Polymorphisms within the *Limulus polyphemus* mitochondrial genome

Megan Cooper, C.C. Chabot; Plymouth State University, Plymouth, NH

Introduction

The horseshoe crab *Limulus polyphemus* is important both ecologically and biomedically. Its eggs are a key food source to migratory birds stopping over the Atlantic coast , and adults are the source of *Limulus* amoebocyte lysate used in the biomedical industry to test for Gram-negative bacterial infection. While the mitochondrial genome for *L. polyphemus* was recently published to NCBI by Lavrov *et. al (2012),* mitochondrial genomes are subject to high rates of mutation. This makes it necessary to analyze multiple mitochondrial genomes within the species to gain a comprehensive view of the mitochondria of the species as a whole. In addition, mitochondrial genomes are transmitted uniparentally from mother to offspring. This tends to create varied mitochondrial genomes amongst maternal lineages, which, separated over evolutionary time, generally become more dissimilar. The goal of this study was to determine sites at which nucleotides within the coding region of mitochondrial genes differ between mitochondrial lineages. We compared the published mitochondrial genome to a *L. polyphemus* mitochondrial genome from a different population to find sites where nucleotides vary within the coding regions, better known as polymorphic residues as a means of finding divergent structures and denoted functions within the genome.

Methods

A *Limulus polyphemus* adult (male/female) was obtained from Woods Hole estuary, and mitochondrial DNA was extracted from a tissue sample and sequenced using Illumina deep sequencing. The sequence data wer imported to CLC Genomics Workbench 5 (CLCbio, http://www.clcbio.com/products/clc-genomics-workbench/) and coding regions were annotated before they were individually aligned to the corresponding coding region of the published *L. polyphemus* mitochondrial genome (NC_003057.1). Polymorphic residues were noted (Table 1) and the entire mitochondrial genome uploaded to NCBI. Transfer RNA secondary structures were derived using tRNAscan-SE v.1.21 (The Lowe Lab, http://lowelab.ucsc.edu/tRNAscan-SE/) (Figure 2). Protein coding gene sequences were translated using CLC Genomics Workbench 5. Amino acid sequences were compared to that of published sequences to find sequence dissimilarities denoting altered function, if any.

References

Avise, J.C., Giblin-Davison, C., Laerm, J., Patton, J.C., Lansman, R.A. "Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, Geomys pinetis", *Proceedings of the National Academy of Sciences* 76 (1979): 6694-6698

Shearer, T.L., Van Oppen, M.J.H., Romano, S.L., Wörheide. "Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria)" *Molecular Ecology* 11.12 (2002):2475-2487

Valverde, J.R., Batuecas, B., Moratilla, C., Marco, R., Garesse, R. "The complete mitochondrial DNA sequence of the crustacean Artemia franciscana", *Journal of Molecular Evolution* 39 (1994): 400-408

Gene Name	Number of Sites	Number of Variable Sites	Number of Substitutions Ts Transitions	Tv Transversions	Number of Frameshift Mutations Insertions	Deletions	Proportion of Variable Sites
ATP6	675	19	16	2	0	1	0.0281
ATP8	156	4	4	0	0	0	0.0256
CYTB	1132	40	35	5	0	0	0.0353
COX1	1536	36	34	2	0	0	0.0234
COX2	685	18	15	3	0	0	0.0263
COX3	784	17	11	6	0	0	0.0217
ND1	933	24	22	2	0	0	0.0257
ND2	1017	31	25	6	0	0	0.0305
ND3	345	13	12	1	0	0	0.0377
ND4	1338	29	24	5	0	0	0.0217
ND4L	300	12	12	0	0	0	0.0400
ND5	1714	47	43	4	0	0	0.0274
ND6	462	12	10	2	0	0	0.0260
rRNA 12S	799	12	10	2	0	0	0.0150
rRNA 16S	1298	26	17	5	3	1	0.0200
Alanine	67	0	0	0	0	0	0.0000
Arginine	63	0	0	0	0	0	0.0000
Asparginine	65	1	1	0	0	0	0.0154
Aspartic Acid	67	1	1	0	0	0	0.0149
Cysteine	64	0	0	0	0	0	0.0000
Glutamic Acid	66	1	1	0	0	0	0.0152
Glutamine	66	2	2	0	0	0	0.0303
Glycine	64	0	0	0	0	0	0.0000
Histidine	69	0	0	0	0	0	0.0000
Isoleucine	67	1	1	0	0	0	0.0149
Leucine CUN	69	0	0	0	0	0	0.0000
Leucine UUR	66	1	1	0	0	0	0.0152
Lysine	70	0	0	0	0	0	0.0000
Methionine	70	0	0	0	0	0	0.0000
Phenylalanine	66	1	1	0	0	0	0.0152
Proline	67	1	1	0	0	0	0.0149
Serine AGN	73	0	0	0	0	0	0.0000
Serine UCN	73	44	13	22	9	0	0.6027
Threonine	69	0	0	0	0	0	0.0000
Tryptophan	68	0	0	0	0	0	0.0000
Tyrosine	67	1	1	0	0	0	0.0149
Valine	69	0	0	0	0	0	0.0000

Table 1. Polymorphic Sites by Type and Gene. Transition and transversion substitutions as well as insertion/deletion mutations were tallied and noted for each of the mitochondrial genes. These mutations were used to calculate the proportion of nucleotides that vary between ours and the published mitochondrial genome.

Results

The total length of this *L. polyphemus* mtDNA variation is 15,012 bp. This is 27 bp longer than the published mitochondrial genome (Lavrov *et al.,* 2000). There are multiple polymorphisms between the two genomes, as well as insertions and deletions. Twelve of the nucleotide insertion mutations occur within coding regions (See Table 1). The coding regions of this mtDNA and that of the published are similar; the tRNAs, for their brevity, are displayed in Figure 1. Polymorphisms were revealed in 25 of the 37 mitochondrial genes, shown with an asterisk following the gene name in Table 1. Transfer RNA-coding genes showed both the highest and lowest variance, as seen in Table 1. None of these nucleotide variations represent a change in anticodon specificity at the secondary RNA structure level from published sequences (Figure 2). One of the two Serine tRNA coding genes was highly mutated, and so a secondary structure was not produced by tRNAscanSe1.21.

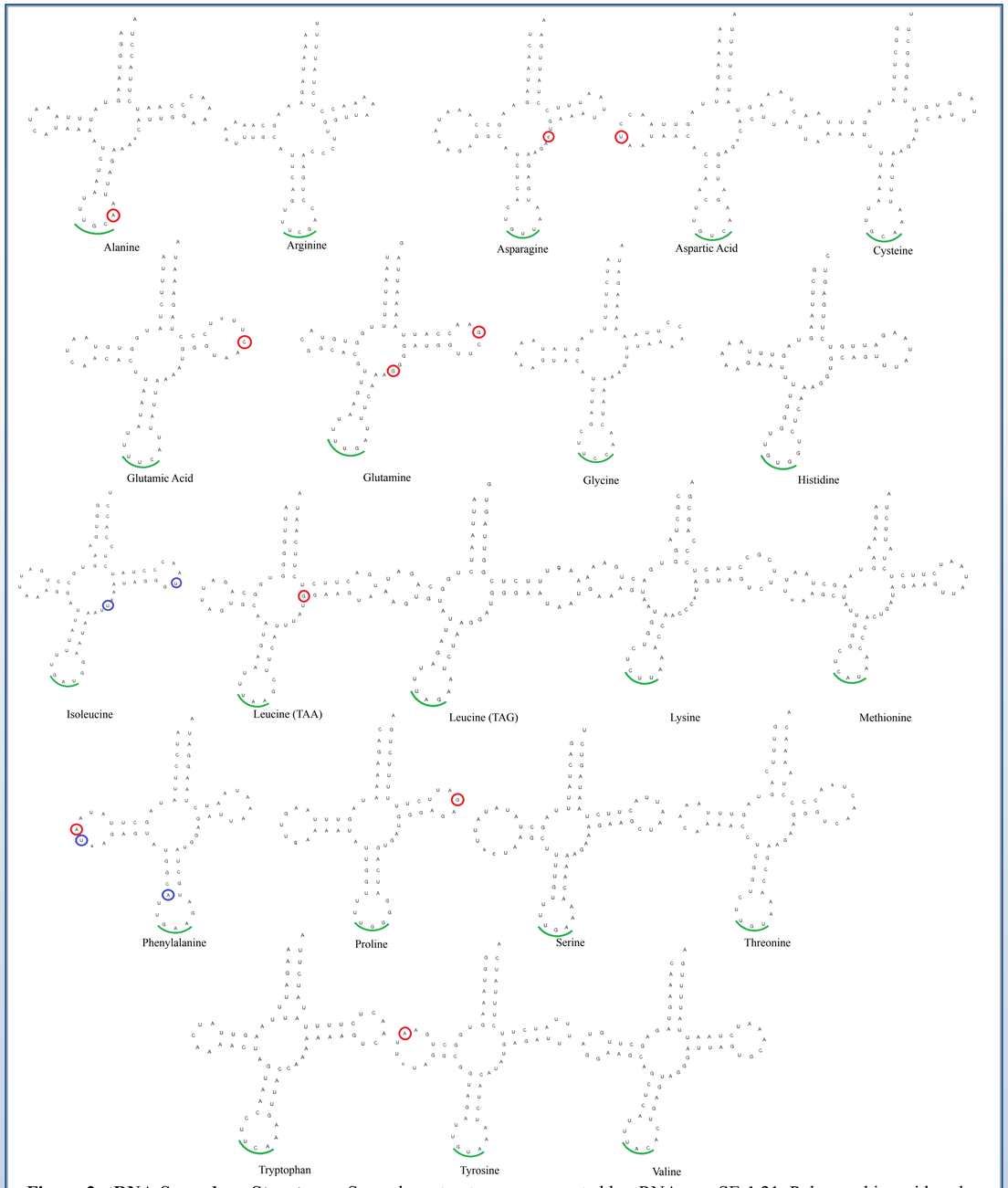

Figure 2. tRNA Secondary Structures. Secondary structures as presented by tRNA-scanSE-1.21. Polymorphic residues between the two genomes are circled by type. Transition mutations, between two purine or two pyrimidine nucleotides, are circled in red. Transversion mutations, mutating from pyrimidine to purine or vice versa, are circled in blue. Anticodons are underlined in green.

Discussion/Future Directions

- Percent similarity lower than expected
  - ◊ Estimate of mtDNA pairwise base substitution per nucleotide 0.018 in mammal (Avise *et al.,* 1979), and an average coding region diversity of 0.17% (Shearer *et al.,* 2008)
- tRNA retain functional properties
  - ◊ All highly conserved except for one of the serine tRNA genes, which was largely mutated except for the anticodon
- Analyze noncoding and protein-coding regions for diversity
  - ◊ Find area of mtDNA control region, a long A+T hairpin-prone noncoding region conserved in many species (Valverde *et al.,* 1994)
  - ◊ Analyze polypeptide sequence for insight into secondary structure for assay antibodies.

◊ **Acknowledgements: The authors would like to thank Kazu Omomoto, Kelley Thomas and Win Watson for their help and expertise.**