

# What makes species different? A study of Unique Genes

## Our Definition of genes with unknown function

Gene models and proteomes for *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Drosophila melanogaster* (Dm), *Anopheles gambiae* (Ag), *Caenorhabditis elegans* (Ce), *Mus musculus* (Mm), *Rattus norvegicus* (Rn), and *Homo sapiens* (Hs) were downloaded from the NCBI website (<ftp.ncbi.nlm.nih.gov>).

HMMPFAM search against several major signature databases- PFAM, TIGRFAM, SMART, and Superfamily

match to one or more of the models in any one of the databases

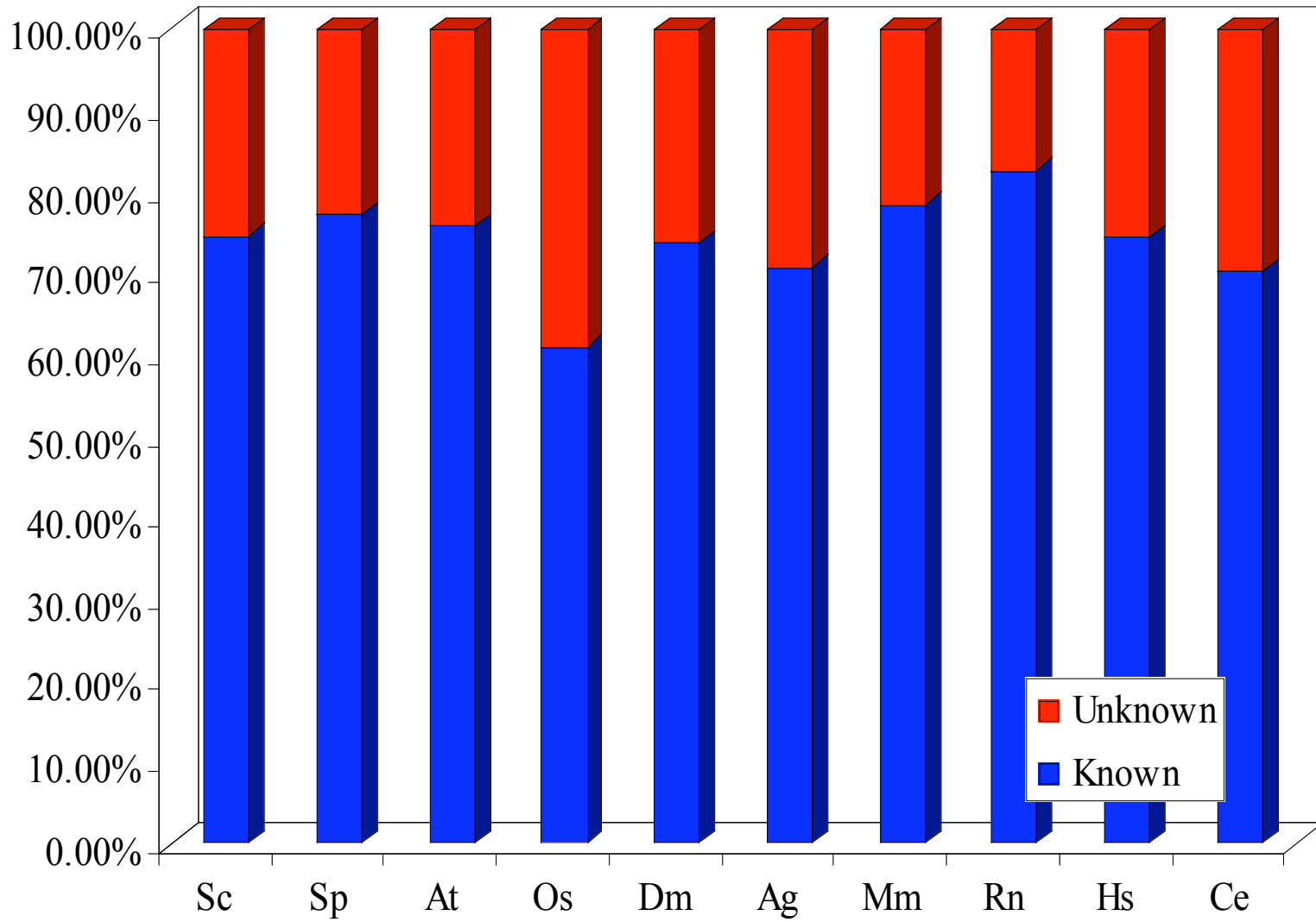
“Known (PDF)”

no matches to any one of the models in any database

“Unknown(POF)”

“Unknowns” account for about 25% of each genome

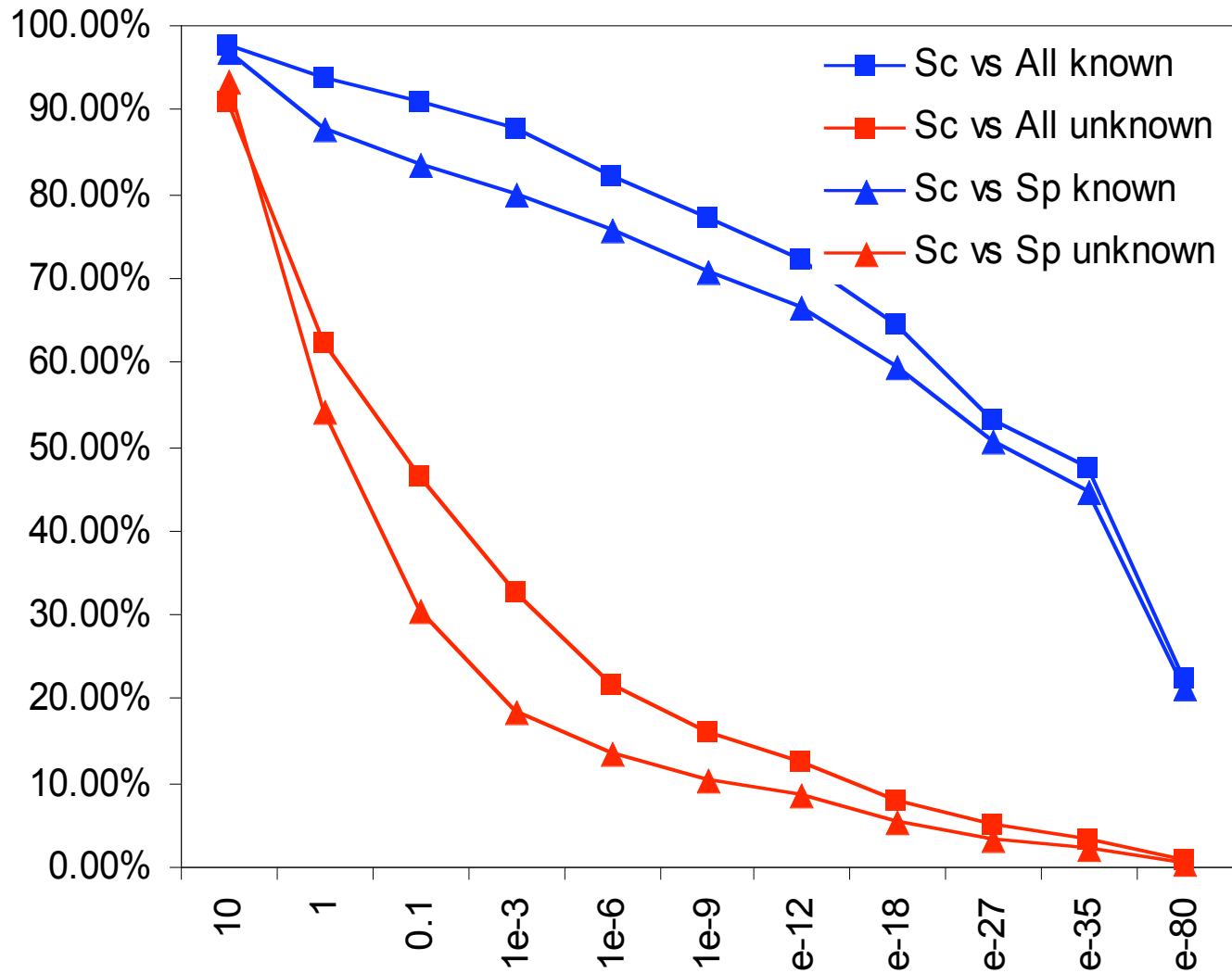
% of Genome



“Unknowns” are not as conserved as “knowns”, even between related organisms!

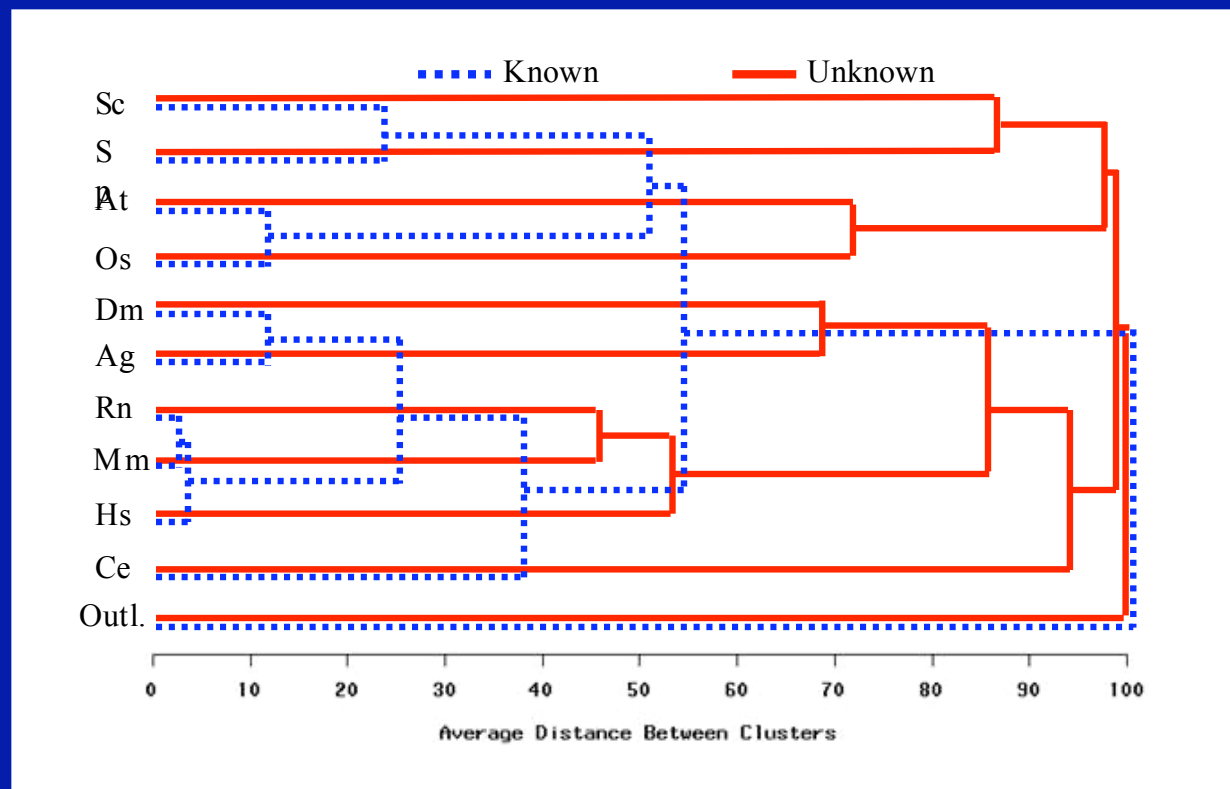
Yeast

% hits at each e-value



BLAST e-value

Relationship tree among the 10 different genomes reveals a high degree of evolutionary divergence among “unknowns” from different species

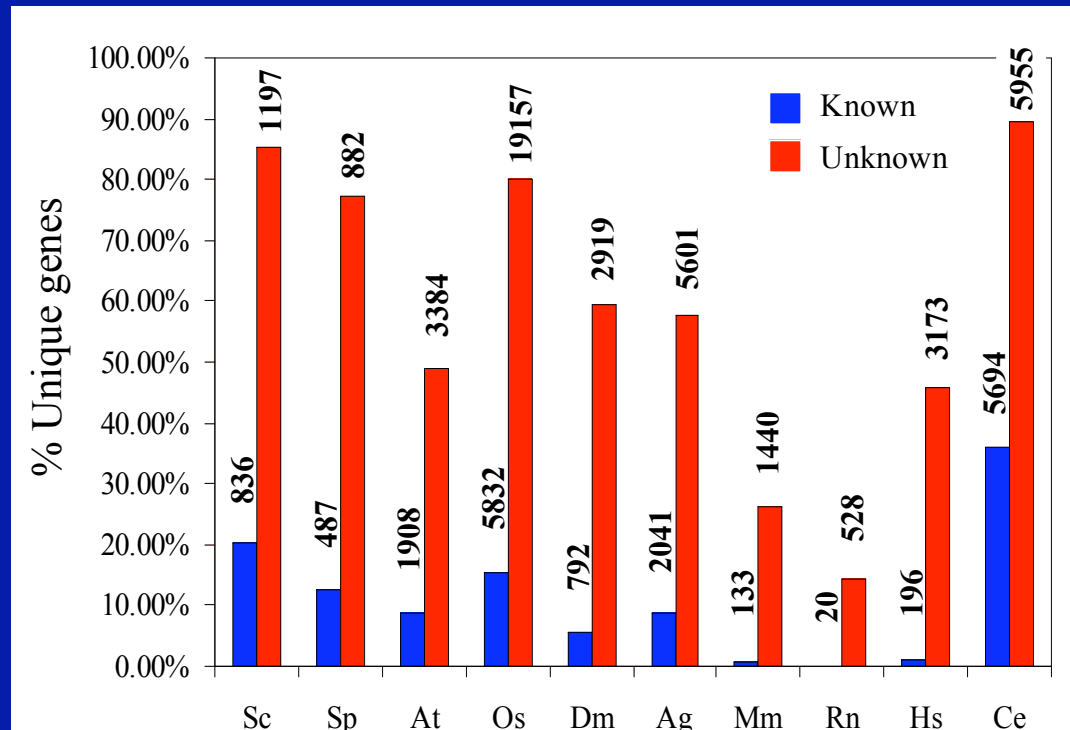


“Unknowns” have a different rate of evolution?

“Unknowns” are new genes?

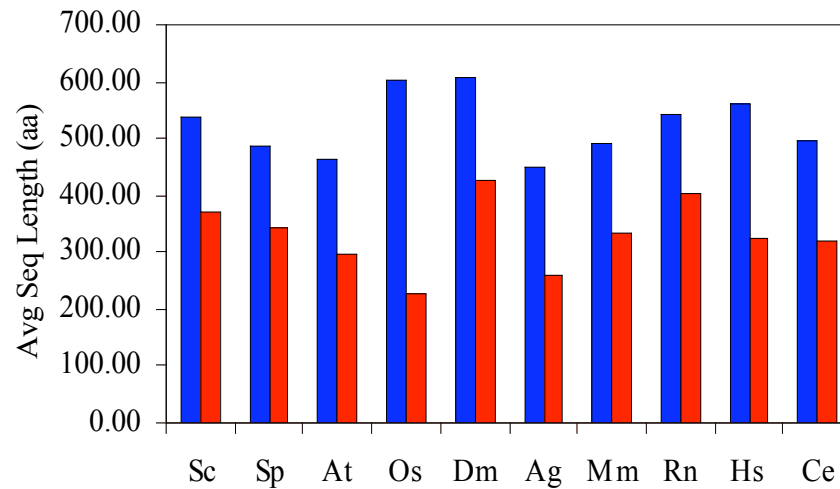
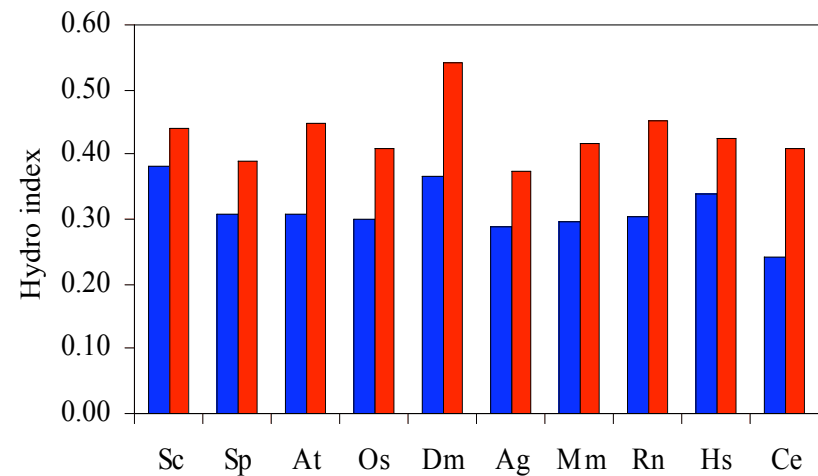
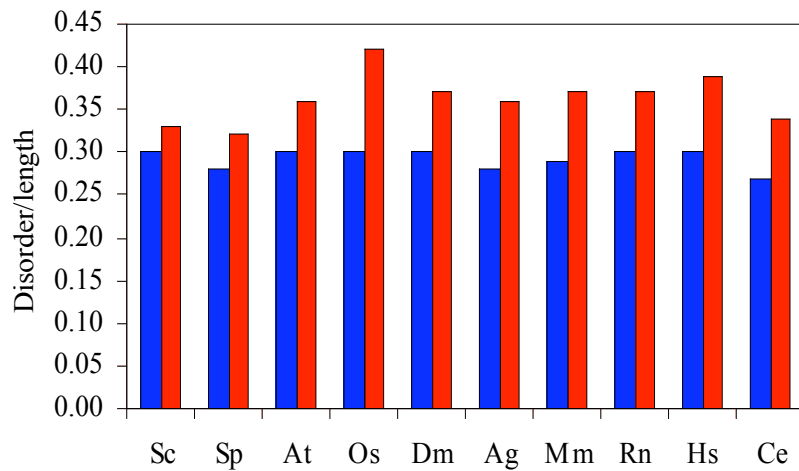
## “Unknowns” are mainly species-specific.

Representation of “unknowns” in the “unique-ome” of different species.



“Unique-ome” was defined by a BLAST cut off of  $10^{-6}$ .  
Between the 10 different genomes!

Compared to “knowns”, “Unknowns” are more disordered, less hydrophobic and shorter.



■ Known  
■ Unknown

# “Unknown” Conclusions

- Unknown genes are typically species-specific and might provide some of the keys that define species-specific differences.
- Unraveling the function of “unknowns” would improve our understanding of species-specific functions.
- Disordered protein functions are thought to include the formation and regulation of large multi-molecular assemblies that participate in important regulatory functions. Disordered regions on proteins have been reported to evolve significantly more rapidly than ordered regions.
- “Unknowns” are likely to be the result of greater evolutionary divergence among species leading to the establishment of new, species-specific regulatory networks.