Genetic footprint of population fragmentation and contemporary collapse in a freshwater cetacean

Minmin Chen[1,2†], Michael C. Fontaine[3†]*, Yacine Ben Chehida[3†], Jinsong Zheng[1]*, Zhigang Mei[1], Yujiang Hao[1], Kexiong Wang[1], Min Wu[1,2], Qingzhong Zhao[1], Ding Wang[1]*

[1] The Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology of Chinese Academy of Sciences, Wuhan 430072, China;

[2] The Key Laboratory of Aquatic Organism Conservation and Water Ecosystem Restoration in Anhui Province, College of Life Science, Anqing Normal University, Anqing 246133, China

[3] Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, PO Box 11103 CC, Groningen, The Netherlands;

[†] These authors contributed equally and should be considered as co-first authors.

*** Corresponding authors:** Michael C Fontaine (mikafontaine@gmail.com); Ding Wang (wangd@ihb.ac.cn); Jinsong Zheng (zhengjinsong@ihb.ac.cn)

Running title: Genetics of the Yangtze Finless Porpoise

Keywords: demographic collapse; extinction dynamics; river dolphin; river connectivity; gene flow; Approximate Bayesian Computation

Abstract

Understanding demographic trends and patterns of gene flow in an endangered species, occupying a fragmented habitat, is crucial for devising conservation strategies. Here, we examined the extent of population structure and recent evolution of the critically endangered Yangtze finless porpoise (YFP, *Neophocaena asiaeorientalis asiaeorientalis*). By analysing genetic variation at the mitochondrial and nuclear microsatellite loci for 153 individuals, we identified 3 populations along the Yangtze River, each one connected to a group of admixed ancestry. Each population displayed extremely low genetic diversity, consistent with extremely small effective size (≤106 individuals) and census sizes. Habitat degradation and distribution gaps correlated with highly asymmetric gene-flow that was inefficient in maintaining connectivity between populations. Genetic inferences of historical demography revealed that the populations in the Yangtze descended from a small number of founders colonizing the river from the sea during the last Ice Age. The colonization was followed by a rapid population split during the last millennium predating the Chinese industrial revolution. However, genetic diversity showed a clear footprint of population contraction over the last 50 years leaving only 2% of the pre-collapsed size, concomitant with the industrial revolution. These results provide background information for devising mitigation strategies to prevent YFP extinction.

1. Introduction

Dispersal and gene flow in a meta-population maintain local demographic and genetic variation, thus increasing the probability of species persistence [1,2]. Persistence of wide-ranging animals occupying fragmented landscapes depends on the matrix quality of the habitat and the ability of individuals to move among habitat patches [3], and corridors facilitate this movement [4-6]. Along the Yangtze River (China), industrial activities of the past 50 years have put intense pressure on the freshwater ecosystem leading to habitat degradation, species range fragmentation and extinction of some emblematic endemic species such as the Yangtze River dolphin or baiji (*Lipotes vexillifer*) [7]. Today, the Yangtze finless porpoise (*Neophocaena asiaeorientalis* spp. *Asiaeorientalis;* YFP) has become the only surviving freshwater cetacean now found in China and the world's only freshwater porpoise species [8].

Endemic to the Yangtze River drainage, the YFP is now primarily restricted to the middle-lower Yangtze Channel and two large appended lakes: Dongting Lake (DT) and Poyang Lake (PY) [9]. The subspecies has occasionally been reported from some of the larger adjacent tributaries (electronic supplementary material (ESM) figure S1) though this is now rare [10-12]. The amount of river and lake habitat available to this subspecies is relatively small compared to that available to marine populations of finless porpoises, which occur in coastal waters from Japan to the Arabian Sea [8]. However, YFP abundance has suffered from dramatic reductions from 2,500 individuals in 1991 [10] to 1,800 in the end of 2006 [13]. More recent surveys conducted during the Yangtze Freshwater Dolphin Expedition conducted in 2012 (YFDE2012) reported that populations declined even further to ~1,040 individuals, including ~500 porpoises in the Yangtze main stream, ~450 in PY and ~90 in DT [14]. With such rapid range contraction, Mei *et al.* [15] estimated that the YFP may become extinct within the next 60 years or less [14]. The YFP was thus recently reclassified as a Critically Endangered population on the IUCN Red List [9]. As a top predator, the survival of the finless porpoise depends heavily on habitat stability, food availability and maintenance of corridors allowing dispersal between populations. However, food and habitat resources for the species have become increasingly scattered and fragmented, and corridors across the landscape have been compromised by the booming of the Chinese economy over the last decades [13]. Despite these imminent threats, we still don't know how reduction in suitable habitats in the Yangtze River has reduced the number of breeding porpoises and how habitat fragmentation has impacted connectivity between population and the population structure

77 itself. This information is extremely difficult to quantify using direct observations. On the
78 other hand, population genetic approaches can provide key insights about historical
79 population demography, structure and connectivity by leaving detectable footprints on the
80 genetic diversity and its geographic structure [16,17].
81
82 Previous phylogeographic analyses based on the mitochondrial Control Region
83 (mtDNA-CR) of the finless porpoises from Chinese and Japanese waters documented
84 evidence of a demographic expansion following the Last Glacial Maximum (LGM, ~24,000
85 yrs BP) and the colonization of Yangtze River from the Yellow Sea ~22,000 yrs BP ago
86 [18,19]. A subsequent mtDNA study within the Yangtze River itself revealed subdivisions
87 within the YFP populations suggestive of population fragmentation [20]. However, the fine
88 scale population structure, connectivity between populations, dynamics of population
89 expansion-contraction and demographic history of the YFP were not investigated further. In
90 the case of the YFP it is of interest to know whether the fragmented population structure and
91 ongoing decline have been recently triggered by human activities during the past 50 years or
92 were initiated earlier and exacerbated during the Anthropocene. We address those questions
93 by analysing a combination of fast and slow evolving genetic markers in order to capture
94 demographic events at different time scales [16,21].
95

96 ## 2. Material & Methods

97 ### *(a) Genetic dataset*

98 Genomic DNA was obtained from 153 finless porpoises sampled in the Yangtze River
99 between 1998 and 2011 (figure 1 and ESM figure S1) as described in [20]. For each sample,
100 nuclear microsatellite genotypes for 11 loci and a 597 base-pairs fragment of the
101 mitochondrial control region (mtDNA-CR, n=129) have been previously described (see
102 details in [20]).

103

104 ### *(b) Population genetic structure*

105 We used the Bayesian model-based clustering of *STRUCTURE* v2.3.4 [22] to estimate the
106 admixture proportions for each individual to each cluster identified in the microsatellite data.
107 *STRUCTURE* cluster individual multilocus genotypes into *K* groups while minimizing
108 departures from Hardy-Weinberg and Linkage Equilibria. We used the admixture *Locprior*
109 model with correlated allele frequencies designed to detect weak signals of genetic structure

without introducing bias or forcing the clustering [22]. We conducted a series of independent runs with different value for K from 1 to 5. Each run used $1x10^6$ iterations after a burn-in of length $2x10^5$. To assess that convergence of the Monte Carlo Markov Chains (MCMCs) had been reached, we performed 10 independent replicates for each K and checked the consistency of results using *CLUMPAK* [23]. We assessed which K value best fit with the data using (1) the likelihood of each K, following *STRUCTURE*'s user manual; (2) its rate of change with increasing K [24]; and (3) visual inspection of newly created clusters with increasing K. Post-processing of the results, including generation of barplots, was conducted using *CLUMPAK* [23]. The geographic distribution of each group was mapped using the R statistical package *MARMAP v.0.9.5* [25] and ETOPO dataset [26].

The genetic structure in the microsatellite data was further inspected using a Principal Component Analysis (PCA). This exploratory method does not rely on any model assumptions and provides a complementary validation of the structure depicted by *STRUCTURE* [27]. This analysis was conducted using *adegenet 1.4-2* package [28] for the R [29] on centred data, with missing data replaced by the mean.

(c) Genetic diversity and differentiation

We compared microsatellite genetic diversity between populations using the allelic richness (A_r) and private A_r (pA_r) computed with *ADZE* [30], and the observed and expected heterozygosity (H_o and H_e) computed with *GenAlEx* v6.5. [31]. Departures from Hardy-Weinberg were tested using 10^4 permutations and quantified using F_{IS} and F_{ST} [32] in *FSTAT* v2.9.3.2 [33]. We applied a Bonferonni correction to correct for multiple tests.

Mitochondrial diversity was originally reported in [20]. Here, we estimated the genetic diversity for the groups identified in the present study. Variation among sequences was measured using the number of segregating sites (S), number of singletons and shared polymorphisms, number of haplotypes, haplotype diversity (H_d) and two estimators of population genetic diversity, π based on the average number of pairwise differences [34] and θ_W based on the number of polymorphic sites [35]. All statistics were calculated using DNASP v5.10.01 [36]. We used the F_{ST} statistics estimated from the average number of differences within and between populations [37]. Significance was tested with 1000 permutations of Hudson's nearest neighbour distance *Snn* statistics, which measures how

143 often the nearest neighbour of a sequence (in sequence space) is from the same population
144 [38].
145
146 ***(e) Contemporary effective population sizes***
147 We used two methods for estimating contemporary effective population size (*Ne*) from
148 microsatellite data. The first method is implemented in *NeEstimator* v2.01 [39] and relies on
149 linkage disequilibrium (LD) between loci, filtering out rare alleles with a frequency $P_{crit} \leq$
150 0.02 that could bias *Ne* estimate [40]. The second method implemented in *ONESAMP* v1.2
151 [41] also use LD as summary statistics among others in an Approximate Bayesian
152 Computation (ABC) to estimate *Ne*, considering uniform prior between 2 and 500 of *Ne*.
153
154 ***(f) Contemporary gene flow between populations***
155 We estimated contemporary effective migration rate (*m*) between populations using *BayesAss*
156 v.3.0.3 [42]. Preliminary runs were performed to adjust the mixing parameters of the MCMC
157 and ensure proposal acceptance rates between 20% and 60% following authors'
158 recommendations. We then performed 12 independent runs with different seeds, a burn-in of
159 $5x10^6$ iterations followed by $2x10^8$ iterations, and a sampling parameter values every 2000
160 iterations. Convergence of the MCMCs was checked by comparing the traces of each run
161 using *Tracer* v1.5 [43] and by evaluating the Effective Sample Sizes (ESSs) of each
162 parameter, keeping only runs where ESS $\geq$ 200. Model fitting to the data was assessed using
163 the Bayesian Deviance Index using the R-script of Meirmans [44]. Runs that converged were
164 combined to estimate the mean, median and 95% Highest Probability Density interval for
165 each parameter in *Tracer* v1.5. The effective number migrants ($N_e.m$) per generation between
166 populations was obtained by combining *Ne* and *m* estimates.
167
168 ***(g) Genetic inference of population demographic history***
169 Departure of the allelic or haplotypic frequency spectrum of microsatellite and mtDNA loci
170 respectively from constant population sizes hypothesis can provide evidence of population
171 size change. For microsatellite loci, we used the M_{GW} ratio of the number of alleles to the
172 range in allele sizes to detect evidence of reductions in *Ne* [45]. We used DIYABC v2.1 [46]
173 to estimate the M_{GW} value and conduct $1x10^6$ coalescent simulations to produce a null
174 distribution against which the observed value could be compared. The *P*-value indicates the
175 proportion of simulations which provide a value below the observed one. For the mtDNA-CR,
176 we used the Tajima's *D* [47] and tested for significant departure from a null expectations

using 10,000 coalescent simulations in *DNAsp* [36].

Next, we investigated the demographic history best describing the genetic diversity of the combined microsatellite and mtDNA markers using a coalescent-based ABC approach [48]. We stratified our workflow in two parts (figure 2): (i) identify the most likely population tree topologies for our dataset among 10 scenarios describing different potential population topologies (figure 2a); and (ii) test for evidence of changes in effective population size in the ancestral and daughter populations (figure 2b). For each part, an ABC analysis was conducted using the program DIYABC v.2.1.0 [46] applying the following steps (ESM figure S2): (1) coalescent simulations of $1x10^6$ pseudo-observed datasets (PODs) for each competing scenario and calculation of summary statistics (SS) describing the observed genetic variation for each POD; (2) select the best model by estimating the posterior probability (PPr) of each scenario using a logistic regression on 1% PODs producing SS values closest to the observed ones; (3) evaluate the confidence in scenario choice by estimating the type-I and type-II error rates based on simulated PODs; (4) estimate the marginal posterior distribution of each parameter based on the best model(s) including (among other) *Ne* and times of population size changes and splits; and finally, (5) evaluate the goodness-of-fit of the fitted model to the data. Details are provided in figure 3 and in ESM (Text S1, Figure S2, table S1, S2).

3. Results

(a) Genetic structure of the Yangtze finless porpoises

The clustering of *STRUCTURE* provided consistent results over the replicated runs performed for each *K* tested (figure 1a). The probability of the data greatly increased when 2 clusters were modelled instead of 1 and showed the highest values on average over 10 replicates (ESM figure S3). At $K>2$, the increase in probability decreased; however some runs displayed the highest probability of all runs at $K=3$ and 4 (ESM figure S3). A visual inspection of the clustering for each *K* value (figure 1a) revealed that porpoises from the PY split from the other individuals of the Yangtze river at K=2, suggesting that these porpoises are highly differentiated from the others. When higher *K* value were tested (K=3 and 4) porpoises from PY, XCSS and TL were all identified as distinct genetic units, while the porpoises from the in-between regions appeared to be an admixed group sharing genetic ancestry with the three distinct populations (figure 1a and 1c). The three individuals at the mouth of the Yangtze River (SH) also seemed to depart from the other groups at K=4, but the

low sample size preclude any definitive conclusions. No further subdivision was observed beyond K=4.

The PCA showed a genetic structure consistent with the results of *STRUCTURE* (figure 1b). Individuals from PY, XCSS and TL were separated by the first two PCs, spreading in three orthogonal directions. Such a pattern is expected when 3 differentiated genetic pools are present and evolved independently from each other by genetic drift [49]. The other porpoises recognized as admixed in *STRUCTURE* were located at the intersection of the 3 groups.

(b) Genetic diversity and differentiation

None of the identified populations displayed significant departures from Hardy–Weinberg expectations as shown by the F_{IS} values (table 1). Porpoises from the admixed and TL groups showed the highest level of microsatellite genetic diversities and XCSS and PY the lowest, as estimated with the values of allelic richness (*Ar*), private allelic richness (*pA*) and expected heterozygosity (H_e) (table 1 and ESM figure S4). Only the two extremes groups, PY and Admix, showed a significant difference in A_r and H_e (Wilcoxon signed-ranked test *p-value* < 0.05). The mitochondrial genetic diversity followed a similar trend with the admixed group showing the highest haplotype and nucleotide diversity followed by TL and PY. We found only one haplotype fixed in the XCSS group (table 1).

All populations showed significant differences in allelic frequencies for the microsatellite loci with F_{ST} values ranging from 0.021 to 0.068 (table 2). Notably F_{ST} values between XCSS, PY and TL were relatively high (>0.05), while F_{ST} values were intermediate between the Admix groups and each distinct population (between 0.02 and 0.03, table 2). For the mitochondrial locus (table 2), F_{ST} values were all significant except one (Admix *vs.* TL), and were especially strong between XCSS and all other groups, due to the fact that one haplotype is fixed in this population (figure 1c).

(c) Contemporary effective population sizes and migration rates

Estimates of contemporary *Ne* for each population were comparable between *NeEstimator* and *ONESAMP* programs (figure 3 and ESM table S3). All values were very low (<106 individuals), with the Admix group displaying the highest estimate, followed by PY, TL and XCSS.

We estimated migration rates (*m*) between groups over the last few generations using *BayesAss*. Out of the 12 repeat runs, 10 showed good mixing properties with Bayesian Deviance values close to each other (mean ± SD: 8,915.28 ± 0.33) and convergent estimates for each parameter. We thus combined them all to estimate the parameter values (ESM table S4). The number of effective migrants (*Ne.m*) per generation over the last generations were obtained by combining *Ne* estimates from *ONESAMP* with *m* estimates of *BayeAss* (figure 3). The three populations – TL, XCSS and PY – did not show any evidence of recent migration among each other, as indicated by the lower bound of the 95% HPD interval equal to 0 (figure 3 and ESM table S4). However, each one is connected to the Admix group with highly asymmetric gene flow. We detected significant unidirectional gene flow from PY to Admix and from Admix to TL and XCSS. *Ne.m* values from the Admix group to TL or XCSS are about the half of those from PY to the Admix group.

(d) Population demographic history

Genetic diversity at the microsatellite markers showed evidence of significant *Ne* contraction in each population, as suggested by the very small M_{GW} value, significantly smaller than expected under the assumption of constant population size (table 1). On the other hand, the Tajima's *D* values (table 1) estimated for the mtDNA-CR in each population did not depart from the constant population size hypothesis.

The ABC analysis showed that out of the 10 scenarios tested (figure 2*a*), the trichotomy (SC1), which assume that XCSS, PY and TL diverged at the same time, received the highest support with a probability of 67% and a 95%CI not overlapping with any other scenarios. The performance analysis for this ABC step (ESM table S5 and S6) showed that more than 61.4% of the simulations under SC1 were correctly identified, leading to a very low average Type-I error rate (false negative) of 4.3% ranging from 1.7% to 9.7%. Simulations under the nine other competing scenarios led to a Type-II error rate of <12.9% incorrect assignment to SC1 (false positives) and a very good power (87.1%) to discriminate the best scenario from the others.

The second ABC step (figure 2*bi*) tested for occurrence of changes in *Ne* during the divergence of the three populations under six scenarios. The scenario assuming a bottleneck in the ancestral population prior to population split (SC3) outcompeted the other five scenarios (i.e., no change (SC1) or simple decline or expansion in the ancestral or daughter

278 populations, SC 2, 4-6), with a probability of 78% and no overlap in 95%CI. The
279 performance analyses (ESM table S7) confirmed the power of this second ABC step to
280 discriminate among the six scenarios tested. The Type-II error rate (false positive) was only
281 7.2% and the power to discriminate this scenario from the others was 92.8%. The sensitivity
282 analysis showed that simulations generated with SC3 were more difficult to identify, with an
283 average Type-I error rate of 12.4% (ranging between 1 and 21% error, depending on the
284 scenario). The goodness-of-fit of this bottleneck scenario (SC3) to the data was also very
285 good, as the simulations using this scenario and posterior distributions for each parameter,
286 were able to reproduce all but one observed summary statistic, in contrast to all other
287 scenarios (ESM table S7 and S8).
288
289 The final step (figure 2*bii* and ESM table S1) tested whether a recent demographic collapse in
290 each population within the last 5 generations (or 50 years [15]) could have produced a
291 detectable genetic footprint. The recent collapse scenario (SC2) received a significantly
292 higher probability (78.2%) compared with the alternative scenario of constant *Ne* since the
293 population split (figure 2 and ESM table S9 and S10). Both Type-I and Type-II error rates
294 were <15% indicating adequate power and sensitivity of our ABC analysis. The estimated
295 parameters under SC2 (figure 2*bii*, ESM table S11 and S12) suggest that the ancestral
296 population would have been large (N_{anc4} = 18,700 individuals; 95%CI: [3,480 – 19,700]), and
297 that a small fraction would have founded the Yangtze River populations ~3,400 95%CI:
298 [1300 – 41,000] years ago (T_{exp4}) and expanded to reach an effective size of about 5,660
299 individuals (N_{exp4}; 95%CI: [2,900 – 9,840]). The three daughter populations would have then
300 split from each other ~1,030 years ago (T_{isol4}; 95%CI: [214 – 4,380]), and reached an
301 effective size of about 2,000 individuals after their split. Each of these daughter populations
302 would have gone through a significant decline leaving less than 2% of their pre-collapse size
303 them during the last 50 years (ESM table S11 and S12).
304

305 4. Discussion

306 Our study shows that the present-day genetic diversity of finless porpoise in the Yangtze
307 River (YFP) has been strongly influenced by an initial founder event that took place several
308 thousand years ago, followed by a population split into 3 populations (XCSS, TL, PY), and a
309 recent demographic collapse within the last 50 years.
310

311 Consistently with previous studies [18,19], the ABC genetic inferences (ESM table S11)
312 showed that a few individuals coming from a large ancestral population, likely a marine
313 population of *N. a. sunameri* in the Yellow Sea, colonized the Yangtze River within the last
314 41kyrs, and most likely during the Last Ice Age. The subsequent population split into three
315 populations occurred between 200 and 5000 years BP. This suggests that the population split
316 was triggered before the intensification of human activities of the last 50 years in the Yangtze
317 River. This event may thus have been related to the colonization process itself and/or other
318 environmental or human-related factors.
319
320 Each population harboured a genetic footprint of dramatic population reduction. This
321 especially affected the microsatellite markers which displayed very small values of the M_{GW}
322 statistic characteristic of significant recent decline [45]. This is also indicated by the ABC
323 analysis supporting a scenario describing a drastic population reduction in each population
324 within the last five generations (figure $2b_{ii}$, SC2). According to this scenario, the reduction in
325 effective size would have been massive since their current sizes would be less than 2% of
326 their pre-collapsed sizes (ESM table S3, S11 and S12). These genetic inferences are in line
327 with the field estimates [14,15,50], reporting a continuous decline of the YFP since the 1980s
328 [10-15,51,52]. Half of the census populations in the main stem of the Yangtze River would
329 have been lost within the past 15 years, with the abundance dropping from 2500 porpoises in
330 1991 [10] to 1225 in 2006 [13]. With the 400 porpoises in the Poyang Lake and the 100 to
331 150 porpoises of the Dongting Lake, the total 2006 census size in the Yangtze River and the
332 two adjacent lakes was only 1800 individuals. The most recent YFDE2012 estimate showed
333 that the porpoises in the main stem of the Yangtze would have been reduced again by half
334 with 505 individuals reported the middle and lower reaches of the Yangtze River and
335 approximately 450 porpoises in PY, and 90 porpoises in DT (Wang D. et al., unpublished
336 data).
337
338 These small estimates of very low census populations sizes (*N*) are consistent with the very
339 small estimates of contemporary effective population size (*Ne*) from genetic (≤106
340 individuals) in each population, with the XCSS population being the smallest of all three
341 differentiated populations (between 13 and 22 individuals, ESM table S3). Comparably low
342 estimates have been reported in other cetacean species, such as the southern Iberian ecotype
343 of harbour porpoise (*Phocoena phocoena meridionalis*) in European waters ($Ne \leq 80$ [17,53])

or the coastal ecotype of bottlenose dolphins (*Tursiops truncatus*) in European waters ($Ne \leq$ 77, [54]). However only highly endangered populations, such as the Borneo Orang-utans (*Pongo pygmaeus*) [55] or the Canadian woodland caribou (*Rangifer tarandus*) [56], have *Ne* values as low as those observed in the XCSS population. Extremely low *Ne* clearly translate the very low genetic diversity observed in each population of the YFP and clearly imply very low numbers of breeding individuals in each populations of the Yangtze River and the adjacent lakes [57]. However, drawing a more direct link between *Ne* and *N* is actually difficult considering our study design. Previous studies have shown that no direct relationship can be expected between *Ne* and *N* when generations are overlapping, sampling spans several years, includes multiple cohorts and age classes, and when immigration may occur [58-60].

The three YFP populations – XCSS, TL and PY– have not exchanged migrants over the last few generations according to our genetic estimates (figure 3 and ESM table S4). However, all three are connected to the admixed group (or have been recently). We observed genetic evidence of unidirectional gene flow from the admixed group to XCSS and TL. This recent migration in the middle section of the Yangtze and XCSS in the upper section may be rare and/or no longer occurring based on the observed fixation of the mitochondrial haplotype in the XCSS population (table 1 and figure 2). This is further supported by the YFDE2006 and YFDE2012 surveys that reported increasing gaps in the distribution of the species in the upper section of the Yangtze River (see ESM figure S1 and [14]). In contrast, gene flow was in the opposite direction from PY to the Admixed group (figure 3 and ESM table S4). This is consistent with field observations reporting groups of porpoises from PY moving to the main river stem in the morning and back to the lake in the afternoon [51]. This result also supports previous assertion [14] that immigration of porpoises from PY to the river may dampen population decline in the Yangtze River. Unfortunately, such migration is likely inadequate given the observed ongoing decline [14,50]. In principle, the admixed group in the middle section of the Yangtze River could serve as a bridge connecting the three differentiated populations. However, our estimates do not support this (figure 3) and are consistent with the increasing observation of gaps in the distribution of the YFP and the loss of connectivity between populations.

Additional populations of YFP may exist in the Yangtze mainstream as YFP is known to occur upstream and downstream from our study area (figure 1 and ESM figure S1). For

example, the three porpoises sampled around Shanghai city (SH) seem to belong to another differentiated unit, but no definitive conclusions can be drawn at this time due to low sample size. Nevertheless, the present study provides a representative view of the population genetic structure, connectivity and demographic trends that can help to define priority areas where conservation measures need to be taken.

Drastic reduction in population abundance has left a clearly detectable genetic footprint on the genetic diversity of the YFP and coincide with the loss of connectivity between population as well as the intensification of human activities along the Yangtze River over the last 50 years. To ensure that connectivity between populations is maintained, mitigation of human impacts need to include the entire river catchment [52]. For example, we suggest restricting fishing and sand-mining activities in the mouth area of Poyang Lake (Hukou, figure 1 and ESM figure S1) in order to restore the lake-river migration of the YFP. We also recommend modification of current *in situ* reserves that will improve connectivity between Ezhou and Zhenjiang. Likewise, more active measures could including a whole year fishing ban in the reserve would certainly help [14] and possible translocation of isolated individuals in the hope of increasing breed opportunities to increase genetic diversity.

Data accessibility. Data supporting this study are deposited at DRYAD (doi: TBA).

Authors' contributions. MCF, JZ, DW designed the study; MC, JZ, ZM, YH, KW, MW, QZ, DW conducted the field expeditions and collected the samples; MC conducted the laboratory experiments and collected the data; MCF and YBC analysed the data; MCF interpreted the results and wrote the manuscript with help from YBC, MC, JZ and DW, and final approval by all co-authors.

Competing interests. There are no competing interests to be declared.

Funding. This work was supported by the National Key Programme of Research and Development (2016YFC0503200) from Ministry of Science and Technology of China, the National Natural Science Foundation of China (Nos. 30730018, 31430080), the Special Fund for Agro-scientific Research in the Public Interest (No. 201203086), and the Special Conservation Fund for the Yangtze finless porpoise from the Ministry of Agriculture of China.

Acknowledgments. We thank Pr. JL Olsen for constructive comments.

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

References

1. Marko, P. B. & Hart, M. W. 2011 The complex analytical landscape of gene flow inference. *Trends Ecol. Evol.* **26**, 448–456. (doi:10.1016/j.tree.2011.05.007)

2. Epps, C. W. & Keyghobadi, N. 2015 Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Mol. Ecol.* **24**, 6021–6040. (doi:10.1111/mec.13454)

3. Fahrig, L. 2003 Effects of Habitat Fragmentation on Biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **34**, 487–515. (doi:10.1146/annurev.ecolsys.34.011802.132419)

4. Hale, M. L., Lurz, P. W., Shirley, M. D., Rushton, S., Fuller, R. M. & Wolff, K. 2001 Impact of landscape management on the genetic structure of red squirrel populations. *Science* **293**, 2246–2248. (doi:10.1126/science.1062574)

5. Mech, S. G. & Hallett, J. G. 2001 Evaluating the effectiveness of corridors: a genetic approach. *Conserv. Biol.* **5**, 467–474.

6. Carroll, C., McRae, B. H. & Brookes, A. 2012 Use of linkage mapping and centrality analysis across habitat gradients to conserve connectivity of gray wolf populations in western North America. *Conserv. Biol.* **26**, 78–87. (doi:10.1111/j.1523-1739.2011.01753.x)

7. Turvey, S. T. et al. 2007 First human-caused extinction of a cetacean species? *Biol. Lett.* **3**, 537–540. (doi:10.1098/rsbl.2007.0292)

8. Gao, A. & Zhou, K. 1995 Geographical variation of external measurements and three subspecies of Neophocaena phocaenoides in Chinese waters. *Acta Theriol. Sin.* **15**, 81–92.

9. Wang, D., Turvey, S. T., Zhao, X. & Mei, Z. 2013 *Neophocaena asiaeorientalis ssp. The IUCN Red List of Threatened Species*. e.T43205774A45893487 (accessed on 14 July 2016.). (doi:10.2305/IUCN.UK.2013-1.RLTS.T43205774A45893487.en)

10. Zhang, X., Renjun, L., Qingzhong, Z., Guocheng, Z., Zhuo, W., Xiaoqiang, W. & Jian, Y. 1993 The population of finless porpoise in the middle and lower reaches of yangtze river. *Acta Theriol. Sin.* **13**, 260–270.

11. Yang, J., Xiao, W., Kuang, X., Wei, Z. & Liu, R. 2000 Studies on the distribution, population size and the active regularity of L*ipotes vexillifer* and *Neophocaena phocaenoides* in Dongting Lake and Boyang Lake. *Resour. Env. Yangtze Valley* **9**, 444–450.

12. Xiao, W. & Zhang, X. 2002 Distribution and population size of Yangtze finless porpoise in Poyang Lake and its branches. *Acta Theriol. Sin.* **22**, 7–14.

13. Zhao, X. et al. 2008 Abundance and conservation status of the Yangtze finless porpoise in the Yangtze River, China. *Biol. Conserv.* **141**, 3006–3018. (doi:10.1016/j.biocon.2008.09.005)

14. Mei, Z. et al. 2014 The Yangtze finless porpoise: On an accelerating path to extinction? *Biol. Conserv.* **172**, 117–123. (doi:10.1016/j.biocon.2014.02.033)

15. Mei, Z., Huang, S.-L., Hao, Y., Turvey, S. T., Gong, W. & Wang, D. 2012 Accelerating population decline of Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*). *Biol. Conserv.* **153**, 192–200. (doi:10.1016/j.biocon.2012.04.029)

16. Fontaine, M. C., Snirc, A., Frantzis, A., Koutrakis, E., Öztürk, B., Oztürk, A. A. & Austerlitz, F. 2012 History of expansion and anthropogenic collapse in a top marine predator of the Black Sea estimated from genetic data. *P. Natl. Acad. Sci. USA* **109**, E2569–76. (doi:10.1073/pnas.1201258109)

17. Fontaine, M. C. et al. 2014 Postglacial climate changes and rise of three ecotypes of harbour porpoises, *Phocoena phocoena*, in western Palearctic waters. *Mol. Ecol.* **23**, 3306–3321. (doi:10.1111/mec.12817)

18. Yang, G., Guo, L., Bruford, M. W., Wei, F. & Zhou, K. 2008 Mitochondrial phylogeography and population history of finless porpoises in Sino-Japanese waters. *Biol. J. Linn. Soc.* **95**, 193–204. (doi:10.1111/j.1095-8312.2008.0932.x)

19. Lin, W., Frère, C. H., Karczmarski, L., Xia, J., Gui, D. & Wu, Y. 2014 Phylogeography of the finless porpoise (genus Neophocaena): testing the stepwise divergence hypothesis in the northwestern Pacific. *Sci. Rep.* **4**, 6572. (doi:10.1038/srep06572)

20. Chen, M., Zheng, J., Wu, M., Ruan, R., Zhao, Q. & Wang, D. 2014 Genetic Diversity and Population Structure of the Critically Endangered Yangtze Finless Porpoise (*Neophocaena asiaeorientalis asiaeorientalis*) as Revealed by Mitochondrial and Microsatellite DNA. *Int. J. Mol. Sci.* **15**, 11307–11323. (doi:10.3390/ijms150711307)

21. Cornuet, J.-M., Ravignie, V. & Estoup, A. 2010 Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinfor.* **11**, 401. (doi:10.1186/1471-2105-11-401)

22. Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. 2009 Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**, 1322–1332. (doi:10.1111/j.1755-0998.2009.02591.x)

23. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. 2015 Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191. (doi:10.1111/1755-0998.12387)

24. 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. (doi:10.1111/j.1365-294X.2005.02553.x)

25. Pante, E. & Simon-Bouhet, B. 2013 marmap: A package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS ONE* **8**, e73051. (doi:10.1371/journal.pone.0073051)

26. Amante, C. & Eatkins, B. W. 2009 ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. (doi:10.7289/V5C8276M)

27. Francois, O. & Durand, E. 2010 Spatially explicit Bayesian clustering models in population genetics. *Mol. Ecol. Resour.* **10**, 773–784. (doi:10.1111/j.1755-0998.2010.02868.x)

28. Jombart, T. 2008 adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. (doi:10.1093/bioinformatics/btn129)

29. R Core Team 2016 R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*

30. Szpiech, Z. A., Jakobsson, M. & Rosenberg, N. A. 2008 ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498–2504. (doi:10.1093/bioinformatics/btn478)

31. Peakall, R. & Smouse, P. E. 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* **28**, 2537–2539. (doi:10.1093/bioinformatics/bts460)

32. Weir, B. S. & Cockerham, C. C. 1984 Estimating F-statistics for the analysis of population structure. *Evolution*, 1358–1370.

33. Goudet, J. 2001 FSTAT, A program to estimate and test gene diversities and fixation indices (version 2.9. 3). Available at http://www2.unil.ch/popgen/softwares/fstat.htm [accessed 2016-09-08].

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

35. Watterson, G. A. 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276.

36. Librado, P. & Rozas, J. 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. (doi:10.1093/bioinformatics/btp187)

37. Hudson, R. R., Boos, D. D. & Kaplan, N. L. 1992 A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**, 138–151.

38. Hudson, R. R. 2000 A new statistic for detecting genetic differentiation. *Genetics* **155**, 2011–2014.

39. Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J. & Ovenden, J. R. 2014 NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (*Ne*) from genetic data. *Mol. Ecol. Resour.* **14**, 209–214. (doi:10.1111/1755-0998.12157)

40. Waples, R. S. & Do, C. 2010 Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol. Appl.* **3**, 244–262. (doi:10.1111/j.1752-4571.2009.00104.x)

41. Tallmon, D. A., Koyuk, A., Luikart, G. & Beaumont, M. A. 2008 ONESAMP: a program to estimate effective population size using approximate Bayesian computation. *Mol. Ecol. Resour.* **8**, 299–301. (doi:10.1111/j.1471-8286.2007.01997.x)

42. Wilson, G. A. & Rannala, B. 2003 Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177–1191.

43. Rambaut, A. 2007 Tracers v1.5 Available from http://tree.bio.ed.ac.uk/software/tracer/.

44. Meirmans, P. G. 2014 Nonconvergence in Bayesian estimation of migration rates. *Mol. Ecol. Resour.* **14**, 726–733. (doi:10.1111/1755-0998.12216)

45. Garza, J. C. & Williamson, E. G. 2001 Detection of reduction in population size using data from microsatellite loci. *Mol. Ecol.* **10**, 305–318. (doi:10.1046/j.1365-294x.2001.01190.x)

46. Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M. & Estoup, A. 2014 DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30**, 1187–1189. (doi:10.1093/bioinformatics/btt763)

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

48. Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & Francois, O. 2010 Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418. (doi:10.1016/j.tree.2010.04.001)

49. Mcvean, G. 2009 A genealogical interpretation of principal components analysis. *PLoS Genet.* **5**, e1000686. (doi:10.1371/journal.pgen.1000686.g006)

50. Dong, L., Wang, D., Wang, K., li, S., Mei, Z., Wang, S., Akamatsu, T. & Kimura, S. 2015 Yangtze finless porpoises along the main channel of Poyang Lake, China: Implications for conservation. *Mar. Mamm. Sci.* **31**, 612–628. (doi:10.1111/mms.12181)

51. Wei, Z., Wang, D., Zhang, Q., Wang, K. & Kuang, X. 2002 Population size, behavior, movement pattern and protection of yangtze finless porpoise at balijiang section of the yangtze river. *Resour. Env. Yangtze Valley* **11**, 427–432.

539 52. Huang, S. L., Mei, Z., Hao, Y., Zheng, J., Wang, K. & Wang, D. 2016 Saving the Yangtze finless
540 porpoise: Time is rapidly running out. *Biol. Conserv.*, in press. (doi:10.1016/j.biocon.2016.05.021)

541 53. Fontaine, M. C. et al. 2010 Genetic and historic evidence for climate-driven population fragmentation in
542 a top cetacean predator: the harbour porpoises in European water. *Proc. R. Soc. B* **277**, 2829–2837.
543 (doi:10.1098/rspb.2010.0412)

544 54. Louis, M. et al. 2014 Habitat-driven population structure of bottlenose dolphins, *Tursiops truncatus*, in
545 the North-East Atlantic. *Mol. Ecol.* **23**, 857–874. (doi:10.1111/mec.12653)

546 55. Goossens, B., Chikhi, L., Ancrenaz, M., Lackman-Ancrenaz, I., Andau, P. & Bruford, M. W. 2006
547 Genetic Signature of Anthropogenic Population Collapse in Orang-utans. *PloS Biol.* **4**, e25.
548 (doi:10.1371/journal.pbio.0040025.st007)

549 56. Weckworth, B. V., Musiani, M., DeCesare, N. J., McDevitt, A. D., Hebblewhite, M. & Mariani, S. 2013
550 Preferred habitat and effective population size drive landscape genetic patterns in an endangered
551 species. *Proc. R. Soc. B* **280**, 20131756. (doi:10.1098/rspb.2013.1756)

552 57. Waples, R. S. 2005 Genetic estimates of contemporary effective population size: to what time periods
553 do the estimates apply? *Mol. Ecol.* **14**, 3335–3352. (doi:10.1111/j.1365-294X.2005.02673.x)

554 58. Palstra, F. P. & Fraser, D. J. 2012 Effective/census population size ratio estimation: a compendium and
555 appraisal. *Ecol. Evol.* **2**, 2357–2365. (doi:10.1002/ece3.329)

556 59. Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., Ruegg, K. &
557 Palstra, F. P. 2011 Understanding and Estimating Effective Population Size for Practical Application in
558 Marine Species Management. *Conserv. Biol.* **25**, 438–449. (doi:10.1111/j.1523-1739.2010.01637.x)

559 60. Waples, R. S. & England, P. R. 2011 Estimating Contemporary Effective Population Size on the Basis
560 of Linkage Disequilibrium in the Face of Migration. *Genetics* **189**, 633–644.
561 (doi:10.1534/genetics.111.132233)

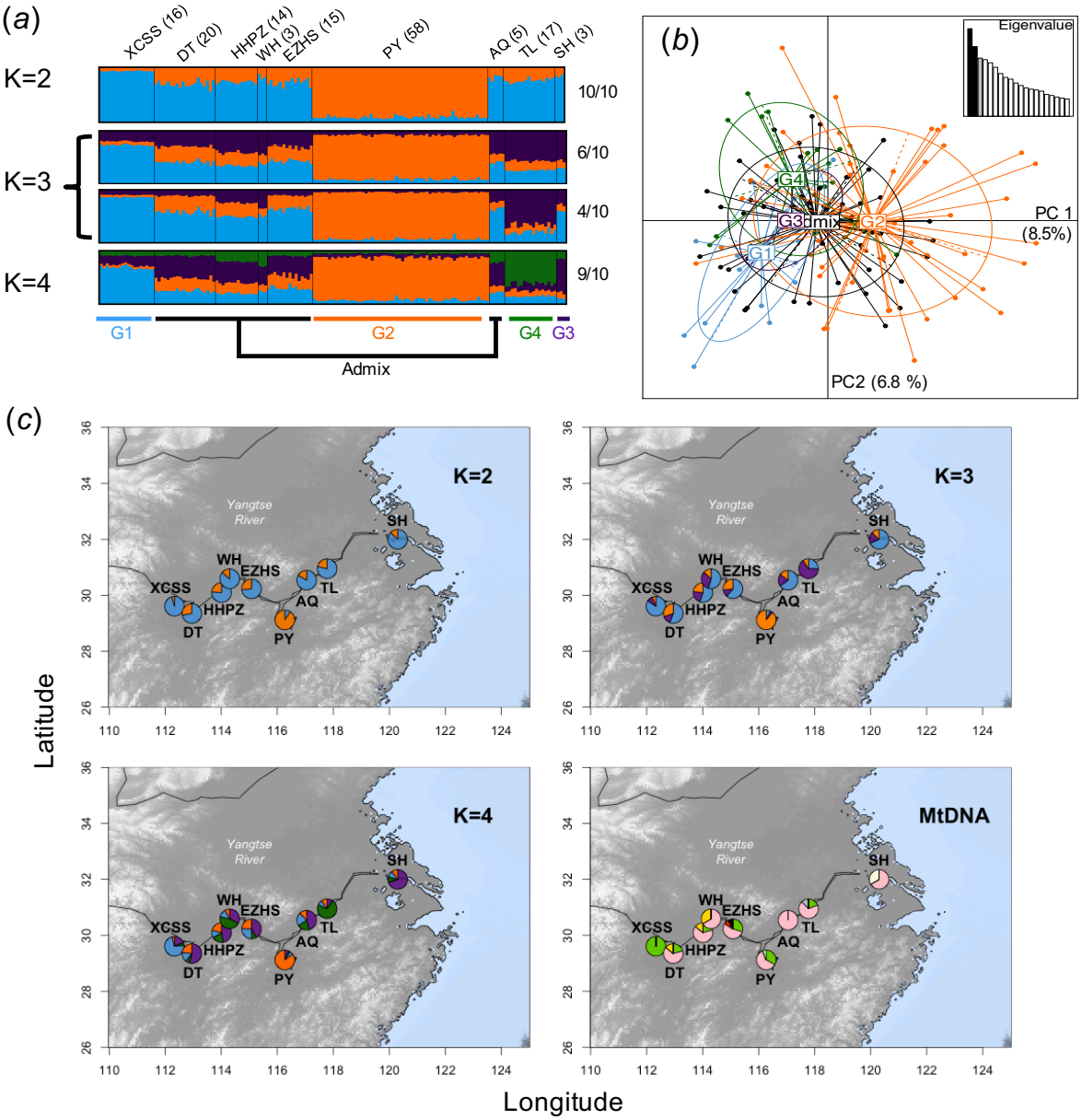
562

Figure legends

Figure 1. (*a*) Population structure estimated using *STRUCTURE*. Each individual is represented by a vertical line divided into *K* segments showing the admixture proportions from each cluster. Sample size in each locality is shown between brackets. Numbers on the right side of the barplot show the number of time this results was found. (*b*) PCA displaying individual scores along the first two components. The proportion of variance explained by each axis and the firsts eigenvalues (top-right inset) are provided. (*c*) Geographical distribution of the *STRUCTURE* admixture proportions and mitochondrial haplotype frequency per localities.

Figure 2. Schematic diagram of ABC analysis to compare evolutionary histories and divergence scenarios generated and tested using the program DIYABC. Each coloured segment depicts a distinct effective population size. (See the text, ESM appendix S1, table S1-S2, S5-S12).

Figure 3: Recent gene flow (*Ne.m*) between populations estimated from *ONESAMP* and *BayeAss*. Confidence intervals are shown between squared brackets. Arrows show the effective migration rate significantly (plain) and not significantly different (dashed) from 0. *Ne* estimates of *ONESAMP* (mean [95%CI]) in each population is provided in the circles.



584

585 **Figure 1**

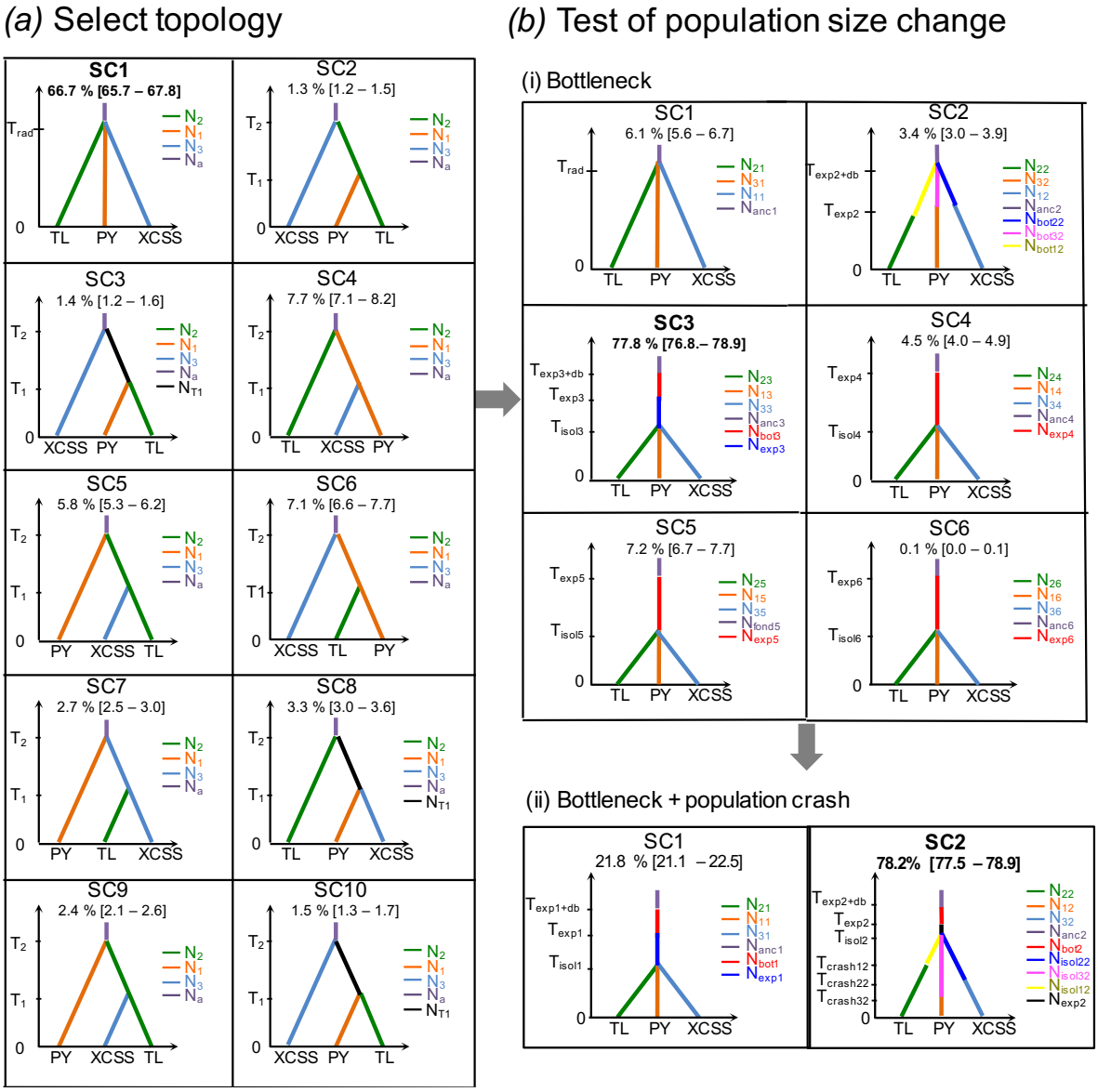
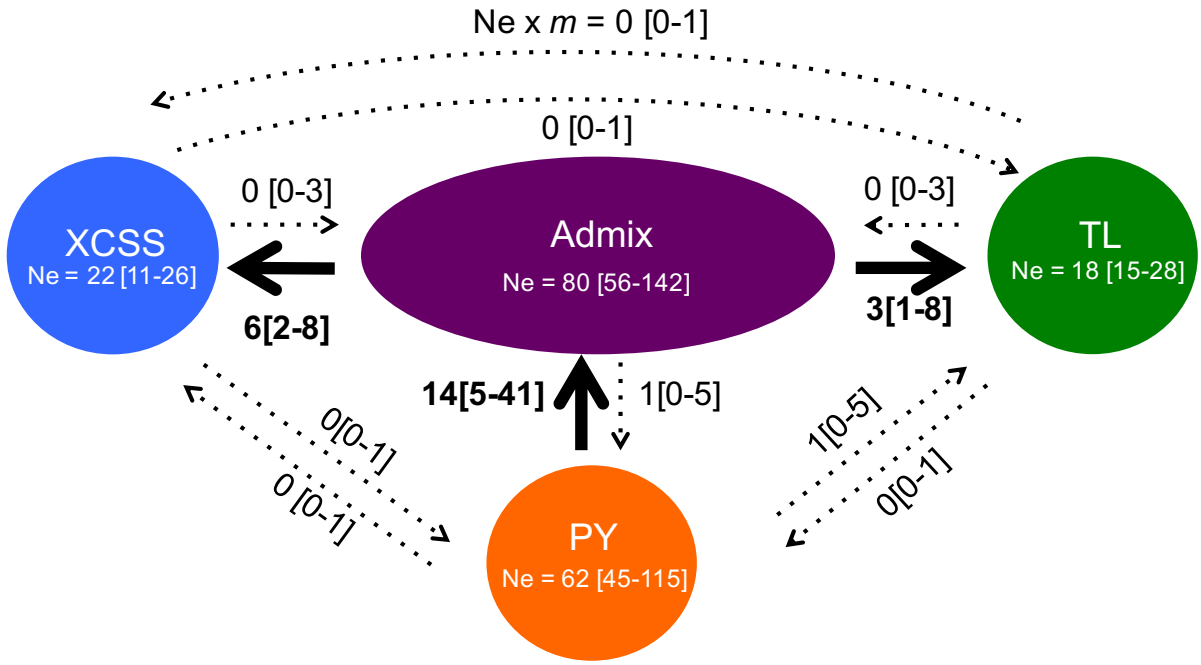


Figure 2



589

590

591 **Figure 3**

592

Table. 1: Genetic variation at the 11 microsatellites and mtDNA control region loci for each distinct population inferred from the *STRUCTURE* analysis.

	Total	PY	TL	XCSS	Admix
Microsatellite loci					
$N_{Mic.}$ ± SD (max)	-	56.3 ± 0.31 (58)	16.1 ± 0.3 (17)	16.4 ± 0.39 (18)	49.9 ± 1.4 (57)
%NA	-	0.01	0.06	0.09	0.14
A_r ± SE	-	4.62 ± 0.39	5.18 ± 0.34	4.88 ± 0.47	5.64 ± 0.43
pA ± SE	-	0.42 ± 0.13	0.52 ± 0.12	0.39 ± 0.13	0.62 ± 0.15
H_o ± SD	-	0.60 ± 0.06	0.71 ± 0.052	0.69 ± 0.04	0.72 ± 0.03
H_e ± SD	-	0.62 ± 0.05	0.68 ± 0.04	0.63 ± 0.04	0.69 ± 0.03
F_{IS} ± SD	-	0.042 ± 0.11^{ns}	-0.048 ± 0.08^{ns}	-0.090 ± 0.18^{ns}	-0.048 ± 0.09^{ns}
M_{GW}†	-	0.51***	0.46**	0.44**	0.53**
MtDNA control region					
N_{mtDNA}	129	56	17	16	37
S	7	2	1	0	5
Singleton	4	0	0	0	3
Shared P.	3	2	1	0	2
#hap.	7	3	2	1	5
Hd	0.57	0.53	0.22	0	0.62
π (per site, %)	0.112	0.096	0.037	0	0.150
Theta-W (per site, %)	0.216	0.073	0.050	0	0.201
D‡	-1.08^{ns}	0.54^{ns}	-0.49^{ns}	0	-0.64^{ns}

$N_{Mic.}$, microsatellite average sample size (max); *%NA*: average proportion of missing data per locus, A_r: allelic richness (estimated for a sample size of 22 individuals); *pA*: Private allelic richness (estimated for a sample size of 3 individuals); *Ho* and *He*: observed and expected heterozygosity; F_{IS}: Inbreeding coefficient; M_{GW}, *M* of Garza and Williamson [62]; N_{mtDNA}, MtDNA sample size; *S*, number of segregating sites; *Share P.*, shared polymorphism; #hap, number of haplotype; *Hd*, haplotype diversity; π, nucleotide diversity; Theta-W, theta from S or Theta-Watterson; *D*, Tajima's *D*. † The significance level of the M_{GW} statistic was evaluated in DIYABC [63] using $1x10^6$ coalescent simulations under a scenario of constant effective population size. The *P*-value indicate the proportion of simulations which provide a value below the observed one. ‡ The significance of *D* values was estimated using 10 000 coalescent simulations in DNAsp [51]. ns: not significant (*p*-value > 0.05); * *p*-value ≤ 0.05; ** *p*-value ≤ 0.01; *** *p*-value ≤ 0.001

605 **Table. 2:** Genetic differentiation between populations identified by *Structure.* Below the diagonal, pairwise
606 F_{ST} values and their 95% CI for microsatellite loci are provided as well as their associated *P*-value. Above
607 the diagonal, F_{ST} values are provided for the mtDNA locus with its corresponding *P*-value.
608

$F_{ST\text{-}mic}$ \ $F_{ST\text{-}mtDNA}$	**XCSS**	**PY**	**TL**	**Admix**
XCSS	–	0.590***	0.875***	0.446***
PY	0.069*** [0.033-0.107]	–	0.109*	0.011**
TL	0.050*** [0.032-0.071]	0.051*** [0.024-0.076]	–	0.131^{ns}
Admix	0.030** [0.009-0.051]	0.021*** [0.013-0.032]	0.023*** [0.010-0.037]	–

609
610 ns: not significant (*p*-value > 0.05); * *p*-value ≤ 0.05; ** *p*-value ≤ 0.01; *** *p*-value ≤ 0.001
611