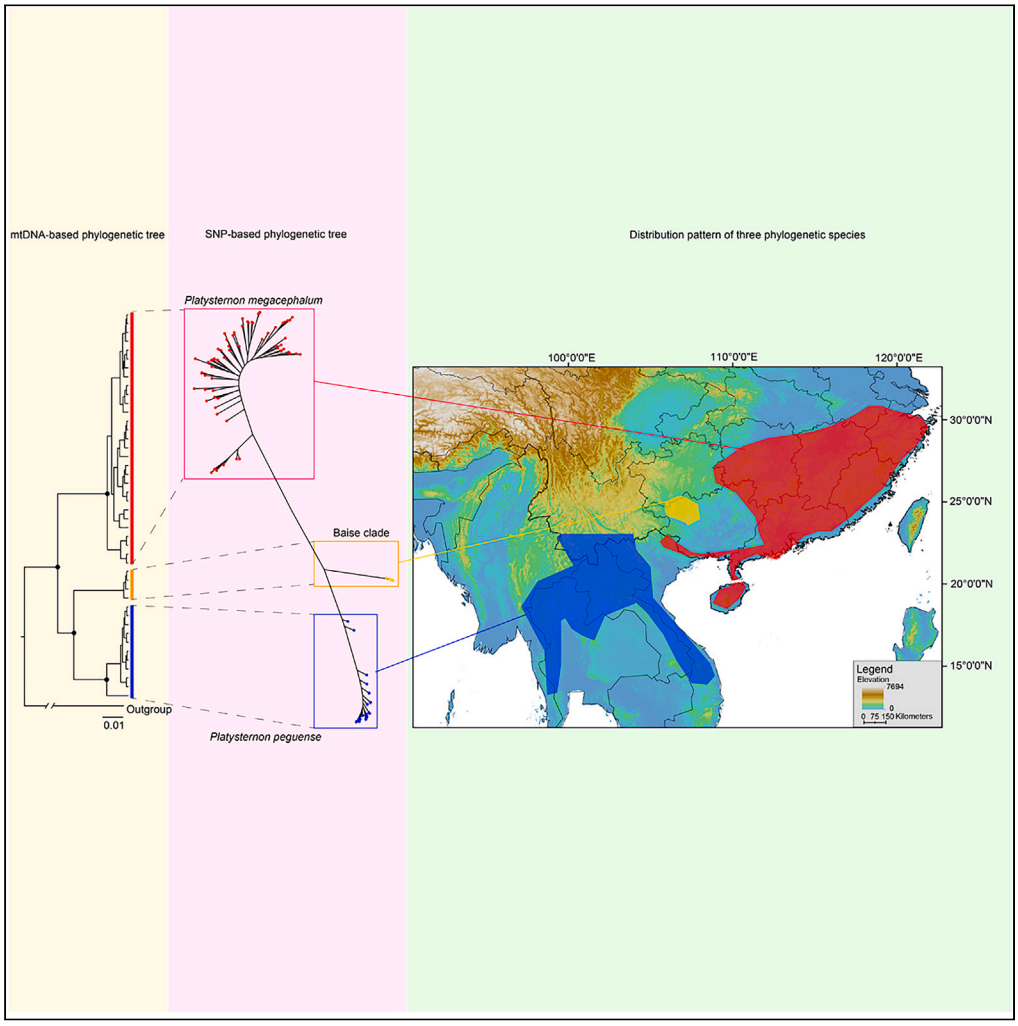
iScience

CellPress
OPEN ACCESS

Article

Genomic analyses reveal three phylogenetic species and their evolutionary histories in the big-headed turtle



Shiping Gong, Chatmongkon Suwannapoom, Minh Le, Truong Quang Nguyen, Yan Ge, Yufeng Wei, Yangchun Gao

gaoyc0412@163.com

Highlights

Three phylogenetic species (*P. megacephalum*, *P. peguense*, and Baise clade)

The key role of lowland areas in driving speciation in the big-headed turtle

Baise clade is more vulnerable to environmental disturbance

Insights into the taxonomy and scientific conservation of the big-headed turtle

Gong et al., iScience 26, 107343
August 18, 2023
https://doi.org/10.1016/j.isci.2023.107343

iScience

CellPress
OPEN ACCESS

Article

Genomic analyses reveal three phylogenetic species and their evolutionary histories in the big-headed turtle

Shiping Gong,[1,2] Chatmongkon Suwannapoom,[3] Minh Le,[4,6] Truong Quang Nguyen,[5] Yan Ge,[2] Yufeng Wei,[2] and Yangchun Gao[2,7,*]

SUMMARY

The critically endangered big-headed turtle (*Platysternon megacephalum*) is currently classified into three subspecies. However, the classification is still controversial and their evolutionary histories are still unclear. Here, multiple genetic analyses consistently revealed three phylogenetic groups with substantial genetic divergences and distinct demographic histories, suggesting three phylogenetic species (*P. megacephalum*, *P. peguense*, and Baise clade). Phylogeographical analyses revealed that the Red River plains and Guangxi basins are largely coincident with the boundaries between the three phylogenetic species, highlighting the key role of lowland areas in driving speciation in the big-headed turtle. The Baise clade is characterized by high-linkage disequilibrium but the lowest effective population size, indicating that the cryptic phylogenetic species is more vulnerable to human activities and environmental disturbance, and urgently needs more protection. Our findings provide fundamental insights into the taxonomy and scientific conservation of the family Platysternidae.

INTRODUCTION

Delimiting threatened species is a precondition for the formulation of an effective conservation policy, as it can inform policy decisions about the high costs of failing to adopt species-specific conservation measures.[1,2] Currently, the delimitation of many threatened species has been mainly based on limited samples and morphological discrepancies, and has barely considered genetic differentiation using genome-wide markers, leading to inappropriate taxonomy and outdated protected species lists, which have been proven to hinder the conservation of threatened species.[3,4] This is especially true for turtles, considering that they face the highest risk of extinction in any sizable vertebrate group.[5]

The big-headed turtle (*Platysternon megacephalum*) is the sole member in the family Platysternidae, with its distribution range including China, Vietnam, Cambodia, Laos, Thailand, and Myanmar.[6] Due to overexploitation (for food, pets, and traditional medicine) and habitat destruction, the wild population of the big-headed turtle has suffered serious decline[7–9] and has been assessed as a critically endangered taxon on the IUCN Red List.[10] For example, 0–0.36 individuals/km of stream were detected in Guangdong and Hainan, while >120 individuals/km of stream with low poaching in Hong Kong.[7] To protect this critically endangered taxon, the big-headed turtle has been listed in Appendix I of CITES. Currently, the captive breeding of the big-headed turtle has been successful in some turtle farms and reintroduction is a promising measure for the recovery of wild populations.[11] However, big-headed turtles on farms are generally collected from markets and their geographical sources are often unknown. As a result, releasing non-native or hybrid individuals into the wild may increase the risk of extinction of local populations, as is the case for the critically endangered Chinese giant salamander (*Andrias davidianus*).[12] Although the importance of genetic assessment before release has been highlighted for threatened species, taxon uncertainty can impede its implementation.[13,14]

Based on morphological characteristics, the big-headed turtle was previously classified into five subspecies: *P. m. megacephalum* Gray, 1831, in southern China; *P. m, peguense* Gray, 1870 from western Vietnam westward to southern Burma; *P. m. vogeli* Wermuth, 1969, in northwestern Thailand; *P. m. tristernalis*, Schleich & Gruber, 1984, in Yunnan in China; and *P. m. shiui* Ernst & McCord, 1987, in northern Vietnam

[1]College of Life Science and Technology, Jinan University, Guangzhou 510632, China

[2]Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Sciences, 105 Xingang Road West, Guangzhou 510260, China

[3]School of Agriculture and Natural Resources, University of Phayao, Phayao 56000, Thailand

[4]Faculty of Environmental Sciences, University of Science & Central Institute for Natural Resources and Environmental Studies (CRES), Vietnam National University, Hanoi, Vietnam

[5]Institute of Ecology and Biological Resources, Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi 10072, Vietnam

[6]Department of Herpetology, American Museum of Natural History, New York, NY, USA

[7]Lead contact

*Correspondence: gaoyc0412@163.com

https://doi.org/10.1016/j.isci.2023.107343

and Yunnan in China. Subsequent reanalysis based on more comprehensive morphological characteristics and geographical distribution revealed that *P. m. vogeli* and *P. m. tristernalis* should be placed together with *P. m. peguense* and *P. m. megacephalum*, respectively.[15] To date, the big-headed turtle is classified into three subspecies, based on morphology: *P. m. megacephalum* Gray, 1831 in southern China; *P. m. shiui* Ernst & McCord, 1987 in northern Vietnam and *P. m. peguense* Gray, 1870 from western Vietnam westward to southern Burma, and Yunnan in China.[6]

Morphologically, *P. m. megacephalum* is mainly characterized by having an unpatterned yellow plastron, most of the dorsal surface of *P. m. shiui* is speckled with colorful spots, and *P. m. peguense* has a dark seam-following plastral pattern.[16] However, the classification has remained controversial, largely due to morphological variation between or within geographic populations and lack of both whole-range samples and genome-wide data.[17] Based on our field investigations, some wild individuals of *P. m. megacephalum* in Heyuan, Meizhou (Guangdong Province, China), Yichun (Jiangxi Province, China) were found with colorful spots on the body surface, which resemble those of *P. m. shiui*,[17] thus challenging the validity of *P. m. shiui*.[16] Although three Hainan samples from the Field Museum of Natural History and the University of Michigan Museum of Zoology were morphologically similar to both *P. m. shiui* and *P. m. peguense*, it is still unclear whether the two subspecies were collected from the wild on Hainan Island.[15] Moreover, no samples of the big-headed turtle were collected from Myanmar, Thailand, Vietnam, Cambodia, and Hainan-China in previous molecular phylogenetic studies with limited genetic markers,[18,19] suggesting an urgent need for a comprehensive genomic analysis based on whole-range samples.

In this study, with a total of 143 samples from across the entire geographic range of the big-headed turtle, we used mitochondrial DNA fragments and genome-wide single nucleotide polymorphisms (SNPs) to reveal phylogenetic relationships, genetic divergences, and demographic history, with the aim of providing fundamental information for the taxonomy, conservation of genetic diversity, and wild population recovery of this critically endangered taxon.

RESULTS

Phylogenetic patterns

The concatenated mtDNA supermatrix (Data S1) resulted in a total of 44 haplotypes (Table S1). On the basis of these haplotypes, Bayesian (BI) phylogenetic analysis resolved five highly supported monophyletic clades (Figure S1). Clade I was comprised of 25 haplotypes from Anhui, Zhejiang, Fujian, Jiangxi, Hunan, Guizhou, Guangdong, Hong Kong, and eastern Guangxi of China. Clade II was comprised of 4 haplotypes from Hainan in China and northern Vietnam (north of the Red River). Clade III consisted of 10 haplotypes from Yunnan in China, Myanmar, Thailand, Laos and southern Vietnam (south of the Red River). Clade IV consisted of 1 haplotype with only two samples from Pingbian, Yunnan in China. Clade V consisted of 4 haplotypes from Baise, Guangxi in China (Figure 1A). Divergence time analysis revealed that the ancestors of the current big-headed turtle were split into two groups around 15.31 Ma. The first group consists of clade I and clade II, and the second group is composed of clade III-clade V, which initially differentiated about 8.81 Ma. (Figure S1).

RAD-seq produced a total of 793,511,052 clean reads with an average sequencing depth of 11.42× and an average mapping rate of 96.20% for each individual. After applying SNP calling and subsequent strict filtering steps, a total of 120,386 reliable SNPs (Data S2) were retained for further analyses. The maximum-likelihood (ML) phylogenetic analysis based on the genome-wide SNPs highly supported three (clade I, clade III, and clade V) of five monophyletic clades, two clades (clade II and clade IV) in BI analysis were not supported by ML analysis (Figures 1B and S1).

Large genetic divergence

Based on the genome-wide SNPs, ADMIXTURE and PCA consistently revealed significant genetic structures (Figures 2A, 2C and 2D). In the ADMIXTURE analysis, although K = 5 had the lowest cross-validation error value, similar values were also found at K = 3–6 (Figure 2B). At K = 3, the whole samples were divided into three clusters. Clade I and Northern Vietnam-Hainan samples formed a cluster; Clade III and Pingbian samples also formed a cluster; Clade V constituted a cluster on its own. At K = 4, Clade I and Northern Vietnam-Hainan samples formed separate clusters; some clade I samples collected from Guangdong and Guangxi of China were clustered with a little of Northern Vietnam-Hainan samples component, supporting the close genetic relationship between the Hainan populations and Guangdong/Guangxi

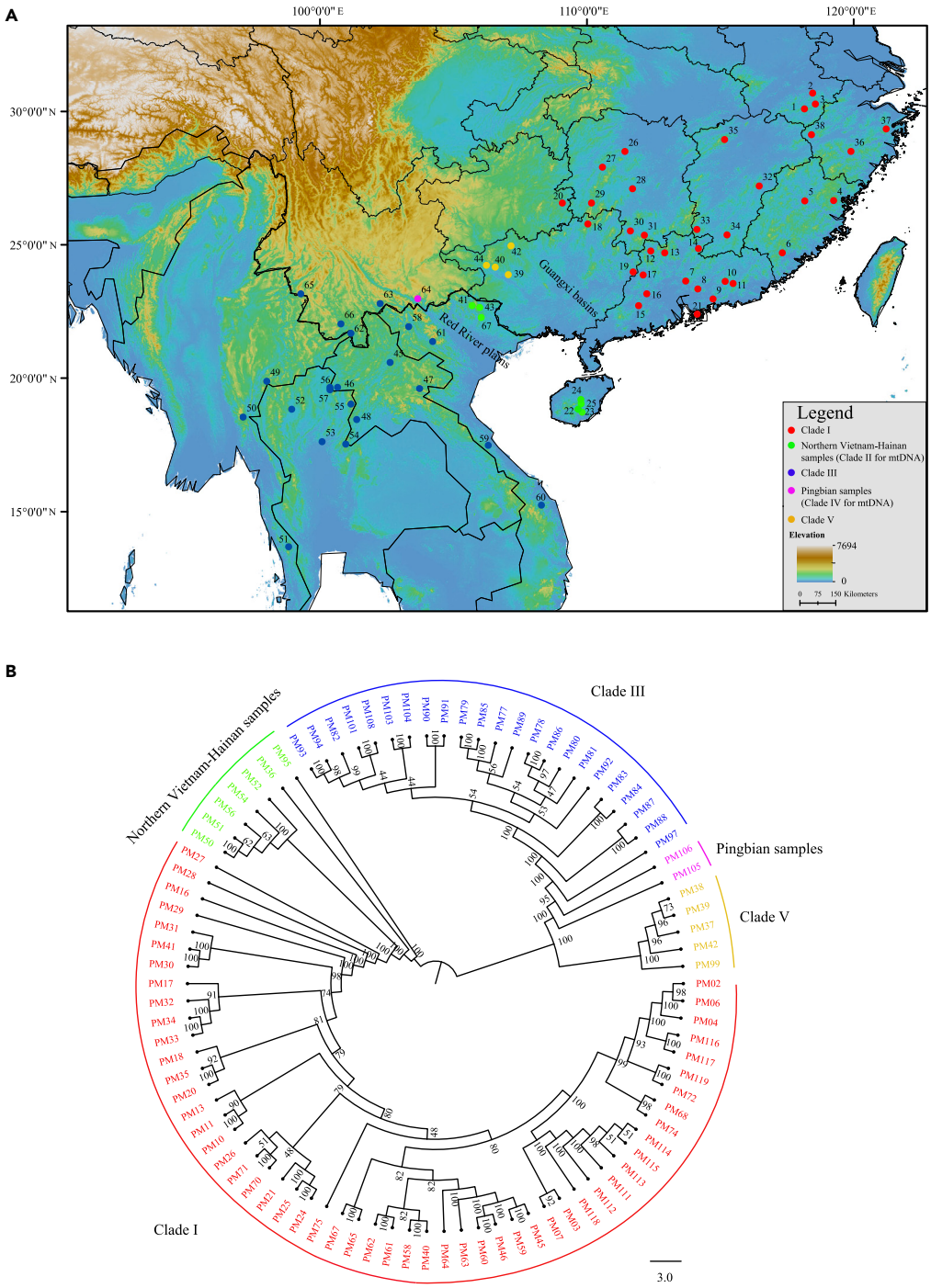


Figure 1. Phylogenetic inferences in big-headed turtles

(A) Sampling sites of the big-headed turtle; sampling site codes are in line with Table S2.

(B) Maximum likelihood tree based on genome-wide SNPs. The values on the tree nodes indicate the bootstrap support. Individuals with red, green, blue, pink, and orange colors indicate clade I, Northern Vietnam-Hainan samples, clade III, Pingbian samples, and clade V, respectively.

populations. At K = 5, one additional genetic cluster at the distribution boundaries (samples from eastern Guizhou and northern Hunan of China) were separated from clade I. At K = 6, clade III and Pingbian samples formed different clusters. Specifically, at K = 3–5, Pingbian samples were mainly found to be an admixture

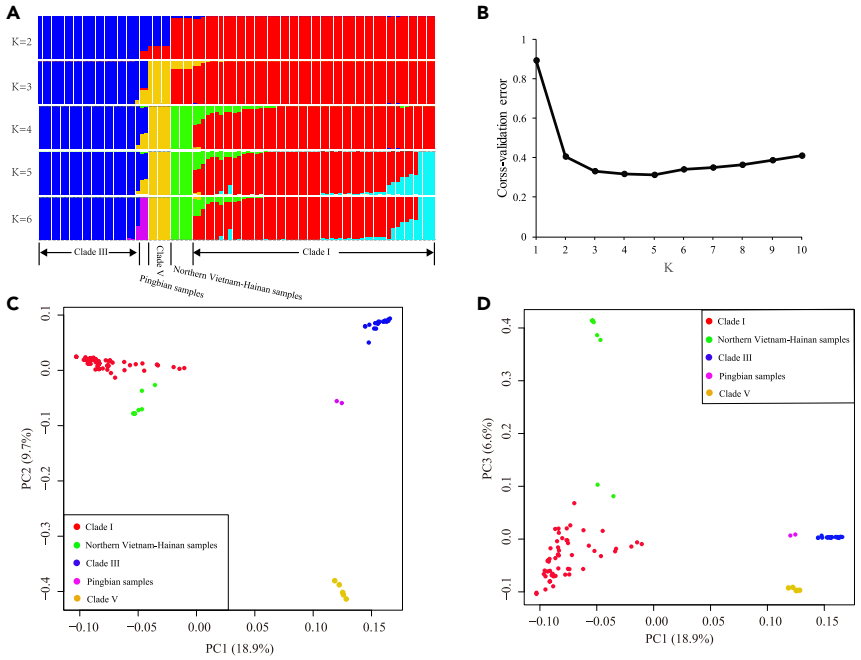


Figure 2. Genetic structure of big-headed turtles

(A) ADMIXTURE result based on genome-wide SNPs with *K* = 2–6.

(B) Cross-validation (CV) errors calculated by ADMIXTURE.

(C) PCA result based on PC1 and PC2.

(D) PCA result based on PC1 and PC3.

of clade III and clade V. PCA also revealed distinct clusters, largely corresponding to the ADMIXTURE result (Figure 2). The first three components (PC1, PC2, and PC3) explained 35.2% of the total variation. PC1 mainly separated clade I and Northern Vietnam-Hainan samples from clade III, Pingbian samples and clade V (Figure 2C). PC2 separated clade III, Pingbian samples and clade V into different clusters (Figure 2C). PC3 separated Northern Vietnam-Hainan samples from clade I (Figure 2D). However, in contrast to the ADMIXTURE result at K = 5 and 6, samples from eastern Guizhou and northern Hunan of China did not form obvious genetic clusters, indicating K = 5 and 6 were not optimal. While, four genetic clusters (K = 4) obtained consistent support from ADMIXTURE and PCA results. Moreover, almost no gene flow was detected among the four genetic clusters, except for a very weak gene flow from clade V to Northern Vietnam-Hainan samples (Figure 3A).

To further quantify genetic differentiation between different clusters, we calculated the average genetic distance indexes (F_{ST}) between the four genetic clusters (Figure 3B). The result showed significant differences ($p < 0.01$) between six pairwise comparisons (clade I vs. Northern Vietnam-Hainan samples, clade I vs. clade III+Pingbian samples, clade I vs. clade V, Northern Vietnam-Hainan samples vs. clade III+Pingbian samples, Northern Vietnam-Hainan samples vs. clade V, clade III+Pingbian samples vs. clade V). However, among these comparisons the last five pairwise F_{ST} values were no smaller than 0.80, which was more than twice as large as the first one (0.39), indicating a small genetic divergence between clade I and Northern Vietnam-Hainan samples. Similarly, based on mtDNA, we also detected very small genetic divergences (F_{ST} = 0.012 and 0.0139) between clade I vs. Northern Vietnam-Hainan samples and between clade III vs. Pingbian samples, respectively, compared to large genetic divergences (F_{ST} = 0.033–0.059) between the rest of the comparisons (Figure S2). Given the very small genetic divergences between clade I vs. Northern Vietnam-Hainan samples and between clade III vs. Pingbian samples, we finally combined the four genetic clusters into three large genetic groups, in line with three clusters at K = 3, and used this for subsequent demographic analyses.

Distinct evolutionary histories

Demographic analyses showed that the three genetic groups (*P. m. megacephalum*, *P. m. peguense*, and the Baise clade) have different demographic histories. Firstly, genome-wide LD analysis revealed that the

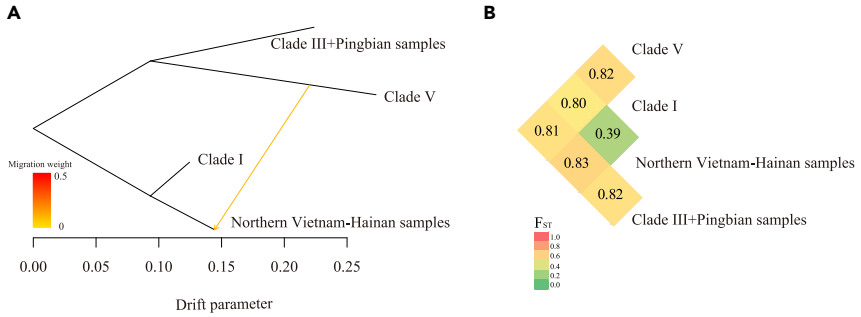


Figure 3. Genetic divergence among the four genetic clusters based on SNPs in the big-headed turtle
(A) Migration dynamics, (B) genetic differentiation.

Baise clade had the highest LD level and slowest LD decay, with a reduced R^2 correlation coefficient becoming stable at a distance of approximately 80 kb, followed by *P. m. peguense*, which showed a middle LD level and middle LD decay with a reduced R^2 becoming stable at a distance of approximately 60 kb, while *P. m. megacephalum* exhibited the lowest LD level and fastest LD decay with a reduced R^2 becoming stable at a distance of approximately 40 kb (Figure 4A). Secondly, PSMC analysis demonstrated that both *P. m. megacephalum* and *P. m. peguense* underwent two obvious expansions/bottlenecks but with a larger effective population size for the former, while the Baise clade underwent only one (Figure 4B). The difference between *P. m. megacephalum* and *P. m. peguense* was initiated at the end of the first population bottleneck, approximately 0.18 Ma. In contrast to the continued population decline of *P. m. peguense*, the effective population size of *P. m. megacephalum* subsequently recovered quickly prior to 30 Ka, after which *P. m. megacephalum* went through a second bottleneck. The difference between *P. m. megacephalum* and the Baise clade began at about 0.12 Ma, after which the effective population size of *P. m. megacephalum* gradually increased until 30 Ka, while the Baise clade continued to decrease mostly.

DISCUSSION

Three phylogenetic species in the big-headed turtle

Delimiting threatened species is a precondition for the formulation of an effective conservation policy.[1,2] Our results revealed three genetic groups with substantial genetic divergences and distinct demographic histories, suggesting three phylogenetic species (*P. m. megacephalum*, *P. m. peguense*, and Baise clade). The first two phylogenetic species corresponded to previously well-described subspecies[3,16] and the last was newly discovered in this study. In addition to the three phylogenetic species, we also detected two small clusters (Northern Vietnam-Hainan samples and Pingbian samples) within *P. m. megacephalum* and *P. m. peguense*, respectively. Although Northern Vietnam-Hainan samples and Pingbian samples showed obvious genetic divergences from clade I and clade III, the average genetic divergence indexes (F_{ST}) in both comparisons (clade I vs. Northern Vietnam-Hainan samples, clade III vs. Pingbian samples) were at least twice as small as that in other pairwise comparisons, indicating that the genetic divergences of both clades have not reached the level of phylogenetic species in the present study. Therefore, from a genetic perspective it is reasonable to consider the two small clusters (Northern Vietnam-Hainan samples and Pingbian samples) as *P. m. megacephalum* and *P. m. peguense*, respectively.

Interestingly, we found morphological variations between individuals collected from northern Vietnam: some individuals had obvious orange spots on their head, carapace, and plastron (Figure S3), which are the main diagnostic criteria for *P. m. shiui*,[16] while others had no such morphological characteristics (Figure S4). However, those individuals showed no genetic differentiation and displayed close phylogenetic relationships with the Hainan population, revealing that the genetic data failed to support the presence of spots as a distinguishing feature, and that populations geographically assigned to *P. m. shiui* are not obviously genetically distinguishable from those assigned to *P. m. megacephalum*. The observation indicated that these morphological characteristics (e.g., spots) cannot be used to distinguish subspecies of the big-headed turtle. We also found some wild individuals in Heyuan and Meizhou (Guangdong Province, China), Yichun (Jiangxi Province, China) with orange spots on the body surface[17] (Figure S5), but which genetically clustered with typical specimens of *P. m. megacephalum*, thus we failed to find support for a northern Vietnamese subspecies, *P. m. shiui*.

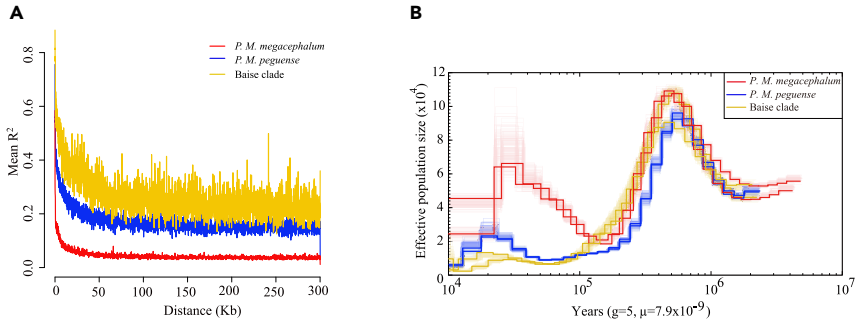


Figure 4. Demographic analyses of three phylogenetic species in the big-headed turtle

(A) Linkage disequilibrium plot of three phylogenetic species (*P. m. megacephalum*, *P. m. peguense*, and Baise clade). (B) PSMC analysis of three big-headed turtle phylogenetic species: the history of the turtle populations spans from 10 ka to Ma, calculations used an estimation of five years per generation with a mutation rate of 7.9×10^{-9}.

Our study further revealed the geographical pattern of the three phylogenetic species (*P. m. megacephalum*, *P. m. peguense*, and the Baise clade). The distribution range of *P. m. megacephalum* differs from previous reports, in which the Hainan population was based on only three specimens; two from the Field Museum of Natural History (FMNH6631, 6632) and one from the University of Michigan Museum of Zoology (UMMZ129833). These were generally considered as *P. m. shiui* or *P. m. peguense*.[3,15] Here, a total of 14 samples were collected from three nature reserves (Wuzhishan, Diaoluoshan, and Limushan) in Hainan of China. Our comprehensive analyses demonstrated that the Hainan population should be assigned to *P. m. megacephalum*. The discrepancy in sample source is likely to be the main reason for the inconsistency between the present and previous studies, because the Hainan samples were collected from wild in the present study and the museum samples may have been from trading market in the previous study.[15] In addition, our molecular analyses confirmed the northern Vietnam populations as *P. m. megacephalum*. Although the geographical pattern of *P. m. peguense* in this study was largely consistent with previous reports,[3,15] our study provided genomic evidence for the geographical pattern. The newly discovered phylogenetic species is only found in Baise, Guangxi in China.

The large amount of genetic variation in the intersection area of Yunnan, Guangxi in China, and Vietnam indicated that the big-headed turtle probably originated in and diversified from the region, which was recently identified as a long-term stable refugium for relict species.[20] The refugium has been largely characterized by mild temperatures, even during the coldest months since the last glacial maximum, thus providing suitable conditions for relict mountain species such as the big-headed turtle. Our phylogenetic analyses based on both mtDNA and SNPs suggest that *P. m. megacephalum* is the basal phylogenetic species, which is a sister phylogenetic species to both *P. m. peguense* and the Baise clade. However, the finding was not accordance with previous studies,[18,19] which suggested that *P. m. peguense* was a sister group to the other two subspecies (*P. m. megacephalum* and *P. m. shiui*). We further checked collection information for *P. m. shiui* and *P. m. peguense* samples, which were used to produce the complete mitochondrial genomes.[18,21] We found that one sample (MVZ 230486) from Hainan in China was preliminarily used to assemble the complete mitochondrial genome,[21] but was mistakenly considered as the complete mitochondrial genome of *P. m. shiui*.[18] The Hainan samples should be classified as *P. m. megacephalum* based on our comprehensive study, and the complete mitochondrial genome (DQ 256377) of MVZ 230486[21] should be re-affiliated to *P. m. megacephalum*.

Demographic history

During the Miocene (23 - 5 Ma), the Himalayan mountains and the Tibetan Plateau were fully uplifted,[22] which simultaneously promoted the formation of steep mountains, rivers, and basins. While steep mountains used to be considered the main barriers to genetic exchange among subspecies, given their weak migration ability,[18] the notion was not supported by our results, as no substantial genetic divergence was detected among *P. m. peguense* populations in Myanmar, Thailand, Laos, and Vietnam, where such extensive barriers exist. A similar pattern was also observed for *P. m. megacephalum* in southeastern China. Big-headed turtles are able to climb steep stones with the aid of their specific morphological features (e.g., long tail and flat body), thus steep mountains may not be major barriers for this species. Similar to high-elevation areas (e.g., mountains), low-land areas (e.g., rivers and basins) were also proposed as barriers

to gene flow for montane reptiles, such as *Plethodon fourchensis*[23] and *Protobothrops jerdonii*.[24] Alternatively, large river plains and basins may be the main barriers limiting genetic exchange among the three phylogenetic species, given specific elevation distributions of 430–1350 m.[25] Several Pearl River plains (e.g., Laibin plain, Youjiang, and Qianjiang River Valley plains) less than 200 m above sea level are located in the middle and south regions (Guangxi basin) of Guangxi in China, where the big-headed turtle has never been reported. Interestingly, the regions are coincident with the geographical boundary between *P. m. megacephalum* and the Baise clade. Similarly, the Red River plain in Vietnam is located at the geographical boundary between *P. m. megacephalum* and *P. m. peguense*. The lowlands (river plains and/or basins) seem to be the major barriers for *P. megacephalum* and may be the main reason for phylogenetic species divergence.

In general, Pingbian samples showed a certain genetic divergence from clade III (other *P. m. peguense* samples) based on genome-wide SNPs and mtDNA analyses. Interestingly, both phylogenetic and PCA analyses based on the former suggested a large genetic differentiation among Pingbian individuals. However, a close relationship was observed among Pingbian individuals using the latter analysis, as a single haplotype was detected among them. The higher resolution for the genome-wide SNPs may be one of the main reasons. The discrepancy in heredity patterns for the two markers may be another reason, given the property of mtDNA maternal inheritance. The results support two potential scenarios: (1) the Pingbian samples may represent the remnant of an ancient lineage, parts of which spread to Vietnam, Laos, Thailand, and Myanmar, finally forming *P. m. peguense* with the current distribution pattern. Some individuals spread to Baise, Guangxi of China, finally evolving into the Baise clade. Some individuals may be ancestors of *P. m. megacephalum*, which firstly dispersed to northern Vietnam, subsequently spread to Hainan in China during the periods when Hainan Island was not separated from northern Vietnam and Guangxi of China, and finally spread to the Chinese mainland. This scenario supports the hypothesis of the intersection area of Yunnan, Guangxi, and Vietnam as the origin of the big-headed turtle (Figure 1), (2) the Pingbian samples may come from novel hybrid individuals between *P. m. peguense* and the Baise clade. They show a consistent maternal origin, probably from Vietnamese or Thai populations, as evidenced by the limited genetic divergence among Pingbian individuals based on mtDNA. However, subsequently introduced individuals of other subspecies, such as the Baise clade, have led to the current genetic pattern within the population. This notion was supported by the observation of a mixture pattern between *P. m. peguense* and the Baise clade at K = 4–6 with low cross-validation errors using ADIMIXTURE analysis.

Suggestions for taxonomic revision and conservation

Currently, two of three phylogenetic species are regarded as subspecies of the big-headed turtle (*P. megacephalum*). However, we found substantial genetic divergence among three phylogenetic species. For example, the genetic differentiation index (F_{ST}) between *P. m. megacephalum* and *P. m. peguense* was highest, with average values of 0.83 based on SNPs and 0.056 based on mtDNA, while the F_{ST} between *P. m. peguense* and the Baise clade was lowest, with an average value of 0.81 based on SNPs and 0.034 based on mtDNA. In addition, based on the same mtDNA (COI and 12S), the F_{ST} (0.034) between *P. m. peguense* and the Baise clade was obviously larger than that (0.023) between two phylogenetically closed turtle species (*Mauremys reevesii* and *Mauremys sinensis*). Given the significant genetic differentiation among the three phylogenetic species of the big-head turtle, we suggest upgrading subspecies *P. m. megacephalum* and *P. m. peguense* to species *P. megacephalum* and *P. peguense* respectively. For the newly discovered phylogenetic species, the Baise clade, it can be taken as an unnamed cryptic species temporarily, and further research is needed to describe and name it.

Wild populations of big-headed turtles have been seriously damaged in many regions throughout their natural ranges, mainly due to human activities such as habitat destruction, trade, and poaching.[7,9] Currently, all wild populations of the big-headed turtle are given the same level of protection in China (grade II national key-protected species), because all are currently regarded as conspecific. However, in fact, we found that the big-headed turtle contains at least three phylogenetic species, representing three evolutionarily significant units. Of these, the newly discovered phylogenetic species, the Baise clade, with the lowest effective population size and a high inbreeding level, is vulnerable to human activities and other threats such as global climate change. Therefore, the Baise clade needs more protection. Given the multiple genetic units, genetic background needs to be considered in conservation activities, such as reintroduction and field release of captive-bred populations or individuals confiscated from illegal markets.

Limitations of the study

The critically endangered big-headed turtle is very rare in the field; we spent about fifteen years to collect range-wide samples in the present study. However, specimen sizes in some locations are still small, such as Baise region, Guangxi of China, where only few individuals were collected and not enough to perform a comprehensive morphological analysis. Hence, without ample morphological evidence, we cannot assign Baise clade as a new species, simply one phylogenetic species in the present study. In addition, some samples were preserved too long time (e.g., at least ten years for PM08 and PM09 collected from Northern Fujian of China), leading to low genome DNA quality. Although the DNA samples with low quality can be used for mtDNA analysis, these samples cannot be used for RAD sequencing-based SNP analysis, hence only 72.59% (98/135) samples were used to SNP-based genome-wide analysis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - The big headed turtle
- METHODS DETAILS
  - DNA extraction and sequencing
  - Data processing and phylogenetic analyses
  - Population genetic divergence
  - Demographic history
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107343.

ACKNOWLEDGMENTS

We thank Markus Auer, Daniel Gaillard, Jian Wang, Song Huang, Fang Zhang, Xiangjing Zhong, Congrong Zheng, Dingqi Rao, Yunming Mo, Liu Lin, Haijing Wang, Dongsheng Pan, Qingsong Xie, Yuanhua Qin, Guocai Zou, Jonathan J. Fong, Cuong T. Pham, Anh V. Pham, Sengvilay Lorphengsy, Kruewan Pipatsawasdikul, Parinya Pawangkhanant, Dzung T. Le and Mali Naiduangchan for their help in the sampling process and in providing some tissue/DNA samples. We also thank Jiaxuan Li for reviewing the manuscript. This study was funded by the National Natural Science Foundation of China (31471966; 31071935; 32270542), the Thousand PhD Program of Guangdong Academy of Sciences (2020GDASYL-20200103099), the Guangdong Basic and Applied Basic Research Foundation (2019A1515110455), the Scientific and Technological Program of Guangzhou (202103000082), GDAS Special Project of Science and Technology Development (2022GDASZH-2022010106), Turtle Conservation Fund, the Biodiversity Conservation Fund of Kadoorie Farm, the Unit of Excellence 2023 on Biodiversity and Natural Resources Management (FF66-UoE003), the Thailand Research Fund 2019 (MRG6280203), the National Science and Technology Development Agency (P-19-50812), Science Fund of the Nanling Mountain National Nature Reserve: for Amphibians and Reptiles, and the Botanic Garden, Ocean Park Conservation Foundation, Hong Kong.

AUTHOR CONTRIBUTIONS

Conceptualization: S.G. and Y.C.G.; Methodology: Y.C.G. and Y.W.; Writing-Original Draft: Y.C.G. and S.G.; Writing-Review & Editing: Y.C.G., S.G., Y.W., Y.G., C.S., M.L., and T.Q.N.; Sample collection: S.G., Y.G., Y.W., C.S., M.L., and T.Q.N. All authors reviewed the whole work and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 9, 2023
Revised: May 21, 2023
Accepted: July 6, 2023
Published: July 10, 2023

REFERENCES

1. Fujisawa, T., and Barraclough, T.G. (2013). Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. Syst. Biol. *62*, 707–724. https://doi.org/10.1093/sysbio/syt033.

2. Perry, A.L., Low, P.J., Ellis, J.R., and Reynolds, J.D. (2005). Climate change and distribution shifts in marine fishes. Science *308*, 1912–1915. https://doi.org/10.1126/science.1111322.

3. Gong, S., Wu, J., Gao, Y., Fong, J.J., Parham, J.F., and Shi, H. (2020). Integrating and updating wildlife conservation in China. Curr. Biol. *30*, R915–R919. https://doi.org/10.1016/j.cub.2020.06.080.

4. Gutiérrez, E.E., and Helgen, K.M. (2013). Mammalogy: Outdated taxonomy blocks conservation. Nature *495*, 314. https://doi.org/10.1038/495314e.

5. Stanford, C.B., Iverson, J.B., Rhodin, A.G.J., van Dijk, P.P., Mittermeier, R.A., Kuchling, G., Berry, K.H., Bertolero, A., Bjorndal, K.A., Blanck, T.E.G., et al. (2020). Turtles and tortoises are in trouble. Curr. Biol. *30*, R721–R735. https://doi.org/10.1016/j.cub.2020.04.088.

6. Turtle Taxonomy Working Group, Rhodin, A.G.K., Iverson, J.B., Bour, R., Fritz, U., Georges, A., and Shaffer, H.B. (2021). Turtles of the World Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status. In Chelonian Research Monographs, 9th Ed., *8*. Chelonian Research Monographs, pp. 1–472. https://doi.org/10.3854/crm.8.checklist.atlas. v9.2021.

7. Gong, S.P., Shi, H.T., Jiang, A.W., Fong, J.J., Gaillard, D., and Wang, J.C. (2017). Disappearance of endangered turtles within China's nature reserves. Curr. Biol. *27*, R170–R171. https://doi.org/10.1016/j.cub.2017.01.039.

8. Sung, Y.H., Karraker, N.E., and Hau, B.C.H. (2013). Demographic evidence of illegal harvesting of an endangered Asian turtle. Conserv. Biol. *27*, 1421–1428. https://doi.org/10.1111/cobi.12102.

9. Wu, J., Wu, Y., Rao, D., Zhou, T., and Gong, S. (2020). China's wild turtles at risk of extinction. Science *368*, 838. https://doi.org/10.1126/science.abc0997.

10. Fong, J., Hoang, H., Li, P., McCormack, T., Rao, D., Timmins, R.J., Wang, L., and van Dijk, P.P. (2021). Platysternon megacephalum. The IUCN Red List of Threatened Species 2021, e.T17585.

11. Gong, S., Hua, L., Ge, Y., and Cao, D. (2019). Unique multiple paternity in the endangered big-headed turtle (*Platysternon megacephalum*) in an ex situ population in South China. Ecol. Evol. *9*, 9869–9877. https://doi.org/10.1002/ece3.5528.

12. Turvey, S.T., Chen, S., Tapley, B., Wei, G., Xie, F., Yan, F., Yang, J., Liang, Z., Tian, H., Wu, M., et al. (2018). Imminent extinction in the wild of the world's largest amphibian. Curr. Biol. *28*, R592–R594. https://doi.org/10.1016/j.cub.2018.04.005.

13. Le, M.D., McCormack, T.E., Van Hoang, H., Duong, H.T., Nguyen, T.Q., Ziegler, T., Nguyen, H.D., and Ngo, H.T. (2020). Threats from wildlife trade: The importance of genetic data in safeguarding the endangered Four-eyed Turtle (*Sacalia quadriocellata*). Nat. Conserv. *41*, 91–111. https://doi.org/10.3897/natureconservation.41.54661.

14. Yan, F., Lü, J., Zhang, B., Yuan, Z., Zhao, H., Huang, S., Wei, G., Mi, X., Zou, D., Xu, W., et al. (2018). The Chinese giant salamander exemplifies the hidden extinction of cryptic species. Curr. Biol. *28*, R590–R592. https://doi.org/10.1016/j.cub.2018.04.004.

15. Ernst, C.H., and Laemmerzahl, A.F. (2002). Geographic variation in the Asian big-headed turtle, *Platysternon megacephalum* (Reptilia: Testudines: Platysternidae). Proc. Biol. Soc. Wash. *115*, 18–24.

16. Ernst, C.H., and McCord, W.P. (1987). Two new turtles from Southeast Asia. Proc. Biol. Soc. Wash. *100*, 624–628.

17. Gong, S. (2016). Marker-based Estimation of Genetic Divergence between Orange-Spotted and Non-spotted Big-Headed Turtle (*Platysternon megacephalum*) from Guangdong, China (The 8th World Congress of Herpetology), p. 596.

18. Luo, H., Li, H., Huang, A., Ni, Q., Yao, Y., Xu, H., Zeng, B., Li, Y., Wei, Z., Yu, G., and Zhang, M. (2019). The complete mitochondrial genome of *Platysternon megacephalum peguense* and molecular phylogenetic analysis. Genes *10*, 487. https://doi.org/10.3390/genes10070487.

19. Zheng, C., Nie, L., Wang, J., Zhou, H., Hou, H., Wang, H., and Liu, J. (2013). Recombination and evolution of duplicate control regions in the mitochondrial genome of the Asian big-headed turtle, *Platysternon megacephalum*. PLoS One *8*, e82854. https://doi.org/10.1371/journal.pone.0082854.

20. Tang, C.Q., Matsui, T., Ohashi, H., Dong, Y.F., Momohara, A., Herrando-Moraira, S., Qian, S., Yang, Y., Ohsawa, M., Luu, H.T., et al. (2018). Identifying long-term stable refugia for relict plant species in East Asia. Nat. Commun. *9*, 4488. https://doi.org/10.1038/s41467-018-06837-3.

21. Parham, J.F., Feldman, C.R., and Boore, J.L. (2006). The complete mitochondrial genome of the enigmatic bigheaded turtle (*Platysternon*): description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. BMC Evol. Biol. *6*, 11. https://doi.org/10.1186/1471-2148-6-11.

22. An, Z., Kutzbach, J.E., Prell, W.L., and Porter, S.C. (2001). Evolution of Asian monsoons and phased uplift of the Himalayan Tibetan plateau since Late Miocene times. Nature *411*, 62–66. https://doi.org/10.1038/35075035.

23. Shepard, D.B., and Burbrink, F.T. (2009). Phylogeographic and demographic effects of Pleistocene climatic fluctuations in a montane salamander. Mol. Ecol. *18*, 2243–2262. https://doi.org/10.1111/j.1365-294X.2009.04164.x.

24. Guo, P., Liu, Q., Li, C., Chen, X., Jiang, K., Wang, Y.Z., and Malhotra, A. (2011). Molecular phylogeography of Jerdon's pitviper (*Protobothrops jerdonii*): importance of the uplift of the Tibetan plateau. J. Biogeogr. *38*, 2326–2336. https://doi.org/10.1111/j.1365-2699.2011.02566.x.

25. Pipatsawasdikul, K., Voris, H.K., and Thirakhupt, K. (2010). Distribution of the big-headed turtle (*Platysternon megacephalum*, Gray 1831) in Thailand. Zool. Stud. *49*, 640–650.

26. Palumbi, S.R. (1996). Nucleic acids II: The polymerase chain reaction. In Molecular Systematics, 2nd edition, D.M. Hillis, C. Moritz, and B.K. Mable, eds. (Sinauer), pp. 205–248.

27. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11 Molecular evolutionary genetics analysis Version 11. Mol. Biol. Evol. *38*, 3022–3027. https://doi.org/10.1093/molbev/msab120.

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

29. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A software platform for bayesian evolutionary analysis. PLoS Comput. Biol. *10*, e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

30. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods *9*, 772. https://doi.org/10.1038/nmeth.2109.

31. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

32. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

33. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

34. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

35. Rochette, N.C., Rivera-Colón, A.G., and Catchen, J.M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. *28*, 4737–4754. https://doi.org/10.1111/mec.15253.

36. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

37. Foll, M., and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics *180*, 977–993. https://doi.org/10.1534/genetics.108.092221.

38. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.

39. Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics *28*, 3326–3328. https://doi.org/10.1093/bioinformatics/bts606.

40. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967.

41. Excoffier, L., and Lischer, H.E.L. (2010). Arlequin Suite ver 3.5, a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x.

42. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics *35*, 1786–1788. https://doi.org/10.1093/bioinformatics/bty875.

43. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496. https://doi.org/10.1038/nature10231.

44. Cao, D., Wang, M., Ge, Y., and Gong, S. (2019). Draft genome of the big-headed turtle *Platysternon megacephalum*. Sci. Data *6*, 60. https://doi.org/10.1038/s41597-019-0067-9.

45. Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One *3*, e3376. https://doi.org/10.1371/journal.pone.0003376.

46. Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., Jiang, C., Guan, N., Ma, C., Zeng, H., et al. (2013). SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PLoS One *8*, e58700. https://doi.org/10.1371/journal.pone.0058700.

47. Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A resource for timelines, timetrees, and divergence Times. Mol. Biol. Evol. *34*, 1812–1819. https://doi.org/10.1093/molbev/msx116.

48. Li, Y., Sun, X., Hu, X., Xun, X., Zhang, J., Guo, X., Jiao, W., Zhang, L., Liu, W., Wang, J., et al. (2017). Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. Nat. Commun. *8*, 1721. https://doi.org/10.1038/s41467-017-01927-0.

49. Zheng, G., Schumann, S., Zhu, X., Liu, Y., Chen, Y., and Luo, J. (2008). Age and growth of *Platysternon megacephalum* (Gay) and difference in growth and morphology between males and females. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. *2008*, *66–69*. (In Chinese with English abstract). 1001-411X(2008)03-0066-05.

50. Green, R.E., Braun, E.L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G., Vandewege, M.W., St, John, J.A., Capella-Gutiérrez, S., Castoe, T.A., et al. (2014). Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science *346*, 1254449. https://doi.org/10.1126/science.1254449.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
The big-head turtle (*Platysternon megacephalum*)	Field collection from China, Vietnam, Laos, Thailand and Myanmar	see Table S2 and Figure 1 for population code, location, sample name and number
The red-eared slider turtle (*Trachemys scripta elegans*)	https://www.ncbi.nlm.nih.gov/	Accession: KM216748.1
Chemicals, peptides, and recombinant proteins		
Ethanol absolute	Guangdong Guanghua Sci-Tech Co., Ltd.	Cat# WHP-1.17113.023
Premix Taq™	TaKaRa	Cat# RR901A
Agarose G-10	Biowest	Cat# BY-R0100
RsaI	NEB	Cat# R0167L
Critical commercial assays		
DNeasy Blood & Tissue Kits	QIAGEN	Cat# 69506
Mastercycler nexus	Eppendorf	Cat# GSX1
Sanger sequencing	PerkinElmer	Cat# ABI3730
High-throughput sequencing	Illumina	Cat# Novaseq 6000
Deposited data		
The sequence of both 12S and COI	This study	Supplementary file
RAD data	NCBI SRA	BioProject number: PRJNA854205 and accession numbers: SRR19903903 - SRR19904000.
Oligonucleotides		
12SA	Palumbi[26]	AAACTGGGATTAGATACCCCACTAT
12SB	Palumbi[26]	GAGGGTGACGGGCGGTGTGT
PmCOIF	This study	TCTACTAACCATAAAGACATTG
PmCOIR	This study	GAATGTATACTTCTGGGTGRC
Software and algorithms		
MEGA 11	Tamura et al.[27]	https://megasoftware.net/dload_win_beta
DnaSP v 5.10	Librado and Rozas[28]	http://www.ub.edu/dnasp/
BEAST v 2.6.3	Bouckaert et al.[29]	https://www.beast2.org/
jmodeltest v 2.1.10	Darriba et al.[30]	https://www.softpedia.com/get/Science-CAD/jModelTest.shtml#download
fastp v 0.23.2	Chen et al.[31]	https://packages.guix.gnu.org/packages/fastp/0.23.2/
Bowtie 2.5.0	Langmead and Salzberg[32]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
SAMtools v 1.16.1	Li et al.[33]	https://sourceforge.net/projects/samtools/files/samtools/1.16/
Sambamba v 0.8.2	Tarasov et al.[34]	https://github.com/biod/sambamba
Stacks v 2.6.2	Rochette et al.[35]	http://catchenlab.life.illinois.edu/stacks/
IQ-TREE v 2.1.1	Minh et al.[36]	https://github.com/iqtree/iqtree2/releases
BayeScan v 2.1	Foll and Gaggiotti[37]	http://cmpg.unibe.ch/software/BayeScan/download.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ADMIXTURE v 1.3.0	Alexander et al.[38]	http://dalexander.github.io/admixture/download.html
SNPRelate	Zheng et al.[39]	https://www.bioconductor.org/packages/release/bioc/html/SNPRelate.html
TreeMix v 1.13	Pickrell and Pritchard[10]	https://software.cqls.oregonstate.edu/updates/treemix-1.13/
Arlequin v 3.5.2.2	Excoffier and Lischer[41]	http://cmpg.unibe.ch/software/arlequin35/
PopLDdecay v 3.31	Zhang et al.[42]	https://github.com/BGI-shenzhen/PopLDdecay/releases
PSMC	Li and Durbin[43]	https://github.com/ChenHuaLab/Beta-PSMC
FigTree v 1.4.3	http://tree.bio.ed.ac.uk/software/Figtree/	http://tree.bio.ed.ac.uk/software/figtree/

RESOURCE AVAILABILITY

Lead contact

Further information should be directed to and will be fulfilled by the Lead Contact, Yangchun Gao (gaoyc0412@163.com).

Materials availability

All materials reported in this paper will be shared by Shiping Gong (gsp621@163.com) upon request.

Data and code availability

- All data reported in this paper are publicly accessible through GenBank: BioProject number: PRJNA854205 and accession numbers: SRR19903903 - SRR19904000. Concatenated mtDNA sequences (Data S1) and genome-wide SNPs (Data S2) are available in the supplementary files.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The big headed turtle

A total of 143 samples of the big-head turtle were collected from 67 localities covering almost the whole distribution range (China, Vietnam, Laos, Thailand and Myanmar) duirng 2006 to 2021 (Figure 1A, Table S2). Most tissue samples were tail tips or buccal swabs collected from live individuals, and the rest of the samples were muscle or liver tissue collected from dead individuals. All tissue samples were preserved in absolute alcohol at -20°C prior to DNA extraction. All experimental protocols were approved by the Ethics Committee of the Institute of Zoology, Guangdong Academy of Sciences (Certificate number: GIABR20200414) and the Institutional Ethical Committee of Animal Experimentation of the University of Phayao, Phayao, Thailand (Certificate number: UP-AE61-01-04-022). Permits for the export of DNA samples from Vietnam were issued by the CITES Management Authority of Vietnam (permit numbers: 16VNOO16 and 18VN1970N).

METHODS DETAILS

DNA extraction and sequencing

The genomic DNA from all tissue samples was extracted using DNeasy Blood & Tissue Kits (Qiagen Ltd., West Sussex, United Kingdom) according to the manufacturer's instructions. The concentration and quality of the genomic DNA were determined using a Nanodrop One spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). Integrity was assessed by inspecting the bands using 1% agarose gel electrophoresis. We found that the genomic DNA for some individuals was heavily degenerated, perhaps due to storage for too long a time, and thus could not be used for high throughput sequencing (e.g. reduced-representation genome sequencing). To overcome the problem, we preliminarily genotyped

almost all the samples (135 samples, 94.40 %), especially those with poor DNA quality and integrity, using two mitochondrial DNA (mtDNA) fragments of genes encoding 12S rRNA (12S) and the cytochrome c oxidase subunit I (COI). Subsequently, samples with high DNA quality and integrity were analyzed using SNPs produced by high throughput sequencing. Analyses based on both mtDNA and SNPs can also provide different insights into evolutionary events, thus also promoting the comprehensive understanding of evolutionary processes in the big-headed turtle.

Mitochondrial 12S fragments were amplified using universal primers, 12SA: AAACTGGGATTAGA TACCCCACTAT and 12SB: GAGGGTGACGGGCGGTGTGT,[26] COI fragments were amplified using a pair of primers, PmCOIF: TCTACTAACCATAAAGACATTG and PmCOIR: GAATGTATACTTCTGGGT GRC designed during this study. The PCR cocktail (25 μl) contained 12.5 μl Premix Taqtrade; (Cat: RR901A, Takara, Dalian, China), 10 pmol forward and reverse primers and 50 ng of genomic DNA. We performed PCRs on a Mastercycler nexus (Eppendorf, Hamburg, Germany) with the following cycle conditions: 95°C for 5 minutes; then 30 cycles at 95°C for 10 seconds, 57°C for 30 seconds and 72°C for 30 seconds; and the final extension at 72°C for 7 minutes. The PCR products were assessed using 2% agarose gel electrophoresis and a gel imaging system, and target PCR products were sequenced on an ABI3730 (PerkinElmer Biosystems, Foster City, CA, USA).

A total of 98 samples showed high DNA quality and integrity, and were used for high throughput sequencing. If we choose genome resequencing to call SNPs, both the massive sample and the large genome size (2.4 Gb) in the big-headed turtle[44] would result in very high costs. To economically obtain genome-wide markers, we selected reduced-representation sequencing (restriction-site associated DNA sequencing, RAD-seq) to call SNPs (Baird et al., 2008).[45] Library construction of RAD-seq was performed using RsaI restriction enzyme digestions, following the protocol described by Sun et al. (2013).[46] The libraries were sequenced with a pair end of 250 bp strategy using the Illumina Novaseq 6000 platform.

Data processing and phylogenetic analyses

To reveal phylogenetic relationships, we performed two separate phylogenetic analyses using Bayesian inference (BI) and maximum likelihood (ML) for mtDNA markers and genome-wide SNPs, respectively. For the former, the raw sequences of both 12S and COI were initially proofread, aligned and trimmed using MEGA 11.[27] The sequences of 12S and COI for each individual were concatenated with a final length of 1006 bp (Supplementary file). The concatenated mtDNA sequences were used to generate haplotypes with DnaSP v 5.10.[28] The BI analyses based on the haplotypes were performed using BEAST v 2.6.3.[29] The red-eared slider turtle (*Trachemys scripta elegans*) was selected as the outgroup, since this species is phylogenetically close to *P. megacephalum*.[18] The best substitution model of GTR + GAMMA was determined based on AIC using jmodeltest v 2.1.10.[30] The coalescent constant model was used as the tree prior with a uniform distribution. Moreover, the divergence time between phylogenetic clades was estimated using BEAST v 2.6.3.[29] The Yule model was used as the tree prior with a uniform distribution. A time constraint was used to calibrate the molecular clock: the most recent common ancestor (MRCA) between *T. scripta elegans* and *P. megacephalum* was estimated to be 70.6-90.3 Ma.[47] The BEAST program was run with 300 million iterations and the phylogenetic trees were sampled every 10,000 iterations, which retained 30,000 trees; the first 25% was discarded as burn-in. Tracer v 1.7.1 was used to assess the running results, which were subsequently annotated using TreeAnnotator v 1.10.4 within BEAST program.

For the ML analyses based on the genome-wide SNPs, raw Illumina sequence reads were first quality-filtered to remove adapters, contamination, and low-quality reads using fastp v 0.23.2[31] with the following major settings: -q 15, -u 30, -x 10, -n 0, –length_required 126. The obtained clean reads were aligned to the reference genome[44] of *P. megacephalum* using Bowtie 2.5.0 with default parameters.[32] We individually converted the resultant .sam files to .bam files and sorted using SAMtools v 1.16.1.[33] Subsequently, we removed PCR duplicates and reads mapped to multiple positions of the reference genome using Sambamba v 0.8.2.[34] The unique alignments were then processed for SNP calling with VCF output format by the ref_map.pl implemented in Stacks v 2.6.2.[35] To obtain reliable SNPs, we performed a further strict filtering step using a populations program implemented in Stacks v 2.6.2[35] with the following parameters: -p 3, -r 0.70, -R 0.70, –min-maf 0.05, –max-obs-het 0.5, –write-random-snp. Specimens representing the mitochondrial populations and major clades were used for phylogenetic analysis based on RAD-seq data. Based on those reliable SNPs, we constructed a maximum likelihood phylogenetic tree using

IQ-TREE v 2.1.1[36] with the following parameters: -m GTR+ASC, -bb 10000, -alrt 10000, -nt 12. Phylogenetic trees with node labels were visualized using FigTree v 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Population genetic divergence

To construct neutral SNP-based genetic structures, we firstly removed potential adaptive loci from all SNPs. Potential adaptive loci were detected using BayeScan v 2.1[37] with a burn-in of 50,000 steps with prior odds of 100, followed by 20 pilot runs with a length of 5,000 iterations. Only SNPs with FDR lower than 5% were considered as potential adaptive SNPs. Two methods based on the neutral SNPs were adopted to characterize the relatedness among populations/samples. Population genetic structure was firstly inferred using ADMIXTURE v 1.3.0[38] with default parameters. The number of genetic clusters was predefined from one to ten, and the lowest cross-validation value of the number assumptions was used to define the optimal K value. Secondly, sample clusters were also visualized with principal component analysis (PCA) using the R package SNPRelate[39] with default settings, which could also facilitate the understanding of population genetic structure. Migration events were inferenced by detecting gene flow among the optimal genetic clusters using TreeMix v 1.13.[40] The genetic differentiation index (F_{ST}) between the optimal genetic cluster pairs was calculated using 1000 permutations in Arlequin v 3.5.2.2.[41]

Demographic history

To assess the linkage disequilibrium (LD) pattern in phylogenetic species, genome-wide LD analysis was performed with calculation of the correlation coefficient (R^2) between any two loci using PopLDdecay v 3.31.[42] To further understand the demographic history, six samples representing three phylogenetic species (two samples for each phylogenetic species) were selected from individual turtles for whole-genome re-sequencing. The protocols for library construction and re-sequencing were according to a previous description.[48] The re-sequencing data was subsequently used for PSMC analysis using the PSMC program.[43] We used an estimation of 5 years per generation (g) for all three phylogenetic species,[49] and the nucleotide mutation rate (μ) was 7.9×10^{-9}, which was estimated using the genomes of other reptiles.[50] The value of μ was calculated using the following formula: $\mu = L/(T \times 10^{-6}) \times g$, where L is the branch length of the big-headed turtle in the time tree, T is the estimated divergence time, and g is the estimated 5 years per generation. One hundred bootstrap replicates per phylogenetic species were used to estimate confidence intervals.

QUANTIFICATION AND STATISTICAL ANALYSIS

To quantify genetic differentiation between different clusters, we calculated the genetic differentiation index (F_{ST}) with statistical significance (*p* value) between the optimal genetic cluster pairs using 1000 permutations in Arlequin v 3.5.2.2.[41]