BHARATHIDASAN UNIVERSITY
Tiruchirappalli- 620024,
Tamil Nadu, India.

Programme : M.Sc., Biomedical Science

Course Title : Bioinformatics
Course Code : BM35S1BI

Unit - IV

TOPIC: GENE FINDERS-GLIMMER

Dr. P. JEGANATHAN
Guest Lecturer
Department of Biomedical Science

GENE FINDERS-GLIMMER

GENE FINDERS

Introduction:

- Gene finding is the process of identifying genome sequence regions representing stretches of DNA that encode biologically active products, such as proteins or functional non coding RNAs.

- Gene finding is one of the first and most important steps in understanding the genome of a species once it has been sequenced.

Importance of bioinformatics:

- With an ever increasing quantity of raw genome sequences.gene finding is an important avenue towards understanding the genetic information of (novel) organisms, as well as learning shared patterns across evolutionarily diverse species.

- Gene finding is the task of identifying the location of coding sequences within the vast amount of genetic contained in the genome.

GLIMMER

- Glimmer is a system for finding genes in microbial DNA, especially the genomes of bacteria,archaea and viruses.

- Glimmer (Gene Locator and Interpolated Markov ModelER) uses interpolated markov models to identify the coding regions and distinguish them from non coding DNA.

GLIMMER

- Glimmer was the primary microbial gene finder used at The Institute for Genomic Research(TIGR) where it was first developed and has been used to annotate the genomes of thousands of bacterial, archaeal and viral genomes around the world.

- Glimmer is used to find genes in prokaryotic DNA.

- It typically finding 98-99% of all relatively long protein coding genes.

How it works?

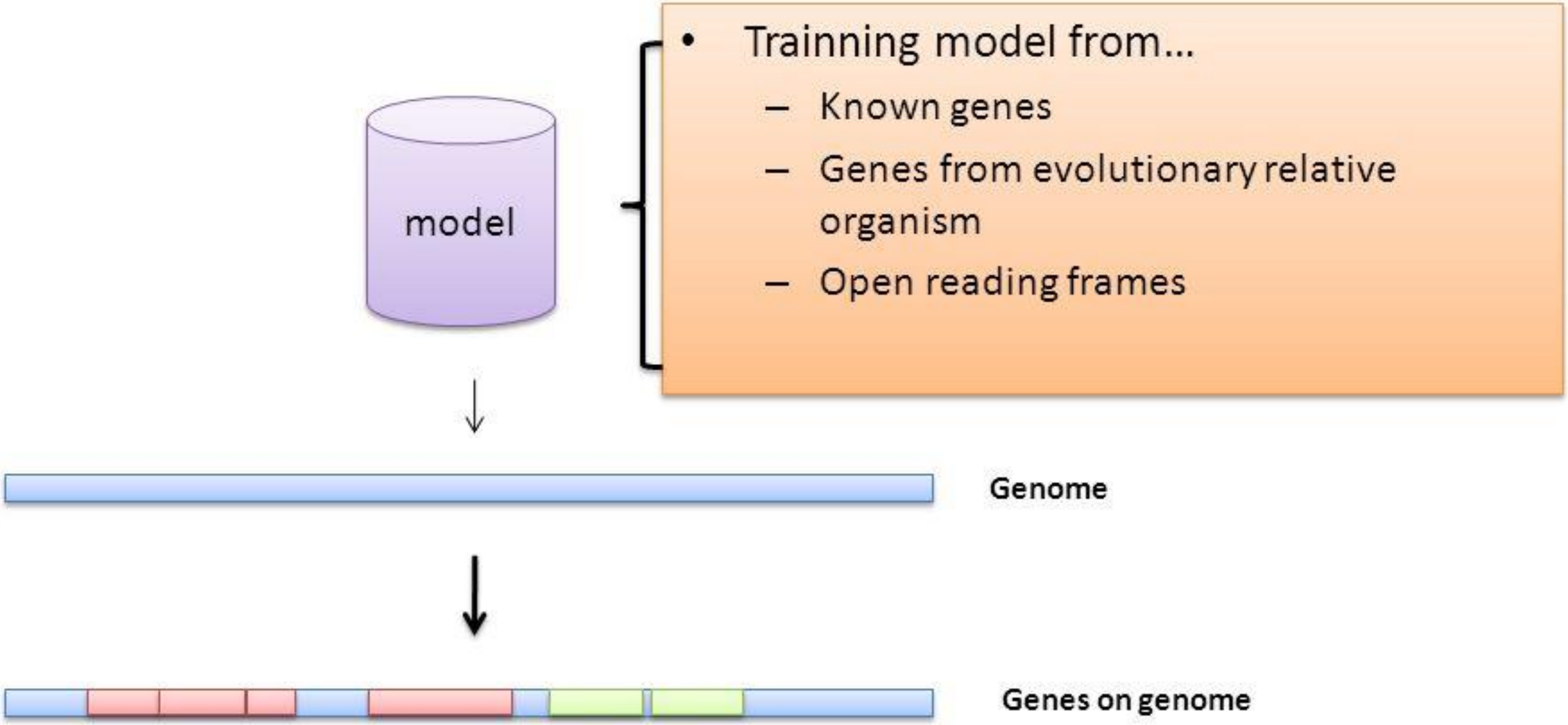
- Glimmer primarily searches for long ORFs. An open reading frame which will be resolved using the technique described in the sub section. Using these long ORFs and following certain amino acid distribution GLIMMER generates training set data.

- Using these training data, GLIMMER trains all the six markov models of coding DNA from zero to eight order and also train the model for non coding DNA.

How it works?

- GLIMMER tries to calculate the probabilities from the data. based on the number of observations, GLIMMER determines whether to use fixed order Markov model or interpolated Markov model.
- GLIMMER obtains score for every long ORF generated using all the six coding DNA models and also using the non coding DNA model.
- If the score is obtained in the previous step is greater than the certain threshold then GLIMMER predicts to be a gene.

Concept of glimmer



model

- Trainning model from…
  - Known genes
  - Genes from evolutionary relative organism
  - Open reading frames

Genome

Genes on genome

Glimmer 2.0

- Output: for each ORF,GLIMMER assigns a likelihood score or probability that the ORF resembles a known gene.

- High scoring ORFs that overlap significantly with other high scoring ORFs are reported but highlighted.

- GLIMMER 2.0 is reported to be 98% accurate on prokaryotic genomes.

GLIMMER 2.0

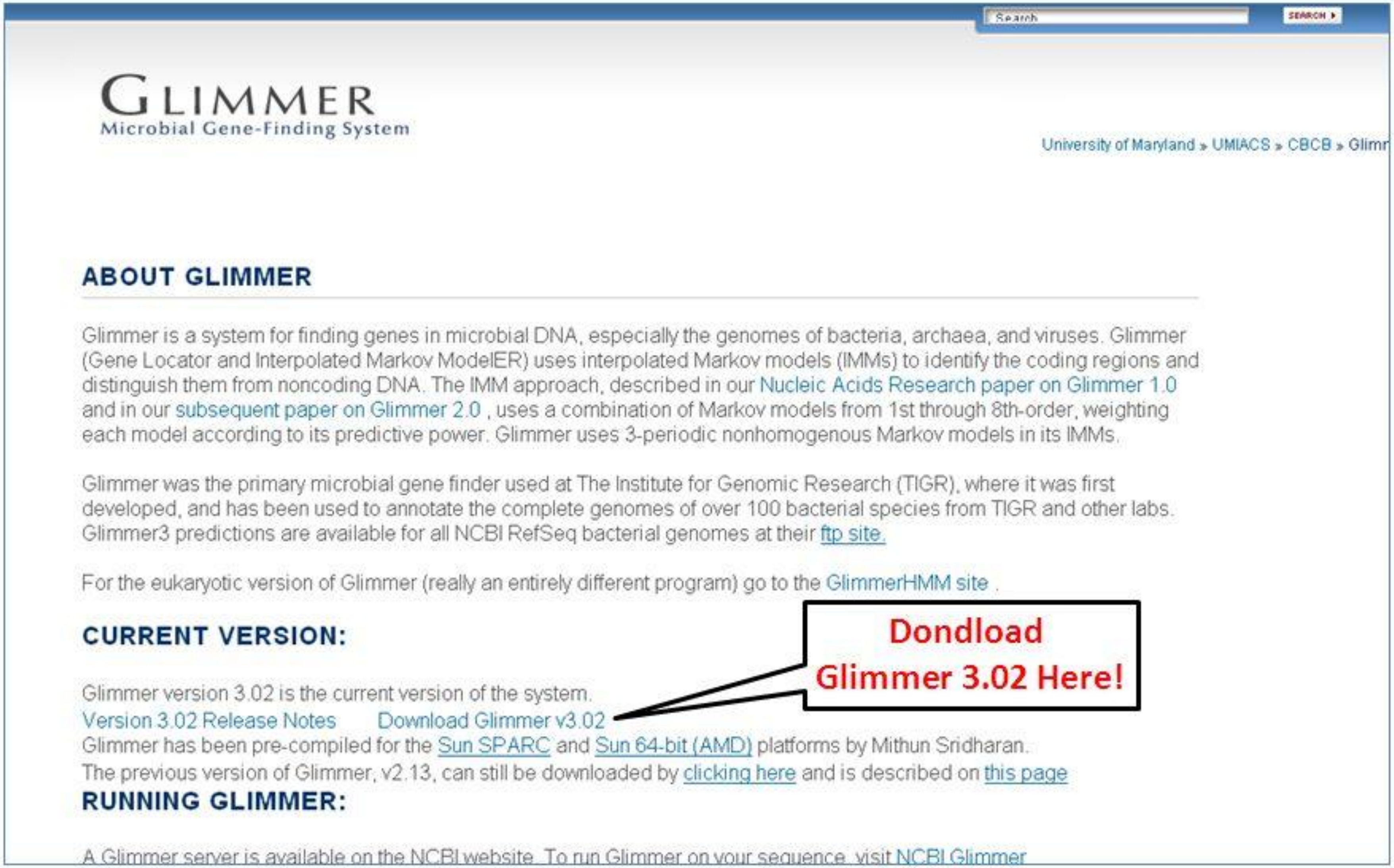
ADVANTAGES

- Fewer false positives because ORFs are evaluated for likelihood for coding.
- Organism specific because model is built on known genes.
- User can modify many parameters during search phase.

DISADVANTAGES

- Requires approximately 500+ known genes for proper training.
- Genuine coding regions with unusual codon composition will be eliminated.
- Reported accuracy difficult to reproduce.

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



GLIMMER
Microbial Gene-Finding System

University of Maryland » UMIACS » CBCB » Glimn

ABOUT GLIMMER

Glimmer is a system for finding genes in microbial DNA, especially the genomes of bacteria, archaea, and viruses. Glimmer (Gene Locator and Interpolated Markov ModelER) uses interpolated Markov models (IMMs) to identify the coding regions and distinguish them from noncoding DNA. The IMM approach, described in our Nucleic Acids Research paper on Glimmer 1.0 and in our subsequent paper on Glimmer 2.0 , uses a combination of Markov models from 1st through 8th-order, weighting each model according to its predictive power. Glimmer uses 3-periodic nonhomogenous Markov models in its IMMs.

Glimmer was the primary microbial gene finder used at The Institute for Genomic Research (TIGR), where it was first developed, and has been used to annotate the complete genomes of over 100 bacterial species from TIGR and other labs. Glimmer3 predictions are available for all NCBI RefSeq bacterial genomes at their ftp site.

For the eukaryotic version of Glimmer (really an entirely different program) go to the GlimmerHMM site .

CURRENT VERSION:

Glimmer version 3.02 is the current version of the system.
Version 3.02 Release Notes Download Glimmer v3.02
Glimmer has been pre-compiled for the Sun SPARC and Sun 64-bit (AMD) platforms by Mithun Sridharan.
The previous version of Glimmer, v2.13, can still be downloaded by clicking here and is described on this page

RUNNING GLIMMER:

A Glimmer server is available on the NCBI website. To run Glimmer on your sequence, visit NCBI Glimmer

Dondload
Glimmer 3.02 Here!

REFERENCES :-

- HTTPS://THEBIOLOGYNOTES.COM/GENE-PREDICTION-IMPORTANCE-AND-METHODS/
- HTTPS://BIOTECHARTICLES.COM/BIOINFORMATICS-ARTICLE/GENSCAN-BIOINFORMATICS-SOFTWARE-FOR-STRUCTURE-PREDICTION-AND-ANALYSIS-OF-GENE-284.HTML
- HTTPS://GENOMEBIOLOGY.BIOMEDCENTRAL.COM/ARTICLES/10.1186/GB-2000-1-1-REPORTS222

THANK YOU