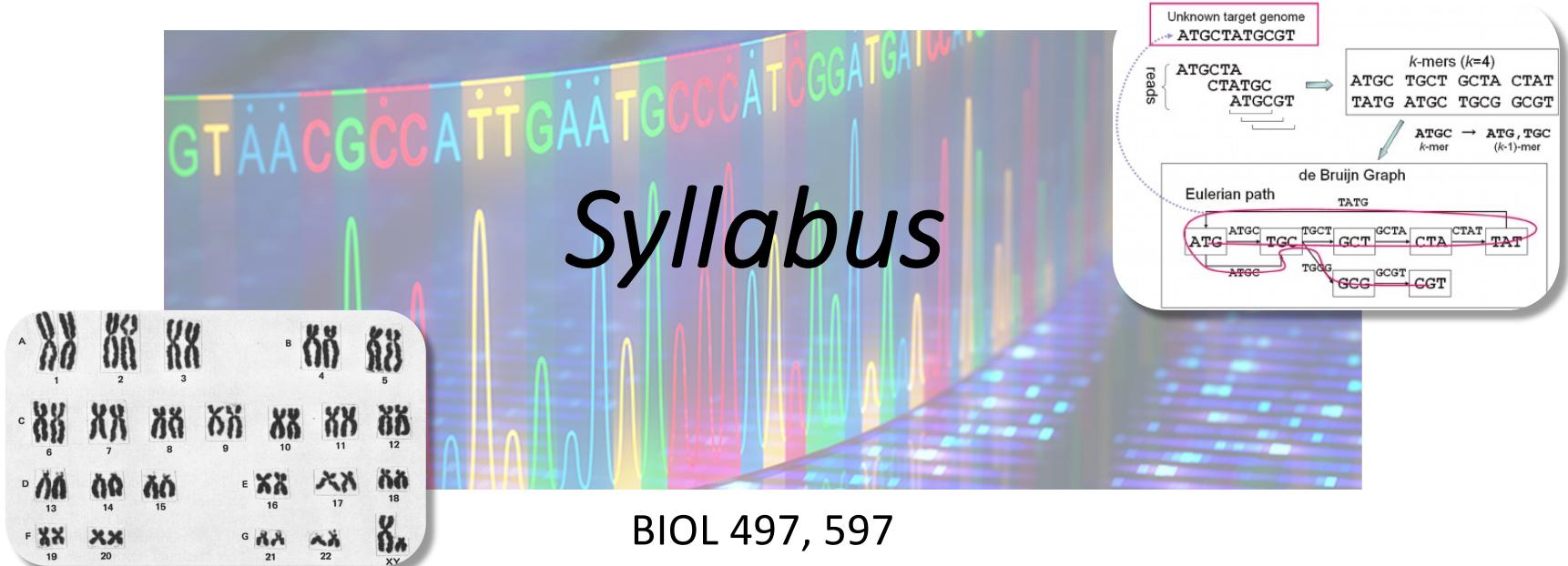


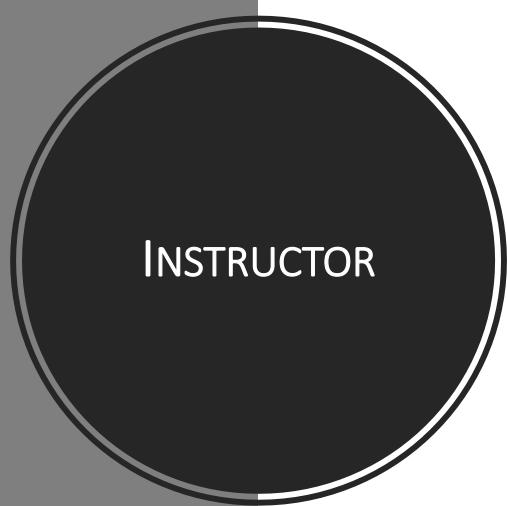
# Genomics & Bioinformatics



BIOL 497, 597

Boise State University

Spring 2022



INSTRUCTOR

- **Name:** Sven BUERKI
- **Office:** Science Building, office 114
- **Email:** [svenbuerki@boisestate.edu](mailto:svenbuerki@boisestate.edu)



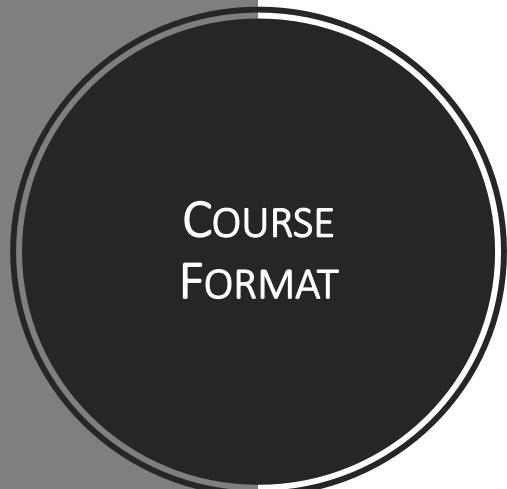
CLASS  
LOCATION &  
MEETING TIMES

- **Location:** SCNC149 computer room.
- **Lectures:** Wednesdays from 9:00-10:50 AM.
- **Labs:** Fridays from 9:00-10:30 AM.
- No office hours, but please contact me if you want to set an appointment.

## COURSE GOAL & DESCRIPTION

The goal is to provide students with the theoretical and applied knowledge in genomics and bioinformatics to sequence, assemble and annotate genomes, especially for non-model organisms.

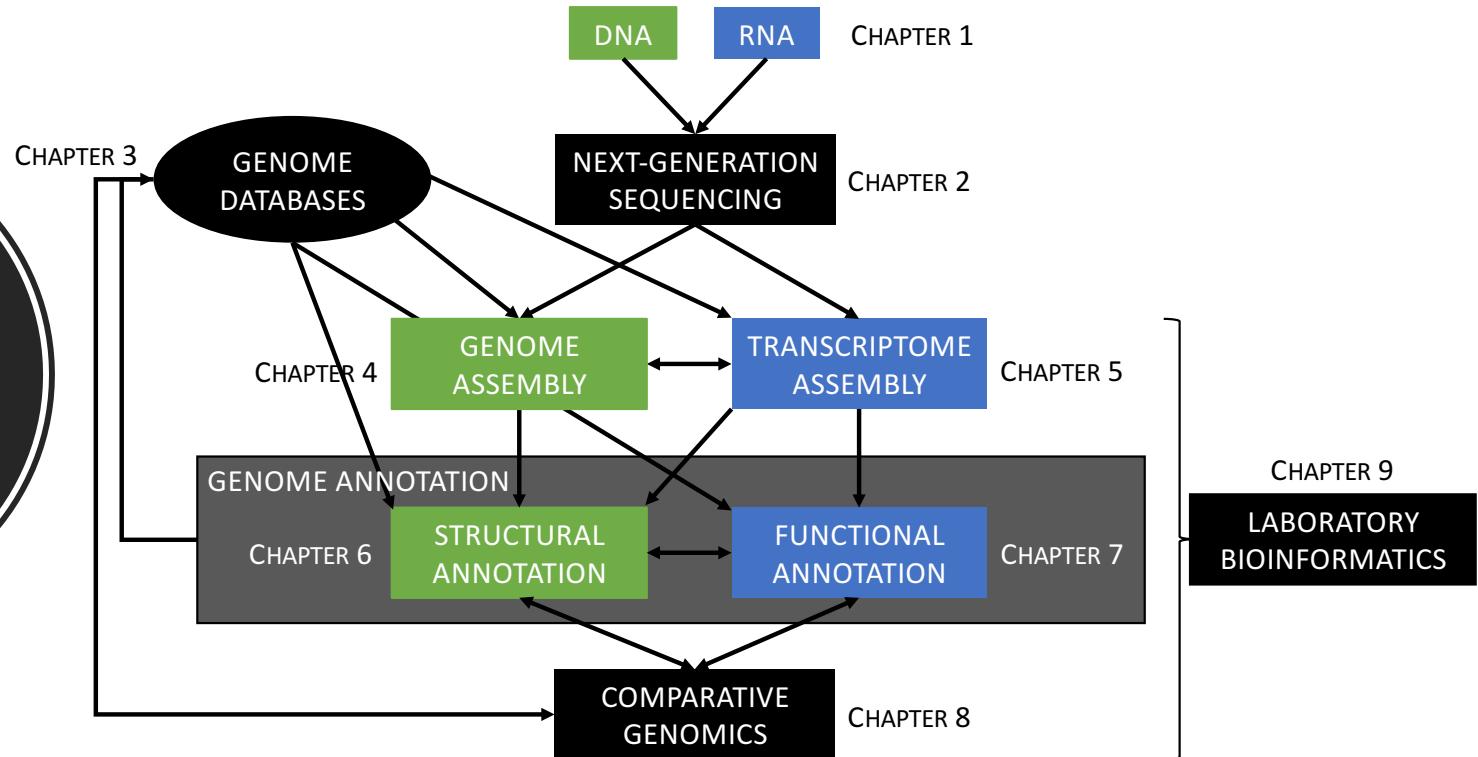




## COURSE FORMAT

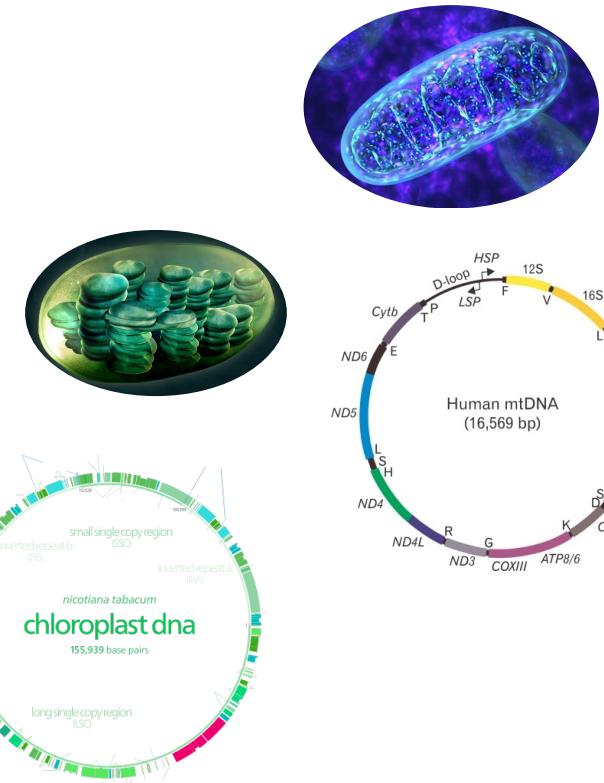
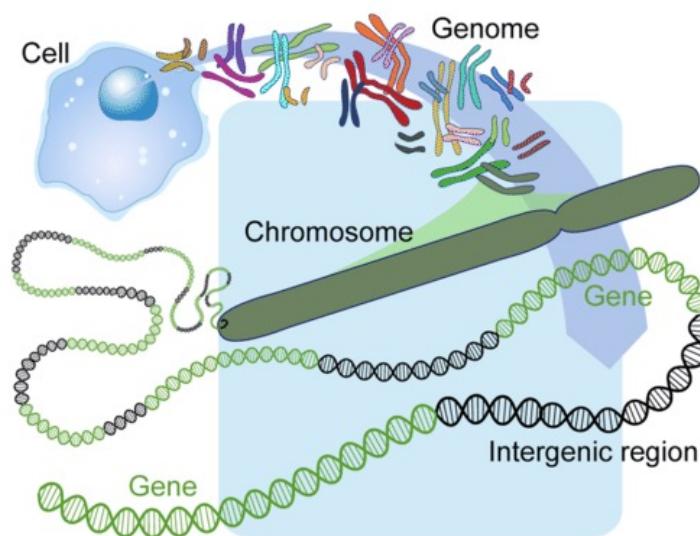
- This class provides a mixture of **classic lectures** together with more applied, **hands-on bioinformatic tutorials**.
- Tutorials are designed to support students in mastering theoretical genomic concepts through their implementations in bioinformatic protocols.
- Genomic field heavily relies on bioinformatic expertise, especially related to unix-based software. We will be working on computers running the **Linux operating system**. These computers will provide opportunities for students to become familiar with the **bash/shell**, **Python** and **R** computing languages.
- Lecture sessions will also serve as a platform to
  - ❖ Work on mini-reports.
  - ❖ Study genomic literature through a journal club.

# COURSE LEARNING OUTCOMES



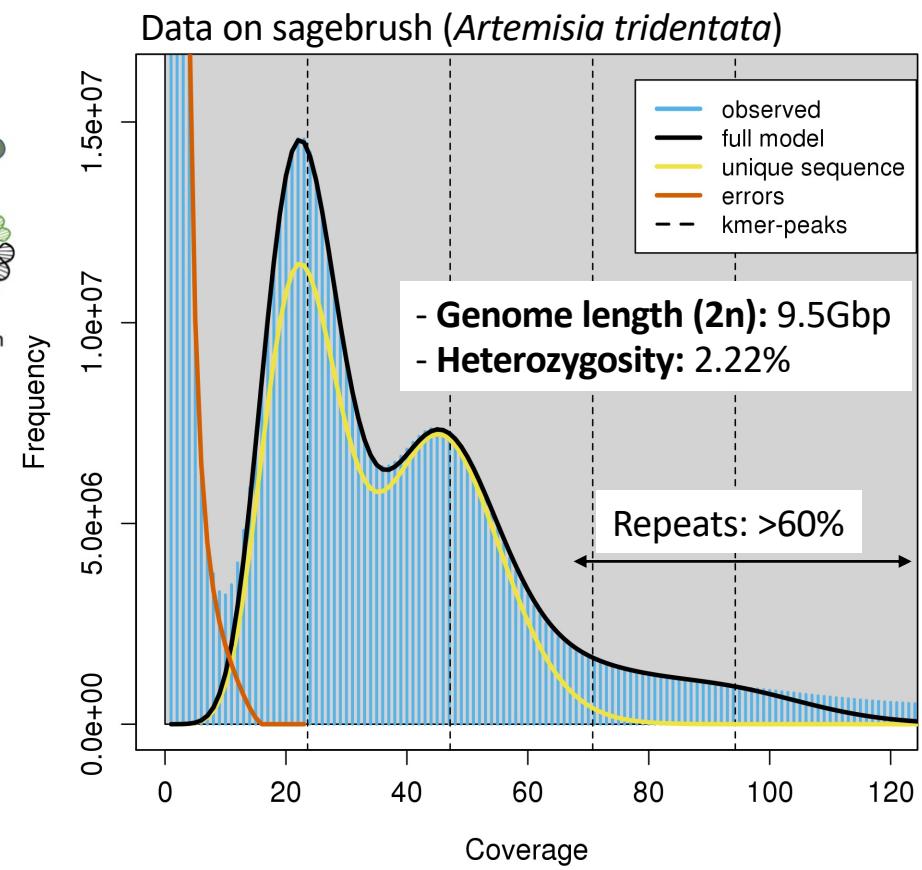
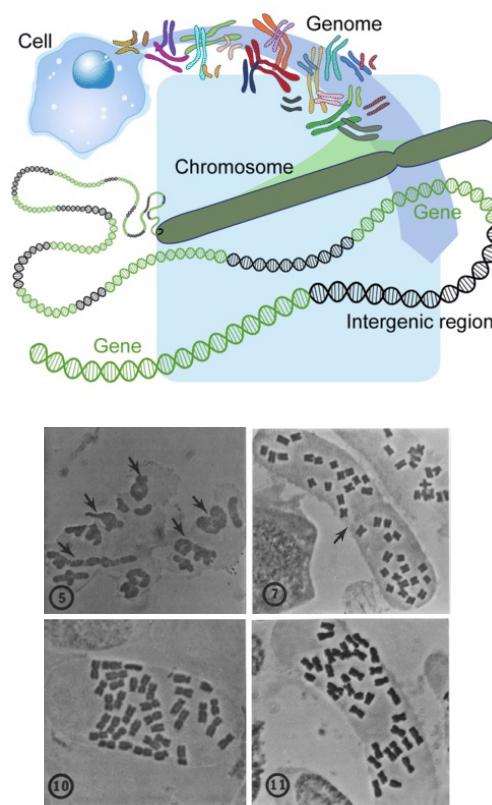
# CHAPTER 1

Moving our mindset from the study of single genes (genetics) to the study of the entire genetic material in a cell (genomics).



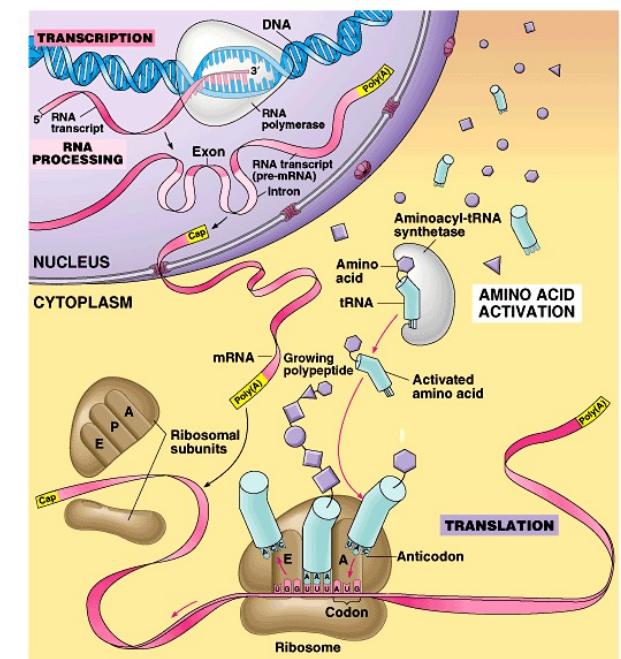
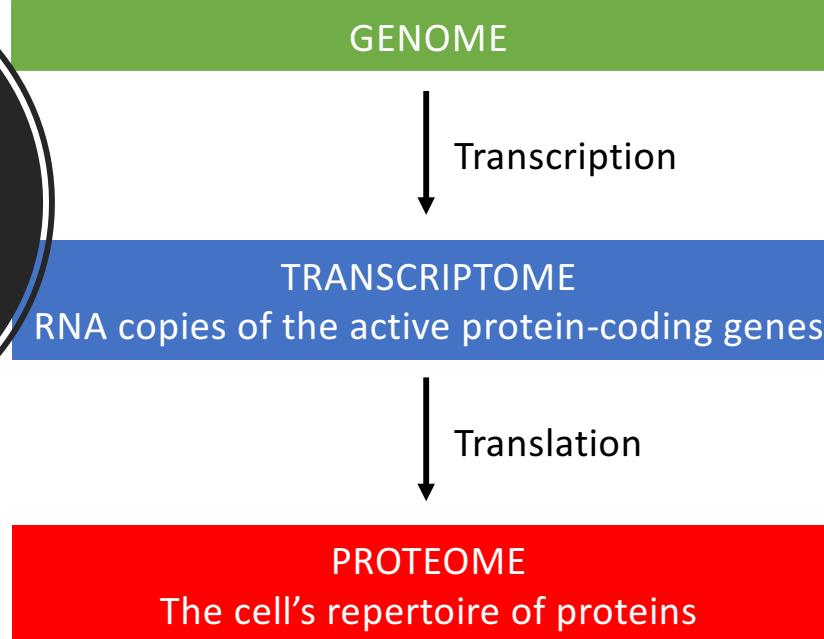
# CHAPTER 1

Appreciate that **eukaryotic genomes contain extensive repetitive regions** of several different kinds. This provides a challenge for genome assembly!



# CHAPTER 1

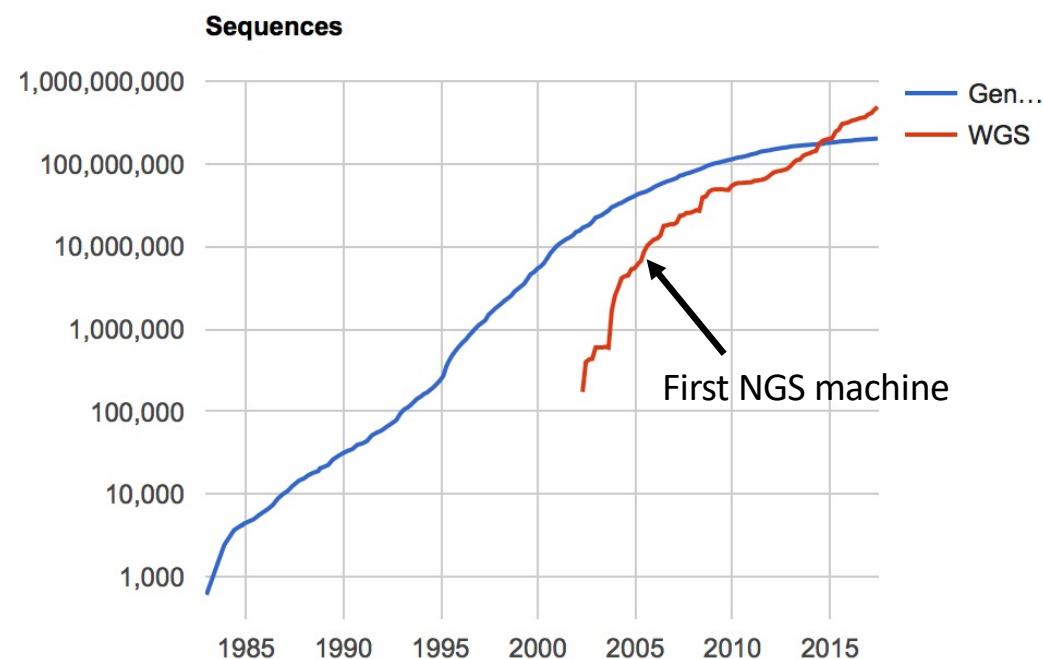
Know the basic dogma that **DNA** is transcribed to **RNA**, which is translated to **protein**.



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

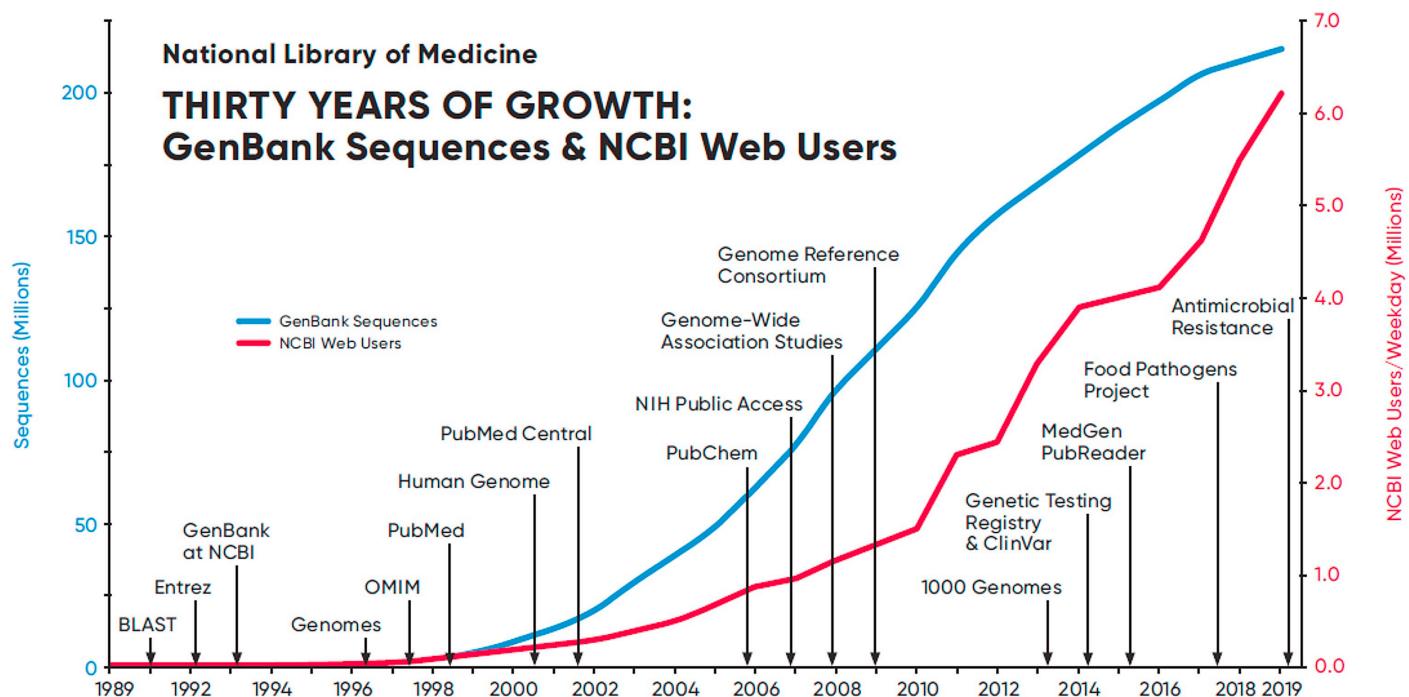
# CHAPTER 1

Perfect moment to study genomics, a plethora of data have been generated in the last 15 years.



# CHAPTER 1

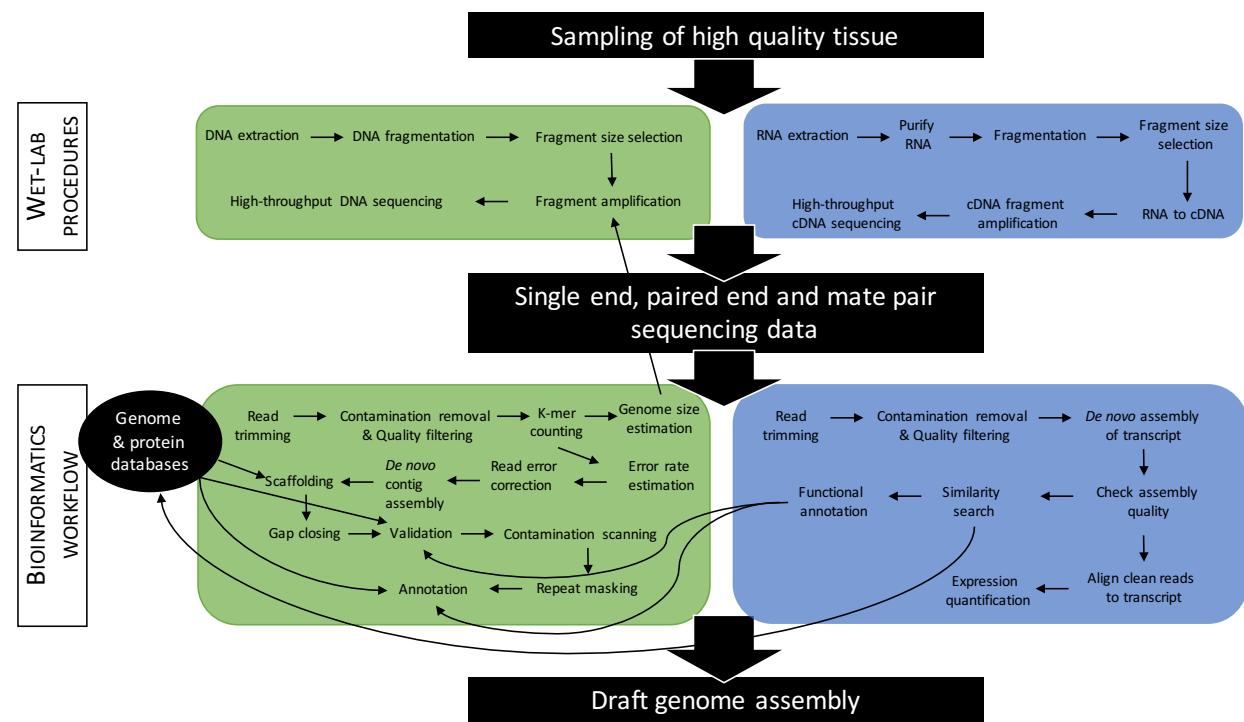
Perfect moment to study genomics, a plethora of data have been generated in the last 15 years.



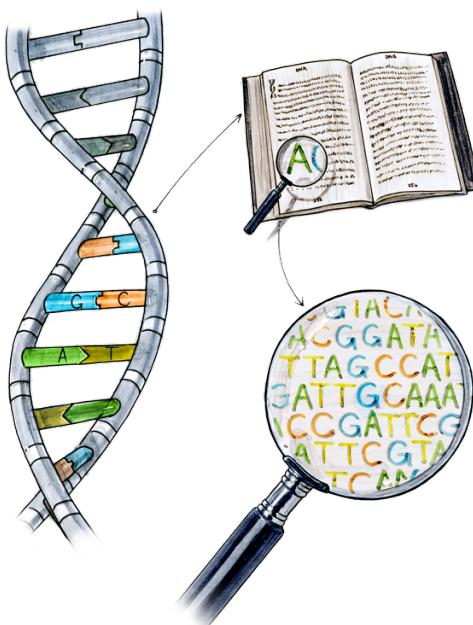
<https://doi.org/10.1016/j.gfs.2020.100411>

# CHAPTER 2

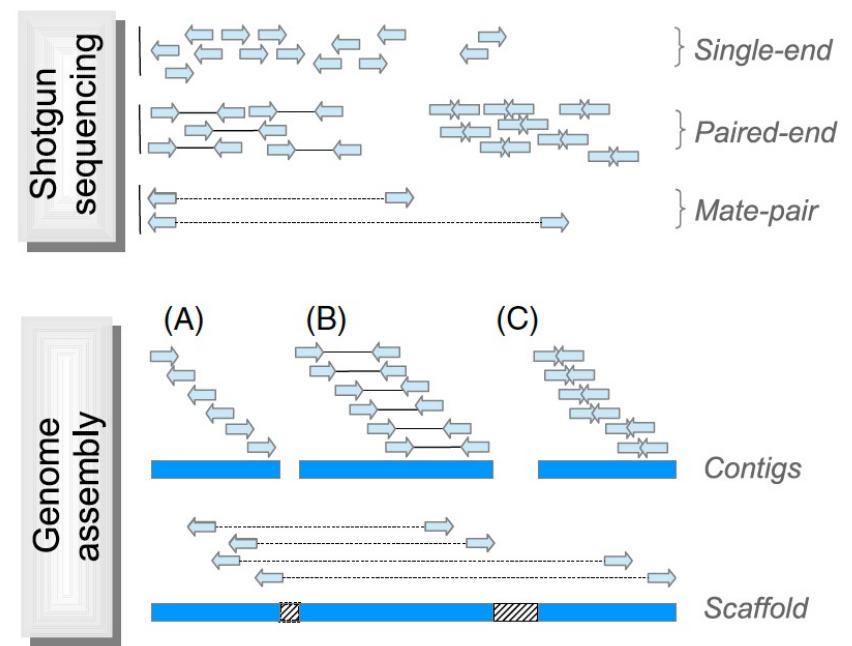
Independently on the approach to produce a whole-genome sequence, all projects share the same major steps:



## CHAPTER 2



Learn the next-generation sequencing (NGS) jargon necessary to assemble and annotate genomes.

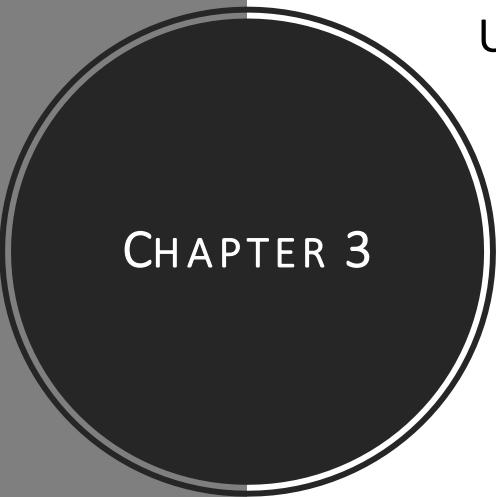


Subject of Mini-report 1

## CHAPTER 2

### Introduction to NGS platforms:





## CHAPTER 3

Understand importance of computer science in:

- Producing raw sequence data (e.g. base-calling),
- Creating databases in molecular biology,
- Archiving and curation of data,
- Distributing data via the Internet,
- Creating information-retrieval tools to allow effective mining of the data for research application.

## CHAPTER 3

Gain knowledge on major molecular biology databases, which are key to genome assembly and annotation:

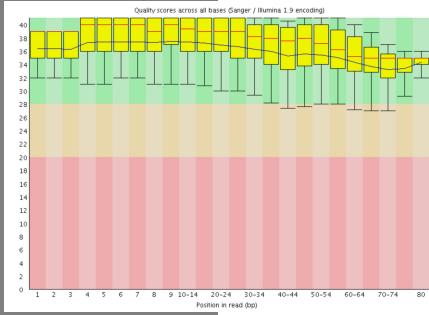
- Nucleic acid sequences databases (e.g. NCBI).
- Protein sequences databases (e.g. Swissprot).
- Gene ontology databases.
- Metabolic pathways databases (e.g. KEGG).
- Specialized annotated genome portals.



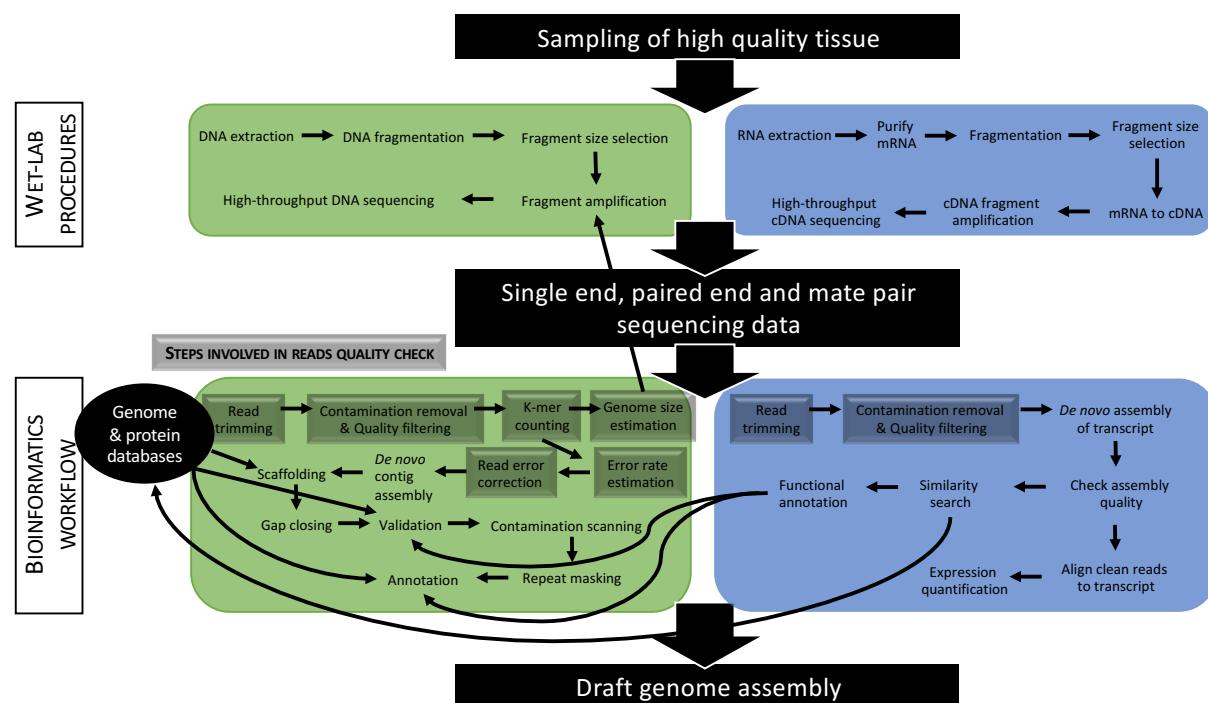
**NCBI**  
National Center for  
Biotechnology Information



# CHAPTER 4

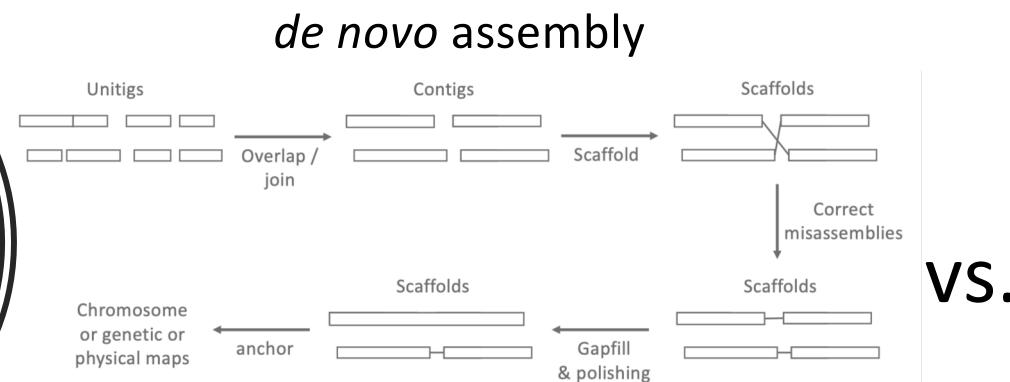


Learn how to conduct reads quality checks on raw NGS data.



# CHAPTER 4

Understand key steps involved in producing a genome assembly and best strategies to get there.



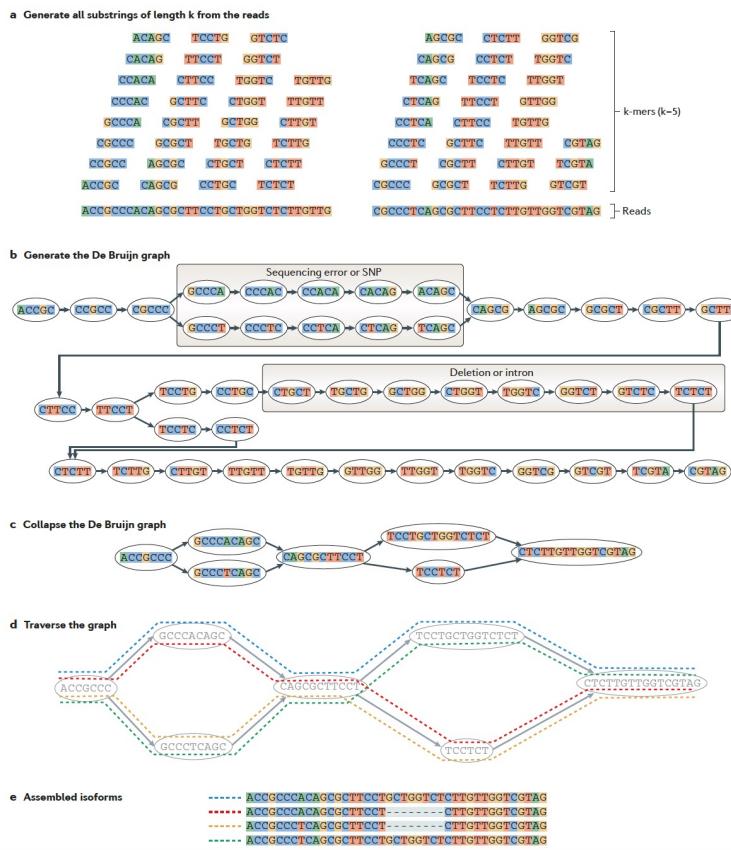
VS.

*Referenced-based assembly*



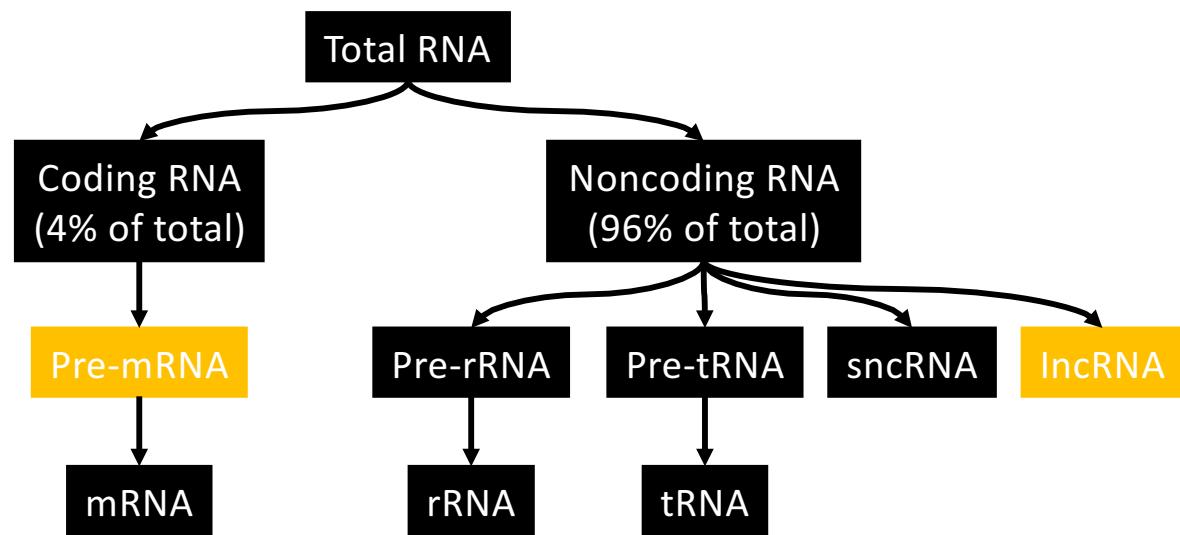
# CHAPTER 4

Study de Bruijn graph procedure for *de novo* genome assembly.



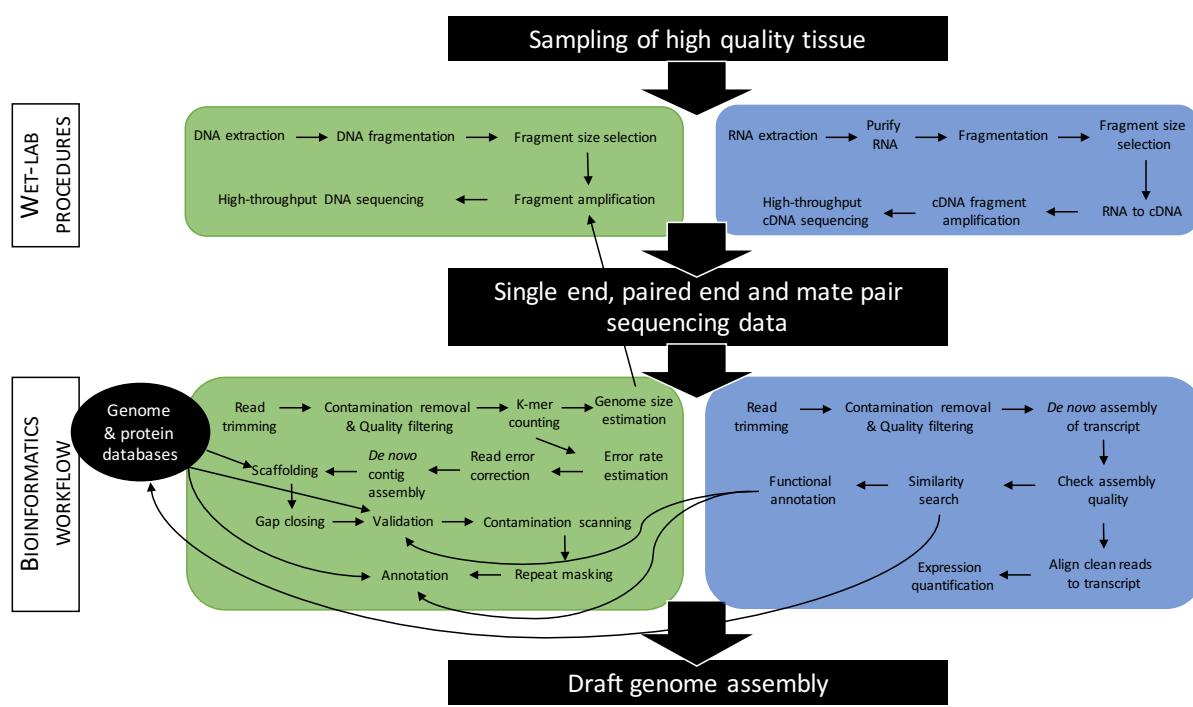
# CHAPTER 5

Learn about the different types of RNA molecules in cell.



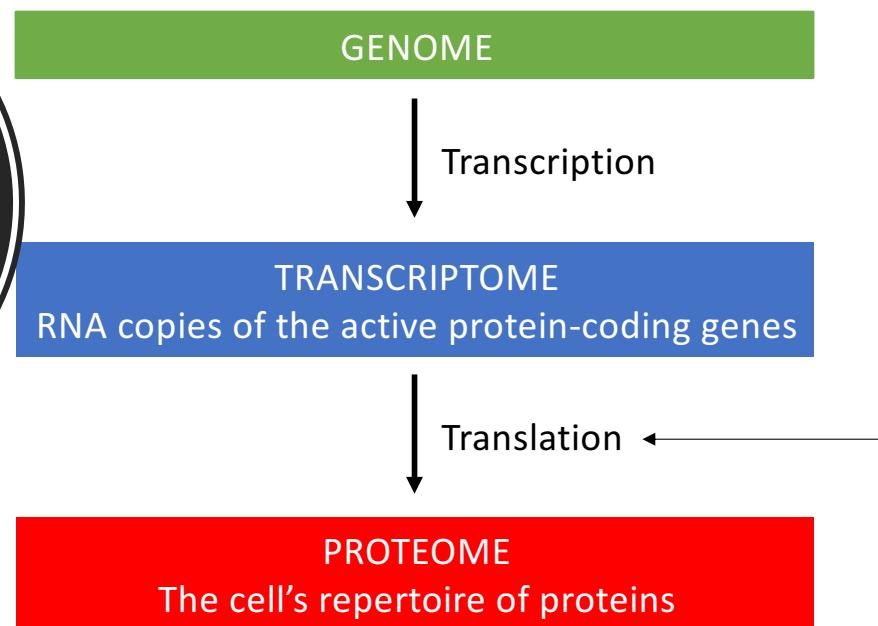
# CHAPTER 5

Adapt wet-lab and bioinformatics workflow according to targeted RNAs.



# CHAPTER 5

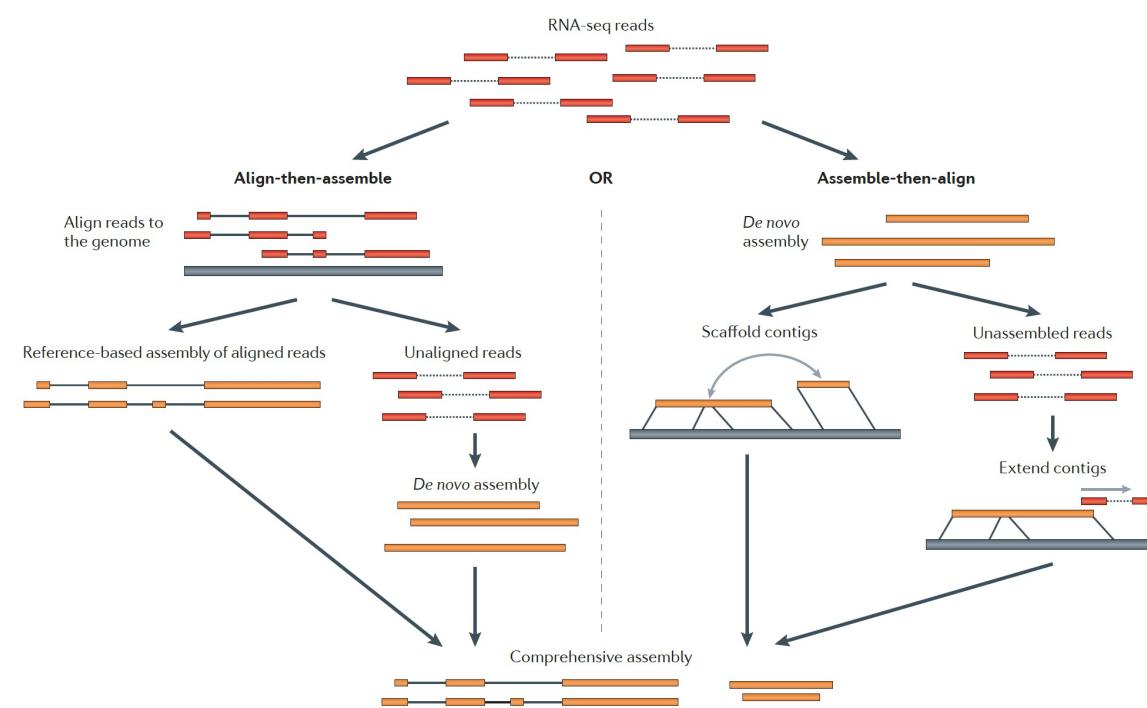
Study the link between the transcriptome and proteome via the genetic code.

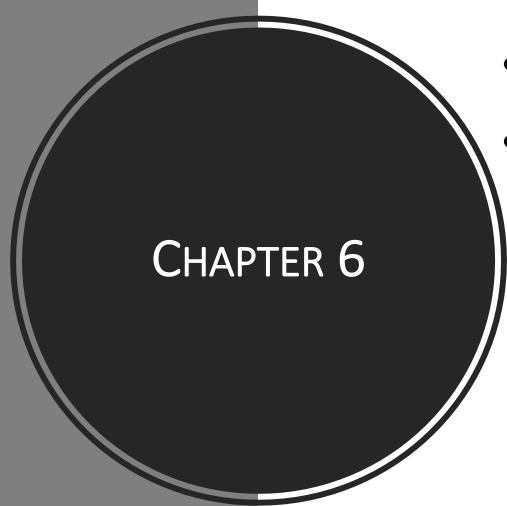


UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC		UCC		UAC		UGC	
UUA	Leu	UCA		UAA	Stop	UGA	Stop
UUG		UCG		UAG		UGG	Trp
CUU		CCU		CAU	His	CGU	
CUC	Leu	CCC		CAC		CGC	
CUA		CCA		CAA	Gln	CGA	
CUG		CCG		CAG		CGG	Arg
AUU		ACU		AAU	Asn	AGU	
AUC	Ile	ACC		AAC		AGC	Ser
AUA		ACA		AAA		AGA	
AUG	Met	ACG		AAG	Lys	AGG	Arg
GUU		GCU		GAU	Asp	GGU	
GUC		GCC		GAC		GGC	
GUA	Val	GCA		GAA	Glu	GGG	Gly
GUG		GCG		GAG			

# CHAPTER 5

Understand key steps involved in **producing a transcriptome assembly** and best strategies to get there.





## CHAPTER 6

- Learn what we exactly mean by the term "genome annotation".
- Know the following key stages of the structural genome annotation process:
  - Repeat identification.
  - Evidence alignment (map transcriptome on genome).
  - *Ab initio* ("from the beginning") gene prediction.
  - Evidence-driven gene prediction (use external info to improve prediction of gene annotations).

## CHAPTER 7

Be aware of challenges to obtain accurate data on gene functions.

Research article | Open Access

# Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa* L.)

Hiroshi Mizuno <sup>†</sup>, Yoshihiro Kawahara <sup>†</sup>, Hiroaki Sakai, Hiroyuki Kanamori, Hironobu Wakimoto, Harumi Yamagata, Youko Oono, Jianzhong Wu, Hiroshi Ikawa, Takeshi Itoh and Takashi Matsumoto 

<sup>†</sup>Contributed equally

BMC Genomics 2010 11:683

<https://doi.org/10.1186/1471-2164-11-683> |

Received: 20 April 2010 | Accepted: 2 Decem-

**A study on the response of rice to salt stress discovered 649 genes that were missing from the rice annotation!**

## CHAPTER 7

Study and compare available pipelines to conduct automated genome annotations.



### Last Software Update

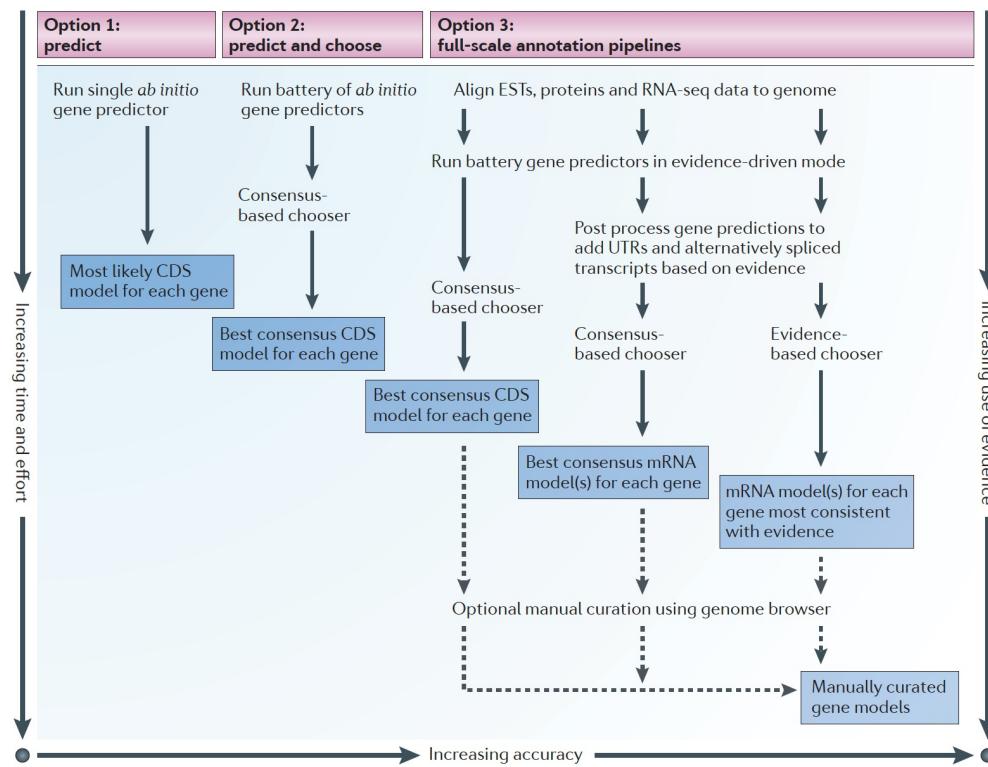
v2.31.9 (Dec 16, 2016)

### Overview

MAKER is a portable and easily configurable genome annotation pipeline. Its purpose is to allow smaller eukaryotic and prokaryotic genome projects to independently annotate their genomes and to create genome databases. MAKER identifies repeats, aligns ESTs and proteins to a genome, produces ab-initio gene predictions and automatically synthesizes these data into gene annotations having evidence-based quality values. MAKER is also easily trainable: outputs of preliminary runs can be used to automatically retrain its gene prediction algorithm, producing higher quality gene-models on subsequent runs. MAKER's inputs are minimal and its outputs can be directly loaded into a GMOD database. They can also be viewed in the Apollo genome browser; this feature of MAKER provides an easy means to annotate, view and edit individual contigs and BACs without the overhead of a database. MAKER should prove especially useful for emerging model organism projects with minimal bioinformatics expertise and computer resources.

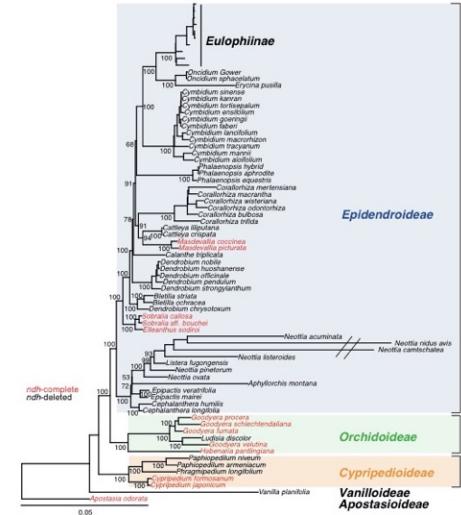
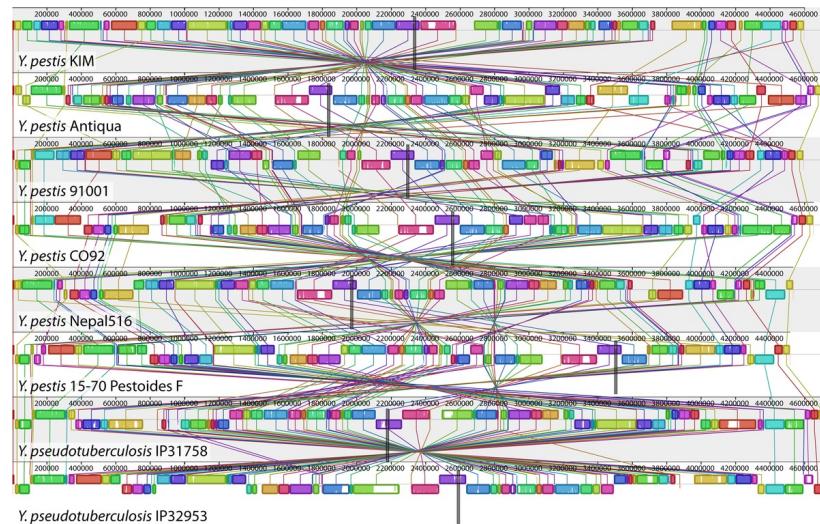
# CHAPTER 7

## Review approaches to assess annotation quality.



# CHAPTER 8

- Learn to read and present a scientific paper reporting new genomic or transcriptomic data.
- Become an expert on a specific area of comparative genomics and share it with your peers.



CLASS  
PROMOTES  
OPEN-SOURCE  
RESOURCES

ubuntu 

Google

 R

LATEX

 YouTube

 python™

 Perl



GitHub