Genome assembly and annotation of the cricket *Gryllus bimaculatus*

***Guillem Ylla*[1], Taro Nakamura[2], Sumihare Noji[3], Taro Mito[4] & Cassandra Extavour[1]**

[1] Department of Organismic and Evolutionary Biology & Department of Molecular and Cellular Biology, Harvard University.
[2] Division of Evolutionary Developmental Biology, National Institute for Basic Biology.
[3] Tokushima University.
[4] Division of Bioscience and Bioindustry, Tokushima University.

Gryllus bimaculatus

Most insect genomes available are from **Holometabolan** insects, which are insects that undergo complete metamorphosis like beetles and flies. In contrast, the **Hemimetabolans**, which undergo gradual metamorphosis, and represent the basal branches of the insect tree, have very few genomic resources available **(Figure 1)**.

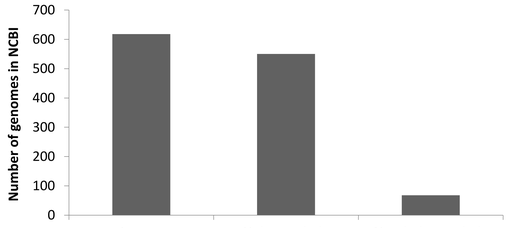


Figure 1: Number of genomes listed in NCBI for all insects, holometabolans ("Endopterygota") and non-holometabolans (insect genomes - holometabolans).

In addition, there is only one available annotated genome for the Orthopteran order, that of *Locusta migratoria*, which is the largest insect genome sequenced until the date with ~7Gb.

Figure 2: Photo of an adult specimen of *Gryllus bimaculatus*. Photo credit to Arpita Kulkarni and Leo Blondel.

Gryllus bimacualtus

- Order: Orthoptera.
- Aka. two-spotted cricket.
- Hemimetabolan insect.
- Experimental model for functional genetics, development and evolution.

Genome annotation Pipeline

Results

Assembly Stats

	Gbi	Lmi	Lko	Bger	Zna
Year	2019	2014	2019	2018	2014
Scaffolds	47,877	564,328	148,784	24,820	31,662
Len. (Gb)	1.66	6.91	1.6	2.04	0.49
Avg.scaff. (Kb)	34.630	12.24	10.72	82.08	15.32
N50 (Kb)	6,287	325	583	1,056	751
N's (Kb)	56,571	1,094,108	31,618	326,821	20,751
L50	71	3,607	753	576	194

Table 1: Assembly statatistics of our newly assembled *Gryllus bimaculatus* genome (Gbi) in comparison with recently published hemimetabolan genomes of *Locusta migratoira* (Lmi), *Blattella germanica* (Bger) and *Zootermopsis nevadensis* (Zna).

Assembly Completeness

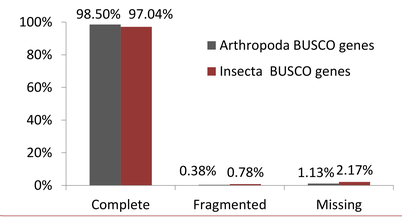


Figure 3: Percentage of complete, fragmented and missing BUSCO genes in the *G. bimaculatus* genome assembly. Analysis performed at arthropod level (1,066 genes) and insect level (1,658 genes).

Transposable Elements

Represent the 34% of the Genome

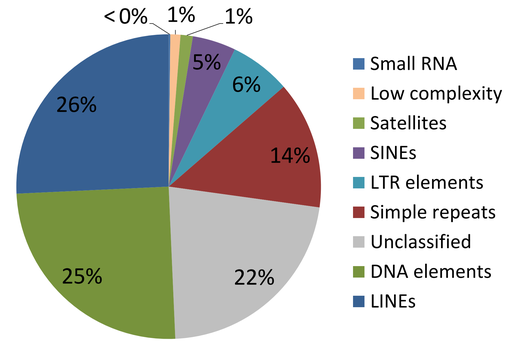


Figure 4: The ~34% of the *G. bimaculatus* genome is composed of transposable elements (TEs). The RepeatMasker software using a custom repeat library reported the percentage of genome accounted for each of the main TE families.

Gene Annotations

17,871 Protein Coding Genes
28,529 Transcripts

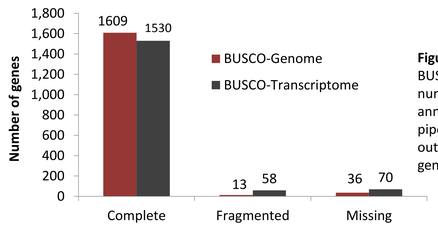


Figure 5: Number of Insect genes that BUSCO found in the genome versus number of insect genes found within the annotated transcripts. The annotation pipeline only failed to annotate 34 genes out of 1,658 (2%) of the BUSCO Insect genes.

Functional Annotations

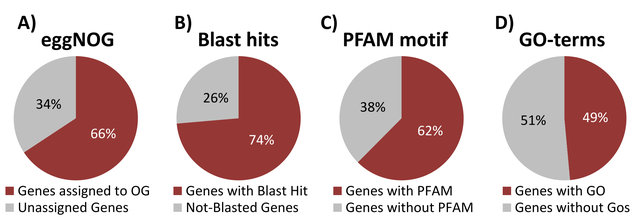
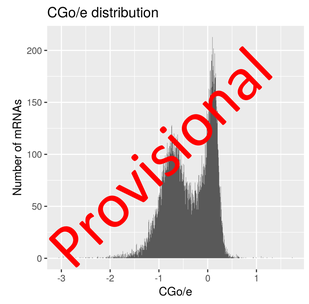
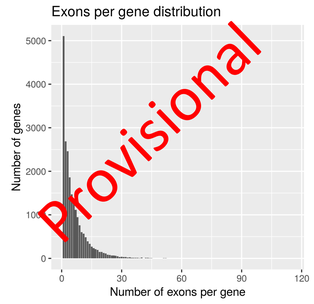


Figure 6: Percentage of genes annotated with; **A)** eggNOG Orthologous Group (OG) **B)** Significant Blastp hit (evalue<XXX) **C)** at least one PFAM motif **D)** At least one GO-term.

Preliminary results

Genome Methylation Evidences





Guillem_ylla@fas.harvard.edu

www.researchgate.net/profile/Guillem_Ylla
www.guillemylla.weebly.com
www.linkedin.com/in/gylla/

References:
A
A
A

Department of Organismic & Evolutionary Biology