

README for *Oncopeltus fasciatus* 454 transcriptome sequences.

Contact bewencampen at oeb dot harvard dot edu with any questions.

Note: These data will soon be searchable on an interactive website, but for the time being we thought a flat file would be useful to other researchers.

A complete description of how these data were generated can be found in this publication:

Ewen-Campen et al. (2011) BMC Genomics 2011, 12:61.

Description

-This .fasta file

'*Oncopeltus_454transcriptome_BLASTnames_01_2011.fasta*' contains gene predictions from the maternal and embryonic transcriptome of the milkweed bug, *Oncopeltus fasciatus*. Both assembled sequences (isotigs/contigs) and singletons are included. The raw reads were sequenced using 454 pyrosequencing and assembled using the 'cdna' algorithm of Newbler 2.3.

-The name of each sequence contains the top 20 BLAST hits against the RefSeq protein database with an e-value cutoff of $1e-10$. You can therefore search through the sequences with any text editor for genes-of-interest.

-In order to keep this file relatively small, we have excluded some sequences:

-We excluded sequences that did not get a BLAST hit at this cutoff value. If you'd like those sequences too, please email bewencampen at oeb dot harvard dot edu.

-We have also only included one predicted transcript per gene (the longest isotig per isogroup, aka the longest predicted transcript per gene), since nearly all transcripts have the same BLAST hits. If you want the alternative isoforms, please email bewencampen at oeb dot harvard dot edu.