

# Integrated Based Gene Prediction Software List

Gianluca de Urzêda Alves

2024-09-25

## Integrated Based Gene Prediction Software List

### Combined Based Gene Prediction

Year	Tool Name	Publication	Method	Organism
2003	GeneComber	Shah, S. P., McVicker, G. P., Mackworth, A. K., Rogic, S., & Ouellette, B. F. F. (2003). GeneComber: Combining outputs of gene prediction programs for improved results. <i>Bioinformatics</i> (Oxford, England), 19(10), 1296–1297. <a href="https://doi.org/10.1093/bioinformatics/btg139">https://doi.org/10.1093/bioinformatics/btg139</a> .	EUI, GI and EUI frame algorithms	
2004	Combined	Allen, J. E., Pertea, M., & Salzberg, S. L. (2004). Computational Gene Prediction Using Multiple Sources of Evidence. <i>Genome Research</i> , 14(1), 142–148. <a href="https://doi.org/10.1101/gr.1562804">https://doi.org/10.1101/gr.1562804</a> .	Combined	
2004	Reganor	McHardy, A. C., Goesmann, A., Pühler, A., & Meyer, F. (2004). Development of joint application strategies for two microbial gene finders. <i>Bioinformatics</i> , 20(10), 1622–1631. <a href="https://doi.org/10.1093/bioinformatics/bth137">https://doi.org/10.1093/bioinformatics/bth137</a> .	Combined	Prokaryotes / Archaea
2005	JIGSAW	Allen, J. E., & Salzberg, S. L. (2005). JIGSAW: Integration of multiple sources of evidence for gene prediction. <i>Bioinformatics</i> , 21(18), 3596–3603. <a href="https://doi.org/10.1093/bioinformatics/bti609">https://doi.org/10.1093/bioinformatics/bti609</a> .	Generalized Hidden Markov Model	
2007	GLEAN	Elsik, C. G., Mackey, A. J., Reese, J. T., Milshina, N. V., Roos, D. S., & Weinstock, G. M. (2007). Creating a honey bee consensus gene set. <i>Genome Biology</i> , 8(1), R13. <a href="https://doi.org/10.1186/gb-2007-8-1-r13">https://doi.org/10.1186/gb-2007-8-1-r13</a> .	Hidden Markov Model	Eukaryotes
2007	Genomix	Coghlan, A., & Durbin, R. (2007). Genomix: A method for combining gene-finders' predictions, which uses evolutionary conservation of sequence and intron-exon structure. <i>Bioinformatics</i> (Oxford, England), 23(12), 1468–1475. <a href="https://doi.org/10.1093/bioinformatics/btm133">https://doi.org/10.1093/bioinformatics/btm133</a> .	Dynamic Programming	Eukaryotes

Year	Tool Name	Publication	Method	Organism
2008	SCGPred	Li, X., Ren, Q., Weng, Y., Cai, H., Zhu, Y., & Zhang, Y. (2008). SCGPred: A score-based method for gene structure prediction by combining multiple sources of evidence. <i>Genomics, Proteomics &amp; Bioinformatics</i> , 6(3–4), 175–185. <a href="https://doi.org/10.1016/S1672-0229(09)60005-X">https://doi.org/10.1016/S1672-0229(09)60005-X</a> .		Eukaryotes
2008	Evidence Modeler (EVM)	Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. <i>Genome Biology</i> , 9(1), R7. <a href="https://doi.org/10.1186/gb-2008-9-1-r7">https://doi.org/10.1186/gb-2008-9-1-r7</a> .		
2008	Evigan	Liu, Q., Mackey, A. J., Roos, D. S., & Pereira, F. C. N. (2008). Evigan: A hidden variable model for integrating gene evidence for eukaryotic gene prediction. <i>Bioinformatics (Oxford, England)</i> , 24(5), 597–605. <a href="https://doi.org/10.1093/bioinformatics/btn004">https://doi.org/10.1093/bioinformatics/btn004</a> .	Dynamic Bayes Networks	Eukaryotes
2008	Maker	Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. <i>Genome Research</i> , 18(1), 188–196. <a href="https://doi.org/10.1101/gr.6743907">https://doi.org/10.1101/gr.6743907</a> .		
2012	eCRAIG (ensemble CRAIG)	Bernal, A., Crammer, K., & Pereira, F. (2012). Automated gene-model curation using global discriminative learning. <i>Bioinformatics (Oxford, England)</i> , 28(12), 1571–1578. <a href="https://doi.org/10.1093/bioinformatics/bts176">https://doi.org/10.1093/bioinformatics/bts176</a> .	Hidden Markov Model Conditional Random Field	
2015	Ipred	Zickmann, F., & Renard, B. Y. (2015). IPred—Integrating ab initio and evidence based gene predictions to improve prediction accuracy. <i>BMC Genomics</i> , 16(1), 134. <a href="https://doi.org/10.1186/s12864-015-1315-9">https://doi.org/10.1186/s12864-015-1315-9</a> .	Combined	

## Pipeline Based Gene Prediction

Year	Tool Name	Publication	Method	Organism
2003	PASA (Program to Assemble Spliced Alignments)	Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. <i>Nucleic Acids Research</i> , 31(19), 5654–5666. <a href="https://doi.org/10.1093/nar/gkg770">https://doi.org/10.1093/nar/gkg770</a>	Pipeline	

Year	Tool Name	Publication	Method	Organism
2006	MaGe	Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., Lajus, A., Pascal, G., Scarpelli, C., & Médigue, C. (2006). MaGe: A microbial genome annotation system supported by synteny results. <i>Nucleic Acids Research</i> , 34(1), 53–65. <a href="https://doi.org/10.1093/nar/gkj406">https://doi.org/10.1093/nar/gkj406</a> .	Pipeline	Bacteria
2006	FGENESH	Selovyev, V., Kosarev, P., Seledsov, I., & Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. <i>Genome Biology</i> , 7(1), S10. <a href="https://doi.org/10.1186/gb-2006-7-s1-s10">https://doi.org/10.1186/gb-2006-7-s1-s10</a> .	Pipeline	
2008	RAST	Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., . . . Zagnitko, O. (2008). The RAST Server: Rapid annotations using subsystems technology. <i>BMC Genomics</i> , 9, 75. <a href="https://doi.org/10.1186/1471-2164-9-75">https://doi.org/10.1186/1471-2164-9-75</a> .	Pipeline	Bacteria / Archaea
2011	MAKER2	Holt, C., & Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. <i>BMC Bioinformatics</i> , 12, 491. <a href="https://doi.org/10.1186/1471-2105-12-491">https://doi.org/10.1186/1471-2105-12-491</a> .	Pipeline	Eukaryotes / Prokaryotes
2011	VMGAP (TheViral MetaGenome Annotation Pipeline)	Lorenzi, H. A., Hoover, J., Inman, J., Safford, T., Murphy, S., Kagan, L., & Williamson, S. J. (2011). TheViral MetaGenome Annotation Pipeline(VMGAP):an automated tool for the functional annotation of viral Metagenomic shotgun sequencing data. <i>Standards in Genomic Sciences</i> , 4(3), 418–429. <a href="https://doi.org/10.4056/sigs.1694706">https://doi.org/10.4056/sigs.1694706</a> .	Pipeline	Viruses
2012	MOCAT	Kultima, J. R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D. R., Arumugam, M., Pan, Q., Liu, B., Qin, J., Wang, J., & Bork, P. (2012). MOCAT: A metagenomics assembly and gene prediction toolkit. <i>PloS One</i> , 7(10), e47656. <a href="https://doi.org/10.1371/journal.pone.0047656">https://doi.org/10.1371/journal.pone.0047656</a> .	Pipeline	Metagenomes
2014	OMIGA (Optimized Maker-Based Insect Genome Annotation)	Liu, J., Xiao, H., Huang, S., & Li, F. (2014). OMIGA: Optimized Maker-Based Insect Genome Annotation. <i>Molecular Genetics and Genomics: MGG</i> , 289(4), 567–573. <a href="https://doi.org/10.1007/s00438-014-0831-7">https://doi.org/10.1007/s00438-014-0831-7</a> .	Pipeline	Insects
2014	Prokka	Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. <i>Bioinformatics (Oxford, England)</i> , 30(14), 2068–2069. <a href="https://doi.org/10.1093/bioinformatics/btu153">https://doi.org/10.1093/bioinformatics/btu153</a> .	Pipeline	Prokaryotes

Year	Tool Name	Publication	Method	Organism
2014	DFAST (DDBJ Fast Annotation and Submission Tool)	Tanizawa, Y., Fujisawa, T., & Nakamura, Y. (2018). DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. <i>Bioinformatics</i> (Oxford, England), 34(6), 1037–1039. <a href="https://doi.org/10.1093/bioinformatics/btx713">https://doi.org/10.1093/bioinformatics/btx713</a> .	Pipeline	Prokaryotes
2016	PGAP	Tatusova T. et al. (2016) NCBI prokaryotic genome annotation pipeline. <i>Nucleic Acids Res.</i> , 44, 6614–6624.	Pipeline	Prokaryotes
2016	CAT Comparative Analysis toolkit	Fiddes, I. T., Armstrong, J., Diekhans, M., Nachtweide, S., Kronenberg, Z. N., Underwood, J. G., Gordon, D., Earl, D., Keane, T., Eichler, E. E., Haussler, D., Stanke, M., & Paten, B. (2018). Comparative Annotation Toolkit (CAT)—Simultaneous clade and personal genome annotation. <i>Genome Research</i> , 28(7), 1029–1038. <a href="https://doi.org/10.1101/gr.233460.117">https://doi.org/10.1101/gr.233460.117</a> .	Pipeline	
2017	funannotate	Jon Love, Jon Palmer, Jason Stajich, Tyler Esser, Erik Kastman, Daniel Bogema, & David Winter. (2019). nextgenusfs/funannotate: funannotate v1.7.0 (1.7.0). Zenodo. <a href="https://doi.org/10.5281/zenodo.3534297">https://doi.org/10.5281/zenodo.3534297</a> .	Pipeline	Fungi
2018	FunGAP	Min, B., Grigoriev, I. V., & Choi, I.-G. (2017). FunGAP: Fungal Genome Annotation Pipeline using evidence-based gene model evaluation. <i>Bioinformatics</i> (Oxford, England), 33(18), 2936–2937. <a href="https://doi.org/10.1093/bioinformatics/btx353">https://doi.org/10.1093/bioinformatics/btx353</a> .	Pipeline	Fungi
2019	VAPiD (Viral Annotation Pipeline and iDentification)	Shean, R. C., Makhsous, N., Stoddard, G. D., Lin, M. J., & Greninger, A. L. (2019). VAPiD: A lightweight cross-platform viral annotation pipeline and identification tool to facilitate virus genome submissions to NCBI GenBank. <i>BMC Bioinformatics</i> , 20(1), 48. <a href="https://doi.org/10.1186/s12859-019-2606-y">https://doi.org/10.1186/s12859-019-2606-y</a> .	Pipeline	Viruses
2019	GAAP	Kong, J., Huh, S., Won, J.-I., Yoon, J., Kim, B., & Kim, K. (2019). GAAP: A Genome Assembly + Annotation Pipeline. <i>BioMed Research International</i> , 2019, 4767354. <a href="https://doi.org/10.1155/2019/4767354">https://doi.org/10.1155/2019/4767354</a> .	Pipeline	
2021	Bakta	Schwengers, O., Jelonek, L., Dieckmann, M. A., Beyvers, S., Blom, J., & Goesmann, A. (2021). Bakta: Rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. <i>Microbial Genomics</i> , 7(11), 000685. <a href="https://doi.org/10.1099/mgen.0.000685">https://doi.org/10.1099/mgen.0.000685</a> .	Pipeline	Bacteria / Plasmids
2021	FINDER	Banerjee, S., Bhandary, P., Woodhouse, M., Sen, T. Z., Wise, R. P., & Andorf, C. M. (2021). FINDER: An automated software package to annotate eukaryotic genes from RNA-Seq data and associated protein sequences. <i>BMC Bioinformatics</i> , 22(1), 205. <a href="https://doi.org/10.1186/s12859-021-04120-9">https://doi.org/10.1186/s12859-021-04120-9</a> .	Pipeline	Eukaryotes

Year	Tool Name	Publication	Method	Organism
2022	Pharokka	Bouras, G., Nepal, R., Houtak, G., Psaltis, A. J., Wormald, P.-J., & Vreugde, S. (2023). Pharokka: A fast scalable bacteriophage annotation tool. <i>Bioinformatics</i> (Oxford, England), 39(1), btac776. <a href="https://doi.org/10.1093/bioinformatics/btac776">https://doi.org/10.1093/bioinformatics/btac776</a> .	Pipeline	Bacteriophages
2023	FLAG	Troy, W., Damas, J., Titus, A. J., & Cantarel, B. L. (2023). FLAG: Find, Label, Annotate Genomes, a fully automated tool for genome gene structural and functional annotation of highly fragmented non-model species (p. 2023.07.14.548907). <i>bioRxiv</i> . <a href="https://doi.org/10.1101/2023.07.14.548907">https://doi.org/10.1101/2023.07.14.548907</a> .	Pipeline	Eukaryotes
2023	Galba	Brüna, T., Li, H., Guhlin, J., Honsel, D., Herbold, S., Stanke, M., Nenasheva, N., Ebel, M., Gabriel, L., & Hoff, K. J. (2023). Galba: Genome annotation with miniprot and AUGUSTUS. <i>BMC Bioinformatics</i> , 24(1), 327. <a href="https://doi.org/10.1186/s12859-023-05449-z">https://doi.org/10.1186/s12859-023-05449-z</a> .	Pipeline	Eukaryotes