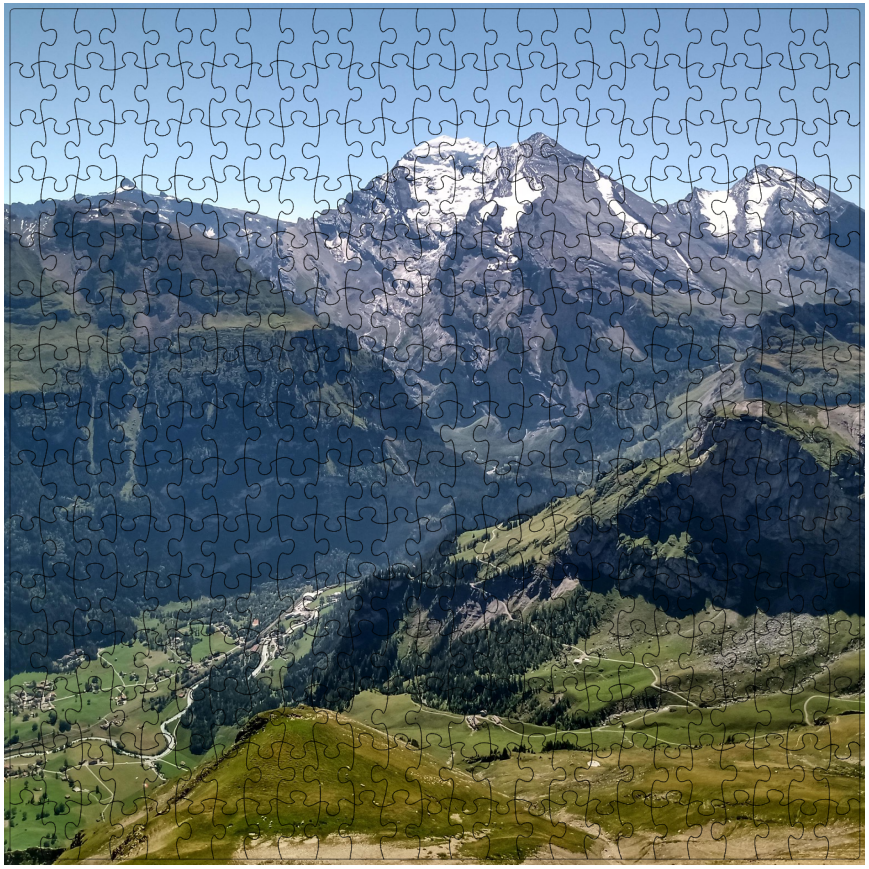
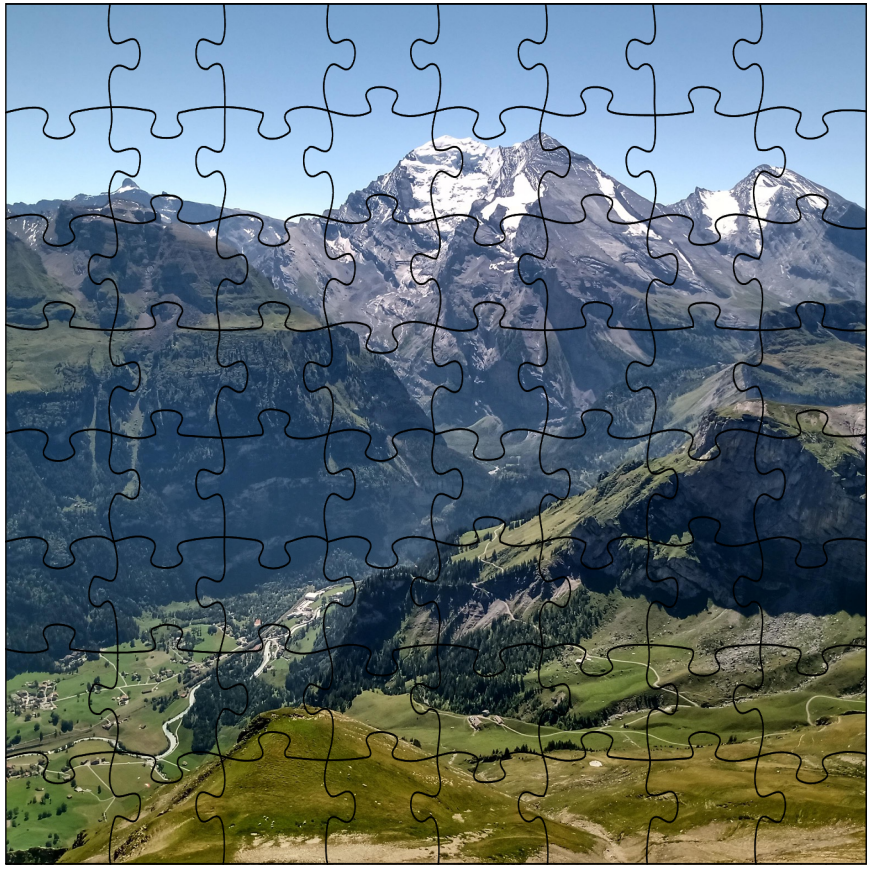
Long-read sequence analysis

Applications

Why long reads?





Applications

- (Genome) assembly
- Variant analysis
- Transcriptome analysis
- Epigenetics
- Metagenomics

Assembly

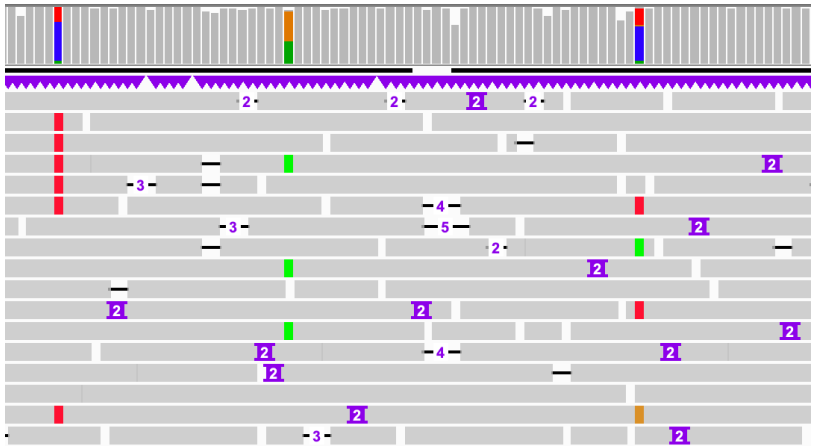
- Reconstructing a DNA sequence out of fragmented sequences:
  - Better understand variations between and within species
  - Reduces resources and increases accuracy for many applications
- Longer reads:
  - Less assembly errors
  - Higher contiguity (longer contigs)
  - Lower computational resources required

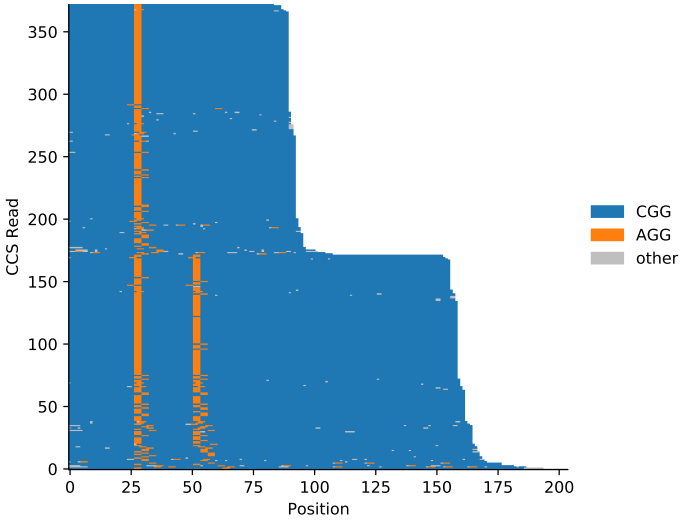
LR assembly software

- Microbial:
  - Flye
  - HGAP
  - Trycycler
- Large(r) genomes:
  - Shasta (ONT only)
  - Falcon (PacBio only)
  - Canu
  - Flye
  - Hifiasm
  - IPA
  - Verkko (integration PacBio and ONT)

Variant analysis

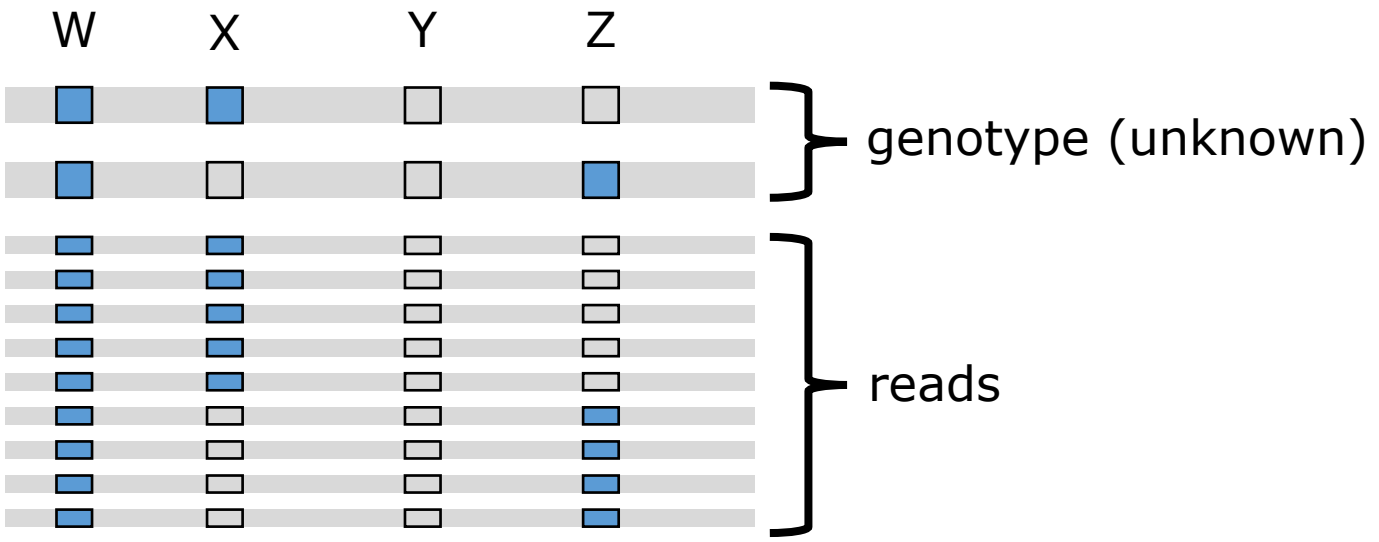
- Low accuracy long reads: not particularly good at variant analysis
- Major applications
  - Structural variation/large repeats
  - Phasing of variants





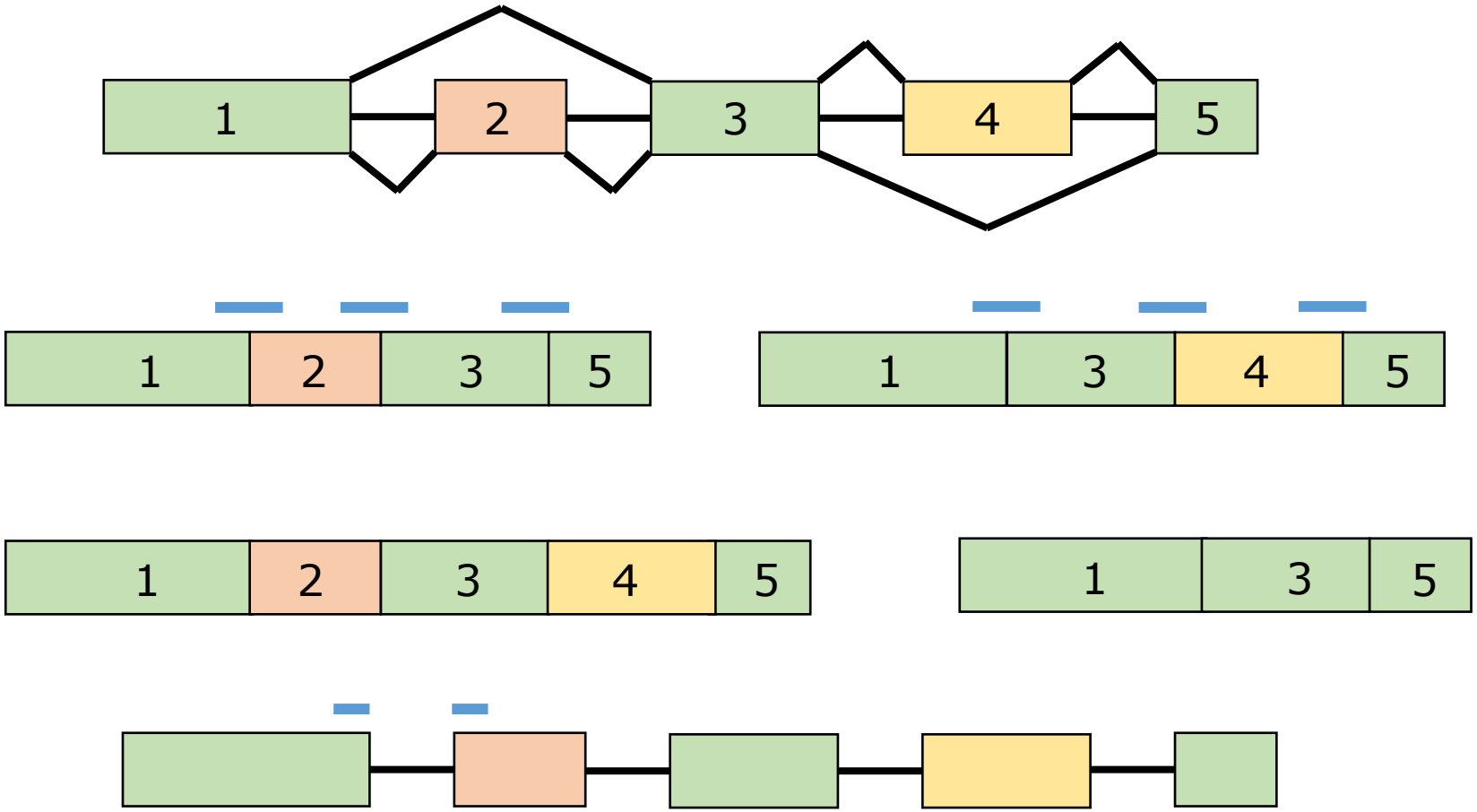
Phasing

- Variant calling:
  - DeepVariant (Pacbio + ONT)
  - Clair3 (ONT)
- Phasing: Whatshap



Repeat expansion - project 2

Transcriptome analysis

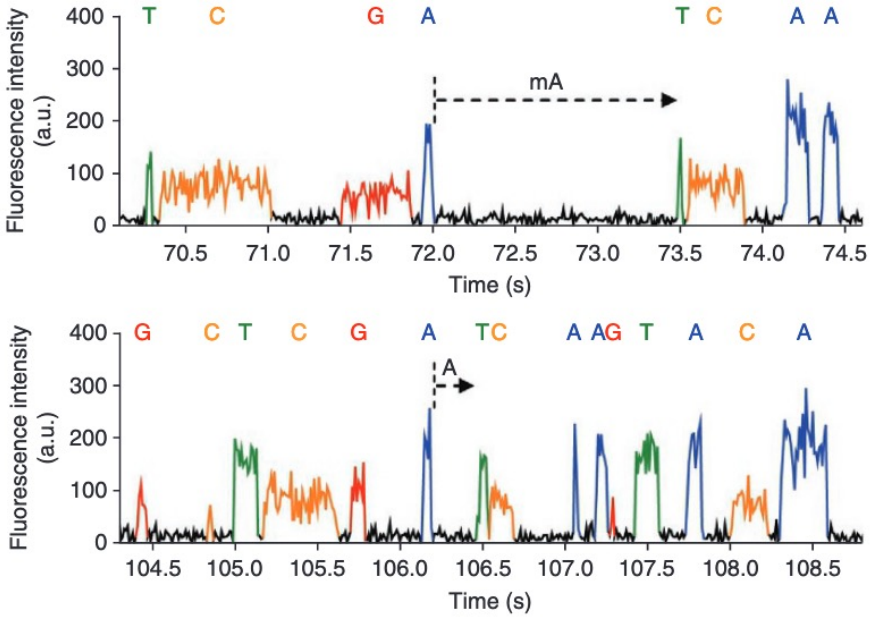


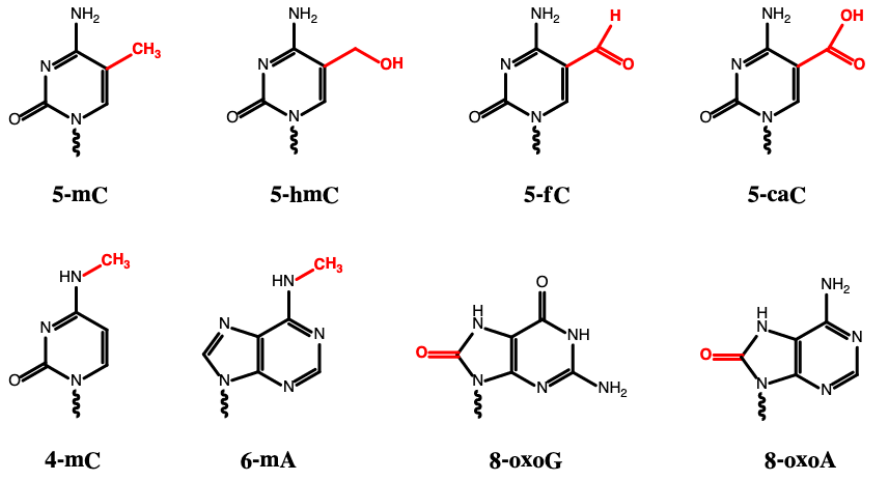
Transcriptome analysis

- Full transcripts: less ambiguous identification + quantification of transcript isoforms
- Better gene model prediction (e.g. for genome annotation)
- ONT direct RNA-seq:
  - No PCR amplification
  - No reverse transcription
  - Base modification identification

Epigenetics

- Base modification in non-amplified libraries (Pacbio and ONT)
- No multiplexing
- Also possible in direct RNA-seq (ONT)





https://www.nature.com/articles/nmeth.1459.pdf

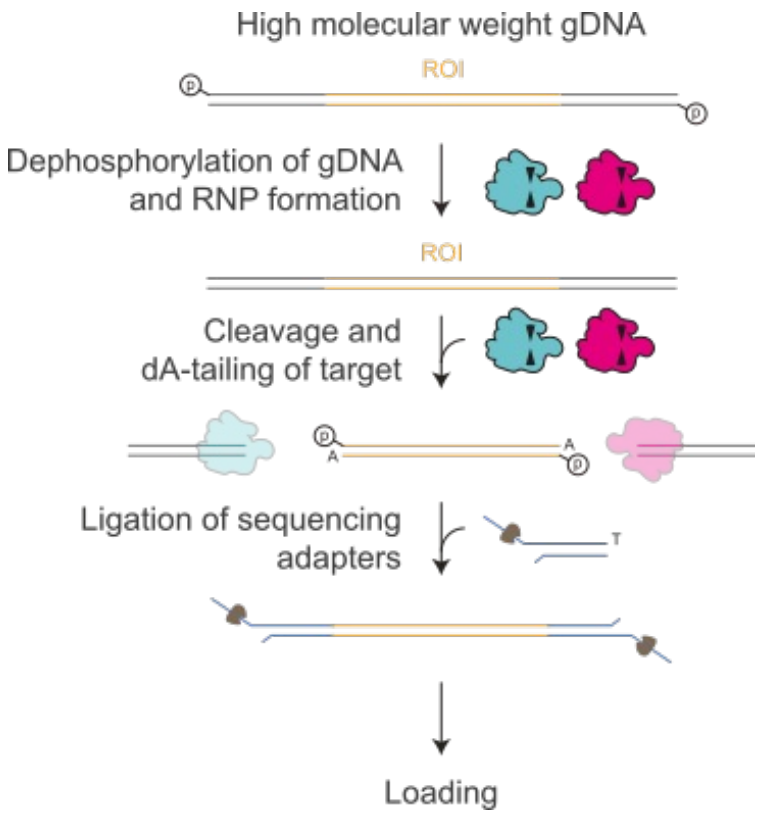
Metagenomics

- Full length 16S genes: better classification (PacBio CCS)
- Metagenome Assembled Genomes (MAGs) -> `flye --meta`

https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-021-02094-5
https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-016-0891-4
https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-07607-0

New possibilities

- Cas9 based targeting (ONT): also base modifications
- Duplex reads (ONT)
- Adaptive sampling (ONT)



https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02582-x
https://www.nature.com/articles/s42003-021-01648-7

image: https://store.nanoporetech.com/catalog/product/view/id/461/s/cas9-sequencing-kit/category/28/

Downstream analysis

- ONT:
  - Workflows on EPI2ME
  - https://nf-co.re/bacass
  - https://nf-co.re/viralrecon
- PacBio:
  - pbbioconda
  - https://nf-co.re/isoseq
- Both:
  - https://nf-co.re/mag