Cuban Tree Frog *De novo* RNA-seq

Cindy Barreto and Mia Nahom

Cuban Tree Frog - *Osteopilus septentrionalis*

dreamstime.com

The largest tree frog in North America (5 to 13 cm or 2 to 5.5 inches)

Native to Cuba, the Bahamas, and the Cayman Islands (invasive to Florida)

They can secrete a toxic mucus from their skin which cause a fiery sensation

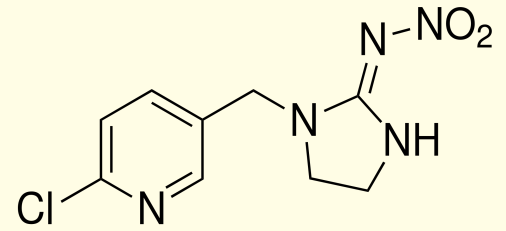Female frogs are larger than males

They can change their colors as a camouflage strategy - young frogs have more green coloration than adults

BackwaterReptiles.com

Cuban tree frog - *Osteopilus septentrionalis*

Imidacloprid

Insecticide that acts as neurotoxin - neonicotinoids

It blocks the nicotinic acetylcholine receptors, preventing acetylcholine from transmitting impulses between nerves
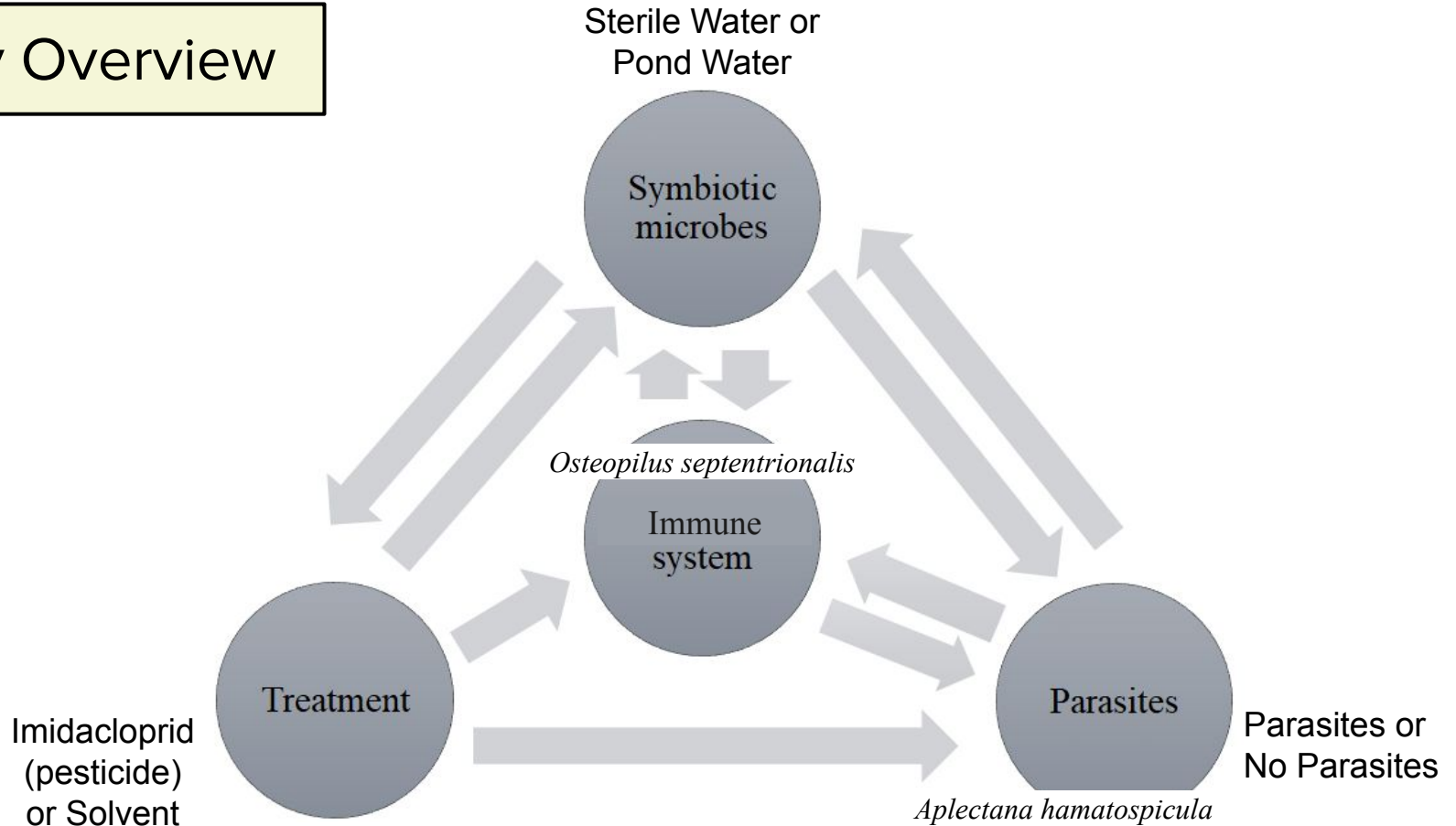
Moderately toxicity (oral) and low toxicity (dermal) - Group E (unlikely carcinogen and weakly mutagenic)

For aquatic life: high toxic on an acute basis EC50 = 3.7 - 115 ug/L

Effects: dizziness, apathy, locomotor effects, labored breathing, transient growth retardation, cardiac and hematological effects

Degenerative changes: testes, thymus, bone marrow, and pancreas

Study Overview

Sterile Water or Pond Water

Symbiotic microbes

Osteopilus septentrionalis

Immune system

Treatment

Imidacloprid (pesticide) or Solvent

Parasites

Parasites or No Parasites

Aplectana hamatospicula

**Samples:** 32 Adult frog blood samples preserved in RNAlater

-4 of each treatment possibility (8 possibilities):

Pesticide or Solvent X Sterile or Pond Water X No Parasites or Parasites
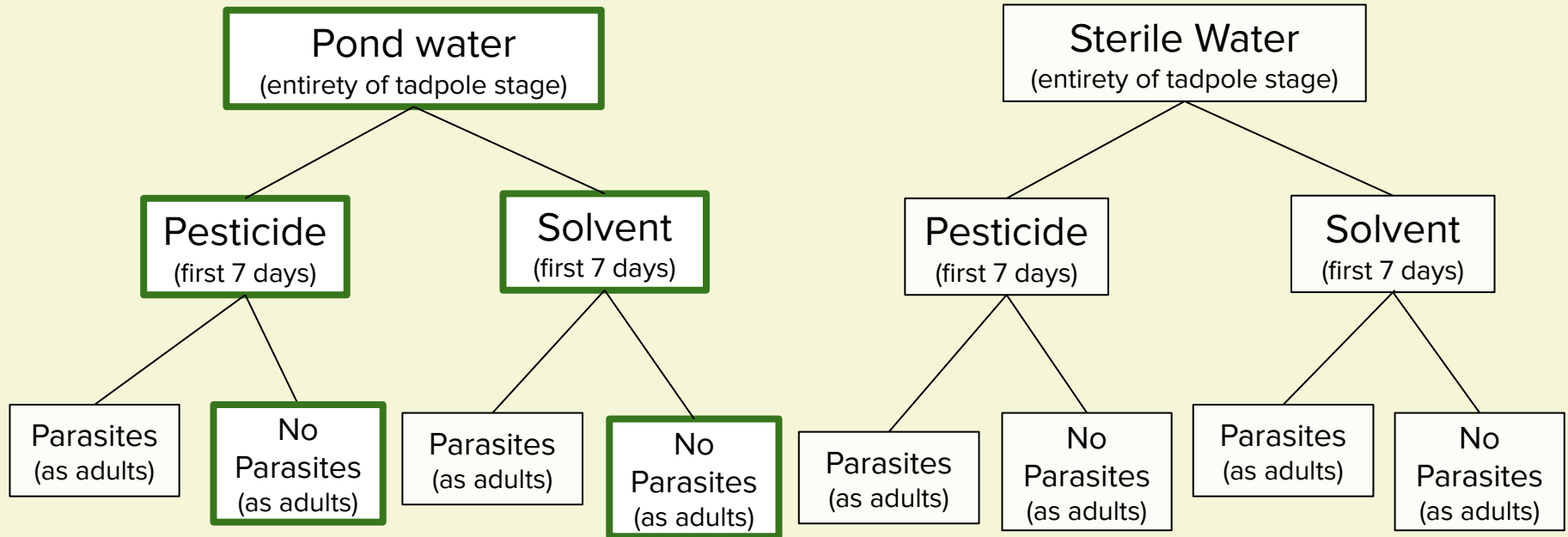
fastqc/multiqc & Trimming

- Trimmed out adapter content- illuminaclip and nextera transposase adapter

- Lost two more samples that were almost entirely adapter content

- Final Multiqc showed all of our samples passing quality scores & adapter content
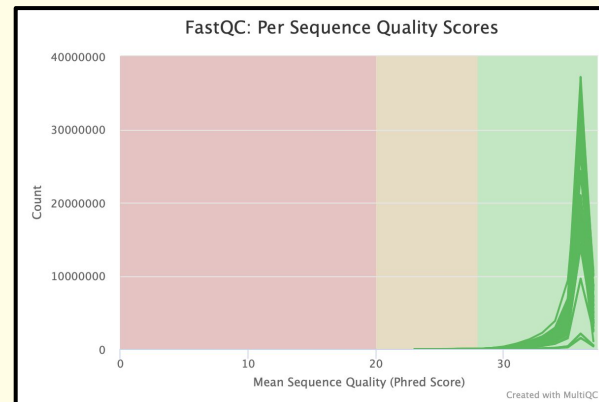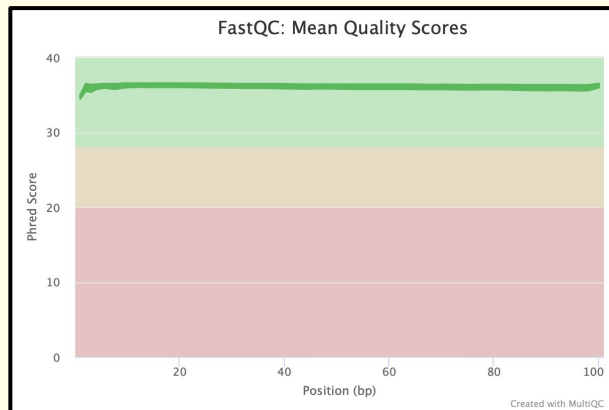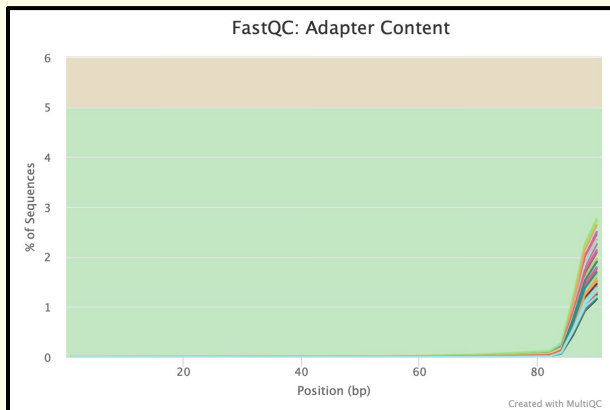
De novo Assembly with Trinity

flaticon.com
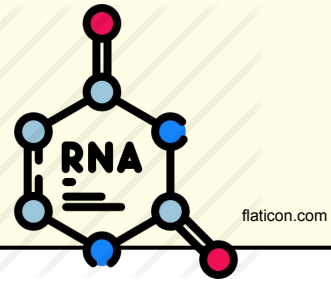
- Assembled each sample separately

- Added a sample prefix to the sequence names in the fasta files (to avoid identical sequence names being input into TransDecoder)

  - typo in one of the sample names resulted in a "prefix" file being blank and not included in our final assembly

- Concatenate the prefix files to create a combined fasta file for transdecoder, clustering, annotation, indexing, and aligning!

Transdecoder

- Ran TransDecoder to identify all long open reading frames
- Used hmmer (command hmmscan) to identify open reading frames with homology to known proteins by searching a protein family database
- Finally, we used TransDecoder.Predict to predict which open reading frames in our transcripts are real by using the output from our hmmscan run

Removing redundant transcripts

- Using vsearch we were able to cluster similar sequences amongst the transcripts and choose a representative to be used in our centroids.fasta file

pngio.com
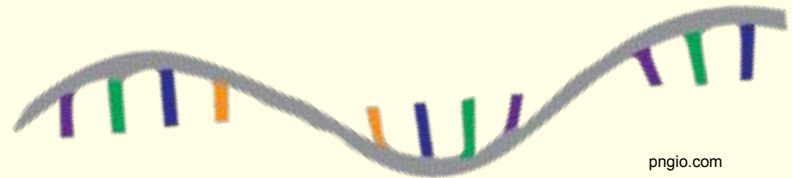
RNAQuast stats

	Transdecoder then Clustering	**Just Clustering**
Transcripts	42,092	264,703
Transcripts > 500 bp	16,521	152,852
Transcripts > 1000 bp	8,357	75,519
Average length of assembled transcripts	772.23	993.978
Longest Transcript	20,622	27,236
Total length	32,404,694	263,208,997
N50	1,350	2,496

EnTAP functional annotation

- Used the shared Diamond databases on xanadu
  - RefSeq, Uniprot, ntnr

- Flagged bacteria and fungi for contamination, and the cuban tree frog genus for the taxon
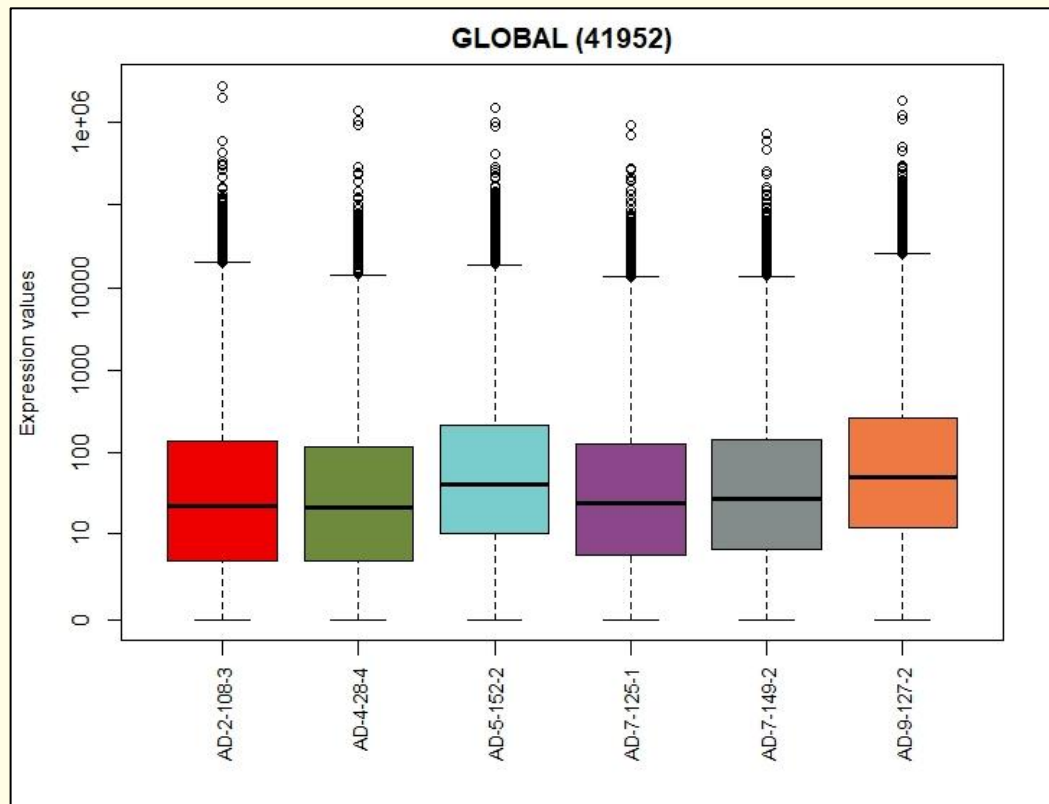 - Minimal contamination hits!

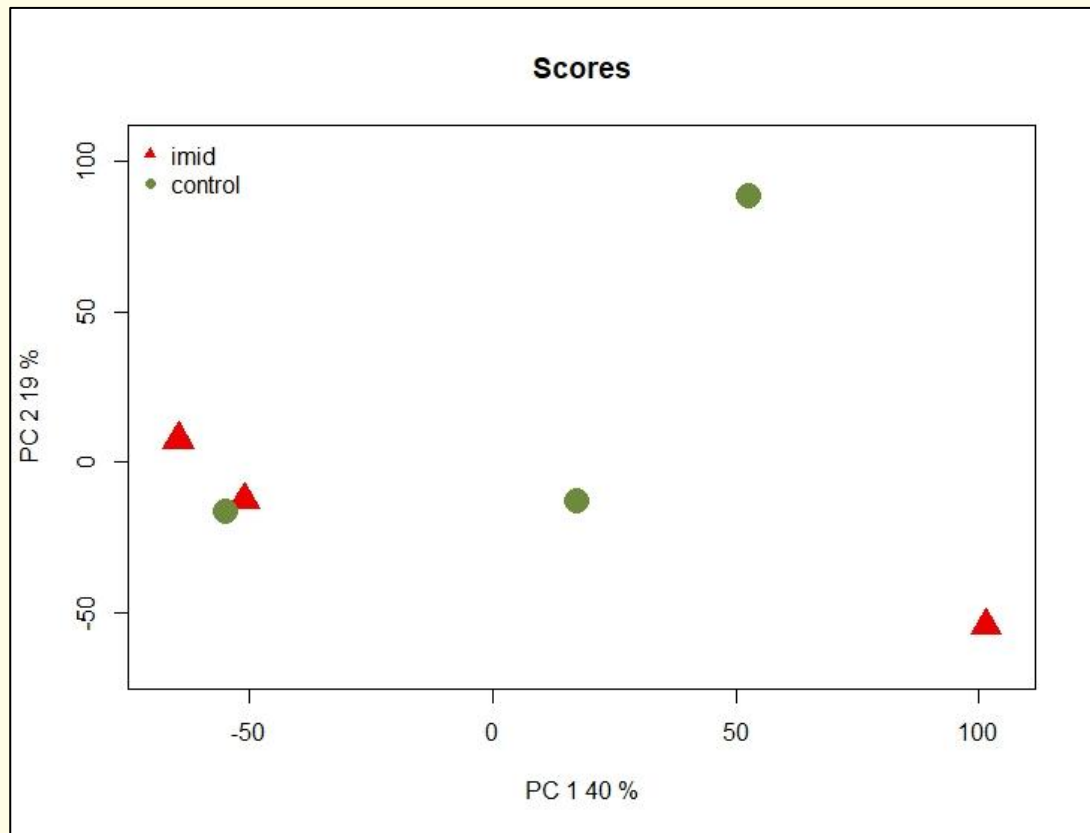| Species	Count
Xenopus tropicalis (Western clawed frog)	5100
X. laevis (African clawed frog)	2250
Nanorana parkeri (High Himalaya frog)	4279
Lithobates catesbeianus (American bullfrog)	1293
Homo sapiens	284
Mus musculus	160
Rhinatrema bivittatum (Two-lined caecilian)	136
Oncorhynchus mykiss (Rainbow trout)	120
Chelonoidis abingdonii (Galapagos-Pinta Island tortoise)	113
Microcaecilia unicolor	97

Indexing and Counting Reads with Kallisto

	Percentage of fragments pseudoaligned	Percentage of fragments pseudoaligned to a unique target sequence
AD-2-108-3	77.1%	67.1%
AD-4-28-4	69.1%	60.6%
AD-5-152-2	64.6%	57.3%
AD-7-125-1	65.2%	57.8%
AD-7-149-2	65.6%	58.8%
AD-9-127-2	65.4%	57.9%

Preliminary Analyses in NOISeq

The samples are clustered according to the experimental design!

control

imid

Mean expression

Length bins

R2 = 98.8%

p-value: 1.6e-116

R2 = 98.8%

p-value: 2.1e-116
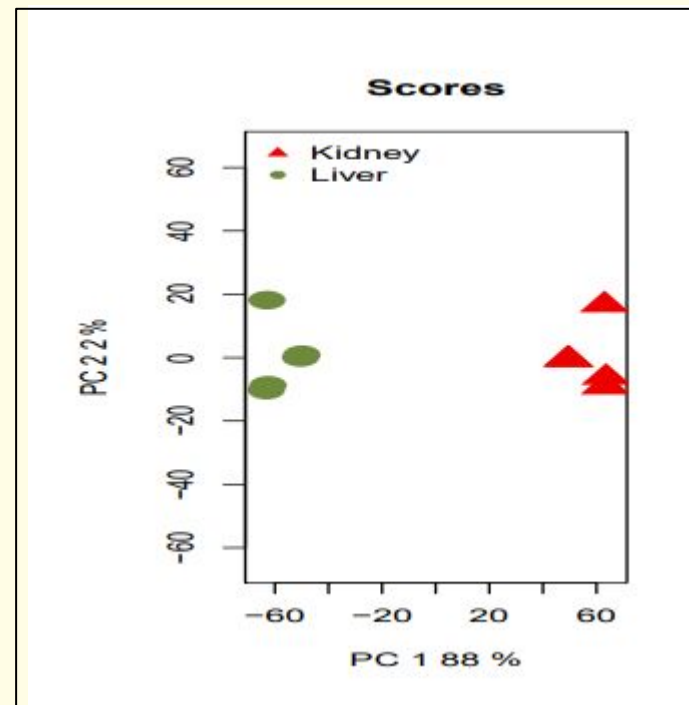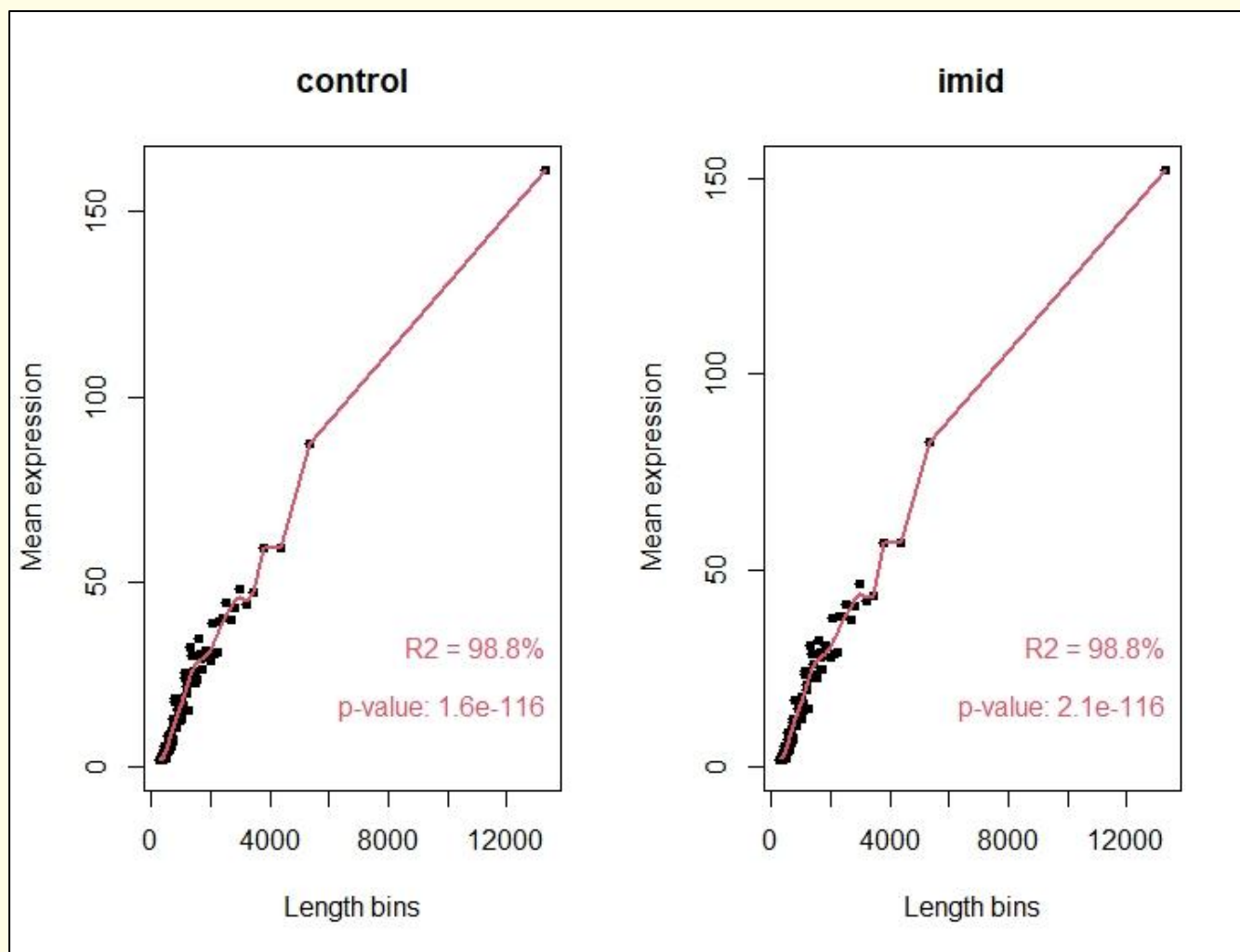
	# Differentially Expressed Features
Total	3751
More expressed in control samples	2755
Under-expressed in control samples	996
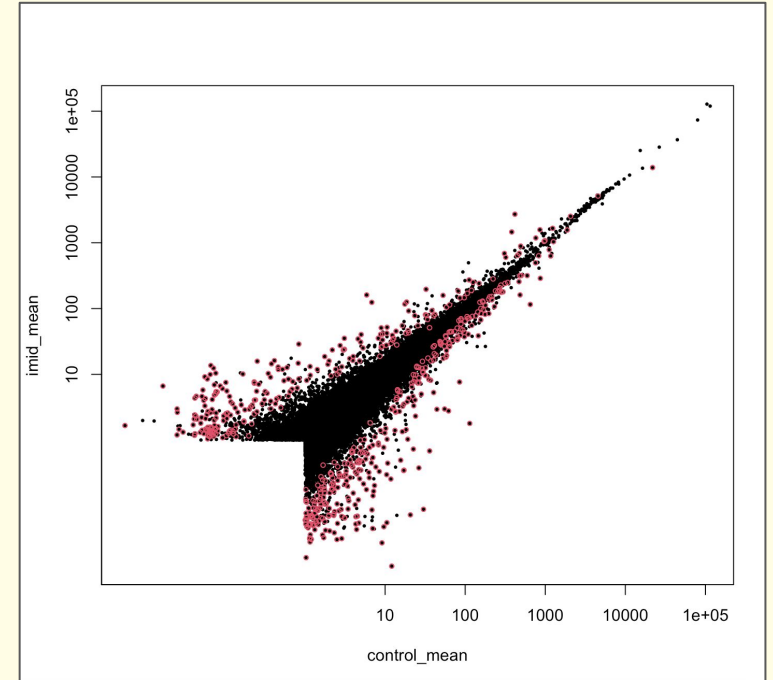
Conclusions & Future Directions

- Imidacloprid influences the gene expression according to preliminary analyses
  - Do the host-associated microbiota mediate the effect of imidacloprid on the tree frogs
- Rerun scripts with all the samples
  - Look at interactions between all the different treatments
  - Identify any effects from extraction date or other sources of bias
- Use our EnTAP outputs to get functional information about the differentially expressed genes

Photo: ©2006 James H. Harding

Cuban Tree Frog *De novo* RNA-seq

Questions?