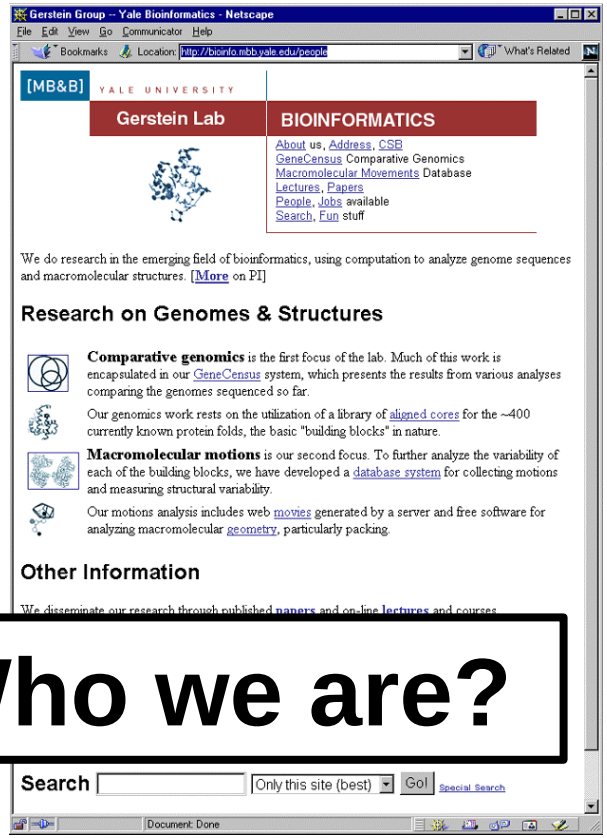
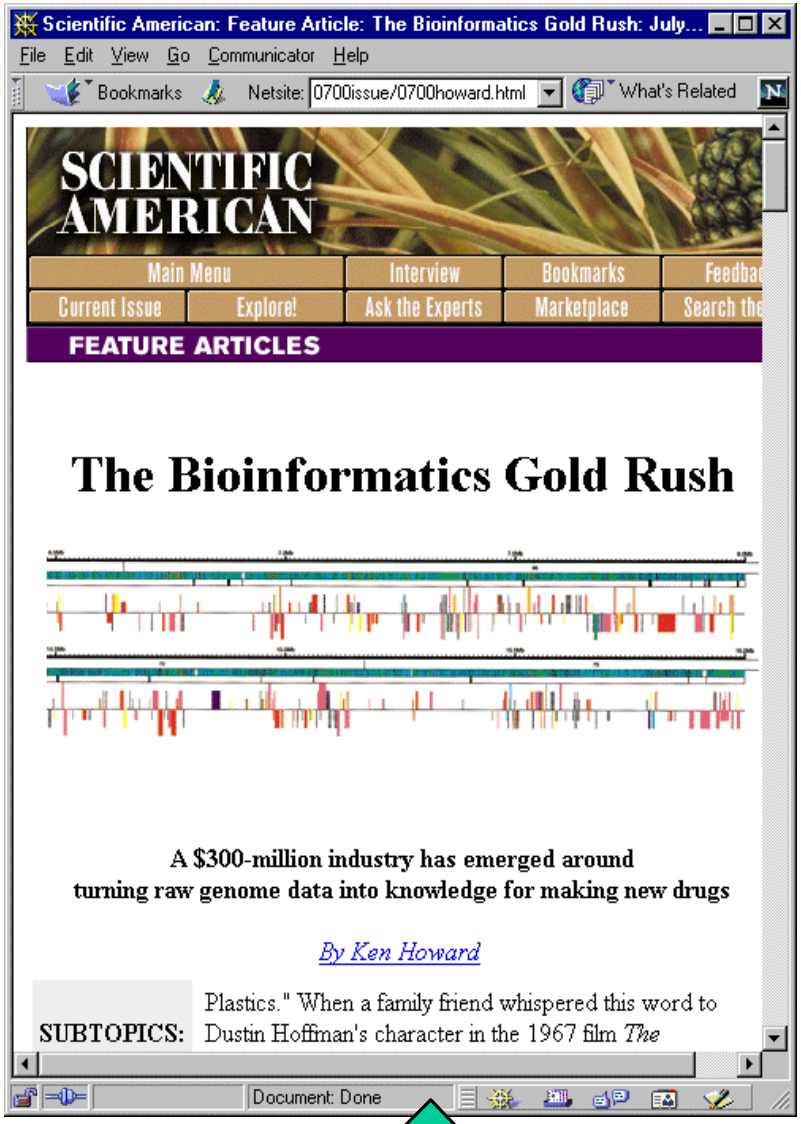
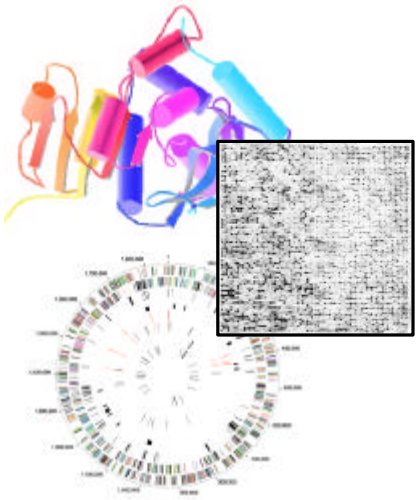
MB&B Bioinformatics Group (Gerstein Lab)



Who we are?



Seems to becoming of interest...



What do we do?

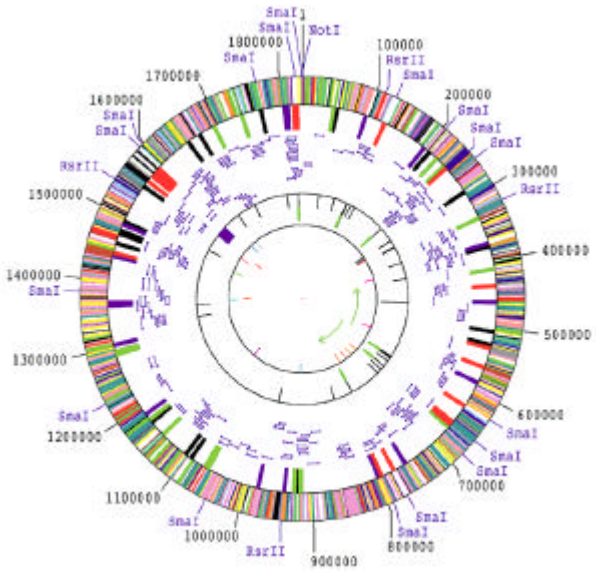
(Molecular) **Bio - informatics** =

Conceptualizing **biology in terms of molecules** (in the physical-chemical sense) and then **applying "informatics"** techniques (derived from disciplines such as CS and statistics) to **organize and analyze** the **information** associated with these molecules, on a **large-scale** and in an **integrative fashion**. Bioinformatics is a practical discipline with many **applications**.

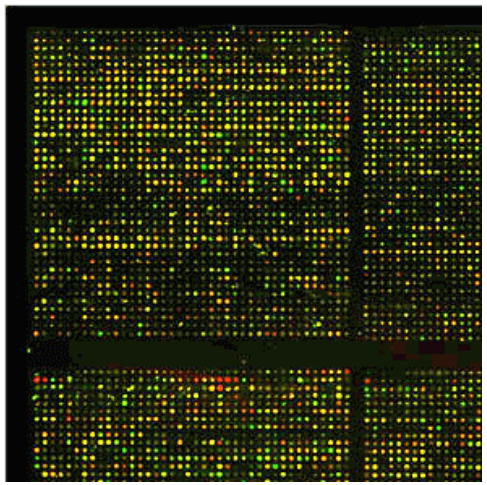
Large-scale, Integrative Analysis of...

Structures

Genomes



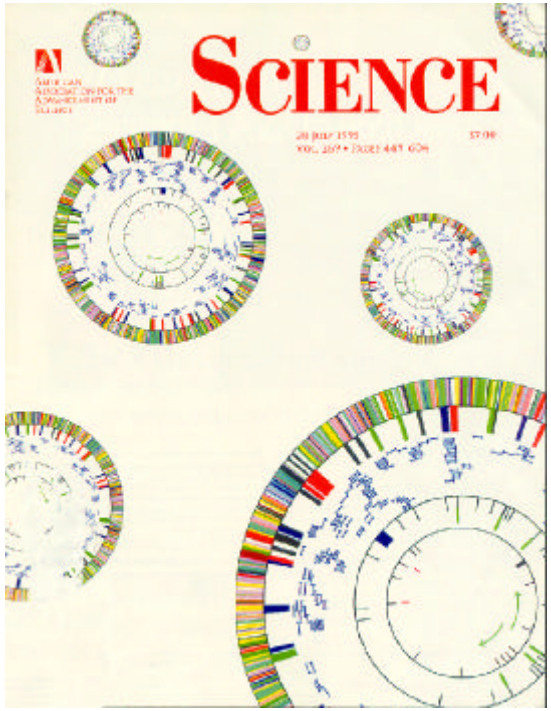
Expression Data



1990 2000 2010

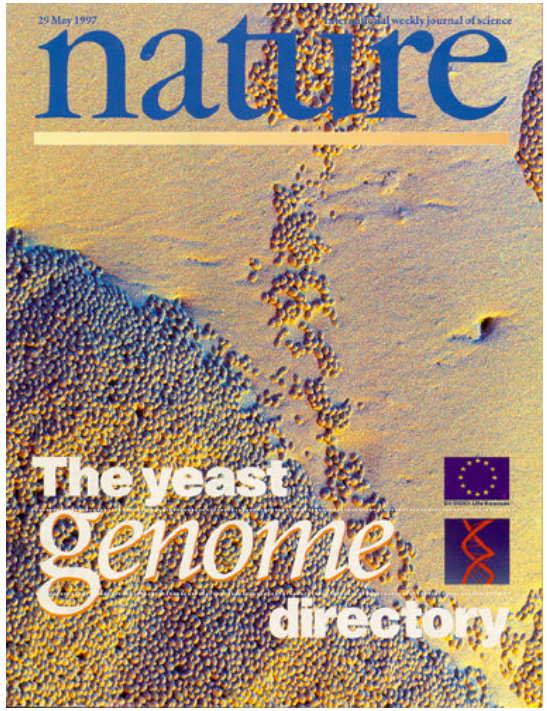
1995

Bacteria,
1.6 Mb,
~1600 genes
[*Science* **269**: 496]



1997

Eukaryote,
13 Mb,
~6K genes
[Nature **387**: 1]

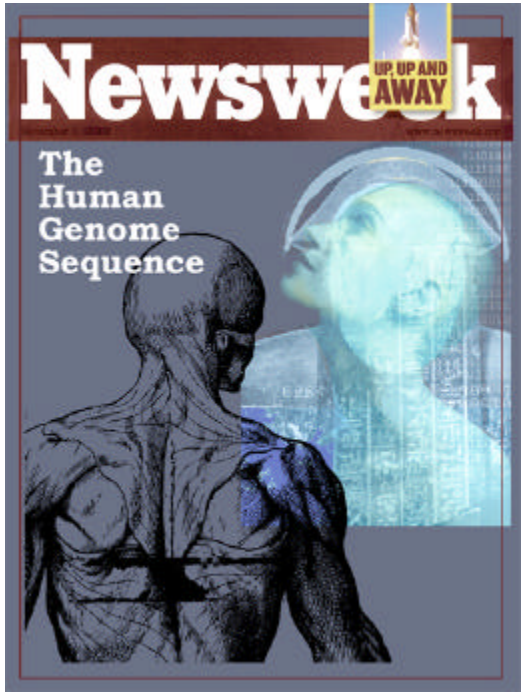


1998

Animal,
~100 Mb,
~20K genes
[*Science* **282**:
1945]

2000

Human,
~3 Gb,
~100K
genes [???]



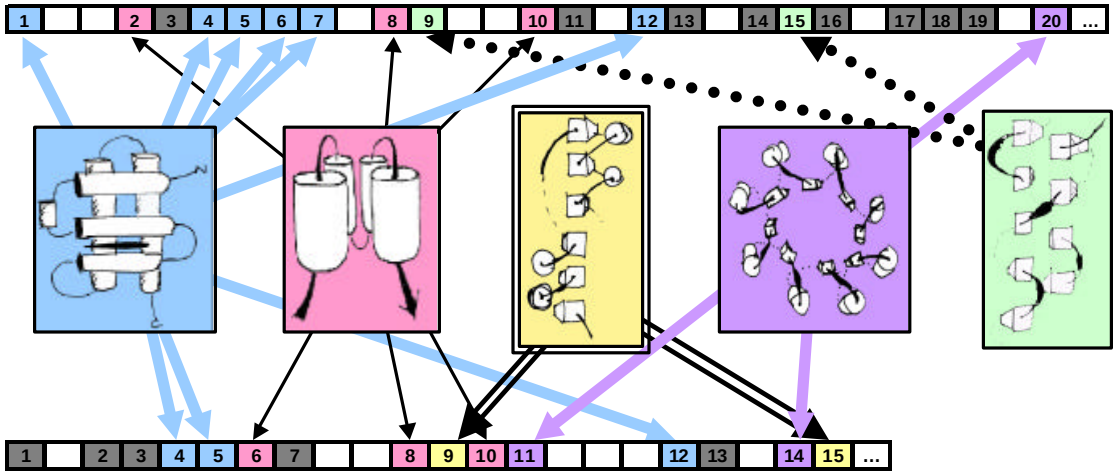
‘**98** spoof real thing, Apr ‘**00**

2000 The Year of Genomics

The Finite Parts List as a Simplifying Theme

(human)

(*T. pallidum*)



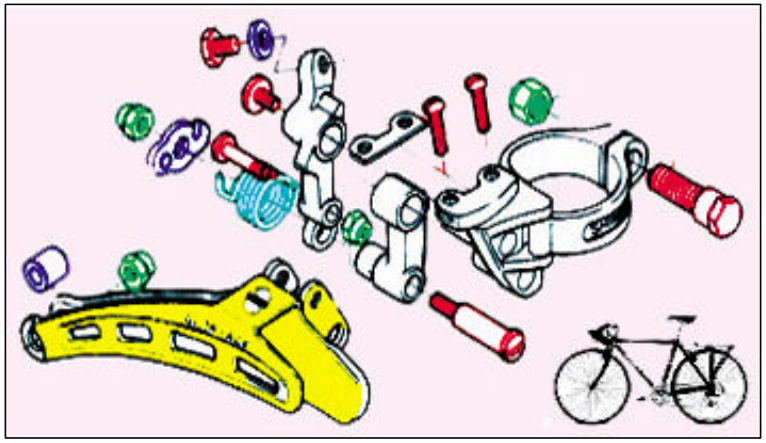
~100000 genes

~1000 folds

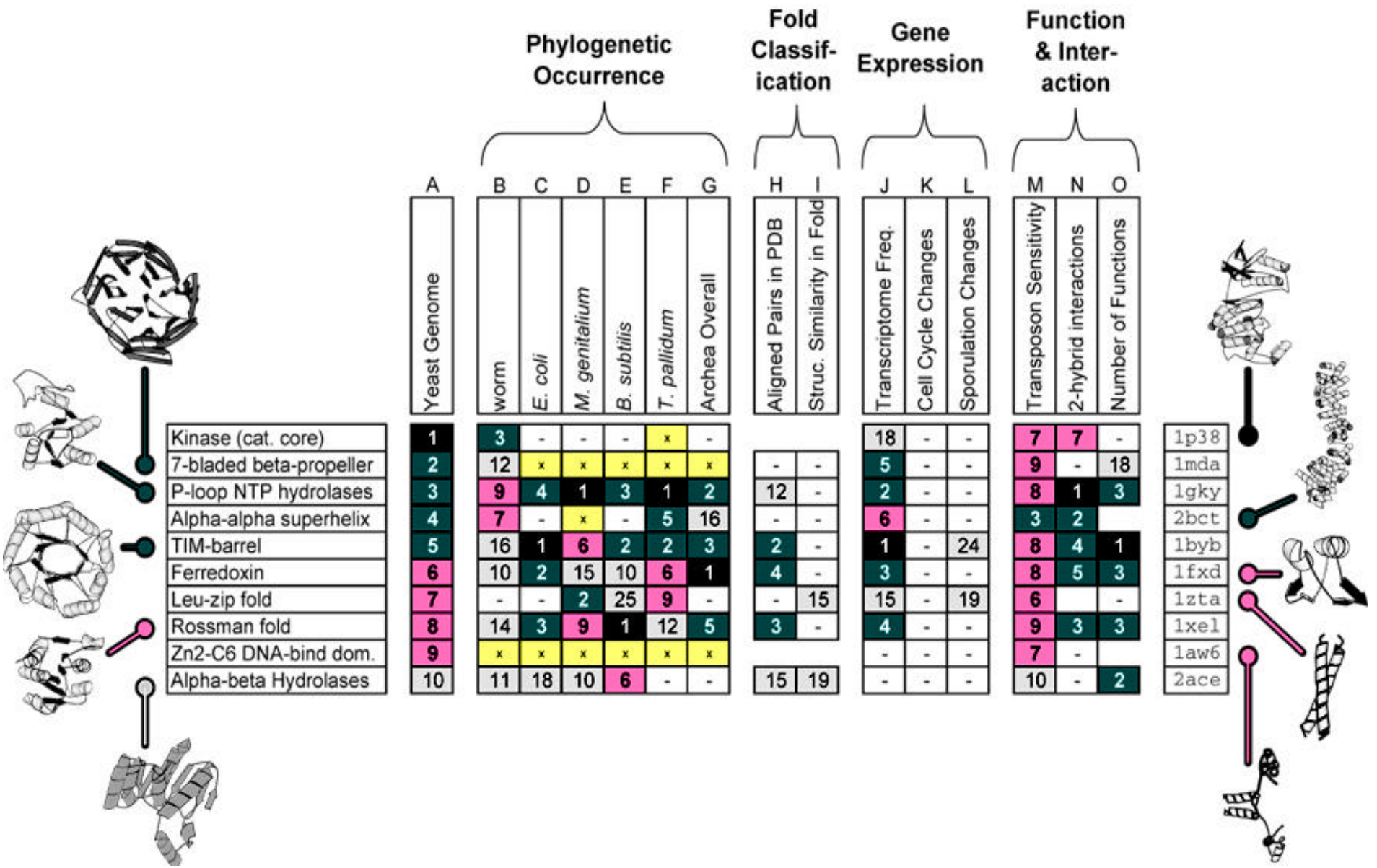
~1000 genes

Parts = **FOLDS**, pathways, functions, sequence families, blocks, motifs....

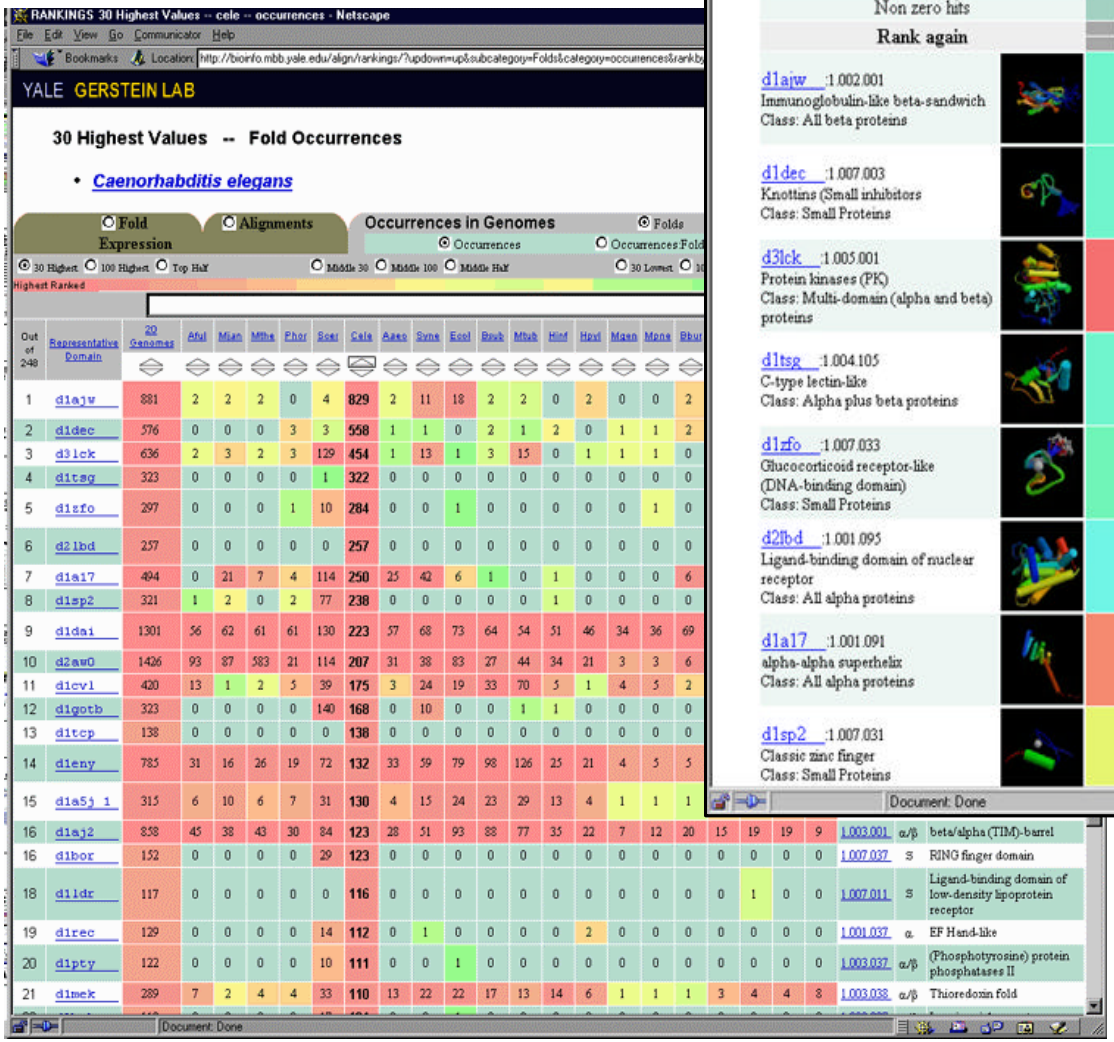
Global Surveys of a Finite Set of Parts from Many Perspectives

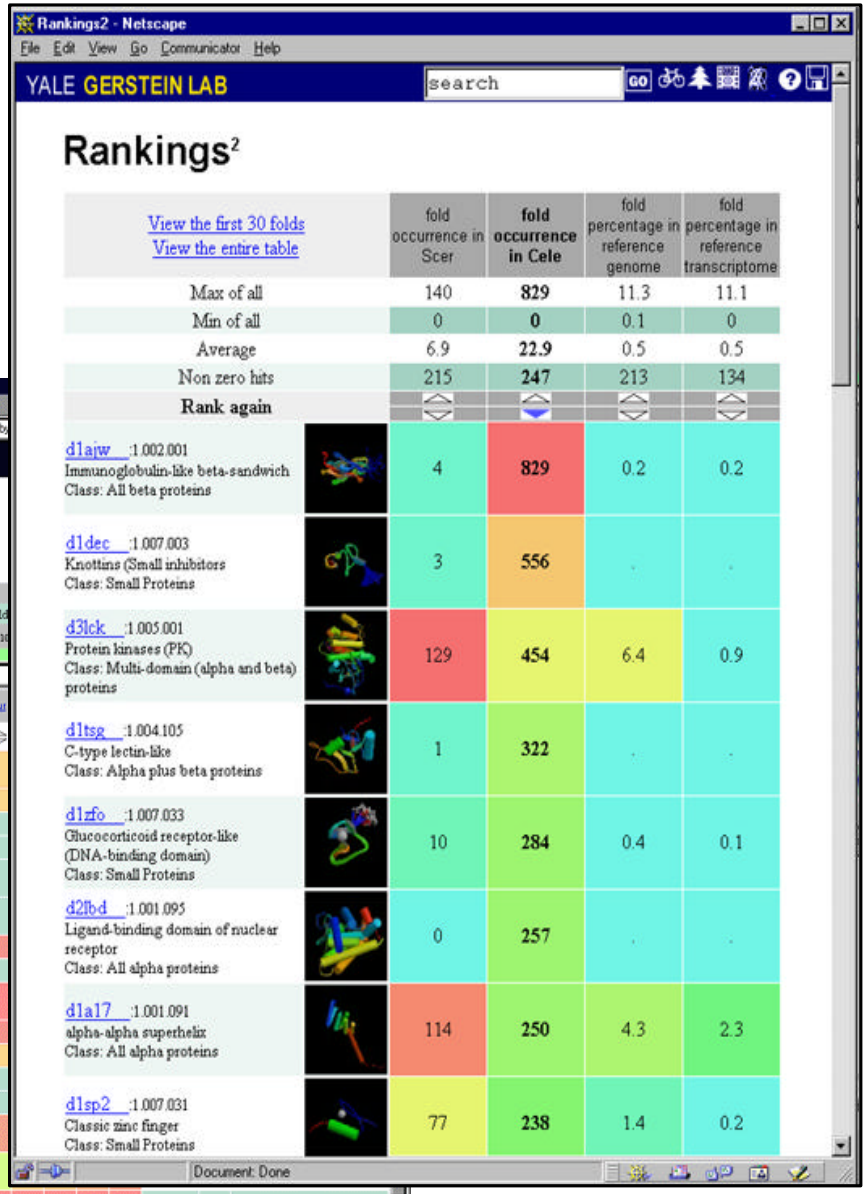


Surveying a Finite Parts List from a Infinite Number of Perspectives



PartsList Ranking Viewers

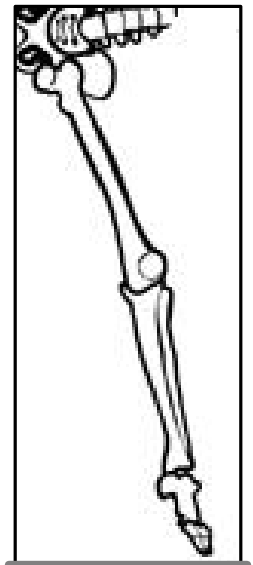


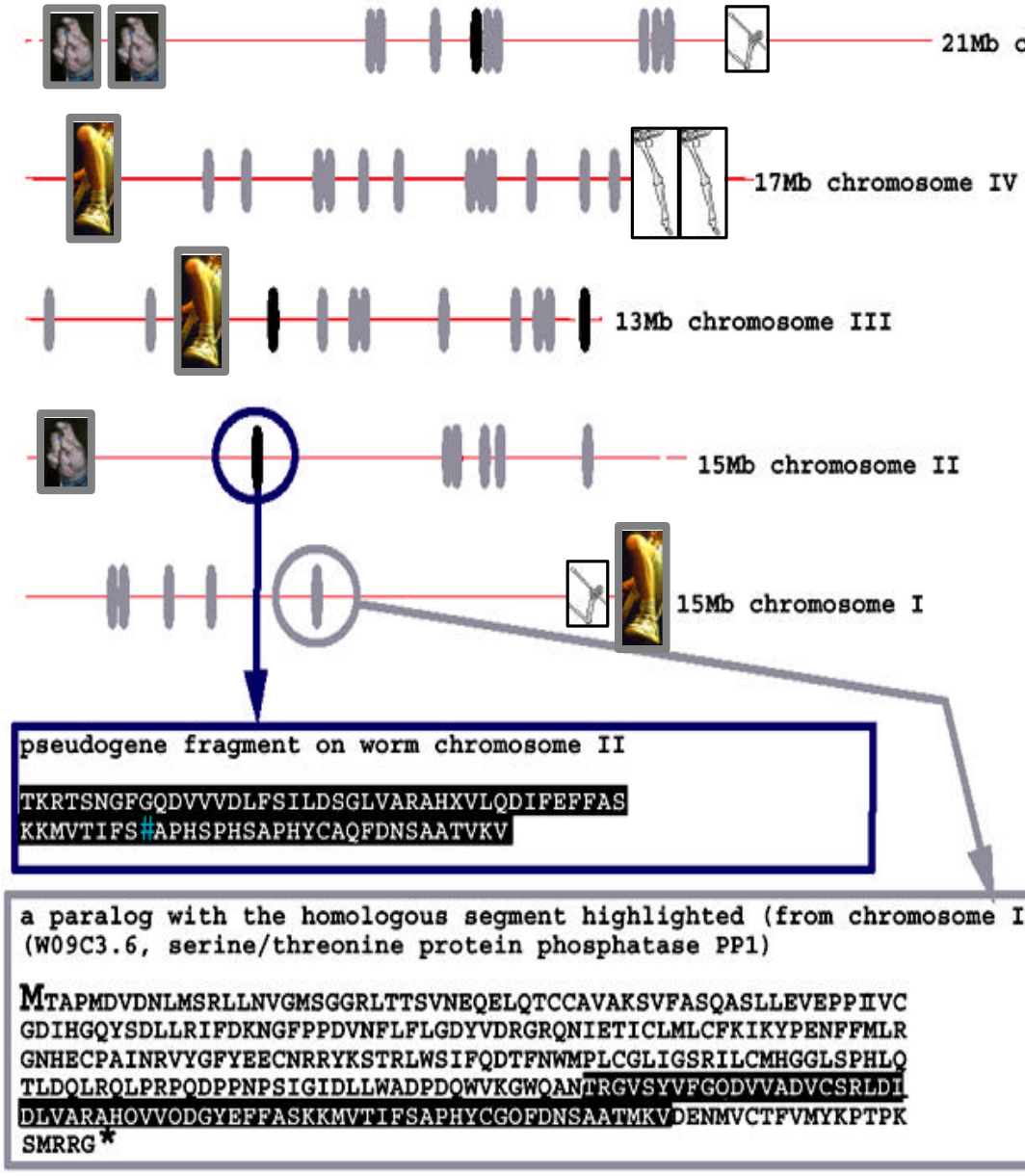


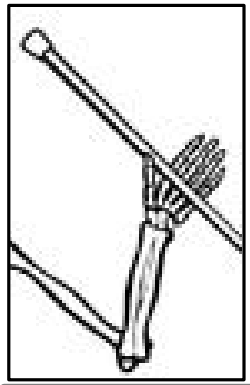
Rank Folds by Genome Occurrence, Expression, Fold Clustering, Length, &c

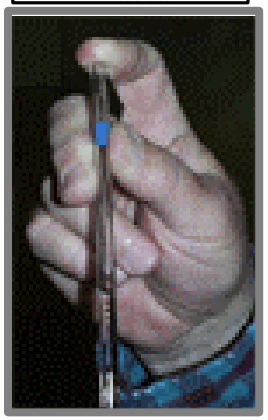
J Qian,
B Stenger,
J Lin....

Pseudogenomics: Surveying "Dead" Parts









Example of a potential ΨG with frameshift in mid-domain