Spring 2024

BIOS 477/877

Bioinformatics and Molecular Evolution

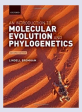
Lecture 3

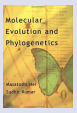
1

TODAY'S TOPICS

➢ Molecular Evolution

- Molecular and Genome Evolution, 1st ed. by Graur, 2016 Sinauer Associates, Inc. [ISBN: 1605354694]
- An Introduction to Molecular Evolution and Phylogenetics, 2nd ed. by Bromham, 2016. Oxford University Press [ISBN: 9780191070693]
- Molecular Evolution and Phylogenetics by Nei & Kumar, 2000 Oxford University Press. [ISBN: 0195135857]
  ➜ Molecular Evolutionary Genetics Analysis (MEGA): free software



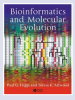


MEGA
Molecular Evolutionary Genetics Analysis

https://www.megasoftware.net/
(for Windows, MacOS, Linux)

2

Molecular Evolution & Bioinformatics



"Bioinformatics and Molecular Evolution"
by Higgs and Attwood (2005) Blackwell Publishing [ISBN: 1405106832]
Ch 1: explains how population genetics and molecular evolution are important and related to bioinformatics
Ch 3: explains the basic mechanisms of molecular evolution
(PDF available on Canvas)

"Fundamental Concepts of Bioinformatics"
by Krane and Raymer (2003) Benjamin Cummings [ISBN: 0805346333]
Ch 3: presents the basics of molecular evolution
(PDF available on Canvas)

"Bioinformatics for Beginners"
by Choudhuri (2014) Elsevier [ISBN: 9780124104716]
Ch 2: presents the foundation of molecular evolution
(PDF available on Canvas)
(The book accessible on Elsevier website from UNL)

*PDF files available in "Molecular Evolution readings" page on Canvas.

3

Molecular Evolution

DNA molecules:

➜ hereditary material for all living organisms

➜ "documents of evolutionary history"
(Emile Zuckerkandle, one of the pioneers of molecular evolution)

The purpose of molecular evolutionary study:

➜ to unravel historical records written in biological molecules,

➜ to fill in the missing gaps found in these records,

➜ to put the information in order, and

➜ to decipher its meaning (Graur and Li, 2000)

4

Molecular Evolution

What we can do:

➜ reconstruct the evolutionary history of genes and genomes (evolution at the molecular level)

➜ reconstruct the evolutionary history of populations and species (evolution at the organismal level)
- can be tracked both in space and in time

➜ attempt to build a classification of the living world

➜ reconstruct the evolution of adaptation (function)

➜ identify the driving forces behind the evolutionary process

5

Molecular Evolution

What's happening in this sequence?

```
>Seq1
ATGGCTCTCACCAACAAGAACATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCAACTTTAATTATTTTGGTT
TTGTTTTCCAAAAAACTTACTTTGTTTTCCCGCTGGTTAGAACTTG
```

It is not easy to find any pattern from a single sequence

6

Molecular Evolution

Comparison gives us a lot more power!

```
>Seq1
ATGGCTCTCACCAACAAGAACATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCAACTTTAATTATTTTGGTT
TTGTTTTCCAAAAAACTTACTTTGTTTTCCCGCTGGTTAGAACTTG

>Seq2
ATGGCACTCACCAACAAAAACGTCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
ACCAGTCGCGAATTGCTTAAGCGGGATCTCAAGGTTTGTTCGGAGAAAAATCTTTGAAAGG
TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```


7

Molecular Evolution

Sequence comparison = Alignment

```
Seq1  ATGGCTCTCACCAACAAGAACATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2  ATGGCACTCACCAACAAAAACGTCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1  ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCAACTTTAATTATTTTGGTT
Seq2  ACCAGTCGCGAATTGCTTAAGCGGGATCTCAAGGTTTGTTCGGAGAAAAATCTTTGAAAGG

Seq1  TTGTTTTCCAAAAAACTTACTTTGTTTTCCCGCTGGTTAGAACTTG
Seq2  TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```


8

Molecular Evolution

Sequence comparison = Alignment

```
Seq1  ATGGCTCTCACCAACAAGAACATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2  ATGGCACTCACCAACAAAAACGTCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1  ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCAACTTTAATTATTTTGGTT
Seq2  ACCAGTCGCGAATTGCTTAAGCGGGATCTCAAGGTTTGTTCGGAGAAAAATCTTTGAAAGG

Seq1  TTGTTTTCCAAAAAACTTACTTTGTTTTCCCGCTGGTTAGAACTTG
Seq2  TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```


9

Molecular Evolution

Can you see any pattern from the alignment below?

```
Seq1  ATGGCTCTCACCAACAAGAACATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2  ATGGCACTCACCAACAAAAACGTCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1  ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCAACTTTAATTATTTTGGTT
Seq2  ACCAGTCGCGAATTGCTTAAGCGGGATCTCAAGGTTTGTTCGGAGAAAATCTTTGAAAGG

Seq1  TTGTTTTCCAAAAAACTTACTTTGTTTTCCCGCTGGTTAGAACTTG
Seq2  TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```

10

Molecular Evolution

Which region is functionally more important?

Region A [more conserved]

```
Seq1  ATGGCTCTCACCAACAAGAACATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2  ATGGCACTCACCAACAAAAACGTCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1  ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCAACTTTAATTATTTTGGTT
Seq2  ACCAGTCGCGAATTGCTTAAGCGGGATCTCAAGGTTTGTTCGGAGAAAAATCTTTGAAAGG

Seq1  TTGTTTTCCAAAAAACTTACTTTGTTTTCCCGCTGGTTAGAACTTG
Seq2  TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```

Region B [less conserved]

A: More conserved
B: Less conserved

12

Molecular Evolution

By comparing sequences, we can speculate functional regions of the gene.

Region A [more conserved]

```
Seq1  ATGGCTCTCACCAACAAGAACATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2  ATGGCACTCACCAACAAAAACGTCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1  ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCAACTTTAATTATTTTGGTT
Seq2  ACCAGTCGCGAATTGCTTAAGCGGGATCTCAAGGTTTGTTCGGAGAAAAATCTTTGAAAGG

Seq1  TTGTTTTCCAAAAAACTTACTTTGTTTTCCCGCTGGTTAGAACTTG
Seq2  TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```

Region B [less conserved]

[Eukaryotic gene: exons + introns]

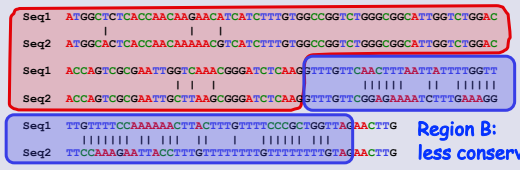
Which is exon, A or B?

14

Molecular Evolution

Exons are functionally more important than introns.

Region A: more conserved



Region B: less conserved

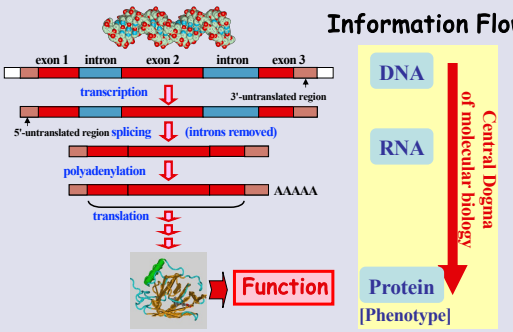
usually

Exon (codes amino acids): functionally more important

Intron (does not code amino acids): less important

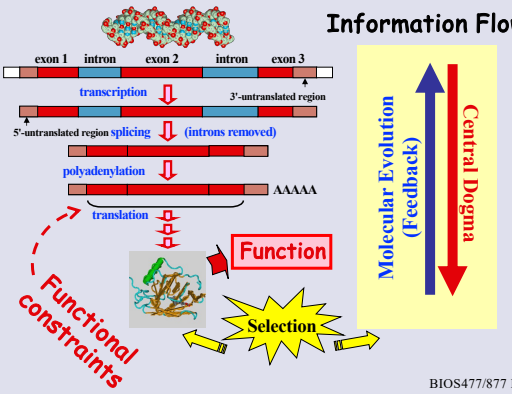
15

Information Flow

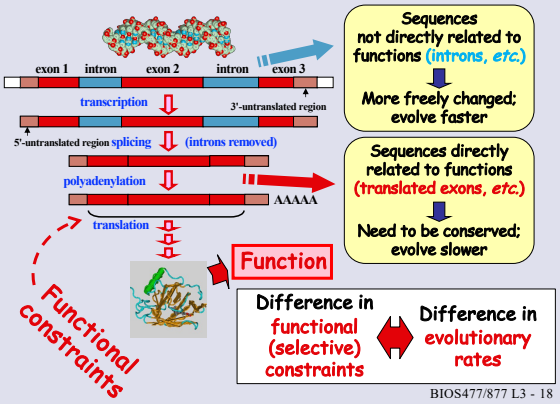


16

Information Flow



17



18

Molecular Evolution

provides foundations and frameworks for bioinformatics methods

"Nothing in biology makes sense except in the light of evolution"

Dobzhansky (1973)

Molecular Evolution:
Mechanisms of sequence evolution

Molecular Evolutionary Methods:
Sequence distance calculation
Phylogenetic reconstruction
Positive selection detection
Population genetics analysis

Bioinformatics Methods:
Similarity search
Functional domain search
Protein family classification
Functional annotation
Gene prediction

19

Molecular Evolution

provides foundations and frameworks for bioinformatics methods

"Nothing in bioinformatics makes sense except in the light of evolution"

Higgs and Attwood (Chapter 1)

Molecular Evolution:
Mechanisms of sequence evolution

Molecular Evolutionary Methods:
Sequence distance calculation
Phylogenetic reconstruction
Positive selection detection
Population genetics analysis

Bioinformatics Methods:
Similarity search
Functional domain search
Protein family classification
Functional annotation
Gene prediction

20

Molecular Evolution

provides foundations and frameworks for bioinformatics methods

More conserved regions are predicted to be more functionally important

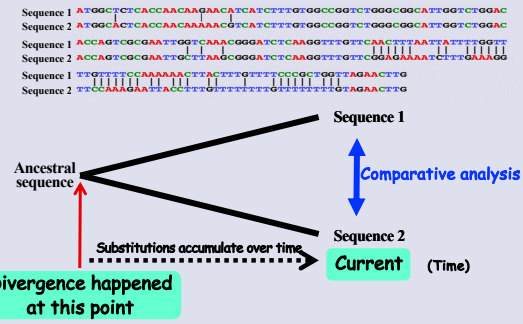
Molecular Evolution:
Mechanisms of sequence evolution

Molecular Evolutionary Methods:
Sequence distance calculation
Phylogenetic reconstruction
Positive selection detection
Population genetics analysis

Bioinformatics Methods:
Similarity search
Functional domain search
Protein family classification
Functional annotation
Gene prediction

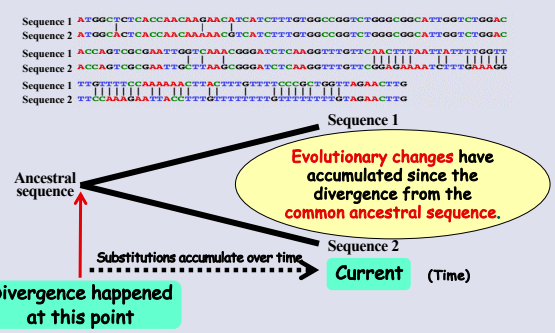
21

Molecular Evolution: Power of Comparison



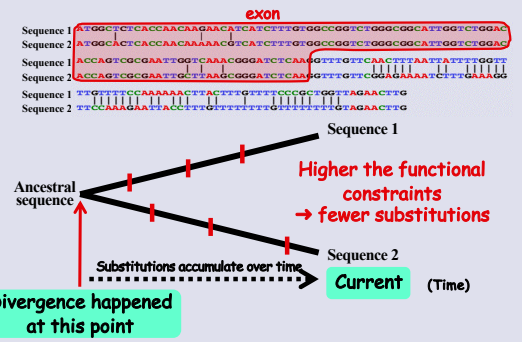
22

Molecular Evolution: Power of Comparison



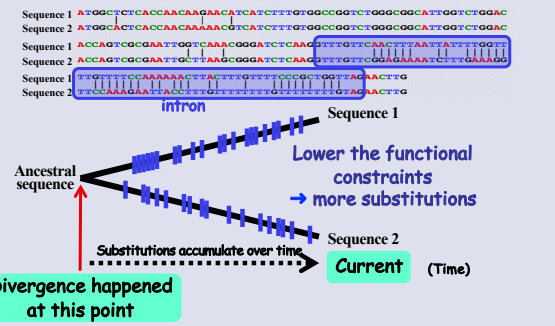
23

Molecular Evolution: Power of Comparison



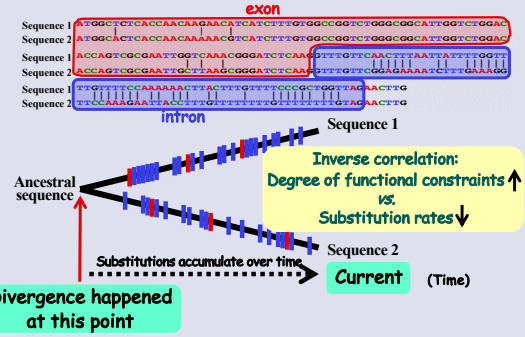
24

Molecular Evolution: Power of Comparison



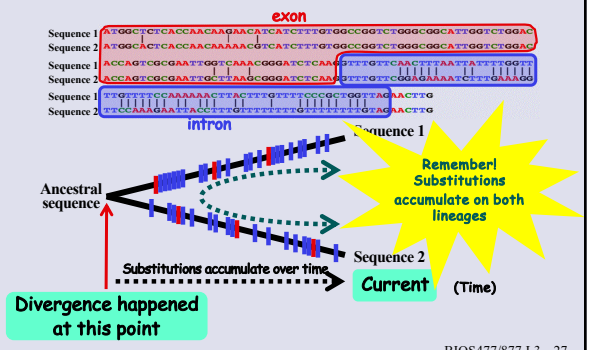
25

Molecular Evolution: Power of Comparison



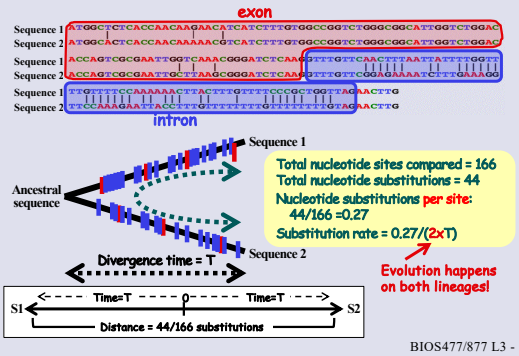
26

Molecular Evolution: Power of Comparison



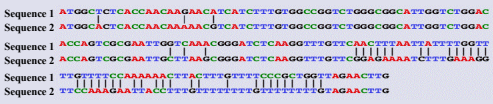
27

Molecular Evolution: Power of Comparison



28

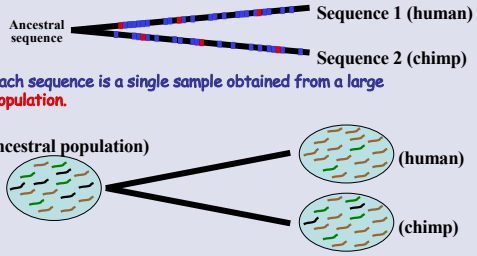
Causes of Molecular Evolution



How do nucleotide or amino acid substitutions happen?
What are substitutions?
What is molecular evolution?

29

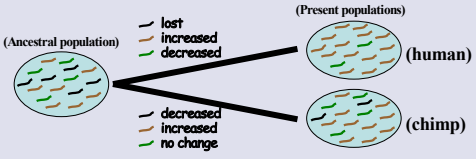
Causes of Molecular Evolution



Evolution happens at the population level.

30

Causes of Molecular Evolution



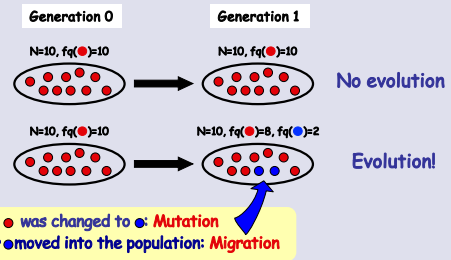
Evolution happens at the population level.

[Definition]
Evolution: any change in the frequency of alleles within a gene pool from one generation to the next.
(population)

31

Causes of Molecular Evolution

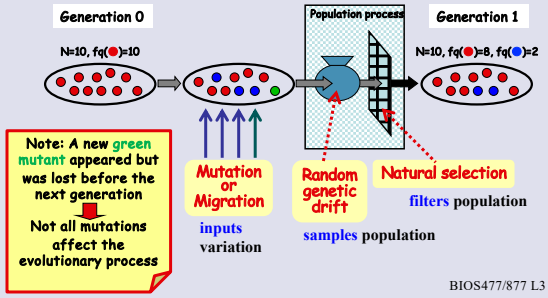
➢ What causes changes in allele frequencies?



32

Causes of Molecular Evolution

➢ What causes changes in allele frequencies?



33

Causes of Molecular Evolution

➢ Three forces that change allele frequencies

1. Mutation
   - ➜ Ultimate source of variation
   - ➜ Point mutations (*e.g.*, A -> T)
   - ➜ Recombinations 
   - ➜ Insertions and deletions (indels) 

(Migration can also introduce new variation)

34

Causes of Molecular Evolution

➢ Three forces that change allele frequencies

1. Mutation
   - ➜ Mutation rates are usually very low (10^{-8}~10^{-10}/base/year)
   - ➜ Higher mutation rates are found in some genomes

     [due to weak (or no) repair systems, *etc.*]

     e.g., animal mitochondrial DNA

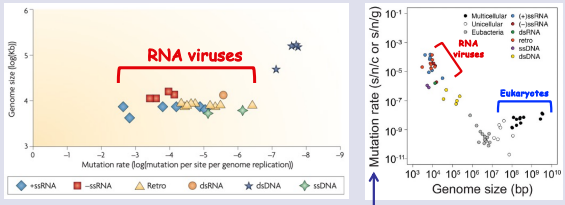
     some RNA viruses (*e.g.*, retroviruses) (10^{-3} ~ 10^{-2} /base/year!)

35

Causes of Molecular Evolution

➢ Three forces that change allele frequencies

1. Mutation



+ssRNA (SARS-CoV-2, West Nile, dengue, Zika, *etc.*)
-ssRNA (Ebola, influenza, *etc.*)
Retroviruses (HIV, *etc.*)

Duffy *et al.* (2008)

Peck & Lauring (2018)

s/n/c: substitutions per nucleotide site per cell infection
s/n/g: substitutions per nucleotide site per generation

36

Causes of Molecular Evolution

➢ Three forces that change allele frequencies

2. Natural Selection

[Definition]
Differential reproduction of genetically distinct individuals or genotypes within a population

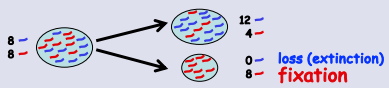
- ➜ Negative (purifying) selection
  Deleterious mutations are selected against and removed from the population quickly
- ➜ Positive (advantageous) selection
  Advantageous mutations will be fixed in the population
- ➜ Neutral mutations: free from natural selection

37

Causes of Molecular Evolution

➢ Three forces that change allele frequencies

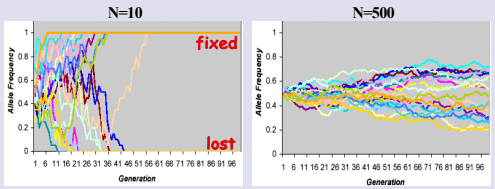
3. Random Genetic Drift
   - ➜ Random fluctuations in allele frequencies produced by random sampling of gametes in the process of reproduction
   - ➜ Direction of the change is random
   - ➜ Its effect is larger in small populations



38

Causes of Molecular Evolution

- **Random Genetic Drift** and population size



N: population size
2 alleles: A or a (both neutral)
Fq(A) = Fq(a) = 0.5 at Generation 0

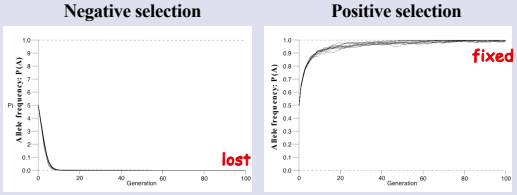
Population size affects how long neutral mutations persist

Simulation program:
POPULUS: https://cbs.umn.edu/populus

39

Causes of Molecular Evolution

- **Natural Selection**

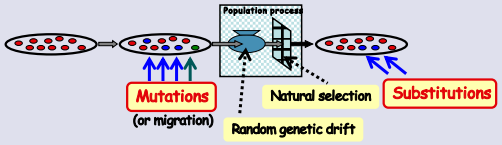


Simulation program:
POPULUS: https://cbs.umn.edu/populus

40

Mutation *vs.* Substitution

- **Mutations** are changes in nucleotide sequences that occur due to **errors in replication or repair.**
- **Substitutions** are mutations that have **passed through population process (drift and selection).**



Mutations ≠ Substitutions !!

41

Mutation *vs.* Substitution

- **Mutations** are changes in nucleotide sequences that occur due to **errors in replication or repair.**
- **Substitutions** are mutations that have **passed through population process (drift and selection).**

Mutations ≠ Substitutions !!

- It is very difficult to directly estimate mutation rates.
- Changes observed by comparing sequences are **substitutions** (after drift and selection).
  e.g. nucleotide substitutions, amino acid substitutions

Only if no drift nor selection, substitution rate = mutation rate

42

Worksheet for calculating the number of nucleotide substitutions

exon

Sequence 1 ATGGCTCTCACCAACAAGAACATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Sequence 2 ATGGCACTCACCAACAAAAACGTCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Sequence 1 ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCAACTTTAATTATTTGGTT
Sequence 2 ACCAGTCGCGAATTGCTTAAGCGGGATCTCAAGGTTTGTTCGGAGAAAATCTTTGAAAGG
Sequence 1 TTGTTTTCCAAAAAACTTACTTTGTTTCCCGCTGGTTAGAACTTG
Sequence 2 TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG

intron

(This is the same alignment used in this lecture)

[exon]

Number of nucleotide sites compared within the exon = _____
Number of nucleotide substitutions within the exon = _____
Number of nucleotide substitutions per site within the exon = _____

[intron]

Number of nucleotide sites compared within the intron = _____
Number of nucleotide substitutions within the intron = _____
Number of nucleotide substitutions per site within the intron = _____

(Downloadable from Canvas)

43

>Adh (partial CDS)
D. melanogaster atg tcg ttt act ttg acc aac aag aac gtg att ttc gtt gcc ggt ctg gga ggc att ggt
D. pseudoobscura --- *** *** t-- c-c --- --- --- --c g-- --- --g --- --- --- --c --- --- --c

D. melanogaster ctg gac acc agc aag gag ctg ctc aag cgc gat ctg aag aac ctg gtg atc ctc gac cgc
D. pseudoobscura --- --- --- --- cg- --- t-- g-- --- --t a-- --- --- --- --- --c --- --g --t ---

D. melanogaster att gag aac ccg gct gcc att gcc gag ctg aag gca atc aat cca aag gtg acc gtc acc
D. pseudoobscura --- --c --t --- --- --- --- --- --a --- --- --- --- --- --c --- --- --- a-- ---

D. melanogaster ttc tac ccc tat gat gtg acc gtg ccc att gcc gag acc acc aag ctg ctg aag acc atc
D. pseudoobscura --- --t --- --c --- --- --t --- --- g-c --t --- --- --- --- --c --- --- --- ---

D. melanogaster ttc gcc cag ctg aag acc gtc gat gtc ctg atc aac gga gct ggt atc ctg gac gat cac
D. pseudoobscura --t --- --- g-- --- --a a-- --- --- --- --- --- --t --- --c --- --- --- --- --t

(*: gap; exclude the gap sites from the calculation)

[1st codon position]
Number of nucleotide sites compared in the 1st codon position = _____
Number of nucleotide substitutions in the 1st codon position = _____
Number of nucleotide substitutions per site in the 1st codon position = _____

[2nd codon position]
Number of nucleotide sites in the 2nd codon position = _____
Number of nucleotide substitutions in the 2nd codon position = _____
Number of nucleotide substitutions per site in the 2nd codon position = _____

[3rd codon position]
Number of nucleotide sites in the 3rd codon position = _____
Number of nucleotide substitutions in the 3rd codon position = _____
Number of nucleotide substitutions per site in the 3rd codon position = _____

(Downloadable from Canvas)

44