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]*, Frédéric Labbé[3,4,5], Zhigang Mei[1], Yujiang Hao[1], Kexiong Wang[1], Min Wu[1], 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] Research Center 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;

[4] Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN 46556, USA

[5] Department of Biological Sciences, University of Notre Dame, Galvin Life Sciences Center, Notre Dame, IN 46556, USA

[†] 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; Yangtze Finless Porpoise; *Neophocaena asiaeorientalis*; baji; *Lipotes vexillifer*

Abstract

Understanding demographic trends and patterns of gene flow in an endangered species is crucial for devising conservation strategies. Here, we examined the extent of population structure and recent evolution of the critically endangered Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*). By analysing genetic variation at the mitochondrial and nuclear microsatellite loci for 148 individuals, we identified three 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 (≤92 individuals). 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 Modern Economy Development. However, genetic diversity showed a clear footprint of population contraction over the last 50 years leaving only ~2% of the pre-collapsed size, consistent with the population collapses reported from field studies. This genetic perspective provides background information for devising mitigation strategies to prevent this species from extinction.

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 facilitating this movement across such landscape[4-6]. Along the Yangtze River (China, Fig. 1), anthropogenic 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 asiaeorientalis* or 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 (Fig. 1), 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 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, as estimated by the Yangtze Freshwater Dolphin Expedition in 2006 (YFDE2006)[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 sub-species on the IUCN Red List[9]. As a top predator, the survival of the finless porpoise depends heavily on habitat suitability, food availability and maintenance of corridors allowing dispersal between populations. However, with increasing underwater noise from boat traffic and incidental catches in fisheries, 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 do not 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 populations and the population structure itself. This information is extremely difficult to quantify using direct observations. On the other hand, population genetic approaches can provide key insights about the current population structure and connectivity and historical population demography by leaving detectable footprints on the genetic diversity and its geographic structure[16,17].

Previous phylogeographic analyses based on the non-coding mitochondrial Control Region (mtDNA-CR) of the finless porpoises from Chinese and Japanese waters documented evidence of a demographic expansion following the Last Glacial Maximum (LGM, ~24,000 yrs BP) and the colonization of Yangtze River from the Yellow Sea ~22,000 yrs BP ago[18,19]. A subsequent study within the Yangtze River by Chen et al.[20] used mtDNA-CR and nuclear microsatellite loci and revealed evidence of genetic subdivisions within the YFP populations suggestive of population fragmentation. However, the fine scale population structure remained unclear and the processes shaping the genetic variation in the YFP unknown. Changes in connectivity between populations, dynamics of population expansion-contraction and demographic history are potent factors shaping the genetic variation in the YFP, but these were not investigated in further details so far.

In this study, we address these above questions by re-analysing the previously published data set of Chen et al.[20] that combined fast evolving microsatellite loci together with more slowly evolving sequences from the mtDNA-CR. Specifically, in contrast with previous studies, we used a statistical population genetic framework in order to (i) resolve the fine-scale population structure using a combination of individual-based multivariate and Bayesian clustering approaches designed for weak genetic structure when it exists; (ii) estimate past and contemporaneous effective population sizes and connectivity among populations; and (iii) reconstruct the demographic history best fitting with the genetic diversity observed in the populations of the YFP using a quantitative model-based inferential population genetic framework relying on Approximate Bayesian Computation (ABC) approaches[16,17,21-23]. Determining and quantifying to which extent populations in the YFP are fragmenting, loosing

connectivity, and the magnitudes of the demographic trends are critical knowledge for designing conservation and management plans. For example, we still do not understand whether population fragmentation and decline have been triggered only recently by human activities during the past 50 years or if these patterns were initiated earlier by more long-term ecological and evolutionary processes and exacerbated lately during the Anthropocene.

Results

Genetic structure and diversity of the Yangtze finless porpoises. The final data set consisted of 148 individuals sampled along the Yangtze River and the two main lakes (Dongting Lake (DT) and Poyang Lake (PY), Fig. 1) genotyped for 11 microsatellite loci and sequenced for a 597 base-pairs fragment of the hyper-variable region 1 (see Table 1 and materials and methods). The clustering of the microsatellite data using the Bayesian clustering algorithm of *STRUCTURE*[24-26] provided consistent results over 10 replicated runs performed for each number *K* of cluster tested (Fig. 2a). The probability of the data greatly increased when two genetic clusters were modelled instead of one and showed the highest values on average over 10 replicates (Fig. S1). At K>2, the increase in probability decreased in average; however some runs displayed the highest probability of all runs at K=3 and 4 (Fig. S1). A visual inspection of the individual clustering for each K value (Fig. 2a) 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 values were tested (K=3 and 4) porpoises from PY, XCSS and TL localities were all identified as differentiated genetic units, while the porpoises from the in-between regions consisted of an admixed group sharing genetic ancestry with the three other populations (Fig. 2a and 2d). The three individuals at the mouth of the Yangtze River close to Shanghai city (SH) also seemed to depart from the other groups at K=4, but the low sample size (n=3) preclude any definitive conclusions. No further subdivision was observed beyond K=4 (result not shown).

145 The results from *STRUCTURE* were further validated using two multivariate approaches that
146 do not rely on model assumptions[27]: a Discriminant Analysis of Principal Components
147 (*DAPC*)[28] and a Principal Component Analysis (PCA)[27,29]. Both methods showed a genetic
148 structure consistent with the results of *STRUCTURE* (Fig. 2b, 2c and Fig. S2). The *DAPC*
149 provided a clear-cut discrimination of the four groups identify by *STRUCTURE* (Fig. 2c). The
150 first discriminant function (DF) discriminated XCSS and PY and the second DF TL and SH. The
151 other porpoises recognized as admixed in *STRUCTURE* were located at the intersection of the
152 four other groups. The proportions of successful reassignment (based on the discriminant
153 functions) of individuals to their original clusters was high (Fig. 2b): 100% for XCSS and SH,
154 94.8% for PY, 94.1% for TL, and 94.3% for the admixed porpoises. These large values indicate
155 clear-cut clusters. Finally, the PCA provided a similar picture as the *DAPC* and *STRUCTURE*, but
156 with more overlap among the groups (Fig. S2).
157
158 None of the identified populations displayed significant departures from Hardy–Weinberg
159 expectations as shown by the F_{IS} values (Table 1). Porpoises from the admixed and TL groups
160 showed the highest level of microsatellite genetic diversities and XCSS and PY the lowest, as
161 estimated with the values of allelic richness (*Ar*), private allelic richness (*pA*) and expected
162 heterozygosity (H_e) (Table 1 and Fig. S3). Only the two extremes groups, PY and Admix,
163 showed a significant difference in A_r and H_e (Wilcoxon signed-ranked test p-value < 0.05). The
164 mitochondrial genetic diversity followed a similar trend with the admixed group showing the
165 highest haplotype and nucleotide diversity followed by TL and PY (Table 1 and Fig. 2d). We
166 found only one haplotype fixed in the XCSS group.
167
168 All populations showed significant differences in allelic frequencies for the microsatellite loci
169 with F_{ST} values ranging from 0.023 to 0.070 (Table 2). Notably F_{ST} values between XCSS, PY
170 and TL were relatively high (>0.05), while F_{ST} values were intermediate between the Admix
171 group and each distinct population (between 0.02 and 0.03, Table 2). For the mitochondrial
172 locus (Table 2), F_{ST} values were all significant except one (Admix vs. TL), and were especially
173 strong between XCSS and all other groups, due to the fact that one haplotype is fixed in this
174 population (Fig. 2d).
175
176

177 **Contemporary effective population sizes and migration rates.** We used two methods for
178 estimating contemporary effective population size (*Ne*) in each population from the
179 microsatellite data: *NeEstimator*[30] based on linkage disequilibrium among loci within
180 population and *ONESAMP*[31] relying on an Approximate Bayesian Computation (ABC) approach.
181 Both approaches provided comparable estimates of *Ne* for each population (Fig. 3 and Table
182 3). All estimated values were very low (<92 individuals), with the Admix and PY groups
183 displaying the highest estimates, followed by TL and XCSS.
184
185 We estimated migration rates (*m*) between pairs of groups over the last few generations using
186 *BayesAss*[32]. Out of the 12 repeat runs, 10 showed good mixing properties with Bayesian
187 Deviance values[33] close to each other (mean ± SD: 8,915.28 ± 0.33) and convergent estimates
188 for each parameter (see the material and methods). We thus combined them all to estimate
189 the parameter values (Table S2). The number of effective migrants (*Ne* x *m*) per generation
190 over the last generations were obtained by combining *Ne* estimates from *ONESAMP*[31] with *m*
191 estimates of *BayesAss*[32] (Fig. 3). The three populations – TL, XCSS and PY – did not show any
192 evidence of recent migration among each other, as indicated by the lower bound of the 95%
193 highest probability density intervals (HPD) interval equal to 0 (Fig. 3 and Table S1). However,
194 each one is connected to the Admix group with highly asymmetric gene flow. We detected
195 significant unidirectional gene flow from PY to Admix and from Admix to TL and XCSS.
196 Estimated *Ne x m* values from the Admix group to TL or XCSS are about the half of those from
197 PY to the Admix group.
198
199 **Population demographic history.** Departure of the allelic or haplotypic frequency spectrum of
200 microsatellite and mtDNA loci, respectively, from those expected under a scenario of constant
201 population size can provide evidence of population size change. For microsatellite loci, we
202 used the Garza and Williamson M_{GW} ratio of the number of alleles to the range in allele size to
203 detect evidence of population size contraction[34]. Genetic diversity at the microsatellite
204 markers showed significant evidence of *Ne* contraction in each population, as suggested by
205 the very small ratio values of the M_{GW} statistic (Table 1). The M_{GW} value estimated in each
206 population was significantly smaller than expected under the assumption of constant
207 population size (Table 1). In contrast, the Tajima's *D*[35] values (Table 1) estimated from the
208 mtDNA-CR sequences in each population did not show any significant departure from the

209 constant population size hypothesis.

210

211 We investigated further the demographic history best fitting with the genetic diversity of the
212 combined microsatellite and mtDNA markers observed in the three distinct populations of
213 the YFP using a coalescent-based ABC approach[22]. See the material and methods, appendix
214 S1, Fig. S4, Table S2 and S3 for further details on the methodology. The first step in the ABC
215 analysis was to identify the population branching order that best fit with the data. Out of the
216 10 scenarios tested (Fig. 4a), the ABC analysis showed that the trichotomy (SC1), which
217 assumes that XCSS, PY and TL populations diverged at the same time, received the highest
218 support. Indeed, the two distinct model choice approaches – a "standard" model choice
219 procedure relying a logistic regression of Linear Discriminant Analysis (ABC-LDA) on the
220 summary statistics[36] and the recently developed Random Forest machine learning
221 classification approach[37,38] (ABC-RF) – identify this scenario SC1 as the best one with a
222 posterior probability respectively of 66.7% with a 95%CI not overlapping with any other
223 scenarios and 55.8 ± 2.5% (Fig. S5a and Table S4). This contrasted with the nine other models
224 where the posterior probabilities estimated using the ABC-LDA were each lower than 8%. The
225 simulation-based performance analysis of this ABC step (Table S4, S5 and Fig. S5) shows that
226 61.4% of the simulated dataset under this SC1 were correctly identified using the ABC-LDA
227 procedure, leading to an average Type-I error rate (false negative) of 4.3% ranging from 1.7%
228 to 9.7% and a total prior error rate (*i.e.* the average misclassification error)[37,38] of 38.6% (62.9%
229 using the ABC-RF). Simulations under the nine other competing scenarios led to a Type-II
230 error rate of <12.9% incorrect assignment to SC1 (false positives) and a power (87.1%) to
231 discriminate the best scenario from the others (Table S4).

232

233 The second ABC step tested for occurrence of simple changes in effective population size (*Ne*)
234 during the divergence of the three populations under six nested competing scenarios (Fig. 4bi,
235 Table S2 and S3). The scenario assuming a bottleneck in the ancestral population prior to
236 population split (SC3) outcompeted the other five scenarios (*i.e.*, no change (SC1) or simple
237 decline or expansion in the ancestral or daughter populations, SC2, SC4-6), with a probability
238 of 78% and no overlap in 95%CI for the ABC-LDA, and 69.8 ± 1.9% under the ABC-RF (Fig. 4bi
239 and Table S6). The sensitivity analysis showed that simulations generated with SC3 were more
240 difficult to identify, with an average Type-I error rate of 12.4% (ranging between 1 and 21%

241 error, depending on the scenario), and a total prior error rate of 62.7% under the ABC-LDA
242 and 41.4% under the ABC-RF (Fig. S6 and Table S6). Nevertheless, the Type-II error rate (false
243 positive) was only 7.2% and the power to discriminate this scenario from the others was
244 92.8%. Furthermore, the assessment of goodness-of-fit of this bottleneck scenario (SC3) to
245 the data showed very good performance, as the simulations using this scenario and posterior
246 distributions for each parameter, were able to reproduce all but one observed summary
247 statistic, in contrast to all other scenarios (Table S6 and S7).
248
249 The final step (Fig. 4bii, Table S2 and S3) tested whether a recent demographic collapse in
250 each population within the last 5 generations (or 50 years as reported in the literature[15])
251 could have produced a detectable genetic footprint. The scenario involving a recent collapse
252 (SC2) in each population received a significantly higher probability (ABC-LDA 78.2%; ABC-RF:
253 69.2 ± 2.0%) compared with the alternative scenario of constant *Ne* since the population split
254 (Fig. 4bii, Fig. S6, Table S8, and S9). Both Type-I and Type-II error rates were <16% indicating
255 adequate power and sensitivity of our ABC analysis.
256
257 The model parameters estimated under the final best demographic scenario (SC2 in Fig. 4bii,
258 Table S10 and S11) suggest that the ancestral population would have been large (N_{anc2} =
259 18,700 individuals; 95%CI: [3,480 – 19,700]), and that a small fraction would have founded
260 the Yangtze River populations ~3,400 95%CI: [1300 – 41,000] years ago (T_{exp2}) and expanded
261 to reach an effective size of about 5,660 individuals (N_{exp2}; 95%CI: [2,900 – 9,840]). The three
262 daughter populations would have then split from each other ~1,030 years ago (T_{isol2}; 95%CI:
263 [214 – 4,380]), and reached an effective size of about 2,000 individuals after their split. Each
264 of these daughter populations would have gone through a significant decline leaving less than
265 2% of their pre-collapse size during the last 50 years (see Table 3, S10, and S11).
266
267

268 Discussion

269
270 Our study shows that the present-day genetic diversity of finless porpoise in the Yangtze
271 River (YFP) has been strongly influenced by an initial founder event that took place several

thousand years ago, followed by a relatively recent split into 3 populations (XCSS, TL, PY), and a recent demographic collapse within the last 50 years. Indeed, consistent with previous studies[18,19], the ABC genetic inferences showed that a few individuals coming from a large ancestral population, likely a marine population of *Neophocaena asiaeorientalis sunameri* in the Yellow Sea, colonized the Yangtze River within the last 41kyrs, and most likely during the Last Ice Age. The subsequent population split into three populations occurred between 200 and 5,000 years BP. This suggests that the population split was triggered before the intensification of human activities of the last 50 years in the Yangtze River. This event may thus have been related to the colonization process itself during the post-glacial period[19] and/or other environmental or human-related factors. Consistent with the hypothesis of population split driven by post-glacial changes, episodes of contraction and expansion of the Yangtze River mainstream and the adjacent lakes occurred during the Holocene period[39]. Interestingly, the Yangtze River mainstream and the appendage lakes, including the Poyang and Dongting lakes, retracted and shrank significantly during the late Holocene (from *ca.* 3000 years BP to now)[39-41]. These environmental changes might have promoted the split of the ancestral population after the colonization of the Yangtze River.

Each population harboured a genetic footprint of dramatic population reduction. This especially affected the genetic diversity at the microsatellite markers which displayed very small values of the M_{GW} statistic (Table 1) characteristic of significant recent decline[34]. This is also indicated by the ABC analysis supporting a scenario describing a drastic population reduction in each population within the last five generations (Fig. 4b_{ii}, SC2). According to this scenario, the reduction in effective size would have been massive since their current sizes would be less than ~2% of their pre-collapsed sizes (Fig. 4, Table 3, S10 and S11). These genetic inferences are in line with the field estimates[14,15,42], reporting a continuous decline of the YFP since the 1980s[10-15,43,44]. Half of the census populations in the main stem of the Yangtze River would have been lost within the past 15 years, with the abundance dropping from 2,500 porpoises in 1991[10] to 1,225 in 2006[13]. With the 400 porpoises in the Poyang Lake and the 100 to 150 porpoises of the Dongting Lake, the total census size in the Yangtze River and the two adjacent lakes observed in 2006 was only 1,800 individuals. The most recent estimates from the YFDE2012 showed that the porpoises in the main stem of the Yangtze would have been reduced again by half with 505 individuals reported the middle and lower

304 reaches of the Yangtze River and approximately 450 porpoises in PY, and 90 porpoises in
305 DT[14].
306 These very small census populations sizes (*N*) are consistent with the very small
307 contemporary effective population size (*Ne*) estimated from genetic data (≤92 individuals) in
308 each population, with the XCSS population being the smallest of all three differentiated
309 populations (between 14 and 22 individuals, Table 3). Comparably low estimates have been
310 reported in other cetacean species, such as the southern Iberian ecotype of harbour porpoise
311 (*Phocoena phocoena meridionalis*) in European waters (Ne ≤ 80[17,45]) or the coastal ecotype of
312 bottlenose dolphins (*Tursiops truncatus*) in European waters (Ne ≤ 77[46]). However only highly
313 endangered populations, such as the Borneo Orangutans (*Pongo pygmaeus*)[47] or the
314 Canadian woodland caribou (*Rangifer tarandus*)[48], have *Ne* values as low as those observed
315 in the XCSS population. Extremely low *Ne* clearly translate the very low genetic diversity
316 observed in each population of the YFP and imply very low numbers of breeding individuals in
317 each populations of the Yangtze River and the adjacent lakes[49]. However, drawing a more
318 direct link between *Ne* and *N* is actually difficult considering our study design. Previous
319 studies have shown that no direct relationship can be expected between *Ne* and *N* when
320 generations are overlapping, sampling spans several years, includes multiple cohorts and age
321 classes, and when immigration may occur[50-52].
322
323 The three YFP populations – XCSS, TL and PY– have not exchanged migrants over the last few
324 generations according to our genetic estimates (Fig. 3 and Table S1). However, all three are
325 or have been recently connected to the admixed group. We observed genetic evidence of
326 unidirectional gene flow from the admixed group to XCSS and TL. This recent migration in the
327 middle section of the Yangtze and XCSS in the upper section may be rare and/or no longer
328 occurring based on the observed fixation of the mitochondrial haplotype in the XCSS
329 population (Table 1 and Fig. 2d). This is further supported by the YFDE2006 and YFDE2012
330 surveys that reported increasing gaps in the distribution of the species in the upper section of
331 the Yangtze River[14]. In contrast to the two other populations in the main stream river, gene
332 flow was in the opposite direction in the Poyang Lake, from PY to the Admixed group (Fig. 3
333 and Table S1). This is consistent with field observations reporting groups of porpoises from PY
334 moving to the main river stem in the morning and back to the lake in the afternoon[43]. This
335 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 insufficient given the observed ongoing decline[14,42]. 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 gene flow estimates do not support this (Fig. 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 the YFP is known to occur upstream and downstream from our study area (Fig. 1). For example, the three porpoises sampled around Shanghai city (SH) seem to belong to another differentiated unit (Fig. 2), but no definitive conclusions can be drawn at this time due to low sample size (n=3). 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 coincides with the loss of connectivity between populations 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[44]. For example, restricting fishing and sand-mining activities in the mouth area of Poyang Lake (Hukou, Fig. 1) could restore the lake-river migration of the YFP. Modification of current *in situ* reserves could improve connectivity between Ezhou and Zhenjiang (Fig. 1). Likewise, more active measures including a whole year fishing ban in the *in situ* natural reserves could certainly help[14] and possible translocation of isolated individuals in the hope of increasing breeding opportunities could increase genetic diversity.

Material & Methods

Samples collection and DNA extraction. A total of 153 Yangtze finless porpoises were sampled

between 1998 and 2011 across the distribution range (Fig. 1), including 3 from Shanghai (SH), 17 from Tongling (TL), 5 from Anqing (AQ), 15 from Ezhou-Huangshi (EZHS), 3 from Wuhan (WH), 14 from Honghu-Paizhou (HHPZ), 2 from Jianli (JL), 16 from Xingchang-Shishou (XCSS), 20 from Dongting Lake (DT) and 58 from Poyang Lake (PY). A detailed description of the sampling procedure and genomic DNA extraction is provided in Chen *et al.*[20]. Briefly, blood samples (n=113) were drawn from the caudal vein of live porpoises, immediately preserved in an Acid-Citrate-Dextrose solution, and stored in liquid nitrogen. Tissue samples (n=40, mainly muscle) were obtained from dead porpoises and preserved in 80% alcohol. Samples from the Yangtze mainstream and Dongting Lake were collected from accidentally killed or stranded individuals over the past decade. Fifty-eight blood samples from PY were collected from live animals during three field surveys conducted in early spring of 2009 (n=28), 2010 (n=17) and 2011 (n=13), under a special permit from the Poyang Lake Fishery and Fishing Administration Office of Jiangxi Province. Gender was identified by visual inspection of the genital parts. In total, 76 males and 77 females were sampled. The sampling was conducted in accordance with the Regulations of the People's Republic of China for the Implementation of Wild Aquatic Animal Protection promulgated in 1993 by the Food and Agriculture Organization of the United Nations (FAO, FAOLEX No. LEX-FAOC011943, http://www.fao.org/faolex/results/details/en/?details=LEX-FAOC011943), adhering to all ethical guidelines and legal requirements in China.

Total genomic DNA was isolated from blood samples using the Whole Genome DNA Extraction Kit (SBS, Shanghai Inc.) following the manufacturer's instructions. For tissue samples, total genomic DNA was extracted using a standard proteinase K digestion and phenol/chloroform extraction protocol[53].

Microsatellite and mitochondrial data set. Microsatellite and mitochondrial (mtDNA) data have been previously obtained and described by Chen *et al.*[20]. All 153 porpoises were genotyped at 11 polymorphic microsatellite loci (YFSSR1, YFSSR42, YFSSR59, YFSSR5, YFSSR40 from *N. p. asiaeorientalis*[54,55], NP391, NP404, NP409, NP464, NP428 from *N. phocaenoides*[56,57], and PPHO130 from *Phocoena phocoena*[58]. The genotyping protocol and quality checks have been described in Chen *et al.*[20]. In the subsequent analyses of the microsatellites, we only kept individuals for which we had at least 50% of the locus available (i.e. at least 5 microsatellite loci and the mtDNA or 6 microsatellite loci). Those filters excluded of 5 individuals, leading to

399 a final data set of 148 individuals. The mtDNA data includes a 597 base-pairs fragment of the
400 hyper-variable region 1 of the control region successfully sequenced for 129 individuals (see
401 Chen *et al.*[20] for details on the PCR and sequencing procedures). The microsatellite and
402 mtDNA-CR data are available in Supplementary Dataset S1.
403
404
405 ***Population genetic structure.*** We used the Bayesian model-based clustering of *STRUCTURE*
406 v2.3.4[26] to estimate the admixture proportions for each individual to each cluster identified in
407 the microsatellite data. *STRUCTURE* cluster individual multilocus genotypes into K groups
408 while minimizing departures from Hardy-Weinberg and Linkage Equilibria. We used the
409 admixture *Locprior* model with correlated allele frequencies designed to detect weak signals
410 of genetic structure without introducing bias or forcing the clustering[26]. We conducted a
411 series of independent runs with different value for *K* from 1 to 5. Each run used $1x10^6$
412 iterations after a burn-in of length $2x10^5$. To assess that convergence of the Monte Carlo
413 Markov Chains (MCMCs) had been reached, we performed 10 independent replicates for
414 each *K* and checked the consistency of results using *CLUMPAK*[59]. We assessed which *K* value
415 best fit with the data using (1) the likelihood of each *K*, following *STRUCTURE*'s user manual;
416 (2) its rate of change with increasing *K*[60]; and (3) visual inspection of newly created clusters
417 with increasing *K*[61]. Post-processing of the results, including generation of barplots, was
418 conducted using *CLUMPAK*[59]. The geographic distribution of each group (Fig. 2) was mapped
419 using the R statistical package *MARMAP* v.0.9.5[62] and ETOPO dataset[63].
420
421 The genetic structure in the microsatellite data was further inspected using a Discriminant
422 Analysis of Principal Components (DAPC)[28] and a Principal Component Analysis (PCA)[27,29].
423 These exploratory methods does not rely on any model assumptions and provides a
424 complementary validation of the structure depicted by *STRUCTURE*[64]. These analysis were
425 conducted using *adegenet* 2.0.1 package[65] for R[66] on centred genetic data (i.e. set to a mean
426 allele frequency of zero), with missing data replaced by the mean, following the authors'
427 recommendations. The PCA was used to display the individual multilocus genotypes into a
428 reduced multidimensional space defined by the first two principal components (PCs),
429 colour-coding each individual according to the clustering identified by *STRUCTURE*, in order to
430 assess the congruence. The DAPC method identifies genetic clusters by optimizing the

difference between predefined groups and minimizing the variation within those groups[28]. It first reduces the number of variables using a PCA and then maximises the differences between groups using a Discriminant Analysis. The DAPC was performed with prior information on groups using the clusters defined by *STRUCTURE*. The number of PCs to retain and the reliability of the DAPC were determined using the cross-validation approach present in the *adegenet* 2.0.1 package[65] for R[66]. As a result of this cross-validation step, a total of 60 PCs and 4 discriminant functions were retained to describe the relationship between the clusters. This number of PCs captured 80% of the total variation and provided the highest percent of correctly predicted subsamples with the lowest error. Finally, the score of each individual for the first two discriminant functions (DFs) were plotted as scatter plot in R and the memberships probability of each individual to the clusters defined by the DAPC were plotted as barplot in R.

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*[67], and the observed and expected heterozygosity (H_o and H_e) computed with *GenAlEx* v6.5[68]. Departures from Hardy-Weinberg were tested using 10^4 permutations with *FSTAT* v2.9.3.2[70] and quantified using F_{IS} and F_{ST}[69] in *GenAlEx* v6.5[68]. We applied a Bonferonni correction to correct for multiple tests.

Mitochondrial genetic diversity was estimated for each of the genetically distinct 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[71] and θ_W based on the number of polymorphic sites[72]. All statistics were calculated using *DNASP* v5.10.01[73]. We used the F_{ST} statistics estimated from the average number of differences within and between populations[74]. Significance was tested with 1,000 permutations of Hudson's nearest neighbour distance *Snn* statistics, which measures how often the nearest neighbour of a sequence (in sequence space) is from the same population[75].

Contemporary effective population sizes. We used *NeEstimator* v2.01[30] as a first approach to estimate contemporary *Ne* based on linkage disequilibrium (LD) between loci, filtering out

463 rare alleles with a frequency $P_{crit} \leq 0.02$ that could bias *Ne* estimate[76]. The second method
464 implemented in *ONESAMP* v1.2[31] also use LD as summary statistics among others in an
465 Approximate Bayesian Computation (ABC) to estimate *Ne*, considering uniform prior between
466 2 and 500 of *Ne*.
467
468 ***Contemporary gene flow between populations.*** We estimated contemporary effective
469 migration rate (*m*) between populations using *BayesAss* v.3.0.3[32]. Preliminary runs were
470 performed to adjust the mixing parameters of the MCMC and ensure proposal acceptance
471 rates between 20% and 60% following authors' recommendations. We then performed 12
472 independent runs with different seeds, a burn-in of $5x10^6$ iterations followed by $2x10^8$
473 iterations, and a sampling parameter values every 2000 iterations. Convergence of the
474 MCMCs was checked by comparing the traces of each run using *Tracer* v1.5[77] and by
475 evaluating the Effective Sample Sizes (ESSs) of each parameter, keeping only runs where ESS ≥
476 200. Model fitting to the data was assessed using the Bayesian Deviance Index using the
477 R-script of Meirmans[33]. Runs that converged were combined to estimate the mean, median
478 and 95% Highest Probability Density interval for each parameter in *Tracer* v1.5. The effective
479 number migrants (N_e x *m*) per generation between populations was obtained by combining
480 *Ne* and *m* estimates.
481
482 ***Genetic inference of population demographic history.*** We used *DIYABC* v2.1[78] to estimate the
483 M_{GW} value and conduct $1x10^6$ coalescent simulations to produce a null distribution against
484 which the observed value could be compared. The *P*-value indicates the proportion of
485 simulations which provide a value below the observed one. For the mtDNA-CR, we used the
486 Tajima's *D*[35] and tested for significant departure from a null expectations using 10,000
487 coalescent simulations in DNAsp[73].
488
489 Next, we investigated the demographic history best describing the genetic diversity of the
490 combined microsatellite and mtDNA markers using a coalescent-based ABC approach[22]. We
491 subdivided our workflow in two nested parts (Fig. 4): identify the most likely population tree
492 topology for our dataset among 10 plausible scenarios describing different potential
493 population branching (Fig. 4a); and then test for evidence of changes in effective population
494 size in the ancestral and daughter populations (Fig. 4b). This later part was further subdivided

in two steps, (i) first testing for simple changes in effective population size (*Ne*) with six nested competing scenarios, and (ii) then testing whether adding the known population decline observed in each population over the last 50 yrs in the best scenario improved significantly the model fit to the data (Fig. 4b, Table S2 and S3).

For each part (Fig. 4), an ABC analysis was conducted using the program DIYABC v.2.1.0[78] applying the following steps (Fig. S4): (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 two approaches: the "standard" procedure relying on a logistic regression on 1% PODs producing SS values closest to the observed ones after a Linear Discriminant Analysis (ABC-LDA) as a pre-processing step[36] and the recently introduced Random Forest (ABC-RF) procedure[37,38]; (3) evaluate the confidence in scenario choice by estimating the type-I and type-II error rates based on simulated PODs using the ABC-LDA, as well as the prior error rate from the of the ABC-LDA and ABC-RF [37,38]; (4) estimate the marginal posterior distribution of each parameter based on the best model including (among other) *Ne* and times of population size changes and splits (*T*); and finally, (5) evaluate the goodness-of-fit of the fitted model to the data. Details are provided in Fig. 4 and in the supplementary materials (Appendix S1, Fig. S4, Table S2, S3).

Acknowledgments.

We thank Pr. JL Olsen for constructive comments. 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. This work was also supported by Major Project of Natural Science Research in Anhui Province (No. KJ2016A863).

Author 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,

YBC and FL 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.

Additional information

Data accessibility. All data generated during this study are included in this article (and its Supplementary Information files).

Competing financial interests: the authors declare no competing financial interests.

Supplementary information accompanies this paper at http://www.nature.com/srep

---

References

1. Marko, P. B. & Hart, M. W. The complex analytical landscape of gene flow inference. ***Trends Ecol. Evol.*** **26,** 448–456 (2011).
2. Epps, C. W. & Keyghobadi, N. Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. ***Mol. Ecol.*** **24,** 6021–6040 (2015).
3. Fahrig, L. Effects of Habitat Fragmentation on Biodiversity. ***Annu. Rev. Ecol. Evol. Syst.*** **34,** 487–515 (2003).
4. Hale, M. L. ***et al.*** Impact of landscape management on the genetic structure of red squirrel populations. ***Science*** **293,** 2246–2248 (2001).
5. Mech, S. G. & Hallett, J. G. Evaluating the effectiveness of corridors: a genetic approach. ***Conserv. Biol.*** **5,** 467–474 (2001).
6. Carroll, C., McRae, B. H. & Brookes, A. 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 (2012).
7. Turvey, S. T. ***et al.*** First human-caused extinction of a cetacean species? ***Biol. Lett.*** **3,** 537–540 (2007).
8. Gao, A. & Zhou, K. Geographical variation of external measurements and three subspecies of ***Neophocaena phocaenoides*** in Chinese waters. ***Acta Theriol. Sin.*** **15,** 81–92 (1995).
9. Wang, D., Turvey, S. T., Zhao, X. & Mei, Z. ***Neophocaena asiaeorientalis ssp. The IUCN Red List of Threatened Species*** e.T43205774A45893487 (accessed on 2017-10-02) (2013). doi:10.2305/IUCN.UK.2013-1.RLTS.T43205774A45893487.en
10. Zhang, X. ***et al.*** The population of finless porpoise in the middle and lower reaches of Yangtze river. ***Acta Theriol. Sin.*** **13,** 260–270 (1993).
11. Yang, J., Xiao, W., Kuang, X., Wei, Z. & Liu, R. Studies on the distribution, population size and the active regularity of ***Lipotes vexillifer*** and ***Neophocaena phocaenoides*** in Dongting Lake and Boyang Lake. ***Resour. Env. Yangtze Valley*** **9,** 444–450 (2000).
12. Xiao, W. & Zhang, X. Distribution and population size of Yangtze finless porpoise in Poyang Lake and its branches. ***Acta Theriol. Sin.*** **22,** 7–14 (2002).
13. Zhao, X. ***et al.*** Abundance and conservation status of the Yangtze finless porpoise in the Yangtze River, China. ***Biol. Conserv.*** **141,** 3006–3018 (2008).
14. Mei, Z. ***et al.*** The Yangtze finless porpoise: On an accelerating path to extinction? ***Biol. Conserv.*** **172,** 117–123 (2014).
15. Mei, Z. ***et al.*** Accelerating population decline of Yangtze finless porpoise (***Neophocaena asiaeorientalis asiaeorientalis***). ***Biol. Conserv.*** **153,** 192–200 (2012).
16. Fontaine, M. C. ***et al.*** 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 (2012).
17. Fontaine, M. C. ***et al.*** Postglacial climate changes and rise of three ecotypes of harbour porpoises, ***Phocoena phocoena***, in western Palearctic waters. ***Mol. Ecol.*** **23,** 3306–3321 (2014).

18. Yang, G., Guo, L., Bruford, M. W., Wei, F. & Zhou, K. Mitochondrial phylogeography and population history of finless porpoises in Sino-Japanese waters. ***Biol. J. Linn. Soc.*** **95,** 193–204 (2008).
19. Lin, W. ***et al.*** Phylogeography of the finless porpoise (genus *Neophocaena*): testing the stepwise divergence hypothesis in the northwestern Pacific. ***Sci. Rep.*** **4,** 6572 (2014).
20. Chen, M. ***et al.*** 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 (2014).
21. Beaumont, M. A. Approximate Bayesian Computation in Evolution and Ecology. ***Annu. Rev. Ecol. Evol. Syst.*** **41,** 379–406 (2010).
22. Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & Francois, O. Approximate Bayesian Computation (ABC) in practice. ***Trends Ecol. Evol.*** **25,** 410–418 (2010).
23. Cornuet, J.-M., Ravignie, V. & Estoup, A. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). ***BMC Bioinfor.*** **11,** 401 (2010).
24. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. ***Genetics*** **155,** 945–959 (2000).
25. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. ***Genetics*** **164,** 1567–1587 (2003).
26. Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. Inferring weak population structure with the assistance of sample group information. ***Mol. Ecol. Resour.*** **9,** 1322–1332 (2009).
27. Jombart, T., Pontier, D. & Dufour, A.-B. Genetic markers in the playground of multivariate analysis. ***Heredity*** **102,** 330–341 (2009).
28. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. ***BMC Genet.*** **11,** 94 (2010).
29. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. ***PLoS Genet.*** **2,** e190 (2006).
30. Do, C. ***et al.*** NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (***Ne***) from genetic data. ***Mol. Ecol. Resour.*** **14,** 209–214 (2014).
31. Tallmon, D. A., Koyuk, A., Luikart, G. & Beaumont, M. A. ONESAMP: a program to estimate effective population size using approximate Bayesian computation. ***Mol. Ecol. Resour.*** **8,** 299–301 (2008).
32. Wilson, G. A. & Rannala, B. Bayesian inference of recent migration rates using multilocus genotypes. ***Genetics*** **163,** 1177–1191 (2003).
33. Meirmans, P. G. Nonconvergence in Bayesian estimation of migration rates. ***Mol. Ecol. Resour.*** **14,** 726–733 (2014).
34. Garza, J. C. & Williamson, E. G. Detection of reduction in population size using data from microsatellite loci. ***Mol. Ecol.*** **10,** 305–318 (2001).
35. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. ***Genetics*** **123,** 585–595 (1989).
36. Estoup, A. ***et al.*** Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. ***Mol. Ecol. Resour.*** **12,** 846–855 (2012).
37. Pudlo, P. ***et al.*** Reliable ABC model choice via random forests. ***Bioinformatics*** **32,** 859–866 (2016).
38. Fraimout, A. ***et al.*** Deciphering the routes of invasion of *Drosophila suzukii* by means of ABC random forest. ***Mol. Biol. Evol.*** **34,** 980–996 (2017).
39. Wu, L., Li, F., Zhu, C., Li, L. & Li, B. Holocene environmental change and archaeology, Yangtze River Valley, China: Review and prospects. ***Geosci. Front.*** **3,** 875–892 (2012).
40. Zhang, X. Y., Cai, S. M. & Sun, S. C. Evolution of Dongting Lake since Holocene. ***J. Lake Sci.*** **60,** 13–21 (1994).
41. Fang, J.-Q. Lake evolution during the last 3000 years in China and its implications for environmental change. ***Quat. Res.*** **39,** 175–185 (1993).
42. Dong, L. ***et al.*** Yangtze finless porpoises along the main channel of Poyang Lake, China: Implications for conservation. ***Mar. Mamm. Sci.*** **31,** 612–628 (2015).
43. Wei, Z., Wang, D., Zhang, Q., Wang, K. & Kuang, X. 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 (2002).
44. Huang, S.-L. ***et al.*** Saving the Yangtze finless porpoise: Time is rapidly running out. ***Biol. Conserv.*** **20**, 40-46 (2017).
45. Fontaine, M. C. ***et al.*** Genetic and historic evidence for climate-driven population fragmentation in a top

cetacean predator: the harbour porpoises in European water. ***Proc. R. Soc. B*** **277,** 2829–2837 (2010).

46. Louis, M. ***et al.*** Habitat-driven population structure of bottlenose dolphins, ***Tursiops truncatus***, in the North-East Atlantic. ***Mol. Ecol.*** **23,** 857–874 (2014).
47. Goossens, B. ***et al.*** Genetic Signature of Anthropogenic Population Collapse in Orang-utans. ***PloS Biol.*** **4,** e25 (2006).
48. Weckworth, B. V. ***et al.*** Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. ***Proc. R. Soc. B*** **280,** 20131756 (2013).
49. Waples, R. S. Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? ***Mol. Ecol.*** **14,** 3335–3352 (2005).
50. Palstra, F. P. & Fraser, D. J. Effective/census population size ratio estimation: a compendium and appraisal. ***Ecol. Evol.*** **2,** 2357–2365 (2012).
51. Hare, M. P. ***et al.*** Understanding and Estimating Effective Population Size for Practical Application in Marine Species Management. ***Conserv. Biol.*** **25,** 438–449 (2011).
52. Waples, R. S. & England, P. R. Estimating Contemporary Effective Population Size on the Basis of Linkage Disequilibrium in the Face of Migration. ***Genetics*** **189,** 633–644 (2011).
53. Sambrook, J. & Russell, D. W. ***Molecular Cloning: A Laboratory Manual***. (CSHL Press, 2012).
54. Zheng, J. S. ***et al.*** Development and characterization of polymorphic microsatellite loci in the endangered Yangtze finless porpoise (***Neophocaena phocaenoides asiaeorientalis***). ***Conserv. Genet.*** **9,** 1007–1009 (2008).
55. Zhou, Z., Zheng, J. S., Chen, M. M., Zhao, Q. Z. & Wang, D. Genetic evaluation and development prognosis on ***ex situ*** conserved Yangtze finless porpoise living in Tian-E-Zhou National Natural Reserve. ***Acta Theriol. Sin.*** **36,** 403–411 (2012).
56. Chen, L., Bruford, M. & Yang, G. Isolation and characterization of microsatellite loci in the finless porpoise (***Neophocaena phocaenoides***). ***Mol. Ecol. Notes*** **7,** 1129–1131 (2007).
57. Chen, L. & Yang, G. Development of tetranucleotide microsatellite loci for the finless porpoise (***Neophocaena phocaenoides***). ***Conserv. Genet.*** **9,** 1033–1035 (2008).
58. Rosel, P. E., France, S. C., Wang, J. Y. & Kocher, T. D. Genetic structure of harbour porpoise ***Phocoena phocoena*** populations in the northwest Atlantic based on mitochondrial and nuclear markers. ***Mol. Ecol.*** **8,** S41–54 (1999).
59. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. ***Mol. Ecol. Resour.*** **15,** 1179–1191 (2015).
60. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. ***Mol. Ecol.*** **14,** 2611–2620 (2005).
61. Vercken, E. ***et al.*** Glacial refugia in pathogens: European genetic structure of anther smut pathogens on Silene latifolia and Silene dioica. ***PLoS Pathog.*** **6,** e1001229 (2010).
62. Pante, E. & Simon-Bouhet, B. marmap: A package for importing, plotting and analyzing bathymetric and topographic data in R. ***PLoS ONE*** **8,** e73051 (2013).
63. Amante, C. & Eatkins, B. W. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. (2009). doi:10.7289/V5C8276M [accessed 2017-10-02].
64. Francois, O. & Durand, E. Spatially explicit Bayesian clustering models in population genetics. ***Mol. Ecol. Resour.*** **10,** 773–784 (2010).
65. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. ***Bioinformatics*** **24,** 1403–1405 (2008).
66. R Core Team. R: A Language and Environment for Statistical Computing. ***R Foundation for Statistical Computing*** (2016).
67. Szpiech, Z. A., Jakobsson, M. & Rosenberg, N. A. ADZE: a rarefaction approach for counting alleles private to combinations of populations. ***Bioinformatics*** **24,** 2498–2504 (2008).
68. Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. ***Bioinformatics*** **28,** 2537–2539 (2012).
69. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. ***Evolution*** 1358–1370 (1984).
70. Goudet, J. 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 2017-10-02]. (2001).
71. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. ***Genetics*** **105,** 437–460 (1983).

691 72. Watterson, G. A. On the number of segregating sites in genetical models without recombination. ***Theor.***
692 ***Popul. Biol.*** **7,** 256–276 (1975).
693 73. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.
694 ***Bioinformatics*** **25,** 1451–1452 (2009).
695 74. Hudson, R. R., Boos, D. D. & Kaplan, N. L. A statistical test for detecting geographic subdivision. ***Mol.***
696 ***Biol. Evol.*** **9,** 138–151 (1992).
697 75. Hudson, R. R. A new statistic for detecting genetic differentiation. ***Genetics*** **155,** 2011–2014 (2000).
698 76. Waples, R. S. & Do, C. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic
699 markers: a largely untapped resource for applied conservation and evolution. ***Evol. Appl.*** **3,** 244–262
700 (2010).
701 77. Rambaut, A. Tracers v1.5 Available from http://tree.bio.ed.ac.uk/software/tracer/ [accessed
702 2017-07-05]. (2007).
703 78. Cornuet, J.-M. ***et al.*** DIYABC v2.0: a software to make approximate Bayesian computation inferences
704 about population history using single nucleotide polymorphism, DNA sequence and microsatellite data.
705 ***Bioinformatics*** **30,** 1187–1189 (2014).
706

Figure legends

Figure 1 | Maps showing the sampling distribution of the Yangtze finless porpoises in the Yangtze River. The top-left insert shows the location of the studied area highlighted by a red rectangle. On the right, the map shows the sampling locations (orange triangles) and their acronyms based on the neighbouring cities. Figure created using ArcGIS 10.3 software using the open source data from the ETOPO1 Global Relief Model[63] (https://www.ngdc.noaa.gov/mgg/global/).

Figure 2 | (a) Population structure estimated using the Bayesian clustering approach of *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 result was found out of the 10 replicates. (b) *DAPC* cluster membership probability plot of the 148 individuals. (c) Scatter plot showing the first two discriminant functions (DFs) of the *DAPC*. (d) Geographical distribution of the *STRUCTURE* admixture proportions and mitochondrial haplotype frequencies per localities. The mtDNA map is modified from Chen *et al.*[20]. Panel (a) was created using CLUMPAK[59], panel (b) and (c) using R v.3.4.0[66], and panel (d) using R v.3.4.0[66], the package MARMAP v.0.9.5[62], and the open source ETOPO1 Global Relief Model[63] (https://www.ngdc.noaa.gov/mgg/global/).

Figure 3 | Recent gene flow (*Ne* x *m*) between populations estimated from *ONESAMP* and *BayesAss*. 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 are provided in the circles.

Figure 4 | Schematic diagram of the 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. The posterior probability estimated using the ABC-LDA procedure is provided for each scenario. * indicates the posterior probability estimated using the ABC-RF is also provided for the best scenario of each step. See the main

738 text, appendix S1, Table S2-S3 for further details.

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)	148	57.2 ± 0.32 (58)	16.0 ± 0.4 (17)	16.4 ± 0.36 (17)	48.1 ± 1.3 (53)
NA	-	1.4%	4.8%	3.7%	9.3%
A_r ± SE	-	4.82 ± 0.40	5.19 ± 0.52	5.25 ± 0.46	5.93 ± 0.45
pA ± SE	-	0.38 ± 0.09	0.55 ± 0.21	0.48 ± 0.12	0.69 ± 0.18
H_o ± SD	-	0.60 ± 0.06	0.70 ± 0.047	0.69 ± 0.05	0.72 ± 0.04
H_e ± SD	-	0.62 ± 0.05	0.62 ± 0.05	0.63 ± 0.04	0.68 ± 0.03
F_{IS} ± SD	-	0.041 ± 0.03^{ns}	-0.123 ± 0.06^{ns}	-0.112 ± 0.03^{ns}	-0.057 ± 0.03^{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
θ_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 26 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[34]; N_{mtDNA}, MtDNA sample size; *S*, number of segregating sites; *Singleton*: rare mutation observed only in one sequence among all; *Share P.*, shared polymorphism (mutation observed in at least two or more sequences) also known as parsimony informative site; *#hap*, number of haplotypes; *Hd*, haplotype diversity; π, nucleotide diversity; θ_W, theta from *S* or Theta-Watterson; *D*, Tajima's *D*. † The significance level of the M_{GW} statistic was evaluated in DIYABC[78] 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[73]. ns: not significant (p-value > 0.05); * p-value ≤ 0.05; ** p-value ≤ 0.01; *** p-value ≤ 0.001

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

$F_{ST\text{-}mtDNA}$ / $F_{ST\text{-}mic}$	XCSS	PY	TL	Admix
XCSS	–	0.590***	0.875***	0.446***
PY	0.070*** [0.038-0.104]	–	0.109*	0.011**
TL	0.052*** [0.034-0.070]	0.050*** [0.024-0.075]	–	0.131^{ns}
Admix	0.029*** [0.011-0.047]	0.023*** [0.014-0.032]	0.023*** [0.010-0.037]	–

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

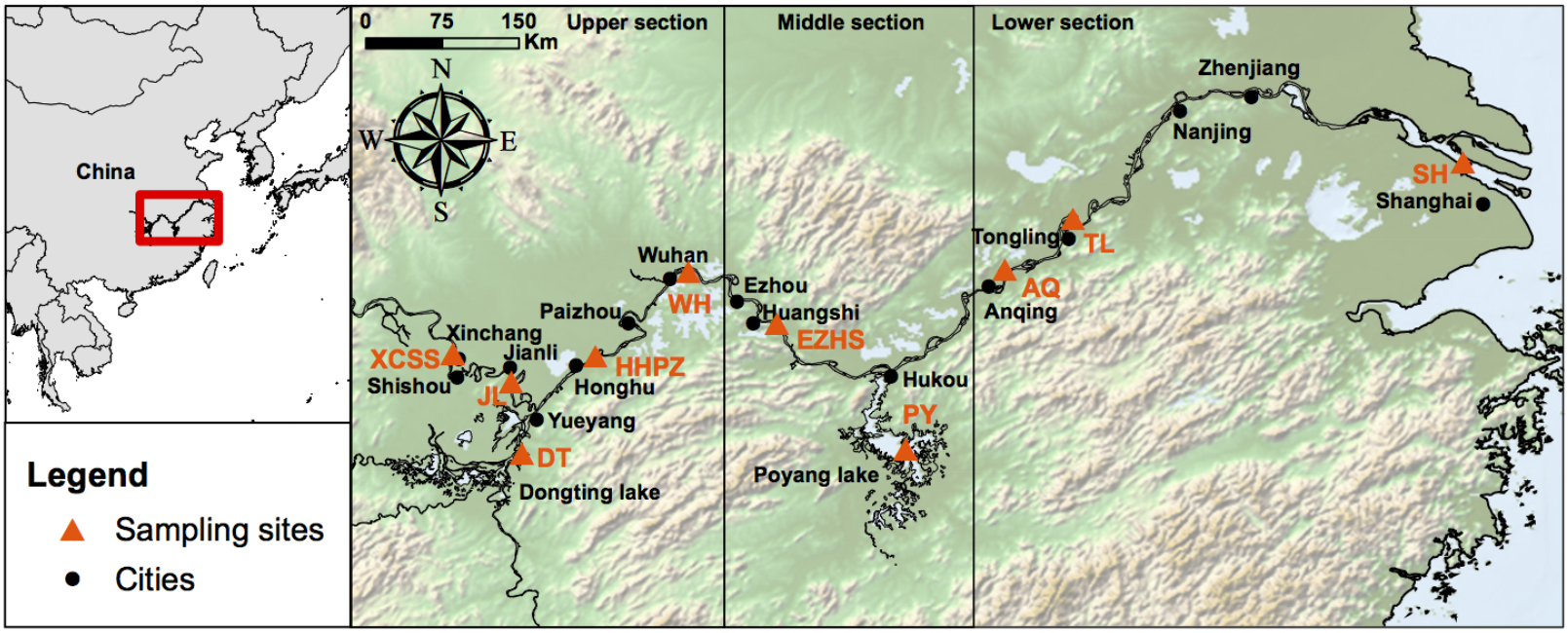
762 **Table 3 |** Effective population size (*Ne*) estimated in each population of the Yangtze finless porpoise.
763 Values have been calculated using an estimator based on linkage disequilibrium between loci in
764 *NeEstimator*[30], using an ABC approach in *ONESAMP*[31], and using DIYABC under SC2

765

	XCSS	TL	PY	Admix	Total
LD–NeEstime *Mean [95%CI]*	16 [7-53]	56 [18-∞]	92 [45-486]	86 [45-308]	251
ONESAMP *Mean [95%CI]*	22 [11-26]	18 [15-28]	62 [45-115]	80 [56-142]	182
DIYABC *Mode (mean) [95%CI]*	14 (42) [7 – 95]	32 (55) [11 – 98]	35 (50) [11 – 66]	-	-

766

767



(a)

XCSS (17) DT (20) HHPZ (13) WH (3) EZHS (13) PY (58) AQ (4) TL (17) SH (3)

K=2 10/10

K=3 9/10

K=4 8/10

(b)

G1 G2 G4 G3

Admix

(c)

G4

G1 Admix G2

DF1

G3

DF2

DA eigenvalues

(d)

K=2

K=3

K=4

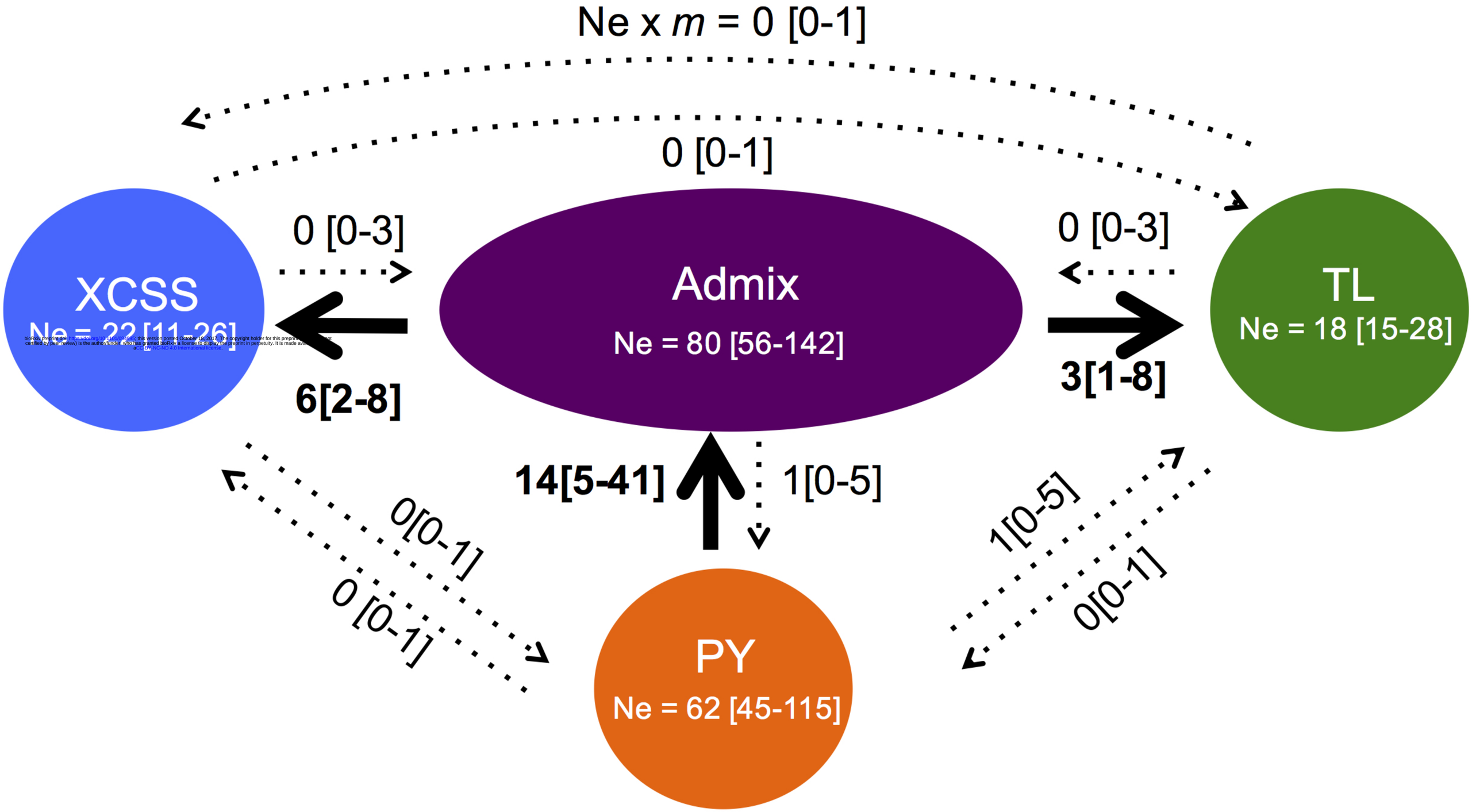
MtDNA

Yangtse River

SH WH EZHS TL AQ XCSS HHPZ DT PY

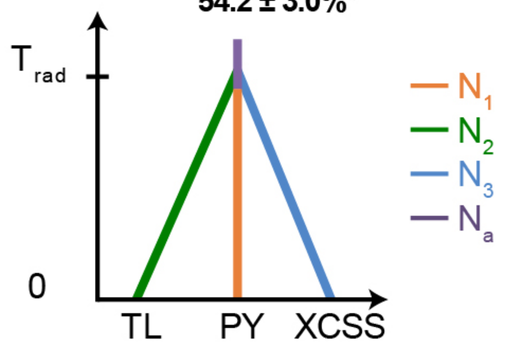
36 34 32 30 28 26

110 112 114 116 118 120 122 124

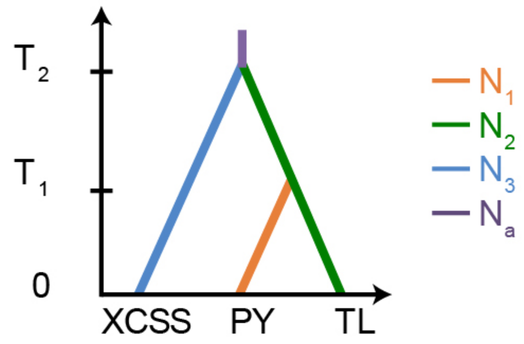


(a) Select topology

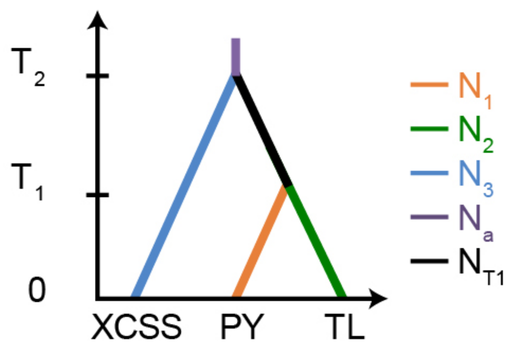
SC1 **66.7% [65.7 - 67.8]** **54.2 ± 3.0%***



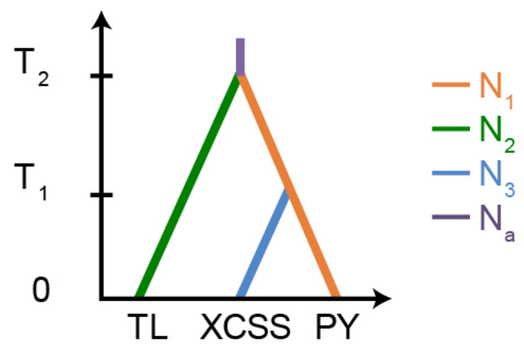
SC2 1.3 % [1.2 - 1.5]



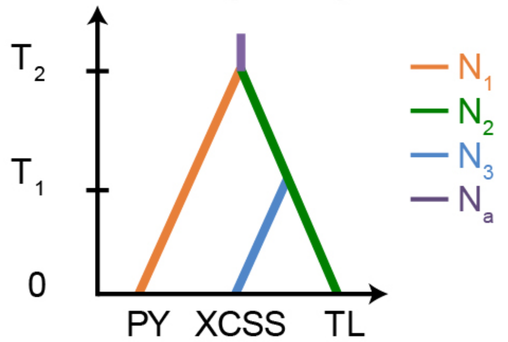
SC3 1.4 % [1.2 - 1.6]



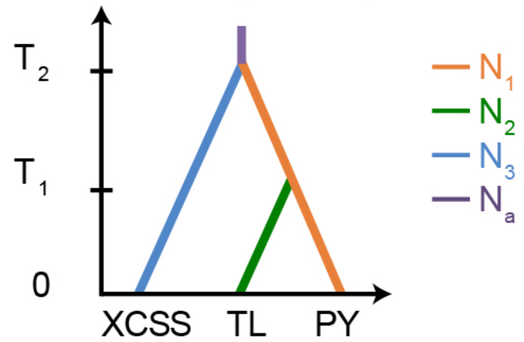
SC4 7.7 % [7.1 - 8.2]



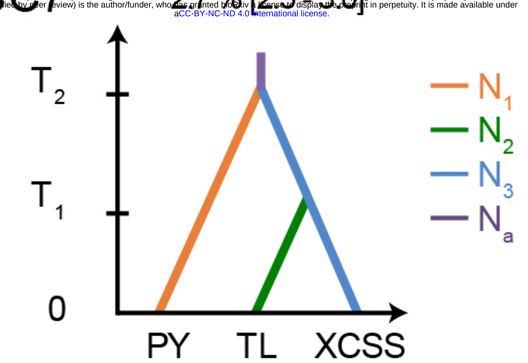
SC5 5.8 % [5.3 - 6.2]



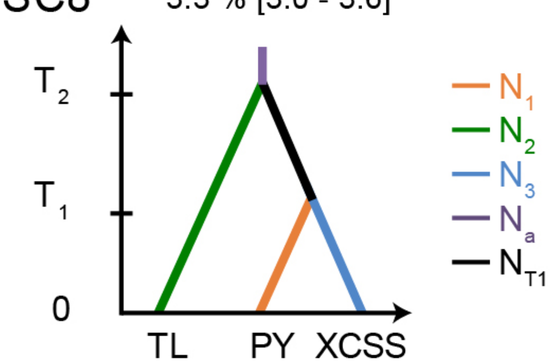
SC6 7.1 % [6.6 - 7.7]



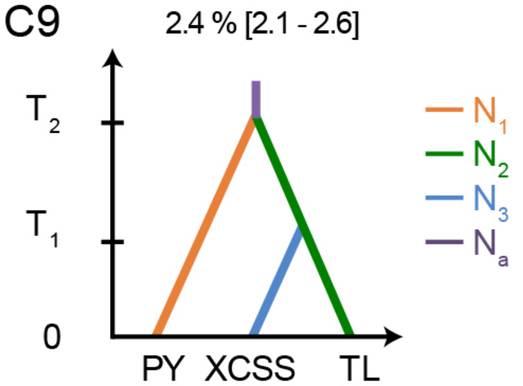
SC7 2.7 % [2.5 - 3.0]



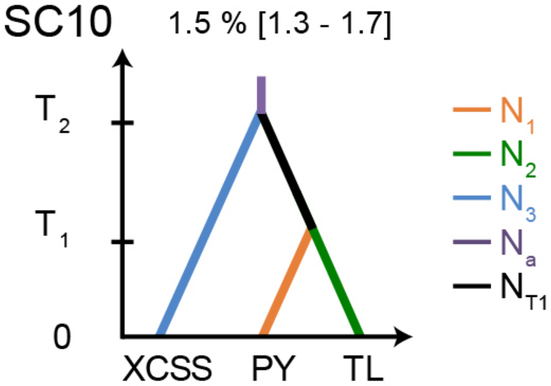
SC8 3.3 % [3.0 - 3.6]



SC9 2.4 % [2.1 - 2.6]



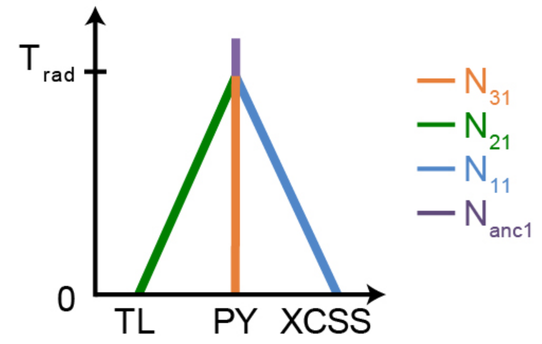
SC10 1.5 % [1.3 - 1.7]



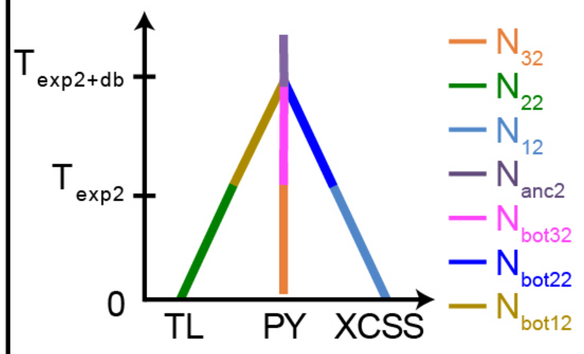
(b) Test of population size change

(i) Bottleneck

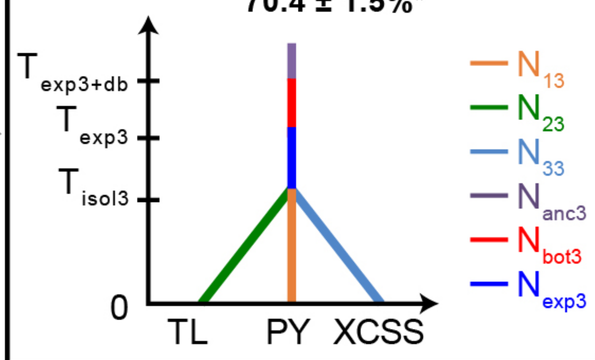
SC1 6.1 % [5.6 - 6.7]



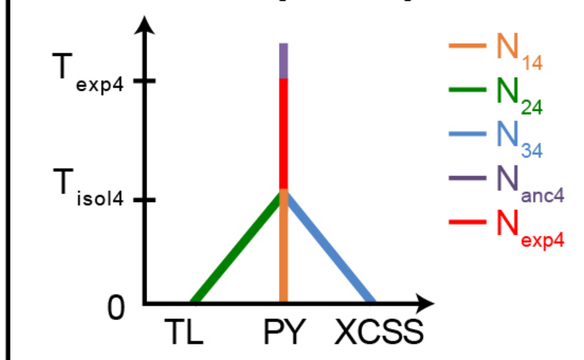
SC2 3.4 % [3.0 - 3.9]



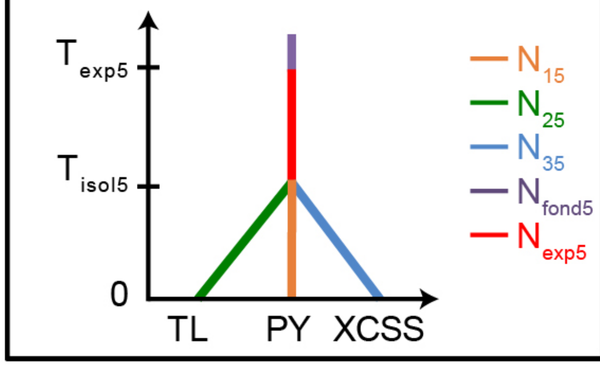
SC3 **77.8 % [76.8 - 78.9]** **70.4 ± 1.5%***



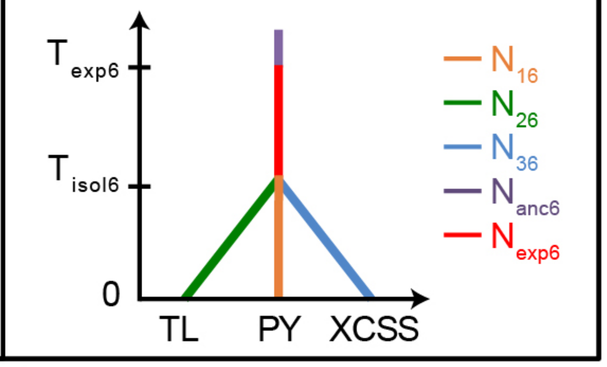
SC4 4.5 % [4.0 - 4.9]



SC5 7.2 % [6.7 - 7.7]

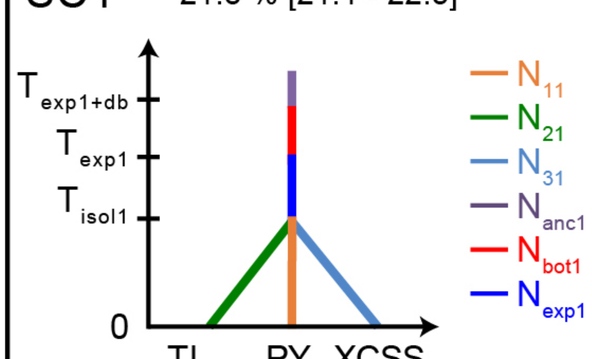


SC6 0.1 % [0.0 - 0.1]



(ii) Bottleneck + population crash

SC1 21.8 % [21.1 - 22.5]



SC2 **78.2 % [77.5 - 78.9]** **68.4 ±2.1%***