

Supplementary Materials

Evolutionary divergence of novel open reading frames in cichlids speciation

Shraddha Puntambekar¹, Rachel Newhouse², Jaime San Miguel Navas², Ruchi Chauhan², Grégoire Vernaz^{2,3,4}, Thomas Willis², Matthew T. Wayland⁵, Yagnesh Urmania⁶, Eric A. Miska^{2,6,4}, and Sudhakaran Prabakaran^{1,2,7*}

¹Department of Biology, Indian Institute of Science Education and Research, Pune, Maharashtra, 411008, India

²Department of Genetics, University of Cambridge, Downing Site, CB2 3EH, UK

³The Wellcome Trust/CRUK Gurdon Institute, University of Cambridge, Cambridge, CB2 1QN, UK

⁴Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK

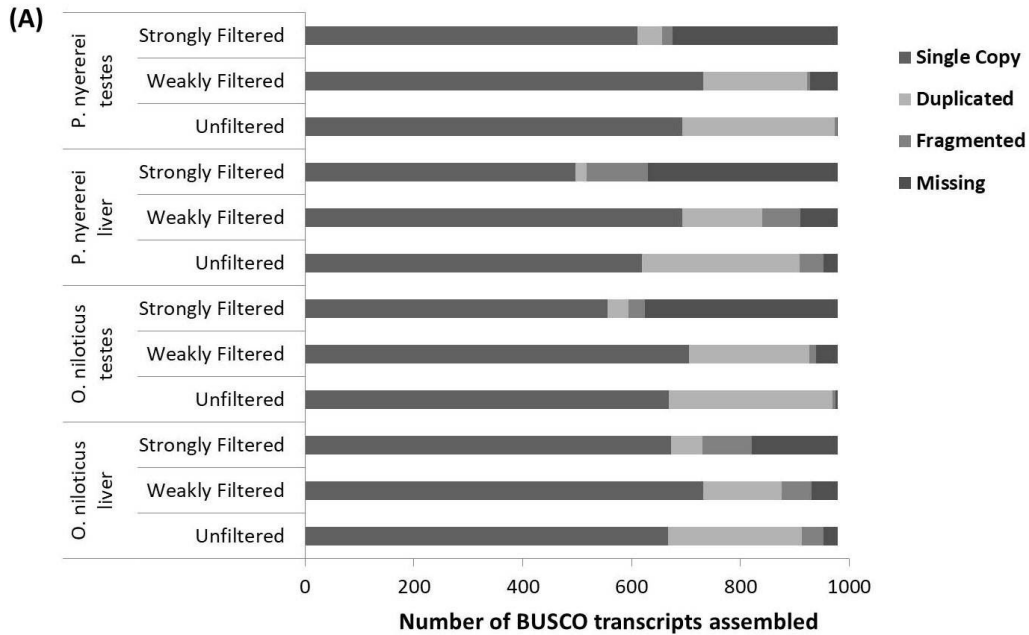
⁵Department of Zoology, University of Cambridge, Downing Site, CB2 3EH, UK

⁶Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QR, United Kingdom

⁷St Edmund's College, University of Cambridge, CB3 0BN, UK

*Corresponding author, email: sp339@cam.ac.uk.

Supplementary Figure 1



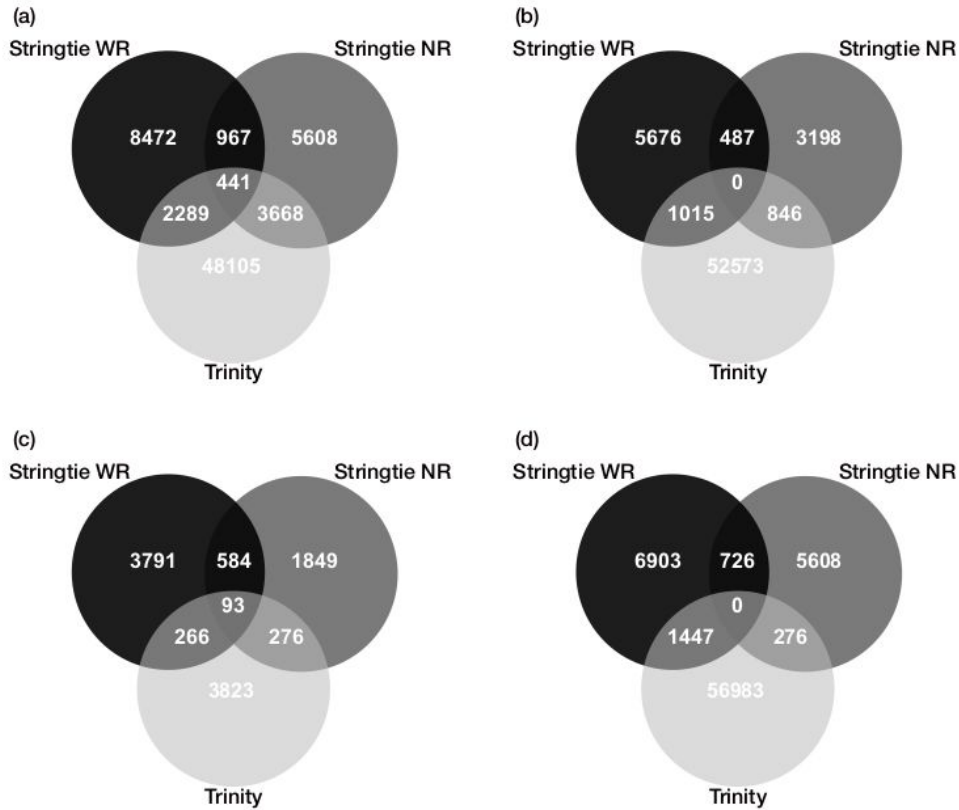
(B)

	<i>O. niloticus</i> liver	<i>O. niloticus</i> testes	<i>P. nyererei</i> liver	<i>P. nyererei</i> liver
Unfiltered	0.159	0.052	0.147	0.054
Weakly filtered	0.297	0.162	0.292	0.196
Strongly filtered	0.324	0.202	0.316	0.236

SF. 1. Processing assembled transcripts.

- A. The number of BUSCO metazoan transcripts present in the unfiltered and filtered Trinity transcriptomes. Weakly filtered: transcripts with a Transrate score of 0.01 or lower removed. Strongly filtered: transcript removal threshold set to optimise the overall assembly Transrate score. Dark gray: single copy. Light gray: duplicated. :fragmented. Black: missing.
- B. The effects of filtering on the whole assembly Transrate scores for each Trinity transcriptome. Weakly filtered: transcripts with a Transrate score of 0.01 or lower removed. Strongly filtered: transcript removal threshold set to optimise the overall assembly Transrate score.

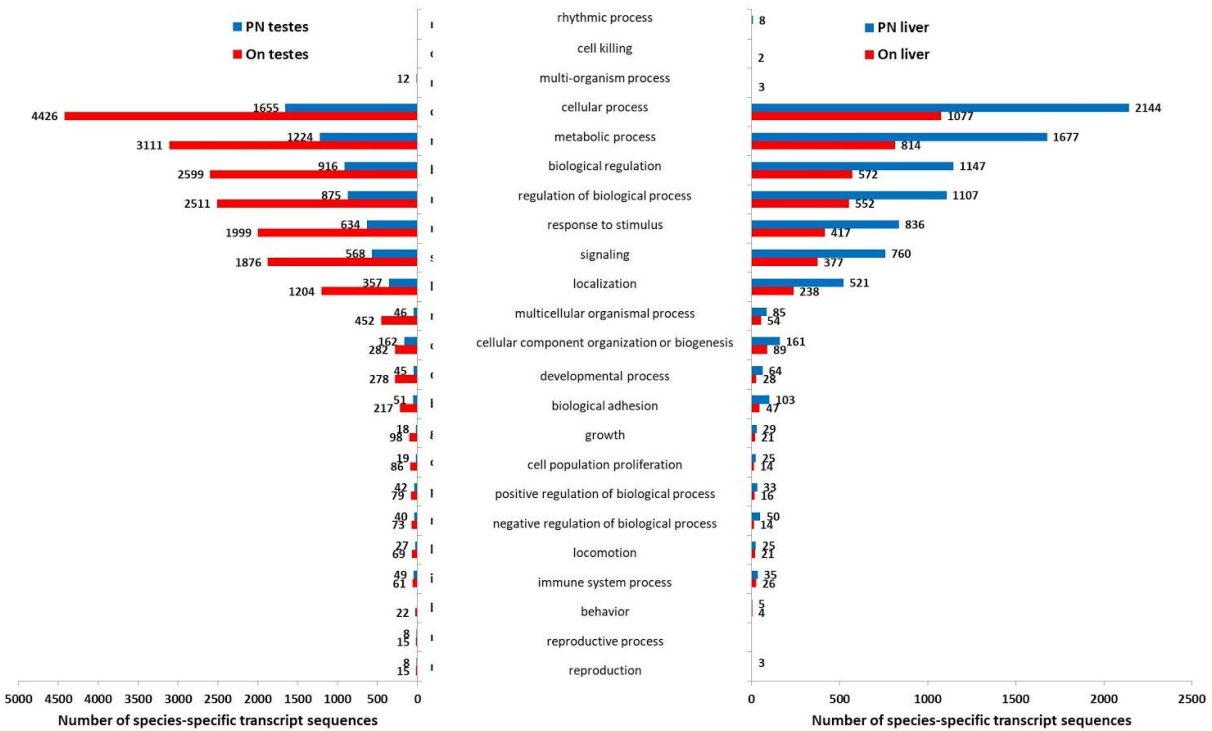
Supplementary Figure 2



SF 2. The overlap in species-specific transcripts identified using each transcriptome assembly method. Species-specific transcripts were identified as those without a match of at least 80% at the nucleotide level in the equivalent transcriptome in the opposing species. The transcripts identified by each method were compared using GFFcompare.

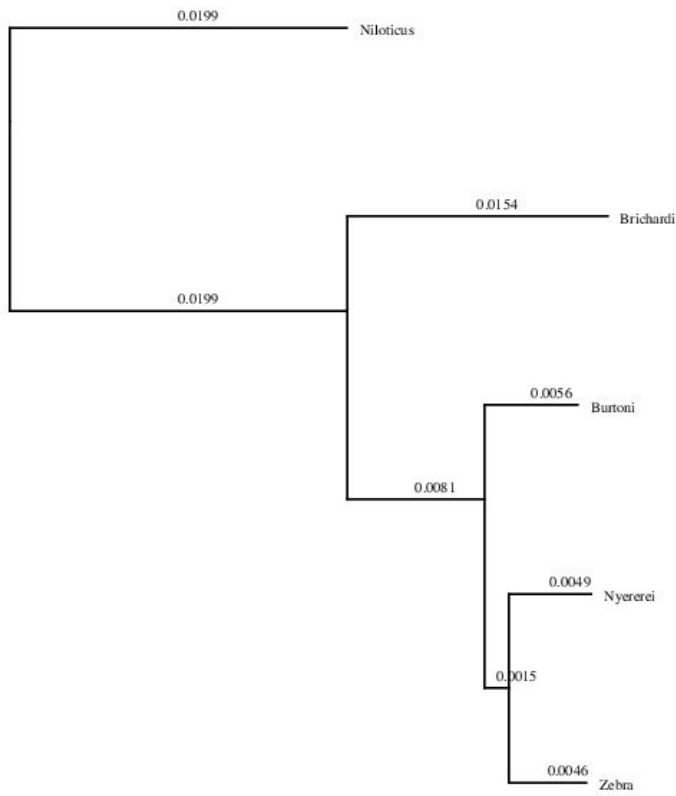
(a) *O. niloticus* testes (b) *O. niloticus* liver (c) *P. nyererei* testes (d) *P. nyererei* liver

Supplementary Figure 3: Functional annotation analysis of species-specific transcripts.



SF. 3: Functional annotation analysis of species-specific transcripts. The Level 2 Biological Process GO Annotations of Species-Specific transcripts for each species and tissue. The union of the species-specific transcripts identified using each transcriptome assembly method was annotated using InterProScan

Supplementary Figure 4



SF. 4 Phylogenetic tree constructed over four-fold degenerate sites from the alignment of five cichlids genome. The numbers on the edge represent the neutral species divergence calculated by phyloFit.