Supplementary Information

Relaxed selective constraints drove functional modifications in peripheral photoreception of the cavefish *P. andruzzii* and provide insight into the time of cave colonization

Luca Calderoni, Omar Rota-Stabelli, Elena Frigato, Alex Panziera, Sandra Kirchner, Nicholas S. Foulkes, Luise Kruckenhauser, Cristiano Bertolucci and Silvia Fuselli

Supplementary Methods

Opsin genes: PCR amplification and sequencing

A fragment of 1187 bp of the *opn4m2* gene spanning nucleotides from 40 to 1226 (CDS length: 1503, *P. andruzzii* GenBank accession number: GQ404489; *Danio rerio* Ensembl ID: ENSDARG00000007553), was obtained using the primers 40F (5'-CCTGGAGACATCAACTGCACCGCA-3', (Cavallari *et al.*, 2011)) and 1277R (5'- GCCTCACTCTCAGTCCAGCAAGAA-3'). The 50-µl reaction mixture contained 1.0 U of Platinum® Taq DNA polymerase High Fidelity (Invitrogen), 2 mM of $MgSO_4$, 0.08 mM each dNTP, 0.4 µM each primer and about 100 ng of genomic DNA. The PCR was conducted as follows: denaturation at 94 °C for 2 min; 30 cycles of 94 °C for 20 s, 53 °C for 30 s, and 68 °C for 1 min and 30 s. Both strands were Sanger sequenced using the PCR primers and two internal primers: 1033F (5'-TGGTCGCCCTATTCTGTGGTG-3') and 2R (5'-TGACTACAAGAGCGACTTTCGCCA-3'). Sequences were deposited in GenBank (accession numbers: KT953343-KT953348).

The entire *rho* gene (CDS: 1068 bp, *P. andruzzii* GenBank accession number: JQ413240.1; *Danio rerio* Ensembl ID: ENSDARG00000070666) was amplified with primers For: 5'-CATCCAACCGCAGCCATGA-3' and Rev: 5'-CGTCGCTGCTTATAGTGC-3'. The reaction mixture and the thermal cycling profile were as described above. Both strands were Sanger sequenced using the PCR primers and two internal primers, cfRHODfor 5'-TCACCATCGAGCACAAGAAG-3' (Tarttelin *et al.*, 2012) and cfRHODrev 5'-

GACCAAGAAGCCGATTACCA-3'. Sequences were deposited in GenBank (accession numbers: KU302709-KU302711)

Cytochrome b: PCR amplification and sequencing

The mitochondrial *cytochrome b* (*cytb*) gene was amplified and sequenced in *P. andruzzii* and *G. barreimiae* with primers Glu32-fwd 5'-GACTTGAAGAACCACCGTTGT-3' and Thr1-rev 5'-GATCTTCGGATTACAAGACCG-3'. The 25-µl reaction mixture contained 1.0 U of *Taq* DNA polymerase (Invitrogen), 2 mM of $MgCl_2$, 0.08 mM each dNTP, 0.4 µM each primer and about 50 ng of extracted DNA. The PCR was conducted as follows: denaturation at 94 °C for 2 min; 30 cycles of 94 °C for 20 s, 52 °C for 10 s, and 72 °C for 1 min and 30 s. Both strands were Sanger sequenced using the PCR primers. The nucleotide sequence of 903 bp from the ATG start codon was obtained and deposited in GenBank (KU926693 and KU738784).

Divergence time estimation between *P. andruzzii* and *G. barreimiae*

Divergence time between *P. andruzzii* and *G. barreimiae* was estimated based on *cytb* alignment using the Bayesian implementation Phylobayes 3.3f package (Lartillot *et al.*, 2009). We used a dating strategy involving multiple calibration points based on 20 fossil constraints obtained from the literature (Near *et al.*, 2012; Wang *et al.*, 2012), Supplementary Table 3) coupled with a root prior for the origin of Neopterygii of 360 mya and a permissive sd 20 mya (Near *et al.,* 2012). The clock was modeled with an Uncorrelated Relaxed Lognormal clock and the replacement process modeled with the CAT model (Lartillot and Philippe, 2004) which has been repeatedly shown to outperform GTR when dealing with ancient phylogenies and saturation events using mtDNA sequences (Bernt *et al.*, 2013). We used the Birth-Death process as tree prior, which was coupled with the use of soft bounds for the calibration allowing a 5% probability of sampling outside bonds (Supplementary table 5 analysis 1). Chains were run for 10000 generations and stopped when posterior divergence times did not change over chains progression as checked by readdiv program. The first 1000 generations were discarded as burn-in, and the chains were sampled every 10 generations. To check robustness of the date estimate, we repeated the analysis varying two key priors, the replacement model and the speciation prior. First we changed the replacement model by using a GTR model and a gamma distribution with four discrete categories instead of the CAT model (Supplementary table 5 analysis 2); second, we changed the tree prior by using a uniform distribution instead of the Birth-Death process, and hard bounds for the calibration instead of

soft bounds (Supplementary table 5 analysis 3). Phylobayes analyses were conducted on a taxonomically broad dataset with 20 calibration points. The dataset included the first 903 nucleotide positions of the *cytb* for 148 Teleost species (26 of which were cyprinids, including the specimens of *P. andruzzii* and *G. barreimiae* which were sequenced in this work) and an Amiid (*Amia calva*) as outgroup (Supplementary Table 4). As a starting tree we used a topology reflecting current state of the art (Ruber *et al*. 2007, Perea *et al.*, 2010; Near *et al.,* 2012; Wang *et al.,* 2012; Farashi *et al.*, 2014).

References

Bernt M, Braband A, Middendorf M, Misof B, Rota-Stabelli O, Stadler PF (2013). Bioinformatics methods for the comparative analysis of metazoan mitochondrial genome sequences. *Mol Phylogenet Evol* **69**: 320–7.

Cavallari N, Frigato E, Vallone D, Fröhlich N, Lopez-Olmeda JF, Foà A, *et al.* (2011). A blind circadian clock in cavefish reveals that opsins mediate peripheral clock photoreception. *PLoS Biol* **9**: e1001142.

Farashi A, Kaboli M, Rezaei HR, Naghavi MR, Rahimian H, Coad BW (2014). Reassessment of the taxonomic position of Iranocypris typhlops Bruun & Kaiser, 1944 (Actinopterygii, Cyprinidae). *Zookeys*: 69–77.

Lartillot N, Lepage T, Blanquart S (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286–8.

Lartillot N, Philippe H (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**: 1095–109.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, *et al.* (2012). Resolution of ray-finned fish phylogeny and timing of diversification. *Proc Natl Acad Sci* **109**: 13698–13703.

Perea S, Böhme M, Zupancic P, Freyhof J, Sanda R, Ozuluğ M, *et al.* (2010). Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. *BMC Evol Biol* **10**: 265.

Ruber L., Kottelat M., Hui Tan H., KL Ng P., Britz R. (2007). Evolution of miniaturization and the phylogenetic position of Paedocypris, comprising the world's smallest vertebrate. *BMC Evolutionary Biology* **7**:38.

Tarttelin EE, Frigato E, Bellingham J, Di Rosa V, Berti R, Foulkes NS, *et al.* (2012). Encephalic photoreception and phototactic response in the troglobiont Somalian blind cavefish Phreatichthys andruzzii. *J Exp Biol* **215**: 2898–2903.

Wang X, Gan X, Li J, Mayden RL, He S (2012). Cyprinid phylogeny based on Bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. *Sci China Life Sci* **55**: 761–73.