SUBSTITUTIONS IN PROTEIN-CODING GENES

may be analysed easily if they are few and absolute

may be analysed in only a complicated way otherwise

**non-, synonymous substitutions
sites
differences**

NON- & SYNONYMOUS SITES

Complications:
site classification changes
site categorisation 'fuzzy'

Step 1: sites
$i / 3$ $(3 - i) / 3$
N_S N_A

Step 2: differences

un- & weighted
0-, 2-, 4-fold degeneracy
M_S M_A

NON- & SYNONYMOUS SITES (2)

$p_S = M_S / N_S$

$K_S = -3Log[1 - ((4 M_S) / (3 N_S))]$

$p_A = M_A / N_A$

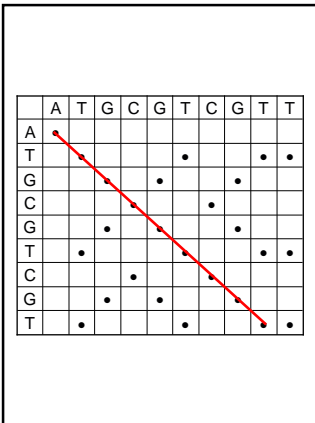$K_A = -3Log[1 - ((4 M_A) / (3 N_A))]$

0, non-   all nonsynonymous

2, two-   (1 / 3) synonymous
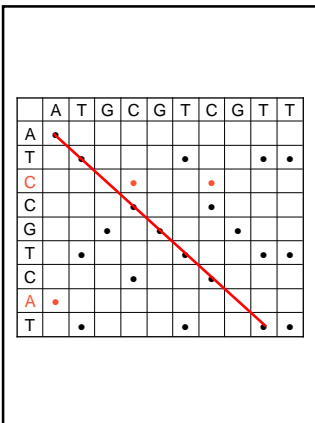
4, four-   all synonymous

AMINO ACID REPLACEMENTS BETWEEN PROTEINS

$p = n / L$

$d = -Log[1 – p]$

Grid 1

	A	T	G	C	G	T	C	G	T	T
A	•									
T		•				•			•	•
G			•		•			•		
C				•			•			
G			•		•			•		
T		•				•			•	•
C				•			•			
G			•		•			•		
T		•				•			•	•

Grid 2

	A	T	G	C	G	T	C	G	T	T
A	•									
T		•				•			•	•
G			•		•			•		
C				•			•			
G			•		•			•		
C				•			•			
G			•		•			•		
T		•				•			•	•
T		•				•			•	•

Grid 3

	A	T	G	C	G	T	C	G	T	T
A	•									
T		•				•			•	•
G			•		•			•		
C				•			•			
G			•		•			•		
T		•				•			•	•
C				•			•			
G			•		•			•		
T		•				•			•	•

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

	A	T	G	C	G	T	C	G	T	T
A	•									
T		•			•			•	•	
C				•		•				
C				•			•			
G			•					•		
C										
G			•							
T		•								•
C				•			•			

_____

_____

_____

_____

_____

_____

_____

_____