

Spring 2024  
**BIOS 477/877**  
*Bioinformatics and Molecular Evolution*  
**Lecture 3**


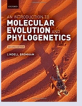
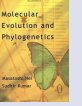

BIOS477/877 L3 - 1

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**, 2<sup>nd</sup> 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







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


BIOS477/877 L3 - 2

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.

BIOS477/877 L3 - 3

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)


BIOS477/877 L3 - 4

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



BIOS477/877 L3 - 5

5

**Molecular Evolution**

**What's happening in this sequence?**

```
>Seq1
ATGGCTCTCAACCAACAGAACATCATCTTTGTGGCCGGTCTGGGGCCATTGCTCTGGAC
ACCAATCGCGAATTTGGTCACACGGGATCTCAAGTTTGTTCACACTTAAATATTTGGTT
TTCTTTTCAGAAAACCTACTTTGTTTCCCGCTGGTGAACCTTG
```

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

BIOS477/877 L3 - 6

6

## Molecular Evolution

Comparison gives us a lot more power!

```
>Seq1
ATGGCTCTCACCAACAAGAATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
ACCACTCGCGAATTGCTTAAAGCGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT
TTGTTTTCCAAAAAAGTACTTTGTTTTTCCCGCTGAGAACTTG

>Seq2
ATGGCACTCACCAACAAAAAGTCACTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
ACCACTCGCGAATTGCTTAAAGCGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT
TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```

BIOS477/877 L3 - 7

7

## Molecular Evolution

Sequence comparison = Alignment

```
Seq1 ATGGCTCTCACCAACAAGAATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2 ATGGCACTCACCAACAAAAAGTCACTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1 ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT
Seq2 ACCACTCGCGAATTGCTTAAAGCGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT

Seq1 TTGTTTTCCAAAAAAGTACTTTGTTTTTCCCGCTGAGAACTTG
Seq2 TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```

BIOS477/877 L3 - 8

8

## Molecular Evolution

Sequence comparison = Alignment

```
Seq1 ATGGCTCTCACCAACAAGAATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2 ATGGCACTCACCAACAAAAAGTCACTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1 ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT
Seq2 ACCACTCGCGAATTGCTTAAAGCGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT

Seq1 TTGTTTTCCAAAAAAGTACTTTGTTTTTCCCGCTGAGAACTTG
Seq2 TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```

BIOS477/877 L3 - 9

9

## Molecular Evolution

Can you see any pattern from the alignment below?

```
Seq1 ATGGCTCTCACCAACAAGAATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2 ATGGCACTCACCAACAAAAAGTCACTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1 ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT
Seq2 ACCACTCGCGAATTGCTTAAAGCGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT

Seq1 TTGTTTTCCAAAAAAGTACTTTGTTTTTCCCGCTGAGAACTTG
Seq2 TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```



BIOS477/877 L3 - 10

10

## Molecular Evolution

Which region is functionally more important?

Region A [more conserved]

```
Seq1 ATGGCTCTCACCAACAAGAATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2 ATGGCACTCACCAACAAAAAGTCACTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1 ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT
Seq2 ACCACTCGCGAATTGCTTAAAGCGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT

Seq1 TTGTTTTCCAAAAAAGTACTTTGTTTTTCCCGCTGAGAACTTG
Seq2 TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```

A: More conserved  
B: Less conserved



Region B [less conserved]

BIOS477/877 L3 - 12

12

## Molecular Evolution

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

Region A [more conserved]

```
Seq1 ATGGCTCTCACCAACAAGAATCATCTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC
Seq2 ATGGCACTCACCAACAAAAAGTCACTTTGTGGCCGGTCTGGGCGGCATTGGTCTGGAC

Seq1 ACCAGTCGCGAATTGGTCAAACGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT
Seq2 ACCACTCGCGAATTGCTTAAAGCGGGATCTCAAGGTTTGTTCACCTTTAATTTTGGTT

Seq1 TTGTTTTCCAAAAAAGTACTTTGTTTTTCCCGCTGAGAACTTG
Seq2 TTCCAAAGAATTACCTTTGTTTTTTTTGTTTTTTTTGTAGAACTTG
```

[Eukaryotic gene: exons + introns]  
Which is exon, A or B?

Region B [less conserved]

BIOS477/877 L3 - 14

14

## Molecular Evolution

Exons are functionally more important than introns.

**Region A: more conserved**

```

Seq1 ATGGCTCTCACCAACAAGAACATCTTTGGCCGGTCTGGCGGCATTGGTCTGGAC
Seq2 ATGGCACTCACCAACAAAAACGTCACTTTGGCCGGTCTGGCGGCATTGGTCTGGAC
Seq1 ACCAGTCGCGAATTGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Seq2 ACCAGTCGCGAATTGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
  
```

**Region B: less conserved**

```

Seq1 TTGTTTCCAAAAACTTACTTTGTTTCCCGCTGGTTAGAACTTG
Seq2 TTCAAAGAATTACCTTTGTTTTTTTGTGTTTTTTGTTAGAACTTG
  
```

usually  
**Exon (codes amino acids): functionally more important**  
**Intron (does not code amino acids): less important**

BIOS477/877 L3 - 15

15

## Information Flow

DNA → RNA → Protein [Phenotype]  
 Central Dogma of molecular biology

BIOS477/877 L3 - 16

16

## Information Flow

Molecular Evolution (Feedback)  
 Central Dogma

BIOS477/877 L3 - 17

17

## Information Flow

**Sequences not directly related to functions (introns, etc.)**  
 More freely changed; evolve faster

**Sequences directly related to functions (translated exons, etc.)**  
 Need to be conserved; evolve slower

Difference in functional (selective) constraints ↔ Difference in evolutionary rates

BIOS477/877 L3 - 18

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

BIOS477/877 L3 - 19

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

BIOS477/877 L3 - 20

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

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

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

BIOS477/877 L3 - 21

21

## Molecular Evolution: Power of Comparison

```

Sequence 1 ATGGCTCTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 2 ATGGCACTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 1 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 2 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 1 TTTTTCCAAAAACTTACTTTGTTTCCCGCGGTAGAACTTG
Sequence 2 TTTCAAGAAATACCTTTGTTTCCCGCGGTAGAACTTG
    
```

BIOS477/877 L3 - 22

22

## Molecular Evolution: Power of Comparison

```

Sequence 1 ATGGCTCTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 2 ATGGCACTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 1 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 2 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 1 TTTTTCCAAAAACTTACTTTGTTTCCCGCGGTAGAACTTG
Sequence 2 TTTCAAGAAATACCTTTGTTTCCCGCGGTAGAACTTG
    
```

BIOS477/877 L3 - 23

23

## Molecular Evolution: Power of Comparison

```

Sequence 1 ATGGCTCTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 2 ATGGCACTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 1 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 2 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 1 TTTTTCCAAAAACTTACTTTGTTTCCCGCGGTAGAACTTG
Sequence 2 TTTCAAGAAATACCTTTGTTTCCCGCGGTAGAACTTG
    
```

BIOS477/877 L3 - 24

24

## Molecular Evolution: Power of Comparison

```

Sequence 1 ATGGCTCTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 2 ATGGCACTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 1 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 2 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 1 TTTTTCCAAAAACTTACTTTGTTTCCCGCGGTAGAACTTG
Sequence 2 TTTCAAGAAATACCTTTGTTTCCCGCGGTAGAACTTG
    
```

BIOS477/877 L3 - 25

25

## Molecular Evolution: Power of Comparison

```

Sequence 1 ATGGCTCTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 2 ATGGCACTCACCAACAAGAACATCATCTTTGGCCGGTCTGGCGGCATTGGCTGGAC
Sequence 1 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 2 ACCAGTCGGGAATGGTCAACGGGATCTCAAGTTTGTTCACCTTAAATATTTGGTT
Sequence 1 TTTTTCCAAAAACTTACTTTGTTTCCCGCGGTAGAACTTG
Sequence 2 TTTCAAGAAATACCTTTGTTTCCCGCGGTAGAACTTG
    
```

BIOS477/877 L3 - 26

26

### Molecular Evolution: Power of Comparison

Sequence 1: ATGGCTCTCACCACAAAGAACATCATCTTTGCGCCGCTCGGGCCGCAATTGGCTGGAC  
 Sequence 2: ATGGCACTCACCACAAAACATCATCTTTGCGCCGCTCGGGCCGCAATTGGCTGGAC  
 Sequence 1: ACCAGTCGCGAATTCCTTAAACGGGATCTCAAGGTTGTTTCACCTTAAATATTTGGT  
 Sequence 2: ACCAGTCGCGAATTCCTTAAACGGGATCTCAAGGTTGTTTCACCTTAAATATTTGGT  
 Sequence 1: TTTGTTTCCAAAACCTTACTTATGTTTCCCGCTGGTTAGACTTG  
 Sequence 2: TTTGTTTCCAAAACCTTACTTATGTTTCCCGCTGGTTAGACTTG

exon  
intron

Ancestral sequence

Substitutions accumulate over time

Remember! Substitutions accumulate on both lineages

Divergence happened at this point

Sequence 1  
Sequence 2  
Current (Time)

BIOS477/877 L3 - 27

27

### Molecular Evolution: Power of Comparison

Sequence 1: ATGGCTCTCACCACAAAGAACATCATCTTTGCGCCGCTCGGGCCGCAATTGGCTGGAC  
 Sequence 2: ATGGCACTCACCACAAAACATCATCTTTGCGCCGCTCGGGCCGCAATTGGCTGGAC  
 Sequence 1: ACCAGTCGCGAATTCCTTAAACGGGATCTCAAGGTTGTTTCACCTTAAATATTTGGT  
 Sequence 2: ACCAGTCGCGAATTCCTTAAACGGGATCTCAAGGTTGTTTCACCTTAAATATTTGGT  
 Sequence 1: TTTGTTTCCAAAACCTTACTTATGTTTCCCGCTGGTTAGACTTG  
 Sequence 2: TTTGTTTCCAAAACCTTACTTATGTTTCCCGCTGGTTAGACTTG

exon  
intron

Ancestral sequence

Divergence time = T

Distance = 44/166 substitutions

Total nucleotide sites compared = 166  
 Total nucleotide substitutions = 44  
 Nucleotide substitutions per site =  $44/166 = 0.27$   
 Substitution rate =  $0.27/(2 \times T)$

Evolution happens on both lineages!

Sequence 1  
Sequence 2

BIOS477/877 L3 - 28

28

### Causes of Molecular Evolution

Sequence 1: ATGGCTCTCACCACAAAGAACATCATCTTTGCGCCGCTCGGGCCGCAATTGGCTGGAC  
 Sequence 2: ATGGCACTCACCACAAAACATCATCTTTGCGCCGCTCGGGCCGCAATTGGCTGGAC  
 Sequence 1: ACCAGTCGCGAATTCCTTAAACGGGATCTCAAGGTTGTTTCACCTTAAATATTTGGT  
 Sequence 2: ACCAGTCGCGAATTCCTTAAACGGGATCTCAAGGTTGTTTCACCTTAAATATTTGGT  
 Sequence 1: TTTGTTTCCAAAACCTTACTTATGTTTCCCGCTGGTTAGACTTG  
 Sequence 2: TTTGTTTCCAAAACCTTACTTATGTTTCCCGCTGGTTAGACTTG

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

BIOS477/877 L3 - 29

29

### Causes of Molecular Evolution

Ancestral sequence

Sequence 1 (human)  
Sequence 2 (chimp)

Each sequence is a single sample obtained from a large population.

(Ancestral population)

(human)  
(chimp)

Evolution happens at the population level.

BIOS477/877 L3 - 30

30

### Causes of Molecular Evolution

(Ancestral population)

(Present populations)

(human)  
(chimp)

lost  
increased  
decreased

decreased  
increased  
no change

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)

BIOS477/877 L3 - 31

31

### Causes of Molecular Evolution

➤ What causes changes in allele frequencies?

Generation 0  
 N=10,  $fq(\bullet)=10$

Generation 1  
 N=10,  $fq(\bullet)=10$  No evolution  
 N=10,  $fq(\bullet)=10$   
 N=10,  $fq(\bullet)=8, fq(\circ)=2$  Evolution!

● was changed to ○: Mutation  
 ○ moved into the population: Migration

BIOS477/877 L3 - 32

32

## Causes of Molecular Evolution

➤ What causes changes in allele frequencies?

**Note:** A new green mutant appeared but was lost before the next generation  
 Not all mutations affect the evolutionary process

BIOS477/877 L3 - 33

33

## Causes of Molecular Evolution

➤ Three forces that change allele frequencies

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

(Migration can also introduce new variation)

BIOS477/877 L3 - 34

34

## Causes of Molecular Evolution

➤ Three forces that change allele frequencies

- 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!)

BIOS477/877 L3 - 35

35

## Causes of Molecular Evolution

➤ Three forces that change allele frequencies

- Mutation**

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

BIOS477/877 L3 - 36

36

## Causes of Molecular Evolution

➤ Three forces that change allele frequencies

- 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

BIOS477/877 L3 - 37

37

## Causes of Molecular Evolution

➤ Three forces that change allele frequencies

- 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

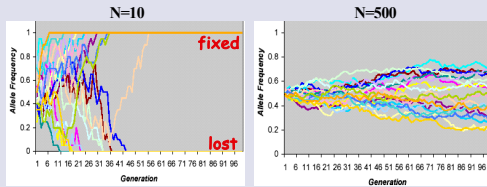
BIOS477/877 L3 - 38

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

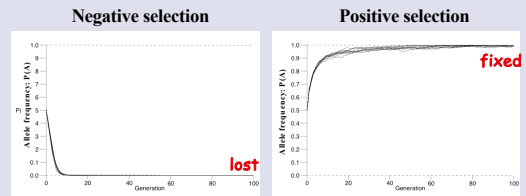
Simulation program:  
POPULUS: <https://cbs.uinn.edu/napulus>

BIOS477/877 L3 - 39

39

## Causes of Molecular Evolution

- **Natural Selection**



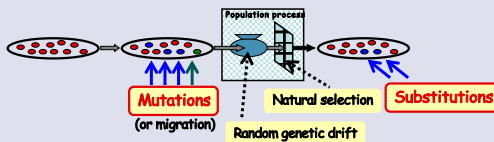
Simulation program:  
POPULUS: <https://cbs.uinn.edu/napulus>

BIOS477/877 L3 - 40

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 !!

BIOS477/877 L3 - 41

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

BIOS477/877 L3 - 42

42

### Worksheet for calculating the number of nucleotide substitutions

Sequence 1 ATGGCTCCACCAACAAGAACATCACTCTTGGCCGGCTGGGGCCATTGGCTCGAAG  
Sequence 2 ATGGCACTCACCAACAACAACATCACTCTTGGCCGGCTGGGGCCATTGGCTCGAAG  
Sequence 1 ACCAGTCGGCAATGGTCAACGGGATCTCAAAATTGTCAACTTAAATTAATTTGGTT  
Sequence 2 ACCAGTCGGCAATGGTCAACGGGATCTCAAAATTGTGTGGGAAATCTTGAAGAAG  
Sequence 1 ATGTTTTCAGAAAATTAATTTACTTGTGTTTCCGCTGGTTACTGT  
Sequence 2 ATCCAAAGAAATTAATTTACTTGTGTTTCCGCTGGTTACTGT

**exon**

**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)

BIOS477/877 L3 - 43

43

```
>M8 (partial CDS)
D. melanogaster  atg  tgg  ttt  acg  ttg  acc  aac  aag  aag  gtg  att  tto  gtt  gac  ggt  ctg  gga  ggc  att  ggt
D. pseudoobscura  ---  ***  t--  c--  c--  ---  ---  ---  g--  ---  ---  ---  ---  ---  ---  ---  ---  ---
D. melanogaster  ctg  gac  aac  aag  aag  gag  ctg  ctg  aag  ggc  gat  ctg  aag  aac  ctg  gtc  ato  ctg  gao  ggc
D. pseudoobscura  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
D. melanogaster  att  gag  aac  cag  gct  gcc  att  ggc  gag  ctg  aag  gca  atc  aat  oca  aag  gtc  acc  gtc  acc
D. pseudoobscura  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
D. melanogaster  ttc  tac  ooc  tat  gat  gtc  aoc  gtc  ooc  att  gcc  gag  aoc  aoc  aag  ctg  gtc  aag  aac  atc
D. pseudoobscura  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
D. melanogaster  ttc  gcc  cag  ctg  aag  acc  gtc  gat  gtc  atc  aac  gga  gct  ggt  atc  ctg  gac  gat  cac
D. pseudoobscura  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---

(*: 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)

BIOS477/877 L3 - 44

44