

Spring 2024
BIOS 477/877
Bioinformatics and Molecular Evolution
Lecture 4

BIOS477/877 L4 - 1

1

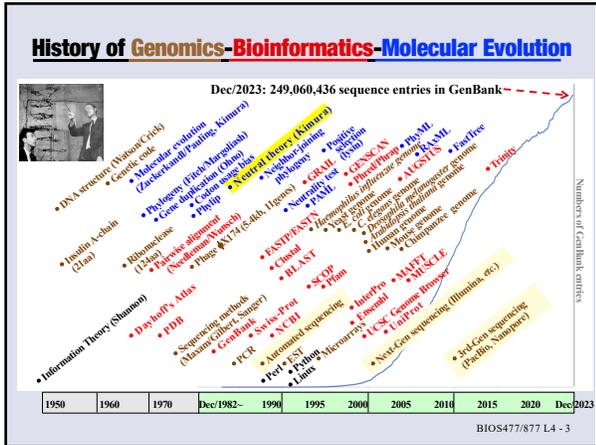
TODAY'S TOPICS

➤ **Molecular Evolution - part 2**

• **Assignment 2**

BIOS477/877 L4 - 2

2



3

The Neutral Theory of Molecular Evolution

- **Kimura, M. (1968)** Evolutionary rate at the molecular level. *Nature* 217: 624-626
- **King, J. L. & Jukes, T. H. (1969)** Non-Darwinian evolution: random fixation of selectively neutral mutations. *Science* 164: 788-798.

➔ **The majority of molecular changes in evolution are due to the random fixation of neutral or nearly neutral mutations**

Selectionist model

Neutralist model

BIOS477/877 L4 - 4

4

The Neutral Theory of Molecular Evolution

The neutral theory of molecular evolution
Motoo Kimura

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution.* Cambridge University Press

- “Functionally less important molecules or parts of molecules evolve faster than more important ones.”
- “Substitutions less disruptive to the existing structure and function of a molecule occur more frequently in evolution than more disruptive ones.”

Inverse correlations between the rate of substitutions and the degree of functional constraints

BIOS477/877 L4 - 5

5

The Neutral Theory of Molecular Evolution

The neutral theory of molecular evolution
Motoo Kimura

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution.* Cambridge University Press

- **The majority of changes found in sequences have little effect in their functions → Observations conform the neutralist model of molecular evolution.**
- **Advantageous mutations happen. They are rare although important.**
- ➔ They are not the driving force of the main part of molecular evolution.

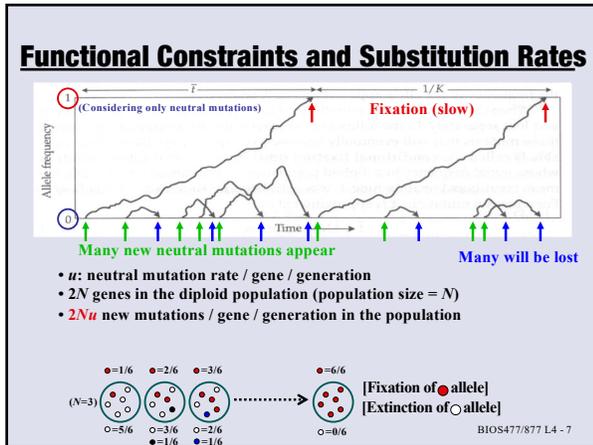
Selectionist model

Neutralist model

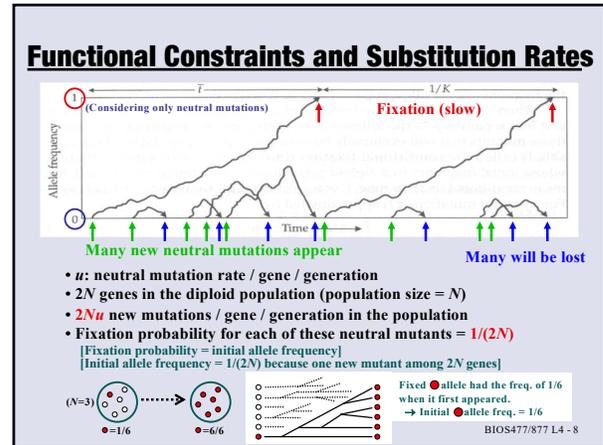
Note: Recent molecular evolutionary studies are more focused on identifying positively selected changes that directly affect the functions/phenotypes.

BIOS477/877 L4 - 6

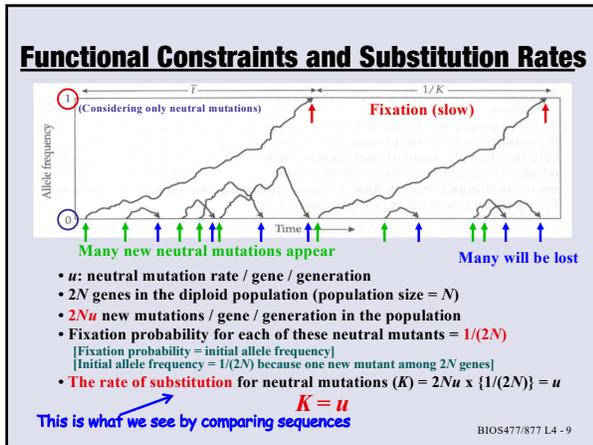
6



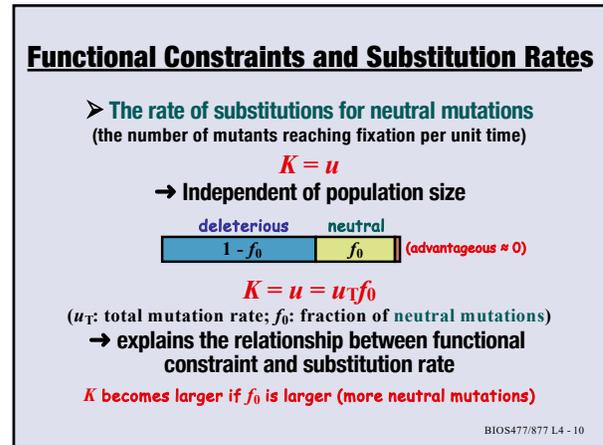
7



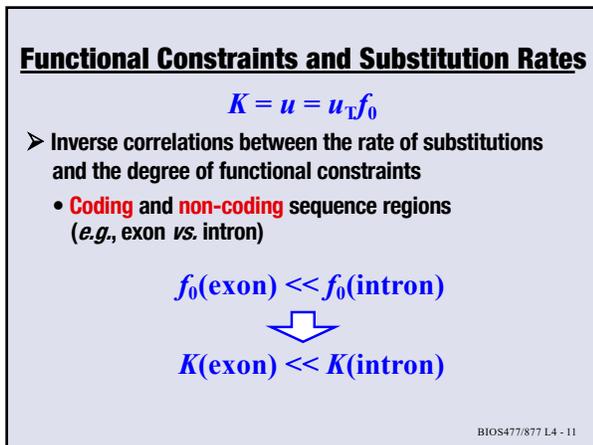
8



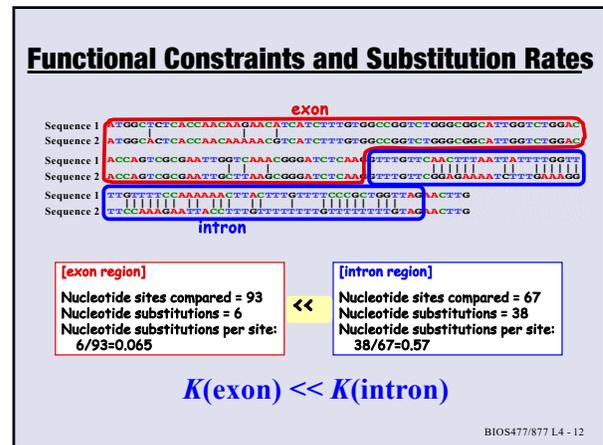
9



10



11



12

Functional Constraints and Substitution Rates

$$K = u = u_{Tf_0}$$

➤ Inverse correlations between the rate of substitutions and the degree of functional constraints

- **Coding and non-coding** sequence regions (e.g., exon vs. intron)
- Nucleotide substitutions at **different codon positions**

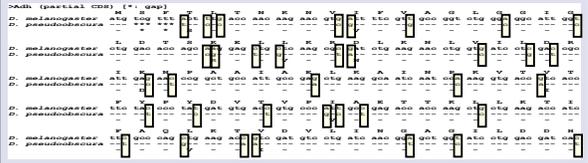


- **Synonymous (silent) substitutions vs. nonsynonymous (replacement) substitutions**

BIOS477/877 L4 - 13

13

Functional Constraints and Substitution Rates



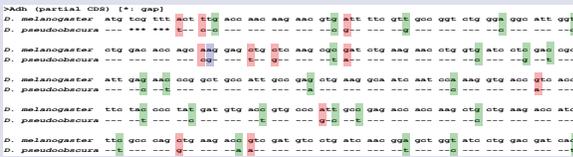
Total nucleotide sites compared = 294
(100 codons, 2 aligned with gaps, 98codons x 3positions = 294bp)

Total nucleotide substitutions = 36
Nucleotide substitutions per site: 36/294=0.12

BIOS477/877 L4 - 14

14

Functional Constraints and Substitution Rates

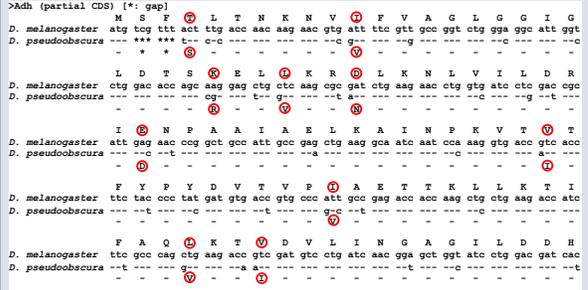


[Codon position]	1st	2nd	3rd	All
# nucleotide sites	98	98	98	294
# nucleotide substitutions	11	1	24	36
Nucleotide substitutions/site	0.11	0.01	0.24	0.12

BIOS477/877 L4 - 15

15

Functional Constraints and Substitution Rates

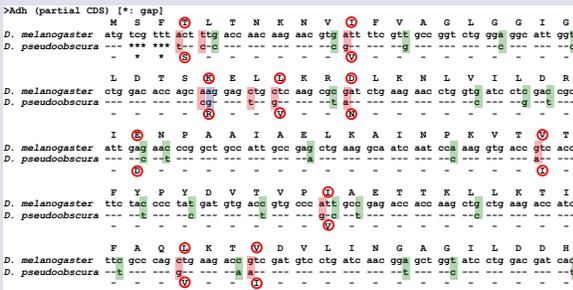


Some nucleotide substitutions change amino acids, and some don't.

BIOS477/877 L4 - 16

16

Functional Constraints and Substitution Rates



(Substitutions at)	(1st)	(2nd)	(3rd)
change amino acids	9/11	1/1	2/24
do not change amino acids	2/11	0/1	22/24

BIOS477/877 L4 - 17

17

Universal Genetic Code

		2nd codon position			
		T	C	A	G
1st codon position	T	TTT Phe	TCT Ser	TAT Tyr	TGT Cys
	C	TCG Ser	TCA Tyr	TAA Stop	TGA Stop
	A	CTT Leu	CCA Pro	CAT His	CGT Arg
	G	CTC Leu	CCC Pro	CAA Gln	CGA Arg
1st codon position	T	ATT Ile	ACT Thr	AAT Asn	AGT Ser
	C	ATC Ile	ACC Thr	AAC Asn	AGC Ser
	A	ATA Ile	ACA Thr	AAA Lys	AGA Arg
	G	ATG Met	ACG Thr	AAG Lys	AGG Arg
1st codon position	T	GTT Val	GCT Ala	GAT Asp	GGT Gly
	C	GTC Val	GCC Ala	GAC Asp	GGC Gly
	A	GTA Val	GCA Ala	GAA Glu	GGA Gly
	G	GTG Val	GCG Ala	GAG Glu	GGG Gly

3rd codon position

BIOS477/877 L4 - 18

18

Functional Constraints and Substitution Rates

Universal Genetic Code

	T	C	A	G
T	TTT Phe	TCT TCC	TAT Tyr	TGT TGG
C	CTA	TCA Ser	TAA Stop	TGA Stop
G	TTG	TGG	TAG Stop	TGG Trp
T	CTT Leu	CCG	CAT His	CGT
C	CTC	CCC	CAC His	CGC
A	CTA	CCA Pro	EAA Gln	CGA Arg
G	CTG	CCG	CAG	CGG
T	ACT Thr	AAT Asp	AAG Arg	AAA Lys
C	ACC	ACA Thr	AAC Asn	AAA Lys
A	ATC Ile	ACU	AUA Met	AUG Met
G	ATG Met	ACG	AUA Met	AUG Met
T	GTT Val	GCC Ala	GAT Asp	GAC
C	GTC	GCC Ala	GAA Glu	GGA Gly
A	GTA	GCA	GAA Glu	GGA Gly
G	GTA	GCA	GAA Glu	GGA Gly

CTA:L

- TTA:L
- CCA:P
- CTT:L
- ATA:I
- CAA:Q
- CTC:L
- GTA:V
- CGA:R
- CTG:L

Substitutions at

	(1st)	(2nd)	(3rd)
No amino acid change	1/3	0/3	< 3/3
With amino acid changes	2/3	3/3	> 0/3

BIOS477/877 L4 - 19

19

Functional Constraints and Substitution Rates

- Nucleotide substitutions in the coding region
 - ~70% of all possible nucleotide changes at the **3rd codon position** do not change amino acids.
 - **Selective constraint is weaker.**
 - **More substitutions.**
- All nucleotide changes at the **2nd codon position** and ~96% at the **1st codon position** change amino acids.
- **Selective constraint is stronger.**
- **Fewer substitutions.**

BIOS477/877 L4 - 20

20

Functional Constraints and Substitution Rates

>Adh (partial CDS) [*: gap]

```
D. melanogaster atg tgg ttt ...
D. pseudoobscura ...
```

[Codon position]	1st	2nd	3rd	All
# nucleotide sites	98	98	98	294
# nucleotide substitutions	11	1	24	36
Nucleotide substitutions/site	0.11	0.01	0.24	0.12

Most conserved ← (1st) (2nd) (3rd) ← Least conserved

$f_0(2nd) < f_0(1st) < f_0(3rd) \Rightarrow K(2nd) < K(1st) < K(3rd)$

BIOS477/877 L4 - 21

21

Functional Constraints and Substitution Rates

- Nucleotide substitutions
 - **Synonymous (silent) substitutions:**
Nucleotide substitutions that **do not change amino acids**
(The majority of nucleotide changes at the 3rd codon position and a few at the 1st codon position)
 - **Nonsynonymous (replacement) substitutions:**
Nucleotide substitutions that **change amino acids**
(All nucleotide changes at the 2nd codon position and the majority at the 1st codon position)

BIOS477/877 L4 - 22

22

>adh

```
D. melanogaster atg tgg ttt ...
D. pseudoobscura ...
```

Legend: █ Synonymous substitutions █ Nonsynonymous substitutions

BIOS477/877 L4 - 23

23

>adh

```
D. melanogaster atg tgg ttt ...
D. pseudoobscura ...
```

Synonymous substitutions: 75*

Nonsynonymous substitutions: 21*

Synonymous > Nonsynonymous

*To simplify, these are the numbers of codons (not numbers of individual positions)

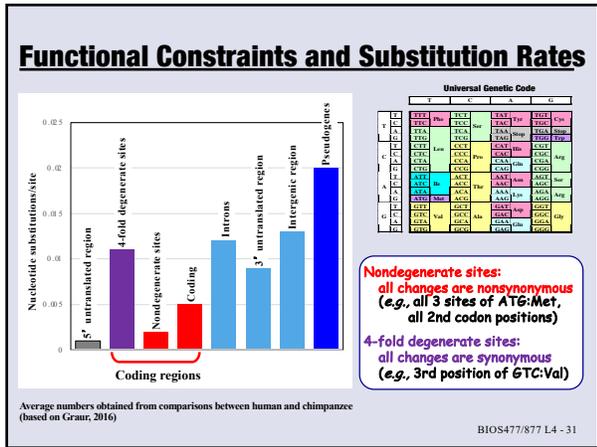
$f_0(\text{nonsynonymous}) < f_0(\text{synonymous})$

$K(\text{nonsynonymous}) < K(\text{synonymous})$

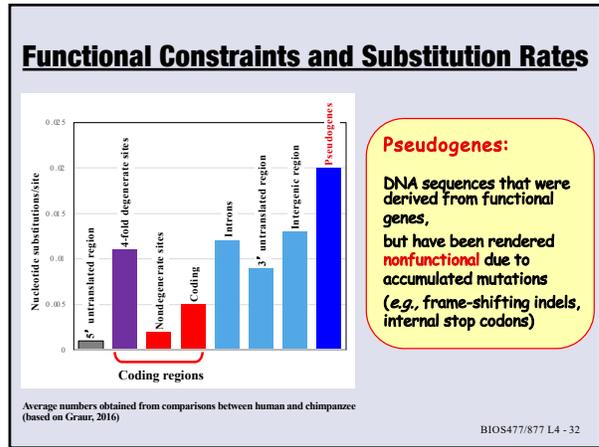
Legend: █ Synonymous substitutions █ Nonsynonymous substitutions

BIOS477/877 L4 - 24

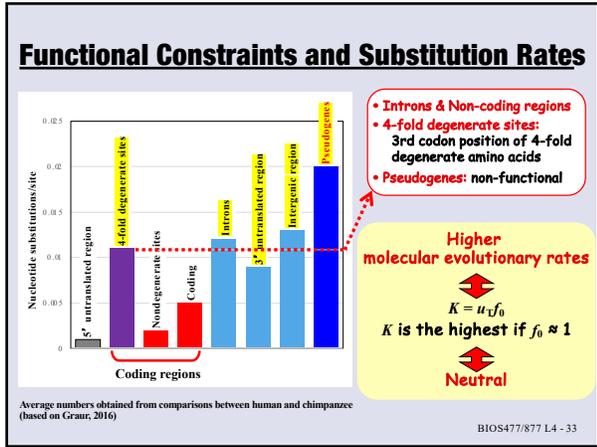
24



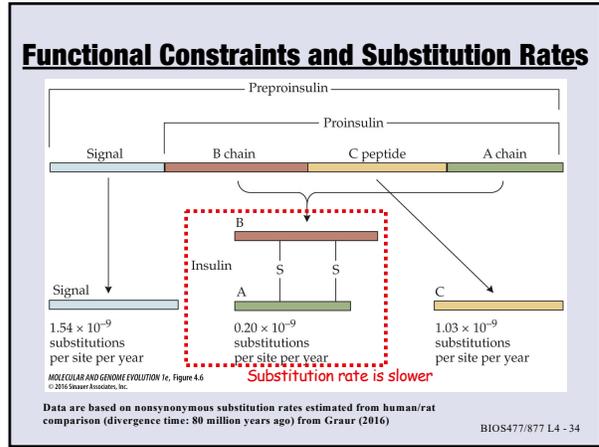
31



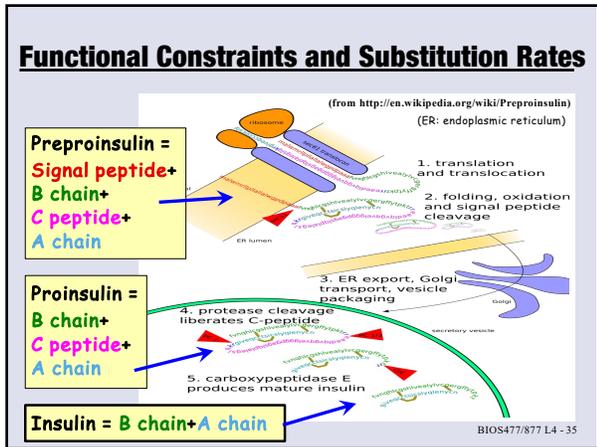
32



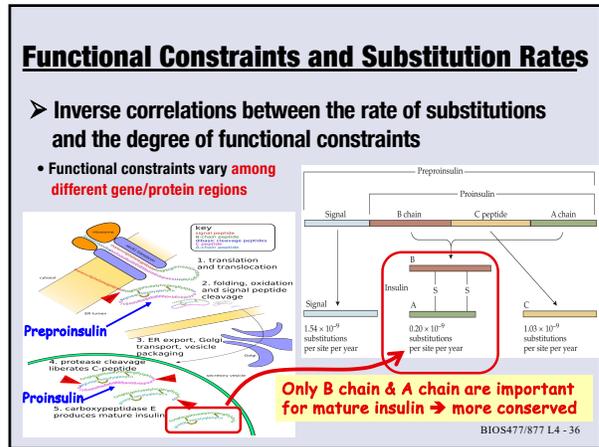
33



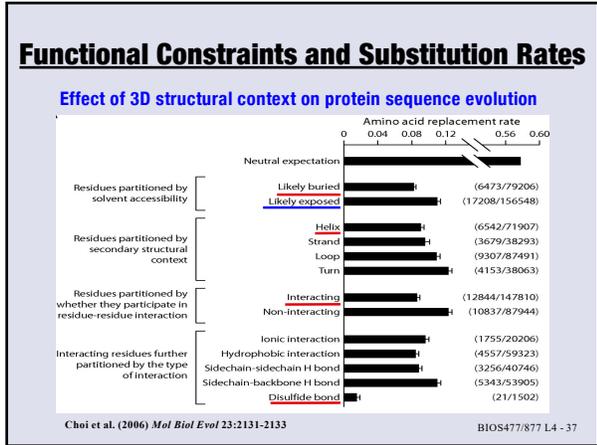
34



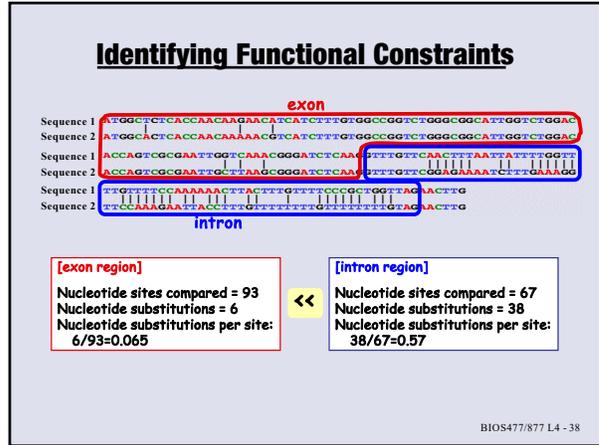
35



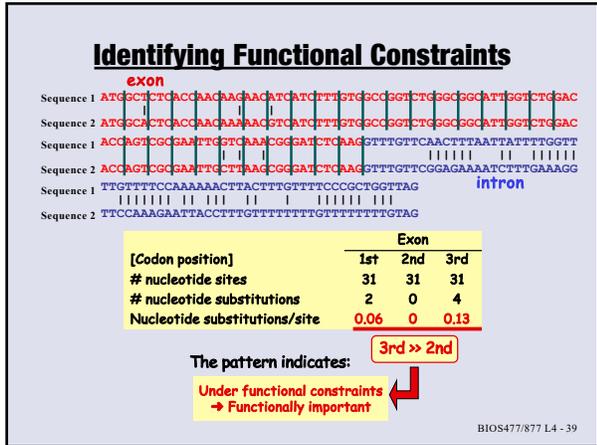
36



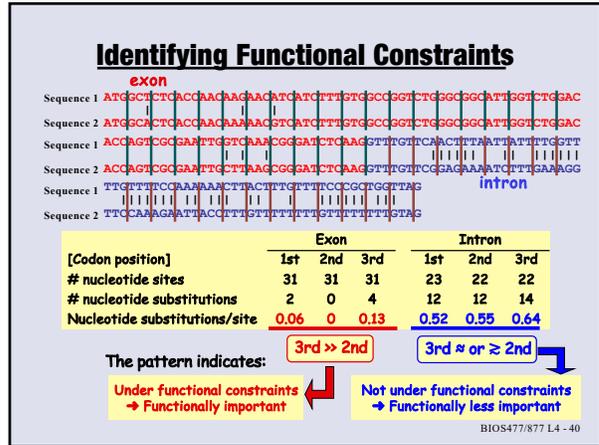
37



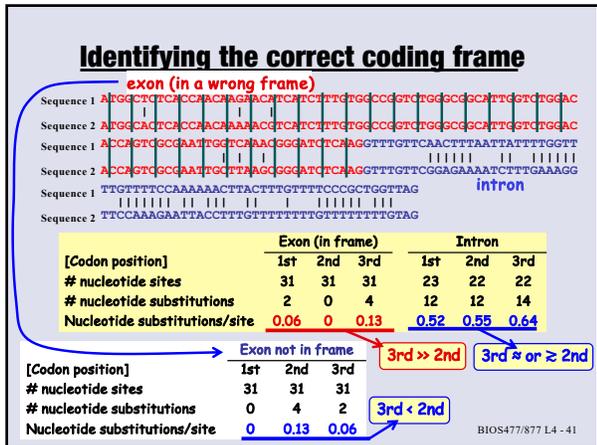
38



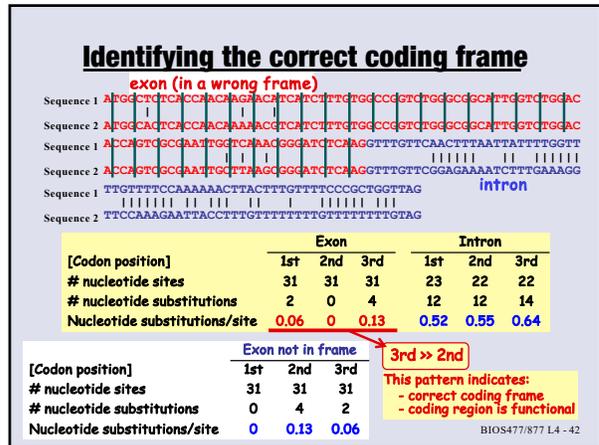
39



40



41



42

Functional Constraints and Substitution Rates

➤ Inverse correlations between the rate of substitutions and the degree of functional constraints

- Functional constraints vary
among different gene/protein regions
among genes/proteins

→ Substitution rates can be used to assess (or predict) functional constraints or functional importance

e.g., exon or intron?

correct coding frame?

functionally important protein region?

functional gene or pseudogene?

BIOS477/877 L4 - 43