Additional Data File 5: Putative Hymenoptera-specific Genes

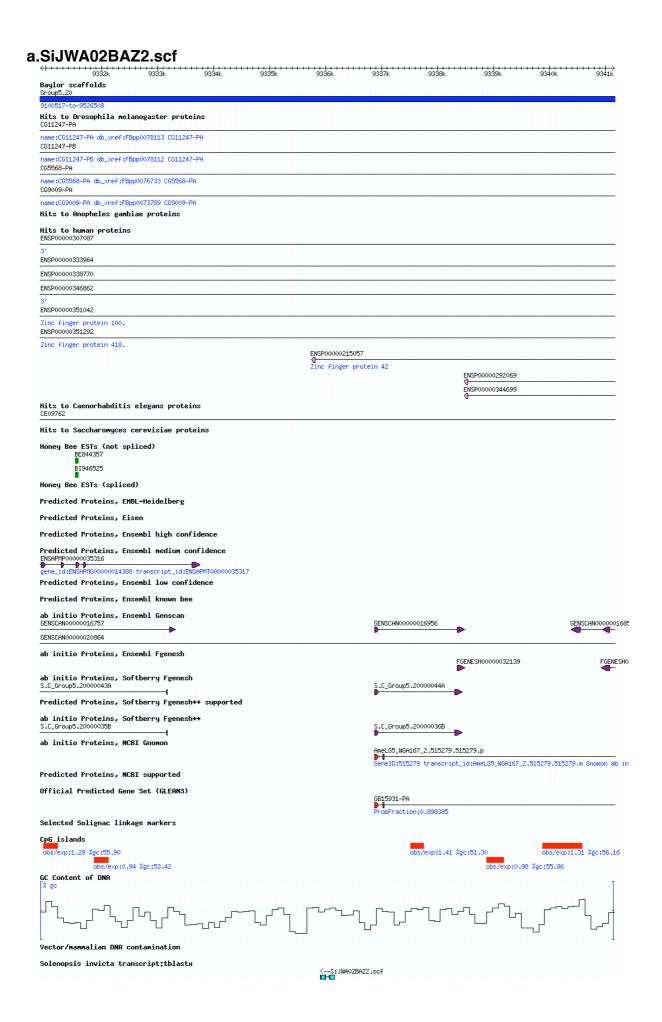
The 10,000 bp region around the ant-bee alignment of every putative Hymenopteraspecific gene is depicted relative to the annotated honey bee genome.

The honey bee genome assembly 2.0 chromosome coordinates are at the top of each image. Below, the alignment of proteins from *A. gambiae, C. elegans, D. melanogaster, H. sapiens* and *S. cerevisiae* are shown. Additionally, honey bee proteins predicted by several different groups (Eisen, EMBL, Ensembl, NCBI) as well as with multiple *ab initio* prediction methods are displayed. Finally, genes from the honey bee Official Gene Set (GLEAN3) are in red. The alignment of fire ant transcripts are in light blue; 5' to 3' direction is indicated by an arrow.

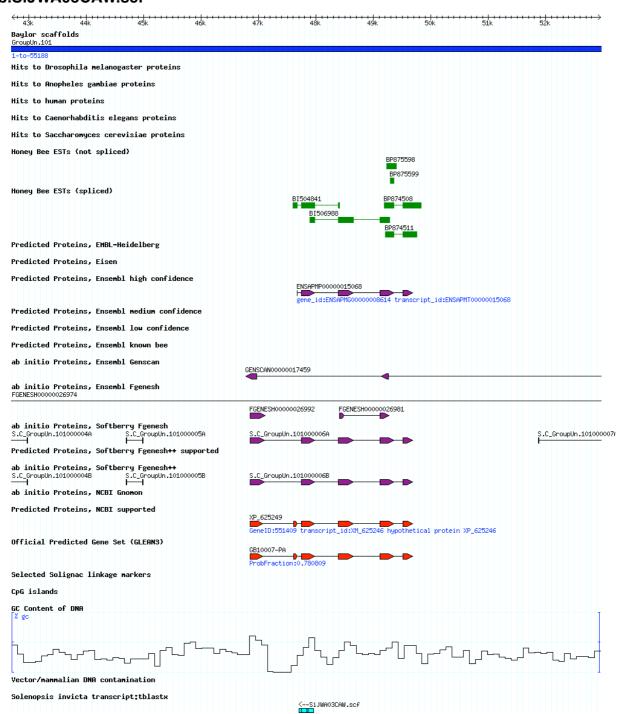
Coordinates may differ from those in Table 3 because Table 3 was generated using assembly 4 of the honey bee genome.

Index of subfigures:

| | page |
|------------------------------------|------|
| a. SiJWA02BAZ2.scf | 2 |
| b. SiJWA03CAW.scf | 3 |
| c. SiJWA12ACK.scf | 4 |
| d. SiJWB12BCQ.tag5_B12_04.scf | 5 |
| e. SiJWC11BAT.scf | 6 |
| f. SI.CL.8.cl.881.Contig1 | 7 |
| g. SI.CL.8.cl.843.SiJWH04BDO2.scf | 8 |
| h. SI.CL.19.cl.1938.Contig1 | 9 |
| i. SI.CL.19.cl.1953.SiJWC11BBX.scf | 10 |
| j. SI.CL.23.cl.2326.Contig1 | 11 |
| k. SI.CL.26.cl.2688.Contig1 | 12 |
| I. SI.CL.33.cl.3311.Contig1 | 13 |
| m.SI.CL.33.cl.3384.Contig1 | 14 |
| n. SI.CL.35.cl.3595.Contig1 | 15 |
| o. SiJWE02BBO2.scf | 16 |
| p. SiJWF07BCC.tag5_F07_11.scf | 17 |
| q. SiJWG01BDU2.scf | 18 |
| r. SiJWG03ACB.scf | 19 |
| s. SiJWH02AAN.scf | 20 |
| t. SiJWH05BDPR5A08.scf | 21 |
| u. SiJWH05BDV2.scf | 22 |
| v. SiJWH08AAT.scf | 23 |
| w. SiJWH08ADY | 24 |

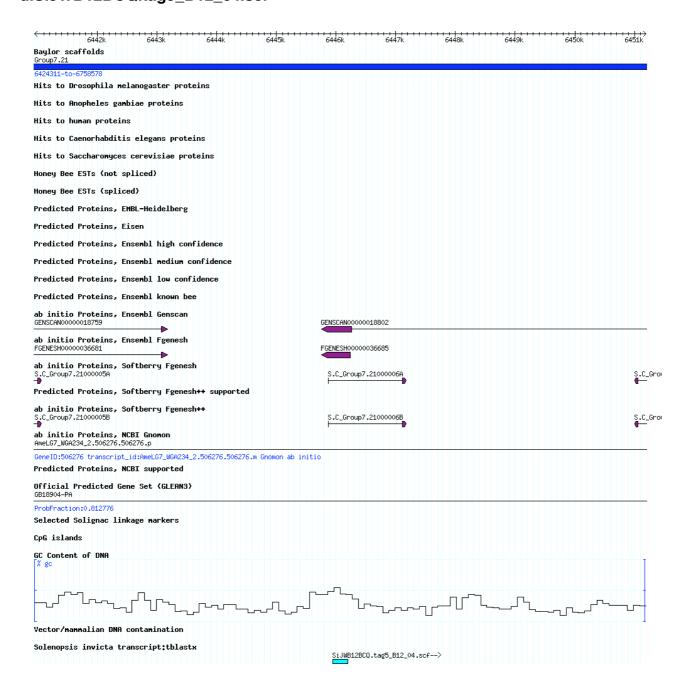


b.SiJWA03CAW.scf



c.SiJWA12ACK.scf 7836k Baylor scaffolds Group3.19 7840k 7841k 7842k 7843k 7844k 7845k Hits to Drosophila melanogaster proteins CG17181-PH name:cest-PA do_xmef:FBpp000293 Escargot pro CG3956-PA name:sna-PA do_xmef:FBpp000298 Snail protei CG4158-PA name:Sur-PA db_xref:FBpp0079553 CG7627-PA name:CG7627-PA db_xref:FBpp00792 Zn-finger, C2H2 type ENSANGP00000021575 Zinc finger protein SLUG ENSP00000270443 ENSP00000282869 ENSP00000313443 Zinc Finger protein 491. ENSP00000340524 PREDICTED: snail homol ENSP00000332595 Zinc finger protein 92 ENSP00000333199 ___a Zinc finger protein 578. Hits to Caenorhabditis elegans proteins CE01196 Guiloshit cyclophi CE01906 CE04674 CE26370 ATP-binding transport family CE37081 aromatic-L-amino-acid decarboxylase Hits to Saccharomyces cerevisiae proteins NP_011797.1 Honey Bee ESTs (not spliced) BI510609 Honey Bee ESTs (spliced) Predicted Proteins, EMBL-Heidelberg Predicted Proteins, Eisen Predicted Proteins, Ensembl high confidence Predicted Proteins, Ensembl medium confidence Predicted Proteins, Ensembl low confidence Predicted Proteins, Ensembl known bee ab initio Proteins, Ensembl Genscan GENSCAN00000020219 GENSCAN00000020248 ab initio Proteins, Softberry Egenesh Predicted Proteins, Softberry Fgenesh++ supported ab initio Proteins, Softberry Fgenesh++ ab initio Proteins, NCBI Gnomon Predicted Proteins, NCBI supported Official Predicted Gene Set (GLEANS) Selected Solignac linkage markers CpG islands GC Content of DNA Solenopsis invicta transcript:tblastx <--SiJWA12ACK.scf

d.SiJWB12BCQ.tag5_B12_04.scf

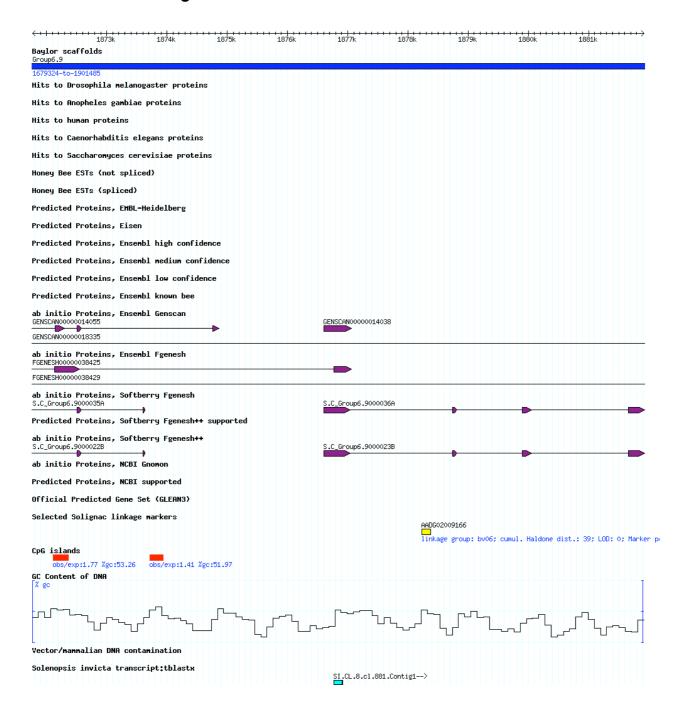


e.SiJWC11BAT.scf 1174k 1175k 1176k 1177k 1178k 1179k 1173k 1171 Baylor scaffolds Group14.2 Hits to Drosophila melanogaster proteins CG2082-PA CG2052-PB 2-PB db_xref:F8pp0088197 name:CG30035-PB db_xref:FBpp0087179 CG30035-PB C08234-PH db_xref:F8pp0087180 RT19440p C08234-PB db_xref:F8pp0087181 RT19440p naxe:C68234-PB db_xref:F8pp0087181 RT19440p Hits to human proteins ENSP00000221735 ENSP00000293471 zinc finger protein 419 ENSP00000306372 Zinc finger protein 38 ENSP00000312519 Zinc Finger protein K0X23 ENSP00000222697 Zinc Finger protein 3228 ENSP00000327928 **nc Finger protein 548 zinc finger prote ENSP00000337555 zinc finger prote ENSP00000339823 zinc finger protei ENSP00000341673 ENSP00000353728 Hits to Caenorhabditis elegans proteins CE22901 CE26263 Hits to Saccharomyces cerevisiae proteins NP_009700.1 Honey Bee ESTs (not spliced) 81512931 81509073 81503506 Honey Bee ESTs (spliced) B1506152 BI514170 Predicted Proteins, EMBL-Heidelberg Predicted Proteins, Ensembl high confidence Predicted Proteins, Ensembl medium confidence Predicted Proteins, Ensembl known bee ab initio Proteins, Ensembl Genscan GENSCHM0000008543 GENSCHM0000008569 GENSCHM00000008544 ab initio Proteins, Ensembl Fgenesh FGENESH00000024685 ab initio Proteins, Softberry Fgenesh Predicted Proteins, Softberry Fgenesh++ supported ab initio Proteins, Softberry Fgenesh++ S.C_Group14.2000106B ab initio Proteins, NCBI Gnomon Predicted Proteins, NCBI supported Official Predicted Gene Set (GLEANS) Selected Solignac linkage markers 51512301 Linkage group: bv14; cuwul. Haldone dist.: 51.2; L00: 6.9; Marker pos: 74 CpG islands

SiJWC11BAT.scf-->

L Vector/mammalian DNA contamination

f. SI.CL.8.cl.881.Contig1

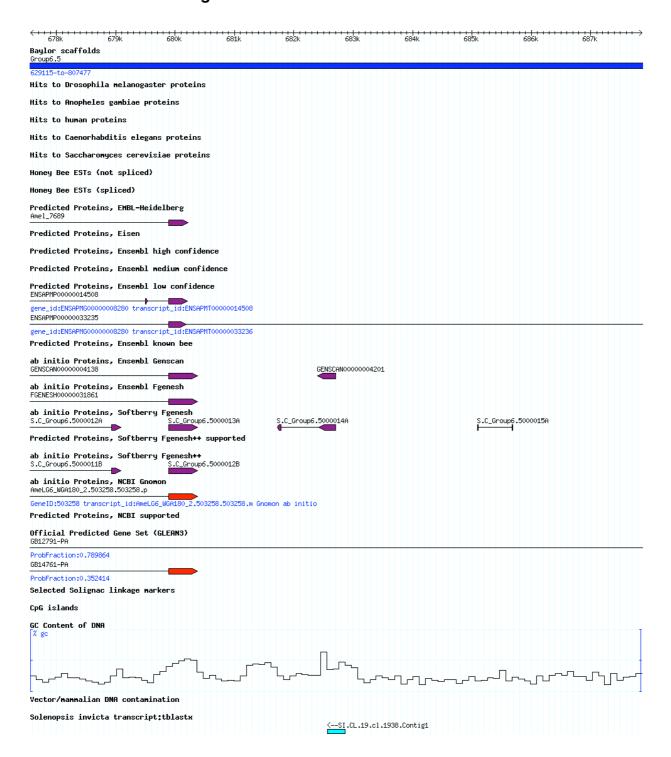


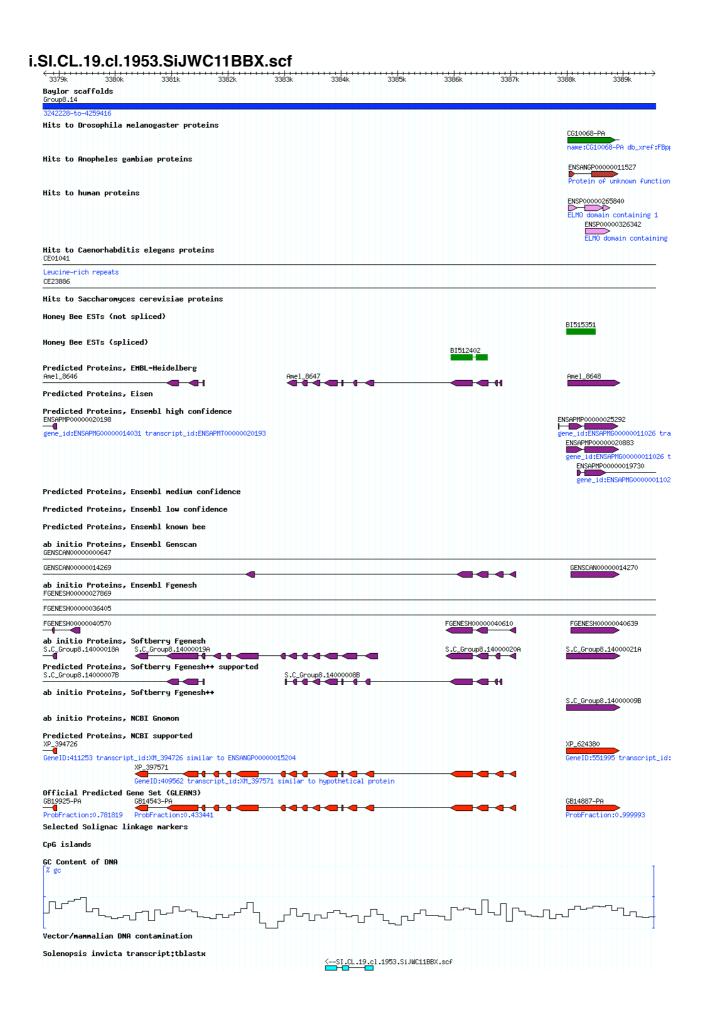
g.SI.CL.8.cl.843.SiJWH04BDO2.scf

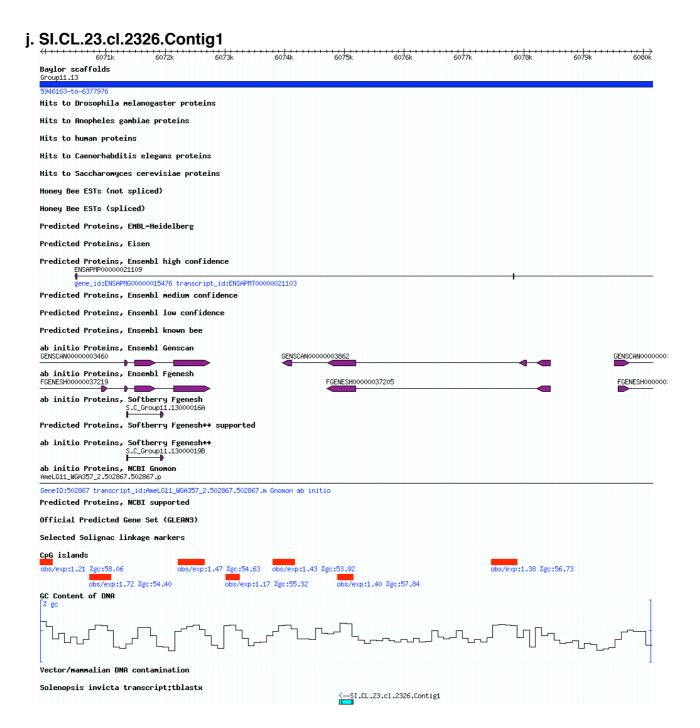


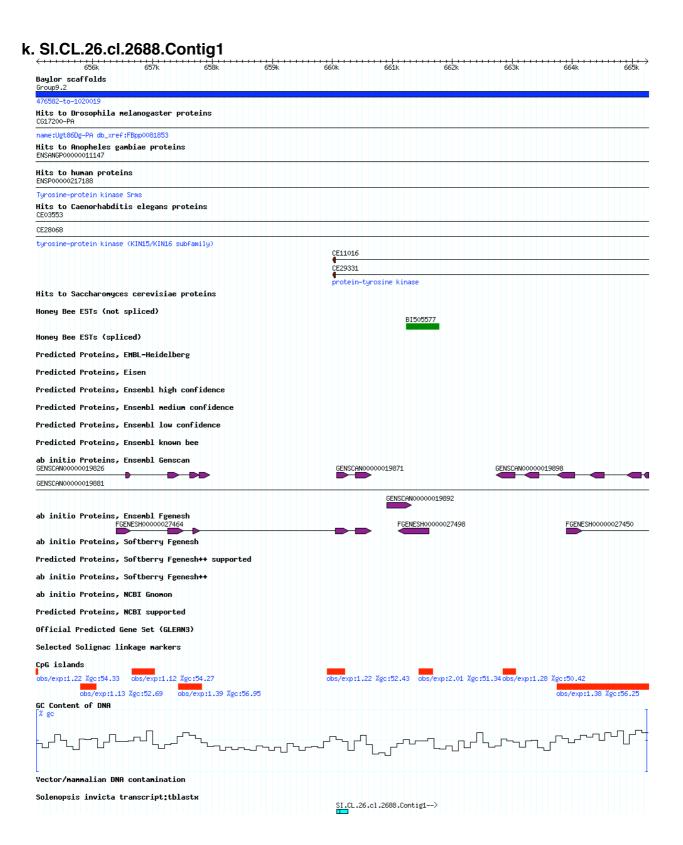
^{*} SI.CL.8.cl.843.SiJWH04BDO2.scf aligns close to one extreme end of the unanchored genomic scaffold GroupUn.742, which is why there is only little annotation data to the right side of SI.CL.8.cl.843.SiJWH04BDO2.scf.

h. SI.CL.19.cl.1938.Contig1





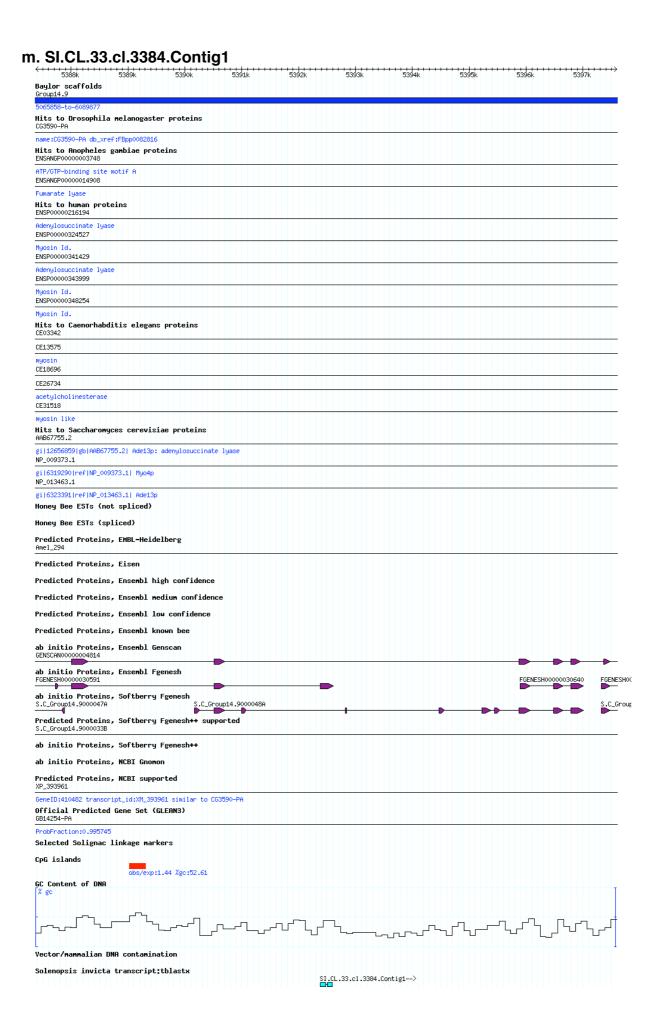


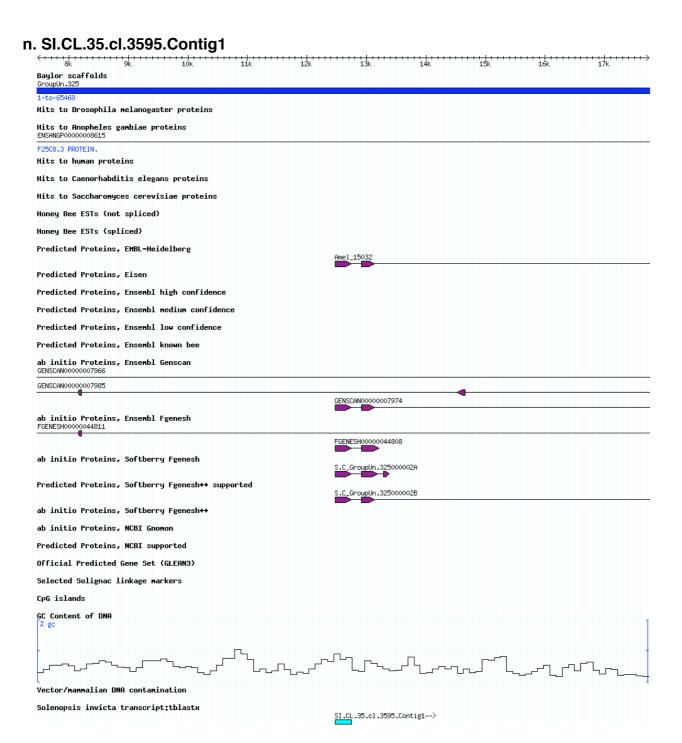


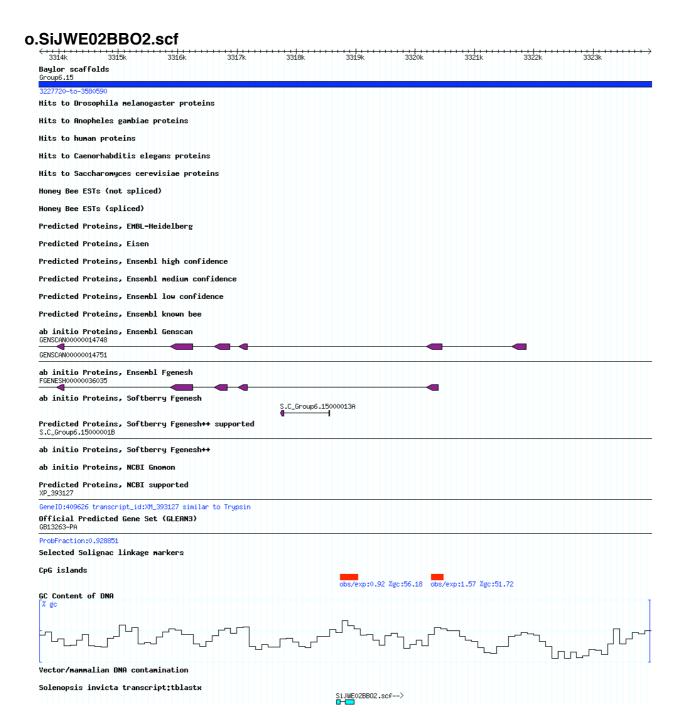
I. SI.CL.33.cl.3311.Contig1 1190k 1191k 1182k Baylor scaffolds Group14.2 Hits to Brosophila melanogaster proteins CG2052-PA Hits to human proteins ENSP00000221735 Zinc finger protes ENSP00000312519 47 kDa protein ENSP00000317125 72 kDa protein ENSP00000318811 Zinc finger protei ENSP00000321812 ENSPOODO Hits to Saccharomyces cerevisiae proteins NP_009700.1 Predicted Proteins, Eisen Predicted Proteins, Ensembl high confidence Predicted Proteins, Ensembl medium confidence Predicted Proteins, Ensembl low confidence Predicted Proteins, Ensembl known bee ab initio Proteins, Ensembl Genscan GENSCAN00000008544 GENSCAN0000000871 S.C_Group14.200 Predicted Proteins, Softberry Fgenesh++ supported ab initio Proteins, Softberry Fgenesh++
S.C.Oroup14.2000106B
ab initio Proteins, NCBI Gnomon Predicted Proteins, NCBI supported Official Predicted Gene Set (GLEANS) CpG islands

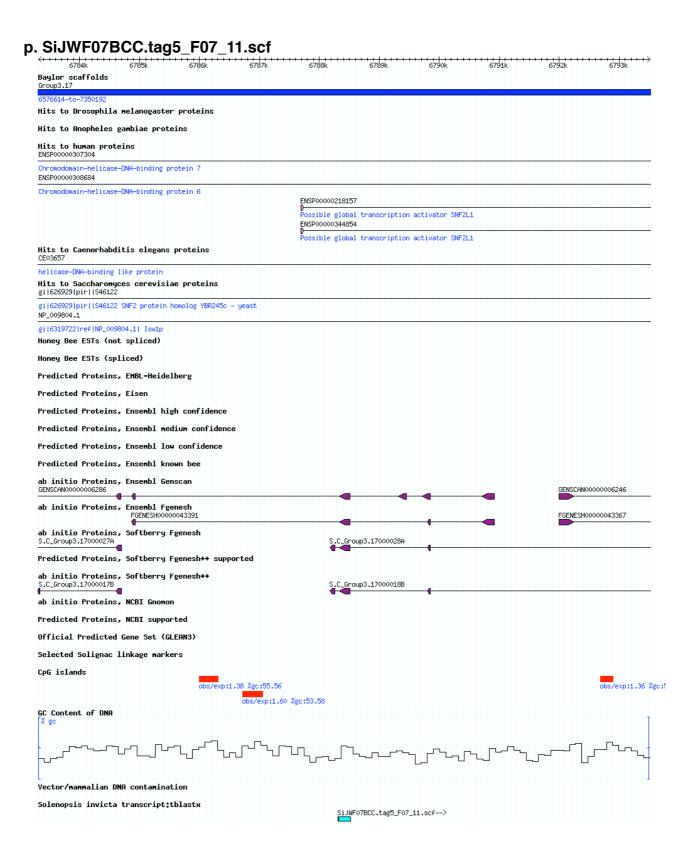
SI.CL.33.c1.3311.Contig1-->

Solenopsis invicta transcript:tblastx

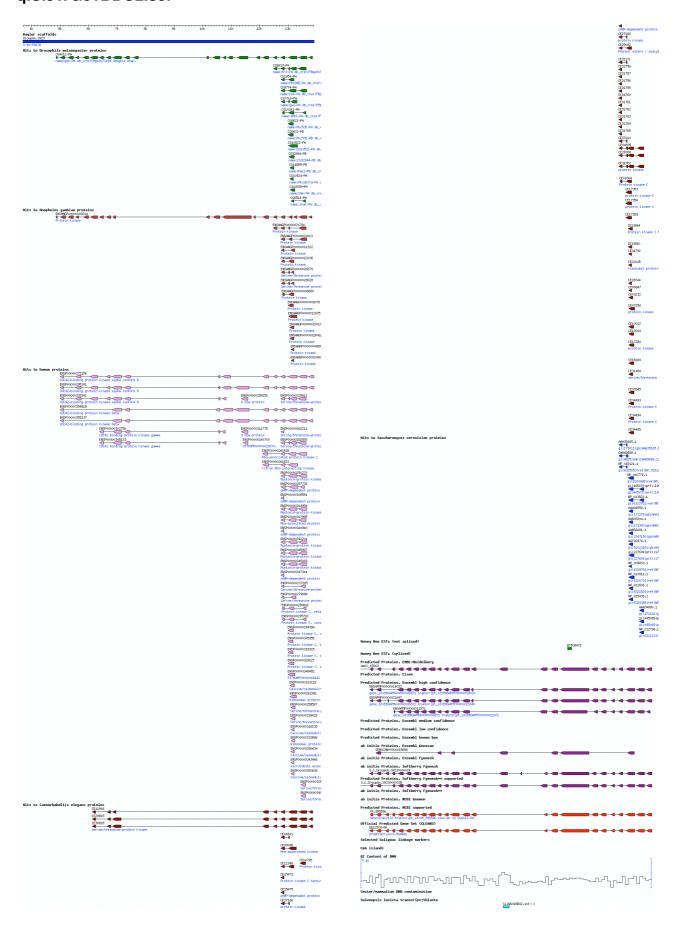




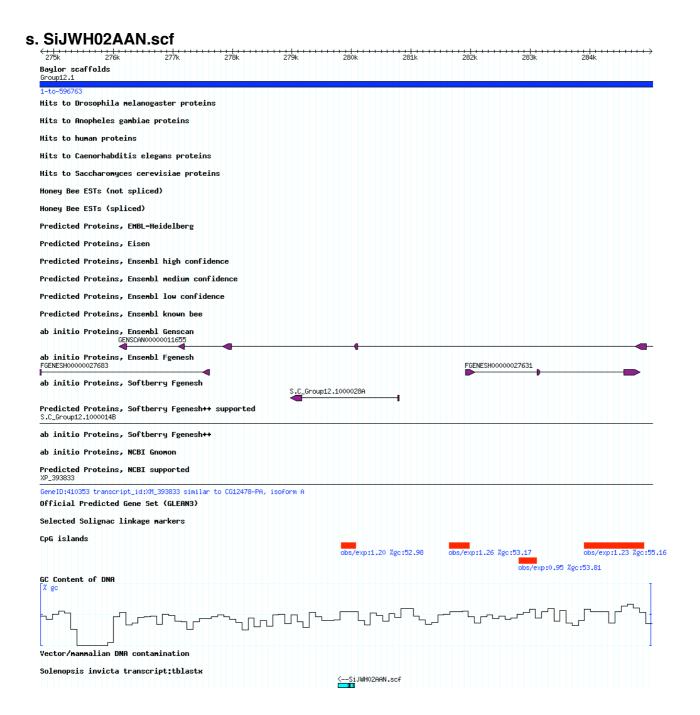


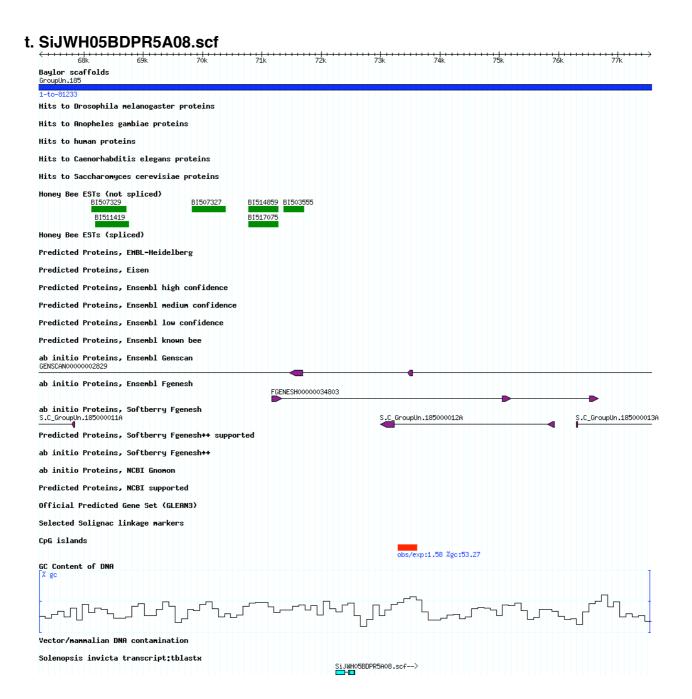


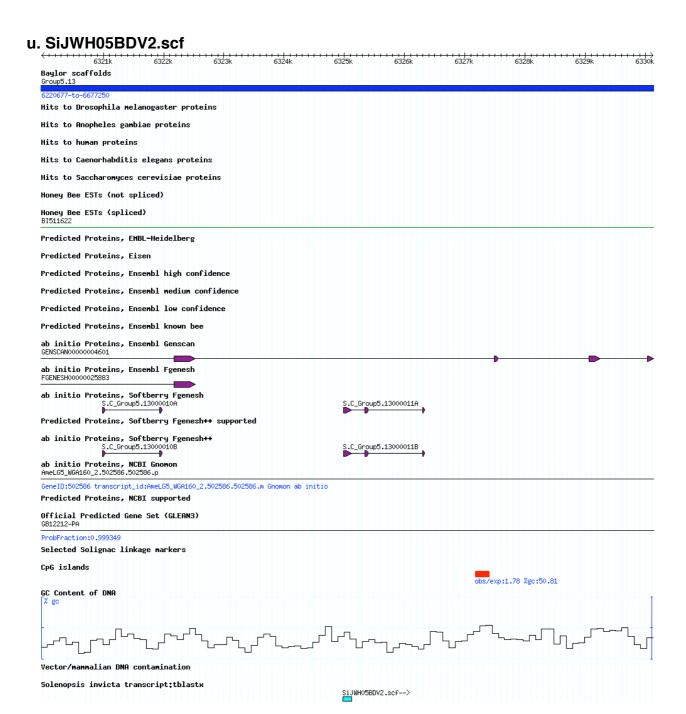
q.SiJWG01BDU2.scf











v Si.IWH08AAT scf

| 785k 786k 787k Baylor scaffolds Group5.2 | <u></u> |
|-----------------------------------------------------------------------------------------|---------|
| 238628-to-979349 Hits to Drosophila melanogaster proteins | |
| | |
| Hits to Anopheles gambiae proteins NSAMEPOOCOCCISISI HTP/GTP-binding site motif A | |
| lits to human proteins NSP00000221735 | |
| cinc finger protein 419 INSP0000242804 | |
| Cinc finger protein 442. NSP00000244316 | |
| linc finger protein 234 NSP00000244563 | |
| cinc finger protein 311 NSP0000247533 | |
| tinc finger protein 484 isoform a NNSP00000253115 | |
| linc finger protein 426. NSP00000261560 | |
| Cinc finger protein 430. PNSP00000282292 | |
| MSP00000285274 Eino finger protein 286. MSP00000292241 | |
| INSP00000292241 Linc finger protein 20 INSP00000293725 | |
| INSP00000293725 | |
| tino Finger protein 563 NSP0000294702 | |
| Eine Finger protein GFI-1 NSP000029755 | |
| odd-skipped related 2 SNSP0000299860 Linc Finder protein 419 | |
| tinc Finger protein 419 ENSP0000300870 Linc Finger protein 267 | |
| linc Finger protein 267 NSP0000031310 | |
| tinc Finger protein 582 NSP000093889 | |
| linc Finger protein 30 NNSP0000303915 | |
| zinc finger protein 561 ENSP0000305373 | |
| Zinc Finger protein 440. ENSP00000308289 | |
| 0 kDa protein NSP00000309357 | |
| linc finger protein ZFD25. NNSP0000311540 | |
| Cinc finger protein 323. NSP00000320627 | |
| Zinc Finger protein 7 ENSP00000321406 | |
| Zinc finger protein 135. ENSP00000322194 | |
| zinc Finger protein 550 ENSP00000324518 | |
| Zino Finger protein 273 ENSP00000325038 | |
| similar to zino finger protein 616 INSP00000326249 | |
| Zino finger protein 454. ENSP00000328166 | |
| PREDICTED: similar to PREDES ENSPOROGO328619 | |
| 50 kDa protein ENSP0000328722 | |
| 70 kBa protein INSP00000329738 | |
| Zinc finger protein. ENSP0000331120 | |
| PREDICTED: similar to PREDSS PNSP00000331577 | |
| linc Finger protein 397 NSP00000332861 | |
| tinc finger protein 530 NSP00000333199 | |
| tinc finger protein 283. NSP00000333434 | <u></u> |
| REDICTED: similar to hypothetical protein | |
| NSP00000334013 | |
| Zinc finger protein 248. ENSP00000334196 Zinc finger protein 347 | |
| Zino finger protein 347 ENSP00000334164 ENSP00000334695 | |
| NSP0000358695 tinc Firger protein 529 NSP0000335042 | |
| MSP0000035002 MEDICTED: similar to Zinc Finger protein 93 MSP00000339767 | |
| NSP00000339767 Cinc finger protein 433. NSP00000340004 | |
| NSP00000340004 Cinc finger protein 564. NSP00000340132 | |
| NSP00000340132 tinc finger protein 383 NSP00000340246 | |
| MSP00000340246 tinc finger protein 550 MSP00000340514 | |
| MSP00000340514 Cinc finger protein 14 MSP00000340683 | |
| MSP00000340683 Lino Finger protein ZFP-36 MSP00000341528 | |
| MSP00000341528 DW74 protein MSP00000342002 | |
| INSP00000342002 Lino Finger protein 93 | |
| Zino Finger protein 93 INSP00000342947 72 kBs protein | |
| 72 kDa protein DNEPOGOODSIA4S Inn Finger protein ZFDZS. DNEPOGOODSIA62 | |
| and resign protects as seen | |

