

BL/CH401 Lec. 3 -- Current Info on Genome Sequencing Projects

Completed microbial genomes -- 11; Microbial genomes in progress -- 33

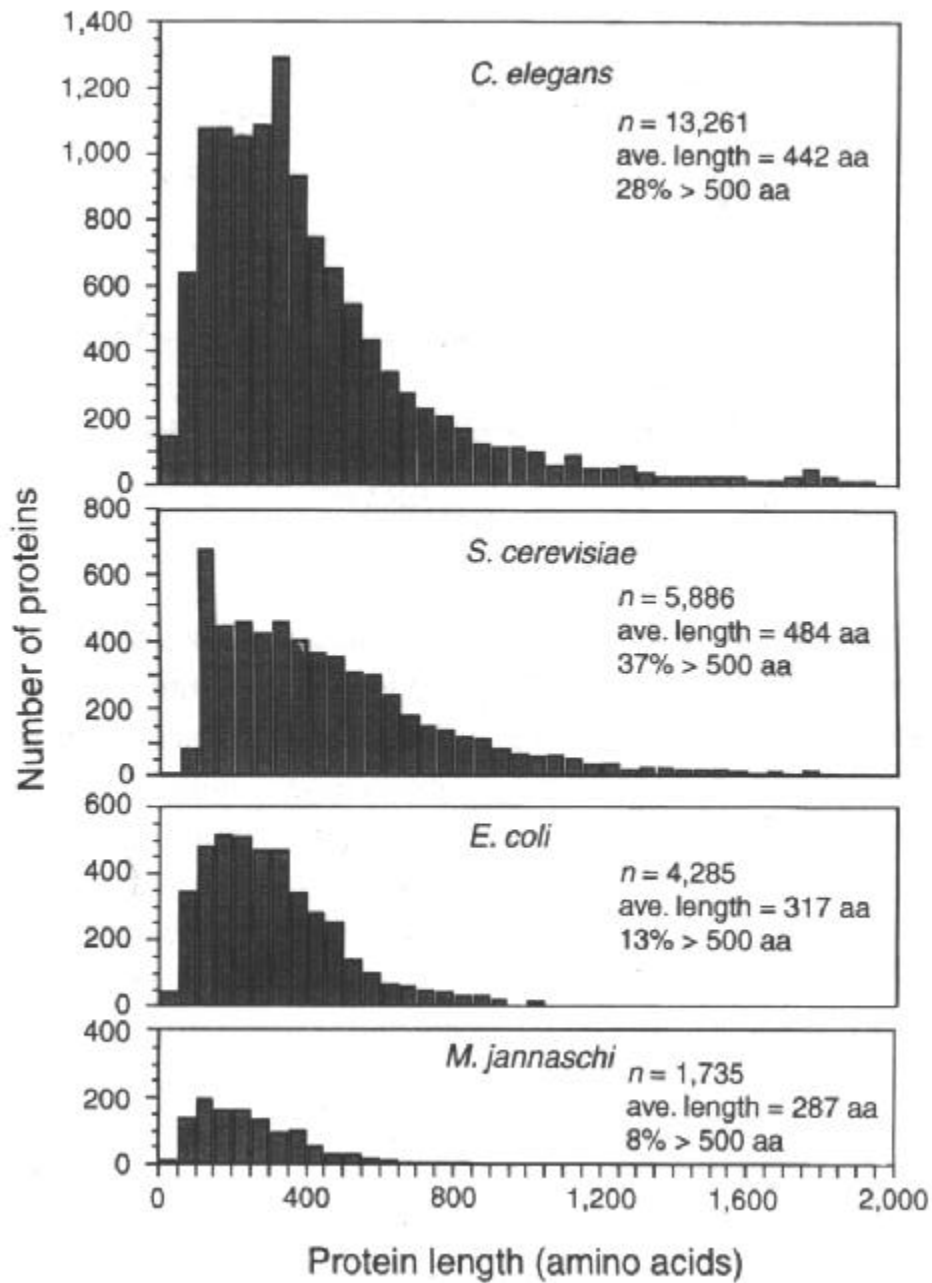
In Table showing info on completed microbial genomes, for Domain (Kingdom of Life): B = Eubacterium; A = Archibacterium; E = Eukaryote

Genome	Strain	Domain	Size (Mb)	Year
<i>Haemophilus influenzae</i> Rd	KW20	B	1.83	1995
<i>Mycoplasma genitalium</i>	G-37	B	0.58	1995
<i>Methanococcus jannaschii</i>		A	1.66	1996
<i>Synechocystis</i> sp.	PCC 6803	B	3.57	1996
<i>Mycoplasma pneumoniae</i>		B	0.81	1996
<i>Saccharomyces cerevisiae</i>		E	13	1997
<i>Helicobacter pylori</i>	26695	B	1.66	1997
<i>Escherichia coli</i>	K-12	B	4.60	1997
<i>Bacillus subtilis</i>		B	4.20	1997-98
<i>Archaeoglobus fulgidus</i>	VC-16, DSM4304	A	2.20	1997-98
<i>Borrelia burgdorferi</i>		B	1.30	1997-98

Source: http://www.tigr.org/tigr_home/tdb/mdb/mdb.html

How do we use Genome sequence data?

One use recently was to compare the size of proteins encoded by individual genomes in terms of the number of amino acids encoded in polypeptide sequences. Reported in Nature Vol 388, pages 343 to 349 by W.J. Netzer and F.U. Hartl in an article entitled, "Recombination of protein domains facilitated by co-translational folding in eukaryotes". Their Fig. 1 is shown below comparing the size of proteins in 4 genomes: *Caenorhabditis elegans* (eukaryote), *Saccharomyces cerevisiae* (eukaryote), *Escherichia coli* = *E. coli* (eubacterium), and *Methanococcus jannaschi* (archibacterium). The authors concluded that part of the evolutionary process involved development of a mechanism for multi-domain proteins to fold at the same time as they are synthesized on ribosomes, which is called co-translational folding. Eukaryotic proteins are larger than prokaryotic (see Fig. 1 below) and many eukaryotic proteins have large independent subparts which we can call domains since they have a distinctly separate folding pattern. Prokaryotes, like *E. coli*, only fold proteins after translation is complete, which is called post-translational folding. This explains part of the reason it is sometimes difficult to express eukaryotic proteins in *E. coli* in their native, naturally folded form. The authors also prove this point by comparing a model protein composed of human H-Ras and mouse dihydrofolate reductase joined by a synthetic linker and show that both proteins fold well in eukaryotic protein synthesis systems but fold poorly in *E. coli*. Read the introduction to the article if you want to learn more about the importance of this discovery to evolutionary thinking about protein adaptation during time.



©Wilbur H. Campbell, 1995, 1996, 1997; wcampbel@mtu.edu

Return to [Lec 2 Home Page](#)