Multiple protein sequence alignment with ClustalW

Multiple sites: pir.georgetown.edu
www.genome.ad.jp
Personal computer: download ClustalW

BLOSUM 62 Scoring Matrix

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	9																			
S	-1	4																		
T	-1	1	5																	
P	-3	-1	-1	7																
A	0	1	0	-1	4															
G	-3	0	-2	-2	0	6														
N	-3	1	0	-2	-2	0	6													
D	-3	0	-1	-1	-2	-1	1	6												
E	-4	0	-1	-1	-1	-2	0	2	5											
Q	-3	0	-1	-1	-1	-2	0	0	2	5										
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8									
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5								
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5							
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5						
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4					
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4				
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4			
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6		
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7	
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11

Example of how to score an alignment: Write down two sequences:

sequence#1	V	D	S	–	C	Y
sequence#2	V	E	S	L	C	Y
Score from sub. Matrix	4	2	4	-11	9	7

Score = Σ(AA pair scores) – gap penalty = 15

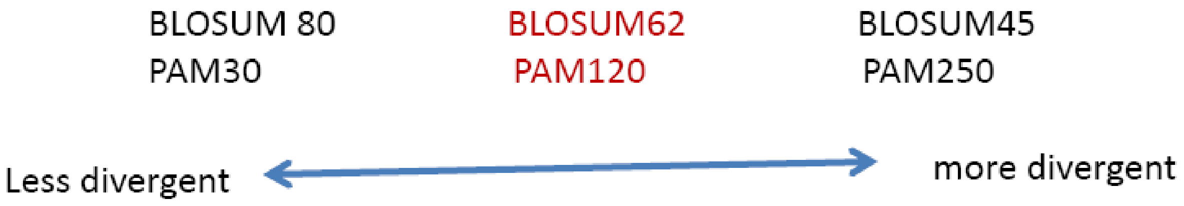
Selecting the right protein-scoring matrix

http://www.ncbi.nlm.nih.gov/pubmed/18792939

- Different scoring matrices are only optimal for detecting different classes of alignments.
- One needs to consider the similarity and amino acid composition of compared sequences.

For general alignment purpose, BLOSUM62 or PAM120 are recommended.

For PAMx matrices, higher x detects more divergent sequences;
For BLOSUMn matrices, lower n detects more divergent sequences.



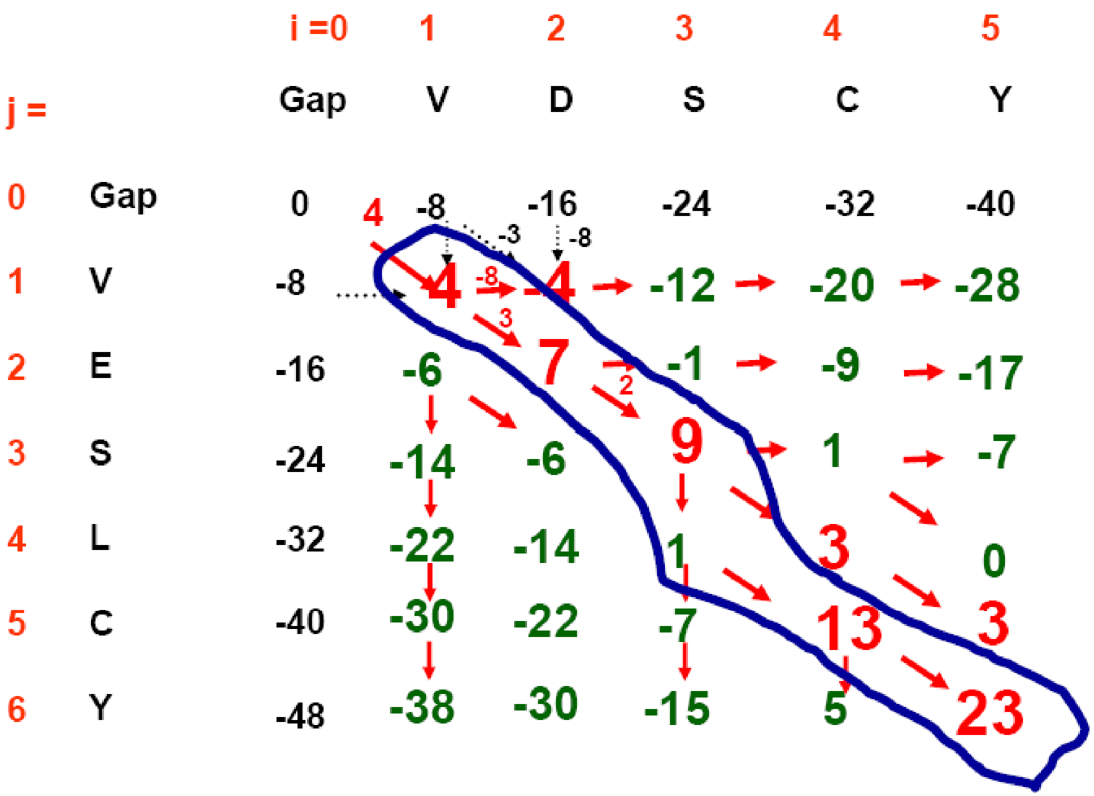
BLOSUM 80
PAM30

BLOSUM62
PAM120

BLOSUM45
PAM250

Less divergent ↔ more divergent

Gap penalty



j =		i =0	1	2	3	4	5
		Gap	V	D	S	C	Y
0	Gap	0	-8	-16	-24	-32	-40
1	V	-8	4	-4	-12	-20	-28
2	E	-16	-6	7	-1	-9	-17
3	S	-24	-14	-6	9	1	-7
4	L	-32	-22	-14	1	3	0
5	C	-40	-30	-22	-7	13	3
6	Y	-48	-38	-30	-15	5	23

http://en.wikipedia.org/wiki/Gap_penalty

Assignment

Confirm your sequences through multiple alignment
Using ClustalW with different matrix (BLOSUM, PAM, GONNET)
Perform Genedoc
Copy and paste for ppt slide