

# Ab Initio Gene Prediction Software List

Gianluca de Urzêda Alves

2024-09-25

## Ab initio Gene Prediction Softwares List

---

### Eukaryotes

Year	Tool Name	Publication	Method	Organism
1994	GeneLang	Dong, S., & Searls, D. B. (1994). Gene structure prediction by linguistic methods. <i>Genomics</i> , 23(3), 540–551. <a href="https://doi.org/10.1006/geno.1994.1541">https://doi.org/10.1006/geno.1994.1541</a> .	Linguistic method, Hidden Markov Model, Dynamic Programming, Weight Array Matrix	Eukaryotes
1995	Fgeneh / GeneFinder	Solovyev, V. V., Salamov, A. A., & Lawrence, C. B. (1995). Identification of human gene structure using linear discriminant functions and dynamic programming. <i>Proceedings. International Conference on Intelligent Systems for Molecular Biology</i> , 3, 367–375.	Hidden Markov Model, Dynamic Programming, Linear Discriminant Analysis	<i>Homo sapiens</i>
1997	HMMGene	Krogh, A. (1997). Two methods for improving performance of an HMM and their application for gene finding. <i>Proceedings. International Conference on Intelligent Systems for Molecular Biology</i> , 5, 179–186.	Hidden Markov Model Conditional Random Field	Vertebrates and <i>C. elegans</i>

Year	Tool Name	Publication	Method	Organism
1997	GenScan	Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. <i>Journal of Molecular Biology</i> , 268(1), 78–94. <a href="https://doi.org/10.1006/jmbi.1997.0951">https://doi.org/10.1006/jmbi.1997.0951</a> .	Generalized Hidden Markov Model	
2003	AUGUSTUS	Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. <i>Bioinformatics (Oxford, England)</i> , 19 Suppl 2, ii215–225. <a href="https://doi.org/10.1093/bioinformatics/btg1080">https://doi.org/10.1093/bioinformatics/btg1080</a>	Hidden Markov Model	Eukaryotes
2004	GlimmerHMM	Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. <i>Bioinformatics (Oxford, England)</i> , 20(16), 2878–2879. <a href="https://doi.org/10.1093/bioinformatics/bth315">https://doi.org/10.1093/bioinformatics/bth315</a>	Generalized Hidden Markov Model	Eukaryotes
2005	GeneMark-ES	Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. <i>Nucleic Acids Research</i> , 33(20), 6494–6506. <a href="https://doi.org/10.1093/nar/gki937">https://doi.org/10.1093/nar/gki937</a> .		Eukaryotes
2005	BGF (Beijing Gene Finder)	Li, H., Liu, J.-S., Xu, Z., Jin, J., Fang, L., Gao, L., Li, Y.-D., Xing, Z.-X., Gao, S.-G., Liu, T., Li, H.-H., Li, Y., Fang, L.-J., Xie, H.-M., Zheng, W.-M., & Hao, B.-L. (2005). Test Data Sets and Evaluation of Gene Prediction Programs on the Rice Genome. <i>Journal of Computer Science and Technology</i> , 20(4), 446–453. <a href="https://doi.org/10.1007/s11390-005-0446-x">https://doi.org/10.1007/s11390-005-0446-x</a> .	Semi Hidden Markov Model	Eukaryotes

Year	Tool Name	Publication	Method	Organism
2010	Gnomon	Souvorov, A., Kapustin, Y., Kiryutin, B., Chetvernin, V., Tatusova, T., & Lipman, D. (2010). Gnomon – NCBI eukaryotic gene prediction tool. Available in: <a href="https://www.ncbi.nlm.nih.gov/core/assets/genome/files/Gnomon-description.pdf">https://www.ncbi.nlm.nih.gov/core/assets/genome/files/Gnomon-description.pdf</a>	Hidden Markov Model, Weight Array Matrix	
2014	GeneMark-ET	Lomsadze, A., Burns, P. D., & Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. <i>Nucleic Acids Research</i> , 42(15), e119. <a href="https://doi.org/10.1093/nar/gku557">https://doi.org/10.1093/nar/gku557</a> .	Hidden Markov Model	Eukaryotes

## Prokaryotes

Year	Tool Name	Publication	Method	Organism
1998	GeneMark.hmm	Lukashin, A. V., & Borodovsky, M. (1998). GeneMark.hmm: New solutions for gene finding. <i>Nucleic Acids Research</i> , 26(4), 1107–1115. <a href="https://doi.org/10.1093/nar/26.4.1107">https://doi.org/10.1093/nar/26.4.1107</a> .	Hidden Markov Model	Prokaryotes / Archaea
2001	GeneHacker Plus	Yada, T., Totoki, Y., Takagi, T., & Nakai, K. (2001). A novel bacterial gene-finding system with improved accuracy in locating start codons. <i>DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes</i> , 8(3), 97–106. <a href="https://doi.org/10.1093/dnares/8.3.97">https://doi.org/10.1093/dnares/8.3.97</a> .	Hidden Markov Model	Prokaryotes

Year	Tool Name	Publication	Method	Organism
2001	GeneMarkS	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. <i>Nucleic Acids Research</i> , 29(12), 2607–2618. <a href="https://doi.org/10.1093/nar/29.12.2607">https://doi.org/10.1093/nar/29.12.2607</a> .	Hidden Markov Model	Prokaryotes
2003	AMIGene	Bocs, S., Cruveiller, S., Vallenet, D., Nuel, G., & Médigue, C. (2003). AMIGene: Annotation of Microbial Genes. <i>Nucleic Acids Research</i> , 31(13), 3723–3726.	Hidden Markov Model	Prokaryotes
2003	EasyGene	Larsen, T. S., & Krogh, A. (2003). EasyGene – a prokaryotic gene finder that ranks ORFs by statistical significance. <i>BMC Bioinformatics</i> , 4(1), 21. <a href="https://doi.org/10.1186/1471-2105-4-21">https://doi.org/10.1186/1471-2105-4-21</a> .	Hidden Markov Model	Prokaryotes / Archaea
2003	ZCurve	Guo, F.-B., Ou, H.-Y., & Zhang, C.-T. (2003). ZCURVE: A new system for recognizing protein-coding genes in bacterial and archaeal genomes. <i>Nucleic Acids Research</i> , 31(6), 1780–1789. <a href="https://doi.org/10.1093/nar/gkg254">https://doi.org/10.1093/nar/gkg254</a> .	Z-curve	Prokaryotes / Archaea
2006	MetaGeneAnnotator (MGA)	Noguchi, H., Taniguchi, T., & Itoh, T. (2008). MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. <i>DNA Research</i> , 15(6), 387–396. <a href="https://doi.org/10.1093/dnares/dsn027">https://doi.org/10.1093/dnares/dsn027</a> .		Prokaryotes

Year	Tool Name	Publication	Method	Organism
2007	GISMO	Krause, L., McHardy, A. C., Nattkemper, T. W., Pühler, A., Stoye, J., & Meyer, F. (2007). GISMO--gene identification using a support vector machine for ORF classification. <i>Nucleic Acids Research</i> , 35(2), 540–549. <a href="https://doi.org/10.1093/nar/gkl1083">https://doi.org/10.1093/nar/gkl1083</a> .	Support Vector Machine	Prokaryotes
2010	Prodigal (PROkaryotic DYnamic programming Gene-finding ALgorithm)	Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. <i>BMC Bioinformatics</i> , 11(1), 119. <a href="https://doi.org/10.1186/1471-2105-11-119">https://doi.org/10.1186/1471-2105-11-119</a> .	Dynamic Programming + Hidden Markov Model	Prokaryotes
2014	ZUPLS	Song, K., Tong, T., & Wu, F. (2014). Predicting essential genes in prokaryotic genomes using a linear method: ZUPLS. <i>Integrative Biology: Quantitative Biosciences from Nano to Macro</i> , 6(4), 460–469. <a href="https://doi.org/10.1039/c3ib40241j">https://doi.org/10.1039/c3ib40241j</a> .	Z-curve	Prokaryotes

## Virus

Year	Tool Name	Publication	Method	Organism
2003	GeneMarkS (virus version)	Mills, R., Rozanov, M., Lomsadze, A., Tatusova, T., & Borodovsky, M. (2003). Improving gene annotation of complete viral genomes. <i>Nucleic Acids Research</i> , 31(23), 7041–7055. <a href="https://doi.org/10.1093/nar/gkg878">https://doi.org/10.1093/nar/gkg878</a> .	Hidden Markov Model	Viruses

Year	Tool Name	Publication	Method	Organism
2006	ZCURVE_V	Guo, F.-B., & Zhang, C.-T. (2006). ZCURVE_V: A new self-training system for recognizing protein-coding genes in viral and phage genomes. <i>BMC Bioinformatics</i> , 7(1), 9. <a href="https://doi.org/10.1186/1471-2105-7-9">https://doi.org/10.1186/1471-2105-7-9</a> .	Z-curve	Viruses
2007	GLIMMER3	Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. <i>Bioinformatics</i> , 23(6), 673–679. <a href="https://doi.org/10.1093/bioinformatics/btm009">https://doi.org/10.1093/bioinformatics/btm009</a> .	Interpolated Markov Model	Bacteria, Archaea, and Viruses
2019	Vgas (Viral Genome Annotation System)	Zhang, K.-Y., Gao, Y.-Z., Du, M.-Z., Liu, S., Dong, C., & Guo, F.-B. (2019). Vgas: A Viral Genome Annotation System. <i>Frontiers in Microbiology</i> , 10. <a href="https://doi.org/10.3389/fmicb.2019.00184">https://doi.org/10.3389/fmicb.2019.00184</a> .	ZCURVE_V + BLASTp	Viruses

## Metagenome

Year	Tool Name	Publication	Method	Organism
2006	Metagene	Noguchi, H., Park, J., & Takagi, T. (2006). MetaGene: Prokaryotic gene finding from environmental genome shotgun sequences. <i>Nucleic Acids Research</i> , 34(19), 5623–5630. <a href="https://doi.org/10.1093/nar/gkl723">https://doi.org/10.1093/nar/gkl723</a> .		Metagenomes
2010	FragGeneScan	Rho, M., Tang, H., & Ye, Y. (2010). FragGeneScan: Predicting genes in short and error-prone reads. <i>Nucleic Acids Research</i> , 38(20), e191. <a href="https://doi.org/10.1093/nar/gkq747">https://doi.org/10.1093/nar/gkq747</a> .	Hidden Markov Model	Metagenomes

Year	Tool Name	Publication	Method	Organism
2010	MetaGeneMark	Zhu, W., Lomsadze, A., & Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. <i>Nucleic Acids Research</i> , 38(12), e132. <a href="https://doi.org/10.1093/nar/gkq275">https://doi.org/10.1093/nar/gkq275</a> .	Hidden Markov Model	Metagenomes
2013	MetaGUN	Liu, Y., Guo, J., Hu, G., & Zhu, H. (2013). Gene prediction in metagenomic fragments based on the SVM algorithm. <i>BMC Bioinformatics</i> , 14(5), S12. <a href="https://doi.org/10.1186/1471-2105-14-S5-S12">https://doi.org/10.1186/1471-2105-14-S5-S12</a> .	Support Vector Machine	Metagenomes