**Methods**

**Transcriptome sequencing and analysis**

A single individual of *Oreobates cruralis* Boulenger, 1902 (TaxonID:448617) collected in XX/XX/XXXX was used for transcriptome sequencing. Samples of three tissues (intestine, liver and spleen) were preserved in RNA Later and whole RNA was extracted using the RNeasy Protect Mini Kit (Qiagen). RNA quality was evaluated with RNA ScreenTape on TapeStation by Agilent, with RIN scores of 6.2, 7.3 and 7.1 respectively. Sequencing libraries were prepared by the SNP&SEQ Technology Platform at Uppsala University from 1μg total RNA using the TruSeq stranded mRNA library preparation kit (Cat# RS-122-2101/2102, Illumina Inc.) and including poly-A selection. The library preparation was performed according to the manufacturers’ protocol (#15031047). The quality of the libraries was evaluated using the Agilent Technologies TapeStation and a DNA 1000-kit Screen Tape. The adapter-ligated fragments were quantified by qPCR using the Library quantification kit for Illumina (KAPA Biosystems) on a StepOnePlus instrument (Applied Biosystems/Life technologies) prior to cluster generation and sequencing. A 14 pM solution of RNA was subjected to cluster generation and paired-end sequencing with 125 bp read length on the HiSeq2500 (Illumina Inc.) using the v4 chemistry according to the manufacturer’s protocols.

Raw transcriptomic data was cleaned following Singhal (2013). In brief, raw fastq reads were filtered using SORTMERNA (Kopylova *et al.* 2012) to remove possible ribosomal RNA (rRNA) contamination, and TRIMMOMATIC (Bolger *et al.* 2014) to perform low quality trimming and adapter removal. Cleaned data of each tissue was normalized using TRINITY *in silico* *normalization* (Grabherr *et al.* 2011) to remove large excess of reads corresponding to moderately and highly expressed transcripts. The resulting data was merged and normalized again prior to assembly. De-novo transcriptome assembly of reads was done using TRINITY (Grabherr *et al.* 2011) and annotated with TRINOTATE (<https://trinotate.github.io/>). A total of 422,999 putative genes were obtained after assembly. Of those, 45,885 transcripts had gene ontology (GO) significant matches. The annotated transcript is available on DRYAD: XXX.

**Sequence capture probe design**

The annotated set of transcripts was filtered prior to bait design following Portik *et al.* (2016). In brief, we first kept transcripts with a GC content between 40% and 60% (40,362 transcripts). Then, we kept transcripts that are 500-850bp in length (21,396). Transcripts were trimmed to 850bp if they were longer. Finally, we used REPEATMASKER (<http://www.repeatmasker.org/>) to remove repetitive elements or low complexity regions (we used ‘vertebrata metazoa’ as database and ‘cross\_match’ as search engine). This resulted in a set of 18,127 transcripts, ranging from 500-850bp, of which 17,879 were unique (based on sequence) with a total length of 14,368,067bp. A FASTA file of the final marker set is available on: XXX. The final filtered dataset was used to design a MYbaits-11 custom bait library (MYcroarray) with 213,879 unique probe sequences. We selected the longer bait length (120nt) to maximizing capture of divergent sequences. The probe design is available as a FASTA file on: XXX.

**Results**

**Transcriptome**

Illumina transcriptome sequencing of *O. cruralis* produced a total of 522,877,358 raw reads (intestine: 193,693,696; liver: 189,463,370; spleen: 139,720,292). Among those, 81.47% were kept after low quality trimming and adapter removal (426,003,462). The number of reads was further reduced to 6.97% after *in silico* *normalization* prior to assembly (36,428,858). A total of 550,871 transcripts were obtained after *de-novo* transcriptome assembly, corresponding to 422,999 putative genes. Transcripts sequences had a N50 length of 731bp and a mean length of 543bp. The assembly included a total of 299,133,111bp.

**Annotation**

|  |  |
| --- | --- |
| **Database** | **Unigenes with significant match** |
| sprot\_Top\_BLASTX\_hit | 50163 (11.86%) |
| sprot\_Top\_BLASTP\_hit | 27481 (6.50%) |
| RNAMMER | 0 |
| Pfam | 23500 (5.56%) |
| SignalP | 0 |
| TmHMM | 5474 (1.29%) |
| eggnog | 37349 (8.83%) |
| Kegg | 38120 (9.01%) |
| gene\_ontology\_blast | 46293 (10.94%) |
| gene\_ontology\_pfam | 15225 (3.60%) |

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, **29**, 644–52.

Kopylova E, Noé L, Touzet H (2012) SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.

Portik DM, Smith LL, Bi K (2016) An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources*, **16**, 1069–1083.

Singhal S (2013) De novo transcriptomic analyses for non-model organisms: An evaluation of methods across a multi-species data set. *Molecular Ecology Resources*, **13**, 403–416.