

Characterisation of transcriptomes of galaxiid fish

Introduction

[Some stuff about why Galaxiids are great]

Rationale of this study - *G. brev* as a high quality reference transcriptome that can be used to underpin future projects. *G. dep* as a test case, lower coverage sequencing to examine genome-wide dNdS and isolate polymorphic microsatellites.

Materials and methods

Before David

- Sample collection
- RNA extraction/normalisation
- Library prep
- Sequencing
- Assembly

David

- Annotation
- Identification of orthologs
- Identification of microsatellites
- Genome-wide dNdS

Results

Sequencing and assembly

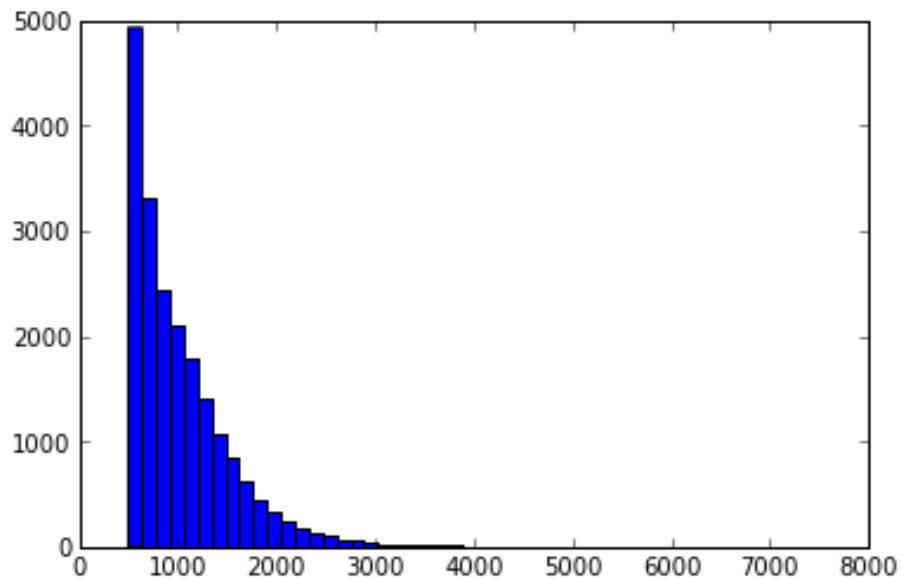
Sequencing summary (needs fastq files)

- Number of reads
- Average read length
- Average quality scores

Assembly summary

- G.brev
 - 20197 contigs > 700 bp

- Median of 24 reads per contig
- Median length of analysed contigs 886 bp
- G.dep
 - 11955 total contigs
 - Median 3 reads pre contig (mean = 10)
 - Median length of contigs 239



Plots of lengths of each contig assembly

Annotation

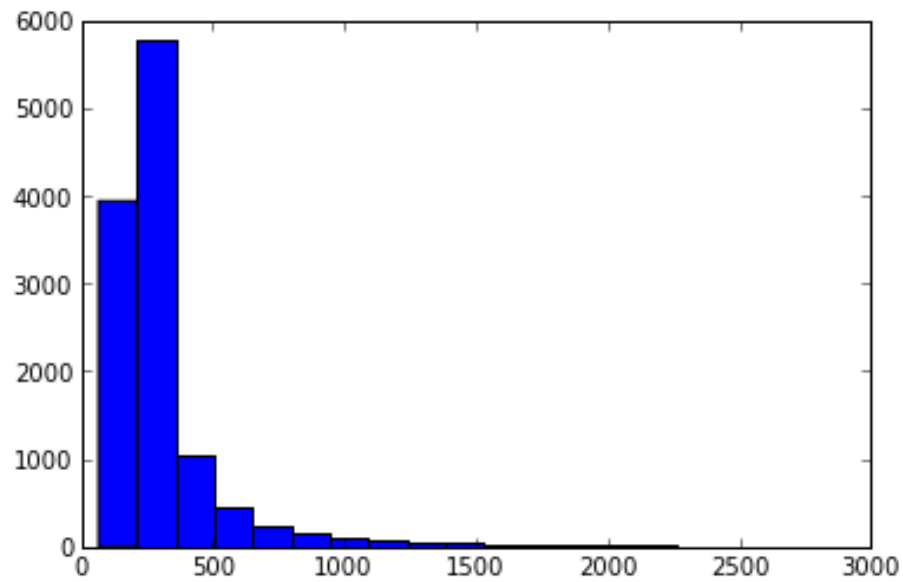
Rationale: Have only performed functional annotation on *G. brev* as it's the highest quality sequence, and we only use *G. dep* sequences that are presumed orthologs of *G. brev*. We also identified ORFs and microsatellites from both datasets.

Results: Have Blasted against NR and Swissprot.

NR shows most hits coming from sensible taxonomic places (i.e. they are galaxiid genes, not bacteria/algae/whatever else)

7525/21000 *G.brev* contigs have hits against Swissprot. 3723 for *dep*.

From these we harvested many many GO terms [summary](#)



Using **orffinder** and BLASTs against chichlid, salmon, trout and zebrafish abgenomes. Using these data we found 8727 ORFs in *G. brev* and 4435 in *G. dep*

summary

species Total contigs swissprot NR ORFs

Gbrev 21000 Gdep 11000

Identification of Orthologues

Genome-wide analysis of selection