The Rest is Silence

Max Chatnoir
(Mary Anne Clark)
Professor of Biology
Texas Wesleyan University
Science Circle, 6 February 2021

What is a Mutation?

- A mutation is an change in a DNA sequence.
- Mutations occur randomly, either during DNA replication or as a consequence of modifying DNA bases by chemicals or radiation.
- Most mutations are single nucleotide replacements.
- Larger scale mutations may occur by errors in crossing over during meiosis.
- Mutations in a protein coding gene can change the protein.
- Mutations that occur in gamete producing cells may be inherited.

Consequences of Mutation

- Many genes are protein-coding, and determine the amino acid sequences of proteins.
- Example: Beta Globin
- Reference DNA coding sequence:

ATGGTGCATCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAA
GGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGCTGCTGGTGGTCT
ACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGAT
GCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGC
CTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACAC
TGAGTGAGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGCTC
CTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCAC
CCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCC
TGGCCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAA

- Translation

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPD
AVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRL
LGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH

Single Nucleotide Changes

Reference DNA sequence (partial) and translation:

ATGGTGCATCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCC

M V H L T P E E K S A V (Normal)

ATGGTGCATCTGACTCCTGTGGAGAAGTCTGCCGTTACTGCC

M V H L T P V E K S A V (AA Sub)

ATGGTGCATCTGACTCCTTAGGAGAAGTCTGCCGTTACTGCC

M V H L T P * E K S A V (Stop)

ATGGTGCATCTGACTCCTGAAGAGAAGTCTGCCGTTACTGCC

M V H L T P E E K S A V (Silent)

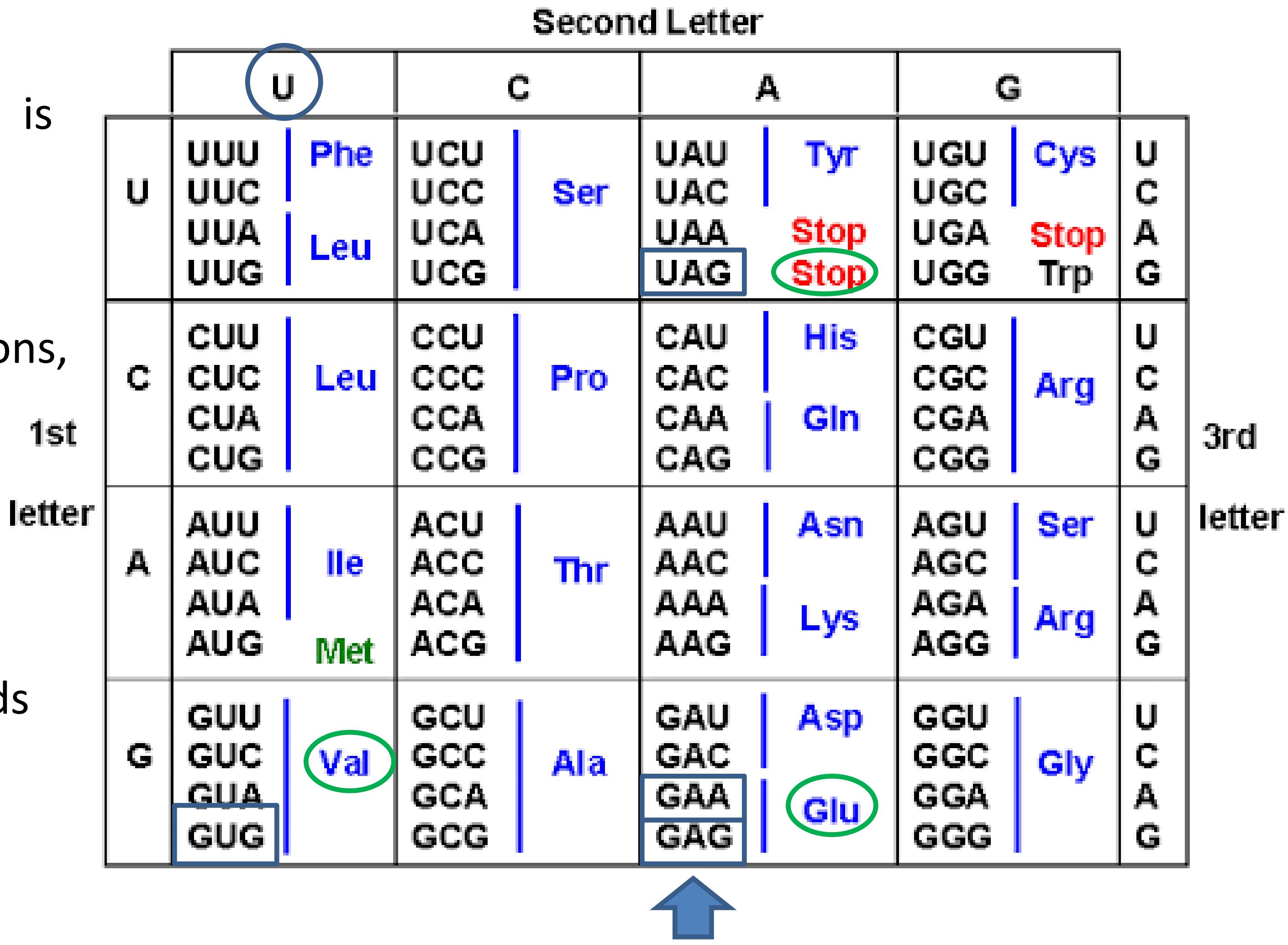
Why Are Some Mutations Silent?

Each amino acid is specified by a 3-base codon.

64 possible codons, all used.

There are only 20 amino acids

Most amino acids have more than one codon.

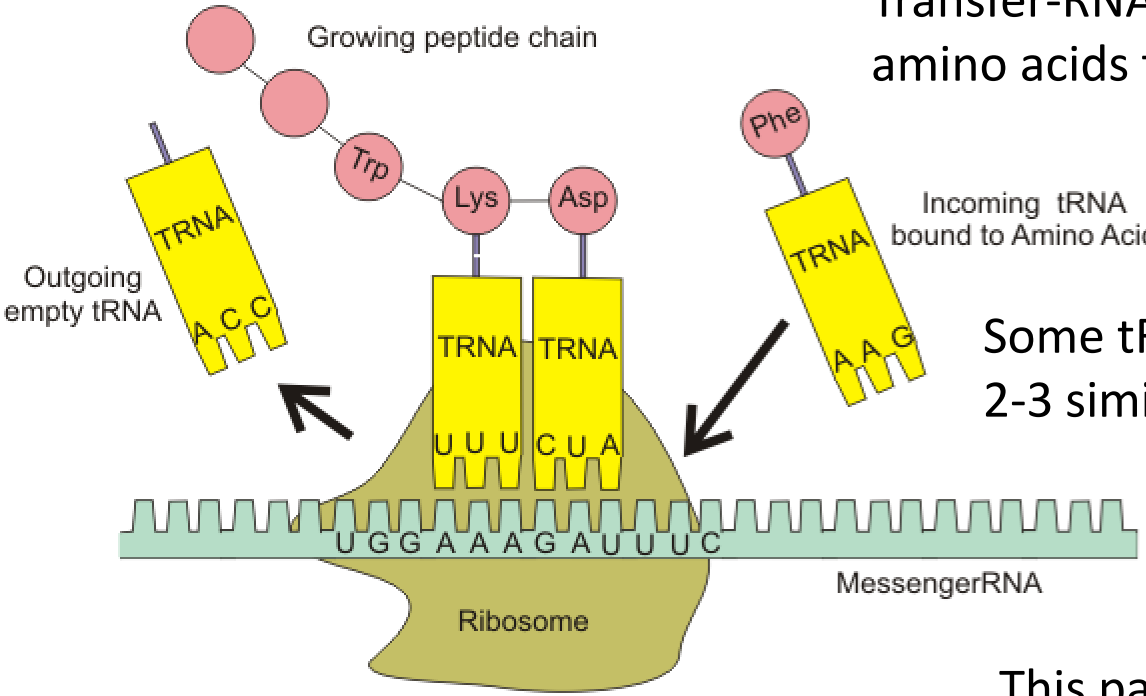


Second Letter

1st letter	U		C		A		G		3rd letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

https://pt.slideshare.net/namarta28/genetic-code-12022220/17

The Code is Translated on Ribosomes



Peptide Synthesis

Transfer-RNAs (tRNAs) match amino acids to specific codons

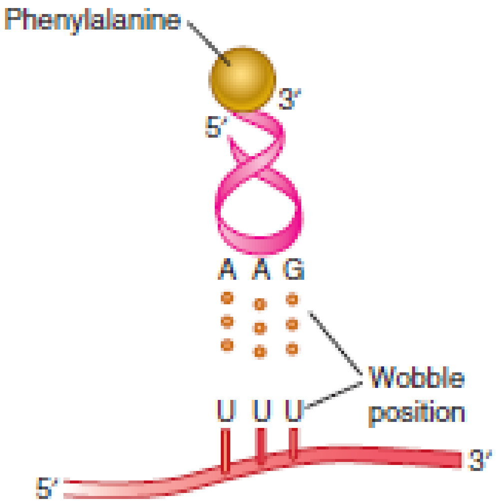
Some tRNAs can recognize 2-3 similar codons.

This partial mismatch is called “wobble.”

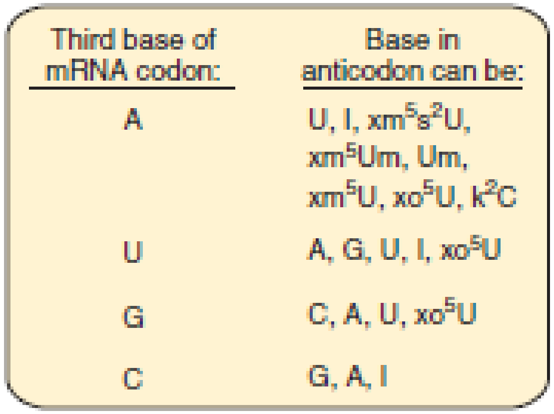
https://en.wikipedia.org/wiki/Transfer_RNA

The Wobble Rules

The first base in the anticodon loop of a tRNA matches the 3rd of the 3 bases in an RNA codon. This is called the wobble position. In tRNAs, this position may be occupied by one of tRNA's many modified bases.



(a) Location of wobble position



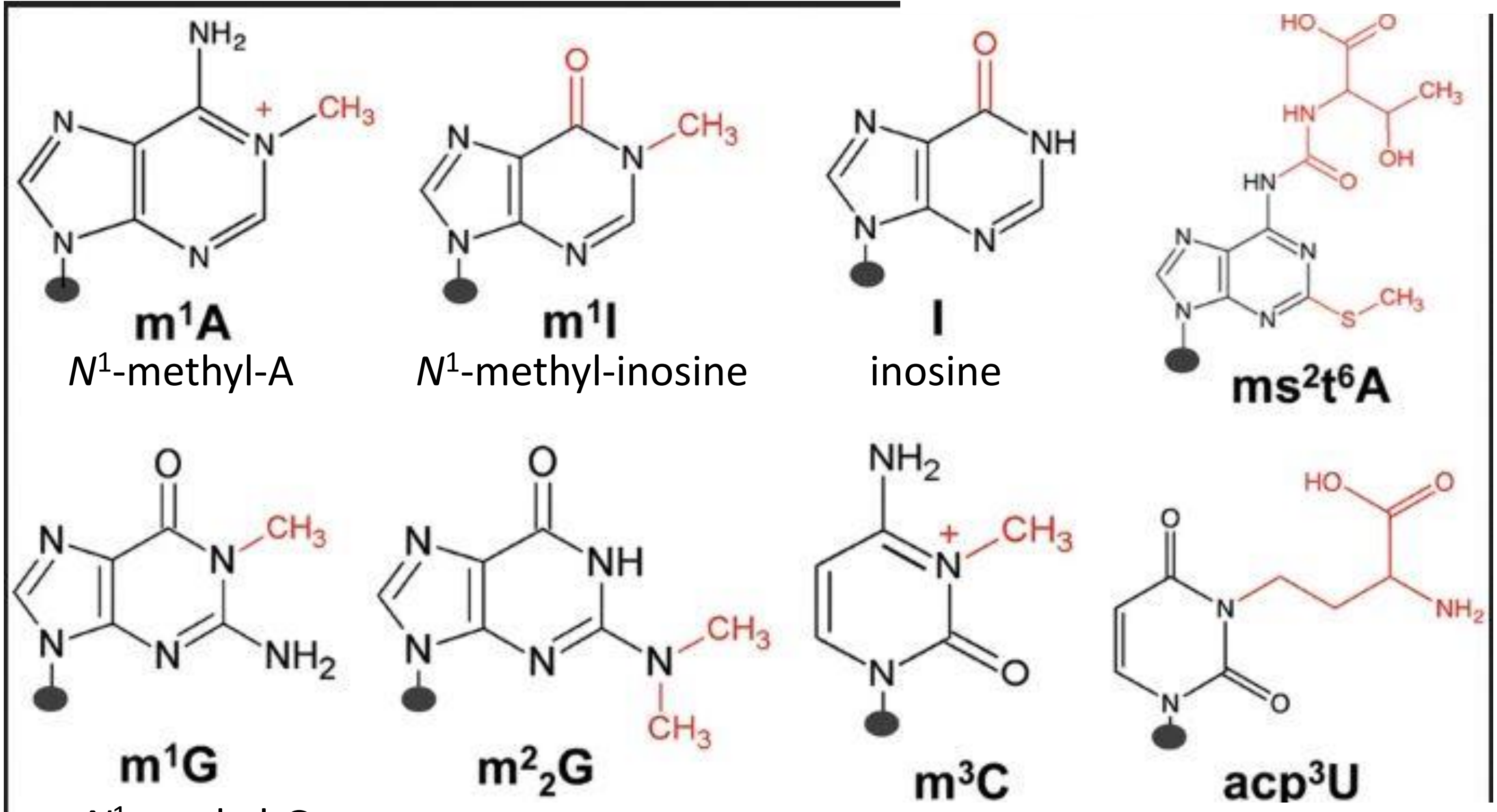
Third base of mRNA codon:	Base in anticodon can be:
A	U, I, xm^{5}s^{2}U, xm^{5}Um, Um, xm^{5}U, xo^{5}U, k^{2}C
U	A, G, U, I, xo^{5}U
G	C, A, U, xo^{5}U
C	G, A, I

(b) Revised wobble rules

https://www.chegg.com/homework-help/

Modified tRNA bases

2-methylthio- N^6-threonylcarbamoyl-A



N^1-methyl-A

N^1-methyl-inosine

inosine

N^1-methyl-G

N^2-dimethyl-G

3-methyl-C

3-(3-amino-
3-carboxypropyl)-U

Pan 2018. https://www.nature.com/articles/s41422-018-0013-y

Why am I even thinking about this?

- Last semester, I asked my students a question: If you look at the same protein-coding gene in two different species, which is likely to be more different: the coding sequence of the gene or the amino acid sequence of the encoded protein?
- In the genetic code, 3 bases specify one amino acid. So a coding sequence of 300 bases would specify a protein of 100 amino acids.
- Assume one base change = one amino acid change, so changing 30 bases would change 30 amino acids.
- 30 bases/300 = 10%. 30 amino acids/100 = 30%. So, as a percentage, the protein changes more than the DNA.

BUT….

- In the genetic code there is more than one codon per amino acid.
- Most amino acids have either 4 codons or 2 codons, for an average number of 3 codons/amino acid.
- Changes in the third base of the codon usually don’t change the amino acid.
- So only 2/3 of the mutations will change the amino acid.
- So of our 30 mutations in the 300 base coding sequence are still 10% of the DNA. But now only 20 of those changes will change the amino acid. So the protein changes by 20%. It still changes more than the DNA.

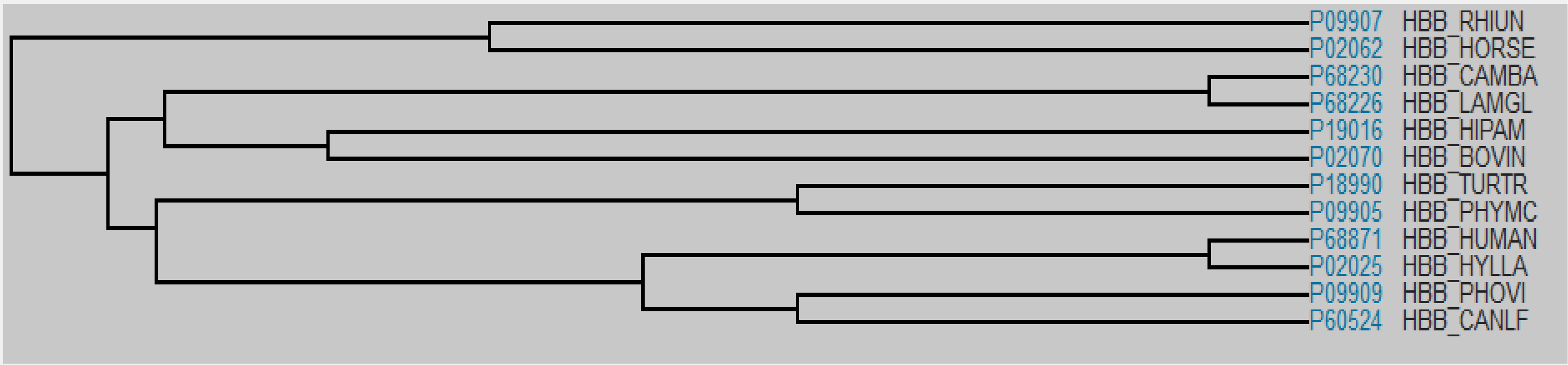
Testing the Hypothesis

- So I have a hypothesis that changing the DNA by 10% will change the protein by 20%.
- I love to snoop in the databases, and this is a question that can be answered in the databases.
- I (or the students) could compare the DNA for a given gene for two different species.
- Compare the protein encoded by that DNA in the same two species.
- This is the story of what I learned by snooping.

Mutations Accumulate over Time

- Mutations are the primary source of evolutionary differences.
- If mutations occur more or less randomly, then the longer two species have been separated, the more mutations will have accumulated between them.
- Note the differences between proteins like the beta globins are more a record of evolutionary diversification than a cause.
- This is the basis for molecular phylogenies.
- Sample phylogeny based on the beta globins of 12 species: human, dog, horse, cow, camel, llama, whale, dolphin, seal, rhino, gibbon.

Beta Globin Phylogeny



We have six sister groups: horse-rhino, camel-llama, hippo-cow, dolphin-whale, human-gibbon and seal-dog. The two closest are camel-llama and human-gibbon. The furthest pair is hippo-cow.

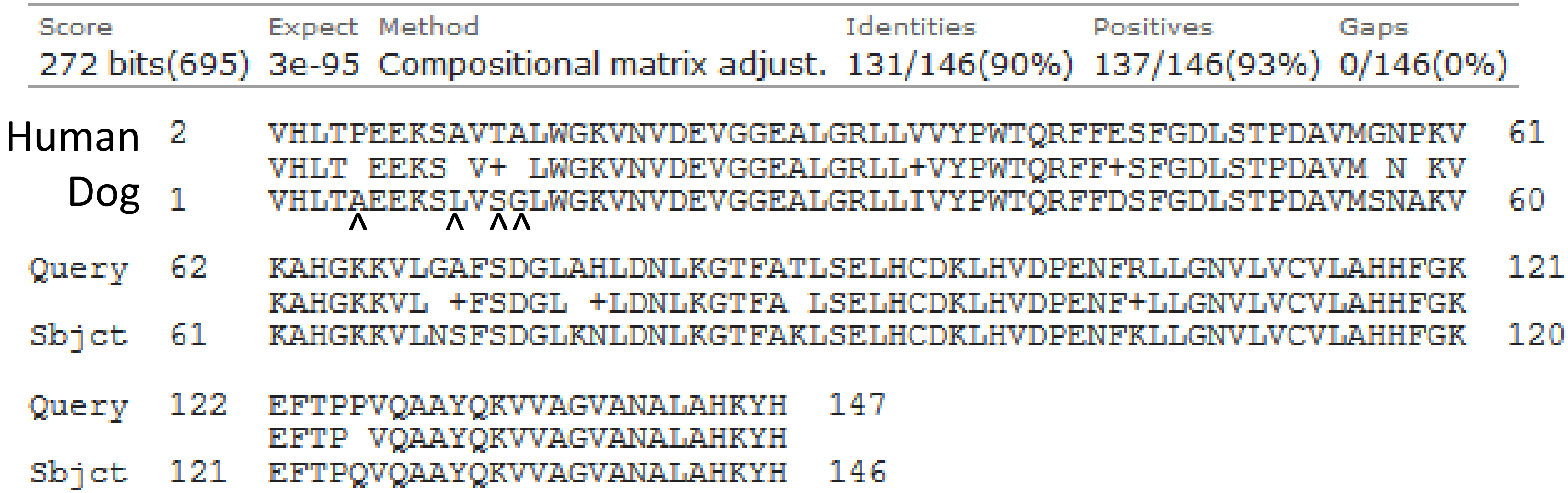
OK, Back to Silence!

- These phylogenies are based on protein differences, i.e. on mutations that are NOT silent.
- How can we detect silent mutations?
- We start with three proteins:
  - Beta globin
  - Amelogenin
  - Rhodopsin
- …and two species: humans and dogs (*Canis lupus familiaris*)

Beta Globin (Human)

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMG
NPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLA
HHFGKEFTPPVQAAYQKVVAGVANALAHKYH

Human vs Dog protein: 15 differences



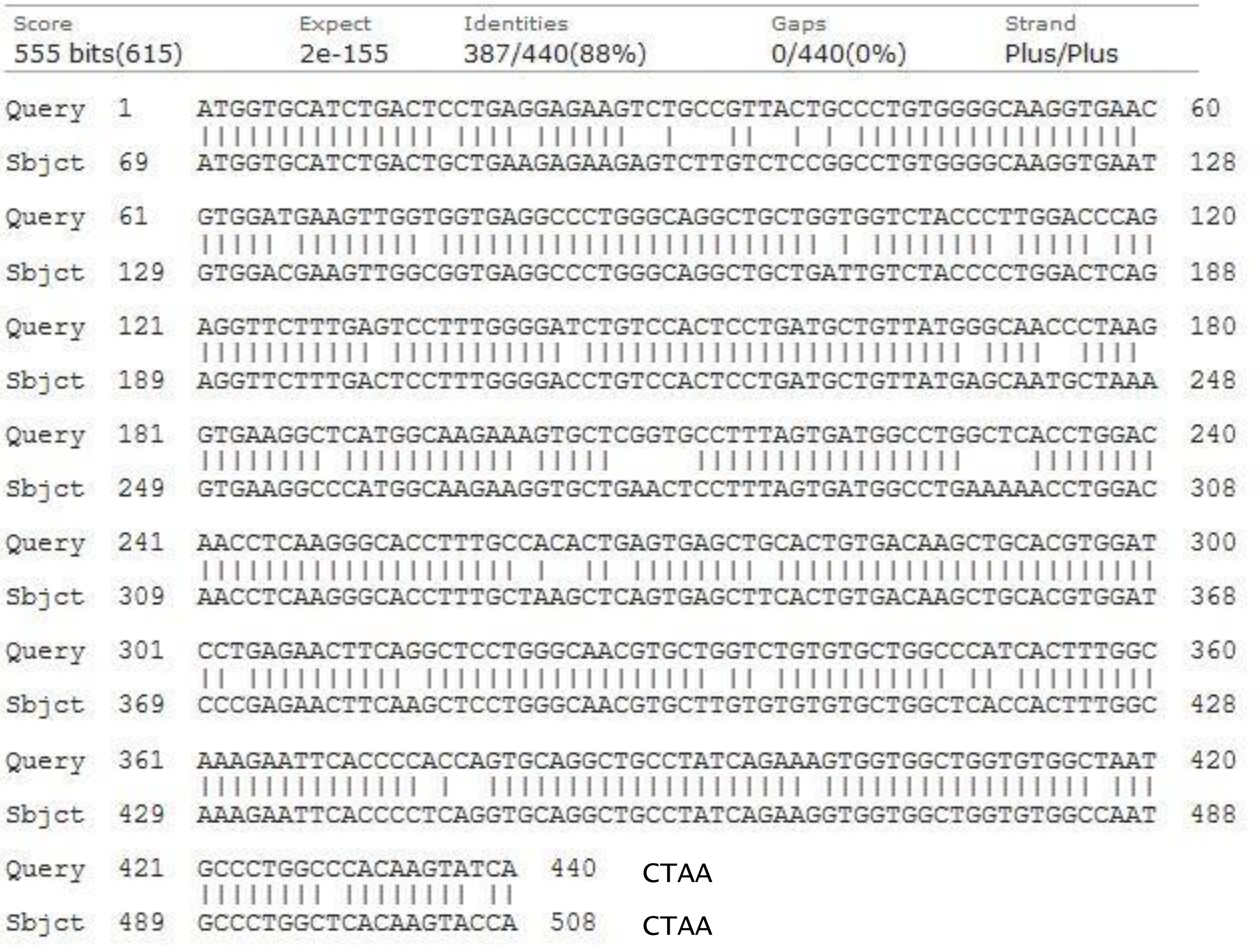
```
Score           Expect  Method                        Identities      Positives       Gaps
272 bits(695)   3e-95   Compositional matrix adjust.  131/146(90%)    137/146(93%)    0/146(0%)

Human  2      VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV  61
              VHLT EEKS V+ LWGKVNVDEVGGEALGRLL+VYPWTQRFF+SFGDLSTPDAVM N KV
  Dog  1      VHLTAEEKSLVSGLWGKVNVDEVGGEALGRLLIVYPWTQRFFDSFGDLSTPDAVMSNAKV  60
                  ^    ^ ^^

Query  62     KAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK  121
              KAHGKKVL +FSDGL +LDNLKGTFA LSELHCDKLHVDPENF+LLGNVLVCVLAHHFGK
Sbjct  61     KAHGKKVLNSFSDGLKNLDNLKGTFAKLSELHCDKLHVDPENFKLLGNVLVCVLAHHFGK  120

Query  122    EFTPPVQAAYQKVVAGVANALAHKYH  147
              EFTP VQAAYQKVVAGVANALAHKYH
Sbjct  121    EFTPQVQAAYQKVVAGVANALAHKYH  146
```

Beta Globin CDS:Human vs Dog

Human vs Dog CDS: 53 differences

15 are reflected by amino acid changes.
The rest (38 differences) are silent.

Human
Dog



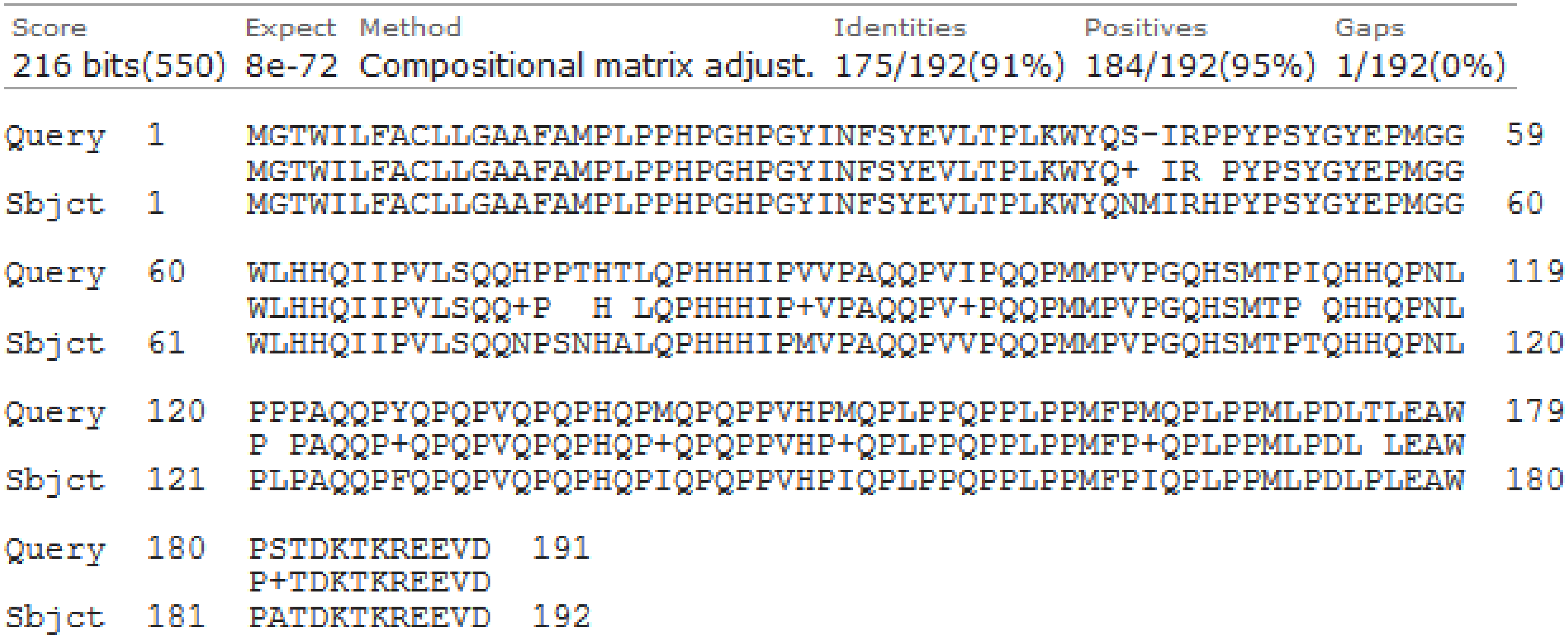
```
Score            Expect     Identities       Gaps          Strand
555 bits(615)    2e-155     387/440(88%)     0/440(0%)     Plus/Plus

Query  1    ATGGTGCATCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAAC  60
            ||||||||||||||| |||| ||||||  |   ||  | | | ||||||||||||||||||
Sbjct  69   ATGGTGCATCTGACTGCTGAAGAGAAGAGTCTTGTCTCCGGCCTGTGGGGCAAGGTGAAT  128

Query  61   GTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGCTGCTGGTGGTCTACCCTTGGACCCAG  120
            ||||| |||||||| |||||||||||||||||||||||| | |||||||| ||||| |||
Sbjct  129  GTGGACGAAGTTGGCGGTGAGGCCCTGGGCAGGCTGCTGATTGTCTACCCCTGGACTCAG  188

Query  121  AGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAG  180
            ||||||||||| ||||||||||| |||||||||||||||||||||||| ||||  |||| 
Sbjct  189  AGGTTCTTTGACTCCTTTGGGGACCTGTCCACTCCTGATGCTGTTATGAGCAATGCTAAA  248

Query  181  GTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGAC  240
            |||||||| ||||||||||| |||||     |||||||||||||||||    ||||||||
Sbjct  249  GTGAAGGCCCATGGCAAGAAGGTGCTGAACTCCTTTAGTGATGGCCTGAAAAACCTGGAC  308

Query  241  AACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT  300
            |||||||||||||||||||| |  || |||||||| ||||||||||||||||||||||||
Sbjct  309  AACCTCAAGGGCACCTTTGCTAAGCTCAGTGAGCTTCACTGTGACAAGCTGCACGTGGAT  368

Query  301  CCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGC  360
            || |||||||||| |||||||||||||||||| || ||||||||||| || |||||||||
Sbjct  369  CCCGAGAACTTCAAGCTCCTGGGCAACGTGCTTGTGTGTGTGCTGGCTCACCACTTTGGC  428

Query  361  AAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAAT  420
            |||||||||||||| |  |||||||||||||||||||| ||||||||||||||||| |||
Sbjct  429  AAAGAATTCACCCCTCAGGTGCAGGCTGCCTATCAGAAGGTGGTGGCTGGTGTGGCCAAT  488

Query  421  GCCCTGGCCCACAAGTATCA  440
            |||||||| |||||||| ||
Sbjct  489  GCCCTGGCTCACAAGTACCA  508
```

CTAA
CTAA

Amelogenin (Human)

MGTWILFACLLGAAFAMPLPPHPGHPGYINFSYEVLTPLKWYQSIRPPYPSYGYEP
MGGWLHHQIIPVLSQQHPPTHTLQPHHHIPVVPAQQPVIPQQPMMPVPGQHS
MTPIQHHQPNLPPPAQQPYQPQPVQPQPHQPMQPQPPVHPMQPLPPQPPLP
PMFPMQPLPPMLPDLTLEAWPSTDKTKREEVD

Human vs Dog protein: 17 differences (1 in a gap, 16 in the CDS)

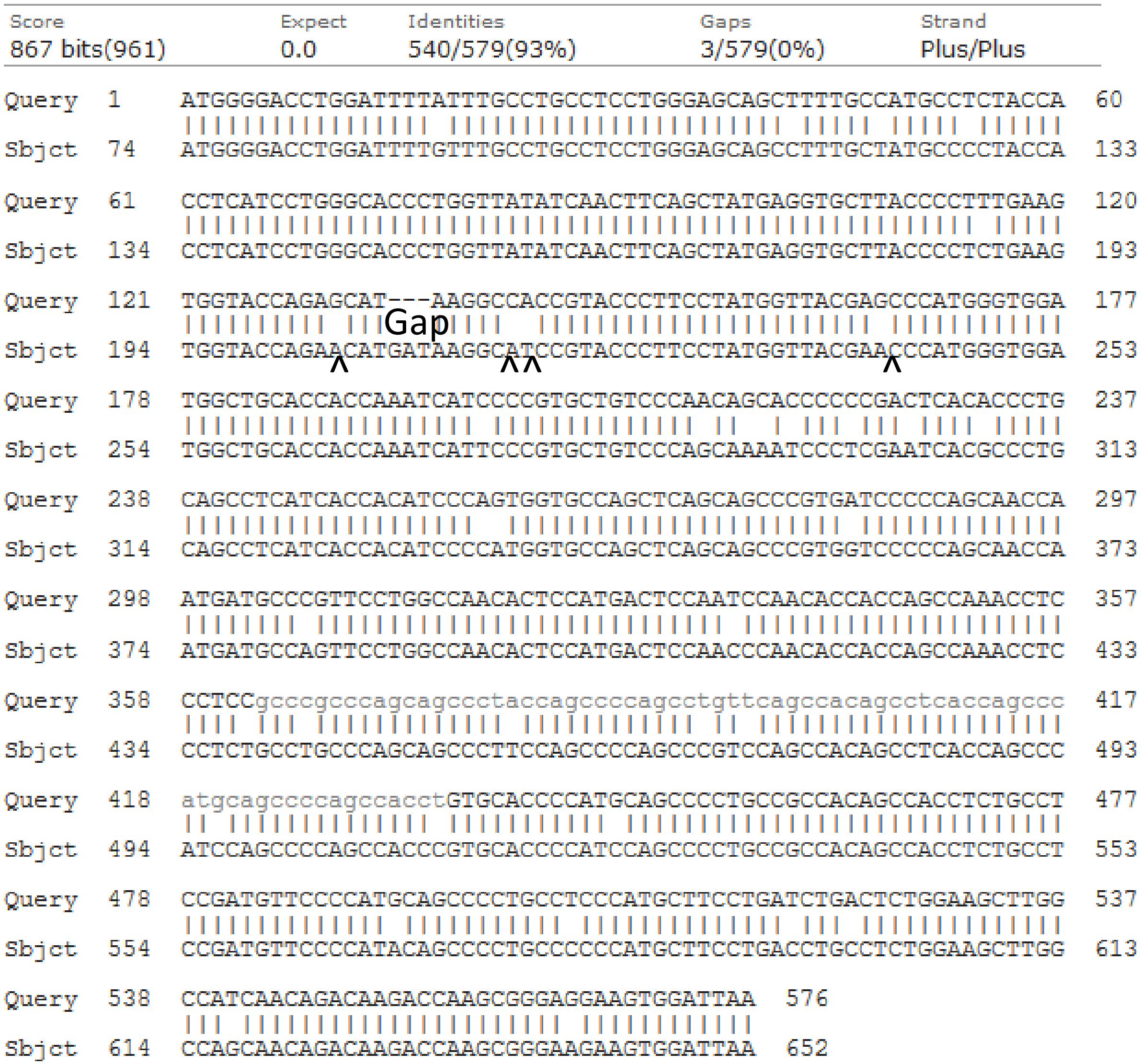


Human

Dog

```
Score            Expect  Method                        Identities      Positives       Gaps
216 bits(550)    8e-72   Compositional matrix adjust.  175/192(91%)    184/192(95%)    1/192(0%)

Query  1    MGTWILFACLLGAAFAMPLPPHPGHPGYINFSYEVLTPLKWYQS-IRPPYPSYGYEPMGG  59
            MGTWILFACLLGAAFAMPLPPHPGHPGYINFSYEVLTPLKWYQ+ IR PYPSYGYEPMGG
Sbjct  1    MGTWILFACLLGAAFAMPLPPHPGHPGYINFSYEVLTPLKWYQNMIRHPYPSYGYEPMGG  60

Query  60   WLHHQIIPVLSQQHPPTHTLQPHHHIPVVPAQQPVIPQQPMMPVPGQHSMTPIQHHQPNL  119
            WLHHQIIPVLSQQ+P  H LQPHHHIP+VPAQQPV+PQQPMMPVPGQHSMTP QHHQPNL
Sbjct  61   WLHHQIIPVLSQQNPSNHALQPHHHIPMVPAQQPVVPQQPMMPVPGQHSMTPTQHHQPNL  120

Query  120  PPPAQQPYQPQPVQPQPHQPMQPQPPVHPMQPLPPQPPLPPMFPMQPLPPMLPDLTLEAW  179
            P PAQQP+QPQPVQPQPHQP+QPQPPVHP+QPLPPQPPLPPMFP+QPLPPMLPDL LEAW
Sbjct  121  PLPAQQPFQPQPVQPQPHQPIQPQPPVHPIQPLPPQPPLPPMFPIQPLPPMLPDLPLEAW  180

Query  180  PSTDKTKREEVD  191
            P+TDKTKREEVD
Sbjct  181  PATDKTKREEVD  192
```

Amelogenin CDS Human vs Dog



Score	Expect	Identities	Gaps	Strand
867 bits(961)	0.0	540/579(93%)	3/579(0%)	Plus/Plus

```
Human  Query  1      ATGGGGACCTGGATTTTATTTGCCTGCCTCCTGGGAGCAGCTTTTGCCATGCCTCTACCA  60
                     ||||||||||||||||| ||||||||||||||||||||||| ||||| ||||| ||||||
  Dog  Sbjct  74     ATGGGGACCTGGATTTTGTTTGCCTGCCTCCTGGGAGCAGCCTTTGCTATGCCCCTACCA  133

       Query  61     CCTCATCCTGGGCACCCTGGTTATATCAACTTCAGCTATGAGGTGCTTACCCCTTTGAAG  120
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
       Sbjct  134    CCTCATCCTGGGCACCCTGGTTATATCAACTTCAGCTATGAGGTGCTTACCCCTCTGAAG  193

       Query  121    TGGTACCAGAGCAT---AAGGCCACCGTACCCTTCCTATGGTTACGAGCCCATGGGTGGA  177
                     |||||||||| |||Gap|||||  ||||||||||||||||||||||| ||||||||||||
       Sbjct  194    TGGTACCAGAACATGATAAGGCATCCGTACCCTTCCTATGGTTACGAACCCATGGGTGGA  253
                               ^           ^^                       ^
       Query  178    TGGCTGCACCACCAAATCATCCCCGTGCTGTCCCAACAGCACCCCCCGACTCACACCCTG  237
                     |||||||||||||||||||| |||||||||||||| ||  | ||| ||| |||| |||||
       Sbjct  254    TGGCTGCACCACCAAATCATTCCCGTGCTGTCCCAGCAAAATCCCTCGAATCACGCCCTG  313

       Query  238    CAGCCTCATCACCACATCCCAGTGGTGCCAGCTCAGCAGCCCGTGATCCCCCAGCAACCA  297
                     ||||||||||||||||||||  ||||||||||||||||||||||||| ||||||||||||
       Sbjct  314    CAGCCTCATCACCACATCCCCATGGTGCCAGCTCAGCAGCCCGTGGTCCCCCAGCAACCA  373

       Query  298    ATGATGCCCGTTCCTGGCCAACACTCCATGACTCCAATCCAACACCACCAGCCAAACCTC  357
                     |||||||| ||||||||||||||||||||||||||| |||||||||||||||||||||||
       Sbjct  374    ATGATGCCAGTTCCTGGCCAACACTCCATGACTCCAACCCAACACCACCAGCCAAACCTC  433

       Query  358    CCTCCgcccgcccagcagccctaccagccccagcctgttcagccacagcctcaccagccc  417
                     |||| ||| ||||||||||||| |||||||||||| || |||||||||||||||||||||
       Sbjct  434    CCTCTGCCTGCCCAGCAGCCCTTCCAGCCCCAGCCCGTCCAGCCACAGCCTCACCAGCCC  493

       Query  418    atgcagccccagccacctGTGCACCCCATGCAGCCCCTGCCGCCACAGCCACCTCTGCCT  477
                     || |||||||||||||| ||||||||||| ||||||||||||||||||||||||||||||
       Sbjct  494    ATCCAGCCCCAGCCACCCGTGCACCCCATCCAGCCCCTGCCGCCACAGCCACCTCTGCCT  553

       Query  478    CCGATGTTCCCCATGCAGCCCCTGCCTCCCATGCTTCCTGATCTGACTCTGGAAGCTTGG  537
                     |||||||||||||| ||||||||||| |||||||||||||| ||| ||||||||||||||
       Sbjct  554    CCGATGTTCCCCATACAGCCCCTGCCCCCCATGCTTCCTGACCTGCCTCTGGAAGCTTGG  613

       Query  538    CCATCAACAGACAAGACCAAGCGGGAGGAAGTGGATTAA  576
                     ||| |||||||||||||||||||||| ||||||||||||
       Sbjct  614    CCAGCAACAGACAAGACCAAGCGGGAAGAAGTGGATTAA  652
```

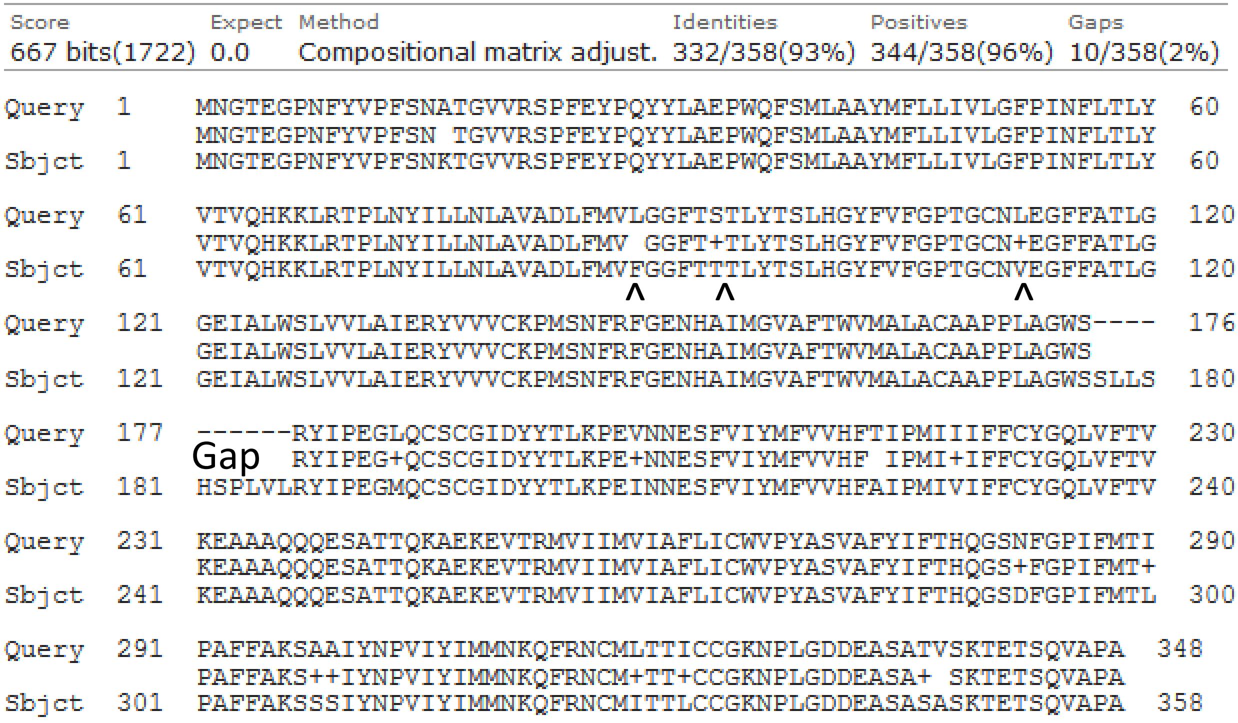
Human vs Dog CDS:
39 differences
(3 in a gap, 36 in the CDS)

20 codon differences are silent.

Rhodopsin (Human)

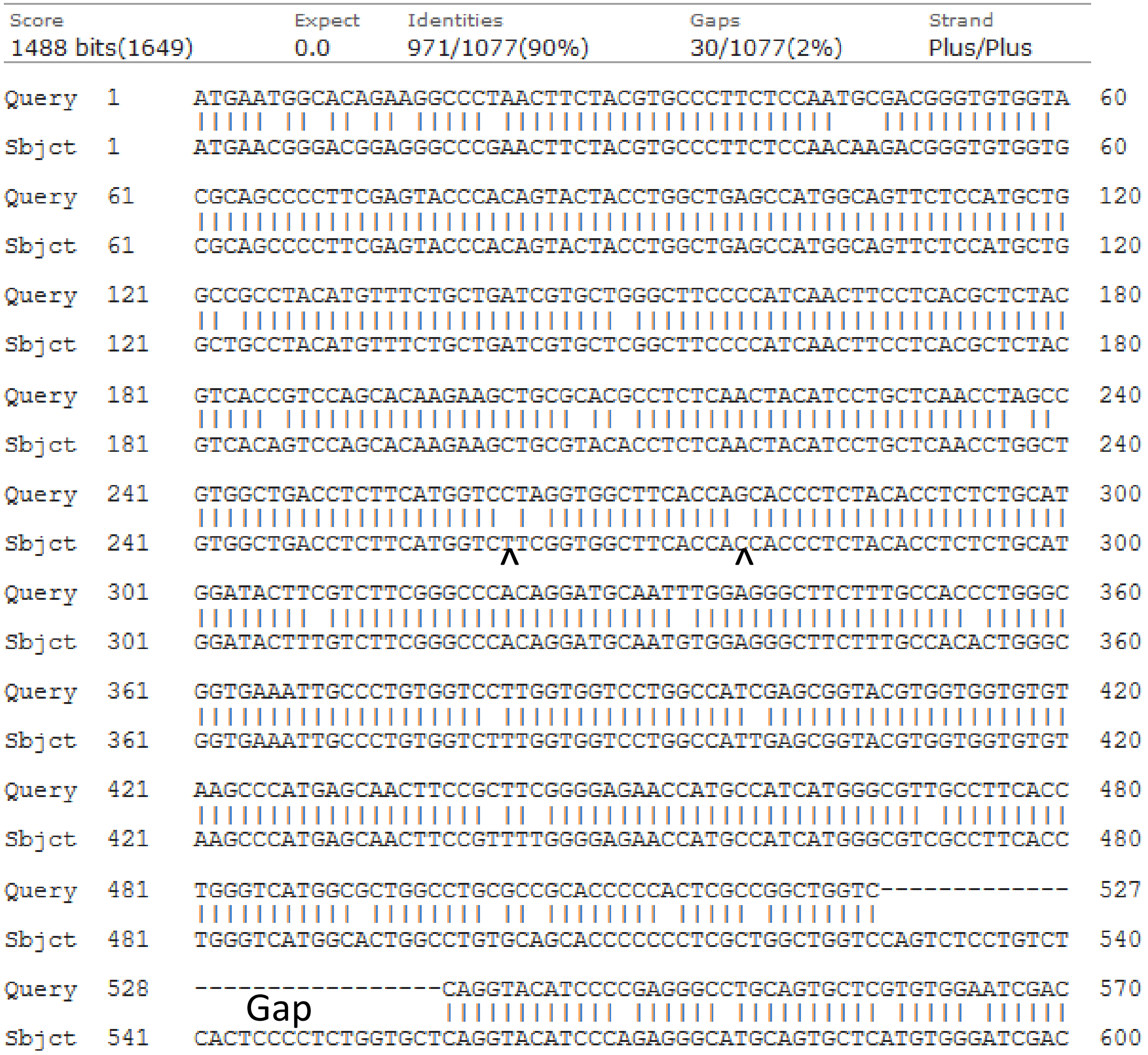
MNGTEGPNFYVPFSNATGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLYVTVQHKKLRTPL
NYILLNLAVADLFMVLGGFTSTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSNF
RFGENHAIMGVAFTWVMALACAAPPLAGWSRYIPEGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMII
IFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQGSNFGPIFMTIPA
FFAKSAAI YNPVIYIMMNKQFRNCMLTTICCGKNPLGDDEASATVSKTETSQVAPA

Human vs dog protein: 26 differences (10 in gap, 16 in CDS)



```
       Score            Expect Method                       Identities      Positives       Gaps
       667 bits(1722)   0.0    Compositional matrix adjust. 332/358(93%)    344/358(96%)    10/358(2%)

Human  Query  1    MNGTEGPNFYVPFSNATGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLY  60
                   MNGTEGPNFYVPFSN TGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLY
  Dog  Sbjct  1    MNGTEGPNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLY  60

       Query  61   VTVQHKKLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGYFVFGPTGCNLEGFFATLG  120
                   VTVQHKKLRTPLNYILLNLAVADLFMV GGFT+TLYTSLHGYFVFGPTGCN+EGFFATLG
       Sbjct  61   VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNVEGFFATLG  120
                                              ^    ^                  ^
       Query  121  GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLAGWS----  176
                   GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLAGWS
       Sbjct  121  GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLAGWSSLLS  180

       Query  177  ------RYIPEGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIIIFFCYGQLVFTV  230
                   Gap   RYIPEG+QCSCGIDYYTLKPE+NNESFVIYMFVVHF IPMI+IFFCYGQLVFTV
       Sbjct  181  HSPLVLRYIPEGMQCSCGIDYYTLKPEINNESFVIYMFVVHFAIPMIVIFFCYGQLVFTV  240

       Query  231  KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQGSNFGPIFMTI  290
                   KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQGS+FGPIFMT+
       Sbjct  241  KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQGSDFGPIFMTL  300

       Query  291  PAFFAKSAAIYNPVIYIMMNKQFRNCMLTTICCGKNPLGDDEASATVSKTETSQVAPA  348
                   PAFFAKS++IYNPVIYIMMNKQFRNCM+TT+CCGKNPLGDDEASA+ SKTETSQVAPA
       Sbjct  301  PAFFAKSSSIYNPVIYIMMNKQFRNCMITTLCCGKNPLGDDEASASASKTETSQVAPA  358
```

Rhodopsin CDS Human vs Dog-1



Human

Dog

```
Score               Expect   Identities          Gaps              Strand
1488 bits(1649)     0.0      971/1077(90%)       30/1077(2%)       Plus/Plus

Query  1      ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTA  60
              ||||| || || || ||||| |||||||||||||||||||||||   ||||||||||||
Sbjct  1      ATGAACGGGACGGAGGGCCCGAACTTCTACGTGCCCTTCTCCAACAAGACGGGTGTGGTG  60

Query  61     CGCAGCCCCTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTG  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61     CGCAGCCCCTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTG  120

Query  121    GCCGCCTACATGTTTCTGCTGATCGTGCTGGGCTTCCCCATCAACTTCCTCACGCTCTAC  180
              || |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Sbjct  121    GCTGCCTACATGTTTCTGCTGATCGTGCTCGGCTTCCCCATCAACTTCCTCACGCTCTAC  180

Query  181    GTCACCGTCCAGCACAAGAAGCTGCGCACGCCTCTCAACTACATCCTGCTCAACCTAGCC  240
              ||||| |||||||||||||||||||| || |||||||||||||||||||||||||| ||
Sbjct  181    GTCACAGTCCAGCACAAGAAGCTGCGTACACCTCTCAACTACATCCTGCTCAACCTGGCT  240

Query  241    GTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACCAGCACCCTCTACACCTCTCTGCAT  300
              ||||||||||||||||||||| | ||||||||||||| ||||||||||||||||||||||
Sbjct  241    GTGGCTGACCTCTTCATGGTCTTCGGTGGCTTCACCACCACCCTCTACACCTCTCTGCAT  300
                                   ^               ^
Query  301    GGATACTTCGTCTTCGGGCCCACAGGATGCAATTTGGAGGGCTTCTTTGCCACCCTGGGC  360
              |||||||| |||||||||||||||||||||||| ||||||||||||||||||| ||||||
Sbjct  301    GGATACTTTGTCTTCGGGCCCACAGGATGCAATGTGGAGGGCTTCTTTGCCACACTGGGC  360

Query  361    GGTGAAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGT  420
              |||||||||||||||||||| ||||||||||||||||| |||||||||||||||||||||
Sbjct  361    GGTGAAATTGCCCTGTGGTCTTTGGTGGTCCTGGCCATTGAGCGGTACGTGGTGGTGTGT  420

Query  421    AAGCCCATGAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACC  480
              |||||||||||||||||||| || |||||||||||||||||||||||||| |||||||||
Sbjct  421    AAGCCCATGAGCAACTTCCGTTTTGGGGAGAACCATGCCATCATGGGCGTCGCCTTCACC  480

Query  481    TGGGTCATGGCGCTGGCCTGCGCCGCACCCCCACTCGCCGGCTGGTC-------------  527
              ||||||||||| |||||||| || |||||||| ||||| ||||||||
Sbjct  481    TGGGTCATGGCACTGGCCTGTGCAGCACCCCCCCTCGCTGGCTGGTCCAGTCTCCTGTCT  540

Query  528    -----------------CAGGTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAATCGAC  570
                  Gap         |||||||||||| |||||| |||||||||| ||||| ||||||
Sbjct  541    CACTCCCCTCTGGTGCTCAGGTACATCCCAGAGGGCATGCAGTGCTCATGTGGGATCGAC  600
```

106 differences (30 in gap, 76 in CDS)

60 codon differences are silent

Rhodopsin CDS Human vs Dog-2

Human
Dog

```
Query  571   TACTACACGCTCAAGCCGGAGGTCAACAACGAGTCTTTTGTCATCTACATGTTCGTGGTC  630
             |||||||| |||||||| ||  ||||||| ||||| || |||||||||||||||||||||
Sbjct  601   TACTACACACTCAAGCCAGAAATCAACAATGAGTCCTTCGTCATCTACATGTTCGTGGTC  660

Query  631   CACTTCACCATCCCCATGATTATCATCTTTTTCTGCTATGGGCAGCTCGTCTTCACCGTC  690
             |||||| |||||||||||||| |||| || ||||||||||| |||||||||||||| |||
Sbjct  661   CACTTCGCCATCCCCATGATTGTCATATTCTTCTGCTATGGACAGCTCGTCTTCACAGTC  720

Query  691   AAGGAGGCCGCTGCCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTC  750
             |||||||| ||||||||||||||||| || |||||||| |||||||| || |||||||||
Sbjct  721   AAGGAGGCAGCTGCCCAGCAGCAGGAATCGGCCACCACCCAGAAGGCTGAAAAGGAGGTC  780

Query  751   ACCCGCATGGTCATCATCATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTACGCCAGC  810
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| 
Sbjct  781   ACCCGCATGGTCATCATCATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTATGCCAGT  840

Query  811   GTGGCATTCTACATCTTCACCCACCAGGGCTCCAACTTCGGTCCCATCTTCATGACCATC  870
             ||||||||||||||||||||||||||||||||| |||| || ||||||||||||||| ||
Sbjct  841   GTGGCATTCTACATCTTCACCCACCAGGGCTCCGACTTTGGCCCCATCTTCATGACCCTC  900

Query  871   CCAGCGTTCTTTGCCAAGAGCGCCGCCATCTACAACCCTGTCATCTATATCATGATGAAC  930
             ||||||||||| ||||||  | || |||||||||||||||||||||||||||||||||||
Sbjct  901   CCAGCGTTCTTCGCCAAGTCCTCCTCCATCTACAACCCTGTCATCTATATCATGATGAAC  960

Query  931   AAGCAGTTCCGGAACTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGAC  990
             ||||||||||||||||||||| |||||||| ||||||| |||||||||||||||||||||
Sbjct  961   AAGCAGTTCCGGAACTGCATGATCACCACCCTCTGCTGTGGCAAGAACCCACTGGGTGAC  1020

Query  991   GATGAGGCCTCTGCTACCGTGTCCAAGACGGAGACGAGCCAGGTGGCCCCGGCCTAA  1047
             || ||||||||||| | ||  |||||||||||||| ||||||||||| |||||||||
Sbjct  1021  GACGAGGCCTCTGCCAGCGCCTCCAAGACGGAGACCAGCCAGGTGGCACCGGCCTAA  1077
```

Partial translation of Rhodopsin

Human	571	TAC	TAC	**ACG**	CTC	AAG	**CCG**	**GAG**	**GTC**	AAC	**AAC**	GAG	**TCT**	**TTT**	GTC	ATC	TAC	ATG	TTC	GTG	GTC
		Y	Y	**T**	L	K	**P**	**E**	V	N	**N**	E	**S**	**F**	V	I	Y	M	F	V	V
Dog	601	TAC	TAC	**ACA**	CTC	AAG	**CCA**	**GAA**	**ATC**	AAC	**AAT**	GAG	**TCC**	**TTC**	GTC	ATC	TAC	ATG	TTC	GTG	GTC
									M												

631	CAC	TTC	**ACC**	ATC	CCC	ATG	ATT	**ATC**	**ATC**	**TTT**	TTC	TGC	TAT	**GGG**	CAG	CTC	GTC	TTC	**ACC**	GTC
	H	F	T	I	P	M	I	I	**I**	**F**	F	C	Y	**G**	Q	L	V	F	**T**	V
661	CAC	TTC	**GCC**	ATC	CCC	ATG	ATT	**GTC**	**ATA**	**TTC**	TTC	TGC	TAT	**GGA**	CAG	CTC	GTC	TTC	**ACA**	GTC
			R					V												

691	AAG	GAG	**GCC**	GCT	GCC	CAG	CAG	CAG	**GAG**	**TCA**	GCC	ACC	**ACA**	CAG	AAG	**GCA**	**GAG**	AAG	GAG	GTC
	K	E	**A**	A	A	Q	Q	Q	**E**	**S**	A	T	**T**	Q	K	**A**	**E**	K	E	V
721	AAG	GAG	**GCA**	GCT	GCC	CAG	CAG	CAG	**GAA**	**TCG**	GCC	ACC	**ACC**	CAG	AAG	**GCT**	**GAA**	AAG	GAG	GTC

751	ACC	CGC	ATG	GTC	ATC	ATC	ATG	GTC	ATC	GCT	TTC	CTG	ATC	TGC	TGG	GTG	CCC	**TAC**	GCC	**AGC**
	T	R	M	V	I	I	M	V	I	A	F	L	I	C	W	V	P	**Y**	A	**S**
781	ACC	CGC	ATG	GTC	ATC	ATC	ATG	GTC	ATC	GCT	TTC	CTG	ATC	TGC	TGG	GTG	CCC	**TAT**	GCC	**AGT**

811	GTG	GCA	TTC	TAC	ATC	TTC	ACC	CAC	CAG	GGC	TCC	**AAC**	**TTC**	**GGT**	CCC	ATC	TTC	ATG	ACC	**ATC**
	V	A	F	Y	I	F	T	H	Q	G	S	N	**F**	**G**	P	I	F	M	T	I
841	GTG	GCA	TTC	TAC	ATC	TTC	ACC	CAC	CAG	GGC	TCC	**GAC**	**TTT**	**GGC**	CCC	ATC	TTC	ATG	ACC	**CTC**
												D								L

In these five lines of sequence, there are 25 mutations (bold) with 5 mutations resulting in substitutions (in red) and 20 that are silent.

Summary of Differences

Protein	Protein Size (Human aa)	CDS Differences	Protein Differences
Beta globin	146	53	15
Amelogenin	191	36	16
Rhodopsin	348	76	16

These are all relatively small proteins. How about a bigger one?

Summary of Differences

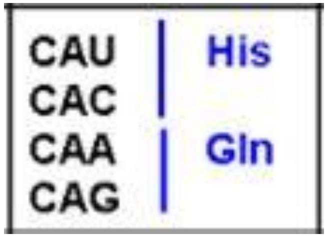
Protein	Protein Size (Human aa)	CDS Differences	Protein Differences
Beta globin	146	53	15 (28%)
Amelogenin	191	36	16 (44%)
Rhodopsin	348	76	16 (21%)
CFTR	1480	468	147 (31%)

Why do silent mutations prevail?

- Silent mutations are more common than chance predicts because of the structure of the code and the chemistry of mutation.
- Silent mutations are more common because if you get side-by-side mutations in the same codon, it will only be expressed once. Not all unexpressed mutations are silent.
- Silent mutations are more common because there is natural selection in favor of silence. If you don’t mess with the protein you are more likely to survive.

Why do silent mutations prevail?

- Silent mutations are more common than chance predicts because of the structure of the code and the chemistry of mutation.



CAU	His
CAC	
CAA	Gln
CAG	

- There are two types of base substitutions:
    - Transitions (purine-purine or pyrmidine-pyrimidine)
    - Transversions (purine <-> pyrimidine).

Because of the chemistry of mutation, transitions are more common than transversions.

- Transitions in the third position are likely to be silent because of the structure of the code.

Transitions and Transversions in the Rhodopsin Sequence

H	571	TAC	TAC	**ACG**	CTC	AAG	**CCG**	**GAG**	**GTC**	AAC	**AAC**	GAG	**TCT**	**TTT**	GTC	ATC	TAC	ATG	TTC	GTG	GTC
		Y	Y	**T**	L	K	**P**	**E**	V	N	**N**	E	**S**	**F**	V	I	Y	M	F	V	V
D	601	TAC	TAC	**ACA**	CTC	AAG	**CCA**	**GAA**	**ATC**	AAC	**AAT**	GAG	**TCC**	**TTC**	GTC	ATC	TAC	ATG	TTC	GTG	GTC
									M												
										TV										TV	
	631	CAC	TTC	**ACC**	ATC	CCC	ATG	ATT	**ATC**	**ATC**	**TTT**	TTC	TGC	TAT	**GGG**	CAG	CTC	GTC	TTC	**ACC**	GTC
		H	F	T	I	P	M	I	I	**I**	**F**	F	C	Y	**G**	Q	L	V	F	**T**	V
	661	CAC	TTC	**GCC**	ATC	CCC	ATG	ATT	**GTC**	**ATA**	**TTC**	TTC	TGC	TAT	**GGA**	CAG	CTC	GTC	TTC	**ACA**	GTC
				R					V												
				TV										TV			TV				
	691	AAG	GAG	**GCC**	GCT	GCC	CAG	CAG	CAG	**GAG**	**TCA**	GCC	ACC	**ACA**	CAG	AAG	**GCA**	**GAG**	AAG	GAG	GTC
		K	E	**A**	A	A	Q	Q	Q	**E**	**S**	A	T	**T**	Q	K	**A**	**E**	K	E	V
	721	AAG	GAG	**GCA**	GCT	GCC	CAG	CAG	CAG	**GAA**	**TCG**	GCC	ACC	**ACC**	CAG	AAG	**GCT**	**GAA**	AAG	GAG	GTC
	751	ACC	CGC	ATG	GTC	ATC	ATC	ATG	GTC	ATC	GCT	TTC	CTG	ATC	TGC	TGG	GTG	CCC	**TAC**	GCC	**AGC**
		T	R	M	V	I	I	M	V	I	A	F	L	I	C	W	V	P	**Y**	A	**S**
	781	ACC	CGC	ATG	GTC	ATC	ATC	ATG	GTC	ATC	GCT	TTC	CTG	ATC	TGC	TGG	GTG	CCC	**TAT**	GCC	**AGT**
																					*TV
	811	GTG	GCA	TTC	TAC	ATC	TTC	ACC	CAC	CAG	GGC	TCC	**AAC**	**TTC**	**GGT**	CCC	ATC	TTC	ATG	ACC	**ATC**
		V	A	F	Y	I	F	T	H	Q	G	S	N	**F**	**G**	P	I	F	M	T	I
	841	GTG	GCA	TTC	TAC	ATC	TTC	ACC	CAC	CAG	GGC	TCC	**GAC**	**TTT**	**GGC**	CCC	ATC	TTC	ATG	ACC	**CTC**
													D								L

Of the 25 mutations, 6 are transversions (TV), but only one (*) was associated with a substitution. The other 4 changes were all due to B1 transitions. So there is no bias toward transversions producing amino acid substitutions.

Why do silent mutations prevail?

- Clustered mutations will be underexpressed if they fall into the same codon.
- In the set of proteins we’ve looked at, the beta globin coding sequence has the most clustering.
- How does the clustering contribute to amino acid changes in the protein?
- How does the clustering contribute to unexpressed base changes?

Mutational Clusters in Human HBB CDS

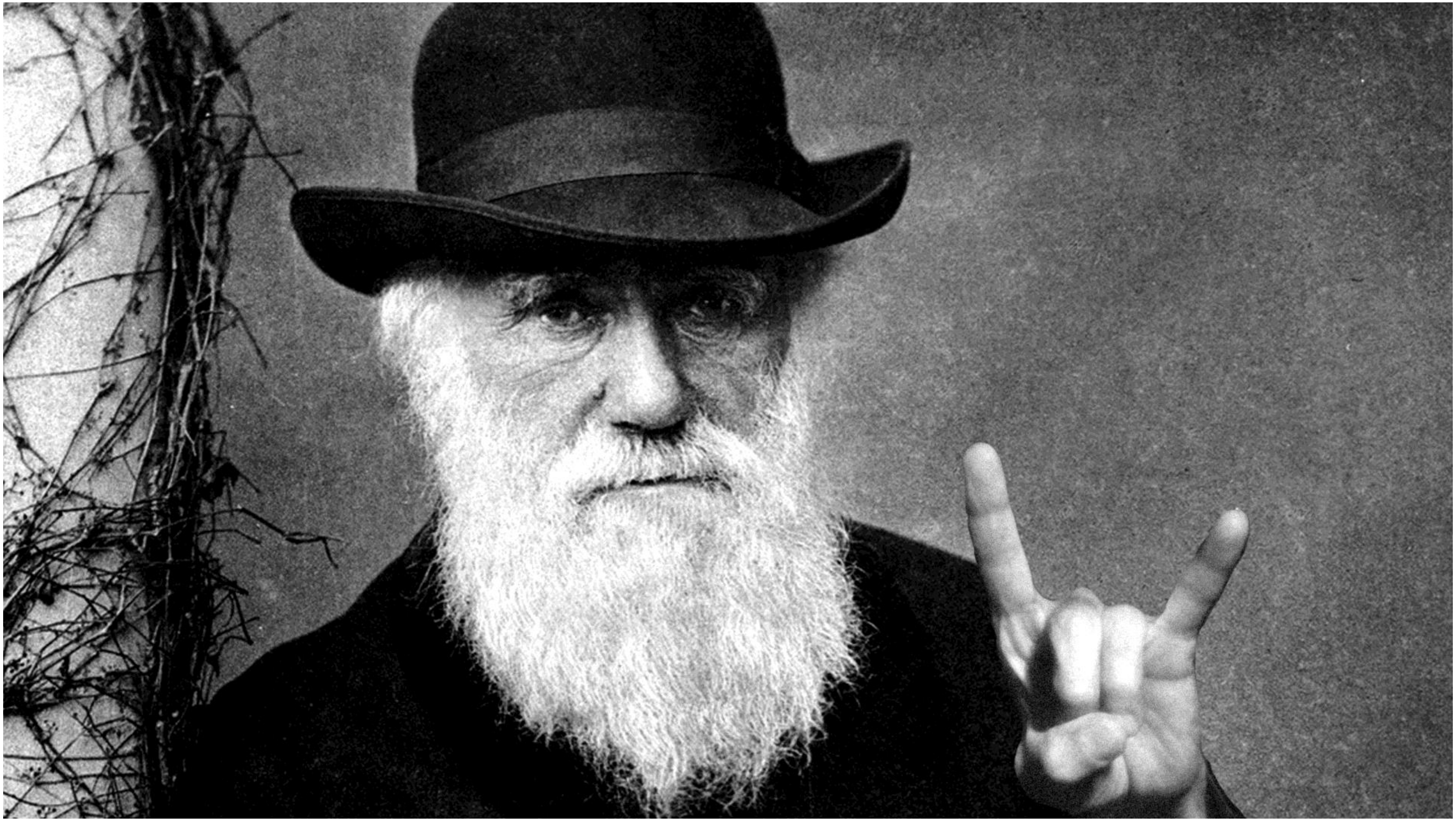
ATG GTG CAT CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT
GCC CTG TGG GGC AAG GTG AAC GTG GAT GAA GTT GGT GGT
GAG GCC CTG GGC AGG CTG CTG GTG GTC TAC CCT TGG ACC
CAG AGG TTC TTT GAG TCC TTT GGG GAT CTG TCC ACT CCT
GAT GCT GTT ATG GGC AAC CCT AAG GTG AAG GCT CAT GGC
AAG AAA GTG CTC GGT GCC TTT AGT GAT GGC CTG GCT CAC
CTG GAC AAC CTC AAG GGC ACC TTT GCC ACA CTG AGT GAG
CTG CAC TGT GAC AAG CTG CAC GTG GAT CCT GAG AAC TTC
AGG CTC CTG GGC AAC GTG CTG GTC TGT GTG CTG GCC CAT
CAC TTT GGC AAA GAA TTC ACC CCA CCA GTG CAG GCT GCC
TAT CAG AAA GTG GTG GCT GGT GTG GCT AAT GCC CTG GCC
CAC AAG TAT CAC TAA

- 53 nucleotide substitutions, with 42 codons changed. 8 codons contain 2-3 substitutions.
- 15 amino acid changes, and 27 silent codon changes. A cluster can hide a nucleotide change, but ALL of the multichange codons will change the amino acid. Of the 34 single nucleotide codon changes, 27 (79%) are silent.

Why do silent mutations prevail?

- Assuming random mutation, about 1/3 of mutations should be silent, mostly those that change only the 3rd nucleotide of a codon.
- In the beta globin gene, there are 10 first position mutations, 8 second position mutations, and 34 third position mutations.
- All of the 8 codons with more than one mutation change their amino acid.
- Of the 34 codons with a single nucleotide change, only 7 change their amino acid, and the remaining 27 are silent.
- Silent mutations appear to be strongly selected, or most mutations that DO change the amino acid will kill the bearer.

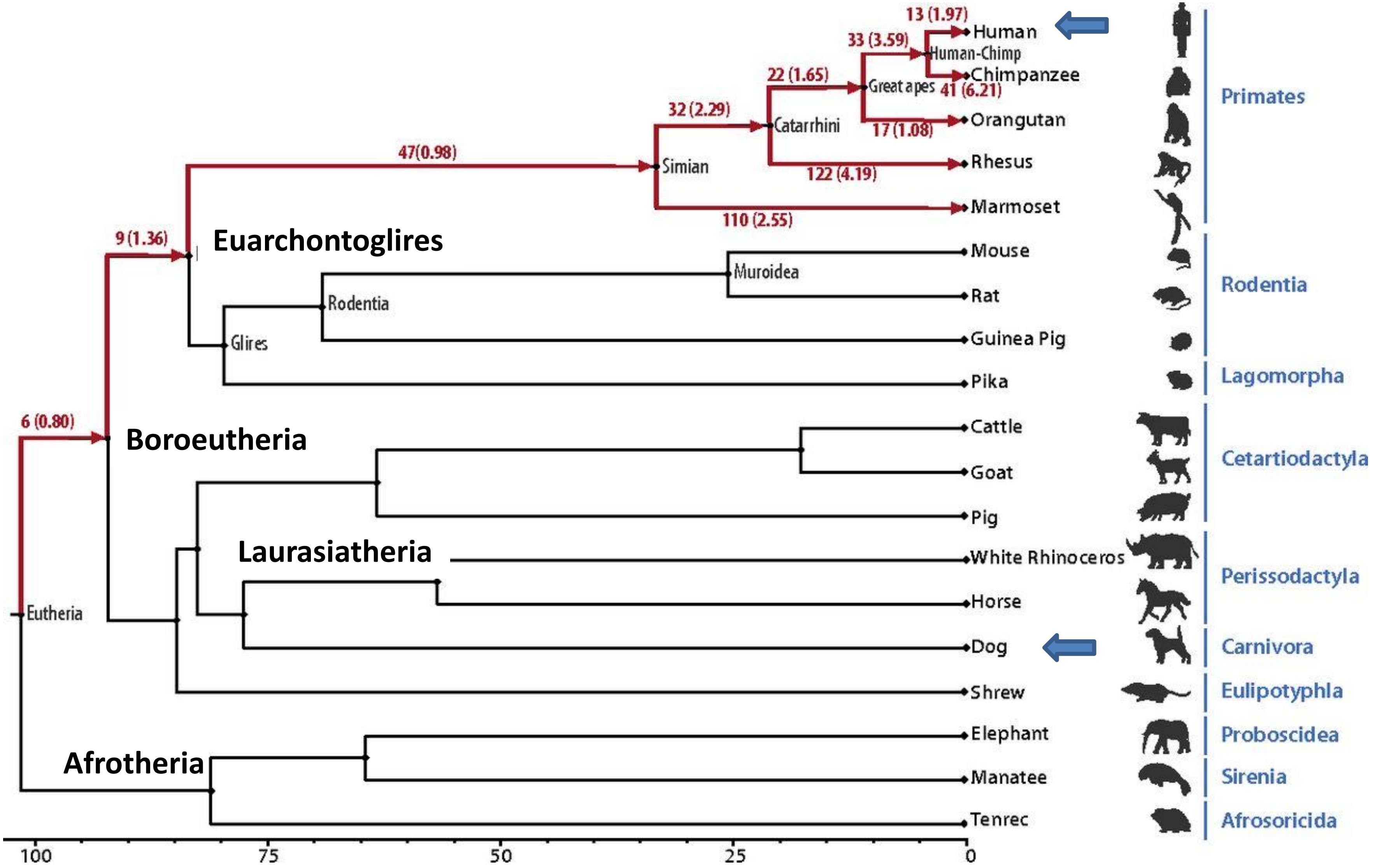
Darwin Wins Again



So it’s Probably Selection

- We’ve looked only at 4 proteins in two species: humans and dogs.
- Humans and dogs are in two different branches of the mammalian phylogenetic tree: humans in the Euarchontoglires and dogs in the Laurasiatheria.

Mammalian phylogeny



Kim et al. 2017. https://doi.org/10.1073/pnas.1702012114

So it’s Probably Selection

- We’ve looked only at 4 proteins in two species: humans and dogs.
- Humans and dogs are in two different branches of the mammalian phylogenetic tree: humans in the Euarchontoglires and dogs in the Laurasiatheria.
- These two lineages separated over 90 MYA, so we’re looking at the mutational accumulation over that time span.
- The silent mutations are much favored over the substitution mutations.
- Some proteins are less tolerant of change than others.

Question

Are silent mutations totally silent?

- They don’t affect the amino acid encoded by the protein.
- But do they affect the production of the protein?
- Translation involves interaction between the tRNA anticodon and the mRNA codon on the ribosome.
- Transfer RNAs have to match two different binding sites: the tRNA binding sites on the ribosome and the tRNA binding sites on the aminoacyl tRNA synthetases that put the amino acid on the tRNA.

Some Codons are Preferred over Others

Example: Human codon bias

UUU	F	0.46	UCU	S	0.19	UAU	Y	0.44	UGU	C	0.46
UUC	F	0.54	UCC	S	0.22	UAC	Y	0.56	UGC	C	0.54
UUA	L	0.08	UCA	S	0.15	UAA	*	0.30	UGA	*	0.47
UUG	L	0.13	UCG	S	0.05	UAG	*	0.24	UGG	W	1.00
CUU	L	0.13	CCU	P	0.29	CAU	H	0.42	CGU	R	0.08
CUC	L	0.20	CCC	P	0.32	CAC	H	0.58	CGC	R	0.18
CUA	L	0.07	CCA	P	0.28	CAA	Q	0.27	CGA	R	0.11
CUG	L	0.40	CCG	P	0.11	CAG	Q	0.73	CGG	R	0.20
AUU	I	0.36	ACU	T	0.25	AAU	N	0.47	AGU	S	0.15
AUC	I	0.47	ACC	T	0.36	AAC	N	0.53	AGC	S	0.24
AUA	I	0.17	ACA	T	0.28	AAA	K	0.43	AGA	R	0.21
AUG	M	1.00	ACG	T	0.11	AAG	K	0.57	AGG	R	0.21
GUU	V	0.18	GCU	A	0.27	GAU	D	0.46	GGU	G	0.16
GUC	V	0.24	GCC	A	0.40	GAC	D	0.54	GGC	G	0.34
GUA	V	0.12	GCA	A	0.23	GAA	E	0.42	GGA	G	0.25
GUG	V	0.46	GCG	A	0.11	GAG	E	0.58	GGG	G	0.25

https://www.kazusa.or.jp/codon/

Switching codons may affect translation or protein folding.

- The preferred codon may have more or more stable tRNAs. Switching from a preferred codon may make the codon-anticodon interaction a little less stable or slower than usual.
- …or vice versa.
- Proteins fold during translation, so speeding up or slowing down translation might interfere with normal protein folding.

References

- Kim, et al. 2017. Reconstruction and evolutionary history of eutherian chromosomes. https://doi.org/10.1073/pnas.1702012114
- Pan, Tau. 2018. Modifications and functional genomics of human transfer RNA. *Cell Research* 28: 395–404. https://doi.org/10.1038/s41422-018-0013-y
- OMIM: https://www.ncbi.nlm.nih.gov/omim
- Codon usage: https://www.kazusa.or.jp/codon/
- BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi