

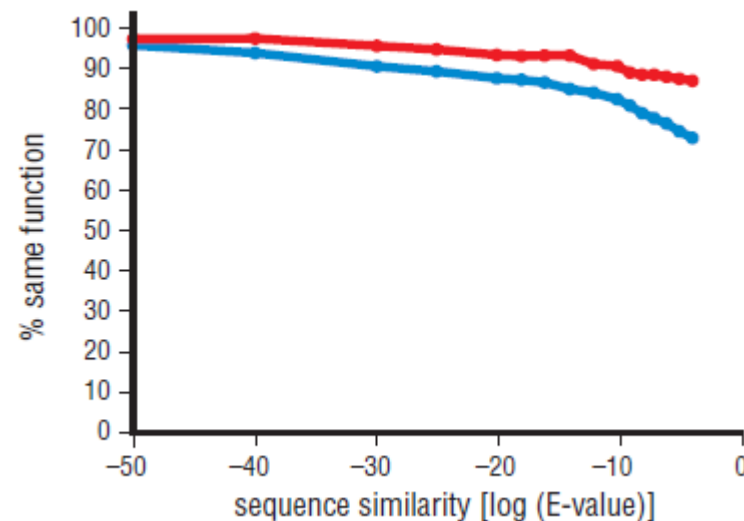
EXPERIMENT 1:

Primary and secondary structure analysis

Dr. Zhiyi Wei
SUSTC

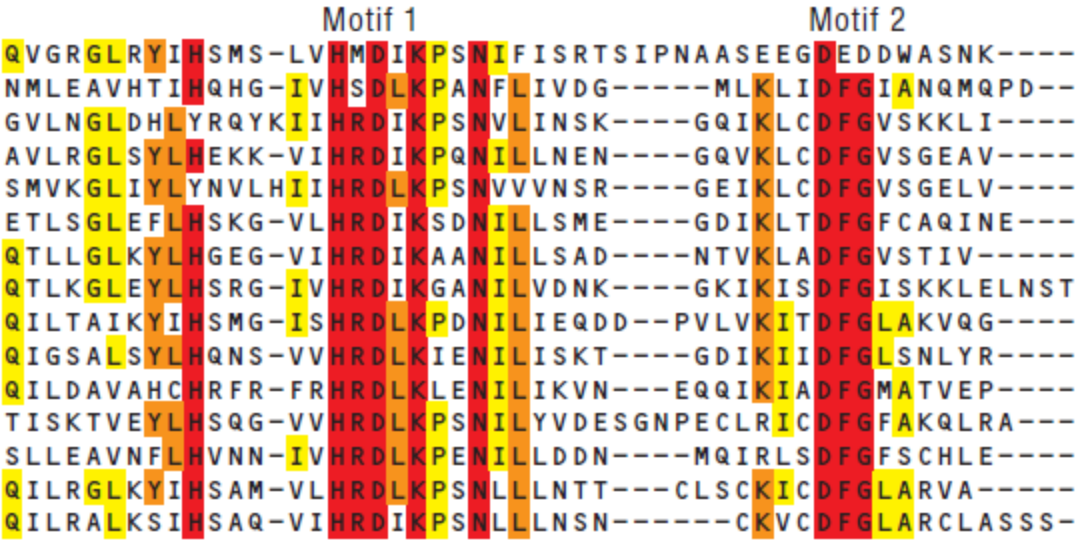
Sequence comparison

- Sequence comparison provides a measure of the relationship between genes



Sequence conservation

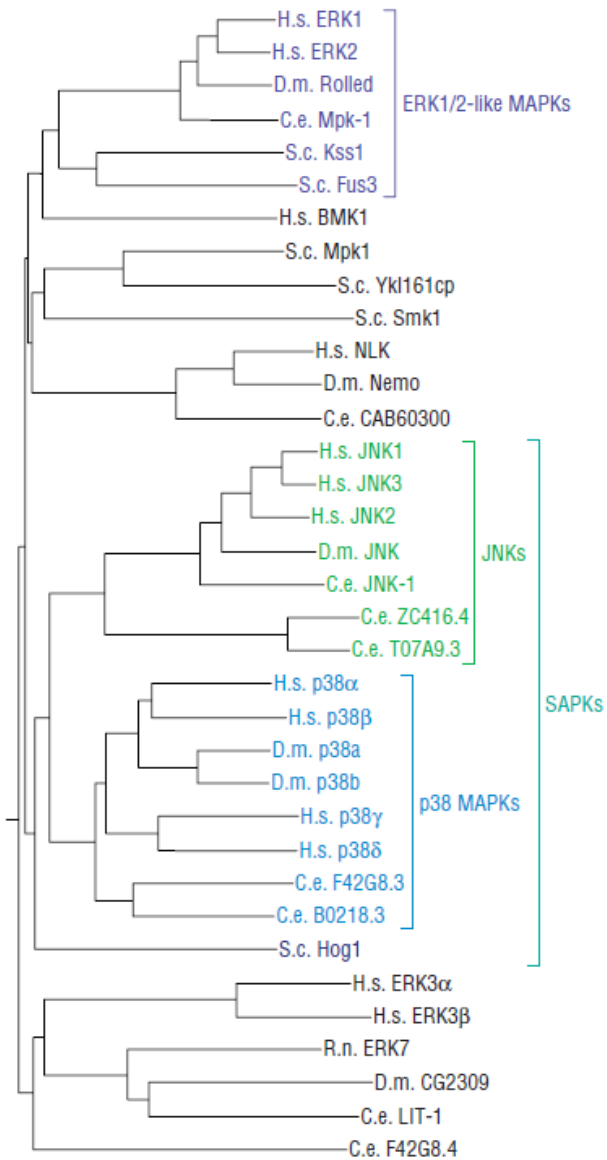
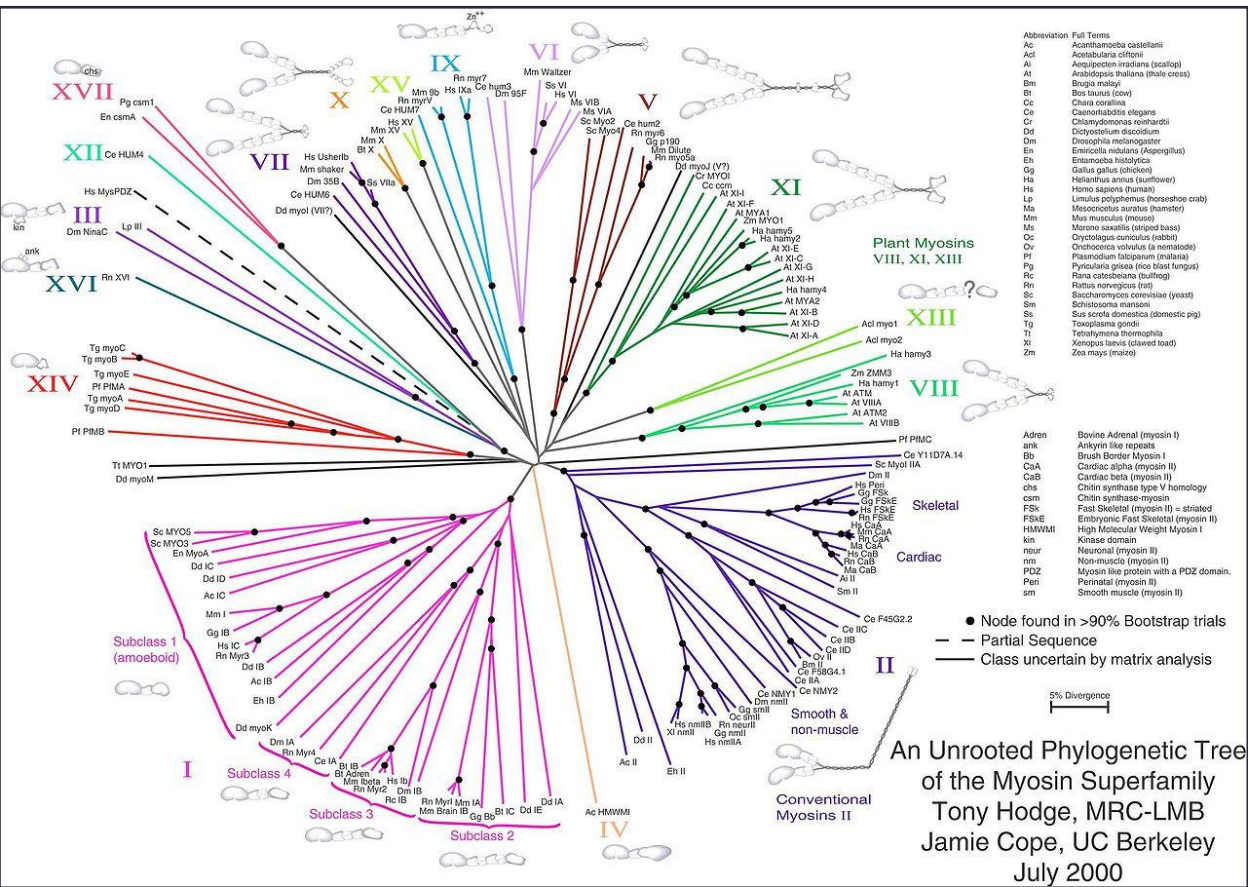
H.s. Wee1 409-457
H.s. Ttk 614-659
S.c. Ste7 313-358
S.c. Mkk1 332-376
S.p. Byr1 168-213
S.c. St20 722-767
S.c. Cc15 129-172
S.p. Byr2 505-553
S.c. Spk1 302-348
S.p. Kin1 249-293
S.p. Cdr1 111-156
M.m. K6a1 507-556
R.n. Kpbh 136-180
H.s. Erk2 132-176
S.c. Kss1 137-182



Multi-sequence alignment

- A target sequence
- Homology sequence selection
 - Search the target sequence against protein databases
 - Uniprot (<http://www.uniprot.org/>)
 - NCBI-Protein Database (<http://www.ncbi.nlm.nih.gov/protein/>)
 - Method: **Blast**
 - Pick protein sequences from model organisms
 - High similarity
 - Diversity
- Alignment programs
 - ClustalW/Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>)
 - Tcoffee (<http://tcoffee.crg.cat/apps/tcoffee/do:regular>)

Phylogenetic tree

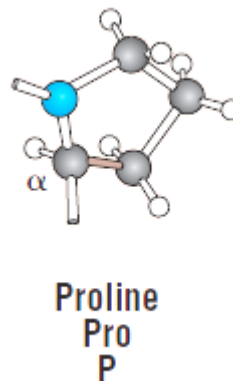
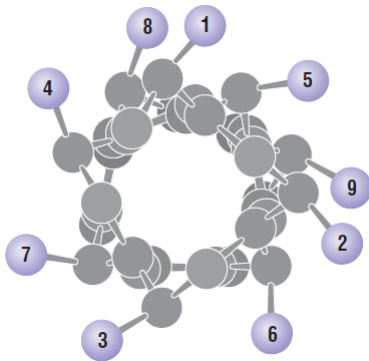


Secondary structure analysis

- Secondary structure assignment
 - Precondition: known 3D structures
 - Program: DSSP
 - Roles
 - Structural genomics
 - Sequence search
 - Structure prediction
- **Secondary structure prediction**
 - Precondition: known protein sequences
 - Programs
 - **JPred** (<http://www.compbio.dundee.ac.uk/jpred/>)
 - PSIPRED
 - ...

Prediction principles

- General preferences
 - Helix: long side chains
 - Strand: branched side chains (at the C_β)
 - Turn: Gly, Pro



- Exception
 - Transmembrane regions
 - Helix: ~20 hydrophobic residues
 - Strand: β -barrel

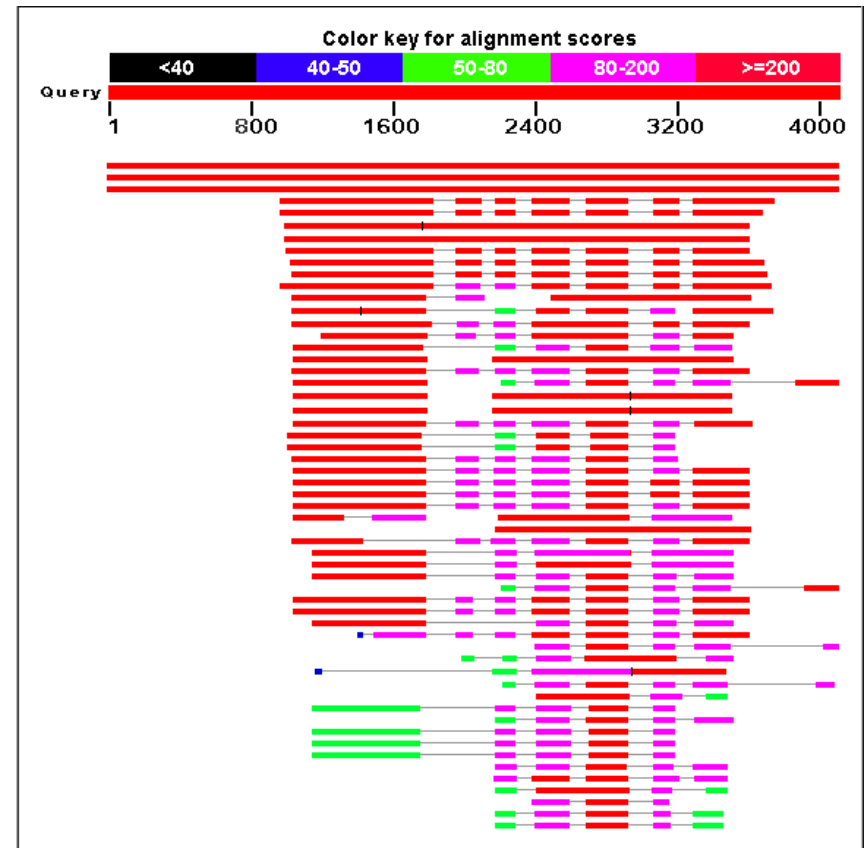
Conformational Preferences of the Amino Acids

Amino acid	Preference		
	α -helix	β -strand	Reverse turn
Glu	1.59	0.52	1.01
Ala	1.41	0.72	0.82
Leu	1.34	1.22	0.57
Met	1.30	1.14	0.52
Gln	1.27	0.98	0.84
Lys	1.23	0.69	1.07
Arg	1.21	0.84	0.90
His	1.05	0.80	0.81
Val	0.90	1.87	0.41
Ile	1.09	1.67	0.47
Tyr	0.74	1.45	0.76
Cys	0.66	1.40	0.54
Trp	1.02	1.35	0.65
Phe	1.16	1.33	0.59
Thr	0.76	1.17	0.90
Gly	0.43	0.58	1.77
Asn	0.76	0.48	1.34
Pro	0.34	0.31	1.32
Ser	0.57	0.96	1.22
Asp	0.99	0.39	1.24

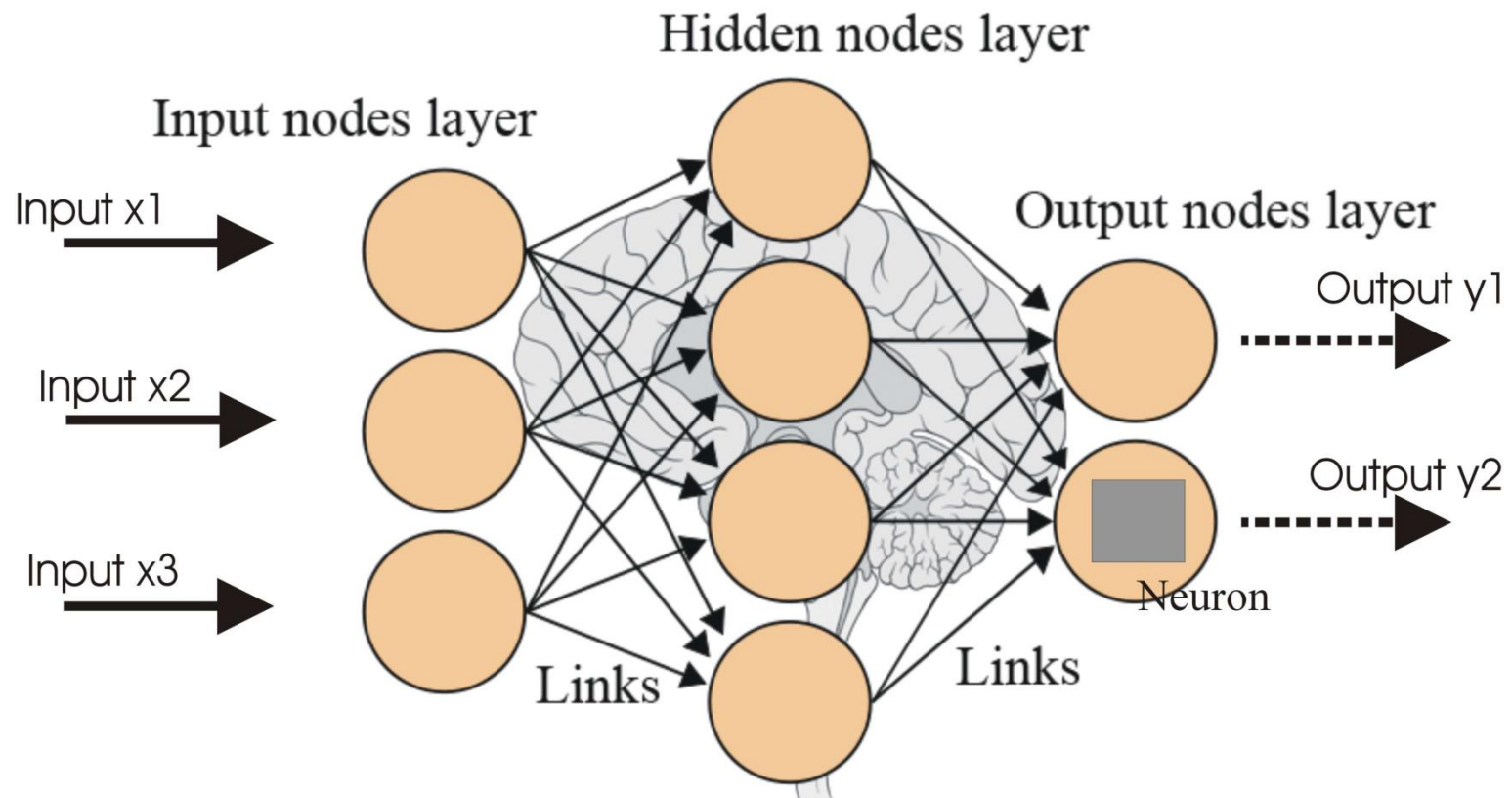
Improved methods with sequence alignment

- PSI-BLAST

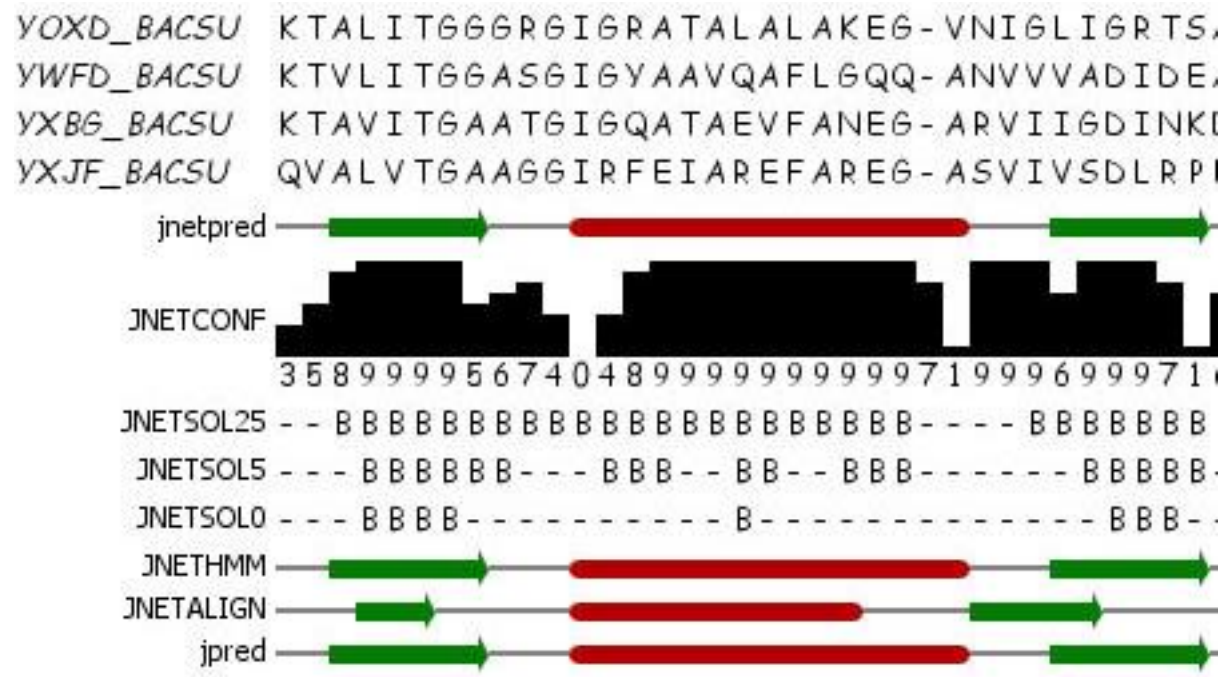
1. Comparing a target sequence to a protein database, using the gapped BLAST program
2. Constructing a multiple alignment and a profile (local alignment)
3. Comparing the profile to the protein database and seeking local alignment
4. Estimating the statistical significance of the local alignments found
5. Iterating by returning to step 2.



Improved methods with artificial neural network



An example of secondary structure prediction by JPred



Tasks

1. Select at least 10 homology sequences of **human UNC5b**
 - Including both orthologues and paralogues
2. Align the selected sequences using ClustalW or Tcoffee
 - Which one give a better result?
3. Generate a phylogenetic tree based on the alignment
 - Can you tell which type dominates, orthologue or paralogue?
4. Perform a secondary structure prediction using JPred
 - Target: the cytoplasmic domain of human UNC5b
5. Compare the predicted result with the X-ray structure (PDB 3G5B)
 - Accuracy calculation
6. Design a truncation construct with minimal size, whereas retaining the overall fold
 - Based on the sequence alignment and secondary structure prediction

Lab report format

- Title
- Your name and student No.
- Introduction
- Methods
- Results
- Conclusions