MPXV ONT Galaxy Workflow

Presenter name

Objectives

- Import and run the MPXV (Mpox) ONT Galaxy workflow

- Understand the outputs of the pipeline

- Download and save the consensus genomes

- Perform quality control (QC) on the consensus genomes

The workflow

edited about 1 month ago

MPXV (Mpox) A 2025-01-29T23:42:10.546Z **NT** ✏

MPXV (Mpox) sequence analysis pipeline for ONT data using the ARTIC minion pipeline (v.1.6.0).
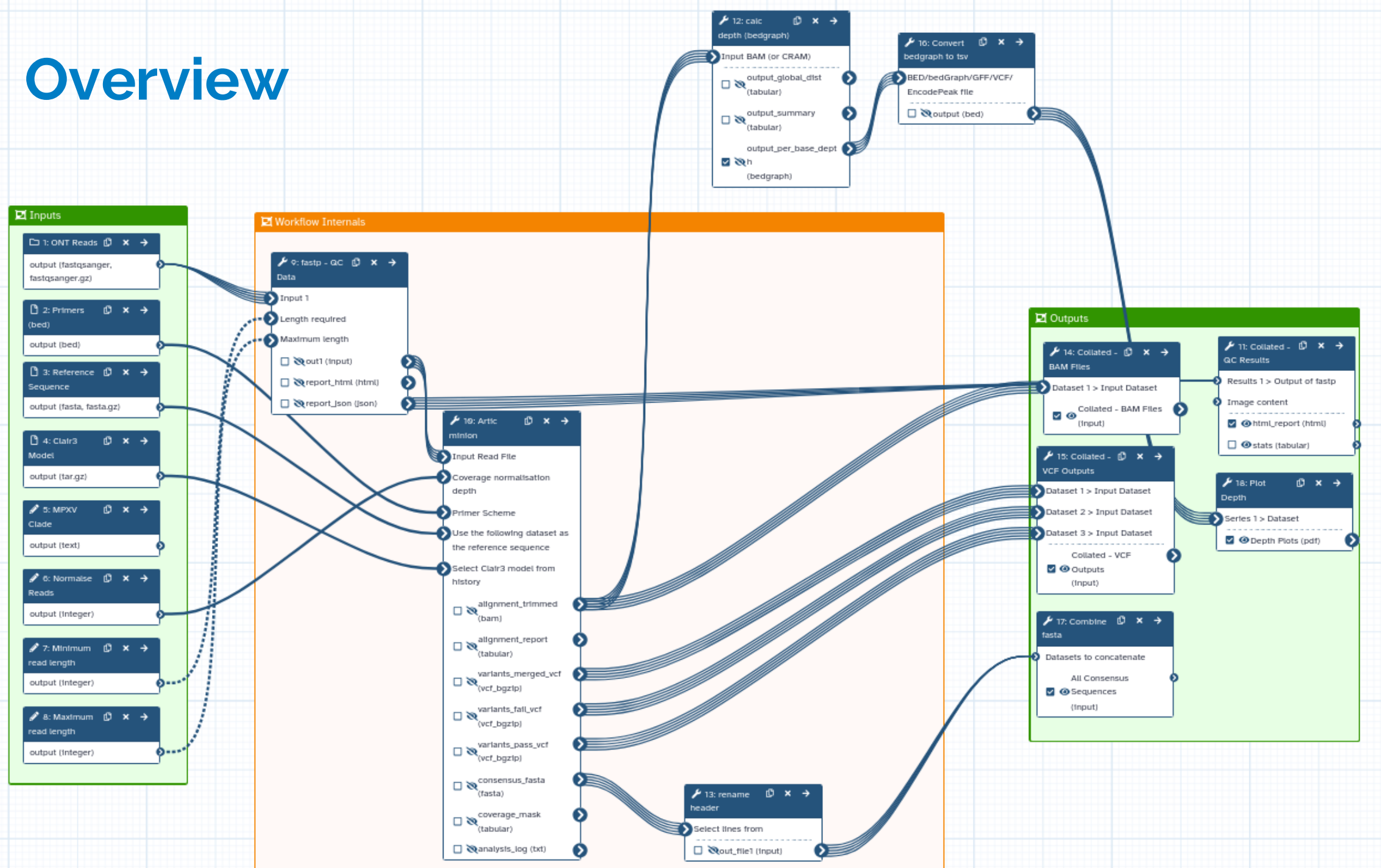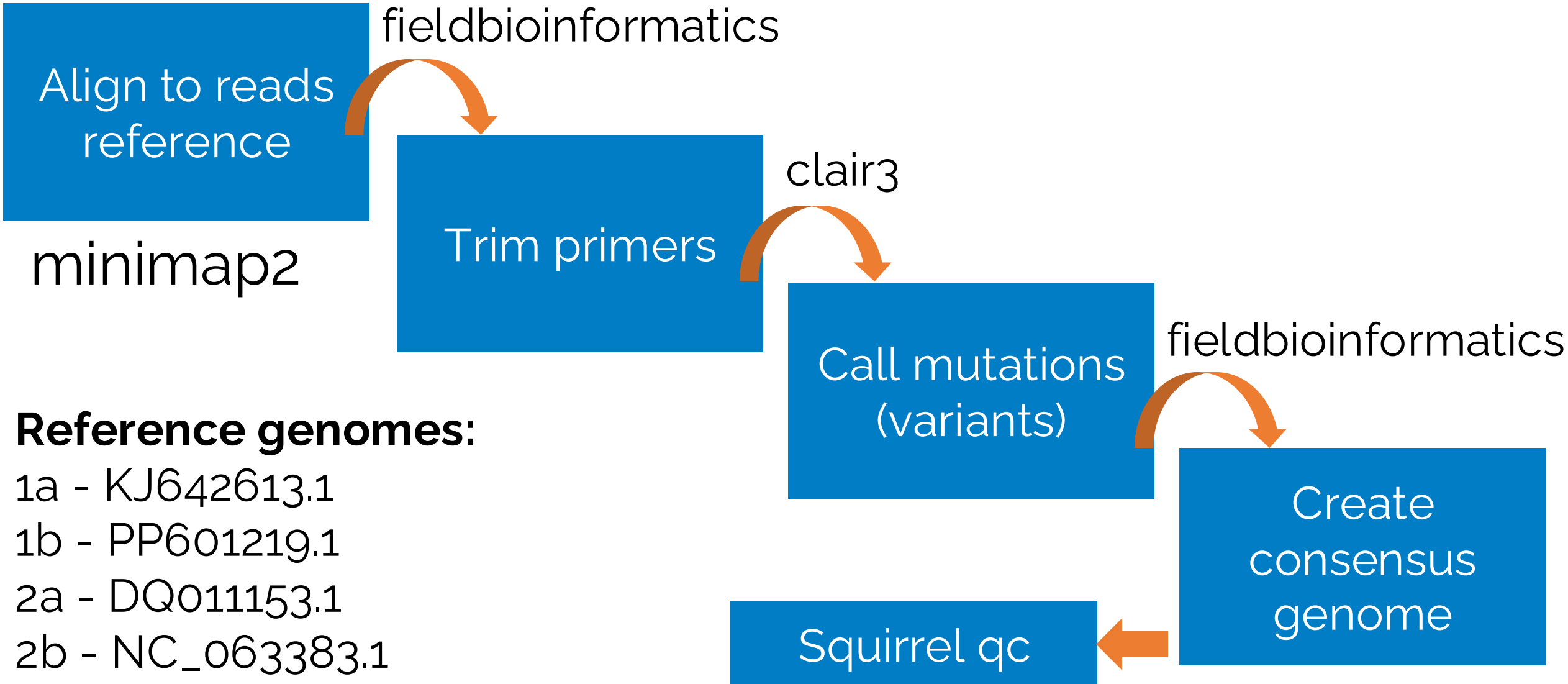
ONT ✖   ARTIC ✖   MPXV ✖   Mpox ✖

Add Tags 🏷

- Design for ONT platform

- Performed genome mapping, variant calling and consensus assembly

Overview

Pipeline Steps - simplified

Align to reads reference

fieldbioinformatics

minimap2

Trim primers

clair3

Call mutations (variants)

fieldbioinformatics

Create consensus genome

Squirrel qc

Reference genomes:
1a – KJ642613.1
1b – PP601219.1
2a – DQ011153.1
2b – NC_063383.1

Let's run the pipeline!

Let's run the pipeline! ... not yet

Pipeline Inputs

Let's run the pipeline! ... not yet

Workflow inputs for MPXV pipeline:
- Reads (fastq.gz)
- Primer file (tab)
- Reference (fasta)
- Clair3 model (zip)
- Specify Clade (Clade1/2)

Pipeline inputs – fastq files

Workflow inputs for MPXV pipeline:

- ✅ Reads (fastq.gz)
- ❏ Primer file (tab)
- ❏ Reference (fasta)
- ❏ Clair3 model (zip)
- ❏ Specify Clade (Clade1/2)

Pipeline inputs – primer/reference files

Workflow inputs for MPXV pipeline:

☑ Reads (fastq.gz)

❑ Primer file (tab) **Get Primers from**

❑ Reference (fasta) https://labs.primalscheme.com

❑ Clair3 model (zip)

❑ Specify Clade (Clade1/2)

Pipeline inputs – primer/reference files

Workflow inputs for MPXV pipeline:

- ✅ Reads (fastq.gz)
- ✅ Primer file (tab) **Get Primers from**
- ✅ Reference (fasta) https://labs.primalscheme.com

IMPORTANT! Ensure you download the fasta/primers of the same scheme.
Do not mix different schemes together!

Pipeline inputs – primer/reference files

https://labs.primalscheme.com/detail/artic-inrb-mpox/2500/v1.0.0/

artic-inrb-mpox / 2500 / v1.0.0

primer.bed ⬤   download

Toggle to show

reference.fasta ⬤   download

Toggle to show

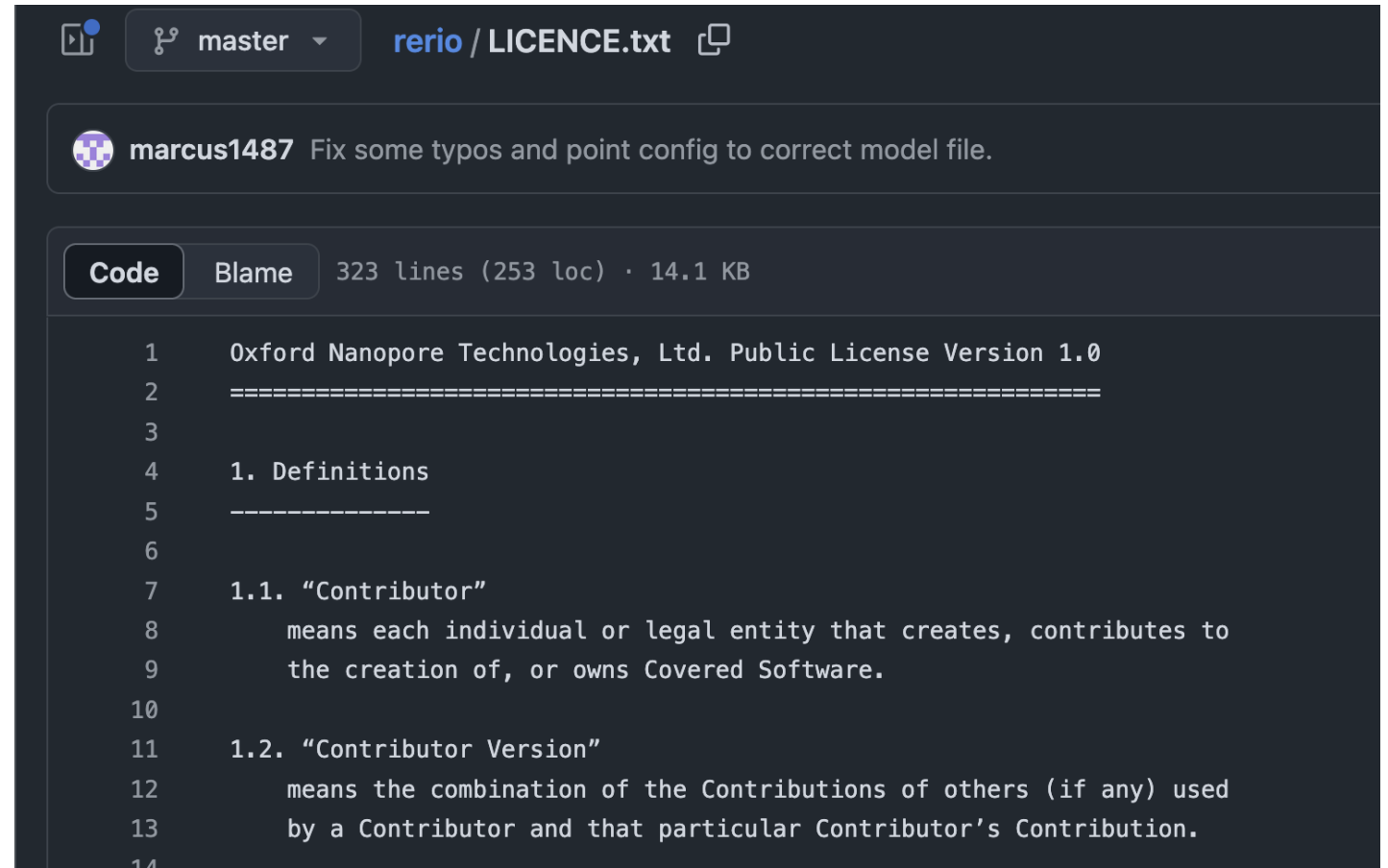# Let's run the pipeline – clair3 models

Workflow inputs for MPXV pipeline:

☑ Reads (fastq.gz)

☑ Primer file (tab)

☑ Reference (fasta)

❑ Clair3 model (zip)

❑ Specify Clade (Clade1/2)

Get clair3 model from

https://cdn.oxfordnanoportal.com/software/analysis/models/c
_e82_400bps_hac_v500.tar.gz (right click save as)

Rerio Clair3 models are not Open Source

Licensed – for research, public health purposes only.

https://github.com/nanoporetech/rerio/blob/master/LICENCE.t

Which model to download?
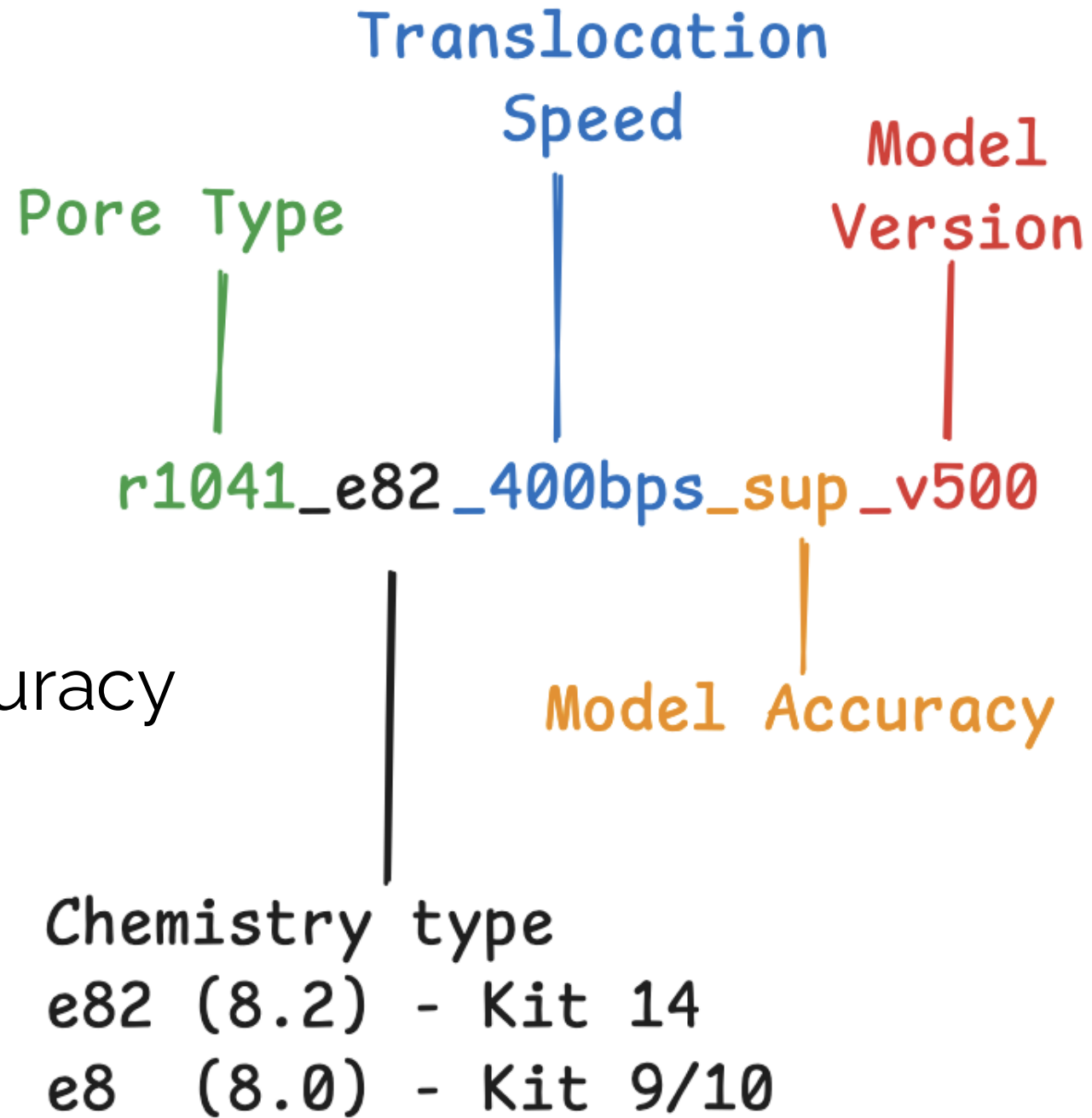
Decoding model names

Which is our model?

We used:
- Chemistry version K14
- Flow cell R10.4
- Model accuracy was high accuracy

Pore Type

Translocation Speed

Model Version

r1041_e82_400bps_sup_v500

Model Accuracy

Chemistry type
e82 (8.2) - Kit 14
e8 (8.0) - Kit 9/10

Where to download?

See table:

https://github.com/nanoporetech/rerio?tab=readme-ov-file#clair3-models

Why do we need a model?

Basecalling is the process of translating the raw, electrical signals captured by our sequencing devices into a sequence of nucleotides (DNA or RNA). This is like voice recognition, where spoken language is translated into written text. In the context of genomic sequencing, our devices detect changes in electrical current caused by DNA or RNA strands passing through a nanopore.
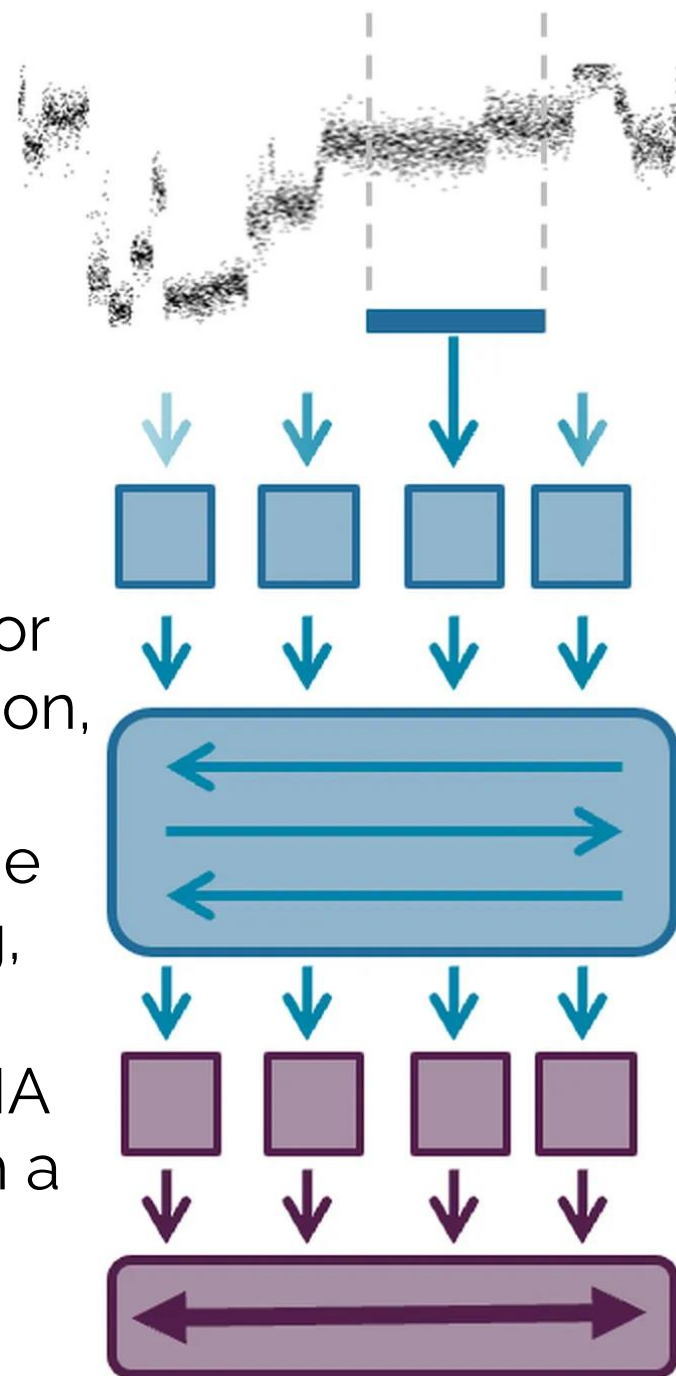
Parameters learned from training data

Extraction of blocks of features

Bidirectional information flow

Multi-base prediction

Decode to sequence

Let's run the pipeline!

Workflow inputs for MPXV pipeline:
- ✅ Reads (fastq.gz)
- ✅ Primer file (tab)
- ✅ Reference (fasta)
- ✅ Clair3 model (zip)
- ❑ Specify Clade (Clade1/2)

Choosing Clade

- Choosing the correct clade is important

  - MPXV clades have major genomic differences.

- If you are unsure, run the pipeline twice:

  - once using clade 1 as reference

  - once using clade 2 as reference

- Compare results

  - Check greatest coverage % and best depth

Let's run the pipeline!

Workflow inputs for MPXV pipeline:

- ✅ Reads (fastq.gz)
- ✅ Primer file (tab)
- ✅ Reference (fasta)
- ✅ Clair3 model (zip)
- ✅ Specify Clade (Clade1/2)

Finally!

Let's run the pipeline!

Your inputs

Need to be this →

History

search datasets

MPXV Training

261 MB | 4 | 137 | 45

143: **r1041_e82_400bps_hac_v5 00.tar.gz**

142: **reference.fasta**

141: **primer.bed**

92: **raw_data**

a list with 2 **fastqsanger.gz** datasets

Let's run the pipeline!

edited about 9 hours ago

≣ workflow runs: 2 ☆ ▾

MPXV (Mpox) Analysis pipeline for ONT

MPXV (Mpox) pipeline for ONT platform. Using the latest Fieldbioinformatics Artic Minion pipeline that uses clair3.

ONT ✖ MPXV ✖ Mpox ✖ #amplicon ✖ Virus ✖

Add Tags 🏷

📋 Copy   ⬇ Download   ◄ Share   ✎ Edit   ▶

Click the play button
To enter workflow

▶

🕐 *edited about 9 hours ago*   ☰ workflow runs: 2

MPXV (Mpox) pipeline for ONT platform. Using the latest Fieldbioinformatics Artic Minion pipeline that uses clair3.
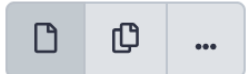
ONT   MPXV   Mpox   #amplicon   Virus

ONT Reads *

| 📁 | 📂 | + | 192: fastp on collection 191: Read 1 output | ▾ |

accepted formats ▾

ONT reads from ARTIC assay with fastqsanger encoding

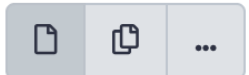**Primers (bed)** *

| 📄 | 🗐 | ... | 141: primer.bed (as bed) | ▾ |

accepted formats ▾

Primer (bed) file for use as reference. Must be the same reference as the input Reference sequence input. Bed files can be found: https://github.com/artic-network/artic-mpxv-nf/tree/master/data/primer_schemes/MPXV
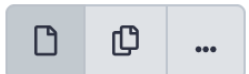
Reference Sequence *

| 📄 | 🗐 | ... | 230: all.consensus.fasta | ▾ |

accepted formats ▾

Fasta file for use as reference. Must be the same reference as the input bed file. Reference sequences files can be found: https://github.com/artic-network/artic-mpxv-nf/tree/master/data/primer_schemes/MPXV

Clair3 Model *

| 📄 | 🗐 | ... | 144: r1041_e82_400bps_hac_v500.tar.gz | ▾ |

accepted formats ▾

Model must be tar.gz

MPXV Clade *

Clade I ▾

Choose a Clade for downstream phylogenetic processing.

Normalse Reads *

400 ⌃⌄

Sample at most this number of reads per amplicon and strand. default is 400

Minimum read length - optional

200 ⌃⌄

Primer scheme

Maximum read length - optional

2600 ⌃⌄

Primer Scheme

Expand to full workflow form.

▶ Run Workflow

Click to run the pipeline
Watch the pipeline deploy!

That was a lot! Well done! 10 minute break.

MPXV pipeline outputs

Workflow inputs for MPXV pipeline:

- ✅ Reads (fastq.gz)
- ✅ Primer file (tab)
- ✅ Reference (fasta)

IMPORTANT! Ensure you download the fasta/primers of the same scheme. **Do not mix different schemes together!**

Questions? + Resources

- ONT - How basecalling works
  - https://nanoporetech.com/platform/technology/basecalling
- Clair3 models:
  - https://github.com/nanoporetech/rerio/tree/master/clair3_models
- To understand model names read this:
  - https://github.com/nanoporetech/dorado?tab=readme-ov-file#decoding-dorado-model-names