Molecular Evolution

Alan R. Rogers

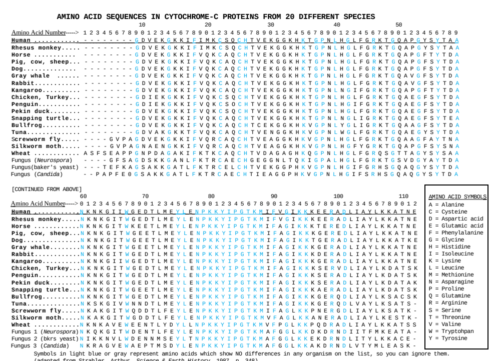
August 31, 2021

Outline

Cytochrome C Amino Acid Sequences

```
AMINO ACID SEQUENCES IN CYTOCHROME-C PROTEINS FROM 20 DIFFERENT SPECIES
                                      10                  20                  30                  40                  50
Amino Acid Number-----> 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
Human ...........       - - - - - - - - G D V E K G K K I F I M K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G Y S Y T A A
Rhesus monkey.....      - - - - - - - - G D V E K G K K I F I M K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G Y S Y T A A
Horse ..........        - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G F T Y T D A
Pig, cow, sheep...      - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G F S Y T D A
Dog.............        - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G F S Y T D A
Gray whale  ......      - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A V G F S Y T D A
Rabbit...........       - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A V G F S Y T D A
Kangaroo..........      - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L N G I F G R K T G Q A P G F T Y T D A
Chicken, Turkey...      - - - - - - - - G D I E K G K K I F V Q K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A E G F S Y T D A
Penguin...........      - - - - - - - - G D I E K G K K I F V Q K C S O C H T V E K G G K H K T G P N L H G I F G R K T G Q A E G F S Y T D A
Pekin duck........      - - - - - - - - G D V E K G K K I F V Q K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A E G F S Y T D A
Snapping turtle...      - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L N G L I G R K T G Q A E G F S Y T E A
Bullfrog..........      - - - - - - - - G D V E K G K K I F V Q K C A Q C H T C E K G G K H K V G P N L Y G L I G R K T G Q A A G F S Y T D A
Tuna.............       - - - - - - - - G D V A K G K K T F V Q K C A Q C H T V E N G G K H K V G P N L W G L F G R K T G Q A E G Y S Y T D A
Screwworm fly.....      - - - - G V P A G D V E K G K K I F V Q R C A Q C H T V E A G G K H K V G P N L H G L F G R K T G Q A A G F A Y T N A
Silkworm moth.....      - - - - G V P A G N A E N G K K I F V Q R C A Q C H T V E A G G K H K V G P N L H G F Y G R K T G Q A P G F S Y S N A
Wheat ...........       A S F S E A P P G N P D A G A K I F K T K C A Q C H T V D A G A G H K Q G P N L H G L F G R Q S G T T A G Y S Y S A A
Fungus (Neurospora)     - - - - G F S A G D S K K G A N L F K T R C A E C H G E G G N L T Q K I G P A L H G L F G R K T G S V D G Y A Y T D A
Fungus(baker's yeast)   - - - T E F K A G S A K K G A T L F K T R C E L C H T V E K G G P H K V G P N L H G I F G R H S G Q A Q G Y S Y T D A
Fungus (Candida)        - - P A P F E O G S A K K G A T L F K T R C A E C H T I E A G G P H K V G P N L H G I F S R H S G Q A Q G Y S Y T D A

[CONTINUED FROM ABOVE]
                        60                  70                  80                  90                  100                 110
Amino Acid Number-----> 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2
Human ............      N K N K G I I W G E D T L M E Y L E N P K K Y I P G T K M I F V G I K K K E E R A D L I A Y L K K A T N E
Rhesus monkey.....      N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F V G I K K K E E R A D L I A Y L K K A T N E
Horse ...........       N K N K G I T W K E E T L M E Y L E N P K K Y I P G T K M I F A G I K K K T E R E D L I A Y L K K A T N E
Pig, cow, sheep...      N K N K G I T W G E E T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R E D L I A Y L K K A T N E
Dog..............       N K N K G I T W G E E T L M E Y L E N P K K Y I P G T K M I F A G I K K T G E R A D L I A Y L K K A T K E
Gray whale........      N K N K G I T W G E E T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R A D L I A Y L K K A T N E
Rabbit...........       N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K D E R A D L I A Y L K K A T N E
Kangaroo..........      N K N K G I I W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R A D L I A Y L K K A T N E
Chicken, Turkey...      N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K S E R V D L I A Y L K D A T S K
Penguin...........      N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K S E R A D L I A Y L K D A T S K
Pekin duck........      N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K S E R A D L I A Y L K D A T A K
Snapping turtle...      N K N K G I T W G E E T L M E Y L E N P K K Y I P G T K M I F A G I K K K A E R A D L I A Y L K D A T S K
Bullfrog..........      N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R Q D L I A Y L K S A C S K
Tuna..............      N K S K G I V W N N D T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R Q D L V A Y L K S A T S -
Screwworm fly.....      N K A K G I T W Q D D T L F E Y L E N P K K Y I P G T K M I F A G L K K P N E R G D L I A Y L K S A T K -
Silkworm moth.....      N K A K G I T W G D D T L F E Y L E N P K K Y I P G T K M V F A G L K K A N E R A D L I A Y L K E S T K -
Wheat ............      N K N K A V E W E E N T L Y D Y L L N P K K Y I P G T K M V F P G L K K P Q D R A D L I A Y L K K A T S S
Fungus 1 (Neurospora)   N K Q K G I T W D E N T L F E Y L E N P K K Y I P G T K M A F G G L K K D K D R N D I I T F M K E A T A -
Fungus 2 (bkrs yeast)   N I K K N V L W D E N N M S E Y L T N P K K Y I P G T K M A F G G L K K E K D R N D L I T Y L K K A C E -
Fungus 3 (Candida)      N K R A G V E W A E P T M S D Y L E N P K K Y I P G T K M A F G G L K K A K D R N D L V T Y M L E A S K -

AMINO ACID SYMBOLS
A = Alanine
C = Cysteine
D = Aspartic acid
E = Glutamic acid
F = Phenylalanine
G = Glycine
H = Histidine
I = Isoleucine
K = Lysine
L = Leucine
M = Methionine
N = Asparagine
P = Proline
Q = Glutamine
R = Arginine
S = Serine
T = Threonine
V = Valine
W = Tryptohpan
Y = Tyrosine

Symbols in light blue or gray represent amino acids which show NO differences in any organism on the list, so you can ignore them.
(adapted from Strahler, Arthur, Science & Earth History, 1987, p. 348)
```


Cytochrome C Amino Acid Sequences



Some parts of the protein vary a lot.

Other parts hardly vary at all.

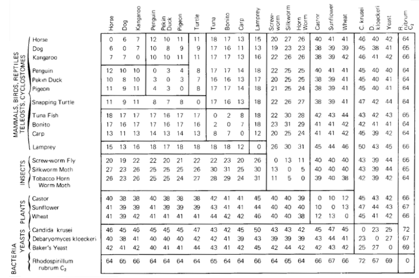
Percent Differences: Cytochrome C

(Dayhoff 1972)

Group	Species	Horse	Dog	Kangaroo	Penguin	Pekin Duck	Pigeon	Turtle	Tuna	Bonito	Carp	Lamprey	Screw-worm	Silkworm	Horn Worm	Castor	Sunflower	Wheat	C. krusei	D. kloeckeri	Yeast	R. rubrum C_2
MAMMALS, BIRDS, REPTILES TELEOSTS, CYCLOSTOMES	Horse	0	6	7	12	10	11	11	18	17	13	15	20	27	26	40	41	41	46	40	42	64
	Dog	6	0	7	10	8	9	9	17	16	11	13	19	23	23	38	39	39	45	38	41	65
	Kangaroo	7	7	0	10	10	11	11	17	17	13	16	22	26	26	38	39	42	46	41	42	66
	Penguin	12	10	10	0	3	4	8	17	17	14	18	22	25	25	40	41	41	45	40	40	64
	Pekin Duck	10	8	10	3	0	3	7	16	16	13	17	20	25	25	38	39	41	45	40	41	64
	Pigeon	11	9	11	4	3	0	8	17	17	14	18	21	25	24	38	39	41	45	40	41	64
	Snapping Turtle	11	9	11	8	7	8	0	17	16	13	18	22	26	27	38	39	41	47	42	44	64
	Tuna Fish	18	17	17	17	16	17	17	0	2	8	18	22	30	28	42	43	44	43	42	43	65
	Bonito	17	16	17	17	16	17	16	2	0	7	18	23	31	29	41	41	42	42	41	41	64
	Carp	13	11	13	14	13	14	13	8	7	0	12	20	25	24	41	41	42	45	39	42	64
	Lamprey	15	13	16	18	17	18	18	18	18	12	0	26	30	31	45	44	46	50	43	45	66
INSECTS	Screw-worm Fly	20	19	22	22	20	21	22	22	23	20	26	0	13	11	40	40	40	43	39	44	66
	Silkworm Moth	27	23	26	25	25	25	26	30	31	25	30	13	0	5	40	40	40	43	39	44	65
	Tobacco Horn Worm Moth	26	23	26	25	25	24	27	28	29	24	31	11	5	0	39	40	38	42	39	42	64
PLANTS	Castor	40	38	38	40	38	38	38	42	41	41	45	40	40	39	0	10	12	45	43	42	66
	Sunflower	41	39	39	41	39	39	39	43	41	41	44	40	40	40	10	0	13	47	44	43	67
	Wheat	41	39	42	41	41	41	41	44	42	42	46	40	40	38	12	13	0	45	41	42	66
YEASTS	Candida krusei	46	45	46	45	45	45	47	43	42	45	50	43	43	42	45	47	45	0	23	25	72
	Debaryomyces kloeckeri	40	38	41	40	40	40	42	42	41	39	43	39	39	39	43	44	41	23	0	27	67
	Baker's Yeast	42	41	42	40	41	41	44	43	41	42	45	42	44	42	42	43	42	25	27	0	69
BACTERIA	Rhodospirillum rubrum C_2	64	65	66	64	64	64	64	65	64	64	66	64	65	64	66	67	66	72	67	69	0

Percent Differences: Cytochrome C

(Dayhoff 1972)

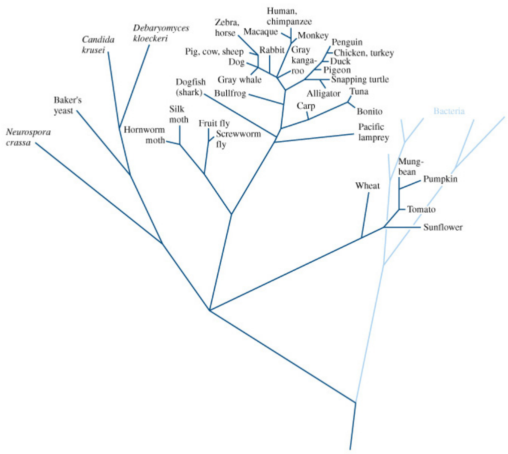


Within each off-diagonal block, percent differences are similar.

Implies equal relatedness between pairs of species.

Which implies a tree of relationships. . .

Cytochrome C phylogeny

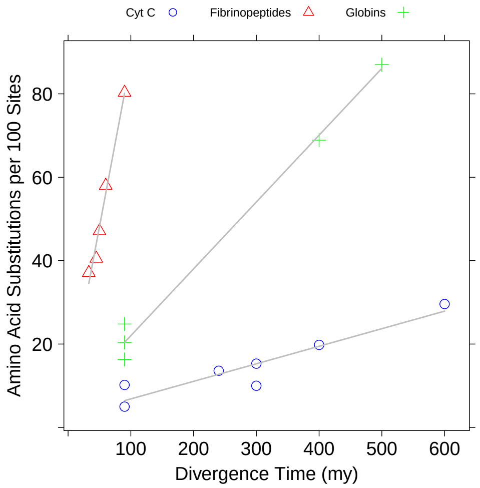


Nested similarities

Groups under groups

Agrees w/ morphology; other proteins.

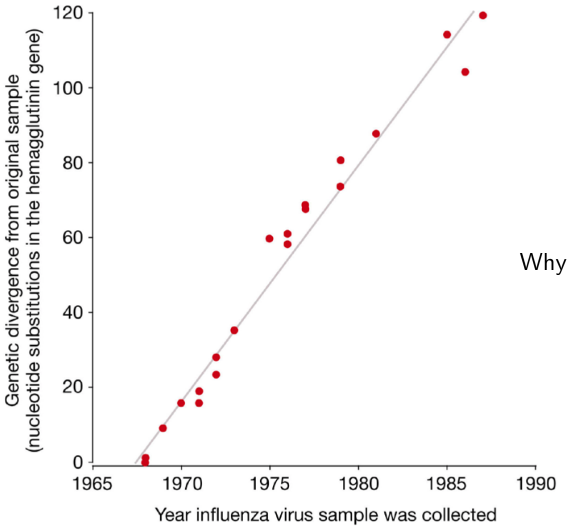
Protein differences versus separation time (Dickerson 1971)



Constant rates of evolution.

Rate differs among proteins.

Recent evolution of influenza virus



Why linear?

Outline

- The pattern in molecular data
- Hypothesis of neutral evolution
- Molecular clock hypothesis
- Why neutral evolution in linear
- Generation time

The neutral theory of molecular evolution

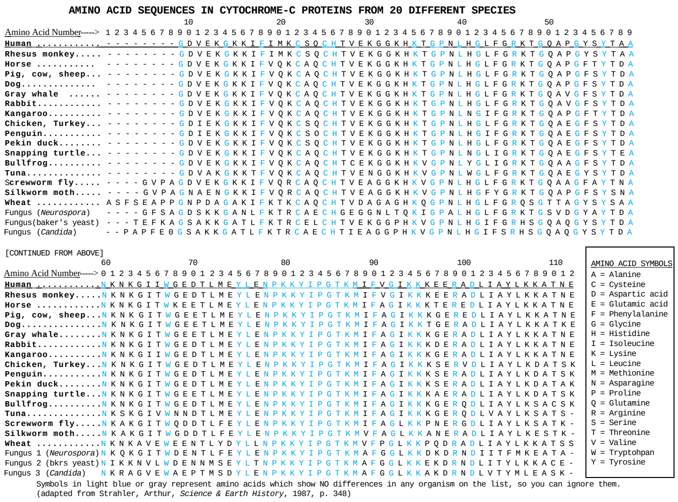
1. Most mutations are harmful and are removed by selection.
2. *Selectively neutral* mutations are not removed.
3. Most molecular variation within species is neutral.
4. Same is true of differences between species.

Evolution is fastest in the parts of the genome that matter least.

The hypothesis of functional constraint

"those amino-acids which are critical to the biological function of a molecule should be strongly conserved by natural selection. Other amino-acids may have general properties which make their presence desirable but not indispensable, and these residues should exhibit a lesser degree of evolutionary restraint. Other amino-aids may simply take up space." (Doolittle and Blombäck 1964)

Cytochrome C Amino Acid Sequences

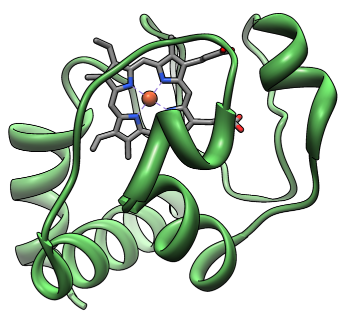


```
AMINO ACID SEQUENCES IN CYTOCHROME-C PROTEINS FROM 20 DIFFERENT SPECIES
                                     10                  20                  30                  40                  50
Amino Acid Number-----> 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
Human ............ - - - - - - - - G D V E K G K K I F I M K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G Y S Y T A A
Rhesus monkey..... - - - - - - - - G D V E K G K K I F I M K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G Y S Y T A A
Horse ...........  - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G F T Y T D A
Pig, cow, sheep... - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G F S Y T D A
Dog..............  - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A P G F S Y T D A
Gray whale  ...... - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A V G F S Y T D A
Rabbit............ - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A V G F S Y T D A
Kangaroo.......... - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L N G I F G R K T G Q A P G F T Y T D A
Chicken, Turkey... - - - - - - - - G D I E K G K K I F V Q K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A E G F S Y T D A
Penguin........... - - - - - - - - G D I E K G K K I F V Q K C S O C H T V E K G G K H K T G P N L H G I F G R K T G Q A E G F S Y T D A
Pekin duck........ - - - - - - - - G D V E K G K K I F V Q K C S Q C H T V E K G G K H K T G P N L H G L F G R K T G Q A E G F S Y T D A
Snapping turtle... - - - - - - - - G D V E K G K K I F V Q K C A Q C H T V E K G G K H K T G P N L N G L I G R K T G Q A E G F S Y T E A
Bullfrog.......... - - - - - - - - G D V E K G K K I F V Q K C A Q C H T C E K G G K H K V G P N L Y G L I G R K T G Q A A G F S Y T D A
Tuna.............. - - - - - - - - G D V A K G K K T F V Q K C A Q C H T V E N G G K H K V G P N L W G L F G R K T G Q A E G Y S Y T D A
Screwworm fly..... - - - - G V P A G D V E K G K K I F V Q R C A Q C H T V E A G G K H K V G P N L H G L F G R K T G Q A A G F A Y T N A
Silkworm moth..... - - - - G V P A G N A E N G K K I F V Q R C A Q C H T V E A G G K H K V G P N L H G F Y G R K T G Q A P G F S Y S N A
Wheat ............ A S F S E A P P G N P D A G A K I F K T K C A Q C H T V D A G A G H K Q G P N L H G L F G R Q S G T T A G Y S Y S A A
Fungus (Neurospora)  - - - - G F S A G D S K K G A N L F K T R C A E C H G E G G N L T Q K I G P A L H G L F G R K T G S V D G Y A Y T D A
Fungus(baker's yeast) - - - T E F K A G S A K K G A T L F K T R C E L C H T V E K G G P H K V G P N L H G I F G R H S G Q A Q G Y S Y T D A
Fungus (Candida)     - - P A P F E 0 G S A K K G A T L F K T R C A E C H T I E A G G P H K V G P N L H G I F S R H S G Q A Q G Y S Y T D A

[CONTINUED FROM ABOVE]
                        60                  70                  80                  90                  100                 110
Amino Acid Number-----> 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2
Human ............ N K N K G I I W G E D T L M E Y L E N P K K Y I P G T K M I F V G I K K K E E R A D L I A Y L K K A T N E
Rhesus monkey..... N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F V G I K K K E E R A D L I A Y L K K A T N E
Horse ............ N K N K G I T W K E E T L M E Y L E N P K K Y I P G T K M I F A G I K K K T E R E D L I A Y L K K A T N E
Pig, cow, sheep... N K N K G I T W G E E T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R E D L I A Y L K K A T N E
Dog............... N K N K G I T W G E E T L M E Y L E N P K K Y I P G T K M I F A G I K K T G E R A D L I A Y L K K A T K E
Gray whale........ N K N K G I T W G E E T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R A D L I A Y L K K A T N E
Rabbit............ N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K D E R A D L I A Y L K K A T N E
Kangaroo.......... N K N K G I I W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R A D L I A Y L K K A T N E
Chicken, Turkey... N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K S E R V D L I A Y L K D A T S K
Penguin........... N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K S E R A D L I A Y L K D A T S K
Pekin duck........ N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K S E R A D L I A Y L K D A T A K
Snapping turtle... N K N K G I T W G E E T L M E Y L E N P K K Y I P G T K M I F A G I K K K A E R A D L I A Y L K D A T S K
Bullfrog.......... N K N K G I T W G E D T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R Q D L I A Y L K S A C S K
Tuna.............. N K S K G I V W N N D T L M E Y L E N P K K Y I P G T K M I F A G I K K K G E R Q D L V A Y L K S A T S -
Screwworm fly..... N K A K G I T W Q D D T L F E Y L E N P K K Y I P G T K M I F A G L K K P N E R G D L I A Y L K S A T K -
Silkworm moth..... N K A K G I T W G D D T L F E Y L E N P K K Y I P G T K M V F A G L K K A N E R A D L I A Y L K E S T K -
Wheat ............ N K N K A V E W E E N T L Y D Y L L N P K K Y I P G T K M V F P G L K K P Q D R A D L I A Y L K K A T S S
Fungus 1 (Neurospora) N K Q K G I T W D E N T L F E Y L E N P K K Y I P G T K M A F G G L K K D K D R N D I I T F M K E A T A -
Fungus 2 (bkrs yeast) N I K K N V L W D E N N M S E Y L T N P K K Y I P G T K M A F G G L K K E K D R N D L I T Y L K K A C E -
Fungus 3 (Candida)    N K R A G V E W A E P T M S D Y L E N P K K Y I P G T K M A F G G L K K A K D R N D L V T Y M L E A S K -
      Symbols in light blue or gray represent amino acids which show NO differences in any organism on the list, so you can ignore them.
      (adapted from Strahler, Arthur, Science & Earth History, 1987, p. 348)

AMINO ACID SYMBOLS
A = Alanine
C = Cysteine
D = Aspartic acid
E = Glutamic acid
F = Phenylalanine
G = Glycine
H = Histidine
I = Isoleucine
K = Lysine
L = Leucine
M = Methionine
N = Asparagine
P = Proline
Q = Glutamine
R = Arginine
S = Serine
T = Threonine
V = Valine
W = Tryptohpan
Y = Tyrosine
```


Some parts of Cyt-C protein vary more than others

- Some positions never change.
- Invariant positions not scattered at random.
- Hypothesis: invariant positions are functionally important.

Why clumping of invariant positions?



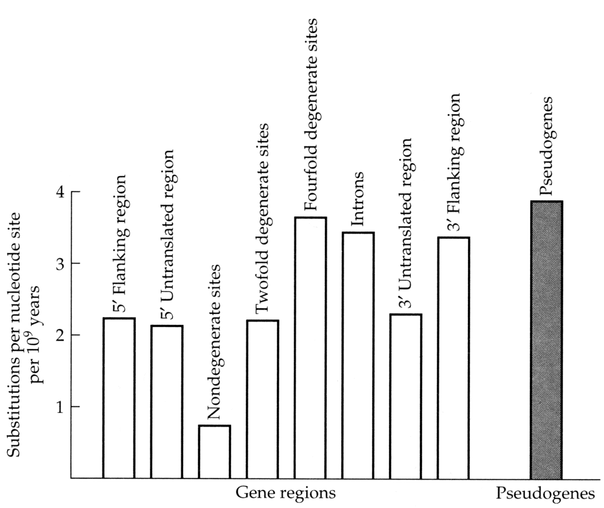
- Proteins fold
- Adjacent positions are close together on folded protein.
- Conserved regions bind with other molecules.

Introns evolve faster than exons (Fairbanks 2009)

1. Verticals link identical sites in pig and human INS genes.
2. More identity in exon (at left) than in intron (at right).
3. Identity in splice junction—1st 6 sites of intron.

Introns were not discovered until 1977—long after the theory of functional constraint.

Pseudogenes evolve faster than functional genes



The genetic code (DNA version)

First Position	Second Position				Third Position	
	A	G	T	C		
A	Phe	Ser	Tyr	Cys	A	(adenine)
	Phe	Ser	Tyr	Cys	G	(guanine)
	Leu	Ser	STOP	STOP	T	(thymine)
	Leu	Ser	STOP	Trp	C	(cytosine)
G	Leu	Pro	His	Arg	A	Few 1st-position changes are synonymous.
	Leu	Pro	His	Arg	G	
	Leu	Pro	Gln	Arg	T	
	Leu	Pro	Gln	Arg	C	
T	Ile	Thr	Asn	Ser	A	Weaker constraint in 2nd position.
	Ile	Thr	Asn	Ser	G	
	Ile	Thr	Lys	Arg	T	
	Met	Thr	Lys	Arg	C	
C	Val	Ala	Asp	Gly	A	Weakest in 3rd-position.
	Val	Ala	Asp	Gly	G	
	Val	Ala	Glu	Gly	T	
	Val	Ala	Glu	Gly	C	

```
1                                                 g
atg.tcg.ttt.act.ttg.acc.aac.aag.aac.gtg.att.ttc.gtt.gcc.ggt.ctg.gga.ggc.att.ggt
Met.Ser.Phe.Thr.Leu.Thr.Asn.Lys.Asn.Val.Ile.Phe.Val.Ala.Gly.Leu.Gly.Gly.Ile.Gly
61
ctg.gac.acc.agc.aag.gag.ctg.ctc.aag.cgc.gat.ctg.aag.aac.ctg.gtg.atc.ctc.gac.cgc
Leu.Asp.Thr.Ser.Lys.Glu.Leu.Leu.Lys.Arg.Asp.Leu.Lys.Asn.Leu.Val.Ile.Leu.Asp.Arg
121
att.gag.aac.ccg.gct.gcc.att.gcc.gag.ctg.aag.gca.atc.aat.cca.aag.gtg.acc.gtc.acc
Ile.Glu.Asn.Pro.Ala.Ala.Ile.Ala.Glu.Leu.Lys.Ala.Ile.Asn.Pro.Lys.Val.Thr.Val.Thr
181                                                         t
ttc.tac.ccc.tat.gat.gtg.acc.gtg.ccc.att.gcc.gag.acc.acc.aag.ctg.ctg.aag.acc.atc
Phe.Tyr.Pro.Tyr.Asp.Val.Thr.Val.Pro.Ile.Ala.Glu.Thr.Thr.Lys.Leu.Leu.Lys.Thr.Ile
241
ttc.gcc.cag.ctg.aag.acc.gtc.gat.gtc.ctg.atc.aac.gga.gct.ggt.atc.ctg.gac.gat.cac
Phe.Ala.Gln.Leu.Lys.Thr.Val.Asp.Val.Leu.Ile.Asn.Gly.Ala.Gly.Ile.Leu.Asp.Asp.His
301
cag.atc.gag.cgc.acc.att.gcc.gtc.aac.tac.act.ggc.ctg.gtc.aac.acc.acg.acg.gcc.att
Gln.Ile.Glu.Arg.Thr.Ile.Ala.Val.Asn.Tyr.Thr.Gly.Leu.Val.Asn.Thr.Thr.Thr.Ala.Ile
361                               t       a
ctg.gac.ttc.tgg.gac.aag.cgc.aag.ggc.ggt.ccc.ggt.ggt.atc.atc.tgc.aac.att.gga.tcc
Leu.Asp.Phe.Trp.Asp.Lys.Arg.Lys.Gly.Gly.Pro.Gly.Gly.Ile.Ile.Cys.Asn.Ile.Gly.Ser
421                       a
gtc.act.gga.ttc.aat.gcc.atc.tac.cag.gtg.ccc.gtc.tac.tcc.ggc.acc.aag.gcc.gcc.gtg
Val.Thr.Gly.Phe.Asn.Ala.Ile.Tyr.Gln.Val.Pro.Val.Tyr.Ser.Gly.Thr.Lys.Ala.Ala.Val
481                                       a       c               g           t
gtc.aac.ttc.acc.agc.tcc.ctg.gcg.aaa.ctg.gcc.ccc.att.acc.ggc.gtg.acc.gct.tac.acc
Val.Asn.Phe.Thr.Ser.Ser.Leu.Ala.Lys.Leu.Ala.Pro.Ile.Thr.Gly.Val.Thr.Ala.Tyr.Thr
541                                              c
gtg.aac.ccc.ggc.atc.acc.cgc.acc.acc.ctg.gtg.cac.aag.ttc.aac.tcc.tgg.ttg.gat.gtt
Val.Asn.Pro.Gly.Ile.Thr.Arg.Thr.Thr.Leu.Val.His.Lys.Phe.Asn.Ser.Trp.Leu.Asp.Val
601   t           c                                       c
gag.ccc.cag.gtt.gct.gag.aag.ctc.ctg.gct.cat.ccc.acc.cag.cca.tcg.ttg.gcc.tgc.gcc
Glu.Pro.Gln.Val.Ala.Glu.Lys.Leu.Leu.Ala.His.Pro.Thr.Gln.Pro.Ser.Leu.Ala.Cys.Ala
661                           a
gag.aac.ttc.gtc.aag.gct.atc.gag.ctg.aac.cag.aac.gga.gcc.atc.tgg.aaa.ctg.gac.ctg
Glu.Asn.Phe.Val.Lys.Ala.Ile.Glu.Leu.Asn.Gln.Asn.Gly.Ala.Ile.Trp.Lys.Leu.Asp.Leu
721
ggc.acc.ctg.gag.gcc.atc.cag.tgg.acc.aag.cac.tgg.gac.tcc.ggc.atc.
Gly.Thr.Leu.Glu.Ala.Ile.Gln.Trp.Thr.Lys.His.Trp.Asp.Ser.Gly.Ile.
```

ADH in *D. melanogaster*

Nucleotides & amino acids.

Most variants in 3rd position.

Only 1 (pos 578) changes amino acid.

Outline

- ○ The pattern in molecular data
- ○ Hypothesis of neutral evolution
- ▶ Why neutral evolution in linear
- ▶ Generation time

Genetic drift

Variant forms of a gene or a nucleotide site are called "alleles."

The "frequency" of an allele is its fraction w/i the population.

Neutral alleles increase or decrease in frequency just by chance. This is called "genetic drift."

Occasionally, one drifts to extinction and is lost.

Eventually, all are lost but one. The remaining allele has frequency 100% and is said to be "fixed" in the population.

If the surviving allele arose from a single mutation, then at this genetic locus, all copies of the gene descend from that ancestral mutation.

Fixation probability of a new neutral mutation

What is the probability that a brand new neutral mutation, which exists only in a single copy, will eventually become 100% of the population?

In a population of size N, there are $2N$ gene copies, one of which is our new mutant. Eventually, all but 1 will be lost.

If they are selectively neutral, each gene copy has the same chance ($1/2N$) of fixation.

The fixation probability of a neutral mutation is $1/2N$.

Why neutral evolution is linear

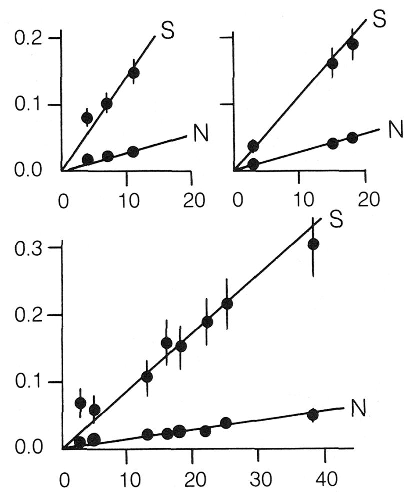
In a population of size N and mutation rate u, there are $2Nu$ new mutations each generation.

A fraction $1/2N$ of these drifts to fixation.

Resulting rate: $2Nu \times \frac{1}{2N} = u$.

Constant mutation rate $\Rightarrow$ const rate of molecular evolution.

Synonymous versus nonsynonymous clocks



Within each protein, synonymous (S) sites evolve faster than non-synonymous (N) sites.

Evolution is linear in either case.

Outline

- ◦ The pattern in molecular data
- ◦ Hypothesis of neutral evolution
- ◦ Why neutral evolution in linear
- ▶ Generation time

The neutral theory again

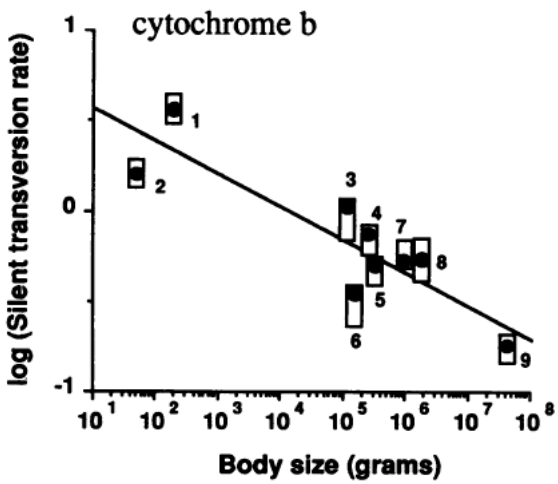
Under strictly neutral evolution, the rate of evolution equals the mutation rate, u.

If u were constant *per generation*, then the rate of evolution should scale with generation time. Large animals have long generation times and should evolve slowly.

Denton (1985, p. 287) argues this isn't so

"Since the rodent order diverged from the primate, it is practically certain that the line leading to mouse has undergone nearly 100 times as many reproductive cycles as that leading to man. If mutation rates are practically constant per generation how then could drift have generated equal rates of genetic divergence in mice and men?"

He's right—large animals evolve more slowly



(Martin & Palumbi 1993)

The body-size effect is not as large as Denton thought

- ▶ Most mutations happen during cell division. The mutation rate per generation depends on the number of cell divisions per generation. This number is larger for large animals than for small ones.
- ▶ Large animals have small populations; this increases fraction of mutations that are effectively neutral.
- ▶ For both reasons, the generation-time effect is smaller than Denton assumed.
- ▶ Clearly detectable only in 1980s.

Summary

1. Molecular differences accumulate at roughly constant rate.
2. Evolution is fastest where selective constraint is weak.
3. Supports idea that most molecular evolution is neutral.
4. Does not imply that neutral evolution at level of organism.