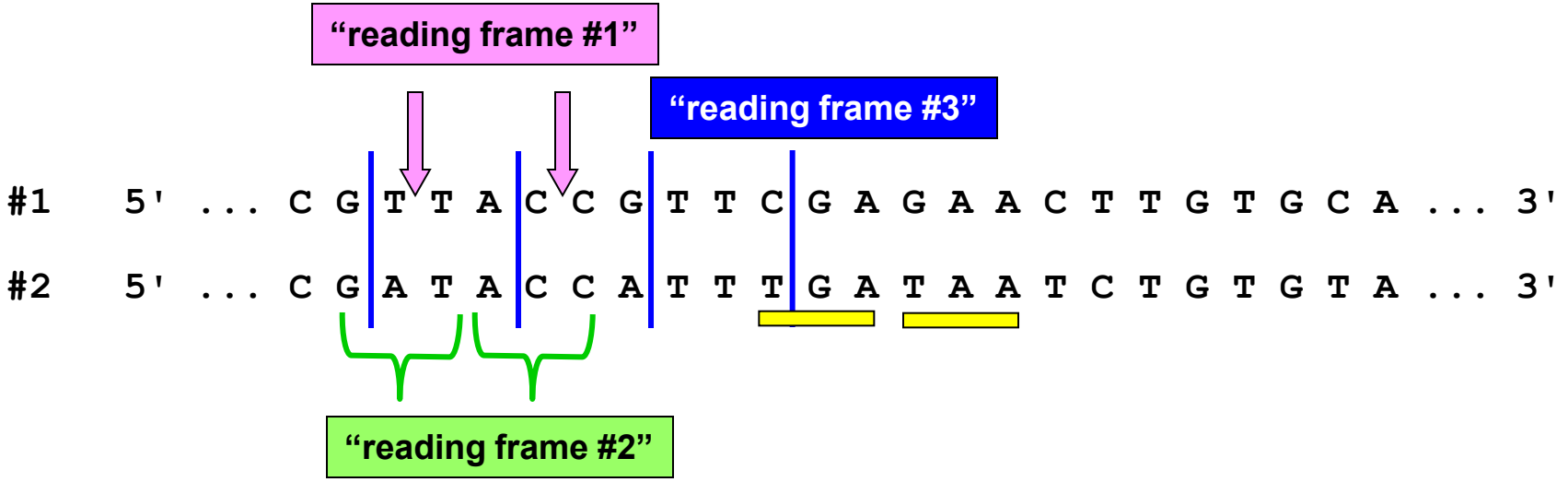
Practice question #1



If protein-coding sequence under strong functional constraint…

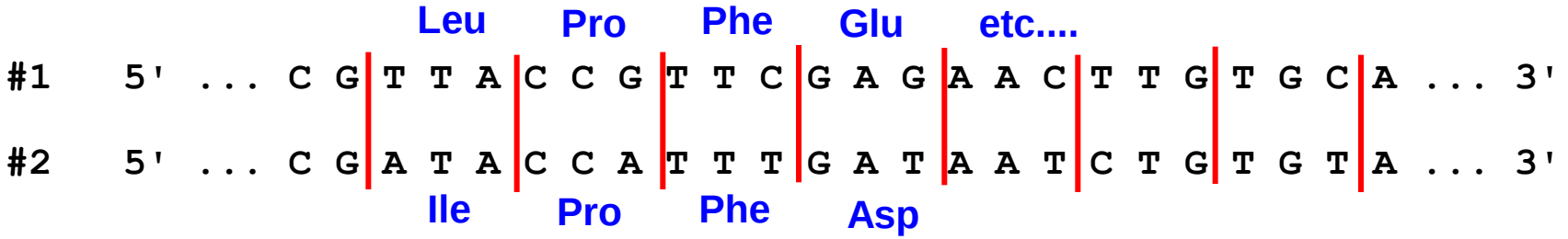
might expect (mostly) silent nt substitutions between #1 and #2…

and conservative amino acid substitutions…

Could look for an initiation codon, but it might be further upstream….

So divide into possible reading frames…

Any potential stop codons (to exclude any frames)??



Practice question #3

TABLE 1.5 Relative frequencies of different types of mutational substitutions in a random protein-coding sequence

Substitution	Number	Percent
Total in all codons	549	100
Synonymous	134	25
Nonsynonymous	415	75

Protein Y coding sequence = 200 x 3 = 600 nt

 # non-syn sites = 600 x 75% = 450 **(approximately)**

What if protein Y is rich in methionine?

then expect higher # non-syn sites...

because any change to AUG triplet will alter amino acid encoded

What type of protein would have a relatively low number of non-synonymous sites?

If high number of codons from 6-member families

Note however, that number of non-synonymous changes will depend on functional constraint on protein...

"changes" ⇨ comparing sequences from 2 organisms...

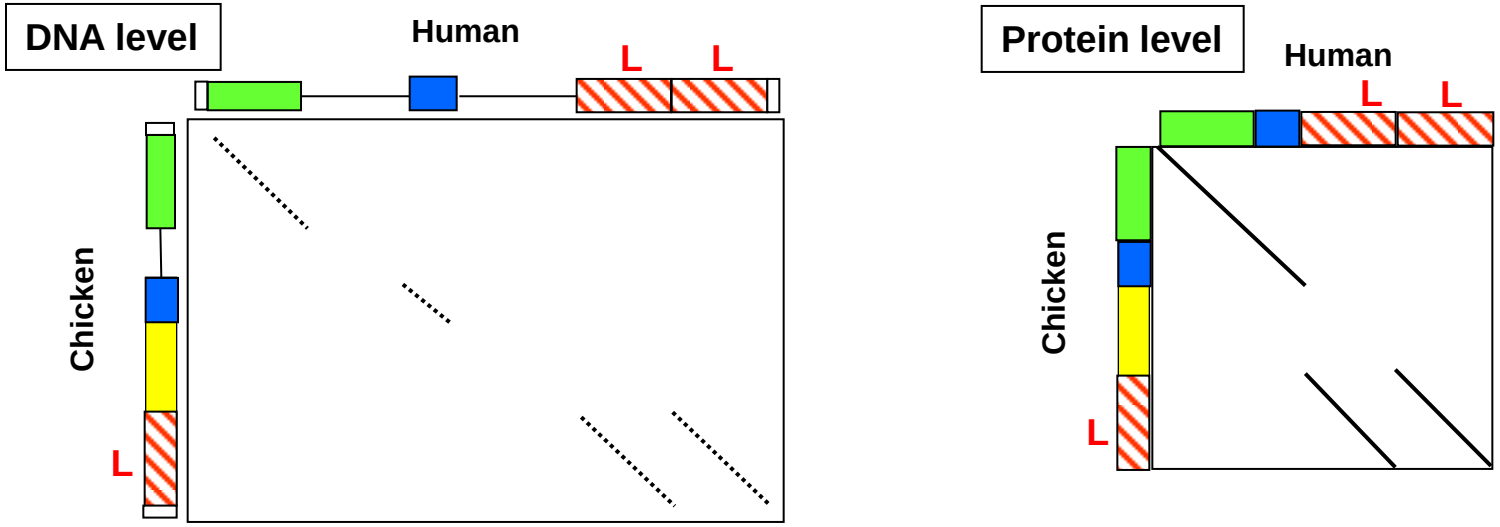
Second letter

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

Practice question #4

Approximately 54% amino acid identity

Practice question #5

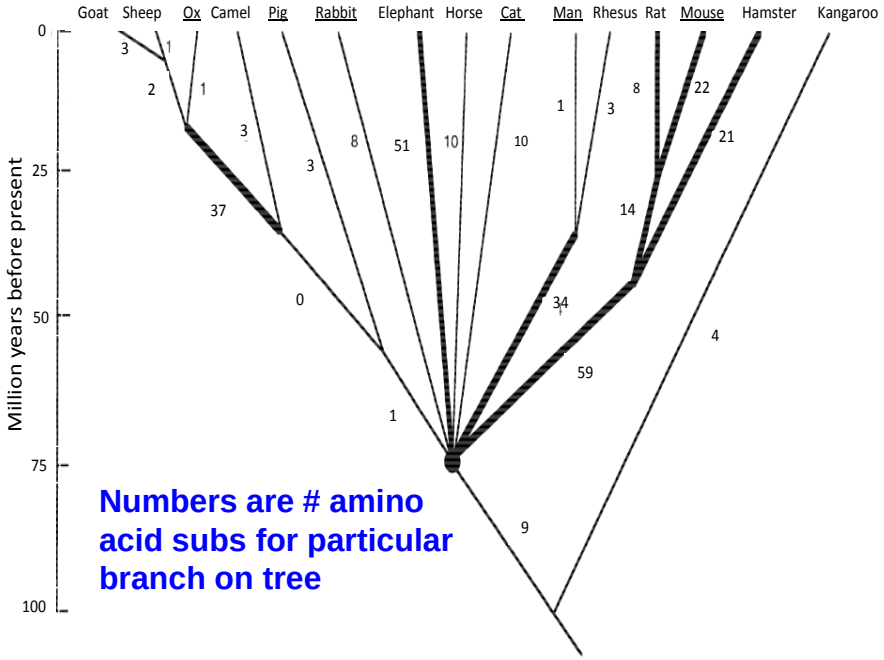


Assumption: human exon 2 closely related to 5' end of (longer) chicken exon 2. What if this was not the case?

Note: white boxes represent UTR regions & coloured blocks are coding regions

Practice question #6

	#1 k_A	#2 k_S
Human	0.111	1.33
Ox	0.131	1.01
Pig	0.040	1.28
Cat	0.052	1.60
Rabbit	0.050	1.21
Mouse	0.261	1.58



- pairwise comparisons of nt sequences with outgroup(s), score differences
- alignment, gap penalty, determine # syn vs. non-syn sites, normalize for length…

Molecular clock hypothesis

… and punctuated (or episodic) evolution

- know that rodents have high sub rate (for syn & non-syn sites) - high metabolic rate…

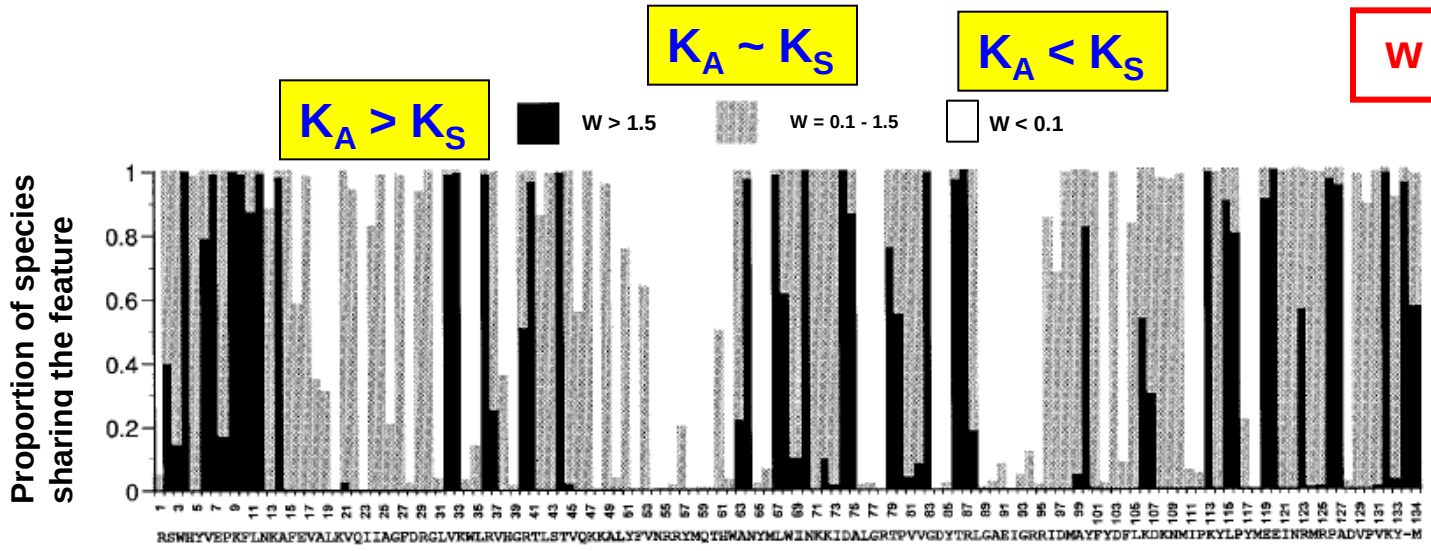
But unequal # aa subs for prolactin since time that rat & mouse shared common ancestor

But elephants have high number of aa subs too (large animal, not high metabolic rate…)

- reduced functional constraint ? (“relaxation of purifying selection”)

or adaptive evolution?

Practice question #7 (Topic 4, p.123 in text)



$w \equiv K_A / K_S$

Mol Biol Evol 17:1446, 2000

Comparison of amino acid sequences of protein E (actually lysin) from 25 abalone species

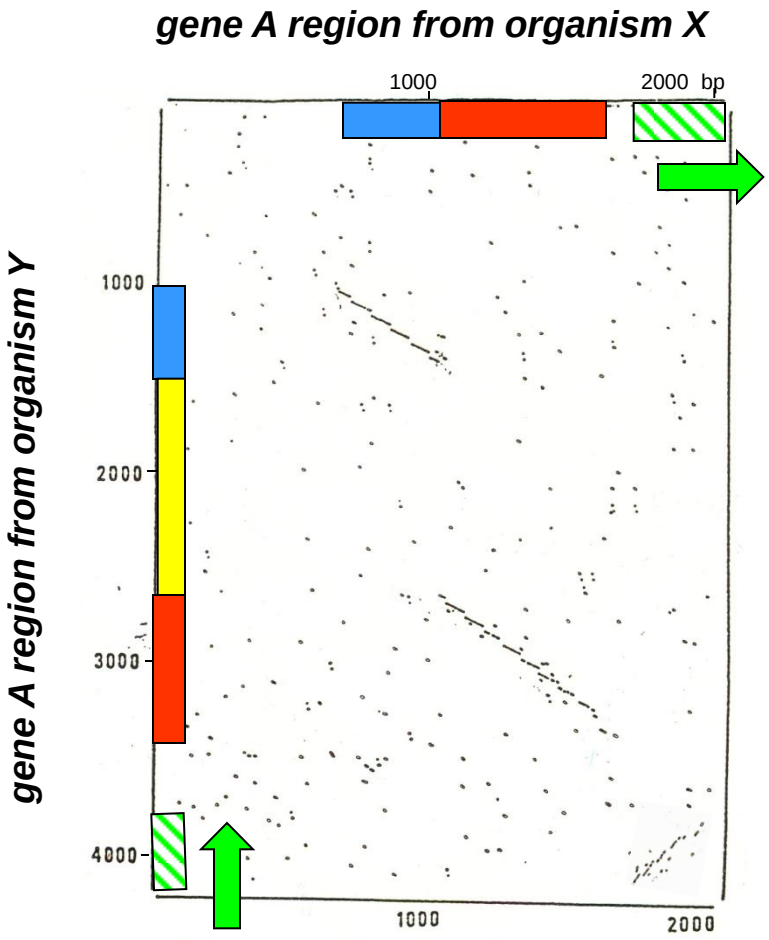
Diversifying selection (adaptive evolution) vs. purifying selection at various sites?

- N-terminal region shows cluster of sites undergoing rapid change (adaptive evolution)

vs. several other regions (eg. amino acid positions 54-60) which show strong functional constraint (important for structural or enzymatic activity of protein?)

> Many amino acid sites are clearly under strong purifying selection, with the nonsynonymous rate close to 0, while some other sites are under diversifying selection, with the nonsynonymous rate elevated to more than three times the synonymous rate.

Practice question #8

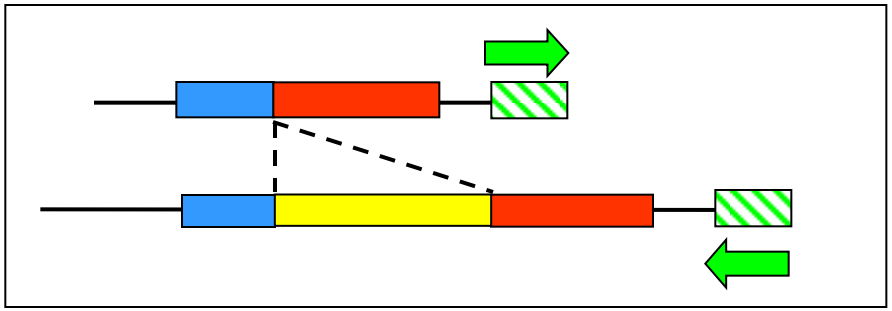


What might yellow region be?

additional protein domain...?

or intron ...?

Length of yellow region ~ 1kb ...



Inversion in downstream flanking region (not within gene)