MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 5936–5948 doi: 10.1111/mec.12535

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?

MARTIN MENDEZ,*† THOMAS A. JEFFERSON,‡ SERGIOS-ORESTIS KOLOKOTRONIS,§ MICHAEL KRÜTZEN,¶ GUIDO J. PARRA,**†† TIM COLLINS,†‡‡ GIANA MINTON,‡‡§§ ROBERT BALDWIN,‡‡ PER BERGGREN,¶¶ ANNA SÄRNBLAD,*** OMAR A. AMIR,††† VIC M. PEDDEMORS,‡‡‡§§§ LESZEK KARCZMARSKI,¶¶¶ ALMEIDA GUISSAMULO,**** BRIAN SMITH,† DIPANI SUTARIA,†††† GEORGE AMATO* and HOWARD C. ROSENBAUM*†

*Sackler Institute for Comparative Genomics, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA, †Wildlife Conservation Society, 2300 Southern Boulevard, Bronx, New York, 10460, USA, ‡Clymene Enterprises, 13037 Yerba Valley Way, Lakeside, CA 92040, USA, §Department of Biological Sciences, Fordham University, 441 East Fordham Road, Bronx, NY 10458, USA, ¶Evolutionary Genetics Group, Anthropological Institute & Museum, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland, **Cetacean Ecology, Behaviour and Evolution Lab, School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia, ††South Australian Research and Development Institute (SARDI), Aquatic Sciences, West Beach, SA 5024, Australia, ‡‡Environment Society of Oman, Al-Khuwair, Muscat, Sultanate of Oman, §§Sarawak Dolphin Project, Institute of Biodiversity and Environmental Conservation, University Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia, ¶¶School of Marine Science and Technology, Newcastle University, Newcastle upon Tyne, UK, ***Department of Zoology, Stockholm University, Stockholm, Sweden, †††Institute of Marine Sciences, University of Dar es Salaam, Zanzibar, Tanzania, ‡‡‡School of Biological and Conservation Sciences, University of KwaZulu-Natal, Durban, South Africa, §§§Cronulla Fisheries Reseach Centre, Cronulla, Sydney, New South Wales, Australia, ¶¶¶The Swire Institute of Marine Science, School of Biological Sciences, The University of Hong Kong, Cape d'Aguilar, Shek O, Hong Kong, ****Universidade Eduardo Mondlane, Museu de Historia Natural, Pracça Travessia do Zambeze, Maputo, Mozambique, ††††School of Earth and Environmental Sciences, James Cook University, Townsville, Australia*

Abstract

The conservation of humpback dolphins, distributed in coastal waters of the Indo-West Pacific and eastern Atlantic Oceans, has been hindered by a lack of understanding about the number of species in the genus (*Sousa*) and their population structure. To address this issue, we present a combined analysis of genetic and morphologic data collected from beach-cast, remote-biopsied and museum specimens from throughout the known *Sousa* range. We extracted genetic sequence data from 235 samples from extant populations and explored the mitochondrial control region and four nuclear introns through phylogenetic, population-level and population aggregation frameworks. In addition, 180 cranial specimens from the same geographical regions allowed comparisons of 24 morphological characters through multivariate analyses. The genetic and morphological data showed significant and concordant patterns of geographical segregation, which are typical for the kind of demographic isolation displayed by species units, across the *Sousa* genus distribution range. Based on our combined genetic and morphological analyses, there is convincing evidence for at least four species within the genus (*S. teuszii* in the Atlantic off West Africa, *S. plumbea* in the central and western Indian Ocean, *S. chinensis* in the eastern Indian and West Pacific Oceans, and a new as-yet-unnamed species off northern Australia).

Correspondence: Martin Mendez and Howard C. Rosenbaum, Fax: 718 220 5100; E-mails: mmendez@wcs.org, hrosenbaum@wcs.org

Keywords: conservation genetics, mammals, molecular evolution, speciation, humpback dolphins, cetaceans

Received 15 March 2013; revision received 12 August 2013; accepted 14 September 2013

Introduction

Understanding evolutionary divergence is essential to our understanding of species and populations, which in turn is paramount for biodiversity conservation (Vogler & DeSalle 1994; Goldstein *et al.* 2000). A variety of species concepts provide theoretically robust frameworks to explore the divergence process using different lines of evidence. In this article, we refer to 'species units' as defined by the phylogenetic species concept (PSC) and the biological species concept (BSC) (Cracraft 1983; De Queiroz 2007). Phylogenetic species units are characterized by evolutionary uniqueness resulting from significant divergence between such units and are usually assessed with phylogenetic methods such as character data or genetic trees (Cracraft 1983). Biological species are units reproductively isolated from one another either by allopatric distribution or by behavioural or physiological mechanisms that prevent gene flow and produce effective isolation even in sympatry (De Queiroz 2007). Cetacean species provide good examples of speciation due to either of these mechanisms. For instance, divergence in right whales, which has been studied using molecular tools, is known to have occurred as a consequence of geographical isolation due to glacial/interglacial periods and suspected antitropical behaviour, with resulting different species in the southern hemisphere, the North Atlantic and North Pacific Oceans (Rosenbaum *et al.* 2000; Gaines *et al.* 2005). Examples of sympatric cetaceans can be found within the Stenella genus, where divergence is presumed to be driven by behavioural and/or ecological factors rather than by a disjoint distribution (Oviedo 2007). Similar to the variety of species concepts, there is a suite of population concepts that are related to ecological, evolutionary and statistical paradigms (Waples & Gaggiotti 2006), all of which assume a group of mating individuals of the same species sharing space and time and, as a result, can be identified when there is significant population structure and negligible gene flow.

Different sources of data have been used for assessments of species and population units. Although morphological data have enabled higher-level taxonomic evaluations (Gatesy & O'Leary 2001; O'Leary *et al.* 2003) characteristic of species units, these data are often unable to resolve recent divergence and cryptic variation normally present between populations (Rosenbaum *et al.* 2000). Genetic data have been proved to be powerful in evaluating taxonomic hypotheses that range from deep (Krause *et al.* 2008; Roca *et al.* 2009) to shallow (Palsbøll *et al.* 2004; Sudarto *et al.* 2010; Welch *et al.* 2011) divergence events, therefore permitting assessments of species to population units. The combination of morphological and genetic data allows for complementary analyses of information subject to different evolutionary forces and can therefore provide a robust picture of divergence patterns (Beasley *et al.* 2005; Chivers *et al.* 2005; Caballero *et al.* 2007). As a result, molecular and morphological data are frequently used in parallel to resolve taxonomic uncertainties and identify cryptic species (Beasley *et al.* 2002, 2005; Lefebure *et al.* 2006; Caballero *et al.* 2007; Charlton-Robb *et al.* 2011).

Unambiguous phylogenetic signals or discrete, fixed characters from genetic data are expected for operational units at the species level (Shaffer & Thomson 2007). However, units that have diverged recently or exhibit ongoing gene flow (i.e. populations) will not be accurately depicted with strictly bifurcating algorithms, due to the presence of incomplete lineage sorting, and should be treated with methods that allow visualizing reticulated relationships (Pearse & Crandall 2004), such as networks (Posada & Crandall 2001). In addition, population-level analytical frameworks can increase our understanding of divergence patterns (i.e. magnitude and directionality of gene flow) and the processes that led to those patterns (e.g. isolation, migration). Combining these approaches to understand evolutionary relationships has been a key to develop meaningful conservation strategies. As an example, the taxonomy of the dolphin genus *Sotalia* was recently revised based on concordant patterns of divergence in genetics and morphology (Caballero *et al.* 2007, 2008) and resulted in the acceptance of two distinct species with different conservation needs.

In this article, we address the controversy regarding the number of species in the dolphin genus *Sousa*. Although the current formal taxonomy of this genus recognizes the existence of *S. teuszii* in the Atlantic Ocean and *S. chinensis* covering the rest of the genus distribution (Rice 1977, 1998), the scientific community has historically considered a range from a single, highly variable species (*S. chinensis*) to four species (*S. teuszii* in the Atlantic Ocean, *S. plumbea* in the Indian Ocean, *S. chinensis* in the Indo-West Pacific Ocean and the possible existence of an as-yet-unnamed species occurring along the coast of northern Australia) (Fig. 1). Humpback dolphins are considered 'vulnerable' (*S. teuszii*)

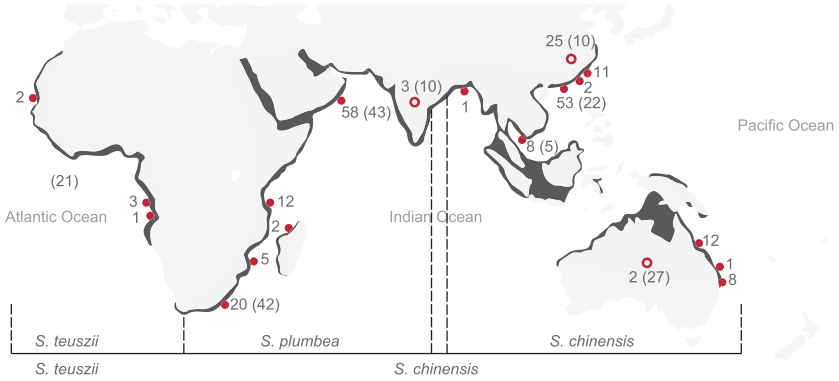

Fig. 1 Study area map. The global distribution of humpback dolphins is shown as shaded along the coasts. Solid circles indicate sampling sites, with genetic sample sizes alongside and morphological sample sizes (in parentheses). Samples with unknown specific location within a region/country are indicated by open circles and plotted on the country/region of origin. Previously recognized taxonomic names are shown at the bottom and represent approximate locations of proposed species distribution limits. Specifically, we show the three-species taxonomy above the bottom line, with an approximate geographical split between *S. teuszii* and *S. plumbea* in western South Africa, and the approximate split between *S. plumbea* and *S. chinensis* as a split range east of India, given current uncertainty about proposals for that split. Below the bottom line, we show the two-species taxonomy including only *S. teuszii* and *S. chinensis*.

and 'near threatened' (*S. chinensis*), both with decreasing population trends in the IUCN Red List (http://www.iucnredlist.org). However, *S. chinesis* comes close to qualifying for vulnerable and should be reassessed following a taxonomic assessment of the genus, especially considering the implications of *S. chinensis* potentially being subdivided into multiple species. *S. chinensis* in the eastern Taiwan Strait has been identified as a discrete population, based on its distinctive colour pattern, and it is considered as critically endangered in the IUCN Red List. Current taxonomic uncertainty and the potential identification of additional demographically isolated populations hinder the development of conservation strategies based on prior operational units (Jefferson 2004; Jefferson & Hung 2004; Jefferson & Van Waerebeek 2004; Frère *et al*. 2008, 2011).

Morphological evidence supports a *S. teuszii–S. chinensis* split, and a potential additional species was included in *S. chinensis* referred to as *S. plumbea* with a known distribution in the western Indian Ocean (Jefferson & Van Waerebeek 2004). West African specimens of humpback dolphins have significantly shorter rostra and lower tooth counts compared with Southeast African, Arabian/Persian Gulf and Indian specimens. The latter have a prominent dorsal hump not present in South-East Asia, which supports *S. plumbea* as a potential third species (Jefferson & Van Waerebeek 2004).

The first lines of genetic data come from a more limited regional data set and one genetic marker, which only included animals from West Africa, South Africa, Hong Kong and Australia ($N_{WA} = 2$, $N_{SA} = 23$, $N_{HK} = 19$, $N_A = 25$, respectively). These mtDNA data showed a monophyletic Australia clade sister to a group of monophyletic clades from Hong Kong and South Africa and those from West Africa as basal to this group (Frère *et al*. 2008). A subsequent phylogenetic analysis, which added a sample from Indonesia ($N_I = 1$), increased the amount of mtDNA sequence data and added sequence information from three nuclear introns, supporting the previous findings about the monophyly of Australian samples (Frère *et al*. 2011). The latter analysis did not include samples from South Africa, hindering direct comparisons with the preceding study.

Recent population-level analyses of mtDNA control region data uncovered further variation within *S. plumbea* in the form of significant genetic structure among putative populations in Oman, Tanzania and an assemblage formed by South Africa and Mozambique (Mendez *et al*. 2011). Such levels of control region genetic differentiation ($\Phi_{ST} > 0.5$) are considered very high compared with other small cetaceans with reported population structure, such as *Stenella frontalis* (Φ_{ST} ~0.3) (Adams & Rosel 2006), *S. longirostris* (Φ_{ST} ~0.25) (Andrews *et al*. 2010), *Phocoenoides dalli* (Φ_{ST} ~0.1) (Escorza Trevino & Dizon 2000) and *Pontoporia blainvillei* (Φ_{ST} ~0.15) (Mendez *et al*. 2008, 2010). Comparably high interpopulation Φ_{ST} levels have been reported for small cetaceans with presumed distribution gaps, such as *Lagenorhynchus obscurus* in Peru, Argentina and South Africa (Φ_{ST} ~0.6) (Cassens *et al*. 2003), or those with known strong female philopatry, such as *Tursiops* sp. in Western Australia (Φ_{ST} ~0.6) (Krützen *et al*. 2004) or *Delphinapterus leucas* in the Nearctic (Φ_{ST} ~0.5) (O' Corry-Crowe *et al*. 1997).

In the present analysis, we combine genetic and morphological data from the most comprehensive range-wide sampling to date of the *Sousa* genus and draw

from phylogenetic and population-level approaches, to assess phylogenetic- and population-level relationships in this group and to address some of the potential taxonomic implications of this evidence.

Methods

Sample collection

A total of 235 tissue samples and 180 cranial specimens were used for genetic and morphological analyses, respectively. The data sets were mostly independent, although there was some overlap from stranded specimens from which we included both genetic and morphological data ($N = 7$ individuals from China). Both data sets represent the entire range of the *Sousa* genus extending along the coasts of the Atlantic, Indian and West Pacific Oceans. Specifically, our genetic data set contains samples from West Africa or 'WA' (Gabon, Congo, Mauritania), Southeast Africa or 'SEA' (South Africa, Mozambique, Tanzania, Madagascar), Arabia or 'OM' (Oman), the Indian subcontinent or 'IN' (India, Bangladesh), Thailand or 'TH' and China or 'CH' in South-East Asia, and Australia 'AUS'. Our morphological data set includes samples from the same general localities except for Bangladesh (Fig. 1, Table S1, Supporting information).

Total genomic DNA was isolated from the 235 incidentally entangled, beach-cast and biopsied humpback dolphins, including all 94 specimens used in the regional-level analysis in Africa and Arabia (Mendez *et al.* 2011) and a representative sample of Australian individuals provided by G.J. Parra. All samples were preserved in ethanol (96% v/v) or in a sodium chloride-saturated 20% dimethyl sulphoxide (DMSO) solution. Total genomic DNA was extracted from tissue samples using the QIAamp Tissue Kit (QIAGEN, Valencia, CA, USA). A fragment of the mitochondrial DNA control region (Baker *et al.* 1993) was amplified from all tissue samples. Preliminary assays to evaluate variation between sampling sites were performed using intron sequences presented in the study by Frère *et al.* (2011) and a suite of other nuclear markers. After this initial screening, we selected a set of three introns that were successfully amplified for most of our samples and that initially presented some visible variation between sampling units. Partial sequences from the parathyroid hormone (*PTH*), proteolipid protein (*PLP*) and esterase D (*ESD*) nuclear gene introns (Lyons *et al.* 1997) were also amplified for a subset of 108 specimens representing all sampling areas. The thermal profile for the mtDNA control region PCR consisted of an initial denaturation for 3 min at 94°C followed by 32 amplification cycles (30 s at 94°C, 30 s at 52°C, 1 min at 72°C) and a final 5 min of extension at 72°C. For all introns, thermal conditions started with an initial denaturation phase for 10 min at 94°C followed by 35 amplification cycles with varying temperatures and ended with a final 10 min of extension at 72°C. The amplification temperatures were as follows: for PTH, we used 30 s at 94°C, 30 s at 55°C and 1 min at 72°C; for PLP, we used 30 s at 94°C, 30 s at 60°C and 1 min at 72°C; and for ESD, we used 30 s at 94°C, 30 s at 56°C and 1 min at 72°C. All loci were sequenced in both directions using BigDye chemistry on a 3730xl DNA Analyzer (Applied Biosystems, Inc. [ABI], Foster City, CA, USA).

The morphological data stem from a cranial morphology data set developed and initially explored by Jefferson and Van Waerebeek (Jefferson & Van Waerebeek 2004). This data set consists of 24 metric and meristic characters from 180 adult humpback dolphin specimens (as determined by ossification level and skull size) collected throughout the range of the genus, closely matching the geographical coverage of the molecular data. The measurements analysed are as follows: upper tooth count (UTC), lower tooth count (LTC), tooth diameter (TD), condylobasal length (CBL), length of rostrum (LR), width of rostrum base (WRB), width of rostrum ½ (WR1/2), width of rostrum ¾ (WR3/4), width of premaxilar ½ (WP1/2), greatest width premaxilar (GWP), preorbital width (PREOW), postorbital width (POSOW), zygomatic width (ZYGW), parietal width (PARW), width of external nares (WEN), width of internal nares (WIN), length of temporal fossa (LTF), height of temporal fossa (HTF), length of orbits (LORB), length of antiorbital process (LAP), length of upper toothrow (LUTR), length of mandible (LMAN), height of mandible (HMAN) and length of mandibular symphysis (LMS). See Supplementary material and Jefferson & Van Waerebeek (2004) for full details on this data set (Table S2, Supporting information).

Genetic data analysis

DNA sequencing chromatograms were inspected and edited in Sequencher 4.8 (Gene Codes, Corp.). Nucleotide sequences were aligned in MUSCLE (Edgar 2004) with a maximum of 10 iterations (Data S1, Supporting information).

The phylogenetic relationships between haplotypes in a concatenated mitochondrial (mtDNA) and nuclear (nuDNA) sequence data set ($N = 105$), and in the mitochondrial data set ($N = 235$) and nuclear sequences ($N = 81$), were examined in a maximum-likelihood (ML) framework in the POSIX Threads build of RAxML 7.2.9–7.3.0 (Stamatakis 2006; Stamatakis & Ott 2008) using an unlinked general-time-reversible (GTR; Lanave *et al.* 1984) and among-site rate heterogeneity modelled by the Γ distribution with four discrete categories (Yang

1994) across loci. One hundred ML inferences on the condensed haplotype alignment were performed, each starting from a random-addition maximum parsimony tree. Node robustness was examined by means of 2000 bootstrap pseudoreplicates (Felsenstein 1985). The number of pseudoreplicates above which node support is not expected to vary was examined using the 'bootstrapping' frequency-based and majority-rule criteria as implemented in RAxML (Pattengale *et al.* 2010). The putative conflicting phylogenetic signal among the respective ML trees of the bootstrap resampled alignments was visualized simultaneously in a single consensus network of all bootstrap trees (Holland & Moulton 2003) in SplitsTree 4 (Huson & Bryant 2006). The length of a network edge represents the number of bootstrap trees that contain the split represented by that given edge. The consensus threshold was set at 0.1, which means that bipartitions that appeared in at least 200 of the 2000 bootstrap trees participated in network construction.

To complement the phylogenetic analysis, a general population aggregation analysis (PAA) framework (Davis & Nixon 1992) was used for evaluating nucleotide substitutions that might be diagnostic of operational units (Sarkar *et al.* 2002). Because character fixation requires a significant amount of divergence time, this analysis allows the identification of nucleotide positions that can be considered diagnostic of putative units, providing an objective and robust framework for visualizing regional divergence. The PAA was carried out excluding alignment gaps to ensure a conservative assessment, and the West African haplotype was used as reference sequence.

We employed additional population-level analyses to assess variation in the mtDNA data set in the geographical context of our sampling units: West Africa, Southeast Africa, Arabia, Thailand, China and Australia (mtDNA sample size in Bangladesh ($N = 1$) and India ($N = 3$) was too small for quantitative analyses of genetic variation, and the Indian subcontinent was therefore excluded below, excepting the construction of genetic networks). We first implemented a median-joining network in the software Network 4.6 (http://www.fluxus-engineering.com) to visualize relationships between the obtained mtDNA haplotypes (Bandelt *et al.* 1999). Our choice of networks to visualize such relationships responds to a general consensus in that networks are especially appropriate for depicting data with reticulations, which is usually the case in population-level situations or recent divergence (Posada & Crandall 2001). Then, we computed descriptive population genetic estimates for the mtDNA data set ($N = 235$), such as nucleotide and haplotype diversity (π and h; Nei 1987) and mean number of pairwise differences (k; Tajima 1983), which were calculated in DnaSP 5.10.1 (Librado & Rozas 2009). Genetic differentiation between our sampling units was estimated via the Φ_{ST} fixation index, Wright's F_{ST} analogue for nucleotide sequence diversity in Arlequin 3.5.1.2 (Excoffier & Lischer 2010) and with the net between-group distance (D_a; Nei & Li 1979), as implemented in MEGA 5 (Tamura *et al.* 2011). This Φ_{ST} estimator uses information from both the frequency distribution and nucleotide divergence in the haplotype data, which makes it especially informative about recent species divergence or pronounced population structure.

Morphological data analysis

Our rationale was to analyse the morphological data using the same regional geographical framework and sampling units as for the genetic data to evaluate whether the putative partitions are supported by both genetic and morphological data. Therefore, we assessed differentiation between West Africa, Southeast Africa, Arabia (Oman), the Indian subcontinent, Thailand, China and Australia. Specifically, a multivariate analysis of the entire morphological data set according to our regional grouping of geographical samples was implemented through a discriminant function analysis (DFA) in JMP 9.0.2 (SAS Institute Inc.). DFA estimates complementary functions of the variables that reflect the differences between groups (in this case, our sampling units). In particular, if the first two or three discriminant functions account for most of the among-group differences, then such differences can be easily visualized in a canonical plot, which is a graphical representation of the values of these variables for each of the samples (Manly 2005). Statistical significance of the DFA, which would indicate that the groupings contain statistically different sets of morphological data, was assessed through the Wilks' λ (the statistic used to evaluate significance in multivariate analyses of variance, MANOVA), Pillai's trace, Hotelling–Lawley and Roy's maximum root multivariate statistics for multiple populations (Manly 2005). This was complemented by separate analyses of variance (ANOVAs) for each individual morphological character, to evaluate their respective contribution to the overall observed pattern of variation between sampling units (Movie S1, Supporting information).

Results

Genetic data

The concatenated mtDNA–nuDNA data set shows a clear geographical partition into geographical clades, as evidenced by the consensus network of the ML bootstrap trees. The most divergent clade appears to be CH+TH (bootstrap support ~70%) and then AUS (bootstrap

support ~51%) with two well-defined haplotype clusters. A divergent and highly variable cluster appears as an African assemblage, with well-defined subclusters WA (bootstrap support ~84%), SEA and a haplotype from Oman ('O8') grouping with a few haplotypes from SEA and Mozambique at the base of the SEA cluster, and two Oman haplotypes (O14 and O15) stemming off the basal stem reticulation before SA haplotypes. All other Arabian haplotypes group in a cluster with low bootstrap support and with significant topological incongruence, as evidenced by the strong pattern of reticulation (Fig. 2a). Overall, although node support was not strong, the frequency-based bootstrapping test showed that more than 600 bootstrap pseudoreplicates would not alter significantly the global node support trend.

The mtDNA data set shows patterns congruent with those of the concatenated mtDNA–nuDNA data set. Although the AUS, BAN, CH, TH, African (with SEA and WA) and Arabian (OM) assemblages are equally distinct and supported in both data sets, the mtDNA data set shows more clear-cut patterns with fewer reticulations. In addition, in this data set, there is an Arabian haplotype that is more closely associated with the African haplotypes than with the other Arabian sequences (Fig. 2b). Although 1200 bootstrap pseudoreplicates were needed to assess node robustness, as shown by the frequency-based bootstrapping criterion, the main geographical clusters were well supported (AUS: 94%, SEA: 86%, WA: 86%, CH+TH: 85%).

The nuDNA data resulted in patterns that were somewhat similar to those presented here, although with significantly lower resolution (CH+TH: 64%). Geographically defined clades were evident from the mtDNA and nuclear *ESD* loci, differentiating AUS, SE Africa and the Indian Ocean. *PLP* and *PTH* were partially informative in differentiating some groups from Australia, China, West Africa, SE Africa and Oman. While these data did not mask or swamp the patterns ascertained from the

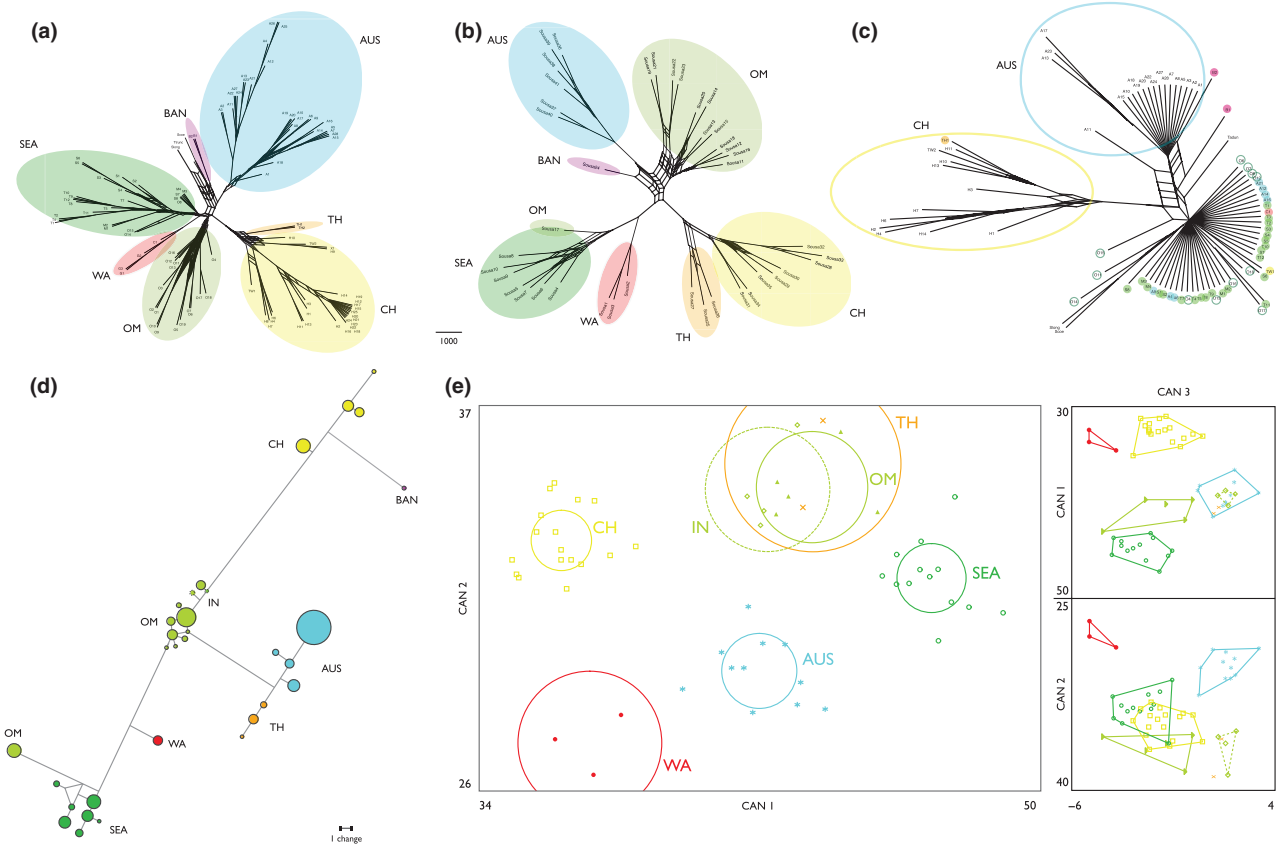

Fig. 2 Molecular phylogenetic relationships and clustering according to morphological characters. (a) Consensus networks built out of 2000 maximum-likelihood phylogenetic bootstrap trees for the concatenated mtDNA and nuDNA data sets; (b) mtDNA data set; (c) nuDNA data set. The scale bar indicates the number of bootstrap trees that contain the split represented by a given edge; (d) median-joining mtDNA haplotype network. Haplotype circle sizes are proportional to their frequencies; (e) scatter plot of canonical scores for morphological characters. Clouds corresponding to the 95% confidence interval are indicated for each region in the chart corresponding to the first two canonical scores. The panels showing canonical scores 1 vs. 3 and 2 vs. 3 display polygons that delimit the outer boundaries of the sample clouds for each region and have no associated statistical value.

phylogenetic reconstruction of mtDNA sequences, they did not significantly increase the informativeness of our overall data (Fig. 2c).

The PAA for the mitochondrial data allowed us to diagnose all regional mtDNA haplotype groupings (West Africa, Southeast Africa, etc) univocally by a series of characters. In some cases, single characters in a specific position in the alignment (indicated by a number between 1 and 441 corresponding to such position) are sufficient for this diagnosis, so we call these 'diagnostic sites'. In other cases, a combination of characters is needed to provide diagnostic power, so we call these groups of sites 'compound diagnostic sites' and indicate with a '+' sign that these sites can be diagnostic when considered together as a group (Fig. 3). For instance, the WA haplotype at the top of the alignment is set apart from the rest by a number of compound characters: sites 7(C)+123(A), 7(C)+189(T), etc. The SEA haplotype assemblage can be diagnosed as unique by sites 123(G) and 377(C) (with the OM haplotype clustering with SEA as the only exception). Similarly, the TH+CH haplotype assemblage can be diagnosed by sites 74(G), 75(C) and 235(T). The AUS haplotype group has distinctive characters at sites 44(G), 76(T), 85(G), 100(T), 253 (A), 267(A) and 391(A). Broader-scale diagnosable groupings are African haplotypes (excepting the OM haplotype mentioned above) [set apart by sites 7(C), 375(T), 396(G)], Arabian + Indian haplotypes [diagnosable by sites 91(G), 101(C), 396(A)] and South-East Asia + Oceania [set apart by site 42(A)]. The nuclear data set concatenating the three intron sequences resulted much less variable than the mitochondrial one, but diagnostic sites could still be found: a single fixed diagnostic character allowed us to diagnose the TH+CH assemblage from all other haplotypes in this analysis, a second character separates the Australian haplotypes from all other haplotypes, and a third character differentiates the African+Arabian haplotypes from all other haplotypes in this analysis (Supporting information).

All mtDNA haplotypes were 'private' to specific regional samples, with no shared haplotypes between regions. The one Bangladesh and three Indian samples presented nucleotide differences from all others in our data set, therefore generating four unique haplotypes.

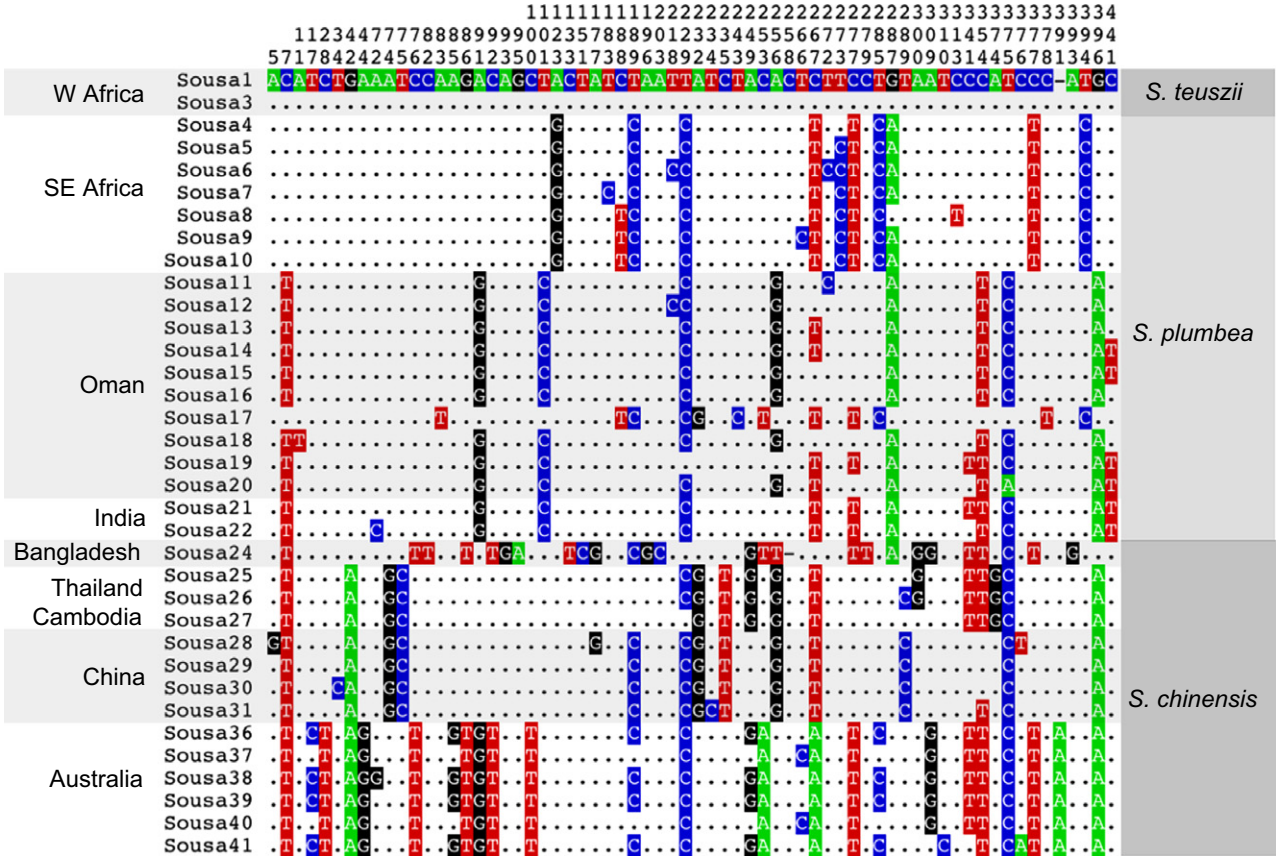

Fig. 3 mtDNA PAA character matrix, displaying only the polymorphic sites that resolve mtDNA haplotypes. Nucleotide position, as referred to in the text, is at the top and should be read vertically from top to bottom. The genus' three-species taxonomy is depicted at the right, with the new proposed species exclusively composed by the Australian haplotypes (Sousa36-Sousa41).

Table 1 Population sample size and genetic diversity of regional samples (WA: West Africa; SEA: South east Africa; OM: Oman; TH: Thailand; CH: China; AUS: Australia)

Region	*N*	*h*	*H* (SD)	*k* (SD)	π (SD)
WA	6	1	0 (0)	0 (0)	0 (0)
SEA	39	8	0.82 (0.03)	2.79 (1.51)	0.04 (0.02)
OM	58	10	0.79 (0.03)	7.29 (3.46)	0.11 (0.06)
TH	8	3	0.61 (0.16)	0.93 (0.71)	0.01 (0.01)
CH	91	8	0.34 (0.06)	1.41 (0.87)	0.02 (0.01)
AUS	23	4	0.55 (0.04)	3.73 (1.94)	0.05 (0.03)

N, number of individuals; *h*, number of haplotypes; *H*, haplotype diversity; *k*, average number of nucleotide differences; π, average number of pairwise nucleotide differences per site; SD, standard deviation.

Table 2 Pairwise genetic distances between regional samples (WA: West Africa; SEA: Southeast Africa; OM: Oman; TH: Thailand; CH: China; AUS: Australia). Above the diagonal are the net between-group genetic distances, and below the diagonal are the Φ_{ST} fixation index distances. All Φ_{ST} values are statistically highly significant ($P < 0.001$)

Sites	WA	SEA	OM	TH	CH	AUS
WA		0.028	0.016	0.039	0.035	0.053
SEA	0.78		0.023	0.055	0.043	0.049
OM	0.50	0.64		0.033	0.032	0.040
TH	0.97	0.89	0.62		0.025	0.060
CH	0.92	0.91	0.76	0.85		0.068
AUS	0.87	0.87	0.73	0.87	0.93	

These samples were only used for the phylogenetic trees and haplotype network as the small sample sizes for these two locations preclude population-level analyses. The median-joining haplotype network of the mtDNA data displays a very clear geographical structure, with all haplotypes from each region forming separate clusters (with the exception of one haplotype from Oman that clustered with WSA haplotypes) (Fig. 2d).

The OM, SEA and AUS regional samples present the highest genetic diversity, while CH displays the lowest, despite its large sample size (Table 1). All putative populations were statistically different in the pairwise genetic comparisons, with high significance values ($P < 0.001$) and large fixation indices ($\Phi_{ST} > 0.5$) (Table 2). The net between-group genetic distances were consistently larger for the comparisons involving the AUS population, followed by those involving WA (Table 2).

Morphological data

When all individual morphological characters were simultaneously evaluated through the DFA, the regional sampling resulted in statistical significance by all statistics used: Wilks' $\lambda = 4.03 \times 10^{-5}$ $P < 0.0001$, Pillai's trace = 4.84 $P < 10^{-4}$, Hotelling–Lawley = 36.22 $P < 10^{-4}$ and Roy's maximum root = 19.55 $P < 10^{-4}$. In addition, all individual character ANOVAs between the regional samples, with the exception of one variable (WR3/4, $P = 0.173$), resulted in statistical significance (TD, WRB, PREOW, HTF, LORB, LMAN and HMAN with $P < 0.05$, and all other variables with $P < 0.0001$). A visual inspection of the scatter plot of the first two canonical scores shows separated clusters at the 95% confidence level of the regional samples WA, SEA, CH and AUS. The samples OM, IN and TH form overlapping clusters. The scatter plots of the first and third, and the second and third canonical scores consistently support the patterns observed with the first two scores (Fig. 2e).

Discussion

Based on the most comprehensive sampling of *Sousa* to date, our study provides genetic and morphological evidence supporting the need for a revision of the current taxonomy of humpback dolphins to consider new putative species.

The evolutionary divergence patterns evaluated in this study result from the combined action of the mutation–drift equilibrium (Futuyma 2005) and the potential presence of barriers to dispersal or philopatric social behaviour. These forces commonly create a continuum of genetic variation from panmixia within populations, to strong population structure, and finally discrete species units. Our ability to detect the latter units is subject to the degree of resolution in the existing data and the power of the statistical methods employed to interpret those data. Our data show that all putative units studied (West Africa, Southeast Africa, Arabia-Oman, the Indian subcontinent, Thailand, China and Australia) exhibited extreme and significant differentiation in our analysis of genetic structure and shared no mtDNA haplotypes, formed well-resolved clusters in the phylogenetic analyses when considering nuDNA and mtDNA jointly and separately, are diagnosable under the PAA approach and exhibited statistically significant morphological differentiation (with the exception of Arabia-Oman, the Indian subcontinent and Thailand, which formed a single cluster).

The position of the Bangladesh samples in the phylogenetic trees of all data sets (concatenated nuDNA–mtDNA, the mtDNA and the nuDNA) is interesting, given its high support and clustering with the outgroup *Tursiops truncatus* and *Stenella* samples. These samples seem as divergent as those from Australia and therefore merit further attention. The Oman mtDNA haplotype that clusters with those of Southeast Africa

occupies an interesting position within the tree topology, given that all other Oman mtDNA haplotypes are grouped in a single Arabian cluster and that no shared haplotypes exist between sampling units. Despite the strong resulting population structure between these units and the otherwise clear phylogenetic pattern (also supported by the morphological data), this topology suggests a degree of connectivity and potential for sympatry between individuals of East Africa and Arabia, which might be historical and has been previously postulated to explain the colonization of new habitats from Arabia into East Africa (Mendez *et al.* 2011).

The different genetic markers employed in our study offered different levels of resolution. Our nuDNA data displayed much lower resolution than the mtDNA data. Such lower resolution in the nuclear DNA data is to be expected, as nuclear introns are more conserved than noncoding mtDNA sequences given their associated regulatory functions (Alberts *et al.* 2008). Comparable discrepancies in resolution between intron and mitochondrial DNA data have also been observed for other cetaceans, such as humpback whales in California and Hawaii, for instance (Palumbi & Baker 1994). The differences between whales in these locations were attributed to differences in genetic drift patterns of nuclear and mitochondrial DNA or, as is common among some cetaceans, differences between male and female dispersal patterns (Palumbi & Cipriano 1998). We acknowledge the presence of ancestral – shared yet incongruent – polymorphisms among taxa probably caused by compacted divergence in time and evolutionary genetic processes, such as incomplete lineage sorting and hybridization. The extent of homoplasy manifested in the nuclear DNA data set did not allow for the confident identification of numerous clear, shared synapomorphies and of monophyletic assemblages. These results hint at a recent origin for the *Sousa* species complex examined here.

Potential drivers of divergence in cetaceans include fragmented distributions, environmental boundaries and complex social behaviour causing local adaptation. *Sousa teuszii* is considered isolated in West Africa, possibly by the Benguela oceanographic system, which explains the strong and significant genetic and morphologic differentiation between this and all other sampling units in our study (Jefferson & Van Waerebeek 2004). In the absence of distributional gaps, it has been proposed that population boundaries for humpback dolphins along Southeast Africa and Arabia might be driven by environmental breaks. Mendez *et al.* (2011) showed that animals from the coasts of South Africa and Mozambique display no population structure, and those from Mozambique and Tanzania, and from Tanzania and Oman, display very strong population structure, suggesting breaks to gene flow along the Tanzanian coast, and in the area between Tanzania and Oman. Interestingly, the observed areas of ongoing gene flow are concordant with a continuous oceanographic regime throughout the Mozambique channel, and areas of genetic boundaries were matched by the presence of oceanographic breaks, for instance, between the Arabian Sea Upwelling Province and the East Africa Coastal Province (Mendez *et al.* 2011). Similar patterns of coincidental environmental and genetic boundaries were observed for another small coastal cetacean in a different ocean basin. The coastal franciscana dolphin (*Pontoporia blainvillei*) displays patterns of population structure that overlap environmental boundaries along its southern distribution range in Argentina (Mendez *et al.* 2010). At a smaller geographical scale in eastern Australian waters, Möller and colleagues postulate that habitat differences could promote localized differentiation of Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) (Möller *et al.* 2007). Alternatively or in addition to environmental factors, social behaviour leading to distinct philopatry has been related to significant levels of population structure in Indo-Pacific bottlenose dolphins in Shark Bay, Australia (Krützen *et al.* 2004; Frère *et al.* 2010). Specialized foraging behaviour has been assumed to drive population structure in odontocete cetaceans with sympatric distributions, particularly in killer whales (*Orcinus orca*) (Ford *et al.* 1998; Matkin *et al.* 2007; Riesch *et al.* 2012). Generally speaking, it has been suggested that local adaptation could be a significant force shaping the observable patterns of genetic divergence (Nosil *et al.* 2008, 2009a,b). It is expected that the same mechanisms that drive population structure, if sustained over a considerable number of generations, also contribute to the development of species boundaries. We therefore postulate that distribution patterns and environmental and behavioural processes probably play significant roles in the emergence of humpback dolphin species.

Taxonomic implications

One of our main goals was to evaluate the potential existence of previously undetected species units based on the phylogenetic and biological species concepts, and for this, we base our assertions on the evidence of differentiation (genetic and morphological) and the potential for reproductive isolation between such units as proxies for phylogenetic and biological species, respectively. While evaluating levels of differentiation is straightforward with the genetic and morphological data and analytical approaches, evaluating reproductive isolation is extremely difficult if not impractical for wild species. Therefore, we work under the assumption that

there is maximum potential for reproductive isolation between genetically differentiated and allopatric units and that such reproductive isolation is less likely between genetically differentiated but fully sympatric units, unless such units depict strong behavioural differences. Under our proposed framework, we suggest that species designations should be considered for revision when there is strong indication of phylogenetic and biological species, as these two aspects incorporate notions of evolutionary significance captured in the genetic and morphological data, and issues of ecological relevance as are the potential sharing of habitat and resources in sympatry.

As studies examining the morphological and genetic variability of humpback dolphins have accumulated, our view of *Sousa* as one of the most highly variable and locally adapted genera of small cetaceans has begun to emerge. Whereas early morphometric studies of *Sousa* did not show significant partitions between Australia and South-East Asia, our current morphological analysis shows a clear distinction between all sampling units except those in Oman, India and Thailand. In addition, recent studies of genetic variation have been consistent in providing evidence for a species split between specimens from Australia and South-East Asia (Rosenbaum *et al.* 2002; Frère *et al.* 2011), and now we can relate this finding to our study taking into account specimens from Africa, which are divergent from all other sampling areas studied to date. This highlights the importance of comprehensive data sets and multiple lines of evidence for a better understanding of divergence patterns. In particular, for taxonomic assessments, genetic and morphological data are suggested as minimum evidence to resolve species boundaries, as morphological traits are typically more conserved than molecular characters and therefore cannot resolve patters that are clear with genetic information (Reeves *et al.* 2004).

With completion of the current study, simultaneously using both morphometric and molecular markers, and incorporating samples from a greater portion of the range of the genus compared with previous efforts, we now have a clearer view of species-level taxonomy within the *Sousa* genus. Independent lines of evidence (mitochondrial and nuclear DNA and morphology) collectively and consistently support our results: clusters from West Africa, Southeast Africa, Arabia (Oman), Bangladesh, Thailand, China and Australia are clearly distinct by their genetic data, and only those from Arabia-Oman, India and Thailand are lumped together by the morphological data. These combined results suggest the existence of at least five assemblages that are evolutionarily unique: (i) West Africa; (ii) Southeast Africa; (iii) one unresolved cluster consisting of populations in Oman, India and Thailand; (iv) China; and (v) Australia. Of these assemblages, there is only confirmed potential for exchange between Southeast Africa and the Arabia samples, as evidenced in our data set by the mtDNA haplotype from Oman clustering with those from Southeast Africa. Lastly, the mtDNA and nuDNA PAA shows the Thailand and China haplotypes as one assemblage and diagnosable from all other regional haplotypes or assemblages. Therefore, taking into consideration both notions of differentiation and potential for reproductive exchange, we propose the recognition of at least four species: *S. teuszii* (in the eastern Atlantic off West Africa), *S. plumbea* (in the central and western Indian Ocean, here encompassing samples from Southeast Africa to Arabia), *S. chinensis* (in the eastern Indian and West Pacific Oceans) and an as-yet-unnamed species off northern Australia (and probably including New Guinea). In the case of *S. teuszii* and the Australian group, allopatry is the most parsimonious explanation for the extreme divergence we observe, as there is no evidence of exchange or contact between these units and any other regional groups of humpback dolphins. In contrast, *S. plumbea* and *S. chinensis* are sympatric from central eastern India to at least Myanmar, where their genetic differences are still extreme and the morphometric characters we study are more homogeneous. Interestingly, these proposed species are notably differentiated by the presence of a dorsal bump in *S. plumbea* and the absence of it in *S. chinensis* and also be a distinct coloration between them. In this case of sympatric speciation, we postulate that behavioural clues probably associated with such morphological differences play an important role in reproductive isolation.

The potential for further differentiation of a Thailand–Bangladesh assemblage, as preliminarily suggested by our molecular data analyses, and the potential for reproductive isolation between Southeast Africa and Arabia are two issues that merit further scrutiny and might result in additional units worth considering as potential species within the genus.

Conservation implications

Knowledge of distinct species or populations, and of evolutionary or ecological drivers of divergence, enhances our ability to design and implement conservation strategies by identifying biologically meaningful conservation units. One of the main implications of species-level conservation units is that they do not entail reproductive exchange and therefore cannot be 'rescued' by migration from other such units in the event of extinction. Although population units generally do exchange migrants that could potentially supplement those units threatened with extinction, evidence of very strong population structure in humpback dolphins indi-

cates that migration events are either very infrequent or may no longer occur.

The evidence we present suggests the existence of at least four species units that must be conserved separately based on threats and challenges idiosyncratic to each. Furthermore, we suggest that attention also be paid to the evolutionary and ecological uniqueness of populations that are clearly divergent, such as those in the central and western Indian Ocean. To continue filling taxonomic gaps of *Sousa* in this geographical region, we recommend increased and targeted sampling efforts and additional analyses of multiple lines of evidence of their evolutionary uniqueness.

Acknowledgements

We thank the Sackler Institute of Comparative Genomics, and the Lewis B. and Dorothy Cullman Program in Molecular Systematics at AMNH for support in the laboratory and data analysis. Specifically, we thank Matt Leslie, Scott Glaberman, Anna Rothschild, Carl Gaines, Jacqueline Loo and Rebecca Hersch for their considerable efforts in the laboratory. K. Van Warebeek assisted with morphometric measurements, and we thank K. Kaliontzopoulou for additional discussion on morphometric analysis. Funding for sample collection in Tanzania was provided by WWF Sweden and Sida/Sarec. Sample collection in Zanzibar was facilitated by the Institute of Marine Sciences, University of Dar es Salaam. Funding for biopsy sampling in Australia was provided by the Sea World Research and Rescue Foundation, WWF Australia, Australian Marine Mammal Centre and Ocean Park Conservation Foundation. Funding for the morphometric and molecular work in this article was mainly provided by the Ocean Park Conservation Foundation, and Agriculture, Fisheries and Conservation Department of the Hong Kong Government, and numerous donors to HCR. This manuscript benefited greatly from feedback provided by the Subject Editor Dr. Stephen R. Palumbi and three anonymous reviewers.

References

Adams LD, Rosel PE (2006) Population differentiation of the Atlantic spotted dolphin (*Stenella frontalis*) in the western North Atlantic, including the Gulf of Mexico. *Marine Biology*, **148**, 671–681.

Alberts B, Wilson JH, Hunt T (2008) *Molecular Biology of the Cell*, 5th edn. Garland Science, New York.

Andrews K, Karczmarski L, Au W *et al.* (2010) Rolling stones and stable homes: social structure, habitat diversity and population genetics of the Hawaiian spinner dolphin (*Stenella longirostris*). *Molecular Ecology*, **19**, 732–748.

Baker CS, Perry A, Bannister JL *et al.* (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proceedings of the National Academy of Sciences, USA*, **90**, 8239–8243.

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

Beasley I, Arnold P, Heinsohn G (2002) Geographical variation in skull morphology of the Irrawaddy Dolphin, Orcaella brevirostris (Owen in Gray, 1866). *Raffles Bulletin of Zoology* **10**, 15–34.

Beasley I, Robertson KM, Arnold P (2005) Description of a new dolphin, the Australian snubfin dolphin *Orcaella heinsohni* sp. nov. (Cetacea, Delphinidae). *Marine Mammal Science*, **21**, 365–400.

Caballero S, Trujillo F, Vianna JA *et al.* (2007) Taxonomic status of the genus *Sotalia*: species level ranking for "tucuxi" (*Sotalia fluviatilis*) and "costero" (*Sotalia guianensis*) dolphins. *Marine Mammal Science*, **23**, 358–386.

Caballero S, Jackson J, Mignucci-Giannoni AA *et al.* (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.

Cassens I, Van Waerebeek K, Best PB *et al.* (2003) The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): a critical examination of network methods and rooting procedures. *Molecular Ecology*, **12**, 1781–1792.

Charlton-Robb K, Gershwin LA, Thompson R *et al.* (2011) A new dolphin species, the Burrunan dolphin *Tursiops australis* sp. nov., endemic to Southern Australian coastal waters. *PLoS ONE*, **6**, e24047.

Chivers SJ, Leduc RG, Robertson KM, Barros NB, Dizon AE (2005) Genetic variation of *Kogia* spp. with preliminary evidence for two species of *Kogia sima*. *Marine Mammal Science*, **21**, 619–634.

Cracraft J (1983) Species concepts and speciation analysis. In: *Current Ornithology* (ed. Ereshefsky M) pp. 159–187. Plenum Press, New York.

Davis JI, Nixon KC (1992) Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology*, **41**, 421–435.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Escorza Trevino S, Dizon AE (2000) Phylogeography, intraspecific structure and sex-biased dispersal of Dall's porpoise, *Phocoenoides dalli*, revealed by mitochondrial and microsatellite DNA analyses. *Molecular Ecology*, **9**, 1049–1060.

Excoffier L, Lischer HE (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 (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Ford JKB, Ellis GM, Barrett-Lennard LG *et al.* (1998) Dietary specialization in two sympatric populations of killer whales (*Orcinus orca*) in coastal British Columbia and adjacent waters. *Canadian Journal of Zoology*, **76**, 1456–1471.

Frère CH, Hale PT, Porter L, Cockcroft VG, Dalebout ML (2008) Phylogenetic analysis of mtDNA sequences suggests revision of humpback dolphin (*Sousa* spp.) taxonomy is needed. *Marine and Freshwater Research*, **59**, 259–268.

Frère CH, Krützen M, Mann J *et al.* (2010) Home range overlap, matrilineal and biparental kinship drive female associations in bottlenose dolphins. *Animal Behaviour*, **80**, 481–486.

Frère C, Seddon J, Palmer C, Porter L, Parra G (2011) Multiple lines of evidence for an Australasian geographic boundary in the Indo-Pacific humpback dolphin (*Sousa chinensis*): population or species divergence? *Conservation Genetics*, **12**, 1633–1638.

Futuyma DJ (2005) *Evolution*, 1st edn. Sinauer Associates, Sunderland, Massachusetts.

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.

Gatesy J, O'Leary MA (2001) Deciphering whale origins with molecules and fossils. *Trends in Ecology & Evolution*, **16**, 562–570.

Goldstein PZ, DeSalle R, Amato G, Vogler AP (2000) Conservation genetics at the species boundary. *Conservation Biology*, **14**, 120–131.

Holland B, Moulton V (2003) Consensus networks: a method for visualizing incompatibilities in collections of trees. In: *Proceedings of Workshop on Algorithms in Bioinformatics* (eds Benson G, Page R), pp. 165–176. Springer, Budapest.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jefferson TA (2004) Special Issue: biology and conservation of humpback dolphins (*Sousa* spp.). *Aquatic Mammals*, **30**, 1–2.

Jefferson TA, Hung SK (2004) A review of the status of the Indo-Pacific humpback dolphin (*Sousa chinensis*) in Chinese waters. *Aquatic Mammals*, **30**, 149–158.

Jefferson TA, Van Waerebeek K (2004) Geographic variation in skull morphology of humpback dolphins (*Sousa* spp.). *Aquatic Mammals*, **30**, 3–17.

Krause J, Unger T, Noçon A *et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology*, **8**, 220.

Krützen 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.

Lanave C, Preparata G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution*, **20**, 86–93.

Lefebure T, Douady CJ, Gouy M, Gibert J (2006) Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, **40**, 435–447.

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

Lyons LA, Laughlin TF, Copeland NG *et al.* (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics*, **15**, 47–56.

Manly BFJ (2005) *Multivariate Statistical Methods: A primer*, 3rd edn. Chapman & Hall/CRC, Boca Raton, Florida.

Matkin CO, Barrett-Lennard LG, Yurk H, Ellifrit D, Trites AW (2007) Ecotypic variation and predatory behavior among killer whales (*Orcinus orca*) off the eastern Aleutian Islands, Alaska. *Fishery Bulletin*, **105**, 74–87.

Mendez M, Rosenbaum HC, Bordino P (2008) Conservation genetics of the franciscana dolphin in Northern Argentina: population structure, by-catch impacts, and management implications. *Conservation Genetics*, **9**, 419–435.

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.

Mendez M, Subramaniam A, Collins T *et al.* (2011) Molecular ecology meets remote sensing: environmental drivers to population structure of humpback dolphins in the Western Indian Ocean. *Heredity*, **107**, 349–361.

Möller LM, Wiszniewski J, Allen SJ, Beheregaray LB (2007) Habitat type promotes rapid and extremely localised genetic differentiation in dolphins. *Marine and Freshwater Research*, **58**, 640–648.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA*, **76**, 5269–5273.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009a) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Nosil P, Harmon LJ, Seehausen O (2009b) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.

O' Corry-Crowe G, Suydam R, Rosenberg A, Frost K, Dizon A (1997) Phylogeography, population structure and dispersal patterns of the beluga whale *Delphinapterus leucas* in the western Nearctic revealed by mitochondrial DNA. *Molecular Ecology*, **6**, 955–970.

O'Leary MA, Gatesy J, Novacek MJ (2003) Are the dental data really at odds with the molecular data? Morphological evidence for whale phylogeny (re)reexamined. *Systematic Biology*, **52**, 853–864.

Oviedo L (2007) Dolphin sympatric ecology in a tropical ford: habitat bathymetry and topography as a strategy to coexist. *Journal of the Marine Biological Association of the United Kingdom*, **87**, 1327–1335.

Palsbøll PJ, Bérubé M, Aguilar A, Notarbartolo-Di-Sciara G, Nielsen R (2004) Discerning between recurrent gene flow and recent divergence under a finite-site mutation model applied to North Atlantic and Mediterranean Sea fin whale (*Balaenoptera physalus*) populations. *Evolution*, **58**, 670–675.

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, **11**, 426–435.

Palumbi SR, Cipriano F (1998) Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation. *Journal of Heredity*, **89**, 459–464.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A (2010) How many bootstrap replicates are necessary? *Journal of Computational Biology*, **17**, 337–354.

Pearse DE, Crandall KA (2004) Beyond F-ST: analysis of population genetic data for conservation. *Conservation Genetics*, **5**, 585–602.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution*, **16**, 37–45.

Reeves RR, Perrin WF, Taylor BL, Baker CS, Mesnick SL (eds) (2004) Report of the Workshop on Shortcomings of Cetacean Taxonomy in Relation to Needs of Conservation and Management (April 30-May 2, 2004, La Jolla, California). In: *NOAA Technical Memorandum NOAA-NMFS-SWFSC-363*, p. 94. National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Fisheries Science Center, La Jolla, California.

Rice DW (1977) A list of the marine mammals of the world. In: *NOAA technical report NMFS SSRF 711*, p. 15. National Oceanic and Atmospheric Administration, Seattle, Washington.

Rice D (1998) *Marine Mammals of the World: Systematics and Distribution*, 1st edn The Society for Marine Mammalogy, Lawrence, Kansas.

Riesch R, Barrett-Lennard LG, Ellis GM, Ford JKB, Deecke VB (2012) Cultural traditions and the evolution of reproductive isolation: ecological speciation in killer whales? *Biological Journal of the Linnean Society*, **106**, 1–17.

Roca A, Ishida Y, Nikolaidis N *et al.* (2009) Genetic variation at hair length candidate genes in elephants and the extinct woolly mammoth. *BMC Evolutionary Biology*, **9**, 232.

Rosenbaum H, Brownell R, Brown M *et al.* (2000) Worldwide genetic differentiation of Eubalaena: questioning the number of right whale species. *Molecular Ecology*, **9**, 1793–1802.

Rosenbaum HC, Glaberman S, Jefferson T *et al.* (2002) Phylogenetic relationships and population structure among humpback dolphins based on mtDNA variation. In: *International Whaling Commission Scientific Committee Report (IWC/54/SM34)*, p. 8. [Unpublished].

Sarkar IN, Thornton JW, Planet PJ *et al.* (2002) An automated phylogenetic key for classifying homeoboxes. *Molecular Phylogenetics and Evolution*, **24**, 388–399.

Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. *Systematic Biology*, **56**, 896–906.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Stamatakis A, Ott M (2008) Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **363**, 3977–3984.

Sudarto, Lalu XC, Kosen JD *et al.* (2010) Mitochondrial genomic divergence in coelacanths (*Latimeria*): slow rate of evolution or recent speciation? *Marine Biology*, **157**, 2253–2262.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Vogler A, DeSalle R (1994) Diagnosing units of conservation management. *Conservation Biology*, **8**, 354–363.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Welch AJ, Yoshida AA, Fleischer RC (2011) Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. *Molecular Ecology*, **20**, 1364–1377.

Yang ZH (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites – approximate methods. *Journal of Molecular Evolution*, **39**, 306–314.

H.C.R. and T.A.J. initially conceived the project; M.M. and H.C.R. designed the present research; T.A.J., M.K., G.J.P., T.C., G.M., R.B., P.B., A.S., O.A.A., V.M.P., L.K., A.G., B.S., D.S. and H.C.R. performed field and museum sampling; H.C.R. and G.A. contributed reagents and analytical tools; M.M. and S.O.K. analysed the data; M.M. wrote the manuscript, and all authors contributed to manuscript revisions.

Data accessibility

All data (morphology, mtDNA alignment, nuDNA alignment) analysed in this manuscript are available as Supporting Information.

Supporting information

Additional supporting information may be found in the online version of this article.

Data S1 Supplementary Sequence files: Alignments of mitochondrial and nuclear DNA.

Movie S1 Dynamic view of the DFA, to bettwer show the differences between putative units.

Table S1 Sample sizes for the phylogenetic analyses using mtDNA, nuDNA, and concatenated mtDNA-nuDNA, the population level analyses using mtDNA data and for the morphologic analysis using cranial morphology data.

Table S2 Morphological dataset.