

Genome sequences

WGS sequence
(454 GS-FLX Titanium)
5,082,983 reads (QV > 20)
1,925 Mbp (coverage 5x)

Celera assembler 6.1 (CABOG)
(de novo assembly)

Assembled WGS sequence
88,530 contigs (N50: 2,273)
246,244 degenerate contigs
106,455 singletons

Repeat Scout
(de novo repeat ident.)

EvidenceModeler
(gene prediction)

Putative repeat sequence
6,310 sequences

Predicted gene CDS
34,890 sequences

Used for repeat masking, etc.

Putative unknown gene set
(no homologous gene/domain)
39,781 sequences

Transcript sequences

RNA-seq
(HiSeq 2000)
201,115,780 reads
(101bp single-read)
20,313 Mbp

Trinity
(de novo assembly)

Assembled RNA-seq
147,370 contigs (N50: 515)

CLOBB2 (clustering) & cap3 (assemble)

Unigene
84,570 sequences

Annotation (BLAST, HMMER3, Blast2GO)

CLOBB2 (clustering) & extract representative sequences

Putative gene set
32,800 sequences

EST (ABI 3730)
35,618 reads (QV > 20)
18.7Mbp
midgut: 12,406 reads
egg: 6,904 reads
testis: 16,308 reads
EST/mRNA (NCBI DB)
1,722 reads
1.1Mbp