

**Pre-PhD Course work
(Zoology)
Maa Shakumbhari University
Saharanpur, Uttar Pradesh**



**Topic :Molecular Biology
Lecture 10: 2.05.2026**

Dr. Piyush Sharma Ph.D. (AIIMS, New Delhi)

Assistant Professor

Department of Zoology

Sanatan Dharm College, Muzaffarnagar

Evidence that a particular ORF actually encodes a protein.

- The ORF in question encodes a protein that is similar to previously described ones (search the protein database for homologs of the given sequence).
- The ORF has a typical GC content, codon frequency, or oligonucleotide composition.
- The ORF is preceded by a typical ribosome-binding site (search for a Shine-Dalgarno sequence in front of the predicted coding sequence).
- The ORF is preceded by a typical promoter

Prokaryotic gene prediction

- Frequency of G and C **FramePlot**, available at the Japanese Institute of Infectious Diseases (<http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl>) and at the **TIGR web site** (http://tigrblast.tigr.org/cmr-blast/GC_Skew.cgi).
- **GeneMark** and **Glimmer** build Markov models of the known coding regions for the given organism and then employ them to **estimate the coding potential of uncharacterized ORFs**.

EasyGene 1.2 - Prokaryotic gene prediction

EasyGene 1.2 produces a list of predicted genes for a chosen prokaryotic genome. Each prediction is attributed with a significance score (R-value) indicating how likely rather than a real gene. The user needs only to specify the output format for the particular genome. The R-value cutoff has a default value which may be altered if needed. The various output formats and features, which can be displayed, are explained in their corresponding links. ✓

For publication of results, please cite:

[EasyGene - a prokaryotic gene finder that ranks ORFs by statistical significance.](#)

Thomas Schou Larsen and Anders Krogh.

BMC Bioinformatics 2003, 4:21

[Large-scale prokaryotic gene prediction and comparison to genome annotation.](#)

P. Nielsen and A. Krogh.

Bioinformatics 2005, 21:4322-4329

[Statistics for the genomes](#)

[EasyGene 1.0 server](#)

Genome

Aeropyrum pernix
Agrobacterium tumefaciens str. C58 (Cereon) circul...
Agrobacterium tumefaciens str. C58 (Cereon) linear...
Agrobacterium tumefaciens str. C58 (Cereon) plasmid...
Agrobacterium tumefaciens str. C58 (Cereon) plasmid...
Agrobacterium tumefaciens str. C58 (U. Washington)...
Agrobacterium tumefaciens str. C58 (U. Washington)...

Format:

- ☒ [Narrow](#)
[HTML](#)
☐ [HTML](#)
☐ [GFF](#)
☐ [CDS](#)

Limitations:

[R cutoff:](#)

2 ▾

[Range:](#) (optional)

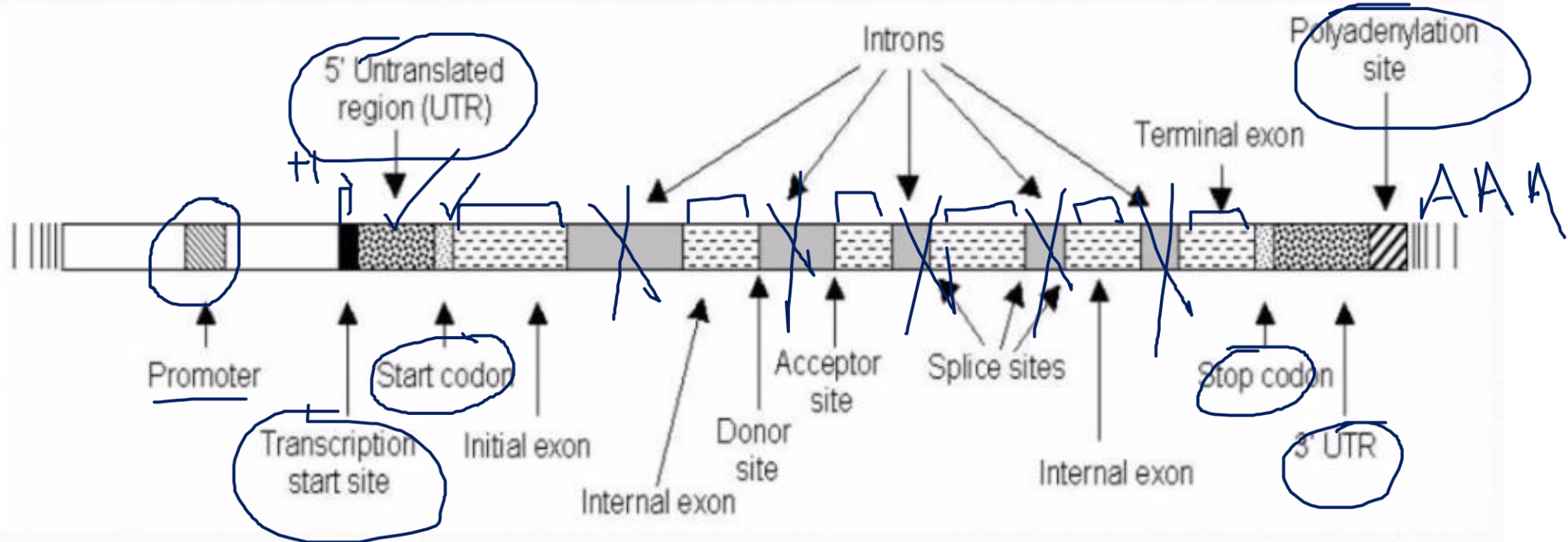
Unicellular eukaryotes

- Genomes of unicellular eukaryotes are extremely diverse in size, the proportion of the genome that is occupied by protein-encoding genes and the frequency of introns.
- Smaller the intergenic regions and the fewer introns are there, the easier it is to identify genes.
- yeast *S. cerevisiae*, at least 67% of the genome is protein-coding, and only 233 genes (less than 4% of the total) appear to have introns

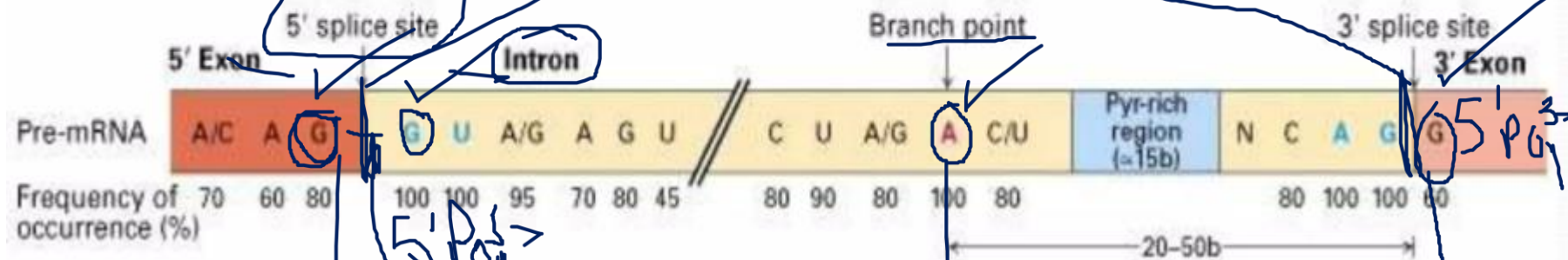
Multicellular eukaryotes

- Coding regions compose only a minor portion of the gene.
- Gene prediction should identify all exons and introns, including those in the 5'-untranslated region (5'-UTR) and the 3'-UTR of the mRNA, in order to precisely reconstruct the predominant mRNA species.
- Correct identification of the exon boundaries relies on the recognition of the splice sites.

EUKARYOTES GENE STRUCTURE



SPLICE SITES



3' OH



↓

Algorithms and software tools for gene identification

- Some of tools perform gene prediction *ab initio*, relying only on the statistical parameters in the DNA sequence for gene identification.
- homology-based methods rely primarily on identifying homologous sequences in other genomes and/or in public databases using BLAST or Smith-Waterman algorithms.
- Many of the commonly used methods combine these two approaches.

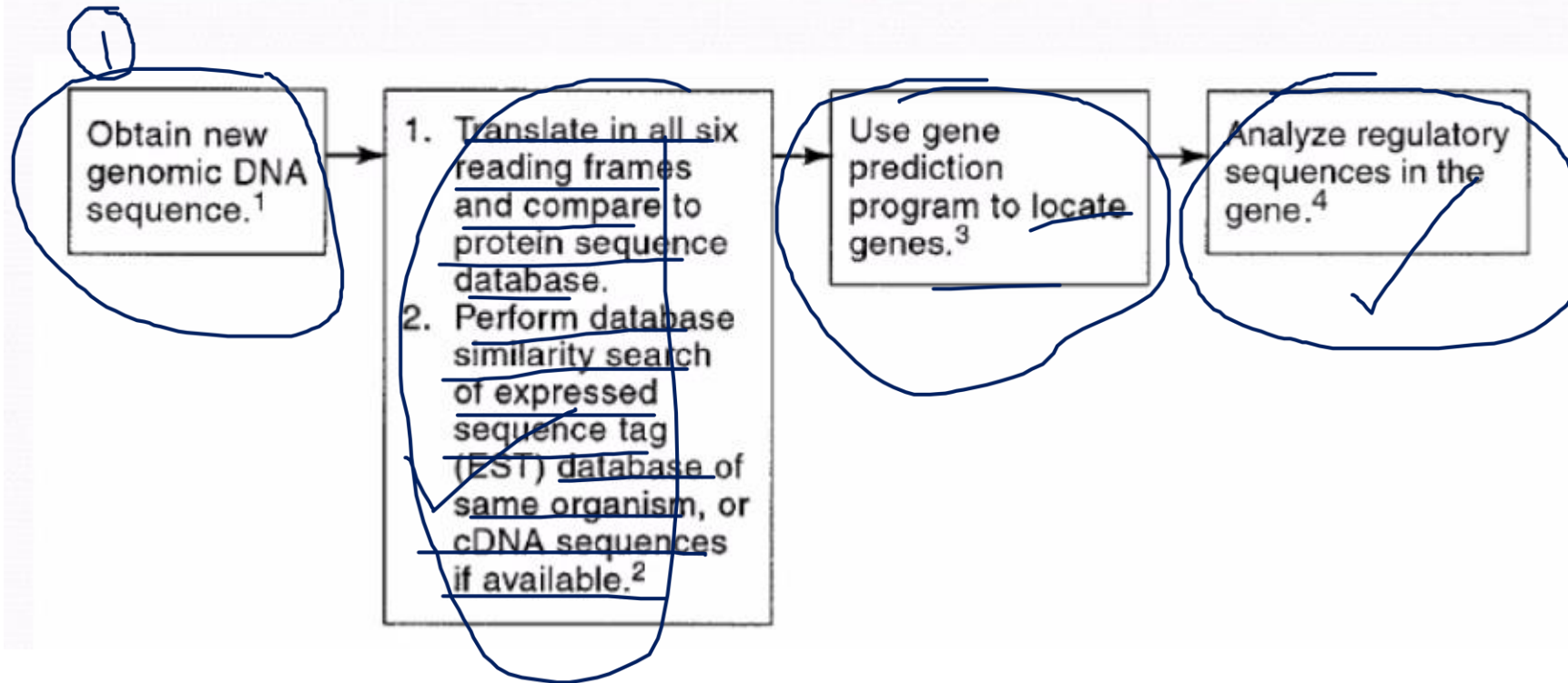
Software tools for *ab initio* gene prediction

Program(s)	Author(s), WWW site	Program versions available for:
GeneMark, GenMark.hmm	Mark Borodovsky, http://opal.biology.gatech.edu/GeneMark	Human, mouse, rat, chicken, <i>C. elegans</i> , <i>Drosophila</i> , rice, <i>Arabidopsis</i> , yeast, many bacteria and archaea
Glimmer, GlimmerM	Steven Salzberg, http://www.tigr.org/softlab	Many bacteria and archaea, <i>Plasmodium</i> , <i>Aspergillus</i> , rice, <i>Arabidopsis</i>
Grail, GrailEXP	Edward Uberbacher, http://compbio.ornl.gov/	Human, mouse, <i>Drosophila</i> , <i>Arabidopsis</i> , <i>E. coli</i>
GenScan	Christopher Burge, http://genes.mit.edu/GENSCAN.html	Human, <i>Arabidopsis</i> , maize
GeneBuilder	Igor Rogozin, Luciano Milanesi http://www.itba.mi.cnr.it/webgene/	Human, mouse, rat, fugu, <i>Drosophila</i> , <i>C. elegans</i> , <i>Arabidopsis</i> , <i>Aspergillus</i>
Genie	David Kulp, David Haussler, http://www.cse.ucsc.edu/~dkulp/cgi-bin/genie	Human
GeneID	Roderick Guigo, http://www1.imim.es/software/geneid	Human, <i>Drosophila</i>
GeneFinder, Fgenes, Fgenesh	Victor Solovyev, http://genomic.sanger.ac.uk/gf/gf.shtml	Human, <i>Drosophila</i> , <i>C. elegans</i> , <i>Arabidopsis</i> , yeast
HMMgene	Anders Krogh, Anders Pedersen, Søren Brunak, http://www.cbs.dtu.dk/services	Human (vertebrates), <i>C. elegans</i> , <i>Arabidopsis</i>
GeneFinder, MZEF	Michael Zhang, http://www.cshl.org/genefinder	Human, mouse, <i>Arabidopsis</i> , <i>S. pombe</i>
GeneParser	Erik Snyder http://mcdb.colorado.edu/~eesnyder/GeneParser.html	Available only for download

Software tools for prediction of splicing sites

Program	Author, WWW site	Training sets
HSPL	Victor Solovyev, http://genomic.sanger.ac.uk/gf/gf.shtml	Human
NNSplice	Martin Reese, http://www.fruitfly.org/seq_tools/splice.html	Human, <i>Drosophila</i>
SpliceView	Igor Rogozin, Luciano Milanesi http://www.itba.mi.cnr.it/webgene/	Human, yeast, <i>Drosophila</i> , <i>Arabidopsis</i>
NetGene2	Søren Brunak, http://www.cbs.dtu.dk/services	Human, <i>C. elegans</i> , <i>Arabidopsis</i>
GeneSplicer	Steven Salzberg, http://www.tigr.org/softlab	Human, <i>Arabidopsis</i>
SpliceProximalCheck	Thangavel Thanaraj http://industry.ebi.ac.uk/~thanaraj/MZEF-SPC.html	Human

GENE PREDICTION METHODS



FUNCTIONAL CLASSIFICATION OF GENES(I)

- An early classification scheme for eight related groups of E. coli genes included categories for
- Enzymes transport elements, regulators, membranes, structural elements, protein factors, leader peptides, and carriers.
- Ninety percent of E. coli genes related by significant sequence similarity fell into these same broad categories

FUNCTIONAL CLASSIFICATION OF GENES(II)

- The EC numbers formulated by the **Enzyme Commission of the International Union of Biochemistry and Molecular Biology** provide a detailed way to classify enzymes based on the **biochemical reactions they catalyze**.
- The designation **ECa.b.c.d** (eg. EC 1.4.3.4) gives the following information:
 - (a) **one of six main classes of biochemical reactions,**
 - (b) **the group of substrate molecule or the nature of chemical bond that is involved in the reaction,**
 - (c) **designation for acceptor molecules (cofactors), and**
 - (d) **specific details of the biochemical reaction.**

FUNCTIONAL CLASSIFICATION OF GENES(III)

- A third measure of functional similarity is based on a physiological characterization of *E. coli* proteins into 118 possible categories (e.g., DNA synthesis, TCA cycle, etc.)
- Approximately one-quarter of *E. coli* genes fall into the same category by this scheme.

FUNCTIONAL CLASSIFICATION OF GENES(IV)

- Other functional classification schemes for genes include a broader category for genes involved in the same biological process, e.g., a three-group scheme for
- Energy-related,
- Information-related, and
- Communication-related genes has also been used.
- By this scheme, plants devote more than one-half of their genome to energy metabolism, whereas animals devote one-half of their genome to communication-related functions

FUNCTIONAL CLASSIFICATION OF GENES(V)

- **Gene Ontology(GO)** classification scheme a collaboration among **yeast, fly, and mouse informatics** groups to develop a general classification scheme useful for several genomes
- This classification scheme provides a description of gene products based on
 - **Function,**
 - **Biological role, and**
 - **Cellular location.**

The Gene Ontology: <http://www.geneontology.org/index.shtml>

[Other files](#)

[FTP and CVS downloads](#)

Tools

Documentation

About GO

Projects

[Contact GO](#)



Ontology File Downloads

Terms, definitions and ontology structure.

Current ontology statistics:

as of March 8, 2010 at 18:00 Pacific time:
29959 terms, 99.2% with definitions.

18581 biological_process

2690 cellular_component

8688 molecular_function

There are 1434 obsolete terms not included in the above statistics.

[Download GO ontology files](#)

Format

OBO v1.2 format

[format guide](#)

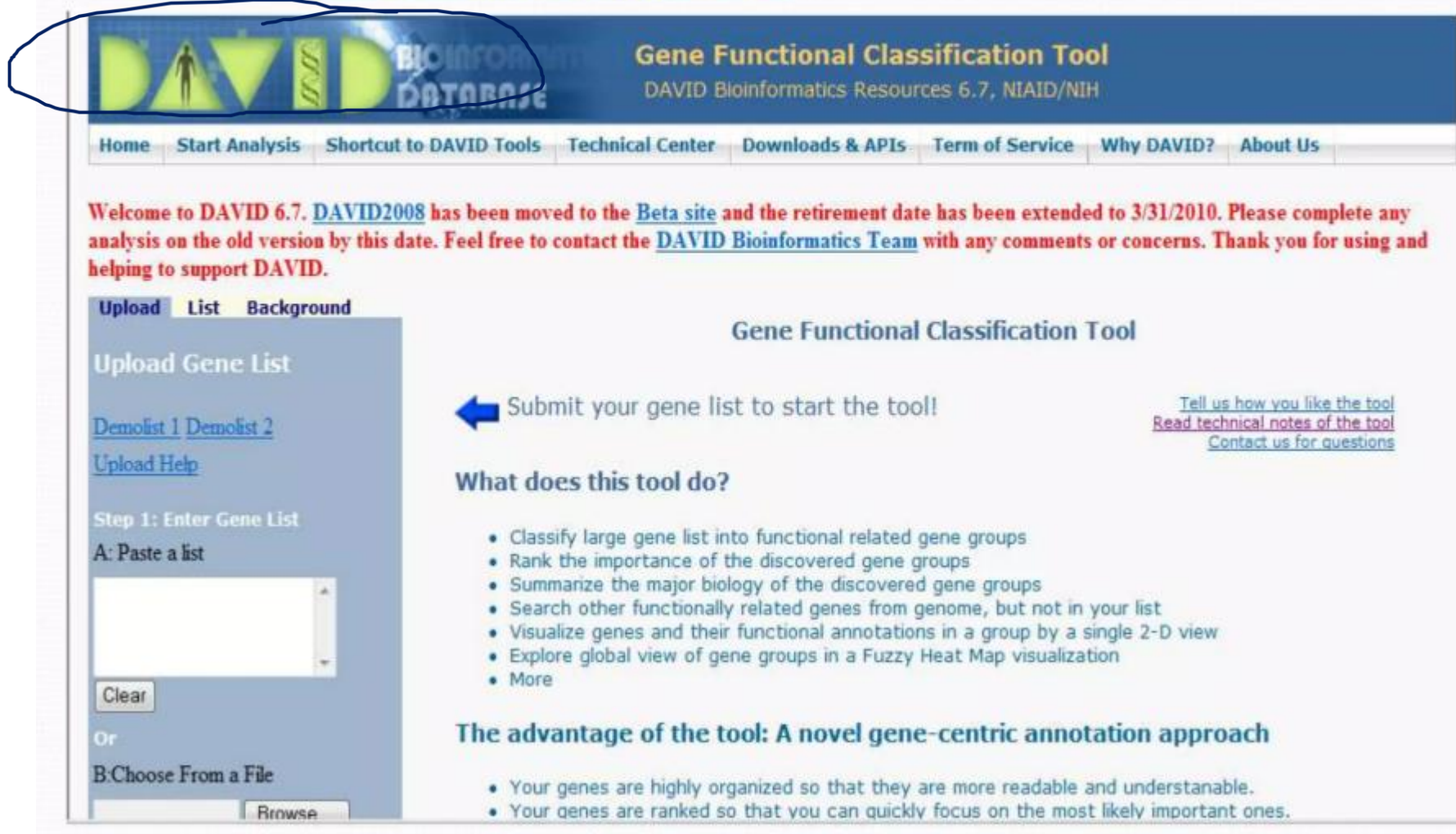
OBO v1.0 format

[format guide](#)

Gene functional classification tool

DAVID : Database for Annotation, Visualization and Integrated Discovery

<http://david.abcc.ncifcrf.gov/home.jsp>



Gene Functional Classification Tool
DAVID Bioinformatics Resources 6.7, NIAID/NIH

[Home](#) [Start Analysis](#) [Shortcut to DAVID Tools](#) [Technical Center](#) [Downloads & APIs](#) [Term of Service](#) [Why DAVID?](#) [About Us](#)

Welcome to DAVID 6.7. [DAVID2008](#) has been moved to the [Beta site](#) and the retirement date has been extended to 3/31/2010. Please complete any analysis on the old version by this date. Feel free to contact the [DAVID Bioinformatics Team](#) with any comments or concerns. Thank you for using and helping to support DAVID.

[Upload](#) [List](#) [Background](#)

Upload Gene List

[Demolist 1](#) [Demolist 2](#)
[Upload Help](#)

Step 1: Enter Gene List

A: Paste a list

Or

B: Choose From a File

Gene Functional Classification Tool

← Submit your gene list to start the tool!

[Tell us how you like the tool](#)
[Read technical notes of the tool](#)
[Contact us for questions](#)

What does this tool do?

- Classify large gene list into functional related gene groups
- Rank the importance of the discovered gene groups
- Summarize the major biology of the discovered gene groups
- Search other functionally related genes from genome, but not in your list
- Visualize genes and their functional annotations in a group by a single 2-D view
- Explore global view of gene groups in a Fuzzy Heat Map visualization
- More

The advantage of the tool: A novel gene-centric annotation approach

- Your genes are highly organized so that they are more readable and understandable.
- Your genes are ranked so that you can quickly focus on the most likely important ones.

MOLECULAR ARCHAEOLOGY

EVOLUTION

1. How we evolved from ape-like ancestors to modern humans, including, when did we diverge from apes, who were the early hominids, and when did modern humans spread out of Africa?
2. What makes us human - what are some of the adaptations that evolved as humans migrated around the globe and established new cultures? In what ways are we still evolving?

In 1871 Charles Darwin proposed in his book "The Descent of Man" that humans evolved in Africa and shared a common ancestor with great apes. This showed remarkable foresight as at that time there were no early human fossils and no DNA evidence.



In the case of human evolution, Charles Darwin noted that

- humans and apes share many biological features .
- He hypothesized that humans and apes share a common ancestor.
- If this hypothesis is true then humans and apes should share a high percentage of their DNA,
- and there should be fossils with humans and ape-like features.
- Darwin wasn't able to gather this sort of evidence to test his hypothesis in his time, but many studies since then have tested these predictions .

Both DNA and fossil evidence support the hypothesis of humans and apes sharing a common ancestor, and the idea that humans evolved from ape-like creatures is now an accepted theory.

