GenSAS tools references:

PRINSEQ-lite

http://prinseq.sourceforge.net/

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27 (6):863-864. doi:10.1093/bioinformatics/btr026

RepeatMasker and RepeatModeler

http://www.repeatmasker.org/

Augustus

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33:W465-W467. doi:10.1093/Nar/Gki458

Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309-W312. doi:10.1093/Nar/Gkh379

GeneMarkES

http://exon.gatech.edu/GeneMark/

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 33 (20):6494-6506. doi:10.1093/nar/gki937

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18 (12):1979-1990. doi:10.1101/gr.081612.108

GeneMarkS

http://exon.gatech.edu/GeneMark/

Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29 (12):2607-2618. doi:10.1093/nar/29.12.2607

Genscan

http://genes.mit.edu/GENSCAN.html

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268 (1):78-94. doi:10.1006/jmbi.1997.0951

Glimmer3

https://ccb.jhu.edu/software/glimmer/

Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23 (6):673-679. doi:10.1093/bioinformatics/btm009

GlimmerM

http://www.cbcb.umd.edu/software/glimmerm/

Salzberg SL, Pertea M, Delcher AL, Gardner MJ, Tettelin H (1999) Interpolated Markov models for eukaryotic gene finding. Genomics 59 (1):24-31. doi:10.1006/geno.1999.5854

SNAP

http://korflab.ucdavis.edu/software.html

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5. doi:10.1186/1471-2105-5-59

BLAST+

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST plus : architecture and applications. BMC Bioinformatics 10. doi:10.1186/1471-2105-10-421

BLAT

https://genome.ucsc.edu/FAQ/FAQblat.html

Kent WJ (2002) BLAT - The BLAST-like alignment tool. Genome Res 12 (4):656-664. doi:10.1101/Gr.229202

PASA

http://pasapipeline.github.io/

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 31 (19):5654-5666. doi:10.1093/Nar/Gkg770

TopHat2

https://ccb.jhu.edu/software/tophat/index.shtml

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14 (4):R36. doi:10.1186/gb-2013-14-4-r36

getorf

http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html

RNAmmer

http://www.cbs.dtu.dk/services/RNAmmer/

Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35 (9):3100-3108. doi:10.1093/nar/gkm160

tRNAScan-SE

http://lowelab.ucsc.edu/tRNAscan-SE/

Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25 (5):955-964. doi:10.1093/Nar/25.5.955

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44 (W1):W54-W57. doi:10.1093/nar/gkw413

EVidenceModeler

http://evidencemodeler.github.io/

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol 9 (1). doi:10.1186/Gb-2008-9-1-R7

InterProScan

http://www.ebi.ac.uk/Tools/pfa/iprscan5/

Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (9):1236-1240. doi:10.1093/bioinformatics/btu031

Pfam

http://pfam.xfam.org/

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44 (D1):D279-D285. doi:10.1093/nar/gkv1344

SignalP

http://www.cbs.dtu.dk/services/SignalP/

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8 (10):785-786. doi:10.1038/nmeth.1701

TargetP

http://www.cbs.dtu.dk/services/TargetP/

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2 (4):953-971. doi:10.1038/nprot.2007.131

Apollo

http://apollo.berkeleybop.org/

Unni DD, N.; Yao, E.; Buels, R.; Li, Y.; Holmes, I.; Elsik, C.; Lewis, S. (2017) GMOD/Apollo: Apollo2.1.0(JB#d3827c) (Version 2.1.0). Zenodo. doi:10.5281/zenodo.1295754

JBrowse

http://jbrowse.org/

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: A next-generation genome browser. Genome Res 19 (9):1630-1638. doi:10.1101/gr.094607.109

GenSAS repeat, transcript and protein database references:

RepBase

https://www.girinst.org/repbase/

Bao WD, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6. doi:10.1186/s13100-015-0041-9

NCBI RefSeq

https://www.ncbi.nlm.nih.gov/refseq/

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44 (D1):D733-745. doi:10.1093/nar/gkv1189

SwissProt/TrEMBL

https://www.uniprot.org/

The UniProt C (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45 (D1):D158-D169. doi:10.1093/nar/gkw1099