Deep Sequencing of Micro-RNAs From the Domestic Turkey (*Meleagris gallopavo)*.

**Abstract:**

**Introduction**

Micro-RNAs (miRNAs) are 22-24nt long non-coding sequences that are increasingly acknowledged to play an important role in regulating gene expression in plants and animals (ref). MiRNAs can be predicted in novel genomes by comparative analysis with well-studied genomes or by computational prediction; eg by looking at the folding of RNA hairpins. More recently, deep sequencing and mapping against databases of known miRNA sequences has proved a useful tool in helping with the annotation of miRNAs from novel genomes.

Turkeys are close relatives of the domestic chicken (estimates of divergence time range from 25 ([Griffin, Robertson et al. 2008](#_ENREF_2)) to 40 ([van Tuinen and Dyke 2004](#_ENREF_4)) mya), and as such are an excellent system for comparative genome analysis. A preliminary analysis of predicted miRNAs in the turkey genome paper found no important differences between the sequences of chicken and turkey miRNAs on the basis of sequence similarity ([Dalloul, Long et al. 2010](#_ENREF_1)). Our analysis extends this investigation of the turkey miRNA complement by performing deep sequencing to confirm the expression of these predicted turkey miRNAs in a cultured cell line.

**Results and Discussion**

**Useful information:**

* Total sequenced reads = 1754082
* Total clipped reads = 1679163
* Total unique reads mapping to turkey precursors = 937442
* Number of reads not mapping to turkey precursors = 741721
* Number of unique sequences (collapsed reads) =
* Number of unmapped reads which map/do not map to genome (if they do not map to the genome, look at their length distribution) = 49879 (6.80%)
* Number of turkey precursors that have a chicken mirBase miRNA mapping = 200
* Number of turkey precursors that have a mirBase miRNA mapping = 249
* Number of predicted turkey precursors that have no mirBase mature miRNA mapping = 180 (where do these come from then??)
* Number of precursors with both reads and a chicken miRNA mapping to the same location:
* Number of precursors with both reads and a mirBase miRNA mapping to the same location:
* Number of turkey precursors with both a chicken 5p and 3p miRNA mapping = 7
* Number of turkey precursors with both a chicken miRNA and miRNA\* mapping = 30
* Number of precursors with isoMirs.
* Counts mapping to each id’d chicken miRNA.
* Compare number of unique sequences with number of reads mapping to known miRNAs (assess diversity of the different classes of read).

*Comparison with chicken (Gallus gallus) micro-RNAs*

467 chicken mature miRNAs in MirBase

77 miRNA\*

11 miRNA precursors annotated with 3p and 5p mature sequences

18226 mature miRNAs total in version of MirBase used (inc miRNA\*s)

As the turkey is closely related to the domestic chicken, a species for which miRNAs have been well-characterised, we were interested in identifying any expressed turkey miRNAs which were homologous to chicken miRNAs. We mapped all mirBase miRNAs to the predicted turkey miRNA precursors from Ensembl, again using Novoalign.

This yielded turkey 249 precursors with a mirBase miRNA mapping to them; 200 of which had chicken miRNAs mapping to them. This somewhat lower estimate than the number of predicted turkey miRNAs reported in the turkey genome paper. One explanation for this might be that not all miRNAs are expressed in all cell-types.

36 precursors had more than one miRNA from chicken mapping to them.

*Novel micro-RNA prediction*

The approximately 750,000 reads that failed to map to a turkey precursor were used as input to MirDeep2 (ref), an algorithm for predicting novel miRNAs. Because most turkey miRNAs were identical to chicken mature sequences, the chicken mature sequences from mirBase were used as a file of known miRNAs. MirDeep2 predicted novel miRNAs from these reads; on closer examination most of these overlapped known miRNAs from either zebrafinch or human and are already annotated in the turkey genome as putative miRNAs. It is possible that we have only sequenced relatively highly expressed miRNAs in this experiment and that sequencing with higher coverage would reveal some novel miRNAs with lower expression. Many of the reads failed to map to the genome entirely; it is likely that these are fragments of degraded mRNAs.

*Molecular evolution of miRNA stem-loops*

**Methods**

Illumina deep sequencing was performed on a small RNA library prepared from a single macrophage-derived turkey cell-line (IAH30), yielding approximately 1.8 million 36 base pair, single-end reads. Adapter sequences and primer-dimer were removed from the raw reads using the CutAdapt package ([Martin 2011](#_ENREF_3)) and reads shorter than 18nt were discarded, as mature miRNAs tend to be in the size range of 20-24nt (ref).

Reads were initially mapped to the set of 429 predicted turkey miRNA precursors downloaded using Ensembl BioMart (<http://www.ensembl.org/biomart/martview>) with Novoalign (Novocraft Technologies). Approximately 900,000 reads mapped to predicted turkey miRNAs; reads which did not map were used as input to miRNA prediction software. MiRNA loci that were expressed in the turkey sample were defined as turkey stem-loops to which both reads and a known chicken miRNA mapped.

Reads that did not map to a turkey precursor were input to MirDeep2 to try and predict novel miRNAs, using chicken mature MirBase miRNAs as the sample of known miRNAs.

**References**

Dalloul, R. A., J. A. Long, et al. (2010). "Multi-Platform Next-Generation Sequencing of the Domestic Turkey (<italic>Meleagris gallopavo</italic>): Genome Assembly and Analysis." PLoS Biol **8**(9): e1000475.

<p>The combined application of next-generation sequencing platforms has provided an economical approach to unlocking the potential of the turkey genome.</p>

Griffin, D., L. Robertson, et al. (2008). "Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution." BMC Genomics **9**(1): 168.

Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads." EMBnet.journal **17**(1).

van Tuinen, M. and G. J. Dyke (2004). "Calibration of galliform molecular clocks using multiple fossils and genetic partitions." Molecular Phylogenetics and Evolution **30**(1): 74-86.