

Supplementary data

The genomic basis of adaptation to the deep water ‘twilight zone’ in Lake Malawi cichlid fishes

Christoph Hahn^{1,2*}, Martin J Genner³, George F Turner⁴, Domino A Joyce¹.

1 Evolutionary and Environmental Genomics Group (@EvoHull), School of Environmental Sciences, University of Hull, Hull HU5 7RX, UK.

2 Institute of Zoology, University of Graz, A-8010 Graz, Austria.

3 School of Biological Sciences, University of Bristol, Bristol Life Sciences Building, 24 Tyndall Avenue, Bristol. BS8 1TQ. UK.

4 School of Biological Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, Wales, UK.

** corresponding author*

Supplementary Data including scripts and jupyter notebooks detailing all analyses are deposited in a dedicated Github repository:

https://github.com/HullUni-bioinformatics/Diplotaxodon_twilight_RAD

Supplementary figures

Supplementary figures are also deposited on Github:

https://github.com/HullUni-bioinformatics/Diplotaxodon_twilight_RAD/tree/master/SUPPLEMENTARY_DATA/SUPPLEMENTARY_FIGURES

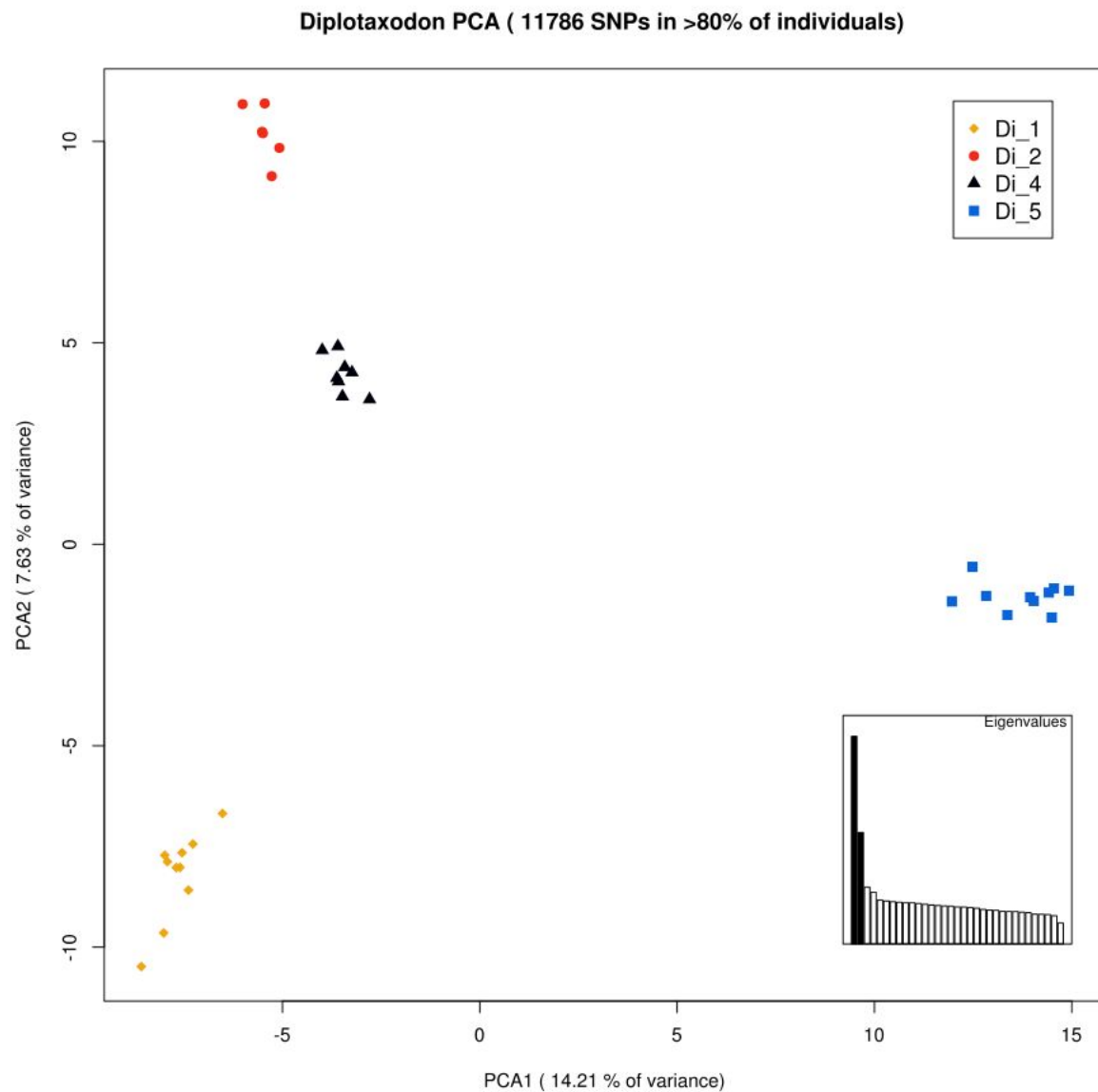


Figure S1. **Principal component analysis based on 11,786 SNPs.** *Diplotaxodon* species. Yellow - *D.* 'macrops black dorsal'; red - *D.* 'limnothrissa black pelvic'; black - *D.* 'macrops offshore'; blue - *D.* 'macrops ngulube'.

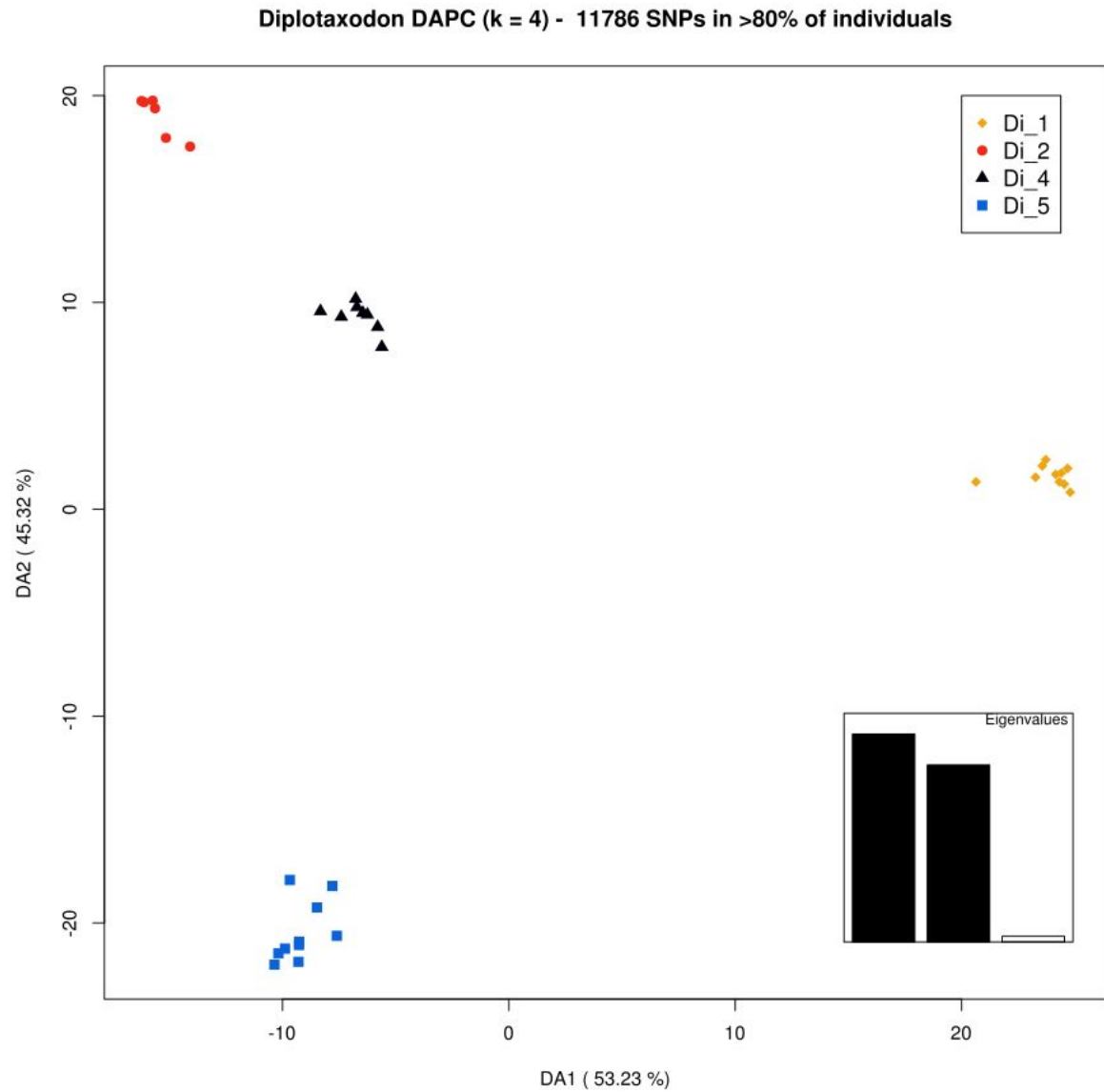


Figure S2. **DAPC based on 11,786 SNPs.** *Diplotaxodon* species. Yellow - *D.* 'macrops black dorsal'; red - *D.* 'limnothrissa black pelvic'; black - *D.* 'macrops offshore'; blue - *D.* 'macrops ngulube'.

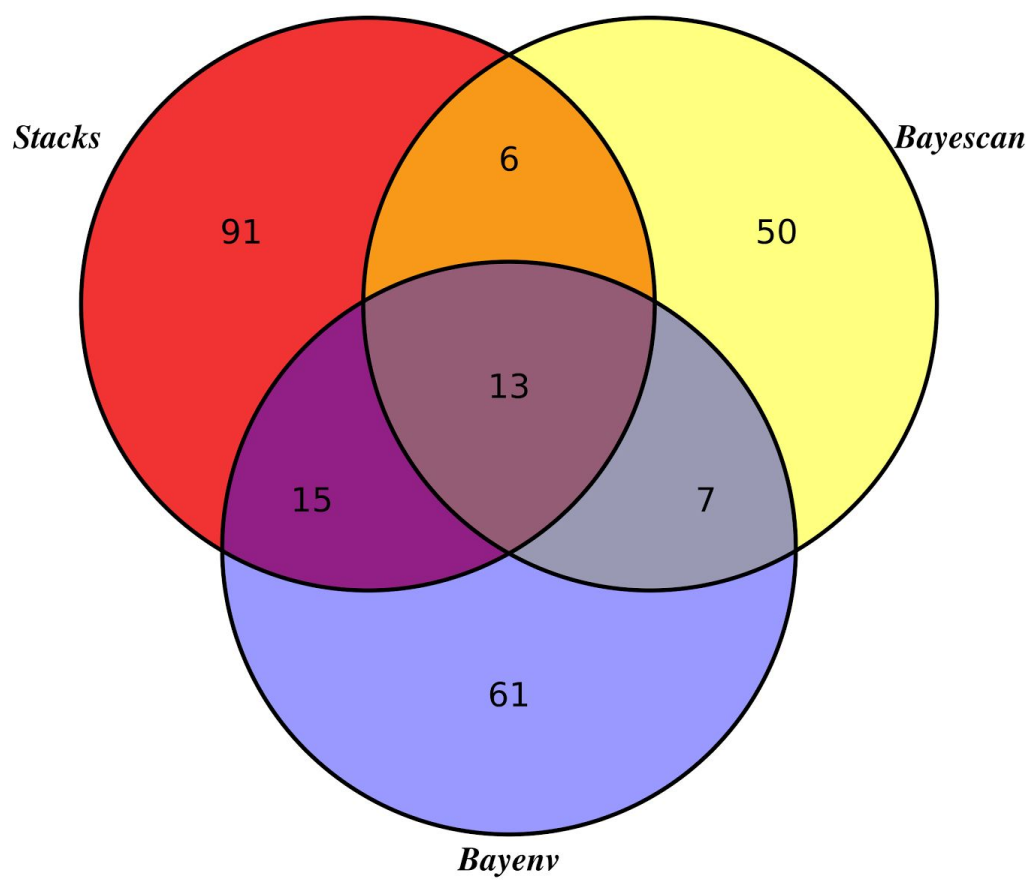


Figure S3. Number and concordance of candidate outlier loci highlighted by three independent approaches.

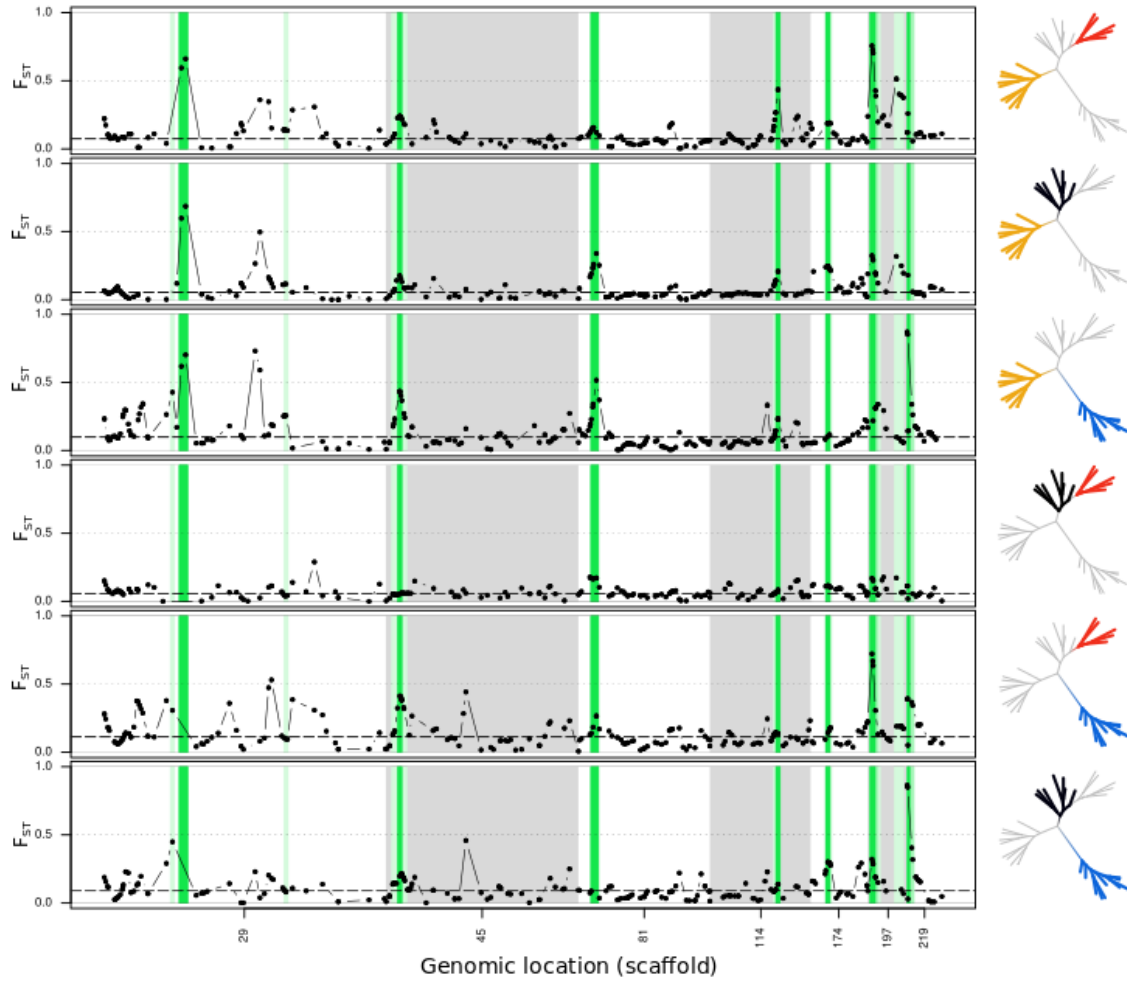


Figure S4. Pairwise F_{ST} divergence at genomic regions highlighted by three independent outlier approaches. Displayed are only the six scaffolds containing a minimum of 5 SNPs - scaffold id on the x axis. Putative candidate regions are highlighted in shades of green. Dark green regions indicate support by three approaches (see tables 1 and S3 for summary of gene complements in highlighted regions). Dots represent the kernel smoothed averages across 50kb windows. Dashed lines indicate the genome wide F_{ST} average. Population pairs are indicated by the highlighted regions in the phylogenetic trees on the right hand side, *Diplotaxodon* species. Yellow - *D.* 'macrops black dorsal'; red - *D.* 'limnothrissa black pelvic'; black - *D.* 'macrops offshore'; blue - *D.* 'macrops ngulube'.

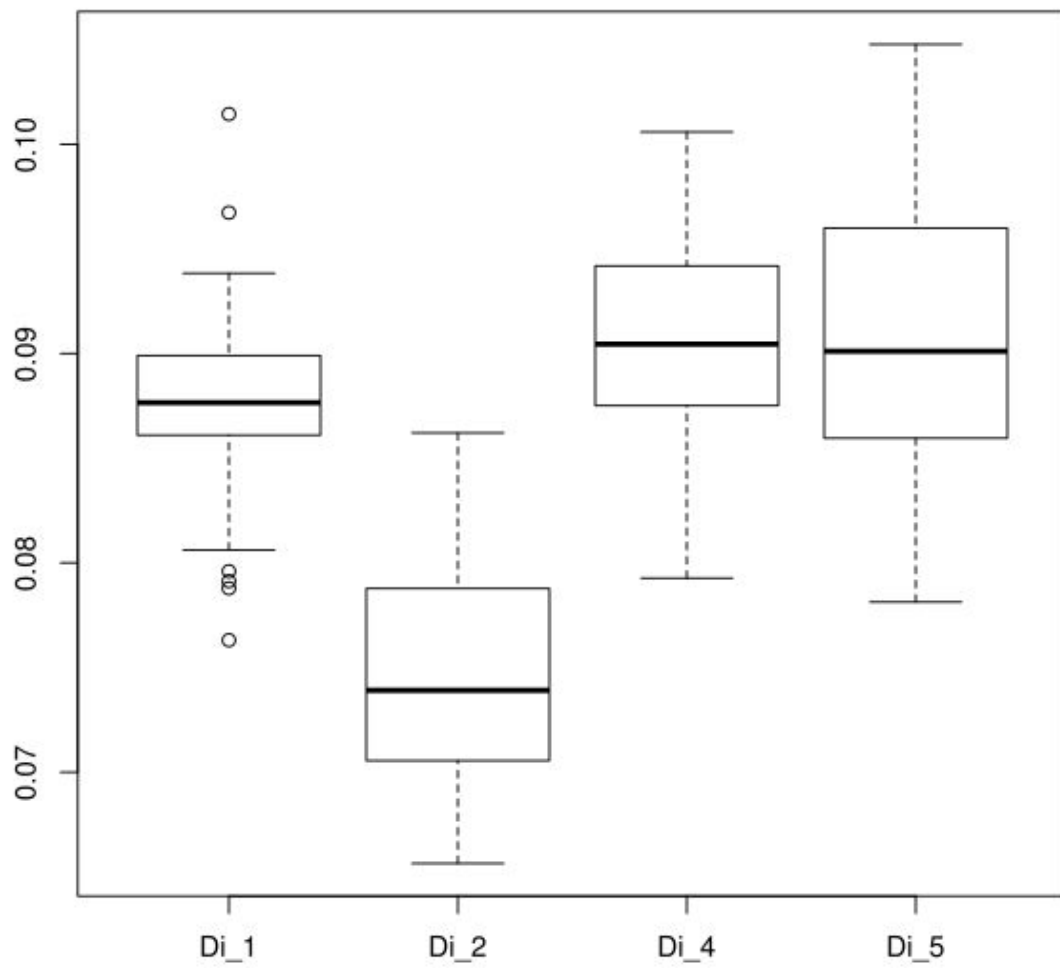


Figure S5. *Diptaxodon* interspecific vertical eye diameter variation. Measurements were normalized by total length of the fish.

Supplementary tables

Supplementary tables are deposited on Github:

https://github.com/HullUni-bioinformatics/Diplotaxodon_twilight_RAD/tree/master/SUPPLEMENTARY_DATA/SUPPLEMENTARY_TABLES

Table S1. **Sample metadata.**

Table S2. **Pairwise global Fst.**

Table S3. **Candidate genes in genomic regions showing signals of selection.** Genomic location (scaffold id), putative functional annotation of genes (if available), original gene identifier and RADtag ids localized in candidate regions under selection. Numbers in parenthesis after the RADtag id indicate whether a given RADtag was considered an outlier by the three independent approaches (1 - outlier; 0 - no outlier) in the following order: Stacks, Bayescan, Bayenv. The number after the 'I' symbol describes the ARR for the RADtag, followed by the smoothed ARR level of significance (assessed only for ARR \geq 0.95; NS: $p > 0.05$; *: $p = 0.05-0.005$; **: $p = 0.005-0.001$; ***: $p < 0.001$), inferred by the eye morphology informed analysis.

Table S4. **Pairwise TukeyHSD for eye size.** TukeyHSD pairwise comparison p-values. Vertical eye diameter - above diagonal; horizontal eye diameter - below diagonal.

Table S5. **Candidate genes in genomic regions strongly associated with eye size differences.** Genomic location (scaffold id), putative functional annotation of genes (if available), original gene identifier and RADtag ids localized in candidate genomic regions associated with eye size differences. Number in parenthesis after the RADtag id describes the ARR for the RADtag, followed by the smoothed ARR level of significance (assessed only for ARR \geq 0.95; NS: $p > 0.05$; *: $p = 0.05-0.005$; **: $p = 0.005-0.001$; ***: $p < 0.001$), inferred by the eye morphology informed analysis.

Table S6. **Detailed summary of polymorphic sites and amino acid changes in candidate genes.**

Further supplementary files are deposited on Github:

https://github.com/HullUni-bioinformatics/Diplotaxodon_twilight_RAD/tree/master/SUPPLEMENTARY_DATA/MISC