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

* Title
* Authors
* Affiliations

**Abstract:**

**Introduction**

Micro-RNAs (miRNAs) are 22-24nt long single-stranded non-coding RNAs that are increasingly acknowledged to play an important role in regulating gene expression in both plants and animals (ref). Mature miRNAs are 22-24 nucleotides long and are excised from longer stem-loop sequences. Within the mature miRNA sequence, a shorter “seed” sequence ~7nt in length regulates gene expression by binding to the 3’UTR of a target transcript. A single miRNA can target many transcripts, so one might expect that there is strong selection for their sequences to remain the same. This is indeed the case, with miRNA sequences showing a high level of conservation across all metazoa, suggesting that miRNA genes tend to be gained but not lost during evolution (Wheeler et al., 2009).

* Expansions in different metazoan lineages
* Mature sequences highly conserved – intense selection on seed regions
* Gain >> loss

The miRNA repertoire of a newly-sequenced genome can be analysed by comparative analysis with well-studied genomes or by computational prediction; for example by looking at the folding of RNA hairpins (references), however such analyses are not informative about differences in miRNA expression. Recently, deep sequencing has proved a useful tool in helping with the annotation of miRNAs from novel genomes and for investigating differences in miRNA expression between tissues or experimental conditions.

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)) million years ago), and as such are an excellent system for comparative genome analysis. A preliminary analysis of miRNAs in the turkey genome found that most known chicken miRNAs are also present in turkey, with few sequence differences ([Dalloul, Long et al. 2010](#_ENREF_1)), however this does not exclude the possibility of novel, turkey-specific miRNAs also being present. Our analysis extends the investigation of the turkey miRNA repertoire by performing deep sequencing to confirm the expression of these predicted turkey miRNAs in a cultured macrophage cell line. By sequencing small RNAs from turkey we also attempt to detect the presence of novel miRNAs. In addition we were interested in further examining patterns of molecular evolution between the microRNAs of these two species.

* Data for exploring avian miRNAs/evolution
* Zebrafinch miRNAs
* Question of whether miRNAs cluster by species or tissue

**Materials and Methods**

Illumina deep sequencing was performed on a small RNA library prepared from IAH30 cells, a cell line derived from turkey macrophages (Lawson et al., 2001) 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 are 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>) using 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. We then mapped known mirBase miRNAs to the turkey stem-loops. Stem-loops that had both sequencing reads and a known miRNA mapping to them were counted as expressed.

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.

**Results and Discussion**

Conclusions

**Useful information for results**

* 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 mirBase miRNA mapping = 249
* Number of turkey stem loops with no mirBase mature miRNA mapping = 180
* 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 chicken miRNA precursors annotated with 3p and 5p mature sequences

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

Our analysis yielded turkey 249 precursors with both reads and a mirBase miRNA mapping to them; 200 of which had chicken miRNAs mapping to them. This is much lower than the 437 predicted turkey miRNAs reported in the turkey genome paper (Dalloul et al., 2010). The likely explanation for this is that our sequencing data comes from macrophages, which are a highly specialized cell type. In fact, most of our reads are taken up with a few highly-expressed miRNAs.

* Enrichment of macrophage-specific miRNAs – Mann-Whitney test??
* Any novel 3p/5p miRNAs? Check
* Human/zebrafinch miRNAs that map where chicken do not – investigate and see if any interesting examples

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*

**References**

Dalloul, R. a, Long, J. a, Zimin, A. V, Aslam, L., Beal, K., Blomberg, L. A., Bouffard, P., et al. (2010). Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis. *PLoS biology*, *8*(9). doi:10.1371/journal.pbio.1000475

Lawson, S., Rothwell, L., Lambrecht, B., Howes, K., Venugopal, K., & Kaiser, P. (2001). Turkey and chicken interferon-gamma, which share high sequence identity, are biologically cross-reactive. *Developmental and comparative immunology*, *25*(1), 69–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10980321

Wheeler, B. M., Heimberg, A. M., Moy, V. N., Sperling, E. a, Holstein, T. W., Heber, S., & Peterson, K. J. (2009). The deep evolution of metazoan microRNAs. *Evolution & development*, *11*(1), 50–68. doi:10.1111/j.1525-142X.2008.00302.x

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.