

# SARS-CoV-2 Phylogenetics in Norwich

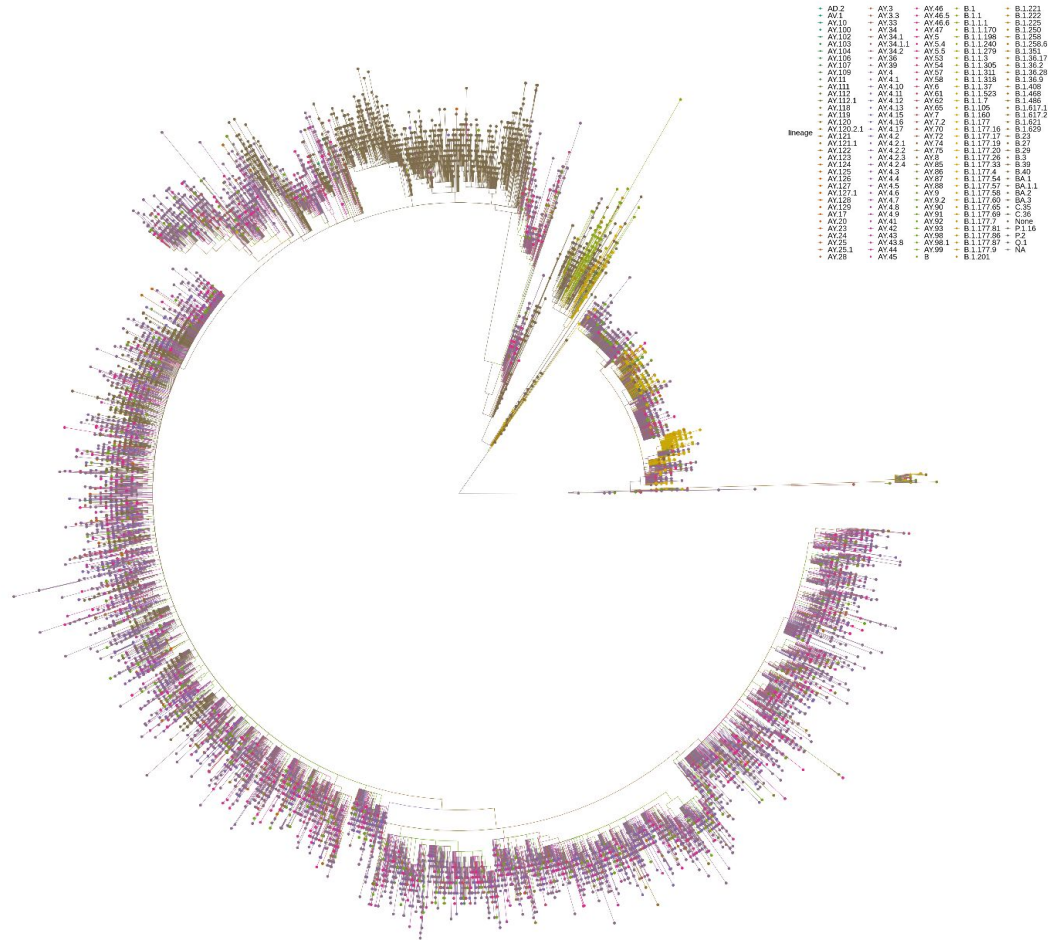
Madeline Iseminger  
May 3, 2022

# Contents

1. Introduction
2. COG-UK data processing pipeline
3. Nextflow-generated ML trees + analysis
  - a. First wave: January to June 2020
  - b. Most recent: January to March 2022
4. Civet Newick tree
5. NORW in context (vs. CAMB)

## Investigate phylogenetic relationships for SARS-CoV-2 samples from Norwich, England

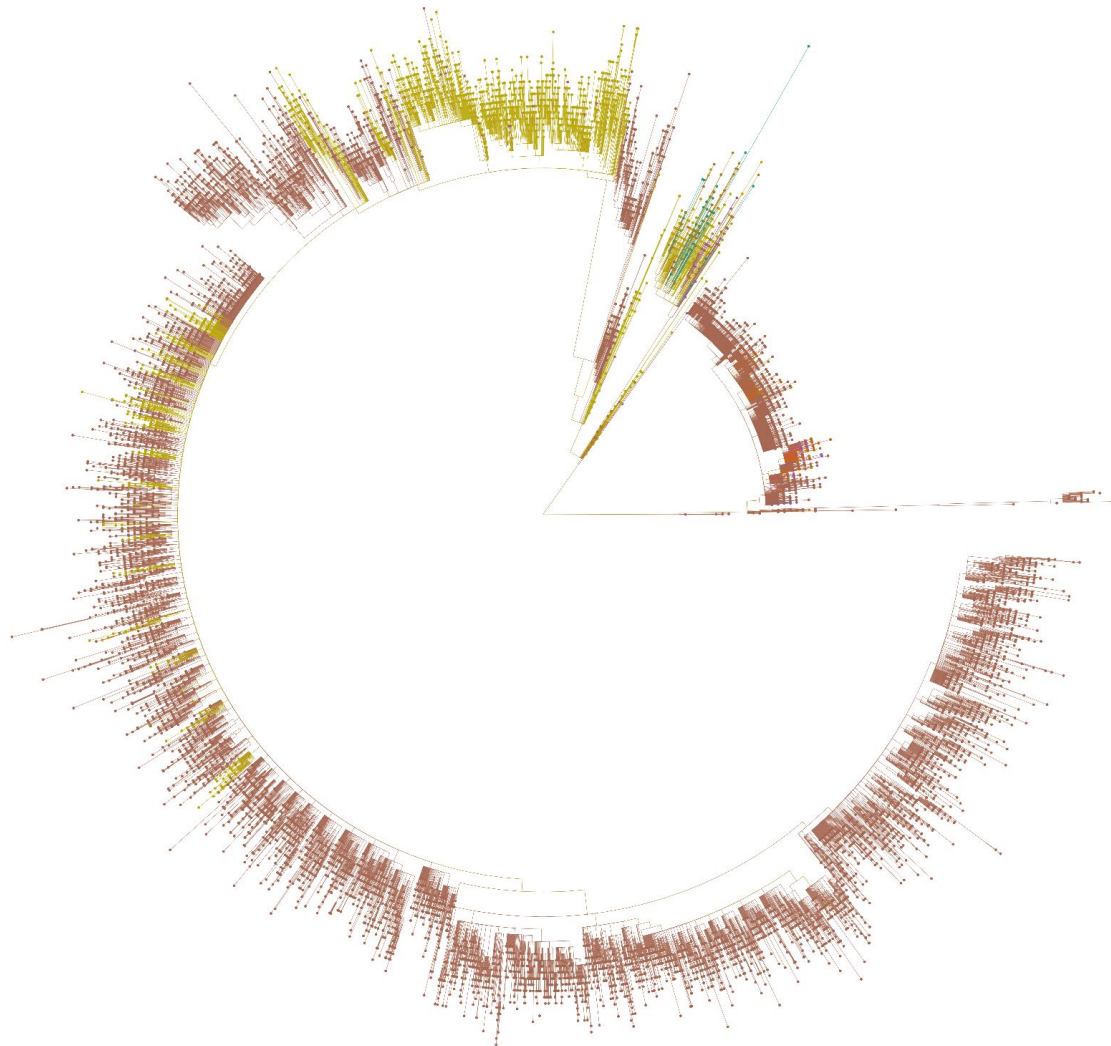
- ❖ the 10,000 most diverse sequences for NORW in the Civet Newick tree (v. 2022-04-18) are shown at right
- ❖ the full tree has 28827 tips from 195 PANGO lineages



# NORW tree

lineage

+	B.1.1	+	B.1.177*	+	Omicron
+	B.1.1.7	+	Delta	+	Other



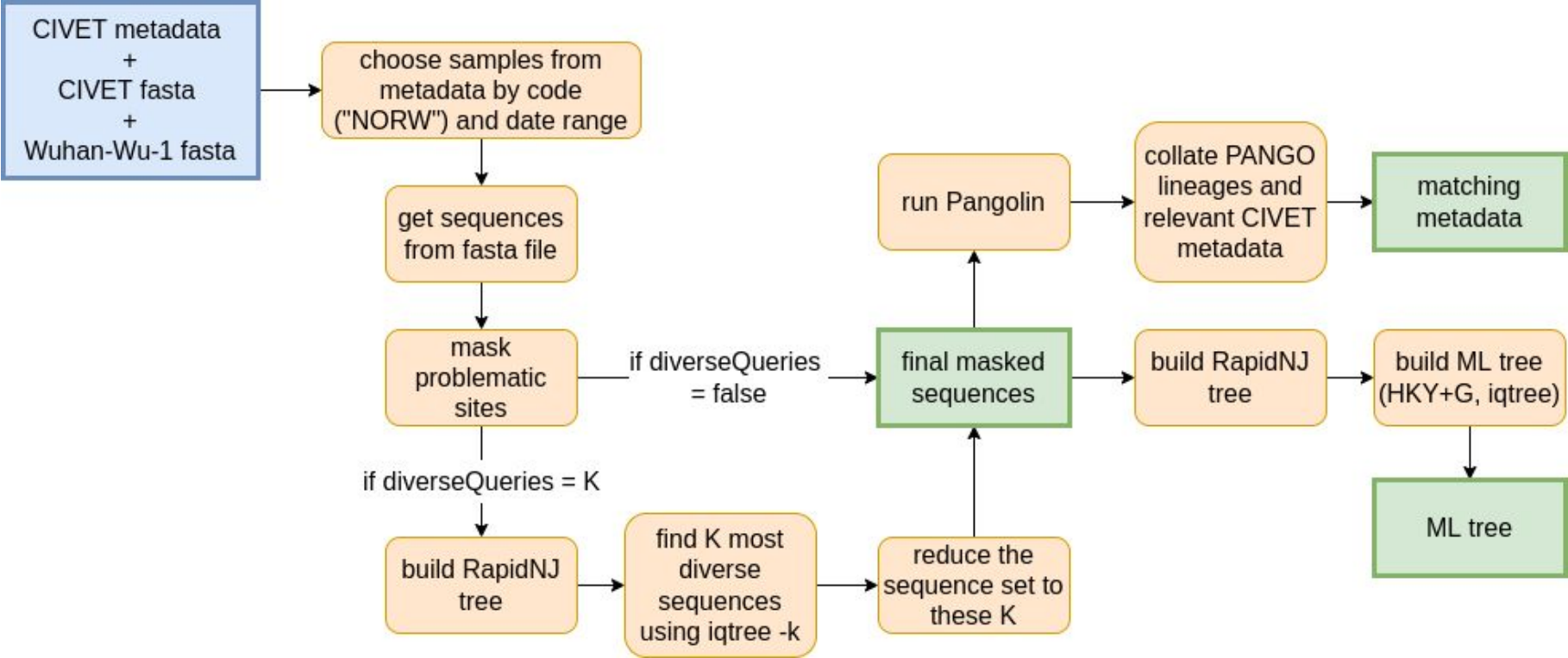
# Data Processing

I wrote a Nextflow pipeline to generate ML trees from the Civet metadata and fasta files.

```
(base) ubuntu@madeline-01:~/nf-test/sequencing_centre_trees$ nohup nextflow run sequencing_centre_trees.nf
--start_date "2020-01-01" --end_date "2020-06-30" --diverseQueries false --centre "NORW" --max_samples_per_month
false
```

```
executor > local (9)
[9c/5c73af] process > chooseSamples (1) [100%] 1 of 1 ✓
[3f/91b862] process > getFastaRecords (1) [100%] 1 of 1 ✓
[be/95bb93] process > maskProblematicSites (1) [100%] 1 of 1 ✓
[cc/89715c] process > removeWuhanHu1 (1) [100%] 1 of 1 ✓
[-] process > reduceQueries -
[9c/46b7f2] process > buildFastTree (1) [100%] 1 of 1 ✓
[b4/37198f] process > refineTree (1) [100%] 1 of 1 ✓
[8e/8155f0] process > runPangolin (1) [100%] 1 of 1 ✓
[ab/dd0377] process > getCOGUKMetadata (1) [100%] 1 of 1 ✓
[e3/ca6bd1] process > joinMetadata (1) [100%] 1 of 1 ✓
Completed at: 28-Mar-2022 22:36:52
Duration : 11h 25m 26s
CPU hours : 14.1
Succeeded : 9
```

# Data Processing

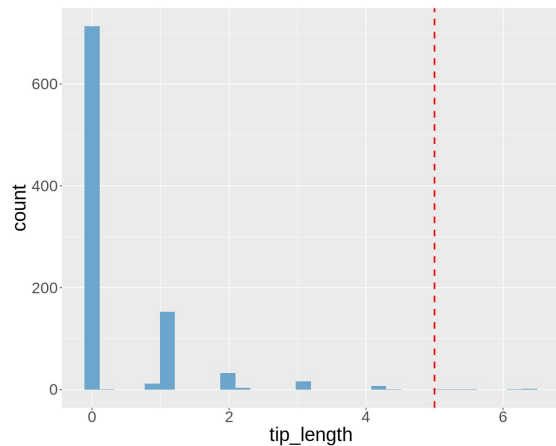
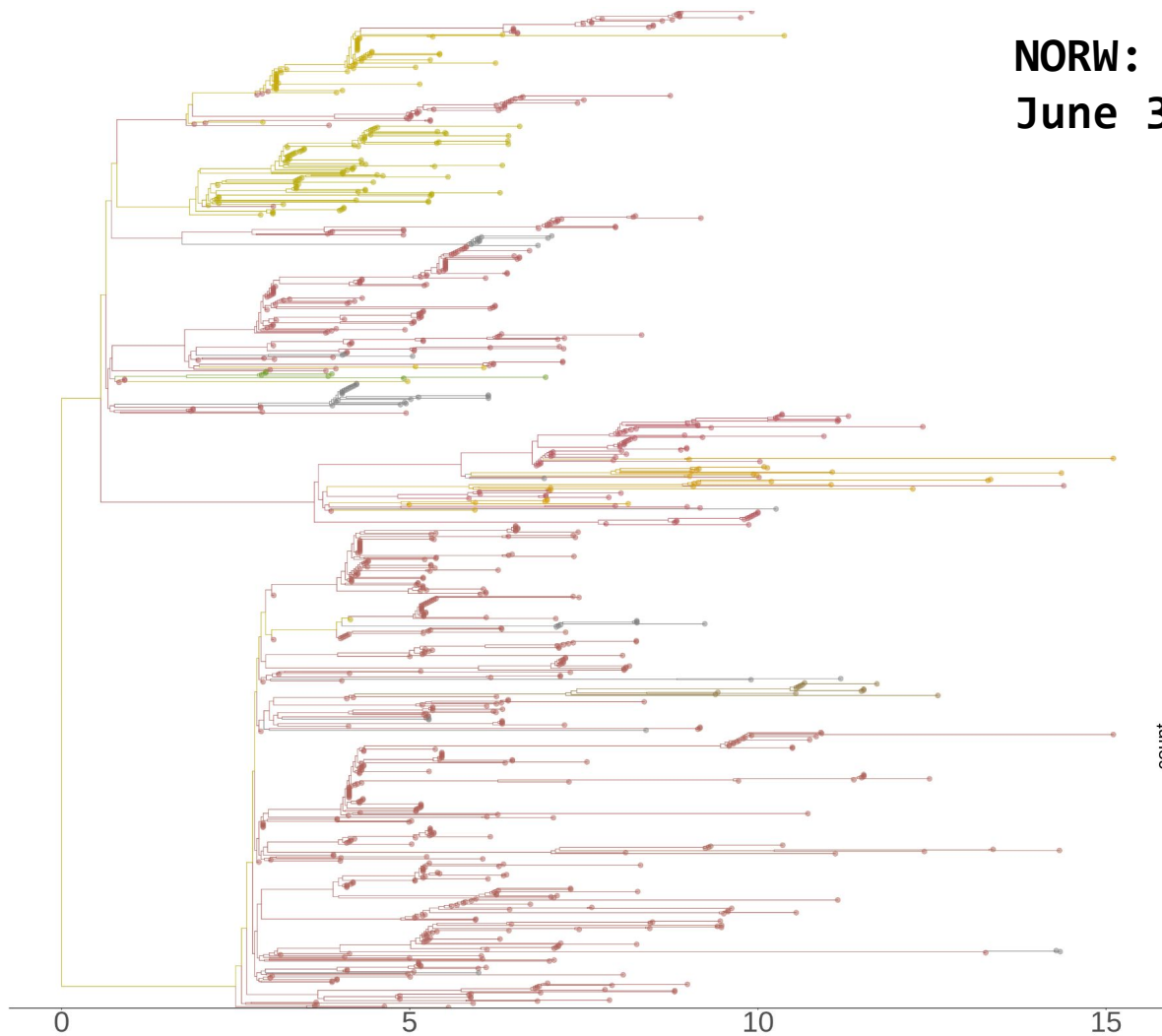


NORW: January 1, 2020 to  
June 30, 2020

lineage

B  
B.1  
B.1.1  
B.1.1.1  
B.1.1.3  
B.1.105  
B.1.201  
B.1.222  
B.1.225  
B.1.250  
B.23  
B.27  
B.29  
B.3  
B.39  
B.40

- ❖ 946 tips
- ❖ Removed all tips with length > 5 SNPs
- ❖ Only 6 tips were removed



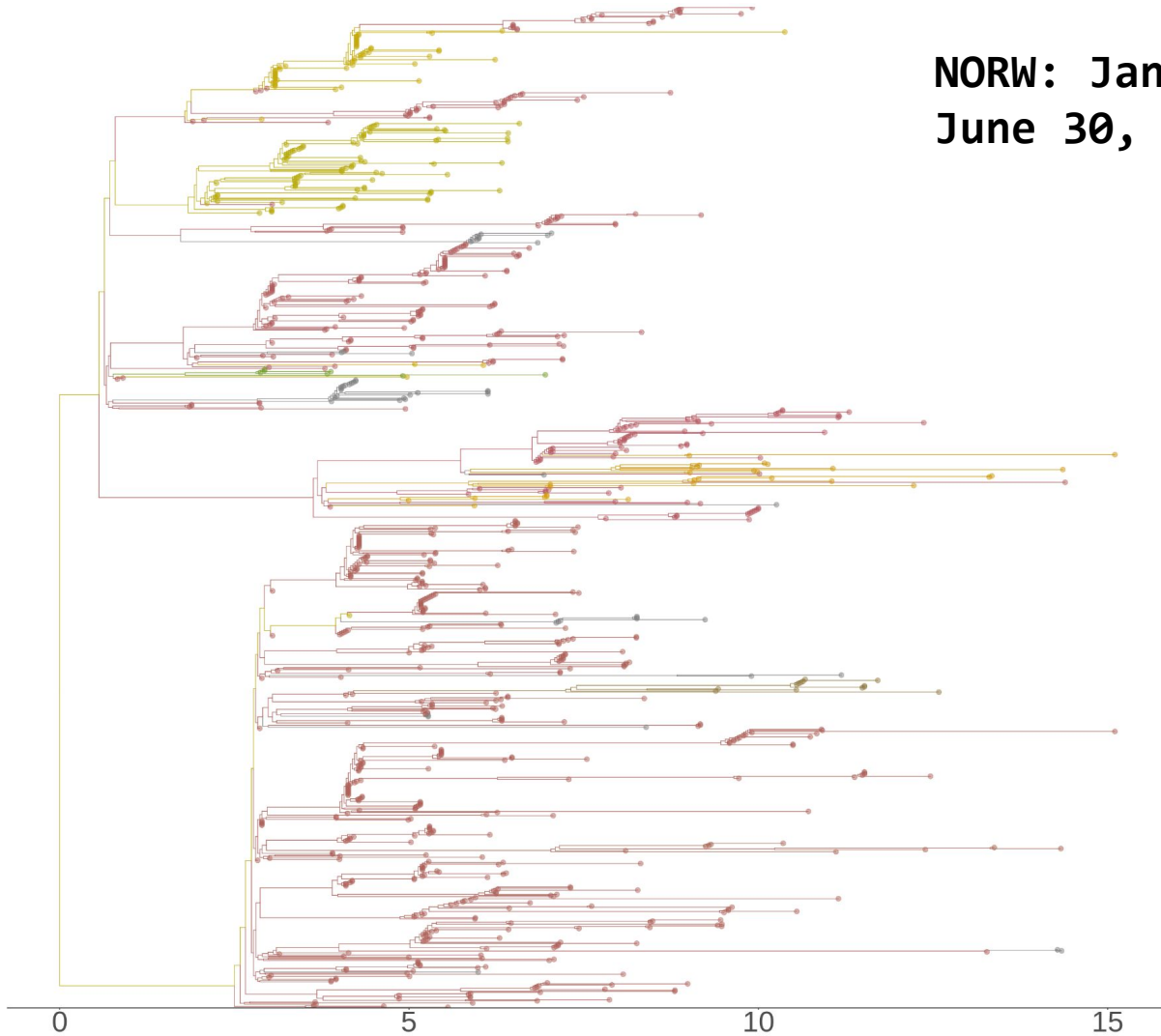


**NORW: January 1, 2020 to  
June 30, 2020**

lineage

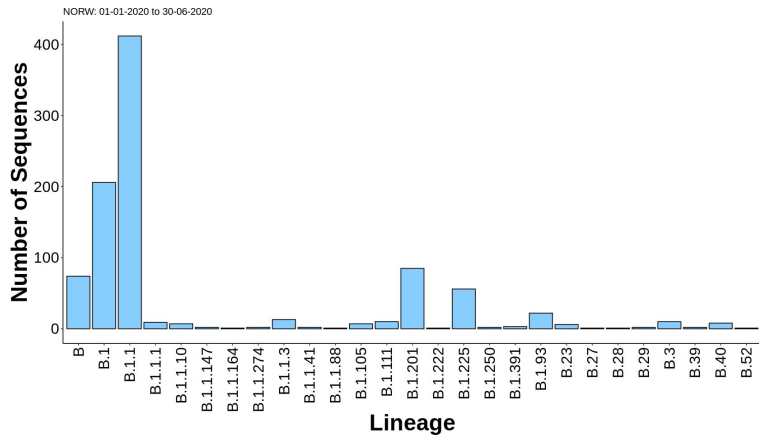
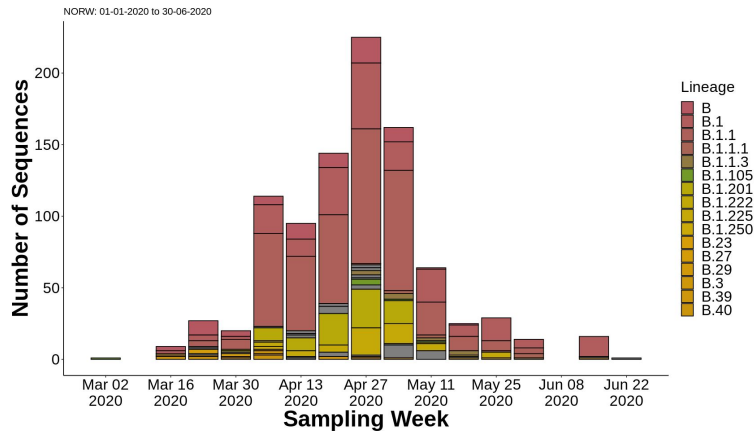
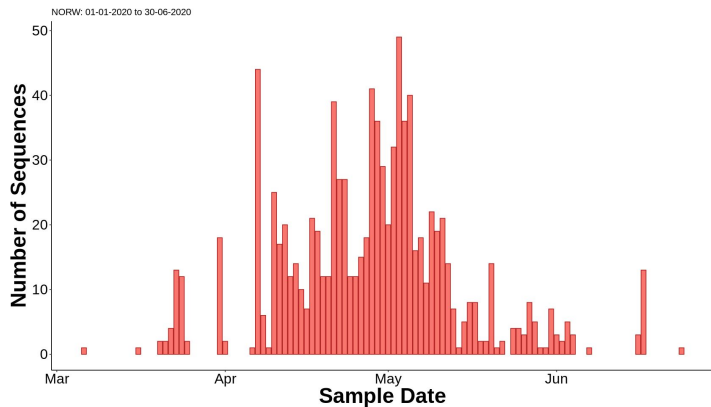
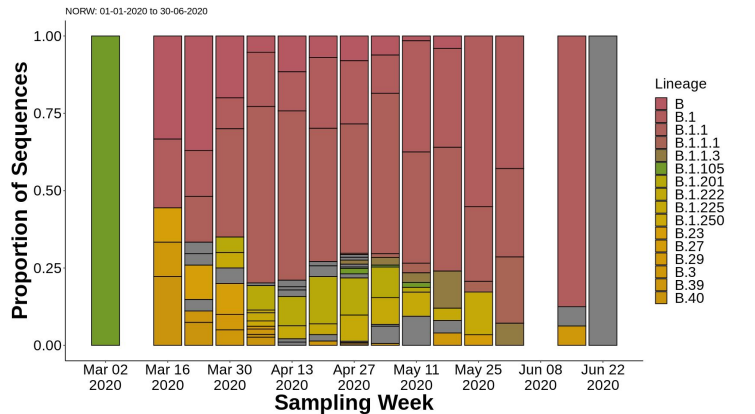
B  
B.1  
B.1.1  
B.1.1.1  
B.1.1.1.3  
B.1.105  
B.1.201  
B.1.222  
B.1.225  
B.1.250  
B.23  
B.27  
B.29  
B.3  
B.39  
B.40

❖ This is the tree  
with outliers  
removed  
❖ 940 tips in total



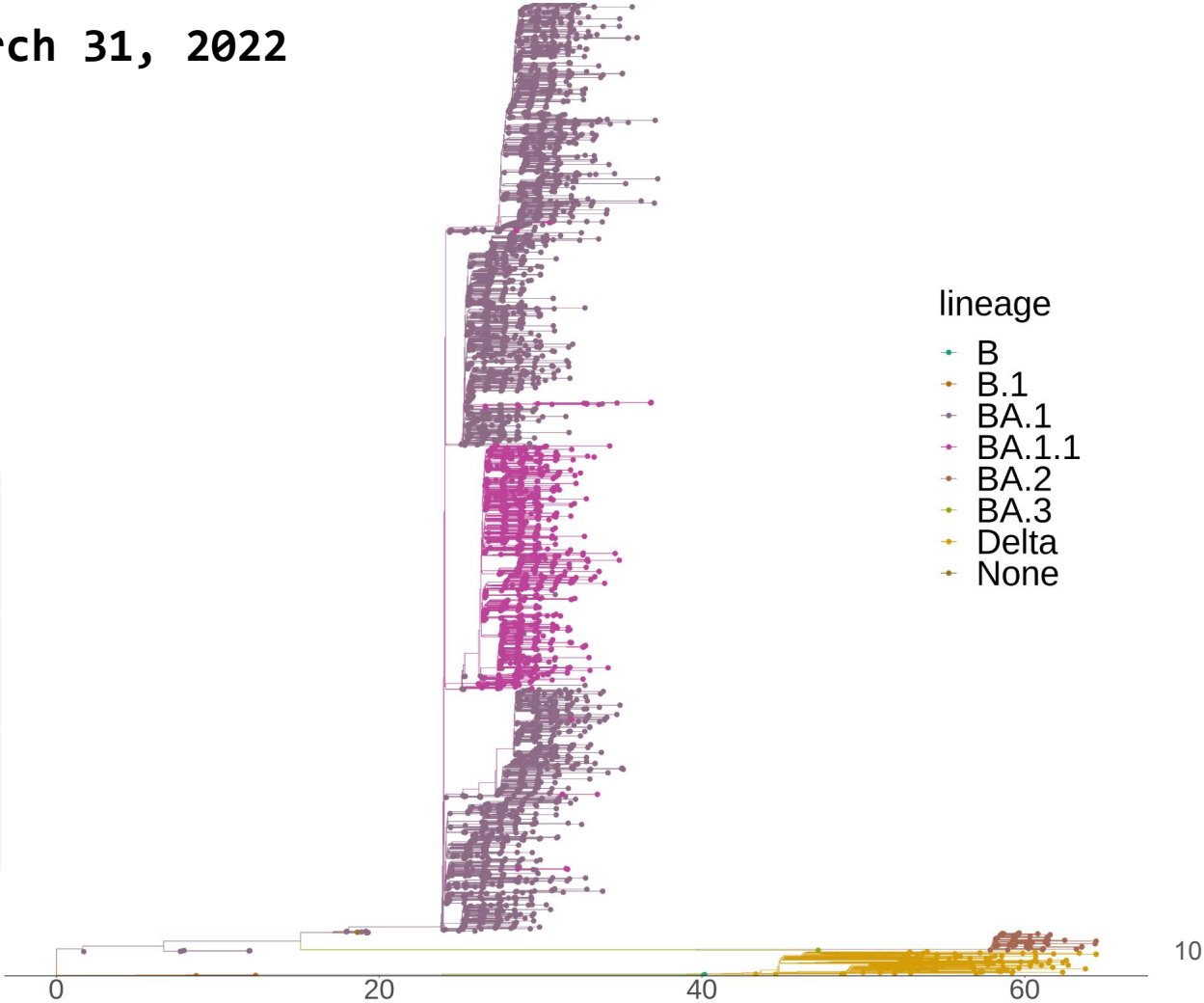
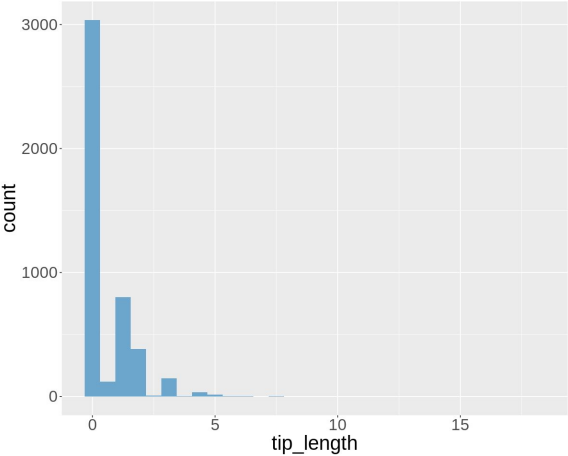


# NORW: January 1 to June 30, 2020



# NORW: January 1 to March 31, 2022

- ❖ 4560 tips
- ❖ Removed all samples with tip length > 5 SNPs
- ❖ Only 7 tips were removed

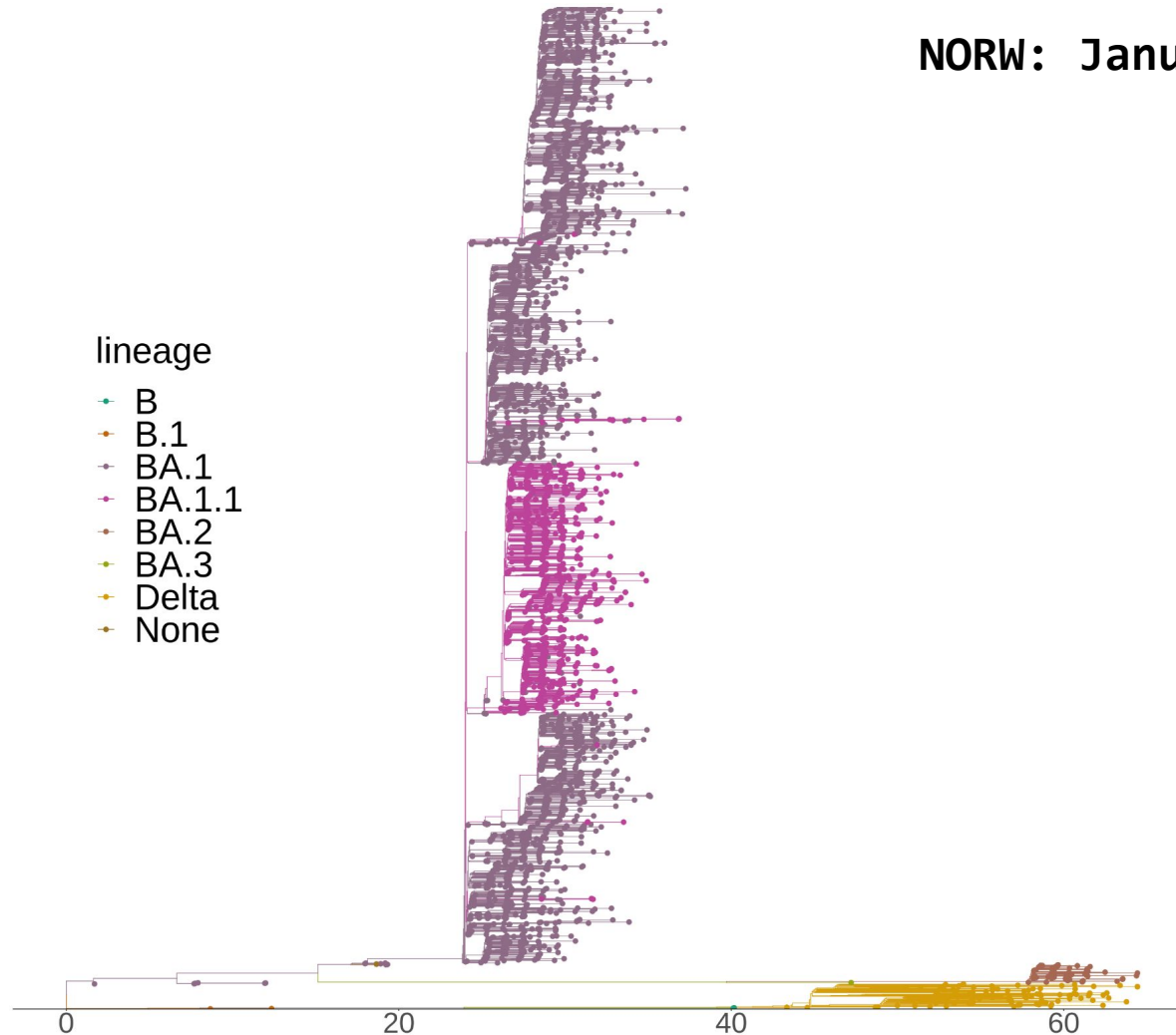


## NORW: January 1 to March 31, 2022

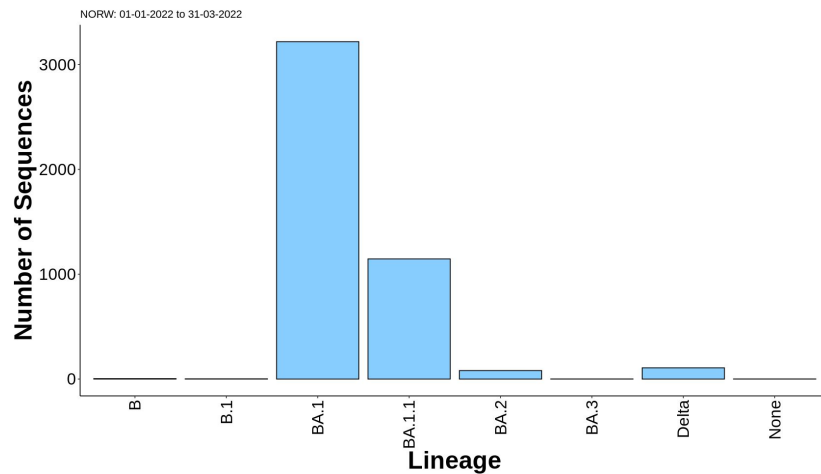
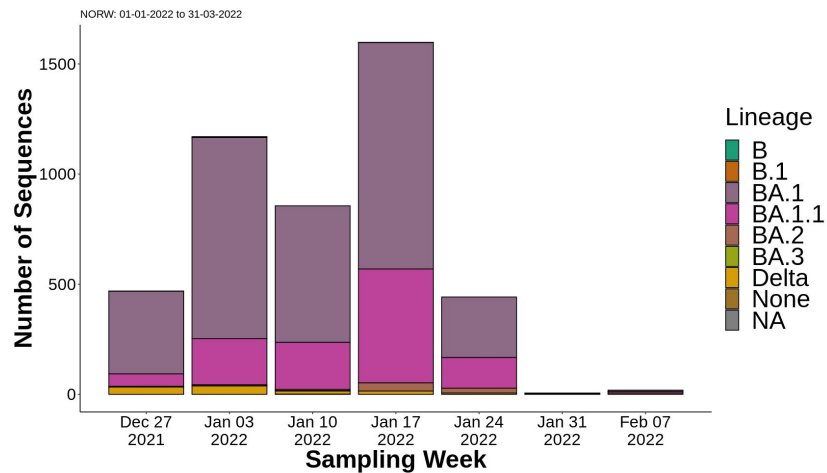
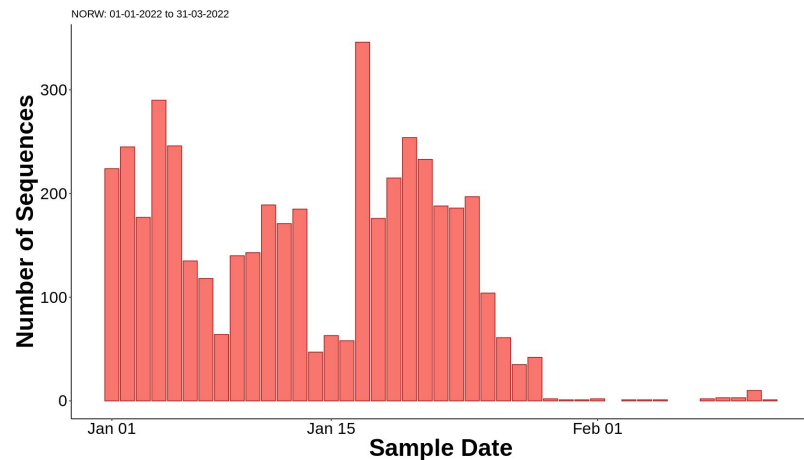
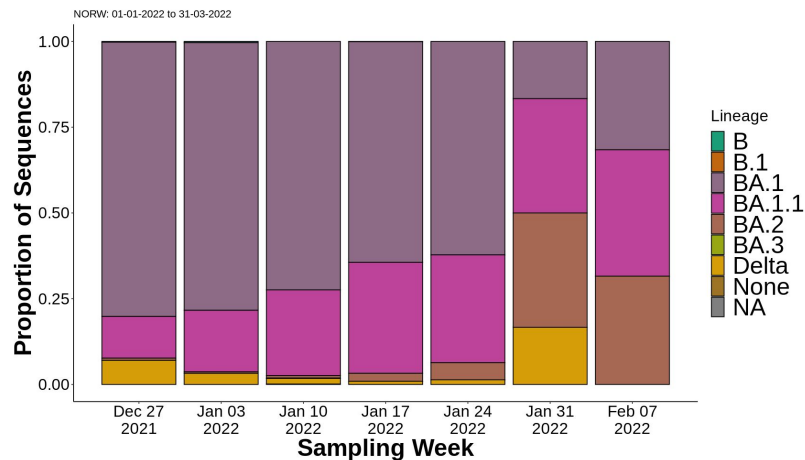
lineage

- B
- B.1
- BA.1
- BA.1.1
- BA.2
- BA.3
- Delta
- None

- ❖ This is the tree with outliers removed
- ❖ 4533 tips in total

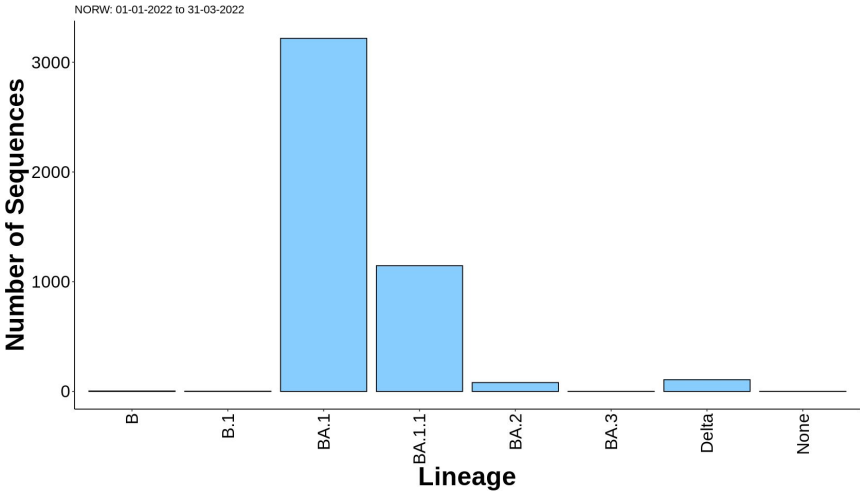
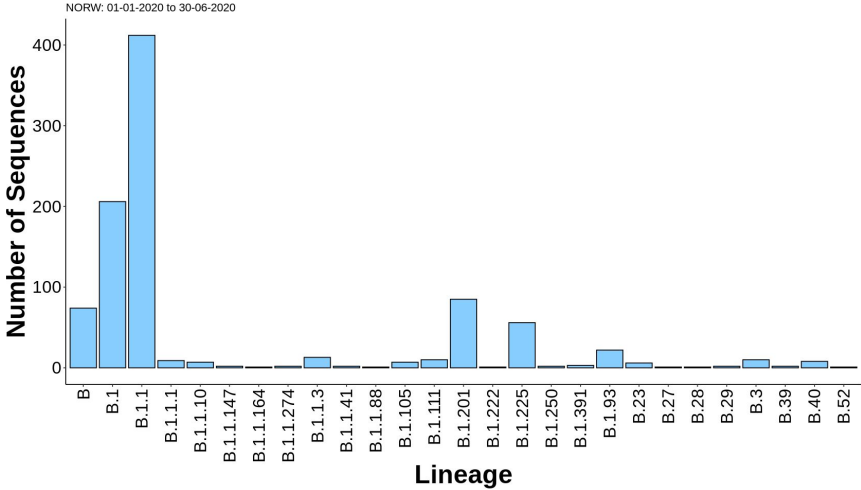


# NORW: January 1 to March 31, 2022



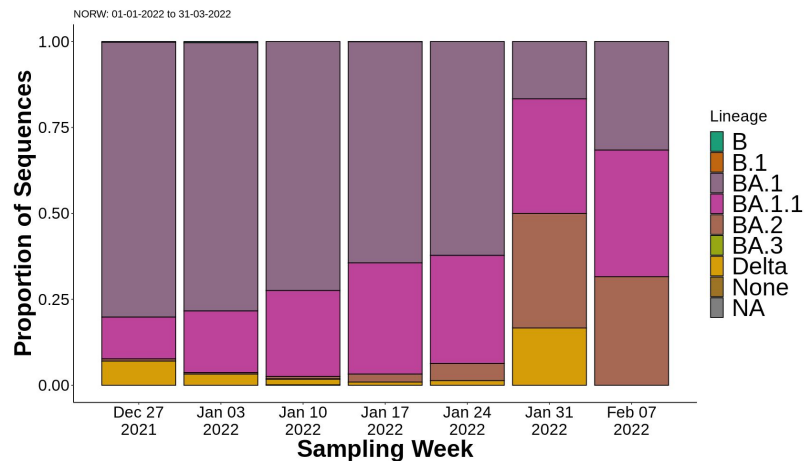
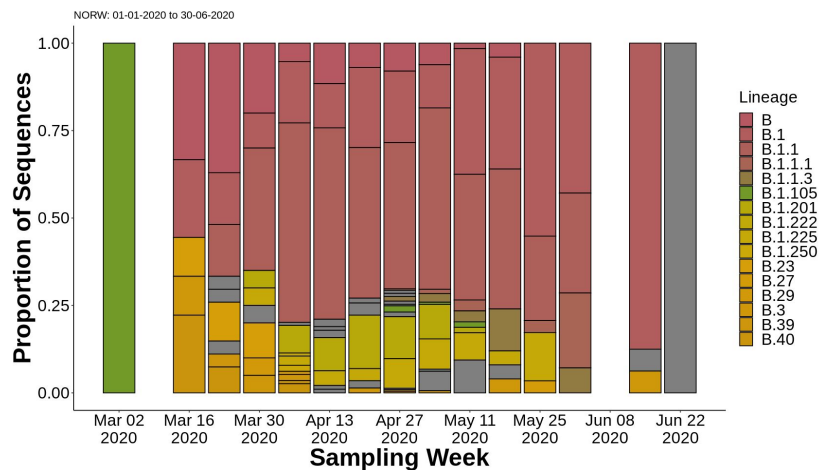
# Comparison between waves

❖ Greater diversity in first wave, despite smaller sample set

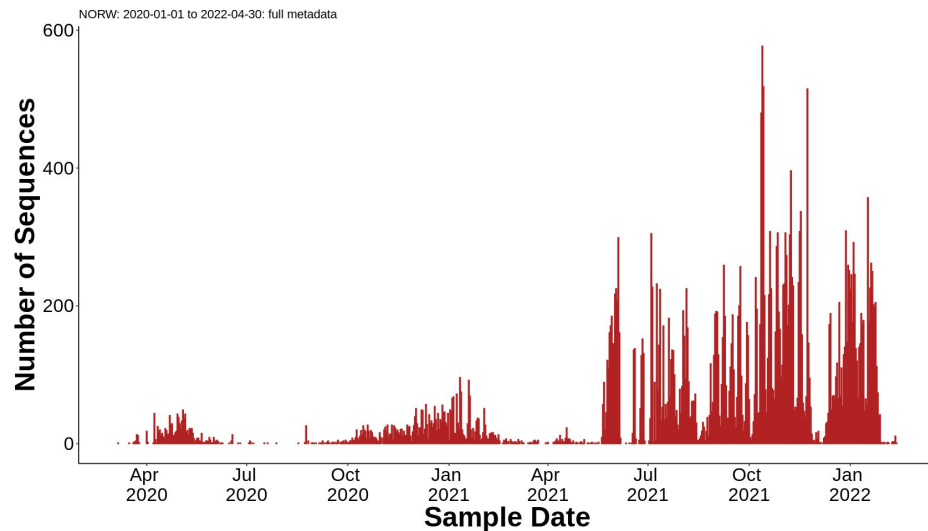
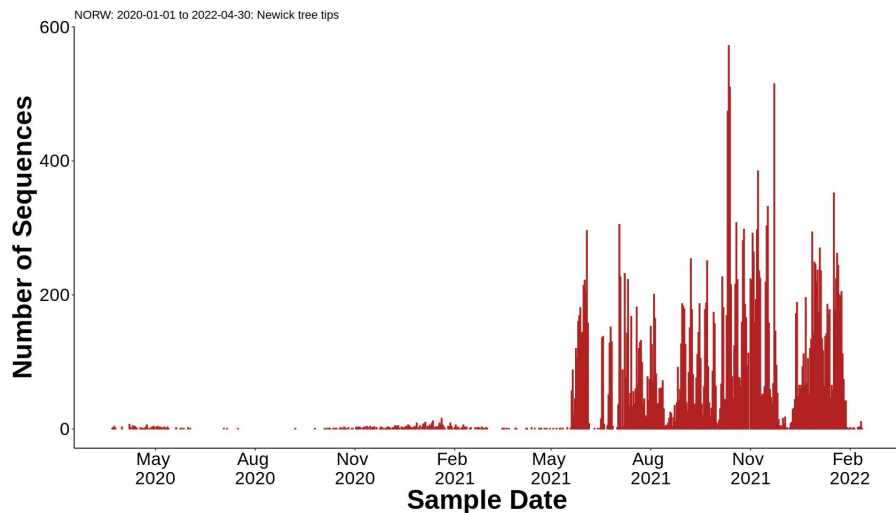


# Comparison between waves

❖ Succession visible in 2022, but not 2020



# Civet Newick Tree

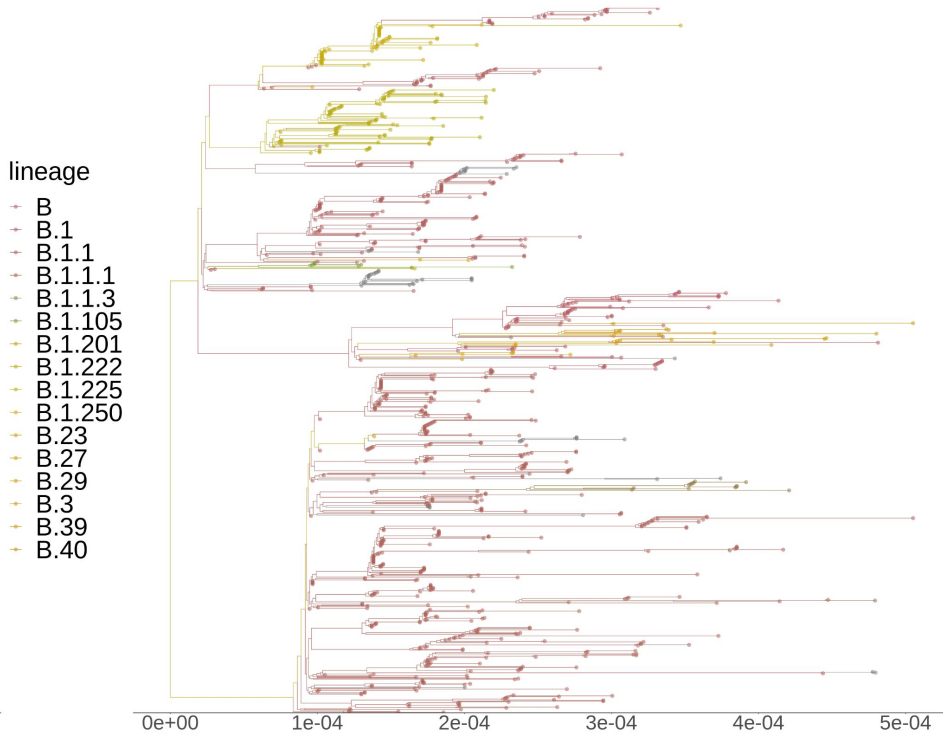
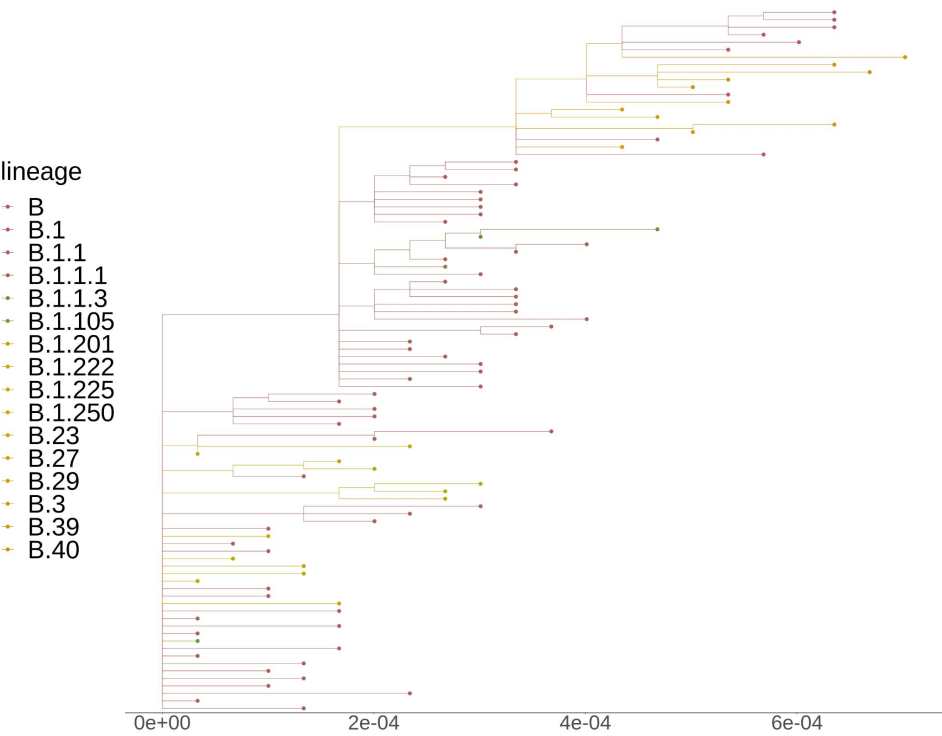


- ❖ The Civet tree is biased towards more recent samples
- ❖ Most of the first two waves are present in the metadata but missing from the tree



# Civet tree vs. Nextflow tree: first wave

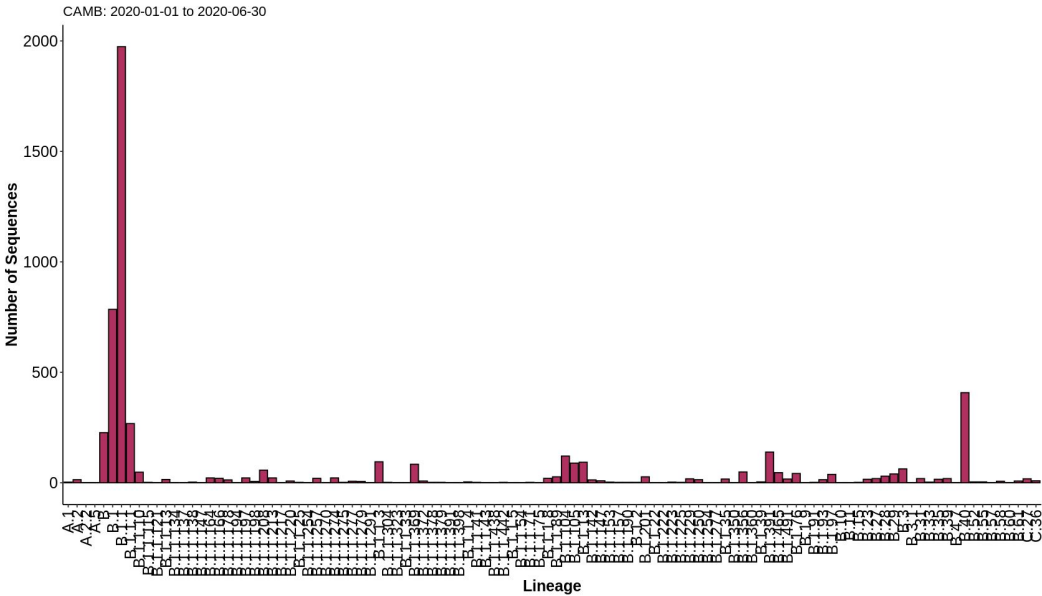
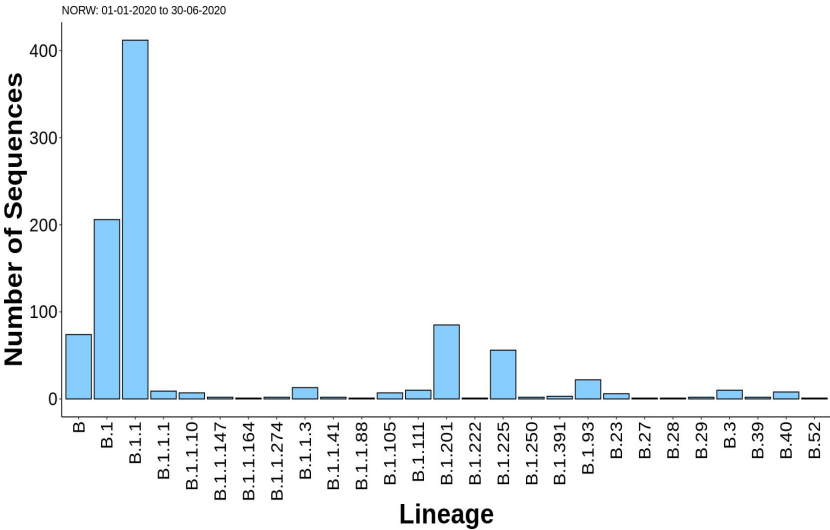
January 1, 2020 to June 30, 2020  
94 tips vs. 940 tips



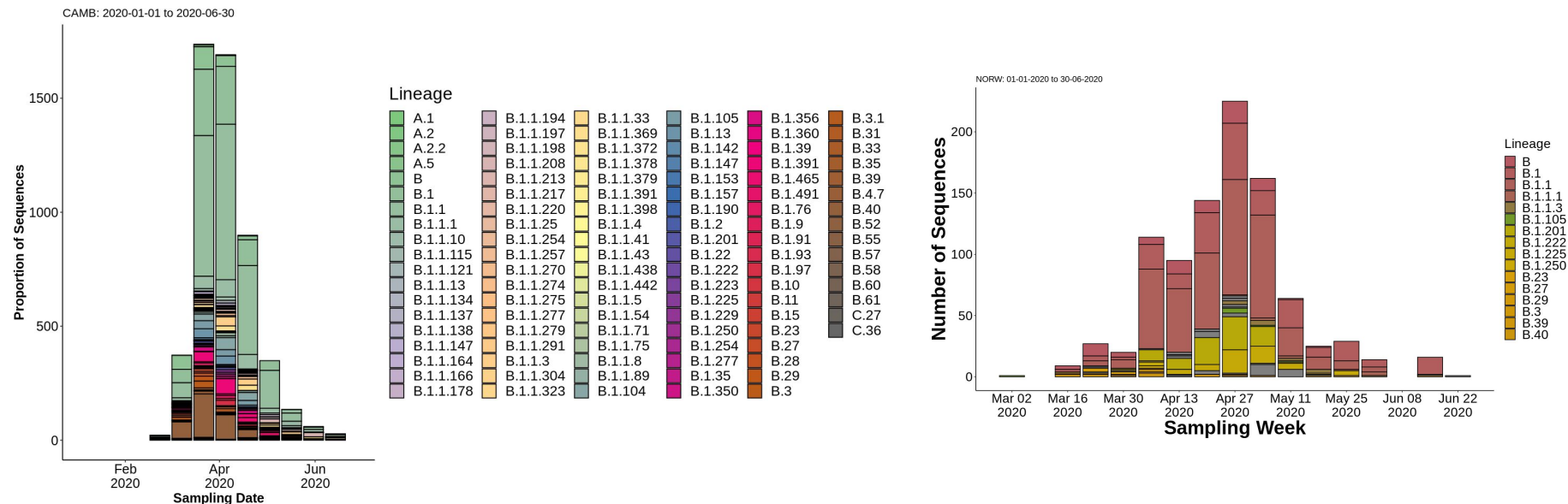
# NORW vs. CAMB: first wave

NORW: 986 seqs, 27 lineages

CAMB: 5302 seqs, 110 lineages



# NORW vs. CAMB: first wave



# Acknowledgements

## Quadram

### *Sequencing*

Justin O'Grady  
Gemma L Kay  
Alexander J Trotter  
Sophie Prosolek  
Ebenezer Foster-Nyarko  
Rebecca Sutcliffe  
Ana Martínez Gascueña  
Darcie Burgess  
David J Baker  
Rhianon Evans

### *Bioinformatics*

Andrew J Page  
Thanh Le-Viet  
Nabil-Fareed Alikhan  
Leonardo de Oliveira Martins  
Martin Lott  
Michaela Matthews

### *NRP BioRepository*

Rachael Stanley  
Roxanne Brunton-Sim  
Joel Wood  
Hannah Limbach  
Caelum Spearing

## Quadram

Alison E Mather  
Robert A Kingsley  
Lizzie Meadows  
Alp Aydin  
Steven Rudder  
Ana P Tedim  
Anastasia Kolyva  
Muhammad Yasir  
Gaetan Thilliez  
Maria Diaz  
Andrew Bell  
Ana Victoria Gutierrez  
Nicholas M Thomson  
Evelien M Adriaenssens  
Rachel AJ Gilroy  
Ian G Charles  
John Wain  
Mark A Webber  
Nicol Janecko  
Andrew Stronach  
Andrew Chapple  
Laura Deakin  
Mike Ridout  
Sandra Stringer

## Norwich and Norfolk University Hospital

Samir Dervisevic  
Emma J Meader  
Will Potter  
Claire Stuart  
Claire Williams  
Catherine Tremlett  
Dheeraj K Sethi  
Lindsay J Coupland  
Ngozi Elumogo  
Reenesh Prakash  
Elizabeth Trotter  
Christopher McDonnell  
Katy Errington  
Stephen Brolly  
Johanna Waters  
Timothy Crouch  
Sara Grundy

## Ipswich Hospital

Luke Bedford  
Gemma McGuire  
**West Hertfordshire  
Hospitals NHS Trust**  
Prema Singh

## Queen Elizabeth Hospital

Chris Jeanes  
**Norfolk County Council**  
S. J. Louise Smith  
Tim Winters

## UEA

Tracey Swingler  
Luke Griffith  
Rose K Davidson  
Orla Jupp  
Caroline Smith

## PHE

Dinesh Aggarwal  
Colin Brown  
Meera Chand  
Richard Myers

We are grateful for  
support from



Biotechnology and  
Biological Sciences  
Research Council



COVID-19  
GENOMICS  
UK CONSORTIUM