoalba*. Esses resultados e o padrão de discordância mito-nuclear sugerem uma introgressão bidirecional, mas com viés de sexo (feminino), que teria substituido completamente (trocado) os genomas mitocondriais originais de *S. longirostris* e *S. coeruleoalba*. Encontramos também evidências de introgressão (13%) dos ancestrais de *S. longirostris*, *S. clymene* e *D. delphis* em *S. coeruleoalba* e um nível muito pequeno (1,4%) de introgressão de *S. frontalis* em *S. longirostris*.

Palavras-chave: Delphininae; Genoma Completo; Filogenômica; Introgressão

Abstract

Hybridization has been shown to be much more prevalent in animal than previously thought, in special with the increasing use of genomic data. It has been suggested recently, based on morphological and molecular (mtDNA and few nuclear loci) data that the Clymene Dolphin, *Stenella clymene*, originated from the hybridization between *Stenella longirostris* and *Stenella coeruleoalba*. Here we use the whole-nuclear and mitochondrial genomes of these species plus three other delphinines to test this hypothesis and reconstruct their evolutionary history. A highly supported nuclear genome species tree shown that *S. clymene* and *S. longirostris* are sister species as well as *S. coeruleoalba* and *T. truncatus*, located in two different clades. In contrast, in mtDNA genome phylogeny *S. longirostris* and *S. coeruleoalba* switched position, in which *S. coeruleoalba* is sister to *S. clymene*, and *S. longirostris* is closer to *T. truncatus*. Admixture analyses shown no evidence of *S. clymene* as a hybrid species, but we found strong evidence of an introgression event in which about 40% of the genome of the ancestor of *S. clymene* and *S. longirostris* came from *S. coeruleoalba*. These results and the pattern of mito-nuclear discordance suggest a bidirectional but sex (female)-biased introgression that completely replaced (exchanged) the original mitochondrial genomes of *S. longirostris* and *S. coeruleoalba*. We find evidence of introgression (13%) of the ancestral of *S. longirostris*, *S. clymene* and *D. delphis* into *S. coeruleoalba* and a very small (1.4%) level of introgression of *S. frontalis* into *S. longirostris.*

Keywords: Delphininae; Whole genome; Phylogenomics; Introgression

Apresentação

A espécie *Stenella clymene* (Gray, 1846) é endêmica do Oceano Atlântico e tem sua distribuição ao longo de águas quentes do Golfo do México e Mar do Caribe, sendo exclusivamente oceânica (Fertl, 2003). Os registros mais ao norte são de Nova Jersey (39° 17' N) no Atlântico ocidental, e no Atlântico oriental na Mauritânia (16° 13' O). Os registros mais ao sul são ao sul do Brasil (29° 58' S), no Atlântico central a 3°40' S da Ilha de Ascensão (Fertl et al., 2003) e a leste próximo ao norte de Angola (Perrin & Mead, 1994). Apenas em 1981 *Stenella clymene* foi considerada uma espécie válida e ainda é uma das espécies menos estudadas do gênero. Até então, os indivíduos da espécie eram considerados como pertencentes à *S. longirostris*, porém como um grupo com diferenças morfológicas, sendo mais robustos menores no comprimento (Perrin, 1981; Jefferson, 1994) e pesando aproximadamente 80 kg (Jefferson et al., 1995). O formato do corpo do animal e as suas extremidades são similares a indivíduos de outras espécies do gênero e seu corpo possui um padrão de cores tripartidas, tendo sua parte dorsal preta, os lados cinza e o ventre branco (Perrin et al., 1981). *Stenella clymene*, apesar de apresentar características muito similares à *S. coeruleoalba* e *S. longisrostris*, se sustenta como espécie plena devido ao fato de que, baseando-se em suas características morfológicas, não é possível atribuí-la exclusivamente a qualquer uma das duas espécies por último mencionadas (Perrin et al., 1981).

Stenella longirostris (Gray, 1828) é o pequeno cetáceo mais comum nas águas pelágicas tropicais. Tem sua distribuição nas zonas tropicais de todo o globo e em algumas zonas subtropicais (Perrin, 2009). Seu tamanho varia entre 1,29m e 2,35m e tem o peso máximo de 79kg (Perrin, 2005). Seu fenótipo é reconhecido por seu padrão de três cores ao longo do dorso, porém pode apresentar variação de cor e tamanho do crânio e rostro (Perrin, 1998). *Stenella coeruleoalba* (Meyen, 1833), assim como as outras espécies, está distribuída na região pelágica e possui sua distribuição similar a *S. longirostris,* se distribuindo nas zonas tropicais e subtropicais do globo. Possui um padrão de coloração característico, que consiste em manto dorsal negro, lateral cinza e ventre branco, além de duas faixas negras que seguem perpendiculares ao corpo, uma do olho ao ânus e outra do olho às nadadeiras peitorais (Maia-

Nogueira et al., 2001). Os indivíduos adultos desta espécie podem medir até 2,6 metros de comprimento, os são machos ligeiramente maiores que as fêmeas e seu peso máximo é cerca de 160 kg (Jeferson, 1994).

A similaridade morfológica entre *S. clymene* e *S. longirostris*, indicava que ambas eram relacionadas com *S. coeruleoalba*, porém, em análises filogenéticas foi possível observar que *S. longirostris* é mais basal que *S. coeruleoalba*, surgindo assim, a hipótese que *S. clymene* é uma espécie que teria se originado de eventos de hibridação (Leduc et al., 1999). Outro estudo discordou desta hipótese, pois não encontrou *S. clymene* como intermediária entre as duas espécies, estando agrupada com *S. longirostris* e *Delphinus delphis*, excluindo *S. coeruleoalba* (Kingston, 2009). Recentemente foram encontradas evidências morfológicas, de DNA mitocondrial e de seis loci de DNA nuclear favoráveis à hipótese de que *S. clymene* é intermediária entre *S. longirostris* e *S. coeruleoalba* (Amaral et al., 2014). O resultado do DNA mitocondrial mostrou que *S. clymene* está mais estreitamente relacionada com *S. coeruleoalba*, enquanto no DNA nuclear mostrou que está mais relacionado com *S. longirostris*. Introgressão ou *ILS* (*Incomplete lineage sorting*) poderiam explicar essas variações no genoma. Estes eventos poderiam explicar os diferentes padrões encontrados na rede de haplótipos nucleares, e talvez ainda afete o genoma destas espécies, dado que este grupo possui uma divergência recente, porém não explica totalmente a discordância entre o DNA mitocondrial, os loci de DNA nuclear e a variação morfológica de *S. clymene*, o que levou os autores a suportar a hipótese de origem híbrida da mesma (Amaral et al., 2014). Estudos recentes (p. ex. vonHoldt et al., 2016; Nilsson et al., 2017; Àrnason et al., 2018) mostraram que análises com genomas completos apontam diferentes cenários evolutivos que apenas alguns genes e marcadores, portanto, análises que levem em consideração o genoma como um todo são necessárias para entender e medir eventos, como no caso deste trabalho, de hibridação, introgressão e *ILS*. Estas análises são necessárias para poder determinar se a espécie *S. clymene* surgiu por especiação híbrida ou se ocorreram outros eventos que poderiam causar a discordância nas filogenias do genoma mitocondrial e do genoma nuclear como visto em Amaral (2014).

O artigo resultante desta dissertação será submetido à revista *Molecular Biology and Evolution* e o manuscrito está de acordo com as normas desta revista.

Referências da Apresentação

Amaral, AR, Lovewell G, Coelho MM, Amato G, Rosenbaum HC. 2014. Hybrid speciation in a marine mammal: the Clymene dolphin (*Stenella clymene*). PLoS ONE. 9: 1 – 8.

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

Fertl D., Jefferson TA, Moreno IB, Zerbini AN, Mullin KD. 2003. Distribution of Clymene dolphin *Stenellaclymene*. Mammal Society, Mammal Review. 33: 253–271.

Gray, J.E. 1846. On the cetaceous animals. In: J. Richardson e J.E. Gray (Eds), The zoology of the voyage of H.M.S Erebus and Terror, under the command of Captain Sir James Clark Ross. E.W. Jason, London. 1 (3): 13 – 53.

Jefferson TA, Leatherwood S, Webber MA. 1994. Species identification guide. Marine mammals of the world. 320: 587.

Jefferson TA, Odell DK. & Prunier, K T. 1995. Notes on the biology of the Clymene dolphin (Stenellaclymene) in the northern Gulf of Mexico. Marine Mammal Science. 11: 564–573.

Kingston SE, Adams LD, Rosel PE. 2009. Testing mitochondrial sequences and anonymous nuclear markers for phylogeny reconstruction in a rapidly radiating group: molecular systematics of the Delphininae (Cetacea: Odontoceti: Delphinidae). BMC Evolutionary Biology. 9: 245.

Leduc RG, Perrin WF, Dizon AE. 1999. Phylogenetic relationships among the Delphinid cetaceans based on full cytochrome b sequences. Marine Mammal Science. 15: 619-648.

Maia-nogueira R, Farias TS, Cunha IF, Dórea-reis LW, Braga FL. 2001. Primeiro registro de Stenella coeruleoalba Meyen, 1883 (cetácea-delphinidae) no litoral do estado da Bahia, incluindo uma revisão da espécie em águas brasileiras. Bioikos, 15: 45- 49.

Nilsson MA, Zheng Y, Kumar V, Phillips MJ, Janke A. 2017. Speciation Generates Mosaic Genomes in Kangaroos. Genome Biol Evol. 10(1): 33-44.

Perrin WF (1998) *Stenella longirostris*. Mammalian Species.599: 1-7

Perrin WF, Dolar MLL, Chan CM, Chivers SJ. 2005. Length-weight relationships in the spinner dolphin. Marine Mammal Science. 21: 765-778.

Perrin WF, Mitchell ED, Mead JG, Caldwell DK, Vanbree PJH. 1981. *Stenella clymene*, a rediscovered tropical dolphin of the Atlantic J Mammal 62: 583–598.

Perrin WF, Rosel PE, Cipriano F. 2013. How to contend with paraphyly in the taxonomy of the delphinine cetaceans. Mar Mammal Sci 29: 567–588

Perrin WF and Mead JG. 1994. Clymene dolphin *Stenella clymene* (Gray, 1846). Handbook of marine mammals: the first book of dolphins (S. H. Ridgway and R. Harrison, eds.). Academic Press, London, United Kingdom 5: 1–416. Pp. 161–171

Perrin WF, Mitchell ED, Mead JG, Caldwell DK, Caldwell MC, Bree PJH, Dawbin WH. 1987. REVISION OF THE SPOTTED DOLPHINS, STENELLA SPP. Marine Mammal Science. 3(2): 99–170.

vonHoldt BM, Cahill JA, Fan Z, Gronau I, Robinson J, Pollinger JP, Shapiro B, Wall J and Wayn RK. 2016. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. 2(7): e1501714.

The Clymene Dolphin (*Stenella clymene*) is not a Hybrid Species (kind of): Whole-Genome Analysis of Dolphin Evolution

Amanda Kessler[1], Fernando Lopes[1], Ana Carolina Oliveira de Meirelles[2], Vitor Luz Carvalho[2], Ana Lúcia Cypriano Souza[1], Sandro Bonatto[1]

[1]Laboratório de Biologia Genômica e Molecular, sala 134, Faculdade de Biociências, PUCRS, Porto Alegre - RS, Brasil.

[2]Associação de Pesquisa e Preservação de Ecossistemas Aquáticos (AQUASIS)

Abstract

Hybridization has been shown to be much more prevalent in animal than previously thought, in special with the increasing use of genomic data. It has been suggested recently, based on morphological and molecular (mtDNA and few nuclear loci) data that the Clymene Dolphin, *Stenella clymene*, originated from the hybridization between *Stenella longirostris* and *Stenella coeruleoalba*. Here we use the whole-nuclear and mitochondrial genomes of these species plus three other delphininae to test this hypothesis and reconstruct their evolutionary history. A highly supported nuclear genome species tree shown that *S. clymene* and *S. longirostris* are sister species as well as *S. coeruleoalba* and *T. truncatus*, located in two different clades. In contrast, in mtDNA genome phylogeny *S. longirostris* and *S. coeruleoalba* switched position, in which *S. coeruleoalba* is sister to *S. clymene*, and *S. longirostris* is closer to *T. truncatus*. Admixture analyses shown no evidence of *S. clymene* as a hybrid species, but we found strong evidence of an introgression event in which about 40% of the genome of the ancestor of *S. clymene* and *S. longirostris* came from *S. coeruleoalba*. These results and the pattern of mito-nuclear discordance suggest a bidirectional but sex(female)-biased introgression that completely replaced (exchanged) the original mitochondrial genomes of *S. longirostris* and *S. coeruleoalba*. We find evidence of introgression (13%) of the ancestral of *S. longirostris*, *S. clymene* and *D. delphis* into *S. coeruleoalba* and a very small (1.4%) level of introgression of *S. frontalis* into *S. longirostris.*

Introduction

Hybridization occurs when individuals from two populations or more generally distinct species cross (Harrison, 1990). The long-term consequences of these events depend largely on the viability and fertility (or not) of the various possible types of hybrid offspring and their frequency. When some hybrids have at least partial reproductive success (including backcrossing), hybrid zones may occur, a range between two areas where parent species reside, in which species meet and reproduce (Harrison & Larson, 2016). In some cases, the effects of hybridization may extend beyond this contact zone, generating introgression of alleles of one species into the genome of individuals of the other species. Many hybrid zones have already been investigated, for example, in *Heliconius* butterflies in French Guiana (Shaak & Counterman, 2017). Examples of ancient introgressions are also well known, one of the most recently studied being that of hominid populations, with evidence of introgressions between modern man, Neanderthal, and Denisova man (McCoy et al., 2017) and between chimpanzees and bonobos (De Manuel et al., 2016). In felids, studies have shown both cases of hybridization and introgression as well as cases of recent hybrid zones (Trigo et al., 2013; Li et al., 2016). In Cetaceans, there are many reports of individuals that may be hybrids between different species of the same genus, or even of different genera (Willis et al., 2004), mainly in captivity (Crossman et al., 2016). While most of these reports are based on morphology, there have been also genetic evidences of more extensive hybridization in dolphins (e.g. Antoniou et al., 2018).

On the other hand, an apparently much rarer situation is hybrid speciation, which is when hybridization played a key role in the formation of a new species, such as when a new species or lineage originated from an or a few hybridization events (Mallet, 2007). When a pattern emerges in hybrids, such as niche differentiation, sexual selection of individuals of the same species and other factors, where parental alleles recombine to form new genetic associations, they lead to new phenotypes that may be better adapted to other niches and resources, managing to remain as a new species (Tobler et al., 2010).

The hybridization speciation phenomenon seems to be relatively common in plants (Vallejo-Marín et al., 2016), but much rarer in animals, especially vertebrates, with reports with different degrees of evidence in fish, birds and mammals (Meier et al., 2016; Peters & Kleindorfer, 2015; Shurtliff, 2013). In the latter, the most studied case of possible hybrid origin is that of the North American red wolf. Several genetic studies in recent years, the most recent being a genomic analysis, have shown that red wolves are highly mixed with varying degrees of mixing with coyotes and gray wolves. These results are not yet conclusive on a completely hybrid origin (between coyote and wolfhound) of this species with subsequent backcrossing at different sites with parental species. Another possibility is that the red wolf was an independent species that was almost extinct and that current individuals are descended from extensive backcrosses of these with coyote or gray wolf (vonHoldt et al., 2016).

It was been hypothesized (Leduc RG. et al., 1999) that *Stenella clymene* was a hybrid species that emerged from the admixture between *Stenella longirostris* and *Stenella coeruleoalba*. The phenotype and the behavior of *S. clymene* was more similar to *S. longirostris,* but in the phylogeny of the mtDNA cytochorme b, *S. coeruleoalba* was closely related to *S. clymene* and more distantly related to *S. longirostris*. A more recent study supported this previous mtDNA results and also showed that using six nuclear genes *S. clymene* was more closely related to *S. longirostris* than to *S. coeruleoalba* (Amaral et al. 2014). However, recent studies (e.g. vonHoldt et al., 2016; Nilsson et al., 2017; Àrnason et al., 2018) shown that a genomic approach is sometimes necessary to disentangled complex evolutionary scenarios caused by past hybridizations and incomplete lineage sorting that produced discordant gene trees and difficult species tree estimation. Here we sequenced the whole-genome of five species (*S. clymene*, *S. longirostris, S. coeruleoalba*, *S. frontalis,* and *D. delphis*) and together with *T. truncatus* reconstructed their species tree, estimated past hybridization events and tested the hybrid origin hypothesis for *S. clymene*.

Results

Genome sequences

We obtained the genome sequence of five Delphinidae species (*S. clymene*, *S. longirostris*, *S. coeruleoalba*, *S. frontalis*, *D. delphis* sequenced in this paper and *T. truncatus* from GenBank), with average coverage of 25X (Table 1) and ~80% of the *reads* were mapped.

Table 1. Coverage and percentage mapped after filtering and mapping against *O. orca* reference genome (size = 2.3Gb).

Species	Coverage	% mapped
Stenella clymene	23X	80%
Stenella longirostris	20X	78%
Stenella coeruleoalba	42X	73%
Stenella frontalis	25X	85%
Delphinus delphis	25X	81%
Tursiops truncatus	21X	83%

The largest scaffold is ~37 Mb and the 25 largest sums around 590 Mb (Table 2). The BEDtools mask pipeline identified ~40% of repetitive regions in the reference *Orcinus orca* assembly. Most of our analyses were performed on the genomes maskered for these repetitive regions.

After removing repetitive sites, missing data and fragments with less than 50% of information (from the multi-species alignments), we obtained 12,799 GFs of 10 kb, 11,411 GFs of 20 kb, 9,294 GFs of 50 kb and 7,143 GFs of 100 kb. The number of average variable sites between *S. clymene* and *S. longirostris* in the GFs of 10kb is 25 (Fig. 1A) and in 50kb is 140, the great majority being larger than 50 (Fig. 1B). Parsimony informative sites (excluding the outgroup) in the alignments presented a very similar pattern (Figs. 1C,D) suggesting all GF sizes are variable enough to resolve with high confidence a phylogeny with only seven taxa.

Table 2. Sizes of the 25 largest scaffolds used in the analyses.

Scaffold	Size (bp)
Scaffold1	37,310,048
Scaffold2	34,915,538
Scaffold3	31,328,679
Scaffold4	29,916,295
Scaffold5	29,316,040
Scaffold6	28,814,452
Scaffold7	26,669,673
Scaffold8	26,420,779
Scaffold9	25,195,141
Scaffold10	24,737,879
Scaffold11	23,960,635
Scaffold12	23,539,721
Scaffold13	20,796,922
Scaffold14	20,204,063
Scaffold15	20,171,943
Scaffold16	20,021,286
Scaffold17	19,848,842
Scaffold18	19,595,934
Scaffold19	19,427,489
Scaffold20	19,204,407
Scaffold21	18,650,016
Scaffold22	18,595,296
Scaffold23	18,110,099

Scaffold24	18,014,134
Scaffold25	17,688,799
Total	592,454,110

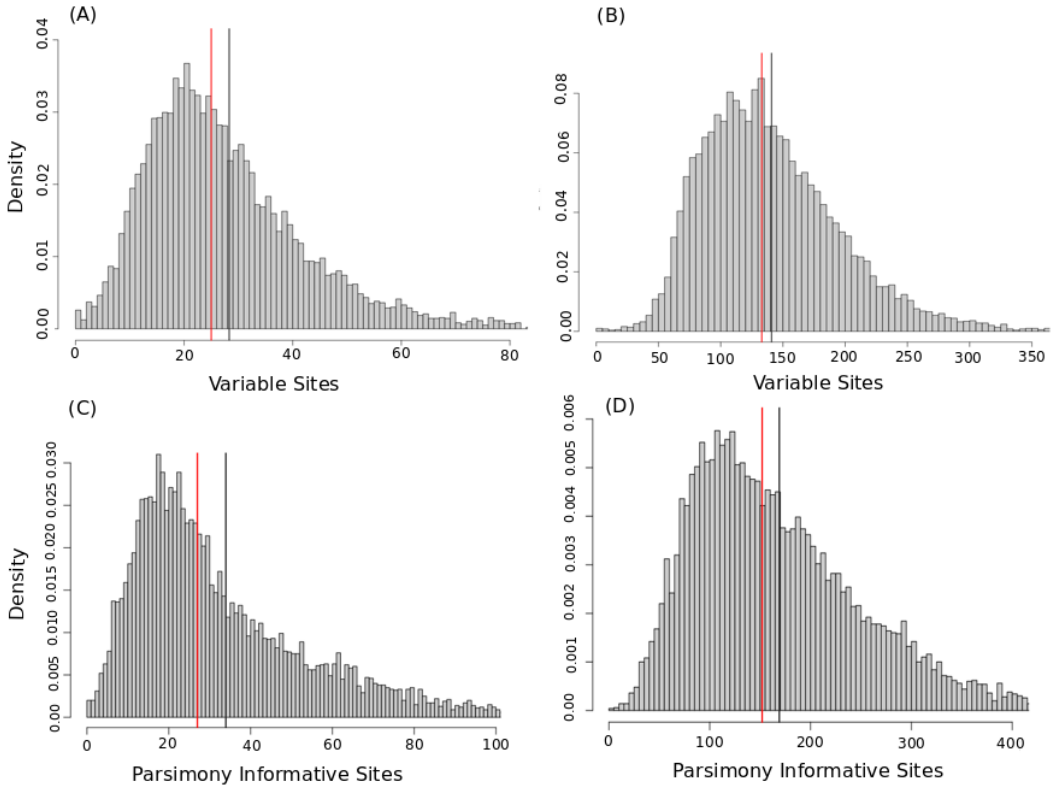

Fig.1. Variability of the GFs. (A,B) Variable sites between *S. clymene* and *S. longirostris* in 5,000 randomly sampled GFs of 10 kb (A) and 50 kb (B). (C,D) Parsimony informative sites in 5,000 (except for the outgroup) in GF of 10 kb (C) and 50kb (D).

Phylogenetic inferences

The species trees for all window sizes reconstructed with ASTRAL III using the ML trees of the GFs produced the same species tree: ((((*S. clymene, S. longirostris*) *D. delphis*), *S. frontalis*), (*S. coeruleoalba*, *T. truncatus*)), *O. orca*), all branches with posterior probability of 1 (Fig. 2A). The species trees estimated with StarBeast2 and BPP presented the same topology as above also with posterior probability of 1 for all

internal branches (Fig. 2B). 13 of the ML trees of the 25 biggest scaffolds also produced the above species tree with high support while the other 12 scaffolds produced an alternative topology, in which *S. frontalis* is sister to the other five species (except for the *O. orca*). Although the species tree is also the most frequent tree in the GFs of 20, 50 and 100 kb (for the GF of 10 kb, the most frequent tree presented *S. frontalis* as sister of *D. delphis*), several other topologies with similarly high frequencies appeared in the genome windows trees (Table 3).

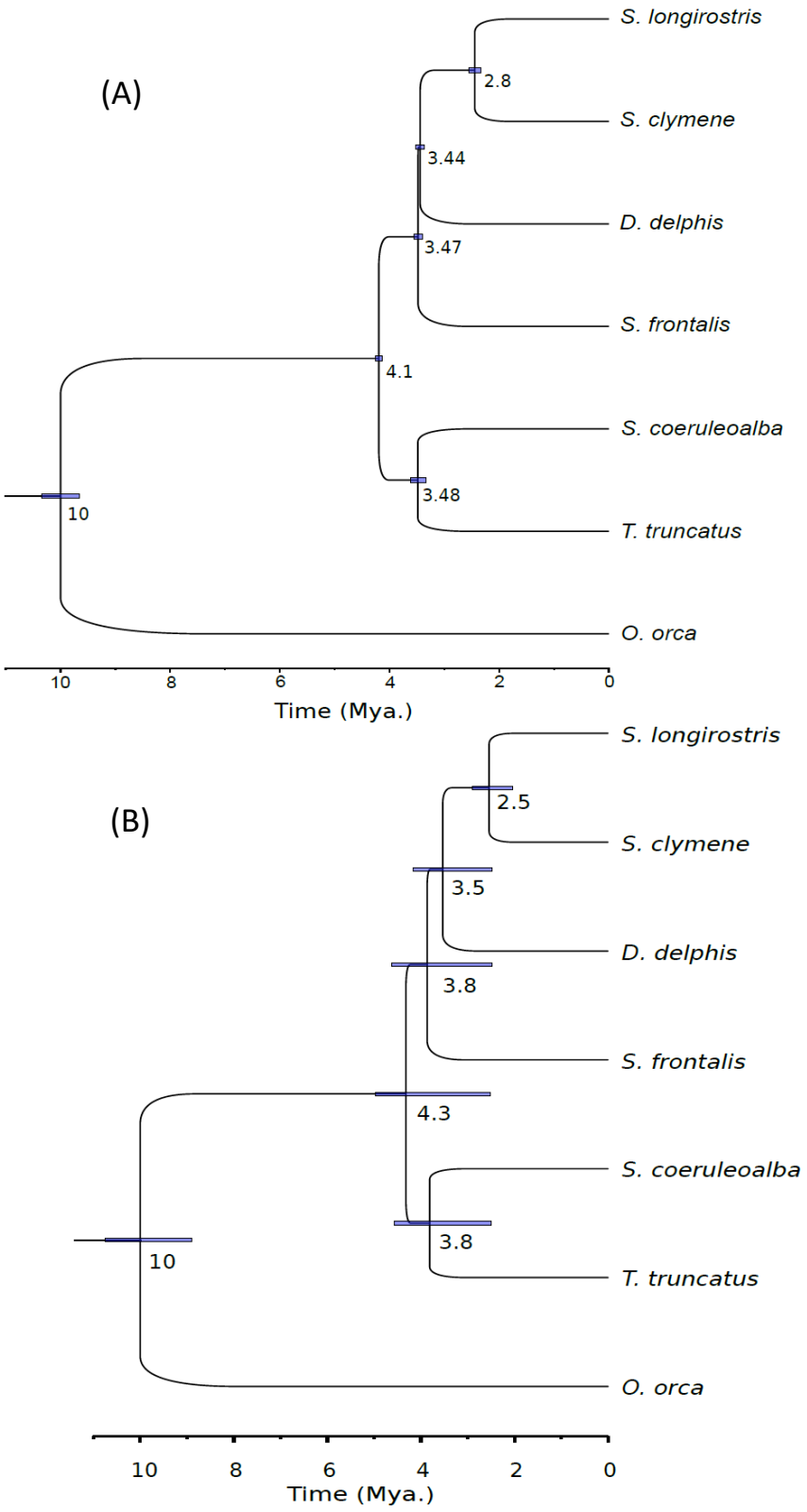

Fig. 2. Species trees divergence time estimated with 300 GFs of 50kb (A) with a Bayesian inference using StarBEAST2 and (B) with BPP, using the species tree topology. Divergence times are near the nodes and the posterior probability for all nodes are 1.

Table 3. The ten most frequent GF trees estimated with different windows sizes in RAxML. The numbers represent absolute frequency. Total frequency is: 10kb = 12,799; 20kb = 11,411; 50kb = 7,143; 100kb = 7,143.

Topology	10kb	20kb	50kb	100kb
(((*D. delphis*,(*S. clymene*,*S. longirostris*)),*S. frontalis*),(*S. coeruleoalba*,*T. truncatus*)),*O. orca*)	243	377	517	565
(((*D. delphis*,(*S. clymene*,*S. longirostris*)),(*S. coeruleoalba*,*T. truncatus*)),*S. frontalis*),*O. orca*)	264	352	454	515
(((*D. delphis*,(*S. clymene*,*S. longirostris*)),((*S. coeruleoalba*,*T. truncatus*),*S. frontalis*)),*O. orca*)	249	295	442	499
(((*D. delphis*,*S. frontalis*),(*S. clymene*,*S. longirostris*)),(*S. coeruleoalba*,*T. truncatus*)),*O. orca*)	231	277	355	378
(((*D. delphis*,(*S. clymene*,*S. longirostris*)),*S. coeruleoalba*),(*S. frontalis*,*T. truncatus*)),*O. orca*)	216	268	357	370
(((*D. delphis*,*S. frontalis*),(*S. coeruleoalba*,*T. truncatus*)),(*S. clymene*,*S. longirostris*)),*O. orca*)	192	218	254	277
(((*D. delphis*,*S. frontalis*),((*S. clymene*,*S. longirostris*),(*S. coeruleoalba*,*T. truncatus*))),*O. orca*)	167	266	355	266
(((*D. delphis*,(*S. clymene*,*S. longirostris*)),(*S. coeruleoalba*,(*S. frontalis*,*T. truncatus*))),*O. orca*)	191	226	245	211
(((*D. delphis*,*S. frontalis*),((*S. clymene*,*S. longirostris*),(*S. coeruleoalba*,*T. truncatus*))),*O. orca*)	167	262	292	266
(((((*D. delphis*,(*S. clymene*,*S. longirostris*)),*S. coeruleoalba*),*T. truncatus*),*S. frontalis*),*O. orca*)	154	143	192	184

The analysis of divergence times of the phylogenies showed that both methods (BPP and Beast) infer similar divergence times (Fig. 2). We set the root as 10 Mya with *O. orca* as outgroup. The divergence time of the ingroup occur ~4.3 Mya, *S. clymene* and *S. longirostris* were the most recent separation, ~2.5 Mya, and *S. coeruleoalba* were more basal than *S. clymene* with divergence time of ~3.8 Mya.

The mitochondrial genome ML tree (Fig. 3), which has high support values for most nodes, with a few exceptions, although presented many similarities with the nuclear genome species tree, has one important difference related to the positions of *S. clymene* and *S. coeruleoalba*. We used a tanglegram (Fig. 4) to highlight the similarities and differences between the nuclear and mitochondrial genome topologies (pruning the mtDNA tree to the same taxa presented in the nuclear tree). We could see that the main difference in that *S. coeruleoalba* and *S. longirostris* switch positions: in nuclear genome, *S. clymene* is sister taxa to *S. longirostris* and in mitochondrial genome it is the sister taxa to *S. coeruleoalba*. The other difference is that in the mtDNA tree *S. longirostris* is not sister to *T. truncatus* but is basal to Delphininae with exception of *S. attenuata*.

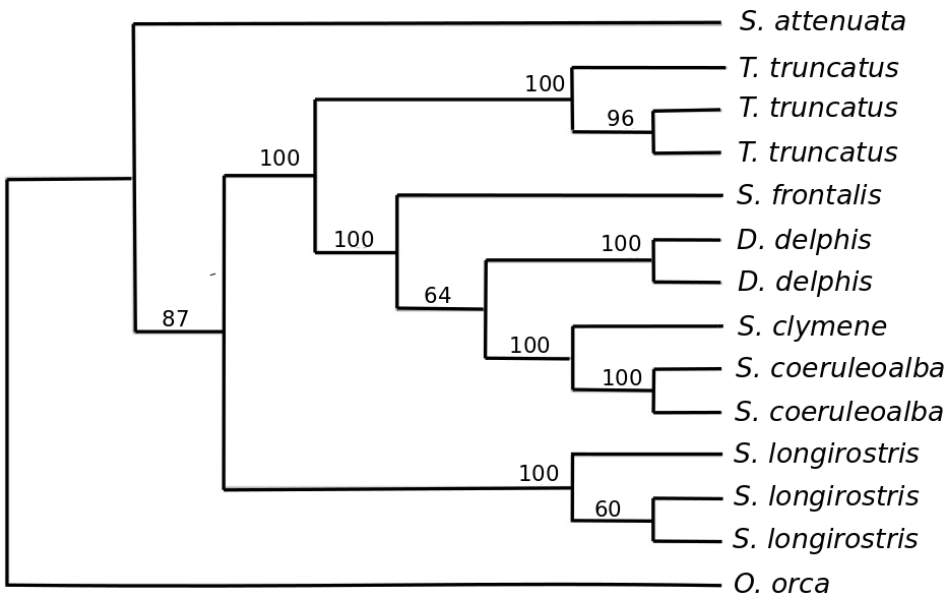

Fig. 3. ML consensus tree of 100 bootstrap trees of the whole mitochondrial genome, with support above the branches.

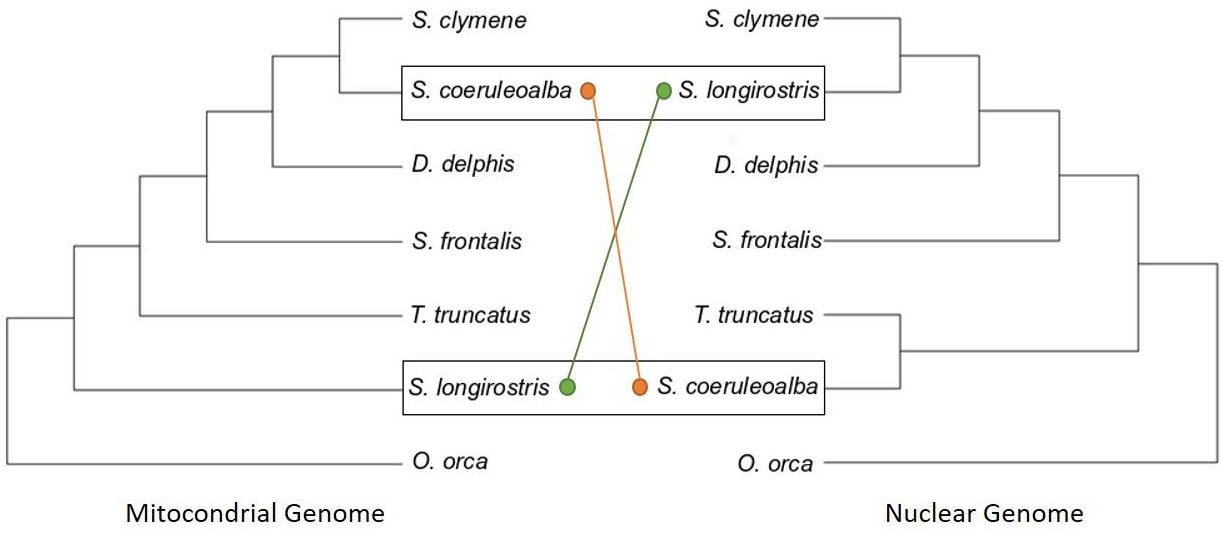

Fig. 4. Comparison of mitochondrial and nuclear genomes topologies. The rectangles show the switch in position between *S. longirostris* and *S. coeruleoalba*.

Introgression Analysis

The TreeMix backbone phylogeny again is identical to the species tree and identified three significant signals of admixture (Fig. 5). The highest migration event was found between *S. coeruleoalba* into the *S. longirostris* and *S. clymene* ancestor, with a value (ancestry proportion) of 0.38. This analysis also suggested introgression (0.13) of the ancestral of *S. longirostris*, *S. clymene* and *D. delphis* into *S. coeruleoalba* and a very small (0.014) level of introgression of *S. frontalis* into *S. longirostris.*

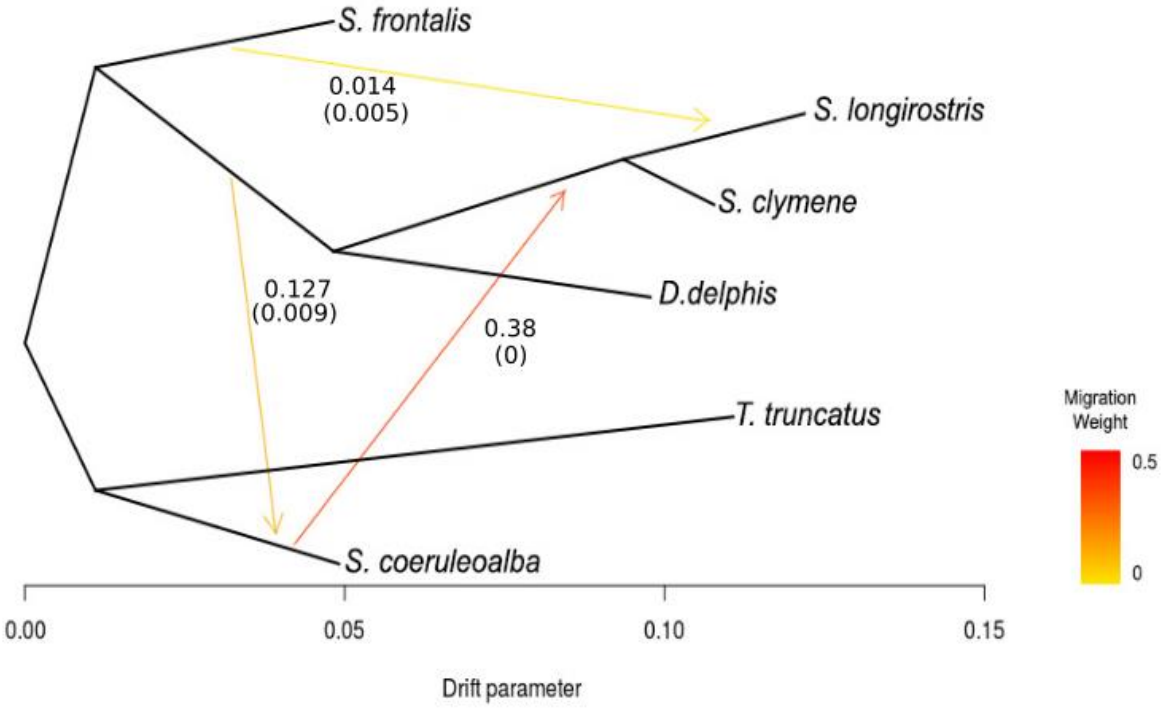

Fig. 5. TreeMix tree with migration events. Arrows are admixture events and color indicate the migration weight (varies between 0 and 0.5), with the values shown above the arrows.

We used admixture graph to estimate admixture proportion in two scenarios, one where there was an admixture between *S. coeruleoalba* and the ancestor of *S. clymene* and *S. longirostris*, as estimated here by TreeMix (Fig. 6A) and the other as suggested in recent literature, where *S. clymene* originated from an admixture between *S. coeruleoalba* and *S. longirostris* (Fig. 6B). Our results clearly reject the hypothesis of the hybrid origin of *S. clymene*, as the proportion of the genome of *S. clymene*, not shared with *S. longirostris*, that came from *S. coeruleoalba* is <0.01 (Fig. 5). On the other hand, we estimated that the admixture proportion of the common ancestor of *S. clymene* and *S. coeruleoalba* that came from *S. coeruleoalba* is 0.39, very similar to the value (0.38) estimated by TreeMix (Fig. 5).

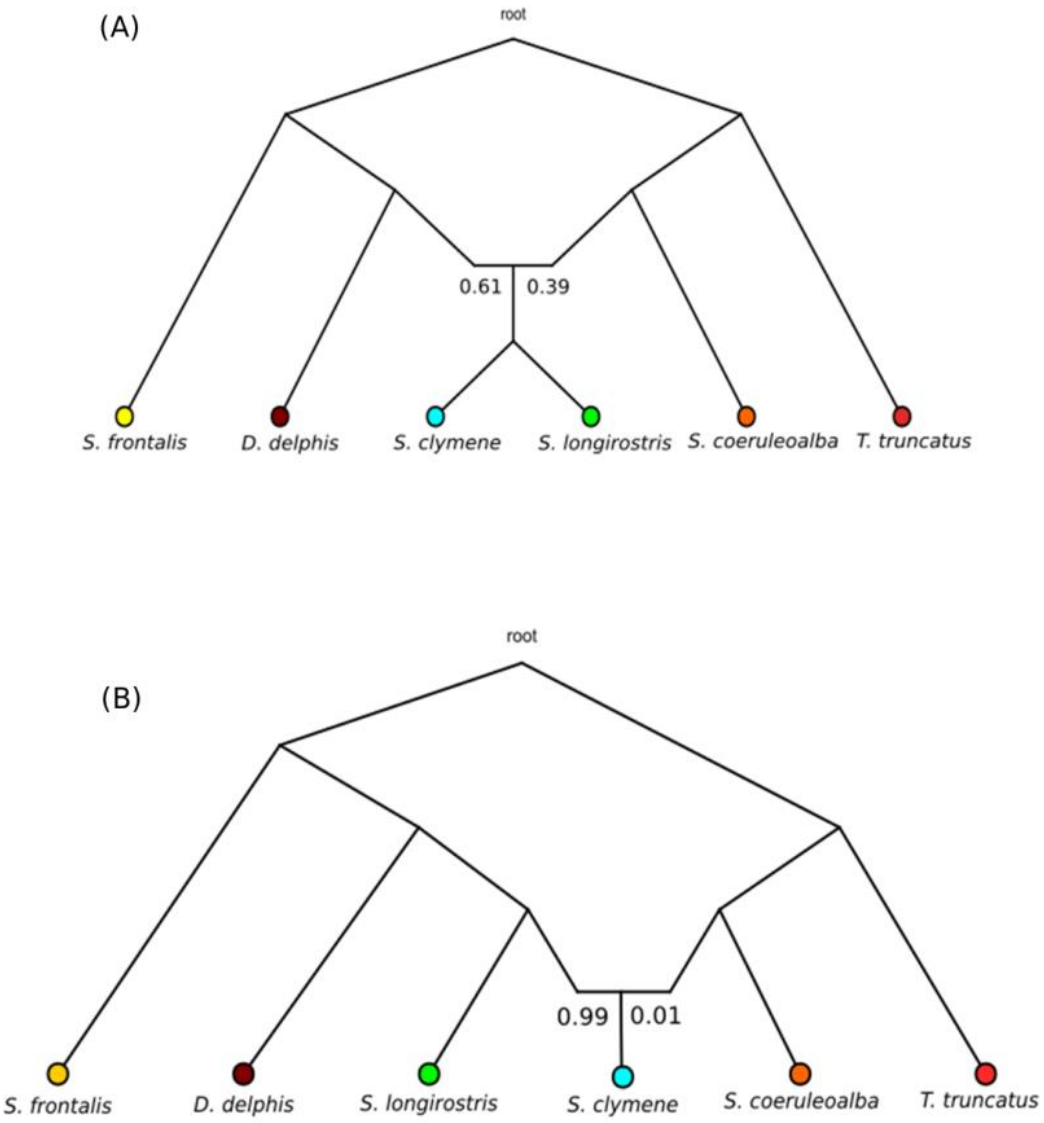

Fig. 6. Admixture graphs based on the two main hypotheses of *S. clymene* origin. (A) Admixture between *S. coeruleoalba* and the ancestor of *S. clymene* and *S. longirostris* (B) *S. clymene* originated from an admixture between *S. coeruleoalba* and *S. longirostris.* Values are admixture proportions based on *f4*-statistics fitting on the admixture graphs.

Admixture results (Fig. 7) with K = 5 genetic components grouped *S. clymene* and *S. longirostris*, with K = 4 grouped the latter two with *S. coeruleoalba* while with K = 3 the groups formed were: *S. clymene* and *S. longirostris*; *S. coeruleoalba* and *T. truncatus*; *S. frontalis* and *D. delphis.*

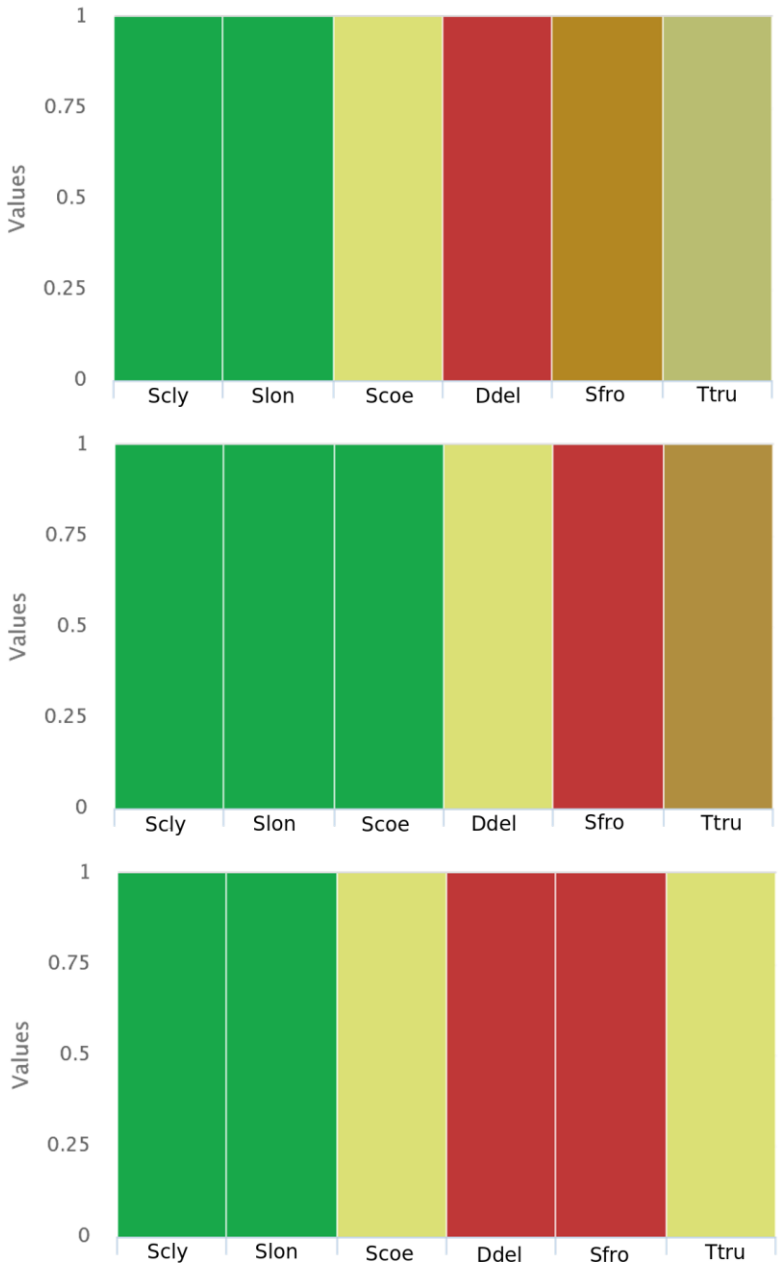

Fig. 7. Admixture plot for K=5-3 (from top to bottom) using 307,200 variants sites.

Discussion

By using whole genomes of several dolphins, in special *S. clymene* and its putative parental species, *S. longirostris* and *S. coeruleoalba*, we presented a first approach for the genomic history of this group and tested the hybrid speciation hypothesis. We showed that there is no evidence that *S. clymene* is a hybrid species between the above-mentioned parental species (Kingston et al., 2009; Amaral et al..,

2014), but that it is the sister species of, and closely related to, *S. longirostris* and that it was their ancestor that hybridized (or introgressed) with *S. coeruleoalba*. We found no evidence of any *S. coeruleoalba* introgressed genetic information into *S. clymene* genome that was not also found in *S. longirostris* genome (Fig. 6). This introgression event seems to have been very important, since about 40% of the genome of present-day *S. clymene* and *S. longirostris* came from *S. coeruleoalba* as opposed to be derived from their common ancestor with *D. delphis*.

Most studies suggested many important topological incongruencies between nuclear and mitochondrial Delphininae phylogenies (e.g. Amaral et al., 2014). Many of these incongruences were likely caused by the wrong nuclear species tree used and most by erroneous mtDNA topologies caused by the use of small fragments and finally some by ILS between close related species or limited recent mtDNA introgression. However, our mitochondrial genome phylogeny strongly corroborates most studies that suggest that at the mtDNA, *S. clymene* and *S. coeruleoalba* are sister species and that *S. longirostris* is very distantly related to them, usually in a basal position and sometimes closer to *T. truncatus* (e.g. Zurano et al., 2019), as found here (Fig. 3). This strong mito-nuclear discordance was one of the main evidences, together with intermediate morphological traits, put forward to support the hybrid origin hypothesis (Amaral et al., 2014). However, when we compare our nuclear vs mitochondrial genome phylogenies (the latter was simplified to facilitate comparison, Fig. 4) we could see that, aside from the non-sister relationship between *S. coeruleoalba* and *T. truncatus*, the only difference between the two trees is that *S. longirostris* and *S. coeruleoalba* changed position with each other. Therefore, the phylogenetic position of *S. clymene* mtDNA is the correct one as expected by its position in the species tree, that is, grouped with *D. delphis* and with *S. frontalis* while the position of *S. longirostris* is not compatible with the species tree. Therefore, contrarily with what have been assumed so far, not only *S. clymene* is not a hybrid species but its mtDNA was not introgressed from *S. coeruleoalba*, but present-day *S. longirostris* mtDNA seems to have been inherited (introgressed) from *S. coeruleobalba* and, at the same time, present-day *S. coeruleoalba* mtDNA seems to have been completely replaced by *S. longirostris* mtDNA. This is a complex scenario that implies a reciprocal exchange of mitochondrial genomes between these species, presumably during *S. coeruleoalba* and *S. longirostris* + *S. clymene* ancestor hybridization. This implies that it should have

occurred a bidirectional but sex-biased (female) introgression that completely replaced the original mtDNA of both species. Introgression hybridization, likely bidirectional, has been described recently between *S. coeruleoalba* and *D. delphis* in the Greek Seas (Antoniou et al., 2018). Complete mtDNA replacement of one species by another species via unidirectional sex-biased introgression associated with hybridization (or gene flow) has been described in several species (e.g. Good et al., 2008, Nevado et al., 2009, Zielinski et al., 2013; Trigo et al., 2013; Llopart et al., 2014; Seixas et al., 2018). However, to our knowledge this seems to be the first example of bidirectional complete replacement of mtDNA between two species. Some authors support an adaptative explanation for complete mtDNA replacement (selective advantage of the introgressed genome, eg. Llopart et al., 2014) while others support purely demographic processes (e.g. Seixas et al., 2018). We have not tested these explanatory hypotheses here. We also do not have at this moment an explanation why *S. clymene* maintained the original (ancestral) mtDNA lineage while it was replaced in *S. longirostris*, although the biogeographical differences between them, the former being restricted to the Atlantic Ocean and the latter being distributed worldwide (together with *S. coeruleoalba*), may have an important role (see e.g. Towes et al., 2012 and Trigo et al., 2013). A phylogeographical approach to deal with population differentiation within species, as seen for example in *S. clymene* (Amaral et al., 2014, Nara et al. 2017) would be important to better understand this complex evolutionary history.

We found a high degree of topological discordance between the GFs trees estimated along the genome (Table 3). This discordance could not be attributed to lack of information of the GFs, since there is a high number of informative and variable sites even in the 10 kb GFs (Fig. 1) and the discordance was not reduced in larger window sizes (50 and 100 kb) or even with the use of scaffolds larger than 15 million bases. Topological discordance between genomic regions is not unusual and is being observed with increasing frequency in recent phylogenomic studies (e.g. Li et al., 2016) including the Cetacean Balaenopteridae (e.g. Àrnason et al., 2018). The sources of these gene-tree discordances are attributed mainly to incomplete lineage sorting (ILS) or to hybridization (Bravo et al. 2019). Here we support at least three hybridization events between these species, the one between *S. coeruleoalba* and *S. longirostris* + *S. clymene* ancestor involved the transference of about 40% of the

genome (Fig. 5). We also found that some internal branches are very small, suggesting short intervals between speciation events, that is one of the main parameters known to increase ILS (Degnan & Rosenberg, 2006), as also suggested previously (Amaral et al., 2012). These two phenomena likely explained the pattern of high topological discordance found here.

This high level of discordance along the genome, with several topologies presenting very similar frequencies may explain the almost complete absence of congruence between phylogenies among and within studies on Delphininae phylogeny (see Perrin et al., 2013). The studies that usually estimated the same phylogeny were those that used the same genes (e.g. cytochrome b). Despite all these gene-tree differences, our whole-genome species tree analyses consistently supported with high confidence a single topology, where *S. clymene* and *S. longirostris* as sister taxa forming a clade with *D. delphis* and *S. frontalis*, and *S. coeruleoalba* sister to *T. truncatus* in basal clade. We did not find this specific topology in any previous phylogeny, although Amaral et al. (2014) species tree (Fig. 5) shared the most relevant clades. Our results also strongly support that *Stenella* is not monophyletic as found in most studies and for most Delphininae genera (see Perrin et al., 2013), although our study is not informative for the other two genera as we have a single species for each of them.

There are surprisingly few studies with divergence time estimates for Delphininae species with some differences between them. Part of the variation may be caused by the different gene regions used that, as we showed here, may have very different topologies and also partially because the set of species used may vary greatly between the phylogenies. Most studies point to an age of the set of species of Delphininae studied here between ~3.5 Mya to ~4.5 Mya (e.g. McGowen et al., 2009; Vilstrup et al., 2011; Banguera et al., 2014; Amaral et al., 2018), with a few extreme dates such as 9.9 Mya (Chen et al., 2011). Our preliminary divergence time analysis suggested the diversification of our studied Delphininae between 4.3-4.1 Mya (Fig. 1), supporting the main results above. Besides, our results are also compatible with the suggested age of the oldest delphinine of approximately 4 Mya (Bianucci, 2013).

Material and Methods

Genome Sequencing and Assembly

Tissue DNA extraction was performed using Qiagen DNAeasy Blood and Tissue Kit following the manufacturer's instructions. The quantity and quality of the samples were tested by agarose gel. One sample each of *Stenella clymene*, *Stenella longirostris*, *Stenella coeruleoalba*, *Stenella frontalis* and *Delphinus delphis* was selected for whole genome sequencing. All the samples are from Brazilian coast (Table 4). The taxonomic identification of these samples was previously confirmed through visual inspection in the field, the control region mtDNA compared with the DNA surveillance site (Ross et al., 2003) and our own database and the genetic profile with eight microsatellites (unpublished results). Whole genome libraries were constructed with TruseqNano and TruSeq DNA PCR Free kits with insert size of 350bp and sequenced using Illumina HiSeqX plataform (paired-end reads with 150bp). *Tursiops truncatus* (GCA_001922835.1) and the reference used for mapping the raw reads, *Orcinus orca* (GCF_000331955.1) were retrieved from GenBank.

Table 4. Sample numbers and locations.

Species	Sample number	Latitude	Longitude
S. longirostris	G103	-23,66771	-41,73
S. frontalis	G167	-27,11759	-47,68
S. clymene	G132	-27,41863	-46,45
D. delphis	G181	-23,76393	-45,23
S. coeruleoalba	295	-4,70238	-37,348

Paleomix v. 1.2.5 (Schubert et al., 2014) pipelines were used to trim the raw reads, discard reads that were <100bp in length and with Phred-score <30 with AdapterRemoval v.2 (Schubert et al., 2016) and single reads were discarded. BWA v. 0.7.17 (Li &Durbin, 2009) was used to map the selected reads against *Orcinus orca* genome. The adapters and PCR duplicates were removed with Picard Tools v. 2.18.15 (broadinstitute.github.io/picard/) and indels were locally realigned by GATK 3.8 (McKenna et al., 2010). The regions annotated as repetitive for *Orcinus orca* genome

were excluded in BEDtoolsv.2.27.0 (Aaron et al., 2010) with maskfasta parameter. The consensus of the filtered scaffolds by Paleomix was made with ANGSD (Korneliussen et al., 2014) with the following parameters: -dofasta 2, that chooses the most common base and -explode, that write out all the scaffolds.

Genomic windows

For some analyses the genomes were separated into genomic fragments (GFs) of size 10, 20, 50 and 100 kb (thousands basepairs), separated by 100 kb to reduce linkage disequilibrium between the fragments. Sites with missing data were removed with trimAl (Capella-Gutierrez et al., 2009) and fragments with <50% of the original window size was excluded. To estimate how many sites are parsimony informative between the species in each window size, 5,000 fragments for each window size were randomly selected and the number of parsimony informative sites among all species was calculated. We also estimated the number of variable sites between the two closest related species, *S. clymene* and *S. longirostris*.

Phylogenetic analysis and divergence time estimation

First, we estimated a maximum likelihood (ML) phylogenetic tree for each of the 25 largest scaffolds in the assembly, using RAxML (Stamakis, 2014) with the following settings: one best tree for each fragment, with 100 bootstrap replicates and the GTR+G model. We also estimated a ML tree for each GF for each window size (10, 20, 50 and 100 kb) with the same settings above. Then, using the best trees for all GFs, a species tree was inferred for each window size using ASTRAL III (Zhang et al., 2018) multi-species coalescent model. Further, a Bayesian species tree was estimated with the StarBeast2 model of BEAST 2 v2.5.2 (Suchard et al., 2018), using 300 randomly choosen gene fragments of size 50 kb. The settings were: a chain length of 5,000,000 storing every 10000, Yule tree Model, HKY substitution model, linked clock models and a strict clock with rate of 1.0, to be calibrated after the analysis (see below). Tracer v.1.7 (Rambaut et al., 2014) was used to visualize the Markov Chains (MCMC) and the ESS values (Estimated Sample Sizes). TreeAnnotator v 2.5.1 (Rambaut and Drummond 2010) was used to estimate the maximum clade credibility (MCC) tree. The MCC species tree was visualized in FigTree v1.4.3 (Rambaut, 2017) and calibrated using 10 million years ago (Mya) as the age of the root (*O. orca* vs.

Delphininae), that is an average of previously studies with nuclear genes (Cunha et al., 2011; Banguera-Hinestroza et al., 2014).

BPP (Flouri et al., 2018), a Bayesian species tree estimation method under the multispecies coalescent model was used with the same 300 random loci of 50 kb used in the StarBeast2 analysis. BPP was used only to estimate divergence times since we used the species tree estimated with the methods described above as a fixed topology. The settings for the BPP run were, a MCMC chain of 400,000 replicates with burnin of 10,000, with theta prior of 0.001 and tau prior of 0.001. 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.

Mitochondrial genome phylogeny

We also reconstruct the mitochondrial genome of the five species we obtained the nuclear genome by mapping the raw reads to a published *Tursiops truncatus* mitochondrial genome using the Paleomix pipelines described above. The mtDNA genomes were aligned in MEGA X (Kumar et al., 2018) with several other Delphininae available genomes, including *Stenella attenuata* (Table 5), resulting in a 16,374 bp alignment. A maximum likelihood tree, with *O. orca* as outgroup, was then inferred using RAxML with the same settings as above and with 100 bootstrap replicates, that were summarized in a majority rule consensus tree.

Table 5. Mitochondrial genomes used in the phylogenetic analyses.

Species	GenBankAC
Delphinus delphis	MF669498.1
Tursiops truncatus	KF570386.1
Tursiops truncatus	MF669486.1
Stenella coeruleoalba	EU557097.1
Stenella longirostris	KX857414.1
Stenella longirostris	KX857394.1
Stenella attenuata	NC012051.1

Introgression and Admixture Analysis

To use the programs of the TreeMix package (v. 1.13) (Pickrell & Pritchard, 2012), a SNP calling was made with Samtools (Li et al., 2009) with mpileup option, with -C50 to reduce the effect of reads with excessive mismatches, skipping alignments with MAPQ smaller than 20, only counting reads with mapping quality greater than 20. A vcf file was created with vcftools, keeping only the variable sites. For the above programs we used 3,720 blocks of 10,000 SNPs (k= 10,000) to standard errors estimation (using block-jackknifing). TreeMix was used to estimate the phylogenetic relationship between the species including up to six migration events. Parameters for this run were: -k= 10,000, -m 6, –noss, –Outgroup and –global, allowing a maximum of six migrations. The clade *S. coeruleoalba* plus *T. Truncatus* was used as outgroup. The *f4*-statistics was calculated with the FOURPOP program to estimate the admixture proportion of one genome to a mixed one (Harris & DeGiorgio, 2016). Admixture graphs, including ancestry proportion for specific hybridization events were calculated in admixturegraph v.1.0.2 R package using as input the f4 results estimated above. Genetic structure among the individuals was estimated using *NGSadmix* (Skotte et al., 2013) with 307,200 SNPs. We estimated the genetic structure considering k=3 to 5 genetic components.

Acknowledgements

This work was supported by grants from the National Council for Scientific and Technological Development (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) to SLB and a Scholarship Grant to AK. We would like to thank the Projeto de Monitoramento de Cetáceos (PMC) for sample collection and to Giovanna Oliveira invaluable help at the laboratory.

References

Amaral AR, Jackson JA, Möller LM, Beheregaray LB, Coelho MM. 2012. Species tree of a recent radiation: The subfamily Delphininae (Cetacea, Mammalia). Molecular Phylogenetics and Evolution. 64(1): 243–253.

Amaral AR, Lovewell G, Coelho MM, Amato G, Rosenbaum HC. 2014. Hybrid speciation in a marine mammal: the Clymene dolphin (*Stenella clymene*). PLoS ONE. 9: 1 – 8.

Amaral KB, Amaral AR, Fordyce RE, Moreno IB. 2018. Historical biogeography of Delphininae dolphins and related taxa (Artiodactyla: Delphinidae). Journal of Mammalian Evolution. 25(2): 241–259.

Antoniou A, Frantzis A, Alexiadou P, Paschou N, Poulakakis N. 2018. Evidence of introgressive hybridization between Stenella coeruleoalba and Delphinus delphis in the Greek Seas. Molecular Phylogenetics and Evolution. 129: 235-337.

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

Banguera-Hinestroza E, Hayano A, Crespo E, Hoelzel AR. 2014. Delphinid systematics and biogeography with a focus on the current genus Lagenorhynchus: Multiple pathways for antitropical and trans-oceanic radiation. Molecular Phylogenetics and Evolution. 80: 217-230

Bianucci G. 2013. *Septidelphis morii*, n. gen. et sp., from the Pliocene of Italy: new evidence of the explosive radiation of true dolphins (Odontoceti, Delphinidae). Journal of Vertebrate Paleontology. 33:722–740.

Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom MPK, Huynh S, Jones G, Knowles LL, Lamichhaney S, Marcussen T, Morlon H, Nakhleh LK,Oxelman B, Pfeil B, Schliep A, Wahlberg N, Werneck FP, Wiedenhoeft J, Willows-Munro S, Edwards SV. 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ 7:e6399

Broad Institute. “Picard Tools.” Broad Institute, GitHub repository. ed. http://broadinstitute.github.io/picard/.

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

Chen Z, Xu S, Zhow K, Yang G. 2011. Whale phylogeny and rapid radiation events revealed using novel retroposed elements and their flanking sequences. BCM Evolutionary Biology. 11:314.

Crossman CA, Taylor EB, Barrett-Lennard LG. 2016. Hybridization in the Cetacea: widespread occurrence and associated morphological, behavioral, and ecological factors. Ecology and Evolution. 6: 1293 – 1303.

Cunha HA, Moraes LC, Medeiros BV, et al. 2011. Phylogenetic status and timescale for the diversification of Steno and Sotalia dolphins. PLoS One. 6(12):e28297.

de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J. 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science. 354: 477 – 481.

Degnan JH, Rosenberg NA. 2006. Discordance of species trees with their most likely gene trees. PLoS Genetics. 2: 762-768.

Flouri T, Jiao X, Rannala B, Yang Z. 2018. Species Tree Inference with BPP using Genomic Sequences and the Multispecies Coalescent. Molecular Biology and Evolution. 35(10):2585-2593.

Good JM, Hird S, Reid N, Demboski JR, Steppan SJ, Martin-Nims TR, Sullivan J. 2008. Ancient hybridization and mitochondrial capture between two species of chipmunks. Molecular Ecolology. 17:1313–1327.

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

Harrison RG and Larson EL. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Molecular Ecology. 25: 2454-2466.

Harrison RG. 1990. Hybrid zones: windows on evolutionary process. Oxford surveys in evolutionary biology. 7: 69–128.

Kingston SE, Adams LD, Rosel PE. 2009. Testing mitochondrial sequences and anonymous nuclear markers for phylogeny reconstruction in a rapidly radiating group: molecular systematics of the Delphininae (Cetacea: Odontoceti: Delphinidae). BMC Evolutionary Biology. 9: 245.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformativs.15: 353.

Kumar S, Stecher G, Li M, Knyaz C, and Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution. 35:1547-1549.

Larsen PA, Marchán-Rivadeneira MR, Baker RJ.2010. Natural hybridization generates mammalian lineage with species characteristics. PNAS.107(25): 11447–11452.

Leduc RG, Perrin WF, Dizon AE. 1999. Phylogenetic relationships among the Delphinid cetaceans based on full cytochrome b sequences. Marine Mammal Science. 15: 619-648.

Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Research. 26: 1 – 11.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25(14): 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N Marth G, Abecasis G, Durbin R, 1000 Genome Project. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16): 2078-2079.

Llopart A, Herrig D, Brud E, Stecklein Z. 2014. Sequential adaptive introgression of the mitochondrial genome in Drosophila yakuba and Drosophila santomea. Molecular ecology. 23(5): 1124-36.

Mallet J. 2007. Hybrid speciation. Nature. 446: 279–283.

McCoy RC, Wakefield J, Akey JM. 2017. Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. Cell. 168: 916-927.

McGowen MR, Spaulding M, Gatesy J. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Molecular Phylogenetics and Evolution. 53(3): 891-906.

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 MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 20(9): 1297 – 1303.

Meier JI, Marques DA, Wagner CE, Excoffier OS. 2018. Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. Molecular biology and evolution. 35: 1489-1506.

Nevado B, Kobmüler C, Sturmbauer J, Snoeks J, Usano-Alemany J, Verheyen E. 2009. Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. Molecular Ecology. 18: 4240-4255

Nilsson MA, Zheng Y, Kumar V, Phillips MJ, Janke A. 2017. Speciation Generates Mosaic Genomes in Kangaroos. Genome Biol Evol. 10(1): 33-44.

Ogilvie HA, Bouckaert RR, Drummond AJ. 2017. StarBEAST2 Brings Faster Species Tree Inference and Accurate Estimates of substitution Rates. Molecular Biology and Evolution. 34: 2101–2114,

Peters KJ, Kleindorfer S. 2015. Divergent foraging behavior in a hybrid zone: Darwin's tree finches (*Camarhynchus* spp.) on Floreana Island. Current Zoology. 61: 181 – 190.

Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. 8(11): e1002967

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics*. 26(6): 841–842.

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

Rambaut A, Suchard MA, Xie D & Drummond AJ. 2014. Tracer v1.6. ed: http://beast.bio.ed.ac.uk/Tracer

Rambaut A. 2017. FigTree. 1.4.3. ed. http://tree.bio.ed.ac.uk/software/figtree/.

Ross HA, Lento GM, Dalebout ML, Goode M, Ewing G, McLaren P, Rodrigo AG, Lavery S, Baker CS. 2003. DNA Surveillance: Web-based molecular identification of whales, dolphins, and porpoises. *Journal of Heredity*. 94:111–114.

Schubert M, Ermini L, Sarkissian CD, Jónsson H, Ginolhac A, Schaefer R, Martin MD, Fernández R, Kircher M, McCue M, Willerslev E, and Orlando L. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nature Protocols. 9: 1056-1084.

Schubert, Lindgreen, and Orlando. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes, 12;9(1):88

Seixas FA, Boursot P, Melo-Ferreira J. 2018. The genomic impact of historical hybridization with massive mitochondiral DNA introgression. Genome Biology. 19:91.

Shaak SG and Counterman BA. 2017. High warning colour polymorphism in *Heliconius* hybrid zone roosts. Ecological Entomology. 42: 315-324.

Shurtliff QR. 2013. Mammalian hybrid zones: a review. Mammal review. 43: 1-21.

Skotte L, Korneliussen ST, Albrechtsen A. 2013. Estimating Individual Admixture Proportions from Next Generation Sequencing Data. Genetics. 195: 693-702

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

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ & Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10.Virus Evolution. 4(1):vey016.

Tobler M, Carson EW. 2010. Environmental variation, hybridization, and phenotypic diversification in Cuatro Cienegas pupfishes. Journal of Evolutionary Biology. 23: 1475–1489.

Toews DP and Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology. 21:3907-3930.

Trigo TC, Schneider A, de Oliveira TG, Lehugeur LM, Silveira L, Freitas TR, Eizirik E. 2013. Molecular data reveal complex hybridization and a cryptic species of neotropical wild cat. Current Biology. 23:2528-2533.

Trigo TC, Tirelli FP, Machado LF, Peters FB, Indrusiak CB et al. 2013. Geographic distribution and food habits of *Leopardus tigrinus* and *L. geoffroyi* (Carnivora, Felidae) at their geographic contact zone in southern Brazil. Studies on Neotropical Fauna and Environment. 48: 56 – 67.

Vallejo-Marín M, Hiscock S. 2016. Hybridization and hybrid speciation under global change. New Phytologist. 211: 1170 – 1187.

Vilstrup JT, Ho SYW, Foote AD, Morin PA, Kreb D, Krützen M, Parra GJ, Robertson KM, Stephanis R, et al. 2011. Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. BCM Evolutionary Biology. 11:65.

vonHoldt BM, Cahill JA, Fan Z, Gronau I, Robinson J, Pollinger JP, Shapiro B, Wall J and Wayn RK. 2016. Whole-genome sequence analysis shows that two

endemic species of North American wolf are admixtures of the coyote and gray wolf. 2(7): e1501714.

Willis PM, Crespi BJ, Dill LM, Baird RW, Hanson MB. 2004. Natural hybridization between Dall's porpoises (*Phocoenoidesdalli*) and harbour porpoises (*Phocoenaphocoena*). Canadian Journal of Zoology.82: 838 – 834.

Zhang, Chao, Rabiee M, Sayyari E, Mirarab S. 2018. “ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees.” BMC Bioinformatics. 19(S6): 153.

Zieliński P , Nadachowska-Brzyska K , Wielstra B , Szkotak R, Covaciu-Marcov SD, Cogălniceanu D, Babik W. 2013. No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). Molecular Ecology. 22: 1884-1903.

Zurano P., Magalhães FM, Asato AE, Silva G, Bidau J, Mesquita DO, Costa GC. 2019. Cetartiodactyla: Updating a time-calibrated molecular phylogeny. Molecular phylogenetics and Evolution. 133:256–262.

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