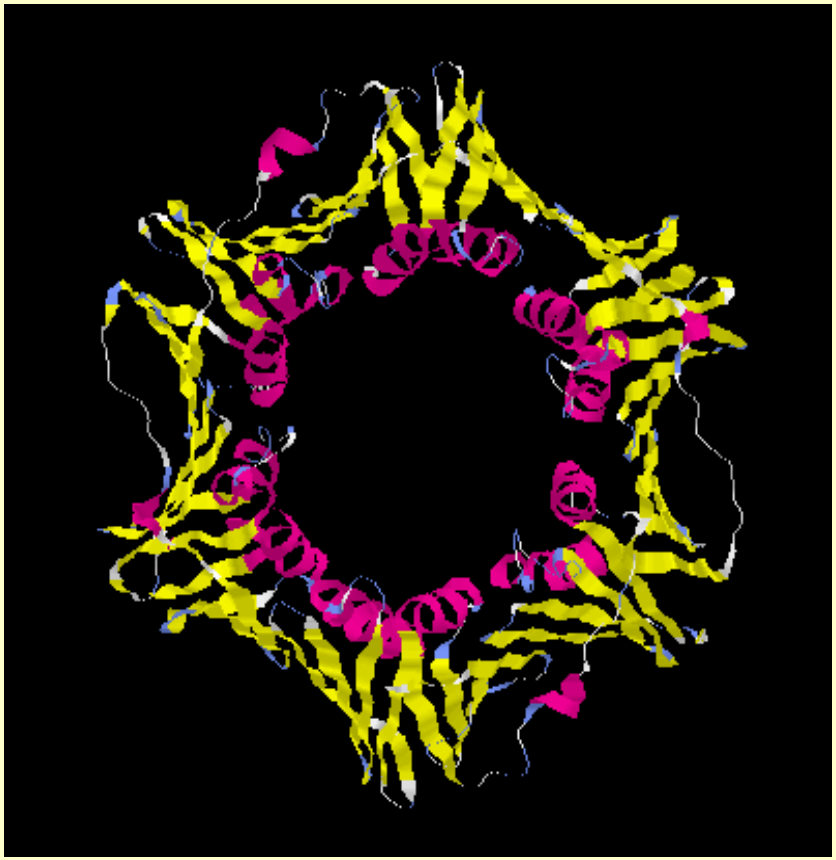
Computational Molecular Biology

Biochem 218 – BioMedical Informatics 231

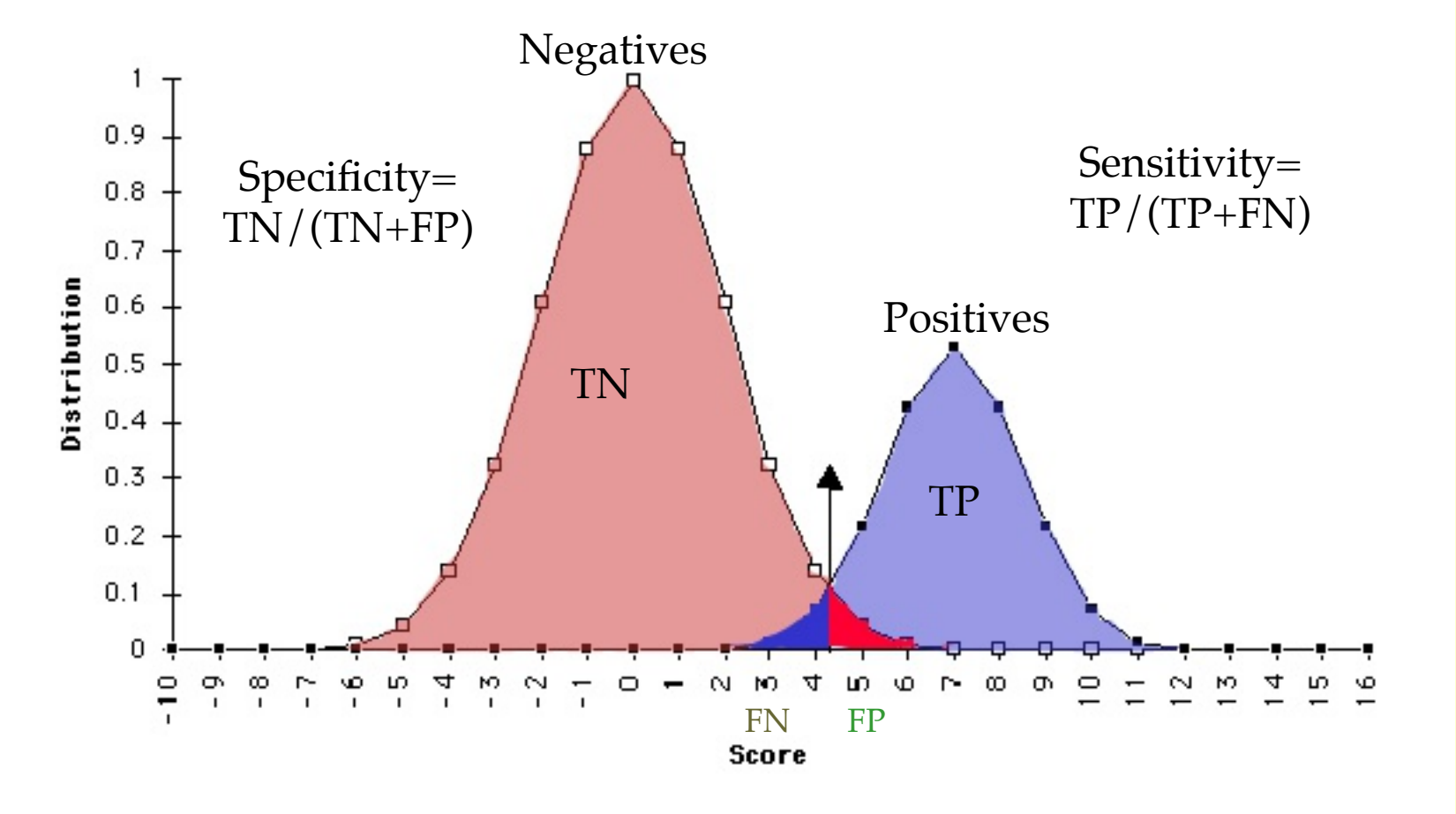
http://biochem218.stanford.edu/

Multiple Sequence Alignment

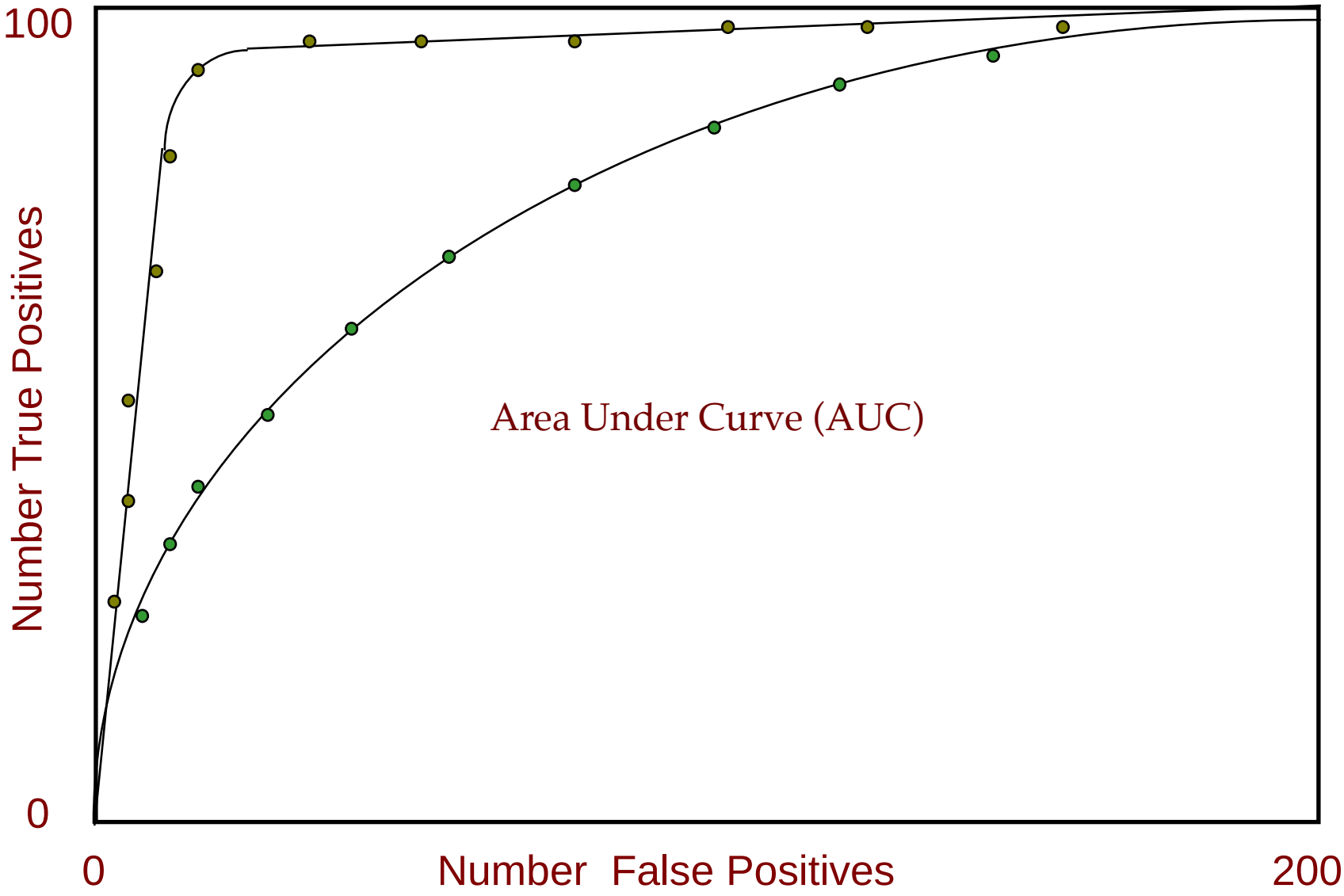


Doug Brutlag
Professor Emeritus
Biochemistry & Medicine (by courtesy)

Evaluation of Search Algorithms



Evaluation of Search Algorithms with Receiver-Operator Characteristic Curve



Pyruvate Dehydrogenase E1 Family (EC 1.2.4.1)

http://uniprot.org/

UniProt › UniProtKB — Downloads · Contact · Documentation/Help

Search | Blast | Align | Retrieve | ID Mapping *

Search in: Protein Knowledgebase (UniProtKB)

Query: ec:1.2.4.1 AND reviewed:yes AND name:alpha

Search | Clear | Fields »

1 - 25 of **47** results for **ec:1.2.4.1** ⊠ **AND reviewed:yes** ⊠ **AND name:alpha** ⊠ in **UniProtKB** sorted by **score** descending ⊠

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% | Customize display — Download...

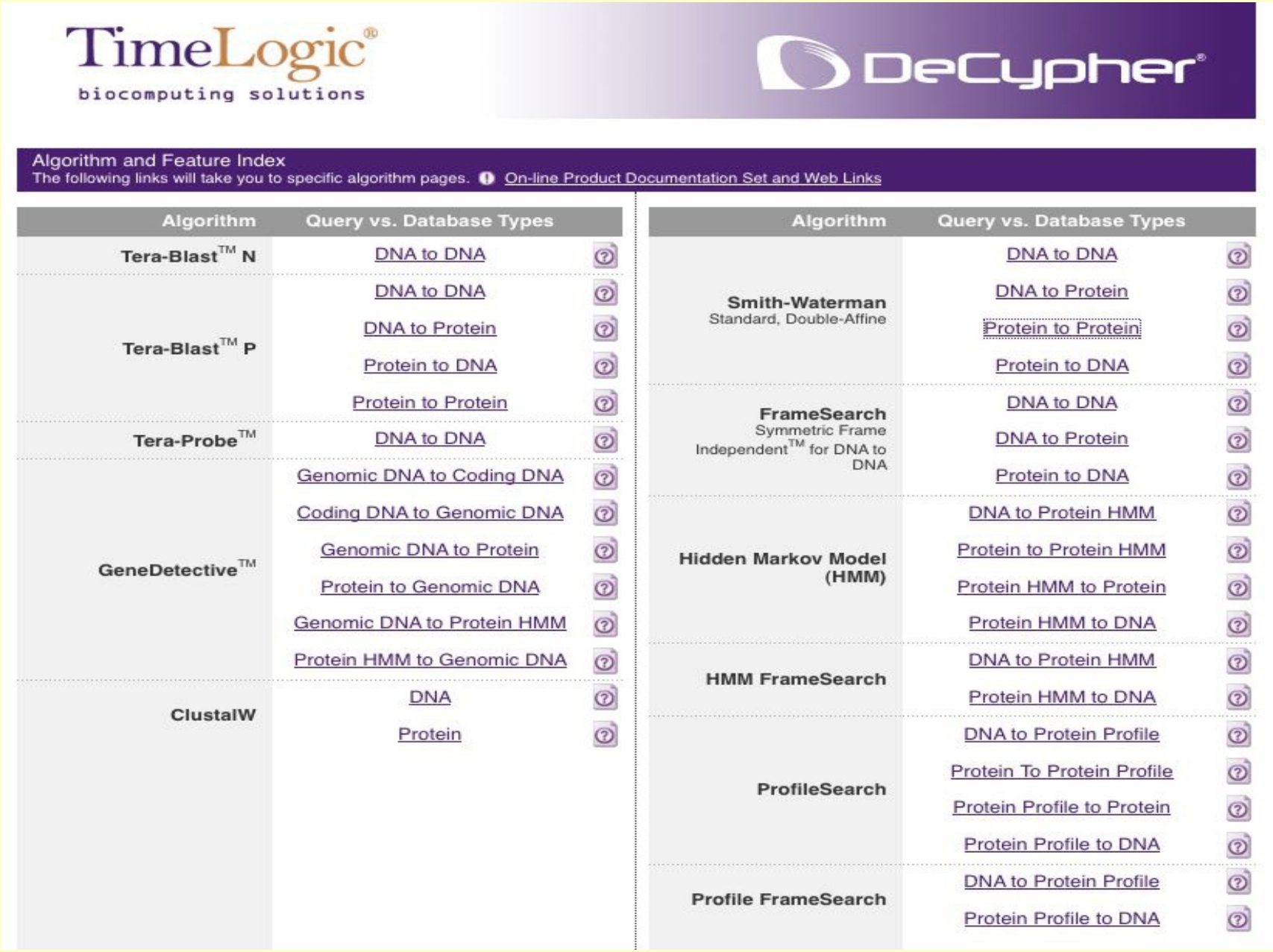
Page 1 of 2 | Next »

	Accession	Entry name	Status	Protein names	Gene names	Organism	Length
☐	P16387	ODPA_YEAST	★	**Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial** (EC 1.2.4.1) (Pyruvate dehydrogenase complex component E1 alpha) (PDHE1-A)	**PDA1** (YER178W)	Saccharomyces cerevisiae (Baker's yeast)	420
☐	P08559	ODPA_HUMAN	★	**Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial** (EC 1.2.4.1) (PDHE1-A type I)	**PDHA1** (PHE1A)	Homo sapiens (Human)	390
☐	P52901	ODPA1_ARATH	★	**Pyruvate dehydrogenase E1 component subunit alpha-1, mitochondrial** (PDHE1-A) (EC 1.2.4.1)	At1g59900 (F23H11.21)	Arabidopsis thaliana (Mouse-ear cress)	389
☐	Q8H1Y0	ODPA2_ARATH	★	**Pyruvate dehydrogenase E1 component subunit alpha-2, mitochondrial** (PDHE1-A) (EC 1.2.4.1)	**IAR4** (At1g24180) (F3I6.11)	Arabidopsis thaliana (Mouse-ear cress)	393
☐	P29803	ODPAT_HUMAN	★	**Pyruvate dehydrogenase E1 component subunit alpha, testis-specific form, mitochondrial** (EC 1.2.4.1) (PDHE1-A type II)	**PDHA2** (PDHAL)	Homo sapiens (Human)	388
☐	P21873	ODPA_BACST	★	**Pyruvate dehydrogenase E1 component subunit alpha** (EC 1.2.4.1)	**pdhA**	Bacillus stearothermophilus (Geobacillus stearothermophilus)	369
☐	P35486	ODPA_MOUSE	★	**Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial** (EC 1.2.4.1) (PDHE1-A type I)	**Pdha1** (Pdha-1)	Mus musculus (Mouse)	390
☐	P26284	ODPA_RAT	★	**Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial** (EC 1.2.4.1) (PDHE1-A type I)	**Pdha1**	Rattus norvegicus (Rat)	390

TimeLogic®
biocomputing solutions

Decypher Home Page

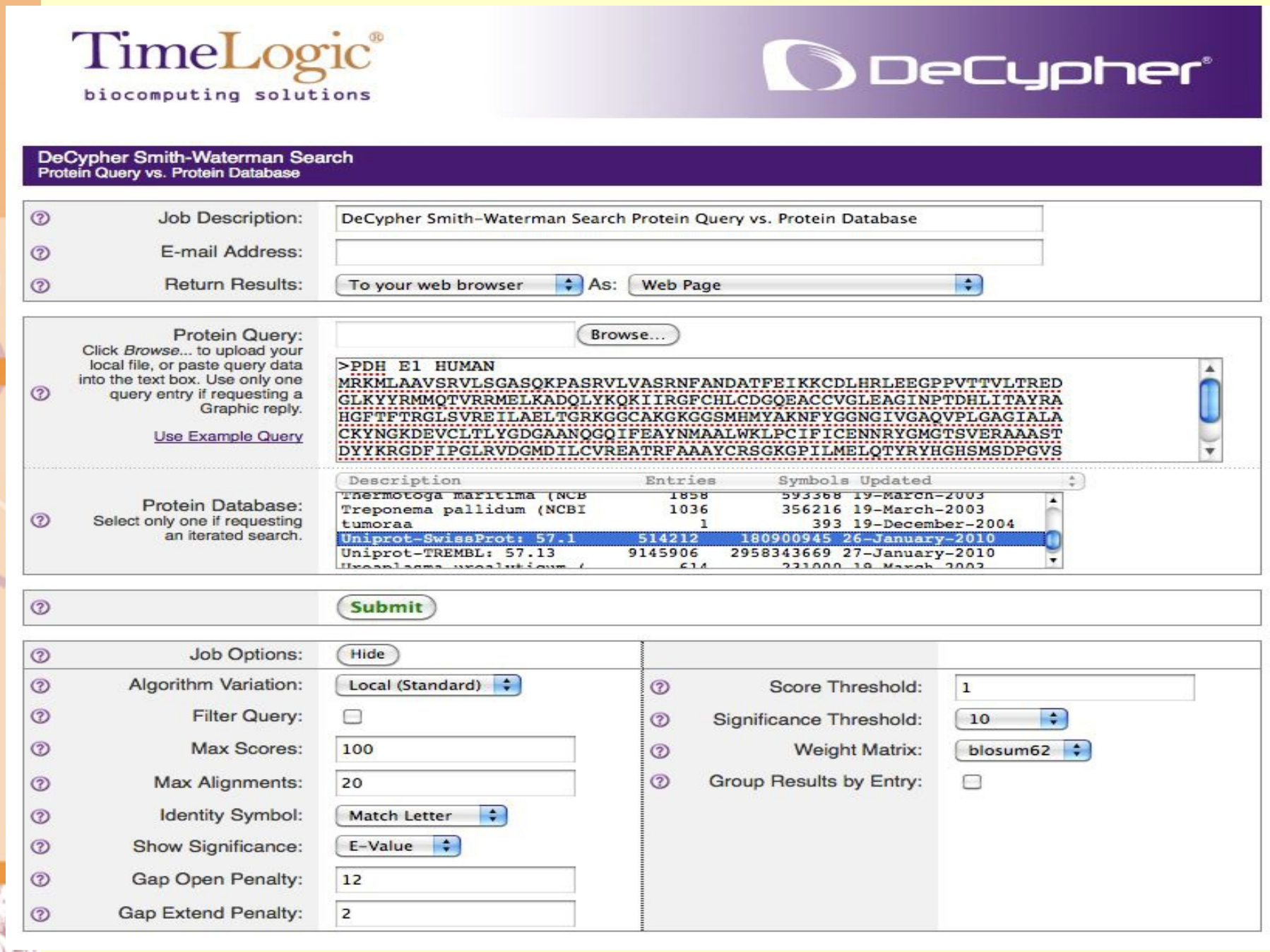
http://decypher.stanford.edu/



TimeLogic®
biocomputing solutions

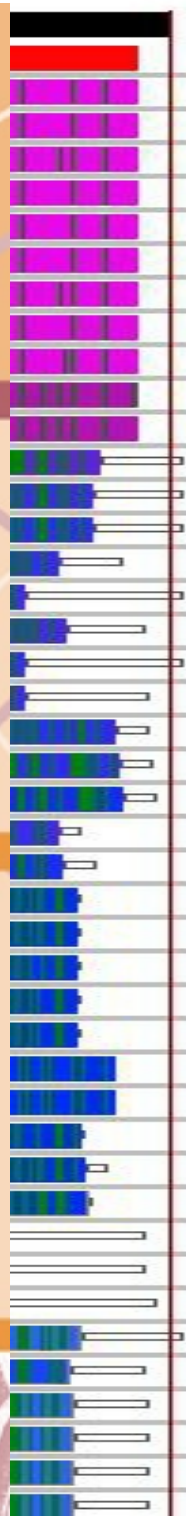
Decypher Search Input

http://decypher.stanford.edu/



r Browsers, Right Click on Hyperlink and Select "Save Target As...")

Smith-Waterman Search Protein Query vs. Protein Data



sp|P29803|ODPAT_HUMAN	388	sp|P29803|ODPAT_HUMAN Pyruvate dehydrogenase
5.4e-272	sp|P29803|ODPAT_HUMAN	Pyruvate dehydrogenase E1 component subunit
8.7e-234	sp|P35486|ODPA_MOUSE	Pyruvate dehydrogenase E1 component subunit
2.3e-233	sp|P08559|ODPA_HUMAN	Pyruvate dehydrogenase E1 component subunit
3.1e-233	sp|A7MB35|ODPA_BOVIN	Pyruvate dehydrogenase E1 component subunit
1.1e-232	sp|P26284|ODPA_RAT	Pyruvate dehydrogenase E1 component subunit a
2.1e-232	sp|A5A6L0|ODPA_PANTR	Pyruvate dehydrogenase E1 component subunit
2.1e-232	sp|Q5R490|ODPA_PONAB	Pyruvate dehydrogenase E1 component subunit
2.9e-232	sp|P29804|ODPA_PIG	Pyruvate dehydrogenase E1 component subunit a
9.4e-231	sp|Q8HXW9|ODPA_MACFA	Pyruvate dehydrogenase E1 component subunit
8.7e-218	sp|P52900|ODPA_SMIMA	Pyruvate dehydrogenase E1 component subunit
1.4e-211	sp|P35487|ODPAT_MOUSE	Pyruvate dehydrogenase E1 component subunit
9.4e-211	sp|Q06437|ODPAT_RAT	Pyruvate dehydrogenase E1 component subunit
8.7e-146	sp|P52899|ODPA_CAEEL	Probable pyruvate dehydrogenase E1 componen
3.6e-135	sp|P26268|ODPT_ASCSU	Pyruvate dehydrogenase E1 component subunit
5.9e-133	sp|P26267|ODPA_ASCSU	Pyruvate dehydrogenase E1 component subunit
5.0e-127	sp|Q54C70|ODPA_DICDI	Pyruvate dehydrogenase E1 component subunit
1.0e-119	sp|P16387|ODPA_YEAST	Pyruvate dehydrogenase E1 component subunit
1.5e-116	sp|P52901|ODPA1_ARATH	Pyruvate dehydrogenase E1 component subunit
5.4e-116	sp|O13366|ODPA_KLULA	Pyruvate dehydrogenase E1 component subunit
1.4e-115	sp|Q10489|ODPA_SCHPO	Pyruvate dehydrogenase E1 component subunit
5.0e-115	sp|Q8H1Y0|ODPA2_ARATH	Pyruvate dehydrogenase E1 component subunit
1.8e-114	sp|P52902|ODPA_PEA	Pyruvate dehydrogenase E1 component subunit a
3.1e-113	sp|P52903|ODPA_SOLTU	Pyruvate dehydrogenase E1 component subunit
1.9e-111	sp|Q9R9N5|ODPA_RHIME	Pyruvate dehydrogenase E1 component subunit
6.9e-091	sp|O66112|ODPA_ZYMMO	Pyruvate dehydrogenase E1 component subunit
5.4e-088	sp|Q4UKQ6|ODPA_RICFE	Pyruvate dehydrogenase E1 component subunit
3.1e-085	sp|Q92IS3|ODPA_RICCN	Pyruvate dehydrogenase E1 component subunit
4.3e-085	sp|Q1RJX4|ODPA_RICBR	Pyruvate dehydrogenase E1 component subunit
1.1e-084	sp|Q68XA9|ODPA_RICTY	Pyruvate dehydrogenase E1 component subunit
7.5e-084	sp|Q92DR4|ODPA_RICPR	Pyruvate dehydrogenase E1 component subunit
8.1e-069	sp|Q1XDM0|ODPA_PORYE	Pyruvate dehydrogenase E1 component subunit
6.4e-066	sp|P51267|ODPA_PORPU	Pyruvate dehydrogenase E1 component subunit
3.7e-063	sp|P27745|ACOA_RALEH	Acetoin:26-dichlorophenolindophenol oxidore
4.3e-061	sp|O31404|ACOA_BACSU	Acetoin:26-dichlorophenolindophenol oxidore
1.0e-035	sp|P37940|ODBA_BACSU	2-oxoisovalerate dehydrogenase subunit alph
2.3e-033	sp|Q72GU1|ODBA_THET2	2-oxoisovalerate dehydrogenase subunit alph
8.1e-033	sp|Q5SLR4|ODBA_THET8	2-oxoisovalerate dehydrogenase subunit alph
1.5e-032	sp|P35485|ODPA_ACHLA	Pyruvate dehydrogenase E1 component subunit
5.4e-032	sp|Q54M22|ODBA_DICDI	2-oxoisovalerate dehydrogenase subunit alph
1.0e-031	sp|P47516|ODPA_MYCGE	Pyruvate dehydrogenase E1 component subunit
1.9e-031	sp|Q5HG21|ODPA_STAAC	Pyruvate dehydrogenase E1 component subunit
1.9e-031	sp|P60089|ODPA_STAAM	Pyruvate dehydrogenase E1 component subunit
1.9e-031	sp|Q820A6|ODPA_STAAN	Pyruvate dehydrogenase E1 component subunit
1.9e-031	sp|Q6GAC1|ODPA_STAAS	Pyruvate dehydrogenase E1 component subunit

General DNA Similarity Search Principles

- Search both Strands
- Translate ORFs and cDNAs
- Use most sensitive search possible
  - UnGapped BLAST for infinite gap penalty (PCR & CHIP oligos)
  - Gapped BLAST for most searches
  - Smith Waterman or megaBLAST or discontinuous MegaBLAST for cDNA/genome comparisons
  - cDNA =>Zero gap-length penalty
  - Consider using transition matrices
  - Ensure that expected value of score is negative
- Examine results with exp. between 0.05 and 10
- Reevaluate results of borderline significance using limited query

General Protein Similarity Search Principles

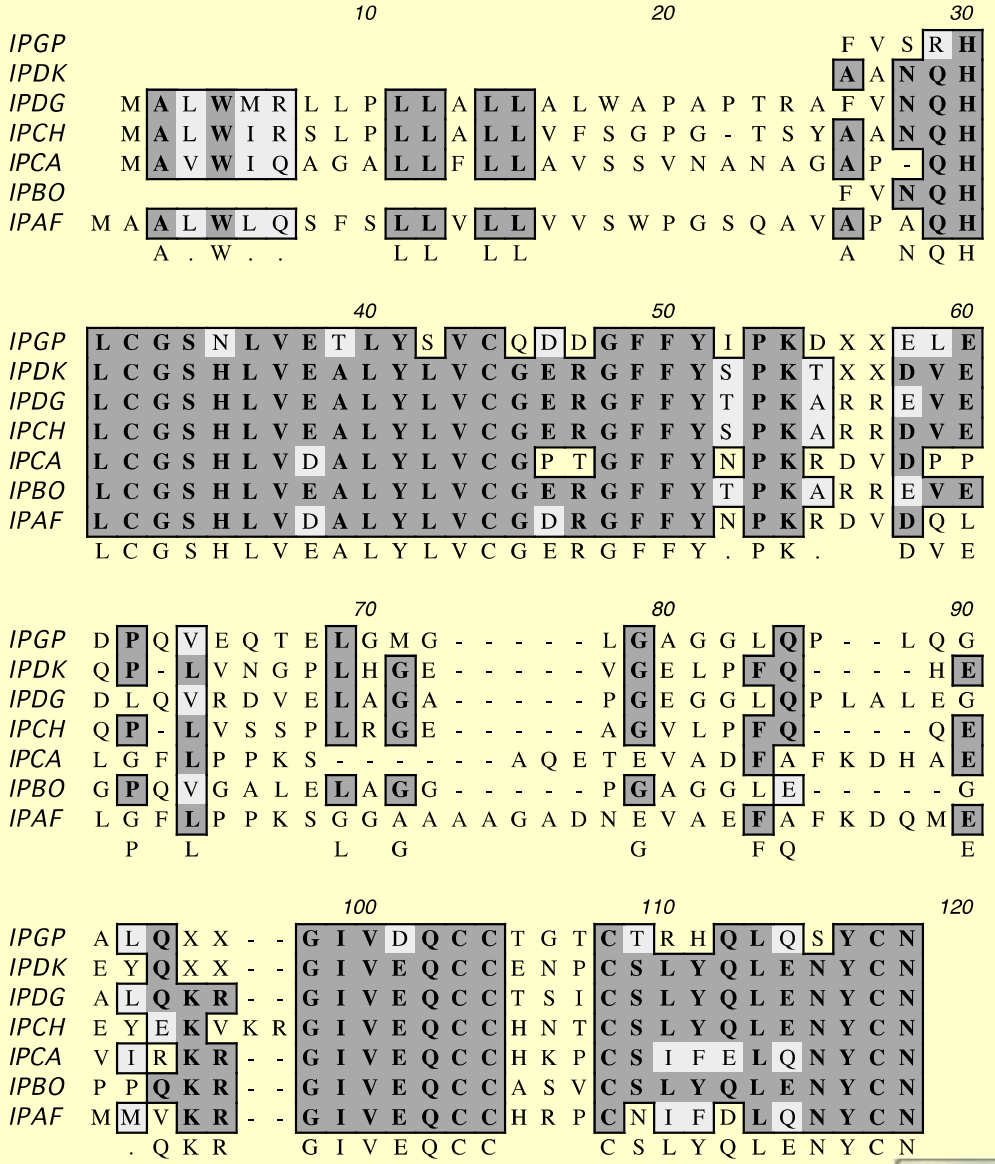
- Chose between local or global search algorithm
- Use most sensitive search algorithm available
  - Original BLAST for no gaps
  - Smith-Waterman for most flexibility
  - Gapped BLAST for well delimited regions
  - PSI-BLAST for families
  - Initially BLOSUM62 and default gap penalties
  - If no significant results, use BLOSUM30 and lower gap penalties
- Examine results between exp. 0.05 and 10 for biological significance
- Beware of long hits or those with unusual amino acid composition
- Reevaluate results of borderline significance using limited query

Goals of Multiple Sequence Alignment

- Determine Consensus Sequences
  - Prosite Patterns
- Building Gene Families
  - InterPro, Prints, ProDom, pFAM, DOMO, COGs, KOGs
- Develop Relationships & Phylogenies
  - Clusters, COGs, KOGs, ClusTR
  - Relationships
  - Evolutionary Models
  - UPGMA, Neighbor Joining, Phylip, GrowTree, PAUP
- Model Protein Structures for Threading and Fold Prediction
  - Profiles, Templates, HSSP, FSSP, SwissModel
  - Hidden Markov Models, pFAM, SAM, SuperFamily
  - Network Models, Neural Nets, Bayesian Networks
  - Statistical Models, Generalized Linear Models

Consensus Sequence From a Multiple Sequence Alignment

ClustalW Insulin Alignments



Block Maker Makes a Multiple Sequence Alignment

http://blocks.fhcrc.org/blocks/blockmkr/make_blocks.html

Sequence Profiles

http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=hmmerpfam

SeqWeb v3.1

accelrys

Programs | Managers | Help Topics | Support

- Programs
- Comparison
- Database Searching
  - Similarity
  - Reference
- Evolution
- Mapping
- Pattern Recognition
- Primer Selection
- Protein Analysis
- Nucleic Acid Secondary Structure
- Translation
- Utilities
- Index

HmmerPfam

Identify known hidden Markov model domains.

Input sequences:

Select From: Default | Project | Local File | Clipboard | Database

Sequence	Description	Type	Length	Range
odpat_human.uniprot_sprot	odpat_human	P	388	1 .. 388
odpa_human.uniprot_sprot	odpa_human	P	390	1 .. 390
odpt_ascsu.uniprot_sprot	odpt_ascsu	P	391	1 .. 391
odpa_yeast.uniprot_sprot	odpa_yeast	P	420	1 .. 420
odpa_rat.uniprot_sprot	odpa_rat	P	390	1 .. 390
odpa_mouse.uniprot_sprot	odpa_mouse	P	390	1 .. 390
odpa_caeel.uniprot_sprot	odpa_caeel	P	397	1 .. 397
odpat_rat.uniprot_sprot	odpat_rat	P	391	1 .. 391
odpa_macfa.uniprot_sprot	odpa_macfa	P	390	1 .. 390
odpa_smima.uniprot_sprot	odpa_smima	P	363	1 .. 363
odpa_ponpy.uniprot_sprot	odpa_ponpy	P	390	1 .. 390
odpa_ascsu.uniprot_sprot	odpa_ascsu	P	396	1 .. 396
odpa_pig.uniprot_sprot	odpa_pig	P	389	1 .. 389
odpat_mouse.uniprot_sprot	odpat_mouse	P	391	1 .. 391

Refresh | Clear

Input Parameters:

HMM Database	Pfam	
Number of processors to use	1	
Turn the XNU filtering of target protein sequences on	☐	
Limit alignment output of best scoring domains to	50	*(range 1 thru 1000)*
Set E-value cutoff for the per-sequence ranked hit list	10.0	*(range 0.0 thru 100.0)*
Set bit score cutoff for the per-sequence ranked hit list	-300	*(range -500.0 thru 500.0)*
Set E-value cutoff for the per-domain ranked hit list	2.0	*(range 0.0 thru 100.0)*
Set bit score cutoff for the per-domain ranked hit list	10	*(range -1000.0 thru 1000.0)*

Run | Reset

SeqWeb v3.1

SeqWeb Sequence Profile Search

http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=hmmerpfar

ProfileSearch

Search a peptide database using a Profile constructed from unaligned peptide sequences.

Input sequences:

Select From: Default | Project | Local File | Clipboard | Database

Sequence	Description	Type	Length	Range
hba_horse.pep	ID HBA_HORSE STANDARD; PRT; 141 AA.	P	141	1 .. 141
lgba_soybn	lgba_soybn	P	143	1 .. 143
hbahuman	hbahuman	P	141	1 .. 141
glb5_petma.pep	ID GLB5_PETMA STANDARD; PRT; 149 AA.	P	149	1 .. 149
lgb1_soybn.pep	- ID LGB1_SOYBN STANDARD; PRT; 143 AA.	P	143	1 .. 143
hba_human	hba_human	P	141	1 .. 141
lgb2_luplu.pep	ID LGB2_LUPLU STANDARD; PRT; 153 AA.	P	153	1 .. 153
hbb_human	hbb_human	P	146	1 .. 146

Refresh | Clear

Input Parameters:

Search set: genpept -- Translated GenBank

Limit for the number of alignments: 15 *(range 0 thru 100000)*

The following matrix and gap weight settings apply to the alignment phase of the analysis.
Select a sequence comparision matrix. This matrix determines how matches and mismatches are scored. The default penalites for gap creation and extension are given after each matrix name.

Scoring Matrix: blosum62

Set gap creation penalty: 8

Set gap extension penalty: 2

The following gap weight settings apply to the search phase of the analysis.

Set maximum gap creation penalty:

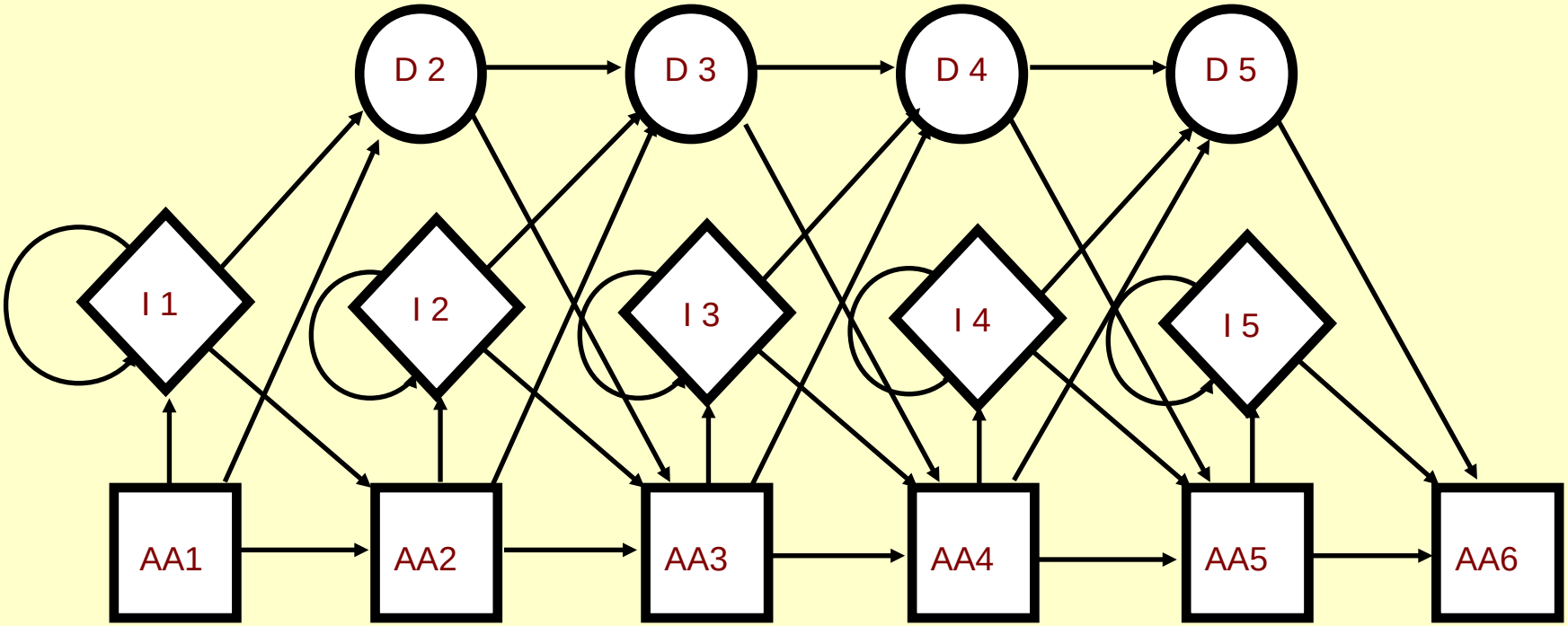
Set maximum gap extension penalty:

A Z score is the number of standard deviations above background.

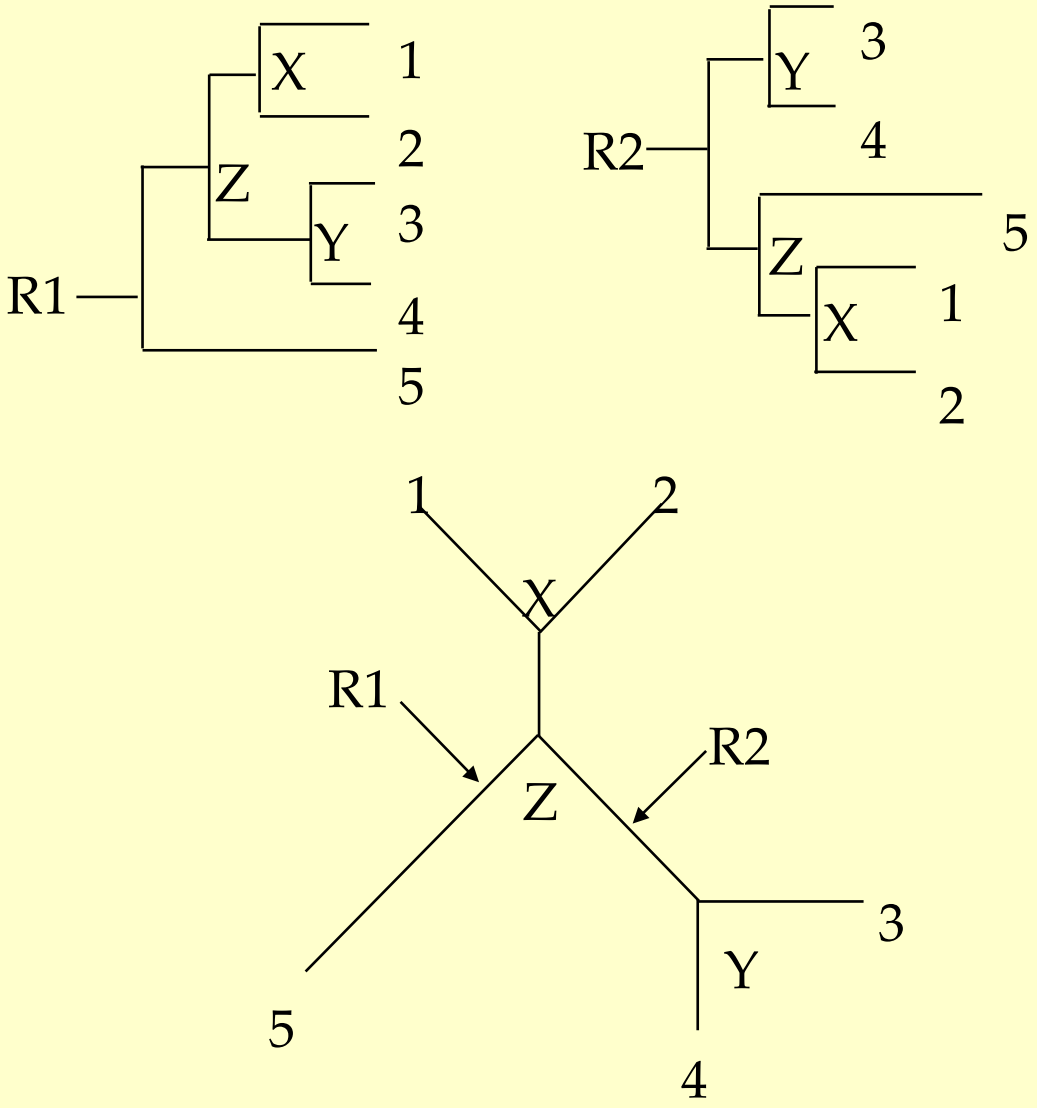
Lowest Z score to report in output list: 2.5 *(range 0.0 thru 100)*

Hidden Markov Models

http://www.cse.ucsc.edu/research/compbio/sam.html



Evolutionary Trees

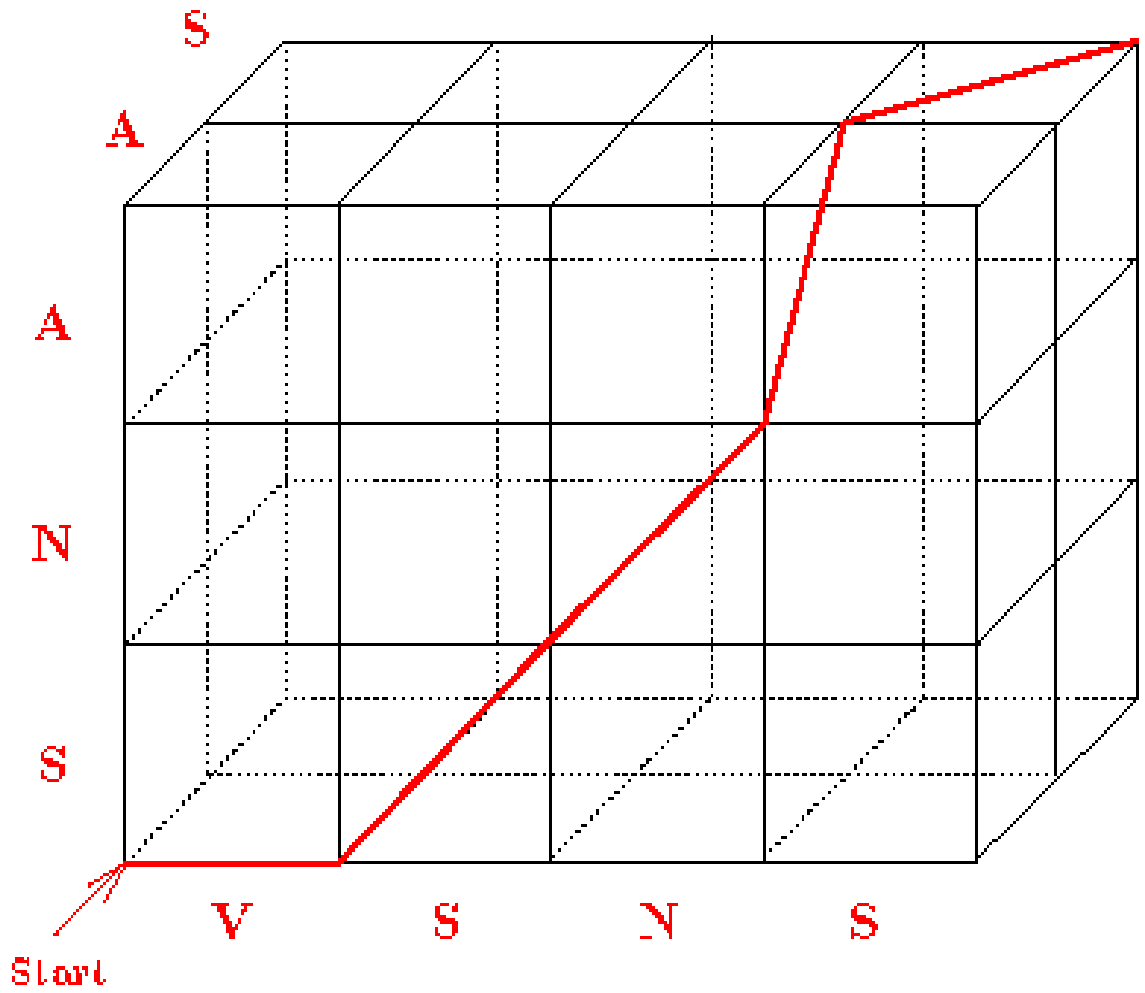


Challenges Aligning Multiple Sequences

- Computational complexity $O(n^k)$ for k sequences n long
- Space requirements $O(n^k)$ for k sequences n long
- Sequence clusters require weighting function
- Weighted alignments tend to overweight erroneous sequences
- Approximations must be used for real world data
  - Linked lists used to find exact words shared between k sequences
  - BLAST can find inexact shared words between k sequences
  - FASTA can be used to do progressive pair-wise alignments
  - HMM Pair models find best overall alignment probabilistically
- Pairwise comparisons followed by Progressive Alignments
- Final alignment is often dependent on order data presented
- Gaps make alignment unnaturally long

Three Protein Alignment
(Murata, Richardson & Sussman)

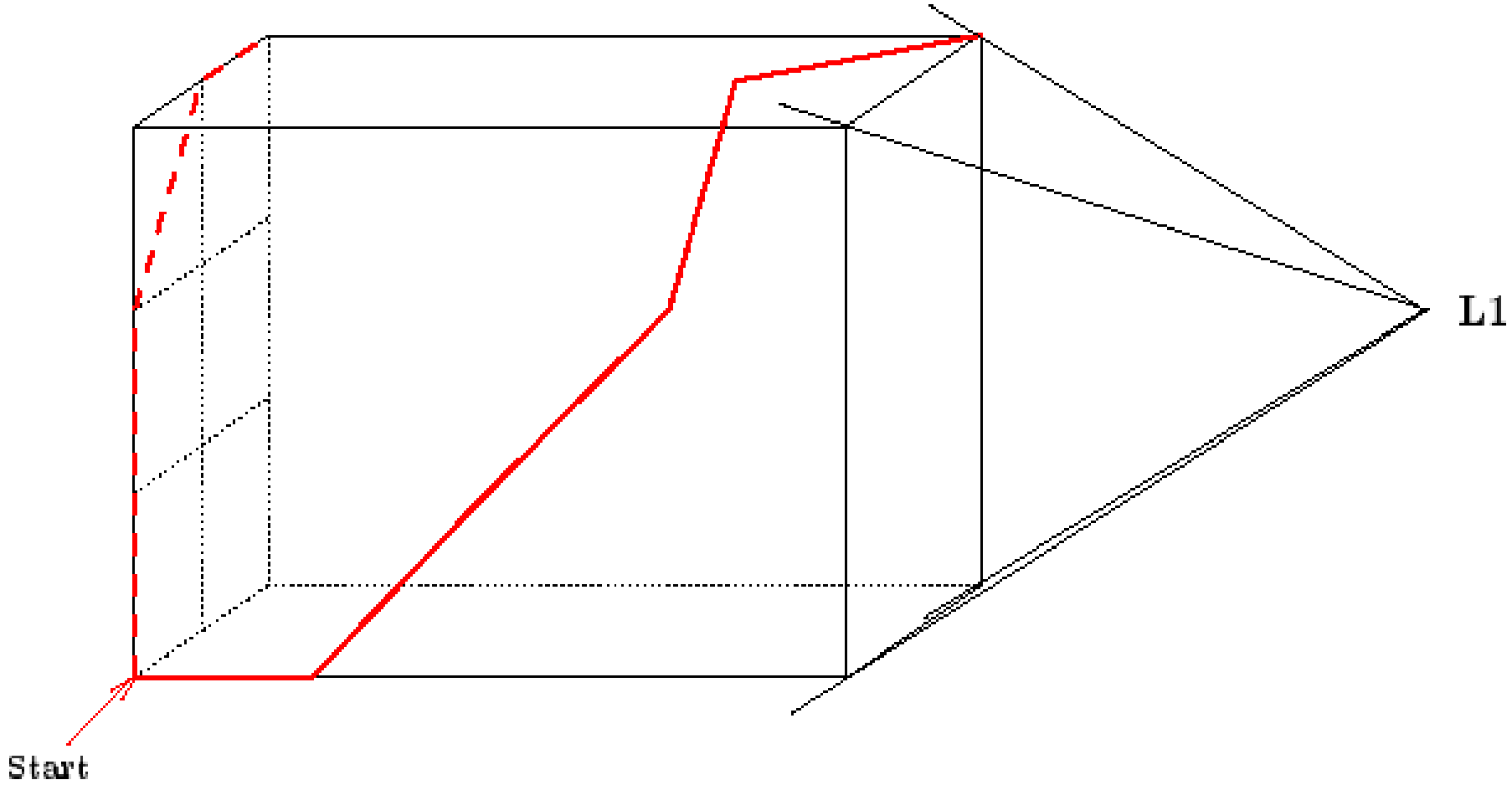
Figure 1: Alignment Path for 3 Sequences.



V S N - S
- S N A -
- - - A S

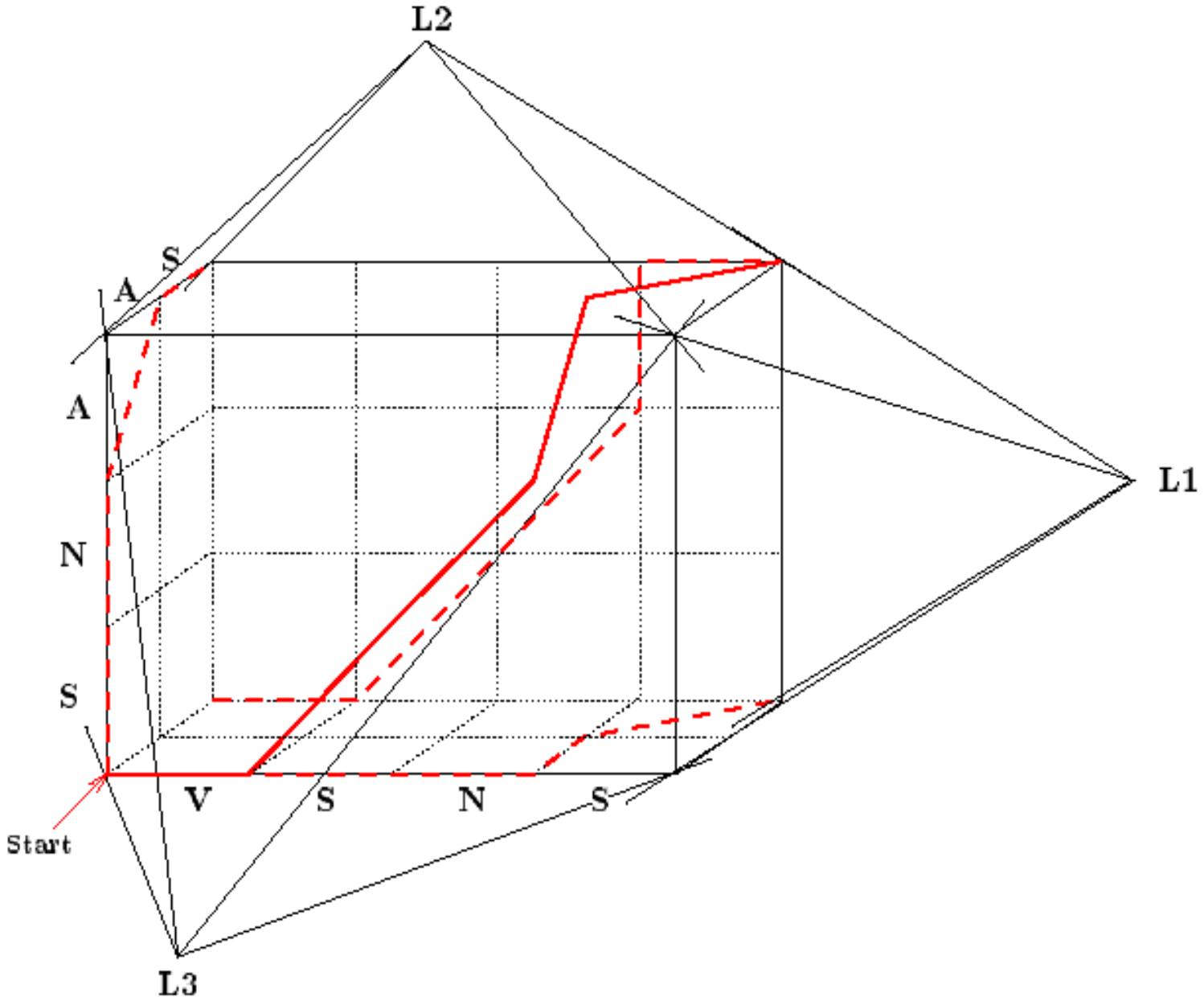
One Pairwise Alignment from the Three-Way Alignment

Figure 2: Projection of the Alignment from the Right-Hand-Side.



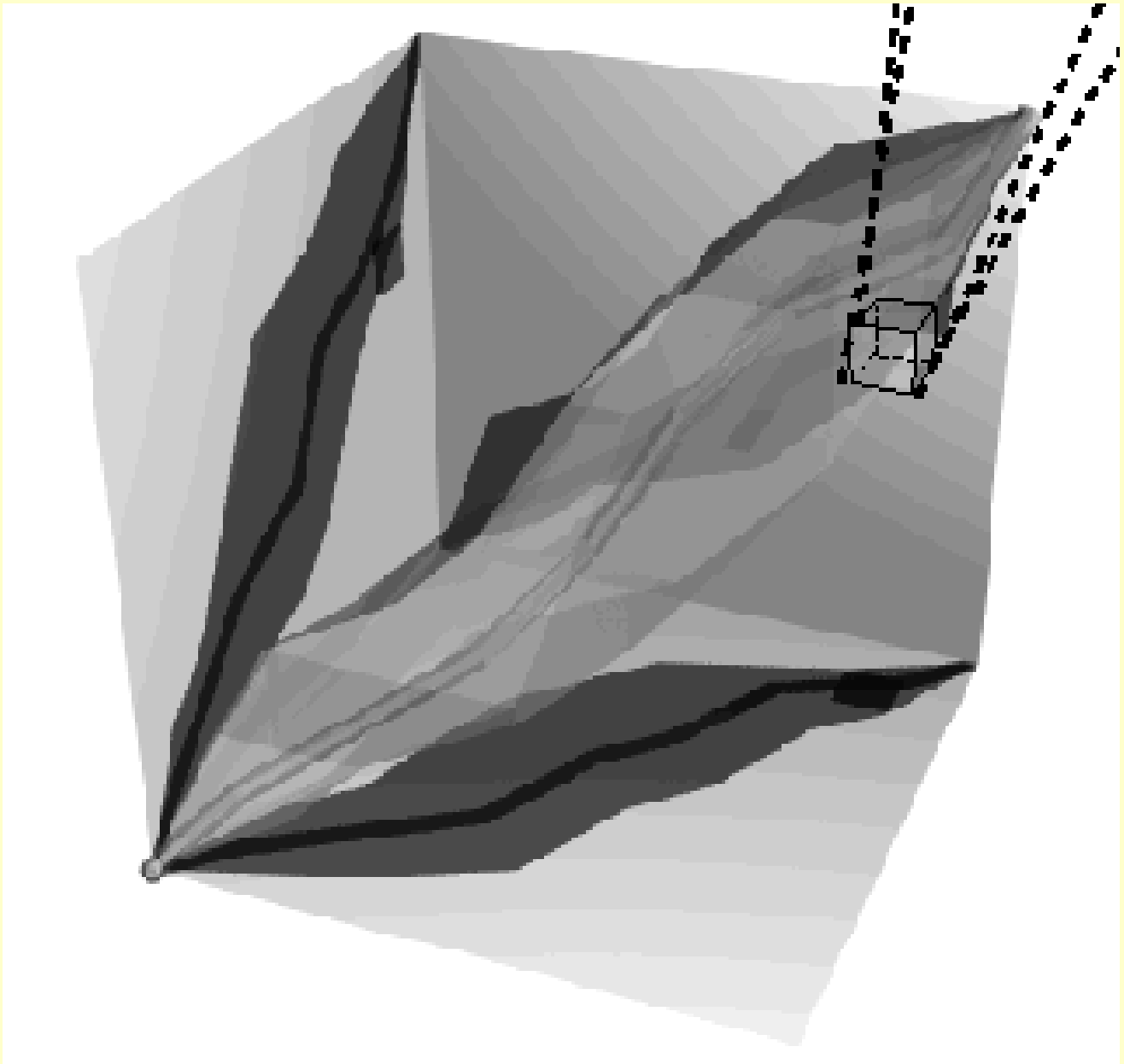
All Pairwise Alignments from the Three-Way Alignment

Figure 3: All 3 Pairwise Projections of the Alignment



Carrillo-Lipman Limits for MSA

http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html

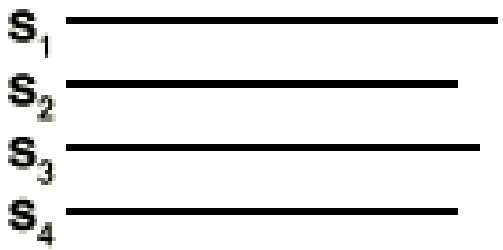


Clustal Progressive Alignment (Step 1)

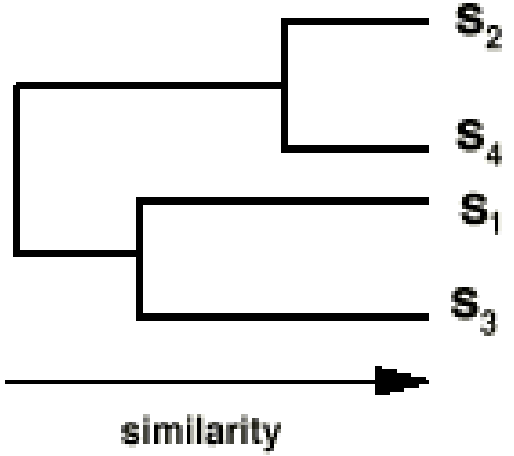
Steps in Multiple Alignment

(A) Pairwise Alignment

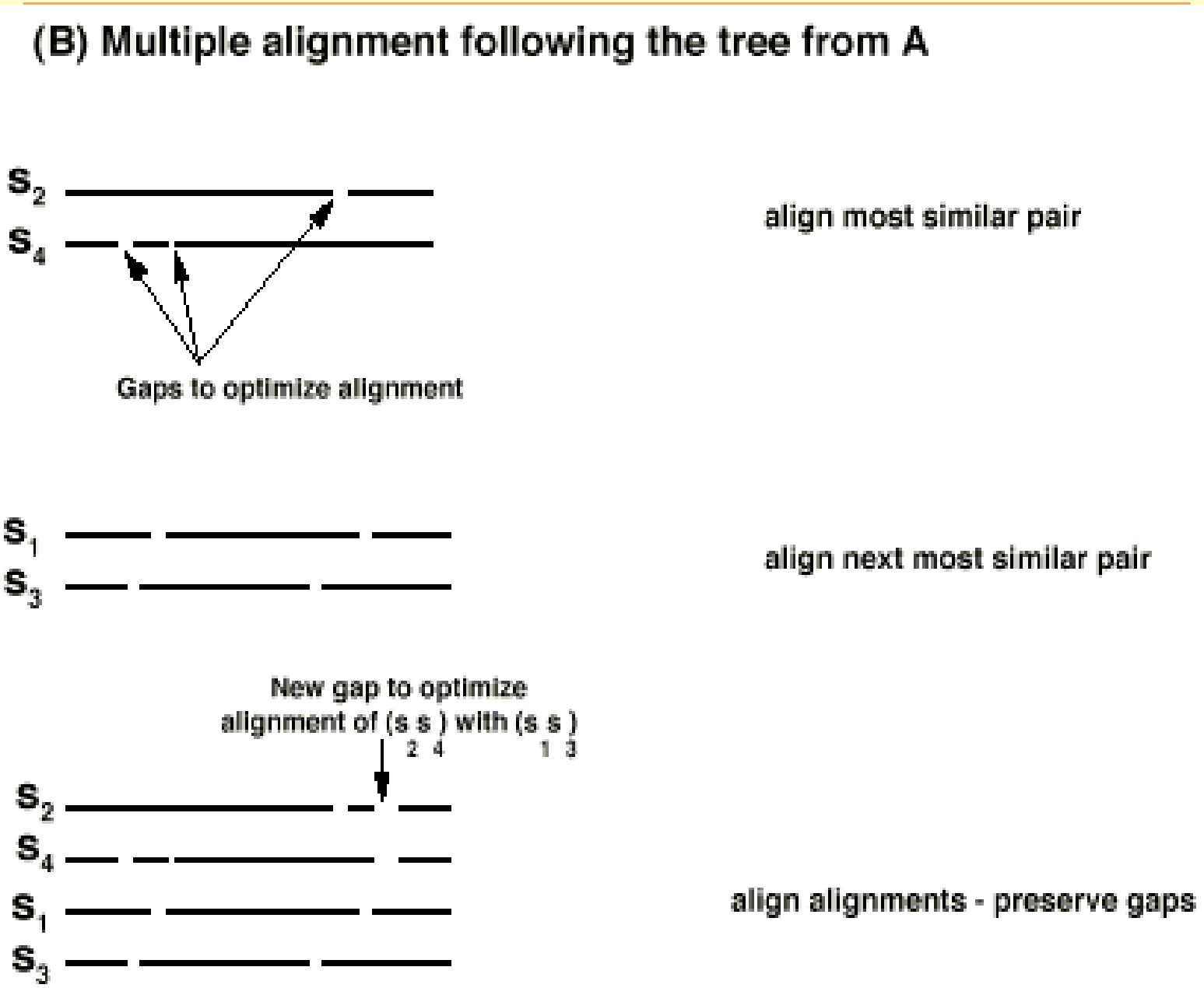
Example - 4 sequences S_1 S_2 S_3 S_4



6 pairwise comparisons
then cluster analysis



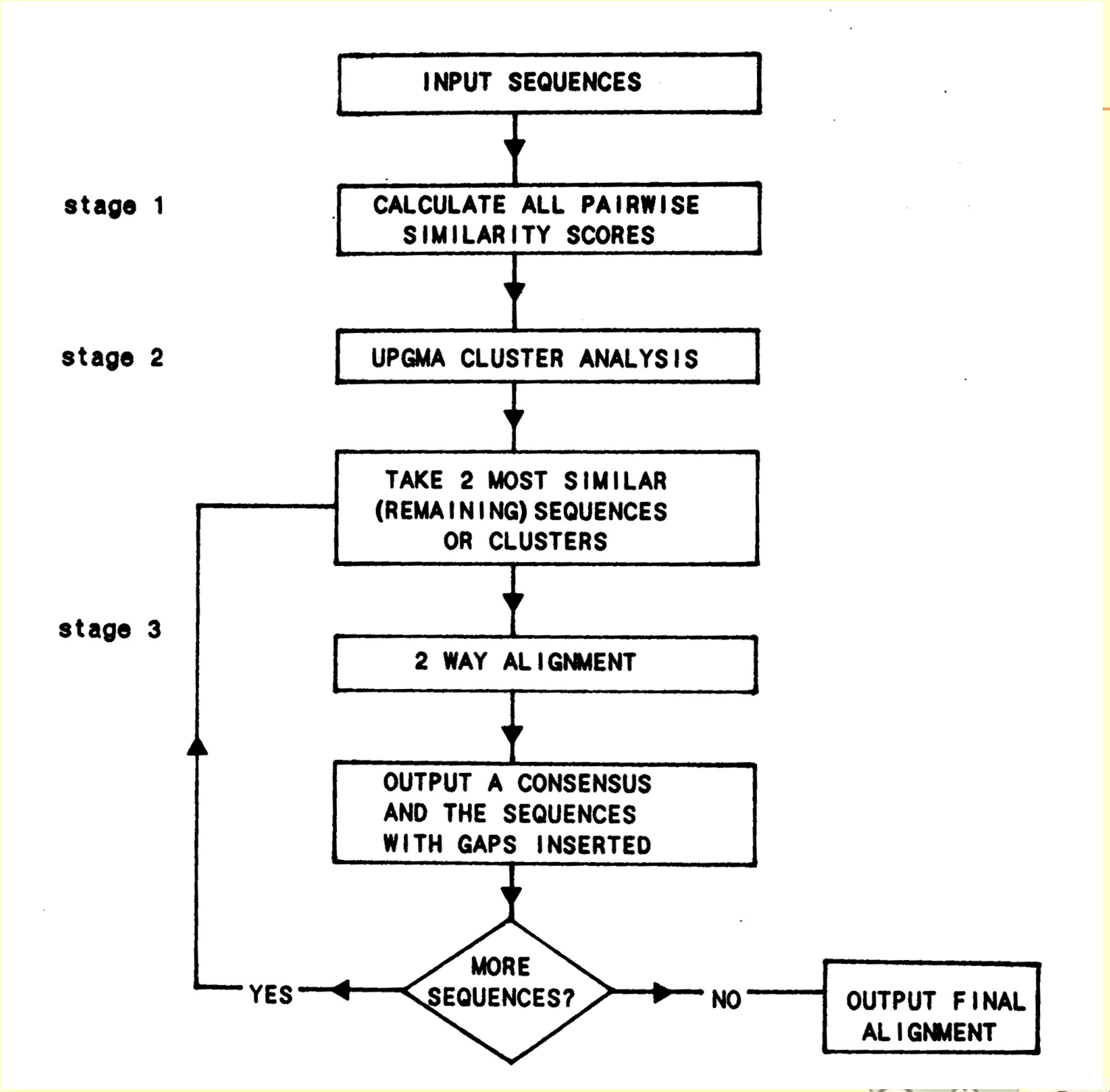
Clustal Progressive Alignment (Step 2)



Gaps Are Propagated To Make Alignment

```
SEQ2:     MQQL-DNPYIVRMIGICEAE-SWM
SEQ1:     MGQF-DHPNIIRLEGVVTKSRPVM

SEQ2:     MQQL-DNPYIVRMIGICEAE-SWM
SEQ3:     MKQL-QHPRLVRLYAVVTQE-PIY

SEQ2:     MQQL-DNPYIVRMIGICEAE-SWM
SEQ4:     MKMIGKHKNIINLLGACTQDGPLY
```

Clustal Procedure



Clustal Dendrogram

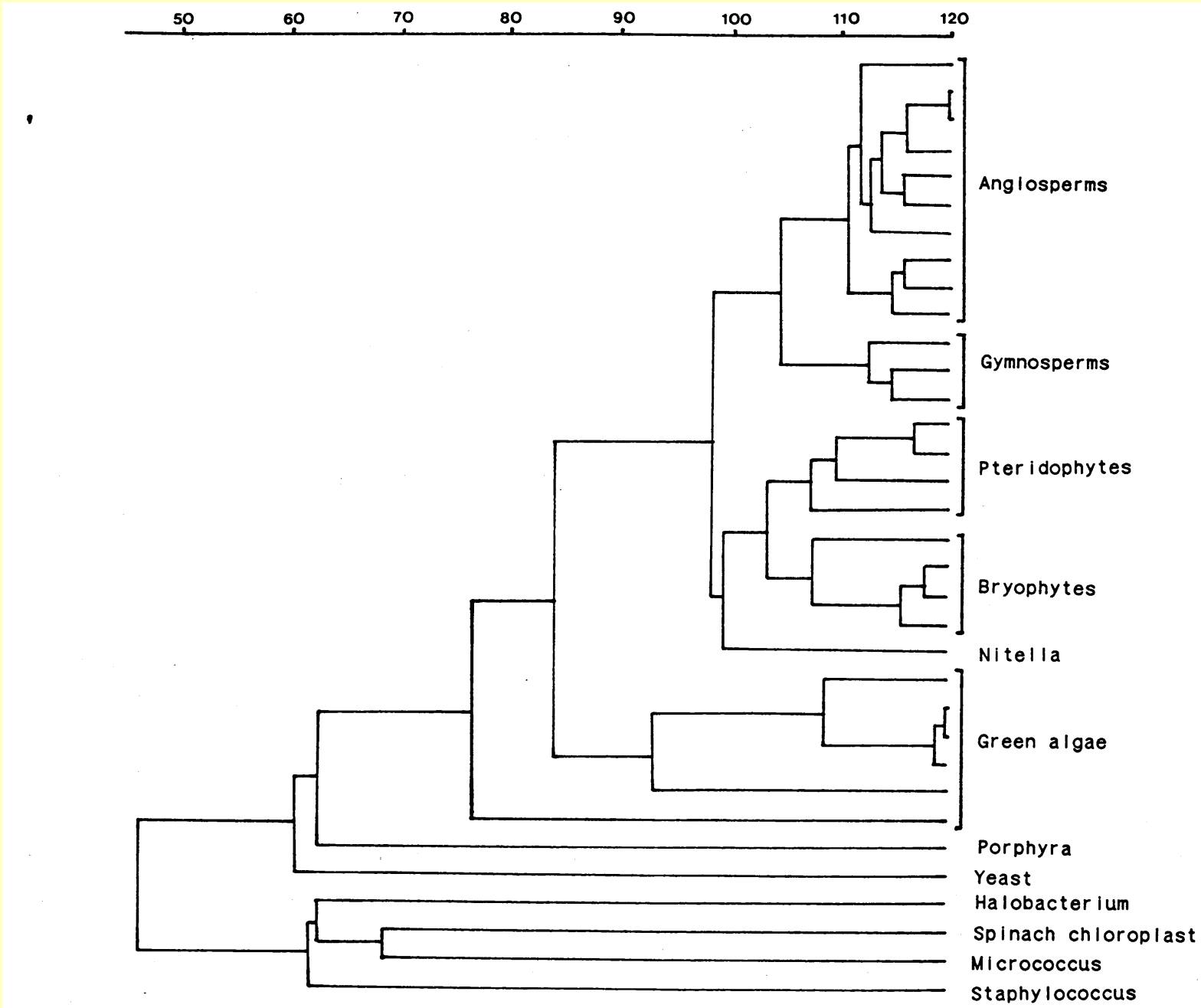


Fig. 4. UPGMA dendrogram of 34 plant, yeast and bacterial 5S RNA sequences. The sequences were taken from Hori et al. (1985) (see RESULTS, section a2). The major plant taxonomic groupings are indicated. The scale across the top margin shows the number of matching nucleotides (after alignment) between two clusters or sequences.

Clustal Globin Alignment

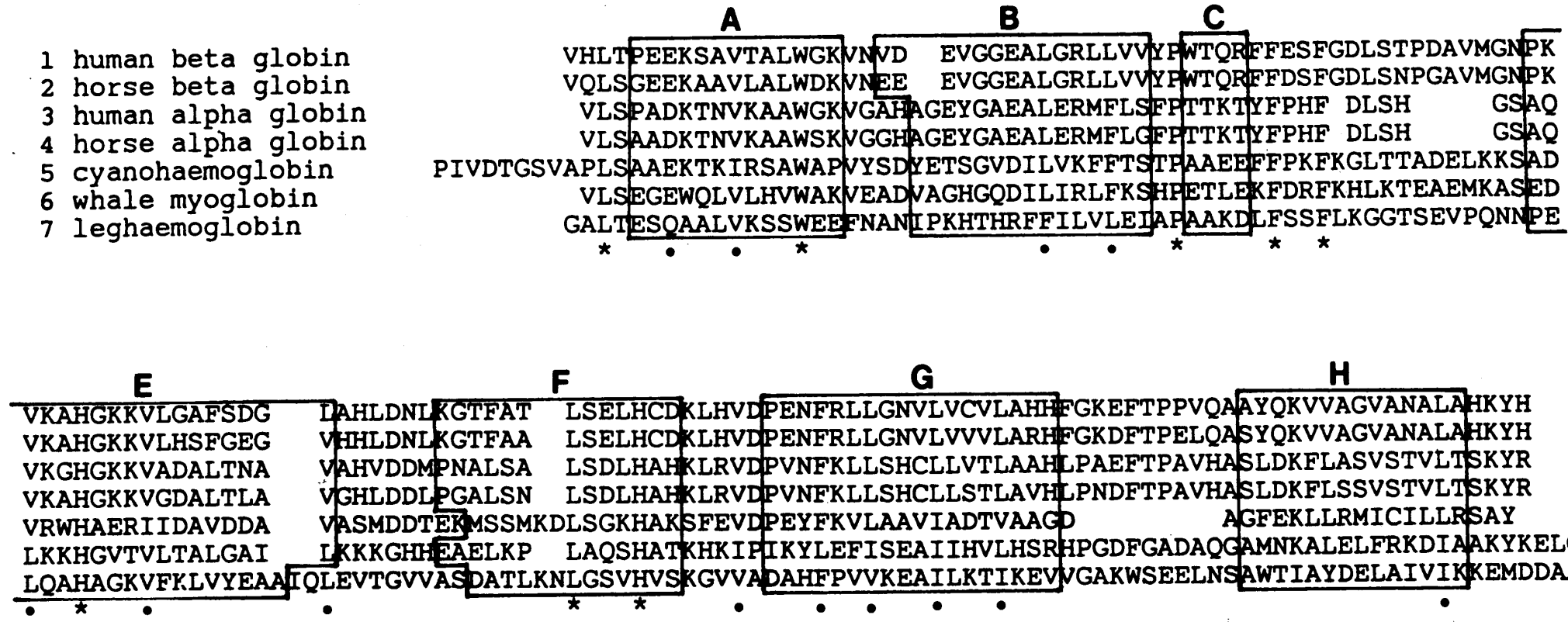


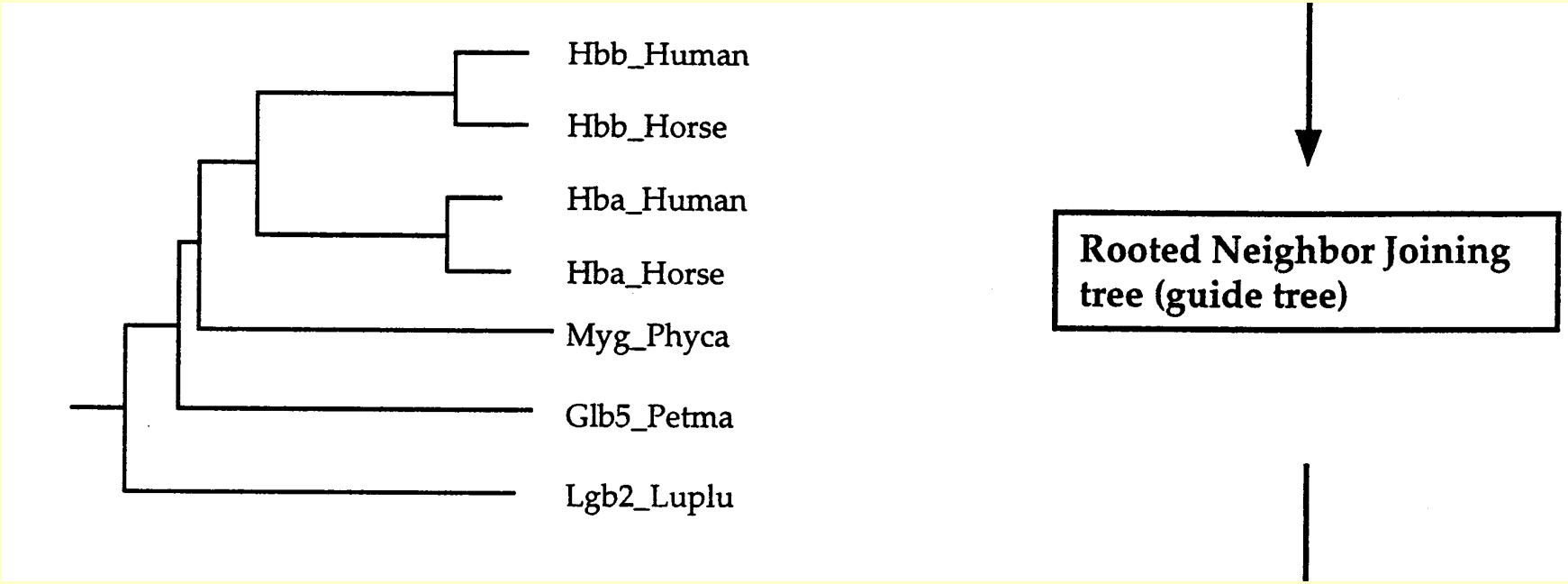
Fig. 3. CLUSTAL-produced multiple alignment of seven globin sequences taken from Lesk and Chothia (1980) (see RESULTS, section

ClustalW Step 1: BLOSUM Distance Matrix

Hbb_Human	1	-					
Hbb_Horse	2	.17	-				
Hba_Human	3	.59	.60	-			
Hba_Horse	4	.59	.59	.13	-		
Myg_Phyca	5	.77	.77	.75	.75	-	
Glb5_Petma	6	.81	.82	.73	.74	.80	-
Lgb2_Luplu	7	.87	.86	.86	.88	.93	.90
		1	2	3	4	5	6

Pairwise alignment:
Calculate distance matrix

ClustalW Step 2: Dendrogram



ClustalW Sequence Weighting

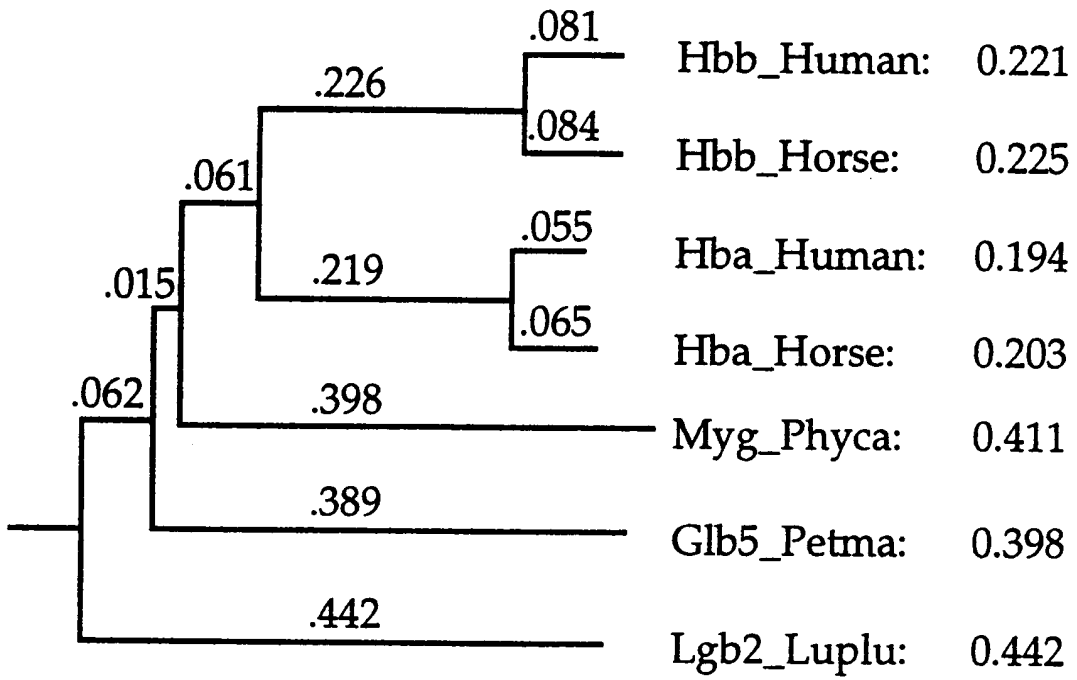


FIG. 3. Sequence weights for the seven globin sequences from Fig. 1. A rooted neighbor-joining tree is shown with branch lengths. The weights are shown for each sequence before normalization (the weights are normalized so as to make the largest equal to 1.0).

ClustalW Residue Specific Gap Penalties

RESIDUE-SPECIFIC GAP OPENING PENALTY FACTORS[a]

Residue	Penalty	Residue	Penalty
A	1.13	M	1.29
C	1.13	N	0.63
D	0.96	P	0.74
E	1.31	Q	1.07
F	1.20	R	0.72
G	0.61	S	0.76
H	1.00	T	0.89
I	1.32	V	1.25
K	0.96	Y	1.00
L	1.21	W	1.23

Position Specific Gap Penalties

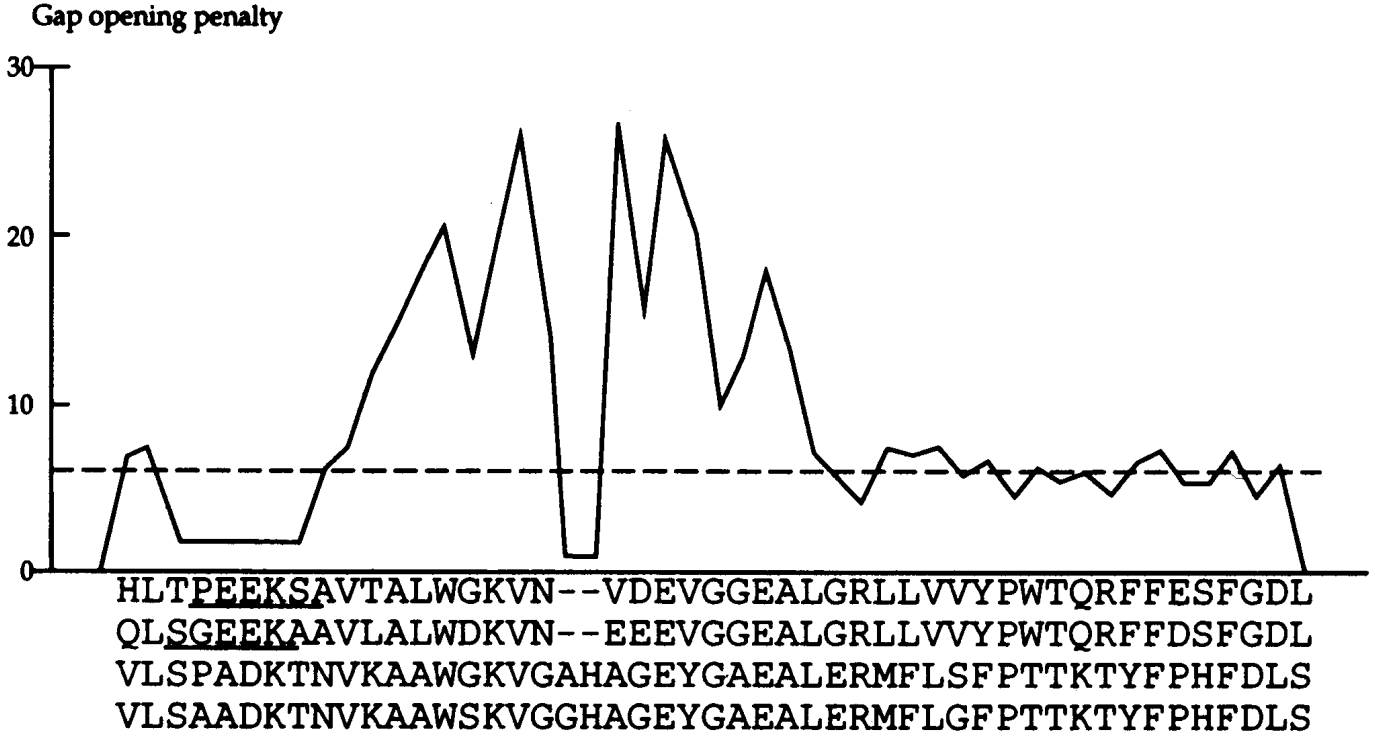
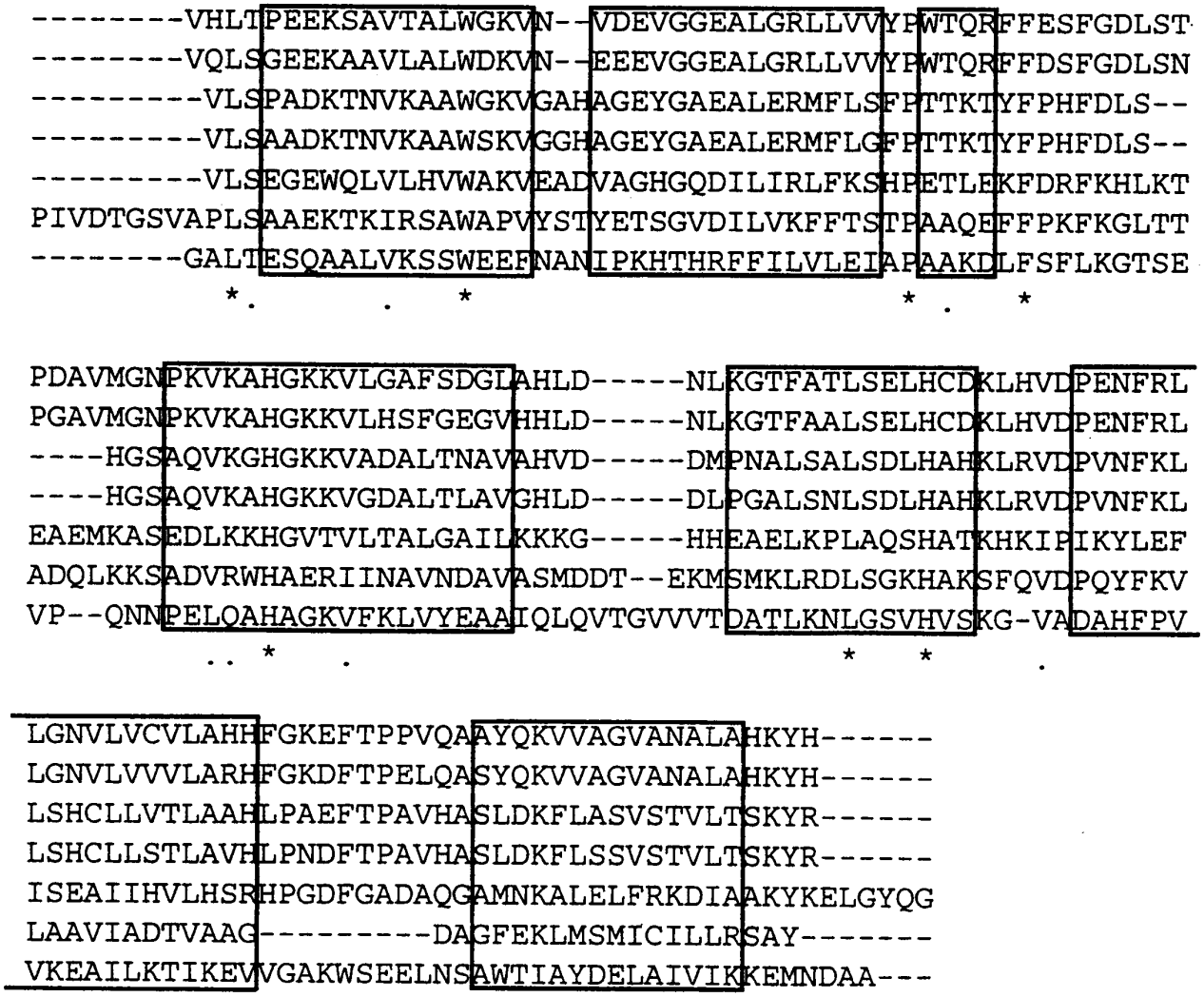


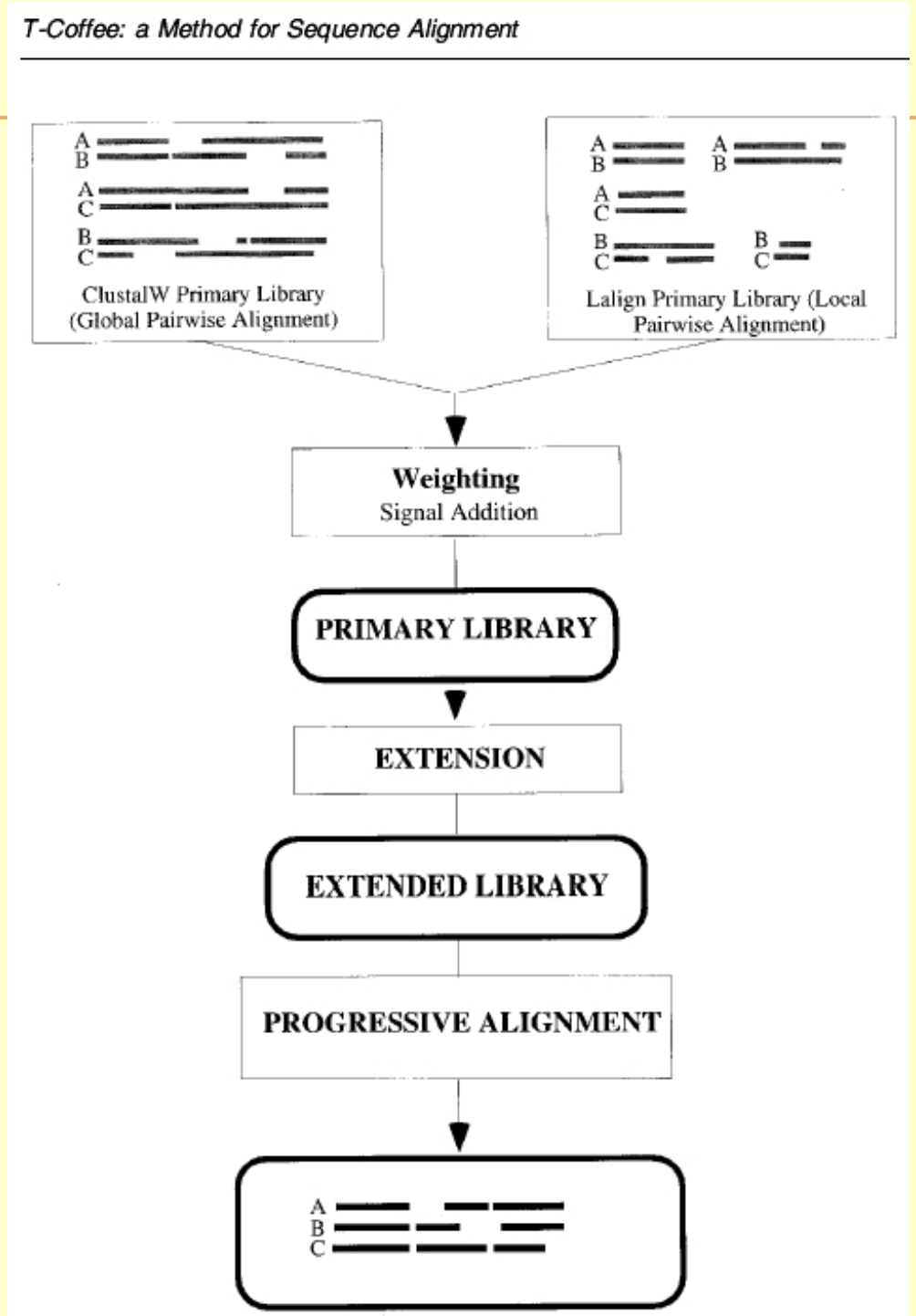
Figure 3. The variation in local gap opening penalty is plotted for a section of alignment. The inital gap opening penalty is indicated by a dotted line. Two hydrophilic stretches are underlined. The lowest penalties correspond to the ends of the alignment, the hydrophilic stretches and the two positions with gaps. The highest values are within 8 residues of the two gap positions. The rest of the variation is caused by the residue specific gap penalties (12).

ClustalW Step 3: Progressive Alignment



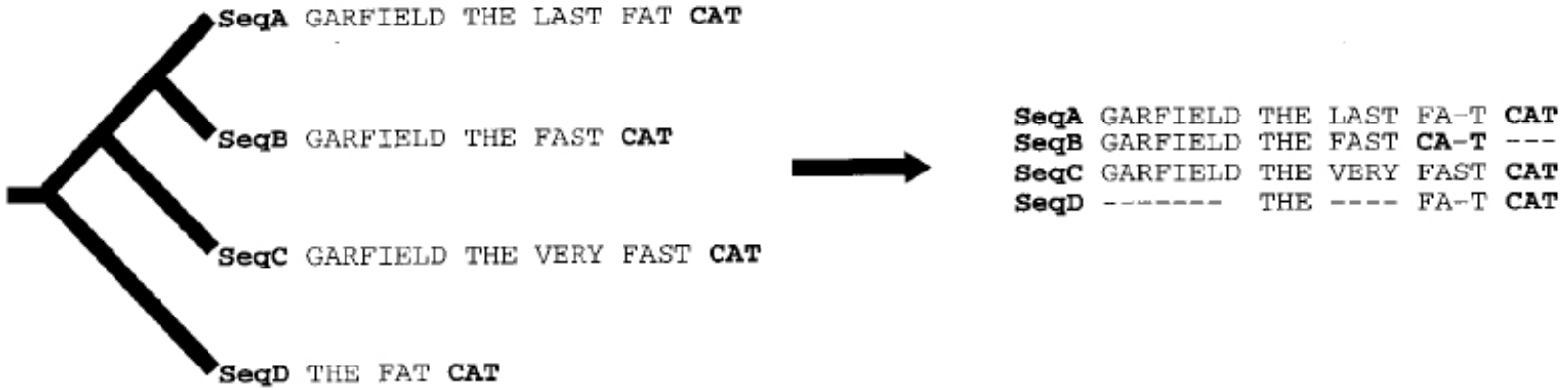
Progressive alignment: Align following the guide tree

T-Coffee Procedure



Regular Progressive Alignment

a)Regular Progressive Alignment Strategy



T-Coffee Primary Alignment Library

b)Primary Library

```
SeqA GARFIELD THE LAST FAT CAT
SeqB GARFIELD THE FAST CAT ---
```
Prim. Weight = 88

```
SeqA GARFIELD THE LAST FA-T CAT
SeqC GARFIELD THE VERY FAST CAT
```
Prim. Weight = 77

```
SeqA GARFIELD THE LAST FAT CAT
SeqD ------- THE ---- FAT CAT
```
Prim. Weight =100

```
SeqB GARFIELD THE ---- FAST CAT
SeqC GARFIELD THE VERY FAST CAT
```
Prim Weight = 100

```
SeqB GARFIELD THE FAST CAT
SeqD ------- THE FA-T CAT
```
Prim. Weight = 100

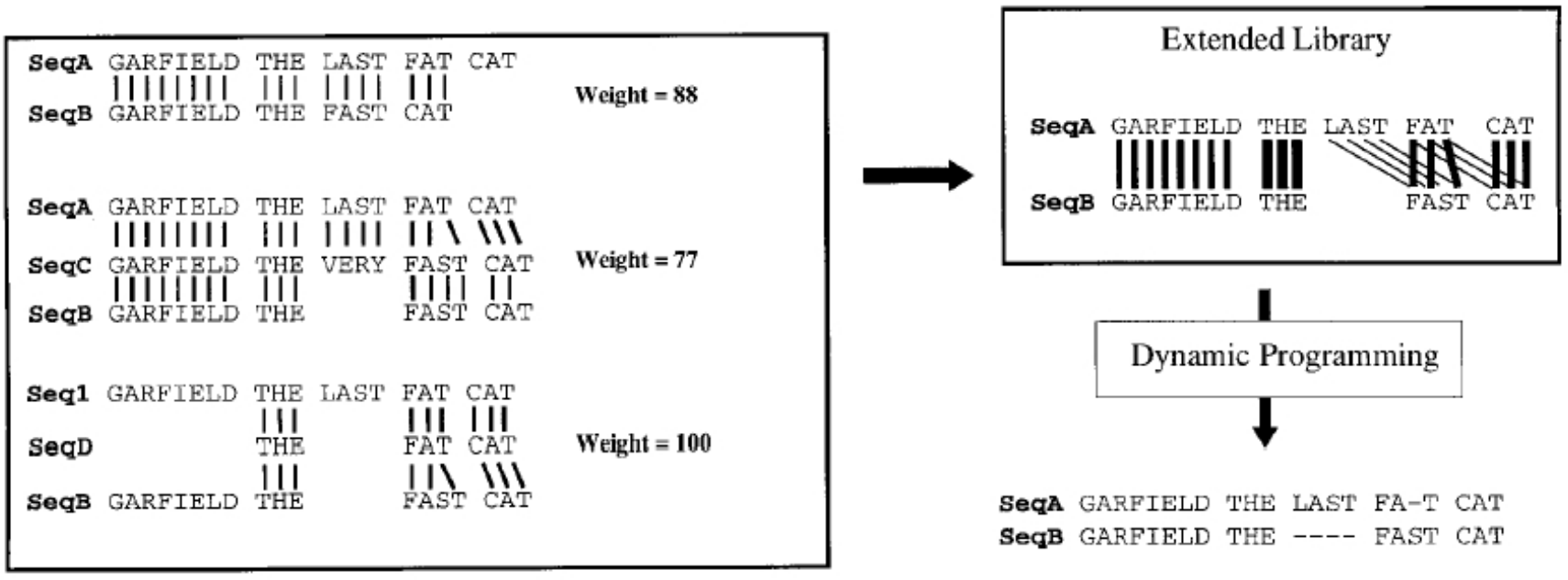
```
SeqC GARFIELD THE VERY FAST CAT
SeqD ------- THE ---- FA-T CAT
```
Prim. Weight = 100

T-Coffee Extended Alignment Library and Progressive Alignment

b)Primary Library

```
SeqA GARFIELD THE LAST FAT CAT
SeqB GARFIELD THE FAST CAT ---        Prim. Weight = 88

SeqA GARFIELD THE LAST FA-T CAT
SeqC GARFIELD THE VERY FAST CAT       Prim. Weight = 77

SeqA GARFIELD THE LAST FAT CAT
SeqD ------- THE ---- FAT CAT         Prim. Weight =100

SeqB GARFIELD THE ---- FAST CAT
SeqC GARFIELD THE VERY FAST CAT       Prim Weight = 100

SeqB GARFIELD THE FAST CAT
SeqD ------- THE FA-T CAT             Prim. Weight = 100

SeqC GARFIELD THE VERY FAST CAT
SeqD ------- THE ---- FA-T CAT        Prim. Weight = 100
```

c)Extended Library for seq1 and seq2

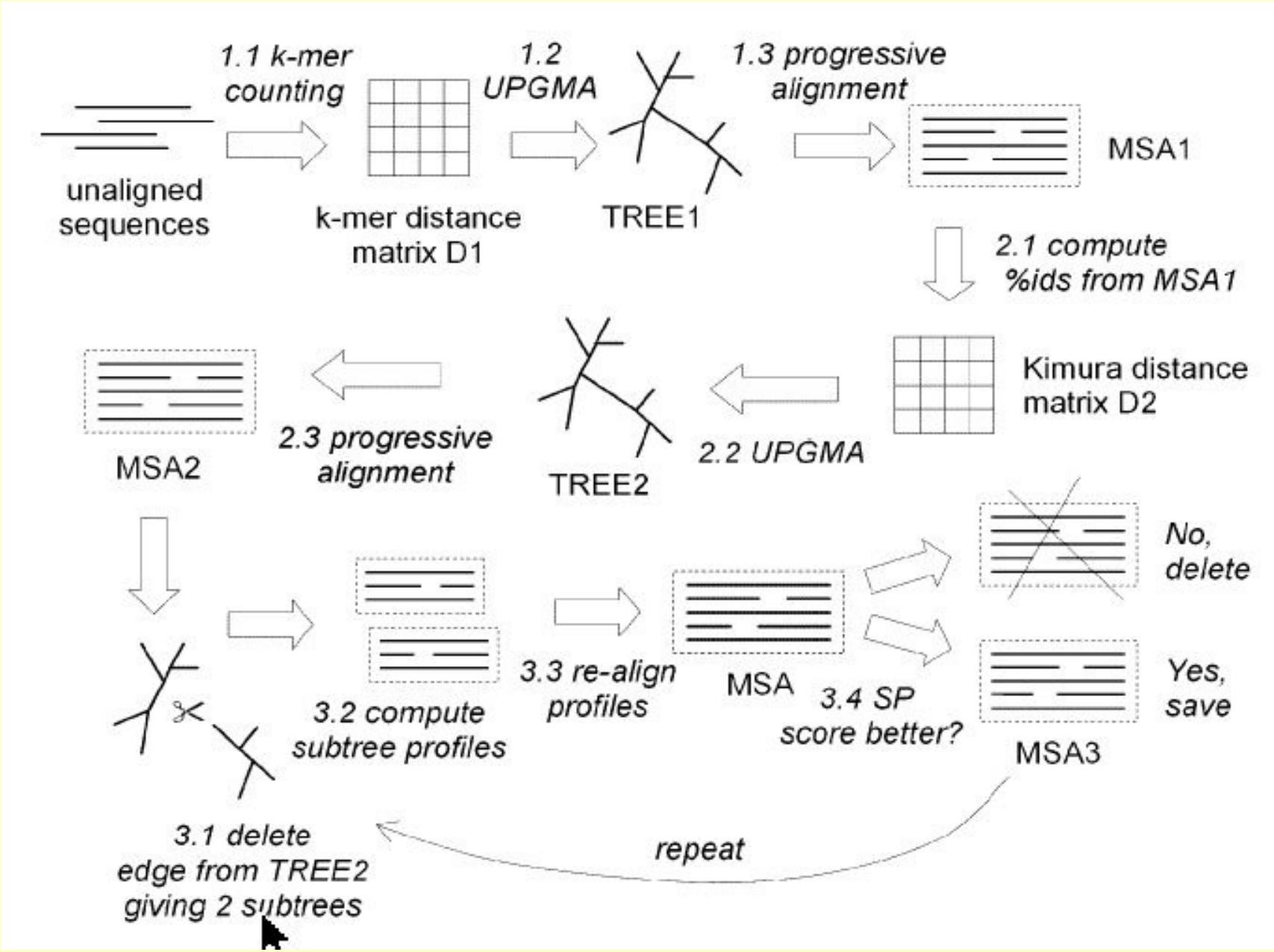


Comparison of T-Coffee to Other MSAs

Table 2. T-Coffee compared with other multiple sequence alignment methods

Method	Cat1 (81)	Cat2 (23)	Cat3 (4)	Cat4 (12)	Cat5 (11)	Total1 (141)
Dialign	71.0	25.2	35.1	74.7	80.4	61.5
ClustalW	78.5	32.2	42.5	65.7	74.3	66.4
Prrp	78.6	32.5	50.2	51.1	82.7	66.4
T-Coffee	**80.7**	**37.3**	**52.9**	**83.2**	**88.7**	**72.1**

MUSCLE
Edgar (2004) NAR 32, 1792-1797



MUSCLE

Edgar (2004) NAR 32, 1792-1797

```
YES_XIPHE    MGCvrSKEaKgPAlKYqpdNsnvvPvSahlgHYGpeptimg
YES_AVISY    -------------dKgPAmKYrtdNtpePiSshvsHYGsd
YES_CHICK    ------MGCikSKEdKgPAmKYrtdNtpePiSshvsHYGsd
YES_HUMAN    ------MGCikSKEnKsPAiKYrpeNtpePvStsvsHYGae
YES_MOUSE    ------MGCikSKEnKsPAiKYtpeNlteP--vSpsasHYG

YES_XIPHE    MGCvrSKEaKgPAlKYqpdNsnvvPvSahlgHYGpeptimg
YES_AVISY    --------dKgPAmKYrtdNtp-ePiSshvsHYGsdssqat
YES_CHICK    MGCikSKEdKgPAmKYrtdNtp-ePiSshvsHYGsdssqat
YES_HUMAN    MGCikSKEnKsPAiKYrpeNtp-ePvStsvsHYGaepttvs
YES_MOUSE    MGCikSKEnKsPAiKYtpeNlt-ePvSpsasHYGvehatva
```

MUSCLE BaliBase Test
Edgar (2004) NAR 32, 1792-1797

Table 2. BAliBASE Q scores on subsets

Method	Ref1	Ref2	Ref3	Ref4	Ref5
MUSCLE	0.887	0.935	0.823	0.876	0.968
MUSCLE-p	0.871	0.928	0.813	0.857	0.974
T-Coffee	0.866	0.934	0.787	0.917	0.957
NWNSI	0.867	0.923	0.787	0.904	0.963
CLUSTALW	0.861	0.932	0.751	0.823	0.859
FFTNS1	0.838	0.908	0.708	0.793	0.947

ProbCons

Do, et al. Genome Research 15, 330-340

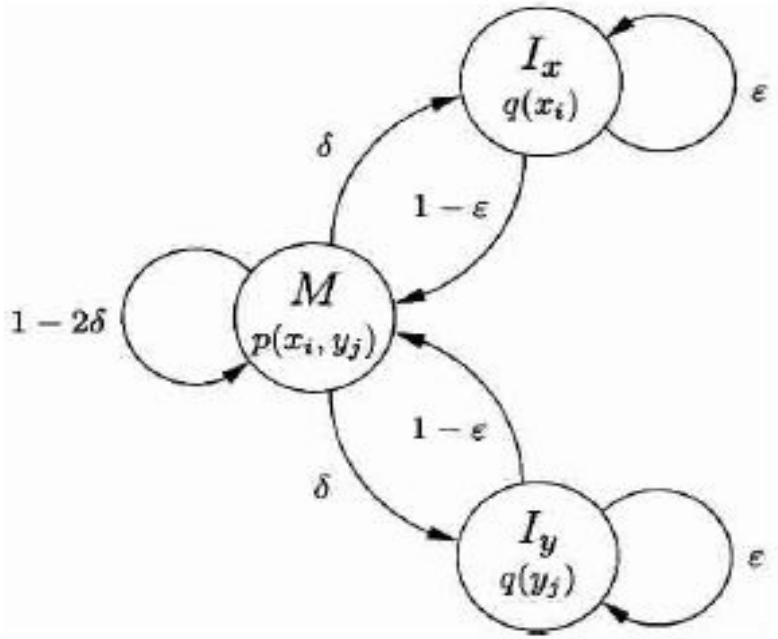


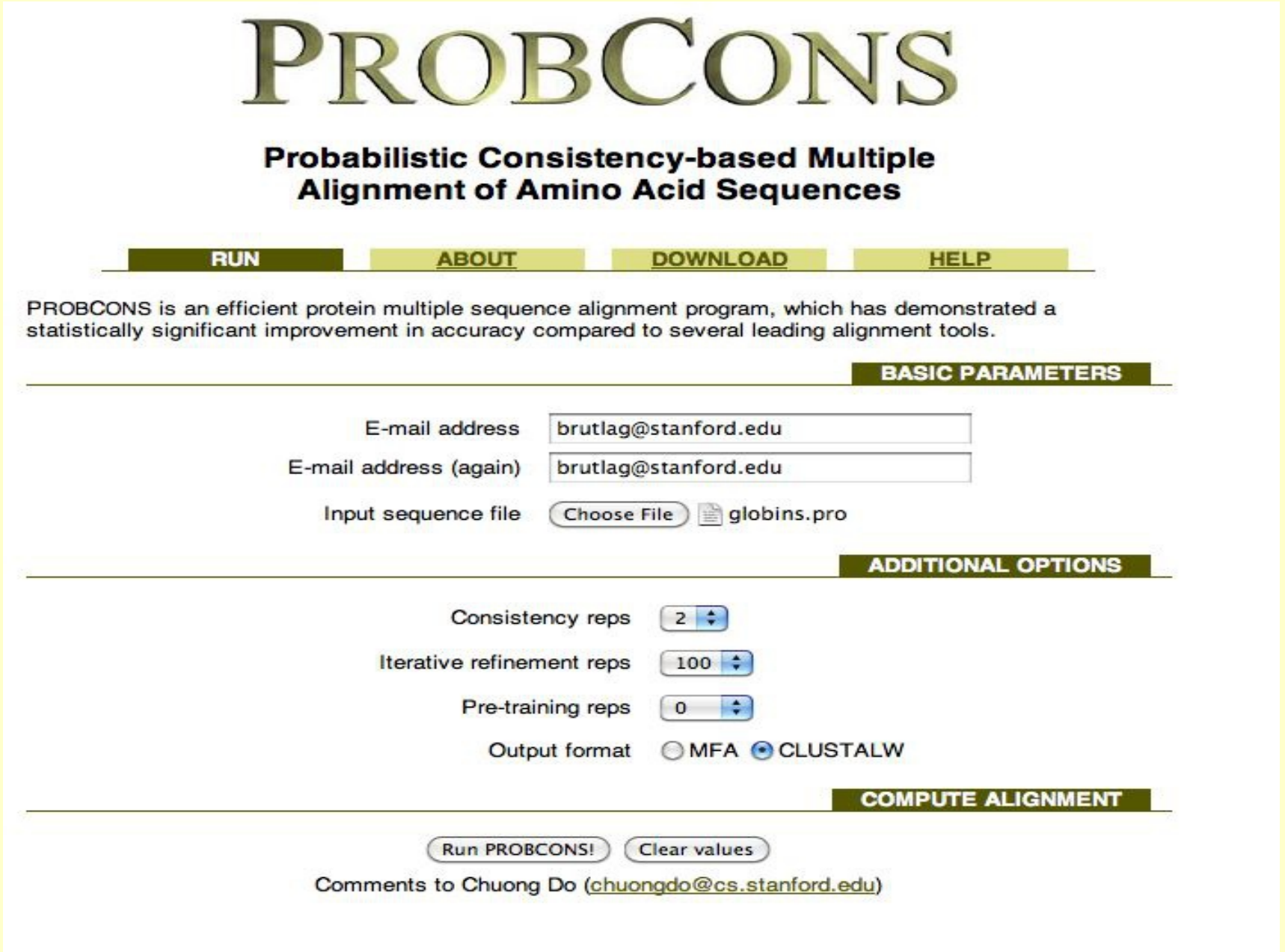
Figure 1. Basic pair-HMM for sequence alignment between two sequences, *x* and *y*. State *M* emits two letters, one from each sequence, and corresponds to the two letters being aligned together. State I_x emits a letter in sequence *x* that is aligned to a gap, and similarly state I_y emits a letter in sequence *y* that is aligned to a gap. Finding the most likely alignment according to this model by using the Viterbi algorithm corresponds to applying Needleman–Wunsch with appropriate parameters. The logarithm of the emission probability function *p*(.,.) at *M* corresponds to a substitution scoring matrix, while affine gap penalty parameters can be derived from the transition probabilities δ and ε (Durbin et al. 1998).

ProbCons
Do, et al. Genome Research 15, 330-340

- Step 1: Compute posterior probability matrices of each pair of aligned sequences from the pair-HMM model
- Step 2: Compute expected accuracies of pairwise alignments.
- Step 3: Probabilistic Consistency Transformation
- Step 4: Calculate Guide Tree using UPGMA from measure of similarities of sequence pairs
- Step 5: Progressive alignment
- Step 6: refinment by dividing sequences into two groups and re-align. Repeat multiple times.

ProbCons

Do, et al. Genome Research 15, 330-340



ProbCons

Do, et al. Genome Research 15, 330-340

```
PROBCONS version 1.12 multiple sequence alignment

GLB5_PETMA     PIVDTGSVAPLSAAEKTKIRSAWAPVYSTYETSGVDILVKFFTSTPAAQEFFPKFKGLTT
HBA_HORSE      ---------VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-DLS-
HBA_HUMAN      ---------VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-
HBB_HORSE      --------VQLSGEEKAAVLALWDKVNE--EEVGGEALGRLLVVYPWTQRFFDSFGDLSN
HBB_HUMAN      --------VHLTPEEKSAVTALWGKVNV--DEVGGEALGRLLVVYPWTQRFFESFGDLST
LGB2_LUPLU     --------GALTESQAALVKSSWEEFNANIPKHTHRFFILVLEIAPAAKDLFSFLKGTS-
MYG_PHYCA      ---------VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKT
                        *:  :   :   *  .          :  .:   * :   *  :   ..

GLB5_PETMA     ADQLKKSADVRWHAERIINAVNDAVASMD-----DTEKMSMKLRDLSGKHAKSFQVDPQY
HBA_HORSE      ----HGSAQVKAHGKKVGDALTLAVGHLD-----DLP---GALSNLSDLHAHKLRVDPVN
HBA_HUMAN      ----HGSAQVKGHGKKVADALTNAVAHVD-----DMP---NALSALSDLHAHKLRVDPVN
HBB_HORSE      PGAVMGNPKVKAHGKKVLHSFGEGVHHLD-----NLK---GTFAALSELHCDKLHVDPEN
HBB_HUMAN      PDAVMGNPKVKAHGKKVLGAFSDGLAHLD-----NLK---GTFATLSELHCDKLHVDPEN
LGB2_LUPLU     -EVPQNNPELQAHAGKVFKLVYEAAIQLQVTGVVVTD---ATLKNLGSVHVSKGVAD-AH
MYG_PHYCA      EAEMKASEDLKKHGVTVLTALGAILKKKG-----HHE---AELKPLAQSHATKHKIPIKY
                    . .:: *.  :   .                      :  *.  *  .

GLB5_PETMA     FKVLAAVIADTVAA---------GDAGFEKLMSMICIL---L-RSA------Y
HBA_HORSE      FKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVSTV---LTSKY------R
HBA_HUMAN      FKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTV---LTSKY------R
HBB_HORSE      FRLLGNVLVVVLARHFGKDFTPELQASYQKVVAGVANA---LAHKY------H
HBB_HUMAN      FRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANA---LAHKY------H
LGB2_LUPLU     FPVVKEAILKTIKEVVGAKWSEELNSAWTIAYDELAIVIKKEMNDA------A
MYG_PHYCA      LEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKD---IAAKYKELGYQG
             : .:   :  .:      .:    ...       .          .
```

Table 1. Performance of aligners on the BAliBASE benchmark alignments database

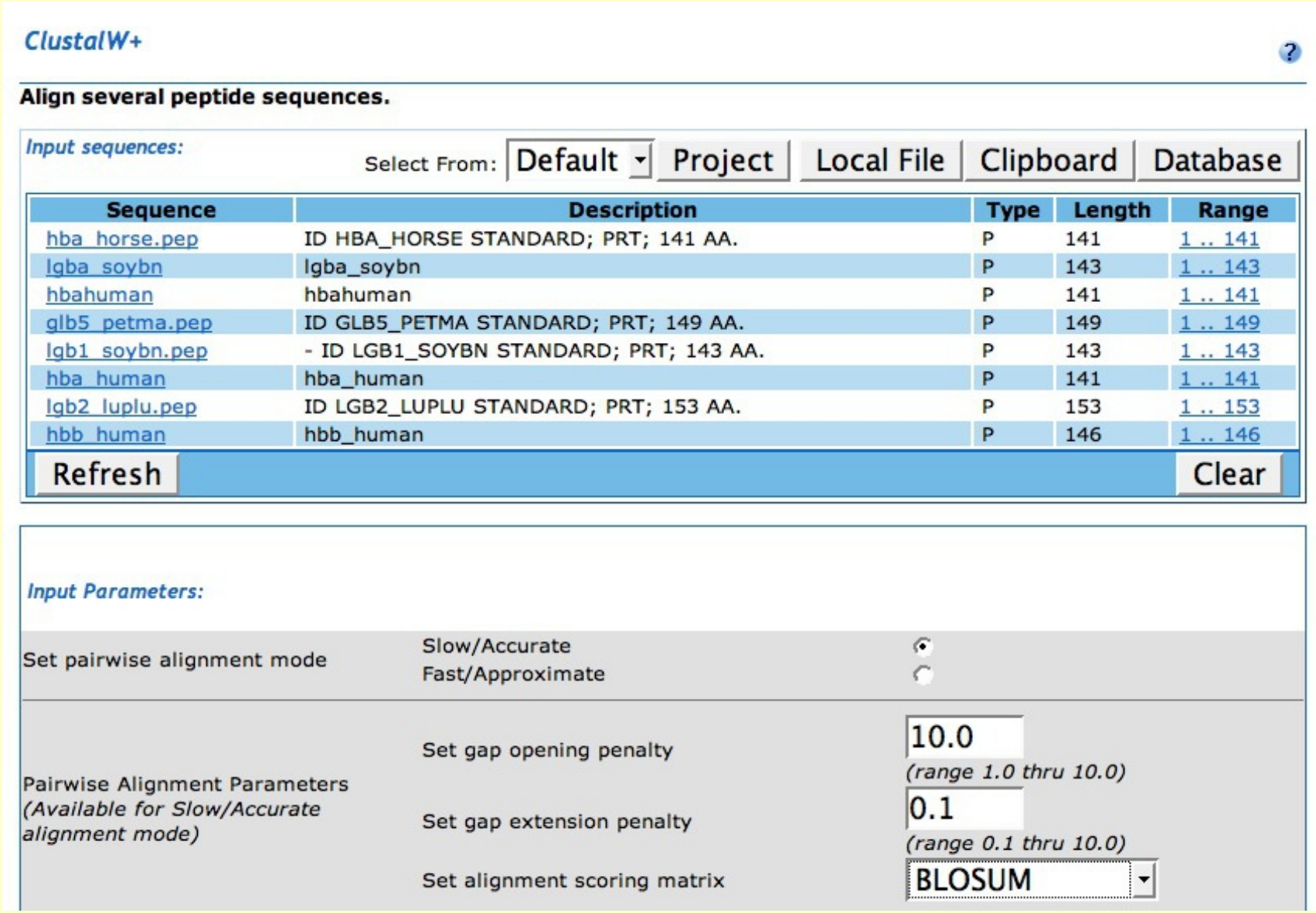
	Ref 1 (82)		Ref 2 (23)		Ref 3 (12)		Ref 4 (12)		Ref 5 (12)		Overall (141)		
Aligner	SP	CS	SP	CS	SP	CS	SP	CS	SP	CS	SP	CS	Time (mm:ss)
Align-m	76.6	n/a	88.4	n/a	68.4	n/a	91.1	n/a	91.7	n/a	80.4	n/a	19:25
DIALIGN	81.1	70.9	89.3	35.9	68.4	34.4	89.7	76.2	94.0	84.3	83.2	63.7	2:53
CLUSTALW	86.1	77.3	93.2	56.8	75.3	46.0	83.4	52.2	85.9	63.8	86.1	68.0	1:07
MAFFT	86.7	78.1	92.4	50.2	78.8	50.4	91.6	72.7	96.3	85.9	88.2	71.4	1:18
T-Coffee	86.6	77.4	93.4	56.1	78.5	48.7	91.8	73.0	95.8	90.3	88.3	72.2	21:31
MUSCLE	88.7	80.8	93.5	56.3	82.5	56.4	87.6	60.9	96.8	90.2	89.6	73.9	**1:05**
ProbCons	**90.1**	**82.6**	**94.4**	**61.3**	84.1	**61.3**	90.1	72.3	97.9	91.9	91.0	77.2	5:32
ProbCons-ext	90.0	82.5	94.2	59.1	**84.3**	61.1	**93.8**	**81.0**	**98.1**	**92.2**	**91.2**	**77.6**	8:02

Columns show the average sum-of-pairs (SP) and column scores (CS) achieved by each aligner for each of the five BAliBASE references. All scores have been multiplied by 100. The number of sequences in each reference is given in parentheses. Overall numbers for the entire database are reported in addition to the total running time of each aligner for all 141 alignments. The best results in each column are shown in bold.

SeqWeb v3.1

SeqWeb ClustalW

http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=clustalw-prot



ClustalW+

Align several peptide sequences.

Input sequences:

Select From: Default | Project | Local File | Clipboard | Database

Sequence	Description	Type	Length	Range
hba_horse.pep	ID HBA_HORSE STANDARD; PRT; 141 AA.	P	141	1 .. 141
lgba_soybn	lgba_soybn	P	143	1 .. 143
hbahuman	hbahuman	P	141	1 .. 141
glb5_petma.pep	ID GLB5_PETMA STANDARD; PRT; 149 AA.	P	149	1 .. 149
lgb1_soybn.pep	- ID LGB1_SOYBN STANDARD; PRT; 143 AA.	P	143	1 .. 143
hba_human	hba_human	P	141	1 .. 141
lgb2_luplu.pep	ID LGB2_LUPLU STANDARD; PRT; 153 AA.	P	153	1 .. 153
hbb_human	hbb_human	P	146	1 .. 146

Refresh | Clear

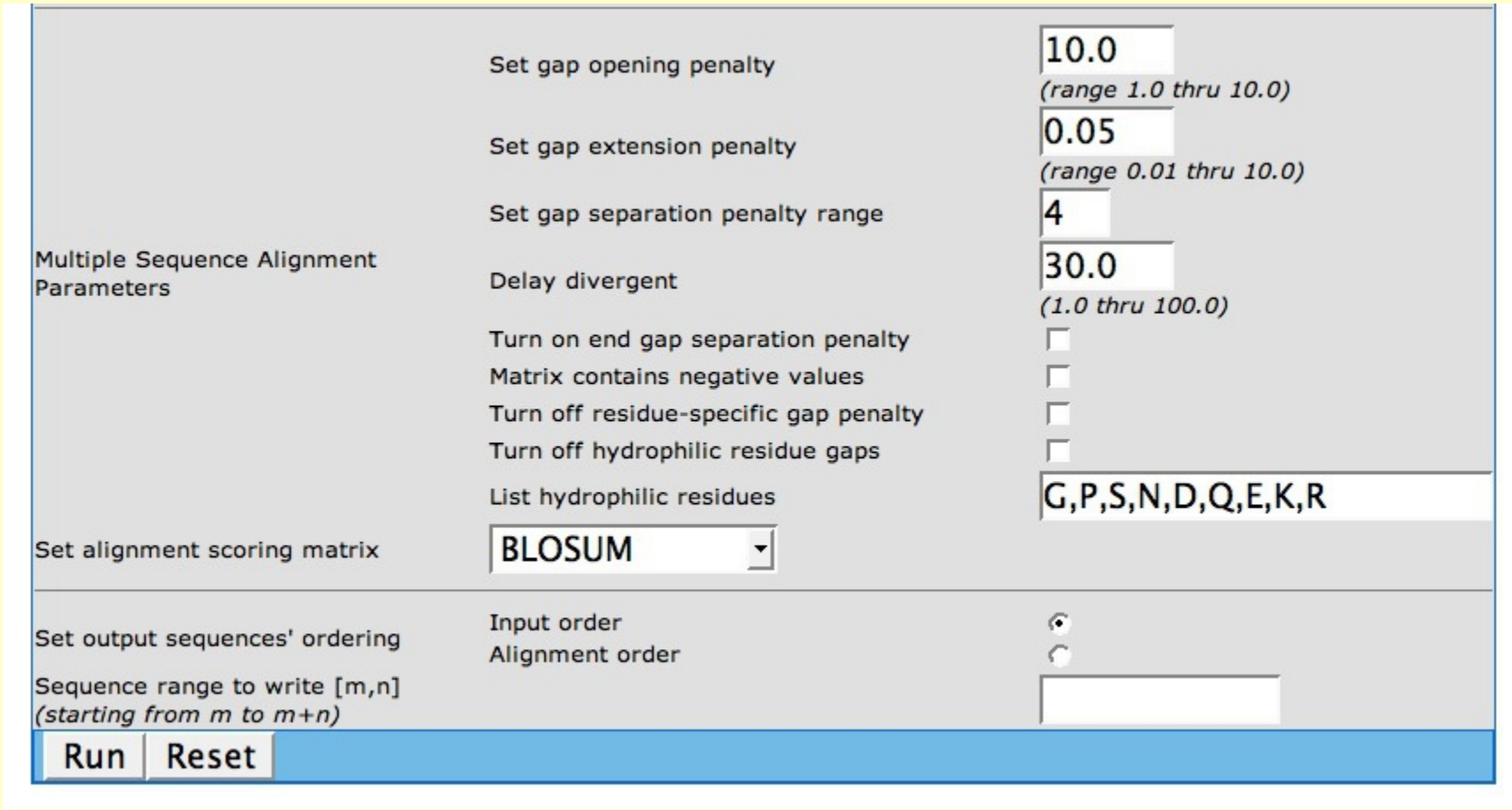
Input Parameters:

Set pairwise alignment mode	Slow/Accurate Fast/Approximate	◉ ○
Pairwise Alignment Parameters *(Available for Slow/Accurate alignment mode)*	Set gap opening penalty	10.0 *(range 1.0 thru 10.0)*
	Set gap extension penalty	0.1 *(range 0.1 thru 10.0)*
	Set alignment scoring matrix	BLOSUM

SeqWeb v3.1

SeqWeb ClustalW MSA Parameters

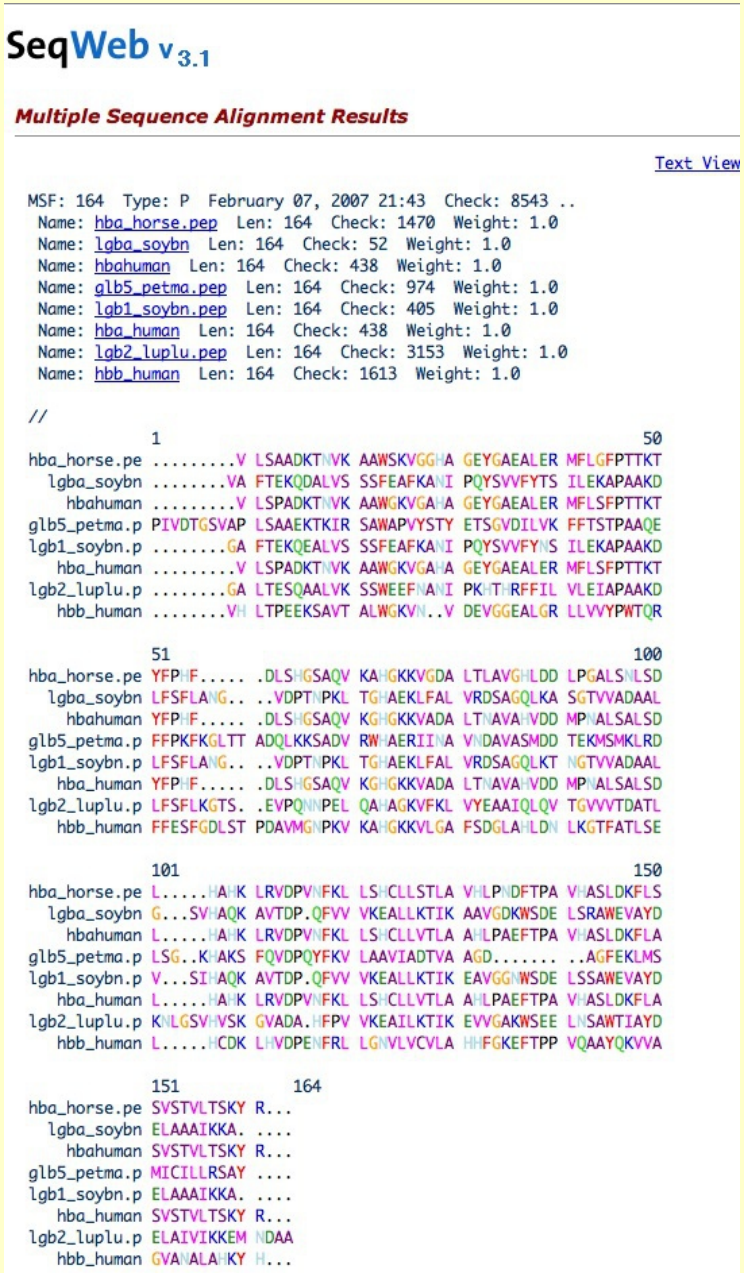
http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=clustalw-prot



SeqWeb v3.1

SeqWeb ClustalW Alignment

http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=clustalw-prot



SeqWeb v3.1

Multiple Sequence Alignment Results

Text View

```
MSF: 164  Type: P  February 07, 2007 21:43  Check: 8543 ..
 Name: hba_horse.pep  Len: 164  Check: 1470  Weight: 1.0
 Name: lgba_soybn  Len: 164  Check: 52  Weight: 1.0
 Name: hbahuman  Len: 164  Check: 438  Weight: 1.0
 Name: glb5_petma.pep  Len: 164  Check: 974  Weight: 1.0
 Name: lgb1_soybn.pep  Len: 164  Check: 405  Weight: 1.0
 Name: hba_human  Len: 164  Check: 438  Weight: 1.0
 Name: lgb2_luplu.pep  Len: 164  Check: 3153  Weight: 1.0
 Name: hbb_human  Len: 164  Check: 1613  Weight: 1.0

//
                1                                                  50
hba_horse.pe .........V LSAADKTNVK AAWSKVGGHA GEYGAEALER MFLGFPTTKT
  lgba_soybn ........VA FTEKQDALVS SSFEAFKANI PQYSVVFYTS ILEKAPAAKD
    hbahuman .........V LSPADKTNVK AAWGKVGAHA GEYGAEALER MFLSFPTTKT
glb5_petma.p PIVDTGSVAP LSAAEKTKIR SAWAPVYSTY ETSGVDILVK FFTSTPAAQE
lgb1_soybn.p ........GA FTEKQEALVS SSFEAFKANI PQYSVVFYNS ILEKAPAAKD
   hba_human .........V LSPADKTNVK AAWGKVGAHA GEYGAEALER MFLSFPTTKT
lgb2_luplu.p ........GA LTESQAALVK SSWEEFNANI PKNTNRFFIL VLEIAPAAKD
   hbb_human ........VH LTPEEKSAVT ALWGKVN..V DEVGGEALGR LLVVYPWTQR

                51                                                 100
hba_horse.pe YFPHF..... .DLSHGSAQV KAHGKKVGDA LTLAVGHLDD LPGALSNLSD
  lgba_soybn LFSFLANG.. ..VDPTNPKL TGHAEKLFAL VRDSAGQLKA SGTVVADAAL
    hbahuman YFPHF..... .DLSHGSAQV KGHGKKVADA LTNAVAHVDD MPNALSALSD
glb5_petma.p FFPKFKGLTT ADQLKKSADV RWHAERIINA VNDAVASMDD TEKMSMKLRD
lgb1_soybn.p LFSFLANG.. ..VDPTNPKL TGHAEKLFAL VRDSAGQLKT NGTVVADAAL
   hba_human YFPHF..... .DLSHGSAQV KGHGKKVADA LTNAVAHVDD MPNALSALSD
lgb2_luplu.p LFSFLKGTS. .EVPQNNPEL QAHAGKVFKL VYEAAIQLQV TGVVVTDATL
   hbb_human FFESFGDLST PDAVMGNPKV KAHGKKVLGA FSDGLAHLDN LKGTFATLSE

                101                                                150
hba_horse.pe L.....HAHK LRVDPVNFKL LSHCLLSTLA VHLPNDFTPA VHASLDKFLS
  lgba_soybn G...SVHAQK AVTDP.QFVV VKEALLKTIK AAVGDKWSDE LSRAWEVAYD
    hbahuman L.....HAHK LRVDPVNFKL LSHCLLVTLA AHLPAEFTPA VHASLDKFLA
glb5_petma.p LSG..KHAKS FQVDPQYFKV LAAVIADTVA AGD..... ..AGFEKLMS
lgb1_soybn.p V...SIHAQK AVTDP.QFVV VKEALLKTIK EAVGGNWSDE LSSAWEVAYD
   hba_human L.....HAHK LRVDPVNFKL LSHCLLVTLA AHLPAEFTPA VHASLDKFLA
lgb2_luplu.p KNLGSVHVSK GVADA.HFPV VKEAILKTIK EVVGAKWSEE LNSAWTIAYD
   hbb_human L.....HCDK LHVDPENFRL LGNVLVCVLA HHFGKEFTPP VQAAYQKVVA

                151           164
hba_horse.pe SVSTVLTSKY R...
  lgba_soybn ELAAAIKKA. ....
    hbahuman SVSTVLTSKY R...
glb5_petma.p MICILLRSAY ....
lgb1_soybn.p ELAAAIKKA. ....
   hba_human SVSTVLTSKY R...
lgb2_luplu.p ELAIVIKKEM NDAA
   hbb_human GVANALAHKY H...
```

SeqWeb v3.1

SeqWeb ClustalW Text Output

http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=clustalw-prot

```
!!AA_MULTIPLE_ALIGNMENT 1.0
MSF: 164  Type: P  February 07, 2007 21:44  Check: 8543 ..
 Name: hbahuman  Len: 164  Check: 438  Weight: 1.0
 Name: hba_human  Len: 164  Check: 438  Weight: 1.0
 Name: hba_horse.pep  Len: 164  Check: 1470  Weight: 1.0
 Name: hbb_human  Len: 164  Check: 1613  Weight: 1.0
 Name: lgba_soybn  Len: 164  Check: 52  Weight: 1.0
 Name: lgb1_soybn.pep  Len: 164  Check: 405  Weight: 1.0
 Name: lgb2_luplu.pep  Len: 164  Check: 3153  Weight: 1.0
 Name: glb5_petma.pep  Len: 164  Check: 974  Weight: 1.0

//

                1                                                    50
hbahuman        .........V LSPADKTNVK AAWGKVGAHA GEYGAEALER MFLSFPTTKT
hba_human       .........V LSPADKTNVK AAWGKVGAHA GEYGAEALER MFLSFPTTKT
hba_horse.pep   .........V LSAADKTNVK AAWSKVGGHA GEYGAEALER MFLGFPTTKT
hbb_human       ........VH LTPEEKSAVT ALWGKVN..V DEVGGEALGR LLVVYPWTQR
lgba_soybn      ........VA FTEKQDALVS SSFEAFKANI PQYSVVFYTS ILEKAPAAKD
lgb1_soybn.pep  ........GA FTEKQEALVS SSFEAFKANI PQYSVVFYNS ILEKAPAAKD
lgb2_luplu.pep  ........GA LTESQAALVK SSWEEFNANI PKHTHRFFIL VLEIAPAAKD
glb5_petma.pep  PIVDTGSVAP LSAAEKTKIR SAWAPVYSTY ETSGVDILVK FFTSTPAAQE

               51                                                   100
hbahuman        YFPHF..... .DLSHGSAQV KGHGKKVADA LTNAVAHVDD MPNALSALSD
hba_human       YFPHF..... .DLSHGSAQV KGHGKKVADA LTNAVAHVDD MPNALSALSD
hba_horse.pep   YFPHF..... .DLSHGSAQV KAHGKKVGDA LTLAVGHLDD LPGALSNLSD
hbb_human       FFESFGDLST PDAVMGNPKV KAHGKKVLGA FSDGLAHLDN LKGTFATLSE
lgba_soybn      LFSFLANG.. ..VDPTNPKL TGHAEKLFAL VRDSAGQLKA SGTVVADAAL
lgb1_soybn.pep  LFSFLANG.. ..VDPTNPKL TGHAEKLFAL VRDSAGQLKT NGTVVADAAL
lgb2_luplu.pep  LFSFLKGTS. .EVPQNNPEL QAHAGKVFKL VYEAAIQLQV TGVVVTDATL
glb5_petma.pep  FFPKFKGLTT ADQLKKSADV RWHAERIINA VNDAVASMDD TEKMSMKLRD

               101                                                  150
hbahuman        L.....HAHK LRVDPVNFKL LSHCLLVTLA AHLPAEFTPA VHASLDKFLA
hba_human       L.....HAHK LRVDPVNFKL LSHCLLVTLA AHLPAEFTPA VHASLDKFLA
hba_horse.pep   L.....HAHK LRVDPVNFKL LSHCLLSTLA VHLPNDFTPA VHASLDKFLS
hbb_human       L.....HCDK LHVDPENFRL LGNVLVCVLA HHFGKEFTPP VQAAYQKVVA
lgba_soybn      G...SVHAQK AVTDP.QFVV VKEALLKTIK AAVGDKWSDE LSRAWEVAYD
lgb1_soybn.pep  V...SIHAQK AVTDP.QFVV VKEALLKTIK EAVGGNWSDE LSSAWEVAYD
lgb2_luplu.pep  KNLGSVHVSK GVADA.HFPV VKEAILKTIK EVVGAKWSEE LNSAWTIAYD
glb5_petma.pep  LSG..KHAKS FQVDPQYFKV LAAVIADTVA AGD.......  ..AGFEKLMS

               151         164
hbahuman        SVSTVLTSKY R...
hba_human       SVSTVLTSKY R...
hba_horse.pep   SVSTVLTSKY R...
hbb_human       GVANALAHKY H...
lgba_soybn      ELAAAIKKA. ....
lgb1_soybn.pep  ELAAAIKKA. ....
lgb2_luplu.pep  ELAIVIKKEM NDAA
glb5_petma.pep  MICILLRSAY ....
```

SeqWeb v3.1

SeqWeb Pileup Input

http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=pileup-prot

PileUp

Align several peptide sequences.

Input sequences:

Select From: Default | Project | Local File | Clipboard | Database

Sequence	Description	Type	Length	Range
myg_phyca	myg_phyca	P	153	1 .. 153
glb5_petma.pep	ID GLB5_PETMA STANDARD; PRT; 149 AA.	P	149	1 .. 149
hba_human	hba_human	P	141	1 .. 141
hba_horse.pep	ID HBA_HORSE STANDARD; PRT; 141 AA.	P	141	1 .. 141
hbb_horse.pep	ID HBB_HORSE STANDARD; PRT; 146 AA.	P	146	1 .. 146
lgb1_soybn.pep	- ID LGB1_SOYBN STANDARD; PRT; 143 AA.	P	143	1 .. 143
hbb_human	hbb_human	P	146	1 .. 146

Refresh Clear

Input Parameters:

Select a sequence comparision matrix. This matrix determines how matches and mismatches are scored. The default penalites for gap creation and extension are given after each matrix name.

Scoring Matrix: blosum62

Set gap creation penalty: 8

Set gap extension penalty: 2

Limit the maximum input sequence range only when needed. Setting a higher limit allows you to align longer sequences while setting a lower limit allows you to add more and longer gaps to each sequence.

Maximum input sequence range: 5000 *(range 1 thru 7000)*

Limit the maximum number of gaps only when needed. Setting a higher limit allows you to add more and longer gaps to each sequence while setting a lower limit allows you to align a greater number of sequences.

Maximum number of gap characters ('.' and '~') added to any sequence: 2000 *(range 0 thru 7000)*

Run Reset

SeqWeb v3.1

http://

SeqWeb v3.1

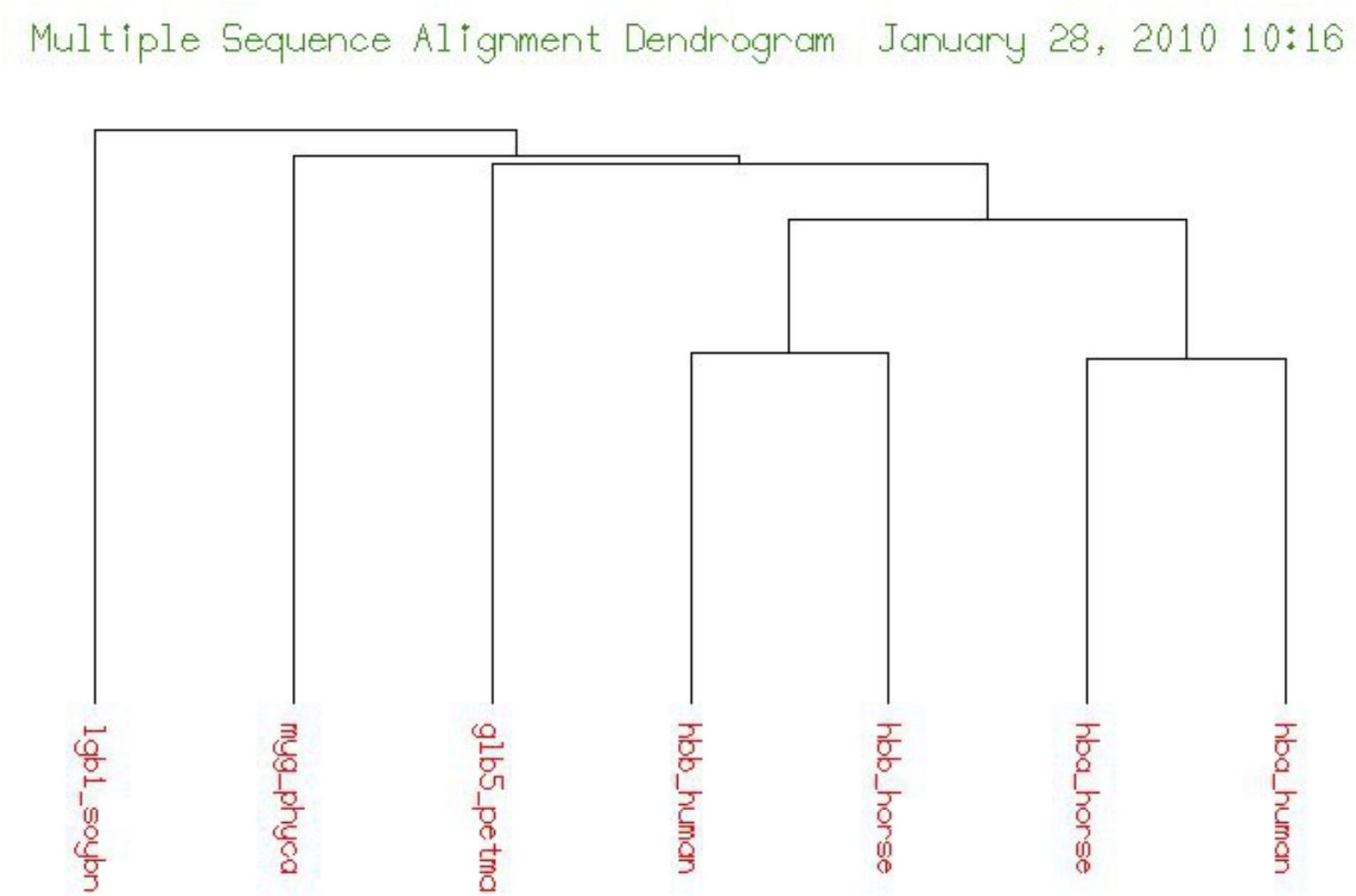
Multiple Sequence Alignment Results

```
Symbol comparison table: share_matrix:blosum62.cmp  CompCheck: 1102

                GapWeight: 8
          GapLengthWeight: 2

myg_phyca_pileup_15277.txt  MSF: 165  Type: P  January 28, 2010 10:16  Check: 6593 ..

Name: hba_human       Len:   165  Check: 1231  Weight:  1.00
Name: hba_horse       Len:   165  Check: 2167  Weight:  1.00
Name: hbb_horse       Len:   165  Check: 9310  Weight:  1.00
Name: hbb_human       Len:   165  Check:  208  Weight:  1.00
Name: glb5_petma      Len:   165  Check: 2079  Weight:  1.00
Name: myg_phyca       Len:   165  Check: 4320  Weight:  1.00
Name: lgb1_soybn      Len:   165  Check: 7278  Weight:  1.00

//
               1                                                  50
  hba_human ~~~~~~~~~~v lspadktnvk aawgkvgaha geygaealer mflsfpttkt
  hba_horse ~~~~~~~~~~V LSAADKTNVK AAWSKVGGHA GEYGAEALER MFLGFPTTKT
  hbb_horse ~~~~~~~~~VQ LSGEEKAAVL ALWDKV..NE EEVGGEALGR LLVVYPWTQR
  hbb_human ~~~~~~~~~vh ltpeeksavt alwgkv..nv devggealgr llvvypwtqr
 glb5_petma PIVDTGSVAP LSAAEKTKIR SAWAPVYSTY ETSGVDILVK FFTSTPAAQE
  myg_phyca ~~~~~~~~~~v lsegewqlvl hvwakveadv aghgqdilir lfkshpetle
 lgb1_soybn ~~~~~~~~~ga ftekqealvs ssfeafkani pqysvvfyns ilekapaakd

               51                                                 100
  hba_human yfphf.dlsh .....gsaqv kghgkkvada ltnavahvdd mpnalsalsd
  hba_horse YFPHF.DLSH .....GSAQV KAHGKKVGDA LTLAVGHLDD LPGALSNLSD
  hbb_horse FFDSFGDLSN PGAVMGNPKV KAHGKKVLHS FGEGVHHLDN LKGTFAALSE
  hbb_human ffesfgdlst pdavmgnpkv kahgkkvlga fsdglahldn lkgtfatlse
 glb5_petma FFPKFKGLTT ADQLKKSADV RWHAERIINA VNDAVASMDD TEKMSMKLRD
  myg_phyca kfdrfkhlkt eaemkasedl kkhgvtvlta lg...ailkk kghheaelkp
 lgb1_soybn lfsflan... .gvdptnpkl tghaeklfal vrdsagql.k tngtvvadaa

               101                                                150
  hba_human l...hahklr vdpvnfklls hcllvtlaah lpaeftpavh asldkflasv
  hba_horse L...HAHKLR VDPVNFKLLS HCLLSTLAVH LPNDFTPAVH ASLDKFLSSV
  hbb_horse L...HCDKLH VDPENFRLLG NVLVVVLARH FGKDFTPELQ ASYQKVVAGV
  hbb_human l...hcdklh vdpenfrllg nvlvcvlahh fgkeftppvq aayqkvvagv
 glb5_petma LSGKHAKSFQ VDPQYFKVLA AVIADTVA.. .......AGD AGFEKLMSMI
  myg_phyca laqshatkhk ipikylefis eaiihvlhsr hpgdfgadaq gamnkalelf
 lgb1_soybn lvsihaqkav tdpq.fvvvk eallktikea vggnwsdels sawevaydel

               151             165
  hba_human stvltskyr~ ~~~~~~
  hba_horse STVLTSKYR~ ~~~~~~
  hbb_horse ANALAHKYH~ ~~~~~~
  hbb_human analahkyh~ ~~~~~~
 glb5_petma CILLRSAY~~ ~~~~~~
  myg_phyca rkdiaakyke lgyqg
 lgb1_soybn aaaikka~~~ ~~~~~~
```

SeqWeb v3.1

SeqWeb Pileup Dendrogram

http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=pileup-prot



SeqWeb v3.1

SeqWeb Pretty Input

http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=pretty-prot

Pretty

Align several peptide sequences and calculate a consensus.

Input sequences:

Select From: Default | Project | Local File | Clipboard | Database

Sequence	Description	Type	Length	Range
myg_phyca	myg_phyca	P	153	1 .. 153
glb5_petma.pep	ID GLB5_PETMA STANDARD; PRT; 149 AA.	P	149	1 .. 149
hba_human	hba_human	P	141	1 .. 141
hba_horse.pep	ID HBA_HORSE STANDARD; PRT; 141 AA.	P	141	1 .. 141
hbb_horse.pep	ID HBB_HORSE STANDARD; PRT; 146 AA.	P	146	1 .. 146
lgb1_soybn.pep	- ID LGB1_SOYBN STANDARD; PRT; 143 AA.	P	143	1 .. 143
hbb_human	hbb_human	P	146	1 .. 146

Refresh | Clear

Input Parameters:

Select a sequence comparision matrix. This matrix determines how matches and mismatches are scored. The default penalites for gap creation and extension are given after each matrix name.

Scoring Matrix: blosum62

Set gap creation penalty: 8

Set gap extension penalty: 2

Limit the maximum input sequence range only when needed. Setting a higher limit allows you to align longer sequences while setting a lower limit allows you to add more and longer gaps to each sequence.

Maximum input sequence range: 5000 *(range 1 thru 7000)*

Limit the maximum number of gaps only when needed. Setting a higher limit allows you to add more and longer gaps to each sequence while setting a lower limit allows you to align a greater number of sequences.

Maximum number of gap characters ('.' and '~') added to any sequence: 2000 *(range 0 thru 7000)*

Display consensus only at completely conserved positions in the alignment: ☐

At each column in the alignment:
- show positions agreeing with the consensus in upper case ○
- display alignment only at positions that disagree with the consensus ○
- none of the above ◉

Count residues toward consensus that
- are identical ◉
- are similar ○
- are somewhat similar ○

Minimum number of votes required for a consensus: 2

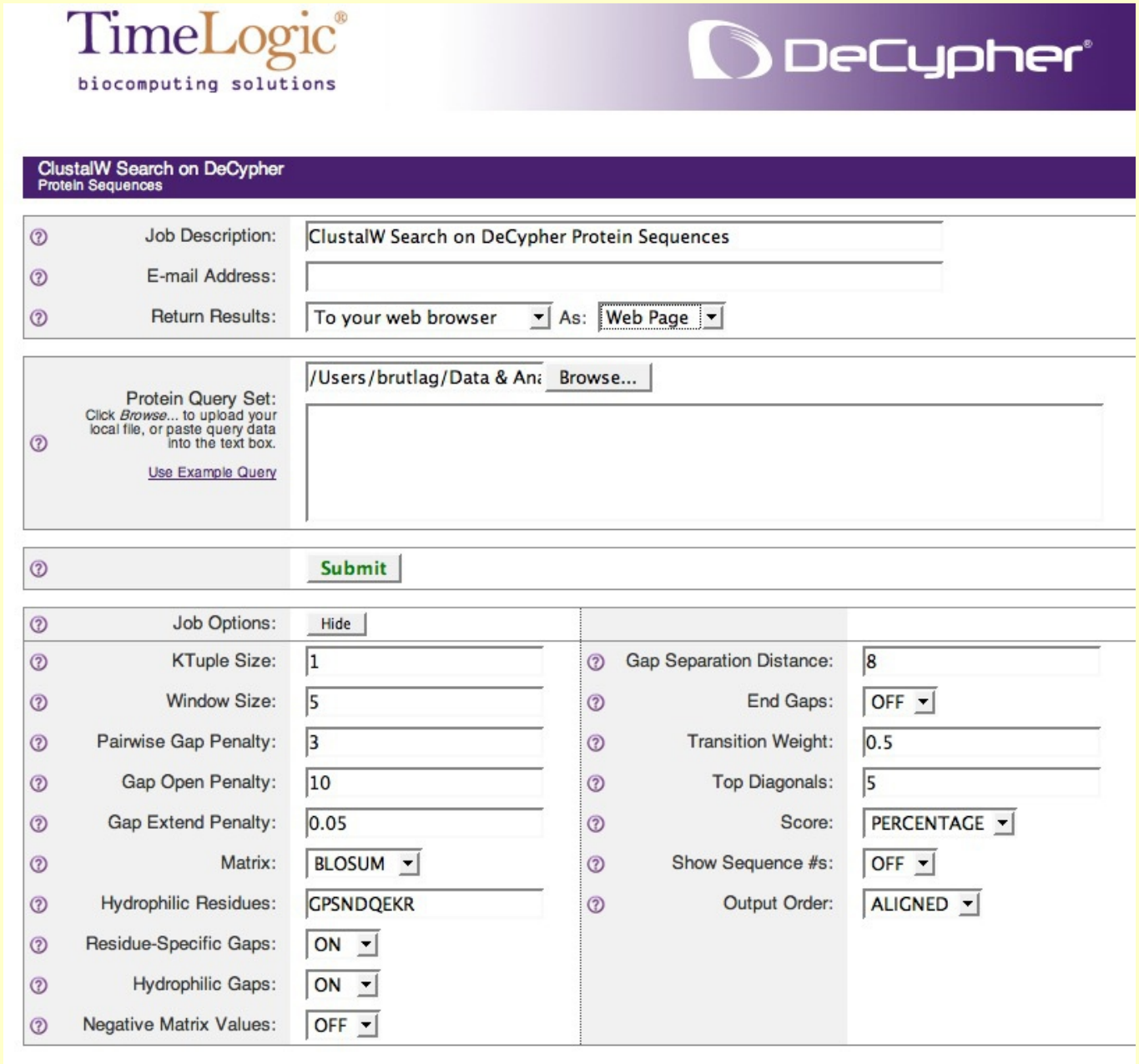
Run | Reset

Pretty Results

```
Plurality: 2.00  Threshold: 4  AveWeight 1.00  AveMatch 2.78  AvMisMat
 Consensus Sequence
 Symbol comparison table: share_matrix:blosum62.cmp  CompCheck: 1102
                  GapWeight: 8
              GapLengthWeight: 2
PileUp  MSF: 165  Type: P  January 28, 2010 10:22  Check: 6593 ..
 Name: hba_human          Len:    165  Check: 1231  Weight:  1.00
 Name: hba_horse.pep      Len:    165  Check: 2167  Weight:  1.00
 Name: hbb_horse.pep      Len:    165  Check: 9310  Weight:  1.00
 Name: hbb_human          Len:    165  Check:  208  Weight:  1.00
 Name: glb5_petma.pep     Len:    165  Check: 2079  Weight:  1.00
 Name: myg_phyca          Len:    165  Check: 4320  Weight:  1.00
 Name: lgb1_soybn.pep     Len:    165  Check: 7278  Weight:  1.00
//
                1                                                  50
   hba_human ~~~~~~~~~~v lspadktnvk aawgkvgaha geygaealer mflsfpttkt
   hba_horse ~~~~~~~~~~V LSAADKTNVK AAWSKVGGHA GEYGAEALER MFLGFPTTKT
   hbb_horse ~~~~~~~~~VQ LSGEEKAAVL ALWDKV..NE EEVGGEALGR LLVVYPWTQR
   hbb_human ~~~~~~~~~vh ltpeeksavt alwgkv..nv devggealgr llvvypwtqr
  glb5_petma PIVDTGSVAP LSAAEKTKIR SAWAPVYSTY ETSGVDILVK FFTSTPAAQE
   myg_phyca ~~~~~~~~~~v lsegewqlvl hvwakveadv aghgqdilir lfkshpetle
  lgb1_soybn ~~~~~~~~~ga ftekqealvs ssfeafkani pqysvvfyns ilekapaakd
   Consensus --------VV LS-AEKT-V- AAW-KVGAN- -EYG-EAL-R LF-S-P-T--

               51                                                 100
   hba_human yfphf.dlsh .....gsaqv kghgkkvada ltnavahvdd mpnalsalsd
   hba_horse YFPHF.DLSH .....GSAQV KAHGKKVGDA LTLAVGHLDD LPGALSNLSD
   hbb_horse FFDSFGDLSN PGAVMGNPKV KAHGKKVLHS FGEGVHHLDN LKGTFAALSE
   hbb_human ffesfgdlst pdavmgnpkv kahgkkvlga fsdglahldn lkgtfatlse
  glb5_petma FFPKFKGLTT ADQLKKSADV RWHAERIINA VNDAVASMDD TEKMSMKLRD
   myg_phyca kfdrfkhlkt eaemkasedl kkhgvtvlta lg...ailkk kghheaelkp
  lgb1_soybn lfsfla....  .gvdptnpkl tghaeklfal vrdsaggl.k tngtvvadaa
   Consensus FFP-F-DLST P-AV-GS-KV KAHGKKVLDA L-DAVAHLDD L-GT-AALSD

              101                                                 150
   hba_human l...hahklr vdpvnfklls hcllvtlaah lpaeftpavh asldkflasv
   hba_horse L...HAHKLR VDPVNFKLLS HCLLSTLAVH LPNDFTPAVH ASLDKFLSSV
   hbb_horse L...HCDKLH VDPENFRLLG NVLVVVLARH FGKDFTPELQ ASYQKVVAGV
   hbb_human l...hcdklh vdpenfrllg nvlvcvlahh fgkeftppvq aayqkvvagv
  glb5_petma LSGKHAKSFQ VDPQYFKVLA AVIADTVA.. ........AGD AGFEKLMSMI
   myg_phyca laqshatkhk ipikylefis eaiihvlhsr hpgdfgadaq gamnkalelf
  lgb1_soybn lvsihaqkav tdpq.fvvvk eallktikea vggnwsdels sawevaydel
   Consensus L---HA-KL- VDP-NFKLLS -VLLVTLA-H ---DFTPAVQ A---K-LA-V

              151              165
   hba_human stvltskyr~ ~~~~~~
   hba_horse STVLTSKYR~ ~~~~~~
   hbb_horse ANALAHKYH~ ~~~~~~
   hbb_human analahkyh~ ~~~~~~
  glb5_petma CILLRSAY~~ ~~~~~~
   myg_phyca rkdiaakyke lgyqg
  lgb1_soybn aaaikka~~~~ ~~~~~~
   Consensus A-ALASKY-- -----
```

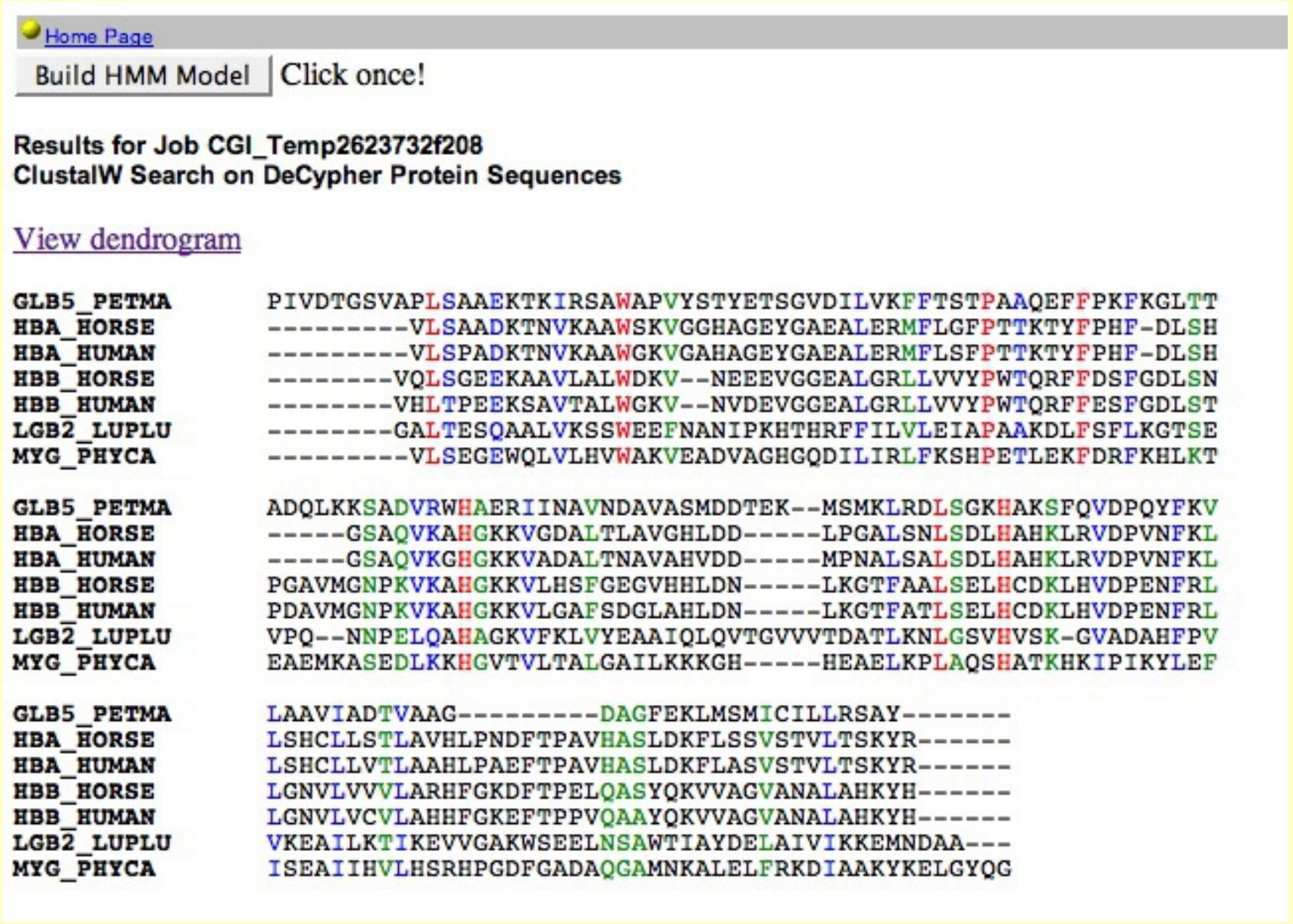
Decypher ClustalW Input

http://decypher.stanford.edu/



Decypher ClustalW Results

http://decypher.stanford.edu/

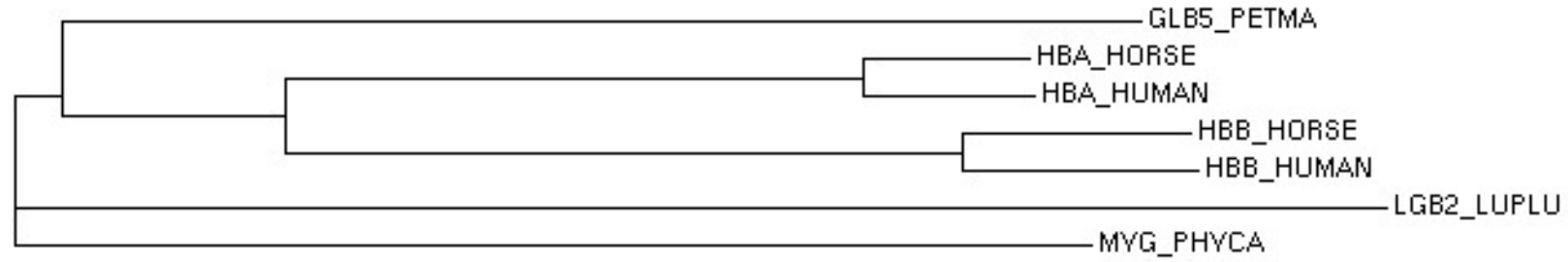


Decypher ClustalW Results

http://decypher.stanford.edu/

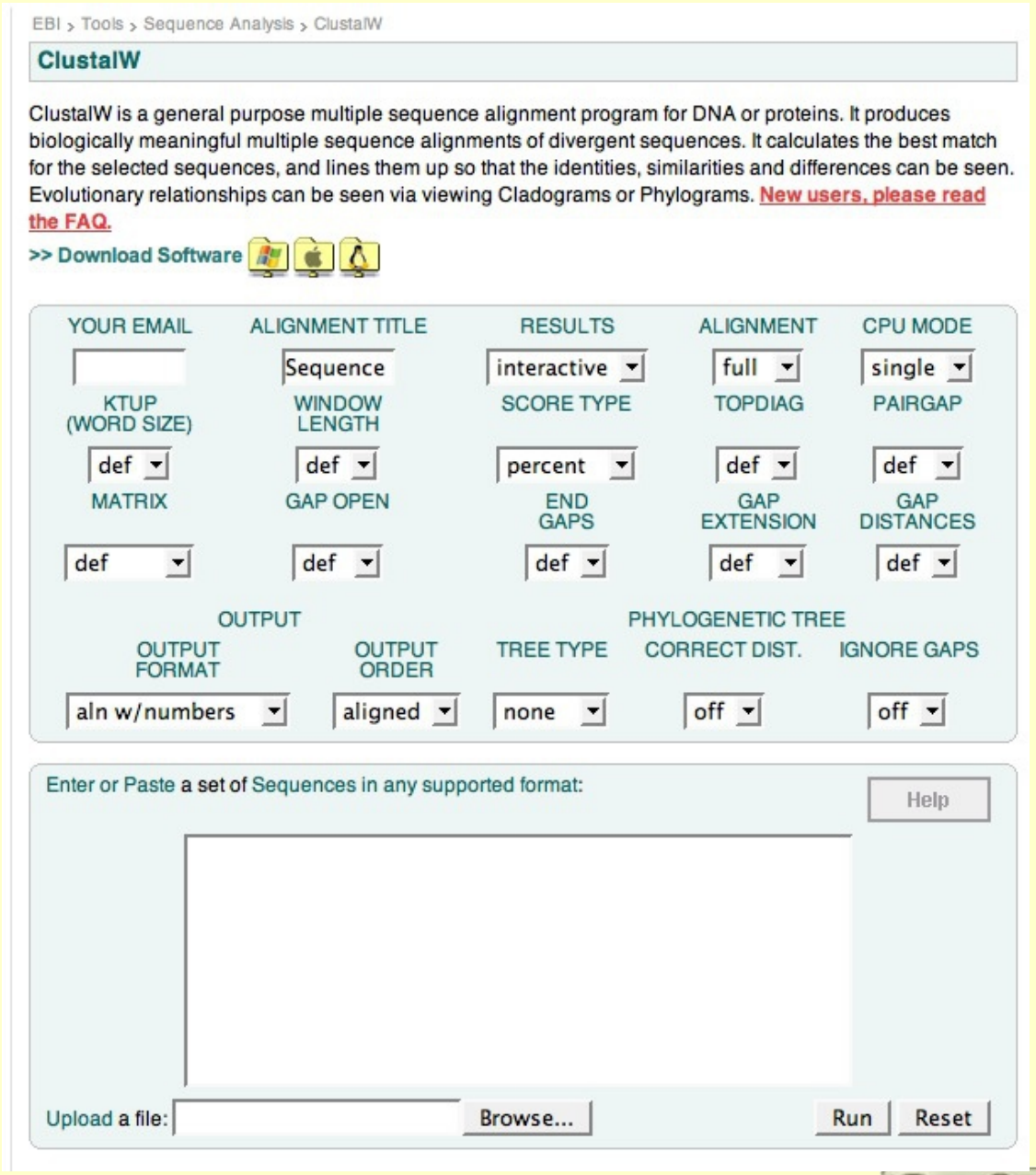
Dendrogram

Return to top



ClustalW @ EBI Input

http://www.ebi.ac.uk/clustalw/



ClustalW @ EBI Results

http://www.ebi.ac.uk/clustalw/

Alignment

Show Colors | View Alignment File

```
CLUSTAL W (1.83) multiple sequence alignment


HBA_HORSE       ---------VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHFDLS-- 49
HBA_HUMAN       ---------VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLS-- 49
HBB_HORSE       --------VQLSGEEKAAVLALWDKVN--EEEVGGEALGRLLVVYPWTQRFFDSFGDLSN 50
HBB_HUMAN       --------VHLTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYPWTQRFFESFGDLST 50
GLB5_PETMA      PIVDTGSVAPLSAAEKTKIRSAWAPVYSTYETSGVDILVKFFTSTPAAQEFFPKFKGLTT 60
MYG_PHYCA       ---------VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKT 51
LGB2_LUPLU      --------GALTESQAALVKSSWEEFNANIPKHTHRFFILVLEIAPAAKDLFSFLKGTSE 52
                         *:  :   :   *  .           :  .:   * :   *  :

HBA_HORSE       ----HGSAQVKAHGKKVGDALTLAVGHLDD-----LPGALSNLSDLHAHKLRVDPVNFKL 100
HBA_HUMAN       ----HGSAQVKGHGKKVADALTNAVAHVDD-----MPNALSALSDLHAHKLRVDPVNFKL 100
HBB_HORSE       PGAVMGNPKVKAHGKKVLHSFGEGVHHLDN-----LKGTFAALSELHCDKLHVDPENFRL 105
HBB_HUMAN       PDAVMGNPKVKAHGKKVLGAFSDGLAHLDN-----LKGTFATLSELHCDKLHVDPENFRL 105
GLB5_PETMA      ADQLKKSADVRWHAERIINAVNDAVASMDDT--EKMSMKLRDLSGKHAKSFQVDPQYFKV 118
MYG_PHYCA       EAEMKASEDLKKHGVTVLTALGAILKKKGH-----HEAELKPLAQSHATKHKIPIKYLEF 106
LGB2_LUPLU      VP--QNNPELQAHAGKVFKLVYEAAIQLQVTGVVVTDATLKNLGSVHVSKGVAD-AHFPV 109
                      . .:: *.  :   .                 :  *.  *  .       : .

HBA_HORSE       LSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVSTVLTSKYR------ 141
HBA_HUMAN       LSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR------ 141
HBB_HORSE       LGNVLVVVLARHFGKDFTPELQASYQKVVAGVANALAHKYH------ 146
HBB_HUMAN       LGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH------ 146
GLB5_PETMA      LAAVIADTVAAG---------DAGFEKLMSMICILLRSAY------- 149
MYG_PHYCA       ISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG 153
LGB2_LUPLU      VKEAILKTIKEVVGAKWSEELNSAWTIAYDELAIVIKKEMNDAA--- 153
                :   :  .:           ...       .   :
```

PLEASE NOTE: Showing colors on large alignments is slow.

Show Colors | View Alignment File

ClustalW @ EBI Results

http://www.ebi.ac.uk/clustalw/