Input DNA

Desire:

1. Clean
2. Good yield
3. High molecular weight (HMW)

Why?

Assessments?

Input DNA

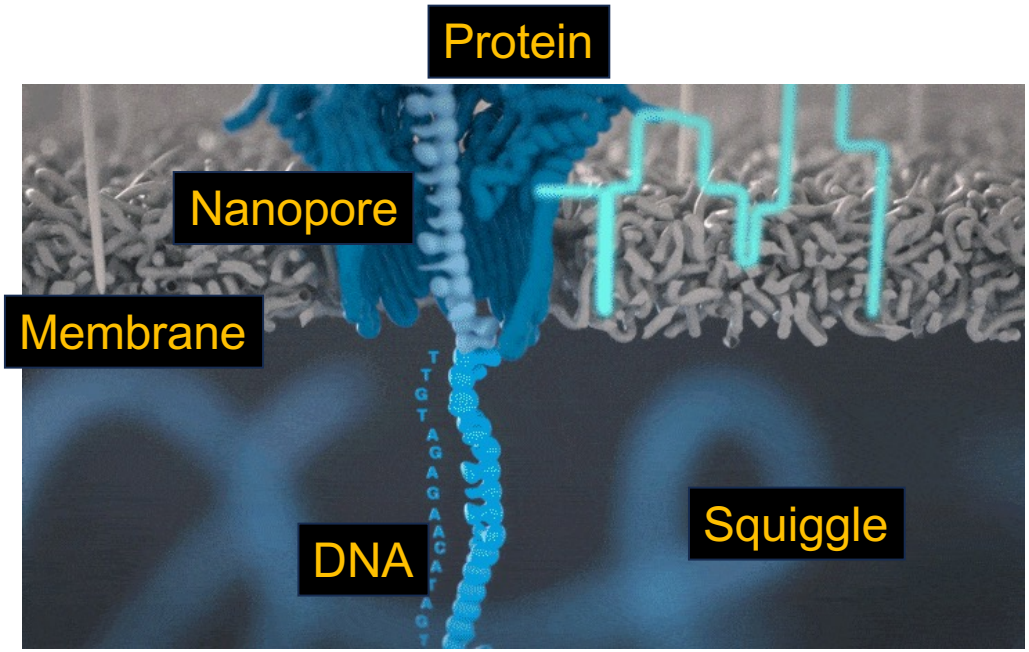
Bacteria:
Precipitation-based kits (Promega Wizard kit)
HMW extraction kits (Monarch, others)
column methods shear DNA, not good

Eukarya:
gentle/old school methods
Phenol/Chloroform
Shepherd's crook
Qiagen genomic tip, HMW kits

post extraction size selection:
Needle shear ultra-long DNA first
then 0.4X SPRI bead clean up

or Blue Pippin size selection
(generally not needed for Bacteria)

Clean DNA



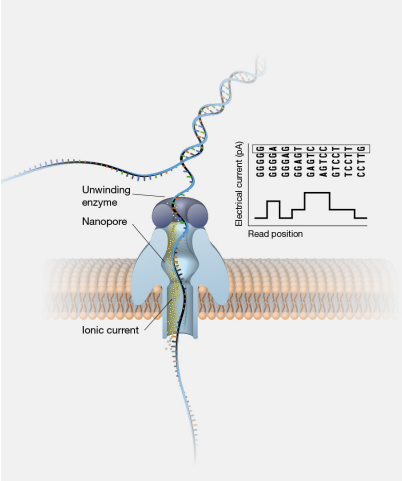
Why clean DNA?

1. Chemicals impact downstream manipulations (library prep)
2. Chemicals or co-precipitated protein may impact proteins on flowcell

Nanopores are proteins

https://nanoporetech.com/platform/technology

Need to balance amount and size



Why?
Goal is to have pores constantly sequencing DNA

Small DNA = preferentially sequenced, but better overall yield
Too (HMW) DNA = ultralong DNA clogs pores quickly

underloaded/little DNA = pores open longer, degrade faster
If HMW DNA, can fragment

Too much DNA = generally not a big problem, can help with yield

No ”one size fits all solution”
Depends on genome size and amount of repeats
Depends on use of genome sequence
But
Can always dilute sample or make HMW smaller, but not the opposite

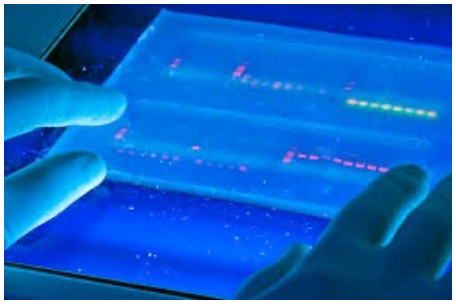
https://www.genome.gov/genetics-glossary/Nanopore-DNA-Sequencing

Methods for assessing DNA

Nanodrop

260/280 ≥ 1.8
260/230 ≥ 2.0

Gel



~Size, ~yield, ~quality

Qubit

Yield

Bioanalyzer

Size

Flongle

~Size, ~yield, ~quality

Library preparation: ligation kit

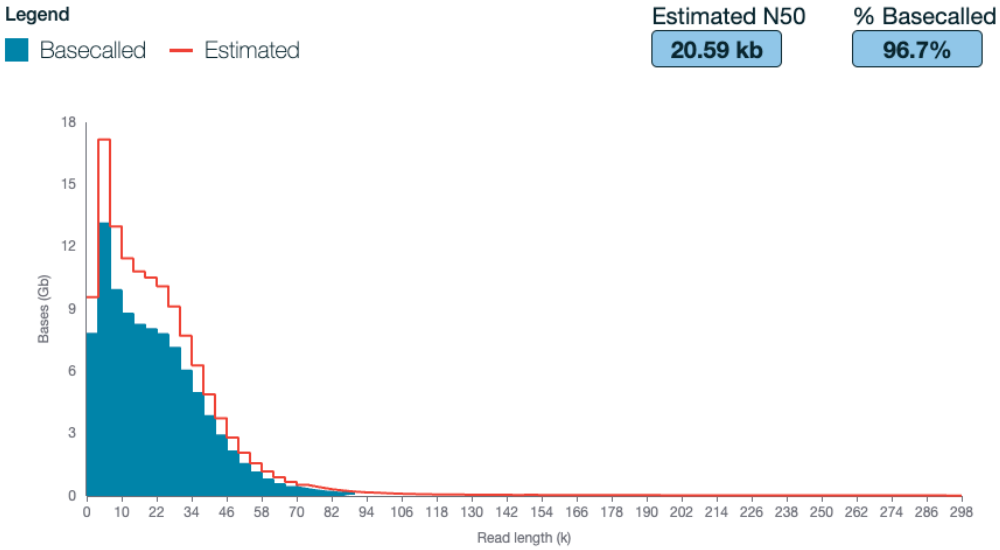
+: generally highest yield (>100-150 Gb on PromethION)
-: need other kits (end-repair & nick repair); cost; time
+/-: generally high seq length, but does not get ultralong reads

DATA OUTPUT

Estimated bases	Reads generated
129.06 Gb	12.59 M
Estimated N50	**Total data produced (pass / fail)**
20.59 kb	1.6 TB

BASECALLING

Reads called	Bases called (min Q score: 10)	
96.7%	97.06 Gb	24.46 Gb
	Pass	Fail



Library preparation: rapid kit

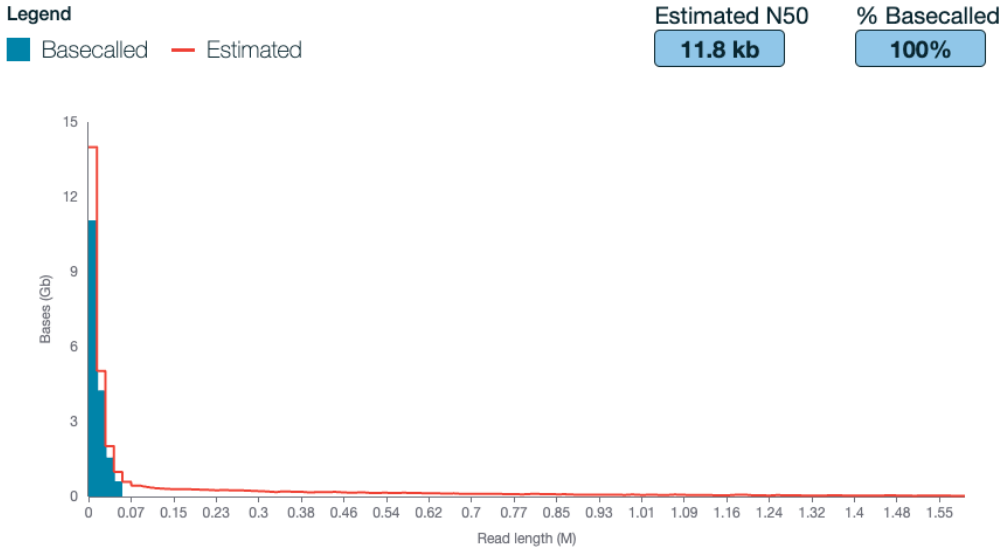
+: faster, easier, and cheaper
-: generally lower yield (~50-80 GB on P2);
-: less control in fragmenting DNA, not good for amplicons if whole length is needed
+/-: generally long reads, can get ultralong reads

DATA OUTPUT

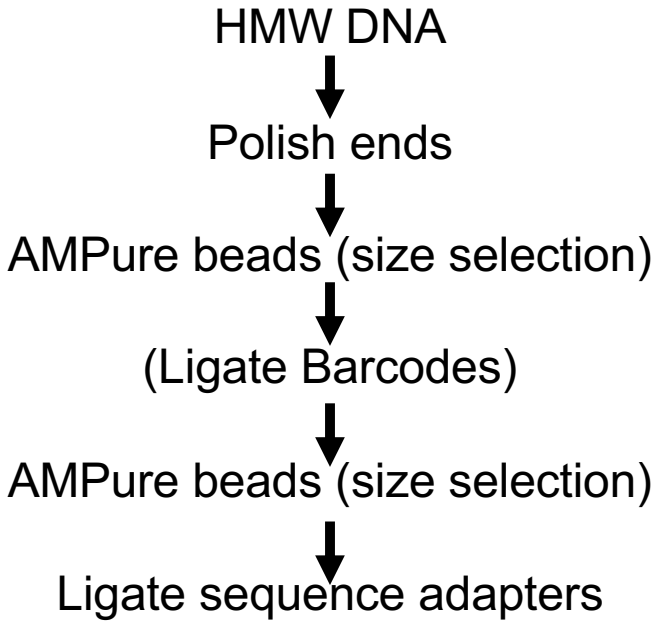
Estimated bases	Reads generated
34.57 Gb	7.8 M
Estimated N50	**Total data produced (pass / fail)**
11.8 kb	423.54 GB

BASECALLING

Reads called	Bases called (min Q score: 10)	
100%	17.68 Gb	5.18 Gb
	Pass	Fail

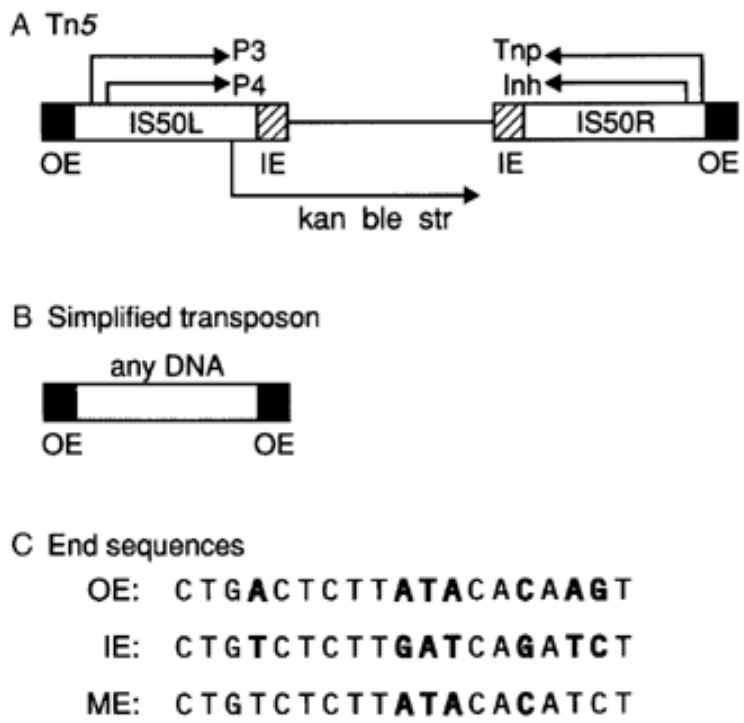


Overview of ligation kit



~9 hrs

Overview of rapid kit



HMW DNA

↓

Add transposome (w/ barcodes)

↓

AMPure beads (size selection)

↓

Add sequence adapters

~2 hrs

Molecular Microbiology, Volume: 47, Issue: 5, Pages: 1199-1206, First published: 19 February 2003, DOI: (10.1046/j.1365-2958.2003.03382.x)

Quality control

What is the end goal?

1. Infer taxonomic identity
2. Functional genomics
3. Genome evolution (structure & variation)
4. MGEs
5. Population genomics

What do you know about the genome(s) of interest?

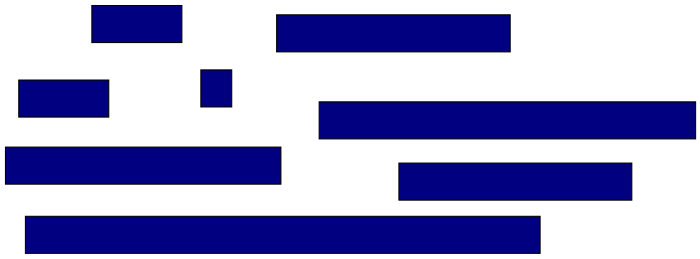
1. Size
2. Complexity (repeat number and size)
3. Segmental duplications
4. Ploidy

Can’t do much w/ bad reads

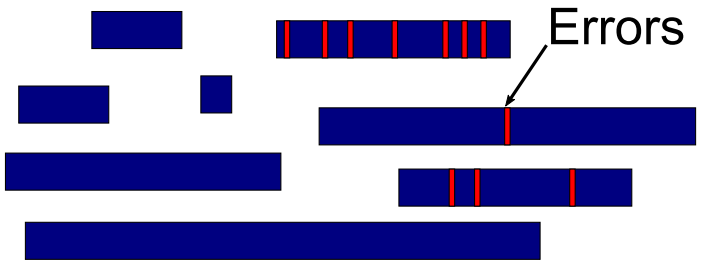
Findings based on genome analyses should be framed on the quality of assemblies

Three important properties of reads

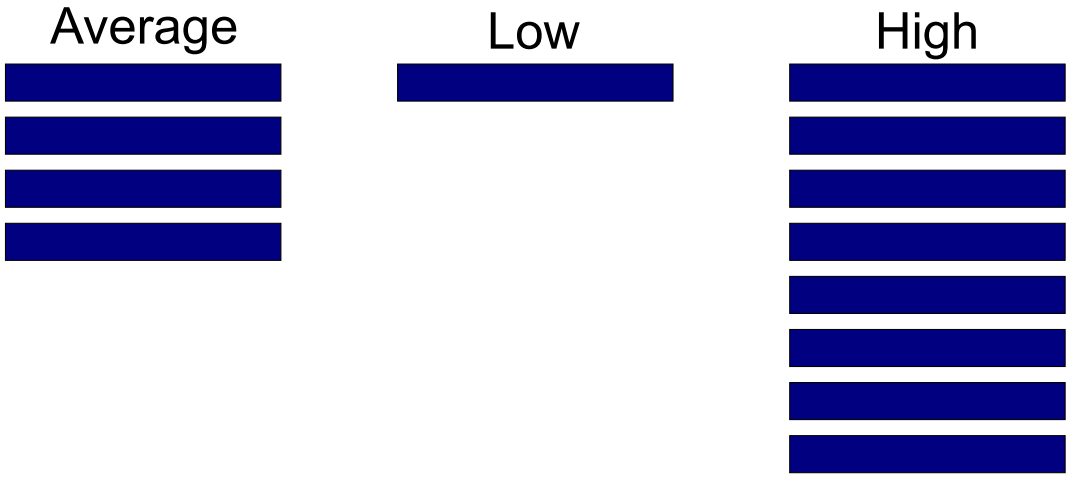
1. Read length



2. Accuracy



3. Evenness



Tools to generate summary statistics

QUAST: quality assessment tool for genome assemblies

Alexey Gurevich[1,*], Vladislav Saveliev[1], Nikolay Vyahhi[1] and Glenn Tesler[2]

MultiQC: summarize analysis results for multiple tools and samples in a single report

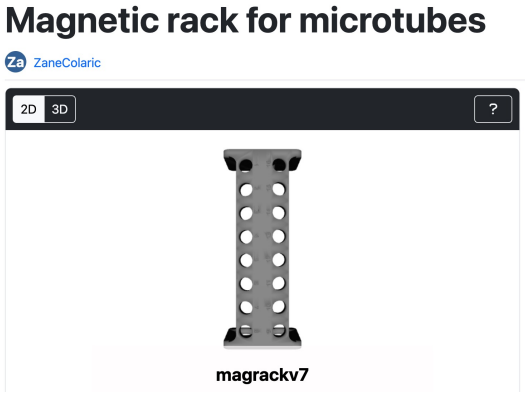
Philip Ewels[1,*], Måns Magnusson[2], Sverker Lundin[3] and Max Käller[3]

QUAST: helps assess quality of genome assembly
MultiQC: gathers, summarizes, and visualizes stats on reads &/or assemblies

doi: 10.1093/bioinformatics/btw354 https://doi.org/10.1093/bioinformatics/btt086

Equipment and computing needs

Magnetic rack for bead cleanup (cheapest to 3D print or order from (Sergi Lab Supply)
https://3d.nih.gov/entries/3DPX-013579
https://www.kjmagnetics.com/proddetail.asp?prod=BZZ082
Qubit not required but highly recommended



Linux computer with GPU for running the sequencer
GPU is the most important part, needed for basecalling
NVIDIA RTX 4090 (or better) is good enough for P2 solo
System76 makes good computers with Linux pre-installed

ONT all-in-one devices (MinIT, Mk1c, P2i…) are convenient but tend to get outdated quickly…