



Public Health
Agency of Canada

Agence de la santé
publique du Canada

Canada

SARS-CoV-2 Recombinant Detection Tutorial

Katherine Eaton

2022 April 27

PROTECTING AND EMPOWERING CANADIANS
TO IMPROVE THEIR HEALTH



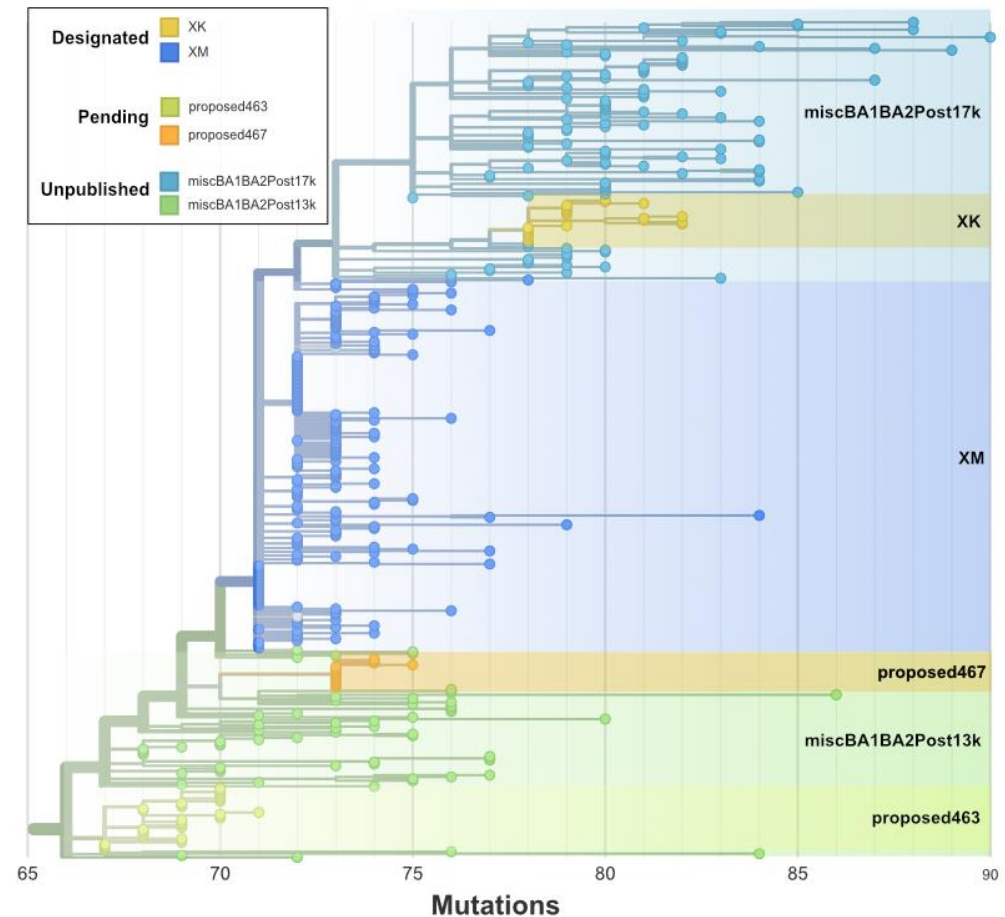
Overview

- **Tutorial Zip**
 - data : sequences and metadata
 - resources: metadata of pango-designation issues
 - results: output of detection tools
 - slides: presentation slides
- **Tool Demo* and Comparison (15 min)**
 - Nextclade: Identify putative recombinants.
 - sc2rf (“scarf”): Plot recombination breakpoints.
 - UShER: Refine lineage assignments.

**I will demo the web browser versions. But all tools have command-line versions for high-throughput analyses.*

Why are recombinants hard to classify?

1. Algorithms assume that recombination is not occurring.
2. Rapidly evolving nomenclature that varies by program.
 - Designated (X*)
 - Pending (proposed*)
 - Unpublished (misc*)
3. Not all recombinant lineages are in public databases.
 - XK, XL



Nextclade: Identify putative recombinants

Nextclade^{v1.14.0}

Clade assignment, mutation calling, and sequence quality checks

<https://clades.nextstrain.org/>

1. Select SARS-CoV-2
2. Upload **data/sequences.fasta**.
3. Click “Run”.

Selected pathogen

SARS-CoV-2

Reference: Wuhan-Hu-1/2019 (genbank: MN908947)

Updated: 2022-04-08 12:00 (UTC)

Change

[Recent dataset updates](#)

➤ Customize dataset files

Provide sequence data

File Link Text

Drag & Drop a file here

or



Select a file

☐ Run automatically

[Load example](#)

Run

Nextclade: Identify putative recombinants

Note: Not the same method as pangolin!

Correct

Wrong

ID	Sequence name	QC	Clade	Pango lineage (Nextclade)	Mut.	non-ACGTN	Ns
0	✓ XM_example_1	N M P C F S	recombinant	XM	63	0	37
1	✓ XM_example_2	N M P C F S	recombinant	XM	62	0	898
2	✓ proposed467_example_1	N M P C F S	recombinant	XM	56	0	880
3	✓ proposed467_example_2	N M P C F S	recombinant	XJ	46	0	3070
4	✓ miscBA1BA2Post17k_example_1	N M P C F S	recombinant	XM	55	11	376
5	✓ miscBA1BA2Post17k_example_2	N M P C F S	recombinant	XM	67	0	0

Private Mutations

Excess Ns



Use the download button in the top right to download the alignment (nextclade.aligned.fasta) for a later step.

sc2rf: Plot recombination breakpoints

INSTALL

```
git clone https://github.com/lenaschimmel/sc2rf  
cd sc2rf  
pip install -r requirements.txt
```

RUN

```
./sc2rf.py results/nextclade.aligned.fasta --csvfile sc2rf.csv
```

OUTPUT

```
genes          1a 1b S 3M 6 N  
  
Omicron / BA.1 / 21K ..G..GA.....G.CT....TT..AC...AAAAT.G....  
Omicron / BA.2 / 21L GT..AT..TGTTTA.T..TTTGT..AG..GAC.....T.TCTCC  
  
XM_example_1    ..G..GA.....G.CT..TGT..AG..GAC.....T.TCTCC  
XM_example_2    ..G..GA.....G.CT..NNT..AG..GAC.....T.TCTCC  
  
proposed467_example_1 ..G..GA.....G.CT.TTTGT..AG..GACNN...T.TCTCC  
proposed467_example_2 ..G..GA.....G.CT.TTNT..AGNNNNNNN...T.TCTCC  
  
miscBA1BA2Post17k_example_1 KYG..KR...YYRYYY.YNNT..AG..GAC.....T.TCTCC  
miscBA1BA2Post17k_example_2 ..G..GA.....G.CT....T..AG..GAC.....T.TCTCC  
  
made with Sc2rf - available at https://github.com/lenaschimmel/sc2rf
```

sc2rf: Plot recombination breakpoints

1. Compare the breakpoints to those in [pango-designation repository](#) issues.
2. Explore the breakpoints from sc2rf with the “breakpoints” column in:
 - **resources/pango-designation_issues_update.tsv**
3. Example
 - miscBA1BA2Post17k has a breakpoint between 20055 and 21618.
 - Matches breakpoints in [Issue#477](#) and [Issue#514](#).

BA.2/BA.1 Recombinant with breakpoint at 20055-21618 detected in travellers arriving in Hong Kong, February 2022 #514

 Open Koohoko opened this issue 25 days ago · 6 comments

Potential BA.1/BA.2 Recombinant Lineage with Likely Breakpoint at NSP16 (4 Seqs in USA-CA as of 2022-03-18) #477

 Closed emily-smith1 opened this issue on Mar 18 · 7 comments

UShER: Refine lineage assignments

1a. Run on UCSC server:

<https://genome.ucsc.edu/cgi-bin/hgPhyloPlace>

1b. Run on local machine:

<https://shusher.gi.ucsc.edu/>

2. Upload **data/sequences.fasta**

UShER: Ultrafast Sample placement on Existing tRee

Place your SARS-CoV-2 sequences in a global phylogenetic tree

Select your FASTA, VCF or list of sequence names/IDs: No file chosen

or paste in sequence names/IDs:

Phylogenetic tree version:

Number of samples per subtree showing sample placement:

[More example files](#)

UShER: Refine lineage assignments

1. Lineages assignments by different algorithms currently differ more than they agree.
2. For pending or unpublished recombinants, UShER is the most accurate.
3. For designated recombinants, UShER and Nextclade are comparable.
4. Pangolin (PLEARN and PUSHER mode) is the most inaccurate.

	UShER Assignment	pangolin Assignment	Nextclade Assignment		
Fasta Sequence	Pango lineage (?)	Lineage of neighbor (?)	Subtree number (?)	#Maximally parsimonious placements (?)	Nextclade Lineage
XM_example_1	XM	XM	1 (view in Nextstrain)	1	XM
XM_example_2	XM	Unassigned	1 (view in Nextstrain)	1	XM
proposed467_example_1	proposed467	BA.1	1 (view in Nextstrain)	1	XM
proposed467_example_2	proposed467	BA.1	1 (view in Nextstrain)	1	XJ
miscBA1BA2Post17k_example_1	miscBA1BA2Post17k	Unassigned	1 (view in Nextstrain)	1	XM
miscBA1BA2Post17k_example_2	miscBA1BA2Post17k	Unassigned	1 (view in Nextstrain)	1	XM

BONUS Slides

- **Updating pango-designations metadata:**

```
pip install click pandas numpy  
resources/pango-designation_issues.py > issues.tsv
```

- **Running the tutorial with [ncov-recombinant](#):**

```
git clone --recursive https://github.com/ktmeaton/ncov-recombinant.git  
cd ncov-recombinant
```

```
mamba env create -f workflow/envs/environment.yaml  
conda activate ncov-recombinant
```

```
snakemake --profile profiles/tutorial
```