

The UniProt Knowledgebase

The UniProt Knowledgebase (UniProtKB) is a freely accessible resource for functional information on proteins with accurate, consistent, rich and comprehensive annotation. In addition to manually reviewed proteins, UniProtKB holds a large number of nucleotide-derived protein sequences, many of which are annotated by automatic annotation procedures.

Complete and reference proteome sets

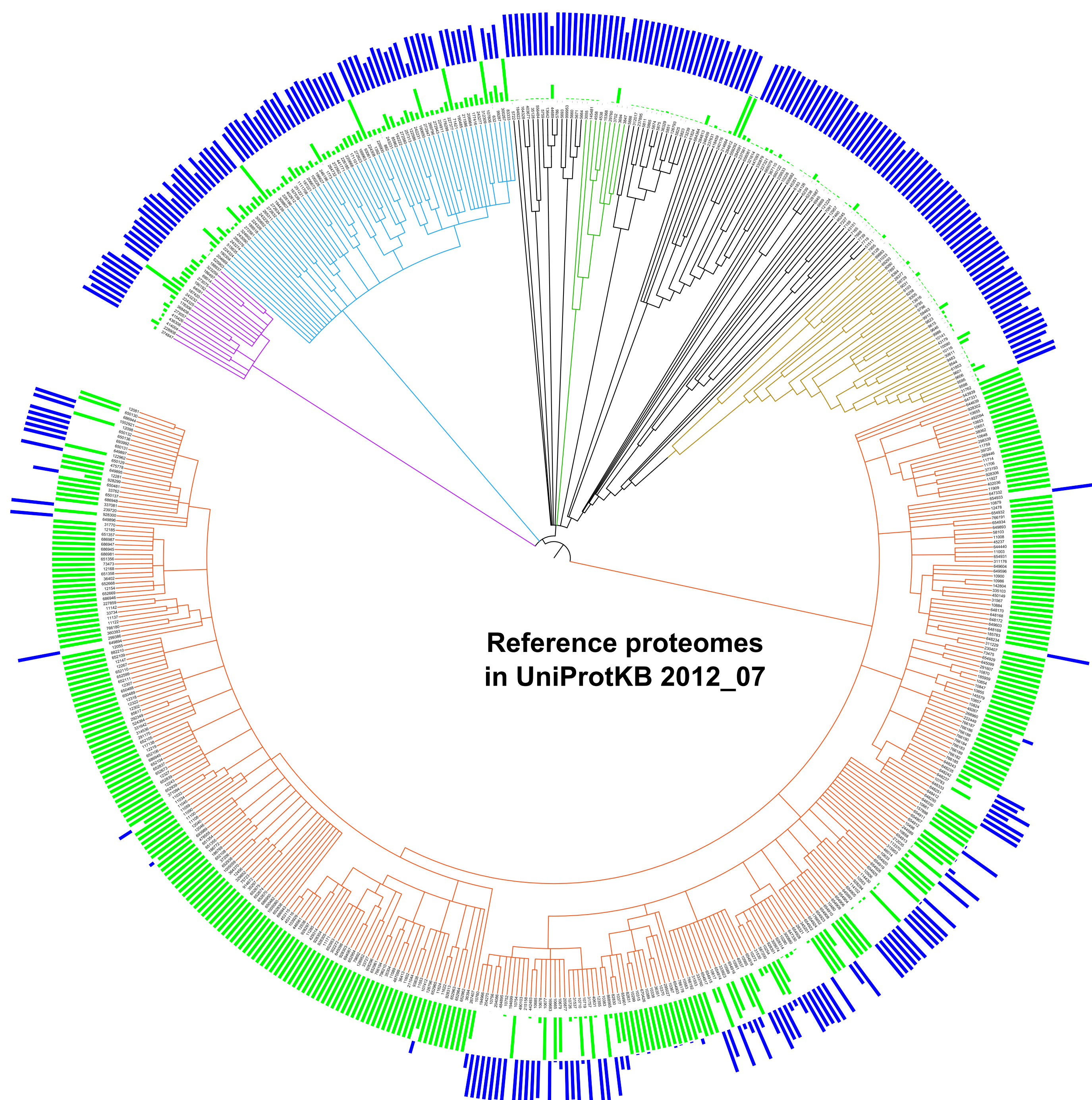
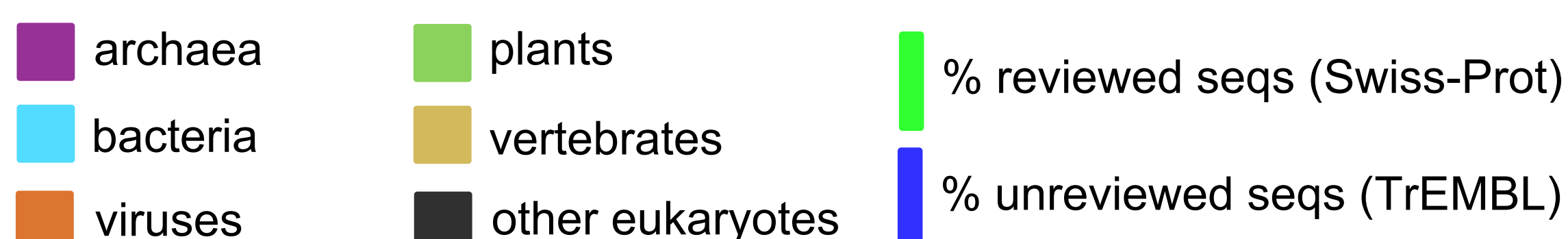
As the cost of sequencing continues to fall, the number of organisms with completely sequenced and annotated genomes is rapidly increasing. UniProtKB endeavours to provide comprehensive protein-centric views of such genomes in the form of complete proteome sets.

What are complete proteome sets?

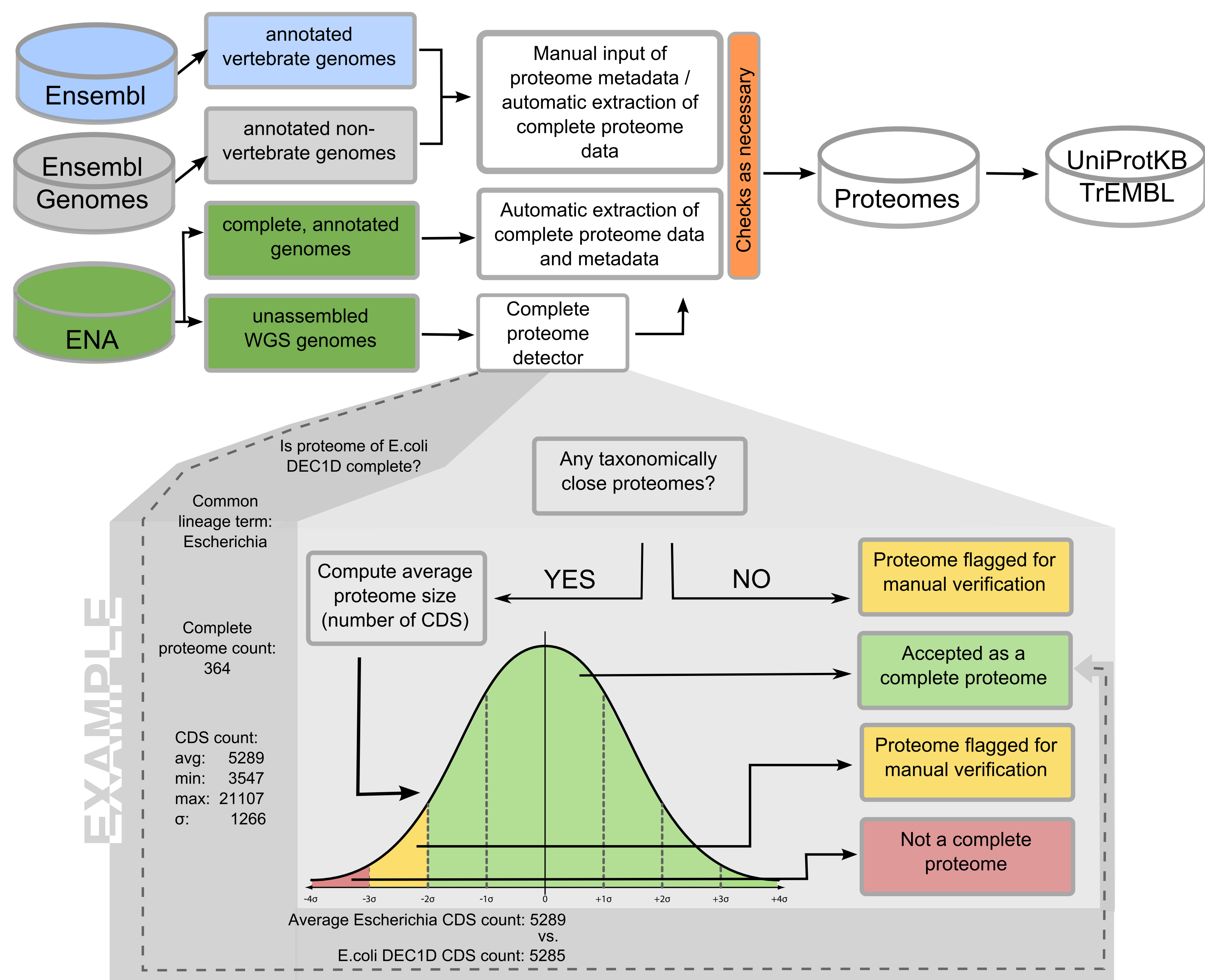
A complete proteome set is defined as the entire set of proteins expressed by a specific organism. The majority of the UniProt complete proteome sets are based on the translation of a completely sequenced genome, and will normally include sequences that derive from extra-chromosomal elements such as plasmids or organellar genomes in organisms where these occur.

What are reference proteome sets?

Reference proteome sets are a subset of complete proteome sets which have been manually selected and include well-studied model organisms and other proteomes of interest for biomedical and biotechnological research. Currently (UniProtKB release 2012_07), there are 556 reference proteome sets. Reference proteomes have been chosen to provide broad coverage of the tree of life, and constitute a representative cross-section of the taxonomic diversity to be found within UniProtKB.



Proteomes pipeline



Challenges of multiple proteomes from one species

Currently, complete proteomes are accessed through the UniProt website using the taxonomy identifier (taxID) of the organism. However, there are now an increasing number of cases where the same strain or isolate, having the same taxID, has been completely sequenced by more than one sequencing project. A new unique proteome identifier is needed which will distinguish between duplicate proteomes of the same organism. The INSDC Genome Assemblies database assigns a unique identifier to each sequencing project. This identifier provides a link between the complete genome of an organism and the corresponding complete proteome. By using the genome collections identifier UniProt can distinguish between duplicate sequencing projects.

How to access proteome sets in UniProtKB

<http://www.uniprot.org>

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: ★ Swiss-Prot, which is manually annotated and reviewed. ★ TrEMBL, which is automatically annotated and is not reviewed. Includes complete and reference proteome sets.
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations, taxonomy, keywords, subcellular locations, cross-referenced databases and more.

COMPLETE PROTEOMES AND REFERENCE PROTEOMES

A complete proteome consists of the set of proteins thought to be expressed by an organism whose genome has been completely sequenced.

A reference proteome is the complete proteome of a representative, well-studied model organism or an organism of interest for biomedical research.

These organisms can be searched via the taxonomy pages, which provide links to download complete and reference proteome sets when available, as well as links to the HAMAP web site.

Browse or list organisms with:

Complete proteomes	Reference proteomes
<ul style="list-style-type: none"> Browse by hierarchy List all Bacteria List all Archaea List all Eukaryota List all Viruses 	<ul style="list-style-type: none"> Browse by hierarchy List all Bacteria List all Archaea List all Eukaryota List all Viruses

Search organisms with complete proteomes:

Search organisms with reference proteomes:

or
FTP download
or
programmatic access

UniProt is funded by the European Molecular Biology Laboratory, the US National Institutes of Health, the European Union and the Swiss Federal Government.