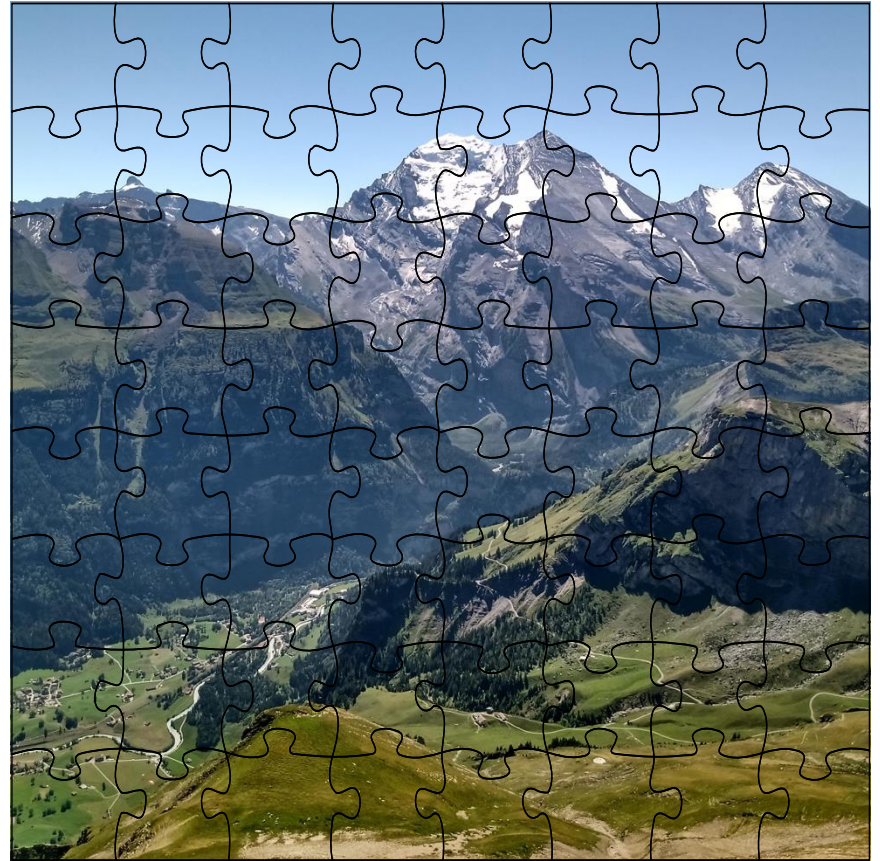


Long-read sequence analysis

Applications

Why long reads?



Applications

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

Assembly

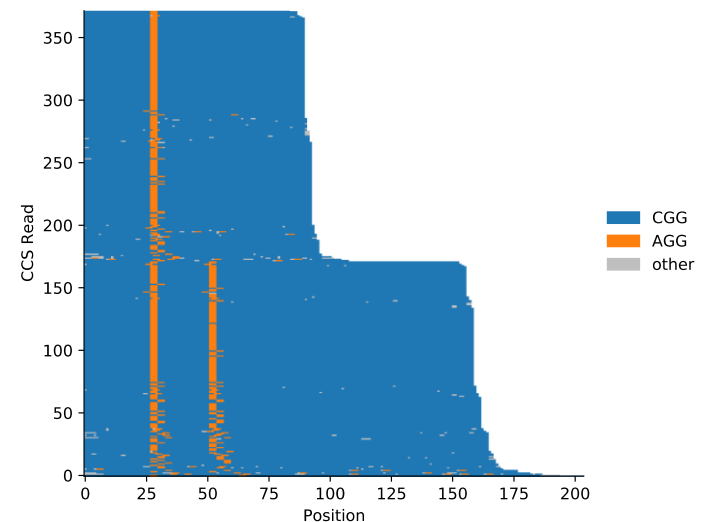
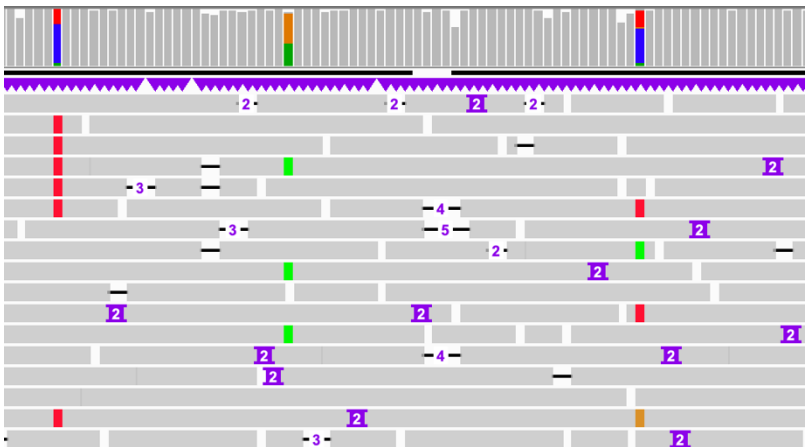
- Reconstructing a DNA sequence out of fragmented sequences:
 - Better understand variations between and within species
 - Reduces resources and increases accuracy for many applications
- Longer reads:
 - Less assembly errors
 - Higher contiguity (longer contigs)
 - Lower computational resources required

LR assembly software

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

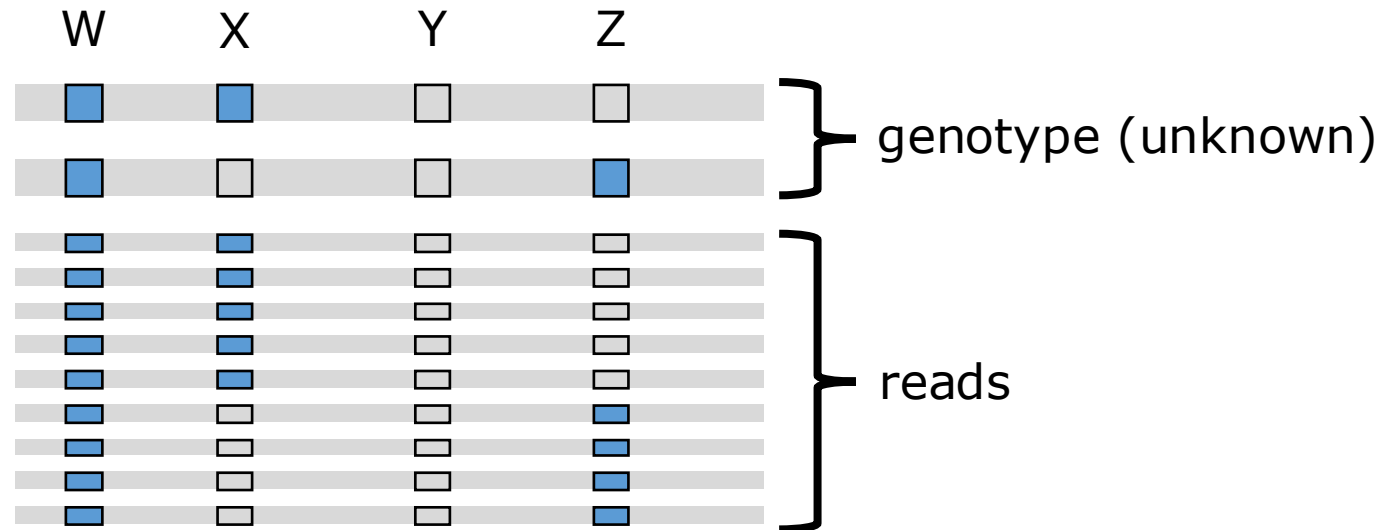
Variant analysis

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



Phasing

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

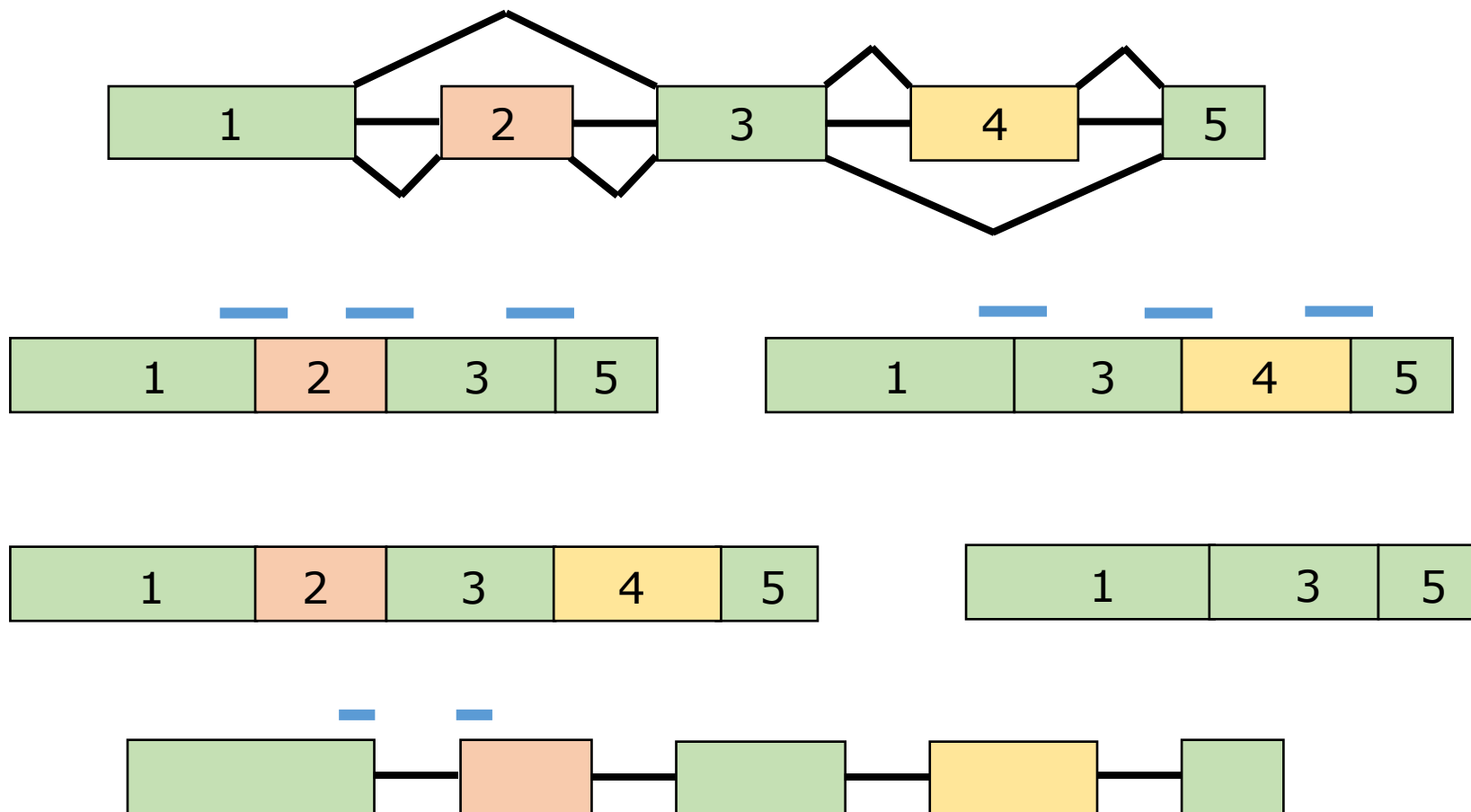


Repeat expansion



- Medaka or PathSTR: ONT
- TRGT: PacBio

Transcriptome analysis

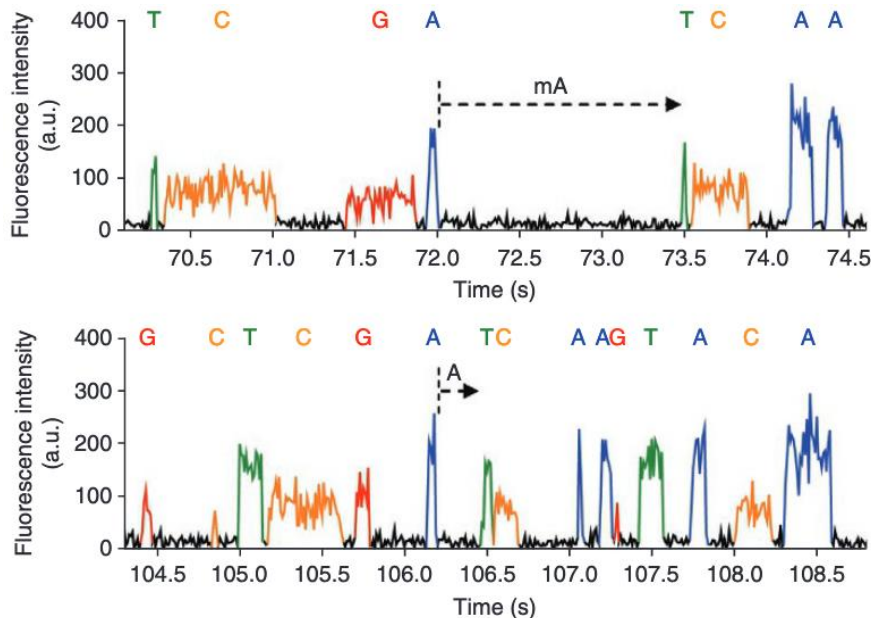


Transcriptome analysis

- 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
- Relevance of RNA integrity
 - RIN > 8 (PacBio and ONT recommendations)
 - RNA degradation can bias results towards 3'

Epigenetics

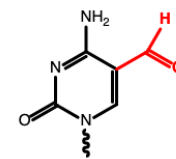
- Base modification in non-amplified libraries (Pacbio and ONT)
- No multiplexing
- Also possible in direct RNA-seq (ONT)



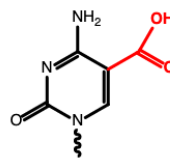
5-mC



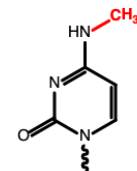
5-hmC



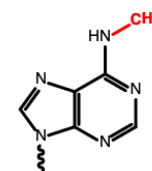
5-fC



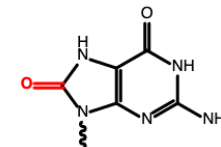
5-caC



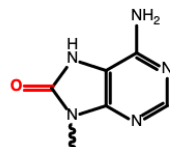
4-mC



6-mA



8-oxoG



8-oxoA

Metagenomics

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

Long read tools

- Evolving field: new tools and updates

Mini-Review

A Hitchhiker's Guide to long-read genomic analysis

Medhat Mahmoud,^{1,4} Daniel P. Agostinho,^{1,4} and Fritz J. Sedlazeck^{1,2,3}

¹Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA; ²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; ³Department of Computer Science, Rice University, Houston, Texas 77005, USA

Table 1. List of methods for long-read analysis and its function

Function	Tool	Technology	GitHub	Citation
Basecalling (A)	CCS	PacBio	https://github.com/PacificBiosciences/ccs	
	Dorado	ONT	https://github.com/nanoporetech/dorado	
	Google Deep Consensus	PacBio	https://github.com/google/deepconsensus	Baid et al. 2023
Reads QC (B)	LongReadSum	ONT/PacBio	https://github.com/WGLab/LongReadSum	
	LongQC	ONT/PacBio	https://github.com/yfukasawa/LongQC	Fukasawa et al. 2020
	NanoPack	ONT/PacBio	https://github.com/wdecoster/nanopack	De Coster et al. 2018
Alignment (C)	LRA	ONT/PacBio	https://github.com/ChaissonLab/LRA	Ren and Chaisson 2021
	minimap2	ONT/PacBio	https://github.com/lh3/minimap2	Li 2018, 2021
	NGMLR	ONT/PacBio	https://github.com/philres/ngmlr	Sedlazeck et al. 2018b
	pbbmm2	PacBio	https://github.com/PacificBiosciences/pbbmm2	
	VACmap	ONT/PacBio	https://github.com/micahvista/VACmap	Ding et al. 2024
	Vulcan	ONT/PacBio	https://github.com/treangenlab/vulcan	Fu et al. 2021
Alignment QC (D)	Winnommap2	ONT/PacBio	https://github.com/marbl/Winnommap2	Jain et al. 2022
	NanoPack	ONT/PacBio	https://github.com/wdecoster/nanopack	De Coster et al. 2018
	Sambamba	ONT/PacBio	https://github.com/biod/sambamba	Tarasov et al. 2015
	SAMtools	ONT/PacBio	https://github.com/samtools/samtools	Li et al. 2009
SNV & indels calling (E)	Clair3	ONT/PacBio	https://github.com/HKU-BAL/Clair3	Zheng et al. 2022
	DeepSomatic	Illumina/ONT/PacBio	https://github.com/google/deepsomatic	Park et al. 2024
	DeepVariant	ONT/PacBio	https://github.com/google/deepvariant	Poplin et al. 2018
	Longshot	ONT/PacBio	https://github.com/pjedge/longshot	Edge and Bansal 2019
SV calling (F)	cuteSV	ONT/PacBio	https://github.com/tjiangHTT/cuteSV	Jiang et al. 2020
	DELLY	ONT/PacBio/Illumina	https://github.com/dellytools/delly?tab=readme-ov-file	Rausch et al. 2012
	NanomonSV	ONT/PacBio	https://github.com/friend1ws/nanomonsv	Shiraishi et al. 2023
	pbsv	PacBio	https://github.com/PacificBiosciences/pbsv	
	Sawfish	PacBio	https://github.com/PacificBiosciences/sawfish	Saunders et al. 2024
	Severus	ONT/PacBio	https://github.com/KolmogorovLab/Severus	Keskus et al. 2024
	Sniffles2	ONT/PacBio	https://github.com/fritzsedlazeck/Sniffles	Smolka et al. 2024
	SAVANA	ONT/PacBio	https://github.com/cortes-ciriano-lab/savana	Erick et al. 2024
	SVision	ONT/PacBio	https://github.com/xjtu-omics/SVvision	Lin et al. 2022b
	HiFiCNV	PacBio	https://github.com/PacificBiosciences/HiFiCNV	
Tandem repeat (TRs) (H)	Spectre	ONT/PacBio	https://github.com/fritzsedlazeck/Spectre	
	Medaka	ONT	https://github.com/nanoporetech/medaka	Lee et al. 2021
	pathSTR	ONT	https://github.com/wdecoster/pathSTR	De Coster et al. 2024
	Straglr	ONT	https://github.com/bcgsc/straglr	Chiu et al. 2021
	StrSpy	ONT	https://github.com/unique379r/strspy	Hall et al. 2022
Genotyping (I)	TRGT	PacBio	https://github.com/PacificBiosciences/trgt	Dolzhenko et al. 2024
	kanpig	ONT/PacBio	https://github.com/ACEnglish/kanpig	English et al. 2024a
Downstream analysis—phasing (J)	SVJedi	ONT/PacBio	https://github.com/llecompte/SVJedi	Romain and Lemaire 2023
	HapCUT2	ONT/PacBio	https://github.com/vibansal/HapCUT2	Edge et al. 2017
	HiPhase	PacBio	https://github.com/PacificBiosciences/HiPhase	Holt et al. 2024
	WhatsHap	ONT/PacBio	https://github.com/whatschap/whatschap	Martin et al. 2023

(continued)

Long read tools

- Evolving field: new tools and updates

nature reviews genetics

<https://doi.org/10.1038/s41576-025-00828-z>

Review article

 Check for updates

Transcriptomics in the era of long-read sequencing

Carolina Monzó ^{1,2}, Tianyuan Liu ^{1,2} & Ana Conesa ^{1,2}

Table 2 | Software tools for long-read RNA sequencing data processing and downstream analysis

Tool	Usage	Platform	Refs.
DeepConsensus	Basecalling	PacBio	82
Dorado	Basecalling	ONT	https://nanoporetech.com/platform/accuracy/
Minimap2	Mapping	PacBio, ONT	84
uLTRA	Mapping	PacBio, ONT	85
Graphmap2	Mapping	PacBio, ONT	86
deSALT	Mapping	PacBio, ONT	87
2passtools	Mapping	PacBio, ONT	88
Splam	Mapping	PacBio, ONT	94
Magic-BLAST	Mapping	PacBio, ONT	90
FLAIR	Read clustering Transcript identification and quantification	PacBio, ONT	92
Mandalorion	Read clustering Transcript identification and quantification	PacBio, ONT	31
IsoQuant	Splicing and intron graph construction Alignment correction Transcript identification and quantification Gene-level visualization	PacBio, ONT	32
IsoTools	Splicing and intron graph construction Transcript identification and quantification Differential isoform expression Differential splicing analysis Quality control visualization Gene-level visualization	PacBio, ONT	37
Isosceles	Splicing and intron graph construction Alignment correction Transcript identification and quantification at single-cell, pseudo-bulk and bulk resolution	PacBio, ONT	93
StringTie2	Splicing and intron graph construction Transcript identification and quantification	PacBio, ONT	94
ESPRESSO	Read classification Splice junction correction Transcript identification and quantification Gene-level visualization	PacBio, ONT	95
TALON	Read classification Transcript identification and quantification	PacBio, ONT	96
FLAMES	Read classification Transcript identification and quantification	PacBio, ONT	97
Bambu	Read classification Machine-learning filtering for true novel transcripts Transcript identification and quantification	PacBio, ONT	33
Freddie	Read clustering Transcript identification and quantification	PacBio, ONT	98
LyRic	Transcript identification and quantification	PacBio, ONT	https://github.com/guigolab/LyRic
IsoSplitter	Read clustering Transcript identification and quantification	PacBio, ONT	99
RNA-Bloom2	Splicing and intron graph construction Transcript identification and quantification	PacBio, ONT	100
rnaSPAdes	Splicing and intron graph construction Transcript identification and quantification	PacBio, ONT	101

Long read tools

<https://long-read-tools.org/>



Downstream analysis

- ONT:
 - Workflows on [EPI2ME](#)
- PacBio:
 - Pbbioconda
- nf-core
 - <https://nf-co.re/bacass>
 - <https://nf-co.re/viralrecon>
 - <https://nf-co.re/isoseq>
 - <https://nf-co.re/mag>
 - <https://nf-co.re/nanoseq>
 - <https://nf-co.re/ampliseq>



PacBio Secondary Analysis Tools on Bioconda

