Mitochondrial Sequences

VectorBase team members @ FORTH-IMBB

Updated by Gloria I. Giraldo-Calderón
December 2013

VectorBase

Bioinformatics Resource for Invertebrate Vectors of Human Pathogens

Outline

1. Why?

2. Accessing mitochondrial sequences

3. Origin of the mitochondrial sequences: GenBank

4. Gene list:
   - Specific gene list
   - Specific species gene list

1. Why ?

Because VectorBase organisms have the <u>nuclear genome</u> but not the <u>mitochondrial genome</u> represented in the <u>gene set</u>.

Released genomes, with gene predictions

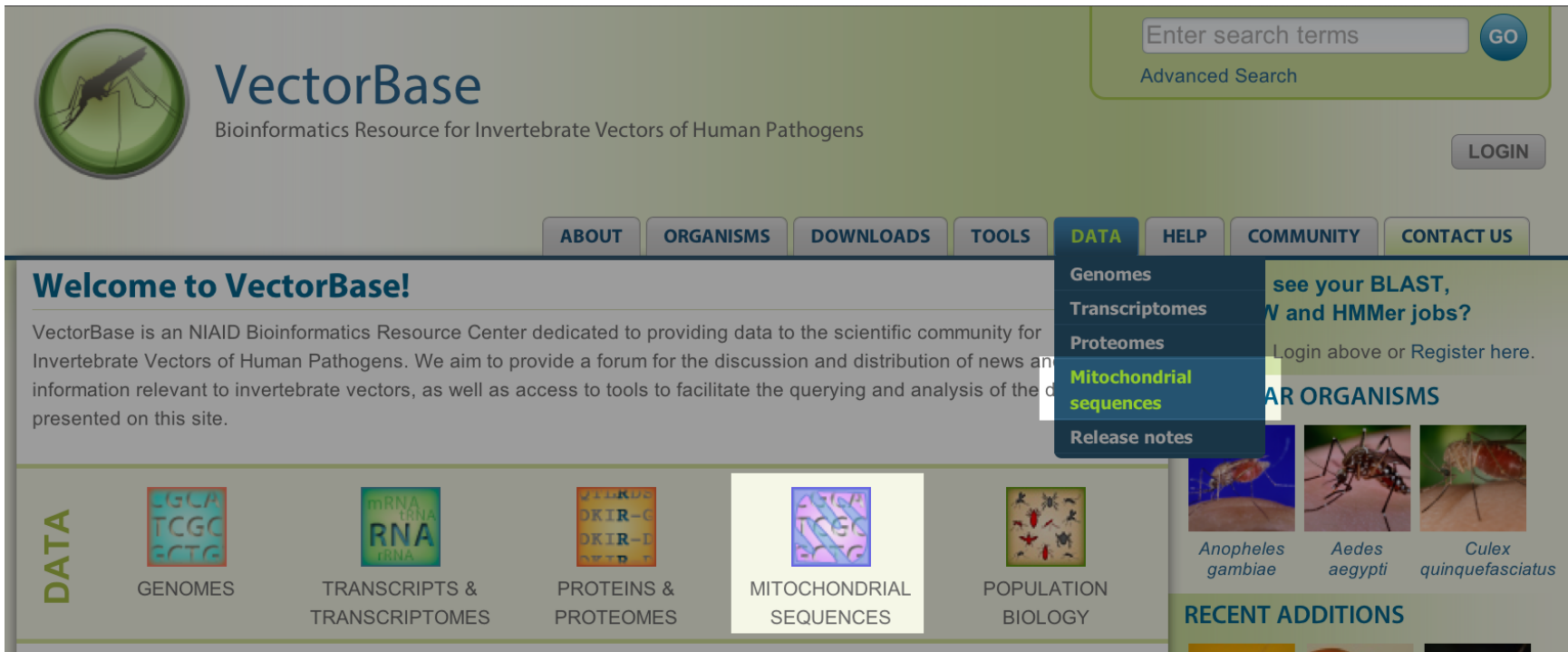| Organism ▲	Strain	Assembly	Gene set	Gene Count	GenBank WGS Project	GenBank Assembly ID	Genome Size (bp)
Anopheles gambiae	PEST	AgamP3	AgamP3.7	13 465	AAAB01	GCA_000005575.1	264 000 000
Aedes aegypti	Liverpool LVP	AaegL1	AaegL1.4	17 391	AAGE02	GCA_000004015.1	1 310 090 344
Ixodes scapularis	Wikel	IscaW1	IscaW1.2	24 925	ABJB01	GCF_000208615.1	1 765 382 190
Culex quinquefasciatus	Johannesburg JHB	CpipJ1	CpipJ1.3	22 985	AAWU01	GCA_000209185.1	539 959 374
Pediculus humanus	USDA	PhumU1	PhumU1.2	10 992	AAZO01	GCA_000006295.1	110 781 312
Rhodnius prolixus	CDC	RproC1	RproC1.1	17 143	ACPB02	GCA_000181055.2	702 648 350
Glossina morsitans	Yale	GmorY1	GmorY1.2	12 362			363 107 930
Anopheles stephensi	Indian	AsteI2	AsteI2.1	11 789	ALPR002	GCA_000300775.2	209 737 895
Anopheles darlingi	Coari	AdarC1	AdarC1.1	11 430	ADMH01	GCA_000211455.2	113 509 265

Pre-released genomes, with mapped data and gene predictions

Organism ▲	Strain	Assembly	Gene set	Gene Count	GenBank WGS Project	GenBank Assembly ID	Genome Size (bp)
Lutzomyia longipalpis	Jacobina	LlonJ1	LlonJ1.0	10 110	AJWK01	GCA_000265325.1	154 229 266
Phlebotomus papatasi	Israel	Ppapl1	Ppapl1.0	11 164	AJVK01	GCA_000262795.1	347 840 937
Anopheles albimanus	STECLA	AalbS1	AalbS1.0	11 911	APCK01	GCA_000349125.1	170 508 315

www.vectorbase.org/genomes

2. Accessing mitochondrial sequences

From the <u>home page</u> or from the <u>navigation tab</u> called DATA.

3. Origin of the mitochondrial sequences: GenBank

- Almost all (>99 %) GenBank sequences should be on VectorBase. Most of those that did not make it into VectorBase contain errors in at least one field and could not be parsed.

- If you are interested in knowing more details about filters used to retrieve the sequences, BLAST parameters and etc., send us an email to info@vectorbase.org. For the reference genomes see next slide.

Mitochondrial data	Number of species	GenBank entries
Genus *Aedes*	93	3,053
Genus *Anopheles*	314	13,679
Genus *Culex*	113	2,133
Order *Ixodida*	328	7,423

3. Origin of the mitochondrial sequences: GenBank

Complete mitochondrial reference genomes
(NCBI Reference Sequence)

Genus *Anopheles*: *Anopheles gambiae* (NC_002084.1)
Beard, C.B., et al. 1993. *Insect Mol. Biol.* 2: 103-124.

Genus *Aedes*: *Aedes albopictus* (NC_006817.1)
Unpublished data

Genus *Culex*: *Culex quinquefasciatus* (NC_014574.1)
Behura, S.K., et al. 2011. *Insect Biochem. Mol. Biol.* 41: 770-777.

Order Ixodida: *Ixodes scapularis*
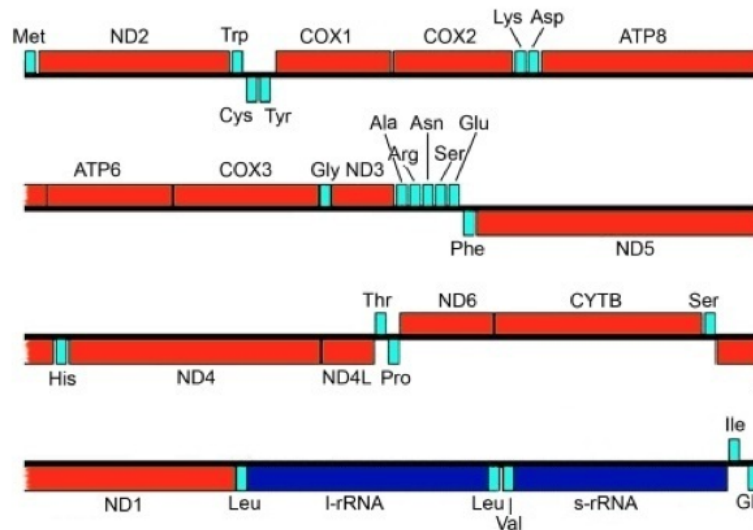Unpublished data provided by Catherine Hill, Purdue University

4. Gene list

1.

ATP6	ATP8	COX1	COX2	COX3	CYTB
l-rRNA	ND1	ND2	ND3	ND4	ND4L
ND5	ND6	s-rRNA	tRNA-Ala	tRNA-Arg	tRNA-Asn
tRNA-Asp	tRNA-Cys	tRNA-Gln	tRNA-Glu	tRNA-Gly	tRNA-His
tRNA-Ile	tRNA-Leu	tRNA-Lys	tRNA-Met	tRNA-Phe	tRNA-Pro
tRNA-Ser	tRNA-Thr	tRNA-Trp	tRNA-Tyr	tRNA-Val	

2.

- A text list (**1**) and a diagram (**2**), both are clickable and can lead you to the respective gene.
- Text list (**3**) for species too.

The clickable figure can also lead you directly to the respective gene (also found in the upper table). The map is a composite approximation for all anopheline mtDNAs and each line represents approximately 4000 base pairs. Turquoise: tRNA genes, dark blue: rDNA, red: other genes. The direction of transcription is indicated as from left to right (above the line) or from right to left (below the line).

3.

Species		
Amblyomma americanum	Amblyomma antillorum	Amblyomma argentinae
Amblyomma aureolatum	Amblyomma auricularium	Amblyomma boeroi
Amblyomma brasiliense	Amblyomma cajennense	Amblyomma calcaratum
Amblyomma coelebs	Amblyomma crassum	Amblyomma darwini

VectorBase
http://www.vectorbase.org

4. Gene list

Specific _gene_ list (1):

- By selecting a gene you get all the instances (orthologs) of this gene from several species.
- The accession number leads to the GenBank entry of the sequence.
- The last column displays the coordinates of the most significant **HSP** (high scoring pair) between the sequence and the complete mitochondrial genome (slide 6).

Mitochondrial sequences containing the ATP6 gene			
Ac. Num.	Description	Gene(s)	Sequence / *A. gambiae* match
☐ JX573138	Haemaphysalis humerosa mitochondrion, partial genome.	ATP6	1-776/10199-10974
☐ JX573138	Haemaphysalis humerosa mitochondrion, partial genome.	ATP6	2425-2445/9476-9496
☐ JX573138	Haemaphysalis humerosa mitochondrion, partial genome.	ATP6	7013-9524/10993-13600

VectorBase
http://www.vectorbase.org

4. Gene list

Specific <u>species</u> gene list (3):

- A list of all mitochondrial genes for each species.
- The accession number leads to the full GenBank record.

Mitochondrial sequences from *Amblyomma americanum*

	Ac. Num.	Description
☐	AF150050	Amblyomma americanum 12S small ribosomal RNA gene, partial sequence; mitochondrial gene for mito
☐	AY342280	Amblyomma americanum 12S ribosomal RNA gene, partial sequence; mitochondrial.
☐	DQ168129	Amblyomma americanum cytochrome b (cytb) mRNA, complete cds; mitochondrial.

- You can select the gene and with a click the fasta file will download to your computer (without going to GenBank).

☐	L34313	Mitochondrion Amblyomma americanum 16S ribosomal RNA (16S rRNA) gene.
☐	L34314	Mitochondrion Amblyomma americanum 16S ribosomal RNA (16S rRNA) gene.
☐	U95849	Amblyomma americanum 12S ribosomal RNA gene, mitochondrial gene for mitochondrial RNA, partial sequence.
☐ Select all		Download selected

How to search for more information or help?

E-mail us at

info@vectorbase.org