BMC Biology

RESEARCH ARTICLE

Open Access

CrossMark

The *Sinocyclocheilus* cavefish genome provides insights into cave adaptation

Junxing Yang[1*†], Xiaoli Chen[2†], Jie Bai[2,3,4†], Dongming Fang[2,6†], Ying Qiu[2,3,5†], Wansheng Jiang[1†], Hui Yuan[2], Chao Bian[2,3], Jiang Lu[2,7], Shiyang He[2,7], Xiaofu Pan[1], Yaolei Zhang[2,8], Xiaoai Wang[1], Xinxin You[2,3], Yongsi Wang[2], Ying Sun[2,5], Danqing Mao[2], Yong Liu[2], Guangyi Fan[2], He Zhang[2], Xiaoyong Chen[1], Xinhui Zhang[2,3], Lanping Zheng[1], Jintu Wang[2], Le Cheng[5,9], Jieming Chen[2,3], Zhiqiang Ruan[2,3], Jia Li[2,3,7], Hui Yu[2,3,7], Chao Peng[2,3], Xingyu Ma[10,11], Junmin Xu[10,11], You He[12], Zhengfeng Xu[13], Pao Xu[14], Jian Wang[2,15], Huanming Yang[2,15], Jun Wang[2,16], Tony Whitten[4*], Xun Xu[2*] and Qiong Shi[2,3,10,11*]

Abstract

Background: An emerging cavefish model, the cyprinid genus *Sinocyclocheilus*, is endemic to the massive southwestern karst area adjacent to the Qinghai-Tibetan Plateau of China. In order to understand whether orogeny influenced the evolution of these species, and how genomes change under isolation, especially in subterranean habitats, we performed whole-genome sequencing and comparative analyses of three species in this genus, *S. grahami*, *S. rhinocerous* and *S. anshuiensis*. These species are surface-dwelling, semi-cave-dwelling and cave-restricted, respectively.

Results: The assembled genome sizes of *S. grahami*, *S. rhinocerous* and *S. anshuiensis* are 1.75 Gb, 1.73 Gb and 1.68 Gb, respectively. Divergence time and population history analyses of these species reveal that their speciation and population dynamics are correlated with the different stages of uplifting of the Qinghai-Tibetan Plateau. We carried out comparative analyses of these genomes and found that many genetic changes, such as gene loss (e.g. opsin genes), pseudogenes (e.g. crystallin genes), mutations (e.g. melanogenesis-related genes), deletions (e.g. scale-related genes) and down-regulation (e.g. circadian rhythm pathway genes), are possibly associated with the regressive features (such as eye degeneration, albinism, rudimentary scales and lack of circadian rhythms), and that some gene expansion (e.g. taste-related transcription factor gene) may point to the constructive features (such as enhanced taste buds) which evolved in these cave fishes.

Conclusion: As the first report on cavefish genomes among distinct species in *Sinocyclocheilus*, our work provides not only insights into genetic mechanisms of cave adaptation, but also represents a fundamental resource for a better understanding of cavefish biology.

Keywords: Cavefish, Genome, Adaptation, Evolution, Qinghai-Tibetan Plateau, *Sinocyclocheilus*

Background

As one of the most successful vertebrate colonizers in subterranean habitats, cavefishes attract interest because of their unusual regressive features, such as the rudimentary eyes and scales, and loss of pigmentation. As possible compensation, some constructive traits have evolved, such as

elongated appendages and non-visual sensory systems [1, 2]. Nevertheless, biologists have long puzzled over these troglomorphic characters [3], and the study of their relationship to their environment has been largely ignored [4]. Although recent work on a traditional model, *Astyanax mexicanus*, revealed some important advances [3], especially in aspects of eye loss [5] and identifying candidate genes underlying quantitative trait loci (QTL) [6], there is no whole genomic data available to unravel the evolution and general adaptation to dark subterranean life among other groups of distinct but closely-related cavefishes.

Sinocyclocheilus (Cypriniformes: Cyprinidae) is endemic to China's massive southwestern karst area, the northern

* Correspondence: yangjx@mail.kiz.ac.cn; tony.whitten@fauna-flora.org;
xuxun@genomics.cn; shiqiong@genomics.cn
†Equal contributors
[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[4]Fauna & Flora International, Cambridge CB1 2JD, UK
[2]BGI-Shenzhen, Shenzhen 518083, China
Full list of author information is available at the end of the article

Yang *et al. BMC Biology* (2016) 14:1
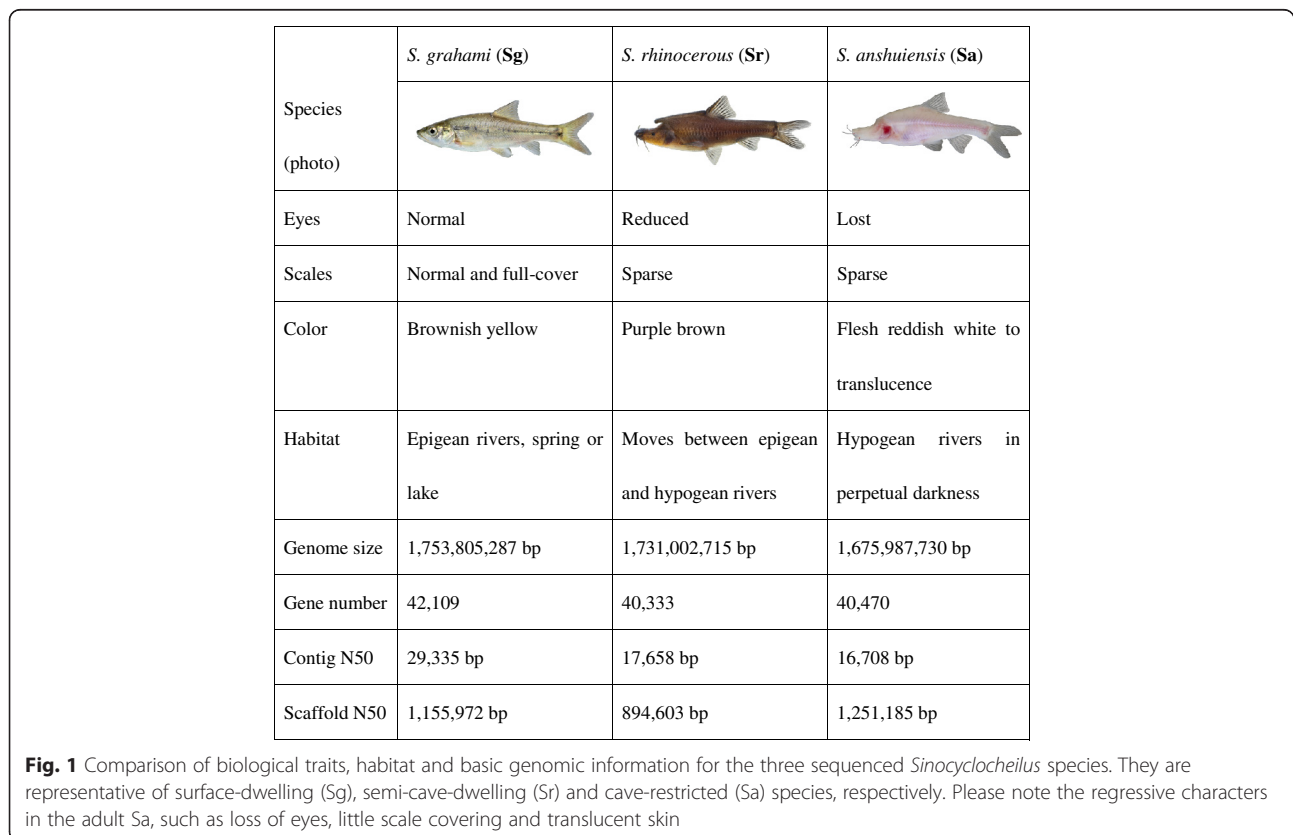
Page 2 of 13

part of which abuts the eastern part of the Qinghai-Tibetan Plateau. This genus serves as an emerging cavefish model for its high species diversity and the phenotypic variation which has evolved in karst river systems. In a previous *Science* letter in 2013 [7], we reported a global hotspot of biodiversity and the threats to it through a survey of cave species in Southwestern China, with an illustration of a *Sinocyclocheilus* cavefish, *S. rhinocerous*, which was collected from Yunnan province. The present study deals with whole genome and transcriptome sequencing of three *Sinocyclocheilus* species (Fig. 1): the surface-dwelling *S. grahami* (Sg); the semi-cave-dwelling *S. rhinocerous* (Sr); and the cave-restricted *S. anshuiensis* (Sa). These species were carefully chosen as representatives of three key nodes on the path to obligate cave life, although almost all *Sinocyclocheilus* species show some cave-related traits and habits [8]. We investigate whether whole genomic data would provide some clues about genetic adaptations to subterranean habitats, and whether the strict co-occurrence of the karstic landscape and *Sinocyclocheilus* species might serve to reveal aspects of the biogeographic history of the uplift of the Qinghai-Tibetan Plateau.

Results and discussion

Genome assembly, assessment and annotation

High-quality genomic DNA was extracted from muscle tissues of the three *Sinocyclocheilus* species, which were collected from Yunnan (Sg and Sr) and Guangxi (Sa) provinces in China (Additional file 1: Figure S1). A series of sequencing libraries (250 bp to 20 kb) were constructed and applied in a whole-genome shotgun sequencing strategy, and a total of 313.3, 174.0 and 188.2 Gb of raw data were obtained for the Sg, Sr and Sa fishes, respectively (Additional file 2: Table S1). The assembled genome sizes are approximately 1.7 Gb (1.75 Gb for Sg, 1.73 Gb for Sr and 1.68 Gb for Sa), and the calculated contig N50 and scaffold N50 values are 17–29 Kb and 0.9–1.3 Mb, respectively (Fig. 1; Additional file 2: Table S3). The quality of the three genome assemblies was evaluated (Additional file 2: Tables S5 and S9, Additional file 3: Figures S6 and S7), and our assessment confirmed that all were high in quality and could be used for further comparative analyses. In addition, we employed a standard annotation pipeline to predict gene sets, resulting in approximately 40,000 genes in the three fish genomes (42,109 for Sg, 40,333 for Sr and 40,470 for Sa; see more details in Additional file 2: Table S15 and Additional

	S. grahami (Sg)	S. rhinocerous (Sr)	S. anshuiensis (Sa)
Species (photo)			
Eyes	Normal	Reduced	Lost
Scales	Normal and full-cover	Sparse	Sparse
Color	Brownish yellow	Purple brown	Flesh reddish white to translucence
Habitat	Epigean rivers, spring or lake	Moves between epigean and hypogean rivers	Hypogean rivers in perpetual darkness
Genome size	1,753,805,287 bp	1,731,002,715 bp	1,675,987,730 bp
Gene number	42,109	40,333	40,470
Contig N50	29,335 bp	17,658 bp	16,708 bp
Scaffold N50	1,155,972 bp	894,603 bp	1,251,185 bp

Fig. 1 Comparison of biological traits, habitat and basic genomic information for the three sequenced *Sinocyclocheilus* species. They are representative of surface-dwelling (Sg), semi-cave-dwelling (Sr) and cave-restricted (Sa) species, respectively. Please note the regressive characters in the adult Sa, such as loss of eyes, little scale covering and translucent skin

Yang *et al. BMC Biology* (2016) 14:1

Page 3 of 13

file 4: Figure S14), which were double that of most diploid species (Additional file 4: Figure S13). The further Hox gene distribution analyses in three *Sinocyclocheilus* species (Additional file 4: Figure S11) and genome alignment between zebrafish and Sg (Additional file 4: Figure S10) provided more evidence to support the tetraploid nature of *Sinocyclocheilus*.

Phylogenetic relationships and divergence time

The collision between India and Asia in the early Cenozoic may have been the largest-ever orogenic event in the Earth's history [9]. It not only triggered the extraordinary uplift of the Qinghai-Tibet Plateau [10], but also profoundly influenced the Asian climate through the genesis of the Asian monsoon [11], the development of large-scale drainage patterns [12], and especially, the speciation and biodiversity of the organisms living on and below the Plateau (e.g. [9, 13]). Because *Sinocyclocheilus*, the most species-rich cyprinid genus, is endemic to the massive southwestern karst area (the northern part of which abuts the eastern part of the Qinghai-Tibetan Plateau), it has been proposed that the development of this diversity was correlated with the cyclic uplift and planation of the Qinghai-Tibet Plateau [14, 15]. Both the phylogenetic analyses of single-copy orthologous genes and of the mitochondrial DNA datasets (Additional file 5: Figures S15 and S16) supported this correlation, recovering the same relationships for the three *Sinocyclocheilus* species studied here (Fig. 2a). The generated chronogram revealed that Sr emerged first from the other two species at 26.3 Ma, and Sg divided from Sa at 17.5 Ma. These divergence times are similar to those obtained for other speciation events of congeners inferred from a more comprehensive study of Cyprininae phylogeny [16]. These splitting events occurred during the initial and final phases, respectively, of the second tectonic uplift of the Qinghai-Tibet Plateau (25–17 Ma), a time when evidence shows that the limestone region they lived in had been experiencing large-scale karst development [17]. The frequent capture and isolation of the subterranean river systems along with the karst development were regarded as the key reasons for speciation in this genus [15], although the complete picture will remain unknown until more species have been examined. The ancient separation revealed here might be one of the reasons to explain the high species diversity within a relatively narrow distribution area (about 270,000 km^2 [15]) (Additional file 1: Figure S1).

Population history

A reconstruction of the population demography of the three *Sinocyclocheilus* species has revealed a similar start but a subsequent divergent trend from 10^4 to 10^7 years ago (Fig. 2b). It seems that the population demography has a greater correlation with the uplifting of the Qinghai-Tibet Plateau than other significant environmental events, such as the Eurasian ice volume or the atmospheric surface air temperature. In a relatively short period of time (from 3.0 to 0.5 Ma), all three species underwent two rounds of population decline, which occurred following two intense uplift phases in the third tectonic uplift of the Qinghai-Tibet Plateau [17, 18]. These two phases of movements at Qingzang (3.6–1.7 Ma) and Kunhuang (1.1–0.6 Ma) uplifted the Plateau from an average of <1,000 m up to 4,000 m, followed by the intensification of the Asian monsoon and increased precipitation [18], in addition to the large-scale development of glaciers [17]. The patterns of coincident population declines in the three species are believed to be a response of a common background following the intense uplift of the Plateau that may have been unfavorable to them (although it is still not clear which kind of environmental changes contributed the most). However, two events of unusual population expansion were also detected after these events. One was recognized in Sr during the interglacial period (0.5–0.15 Ma), a period during which Sg and Sa kept their low populations, which hints at the presence of extensive subterranean water systems in the Sr-native Luoping Basin at this time. Another was observed for Sg from 0.023 Ma when its main distribution area, the paleo-lake Dianchi, was coincidentally documented from geological sediments to have been some three times larger than at present [19]. Range expansion along new river channels might accompany Plateau uplift, thereby providing some fish the opportunity to increase their population sizes as reported in the typical Plateau schizothoracine fish *Schizopygopsis pylzovi* [20] and *Gymnocypris chilianensis* [21].

Variations of eye structures and related genetic and expression responses

Cavefishes often display regressive features, such as the degeneration of the eyes, which often develop normally during embryogenesis but subsequently arrest, degenerate or sink beneath the skin [22]. In the three adult *Sinocyclocheilus* fishes in this study, Sg has normal-sized eyes with a lens and retinal structures; Sr has small eyes with a reduced lens and retinal cell density; and Sa has lost its external eyeballs and lens, and the retina has degenerated to become merely disorganized and indistinguishable cellular layers (Fig. 3). The recently published *A. mexicanus* genome identified many specific candidate genes under eye-related QTL, and revealed expression differences in several development stages [6]. In our study, we analyzed the genetic changes of these existing crystallin and opsin genes and also presented copy number variations among cave-dwelling and surface-dwelling fishes. Our comparative genomic data suggested that several opsin genes, including *Lws2* (long wavelength-sensitive), *Rh2-1* and *Rh2-2* (middle wavelength-sensitive), have been lost in *Sinocyclocheilus*, and that *Rh2-4* has

Yang *et al. BMC Biology* (2016) 14:1

Page 4 of 13
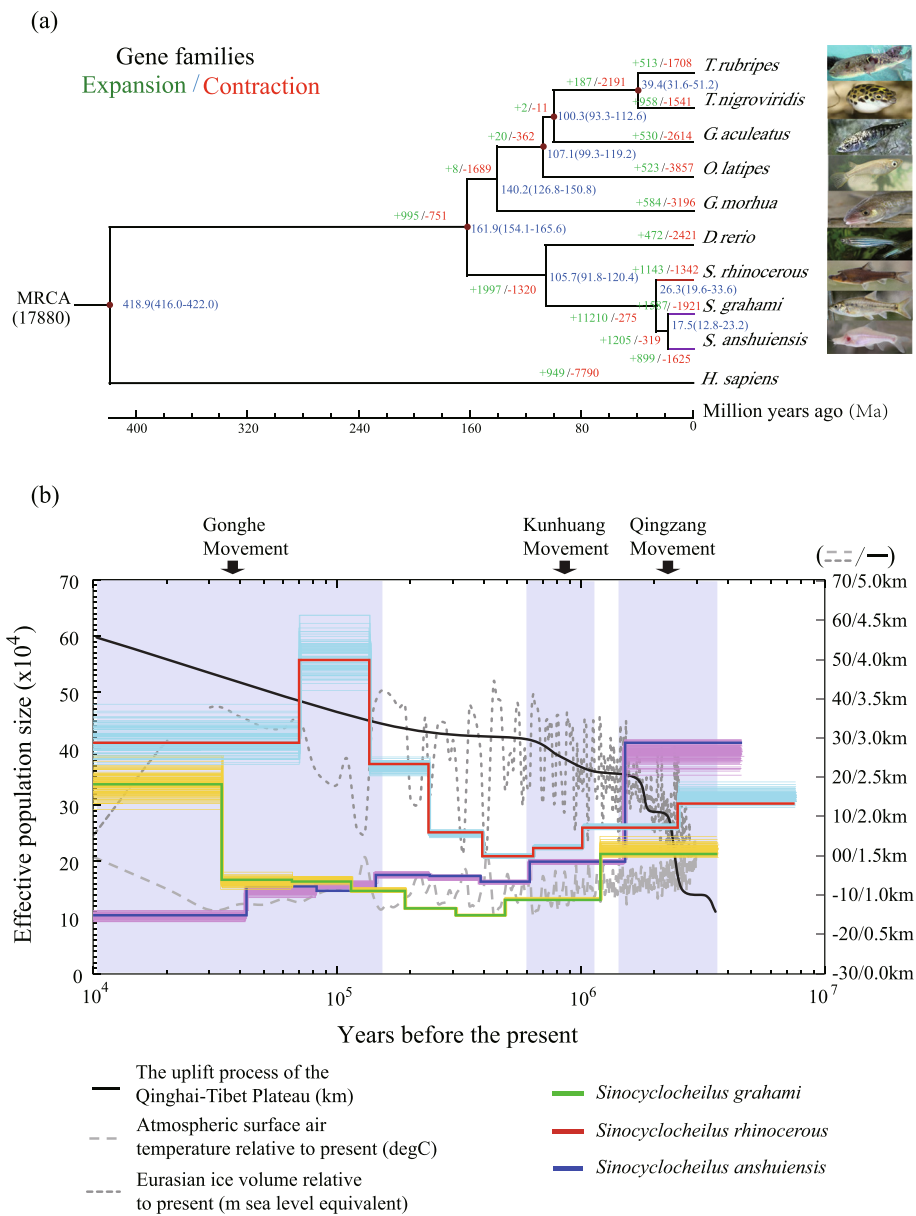
Fig. 2 Phylogenomic analysis and demographic histories of the three *Sinocyclocheilus* species. **a** Phylogenomic relationships inferred from 3,181 orthologous genes of the three *Sinocyclocheilus* and six other teleost species (*Homo sapiens* was the outgroup), with the branch lengths scaled to estimated divergence times (numbers in blue show median and range values). The numbers besides the branch indicated expanded (green) and contracted (red) gene families since the split from a most recent common ancestor (MRCA). **b** Demographic histories were reconstructed using the pairwise sequentially Markovian coalescent (PSMC) model. The uplift process of the Qinghai-Tibet Plateau since 3.6 Ma was obtained from a published paper [11] and other significant environmental events, such as the atmospheric surface air temperature and Eurasian ice volume for the past 3 Ma, were taken from the National Centers for Environmental Information (NCEI; http://www.ncdc.noaa.gov/). The time range of three rounds of intense uplift (Qingzang, Kunhuang and Gonghe Movement) is highlighted in lilac. All three species have similar patterns of decrease under the Qingzang and Kunhuang Movement, but a subsequent divergent trend under the Gonghe Movement

been lost in Sa (Additional file 2: Table S26). In dim conditions, long wavelength light is rapidly attenuated [23], and it seems rational that the long and middle wavelength-sensitive genes were lost specifically in Sa. The results from transcriptomic analyses of eyes also unsurprisingly showed significantly lower expression levels

of most visual opsin genes in Sa when compared with Sg and Sr (Additional file 2: Table S27). Development and maintenance of photoreceptors requires a series of transcriptional factors. As expected, nine important transcriptional factors, including *Crx*, *Nrl*, *Otx2*, *Otx5*, *Nr2e3*, *Ggca1A*, *Gnat1*, *Gnat2* and *Rorb*, were significantly down-
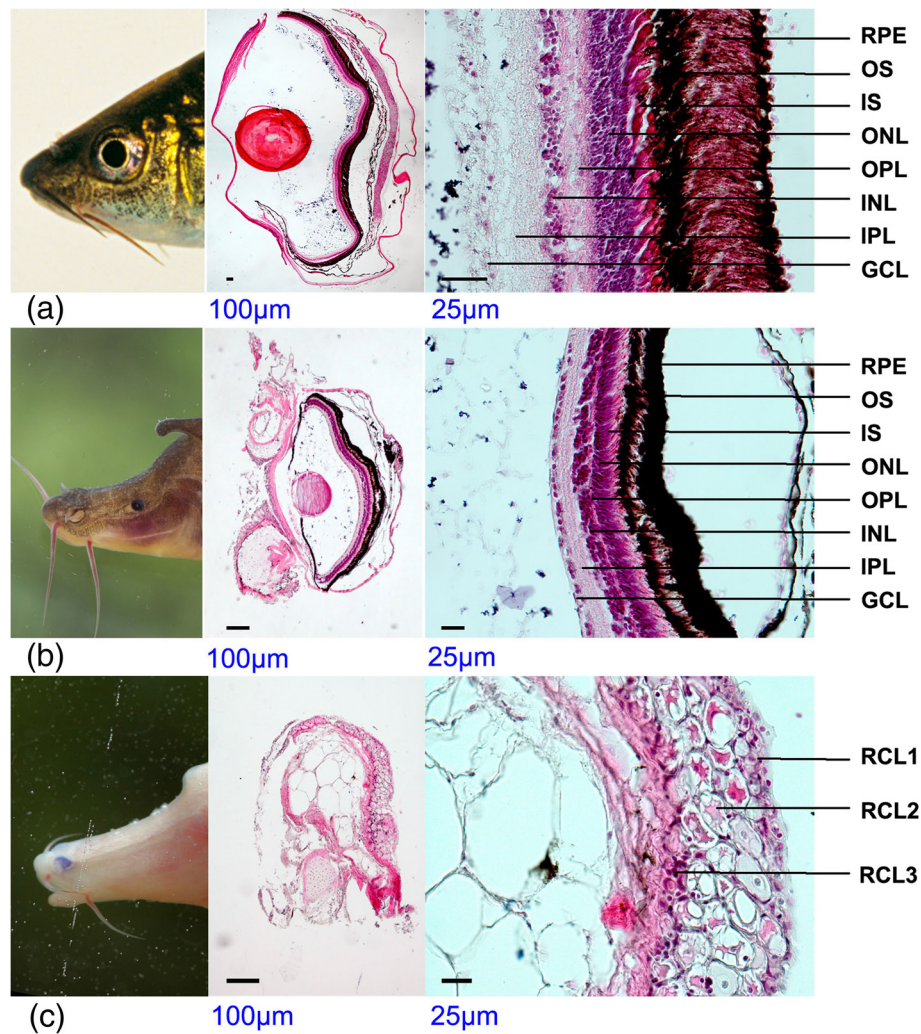
Fig. 3 Comparison of retinal structures among the three *Sinocyclocheilus* species. Phenotypes and H&E stained sections of eyes from top to bottom are those in **a** Sg, **b** Sr and **c** Sa, respectively. GCL, ganglion cell layer; INL, inner nuclear layer; IPL, inner plexiform layer; IS, inner segment; ONL, outer nuclear layer; OPL, outer plexiform layer; OS, outer segment; RCL, relict cell layer; RPE, retinal pigmented epithelium

regulated in Sa (Additional file 2: Table S27 and Additional file 6: Figure S25). These results are consistent with those found for *Astyanax* cavefishes [24] and for other *Sinocyclocheilus* species [25, 26], which supports the hypothesis that down-regulation of rhodopsin might play a critical role in eye degradation of cavefishes [25, 26].

By screening crystallin genes in the three *Sinocyclocheilus* genomes, we found that copy numbers of most genes were lower than those of the diploid zebrafish (Additional file 2: Table S26 and Additional file 6: Figure S28). Transcriptomic data further demonstrated that the expression levels of most crystallin genes were maintained at high levels in Sg, but were not expressed in Sa (Additional file 2: Table S27). We also observed that several crystallin genes in Sa (such as *Cryball1*, *Crygm2d12* and *Crygm7*) have evolved into pseudogenes due to existing premature stop codons

(Additional file 2: Table S25E). Previous studies in *Astyanax* found that the expression of crystallin genes were down-regulated in the development of cavefish lens [27], suggesting that the lens plays a critical role in promoting cell survival in the development of eyes [28]. In particular, *β*- and *γ*-crystallins play a pivotal role in retinal tissue remodeling and repair, and also strongly enhance axon regeneration of retinal ganglion cells [29]. Hence, the lack, or down-regulation, of crystallin gene expression in Sa supports its phenotype of a visual system in which both the lens and retina have degenerated. Interestingly, we also found that Sa has two copies of *Hsp90α1.1* and *Hsp90α1.2* (heat shock protein 90*α*) genes, while both Sg and Sr have only one. Meanwhile, the expression levels of *Hsp90α* in Sa eyes were much higher than those in Sg and Sr (Additional file 2: Table S27). These observations provide

Yang *et al. BMC Biology* (2016) 14:1

Page 6 of 13

further evidence to support a novel role of *Hsp90α* in lens apoptosis and eye degeneration of cavefishes [30].

Different mechanisms of albinism compared with *Astyanax*

In order to check the mechanisms of loss of pigmentation in Sa skin, we compared melanogenesis-related genes in the three *Sinocyclocheilus* genomes. In *Astyanax*, the albinism of some cave populations is caused by the deletions of *Oca2* gene exon regions, which disturbs the upstream steps of the melanin synthesis pathway [31]. Although similar deletions of *Oca2* genes were not found in either of the two copies in Sa (Additional file 6: Figure S21), nor were any identical mutations specifically present in Sa and *Astyanax* cave populations, some other new mutations were identified (Additional file 6: Figure S21), and the transcriptome analysis performed on the skin of Sg, Sr and Sa indicated that the expression of *Oca2* gene in Sa was the lowest (Additional file 2: Table S20). In the melanin synthesis pathway, several enzymes work downstream, such as Tyr (tyrosinase), a key rate-limiting enzyme, and Tyrp1 (tyrosinase-related protein 1) [32]. More interestingly, we found that Tyr has an amino acid mutation (G420R) in Sa (Additional file 6: Figure S18), which was identical to that identified in Caucasian human patients (G419R, http://www.ifpcs.org/albinism/oca1mut.html) [33, 34]. Meanwhile, the expression levels of Sa genes in the melanogenesis pathway in the skin transcriptomes were the lowest, especially *Tyr* and *Tyrp1*, with significantly lower expression levels in Sa compared with Sg (Additional file 2: Table S20). It seems that similar phenotypes in *Sinocyclocheilus* cavefish might evolve by different mechanisms from *Astyanax*. Furthermore, we found that the Mpv17 protein in Sa had a deletion in the signal region compared with zebrafish, Sg and Sr (Additional file 6: Figure S22), and the expression level of *Mpv17* in Sa is the lowest (Additional file 2: Table S20). A previous study shows that the deletion in *Mpv17* can cause the *tra* mutant phenotype in zebrafish, and cause a loss (or strong reduction) of iridophores throughout larval and adult stages [35]. This study also pointed out that differentiated iridophores were required for the accumulation and maintenance of melanophores during pigment pattern formation [35], and a parallel study showed that the interaction between iridophores and other chromatophores is critical in the stripe formation of zebrafish [36]. Thus we infer that the deletion of *Mpv17* might cause the loss of iridophores, which affected the formation of melanophores, and consequently played a role in the albinism of Sa.

Gene mutation and loss in scale degeneration

A previous study has indicated that mutations in the ectodysplasin-A receptor (*Edar*) encoding locus can lead to complete scale loss in fish such as medaka [37]. For this reason, two copies of *Edar* gene (named as *Edar1* and *Edar2*, respectively) in the three *Sinocyclocheilus* genomes were identified and checked. Interestingly, one of the proteins, Edar1, has deletions in the signal peptide and partial extracellular regions in all three *Sinocyclocheilus* species, which may lead to functional changes or loss in this copy (Additional file 6: Figure S20). For the other protein copy, Edar2, only Sa has the signal peptide region and partial extracellular regions totally deleted when compared with Sg and Sr (Additional file 6: Figure S20). This deletion in Sa may lead to a functional disorder in guiding the Edar protein transfer across the membrane, thus generating fewer scales at the skin surface of Sa (Additional file 7: Figure S29). Coincidentally, two important genes, *Lamb3* and *Col7a*, were also lost from the Sa genome, which may cause a defect in the anchoring between the epidermis and dermis, resulting in friction and skin fragility [38, 39] in the scale covering.

Possible hearing loss in Sa

The hearing of cavefishes is interesting but less studied than the other functions discussed above. We found the deletion of *Mpv17* gene, which we mentioned in the section on albinism, may also have some influence in hearing. A previous study reported that *Mpv17*-deficient mice suffered from degeneration of cochlea and loss of sensorineural hearing at 2 months old [40]. Another gene, *Ush2a*, has also changed in Sa, especially two amino acid sites, R334S [41, 42] and V382A [43] (referred to the human sites), which may affect splicing and the termination codon (Additional file 6: Figure S19). It has been proved that the encoded protein of *Ush* is found in the basement membrane and may be related to the development of the inner ear [44, 45]. These mutations in Sa may cause its sensorineural deafness. There are no similar changes in Sg and Sr (Additional file 6: Figure S22). Reconstructions of the saccular otolith morphology using synchrotron X-ray microtomography among these three species show that the ventral surface of this otolith in Sa is seriously aberrant, and the degree of corrosion has distinctly increased in the following order: Sg < Sr < Sa (Fig. 4). Both the anatomy of the swim bladder (Additional file 7: Figure S30) and the numbers of neuromasts and scales of the trunk lateral line system (Additional file 7: Figure S31) also indicate these hearing-related organs in Sa have different degrees of weakness. Our data together indicated that the cave-restricted species Sa might have reduced hearing, which could be similar with that demonstrated in amblyopsid cavefishes [46]. Furthermore, a comparison of the distribution of neuromasts on the head (Fig. 4) also suggested that the response to vibration of these three species was Sg > Sr > Sa, which is different from the general pattern in *Astyanax* [47].

Yang *et al. BMC Biology* (2016) 14:1

Page 7 of 13
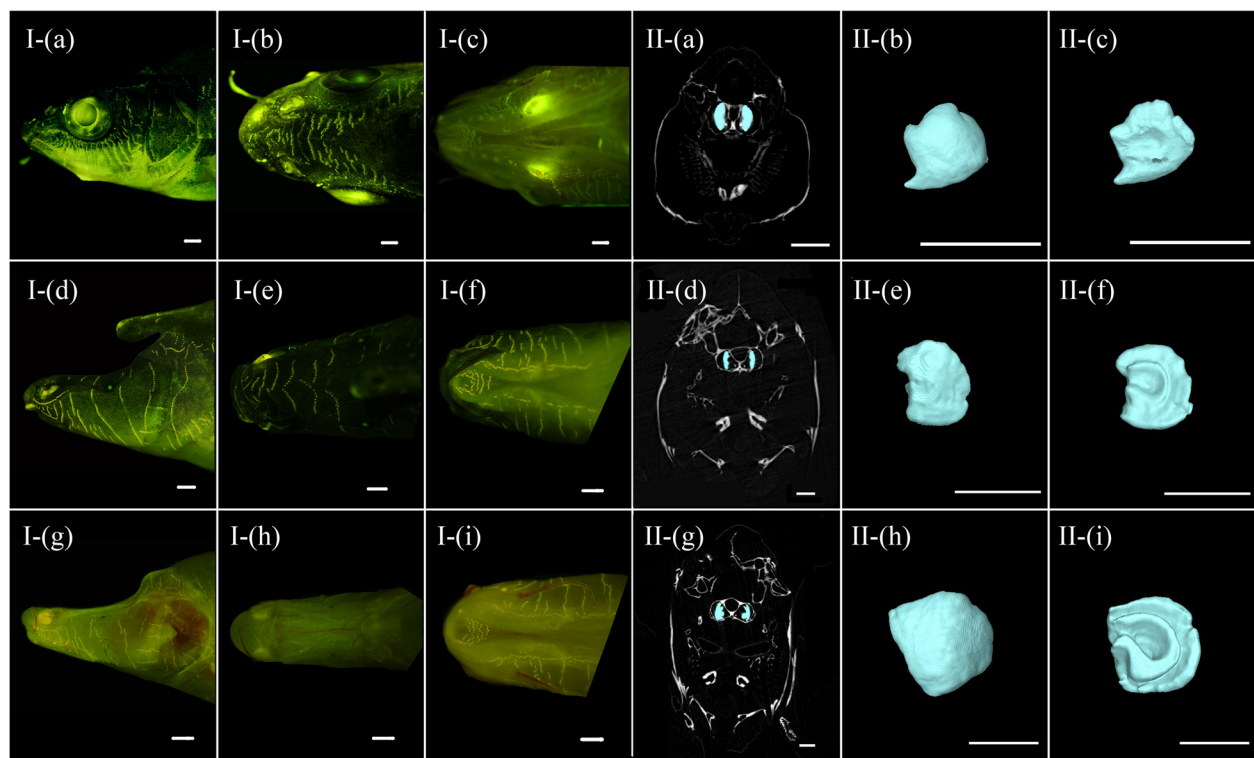
Fig. 4 I The distributions of superficial neuromasts on the head and **II** morphology of saccular otolith in the inner ear among the three *Sinocyclocheilus* species. The superficial neuromasts after DASPEI staining from the plates I-(a–c), I-(d–f) and I-(g–i) represent Sg, Sr and Sa, respectively. The photos from left to right show the lateral view, dorsal view and ventral view. These figures show that the numbers of neuromasts in the adult fishes decline in the following order: Sg > Sr > Sa. The morphology of the saccular otoliths was reconstructed based on synchrotron X-ray microtomography. The plates II-(a–c), II-(d–f) and II-(g–i) represent Sg, Sr and Sa, respectively. The photos from left to right show the location of saccular otoliths in the inner ear, the dorsal view and ventral view of its morphology. The ventral of saccular otolith in Sa is seriously aberrant, with a deep and expanded central pit, encircled by another lateral sulcus. The degree of corrosion increase is in the following order: Sg < Sr < Sa. Scale bar: 1 mm

Different immune responses to specific habitats

The immune activities of Sa may be lower than its epigean counterparts because it probably lives in a less diverse microbial environment. The fact that Sa is more susceptible to disease in captivity might support this inference (unpublished data). We found relatively fewer copies of immune genes in Sa when compared with those in Sg and Sr (Additional file 2: Table S22). However, one important innate immune group, the *Tlr* (toll-like receptor) gene family, showed some degree of expansion in Sa (Additional file 2: Table S23), through the duplication of *Tlr8* and *Tlr18* in the Sa genome. This suggests that these three species may have evolved differential immune activities for innate immunity and adaptive immunity according to their different habitats. Interestingly, the semi-cave-dwelling Sr had substantially more copy numbers of immune genes than Sg and Sa (Additional file 2: Table S22), which may be a result of an adaptation to heterogeneous elements between epigean and hypogean habitats, as found in the amphibious mudskippers [48].

Lack of diurnal rhythms

Previous research reported that some cavefishes lack diurnal rhythms when living in perpetual darkness [49]. For example, cave populations of *A. mexicanus* have a phenotype of reduced sleep in comparison to their surface relatives [50]. We found that both the two copies of *Skp1* in Sa had deletions in the N-terminal end of its protein (Additional file 6: Figure S24). Since Skp1 is one of the components of the Skp1-Cul1-Fbxl3 (SCF) protein complex, and the SCF complex is most relevant in the mammalian clock mechanism [51], the deletions in Skp1 might lead to some dysfunction in SCF, which suggests weaker light rhythms in Sa. Meanwhile, the transcriptomic analysis of the eye demonstrated that expression levels of the rhythm pathway genes decrease in the order Sg > Sr > Sa (Additional file 2: Table S28 and Additional file 6: Figure S26).

Low fecundity in Sa

Although low fecundity is often assumed to be normal in cave species [4], there is little empirical evidence, and

Yang *et al. BMC Biology* (2016) 14:1

Page 8 of 13

the different fecundity levels between surface and cave forms sometimes seem to be habitat plastic (e.g. [52]). We know of no study on the fecundity of the cave-dwelling *Sinocyclocheilus* species. Our analysis of the absolute fecundity (number of mature eggs) of Sg and Sa (2,402.9 ± 881.9 in Sg [53] vs. 143 ± 116 in Sa (count from four specimens with mature eggs in this study)) indicated that the fecundity in cave *Sinocyclocheilus* species is much less than surface congeners. Interestingly, one related gene, *Creb3l4*, was found to have been lost in the Sa genome. It has been reported that *Creb3l4* can regulate the expression of genes required for germ cell survival, although it is insufficient to disrupt the normal fertility in mice [54].
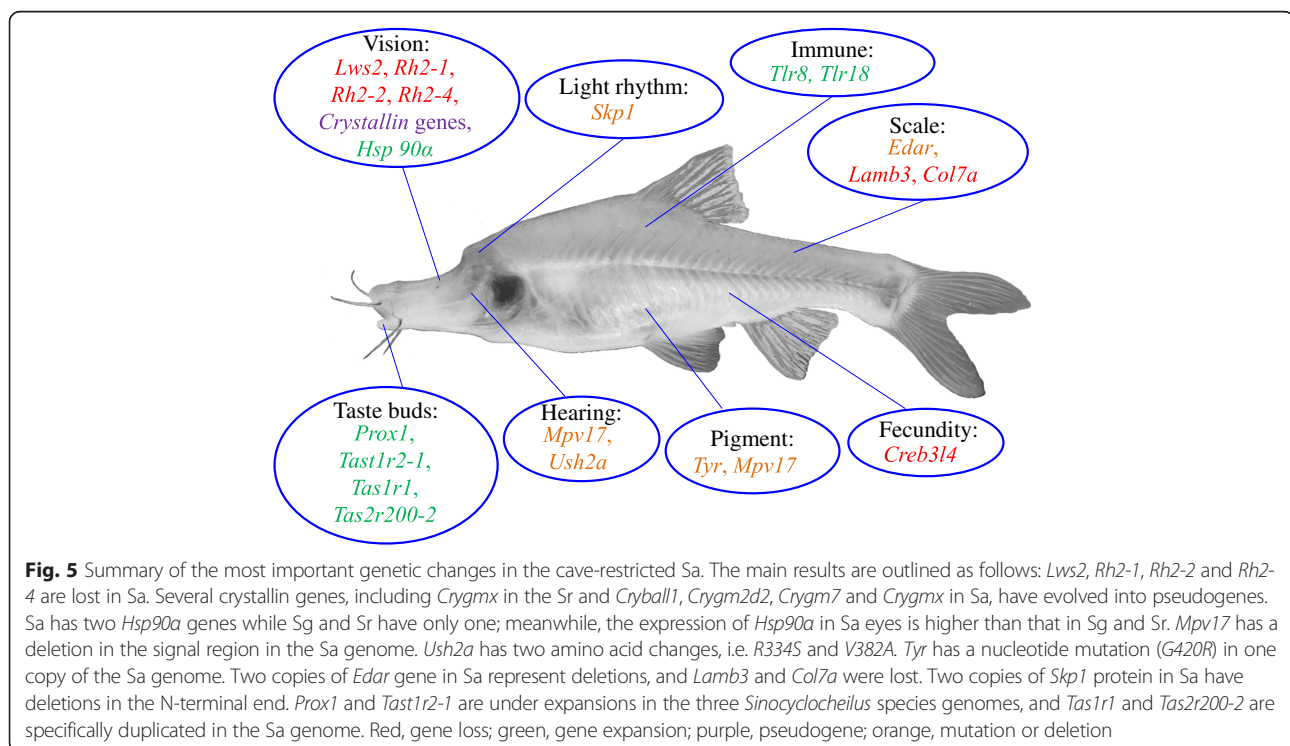
Enhancement of taste

Taste buds are enhanced in some cavefishes, such as *Astyanax* [55]. We applied the taste-related gene sequences of zebrafish to BLAST to the three *Sinocyclocheilus* genomes, and unexpectedly found that one taste receptor gene, the *Tast1r2-1*, was significantly expanded (almost fourfold) and one important transcription factor, the *Prox1* gene, had threefold copies in these three species compared with zebrafish (Additional file 8: Table S29). However, if these expansions were correlated to the overall enhancement of taste in the three *Sinocyclocheilus* species (benthic and cave-preferred) we still need further testing. Some taste receptor genes, such as *Tas1r1* and *Tas2r200-2*, were

specifically duplicated in the Sa genome (vs. Sg and Sr) (Additional file 8: Table S29), which suggests a further improvement in the sense of taste in the cave-restricted Sa. A preliminary study of the distribution of taste buds within the jaws of these three species also indicated that their numbers increase in a sequence from Sg < Sr < Sa (Additional file 7: Figure S32).

Conclusion

This paper focused on comparative genomic and transcriptomic studies of three *Sinocyclocheilus* species, representative of surface-dwelling (Sg), semi-cave-dwelling (Sr) and cave-restricted (Sa) species. We found that speciation and population dynamics of these fishes are closely related to the uplifting stages of the Qinghai-Tibetan Plateau. Comparative genome analysis revealed many genetic changes, such as gene loss, pseudogenes, mutations and down-regulation, which were associated with regressive features (eye degeneration, albinism, rudimentary scales and low fecundity), and some gene expansions suggest that constructive features (such as the sense of taste) have evolved in caves (a few supplementary notes can be seen in Additional file 9: Note S6). The most important genetic changes in the cave-restricted Sa are summarized in Fig. 5. Among them, some changes are similar to those in other cavefishes (e.g. *A. mexicanus*) from very different parts of the world (such as the down-regulation of some rhodopsin-related genes), but some are

Fig. 5 Summary of the most important genetic changes in the cave-restricted Sa. The main results are outlined as follows: *Lws2, Rh2-1, Rh2-2* and *Rh2-4* are lost in Sa. Several crystallin genes, including *Crygmx* in the Sr and *Cryball1, Crygm2d2, Crygm7* and *Crygmx* in Sa, have evolved into pseudogenes. Sa has two *Hsp90α* genes while Sg and Sr have only one; meanwhile, the expression of *Hsp90α* in Sa eyes is higher than that in Sg and Sr. *Mpv17* has a deletion in the signal region in the Sa genome. *Ush2a* has two amino acid changes, i.e. *R334S* and *V382A*. *Tyr* has a nucleotide mutation (*G420R*) in one copy of the Sa genome. Two copies of *Edar* gene in Sa represent deletions, and *Lamb3* and *Col7a* were lost. Two copies of *Skp1* protein in Sa have deletions in the N-terminal end. *Prox1* and *Tast1r2-1* are under expansions in the three *Sinocyclocheilus* species genomes, and *Tas1r1* and *Tas2r200-2* are specifically duplicated in the Sa genome. Red, gene loss; green, gene expansion; purple, pseudogene; orange, mutation or deletion

Yang *et al. BMC Biology* (2016) 14:1

Page 9 of 13

not (such as without vs. with the exon deletion of *Oca2* gene in Sa vs. some *A. mexicanus* cave populations). Although the incidence of gene losses and expansions in cave-restricted Sa corresponds to the distinct phenotype variations, the changes seem to be fixed in its genome rather than in other mechanisms. The genetic changes found at the genome level in this study, although needing further functional confirmation, still give some good clues to understanding the mysteries of cave adaptation.

In the history of cave biology research, people have often been confused by the plethora of views and terms, such as disuse, preadaptation, opportunism, compensation, regressive evolution, etc. Adaptations to cave living are complicated because there is no morphological 'archetype' for cave animals, and the characters show a highly diverse mix [4]. There is still a long way to go before this work can be applied to general evolutionary theory. As the first report on cavefish genomes among distinct species in *Sinocyclocheilus*, our work provides not only insights into genetic mechanisms of cave adaptation, but is also a fundamental resource for better understanding of cavefish biology.

Methods

Genome sequencing and assembly

Three female adult *Sinocyclocheilus* fishes, collected from Yunnan province (Sg and Sr) and Guangxi province (Sa) of China, were used for sequencing. The research protocol and treatment of experimental fishes have been reviewed and approved by the internal review board of the Kunming Institute of Zoology, Chinese Academy of Sciences (approval ID: SYDW-2014020). High-quality genomic DNA was extracted from muscle tissues using Puregene Tissue Core Kit A (Qiagen, MD, USA) for construction of libraries with different inserted sizes (250 bp to 20 kb) (Additional file 2: Table S1). In total, 25 paired-end libraries (11 for Sg, 7 for Sr and 7 for Sa, respectively) were generated with the Illumina standard operating procedure. HiSeq 2000 shotgun sequencing and assembling by SOAPdenovo assembler were performed as previously reported [56]. Artificial and low-quality reads were filtered first and then sequencing errors with 17-mer frequency lower than four were collected using a method described in a previous study [57]. Next, 88.05-, 48.31- and 60.54-fold coverage of *Sinocyclocheilus* genomes were used for assembly (Additional file 2: Table S2). In addition, we estimated genome sizes of the three *Sinocyclocheilus* fishes using the 17-mer depth frequency distribution method: G (Genome size) = K-mer_num/Peak_depth. The estimated genome sizes are 1.79, 1.89 and 1.81 Gb, respectively (Additional file 2: Table S4). Subsequently, the filtered reads were assembled into contigs and scaffolds with SOAPdenovo [58] and gaps were fulfilled with GapCloser

[58]. Finally, a primary assessment was performed on the genome assemblies (Additional file 9: Note S1).

Genome annotation

We performed homology-based predictions by running RepeatMasker (version 3.3.0) [59] against the RepBase [60], and identified repeat sequences at DNA and protein levels using TE library (version 16.10) and Repeat-ProteinMask. We searched the *de novo* prediction build repeat library using RepeatModeler (version 1.0.5) and generated TE results with classification information for each repeat family by running RepeatMasker on the genome sequences subsequently. Tandem repeats were also searched using Tandem Repeats Finder (version 4.04) [61]. The protein coding genes were obtained using a combination of the *de novo* method, homology-based gene prediction and RNA-Seq data (Additional file 9: Note S2). All predicted gene evidence was integrated by GLEAN [62] to get non-redundant data [63].

Transcriptome analysis

RNA was isolated for sequencing from four tissues (eye, skin, liver and ovary) of Sg, Sr and Sa, respectively. We applied an in-house C++ program to filter raw reads and then obtain high-quality reads (Additional file 9: Note S3). All of the clean RNA-Seq reads were mapped onto the corresponding reference genomes (Sg, Sr and Sa) using TopHat (version 2.0.4) [64]. According to the mapped results, transcripts were constructed using Cufflinks (version 2.0.0) [65]. The *de novo* transcriptomes of the four tissues were assembled by Trinity with filtered reads from each tissue separately into contigs and scaffolds. Trinity contains Inchworm, Chrysalis and Butterfly, which were employed sequentially to process large volumes of RNA-Seq reads.

Evolutionary analysis

(1) Gene family cluster: we defined gene families using TreeFam (http://www.treefam.org) among the three *Sinocyclocheilus* fishes (Sg, Sr and Sa) and seven other vertebrates, including fugu (*Takifugu rubripes*), green spotted puffer (*Tetraodon nigroviridis*), three-spined stickleback (*Gasterosteus aculeatus*), Atlantic cod (*Gadus morhua*), medaka (*Oryzias latipes*), zebrafish (*Danio rerio*) and human (*Homo sapiens*). A total of 17,883 gene families and 210 single-copy gene families were identified. The numbers of orthologous genes across the ten species were counted (Additional file 2: Table S19 and Additional file 4: Figure S13) and plotted in a Venn diagram (Additional file 4: Figure S14). (2) Phylogenetic analysis: phylogenetic relationships were established using 3,181 single-copy orthologous genes shared among nine teleost fish genomes (Additional file 5: Figure S15) using maximum likelihood (ML) method in PhyML [66, 67] and Bayesian

Yang *et al. BMC Biology* (2016) 14:1

Page 10 of 13

inference method in MrBayes [68]. An additional dataset of six mitochondrial gene sequences were also used to reconstruct the phylogenetic trees (Additional file 5: Figure S16). (**3**) Divergence time estimation: the divergence times were estimated using the mcmctree [69] in PAML [70] and recalculated using the multidivtime [71, 72] program, and all were calibrated by five fossil records [73] (Additional file 5: Figure S17 and Additional file 9: Note S4). (**4**) Demographic history: the distribution of time to TMRCA (the most recent common ancestor) between two alleles in an individual can be related to the history of population size fluctuation. The population size histories of Sg, Sr and Sa were inferred using the pairwise sequentially Markovian coalescent (PSMC) model [74] on heterozygous sites with the generation time ($g = 1$ year) (according to artificial breeding of Sg) and the mutation rate ($m = 3.51 \times 10^{-9}$ per year per nucleotide) [75]. Reconstructed population history was plotted for Sg, Sr and Sa separately using gnuplot (version 4.4) [76]. In addition, we obtained the atmospheric surface air temperature (°C) and Eurasian ice volume (m sea level equivalent) data for the past 3 million years from the NCEI (http://www.ncdc.noaa.gov/). (**5**) Gene family contraction and expansion: we performed *Sinocyclocheilus* lineage-specific expansion and contraction analysis using the CAFE program [77]. Based on random birth and death model [78], a global parameter λ was estimated using maximum likelihood. Comparing each branch and their ancestor branch, we calculated a conditional *P* value and marked families with a *P* value of less than 0.05 as a significant change [79], which means it underwent contraction or expansion during evolution. These families were then subjected to GO/KEGG/IPR enrichment analyses along each *Sinocyclocheilus* lineage, respectively (Additional file 2: Table S25). More details are shown in Additional file 8: Table S29. (**6**) Genes with accelerated evolutionary rate: positive Darwinian selection at the DNA sequence level has been tested by estimating the ratio (ω) of nonsynonymous nucleotide substitutions (dN) to synonymous nucleotide substitutions (dS) between ortholog genes [80]. A branch-site model was used to search for the positive selection genes (PSG) [81]. After obtaining the *Sinocyclocheilus* PSG list (Additional file 10: Table S30), we converted it to the corresponding human orthologs as the input against a background of human genes [80] using the DAVID Functional Annotation [82] tool. (**7**) Evolution of Hox clusters: to define Hox genes in the three *Sinocyclocheilus* genomes, the Hox genes of zebrafish were downloaded from the ensemble as cross-references. The Hox gene numbers (Additional file 2: Table S18) and the order along the scaffolds (Additional file 4: Figure S11) in *Sinocyclocheilus* indicated that these fishes are indeed tetraploids when compared to diploid zebrafish. (**8**) Loss of Sa-specific gene families: in order to identify Sa-specific gene family loss, we extracted gene families that have no member in Sa while more than zero in the other nine species. The lost gene family list is included in Additional file 2: Table S24.

Morphological comparison

(**1**) Paraffin sections of the eyes of three *Sinocyclocheilus* species: histological studies on eyes were performed from paraffin sections and hematoxylin and eosin (H&E) staining. (**2**) Immunocytochemistry of taste buds: this analysis focused on distributions of taste buds on upper and lower jaws, using a primary antibody (rabbit against calretinin, labeling entire receptor cells within taste buds) according to a standard protocol (Additional file 9: Note S5). (**3**) Anatomy of gas bladder, absolute fecundity and other measurements: all samples for anatomy and measurements were from specimens immersed in 90 % ethanol. (**4**) Synchrotron X-ray microtomography of the saccular otolith: this experiment was performed at BL13W1 beamline at the Shanghai Synchrotron Radiation Facility (Shanghai, China); the slices were reconstructed using FBP algorithm, and 3D renderings were created and manipulated in VGStudio 2.1 software.

Data availability

The Whole Genome Shotgun projects have been deposited at GenBank under accession numbers LCYQ00000000 (Sg), LAVF00000000 (Sr) and LAVE00000000 (Sa), which are the same versions described in this paper. The RNA-Seq data from four tissues (eye, skin, liver and ovary) have been deposited at GenBank under accession numbers SRS1179797–SRS1179800 (Sg), SRS1179996–SRS1179999 (Sr) and SRS1180000–SRS1180003 (Sa).

Additional files

Additional file 1: Figure S1. The collection sites of three *Sinocyclocheilus* species in this study. (PDF 415 kb)

Additional file 2: Table S1–S28. The supplementary data and statistics, including the information of genome sequencing and assembly (Table S1–S9), genome annotation (Table S10–S17), evolutionary analyses (Table S18–S24), and cave adaption analysis (Table S25–S28). (XLSX 134 kb)

Additional file 3: Figures S2–S7. The characteristics of genomes, including the 17-mer frequency distribution (Figure S2), GC content distributions (Figure S3), correlation of GC content and sequencing depth (Figure S4), sequencing depth (Figure S5), comparison of the assembled Sg genome with three fosmid sequences (Figure S6), and conserved sequences (Figure S7). (PDF 468 kb)

Additional file 4: Figure S8–S14. The characteristics of repeats and genes, including the distributions of TEs (Figure S8), mRNA, CDS, exon and intron lengths (Figure S9), double-conserved synteny between Dr and Sg genomes (Figure S10), the evolution of Hox genes (Figure S11) and the similarity analysis (Figure S12), protein orthology comparison (Figure S13), and shared orthologous groups among Dr and three *Sinocyclocheilus* genomes (Figure S14). (PDF 666 kb)

Additional file 5: Figure S15–S17. The phylogenetic trees and divergence time analyses, including the phylogenetic trees based on CDS, protein, 1st, 2nd, 1st and 2nd, 3rd, 4d codon position of codon sequences of 3,181 genes (Figure S15), mitochondrial CDS sequences

Yang *et al. BMC Biology* (2016) 14:1

Page 11 of 13

(Figure S16), and divergence times calibrated by five fossil records (Figure S17). (PDF 170 kb)

Additional file 6: Figure S18–S28. Analyses to the specific genes, including *Tyr* (Figure S18), *Ush2a* (Figure S19), *Edar* (Figure S20), *Oca2* (Figure S21), *Mpv17* (Figure S22), *Rom1b* (Figure S23), *Skp1* (Figure S24), and differential expression of eye-related genes, transcriptional factors (Figure S25) and circadian rhythm pathway genes (Figure S26), and similarity cluster of the *Tlr* gene family (Figure S27) and crystallins (Figure S28). (PDF 1295 kb)

Additional file 7: Figure S29–S32. Morphological analyses of the three *Sinocyclocheilus* species, including scales (Figure S29), swim bladder (Figure S30), lateral line system (Figure S31) and taste buds (Figure S32). (PDF 434 kb)

Additional file 8: Table S29. The results of gene family contraction and expansion. (XLSX 75 kb)

Additional file 9: Notes S1–S6. The supplementary notes to the article, including the organism background, genome sequencing and assembly (Note S1), genome annotation (Note S2), transcriptome analysis (Note S3), evolutionary analyses (Note S4), morphological comparison analysis (Note S5), and cave adaption analysis in the aspects of vision, pigmentation, scale development, hearing, immune response, circadian rhythm and sense of taste (Note S6). (PDF 668 kb)

Additional file 10: Table S30. The *Sinocyclocheilus* positive selection gene (PSG) list. (XLSX 153 kb)

Authors' contributions
QS, JY, JB and WJ conceived the project and designed scientific objectives. XP, XW, XY, ZR, JC, Xiaoyong C, LZ, H Yu, YS, Le C, XM, CP and JX collected and prepared the fish samples. Xiaoli C, DF, H Yuan, Jintu W, SH, Jiang L, YZ, YW, YL, CB, GF, HZ, DM, XZ, JL and YQ conducted the genome assembly, annotation and bioinformatics analysis. WJ, XW, XP and YH did the morphological comparison analysis. JB, WJ, Xiaoli C, DF, YQ, H Yuan, QS, JY and TW prepared the manuscript. JX, ZX, PX, Jian W, H Yang, Jun W and XX participated in discussions and provided suggestions. All authors read and approved the final manuscript.

Author details
[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. [2]BGI-Shenzhen, Shenzhen 518083, China. [3]Shenzhen Key Lab of Marine Genomics, State Key Laboratory of Agricultural Genomics, Shenzhen 518083, China. [4]Fauna & Flora International, Cambridge CB1 2JD, UK. [5]China National Genebank, Shenzhen 518083, China. [6]Agricultural Genomes Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China. [7]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [8]School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China. [9]BGI-Yunnan, Kunming 650106, China. [10]Shenzhen BGI Fisheries Sci & Tech Co. Ltd., Shenzhen 518083, China. [11]Zhenjiang BGI Fisheries Science & Technology Industrial Co. Ltd., Zhenjiang 212000, China. [12]Shanghai Synchrotron Radiation Facility, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201204, China. [13]State Key Laboratory of Reproductive Medicine, Department of Prenatal Diagnosis, Nanjing Maternity and Child Health Care Hospital, Nanjing Medical University, Nanjing 210029, China. [14]Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi 214081, China. [15]James D. Watson Institute of Genome Science, Hangzhou 310008, China. [16]Department of Biology, Ole Maaløes Vej 5, University of Copenhagen, DK-2200 Copenhagen, Denmark.

References

1. Jeffery WR. Cavefish as a model system in evolutionary developmental biology. Dev Biol. 2001;231(1):1–12.
2. Soares D, Niemiller ML. Sensory adaptations of fishes to subterranean environments. Bioscience. 2013;63(4):274–83.
3. Jeffery WR. Regressive evolution in *Astyanax* cavefish. Ann Rev Gene. 2009;43:25–47.
4. Romero A. Cave biology, life in darkness. New York: Cambridge University Press; 2009.
5. Rohner N, Jarosz DF, Kowalko JE, Yoshizawa M, Jeffery WR, Borowsky RL, et al. Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. Science. 2013;342(6164):1372–5.
6. McGaugh SE, Gross JB, Aken B, Blin M, Borowsky R, Chalopin D, et al. The cavefish genome reveals candidate genes for eye loss. Nat Commun. 2014;5:5307.
7. Shu S, Jiang W, Whitten T, Yang J, Chen X. Drought and China's cave species. Science. 2013;340:272.
8. Wang D, Chen Y. The origin and adaptive evolution of the genus *Sinocyclocheilus*. Acta Hydrobiol Sinica. 2000;24(6):630–4.
9. Che J, Zhou WW, Hu JS, Yan F, Papenfuss TJ, Wake DB, et al. Spiny frogs (Paini) illuminate the history of the Himalayan region and Southeast Asia. Proc Natl Acad Sci U S A. 2010;107(31):13765.
10. Harrison T, Copeland P, Kidd W, Yin A. Raising Tibet. Science. 1992;255(5052):1663–70.
11. Zhisheng A, Kutzbach JE, Prell WL, Porter SC. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. Nature. 2001;411(6833):62–6.
12. Clark M, Schoenbohm L, Royden L, Whipple K, Burchfiel B, Zhang X, et al. Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. Tectonics. 2004;23(1):1006–29.
13. He DK, Chen YF. Biogeography and molecular phylogeny of the genus *Schizothorax* (Teleostei: Cyprinidae) in China inferred from cytochrome b sequences. J Biogeogr. 2006;33(8):1448–60.
14. Xiao H, Chen S, Liu Z, Zhang R, Li W, Zan R, et al. Molecular phylogeny of *Sinocyclocheilus* (Cypriniformes: Cyprinidae) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol. 2005;36(1):67–77.
15. Zhao Y, Zhang C. Endemic fishes of *Sinocyclocheilus* (Cypriniformes: Cyprinidae) in China-species diversity, cave adaptation, systematics and zoogeography. Beijing: Science Press; 2009.
16. Pasco-Viel E, Yang L, Veran M, Balter V, Mayden RL, Laudet V, et al. Stability versus diversity of the dentition during evolutionary radiation in cyprinine fish. Proc Biol Sci. 2014;281(1780):20132688.
17. Shi Y, Li J, Li B, Yao T, Wang S, Li S, et al. Uplift of the Qinghai-Xizang (Tibetan) plateau and east Asia environmental change during late Cenozoic. Acta Geogr Sin. 1999;54(1):10–20.
18. Li J, Fang X. Uplift of the Tibetan Plateau and environmental changes. Chinese Sci Bull. 1999;44(23):2117–24.
19. Zhu H, Chen Y, Pu P, Wang S, Zhuang D. Environments and sedimentation of fault lakes, Yunnan province. Beijing: Science Press; 1989.
20. Qi D, Guo S, Zhao X, Yang J, TANG W. Genetic diversity and historical population structure of *Schizopygopsis pylzovi* (Teleostei: Cyprinidae) in the Qinghai-Tibetan Plateau. Freshwater Biol. 2007;52(6):1090–104.
21. Zhao K, Duan Z, Peng Z, Gan X, Zhang R, He S, et al. Phylogeography of the endemic *Gymnocypris chilianensis* (Cyprinidae): sequential westward colonization followed by allopatric evolution in response to cyclical Pleistocene glaciations on the Tibetan Plateau. Mol Phylogenet Evol. 2011;59(2):303–10.
22. Langecker TG, Neumann B, Hausberg C, Parzefall J. Evolution of the optical releasers for aggressive behavior in cave-dwelling *Astyanax fasciatus* (Teleostei, Characidae). Behav Processes. 1995;34(2):161–7.

Yang *et al. BMC Biology* (2016) 14:1

Page 12 of 13

23. Tobler M, Coleman SW, Perkins BD, Rosenthal GG. Reduced opsin gene expression in a cave-dwelling fish. Biol Lett. 2010;6(1):98–101.

24. Strickler AG, Jeffery WR. Differentially expressed genes identified by cross-species microarray in the blind cavefish *Astyanax*. Integr Zool. 2009;4(1):99–109.

25. Meng F, Zhao Y, Postlethwait JH, Zhang C. Differentially-expressed genes identified in cavefish endemic to China. Curr Zool. 2013;59(2):170–4.

26. Meng F, Braasch I, Phillips JB, Lin X, Titus T, Zhang C, et al. Evolution of the eye transcriptome under constant darkness in *Sinocyclocheilus* cavefish. Mol Bio Evol. 2013;30(7):1527–43.

27. Strickler AG, Byerly MS, Jeffery WR. Lens gene expression analysis reveals downregulation of the anti-apoptotic chaperone αA-crystallin during cavefish eye degeneration. Dev Genes Evol. 2007;217(11–12):771–82.

28. Strickler AG, Yamamoto Y, Jeffery WR. The lens controls cell survival in the retina: evidence from the blind cavefish *Astyanax*. Dev Biol. 2007;311(2):512–23.

29. Thanos S, Bohm MR, Meyer Zu Horste M, Prokosch-Willing V, Hennig M, Bauer D, et al. Role of crystallins in ocular neuroprotection and axonal regeneration. Prog Retin Eye Res. 2014;42C:145–61.

30. Hooven TA, Yamamoto Y, Jeffery WR. Blind cavefish and heat shock protein chaperones: a novel role for hsp90alpha in lens apoptosis. Int J Dev Biol. 2004;48(8–9):731–8.

31. Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, et al. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat Genet. 2005;38(1):107–11.

32. Bilandžija H, Ma L, Parkhurst A, Jeffery WR. A potential benefit of albinism in Astyanax cavefish: downregulation of the oca2 gene increases tyrosine and catecholamine levels as an alternative to melanin synthesis. PLoS One. 2013;8(11):e80823.

33. King RA, Mentink MM, Oetting WS. Non-random distribution of missense mutations within the human tyrosinase gene in type I (tyrosinase-related) oculocutaneous albinism. Mol Biol Med. 1991;8(1):19–29.

34. Hutton SM, Spritz RA. Comprehensive analysis of oculocutaneous albinism among non-Hispanic caucasians shows that OCA1 is the most prevalent OCA type. J Invest Dermatol. 2008;128(10):2442–50.

35. Krauss J, Astrinides P, Frohnhöfer HG, Walderich B, Nüsslein-Volhard C. Transparent, a gene affecting stripe formation in Zebrafish, encodes the mitochondrial protein Mpv17 that is required for iridophore survival. Biol Open. 2013;2:703–10.

36. Frohnhöfer HG, Krauss J, Maischein HM, Nüsslein-Volhard C. Iridophores and their interactions with other chromatophores are required for stripe formation in zebrafish. Development. 2013;140(14):2997–3007.

37. Kondo S, Kuwahara Y, Kondo M, Naruse K, Mitani H, Wakamatsu Y, et al. The medaka rs-3 locus required for scale development encodes ectodysplasin-A receptor. Curr Biol. 2001;11(15):1202–6.

38. Pulkkinen L, Gerecke D, Christiano A, Wagman D, Burgeson R, Uitto J. Cloning of the beta 3 chain gene (LAMB3) of human laminin 5, a candidate gene in junctional epidermolysis bullosa. Genomics. 1995;25(1):192–8.

39. Mecklenbeck S, Hammami-Hauasli N, Höpfner B, Schumann H, Kramer A, Küster W, et al. Clustering of COL7A1 mutations in exon 73: implications for mutation analysis in dystrophic epidermolysis bullosa. J Invest Dermatol. 1999;112(3):398–400.

40. Muller M, Smolders JW, Meyer zum Gottesberge AM, Reuter A, Zwacka RM, Weiher H, et al. Loss of auditory function in transgenic Mpv17-deficient mice. Hear Res. 1997;114(1–2):259–63.

41. Dreyer B, Tranebjaerg L, Rosenberg T, Weston MD, Kimberling WJ, Nilssen O. Identification of novel USH2A mutations: implications for the structure of USH2A protein. Eur J Hum Genet. 2000;8(7):500–6.

42. Baux D, Larrieu L, Blanchet C, Hamel C, Ben Salah S, Vielle A, et al. Molecular and in silico analyses of the full-length isoform of usherin identify new pathogenic alleles in Usher type II patients. Hum Mutat. 2007;28(8):781–9.

43. Garcia-Garcia G, Aparisi MJ, Jaijo T, Rodrigo R, Leon AM, Avila-Fernandez A, et al. Mutational screening of the USH2A gene in Spanish USH patients reveals 23 novel pathogenic mutations. Orphanet J Rare Dis. 2011;6:65.

44. Van WE, Pennings R, te Brinke H, Claassen A, Yntema H, Hoefsloot L, et al. Identification of 51 novel exons of the Usher syndrome type 2A (USH2A) gene that encode multiple conserved functional domains and that are mutated in patients with Usher syndrome type II. Am J Hum Genet. 2004;74(4):738–44.

45. Bhattacharya G, Kalluri R, Orten D, Kimberling W, Cosgrove D. A domain-specific usherin/collagen IV interaction may be required for stable integration into the basement membrane superstructure. J Cell Sci. 2004;15(117):233–42.

46. Niemiller ML, Higgs DM, Soares D. Evidence for hearing loss in amblyopsid cavefishes. Bio Lett. 2013;9(3):20130104.

47. Yoshizawa M, Gorički Š, Soares D, Jeffery WR. Evolution of a behavioral shift mediated by superficial neuromasts helps cavefish find food in darkness. Curr Biol. 2010;20(18):1631–6.

48. You X, Bian C, Zan Q, Xu X, Liu X, Chen J, et al. Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. Nat Commun. 2014;5(5594):1–8.

49. Cavallari N, Frigato E, Vallone D, Frohlich N, Lopez-Olmeda JF, Foa A, et al. A blind circadian clock in cavefish reveals that opsins mediate peripheral clock photoreception. PLoS Bio. 2011;9(9):e1001142.

50. Duboue ER, Keene AC, Borowsky RL. Evolutionary convergence on sleep loss in cavefish populations. Curr Biol. 2011;21(8):671–6.

51. Lowrey PL, Takahashi JS. Genetics of circadian rhythms in mammalian model organisms. Adv Genet. 2011;74:175–230.

52. Riesch R, Tobler M, Plath M, Schlupp I. Offspring number in a livebearing fish (*Poecilia mexicana*, Poeciliidae): reduced fecundity and reduced plasticity in a population of cave mollies. Environ Biol Fish. 2009;84(1):89–94.

53. Pan X, Yang J, Chen X, Li Z. Broodstocks management, fecundity and the relationship between egg size and embryo survival ability of *Sinocyclocheilus grahami*. Zoological Res. 2011;32(2):196–203.

54. Adham IM, Eck TJ, Mierau K, Müller N, Sallam MA, Paprotta I, et al. Reduction of spermatogenesis but not fertility in Creb3l4-deficient mice. Mol Cell Biol. 2005;25(17):7657–64.

55. Yamamoto Y, Byerly MS, Jackman WR, Jeffery WR. Pleiotropic functions of embryonic sonic hedgehog expression link jaw and taste bud amplification with eye loss during cavefish evolution. Dev Biol. 2009;330(1):200–11.

56. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1(1):18.

57. Li R, Fan W, Tian G, Zhu H, He L, Cai J, et al. The sequence and de novo assembly of the giant panda genome. Nature. 2010;463(7279):311–7.

58. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: An improved ultrafast tool for short read alignment. Bioinformatics. 2009;25(15):1966–7.

59. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009;Chapter 4: Unit 4.10. doi:10.1002/0471250953.bi0410s25.

60. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110(1–4):462–7.

61. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.

62. Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS, Weinstock GM. Creating a honey bee consensus gene set. Genome Biol. 2007;8(1):R13.

63. Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, et al. The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun. 2013;4:2433.

64. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.

65. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31(1):46–53.

66. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307–21.

67. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol. 2009;537:113–37.

68. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17(8):754–5.

69. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.

70. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13(5):555–6.

71. Wikstrom N, Savolainen V, Chase MW. Evolution of the angiosperms: calibrating the family tree. Proc Biol Sci. 2001;268(1482):2211–20.

72. Wang ZQ. A new Permian gnetalean cone as fossil evidence for supporting current molecular phylogeny. Ann Bot. 2004;94(2):281–8.

73. Benton MJ, Donoghue PC. Paleontological evidence to date the tree of life. Mol Bio Evol. 2007;24(1):26–53.

74. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475(7357):493–6.

Yang *et al. BMC Biology* (2016) 14:1

Page 13 of 13

75. Graur D, Li WH. Fundamentals of molecular evolution. Sunderland: Sinauer Associates; 2000.

76. Philipp JK. Gnuplot in action: understanding data with graphs. New York: Manning Publications; 2009.

77. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 2006;22(10):1269–71.

78. Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. Estimating the tempo and mode of gene family evolution from comparative genomic data. Genome Res. 2005;15(8):1153–60.

79. Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. The evolution of mammalian gene families. PLoS One. 2006;1:e85.

80. Sun YB, Zhou WP, Liu HQ, Irwin DM, Shen YY, Zhang YP. Genome-wide scans for candidate genes involved in the aquatic adaptation of dolphins. Genome Biol Evol. 2013;5(1):130–9.

81. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 2005;22(12):2472–9.

82. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.