

SUPPLEMENTARY MATERIALS

Evolution of the eye transcriptome under constant darkness in

Sinocyclocheilus cavefish

Fanwei Meng^{1,2}, Ingo Braasch², Jennifer B. Phillips², Xiwen Lin¹, Tom Titus², Chunguang
Zhang^{1§} and John H. Postlethwait^{2§}

1 Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, P.R. China.

2 Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA

§ Corresponding author: John H. Postlethwait, E-mail: jpostle@uoneuro.uoregon.edu;

co-corresponding author: Chunguang Zhang E-mail: fish@ioz.ac.cn

Table S1: Summary of RNA-seq and *de novo* assembly of *Sinocyclotheilus*

	Number	Total nucleotides	N50	Mean size	N90
Reads	132,682,597	9,179,015,231	69	69	69
Contig of surface fish	156,118	82,820,563	534	531	332
Contig of cave fish	152,464	80,128,332	525	526	332
Merged contig	56,074	52,620,732	1,030	939	539

Table S3: Enriched Pathways and Diseases Identified by KOBAS

Pathways or Diseases	Database	Pathway Id	Sample number	Background number	P-Value
Down-regulated group (Enriched Pathway)					
Visual signal transduction: Cones	PID Curated	cone_pathway	22 / 211	31 / 1685	3.35E-11
Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction	PANTHER	P00028	27 / 185	47 / 1442	5.45E-11
Phototransduction	KEGG PATHWAY	hsa04744	25 / 404	42 / 3361	5.45E-11
Visual signal transduction: Rods	PID Curated	rhodopsin_pathway	20 / 211	32 / 1685	4.02E-09
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	Reactome	REACT:6305	26 / 315	81 / 2914	1.58E-05
Synaptic Transmission	Reactome	REACT:13685	42 / 315	194 / 2914	0.000393
Oxidative phosphorylation	KEGG PATHWAY	hsa00190	30 / 404	108 / 3361	0.000453
Parkinson's disease	KEGG PATHWAY	hsa05012	28 / 404	105 / 3361	0.001992
g-protein signaling through tubby proteins	PID BioCarta	100009	6 / 91	9 / 896	0.004251
Signaling by GPCR	Reactome	REACT:14797	37 / 315	187 / 2914	0.007512
Purine metabolism	KEGG PATHWAY	hsa00230	24 / 404	98 / 3361	0.011324
regulation of spermatogenesis by crem	PID BioCarta	100197	5 / 91	8 / 896	0.011986
chrebp regulation by carbohydrates and camp	PID BioCarta	100203	7 / 91	16 / 896	0.012695

Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	PANTHER	P00026	24 / 185	95 / 1442	0.012695
Pyrimidine catabolism	PID Reactome	500232	3 / 110	3 / 1236	0.017205
corticosteroids and cardioprotection	PID BioCarta	100156	7 / 91	17 / 896	0.017563
cystic fibrosis transmembrane conductance regulator (cftr) and beta 2 adrenergic receptor (b2ar) pathway	PID BioCarta	100205	7 / 91	17 / 896	0.017563
Muscarinic acetylcholine receptor 2 and 4 signaling pathway	PANTHER	P00043	16 / 185	56 / 1442	0.023385
Gastric acid secretion	KEGG PATHWAY	hsa04971	16 / 404	59 / 3361	0.023855
attenuation of gpcr signaling	PID BioCarta	100253	6 / 91	14 / 896	0.029359
Alzheimer's disease	KEGG PATHWAY	hsa05010	33 / 404	164 / 3361	0.031419
Olfactory transduction	KEGG PATHWAY	hsa04740	10 / 404	30 / 3361	0.035676
Down-regulated group (Enriched Diseases)					
Retinitis pigmentosa	FunDO	1705	18 / 281	35 / 2232	3.47E-06
Retinal disease	FunDO	2178	22 / 281	62 / 2232	0.000248
VISION	GAD		18 / 220	51 / 1707	0.002021
Retinitis punctata albescens	OMIM	None	8 / 219	13 / 1577	0.005431
cone-rod dystrophy	GAD		5 / 220	6 / 1707	0.007512
Macular dystrophy, patterned	OMIM	None	5 / 219	6 / 1577	0.008204

Leber optic atrophy	KEGG DISEASE	H00068	6 / 57	10 / 462	0.011324
Mitochondrial complex I deficiency	OMIM	None	5 / 219	7 / 1577	0.018464
Pheochromocytoma	OMIM	None	5 / 219	7 / 1577	0.018464
achromatopsia	GAD		3 / 220	3 / 1707	0.038897
paragangliomas, head and neck	GAD		3 / 220	3 / 1707	0.038897
Bradyopsia	OMIM	None	3 / 219	3 / 1577	0.043356
Pachyonychia congenita, Jackson-Lawler type	OMIM	None	3 / 219	3 / 1577	0.043356
Up-regulated group (Enriched Pathway)					
Intestinal immune network for IgA production	KEGG PATHWAY	hsa04672	3 / 157	7 / 3361	0.044873
Cytokine-cytokine receptor interaction	KEGG PATHWAY	hsa04060	10 / 157	48 / 3361	0.004537
Beta2 integrin cell surface interactions	PID Curated	integrin2_pathway	5 / 74	13 / 1685	0.006697
Pertussis	KEGG PATHWAY	hsa05133	7 / 157	40 / 3361	0.035581
Staphylococcus aureus infection	KEGG PATHWAY	hsa05150	7 / 157	12 / 3361	4.43E-05
Muscle contraction	Reactome	REACT:17044	15 / 147	55 / 2914	8.57E-06
glycolysis I	BioCyc	GLYCOLYSIS	5 (29)	23 / 562	0.044873
intrinsic prothrombin activation pathway	PID BioCarta	100122	3 (31)	6 / 896	0.017292
EphrinA-EPHA pathway	PID Curated	ephrina_ephapat hway	3 / 74	5 / 1685	0.017457
Antigen processing and presentation	KEGG PATHWAY	hsa04612	8 / 157	36 / 3361	0.008183

tyrosine degradation I	BioCyc	TYRFUMCAT-PWY	3 (29)	4 / 562	0.016436
Cell adhesion molecules (CAMs)	KEGG PATHWAY	hsa04514	10 / 157	76 / 3361	0.039406
classical complement pathway	PID BioCarta	100200	3 (31)	5 / 896	0.013245
Complement and coagulation cascades	KEGG PATHWAY	hsa04610	11 / 157	24 / 3361	1.16E-06
Common Pathway	PID Reactome	500303	3 / 49	5 / 1236	0.017297
Beta1 integrin cell surface interactions	PID Curated	integrin1_pathway	7 / 74	46 / 1685	0.044873
Hypertrophic cardiomyopathy (HCM)	KEGG PATHWAY	hsa05410	10 / 157	65 / 3361	0.017293
Wnt signaling pathway	PANTHER	P00057	16 / 63	188 / 1442	0.045768
Up-regulated group (Enriched Diseases)					
smoking cessation	GAD		21 / 116	104 / 1707	0.000282
Systemic infection	FunDO	2111	7 / 123	21 / 2232	0.004760
Immune system diseases	KEGG DISEASE	None	12 / 33	54 / 462	0.005563
REPRODUCTION	GAD		14 / 116	73 / 1707	0.009909
Cardiac diseases	KEGG DISEASE	None	7 / 33	24 / 462	0.017292
Classic complement pathway component defects	KEGG DISEASE	H00102	4 / 33	7 / 462	0.017292
pregnancy loss, recurrent	GAD		4 / 116	7 / 1707	0.017292
Cataract, pulverulent	OMIM	None	3 / 102	4 / 1577	0.019259
CHEMDEPENDENCY	GAD		22 / 116	164 / 1707	0.019259

HIV	GAD		4 / 116	8 / 1707	0.020697
antiphospholipid syndrome	GAD		2 / 116	2 / 1707	0.044873
C-reactive protein	GAD		2 / 116	2 / 1707	0.044873
Cataract, nuclear pulverulent	OMIM	None	2 / 102	2 / 1577	0.044873
cerebrovascular disease, ischemic	GAD		2 / 116	2 / 1707	0.044873
Holoprosencephaly-5	OMIM	None	2 / 102	2 / 1577	0.044873
Polymyositis	FunDO	2046	2 / 123	2 / 2232	0.044873
Retinopathy of Prematurity	GAD		2 / 116	2 / 1707	0.044873
Afibrinogenemia	KEGG DISEASE	H00222	2 / 33	2 / 462	0.045768
Tyrosinemia	KEGG DISEASE	H00165	2 / 33	2 / 462	0.045768
RENAL	GAD		9 / 116	50 / 1707	0.047967
Stroke	GAD		5 / 116	18 / 1707	0.049472
Alzheimer's disease	GAD		12 / 116	79 / 1707	0.049615
Circulatory system diseases	KEGG DISEASE	None	9 / 33	51 / 462	0.049669

Table S4: Real-time PCR primers used for the cDNA of both sepcies

Symbol	Sequence (5`-3`)
Atp5a1-F	CAAAGATGAACGACAACCTTCGG
Atp5a1-R	CGGATACCCTTGTAGAACAACCTC
Bactin1-F	GAAGATCAAGATCATTGCTCCC
Bactin1-R	ATGTCATCTTGTTTCGAGAGGT
Crx-a-F	GCGTCGGGAGCGCACTAC
Crxaa-F	TCGGGAGCGCACTACCTTC
Crxab-F	TCGGGAGCGCACTACTTTT
Crx-a-R	CGGCATTTAGCACGACGGT
Crybb1-F	CGCAGTGATTGTCTCATGTC
Crybb1-R	TCAAACAGATACTGGTAGCCTC
Crygs1-F	GGCTACCAGTATGTTCTGACC
Crygs1-R	GATCCCATCTAGCACCTTACAG
Hsp47-F	CAGCAAGAAGCACTACAACCTACG
Hsp47-R	TAGAAGCCGTAGATACCAGTGCG
Hsp90ab1-F	TACATGATGGCCAAGAAACAC
Hsp90ab1-R	CATCTTCATCAATACCCAAACCC
Lrp5-F	TTGGAGTTGCTGAGTGATCTG
Lrp5-R	ATAGTCTTGGTGATGACATCGG
Nr2e3-F	TTAAACCAGAAACACGAGGAC
Nr2e3-R	CAAAGTGAAGGGATGGAAGGA
Nrl-F	GCCTATACTGGTCGTTTCTTCC
Nrl-R	CGCAAATGTCTGTTTCAGCTC
Opn1mw4-F	GCTTGTTTGCACAGTCAAGG
Opn1mw4-R	ATTAAAGAAGATCCAGGCAGCA
Otx-5-F	TACTCGGTGAACGGCTTGA
Otx-5-R	GCTGCTGTTGACGGCACTT
Otx2-F	ATGTCGTATCTCAAGCAACCA
Otx2-R	CAAACCTGTACACGGGACTC
Pde5aa-F	CATCCCATCAGCAGATACCCTCT
Pde5aa-R	CTCTTTAGTTCCTTTGTCTGCAGC

Symbol	Sequence (5`-3`)
Rho-F	AACCCGTGCATCTACATCTG
Rho-R	CTTTGTGGTCTCTGTGTCGT
Rps11-F	CCAAGATGAAGATGCAGAGGA
Rps11-R	TTCAGGACGTTGAACCTCAC
Sox2-F	CGGAAATAAGCAAGCGACTC
Sox2-R	CTTGCCTTCTTCATGAGGGT
Wnt-2-F	TCATGAACTTGCACAACAACC
Wnt-2-R	TCTTAAACATCCTGTAGGCACTG

Table S5: Genbank/Ensembl accession numbers for *zic1* and *gpr85* nucleotide sequences

Species	Common name	Accession numbers	
		<i>zic1</i>	<i>gpr85</i>
<i>Homo sapiens</i>	Human	NM_003412	NM_001146265
<i>Gallus gallus</i>	Chicken	NM_204254	XM_416019
<i>Xenopus tropicalis</i>	Western clawed frog	ENSXETT00000033767	NM_001078822
<i>Lepisosteus osseus</i>	Longnose gar	EU001886	EU002141
<i>Amia calva</i>	Bowfin	EF032909	EF033026
<i>Hiodon alosoides</i>	Goldeye	EU001866	EU002120
<i>Oncorhynchus mykiss</i>	Rainbow trout	EF032911	EF033028
<i>Gadus morhua</i>	Atlantic cod	EU001871	EU002126
<i>Oryzias latipes</i>	Medaka	ENSORLT00000012607	ENSORLT00000013169
<i>Oreochromis niloticus</i>	Nile tilapia	EF032915	EF033032
<i>Gasterosteus aculeatus</i>	Three-spined stickleback	ENSGACT00000004841	ENSGACT00000025057
<i>Tetraodon nigroviridis</i>	Spotted green pufferfish	ENSTNIT00000015940	ENSTNIT00000021929
<i>Takifugu rubripes</i>	Fugu	ENSTRUT00000045674	ENSTRUT00000005315
<i>Ictalurus punctatus</i>	Channel catfish	EF032916	EF033033
<i>Danio rerio</i>	Zebrafish	NM_130933	NM_131499
<i>Sinocyclocheilus anophthalmus</i>	blind goldenline barbel	zic1a: JX403564	gpr85a: JX403568
		zic1b: JX403565	gpr85b: JX403569
<i>Sinocyclocheilus angustiporus</i>	small gill opening goldenline barbel	zic1a: JX403566	gpr85a: JX403570
		zic1b: JX403567	gpr85b: JX403571

Figure S1: Phylogeny of species and families referred to in this paper. (Adapted from (Tang et al. 2010; Wu et al. 2010; Near et al. 2012)).

Figure S2: Contig size distribution of *Sinocyclocheilus* transcriptomes.

Figure S3: Gene expression levels were detected by real-time PCR using RNA isolated from eyes of each species. The expression levels of *crybb1*, *hsp90ab1*, *lrp5*, *nrl*, *opn1mw4*, *otx5*, *rps11* and *wnt2* were quantified and normalized to *beta-actin1*. Relative expression values are mean \pm s.d. of at least three independent experiments.

Figure S4: Gene ontology (GO) ID representations for down-regulated genes (green) and up-regulated group (red). Three comparisons are shown: (A) cellular component ontology; (B) biological processed ontology; (C) molecular function ontology.

Figure S5: Collection sites for surface fish *Sinocyclocheilus angustiporus* (A) and cavefish *Sinocyclocheilus anophthalmus*. The inset is a map of China with a filled rectangle representing the area of the larger map. The red circle indicates the collection location of cavefish *Sinocyclocheilus anophthalmus* (N 25.05478°, E 103.37975°). Red triangles indicate the locations of surface fish *Sinocyclocheilus*

angustiporus. Surface fish were collected from Huangnihe River in Agang Town, Luoping (N 25.00905°, E 103.59256°). Longitude and latitude: 105E and 25N.

Figure S6: Average identities of paralogs or orthologs comparing surface fish, cave fish, zebrafish and human for 16 genes.

Reference:

- Near, TJ, RI Eytan, A Dornburg, KL Kuhn, JA Moore, MP Davis, PC Wainwright, M Friedman, WL Smith. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc Natl Acad Sci U S A* 109:13698-13703.
- Tang, KL, MK Agnew, MV Hirt, et al. 2010. Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). *Mol Phylogenet Evol* 57:189-214.
- Wu, X, L Wang, S Chen, R Zan, H Xiao, YP Zhang. 2010. The complete mitochondrial genomes of two species from *Sinocyclocheilus* (Cypriniformes: Cyprinidae) and a phylogenetic analysis within Cyprininae. *Mol Biol Rep* 37:2163-2171.

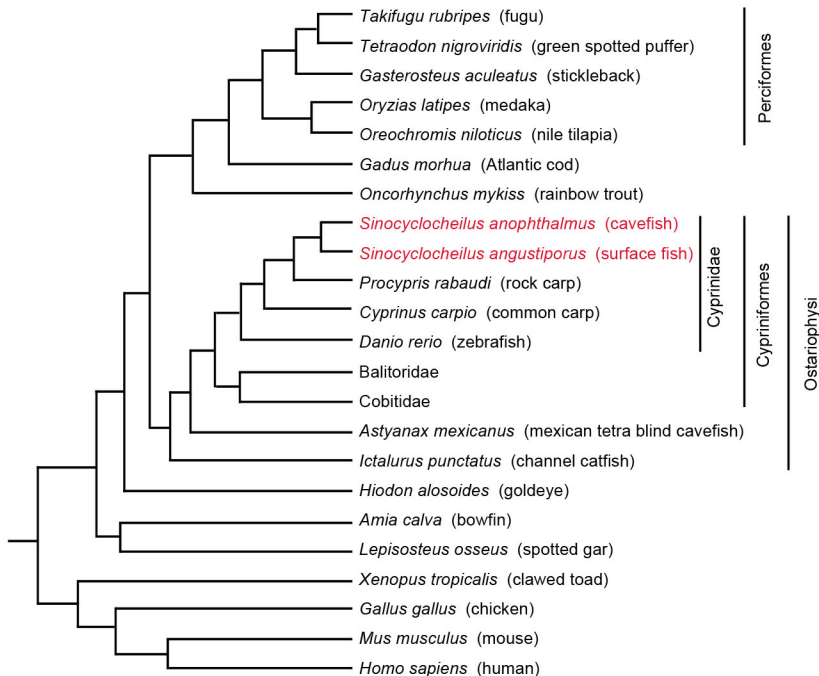


Figure S1

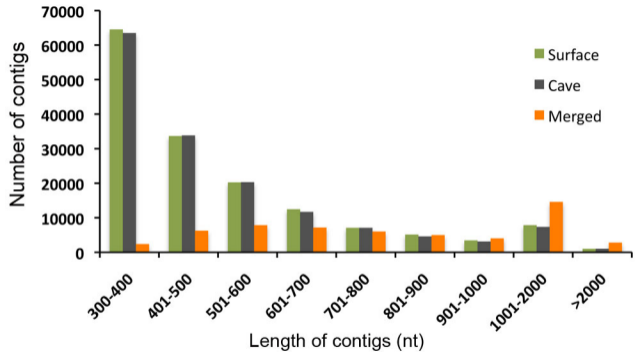
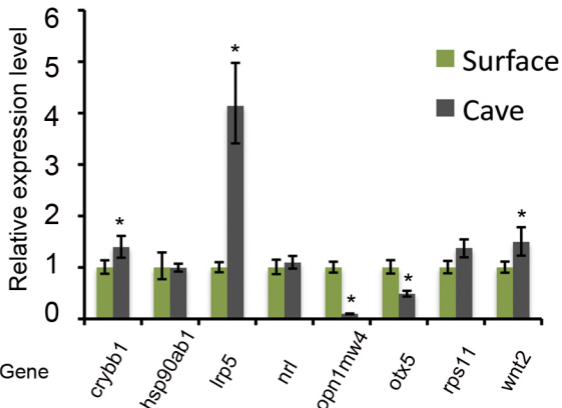


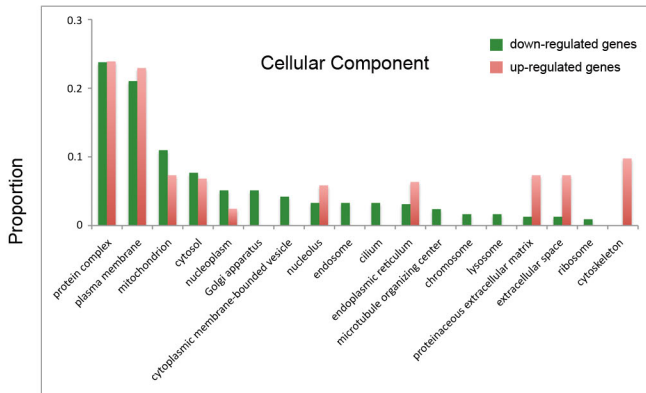
Figure S2



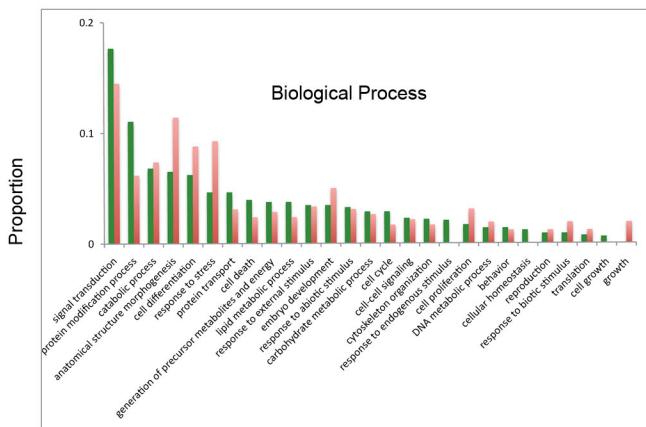
Real-time PCR	0.48	-0.01	2.05	0.13	-3.35	-1.04	0.46	-0.58
RNA-seq	1.15	0.58	0.86	0.32	-0.92	-1.17	0.35	-2.01
Fold change of Cave to Surface fish								

Figure S3

A



B



C

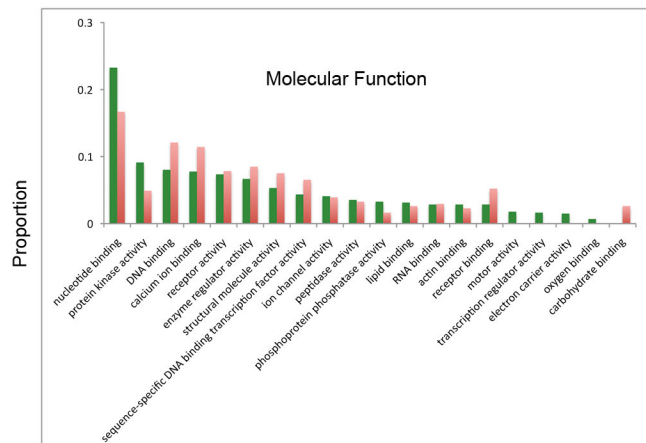


Figure S4

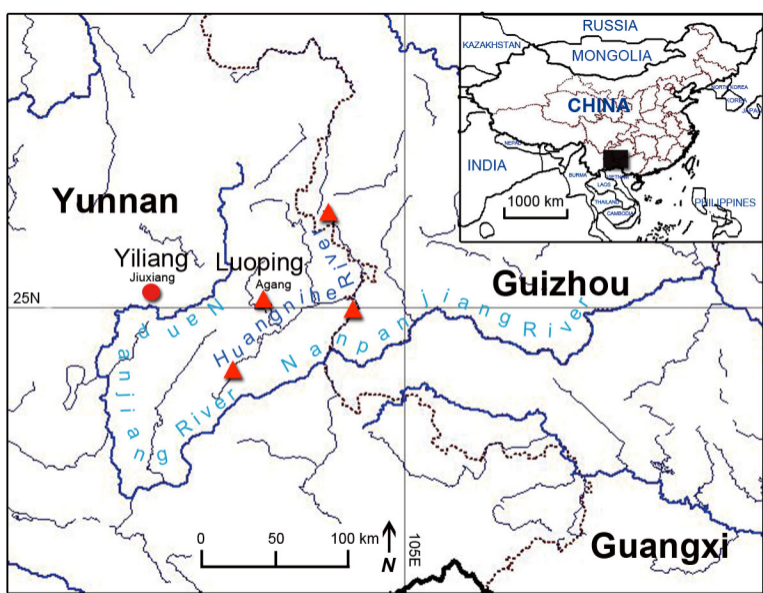


Figure S5

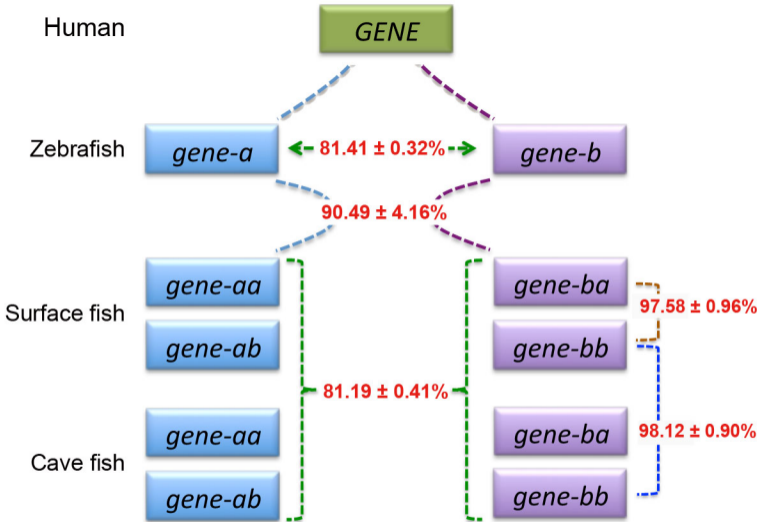


Figure S6