

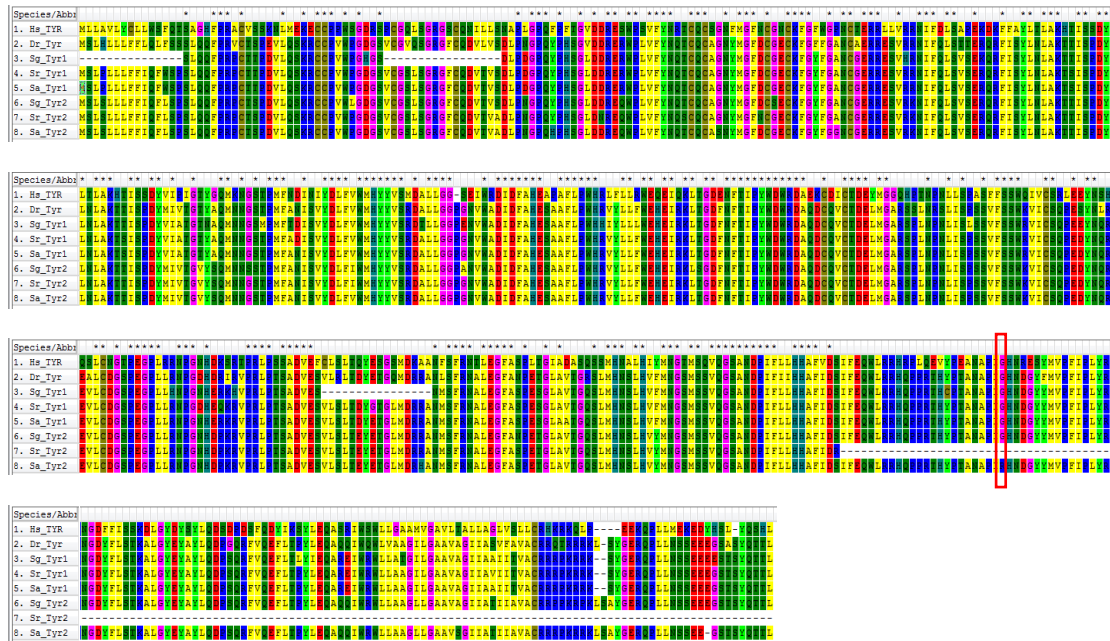


Figure S18. Tyr protein alignment. The Tyr protein from human (Hs), Dr, Sg, Sr, Sa were aligned by MUSCLE. Two copies of *Tyr* genes were found in *Sinocyclocheilus*, encoding proteins named as Tyr1 and Tyr2. Three small deletions were identified in Sg Tyr1, and the premature termination was occurred in Sr Tyr2. Besides that, a G420R mutation was identified in Sa Tyr2 (shown in red rectangle), which was also reported in the Caucasian patients. This mutation site located at exon IV, and previous studies reported that a cluster including this site could be an important functional domain in thse Tyr.

A

Human	EILEVFSGDLLRLHAQSHCRCPGSHPRVHPLAQRYPNDAGDTADNRVSR
Sg	EIEEVYTGQFPPLHTQSTCRCPASHPKVHPLVERYCIPNAANDTTHNKVLR
Sr	EIEEVYTGQFPPLHTQSTCRCPASHPKVHPLVERYCIPNAANDTTHNKVLR
Sa	EIEEVYTGQFPH-HTQSTCRCPASHPRVHPLVERYCIPNAANDTTHNKVLS

B

Human	FVNDNDVGTSWVSNVFTNITQLNQGVTTISVDLENGQYQ
Sg	YINDNDIGTTWISSIFPTLELLDKGITISIDLENGQYQ
Sr	YINDNDIGTTWISSIFPTLELLDKGITITIDLENGQYQ
Sa	YINDNDIGTTWNSSVFPTLELLDKGITITIDLENGQYQ

C

Human	NQCEPHRYNLTIDNFQH-CQMCECDSLGTLPGTICDPISGQCLCVPMRQGR
Sg	NRCSSHTYNLSSANATHGCQPCHDHMGTVPATVCNPVSGQCICLPTRYG
Sr	NRCSSHTYNLSSANATHGCQPCHDHMGTVPATVCNPVSGQCVCLPTRNG
Sa	NRCSSHTYNLSSANAMHGCQPCHDHMGTVPATVCDPVSGQCFCLPTRYD

D

Human	LRPPLVKGINSTTIHLRWFPPEELNGPSP
Sg	QSAPTLRPLNMTVMKVSWDAPAEELNGPPPL
Sr	LSAPTLRPLNMTVMKVSWDAPAEELNGPPPL
Sa	LRAPALRPLNMTVMKVSWDASAEELNGPPPL

Figure S19. Ush2a protein alignment. The Ush2a protein from human, Sg, Sr, Sa were aligned by MUSCLE. The mutations marked by dark colors are (A) R334S, (B) V382A, (C) the first termination codon, and (D) the second termination codon. Previous studies showed that the former two mutation sites may slightly affect splicing, while the existence of several termination codons indicate that Sa *Ush2a* may be a pseudogene.

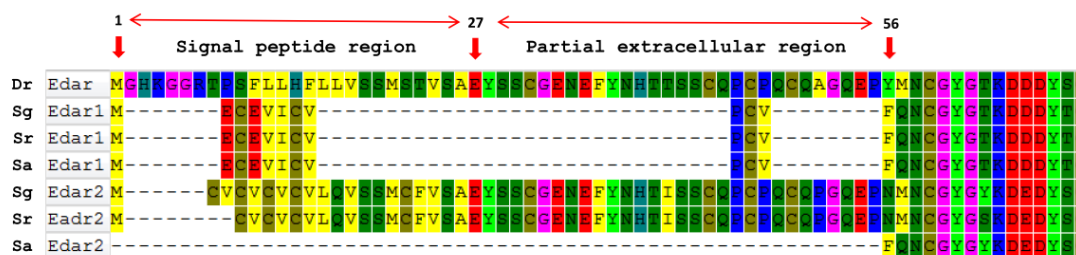


Figure S20. Deletions in Edar (Ectodysplasin-A receptor) of Sa. The Edar protein from Dr, Sg, Sr, Sa were aligned by MUSCLE. Compared with the *Edargene* in Dr, two copies of *Edargenes* were identified in three *Sinocyclocheilus* species, encoding proteins named Edar1 and Edar2 respectively. There were deletions in the signal peptide region and partial extracellular region in all three Edar1, which may lead to change or loss of function. The structures of Edar2 in Sg and Sr were relatively complete compared with Sa, while the signal peptide region and partial extracellular region were totally deleted in Edar2 of Sa, which may lead to function loss in guiding the Edar protein transfer across the membrane, and subsequently generate few scales at the skin surface of Sa.

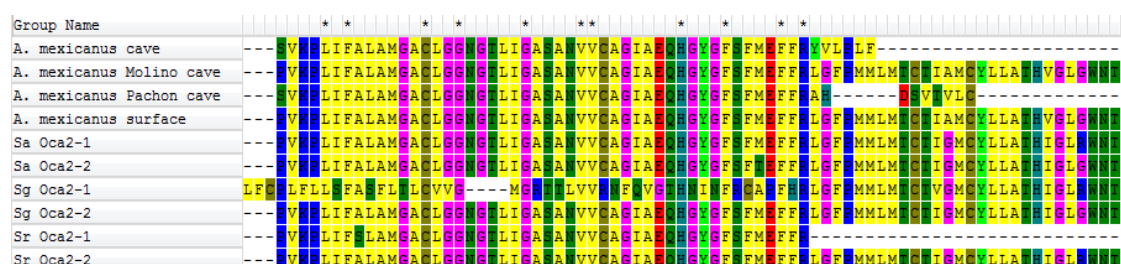


Figure S21. Comparison of Oca2 of A. mexicanus cavefish with the three Sinocyclocheilus species. The Oca2 protein from A. mexicanus, Sg, Sr, Sa were aligned by MUSCLE. The deletions of A. mexicanus cavefish Pachon population may produce function loss in pigmentation. No similar results were found in this population and Sa.

[illegible]

Figure S22. Mutation of Mpv17 protein in Sa. The Mpv17 protein from Dr, Sg, Sr, Sa were aligned by MUSCLE. The mutant and deleted sequence of Sa may lead to functional decline in Mpv17 protein, which may influence listening, or pigment formation.

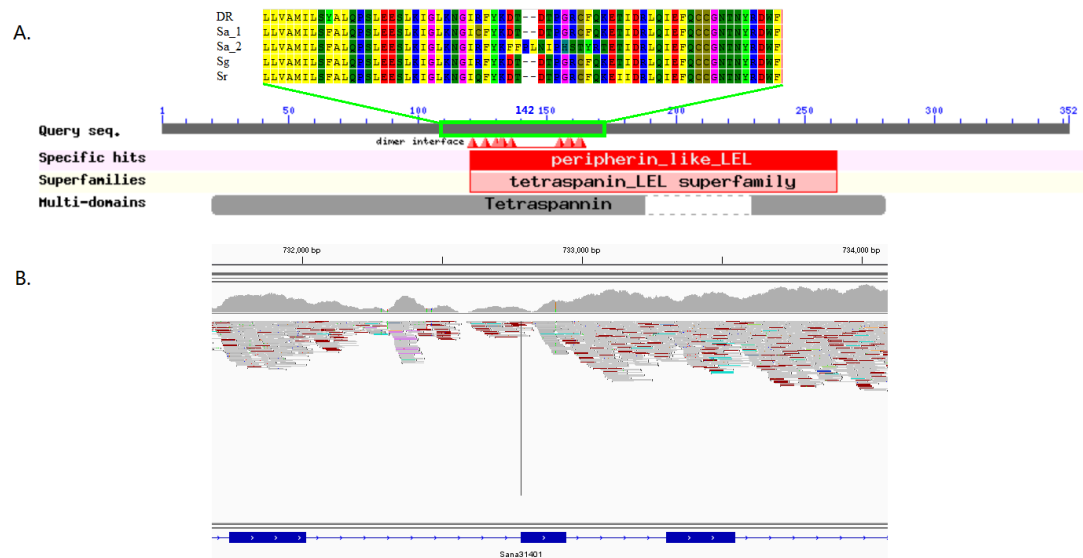


Figure S23. A 2-amino-acid insertion in the Sa Rom1b protein. (A) The structural variation in the Rom1b protein was compared with those in Dr and two *Sinocyclocheilus* species (Sg and Sr). The inserted site is located in Tetraspanin family conserved domain [NCBI accession: [pfam000335](https://www.ncbi.nlm.nih.gov/pfam/fam000335)]. (B) IGV (Integrative Genomics Viewer) shows the breaking point of sequencing depth nearby the insertion site.

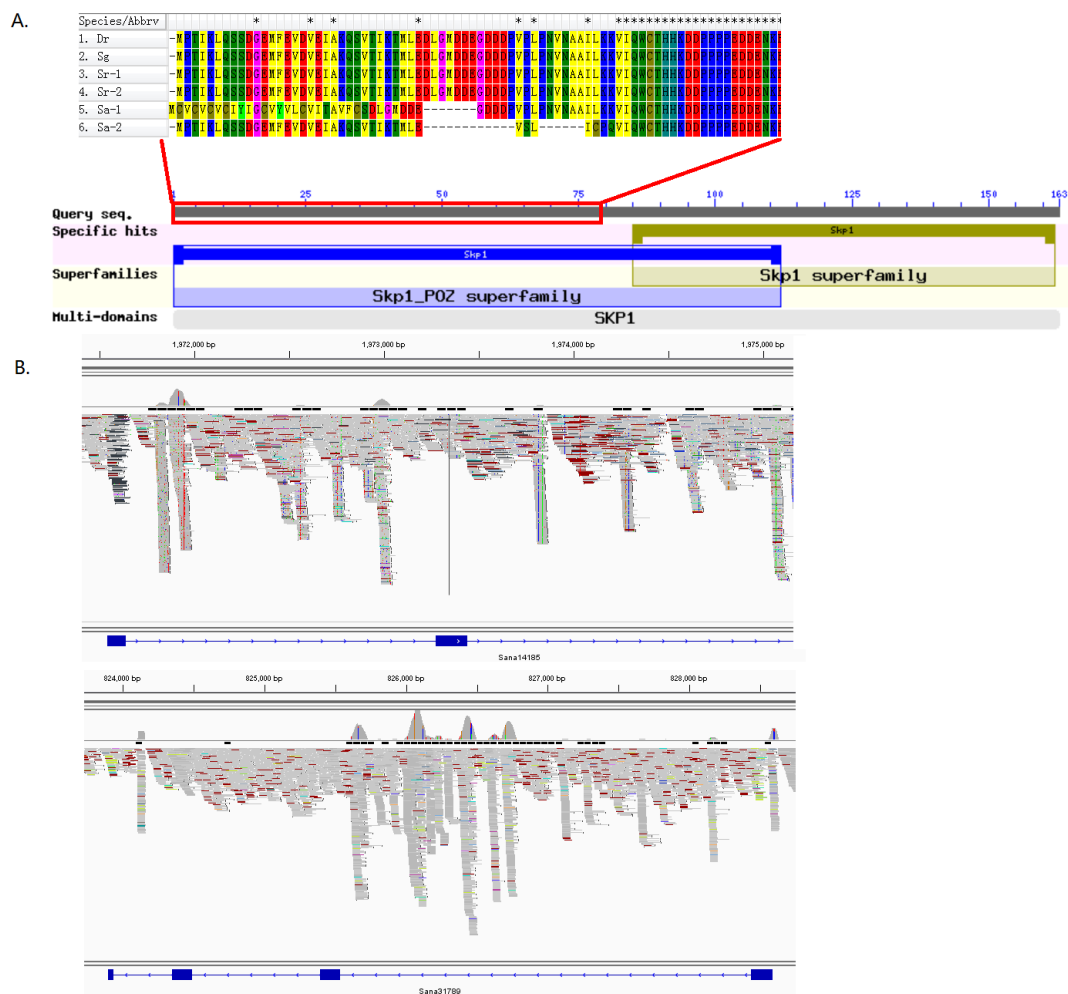


Figure S24. Structural variation in the Skp1 protein. (A) Comparison among Dr and the sequenced *Sinocyclocheilus* species reveals the Skp1 structural variation in Sa. The N-terminal deletions of Sa Skp1 proteins (Sa-1 and Sa-2) are located in the Skp1_Poz superfamily conserve domain. (B) IGV (Integrative Genomics Viewer) shows the breaking point of sequencing depth near the deletion sites.

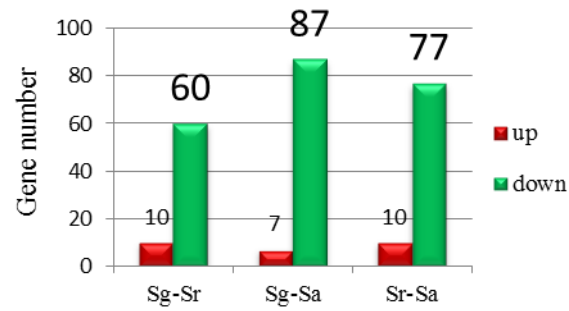


Figure S25A. Differential expression of eye-related genes in the three *Sinocyclocheilus* species among the 140 examined genes. All of the differential expression statuses were tested with FDR less than 0.01. The data demonstrate that the status of the transcriptional factors' expression declines as Sg>Sr>Sa. See more details in Table S26.

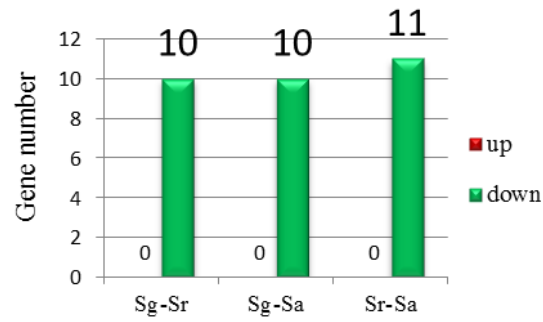


Figure S25B. Differential expression of eye development related transcriptional factors in the three *Sinocyclocheilus* species. Within the 12 analyzed transcriptional factors, compared with Sg, both Sr and Sa have 10 down-regulated ($FDR \leq 0.01$) factors; compared with Sr, Sa has 11 down-regulated ($FDR \leq 0.01$) factors. In general, the status of the transcriptional factors' expression declines as: $Sg > Sr > Sa$. See more details in Table S26.

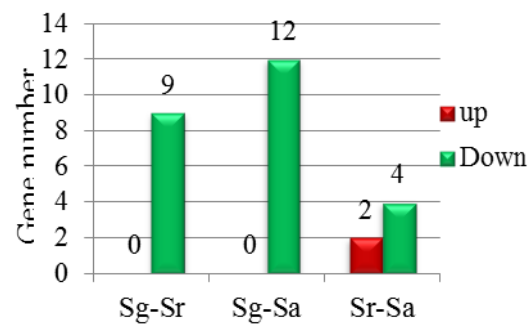


Figure S26. Differential expression of circadian rhythm pathway genes. Among the 17 examined genes, compared with Sg, Sr has 9 down-regulated genes, while Sa has 12 down-regulated genes. Compared with Sr, Sa has 4 down-regulated genes, but only 2 up-regulated genes. All of the differential expression statuses were tested with FDR less than 0.01. Generally speaking, the status of the rhythm pathway genes' expression declines as Sg>Sr>Sa, implying that the cavefish have no or low circadian rhythms.

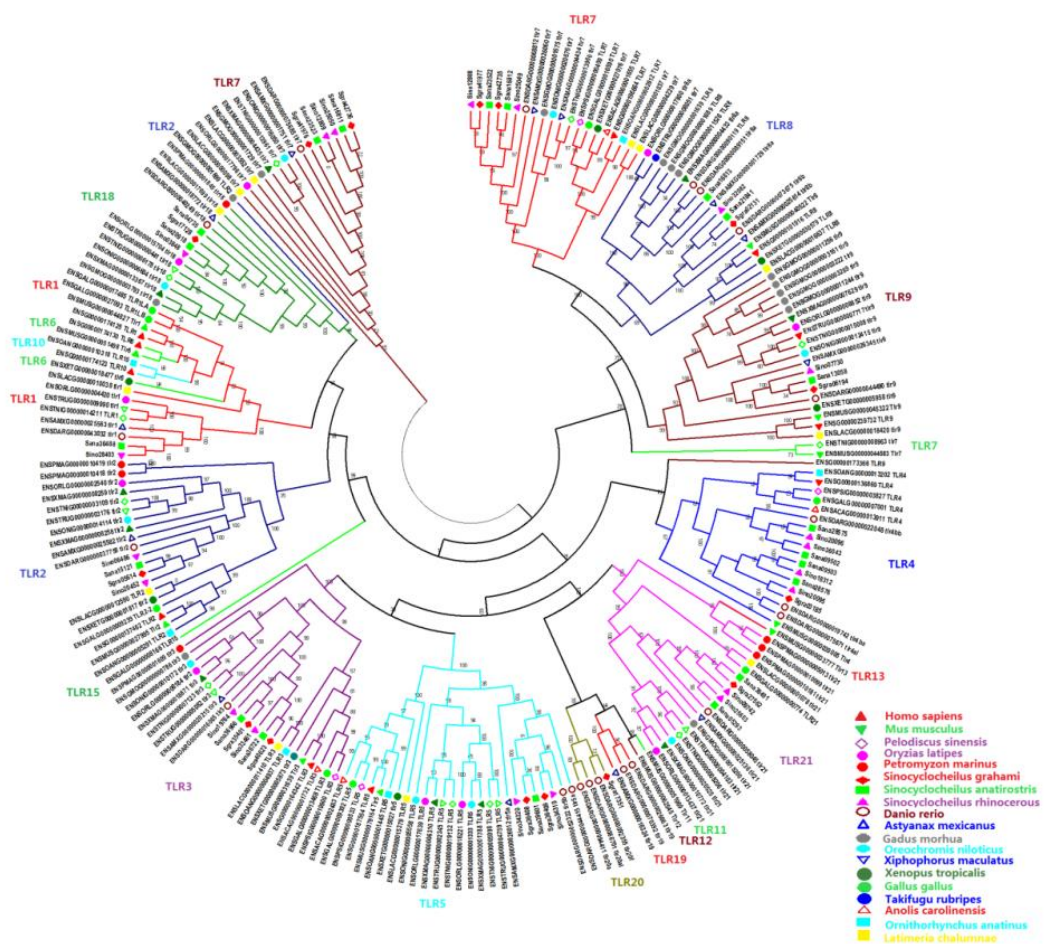


Figure S27. Similarity cluster of the *Tlr* gene family. The *Sinocyclocheilus Tlr* genes are closer to those of Dr. Each category is marked in a different color.

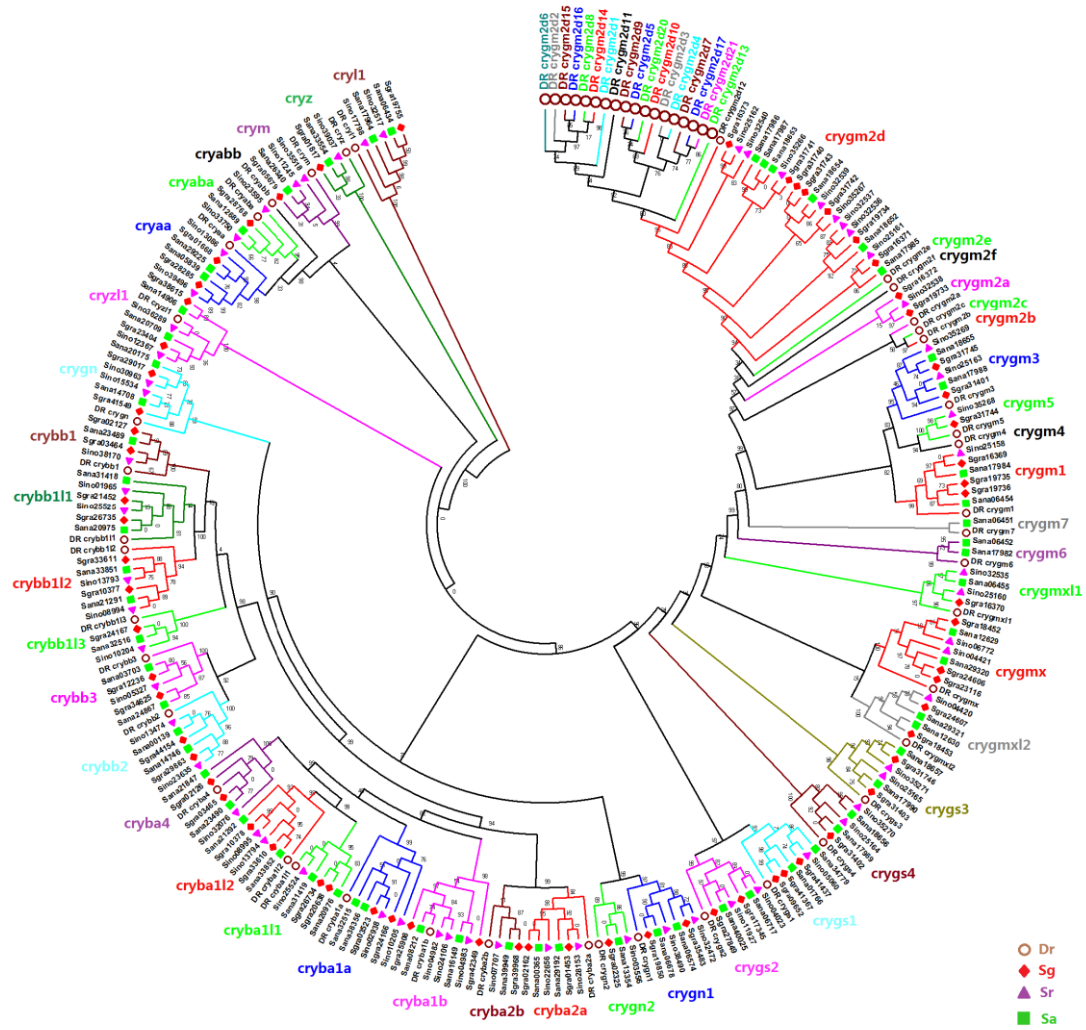


Figure S28. Similarity cluster of crystallins among Dr and the three *Sinocyclocheilus* species. The *Sinocyclocheilus* genes are classified referring to the Dr gene nomenclature. Each branch is marked by a different color. Among the 60 examined crystallin genes, 29 Dr genes are defined as outgroups compared to the *Sinocyclocheilus* genes, and 4 Dr genes cannot be distinguished from those in the *Sinocyclocheilus* species.