UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

MACQUARIE UNIVERSITY

INSIGHTS INTO THE EVOLUTIONARY RELATIONSHIPS WITHIN THE DELPHININAE, WITH FOCUS ON THE GENUS *DELPHINUS*: A MULTI-LOCUS, MULTI-DISCIPLINARY APPROACH

Ana Rita Gabirro da Silva Teixeira do Amaral

DOUTORAMENTO EM BIOLOGIA
(Biologia Evolutiva)
2011

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

MACQUARIE UNIVERSITY

INSIGHTS INTO THE EVOLUTIONARY RELATIONSHIPS WITHIN THE DELPHININAE, WITH FOCUS ON THE GENUS *DELPHINUS*: A MULTI-LOCUS, MULTI-DISCIPLINARY APPROACH

Ana Rita Gabirro da Silva Teixeira do Amaral

Tese orientada por:
Professora Doutora M. Manuela Coelho
Doutora Luciana Möller

DOUTORAMENTO EM BIOLOGIA
(Biologia Evolutiva)
2011

This study was supported by Fundação para a Ciência e Tecnologia (FCT), through a PhD grant (SFR/BD/27245/2006) co-financed by the Science and Innovation Operational Programme 2010 (POCI 2010) and by the Ministry of Science, Technology and Higher Education, attributed to Ana Rita Amaral.

This dissertation should be cited as:

Amaral AR (2011) *Insights into the Evolutionary Relationships Within the Delphininae, with Focus on the Genus* Delphinus*: A Multi-Locus, Multi-Disciplinary Approach.* PhD Thesis, University of Lisbon, Portugal.

NOTA PRÉVIA

Na elaboração da presente dissertação, e nos termos do Nº1 do Artigo 41, do Regulamento de Estudos Pós-Graduados da Universidade de Lisboa, publicado no *Diário da República* nº209, II Série de 30 de Outubro de 2006, foram usados integralmente artigos científicos publicados, em publicação ou submetidos para publicação, em revistas internacionais indexadas. Tendo os trabalhos referidos sido efectuados em colaboração, a autora da dissertação esclarece que participou integralmente na concepção e execução do trabalho experimental, na análise, interpretação e discussão dos resultados, e na redacção de todos os manuscritos.

Em memória da minha avó Gabi

TABLE OF CONTENTS

AGRADECIMENTOS I ACKNOWLEDGEMENTS

Esta caminhada não teria sido possível sem a ajuda de muitas pessoas, por isso desde já um muito obrigado a todos aqueles que me acompanharam e me ajudaram a chegar aqui. Foi difícil, mas foi a concretização de um sonho! Gostaria de agradecer em particular:

À Professora Doutora Manuela Coelho, minha co-orientadora, pelo apoio incondicional que sempre me demonstrou e sem o qual este trabalho não teria sido possível. Muito obrigada por me ter deixado seguir o meu caminho e, por sempre que precisei, me ter guiado na direcção certa.

À Doutora Luciana Möller, minha co-orientadora, por ter aceite orientar este trabalho, pelo incentivo, pelas ideias, por tudo. Obrigada por me ter recebido nas minhas visitas a Sydney e a Adelaide e por me ter proporcionado experiências que contribuíram muito para a minha aprendizagem sobre cetáceos.

Ao Doutor Luciano Beheregaray, meu co-orientador, por ter aceite orientar este trabalho, por me receber no seu laboratório de ecologia molecular na Macquarie University, em Sydney, e depois na Flinders University, em Adelaide. Obrigada pelo espírito crítico com que sempre comentou o meu trabalho, o que me fez aprender muito e ensinou a pensar para além do mundo dos golfinhos.

To Professor Lesley Hughes, for all the support given during the last stages of my thesis.

À Doutora Mónica Silva por me ter ensinado as técnicas laboratoriais necessárias ao desenvolvimento de marcadores moleculares anónimos e por ter disponibilizado os meios necessários para eu o fazer; por sempre se ter mostrado disponível para longas conversas científicas e pela partilha incessante de conhecimento.

To Dr. William Perrin, for his interest in my research and for helpful discussions during the progress of my thesis.

To Dr. Phil Morin, for receiving me in the Southwest Fisheries Science Center and for having the time to discuss the results of my research and provide me with useful comments.

To Kelly Robertson, for her endless effort in providing me with samples that were essential for this study, and for having received me during my visit to the Southwest Fisheries Science Center.

To Dr. Jennifer Jackson, for her valuable contribution to the phylogenetic analyses, and for all the time spent explaining me the statistical issues of species tree estimation methods.

Ao Dmitri Boutov, pela sua enorme ajuda e contribuição com os dados oceanográficos.

To all those who contributed with biological samples. I am specially indebted to: Jennifer Learmonth (SAC-Scottish Agricultural College and DEFRA), Luca Mirimin (University of Cork, Ireland), Marisa Ferreira (Sociedade Portuguesa de Vida Selvagem), S. Siciliano and J.F. Moura (Marine Mammal Group Study, GEMM-Lagos and FIOCRUZ, Rio de Janeiro, Brazil), M.C. Oliveira (Instituto Oceanográfico da Universidade de São Paulo), Robert Baldwin, Gianna Minton and Tim Collins (Envrionment Society of Oman), Luís Freitas (Museu da Baleia, Madeira), Karen Stockin (Massey University, New Zealand).

A todos os meus colegas de laboratório na Faculdade de Ciências da Universidade de Lisboa, que muito me ajudaram ao longo dos anos e com quem partilhei momentos de diversão inesquecíveis: Maria Ana Aboim, Ângela Inácio, Carina Cunha, Cristina Luís, Isa Matos, Irene Pala, Mónica Silva, Cristiane Silveira, Silke Waap. Um especial obrigado à Maria Ana, à Mónica, à Irene, à Carina e à Cristiane por me terem ensinado muito do que sei hoje.

To all my colleagues at the Molecular Ecology Lab, Macquarie University, Sydney, Catherine Attard, Kerstin Bilgmann, Georgina Cooke, Shannon Corrigan, Leanne Faulks, Jonathan Sandoval-Castillo, Peter Teske, Jo Wiszniewski, for showing me around and for helping me in everything I needed.

To my office mates at Macquarie University, Catherine Attard, Jonathan Davis, Mattias Roth, Daisy Taylor, Emma Thompson, Emma Wilkie, for all the fun.

To all the staff at the Department of Biological Sciences, Macquarie University, but especially to Marie Howitt for all her help with bureaucracy, for the patience and for caring.

To my colleagues at Flinders University, Adelaide, Jonathan Sandoval-Castillo, Katharina Peters, Jimena Rodriguez, Minami Sasaki, for all the help, advice, friendship and fun.

To all the friends I made in Australia, whom I will certainly miss a lot, but specially to Julia Cooke, who made my life so much easier during my first visit to Sydney; to Kerstin Bilgmann, for all our long conversations; to Jo Wiszniewski for the fun trips and for opening the door of her apartmement to us during the last visit to Sydney; to Jimena and Yuma for all the support given during my stay in Adelaide; to Kathy for all the wonderful time we spent together in Adelaide.

To Karen Stockin, for all her support throughout my thesis and for the friendship.

Às minhas queridas amigas, pelo apoio incondicional, pelo carinho, pelas risadas, pelos almoços e jantares, pelos momentos que garantiram o meu equilíbrio emocional: Maria Ana Aboim, Cristina Brito, Inês Carvalho, Joana Marques, Mónica Silva, Liliana Matias, Rita Vasconcelos, Susana França, Marisa Batista, Nina Vieira.

À Cristina e à Inês, pelos momentos divertidos e inesquecíveis nas nossas muitas viagens, conferências e saídas de mar; muito obrigado por me ajudarem a entrar no mundo dos cetáceos.

À Cristina Brito por estar sempre presente, por me ouvir e por me aconselhar.

À Inês Carvalho, pela amizade, pela força, pelo incentivo.

À minha família, por todo o apoio.

Ao meu avô Fernando, um obrigado especial por tudo.

Aos meus pais e à minha irmã, pelo apoio incondicional, por tornarem os meus sonhos possíveis, por sempre terem acreditado em mim, por suportarem as minhas muitas ausências: se consegui chegar até aqui, foi graças a vocês.

Ao Valter, por estar sempre ao meu lado, por achar que eu sou sempre melhor do que realmente sou, por me aturar nos bons e nos maus momentos, por me seguir até ao outro lado do mundo: sem ti esta caminhada teria certamente sido muito mais difícil.

RESUMO

Desde a antiguidade que os golfinhos sempre cativaram a nossa atenção mas no entanto, vários aspectos da sua biologia, ecologia, estrutura populacional e história evolutiva, permanecem desconhecidos. Os golfinhos pertencem à Família Delphinidae, que surgiu no fim do Miocénico (11-12 milhões de anos atrás). Esta família contém cerca de 35 espécies, desde animais de pequeno porte e com distribuição restrita como o golfinho de Hector (*Cephalorhynchus hectorii*), endémico da Nova Zelândia, até animais de grande porte como a orca (*Orcinus orca*), com distribuição cosmopolita. As relações evolutivas dentro da Família Delphinidae são incertas, nomeadamente as que envolvem as diferentes subfamílias. A subfamília Delphininae, que compreende os géneros *Tursiops, Stenella, Sousa, Lagenodelphis* e *Delphinus* é um desses casos, com uma confusa e controversa taxonomia. Isto deve-se, em grande parte, ao desacordo entre a taxonomia que foi originalmente estabelecida com base em caracteres morfológicos, e as relações filogenéticas subsequentemente suportadas por estudos moleculares.

O principal objectivo desta dissertação foi o de contribuir para o esclarecimento das relações evolutivas dentro da subfamília Delphininae e, em particular, das espécies do género *Delphinus*, utilizando para tal uma abordagem multidisciplinar e multi-locus. Foram objectivos mais específicos: 1) o desenvolvimento de novos marcadores moleculares para o estudo genético de cetáceos; 2) o estudo dos padrões de evolução de proteínas ditas reprodutoras de forma a esclarecer se estas proteínas evoluem rapidamente como descrito em outros mamíferos, e também desempenham algum papel significativo no estabelecimento de barreiras reprodutoras, nomeadamente nas espécies evolutivamente mais próximas dentro da subfamília Delphininae; 3) estimar uma "árvore de espécie" da subfamília Delphininae de forma a esclarecer as relações filogenéticas e taxonomia; 4) avaliar a influência das oscilações climáticas durante o período Pleistocénico na história evolutiva do género *Delphinus*, nomeadamente no que se refere à sua demografia, distribuição geográfica e especiação, e avaliando ainda o efeito de

um conjunto de variáveis oceanográficas na subdivisão populacional do golfinho-comum de bico curto (*D. delphis*), em diferentes escalas espaciais.

Com o objectivo de ampliar o leque de marcadores moleculares disponíveis para o estudo deste grupo de cetáceos, foram desenvolvidos 17 marcadores moleculares anónimos, a partir de uma biblioteca genómica criada para o golfinho-comum, *Delphinus delphis*. Estes *loci* anónimos, que são marcadores espalhados pelo genoma, têm como vantagem, em relação a outros marcadores nucleares, o facto de fornecerem uma medida de diversidade genética mais abrangente. Foram testados com sucesso em várias espécies de cetáceos o que fortemente sugere que podem contribuir para estudos genéticos nestas espécies que requeiram uma abordagem multi-locus.

O estudo de genes comprovadamente ligados ao estabelecimento de isolamento reprodutor certamente contribui para a melhor compreensão dos processos de especiação, e, consequentemente para o estudo da história evolutiva de um dado grupo de espécies. As "proteínas reprodutoras" são proteínas que estão envolvidas no processo de fertilização, mediando a interacção entre o oócito e o espermatozóide. Pensa-se que estas proteínas possam, por isso, desempenhar um papel crucial no estabelecimento de isolamento reprodutor. A sua rápida evolução tem sido descrita em várias espécies de animais. Tem também sido sugerido que em sistemas reprodutores promíscuos, onde existe competição espermática, em princípio acentuada, que esta selecção sexual seja a responsável por essa rápida evolução. Nos cetáceos, onde vários sistemas reprodutores têm sido descritos como promíscuos, ao contrário do que se esperaria, o estudo do padrão de evolução de duas "proteínas reprodutoras", ZP3 e PKDREJ, revelou fraco sinal de selecção positiva. Isto pode explicar-se pelas baixas taxas de evolução encontradas neste tipo de proteínas nestes animais, em comparação com as de outros mamíferos, o que pode ser uma consequência directa da lenta evolução do genoma nuclear dos cetáceos, já anteriormente descrita. É também possível, no entanto, que exista um qualquer relaxamento no mecanismo de reconhecimento específico, ou que este seja comportamental e portanto não inteiramente molecular. Esse relaxamento explicaria os casos de

hibridação que têm sido descritos nas espécies de golfinhos estudadas.

Com o objectivo de esclarecer as relações evolutivas entre os membros da subfamília Delphininae, foi estimada uma "árvore de espécie" com base em métodos de coalescência e utilizando vários marcadores nucleares, incluindo 10 dos marcadores anónimos acima referidos. As relações evolutivas estabelecidas nesta árvore são concordantes com as relações evolutivas estabelecidas com base na morfologia destas espécies. Esta concordância entre caracteres moleculares e morfológicos nunca tinha sido obtida em estudos anteriores. Embora a monofilia do género *Tursiops* e a relação próxima entre *Stenella frotalis* e *S. attenuata* tivessem já sido reportadas em estudos anteriores, nunca foram obtidas em conjunto na mesma árvore filogenética, partilhando o mesmo ancestral comum. A parafilia do género *Stenella* foi confirmada, indicando que a taxonomia do género necessita de uma revisão, uma vez que parece ser um conjunto "artificial", não filogenético, de espécies. Nesta análise foram também observadas incongruências entre marcadores mitocondriais e nucleares, o que se poderá dever a um processo de "incomplete lineage sorting" ou à existência de hibridação entre algumas espécies. Foi possível, no entanto, mostrar que recorrendo a novos métodos filogenéticos e utilizando vários marcadores moleculares, é possível obter-se uma "árvore de espécie" congruente com as relações estabelecidas com base na morfologia.

O género *Delphinus* inclui actualmente duas espécies, subdivididas em quatro subespécies de golfinho-comum. Estas apresentam uma grande variabilidade morfológica em toda a sua área de distribuição, o que tem sido a causa de alguma incerteza no seu arranjo taxonómico, deixando em aberto várias questões relacionadas com a sua filogeografia, estrutura populacional e história evolutiva. Através da utilização de um marcador mitocondrial e vários marcadores nucleares foi possível mostrar que as flutuações de temperatura nos oceanos provocadas pelas oscilações climáticas do Pleistocénico e consequentes alterações nas intensidades dos sistemas de afloramento costeiro e disponibilidade de recursos, influenciaram a demografia, dispersão e diferenciação das espécies do género *Delphinus*. Tendo em conta os resultados

obtidos, foi proposto um cenário evolutivo que sugere que o golfinho-comum de bico curto teria tido origem no oceano Pacífico durante o período Pleistocénico e, posteriormente, se teria dispersado para os Oceanos Índico e Atlântico. Este padrão filogeográfico tinha sido já proposto como estando na base da evolução das formas do género *Stenella* e foi também proposto para vários grupos de organismos marinhos, como peixes, tartarugas e aves marinhas. Foi ainda sugerido que a origem do morfotipo do golfinho-comum de bico comprido teria ocorrido mais tarde, no Pleistocénico, na região do Nordeste Pacífico. Esta origem estaria associada à exploração de novos *habitats* costeiros, resultante de uma diminuição dos recursos disponíveis que se seguiu ao fim de uma época de intenso afloramento costeiro. Eventos de diferenciação independentes teriam ocorrido mais tarde no oceano Atlântico originando o morfotipo de bico comprido que aí ocorre.

Com o objectivo de identificar os padrões da estrutura populacional no golfinho-comum de bico curto, e quais os factores ambientais responsáveis por essa estrutura, foram utilizados 14 microssatélites e dados oceanográficos obtidos a partir de satélite. Foi obtida elevada diferenciação populacional entre as populações amostradas nos oceanos Atlântico, Pacífico e Índico, que parece ser explicada por um processo de isolamento pela distância. No entanto, a diferenciação populacional obtida dentro de cada oceano parece ser explicada por diferenças nas variáveis oceanográficas testadas, nomeadamente da produtividade primária e da temperatura da água. Uma vez que é pouco provável que estas variáveis afectem directamente os golfinhos, foi sugerido que seria o comportamento das suas presas, mais frequentemente afectadas por estes factores, que explicaria aquele padrão. A distribuição do golfinho-comum é em parte, determinada pelas distribuições das suas presas, o que explicaria a sua preferência por determinadas condições oceanográficas.

As variáveis ambientais parecem assim ter desempenhado um papel importante não só na origem, expansão e diferenciação do golfinho-comum durante o Pleistocénico, mas também, mais recentemente, em estabelecer e manter a estrutura populacional das populações de golfinho-comum de bico curto. Embora

se soubesse que a produtividade primária e as temperaturas marinhas influenciam a ocupação dos diversos habitats e os movimentos migratórios das várias espécies de cetáceos, a influência directa destas variáveis na história evolutiva e na estrutura populacional de uma dada espécie de cetáceo, com distribuição cosmopolita, nunca tinha antes sido reportada. Este estudo vem ainda salientar a importância de se juntarem dados ambientais a dados genéticos quando se pretende desenhar áreas marinhas protegidas e também proteger, pela sua importância trófica, os predadores de topo da cadeia alimentar marinha.

Palavras-chave: Delphininae; *Delphinus*; golfinho-comum; evolução; "árvore de espécie"; filogeografia.

ABSTRACT

Evolutionary relationships within the subfamily Delphininae remain contentious mainly due to the disagreement found between taxonomy originally established by morphological characters and phylogenetic relationships subsequently supported by molecular studies. This dissertation is a contribution to the clarification of the evolutionary history of this subfamily, and in particular, of the genus *Delphinus*, using a multi-locus, multi-disciplinary approach. The study of the pattern of evolution in two reproductive proteins across several cetacean species revealed a weak signal of positive selection. This is likely a consequence of their slower rate of evolution when compared to other mammals, which may have consequences to the mechanism of species recognition and explain the hybridization cases reported in several cetaceans. A species tree for the Delphininae was obtained with coalescent-based methods and 13 nuclear DNA loci. This tree agrees with morphology-based species relationships, highlighting the importance of using methods that account for gene tree heterogeneity in obtaining a better estimate of evolutionary relationships. Pleistocene climatic oscillations were shown to have possibly influenced the demography, dispersal and speciation of *Delphinus*. A scenario for the origin of the short-beaked morphotype in the Pacific Ocean during early Pleistocene, with subsequent dispersal into the Atlantic, through the Indian Ocean was proposed. The long-beaked morphotype originated later, possibly by exploring new coastal habitats, in the Northeast Pacific and independent events originated the long-beaked morphotype occurring in the Atlantic Ocean. Through a seascape genetics approach, marine productivity and temperature were found to likely play a role in driving and maintaining population divergence in common dolphins. A direct influence of oceanographic variables on the evolutionary history and population structure of a widely distributed cetacean species had never been reported. These findings highlight the importance of considering such variables in the assessment of distribution and connectivity of top marine predators, particularly in a scenario of ongoing climate change.

Key Words: Delphininae; *Delphinus*; common dolphin; evolution; species tree; phylogeography; seascape genetics.

Chapter I
General Introduction

Dolphins have always captivated people's attention and imagination since ancient times. However, many aspects of their biology, ecology, population structure, and in particular of their evolutionary history remain poorly understood.

Dolphins form part of the family Delphinidae, which likely arose in the late Miocene, 11-12 million years ago (Mya), concomitantly with the two other extant families in the superfamily Delphinoidea: Phocoenidae and Monodontidae (Barnes *et al.* 1985). Delphinidae is the most speciose family of marine mammals, comprising over 35 different species, distributed in all oceans and most seas of the world (LeDuc 2009). This family is also extremely diverse, and includes smaller dolphins, such as the New Zealand endemic Hector's dolphin (*Cephalorhynchus hectorii*), and the much larger and widely distributed killer whale (*Orcinus orca*).

Global past climatic changes, ocean currents, oceanic food chains, primary productivity and tectonically-driven vicariant events are thought to have driven the rapid radiation of species in this family (Nikaido *et al.* 2001; Steeman *et al.* 2009). Evolutionary relationships among species of delphinids, and in particular those involving the different subfamilies, are presently uncertain. Much of this uncertainty is due to the recent use of molecular tools, which nonetheless brought new insights into relationships that had initially been established based on morphological characters alone (e.g. LeDuc *et al.* 1999). Two classifications, one representing a more traditional view of dolphin taxonomy, based on morphology, and a revised classification based on recent molecular analyses are presented in Table 1.1.

1.1. THE SUBFAMILY DELPHININAE

The subfamily Delphininae comprises the genera *Tursiops, Stenella, Sousa, Lagenodelphis* and *Delphinus.* The delphinines are also involved in taxonomic uncertainty. The evolutionary relationships within the subfamily have been largely unresolved partly because of the apparent paraphyly of *Tursiops* and *Stenella*, but

also due to the uncertainty in the inclusion of *Sousa* (LeDuc *et al.* 1999). The latter issue appears to have been resolved by recent molecular phylogenetic analyses that placed *Sousa* sp. within the delphinines (Agnarsson & May-Collado 2008; Caballero *et al.* 2008; May-Collado & Agnarsson 2006; McGowen *et al.* 2009; Steeman *et al.* 2009).

The paraphyly of *Tursiops* and *Stenella* has, nevertheless, remained contentious. The genus *Tursiops* was monotypic, including *T. truncatus* (Montagu 1821), until full species status for the Indo-Pacific bottlenose dolphin, *T. aduncus* (Ehrenberg 1832) was recognized in recent years (Ross 1977). However, a phylogenetic study of Delphinidae using the mitochondrial cytochrome *b* gene uncovered the paraphyly of the genus, with *T. aduncus* being more closely related to *S. frontalis* than to *T. truncatus* (LeDuc *et al.* 1999). The authors of such study proposed that these could actually represent sister taxa because of the morphological similarities recognized to exist between *T. aduncus* and *S. frontalis* (Perrin *et al.* 1987). Nevertheless, a more recent phylogenetic study, using a supertree approach and both mitochondrial and nuclear DNA loci, uncovered a weak signal for the monophyly of *Tursiops*, indicating the need for a thorough revision of this genus (McGowen *et al.* 2009). In addition, it has been suggested that another two species may exist in the genus, one in coastal waters of southern Australia (Möller *et al.* 2008) and another in South Africa (Natoli *et al.* 2004).

The genus *Stenella* currently comprises five species (Table 1.1). Using different molecular markers and methods, every molecular phylogenetic study conducted so far has uncovered the paraphyly of the genus, with some members more closely related to *Delphinus*, *Lagenodelphis* or *Tursiops,* than to nominal congeners (Agnarsson & May-Collado 2008; Caballero *et al.* 2008; Kingston *et al.* 2009; LeDuc *et al.* 1999; May-Collado & Agnarsson 2006; McGowen *et al.* 2009; Steeman *et al.* 2009). In fact, morphological studies using cranial characters and pigmentation patterns have shown clear similarities between *Delphinus*, *S. coeruleoalba*, *S. clymene*, *S. longirostris* and *Lagenodelphis* and between *S. attenuata* and *S. frontalis* (Perrin 1997; Perrin *et al.* 1981), indicating that *Stenella* may indeed be an artificial assemblage of species (LeDuc *et al.* 1999).

Table 1.1. Two classifications of the Family Delphinidae.

a) Traditional dolphin taxonomy. From LeDuc 2009.	b) Revised classification based on molecular analyses. From LeDuc *et al*. 1999.
Family Delphinidae	Family Delphinidae
Subfamily Stenoninae	Subfamily Stenoninae
Steno bredanensis	*Steno bredanensis*
Sousa chinensis	*Sotalia fluviatilis*
S. teuszii	*S. guianensis*
Sotalia fluviatilis	Subfamily Delphininae
Subfamily Delphininae	*Sousa chinensis*
Lagenorhynchus albirostris	*Stenella clymene*
L. acutus	*S. coeruleoalba*
L. obscurus	*S. frontalis*
L. obliquidens	*S. attenuata*
L. cruciger	*S. longirostris*
L. australis	*Delphinus delphis*
Grampus griseus	*D. capensis*
Tursiops truncatus	*Tursiops truncatus*
Stenella frontalis	*T. aduncus*
S. attenuata	*Lagenodelphis hosei*
S. longirostris	Subfamily Lissodelphininae
S. clymene	*Lissodelphis borealis*
S. coeruleoalba	*L. peronii*
Delphinus delphis	*Cephalorhynchus heavisidii*
D. capensis	*C. hectori*
Lagenodelphis hosei	*C. eutropia*
Subfamily Lissodelphininae	*C. commersonii*
Lissodelphis borealis	*Sagmatias obscurus*
L. peronii	*S. obliquidens*
Subfamily Cephalorhynchinae	*S. cruciger*
Cephalorhynchus commersonii	*S. australis*
C. eutropia	Subfamily Globicephalinae
C. heavisidii	*Feresa attenuata*
C. hectori	*Peponocephala electra*
Subfamily Globicephalinae	*Globicephala melas*
Peponocephala electra	*G. macrorhynchus*
Feresa attenuata	*Pseudorca crassidens*
Pseudorca crassidens	*Grampus griseus*
Orcinus orca	Subfamily Orcininae
Globicephala melas	*Orcinus orca*
G. macrorhynchus	*Orcaella brevirostris*
Subfamily Orcaellinae	*O. heinsohni*
Orcaella brevirostris	*Insertae sedis*
	Lagenorhynchus albirostris
	Leucopleurus acutus

Lagenodelphis is a monotypic genus with its single species being morphologically closely related to *Delphinus*. Individuals of both genera possess deep palatal grooves, which is a unique characteristic amongst the Delphinidae, but they are

also related to *S. longirostris*, *S. coeruleoalba* and *S. clymene* in other morphological characteristics (Dolar 2009). Phylogenetically, its position is uncertain with studies based on mitochondrial DNA not having enough resolving power (LeDuc *et al.* 1999; May-Collado & Agnarsson 2006) and studies based on nuclear DNA placing it closer to *S. longirostris* (McGowen *et al.* 2009; Steeman *et al*. 2009). A genetic study of coastal Indo-Pacific bottlenose dolphins from southern Australia based on mitochondrial DNA and microsatellites has uncovered a sister taxa relationship with *Lagenodelphis* (Möller *et al.* 2008).

In summary, the systematic and phylogenetic relationships within this subfamily remain contentious to date. There is growing necessity to clarify these matters, since knowledge of evolutionary relationships is essential for the establishment of species boundaries and consequent implementation of management and conservation policies. Several methods for estimating species trees based on individual gene trees obtained from different nuclear DNA loci have been recently developed (Belfiore *et al.* 2008; Brumfield *et al.* 2008; Cranston *et al.* 2009; Liu *et al.* 2008). These methods incorporate the stochastic sorting of lineages in the estimation of species trees and have proven useful to clarify relationships in species complexes that show rapid diversification and confusing taxonomy (e.g. Belfiore *et al.* 2008; Dolman & Hugall 2008). Such an approach could therefore provide a better insight into the species tree and phylogenetic relationships of the subfamily Delphininae, and ultimately in clarifying its taxonomy.

1.2. THE GENUS *DELPHINUS*

1.2.1. *Distribution and taxonomy*

The dolphins in this genus are known as common dolphins. They are widely distributed, occurring in warm-temperate waters of the Atlantic, Pacific and Indian Oceans and in many enclosed and semi-enclosed basins such as the Mediterranean Sea, the Black Sea, the Caribbean Sea, Persian Gulf, Gulf of Thailand, Sea of Japan and Gulf of California (Heyning & Perrin 1994; Jefferson *et al.* 2008; Perrin 2009). The genus presently comprises two species and four subspecies: the short-beaked common dolphin, *D. delphis delphis* Linnaeus,

1758, distributed in continental shelves and pelagic waters of the Atlantic and Pacific Oceans; the long-beaked common dolphin, *D. capensis capensis* Gray, 1828, distributed in nearshore tropical and temperate waters of the Pacific and South Atlantic Oceans; the Black-Sea common dolphin, *D. d. ponticus* Barabash, 1935; and the Indo-Pacific common dolphin, *D. c. tropicalis* van Bree, 1971 (Perrin 2009). This classification was mainly based on morphological and skeletal characters such as coloration, overall body size, length of the rostrum and tooth counts (Heyning & Perrin 1994; Jefferson & Van Waerebeek 2002).

Initial descriptions of the short and long-beaked forms referred to dolphins occurring off coastal Californian waters, in the Northeast Pacific, and to *D. delphis* and *D. bairdii* Dall, 1873 (the nominal species from California), respectively (Heyning & Perrin 1994). After reviewing existing literature and a number of other nominal species these authors concluded that based on the relative size and proportions of the rostrum, *D. bairdii* appeared to be a junior synonym of *D. capensis* Gray, 1828 from South Africa. They followed to conclude that the short-beaked and long-beaked morphotypes occurred around the world as two species, *D. delphis* and *D. capensis.* Nevertheless, the extensive geographic variation in external appearance of common dolphins has lead to some confusion in defining the areas where the morphotypes occur (e.g. Jefferson *et al.* 2009). In fact, recent morphological studies of common dolphins inhabiting regions of the North Atlantic and Southwest Pacific have shown populations with measures of relative rostrum length and tooth counts not matching those of the short and long-beaked forms initially described from the Northeast Pacific, raising taxonomic issues (Bell *et al.* 2002; Murphy *et al.* 2006; Westgate 2007). Although the separation of the two species on the coast of California was confirmed by a genetic study based on the mitochondrial DNA control region (Rosel *et al.* 1994), genetic studies including additional molecular markers and populations from other regions have found disagreement between morphological and genetic characters (Amaral *et al.* 2007; Kingston & Rosel 2004; Natoli *et al.* 2006). As a result, the taxonomy of the genus Delphinus remains uncertain.

1.2.2. *Ecology*

The ecology and behaviour of common dolphins is still poorly known. They occur in large groups of dozens to hundreds of individuals, likely composed by smaller subunits of 20-30 individuals (Perrin 2009). It has been suggested that these groupings occur regardless of genetic relationships, but with possible age and gender segregation (Viricel *et al.* 2008).

The short and long-beaked forms are known to occur in sympatry in some areas (Perrin 2009). In the Northeast Pacific, the long-beaked form seems to prefer warmer and more coastal waters, while the short-beaked form is known to migrate according to oceanographic conditions and prefers upwelling-modified waters (Ballance *et al.* 2006). In certain regions, the short-beaked common dolphin has also been associated to particular water masses, generated by different temperature regimes (Möller *et al.* 2011). Their movements seem to be largely determined by those of their potential prey (e.g. Young & Cockcroft 1994), and thus their diets vary according to location and season, although they generally prey on small, mesopelagic schooling fish species, such as sardines and anchovies (Young & Cockcroft 1994, Pusineri *et al.* 2007). In some areas, squids can also be an important component of their diet (Jefferson *et al.* 2008; Perrin 2009). A dietary segregation between short-beaked common dolphins occupying oceanic and neritic waters has been reported for the Bay of Biscay, Northeast Atlantic (Lahaye *et al.* 2005; Pusineri *et al.* 2007).

1.2.3. *Phylogeography*

The only broad phylogeographic study on common dolphins conducted to date showed overall low levels of genetic differentiation among short-beaked populations across oceans, and high levels of genetic differentiation between long-beaked populations inhabiting the Northeast Pacific and southern Africa (Natoli *et al.* 2006). The generalised lack of agreement between morphological and genetic differentiation was attributed to local adaptation, and the high divergence of long-beaked morphotypes was attributed to independent evolution events converging on the same morphotype (Natoli *et al*. 2006). However, this

study did not include short-beaked populations from the Indo-Pacific region, neither the *tropicalis* morphotype that inhabits the Indian Ocean basin, or long-beaked populations that occur in other regions. A global study, using several molecular markers and all described morphotypes would certainly contribute to the clarification of the phylogeography and evolutionary history of the genus.

1.2.4. *Population structure*

High levels of genetic diversity have been reported for common dolphin populations worldwide (Amaral *et al.* 2007; Bilgmann *et al.* 2008; Natoli *et al.* 2006; Rosel *et al.* 1994; Viricel *et al.* 2008). On a broad scale, genetic differentiation in short-beaked populations has been found to be lower across larger geographical scales when comparing populations from the Atlantic and Northeast Pacific Oceans (Natoli *et al.* 2006). Regional genetic studies have been conducted in the Atlantic and Pacific Oceans (Amaral *et al.* 2007; Bilgmann *et al.* 2008; Mirimin *et al.* 2009; Querouil *et al.* 2010; Möller *et al*. 2011). Within the Atlantic Ocean, genetic differentiation in short-beaked common dolphins has only been found between the eastern and western North Atlantic, with no structure found within each region, or around the Azores archipelago (Amaral *et al.* 2007; Mirimin *et al.* 2009; Querouil *et al.* 2010). Within the Pacific Ocean, different patterns have been reported, with fine scale population structure in short-beaked common dolphins found to occur off the West USA coast (Chivers *et al.* 2009), off eastern Australia coast (Möller *et al.* 2011) and around New Zealand (Stockin *et al*. in prep). Particular oceanographic characteristics, such as currents, temperature and salinity, have been suggested to limit the movement of short-beaked common dolphins and promote such patterns of genetic differentiation (Bilgmmann *et al.* 2008; Chivers *et al.* 2009; Möller *et al.* 2011). A global genetic study including short-beaked common dolphin populations from all oceans where it occurs has never been conducted. Direct evaluation of the influence of oceanographic variables on the genetic structure of short-beaked common dolphins is also still to be undertaken. This kind of approach, where environmental variables are used to complement genetic data, has provided invaluable insights into which factors may be driving population divergence in both terrestrial and

marine environments (Manel *et al.* 2003; Selkoe *et al.* 2008). Since the distribution of short-beaked common dolphins seems to coincide with certain oceanographic conditions and prey distribution, such integrative approach would improve our understanding on the mechanisms governing the global patterns of population structure in these species.

1.3. Molecular Markers and Their Application in Cetacean Genetic Studies

The marker of choice for genetic studies focusing on cetacean species has been mitochondrial DNA (mtDNA). It has been used for species and stock identification (e.g. Baker & Dalebout 2009), and also for population structure, phylogeographic and phylogenetic studies (e.g. LeDuc *et al.* 1999; Oremus *et al.* 2009). Some of the properties that have made this an appealing marker are its maternal inheritance, lack of recombination, nearly neutral evolution and a clock-like evolutionary rate (Avise *et al.* 1987; Zink & Barrowclough 2008). Nevertheless, recent research has questioned these characteristics (e.g. Ballard & Whitlock 2004). Moreover, there is an increasing awareness for the errors that can be made when using a single molecule to assess population structure and the evolutionary and demographic past of species and populations (Ballard & Whitlock 2004; Brito & Edwards 2009). Despite these criticisms, mtDNA will certainly continue to be used in phylogeographic and demographic studies as a complement to nuclear data, since it reveals different aspects of the evolutionary history of species or populations and, in some cases, it can be the only informative marker available (for example, when nuclear genes are not informative and/or exhibit paraphyly patterns (Brito & Edwards 2009; Zink & Barrowclough 2008)).

With the development of amplification (through polymerase chain reaction - PCR) and sequencing technologies, and the increase in availability of sequenced genomes in public databases, the development of new nuclear molecular markers for non-model taxa has become more accessible. The study of genes and proteins, and of their evolution and function, has also become feasible in

organisms with no genomic resources. This has driven a transition in genetic studies from single- to multi-locus approaches (Brito & Edwards 2009). Nevertheless, the use of nuclear sequence data in cetaceans has been limited to only a few phylogenetic studies that have used introns (e.g. Baker *et al.* 1998; Caballero *et al.* 2008; Cassens *et al.* 2000; Gaines *et al.* 2005). One of the reasons for this may be the low level of polymorphism in nuclear genes that have been reported for cetaceans, which are likely due to the slower evolution of the cetacean nuclear genome when compared to other mammals (Jackson *et al.* 2009).

A class of nuclear markers that has been seldom used in genetic studies, and apparently never used in cetaceans, are anonymous markers. These are non-coding regions of the genome, randomly collected and presumably dispersed across the chromosomes, thereby representing wide and potentially unbiased variation across the genome (Sunnucks 2000). In addition, they provide the power to obtain multiple, replicate gene histories and have been shown to be more variable than introns (Jennings & Edwards 2005). They are also relatively easy and affordable to develop and likely to be informative across species. Overall, these markers seem promising tools to address several questions of phylogenetics and phylogeography in cetacean species.

1.3.1. *Reproductive proteins and reproductive isolation*

The study of reproductive isolation can be the key to understand the process of speciation (Coyne & Orr 2004). Although morphological characters and ecological aspects can be used as proxies for reproductive isolation, recent research has focused on the genetic basis of postmating prezygotic isolation through the study of reproductive proteins (e.g. Berlin *et al.* 2008; Berlin & Smith 2005; Calkins *et al.* 2007; Clark *et al.* 2006; Galindo *et al.* 2003; Martin-Coello *et al.* 2009; Palumbi 2009; Turner & Hoekstra 2006, 2008). These proteins, which are involved in sperm-egg interactions, have been found to be evolving rapidly in several animal groups (Calkins *et al.* 2007; Metz *et al.* 1998; Swanson *et al.* 2001; Turner & Hoekstra 2006). This rapid evolution is thought to be driven by sperm competition,

cryptic female choice, sexual conflict and avoidance of heterospecific fertilization (Birkhead & Pizzari 2002; Swanson & Vacquier 2002; Turner & Hoekstra 2008).

In mammals, several of these proteins have been studied. Amongst them, ZP3, zona pellucida 3, is one of the best characterized. It is located in the egg coat, contains a region described as the "sperm-combining" region (Chen *et al.* 1998; Wassarman *et al.* 2004) and is generally accepted to be the natural agonist that initiates the acrosome reaction (Wassarman & Litscher 2001). On the male side, one protein that has been studied is PKDREJ, which is a protein localized in the plasma membrane of the acrosomal crescent region of the sperm head (Sutton *et al.* 2006). PKDREJ is a candidate egg-sperm binding protein that controls acrosome exocytosis through the process of capacitation, which represents a time delay between insemination and fertilization (Sutton *et al.* 2008). This protein has been suggested to play an important role in cases of postcopulatory sperm competition (Birkhead & Pizzari 2002; Sutton *et al.* 2008).

The study of reproductive proteins has shown interesting patterns of rapid evolution and divergence between closely related species in rodents and primates, suggesting their role in the establishment of reproductive isolation (Turner & Hoekstra 2006, 2008). Moreover, it is expected that the forces of sexual selection driving the rapid evolution of these proteins will be particularly strong in promiscuous mating systems (Birkhead & Pizzari 2002).

In cetaceans, several mating systems have been reported to be promiscuous and thus likely characterized by sperm competition and sexual conflict (Berta & Sumich 1999; Connor *et al.* 2000). Female promiscuity has been documented both in whales (minke whales (Skaug *et al.* 2008), humpback whales (Clapham & Palsboll 1997), gray whales (Swartz *et al.* 2006), right whales (Frasier *et al.* 2007)) and in dolphins (Connor *et al.* 2000; Murphy *et al.* 2005). However, the extent to which this promiscuity is related to reproductive protein evolution or to the establishment of reproductive isolation has never been assessed. Nevertheless, with the increased availability of genomic resources it is now possible to develop primers and isolate the genes of interest to address such questions. The study of ZP3 and PKDREJ in cetaceans would further allow testing

whether the model of sexual conflict, as an evolutionary force driving the rapid evolution of these proteins in mammals, also applies in this group.

1.4. AIMS

The general aim of this thesis is to contribute to a better understanding of the evolutionary history of the subfamily Delphininae, by targeting key unresolved issues, with particular focus on the genus *Delphinus*, and by using a multi-locus, multi-disciplinary approach. Specific aims include:

1) the development of new nuclear molecular markers for genetic studies of cetacean species;

2) the study of the patterns of evolution of reproductive proteins in cetaceans in order to understand if their role in the sperm-binding process and establishment of reproductive isolation can be confirmed in the closely related members of the subfamily Delphininae;

3) the estimation of a multi-locus species tree for the subfamily Delphininae in order to elucidate phylogenetic relationships and shed light into the taxonomy of this group;

4) the assessment of the evolutionary history of the genus *Delphinus* by: (i) investigating the influence of Pleistocene climatic oscillations on population demography, geographic distribution and speciation of common dolphins; and (ii) testing the effect of a set of oceanographic variables on the contemporary population subdivision of short-beaked common dolphins at different spatial scales (across oceans and within ocean basins).

1.5. THESIS STRUCTURE

The thesis comprehends seven chapters. Chapter I corresponds to the general introduction. Chapters II, III, IV, V and VI correspond to five specific investigations that were conducted to address the specific aims and that resulted in scientific articles. Chapter VII corresponds to the general discussion and final remarks.

The first aim of the thesis is addressed in Chapter II with article 1. This article reports the development and characterization of a set of anonymous nuclear markers for cetacean species, which can be used for population structure, phylogeographic or phylogenetic studies. Article 1 is published in *Conservation Genetics*. (Amaral AR, Silva MC, Möller LM, Beheregaray LB, Coelho MM (2010) Anonymous nuclear markers for cetacean species. *Conservation Genetics* 11, 1143-1146).

The second aim of the thesis is addressed in Chapter III. Article 2 investigates the evolution of two reproductive proteins, ZP3 and PKDREJ, in cetaceans. Evidence of positive selection acting on these proteins is tested. In addition, the rates of evolution of ZP3 and PKDREJ in cetaceans, primates and rodents are compared with those of two non-reproductive proteins (MC1R and BMP4) to assess whether reproductive proteins are evolving faster in cetaceans. Article 2 is published in *Journal of Heredity* (Amaral AR, Möller LM, Beheregaray LB, Coelho MM (2011) Evolution of 2 reproductive proteins, ZP3 and PKDREJ, in cetaceans. *Journal of Heredity* 102, 275-282).

Chapter IV addresses the third aim of the thesis. In Article 3 a species tree of the subfamily Delphininae is estimated using 13 nuclear loci sequences (including 10 of the anonymous markers developed in Article 1). The performance of different species tree estimation methods is compared and the phylogenetic relationships among members of the subfamily are discussed. Article 3 is currently under review in *Molecular Phylogenetics and Evolution*.

The last aim of the thesis is addressed in Chapters V and VI. In Chapter V, Article 4 assesses the influence of Pleistocene climatic oscillations in the phylogeography and demographic history of the genus *Delphinus* using one mitochondrial DNA gene and five nuclear loci. Populations of short-beaked and long-beaked common dolphins from the Atlantic, Pacific and Indian Oceans are analysed, and insights into the geographical origin and dispersal of the genus are discussed in light of marine biogeographic models and paleoceanography. Article 4 is currently under review in *Molecular Ecology*.

In Chapter VI, Article 5 analyses the worldwide population genetic structure of short-beaked common dolphins using 14 microsatellite loci and examines the influence that a set of oceanographic variables (sea surface temperature, chlorophyll concentration and water turbidity) have in the establishment of population divergence under a seascape genetics approach. Article 5 has been submitted to *Plos One*.

Chapter VII corresponds to the general discussion where the main findings of articles 1 to 5 are summarized and integrated in a final conclusion. Particular focus is given to articles 4 and 5 where insights are given into the evolutionary history and taxonomy of common dolphins.

1.6. REFERENCES

Agnarsson I, May-Collado LJ (2008) The phylogeny of Cetartiodactyla: The importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome *b* to provide reliable species-level phylogenies. *Molecular Phylogenetics and Evolution* **48**, 964-985.

Amaral AR, Sequeira M, Cedeira-Martínez J, Coelho MM (2007) New insights on population genetic structure of *Delphinus delphis* from the Northeast Atlantic and phylogenetic relationships within the genus inferred from two mitochondrial markers. *Marine Biology* **151**, 1967-1976.

Avise JC, Arnold J, Ball RM*,* Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography - The mitochondrial-DNA bridge between population-genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Baker CS, Dalebout ML (2009) Forensic Genetics. In: *Encyclopedia of Marine Mammals*, pp. 453-459. Academic Press, New York.

Baker CS, Medrano-Gonzalez L, Calambokidis J*,* Perry A, Pichler F, Rosenbaum H, Straley JM, Urban-Ramirez J, Yamaguchi M, Von Ziegesar O (1998) Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. *Molecular Ecology* **7**, 695-707.

Ballance LT, Pitman RL, Fiedler PC (2006) Oceanographic influences on seabirds and cetaceans of the Eastern Tropical Pacific: A review. *Progress in Oceanography* **69**, 360-390.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.

Barnes LG, Domning DP, Ray CE (1985) Status of studies on fossil marine mammals. *Marine Mammal Science* **1**, 15-53.

Belfiore NM, Liu L, Moritz C (2008) Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia : Geomyidae). *Systematic Biology* **57**, 294-310.

Bell CH, Kemper CM, Conran JG (2002) Common dolphins *Delphinus delphis* in southern Australia: A morphometric study. *Australian Mammalogy* **24**, 1-10.

Berlin S, Qu LJ, Ellegren H (2008) Adaptive evolution of gamete-recognition proteins in birds. *Journal of Molecular Evolution* **67**, 488-496.

Berlin S, Smith NGC (2005) Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. *Bmc Evolutionary Biology* **5**, 65.

Berta A, Sumich JL (1999) *Marine Mammals: Evolutionary Biology*. Academic Press, San Diego.

Bilgmann K, Möller LM, Harcourt RG, Gales R, Beheregaray LB (2008) Common dolphins subject to fisheries impacts in southern Australia are genetically differentiated: Implications for conservation. *Animal Conservation* **11**, 518-528.

Birkhead TR, Pizzari T (2002) Postcopulatory sexual selection. *Nature Reviews Genetics* **3**, 262-273.

Brito P, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-455.

Brumfield RT, Liu L, Lum DE, Edwards SV (2008) Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae, *Manacus*) from multilocus sequence data. *Systematic Biology* **57**, 719-731.

Caballero S, Jackson J, Mignucci-Giannoni AA, Barrios-Garrido H, Beltran-

Pedreros S, Montiel-Villalobos MG, Robertson KM, Baker CS (2008) Molecular systematics of South American dolphins *Sotalia*: Sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. *Molecular Phylogenetics and Evolution* **46**, 252-268.

Calkins JD, El-Hinn D, Swanson WJ (2007) Adaptive evolution in an avian reproductive protein: ZP3. *Journal of Molecular Evolution* **65**, 555-563.

Cassens I, Vicario S, Waddell VG*,* Balchowsky H, Van Belle D, Ding W, Fan C, Mohan RSL, Simo EC, Bastida R, Meyer A, Stanhope MJ, Milinkovitch MC (2000) Independent adaptation to riverine habitats allowed survival of ancient cetacean lineages. *Proceedings of the National Academy of Sciences of the United States of America.* **97**, 11343-11347.

Chen J, Litscher ES, Wassarman PM (1998) Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6193-6197.

Chivers SJ, Hedrick NM, LeDuc CA (2009) Genetic Evidence for Population structure Structure in eastern North Pacific *Delphinus delphis*. International Whaling Commission SC/61/For Info 57.

Clapham PJ, Palsboll PJ (1997) Molecular analysis of paternity shows promiscuous mating in female humpback whales (*Megaptera novaengliae*, Borowski). *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 95-98.

Clark NL, Aagaard JE, Swanson WJ (2006) Evolution of reproductive proteins from animals and plants. *Reproduction* **131**, 11-22.

Connor RC, Read AJ, Wranhgam R (2000) Male Reproductive Strategies and Social Bonds. In: *Cetacean Societies: Field Studies of Dolphin and Whales* (eds. Mann J, Connor RC, Tyack PL, Whitehead H). The Chicago University Press, Chicago.

Coyne JA, Orr HA (2004) *Speciation.* Sinauer Associates, Sunderland, Massachusetts.

Cranston KA, Hurwitz B, Ware D, Stein L, Wing RA (2009) Species trees from highly incongruent gene trees in Rice. *Systematic Biology* **58**, 489-500.

Dolar MLL (2009) Fraser's Dolphin *Lagenodelphis hosei*. In: *Encyclopdia of Marine Mammals* (eds. Perrin WF, Wursig B, Thewissen JGM), pp. 469-472. Academic Press, New York.

Dolman G, Hugall AF (2008) Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). *Molecular Phylogenetics and Evolution* **49**, 782-794.

Frasier TR, Hamilton PK, Brown MW*,* Conger LA, Knowlton AR, Maryx MK, Slay CK, Kraus SD, White BN (2007) Patterns of male reproductive success in a highly promiscuous whale species: The endangered North Atlantic right whale. *Molecular Ecology* **16**, 5277-5293.

Gaines CA, Hare MP, Beck SE, Rosenbaum HC (2005) Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. *Proceedings of the Royal Society B-Biological Sciences* **272**, 533-542.

Galindo BE, Vacquier VD, Swanson WJ (2003) Positive selection in the egg receptor for abalone sperm lysin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4639-4643.

Heyning JE, Perrin WF (1994) Evidence for two species of common dolphins (genus *Delphinus*) from the eastern North Pacific. *Contributions in Science* **442**, 1-35.

Jackson JA, Baker CS, Vant M*,* Steel DJ, Medrano-Gonzalez L, Palumbi SR (2009) Big and slow: Phylogenetic estimates of molecular evolution in baleen whales (Suborder Mysticeti). *Molecular Biology and Evolution* **26**, 2427-2440.

Jefferson TA, Fertl D, Bolanos-Jimenez J, Zerbini AN (2009) Distribution of common dolphins (*Delphinus* spp.) in the western Atlantic Ocean: A critical re-examination. *Marine Biology* **156**, 1109-1124.

Jefferson TA, Van Waerebeek K (2002) The taxonomic status of the nominal dolphin species *Delphinus tropicalis* van bree, 1971. *Marine Mammal Science* **18**,

787-818.

Jefferson TA, Webber MA, Pitman RL (2008) *Marine Mammals of the World: A Comprehensive Guide to Their Identification.* Academic Press, San Diego.

Jennings WB, Edwards SV (2005) Speciational history of Australian grass finches (*Poephila*) inferred from thirty gene trees. *Evolution* **59**, 2033-2047.

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.

Kingston SE, Rosel PE (2004) Genetic differentiation among recently diverged delphinid taxa determined using AFLP markers. *Journal of Heredity* **95**, 1-10.

Lahaye V, Bustamante P, Spitz J*,* Dabin W, Das K, Pierce GJ, Caurant F (2005) Long-term dietary segregation of common dolphins *Delphinus delphis* in the Bay of Biscay, determined using cadmium as an ecological tracer. *Marine Ecology-Progress Series* **305**, 275-285.

LeDuc RG (2009) Delphinids, Overview. In: *Encyclopedia of Marine Mammals* (eds. Perrin WF, Wursig B, Theweissen JGM), pp. 298-302. Academic Press, New York.

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.

Liu L, Pearl DK, Brumfield RT, Edwards SV (2008) Estimating species trees using multiple-allele DNA sequence data. *Evolution* **62**, 2080-2091.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Martin-Coello J, Benavent-Corai J, Roldan ERS, Gomendio M (2009) Sperm competition promotes asymmetries in reproductive barriers between closely related species. *Evolution* **63**, 613-623.

May-Collado L, Agnarsson I (2006) Cytochrome b and Bayesian inference of whale phylogeny. *Molecular Phylogenetics and Evolution* **38**, 344-354.

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

Metz EC, Gomez-Gutierrez G, Vacquier VD (1998) Mitochondrial DNA and bindin gene sequence evolution among allopatric species of the sea urchin genus Arbacia. *Molecular Biology and Evolution* **15**, 185-195.

Mirimin L, Westgate A, Rogan E*,* Rosel P, Read A, Coughlan J, Cross T (2009) Population structure of short-beaked common dolphins (*Delphinus delphis*) in the North Atlantic Ocean as revealed by mitochondrial and nuclear genetic markers. *Marine Biology* **156**, 821-834.

Möller LM, Kerstin B, Charlton-Robb K, Beheregaray L (2008) Multi-gene evidence for a new bottlenose dolphin species in southern Australia. *Molecular Phylogenetics and Evolution* **49**, 674-681.

Möller LM, Valdez FP, Allen S*,* Bilgmann K, Corrigan S, Beheregaray LB (2011) Fine-scale genetic structure in short-beaked common dolphins (*Delphinus delphis*) along the East Australian Current. *Marine Biology* **158**, 113-126.

Murphy S, Collet A, Rogan E (2005) Mating strategy in the male common dolphin (*Delphinus delphis*): What gonadal analysis tells us. *Journal of Mammalogy* **86**, 1247-1258.

Murphy S, Herman JS, Pierce GJ, Rogan E, Kitchener AC (2006) Taxonomic status and geographical cranial variation of common dolphins (*Delphinus*) in the eastern North Atlantic. *Marine Mammal Science* **22**, 573-599.

Natoli A, Cañadas A, Peddemors VM*,* Aguilar A, Vaquero C, Fernandez-Piqueras P, Hoelzel AR (2006) Phylogeography and alpha taxonomy of the common dolphin (*Delphinus* sp.). *Journal of Evolutionary Biology* **19**, 943-954.

Natoli A, Peddemors VM, Hoelzel AR (2004) Population structure and speciation in the genus *Tursiops* based on microsatellite and mitochondrial DNA analyses. *Journal of Evolutionary Biology* **17**, 363-375.

Nikaido M, Matsuno F, Hamilton H, Brownell RL, Cao Y, Ding W, Zuoyan Z, Shedlock AM, Fordyce RE, Hasegawa M, Okada N (2001) Retroposon analysis of major cetacean lineages: The monophyly of toothed whales and the paraphyly of river dolphins. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7384-7389.

Oremus M, Gales R, Dalebout ML, Funahashi N, Endo T, Kage T, Steel D, Baker CS (2009) Worldwide mitochondrial DNA diversity and phylogeography of pilot whales (*Globicephala* spp.). *Biological Journal of the Linnean Society* **98**, 729-744.

Palumbi SR (2009) Speciation and the evolution of gamete recognition genes: Pattern and process. *Heredity* **102**, 66-76.

Perrin WF (1997) Development and homologies of head stripes in the delphinoid cetaceans. *Marine Mammal Science* **13**, 1-43.

Perrin WF (2009) Common dolphins, *Delphinus delphis* and *D. capensis*. In: *Encyclopedia of Marine Mammals* (eds. Perrin WF, Wursig B, Thewissen JGM), pp. 255-259. Academic Press, New York.

Perrin WF, Mitchell ED, Mead JG, Caldwell DK, Caldwell MC, Vanbree PJH, Dawbin WH (1987) Revision of the spotted dolphins, *Stenella* Spp. *Marine Mammal Science* **3**, 99-170.

Perrin WF, Mitchell ED, Mead JG, Caldwell DK, Vanbree PJH (1981) *Stenella clymene*, a rediscovered tropical dolphin of the Atlantic. *Journal of Mammalogy* **62**, 583-598.

Pusineri C, Magnin V, Meynier L, Spitz J, Hassani S, Ridoux V (2007) Food and feeding ecology of the common dolphin (*Delphinus delphis*) in the oceanic Northeast Atlantic and comparison with its diet in neritic areas. *Marine Mammal Science* **23**, 30-47.

Querouil S, Freitas L, Cascao I, Alves F, Dinis A, Almeida JR, Prieto R, Borras S, Matos JA, Mendonca D, Santos RS (2010) Molecular insight into the population structure of common and spotted dolphins inhabiting the pelagic waters of the Northeast Atlantic. *Marine Biology* **157**, 2567-2580.

Rosel PE, Dizon AE, Heyning JE (1994) Genetic analysis of sympatric morphotypes of common dolphins (genus *Delphinus*). *Marine Biology* **119**, 159-168.

Ross GJB (1977) The taxonomy of bottlenose dolphins *Tursiops* species in South African waters, with notes on their biology. *Annals of the Cape Provincial Museum* **11**, 135-194.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries* **9**, 363-377.

Skaug HJ, Bérubé M, Rew MB, Palsboll PJ (2008) Genetic analysis reveal promiscuous mating in female common minke whales, *Balaenoptera acutorostrata*. *Journal of Cetacean Research and Management* **9**, 249-251.

Steeman ME, Hebsgaard MB, Fordyce RE, Ho SYW, Rabosky DL, Nielsen R, Rahbek C, Glenner H, Sorensen MV, Willerslev E (2009) Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology* **58**, 573-585.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**, 199-203.

Sutton KA, Jungnickel MK, Florman HM (2008) A polycystin-1 controls postcopulatory reproductive selection in mice. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8661-8666.

Sutton KA, Jungnickel MK, Ward CJ, Harris PC, Florman HM (2006) Functional characterization of PKDREJ, a male germ cell-restricted polycystin. *Journal of Cellular Physiology* **209**, 493-500.

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**, 137-144.

Swanson WJ, Zhang ZH, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2509-2514.

Swartz SL, Taylor BL, Rugh DJ (2006) Gray whale *Eschrichtius robustus*

population and stock identity. *Mammal Review* **36**, 66-84.

Turner LM, Hoekstra HE (2006) Adaptive evolution of fertilization proteins within a genus: Variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Molecular Biology and Evolution* **23**, 1656-1669.

Turner LM, Hoekstra HE (2008) Causes and consequences of the evolution of reproductive proteins. *International Journal of Developmental Biology* **52**, 769-780.

Viricel A, Strand AE, Rosel PE, Ridoux V, Garcia P (2008) Insights on common dolphin (*Delphinus delphis*) social organization from genetic analysis of a mass-stranded pod. *Behavioral Ecology and Sociobiology* **63**, 173-185.

Wassarman PA, Jovine L, Litscher ES, Qi HY, Williams Z (2004) Egg-sperm interactions at fertilization in mammals. *European Journal of Obstetrics Gynecology and Reproductive Biology* **115**, S57-S60.

Wassarman PM, Litscher ES (2001) Towards the molecular basis of sperm and egg interaction during mammalian fertilization. *Cells Tissues Organs* **168**, 36-45.

Westgate AJ (2007) Geographic variation in cranial morphology of short-beaked common dolphins (*Delphinus delphis*) from the North Atlantic. *Journal of Mammalogy* **88**, 678-688.

Young DD, Cockcroft VG (1994) Diet of common dolphins (*Delphinus delphis*) off the South-east coast of southern Africa - Opportunism or specialization. *Journal of Zoology* **234**, 41-53.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107-2121.

Chapter II (Article 1)
Anonymous Nuclear Markers for Cetacean Species

Ana R. Amaral[1,2], Mónica C. Silva[1], Luciana M. Möller[2,3], Luciano B. Beheregaray[2,3], M. Manuela Coelho[1]

[1] Universidade de Lisboa, Faculdade de Ciências, Centro de Biologia Ambiental, Campo Grande, 1749-016 Lisbon, Portugal
[2] Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
[3] School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia

Conservation Genetics (2010) 11, 1143-1146

Anonymous nuclear markers for cetacean species

2.1. ABSTRACT

Here we report the development and characterization of 17 anonymous nuclear markers for cetacean species. These markers were isolated from a genomic library built from a common dolphin (genus *Delphinus*), and tested across several families within Cetacea. An average of 1 SNP per 272 bp was found in 10 anonymous markers screened for polymorphism within the genus *Delphinus* (total of 6537 bp sequenced). These markers represent a significant addition to the set of tools used in genetic studies of cetaceans where population and species boundaries have to be inferred in order to implement proper conservation strategies.

2.2. INTRODUCTION

Mitochondrial DNA (mtDNA) and microsatellites are amongst the most common classes of markers in ecological and conservation genetic studies (Beheregaray 2008). However, some caveats exist when mtDNA and microsatellite data are compared. The mtDNA genome evolves as a single unit, which yields a single gene tree, no matter how many base pairs or genes are sequenced. This warrants the need to use multiple nuclear sequence loci in studies where parameters such as effective population sizes and coalescent times are to be estimated (Ballard & Whitlock 2004). Microsatellites, on the other hand, are very popular because of their high variability and power to resolve population structure. However, the mutation models associated with these markers are not well understood in some cases and are not comparable to the mutation model of single nucleotide substitutions per nucleotide of mtDNA genes (Takezaki & Nei 1996). Additionally, using microsatellite frequencies to infer the phylogeny as part of a phylogeographic study involves making a number of heroic assumptions involving clustering individuals into populations for analysis, and rooting the resulting distance trees (Zink & Barrowclough 2008).

Recently, with an increasing number of genome sequencing projects underway for model organisms, single nucleotide polymorphisms (SNPs) have become markers of choice for a number of studies due to several advantages compared to microsatellites (Brumfield *et al*. 2003; Morin *et al*. 2004). In non-model organisms, SNPs have been traditionally obtained following two different approaches: a targeted gene approach in which primers can be designed from conserved regions of aligned sequences of at least two species in order to amplify a less conserved region [CATS, (Lyons *et al.* 1999)]; and an alternative approach where a genomic library is constructed with randomly sheared DNA and loci obtained by cloning sequencing (Rosenblum *et al.* 2007). This latter approach is more appealing for non-model organisms where hardly any sequence information exists.

The Family Delphinidae is the largest and most diverse family within the Order Cetacea, with currently 37 species (Caballero *et al*. 2008). It is a group with a complex evolutionary history, with many phylogenetic relationships still unresolved. Several species and/or populations are facing threats such as pollution, by-catch, food depletion and global warming, which warrant the need to develop new molecular markers that can help to understand and define biological boundaries, so that proper conservation strategies can be designed and implemented. Here we describe the development and characterization of multiple anonymous nuclear markers that have the potential to be used for phylogenetic, phylogeographic and population genetic structure studies of several cetacean species.

2.3. MATERIALS AND METHODS

We built a genomic library for an individual short-beaked common dolphin (*Delphinus delphis*), which stranded in the Portuguese coast, using the TOPO® Shotgun Subcloning Kit (Invitrogen). Genomic DNA was extracted following standard phenol-chloroform procedures and was then sheared with a nebulizer. DNA fragments, 2 to 4 kb in size, were then blunted with T4 DNA and Klenow polymerases, dephospholylated with calf intestinal phosphatase, ligated into a

vector (pCR®4Blunt-TOPO®) and then transformed into *Escherichia coli* competent cells.

We sequenced 30 random clone inserts with vector primers, used Sequencher (v.4.2 Gene Codes Corporation) to visualize sequences and performed a BLAST search in GenBank to confirm their suitability for population genetic studies by ruling out the possibility they encode proteins or other conserved regions. Most loci remained anonymous with the exception of one that exhibited high percentage match to a known gene in other mammals. We also used the RepeatMasker program to screen sequences for interspersed repeats (http://www.repeatmasker.org). In several loci, repetitive elements were found, namely short interspersed elements (SINEs) and long interspersed elements (LINEs). Roughly half of the higher eukaryotic genome is composed of a variety of repetitive sequences with no obvious function (Ray 2007), so this finding was not surprising.

We designed primers for 17 clones using the program Primer 3 (Rozen & Skaletsky 2000), which were then tested using a polymerase chain reaction (PCR) gradient thermocycle (MyCycler, Biorad) with annealing temperatures ranging from 55ºC to 64ºC. Primer sequences, annealing temperatures and approximate fragment length can be found in Table 2.1. Standard PCR conditions were applied to all reactions: 25-μL reactions containing 10-100 ng DNA, 0.2 mM each dNTP, 0.3 μM each primer, 1 U *Taq* Polymerase and 1 X *Taq* buffer. Thermocycler profiles included 5 min at 94ºC, followed by 35 cycles of 30 s at 94ºC, 30 s at specified annealing temperature, 45 s at 72ºC and a final extension step of 10 min at 72ºC.

Amplification was tested in a panel of 41 individuals from several delphinid cetacean species: *D. delphis*, *D. capensis, Stenella coeruleoalba, S. frontalis, S. attenuata, S. longirostris, Tursiops truncates, T. aduncus, Lagenodelphis hosei, Sousa chinensis, Sotalia fluviatilis* and *Globicephala melas*. We also tested amplification in representatives of other cetacean families: *Phocoena phocoena* (Phocoenidae), *Kogia breviceps* (Kogidae) and *Balaenoptera acutorostrata* (Balaenopteridae), to determine how primers would perform across families.

When specific annealing temperatures failed to amplify in more distantly related species, a touchdown PCR profile was used [5 min at 94ºC, followed by 10 cycles of 30 s at 94ºC, 30 s at 64-60ºC (decrease of 0.5ºC per cycle) and 45 s at 72ºC, followed by 35 cycles of 30 s at 94ºC, 30 s at 60ºC and 45 s at 72ºC; and final extension step of 10 min at 72ºC]. We obtained high levels of cross-amplification success, with almost all loci being amplified in all species tested. Ten loci were randomly chosen to screen for a panel of individuals from the genus *Delphinus*. This panel consisted of 12 *D. delphis* (4 from NE Atlantic, 2 from NW Atlantic, 2 from West Pacific and 4 from Eastern Tropical Pacific), 4 *D. capensis* (from the Eastern Tropical Pacific) and 1 *D. capensis tropicalis* (from the Arabian Sea). These loci were directly sequenced in both directions (BigDye Terminator CycleSequencing: Applied Biosystems) on an ABI 3730xl automated sequencer (Applied Biosystems). Sequencher (v.4.2, Gene Codes Corporation) was used to visualize sequences. Gametic phase was resolved computationally using PHASE v.2.1.1 (Stephens & Donnelly 2003).

2.4. RESULTS AND DISCUSSION

A total of 24 variable sites, or SNPs, were found in 6537 bp sequenced with an average of 1 SNP/272 bp (Table 2.2). This level of nucleotide variation is higher to the one described for sperm whales [average of 1 SNP/540 bp (Morin et al. 2007)], but it falls within the range of 1 SNP in every 200-500 bp given by (Morin *et al.* 2004) for terrestrial mammals. However it is lower than the levels described for birds [1 SNP/175bp, (Primmer *et al*. 2002)] and for reptiles [1 SNP/30bp, (Rosenblum *et al*. 2007)]. Polymorphism was not uniformly distributed across all loci, with some loci showing no polymorphism (Table 2.2). This pattern may be due to differences in nucleotide composition and genomic location, which may influence substitution rates among loci. The two loci showing higher levels of polymorphism also had the higher percentage of CG content (Table 2.2). Genomic regions rich in CpG islands are highly susceptible to mutation through methylation, where the cytosine mutates to a thymine, as opposed to genomic regions rich in A/T, which are less prone to mutations (Han *et al*. 2008).

Table 2.1. Primer sequences (5'-3') for 17 anonymous nuclear loci and PCR annealing temperatures (ºC) across several cetacean species. Del. - Delphinidae; *P.pho* - *Phocoena phocoena*; *K.bre* - *Kogia breviceps*; *B.acu* - *Balaenoptera acutorostrata*.

Locus ID	Accession Numbers	Primer Sequence (5'-3')	PCR annealing temperature			
			Del.	*P.pho*	*K.bre*	*B.acu*
Del_01	FJ490557	GAGCCTCACTTGGAAACTGG GCTGGTGGATAGGCAAAATG	64	64	64	64
Del_02	FJ490558	TGACTCCATGCCTCCTCTCT AGACGGTGAGGCCAATTTTT	64	64	64	64
Del_03	FJ490559	TAGGGAGTGAGGGAGCTCAG TCTTCACCAACCCTTCCAGT	64	64	64	64
Del_04	FJ490560	GCTGTATAACAAATGACCCCAGT CAGATCACATCTGGGGGAAC	64-60	64-60	n/a	n/a
Del_05	FJ490561	TACAGAAAGCCCATGTGCAG CGGTGGCATTTCTAAAAGGA	60	60	n/a	60
Del_06	FJ490562	TAAAGCCCCAGAGATTTGGA CGAATTCGCCCTTCACTTA	64	64	64	64
Del_07	FJ490563	TCGCAGCTGCTGTTTGTTAG TGGCTTGGTAGTTCAGAGACC	64	64	64	64
Del_08	FJ490564	TAGCTCTTGAGCGAATGCAA TGGACCTAGCCTTGTTAATGC	64-60	64-60	64-60	n/a
Del_09	FJ490565	TTCAAATTGGAAAGGAAGAGG GTGGAATTGGGAGCAATGAT	60	60	60	60
Del_10	FJ490566	CAGATATTGGAACTTCCCTGGT TTTCCAAAAAGCCAGATGGT	60	60	60	n/a
Del_11	FJ490567	CACAAATCTGAGGAACACACAAA TTGTAAAGCCTTATAAATTTCAGGTTA	64-60	64-60	64-60	64-60
Del_12	FJ490568	GGAGGTAGGGACCACACTGA AGAATGATCCGCTCCAAATG	60	60	60	n/a
Del_13	FJ490569	ACAAAATGTCCCACAGCGTA TTAATAGCTTCCGGGGATGG	60	60	60	60
Del_14	FJ490570	TGGGTCCCAGAAGAAGAACA TCTCTTAGCTTTTGCTTGTCTGT	64	64	64	64
Del_15	FJ490571	ACAAAACCTCGTTGGTCCAG GGTTGACAGCTTGCCATGT	64	64	64	64
Del_16	FJ490572	TCTATATAAAATCTGTTGAGTCCCTTT CAGAGCAACAACACATTTAGGG	60	60	60	60
Del_17	FJ490573	TTCTCTGTCTGACTGACTTCACTG CCATCCTGTTAAATGCCTTG	60	60	60	n/a

The minimum number of recombination events within loci (Hudson & Kaplan 1985) was evaluated using DNAsp v.4.5 (Rozas *et al*. 2003), but no recombination was detected.

Table 2.2. Summary statistics for the 10 anonymous nuclear loci screened for variation in 17 *Delphinus* species. *S*, segregating sites, π, nucleotide diversity.

Locus ID	Product size (bp)	*S*	π	GC content (%)
Del_02	829	2	0.00030	31.26
Del_04	636	0	0.0000	47.67
Del_05	723	1	0.00016	41.35
Del_08	768	3	0.00106	28.94
Del_10	401	3	0.00180	35.97
Del_11	571	1	0.00021	32.95
Del_12	736	5	0.00148	53.73
Del_15	356	0	0.0000	36.40
Del_16	782	7	0.00350	56.97
Del_18	735	2	0.00090	46.75
Total	6537	24	0.00101	41.20

The non-random association between polymorphisms at different loci was measured by the degree of linkage disequilibrium (LD), also using DNAsp. After correcting for multiple comparisons, no significant LD was observed among loci, indicating that the 10 loci analysed are not likely to be linked.

The primers presented here, obtained with a random fragment genomic library approach, represent a good alternative for rapid marker development and SNP discovery in cetaceans and will help addressing questions in phylogenetics and population structure in cetaceans.

2.5. ACKNOWLEDGEMENTS

We thank Kelly Robertson (Southwest Fisheries Research Center, NOAA) for providing samples of several delphinid cetacean species. A. R. Amaral is supported by a PhD grant (SFR/BD/27245/2006) and M. C. Silva is supported by a post-doctoral grant (SFRH/BPD/22276/2005), both funded by FCT (Portugal).

2.6. REFERENCES

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.

Beheregaray LB (2008) Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology* **17**, 3754-3774.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* **18**, 249-256.

Caballero S, Jackson J, Mignucci-Giannoni AA, Barrios-Garrido H, Beltran-Pedreros S, Montiel-Villalobos MG, Robertson KM, Baker CS (2008) Molecular systematics of South American dolphins *Sotalia*: Sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. *Molecular Phylogenetics and Evolution* **46**, 252-268.

Han L, Su B, Li WH, Zhao ZM (2008) CpG island density and its correlations with genomic features in mammalian genomes. *Genome Biology* **9**, 12.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. *Genetics* **111**, 147-164.

Lyons LA, Kehler JS, O'Brien SJ (1999) Development of comparative anchor tagged sequences (CATS) for canine genome mapping. *Journal of Heredity* **90**, 15-26.

Morin PA, Aitken NC, Rubio-Cisneros N, Dizon AE, Mesnick S (2007) Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*). *Molecular Ecology Notes* **7**, 626-630.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* **19**, 208-216.

Primmer CR, Borge T, Lindell J, Saetre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: High nucleotide diversity revealed in the avian genome. *Molecular Ecology* **11**, 603-612.

Ray DA (2007) SINEs of progress: Mobile element applications to molecular ecology. *Molecular Ecology* **16**, 19-33.

Rosenblum EB, Belfiore NM, Moritz C (2007) Anonymous nuclear markers for the eastern fence lizard, *Sceloporus undulatus*. *Molecular Ecology Notes* **7**, 113-116.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Rozen S, Skaletsky HJ (2000) Primer3 on WWW for General Users and for Biologist Programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds. Krawetz S, Misener S), pp. 365-386. Humana Press, Totowa, NJ.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73**, 1162-1169.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**, 389-399.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107-2121.

Chapter III (Article 2)
Evolution of Two Reproductive Proteins, ZP3 and PKDREJ, in Cetaceans

---

Ana R. Amaral[1,2], Luciana M. Möller[3], Luciano B. Beheregaray[2,3] and M. Manuela Coelho[1]

[1] Universidade de Lisboa, Faculdade de Ciências, Centro de Biologia Ambiental, Campo Grande, 1749-016 Lisbon, Portugal
[2] Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
[3] School of Biological Sciences, Flinders University, Adelaide, SA, 5001, Australia

Journal of Heredity (2011) 102, 275-282

Evolution of two reproductive proteins, ZP3 and PKDREJ, in cetaceans

3.1. ABSTRACT

The rapid evolution of proteins involved in reproduction has been documented in several animal taxa. This is thought to be the result of forces involved in sexual selection and is expected to be particularly strong in promiscuous mating systems. In this study, a range of cetacean species were used to analyse the patterns of evolution in 2 reproductive proteins involved in fertilization: the zona pellucida 3 (ZP3), present in the egg coat, and PKDREJ, localized in the sperm head. We targeted exons 6 and 7 of ZP3 and a part of the REJ domain in PKDREJ for a total of 958 bp in 18 species. We found very low levels of amino acid sequence divergence in both proteins, a very weak signal of positive selection in ZP3 and no signal in PKDREJ. These results were consistent with previous reports of a slow rate of molecular evolution in cetaceans but unexpected due to the existence of promiscuous mating systems in these species. The results raise questions about the evolution of reproductive isolation and species recognition in whales and dolphins.

Key Words: PKDREJ, ZP3, sexual selection, Cetacea

3.2. INTRODUCTION

Recent research on fertilization proteins, those mediating sperm-egg interactions, has revealed a pattern of rapid adaptive evolution in several animal groups, such as in marine invertebrates, birds and mammals (Calkins *et al.* 2007; Metz *et al.* 1998; Swanson *et al.* 2003; Turner & Hoekstra 2006). This widespread phenomenon may have important consequences, like the establishment of barriers to fertilization that could lead to speciation (Swanson & Vacquier 2002). The selective forces of sperm competition, sexual selection and sexual conflict have been suggested as drivers of the rapid evolution of these proteins (Swanson and Vacquier 2002). In mammals, the initial binding of sperm to the egg coat is thought to be the critical step of sperm-egg recognition (Wassarman & Litscher

2001). The egg coat comprises at least three glycoproteins with zona pellucida (ZP) domains: ZP1, ZP2 and ZP3, the latter being generally accepted to be the natural agonist that initiates the acrosome reaction upon binding of sperm to egg (Wassarman *et al.* 2001). Moreover, ZP3 is one of the best characterized mammalian fertilization proteins, containing a region described as the "sperm-combining" region (Chen *et al.* 1998; Wassarman *et al.* 2004).

PKDREJ is a protein localized in the plasma membrane of the acrosomal crescent region of the sperm head, whose expression has only been detected in the spermatogenic lineage (Butscheid *et al.* 2006). It has been recently shown that this protein controls acrosome exocytosis through the process of capacitation (Sutton *et al.* 2008), which represents a time delay between insemination and fertilization. In species where sperm competition exists as a form of postcopulatory sexual selection, genes that control the duration of capacitation could provide a selective paternal advantage and therefore could be targets of positive selection (Birkhead & Pizzari 2002). PKDREJ is therefore a candidate egg-binding sperm protein with a presumed role in cases of postcopulatory sperm competition (Sutton *et al.* 2008).

In mammals, initial studies of reproductive protein evolution used gene sequences from relatively distant species (Swanson *et al.* 2001). However, it has been suggested that an understanding of how amino acid changes affect fertilization, and consequently reproductive isolation, will only be possible by studying the patterns of evolution of these proteins in closely related species (Turner & Hoekstra 2008). Such studies have only been conducted in rodents for ZP3 (*Mus,* (Jansa *et al.* 2003); *Peromyscus* (Turner & Hoekstra 2006); Australasian rodents, (Swann *et al.* 2007)) and primates for PKDREJ (Hamm *et al.* 2007). Patterns of positive selection were documented for both proteins in all these studies suggesting their key role in the egg-sperm binding process. Although not the aim of these studies, the authors have also found no relation between mating strategies, that is, different levels of sperm competition, in the studied species and the pattern of evolution of these proteins. Nevertheless, investigation of additional

taxa is needed to confirm if this pattern of rapid evolution can be generalized across closely related and recently diverged species.

Here we investigate the evolution of ZP3 and PKDREJ in cetaceans. These proteins were chosen based on their putative role in egg-sperm interaction as mentioned above.

Cetaceans are thought to have diverged from *Hippopotamus* 53 million years ago (Mya) (Arnason *et al.* 2004). Extant species have split into two main groups around 35 Ma: the Mysticeti (baleen whales) and the Odontoceti (toothed whales). The explosive radiation of delphinoids (especially the family Delphinidae) occurred 11-12 Mya, with some dolphin species having originated as recently as 1-3 Mya (Caballero *et al.* 2008; McGowen *et al.* 2009). This group of dolphins, referred to as the STDL species complex, includes the genera *Stenella-Sousa-Tursiops-Delphinus-Lagenodelphis* (Perrin & Reeves 2004) and provides an excellent case to test whether the rapid evolution of reproductive proteins is a phenomenon generalized across different closely related taxa. Several mating systems in both the mysticetes and the odontocetes have been reported to be promiscuous, and thus characterised by sperm competition and sexual conflict (Berta & Sumich 1999). Nevertheless, the different life history patterns of the 2 groups likely resulted in different mating strategies that could have influenced the evolution of reproductive proteins. It has also been reported that at least some cetaceans have a slow rate of molecular evolution (Jackson *et al.* 2009; Martin & Palumbi 1993), potentially limiting the adaptive potential of those genes. Our aim in this study was to test the hypothesis that positive Darwinian selection is acting on female and male reproductive proteins in cetaceans. Such result would lend support to models that propose sexual conflict and sperm competition as selective forces driving the divergence of these proteins and confirming their role in the sperm-binding process. For that, we studied patterns of evolution in 2 reproductive proteins, ZP3 and PKDREJ, and 2 non-reproductive proteins, MC1R and BMP4, in several cetacean species.

3.3. MATERIALS AND METHODS

Genomic DNA was extracted from ethanol-preserved tissue using standard phenol-chloroform extraction. Species used in this study are specified in Supplementary Table S3.1. Comparisons of published DNA sequences from *Bos taurus*, *Sus scrofa*, *Ovis aries* and *Mus musculus* were used to design primers for exons 6 and 7 of ZP3 (ZP3-F1, 5'- CTGCCACCTGAAGGTCACTC-3' and ZP3-R1, 5'-GCGACTTCGGGGAACAGA-3'). These regions were chosen because they contain several sites identified as targets of selection in an analysis of divergent mammalian species, namely the sperm-binding region (Swanson *et al.* 2001). For PKDREJ, we used published primers (Demere *et al.* 2008). Primers for BMP4 exon 3 (BMP4-F3 5'-CCACCTTGTCATACTCATCCAG-3'; BMP4-R3 5' AGAACATCCCAGGGACCAG-3') were designed based on an alignment of published DNA sequences. For MC1R, sequences deposited in GenBank were used (accession numbers FJ773287-FJ773291, FJ773294, FJ773296, FJ773305, FJ773313). Polymerase chain reactions (PCRs) were performed in 25 μL reactions containing 10-100 ng DNA, 0.2 mM each dNTP, 0.3 μM each primer, 1 U *Taq* Polymerase and 1 X *Taq* buffer. For ZP3 and BMP4, the thermocycle profile included one cycle of 95ºC for 2 min, followed by 20 cycles of 95ºC for 30 s and a touchdown from 65ºC to 55ºC for 1 min decreasing by 0.5ºC/cycle, and then 72ºC for 1.50 min. This was followed by 20 cycles of 95ºC for 30 s, 55ºC for 1 min and 72ºC for 1.50 and a final extension step of 72ºC for 10 min. For PKDREJ, the thermocycle profile consisted of an initial denaturation step at 94ºC for 2 min followed by 35 cycles of 94ºC for 30 s, 60ºC for 30 s and 72ºC for 45 s, followed by a final extension step at 72ºC for 7 min. PCR products were separated on 1.0% agarose gels, stained with ethidium bromide and visualized with ultraviolet light. PCR products were cleaned with Exonuclease I and Shrimp Alkaline Phosphatase and directly sequenced in both directions. Sequences were aligned in Sequencher v.4.2, with heterozygous nucleotide sites being coded as ambiguities (position 192 in exons 6 and 7 of ZP3 and positions 153, 219, 407 and 550 in PKDREJ, see Supplementary Figure S3.1). The identity of the sequenced fragments was confirmed by performing a basic alignment search tool (BLAST) of amino acid

sequences obtained for all genes using the BLASTp algorithm (NCBI). Phylogenetic relationships for ZP3 and PKDREJ were constructed using the Maximum Likelihood (ML) method as implemented in PAUP* (v.4b10, Swofford 2003). The best evolutionary model for each gene was determined using the Akaike Information Criterion in Modeltest (v.3.7, Posada & Buckley 2004). Bayesian trees were constructed using MrBayes (Huelsenbeck & Ronquist 2001), and 10,000,000 generations of Monte Carlo Markov Chain (MCMC) were run using the program default priors as starting values for the model. Trees were sampled every 100 generations during the analysis. The first 600,000 generations were excluded as burn-in after examining the variation in log-likelihood scores over time.

Evidence of positive selection in ZP3 and PKDREJ was tested using different maximum-likelihood methods as implemented in CODEML, as part of the PAML package (v.4, Yang 2000). A likelihood ratio test (LRT) was used to examine the data for individual codons with d_N/d_S ratios (ω) significantly > 1. This was done by comparing a null (neutral) model that does not allow $\omega > 1$, with an alternative model that does. The null models included a model with a d_N/d_S class between 0 and 1 and a class with $d_N/d_S = 1$ (M1a) and a model which assumes a beta distribution for d_N/d_S in the interval 0,1 (M7). The alternative models include an additional class of sites with $d_N/d_S > 1$ estimated from the data set (M2a and M8). An additional test comparing results from M8 to a modified version of the model where the selection class has d_N/d_S set to 1 (model M8a) was performed. This test rules out the possibility that the neutral model M7 is rejected due to a poor fit of the beta distribution for neutral and negatively selected sites. The test statistic follows a 50:50 mix of a χ^2 distribution with one *df* and a point mass of zero. If the LRT is significant, positive selection is inferred. A Bayesian analysis was used to calculate the posterior probability that each site is from a particular site class, and sites with high posterior probabilities coming from the class with $d_N/d_S > 1$ ($P >$ 95%) were considered to be under positive selection. This Bayes empirical Bayes approach performs best when the data set is small and lacks information (Yang *et al.* 2005).

We also used other methods that account for variation in both synonymous and nonsynonymous rates; the single-likelihood ancestor counting (SLAC) method, the fixed effects likelihood (FEL) method, and the random effects likelihood (REL) method (see Pond & Frost 2005b). These methods were implemented using the web interface DATAMONKEY (Pond & Frost 2005a). Additionally, we compared synonymous and nonsynonymous substitution rates of ZP3 and PKDREJ in rodents and primates with those of cetaceans in order to assess whether reproductive proteins are evolving slower or faster in these species. Such comparisons were also performed using the non-reproductive proteins, MC1R and BMP4. Sequences were retrieved from available databases (NCBI and Ensembl, Accession Numbers in Supplementary Table S3.2) and truncated to correspond to the region amplified in cetaceans. Pairwise comparisons of d_N and d_S were obtained using the runmode = -2 option in CODEML and mean estimations were then calculated over all species. Overall levels of nucleotide divergence amongst cetaceans were estimated using MEGA 4.0 (Tamura *et al.* 2007).

3.4. RESULTS

A 355-bp fragment of ZP3 was sequenced, including exon 6 (92 bp), intron 6 (136 bp) and exon 7 (127 bp). Translation resulted in a fragment of 72 amino acids in total, corresponding to positions 279-354 of *Mus musculus* (NP_035906, 48% identity), which includes the sperm-combining region (328-343) (Chen *et al.* 1998).

Alignment with *Mus* ZP3 revealed a 3 amino acid deletion. The existence of some conserved regions suggest that some domain structures predicted in *Mus* are likely retained in cetaceans. However, one (Ser-332) of the two serine residues identified to be essential for sperm receptor activity in mouse, rat and human ZP3 (Chen *et al.* 1998) has been lost in cetaceans, while the other (Ser-334) has been retained in *Balaenoptera acutorostrata, B. musculus* (both Balaenopteridae) and *Phocoena phocoena* (Phocoenidae). We found very low levels of amino acid sequence divergence in cetaceans (only 6.9% of sites differed among species) with all dolphin species of the STDL complex having identical amino acid sequences (Figure 3.1). For PKDREJ, a 603-bp fragment was sequenced.

Translation resulted in a fragment of 200 amino acids in total, corresponding to positions 217-419 of *Homo sapiens* (NP_006062.1, 53% identity), which falls in the REJ domain, a region predicted to be functionally important in the sperm-egg recognition process (Sutton *et al.* 2006). As in ZP3, amino acid sequence divergence was very low, with all dolphin species of the STDL complex having identical amino acid sequences (Figure 3.1). Nucleotide divergence was also very low. For BMP4, a 771-bp fragment was sequenced. Translation resulted in a fragment of 257 amino acids in total, corresponding to positions 125-381 of *Homo sapiens* (BAA06410.1).

a)

	292	308	327	334	339
	1	3	**4**	**5**	**5**
	4	0	**8**	**5**	**9**
Consensus	**R**	N	**S**	**Y**	**R**
B. acutorostrata	**.**	S	**H**	**S**	**T**
B. musculus	**.**	.	**R**	**S**	**T**
B. physalus	**.**	.	**.**	**.**	**.**
K. breviceps	**.**	.	**.**	**.**	**.**
P. phocoena	**Q**	.	**.**	**S**	**.**
G. melas	**.**	.	**.**	**.**	**.**
S. fluviatilis	**.**	.	**.**	**.**	**.**
S. chinensis	**.**	.	**.**	**.**	**.**
S. coeruleoalba	**.**	.	**.**	**.**	**.**
S. attenuata	**.**	.	**.**	**.**	**.**
S. frontalis	**.**	.	**.**	**.**	**.**
S. longirostris	**.**	.	**.**	**.**	**.**
L. hosei	**.**	.	**.**	**.**	**.**
D. delphis	**.**	.	**.**	**.**	**.**
D. capensis	**.**	.	**.**	**.**	**.**
D. c. tropicalis	**.**	.	**.**	**.**	**.**
T. truncatus	**.**	.	**.**	**.**	**.**
T. aduncus	**.**	.	**.**	**.**	**.**

b)

	2	3	3	3	3	4	6	6	7	7	7	9	9	1 0	1 **0**	1	1	1	1	1	1	1	1
	9	0	3	5	7	4	4	**7**	**3**	4	**7**	6	9	5	**9**	2	5	9	7	9	6	2	6
Consensus	A	H	N	R	F	Y	F	**E**	**I**	R	**T**	Y	T	D	**I**	I	K	Q	E	G	H	E	R
B. acutorostrata	.	R	I	S	.	F	.	**G**	.	H	**A**	.	K	.	**M**	.	R	E	D	E	.	D	.
B. musculus	.	R	I	S	.	F	.	**G**	.	H	**A**	.	K	.	.	.	R	E	D	E	.	D	.
B. physalus	.	R	I	S	.	F	.	**G**	.	H	**A**	.	K	.	.	.	R	E	D	E	.	D	.
K. breviceps	.	R	I	.	L	.	.	.	**V**	H	**A**	.	K	G	**M**	S	R	.	D	E	.	.	G
P. phocoena	V	.	I	.	.	.	.	.	.	.	.	H	K	.	.	.	.	.	D	E	Q	.	.
G. melas	.	.	.	.	.	.	L	**G**	.	.	**A**	.	.	.	.	.	.	.	.	.	.	.	.
S. fluviatilis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
S. chinensis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
S. coeruleoalba	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
S. attenuata	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
S. frontalis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
S. longirostris	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
L. hosei	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. delphis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. capensis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. c. tropicalis	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T. truncatus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
T. aduncus	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Figure 3.1. Amino acid sequence alignment of variable amino acid sites in (a) exons 6-7 (aa 1-72, corresponding to aa 279-354 of *Mus musculus*) of ZP3, and (b) PKDREJ (aa 1-200). Sites identified by ML methods as being likely subjected to positive selection are in bold.

Phylogenetic trees obtained for ZP3 and PKDREJ with ML and Bayesian methods were concordant (Figure 3.2). In all trees, phylogenetic relationships among STDL species were unresolved. Despite this low level of resolution, overall topology is in agreement with published cetacean phylogenies (McGowen *et al.* 2009). Moreover, it has been suggested that the detection of positive selection is largely unaffected by possible uncertainties in underlying phylogenies (Pie 2006).

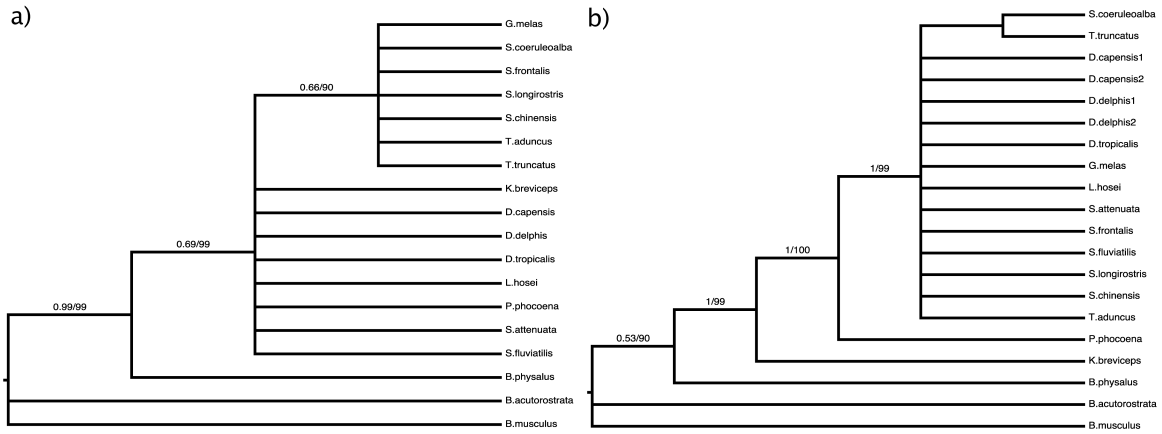


Figure 3. 2. Phylogenetic trees of cetaceans based on nucleotide sequences of (a) ZP3 and (b) PKDREJ genes. Individuals are identified by species name. Values above branches represent Bayesian posterior probabilities and Maximum Parsimony bootstrap support values, respectively.

The average d_N/d_S across all lineages and codon sites for both ZP3 and PKDREJ was calculated. When using the codon evolutionary model M0, which estimates a single d_N/d_S value across the whole tree, values were < 1 in all analyses, suggesting that these genes are evolving under selective constraints (Table 3.1). However, these proteins could contain amino acid sites subjected to positive selection that would be masked by a higher proportion of sites under purifying selection with ω close to zero. We thus compared three pairs of models (M1a vs M2a; M7 vs M8 and M8 vs M8a) using results from CODEML (Table 3.1). In both ZP3 and PKDREJ, the neutral models (M1a and M7) were not significantly different from the selection models (M2a and M8). However, in ZP3, the LRT test comparing the null model M8a with the selection model M8 was statistically significant at 5% if we consider the degree of freedom for the χ^2 statistic to be the 50:50 mixture of point mass 0 and χ_1^2 with a critical value of 2.71. The Bayes empirical Bayes (BEB) approach identified sites 14, 48, 55 and 59 as likely targets of positive selection with posterior probability values ranging from 0.54 to 0.78. In

PKDREJ, even though none of the LRT tests was statistically significant, the BEB approach identified sites 67, 73, 77 and 109 as likely targets of selection with posterior probability values ranging from 0.53 to 0.80. The methods SLAC and FEL failed to identify any sites in ZP3 and PKDREJ under positive selection. Nevertheless, the REL method identified sites 14, 48, 55 and 59 in ZP3 as positively selected with posterior probabilities > 0.95 (the same sites identified by the BEB approach), but failed to identify such sites in PKDREJ.

Table 3.1. Tests of adaptive evolution for (a) ZP3 and (b) PKDREJ using site models.

(a) ZP3 Model Code	*p*	d_N/d_S	*l*	Parameters	Positively Selected Sites BEB	LRT
M0	1	0.497		$\omega = 0.497$		
M1a	2	0.288	-340.476	$p_0 = 0.712$ ($p_1 = 0.288$) $w_0 = 0.000$ ($\omega_1 = 1.000$)		M1a-M2a 2.768
M2a	4	0.542	-339.092	p_0=0.847 p_1=0.000 p_2=0.152 ω_0=0.000 (ω_1=1.000) ω_2=3.554	48H(0.602); 55S(0.665); 59T (0.594)	
M7	2	0.2	-340.479	β (0.012, 0.005)		M7-M8 2.774
M8	4	0.621	-339.092	$p_0 = 0.847$ ($p_1 = 0.152$) β (0.005, 76.667) $\omega = 3.555$	14R(0.545); 48H(0.729); 55S (0.783); 59T (0.722)	
M8a	5	0.288	-340.475	$p_0 = 0.711$ ($p_1 = 0.288$) ω =1.000 β (0.007, 1.486)		M8-M8a 2.766*
(b) PKDREJ						
M0	1	0.604		$\omega = 0.604$		
M1a	2	0.578	-1123.576	$p_0 = 0.422$ ($p_1 = 0.578$) $\omega_0 = 1.000$ ($\omega_1 = 1.000$)		M1a-M2a 0.576
M2a	4	0.623	-1123.288	p_0=0.897 p_1=0.000 p_2=0.103 ω_0=0.384 (ω_1=1.000) ω_2=2.710	67G (0.706); 77A (0.539)	
M7	2	0.6	-1123.588	β (0.007, 0.005)		M7-M8 0.6
M8	4	0.623	-1123.288	$p_0 = 0.898$ ($p_1 = 0.102$) β (62.042, 99.000) $\omega = 2.714$	67G (0.804); 73I (0.530); 77A (0.648); 109I (0.587)	
M8a	5	0.577	-1123.576	$p_0 = 0.422$ ($p_1 = 0.578$) β (0.007, 1.988) $\omega = 1.000$		M8-M8a 0.576

Note: *p*, number of parameters relating to variation in d_N/d_S for each model; d_N/d_S, ratio averaged across all sites and lineages; $\omega = d_N/d_S$; *p* - proportion of codon sites in each class of ω; β = parameters for beta distribution; *l*, log likelihood of the model. *significant ($P < 0.05$).

The low number of amino acid substitutions in ZP3 and PKDREJ found among cetacean species contrasts to what has been described for rodents and primates. This observation is supported when d_N and d_S values are averaged across these groups, with cetaceans showing much lower synonymous and nonsynonymous substitution rates than rodents or primates (Table 3.2). However, when d_N/d_S is compared instead, values are actually higher in cetaceans for ZP3. In PKDREJ, rodents present the highest ratio, followed by cetaceans and then primates (Table 3.2). Within cetaceans, if mysticetes and odontocetes are considered separately,

d_N values are slightly higher than d_S values for mysticetes in ZP3 sequences, resulting in $\omega > 1$. When rates of protein evolution given by the d_N/d_S values are compared for the non reproductive proteins, BMP4 and MC1R, cetaceans show lower or similar values to the ones obtained for primates and rodents, which is in contrast to the pattern seen for the reproductive proteins, as described above (Table 3.2).

Table 3. 2. Nonsynonymous (d_N) and synonymous (d_S) substitution rates estimated based on pairwise comparisons of (a) ZP3, (b) PKDREJ, (c) MC1R and (d) BMP4 sequences from cetaceans, rodents and primates. Nucleotide divergence (d) estimated for cetacean species.

	d_N/d_S	d_N	d_S	d
(a) ZP3				
Primates	0.286	0.1407	0.1025	
Rodents	0.307	0.2029	0.6605	
Cetaceans	0.425	0.0068	0.0161	0.01
Mysticetes	1.112	0.0175	0.0157	0.016
Odontocetes	0.4096	0.0066	0.016	0.005
(b) PKDREJ				
Primates	0.264	0.0205	0.0777	
Rodents	0.875	0.1089	0.1245	
Cetaceans	0.623	0.0162	0.0259	0.005
Mysticetes	0.295	0.003	0.0102	0.006
Odontocetes	0.626	0.0164	0.0263	0.009
(c) BMP4				
Primates	0.066	0.0017	0.026	
Rodents	0.069	0.0297	0.433	
Cetaceans	0.049	0.0011	0.022	
(d) MC1R				
Primates	0.0902	0.0375	0.4153	
Rodents	0.1191	0.0626	0.5252	
Cetaceans	0.1088	0.01	0.0915	

3.5. DISCUSSION

In cetaceans, the occurrence of female promiscuity, leading to the existence of sperm competition and sexual conflict, would suggest that the evolution of reproductive proteins would be rapid, driven by positive selection, since an increased mating rate escalates sexual conflict (Birkhead & Pizzari 2002; Swanson & Vacquier 2002). However, in this study, we found very low levels of amino acid divergence in ZP3 and PKDREJ between species, particularly among delphinines. This lack of polymorphism resulted in the failure to reject the null model in favour of the positive selection model for PKDREJ, despite the fact that one of the codons (67) identified as likely target of positive selection, is very close

to the codons indentified as under positive selection in the human REJ domain (codon 285 in the human PKDREJ, Hamm *et al.* 2007). For ZP3, only one test (M8-M8a) out of five was statistically significant, identifying a few codons as likely targets of selection. One of such codons (55) corresponds to a serine residue (Ser-334) identified to be essential for sperm receptor activity in mouse, rat and human ZP3 (Chen *et al.* 1998). This residue, however, has only been retained in three species: *Balaenoptera acutorostrata, B. musculus* and *Phocoena phocoena*. Nevertheless, we would expect a strong signal of selection to be present in these proteins for the reasons described above.

In delphinines, a group of species recently diverged, it is possible that fertilization specificities are evolving slowly as a result of greater functional constraint on these reproductive proteins or reduced selective pressures for species recognition. In fact, hybridization among dolphin species seems to occur, both in captivity (Zornetzer & Duffield 2003) and in the wild (Bérubé 2002). It would, however, be expected in closely related species, where lineages have not yet sorted out completely, that differences in the genome would be found in the so-called speciation genes, those affecting target phenotypic traits or those involved in species recognition, such as ZP3 and PKDREJ (Wu 2001).

Other factors such as long generation times and intrinsic demographic features could also be dictating the slower evolution of proteins across the genome in cetaceans, and therefore affecting reproductive proteins as well. This is supported by our results, where both reproductive and non-reproductive proteins show overall low number of synonymous and nonsynonymous substitutions in cetaceans. It has been generally accepted that long-lived, larger mammals experience a slower mutation rate than small-bodied mammals due to several genomic features relating substitution rates with generation times and life-history traits (e.g. Bromham 2009; Jackson *et al.* 2009; Martin & Palumbi 1993). Our estimates of the rates of amino acid evolution appear to support this theory, since rodents present a consistently higher rate of amino acid evolution when compared to primates and cetaceans for all proteins, with the exception of ZP3. Although cetaceans present a marginally higher d_N/d_S ratio for this protein, this likely results

from the overall low number of both nonsynonymous and synonymous substitutions. When the mysticetes are analysed separately, the ratio is even higher due to higher number of nonsynonymous changes present in this lineage when compared to the odontocetes (see Figure 3.1). This is likely explained by noise at low levels of divergence, and not necessarily by the signal left by positive selection.

Studies in primates have shown a positive correlation between the intensity of sperm competition and degree of polyandry and the strength of positive selection in genes encoding for structural components of semen coagulum (semenogelin I, (Kingan *et al.* 2003) and semenogelin II (Dorus *et al.* 2004)), therefore supporting this theory. However, such association was not found between adaptive evolution of ZP3 and PKDREJ and the potential for sperm competition in rodents and primates (Turner and Hoekstra 2006; Swann *et al.* 2007; Hamm *et al.* 2007). The same appears to apply for cetaceans. Although varying degrees of sperm competition have been suggested across taxa (e.g. Connor *et al.* 2002), we did not find a strong overall pattern of positive selection in ZP3 and PKDREJ. However, it should be noted that within mysticetes, Balaenids appear to show higher levels of sperm competition than Balaenopterids. Female promiscuity has in fact been well documented in some species (e.g. minke whales (Skaug *et al.* 2008); humpback whales (Clapham and Palsboll 1997); gray whales (Swartz *et al.* 2006) and right whales (Frasier *et al.* 2007)). Unfortunately it was not possible to include all these species in this study, and its inclusion could have changed the results presented here, namely in the signal for positive selection in ZP3. Within odontocetes, our sampling of Delphinid species, for which more information on levels of sperm competition is available, would have enabled us to make such comparisons, if it was not for the complete lack of amino acid substitutions we observed.

It should be mentioned that the reduced statistical power to detect positive selection due to low polymorphism among our study species may have influenced our results, particularly because the use of χ^2 distribution makes the likelihood ratio tests very conservative for short, closely related sequences (Anisimova *et al.*

2001). Additionally, we cannot rule out that for PKDREJ, other regions than the one analysed here could be targets of positive selection in cetaceans. Nonetheless, the results obtained in this study were surprising and should initiate a discussion on the evolutionary forces driving the evolution of reproductive proteins in cetaceans and processes that may be dictating the establishment of reproductive isolation and species recognition. Future studies should focus on the study of additional species and reproductive proteins.

3.6. ACKNOWLEDGEMENTS

The authors wish to thank S. L. Kosakovsky Pond for guidance with data analyses in Datamonkey; M. C. Silva and J. R. Powell for helpful discussions; and two anonymous reviewers for helpful comments on previous versions of the manuscript. Kelly Robertson (SWFSC) and Marina Sequeira (ICN) provided tissue samples. A. R. Amaral is supported by a PhD grant (SFR/BD/27245/2006) funded by Fundação para Ciência e Tecnologia (Portugal).

3.7. REFERENCES

Anisimova M, Bielawski JP, Yang ZH (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* **18**, 1585-1592.

Arnason U, Gullberg A, Janke A (2004) Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* **333**, 27-34.

Berta A, Sumich JL (1999) *Marine Mammals: Evolutionary Biology.* Academic Press, San Diego.

Birkhead TR, Pizzari T (2002) Postcopulatory sexual selection. *Nature Reviews Genetics* **3**, 262-273.

Bromham L (2009) Why do species vary in their rate of molecular evolution? *Biology Letters* **5**, 401-404.

Butscheid Y, Chubanov V, Steger K, Meyer D, Dietrich A, Gudermann T (2006) Polycystic kidney disease and receptor for egg jelly is a plasma membrane protein of mouse sperm head. *Molecular Reproduction and Development* **73**, 350-360.

Caballero S, Jackson J, Mignucci-Giannoni AA, Barrios-Garrido H, Beltran-Pedreros S, Montiel-Villalobos MG, Robertson KM, Baker CS (2008) Molecular systematics of South American dolphins *Sotalia*: Sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. *Molecular Phylogenetics and Evolution* **46**, 252-268.

Calkins JD, El-Hinn D, Swanson WJ (2007) Adaptive evolution in an avian reproductive protein: ZP3. *Journal of Molecular Evolution* **65**, 555-563.

Chen J, Litscher ES, Wassarman PM (1998) Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6193-6197.

Clapham PJ, Palsboll PJ (1997) Molecular analysis of paternity shows promiscuos mating in female humpback whales (*Megaptera novaengliae*, Borowski). *Proceedings of the Royal Society of London, Series B* **264**, 95-98.

Connor RC, Read AJ, Wrangham R (2000) Male Reproductive Strategies and Social Bonds. *In* Mann J, Connor RC, Tyack PL, Whitehead H, eds. *Cetacean Societies: Field studies of Dolphins and Whales*. 247-270 pp. The Chicago University Press, Chicago.

Demere TA, McGowen MR, Berta A, Gatesy J (2008) Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. *Systematic Biology* **57**, 15-37.

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature Genetics* **36**, 1326-1329.

Frasier TR, Hamilton PK, Brown MW, Conger LA, Knowlton AR, Maryx MK, Slay CK, Kraus SD, White BN (2007) Patterns of male reproductive success in a highly promiscuous whale species: The endangered North Atlantic right whale. *Molecular Ecology* **16**, 5277-5293.

Hamm D, Mautz BS, Wolfner MF, Aquadro CF, Swanson WJ (2007) Evidence of amino acid diversity-enhancing selection within humans and among primates at

the candidate sperm-receptor gene PKDREJ. *American Journal of Human Genetics* **81**, 44-52.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-Gonzalez L, Palumbi SR (2009) Big and slow: Phylogenetic estimates of molecular evolution in baleen whales (Suborder Mysticeti). *Molecular Biology and Evolution* **26**, 2427-2440.

Jansa SA, Lundrigan BL, Tucker PK (2003) Tests for positive selection on immune and reproductive genes in closely related species of the murine genus *Mus*. *Journal of Molecular Evolution* **56**, 294-307.

Kingan SB, Tatar M, Rand DM (2003) Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. *Journal of Molecular Evolution* **57**, 159-169.

Martin AP, Palumbi SR (1993) Body size, metabolic-rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America.* **90**, 4087-4091.

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

Metz EC, Robles-Sikisaka R, Vacquier VD (1998) Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10676-10681.

Perrin WF, Reeves RR (2004) Appendix 5: Report of the Working Group on Species- and Subspecies-level Taxonomy. In: *Report of the Workshop on Shortcomings of Cetacean Taxonomy in Relation to Needs of Conservation and Management, April 30-May 2* (eds. Reeves RR, Perrin WF, Taylor BL, Baker CS, Mesnick SL), pp. 25-60. NOAA technical memorandum, La Jolla, California.

Pie MR (2006) The influence of phylogenetic uncertainty on the detection of positive darwinian selection. *Molecular Biology and Evolution* **23**, 2274-2278.

Pond SLK, Frost SDW (2005a) Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533.

Pond SLK, Frost SDW (2005b) Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**, 1208-1222.

Posada D, Buckley T (2004) Model selection and model averaging in phylogenetics: Advantages of Akaike Information Criterion and Bayesian approaches over Likelihood Ratio Tests. *Systematic Biology* **53**, 793-808.

Skaug HJ, Bérubé M, Rew MB, Palsboll PJ (2008) Genetic analyses reveal promiscuous mating in female common minke whales, *Balaenoptera acutorostrata*. *Journal of Cetacean Research and Management* **9**, 249-251.

Sutton KA, Jungnickel MK, Florman HM (2008) A polycystin-1 controls postcopulatory reproductive selection in mice. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8661-8666.

Sutton KA, Jungnickel MK, Ward CJ, Harris PC, Florman HM (2006) Functional characterization of PKDREJ, a male germ cell-restricted polycystin. *Journal of Cellular Physiology* **209**, 493-500.

Swann CA, Cooper SJB, Breed WG (2007) Molecular evolution of the carboxy terminal region of the zona pellucida 3 glycoprotein in murine rodents. *Reproduction* **133**, 697-708.

Swanson WJ, Nielsen R, Yang QF (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution* **20**, 18-20.

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**, 137-144.

Swanson WJ, Zhang ZH, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2509-2514.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other

Methods). Sinauer Associates, Sunderland, Massachusetts.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**,1596-1599.

Turner LM, Hoekstra HE (2006) Adaptive evolution of fertilization proteins within a genus: Variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Molecular Biology and Evolution* **23**, 1656-1669.

Turner LM, Hoekstra HE (2008) Causes and consequences of the evolution of reproductive proteins. *International Journal of Developmental Biology* **52**, 769-780.

Wassarman PM, Jovine L, Litscher ES (2001) A profile of fertilization in mammals. *Nature Cell Biology* **3**, E59-E64.

Wassarman PM, Jovine L, Litscher ES (2004) Mouse zona pellucida genes and glycoproteins. *Cytogenetic and Genome Research* **105**, 228-234.

Wassarman PM, Litscher ES (2001) Towards the molecular basis of sperm and egg interaction during mammalian fertilization. *Cells Tissues Organs* **168**, 36-45.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.

Yang Z (2000) Phylogenetic Analysis by Maximum Likelihood (PAML). University College, London.

Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22**, 1107-1118.

Zornetzer HR, Duffield DA (2003) Captive-born bottlenose dolphin x common dolphin (*Tursiops truncatus* x *Delphinus capensis)* intergeneric hybrids. *Canadian Journal of Zoology* **81**, 1755-1762.

3.8. SUPPLEMENTARY MATERIAL

Table S3.1. List of cetacean species included in this study.

Species	Accesion Numbers ZP3	PKDREJ	Geographic Location	Institution[a]
Suborder Mysticeti				
Family Balaenopteridae				
Balaenoptera musculus (Blue whale)	FJ490593	FJ490575	Australia	Macquarie University
B. physalus (Fin whale)	FJ490594	FJ490576	Portugal	ICN
B. acutorostrata (Minke whale)	FJ490592	FJ490574	Portugal	ICN
Suborder Odontoceti				
Family Kogidae				
Kogia breviceps (Pigmy sperm whale)	FJ490599	FJ490581	Portugal	ICN
Family Phocoenidae				
Phocoena phocoena (Harbour porpoise)	FJ490601	FJ490583	Portugal	ICN
Family Delphinidae				
Subfamily Globicephalinae				
Globicephala melas (Pilot whale)	FJ490598	FJ490580	Portugal	ICN
Subfamily Stenoninae				
Sotalia fluviatilis (Tucuxi)	-	FJ490587	Brazil	SWFSC
Subfamily Delphininae				
Sousa chinensis (Indo-Pacific humpback dolphin)	FJ490604	FJ490589	Hong Kong	SWFSC
Lagenodelphis hosei (Fraser's dolphin)	FJ490600	FJ490582	Hawaii	SWFSC
Stenella longirostris (Spinner dolphin)	FJ490606	FJ490588	East Pacific	SWFSC
S. attenuata (Pantropical spotted dolphin)	FJ490602	FJ490584	Mexico	SWFSC
S. coeruleoalba (Striped dolphin)	FJ490603	FJ490585	Portugal	ICN
S. frontalis (Atlantic spotted dolphin)	FJ490605	FJ490586	Portugal	ICN
Delphinus delphis (Short-beaked common dolphin)	FJ490596	FJ490578	Portugal	ICN
D. capensis (Long-beaked common dolphin)	FJ490595	FJ490577	East Pacific	SWFSC
D. c. tropicalis (Arabian common dolphin)	FJ490597	FJ490579	Arabian Sea	SWFSC
Tursiops truncatus (Bottlenose dolphin)	FJ490607	FJ490591	Portugal	ICN
T. aduncus (Indo-Pacific bottlenose dolphin)	-	FJ490590	Australia	Macquarie University

Table S3.2. GenBank and Ensembl Accession Numbers for sequences used in the comparison of synonymous and nonsynonymous substitution rates in ZP3, PKDREJ, MC1R and BMP4 from rodents and primates.

	Accession Numbers			
Species	ZP3	PKDREJ	BMP4	MC1R
Homo sapiens	ENST00000394857	EF517278.1	NM_130851.2	AY363627.1
Gorilla gorilla	ENSGGOG00000005450	EF517281.1	-	AY205088.1
Macaca mulatta	ENSMMUT00000003633	EF517284.1	XM_001084200.1	AY205103.1
Macaca nigra	-	EF517285.1	-	AY205102.1
Pan troglodytes	ENSPTRT00000035756	EF517279.1	XM_509954.2	AY205086.1
Pan paniscus	-	EF517280.1	-	AB296237.1
Pongo pygmaeus	-	EF517282.1	ENSPPYG00000005826	AY205087.1
Callithrix jacchus	ENSCJAG00000016252	-	-	AY205120.1
Rattus norvegicus	NM_053762.1	NM_001134866	NC_005114	AB306978.1
Mus musculus	AY057779.1	NM_011105.2	NC_000080.5	BC119294
Peromyscus polionatus	DQ668304.1	-	-	FJ389440.1
Peromyscus truei	EU568744.1	-	-	-
Peromyscus maniculatus	-	-	-	GQ337978
Meriones unguiculatus	-	AB201310.1	-	-
Dipodomys ordii	-	ENSDORT00000010496	-	-
Cavia porcellus	-	ENSCPOG0000000688	-	-
Eothenomys melanogaster	-	-	-	GU001572.1

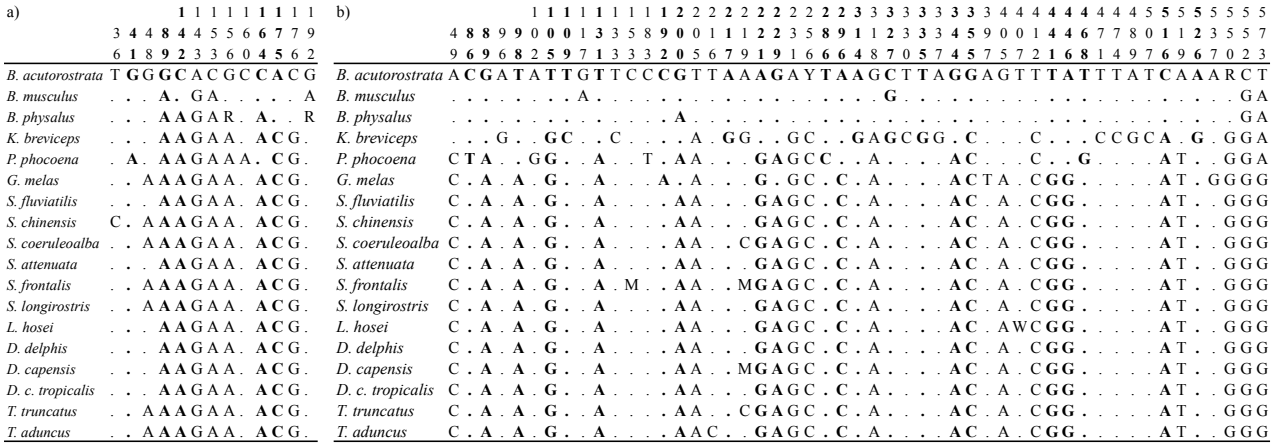


Figure S3.1. Nucleotide sequence alignment of the observed variable sites in a) exons 6 and 7 of the gene ZP3 and in b) the gene PKDREJ of the cetacean species analysed. Nucleotide changes that result in amino acid changes are highlighted in bold.

Chapter IV (Article 3)
Species Tree of a Rapid Radiation: The Subfamily Delphininae (Cetacea, Mammalia)

Ana R. Amaral[1,2*], Jennifer A. Jackson[3], Luciana M. Möller[4], Luciano B. Beheregaray[2,4] and M. Manuela Coelho[1]

[1] Universidade de Lisboa, Faculdade de Ciências, Centro de Biologia Ambiental, Campo Grande, 1749-016 Lisbon, Portugal
[2] Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
[3] British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0Et, England
[4] School of Biological Sciences, Flinders University, Adelaide, SA, 5001, Australia

Molecular Phylogenetics and Evolution (in review)

Species tree of a rapid radiation: The subfamily Delphininae (Cetacea, Mammalia)

4.1. ABSTRACT

Species undergoing rapid radiations provide exceptional cases for the study of processes of speciation and adaptation, but also represent a challenge for molecular systematics because ancestral retention of polymorphisms and the occurrence of hybridization can obscure relationships among lineages. Dolphins in the subfamily Delphininae are one such case. Non-monophyly, rapid speciation events, and discordance between morphological and molecular characters have made the inference of phylogenetic relationships within this subfamily very difficult. Here we compare the performance of different methods intended to estimate species trees using a multi-gene dataset for the Delphininae (*Sousa, Sotalia, Stenella, Tursiops, Delphinus* and *Lagenodelphis*). Incongruent gene trees obtained indicate that incomplete lineage sorting and possibly hybridization are confounding factors for the inference of species history in this group. Nonetheless, using coalescent-based methods, we have been able to extract an underlying species-tree signal from divergent histories of independent genes, which supports relationships based on morphology. This is the first time that a molecular study has provided support for such relationships. This study further illustrates how methods for species-tree inference can be very sensitive both to the characteristics of the dataset and the evolutionary processes affecting the evolution of the group under study.

Key Words: Delphinidae; incomplete lineage sorting; hybridization; rapid radiation; speciation; species tree

4.2. INTRODUCTION

Species complexes undergoing rapid radiation provide an exceptional opportunity to understand the processes of speciation and adaptation. They also represent a major challenge in molecular systematics because relationships among lineages

can be hidden either by incomplete lineage sorting or introgressive hybridization (Maddison & Knowles 2006; Wiens *et al.* 2006).

During a rapid radiation, gene copies may not diverge in parallel with speciation events, with the consequence that the coalescence pattern of individual gene phylogenies may not match the true pattern of speciation (Hudson 1992). As a result, many gene trees will be discordant between each other and from the actual species tree (reviewed in Degnan & Rosenberg 2009; Knowles 2009).

Furthermore, in cases of rapid radiations, the intrinsic barriers that prevent gene flow between species may have insufficient time to develop fully, leading to hybridization among recently evolved lineages (Seehausen 2004). The extent to which gene flow persists throughout the process of speciation remains unclear, although it has been documented in a number of recent studies (Niemiller *et al.* 2008; Quesada *et al.* 2007; Savolainen *et al.* 2006). The existence of hybridization events in the evolutionary history of a group means that such taxa will not follow the usual procedure of divergence from a common ancestor through a bifurcating tree (Hennig 1966), resulting in discordant gene trees.

Incomplete lineage sorting and hybridization are only two of the evolutionary processes that can lead to discordance between gene trees and species trees. Horizontal gene transfer and gene duplication can also lead to such incongruence (Maddison 1997). However, incomplete lineage sorting and hybridization are the processes that have been more thoroughly studied in a phylogenetic context, leading to the development of several methods that incorporate the stochastic sorting of lineages in the estimation of species trees from gene trees (Kubatko *et al.* 2009; Liu 2008; Liu & Pearl 2007; Maddison 1997; Maddison & Knowles 2006). Analyses of empirical and simulated data suggest that these methods can accurately estimate species trees even when high levels of discordance between gene trees exist (Brumfield *et al.* 2008; Linnen & Farrell 2008; Liu *et al.* 2008), and are therefore useful for investigating relationships in species complexes that show rapid diversification and very confusing taxonomy (Belfiore *et al.* 2008; Dolman & Hugall 2008).

In this study we use three different methods for estimating species trees, applying these to a subfamily of dolphins that has recently radiated and presents a highly confounding and controversial taxonomy. Firstly, we use a concatenation approach combining multigene data and standard methods of phylogenetic inference to estimate a species tree. The reasoning behind this approach is that by combining the data, a dominant signal will emerge, resulting in a more strongly supported phylogenetic estimate, assumed to be the species tree (de Queiroz & Gatesy 2007; Gadagkar *et al.* 2005; Huelsenbeck *et al.* 1996; Rokas & Carroll 2005). This approach has recently been criticized (Degnan & Rosenberg 2006; Kubatko & Degnan 2007), mainly because if substantial variation in single-gene histories exists, this variance is not incorporated, and phylogenetic signals from the most variable loci will tend to dominate, misleading inference of the true species evolutionary history. Therefore, we also use three different methods that incorporate the coalescent in the estimation of species trees: a summary statistic method, the 'minimize deep coalescence' method (Maddison & Knowles 2006) and two probabilistic methods that combine Bayesian models in a coalescent framework as implemented in the programs BEST (Liu 2008) and *BEAST (Heled & Drummond 2010). These methods differ in the way the information from the coalescent is incorporated, in how coalescent times are summarized, and in the incorporation of uncertainty in the estimated species tree (Knowles 2009; Liu *et al.* 2009). The comparison between these different methods allows us to explore the evolutionary processes that have shaped the evolution of this dolphin group, as well as to assess their performance in estimating a species tree from a recent and likely rapid radiation.

The subjects of this study, dolphins of the subfamily Delphininae (family Delphinidae) have likely evolved through a rapid radiation (McGowen *et al.* 2009; Kingston *et al.* 2009; Steeman *et al.* 2009; Slater *et al.* 2010). Incongruence between mtDNA and nuclear phylogenies and incomplete lineage sorting (Kingston *et al.* 2009), uncertainty in the placement of taxa (Xiong *et al.* 2009) and a rise in net diversification rate within Delphinidae (Slater *et al.* 2010), all support a rapid radiation of these taxa. The Delphininae therefore exemplify the challenges of inferring species boundaries and relationships described above.

The rapid radiation originating the Delphinidae is estimated to have occurred during the mid to late Miocene (11-15 Mya) (Barnes *et al.* 1985). The potential drivers of this explosive radiation are related to social structure, growth and reproductive characteristics (Gygax 2002), trophic diversification (Lipps & Mitchell 1976), and/or climatic changes during the glacial periods of the Pleistocene (Steeman *et al*. 2009). Delphinidae is the largest cetacean family, comprising at least 37 species (Caballero *et al.* 2008). Although several phenetic morphological (e.g. Flower 1883; Mead 1975; Muizon 1988; True 1889) and cladistic molecular (Caballero *et al.* 2008; LeDuc *et al.* 1999; May-Collado & Agnarsson 2006) studies have been conducted, the evolutionary relationships within this family remain unclear, particularly within one of its subfamilies: the Delphininae (*Sotalia*, *Sousa, Stenella, Tursiops, Delphinus* and *Lagenodelphis*) (Figure 4.1, Table 4.1).

The monophyly of the *Stenella* and *Tursiops* genera has been questioned for more than a century (True 1889) due to a complex of cranial characters not shared by all species of the genus *Stenella*, some of which may actually be more closely related to *Tursiops* or *Delphinus* than to their congeners. Recent phylogenetic studies based on newly sequenced mitochondrial genomes and AFLP markers have supported the polyphyly of *Stenella* and *Tursiops* (Kingston *et al.* 2009; Xiong *et al.* 2009). However, two other studies using nuclear gene sequences and a supermatrix approach to infer the phylogeny of whales and dolphins have recovered *Tursiops* as monophyletic (McGowen *et al.* 2009; Steeman *et al.* 2009). The number of species within the genera *Tursiops* and *Delphinus* has also been a point of contention. Within *Tursiops*, most recent studies recognize two species, the common bottlenose dolphin (*T. truncatus*) and the Indo-Pacific bottlenose dolphin (*T. aduncus*). However, recent molecular evidence based on mitochondrial DNA suggested that *aduncus*-type dolphins occurring off South Africa (Natoli *et al*., 2004), as well as coastal bottlenose dolphins from southern Australia (Möller *et al.* 2008; Natoli *et al.* 2004) may actually be different species, with the latter being more closely related to *Lagenodelphis hosei* than to the *truncatus* or *aduncus* types (see Figure 4.1 and Appendix 4B). Within *Delphinus*, there are two currently recognized species, the short-beaked common dolphin (*D. delphis*) and the long-beaked common dolphin

(*D. capensis*). In addition, an extremely long-beaked form from the Arabian Sea is considered a subspecies (*D. capensis tropicalis*) (Jefferson & Van Waerebeek 2002), and common dolphins in the Black Sea are also recognized as a subspecies (*D. delphis ponticus*) (Perrin 2009). However, recent morphological and molecular evidence suggest that at least in some geographical areas this classification (which is based primarily on morphological characters such as beak length and coloration) may not be valid (Amaral *et al.* 2007a; Murphy *et al.* 2006).

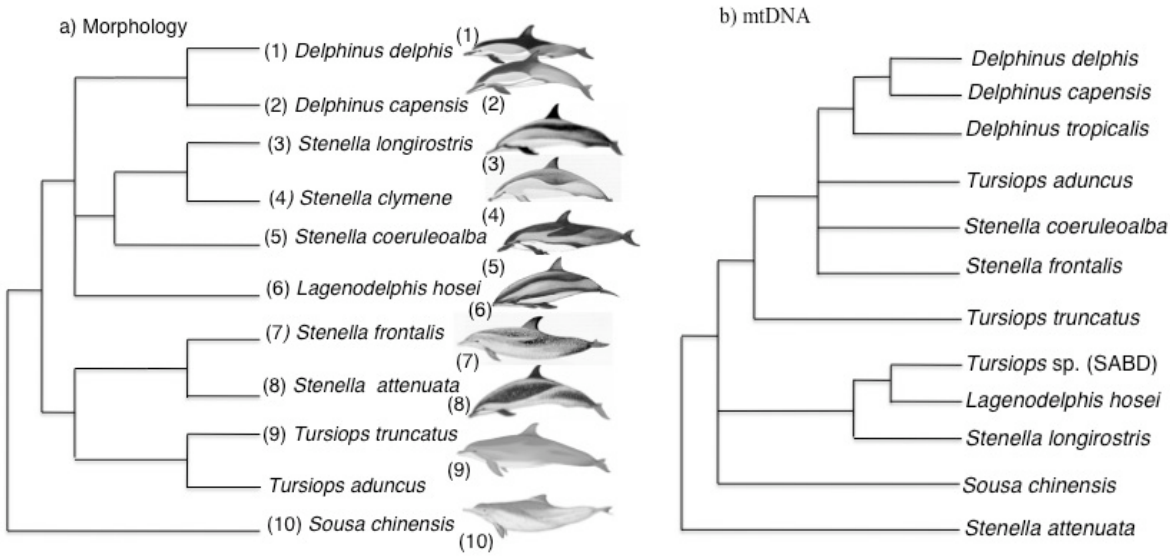


Figure 4. 1. Relationships within the nominal species in the subfamily Delphininae based on a) recent morphological analyses (Perrin *et al*., 1987; Perrin 2009) and b) mitochondrial DNA cytochrome *b* gene (adapted from LeDuc *et al*., 1999; Möller *et al*., 2008).

The use of inadequate phylogenetic methods that do not capture the complex nature of DNA evolution in cetaceans has been cited as a reason for the poorly clarified evolutionary relationships within the Delphininae (Xiong *et al*. 2009; May-Collado & Agnarsson 2006). Moreover, the datasets used so far have proved to be insufficient to resolve the phylogenetic tree of the subfamily. Nearly all molecular studies conducted so far have focussed on the mitochondrial genome (using the cytochrome *b* gene or full genome, e.g. Xiong *et al*., 2009; May-Collado & Agnarsson 2006; LeDuc *et al*. 1999). Four other studies have included DNA sequences from multiple nuclear loci. Two were aimed at elucidating phylogenetic relationships within the order Cetacea and used supermatrix approaches (McGowen *et al.* 2009; Steeman *et al.* 2009), while the third applied phylogenetic

methods to single-locus and concatenated datasets (Caballero *et al.,* 2008) and a fourth used Amplified Fragment Length Polymorphism (AFLP) markers (Kingston *et al.* 2009).

Table 4.1. List of cetacean species included in this study.

Species	Number of samples sequenced	Geographic Location	Institution[a]
Family Delphinidae			
Subfamily Delphininae			
Sotalia fluviatilis (Tucuxi)	1	Brazil	SWFSC
Sousa chinensis (Indo-Pacifichumpback dolphin)	2	Hong Kong	SWFSC
Lagenodelphis hosei (Fraser's dolphin)	1	Hawaii	SWFSC
Stenella longirostris (Spinner dolphin)	2	East Pacific	SWFSC
S. attenuata (Pantropical spotted dolphin)	2	Mexico	SWFSC
S. coeruleoalba (Striped dolphin)	2	Portugal	ICN
S. frontalis (Atlantic spotted dolphin)	1	Portugal	ICN
Delphinus delphis (Short-beaked common dolphin)	2	Portugal	ICN
D. capensis (Long-beaked common dolphin)	2	East Pacific	SWFSC
D. c. tropicalis (Arabian common dolphin)	1	Arabian Sea	SWFSC
Tursiops truncatus (Common bottlenose dolphin)	2	Portugal	ICN
T. aduncus (Indo-Pacific bottlenose dolphin)	2	Australia	MQ
Southern Australian Bottlenose Dolphin	2	Australia	MQ
Subfamily Globicephalinae			
Globicephala melas (Long-finned pilot whale)	1	Portugal	ICN
Family Phocoenidae			
Phocoena phocoena (Harbour porpoise)	1	Portugal	ICN

Note: [a] Institutional abbreviations: ICN – Instituto de Conservação da Natureza (Portugal); SWFSC – Southwest Fisheries Science Center (U.S.A.); MQ – Macquarie University (Australia).

Here we utilize a multi-gene dataset, including 1 mitochondrial gene (cytochrome *b*) and 13 nuclear loci (3 introns and 10 anonymous) to estimate, for the first time, a species tree for the subfamily Delphininae. For the first time, coalescent-based methods that account for gene tree heterogeneity were used. The following questions were addressed: (1) Are the coalescent-based methods for species tree estimation able to consistently resolve relationships within the Delphininae? (2) Do these relationships differ from previously published mtDNA and nuDNA phylogenies, namely in resolving the polyphyly of the genera *Tursiops* and *Stenella*? (3) Are there differences in the species tree topology obtained with the different methods used? In a broader context, our study contributes to recent analytical debates concerning gene trees and species trees and helps to clarify the evolutionary history of a likely rapid radiation, of a globally distributed and charismatic group of organisms.

4.3. MATERIALS AND METHODS

4.3.1. *Sample acquisition, DNA extraction, amplification and sequencing*

A total of 21 individual samples comprising 12 species belonging to the subfamily Delphininae were used (Table 4.1). Additionally, *Globicephala melas* (Subfamily Globicephalinae) and *Phocoena phocoena* (Family Phocoenidae) were used as outgroups in the phylogenetic analyses (Caballero *et al*., 2008). Samples were obtained as skin or muscle tissue from dead stranded animals or from free-ranging animals using biopsy darts. Some samples were received from the Southwest Fisheries Science Center, Marine Mammal and Turtle Research Sample Collection (SWFSC-NOAA, La Jolla, CA) as extracted DNA. Recognized experts made all species identifications.

DNA was extracted from skin or muscle tissue following standard phenol-chloroform extraction protocols (Sambrook *et al*. 1989). The Polymerase Chain Reaction (PCR) was used to amplify one fragment of the mitochondrial genome (the cytochrome *b* gene), three nuclear introns (BTN, CHRNA1 and PLP) and ten nuclear anonymous loci (Table 4.2). Anonymous markers are non-coding regions of the genome, randomly collected and presumably dispersed across the chromosomes, thereby representing wide and potentially unbiased variation across the genome. These loci were developed from clone sequences selected from a genomic library created for the common dolphin *D. delphis* (Amaral *et al.* 2010). The PCR reactions were performed in 25-μL reactions containing 10-100 ng DNA, 0.2 mM each dNTP, 0.3 *μ*M each primer, 1 U *Taq* Polymerase and 1 X *Taq* buffer. PCR products were separated on 1.0% agarose gels, stained with ethidium bromide and visualized with ultraviolet light. PCR products were cleaned with Exonuclease I and Shrimp Alkaline Phosphatase to remove free nucleotides and primers, and sequenced in both directions (BigDye Terminator CycleSequencing: Applied Biosystems) on an ABI 3730xl automated sequencer (Applied Biosystems).

Sequences were manually edited and aligned using Sequencher v.4.2 (Gene Code Corporation). Some alignments for the nuclear loci required gaps: in Del_02

a gap of 90 bp was required; in Del_05 a gap of 8 bp was required; in Del_08 several simple gaps were required due to length variations of AT repeats, as well as an additional gap of 36 bp; and in Del_11 a gap of 19 bp was required. In BTN a simple gap of 2 bp was required. Alignments were confirmed using CLUSTALX v.2.0.10 with the default parameter settings.

Table 4.2. List of loci and primers used in this study.

Locus	Primers	Approximate product size (bp)	Reference
Cyt*b*	L14724 P2	1120	LeDuc et al. 1999
BTN	But-b1s BTNr4	754	Lyons et al. 1997
PLP	PLP-F PLP-R	750	Lyons et al. 1997
CHRNA1	CHRNA1F CHRNA1R	357	Roca et al. 2001
Del_02	Del_02F Del_02R	923	Amaral et al. 2010
Del_04	Del_04F Del_04R	636	Amaral et al. 2010
Del_05	Del_05F Del_05R	750	Amaral et al. 2010
Del_08	Del_08F Del_08R	806	Amaral et al. 2010
Del_10	Del_10F Del_10R	402	Amaral et al. 2010
Del_11	Del_11F Del_11R	572	Amaral et al. 2010
Del_12	Del_12F Del_12R	729	Amaral et al. 2010
Del_14	Del_14F Del_14R	318	Amaral et al. 2010
Del_15	Del_15F Del_15R	780	Amaral et al. 2010
Del_17	Del_17F Del_17R	739	Amaral et al. 2010

The direct sequencing of the PCR products for the anonymous loci Del_02, Del_05, Del_10 and Del_12 frequently generated continuous overlap of signals between sequences amplified with the forward and reverse primers. This indicated the existence of a length polymorphism in the amplified region. We cloned the PCR product (performed at Macrogen, Inc) and sequenced between 8-10 cloned fragments to determine the two allelic sequences. We then used the program Indelligent (http://ctap.inhs.uiuc.edu/dmitriev/indel.asp) to check if the two allelic sequences identified matched the expected ones. Although an error in PCR can create artificial heterozygosity, we performed PCRs for cloning and direct sequencing independently, and still found consistency between the

overlapping signals on the direct and cloned sequences. The results thus appeared to be unaffected by PCR artefacts. The program PHASE v.2.1 (Stephens & Donnelly 2003; Stephens *et al.* 2001) was used to infer alleles from heterozygous individuals, setting the phase-certainty threshold at 90%.

4.3.2*. Dataset construction*

Both combined and separate analyses of the major partitions of loci were conducted. Three different datasets were constructed: "mtDNA" including the cytochrome *b* gene; "nuDNA" including the three introns and the 10 anonymous loci; and "mtDNA+nuDNA" including the cytochrome *b* gene, the three introns and the 10 anonymous loci.

In order to test whether heterozygote sites, allele size polymorphisms and insertions/deletions would have any influence on the estimated phylogenetic trees, we performed analyses considering: i) heterozygous sites coded using the IUPAC ambiguity code and gaps treated as missing "new states" or coded with a binary code (0 or 1) indicating their presence or absence; and ii) including information from heterozygotes and allele size polymorphisms by including all allelic sequences from each individual in the input matrix. Gene tree topologies did not vary substantially with the inclusion of gap partitions and all allelic sequences. Numbers of equally parsimonious trees obtained, tree lengths, bootstrap nodal support and posterior probability values were altered only slightly. Therefore subsequent analyses were performed with heterozygous sites coded as IUPAC ambiguity codes and gaps treated as missing data. Using all allele sequences would be computationally very demanding, particularly for estimating species trees using the program BEST.

4.3.3. *Phylogenetic analyses*

MrBayes v.3.1 (Huelsenbeck & Ronquist 2001) was used to estimate Bayesian phylogenetic trees for each locus and for the concatenated dataset. Four simultaneous Metropolis-Coupled MCMC chains (one cold and three heated) were run for 2 million generations, with trees sampled at intervals of 100 generations. Random trees were used to begin each Markov chain, and a molecular clock was

not enforced. Convergence was assessed by the standard deviation of split frequencies, and by the achievement of stationarity of the log-likelihood values of the cold chain. The first 2000 trees were discarded as "burn-in" after examining the variation in log-likelihood scores over time. The cytochrome *b* dataset was partioned by codon positions in the Bayesian analysis, assuming that there might be differences in the molecular evolution of the different positions.

Modeltest v.3.7 (Posada & Crandall 1998) was used to infer the best-fitting evolutionary model for each locus. Models of evolution were chosen for subsequent analyses according to a second-order Akaike Information Criterion (AIC_c), with branch lengths included as additional parameters and a correction for small sample sizes employed (Hurvich & Tsai 1995; Posada & Buckley 2004). Nucleotide substitution models for each locus are given in Table 4.3. We tested for incongruence among loci by performing 'crossed' Shimodaira-Hasegawa tests in PAUP* (Shimodaira & Hasegawa 1999), whereby the highest likelihood topologies obtained with individual datasets were compared against each other (using the AIC_c-preferred evolutionary model for each dataset) (Table 4.4), and including comparisons of individual datasets against the ML phylogenies obtained by all nuclear loci concatenated, and all loci concatenated (e.g. Delsuc *et al.* 2002). With this analysis we intend to test whether the mitochondrial gene tree is more discordant with nuclear gene trees than nuclear gene trees are with one another. Phylogenetic trees for each locus were also obtained using maximum parsimony (MP) method. Details of this analysis and results are included in Appendix 4A.

4.3.4. *Estimation of species trees*

In addition to the concatenated analysis performed using Bayesian methods and maximum parsimony (Appendix 4A), three different coalescent-based methods were used to estimate the species tree from the 13 nuclear gene trees obtained: Minimize Deep Coalescence (MDC), Bayesian Estimation of Species Trees (BEST) and *BEAST. The MDC approach seeks the species tree that minimizes the number of incomplete lineage sorting (deep coalescence) events, which must be inferred to explain observed gene trees (Maddison 1997). This approach was

implemented in Mesquite v.2.72 (Maddison & Maddison 2009) using the individual gene trees estimated using MP and Bayesian inference as described above. Following the methods proposed by Linnen & Farrell (2008), tree searches were first performed using the following options: contained polytomies automatically resolved, branch lengths of contained trees included, and tree rearrangements made by subtree pruning and regrafting (SPR) swapping. A second search was performed without automatically resolving polytomies and without including branch lengths, so as to evaluate the sensitivity of the species tree to this search strategy.

Table 4.3. Modeltest minimum AIC models (taxa and character-corrected) for each locus. *AIC was incalculable for this locus.

Locus	Nucleotide substitution model
Del_02	HKY
Del_04	JC
Del_05	TIM
Del_08	JC
Del_10	GTR+I
Del_11	K81f
Del_12	JC
Del_14	N/A*
Del_15	HKY
Del_17	HKY
BTN	JC
CHRNA1F	GTR
PLP	K80
CYTB	GTR+G (0.418)

The Bayesian Estimation of Species Trees (BEST) uses a Bayesian hierarchical model to estimate a distribution of species trees from vectors of estimated gene trees across multiple loci, under a multispecies coalescent model (Liu & Pearl 2007). BEST uses MrBayes to generate a posterior distribution of gene trees across loci using a prior based on an approximate species tree; it then estimates a species tree from the joint posterior distribution of gene trees using a uniform prior method. The analysis was implemented in BEST v.2 using the partitioned combined dataset described above. In order to minimize over-parameterization of the individual datasets in the shared Bayesian framework, one transition:transversion ratio and set of base frequencies was co-estimated for all

datasets combined. Sensitivity of each dataset to this approach was determined by comparing the maximum likelihood phylogeny from each dataset (as estimated in PAUP* using the AICc-preferred evolutionary models summarized in Table 4.3) with the phylogeny produced under the parameters of the combined evolutionary model using Shimodaira-Hasegawa testing (Shimodaira & Hasegawa 1999): this evolutionary model was not found to be a significantly worse fit for any dataset. In order to investigate the ability of BEST to achieve convergence, we performed one long analysis of 100 million generations, using an inverse gamma prior of α = 3 and β = 0.003 (theta (θ) = 0.0015) and using two independent runs of four simultaneous Metropolis-coupled MCMC chains (one cold and three heated), sampling every 1000th tree with the first 10 million generations discarded as "burn-in". To investigate the impact of population size prior choice on species-tree inference, two further analyses were performed with alternative inverse gamma priors, chosen so as to bound the sequence-based estimates of θ calculated in DNAsp (Rozas *et al.* 2003). These were α = 2 and β = 0.001 and 0.002, corresponding to θ = 0.001 and 0.002, respectively. Each analysis was run for 20 million generations and summarized over ten to twelve independent runs, with 2 million generations discarded as "burn-in". In order to examine the impact of using different branch-length priors, we further performed two analyses using a coalescent (uniform clock) branch-length prior and inverse gamma priors of θ = 0.001 and θ = 0.002. We also tested the phylogenetic impact of widening the range of the mutation rate prior for the coalescent prior analyses, since inspection of the posterior mutation rates of these analyses revealed some median values close to the prior boundaries.

The standard prior (which allowed relative rates to be uniform over 0.5-1.5 of the mean value) was therefore modified to a wider range of (0.1-2). A BEST analysis using the inverse gamma prior of θ = 0.0015 (and the coalescent tree prior and wider mutation rate prior) was also carried out using ten independent runs over 20 million generations, since this theta value is closest to the true estimate derived for the population. Posterior summary distributions were inspected for

convergence and mixing using the program TRACER v.1.5 (Rambaut & Drummond 2007).

Table 4.4. Crossed Shimodaira-Hasegawa (SH) tests for congruence among nuclear and mitochondrial loci. Maximum likelihood (ML) Log-likelihood (LnL) scores for each dataset are shown along diagonal. Values in the rest of the matrix show the LnL difference between the ML tree estimated using the dataset indicated along the row (with the best fitting evolutionary rate parameters indicated using ModelTest) and each ML topology supported by the dataset indicated down the column. Values in bold are significant at $p < 0.05$ and total numbers of significant values are given in parentheses next to each locus.

		CYTB	PLP	MBTN	LCHRNA1	Del02	TDel04	ODel05	PDel08	ODel10	LDel11	ODel12	GDel14	YDel15	Del17	Nuc loci	All loci
	CYTB (10)	3722	**68**	**282**	**95**	**132**	**141**	40	**227**	**80**	61	35	**102**	**211**	**198**	**89**	1
D	PLP (2)	30	1187	14	16	8	19	6	**46**	6	11	8	6	**30**	28	9	29
A	BTN (4)	63	18	1329	**59**	34	30	0	**75**	26	24	25	29	**63**	**60**	53	**59**
T	CHRNA1 (5)	23	23	**45**	562	18	30	17	**45**	**34**	12	18	0	**52**	**45**	23	23
A	Del02 (0)	21	0	21	16	1363	21	0	29	29	21	21	0	30	15	15	21
S	Del04 (0)	7	9	16	17	9	963	0	7	7	6	0	9	24	17	0	16
E	Del05 (0)	0	0	0	0	7	7	1089	0	0	0	0	0	0	0	0	0
T	Del08 (11)	**101**	**73**	**132**	**55**	**104**	**75**	3	1406	28	**88**	**62**	**57**	**106**	**132**	**62**	**91**
	Del10 (1)	22	3	22	15	12	19	0	24	625	12	15	10	24	**34**	11	22
	Del11 (12)	**71**	**71**	**86**	**71**	**71**	**94**	**71**	**186**	**71**	914	**71**	0	**106**	**96**	**42**	**71**
	Del12 (0)	14	3	14	7	9	19	11	29	15	22	1245	13	26	36	5	14
	Del14 (0)	8	0	0	0	0	8	0	8	0	0	8	447	8	8	6	8
	Del15 (8)	**105**	21	**81**	**71**	**59**	**83**	23	**89**	**48**	18	26	26	1266	**71**	**39**	**100**
	Del17 (2)	47	16	24	32	0	29	0	**42**	24	0	23	8	**38**	1096	23	47

The third method used was implemented in the software package *BEAST (Heled & Drummond 2010). Although *BEAST also estimates species trees under the multispecies coalescent, there are several modelling differences when compared to the method implemented in BEST. *BEAST coestimates the species tree and all gene trees simultaneously in just one Bayesian MCMC analysis, instead of the two steps required by BEST. Moreover, an outgroup is not required, population size can or not be assumed constant over the branches of the trees and two different species tree priors are available: Yule process and Birth-Death process (Heled & Drummond 2010). We ran 100 million MCMC generations sampling every 10,000 generations, choosing the Yule process as species tree prior and the Piecewise constant and linear model for population size estimates. The HKY model of nucleotide substitution was chosen for all loci with the exception of Del_10 and BTN, for which the GTR model was chosen. A relaxed molecular clock with an uncorrelated lognormal distribution was chosen. The program TRACER v.1.5 was run to ensure mixing and convergence of the posterior distribution and parameters by examining effective sample size (ESS) values.

TreeAnnotator v.1.6.1 (Rambaut & Drummond 2010) was subsequently used to summarize the obtained trees in a single tree that best represents the posterior distribution. The default options were used, with the exception of the node heights option, where the option “mean heights” was chosen.

4.4. RESULTS

4.4.1. *nuDNA - Separate analyses of each nuclear locus*

A total of 8516 bp of nuclear DNA was obtained for the 15 species analysed across 13 nuclear loci. Amplification of some loci was unsuccessful for some species (Del_04, Del_05 and Del_08 for *Lagenodelphis hosei*; Del_05 and Del_17 for *Phocoena phocoena* and Del_17 for *Globicephala melas*). Initial phylogenetic analyses were conducted considering each locus separately, since the partition homogeneity test identified a conflict of signal among the data partitions (P = 0.01), such a result was also observed in the crossed SH tests (Table 4.4). Fragment lengths obtained varied from 357 bp (CHRNA1) to 923 bp (Del_02). Levels of polymorphism obtained were low, with parsimony-informative sites varying from 1 to 17 (Table 4.5). MP and Bayesian trees obtained for each locus had low resolution and presented highly discordant genealogies (Supplementary Material, Figure S4.1, Bayesian trees only).

4.4.2. *mtDNA*

A 1120-bp fragment of the cytochrome *b* gene (hereafter Cyt*b*) was sequenced for the 15 species analysed in this study. The Bayesian tree obtained resulted in nearly the same topology as the MP tree. Here we present the Bayesian tree (Figure 4.3). This tree is similar to the one presented in previous studies (Möller *et al.* 2008 (Figure 4.1a; LeDuc *et al.* 1999). The genus *Delphinus* was the only genus that was rendered monophyletic. The genus *Stenella* was paraphyletic, with *S. coeruleoalba* and *S. frontalis* more closely related to *Delphinus* spp. than with its congeners. The genus *Tursiops* was also paraphyletic, with the southern Australian bottlenose dolphin (hereafter *Tursiops* sp.) clustering with *L. hosei*.

Table 4.5. Total number of sites analysed, base composition bias (A/T %), parsimony informative sites (PI), tree length, number of equally good parsimonious trees (MP trees), retention index (RI), consistency index (CI), rate heterogeneity (alpha) and proportion of invariable sites (Pi) for the Maximum Parsimony phylogenetic trees obtained for each single dataset and for combined datasets.

Data Partitions	Total sites	Variable Sites	A/T %	PI sites	Tree length	MP trees	RI	CI
Single dataset								
Del_02	923	25	69.6	6	23	204	1.000	1.000
Del_04	636	26	50	5	18	79947	1.000	1.000
Del_05	750	16	58.5	3	9	642	1.000	1.000
Del_08	806	41	70.6	17	39	2586	0.938	0.897
Del_10	402	15	64	6	14	3	0.900	0.929
Del_11	572	26	67.1	17	24	2	1.000	1.000
Del_12	729	40	46.1	7	33	21644	0.875	0.939
Del_14	318	2	62.3	1	1	3	1.000	1.000
Del_15	780	32	43.1	10	27	45	0.935	0.926
Del_17	739	21	53.7	6	11	30	1.000	1.000
BTN	754	44	58.2	13	40	31118	0.892	0.900
CHRNA1F	357	14	50	5	11	425	1.000	1.000
PLP	750		50	5	20	2	1.000	1.000
CYTB	1120		56	170	470	5	0.684	0.653
All nuclear								
Anloci + introns	8516			101	327	1	0.740	0.783
All data								
Anloci+introns+mtDNA	9636			271	817	4	0.678	0.689

4.4.3. *nuDNA (introns + anonymous)*

The concatenation of all nuDNA loci resulted in a total of 8516 bp. Bayesian and MP trees resulted in very similar topologies, with all but three branches in the Bayesian tree having posterior probabilities of 100% (Figure 4.4). The nuDNA phylogeny differed from the one obtained with mtDNA in the order of branching relationships, but the genera *Stenella* and *Tursiops* were still not monophyletic. southern Australian *Tursiops* sp. clustered with *T. aduncus* and not with *L. hosei* as it did in the mtDNA tree. *L. hosei* was more closely related with *S. coeruleoalba* and *Delphinus* spp. in this nuDNA phylogeny. The differences in taxon position between mtDNA and nuDNA trees are represented in the trees (Figures 4.3 and 4.4, respectively). The sister taxon of the species of *Delphinus* was not resolved.

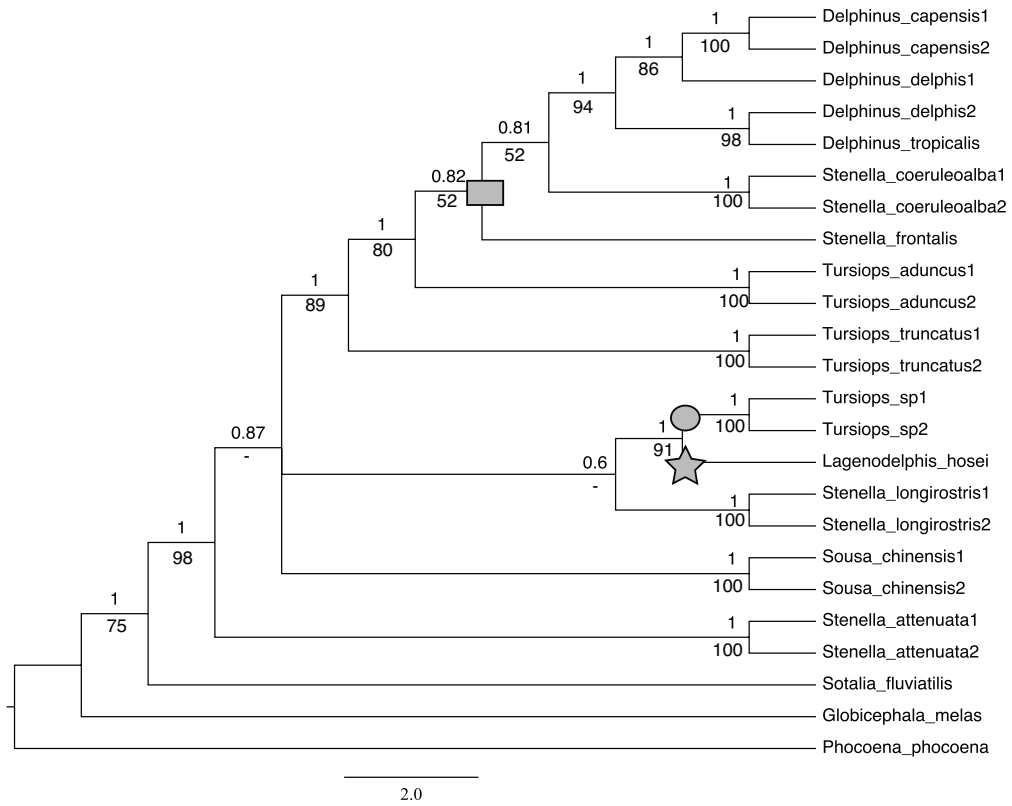


Figure 4.3. Majority-rule consensus tree generated in MrBayes for the cytochrome *b* gene. Posterior probability values are above nodes and bootstrap support values obtained in the Maximum-Parsimony analysis are below nodes, with (-) indicating lack of support. Symbols indicate clades that differ from the ones recovered in the nuDNA dataset (see text for further details). Branch lengths are in substitutions/site.

4.4.4. *mtDNA + nuDNA*

The concatenation of mtDNA and nuDNA sequences resulted in a total of 9636 bp. Bayesian and MP trees resulted in similar topologies, again with the Bayesian tree showing more resolved branches. The topology of this tree is more similar to the mtDNA tree than to the concatenated nuDNA tree, although some of the relationships are in a different order. This is explained by the higher proportion of parsimony-informative sites contributed by the mtDNA dataset (Table 4.5), thereby providing a stronger phylogenetic signal to the consensus tree. This can be seen clearly by the crossed SH tests, where the Cyt*b* locus is highly congruent with the ML topology of all loci combined (Table 4.4), yet significantly incongruent with the topology of all nuclear loci combined. In this case, the analysis of such a combined mtDNA and nuDNA tree will not elucidate the phylogenetic relationships of this group. Hence, the following analyses and discussion will be based on the comparison of the separate mtDNA and nuDNA trees.

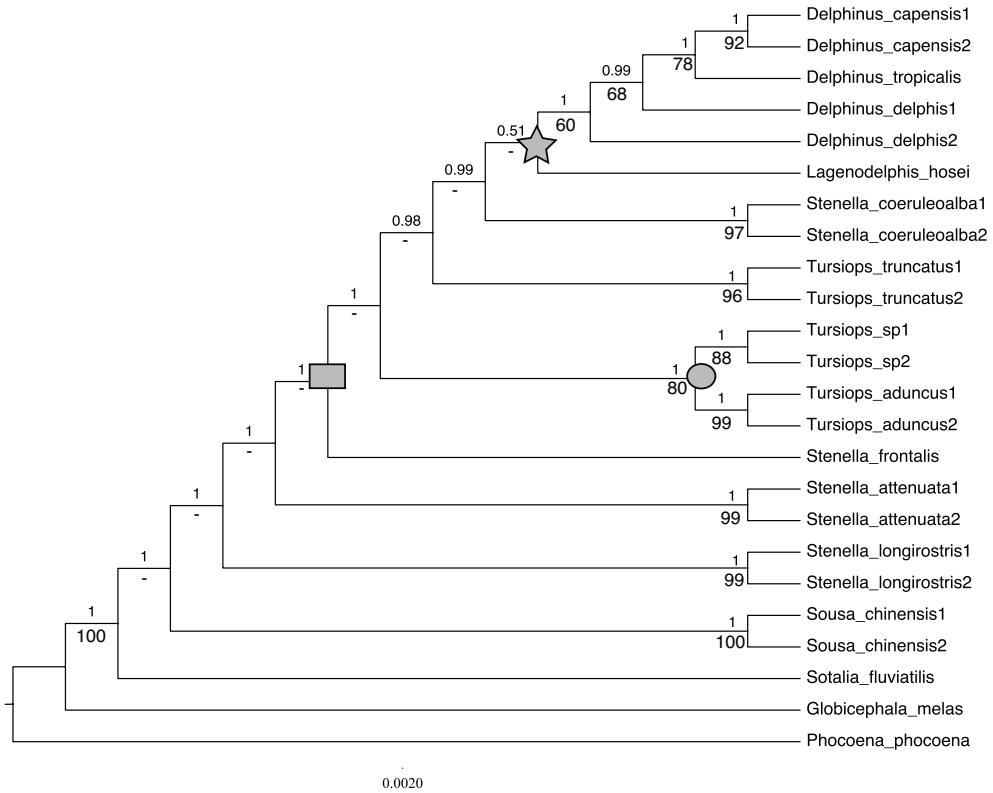


Figure 4.4. Majority-rule consensus tree generated in MrBayes from thirteen concatenated nuclear loci. Posterior probability values are above nodes and bootstrap support values obtained in the Maximum-Parsimony analysis are below nodes, with (-) indicating lack of support. Symbols indicate clades that differ from the ones recovered in the mtDNA dataset (see text for further details). Branch lengths are in substitutions/site.

4.4.5. *Species trees*

For the MDC analyses we decided to discard locus Del_14 since the very low number of variable sites available (two) provided insufficient information to estimate phylogeny. No individual loci rejected the evolutionary model estimated from the combined dataset, so this single evolutionary model (transition: transversion ratio = 4.4915, rates = equal, base frequencies from combined dataset) was applied to all loci in the BEST framework.

The MDC method for species tree estimation, as implemented in Mesquite, yielded 3 equally good species trees with score 94 when individual gene trees estimated with the MP method were used (Figure 4.5a, Table 4.6), and 4 equally good species trees with a score of 71 when Bayesian individual gene trees were used (Figure 4.5b, Table 4.6).

Table 4.6. Deep coalescence scores obtained for each locus in the analysis based on individual gene trees obtained with maximum parsimony (MP) and Bayesian Inference (BI) a) using the auto-resolve polytomies option and b) non auto-resolving polytomies.

a)Species trees	Nuclear Loci												Total score
	Del 02	Del 04	Del 05	Del 08	Del10	Del11	Del12	Del15	Del17	BTN	CHRNA1	PLP	
MP													
1	8	1	0	20	8	4	6	12	7	9	16	3	97
2	8	1	0	21	8	4	8	13	6	9	13	3	97
3	7	1	0	21	8	4	9	14	6	8	13	3	97
BI													
1	2	1	5	21	0	3	6	11	7	10	3	2	71
2	2	1	6	20	0	4	6	10	7	10	3	2	71
3	2	2	5	20	0	3	6	12	6	10	3	2	71
4	2	2	6	19	0	4	6	11	6	10	3	2	71
b) nonAR													
MP													
	58	53	58	49	63	54	62	36	47	45	60	49	634
BI													
	40	46	55	43	48	50	50	40	43	41	57	36	549

Topologies obtained were similar, differing only in the branching order within Clade A (Figures 4.5a and 4.5b), which includes the genus *Delphinus*, *S. coeruleoalba*, *S. longirostris* and *L. hosei.* In both trees the genus *Tursiops* is rendered monophyletic, clustering with *S. attenuata* and *S. frontalis* (Clade B). *Sousa chinensis* and *S. fluviatilis* occupy a basal position in both trees. Not resolving polytomies automatically resulted in one tree with score 549 when MP gene trees were used and one tree with score 634 when Bayesian gene trees were used. These trees differed from the ones obtained when polytomies were automatically resolved (Appendix 4B) and the MDC scores were much higher (Table 4.6). Inclusion of branch length information in the estimation of species tree had no effect on the topology of the tree. The species trees presented here were estimated using branch lengths since this is usually recommended in order for the fit to reflect the actual history.

The BEST analyses averaged over 10-12 runs did not achieve full convergence (effective sample sizes, ESS, higher than 100) for all parameters, suggesting that incomplete mixing was achieved for most analyses. However nearly all parameters in the coalescent tree prior analysis achieved convergence (with some mutation rate priors not converging), while neither mutation rates nor tree priors converged adequately in the exponential tree prior analysis (Appendix 4D).

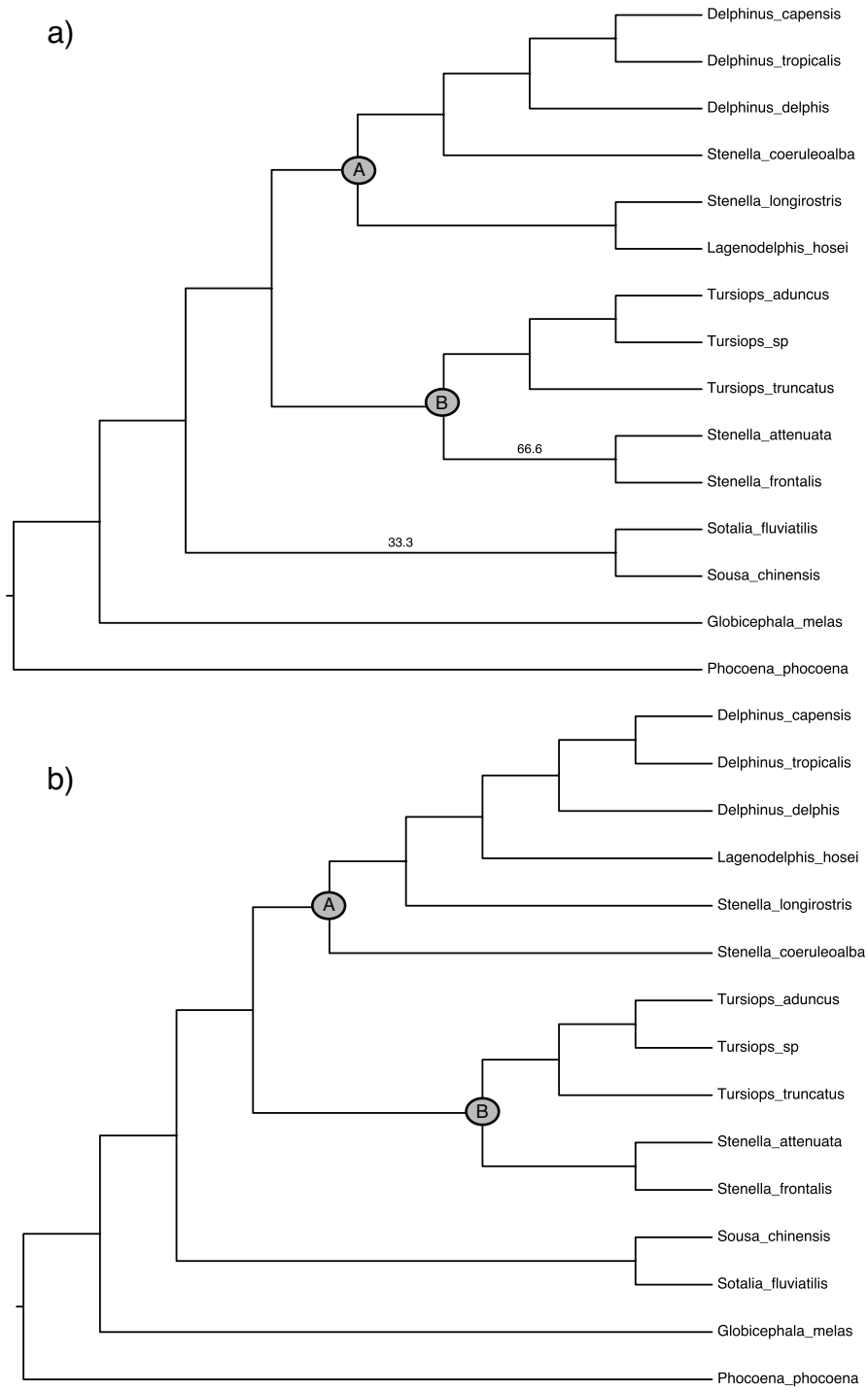


Figure 4.5. Species tree estimated with minimize deep coalescence method for a) individual trees obtained with Maximum-Parsimony and b) individual trees obtained with Bayesian Inference method. In 'a', number above node indicates the percentage of Minimizing Deep Coalescence trees that contained that clade. Clade labels A and B indicate clades that are discussed further in the text.

Choice of different population size, branch-length and range of mutation rate priors had some influence on the topology of the species trees estimated by BEST. The analyses that resulted in the highest levels of convergence of the

likelihood parameter (as given by ESS values, Appendix 4D) were the ones where the coalescent branch-length prior and a wider ranging mutation rate prior were used (Figure 4.6). Here, using different theta priors resulted in an identical species tree topology. Differences in the tree topology obtained with other priors relate mostly to the position of the genus *Tursiops* and *Lagenodelphis hosei* (results not shown). Some differences in the basal group species were also observed. Although support for most branches is quite low, these trees are very similar to the MDC species trees in that Clades A and B are also recovered. The only difference is that the spotted dolphins, *Stenella attenuata* and *S. frontalis*, cluster together (Clade C) but do not have a sister taxon relationship with *Tursiops* (Figure 4.6).

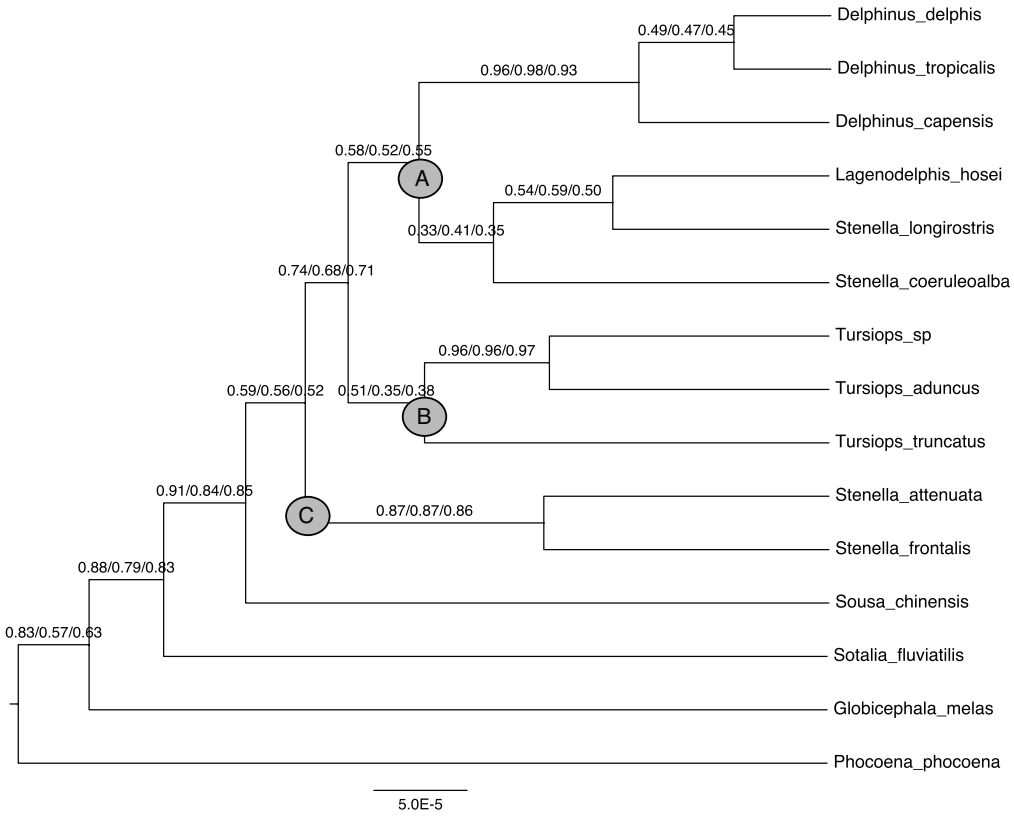


Figure 4.6. Species trees estimated with the BEST method using a coalescent branch-length prior, a wider mutation rate prior and population size priors of $\theta = 0.001$, $\theta = 0.0015$ and $\theta = 0.002$. Posterior probability values are above nodes and correspond to the trees obtained with the different theta values. Clade labels A and B indicate clades that are discussed further in the text.

The *BEAST analysis achieved convergence, with the posterior distribution and all parameters having ESS values higher than 100. The tree obtained has exactly the same topology as the tree obtained in BEST, but most branches are supported by

was very low. This may reflect the uncertainty in resolving taxa position within this clade since the tree obtained with BEST resulted in the same low support for these relationships. higher posterior probability values (Figure 4.7). Nevertheless, support for Clade A was very low. This may reflect the uncertainty in resolving taxa position within this clade since the tree obtained with BEST resulted in the same low support for these relationships.

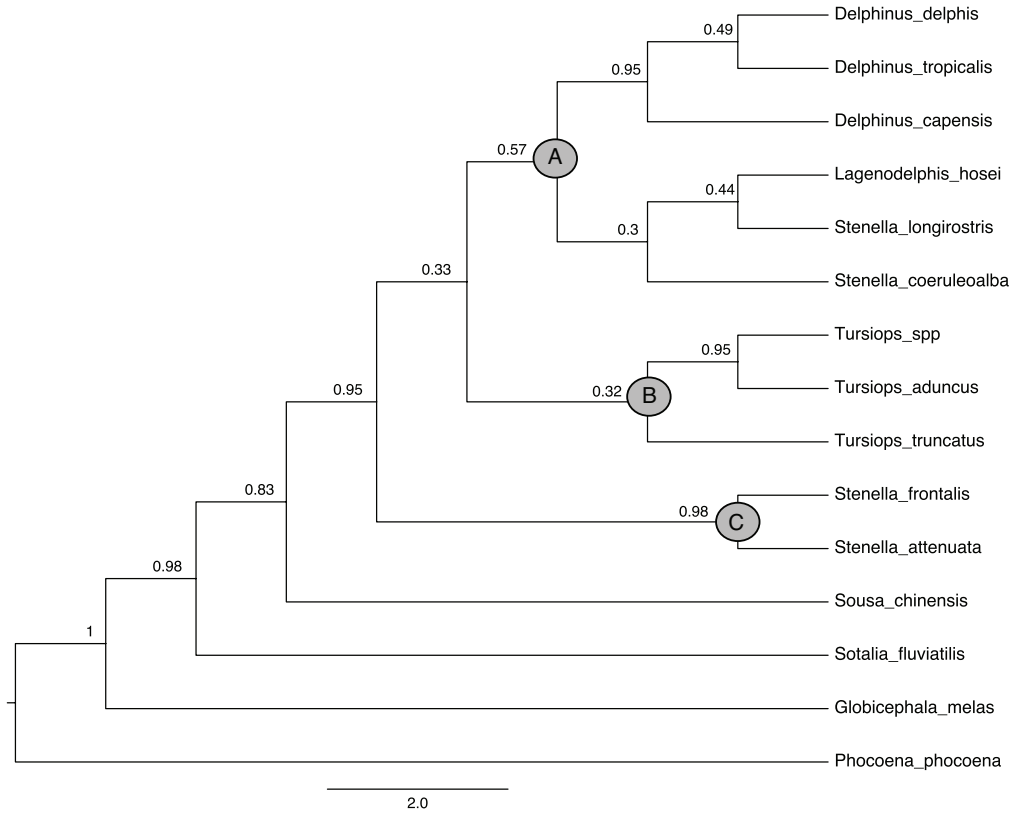


Figure 4.7. Species trees estimated with the *BEAST method. Posterior probability values are above nodes. Clade labels A, B and C indicate clades that are discussed further in the text.

4.5. DISCUSSION

This study is the first to use a species tree approach that accounts for gene tree heterogeneity, to infer phylogenetic relationships for the Delphininae using DNA sequences from several nuclear loci. Although individual gene trees were unresolved and highly incongruent, using coalescent-based methods we have been able to recover a species tree that supports morphology-based species relationships (Figure 4.1; Perrin *et al*. 1987; Perrin 2009). This finding has never been entirely recovered in previous molecular studies (LeDuc *et al*. 1999;

Caballero *et al*. 2008; Möller *et al*. 2008; Xiong *et al*. 2009; Kingston *et al*. 2009; Steeman *et al*. 2009; McGowen *et al*. 2009) and highlights the importance of using the coalescent model to estimate species trees in recent, likely rapid radiations. Of the analytical methods used, we consider the minimize deep coalescence method and *BEAST, to perform better in estimating the Delphininae species tree than BEST and the concatenation approach. Although the species trees estimated with BEST showed similar topologies to the MDC and *BEAST trees, we found this method to be highly susceptible to prior choice.

4.5.1. *Phylogenetic relationships*

We provide an example of how phylogenetic studies based on single gene trees may prove misleading, particularly in recent species radiations. The phylogeny obtained with mtDNA differed from the phylogeny obtained by concatenating all nuclear loci, and both differed from the species trees estimated by methods that account for gene tree heterogeneity. Although MDC and the Bayesian species trees obtained in *BEAST and BEST differed in the placement of *Tursiops*, the overall topology obtained is in agreement with morphology-based relationships (Figure 4.1). Therefore it is likely these species trees reflect the true evolutionary history of the Delphininae. Uncertainty in resolving positioning of taxa within Clade A, together with the short branch lengths obtained, reflects the early history of these species. It further suggests that they have recently and rapidly radiated.

Most of the controversy surrounding the taxonomy of members of the subfamily Delphininae has arisen from the disagreement between the taxonomy originally established by morphological characters (e.g. Flower 1883), and the phylogenetic relationships subsequently supported by molecular studies (e.g. LeDuc *et al*. 1999). Such studies, however, were based on mtDNA, single-locus phylogenies or AFLPs (Kingston *et al*. 2009) and likely recovering an incomplete phylogeny.

4.5.1.1. *Genus* Tursiops

The two species presently included in the genus *Tursiops* share several morphological similarities, with a short beak distinctly marked off from the prenarial adipose elevation and less numerous and larger teeth that distinguish

them from other Delphininae genera (Flower 1883; Perrin *et al.* 2007; True 1889; Wang *et al.* 2000). However, all molecular studies conducted using mtDNA and AFLPs have recovered the genus as polyphyletic (e.g. Kingston *et al*. 2009; LeDuc *et al*. 1999). The species trees obtained in this study with the coalescent-based methods have all recovered the genus as monophyletic, supporting the results obtained in two recent studies that used a supermatrix approach (McGowen *et al*. 2009; Steeman *et al*. 2009). The recently proposed new species of bottlenose dolphin (*Tursiops* sp.) from southern Australia (Möller *et al*. 2008) clustered with *T. aduncus* in all analyses that included the nuDNA dataset. This relationship is strongly discordant with the mtDNA tree but supports the revision of its taxonomic status. This discordance between nuDNA and mtDNA should be further explored to clarify, for instance, whether the new species may have arisen through a process of hybridization.

4.5.1.2*. Genus* Stenella

The spotted dolphins, *S. attenuata* and *S. frontalis*, are morphologically very similar both in coloration and in skull characters (Perrin *et al.* 1987), but most molecular phylogenetic studies conducted to date have failed to cluster them (LeDuc *et al*. 1999; this study, Figure 4.1). However, all the species trees obtained in this study have recovered a sister relationship between these two taxa, supporting a recent phylogenetic study based on AFLP markers (Kingston *et al.* 2009). More importantly, the MDC species trees clustered these two species with the genus *Tursiops*, a relationship that has never before been recovered in a molecular study. Interestingly, the two studies referred to above, and that used a supermatrix approach, failed to recover this relationship, as did the phylogenetic tree resulting from the concatenation approach presented in this study (Figure 4.4).

The genus *Stenella* is rendered polyphyletic in all analyses, supporting previous molecular phylogenetic studies (LeDuc *et al*. 1999; Caballero *et al*. 2008; Xiong *et al*. 2009, Kingston *et al*. 2009; McGowen *et al*. 2009; Steeman *et al*. 2009) and indicating that this group needs considerable taxonomic revision. Species within this genus are both morphologically and genetically very dissimilar (LeDuc *et al.*

1999; Perrin 1997), suggesting that this is an artificial assemblage. As can be seen in all trees, *S. coeruleoalba* is more closely related with the genus *Delphinus* than with its congeners, and *S. frontalis* and *S. attenuata* form a different, divergent group from *S. longirostris*.

4.5.1.3. *Genera* Delphinus, Lagenodelphis *and* Sousa

The genus *Delphinus* is rendered monophyletic in all analyses. However, in the mtDNA and nuDNA phylogenies, the position of the *tropicalis* form varied. In the species tree framework, all individuals from one species are "forced" to be monophyletic by the structure of the data input, so these relationships could not be clarified. The sister taxon affinities of this genus could also not be elucidated. The skulls of *L. hosei* and *S. coeruleoalba* show a strong resemblance with that of *Delphinus* spp. with regards to the presence of deep palatal grooves, a derived characteristic that no other Delphinid species hold ((Dolar 2009); personal communication, W. F. Perrin). The skull of *S. coeruleoalba* shares additional similarities with that of species of *Delphinus* (Amaral *et al.* 2009). In fact, most phylogenetic trees obtained in this study, including mtDNA and nuDNA trees, place *S. coeruleoalba* as the sister taxon of *Delphinus*.

The position of *Sousa chinensis* varied between the phylogenetic analyses but its inclusion in the subfamily Delphininae is supported, as suggested by other molecular phylogenies (LeDuc *et al*.1999; Caballero *et al*. 2008).

4.5.2. *Comparison of methods*

The different species tree methods used in this study resulted in somewhat different topologies. The tree obtained with the concatenation approach (Figure 4.4) differed from the tree obtained with the coalescent-based methods that take into account gene tree heterogeneity, despite having highly supported branches. It has been suggested that the statistical advantage conferred by increasing sample size (number of sites) may result in a presumed improvement in phylogenetic accuracy and branch support (Gadagkar *et al.* 2005). This possibly explains the fact that the two studies using a supermatrix approach (McGowen *et al*. 2009; Steeman *et al*. 2009) recovered the genus *Tursiops* as monophyletic since they

used a considerably higher number of sites than the present study. However, the failure of such approaches to explicitly model relationships between gene trees and species trees will likely result in an incorrect phylogeny estimate (Degnan *et al.* 2006; Kubatko & Degnan 2007; Degnan & Rosenberg 2009). This may explain why the coalescent-based methods used in this study resulted in phylogenies that recovered most of the relationships supported by morphology, as opposed to those recovered by the supermatrix approach mentioned above.

The trees obtained with the MDC method (BI and MP, automatically resolving polytomies) are the ones that more strongly agree with morphology as referred before. This difference between a method based on summary statistics and parameter-rich Bayesian probabilistic model, as implemented in BEST and *BEAST, is likely explained by the characteristics of the dataset used, particularly the low number of variable sites obtained. This lack of variability caused problems with parameter convergence in the BEST analysis, which in turn was reflected in the low support obtained for most branches. Similar results were obtained in a recent study of the genus *Oriza,* which was also characterized by low levels of sequence divergence (Cranston *et al.* 2009). Although the analysis implemented in *BEAST achieved convergence much better than the analysis implemented in BEST, the resulting tree still had some branches with low support values. In contrast, MDC gene trees are obtained using maximum parsimony (Maddison and Knowles 2006), which performs optimally under conditions of relatively low sequence divergence, since it cannot account for unobserved substitutions (e.g. Steel & Penny 2000).

Additionally, the multispecies coalescent model implemented in BEST and *BEAST assumes that incomplete lineage sorting (deep coalescence) is the only evolutionary process causing the incongruence between gene trees. However, it is possible that hybridization is also playing a role in the evolutionary history of Delphininae, which may compromise the performance of these methods by altering gene tree branch-lengths, which in turn will restrict the corresponding speciation times and mislead the species tree estimation (Liu & Pearl 2007). Although hybridization can also compromise the performance of the MDC method,

it has been suggested that this method is more robust to the presence of gene flow as long as it is not the major force driving the evolutionary history of the species (Maddison & Knowles 2006; Liu *et al*. 2009). However, this method still has some caveats. Firstly, there were differences in the estimated trees when polytomies were not automatically resolved (although MDC scores were substantially worse), which suggests that the method still has difficulties in handling uncertainties in individual gene trees. Secondly, this method is unable to provide a measure of support for the relationships.

Another factor that may be compromising the performance of BEST is its sensitivity to prior choice in estimating the species-tree. We found that using different branch-length, range of mutation rates and theta priors has highly influenced the resulting trees. Although we have managed to find the "combination" of priors that lead to a more robust species tree estimation, where theta values no longer influenced tree topology, this process is time-consuming and does not guarantee that the best species tree is estimated. Although some studies have been quite robust to prior choice (Liu & Pearl 2007; Brumfield *et al*. 2008), others have also found BEST to be sensitive to the choice of theta for estimating species tree in their dataset (Linnen & Farrel 2008). We therefore suggest that a precautionary approach should be taken when using BEST and a thorough exploration of the priors choices, as these can be highly dependent on the dataset used. Although *BEAST may also be sensitive to prior choice, it is computational not so demanding and time consuming as BEST and seems to achieve convergence faster.

Finally, sampling may have also influenced the differences obtained with the methods used. It has been shown that the species tree methods used in this study can in fact be sensitive to sampling schemes (e.g. Linnen & Farrel 2008). Several species within Delphinidae present cosmopolitan distributions and higher intraspecific than interspecific genetic variability (e.g. Amaral *et al.* 2007b; Forcada 2009). It is possible that having sampled individuals from different geographical locations or even more individuals per species could have resulted in different tree topologies. The sensitivity of the methods to different sampling

schemes could not be evaluated in this study due to the difficulty of accessing samples, but the fact that each method resulted in a different topology indicates that such sensitivity may exist.

4.5.3. *Incomplete lineage sorting, hybridization, or both*

Incongruence between mtDNA and nuDNA phylogenies has been described in other animal groups (Goncalves *et al.* 2007; McCracken & Sorenson 2005; Peters *et al.* 2007; Shaw 2002) and may be due to incomplete lineage sorting, hybridization, or both. Gene tree heterogeneity is common in cases of rapid speciation such as the one that has likely given rise to the Delphininae. Incomplete lineage sorting will make the genealogical histories of individual gene loci appear misleading or uninformative about the relationships among species due to retention and stochastic sorting of ancestral polymorphism (Pamilo & Nei 1988). However, a genetic polymorphism shared among lineages can also result from a gene copy introduced to the population via gene flow if the lineages exchange members, which can be particularly common if they occur in sympatry. It is often very difficult to distinguish between these two processes, and methods that estimate species trees taking into account both the presence of incomplete lineage sorting and hybridization are still in their infancy (Kubatko 2009). Our multiple, independent loci suggest that a rapid series of divergences, characterized by short internodes, occurred during the early stages of diversification of the Delphininae (Figure S4.1), which suggests that incomplete lineage sorting is affecting the inference of phylogenetic relationships in this group. This was also clear in other molecular studies (Amaral *et al.* 2007b), where the failure to recover monophyletic groups was attributed to that process. However, using phylogenetic methods that account for this process has not yielded a fully resolved species phylogeny. One of the reasons may be that other factors such as hybridization are affecting species history, thereby further confounding the inference of phylogenetic relationships. Hybridization in cetaceans has been reported to occur both in captivity and in the wild (Bérubé 2002), and could be in fact more common than previously thought.

4.6. CONCLUSION

This study illustrates the complexity of inferring phylogenetic relationships in a group likely subject to a rapid radiation (e.g. the Delphininae), where incomplete lineage sorting, possibly coupled with hybridization events, is confounding the species history. By comparing three different coalescent-based methods to infer species trees, we were able to conclude that the MDC method seems to perform better in cases of low sequence divergence and highly incongruent gene trees when compared to the Bayesian probabilistic models implemented in BEST and *BEAST. Between the two Bayesian approaches, *BEAST seems to perform better than BEST in achieving convergence and higher support values for branches. These results should, nevertheless, be confirmed by simulations that compare the performance of the methods. In the case of the Delphininae, we were able to extract an underlying species-tree signal from divergent histories of independent genes using these methods and, for the first time, provide molecular support for relationships supported by morphology. Our study thus shows that the use of multiple loci is likely to result in a more realistic depiction of lineage history than the use of one or a few loci, particularly if analysed in a coalescent context. Our results also emphasize the need for coalescent methods that can be applied at the interface of phylogenetic and population processes and that account for both recent rapid speciation events and gene flow between lineages. Finally, this study illustrates how methods for species-tree inference can be dependent on the dataset and on the biological processes dictating the evolution of a particular group.

4.7. ACKNOWLEDGEMENTS

The authors would like to thank Kelly Robertson (SWFSC) and Marina Sequeira (ICNB) for providing tissue samples. We also thank W. F. Perrin, E. Jockusch and two anonymous reviewers for comments on the manuscript. Research was supported by Macquarie University. A. R. Amaral acknowledges a PhD grant (SFR/BD/27245/2006) from Fundação para a Ciência e Tecnologia (Portugal). Part of this work was carried out by using the resources of the Computational

Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation.

4.8. REFERENCES

Amaral AR, Coelho MM, Marugan-Lobon J, Rohlf FJ (2009) Cranial shape differentiation in three closely related delphinid cetacean species: Insights into evolutionary history. *Zoology* **112**, 38-47.

Amaral AR, Sequeira M, Cedeira-Martínez J, Coelho MM (2007a) New insights on population genetic structure of *Delphinus delphis* from the Northeast Atlantic and phylogenetic relationships within the genus inferred from two mitochondrial markers. *Marine Biology* **151**, 1967-1976.

Amaral AR, Sequeira M, Coelho MM (2007b) A first approach to the usefulness of COI barcodes in the identification of closely related delphinid cetacean species. *Marine and Freshwater Research* **58**, 505-510.

Amaral AR, Silva MC, Möller LM, Beheregaray LB, Coelho MM (2010) Anonymous nuclear markers for cetacean species. *Conservation Genetics* **11**, 1143-1146.

Barnes LG, Domning DP, Ray CE (1985) Status of studies on fossil marine mammals. *Marine Mammal Science* **1**, 15-53.

Belfiore NM, Liu L, Moritz C (2008) Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). *Systematic Biology* **57**, 294-310.

Bérubé M (2002) Hybridism. In: *Encyclopedia of marine mammals* (eds. Perrin WF, Wursig B, Theweissen JGM), pp. 596-600. Academic Press Inc., San Diego.

Brumfield RT, Liu L, Lum DE, Edwards SV (2008) Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae, *Manacus*) from multilocus sequence data. *Systematic Biology* **57**, 719-731.

Caballero S, Jackson J, Mignucci-Giannoni AA, Barrios-Garrido H, Beltran-Pedreros S, Montiel-Villalobos MG, Robertson KM, Baker CS (2008) Molecular systematics of South American dolphins *Sotalia*: Sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the

Delphinidae. *Molecular Phylogenetics and Evolution* **46**, 252-268.

Cranston KA, Hurwitz B, Ware D, Stein L, Wing RA (2009) Species trees from highly incongruent gene trees in Rice. *Systematic Biology* **58**, 489-500.

de Queiroz A, Gatesy J (2007) The supermatrix approach to systematics. *Trends in Ecology & Evolution* **22**, 34-41.

Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *Plos Genetics* **2**, 762-768.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* **24**, 332-340.

Delsuc F, Scally M, Madsen O*,* Stanhope MJ, de Jong WW, Catzeflis FM, Springer MS, Douzery EJP (2002) Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Molecular Biology and Evolution* **19**, 1656-1671.

Dolar MLL (2009) Fraser's Dolphin *Lagenodelphis hosei*. In: *Encyclopdia of Marine Mammals* (eds. Perrin WF, Wursig B, Thewissen JGM), pp. 469-472. Academic Press, New York.

Dolman G, Hugall AF (2008) Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). *Molecular Phylogenetics and Evolution* **49**, 782-794.

Flower WH (1883) On the characters and divisions of the family *Delphinidae*. *Proceedings of the Zoological Society of London*, 466-513.

Gadagkar SR, Rosenberg MS, Kumar S (2005) Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* **304B**, 64-74.

Goncalves H, Martinez-Solano I, Ferrand N, Garcia-Paris M (2007) Conflicting phylogenetic signal of nuclear vs mitochondrial DNA markers in midwife toads (Anura, Discoglossidae, *Alytes*): Deep coalescence or ancestral hybridization? *Molecular Phylogenetics and Evolution* **44**, 494-500.

Gygax L (2002) Evolution of group size in the superfamily Delphinoidea (Delphinidae, Phocoenidae and Monodontidae): A quantitative comparative analysis. *Mammal Review* **32**, 295-314.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**, 570-580.

Hennig, W. 1966. Phylogenetic Systematics. Urbana-Champaign (IL): Illinois Press, Illinois.

Hudson RR (1992) Gene trees, species tree and the segregation of ancestral alleles. *Genetics* **131**, 509-512.

Huelsenbeck JP, Bull JJ, Cunningham CW (1996) Combining data in phylogenetic analysis. *Trends in Ecology & Evolution* **11**, 152-158.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Hurvich CM, Tsai CL (1995) Model selection for extended quasi-likelihood models in small samples. *Biometrics* **51**, 1077-1084.

Jefferson TA, Van Waerebeek K (2002) The taxonomic status of the nominal dolphin species *Delphinus tropicalis* van bree, 1971. *Marine Mammal Science* **18**, 787-818.

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.

Knowles LL (2009) Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Systematic Biology* **58**, 463-467.

Kubatko LS (2009) Identifying hybridization events in the presence of coalescence via model selection. *Systematic Biology* **58**, 478-488.

Kubatko LS, Carstens BC, Knowles LL (2009) STEM: Species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* **25**, 971-973.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* **56**, 17-24.

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.

Linnen C, Farrell B (2008) Comparison of methods for species-tree inference in the sawfly genus *Neodiprion* (Hymenoptera: Diprionidae). *Systematic Biology* **57**, 876-890.

Lipps JH, Mitchell E (1976) Model for the adaptive radiations and extinctions of pelagic marine mammals. *Paleobiology* **2**, 147-155.

Liu L (2008) BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* **24**, 2542-2543.

Liu L, Pearl DK (2007) Species trees from gene trees: Reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology* **56**, 504-514.

Liu L, Pearl DK, Brumfield RT, Edwards SV (2008) Estimating species trees using multiple-allele DNA sequence data. *Evolution* **62**, 2080-2091.

Liu L, Yu LL, Kubatko L, Pearl DK, Edwards SV (2009) Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution* **53**, 320-328.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology* **46**, 523-536.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55**, 21-30.

Maddion WP, Maddison DR (2009) Mesquite: a modular system for evolutionary analysis. Version 2.72 http://mesquiteproject.org.

May-Collado L, Agnarsson I (2006) Cytochrome b and Bayesian inference of whale phylogeny. *Molecular Phylogenetics and Evolution* **38**, 344-354.

McCracken KG, Sorenson MD (2005) Is homoplasy or lineage sorting the source

of incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (*Nomonyx-Oxyura*)? *Systematic Biology* **54**, 35-55.

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

Mead JG (1975) Anatomy of the external nasal passages and facial complex in the Delphinidae (Mammalia: Cetacea). *Smithsonian Contributions in Zoology* **207**, 1-72.

Möller LM, Kerstin B, Charlton-Robb K, Beheregaray L (2008) Multi-gene evidence for a new bottlenose dolphin species in southern Australia. *Molecular Phylogenetics and Evolution* **49**, 674-681.

Muizon Cd (1988) Les relations phylogénétiques des Delphinida (Cetacea, Mammalia). *Annales de Paléontologie* **74**, 159-227.

Murphy S, Herman JS, Pierce GJ, Rogan E, Kitchener AC (2006) Taxonomic status and geographical cranial variation of common dolphins (*Delphinus*) in the eastern North Atlantic. *Marine Mammal Science* **22**, 573-599.

Natoli A, Peddemors VM, Hoelzel AR (2004) Population structure and speciation in the genus *Tursiops* based on microsatellite and mitochondrial DNA analyses. *Journal of Evolutionary Biology* **17**, 363-375.

Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred from gene genealogies. *Molecular Ecology* **17**, 2258-2275.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**, 568-583.

Perrin WF (1997) Development and homologies of head stripes in the delphinoid cetaceans. *Marine Mammal Science* **13**, 1-43.

Perrin WF (2009) Common Dolphins, *Delphinus delphis* and *D. capensis*. In: *Encyclopedia of Marine Mammals* (eds. Perrin WF, Wursig B, Thewissen JGM), pp. 255-259. Academic Press, New York.

Perrin WF, Mitchell ED, Mead JG, Caldwell DK, Caldwell MC, Vanbree PJH, Dawbin WH (1987) Revision of the spotted dolphins, *Stenella* Spp. *Marine Mammal Science* **3**, 99-170.

Perrin WF, Robertson KM, Van Bree PJH, Mead JG (2007) Cranial description and genetic identity of the holotype specimen of *Tursiops aduncus* (Ehrenberg, 1832). *Marine Mammal Science* **23**, 343-357.

Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). *Evolution* **61**, 1992-2006.

Posada D, Buckley T (2004) Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**, 793-808.

Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.

Quesada H, Posada D, Caballero A, Moran P, Rolan-Alvarez E (2007) Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* **61**, 1600-1612.

Rambaut A, Drummond AJ (2007) Tracer v.1.5, Available from http://beast.bio.ed.ac.uk/Tracer.

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

Rokas A, Carroll SB (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* **22**, 1337-1344.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Savolainen V, Anstett MC, Lexer C, Hutton I, Clarkson JJ, Norup MV, Powell MP,

Springate D, Salamin N, Baker WJ (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**, 198-207.

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16122-16127.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**, 1114-1116.

Slater GJ, Price SA, Santini F, Alfaro ME (2010) Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B-Biological Sciences* **277**, 3097-3104.

Steel, M. and Penny, D. 2000. Parsimony, likelihood and the role of models in molecular phylogenetics. *Molecular Biology and Evolution* **17**, 839-850.

Steeman ME, Hebsgaard MB, Fordyce RE, Ho SYW, Rabosky DL, Nielsen R, Rahbek C, Glenner H, Sorensen MV, Willerslev E (2009) Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology* **58**, 573-585.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73**, 1162-1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978-989.

True FW (1889) Contributions to the natural history of the cetaceans: A review in the family Delphinidae. *Bulletin of the United States National Museum* **36**, 191.

Wang JY, Chou LS, White BN (2000) Osteological differences between two sympatric forms of bottlenose dolphins (genus *Tursiops*) in Chinese waters.

Journal of Zoology **252**, 147-162.

Wiens JJ, Engstrom TN, Chippindale PT (2006) Rapid diversification, incomplete isolation, and the "speciation clock" in North American salamanders (Genus *Plethodon*): Testing the hybrid swarm hypothesis of rapid radiation. *Evolution* **60**, 2585-2603.

Xiong Y, Brandley MC, Xu SX, Zhou KY, Yang G (2009) Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. *Bmc Evolutionary Biology* **9**, 13.

4.9. APPENDIXES

Appendix 4A. Maximum Parsimony Analysis

Materials and Methods

PAUP* 4.01b (Swofford 2003) was used to obtain maximum parsimony (MP) trees for each locus, by heuristic search of 100 random addition analyses with tree bisection-reconnection (TBR) branch swapping. The support for each node on a strict consensus tree was evaluated by nonparametric bootstrap (1000 replications; heuristic analysis of 10 random addition analyses with TBR branch swapping). The consistency index (CI), the retention index (RI), the number of parsimony informative sites and the number of equally parsimonious trees obtained for each locus were also calculated in PAUP*. Tests for horizontal gene transfer (recombination) were also performed for each locus, using the *Maximum Chisquare* test of Smith (1992).

Incongruence between the smallest data partitions (i.e. between loci) was assessed with the partition homogeneity test (Farris *et al*. 1994; 1995) as implemented in PAUP*. In order to assess the relative contribution of each locus to the simultaneous analysis results, Partitioned Bremer Support (PBS, Baker & DeSalle 1997; Bremer 1994) was calculated using TreeRot v.3 (Sorenson & Franzosa 2007), where the PBS value for a given node is obtained by subtracting the number of steps for the partition of the simultaneous analysis tree from the number of steps for a partition on a tree constrained to not contain the node of interest. A positive PBS indicates support for the node by the data partition, whereas a negative PBS indicates that evidence in the data partition is inconsistent with that node. PBS values were then standardized for each locus by dividing each value by the minimum number of steps contributed by that locus (Baker *et al*. 2001). This controls for differences in size in data partitions (locus). This measure (PBS/min steps) provides a quantitative measure of the overall contribution of each locus to tree resolution.

Results

Separate analysis for each nuclear locus resulted in a high number of equally parsimonious trees. CI values varied from 0.897 to 1.000, indicating low levels of homoplasy, and RI values varied from 0.875 to 1.000, indicating a high number of informative shared-characters (synapomorphies) concentrated on internal nodes (Table 4.5).

For the cytochrome *b* dataset, there were 170 parsimony-informative characters. Five equally parsimonious trees were recovered with a length of 470. The CI was 0.653, revealing some level of homoplasy, and RI was 0.684, indicating that the phylogenetic signal is concentrated along the terminal branches of the tree (Table 4.5). Bootstrap support values were > 50% for all branches. The shape parameter of the gamma distribution (α) was 0.23, indicating strong heterogeneity of substitution rates across sites.

For the concatenated nuclear loci dataset (8516 bp in total), 101 sites were parsimony-informative. A single MP phylogenetic tree was obtained with a length of 327. The CI was 0.783, indicating a lower level of homoplasy when compared with the mtDNA tree. The RI was 0.740, indicating that the phylogenetic signal is concentrated along the terminal branches (Table 4.5). The relative contribution of each locus to the MP consensus tree was assessed by calculating Partitioned Bremer Support (PBS) divided by the minimum number of steps for each partition (locus) in a parsimony framework (Table 4A.1). The highest PBS/minimum steps values were obtained for CHRNA1, PLP, Del_12, Del_14 and Del_15, thereby reflecting their higher relative support for the combined nuDNA tree topology. On the other hand, locus Del_11 seemed to show a major discrepancy from the rest of the dataset; this dataset was a significantly poorer fit to the ML topologies obtained from all other loci except Del_14, (Table 4.4). This is likely due to the fact that there is a 19-bp indel in the alignment that separates *S. coeruleoalba* from the other taxa. This results in the MP tree placing *S. coeruleoalba* as a basal group, which is in disagreement with trees obtained with the other loci.

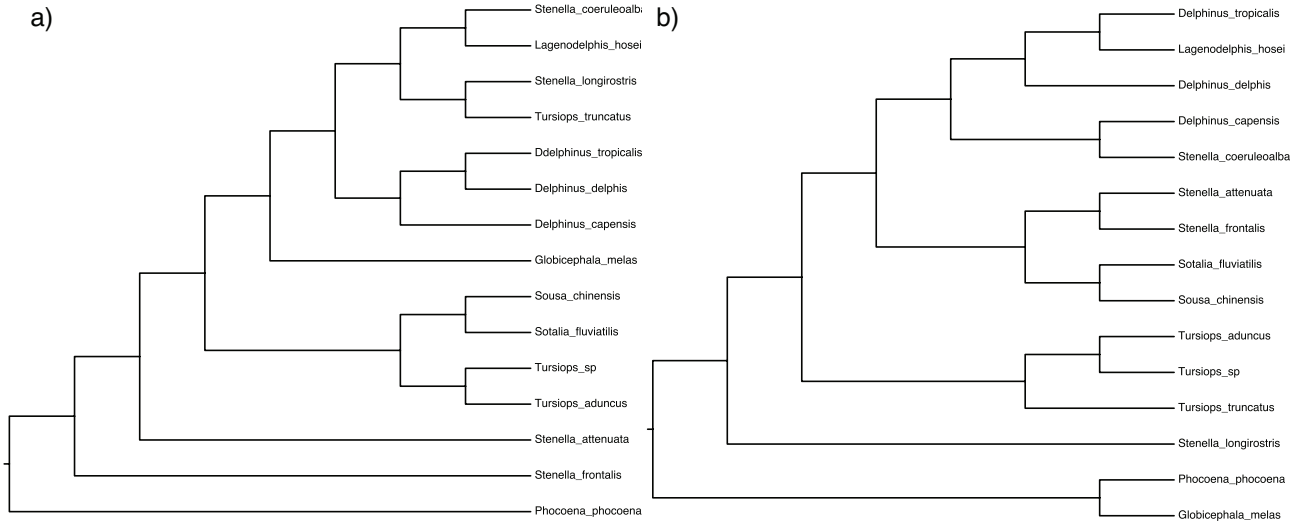
The concatenation of the mtDNA and nuclear loci resulted in a total of 9636 bp, with 271 characters being parsimony-informative. Four equally parsimonious trees

were obtained with a length of 817. A CI value of 0.689 was obtained, indicating a higher level of homoplasy than for the combined nuDNA tree. A RI value of 0.678 indicates that the phylogenetic signal was concentrated along the terminal branches of the tree (Table 4.5).

Table 4A1. Partitioned branch support (PBS) values for each data partition. These values were summed across all the nodes on the combined analysis tree and standardized by the minimum possible number of steps for each partition.

Data Partitions	Summed PBS	Min steps	PBS/min steps
Anon Loci+Introns			
Del_02	0.200	26	0.008
Del_04	0.400	10	0.040
Del_05	0.310	9	0.034
Del_08	5.940	52	0.114
Del_10	4.600	16	0.288
Del_11	-18.000	38	-0.474
Del_12	12.840	34	0.378
Del_14	1.000	2	0.500
Del_15	12.850	34	0.378
Del_17	3.120	14	0.223
BTN	15.090	44	0.343
CHRNA1F	14.120	17	0.831
PLP	11.530	21	0.549

Appendix 4B. Species tree estimated with minimize deep coalescence method, using the non-autoresolve polytomies option, for a) individual trees obtained with Maximum-Parsimony and b) individual trees obtained with Bayesian Inference method.



Appendix 4C. Classification of the subfamily Delphininae based on Rice (1998) and Perrin *et al.* (2010).

Classification of the subfamily Delphininae	
Genera	
Delphinus	
	D. delphis delphis Linnaeus, 1758
	D. delphis pontincus Barabash, 1935
	D. capensis capensis Gray, 1828
	D. capensis tropicalis van Bree, 1971
Tursiops	
	T. truncatus truncatus (Montagu, 1821)
	T. truncatus ponticus Barabash-Nikiforov, 1940
	T. aduncus (Ehrenberg, 1833)
	Tursiops sp. from southern Australia (Möller *et al*. 2008)
	Tursiops sp. from Southeast Asia (Natoli *et al*. 2004)
Stenella	
	S. attenuatagraffmani (Lonnberg, 1934)
	S. attenuata attenuata (Gray, 1846)
	S. longirostris longirostris (Gray, 1928)
	S. longirostris orientalis Perrin, 1990
	S. longirostris centroamericana Perrin, 1990
	S. longirostris roseiventris (Wagner, 1846)
	S. clymene (Gray, 1850)
	S. coeruleoalba (Meyen, 1833)
	S. frontalis (G. Cuvier, 1829)
Lagenodelphis	
	L. hosei Fraser, 1956
Sousa	
	S. chinensis (Osbeck, 1765)
	S. teuszii (Kukenthai, 1892
Sotalia	
	S. fluviatilis (Gervais and Deville, 1853)
	S. guianensis (Van Bénéden 1864)

Appendix 4D. Summary statistics, including effective sample size, obtained with the BEST method for the different priors used.

Coalescent theta = 0.002	**LnL**	**LnJointGenePr**	**TL(1)**	**TL(2)**	**TL(3)**	**TL(4)**	**TL(5)**	**TL(6)**	**TL(7)**	**TL(8)**	**TL(9)**	**TL(10)**	**TL(11)**	**TL(12)**	**TL(13)**	**kappa(all)**
mean	-14012.4085	1991.0621	7.99E-03	4.77E-03	4.61E-03	1.17E-02	6.05E-03	9.99E-03	1.01E-02	4.21E-03	8.87E-03	5.00E-03	1.13E-02	5.42E-03	9.57E-03	4.4767
stderr of mean	1.9106	1.7127	1.39E-04	8.23E-05	7.64E-05	1.01E-04	1.38E-04	1.71E-04	1.03E-04	4.30E-05	8.29E-05	9.03E-05	1.28E-04	9.10E-05	2.14E-03	
median	-14011.989	1991.608	8.00E-03	5.00E-03	4.00E-03	1.20E-02	6.00E-03	1.00E-02	1.00E-02	4.00E-03	9.00E-03	5.00E-03	1.10E-02	5.00E-03	9.00E-03	4.4307
geometric mean	n/a	1990.8517	7.83E-03	4.65E-03	4.51E-03	1.16E-02	5.87E-03	9.79E-03	9.93E-03	4.15E-03	8.72E-03	4.88E-03	1.12E-02	5.26E-03	9.45E-03	4.4372
95% HPD lower	-14061.676	1933.795	5.00E-03	3.00E-03	3.00E-03	8.00E-03	4.00E-03	6.00E-03	7.00E-03	3.00E-03	6.00E-03	3.00E-03	8.00E-03	4.00E-03	7.00E-03	3.3428
95% HPD upper	-13963.918	2047.335	1.10E-02	7.00E-03	6.00E-03	1.40E-02	9.00E-03	1.30E-02	1.30E-02	5.00E-03	1.20E-02	7.00E-03	1.40E-02	9.00E-03	1.20E-02	5.6633
auto-correlation time (ACT)	1.26E+06	7.57E+05	1.64E+06	1.27E+06	1.31E+06	8.82E+05	1.87E+06	1.62E+06	8.82E+05	7.33E+05	1.21E+06	1.15E+06	7.18E+05	1.92E+06	7.09E+05	2742.1537
effective sample size (ESS)	**171.1099**	**285.2942**	**131.4706**	**170.2489**	**164.4739**	**244.972**	**115.8154**	**133.5328**	**244.8329**	**294.7317**	**177.8834**	**187.5822**	**300.7755**	**112.7756**	**304.7615**	**78778.954**
Coalescent theta = 0.001																
mean	-14056.8884	2112.0527	6.24E-03	3.73E-03	3.79E-03	9.38E-03	4.83E-03	8.08E-03	8.05E-03	3.33E-03	7.27E-03	3.96E-03	9.09E-03	4.31E-03	7.72E-03	4.4709
stderr of mean	2.3972	2.2749	1.54E-04	6.52E-05	7.50E-05	1.13E-04	1.32E-04	1.41E-04	1.09E-04	4.10E-05	1.17E-04	6.88E-05	1.03E-04	9.48E-05	8.95E-05	2.25E-03
median	-14057.005	2112.857	6.00E-03	4.00E-03	4.00E-03	9.00E-03	5.00E-03	8.00E-03	8.00E-03	3.00E-03	7.00E-03	4.00E-03	9.00E-03	4.00E-03	8.00E-03	4.4276
geometric mean	n/a	2111.7935	6.08E-03	3.64E-03	3.69E-03	9.26E-03	4.65E-03	7.92E-03	7.94E-03	3.27E-03	7.13E-03	3.85E-03	8.98E-03	4.17E-03	7.60E-03	4.4316
95% HPD lower	-14113.569	2045.25	4.00E-03	3.00E-03	3.00E-03	7.00E-03	2.00E-03	5.00E-03	6.00E-03	2.00E-03	4.00E-03	3.00E-03	7.00E-03	3.00E-03	6.00E-03	3.3703
95% HPD upper	-13999.085	2175.474	9.00E-03	5.00E-03	6.00E-03	1.20E-02	7.00E-03	1.10E-02	1.10E-02	4.00E-03	9.00E-03	6.00E-03	1.20E-02	6.00E-03	1.10E-02	5.6844
auto-correlation time (ACT)	1.35E+06	9.61E+05	2.48E+06	1.36E+06	1.34E+06	1.15E+06	1.99E+06	1.52E+06	1.33E+06	9.34E+05	1.49E+06	1.10E+06	1.07E+06	1.48E+06	8.40E+05	2869.1924
effective sample size (ESS)	**150.2401**	**211.2387**	81.9635	**149.0375**	**151.2076**	**176.1078**	**102.2643**	**133.9396**	**152.4752**	**217.3467**	**136.2308**	**184.4789**	**190.4857**	**136.834**	**241.6804**	**70762.7705**
Theta = 0.002																
mean	-13889.8466	1569.2895	5.78E-02	0.1095	9.86E-02	0.151	8.06E-02	7.71E-02	7.30E-02	7.49E-02	6.61E-02	0.1169	0.1359	8.32E-02	5.77E-02	4.4965
stderr of mean	3.7194	14.2642	8.56E-04	1.26E-02	9.81E-03	1.75E-02	1.55E-03	1.30E-03	1.06E-03	1.22E-03	1.58E-03	1.60E-02	1.20E-02	1.32E-03	7.80E-04	3.77E-03
median	-13887.269	1581.991	5.80E-02	6.80E-02	6.50E-02	9.00E-02	8.00E-02	7.60E-02	7.30E-02	7.40E-02	6.70E-02	6.20E-02	9.40E-02	8.20E-02	5.80E-02	4.4496
geometric mean	n/a	1566.3391	5.74E-02	8.45E-02	7.92E-02	0.1203	7.97E-02	7.64E-02	7.25E-02	7.43E-02	6.53E-02	8.12E-02	0.1148	8.26E-02	5.72E-02	4.4568
95% HPD lower	-13941.589	1383.365	4.50E-02	3.40E-02	3.50E-02	5.80E-02	5.70E-02	5.60E-02	5.70E-02	4.40E-02	3.60E-02	6.00E-02	6.50E-02	4.40E-02	3.3718	
95% HPD upper	-13843.149	1738.392	7.00E-02	0.316	0.275	0.414	0.102	9.80E-02	8.90E-02	9.30E-02	8.00E-02	0.42	0.353	0.105	7.10E-02	5.7085
auto-correlation time (ACT)	3.87E+06	4.02E+06	3.05E+06	2.16E+06	1.75E+06	3.03E+06	3.18E+06	2.87E+06	2.83E+06	2.88E+06	5.16E+06	1.62E+06	2.42E+06	2.88E+06	2.11E+06	7019.2931
effective sample size (ESS)	46.5052	44.8337	59.1109	83.3086	**103.1118**	59.4648	56.5972	62.8315	63.5582	62.504	34.8984	**111.0893**	74.4585	62.5088	85.2555	**25646.4574**

Coalescent theta = 0.002	**pi(A)(all)**	**pi(C)(all)**	**pi(G)(all)**	**pi(T)(all)**	**GeneMu(1)**	**GeneMu(2)**	**GeneMu(3)**	**GeneMu(4)**	**GeneMu(5)**	**GeneMu(6)**	**GeneMu(7)**	**GeneMu(8)**	**GeneMu(9)**	**GeneMu(10)**	**GeneMu(11)**	**GeneMu(12)**
mean	0.3036	0.2204	0.2041	0.272	1.0651	0.6686	1.3872	1.1506	0.8258	1.2951	0.5633	1.18	0.661	1.3946	0.7199	
stderr of mean	7.00E-05	6.15E-05	6.10E-05	6.31E-05	1.86E-02	1.15E-02	7.94E-03	4.80E-03	1.54E-02	1.97E-02	1.00E-02	2.88E-03	1.56E-02	1.02E-02	5.35E-03	1.52E-02
median	0.3036	0.2203	0.204	0.272	1.0532	0.6434	0.6062	1.4069	0.8101	1.1577	1.3124	0.5454	1.1782	0.6351	1.4163	0.6928
geometric mean	0.3035	0.2203	0.204	0.272	1.0504	0.6575	0.6201	1.3842	0.8088	1.1354	1.2875	0.5604	1.1674	0.6503	1.3915	0.7048
95% HPD lower	0.2933	0.2113	0.1956	0.2629	0.7622	0.5	0.5	1.2115	0.5115	0.8445	1.0499	0.5	0.8963	0.5	1.2214	0.5
95% HPD upper	0.3127	0.2289	0.2124	0.2818	1.4378	0.916	1.5	1.2221	1.1332	1.5	1.5	0.6914	1.5	0.9065	1.5	1.0195
auto-correlation time (ACT)	42897.1383	40160.317	42898.5746	36639.8328	2.40E+06	1.76E+06	1.35E+06	6.35E+05	1.81E+06	2.52E+06	1.17E+06	5.00E+05	1.82E+06	1.49E+06	7.76E+05	2.11E+06
effective sample size (ESS)	**5035.8604**	**5379.0412**	**5035.6918**	**5895.8784**	89.8994	**123.0925**	**159.8107**	**340.0787**	**119.2494**	85.5832	**184.0132**	**431.8635**	**118.6555**	**151.5617**	**278.3277**	**102.5123**
Coalescent theta = 0.001																
mean	0.3035	0.2204	0.2041	0.272	1.0376	0.6512	0.6484	1.387	0.8225	1.1645	1.3021	0.5625	1.1967	0.655	1.3939	0.7163
stderr of mean	6.93E-05	5.89E-05	5.67E-05	6.76E-05	2.08E-02	8.27E-03	1.05E-02	7.61E-03	1.89E-02	1.77E-02	1.15E-02	3.06E-03	1.72E-02	9.00E-03	5.84E-03	1.27E-02
median	0.3035	0.2203	0.2041	0.272	1.0334	0.6365	0.6194	1.4128	0.8053	1.1705	1.3269	0.5441	1.2113	0.6295	1.4168	0.6996
geometric mean	0.3035	0.2204	0.204	0.272	1.0209	0.6436	0.6383	1.3834	0.803	1.1512	1.2943	0.5597	1.1842	0.6448	1.3909	0.703
95% HPD lower	0.2944	0.2118	0.1956	0.2631	0.7067	0.5	0.5	1.1909	0.5	0.857	1.0451	0.5	0.9176	0.5	1.2186	0.5
95% HPD upper	0.3138	0.229	0.2125	0.2813	1.4138	0.8416	1.5	1.1651	1.5	1.5	1.5	0.6806	1.5	0.8792	1.5	0.9709
auto-correlation time (ACT)	39466.0971	35846.4679	35461.5061	42549.6172	2.60E+06	1.32E+06	1.57E+06	1.24E+06	2.18E+06	2.17E+06	1.40E+06	5.58E+05	2.14E+06	1.10E+06	8.59E+05	1.65E+06
effective sample size (ESS)	**5144.4661**	**5663.9332**	**5725.4195**	**4771.6528**	78.0659	**153.4894**	**128.9169**	**163.9304**	93.1942	93.6226	**144.725**	**363.7495**	95.0976	**184.3739**	**236.3082**	**122.984**
Theta = 0.002																
mean	0.3018	0.2214	0.2057	0.2712	0.8262	0.9135	0.8917	1.0655	1.1342	1.0756	1.0671	1.0461	0.9277	0.8746	1.1506	1.1673
stderr of mean	6.35E-05	6.09E-05	5.67E-05	6.61E-05	1.63E-02	2.05E-02	2.18E-02	2.44E-02	2.42E-02	2.04E-02	2.13E-02	1.80E-02	1.77E-02	1.76E-02	2.29E-02	1.55E-02
median	0.3018	0.2213	0.2056	0.2713	0.8251	0.9205	0.8818	1.0711	1.1324	1.0573	1.0619	1.0432	0.9289	0.8635	1.1424	1.1641
geometric mean	0.3017	0.2213	0.2056	0.2711	0.8196	0.9033	0.8808	1.0537	1.1222	1.0669	1.0573	1.0389	0.9211	0.8646	1.1405	1.1599
95% HPD lower	0.2926	0.2125	0.1975	0.2619	0.6169	0.6159	0.5896	0.7965	0.8277	0.8299	0.7771	0.8195	0.684	0.6357	0.8753	0.9237
95% HPD upper	0.3114	0.2296	0.2139	0.2804	1.0242	1.1651	1.1496	1.3848	1.4281	1.3872	1.3594	1.2938	1.1237	1.1183	1.4615	1.4511
auto-correlation time (ACT)	31619.9206	34802.4877	32762.5858	35554.4893	4.31E+06	4.18E+06	4.32E+06	4.47E+06	3.95E+06	3.90E+06	4.04E+06	3.88E+06	4.72E+06	3.12E+06	4.12E+06	2.50E+06
effective sample size (ESS)	**5693.2464**	**5172.6187**	**5494.6823**	**5063.2143**	41.7298	43.0333	41.6368	40.2391	45.5838	46.1841	44.5866	46.4011	38.1512	57.629	43.6688	71.9969

Appendix 4D. cont. Summary statistics, including effective sample size, obtained with the BEST method for the different priors used.

Theta = 0.001	LnL	LnJointGenePr	TL(1)	TL(2)	TL(3)	TL(4)	TL(5)	TL(6)	TL(7)	TL(8)	TL(9)	TL(10)	TL(11)	TL(12)	TL(13)	kappa(all)
mean	-13870.6866	1560.2717	5.79E-02	0.1535	0.1315	0.2138	7.87E-02	7.50E-02	6.55E-02	6.97E-02	6.23E-02	0.1844	0.1733	8.10E-02	5.99E-02	4.4829
stderr of mean	4.3317	18.9692	1.53E-03	1.82E-02	1.71E-02	3.44E-02	1.35E-03	1.50E-03	1.58E-03	2.20E-03	1.57E-03	3.21E-02	2.51E-02	2.12E-03	9.22E-04	2.30E-03
median	-13868.019	1554.367	5.90E-02	8.30E-02	7.20E-02	9.80E-02	7.80E-02	7.40E-02	6.60E-02	6.80E-02	6.30E-02	7.10E-02	9.70E-02	8.00E-02	5.90E-02	4.4371
geometric mean	n/a	1556.1179	5.73E-02	0.1161	9.51E-02	0.1513	7.80E-02	7.43E-02	6.47E-02	6.86E-02	6.16E-02	0.1095	0.1331	8.01E-02	5.96E-02	4.4432
95% HPD lower	-13919.811	1360.966	3.90E-02	4.50E-02	3.50E-02	6.30E-02	5.60E-02	4.50E-02	4.90E-02	4.10E-02	4.30E-02	6.10E-02	5.90E-02	4.70E-02	3.50E-02	3.3488
95% HPD upper	-13825.359	1749.676	7.10E-02	0.404	0.379	0.607	9.70E-02	9.50E-02	8.00E-02	9.50E-02	7.80E-02	0.633	0.5	0.103	7.30E-02	5.6763
auto-correlation time (ACT)	4.57E+06	5.02E+06	6.05E+06	2.74E+06	2.64E+06	4.41E+06	3.11E+06	3.67E+06	4.58E+06	5.48E+06	5.24E+06	2.07E+06	4.80E+06	5.08E+06	3.34E+06	2636.3144
effective sample size (ESS)	39.3839	35.8698	29.7634	65.76	68.2827	40.8357	57.8041	49.0584	39.2788	32.8771	34.3434	86.9309	37.5272	35.4606	53.8291	**68284.724**
Coalescent theta = 0.002 wide mtprior																
mean	-14014.5535	1995.3624	7.43E-03	3.93E-03	3.49E-03	1.33E-02	5.63E-03	9.49E-03	1.06E-02	1.70E-03	8.70E-03	4.06E-03	1.30E-02	4.57E-03	1.18E-02	4.474
stderr of mean	1.3161	1.1295	1.58E-04	1.23E-04	8.46E-05	1.50E-04	1.69E-04	2.09E-04	1.67E-04	6.93E-05	1.47E-04	1.12E-04	1.97E-04	9.76E-05	7.61E-05	2.10E-03
median	-14014.072	1995.634	7.00E-03	4.00E-03	3.00E-03	1.30E-02	5.00E-03	9.00E-03	1.10E-02	2.00E-03	9.00E-03	4.00E-03	1.30E-02	4.00E-03	1.20E-02	4.4296
geometric mean	n/a	1995.1716	7.26E-03	3.73E-03	3.30E-03	1.31E-02	5.38E-03	9.23E-03	1.05E-02	n/a	8.52E-03	3.86E-03	1.28E-02	4.38E-03	1.16E-02	4.4345
95% HPD lower	-14062.336	1941.602	5.00E-03	2.00E-03	1.00E-03	1.00E-02	3.00E-03	5.00E-03	8.00E-03	1.00E-03	6.00E-03	2.00E-03	9.00E-03	3.00E-03	9.00E-03	3.3348
95% HPD upper	-13968.789	2049.918	1.10E-02	6.00E-03	5.00E-03	1.80E-02	9.00E-03	1.30E-02	1.50E-02	3.00E-03	1.20E-02	6.00E-03	1.70E-02	7.00E-03	1.50E-02	5.6486
auto-correlation time (ACT)	6.56E+05	3.62E+05	1.99E+06	2.15E+06	1.19E+06	1.08E+06	2.06E+06	1.88E+06	1.56E+06	1.77E+06	1.46E+06	1.65E+06	1.65E+06	1.24E+06	4.21E+05	2659.8156
effective sample size (ESS)	**329.3756**	**596.3597**	**108.7126**	**100.3848**	**181.2924**	**200.3273**	**104.8296**	**114.778**	**138.0885**	**122.1677**	**148.3131**	**131.2413**	**131.2564**	**174.7985**	**513.4187**	**81217.661**
Coalescent theta = 0.0015 wide mtprior																
mean	-14002.8524	1965.0787	7.99E-03	3.99E-03	3.86E-03	1.40E-02	5.62E-03	9.59E-03	1.19E-02	1.72E-03	9.06E-03	4.16E-03	1.37E-02	5.26E-03	1.25E-02	4.4761
stderr of mean	1.327	1.2108	1.65E-04	7.62E-05	9.18E-05	1.96E-04	1.40E-04	2.01E-04	1.76E-04	7.26E-05	1.83E-04	9.60E-05	1.89E-04	1.44E-04	9.49E-05	2.33E-03
median	-14002.397	1965.262	8.00E-03	4.00E-03	4.00E-03	1.40E-02	5.00E-03	9.00E-03	1.20E-02	2.00E-03	9.00E-03	4.00E-03	1.40E-02	5.00E-03	1.20E-02	4.4336
geometric mean	n/a	1964.8961	7.82E-03	3.85E-03	3.68E-03	1.38E-02	5.39E-03	9.39E-03	1.17E-02	1.56E-03	8.86E-03	3.99E-03	1.35E-02	5.04E-03	1.24E-02	4.4369
95% HPD lower	-14050.3	1912.584	5.00E-03	2.00E-03	2.00E-03	1.00E-02	3.00E-03	6.00E-03	8.00E-03	1.00E-03	6.00E-03	2.00E-03	9.00E-03	3.00E-03	9.00E-03	3.3692
95% HPD upper	-13958.939	2017.584	1.10E-02	6.00E-03	6.00E-03	1.80E-02	9.00E-03	1.30E-02	1.60E-02	3.00E-03	1.30E-02	6.00E-03	1.80E-02	8.00E-03	1.60E-02	5.6842
auto-correlation time (ACT)	5.63E+05	3.68E+05	1.74E+06	9.46E+05	1.08E+06	1.26E+06	1.31E+06	1.88E+06	1.11E+06	1.51E+06	1.59E+06	1.18E+06	1.16E+06	1.65E+06	4.74E+05	2736.1945
effective sample size (ESS)	**319.8999**	**489.244**	**103.559**	**190.2056**	**166.5611**	**142.7667**	**137.7372**	96.0077	**162.1552**	**119.1184**	**112.9097**	**152.3445**	**155.0952**	**108.9761**	**379.7914**	**65792.1067**
Coalescent theta = 0.001 wide mtprior																
mean	-14056.0418	2111.8597	6.15E-03	3.21E-03	2.90E-03	1.04E-02	4.25E-03	7.69E-03	8.81E-03	1.45E-03	6.91E-03	3.26E-03	1.05E-02	3.70E-03	9.54E-03	4.4754
stderr of mean	1.658	1.5439	1.19E-04	1.00E-04	9.22E-05	1.60E-04	1.20E-04	1.53E-04	1.57E-04	5.33E-05	1.27E-04	9.41E-05	1.35E-04	1.05E-04	8.37E-05	2.12E-03
median	-14055.979	2112.443	6.00E-03	3.00E-03	3.00E-03	1.00E-02	4.00E-03	8.00E-03	9.00E-03	1.00E-03	7.00E-03	3.00E-03	1.00E-02	4.00E-03	9.00E-03	4.4332
geometric mean	n/a	2111.641	6.00E-03	3.04E-03	2.76E-03	1.03E-02	4.06E-03	7.51E-03	8.63E-03	n/a	6.76E-03	3.10E-03	1.03E-02	3.54E-03	9.42E-03	4.4361
95% HPD lower	-14108.857	2052.975	4.00E-03	1.00E-03	2.00E-03	7.00E-03	2.00E-03	5.00E-03	6.00E-03	1.00E-03	4.00E-03	2.00E-03	8.00E-03	2.00E-03	7.00E-03	3.3638
95% HPD upper	-14004.563	2171.422	9.00E-03	5.00E-03	5.00E-03	1.40E-02	7.00E-03	1.10E-02	1.30E-02	3.00E-03	9.00E-03	5.00E-03	1.50E-02	6.00E-03	1.20E-02	5.6795
auto-correlation time (ACT)	8.28E+05	5.58E+05	1.65E+06	1.80E+06	2.12E+06	1.58E+06	1.94E+06	1.78E+06	1.67E+06	1.53E+06	1.65E+06	1.95E+06	1.17E+06	1.94E+06	5.97E+05	2709.7231
effective sample size (ESS)	**260.7741**	**386.8843**	**130.8078**	**119.7955**	**101.866**	**136.909**	**111.3194**	**121.4849**	**129.707**	**141.3856**	**130.7274**	**110.796**	**184.8808**	**111.1737**	**361.5574**	**79721.7981**

Theta = 0.001	pi(A)(all)	pi(C)(all)	pi(G)(all)	pi(T)(all)	GeneMu(1)	GeneMu(2)	GeneMu(3)	GeneMu(4)	GeneMu(5)	GeneMu(6)	GeneMu(7)	GeneMu(8)	GeneMu(9)	GeneMu(10)	GeneMu(11)	GeneMu(12)
mean	0.3017	0.2215	0.2056	0.2712	0.891	0.9717	0.8352	1.0061	1.1651	1.0216	0.9909	0.9869	0.9519	0.9795	1.1004	1.1669
stderr of mean	6.18E-05	5.86E-05	5.60E-05	6.19E-05	2.43E-02	2.15E-02	2.52E-02	3.11E-02	2.51E-02	1.87E-02	2.65E-02	2.28E-02	1.93E-02	2.77E-02	2.33E-02	2.94E-02
median	0.3016	0.2215	0.2056	0.2713	0.9066	0.9655	0.8269	1.0097	1.1648	1.0209	1.0048	1.0007	0.9614	0.9693	1.0944	1.1804
geometric mean	0.3016	0.2215	0.2056	0.2712	0.8806	0.9628	0.8242	0.9912	1.1548	1.0148	0.9794	0.9775	0.945	0.9674	1.0909	1.1551
95% HPD lower	0.2921	0.2128	0.1975	0.262	0.5946	0.7267	0.5695	0.7394	0.8842	0.8018	0.7461	0.7478	0.7185	0.7179	0.8036	0.875
95% HPD upper	0.3112	0.2302	0.2145	0.2806	1.1101	1.2236	1.0991	1.3025	1.4569	1.2449	1.2845	1.2189	1.1422	1.2466	1.3657	1.4567
auto-correlation time (ACT)	28326.0642	30947.9828	30236.6861	30379.94	6.28E+06	4.88E+06	6.22E+06	5.92E+06	4.80E+06	4.56E+06	5.63E+06	5.22E+06	5.31E+06	5.82E+06	4.70E+06	5.81E+06
effective sample size (ESS)	**6355.2776**	**5816.8573**	**5953.6948**	**5925.6207**	28.6788	36.9174	28.9286	30.4002	37.4927	39.4748	32.0025	34.5116	33.9214	30.9555	38.3085	30.9863
Coalescent theta = 0.002 wide mtprior																
mean	0.3035	0.2204	0.2042	0.2719	0.996	0.5523	0.4764	1.6107	0.7703	1.1101	1.393	0.2285	1.1623	0.5368	1.6372	0.6116
stderr of mean	6.19E-05	5.75E-05	5.38E-05	6.54E-05	1.96E-02	1.71E-02	1.05E-02	1.82E-02	2.00E-02	2.36E-02	2.32E-02	8.09E-03	2.09E-02	1.35E-02	2.20E-02	1.23E-02
median	0.3035	0.2204	0.2042	0.2718	0.9799	0.5356	0.4623	1.6098	0.7436	1.0873	1.379	0.2121	1.1511	0.5232	1.6562	0.5981
geometric mean	0.3035	0.2204	0.2041	0.2719	0.9782	0.5297	0.4571	1.6003	0.7453	1.0871	1.3766	0.2134	1.1452	0.517	1.6234	0.5914
95% HPD lower	0.2939	0.2116	0.1959	0.2627	0.6475	0.2647	0.2295	1.2715	0.4223	0.6772	1.0147	0.1	0.7777	0.2749	1.2753	0.3361
95% HPD upper	0.3132	0.2291	0.2126	0.2813	1.3789	0.8815	0.7422	1.952	1.1971	1.5389	1.8282	0.3949	1.566	0.8314	2	0.9234
auto-correlation time (ACT)	34121.4555	36610.944	34023.1873	41064.1652	2.34E+06	2.55E+06	1.26E+06	2.18E+06	2.17E+06	2.35E+06	2.52E+06	1.88E+06	2.36E+06	1.84E+06	2.44E+06	1.35E+06
effective sample size (ESS)	**6331.0312**	**5900.5307**	**6349.317**	**5260.6451**	92.1757	84.8602	**170.825**	98.921	99.7075	91.9922	85.8081	**115.0195**	91.4027	**117.2331**	88.6857	**159.8876**
Coalescent theta = 0.0015 wide mtprior																
mean	0.3035	0.2204	0.2041	0.272	1.0147	0.5304	0.5002	1.5958	0.7292	1.0597	1.4766	0.2198	1.1456	0.5242	1.6299	0.6615
stderr of mean	6.89E-05	6.44E-05	6.35E-05	6.55E-05	2.07E-02	9.87E-03	1.16E-02	2.36E-02	1.62E-02	1.91E-02	2.38E-02	8.50E-03	2.39E-02	1.28E-02	1.96E-02	1.64E-02
median	0.3036	0.2204	0.204	0.2721	0.9997	0.523	0.4842	1.5911	0.7159	1.0492	1.4747	0.2006	1.1351	0.5123	1.6392	0.6584
geometric mean	0.3035	0.2204	0.204	0.272	0.9959	0.5162	0.4818	1.5831	0.7069	1.0443	1.4602	0.2044	1.1274	0.5057	1.6188	0.6402
95% HPD lower	0.2941	0.2118	0.1955	0.2627	0.6512	0.2991	0.2626	1.2714	0.3964	0.7281	1.0526	0.1	0.7518	0.2705	1.3144	0.3566
95% HPD upper	0.3134	0.2291	0.2126	0.2811	1.4244	0.7701	0.7739	2	1.072	1.4278	1.8872	0.3892	1.5547	0.8112	2	0.9869
auto-correlation time (ACT)	34851.4267	37845.5412	38137.5165	34130.5717	1.99E+06	1.18E+06	1.30E+06	2.53E+06	1.42E+06	2.02E+06	2.17E+06	1.62E+06	2.45E+06	1.52E+06	2.00E+06	1.78E+06
effective sample size (ESS)	**5165.3553**	**4756.703**	**4720.2864**	**5274.4502**	90.4936	**152.4244**	**138.1348**	71.2896	**126.9522**	89.1503	83.0483	**111.4624**	73.542	**118.3761**	90.2279	**101.0204**
Coalescent theta = 0.001 wide mtprior																
mean	0.3036	0.2204	0.2041	0.2719	1.025	0.5574	0.4944	1.57	0.7227	1.1166	1.434	0.241	1.1387	0.536	1.6358	0.6151
stderr of mean	6.56E-05	6.14E-05	5.55E-05	6.40E-05	2.00E-02	1.54E-02	1.41E-02	2.42E-02	1.76E-02	2.05E-02	2.44E-02	8.34E-03	2.00E-02	1.40E-02	1.85E-02	1.55E-02
median	0.3034	0.2203	0.2041	0.2719	1.0218	0.5309	0.4817	1.5753	0.7069	1.1071	1.4306	0.2247	1.1317	0.5261	1.6463	0.6003
geometric mean	0.3035	0.2204	0.2041	0.2719	1.0057	0.536	0.4759	1.5539	0.6999	1.0989	1.4159	0.2257	1.1232	0.5174	1.6236	0.5956
95% HPD lower	0.294	0.2115	0.1958	0.2626	0.6444	0.29	0.2556	1.2015	0.3899	0.7517	1.0284	0.1	0.7582	0.249	1.3021	0.3323
95% HPD upper	0.3132	0.2289	0.2128	0.281	1.4344	0.8887	0.7638	2	1.0937	1.5249	1.9117	0.4122	1.5172	0.8259	2	0.9287
auto-correlation time (ACT)	38658.0244	40845.4502	35771.0607	39066.0753	2.24E+06	2.02E+06	2.27E+06	2.64E+06	2.03E+06	2.28E+06	2.54E+06	1.80E+06	2.44E+06	2.10E+06	1.92E+06	2.16E+06
effective sample size (ESS)	**5588.0766**	**5288.8143**	**6039.0717**	**5529.7083**	96.604	**106.9998**	94.9673	81.8916	**106.5708**	94.7027	85.0501	**120.3238**	88.5991	**102.6477**	**112.2899**	99.929

Chapter V (Article 4)
Influences of Past Climatic Changes in the Phylogeography of a Cosmopolitan Marine Top Predator, the Common Dolphin (Genus *Delphinus*)

Ana R. Amaral[1,2,*], Luciano B. Beheregaray[2,3], Kerstin Bilgmann[4], Luís Freitas[5], Kelly M. Robertson[6], Marina Sequeira[7], Karen A. Stockin[8], M. Manuela Coelho[1] and Luciana M. Möller[3]

[1] Universidade de Lisboa, Faculdade de Ciências, Centro de Biologia Ambiental, Campo Grande, 1749-016 Lisbon, Portugal
[2] Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
[3] School of Biological Sciences, Flinders University, Adelaide SA 5001, Australia
[4] Graduate School of the Environment, Macquarie University, Sydney, NSW 2109, Australia
[5] Museu da Baleia da Madeira, 9200-031, Caniçal, Madeira, Portugal
[6] National Marine Fisheries Service, Southwest Fisheries Science Center, 3333 N. Torrey Pines Ct., La Jolla, CA 92037, USA
[7] Instituto de Conservação da Natureza e Biodiversidade, Rua de Santa Marta, 55, 1150-294 Lisbon, Portugal
[8] Coastal-Marine Research Group, Institute of Natural Sciences, Massey University, Private Bag 102 904, North Shore MSC, Auckland, New Zealand

Molecular Ecology (in review)

Influences of past climatic changes in the phylogeography of a cosmopolitan marine top predator, the common dolphin (genus *Delphinus*)

5.1. ABSTRACT

Climatic oscillations during the Pleistocene have greatly influenced the distribution and connectivity of many organisms, leading to extinctions but also generating biodiversity. While the effects of such changes have been extensively studied in the terrestrial environment, studies focusing on the marine realm are still scarce. Here we used sequence data from one mitochondrial and five nuclear loci to assess the influence of Pleistocene climatic changes on the phylogeography and demographic history of a cosmopolitan top predator, the common dolphin (genus *Delphinus*). Population samples representing the three major morphotypes of *Delphinus* were obtained from 10 oceanic regions. Our results indicate that short-beaked common dolphins originated in the Pacific Ocean during the Pleistocene and expanded into the Atlantic through the Indian Ocean. On the other hand, long-beaked common dolphins appeared to have evolved more recently and independently in different oceans by exploring new coastal habitats, following periods of intensive upwelling, which made resources scarce. Short-beaked common dolphins also show recurrent demographic expansions concomitant with changes in sea surface temperature during the Pleistocene and associated increases in resource availability, which varied in the North Atlantic and Pacific Ocean basins. Phylogeographic patterns associated with vicariant events during the Pleistocene and with different effects that glaciations had in different ocean basins are also described for common dolphins. By proposing how past environmental changes had an impact on the demography and speciation of a widely distributed, top marine predator, we highlight the impacts that climate change may have on the distribution and abundance of marine predators and its ecological consequences for marine ecosystems.

Key Words: Cetaceans; adaptive evolution, speciation; evolutionary biology; taxonomy.

5.2. INTRODUCTION

Strong selection pressures and vicariant events due to past climatic changes have influenced biodiversity patterns at both regional and global scales. In the marine realm, global fluctuations in climate during the Pleistocene, and consequent sea level changes due to expansion and recession of ice sheets have not only produced land bridges but also changed sea surface temperatures and the flow of ocean currents (Hewitt 1996; Hewitt 2000). Such historical events have influenced the distribution and connectivity of marine organisms, leading species to extinction but also acting as drivers of biodiversity (Hewitt 1996). Understanding how such shifts in the geographic distribution of organisms have impacted intraspecific genetic diversity and studying the underlying processes generating these patterns have long been of interest to evolutionary biologists and more recently, to phylogeographers (Avise 2000; Beheregaray 2008).

However, the majority of studies assessing the influence of climatic oscillations on the phylogeography of marine organisms were conducted either on regional scales or on organisms with larval dispersal (e.g. Lamurseau *et al.* 2009; Lopez *et al.* 2010). In contrast, such studies are rare in actively dispersing top marine predators, particularly for those distributed over global scales (e.g. Duncan *et al.* 2006).

Cetaceans are a group of marine mammals that radiated from their terrestrial ancestors around 53 Mya (Arnason *et al.* 2004). Patterns of ocean restructuring during climatic oscillations in the Oligocene and Miocene have been suggested to account for the radiation of extant cetacean species (Steeman *et al.* 2009). Peaks in availability of resources caused by changes in upwelling intensities, in particular, have been associated with peaks in cetacean species diversity (Lipps & Mitchell 1976). Throughout their evolutionary history, dietary specializations are thought to have lead to ecomorphological diversity (Slater *et al.* 2010) and to convergent evolution (Natoli *et al.* 2006; Natoli *et al.* 2004), particularly in the Delphinidae, the most speciose family of marine mammals (LeDuc 2009). Within the Delphinidae, the subfamily Delphininae includes the recently evolved and closely related polytypic genera *Tursiops*, *Stenella* and *Delphinus*, whose

phylogenetic relationships, phylogeography and evolutionary history are still under debate (e.g. Amaral *et al*. in review; Caballero *et al.* 2008; LeDuc *et al.* 1999; Möller *et al.* 2008; Natoli *et al.* 2006; Natoli *et al.* 2004).

Common dolphins are widely distributed and abundant small cetaceans that present great morphological variability throughout their distribution. Two species and four subspecies within the genus are currently recognized: the short-beaked common dolphin, *Delphinus delphis* Linnaeus, 1758, distributed in tropical and temperate continental shelf and pelagic waters of the Atlantic, Pacific and Southeast Indian Oceans (hereinafter referred to as the short-beaked morphotype), the long-beaked common dolphin, *Delphinus capensis* Gray, 1828, distributed in nearshore tropical and temperate waters of the Pacific and southern Atlantic waters (hereinafter referred to as the long-beaked morphotype), the Arabian common dolphin, *D. c. tropicalis* van Bree, 1971 restricted to the Indian Ocean (hereinafter referred to as the *tropicalis*-form), and the Black Sea common dolphin, *D. d. ponticus* Barabash, 1935, restricted to the Black Sea (Perrin 2009). Previous molecular studies based on mitochondrial DNA (mtDNA) data corroborated the separation of the short and long-beaked morphotypes occurring in California as two species (Rosel *et al.* 1994). However, when populations from other regions were analysed, a disagreement between morphological and genetic characters was found (Amaral *et al.* 2007; Kingston & Rosel 2004; LeDuc *et al.* 1999; Natoli *et al.* 2006). A highly divergent mtDNA clade including short-beaked individuals from the Northeast Atlantic and *tropicalis* individuals from the Indian Ocean was reported (Amaral *et al.* 2007) and long-beaked populations from the Northeast Pacific and off South Africa showed high levels of differentiation, suggested as an independent process of evolution and convergence on the same morphotype (Natoli *et al.* 2006). As for short-beaked populations, patterns of genetic differentiation varied from low levels of differentiation found in the North Atlantic (Amaral *et al.* 2007; Mirimin *et al.* 2009; Natoli *et al.* 2006; Querouil *et al.* 2010) to fine-scale population structure found along the eastern (Möller *et al.* 2011) and southern Australian coasts (Bilgmann *et al.* 2008).

Nonetheless, no studies to date have investigated the phylogeography of common dolphins in the light of past climatic changes or tested general marine biogeographic models that could have accounted for the origin of *Delphinus*. An origin in the Indo-Pacific region with subsequent dispersal into the Atlantic via southern Africa is a biogeographic scenario reported for several tropical and subtropical marine animals (Bowen *et al.* 2001; Bowen *et al.* 1997; Bowen & Grant 1997; Briggs 1974; Duncan *et al.* 2006; Graves 1998), including dolphins of the genus *Stenella* (Perrin 2007; Perrin *et al.* 1987). The dispersal between Indian and Atlantic Ocean basins is facilitated by the Agulhas current system, which occasionally projects warm water masses westward around the horn of South Africa (Peeters *et al.* 2004). This passage was however intermittent, with long-term periods of isolation between the Atlantic and Indian Ocean basins (Peeters *et al*. 2004) leaving a signal in the mtDNA phylogenies of many fishes (Bremer *et al.* 1998; Grant & Bowen 1998; Martinez *et al.* 2006), prawns (Teske *et al.* 2009), marine turtles (Bowen *et al.* 1997), sea birds (Avise *et al.* 2000) and sharks (Duncan *et al.* 2006).

Here we use an ideal study system to assess the influence of past climatic changes on population history and diversification of a globally distributed top marine predator. Common dolphins occupy a top position in the marine food chain and their distribution is thought to be associated with that of their prey and with specific water masses characterized by different temperature regimes (Ballance *et al.* 2006; Möller *et al.* 2011), including “upwelling-modified” waters in both tropical (Ballance *et al.* 2006) and temperate regions (Möller pers. obs.). These “upwelling-modified” waters are regions of highly variable oceanographic features, characterized by year round or seasonal rising of cool nutrient-rich waters from the bottom of the ocean towards the surface (Au & Perryman 1985). Our study contrasts with previous molecular surveys of the genus since it is the first to specifically address phylogeographic hypotheses using coalescent-based multilocus methods and a large dataset in terms of population samples and genetic markers. The latter includes population samples representing three common dolphin morphotypes collected in ten oceanic regions across the Pacific, Indian and Atlantic Oceans, and sequence data from six loci (one mitochondrial

and five nuclear markers). Multilocus datasets offer greater power to estimate demographic parameters by providing replicate samples of the underlying history affecting the genome and replicate samples of the coalescent process (Carling & Brumfield 2007; Felsenstein 2006; Lee & Edwards 2008). Our analysis focuses on assessing the influence of Pleistocene climatic oscillations on population demography, geographic distribution and speciation in common dolphins, a highly mobile, top marine predator. Moreover, we test a historical scenario for the origin of the genus that involves an origin in the Indo-Pacific region with subsequent dispersal into the Atlantic Ocean basin via southern Africa. We predict that populations across the different oceans will differ in genetic diversity, divergence times and demographic and genealogical patterns. An understanding of the phylogeography and historical dispersal patterns of common dolphins will certainly contribute to our knowledge of how glaciations have influenced marine populations at a global scale.

5.3. MATERIALS AND METHODS

5.3.1. *Sampling and DNA extraction*

In total, we analysed 343 common dolphin samples representing 10 oceanic regions and all main morphotypes of *Delphinus* (Figure 5.1). For the short-beaked morphotype, the sampled regions were the Northeast Atlantic (NEATL), *n* = 63; the Central Eastern Atlantic (CEATL), *n* = 21; the Northwest Atlantic (NWATL), *n* =27; the Northeast Pacific (NEPAC), *n* = 26; the Southwest Pacific, *n* = 41 (encompassing eastern Australian waters, SWPAC_AUS) and *n* = 40 (encompassing New Zealand waters, SWPAC_NZ) and the Southeast Indian Ocean (southern Australian waters, SEIND), *n* = 27 (Figure 5.1). For the long-beaked morphotype, the sampled regions were the Northeast Pacific, *n* = 40; the Southeast Atlantic (SEATL), off South Africa, *n* = 26 (these samples are here classified as long-beaked following Samaai *et al.* (2005) and P. Best (pers. comm.)); and the Southwest Atlantic (SWATL), off Brazil, *n* =7. Finally, for the *tropicalis*-form, *n* = 25, samples were obtained from the Arabian Sea in the Western Indian Ocean (WIND). Tissue samples were obtained either from stranded animals or from dart biopsying live animals and preserved in 99%

ethanol. 144 samples from NWATL, NEPAC, SEATL and WIND were received from the Southwest Fisheries Science Center, Marine Mammal and Turtle Research Sample Collection (SWFSC-NOAA, La Jolla, CA) as extracted DNA. DNA from remaining samples was extracted from muscle or skin tissue using either a standard proteinase K and two phenol-chloroform-isoamyl extractions (Rosel & Block 1996), or a salting-out method (Sunnucks & Hales 1996).

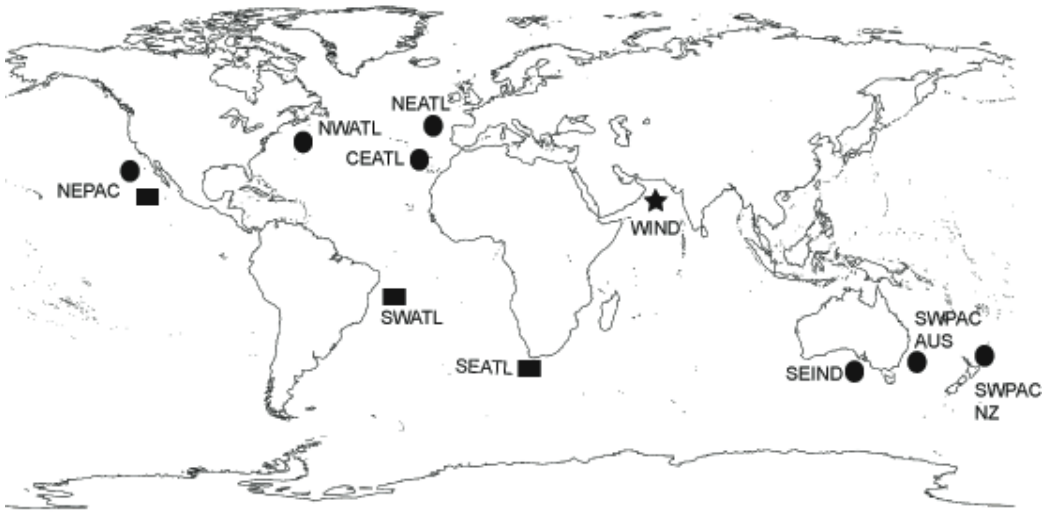


Figure 5.1. Map showing sampling locations for the common dolphin populations analysed in this study. (NEPAC – Northeast Pacific; NWATL – Northwest Atlantic; CEATL – Central Eastern Atlantic; SWATL – Southwest Atlantic; SEATL – Southeast Atlantic; WIND – Western Indian Ocean; SEIND – Southeast Indian Ocean; SWPAC_AUS – Southwest Pacific Australia; SWPAC_NZ – Southwest Pacific New Zealand). ● Short-beaked morphotype ■ Long-beaked morphotype★ *tropicalis* morphotype.

5.3.2. *Sequencing*

mtDNA

The cytochrome *b* gene was amplified and sequenced (1121 bp) using primers on the transfer RNA (tRNA) genes for the 343 samples (GenBank Accession Numbers XXXX-XXXX). The L-strand primer was on tRNA glutamine (L14724, 5'-TGACTTGAARAACCAYCG TTG 3') and the H-strand primer on tRNA threonine (5'CCTTTTCCGGTTTACAAGAC 3') (LeDuc *et al.* 1999). The thermocycle profile and PCR conditions used are described in Amaral *et al.* (2007). The PCR products were cleaned by adding 0.5 U of Shrimp Alkaline Phosphatase and 5 U of Exonuclease I and incubating at 37ºC for 30 min and 80ºC for 15 min. Both

strands were directly sequenced (BigDye Terminator CycleSequencing; Applied Biosystems) on an ABI 3730 automated sequencer (Applied Biosystems). All sequences obtained were aligned using the software Sequencher, v.4.2 (Gene Codes Corporation).

Nuclear Loci

Three anonymous nuclear loci developed from *Delphinus delphis* [Del_12, Del_15 and Del_18 (Amaral *et al.* 2010)] and two introns [CHRNA1 (Roca *et al.* 2001) and PLP (Lyons *et al.* 1999)] were PCR amplified and sequenced for 92 common dolphin samples (Table 5.1) (short-beaked morphotype: NE Atlantic, *n* = 9; CE Atlantic, *n* = 10; NW Atlantic, *n* = 9; SW Pacific Australia, *n* =6, SW Pacific New Zealand, *n* = 10; NE Pacific, *n* =11, SE Indian, *n* = 5; long-beaked morphotype: SE Atlantic, *n* = 7; SW Atlantic, *n* = 7; NE Pacific, *n* = 11; *tropicalis*-form: W Indian, *n* = 5) (GenBank Accession Numbers XXXX-XXXX). The PCRs were performed in 25-μL reactions containing 10-100 ng DNA, 0.2 mM each dNTP, 0.3 μM each primer, 1 U *Taq* Polymerase and 1 X *Taq* buffer. PCR products were cleaned, sequenced and aligned as above. In order to obtain all haplotypes for each nuclear locus, we used the Bayesian approach implemented in the software Phase 2.1.1 (Stephens & Donnelly 2003; Stephens *et al.* 2001). The default settings were used, except that we only accepted haplotype reconstructions with Bayesian posterior probabilities of ≥ 95%. Tests for recombination were performed for each locus, using the *Maximum Chisquare* test of Smith (1992).

5.3.3. *Statistical analyses*

Genetic diversity and demography

In all analyses, each sample from an oceanic region and each morphotype was considered a putative population. Sequence diversity measures for mtDNA and each nuclear locus (nucleotide and haplotype diversities) were estimated in DNAsp v. 5.10.01 (Librado & Rozas 2009). To detect departures from neutrality or from constant population sizes, Tajima's *D* (Tajima 1989), Fu's *Fs* (Fu 1997) and R^2 (Ramos-Onsis & Rozas 2002) were estimated in DNAsp. Significant negative values of Tajima's *D*, as well as large negative values of *Fs* and small positive

values of R^2 indicate population growth, whereas significant positive values of Tajima's *D* are a signature of genetic subdivision (Ramos-Onsins & Rozas 2002). To further investigate past population dynamics we calculated mismatch distributions (MMD; Rogers & Harpending 1992) and estimated the demographic parameters τ, θ_0 and θ_1 in Arlequin v. 3.5 (Excoffier & Lischer 2010). Goodness of fit was assessed by the sum of square deviations (SSD) and the Harpending raggedness indices between the observed and the expected mismatch with their significance determined by a parametric bootstrap. The relationship Tau, $\tau = 2\mu kt$ was used to estimate the time of expansion (*t*), where *k* is number of nucleotides sequenced and μ is the mutation rate per nucleotide. For the cytochorme *b* gene we used the average mutation rate estimated for the delphinid mitochondrial genome, 9.86 x 10^{-9} substitutions per nucleotide per year (Vilstrup *et al.* 2011). For the nuclear loci, since no estimates exist for delphinids, we used the mutation rates estimated for whales: 4.79 x 10^{-10} substitutions per nucleotide site per year for the nuclear loci (averaged substitution rates of seven autosomal nuclear introns; Alter *et al*. 2007). In addition we also used a coalescent-based multi-locus method, the Extended Bayesian Skyline Plot (EBSP), which takes genealogy into account and is expected to provide a better estimate of demographic history than other methods (Drummond *et al*. 2005). Moreover, by allowing the analysis of multiple loci, this nonparametric Bayesian MCMC method is more powerful than previous coalescent-based methods in estimating changes in population size through time (Drummond *et al.* 2005). Analyses were performed for each putative population using sequence data for the cytochrome *b* gene and for the five nuclear loci. The same mutation rates as described above were used. 1 x 10^7 MCMC generations were run in the program BEAST v.1.6.1 (Drummond & Rambaut 2007), where this method is implemented. Since not all models of nucleotide evolution are available to choose in BEAST, the HKY model was chosen because it was the most approximate model to the ones we obtained in jModeltest (Posada 2008) for each locus (results not shown); all other parameters were set as suggested by the authors (Heled & Drummond 2008). Convergence of the MCMC chains was inspected using Tracer v.1.5 (Rambaut & Drummond 2007) by visually checking the Effective Sample Size (ESS) values.

Table 5.1. Molecular markers sequenced for this study with information on the number of base pairs sequenced, variable sites and total haplotypes obtained.

Marker	Length (bp)	Variable sites	Total Haplotypes
Cytochrome *b*	1121	192	175
CHRNA1	379	12	26
Del_12	801	21	33
Del_15	746	20	33
Del_17	733	13	21
PLP	775	13	16
Total	4555	271	304

Population differentiation and phylogeography

Population differentiation was tested by calculating pairwise ϕ_{ST} using Tamura-Nei distances for both mtDNA and nuclear DNA datasets in Arlequin v.3.5 (Excoffier & Lischer 2010). Significance was tested through 10000 permutations and significance levels of multiple tests were adjusted using sequential Bonferroni corrections (Rice 1989). An analysis of molecular variance (AMOVA) was also computed using the following hierarchical levels: (1) all putative populations of short-beaked and long-beaked morphotypes; (2) short-beaked populations from different ocean basins (i.e. the Atlantic, the Pacific and SE Indian Ocean); (3) short-beaked populations within each basin, in which populations from each ocean basin were subdivided in western and eastern regions; (4) long-beaked from different ocean basins (i.e. the Atlantic, the Pacific and Western Indian Ocean). Significance was also tested through 10,000 permutations in Arlequin.

Genealogical relationships among haplotypes were inferred for each nuclear locus and for the cytochrome *b* through median-joining networks as implemented in Network v.4.6.0.0 (Bandelt *et al.* 1999). In addition, to generate a nuclear multi-locus phylogeny, we used the POFAD software (Phylogeny of Organisms From Allelic Data; Joly & Bruneau 2006). The algorithm implemented in POFAD combines genetic distance matrices generated from allelic data from individual loci into a single genetic distance matrix. Distance matrices for each of the five nuclear loci were generated in PAUP* (Swofford 2003) using uncorrected *p* distances and then inserted in POFAD to generate a combined-locus matrix. A

neighbor-joining phylogram was then reconstructed in PAUP* using the combined-locus genetic distance matrix.

Population divergence times and migration rates based on mtDNA

Divergence time estimates and migration rates between each putative population pair were obtained using the cytochrome *b* data set and the MCMC approach implemented in the program MDIV (Nielsen & Wakeley 2001). The estimated parameters were θ ($\theta = 2N_{ef}\mu$) where N_{ef} is the effective population size and μ is the mutation rate, *M* ($M = 2N_{ef}m$) where *m* is the migration rate, and *T* ($T = t/N_{ef}$) where *t* is the divergence time. These parameters were obtained using a finite sites model (HKY) to allow for the possibility of multiple mutations per site. We ran 1×10^6 cycles with a burn-in of 1×10^5. Maximum values for *T* and *M* were set at 5 and 40, respectively. Five runs with different random seeds were run for each population comparison. Divergence time (*t*) was calculated as in Brown *et al.* (2007), using the formula $t = T * \theta/(2u) * g$. T and θ are generated by the program, u is the mutation rate and *g* is the generation time. *u* was calculated as $2*\mu*k$, where μ is the mutation rate per nucleotide and *k* is the length of the sequence. As above, 9.86×10^{-9} substitutions per nucleotide site per year was the mutation rate used for the cytochrome *b* gene (Vilstrup *et al.* 2011). A generation time of 7 years was considered since it is within the range of the age of sexual maturity described for female common dolphins for the Pacific and Atlantic oceans (Murphy *et al.* 2009; Perrin 2009).

5.4. RESULTS

5.4.1. *Genetic diversity*

Haplotype and nucleotide diversities for the mtDNA were high for most populations, with short-beaked populations from the Pacific Ocean showing higher diversity than those from the Atlantic (Table 5.2). In comparison, long-beaked common dolphins showed lower diversity for all populations, except SWATL for which we had a small sample size. For the nuclear loci, after phasing of heterozygous sites, the final dataset comprised 900 alleles. The five nuclear loci

were polymorphic across the entire dataset but not for every surveyed putative population sample (Table 5.2). Overall, levels of haplotypic and nucleotide diversity at nuclear loci were lower than for cytochrome *b* (Table 5.2). The short-beaked populations from the Pacific Ocean showed the highest nucleotide diversity in CHRNA1 and PLP (0.00721 for SWPAC_NZ and 0.00166 for NEPAC, respectively) (Table 5.2).

5.4.2. *Population differentiation and phylogeography*

Pairwise ϕ_{ST} values obtained for cytochrome *b* showed significant genetic differentiation between most putative populations (Table 5.3). Overall, high differentiation was found between long-beaked and short-beaked populations, except between long-beaked SEATL and short-beaked populations of the Atlantic Ocean. High differentiation was observed between all pairwise long-beaked population comparisons. On the other hand, high structure was found between for short-beaked populations across different oceans, but low or no differentiation was detected between populations from the same ocean or in close geographic proximity (e.g. between CEATL and NEATL (ϕ_{ST} = 0.0167, P>0.05) and SWPAC_AUS and SWPAC_NZ (ϕ_{ST} = 0.0048, P>0.05)).

Pairwise θ_{ST} values obtained for the five nuclear loci differed, reflecting the stochasticity of the nuclear genome (Tables S5.1 – S5.3). Nevertheless, despite differences in the magnitude of levels of differentiation between loci, overall results were concordant to those reported for the cytochrome *b*. The AMOVA analyses showed significant genetic structure among all populations for all molecular markers (Tables 5.4 and 5.5). However, for the cytochrome *b* gene, no significant differences between oceans were detected for short-beaked and long-beaked populations (ϕ_{CT} = 0.029931072, P = 0.16820092; ϕ_{CT} = 0.5642, P = 0.1634, respectively) (Table 5.4). In contrast, significant differences were found among populations within oceans for short-beaked populations (ϕ_{SC} = 0.10724, P = 0.0092) (Table 5.4).

Table 5.2. Indices of genetic diversity and neutrality tests for common dolphin populations: *N*, number of individuals sequenced; *Nh*, number of haplotyopes; *h*, haplotype diversity; π, nucleotide diversity; *D*, Tajima's *D*; *Fs*, Fu's *Fs*.

Marker	Statistics	Short-beaked							Long-beaked			*tropicalis*-form
		NEATL	CEATL	NWATL	NEPAC	SWPAC_AUS	SWPAC_NZ	SEIND	NEPAC	SEATL	SWATL	WIND
mtDNA	*N*	63	21	27	26	41	40	27	40	26	7	25
	Hd	0.911	0.833	0.960	0.991	0.948	0.988	0.969	0.755	0.809	0.286	0.587
	pi	0.00478	0.00273	0.00500	0.00937	0.00691	0.00871	0.00531	0.0038	0.0056	0.00260	0.00522
	Tajima's *D*	-1.3688	-1.1638	-0.3884	-1.71595	-1.71594	-1.72701	-1.04668	-0.18415	-0.71714	-1.60974*	-1.19765
	Fu's Fs	-8.756***	-2.512*	-6.958***	-10.521***	-16.185***	-14.786***	-6.418**	0.5340	-0.2440	4.273	2.8390
	R^2	0.0592	0.0842	0.1012	0.0540	0.0494	0.0546	0.0878	0.1069	0.0981	0.3499	0.0978
CHRNA1	*N*	9	10	9	11	6	10	5	11	7	7	5
	Hd	0.837	0.821	0.745	0.848	0	0.889	0.711	0.623	0.626	0	0.867
	pi	0.00426	0.00451	0.00407	0.00538	0	0.00721	0.00563	0.00423	0.00206	0	0.00596
	Tajima's *D*	0.3499	0.6539	-0.8215	-0.23277	-	0.71238	1.95339	0.49867	-0.52939	-	0.42681
	Fu's Fs	-2.108	-2.769*	-3.475*	-4.110**	-	-3.145**	2-146	0.229	-0.959	-	-3.347*
	R2	0.1572	0.1625	0.1015	0.1215	-	0.1626	0.2667	0.1567	0.1428	-	0.1749
PLP	*N*	9	10	9	11	6	10	5	11	7	7	5
	Hd	0.366	0	0.209	0.801	0.742	0.537	0.356	0.593	0.495	0	0
	pi	0.00047	0	0.00027	0.00166	0.00121	0.00077	0.00046	0.00088	0.00138	0	0
	Tajima's *D*	0.4881	-	-0.5290	-0.67006	-1.02271	-0.79238	0.01499	-0.45042	0.39778	-	-
	Fu's Fs	0.796	-	-0.011	-1.348	-2.059	-1.177	0.417	-0.788	-0.281	-	-
	R2	0.1830	-	0.1046	0.1058	0.1201	0.1197	0.1778	0.1202	0.1777	-	-
Del_12	*N*	9	10	9	11	6	10	5	11	7	7	5
	Hd	0.810	0.842	0.719	0.472	0.879	0.642	0.711	0	0.846	0.824	0.689
	pi	0.00264	0.00294	0.00250	0.00121	0.00227	0.00097	0.00158	0	0.00236	0.00285	0.00186
	Tajima's *D*	0.1267	-0.2568	-0.0586	-0.97093	-0.32373	0.89508	-1.13610	-	0.00646	0.12331	1.45356
	Fu's Fs	-1.270	-0.7210	-0.461	-1.072	-1.585	-0.575	-0.384	-	-2.250	-1.6830	1.2930
	R2	0.1422	0.1272	0.1506	0.0975	0.1480	0.1947	0.1892	-	0.1459	0.1541	0.2481
Del_15	*N*	9	10	9	11	6	10	5	11	7	7	5
	Hd	0.889	0	0	0	0	0	0	0.727	0.758	0	0
	pi	0.00216	0	0	0	0	0	0	0.00341	0.00355	0	0
	Tajima's *D*	0.2846	-	-	-	-	-	-	0.09806	0.74965	-	-
	Fu's Fs	-1.700	-	-	-	-	-	-	-0.284	-1.240	-	-
	R2	0.1573	-	-	-	-	-	-	0.1412	0.1760	-	-
Del_17	*N*	9	10	9	11	6	10	5	11	7	7	5
	Hd	0.693	0	0	0	0	0	0	0	0	0	0.2
	pi	0.0018	0	0	0	0	0	0	0	0	0	0.0006
	Tajima's *D*	0.1360	-	-	-	-	-	-	-	-	-	-1.40085
	Fu's Fs	-0.762	-	-	-	-	-	-	-	-	-	0.5860
	R2	0.1495	-	-	-	-	-	-	-	-	-	0.3000

*$P < 0.05$, **$P < 0.01$, ***$P < 0001$

Table 5.3. Pairwise ϕ_{ST} values obtained for the cytochrome *b* gene for the different putative populations analysed in this study.

	Sb_NEATL	Sb_CEATL	Sb_NWATL	Sb_NEPAC	Sb_SWPAC AUS	Sb_SWPAC NZ	Sb_SEIND	Lb_NEPAC	Lb_SEATL	Lb_SWATL	Tro_WIND
Sb_NEATL											
Sb_CEATL	0.0167										
Sb_NWATL	0.1811***	0.2504***									
Sb_NEPAC	0.19578***	0.22727***	0.12379***								
Sb_SWPACAUS	0.09497***	0.11833***	0.16214***	0.05461***							
Sb_SWPACNZ	0.07988***	0.10587***	0.1153***	0.02863**	0.0048						
Sb_SEIND	0.04162**	0.07964**	0.18231***	0.12911***	0.0196	0.03593*					
Lb_NEPAC	0.56213***	0.59951***	0.48429***	0.41691***	0.47156***	0.43553***	0.52315***				
Lb_SEATL	0.04502*	0.07821**	0.17955***	0.20503***	0.10408***	0.08833***	0.03843*	0.55752***			
Lb_SWATL	0.12793*	0.2127**	0.20225*	0.23789***	0.18411***	0.12634**	0.14996**	0.59542***	0.0514		
Tro_WIND	0.5716***	0.63669***	0.25691***	0.38962***	0.49458***	0.4346***	0.56455***	0.60332***	0.56381***	0.60762***	

$^{*}P<0.05$; $^{**}P<0.01$, $^{***}P<0.001$

NEATL – Northeast Atlantic; CEATL – Central Eastern Atlantic; NWATL – Northwest Atlantic; SWATL – Southwest Atlantic; SWPAC – Southwest Pacific; NEPAC – Northeast Pacific; SEIND – Southeast Indian Ocean; WIND – Western Indian Ocean; Sb – short-beaked common dolphin; Lb – long-beaked common dolphin; Tro – *tropicalis* form.

Table 5.4. Results from analysis of molecular variance (AMOVA) of population structure in common dolphins obtained for the cytochrome *b* gene. Statistically significant values are highlighted in bold.

	Comparison levels	Source of variation	%variation	ϕ-statistics	*P*
mtDNA	All populations	Among populations	30.48		
		within populations	69.52	ϕ_{ST} = 0.30481	**0.0000**
	Short-beaked ocean basins	Among oceans (ATL, PAC, SEIND)	2.99	ϕ_{CT} = 0.02993	0.1682
		Among populations within oceans	7.92	ϕ_{SC} = 0.08167	**0.0000**
		Within populations	89.08	ϕ_{ST} = 0.10915	**0.0000**
	Short-beaked regions	Among regions (Western, Eastern)	10.72	ϕ_{CT} = 0.10724	**0.0092**
		Among populations within regions	0.54	ϕ_{SC} = 0.00608	0.1273
		Within populations	88.73	ϕ_{ST} = 0.11267	**0.0000**
	Long-beaked ocean basins	Among oceans (ATL, PAC, WIND)	56.48	ϕ_{CT} = 0.56482	0.1634
		Among populations within oceans	1.90	ϕ_{SC} = 0.04371	0.1285
		Within populations	41.62	ϕ_{ST} = 0.58384	**0.0000**

Table 5.5. Results from the analysis of molecular variance (AMOVA) of population structure in common dolphins obtained for the five nuclear loci. Statistically significant values are highlighted in bold.

Marker	Comparison levels	Source of variation	% variation	Fixation indices	*F*-statistics	*P*
Del_12	All populations	Among populations	17.55			
		within populations	82.45	ϕST	0.1755	**0.000**
	Short-beaked ocean basins	Among oceans (ATL, PAC, SEIND)	9.75	ϕCT	0.0975	**0.045**
		Among populations within oceans	2.85	ϕSC	0.0316	0.100
		Within populations	87.4	ϕST	0.1260	**0.000**
	Short-beaked regions	Among regions (Western, Eastern)	10.69	ϕCT	0.1069	**0.046**
		Among populations within regions	-0.01	ϕSC	-0.0001	0.466
		Within populations	89.32	ϕST	0.1068	**0.000**
	Long-beaked ocean basins	Among oceans (ATL, PAC, WIND)	-23.97	ϕCT	-0.2397	0.834
		Among populations within oceans	52.31	ϕSC	0.4220	**0.000**
		Within populations	71.66	ϕST	0.2834	**0.000**
Del_15	All populations	Among populations	19.96			
		within populations	80.04	ϕST	0.1996	**0.000**
	Short-beaked ocean basins	Among oceans (ATL, PAC, SEIND)	9.14	ϕCT	0.0914	0.029
		Among populations within oceans	-0.94	ϕSC	-0.0104	0.577
		Within populations	91.81	ϕST	0.0819	**0.039**
	Short-beaked regions	Among regions (Western, Eastern)	6.01	ϕCT	0.0601	0.245
		Among populations within regions	-0.16	ϕSC	-0.0017	0.418
		Within populations	94.16	ϕST	0.0584	**0.040**
	Long-beaked ocean basins	Among oceans (ATL, PAC, WIND)	-60.58	ϕCT	-60.5830	0.673
		Among populations within oceans	96.53	ϕSC	0.6012	**0.000**
		Within populations	64.05	ϕST	0.3595	**0.000**
Del_17	All populations	Among populations	21.04			
		within populations	78.96	ϕST	0.2104	**0.000**
	Short-beaked ocean basins	Among oceans (ATL, PAC, SEIND)	14.41	ϕCT	0.1441	**0.014**
		Among populations within oceans	4.99	ϕSC	0.0583	**0.026**
		Within populations	80.6	ϕST	0.1940	**0.000**
	Short-beaked regions	Among regions (Western, Eastern)	8.81	ϕCT	0.0881	0.113
		Among populations within regions	7.28	ϕSC	0.0799	**0.038**
		Within populations	83.9	ϕST	0.1610	**0.000**
	D. capensis ocean basins	Among oceans (ATL, PAC, WIND)	33.2	ϕCT	0.3320	0.169
		Among populations within oceans	9.1	ϕSC	0.1362	**0.023**
		Within populations	57.7	ϕST	0.4230	**0.000**

Table 5.5. cont. Results from the analysis of molecular variance (AMOVA) of population structure in common dolphins obtained for the five nuclear loci. Statistically significant values are highlighted in bold.

Marker	Comparison levels	Source of variation	% variation	Fixation indices	*F*-statistics	*P*
CHRNA1	All populations	Among populations	22.24			
		within populations	77.76	ϕST	0.2224	**0.000**
	Short-beaked ocean basins	Among oceans (ATL, PAC, SEIND)	2.81	ϕCT	0.0281	0.201
		Among populations within oceans	6.23	ϕSC	0.0641	**0.010**
		Within populations	90.96	ϕST	0.0904	**0.001**
	Short-beaked regions	Among regions (Western, Eastern)	0.63	ϕCT	0.0063	0.431
		Among populations within regions	7.63	ϕSC	0.0768	**0.048**
		Within populations	91.74	ϕST	0.0826	**0.001**
	Long-beaked ocean basins	Among oceans (ATL, PAC, WIND)	-44.64	ϕCT	-0.4465	1.000
		Among populations within oceans	92.14	ϕSC	0.6370	**0.000**
		Within populations	52.5	ϕST	0.4750	**0.000**
PLP	All populations	Among populations	16.28			
		within populations	83.72	ϕST	0.1628	**0.000**
	Short-beaked ocean basins	Among oceans (ATL, PAC, SEIND)	1.78	ϕCT	0.0178	0.271
		Among populations within oceans	4.17	ϕSC	0.0425	**0.045**
		Within populations	94.05	ϕST	0.0595	**0.005**
	Short-beaked regions	Among regions (Western, Eastern)	-0.03	ϕCT	-0.0003	0.541
		Among populations within regions	5.43	ϕSC	0.0543	0.005
		Within populations	94.6	ϕST	0.0541	0.042
	Long-beaked ocean basins	Among oceans (ATL, PAC, WIND)	28.77	ϕCT	0.2877	0.333
		Among populations within oceans	10.13	ϕSC	0.1423	0.104
		Within populations	61.09	ϕST	0.3891	**0.000**

For the nuclear markers, although results varied slightly across markers, overall, the main pattern of differentiation obtained supported a differentiation among oceans for the short-beaked but not for the long-beaked morphotype. In addition, differentiation within each ocean for the short-beaked morphotype was also supported (Table 5.5).

The median-joining network reconstructions for the cytochrome *b* gene, the five nuclear loci and the nuclear multi-locus phylogeny resulted in complex phylogeographic patterns and no reciprocal monophyly of species or morphotypes (Figures 5.2, 5.3 and S5.1, respectively). For the more resolving mitochondrial network, most high frequency haplotypes are clustered together and were sampled mostly in the Atlantic Ocean (Figure 5.2). This includes short-beaked populations inhabiting the North Atlantic and long-beaked populations inhabiting the South Atlantic. Some of these haplotypes were also sampled in short-beaked populations from the Indo-Pacific region (SWPAC_AUS, SWPAC_NZ and SEIND), one haplotype was sampled in the short-beaked dolphins from the

Northeast Pacific (NEPAC) and one from the *tropicalis* population (WIND). We herein loosely refer to these as the Atlantic/Indo-Pacific cluster. Importantly, several single frequency haplotypes from this cluster are nested with the central haplotypes, forming a star phylogeny pattern normally seen in population genealogies impacted by recent demographic expansions. In marked contrast to the cluster described above, most haplotypes sampled in the Southwest and Northeast Pacific short-beaked populations have long branches, are not arranged in star phylogenies and show relatively high divergence from the Atlantic/Indo-Pacific cluster. A highly divergent cluster is located at a tip of the network and contains mostly haplotypes that were sampled in short and long-beaked populations from the Atlantic and Indian Oceans but also short-beaked individuals from the Pacific Ocean. Long-beaked dolphins from NEPAC show a distinct pattern from those described above. Here, most haplotypes form a divergent lineage composed of relatively related haplotypes, except for two that cluster with the *tropicalis* population from WIND. The remaining haplotypes that have been sampled in the *tropicalis* population from WIND are found in the Atlantic/Indo-Pacific cluster, and also in the highly divergent cluster (Figure 5.2).

Although the median-joining networks based on nuclear loci are less variable and generally less informative than that obtained with mtDNA, they provide strong support for a recent evolutionary history of the genus and historical gene flow among morphotypes and species (Figure 5.3). Patterns of high frequency haplotypes that show wide distribution closely linked to low frequency haplotypes are seen in most networks, especially in PLP, where the central haplotype was sampled in all short-beaked and long-beaked populations (Figure 5.3b). The network obtained for locus Del_15 shows two haplogroups with high frequency haplotypes, which differ in the placement of short-beaked and long-beaked populations from the Atlantic Ocean (Figure 5.3d). Long-beaked haplotypes from SWATL, SEATL and NEPAC are preferentially located at the tips. This same pattern is seen in the Del_12 network (Figure 5.3c). Here, the geographic distribution of haplotypes is not so clear with the exception of the cluster of long-beaked individuals located at the tip.

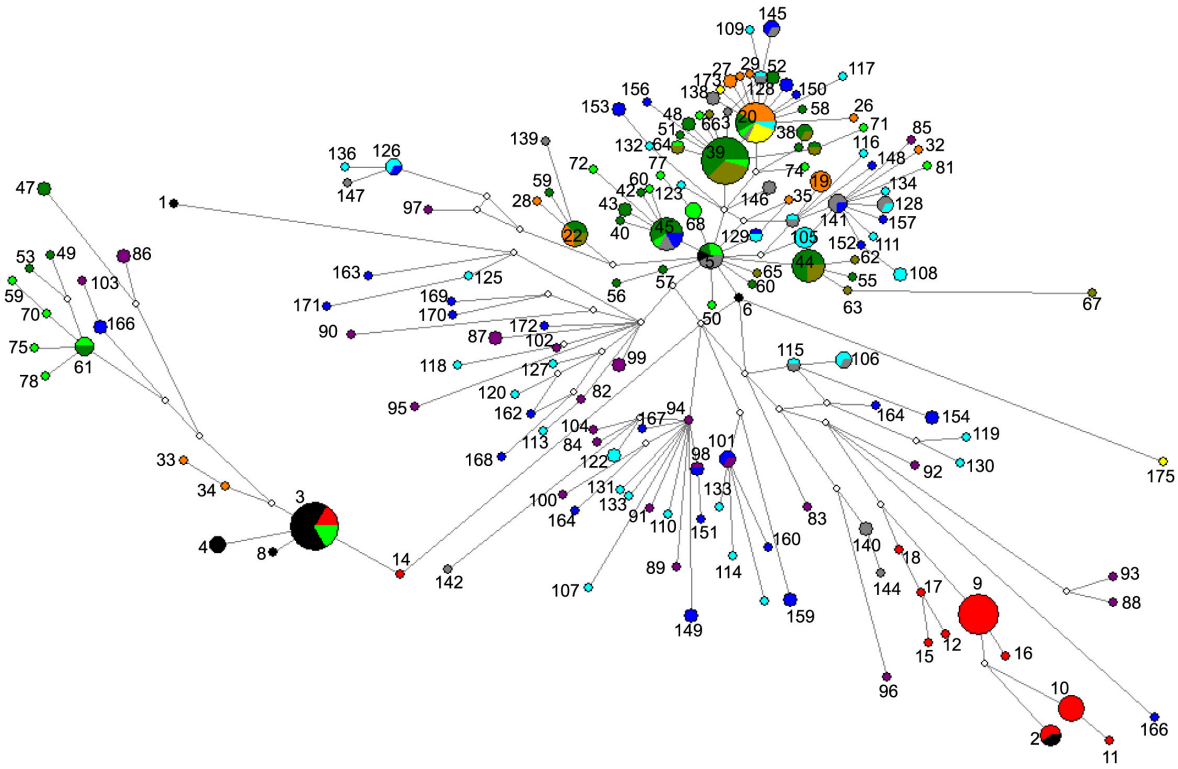


Figure 5.2. Median-joining network of cytochrome *b* gene haplotypes of common dolphins. Circle size is proportional to the number of individuals exhibiting the corresponding haplotype. Each population within each haplotype is coloured according to the legend. Length of lines is proportional to the number of mutational steps separating haplotypes. White circles indicate missing, intermediate haplotypes. ● Western Indian Ocean *tropicalis* form; ● short-beaked Northeast Atlantic; ● short-beaked Northwest Atlantic; ● short-beaked Central Eastern Atlantic; ● short-beaked Northeast Pacific; ● short-beaked Southwest Pacific Australia; ● short-beaked Southwest Pacific New Zealand; ● short-beaked Southeast Indian; ● long-beaked Northeast Pacific; ● long-beaked Southeast Atlantic; ● long-beaked Southwest Atlantic.

Moreover, the central haplotypes in this network, which are located intermediate between the most high frequency haplotypes, have mostly been sampled in short-beaked and long-beaked populations from the Pacific Ocean. Networks obtained for CHRNA1 and Del_17 do not show clear geographical patterns, with high frequency haplotypes located at the centre that have been sampled in most common dolphin populations, and lower frequency haplotypes located at the tips (Figures 5.3a and 5.3e, respectively). Finally, the multi-locus nuclear phylogeny obtained in POFAD supports the close evolutionary relationships and putative young age of *Delphinus* lineages, with branches clustering alleles with no geographical or taxonomical association (Supplementary Material, Figure S5.1).

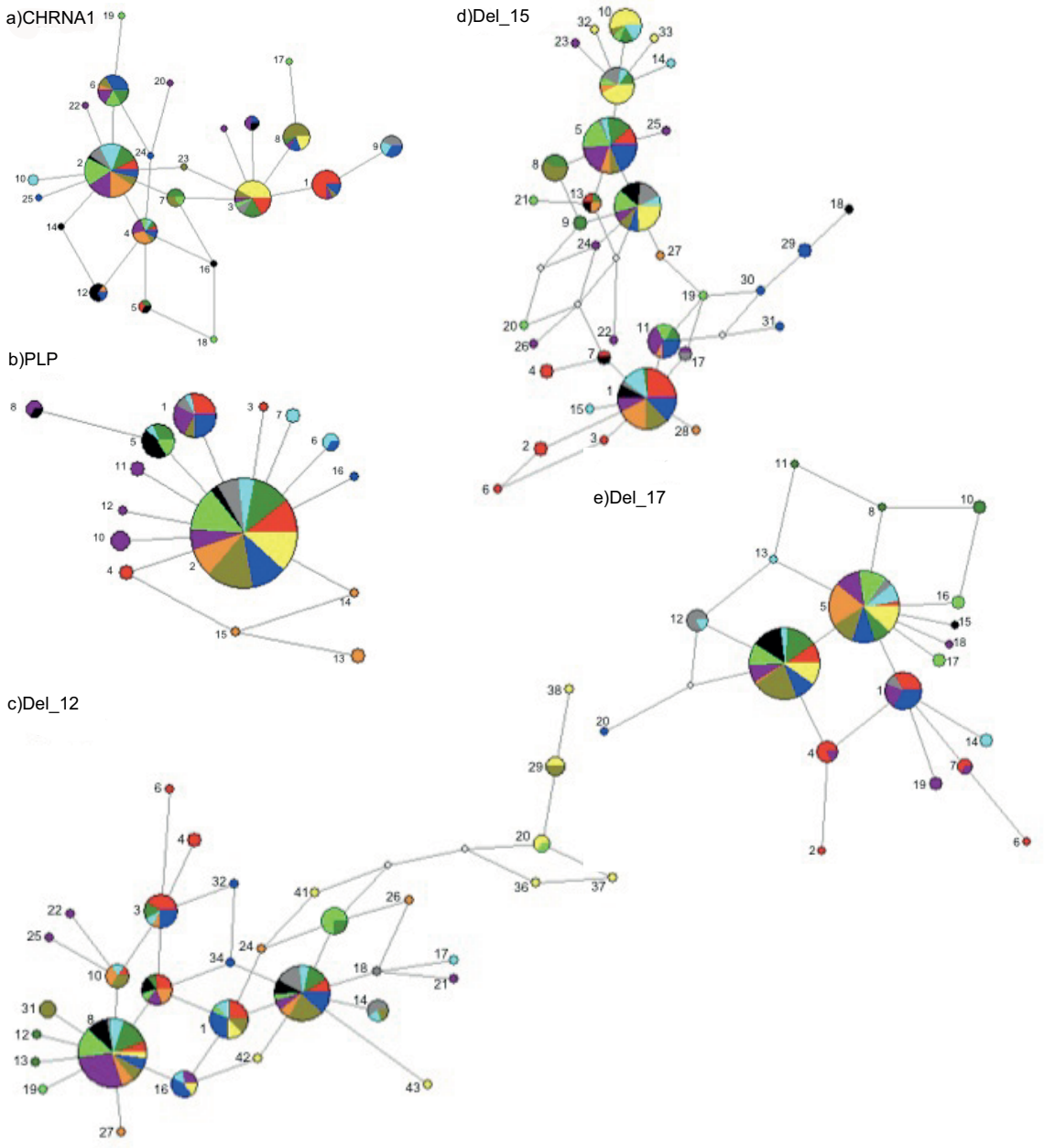


Figure 3. Median-joining networks of nuclear gene haplotypes of common dolphins: a) CHRNA1, b) PLP, c) Del_12, d) Del_15, e) Del_17. Circle size is proportional to the number of individuals exhibiting the corresponding haplotype. Each population within each haplotype is coloured according to the legend. Length of lines is proportional to the number of mutational steps separating haplotypes. White circles indicate missing intermediate haplotypes. ● Western Indian Ocean *tropicalis* form; ● short-beaked Northeast Atlantic; ● short-beaked Northwest Atlantic; ● short-beaked Central Eastern Atlantic; ● short-beaked Northeast Pacific; ● short-beaked Southwest Pacific Australia; ● short-beaked Southwest Pacific New Zealand; ● short-beaked Southeast Indian; ● long-beaked Northeast Pacific; ● long-beaked Southeast Atlantic; ● long-beaked Southwest Atlantic.

5.4.3. *Divergence time estimates*

Estimates obtained in the program MDIV varied from 0.021 million years (Ma) (between short-beaked CEATL and NEATL) to 1.590 Ma (between short-beaked NEPAC and CEATL), suggesting that divergences within *Delphinus* took place within the Pleistocene period (Table 5.6). The divergence between short and long-beaked populations from NEPAC was estimated at 1.163 Ma. Within short-beaked populations, divergence time estimates between populations from different oceans were higher than between populations within the same ocean. Within long-beaked populations, the SEATL population diverged more recently than the NEPAC population. MDIV also generated estimates of the ancestral populations sizes (θ) and migration (M) for all pairwise comparisons (Table 5.6). Theta values varied from 2.59 to 37.83 with all pairwise comparisons with high theta values involving the short-beaked populations from NEPAC and SWPAC_NZ. Migration rates obtained with MDIV are generally consistent with ϕ_{ST} results, showing that levels of gene flow are higher within ocean basins (e.g. short-beaked populations from NEATL and CEATL or from SWPAC_AUS and SWPAC_NZ), and lower between populations inhabiting different ocean basins. Levels of gene flow were relatively low between the different morphotypes, with the exception of the long-beaked population from SWATL that appears to have recently separated from the short-beaked populations from the North Atlantic.

5.4.4. *Demography*

As expected based on visual assessment of nuclear and mtDNA genealogies, our statistical framework used to investigate demographic history (i.e. summary statistics, mismatch analysis and Bayesian skyline plots) revealed that most common dolphins populations do not conform with a model of constant size through time. The neutrality tests based on summary statistics (Fu's *Fs* and R^2) revealed population expansions in the cytochrome *b* for all short-beaked populations and for the long-beaked population from SWATL (Table 5.2). While this result was also obtained for some short-beaked populations in CHRNA1 and for the *tropicalis*-form, no signal of expansion was recovered in other nuclear loci (Table 5.2). Only one statistically significant negative value of Tajima's *D* was

obtained, for the cytochrome *b* gene, for the SWATL long-beaked population (Table 5.2). Mismatch analysis support demographic expansions for almost every short-beaked population analysed and for some of the long-beaked populations in both mitochondrial and nuclear markers (Supplementary Material, Table S5.4, Figure S5.2). As expected, estimates of demographic parameters varied across nuclear loci (Table S5.4). The Bayesian skyline plot analyses, which are coalescent-based and utilised the combined mitochondrial and nuclear datasets, recovered a more detailed picture of the historical demography of *Delphinus* (Figure 5.4). Here, differences were suggested between short-beaked populations from the Pacific and Atlantic Oceans (Figure 5.4). Pacific Ocean populations showed older (starting between 0.35 (for NEPAC) and 0.40 (for SWPAC_AUS and SWAC_NZ) Ma) and extreme demographic expansion after a long period of constant population, while those from the Atlantic showed more gradual population expansions that started between 0.3 and 0.07 Ma. These results obtained with the mismatch analysis resulted in much wider intervals, with the cytochrome *b* gene and the nuclear loci placing the time of expansion for short-beaked Atlantic populations between 0.035-0.242 and 0.297-9.958 Ma, respectively (Table S5.4). For short-beaked Pacific populations the estimated time of expansion for the same loci was between 0.347-0.402 Ma and 1.013-11.044 Ma, respectively (Table S5.4). Estimates of effective population size were also different between oceans, being higher for populations inhabiting the Pacific Ocean, and lower for populations inhabiting the Atlantic. These results are concordant with those based on MDIV (Table 5.6). Long-beaked common dolphin populations showed different demographic patterns. The SEATL population was the only showing a sign of expansion (Figure 5.4h). The NEPAC population showed a constant population size through time (Figure 5.4g), a result consistent with summary statistics and with most MMD analyses (Table S5.4, Figure S5.2), while the population from SWATL showed a population decline followed by a signal of recent population expansion (Figure 5.4i). The *tropicalis*-form population from WIND showed signs of an old population expansion that ended at around 0.2 MA, followed by a population decline that seems to be stabilizing (Figure 5.4j). All long-beaked populations showed effective population sizes comparable to those

of the short-beaked populations from the Atlantic and SE Indian Ocean. In summary, all analyses showed that common dolphins are a group with young coalescence and multiple localized demographic expansions.

Table 5.6. Divergence times between common dolphin populations obtained in MDIV. Maximum likelihood estimates of divergence times, effective population sizes and migration rates based on the mitochondrial cytochrome *b* gene. θ, effective population size; M (2N*m*); *T* (*t*/2N) . Time divergence values (in million years, Ma) are given for a mutation rate of 9.86 x 10^{-9} substitutions per nucleotide site per year and a generation time of 7 years.

Population comparison	θ	M	T	t (Ma)
Sb_NEATL / Sb_CEATL	6.61 (0.14)	36.20 (7.43)	0.02 (0.01)	0.021
Sb_NEATL / Sb_NWATL	8.89 (0.05)	2.36 (0.01)	0.13 (0.03)	0.183
Sb_NEATL / Lb_SEATL	3.95 (0.03)	0.89 (0.04)	0.19 (0.02)	0.119
Sb_NEATL / Sb_NEPAC	20.60 (0.53)	0.62 (0.03)	0.37 (0.14)	1.207
Sb_NEATL / Sb_SWPACAUS	12.55 (0.25)	0.59 (0.03)	0.30 (0.11)	0.596
Sb_NEATL / Sb_SWPACNZ	22.81 (0.21)	1.39 (0.06)	0.34 (0.12)	1.228
Sb_NEATL /Sb_SEIND	10.79 (0.20)	1.62 (0.06)	0.12 (0.02)	0.205
Sb_NEATL / Lb_NEPAC	7.00 (0.20)	0.02 (0.00)	0.91 (0.08)	1.009
Sb_NEATL / Tro_WIND	8.40 (0.20)	0.26 (0.02)	0.56 (0.05)	0.745
Sb_SWPAC_AUS / Sb_CEATL	11.28 (0.39)	1.12 (0.11)	0.27 (0.08)	0.482
Sb_SWPAC_AUS / Sb_NWATL	12.60 (0.14)	1.30 (0.05)	0.21 (0.06)	0.419
Sb_SWPAC_AUS / Lb_SEATL	10.68 (0.43)	1.85 (0.15)	0.21 (0.08)	0.355
Sb_SWPAC_AUS / Sb_NEPAC	21.04 (0.53)	4.22 (0.18)	0.10 (0.04)	0.333
Sb_SWPAC_AUS / Sb_SWPACNZ	19.65 (0.39)	20.96 (3.22)	0.04 (0.03)	0.124
Sb_SWPAC_AUS / Sb_SEIND	11.02 (0.09)	27.07 (3.57)	0.02 (0.00)	0.035
Sb_SWPAC_AUS / Lb_NEPAC	10.49 (0.19)	0.02 (0.00)	0.60 (0.01)	0.997
Sb_SWPAC_AUS / Tro_WIND	12.15 (0.26)	0.76 (0.07)	0.21 (0.01)	0.404
Sb_SWPAC_NZ / Sb_CEATL	22.72 (0.28)	2.38 (0.18)	0.11 (0.06)	0.396
Sb_SWPAC_NZ / Sb_NWATL	22.45 (0.72)	3.18 (0.39)	0.18 (0.13)	0.640
Sb_SWPAC_NZ /Lb_SEATL	13.34 (0.35)	2.76 (0.29)	0.10 (0.02)	0.211
Sb_SWPAC_NZ / Sb_NEPAC	37.83 (1.29)	6.20 (0.60)	0.10 (0.08)	0.599
Sb_SWPAC_NZ / Sb_SEIND	24.68 (0.24)	18.38 (3.34)	0.04 (0.01)	0.156
Sb_SWPAC_NZ / Lb_NEPAC	20.48 (0.44)	0.32 (0.04)	0.45 (0.02)	1.459
Sb_SWPAC_NZ / Tro_WIND	22.64 (0.49)	1.60 (0.25)	0.16 (0.02)	0.574
Lb_NEPAC / Sb_CEATL	4.35(0.13)	0.02 (0.00)	1.54 (0.08)	1.061
Lb_NEPAC / Sb_NWATL	5.00 (0.17)	0.20 (0.02)	1.44 (0.15)	1.140
Lb_NEPAC / Lb_SEATL	2.59 (0.04)	0.11 (0.01)	1.42 (0.11)	0.582
Lb_NEPAC / Sb_NEPAC	17.09 (0.64)	0.48 (0.02)	0.43 (0.06)	1.163
Lb_NEPAC / Sb_SEIND	8.06 (0.37)	0.02 (0.00)	0.91 (0.06)	1.161
Lb_NEPAC / Tro_WIND	2.95 (0.06)	0.54 (0.04)	2.04 (0.35)	0.953
Sb_NWATL / Sb_CEATL	7.27 (0.10)	1.38 (0.09)	0.32 (0.11)	0.378
Sb_NWATL / Lb_SEATL	4.26 (0.09)	1.19 (0.01)	0.18 (0.02)	0.121
Sb_NWATL / Sb_NEPAC	22.00 (0.62)	0.73 (0.04)	0.36 (0.01)	1.254
Sb_NWATL / Sb_SEIND	9.81 (0.23)	2.30 (0.24)	0.11 (0.03)	0.171
Sb_NWATL / Tro_WIND	6.50 (0.13)	0.67 (0.02)	0.49 (0.03)	0.504
Sb_SEIND / Sb_CEATL	8.22 (0.20)	1.35 (0.14)	0.17 (0.05)	0.221
Sb_SEIND / Lb_SEATL	5.08 (0.08)	1.48 (0.07)	0.13 (0.02)	0.105
Sb_SEIND / Sb_NEPAC	25.97 (0.48)	1.40 (0.16)	0.30 (0.04)	1.234
Sb_SEIND / Tro_WIND	9.13 (0.31)	0.42 (0.02)	0.58 (0.03)	0.838
Sb_NEPAC /Sb_CEATL	22.83 (0.40)	0.43 (0.03)	0.44 (0.03)	1.590
Sb_NEPAC / Lb_SEATL	12.89 (0.36)	0.64 (0.02)	0.41 (0.05)	1.411

Table 5.6. cont. Divergence times between common dolphin populations obtained in MDIV. Maximum likelihood estimates of divergence times, effective population sizes and migration rates based on the mitochondrial cytochrome *b* gene. θ, effective population size; M (2N*m*); *T* (*t*/2N) . Time divergence values (in million years, Ma) are given for a mutation rate of 9.86 x 10^{-9} substitutions per nucleotide site per year and a generation time of 7 years.

Population comparison	θ	M	T	t (Ma)
Lb_SWATL / Sb_NWATL	4.50 (0.20)	1.90 (0.38)	0.31 (0.06)	0.221
Lb_SWATL / Sb_SEIND	5.87 (0.13)	3.37 (0.48)	0.19 (0.09)	0.177
Lb_SWATL / Sb_NEPAC	21.74 (1.16)	0.42 (0.05)	0.36 (0.02)	1.239
Lb_SWATL / Lb_SEATL	2.66 (0.15)	1.66 (0.44)	0.24 (0.06)	0.101
Lb_SWATL / Sb_CEATL	2.89 (0.12)	13.66 (6.96)	0.14 (0.10)	0.064
Lb_SWATL / Tro_WIND	4.18 (0.08)	0.38 (0.04)	1.93 (0.19)	1.277
Sb_NWATL / Sb_CEATL	7.27 (0.10)	1.38 (0.09)	0.32 (0.11)	0.378
Sb_NWATL / Lb_SEATL	4.26 (0.09)	1.19 (0.01)	0.18 (0.02)	0.121
Sb_NWATL / Sb_NEPAC	22.00 (0.62)	0.73 (0.04)	0.36 (0.01)	1.254
Sb_NWATL / Sb_SEIND	9.81 (0.23)	2.30 (0.24)	0.11 (0.03)	0.171
Sb_NWATL / Tro_WIND	6.50 (0.13)	0.67 (0.02)	0.49 (0.03)	0.504
Sb_SEIND / Sb_CEATL	8.22 (0.20)	1.35 (0.14)	0.17 (0.05)	0.221
Sb_SEIND / Lb_SEATL	5.08 (0.08)	1.48 (0.07)	0.13 (0.02)	0.105
Sb_SEIND / Sb_NEPAC	25.97 (0.48)	1.40 (0.16)	0.30 (0.04)	1.234
Sb_SEIND / Tro_WIND	9.13 (0.31)	0.42 (0.02)	0.58 (0.03)	0.838
Sb_NEPAC /Sb_CEATL	22.83 (0.40)	0.43 (0.03)	0.44 (0.03)	1.590
Sb_NEPAC / Lb_SEATL	12.89 (0.36)	0.64 (0.02)	0.41 (0.05)	1.411
Sb_NEPAC / Tro_WIND	21.74 (0.82)	1.37 (0.15)	0.23 (0.02)	0.792
Lb_SEATL /Sb_CEATL	3.09 (0.06)	0.76 (0.03)	0.42 (0.04)	0.205
Lb_SEATL / Tro_WIND	3.79 (0.10)	0.36 (0.04)	0.45 (0.07)	0.270
Sb_CEATL / Tro_WIND	5.23 (0.21)	0.24 (0.01)	0.75 (0.08)	0.621
Lb_SWATL / Sb_NEATL	5.66 (0.16)	16.50 (3.62)	0.05 (0.06)	0.045
Lb_SWATL / Sb_SWPACAUS	10.68 (0.13)	3.55 (1.74)	0.28 (0.07)	0.473
Lb_SWATL / Sb_SWPACNZ	19.65 (0.46)	17.63 (1.77)	0.02 (0.01)	0.062
Lb_SWATL / Lb_NEPAC	2.73 (0.12)	0.08 (0.00)	1.34 (0.10)	0.579
Lb_SWATL / Sb_NWATL	4.50 (0.20)	1.90 (0.38)	0.31 (0.06)	0.221
Lb_SWATL / Sb_SEIND	5.87 (0.13)	3.37 (0.48)	0.19 (0.09)	0.177
Lb_SWATL / Sb_NEPAC	21.74 (1.16)	0.42 (0.05)	0.36 (0.02)	1.239
Lb_SWATL / Lb_SEATL	2.66 (0.15)	1.66 (0.44)	0.24 (0.06)	0.101
Lb_SWATL / Sb_CEATL	2.89 (0.12)	13.66 (6.96)	0.14 (0.10)	0.064
Lb_SWATL / Tro_WIND	4.18 (0.08)	0.38 (0.04)	1.93 (0.19)	1.277

5.5. DISCUSSION

This study used a multilocus dataset to reconstruct the history of evolutionary diversification and population dynamics of a group of cosmopolitan top marine predators, the common dolphins (genus *Delphinus*). Our analyses indicate that this widely distributed group is composed of very closely related lineages that show young age, rapid morphologic diversification and strong signatures of

regional demographic expansions. Here we present a biogeographic scenario accounting for the origin of common dolphins and propose that Pleistocene changes in climate and oceanography were the main drivers influencing the demography, dispersal and speciation in *Delphinus*.

5.5.1. *Origin, range expansion and speciation of common dolphins*

Genealogical relationships, estimates of effective population sizes and divergence times indicate that the ancestral common dolphin populations are in the Pacific Ocean (Figure 5.5). Evidence for this hypothesis comes from the unequal frequency and distribution of haplotypes from the Atlantic and the Pacific oceans in the mtDNA network, the higher estimates of effective population sizes and higher nucleotide diversities obtained for the short-beaked populations inhabiting the Pacific Ocean, namely for the NEPAC and SWPAC_NZ populations, and the older divergence times obtained for NEPAC. We therefore envision two possible scenarios for the origin of the genus *Delphinus* in the Pacific Ocean. The first would be common dolphins originating in the NEPAC during the Pleistocene, as suggested by the highest divergence time estimates obtained in MDIV. The North Pacific is considered a marine centre of evolutionary origin (Briggs 2003). It has originated biota able to transgress and successfully colonize the Arctic, Atlantic and South Pacific Oceans, and its biota has also remained permeable to invasions by taxa from other regions (Briggs 2003). A cooling of the tropical Pacific would then favour the dispersal across equatorial waters to the Southern Hemisphere (Lawrence *et al.* 2006; Lee & Poulsen 2005; Lindberg 1991). Alternatively, one could envisage a scenario with origin in the Southwest Pacific and subsequent migration towards the Northern Hemisphere. This would be possible since during the Pleistocence period migrations across equatorial waters were almost symmetrical (Lindberg 1991). Moreover, it has been suggested that the climatic oscillations that occurred during this period were not so severe in the Southern Hemisphere (Lee & Poulsen 2005), which would make this region more stable. Our results do not allow a clear distinction between which of these two scenarios would be more plausible.

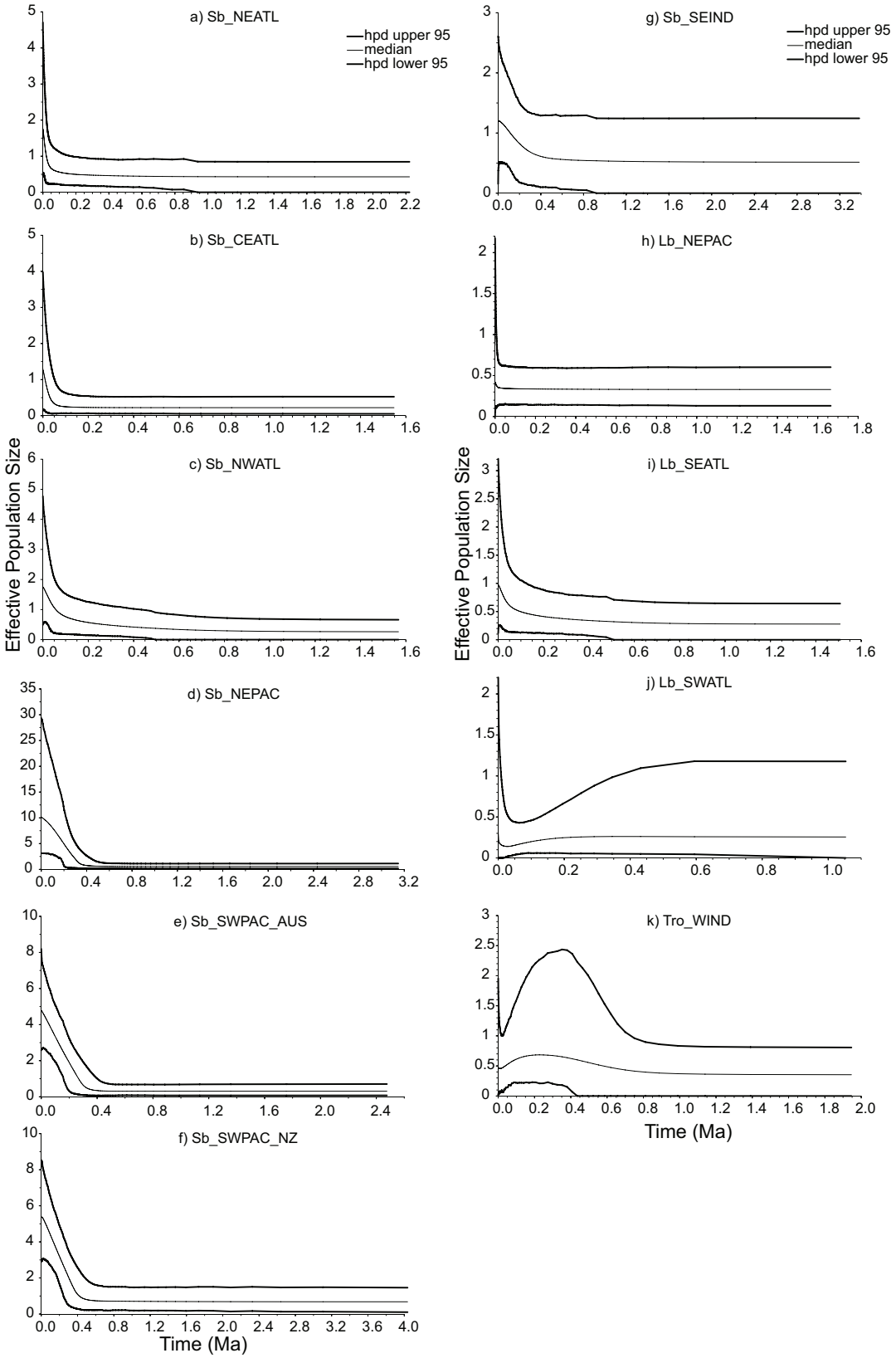


Figure 5.4. Extended Bayesian Skyline Plots showing changes in population size through time (Ma): a) Sb_NEATL, b) Sb_CEATL, c) Sb_NWATL, d) Sb_NEPAC, e) Sb_SWPA_AUS, f) Sb_SWPAC_NZ, g) Sb_SEIND, h) Lb_NEPAC, i) Lb_SEATL, j) Lb_SWATL, k) *tropicalis*_WIND.

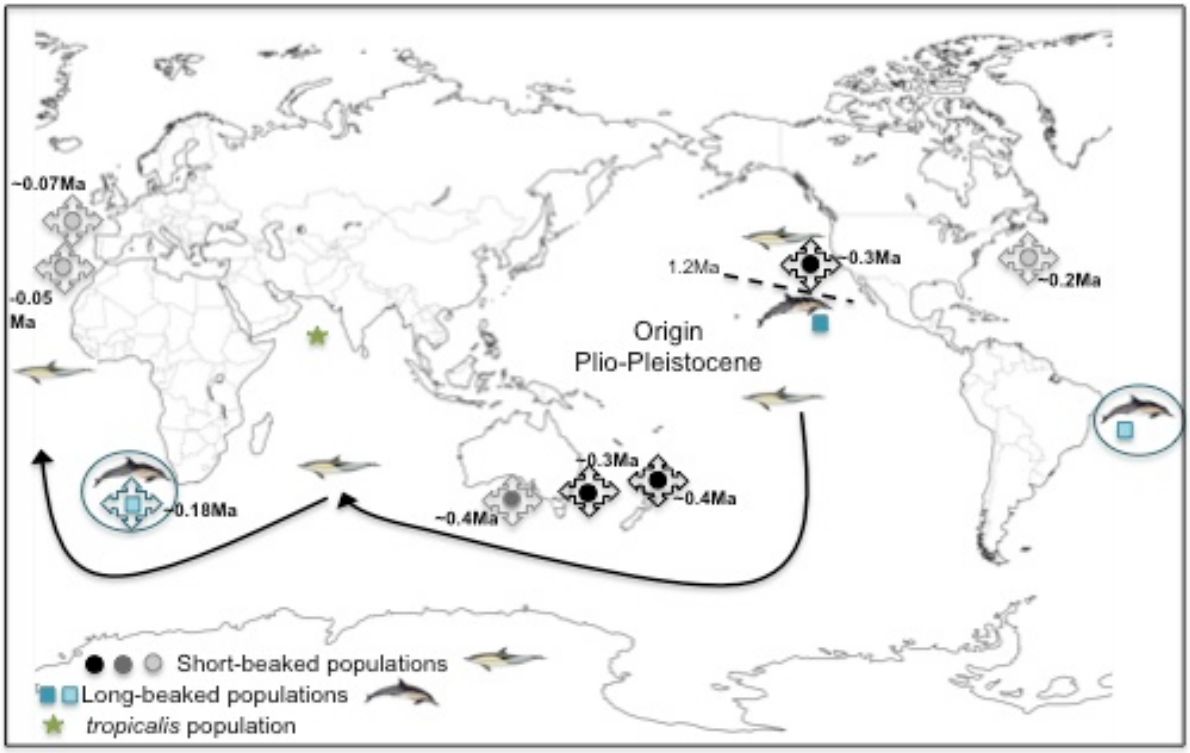


Figure 5.5. Map summarizing the proposed origin, expansion and speciation of the genus *Delphinus.* The different shades of grey and blue illustrate the older (darker) and younger (lighter) populations of the short and long-beaked morphotypes. Arrows indicate the route of colonization of the short-beaked morphotype. The dash line represents the time of speciation/origin of the long-beaked morphotype in the Northeast Pacific. Blue circles indicate the independent origin of the long-beaked populations in the Atlantic Ocean. Centered arrows represent population expansions with approximate times of expansion obtained with the Extended Bayesian Skyline Plot method in bold.

Our results further suggest that common dolphins would have then dispersed westerly, across the Indian Ocean and into the Atlantic Ocean around the tip of South Africa (Figure 5.5). The exact timing of this dispersal could not be estimated, but the presence of haplotypes from the South Pacific nested within those from the Atlantic Ocean supports this hypothesis (Figure 5.2). Moreover, Southeast Indian haplotypes (SEIND) are mostly clustered with those from the Atlantic, and not dispersed with South Pacific haplotypes. This same route of dispersal has been previously described for several marine organisms, from teleost fishes (e.g. Bremer *et al*. 1998) to marine turtles (Bowen *et al.* 1997) and sea birds (Avise *et al*. 2000) and has also been suggested for the dusky dolphin, *Lagenorhynchus obscurus* (Harlin-Cognato *et al.* 2007) and its prey, *Engraulis* sp. (Grant & Bowen 1998), and the species of *Stenella* (Perrin 2007). In fact, the phylogeography of dusky dolphins correlates to that of *Engraulis* sp., suggesting

that primary productivity and prey abundance have played a role in the species history (Harlin-Cognato *et al.* 2007). Common dolphins prey on the same small schooling fish and therefore it is likely that trophic changes have also played a role in the evolutionary history of the genus.

After dispersal of short-beaked populations, exploration of new habitats in more coastal waters could have led to the speciation and the origin of the long-beaked morphotype through feeding specialization. This likely occurred independently in different oceans, as previously suggested (Natoli *et al.* 2006). In the Northeast Pacific, the divergence of the two morphotypes is estimated to have occurred during the Pleistocene (1.2 Ma). This was after a period of maximum primary productivity in this region, which was caused by a cooling of the surface temperature (Lawrence *et al.* 2006) (Figure 5.5). The decrease in abundance of resources could have lead to the search for new habitat, leading to niche partitioning and the consequent evolution of the long-beaked morphotype. In the Atlantic Ocean, the long-beaked morphotype appears to have evolved much more recently, as indicated by younger estimates of divergence times and the position of haplotypes at the tips of mtDNA and nuclear genealogies (Figures 5.2, 5.3, 5.5 and S5.4).

5.5.2. *Phylogeography*

The existence of highly divergent mitochondrial clades in marine animals has been associated with scenarios of vicariance during the Pleistocene – a period in which temperature fluctuations temporarily impeded regional migrations (Bremer *et al.* 1998; Buonaccorsi *et al.* 2001; Graves & McDowell 1995; Martinez *et al.* 2006; Vinas *et al.* 2004). Secondary contact and subsequent unidirectional migration would result in contemporary asymmetrical distribution of mitochondrial clades (e.g. Bremer *et al.* 1998; Martinez *et al.* 2006; Peeters *et al.* 2004). The intermittent isolation between the Atlantic and Indo-Pacific Ocean basins during the Pleistocene has been suggested to explain phylogeographic patterns in large migratory bony fishes such as Atlantic big-eyed tuna (Bremer *et al.* 1998; Martinez *et al.* 2006), Atlantic bonito (Vinas *et al.* 2004) and swordfish (Bremer *et al.* 2005), as well and in hammerhead sharks (Duncan *et al.* 2006). A similar scenario of

vicariance and secondary contact could therefore account for the highly divergent lineage observed in the mtDNA network obtained for common dolphins, where the predominance of haplotypes is from the Atlantic and Indian Oceans.

A common phylogeographic pattern across many widespread teleost fishes and sharks is lower genetic diversity in populations inhabiting the Atlantic Ocean compared to those from the Pacific, as well as significant differences in haplotype frequencies between the two ocean basins (e.g. Castro *et al.* 2007; Duncan *et al.* 2006; Vinas *et al.* 2004). Our analyses have also disclosed such pattern among short-beaked common dolphins inhabiting the two ocean basins, suggesting a recent colonization of the Atlantic Ocean by taxa originating in more diverse ecosystems such as the Indo-West Pacific or the North Pacific Ocean (Briggs 2000, 2003). Moreover, temperature fluctuations in the North Pacific were not so drastic as in the North Atlantic during the Quaternary glaciations. Firstly, the Pacific basin is larger and therefore more climatic stable and, secondly, glaciations were more intense in the North Atlantic (Briggs 1974). These different climatic regimes may therefore account for the markedly different phylogeographic patterns obtained in marine organisms occurring in the Atlantic and in the Pacific Ocean basins.

5.5.3. *Historical demography*

Overall, despite some differences across loci and methods, we recovered strong signals of demographic expansion for all short-beaked populations and for the SEATL long-beaked population. Differences in estimates of effective population sizes and time of expansion between mitochondrial and nuclear markers are expected due to the inherent properties of these genomes. However, differences across nuclear loci are most likely explained by coalescent stochasticity, since different markers in the nuclear genome can have different times to their most recent common ancestor and different evolutionary rates (Edwards & Beerli 2000). Moreover, estimates of the time of population expansion should be interpreted cautiously, since departures from the mutation rate may be expected and could cause an error in the estimation. Nevertheless, values obtained can be considered a rough approximation at which period the historical events occurred.

Our estimates place all population expansions in the Pleistocene, though the actual ages since expansion differed between Atlantic and Pacific Ocean populations. Older expansions occurred in the Pacific Ocean, which is in agreement with these populations being the oldest. Nevertheless, it is possible that our nuclear dataset lacks power to recover even older expansions in the Pacific populations. Recurrent processes of extinction and expansion caused by climatic oscillations can cause shallow coalescences and explain the lack of signal in the data (e.g. Grant & Bowen 1998). During the Pleistocene period glaciations in the Northern Hemisphere caused temperature fluctuations that have influenced upwelling systems and consequently favoured the availability of resources, which could lead to population expansions (Lindberg 1991; Lawrence *et al.* 2006). This same pattern has also been described for other marine taxa in the Atlantic (e.g. Aboim *et al.* 2005; Larmuseau *et al.* 2009) and in the Pacific Ocean (e.g. Diaz-Jaimes *et al.* 2010; Lopez *et al.* 2010). The North Atlantic was however subject to more severe temperature cycles than the North Pacific during the Pleistocene, as mentioned above (Briggs 1974) and this may account for the different patterns of population expansion seen in the short-beaked populations inhabiting these two ocean basins. Nevertheless, long-beaked populations from NEPAC, SWATL and WIND did not appear to have experienced demographic expansions according to our results. While the NEPAC population showed a constant population size through time, the SWATL and the WIND population appear to have suffered population declines, which were dated from the late Pleistocene. We suggest that these declines could be related to the fact that these long-beaked populations were not so well adapted to the new coastal habitats explored.

5.5.4. *Population differentiation*

In short-beaked common dolphins, fixation indices showed a pattern of higher genetic differentiation across larger geographical scales (e.g. between populations inhabiting different oceans) and lower differentiation among populations inhabiting the same ocean basin. Although including a broader geographic sampling, these results support previous findings (Natoli *et al.* 2006).

Nonetheless, AMOVA analyses also supported higher levels of partition for populations within ocean basins for both mitochondrial and nuclear markers. As for long-beaked populations, high levels of differentiation were found at both mitochondrial and nuclear markers, between the populations inhabiting SEATL and NEPAC, similar to that reported for mitochondrial control region (Natoli *et al*. 2006). By analysing another long-beaked population inhabiting the South Atlantic, which also showed high levels of differentiation when compared to the other long-beaked populations, and by including nuclear loci in these analyses, we reinforce the hypothesis that the long-beaked morphotype has originated independently in different regions and recommend that the taxonomic status of these populations be reviewed. Although the presently recognized two species may prove to be invalid, it seems unlikely, despite their close genetic relationship that all ecologically and morphologically distinct *Delphinus* populations belong to the same species.

5.6. CONCLUSION

Using multilocus sequence data from a global sample and analyses based on coalescent and traditional statistical methods, we show that the phylogeographic history, historical demography and local adaptation of common dolphins have been likely influenced by the Pleistocene climatic oscillations. Studies demonstrating the impacts of past environmental changes on the demography and speciation of widely distributed top marine predators are very rare (e.g. Lamurseau *et al.* 2009). Our study highlights the potential role of ongoing climate change on the distribution and abundance of top marine predators, such as those reported over contemporary time-scales for pelagic birds (e.g. Veit *et al*. 1997*)*. Changes in population dynamics of marine predators can have massive ecological consequences, including ecosystem-level transformations in terms of restructuring of trophic cascades associated with declines of top predators (Myers *et al.* 2007). Furthermore, we provided insights into the evolutionary history of the genus *Delphinus,* showing that the route of dispersal into the Atlantic Ocean coincides with a biogeographical model proposed for other marine organisms such as teleosts, turtles, sharks and sea birds.

5.7. ACKNOWLEDGEMENTS

We would like to acknowledge all who kindly provided tissue samples: Jennifer Learmonth (SAC-Scottish Agricultural College and DEFRA), Luca Mirimin (University of Cork, Ireland), Marisa Ferreira (Sociadede Portuguesa de Vida Selvagem), S. Siciliano and J.F. Moura (Marine Mammal Group Study, GEMM-Lagos and FIOCRUZ, Rio de Janeiro, Brazil), M.C. Oliveira (Instituto Oceanográfico da Universidade de São Paulo), Robert Baldwin, Gianna Minton and Tim Collins (Envrionment Socity of Oman). A. R. Amaral acknowledges a PhD grant (SFR/BD/27245/2006) from Fundação para a Ciência e Tecnologia (Portugal). Sample collection in Australia was funded by Macquarie University through research grants to L. Möller, L. Beheregaray and K. Bilgmann. Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation.

5.8. REFERENCES

Aboim MA, Menezes GM, Schlitt T, Rogers AD (2005) Genetic structure and history of populations of the deep-sea fish *Helicolenus dactylopterus* (Delaroche, 1809) inferred from mtDNA sequence analysis. *Molecular Ecology* **14**, 1343-1354.

Amaral AR, Sequeira M, Cedeira-Martínez J, Coelho MM (2007) New insights on population genetic structure of *Delphinus delphis* from the Northeast Atlantic and phylogenetic relationships within the genus inferred from two mitochondrial markers. *Marine Biology* **151**, 1967-1976.

Amaral AR, Silva MC, Möller LM, Beheregaray LB, Coelho MM (2010) Anonymous nuclear markers for cetacean species. *Conservation Genetics* **11**, 1143-1146.

Arnason U, Gullberg A, Janke A (2004) Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* **333**, 27-34.

Au DWK, Perryman WL (1985) Dolphin habitats in the Eastern Tropical Pacific. *Fishery Bulletin* **83**, 623-643.

Avise JC (2000) *Phylogeography: The History and Formation of Species.* Harvard

University Press, Cambridge, Massachusetts.

Avise JC, Nelson WS, Bowen BW, Walker D (2000) Phylogeography of colonially nesting seabirds, with special reference to global matrilineal patterns in the sooty stern (*Sterna fuscata*). *Molecular Ecology* **9**, 1783-1792.

Ballance LT, Pitman RL, Fiedler PC (2006) Oceanographic influences on seabirds and cetaceans of the Eastern Tropical Pacific: A review. *Progress in Oceanography* **69**, 360-390.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Beheregaray LB (2008) Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology* **17**, 3754-3774.

Bilgmann K, Möller LM, Harcourt RG, Gales R, Beheregaray LB (2008) Common dolphins subject to fisheries impacts in southern Australia are genetically differentiated: implications for conservation. *Animal Conservation* **11**, 518-528.

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR (2001) Phylogeography of the trumpetfishes (Aulostomus): Ring species complex on a global scale. *Evolution* **55**, 1029-1039.

Bowen BW, Clark AM, Abreu-Grobois FA, Chaves A, Reichart, HA, Ferl RJ (1997) Global phylogeography of the ridley sea turtles (*Lepidochelys* spp.) as inferred from mitochondrial DNA sequences. *Genetica* **101**, 179-189.

Bowen BW, Grant WS (1997) Phylogeography of the sardines (*Sardinops* spp): Assessing biogeographic models and population histories in temperate upwelling zones. *Evolution* **51**, 1601-1610.

Bremer JRA, Stequert B, Robertson NW, Ely B (1998) Genetic evidence for inter-oceanic subdivision of bigeye tuna (*Thunnus obesus*) populations. *Marine Biology* **132**, 547-557.

Bremer JRA, Vinas J, Mejuto J, Ely B, Pla C (2005) Comparative phylogeography of Atlantic bluefin tuna and swordfish: The combined effects of vicariance,

secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Molecular Phylogenetics and Evolution* **36**, 169-187.

Briggs JC (1974) *Marine Zoogeography.* McGraw-Hill, USA.

Briggs JC (2000) Centrifugal speciation and centres of origin. *Journal of Biogeography* **27**, 1183-1188.

Briggs JC (2003) Marine centres of origin as evolutionary engines. *Journal of Biogeography* **30**, 1-18.

Brown DM, Brenneman RA, Koepfli KP*,* Pollinger JP, Mila B, Georgiadis NJ, Louis EE, Grether GF, Jacobs DK, Wayne RK (2007) Extensive population genetic structure in the giraffe. *Bmc Biology* **5**, 13.

Buonaccorsi VP, McDowell JR, Graves JE (2001) Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). *Molecular Ecology* **10**, 1179-1196.

Caballero S, Jackson J, Mignucci-Giannoni AA*,* Barrios-Garrido H, Beltran-Pedreros S, Montiel-Villalobos MG, Robertson KM, Baker CS (2008) Molecular systematics of South American dolphins *Sotalia*: Sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. *Molecular Phylogenetics and Evolution* **46**, 252-268.

Carling MD, Brumfield RT (2007) Gene sampling strategies for multi-Locus population estimates of genetic diversity (theta). *Plos One* **2**, e160.

Castro ALF, Stewart BS, Wilson SG*,* Hueter RE, Meekan MG, Motta PJ, Bowen BW, Karl SA (2007) Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). *Molecular Ecology* **16**, 5183-5192.

Diaz-Jaimes P, Uribe-Alcocer M, Rocha-Olivares A*,* Garcia-de-Leon FJ, Nortmonn P, Durand JD (2010) Global phylogeography of the dolphinfish (*Coryphaena hippurus*): The influence of large effective population size and recent dispersal on the divergence of a marine pelagic cosmopolitan species. *Molecular Phylogenetics and Evolution* **57**, 1209-1218.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7**, 214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185-1192.

Duncan KM, Martin AP, Bowen BW, De Couet HG (2006) Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). *Molecular Ecology* **15**, 2239-2251.

Edwards SV, Beerli P (2000) Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839-1854.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* **23**, 691-700.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.

Grant WS, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. *Journal of Heredity* **89**, 415-426.

Graves JE (1998) Molecular insights into the population structures of cosmopolitan marine fishes. *Journal of Heredity* **89**, 427-437.

Graves JE, McDowell JR (1995) Inter-ocean genetic-divergence of isttiophorid bullfishes. *Marine Biology* **122**, 193-203.

Harlin-Cognato AD, Markowitz T, Wuersig B, Honeycutt RL (2007) Multi-locus phylogeography of the dusky dolphin (*Lagenorhynchus obscurus*): Passive dispersal via the west-wind drift or response to prey species and climate change?

Bmc Evolutionary Biology **7**, 17.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *Bmc Evolutionary Biology* **8**, 289.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Joly S, Bruneau A (2006) Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: An example from Rosa in North America. *Systematic Biology* **55**, 623-636.

Kingston SE, Rosel PE (2004) Genetic differentiation among recently diverged delphinid taxa determined using AFLP markers. *Journal of Heredity* **95**, 1-10.

Larmuseau MHD, Van Houdt JKJ, Guelinckx J, Hellemans B, Volckaert FAM (2009) Distributional and demographic consequences of Pleistocene climate fluctuations for a marine demersal fish in the North-eastern Atlantic. *Journal of Biogeography* **36**, 1138-1151.

Lawrence KT, Liu ZH, Herbert TD (2006) Evolution of the Eastern Tropical Pacific through Plio-Pleistocene glaciation. *Science* **312**, 79-83.

LeDuc RG (2009) Delphinids, Overview. In: *Encyclopedia of Marine Mammals* (eds. Perrin WF, Wursig B, Theweissen JGM), pp. 298-302. Academic Press, New York.

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.

Lee JY, Edwards SV (2008) Divergence across Australia's Carpentarian Barrier: Statistical phylogeography of the red-Backed fairy wren (*Malurus melanocephalus*). *Evolution* **62**, 3117-3134.

Lee SY, Poulsen CJ (2005) Tropical Pacific climate response to obliquity forcing in the Pleistocene. *Paleoceanography* **20**, 10.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Lindberg DR (1991) Marine biotic interchange between the Northern and Southern Hemispheres. *Paleobiology* **17**, 308-324.

Lipps JH, Mitchell E (1976) Model for the adaptive radiations and extinctions of pelagic marine mammals. *Paleobiology* **2**, 147-155.

Lopez MD, Alcocer MU, Jaimes PD (2010) Phylogeography and historical demography of the Pacific sierra mackerel (*Scomberomorus sierra*) in the eastern Pacific. *Bmc Genetics* **11**, 34.

Lyons LA, Kehler JS, O'Brien SJ (1999) Development of comparative anchor tagged sequences (CATS) for canine genome mapping. *Journal of Heredity* **90**, 15-26.

Martinez P, Gonzalez EG, Castilho R, Zardoya R (2006) Genetic diversity and historical demography of Atlantic bigeye tuna (*Thunnus obesus*). *Molecular Phylogenetics and Evolution* **39**, 404-416.

Mirimin L, Westgate A, Rogan E, Rosel P, Read A, Coughlan J, Cross T (2009) Population structure of short-beaked common dolphins (*Delphinus delphis*) in the North Atlantic Ocean as revealed by mitochondrial and nuclear genetic markers. *Marine Biology* **156**, 821-834.

Möller LM, Kerstin B, Charlton-Robb K, Beheregaray L (2008) Multi-gene evidence for a new bottlenose dolphin species in southern Australia. *Molecular Phylogenetics and Evolution* **49**, 674-681.

Möller LM, Valdez FP, Allen S, Bilgmann K, Corrigan S, Beheregaray LB (2011) Fine-scale genetic structure in short-beaked common dolphins (*Delphinus delphis*) along the East Australian Current. *Marine Biology* **158**, 113-126.

Murphy S, Winship A, Dabin W, Jepson PD, Deaville R, Reid RJ, Spurrier C, Rogan E, Lopez A, Gonzalez A, Read FL, Addink M, Silva M, Ridoux V, Learmonth JA, Pierce GJ, Northridge SP (2009) Importance of biological parameters in assessing the status of *Delphinus delphis*. *Marine Ecology-Progress Series* **388**, 273-291.

Myers RA, Baum JK, Shepherd TD, Powers SP, Peterson CH (2007) Cascading effects of the loss of apex predatory sharks from a coastal ocean. *Science* **315**, 1846-1850.

Natoli A, Cañadas A, Peddemors VM, Aguila A, Vaquero C, Fernadez-Piqueras P, Hoelzel AR (2006) Phylogeography and alpha taxonomy of the common dolphin (*Delphinus* sp.). *Journal of Evolutionary Biology* **19**, 943-954.

Natoli A, Peddemors VM, Hoelzel AR (2004) Population structure and speciation in the genus *Tursiops* based on microsatellite and mitochondrial DNA analyses. *Journal of Evolutionary Biology* **17**, 363-375.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov Chain Monte Carlo approach. *Genetics* **158**, 885-896.

Peeters FJC, Acheson R, Brummer GJA, de Ruitjer, WPM, Schneider RR, Ganssen GM, Ufkes E, Kroon D (2004) Vigorous exchange between the Indian and Atlantic Oceans at the end of the past five glacial periods. *Nature* **430**, 661-665.

Perrin WF (2007) The South Africa Species Gate. In: *Whales and dolphins of the Southern Africa Subregion* (ed. Best PB). Cambridge University Press, New York.

Perrin WF (2009) Common dolphins, *Delphinus delphis* and *D. capensis*. In: *Encyclopedia of Marine Mammals* (eds. Perrin WF, Wursig B, Thewissen JGM), pp. 255-259. Academic Press, New York.

Perrin WF, Mitchell ED, Mead JG, Caldwell DK, Caldwell MC, Vanbree PJH, Dawbin WH (1987) Revision of the spotted dolphins, *Stenella* Spp. *Marine Mammal Science* **3**, 99-170.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Querouil S, Freitas L, Cascao I, Alves F, Dinis A, Almeida JR, Prieto R, Borras S, Matos JA, Mendonca D, Santos RS (2010) Molecular insight into the population structure of common and spotted dolphins inhabiting the pelagic waters of the Northeast Atlantic. *Marine Biology* **157**, 2567-2580.

Rambaut A, Drummond AJ (2007) Tracer v.1.5, Available from http://beast.bio.ed.ac.uk/Tracer.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**, 2092-2100.

Rice WR (1989) Analysing tables of statisticals tests. *Evolution* **43**, 223-225.

Roca AL, Georgiadis N, Pecon-Slattery J, O'Brien SJ (2001) Genetic evidence for two species of elephant in Africa. *Science* **293**, 1473-1477.

Rogers AR, Harpending H (1992) Population-growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-569.

Rosel PE, Block BA (1996) Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius*. *Marine Biology* **125**, 11-22.

Rosel PE, Dizon AE, Heyning JE (1994) Genetic analysis of sympatric morphotypes of common dolphins (genus *Delphinus*). *Marine Biology* **119**, 159-168.

Slater GJ, Price SA, Santini F, Alfaro ME (2010) Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B-Biological Sciences* **277**, 3097-3104.

Smith JM (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34**, 126-129.

Steeman ME, Hebsgaard MB, Fordyce RE, Ho SYW, Rabosky DL, Nielsen R, Rahbek C, Glenner H, Sorensen MV, Willerslev E (2009) Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology* **58**, 573-585.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73**, 1162-1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978-989.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution* **13**, 510-524.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Teske PR, Winker H, McQuaid CD, Barker NP (2009) A tropical/subtropical biogeographic disjunction in southeastern Africa separates two evolutionarily significant units of an estuarine prawn. *Marine Biology* **156**, 1265-1275.

Veit RR, McGowan JA, Ainley DG, Wahls TR, Pyle P (1997) Apex marine predator declines ninety percent in association with changing oceanic climate. *Global Change Biology* **3**, 23-28.

Vilstrup JT, Ho SYW, Foote AD, Morin PA, Kreb D, Krutzen M, Parra GJ, Robertson KM, de Stephanis R, Verborgh P, Willerslev E, Orlando L, Gilbert MTP (2011) Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *Bmc Evolutionary Biology* **11**, 65.

Vinas J, Bremer JRA, Pla C (2004) Inter-oceanic genetic differentiation among albacore (*Thunnus alalunga*) populations. *Marine Biology* **145**, 225-232.

5.9. SUPPLEMENTARY MATERIAL

Figure S5.1. Neighbour-joining phylogram based on a combined distance matrix of all five nuclear loci generated using the program POFAD. Branches are coloured according to morphotype and geographical origin.

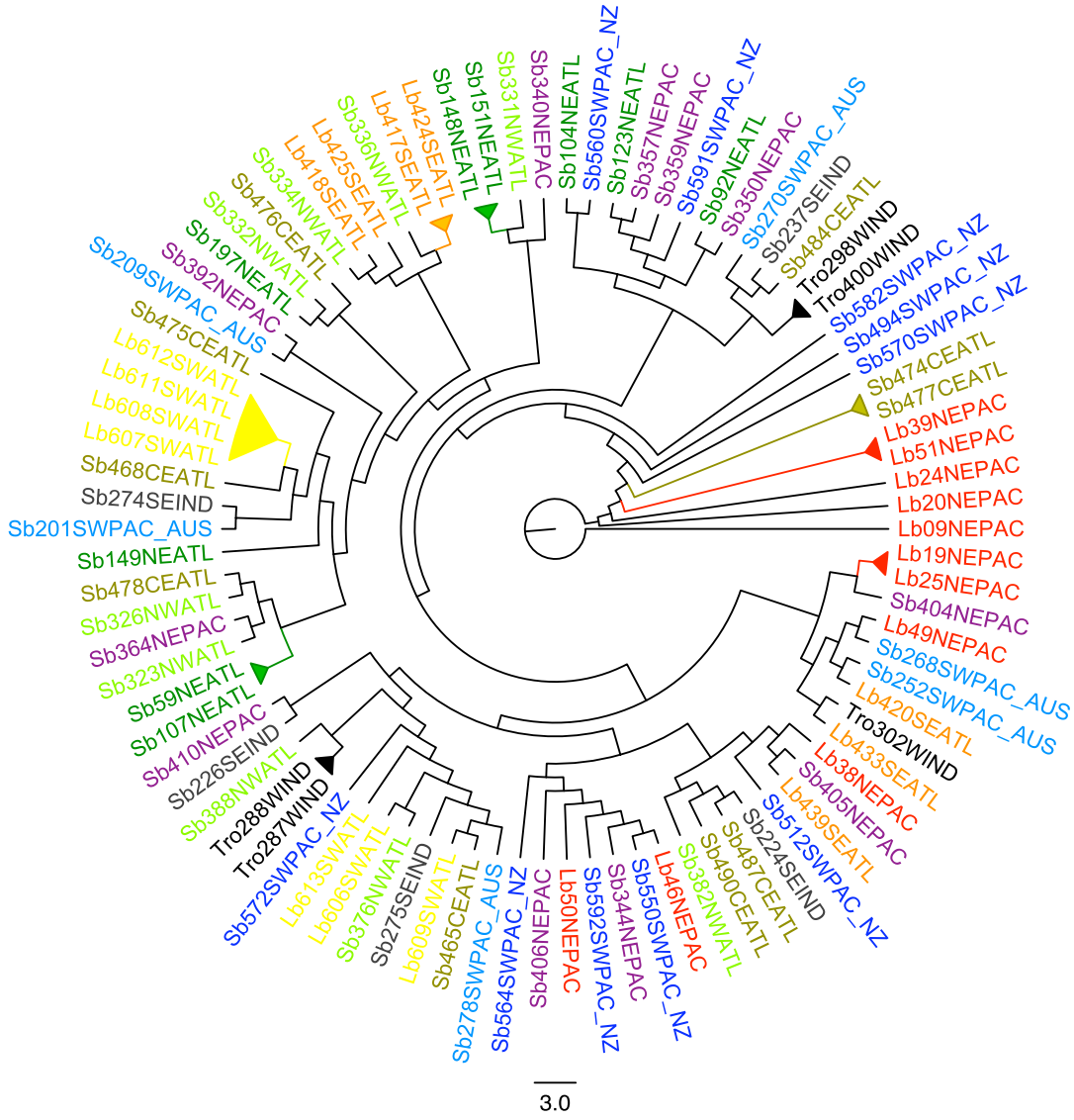


Figure S5.2. Mismatch distributions obtained for (a) the cytochrome *b* gene, (b) CHRNA1, (c) PLP, (d) Del_12, (e), Del_15, (f) Del_17.

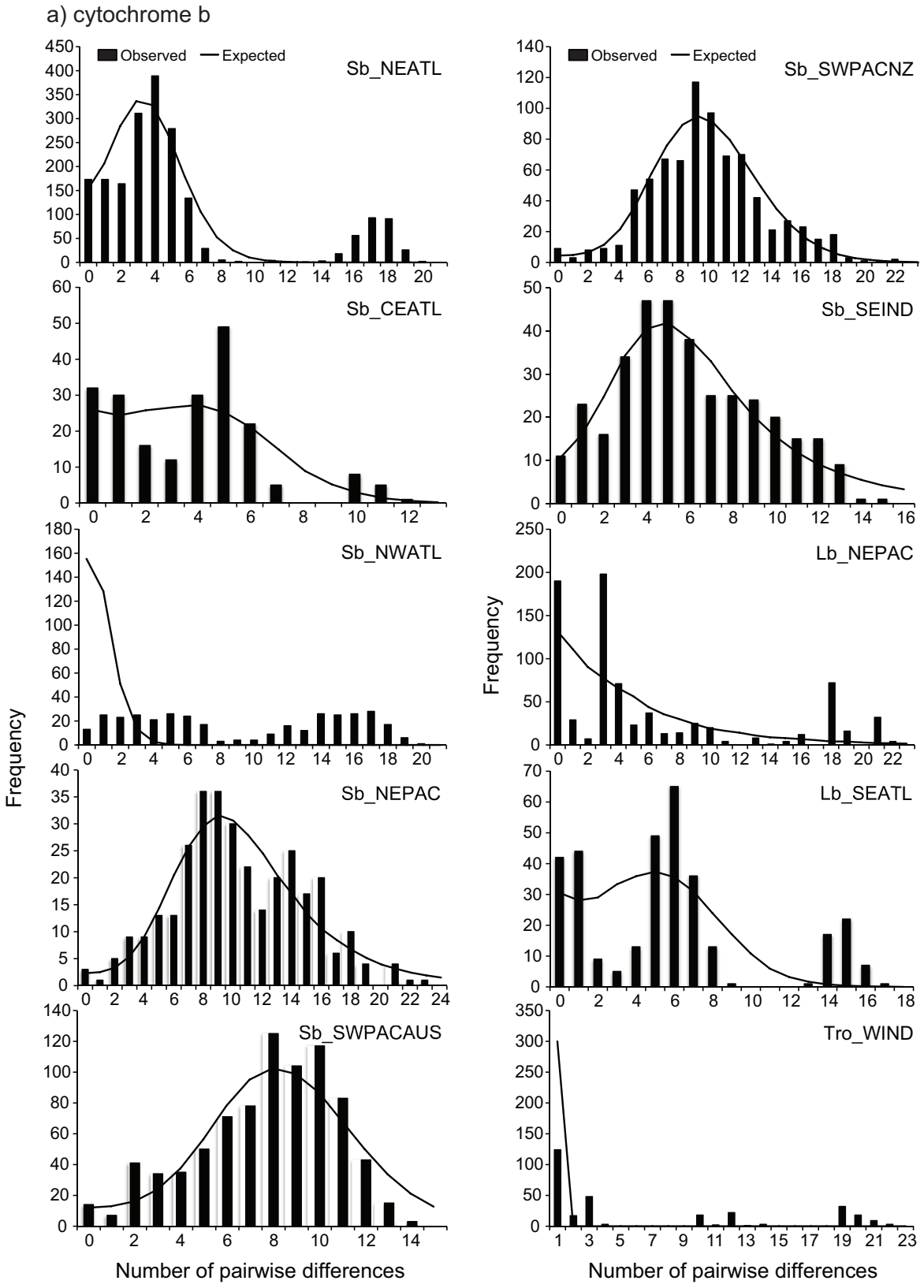


Figure S5.2. cont. Mismatch distributions obtained for (a) the cytochrome *b* gene, (b) CHRNA1, (c) PLP, (d) Del_12, (e), Del_15, (f) Del_17.

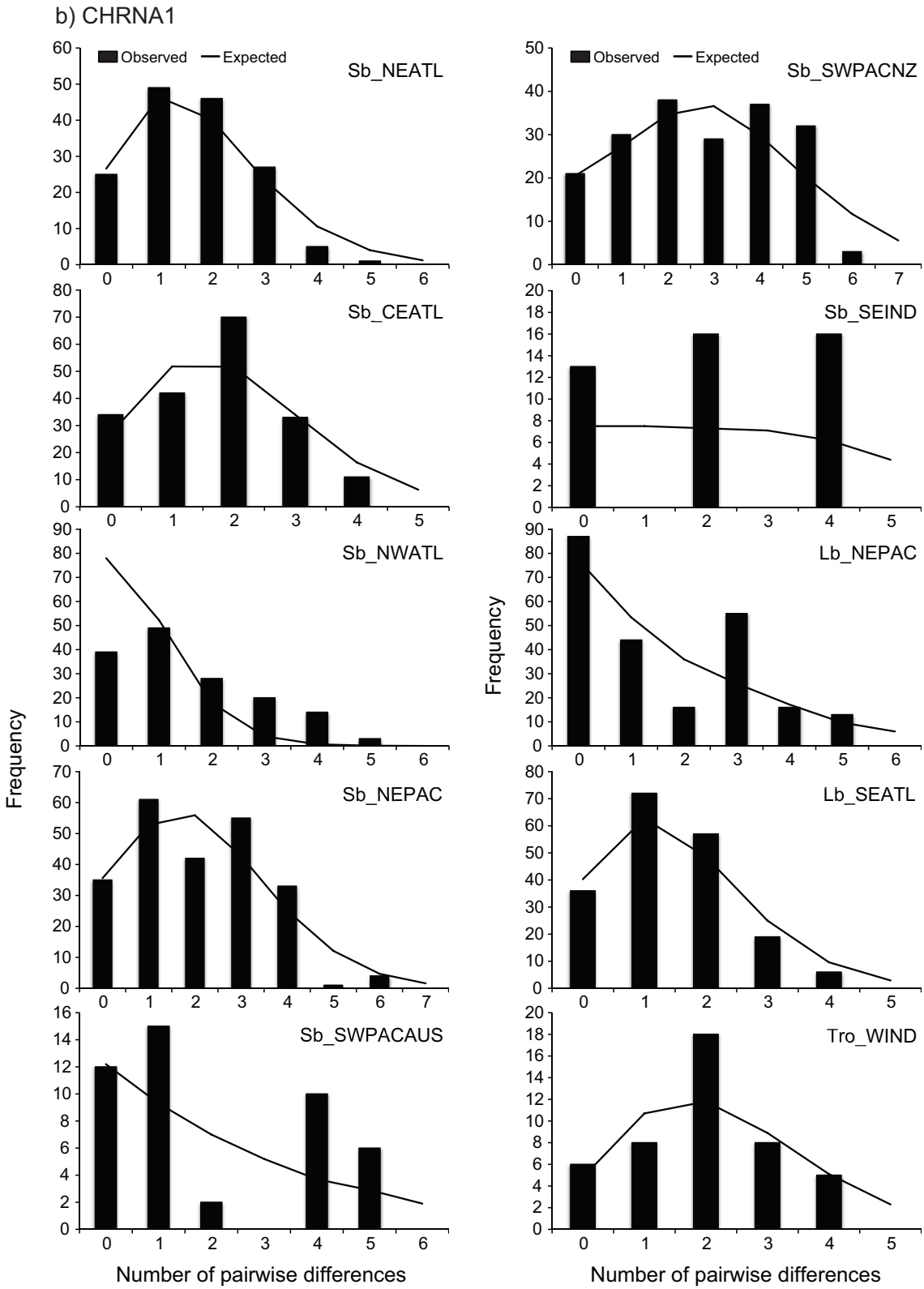


Figure S5.2. cont. Mismatch distributions obtained for (a) the cytochrome *b* gene, (b) CHRNA1, (c) PLP, (d) Del_12, (e), Del_15, (f) Del_17.

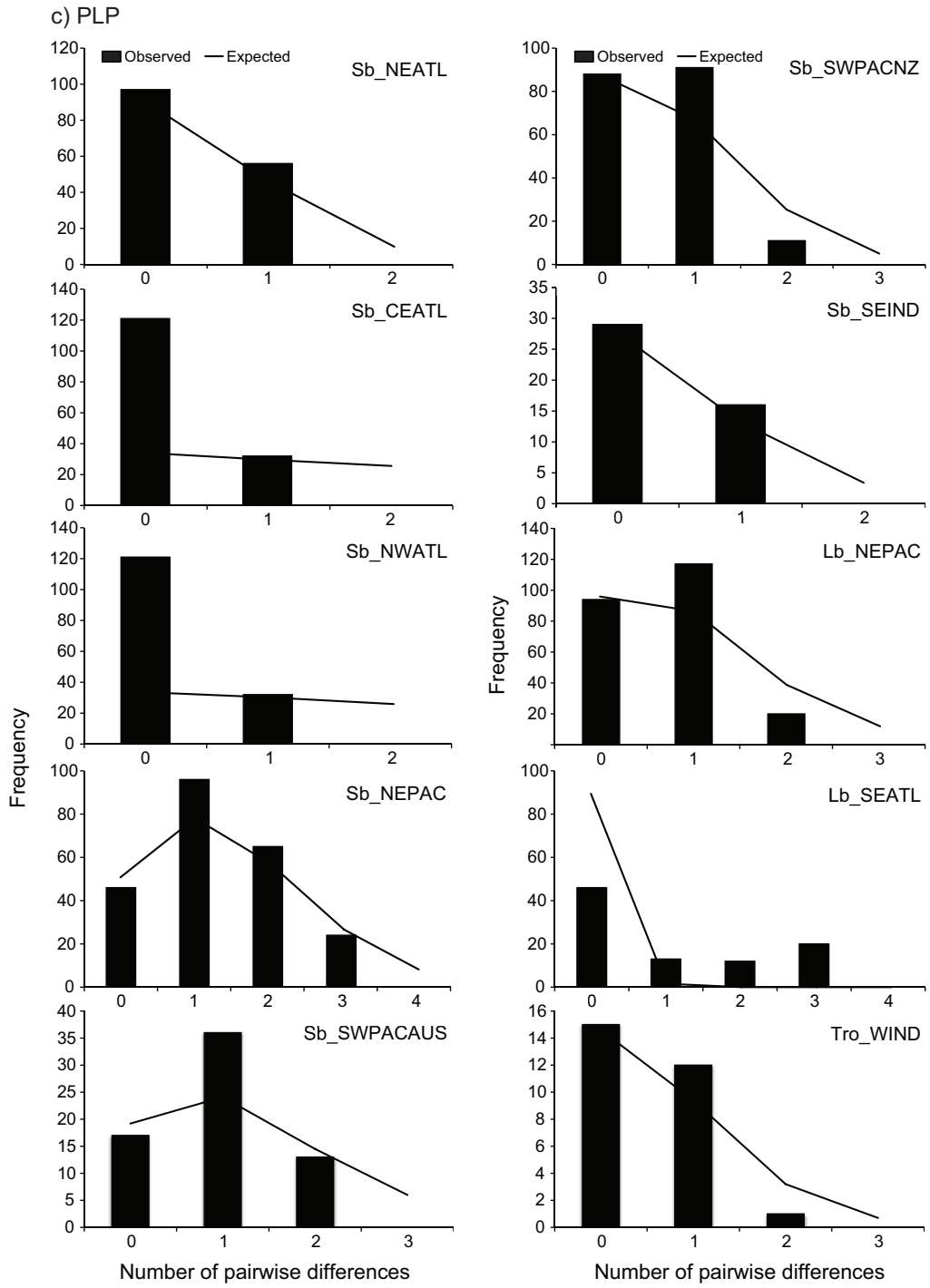


Figure S5.2. cont. Mismatch distributions obtained for (a) the cytochrome *b* gene, (b) CHRNA1, (c) PLP, (d) Del_12, (e), Del_15, (f) Del_17.

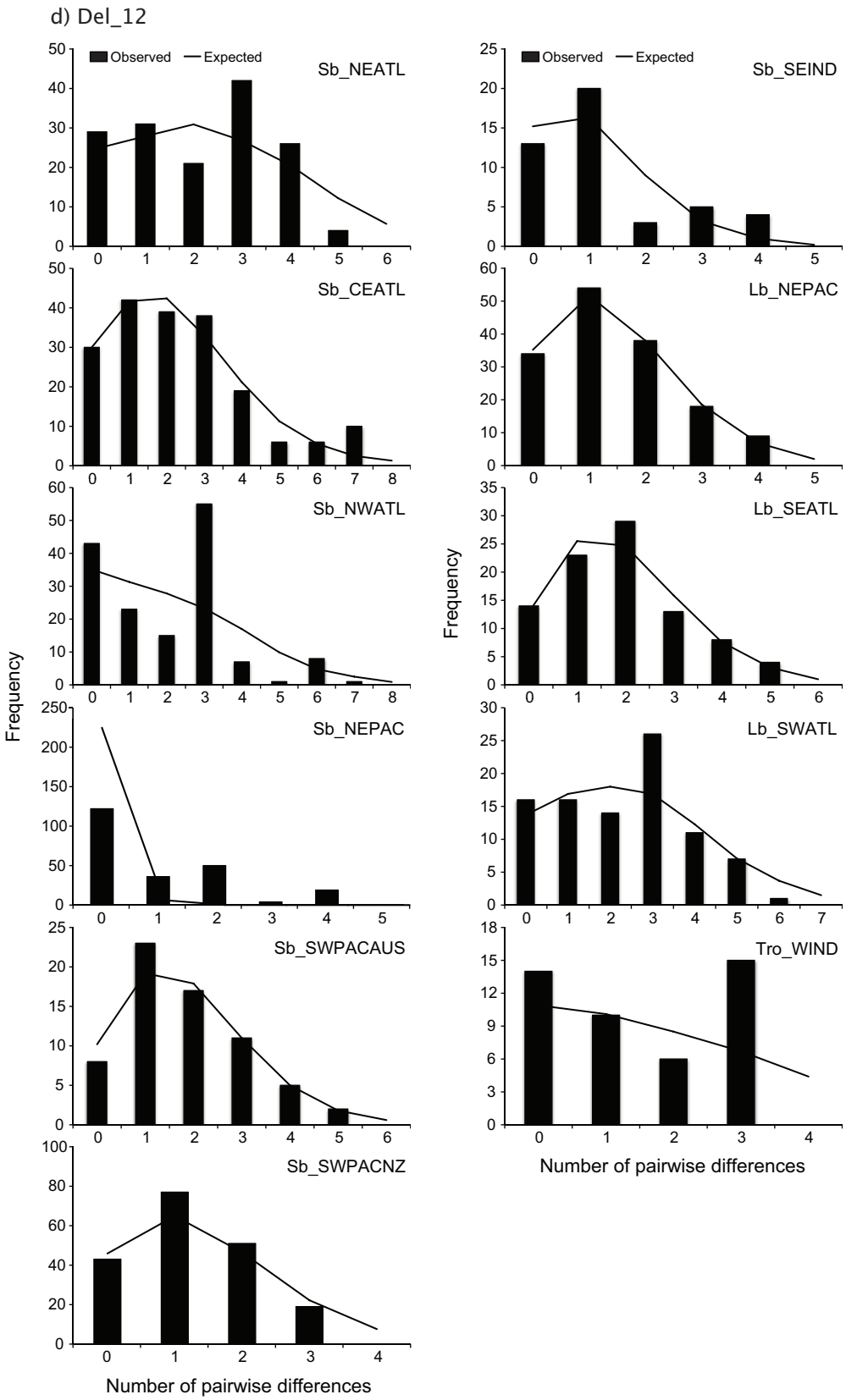


Figure S5.2. Mismatch distributions obtained for (a) the cytochrome *b* gene, (b) CHRNA1, (c) PLP, (d) Del_12, (e), Del_15, (f) Del_17.

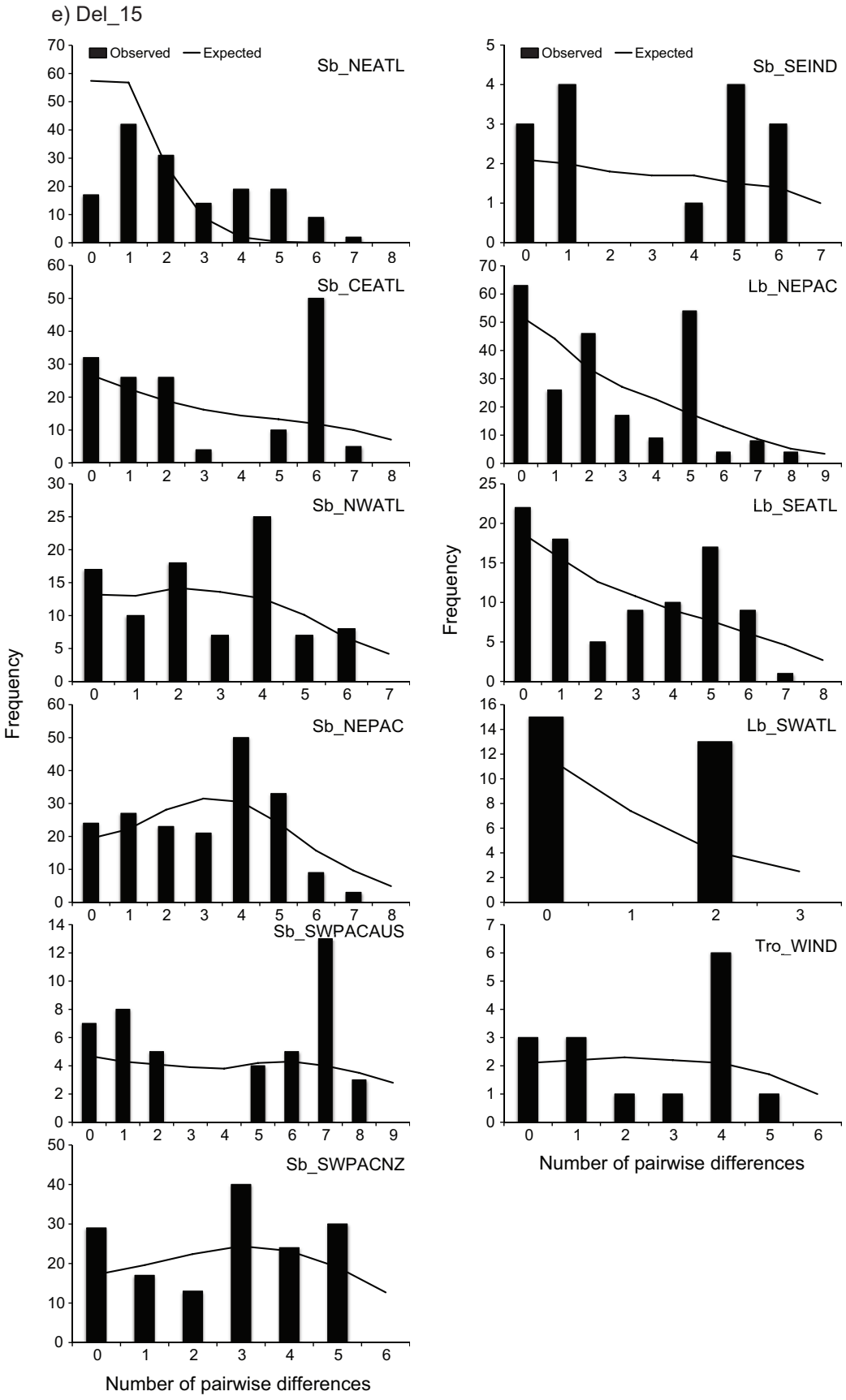


Figure S5.2. Mismatch distributions obtained for (a) the cytochrome *b* gene, (b) CHRNA1, (c) PLP, (d) Del_12, (e), Del_15, (f) Del_17.

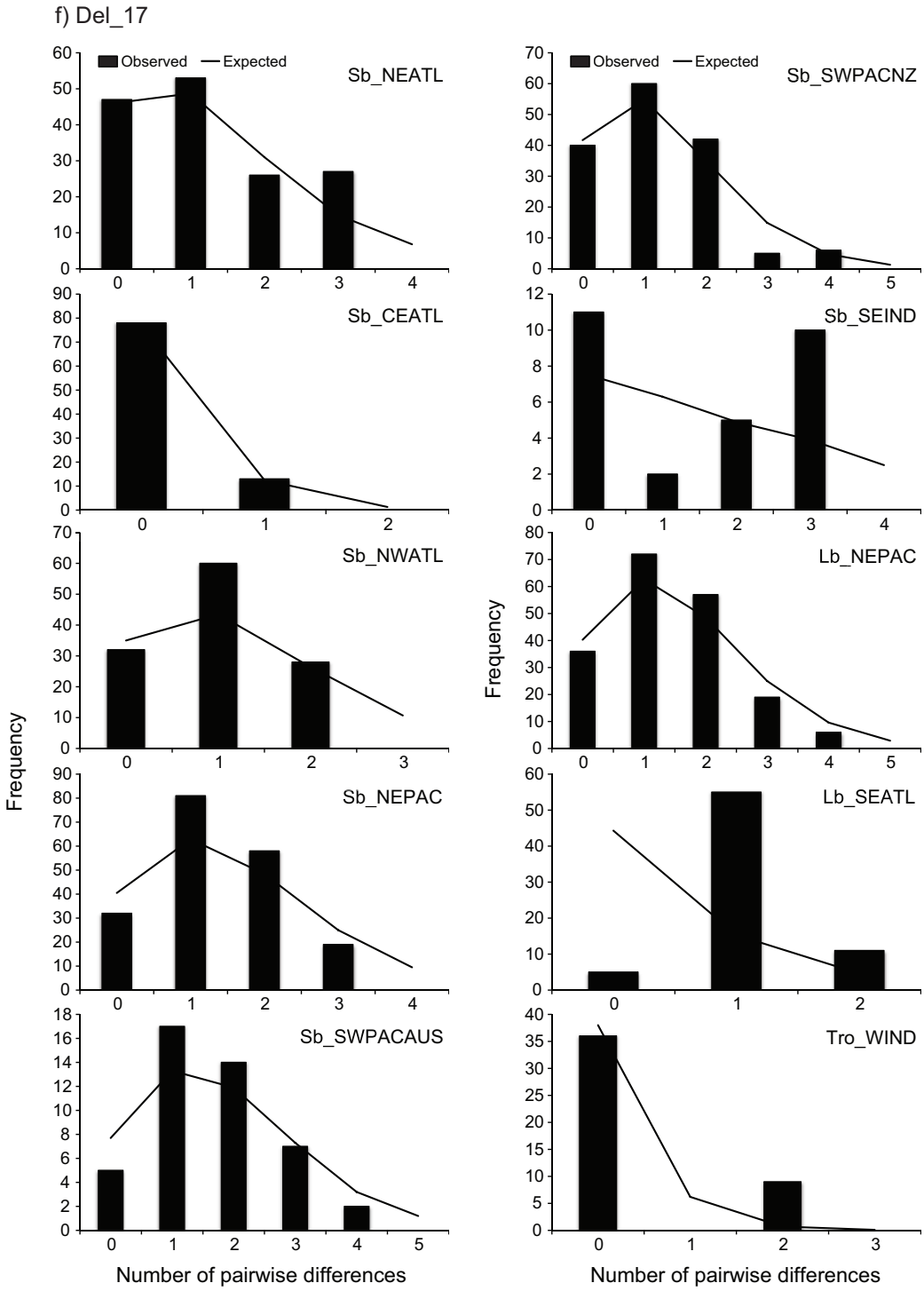


Table S5.1. Pairwise ϕ_{ST} values obtained for locus CHRNA1 (above diagonal) and locus PLP (above diagonal) for the different putative populations analysed in this study.

	Sb_NEATL	Sb_CEATL	Sb_NWATL	Sb_NEPAC	Sb_SWPACAUS	Sb_SWPACNZ	Sb_SEIND	Lb_NEPAC	Lb_SEATL	Lb_SWATL	Tro_WIND
Sb_NEATL		0.13725	-0.01279	0.08004	0.06445	0.1578*	0.14447	0.16536*	0.18788*	0.13253	0.50005*
Sb_CEATL	0.09294		0.05882	0.03894	0.05492	0.00605	-0.04972	0.02647	0.17539	0.02703	0.72293***
Sb_NWATL	-0.01471	0.19724*		0.07244	0.05492	0.12336	0.11007	0.13522*	0.17539	0.02703	0.66076***
Sb_NEPAC	-0.02842	0.12539	-0.01717		0.0431	0.02004	-0.01281	0.02721	0.13573*	0.06314	0.36125*
Sb_SWPACAUS	0.04341	0.24494*	0.04686	0.01283		0.04063	0.02039	0.07147	0.12151	0.06061	0.44843*
Sb_SWPACNZ	0.02767	0.03851	0.06403	0.02725	0.05173		-0.06442	-0.02348	0.17584*	0.10115	0.58395**
Sb_SEIND	0.17995	0.1777	0.27559*	0.1797	0.10628	0.0669		-0.04812	0.13548	0.13669	0.64896***
Lb_NEPAC	0.283*	0.22846**	0.3855***	0.28019**	0.27986*	0.18376*	0.05933		0.14944*	0.1153	0.56122***
Lb_SEATL	0.16093*	0.42556***	0.07778	0.08166	0.09857	0.19596*	0.43003*	0.52355***		0.15174	0.49302***
Lb_SWATL	0.44694*	0.18584**	0.57345*	0.42858	0.58752	0.3054	0.3472*	0.32059**	0.79859		0.79649**
Tro_WIND	0.19274**	0.3475*	0.19438***	0.14023***	0.09538***	0.13718**	0.24744***	0.43297**	0.16403***	0.64584***	

$^{*}P<0.05$; $^{**}P<0.01$, $^{***}P<0.001$

NEATL – Northeast Atlantic; CEATL – Central Eastern Atlantic; NWATL – Northwest Atlantic; SWATL – Southwest Atlantic; SWPACAUS – Southwest Pacific Australia; SWPACNZ – Southwest Pacific New Zealand; NEPAC – Northeast Pacific; SEIND – Southeast Indian Ocean; WIND – Western Indian Ocean; Sb – short-beaked common dolphin; Lb – long-beaked common dolphin; Tro – *tropicalis* form.

Table S5.2. Pairwise ϕ_{ST} values obtained for locus Del_12 (above diagonal) and locus Del_15 (above diagonal) for the different putative populations analysed in this study.

	Sb_NEATL	Sb_CEATL	Sb_NWATL	Sb_NEPAC	Sb_SWPACAUS	Sb_SWPACNZ	Sb_SEIND	Lb_NEPAC	Lb_SEATL	Lb_SWATL	Tro_WIND
Sb_NEATL		-0.02084	-0.0435	0.09783	0.12425	0.21135*	-0.00717	0.43792***	0.31449*	0.3507**	0.36365*
Sb_CEATL	0.02187		-0.0316	0.04793	0.01993	0.13173	-0.08956	0.34137**	0.20612	0.30797*	0.23493
Sb_NWATL	0.02317	0.01147		0.04259	0.05964	0.14911	-0.05748	0.3902***	0.25503*	0.33792*	0.30201
Sb_NEPAC	0.00822	0.1393*	0.16026*		-0.00368	0.01034	0.01034	0.00643*	0.17976	0.47583***	0.01674
Sb_SWPACAUS	-0.00683	-0.01583	0.04815	0.07852		0.01934	-0.05771	0.13685	0.01507	0.56908	0.04752
Sb_SWPACNZ	0.00799	0.01388	0.08309	0.09146	-0.05395		0.09827	0.14015	0.01675	0.56908***	0.04752
Sb_SEIND	0.32115*	0.11431	0.22992*	0.51636***	0.19899	0.30325*		0.32543*	0.17051	0.36607**	0.20952
Lb_NEPAC	0.09412	0.18595*	0.23286*	0.10065	0.130064	0.1409	0.50562***		-0.00093	0.68306***	-0.04753
Lb_SEATL	-0.02081	0.06969	0.05315	0.01781	0.02019	0.03985	0.37649**	0.01813		0.63238***	-0.08232
Lb_SWATL	0.3474***	0.20557*	0.21355*	0.48165***	0.30465*	0.37297**	0.32336*	0.47346***	0.34697**		0.71839**
Tro_WIND	-0.05641	0.03223	0.05647	-0.02707	-0.00012	0.02498	0.4186*	0.13829	0.01112	0.3783*	

*P<0.05; **P<0.01, ***P<0.001

NEATL – Northeast Atlantic; CEATL – Central Eastern Atlantic; NWATL – Northwest Atlantic; SWATL – Southwest Atlantic; SWPACAUS – Southwest Pacific Australia; SWPACNZ – Southwest Pacific New Zealand; NEPAC – Northeast Pacific; SEIND – Southeast Indian Ocean; WIND – Western Indian Ocean; Sb – short-beaked common dolphin; Lb – long-beaked common dolphin; Tro – *tropicalis* form.

Table S5.3. Pairwise ϕ_{ST} values obtained for locus Del_17 for the different putative populations analysed in this study.

	Sb_NEATL	Sb_CEATL	Sb_NWATL	Sb_NEPAC	Sb_SWPACAUS	Sb_SWPACNZ	Sb_SEIND	Lb_NEPAC	Lb_SEATL	Lb_SWATL	Tro_WIND
Sb_NEATL											
Sb_CEATL	0.21103										
Sb_NWATL	0.01703	0.33823*									
Sb_NEPAC	0.14229*	0.37834**	0.11055								
Sb_SWPACAUS	0.09698	0.39783**	0.08105	-0.00187							
Sb_SWPACNZ	0.09349	0.32833*	0.07201	-0.02711	-0.01581						
Sb_SEIND	0.21081*	0.43956***	0.27928*	0.23021*	0.0871	0.19378					
Lb_NEPAC	0.28633***	0.39238***	0.30676***	0.09501	0.14499	0.10221	0.25937*				
Lb_SEATL	0.18518	0.82042***	0.10092	0.15098	0.14471	0.16432	0.49631*	0.4329***			
Lb_SWATL	0.00588	0.37179	-0.00747	0.13733	0.12305	0.07677	0.33091*	0.3194**	0.33437		
Tro_WIND	0.14807	-0.04396	0.25313*	0.31092*	0.29985*	0.25611*	0.33282*	0.3334*	0.7059***	0.24768	

*P<0.05; **P<0.01, ***P<0.001

NEATL – Northeast Atlantic; CEATL – Central Eastern Atlantic; NWATL – Northwest Atlantic; SWATL – Southwest Atlantic; SWPACAUS – Southwest Pacific Australia; SWPACNZ – Southwest Pacific New Zealand; NEPAC – Northeast Pacific; SEIND – Southeast Indian Ocean; WIND – Western Indian Ocean; Sb – short-beaked common dolphin; Lb – long-beaked common dolphin; Tro – *tropicalis* form.

Table S5.4. Estimation of time since beginning of expansion using the Mismatch distributions (MMD) method.

	SSD	*Hri*	θ_0	θ_1	t	t (9.86×10^{-9})
mtDNA						
Sb_NEATL	0.014	0.021	0.000	13.760	4.301	0.195
Sb_CEATL	0.028	0.046	0.000	7.068	5.350	0.242
Sb_NWATL	0.311***	0.008	0.000	99999.000	0.777	0.035
Sb_NEPAC	0.005	0.009	3.683	143.203	7.666	0.347
Sb_SWPACAUS	0.005	0.013	0.000	67.969	8.881	0.402
Sb_SWPACNZ	0.003	0.011	1.113	175.859	8.777	0.397
Sb_SEIND	0.003	0.009	2.983	36.426	3.871	0.175
Lb_NEPAC	0.065	0.153	0.002	4.903	19.832	0.897
Lb_SEATL	0.041	0.047	0.000	9.478	6.682	0.302
Tro_WIND	0.398***	0.191	0.000	428.125	0.000	0.000
	SSD	*Hri*	θ_0	θ_1	t	t (4.79×10^{-10})
CHRNA1						
Sb_NEATL	0.062	0.004	0.000	99999.000	1.742	4.798
Sb_CEATL	0.015	0.078	0.007	102.188	1.967	5.418
Sb_NWATL	0.095*	0.033	0.000	99999.000	0.654	1.801
Sb_NEPAC	0.011	0.051	0.009	9.647	2.457	6.767
Sb_SWPACAUS	0.065	0.165	2.634	2.634	0.000	0.000
Sb_SWPACNZ	0.009	0.032	0.011	10.605	3.564	9.816
Sb_SEIND	0.152	0.589*	0.000	5.120	4.041	11.130
Lb_NEPAC	0.028	0.110	0.000	1.969	4.326	11.915
Lb_SEATL	0.017	0.150	0.000	99999.000	0.926	2.550
Tro_WIND	0.023	0.118	0.000	99999.000	2.236	6.158
Del_12						
Sb_NEATL	0.019	0.056	0.000	5.982	3.160	4.118
Sb_CEATL	0.003	0.022	0.858	9.316	1.785	2.326
Sb_NWATL	0.064	0.192	0.000	3.488	3.346	4.360
Sb_NEPAC	0.301***	0.193	0	99999.000	0	0.000
Sb_SWPACAUS	0.005	0.079	0.004	99999.000	1.814	2.364
Sb_SWPACNZ	0.005	0.089	0.000	99999.000	1.402	1.827
Sb_SEIND	0.032	0.177	0.146	99999.000	0.953	1.242
Lb_NEPAC	0.000	0.052	0.000	99999.000	3.160	4.118
Lb_SEATL	0.004	0.052	0.000	353.125	1.949	2.540
Tro_WIND	0.042	0.167	0.000	3.142	3.104	4.045

Table S5.4. cont. Estimation of time since beginning of expansion using the Mismatch distributions (MMD) method.

	SSD	*Hri*	θ_0	θ_1	t	t (4.79×10^{-10})
Del_15						
Sb_NEATL	0.124*	0.052	0.000	99999.000	0.922	1.290
Sb_CEATL	0.084	0.183	0.002	4.808	7.117	9.958
Sb_NWATL	0.028	0.106	0.000	6.001	4.363	6.105
Sb_NEPAC	0.021	0.049	0.000	9.551	4.270	5.975
Sb_SWPACAUS	0.066	0.111	0.002	8.555	7.893	11.044
Sb_SWPACNZ	0.026	0.089	0.002	8.203	4.340	6.073
Sb_SEIND	0.089	0.164	0.000	5.903	6.498	9.092
Lb_NEPAC	0.041	0.136	0.004	3.346	5.574	7.799
Lb_SEATL	0.022	0.046	0.002	3.879	5.721	8.005
Tro_WIND	0.091	0.244	0.000	5.773	4.553	6.371
Del_17						
Sb_NEATL	0.008	0.064	0.758	78.343	0.666	0.948
Sb_CEATL	0.000	0.531	0.000	0.424	0.207	0.295
Sb_NWATL	0.021	0.180	0.005	99999.000	1.223	1.742
Sb_NEPAC	0.015	0.133	0.000	99999.000	1.529	2.177
Sb_SWPACAUS	0.012	0.114	0.002	99999.000	1.760	2.506
Sb_SWPACNZ	0.008	0.091	0.000	99999.000	1.277	1.819
Sb_SEIND	0.085	0.274	0.002	2.748	3.502	4.987
Lb_NEPAC	0.006	0.088	0.000	99999.000	1.520	2.165
Lb_SEATL	0.023	0.472	0.450	0.450	2.965	4.222
Tro_WIND	0.051	0.720	0.000	0.180	3.000	4.272
PLP						
Sb_NEATL	0.005	0.206	0.000	99999.000	0.502	0.676
Sb_CEATL	0.321	0.382	0.900	3.600	2.930	3.946
Sb_NWATL	0.321	0.382	0.900	3.600	2.930	3.946
Sb_NEPAC	0.008	0.107	0.000	99999.000	1.465	1.973
Sb_SWPACAUS	0.036	0.243	0.000	99999.000	1.195	1.610
Sb_SWPACNZ	0.021	0.181	0.000	99999.000	0.752	1.013
Sb_SEIND	0.004	0.210	0.000	99999.000	0.486	0.655
Lb_NEPAC	0.025	0.194	0.000	99999.000	0.863	1.162
Lb_SEATL	0.331***	0.188	0.000	99999.000	1.529	2.059
Tro_WIND	0.014	0.167	0.000	99999.000	0.627	0.845

Chapter VI (Article 5)
Seascape Genetics of a Globally Distributed, Highly Mobile Marine Mammal: The Short-Beaked Common Dolphin (Genus *Delphinus*)

Ana R. Amaral[1,2,*], Luciano B. Beheregaray[2,3], Kerstin Bilgmann[4], Dmitri Boutov[5], Luís Freitas[6], Kelly M. Robertson[7], Marina Sequeira[8], Karen A. Stockin[9], M. Manuela Coelho[1] and Luciana M. Möller[3]

[1] Universidade de Lisboa, Faculdade de Ciências, Centro de Biologia Ambiental, Campo Grande, 1749-016 Lisbon, Portugal
[2] Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
[3] School of Biological Sciences, Flinders University, Adelaide SA 5001, Australia
[4] Graduate School of the Environment, Macquarie University, Sydney, NSW 2109, Australia
[5] Universidade de Lisboa, Faculdade de Ciências, Centro de Oceanografia, Campo Grande, 1749-016 Lisbon, Portugal
[6] Museu da Baleia da Madeira, 9200-031, Caniçal, Madeira, Portugal
[7] National Marine Fisheries Service, Southwest Fisheries Science Center, 3333 N. Torrey Pines Ct., La Jolla, CA 92037, USA
[8] Instituto de Conservação da Natureza e Biodiversidade, Rua de Santa Marta, 55, 1150-294 Lisbon, Portugal
[9] Coastal-Marine Research Group, Institute of Natural Sciences, Massey University, Private Bag 102 904, North Shore MSC, Auckland, New Zealand

Plos One (submitted)

Seascape genetics of a globally distributed, highly mobile marine mammal: The short-beaked common dolphin (genus *Delphinus*)

6.1. ABSTRACT

Identifying which factors shape the distribution of intraspecific genetic diversity is central in evolutionary and conservation biology. In the marine realm, the absence of obvious barriers to dispersal can make this task more difficult. Nevertheless, recent studies have provided valuable insights into which factors may be shaping genetic structure in the world's oceans. These studies were, however, generally conducted on marine organisms with larval dispersal. Here, using a seascape genetics approach, we show that marine productivity and sea surface temperature are correlated with genetic structure in a highly mobile, widely distributed marine mammal species, the short-beaked common dolphin. Isolation by distance also appears to influence population divergence over larger geographical scales (i.e. across different ocean basins). We suggest that the relationship between environmental variables and population structure may be caused by prey behaviour, which is believed to determine common dolphins' movement patterns and preferred associations with certain oceanographic conditions. Our study highlights the role of oceanography in shaping genetic structure of a highly mobile and widely distributed top marine predator. Thus, seascape genetic studies can potentially track the biological effects of ongoing climate-change at oceanographic interfaces and also inform marine reserve design in relation to the distribution and genetic connectivity of charismatic and ecologically important megafauna.

Key Words: *Delphinus delphis*; marine connectivity; habitat heterogeneity; feeding specialization; conservation genetics.

6.2. INTRODUCTION

Identifying environmental conditions underlying the division of natural populations into smaller units is central for understanding ecological and evolutionary processes and for the conservation management of biodiversity. In highly mobile

species that are distributed across continuous environments with few barriers to dispersal, it is expected that persistent gene flow will stifle genetic differentiation and speciation. Nevertheless, there is growing recognition that gene flow can be limited even in the absence of geographical barriers, both in terrestrial and aquatic environments (Brown *et al.* 2007; Hellberg 2009). A detailed knowledge of how landscape characteristics structure populations has therefore become an important focus of molecular ecological research (Manel *et al.* 2003), leading to the emerging field of landscape genetics (Manel *et al.* 2003; Storfer *et al.* 2010). This multidisciplinary approach aims to complement genetic data with lines of evidence from other areas such as spatial statistics and landscape ecology in order to understand the effects of the landscape on the spatial distribution of genetic diversity (Holderegger & Wagner 2008; Manel *et al.* 2003; Storfer *et al.* 2007). Although extensively applied in terrestrial systems, this approach has been seldom used in the marine environment (Storfer *et al*. 2010; but see Galindo *et al.* 2006 and Selkoe *et al.* 2008).

The study of connectivity in marine systems can be challenging due to the absence of obvious barriers to dispersal and generally large population sizes of marine organisms that often resist genetic divergence, leading to low statistical power to detect population structure (Selkoe *et al.* 2008; Selkoe *et al.* 2010). Therefore, the use of an integrative approach such as the one used in landscape genetics (or 'seascape genetics' when applied to the marine environment) has provided valuable insights into which factors may be shaping genetic structure in the world's oceans (Banks *et al.* 2007; Galindo *et al.* 2006). Biogeographic barriers and environmental variables such as ocean currents, upwelling, variation in sea surface temperature and salinity are some of the factors that have been proposed to explain genetic diversity and structure in marine organisms (Banks *et al*. 2007; Banks *et al.* 2010; Selkoe *et al.* 2010). However, most of these studies have been conducted in organisms with larval dispersal. In active marine dispersers such as sharks and dolphins, where dispersal potential is dependent upon adult vagility, the interplay of environmental features and genetic structure has remained largely untested (but see Mendez *et al.* 2010). Although differences in salinity, temperature and productivity levels have been suggested to explain

genetic discontinuities in dolphins (Bilgmann *et al.* 2007; Fullard *et al.* 2000; Möller *et al.* 2011; Natoli *et al.* 2005), a direct relationship between such oceanographic features and genetic structure has only been recently evaluated for two coastal dolphin species with limited distribution: the franciscana (*Pontoporia blainvillei*) (Mendez *et al.* 2010) and the humpback dolphin (Mendez *et al.* 2011). These authors found that heterogeneity in chlorophyll concentration, water turbidity and temperature likely influenced the occurrence of genetically distinct populations of these species along the coast of Argentina and in the Western Indian Ocean, respectively.

In this study we use as model a highly mobile, widely distributed cetacean species belonging to the genus *Delphinus*, the short-beaked common dolphin. Common dolphins occur in all oceans from tropical to temperate waters. Two species and four subspecies are currently recognized: the short-beaked common dolphin, *Delphinus delphis* Linnaeus, 1758, distributed in continental shelf and pelagic waters of the Atlantic and Pacific Oceans; the long-beaked common dolphin, *Delphinus capensis* Gray, 1828, distributed in nearshore tropical and temperate waters of the Pacific and southern Atlantic waters; *D. d. ponticus* Barabash, 1935, restricted to the Black sea; and *D. c. tropicalis* van Bree, 1971, restricted to the Indian Ocean (Perrin 2009). However, due to discordance between morphological and genetic characters, the phylogenetic relationships and taxonomy within the genus, particularly in regard to the specific status of the long-beaked form, are still under debate (Natoli *et al*. 2006; Amaral *et al*. 2009; Amaral *et al.* in review).

Short-beaked common dolphins are known to occur in large groups of dozens to hundreds of individuals. Although their social structure is still poorly understood, individuals seem to group irrespective of genetic relationships, with possible gender and age segregation (Viricel *et al.* 2008). However, there is a gap in knowledge if these findings are representative for common dolphins in other geographic regions. The movements of common dolphins are thought to be largely determined by those of their potential prey (e.g. Young & Cockcroft 1994) and their diet vary between locations and seasons. Nonetheless, they generally depend on small, mesopelagic shoaling fishes such as scombroids and clupeoids,

and squids (Young & Cockcroft 1994, Pusineri *et al.* 2007). It has been suggested that short-beaked common dolphins often prefer specific water masses (Ballance *et al.* 2006; Doksaeter *et al.* 2008; Möller *et al.* 2011) and in the Eastern Tropical Pacific they occur preferentially in upwelling-modified waters (Balance *et al*. 2006).

Genetic studies conducted so far have shown significant genetic differentiation among populations inhabiting different oceans and different coasts of the Atlantic Ocean (Natoli *et al.* 2006; Mirimin *et al*. 2009; Amaral *et al.* in review). However, within each side of the Atlantic Ocean, no genetic structure has been detected, suggesting a lack of strong dispersal barriers in these areas (Amaral *et al.* 2007; Mirimin *et al.* 2009). Within the Pacific Ocean, results from regional studies have reported fine-scale population genetic structure in short-beaked common dolphins occurring off the USA coast (Chivers *et al*. 2009), off the Eastern (Möller *et al.* 2011) Australian Coast and around New Zealand (Stockin *et al.* unpublished data). Particular oceanographic characteristics, such as ocean currents and temperature and salinity differences have been pointed out as likely factors limiting movement of short-beaked common dolphins (Chivers *et al.* 2009; Bilgmman *et al.* 2008; Möller *et al*. 2011). However, a direct evaluation of the influence of oceanographic variables on the genetic structure of this species has never been carried out.

Our aim is to assess the relative influence of key oceanographic variables on population subdivision of short-beaked common dolphins at a range of medium to large spatial scales, including across oceans and within ocean basins. To achieve this aim we have sampled populations inhabiting the Atlantic, Pacific and Indian Oceans and used remote sensing data under a seascape genetics approach. The global distribution, high mobility, and putatively close association of short-beaked common dolphins with water masses, makes them an excellent model species to test for interactions between variation in environmental factors and genetic structure, contributing towards an understanding of ecological processes affecting population connectivity in the sea.

6.3. MATERIALS AND METHODS

6.3.1. *Sampling*

We used samples from seven oceanic regions (Figure 6.1): the Northeast Atlantic (NEATL), *n* = 75; the Central Eastern Atlantic (CEATL), *n* = 29; the Northwest Atlantic (NWATL), *n* =38; the Northeast Pacific (NEPAC), *n* = 40; the Southwest Pacific, *n* = 35 (encompassing eastern Australian waters, SWPAC_AUS) and *n* = 39 (encompassing New Zealand waters, SWPAC_NZ) and the Southeast Indian Ocean (southern Australian waters, SEIND), *n* = 27 (Table 6.1). All tissue samples were obtained from either stranded animals or from skin biopsies collected from free-ranging dolphins. Tissues were stored either in ethanol or in 20% DMSO/saturated NaCl.

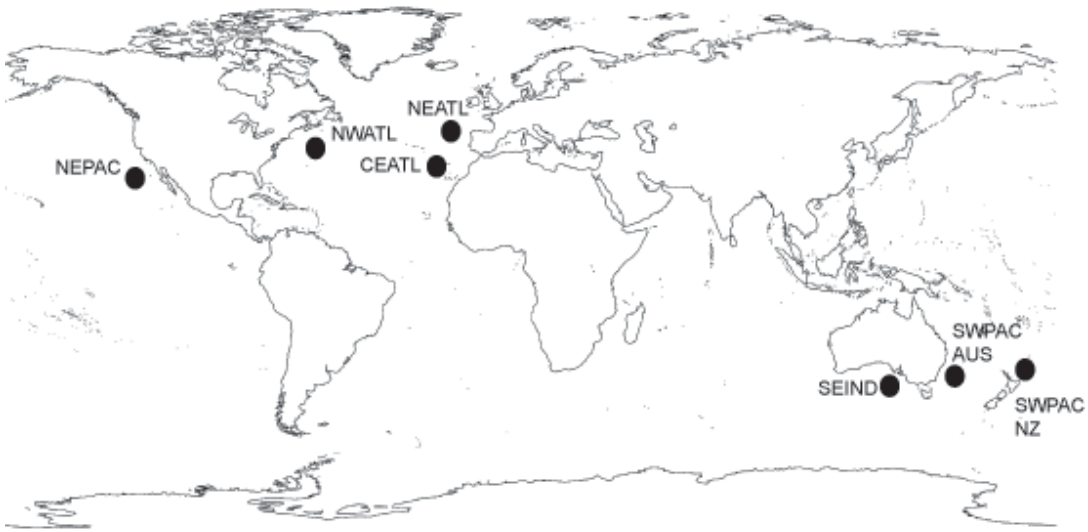


Figure 6.1. Map showing sampling locations for the short-beaked common dolphin populations analysed in this study. (NEPAC – Northeast Pacific; NWATL – Northwest Atlantic; CEATL – Central Eastern Atlantic; SEIND – Southeast Indian Ocean; SWPAC_AUS – Southwest Pacific Australia; SWPAC_NZ – Southwest Pacific New Zealand).

6.3.2. *DNA extraction and microsatellite genotyping*

Genomic DNA was isolated from skin or muscle using a standard proteinase K digestion and two phenol-chloroform and one chlorofom-isoamyl extractions followed by ethanol precipitation (Rosel & Block 1996) or, alternatively, using a salting-out protocol (Sunnucks & Hales 1996). DNA quality and concentration was

verified using Thermo Scientifc NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific Inc.). Samples from NEPAC and NWATL were provided as DNA by the Southwest Fisheries Science Center, Marine Mammal and Turtle Research Sample Collection (SWFSC-NOAA, La Jolla, CA).

All samples were genotyped at 14 polymorphic microsatellite loci: 7 tetranucleotide (Tur4_80, Tur4_87, Tur4_92, Tur4_105, Tur4_141, Tur4_142; (Nater *et al.* 2009) and Dde59 (Coughlan *et al.* 2006) and 7 dinucleotide (Dde66, Dde70; Coughlan *et al*. 2006), KW2, KW12 (Hoelzel *et al.* 1998), EV1 (Valsecchi & Amos 1996), MK6 and MK8 (Krutzen *et al.* 2001). The forward primer for each primer pair was labelled with a M13 tag (Schuelke 2000). Fluorescent dyes were also labelled with the M13 tag. Amplification reactions contained 50-100 ng DNA, 1 x reaction buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 μM of each primer and 1 U *Taq* DNA polymerase. The thermal cycler profile for the tetranucleotide loci and Dde66 and Dde70 consisted of initial denaturation at 94ºC for 3 min followed by a touchdown profile for 5 cycles with the annealing temperature starting at 63ºC and decreasing 2ºC per cycle, followed by 30 cycles with an annealing temperature of 53ºC, and a final extension step at 72ºC for 10 min. For the remaining dinucleotide loci, conditions followed the original publications. All reactions included both positive and negative controls. Following amplification, samples were mixed with an internal size standard and run on an ABI 3130 Genetic Analyzer. The GeneMapper v.4.1 software (Applied Biosystems, CA) was used for sizing of allele fragments.

6.3.3. *Data analysis*

Genetic diversity

The program Micro-checker v.2.2.3 (Oosterhout *et al.* 2004) was used to check for the presence of genotyping errors such as scoring errors due to stuttering, large allele dropout or evidence for null alleles. Departures from Hardy-Weinberg Equilibrium were tested for each population using the Fisher exact test in Genepop v.4.0 (Rousset 2008). Genepop was also used to test for linkage disequilibrium between loci. Samples were grouped into 7 putative populations

according to their geographical origin as described above. Genetic diversity measures such as mean number of alleles per locus and observed (H_O) and expected (H_E) heterozygosities were calculated in Arlequin v.3.5.1 (Excoffier & Lischer 2010) and allelic richness (A_R) calculated using FSTAT v.2.9.3 (Goudet 1995).

Genetic differentiation

Three different measures of population differentiation were used: the fixation index F_{ST}, estimated using FSTAT (Goudet 1995); the analogous R_{ST}, estimated using Genepop v.4.0 (Rousset 2008); and the statistic *D* (Jost 2008), estimated using SMOGD v.1.2.5 (Crawford 2010). The latter has been shown to provide a more accurate measure of differentiation when using highly polymorphic microsatellite loci (Jost 2008). Additionally, we tested for a mutation effect on genetic structure by randomly reassigning allele sizes while keeping allele identity the same (Hardy *et al.* 2003). The test was conducted in SPAGEDI v.1.3 through 10,000 permutations. R_{ST} values significantly larger then F_{ST} values indicate that mutation, in addition to drift and gene flow, has contributed to frequency differences among samples, which in some cases can be interpreted as phylogeographic signal (Hardy *et al.* 2003).

In order to visualize relationships among putative populations based on genetic variation, we performed a principal component analysis (PCA) on a table of standardised allele frequencies using the adegenet and ade4 packages in R (Jombart *et al.* 2009). In addition, we performed an analysis of nonmetric multidimensional scaling (MDS, Kruskal & Wish, 1978) on each of the genetic distance matrices using the PRIMER computer package (Clarke & Warwick 2001).

An analysis of molecular variance, AMOVA (Excoffier *et al.* 1992) was conducted in Arlequin to assess population structure. Different hierarchical levels were tested, considering differences occurring between populations in different oceans and within the same ocean basin.

A Bayesian approach to identify the number of populations (*K*) present in the dataset was implemented in the program STRUCTURE v.2.3.3 (Falush *et al.*

2003; Pritchard *et al.* 2000). The admixture and the correlated allele frequencies models were implemented since we expect that allele frequencies in the different populations are likely to be similar due to migration or shared ancestry. Sampling locations were used as prior to help detect population structure (Hubisz *et al.* 2009). Ten independent runs of *K* between 1 and 10 were run with 50 000 "burn in" and 300,000 MCMC replicates. The maximum log-likelihood values from all runs corresponding to each given *K* were checked for consistency and averaged. The *K* with the highest averaged maximum log-likelihood was considered the most likely number of clusters that better explains our dataset. *CLUMMP* v.1.1.2 (Jakobsson & Rosenberg 2007) was used to summarize parameters across 10 runs and *distruct* v.1.1 (Rosenberg 2004) was used to produce the corresponding graphical output.

Isolation by distance

Isolation by distance (IBD) was evaluated using a Mantel test implemented in the program IBDWS v.3.16 (Jensen *et al.* 2005). Genetic distance matrices given by $F_{ST}/(1- F_{ST})$ were regressed against the logarithm of geographical distances following a two-dimensional model (Rousset 1997). R_{ST} and *D* values were also used. Geographic distances were measured in Google Earth by using set points and measuring either straight-line distance across oceans, or the shortest geographical distance along continental margins.

Environmental predictors of genetic structure

Three different oceanographic variables were used as predictors of the observed genetic differences between short-beaked common dolphin populations. These were night-time sea surface temperature (SST, ºC), chlorophyll concentration (CHL, mg/m^3) and water turbidity measured as diffuse attenuation coefficient at 490 nm (KD490, m^{-1}). These variables, here obtained from remote sensing data, have been previously related to habitat heterogeneity (Bost *et al.* 2009) and associated with genetic differences in other dolphin species (Mendez *et al.* 2010; Mendez *et al.* 2011). Furthermore, the oceanographic variables chosen have wide a geographic coverage through remote sensing, making them ideal for a global

approach. Seven oceanic regions, corresponding to the sampling areas for short-beaked common dolphins, were used for the extraction of these oceanographic variables to assess association with patterns of genetic differentiation. Polygons were defined considering the possible range of common dolphins within that oceanic region. For NWATL the area was defined between 46ºN, 38ºN and 57ºW; for CEATL between 34ºN, 32ºN and 16ºW; for NEATL between 60ºN, 35ºN and 0º; for NEPAC between 45ºN, 25ºN and 108ºW; for SWPAC_NZ between 32ºS, 44ºS and 180ºW; for SWPAC_AUS between 26ºS, 44ºS and 156ºE; and for SEIND between 31ºS, 37ºS and 140ºE. Monthly averaged data of the three variables, with a 4 km spatial resolution was obtained from Ocean Color Web (http://oceancolor.gsfc.nasa.gov/) for the period from July 2002 to October 2010 and processed using MATLAB software (www.mathworks.com). Data collected during this time period provide a characterization of the oceanographic features for each region and are robust to inter-annual oscillations (Supplementary Material, Figure S6.1). Data analysis included the construction of temperature, chlorophyll and turbidity maps for each region, where each pixel of the map corresponds to the eight-year average value for a 4 km grid. These maps were visually inspected to detect geographical areas of environmental heterogeneity. Monthly averages for each oceanic region were then statistically analysed using a paired t-test to detect differences among those regions. Total averages for the 8 year-period for each factor and each sampled region were subsequently used to examine environmental and genetic associations (details below). Environmental distances were calculated as pairwise differences in mean temperature, chlorophyll and turbidity between regions. Pairwise F_{ST}, R_{ST} and D were used as genetic distances.

All analyses were carried out at different spatial scales: at a large scale, all oceans included; each ocean considered in separate, i.e. all populations within the Atlantic and all populations within the Pacific Ocean and the population in the Southeast Indian Ocean; and at a medium scale, the North and Central Atlantic populations (hereinafter referred to as North Atlantic) and the South Pacific and Southeast Indian Ocean populations (hereinafter referred to as South Indo-Pacific).

Seascape genetics

Associations between genetic and environmental factors were examined using a hierarchical Bayesian method implemented in GESTE (Foll & Gaggiotti 2006), which estimates individual F_{ST} values for each local population and then relates them to environmental factors via a generalized linear model. Here we used 10 pilot runs of 1,000 iterations to obtain the parameters of the proposal distribution used by the MCMC, and an additional burn-in of 5 x 10^6 iterations with a thinning interval of 20. The model with the highest posterior probability is the one that best explains the data (Foll & Gaggiotti 2006).

Additionally, we used the BIOENV procedure of (Clarke & Ainsworth 1993) as implemented in PRIMER v.5 (Clarke & Warwick 2001) and as described in Geffen *et al.* (2004) to examine which predictor variable would provide the best model to explain the population genetic structure observed in the data. This procedure calculates the value of Spearman's rank correlation coefficient (ρ) between a genetic distance matrix (response matrix) with a distance matrix calculated as the Euclidean distance among one or more predictor variables. It then calculates the value of ρ using every possible combination of predictor variables until it finds the "best fit", corresponding to the combination of predictor variables whose Euclidean distance matrix yields the highest value of ρ (Clarke & Ainsworth 1993). We used three different response matrices corresponding to F_{ST}, R_{ST} and D distance matrices to identify the best one, two or three-variable fits.

Mantel tests (Mantel 1967) were also used to test for correlations between the pairwise genetic and environmental distances. Partial Mantel tests were used to control the effect of geographical distances in these potential correlations. These tests were performed using the package vegan in R.

6.4. RESULTS

6.4.1. *Genetic diversity*

In total 281 short-beaked common dolphin samples were genotyped at 14 microsatellite loci (Table 6.1). Deviations from Hardy-Weinberg equilibrium (HWE)

were found at 4 loci. Two of these (Tur91 and Tur80) showed deviations in only one population each and were therefore included in subsequent analyses, whereas the other two (Tur141 and Dde66) showed deviations in 4 and 2 populations, respectively. These deviations are due to a deficit of heterozygotes (significant F_{IS} values, Table 6.1). To test whether results would be affected by the inclusion of these two loci, estimates of genetic variability and differentiation were carried out with and without them. Since no major differences in results were observed (data not shown), all 14 loci were used in subsequent analyses.

Table 6.1. Genetic diversity measures obtained for the short-beaked common dolphin populations analysed in this study for 14 microsatellite loci.

Region	***N***	***Na***	***Ar***	H_E	H_O	F_{IS}
NE Atlantic (NEATL)	75	10.500	8.371	0.789	0.774	0.020
CE Atlantic (CEATL)	29	8.214	7.511	0.739	0.687	0.072
NW Atlantic (NWATL)	38	9.286	8.184	0.785	0.745	0.051
NE Pacific (NEPAC)	40	11.643	9.424	0.784	0.730	0.069*
SW Pacific Australia (SWPAC_AUS)	35	10.643	8.485	0.782	0.726	0.073*
SW Pacific New Zealand (SWPAC_NZ)	39	10.500	9.130	0.792	0.697	0.121*
SE Indian (SEIND)	25	7.571	7.163	0.700	0.696	0.006
Total / Mean	281	9.765	8.324	0.767	0.722	

N - sample size; *N*a - mean number of alleles; *A*r - allelic richness; H_E - expected heterozygosity; H_O - observed heterozygosity; F_{IS} - inbreeding coefficient. *value statistically significant at $P < 0.05$.

Levels of genetic diversity, given by mean number of alleles, allelic richness and expected and observed heterozygosities were high for most populations (Table 6.1). Significant F_{IS} values were obtained for populations from NE Pacific and SW Pacific Australia and New Zealand, which can be due to the presence of population sub-structure (i.e. Wahlund effect). In fact, this is known to be the case for common dolphins inhabiting those regions (Bilgmann *et al.* 2008; Chivers *et al*. 2009; Möller *et al*. 2011; Stockin *et al.* unpublished).

6.4.2. *Genetic differentiation*

Pairwise F_{ST}, R_{ST} and *D* comparisons showed significant levels of differentiation among all putative populations (Table 6.2), although the extent of that differentiation differed for each index. Overall *D* values were higher than F_{ST} and R_{ST} values. R_{ST} also tended to be higher than F_{ST}. Since R_{ST} is based on allele size, the differences observed indicate that mutation, in addition to drift or gene

flow may be affecting the differentiation between these populations. This result was confirmed using SPAGEDI. The overall R_{ST} value was significantly higher than the overall F_{ST} value ($P = 0.042$).

Table 6.2. Pairwise fixation index values obtained between short-beaked common dolphins populations for 14 microsatellite loci. a) F_{ST}; b) R_{ST} and c) D.

a) F_{ST}	NEATL	CEATL	NWATL	NEPAC	SWPACAUS	SWPACNZ	SEIND
NEATL							
CEATL	0.0150						
NWATL	0.0051	0.0151					
NEPAC	0.0313	0.0439	0.0284				
SWPACAUS	0.0267	0.0464	0.0228	0.0117			
SWPACNZ	0.0268	0.0471	0.0239	0.0211	0.0137		
SEIND	0.0680	0.0896	0.0716	0.0663	0.0473	0.0386	
b) R_{ST}	NEATL	CEATL	NWATL	NEPAC	SWPACAUS	SWPACNZ	SEIND
NEATL							
CEATL	0.0099						
NWATL	-0.0026	0.0069					
NEPAC	0.0341	0.0434	0.0335				
SWPACAUS	0.0122	0.0280	0.0059	0.0114			
SWPACNZ	0.0373	0.0671	0.0336	0.0720	0.0668		
SEIND	0.0430	0.0656	0.0419	0.0976	0.0497	0.0923	
c) D	NEATL	CEATL	NWATL	NEPAC	SWPACAUS	SWPACNZ	SEIND
NEATL							
CEATL	0.0082						
NWATL	0.0119	0.0103					
NEPAC	0.1136	0.1422	0.1090				
SWPACAUS	0.0687	0.1142	0.0673	0.0293			
SWPACNZ	0.0921	0.1398	0.0814	0.0234	0.0135		
SEIND	0.1479	0.1795	0.1670	0.1542	0.0835	0.0736	

Taken as a whole, the fixation indices showed high levels of differentiation between short-beaked populations inhabiting different ocean basins. The SEIND and NEPAC populations showed the highest levels of differentiation when compared with all other short-beaked populations. Contrasting to the inter-ocean basin differentiation, lower levels of differentiation were observed between short-beaked populations inhabiting the same ocean basins.

The first two principal components of the PCA analysis explained 84.35% of the variance in allele frequencies among putative populations (Figure 6.2). The first principal component shows a clear separation between populations inhabiting the Indo-Pacific and the Atlantic Oceans. The second principal component further shows some structure within the Indo-Pacific region, with the SEIND and NEPAC populations appearing separated from the SWPAC_AUS and SWPAC_NZ populations.

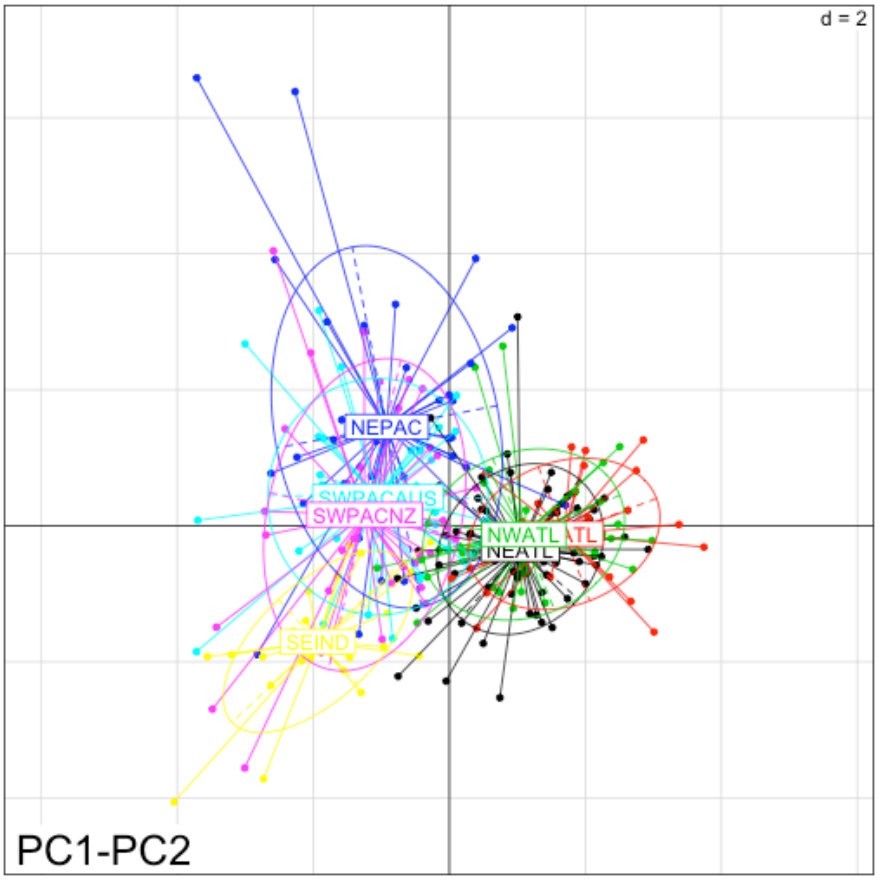


Figure 6.2. Principal component analysis principal component analysis (PCA) performed on a table of standardised allele frequencies of the short-beaked populations analysed in this study for 14 microsatellite loci.

Non metric MDS analyses using the three different genetic indices also show a clear separation from populations inhabiting the Atlantic, the Pacific and Indian oceans, with the exception of the analysis using R_{ST}, which grouped the NEPAC population with Atlantic ones (Figure 6.3). The analyses using F_{ST} and D show a closer proximity among the short-beaked populations inhabiting the North Atlantic, and also of the populations inhabiting the Pacific Ocean.

Results obtained in STRUCTURE using the correlated allele frequency model resulted in a peak of maximum ln $P(K)$ at K =3 (Figure 6.4). These clusters correspond to populations inhabiting the three ocean basins: the Atlantic (including the NEATL, NWATL and CEATL populations), the Pacific (including the NEPAC, SWPAC_AUS and SWPAC_NZ populations) and the Indian Ocean including the SEIND population (Figure 6.4).

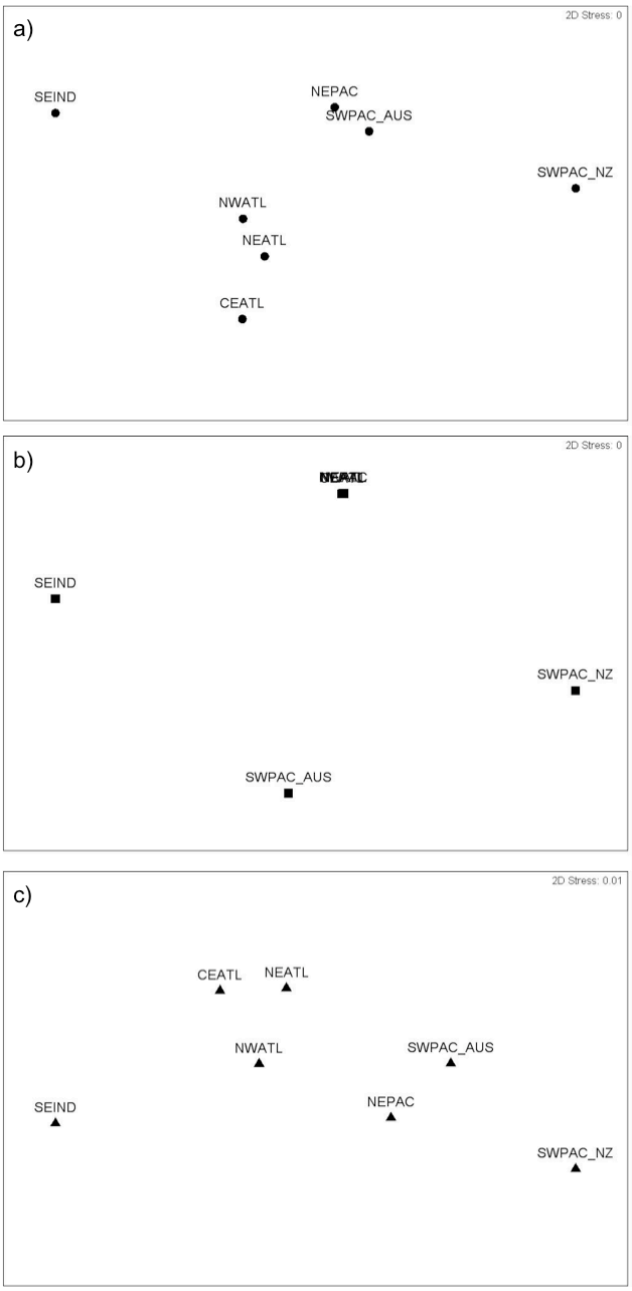


Figure 6.3. Non-metric MDS plots of short-beaked common dolphin populations on the basis of genetic distances using a) F_{ST}, b) R_{ST} or c) D. Stress values are indicated.

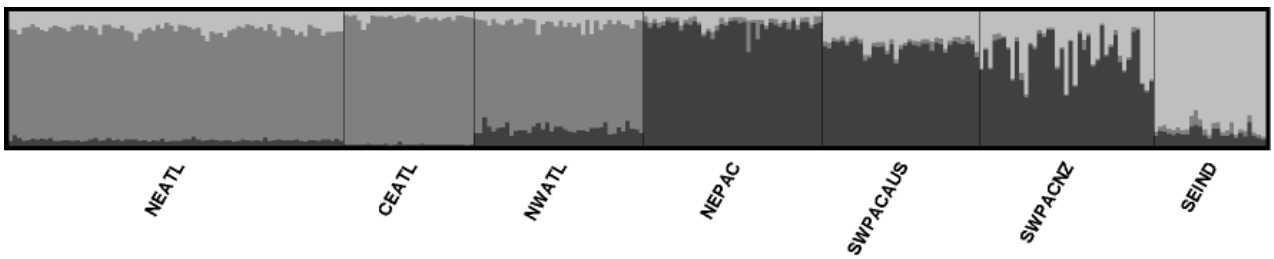


Figure 6.4. Results from the program STRUCTURE showing individual assignment values for $K = 3$. Each colour depicts the relative contribution of each of the three clusters to the genetic constitution of each individual.

The AMOVA analysis showed that the highest levels of differentiation were obtained when populations were divided by eastern versus western regions within ocean basins (F_{CT} = 0.03425, $P < 0.0001$) (Table 6.3).

Table 6.3. Analysis of hierarchical variance (AMOVA) results obtained for the short-beaked common dolphin populations.

Source of variation	%variation	*F*-statistics	*P*
Among ocean basins	2.71	F_{CT} = 0.02710	0.0000
Among groups within populations	1.35	F_{SC} = 0.01386	0.0000
Within populations	95.94	F_{ST} = 0.04058	0.0000
Among regions	1.92	F_{CT} = 0.03425	0.0001
Among groups within populations	1.5	F_{SC} = 0.01532	0.0000
Within populations	96.58	F_{ST} = 0.03425	0.0000

6.4.3. *Isolation by distance*

The relationship between geographic and genetic distance was only observed when populations inhabiting all oceans were considered in the analysis and when F_{ST} and *D* values were used (Table 6.4). This relationship was not detected when R_{ST} values were used, nor when finer spatial scales were considered.

Table 6.4. Summary results for Isolation by Distance tests conducted for all short-beaked common dolphin populations in all oceans, for North Atlantic populations only, for Pacific populations only, and for South Indo-Pacific populations only. Values in bold were statistically significant ($P < 0.05$).

	P	*r* (slope)	R^2
All oceans			
Fst	**0.0196**	0.0502	0.1560
Rst	0.9072	-0.0657	0.0416
D	**0.0091**	0.1240	0.4660
North Atlantic			
Fst	0.4995	-0.0211	0.2010
Rst	0.8351	-0.0239	0.4210
D	0.3316	0.0068	0.7740
Pacific			
Fst	0.3364	0.0573	0.0483
Rst	0.6241	-0.0840	0.0024
D	0.3328	0.1410	0.1150
South Indo-Pacific			
Fst	0.3310	0.0984	0.7860
Rst	0.4980	0.1209	0.1130
D	0.3321	0.2137	0.8760

6.4.4. *Oceanographic predictors*

Data on sea surface temperature (SST), chlorophyll concentration (CHL) and water turbidity (KD490) was gathered for the seven oceanic regions where short-beaked common dolphins were sampled: NEATL, CEATL, NWATL, NEPAC, SWPAC_AUS, SWPAC_NZ and SEIND (Figure 6.5). Paired t-tests showed significant differences in the 8 year average values of SST between most regions with exception of the comparison between NEATL and NWATL, between NEPAC and SWPAC (both AUS and NZ), and between NEPAC and SEIND, where differences were not statistically significant ($P < 0.01$, see Supplementary Material, Table S6.1). In the SST maps, all regions are heterogeneous, having regions of colder and warmer waters (Figure 6.5). Nevertheless, NEATL and NWATL regions are dominated by colder waters when compared with other regions, which are dominated by warmer waters, such as SWPAC_AUS and SWPAC_NZ. Significant differences were not detected in mean CHL values between NEPAC and SWPAC (both AUS and NZ) and between NEPAC and SEIND, as well as among SEIND, SWPAC_AUS and SWPAC_NZ. All other comparisons were significant. Despite this, in the CHL maps, clear differences can be seen among the regions located in the Pacific Ocean. Chlorophyll concentrations are higher in the NEPAC region closer to the coast when compared to the SWPAC_AUS and SWPAC_NZ regions. Regarding turbidity mean values, these were only not significant in the comparisons among SWPAC_AUS, SWPAC_NZ and SEIND (Table S6.1). Patterns seen in the maps are similar to the ones obtained for the CHL maps (Figure 6.5).

6.4.5. *Seascape genetics*

Hierarchical Bayesian analyses implemented in GESTE identified the model including the constant as the best one in all spatial scales considered (Table 6.5). The second best model for all analyses was the one including KD490, though the third and fourth models (including CHL and SST) all had very similar posterior probability values. Higher posterior probabilities were obtained when medium spatial scales were analysed. Positive signals of the regression coefficients were obtained for the association between CHL and genetic differentiation in the Pacific

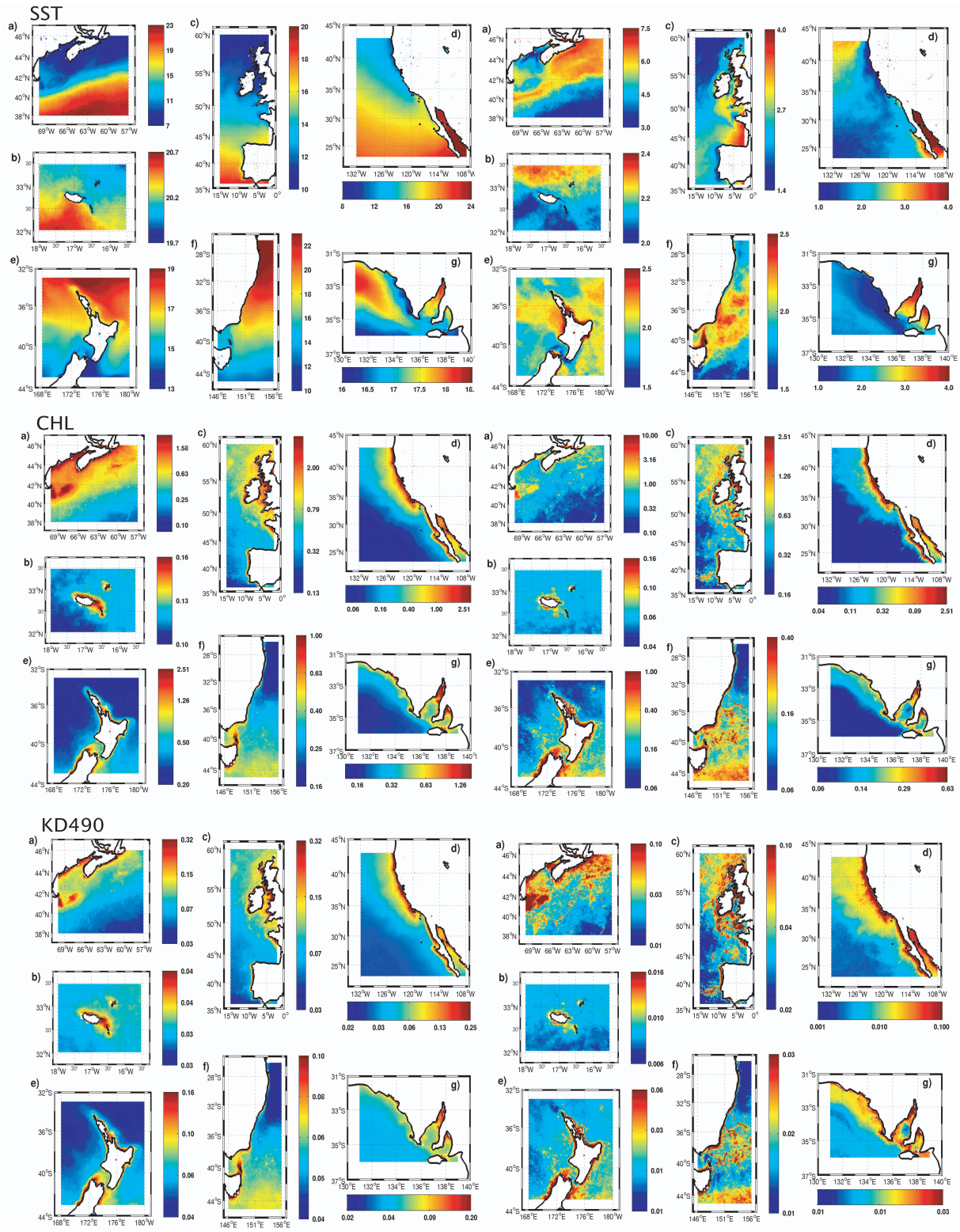


Figure 6.5. Regional maps showing 8-year average values for sea surface temperature (SST), chlorophyll concentration (CHL) and water turbidity (KD490) on the left and standard deviation values on the right for the oceanic regions where the short-beaked common dolphin populations analysed in this study were sampled: a) Northwest Atlantic; b) Central Eastern Atlantic; c) Northeast Atlantic; d) Northeast Pacific; e) Southwest Pacific New Zealand; f) Southwest Pacific Australia; g) Southeast Indian.

Table 6.5. Posterior probabilities of the four most probable models for the GESTE analysis of environmental associations with genetic structure (population specific F_{ST}) of common dolphins.

Model	Factors included	*P*	Coefficient	Mean	Mode	95% HPDI
All Oceans						
1	Constant	0.702	a0	-3.02	-3.01	-3.60; -2.43
			sigma	0.591	0.378	0.125; 1.319
2	Constant, SST	0.067	a0	-3.01	-2.99	-3.61; -2.33
			a1	0.13	0.12	-0.52; 0.73
			sigma	0.708	0.422	0.125; 1.70
3	Constant, CHL	0.0649	a0	-3	-3	-3.66; -2.36
			a2	-0.13	-0.11	-0.69; 0.56
			sigma	0.679	0.367	0.123; 1.501
5	Constant, KD490	0.0707	a0	-3.03	-3.05	-3.60; -2.32
			a3	-0.1	-0.1	-0.80; 0.53
			sigma	0.694	0.4	0.113; 1.726
Pacific						
1	Constant	0.628	a0	-3.08	-3.12	-4.02; -1.97
			sigma	1.094	0.701	0.173; 2.88
2	Constant, SST	0.092	a0	-3.1	-3.16	-4.30; 2.02
			a1	-0.04	-0.12	-1.26; -1.10
			sigma	1.42	0.695	0.198; 4.102
3	Constant, CHL	0.0991	a0	-3.04	-3.1	-4.16; -1.61
			a2	0.13	0.06	-1.07; 1.25
			sigma	1.63	0.713	0.140; 4.47
5	Constant, KD490	0.104	a0	-3.04	-3.17	-4.16; -1.85
			a3	0.14	0.16	-1.10; 1.23
			sigma	1.534	0.68	0.199; 4.601
North Atlantic						
1	Constant	0.496	a0	-3.25	-3.33	-4.52; -2.05
			sigma	1.14	0.677	0.097; 3.27
2	Constant, SST	0.101	a0	-3.22	-3.28	-4.59; -1.61
			a1	0.29	0.31	-0.97; 1.9
			sigma	1.557	0.774	0.114; 4.876
3	Constant, CHL	0.1	a0	-3.22	-3.3	-4.46; 1.63
			a2	-0.25	-0.25	-1.55; -1.08
			sigma	1.547	0.783	0.135; 5.112
5	Constant, KD490	0.103	a0	-3.19	-3.32	-4.45; -1.65
			a3	-0.27	-0.29	-1.85; -1.11
			sigma	1.694	0.86	0.134; 5.4
South Indo-Pacific						
1	Constant	0.501	a0	-2.95	-3	-4.26; -1.63
			sigma	1.481	0.825	0.146; 4.305
2	Constant, SST	0.0946	a0	-2.87	-3.1	-4.25; 0.95
			a1	0.14	0.19	-1.52; 1.64
			sigma	2.246	1.195	0.163; 7-064
3	Constant, CHL	0.0969	a0	-2.93	-2.99	-4.43; -1.06
			a2	0.08	0.13	-1.70; 1.65
			sigma	2.331	0.933	0.169; 7.64
5	Constant, KD490	0.171	a0	-2.96	-3.07	-4.27; -1.61
			a3	-0.54	-0.59	-1.84; 0.91
			sigma	1.678	0.765	0.124; 5.344

Ocean and South Indo-Pacific Ocean populations, and for the association between KD490 and genetic differentiation in the Pacific Ocean populations (Table 6.5). Regarding SST, positive signals of the regression coefficients were

obtained for all populations across all oceans, for the North Atlantic populations, and for the South Indo-Pacific populations (Table 6.5). Therefore, genetic isolation of populations within the Pacific Ocean increases with differences in CHL and KD490 among regions, whereas genetic isolation of populations within the Atlantic Ocean increases with differences in SST among regions. In the South Indo-Pacific region, both CHL and SST increase genetic isolation among populations. The percentage of variation that remained to be explained (indicated by sigma values) was however moderate (Table 6.5).

Table 6.6. Results of the BIOENV procedure, showing the best fit obtained, for all short-beaked common dolphin populations, North Atlantic populations only, Pacific populations only, and South Indo-Pacific populations only, in the case of one, two and three predictor variables for each genetic distance matrix.

Number variables	Spearman's rho	Variables chosen	Number variables	Spearman's rho	Variables chosen
All Oceans			North Atlantic		
Fst			Fst		
1	-0.341	CHL	1	1	KD490
2	-0.356	CHL, KD490	2	1	CHL, KD490
3	-0.227	SST, CHL, KD490	3	0.5	SST, CHL, KD490
JostD			JostD		
1	-0.366	CHL	1	-0.5	KD490
2	-0.374	CHL, KD490	2	-0.5	CHL, KD490
3	-0.31	SST, CHL, KD490	3	-1	SST, CHL, KD490
Rst			Rst		
1	-0.713	CHL	1	1	SST
2	-0.703	CHL, KD490	2	1	SST, CHL
3	-0.573	SST, CHL, KD490	3	1	SST, CHL, KD490
Pacific			South Indo-Pacific		
Fst			Fst		
1	-0.314	CHL	1	1	KD490
2	-0.371	CHL, KD490	2	-0.5	CHL, KD490
3	-0.029	SST, CHL, KD490	3	-0.5	SST, CHL, KD490
JostD			JostD		
1	-0.314	CHL	1	1	KD490
2	-0.714	CHL, KD490	2	0.5	CHL, KD490
3	-0.714	SST, CHL, KD490	3	-1	SST, CHL, KD490
Rst			Rst		
1	0.029	CHL	1	0.5	KD490
2	0.086	CHL, KD490	2	0.5	SST, KD490
3	-0.2	SST, CHL, KD490	3	0.5	SST, CHL, KD490

The BIOENV procedure found strong positive correlations between oceanographic predictors and genetic differentiation for the analyses conducted at medium spatial scales (Table 6.6). For the populations within the Atlantic Ocean and within the South Indo-Pacific, CHL and KD490 showed stronger correlation with genetic distance. For the larger spatial scales considered (across all oceans and within the Pacific Ocean), a strong negative correlation between CHL and KD490 with rank genetic distance was found (Table 6.6).

Mantel tests and Partial Mantel tests between genetic and environmental distances were not statistically significant for any comparison, even considering different spatial scales (results not shown). Failures of these tests to detect relationships between genetic and environmental data have been previously described (Legendre & Fortin 2010; Raufaste & Rousset 2001) and could explain the unsuccessful use with our datasets.

6.5. DISCUSSION

We used a seascape approach to investigate the interaction between a set of oceanographic variables and population structure in a highly mobile, widely distributed top marine predator, the short-beaked common dolphin. We show that SST, chlorophyll concentration and water turbidity seem to be important factors in explaining the observed patterns of genetic structure in these dolphins, more than geographical distance alone, particularly when medium spatial scales were considered.

6.5.1. *Genetic structure*

The overall global pattern of genetic structure obtained here supports previous studies (Amaral *et al.* in review; Natoli *et al.* 2006): higher levels of differentiation were obtained across large geographical scales, between different ocean basins, and lower levels were obtained when medium geographical scales were considered, within the same ocean basin. While results from STRUCTURE showed a clear differentiation between ocean basins, the AMOVA analysis resulted in higher F_{CT} estimates for partitioning of short-beaked populations among regions within each ocean basin. The low levels of divergence found

between populations inhabiting the same ocean basin may have affected the power of the program STRUCTURE to detect such differentiation, even using recently developed algorithms that account for weak differentiation (Hubisz *et al.* 2009). Nonetheless, the PCA and the NMDS plots also indicate some level of differentiation within ocean basins, which seems to be stronger among the Pacific Ocean populations. Multivariate analysis does not require strong assumptions about the underlying genetic model, such as Hardy-Weinberg equilibrium or the absence of linkage disequilibrium (Jombart *et al.* 2009). The high levels of differentiation found for the SEIND population (southern Australia) were surprising given the comparatively shorter distance separating this population from the Southwest Pacific populations (off New South Wales, southeastern Australia), even considering that the region where the SEIND population was sampled (off South Australia) falls into a different biogeographic region (see Waters *et al.* 2010) to the one of the SWPAC_AUS population. Such high differentiation was also reported by Bilgmann *et al.* (2008) when comparing individuals from this region to individuals from southeastern Tasmania (Southwest Pacific) – in that case oceanographic features affecting the distribution of target prey were suggested to be the likely explanation for the genetic differentiation found. Our study corroborates this previous finding (see below).

6.5.2. *Isolation by distance*

A pattern of isolation by distance was only observed when large spatial scales were considered, indicating that the stronger genetic differentiation observed in short-beaked common dolphins from different oceans may be an effect of geographic distance. Isolation by distance has been reported for other cetacean species, such as in the harbour porpoise (Fontaine *et al.* 2007) and in bottlenose dolphins (Krutzen *et al.* 2004). Conversely, when medium geographic scales were considered (i.e. within each ocean basin), no isolation by distance effect was detected, and genetic differentiation could be explained by oceanographic variables. This pattern has also been described for common dolphins at small geographical scales, along the eastern Australian coast (Möller *et al.* 2011), for bottlenose dolphins in South Australia where a temperature and salinity front

coincides with the boundary between two distinct genetic populations (Bilgmann *et al.* 2007), and for pilot whales, where ecological factors, such as SST, were more important in explaining genetic structure than geographic separation (Fullard *et al.* 2000). In franciscana and humpback dolphins, environmental factors were also more important in explaining genetic structure than distance at small geographical scales (Mendez *et al.* 2010; Mendez *et al*. 2011).

6.5.3. *Oceanographic predictors*

All oceanographic variables tested, CHL, KD490 and SST, showed an association with population genetic structure in short-beaked common dolphins. These associations were strongest at the medium spatial scales considered. In the Pacific Ocean, CHL and KD490 were the environmental predictors that were most strongly associated with increased genetic isolation in short-beaked common dolphins. Conversely, in the Atlantic Ocean, SST was the strongest predictor associated with population divergence. Although no significant statistical differences in the 8-year average values of CHL and KD490 were detected among regions in the Pacific Ocean, a visual inspection of the regional maps shows heterogeneity in these variables among regions (Figure 6.5). Heterogeneity in SST, CHL and KD490 is also seen among Atlantic Ocean regions, although our results suggest that only SST seems to explain genetic differentiation of short-beaked common dolphins in this area. Marine productivity and SST are important variables for habitat occupancy and dispersal in cetaceans (Forney 2000; Hamazaki 2002) and have been shown to influence population structure in franciscana (Mendez *et al*. 2010) and in humpback dolphins (Mendez *et al.* 2011). Here, we suggest that they are also important drivers of population structure in common dolphins. A direct causality is however difficult to establish. For example, it has been suggested that ecological factors such as prey behaviour rather than inherent sensitivity to environmental factors, could account for the relationship between SST and population structure in pilot whales (Fullard *et al.* 2000; Kasuya *et al.* 1988). Similarly, differences in prey distribution and abundance between regions rather than SST differences themselves are suggested to account for genetic differentiation of bottlenose dolphins in South Australia (Bilgmann *et al.*

2007) and short-beaked common dolphins in southern (Bilgmann *et al.* 2008) and southeastern Australia (Möller *et al.* 2011). We suggest that a similar process may account for the patterns obtained in this study. Since dolphins feed high in the food chain, a statistical association with oceanographic variables that do not directly affect the individuals, but rather affect their prey, is expected to be weak (Ballance *et al.* 2006). This could also explain the fact that analyses performed in GESTE did not result in a single best-chosen model and that the percentage of variability that remained to be explained in the data was moderate.

Chlorophyll concentration, water turbidity and SST are routinely used to map ocean primary productivity (e.g. Gremillet *et al.* 2008). Due to the bottom-up processes that control marine ecosystems (Frank *et al.* 2007), these variables have been related to prey distribution and abundance, and to the occurrence of top marine predators (e.g. Bailleul *et al.* 2005; Pinaud & Weimerskirch 2007). Distribution and abundance of prey has been suggested as the main factor dictating seasonal migrations in several species of delphinids, including short-beaked common dolphin (e.g. Young & Cockcroft 1994). Moreover, short-beaked common dolphins feed primarily on small mesopelagic schooling fish such as sardines and anchovies (Pusineri *et al.* 2007; Young & Cockcroft 1994). These fishes are filter feeders and occur in association with nutrient rich waters (e.g. Bowen & Grant 1997), and could explain the dolphins' preference for certain oceanographic conditions.

We further suggest that a behavioural mechanism such as specialization for local resources could also explain the patterns observed. Resource specialization is a common mechanism driving population structure in delphinds (Hoelzel 2009). Moreover, dietary segregation is known to occur in short-beaked common dolphins. In the Bay of Biscay, Northeast Atlantic Ocean, common dolphins inhabiting neritic and oceanic waters feed on different prey species (Lahaye *et al.* 2005). Feeding specialization leading to local adaptation has also been suggested as driving speciation of the short and long-beak forms (Amaral *et al.* in review; Natoli *et al.* 2006), and as important triggers for the process of population divergence and speciation in the genera *Tursiops* and *Stenella* (Natoli *et al.* 2004;

Perrin *et al.* 1987). Perhaps the best studied example within delphinids are killer whales (*Orcinus orca*), where resource partitioning and foraging specializations of sympatric populations occurring in the North Pacific have lead to the evolution of distinct lineages (Morin *et al.* 2010). Short-beaked common dolphins could therefore be locally adapted to the existent prey species and only move within certain regions following prey migration. Seasonal migrations are known to occur in the Northeast Pacific (Forney & Barlow 1998) and Southwest Indian Ocean (Cockcroft & Peddemors 1990). Further investigation is however required to support this hypothesis.

There are also other factors that may account for population divergence in common dolphins that were not assessed in this study. Fine-scale oceanic processes, for example, have recently been suggested to affect connectivity in common dolphins (Möller *et al*. 2011). A proper assessment of its direct relationship with genetic structure requires knowledge on hydrodynamic modelling and will certainly be the aim of forthcoming studies. Demographic and historical processes can also contribute to population structure and should also be integrated in future analyses.

6.5.4. *Implications for conservation and management*

The results presented here are of particular importance for marine conservation management and design of marine protected areas (MPA). MPAs are usually designed to protect coastal regions that are either important habitats, as part of the marine ecosystem, or biodiversity hotspots (Agardy 1994). Marine predators are often used as indicators for MPA design, because their protection aids in protecting the more complex environments they use (Bailey & Thompson 2009; Hooker & Gerber 2004; Zacharias & Roff 2001). Although several studies have described the distribution and occurrence of cetacean species in relation to different habitat variables (e.g. Canadas *et al.* 2005; Canadas *et al.* 2002; Panigada *et al.* 2008), only a few have found a direct correlation between oceanographic variables and population structure (Mendez *et al.* 2010; Mendez *et al.* 2011). In this study, by showing how marine productivity correlate with population structure in short-beaked common dolphins, we highlight the

importance of using seascape genetic studies to inform MPA design in relation to distribution and genetic connectivity of charismatic and ecologically important megafauna. Furthermore, we highlight how such an approach can track the biological effects of ongoing climate-change and prevent the loss of top marine predators (Myers *et al.* 2007).

6.6. CONCLUSION

Understanding which factors shape the distribution of intraspecific genetic diversity is central in evolutionary and conservation biology. In the marine realm, most such studies have focused on organisms with larval dispersal. Here, using a seascape approach, we show that marine productivity and sea surface temperature are correlated with population structure in a highly mobile, widely distributed marine mammal species, the short-beaked common dolphin. We also highlight how this kind of approach can inform MPA design and consequently track the ongoing effects of climate-change on the distribution and connectivity of top marine predators.

6.7. ACKNOWLEDGEMENTS

We would like to acknowledge all who kindly provided tissue samples: Jennifer Learmonth (SAC-Scottish Agricultural College and DEFRA), Luca Mirimin (University of Cork, Ireland), Marisa Ferreira (Sociade Portuguesa de Vida Selvagem). A. R. Amaral acknowledges a PhD grant (SFR/BD/27245/2006) from Fundação para a Ciência e Tecnologia (Portugal). Sample collection in Australia was funded by Macquarie University through research grants to L. Möller, L. Beheregaray (MQ A006162) and K. Bilgmann. Statistical analyses were partially conducted using the Computational Biology Service Unit from Cornell University. This represents article contribution #41 of the Molecular Ecology Group for Marine Research (MEGMAR).

6.8. REFERENCES

Agardy MT (1994) Advances in marine conservation - The role of marine protected areas. *Trends in Ecology & Evolution* **9**, 267-270.

Amaral AR, Sequeira M, Cedeira-Martínez J, Coelho MM (2007) New insights on population genetic structure of *Delphinus delphis* from the northeast Atlantic and phylogenetic relationships within the genus inferred from two mitochondrial markers. *Marine Biology* **151**, 1967-1976.

Amaral AR, Beheregaray LB, Sequeira M, Robertson KM, Coelho MM, Möller LM (2009) Worldwide Phylogeography of the Genus *Delphinus* Revisited. International Whaling Commission Report SC/61/SM11.

Bailey H, Thompson PM (2009) Using marine mammal habitat modelling to identify priority conservation zones within a marine protected area. *Marine Ecology-Progress Series* **378**, 279-287.

Bailleul F, Luque S, Dubroca L, Arnould JPY, Guinet C (2005) Differences in foraging strategy and maternal behaviour between two sympatric fur seal species at the Crozet Islands. *Marine Ecology-Progress Series* **293**, 273-282.

Ballance LT, Pitman RL, Fiedler PC (2006) Oceanographic influences on seabirds and cetaceans of the eastern tropical Pacific: A review. *Progress in Oceanography* **69**, 360-390.

Banks SC, Piggott MP, Williamson JE, BoveU, Holbrook NJ, Beheregaray LB (2007) Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology* **88**, 3055-3064.

Bilgmann K, Möller LM, Harcourt RG, Gales R, Beheregaray LB (2008) Common dolphins subject to fisheries impacts in Southern Australia are genetically differentiated: Implications for conservation. *Animal Conservation* **11**, 518-528.

Bilgmann K, Möller LM, Harcourt RG, Gibbs SE, Beheregaray LB (2007) Genetic differentiation in bottlenose dolphins from South Australia: Association with local oceanography and coastal geography. *Marine Ecology-Progress Series* **341**, 265-276.

Bost CA, Cotte C, Bailleul F, Charrassin JB, Guinet C, Ainley DG, Weimerskirch H (2009) The importance of oceanographic fronts to marine birds and mammals of the southern oceans. *Journal of Marine Systems* **78**, 363-376.

Bowen BW, Grant WS (1997) Phylogeography of the sardines (*Sardinops* spp): Assessing biogeographic models and population histories in temperate upwelling zones. *Evolution* **51**, 1601-1610.

Brown DM, Brenneman RA, Koepfli KP, Pollinger JP, Mila B, Georgiadis NJ, Louis EE, Grether GF, Jacobs DK, Wayne RK (2007) Extensive population genetic structure in the giraffe. *Bmc Biology* **5**, 13.

Canadas A, Sagarminaga R, De Stephanis R, Urquiola E, Hammond PS (2005) Habitat preference modelling as a conservation took proposals for marine protected areas for cetaceans in southern Spanish waters. *Aquatic conservation: marine and freshwater ecosystems* **15**, 495-521.

Canadas A, Sagarminaga R, Garcia-Tiscar S (2002) Cetacean distribution related with depth and slope in the Mediterranean waters off southern Spain. *Deep-Sea Research Part I-Oceanographic Research Papers* **49**, 2053-2073.

Chivers SJ, Hedrick NM, LeDuc CA (2009) Genetic Evidence for Population Structure in Eastern North Pacific *Delphinus delphis*. International Whaling Commission. SC/61/For Info 57.

Clarke KR, Ainsworth M (1993) A method of linking multivariate community structure to environmental variables. *Marine Ecology-Progress Series* **92**, 205-219.

Clarke, KR, Warwick RM (2001). Change in marine communities: An approach to statistical analysis and interpretation, 2nd edition. PRIMER-E, Plymouth.

Cockcroft VG, Peddemors VM (1990). Seasonal distribution and density of common dolphins *Delphinus delphis* off the south-east coast of southern Africa. *South African Journal of Marine Science* **9**: 371-377.

Coughlan J, Mirimin L, Dillane E, Rogan E, Cross TF (2006) Isolation and characterization of novel microsatellite loci for the short-beaked common dolphin (*Delphinus delphis*) and cross-amplification in other cetacean species. *Molecular*

Ecology Notes **6**, 490-492.

Crawford NG (2010) smogd: software for the measurement of genetic diversity. *Molecular Ecology Resources* **10**, 556-557.

Doksaeter L, Olsen E, Nottestad L, Ferno A (2008) Distribution and feeding ecology of dolphins along the Mid-Atlantic Ridge between Iceland and the Azores. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **55**, 243-253.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of Populations. *Genetics* **174**, 875-891.

Fontaine MC, Baird SJE, Piry S, Ray N, Tolley KA, Duke S, Birkun A, Ferreira M, Jauniaux T, Llavona A, Ozturk B, Ozturk AA, Ridoux V, Rogan E, Sequeira M, Siebert U, Vikingsson GA, Bouquegneau JM, Michaux JR (2007) Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *Bmc Biology* **5**, 30.

Forney KA (2000) Environmental models of cetacean abundance: Reducing uncertainty in population trends. *Conservation Biology* **14**, 1271-1286.

Forney KA, Barlow J (1998) Seasonal patterns in the abundance and distribution of California cetaceans, 1991-1992. *Marine Mammal Science* **14**, 460-489.

Frank KT, Petrie B, Shackell NL (2007) The ups and downs of trophic control in continental shelf ecosystems. *Trends in Ecology & Evolution* **22**, 236-242.

Fullard KJ, Early G, Heide-Jorgensen MP, Bloch D, Rosing-Asvid A, Amos W

(2000) Population structure of long-finned pilot whales in the North Atlantic: A correlation with sea surface temperature? *Molecular Ecology* **9**, 949-958.

Galindo HM, Olson DB, Palumbi SR (2006) Seascape genetics: A coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology* **16**, 1622-1626.

Geffen E, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology* **13**, 2481-2490.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Gremillet D, Lewis S, Drapeau L, van Der Lingen CD, Huggett JA, Coetzee JC, Verheye HM, Daunt F, Wanless S, Ryan PG (2008) Spatial match-mismatch in the Benguela upwelling zone: Should we expect chlorophyll and sea-surface temperature to predict marine predator distributions? *Journal of Applied Ecology* **45**, 610-621.

Hamazaki T (2002) Spatiotemporal prediction models of cetacean habitats in the mid-western North Atlantic Ocean (from Cape Hatteras, North Carolina, USA to Nova Scotia, Canada). *Marine Mammal Science* **18**, 920-939.

Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. *Genetics* **163**, 1467-1482.

Hellberg ME (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology Evolution and Systematics* **40**, 291-310.

Hoelzel AR, Dahlheim M, Stern SJ (1998) Low genetic variation among killer whales (*Orcinus orca*) in the eastern North Pacific and genetic differentiation between foraging specialists. *Journal of Heredity* **89**, 121-128.

Hoelzel AR (2009) Evolution of Population Genetic Structure in Marine Mammal Species. Pages 294-318. *in* Bertorelle G, Bruford MW, Hauffe HC eds. *Population Genetics for Animal Conservation*. Cambridge University Press.

Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* **58**, 199-207.

Hooker SK, Gerber LR (2004) Marine reserves as a tool for ecosystem-based management: The potential importance of megafauna. *Bioscience* **54**, 27-39.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.

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

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *Bmc Genetics* **6**, 13.

Jombart T, Pontier D, Dufour AB (2009) Genetic markers in the playground of multivariate analysis. *Heredity* **102**, 330-341.

Jost L (2008) G(ST) and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015-4026.

Kasuya T, Myashita T, Kasamatsu F (1988) Segregation of two forms of short-finned pilot whales off the Pacific coast of Japan. *Scientific Report of the Whales Research Institute* **39**, 77-90.

Krutzen M, Sherwin WB, Berggren P, Gales N (2004) Population structure in an inshore cetacean revealed by microsatellite and mtDNA analysis: Bottlenose dolphins (*Tursiops* sp.) in Shark Bay, Western Australia. *Marine Mammal Science* **20**, 28-47.

Krutzen M, Valsecchi E, Connor RC, Sherwin WB (2001) Characterization of microsatellite loci in *Tursiops aduncus*. *Molecular Ecology Notes* **1**, 170-172.

Kruskal JB, Wish M (1978) *Multidimensional Scaling*. Sage University Paper series on Quantitative Applications in the Social Sciences, number 07-011. Sage Publications, Newbury Park, California.

Lahaye V, Bustamante P, Spitz J, Dabin W, Das K, Pierce GJ, Caurant F (2005) Long-term dietary segregation of common dolphins *Delphinus delphis* in the Bay of Biscay, determined using cadmium as an ecological tracer. *Marine Ecology-*

Progress Series **305**, 275-285.

Legendre P, Fortin MJ (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* **10**, 831-844.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Mendez M, Subramaniam A, Collins T, Minton G, Baldwin R, Berggren P, Sarnblad A, Amir OA, Peddemors VM, Karczmarski L, Guissamulo A, Rosenbaum HC (2011) Molecular ecology meets remote sensing: Environmental drivers to population structure of humpback dolphins in the Western Indian Ocean. *Heredity* doi:10.1038hdy.2011.21.

Mendez M, Rosenbaum HC, Subramaniam A, Yackulic C, Bordino P (2010) Isolation by environmental distance in mobile marine species: molecular ecology of franciscana dolphins at their southern range. *Molecular Ecology* **19**, 2212-2228.

Mirimin L, Westgate A, Rogan E, Rosel P, Read A, Coughlan J, Cross T (2009) Population structure of short-beaked common dolphins (*Delphinus delphis*) in the North Atlantic Ocean as revealed by mitochondrial and nuclear genetic markers. *Marine Biology* **156**, 821-834.

Möller LM, Valdez FP, Allen S, Bilgmann K, Corrigan S, Beheregaray LB (2011) Fine-scale genetic structure in short-beaked common dolphins (*Delphinus delphis*) along the East Australian Current. *Marine Biology* **158**, 113-126.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Nielsen SCA, Rasmussen M, Willerslev E, Gilbert MTP, Harkins T (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* **20**, 908-916.

Myers RA, Baum JK, Shepherd TD, Powers SP, Peterson CH (2007) Cascading effects of the loss of apex predatory sharks from a coastal ocean. *Science* **315**, 1846-1850.

Nater A, Kopps AM, Krutzen M (2009) New polymorphic tetranucleotide microsatellites improve scoring accuracy in the bottlenose dolphin *Tursiops aduncus. Molecular Ecology Resources* **9**, 531-534.

Natoli A, Birkun A, Aguilar A, Lopez A, Hoelzel AR (2005) Habitat structure and the dispersal of male and female bottlenose dolphins (*Tursiops truncatus*). *Proceedings of the Royal Society of London Series B*-Biological Sciences **272**, 1217-1226.

Natoli A, Cañadas A, Peddemors VM*, et al.* (2006) Phylogeography and alpha taxonomy of the common dolphin (*Delphinus* sp.). *Journal of Evolutionary Biology* **19**, 943-954.

Natoli A, Peddemors VM, Hoelzel AR (2004) Population structure and speciation in the genus *Tursiops* based on microsatellite and mitochondrial DNA analyses. *Journal of Evolutionary Biology* **17**, 363-375.

Oosterhout CV, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Panigada S, Zanardelli M, MacKenzie M*,* Donovan C, Melin F, Hammond PS (2008) Modelling habitat preferences for fin whales and striped dolphins in the Pelagos Sanctuary (Western Mediterranean Sea) with physiographic and remote sensing variables. *Remote Sensing of Environment* **112**, 3400-3412.

Perrin WF (2009) Common Dolphins, *Delphinus delphis* and *D. capensis*. In: *Encyclopedia of Marine Mammals* (eds. Perrin WF, Wursig B, Thewissen JGM), pp. 255-259. Academic Press, New York.

Perrin WF, Mitchell ED, Mead JG*,* Caldwell DK, Caldwell MC, Vanbree PJH, Dawbin WH (1987) Revision of the spotted dolphins, *Stenella* Spp. *Marine Mammal Science* **3**, 99-170.

Pinaud D, Weimerskirch H (2007) At-sea distribution and scale-dependent foraging behaviour of petrels and albatrosses: a comparative study. *Journal of Animal Ecology* **76**, 9-19.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure

using multilocus genotype data. *Genetics* **155**, 945-959.

Pusineri C, Magnin V, Meynier L, Spitz J, Hassani S, Ridoux V (2007) Food and feeding ecology of the common dolphin (*Delphinus delphis*) in the oceanic Northeast Atlantic and comparison with its diet in neritic areas. *Marine Mammal Science* **23**, 30-47.

Raufaste N, Rousset F (2001) Are partial mantel tests adequate? *Evolution* **55**, 1703-1705.

Rosel PE, Block BA (1996) Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius*. *Marine Biology* **125**, 11-22.

Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.

Rousset F (2008) GENEPOP ' 007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments. *Nature* **18**, 233–234.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries* **9**, 363-377.

Selkoe KA, Watson JR, White C*,* Ben Horin T, Iacchei M, Mitarai S, Siegel DA, Toonen RJ (2010) Taking the chaos out of genetic patchiness: Seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology* **19**, 3708-3726.

Storfer A, Murphy MA, Evans JS*,* Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierlig L, Waits LP (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128-142.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: Where are we now? *Molecular Ecology* **19**, 3496-3514.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution* **13**, 510-524.

Valsecchi E, Amos W (1996) Microsatellite markers for the study of cetacean populations. *Molecular Ecology* **5**, 151-156.

Viricel A, Strand AE, Rosel PE, Ridoux V, Garcia P (2008) Insights on common dolphin (*Delphinus delphis*) social organization from genetic analysis of a mass-stranded pod. *Behavioral Ecology and Sociobiology* **63**, 173-185.

Waters JM, Wernberg T, Connell SD, Thomsen MS, Zucarello GC, Kraft GT, Sanderson JC, West JA, Gurgel CFD (2010) Australia's marine biogeography revisited: Back to the future? *Austral Ecology* **35**, 988-992.

Young DD, Cockcroft VG (1994) Diet of common dolphins (*Delphinus delphis*) off the southeast coast of southern Africa - Opportunism or specialization. *Journal of Zoology* **234**, 41-53.

Zacharias MA, Roff JC (2001) Use of focal species in marine conservation and management: a review and critique. *Aquatic Conservation-Marine and Freshwater Ecosystems* **11**, 59-76.

6.9. SUPPLEMENTARY MATERIAL

Table S6.1. Mean pairwise difference between average values of a) sea surface temperature (SST), b) chlorophyll concentration (CHL) and c) water turbidity (KD490) obtained for each oceanographic region where short-beaked common dolphins were sampled for this study, with significant values of paired t-tests indicated in bold.

a) SST

	NEATL	CEATL	NWATL	NEPAC	SWPACAUS	SWPACNZ	SEIND
NEATL							
CEATL	**6.47886**						
NWATL	0.38654	**6.09231**					
NEPAC	**4.00799**	**2.47087**	**-3.62145**				
SWPACAUS	**3.98241**	**-2.49645**	**3.59586**	-0.02558			
SWPACNZ	**2.88076**	**3.59809**	**2.49422**	-1.12723	**1.10164**		
SEIND	**3.44226**	**3.03660**	**3.05571**	-0.56573	**0.54015**	**-0.56149**	

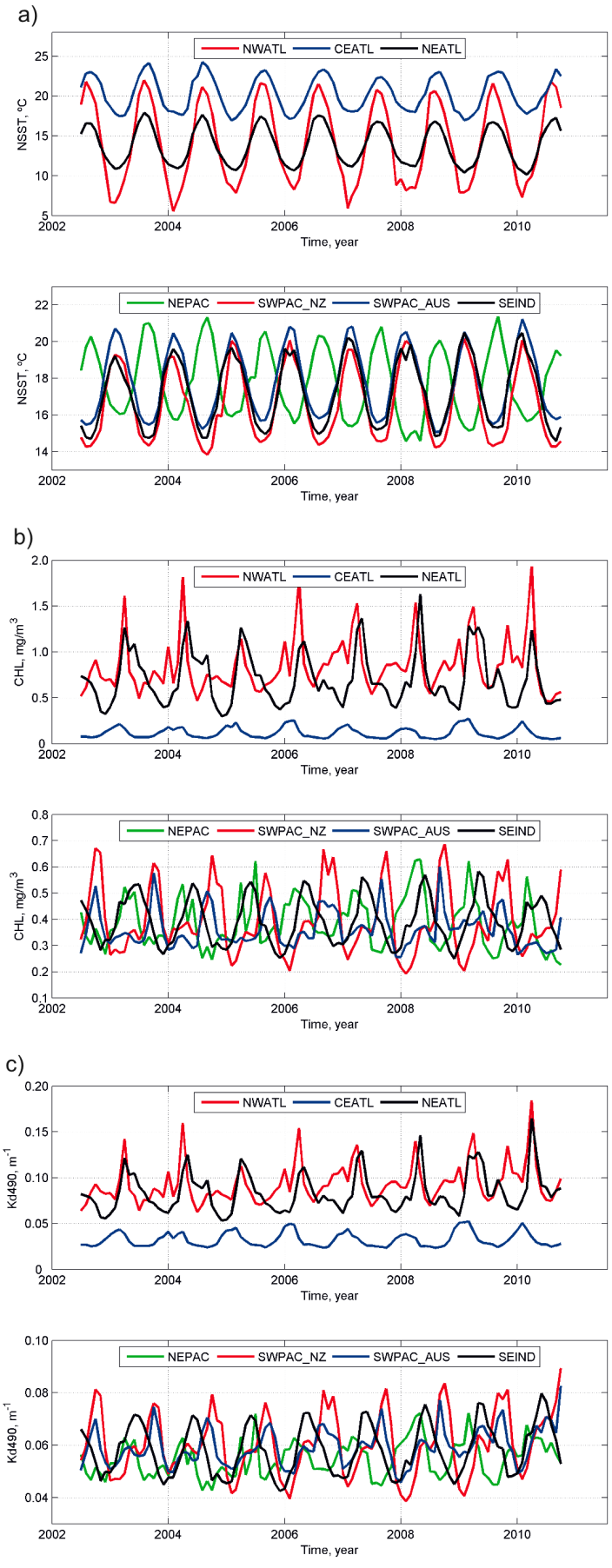
b) CHL

	NEATL	CEATL	NWATL	NEPAC	SWPACAUS	SWPACNZ	SEIND
NEATL							
CEATL	**-0.58515**						
NWATL	**0.14719**	**-0.73234**					
NEPAC	**-0.31631**	**-0.26884**	**0.46350**				
SWPACAUS	**-0.33893**	**0.24622**	**-0.48612**	-0.02262			
SWPACNZ	**-0.30756**	**-0.27759**	**-0.45475**	0.00875	**-0.03137**		
SEIND	**-0.30755**	**-0.27761**	**-0.45473**	0.00877	-0.03138	-0.00001	

c) KD490

	NEATL	CEATL	NWATL	NEPAC	SWPACAUS	SWPACNZ	SEIND
NEATL							
CEATL	**-0.05175**						
NWATL	**0.00928**	**-0.06103**					
NEPAC	**-0.02937**	**-0.02238**	**0.03865**				
SWPACAUS	**-0.02481**	**0.02694**	**-0.03409**	**0.00456**			
SWPACNZ	**-0.02390**	**-0.02785**	**-0.03318**	**0.00547**	-0.00091		
SEIND	**-0.02593**	**-0.02583**	**-0.03521**	**0.00344**	0.00112	0.00203	

Figure S6.1. Annual average values for (a) sea surface temperature, (b) chlorophyll concentration and (c) water turbidity for the different oceanographic regions.



Chapter VII
General Discussion

7.1. GENERAL DISCUSSION

This dissertation combines the use of several types of molecular markers and innovative analytical approaches in order to contribute to a better understanding of the evolutionary history of the subfamily Delphininae, with particular focus on the genus *Delphinus*.

Molecular studies of cetaceans, including delphinids, have generally been based on the use of mitochondrial DNA (mtDNA) and microsatellite loci, with the exception of a few studies that have used Amplified Fragment Length Polymorphisms (Kingston *et al.* 2009; Kingston & Rosel 2004) and nuclear intron sequences (Caballero *et al.* 2008). Nonetheless, there is growing recognition that many questions related to phyogenetics and phylogeography should not be addressed solely by mtDNA markers, but also include multiple nuclear DNA markers. This is because a single locus with maternal inheritance such as the mtDNA, can be prone to errors when representing the entire population or species history, while multiple nuclear DNA loci provide replicate samples of the coalescent process, thus offering greater power to estimate demographic parameters and/or species trees (Carling & Brumfield 2007; Felsenstein 2006; Lee & Edwards 2008). A set of 17 anonymous nuclear markers was developed and characterized from a common dolphin (genus *Delphinus*) genomic library, and was successfully tested across several cetacean species, thus representing a significant addition to the set of tools for use in genetic studies of cetaceans. A subset of 10 markers, which were screened for polymorphism within *Delphinus,* revealed an average of 1 single nucleotide polymorphism (SNP) per 272 bp sequenced, demonstrating the utility of these markers for rapid SNP discovery in cetaceans (Chapter II). Moreover, these markers proved to be informative for phylogenetic (Chapter IV) and phylogeographic (Chapter V) studies of delphinids.

The study of genes likely involved in the establishment of reproductive isolation can provide insights into the process of speciation and on the evolutionary history of a group of species (Wu & Ting 2004). Recent research has focused on the

study of reproductive proteins involved in the fertilization process, which mediate the sperm-egg interaction (Turner & Hoekstra 2008). The rapid evolution of these proteins has been documented in several animal taxa (Calkins *et al.* 2007; Metz *et al.* 1998; Turner & Hoekstra 2006), which is thought to be the result of forces involved in sexual selection (Swanson & Vacquier 2002). In cetaceans, several mating systems have been reported as promiscuous, leading to the likely existence of sperm competition (Connor *et al.* 2000). This would suggest that the evolution of reproductive proteins is also likely rapid in this group and driven by positive selection. Nevertheless, the study of the pattern of evolution in two reproductive proteins, ZP3 and PKDREJ, across 18 cetacean species revealed very low levels of amino acid sequence divergence, a very weak signal of positive selection for ZP3 and no signal of positive selection for PKDREJ (Chapter III). The slower rate of evolution of these proteins in cetaceans when compared to other mammals is consistent with previous reports of a general slower evolution of the cetacean genome (Jackson *et al.* 2009). Nevertheless, some pressure for the rapid evolution of these proteins would still be expected in promiscuous systems, such as those reported for several cetacean species. The results obtained in this study launch a discussion on the evolutionary forces driving the evolution of reproductive proteins in cetaceans and on the potential alternative processes that may be dictating the establishment of reproductive isolation and species recognition. It may be that these mechanisms are not entirely molecular but are also behavioural. Such an understanding is, however, crucial for the study of speciation and ultimately for the establishment of species boundaries in taxonomic confusing, closely related groups, such as the Delphininae.

The dolphins within the subfamily Delphininae present a confusing and controversial taxonomy, which is mainly due to the disagreement found between the taxonomy originally established by morphological characters (e.g. Flower 1883), and the phylogenetic relationships subsequently supported by molecular studies (e.g. LeDuc *et al.* 1999). Using coalescent-based methods based on 13 nuclear DNA loci sequences, a species tree that agrees with morphology-based

species relationships within the subfamily Delphininae was obtained (Chapter IV). To the best of my knowledge this was the first time that both such an approach was undertaken and that such result was obtained (Figure 7.1).

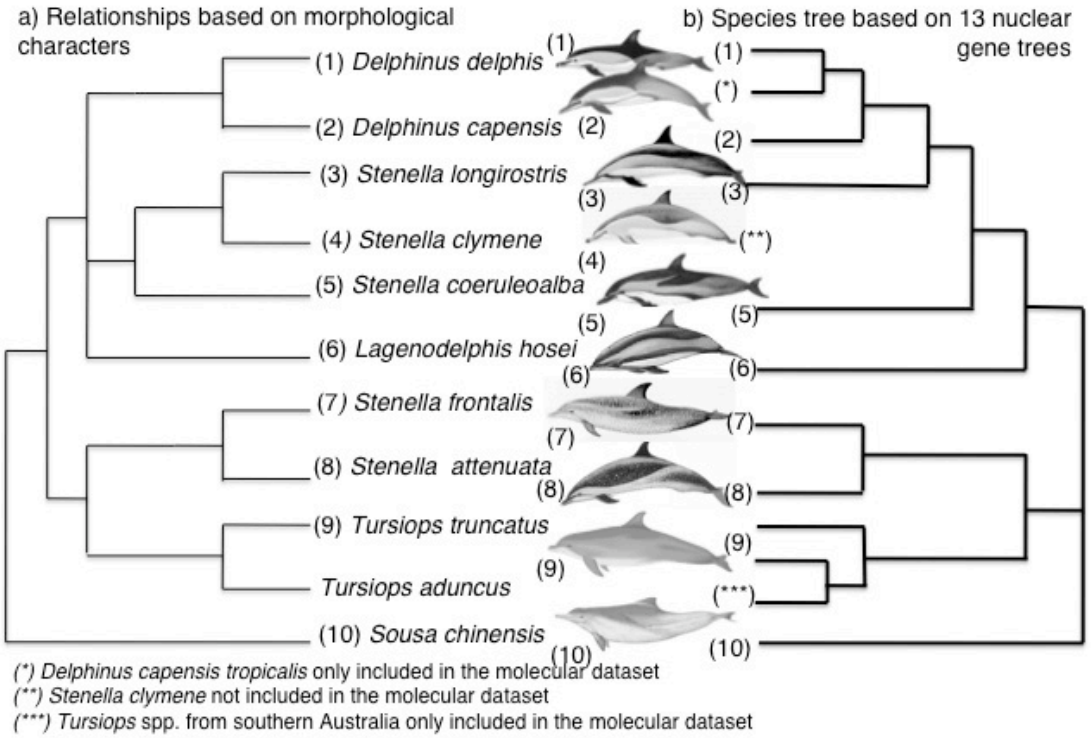


Figure 7.1. Species trees based on a) recent morphological analyses (Perrin *et al.* 1987) Perrin 2009; and b) on 13 nuclear gene trees (Amaral *et al*. in review).

Although the monophyly of the genus *Tursiops* had been previously reported (McGowen *et al.* 2009) as well as the sister taxa relationship between the spotted dolphins *Stenella attenuata* and *S. frontalis* (Kingston *et al.* 2009), these two groups had never been clustered together in a phylogeny before. However, morphologically, they are very similar (Perrin *et al.* 1987). Some of the recovered phylogenetic relationships, although in agreement with morphological characters, still do not resolve the confusing taxonomy. A clear example is the paraphyly of the genus *Stenella*, which supports previous suggestions that the genus is in fact an artificial assemblage of species and that its taxonomy needs revision.

Incongruence between mtDNA and nuDNA delphininae phylogenies was found. Similar incongruence has also been found in other animal groups (McCracken & Sorenson 2005; Peters *et al.* 2007) and is likely caused by incomplete lineage sorting or by hybridization. It is however difficult to tell these two processes apart.

The multiple, independent gene trees obtained in this study suggest that a rapid series of divergences, characterized by short internodes, occurred during the early stages of the Delphininae evolution, which suggests that incomplete lineage sorting may affect phylogenetic inference for this group. The occurrence of hybridization is also possible. It has been described to occur both in the wild and in captivity (Bérubé 2002). Moreover, as the study of reproductive proteins suggested (Chapter III), it may be possible that the molecular mechanisms for species recognition in dolphins is relaxed. Although the comparison of several species tree methods have not completely elucidated the species history of the Delphininae, this approach showed how the use of multiple nuclear loci is likely to result in a more realistic depiction of lineage history than the use of one or a few loci, particularly when analysed under a coalescent-based species tree approach.

Common dolphins within the genus *Delphinus* are widely distributed and present a great morphological variability throughout their distribution, which has caused some uncertainty in their taxonomy. In addition, several aspects of their phylogeography, population structure and evolutionary history have remained unanswered. Using sequences from the mitochondrial cytochrome *b* gene and from five nuclear DNA loci, the investigation in Chapter V showed that temperature fluctuations caused by the Pleistocene climatic oscillations and consequent changes in upwelling intensities and availability of resources may have had an influence in the demography, dispersal and speciation of *Delphinus*. A scenario for the origin of the short-beaked morphotype in the Pacific Ocean basin during the Pleistocene and consequent dispersal into the Atlantic through the Indian Ocean was proposed. This biogeographic model has been suggested to explain dispersal in the genus *Stenella* as well as in several other marine organisms, such as sea turtles, sharks, sea birds and teleosts. The origin of the long-beaked morphotype was suggested to be associated with the exploration of new available coastal habitats during mid-Pleistocene. During this period, a decrease in upwelling intensity and consequent decrease in availability of resources could have caused some short-beaked common dolphin populations to

explore new habitats and originate the long-beaked morphotype. Later, independent events in the Atlantic Ocean likely originated the long-beaked morphotype occurring in that ocean basin.

Using 14 microsatellite loci, analyses undertaken in Chapter VI showed a strong pattern of population genetic structure across short-beaked common dolphin populations inhabiting the Pacific, Indian and Atlantic Ocean basins. While this pattern seemed to be explained by a process of isolation by distance, the divergence found within each ocean basin is more likely explained by oceanographic features such as marine productivity and sea surface temperature. Prey behaviour and feeding specializations are also possible explanations for the associations found between environmental variables and population structure (Chapter VI). Furthermore, this study highlights the importance of using a seascape genetics approach to infer distribution and connectivity of top marine predators and its use for designing marine protected areas.

Genetic differentiation in short-beaked common dolphin populations has been shown to be higher across larger geographical scales and lower within each ocean basin (Natoli *et al.* 2006; Chapters V and VI). Higher levels of divergence have been reported for populations inhabiting the Pacific Ocean when compared to those inhabiting the Atlantic Ocean (Amaral *et al.* 2007; Bilgmann *et al.* 2008; Mirimin *et al.* 2009; Möller *et al.* 2011). This pattern was not entirely evident in the dataset used for this study (due to the sample size used that was not intended for fine scale analyses), but became apparent in comparisons with other studies. One hypothesis accounting for this pattern would be that populations inhabiting the Pacific Ocean are older, as suggested by results obtained in Chapter V, and would therefore have had more time to diverge. Alternatively, it may be that oceanographic conditions in the Pacific Ocean have created more barriers to dispersal and or to local adaptation. Paleoceanographic data indicate that the Pacific Ocean basin was more stable to temperature fluctuations than the North Atlantic basin due to its larger size and due to the fact that glaciations were more severe in the North Atlantic (Briggs 1974). This made the Pacific Ocean a much

richer ecosystem, which likely explains the highest levels of genetic diversity found in populations of short-beaked common dolphins from the Pacific Ocean. This may also provide an alternative explanation for the existence of finer scale population structure. A comparison with other marine organisms has revealed a similar pattern of higher genetic diversity in the Pacific Ocean populations when compared with Atlantic Ocean populations (e.g. Duncan *et al.* 2006; Vinas *et al.* 2004).

Until now, few genetic studies have confirmed the taxonomic status of long-beaked populations distributed worldwide. In this study (Chapter V), we provide evidence that long-beaked common dolphin populations are not a single, globally distributed species, but most likely the result of independent local adaptations, thus supporting previous findings made based only on two long-beaked populations (Natoli *et al.* 2006). The Northeast Pacific population is genetically well differentiated from all other common dolphin populations and therefore likely constitutes a different species. Consequently, its current classification as *Delphinus capensis* would be incorrect, since this is the designation of the holotype specimen for the South African long-beaked common dolphin. Accordingly, it should be changed to *Delphinus bardii* Dall, 1873 as initially proposed (Heyning & Perrin 1994). Regarding the Southeast (off South Africa) and Southwest (off Brazil) Atlantic long-beaked populations, these still shared polymorphisms with the short-beaked populations from the Atlantic Ocean, suggesting that they are still in the process of speciation. A separate species status would probably not be appropriate for these populations. However, since there is evidence from multiple molecular markers and morphological characters that they represent a different evolutionary lineage, they should be considered different subspecies as suggested by Perrin & Reeves 2004 and Perrin *et al*. 2010.

The *tropicalis* population inhabiting the Western Indian Ocean showed high levels of differentiation based on fixation indices but not in genealogical lineages, either in the mitochondrial or the nuclear haplotype networks (Chapter V). Although its taxonomic status was revised based on morphology (Jefferson & Van Waerebeek

2002), its phylogenetic position remains uncertain, warranting further investigation.

7.2. Final Remarks

Through a multi-locus, multi-disciplinary approach, insights were gained into the evolution of delphinid species and populations, and ultimately into some of the factors driving their evolution. The set of nuclear molecular markers used in this study, as well as the several methodologies used, have significantly advanced our understanding of the phylogenetic relationships within the Delphininae, and of the demography and population structure of *Delphinus*.

The species tree obtained for the subfamily Delphininae revealed how the inference of molecular phylogenies can be complex in cases of rapid species radiations, where processes like incomplete lineage sorting and hybridization may affect the evolutionary history. Using methods that account for gene tree heterogeneity, a nuclear species tree in agreement with morphology-based relationships was obtained. This highlights the importance of using such methods in obtaining a better estimate of the evolutionary relationships of a group when the existence of incomplete lineage sorting results in incongruent gene trees. Nonetheless, the incongruent patterns obtained between mitochondrial and nuclear phylogenies likely reflect not only the existence of incomplete lineage sorting but also hybridization. It is possible that the molecular mechanism of species recognition in cetaceans is relaxed. The slow rate of evolution detected in reproductive proteins across several cetacean species support this. Although this mechanism may also be behavioural, a relaxation of the species recognition process could certainly explain the hybridization cases that have been reported and such incongruent patterns between mitochondrial and nuclear DNA phylogenies.

The demography, dispersal and speciation of the genus *Delphinus* were shown to have been likely influenced by Pleistocene climatic oscillations. The use of several molecular markers and coalescent-based methods allowed the inference of an origin and route of dispersal for these species, shedding light into their

evolutionary history. The phylogeographic patterns obtained for *Delphinus* agree with the patterns described for other highly mobile marine organisms, suggesting that biogeographic models are similar across several marine taxa. These findings highlight the potential role of ongoing climate change on the distribution and abundance of top marine predators. Environmental features seem to have not only influenced the origin, range expansion and speciation of *Delphinus* during the Pleistocene period, but appear to be playing a role in driving and maintaining population divergence in short-beaked common dolphin populations in more recent times. Although marine productivity and sea surface temperature have long been recognized to influence habitat occupancy and dispersal in cetaceans (Forney 2000; Hamazaki 2002), a direct influence of such oceanographic features on the evolutionary history and population structure of a globally distributed cetacean species had never been reported before.

7.3. FUTURE RESEARCH

In all chapters of this dissertation, some questions remained unanswered. Future research should focus on:

- Studying other reproductive proteins, mainly those that have been reported to be directly linked to sperm competition in primates, the protamines, and extend the number of cetacean species sequenced. Right whales would be extremely interesting species to study due to the high levels of promiscuity that have been reported for these species;
- Including *Stenella clymene*, *Sousa teuszii* and all *Tursiops* spp. morphotypes in a new Delphininae species tree estimate and using methods at the interface of phylogenetic and population processes;
- Broadening the sampling of common dolphin populations by including samples from geographical regions that were not available to this study (e.g., West Africa, Peru, Japan/Taiwan);
- Obtaining genome-wide molecular markers to expand inferences about population demography and evolutionary history of common dolphins and identify genomic regions with a putative adaptive role. This would allow a deeper

understanding of the processes leading to population divergence and local adaptation in these and other delphinid species.

7.4. REFERENCES

Amaral AR, Sequeira M, Cedeira-Martínez J, Coelho MM (2007) New insights on population genetic structure of *Delphinus delphis* from the northeast Atlantic and phylogenetic relationships within the genus inferred from two mitochondrial markers. *Marine Biology* **151**, 1967-1976.

Bérubé M (2002) Hybridism. In: *Encyclopedia of Marine Mammals* (eds. Perrin WF, Wursig B, Theweissen JGM), pp. 596-600. Academic Press Inc., San Diego, California.

Bilgmann K, Möller LM, Harcourt RG, Gales R, Beheregaray LB (2008) Common dolphins subject to fisheries impacts in Southern Australia are genetically differentiated: Implications for conservation. *Animal Conservation* **11**, 518-528.

Briggs JC (1974) *Marine Zoogeography*. McGraw-Hill, USA.

Caballero S, Jackson J, Mignucci-Giannoni AA, Barrios-Garrido H, Beltran-Pedreros S, Montiel-Villalobos MG, Robertson KM, Baker CS (2008) Molecular systematics of South American dolphins *Sotalia*: Sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. *Molecular Phylogenetics and Evolution* **46**, 252-268.

Calkins JD, El-Hinn D, Swanson WJ (2007) Adaptive evolution in an avian reproductive protein: ZP3. *Journal of Molecular Evolution* **65**, 555-563.

Carling MD, Brumfield RT (2007) Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). *Plos One* **2**, e160.

Connor RC, Read AJ, Wranhgam R (2000) Male Reproductive Strategies and Social Bonds. In: *Cetacean Societies: Field Studies of Dolphin and Whales* (eds. Mann J, Connor RC, Tyack PL, Whitehead H). The Chicago University Press. Chicago.

Duncan KM, Martin AP, Bowen BW, De Couet HG (2006) Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). *Molecular Ecology* **15**,

2239-2251.

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* **23**, 691-700.

Flower WH (1883) On the characters and divisions of the family *Delphinidae*. *Proceedings of the Zoological Society of London*, 466-513.

Forney KA (2000) Environmental models of cetacean abundance: Reducing uncertainty in population trends. *Conservation Biology* **14**, 1271-1286.

Hamazaki T (2002) Spatiotemporal prediction models of cetacean habitats in the mid-western North Atlantic Ocean (from Cape Hatteras, North Carolina, USA to Nova Scotia, Canada). *Marine Mammal Science* **18**, 920-939.

Heyning JE, Perrin WF (1994) Evidence for two species of common dolphins (Genus *Delphinus*) from the eastern north Pacific. *Contributions in Science* **442**, 1-35.

Jackson JA, Baker CS, Vant M*,* Steel DJ, Medrano-Gonzalez L, Palumbi SR (2009) Big and slow: Phylogenetic estimates of molecular evolution in baleen whales (Suborder Mysticeti). *Molecular Biology and Evolution* **26**, 2427-2440.

Jefferson TA, Van Waerebeek K (2002) The taxonomic status of the nominal dolphin species *Delphinus tropicalis* van bree, 1971. *Marine Mammal Science* **18**, 787-818.

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.

Kingston SE, Rosel PE (2004) Genetic differentiation among recently diverged delphinid taxa determined using AFLP markers. *Journal of Heredity* **95**, 1-10.

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.

Lee JY, Edwards SV (2008) Divergence across Australia's Carpentarian Barrier: Statistical phylogeography of the red-backed fairy wren (*Malurus Melanocephalus*). *Evolution* **62**, 3117-3134.

McCracken KG, Sorenson MD (2005) Is homoplasy or lineage sorting the source of incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (*Nomonyx-Oxyura*)? *Systematic Biology* **54**, 35-55.

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

Metz EC, Gomez-Gutierrez G, Vacquier VD (1998) Mitochondrial DNA and bindin gene sequence evolution among allopatric species of the sea urchin genus *Arbacia*. *Molecular Biology and Evolution* **15**, 185-195.

Mirimin L, Westgate A, Rogan E, Rosel P, Read A, Coughlan J, Cross T (2009) Population structure of short-beaked common dolphins (*Delphinus delphis*) in the North Atlantic Ocean as revealed by mitochondrial and nuclear genetic markers. *Marine Biology* **156**, 821-834.

Möller LM, Valdez FP, Allen S, Bilgmann K, Corrigan S, Beheregaray LB (2011) Fine-scale genetic structure in short-beaked common dolphins (*Delphinus delphis*) along the East Australian Current. *Marine Biology* **158**, 113-126.

Natoli A, Cañadas A, Peddemors VM, Aguila A, Vaquero C, Fernadez-Piqueras P, Hoelzel AR (2006) Phylogeography and alpha taxonomy of the common dolphin (*Delphinus* sp.). *Journal of Evolutionary Biology* **19**, 943-954.

Perrin WF, Mitchell ED, Mead JG, Caldwell DK, Caldwell MC, Vanbree PJH, Dawbin WH (1987) Revision of the spotted dolphins, *Stenella* spp. *Marine Mammal Science* **3**, 99-170.

Perrin WF, Reeves RR (2004) Appendix 5: Report of the Working Group on Species- and Subspecies-level Taxonomy. In: *Report of the Workshop on Shortcomings of Cetacean Taxonomy in Relation to Needs of Conservation and Management, April 30-May 2* (eds. Reeves RR, Perrin WF, Taylor BL, Baker CS, Mesnick SL), pp. 25-60. NOAA technical memorandum, La Jolla, California.

Perrin WF, Mead JG, Brownell Jr, RL (2010) Review of the Evidence Used in the Description of Currently Recognized Cetacean Subspecies. Technical Memorandum NOAA-TM-NMFS-SWFSC-450.

Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). *Evolution* **61**, 1992-2006.

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**, 137-144.

Turner LM, Hoekstra HE (2006) Adaptive evolution of fertilization proteins within a genus: Variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Molecular Biology and Evolution* **23**, 1656-1669.

Turner LM, Hoekstra HE (2008) Causes and consequences of the evolution of reproductive proteins. *International Journal of Developmental Biology* **52**, 769-780.

Vinas J, Bremer JRA, Pla C (2004) Inter-oceanic genetic differentiation among albacore (*Thunnus alalunga*) populations. *Marine Biology* **145**, 225-232.

Wu CI, Ting CT (2004) Genes and speciation. *Nature Reviews Genetics* **5**, 114-122.