**Table S2** Summary of the programs used in the pipeline

|  |  |  |
| --- | --- | --- |
| **Name** | **Function in pipeline** | **Website/Reference** |
| SEQCLEAN | EST quality control | http://compbio.dfci.harvard.edu/tgi/software/ |
| RepeatMasker | Screen repeats | http://www.repeatmasker.org/ |
| WU\_BLAST\* | Similarity search | http://blast.wustl.edu/ |
| TGICL | EST clustering and assembly | http://compbio.dfci.harvard.edu/tgi/software/ |
| GMAP | Genome mapping | [[6](#_ENREF_6)]http://research-pub.gene.com/gmap/ |
| NCBI\_BLAST | Similarity search | ftp://ftp.ncbi.nih.gov/blast/ |
| EMBOSS (getorf) | ORF prediction | [[7](#_ENREF_7)]http://emboss.sourceforge.net/ |

\* WU\_BLAST has been renamed to AB\_BLAST, which is not freely accessible. We used version 2 of WU\_BLAST in the pipeline.