Genome-Wide SNP and Population Divergence of Finless Porpoises

Shuzhen Li[1,2], Shixia Xu[1], Huirong Wan[1], Heyi Ji[1], Kaiya Zhou[1], and Gang Yang[1,*]

[1]Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, China

[2]Present address: Nanjing Children's Hospital, Affiliated to Nanjing Medical University, Nanjing, China

*Corresponding author: E-mail: gyang@njnu.edu.cn.

Accepted: March 25, 2013

Data deposition: This project has been deposited at GSS under accession nos. FI592654, FI592658–FI592662, FI592665, FI592667, FI592668, FI592670, FI592671, FI592673–FI592675, FI592677, FI592680, FI592682–FI592688, FI592690, FI592691, FI592694–FI592696, FI592698, FI592699, FI592701–FI592705, FI592707–FI592711, FI592713, FI592714, FI592717, FI592720, FI592722, GS884018–GS884021, GS884023–GS884025, GS884027, GS884028, GS884030, GS884033, GS884034, GS884036–GS884039, GS884042–GS884044, GS884046, GS884049, GS884051–GS884059, GS884061, GS884063, GS884068–GS884070, GS884074, GS884077, GS884079, GS884082–GS884086, GS884088, GS884091, GS887694, GS887695, FJ176332, FJ176335, FJ176338, FJ176341, FJ176346, FJ176352, FJ176357, FJ176361, FJ176362, FJ176366, FJ176368, FJ176372, FJ176375, FJ176376, and FJ176378.

Abstract

Single nucleotide polymorphisms (SNPs) are rapidly becoming the population genomic markers in addressing ecology, evolution, and conservation issues for their high capacity to access variability across the genome. We isolated a total of 140 ideal SNPs from the finless porpoise and used 78 (under Hardy–Weinberg equilibrium) of them to conduct those issues especially for addressing population genetic differentiation. Bayesian clustering and principal component analyses all suggested that finless porpoises in Chinese waters could be divided into three distinct genetic groupings. Low levels of within-population genetic variation (mean $H_E = 0.3405$, standard deviation $= 0.1188$) and significant differentiation among populations ($F_{ST} = 0.1050$–0.1628, $P < 0.01$) were confirmed. Limited gene flow was found especially between the freshwater Yangtze River porpoise and the oceanic Yellow Sea and South China Sea populations, which strongly suggested that some barriers might have restricted their genetic exchange. These evidences not only support a recent subdivision of the finless porpoise into two species but also suggest a full species status for the Yangtze finless porpoise, especially considering the significant genetic divergence between freshwater and marine porpoises, in combination with the unique distribution of Yangtze finless porpoises in freshwater and their distinctness in physiological and morphological features.

Key words: finless porpoise, SNPs, genetic diversity, population divergence.

Introduction

Recent theoretical (Morin et al. 2004) and empirical studies (Seddon et al. 2005; Narum et al. 2008) suggest that the single nucleotide polymorphisms (SNPs) is becoming particularly useful in evolution, ecology, and conservation, especially with the accumulation of genomic sequence information in the past decades. However, application of SNPs in wild populations addressing ecological or conservation issues is still limited to a few cases. Until recently, Choi et al. (2007) have conducted genome-wide SNP discovery projects either using existing genomic data which are, although very limited, becoming increasingly available in some non-model organisms, or using basic cloning and sequencing techniques. Genome-wide data sets have the potential to improve the inference of population parameters and to reliably reconstruct population demography and evolutionary history.

The finless porpoise, *Neophocaena phocaenoides*, is one of the smallest cetaceans inhabiting shallow and often partially enclosed marine waters along the coasts of southern and eastern Asia (Reeves et al. 2003). In China, its distribution covers all Chinese coastal waters and the freshwater Yangtze River, with the Yangtze finless porpoise as the sole freshwater population. Gao and Zhou (1993) have noted that there are geographic variations in external morphology and differences in growth and reproduction pattern among finless porpoises from different parts of Chinese waters for

© The Author(s) 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

a long period of time, and three populations (or ecological/morphological forms) from the very beginning: the Yellow Sea population, the South China Sea population, and the freshwater Yangtze River population. Thus, they also regarded these three populations as respective subspecies (Gao and Zhou 1995a, 1995b, 1995c): 1) *N. p. phocaenoides* is characterized with a wide area of tubercules and more than 10 rows of denticles on the dorsal surface. This made them being called "wide-ridge form" or "wide form" as well. This subspecies has a range not only in the southern part of East China Sea and the South China Sea but also extends their habitat along the mainland coast of southern Asia west to the Persian Gulf (Rice 1998); 2) *N. p. sunameri* is characterized by a narrow tuberculed area on the dorsal ridge ("narrow-ridge form" or "narrow form") and is found in the northern part of East China Sea, the Yellow/Bohai Seas, and the waters of Korea and Japan. This subspecies is regarded as potentially sympatric with *N. p. phocaenoides* in the Taiwan Strait; and 3) *N. p. asiaeorientalis* is also characterized by a similar dorsal surface with *N. p. sunameri* but has a distinct distribution in freshwaters, that is, the middle and lower reaches of the Yangtze River, including Lakes Poyang and Dongting and their tributaries, the Ganjiang River and Xiangjiang River (Rice 1998). Jefferson (2002) examined the geographical variation in cranial morphometrics of the finless porpoise and suggested that the recognition of two species, a hypothesis originally proposed by Pilleri and Gihr (1972), might be warranted. Wang et al. (2008) further supported this hypothesis, which was demonstrated that two morphological forms of finless porpoises, although sympatric with each other in the Taiwan Straits, have been reproductively isolated for approximately 18,000 years and thus represented two distinct biological species as suggested by Jefferson (2002): the Indo-Pacific finless porpoise, *N. phocaenoides*; and the narrow-ridged finless porpoise, *N. asiaeorientalis* (Pilleri and Gihr 1972). Wang et al. (2010) further demonstrated that the two species can be differentiated at sea by direct observation and examination of high-quality photographs. However, Jefferson (2002) and Jefferson and Hung (2004) did not support the further subdivision of the narrow form into the *sunameri* (Yellow Sea–Japanese Sea finless porpoise) and *asiaeorientalis* (Yangtze finless porpoise) subspecies as suggested by Gao and Zhou (1995a, 1995b, 1995c). Wang et al. (2008) was also unable to answer this question due to unavailability of the freshwater samples from the Yangtze River in their study. Molecular markers such as mitochondrial control region and nuclear microsatellite have been applied to address genetic diversity, population structure and phylogeography of finless porpoises in the past decades (Yoshida et al. 2001; Yang et al. 2002, 2003, 2008; Wang et al. 2008; Chen et al. 2010), uncovering significant genetic differentiation among populations. However, inferences about the level of population genetic variability were not always congruent. Although mitochondrial haplotype diversity and nucleotide variability of finless porpoises were overall low compared with other cetaceans (Yoshida et al. 2001; Yang et al. 2002, 2003, 2008), relatively high levels of genetic variation in nuclear microsatellite profiles as shown with H_E (0.732–0.795) were found comparable with that of some nonendangered cetacean species (Chen et al. 2010).

When the three populations in Chinese waters were specially taken into consideration, significant differentiation was found between either the Yangtze River or the Yellow Sea population and the South China Sea population, whereas the differentiation level between the Yangtze freshwater population and the adjacent Yellow Sea population was, albeit significant, relatively smaller. In addition, although the Yellow Sea population and the South China Sea population were regarded to be potentially sympatric in the Taiwan Straits, combined genetic and morphological evidence suggest little or no genetic exchange between them, and likely (although relatively recent) species-level differentiation (Wang et al. 2008). This was further evidenced by Chen et al. (2010) using microsatellite analyses of a larger size of samples from Chinese coastal waters and the freshwater Yangtze River, which revealed no or very limited gene flow among subspecies or morphological forms even in the overlapping areas. However, an inference from genome-wide analyses of genetic diversity would be less subject to the inherent biases due to individual or small number of loci such as mtDNA control region and microsatellite, and thus will reflect more accurate overall pattern of genetic variation in the finless porpoise.

In this study, we developed a series of SNP loci via random whole-genome "shotgun" sequencing technique (Staden 1979) and a targeted gene approach (e.g., CATS, Aitken et al. 2004). We finally chose a total of 140 SNPs to genotype 202 individuals representing 15 sample sites from a relatively wide geographic range across Chinese coastal waters and the Yangtze River. We used the resultant data sets to address genetic diversity and population structure of finless porpoises at the genomic level. This genome-wide picture of standing variation can certainly provide broader understanding of the taxonomic status of *Neophocaena* populations, and more importantly in assessing conservation priority and designing conservation programs for this highly endangered species.

Materials and Methods

Sample Collection and DNA Isolation

Two hundred and two tissue samples of finless porpoises collected from the middle and lower reaches of the Yangtze River and the coastal China seas were available for this study (fig. 1 and table 1). Because all the finless porpoise specimens analyzed in this study were killed incidentally in fishing nets or were found stranded between 1979 and 2009, no ethical approval is necessary in such a case. We collected these samples from 15 sites, with 1 to 72 specimens in each site. Although there were only very few specimens in some sites,

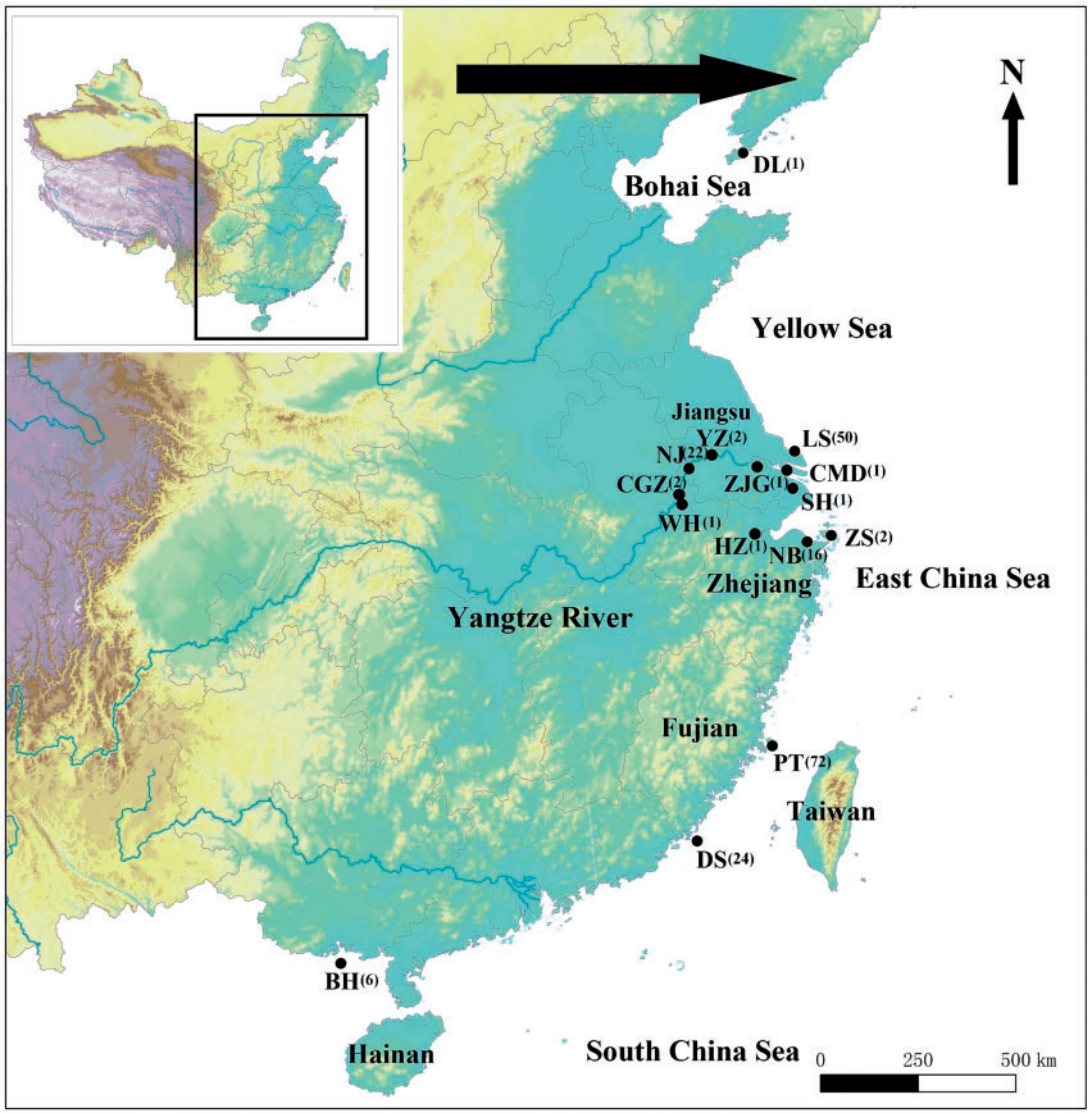

Fig. 1.—Schematic map showing finless porpoises sampled in this study, with sample size for each locality shown in table 1. Sampling locality abbreviations are as follows: WH (Wuhu); CGZ (Caoguzhou); NJ (Nanjing and Jiangpu); YZ (Yizheng); ZJG (Zhangjiagang); SH (Shanghai or Yangtze River mouth); CMD (Chongmingdao); DL (Dalian); LS (Lusi); HZ (Hangzhou); ZS (Zhoushan); NB (Ningbo); PT (Pingtan); DS (Dongshan); and BH (Beihai).

the total sample sizes for the three major geographic regions of finless porpoises in China were moderately large (30 in the Yangtze River, 70 in the Yellow/Bohai Sea, and 102 in the South China Sea). Voucher specimens were preserved in Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University. Total genomic DNA from muscles or skeleton samples was extracted by using the DNeasy tissue kit (Qiagen) following the manufacturer's instruction.

SNP Screening and Genotyping

We used a shotgun library as the major approach for producing SNPs in the finless porpoise. Seven hundred twenty colonies were screened, 502 of which were sequenced on an ABI PRISM 3730 automated DNA sequencer (Applied Biosystems). Two hundred ninety sequences of sufficient length and ideal quality if in both directions were chosen for designing primers (Li et al. 2009). In addition, we also used a total of 202 CATS (comparative anchor tagged sequences) primers to discover SNPs following the methods described in Aitken et al. (2004). The targeted gene approach was used to amplify a less conserved region (e.g., an intron or 3′-UTR) with primers designed from conserved regions of aligned genes of at least two species (e.g., mouse and human) (Lyons et al. 1997). The advantages of this approach include wide, current availability of primers, knowledge of the gene

Table 1
List of Sampling Sites of Finless Porpoises Examined in This Study

Geographical Region	Sample Site No.	Sample Site Name	Sample Size
Yangtze River (30/7)			
Anhui Province	1	Wuhu	1
	2	Caoguzhou	2
Jiangsu Province	3	Nanjing	22
	4	Yizheng	2
	5	Zhangjiagang	1
Shanghai City	6	Chongming Island	1
	7	Yangtze River mouth waters	1
Yellow/Bohai Sea (70/5)			
Liaoning Province	8	Dalian	1
Jiangsu Province	9	Lusi	50
Zhejiang Province	10	Hangzhou	1
	11	Ningbo	16
	12	Zhoushan	2
South China Sea (102/3)			
Fujian Province	13	Pingtan	72
	14	Dongshan	24
Guangxi Province	15	Beihai	6

NOTE.—The values in the parenthesis refer to the total number of samples in the region/number of sample sites in that region.

ortholog in which the SNPs are found, which could be useful for potentially broad application over a group of species with less per-sequence initial effort to find SNP loci than might be required using a random sequence approach (Aitken et al. 2004).

We selected primers (from shotgun library or CATS method) yielding a single polymerase chain reaction (PCR) product for amplification in 24 individual DNA samples (eight specimens from the Yangtze River, the Yellow/Bohai Sea, and the South China Sea). We selected optimal SNPs for further genotyping and characterization by the fragment length discrepant allele-specific PCR (FLDAS-PCR) scored on a LI-COR 4300 DNA Analyzer (LI-COR Biosciences) (Li et al. 2009) according to a variety of practical factors, such as whether sufficient flanking sequences were available for primer design and whether the designed primers could genotype individuals very well (Morin et al. 2007).

Data Analysis

Genetic Variation and Population Differentiation

To assess how SNP diversity varied among different geographical regions, we grouped three populations according to their sampling localities (fig. 1 and table 1) and calculated the average expected (H_E) heterozygosities and the within-population fixation index (F). Large allele dropout or stuttering was detected with MICRO-CHECKER utility (van Oosterhout et al. 2004). We also tested population departure from Hardy–Weinberg equilibrium (HWE). Whenever pertinent, the significance of P values was adjusted following Bonferroni sequential corrections for multiple simultaneous statistical tests (Rice 1989). These parameters were calculated with GENEPOP v3.4 (Raymond and Rousset 1995). The difference between populations was calculated using Arlequin v3.1 (Excoffier et al. 2005) exact test of population differentiation of pairwise weighted mean F_{ST} (Weir and Cockerham 1984).

Population Demography Inference

We tested for recent population bottlenecks in the three finless porpoise populations using BOTTLENECK v1.2.02 program (Piry et al. 1999). To determine whether a population exhibits a significant number of loci with heterozygosity excess, BOTTLENECK proposes three tests: Sign test, Standardized differences test, and Wilcoxon sign-rank test. The allele frequency distribution is established to see whether it is approximately L shaped (as expected under mutation-drift equilibrium) or not (recent bottlenecks provoke a mode shift).

Bayesian Analysis with STRUCTURE

We used STRUCTURE v2.0 (Pritchard et al. 2000) for searching for the occurrence of independent populations (clusters, K) in the data set. This approach avoids a priori population classifications, and instead estimates the shared population ancestry of individuals based solely on their genotypes assuming HWE and linkage equilibrium in ancestral populations. Calculations were conducted under the admixture model and the assumption of correlated allele frequencies (F-model), with a burn-in of 100,000 followed by 1,000,000 iterations. We set the number of populations, K, from 1 to 8. Ten independent runs of the Markov chain were performed for checking the convergence of the chain and homogeneity among runs for each K. The true number of populations (K) was usually identified using the maximal value of ln Pr ($X|K$) [$L(K)$] returned by structure (Pritchard et al. 2000). However, Evanno et al. (2005) observed no clear mode of the distribution of $L(K)$ for the true K, whereas they found that an ad hoc quantity based on the second order rate of change of the likelihood function with respect to K (ΔK) did show a clear peak at the true value of K.

To examine recent migration among populations and for comparisons with the historical migration estimated under the coalescent, we conducted assignment tests on the SNP genotypes in a separate STRUCTURE run based on the value of K resulting from model selection. We allowed for detection of migrants up to four generations before present (option GENSBACK = 4). To ensure a strong statistical support for any inference of mixed ancestry, we further set MIGRPRIOR = 0.1.

Principal Component Analysis

Principal component analysis (PCA) has similar power to detect population structure as STRUCTURE (Patterson et al. 2006). In addition, an estimation of the maximal number of subpopulations that can be found within a data set was achieved by determining the number of statistically significant principal components (PCs). For the 202 finless porpoises, genotypes were coded as −1, 0, or 1. PCs and PC variances were calculated from the singular value decomposition using R 3.0.0 (http://cran.r-project.org/bin/windows/base/).

Results

SNP Discovery and Genotyping Success

A total of 138,434 bp high-quality genomic sequences from CATS and shotgun libraries were screened and used to detect SNPs, with 380 SNPs found at 168 loci, equivalent to the mean value of one SNP per 364 bp. We also identified 51 ideally amplified SNPs from the 55 CATS primer pairs (with a total length of 28,107 bp), showing an average frequency of one SNP every 551 bp. The density was lower within the nuclear introns than that within the randomly cloned "shotgun" sequences (1/335 bp). We then chose a subset of approximately 228 optimal SNPs for the genotyping assay and 58 of them failed to amplify consistent products, whereas 19 yielded monomorphic profiles in the samples (i.e., <1% frequency of the minor allele), 11 produced predominantly heterozygote-like signals most probably due to hypervariable features or co-amplification of duplicated genes, and the remaining 140 (61%) were polymorphic and were successfully genotyped in more than 97% of the sampled individuals.

Genetic Diversity and Population Differentiation

All the 140 SNP loci were found highly polymorphic in finless porpoises. No more than two nucleotides could be detected per SNP, and accordingly the mean number of alleles per polymorphic locus (A) per population ranged from 1 to 2 with a mean value of 1.52. Approximately 99% SNPs had an overall frequency equal to or greater than 0.10 (fig. 2) and thus were considered common SNPs. As shown in table 2, the overall genetic diversity (i.e., considering all polymorphic SNPs together) expressed by the average observed heterozygosity H_O ranged from 0.2326 to 0.2920 among populations with a grand mean of 0.2609 (standard deviation [SD] = 0.1190). Average unbiased H_E was generally higher than H_O and ranged from 0.3116 to 0.3602 per population with a grand mean of 0.3405 (SD = 0.1188). The average value of H_O was lower than H_E and the average $F_{IS} = 0.1840$ was significantly positive ($P < 0.05$, table 2), indicating that there were fewer heterozygotes than expected in the total sample (table 2). After sequential Bonferroni correction (Rice 1989), we found that all three population units showed significant deviations from HWE in 62 loci, which would be ruled out for the subsequent detection (supplementary table S1, Supplementary Material online). Micro-Checker assessment for all of the isolated loci indicated that there were no scoring errors attributable to stuttering or large allele dropout, which suggested there were some other reasons might cause deviation from HWE. One of the reasons is probably the loss of

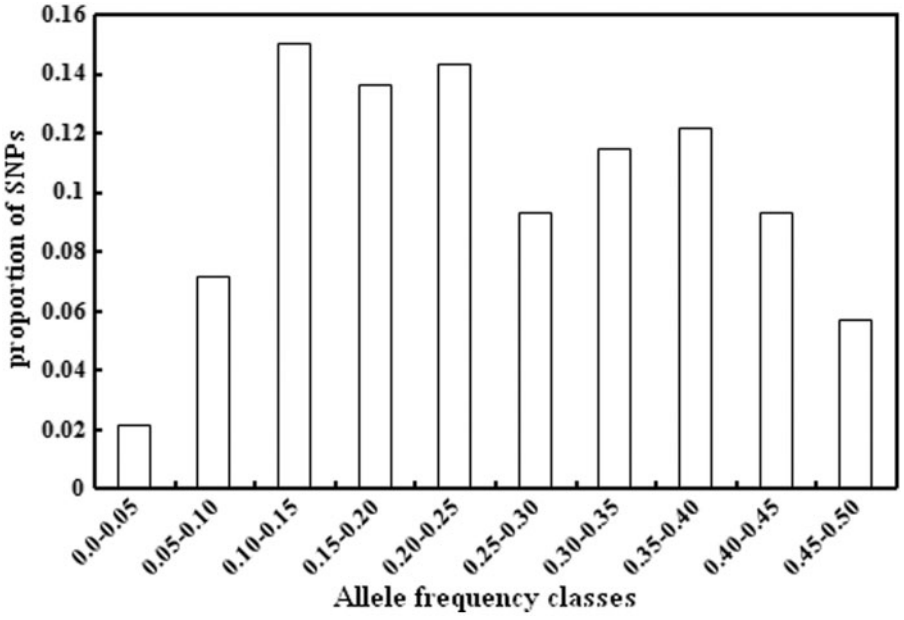

Fig. 2.—Allele frequency distribution at 140 SNP loci in 202 finless porpoises.

Table 2
Genetic Diversity Parameters for Three Finless Porpoise Populations Based on 140 Polymorphic SNP Loci

Population	*n*	A^a (SD)	$A_e{}^b$ (SD)	$H_O{}^c$ (SD)	$H_E{}^d$ (SD)	$F_{IS}{}^e$
Yangtze River	30	1.9929 (0.0845)	0.0845 (0.2920)	0.2920 (0.1471)	0.3402 (0.1280)	0.1475
Yellow/Bohai Sea	70	2.0000 (0.0000)	1.5686 (0.2847)	0.2869 (0.1614)	0.3602 (0.1251)	0.1568
South China Sea	102	2.0000 (0.1200)	1.5128 (0.3008)	0.2326 (0.1457)	0.3116 (0.1409)	0.2537
Total	202	2.0000 (0.0845)	1.5938 (0.2745)	0.2609 (0.1254)	0.3405 (0.1188)	0.1840

[a]Mean number of alleles.
[b]Effective number of alleles.
[c]Observed heterozygosity.
[d]Expected heterozygosity (H_E).
[e]Fixation index (F_{IS}).

Table 3
Pair-Wise F_{ST} Comparisons (below Diagonal) and Relevant *P* Values (above Diagonal) between Geographical Populations of Finless Porpoises in Chinese Waters

Population	Yangtze River	Yellow Sea	South China Sea
Yangtze River		$P < 0.0001$	$P < 0.0001$
Yellow Sea	0.1050		$P < 0.0001$
South China Sea	0.1628	0.1151	

heterozygotes, which may be explained by the habitat fragmentation leading to further division of each population into small subpopulations causing an increase of homozygotes due to genetic drift and endogamy. As each population was sampled from several localities, with small numbers of individuals in some localities, different allele frequencies in the populations would cause a Wahlund effect, which might also result in deviations from expectation (Wahlund 1928).

Pairwise population divergence (F_{ST}) was low between the Yangtze River and Yellow Sea populations ($F_{ST} = 0.1050$, $P < 0.0001$), whereas the largest divergence was found between the Yangtze River and South China Sea populations ($F_{ST} = 0.1628$, $P < 0.0001$) (table 3). Divergence in SNP frequencies among present-day populations was low but significant, and showed a multimodal distribution, suggesting the presence of higher order genetic structure.

Bottleneck Detection

To detect whether the finless porpoises in Chinese Waters have experienced a population reduction in size, we detected excess heterozygosity in a population at mutation-drift equilibrium (H_{eq}) under the following two models of mutation: the IAM and the SMM by using the program BOTTLENECK. Three populations showed significant ($P < 0.05$) heterozygosity excess under all the two models as an indication of recent demographic contraction.

Population Structure

We used the total sample set, 78 loci (under HWE) and $K = 1$–8, Bayesian analysis of population structure generated

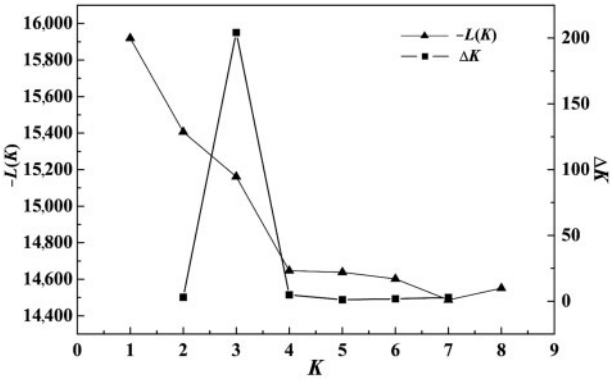

FIG. 3.—Magnitude of ΔK as a function of K (mean ± SD over 10 replicates) calculated by using the $L(K)$ [the log-likelihood value; ln Pr (X/K)]. The real number of groups is best detected by the highest value of ΔK, a quantity based on the second-order rate of change with respect to K of the likelihood function.

a maximum probability of the data with $K = 7$, but the modal value of ΔK was at $K = 3$ (fig. 3). Samples from Chinese waters were separated into three distinct clusters, which indicated that the individuals examined were most likely to be divided into three genetically distinct groups. All individuals from the Yangtze River were grouped into a single cluster. The second cluster contained mostly the individuals from the Yellow/Bohai Sea and the East China Sea, whereas the third was mainly comprised of the South China Sea samples from Dongshan, Pingtan, and Beihai (fig. 4). As shown in figure 4, individuals from the Yangtze River and the Yellow/Bohai Sea were almost grouped together with some admixture between the South China Sea clusters. We also conducted a PCA on the population allele frequencies, which also revealed three major clusters (fig. 5) broadly consistent with assignments in the above STRUCTURE analyses. It implied the highest variation between Yellow Sea and South China Sea samples, since these were placed at opposite ends of PC1, which harbored 15.98% of the variation (fig. 5). PC2 split samples from Yangtze River and Yellow Sea clusters and explained 5.77% of the variation found in the data set (fig. 5).

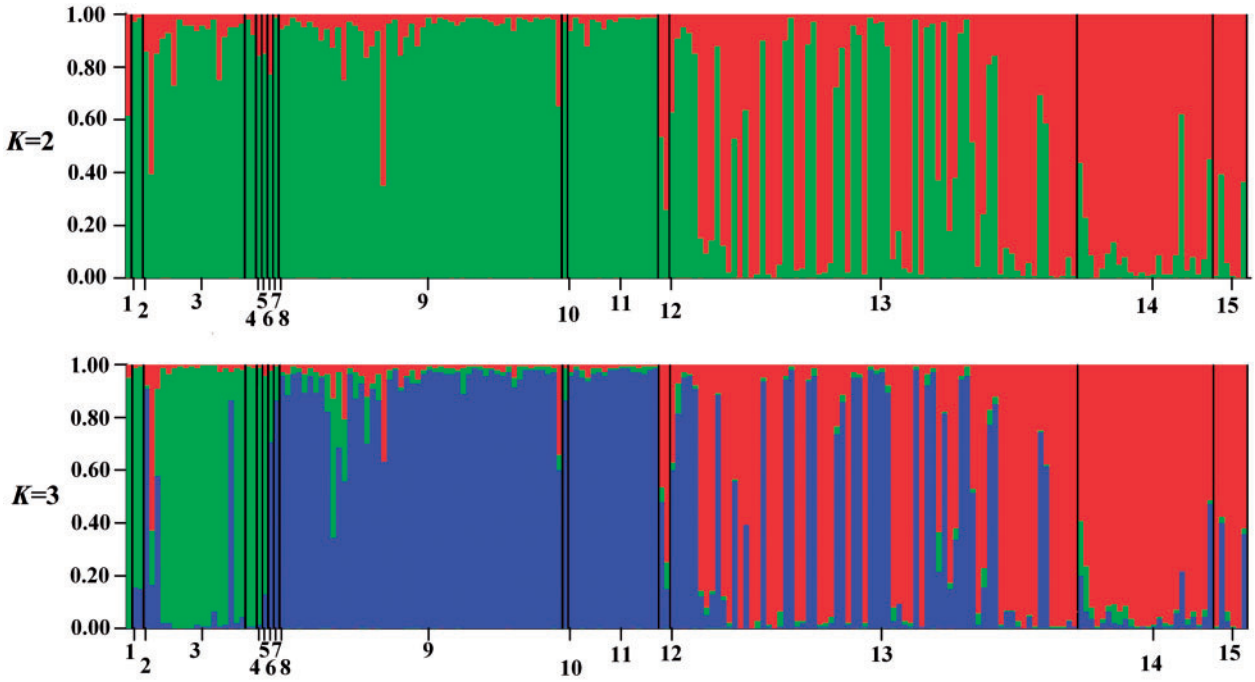

Fig. 4.—Estimated population structure for finless porpoises in Chinese waters. Sampling locations are labeled in numbers below the figure. Each individual of those populations is represented by a vertical line, which is partitioned into *K* colored clusters representing the individual's estimated membership fractions. $K = 3$ was shown based on the ΔK described in Evanno et al. (2005) and the colors for it correspond to colors used in PCA (fig. 5). 1, Wuhu; 2, Caoguzhou; 3, Nanjing; 4, Yizheng; 5, Zhangjiagang; 6, Chongmingdao; 7, Yangtze River mouth; 8, Dalian; 9, Lusi; 10, Hangzhou; 11, Ningbo; 12, Zhoushan; 13, Pingtan; 14, Dongshan; 15, Beihai.

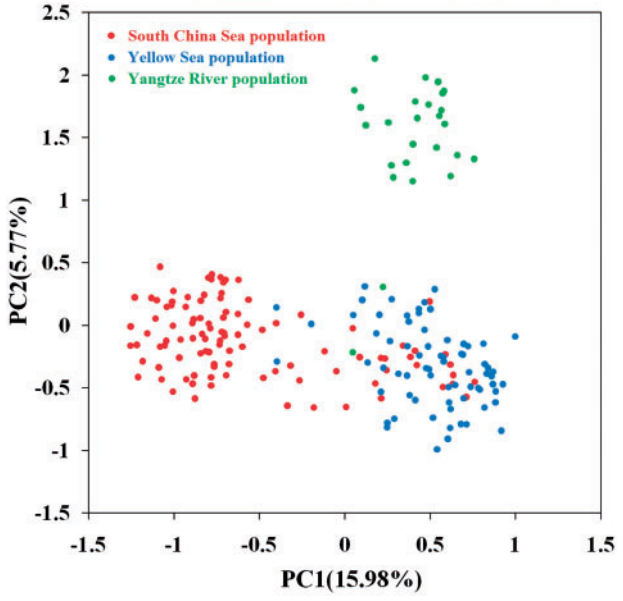

Fig. 5.—PCA showing all the individuals based on the 78 SNP loci. Plotting PC1 against PC2 revealed three major geographical groups (populations).

However, both clustering analyses all showed some admixture between populations. Three individuals from Nanjing and the Yangtze River mouth waters were also found in the Yellow Sea population cluster with a probability of more than 60.0% (table 4). In addition, one individual from Zhoushan and two others from Lusi of the Yellow Sea were found intermixed with the South China Sea samples. More individuals ($n = 30$) collected in the South China Sea were genetically assigned to the Yellow Sea cluster. Especially, 29 of 72 individuals from Pingtan, which is generally regarded as the overlapping area of two finless porpoise forms (i.e., wide form and narrow form), were identified to be intermixed with those individuals from the Yellow Sea population. In total, 36 individuals were identified as migrants between populations.

Of the 36 migrants detected using STRUCTURE, 31 were further identified as first-generation migrants which all had a probability of 1.0 except for two samples that had pure ancestry proportion of the Yellow Sea population and a probability of less than 0.9 ($P = 0.702$), whereas three were identified as second- to fourth-generation migrants but all of those had very low probabilities (0.434, 0.505, and 0.513, respectively, for the second, third, and fourth generation). This provided weak evidence that these later

Table 4

Bayesian Clustering Analysis of Finless Porpoises Performed Using STRUCTURE (Pritchard et al. 2000) with the Total Sample Set (202 Samples; 78 SNP Loci; and 15 Sampling Locations)

Locality	Proportion of Membership for $K = 3$		
	Cluster 1	Cluster 2	Cluster 3
Wuhu (1)	0.100 (1)	0.000 (0)	0.000 (0)
Caoguzhou (2)	0.908 (2)	0.092 (0)	0.000 (0)
Nanjing (22)	0.889 (20)	0.090 (2)	0.021 (0)
Yizheng (2)	0.981 (2)	0.011 (0)	0.008 (0)
Zhangjiagang (1)	0.938 (1)	0.013 (0)	0.049 (0)
Chongmingdao (1)	0.868 (1)	0.103 (0)	0.029 (0)
Yangtze River mouth (1)	0.067 (0)	0.690 (1)	0.243 (0)
Dalian (1)	0.138 (0)	0.854 (1)	0.008 (0)
Lusi (50)	0.081 (0)	0.872 (48)	0.047 (2)
Hangzhou (1)	0.149 (0)	0.837 (1)	0.014 (0)
Ningbo (16)	0.016 (0)	0.970 (16)	0.014 (0)
Zhoushan (2)	0.085 (0)	0.314 (1)	0.601 (1)
Pingtan (72)	0.031 (0)	0.302 (29)	0.667 (43)
Dongshan (24)	0.050 (0)	0.074 (1)	0.876 (23)
Beihai (6)	0.017 (0)	0.133 (0)	0.850 (6)

NOTE.—The posterior probability of the number of populations in the sample set was maximum with $K = 3$ as prior population information based on the STRUCTURE analysis. The table shows the proportion of membership of each sampled site in each of three inferred clusters. The values in the parentheses refer to the number of samples in each locality or in the relevant cluster.

generation migrants might be hybrids between different populations due to recent migration.

Discussion

Genome Scans in Finless Porpoises

In this study, we isolated a series of SNP markers by utilizing shotgun library approach (Staden 1979) and CATS method (Aitken et al. 2004; Li et al. 2009), both of which have been considered as efficient mechanisms for SNP discovery at the whole-genome level. 380 SNPs were found from 138,434 bp of high-quality sequences, with a frequency of one SNP per 364 bp, which was of the same order of magnitude as those in human (1/300–1,000 bp; International SNP Map Working Group 2001) and wolf (1/306 bp; Seddon et al. 2005). However, the present SNP density is higher than the correspondent value of approximately 1/1,500 bases in the dog (Kirkness et al. 2003) and lower than the observed value in the eastern fence lizard (4.4/100 bp; Rosenblum and Novembre 2007). This is not very surprising because differences in SNP density may occur as a result of detection method and species specificity (Brumfield et al. 2003), and is also obviously dependent on the samples used for SNP discovery.

The present estimate of genetic diversity with SNPs revealed an average observed heterozygosity of $H_E = 0.3405 \pm 0.1188$. This value was much lower than those estimated with microsatellite markers ($H_E = 0.732$–0.795) (Chen et al. 2010). However, such a direct comparison of heterozygosity estimates between different markers can be misleading considering that SNPs have their own distinctiveness such as diallelic nature and relatively low mutation rates (10^{-8}–10^{-9}; Brumfield et al. 2003). In contrast with the differences between markers in the same species, the present average H_E in the finless porpoise fell in the similar range of those observed in other species such as humans ($H_E = 0.199$–0.493, Beaty et al. 2005), white spruce ($H_E = 0.266$–0.274, Namroud et al. 2008), and so forth. One reason which might cause deviation from HWE is probably due to the loss of heterozygotes, which might be explained by the habitat fragmentation leading to further division of each population into small subpopulations causing an increase of homozygotes due to genetic drift and endogamy. As each population was sampled from several localities, with small numbers of individuals in some localities, different allele frequencies in the populations would cause a Wahlund effect, which might also result in deviation from expectation (Wahlund 1928).

Bayesian clustering and PCA all revealed significant structure in the finless porpoise in Chinese waters. Especially, the STRUCTURE analysis provided strong support for the subdivision of finless porpoises in Chinese waters into genetically differentiated populations, subspecies or species as suggested in previous studies (Gao and Zhou 1995a, 1995b, 1995c; Yang et al. 2002, 2003, 2008; Chen et al. 2010). Moreover, STRUCTURE analysis with GENSBACK = 4 and MIGRPRIOR = 0.1 also suggested little evidence of genetic exchange even in the sympatric area of the Yellow Sea

population and the South China Sea population (e.g., Pingtan of the Taiwan Straits). Significant F_{ST} values ranging from 0.1050 between the Yangtze River population and the Yellow Sea population to 0.1628 between the Yangtze River population and the South China Sea population further supported that these populations might have been genetically isolated especially between the freshwater and marine porpoises, which were overall comparable with those reported with mitochondrial DNA and nuclear microsatellite markers (Yang et al. 2002, 2003, 2008; Chen et al. 2010).

Potential Demographic History of Finless Porpoises in Chinese Waters

Predicting the genetics consequences of small population size has been one of the major tasks of conservation biology, because a reduction can result in inbreeding, loss of genetic variation and fixation of deleterious alleles, and thereby decrease the evolutionary potential and increase the probability of population extinction (Maudet et al. 2002). The effects of population bottlenecks are directly related to the increase of stochastic events associated with small population size. Our data exhibit a heterozygosity excess than a heterozygosity deficiency under the IAM and the SMM models of mutation. This indicated that the finless porpoise in Chinese waters had encountered a recent population bottleneck. Since the Pleistocene of the Quaternary Period, the marine environment of coastal China has changed dramatically, along with the periodical change of global climatic condition, and the advance and retreat of glacier (Jin 1985). As discussed in previous studies (e.g., Yang et al. 2008), in the climax of the Dali Ice Age, the continent shelf in the Yellow/Bohai Seas, the East China Sea, and the majority of the South China Sea were exposed. In addition, the finless porpoise inhabits the coastal waters and thus could be easily influenced by various human activities, such as directed and/or incidental catches, habitat degradation caused by water contamination, and so forth. These geological events and anthropogenic disturbance might have some impacts on the finless porpoises resulting in a natural population bottleneck. Furthermore, limited sample sizes in some populations used in our study may also have affected the detection of recent population declines. Further studies with more systematic sampling and larger sample size are urgently needed.

Taxonomic Implications and Conservation Recommendations Especially for the Yangtze Finless Porpoise

Traditionally, the finless porpoise (genus *Neophocaena*) was widely recognized as comprising a single species, although distinct morphological forms of this species have been known for some time (Wang 1992). Three subspecies have been identified based upon external morphology, craniometrics, and distribution (Gao and Zhou 1995a, 1995b, 1995c; Rice 1998). Moreover, recent studies all supported the hypothesis originally proposed by Pilleri and Gihr (1972) who suggested that the "wide" (*phocaenoides*) and "narrow" forms were differentiated enough to warrant separate species status; however, the further subdivision of the "narrow" form into the *asiaeorientalis* and *sunameri* subspecies was not supported (Jefferson 2002; Jefferson and Hung 2004; Wang et al. 2008, 2010).

In a very recent study, Chen et al. (2010) not only supported the significant genetic differentiation between the wide-ridged (*N. phocaenoides*) and narrow-ridged (*N. asiaeorientalis*) finless porpoises as suggested by Wang et al. (2008) but also provided evidence for the further subdivision within narrow-ridged form of the finless porpoise. The earlier findings were further corroborated from a genome-wide analysis of SNPs in this study. The STRUCTURE, PCA, and AMOVA analyses (figs. 3–5; table 3) all supported significant genetic divergence not only between two marine populations overlapping in the Taiwan Straits but also between the Yellow/Bohai Sea and the Yangtze freshwater population.

Although the F_{ST} value between the Yangtze River population and the Yellow Sea population was relatively lower than any other population-pair comparison (all having the same level of statistical significance), limited gene flow was found especially between the freshwater and the marine porpoises, which suggested that some barriers might have restricted their genetic exchange. Now that the South China Sea population and the Yellow Sea population have been regarded as genetically differentiated species, the Yangtze finless porpoise, having a significant genetic variation and limited gene flow between the oceanic populations, in combination with its unique distribution in the Yangtze River and physiological (Ni and Zhou 1988) and morphological (Gao and Zhou 1995a, 1995b, 1995c) distinctness, should also be given separate species status. It is noteworthy that although Jefferson (2002) and Jefferson and Hung (2004) did not find convincing morphological evidence to differentiate populations within the narrow form, Gao and Zhou (1995a, 1995b, 1995c) could reliably distinguish the Yangtze finless porpoise from other populations with stepwise discriminant analysis of skeletal morphology. This illustrates the necessity to further examine morphometrics to reveal potential distinctive diagnostic features of the Yangtze finless porpoise.

The Yangtze finless porpoise is the sole freshwater-adapted porpoise population. Recent surveys indicated that this population decreased in size to approximately 1,000 individuals (Wang et al. 2006), less than half the 1990s' level (Zhang et al. 1993). Population viability analysis also suggested that this population will go extinction in 100 years if no effective protection is taken (Zhang and Wang 1999). The finless porpoise was listed as "vulnerable" on the IUCN "red list" database of the endangered species, and the Yangtze River population was even characterized as "endangered" due to its apparent decline in wild population size (Reeves et al.

2008). The conservation of the Yangtze finless porpoise has attracted worldwide attention especially along with the near-extinction of the Baiji that is sympatric with the Yangtze finless porpoise in the middle and lower reaches of the Yangtze River. The Chinese government is now considering to upgrade the conservation rank of the finless porpoise from the current National II to National I. Obviously, the discovery of the significant genetic differentiation revealed that the use of a genome-level marker provide an important basis not only for reconsidering its taxonomic status, but also more importantly for enhancing its conservation. The Yangtze finless porpoise represents an important component not only of the finless porpoise but also of the Yangtze freshwater biodiversity. Its extinction will be a sign of the disappearance of a distinct aquatic mammal species that has special adaptation to the freshwater environment. Thus, it is urgently recommended to raise its conservation priority and design some special programs for this population or species in the very near future.

Future Outlook

This study represents a successful development and application of SNPs in population biology and conservation of an endangered aquatic mammal. We isolated a series of genome-wide SNPs to provide novel insights into genetic diversity, population structure, and genetic divergence of finless porpoises.

However, considering current technical limitations and laboratory difficulties studying nonmodel organisms with SNP, only a relatively small number of SNPs were available in this study. In the future, we will still need to detect more loci for a finer resolution of population and conservation genetics of finless porpoises, especially when the next-generation sequencing methods become available, and comprehensive SNP data sets is becoming an efficient and cost-effective genetic tool. Although a SNP panel has been identified from some functionally known genes, more such SNPs are still necessary to give us a comprehensive understanding of population structure, adaptive divergence, and some other associated genetic studies of finless porpoises in the future.

Supplementary Material

Supplementary table S1 is available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) key project (grant number 30830016), the Program for New Century Excellent Talents in University (grant number NCET-07-0445), the Ministry of Education of China, the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) to G.Y., and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (grant number 10KJB180002).The authors thank Dr. Anli Gao, Prof. Qing Chang, Mr. Xinrong Xu, and some students who have studied or are studying at NJNU for their assistance in porpoise sample collection.

Literature Cited

Aitken N, Smith S, Schwarz C. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol Ecol. 13:1423–1431.

Beaty TH, Fallin MD, Hetmanski JB. 2005. Haplotype diversity in 11 candidate genes across four populations. Genetics 171:259–267.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol Evol. 18:249–256.

Chen L, Bruford MW, Xu SX, Zhou KY, Yang G. 2010. Microsatellite variation and significant population genetic structure of endangered finless porpoises (*Neophocaena phocaenoides*) in Chinese coastal waters and the Yangtze River. Mar Biol. 157:1453–1462.

Choi IY, et al. 2007. A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176: 685–696.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14:2611–2620.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver.3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online. 1:47–50.

Gao AL, Zhou KY. 1993. Growth and reproduction of three populations of finless porpoise, *Neophocaena phocaenoides*, in Chinese waters. Aquat Mammals. 19:3–12.

Gao AL, Zhou KY. 1995a. Geographical variation of external measurements and three subspecies of *Neophocaena phocaenoides* in Chinese waters. Acta Theriol Sin. 15:81–92.

Gao AL, Zhou KY. 1995b. Geographical variation of skull among the populations of *Neophocaena phocaenoides* in Chinese waters. Acta Theriol Sin. 15:161–169.

Gao AL, Zhou KY. 1995c. Geographical variation of postcranial skeleton among the populations of *Neophocaena phocaenoides* in Chinese waters. Acta Theriol Sin. 15:246–253.

International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933.

Jefferson TA. 2002. Preliminary analysis of geographic variation in cranial morphometrics of the finless porpoise (*Neophocaena phocaenoides*). Raffles Bull Zool. 10(Suppl.):3–14.

Jefferson TA, Hung SK. 2004. *Neophocaena phocaenoides*. Mammal Species. 746:1–12.

Jin X. 1985. Evolvement of ancient climate and oceans in Pliocene and the Quaternary Period. In: The Department of Oceanography and Geology, editor. Fundamentals of ancient oceanography. Shanghai: Publishing House of Tongji University. p. 1-225.

Kirkness EF, et al. 2003. The dog genome: survey sequencing and comparative analysis. Science 301:1898–1903.

Li SZ, Wan HR, Ji HY, Zhou KY, Yang G. 2009. SNP discovery based on CATS and genotyping in the finless porpoise (*Neophocaena phocaenoides*). Conserv Genet. 10:2013–2019.

Lyons LA, et al. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat Genet. 15: 47–56.

Maudet C, et al. 2002. Microsatellite DNA and recent statistical methods in wildlife conservation management: applications in *Alpine ibex* (*Capra ibex* (ibex)). Mol Ecol. 11:421–436.

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

Morin PA, Luikart G, Wayne RK, SNP workshop group. 2004. SNPs in ecology, evolution and conservation. Trends Ecol Evol. 19: 208–216.

Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. Mol Ecol. 17: 3599–3613.

Narum SR, et al. 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Mol Ecol. 17:3464–3477.

Ni JY, Zhou KY. 1988. Rencular structural indices and urinary concentrating capacity of *Neophcaena phocaenoides*. Acta Zool Sin. 34:46–53.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet. 2:2074–2093.

Pilleri G, Gihr M. 1972. Contribution to the knowledge of the cetaceans of Pakistan with particular reference to the genera Neomeris, Sousa, Delphinus and Tursiops and description of a new Chinese porpoise (*Neomeris asiaeorientalis*). In: Pilleri G, editor. Investigations on Cetacea. Vol. 4. Bern (Switzerland). p. 107–162.

Piry S, Luikart G, Cournet JM. 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered. 90:502–503.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Raymond M, Rousset F. 1995. Genepop (version 1.2): population genetics software for exact tests and ecumenicism. J Hered. 86:248–249.

Reeves RR, et al. 2008. *Neophocaena phocaenoides*. In: IUCN 2008, IUCN red list of threatened species. Gland (Switzerland) and Cambridge (UK): IUCN.

Reeves RR, Smith BD, Crespo EA, Notarbartolo di Sciara G, editors. 2003. Dolphins, whales and porpoises: 2002-2010 conservation action plan for the world's Cetaceans. Gland (Switzerland) and Cambridge (UK): IUCN, p. 139.

Rice DW. 1998. Marine Mammals of the World: Systematics and Distribution (Special Publication Number 4). Lawrence (KS): The Society for Marine Mamalogy.

Rice WR. 1989. Analyzing tables of statistical tests. Evolution 43:223–225.

Rosenblum EB, Novembre J. 2007. Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. J Hered. 98: 331–336.

Seddon JM, Parker HG, Ostrander EA, Ellegren H. 2005. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Mol Ecol. 14:503–511.

Staden R. 1979. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 6:2601–2610.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in mirosatellite data. Mol Ecol Resour. 4:535–538.

Wahlund S. 1928. Zusammensetzung von Populationen und Korrelationsers-chinungen von Standpunkt der Vererbungslehre aus betrachtet. Hereditas 11:65–106.

Wang JY, Frasier TR, Yang SC, White BN. 2008. Detecting recent speciation events: the case of the finless porpoise (genus *Neophocaena*). Heredity 101:145–155.

Wang JY, Yang SC, Wang BJ, Wang LS. 2010. Distinguishing between two species of finless porpoises (*Neophocaena phocaenoides* and *N. asiaeorientalis*) in areas of sympatry. Mammalian 74:305–310.

Wang KX, Wang D, Zhang XF, Pfluger A, Barrett L. 2006. Range-wide Yangtze freshwater dolphin expedition: the last chance to see Baiji? Environ Sci Pollut R. 13:418–424.

Wang P. 1992. The morphological characters and the problem of subspecies identifications of the finless porpoise. Fish Sci. 11:4–9.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Yang G, et al. 2002. Population genetic structure of finless porpoises, *Neophocaena phocaenoides*, in Chinese waters, inferred from mitochondrial control region sequences. Mar Mamm Sci. 18: 336–347.

Yang G, Guo L, Bruford M, Wei FW, Zhou KY. 2008. Mitochondrial phylogeography and population history of finless porpoises in Sino-Japanese waters. Biol J Linn Soc. 95:193–204.

Yang G, Liu S, Ren WH, Zhou KY, Wei FW. 2003. Mitochondrial control region variability of Baiji and the Yangtze finless porpoises, two sympatric small cetaceans in the Yangtze River. Acta Theriol. 48: 469–483.

Yoshida H, Yoshioka M, Shirakihara M, Chow S. 2001. Population structure of finless porpoises (*Neophocaena phocaenoides*) in coastal waters of Japan based on mitochondrial DNA sequences. J Mammal. 82:123–130.

Zhang XF, et al. 1993. The population status of finless porpoise in the middle and lower reach of the Yangtze River. Acta Theriol Sin. 13: 260–270.

Zhang XF, Wang KX. 1999. Population variability analysis of the Yangtze finless porpoises. Acta Theriol Sin. 19:529–533.

Associate editor: Bill Martin