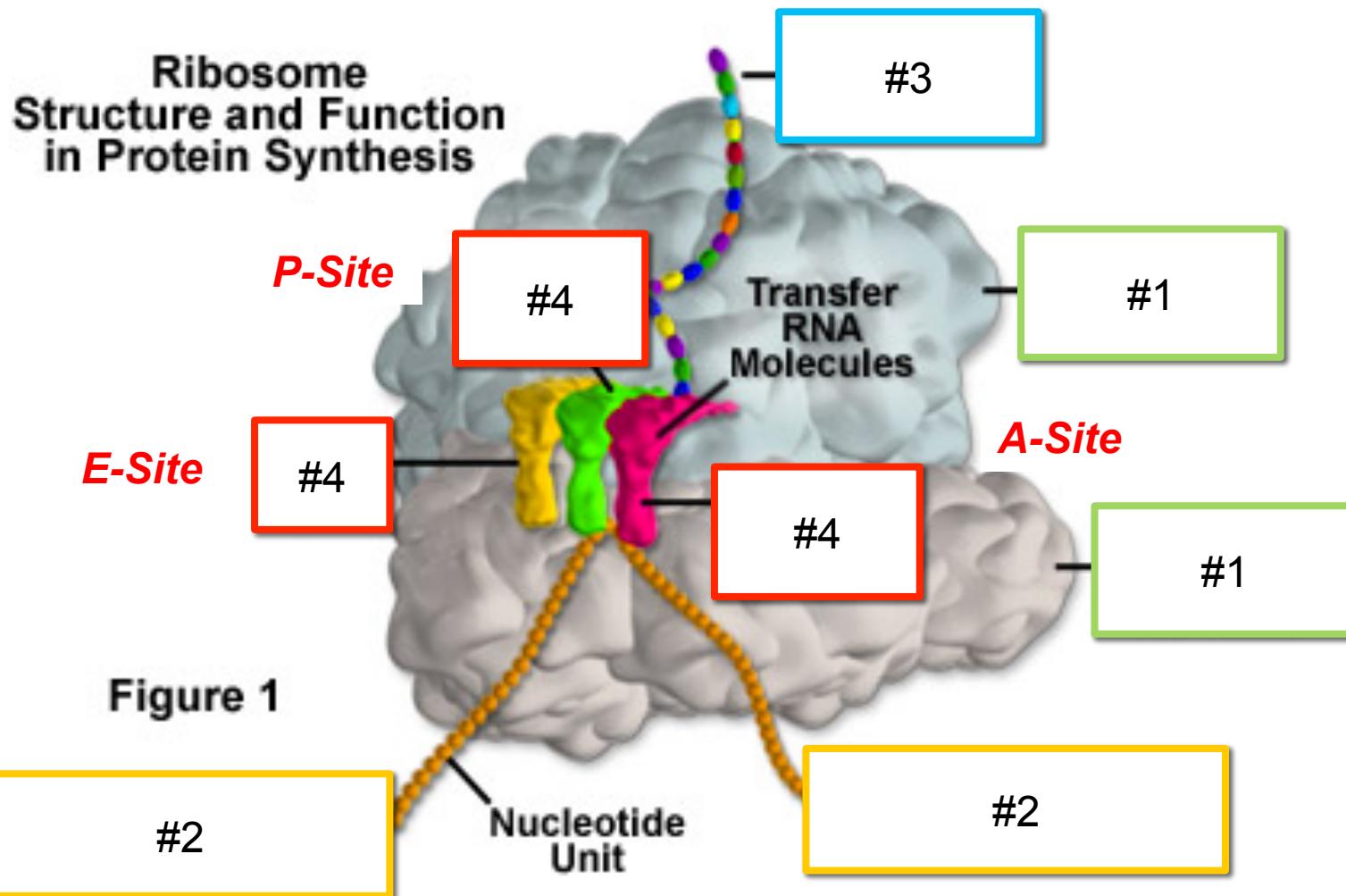


HW3 DETAILS & REPLICATION

Peptides are Translated by the Ribosome



HW3: Programming the Central Dogma

- A DNA sequence (a string).
- The *strand* the gene appears on (a string). If the string is '+', then the DNA is from the positive strand. If the string is '-', then the DNA is from the reverse complement (negative) strand.
- A list of exon start coordinates (a list of integers).
- A list of intron start coordinates (a list of integers).

*Extra print statements
are encouraged!*

Dataset: testPos

DNA: CCCATGGTCGGGGGGGGGGAGTCCATAACCC

Num
str
RNA
pep

Lastly...make your program
work for genes on the
negative strand ("")

Computed mRNA: AUGGUCAGGUCCAUAA
mRNA matches!

Computed Peptide: MVSP*
Peptide matches!

Compare to sequences

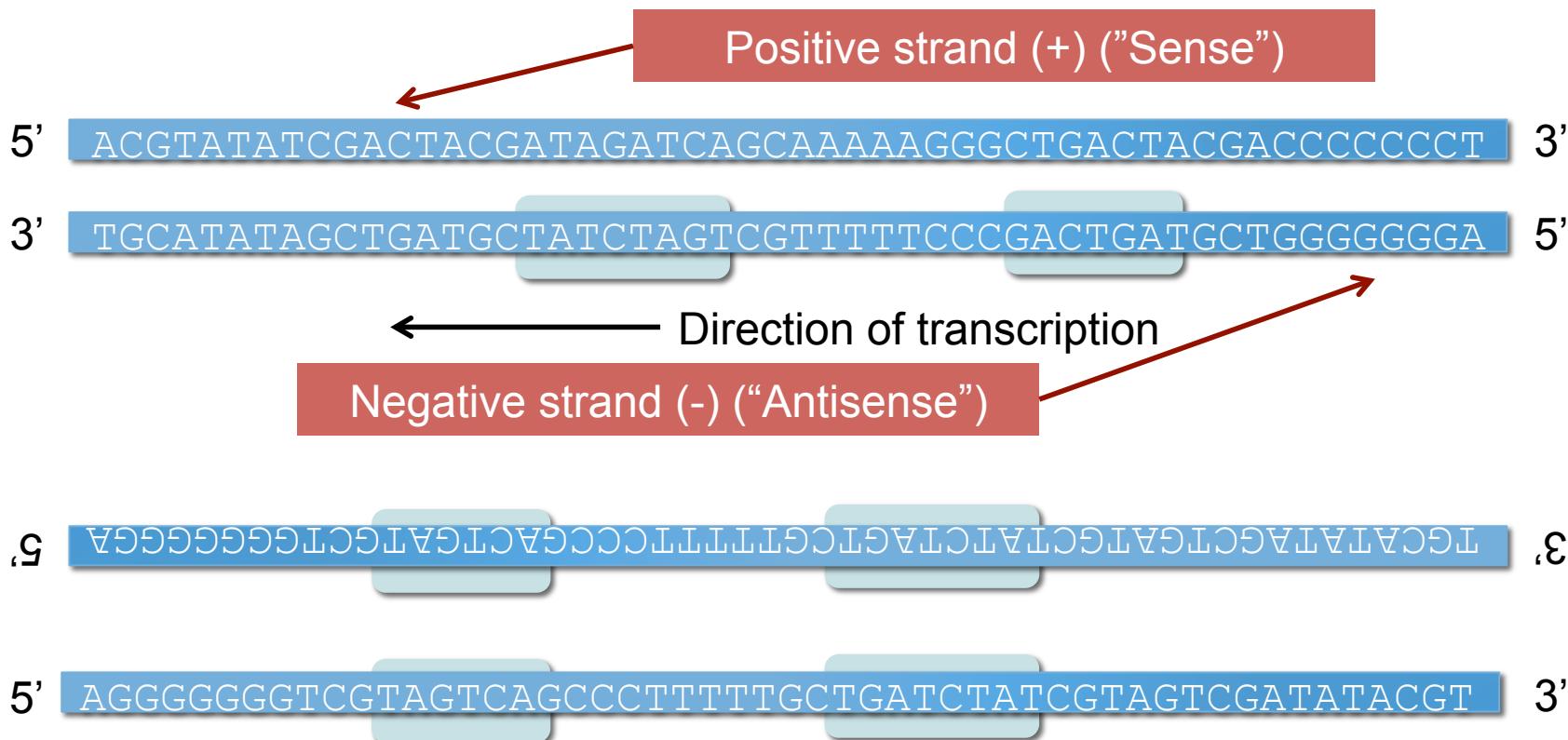
Read and print Information
about the Input

transcribe() function (returns mRNA)

translate() function (returns peptide)

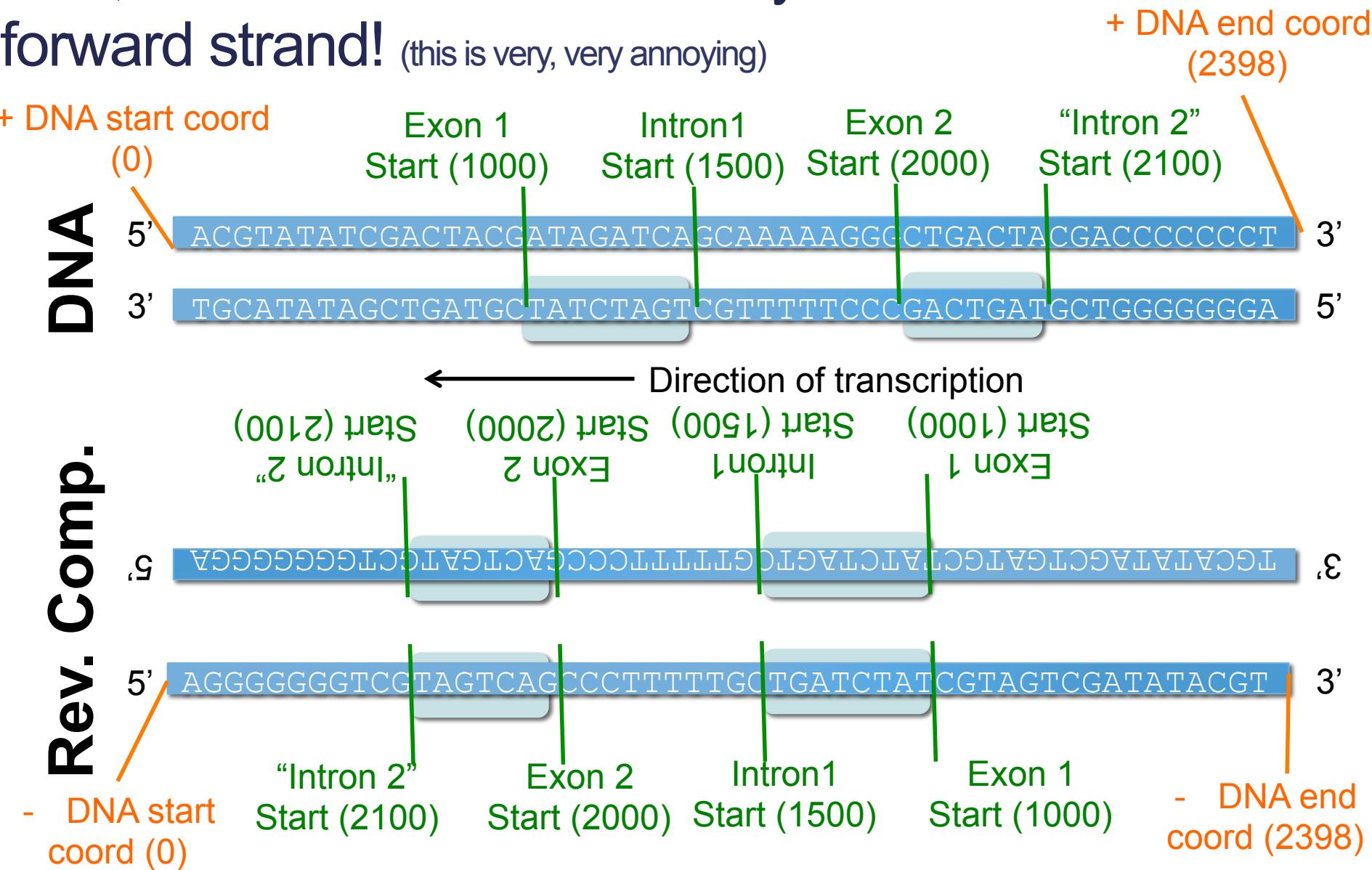
The Reverse Complement Strand

Rev. Comp. DNA

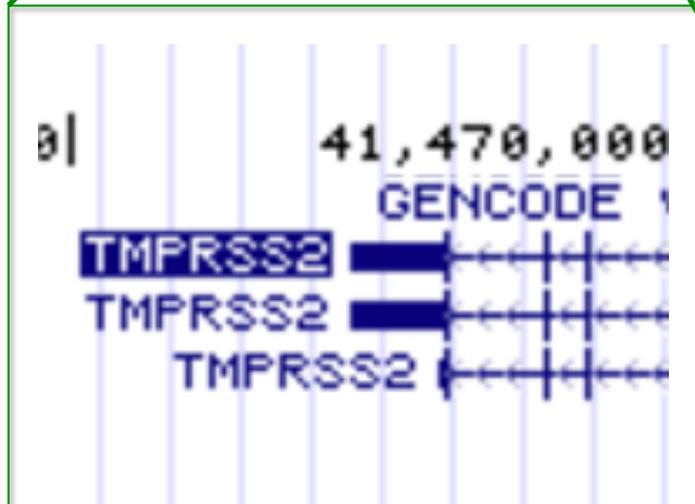
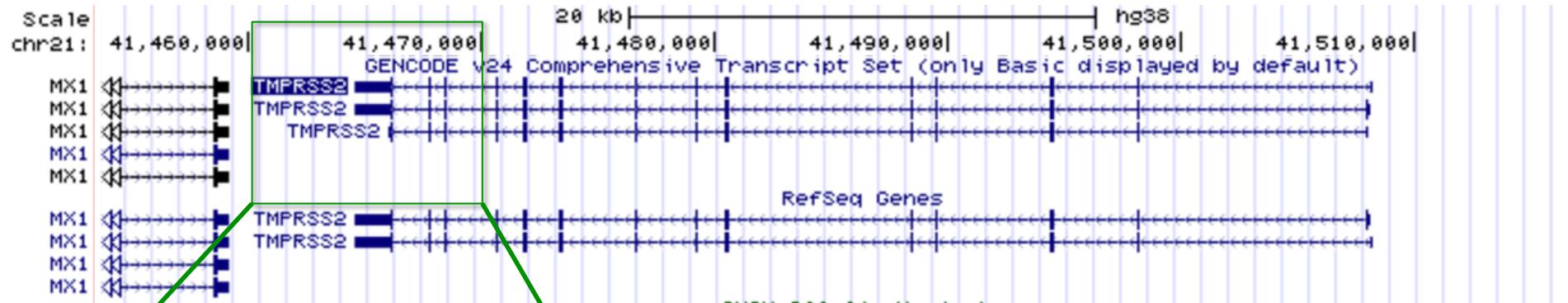


Sequences are always reported
in the 5' to 3' direction.

BUT, the coordinates are always relative to the forward strand! (this is very, very annoying)



The Data: TMPRSS2 Gene

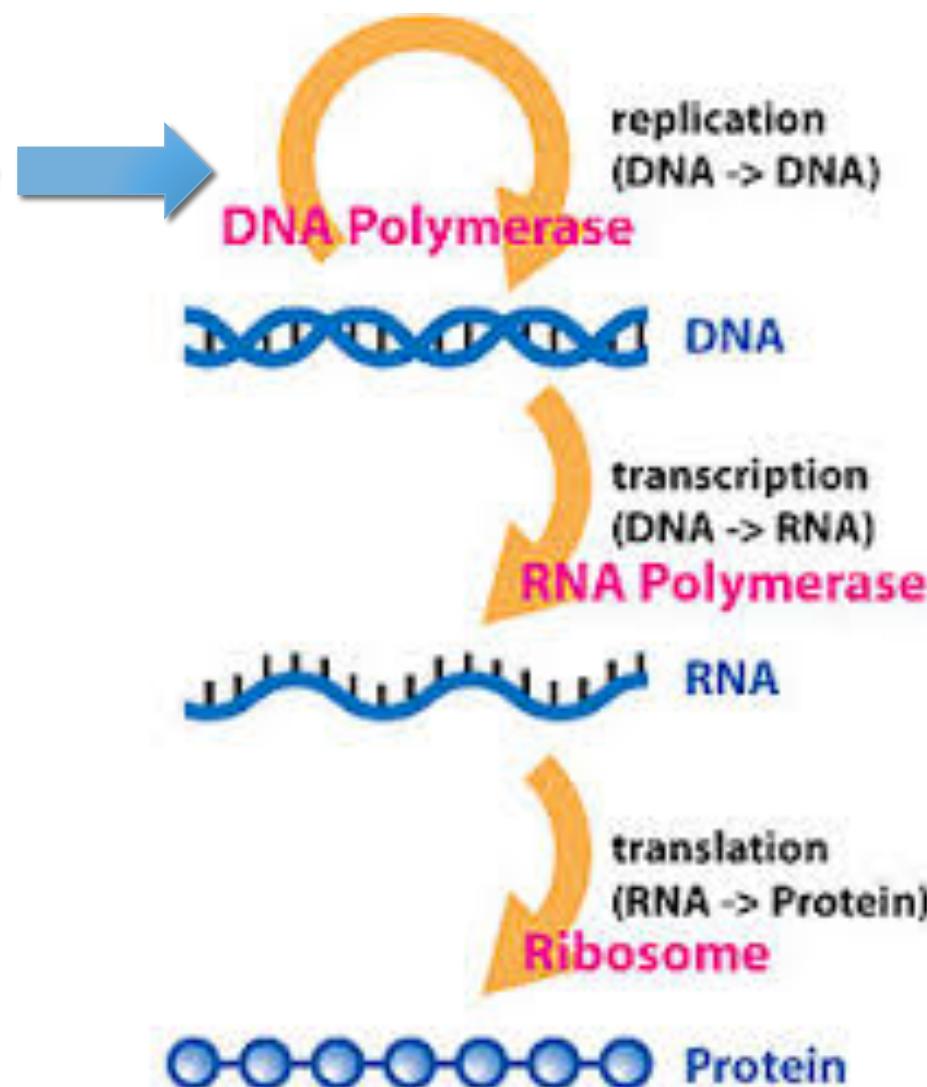


Data From Genome Browser

- DNA string (slice of chromosome)
- Strand = '-'
- Exon/Intron Start Positions shifted to index into DNA string ***on the positive strand.***

Switching Gears: DNA Replication

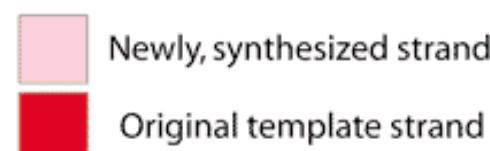
Start Here



Three Models of DNA Replication

How?
**Meselson-Stahl Experiment
(1958)**

Three postulated methods of DNA Replication

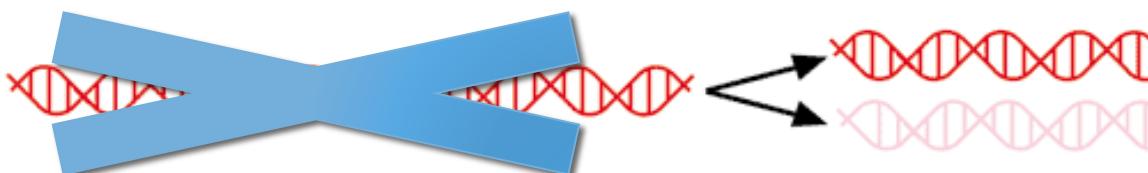


Three Models of DNA Replication

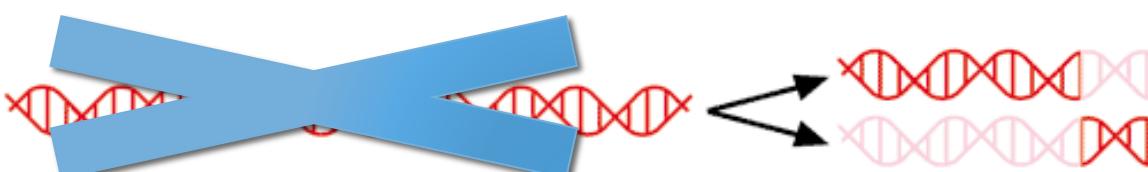
Three postulated methods of DNA Replication



Semi-Conservative



Conservative



Dispersive

Lighter Nitrogen (^{14}N)

Heavier Nitrogen (^{15}N)



Newly synthesized strand

Original template strand

How?

Meselson-Stahl Experiment (1958)

Idea: make *E. Coli* that have DNA with ^{15}N , then transfer them to a ^{14}N medium.

After **one round** of replication: both strands had the same density

After **two rounds** of replication: DNA had two different densities

Three Models of DNA Replication

Three postulated methods of DNA Replication



Semi-Conservative



Conservative*



Dispersive*



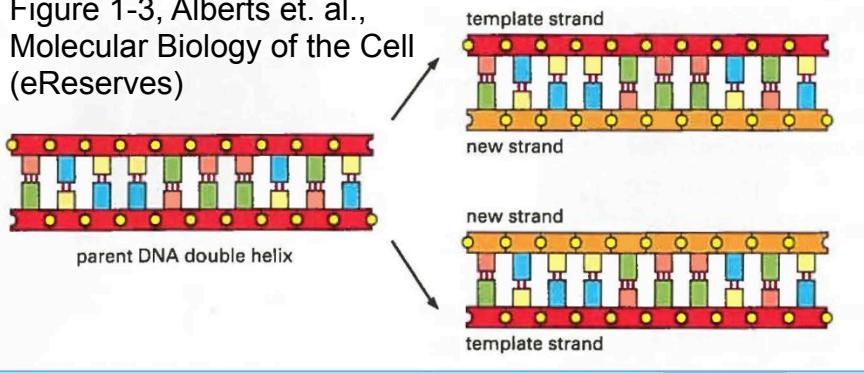
Newly, synthesized strand

Original template strand

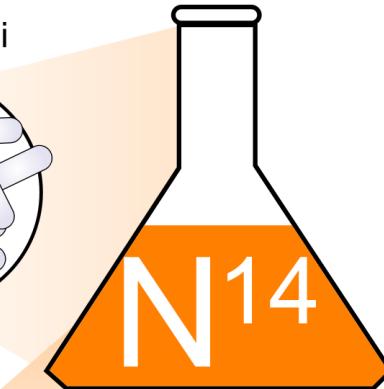
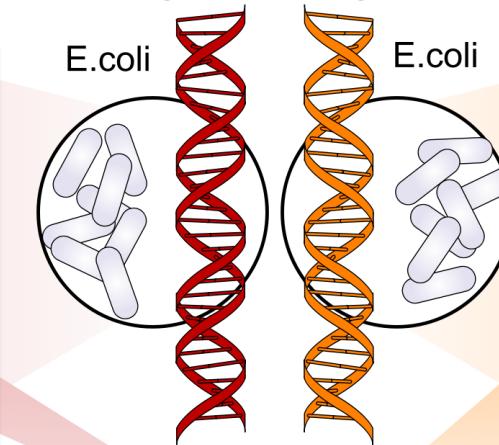
* not found to be
biologically significant

DNA Replication is Semiconservative

Figure 1-3, Alberts et. al.,
Molecular Biology of the Cell
(eReserves)

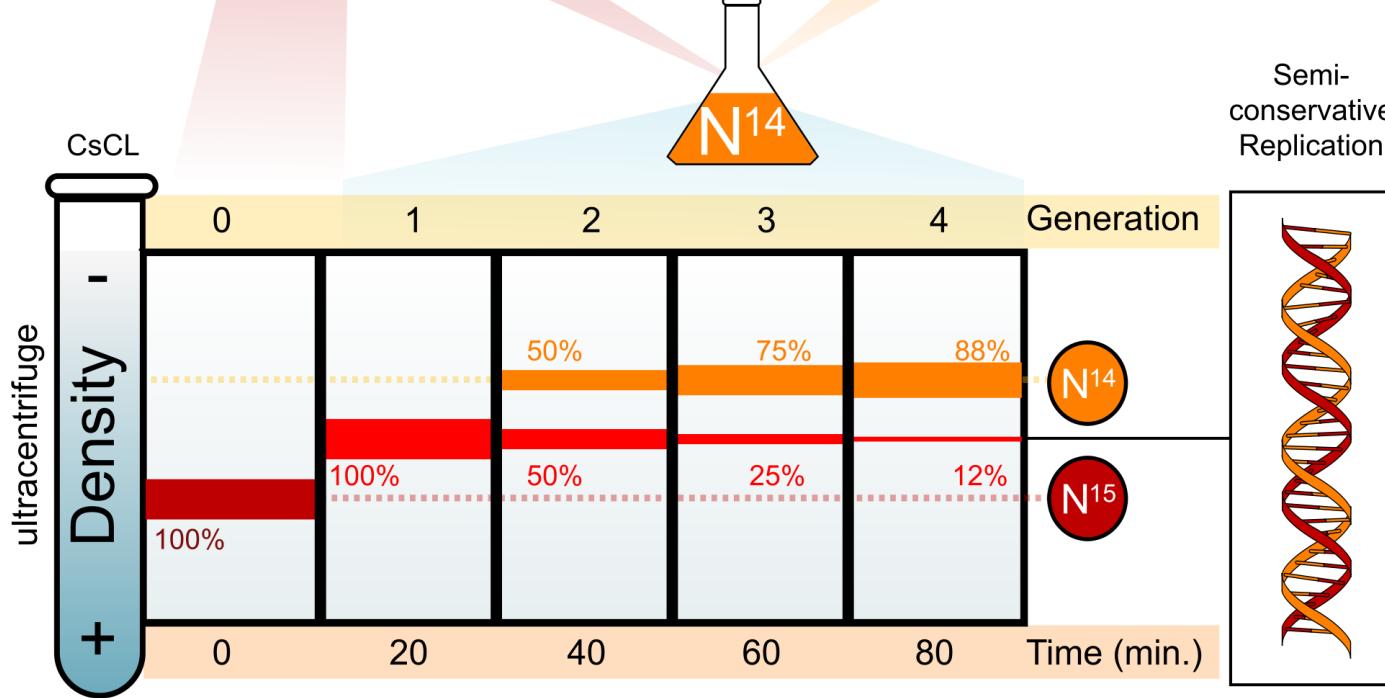


14 generations of growth



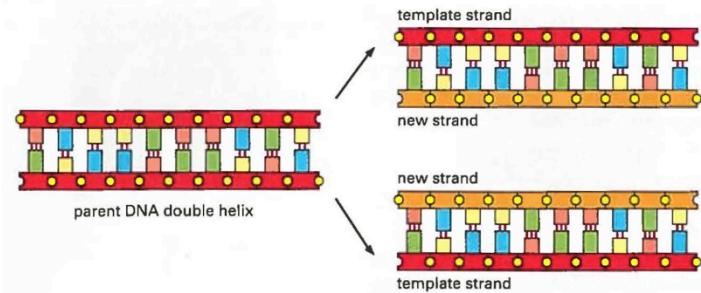
N14

Semi-conservative
Replication



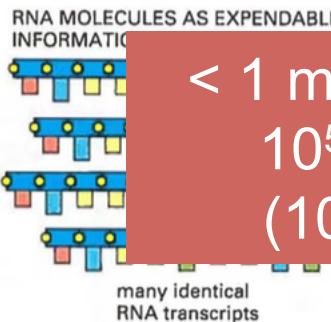
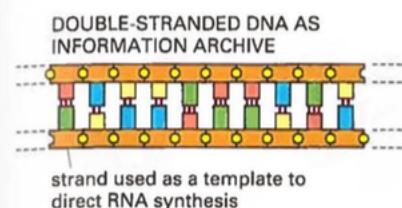
"Meselson-stahl experiment diagram en" by LadyofHats - did myself based on the information in wikipedia plus the following websites:[1], [2], [3], [4], [5] and [6]. Licensed under Public Domain via Commons - https://commons.wikimedia.org/wiki/File:Meselson-stahl_experiment_diagram_en.svg#media/File:Meselson-stahl_experiment_diagram_en.svg

Machinery Errors



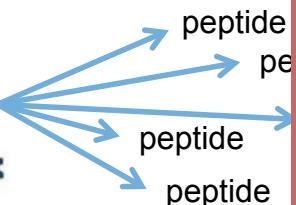
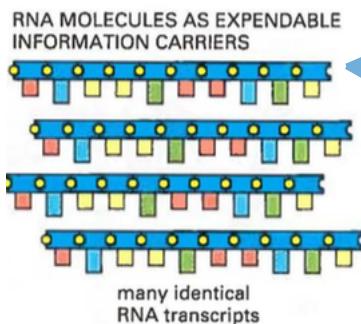
< 1 mistake per
 10^7 bases
(10,000,000)

Proofreading
Ability to
Correct Errors



< 1 mistake per
 10^5 bases
(100,000)

Proofreading
Ability to
Correct Errors



< 1 mistake per
 10^4 bases
(10,000)

Proofreading
Ability to
Correct Errors



The Textbook

Phillip Compeau and Pavel Pevzner

Bioinformatics Algorithms: an Active Learning Approach

©2013 by Compeau and Pevzner. All rights reserved

(Slides kindly provided by Phillip Compeau)

UC San Diego
Department of Computer Science & Engineering

Book from a MOOC (Massive Open Online Course)

YouTube Channel:

<https://www.youtube.com/user/bioinfalgorithms>

Website with slides:

<http://bioinformaticsalgorithms.com/index.htm>

Web Interface to Submit Problems:

<http://rosalind.info/problems/locations/>

Book is designed for **computer scientists** with
little/no background in **biology**.



Pevzner



Compeau

A New Idea: Special “Conference”

- Opt-In (email soon)
- One hour every other week or so to discuss computer science concepts
- These concepts can be applied to extra problems on the homeworks
- You must commit to attending the meetings (you don't have to do the extra problems).

Upcoming Schedule

Week 3	
Mon 2/6	Lecture, HW2 Due, HW3 Out
Mon/Tues Lab: Reverse Complement Function	
Tue 2/7 12p-1p: <i>ES-Bio Faculty Candidate Talk</i>	
Elaine Dojo Hours (7-9pm Tues)	Lecture
Wed 2/8 12p-1p: <i>ES-Bio Faculty Candidate Talk</i>	
Fri 2/10	Lecture, Internal HW3.1 Deadline
Rose Dojo Hours (8-10pm Thurs)	
Sun 2/12	Lecture, Internal HW3.2 Deadline
Mon 2/13 Lab: Rosalind Tutorial	
Anna Office Hours 11a—12pm Wed/Thurs	Lecture
Tue 2/14	Lecture, HW3 Due

HW1 Back
(Email in
next hour)

Obtain
Textbook

Function Practice

- Download 2017_02_10_worksheet.py from Moodle

Create a variable `dna` with a DNA string (e.g., `'ACTGACTACGATC'`). Write the following functions:

`proportion()`: *Compute the proportion of A's, C's, G's, and T's in a string*

1. Inputs: a string of A/C/G/T.
Returns: four floats.

You can return multiple variables by separating them with commas, e.g.,

```
return numAs, numCs, numGs, numTs
```

If you need some guidance, here's how I break down this problem. Work on each part sequentially.

- Write a `for` loop that prints the A's in the string `dna`.
- Modify the loop to count the number of A's and print the result.
- Count the number of C's, G's, and T's and print the results.
- Convert the counts to fractions by dividing by the length of the string `dna`.