

Table S4. Gene repertoire of BSF genome

Gene prediction	
Method	# predicted genes
<i>Ab initio</i> - AUGUSTUS	16,571
<i>Ab initio</i> - SNAP	39,365
<i>Ab initio</i> - Genscan	29,987
Transcriptome - Stringtie	15,324
Homolog – Fruit fly	13,068
Homolog – House fly	14,338
Homolog – Tsetse fly	13,308
Homolog – Stable fly	14,147
Homolog – Sand fly	12,561
Homolog – Anopheles mosquito	12,769
Homolog – UniProt	15,377
Combined - Maker	15,054
Combined – GLEAN	17,033
OGS (filtered GLEAN)	16,770
Gene annotation	
Database	# genes with hit
Transcriptome evidence	16,650 (99.3%)
Dipteran homology	13,977 (83.3%)
Drosophila homology	13,082 (78.0%)
KEGG KO	9596 (57.2%)
GeneOntology	8837 (52.7%)
InterPro domain	12,022 (71.7%)
Pfam domain	11,477 (68.4%)
NCBI RefSeq (invertebrate)	14,055 (83.8%)
UniProt (Diptera)	14,214 (84.8%)
UniRef50	14,218 (84.8%)
Any homology	14,300 (85.3%)
Statistics based on orthology grouping	
Categories	Gene numbers
Universal single copy	2,950
Mutiple copy	6,140
Species-specific	1,818
Species-specific duplicated	1,798
Brachycera only	797
Homology	587
Patch	2,680