

BLOSUM Matrices

- Trusted data
 - 2000 blocks of conserved regions in ~500 groups of proteins
- Count amino acid pairs: A_{xy}^N
 - Parameterize by evolutionary distance, N**
 - Correct for sample bias**
- Calculate amino acid frequencies:
 - Related pairs: q_{xy}^N
 - Background pair frequencies: E_{xy}
- Log likelihood scoring matrix
 - $S^N = 2 \log_2 \frac{q_{xy}^N}{E_{xy}}$

Count amino acid pairs: A_{xy}^N

- Parameterize by evolutionary distance, N**
Correct for sample bias
- Cluster sequences such that if $s1$ and $s2$ are in different clusters, then $identity(s1, s2) < N\%$
 - Count amino acid pairs in $s1$ aligned with $s2$ only if $s1$ and $s2$ are in different clusters
 - Normalize for cluster size

BLOSUM clustering example

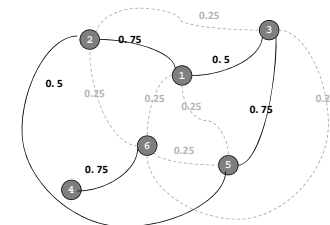
- 1: KKRK
- 2: KKKK
- 3: KNRN
- 4: NRNR
- 5: KNKN
- 6: KRNR

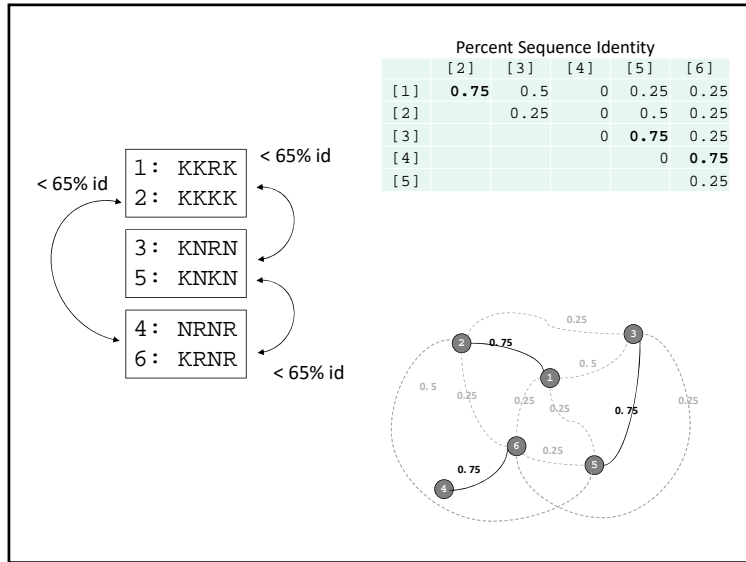
		Percent Sequence Identity				
	[2]	[3]	[4]	[5]	[6]	
[1]	0.75	0.5	0	0.25	0.25	
[2]		0.25	0	0.5	0.25	
[3]			0	0.75	0.25	
[4]				0	0.75	
[5]					0.25	

Unclustered sequences: Every sequence is at least 25% identical

		Percent Sequence Identity				
	[2]	[3]	[4]	[5]	[6]	
[1]	0.75	0.5	0	0.25	0.25	
[2]		0.25	0	0.5	0.25	
[3]			0	0.75	0.25	
[4]				0	0.75	
[5]					0.25	

- 1: KKRK
 - 2: KKKK
 - 3: KNRN
 - 5: KNKN
- < 45% identical
- 4: NRNR
 - 6: KRNR





	PAM	BLOSUM
Evolutionary model	Explicit evolutionary model	None
Data	Full length MSAs of closely related sequences.	Conserved blocks. i.e., ungapped local MSAs
Bias correction	Trees	Clustering
Multiple substitutions	Markov model: $P^n = (P^1)^n$	Implicitly represented in data (clustering)
Evolutionary distance	Markov model: $P^n = (P^1)^n$	Clustering
Matrices	Transition and log odds scoring matrices	Log odds scoring matrix only.
Parameter n	Distance increases with n	Distance decreases with n
Biophysical properties	Derived indirectly from data	Derived indirectly from data

Comparing PAM and BLOSUM matrices

PAM	Sequence identity	BLOSUM
20	83%	
30		
60	63%	
70		
100	43%	90
120	38%	80
160	30%	60
200	25%	50
250	20%	45

Less divergent ↑

↓ More divergent

