Beyond the Genome

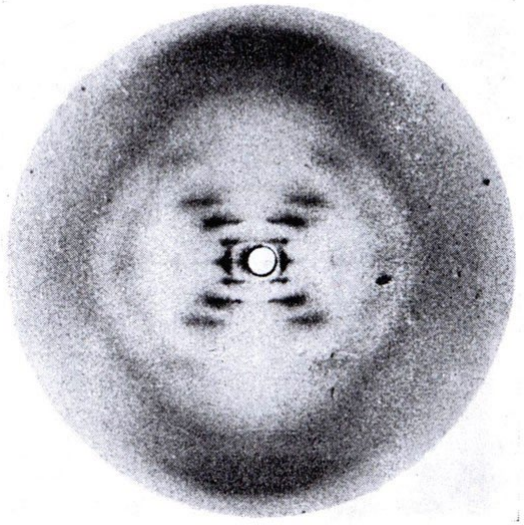
Chandler Sutherland

Krasileva Lab, Department of Plant and Microbial Biology, UC Berkeley

Letter	**Nucleotide:** building block of DNA, consisting of sugar molecule and a nucleotide base
Word	**Sequence:** a string of nucleotides
Book	**Genome:** the complete set of genetic material (DNA) in an organism
Library	**Genomics:** study of the structure, function, evolution, and editing of genomes

Brief History of Sequencing

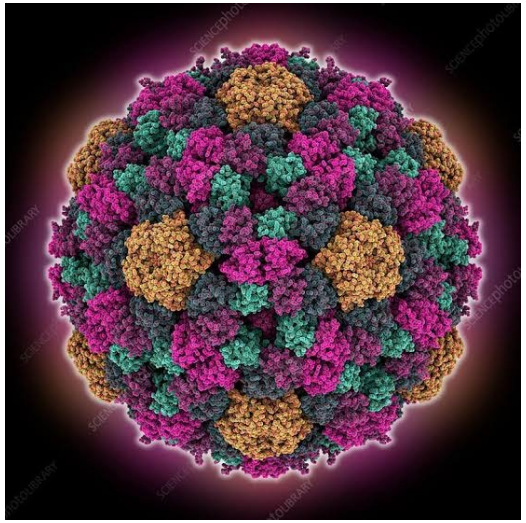
1869: Discovery and Isolation of DNA



1953: Structure of DNA determined

1965: First DNA Sequence

1977: First genome

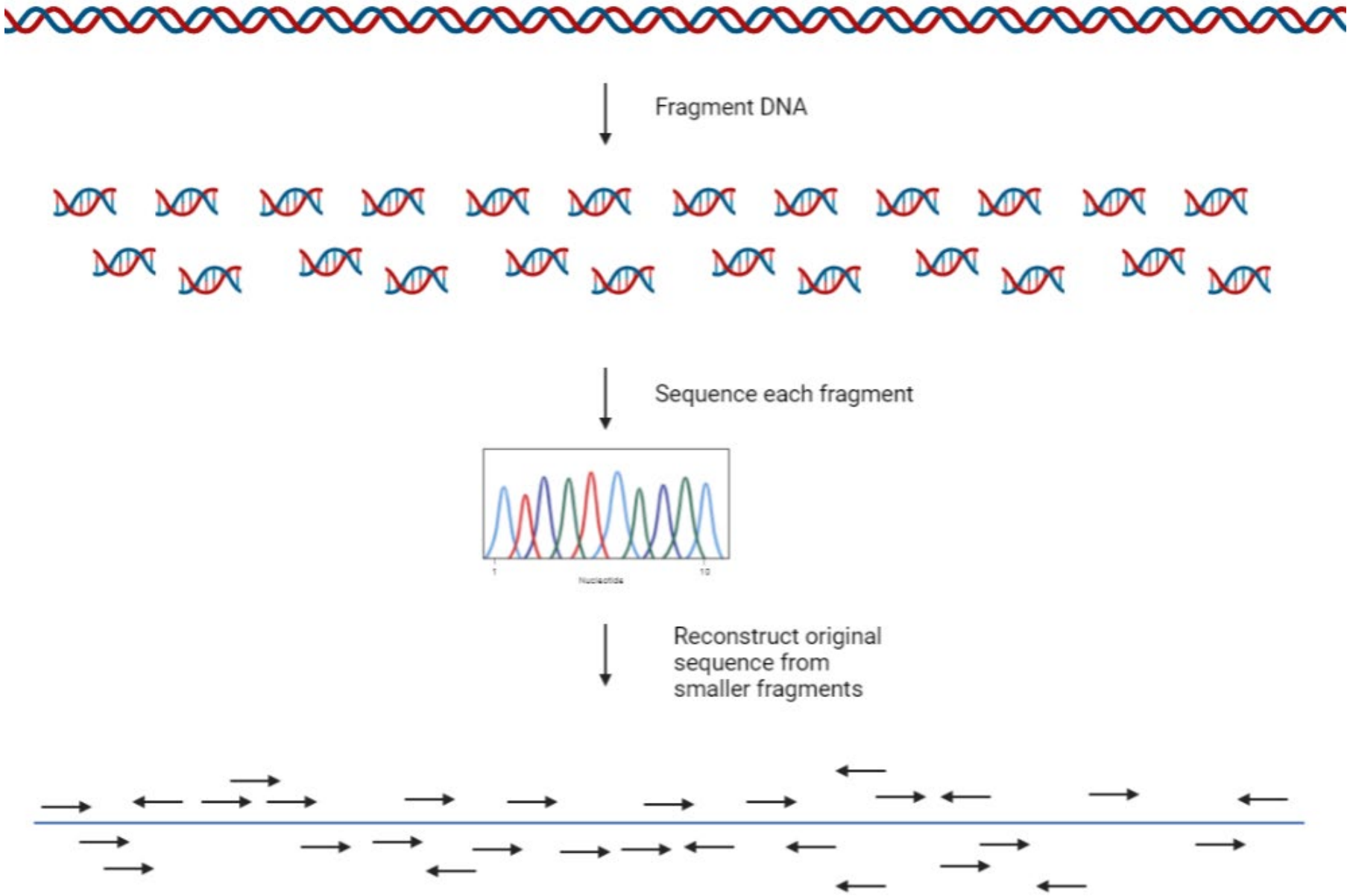


1996: Sequencing by fluorescently labeled dyes

2000: First plant genome sequenced

2014: Nanopore long read sequencing

2021: Human genome finished end to end


Fragment DNA
Sequence each fragment
Reconstruct original sequence from smaller fragments

Dominant Sequencing Technologies

1

Short read sequencing by synthesis (Illumina)

- >90% of world's sequencing data
- Each read (sequence-able fragment of DNA) is only 100-350 nucleotides long
- Very accurate

2

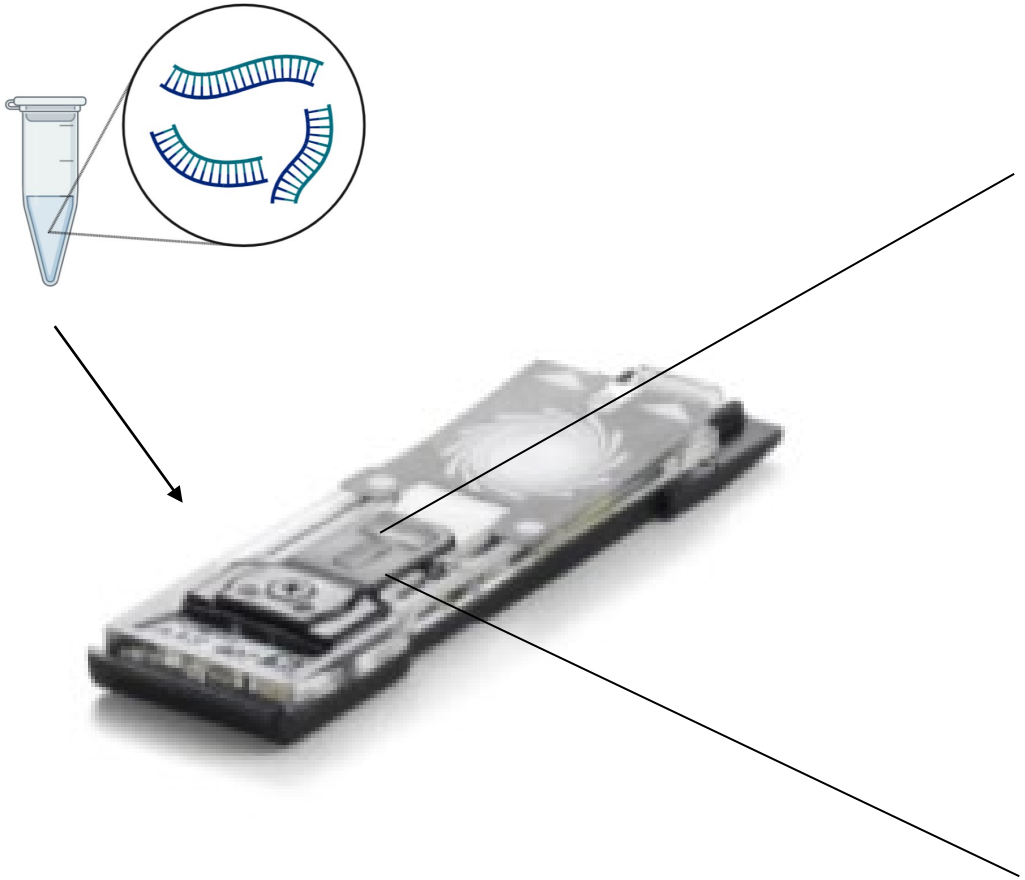
Long read sequencing (PacBio)

- Rapid sequencing (3 DNA molecules per second)
- Long reads – up to 20,000 nucleotides
- Somewhat accurate

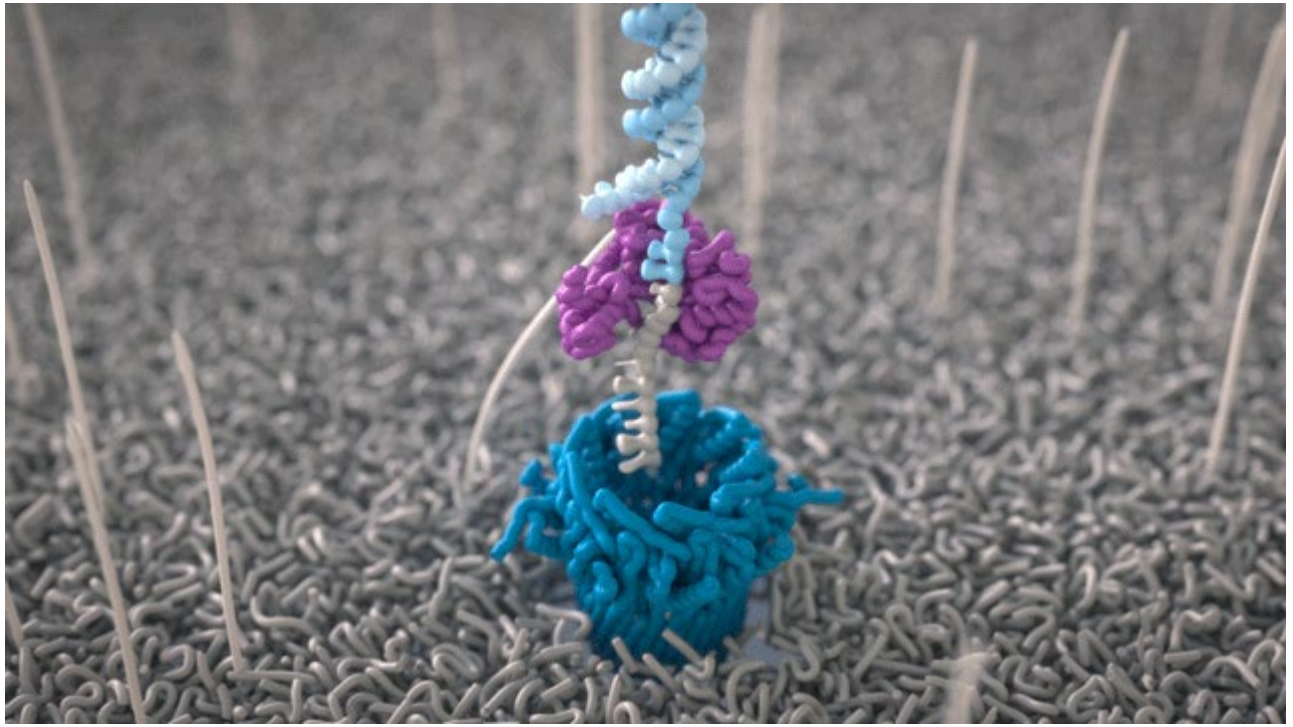
3

Long read sequencing (Oxford Nanopore)

- Small, portable, USB-powered
- Long reads (up to 1,000,000s nucleotides)
- Field applications, including missions in space

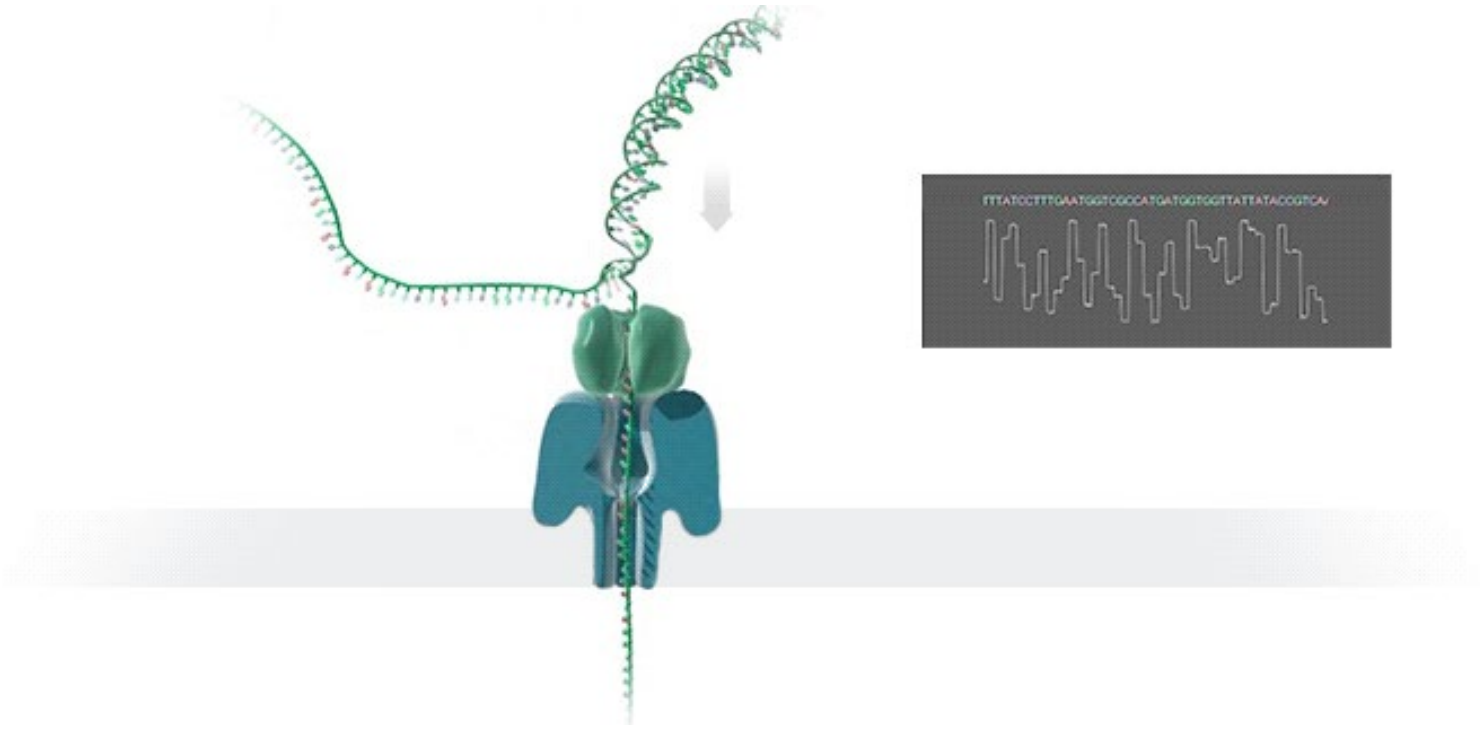


Flow Cell

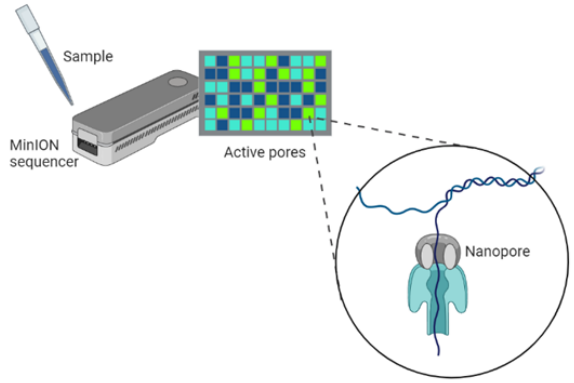


Nanopores

As each strand of DNA is passed through a nanopore, an electrical signal is recorded that can be used to decode the sequence



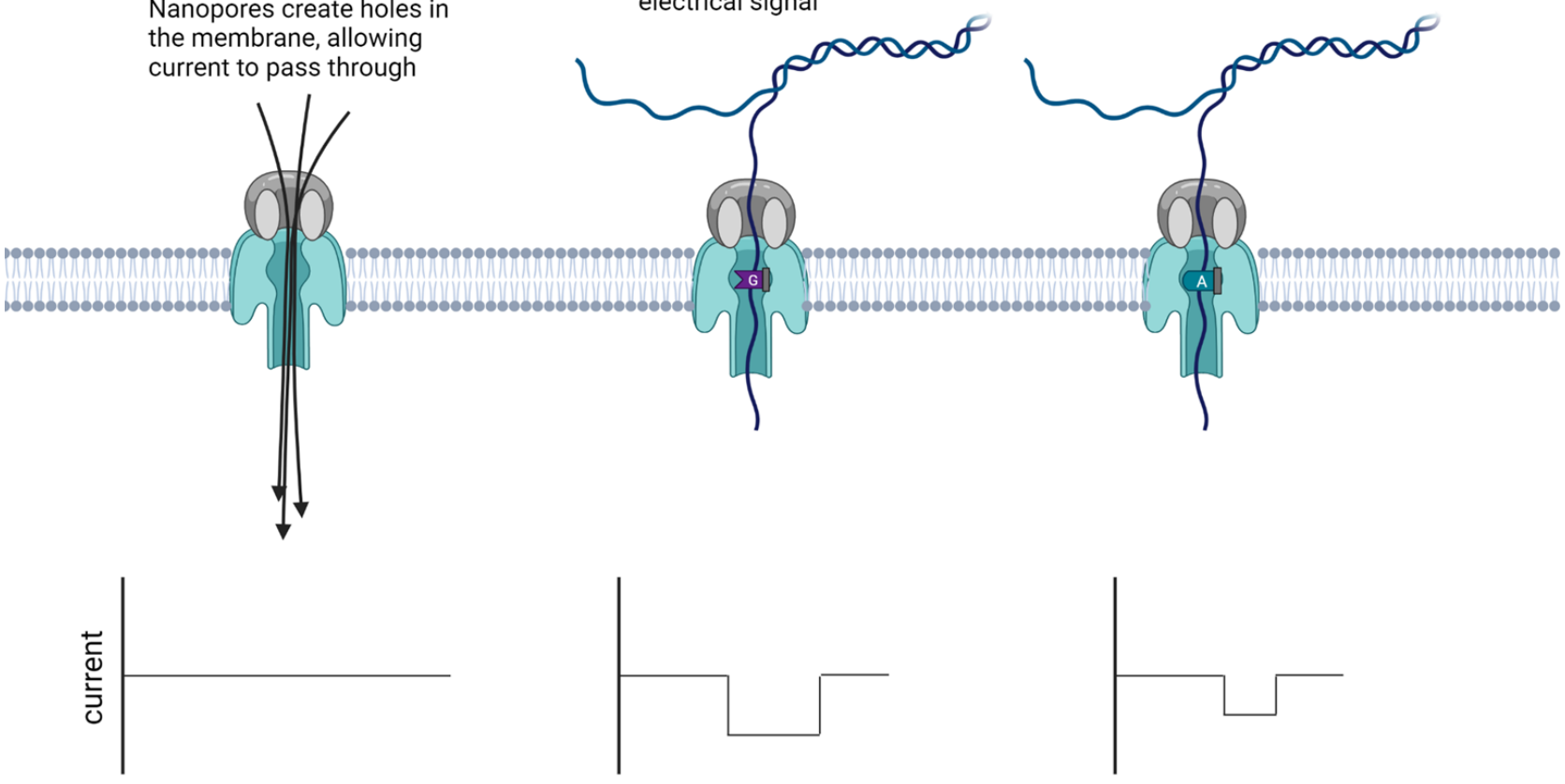
Nanopore sensing



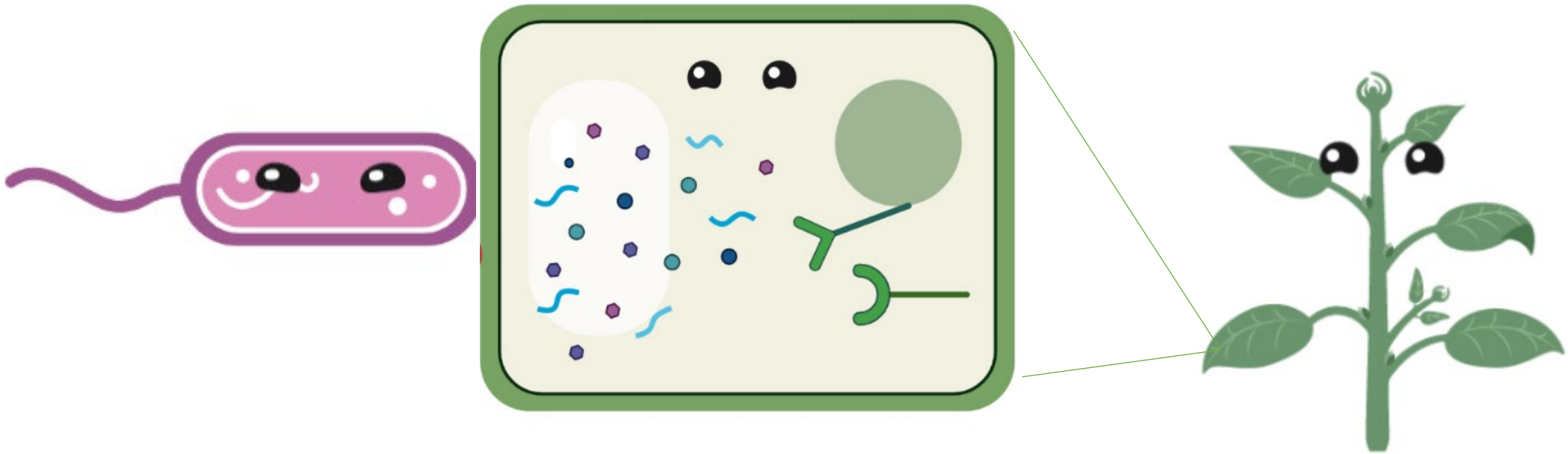
Nanopores create holes in the membrane, allowing current to pass through

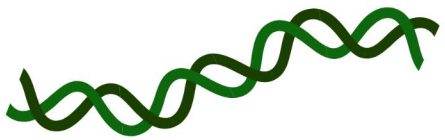
As a DNA sequence passes through the pore, disruptions are characterized by a distinct electrical signal

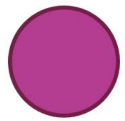
Current distruption is interpreted to understand the identity of the base pair

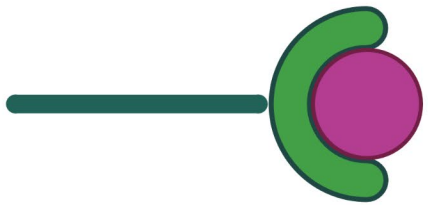


Plant Recognition of Pathogens

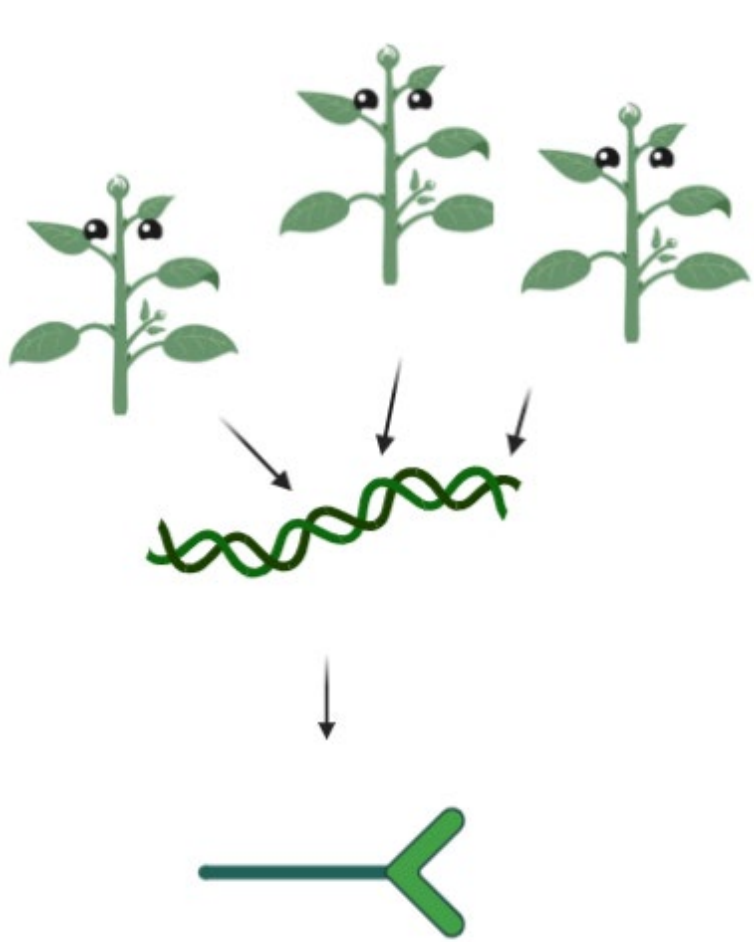


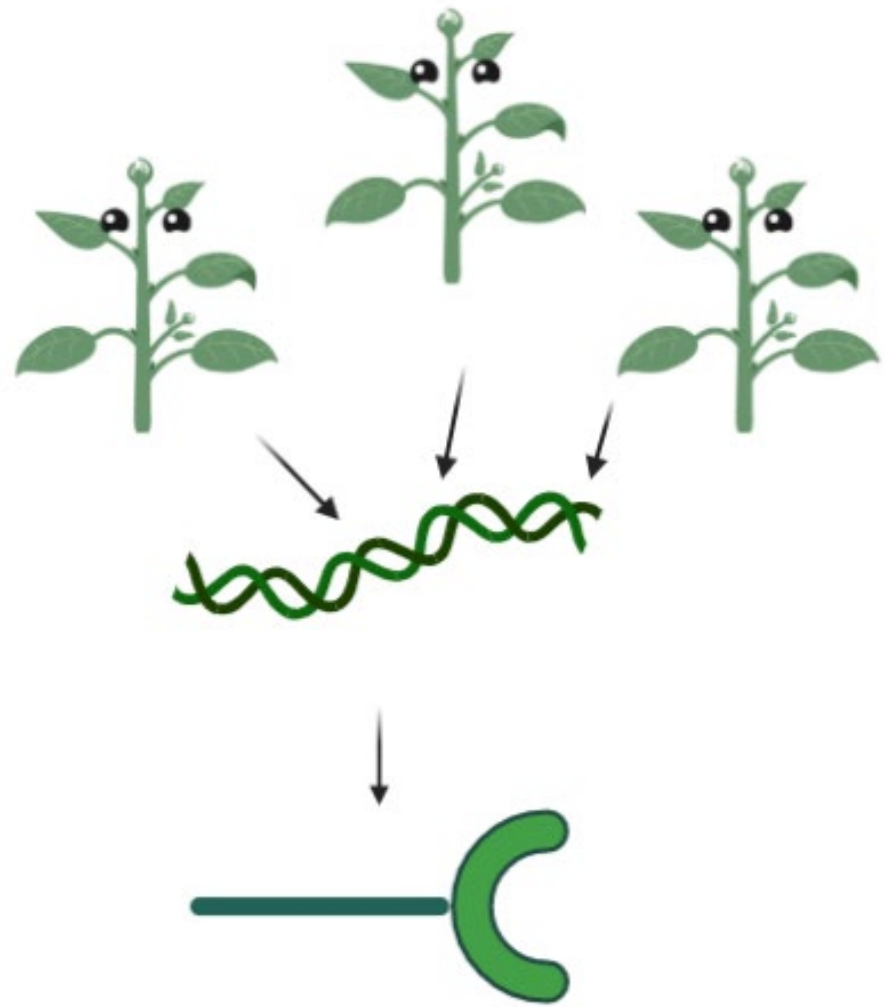


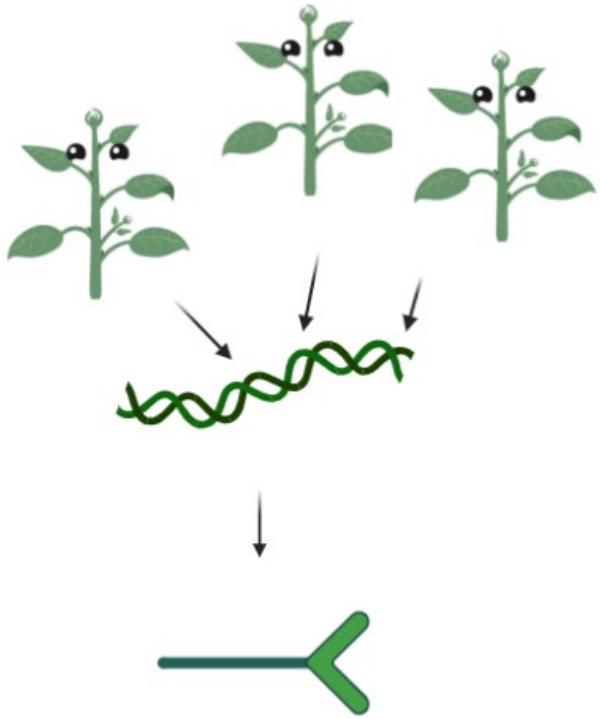




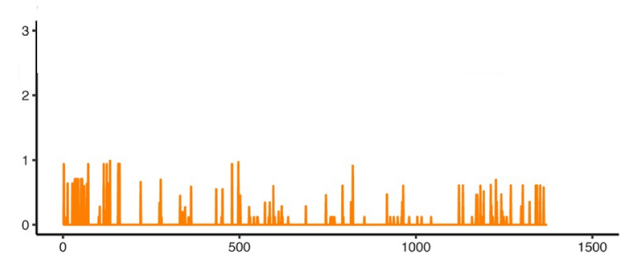
Recognition =
Immunity!



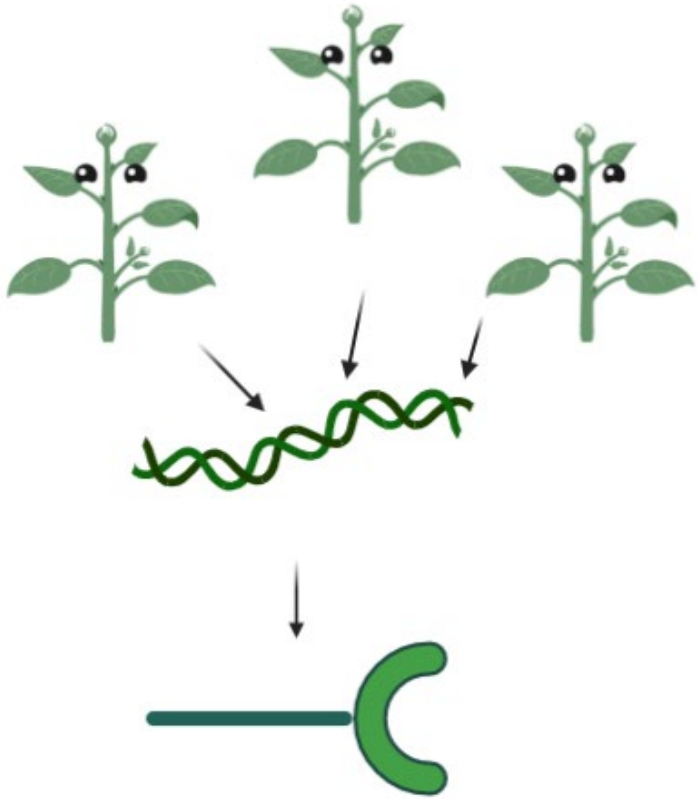




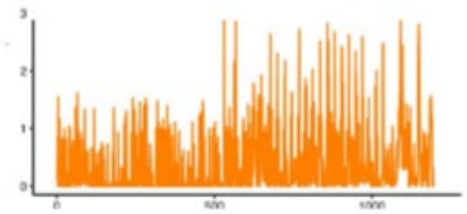
Diversity



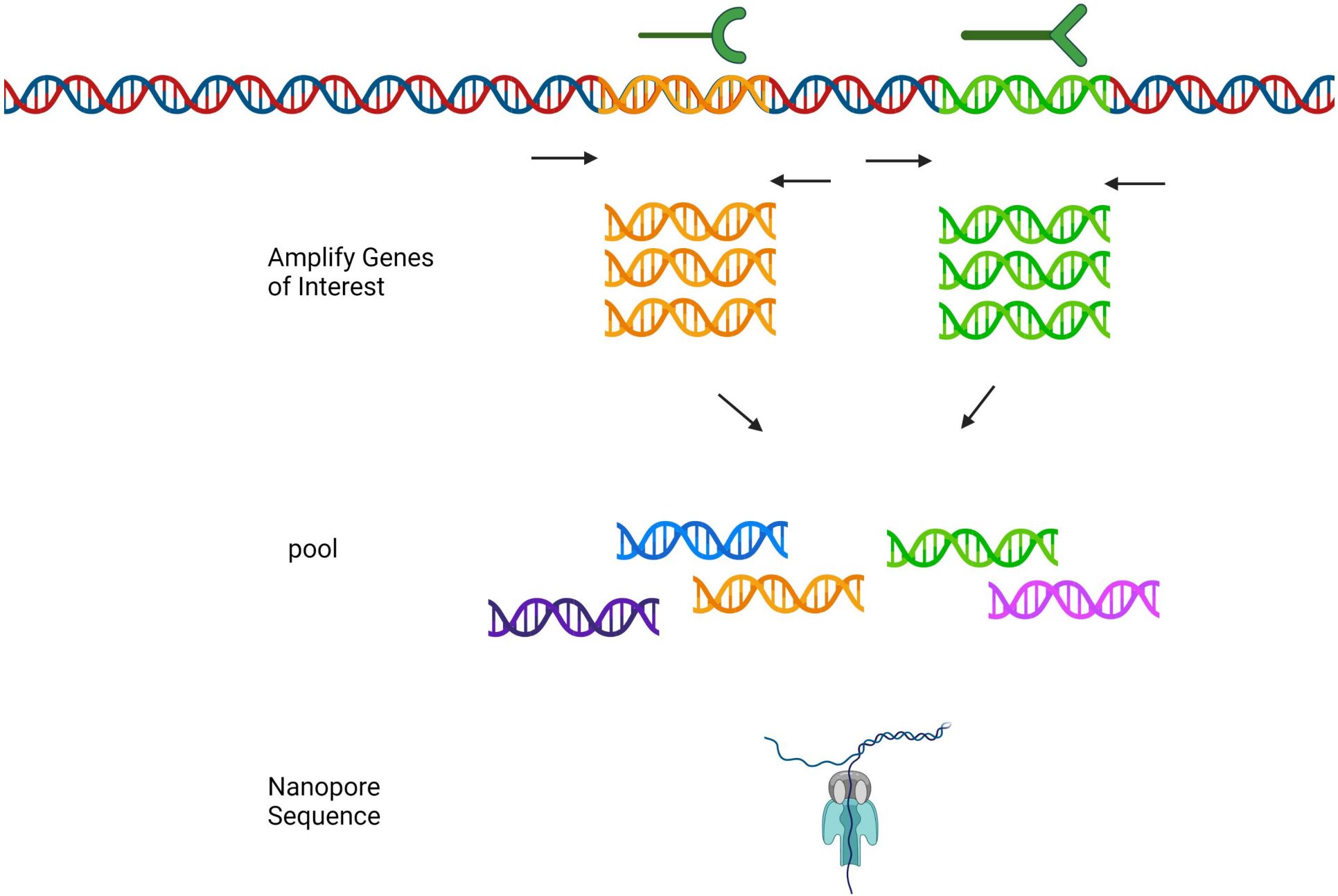
DNA sequence

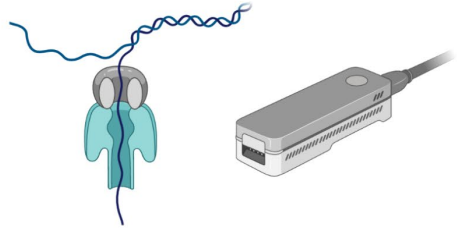


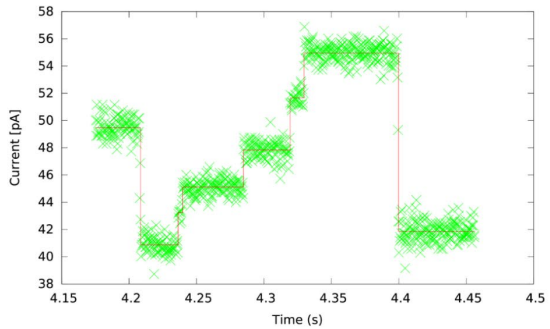
Diversity



DNA sequence


Amplify Genes
of Interest
pool
Nanopore
Sequence



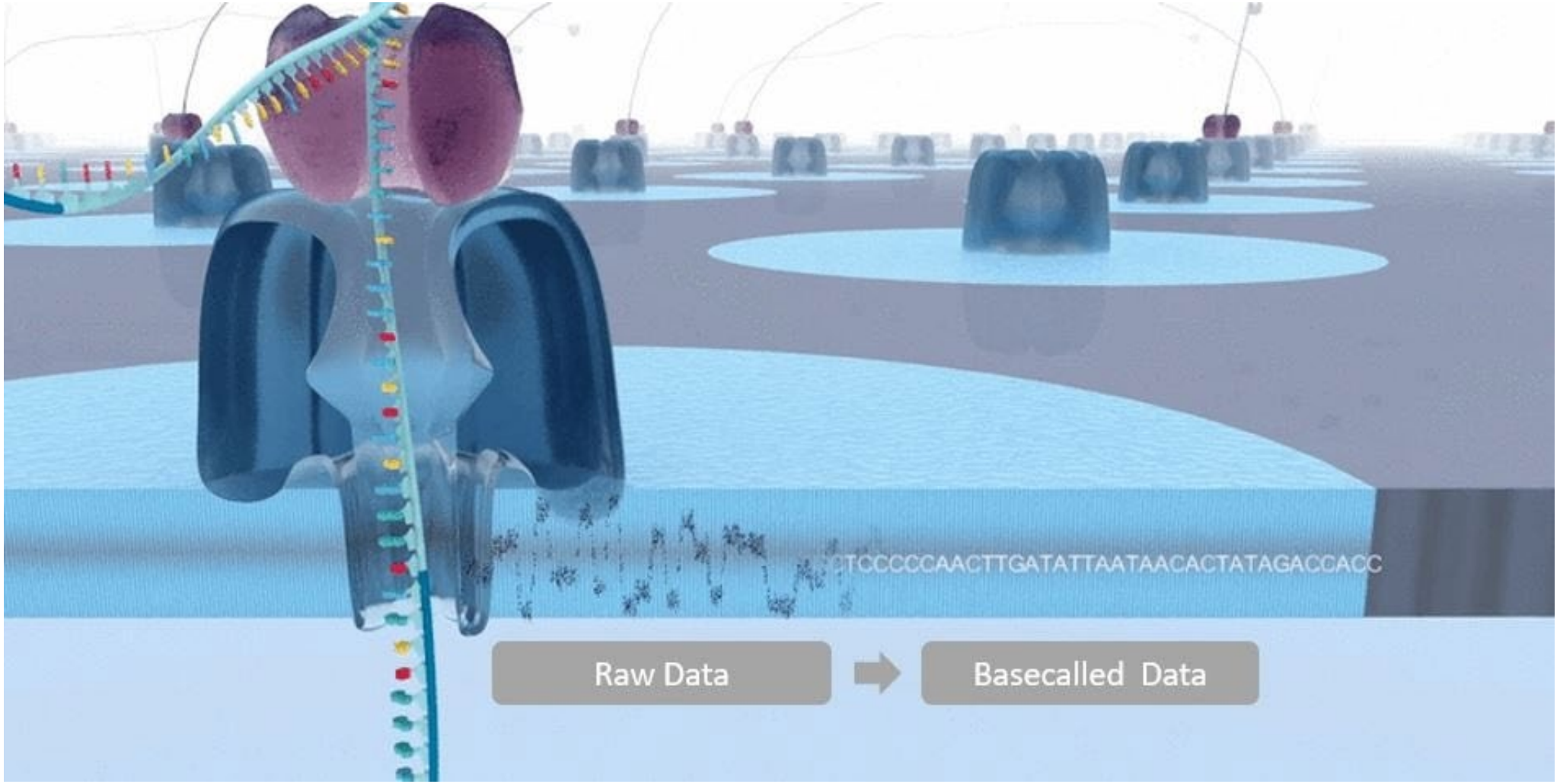


.fast5

Basecalling

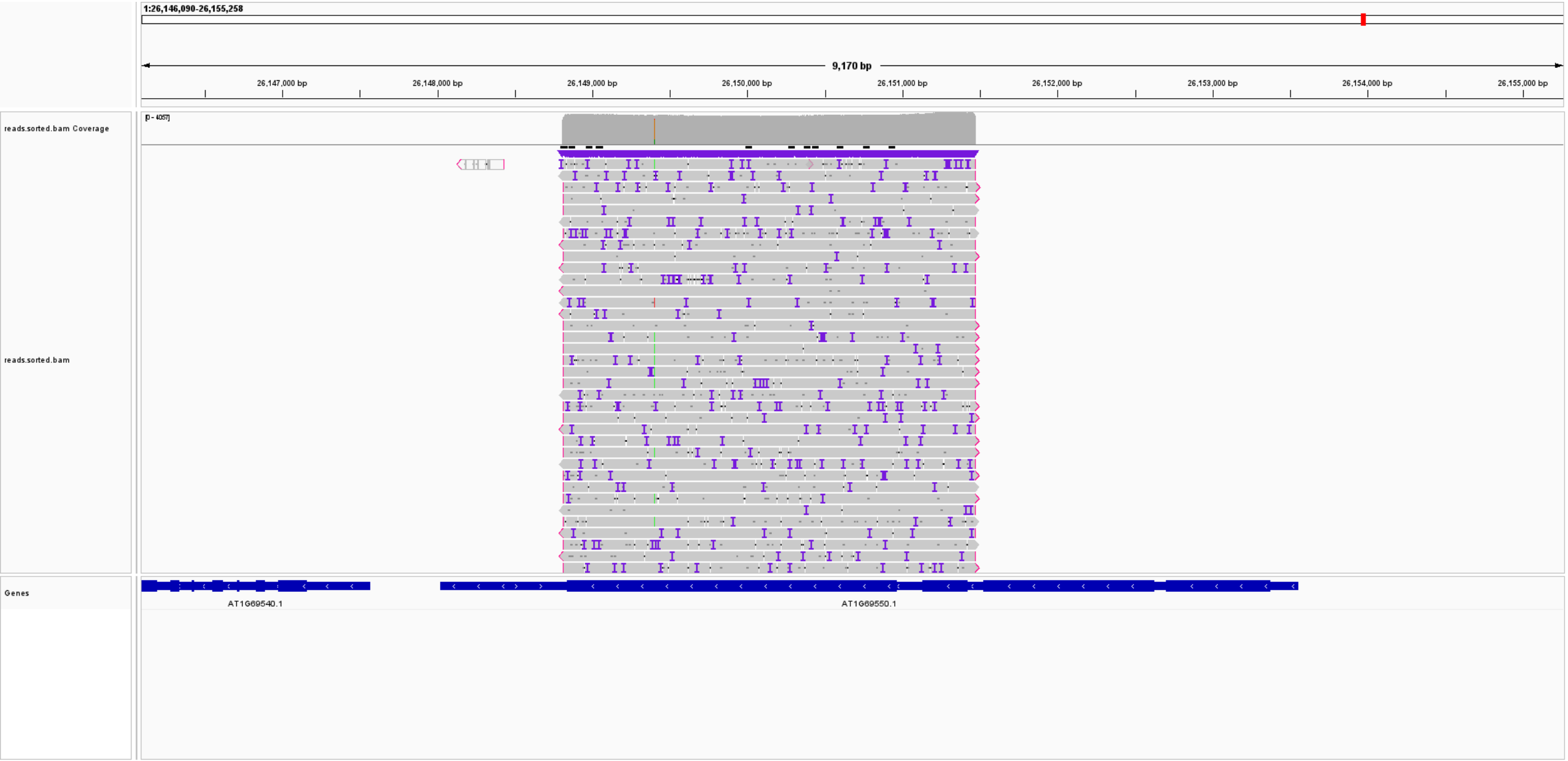
.fastq

Basecalling

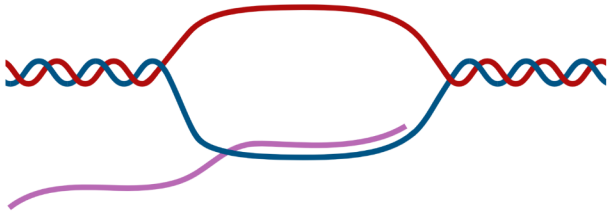


Time to process some sequenced reads!

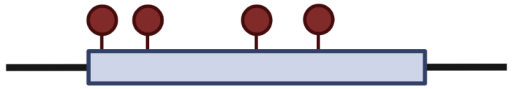
tinyurl.com/nuevagenome


1:26,146,090-26,155,258
9,170 bp
26,147,000 bp
26,148,000 bp
26,149,000 bp
26,150,000 bp
26,151,000 bp
26,152,000 bp
26,153,000 bp
26,154,000 bp
26,155,000 bp
reads.sorted.bam Coverage
[p - 4057]
reads.sorted.bam
Genes
AT1G69540.1
AT1G69550.1

What other information can we evaluate?



Transcription

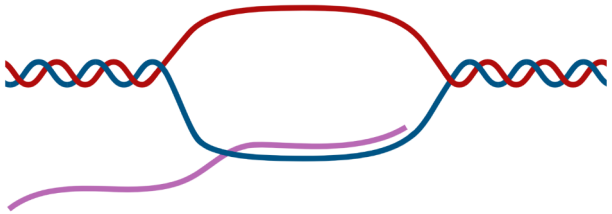


Methyl groups

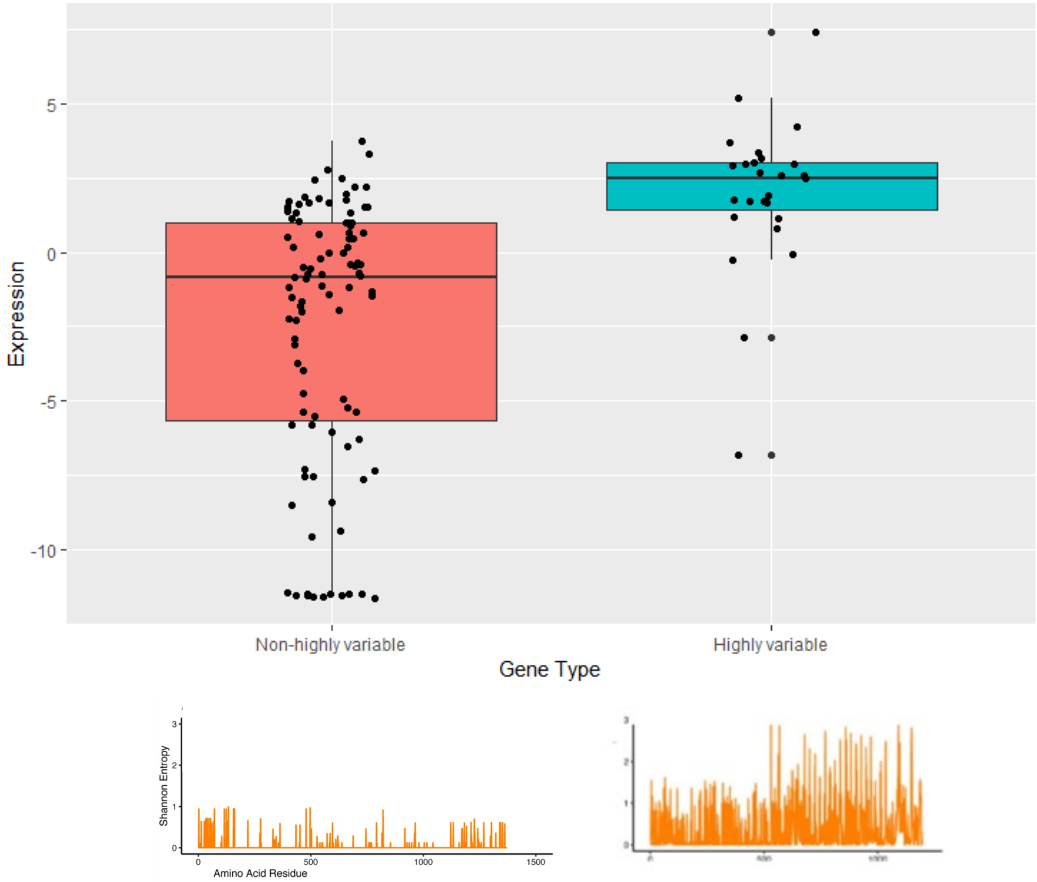
Mutation rate

Transposable Elements

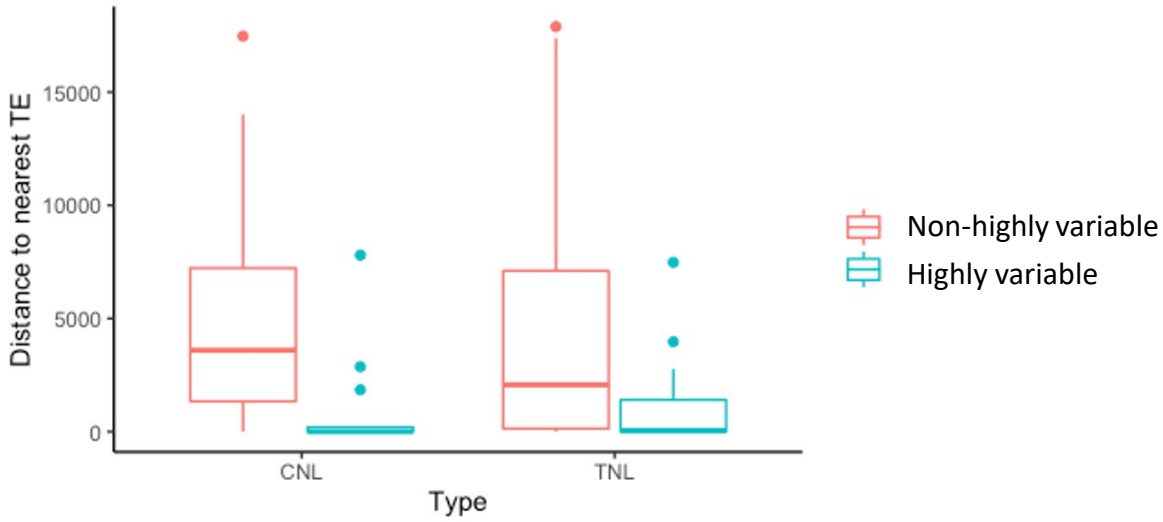
Highly variable immune receptors are expressed more than non-highly variable immune receptors



Transcription

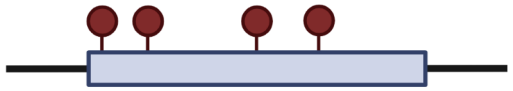


Highly variable immune receptors are closer to transcriptional elements than non-highly variable immune receptors

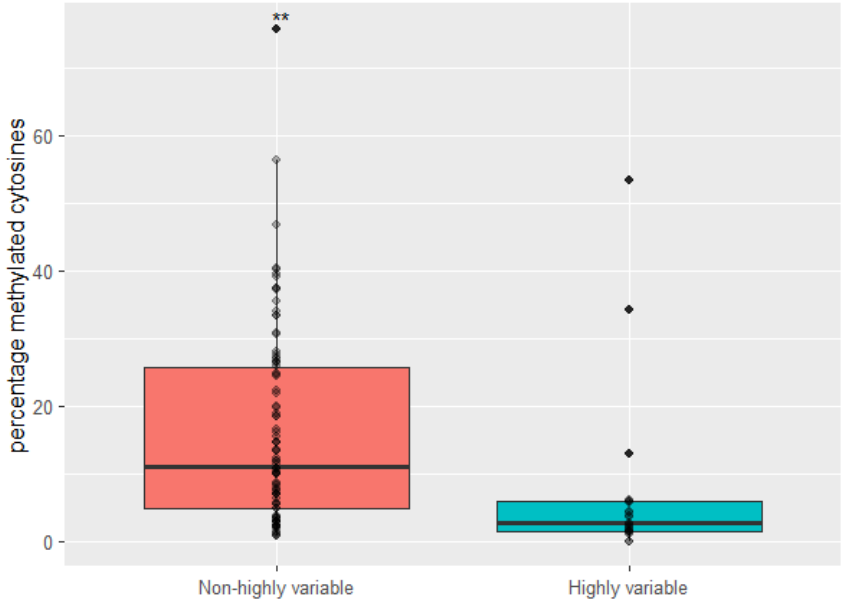


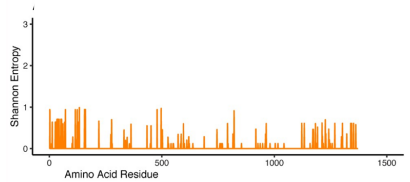
Transposable Elements

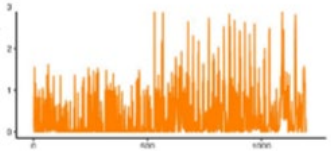
Highly variable immune are more methylated than non-highly variable immune receptors



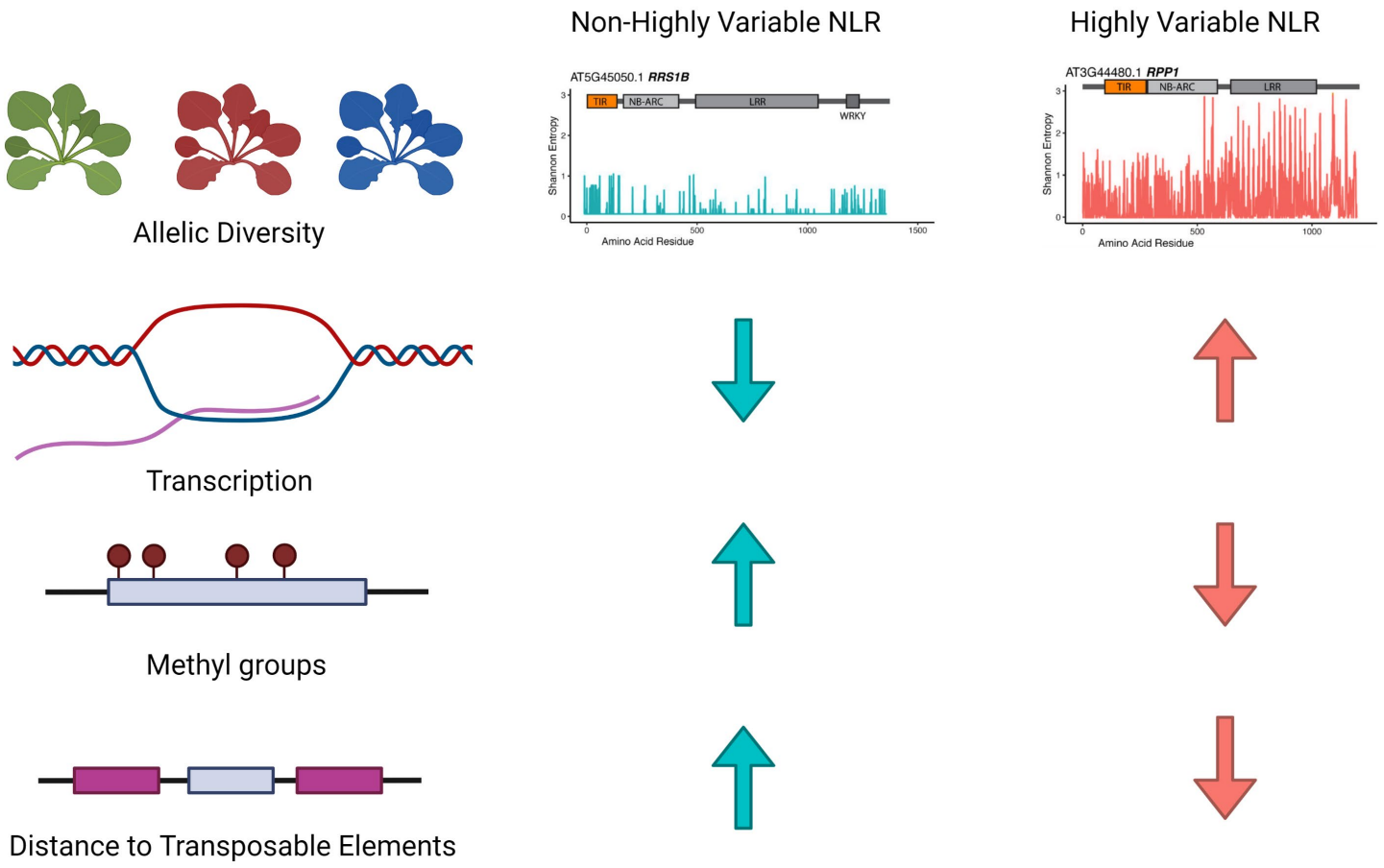
Methyl groups







An individual's methylation status, expression, and TE distance are correlated with intraspecies allelic diversity.



Berkeley Plant & Microbial Biology

@chandler_suth

chandlersutherland@berkeley.edu