Molecular Ecology (2003) **12**, 1781–1792 doi: 10.1046/j.1365-294X.2003.01876.x

The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): a critical examination of network methods and rooting procedures

I. CASSENS,* K. VAN WAEREBEEK,§ P. B. BEST,† E. A. CRESPO,¶ J. REYES‡ and M. C. MILINKOVITCH*

**Unit of Evolutionary Genetics, Institute of Molecular Biology and Medicine, Free University of Brussels, rue Jeener & Brachet 12, 6041 Gosselies, Belgium, §Peruvian Centre for Cetacean Research (CEPEC), Jorge Chavez 302, Pucusana, Lima-20, Peru, †Mammal Research Institute, University of Pretoria, Pretoria, 0002 South Africa, ¶Laboratorio de Mamiferos Marinos, Centro Nacional Patagonico, Boulevard Brown S/N, Puerto Madryn, Chubut, Argentina, ‡Areas Costeras y Recursos Marinos (ACOREMA), Avenue, San Martin 1471, Pisco, Peru*

Abstract

We investigated the phylogeography and evolutionary history of dusky dolphins (*Lagenorhynchus obscurus*) using DNA sequences of the full mitochondrial cytochrome *b* gene in 124 individuals from the putative stocks off Peru, Argentina and Southwest Africa. While genetic differentiation within oceans is surprisingly low, there is no evidence for recent female gene flow between Atlantic and Pacific waters. Highest genetic variability in terms of sequence divergence and number of haplotypes is found in the Atlantic. Our analyses also indicate that the eastern South Pacific dusky dolphins stock should be considered a separate management unit. Given the high level of mortality experienced by the Peruvian dusky dolphin in local fishery activities, these findings have important implications for an objective management of the species. Furthermore, we analysed our mitochondrial sequence data with several widely used network estimation and rooting methods. The resulting intraspecific gene genealogies and rooting inferences exhibited substantial differences, underlying the limitations of some algorithms. Given that scientific hypotheses and management decisions depend strongly on inferred tree or network topologies, there is a clear need for a systematic comparative analysis of available methods. Finally, the present study indicates that (i) the dusky and the Pacific white-sided dolphins are sister species and (ii) not only the Westwind Drift hypothesis but also other models of dispersion are compatible with the current geographical distribution of dusky dolphins.

Keywords: Cetacea, conservation genetics, mtDNA, network methods, phylogeography, rooting techniques

Received 18 December 2002; revision received 26 February 2003; accepted 27 March 2003

Introduction

The dusky dolphin *Lagenorhynchus obscurus* (Gray 1828) is distributed in cool temperate waters of the Southern Hemisphere. Its occurrence is well documented along the coasts off Southwest Africa, Argentina, Chile, Peru, New Zealand and, recently, Australia (Van Waerebeek, pers. comm.). Although the species has been positively recorded from the vicinity of many oceanic island groups (e.g. Campbell, Auckland, Chatham islands in the western South Pacific; Gough and Falkland islands in the South Atlantic; Amsterdam and Prince Edward islands in the Indian Ocean), its pelagic occurrence in the southern oceans remains unconfirmed (Van Waerebeek *et al*. 1995). Distributional information and morphological studies, which have revealed significant differentiation in cranial characteristics and body size among geographical regions (Van Waerebeek 1993a,b), led authors to propose a disjunct distribution with discrete dusky dolphin stocks confined to continental shelves and islands (Van Waerebeek *et al*. 1995). However, more data are needed in order to precisely describe population boundaries, e.g. Argentinean individuals have not yet been included in comparative analyses

Correspondence: Michel C. Milinkovitch. Fax: 00332 2 650 9956; E-mail: mcmilink@ulb.ac.be (www.ulb.ac.be/sciences/ueg)

(Van Waerebeek 1993b; Van Waerebeek *et al*. 1993) and little is known on how they relate to other dusky dolphin populations.

Moreover, a better understanding of dusky dolphin population structure is of major importance with respect to species conservation. Especially in Peruvian fisheries dusky dolphins are, along with other small cetacean species, subjected to high mortality levels from by-catches and direct take (Van Waerebeek & Reyes 1990; Van Waerebeek & Reyes 1994a,b; Van Waerebeek 1994; Van Waerebeek *et al*. 1997). Although takes of cetaceans have been banned since 1990, the species continue to be exploited for human consumption and increasingly for bait in shark fisheries. Extensive monitoring of fish landing sites over the past two decades revealed a decrease in the proportional representation of dusky dolphins in cetacean catches off central Peru (Van Waerebeek & Reyes 1994a; Van Waerebeek 1994; Van Waerebeek *et al*. 1997), which could reflect a severe decline in abundance due to fishery activities. However, for assessing objectively the impact of this exploitation on the species, stock identity needs to be determined.

The investigation of a species population structure, phylogeography and evolutionary history, however, most often requires the estimation and polarization of gene genealogies. At the population level, phylogenetic networks are more convenient than strictly hierarchical trees to represent relationships among closely related sequences because the former allow the display of all equally parsimonious hypotheses (i.e. ambiguous relationships) on a single figure. Furthermore, network construction methods incorporate specifically the possibility for the persistence of many ancestral haplotypes in the population and for the occurrence of recombination events. A recent review described the large diversity of methods available currently for network construction (Posada & Crandall 2001). Although some of these methods are used widely in population genetic studies, little work has been conducted to test the reliability of different algorithms. Similarly, estimating the relative age of haplotypes in intraspecific genealogies has proved to be a highly problematic task for two reasons. First, rooting analysis using an outgroup is challenging. Indeed, the likelihood that molecular character states shared by one taxon and the outgroup will be based on random similarity rather than on history increases with increasing divergence between the outgroup and the ingroup taxa (Wheeler 1990; Templeton 1992, 1993; Milinkovitch *et al*. 1996). Furthermore, reproductively isolated populations do not always form reciprocally monophyletic groups (e.g. (Avise 1994; Milinkovitch *et al*. 2002)). When the ingroup is paraphyletic with respect to the outgroup, the basic assumption of outgroup analysis will be violated and the polarization of the ingroup genealogy will be in error (Templeton 1992, 1993; Milinkovitch *et al*. 2002). Second, coalescence theory predicts that, on average, the most ancestral haplotype in a sample is the most frequent one and displays the highest number of connections to other haplotypes (Donnelly & Tavaré 1986). However, existing algorithms (e.g. Castelloe & Templeton 1994) that use these predictions for estimating root probabilities on each haplotype have been barely tested. Given that hypotheses often depend strongly on the polarization of the inferred trees or network topologies, there is a clear need for the identification of potential artefacts and systematic biases of the different algorithms available. In the study reported here, information on the relative age of the sampled haplotypes is indeed necessary for studying the evolutionary history of dusky dolphins. The most likely sister species of the dusky dolphin, the Pacific white-sided dolphin (*Lagenorhynchus obliquidens*, Gill 1865) is found in North Pacific waters (Van Waerebeek 1993b; LeDuc *et al*. 1999). In agreement with the antitropical distribution of *L. obscurus*/*L. obliquidens* and morphological studies, it has been suggested that the Peruvian dusky dolphin represents the most basal lineage within the species, while the other populations off Argentina, Southwest Africa and New Zealand would have originated successively through dispersion via the eastflowing Westwind Drift current (Van Waerebeek 1993b). To date, this hypothesis has not been tested using molecular markers.

In this study, we report on DNA sequence analyses of the full mitochondrial (mt) cytochrome *b* gene in 124 dusky dolphins from the putative stocks off Peru, Argentina and Southwest Africa, allowing us to address four main issues. First, the genetic population structure of the species is assessed. Second, considering that the species suffers considerable mortality off Peru and Chile, the mt sequence data allow for a first identification of management units in dusky dolphins. Third, we compare several network construction and rooting methods for estimating the ancestor–descendent relationships among haplotypes, and discuss the conflicting results. Fourth, the probable geographic origin of the dusky dolphin lineage and the mode of its subsequent dispersion are discussed based on the rooted intraspecific genealogy, together with estimates of genetic variation and its geographical distribution.

Materials and methods

Data collection

Skin, liver and bone samples were collected from stranded, incidentally or directly caught dusky dolphin individuals (*L. obscurus*) from Pacific waters off Peru ($n = 78$) as well as from the Atlantic populations off Argentina ($n = 19$) and South Africa ($n = 27$). Additionally, three biopsy samples from the Pacific white-sided dolphin (*L. obliquidens*) were included in the analyses. Genomic DNA was extracted following standard proteinase digestion and phenol–chloroform

extraction procedures (Hillis *et al*. 1996). The bone samples were treated first with emery paper on the outer surface in order to reduce contamination sources, and then powdered using liquid nitrogen before adding the digestive solution. The full cytochrome *b* gene [1140 base pairs (bp)] was amplified and sequenced directly (BigDye Terminator Cycle Sequencing; Applied Biosystems) on both strands (two pairs of primers for polymerase chain reaction (PCR) and sequencing from Cassens *et al*. (2000), on an ABI 377 automated sequencer (Applied Biosystems). For outgroup rooting, full cytochrome *b* sequences from other delphinid species were downloaded from GenBank and added to the data set: *L. obliquidens* (AF084067), *L. cruciger* (AF084068), *L. australis* (AF084069), *Cephalorhynchus eutropia* (AF084072), *Lissodelphis peronii* (AF084065) and *Delphinus delphis* (AF084084).

Genetic diversity and population structure

Levels of genetic variation within the Peruvian, Argentinean and South African dusky dolphin populations were measured in terms of haplotype diversity (H) and nucleotide diversity (π_n) (Nei 1987). The significance of these statistics among populations was tested applying the parametric GT2-method (Hochberg 1974). The geographical differentiation of haplotypes was quantified using a hierarchical analysis of variance (AMOVA; (Excoffier *et al*. 1992)). We assigned portions of total genetic variation to divergence either among oceans (Atlantic, Pacific), among populations within oceans (South African, Argentinean waters within the Atlantic) or within populations. This approach incorporates information on the absolute number of differences among haplotypes as well as on haplotype frequencies. The significance of variance components and *F*-statistic analogues, designated Φ-statistics, was tested by multiple (1000) random permutations. All estimations were performed using ARLEQUIN, version 2.000 (Schneider *et al*. 2000).

Network estimation

Intraspecific gene genealogies were inferred using four different network construction methods, all of them implemented in freely available software packages.

First, the method of split decomposition ((Bandelt & Dress 1992); SPLITSTREE, VERSION 2.4 (Huson 1998)) produces a split graph representing all 'weakly compatible' (Huson 1998) partitions of sequences. Conflicts among partitions are represented in the form of reticulations. Because the method demonstrates principally the amount of homoplasy in the data set, it is indicative of how 'tree-like' the network of interest is. Hence, in population genetic studies, this approach might be less suitable, especially when large data sets with many taxa are analysed.

Second, the algorithm for constructing minimum spanning trees (MSTs) from a matrix of pairwise distances (absolute number of differences) among haplotypes (Prim 1957; Rohlf 1973) has been modified in order to include all possible MSTs within a single graph, the minimum spanning network (MSN) [(Excoffier & Smouse 1994); ARLEQUIN, version 2.000 (Schneider *et al*. 2000)]. MSTs rarely correspond to maximum parsimony (minimum length) trees. We included this method in our comparative approach because of its implementation in a widely used population genetic program (ARLEQUIN).

Third, the method of statistical parsimony ((Templeton *et al*. 1992); TCS (Clement *et al*. 2000), version 1.14beta provided by D. Posada) defines first the uncorrected distance above which the parsimony criterion is violated with more than 5% probability (parsimony limit). Then, all connections are established among haplotypes starting with the smallest distances and ending either when all haplotypes are connected or the distance corresponding to the parsimony limit has been reached.

Fourth, in the median-joining network approach [(Bandelt *et al*. 1999), NETWORK, version 2.0, available at http://www.fluxus-engineering.com/sharenet.htm], all MSTs are first combined within a single network (MSN) following an algorithm analogous to that proposed by Excoffier & Smouse (1994). Then, using the parsimony criterion, inferred intermediate haplotypes are added to the network in order to reduce overall tree length. In addition, by setting a parameter ε to values > 0, less parsimonious pathways can be included in the inferred network (Bandelt *et al*. 1999). However, it should be noted that such an analysis often yields a graph which is too complex for visualization and data interpretation, especially in the presence of substantial amounts of homoplasy.

In addition to network construction, maximum parsimony (MP) analyses were performed using the program PAUP*4.0b8 (Swofford 2000) on all dusky dolphin haplotypes (heuristic search with TBR branch swapping; 10 000 replicates, stepwise addition starting tree, random sequence addition).

Rooting techniques

Phylogenetic analyses were performed using the program PAUP*4.0b8 (Swofford 2000). Maximum parsimony (MP) and neighbour-joining (NJ) analyses of the Delphinidae (true dolphins), based on the full cytochrome *b* gene (LeDuc *et al*. 1999), strongly support a clade (Lissodelphininae) containing the genera *Cephalorhynchus*, *Lissodelphis* and four of the six species of *Lagenorhynchus*, namely *L. obscurus*, *L. obliquidens*, *L. australis* and *L. cruciger*. As many of the relationships within this group remain largely unresolved, we analysed under MP (heuristic search with TBR branch swapping, 1000 replicates, stepwise addition starting tree, random sequence addition) and NJ (using three different substitution models: Jukes-Cantor, HKY85 and LogDet distances) all dusky dolphin haplotypes together with

homologous sequences from representatives of six species of the above-mentioned genera. The Pacific white-sided dolphin (*L. obliquidens*), probably the sister species of *L. obscurus* (Van Waerebeek 1993b; LeDuc *et al*. 1999), was represented by several individuals. Furthermore, a maximum-likelihood (ML) analysis was carried out with PAUP*4.0b8 (Swofford 2000) on a smaller subset of taxa (proportion of invariant characters, γ shaped distribution parameter and Ti/Tv ratio estimated by ML). In all analyses, *D. delphis* was used as the single outgroup to determine whether *L. obscurus* is monophyletic with respect to closely related species (especially *L. obliquidens*) and to root the dusky dolphin species group.

Alternatively, based on coalescence theory, each haplotype in the statistical parsimony network was assigned a so-called 'outgroup probability' ((Donnelly & Tavaré 1986; Castelloe & Templeton 1994), TCS version 1.13 (Clement *et al*. 2000)). The likelihood is calculated as a function of the position of the haplotype in the network, its frequency and its number of connections with neighbour haplotypes. However, the algorithm as currently implemented in the TCS software ignores connections to unsampled (inferred) haplotypes (K. Crandall & D. Posada, pers. comm.). Therefore, to estimate the most probable location of the root, we computed manually, for each haplotype, the likelihood of rooting considering all connections, i.e. both connections to observed haplotypes and to missing intermediates. More specifically, we increase outgroup probability of an observed haplotype for each additional observed haplotype it is connected to, independently of the connecting branch length. When a missing intermediate is encountered before an observed haplotype, all branches diverging from that missing haplotypes are trailed.

Results

mtDNA sequence diversity and population structure

Sixty variable positions in the mt cytochrome *b* gene define 36 haplotypes in our sample of 124 dusky dolphins (see Appendix 1). All sequences have been deposited in the GenBank database under Accession nos AY257126–AY257161. Each haplotype is restricted to one of the three predefined populations (Peru, Argentina and South Africa) except for haplotype A1.1, that is shared between the two Atlantic populations (Fig. 1). The nucleotide and haplotype diversities exhibited by the Peruvian population are significantly ($P < 0.01$) lower than in the Argentinean and South African populations (Table 1); 50% of the Peruvian individuals share the same haplotype P5.1 (Fig. 1). While pairwise Φ_{ST} values indicate significant differentiation among all three populations of dusky dolphins (Table 1), the AMOVA analysis indicates that most of the population partitioning (about 53%) is attributable to differences between Atlantic and Pacific haplotypes (Table 2). Network reconstruction reveals that this is due largely to three fixed, nonsynonymous substitutions at positions 49, 326, and 611 (cf. underlined numbers in Fig. 1).

Comparison of network estimation methods

All algorithms yield genealogies revealing an unambiguous differentiation between Pacific (Peru) and Atlantic (Argentina, South Africa) haplotypes (Fig. 2A–D). On the other hand, the four methods produce different topologies within these groups. While the split decomposition and the MSN constructions (Fig. 2A,B) exhibit ambiguous relationships

	Peru		Argentina		South Africa	
Peru ($n = 78$)	0.713 (± 0.048)	0.236 (± 0.141)	***	***	***	***
Argentina ($n = 19$)	0.627***		0.930 (± 0.036)	0.442 (± 0.251)	NS	NS
South Africa ($n = 27$)	0.594***		0.065*		0.903 (± 0.032)	0.503 (± 0.277)

NS = $P > 0.05$, * = $P < 0.05$, ***$P < 0.001$, n = sample size.

Table 1 Genetic diversities within the three dusky dolphin populations in terms of haplotype (H, diagonal/left side) and nucleotide diversities (π_n in percentage, diagonal/right side). Standard deviations are given between parentheses. Significance levels of pairwise population comparisons are indicated above diagonal (left and right side for H and π_n, respectively). Genetic differentiation among populations in terms of pairwise Φ_{ST}-values is shown below diagonal

	(df)	Variance components	Variation (%)	Fixation index
Between oceans	1	2.335	52.94	
Among populations/within oceans	1	0.228	5.18	
Within populations	123	1.847	41.88	$\Phi_{ST} = 0.581$***

***$P < 0.001$, probability of having more extreme Φ_{ST}-statistic by chance alone (based on 1000 random permutations).

Table 2 Analysis of variance of pairwise genetic distances among mtDNA cytochrome *b* sequences of dusky dolphins. Oceans: Atlantic, Pacific. Populations: Peru, Argentina, South Africa

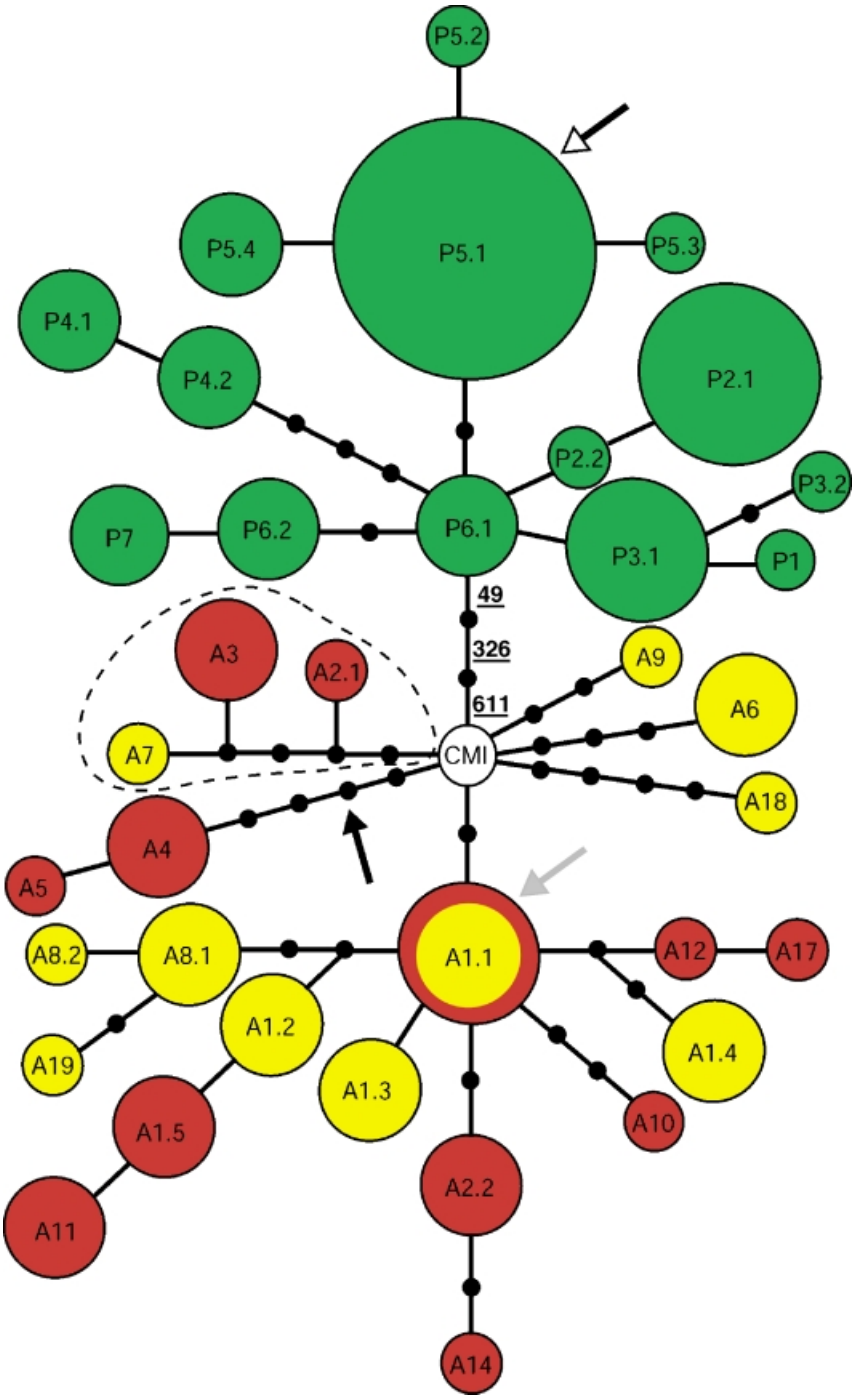

Fig. 1 Median-joining network depicting the phylogenetic relationships among, and geographical assignment of, all dusky dolphin mtDNA haplotypes (filled coloured circles) based on cytochrome *b* DNA sequences: green, Peru; yellow, Argentina; red, South Africa. The size of each circle is proportional to the corresponding haplotype frequency. Missing intermediates are indicated by black circles except for the most central missing intermediate, noted by 'CMI'. Each branch between two (sampled or missing) haplotypes indicates a single mutational step. The three underlined numbers indicate fixed substitutions separating Atlantic and Pacifc haplotypes. Arrows indicate alternative rooting positions: black arrow, outgroup rooting; empty arrow, coalescence/TCS analysis; grey arrow, coalescence/our alternative method (cf. Material and methods). Note that the latter two rooting locations have been inferred using TCS, but transferred to the median-joining network for comparison.

among haplotypes (i.e. reticulations), the statistical parsimony and median-joining (with $\varepsilon = 0$) approaches produce nonreticulated networks (Fig. 2C,D). Three of the four methods infer a missing intermediate haplotype (indicated by arrows in Fig. 2), to which some of the Argentinean and South African haplotypes are connected by long branches. On the other hand, the MSN approach (Fig. 2B) yields a star-like phylogeny. The two unreticulated networks differ by the position of a single branch (circled in Fig. 2C,D) grouping three sampled haplotypes. The MP approach yields six trees of size 71. The minimum number of mutations necessary to explain the data given the genealogies produced by the MSN, statistical parsimony and median-joining approaches are 85, 73 and 71, respectively. Hence, under the strict parsimony criterion, the median-joining network method yields the best genealogy. Increasing the value of the parameter ε in the median-joining procedure increases the number of reticulations in

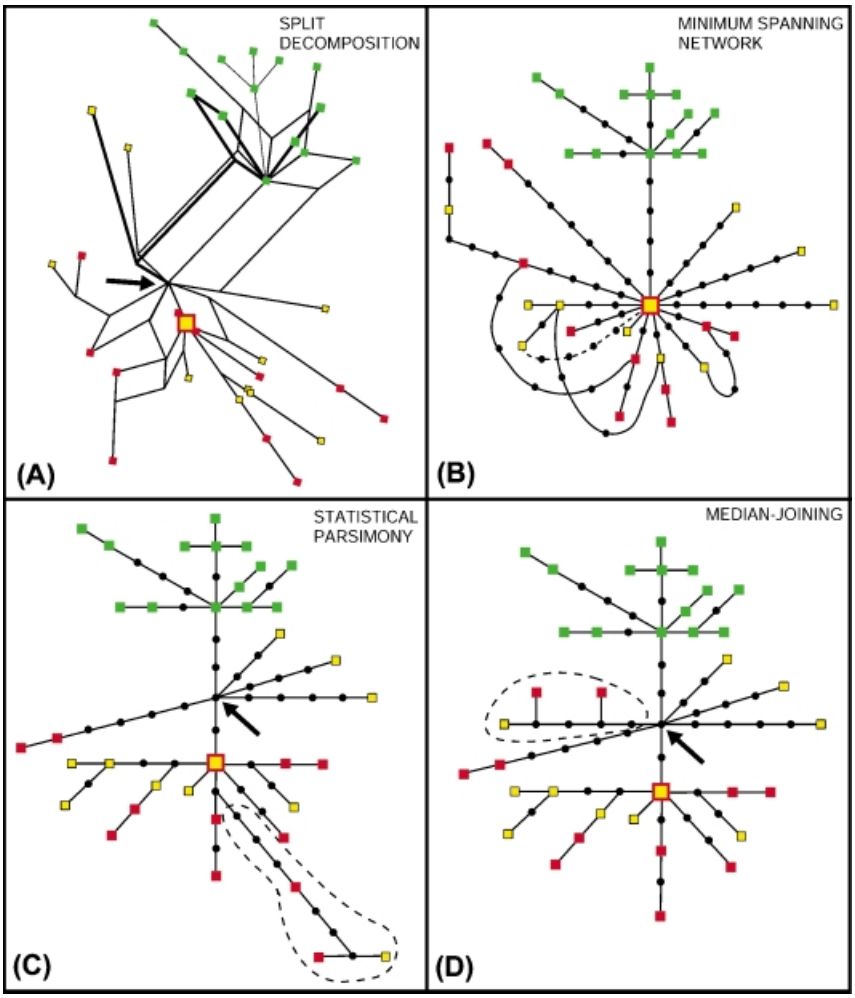

Fig. 2 Cytochrome *b* gene genealogies of dusky dolphins using four different network construction methods: (A) split decomposition [(Bandelt & Dress 1992); SPLITSTREE, version 2.4 (Huson 1998), both Jukes-Cantor and Hamming distances yielded the same topology]; (B) minimum spanning network [(Excoffier & Smouse 1994); ARLEQUIN, version 2.000 (Schneider *et al*. 2000)]; (C) statistical parsimony [(Templeton *et al*. 1992); TCS (Clement *et al*. 2000), version version 1.14b]; (D) median-joining network [(Bandelt *et al*. 1999), NETWORK, version 2.0, available at http://www.fluxus-engineering.com/sharenet.htm; with ε = 0]. Black circles indicate missing intermediates; coloured squares indicate observed haplotypes from three geographical regions: green, Peru; yellow, Argentina; red, South Africa. Each branch represents a single mutational step. The direct five-step link between A1.1 and A19 (in B), which is not consistent with the MSN algorithm, is shown as a dotted line (see text for further explanation). Networks shown in (C) and (D) differ from each other by the placement of the lineage circled with a dotted line.

the resulting topologies (data not shown). With ε = 1, the separation between Pacific and Atlantic haplotypes is still detectable, but 17 loops (16 among Atlantic haplotypes) allow for a very large number of alternative pathways within the network (including the less parsimonious solution yielded by the statistical parsimony approach).

Rooting of the intraspecific genealogy

All outgroup analyses (under NJ, MP and ML) confirm the sister relationship between the Pacific white-sided and dusky dolphins and indicate their reciprocal monophyly (Fig. 3). More surprisingly, the dusky dolphin clade is rooted consistently within a group of Atlantic haplotypes, most often along a branch leading to two South African samples (cf. black arrow in Fig. 1 and Fig. 3). This result suggests that Peruvian dusky dolphins form a monophyletic group nested among Atlantic haplotypes. On the other hand, rooting the network on the most ancient haplotype (as estimated by the algorithm implemented in TCS) groups the Atlantic haplotypes into a clade nested among the Peruvian lineages (Fig. 1, empty arrow). When all connections (both to sampled haplotypes and to missing intermediates) branching from interior haplotypes are included in the calculations (cf. Material and methods), the highest rooting probability is assigned to the only haplotype found in both of the Atlantic populations (Fig. 1, grey arrow). Indeed, if we assume equal frequencies for all haplotypes, the weights calculated for P5.1 and A1.1 amount to 5 (its own frequency plus 1 for each of the connections to P5.2, P5.3, P5.4, and P6.1) and 14, respectively, rather than 4 and 2 under the algorithm implemented in TCS.

Discussion

Population structure and management implications

Dusky dolphins occur in all three oceans of the Southern Hemisphere (Rice 1998) but appear to be restricted to continental shelves and the vicinity of oceanic islands (Van Waerebeek 1992; Van Waerebeek *et al*. 1995; Crespo *et al*. 1997a). While long-distance movements have been recorded

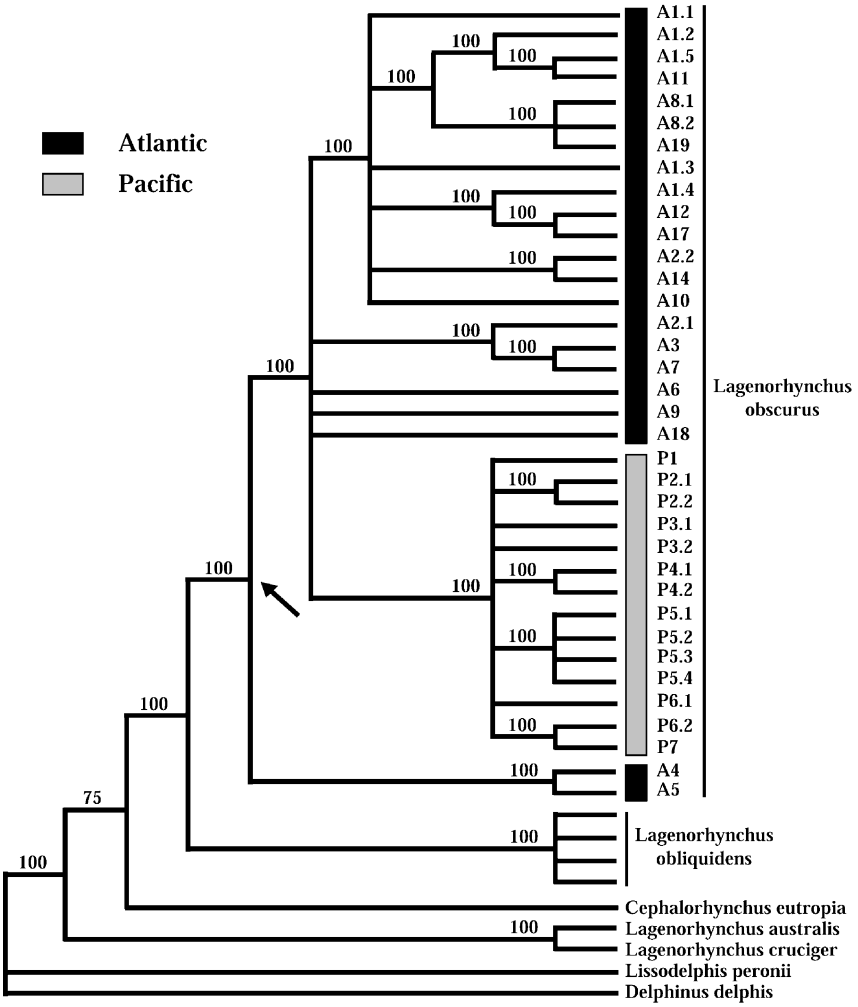

Fig. 3 Fifty per cent majority-rule consensus among the 48 maximum parsimony trees inferred from cytochrome *b* DNA sequences of dusky dolphins (all 36 haplotypes) and eight other Lissodelphininae (including four *L. obliquidens* individuals). *Delphinus delphis* is used as an outgroup. Numbers above branches show frequency of bipartitions observed among the 48 MP trees.

over coastal habitat off Argentina (Crespo *et al*. 1997a), the occurrence of dusky dolphins in deep, far offshore waters has not been confirmed (Van Waerebeek 1993b; Van Waerebeek *et al*. 1995). Surprisingly, our analyses did not indicate a pronounced differentiation between Argentinean and Southwest African populations. Although Φ_{ST} values indicated some differentiation, the two geographical groups share a common haplotype. Furthermore, our analyses reveal a matrilineal genealogy with no obvious geographical partitioning between the eastern and western Atlantic (Fig. 1). This result suggests that either the two populations have not been isolated long enough to be genetically differentiated or that recent gene flow has occurred with individuals occasionally crossing pelagic waters. On the other hand, our phylogeographic analyses indicate a clear separation between a Pacific and an Atlantic lineage, a result consistent with significant differences found in body size (Van Waerebeek 1993a), cranial characters (Van Waerebeek 1993b) and roundworm load (*Crassicauda* sp.) (Van Waerebeek *et al*. 1993). Furthermore, the population off Peru appears to exhibit reproductive segregation with shifted breeding seasons in comparison to the Argentine population (Van Waerebeek & Read 1994). Hence, our mtDNA analyses, in concordance with morphological and other nonmolecular findings, suggest strongly that dusky dolphins from the eastern South Pacific should be considered a discrete management unit (or 'stock') that is isolated from Atlantic populations. This result is of special concern when considered in perspective with previously published field studies. While catches of dusky dolphins in Argentinean waters are mostly rare and accidental (Crespo *et al*. 1994; Crespo *et al*. 1997b; Dans *et al*. 1997), extensive and directed captures off Peru and, to a lesser degree, off Chile are well documented (Van Waerebeek & Reyes 1990, 1994a,b; Van Waerebeek *et al*. 1997). The proportion of dusky dolphins in Peruvian catch statistics has declined continuously in the last two decades (Van Waerebeek & Reyes 1994a; Van Waerebeek 1994; Van Waerebeek *et al*. 1997), possibly reflecting a severe decrease in the abundance of the population. Given that all molecular and nonmolecular datasets collected thus far indicate that the eastern South Pacific dusky dolphin stock is reproductively

isolated, we promote a re-evaluation of present-day catch rates, fisheries practices and currently employed management measures in order to ensure the population's long-term survival. Furthermore, human exploitation might explain particularities exhibited by Peruvian dusky dolphins, such as their shorter reproductive cycle and higher fecundity in comparison with other dolphins of similar size (Van Waerebeek & Read 1994; Berta & Sumich 1999). Our molecular analyses indicate that Peruvian dusky dolphins exhibit levels of genetic variation that are significantly lower than those of any other population of the species examined thus far. However, given that severe exploitation of the Peruvian dusky dolphin population is probably not older than 30 years (corresponding to approximately four generations), it is more probable that the observed low genetic diversity might have been caused by regular El Niño events that severely disturb dusky dolphin's cool-water habitat. Only an accurate modelling of the relative impacts of human activities and natural El Niño episodes would allow the definition of a fine-scale management of the population.

Network methods

Contrary to phylogeny inference methods, the series of algorithms available for network estimation (reviewed in Posada & Crandall 2001) have not been compared in a systematic fashion in order to identify their possible specific artefacts as well as the impact of implicit and explicit assumptions of each approach. However, conflicting results, as shown in this study, emphasize clearly the need for more research on the relative performance of different network estimation methods. Moreover, we suggest that our exploratory comparative analysis identifies some major methodological issues described below.

Split decomposition. The usefulness of this method for reconstructing phylogenetic relationships among haplotypes is certainly limited when using large data sets, especially under high levels of homoplasy. Given this limitation, our split diagram, including 36 mostly closely related sequences, is still surprisingly informative, revealing a clear partition between Atlantic and Pacific dolphins. However, a considerable amount of homoplasy is evidenced. For example, the Atlantic haplotypes (regardless of the inclusion or exclusion of Pacific haplotypes) form a split diagram that includes five loops (Fig. 2A).

Minimum spanning network. The application of the MSN technique to our data set points out two potential methodological problems. First, the MSN approach seems to artificially favour star-like reconstructions of the genealogies. We propose that this systematic bias originates from the explicit assumption that 'the direct common ancestor of all observed haplotypes is itself present in the sample' (Excoffier & Smouse 1994), preventing the lineages of observed haplotypes to branch off from inferred intermediates. Hence, when haplotypes are sufficiently divergent such that a large number of missing intermediates must be inferred, most of the lineages in MSTs are constrained typically to diverge simultaneously from a 'central' observed haplotype. Hence, genealogies produced by the MSN method (i) probably overestimate overall tree length significantly (cf. we observe an increase from 71 to 85 mutational steps in the case of our data set), and (ii) can bias considerably and systematically interpretation of the data. Indeed, population genetic theory predicts that genealogical topologies reveal past demographic events (Avise 2000) and star-like phylogenies are usually interpreted as the signature of recent and rapid increases in population size (e.g. O'Corry-Crowe *et al.* 1997; Grant & Bowen 1998; Rosel *et al.* 1999). We suggest that the issue of star-like genealogies validation has been overlooked. The second possible shortcoming of the MSN approach is that some connections depend simply on the order in which the haplotypes are written in the input file. Under some haplotype orders, the algorithm can even result in the production of suboptimal alternative connections (see e.g. the direct five-step link between the central A1.1 haplotype and the peripheral A19 haplotype, highlighted as a dotted line in Fig. 2B), significantly limiting the utility of this method.

Statistical parsimony. It has been argued that this approach outperforms MP when applied to intraspecific data sets (Crandall 1996; Crandall & Templeton 1996), especially because the low number of mutational changes can cause MP to yield an astronomical number of equally parsimonious trees. Furthermore, the method incorporates the estimation (at the 95% confidence level) of a 'parsimony limit' (i.e. the maximum number of substitutions below which there is no multiple hit) yielding a quantitative assessment of the reliability of connections (Templeton *et al.* 1992; Crandall 1996). In the case of our mt data set, however, the statistical parsimony approach misplaced one lineage, such that the resulting network (Fig. 2C) is two steps longer than an alternative solution (Fig. 2D). Interestingly, however, when we add the central missing haplotype (CMI in Fig. 1) to the data set, TCS yields a network topology identical to that estimated by the median-joining approach. Although statistical parsimony seems to outperform the MSN method, further work will be needed to assess how inferred missing intermediates influence the former algorithm. Indeed, even small differences among topologies inferred with alternative approaches can have a large impact on the interpretation of the data. This is especially important with respect to the statistical parsimony, as this method is used increasingly for the inference of haplotypic networks in nested clade analyses (e.g. Bouzat *et al.* 1998; Good & Sullivan 2001;

Hurwood & Hughes 2001; Mardulyn 2001), although the exact algorithm implemented in TCS is not described fully.

Median-joining. Only this method yielded a network of length identical to that of the strict consensus among most parsimonious trees. Furthermore, the median-joining topology is identical to that of the strict consensus among three of the six MP trees. In other words, this approach fails to identify three among the six MP trees. The inclusion of less parsimonious pathways within the topology (by increasing parameter ε) did not seem to be very useful, at least as far as our data set is concerned. Indeed, even with reasonably low values of ε (e.g. 1), the resulting network included more reticulations than the split diagram, rendering data visualization and interpretation difficult.

In summary, network methods are useful because they allow to efficiently summarize the alternative (i.e. ambiguous) relationships within a genealogy, but some of the underlying assumptions in currently used network reconstruction approaches probably cause random or systematic artefacts that warrant extensive comparative analyses (possibly through simulation studies). Today, the user is faced simply with alternative methods yielding different topologies without the possibility to objectively evaluate their respective accuracies.

Rooting

As outlined above (cf. Introduction), outgroup rooting of intraspecific genealogies is problematic for two diametrically different reasons: an outgroup can be either too closely related to or too diverged from the ingroup. None of these two potential problems seem to be relevant to the present study, as phylogenetic analyses of the Lissodelphininae clade (using *D. delphis* as an outgroup), incorporating multiple *L. obliquidens* and *L. obscurus* individuals, indicates that the two species form reciprocally monophyletic sister groups (Fig. 3) separated by a short branch (20 substitutions on one of the 48 most parsimonious tree topologies). Furthermore, these phylogenetic analyses indicate that Peruvian dusky dolphins form a clade nested within a paraphyletic group of Atlantic haplotypes. On the contrary, the algorithm implemented in the TCS program for calculating rooting weights (Castelloe & Templeton 1994) — using the frequency of each candidate root haplotype, but also the number and frequencies of haplotypes to which it is connected — locates the root within the clade of Peruvian haplotypes (empty arrow, Fig. 1). However, as the frequencies of missing intermediates are unknown, the method disregards connections to unsampled haplotypes when calculating rooting probabilities. Potential problems, caused by the implementation of this algorithm, are particularly perceptible in the analysis of our data set. For example, assignment of the highest rooting probability to P5.1 (Fig. 1) seems artefactual. Indeed, we thought first that this rooting was due to the large frequency of the P5.1 haplotype, but rooting remains unchanged even after giving equal frequencies to all haplotypes. Haplotype P5.1 is assigned the highest root probability simply because it displays the highest number of connections (three) to sampled haplotypes. Several other haplotypes harbour three or more connections but all or some of these connections are to missing intermediates. Hence, when missing intermediates are not distributed homogeneously within the network, the method will suffer from a bias: e.g. a haplotype of a given frequency and with *n* connections to sampled haplotypes will be more likely to be assigned the root than another haplotype of identical frequency with n connections to haplotypes of which at least one is a missing intermediate. After assigning root probabilities to haplotypes, statistical tests are certainly needed in order to tell whether differences among candidate root haplotypes are significant. Furthermore, the level of robustness of the statistical parsimony method to violation of its basic assumptions (marker neutrality, homogeneous sampling, etc.) warrants further investigation. We suggest that inclusion of connections to missing haplotypes in the computation of relative rooting probabilities already increases the reliability of this estimation. Hence, we recomputed manually rooting weights of every candidate root haplotype using an alternative algorithm in which we considered both connections to sampled and to missing haplotypes. Using this approach, A1.1 is the most likely root haplotype, a result more consistent with outgroup rooting (i.e. placing the root among Atlantic haplotypes). Looking at haplotypes that are any number of steps away (focusing on the number of connections of interior haplotypes) seems therefore reasonable, although there are many different possible modifications of the original algorithm and simulation studies should be carried out to test them thoroughly. Clearly, reliable methods for estimating the placement of a network root warrant further developments. For example, in our analysis the most likely candidate for the root could be the central node (CMI, from which seven branches diverge) but, by definition, it is excluded from calculations requiring haplotype frequencies. Furthermore, additional parameters such as the geographical distribution of haplotypes might be of interest for efficiently estimating rooting probabilities.

Species origin and dispersion

Given (i) the observed levels of population subdivision among and genetic diversity within the Peruvian, Argentinean and Southwest African populations; (ii) the *L. obscurus* genealogy, as constructed by the median-joining network approach (Fig. 1); and (iii) the placement of the root (cf. above, outgroup rooting) along an Atlantic lineage, two evolutionary hypotheses regarding species origin and dispersion can be proposed. The two alternative hypotheses

disagree on the geographical identity of the missing ancestral haplotypes in the centre of the network (CMI and all missing intermediates connected to it, cf. Fig. 1). If all extant dusky dolphin lineages originated in the Atlantic Ocean (i.e. if their most recent common ancestor inhabited Atlantic waters), then the actual Peruvian population has probably been founded by a single dispersal event from an Atlantic lineage. This scenario would be compatible with the differences in levels of haplotype and nucleotide diversity revealed in this study: the ancestral populations (in Argentinean and South African waters) would, as expected, exhibit larger levels of genetic diversity than that of a more recent population founded from a very low number of possibly closely related individuals. Under the alternative scenario, most of the lineages observed today would have originated in the Pacific, multiple migration events from the Pacific to the Atlantic would have then occurred, and the fixation of a single well-defined and poorly diversified lineage of dusky dolphins in the Pacific would have been facilitated by severe and successive population bottlenecks caused by numerous and severe El Niño events. The likelihood of the second scenario might increase when considering the possible influence of an east-flowing Westwind Drift in the dispersion of the various populations of dusky dolphins (Van Waerebeek 1993b), a pattern of migration that has also been suggested for the genus *Cephalorhynchus* (Pichler *et al*. 2001). Additional data are clearly needed in order to discriminate between the two alternative evolutionary hypotheses described above. However, it should be noted that only statistical parsimony and median-joining network approaches, i.e. methods that inferred the central missing intermediate (CMI), yield sufficient resolution for testing these hypotheses. Indeed, in the star-like topology as inferred by the MSN, with rooting on one of the long Atlantic lineages, the evolutionary hypothesis suggesting species' origin in Pacific waters and subsequent dispersion under the influence of the Westwind Drift current becomes highly unlikely.

Acknowledgements

We wish to thank Rick LeDuc and Kelly Robertson from the SW Fisheries Center, La Jolla for kindly providing *Lagenorhynchus obliquidens* samples. The study has benefited from discussions with Keith Crandall and David Posada on outgroup weight probabilities. Comments from four reviewers and the subject Editor significantly improved the manuscript. Fieldwork in Peru (KVW) received long-term support from the Leopold III Fund for Nature Research and Conservation (Belgium), the IUCN/CSG Cetacean Specialist Group and the Van Tienhoven Foundation. IC is supported by the 'DAAD Doktorandenstipendium im Rahmen des gemeinsamen HSP III von Bund und Ländern'. The molecular work was funded by the National Fund for Scientific Research Belgium (FNRS), the Free University of Brussels (ULB), the Fonds Van Buuren, the Fonds E. Defay and the 'Communauté Française de Belgique' (ARC 11649/20022770).

References

Avise JC (1994) *Molecular Markers, Natural History, and Evolution*. Chapman & Hall, New York.

Avise JC (2000) *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge, MA.

Bandelt HJ, Dress AWM (1992) A canonical decomposition theory for metrics on a finite set. *Advances in Mathematics*, **92**, 47–105.

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

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

Bouzat JL, Lewin HA, Paige KN (1998) The ghost of genetic diversity past: historical DNA analysis of the greater prairie chicken. *American Naturalist*, **152**, 1–6.

Cassens I, Vicario S, Waddell VG *et al.* (2000) Independent adaptation to riverine habitats allowed survival of ancient cetacean lineages. *Proceedings of the National Academy of Science USA*, **97**, 11343–11347.

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution*, **3**, 102–113.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Crandall KA (1996) Multiple interspecies transmissions of human and simian T-cell leukemia/lymphoma virus type I sequences. *Molecular Biology and Evolution*, **13**, 115–131.

Crandall KA, Templeton AR (1996) Applications of intraspecific phylogenetics. In: *New Uses for New Phylogenies* (eds Harvey PH, Leigh Brown AJ, Maynard Smith J, Nee S), pp. 81–99. Oxford University Press, New York.

Crespo EA, Corcuera JF, Lopez Cazorla A (1994) Interactions between marine mammals and fisheries in some coastal fishing areas of Argentina. *Report of the International Whaling Commission (Special Issue)*, **15**, 269–281.

Crespo EA, Pedraza SN, Coscarella M *et al.* (1997a) Distribution and school size of dusky dolphins, *Lagenorhynchus obscurus* (Gray, 1828), in the southwestern South Atlantic Ocean. *Report of the International Whaling Commission (Special Issue)*, **47**, 693–697.

Crespo EA, Pedraza SN, Dans SL *et al.* (1997b) Direct and indirect effects of the highseas fisheries on the marine mammal populations in the northern and central Patagonian coast. *Journal of Northwest Atlantic Fishery Science*, **22**, 189–207.

Dans SL, Crespo EA, Garcia NA *et al.* (1997) Incidental mortality of Patagonian dusky dolphins in mid-water trawling: retrospective effects from the early 1980s. *Report of the International Whaling Commission (Special Issue)*, **47**, 699–703.

Donnelly P, Tavaré S (1986) The ages of alleles and a coalescent. *Advances in Applied Probability*, **18**, 1–19.

Excoffier L, Smouse PE (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics*, **136**, 343–359.

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

Good JM, Sullivan J (2001) Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern Rocky Mountain endemic. *Molecular Ecology*, **10**, 2683–2695.

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

Hillis DM, Moritz C, Mable BK (1996) *Molecular Systematics*. Sinauer Associates, Inc, Sunderland, MA.

Hochberg Y (1974) Some generalizations of the T-method in simultaneous inference. *Journal of Multivariance Analysis*, **4**, 224–234.

Hurwood DA, Hughes JM (2001) Nested clade analysis of the freshwater shrimp, *Caridina zebra* (Decapoda: Atyidae), from north-eastern Australia. *Molecular Ecology*, **10**, 113–125.

Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68–73.

LeDuc RG, Perrin WF, Dizon AE (1999) Phylogenetic relationships among the delphinid cetaceans based on full cytochrome *b* sequences. *Marine Mammal Science*, **15**, 619–648.

Mardulyn P (2001) Phylogeography of the Vosges mountains populations of *Gonioctena pallida* (Coleoptera: Chrysomelidae): a nested clade analysis of mitochondrial DNA haplotypes. *Molecular Ecology*, **10**, 1751–1763.

Milinkovitch MC, LeDuc RG, Adachi J *et al.* (1996) Effects of character weighting and species sampling on phylogeny reconstruction: a case study based on DNA sequence data in cetaceans. *Genetics*, **144**, 1817–1833.

Milinkovitch MC, LeDuc RG, Tiedemann R, Dizon AE (2002) Applications of molecular data in cetacean taxonomy and population genetics with special emphasis on defining species boundaries. In: *Marine Mammals: Biology and Conservation* (eds Evans PGH, Raga JA), pp. 325–359. Kluwer Academic/Plenum Publishers, New York.

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

O'Corry-Crowe GM, Suydam RS, Rosenberg A, Frost KJ, Dizon AE (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.

Pichler FB, Robineau D, Goodall RNP *et al.* (2001) Origin and radiation of Southern Hemisphere coastal dolphins (genus Cephalorhynchus). *Molecular Ecology*, **10**, 2215–2223.

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

Prim RC (1957) Shortest connection networks and some generalizations. *Bell Systems Technical Journal*, **36**, 1389–1401.

Rice DW (1998) *Marine Mammals of the World Systematics and Distribution*. Allen Press, Inc., Lawrence, KS.

Rohlf FJ (1973) Algorithm 76. Hierarchical clustering using the minimum spanning tree. *Computer Journal*, **16**, 93–95.

Rosel PE, Tiedemann R, Walton MJ (1999) Genetic evidence for restricted trans-Atlantic movements of the harbour porpoise, *Phocoena phocoena*. *Marine Biology*, **133**, 583–591.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN, version 2.000: a Software for Population Genetics Data Analysis Genetics and Biometry Laboratory*. University of Geneva, Geneva, Switzerland.

Swofford DL (2000) *PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

Templeton AR (1992) Human origins and analysis of mitochondrial DNA sequences. *Science*, **255**, 737.

Templeton AR (1993) The 'Eve' hypotheses: a genetic critique and reanalyis. *American Anthropologist*, **95**, 51–72.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Van Waerebeek K (1992) Records of dusky dolphins *Lagenorhynchus obscurus* (Gray, 1828) in the eastern South Pacific. *Beaufortia*, **43**, 45–61.

Van Waerebeek K (1993a) External features of the dusky dolphin *Lagenorhynchus obscurus* (Gray, 1828) from Peruvian waters. *Estudios Oceanologicos*, **12**, 37–53.

Van Waerebeek K (1993b) Geographic variation and sexual dimorphism in the skull of the dusky dolphin, *Lagenorhynchus obscurus* (Gray, 1828). *Fishery Bulletin*, **91**, 754–774.

Van Waerebeek K (1994) A note on the status of the dusky dolphin (*Lagenorhynchus obscurus*) off Peru. *Report of the International Whaling Commission (Special Issue)*, **15**, 525–527.

Van Waerebeek K, Read AJ (1994) Reproduction of dusky dolphins, *Lagenorhynchus obscurus*, from coastal Peru. *Journal of Mammalogy*, **75**, 1054–1062.

Van Waerebeek K, Reyes JC (1990) Catch of small cetaceans at Pucusana port, central Peru, during 1987. *Biological Conservation*, **51**, 15–22.

Van Waerebeek K, Reyes JC (1994a) Interactions between small cetaceans and Peruvian fisheries in 1988–89 and analysis of trends. *Report of the International Whaling Commission (Special Issue)*, **15**, 495–502.

Van Waerebeek K, Reyes JC (1994b) Post-ban small cetacean takes off Peru: a review. *Report of the International Whaling Commission (Special Issue)*, **15**, 503–520.

Van Waerebeek K, Reyes JC, Alfaro J (1993) Helminth parasites and phoronts of dusky dolphins *Lagenorhynchus obscurus* (Gray, 1828) from Peru. *Aquatic Mammals*, **19**, 159–169.

Van Waerebeek K, Van Bree PJH, Best PB (1995) On the identity of *Prodelphinus petersii* Luetken, 1889 and records of dusky dolphin *Lagenorhynchus obscurus* (Gray, 1828) from the southern Indian and Atlantic Oceans. *South African Journal of Marine Science*, **16**, 25–35.

Van Waerebeek K, Van Bressem M-F, Felix F *et al.* (1997) Mortality of dolphins and porpoises in coastal fisheries off Peru and southern Ecuador in 1994. *Biological Conservation*, **81**, 43–49.

Wheeler WC (1990) Nucleic acid sequence phylogeny and random outgroups. *Cladistics*, **6**, 363–367.

All authors share a common interest in the conservation of small cetacean natural populations. This work is part of I. Cassens' doctoral thesis on molecular phylogenetics and conservation genetics of cetaceans. M. C. Milinkovitch is associate professor at the Free University of Brussels (ULB), heads the laboratory of Evolutionary Genetics at the Institute of Molecular Biology and Medicine and works on various aspects of conservation genetics, phylogenetics, and phyloinformatics. K. Van Waerebeek, who is director of the Peruvian Centre for Cetacean Research, coordinates marine mammal conservation and research projects in the SE Pacific, West Africa and the Arabian Sea. P. B. Best is a Senior Research Officer with the Mammal Research Institute of the University of Pretoria, although based at the South African Museum in Cape Town. E. A. Crespo and J. Reyes lead South American research programmes in conservation biology.

Appendix I

	11111111 11111111223334555666666666777777777888899999900000011 24469900134689021242003011145789124577899157900368881245671 2 46966969484268488656474915887283211901856984749000598333580 2
A1.1	CGGCCTTCTATCTCTCTTCTTCTAGCTCACCTAGTTCCGCGTTTTGTTTCACGATTTTCC
A1.2	.....T.....C.....
A1.3	.....C.....
A1.4	.....T.....T.....T.
A1.5	.....T.....A.....C.....
A2.1	.....T....C.GT..G.....
A2.2	.....T.....T.....
A3	.....T.....C...C.....C.GT..G.....
A4	.....G.....C....CG....T...C.....C...
A5	T.....G.....C....CG....T...C.....c...
A6	.....T.....C.....G.....C.....A....C..
A7	.....C.T.C.....C.GT..G.....
A8.1	.....C.....C.....C.....
A8.2	.....C.....C.....C...A.....
A9	.....C.C.....C.....G.....C.....
A10	.....C.....C.....C.....
A11	.....C.....T.....A.....C.....
A12	.....C.....T.....
A14	.....C.....T.....A...T.....
A17	.....C.....T.....C.....
A18	...T.....C....C...T..C...T.G.....
A19	.....C.....C.....A.C.....C.....
P1	..A.T.....T.C.....A.C.....G.....
P2.1	..A.....T...C.....A.C.....G.....G.....
P2.2	..A.....T...C.....A.C.....G.....
P3.1	..A.T.....C.....A.C.....G.....
P3.2	..A.T.....C.....A.C.....GA.....C.....
P4.1	..A.....CT.....C.....A.C.....GAC....A.....
P4.2	..A.....CT.....C.....A.C.....GA.....A.....
P5.1	.AA.....C.....A.C.....G...T.....
P5.2	.AA.....C.....A.C.....G...T.....T
P5.3	.AA.....C.....A.C.....G...T.....T.....
P5.4	.AA.....C.....A.C.....G...T.....C.....
P6.1	..A.....C.....A.C.....G.....
P6.2	..A.....C.....A.C.....G.....G.C....
P7	..A.....C.C....A.C.....G.....G.C....