

Thomas J Matthew

11/06/15

BME 205 HW6

Annotate Banana Slug mRNA

11/01/2015 15:55

[BlastX](#)¹ (search protein databases using a translated nucleotide query)

ran mitoDNA sequence mito_hand11

Request: 3D9W98T001R

database: "Non-redundant protein sequences (All non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF excluding environmental samples from WGS projects)

Max Target sequences: 100

Expectation Threshold: 10

Word size: 6

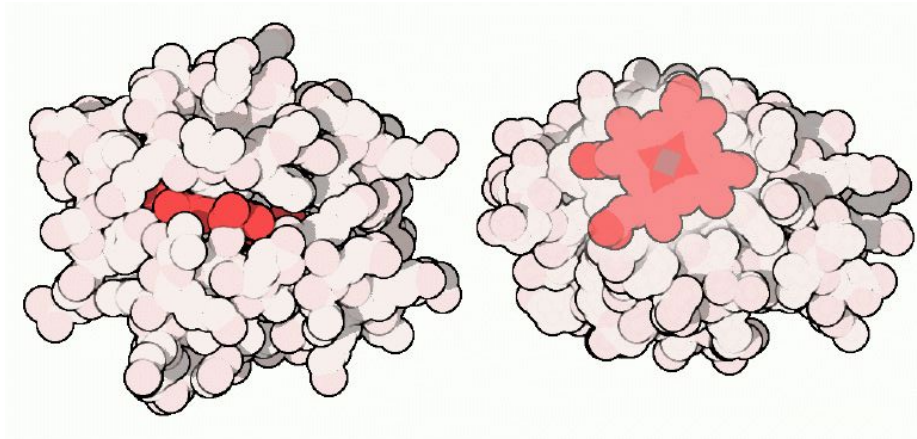
Matrix: BLOSUM62

Gap costs: Existence: 11, Extension: 1

Conditional compositional score matrix adjustment

Filter low complexity regions

Results: All top 100 homology results were from cytochrome C in land-based mollusks. Cytochrome C is expected to be encoded in the mitochondria due to its function as an electron carrier in oxidative phosphorylation, and also expected to be well-conserved in many other organisms. Unfortunately, all the highest sequence similarities returned by BlastX were restricted to Cytochrome C. BlastX was rerun with increased Max Target sequences parameter to capture a higher number of interesting protein alignments than what was gleaned here.



Cytochrome C illustration by David S. Goodsell courtesy [RCSB PDB](#)².

Request: 3FDT6PR9014

¹ Gish, Warren, and David J States. "Identification of protein coding regions by database similarity search." *Nature genetics* 3.3 (1993): 266-272.

² "RCSB Protein Data Bank - RCSB PDB." 2002. 6 Nov. 2015 <<http://www.rcsb.org/>>

database: "Non-redundant protein sequences (All non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF excluding environmental samples from WGS projects)

Max Target sequences: 5000

Expectation Threshold: 10

Word size: 6

Matrix: BLOSUM62

Gap costs: Existence: 11, Extension: 1

Conditional compositional score matrix adjustment

Filter low complexity regions

Results: After running with a much larger max target sequences, there was still more Cytochrome C returned as top hits. Returning more sequences was done based on the assumption that sequences can be filtered to reveal more diverse results. However, filtering hits on function or gene module does not appear to be possible in this interface of Blast. Since there are too many subsequent hits to adequately cover in less than 10 hours, a new tool needs to be used to find more informative sequence similarities for a more complete annotation of the mtDNA sample.

11/02/2015 10:45

[MITOS Web Server](#)³

Allows annotation of proteins and non-coding RNA of metazoan mitochondrial genomes (MGGs)

MGGs found by BLASTx searches of sequences against annotated proteins in NCBI RefSeq 39

Annotation of tRNA and rRNA accomplished using MITFI, [a structure-based covariance model](#)⁴

Inputs:

- 1) Fasta file: [draft assembly](#) of *A. dolichophallus* mitochondria (file upload)
- 2) Genetic code of [appropriate taxonomy](#) (drop down menu)

11/05/2015 10:00

Submitted MITOS job: MljzhWNT

Performed literature survey for land molluscan mitochondrial DNA, but was unsuccessful in identifying any new proteins over expected list of metabolic proteins conserved in humans.

11/05/2015 13:10

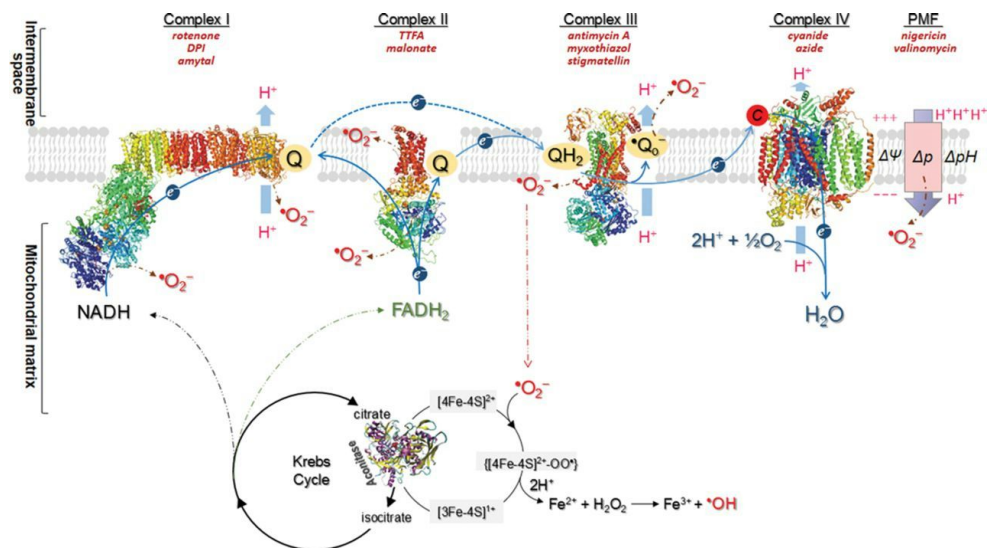
MITOS run complete

Results:

List of annotated genes from the mitochondrial genome fall into two main categories: housekeeping genes and metabolic genes. Housekeeping genes include large and small mitochondrial ribosome subunits and mitochondrial tRNA for 20 amino acids, allowing for self contained transcription consistent with the endosymbiotic hypothesis. All other genes fit into a common metabolic function, especially those playing a role in the electron transport chain and oxidative phosphorylation.

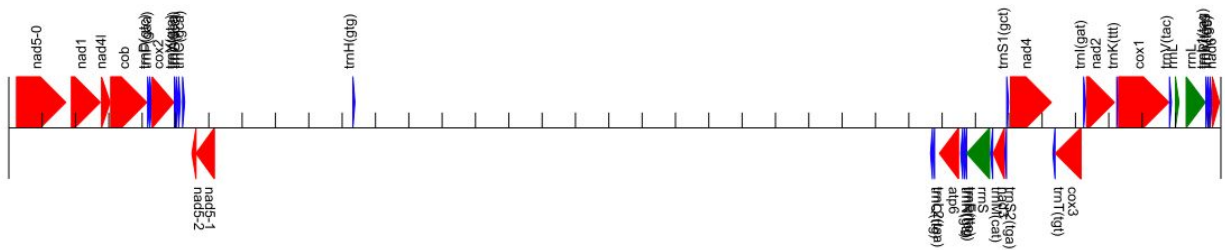
³ "MITOS Web Server." 2011. 5 Nov. 2015 <<http://mitos.bioinf.uni-leipzig.de/>>

⁴ Jühling, F. "Improved systematic tRNA gene annotation allows new ..." 2011. <<http://nar.oxfordjournals.org/content/early/2011/12/01/nar.gkr1131>>



Electron Transport Chain illustration by Jay L. Zeier⁵

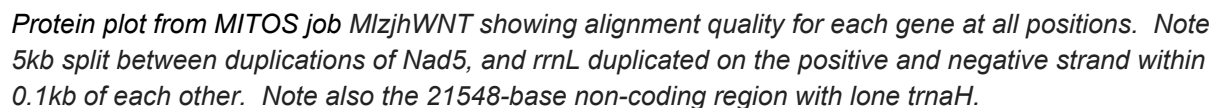
NADH dehydrogenase (Nad^{*}) plays a role in dehydrogenating the electron carrier NADH to create a proton gradient across the inner mitochondrial membrane. This proton motive force is harnessed by ATP synthase which translates kinetic energy of protons moving back to the matrix into chemical energy in the form of oxidative phosphorylation of ADP with inorganic phosphate. ATP synthase shows as ATP6 for its F₀ (inner membrane) subunit, and as (Nad5-*) for complex V.



Linearized plot of *A. dolichophallus* mtDNA with annotations from MITOS job MlzhWNT

It is interesting to note that ATP8, the mitochondrial gene associated with ATP synthase subunit 8 is expected but not present. The genes *rrnL*, mitochondrial 23S ribosomal subunit, are duplicated across the positive and negative strands from one another. The gene *Nad5*, electron acceptor complex V, has three transplises and a single duplication on the the same strand nearly 5kb apart from each other. Another interesting observation is the lack of any genes (with the clear exception of *trnH*) in the middle 21548-base stretch. A next step could be to investigate the location of histidine tRNA in other species to determine if this feature is specific to the banana slug.

⁵ Chen, Yeong-Renn, and Jay L. Zweier. "Cardiac mitochondria and reactive oxygen species generation." *Circulation research* 114.3 (2014): 524-537.



Annotated mtDNA genes from <i>A. Dolichophallus</i>				
gene	start loc	stop loc	strand	product
nad5-0	231	1712	+	NADH dehydrogenase Subunit V, transplice 1/3
nad1	1877	2746	+	NADH dehydrogenase Subunit I
cob	3057	4142	+	Cytochrome B Oxidase Core
trnD(gtc)	4166	4229	+	tRNA Aspartic Acid
trnF(gaa)	4232	4295	+	tRNA Phenylalanine
cox2	4299	4955	+	Cytochrome C Oxidase Subunit II
trnY(gta)	4968	5023	+	tRNA Tyrosine
trnW(tca)	5024	5088	+	tRNA Tryptophan
trnG(tcc)	5096	5147	+	tRNA Glycine
trnC(gca)	5214	5277	+	tRNA Cysteine
nad5-2	5618	5493	-	NADH dehydrogenase Subunit V, transplice 3/3
nad5-1	6169	5615	-	NADH dehydrogenase Subunit V, transplice 2/3
trnH(gtg)	10324	10384	+	tRNA Histidine
trnL2(taa)	27776	27717	-	tRNA Leucine 2
atp6	28495	27917	-	ATP synthase F0 sector
trnN(gtt)	28618	28559	-	tRNA Asparagine
trnR(tcg)	28677	28620	-	tRNA Arginine
trnE(ttc)	28737	28673	-	tRNA Glutamic Acid
trnK(ttt)	29426	28737	+	tRNA Lysine
rrnS	29426	28737	-	16S Ribosomal RNA
trnQ(ttg)	29520	29455	-	tRNA Glutamine
trnM(cat)	29520	29455	-	tRNA Methionine
nad3	29870	29532	-	NADH dehydrogenase Subunit III
trnS2(tga)	29933	29883	-	tRNA Serine 2
trnS1(gct)	29939	29995	+	tRNA Serine 1
nad4	30051	31280	+	NADH dehydrogenase Subunit IV
nad4	30051	31280	+	NADH dehydrogenase Subunit IV
trnT(tgt)	31389	31325	-	tRNA Threonine
cox3	32170	31394	-	Cytochrome C Oxidase Subunit III
trnI(gat)	32247	32310	+	tRNA Isoleucine
nad2	32332	33171	+	NADH dehydrogenase Subunit II
cox1	33291	34805	+	Cytochrome C Oxidase Subunit I

trnV(tac)	34823	34881	+	tRNA Valine
rrnL	35000	35108	+	23S Ribosomal RNA
rrnL	35322	35897	+	23S Ribosomal RNA
trnL1(tag)	35914	35970	+	tRNA Leucine 1
trnP(tgg)	35971	36036	+	tRNA Proline
trnA(tgc)	36037	36102	+	tRNA Alanine
nad6	36107	36322	+	NADH dehydrogenase Subunit VI

Genes annotated from *A. dolichophallus* mtDNA, with gene products taken from [wikigene](#)⁶ searches

⁶ Maier, Holger et al. "LitMiner and WikiGene: identifying problem-related key players of gene regulation using publication abstracts." *Nucleic acids research* 33.suppl 2 (2005): W779-W782.