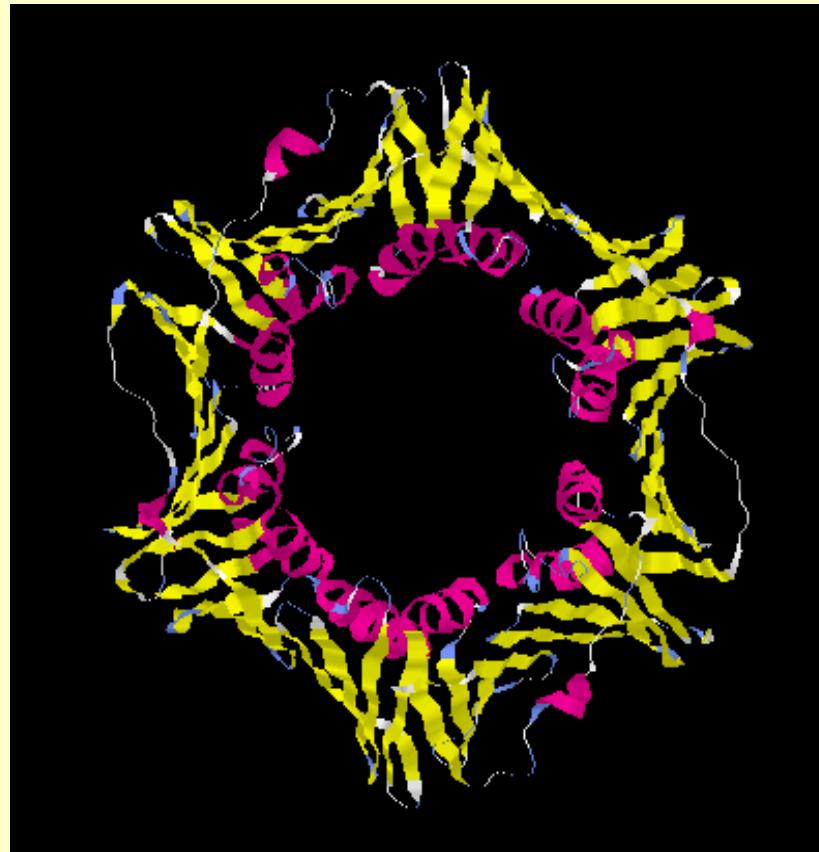


Computational Molecular Biology

Biochem 218 – BioMedical Informatics 231

<http://biochem218.stanford.edu/>

Multiple Sequence Alignment

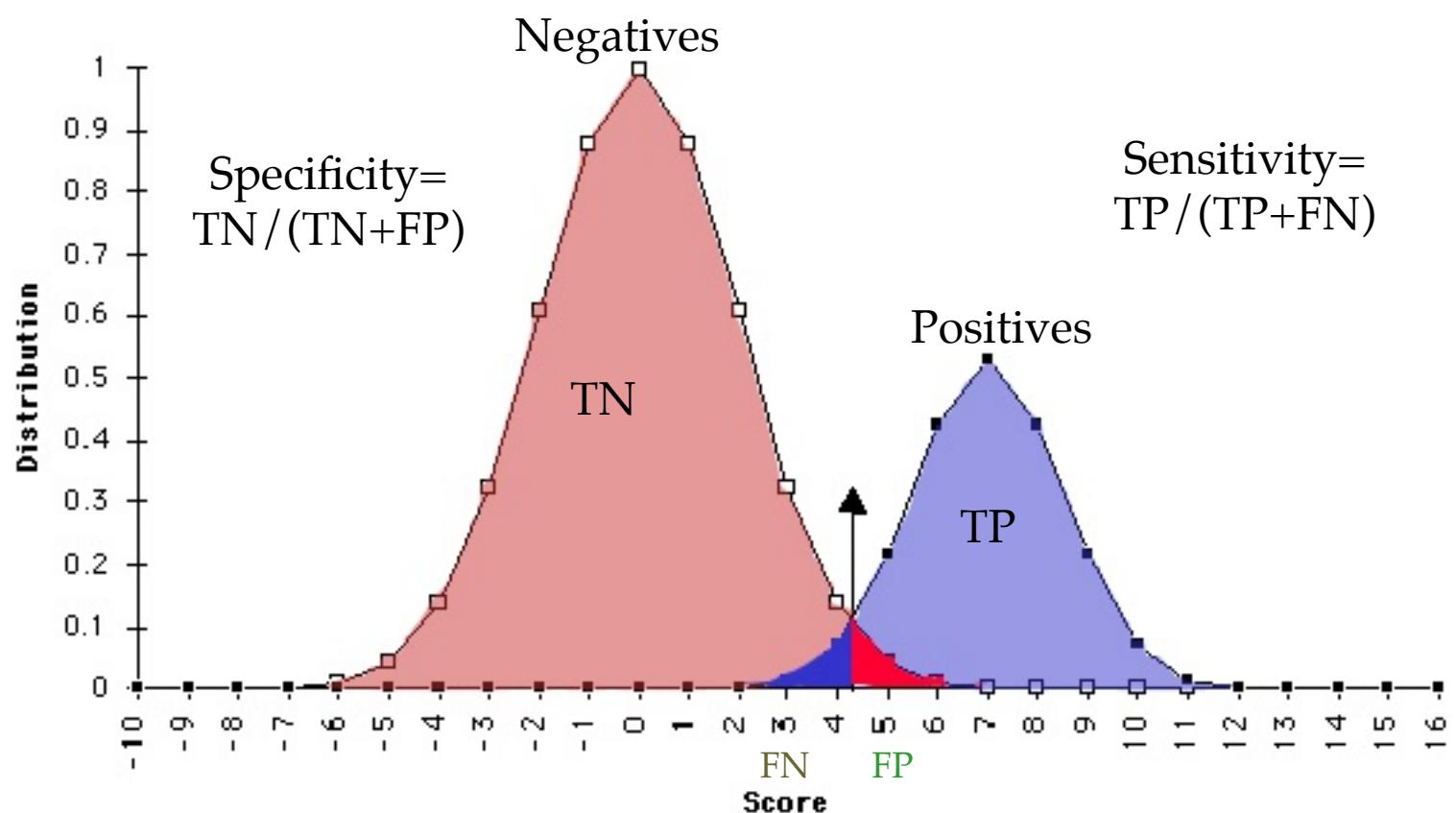


Doug Brutlag
Professor Emeritus
Biochemistry & Medicine (by courtesy)

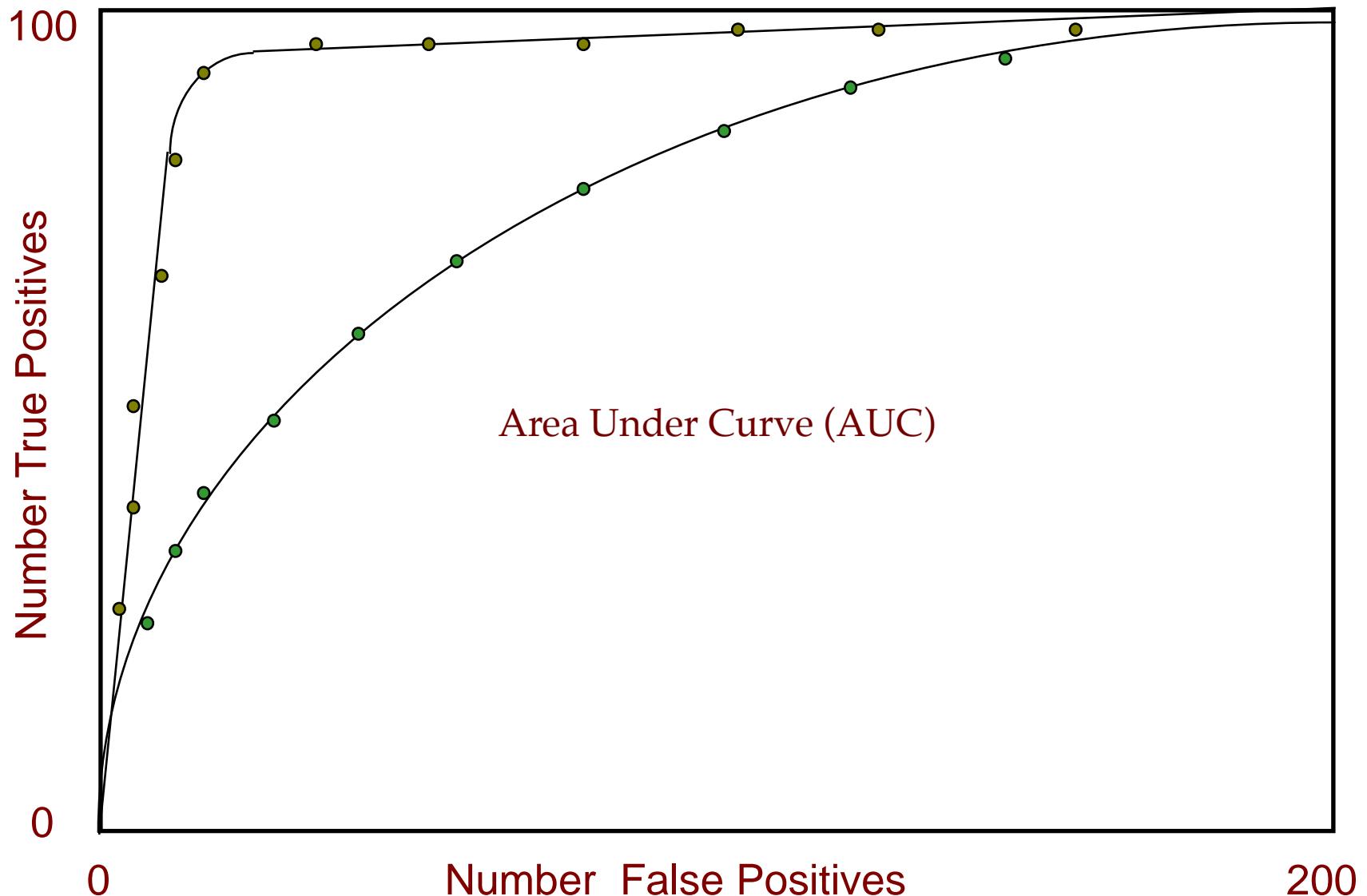


Doug Brutlag 2010

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/>

UniProtKB > UniProtKB

Downloads · Contact · Documentation/Help

Search Blast Align Retrieve ID Mapping *

Search in Query

Protein Knowledgebase (UniProtKB) 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

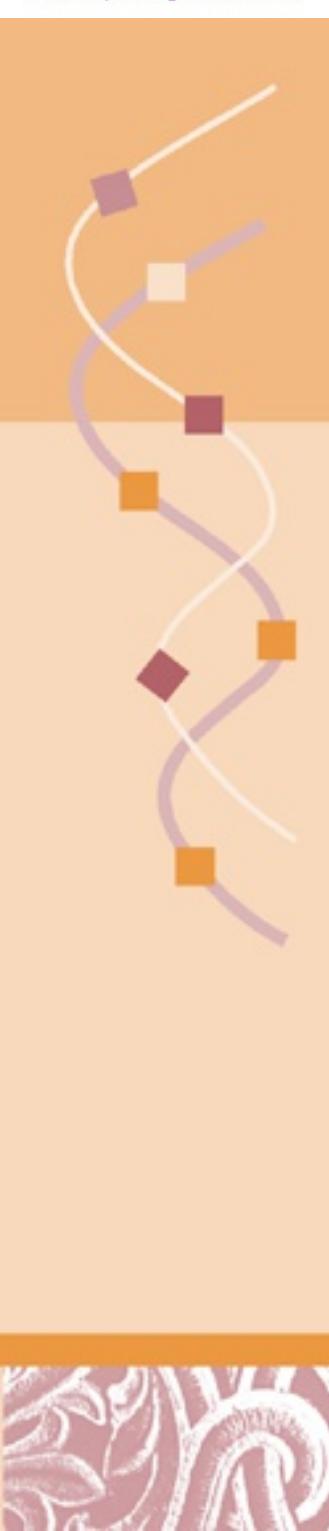
Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% | Customize display | Download...

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

Decypher Home Page

<http://decypher.stanford.edu/>



TimeLogic®
biocomputing solutions

Algorithm and Feature Index
The following links will take you to specific algorithm pages. ⓘ [On-line Product Documentation Set and Web Links](#)

Algorithm	Query vs. Database Types	Algorithm	Query vs. Database Types
Tera-Blast™ N	DNA to DNA ⓘ DNA to DNA ⓘ DNA to Protein ⓘ Protein to DNA ⓘ Protein to Protein ⓘ	Smith-Waterman <small>Standard, Double-Affine</small>	DNA to DNA ⓘ DNA to Protein ⓘ Protein to Protein ⓘ Protein to DNA ⓘ DNA to DNA ⓘ DNA to Protein ⓘ Protein to DNA ⓘ DNA to Protein HMM ⓘ
Tera-Blast™ P		FrameSearch <small>Symmetric Frame Independent™ for DNA to DNA</small>	DNA to DNA ⓘ DNA to Protein ⓘ Protein to DNA ⓘ DNA to Protein HMM ⓘ Protein to Protein HMM ⓘ Protein HMM to Protein ⓘ Protein HMM to DNA ⓘ DNA to Protein HMM ⓘ Protein HMM to DNA ⓘ DNA to Protein Profile ⓘ Protein To Protein Profile ⓘ Protein Profile to Protein ⓘ Protein Profile to DNA ⓘ DNA to Protein Profile ⓘ Protein Profile to DNA ⓘ
Tera-Probe™	DNA to DNA ⓘ Genomic DNA to Coding DNA ⓘ Coding DNA to Genomic DNA ⓘ Genomic DNA to Protein ⓘ Protein to Genomic DNA ⓘ	Hidden Markov Model (HMM)	
GeneDetective™	Genomic DNA to Protein HMM ⓘ Protein HMM to Genomic DNA ⓘ	HMM FrameSearch	
ClustalW	DNA ⓘ Protein ⓘ	ProfileSearch	
		Profile FrameSearch	

Decypher Search Input

<http://decypher.stanford.edu/>



DeCypher Smith-Waterman Search Protein Query vs. Protein Database

Job Description: DeCypher Smith-Waterman Search Protein Query vs. Protein Database

E-mail Address:

Return Results: To your web browser As: Web Page

Protein Query:
Click Browse... to upload your local file, or paste query data into the text box. Use only one query entry if requesting a Graphic reply.

Use Example Query

Protein Database:
Select only one if requesting an iterated search.

Browse...

```
>PDH E1 HUMAN
MRKMLAAVSRVLSGASQKQPASRVLVASRNPFANDATFEIKKCDLHRLEEGPPVTTLTRED
GLKYYRMMQTVRRMELKADOLYKOKIIRGFCHLCGQEAACVGLEAGINPTDHЛИTAYRA
HGFTFTRGLSVREILAELETGRKGCCAKGGSMHYAKNFYGGNGIVGAQVPLGAGIALA
CKYNGKDEVCLTLYGDGAANQGQIFEAYNMAALWKLPCIFICENNRYGMGTSVERAAAST
DYKRGDFIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRYHGHMSDPGVs
```

Description	Entries	Symbols	Updated
Thermotoga maritima (NCB	1858	593368	19-March-2003
Treponema pallidum (NCBI	1036	356216	19-March-2003
tumora	1	393	19-December-2004
Uniprot-SwissProt: 57.1	514212	180900945	26-January-2010
Uniprot-TREMBL: 57.13	9145906	2958343669	27-January-2010
<i>Escherichia coli K12</i>	614	231600	16-May-2003

Submit

Job Options: Hide

Algorithm Variation: Local (Standard)

Filter Query:

Max Scores: 100

Max Alignments: 20

Identity Symbol: Match Letter

Show Significance: E-Value

Gap Open Penalty: 12

Gap Extend Penalty: 2

Score Threshold: 1

Significance Threshold: 10

Weight Matrix: blosum62

Group Results by Entry:

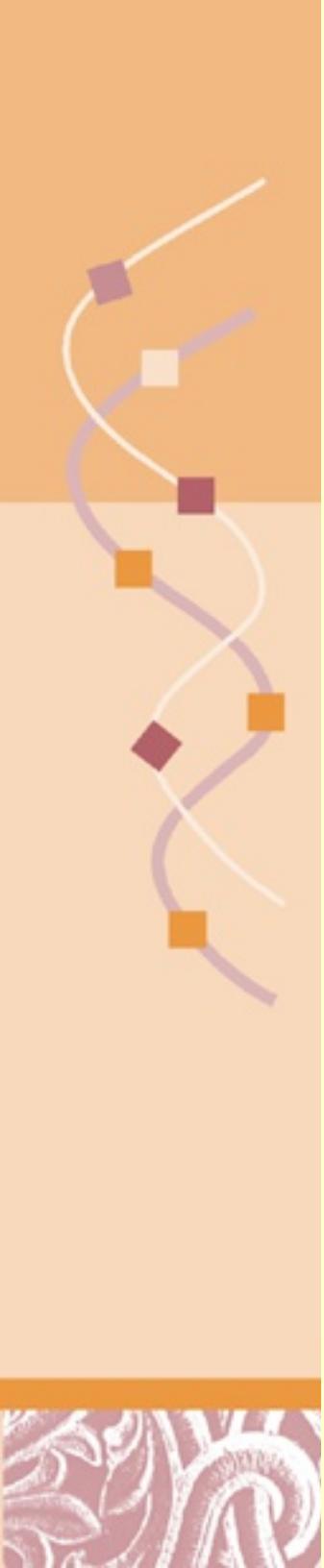
Smith-Waterman Search Protein Query vs. Protein Data

sp P29803 ODPAT_HUMAN	388	sp P29803 ODPAT_HUMAN	Pyruvate dehydrogenase E1 component subunit alpha
5.4e-272	sp P29803 ODPAT_HUMAN	Pyruvate dehydrogenase E1 component subunit alpha	
8.7e-234	sp P35486 ODPA_MOUSE	Pyruvate dehydrogenase E1 component subunit alpha	
2.3e-233	sp P08559 ODPA_HUMAN	Pyruvate dehydrogenase E1 component subunit alpha	
3.1e-233	sp R7MB35 ODPA_BOVIN	Pyruvate dehydrogenase E1 component subunit alpha	
1.1e-232	sp P26284 ODPA_RAT	Pyruvate dehydrogenase E1 component subunit alpha	
2.1e-232	sp A5A6L0 ODPA_PANTR	Pyruvate dehydrogenase E1 component subunit alpha	
2.1e-232	sp Q5R490 ODPA_PONAB	Pyruvate dehydrogenase E1 component subunit alpha	
2.9e-232	sp P29804 ODPA_PIG	Pyruvate dehydrogenase E1 component subunit alpha	
9.4e-231	sp Q8HGW9 ODPA_MACFR	Pyruvate dehydrogenase E1 component subunit alpha	
8.7e-218	sp P52900 ODPA_SMIMA	Pyruvate dehydrogenase E1 component subunit alpha	
1.4e-211	sp P35487 ODPAT_MOUSE	Pyruvate dehydrogenase E1 component subunit alpha	
9.4e-211	sp Q06437 ODPAT_RAT	Pyruvate dehydrogenase E1 component subunit alpha	
8.7e-146	sp P52899 ODPA_CAEEL	Probable pyruvate dehydrogenase E1 component subunit alpha	
3.6e-135	sp P26268 ODPT_ASCSU	Pyruvate dehydrogenase E1 component subunit alpha	
5.9e-133	sp P26267 ODPA_ASCSU	Pyruvate dehydrogenase E1 component subunit alpha	
5.0e-127	sp Q54C70 ODPA_DICDI	Pyruvate dehydrogenase E1 component subunit alpha	
1.0e-119	sp P16387 ODPA_YEAST	Pyruvate dehydrogenase E1 component subunit alpha	
1.5e-116	sp P52901 ODPA1_ARATH	Pyruvate dehydrogenase E1 component subunit alpha	
5.4e-116	sp Q13366 ODPA_KLULR	Pyruvate dehydrogenase E1 component subunit alpha	
1.4e-115	sp Q10489 ODPA_SCHPO	Pyruvate dehydrogenase E1 component subunit alpha	
5.0e-115	sp Q8H1Y0 ODPA2_ARATH	Pyruvate dehydrogenase E1 component subunit alpha	
1.8e-114	sp P52902 ODPA_PEA	Pyruvate dehydrogenase E1 component subunit alpha	
3.1e-113	sp P52903 ODPA_SOLTU	Pyruvate dehydrogenase E1 component subunit alpha	
1.9e-111	sp Q9R9H5 ODPA_RHIME	Pyruvate dehydrogenase E1 component subunit alpha	
6.9e-091	sp Q66112 ODPA_ZYMMO	Pyruvate dehydrogenase E1 component subunit alpha	
5.4e-088	sp Q4UKQ6 ODPA_RICFE	Pyruvate dehydrogenase E1 component subunit alpha	
3.1e-085	sp Q92IS3 ODPA_RICCH	Pyruvate dehydrogenase E1 component subunit alpha	
4.3e-085	sp Q1RJX4 ODPA_RICBR	Pyruvate dehydrogenase E1 component subunit alpha	
1.1e-084	sp Q68XA9 ODPA_RICTY	Pyruvate dehydrogenase E1 component subunit alpha	
7.5e-084	sp Q9ZDR4 ODPA_RICPR	Pyruvate dehydrogenase E1 component subunit alpha	
8.1e-069	sp Q1XDM0 ODPA_PORYE	Pyruvate dehydrogenase E1 component subunit alpha	
6.4e-066	sp P51267 ODPA_PORPU	Pyruvate dehydrogenase E1 component subunit alpha	
3.7e-063	sp P27745 ACOA_RALEH	Acetooin:26-dichlorophenol lindophenol oxidoreductase	
4.3e-061	sp Q31404 ACOA_BACSU	Acetooin:26-dichlorophenol lindophenol oxidoreductase	
1.0e-035	sp P37940 ODBA_BACSU	2-oxoisovalerate dehydrogenase subunit alpha	
2.3e-033	sp Q72GU1 ODBA_THET2	2-oxoisovalerate dehydrogenase subunit alpha	
8.1e-033	sp Q5SLR4 ODBA_THET8	2-oxoisovalerate dehydrogenase subunit alpha	
1.5e-032	sp P35485 ODPA_ACHLA	Pyruvate dehydrogenase E1 component subunit alpha	
5.4e-032	sp Q54M22 ODBA_DICDI	2-oxoisovalerate dehydrogenase subunit alpha	
1.0e-031	sp P47516 ODPA_MYCGE	Pyruvate dehydrogenase E1 component subunit alpha	
1.9e-031	sp Q5HG21 ODPA_STAAC	Pyruvate dehydrogenase E1 component subunit alpha	
1.9e-031	sp P60089 ODPA_STARM	Pyruvate dehydrogenase E1 component subunit alpha	
1.9e-031	sp Q820A6 ODPA_STRAAH	Pyruvate dehydrogenase E1 component subunit alpha	
1.9e-031	sp Q6GAC1 ODPA_STARS	Pyruvate dehydrogenase E1 component subunit alpha	



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

- Choose 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



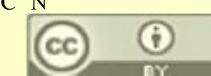
Goals of Multiple Sequence Alignment

- 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

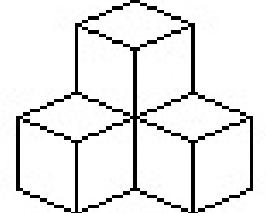
Consensus Sequence From a Multiple Sequence Alignment

ClustalW Insulin Alignments

	10	20	30
<i>IPGP</i>			
<i>IPDK</i>			
<i>IPDG</i>	M A L W M R L L P L L A L L A L W A P A P T R A F V N Q H		
<i>IPCH</i>	M A L W I R S L P L L A L L V F S G P G - T S Y A A N Q H		
<i>IPCA</i>	M A V W I Q A G A L L F L L A V S S V N A N A G A P - Q H		
<i>IPBO</i>			F V N Q H
<i>IPAF</i>	M A A L W L Q S F S L L V L L V V S W P G S Q A V A P A Q H		
	A . W . .	L L L L	A N Q H
	40	50	60
<i>IPGP</i>	L C G S N L V E T L Y S V C Q D D G F F Y I P K D X X E L E		
<i>IPDK</i>	L C G S H L V E A L Y L V C G E R G F F Y S P K T T X X D V E		
<i>IPDG</i>	L C G S H L V E A L Y L V C G E R G F F Y T P K A R R E V E		
<i>IPCH</i>	L C G S H L V E A L Y L V C G E R G F F Y S P K A R R D V E		
<i>IPCA</i>	L C G S H L V D A L Y L V C G P T G F F Y N P K R D V D P P		
<i>IPBO</i>	L C G S H L V E A L Y L V C G E R G F F Y T P K A R R E V E		
<i>IPAF</i>	L C G S H L V D A L Y L V C G D R G F F Y N P K R D V D Q L		
	L C G S H L V E A L Y L V C G E R G F F Y . P K .	D V E	
	70	80	90
<i>IPGP</i>	D P Q V E Q T E L G M G - - - - L G A G G L Q P - - L Q G		
<i>IPDK</i>	Q P - L V N G P L H G E - - - - V G E L P F Q - - - H E		
<i>IPDG</i>	D L Q V R D V E L A G A - - - - P G E G G L Q P L A L E G		
<i>IPCH</i>	Q P - L V S S P L R G E - - - - A G V L P F Q - - - Q E		
<i>IPCA</i>	L G F L P P K S - - - - A Q E T E V A D F A F K D H A E		
<i>IPBO</i>	G P Q V G A L E L A G G - - - - P G A G G L E - - - G		
<i>IPAF</i>	L G F L P P K S G G A A A A G A D N E V A E F A F K D Q M E		
	P L L G G	G F Q	E
	100	110	120
<i>IPGP</i>	A L Q X X - - G I V D Q C C T G T C T R H Q L Q S Y C N		
<i>IPDK</i>	E Y Q X X - - G I V E Q C C E N P C S L Y Q L E N Y C N		
<i>IPDG</i>	A L Q K R - - G I V E Q C C T S I C S L Y Q L E N Y C N		
<i>IPCH</i>	E Y E K V K R G I V E Q C C H N T C S L Y Q L E N Y C N		
<i>IPCA</i>	V I R K R - - G I V E Q C C H K P C S I F E L Q N Y C N		
<i>IPBO</i>	P P Q K R - - G I V E Q C C A S V C S L Y Q L E N Y C N		
<i>IPAF</i>	M M V K R - - G I V E Q C C H R P C N I F D L Q N Y C N		
	. Q K R G I V E Q C C C S L Y Q L E N Y C N		



BLOCK MARKER Makes a Multiple Sequence Alignment



BLOCKS

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

10-45

25-55

40

NLQGYMLGNP
NFMGYMVNG
NLKGFLVGNA
NLKGILIGNA
NLKGFAIGNG
NFKGYLVNG
NLKGFIIVGNP
NIKGYIQGNA
NLKGFMIGNA
NLQGYILGNP
NFKGFMVGN
NLQGYVLGNP

PLLLWLNGGPGCSSLIGYGASEEIG
PLVLWFNGGPGCSSLVGFGAEEELG
PLMIWLTGGPGCSGLSSFVYEIGP
PLMIWLTGGPGCSGLSTFLYEFGP
PLLLWLSSGGPGCSSLTGLLFENG
PLVLWLNGGPGCSSLVAYGAAEEIG
PVVIWLTGGPGCSSELALFYENGP
PLVIWFNGGPGCSSLGGAFKELGP
PLVIWFNGGPGACSSLGGAFLELG
PLVLWLNGGPGCSSLIGAFQELGP
PLVLWLNGGPGCSSIAYGASEEVG
PLTLWLNGGPGCSSLVGGAFTELG

TVKQWSGYMDYKDS
GVNQYSGYLSVGSN
SFAHYAGYVTVSED
DFAQYAGYVTVDAA
DLGHHAGYYKLPKS
SVESYSGFMTVDAK
GVKSYTGYLLANAT
NFKQYSGYYNVGTK
NFKSYSGYVDANAN
NFKHYSGFFQVSDN
DFFHYSGYLRAWTD
TVKQYTGYLDVEDD

Sequence Profiles

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

	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	<h3>HmmерPfam</h3> <p>Identify known hidden Markov model domains.</p> <p><i>Input sequences:</i></p> <table border="1"> <thead> <tr> <th>Sequence</th> <th>Description</th> <th>Type</th> <th>Length</th> <th>Range</th> </tr> </thead> <tbody> <tr><td>odpat_human.uniprot_sprot</td><td>odpat_human</td><td>P</td><td>388</td><td>1 .. 388</td></tr> <tr><td>odpa_human.uniprot_sprot</td><td>odpa_human</td><td>P</td><td>390</td><td>1 .. 390</td></tr> <tr><td>odpt_ascsu.uniprot_sprot</td><td>odpt_ascsu</td><td>P</td><td>391</td><td>1 .. 391</td></tr> <tr><td>odpa_yeast.uniprot_sprot</td><td>odpa_yeast</td><td>P</td><td>420</td><td>1 .. 420</td></tr> <tr><td>odpa_rat.uniprot_sprot</td><td>odpa_rat</td><td>P</td><td>390</td><td>1 .. 390</td></tr> <tr><td>odpa_mouse.uniprot_sprot</td><td>odpa_mouse</td><td>P</td><td>390</td><td>1 .. 390</td></tr> <tr><td>odpa_caeel.uniprot_sprot</td><td>odpa_caeel</td><td>P</td><td>397</td><td>1 .. 397</td></tr> <tr><td>odpat_rat.uniprot_sprot</td><td>odpat_rat</td><td>P</td><td>391</td><td>1 .. 391</td></tr> <tr><td>odpa_macfa.uniprot_sprot</td><td>odpa_macfa</td><td>P</td><td>390</td><td>1 .. 390</td></tr> <tr><td>odpa_smima.uniprot_sprot</td><td>odpa_smima</td><td>P</td><td>363</td><td>1 .. 363</td></tr> <tr><td>odpa_ponpy.uniprot_sprot</td><td>odpa_ponpy</td><td>P</td><td>390</td><td>1 .. 390</td></tr> <tr><td>odpa_ascsu.uniprot_sprot</td><td>odpa_ascsu</td><td>P</td><td>396</td><td>1 .. 396</td></tr> <tr><td>odpa_pig.uniprot_sprot</td><td>odpa_pig</td><td>P</td><td>389</td><td>1 .. 389</td></tr> <tr><td>odpat_mouse.uniprot_sprot</td><td>odpat_mouse</td><td>P</td><td>391</td><td>1 .. 391</td></tr> </tbody> </table> <p>Refresh Clear</p> <p><i>Input Parameters:</i></p> <p>HMM Database: Pfam ▼</p> <p>Number of processors to use: 1 ▼</p> <p>Turn the XNU filtering of target protein sequences on <input type="checkbox"/></p> <p>Limit alignment output of best scoring domains to 50 (range 1 thru 1000)</p> <p>Set E-value cutoff for the per-sequence ranked hit list 10.0 (range 0.0 thru 100.0)</p> <p>Set bit score cutoff for the per-sequence ranked hit list -300 (range -500.0 thru 500.0)</p> <p>Set E-value cutoff for the per-domain ranked hit list 2.0 (range 0.0 thru 100.0)</p> <p>Set bit score cutoff for the per-domain ranked hit list 10 (range -1000.0 thru 1000.0)</p> <p>Run Reset</p>							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
	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																																																																														



SeqWeb Sequence Profile Search

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

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
lgb1_soybn	lgb1_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 comparison matrix. This matrix determines how matches and mismatches are scored. The default penalties 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

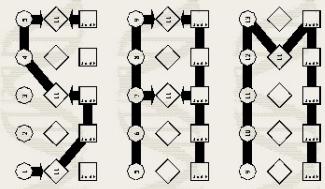
Set maximum gap extension penalty

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

Lowest Z score to report in output list

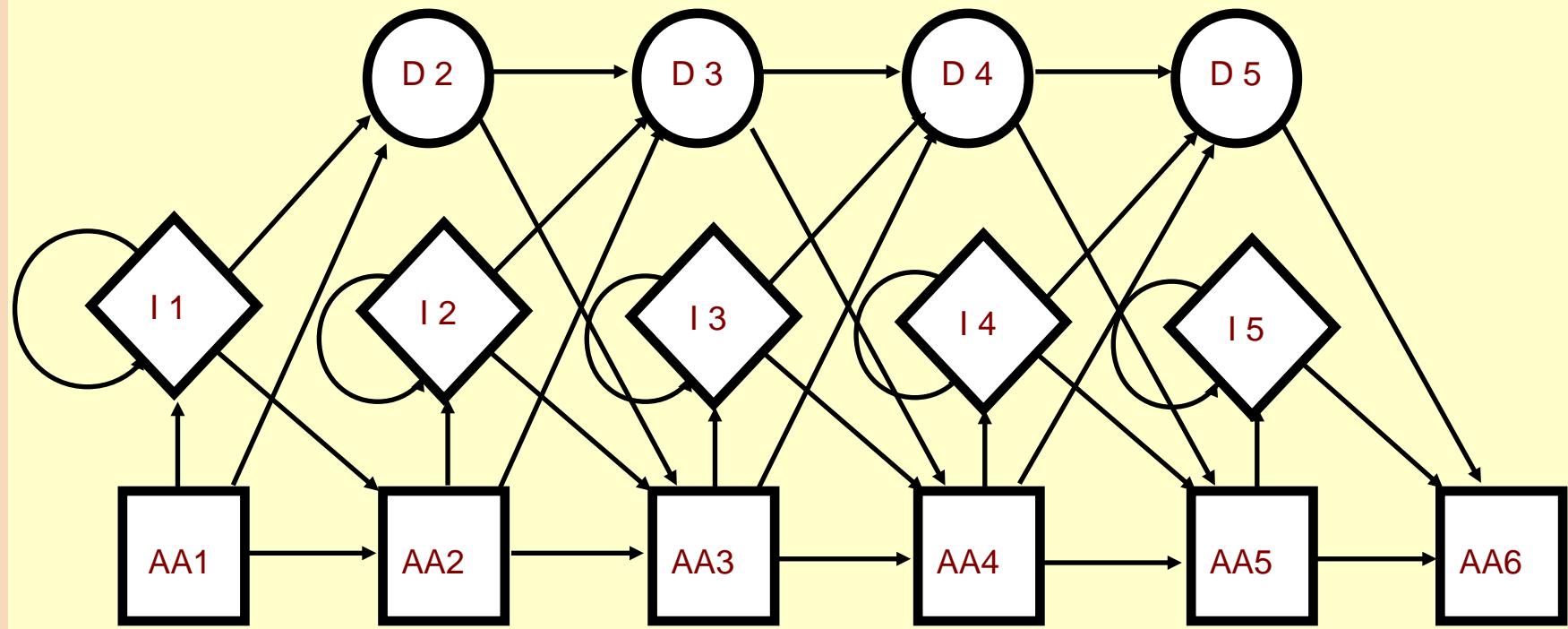
2.5 (range 0.0 thru 100)



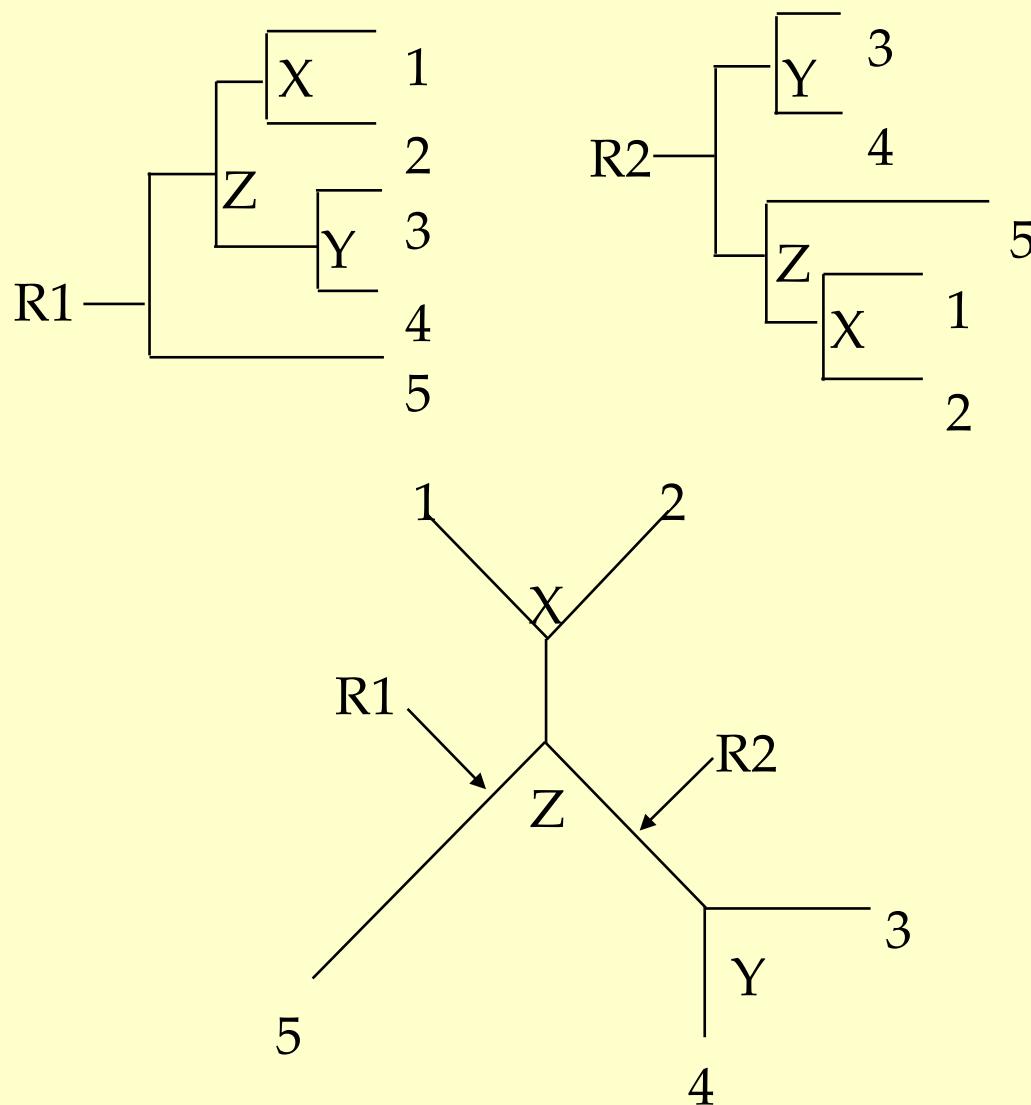


Hidden Markov Models

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



Evolutionary Trees

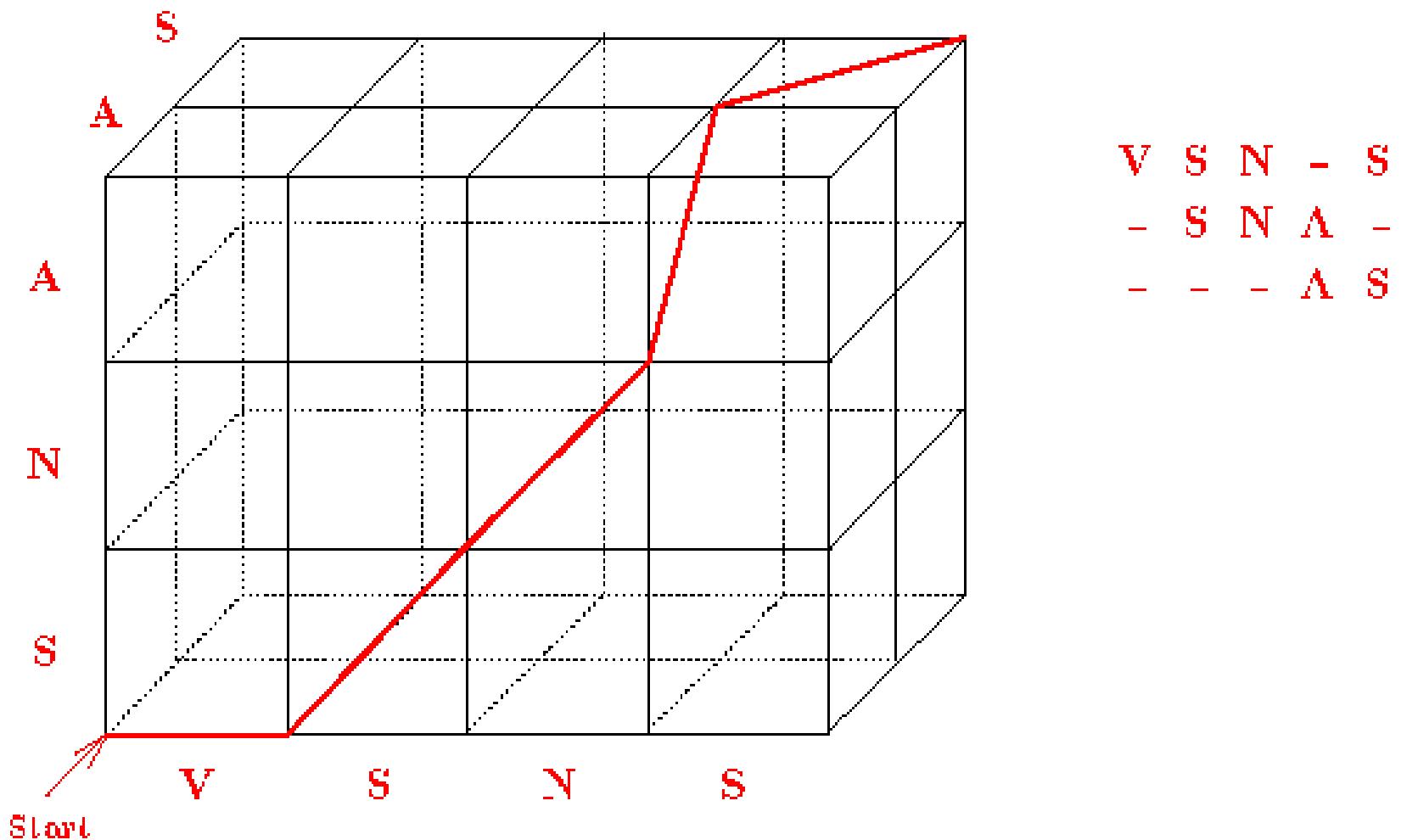


Challenges Aligning Multiple Sequences

- 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

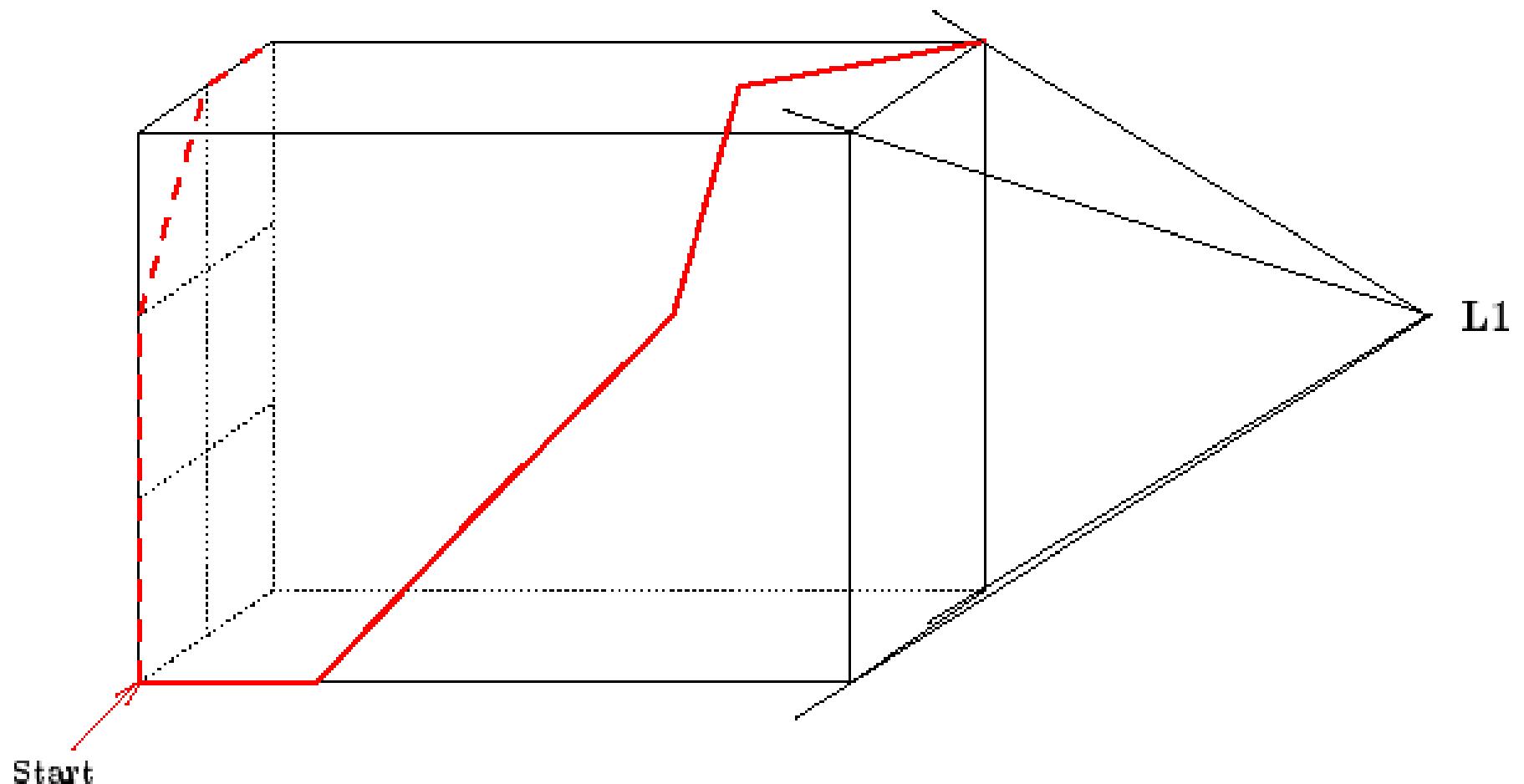
Three Protein Alignment (Murata, Richardson & Sussman)

Figure 1: Alignment Path for 3 Sequences.



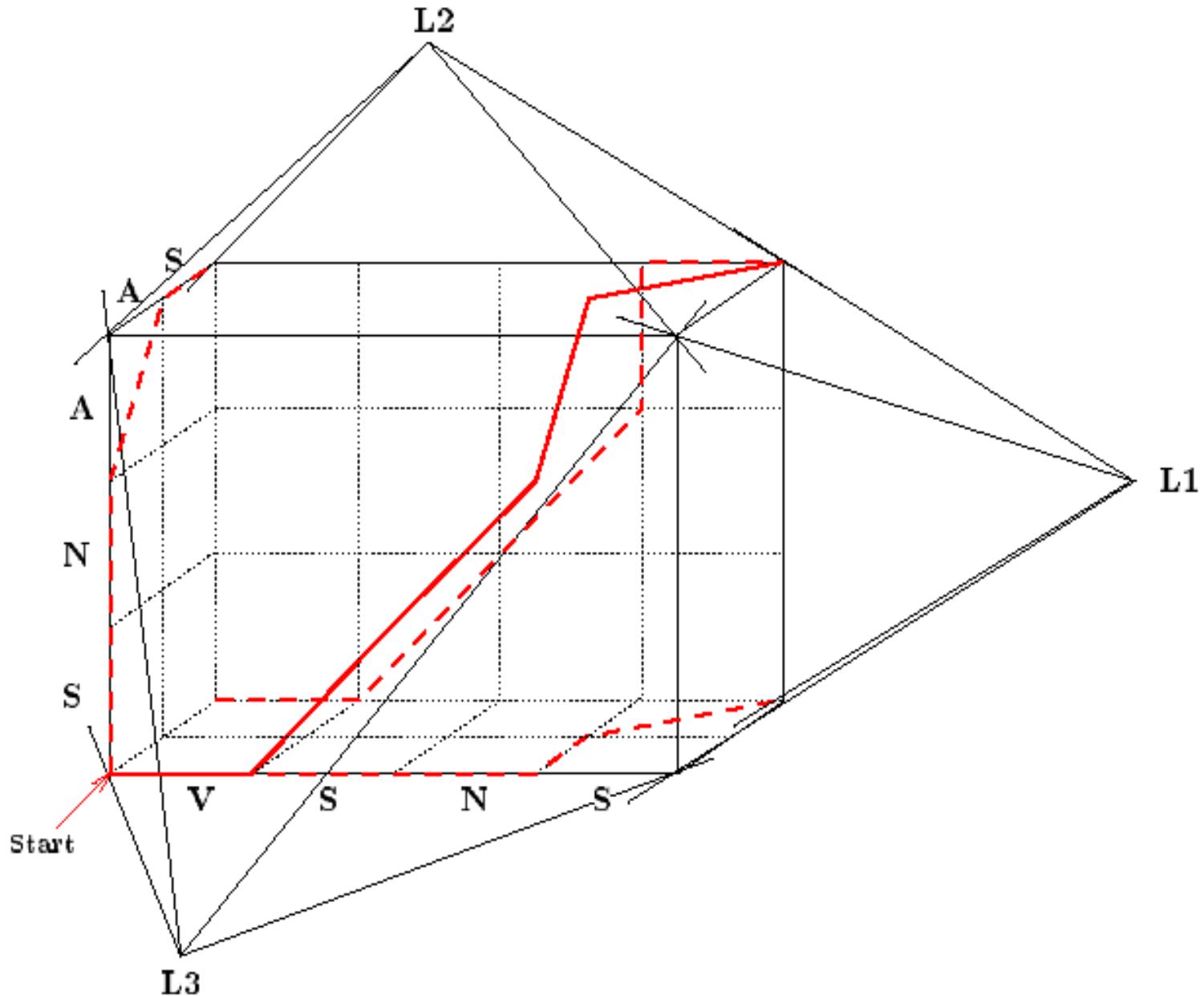
One Pairwise Alignment from the Three-Way Alignment

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



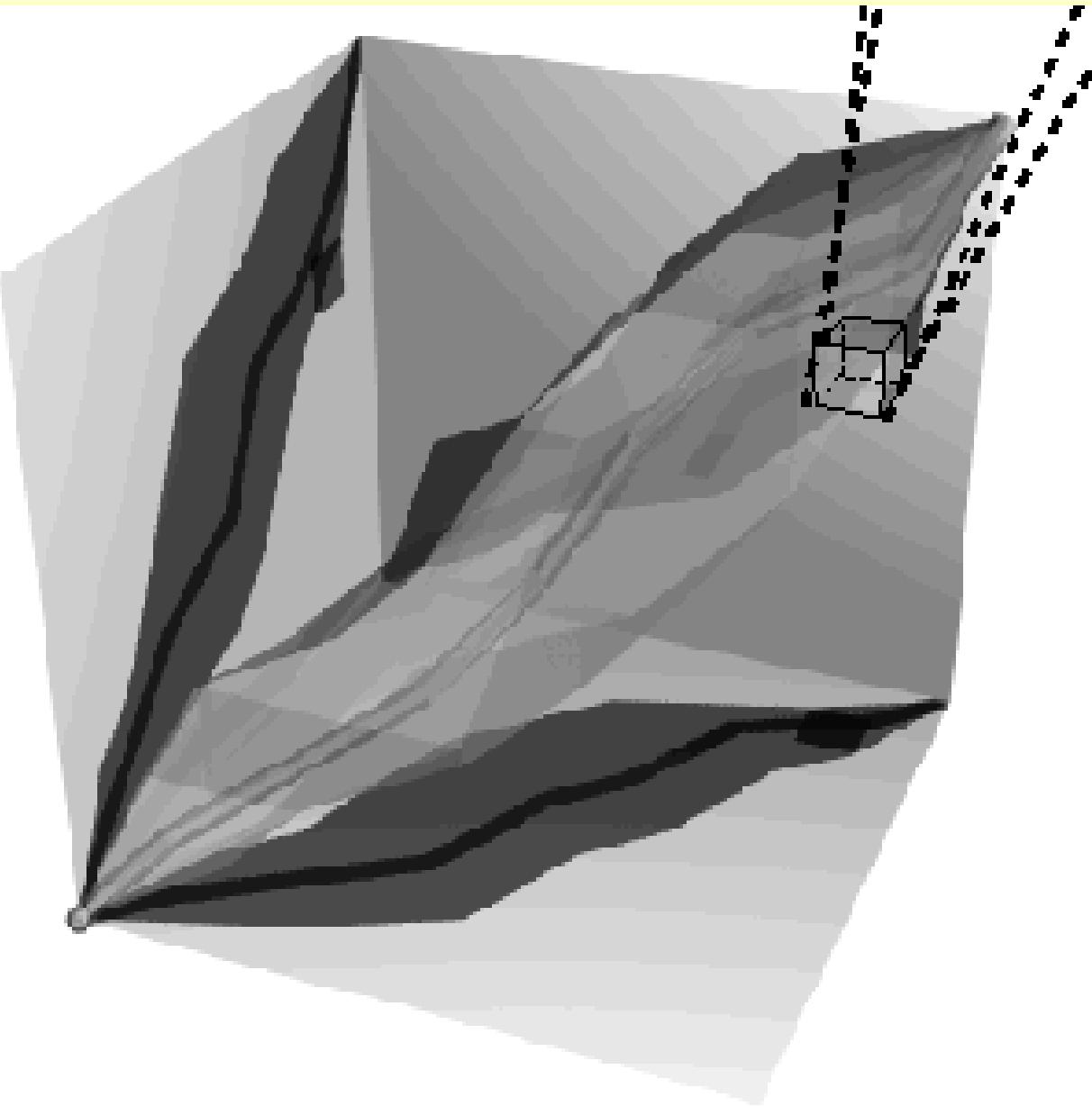
All Pairwise Alignments from the Three-Way Alignment

Figure 3: All 3 Pairwise Projections of the Alignment



Carrillo-Lipman Limits for MSA

<http://searchlauncherbcm.tmc.edu/multi-align/multi-align.html>



Clustal Progressive Alignment (Step 1)

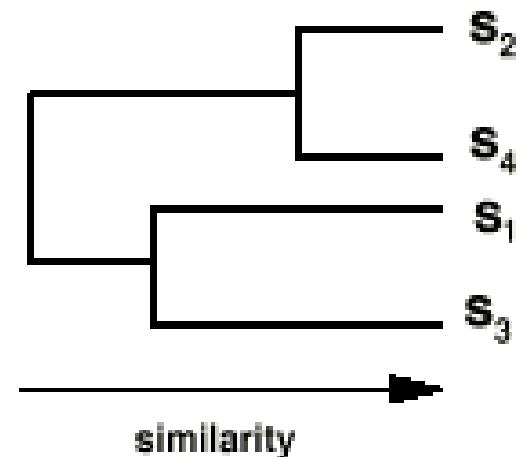
Steps in Multiple Alignment

(A) Pairwise Alignment

Example - 4 sequences S_1, S_2, S_3, S_4

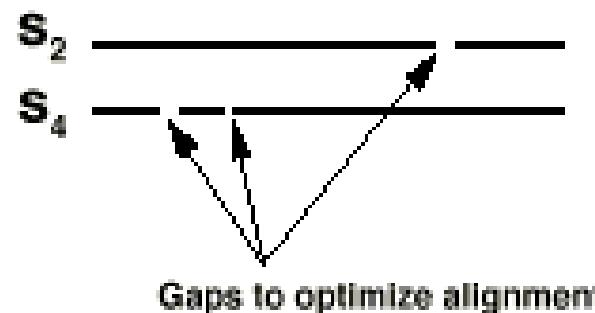
S_1 _____
 S_2 _____
 S_3 _____
 S_4 _____

6 pairwise comparisons
then cluster analysis

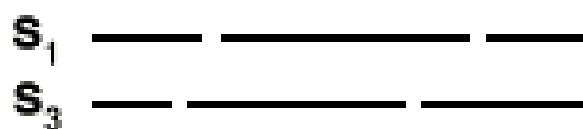


Clustal Progressive Alignment (Step 2)

(B) Multiple alignment following the tree from A

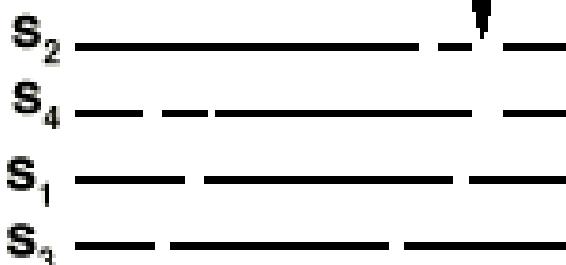


align most similar pair

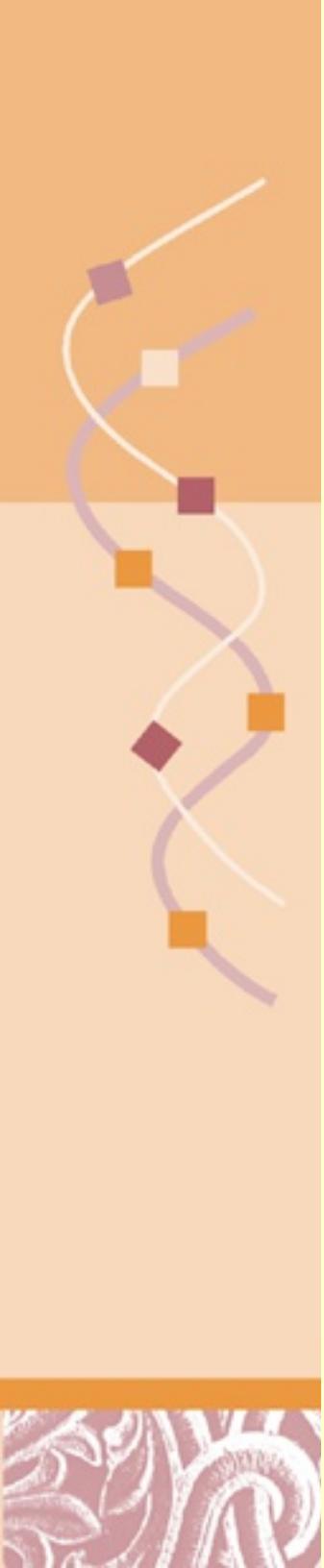


align next most similar pair

New gap to optimize
alignment of (S_2 S_4) with (S_1 S_3)



align alignments - preserve gaps



Gaps Are Propagated To Make Alignment

SEQ2 : MQQL- DNPYIVRMIGICEAE-SWM

SEQ1 : MGQF- DHPNITIRLEGVVTKSRPVM

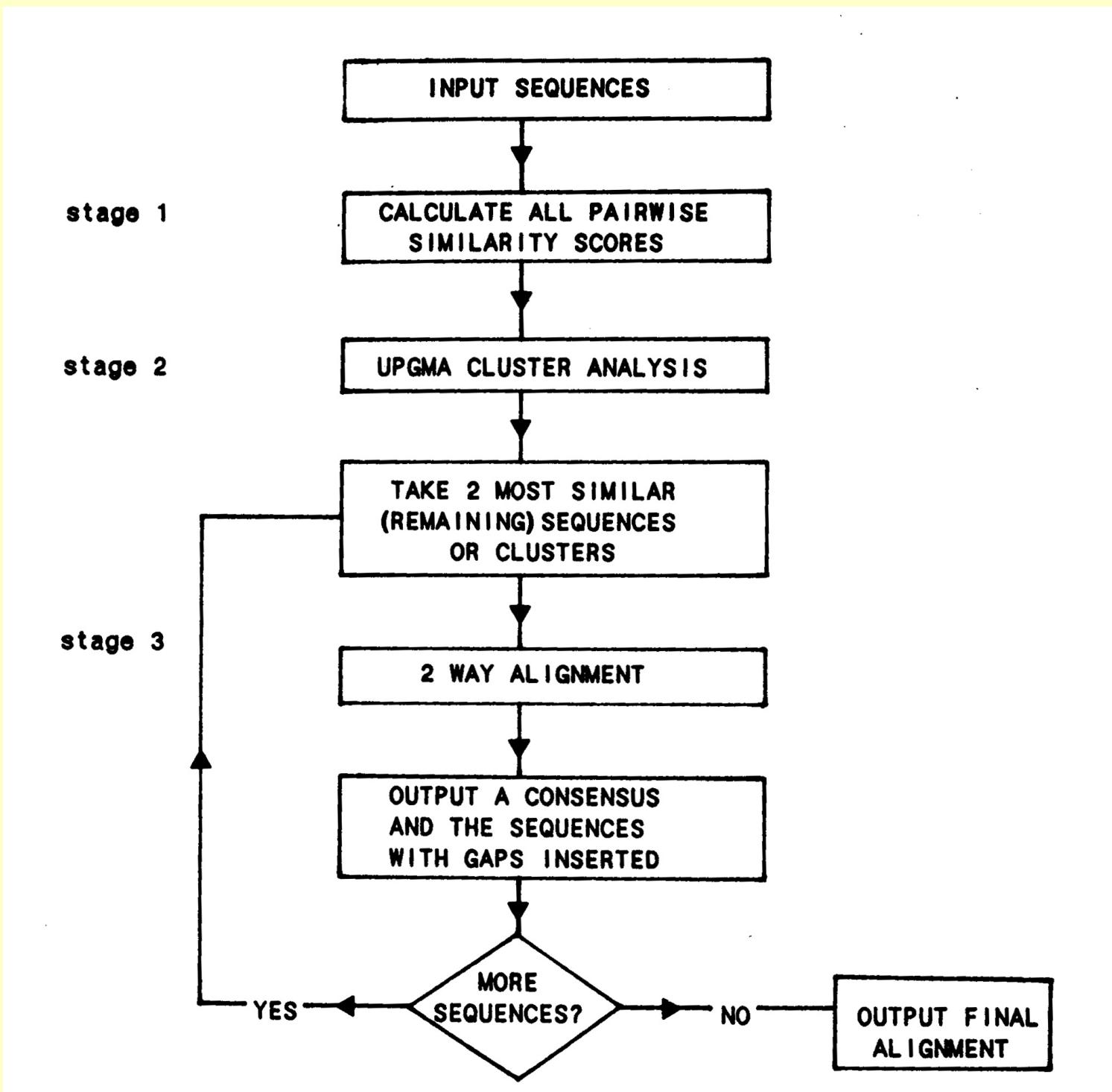
SEQ2 : MQQL- DNPYIVRMIGICEAF-SWM

SEQ3 : MKQL- QHPRLVRLYAVVTQF-PIY

SEQ2 : MQQL- DNPYIVRMIGICEAE-SWM

SEQ4 : MKMIGKHKNIINLLGACTQDGPLY

Clustal Procedure



Clustal Dendrogram

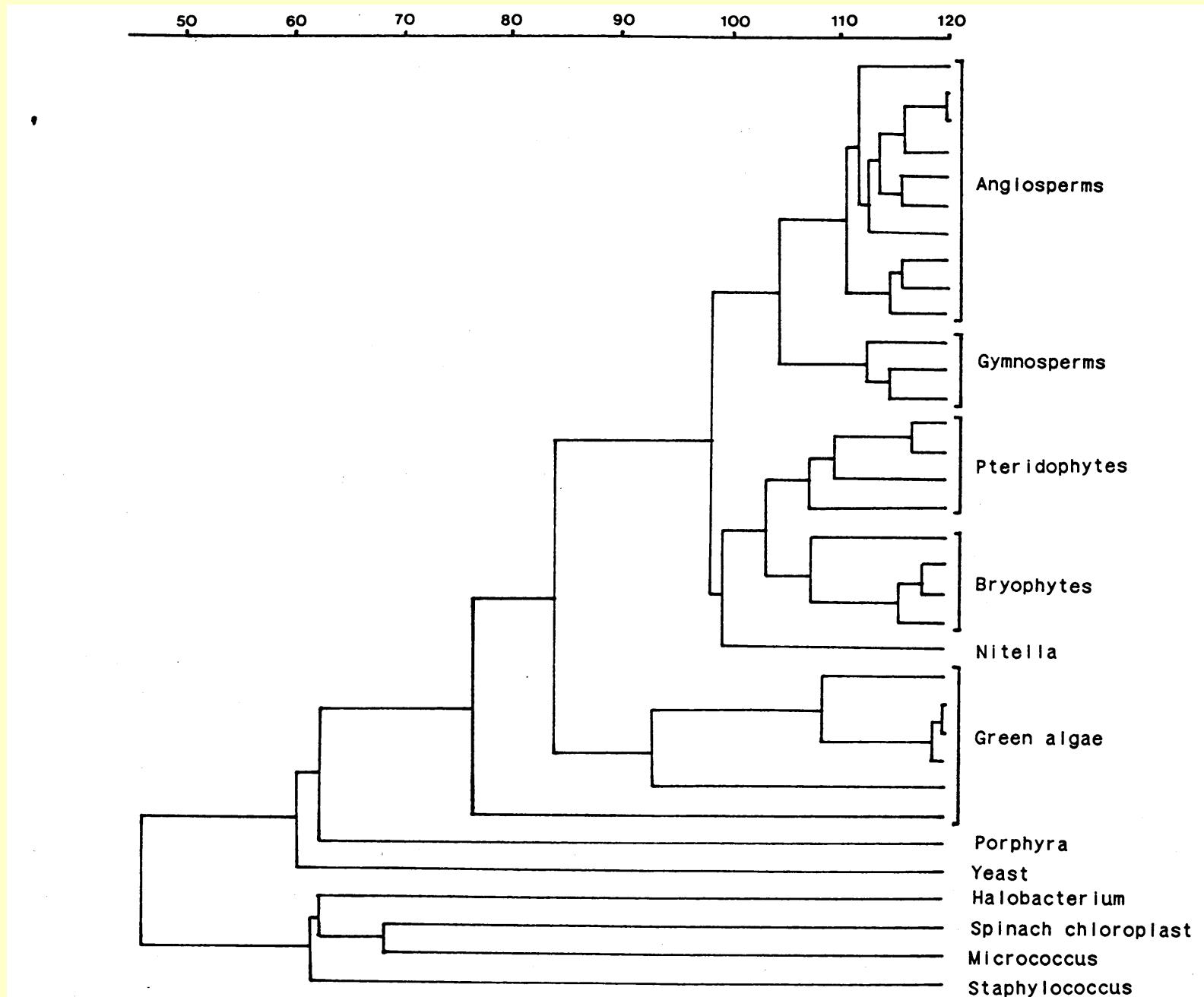


Fig. 4. UPGMA dendrogram of 34 plant, yeast and bacterial 5S RNA sequences. The sequences were taken from Hori et al. (1985) (see RESULTS, section a2). The major plant taxonomic groupings are indicated. The scale across the top margin shows the number of matching nucleotides (after alignment) between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

Clustal Globin Alignment

- 1 human beta globin
- 2 horse beta globin
- 3 human alpha globin
- 4 horse alpha globin
- 5 cyanohaemoglobin
- 6 whale myoglobin
- 7 leghaemoglobin

	A	B	C		E	F	G	H
1	VHLTPEEKSAVTALWGKVNV	EVGGEALGRLLVVY	PWTQRFFESFGDLSTPD	AVMGNPK	VKAHGKKVLGAFSDG	DAHLDNLKGTFAT	LSELHCDKLHVDPENFRLLGNVL	VCAHGFVLAHKYH
2	VQLSGEEKAAVLALWDKVNEE	EVGGEALGRLLVVY	PWTQRFFDSFGDLSNPG	AVMGNPK	VKAHGKKVLHSFGE	VHLDNLKGFAA	CVLVAHFGKEFTP	VCAHGFVLAHKYH
3	VLSPADKTNVKAAGKVGAG	EVGAEALERMFLSFPTTKT	YFPHFDLSH	GSAQ	VKGHGKKVADALTNA	VAHVDDMPNALSA	YFPHFDLSH	GSAQ
4	VLSAADKTNVKAAWSKVGGH	AGEYGAEALERMFLGFP	TTKTYFPHFDLSH	GSAQ	VKAHGKKVGDALT	VGHLDLPGALSN	TTKTYFPHFDLSH	GSAQ
5	PIVDTGSVAPLSAAEKT	KIRSAWAPVYS	DYETSGVDILVKFFT	TPAAEEFFPKFKGLT	AAKDLFSSFLKG	PIVDTGSVAPLSAAEKT	KIRSAWAPVYS	AAKDLFSSFLKG
6	VLSEGEWQLVLHWAKVEAD	VAGHQDILIRLFKSHP	ETLEKFDRFKHLKTEAEM	KASED	VLSEGEWQLVLHWAKVEAD	VLSEGEWQLVLHWAKVEAD	ETLEKFDRFKHLKTEAEM	KASED
7	GALTESQAALVKSSWEEFN	ANIPKHTHRFFILVLEI	PAAKDLPFSSFLKG	GTSEVPQNNPE	GALTESQAALVKSSWEEFN	GALTESQAALVKSSWEEFN	ANIPKHTHRFFILVLEI	PAAKDLPFSSFLKG

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

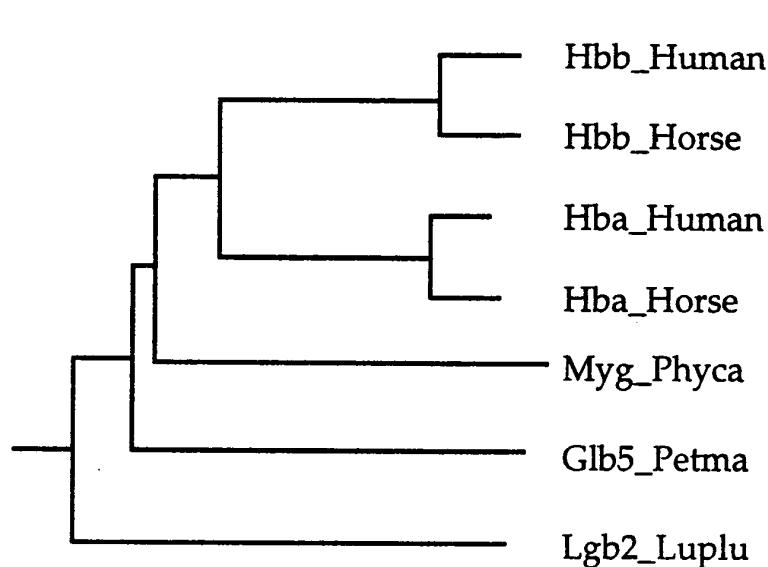
ClustalW Step 1: BLOSUM Distance Matrix



Hbb_Human	1	-				
Hbb_Horse	2	.17	-			
Hba_Human	3	.59	.60	-		
Hba_Horse	4	.59	.59	.13	-	
Myg_Phyc	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



Rooted Neighbor Joining
tree (guide tree)

ClustalW Sequence Weighting

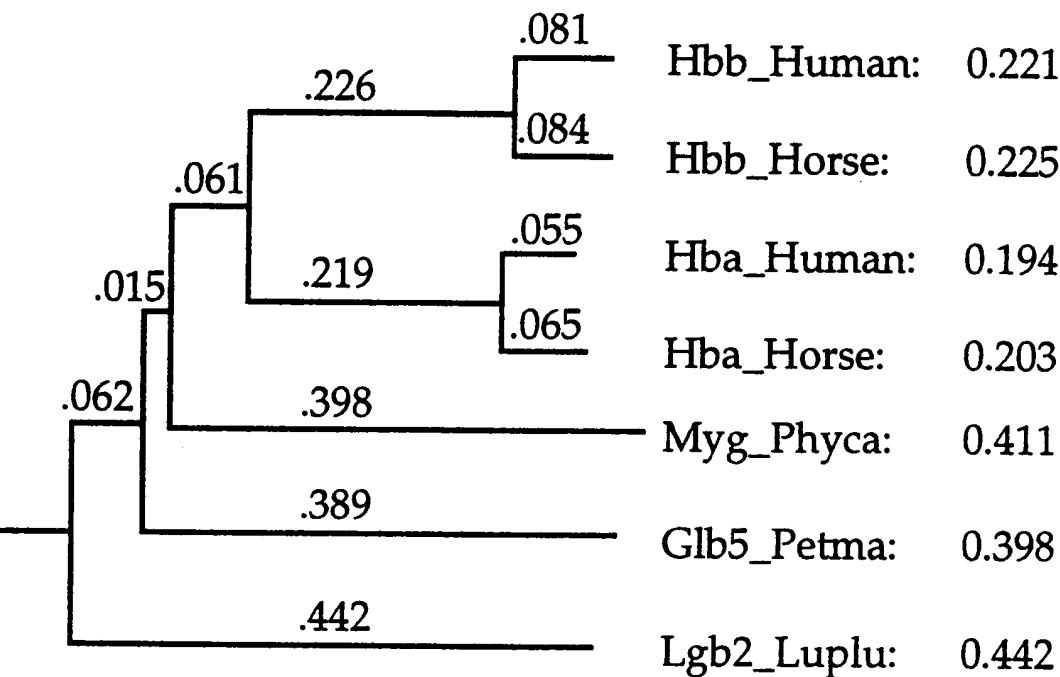


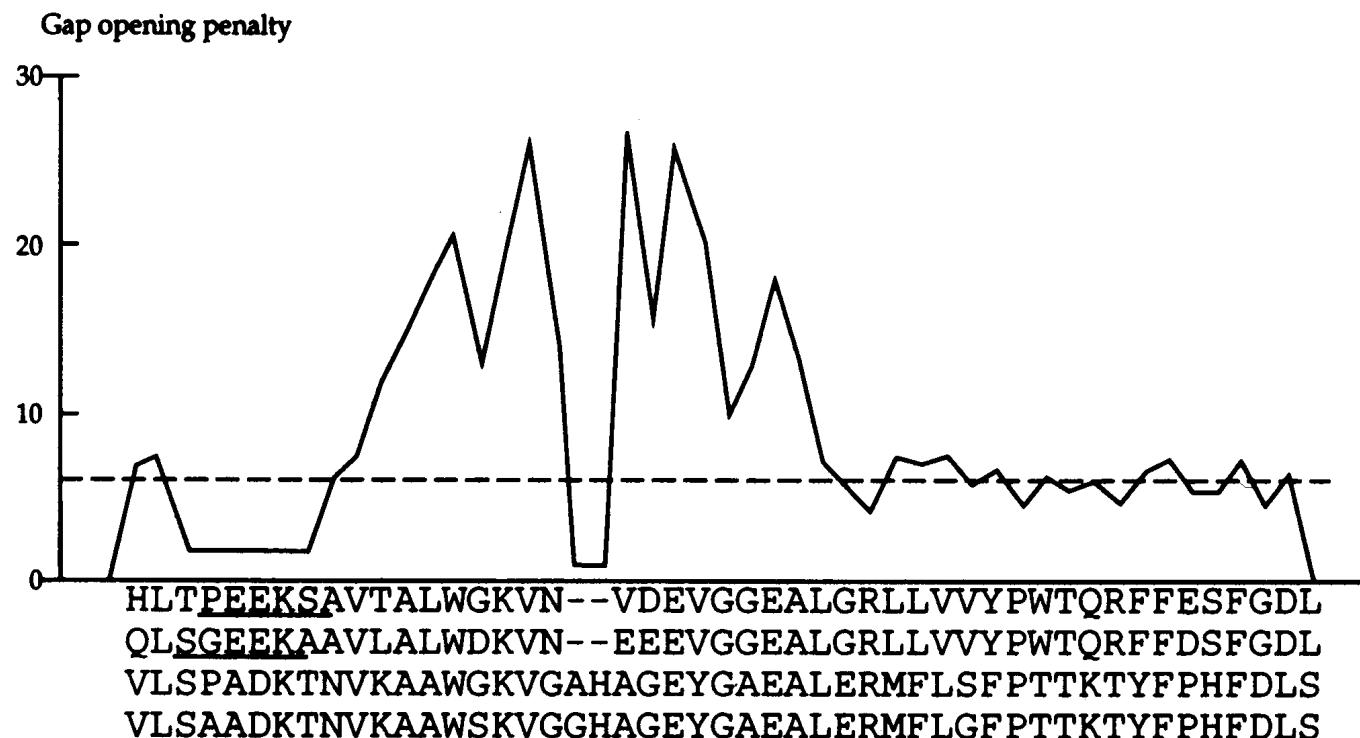
FIG. 3. Sequence weights for the seven globin sequences from Fig. 1. A rooted neighbor-joining tree is shown with branch lengths. The weights are shown for each sequence before normalization (the weights are normalized so as to make the largest equal to 1.0).

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



ClustalW Step 3: Progressive Alignment



```
-----VHLTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYWTQRFFESFGDLST
-----VQLSGEEKAAVLALWDKVN--EEEVGGEALGRLLVVYWTQRFFDSFGDLSN
-----VLSPADKTNVKAAWGKVGAGAGEYGAELERMFLSFPTTKTYFPHFDSL-
-----VLSAADKTNVKAAWSKVGGHAGEYGAELERMFLGFPTTKTYFPHFDSL-
-----VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKT
PIVDTGSVAPLSAAEKT KIRSAWAPVYSTYETSGVDILVKFFTSTPAAQEFFPKFKGLTT
-----GALTESQAALVKSSWEENANIPKHTHRFFILVLEIAPA AKDLFSFLKG TSE
```

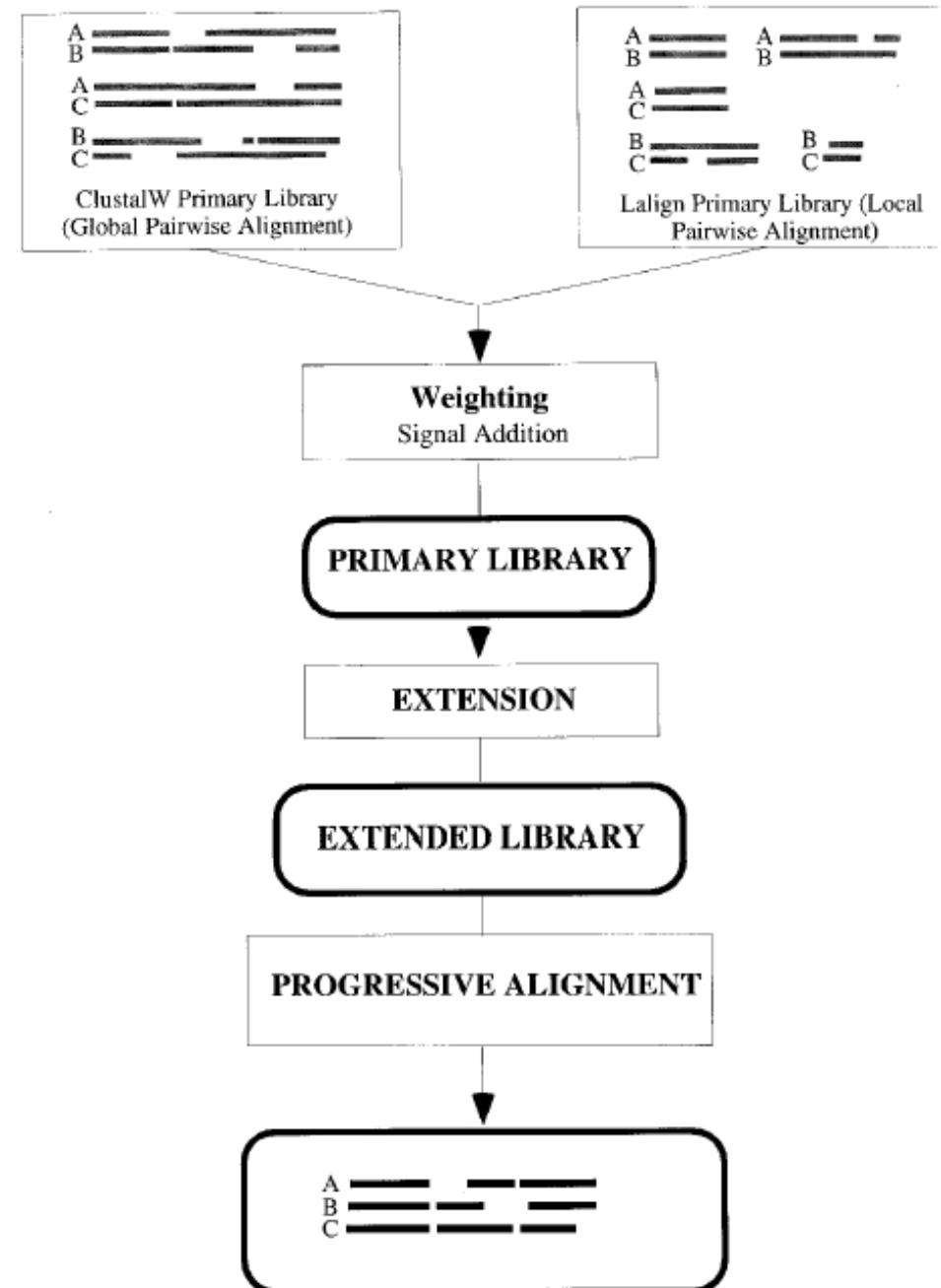
Progressive alignment:
Align following the guide tree

```
PDAVMGNPKVKAHGKKVLGAFSDGIAHLD-----NLKGTFAATLSELHCDKLHVDPENFRL
PGAVMGNPKVKAHGKKVLHSFGEGVHHLD-----NLKGTFAALSELHCDKLHVDPENFRL
-----HGSAQVKGHGKKVADALTNAVAHVD-----DMPNALSALSDLHAHKLRVDPVNFKL
-----HGSAQVKAHGKKVG DALT LAVGHLD-----DLPGALSNLSDLHAHKLRVDPVNFKL
EAEMKASEDLKKHGVTVL TALGAI LKKKG-----HHEAEELKPLAQSHATKHKIPIKYLEF
ADQLKKSA D VRWHAERI I NAVND AVASMDT-- EKMSM KLRDLSGKHAKSFQVD PQYFKV
VP--QNNPELOAHAGKVFKLVYEAAIQLQVTGVVVT DATLKNLGSVHVS KG-VADAHPV
```

```
LGNVLVCVLAHHFGKEFTP PVQAAYQKV VAGVANALAHKYH-----
LGNVLVVVLARHFGKDFTPELQASYQKV VAGVANALAHKYH-----
LSHCLLVTLA AHLPAEFTP AVHASLDKFLASVSTVLT SKYR-----
LSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVSTVLT SKYR-----
ISEAIIHV LHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
LAAVIADTV AAG-----DAGFEKLM SMICILLRSAY-----
VKEAILKTIKEVGA KWSEELNSAWTIAYDELAIVIKKEMNDAA---
```

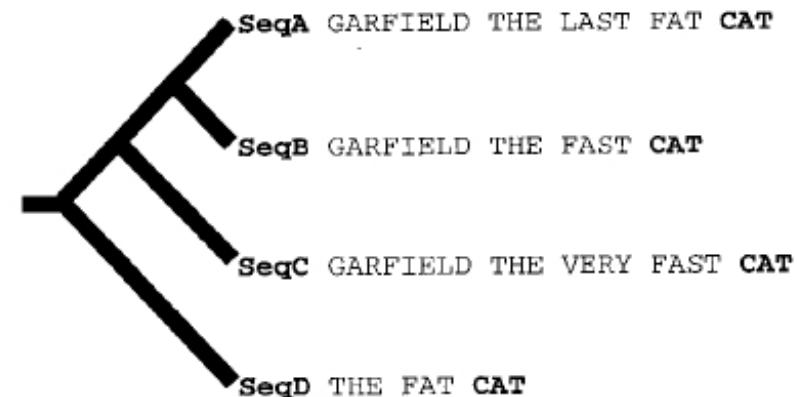
T-Coffee Procedure

T-Coffee: a Method for Sequence Alignment



Regular Progressive Alignment

a) Regular Progressive Alignment Strategy



SeqA GARFIELD THE LAST FA-T CAT
SeqB GARFIELD THE FAST CA-T ---
SeqC GARFIELD THE VERY FAST CAT
SeqD ----- THE --- FA-T CAT

T-Coffee Primary Alignment Library

b)Primary Library

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

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

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

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

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

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

T-Coffee Extended Alignment Library and Progressive Alignment

b) Primary Library

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

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

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

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

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

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

c) Extended Library for seq1 and seq2

SeqA GARFIELD THE LAST FAT CAT
||||| ||||| ||||| |||||
SeqB GARFIELD THE FAST CAT

Weight = 88

SeqA GARFIELD THE LAST FAT CAT
||||| ||||| ||||| |||||
SeqC GARFIELD THE VERY FAST CAT
||||| ||||| ||||| |||||
SeqB GARFIELD THE FAST CAT

Weight = 77

Seq1 GARFIELD THE LAST FAT CAT
||| ||||| |||||
SeqD THE FAT CAT
||| ||||| |||||
SeqB GARFIELD THE FAST CAT

Weight = 100

Extended Library

SeqA GARFIELD THE LAST FAT CAT
||||| ||||| ||||| |||||
SeqB GARFIELD THE FAST CAT
||||| ||||| ||||| |||||

Dynamic Programming

SeqA GARFIELD THE LAST FA-T CAT
SeqB GARFIELD THE ---- FAST CAT

Comparison of T-Coffee to Other MSAs

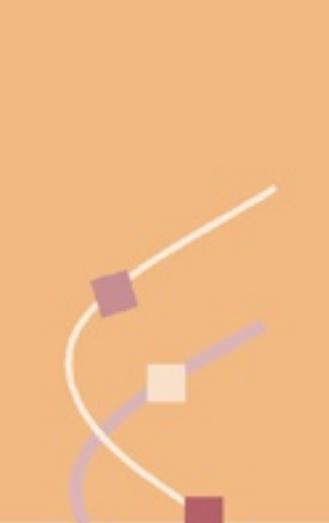


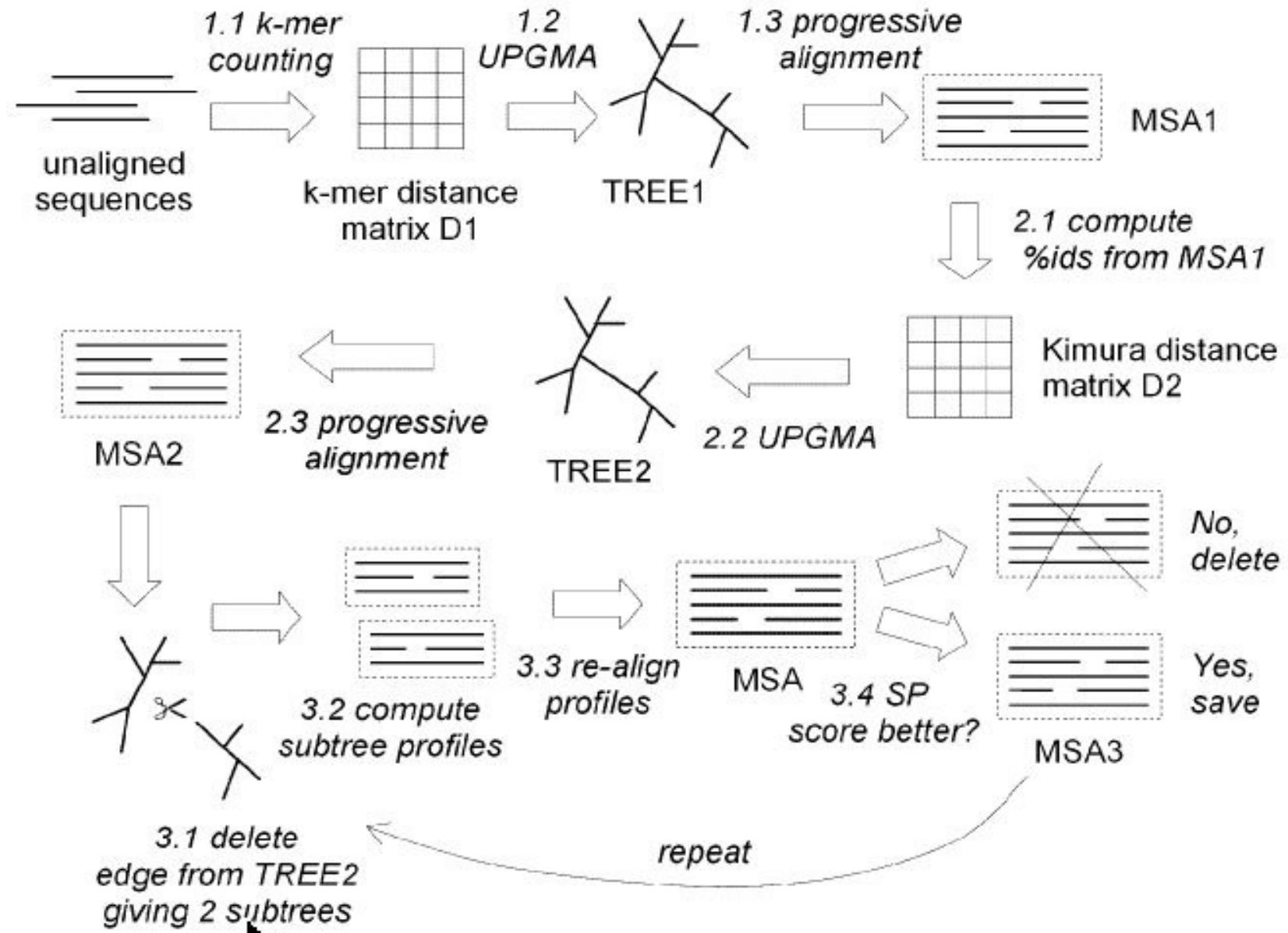
Table 2. T-Coffee compared with other multiple sequence alignment methods

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	<u>80.7</u>	<u>37.3</u>	<u>52.9</u>	<u>83.2</u>	88.7	<u>72.1</u>



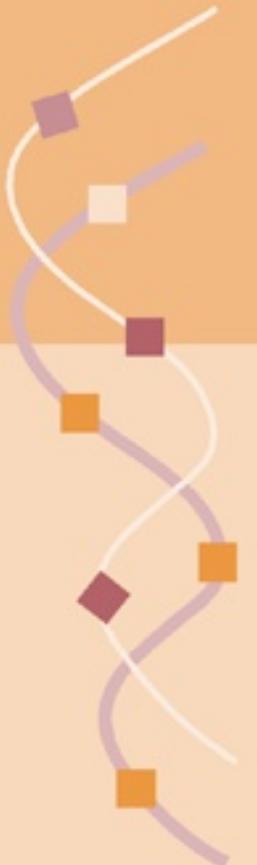
MUSCLE

Edgar (2004) NAR 32, 1792-1797



MUSCLE

Edgar (2004) NAR 32, 1792-1797



YES_XIPHE	MGCvrSKEaKgPA lKYqpdnsvvPvSahlgHYGpeptimg
YES_AVISY	-----d KgPA mKYrtdntp-ePiSshvsHYGsd
YES_CHICK	----- MGCikSKE d KgPA mKYrtdntp-ePiSshvsHYGsd
YES_HUMAN	----- MGCikSKE n KsPA iKYrpeNtp-ePvStsvsHYGae
YES_MOUSE	----- MGCikSKE n KsPA iKYtpeNlt-eP--vSpsasHYG
YES_XIPHE	MGCvrSKEaKgPA lKYqpdnsvvPvSahlgHYGpeptimg
YES_AVISY	-----d KgPA mKYrtdntp-ePiSshvsHYGsdssqat
YES_CHICK	MGCikSKE d KgPA mKYrtdntp-ePiSshvsHYGsdssqat
YES_HUMAN	MGCikSKE n KsPA iKYrpeNtp-ePvStsvsHYGaepttvs
YES_MOUSE	MGCikSKE n KsPA iKYtpeNlt-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
FFTNSI	0.838	0.908	0.708	0.793	0.947

ProbCons

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

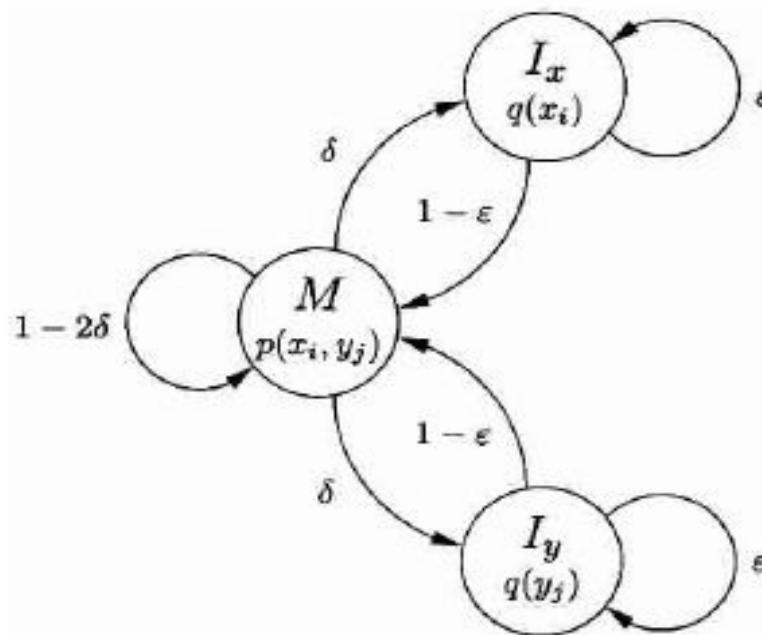
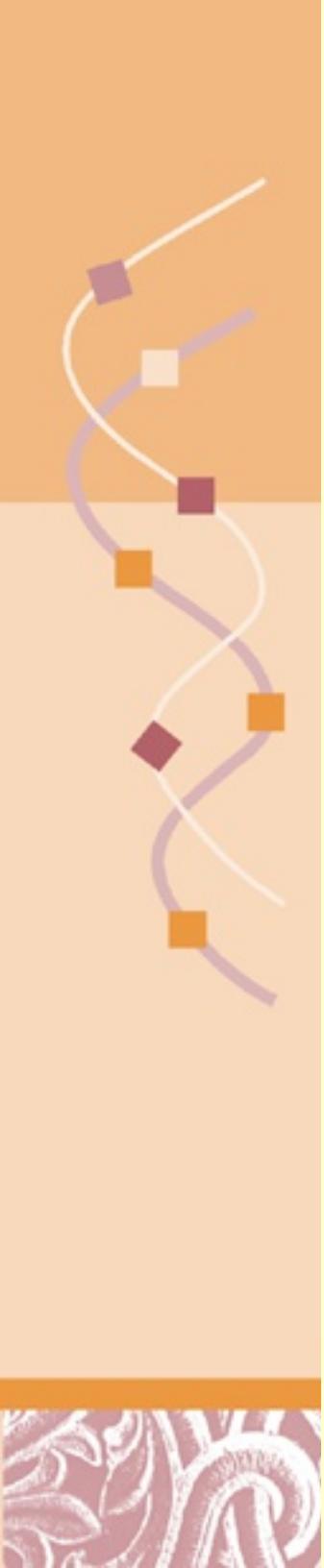


Figure 1. Basic pair-HMM for sequence alignment between two sequences, x and y . State M emits two letters, one from each sequence, and corresponds to the two letters being aligned together. State I_x emits a letter in sequence x that is aligned to a gap, and similarly state I_y emits a letter in sequence 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).

A decorative graphic of a DNA double helix on the left side of the slide. The helix is composed of two strands, one purple and one white, forming a spiral shape. There are several small colored squares (purple, orange, and yellow) placed along the strands, likely representing specific nucleotide positions or mutations.

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: refinement by dividing sequences into two groups and re-align. Repeat multiple times.

ProbCons

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

PROBCONS

Probabilistic Consistency-based Multiple Alignment of Amino Acid Sequences

RUN

ABOUT

DOWNLOAD

HELP

PROBCONS is an efficient protein multiple sequence alignment program, which has demonstrated a statistically significant improvement in accuracy compared to several leading alignment tools.

BASIC PARAMETERS

E-mail address

E-mail address (again)

Input sequence file

ADDITIONAL OPTIONS

Consistency reps

Iterative refinement reps

Pre-training reps

Output format MFA CLUSTALW

COMPUTE ALIGNMENT

Comments to Chuong Do (chuongdo@cs.stanford.edu)



Doug Brutlag 2010

ProbCons

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

PROBCONS version 1.12 multiple sequence alignment

GLB5_PETMA	PIVDTGSVAPLSAAEKT	KIRSAWAPVYSTYETSGVDILVKFFTSTPAAQEFFFPKFKGLTT
HBA_HORSE	-----VLSAADKTNVKA	AWSKVGGHAGEYGAEALEMFLGFPTTKTYFPHF-DLS-
HBA_HUMAN	-----VLSPADKTNVKA	AWGKVGAGHEYGAEALEMFLSFPTTKTYFPHF-DLS-
HBB_HORSE	-----VQLSGEEKAAVLAL	WDKVNE--EEVGGEALGRLLVVYPWTQRFFDSFGDLSN
HBB_HUMAN	-----VHLTPEEKSAVTAL	WGKVNV--DEVGGEALGRLLVVYPWTQRFFESFGDLST
LGB2_LUPLU	-----GALT	ESQAALVKSSWEFNANIPKHTHRFFILVLEIAPAAKDLFSFLKGTS-
MYG_PHYCA	-----VLSEG	EWEQQLVLHVWAKVEADVAGHQDILIRLFKSHPETLEKFDRFKHLKT
	*: : : * .	: .. * : * : ..
GLB5_PETMA	ADQLKKSA	DVRWHAERIINAVNDAVASMD-----DTEKMSMKLRDL SGKHAKSFQVDPQY
HBA_HORSE	----HGSAQVKA	HGKKVGDALTAVGHLD-----DLP---GALSNLSDLHAHKL RVDPVN
HBA_HUMAN	----HGSAQVKGHGKKV	ADALTNAVAHVD-----DMP---NAL SALSDLHAHKL RVDPVN
HBB_HORSE	PGAVMGNPKVKA	HGKKVLHSFGEGVHLD-----NLK---GTFAALSELHCDKLHVD PEN
HBB_HUMAN	PDAVMGNPKVKA	HGKKVLGAFSDGLAHLD-----NLK---GTFA TLSELHCDKLHVD PEN
LGB2_LUPLU	-EV	PQNNPELQAHAGKVFKLVYEAAIQLQVTGVVVTD---ATLKNLGSVHSV SKVAD-AH
MYG_PHYCA	EAEMKASEDLKKHGVTV	LTA GAILKKKG-----HHE---AELKPLAQSHATKHKIPIKY
	. . : * . : .	: * . * .
GLB5_PETMA	FKVLA	AVIAIDTVAA-----GDAGFEKLMSMICIL---L-RSA-----Y
HBA_HORSE	FKLLSHCLL	STLAVHLPNDFTP AVHASLDKFLSSVSTV---LTSKY-----R
HBA_HUMAN	FKLLSHCLL	VTLAAHLPAEFTP AVHASLDKFLASVSTV---LTSKY-----R
HBB_HORSE	FRLLGNVL	VVVLARHFGKDFTPELQASYQKV VAGVANA---LAHKY-----H
HBB_HUMAN	FRLLGNVL	CVLAHHFGKEFTP VQAA YQKV VAGVANA---LAHKY-----H
LGB2_LUPLU	FPVVKEA	ILKTIKEVVGAKWSEELNSAWTIAYDELAIVIKKEMNDA-----A
MYG_PHYCA	LEFISEAIIHVLHSRHPGDF	GADAQGAMNKALELFRKD---IAAKYKELGYQG
	: . : . :

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

Aligner	Ref 1 (82)		Ref 2 (23)		Ref 3 (12)		Ref 4 (12)		Ref 5 (12)		Overall (141)		Time (mm:ss)
	SP	CS	SP	CS									
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 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)

0.1

(range 0.1 thru 10.0)

Set gap extension penalty

BLOSUM

Set alignment scoring matrix



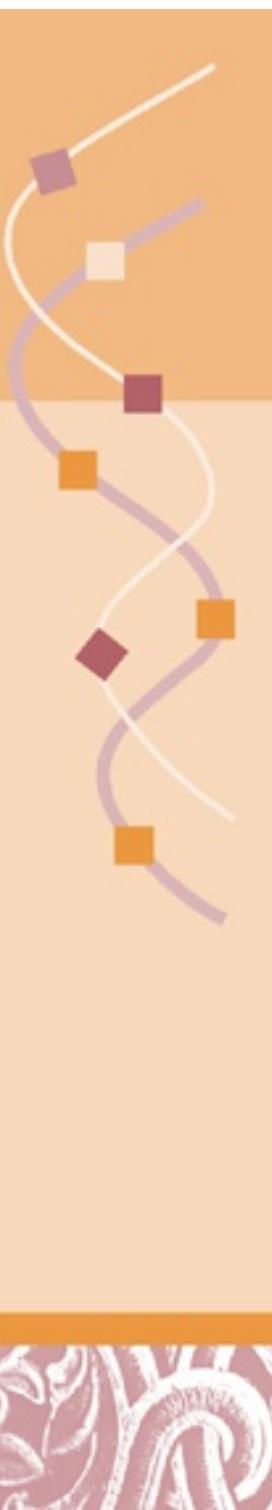
SeqWeb ClustalW MSA Parameters

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

Multiple Sequence Alignment Parameters

Set gap opening penalty	<input type="text" value="10.0"/> (range 1.0 thru 10.0)
Set gap extension penalty	<input type="text" value="0.05"/> (range 0.01 thru 10.0)
Set gap separation penalty range	<input type="text" value="4"/> <input type="text" value="30.0"/> (1.0 thru 100.0)
Delay divergent	<input type="checkbox"/>
Turn on end gap separation penalty	<input type="checkbox"/>
Matrix contains negative values	<input type="checkbox"/>
Turn off residue-specific gap penalty	<input type="checkbox"/>
Turn off hydrophilic residue gaps	<input type="checkbox"/>
List hydrophilic residues	G,P,S,N,D,Q,E,K,R
Set alignment scoring matrix	BLOSUM
Set output sequences' ordering	Input order Alignment order
Sequence range to write [m,n] (starting from m to m+n)	<input type="text"/>

Run | **Reset**



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
```

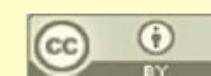
//

1	50
hba_horse.peV LSAADKTNVK AAWSKVGGHA GEYGAELER MFLGPTTKT
lgba_soybnVA FTEKQDALVS SSFEAFKANI PQYSVVFYTS ILEKAPAAKD
hbahumanV LSPADKTNVK AAWGKVGAIA GEYGAELER MFLSFPTTKT
glb5_petma.p	PIVDTGSVAP LSAAEKTKIR SAWAPVYSTY ETSGVDLVK FFTSTPAQE
lgb1_soybn.pGA FTEKQEAALVS SSFEAFKANI PQYSVVFYNS ILEKAPAAKD
hba_humanV LSPADKTNVK AAWGKVGAIA GEYGAELER MFLSFPTTKT
lgb2_luplu.pGA LTESQAALVK SSWEENFANI PKTHRFFIL VLEIAPAAKD
hbb_humanVH LTPEEKSAVT ALWGKVNN..V DEVGGEALCR LLVYYPWTQR

51	100
hba_horse.pe YFPHF.....	.DLSHGSAQV KAHGKKVGD A LTIAVGHLD LPGALSNLSD
lgba_soybn LFSFLANG...	.VDPTRPKL TGHAEKLFA L VRDSAGOLKA SGTVVADAL
hbahumanDLSHGSAQV KGHGKKVADA LTNAVAVVD MPNALSALSD
glb5_petma.p FFPFKKGTLT	ADQLKKSAADV RWIAERIIINA VNDAVASMD TEKMSMLRD
lgb1_soybn.p LFSFLANG...	.VDPTRPKL TGHAEKLFA L VRDSAGQLKT NGTVVADAL
hba_human YFPHF.....	.DLSHGSAQV KGHGKKVADA LTIAVAVVD MPNALSALSD
lgb2_luplu.p LFSFLKGTS.	.EVQPQNPTEL QAHACKVFKL VYEAIQLOV TGVVTDATL
hbb_human FESFGDLST	PDAVMGRPKV KAHGKKVGLA FSDGLAHLDN LKGTFATLSE

101	150
hba_horse.pe L.....HAK LRVDPVNFKL LSHCLLSTLA VHLPNDFTPA VHASLDKFLS	
lgba_soybn G....SVHAQK AVTDP.QFVV VKEALLKTIK AAVGDKNSDE LSRAWEVAYD	
hbahuman L.....HAK LRVDPVNFKL LSHCLLVTLA AHLPAEFTPA VHASLDKFLA	
glb5_petma.p LSG..KAKS FQVDPQYFKV LAIAVIADTV A GD..... AGFEKLM	
lgb1_soybn.p V....SIHAQK AVTDP.QFVV VKEALLKTIK EAVGGIWSDE LSSAWEVAYD	
hba_human L.....HAK LRVDPVNFKL LSHCLLVTLA AHLPAEFTPA VHASLDKFLA	
lgb2_luplu.p KNLGSVHVSK GVADA.HFPV VKEAILKTIK EVVGAKWSEE LNSAWTIAYD	
hbb_human L.....HCDK LHVDPENFRL LGNVLCVLA HHFGKEFTP P VQAAYQKVA	

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



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: lgb1_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 DEVGEALGR LLVVYPWTQR
lgb1_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 LSAAEKTIR SAWAPVYSTY ETSGVDILVK FFTSTPAAQE

      51                                100
hbahuman YFPHF..... DLSHGSAQV KGHGKKVADA LTNAVAHVDD MPNALSA LSD
hba_human YFPHF..... DLSHGSAQV KGHGKKVADA LTNAVAHVDD MPNALSA LSD
hba_horse.pep YFPHF..... DLSHGSAQV KAHGKKVGDA LTLAVGHLDD LPGALSNLSD
hbb_human FFESFGDLST PDAVMGNPKV KAHGKKVLGA FSDGLAHLDN LKGTFTATLSE
lgb1_soybn LFSFLANG... .VDPTNPKL TGHAEKLFA L VRDSAGQLKA SGTVVADAAL
lgb1_soybn.pep LFSFLANG... .VDPTNPKL TGHAEKLFA L VRDSAGQLKT NGTVVADAAL
lgb2_luplu.pep LFSFLKGTS. .EVPAQNNEP QAHAGKVFKL VYEAAIQLQV TGVVVTDA TL
glb5_petma.pep FFPKFKGLT ADQLKKSA DV RWHAERIINA VNDAVASMDD TEKMSMKLRD

      101                               150
hbahuman L.....HAHK LRVDPVNFKL LSHCLLVTLA AHLPAEFTP A VHASLDKFLA
hba_human L.....HAHK LRVDPVNFKL LSHCLLVTLA AHLPAEFTP A VHASLDKFLA
hba_horse.pep L.....HAHK LRVDPVNFKL LSHCLLSTLA VHLPNDFTP A VHASLDKFLS
hbb_human L.....HCDK LHVDPENFRL LGNVLVCVLA HHFGKEFTP VQAAYQKVVA
lgb1_soybn G....SVHAQK AVTDP.QFVV VKEALLKTIK AAVGDKWSDE LSRAWEVAYD
lgb1_soybn.pep V....SIHAQK AVTDP.QFVV VKEALLKTIK EAVGGNWSDE LSSAWEVAYD
lgb2_luplu.pep KNLGSVHVK C VADA.HFPV VKEAILKTIK EVVGAKWSEE LNSAWTIAYD
glb5_petma.pep LSG..KHAKS FQVDPQYFKV LAAVIADTV AGD..... .AGFEKLMS

      151          164
hbahuman SVSTVLTSKY R...
hba_human SVSTVLTSKY R...
hba_horse.pep SVSTVLTSKY R...
hbb_human CVANALAHKY H...
lgb1_soybn ELAAAIIKKA. ....
lgb1_soybn.pep ELAAAIIKKA. ....
lgb2_luplu.pep ELAIVIKKEM NDAA
glb5_petma.pep MICILLRSAY ....
```

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

Input Parameters:

Select a sequence comparison matrix. This matrix determines how matches and mismatches are scored. The default penalties 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)

http://

Multiple Sequence Alignment Results

y-prot



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

//

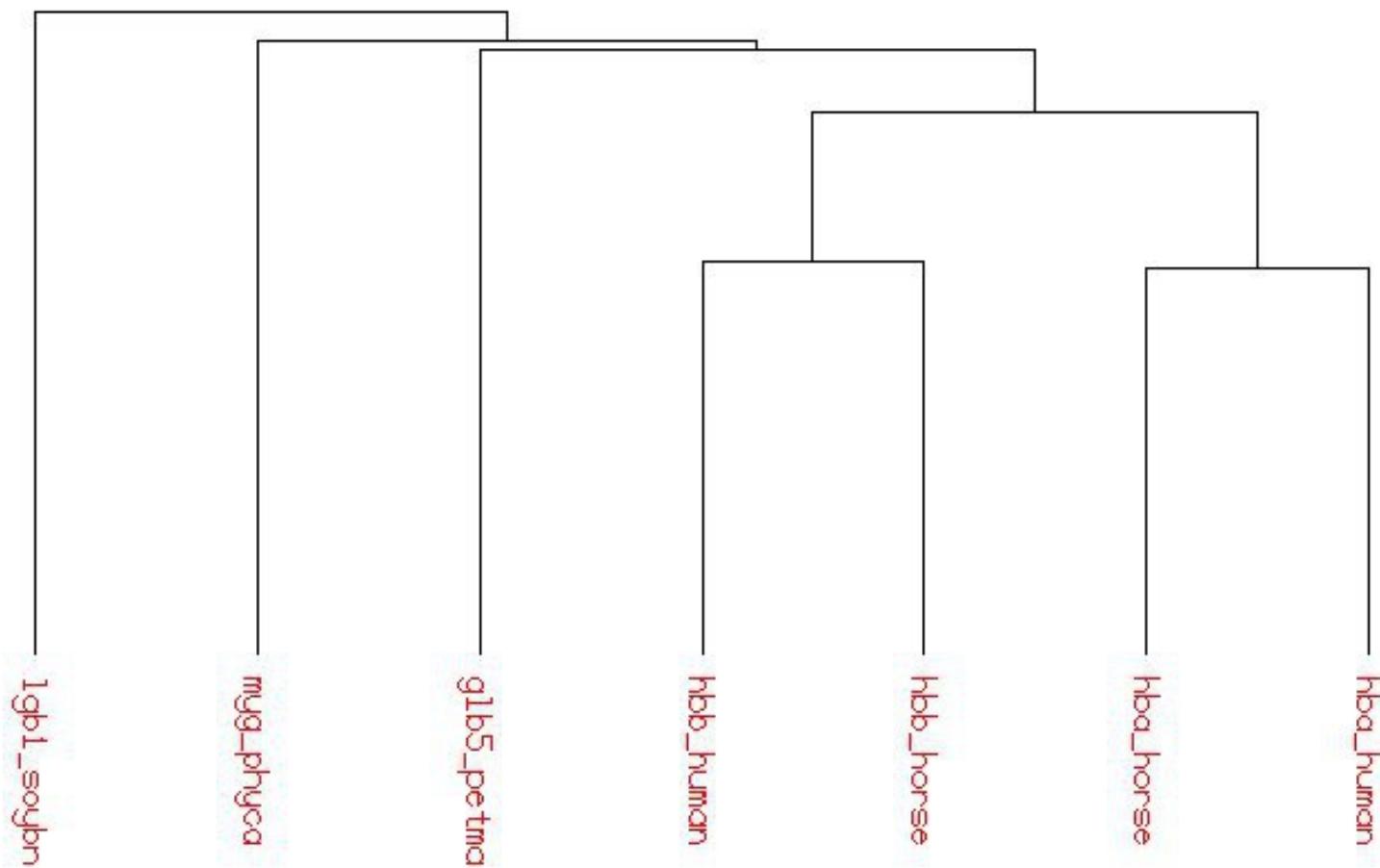
	1	50
hba_human	~~~~~v 1spadktrvk aawgkvgaaha geygaeler mflsfpttkt	
hba_horse	~~~~~V LSAADKTNVK AAWSKVGGHA GEYCAEALER MFLGFPTTKT	
hbb_horse	~~~~~VQ LSGEEKAAVL ALWDKV..NE EEVGGEALGR LLVVYPWTQR	
hbb_human	~~~~~vh 1tpeeksavt alwgkv..nv devggealgr llvvypwtqr	
glb5_petma	PIVDTGSVAP LSAAEKTIR SAWAPVYSTY ETSGVDILVK FFTSTPAAQE	
myg_phyca	~~~~~v 1segewqlvl hhwakveadv aghgqddilir lfkshpetle	
lgb1_soybn	~~~~~ga ftekqealvs ssfeafkani pgysvvfyins ilekapaakd	
	51	100
hba_human	yfphf.dlsh gsaqv kg hgkkvada 1trava hvdd mpnalsalsd	
hba_horse	YFPHF.DLSH GSAQV KA HGKKVGDA LT LAVGHLD LPGALSNLSD	
hbb_horse	FFDSFGDLSN PGAVMGNPKV KA HGKKVLSN FGECHHLDN LKGTFALSE	
hbb_human	ffesfgd1st pdavmgnpkv ka hgkkvlg a fsdglah1dn lkgtfatlse	
glb5_petma	FFPKFKGLTT ADQLKKKSADV RW HAERIINA VNDAVASMDD TEKMSMKLRD	
myg_phyca	kfdrfkh1kt eaemkased1 kk hgvtv1ta lg...ai1kk kghheae1kp	
lgb1_soybn	1fsflan... .gvdp1rpk1 tghaek1fal vrdsagg1.k tngtvvadaa	
	101	150
hba_human	1...hank1r vdpvnfk11s hc11vtlaah 1paeftpavh as1dkflasv	
hba_horse	L...HAKLKR VDPVNFKLLS HCLLSTLAHV LPNDFTPAVH ASLDKFLLSSV	
hbb_horse	L...CDKLH VDPENFRLLG NVLVVVLARH FGKDFTPELQ ASYQKVVAGV	
hbb_human	1...hcd1k1h vdpvnfr11g nv1vcv1ahh fgkeftppvq aayqkvvagv	
glb5_petma	LSGKMAKSFQ VDPQYFKVLA AVIADTV...AGD AGFEKLMSMI	
myg_phyca	laqshatkhk ipikylefis eaiihvlsr hpgdfgadaq gammkalelf	
lgb1_soybn	1vs1haqkav tdpq. fvvvk eallktikea vggriwsdels sawevayde1	

	151	165
hba_human	stv1tskyr~ ~~~~	
hba_horse	STVLT SKYR~ ~~~~	
hbb_horse	ANALAHKYH~ ~~~~	
hbb_human	ana lankyh~ ~~~~	
glb5_petma	CILLRSAY~ ~~~~	
myg_phyca	rkdiaakyke 1gyqg	
lgb1_soybn	aaaikka~~~ ~~~~	

SeqWeb Pileup Dendrogram

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

Multiple Sequence Alignment Dendrogram January 28, 2010 10:16



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 comparison matrix. This matrix determines how matches and mismatches are scored. The default penalties 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](#)

show positions agreeing with the consensus in upper case

display alignment only at positions that disagree with the consensus

none of the above

are identical

are similar

are somewhat similar

[Count residues toward consensus that](#)

2

[Minimum number of votes required for a consensus](#)

Run Reset



Plurality: 2.00 Threshold: 4 AveWeight 1.00 AveMatch 2.78 AvMisMatch 0.00

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:	<u>hba_horse.pep</u>	Len:	165	Check:	2167	Weight:	1.00
Name:	<u>hbb_horse.pep</u>	Len:	165	Check:	9310	Weight:	1.00
Name:	<u>hbb_human</u>	Len:	165	Check:	208	Weight:	1.00
Name:	<u>g1b5_petma.pep</u>	Len:	165	Check:	2079	Weight:	1.00
Name:	<u>myg_phyca</u>	Len:	165	Check:	4320	Weight:	1.00
Name:	<u>1gb1_soybn.pep</u>	Len:	165	Check:	7278	Weight:	1.00

//

	1	50						
hba_human	-----v	1spadktnvk	aawgkvga	ha	geygaea	ter	mflsf	pttkt
hba_horse	-----v	LSAADKTNVK	AAWSKVGGHA		GEYGAEALER		MFLGF	PTTKT
hbb_horse	-----VQ	LSGEEKAAVL	ALWDKV..NE		EEVGGEALGR		LLWVY	PWTQR
hbb_human	-----vh	1tppeeksavt	a1wgkv..nv		devggealgr		11vvvypwtqr	
g1b5_petma	PIVDTGSVAP	LSAAEKTKIR	SAWAPVYSTY		ETSGVDILVK		FFTSTPAAQE	
myg_phyca	-----v	1segewq1v1	hwakveadv		aghgqdilir		1fkshpetle	
1gb1_soybn	-----ga	ftekqealvs	ssfeafkani		pqysvvfy	ns	11ekapaakd	
Consensus	-----W	LS-AEKT-V-	AAW-KVGAN-		-EYG-EAL-R		LF-S-P-T--	

	51	100						
hba_human	yfphf.d1shgsaqv	kghgkkvada		1tnavahvdd		mpnalsalsd	
hba_horse	YFPHF.DLSHGSAQV	KAHGKKVGD	A	LTLAVGHLD	D	LPGALSNLSD	
hbb_horse	FFDSFGDLSN	PGAVMGNPKV	KAHGKKVHL	S	FGEGVHHLD	N	LKGTF	FAALSE
hbb_human	ffesfgd1st	pdavmgnpkv	kahgkkv1ga		fsdg1ah1dn		1Kgtfatl	se
g1b5_petma	FFPKFKGLTT	ADQLKKSSADV	RWHAERIINA		VNDAVASMD	D	TEKMSMKLRD	
myg_phyca	kfdrfkh1kt	eaeenkased1	kkhgvtv1ta		Tg...ai1kk		Kghheaelkp	
1gb1_soybn	1fsflan...	.gvdpnlpk1	tghaek1fa1		vrdsagg1.k		tngtvvadaa	
Consensus	FFP-F-DLST	P-AV-GS-KV	KAHGKKVLD	A	L-DAVAHLD	D	L-GT-AALSD	

	101	150						
hba_human	1...hahk1r	vdpvnfk11s	hc11vtlaah		1paeftpavh		as1dkf1asv	
hba_horse	L...HAKLR	VDPVNFKLLS	HCLLSTLAHV		LPNDFTPAVH		ASLDKF1LSSV	
hbb_horse	L...HCDKLH	VDPENFRLLG	NVLVVVLAHV		FGKDFTPELQ		ASYQKVVAGV	
hbb_human	L...hcdk1h	vdpnf11g	nvlvcvlahh		fgkeftppvq		aayqkvvagv	
g1b5_petma	LSGKHAKSFQ	VDPQYFKVLA	AVIADTV...	AGD		AGFEKLMSMI	
myg_phyca	1aqshatkhk	ipikylefis	ea11hvhsr		hpqdfgadaq		gamnkate1f	
1gb1_soybn	1vsihaqkav	tdpq.fvvvk	ea11ktikea		vggnw	sawevayde1		
Consensus	L---HA-KL-	VDP-NFKLLS	-VLLVTLA-H		--DFTPAVQ		A---K-LA-V	

	151	165						
hba_human	Stv1tskyr~	-----						
hba_horse	STVLTTSKYR~	-----						
hbb_horse	ANALAHKY~	-----						
hbb_human	analankhy~	-----						
g1b5_petma	CILLRSAY~	-----						
myg_phyca	rkdiaakyke	1gyqq						
1gb1_soybn	aaaikkka~~~	-----						
Consensus	A-ALASKY--	-----						

DeCypher ClustalW Input

<http://decypher.stanford.edu/>

TimeLogic®
biocomputing solutions

DeCypher®

ClustalW Search on DeCypher Protein Sequences

Job Description: ClustalW Search on DeCypher Protein Sequences

E-mail Address:

Return Results: To your web browser As: Web Page

Protein Query Set:
Click Browse... to upload your local file, or paste query data into the text box.
[Use Example Query](#)

/Users/brutlag/Data & Analysis/Browse...

Submit

Job Options: Hide

KTuple Size: 1	Gap Separation Distance: 8
Window Size: 5	End Gaps: OFF
Pairwise Gap Penalty: 3	Transition Weight: 0.5
Gap Open Penalty: 10	Top Diagonals: 5
Gap Extend Penalty: 0.05	Score: PERCENTAGE
Matrix: BLOSUM	Show Sequence #:s: OFF
Hydrophilic Residues: GPSNDQEKR	Output Order: ALIGNED
Residue-Specific Gaps: ON	
Hydrophilic Gaps: ON	
Negative Matrix Values: OFF	

Decypher ClustalW Results

<http://decypher.stanford.edu/>

[Home Page](#)

Build HMM Model | Click once!

Results for Job CGI_Temp2623732f208
ClustalW Search on DeCypher Protein Sequences

[View dendrogram](#)

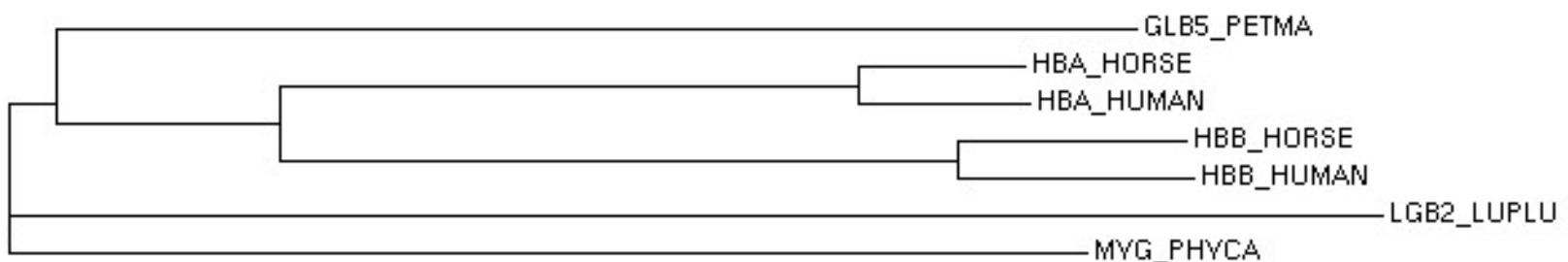
GLB5_PETMA	PIVDTGSVAPLSAAEKT	KIRSAWAPVYSTYETSGVDILVKFFTSTPAAQEFPKF	KGLTT
HBA_HORSE	-----VLSAADKTNVKA	AWSKVGGHAGEYGAEALEMFLGFP	TTKTYFPHF-DLSH
HBA_HUMAN	-----VLSPADKTNVKA	AWGKVGAHAGEYGAEALEMFLSF	PTTKTYFPHF-DLSH
HBB_HORSE	-----VQLSGEEKAAVLALWDKV	--NEEEVGGEALGRLLVVYPWTQRF	FDSFGDLSN
HBB_HUMAN	-----VHLTPEEKSAVTALWGKV	--NVDEVGGEALGRLLVVYPWTQRF	FESFGDLST
LGB2_LUPILU	-----GALTESQAALVKSSWEEF	NANIPKHTHRFFILVLEIAPAAKDLF	SFLKG
MYG_PHYCA	-----VLSEGEWQLVLHVWAKV	EADVAGHQDILIRLFKSHPET	LEKFDRFKHLKT
GLB5_PETMA	ADQLKK	SADVRWHAERI	IINAVNDAVASMDDTEK--MSM
HBA_HORSE	-----GSAQVKAH	GKKVGDALTLAVGHLDD	LSGKHAKSFQVDPQYFKV
HBA_HUMAN	-----GSAQVKG	HGKKVADALTNAVAHVDD	LPGALSNLSDLHAHKLRVDPVNFKL
HBB_HORSE	-----PGAVMGNPKVKAH	GKKVLHSFGEGVHHLDN	MPNALSALSDLHAHKLRVDPVNFKL
HBB_HUMAN	-----PDAVMGNPKVKAH	GKKVLGAFSDGLAHLDN	LKGTFAALSELHCDKLHVDPENFRL
LGB2_LUPILU	-----VPQ--NNPELQAHACKV	FKLVYEAAIQLQVTGVVVTDATLKNLGSVHVS	K-GVADAHFPV
MYG_PHYCA	-----EAEMKASEDLKKHGVT	VLTALGAILKKGH	-HEAELKPLAQSHATKHKIPIKYLEF
GLB5_PETMA	LAAVIAD	TVAAG-----DAGFEKLM	SMICILLRSAY-----
HBA_HORSE	LSHCLLST	TLAVHLPNDFTPAVHASLDKFLSS	VSTVLTSKYR-----
HBA_HUMAN	LSHCLLV	TAAHLPAEFTPASHASLDKFLAS	VSTVLTSKYR-----
HBB_HORSE	LGNVLVV	VLAHFGKDFTPELQASYQKV	VAGVANALAHKYH-----
HBB_HUMAN	LGNVLVC	VLAHHFGKEFTPVQAAYQKV	VAGVANALAHKYH-----
LGB2_LUPILU	VKEAILKT	KEVVGAKWSEELNSAWTIAYDEL	AIVIKKEMNDAA-----
MYG_PHYCA	ISEAIIH	VLSRHPGDFGADAQGAMNKALEL	FRKDIAAKYKELGYQG

Decypher ClustalW Results

<http://decypher.stanford.edu/>

Dendrogram

[Return to top](#)



ClustalW @ EBI Input

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

EBI > Tools > Sequence Analysis > ClustalW

ClustalW

ClustalW is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylogenograms. [New users, please read the FAQ.](#)

>> Download Software

YOUR EMAIL	ALIGNMENT TITLE	RESULTS	ALIGNMENT	CPU MODE
<input type="text"/>	Sequence	interactive <input type="button" value="▼"/>	full <input type="button" value="▼"/>	single <input type="button" value="▼"/>
KTUP (WORD SIZE)	WINDOW LENGTH	SCORE TYPE	TOPDIAG	PAIRGAP
def <input type="button" value="▼"/>	def <input type="button" value="▼"/>	percent <input type="button" value="▼"/>	def <input type="button" value="▼"/>	def <input type="button" value="▼"/>
MATRIX	GAP OPEN	END GAPS	GAP EXTENSION	GAP DISTANCES
def <input type="button" value="▼"/>	def <input type="button" value="▼"/>	def <input type="button" value="▼"/>	def <input type="button" value="▼"/>	def <input type="button" value="▼"/>
OUTPUT		PHYLOGENETIC TREE		
OUTPUT FORMAT	OUTPUT ORDER	TREE TYPE	CORRECT DIST.	IGNORE GAPS
aln w/numbers <input type="button" value="▼"/>	aligned <input type="button" value="▼"/>	none <input type="button" value="▼"/>	off <input type="button" value="▼"/>	off <input type="button" value="▼"/>

Enter or Paste a set of Sequences in any supported format: Help

Upload a file: Browse... Run Reset

ClustalW @ EBI Results

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

Alignment

[Show Colors](#)[View Alignment File](#)

CLUSTAL W (1.83) multiple sequence alignment

HBA_HORSE	-----VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFGLGPTTKTYFPHFDLS--	49
HBA_HUMAN	-----VLSPADKTNVKAAWGVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLS--	49
HBB_HORSE	-----VQLSGEEKAAVLALWDKVN--EEEVGGEALGRLLVVYPWTQRFFDSFGDLSN	50
HBB_HUMAN	-----VHLTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYPWTQRFFESFGDLST	50
GLB5_PETMA	PIVDTGSVAPLSAAEKTKIRSAWAVPVYSTYETSGVDILVKFFTSTPAAQEFFFPKFKGLTT	60
MYG_PHYCA	-----VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKT	51
LGB2_LUPLU	-----GALTESQAALVKSSWEENANIPKHTHRFFILVLEIAPAACKDLFSFLKGTSE	52
	*: : : * . : .. * : * :	
HBA_HORSE	----HGSAQVKAHGKKVGDALTAVGHLD-----LPGALSNLSDLHAHKLRDPVNFKL	100
HBA_HUMAN	----HGSAQVKHGKKVADALTNAVAHVDD-----MPNALSALSNDLHAHKLRDPVNFKL	100
HBB_HORSE	PGAVMGNPKVKAHGKKVHLHSFGEGVHLDN-----LKGTFAALSELHCDKLHVDPENFRL	105
HBB_HUMAN	PDAVMGNPKVKAHGKKVLGAFSDGLAHLDN-----LKGTFAALSELHCDKLHVDPENFRL	105
GLB5_PETMA	ADQLKKSADVRWHAERIIINAVNDAVASMDT--EKMSMCLRDLSGKHAKSFQVDPQYFKV	118
MYG_PHYCA	EAEMKASEDLKKHGTVTLTALGAILKKKGH-----HEAEELKPLAQSHATKHKIPIKYLEF	106
LGB2_LUPLU	VP--QNNPELQAHAGKVFKLVYEAAIQLQVTGVVVTATLKNLGSVHVSKGVAD-AHFPV	109
	. . : * . : . : * . * . : .	
HBA_HORSE	LSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVSTVLTSKYR-----	141
HBA_HUMAN	LSHCLLVTAAHLPAEFTPAPVHASLDKFLASVSTVLTSKYR-----	141
HBB_HORSE	LGNVLVVVLARHFGKDFTPELQASYQKVVACVANALAHKYH-----	146
HBB_HUMAN	LGNVLVCVLAHHFGKEFTPPVQAAYQKVVACVANALAHKYH-----	146
GLB5_PETMA	LAAVIADTVAAAG-----DAGFEKLMSCMICILLRSAY-----	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/>

