Iso-seq Transcriptome Analysis of Hummingbird Archilochis Colubris

Rachael Workman[1], Alex Myrka[2], Kenneth Welch[2], G. William Wong[3], Winston Timp[1]

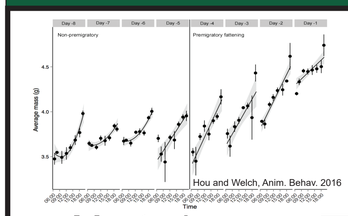
[1]Johns Hopkins University Department of Biomedical Engineering
[2]University of Toronto Scarborough Department of Biological Sciences
[3]Johns Hopkins University Department of Physiology

Abstract

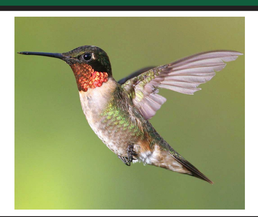
The hummingbird occupies a unique place in the vertebrate world. It has the highest known metabolic rate, needed to fuel incredible energetic demands of hovering flight the bird performs daily to collect nectar from flowers. To sustain hovering flight, a hummingbird needs to maintain a wing beat up to 80 beats per second. This remarkable feat is made possible by extremely high metabolic rates in the liver, with overall enzymatic activity operating at the peak of catalytic efficiency. Understanding the molecular basis of such extreme physiology will provide foundational knowledge to enable rational engineering of metabolic circuits in mammalian cells. To do this, we generated a de novo transcriptome of the hummingbird liver using PacBio IsoSeq, yielding a total of 8.6Gb of sequencing data, or 2.6M reads from 4 different size fractions. We analyzed data with the SMRTAnalysis IsoSeq 3.0 platform, including classification of reads, clustering of isoforms (ICE) followed by error-correction (Arrow). Here we describe our process of data QC, transformation (redundancy reduction, ORF prediction and translation), annotation, and orthology prediction, and work on pathway analysis and proteins of interest.

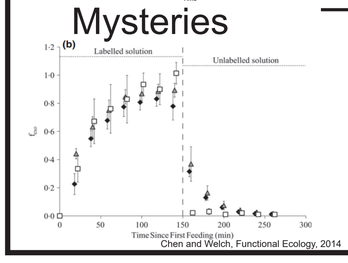
From series "Gems of Brazil" by Martin Johnson Heade

Introduction



Hou and Welch, Anim. Behav. 2016





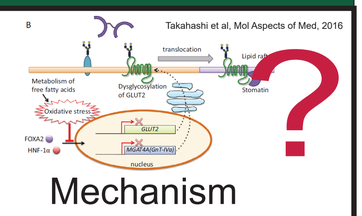
Chen and Welch, Functional Ecology, 2014

Mysteries

RNAseq

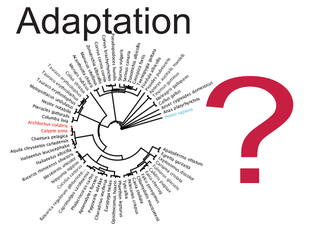
Pacbio Isoform Sequencing
Assembly-free full-length transcripts

Hummingbirds perform extraordinary metabolic feats, including massive daily weight fluctuations, rapid fuel-switching and fuel utilization. We sought to probe mechanism and adaptations present in this species through whole-transcriptome sequencing of the hummingbird liver using Pacbio Iso-seq.



Mechanism

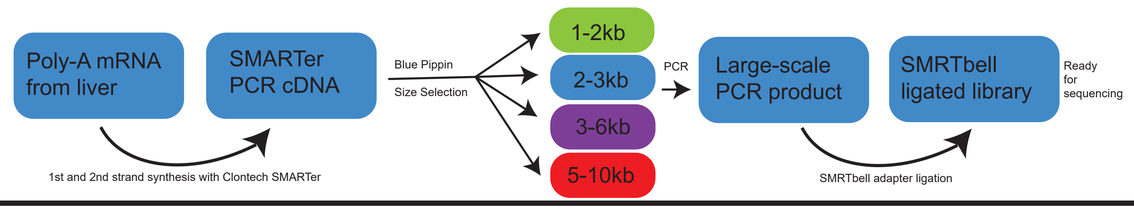
Adaptation



Methods

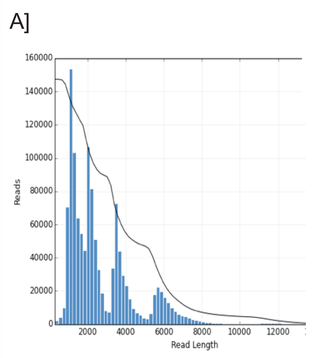
Sample prep:
Wild adult male ruby-throated hummingbirds (Archilochus colubris) were captured at the University of Toronto Scarborough using modified box traps. Birds were housed in the University of Toronto Scarborough vivarium and fed NEKTON-Nectar-Plus (Nekton, Tarpon Springs, FL, USA) ad libitum. Birds were sacrificed after ad libitum feeding, and tissues were sampled immediately after euthanization using RNAse-free tools. Six liver tissue samples were collected from six birds. Tissues were homogenized at 4ºC in 1 ml cold Tri Reagent using an RNase free glass tissue homogenizer and RNase free syringes of increasing needle gauge. Up to 100 mg of tissue was used per 1 ml of Tri Reagent, and chloroform extraction was performed twice to ensure quality. RNA was precipitated, centrifuged down, washed with ethanol, vacuum dried and eluted in RNAse free water. DNAse I digestion and spin column cleanup was performed. RNA concentration and RIN was determined with RNA Bioanalyzer.

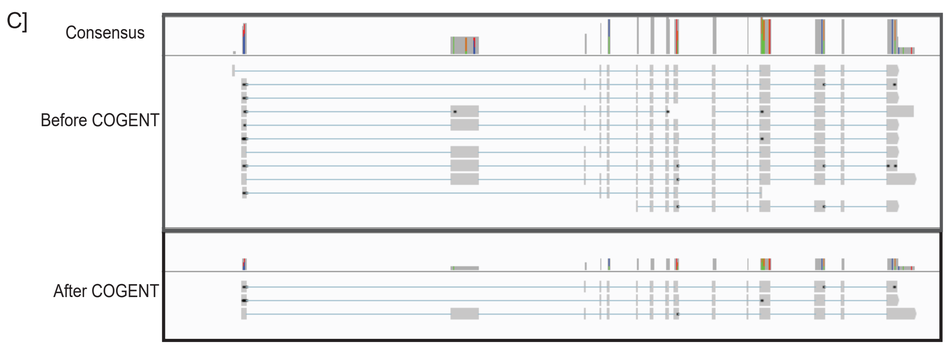
cDNA synthesis and SMRTbell template preparation:



Results

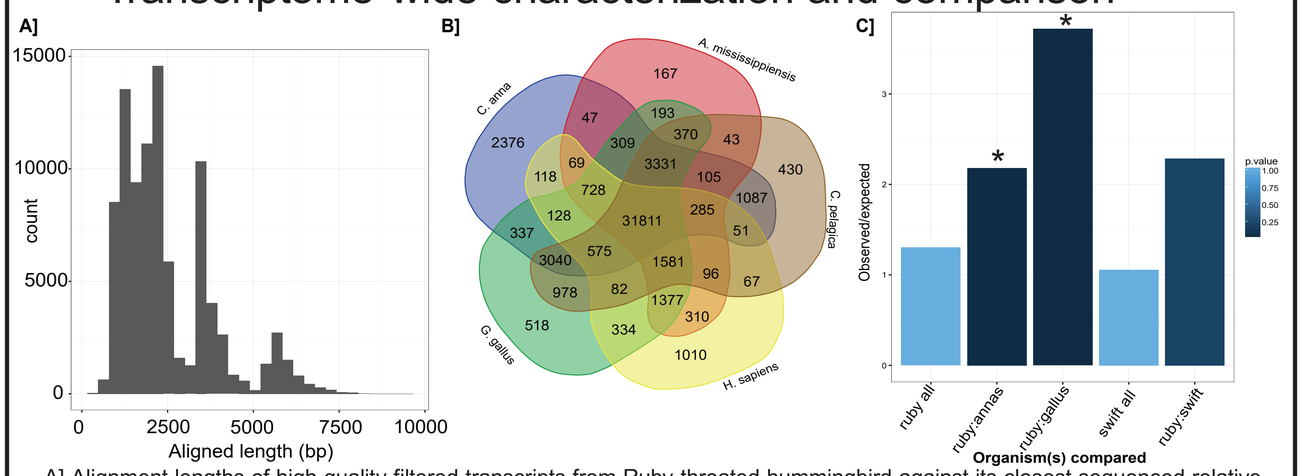
Data QC and Redundancy Reduction





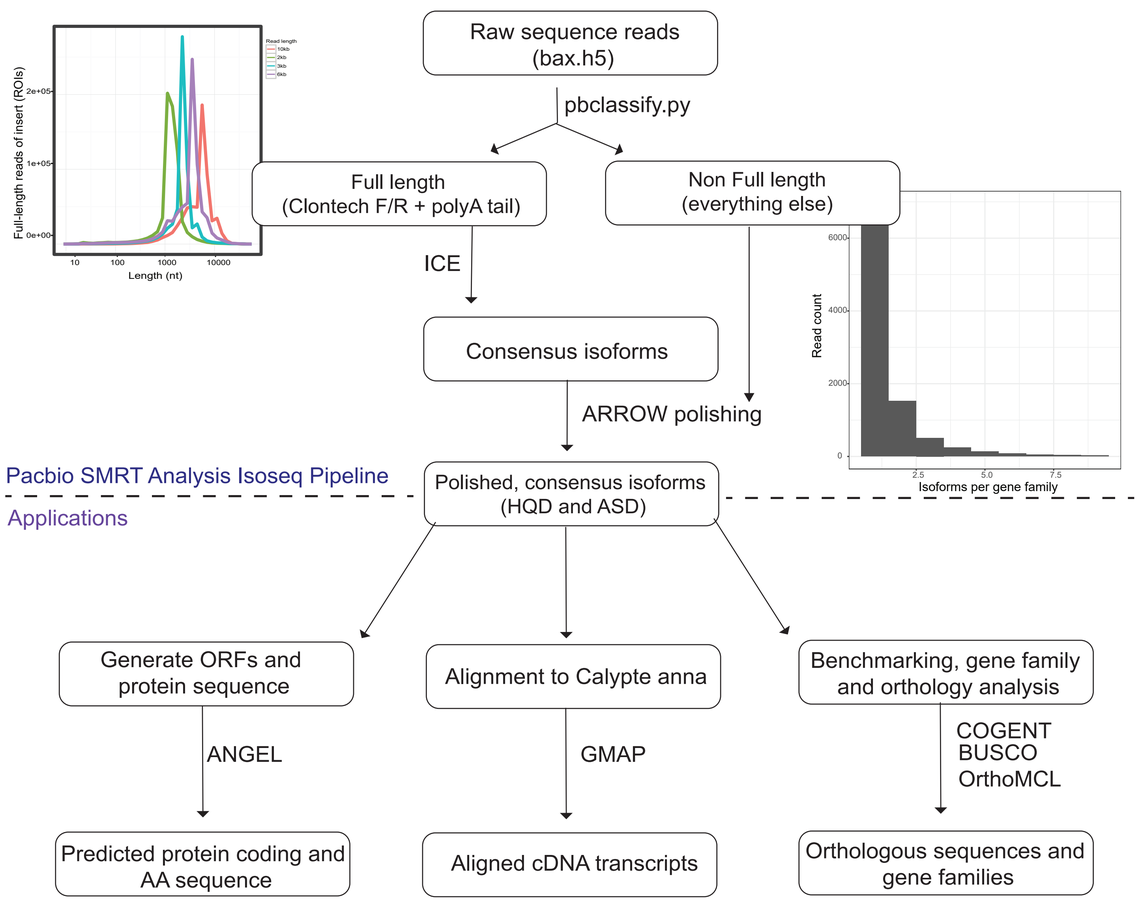
A] Data length distribution from 40 SMRT cells, black line represents Mb of reads in dataset longer than the given read length, demonstrates excellent read length generation. B] BUSCO benchmarking using Aves and Metazoan ortholog sets shows success in capturing predicted transcript diversity, with deficiencies potentially a consequence of single-tissue analysis. C] Cogent collapse. MATR3 example demonstrates Cogent success in reducing redundant isoforms and collapsing to unique isoforms.

Transcriptome-wide characterization and comparison



A] Alignment lengths of high quality filtered transcripts from Ruby-throated hummingbird against its closest sequenced relative, the Anna's hummingbird (Calypte anna) show full-length transcript mapping without need for assembly. B] Orthology analysis using OrthoMCL reveals large degree of transcript conservation and hints at interesting proteins for further investigation. C] GO analysis from 1:1 orthologs from OrthoMCL reveal larger than expected numbers of genes pertaining to lipid metabolism represented in the 1:1 Anna's and chicken comparison groups.

Analysis

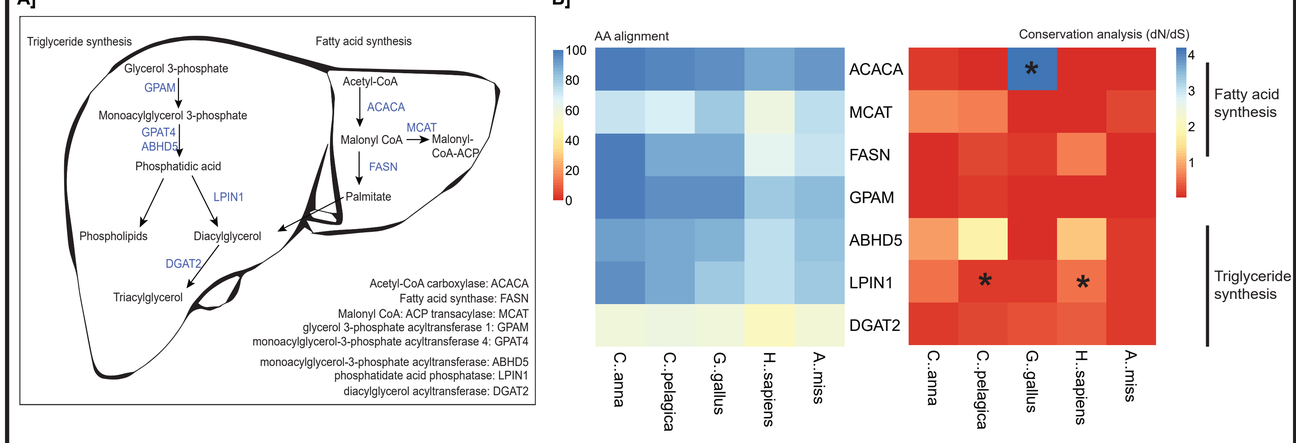


Discussion

Isoform sequencing is in the unique position of providing full length mRNA transcripts, giving clear insight into coding sequences for novel protein products. We have been able to identify some homologies between the ruby-throated hummingbird and other species, and see interesting differences that hint as to how extreme metabolisms function. Orthology analysis coupled to GO annotation has revealed interesting putative functional characterization for divergent hummingbird orthologs, and gene-level analysis of metabolic genes of interest provides key insight into sequence changes which will be investigated for functional significance.

Future work includes Iso-seq analysis of additional tissue types, completion of an Archilochis colubris genome, as degree of divergence between additional hummingbirds and established reference Calypte anna is unknown, and the cloning, enzymatic characterization, and Cryo-EM visualization of proteins of interest.

A Case Study: Hepatic lipogenic pathway



A] Diagram of fatty acid and triglyceride synthesis pathways, with investigated enzymes in blue. B] Heat map illustrating percent alignment identity between the organisms on x-axis and ruby-throated hummingbird, with each row representing the identities for the given enzyme. Percent identity more variable in enzymes involved in the triglyceride synthesis pathway relative to fatty acid synthesis, suggesting divergence to be investigated in future studies.

Acknowledgments

PacBio for reagents and protocol support and Haiping Hao of the JHU HIT Center for sequencing. The protocols in this study were approved by and performed in accordance with the requirements of U. Toronto Animal Care Committee. This study was financially supported by awards from PacBio's "World's Most Interesting Genome" grant and Human Frontiers in Science Grant (#RGP0062/2016).