Question 2a & 2b

Ala

GCN

5'.. C | G A C | A G G | G A G | C C G | C T A | A G C | T G T | G A C | G A G | G C g t a c t t a g c g g t c c a c t c g c t c t c t a c g t g .. 3'

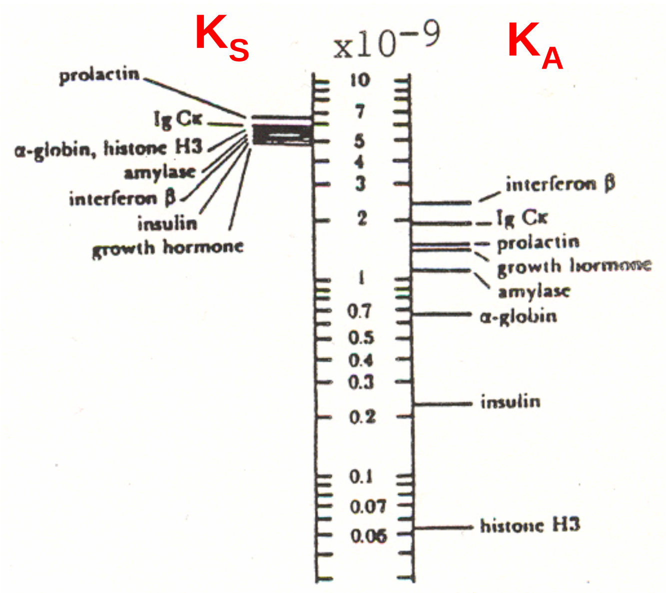
Asp Arg Glu Pro Leu Ser Cys Asp Glu

☆ ☆ ☆ ☆

% nt identity?

$r = k / 2T$

Rates



Total # non-syn subs and # syn subs gives % nt divergence

Solve for k_A (# non-syn subs/non-syn site)....

and for k_S (# syn subs/syn site)

Then multiply by number of sites...

eg. 660 x 0.25 = 165 syn sites

Total # subs expected = 129 syn +89 non-syn

69% nt identity

Aside: Figure is similar to question in Topic 4, convert Table 4.1 info to log format

Question 2c

Asp Arg Glu Pro Leu Ser Cys Asp Glu Ala

5'.. CGACAGGGAGCCGCTAAGCTGTGACGAGGC gtacttagcggtccactcgctctctacgtg .. 3'

30 nt exon **30 nt intron**

What would you expect to see after 100 million years?

(compared to 120 million years of evolutionary time in part 2b)

For coding:

 # synonymous subs > # non-syn subs

Glu and Asp residues conserved (since “important acidic domain”)

Cys conserved? Leu-to-Ile change?

For intron:

 # intron changes > # coding sequence changes

rate ~ same as synonymous rate

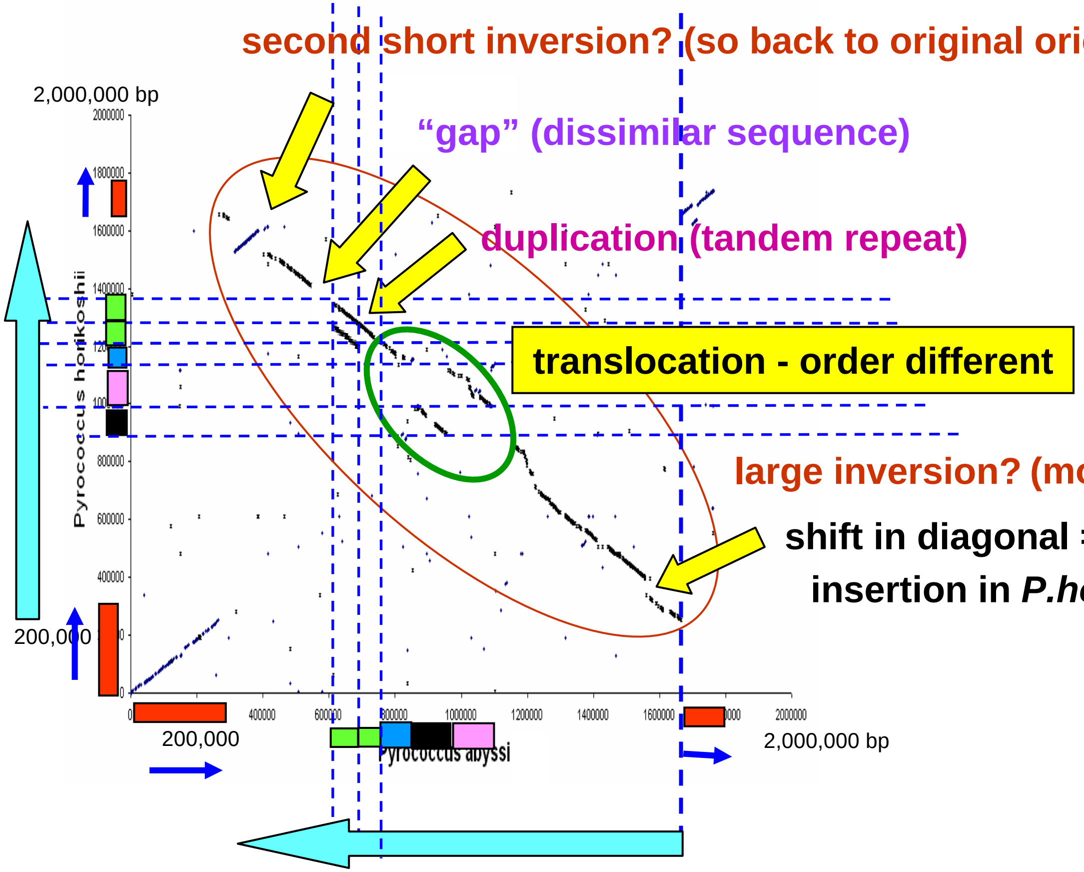
expect indels as well as nt subs

For all:

Transitions > transversions

N.B. Cannot use terms like “synonymous” or “non-synonymous” when talking about non-coding sequences (introns, UTRs, non-genic....)

Question 3



second short inversion? (so back to original orientation)

“gap” (dissimilar sequence)

duplication (tandem repeat)

translocation - order different

large inversion? (more parsimonious than many short ones)

shift in diagonal = indel

insertion in *P.horikpshii* or deletion in *P.abyssi*

Researchers interested in whether genome rearrangements might be driven by extreme lifestyle (hyperthermophilic: hot springs, thermal vents...)

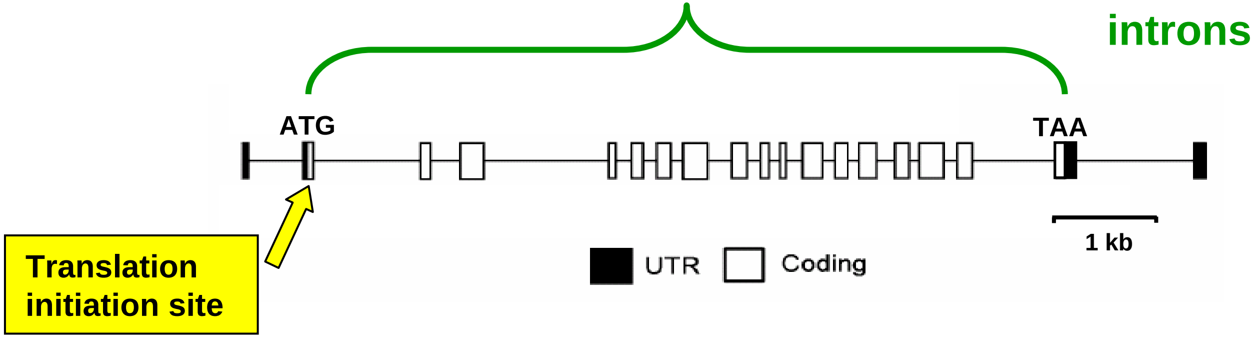
Joseph & Sasikumar *BMC Bioinformatics* 7:243, 2006

Question 4

SPP120 gene

gene ~ 7.1 kb long

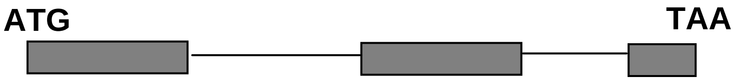
introns ~ 68% of gene length



Translation initiation site

Definitions:

gene = coding region + introns within it **(in eukaryotes, because virtually all bacterial genes lack introns)**

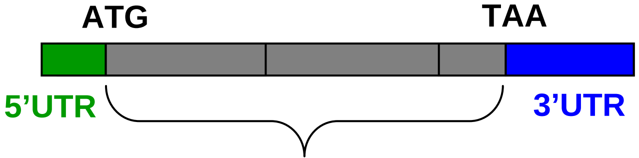


Gene X

Does not depend on knowing anything about regulatory sequences or ends of the mRNA....

“Gene X is 5 kb long and it has two introns ”

mRNA = exons (comprised of coding sequence plus UTRs)



coding sequence

mRNA X

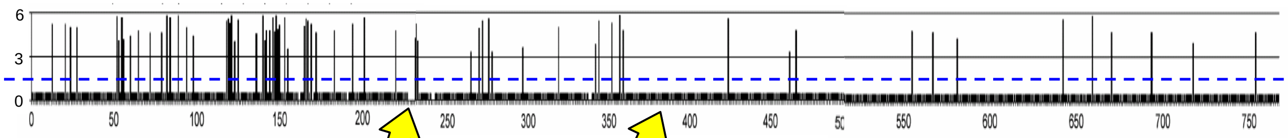
“mRNA X has 3 exons, with exon 1 & exon 3 containing the 3’UTR & 5’UTR, respectively”

Question 4

Adaptive evolution at these sites?

B

K_A / K_S



$K_A = K_S$

If “no constraint” (blue dotted line)

$K_A / K_S = 0$

Invariant sites? (eg. Cys)

Strong functional constraint

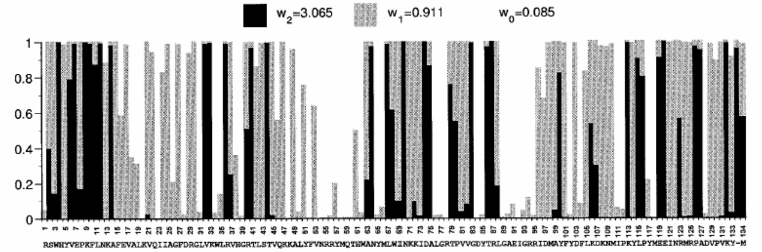
$K_A < K_S$

Most codons have K_A / K_S values less than 1

but certain regions (eg. N-terminal) have high non-syn rate

- positive Darwinian selection?

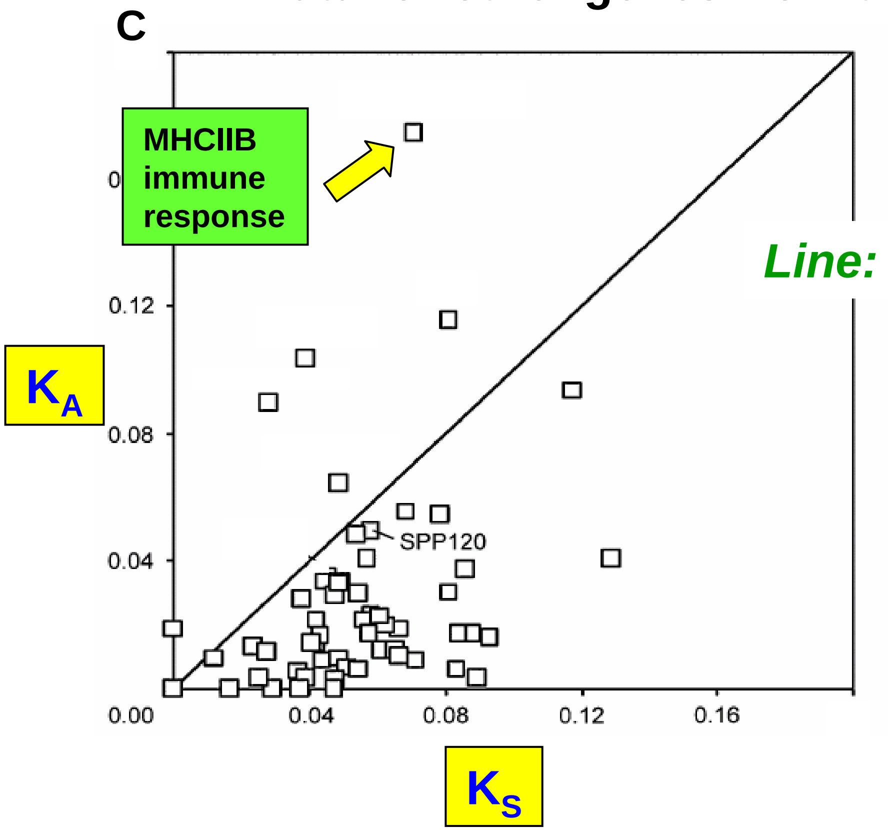
selective pressure changing a few key sites may lead to major change in phenotype (eg. lysozyme in ruminants...)



Aside: Figure B is similar to Practice question #6

Question 4

Data for other genes from these fish...



- shows overall syn & non-syn rates for each gene (vs. along length of one gene in part b)

Line: $K_S = K_A$

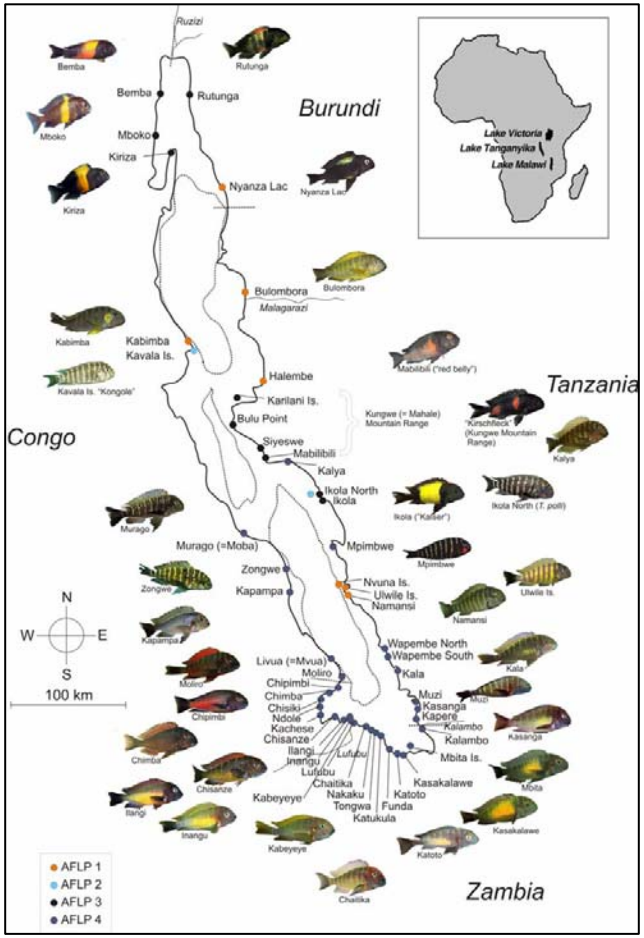
For most genes shown, $K_S > K_A$

But several have overall $K_A > K_S$

Very strong candidates for adaptive evolution?

Wide range in K_S values among genes (several high values??)

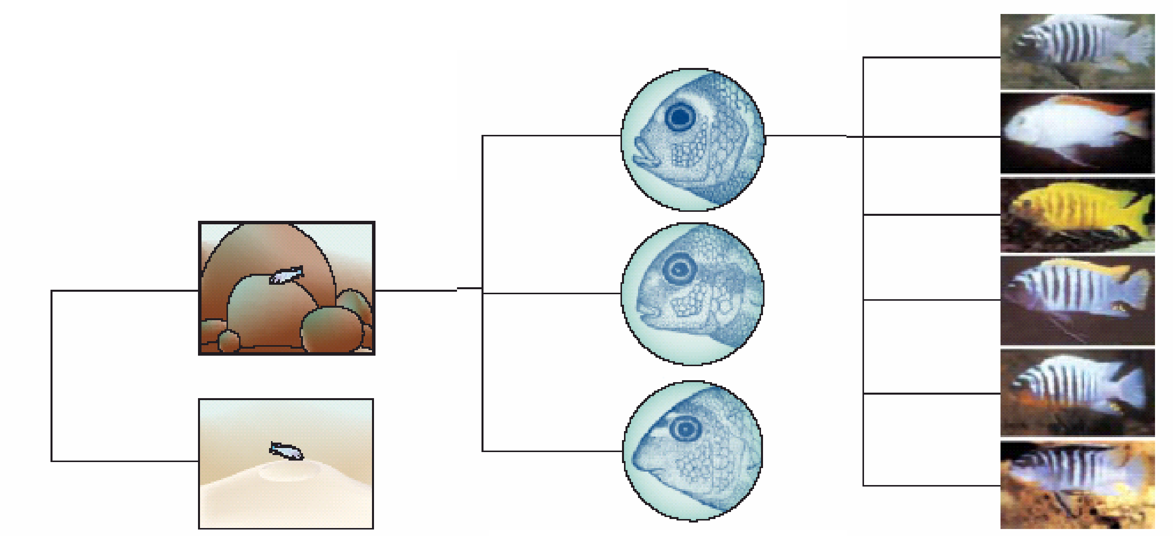
East African lake fish



Lake Tanganyika

Egger *BMC Evol Biol.* 2007

"Adaptive evolution and explosive speciation: the cichlid fish model"



Habitat adaptation

Diversification of feeding apparatus

Diversification of male colour pattern

Researchers interested in identifying genes important for speciation...

SPP120 is "fertilization-related" gene

Kocher *Nature Rev Genet* 2004