Supplementary Figure Legends

Figure S1

Comparison of noise in technical replicates to Poisson distribution. Histogram of deviation from the mean, scaled by the expected standard deviation for four PF technical replicates. The expected mean was approximated as the weighted scaled count of the four replicates.

Figure S2

Comparison of biological replicates. Histogram of the deviation of tag counts between pairs of biological replicates, averaged over technical replicates where possible, scaled by the expected standard deviation based on only technical noise (i.e. counting statistics). Magenta line: expected histogram based on only technical noise. Red line: Gaussian fit of histogram. A: Comparison of two wild-type BF replicates; B: comparison of the first wild-type BF replicate to the one SM BF run; C: comparison of the second "no marker" BF replicate to "Single marker" BF, D: comparison of two PF biological replicates. Specifically, we used the test quantity $T_{ik} = \left(\tilde{X}_i - \tilde{X}_k\right) \Big/ \sqrt{\sigma_i^2 + \sigma_k^2}$, with $\sigma_i^2 = \left(\sum_j X_{ij}\right) \Big/ \sum_j a_j^{-2}$. The peaks at $T_{ik} = 0$ are caused by the discrete nature of the actual counts.

Figure S3

Comparison of RNA-seq with results from Koumandou et al., 2008. N-fold regulation for a selection of genes as determined by Koumandou et al., 2008 using microarray (upper panel) and real time PCR (lower panel) was plotted against RNA-seq data.

Figure S4

Alignment of sequenced tags that identify the location of introns in Tb927.3.3160 and Tb927.8.1510. Tag sequences in red are as sequenced: tag sequences in green are the reverse complement of the actual sequences.

Figure S1

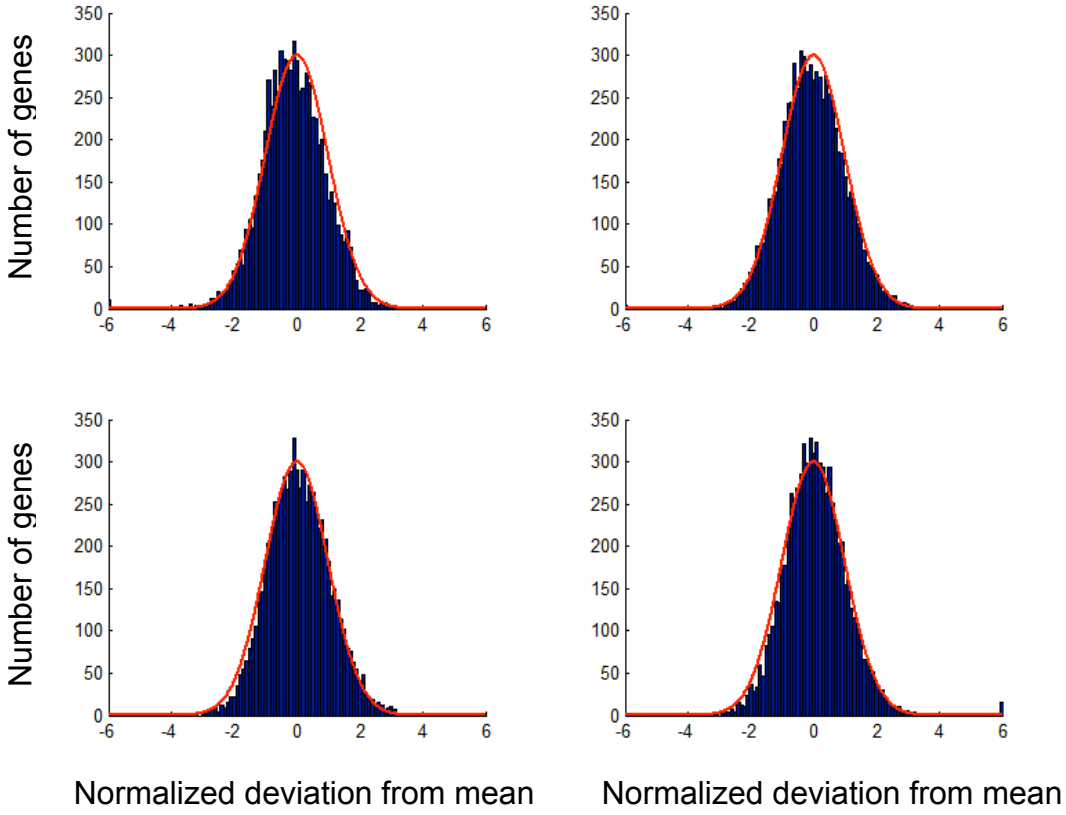

Number of genes
Number of genes
Normalized deviation from mean
Normalized deviation from mean

Figure S2

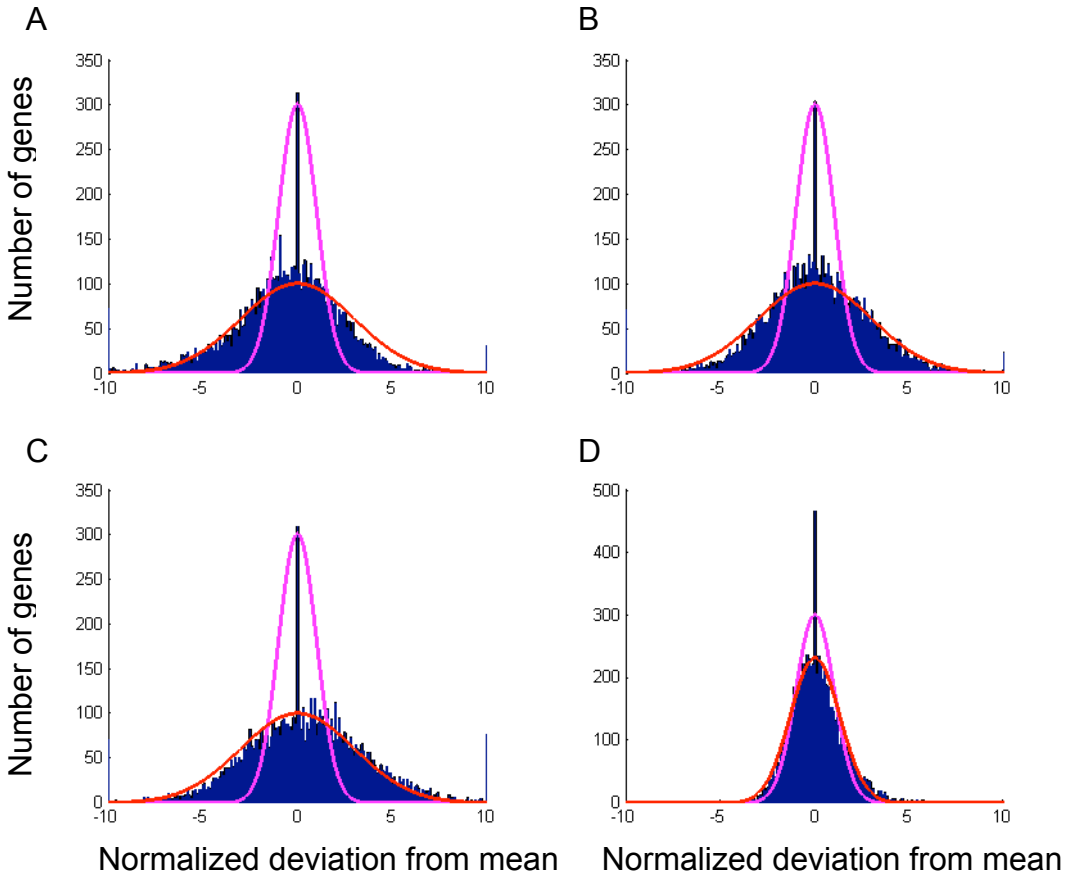
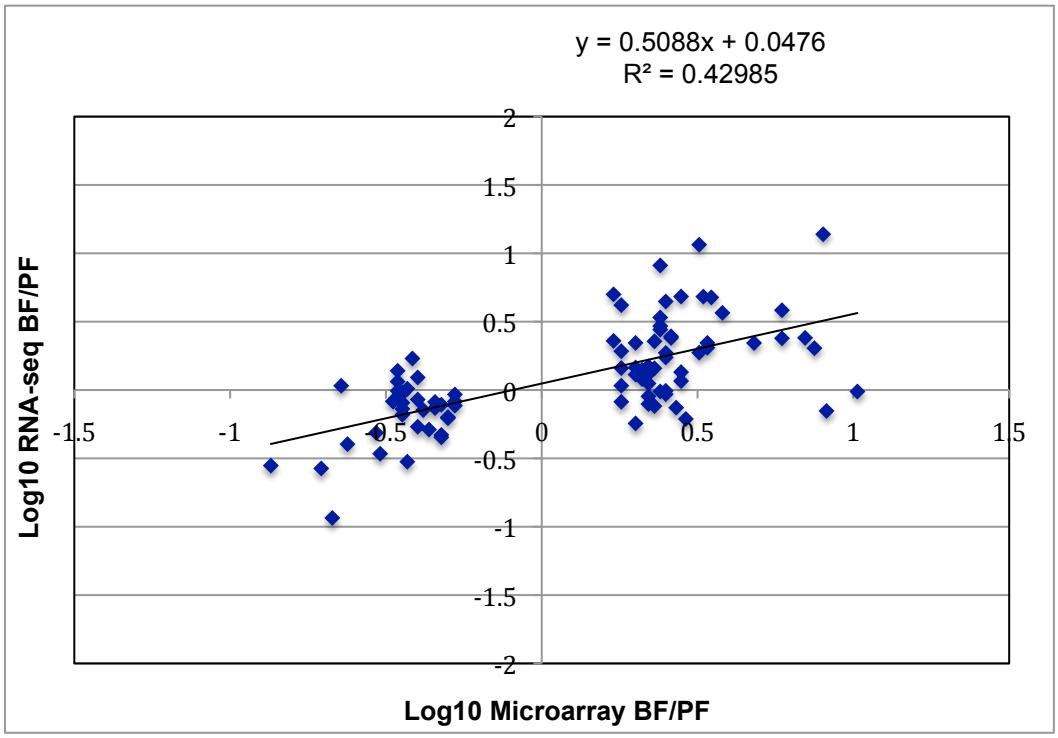

A
B
C
D
Number of genes
Number of genes
Normalized deviation from mean
Normalized deviation from mean

FIGURE S3


y = 0.5088x + 0.0476
R² = 0.42985
Log10 RNA-seq BF/PF
Log10 Microarray BF/PF
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5

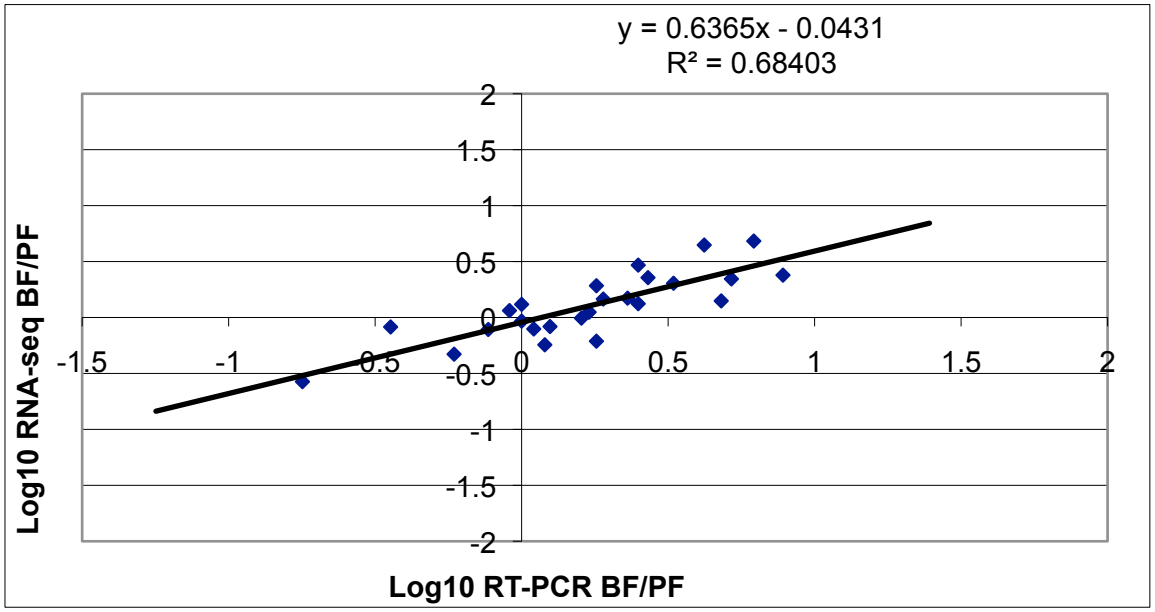

y = 0.6365x - 0.0431
R² = 0.68403
Log10 RNA-seq BF/PF
Log10 RT-PCR BF/PF
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

FIGURE S4

Tb927.3.3160: EXON sequences are black, INTRON blue, TAG sequences red or green (reverse complement of actual tag sequences)

```
ATGGAGTGGATGTATGGCCCAACGAAACCACTAGAGGTTCCCCGTCCACAAGACCATGACTACGACGAGTCTGAGATGCTTTACGGCGTGTTGGCCGAAT
GCCCTTCCATTAGTCCCAATCCGGTTTTAACGCTAGTGGAGTCGATGGCGCTTCGCATTGTGCAGCGGCACTCCCAAAACACTCAGGAACAATGGGCACG
TGCCTACCCATTTGGTTCATGCGGTCTTTCCGCTTCAGTTGCAGAGAGTGATCTCGATGTgtatggtgagttctttccttaatccctcctttctactccc
                                  GTTGCAGAGAGTGATCTCGATGT----------------------------------------
                              TTTCCGCTGCAGTTGCAGAGAGTGATCTCGATGT----------------------------------------
                             CTTTCCGCTGCAGTTGCAGAGAGTGATCTCGATGT----------------------------------------
                    TCATGCGGTCTTTCCGCTGCAGTTGCAGAGAGTGATCTCGATGT----------------------------------------
                                    TTGCAGAGAGTGATCTCGATGT----------------------------------------
                         CGGTCTTTCCGCTGCAGTTGCAGAGAGTGATCTCGATGT----------------------------------------
                     CATGCGGTCTTTCCGCTGCAGTTGCAGAGAGTGATCTCGATGT----------------------------------------
                                  AGTTGCAGAGAGTGATCTCGATGT----------------------------------------
                                    TTGCAGAGAGTGATCTCGATGT----------------------------------------
                       GCGGTCTTTCCGCTGCAGTTGCAGAGAGTGATCTCGATGT----------------------------------------
                                 CAGTTGCAGAGAGTGATCTCGATGT----------------------------------------
                               CGCTGCAGTTGCAGAGAGTGATCTCGATGT----------------------------------------
           CATTTGGTTCATGCGGTCTTTCCGCTGCAGTTGCAGAGAGTGATCTCGATGT----------------------------------------

ttcttcttttctcctctctttgtatatgcatctcccctcttcttcctctctttctttactattcttttccttttggtgttctgggtccttttttcatttt
atgttggatcatgtgaacccttaaaggaacgtgcttctgctttacgtgtaatcttttccctgctcttctcccgtcctcttctttttcttcccctcctcaca
tcgttccctgttcctcgtttcacgtgccgccatcgccttgtcgacgtctggacgcgttcatcttgccgtggttgtcgaaccttcgtggctgatctgcgtg
ctgtttcgctttcactgtctatacatttgtgtttatttccaacattgaccattgcttcttccttcgttaatgatgtgattatctaaatatgcatatacat
gtataatcatatatgtacacgtatacatttgtgtatatgtgtatattttgttgatgtgagtctgttcgtcgtcgtggcactcgtgacccgcacatggtac
tgacggcaaacacggatgctgattttccatttttttttttgtcaaaaccccgcaaccataccatttcaaccaacaacaatcacaacgccgaattacgttga
atttgtgaagCGCGCTGATTTGTCCATCTGTTTTAACGACTGATATCTTTTTCGATGAGTTTCCGAGGTTGTTATATAACTCTGTCGGTCCCGTGTCTGT
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGATATTTTTTTCGATGAGTTTC
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGATATCTTTTT
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGATATCTTTT
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTG
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGATATCTTTTTCGATGAGTTTCC
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGAGATC
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGA
----------CGCGCTGATTTGTACATCTGTTTTAACGACTGATATCTTTTTCGATGAGTTT
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGATATCTTTTTCGATGAGTTTCC
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGATAT
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGATATCTTTTTCGATGAGTT
----------CGCGCTGATTTGTCCATCTGTTTTAACGACTGATATCTTTTTCGAT
----------CGCGCTGATTTGTCCATCTGTTTT

AGTCGCTGCCAGAGTTCCCGTTGTTAAATTTGAGTACAGAGGCACAGCTGTCGATGTTGTGTTTGTTTCTGTTGGACTGCCACAGCCGCCTACGGAGGAG
CAGATGCTGGATGATTCGTTTCTACTGAAGGTGGCTCGAGAAACCCGACCCAGTGCGAACGGAATCCGTTTCACGTTTGAGATTAAGCGGCGACTTCCGG
TGCCGTACGATGTCTTCACTGCCGTTCTGAAGACAGTAAAGCTATGGGCGATGCGTCGGATGGTTTACGGCAATGTCTACACGTATCCGAACGGCGCGGT
GCTTGCCATCATGGTGGCTAGGGTTTGTCAGGTGCTTCCATCTTCTCATCCCAGTACTCTTCTTCGATTTTTCTTTCTTTTCTACACGCAATGGATGAGC
CGCCATGATCGTATTTCACCCGTGTATTTGACTGCGACTCTGGAGGGTCGCGGTCGCATTCCCGGGCTGCCAGACTCATGGAAGCCGTCGAAAGATAAAA
TTTGCTGCGATTTGTTTCCCGTGATCAGTCCGGCGTATCCATATGTGAACGATGCAAGTAGCGTTGGGCGATGTGGGCTGGAGGCGCTTTACAGCGAGAT
CACACGGGTGCAATGCATTCTGACGGAGGCGCGGACTCTCCCGTTGGAAGAGATGTGGGAACCGTATAGAATAGAGGAAGAGTATTCAACCTTTCTTGTT
GTGGGTGTAAGTTGTGCGGGCCACAGTATGGCGGAGACTGAGCAGGCGCTTTCTGTGTGGAGTTCCTATGTGGCCAGCAGGTTACGAATCCTCATATACT
CCATAGAGCGTCATGCACAGGCCCGTCCCTGTCCGCGTAAGATAAGGCCCAAAAAGGATAATCACGTGGGGGGTTCTTCATGTTTTTTGTCTGTCAATTT
CCTCATAGGTGTTAAGGCAAAGGAGGGGGGTCCCACCCCACAGCCGTCTCTGTTTACTGAGGCGTGTGGCGAGTTTCACCACGCTGTGAAGGAGGGATGT
AACAATGACTCCATTCCGTGGTCATTTCAGCGGAATGATCGTGCAATGCATTGGCCCCGTGTCAAATTTTTAGAGGTGCATCAGGTGCTGCACCTGTTAC
AAGACCTGAAGTAA
```

Tb927.8.1510

ATGTATGGGATGAACAGAGGAGGAGGCGGTGGATACCGCAACGACCGGATGGGTGCACTCGGTGCCAACCTACAGGATGTAAAGTGGGACCAGGTGAATG
TTGTGGCTCCTCAGTGGAACTATTACAAACCCCAGGAGCGTTGCAGTGACGAGGAGGTGGCACAATACATGCGCGACAACCACATTACTATTTATGGTGA
TTCCGTGCCCCAGCCAATGCTCAAGTTCTCTGACCTCGTTGCGCCGGACACCATTCATCAGGCATTCATAGACCTCGGCTACAAGAGCCCGACCCCAATT
CAGTCTATCGCCTGGCCAATCCTTCTCAACTCACGCGACCTTGTCGGCGTGGCGAAGACCGGATCCGGCAAGACCATGGGTTTCATGGTCCCTGCCGCCC
TTCATATAATGGCGCAGCCGCCCATCCGCGTGGGTGAGGGTCCCATTGCTTTGGTGCTCGCGCCGACGAGGGAACTTGCGGTGCAGATCGAAGAGGAGAC
GCGTAAGGTGCTACGACGTGTGCCGACCATCACGACGGCGTGTTTGTATGGCGGTACACCAAAGGGGCCGCAAATCCGTGCCCTTCGAGCTGGGGTACAT
GTCTGCATTGCAACTCCCGGCCGCCTCATCGACCTGCTGGAGATTCGCGCTGCCAATCTTCTGCGCGTGACGTTCCTTGTGCTTGACGAAGCGGACCGTA
TGCTTGATATGGGGTTCGAGGTGCAAATTCGAAAGATTTGCCAGCAGATTCGCAAGGATCGGCAGACGCTTATGTTCTCGGCAACGTGGCCCCAAGAAAT
AAGGAACCTCGCTGCTAGTTTTCAACGTGACTTCATCCGTGTTCATGTTGGATCTGAGGACCTAATTGCGAACGCTGATGTTACTCAGCACGTGAGTGTG
GTGGAAGATTACGACAAACAGCGGCGGCTGGAGGAGATTCTGCAAAAGGTTGGCAAGCAGAGGGTTCTCATCTTTGTAAAGACGAAGCGTACCGCTGATT
CCCTGCACCATTCGCTACAGCGCCTCATTGGAGGCTCAGTGATGGCCATCCATGGTGATAAGGAACAATCGCAGCGCGATTATGTGCTTGACCGCTTCCG
TCGCGACGAAAGGTCCGTGCTGGTAGCAACGGACGTTGCCGCCCGCG
GGCTTGACATTAAGAACCTCGATGTGGTGATTAACTTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTGgtatgtgaattgcaatccggtt
AAACATAGAAGATTACGTGCATCGTATTG----------
AAACATAGAAGATTACGTGCATCGTATTG----------
ACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
ATAGAAGATTACGTGCATCGTATTG----------
ATAGAAGATTACGTGCATCGTATTG----------
ATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
ATTAACTTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
CACAAACATAGAAGATTACGTGCATCGTATTG----------
CATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
CTTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
CTTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
TGATTAACTTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
TTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
TTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
CATGCCCACAAACATAGAAGATTACTTGCATCGTATTG----------
GTGATTAACTTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
GGTGATTAACTTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
TGTGATTAACTTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
TTCGACATGCCCACAAACATAGAAGATTACGTGCATCGTATTG----------
AAACATAGAAGATTACGTGCATCGTATTG----------
CACAAACATAGAAGATTACGTGCATCGTATTG----------
TCTAAACATAGAAGATTACGTGCATCGTATTG----------

gagttcaactacgtgctttcccctcttcgctttgcttccccttctctcaccccctctccactgcataggcgcaaaacataggcacattccacaacgccag
aagaaacacacacacacacacacacacacacacacacacacacacacacacacacacacatacacttacacgtatgtacagtggtgtgtaaacctatatatgggtttc
tgccactgtattcttctctgttttcggttgtgcgttttgttcttacccccgatttggttctcatgtcttccatgcacacgtgcagcaggcactaggggaa
ttagacgtaaatactggtaggttaacgcaataggaagcttctactgatttttttgatatatatatatatatatatatatgtatatgcgtatgtgtgtgctgcct
atctctgtatatagggtgtgacgcggggatagccaaagaagtagtgtgtttcgacgcgtgtgtaggatgtgcctttatgtttgtgtttgtgtatgtgtct
ttcccctccccgcctcctctctccgtggcagtcggctaacggcaggcgactcactttcgcttcagtgcagcgcgttgtgcacgcgcaatgggttagtggt
agttgccaggcgcttagggagtaaattgttgcacatctatcgttttggtacttcagttatacatgtgtgtgtgtggagatagtttctcccactttacgta
tgtagggatgcttccacgtgggaatatacgtgcgaacgtaccgtgcctgaagtagtgcggtagcaacaatccctttgttcgtgcatgtgtgcgtgtttgg
tacatctgaatgaatgaattctctgaagtggctatacgcaaggaagtgtcccataggtttgcaaggatgagggtcgctcgggcctgtacctgtggcacat
ttccctctgtgtgtgtgtgtatgcgtttttgttttgcgtcggccacagaaccagagttcacatgtaccattaatgcgcaagcacatatggcactccgctc
tctttccgcttgttggatacagagtctccccctgcgcggcgcagcgaagtgaggtagtacgctggtgtcgccggctgtcacgatgctcccttcttcaaca
acttttagaagttggttacttctgtacatttgctgaaaactttcttttctcttctcctttctctaatttcacttccccttgacattgccgttgctctcat
tgtgatgtggtgttgttgtgcgcaacgataccatttcaattggcgacatgttt
gacaaactgccgcacatcaatgtcaacttccttactgtcatatgaaaagGGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTACACATTTGTTAGCG
----------GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTACACATTTGTT
----------GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTACACATTTTTT
----------GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTA
----------GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTACACATTTGTTAGCG
----------GGCGTACCGGTCGTGCCGGGCAGGGTGGTGATGCTTAAAAATTTGTTAGCG
----------GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTACA
----------GGCGTACCGGTCGTGCCGGGCAGCGTG
----------GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTAAACATTT

--GGCGTACCGGTCGTGCCGGGCAGGGTGGTGATGCTTCC
--GGCGTACCGGTCGTGCCGGGCAGCGTGGTGAT
--GGCGTACCGGTCGTGCCGGGCAGAGTGGTGAT
--GGCGTACCGGTCGTGCCGGGCAGCG
--GGCGTACCGGTCGTGCCGGGCAGCGTGGGGATG
--GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATG
--GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTAC
--GGCGTACCGGTCGTGCCGGGCAGC
--GGCGTACCGGTCGTGCCGGGCAG
--GGCGTACCGGTCGTGCCGGGCAG
--GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATG
--GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTACACATTTGTT
--GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTACACATTT
--GGCGTACCGGTCGTGCCGGGCAGCGTGGTGATGCTTACACATTT

GTGCCGATCCGTCCAAGACTGTTCGTGACTTGATCGACATTCTTCGCCGCGCCAATCAGGAGATTCCACCAGGGCTTCACTCGCTTGCTGGCCGCGGTGG
CGGTGGCAGTGGCGGTGGCGGCTTCCGTGGGGGCCGTCGCGGAGGGTACGGTGGGGGCGGTGGCGGTTACGGCTACGGCCGCTCTGGTGCGCAGTACGGC
TCTGGGCCCACATACGGCAGGGATTACCGTGGCGCTCCATCCCACCATCAATCGGGTATTTCTGTTGGTGGCAATACATCTGATGCCTATTGGAGAGGGG
GAGCGCCACCGCCTGGTCAACAGCCTTACTATCCGGTGTGCGGAGGTAATAAACGGACCCGTGAGGTGTCCGGCGGCTTTCACTCAGATCCTCAGAACTC
GTACCGTCAGCAACCGACGGGTGGTTCTTATTAG

Genes reported to be upregulated in PF (Brems, Guilbride, Gundlesdodjir-Planck, Busold, Luu, Schanne, Hoheisel & Clayton (2005) The transcriptomes of Trypanosoma brucei Lister 427 and TREU927 bloodstream and procyclic trypomastigotes. Mol Biochem Parasitol 139, 163-172.) are largely consistent with our RNA-seq data

Highlighted fields indicate genes for which our RNA-seq data are inconsistent the data reported by Brems et al.

Group	Old Gene ID	New Gene ID	Description	RNA-seq BF/PF Ratio	RNA-seq p Value	Comments
Membrane	Tb10.6k15.3510		CRAM	not analyzed (repetitive)		
	Tb07.25D22.110	Tb927.7.6850	trans-sialidase	0.12	0.00	
	Tb08.29O9.350	Tb927.8.1610	major surface protease gp6	0.39	0.00	
	Tb05.29K2.440	Tb927.5.440	trans-sialidase	0.1	0.00	
	Tb05.3C6.130	Tb927.5.2170	META1-like	not analyzed (repetitive)		
	Tb09.211.4070		GPI anchored protein	0.39	0.01	
	Tb927.1.2310		Membrane protein	0.79	0.16	
	Tb927.2.5290		Mucin-like, 8 copies	not analyzed (repetitive)		
	Tb05.29K2.440	Tb927.5.440	Trans-sialidase-like	0.1	0.00	
	Tb11.0310		sequence orphan			
Transporter	Tb03.6N20.700	Tb927.3.590	Adenosine transporter	0.6	0.08	
	Tb04.3I12.190	Tb927.4.4730	Amino acid transporter	not analyzed (repetitive)		
	Tb07.10C21.10/20	Tb927.7.5940	hypothetical protein	not analyzed (repetitive)		
	Tb07.2F2.110-150	Tb927.7.5970	hypothetical protein	not analyzed (repetitive)		
	Tb08.4A8.350	Tb927.8.4700	Amino acid transporter	0.56	0.00	
	Tb10.6k15.1350		Pteridine transporter	0.27	0.00	
	Tb09.160.5480		Nucleoside transporter	0.12	0.00	
Glycosome	Tb11.02.4150		pyruvate phosphate dikinase	0.12	0.00	
	Tb10.61.0980		glycosomal malate dehydrogenase	0.14	0.00	
	Tb927.2.4210		glycosomal phosphoenolpyruvate carboxykinase,glycosomal protein P60	0.28	0.00	
	Tb03.26J7.860	Tb927.3.4500	fumarate hydratase, putative	0.24	0.00	
Mitochondrion	Tb07.30D13.350/60	Tb927.7.7420	ATP synthase alpha	0.32	0.00	
	Tb03.1J15.500	Tb927.3.1380	ATP synthase beta	0.18	0.00	
	Tb09.160.4310		Glutamate dehydrogenase	0.22	0.00	
	Tb09.160.1820		COX V cytochrome c oxidase subunit V	0.44	0.10	
	Tb03.1J15.440	Tb927.3.1410	COX VII cytochrome c oxidase subunit VII	0.41	0.10	
	Tb04.3I12.400	Tb927.4.4620	COX VIII cytochrome c oxidase subunit VIII	0.21	0.00	
	Tb11.01.4702		COX X cytochrome c oxidase subunit X	not analyzed (repetitive)		
	Tb08.26A17.490	Tb927.8.2530	Cytochrome c oxidase copper chaperon domain	1.32	0.84	
	Tb07.13M20.760	Tb927.7.2700	NADH-cytochrome b5 reductase	0.38	0.00	
	Tb09.160.4380		Succinate dehydrogenase	0.43	0.02	
	Tb09.211.4700		reiske iron-sulfur protein, mitochondrial precursor	0.28	0.00	
	Tb09.244.2620		hypothetical protein, conserved	0.7	0.07	
	Tb10.1650	Tb10.61.2880	Aconitase	0.24	0.00	
	Tb11.01.1740		2-oxoglutarate dehydrogenase E1 component, putative	0.4	0.00	
	Tb11.01.3550		2-oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase, putative	0.41	0.00	
	Tb11.02.0290		succinyl-coA:3-ketoacid-coenzyme A transferase, mitochondrial precursor, putative	0.88	0.15	
	Tb10.6k15.3250		succinyl-CoA ligase [GDP-forming] beta-chain, putative	0.5	0.00	

	Tb11.03.0870		mitochondrial carrier protein, putative	0.7	0.10	
	Tb05.28F8.230	Tb927.5.1060	mitochondrial processing peptidase, beta subunit, putative,metallo-peptidase, Clan ME, Family M16	0.20	0.00	
	Tb927.1.1690		RNA-editing complex protein,nuclease, putative,KREPB1	0.81	0.15	
	Tb07.27M11.900	Tb927.7.1550	RNA-editing complex protein,RNA-editing 3' terminal uridylyl transferase 2,KRET2	0.65	0.05	
	Tb11.03.0940		elongation factor, putative	0.73	0.04	
Enzyme						
	Tb11.02.2310		Prostaglandin F synthase	0.34	0.00	
	Tb11.02.5410		Ubiquitin activating enzyme	0.95	0.44	
	Tb927.1.2230		Calpain-like fragment	not analyzed (repetitive)		
	Tb07.26A24.1130	Tb927.7.4390	Threonine synthase	0.31	0.00	
	Tb10.389.1480		Metalloprotease	0.19	0.00	
	Tb927.2.2500		hypothetical protein, conserved	0.71	0.13	
	Tb927.2.2510		hypothetical protein, conserved	0.19	0.00	
	Tb03.27F10.890	Tb927.3.690	Protein kinase	0.93	0.35	
Cytoskeleton	Tb04.1D20.740	Tb927.4.3950	CAP5.5	not analyzed (repetitive)		
	Tb08.10J17.170	Tb927.8.4010	FLA1 flagellar adhesion protein	not analyzed (repetitive)		
	Tb07.22O10.840	Tb927.7.2640	Myosin-like	not analyzed (repetitive)		
	Tb10.406.0650		Microtubule-associated	not analyzed (repetitive)		
	Tb08.4A8.660		Kinesin	not analyzed (repetitive)		
	Tb11.01.7880		microtubule-associated protein,corset-associated protein 17	0.34	0.00	
	Tb04.5E12.340	Tb927.4.1110	hypothetical protein, conserved	1.04	0.58	Brems ("weak data")
Gene expression	Tb08.10J17.140	Tb927.8.4040	Endonuclease	not analyzed (repetitive)		
	Tb05.26K5.1010	Tb927.5.2570	Translation initiation factor	0.58	0.01	
	Tb10.70.4600		Elongation factor	0.71	0.04	
	Tb10.26.1080		*HSP83	0.35	0.00	
	Tb07.13M20.480	Tb927.7.2970	ATP-dependent DEAD/H RNA helicase, putative	1.23	0.93	Brems ("weak data")
	Tb11.01.8290		ATP-dependent DEAD/H RNA helicase, putative	1.22	0.80	Brems ("weak data")
	Tb11.02.2930		SNF2 DNA repair protein, putative,SNF2 family protein	0.8	0.20	
Other	Tb07.27M11.780	Tb927.7.1460	Importin	0.87	0.20	
Unknown	Tb08.26N11.180	Tb927.8.1790	hypothetical protein, conserved	0.61	0.04	
	Tb03.1J15.550	Tb927.3.1350	hypothetical protein, conserved	0.39	0.01	
	Tb09.244.2660		hypothetical protein, conserved	0.57	0.01	
	Tb07.21H15.100	Tb927.7.7220	hypothetical protein, conserved	1.06	0.62	
	Tb11.01.7880		microtubule-associated protein,corset-associated protein 17	0.34	0.00	
	Tb07.22O10.800	Tb927.7.2630	hypothetical protein, conserved	1.12	0.78	Brems ("weak data")
	Tb927.1.2260		calpain-like protein fragment, putative	not analyzed (repetitive)		
	Tb07.13M20.380	Tb927.7.2980	hypothetical protein, conserved	0.38	0.00	
	Tb05.3C6.130	Tb927.5.2170		not analyzed (repetitive)		
	Tb08.10J17.80	Tb927.8.4050	hypothetical protein, conserved	0.24	0.00	

Genes reported to be upregulated in BF (Brems et al 2005) are largely consistent with our RNA-seq data

Membrane	Tb927.5.370	75 kDa invariant surface glycoprotein, putative	not analyzed (repetitive)	
	Tb927.2.3270	65 kDa invariant surface glycoprotein, putative	not analyzed (repetitive)	
	Tb11.0360	could not identify gene		

	Tb11.33.0013		could not identify gene			
	Tb927.2.6000		glycosylphosphatidylinositol-specific phospholipase C,VSG lipase	13	1.00	
	Tb11.01.3790		Membrane protein	1.2	0.70	
Transporter	Tb11.01.0725-30		Cation transporter	not analyzed (repetitive)		
	Tb927.2.6150		P1-type nucleoside transporter	not analyzed (repetitive)		
	Tb03.26J7.120-160	Tb927.3.4070	13 trans-membrane domains	not analyzed (repetitive)		
	Tb04.3M17.20-69		could not identify gene			
	Tb927.2.6200		TbNT3 adenosine transporter 2	not analyzed (repetitive)		
Glycosome	Tb10.70.1370		Aldolase	5.03	1.00	
	Tb10.70.5800		Hexokinase	3.67	1.00	
	Tb03.3K10.320	Tb927.3.3270	ATP-dependent phosphofructokinase	5.13	1.00	
	Tb09.211.3540-3580		Glycerol kinase	0.99	0.34	
	Tb10.1550		could not identify gene			
	Tb927.1.3830		PGI glucose-6-phosphate isomerase	3	1.00	
Mitochondrion	Tb10.6k15.3640		Alternative oxidase	4.97	1.00	
Cytoskeleton	Tb10.61.1750		Kinesin	5.26	1.00	
	Tb11.01.3805		CAP15	not analyzed (repetitive)		
Gene expression	Tb927.2.3880		HNRNPH	2.57	0.99	
	Tb05.45E22.550	Tb927.5.4250	Histone H4	0.87	0.02	Brems ("weak data")
	Tb07.8P12.500	Tb927.7.320	RBP8	0.92	0.36	
Enzyme	Tb07.5F10.450	Tb927.7.4130	Glycosyl hydrolase	not analyzed (repetitive)		
	Tb10.70.5230		Metacaspase MCA4	1.01	0.52	
	Tb11.02.0730		Metacaspase	8.42	1.00	
Unknown	Tb05.25N21.20	Tb927.5.310	hypothetical protein	11.57	1.00	
	Tb05.26C7.190,220	Tb927.5.4570	hypothetical protein	not analyzed (repetitive)		
	Tb10.6k15.0940		hypothetical protein	19.23	1.00	
	Tb11.01.3810		hypothetical protein	0.45	0.38	
	Tb08.26A17.40	Tb927.8.2260	hypothetical protein	not analyzed (repetitive)		
	Tb07.26A24.60	Tb927.7.4920	hypothetical protein	2.17	1.00	
	Tb10.70.3090			0.75	0.03	Brems ("weak data")

ALTERNATIVE CDS COORDINATES SUMMARY based on splice-site analyses

These summary data for revised ATG predictions are organized into three categories. The detailed data (XLSX files) can be obtained on request and will be available on the lab website soon. The last page has a short technical note on how the pdf file was made.

CATEGORY 1

The tagging of a splice site downstream of the TrypDB-predicted ATG indicates that a downstream in-frame ATG must form the N-terminus of the protein. This situation could not have been predicted from the sequence alone. The number of sequence tag hits per gene overall ranges from 2 to 1,986. Data for genes with > 4 hits per gene or per SAS are considered highly reliable, but all the hits are unique for each sequenced tag, or quasi-unique where genes are present in multiple (usually but not always tandem) copies.

The table has been sorted first by the % of SAS hits that predict one specific internal ATG, then by chromosome and GeneID. In cases where there are only a few sequence tag hits per gene (or per SAS), the data should be viewed with caution and checked if it is an important gene for you.

There are 488 genes for which 609 SAS predict ORF start sites internal to the originally predicted ORF.

Of these 488, there are 321 genes where > 95% of the predicted SAS (2–1,986 sequence tag hits per gene) are inside the originally predicted ORF. Of these, 281 have more than 4 hits per gene, so are likely to be highly reliable. On the other hand, there could be other undetected SAS that are consistent with the originally predicted ATG, in which case there would be the potential for alternative forms of the protein with different initiation sites. This is the case for 167 of the genes (towards the bottom of the table) for which the first internal in-frame ATG is predicted but this does not account for 100% of the predictions: the others were consistent with the originally predicted ATG. There are 4 genes that have 2 entries in the table, where there are two predicted alternative internal ATGs. These are highlighted.

Chr	Gene	TrypDB	TriTrypDB	SAS per Gene	SAS % for ATG	Hits per Gene	ATG Diff	ORF Length True	ORF Start True	Abbreviated Description from TrypDB
1	Tb927.1.3810	Tb927.1.3810	Tb927.1.3810	1	100	55	+96	363	815129	hypothetical protein, conserved
1	Tb927.1.3840	Tb927.1.3840	Tb927.1.3840	1	100	2	+141	1830	818737	hypothetical protein, conserved
1	Tb927.1.5100	Tb927.1.5100	Tb927.1.5100	1	100	6	+15	1362	1018107	expression site-associated gene 2protein
1	Tb927.1.880	Tb927.1.880	Tb927.1.880	1	100	7	+24	13779	278162	hypothetical protein, conserved
2	Tb927.2.2390	Tb927.2.2390	Tb927.2.2390	1	100	74	+33	486	469891	hypothetical protein, conserved
2	Tb927.2.2950	Tb927.2.2950	Tb927.2.2950	1	100	4	+75	2868	560909	hypothetical protein, conserved
2	Tb927.2.3400	Tb927.2.3400	Tb927.2.3400	1	100	10	+147	1005	629941	hypothetical protein, conserved

2	Tb927.2.5080	Tb927.2.5080	Tb927.2.5080	1	100	10	+138	1410	898797	hypothetical protein, conserved
2	Tb927.2.5220	Tb927.2.5220	Tb927.2.5220	1	100	14	+108	1056	918441	hypothetical protein, conserved
2	Tb927.2.5900	Tb927.2.5900	Tb927.2.5900	2	100	14	+78	3174	1073575	hypothetical protein, conserved
2	Tb927.2.6070	Tb927.2.6070	Tb927.2.6070	1	100	52	+297	858	1096360	hypothetical protein, conserved
3	Tb927.3.1260	Tb927.3.1260	Tb927.3.1260	2	100	44	+39	582	312271	hypothetical protein, conserved
3	Tb927.3.1280	Tb927.3.1280	Tb927.3.1280	1	100	4	+276	1227	315296	hypothetical protein, conserved
3	Tb927.3.2130	Tb927.3.2130	Tb927.3.2130	1	100	14	+246	2241	562629	hypothetical protein, conserved
3	Tb927.3.2850	Tb927.3.2850	Tb927.3.2850	1	100	3	+222	615	737813	hypothetical protein, conserved
3	Tb927.3.3050	Tb927.3.3050	Tb927.3.3050	1	100	5	+60	3036	781790	hypothetical protein, conserved
3	Tb927.3.3200	Tb927.3.3200	Tb927.3.3200	1	100	51	+144	1053	828109	hypothetical protein, conserved
3	Tb927.3.3250	Tb927.3.3250	Tb927.3.3250	1	100	3	+33	1011	837376	hypothetical protein, conserved
3	Tb927.3.3350	Tb927.3.3350	Tb927.3.3350	1	100	2	+195	1557	860253	hypothetical protein, conserved
3	Tb927.3.3650	Tb927.3.3650	Tb927.3.3650	1	100	8	+135	1674	1022896	hypothetical protein, conserved
3	Tb927.3.3860	Tb927.3.3860	Tb927.3.3860	1	100	20	+99	3750	1071443	lipase, ,triacylglycerol lipase
3	Tb927.3.4030	Tb927.3.4030	Tb927.3.4030	1	100	29	+15	3642	1139026	hypothetical protein, conserved
3	Tb927.3.4120	Tb927.3.4120	Tb927.3.4120	1	100	16	+39	2034	1162824	hypothetical protein, conserved
3	Tb927.3.4190	Tb927.3.4190	Tb927.3.4190	1	100	12	+252	1938	1178152	endosomal integral membrane protein
3	Tb927.3.4210	Tb927.3.4210	Tb927.3.4210	1	100	10	+117	5313	1185220	hypothetical protein, conserved
3	Tb927.3.4350	Tb927.3.4350	Tb927.3.4350	1	100	3	+69	2325	1223193	hypothetical protein, conserved
3	Tb927.3.4400	Tb927.3.4400	Tb927.3.4400	1	100	6	+81	1083	1240558	hypothetical protein, conserved
3	Tb927.3.5170	Tb927.3.5170	Tb927.3.5170	1	100	12	+42	1242	1452148	hypothetical protein, conserved
3	Tb927.3.650	Tb927.3.650	Tb927.3.650	1	100	64	+69	189	161942	hypothetical protein, conserved
3	Tb927.3.840	Tb927.3.840	Tb927.3.840	1	100	2	+99	1068	196552	hypothetical protein, conserved
4	Tb927.4.1630	Tb927.4.1630	Tb927.4.1630	1	100	40	+99	2112	418548	ribosomal RNA processing protein 6
4	Tb927.4.1790	Tb927.4.1790	Tb927.4.1790	1	100	966	+153	1290	449754	ribosomal protein L3,
4	Tb927.4.1960	Tb927.4.1960	Tb927.4.1960	1	100	13	+81	1044	496987	hypothetical protein, conserved
4	Tb927.4.2340	Tb927.4.2340	Tb927.4.2340	1	100	9	+39	1539	606859	hypothetical protein, conserved
4	Tb927.4.2980	Tb927.4.2980	Tb927.4.2980	1	100	2	+165	660	790179	hypothetical protein, conserved

4	Tb927.4.3500	Tb927.4.3500	Tb927.4.3500	3	100	144	+198	531	905104	hypothetical protein, conserved
4	Tb927.4.3720	Tb927.4.3720	Tb927.4.3720	2	100	30	+174	690	938762	hypothetical protein, conserved
4	Tb927.4.4160	Tb927.4.4160	Tb927.4.4160	1	100	14	+57	2691	1103944	hypothetical protein, conserved
4	Tb927.4.4630	Tb927.4.4630	Tb927.4.4630	2	100	35	+132	366	1278372	hypothetical protein, conserved
4	Tb927.4.4680	Tb927.4.4680	Tb927.4.4680	1	100	12	+123	537	1291889	hypothetical protein, conserved
4	Tb927.4.4950	Tb927.4.4950	Tb927.4.4950	1	100	19	+165	1713	1360338	hypothetical protein, conserved
4	Tb927.4.510	Tb927.4.510	Tb927.4.510	2	100	24	+120	318	144483	hypothetical protein, conserved
4	Tb927.4.5360	Tb927.4.5360	Tb927.4.5360	1	100	6	+75	531	1459777	hypothetical protein, conserved
4	Tb927.4.570	Tb927.4.570	Tb927.4.570	1	100	5	+45	1752	166888	hypothetical protein, conserved
4	Tb927.4.720	Tb927.4.720	Tb927.4.720	1	100	107	+192	318	203329	hypothetical protein, conserved
4	Tb927.4.830	Tb927.4.830	Tb927.4.830	2	100	47	+192	291	233063	hypothetical protein, conserved
5	Tb927.5.1140	Tb927.5.1140	Tb927.5.1140	1	100	11	+90	1887	388120	choline/ethanolamine kinase,
5	Tb927.5.120	Tb927.5.120	Tb927.5.120	1	100	3	+102	774	12993	expression site-associated gene 9
5	Tb927.5.1900	Tb927.5.1900	Tb927.5.1900	3	100	35	+216	1833	589244	hypothetical protein, conserved
5	Tb927.5.2580	Tb927.5.2580	Tb927.5.2580	1	100	15	+276	888	828611	hypothetical protein, conserved
5	Tb927.5.2740	Tb927.5.2740	Tb927.5.2740	1	100	14	+186	1440	865523	hypothetical protein, conserved
5	Tb927.5.3180	Tb927.5.3180	Tb927.5.3180	2	100	9	+42	540	994488	hypothetical protein, conserved
5	Tb927.5.3680	Tb927.5.3680	Tb927.5.3680	1	100	3	+291	771	1155314	hypothetical protein, conserved
5	Tb927.5.3700	Tb927.5.3700	Tb927.5.3700	1	100	46	+87	879	1158374	hypothetical protein, conserved
5	Tb927.5.4100	Tb927.5.4100	Tb927.5.4100	1	100	6	+102	1443	1257284	hypothetical protein, conserved
5	Tb927.5.4340	Tb927.5.4340	Tb927.5.4340	1	100	2	+93	1758	1296030	hypothetical protein, conserved
5	Tb927.5.660	Tb927.5.660	Tb927.5.660	1	100	9	+267	1206	230069	hypothetical protein
5	Tb927.5.920	Tb927.5.920	Tb927.5.920	2	100	15	+165	1350	307331	hypothetical protein, conserved
6	Tb927.6.1230	Tb927.6.1230	Tb927.6.1230	1	100	9	+96	1251	467087	kinetoplastid-specific phosphatase
6	Tb927.6.1280	Tb927.6.1280	Tb927.6.1280	2	100	47	+81	900	472687	translation initiation factor eIF-2B alpha
6	Tb927.6.1420	Tb927.6.1420	Tb927.6.1420	1	100	47	+315	432	508600	hypothetical protein, conserved
6	Tb927.6.1440	Tb927.6.1440	Tb927.6.1440	2	100	17	+132	777	512568	hypothetical protein, conserved
6	Tb927.6.1680	Tb927.6.1680	Tb927.6.1680	1	100	11	+99	1575	569285	zinc finger protein,

6	Tb927.6.1860	Tb927.6.1860	Tb927.6.1860	1	100	60	+204	825	620391	hypothetical protein, conserved
6	Tb927.6.2150	Tb927.6.2150	Tb927.6.2150	1	100	17	+156	1644	685795	cell division cycle protein 16,
6	Tb927.6.2480	Tb927.6.2480	Tb927.6.2480	1	100	52	+123	1038	757526	chaperone protein DNAJ,
6	Tb927.6.2580	Tb927.6.2580	Tb927.6.2580	1	100	5	+279	1815	785293	hypothetical protein, conserved
6	Tb927.6.2730	Tb927.6.2730	Tb927.6.2730	1	100	14	+177	5358	817710	hypothetical protein, conserved
6	Tb927.6.3200	Tb927.6.3200	Tb927.6.3200	1	100	29	+252	336	951442	hypothetical protein, conserved
6	Tb927.6.3210	Tb927.6.3210	Tb927.6.3210	1	100	28	+144	534	952233	hypothetical protein, conserved
6	Tb927.6.3620	Tb927.6.3620	Tb927.6.3620	1	100	4	+42	1230	1081287	zinc finger protein,
6	Tb927.6.3700	Tb927.6.3700	Tb927.6.3700	1	100	31	+60	771	1104213	hypothetical protein, conserved
6	Tb927.6.4120	Tb927.6.4120	Tb927.6.4120	3	100	12	+234	1182	1191616	GTPase activating protein,
6	Tb927.6.4150	Tb927.6.4150	Tb927.6.4150	2	100	26	+162	642	1197079	hypothetical protein, conserved
6	Tb927.6.4170	Tb927.6.4170	Tb927.6.4170	1	100	5	+108	1194	1200553	hypothetical protein, conserved
6	Tb927.6.4190	Tb927.6.4190	Tb927.6.4190	1	100	3	+9	1647	1204255	hypothetical protein, conserved
6	Tb927.6.4200	Tb927.6.4200	Tb927.6.4200	1	100	42	+315	1335	1205935	hypothetical protein, conserved
6	Tb927.6.4330	Tb927.6.4330	Tb927.6.4330	2	100	39	+99	543	1225722	hypothetical protein, conserved
6	Tb927.6.4350	Tb927.6.4350	Tb927.6.4350	2	100	17	+114	864	1228040	ribosomal RNA assembly protein,
6	Tb927.6.4370	Tb927.6.4370	Tb927.6.4370	1	100	9	+90	1521	1231363	eukaryotic translation initiation factor 3
6	Tb927.6.4600	Tb927.6.4600	Tb927.6.4600	1	100	19	+129	2901	1287787	pre-mRNA splicing RNA helicase
6	Tb927.6.4960	Tb927.6.4960	Tb927.6.4960	1	100	34	+156	786	1366188	zinc finger-domain protein,
6	Tb927.6.880	Tb927.6.880	Tb927.6.880	1	100	13	+216	1395	379018	hypothetical protein, conserved
7	Tb927.7.1700	Tb927.7.1700	Tb927.7.1700	2	100	12	+276	1659	423088	hypothetical protein, conserved
7	Tb927.7.1770	Tb927.7.1770	Tb927.7.1770	1	100	5	+12	1818	434546	hypothetical protein, conserved
7	Tb927.7.190	Tb927.7.190	Tb927.7.190	2	100	26	+174	2058	34206	thimet oligopeptidase A, putativ
7	Tb927.7.270	Tb927.7.270	Tb927.7.270	1	100	15	+201	1113	47961	ribosome biogenesis protein,
7	Tb927.7.3310	Tb927.7.3310	Tb927.7.3310	1	100	5	+84	987	854846	hypothetical protein, conserved
7	Tb927.7.340	Tb927.7.340	Tb927.7.340	1	100	2	+105	984	62510	hypothetical protein, conserved
7	Tb927.7.3870	Tb927.7.3870	Tb927.7.3870	1	100	6	+63	771	1010286	hypothetical protein, conserved
7	Tb927.7.4220	Tb927.7.4220	Tb927.7.4220	1	100	25	+15	3318	1121347	hypothetical protein, conserved

7	Tb927.7.4970	Tb927.7.4970	Tb927.7.4970	1	100	29	+120	1146	1318106	glutamine synthetase,
7	Tb927.7.5700	Tb927.7.5700	Tb927.7.5700	1	100	28	+255	963	1540476	hypothetical protein, conserved
7	Tb927.7.5740	Tb927.7.5740	Tb927.7.5740	1	100	12	+207	1500	1548341	hypothetical protein, conserved
7	Tb927.7.5890	Tb927.7.5890	Tb927.7.5890	1	100	9	+231	2001	1582960	hypothetical protein, conserved
7	Tb927.7.720	Tb927.7.720	Tb927.7.720	1	100	10	+165	5214	163998	hypothetical protein, conserved
8	Tb927.8.1270	Tb927.8.1270	Tb927.8.1270	1	100	37	+228	1461	402970	hypothetical protein, conserved
8	Tb927.8.1730	Tb927.8.1730	Tb927.8.1730	1	100	27	+69	384	572989	hypothetical protein, conserved
8	Tb927.8.1800	Tb927.8.1800	Tb927.8.1800	1	100	15	+243	1506	591487	hypothetical protein, conserved
8	Tb927.8.2210	Tb927.8.2210	Tb927.8.2210	2	100	31	+303	807	677937	pteridine reductase
8	Tb927.8.2230	Tb927.8.2230	Tb927.8.2230	1	100	19	+114	1023	684928	hypothetical protein, conserved
8	Tb927.8.2870	Tb927.8.2870	Tb927.8.2870	1	100	4	+558	2247	870833	hypothetical protein, conserved
8	Tb927.8.3100	Tb927.8.3100	Tb927.8.3100	3	100	58	+72	1500	929087	coronin,
8	Tb927.8.3320	Tb927.8.3320	Tb927.8.3320	3	100	250	+201	1209	997489	hypothetical protein, conserved
8	Tb927.8.3860	Tb927.8.3860	Tb927.8.3860	1	100	15	+24	1251	1151744	hypothetical protein, conserved,WD-repeat
8	Tb927.8.4030	Tb927.8.4030	Tb927.8.4030	1	100	33	+60	639	1200385	class I transcription factor A, subunit 5a (CITFA-5a)
8	Tb927.8.4080	Tb927.8.4080	Tb927.8.4080	1	100	28	+60	639	1210003	class I transcription factor A (CITFA-5a)
8	Tb927.8.4130	Tb927.8.4130	Tb927.8.4130	1	100	33	+60	636	1219615	class I transcription factor A (CITFA-5b)
8	Tb927.8.4140	Tb927.8.4140	Tb927.8.4140	2	100	4	+33	1215	1221082	hypothetical protein, conserved
8	Tb927.8.4490	Tb927.8.4490	Tb927.8.4490	1	100	10	+90	1266	1328621	hypothetical protein
8	Tb927.8.5810	Tb927.8.5810	Tb927.8.5810	1	100	3	+57	858	1710893	mitochondrial carrier protein,
8	Tb927.8.5980	Tb927.8.5980	Tb927.8.5980	1	100	12	+48	2412	1751038	TFIIH complex helicase
8	Tb927.8.5990	Tb927.8.5990	Tb927.8.5990	1	100	26	+342	1575	1753999	hypothetical protein, conserved
8	Tb927.8.6960	Tb927.8.6960	Tb927.8.6960	1	100	8	+75	1158	2006213	hypothetical protein, conserved
8	Tb927.8.7010	Tb927.8.7010	Tb927.8.7010	1	100	38	+84	609	2017120	chaperone protein DNAJ,
8	Tb927.8.7190	Tb927.8.7190	Tb927.8.7190	1	100	10	+153	939	2065377	hypothetical protein
8	Tb927.8.740	Tb927.8.740	Tb927.8.740	1	100	6	+261	159	206277	nucleolar RNA-binding protein, truncated
8	Tb927.8.8070	Tb927.8.8070	Tb927.8.8070	2	100	10	+72	1581	2388868	hypothetical protein, conserved

8	Tb927.8.8110	Tb927.8.8110	Tb927.8.8110	2	100	11	+72	1728	2396775	hypothetical protein, conserved
8	Tb927.8.920	Tb927.8.920	Tb927.8.920	1	100	29	+36	483	286356	ubiquitin-protein ligase
9	Tb09.160.0620	Tb09.160.0620	Tb09.160.0620	1	100	41	+252	672	384124	peroxisomal membrane protein 4,
9	Tb09.160.0890	Tb09.160.0890	Tb09.160.0890	1	100	11	+270	1245	469540	hypothetical protein, conserved
9	Tb09.160.1020	Tb09.160.1020	Tb09.160.1020	2	100	20	+282	2190	498589	hypothetical protein, conserved
9	Tb09.160.2210	Tb09.160.2210	Tb09.160.2210	1	100	59	+87	555	672585	glutaredoxin-like protein
9	Tb09.160.3490	Tb09.160.3490	Tb09.160.3490	1	100	7	+51	669	842669	hypothetical protein, conserved
9	Tb09.160.3560	Tb09.160.3560	Tb09.160.3560	1	100	15	+198	1158	850860	hypothetical protein, conserved
9	Tb09.160.3730	Tb09.160.3730	Tb09.160.3730	1	100	52	+150	1734	867682	glutaminyl-tRNA synthetase,
9	Tb09.160.3910	Tb09.160.3910	Tb09.160.3910	3	100	16	+51	657	888993	hypothetical protein, conserved
9	Tb09.160.4690	Tb09.160.4690	Tb09.160.4690	1	100	3	+192	255	1007470	hypothetical protein, conserved
9	Tb09.160.5040	Tb09.160.5040	Tb09.160.5040	1	100	48	+216	204	1078079	hypothetical protein, conserved
9	Tb09.160.5130	Tb09.160.5130	Tb09.160.5130	1	100	27	+129	1485	1093818	hypothetical protein, conserved
9	Tb09.160.5180	Tb09.160.5180	Tb09.160.5180	2	100	12	+246	1104	1098892	hypothetical protein, conserved
9	Tb09.160.5310	Tb09.160.5310	Tb09.160.5310	1	100	18	+342	552	1124814	hypothetical protein, conserved
9	Tb09.211.0200	Tb09.211.0200	Tb09.211.0200	1	100	11	+111	1497	1301035	hypothetical protein, conserved
9	Tb09.211.0210	Tb09.211.0210	Tb09.211.0210	1	100	4	+150	1296	1302639	atypical dual specificity phosphatase
9	Tb09.211.0800	Tb09.211.0800	Tb09.211.0800	1	100	5	+135	1146	1405571	hypothetical protein, conserved
9	Tb09.211.0990	Tb09.211.0990	Tb09.211.0990	1	100	23	+150	774	1445866	hypothetical protein, conserved
9	Tb09.211.1270	Tb09.211.1270	Tb09.211.1270	1	100	19	+81	705	1491509	hypothetical protein, conserved
9	Tb09.211.1600	Tb09.211.1600	Tb09.211.1600	1	100	58	+81	867	1540043	hypothetical protein, conserved
9	Tb09.211.1820	Tb09.211.1820	Tb09.211.1820	1	100	14	+18	6723	1587284	DNA polymerase epsilon catalytic subunit
9	Tb09.211.1850	Tb09.211.1850	Tb09.211.1850	4	100	31	+189	1032	1593590	hypothetical protein, conserved
9	Tb09.211.2300	Tb09.211.2300	Tb09.211.2300	1	100	35	+282	2106	1678138	ATP-dependent DEAD/H RNA helicase
9	Tb09.211.2620	Tb09.211.2620	Tb09.211.2620	3	100	47	+150	438	1737006	hypothetical protein, conserved
9	Tb09.211.2680	Tb09.211.2680	Tb09.211.2680	1	100	16	+234	1773	1747431	hypothetical protein, conserved
9	Tb09.211.2810	Tb09.211.2810	Tb09.211.2810	1	100	68	+150	612	1770240	hypothetical protein, conserved
9	Tb09.211.3290	Tb09.211.3290	Tb09.211.3290	1	100	2	+207	399	1858531	hypothetical protein, conserved

9	Tb09.211.3330	Tb09.211.3330	Tb09.211.3330	3	100	146	+105	1233	1867992	cystathione gamma lyase,
9	Tb09.211.3730	Tb09.211.3730	Tb09.211.3730	2	100	15	+138	3300	1962089	hypothetical protein, conserved
9	Tb09.211.4150	Tb09.211.4150	Tb09.211.4150	2	100	14	+192	1389	2047339	hypothetical protein, conserved
9	Tb09.211.4290	Tb09.211.4290	Tb09.211.4290	1	100	11	+357	459	2083049	hypothetical protein, conserved
9	Tb09.211.4360	Tb09.211.4360	Tb09.211.4360	1	100	27	+93	2727	2097146	hypothetical protein, conserved
9	Tb09.211.4550	Tb09.211.4550	Tb09.211.4550	1	100	340	+165	495	2136533	60S ribosomal protein L12,
10	Tb10.05.0040	Tb10.05.0040	Tb10.05.0040	1	100	31	+219	2832	3252326	hypothetical protein, conserved,
10	Tb10.100.0150	Tb10.100.0150	Tb10.100.0150	1	100	5	+105	234	81558	hypothetical protein, conserved
10	Tb10.100.0190	Tb10.100.0190	Tb10.100.0190	1	100	35	+81	1677	86582	pumillio RNA binding protein,
10	Tb10.100.0200	Tb10.100.0200	Tb10.100.0200	1	100	22	+183	396	91041	hypothetical protein, conserved
10	Tb10.20.0130	Tb10.20.0130	Tb10.20.0130	1	100	6	+75	885	2059677	legume-like lectin,
10	Tb10.26.0570	Tb10.26.0570	Tb10.26.0570	1	100	22	+123	1173	2706162	hypothetical protein, conserved
10	Tb10.26.0910	Tb10.26.0910	Tb10.26.0910	2	100	12	+66	2934	2651263	hypothetical protein, conserved
10	Tb10.389.0270	Tb10.389.0270	Tb10.389.0270	1	100	10	+252	1158	3138542	actin related protein 2/3 complex
10	Tb10.389.0610	Tb10.389.0610	Tb10.389.0610	1	100	3	+24	2793	3067008	hypothetical protein, conserved
10	Tb10.389.1330	Tb10.389.1330	Tb10.389.1330	2	100	40	+156	1449	2937200	membrane transporter protein,
10	Tb10.389.1340	Tb10.389.1340	Tb10.389.1340	2	100	52	+321	777	2934377	hypothetical protein, conserved
10	Tb10.61.0100	Tb10.61.0100	Tb10.61.0100	2	100	25	+117	903	3900722	protein kinase,
10	Tb10.61.2720	Tb10.61.2720	Tb10.61.2720	1	100	68	+228	360	3360517	hypothetical protein, conserved
10	Tb10.6k15.0120	Tb10.6k15.0120	Tb10.6k15.0120	1	100	2	+510	867	2462539	hypothetical protein, conserved
10	Tb10.6k15.0290	Tb10.6k15.0290	Tb10.6k15.0290	2	100	27	+258	606	2426357	hypothetical protein, conserved
10	Tb10.6k15.0300	Tb10.6k15.0300	Tb10.6k15.0300	2	100	14	+240	1026	2425066	hypothetical protein, conserved
10	Tb10.6k15.1380	Tb10.6k15.1380	Tb10.6k15.1380	2	100	35	+57	2001	2205911	hypothetical protein, conserved
10	Tb10.6k15.1910	Tb10.6k15.1910	Tb10.6k15.1910	1	100	63	+96	768	2100023	hypothetical protein, conserved
10	Tb10.6k15.2360	Tb10.6k15.2360	Tb10.6k15.2360	2	100	34	+51	324	2006431	hypothetical protein, conserved
10	Tb10.6k15.2390	Tb10.6k15.2390	Tb10.6k15.2390	2	100	28	+213	1230	2003454	hypothetical protein, conserved
10	Tb10.6k15.3480	Tb10.6k15.3480	Tb10.6k15.3480	1	100	13	+270	687	1794645	hypothetical protein, conserved
10	Tb10.6k15.3740	Tb10.6k15.3740	Tb10.6k15.3740	1	100	2	+276	1641	1745095	hypothetical protein, conserved

10	Tb10.6k15.3750	Tb10.6k15.3750	Tb10.6k15.3750	2	100	12	+189	999	1743357	hypothetical protein, conserved
10	Tb10.70.0630	Tb10.70.0630	Tb10.70.0630	3	100	49	+180	1512	1568080	aldehyde dehydrogenase,
10	Tb10.70.0770	Tb10.70.0770	Tb10.70.0770	1	100	4	+198	1308	1535568	hypothetical protein, conserved
10	Tb10.70.1070	Tb10.70.1070	Tb10.70.1070	1	100	9	+66	1371	1469019	hypothetical protein, conserved
10	Tb10.70.1510	Tb10.70.1510	Tb10.70.1510	2	100	19	+144	1695	1381318	endonuclease/exonuclease/phosphatase
10	Tb10.70.2260	Tb10.70.2260	Tb10.70.2260	2	100	58	+99	1314	1229282	serine/threonine protein kinase
10	Tb10.70.2280	Tb10.70.2280	Tb10.70.2280	1	100	19	+120	1341	1226497	hypothetical protein, conserved
10	Tb10.70.2720	Tb10.70.2720	Tb10.70.2720	1	100	44	+36	939	1156708	mRNA capping methyltransferase,
10	Tb10.70.2750	Tb10.70.2750	Tb10.70.2750	1	100	18	+210	3156	1149961	hypothetical protein, conserved
10	Tb10.70.2760	Tb10.70.2760	Tb10.70.2760	1	100	21	+81	903	1146336	leucine carboxyl methyltransferase,
10	Tb10.70.3440	Tb10.70.3440	Tb10.70.3440	1	100	5	+156	2520	1029654	phosphatidylinositol-4-phosphate 5-kinase
10	Tb10.70.3840	Tb10.70.3840	Tb10.70.3840	1	100	41	+210	975	926750	hypothetical protein, conserved
10	Tb10.70.3860	Tb10.70.3860	Tb10.70.3860	2	100	22	+135	1170	921068	arginine N-methyltransferase,
10	Tb10.70.4150	Tb10.70.4150	Tb10.70.4150	1	100	9	+195	822	847231	hypothetical protein, conserved
10	Tb10.70.4430	Tb10.70.4430	Tb10.70.4430	1	100	32	+228	717	801615	hypothetical protein, conserved
10	Tb10.70.5350	Tb10.70.5350	Tb10.70.5350	1	100	5	+423	2751	608536	hypothetical protein, conserved
10	Tb10.70.6150	Tb10.70.6150	Tb10.70.6150	1	100	20	+105	729	463611	hypothetical protein, conserved
10	Tb10.70.6770	Tb10.70.6770	Tb10.70.6770	1	100	47	+198	252	343784	hypothetical protein, conserved
10	Tb10.70.7500	Tb10.70.7500	Tb10.70.7500	1	100	13	+39	861	180599	hypothetical protein, conserved
11	Tb11.01.0280	Tb11.01.0280	Tb11.01.0280	1	100	26	+126	2292	2264498	hypothetical protein, conserved
11	Tb11.01.0780	Tb11.01.0780	Tb11.01.0780	1	100	8	+24	342	2425298	hypothetical protein, conserved
11	Tb11.01.0890	Tb11.01.0890	Tb11.01.0890	1	100	5	+273	5187	2451477	hypothetical protein, conserved
11	Tb11.01.1080	Tb11.01.1080	Tb11.01.1080	1	100	8	+69	2703	2504312	tRNA pseudouridine synthase,
11	Tb11.01.1120	Tb11.01.1120	Tb11.01.1120	1	100	21	+72	1194	2513538	phosphonopyruvate decarboxylase-like
11	Tb11.01.1620	Tb11.01.1620	Tb11.01.1620	1	100	22	+129	363	2608128	hypothetical protein, conserved
11	Tb11.01.1880	Tb11.01.1880	Tb11.01.1880	2	100	28	+189	2409	2656719	hypothetical protein, conserved
11	Tb11.01.2580	Tb11.01.2580	Tb11.01.2580	1	100	2	+135	2541	2831958	hypothetical protein, conserved
11	Tb11.01.2640	Tb11.01.2640	Tb11.01.2640	1	100	4	+123	729	2844911	hypothetical protein, conserved

11	Tb11.01.2650	Tb11.01.2650	Tb11.01.2650	1	100	31	+117	1296	2846764	hypothetical protein, conserved
11	Tb11.01.2660	Tb11.01.2660	Tb11.01.2660	1	100	3	+36	1293	2848978	hypothetical protein, conserved
11	Tb11.01.3050	Tb11.01.3050	Tb11.01.3050	1	100	15	+141	1341	2955686	hypothetical protein, conserved
11	Tb11.01.3060	Tb11.01.3060	Tb11.01.3060	2	100	13	+303	1407	2957542	hypothetical protein, conserved
11	Tb11.01.3150	Tb11.01.3150	Tb11.01.3150	1	100	3	+51	2349	2973329	gamma-tubulin complex subunit
11	Tb11.01.3660	Tb11.01.3660	Tb11.01.3660	1	100	20	+183	2256	3098123	DNA helicase PIF1
11	Tb11.01.3860	Tb11.01.3860	Tb11.01.3860	2	100	65	+156	324	3157934	hypothetical protein, conserved
11	Tb11.01.3890	Tb11.01.3890	Tb11.01.3890	2	100	28	+60	876	3162417	hypothetical protein, conserved
11	Tb11.01.4280	Tb11.01.4280	Tb11.01.4280	2	100	73	+39	1185	3263627	hypothetical protein, conserved
11	Tb11.01.4450	Tb11.01.4450	Tb11.01.4450	1	100	13	+162	1005	3312088	cyclin 1,serine peptidase family S51, peptidase E,
11	Tb11.01.4530	Tb11.01.4530	Tb11.01.4530	1	100	8	+72	1056	3330430	hypothetical protein, conserved
11	Tb11.01.4950	Tb11.01.4950	Tb11.01.4950	1	100	22	+84	849	3525361	hypothetical protein, conserved
11	Tb11.01.4990	Tb11.01.4990	Tb11.01.4990	2	100	31	+90	504	3533932	hypothetical protein, conserved
11	Tb11.01.5160	Tb11.01.5160	Tb11.01.5160	1	100	14	+171	1635	3566678	hypothetical protein, conserved
11	Tb11.01.5180	Tb11.01.5180	Tb11.01.5180	2	100	23	+285	750	3569102	ubiquinone biosynthesis protein-like protein
11	Tb11.01.5190	Tb11.01.5190	Tb11.01.5190	1	100	41	+177	1065	3571842	hypothetical protein, conserved
11	Tb11.01.5400	Tb11.01.5400	Tb11.01.5400	1	100	6	+75	2088	3622015	hypothetical protein, conserved
11	Tb11.01.5560	Tb11.01.5560	Tb11.01.5560	1	100	2	+117	960	3648605	GDP-L-fucose synthetase,
11	Tb11.01.5800	Tb11.01.5800	Tb11.01.5800	1	100	6	+144	4371	3707260	calpain-like cysteine peptidase
11	Tb11.01.5850	Tb11.01.5850	Tb11.01.5850	1	100	3	+225	2961	3730467	hypothetical protein, conserved
11	Tb11.01.6210	Tb11.01.6210	Tb11.01.6210	1	100	2	+42	1212	3818943	procyclin-associated gene 2-like protein
11	Tb11.01.6230	Tb11.01.6230	Tb11.01.6230	1	100	8	+90	1377	3821496	expression site-associated gene 2
11	Tb11.01.6290	Tb11.01.6290	Tb11.01.6290	1	100	13	+174	489	3839485	hypothetical protein, conserved
11	Tb11.01.6310	Tb11.01.6310	Tb11.01.6310	1	100	11	+207	813	3849673	hypothetical protein, conserved
11	Tb11.01.6440	Tb11.01.6440	Tb11.01.6440	1	100	3	+207	735	3877601	hypothetical protein, conserved
11	Tb11.01.6760	Tb11.01.6760	Tb11.01.6760	1	100	11	+66	1722	3960320	hypothetical protein, conserved
11	Tb11.01.6940	Tb11.01.6940	Tb11.01.6940	1	100	3	+24	801	3995831	hypothetical protein, conserved

11	Tb11.01.7540	Tb11.01.7540	Tb11.01.7540	1	100	4	+246	2865	4150293	hypothetical protein, conserved
11	Tb11.01.7710	Tb11.01.7710	Tb11.01.7710	1	100	31	+63	1809	4192006	hypothetical protein, conserved
11	Tb11.01.7930	Tb11.01.7930	Tb11.01.7930	1	100	23	+126	1695	4243849	hypothetical protein, conserved
11	Tb11.02.0351	Tb11.02.0351	Tb11.02.0351	2	100	51	+120	1071	789777	hypothetical protein, conserved
11	Tb11.02.0540	Tb11.02.0540	Tb11.02.0540	2	100	46	+162	231	856145	hypothetical protein
11	Tb11.02.0650	Tb11.02.0650	Tb11.02.0650	1	100	3	+123	1131	877607	hypothetical protein, conserved
11	Tb11.02.0940	Tb11.02.0940	Tb11.02.0940	1	100	43	+120	2070	1001530	hypothetical protein, conserved
11	Tb11.02.1180	Tb11.02.1180	Tb11.02.1180	1	100	36	+90	1650	1066728	hypothetical protein, conserved
11	Tb11.02.1390	Tb11.02.1390	Tb11.02.1390	1	100	6	+234	2175	1119315	dimethylguanosine tRNA methyltransferase
11	Tb11.02.1400	Tb11.02.1400	Tb11.02.1400	1	100	38	+246	828	1121828	glycosyltransferase family-like protein
11	Tb11.02.1610	Tb11.02.1610	Tb11.02.1610	1	100	2	+210	432	1185881	hypothetical protein, conserved
11	Tb11.02.1710	Tb11.02.1710	Tb11.02.1710	1	100	17	+66	1350	1212458	hypothetical protein, conserved
11	Tb11.02.1890	Tb11.02.1890	Tb11.02.1890	1	100	5	+105	2673	1237079	hypothetical protein, conserved
11	Tb11.02.1910	Tb11.02.1910	Tb11.02.1910	2	100	23	+180	774	1241438	hypothetical protein, conserved
11	Tb11.02.1930	Tb11.02.1930	Tb11.02.1930	1	100	8	+135	6456	1249296	DEAD/DEAH box RNA helicase,
11	Tb11.02.2050	Tb11.02.2050	Tb11.02.2050	1	100	8	+72	3270	1269391	threonine protein kinase
11	Tb11.02.2760	Tb11.02.2760	Tb11.02.2760	3	100	6	+207	2694	1433590	hypothetical protein, conserved
11	Tb11.02.3900	Tb11.02.3900	Tb11.02.3900	1	100	16	+153	1155	1692776	hypothetical protein, conserved
11	Tb11.02.4205	Tb11.02.4205	Tb11.02.4205	1	100	30	+189	1203	1760476	hypothetical protein, conserved
11	Tb11.02.4460	Tb11.02.4460	Tb11.02.4460	1	100	26	+90	1095	1843591	hypothetical protein, conserved
11	Tb11.02.4540	Tb11.02.4540	Tb11.02.4540	1	100	27	+51	1146	1869516	hypothetical protein, conserved
11	Tb11.02.5210	Tb11.02.5210	Tb11.02.5210	1	100	26	+84	801	2040018	RNA binding protein,
11	Tb11.02.5340	Tb11.02.5340	Tb11.02.5340	1	100	27	+60	579	2068694	hypothetical protein, conserved
11	Tb11.02.5440	Tb11.02.5440	Tb11.02.5440	1	100	12	+102	1614	2093187	hypothetical protein, conserved
11	Tb11.03.0060	Tb11.03.0060	Tb11.03.0060	1	100	19	+147	1563	267369	hypothetical protein, conserved
11	Tb11.03.0470	Tb11.03.0470	Tb11.03.0470	1	100	14	+75	996	149244	hypothetical protein, conserved
11	Tb11.03.0500	Tb11.03.0500	Tb11.03.0500	3	100	20	+33	549	145098	lipid-like atypical dual specifi city phosphatase,

11	Tb11.03.0940	Tb11.03.0940	Tb11.03.0940	1	100	54	+18	1944	45067	elongation factor,
11	Tb11.12.0005	Tb11.12.0005	Tb11.12.0005	1	100	3	+135	1398	2346997	hypothetical protein, conserved
11	Tb11.12.0006	Tb11.12.0006	Tb11.12.0006	1	100	2	+273	1581	2348916	major surface protease gp63, ,surface protease homologue
11	Tb11.12.0007	Tb11.12.0007	Tb11.12.0007	1	100	15	+144	1959	2351823	S. cerevisiae CDC 48 homolog
11	Tb11.12.0015	Tb11.12.0015	Tb11.12.0015	1	100	33	+132	1338	2365787	hypothetical protein, conserved
11	Tb11.18.0006	Tb11.18.0006	Tb11.18.0006	1	100	59	+93	2214	610570	carnitine O-acetyltransferase,
11	Tb11.46.0011	Tb11.46.0011	Tb11.46.0011	1	100	33	+69	1179	551428	leucine-rich repeat protein (LRRP),
11	Tb11.47.0005	Tb11.47.0005	Tb11.47.0005	1	100	51	+147	507	417979	hypothetical protein, conserved
11	Tb11.47.0019	Tb11.47.0019	Tb11.47.0019	2	100	26	+381	1083	372953	hypothetical protein, conserved
11	Tb11.55.0005	Tb11.55.0005	Tb11.55.0005	1	100	34	+261	984	510408	hypothetical protein, conserved
11	Tb11.55.0014	Tb11.55.0014	Tb11.55.0014	1	100	3	+132	1989	490984	vesicular transport protein (CDC48),
11	Tb11.55.0020	Tb11.55.0020	Tb11.55.0020	1	100	39	+90	1065	477940	tatD related deoxyribonuclease,
7	Tb927.7.230	Tb927.7.230	Tb927.7.230	2	99	1110	+18	294	43411	40S ribosomal protein S33,
7	Tb927.7.240	Tb927.7.240	Tb927.7.240	2	99	1131	+18	294	43937	40S ribosomal protein S33,
8	Tb927.8.2000	Tb927.8.2000	Tb927.8.2000	2	99	100	+117	789	635332	cyclophilin type peptidyl-prolyl isomerase
8	Tb927.8.4700	Tb927.8.4700	Tb927.8.4700	2	99	172	+222	1377	1381499	amino acid transporter,
9	Tb09.160.4450	Tb09.160.4450	Tb09.160.4450	4	99	1021	+156	645	957413	40S ribosomal protein S3,
9	Tb09.244.2720	Tb09.244.2720	Tb09.244.2720	3	99	736	+51	615	2308692	ribosomal protein L15,
10	Tb10.6k15.1480	Tb10.6k15.1480	Tb10.6k15.1480	2	99	192	+180	381	2184558	kinetoplast DNA-associated protein
10	Tb10.6k15.3250	Tb10.6k15.3250	Tb10.6k15.3250	2	99	235	+261	1260	1850142	succinyl-CoA ligase [GDP-forming] beta
10	Tb10.70.3160	Tb10.70.3160	Tb10.70.3160	7	99	1986	+27	318	1082221	60S ribosomal protein L30
11	Tb11.01.2680	Tb11.01.2680	Tb11.01.2680	2	99	1294	+99	735	2853626	40S ribosomal protein SA,
11	Tb11.01.5280	Tb11.01.5280	Tb11.01.5280	2	99	173	+81	558	3591959	hypothetical protein, conserved
11	Tb11.02.2730	Tb11.02.2730	Tb11.02.2730	2	99	73	+360	243	1427806	hypothetical protein, conserved
2	Tb927.2.1560	Tb927.2.1560	Tb927.2.1560	3	98	294	+234	591	305260	cyclophilin type peptidyl-prolyl isomerase
3	Tb927.3.3310	Tb927.3.3310	Tb927.3.3310	4	98	990	+33	657	851554	60S ribosomal protein L13,
3	Tb927.3.5130	Tb927.3.5130	Tb927.3.5130	3	98	40	+222	1137	1438364	hypothetical protein, conserved

4	Tb927.4.1420	Tb927.4.1420	Tb927.4.1420	2	98	53	+300	405	368344	hypothetical protein, conserved
6	Tb927.6.2000	Tb927.6.2000	Tb927.6.2000	3	98	43	+63	1371	653078	spliceosome-associated protein,
6	Tb927.6.4620	Tb927.6.4620	Tb927.6.4620	3	98	54	+144	468	1292504	hypothetical protein, conserved
6	Tb927.6.660	Tb927.6.660	Tb927.6.660	2	98	41	+258	1812	281044	hypothetical protein, conserved
7	Tb927.7.1740	Tb927.7.1740	Tb927.7.1740	5	98	792	+45	729	430058	60S ribosomal protein L7,
7	Tb927.7.1750	Tb927.7.1750	Tb927.7.1750	5	98	703	+45	729	431071	60S ribosomal protein L7,
7	Tb927.7.7380	Tb927.7.7380	Tb927.7.7380	2	98	47	+30	246	2123051	U6 snRNA-associated Sm-like protein
7	Tb927.7.930	Tb927.7.930	Tb927.7.930	3	98	58	+159	1233	235926	hypothetical protein, conserved, zinc
10	Tb10.70.1640	Tb10.70.1640	Tb10.70.1640	2	98	42	+120	1725	1359094	hypothetical protein, conserved
11	Tb11.01.1990	Tb11.01.1990	Tb11.01.1990	2	98	58	+90	1086	2681686	hypothetical protein, conserved
11	Tb11.02.0530	Tb11.02.0530	Tb11.02.0530	2	98	134	+255	1071	854842	phosphoribosylpyrophosphate synthetase,
11	Tb11.02.3770	Tb11.02.3770	Tb11.02.3770	4	98	338	+105	324	1668423	hypothetical protein, conserved
2	Tb927.2.5160	Tb927.2.5160	Tb927.2.5160	4	97	265	+12	1203	910854	chaperone protein DnaJ,
4	Tb927.4.2310	Tb927.4.2310	Tb927.4.2310	3	97	69	+237	2004	601242	asparaginyl-tRNA synthetase,
4	Tb927.4.3570	Tb927.4.3570	Tb927.4.3570	3	97	770	+105	681	915785	translation elongation factor 1-beta,
4	Tb927.4.3590	Tb927.4.3590	Tb927.4.3590	3	97	839	+105	681	917842	translation elongation factor 1-beta
7	Tb927.7.2370	Tb927.7.2370	Tb927.7.2370	2	97	682	+60	459	611752	40S ribosomal protein S15,
8	Tb927.8.1610	Tb927.8.1610	Tb927.8.1610	2	97	263	+81	1665	537142	major surface protease gp63
8	Tb927.8.1630	Tb927.8.1630	Tb927.8.1630	2	97	293	+99	1665	541528	major surface protease gp63
8	Tb927.8.1640	Tb927.8.1640	Tb927.8.1640	2	97	259	+99	1665	543721	major surface protease gp63
8	Tb927.8.7700	Tb927.8.7700	Tb927.8.7700	3	97	132	+177	1464	2216565	amino acid transporter,
9	Tb09.160.3820	Tb09.160.3820	Tb09.160.3820	2	97	59	+222	1284	878846	nucleolar RNA binding protein,
9	Tb09.211.1670	Tb09.211.1670	Tb09.211.1670	2	97	33	+99	2022	1556615	hypothetical protein, conserved
10	Tb10.70.1690	Tb10.70.1690	Tb10.70.1690	5	97	1095	+141	519	1347452	40S ribosomal protein S10,
11	Tb11.01.4750	Tb11.01.4750	Tb11.01.4750	4	97	690	+384	1215	3479979	elongation factor 1 gamma,
11	Tb11.01.5300	Tb11.01.5300	Tb11.01.5300	2	97	118	+66	1272	3596349	ornithine decarboxylase
11	Tb11.03.0475	Tb11.03.0475	Tb11.03.0475	3	97	254	+42	279	147240	hypothetical protein, conserved
4	Tb927.4.3550	Tb927.4.3550	Tb927.4.3550	6	96	711	+360	669	912419	60S ribosomal protein L13a,

6	Tb927.6.1070	Tb927.6.1070	Tb927.6.1070	2	96	23	+204	2592	427443	hypothetical protein, conserved
7	Tb927.7.5180	Tb927.7.5180	Tb927.7.5180	4	96	318	+168	495	1364071	60S ribosomal protein L23a,
8	Tb927.8.1620	Tb927.8.1620	Tb927.8.1620	2	96	250	+99	1665	539335	major surface protease gp63
8	Tb927.8.1930	Tb927.8.1930	Tb927.8.1930	2	96	23	+45	786	621931	hypothetical protein, conserved
9	Tb09.160.0400	Tb09.160.0400	Tb09.160.0400	2	96	27	+135	1842	339615	hypothetical protein, conserved
9	Tb09.160.3060	Tb09.160.3060	Tb09.160.3060	2	96	25	+195	1971	779001	hypothetical protein
9	Tb09.211.0530	Tb09.211.0530	Tb09.211.0530	3	96	46	+288	510	1353390	hypothetical protein, conserved
10	Tb10.406.0290	Tb10.406.0290	Tb10.406.0290	2	96	92	+111	690	2543623	protein tyrosine phosphatase,
11	Tb11.02.0580	Tb11.02.0580	Tb11.02.0580	2	96	70	+132	615	862920	vesicular protein trafficking mediator
11	Tb11.03.0700	Tb11.03.0700	Tb11.03.0700	2	96	109	+207	882	98820	hypothetical protein, conserved
1	Tb927.1.3950	Tb927.1.3950	Tb927.1.3950	2	95	43	+195	1515	840877	alanine aminotransferase,
2	Tb927.2.4460	Tb927.2.4460	Tb927.2.4460	3	95	18	+180	1791	790729	hypothetical protein, conserved
3	Tb927.3.3370	Tb927.3.3370	Tb927.3.3370	3	95	41	+177	1173	863609	hypothetical protein, conserved
4	Tb927.4.4820	Tb927.4.4820	Tb927.4.4820	3	95	61	+51	1422	1320642	amino acid transporter 10,
4	Tb927.4.4840	Tb927.4.4840	Tb927.4.4840	3	95	62	+51	1422	1324631	amino acid transporter 7,
4	Tb927.4.4890	Tb927.4.4890	Tb927.4.4890	3	95	20	+72	1584	1339222	hypothetical protein, conserved
5	Tb927.5.2640	Tb927.5.2640	Tb927.5.2640	4	95	21	+231	834	845231	hypothetical protein, conserved
6	Tb927.6.4630	Tb927.6.4630	Tb927.6.4630	5	95	35	+234	1101	1293226	kinetoplastid-specific phospho-protein phosphatase
8	Tb927.8.6520	Tb927.8.6520	Tb927.8.6520	2	95	19	+60	1245	1884528	hypothetical protein, conserved
10	Tb10.389.0150	Tb10.389.0150	Tb10.389.0150	3	95	17	+135	7821	3174532	hypothetical protein, conserved
11	Tb11.01.2700	Tb11.01.2700	Tb11.01.2700	2	95	19	+249	852	2859744	hypothetical protein, conserved
11	Tb11.01.8620	Tb11.01.8620	Tb11.01.8620	3	95	37	+330	1449	4434141	hypothetical protein, conserved
1	Tb927.1.1990	Tb927.1.1990	Tb927.1.1990	2	94	16	+81	597	517835	hypothetical protein, conserved
1	Tb927.1.3280	Tb927.1.3280	Tb927.1.3280	2	94	18	+102	756	712980	septum formation protein MAF homologue,
3	Tb927.3.2010	Tb927.3.2010	Tb927.3.2010	2	94	16	+84	1665	524249	hypothetical protein, conserved
4	Tb927.4.930	Tb927.4.930	Tb927.4.930	3	94	35	+84	570	260957	50S ribosomal protein L14,
5	Tb927.5.2260	Tb927.5.2260	Tb927.5.2260	3	94	642	+54	351	704880	hypothetical protein, conserved

8	Tb927.8.7680	Tb927.8.7680	Tb927.8.7680	3	94	137	+177	1464	2212523	amino acid transporter,
10	Tb10.70.6320	Tb10.70.6320	Tb10.70.6320	2	94	17	+57	1074	435699	hypothetical protein, conserved
11	Tb11.01.3530	Tb11.01.3530	Tb11.01.3530	2	94	18	+165	1089	3065417	hypothetical protein, conserved
11	Tb11.01.3700	Tb11.01.3700	Tb11.01.3700	4	94	38	+54	2094	3111754	leucine-rich repeat protein (LRRP),
11	Tb11.03.0840	Tb11.03.0840	Tb11.03.0840	2	94	31	+66	555	68409	hypothetical protein, conserved
4	Tb927.4.3820	Tb927.4.3820	Tb927.4.3820	3	93	14	+69	1281	1000983	hypothetical protein, conserved
5	Tb927.5.2160	Tb927.5.2160	Tb927.5.2160	4	93	605	+54	351	690857	hypothetical protein, conserved
5	Tb927.5.2170	Tb927.5.2170	Tb927.5.2170	5	93	630	+54	351	692334	hypothetical protein, conserved
5	Tb927.5.2230	Tb927.5.2230	Tb927.5.2230	3	93	616	+54	351	700698	hypothetical protein, conserved
6	Tb927.6.4780	Tb927.6.4780	Tb927.6.4780	2	93	45	+141	2100	1334045	DNA ligase I,
9	Tb09.160.1780	Tb09.160.1780	Tb09.160.1780	2	93	27	+30	3042	630408	protein kinase,
10	Tb10.6k15.2770	Tb10.6k15.2770	Tb10.6k15.2770	3	93	14	+177	1164	1935472	COP9 signalosome subunit
10	Tb10.70.1550	Tb10.70.1550	Tb10.70.1550	2	93	29	+24	1413	1374453	hypothetical protein, conserved
11	Tb11.01.0710	Tb11.01.0710	Tb11.01.0710	3	93	14	+21	2454	2409775	elongation factor G2-like protein
3	Tb927.3.3530	Tb927.3.3530	Tb927.3.3530	2	92	12	+210	1143	993469	hypothetical protein, conserved
5	Tb927.5.2200	Tb927.5.2200	Tb927.5.2200	3	92	559	+54	351	696516	hypothetical protein, conserved
8	Tb927.8.1370	Tb927.8.1370	Tb927.8.1370	2	92	13	+45	462	445512	hypothetical protein, conserved
9	Tb09.160.1030	Tb09.160.1030	Tb09.160.1030	2	92	75	+81	300	499249	hypothetical protein, conserved
9	Tb09.211.2080	Tb09.211.2080	Tb09.211.2080	2	92	13	+36	1794	1639853	hypothetical protein, conserved
10	Tb10.389.0630	Tb10.389.0630	Tb10.389.0630	4	92	179	+279	2148	3062948	bifunctional aminoacyl-tRNA synthetase
10	Tb10.70.0930	Tb10.70.0930	Tb10.70.0930	2	92	13	+48	1701	1497177	hypothetical protein, conserved zinc finger
10	Tb10.70.2400	Tb10.70.2400	Tb10.70.2400	2	92	12	+171	1746	1205830	hypothetical protein, conserved
11	Tb11.01.3340	Tb11.01.3340	Tb11.01.3340	3	92	26	+141	2103	3021928	phosphoglycerate mutase-like protein
10	Tb10.6k15.3700	Tb10.6k15.3700	Tb10.6k15.3700	2	91	11	+33	1380	1751040	hypothetical protein, conserved
10	Tb10.70.2620	Tb10.70.2620	Tb10.70.2620	2	91	11	+27	1374	1173547	hypothetical protein, SET domain
11	Tb11.01.0170	Tb11.01.0170	Tb11.01.0170	2	91	43	+18	1890	2290595	NADPH--cytochrome P450 reductase
11	Tb11.v4.0008	Tb11.v4.0008	Tb11.v4.0008	2	91	11	+207	1866	4003354	hypothetical protein
2	Tb927.2.4150	Tb927.2.4150	Tb927.2.4150	5	90	29	+147	960	730121	hypothetical protein, conserved

2	Tb927.2.4390	Tb927.2.4390	Tb927.2.4390	2	90	10	+33	2259	777424	endo/exonuclease Mre11
3	Tb927.3.1710	Tb927.3.1710	Tb927.3.1710	2	90	20	+102	1137	449751	hypothetical protein, conserved
6	Tb927.6.2980	Tb927.6.2980	Tb927.6.2980	2	90	10	+39	3078	887744	protein kinase,
8	Tb927.8.7540	Tb927.8.7540	Tb927.8.7540	2	90	10	+138	3330	2169925	hypothetical protein
6	Tb927.6.4590	Tb927.6.4590	Tb927.6.4590	3	89	44	+78	1746	1285659	glutamyl-tRNA synthetase,
11	Tb11.03.0370	Tb11.03.0370	Tb11.03.0370	2	89	18	+300	738	192839	hypothetical protein, conserved
11	Tb11.03.0540	Tb11.03.0540	Tb11.03.0540	2	89	18	+78	1998	137442	ABC transporter,
1	Tb927.1.1720	Tb927.1.1720	Tb927.1.1720	5	88	33	+66	1227	436940	hypothetical protein, conserved
2	Tb927.2.5440	Tb927.2.5440	Tb927.2.5440	4	87	24	+246	2643	971330	hypothetical protein, conserved
5	Tb927.5.940	Tb927.5.940	Tb927.5.940	4	86	15	+198	2436	314335	NADH-dependent fumarate reductase
7	Tb927.7.3890	Tb927.7.3890	Tb927.7.3890	3	86	7	+78	843	1019389	hypothetical protein, conserved
7	Tb927.7.4590	Tb927.7.4590	Tb927.7.4590	2	86	7	+111	1530	1220705	chaperone protein DNAJ,
8	Tb927.8.3020	Tb927.8.3020	Tb927.8.3020	2	86	14	+225	1812	907278	hypothetical protein, conserved
8	Tb927.8.3330	Tb927.8.3330	Tb927.8.3330	2	86	7	+99	924	999073	mitochondrial carrier protein,
10	Tb10.6k15.3720	Tb10.6k15.3720	Tb10.6k15.3720	3	86	35	+24	1284	1747266	acid phosphatase,
7	Tb927.7.4670	Tb927.7.4670	Tb927.7.4670	2	85	71	+69	906	1238373	exosome complex exonuclease
10	Tb10.61.0480	Tb10.61.0480	Tb10.61.0480	3	85	14	+30	795	3834968	hypothetical protein, conserved
6	Tb927.6.4160	Tb927.6.4160	Tb927.6.4160	4	84	62	+225	1089	1198438	hypothetical protein, conserved
8	Tb927.8.3900	Tb927.8.3900	Tb927.8.3900	3	84	58	+186	927	1163808	hypothetical protein, conserved
11	Tb11.01.6700	Tb11.01.6700	Tb11.01.6700	5	84	88	+18	381	3945507	hypothetical protein, conserved
6	Tb927.6.2490	Tb927.6.2490	Tb927.6.2490	3	83	6	+153	723	759002	hypothetical protein, conserved
10	Tb10.70.5630	Tb10.70.5630	Tb10.70.5630	2	83	6	+204	2526	551388	hypothetical protein, conserved
8	Tb927.8.2390	Tb927.8.2390	Tb927.8.2390	2	82	11	+198	4398	717543	hypothetical protein, conserved
9	Tb09.211.2460	Tb09.211.2460	Tb09.211.2460	2	82	11	+222	2424	1717653	hypothetical protein, conserved
1	Tb927.1.290	Tb927.1.290	Tb927.1.290	2	80	10	+165	4182	113137	leucine-rich repeat protein (LRRP),
2	Tb927.2.2410	Tb927.2.2410	Tb927.2.2410	2	80	5	+192	1944	475894	hypothetical protein, conserved
3	Tb927.3.1070	Tb927.3.1070	Tb927.3.1070	2	80	15	+45	1335	263824	peptide chain release factor 1,
3	Tb927.3.1460	Tb927.3.1460	Tb927.3.1460	6	80	45	+126	375	372975	hypothetical protein

9	Tb09.160.4290	Tb09.160.4290	Tb09.160.4290	3	78	18	+276	1377	940660	hypothetical protein, conserved
9	Tb09.211.1030	Tb09.211.1030	Tb09.211.1030	3	78	111	+54	1014	1454758	phosphatidylcholine:ceramide cholinephosphotransferase 2
10	Tb10.05.0090	Tb10.05.0090	Tb10.05.0090	4	78	50	+210	1371	3263167	hypothetical protein, conserved
5	Tb927.5.1990	Tb927.5.1990	Tb927.5.1990	2	75	12	+210	795	616011	hypothetical protein, conserved
8	Tb927.8.7600	Tb927.8.7600	Tb927.8.7600	4	75	165	+216	1404	2194319	amino acid transporter,
11	Tb11.01.0210	Tb11.01.0210	Tb11.01.0210	2	75	4	+54	444	2283676	hypothetical protein
11	Tb11.01.3760	Tb11.01.3760	Tb11.01.3760	3	73	41	+129	423	3128917	hypothetical protein, conserved
8	Tb927.8.2690	Tb927.8.2690	Tb927.8.2690	3	71	14	+90	1110	800889	hypothetical protein, conserved
10	Tb10.6k15.3370	Tb10.6k15.3370	Tb10.6k15.3370	2	71	7	+186	2484	1827592	terminal uridylyltransferase 3,
4	Tb927.4.4010	Tb927.4.4010	Tb927.4.4010	4	70	46	+84	1371	1067782	amino acid transporter,
11	Tb11.52.0004	Tb11.52.0004	Tb11.52.0004	4	70	23	+201	708	3384321	hypothetical protein, conserved
11	Tb11.02.2250	Tb11.02.2250	Tb11.02.2250	4	69	16	+3789	780	1319085	hypothetical protein, conserved
4	Tb927.4.4000	Tb927.4.4000	Tb927.4.4000	4	68	47	+84	1371	1064468	amino acid transporter,
9	Tb09.211.1000	Tb09.211.1000	Tb09.211.1000	6	68	110	+54	1044	1447601	phosphatidylcholine:ceramide cholinephosphotransferase 2
10	Tb10.6k15.3100	Tb10.6k15.3100	Tb10.6k15.3100	3	68	22	+48	1623	1880531	hypothetical protein, conserved
7	Tb927.7.6410	Tb927.7.6410	Tb927.7.6410	2	67	3	+2451	1920	1739255	hypothetical protein, conserved
10	Tb10.406.0140	Tb10.406.0140	Tb10.406.0140	2	67	3	+231	1704	2568891	hypothetical protein, conserved
11	Tb11.02.0280	Tb11.02.0280	Tb11.02.0280	2	67	3	+420	2097	767487	hypothetical protein, conserved
11	Tb11.02.2550	Tb11.02.2550	Tb11.02.2550	2	67	3	+60	1572	1383708	AAA ATPase,
11	Tb11.02.1020	Tb11.02.1020	Tb11.02.1020	4	65	31	+135	543	1024046	hypothetical protein, conserved
4	Tb927.4.3990	Tb927.4.3990	Tb927.4.3990	4	63	40	+84	1371	1061159	amino acid transporter,
6	Tb927.6.2180	Tb927.6.2180	Tb927.6.2180	4	63	76	+222	519	695798	hypothetical protein, conserved
9	Tb09.244.2830	Tb09.244.2830	Tb09.244.2830	5	63	24	+156	885	2292299	hypothetical protein, conserved
10	Tb10.70.6030	Tb10.70.6030	Tb10.70.6030	2	63	19	+108	795	476970	hypothetical protein, conserved
8	Tb927.8.1700	Tb927.8.1700	Tb927.8.1700	3	62	13	+246	2010	566789	hypothetical protein, conserved
11	Tb11.03.0300	Tb11.03.0300	Tb11.03.0300	6	62	58	+126	2280	214244	hypothetical protein, conserved
8	Tb927.8.3880	Tb927.8.3880	Tb927.8.3880	5	61	83	+186	726	1158421	hypothetical protein, conserved

1	Tb927.1.480	Tb927.1.480	Tb927.1.480	2	60	10	+165	4185	181845	leucine-rich repeat protein (LRRP)
4	Tb927.4.910	Tb927.4.910	Tb927.4.910	3	60	10	+141	777	258442	hypothetical protein, conserved
11	Tb11.01.8600	Tb11.01.8600	Tb11.01.8600	2	60	5	+30	2733	4428642	hypothetical protein, conserved
8	Tb927.8.3890	Tb927.8.3890	Tb927.8.3890	5	59	78	+186	792	1161047	hypothetical protein, conserved
11	Tb11.01.1340	Tb11.01.1340	Tb11.01.1340	2	59	44	+36	396	2556159	hypothetical protein, conserved
7	Tb927.7.3390	Tb927.7.3390	Tb927.7.3390	4	58	26	+135	1047	884821	hypothetical protein, conserved
8	Tb927.8.5610	Tb927.8.5610	Tb927.8.5610	3	58	31	+81	1416	1665119	hypothetical protein, conserved
10	Tb10.70.5970	Tb10.70.5970	Tb10.70.5970	8	58	91	+90	456	486205	hypothetical protein, conserved
10	Tb10.6k15.1220	Tb10.6k15.1220	Tb10.6k15.1220	4	55	87	+198	3234	2236267	isoleucyl-tRNA synthetase,
5	Tb927.5.1150	Tb927.5.1150	Tb927.5.1150	5	54	62	+39	2169	393097	pre-mRNA RNA helicase
10	Tb10.61.2130	Tb10.61.2130	Tb10.61.2130	3	54	35	+126	1857	3495367	ATP-dependent DEAD/H RNA helicase
11	Tb11.01.4290	Tb11.01.4290	Tb11.01.4290	4	53	15	+72	1635	3266377	hypothetical protein, conserved
11	Tb11.02.4890	Tb11.02.4890	Tb11.02.4890	4	53	98	+33	342	1956475	pterin-4-alpha-carbinolamine dehydratase
4	Tb927.4.4810	Tb927.4.4810	Tb927.4.4810	2	52	61	+399	696	1315998	hypothetical protein, conserved
1	Tb927.1.370	Tb927.1.370	Tb927.1.370	2	50	4	+165	4182	144395	leucine-rich repeat protein (LRRP),
9	Tb09.160.0920	Tb09.160.0920	Tb09.160.0920	3	50	12	+3273	1005	473133	hypothetical protein, conserved
9	Tb09.211.3760	Tb09.211.3760	Tb09.211.3760	2	50	4	+735	861	1974713	poly(ADP-ribose) glycohydrolase,
11	Tb11.01.1380	Tb11.01.1380	Tb11.01.1380	2	50	30	+42	2253	2564020	hypothetical protein, conserved
11	Tb11.02.0820	Tb11.02.0820	Tb11.02.0820	4	50	34	+156	1068	951987	ras-family member, GTP-binding protein
11	Tb11.22.0011	Tb11.22.0011	Tb11.22.0011	4	50	20	+78	3066	639451	hypothetical protein, conserved
11	Tb11.02.5470	Tb11.02.5470	Tb11.02.5470	4	48	100	+159	600	2098252	vacuolar type H+ ATPase subunit,
10	Tb10.70.5830	Tb10.70.5830	Tb10.70.5830	2	47	15	+147	1530	523989	actin-like protein,
10	Tb10.6k15.2300	Tb10.6k15.2300	Tb10.6k15.2300	4	45	62	+132	390	2020822	hypothetical protein, conserved
7	Tb927.7.6620	Tb927.7.6620	Tb927.7.6620	4	44	120	+141	1482	1810332	hypothetical protein, conserved
8	Tb927.8.4470	Tb927.8.4470	Tb927.8.4470	4	44	9	+57	657	1325310	chaperone protein DNAJ,
7	Tb927.7.7510	Tb927.7.7510	Tb927.7.7510	3	43	54	+39	699	2176343	hypothetical protein
10	Tb10.70.7170	Tb10.70.7170	Tb10.70.7170	3	43	19	+123	1821	267401	hypothetical protein, conserved
3	Tb927.3.2510	Tb927.3.2510	Tb927.3.2510	2	40	5	+345	1026	633857	expression site-associated gene 2,

3	Tb927.3.920	Tb927.3.920	Tb927.3.920	3	40	5	+342	3468	213287	hypothetical protein, conserved
4	Tb927.4.2900	Tb927.4.2900	Tb927.4.2900	2	40	5	+159	2757	768136	hypothetical protein, conserved
10	Tb10.70.5230	Tb10.70.5230	Tb10.70.5230	4	39	18	+183	1971	637485	hypothetical protein, conserved
11	Tb11.01.4590	Tb11.01.4590	Tb11.01.4590	3	39	18	+90	960	3344853	hypothetical protein, conserved
11	Tb11.01.4840	Tb11.01.4840	Tb11.01.4840	3	38	8	+72	933	3496954	hypothetical protein, conserved
10	Tb10.26.0050	Tb10.26.0050	Tb10.26.0050	2	36	22	+207	798	2817908	hypothetical protein, conserved
1	Tb927.1.2980	Tb927.1.2980	Tb927.1.2980	5	34	29	+123	3417	658938	hypothetical protein, conserved
10	Tb10.61.0370	Tb10.61.0370	Tb10.61.0370	2	33	21	+135	384	3853890	hypothetical protein, conserved
7	Tb927.7.450	Tb927.7.450	Tb927.7.450	2	30	23	+1383	930	80862	hypothetical protein, conserved
1	Tb927.1.4220	Tb927.1.4220	Tb927.1.4220	5	29	118	+87	1200	871865	hypothetical protein, conserved
3	Tb927.3.4810	Tb927.3.4810	Tb927.3.4810	4	29	28	+42	810	1355759	hypothetical protein, conserved
10	Tb10.6k15.2630	Tb10.6k15.2630	Tb10.6k15.2630	5	29	137	+93	600	1958565	hypothetical protein, conserved
1	Tb927.1.3410	Tb927.1.3410	Tb927.1.3410	3	28	50	+183	1314	735770	hypothetical protein, conserved
3	Tb927.3.4870	Tb927.3.4870	Tb927.3.4870	3	25	40	+69	972	1370695	hypothetical protein, conserved
10	Tb10.100.0045	Tb10.100.0045	Tb10.100.0045	6	25	12	+78	525	67786	hypothetical protein, conserved
11	Tb11.01.0120	Tb11.01.0120	Tb11.01.0120	4	25	8	+27	837	2304863	haloacid dehalogenase-like hydrolase,
11	Tb11.01.3900	Tb11.01.3900	Tb11.01.3900	3	25	12	+138	621	3163689	N-acetylglucosaminyl PI deacetylase
4	Tb927.4.3980	Tb927.4.3980	Tb927.4.3980	2	23	43	+27	1461	1058059	chaperone protein DNAJ,
7	Tb927.7.7110	Tb927.7.7110	Tb927.7.7110	2	23	74	+294	1938	2035069	leucine-rich repeat protein 1 (LRRP1)
4	Tb927.4.730	Tb927.4.730	Tb927.4.730	3	22	9	+246	1233	204893	hypothetical protein, conserved
11	Tb11.02.0353	Tb11.02.0353	Tb11.02.0353	5	22	43	+858	1128	792927	hypothetical protein, conserved
3	Tb927.3.3160	Tb927.3.3160	Tb927.3.3160	2	20	15	+1103	1011	814216	poly(A) polymerase
7	Tb927.7.6470	Tb927.7.6470	Tb927.7.6470	4	20	20	+105	1119	1756990	O-sialoglycoprotein endopepti dase
7	Tb927.7.790	Tb927.7.790	Tb927.7.790	3	19	16	+96	1608	180728	hypothetical protein, conserved
3	Tb927.3.4700	Tb927.3.4700	Tb927.3.4700	4	18	11	+36	1152	1328737	hypothetical protein, conserved
4	Tb927.4.4370	Tb927.4.4370	Tb927.4.4370	5	18	76	+120	1653	1153147	hypothetical protein, conserved
8	Tb927.8.2140	Tb927.8.2140	Tb927.8.2140	2	18	17	+81	618	665195	hypothetical protein, conserved
8	Tb927.8.4760	Tb927.8.4760	Tb927.8.4760	4	18	65	+63	735	1397151	hypothetical protein, conserved

9	Tb09.211.2460	Tb09.211.2460	Tb09.211.2460	2	18	11	+360	2286	1717515	hypothetical protein, conserved
3	Tb927.3.2620	Tb927.3.2620	Tb927.3.2620	3	15	82	+204	5850	668683	hypothetical protein, conserved
2	Tb927.2.4420	Tb927.2.4420	Tb927.2.4420	2	14	14	+246	888	785863	hypothetical protein, conserved
11	Tb11.01.5260	Tb11.01.5260	Tb11.01.5260	6	13	63	+81	510	3587767	radial spoke protein RSP11,
11	Tb11.02.3060	Tb11.02.3060	Tb11.02.3060	2	13	30	+303	546	1509398	hypothetical protein, conserved
9	Tb09.244.2840	Tb09.244.2840	Tb09.244.2840	2	11	65	+147	468	2290563	hypothetical protein, conserved
11	Tb11.03.0370	Tb11.03.0370	Tb11.03.0370	2	11	18	+63	975	192602	hypothetical protein, conserved
3	Tb927.3.2170	Tb927.3.2170	Tb927.3.2170	3	10	20	+765	1908	571964	translation elongation factor EF-2,
10	Tb10.6k15.2130	Tb10.6k15.2130	Tb10.6k15.2130	2	10	20	+90	1440	2044045	ribonuclease,
3	Tb927.3.2030	Tb927.3.2030	Tb927.3.2030	4	9	33	+51	282	528128	acylphosphatase,
7	Tb927.7.6760	Tb927.7.6760	Tb927.7.6760	3	8	26	+306	3405	1871025	hypothetical protein, conserved
9	Tb09.211.3610	Tb09.211.3610	Tb09.211.3610	7	8	305	+138	3507	1935292	ubiquitin-activating enzyme E1,
3	Tb927.3.5430	Tb927.3.5430	Tb927.3.5430	2	7	28	+21	795	1519786	hypothetical protein, conserved
9	Tb09.160.1780	Tb09.160.1780	Tb09.160.1780	2	7	27	+105	2967	630333	protein kinase,
8	Tb927.8.4760	Tb927.8.4760	Tb927.8.4760	4	5	65	+162	636	1397052	hypothetical protein, conserved

CATEGORY 2

There are 178 genes where my analysis of splice-site data indicates that the ORF assigned in TrypDB is incorrect. For > 90% of the tagged splice sites, there is an upstream in-frame ATG that would be expected to encode the start of the protein.

For an additional 23 genes, where between 25 and 88% of the predicted SAS precede an upstream ATG, alternative splicing could produce different versions of the protein. I do not know why the gene-prediction algorithms did not predict the longest ORF from the first in-frame ATG. The table lists the newly predicted ATG (relative to the original prediction), ORF start and ORF length.

The table has been sorted first by the % of SAS hits that predict one specific upstream ATG, then by chromosome and GeneID. In cases where there are only a few sequence tag hits per gene (or per SAS), the data should be viewed with caution and checked if it is an important gene for you.

Chr	GeneID	TrypDB	TriTrypDB	SAS per Gene	SAS % for ATG	Hits per Gene	ATG Diff	ORF Length True	ORF Start True	Abbreviated Description from TrypDB
2	Tb927.2.2550	Tb927.2.2550	Tb927.2.2550	1	100	7	-183	1728	503563	hypothetical protein, conserved
2	Tb927.2.2880	Tb927.2.2880	Tb927.2.2880	3	100	4	-9	1302	551545	hypothetical protein, conserved
2	Tb927.2.5200	Tb927.2.5200	Tb927.2.5200	3	100	12	-114	2142	915057	hypothetical protein, conserved
2	Tb927.2.940	Tb927.2.940	Tb927.2.940	3	100	6	-66	363	158095	hypothetical protein
3	Tb927.3.1620	Tb927.3.1620	Tb927.3.1620	1	100	10	-96	4896	428061	hypothetical protein, conserved
3	Tb927.3.1680	Tb927.3.1680	Tb927.3.1680	2	100	144	-60	1017	444390	hypothetical protein, conserved
3	Tb927.3.1820	Tb927.3.1820	Tb927.3.1820	1	100	5	-132	870	473676	hypothetical protein, conserved
3	Tb927.3.2610	Tb927.3.2610	Tb927.3.2610	1	100	7	-102	2448	662234	hypothetical protein, conserved
3	Tb927.3.2720	Tb927.3.2720	Tb927.3.2720	1	100	4	-363	1305	695060	hypothetical protein
3	Tb927.3.4070	Tb927.3.4070	Tb927.3.4070	1	100	257	-30	1788	1150052	hypothetical protein, conserved
3	Tb927.3.4080	Tb927.3.4080	Tb927.3.4080	1	100	252	-36	1752	1152313	hypothetical protein, conserved
3	Tb927.3.4090	Tb927.3.4090	Tb927.3.4090	1	100	222	-36	1788	1154649	hypothetical protein, conserved
3	Tb927.3.4100	Tb927.3.4100	Tb927.3.4100	1	100	252	-30	1758	1156968	hypothetical protein, conserved
3	Tb927.3.4280	Tb927.3.4280	Tb927.3.4280	1	100	18	-174	2373	1208542	mismatch repair protein MSH5,
3	Tb927.3.4360	Tb927.3.4360	Tb927.3.4360	3	100	793	-120	393	1223845	40S ribosomal protein S15a,
3	Tb927.3.4570	Tb927.3.4570	Tb927.3.4570	1	100	15	-27	2190	1279149	N-acetylglucosaminyltransferase,

3	Tb927.3.4950	Tb927.3.4950	Tb927.3.4950	3	100	27	-75	1500	1387003	hypothetical protein, conserved
3	Tb927.3.5030	Tb927.3.5030	Tb927.3.5030	1	100	12	-24	2238	1412962	KU70 protein
3	Tb927.3.5150	Tb927.3.5150	Tb927.3.5150	1	100	43	-114	990	1445889	exonuclease,
3	Tb927.3.5320	Tb927.3.5320	Tb927.3.5320	5	100	6	-27	1212	1496089	hypothetical protein, conserved
3	Tb927.3.990	Tb927.3.990	Tb927.3.990	1	100	32	-132	1572	243247	hypothetical protein, conserved
4	Tb927.4.1230	Tb927.4.1230	Tb927.4.1230	1	100	30	-207	996	330303	hypothetical protein
4	Tb927.4.1310	Tb927.4.1310	Tb927.4.1310	2	100	25	-165	1416	344695	hypothetical protein, conserved, ZFP family
4	Tb927.4.1350	Tb927.4.1350	Tb927.4.1350	2	100	14	-441	1434	356010	glyoxalase
4	Tb927.4.1980	Tb927.4.1980	Tb927.4.1980	1	100	3	-51	2547	502463	hypothetical protein, conserved
4	Tb927.4.2710	Tb927.4.2710	Tb927.4.2710	1	100	38	-15	492	715143	ubiquitin-conjugating enzyme E2
4	Tb927.4.3530	Tb927.4.3530	Tb927.4.3530	2	100	19	-132	2265	909546	hypothetical protein, conserved
4	Tb927.4.3710	Tb927.4.3710	Tb927.4.3710	1	100	22	-156	510	937775	hypothetical protein, conserved
4	Tb927.4.3840	Tb927.4.3840	Tb927.4.3840	4	100	57	-162	1740	1003799	nucleolar protein,
4	Tb927.4.4300	Tb927.4.4300	Tb927.4.4300	1	100	3	-159	1851	1139536	hypothetical protein, conserved
4	Tb927.4.4740	Tb927.4.4740	Tb927.4.4740	1	100	35	-24	1203	1304917	hypothetical protein, conserved
4	Tb927.4.4920	Tb927.4.4920	Tb927.4.4920	1	100	76	-90	777	1347976	hypothetical protein, conserved
4	Tb927.4.5170	Tb927.4.5170	Tb927.4.5170	1	100	8	-138	642	1413846	ribosomal protein L7/L12,
4	Tb927.4.870	Tb927.4.870	Tb927.4.870	1	100	3	-453	13800	252137	dynein heavy chain,
4	Tb927.4.970	Tb927.4.970	Tb927.4.970	2	100	23	-105	669	267912	hypothetical protein, conserved
5	Tb927.5.1380	Tb927.5.1380	Tb927.5.1380	1	100	2	-282	858	444039	hypothetical protein
5	Tb927.5.1980	Tb927.5.1980	Tb927.5.1980	2	100	16	-234	1950	611407	hypothetical protein, conserved
5	Tb927.5.2360	Tb927.5.2360	Tb927.5.2360	2	100	4	-156	2262	749008	hypothetical protein, conserved
5	Tb927.5.3170	Tb927.5.3170	Tb927.5.3170	1	100	2	-888	1962	992910	ribose-phosphate pyrophosphokinase
5	Tb927.5.3300	Tb927.5.3300	Tb927.5.3300	2	100	80	-156	648	1038903	hypothetical protein, conserved
5	Tb927.5.3780	Tb927.5.3780	Tb927.5.3780	1	100	3	-90	1254	1175492	hypothetical protein, conserved
5	Tb927.5.650	Tb927.5.650	Tb927.5.650	4	100	9	-96	3846	223645	receptor-type adenylate cyclase GRESAG 4
5	Tb927.5.820	Tb927.5.820	Tb927.5.820	3	100	7	-408	1035	282974	hypothetical protein, conserved
6	Tb927.6.1590	Tb927.6.1590	Tb927.6.1590	1	100	24	-141	1545	548029	hypothetical protein, conserved

6	Tb927.6.2550	Tb927.6.2550	Tb927.6.2550	2	100	27	-93	2268	771666	RNA-binding protein,
6	Tb927.6.2660	Tb927.6.2660	Tb927.6.2660	2	100	11	-177	3327	800724	hypothetical protein, conserved
6	Tb927.6.2690	Tb927.6.2690	Tb927.6.2690	1	100	5	-123	2043	807701	ubiquitin carboxyl-terminal hydrolase
6	Tb927.6.2920	Tb927.6.2920	Tb927.6.2920	1	100	6	-294	2253	874885	hypothetical protein, conserved
6	Tb927.6.3060	Tb927.6.3060	Tb927.6.3060	2	100	33	-150	711	907031	hypothetical protein, conserved
6	Tb927.6.3130	Tb927.6.3130	Tb927.6.3130	1	100	10	-120	1344	923896	queuine tRNA-ribosyltransferase
6	Tb927.6.3350	Tb927.6.3350	Tb927.6.3350	1	100	4	-183	1596	982414	hypothetical protein, conserved
6	Tb927.6.4360	Tb927.6.4360	Tb927.6.4360	3	100	29	-21	405	1229188	hypothetical protein, conserved
6	Tb927.6.4550	Tb927.6.4550	Tb927.6.4550	2	100	5	-132	2406	1275264	hypothetical protein, conserved
6	Tb927.6.4950	Tb927.6.4950	Tb927.6.4950	3	100	40	-60	504	1365428	mago nashi-like protein,
6	Tb927.6.570	Tb927.6.570	Tb927.6.570	1	100	81	-51	366	239804	hypothetical protein, conserved
7	Tb927.7.1030	Tb927.7.1030	Tb927.7.1030	2	100	27	-99	2613	265231	heat shock 70 kDa protein,
7	Tb927.7.1090	Tb927.7.1090	Tb927.7.1090	4	100	12	-213	4839	277535	hypothetical protein, conserved
7	Tb927.7.1610	Tb927.7.1610	Tb927.7.1610	1	100	18	-15	1962	404547	6-phosphofructo-2-kinase
7	Tb927.7.2610	Tb927.7.2610	Tb927.7.2610	1	100	2	-291	1179	670966	hypothetical protein, conserved
7	Tb927.7.2630	Tb927.7.2630	Tb927.7.2630	1	100	14	-63	2766	675448	hypothetical protein, conserved
7	Tb927.7.3280	Tb927.7.3280	Tb927.7.3280	3	100	14	-198	2364	846746	translation initiation factor IF-2,
7	Tb927.7.3340	Tb927.7.3340	Tb927.7.3340	1	100	6	-249	1782	874847	hypothetical protein, conserved
7	Tb927.7.3970	Tb927.7.3970	Tb927.7.3970	1	100	13	-90	963	1039338	hypothetical protein, conserved
7	Tb927.7.4110	Tb927.7.4110	Tb927.7.4110	2	100	11	-21	2706	1088478	kinesin,
7	Tb927.7.4310	Tb927.7.4310	Tb927.7.4310	1	100	2	-108	1542	1144820	hypothetical protein, conserved
7	Tb927.7.5280	Tb927.7.5280	Tb927.7.5280	1	100	8	-177	6207	1395554	hypothetical protein, conserved
7	Tb927.7.6270	Tb927.7.6270	Tb927.7.6270	2	100	6	-42	1263	1698787	peptidase T,
7	Tb927.7.6690	Tb927.7.6690	Tb927.7.6690	1	100	10	-258	1467	1849254	hypothetical protein
7	Tb927.7.6750	Tb927.7.6750	Tb927.7.6750	2	100	6	-501	2187	1866795	hypothetical protein, conserved
7	Tb927.7.7050	Tb927.7.7050	Tb927.7.7050	1	100	39	-165	2898	2017419	hypothetical protein, conserved
7	Tb927.7.7440	Tb927.7.7440	Tb927.7.7440	4	100	84	-48	420	2137654	hypothetical protein, conserved
8	Tb927.8.1770	Tb927.8.1770	Tb927.8.1770	1	100	8	-18	1581	581271	hypothetical protein, conserved

8	Tb927.8.3600	Tb927.8.3600	Tb927.8.3600	4	100	32	-72	339	1081815	hypothetical protein, conserved
8	Tb927.8.4250	Tb927.8.4250	Tb927.8.4250	1	100	24	-168	1401	1264723	hypothetical protein, conserved
8	Tb927.8.5230	Tb927.8.5230	Tb927.8.5230	1	100	15	-117	849	1563400	cyclophilin-type peptidyl-prolyl isomerase
8	Tb927.8.850	Tb927.8.850	Tb927.8.850	2	100	18	-69	2337	249641	hypothetical protein, conserved
9	Tb09.160.0360	Tb09.160.0360	Tb09.160.0360	8	100	7	-399	1014	332395	hypothetical protein, conserved
9	Tb09.160.0460	Tb09.160.0460	Tb09.160.0460	3	100	23	-345	1494	350355	hypothetical protein, conserved
9	Tb09.160.0780	Tb09.160.0780	Tb09.160.0780	1	100	15	-3	1929	437909	syntaxin binding protein 1,
9	Tb09.160.0790	Tb09.160.0790	Tb09.160.0790	1	100	52	-36	2115	440569	hypothetical protein, conserved
9	Tb09.160.0810	Tb09.160.0810	Tb09.160.0810	1	100	26	-84	1392	444222	kynureninase,
9	Tb09.160.0840	Tb09.160.0840	Tb09.160.0840	3	100	33	-171	1251	450997	membrane protein YIP1,
9	Tb09.160.2110	Tb09.160.2110	Tb09.160.2110	1	100	26	-120	552	663196	hypothetical protein, conserved
9	Tb09.160.4770	Tb09.160.4770	Tb09.160.4770	1	100	3	-102	2181	1026509	protein kinase,
9	Tb09.160.5670	Tb09.160.5670	Tb09.160.5670	1	100	2	-33	4089	1218037	hypothetical protein, conserved
9	Tb09.211.0590	Tb09.211.0590	Tb09.211.0590	1	100	16	-255	1581	1369056	tRNA nucleotidyltransferase,
9	Tb09.211.1150	Tb09.211.1150	Tb09.211.1150	1	100	48	-105	639	1471782	hypothetical protein, conserved
9	Tb09.211.1190	Tb09.211.1190	Tb09.211.1190	1	100	5	-36	2322	1478511	minichromosome maintenance complex
9	Tb09.211.1330	Tb09.211.1330	Tb09.211.1330	2	100	45	-48	684	1503481	membrane-trafficking protein,
9	Tb09.211.2160	Tb09.211.2160	Tb09.211.2160	1	100	14	-315	918	1654887	hypothetical protein
9	Tb09.211.2960	Tb09.211.2960	Tb09.211.2960	1	100	40	-51	654	1795318	hypothetical protein, conserved
9	Tb09.244.2670	Tb09.244.2670	Tb09.244.2670	2	100	14	-24	1632	2315907	hypothetical protein, conserved
10	Tb10.26.0040	Tb10.26.0040	Tb10.26.0040	2	100	8	-186	1233	2819675	hypothetical protein, conserved
10	Tb10.26.0200	Tb10.26.0200	Tb10.26.0200	2	100	39	-84	621	2789465	guanylate kinase,
10	Tb10.26.0220	Tb10.26.0220	Tb10.26.0220	1	100	18	-132	2958	2783540	chloride channel protein,
10	Tb10.26.0300	Tb10.26.0300	Tb10.26.0300	1	100	37	-273	690	2763123	hypothetical protein, conserved
10	Tb10.26.0330	Tb10.26.0330	Tb10.26.0330	2	100	9	-102	5094	2751262	hypothetical protein, conserved, WD40
10	Tb10.26.0340	Tb10.26.0340	Tb10.26.0340	1	100	11	-42	2202	2748627	hypothetical protein, conserved
10	Tb10.26.0350	Tb10.26.0350	Tb10.26.0350	1	100	4	-33	1233	2747020	hypothetical protein, conserved
10	Tb10.26.0400	Tb10.26.0400	Tb10.26.0400	2	100	10	-210	1281	2738398	hypothetical protein, conserved

10	Tb10.26.0410	Tb10.26.0410	Tb10.26.0410	1	100	4	-366	1014	2736441	hypothetical protein, conserved
10	Tb10.26.0430	Tb10.26.0430	Tb10.26.0430	1	100	11	-90	1914	2733779	hypothetical protein, conserved
10	Tb10.26.0520	Tb10.26.0520	Tb10.26.0520	1	100	4	-81	1002	2716545	hypothetical protein, conserved
10	Tb10.26.0550	Tb10.26.0550	Tb10.26.0550	2	100	15	-279	1518	2708676	hypothetical protein, conserved
10	Tb10.26.0590	Tb10.26.0590	Tb10.26.0590	1	100	18	-336	2034	2702920	protein phosphatase 2a regulatory subunit
10	Tb10.26.0600	Tb10.26.0600	Tb10.26.0600	4	100	35	-60	462	2701792	hypothetical protein, conserved
10	Tb10.26.0690	Tb10.26.0690	Tb10.26.0690	1	100	4	-258	1302	2689136	hypothetical protein, conserved
10	Tb10.26.0770	Tb10.26.0770	Tb10.26.0770	3	100	31	-93	1287	2679398	hypothetical protein, conserved
10	Tb10.26.0860	Tb10.26.0860	Tb10.26.0860	1	100	27	-282	1014	2661341	HIRA-interacting protein 5,
10	Tb10.26.0930	Tb10.26.0930	Tb10.26.0930	2	100	28	-216	2538	2646857	hypothetical protein, conserved
10	Tb10.26.1000	Tb10.26.1000	Tb10.26.1000	2	100	11	-6	945	2628693	hypothetical protein, conserved
10	Tb10.389.0210	Tb10.389.0210	Tb10.389.0210	1	100	7	-72	1860	3151841	hypothetical protein, conserved
10	Tb10.389.0340	Tb10.389.0340	Tb10.389.0340	3	100	12	-171	1206	3122346	mitochondrial carrier protein
10	Tb10.389.1270	Tb10.389.1270	Tb10.389.1270	1	100	9	-105	2604	2944462	kinesin,
10	Tb10.389.1400	Tb10.389.1400	Tb10.389.1400	2	100	5	-33	2712	2924574	hypothetical protein, conserved
10	Tb10.389.1770	Tb10.389.1770	Tb10.389.1770	2	100	25	-30	1170	2850529	hypothetical protein, conserved
10	Tb10.389.1780	Tb10.389.1780	Tb10.389.1780	2	100	19	-36	432	2849589	hypothetical protein, conserved
10	Tb10.389.1870	Tb10.389.1870	Tb10.389.1870	1	100	3	-153	1455	2833645	hypothetical protein, conserved
10	Tb10.406.0190	Tb10.406.0190	Tb10.406.0190	2	100	41	-54	1128	2559577	hypothetical protein, conserved
10	Tb10.61.0260	Tb10.61.0260	Tb10.61.0260	1	100	2	-120	1062	3876192	hypothetical protein, conserved
10	Tb10.61.0340	Tb10.61.0340	Tb10.61.0340	1	100	29	-315	1416	3857674	cation transporter,
10	Tb10.61.0690	Tb10.61.0690	Tb10.61.0690	1	100	13	-27	1767	3765055	tRNA pseudouridine synthase A-like protein
10	Tb10.61.0860	Tb10.61.0860	Tb10.61.0860	1	100	39	-153	759	3735051	hypothetical protein, conserved
10	Tb10.61.2510	Tb10.61.2510	Tb10.61.2510	1	100	10	-342	912	3400076	mitochondrial carrier protein,
10	Tb10.61.2900	Tb10.61.2900	Tb10.61.2900	1	100	42	-6	627	3334072	hypothetical protein, conserved
10	Tb10.61.3020	Tb10.61.3020	Tb10.61.3020	2	100	8	-12	1824	3320463	hypothetical protein, conserved
10	Tb10.61.3120	Tb10.61.3120	Tb10.61.3120	4	100	64	-324	1296	3302533	hypothetical protein, conserved
10	Tb10.61.3160	Tb10.61.3160	Tb10.61.3160	2	100	38	-141	654	3288731	hypothetical protein, conserved

10	Tb10.70.4940	Tb10.70.4940	Tb10.70.4940	1	100	19	-516	1722	698117	hypothetical protein, conserved
10	Tb10.70.5020	Tb10.70.5020	Tb10.70.5020	1	100	22	-84	2307	680402	intraflagellar transport protein IFT81
10	Tb10.70.7450	Tb10.70.7450	Tb10.70.7450	2	100	13	-234	1059	196401	hypothetical protein, conserved
10	Tb10.70.7880	Tb10.70.7880	Tb10.70.7880	2	100	240	-168	1926	115804	hypothetical protein, conserved
11	Tb11.01.2270	Tb11.01.2270	Tb11.01.2270	1	100	6	-297	1590	2741474	hypothetical protein, conserved
11	Tb11.01.2820	Tb11.01.2820	Tb11.01.2820	1	100	17	-15	699	2893573	exosome-associated protein 4
11	Tb11.01.3410	Tb11.01.3410	Tb11.01.3410	2	100	24	-39	1263	3042341	hypothetical protein, conserved
11	Tb11.01.7050	Tb11.01.7050	Tb11.01.7050	2	100	53	-15	3657	4033627	hypothetical protein, conserved
11	Tb11.01.7060	Tb11.01.7060	Tb11.01.7060	1	100	12	-21	1029	4035771	hypothetical protein, conserved
11	Tb11.01.8510	Tb11.01.8510	Tb11.01.8510	2	100	129	-141	1659	4400935	t-complex protein 1, alpha subunit
11	Tb11.02.1660	Tb11.02.1660	Tb11.02.1660	1	100	23	-90	1524	1199825	hypothetical protein, conserved
11	Tb11.02.3350	Tb11.02.3350	Tb11.02.3350	1	100	4	-99	2874	1568974	hypothetical protein, conserved
11	Tb11.12.0010	Tb11.12.0010	Tb11.12.0010	1	100	12	-279	1863	2355969	hypothetical protein, conserved
11	Tb11.18.0007	Tb11.18.0007	Tb11.18.0007	1	100	9	-135	1338	608977	hypothetical protein, conserved
11	Tb11.18.0010	Tb11.18.0010	Tb11.18.0010	1	100	31	-3	1269	595624	hypothetical protein, conserved
11	Tb11.22.0004	Tb11.22.0004	Tb11.22.0004	3	100	173	-174	1728	660345	hypothetical protein, conserved
11	Tb11.55.0023	Tb11.55.0023	Tb11.55.0023	2	100	26	-432	1419	471227	hypothetical protein, conserved
6	Tb927.6.2020	Tb927.6.2020	Tb927.6.2020	3	99	9	-27	2238	658099	intergrin alpha chain protein,
3	Tb927.3.4170	Tb927.3.4170	Tb927.3.4170	2	98	43	-54	414	1172073	hypothetical protein, conserved
3	Tb927.3.5620	Tb927.3.5620	Tb927.3.5620	3	98	248	-30	3069	1568735	hypothetical protein, conserved
4	Tb927.4.2450	Tb927.4.2450	Tb927.4.2450	3	98	54	-6	1242	636840	thioredoxin,
4	Tb927.4.2660	Tb927.4.2660	Tb927.4.2660	2	98	65	-39	309	706596	hypothetical protein, conserved
6	Tb927.6.2230	Tb927.6.2230	Tb927.6.2230	4	98	78	-57	2385	704537	hypothetical protein, conserved
8	Tb927.8.6630	Tb927.8.6630	Tb927.8.6630	2	98	85	-84	639	1916177	hypothetical protein
9	Tb09.160.3090	Tb09.160.3090	Tb09.160.3090	5	98	41	-6	2508	783199	HSP70-like protein
9	Tb09.211.0330	Tb09.211.0330	Tb09.211.0330	3	98	45	-378	1383	1322010	chaperone protein DNAJ,
9	Tb09.v1.0420	Tb09.v1.0420	Tb09.v1.0420	4	98	119	-6	603	1234736	hypothetical protein, conserved
7	Tb927.7.4770	Tb927.7.4770	Tb927.7.4770	4	97	207	-75	597	1265132	cyclophilin-type peptidyl-prolyl isomerase

7	Tb927.7.4920	Tb927.7.4920	Tb927.7.4920	4	97	86	-405	2226	1299282	hypothetical protein, conserved
9	Tb09.211.1010	Tb09.211.1010	Tb09.211.1010	6	97	133	-21	990	1449935	cholinephosphotransferase 2,
9	Tb09.211.2230	Tb09.211.2230	Tb09.211.2230	2	97	35	-33	723	1664452	hypothetical protein, conserved
10	Tb10.26.0990	Tb10.26.0990	Tb10.26.0990	3	97	81	-369	591	2629586	hypothetical protein, conserved
10	Tb10.61.0150	Tb10.61.0150	Tb10.61.0150	2	97	69	-195	1539	3893675	inosine-5'-monophosphate dehydrogenase
3	Tb927.3.5490	Tb927.3.5490	Tb927.3.5490	2	96	24	-42	2070	1535250	flagellar transport protein,
5	Tb927.5.480	Tb927.5.480	Tb927.5.480	2	96	23	-96	768	150618	hypothetical protein, conserved
9	Tb09.160.1180	Tb09.160.1180	Tb09.160.1180	7	96	70	-9	1740	523916	nucleolar protein
7	Tb927.7.4880	Tb927.7.4880	Tb927.7.4880	3	95	81	-36	684	1281517	hypothetical protein, conserved
9	Tb09.160.1160	Tb09.160.1160	Tb09.160.1160	7	95	72	-9	2283	521183	nucleolar protein
9	Tb09.211.1020	Tb09.211.1020	Tb09.211.1020	6	95	118	-153	972	1452422	cholinephosphotransferase 2,
10	Tb10.26.0230	Tb10.26.0230	Tb10.26.0230	4	95	22	-324	1083	2781988	hypothetical protein, conserved
11	Tb11.02.3790	Tb11.02.3790	Tb11.02.3790	3	95	20	-84	1155	1673109	hypothetical protein, conserved
11	Tb11.02.3810	Tb11.02.3810	Tb11.02.3810	3	95	19	-324	1380	1677123	hypothetical protein, conserved
6	Tb927.6.3170	Tb927.6.3170	Tb927.6.3170	2	94	16	-72	1446	944321	hypothetical protein, conserved
6	Tb927.6.5100	Tb927.6.5100	Tb927.6.5100	7	94	667	-21	2526	1394258	serine/threonine-protein kinase,
7	Tb927.7.980	Tb927.7.980	Tb927.7.980	3	94	17	-255	2877	252806	hypothetical protein, conserved
8	Tb927.8.5070	Tb927.8.5070	Tb927.8.5070	4	94	189	-15	393	1497534	hypothetical protein, conserved
9	Tb09.160.1140	Tb09.160.1140	Tb09.160.1140	3	94	18	-189	1161	517076	electon transport protein,
2	Tb927.2.6050	Tb927.2.6050	Tb927.2.6050	3	92	26	-12	2568	1092934	beta prime COP protein
3	Tb927.3.5070	Tb927.3.5070	Tb927.3.5070	3	92	14	-111	1938	1421239	hypothetical protein, conserved
10	Tb10.389.1390	Tb10.389.1390	Tb10.389.1390	5	92	27	-426	798	2925678	hypothetical protein, conserved
4	Tb927.4.1590	Tb927.4.1590	Tb927.4.1590	2	88	42	-237	1293	410232	hypothetical protein, conserved
4	Tb927.4.680	Tb927.4.680	Tb927.4.680	3	88	25	-261	1140	194973	hypothetical protein, conserved
10	Tb10.26.0920	Tb10.26.0920	Tb10.26.0920	2	88	8	-93	750	2648015	hypothetical protein, conserved
10	Tb10.26.0900	Tb10.26.0900	Tb10.26.0900	7	87	52	-45	4671	2656714	hypothetical protein, conserved
7	Tb927.7.1550	Tb927.7.1550	Tb927.7.1550	2	86	7	-165	1464	386392	RNA-editing 3' terminal uridylyl transferase
11	Tb11.03.0990	Tb11.03.0990	Tb11.03.0990	2	85	117	-192	3708	28840	expression site-associated gene 4,

4	Tb927.4.2550	Tb927.4.2550	Tb927.4.2550	5	84	12	-90	909	671279	hypothetical protein, conserved
10	Tb10.70.1440	Tb10.70.1440	Tb10.70.1440	2	75	4	-93	1770	1391982	GPI anchor biosynthesis protein
7	Tb927.7.7290	Tb927.7.7290	Tb927.7.7290	2	67	3	-342	3048	2106650	hypothetical protein, conserved
10	Tb10.26.0980	Tb10.26.0980	Tb10.26.0980	2	67	3	-198	1452	2631353	hypothetical protein, conserved
11	Tb11.01.4690	Tb11.01.4690	Tb11.01.4690	6	67	21	-330	2124	3443911	hypothetical protein, conserved
11	Tb11.02.2740	Tb11.02.2740	Tb11.02.2740	2	67	3	-87	1254	1429902	aspartate aminotransferase, mitochondrial
10	Tb10.389.0710	Tb10.389.0710	Tb10.389.0710	4	64	22	-90	1011	3047285	acetyltransferase,
10	Tb10.26.0240	Tb10.26.0240	Tb10.26.0240	3	60	5	-159	1362	2780365	hypothetical protein, conserved
8	Tb927.8.5900	Tb927.8.5900	Tb927.8.5900	4	57	7	-6	396	1735242	hypothetical protein, conserved
5	Tb927.5.2890	Tb927.5.2890	Tb927.5.2890	3	56	18	-102	6594	913081	hypothetical protein, conserved
6	Tb927.6.1460	Tb927.6.1460	Tb927.6.1460	5	56	39	-111	1356	515869	cyclin 3,mitotic cyclin,
6	Tb927.6.1490	Tb927.6.1490	Tb927.6.1490	2	56	16	-126	1311	524460	leucine-rich repeat protein (LRRP)
4	Tb927.4.4910	Tb927.4.4910	Tb927.4.4910	2	51	39	-210	1413	1345523	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor,
6	Tb927.6.2960	Tb927.6.2960	Tb927.6.2960	2	50	6	-129	3207	882763	epsilon-adaptin,
8	Tb927.8.1600	Tb927.8.1600	Tb927.8.1600	2	47	36	-228	1983	531183	lysyl-tRNA synthetase,
9	Tb09.244.2830	Tb09.244.2830	Tb09.244.2830	5	29	24	-42	1083	2292101	hypothetical protein, conserved
7	Tb927.7.7490	Tb927.7.7490	Tb927.7.7490	2	25	8	-204	1041	2165127	hypothetical protein, conserved

CATEGORY 3

After obtaining ORF extension data for genes where I had upstream SAS predictions, I ran a predictive algorithm on 1,768 protein-coding genes for which I had no SAS data. 47 genes were predicted to have upstream ORF extensions, compared to the TrypDB ATG predictions. Nine of the 47 were predicted to be very unlikely extensions because polyY tracts that probably represent splice signals were apparent by visual inspection of the predicted extension sequences. (Prior to obtaining additional SAS data there were 74 genes predicted to have upstream extensions: the additional data, from 76-bp reads, showed SAS that are consistent with the use of the upstream ATG.) It is not known why the first in-frame ATG was not used to predict these ORFs in TrypDB. Some of these predicted extensions might be incorrect, if the upstream SAS represents an intermediate or aberrant splice site and the tag abundance was too low to find a further downstream SAS.

Chr	GeneID	TryDB	TriTryDB	Relative ATG Position	Predicted ORF Start	Predicted ORF Length	Original ORF Length	Abbreviated Description from TrypDB
1	Tb927.1.3730	Tb927.1.3730	Tb927.1.3730	-93	793438	267	174	hypothetical protein
3	Tb927.3.2500	Tb927.3.2500	Tb927.3.2500	-102	632487	1290	1188	hypothetical protein
3	Tb927.3.3230	Tb927.3.3230	Tb927.3.3230	-138	835188	1968	1830	hypothetical protein, conserved
4	Tb927.4.1410	Tb927.4.1410	Tb927.4.1410	-297	367410	1845	1548	hypothetical protein, conserved
4	Tb927.4.2400	Tb927.4.2400	Tb927.4.2400	-117	619197	4404	4287	hypothetical protein, conserved
4	Tb927.4.2490	Tb927.4.2490	Tb927.4.2490	-105	652953	1353	1248	hypothetical protein, conserved
4	Tb927.4.2580	Tb927.4.2580	Tb927.4.2580	-27	677636	3048	3021	hypothetical protein, conserved
4	Tb927.4.2720	Tb927.4.2720	Tb927.4.2720	-90	716452	2310	2220	hypothetical protein, conserved
4	Tb927.4.4130	Tb927.4.4130	Tb927.4.4130	-213	1094078	2784	2571	hypothetical protein, conserved
4	Tb927.4.4670	Tb927.4.4670	Tb927.4.4670	-87	1291160	1539	1452	hypothetical protein, conserved
5	Tb927.5.2800	Tb927.5.2800	Tb927.5.2800	-111	888774	3255	3144	hypothetical protein, conserved
5	Tb927.5.3760	Tb927.5.3760	Tb927.5.3760	-189	1170071	1485	1296	meiotic recombination protein SPO11,
7	Tb927.7.310	Tb927.7.310	Tb927.7.310	-141	58581	1029	888	hypothetical protein, conserved
7	Tb927.7.4150	Tb927.7.4150	Tb927.7.4150	-255	1104777	3756	3501	hypothetical protein, conserved
7	Tb927.7.7010	Tb927.7.7010	Tb927.7.7010	-105	2007048	570	465	hypothetical protein, conserved
7	Tb927.7.7550	Tb927.7.7550	Tb927.7.7550	-111	2203046	426	315	hypothetical protein

8	Tb927.8.4560	Tb927.8.4560	Tb927.8.4560	-21	1354850	3777	3756	hypothetical protein, conserved
8	Tb927.8.5560	Tb927.8.5560	Tb927.8.5560	-192	1638560	1794	1602	hypothetical protein, conserved
9	Tb09.142.0310	Tb09.142.0310	Tb09.142.0310	-45	158876	1029	984	expression site-associated gene 1 (ESAG1) protein
9	Tb09.160.0940	Tb09.160.0940	Tb09.160.0940	-120	481147	396	276	hypothetical protein
9	Tb09.160.3980	Tb09.160.3980	Tb09.160.3980	-360	894498	2997	2637	hypothetical protein, conserved
9	Tb09.244.0340	Tb09.244.0340	Tb09.244.0340	-42	2955321	276	234	hypothetical protein
9	Tb09.244.2070	Tb09.244.2070	Tb09.244.2070	-189	2472102	1020	831	GTP-binding protein,
9	Tb09.v1.0030	Tb09.v1.0030	Tb09.v1.0030	-54	274471	390	336	hypothetical protein, conserved
9	Tb09.v1.0370	Tb09.v1.0370	Tb09.v1.0370	-273	1224471	2772	2499	hypothetical protein, conserved
9	Tb09.v1.0450	Tb09.v1.0450	Tb09.v1.0450	-51	1245366	969	918	hypothetical protein, conserved
9	Tb09.v1.0470	Tb09.v1.0470	Tb09.v1.0470	-180	1251263	1110	930	hypothetical protein, conserved
10	Tb10.26.0260	Tb10.26.0260	Tb10.26.0260	-189	2775542	1476	1287	hypothetical protein, conserved
10	Tb10.26.0780	Tb10.26.0780	Tb10.26.0780	-33	2677861	1005	972	hypothetical protein, conserved
10	Tb10.389.0470	Tb10.389.0470	Tb10.389.0470	-204	3095858	2241	2037	calpain-like cysteine peptidase,
10	Tb10.389.0600	Tb10.389.0600	Tb10.389.0600	-192	3070836	3108	2916	hypothetical protein, conserved
10	Tb10.61.1190	Tb10.61.1190	Tb10.61.1190	-99	3659800	5037	4938	hypothetical protein, conserved
10	Tb10.61.2747	Tb10.61.2747	Tb10.61.2747	-3	3358370	1584	1581	transporter,
10	Tb10.6k15.3320	Tb10.6k15.3320	Tb10.6k15.3320	-54	1839419	1995	1941	ATP-binding cassette protein,
10	Tb10.70.1380	Tb10.70.1380	Tb10.70.1380	-39	1402441	612	573	40S ribosomal protein S9,
11	Tb11.02.2240	Tb11.02.2240	Tb11.02.2240	-183	1311482	3339	3156	calpain-like protein,
11	Tb11.02.2950	Tb11.02.2950	Tb11.02.2950	-99	1483471	456	357	ATPase subunit 9,
11	Tb11.02.3720	Tb11.02.3720	Tb11.02.3720	-351	1655173	2532	2181	hypothetical protein, conserved

The following predicted extensions are unlikely: they contain a probable SAS.

3	Tb927.3.2520	Tb927.3.2520	Tb927.3.2520	-114	636507	1089	975	expression site-associated gene 1,

5	Tb927.5.1670	Tb927.5.1670	Tb927.5.1670	-363	526728	1410	1047	hypothetical protein, conserved
6	Tb927.6.1450	Tb927.6.1450	Tb927.6.1450	-126	514081	969	843	hypothetical protein, conserved
6	Tb927.6.2920	Tb927.6.2920	Tb927.6.2920	-294	874885	2253	1959	hypothetical protein, conserved
6	Tb927.6.2960	Tb927.6.2960	Tb927.6.2960	-129	882763	3207	3078	epsilon-adaptin,
8	Tb927.8.6670	Tb927.8.6670	Tb927.8.6670	-252	1933066	1785	1533	hypothetical protein, conserved
9	Tb09.160.0370	Tb09.160.0370	Tb09.160.0370	-144	335139	1284	1140	hypothetical protein, conserved
9	Tb09.160.5070	Tb09.160.5070	Tb09.160.5070	-330	1081428	1152	822	DNA topoisomerase type IB small subunit
8	Tb927.8.2810	Tb927.8.2810	Tb927.8.2810	-381	839490	2988	2607	5'-3' exonuclease XRNC,