Genomics of population differentiation in humpback dolphins, *Sousa* spp. in the Indo-Pacific Ocean

Ana R. Amaral[*1,2¶] Cátia Chanfana[*2] Brian D. Smith[3], Rubaiyat Mansur[3], Tim Collins[3], Robert Baldwin[4], Gianna Minton[5], Guido J. Parra[6], Michael Krützen[7], Thomas A. Jefferson[8], Leszek Karczmarski[9,10], Almeida Guissamulo[11], Robert L. Brownell Jr.[12], Howard C. Rosenbaum[3,1]

[1] *Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th Street and Central Park West, New York, NY 10024, United States of America.*
[2] *Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal*
[3] *Wildlife Conservation Society, Ocean Giants Program, 2300 Southern Boulevard, Bronx, New York 10460, United States of America*
[4] *Five Oceans Environmental Services, P.O. Box 660, Postal Code 131, Sultanate of Oman.*
[5] *Megaptera Marine Conservation, The Hague, The Netherlands*
[6] *Cetacean Ecology, Behaviour and Evolution Lab, College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia*
[7] *Evolutionary Genetics Group, Department of Anthropology, University of Zurich, Winterthurerstr. 190, CH 8057 Zurich, Switzerland, ORCID-ID: 0000-0003-1055-5299*
[8] *Clymene Enterprises, 13037 Yerba Valley Way, Lakeside, CA 92040 USA*
[9] *Cetacean Ecology Lab, Cetacea Research Institute, Lantau, Hong Kong*
[10] *Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Hatfield, Pretoria, South Africa*
[11] *Universidade Eduardo Mondlane, Museu de Historia Natural, 104, Praca Travessia do Zambeze. Maputo. Mozambique*
[12] *NOAA Fisheries, Southwest Fisheries Science Center, 8901 La Jolla Shores Drive, La Jolla, CA, 92037, USA*

[*]authors contributed equally

¶ Corresponding author

Address: Faculdade de Ciências da Universidade de Lisboa, Departamento de Biologia Animal, Edifício C2, Campo Grande, 1749-016, Lisbon, Portugal
Telephone number: +351217500000 ext.22312
E-mail: aramaral@fc.ul.pt

Abstract

Speciation is a fundamental process in evolution and crucial to the formation of biodiversity. It is a continuous and complex process, which can involve multiple interacting barriers leading to heterogeneous genomic landscapes with various peaks of divergence among populations. In this study, we used a population genomics approach to gain insights on the speciation process and to understand the population structure within the genus *Sousa* across its distribution in the Indo-Pacifc region. We found 5 distinct clusters, corresponding to *S. plumbea* along the eastern African coast and the Arabian Sea, the Bangladesh population, *S. chinensis* off Thailand and *S. sahulensis* off Australian waters. We suggest that the high level of differentiation found, even across geographically close areas, is likely determined by different oceanographic features such as sea surface temperature and primary productivity.

Keywords: Speciation, Marine Mammals, Delphinids, Genotyping-by-sequencing

Introduction

Understanding drivers of population divergence and speciation is a central question in evolutionary biology. This is especially true in the marine environment where barriers to dispersal are not as obvious as in the terrestrial environment. A central paradigm in marine systems is that populations are typically characterized by weak genetic differentiation due to the potential for long-distance dispersal favouring high levels of gene flow (Palumbi, 1992). However, several studies have shown that marine megafauna show high levels of genetic differentiation (e.g. Hess et al, 2013), as is the case for inshore populations (e.g. Tezanos-Pinto et al, 2009). There are neutral and adaptive processes that

63 can lead to higher than expected differentiation in the marine environment. Neutral
64 processes include population dynamics caused by birth, death and dispersal of organisms
65 through different regions and environments, causing genetic drift. Adaptive processes
66 include local adaptation, where organisms have higher average fitness in their local
67 environment when compared to individuals elsewhere.
68 Cetaceans are a unique taxonomic group in that species underwent drastic evolutionary
69 transitions from terrestrial to marine environments (Steeman et al, 2009). Delphinids, in
70 particular, have radiated very recently (at around 10-12 Ma, McGowen et al. 2009) and
71 have populated many different habitats and environments, providing a unique opportunity
72 to study the role of different evolutionary processes in shaping population structure and
73 genetic diversity at large spatial, but relatively short temporal scales.
74 Several factors and mechanisms have been suggested as likely to influence and drive
75 genetic differentiation and speciation in cetacean species. Despite being marine predators
76 with high mobility and few obvious barriers to dispersal, environmental factors like sea
77 surface temperature, salinity and ocean currents have been shown to influence patterns of
78 population structure, as these dictate prey dispersal and availability (e.g. Amaral et al,
79 2012; Mendez et al, 2011). Other mechanisms such as social interactions, behaviour and
80 culture, have also been suggested to shape population structure and genetic diversity
81 (Alexander et al, 2016; Carroll et al, 2015; Kopps et al, 2014; Riesch et al, 2012).
82 Humpback dolphins (*Sousa* spp.) are distributed discontinuously in coastal waters of
83 West Africa and in the Indian and Western Pacific Oceans and all populations are
84 currently facing anthropogenic pressures, raising conservation concerns (Braulik et al,
85 2015; Jefferson and Smith, 2016b; Parra and Cagnazzi, 2016). This genus comprises four
86 species: *S. teuszii* in the Eastern Atlantic Ocean along the west African coast, *S. plumbea*

87 in the Western Indian Ocean, *S. chinensis* distributed in the Eastern Indian and Western
88 Pacific Oceans and *S. sahulensis* in Northern Australia and New Guinea (Jefferson and
89 Rosenbaum, 2014) (Figure 1). However, the exact eastern limit of *S. plumbea* in the Bay
90 of Bengal and the western limit of *S. chinensis* are poorly known. In terms of external
91 appearance,; *S. plumbea* has a darker coloration with little spotting and a prominent dorsal
92 fin hump; *S. teuszii* has a similar appearance to that of *S. plumbea* but with significantly
93 shorter rostra and lower tooth counts; *S. chinensis* has light adult coloration, often with
94 bluish gray spotting and lacks the prominent dorsal hump; *S. sahulensis* has no visible
95 dorsal fin hump and the dorsal fin is low and triangular, with adults having a dark grey to
96 grey back and a lighter belly (Jefferson and Rosenbaum 2014). Analyses conducted to
97 date suggest high levels of population genetic structure within both *S. plumbea* and *S.*
98 *chinensis* and a highly differentiated population in the Bay of Bengal (Amaral et al, 2017;
99 Mendez et al, 2013). Oceanographic features such as sea surface temperature and primary
100 productivity have been suggested as important drivers of population differentiation in
101 these animals (Amaral et al., 2017; Mendez et al., 2011).
102 The Bay of Bengal is a marine region in the Northern Indian Ocean that supports an
103 impressive variety of cetaceans, but with little knowledge on the evolutionary processes
104 acting on those species (Mansur et al, 2012; Smith et al, 2008). The extreme infusion and
105 redistributive dynamism of biological productivity in this region is a rare ecological
106 condition that supports cetaceans in numbers generally much larger than other
107 populations in the region (Mansur et al., 2012). While little is known about the
108 morphological differences in the highly-differentiated humpback dolphin population
109 occurring in this region, it has been hypothesized that the relatively rare environmental
110 conditions in the Bay of Bengal explains its genetic distinctiveness (Amaral et al,. 2017).

111 Other marine species occurring in this area have also shown high levels of genetic
112 differentiation (e.g.Li et al, 2015).
113 In this study we aim to build on our previous work that used mtDNA and three nuclear
114 markers to investigate genetic connectedness of Indo-Pacific humpback dolphin
115 populations. Using a population genomics approach, we aim to investigate patterns of
116 genome wide differentiation in Indo-Pacific humpback dolphins across the Indian and
117 West Pacific Oceans.
118

119 **Material and Methods**

120

121 **Sample collection and Sequencing**

122 Our total data set consisted of 30 samples obtained from stranded or biopsied humpback
123 dolphins, which were selected from a set of samples already used in previous studies
124 (Mendez et al., 2013, Amaral et al., 2017). Representing the entire distribution range of
125 the *Sousa* genus in the Indo-pacific region, our data set contains samples from Southeast
126 Africa (SEA - South Africa and Mozambique n=6), Arabian Sea (OM - Oman, n=8), Bay
127 of Bengal (BAN - Bangladesh, n=10), Indo-China (CHI - Thailand, Hong Kong and
128 Taiwan, n=4) and Northern Australia (AUS, n=2) (Figure 1).
129 The genomic DNA from tissues samples already preserved in ethanol (96% v/v) or in
130 sodium chloride-saturated 20% dimethyl sulphoxide (DMSO) solution, was extracted
131 using QIAamp Tissue Kit (QIAGEN, Valencia, CA, USA) and its concentration
132 measured using a Qubit Florometric Quantitation (ThermoFisher). The samples were then
133 shipped to the Cornell University Institute of Biotechnology's Genomic Diversity Facility
134 (http://www.biotech.cornell.edu/brc/genomic-diversity-facility) where the GBS

(genotyping-by-sequencing) data was generated. Sequencing libraries were constructed using the restriction enzyme *Pst*I (CTGCAG) by a genotype-by-sequencing protocol (Elshire et al, 2011). Unique oligonucleotide barcodes were added to each sample for multiplexed sequencing on an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA). Template-controls were included with the batch of samples. Single-end reads were generated with an average length of approximately 100 bp.

Data processing

Demultiplexing, initial quality control, assembly, and SNP discovery were completed in the TASSEL pipeline v3.0.174 (Glaubitz et al, 2014), which was specifically designed for GBS datasets. The killer whale genome was used as a reference to identify single nucleotide polymorphisms (SNPs) (*O. orca*, Oorc_1.1, 200.0x coverage, (Foote et al, 2015; Morin et al, 2010) using bwa (v0.7.8-r455; Li and Durbin, 2009).

The TASSEL pipeline relies on the number of times a given tag has been observed as an indicator of sequence quality, and not quality scores, as these are frequently not indicative of sequence quality in short reads as those obtained in a GBS approach (Dohm et al, 2008; Eren et al, 2013; Glaubitz et al, 2014). The first step of the pipeline consists in processing and collapsing all barcoded reads into a set of unique sequence tags, with one TagCounts file produced per input FASTQ file. These separate files are then merged into a single master file and the tag list is aligned to the reference genome. The barcode information in the original FASTQ files is used to infer the number of times each tag in the master file is observed in each sample and these counts are stored in a different TagsByTaxa file. This information is then used to discover SNPs at each set of tags with the same genomic position and filter the SNPs based upon the proportion of taxa covered, minor allele

159 frequency (MAF = 0.1), linkage disequilibrium (minimum median population LD(R^2)
160 was set to 0.1) and inbreeding coefficient (F = 1-Ho/He, where Ho- observed
161 heterozygosity and He – expected heterozygosity) (Glaubitz et al, 2014).
162 After the SNP calling obtained with the TASSEL pipeline, blank-controls and 3
163 individuals were excluded due to missing data, producing a final data set of 27 individuals
164 (Table S1). For these individuals, we applied additional filters to further reduce false
165 positive SNPs for subsequent analysis. Firstly, limits for the genomic depth of coverage
166 were calculated and applied for each individual in RStudio (v1.0.136; RStudio Team
167 (2016); R Core Team (2016)) using a custom script (V. Sousa). The calculation
168 corresponded to 1/3 of the mean-depth for the minimum limit and the double of the mean-
169 depth for the maximum limit. This calculation was applied because it considers the
170 average coverage of each individual. Secondly, to minimize the genotyping error that
171 could come from a heterozygosity excess, we performed the Hardy-Weinberg
172 Equilibrium test using the hardy option in VCFtools v0.1.15 (Danecek et al, 2011). The
173 sites with *P*-values significant at the 0.01 level were excluded. Non bi-allelic sites as well
174 as sites with missing data higher than 50% were also removed using VCFtools.
175 A filter for Minimum Allele Frequency (hereafter MAF) was also applied to the raw data
176 as the initial filter of MAF = 0.1 applied in TASSEL seemed very conservative. We used
177 two different values of MAF (2 and 5%) to understand how this choice would affect
178 subsequent analyses, since rare variants could be false positives of the sequencing
179 protocol but could also be important genetic variation that can have true genetic effects
180 in the population (Nielsen et al, 2012; Whitlock and Lotterhos, 2015).
181 We used two different datasets in all the population structure analyses described in the
182 next section. After this step, each data set was converted to various formats using

PGDSpider (v2.1.1.3; Lischer and Excoffier, 2012) for subsequent analyses. The application of the MAF filter greatly reduced the number of SNPs to analyze, but had no effect on the patterns obtained, therefore we chose to use the dataset with the highr number of SNPs (19 462) to generate the results presented in Figures 2-5.

In order to measure the genetic differentiation between populations, we used the snpgdsFst function in the SNPRelate package (Zheng *et al*, 2012). The estimator of (Wright, 1951) F_{ST} (hereafter F_{ST}) was calculated following the approach of (Weir and Cockerham, 1984). We only compared populations with sample sizes higher than 5, Bangladesh, Arabian Sea and East coast of Africa, and results are just preliminary.

Population structure

To infer population structure in the genus *Sousa*, we first used a discriminant analysis of principal components (DAPC) to identify genetic clusters. DAPC is a multivariate approach that transforms individual genotypes using principal components analysis (PCA) prior to a discriminant analysis (DA) (Jombart *et al*, 2010). This maximizes the differentiation between groups while minimizing variation within groups and was conducted using the dapc function in the *Adegenet* package (v2.1.1; Jombart et al, 2008). Since DAPC requires group assignment *a priori*, we employed a K-means clustering algorithm implemented in *Adegenet* to identify the optimal number of clusters from K = 1 to K = 10. Different clustering solutions were then compared using Bayesian Information Criterion (BIC), and to avoid over-fitting of discriminant functions, we used Alpha-score optimization to evaluate the optimal number of principle components (PCs) to retain in the analysis, as described in Jombart et al, (2010).

206 Second, we estimated individual genetic ancestry using sNMF (Frichot et al, 2014)
207 through snmf function in the *LEA* package (v1.6.0; Frichot and François 2015), and the
208 program STRUCTURE (v2.3.2) (Pritchard et al, 2000). Both programs compute
209 proportion quantities called ancestry coefficients that represent the proportion of an
210 individual genome that originate from multiple ancestral gene pools (Pritchard et al.,
211 2000; Frichot et al., 2014). While sNMF generates comparable results to those obtained
212 from STRUCTURE, it does not require Hardy-Weinberg equilibrium assumptions
213 (Frichot et al., 2014).
214 The ancestry coefficients were estimated from a specified number of ancestral
215 populations (K). For sNMF, the ancestry coefficient was calculated for K 1 to 10 using
216 100 replicates for each K. The preferred number of K was chosen using a cross-entropy
217 criterion based on the prediction of masked genotypes to evaluate the error of ancestry
218 estimation. For STRUCTURE, a correlated allele frequency model with no admixture
219 was used (Hubisz et al, 2009). We conducted 20 runs for each K value (1-6) with a burn-
220 in of 10,000 repetitions for each value of K followed by 100,000 MCMC repetitions. To
221 determine the best value of K we employed two approaches. We used an iterative
222 approach based on the ΔK statistic (Evanno et al, 2005) and also used the ln(Pr(X|K)
223 values in order to identify the K for which Pr(K=k) is highest, as described in Pritchard
224 et al. 2000. Both approaches were conducted using CLUMPAK (Kopelman et al, 2015)
225 and STRUCTURE HARVESTER (v0.6.94; Earl and vonHoldt, 2012).
226 A maximum-likelihood framework was also applied to infer phylogenetic relationships
227 between populations. The analysis was implemented using RAxML (v8.2.11;
228 Stamatakis, 2014) in which we carried out 1,000 inferences using the GTR model with
229 no rate heterogeneity modelled (ASC_GTRCAT). The branch support was estimated

using bootstrap by a majority-rule criteria as implemented in RAxML and visualized simultaneously in a single consensus tree (Holland et al, 2005) in Figtree (v1.4.3; Rambaut 2016). The consensus tree was set at 0.1, which means that bipartitions that appeared in at least 200 of the 2,000 bootstrap trees participated in network construction. RAxML was run using the two data sets (Table 1).

Results

We generated genome-wide SNPs for 30 individuals, 3 of which were excluded due to high levels of missing data (higher than 90% of missing SNPs - CHI12,14, 13), producing a final data set of 27 individuals (Table S1) that were used for the downstream analysis: Southeast Africa (SEA - South Africa and Mozambique n=6), Arabian Sea (OM - Oman, n=8), Bay of Bengal (BAN - Bangladesh, n=10), Thailand n=1 and Northern Australia (AUS, n=2) After the TASSEL pipeline, 55615 SNPs were obtained, and this number was reduced to a range of 11591 – 19 462 SNPs, depending on the value of MAF used.

Population structure and differentiation

No differences were found in the initial exploratory analyses using the two datasets obtained using different filters. All the results presented below correspond to the results obtained with 19 462 SNPs.

The clustering analysis performed in STRUCTURE resulted in the best value of K=3 if we consider the Evanno method and of K=4 if we consider the highest value of ln(Pr(X|K) (Table 2). The results obtained using sNMF showed K=4 as the best fitting number of clusters (Figure 2). The overall pattern obtained in these two clustering methods corresponds to the separation of the three species, *S. sahulensis*, *S. plumbea* and *S.*

chinensis (Figures 2, 3, 4 and 5) and a fourth cluster including the subdivision within *S. plumbea* separating the populations from the African coast and the Arabian Sea. The *S. chinensis* population of Bangladesh is clearly separated from all other populations. In addition, both STRUCTURE and SNMF analyses showed the individual from Thailand as an individual with a mixed ancestry from Bangladesh, Oman, East African coast and Australia (Figure 2). The DAPC results show five clearly separated clusters (Figure 3). For this analysis, 5 PCs were retained as indicated by the a-score (Figure S1) and the best-fitting value of K was chosen according to the BIC plot (Figure S2).

The preliminary F_{ST} analysis show results consistent with those described above, with high levels of genetic differentiation found between the Bangladesh population and the Arabian Sea and the African coast populations. The lowest value of differentiation is seen between the Arabian Sea and the African coast populations (Table 1).

Phylogenetic relationships

Using the ML method, the phylogenetic tree showed the same pattern mentioned above. Three main and highly supported clusters, corresponding to the three described species are seen. The subdivision within *S. plumbea* is also identified and supported with bootstrap values of 100 (Figure 5). The individuals from Bangladesh and the individual from Thailand are also found in separate highly supported clusters.

Discussion

In this study, we conducted for the first time a genome-wide population analysis of humpback dolphins occurring in the Indo-Pacific Ocean. We found high levels of species and within-species divergence consistent with previous studies using mitochondrial DNA

278 and five nuclear loci, that support the currently recognized species of *Sousa* as well as
279 strong genetic subdivisions within species.
280
281 **Population structure and environmental drivers**
282 Our study supports previous findings that humpback dolphins in the Indo-Pacific region
283 appear to be divided in five main genetic clusters. The three species already described (*S.*
284 *plumbea*, *S. chinensis* and *S. sahulensis*) and the Bangladesh population are strongly
285 differentiated. Within *S. plumbea*, we further obtained a genetic division, albeit weaker,
286 between the African coast and the Arabian Sea. The Bangladesh population in the Bay of
287 Bengal seems to be genetically more similar to *S. sahulensis* and *S. chinensis*, even though
288 the dolphin's outer body morphology is similar to the other species, *S. plumbea*. Since
289 this population is located in a transition region between *S. plumbea* and *S. chinensis* and
290 shows morphological characters of both species, hybridization between the two types was
291 hypothesized (Jefferson and Rosenbaum, 2014; Mendez et al., 2013). However, both
292 previously obtained mitochondrial DNA data and the genomic DNA obtained in this
293 study show congruent results, ruling out the hybridization scenario (Amaral et al., 2017).
294 Based on our previous results with the mithochondrial DNA and those obtained in this
295 study, we suggest that this population may constitute a separate taxonomic entity, but
296 additional evidence with samples from surrounding areas is needed. This region seems to
297 harbour a strong potential for endemism and speciation, as seen in the high levels of
298 genetic differentiation obtained for a sympatric dolphin species, the Indo-Pacific
299 bottlenose dolphin (Amaral et al., 2017), as was well as other mobile marine species (e.g.
300 Li et al., 2015). The northern Bay of Bengal is located in an ecological "cul-de-sac" and
301 has extraordinary oceanographic conditions, including intrusion of massive and dynamic

freshwater and sediment flow from among the world's largest river systems, leaf litter and other bio-productivity from a large mangrove forest. In addition, this region has an upwelling from a deep submarine canyon which supports a large sediment fan and a seasonally reversing current gryre with associated meso-eddies that retain and redistribute nutrients (Cheng et al, 2013; Hussain and Acharya, 1994). Together these local conditions are unique in terms of their dynamics and scale, and likely explain the genetic distinctiveness found in marine organisms occurring in the northern Bay of Bengal.

The sample from Thailand showed a mixed ancestry with genetic contributions from *S. sahulensis* and the Bangladesh population, and on a much lower level with *S. plumbea.* This suggests that it could be a hybrid individual and more samples are needed to understand the level of genetic distinctiveness of individuals occurring in this region.

The genetic division obtained with *S. plumbea*, separating the southeast South Africa population from the Arabian Sea population, has already been described using mtDNA and a few nuclear markers (Mendez et al., 2013; Amaral et al., 2017). Both these regions are characterized by unique oceanographic features that could explain this pattern. The coast of Oman is part of the Arabian Sea Upwelling Province, where the annual monsoon influences the system of currents and the occurrence of rich upwelling areas (Longhurst, 2006). The coasts of Mozambique and South Africa are part of the Eastern African Coastal Province, which includes the Mozambique Channel, and is also influenced by a series of gyres and currents, creating unique environmental conditions. Surface currents and other oceanographic variables such as water turbidity and chlorophyl concentration are known to influence and drive distribution patterns in mobile marine species, such as turtles (e.g. Bass et al, 2006), common dolphins (Amaral et al*.,* 2012) and franciscana

dolphins (Mendez et al, 2010) and could therefore also determine the patterns of genetic differentiation seen in humpback dolphins.

The overall phylogeographic pattern obtained in this study, with distinct lineages in the east and west of the Indo-Pacific Ocean has also been described in other marine species (Ahti et al, 2016; Bowen et al, 2016; Farhadi et al, 2017; Li et al, 2015). This pattern may have resulted from restricted connectivity of populations across the Sunda shelf (southeast extension of the continental shelf of Southeast Asia comprising the Malay Peninsula, Sumatra, Borneo, Java and Bali) during periods of low sea level in the glacial periods of the Pleistocene (Voris, 2000).

Final considerations

In the present study we analysed 19 462s genome-wide SNPs, following a population genomics approach to evaluate the variability and differentiation in Indo-Pacific populations of the genus *Sousa*. Our work supports previous studies where five clusters were observed. The three Indo-Pacific species, *S. sahulensis*, *S. plumbea* and *S. chinensis* were clearly separated from each other with absence of gene flow between them. Genetic segregation within *S. plumbea* was also observed separating the African Coast population from the Arabian Sea and the Bangladesh population was highly differentiated from the other species with little gene flow between them pointing towards the possibility of a fifth species of humpback dolphin. Oceanographic features have been suggested as important factors driving the divergence of these populations. The discontinuous range resulting from past sea level rise has also likely contributed to population isolation. Future studies need to investigate molecular dating to estimate the time of dispersal events and a biogeographical analysis to study the origin and dispersal of the various populations of

Sousa. This was the first study to our knowledge, to use genome-wide markers to analyse the population divergence in these dolphin species.

The clarification of the population structure within *Sousa* and the processes involved in this differentiation is extremely important for the conservation of these species. Living in nearshore habitats with freshwater input and in developing nations heavily influenced by human activities, makes the genus extremely vulnerable to fatal entanglements in fishing gear, impacts of vessel traffic and the increasing degradation of their habitat.

Acknowledgements

This research would not have been possible without the dedicated effort of WCS field assistants and our research boat crew especially Musa Kalimullah. Samples for this study were collected under permit from the Ministry of Environment and Forest, Bangladesh. We are grateful to Mr. Yunus Ali, Chief Conservator of Forest, Bangladesh, for his help with obtaining the CITES export permit for the skin samples used in this study. Funding for this work was provided by the IWC Small Cetacean Conservation Fund Awarded and Ocean Park Conservation Foundation Hong Kong to BDS and HCR. We are also grateful to Dr.Vic Peddemors and the KwaZulu-Natal Sharks Board for contributing with the samples from South Africa used in this study. The samples from Mozambique were collected with a financial support from German Dolphin Conservation Society awarded to LK. We thank the Ministry of Environment and Climate Affairs in Oman for permission to survey and sample cetaceans in the Arabian Sea. A.R. Amaral was supported by a grant (SFRH/BPD/79002/2011) from the Portuguese Science Foundation.

Author contribution

A.R.A. and H.C.R. conceived the study. A.R.A. and C. C. analysed the data and wrote the manuscript. B.D.S., R.M., T.C., R. B., G.M., G.J.P., M.K., T.A.J., L.K., A.G. and R.LB Jr., were involved in sample collection.

Data Availability

All the primary data used in this study is deposited in DRYAD.

References

Ahti PA, Coleman RR, DiBattista JD, Berumen ML, Rocha LA, Bowen BW (2016). Phylogeography of Indo-Pacific reef fishes: sister wrasses Coris gaimard and C. cuvieri in the Red Sea, Indian Ocean and Pacific Ocean. J Biogeogr **43**(6)**:** 1103-1115.

Alexander A, Steel D, Hoekzema K, Mesnick SL, Engelhaupt D, Kerr I *et al* (2016). What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? Molecular Ecology **25**(12)**:** 2754-2772.

Amaral AR, Beheregaray LB, Bilgmann K, Boutov D, Freitas L, Robertson KM *et al* (2012). Seascape Genetics of a Globally Distributed, Highly Mobile Marine Mammal: The Short-Beaked Common Dolphin (Genus *Delphinus*). Plos One **7**(2).

Amaral AR, Moeller LM, Beheregaray LB, Coelho MM (2011). Evolution of 2 Reproductive Proteins, ZP3 and PKDREJ, in Cetaceans. Journal of Heredity **102**(3)**:** 275-282.

Amaral AR, Smith BD, Mansur RM, Brownell RL, Jr., Rosenbaum HC (2017). Oceanographic drivers of population differentiation in Indo-Pacific bottlenose (*Tursiops aduncus*) and humpback (*Sousa* spp.) dolphins of the northern Bay of Bengal. Conservation Genetics **18**(2)**:** 371-381.

Bass AL, Epperly SP, Braun-McNeil J (2006). Green turtle (*Chelonia mydas*) foraging and nesting aggregations in the Caribbean and Atlantic: impact of currents and behavior on dispersal. Journal of Heredity **97:** 346-354.

Bowen BW, Gaither MR, DiBattista JD, Iacchei M, Andrews KR, Grant WS *et al* (2016). Comparative phylogeography of the ocean planet. Proc Natl Acad Sci USA **113**(29)**:** 7962-7969.

411 Braulik GT, Findlay K, Cerchio S, Baldwin R (2015). Assessment of the Conservation
412 Status of the Indian Ocean Humpback Dolphin (Sousa plumbea) Using the IUCN Red
413 List Criteria. In: Jefferson TA and Curry BE (eds) *Humpback Dolphins*. Vol. 72, pp 119-
414 141.
415
416 Carroll EL, Baker CS, Watson M, Alderman R, Bannister J, Gaggiotti OE *et al* (2015).
417 Cultural traditions across a migratory network shape the genetic structure of southern
418 right whales around Australia and New Zealand. Scientific Reports **5**.
419
420 Cheng XH, Xie SP, McCreary JP, Qi YQ, Du Y (2013). Intraseasonal variability of sea
421 surface height in the Bay of Bengal. J Geophys Res-Oceans **118**(2)**:** 816-830.
422
423 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA *et al* (2011). The
424 variant call format and VCFtools. Bioinformatics **27**(15)**:** 2156-2158.
425
426 Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008). Substantial biases in ultra-short
427 read data sets from high-throughput DNA sequencing. Nucleic Acids Res 36: e105.
428
429 Earl DA, vonHoldt BM (2012). STRUCTURE HARVESTER: a website and program for
430 visualizing STRUCTURE output and implementing the Evanno method. Conservation
431 Genetics Resources **4**(2)**:** 359-361.
432
433 Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES *et al* (2011). A
434 Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.
435 Plos One **6**(5).
436
437 Eren AM, Vineis JH, Morrison HG, Sogin ML (2013). A filtering method to generate
438 high quality short reads using illumina paired-end technology. Plos One 8: e66643.
439
440 Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals
441 using the software STRUCTURE: a simulation study. Molecular Ecology **14**(8)**:** 2611-
442 2620.
443
444 Farhadi A, Jeffs AG, Farahmand H, Rejiniemon TS, Smith G, Lavery SD (2017).
445 Mechanisms of peripheral phylogeographic divergence in the indo-Pacific: lessons from
446 the spiny lobster Panulirus homarus. BMC Evol Biol **17**.
447
448 Foll M, Gaggiotti O (2008). A Genome-Scan Method to Identify Selected Loci
449 Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective.
450 Genetics **180**(2)**:** 977-993.
451
452 Foote AD, Liu Y, Thomas GWC, Vinar T, Alfoldi J, Deng JX *et al* (2015). Convergent
453 evolution of the genomes of marine mammals. Nature Genet **47**(3)**:** 272-+.
454
455 Foote AD, Vijay N, Avila-Arcos MC, Baird RW, Durban JW, Fumagalli M *et al* (2016).
456 Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. Nature
457 Communications **7**.
458

Frichot E, Mathieu F, Trouillon T, Bouchard G, Francois O (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. Genetics **196**(4)**:** 973-+.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012). Harnessing genomics for delineating conservation units. Trends Ecol Evol **27**(9)**:** 489-496.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q *et al* (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. Plos One **9**(2).

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. Molecular Ecology **22**(11)**:** 2898-2916.

Holland BR, Delsuc F, Moulton V (2005). Visualizing conflicting evolutionary hypotheses in large collections of trees: Using consensus networks to study the origins of placentals and hexapods. Systematic Biology **54**(1)**:** 66-76.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources **9**(5)**:** 1322-1332.

Hussain Z, Acharya G. (1994). *Vol. 2*.

Jefferson TA, Rosenbaum HC (2014). Taxonomic revision of the humpback dolphins (Sousa spp.), and description of a new species from Australia. Marine Mammal Science **30**(4)**:** 1494-1541.

Jefferson TA, Smith BD (2016a). Re-assessment of the Conervation Status of the Indo-Pacific Humpback Dolphin (*Sousa chinensis*) Using the IUCN Red List Criteria. Advances in Marine Biology **73:** 1-26.

Jefferson TA, Smith BD (2016b). Re-assessment of the Conservation Status of the Indo-Pacific Humpback Dolphin (Sousa chinensis) Using the IUCN Red List Criteria. In: Jefferson TA and Curry BE (eds) *Humpback Dolphins*. Vol. 73, pp 1-26.

Jombart T, Devillard S, Balloux F (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet **11**.

Jombart T, Devillard S, Dufour AB, Pontier D (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity **101**(1)**:** 92-103.

Karczmarski L (1999). Group dynamics of humpback dolphins (Sousa chinensis) in the Algoa Bay region, South Africa. J Zool **249:** 283-293.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources **15**(5)**:** 1179-1191.

Kopps AM, Krutzen M, Allen SJ, Bacher K, Sherwin WB (2014). Characterizing the socially transmitted foraging tactic "sponging" by bottlenose dolphins (Tursiops sp.) in the western gulf of Shark Bay, Western Australia. Marine Mammal Science **30**(3)**:** 847-863.

Li CH, Corrigan S, Yang L, Straube N, Harris M, Hofreiter M *et al* (2015). DNA capture reveals transoceanic gene flow in endangered river sharks. Proc Natl Acad Sci USA **112**(43)**:** 13302-13307.

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

Lischer HEL, Excoffier L (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics **28**(2)**:** 298-299.

Longhurst AR (2006). *Ecological Geography of the Sea*, 2nd edn. edn. Academic Press: San Diego.

Lotterhos KE, Whitlock MC (2014). Evaluation of demographic history and neutral parameterization on the performance of F-ST outlier tests. Molecular Ecology **23**(9)**:** 2178-2192.

Mansur RM, Strindberg S, Smith BD (2012). Mark-resight abundance and survival estimation of Indo-Pacific bottlenose dolphins, Tursiops aduncus, in the Swatch-of-No-Ground, Bangladesh. Marine Mammal Science **28**(3)**:** 561-578.

Mendez M, Jefferson TA, Kolokotronis S-O, Kruetzen M, Parra GJ, Collins T *et al* (2013). Integrating multiple lines of evidence to better understand the evolutionary divergence of humpback dolphins along their entire distribution range: a new dolphin species in Australian waters? Molecular Ecology **22**(23)**:** 5936-5948.

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**(11)**:** 2212-2228.

Mendez M, Subramaniam A, Collins T, Minton G, Baldwin R, Berggren P *et al* (2011). Molecular ecology meets remote sensing: environmental drivers to population structure of humpback dolphins in the Western Indian Ocean. Heredity **doi:10.1038/hdy.2011.21**.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P *et al* (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (Orcinus orca) indicates multiple species. Genome Res **20**(7)**:** 908-916.

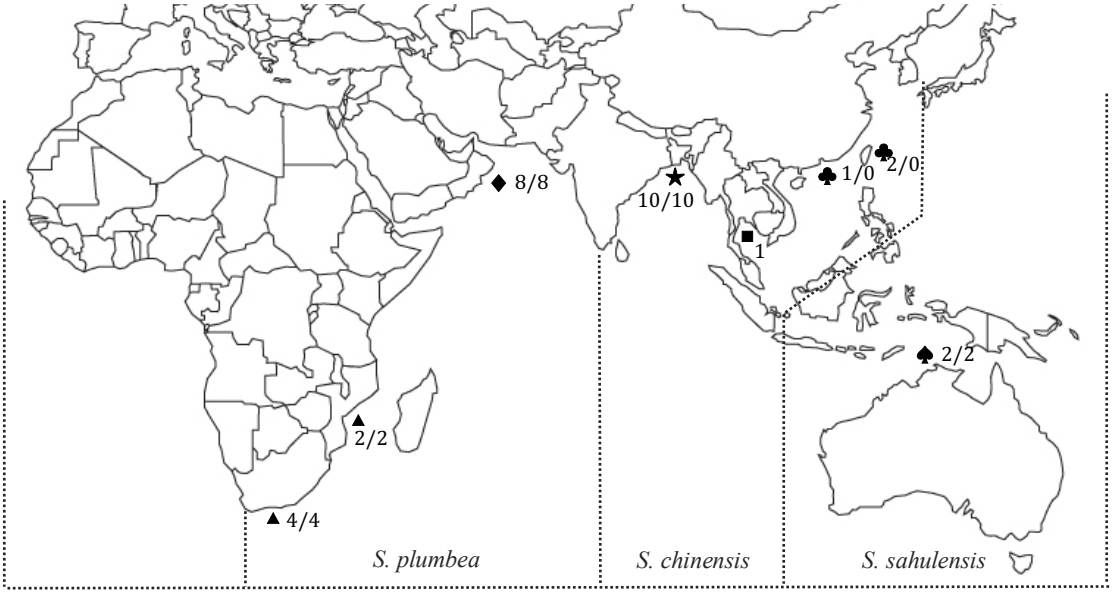
Nielsen R, Korneliussen T, Albrechtsen A, Li YR, Wang J (2012). SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. Plos One **7**(7).

555 Palumbi SR (1992). Marine speciation on a small planet. Trends Ecol Evol **7**(4)**:** 114-
556 118.
557
558 Parra GJ, Cagnazzi D (2016). Conservation Status of the Australian Humpback Dolphin
559 (Sousa sahulensis) Using the IUCN Red List Criteria. Humpback Dolphins (Sousa Spp):
560 Current Status and Conservation, Pt 2 **73:** 157-192.
561
562 Parra GJ, Corkeron PJ, Arnold P (2011). Grouping and fission-fusion dynamics in
563 Australian snubfin and Indo-Pacific humpback dolphins. Animal Behaviour **82**(6)**:** 1423-
564 1433.
565
566 Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using
567 multilocus genotype data. Genetics **155**(2)**:** 945-959.
568
569 Riesch R, Barrett-Lennard LG, Ellis GM, Ford JKB, Deecke VB (2012). Cultural
570 traditions and the evolution of reproductive isolation: ecological speciation in killer
571 whales? Biol J Linnean Soc **106**(1)**:** 1-17.
572
573 Savolainen O, Lascoux M, Merila J (2013). Ecological genomics of local adaptation.
574 Nature Reviews Genetics **14**(11)**:** 807-820.
575
576 Smith BD, Ahmed B, Mansur R, Strindberg S (2008). Species occurrence and
577 distributional ecology of nearshore cetaceans in the Bay of Bengal, Bangladesh, with
578 abundance estimates for Irrawady dolphins *Orcaella brevirostris* and finless porpoises
579 *Neophocoena phocaenoides*. Journal of Cetacean Research and Management **10**(1)**:** 45-
580 58.
581
582 Stamatakis A (2014). RAxML version 8: a tool for phylogenetic analysis and post-
583 analysis of large phylogenies. Bioinformatics **30**(9)**:** 1312-1313.
584
585 Steeman ME, Hebsgaard MB, Fordyce RE, Ho SYW, Rabosky DL, Nielsen R *et al*
586 (2009). Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. Systematic
587 Biology **58**(6)**:** 573-585.
588
589 Tezanos-Pinto G, Baker CS, Russell K, Martien K, Baird RW, Hutt A *et al* (2009). A
590 Worldwide Perspective on the Population Structure and Genetic Diversity of Bottlenose
591 Dolphins (Tursiops truncatus) in New Zealand. Journal of Heredity **100**(1)**:** 11-24.
592
593 Voris HK (2000). Maps of Pleistocene sea levels in Southeast Asia: shorelines, river
594 systems and time durations. J Biogeogr **27**(5)**:** 1153-1167.
595
596 Whitlock MC, Lotterhos KE (2015). Reliable Detection of Loci Responsible for Local
597 Adaptation: Inference of a Null Model through Trimming the Distribution of F-ST. Am
598 Nat **186:** S24-S36.
599
600

List of Figures

List of Tables

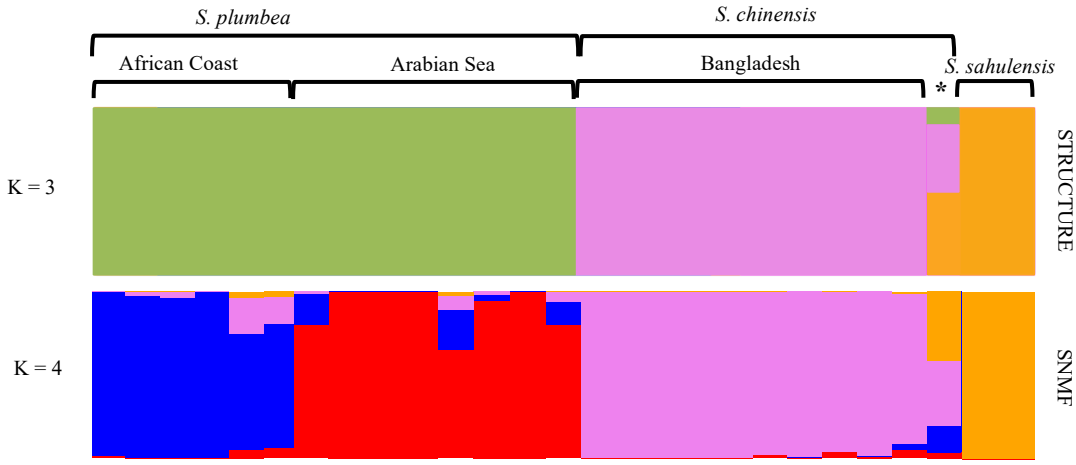
643
644
645 Figure 1



646

647 Figure 2

648

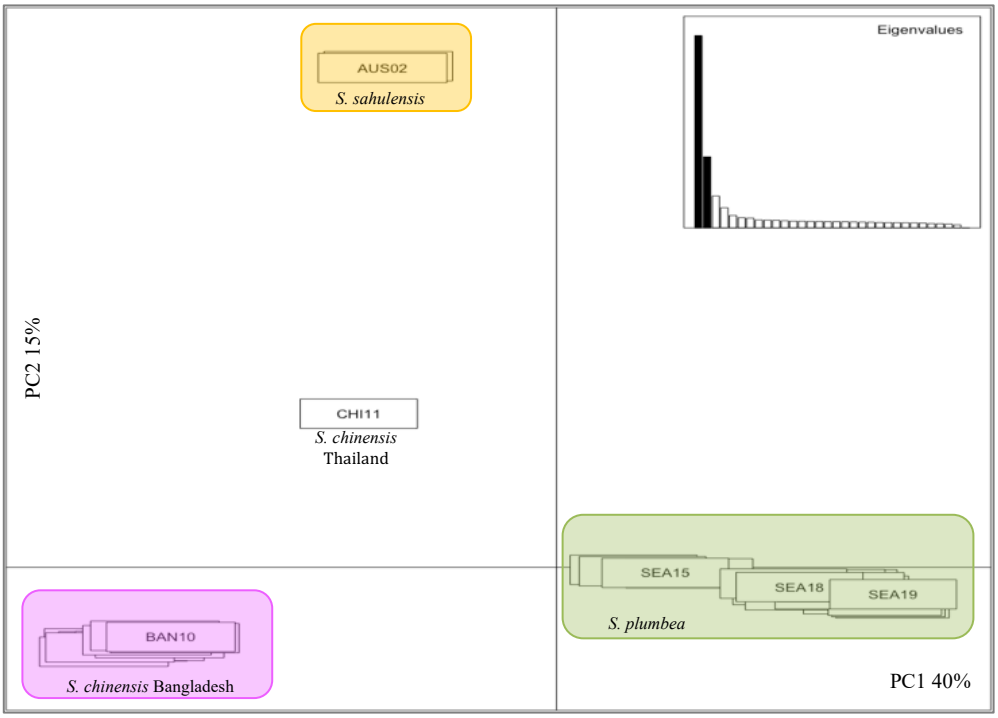


649

650

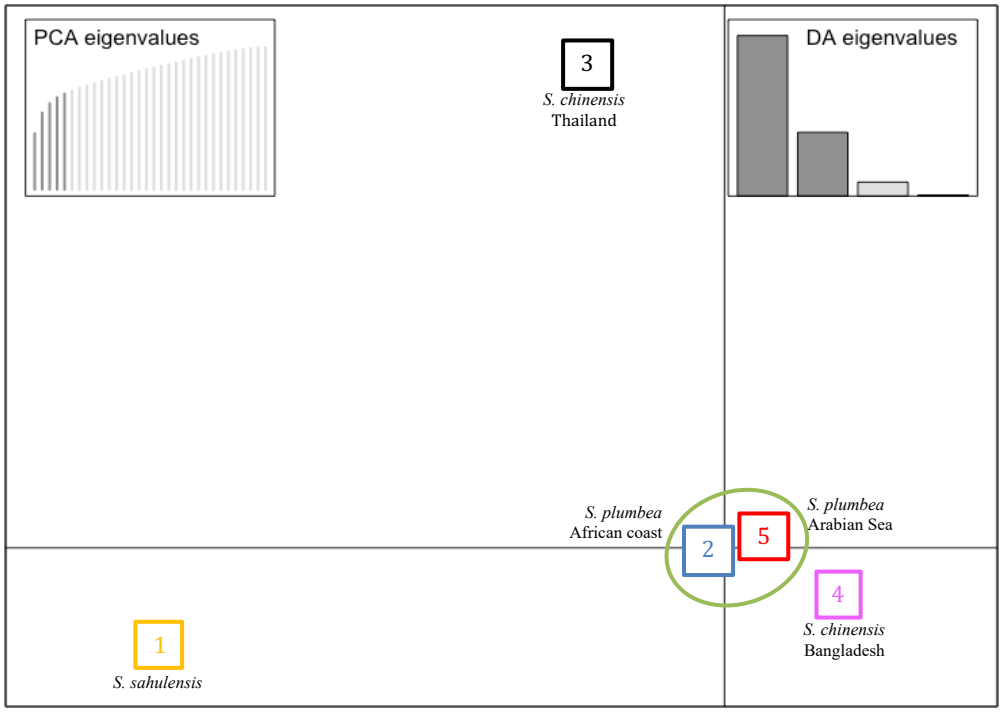
651

652 Figure 3



653

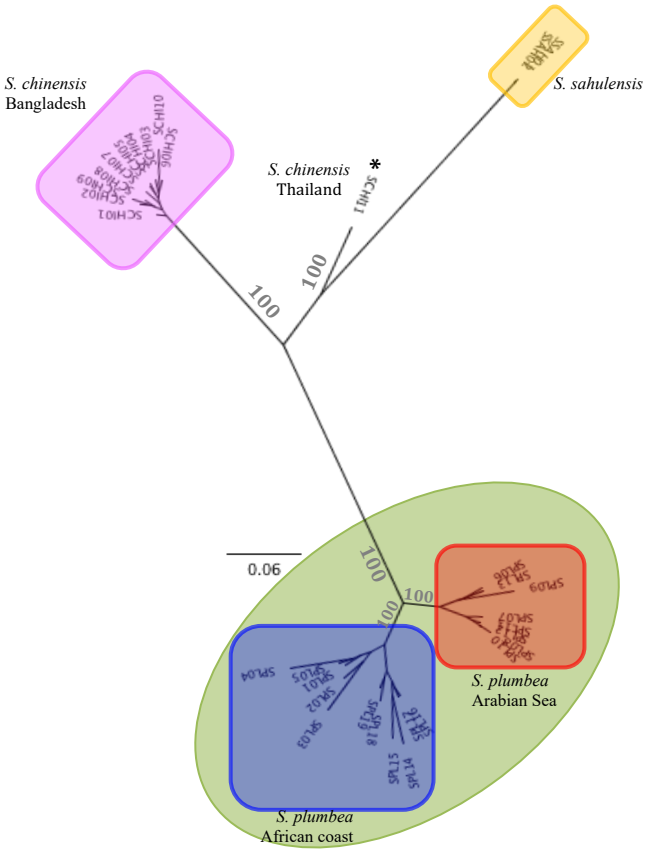
654 Figure 4



655

656

657 Figure 5



658
659

660 Table 1

	Bangladesh	African Coast	Arabian Sea
Bangladesh	-	0.7142	0.6698
African Coast	-	-	0.3385
Arabian Sea	-	-	-

661

662

663

664

665

666

667 Table 2

K	Reps	Mean LnP (K)	Delta K
2	20	-142077.4400	-
3	20	-124200.8100	12.7604
4	20	-124508.5200	0.4863
5	20	-124925.4600	0.2613
6	20	-124940.8000	-

668 *K – number of clusters tested; Reps – number of repetitions.*