IBIOM
Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari

DBS
Dipartimento di Bioscienze

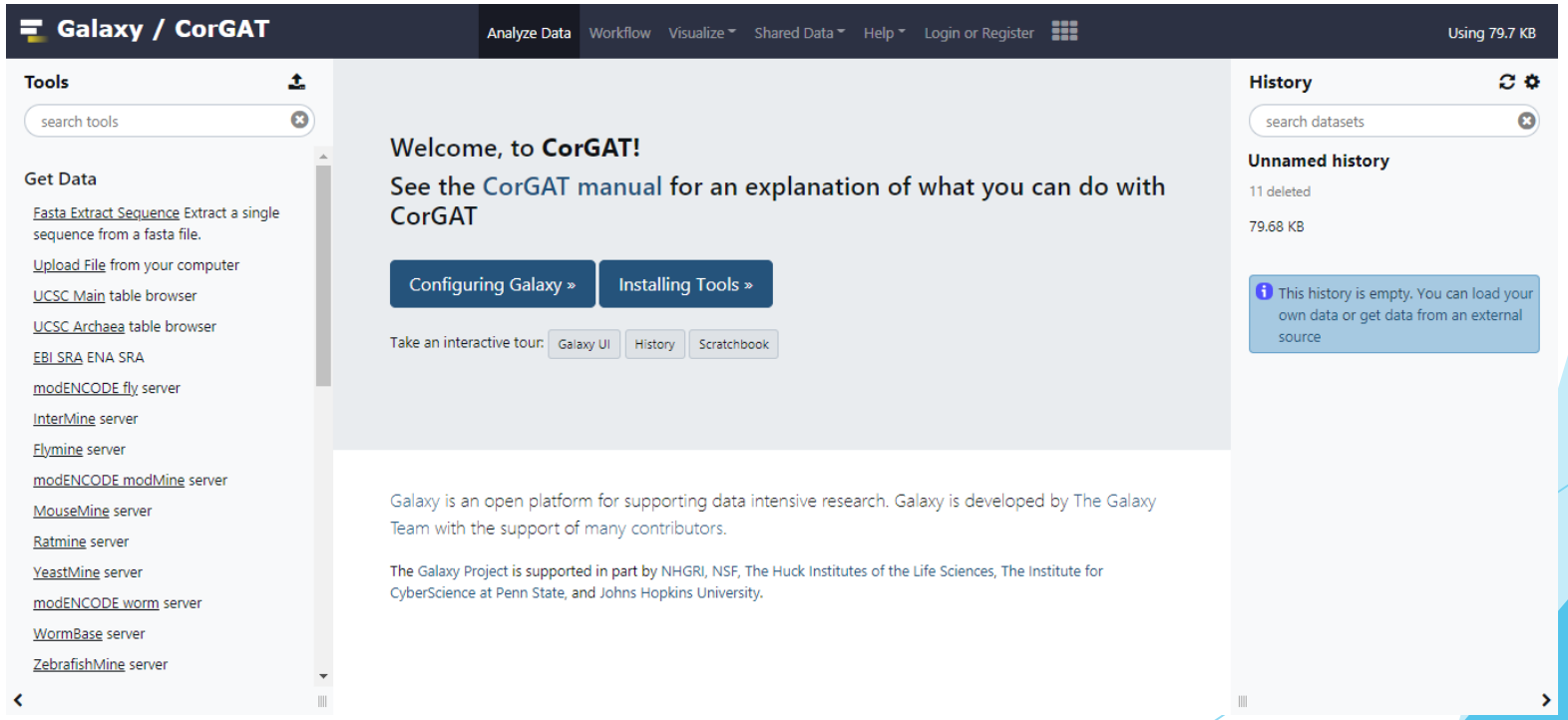
UNIVERSITÀ DEGLI STUDI DI MILANO

CorGAT and CorGAT-tracker

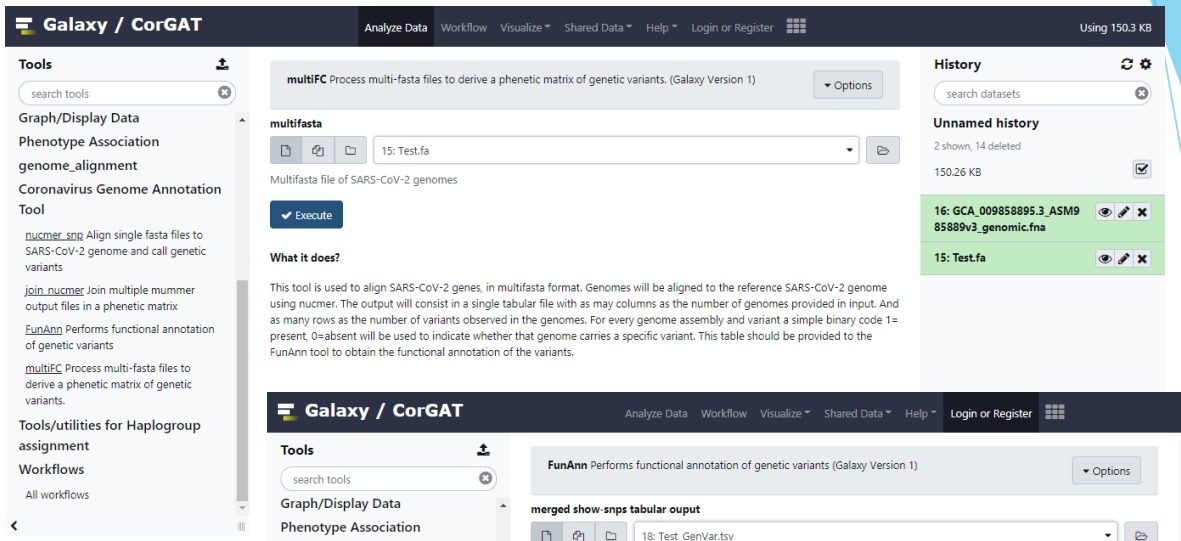
Functional annotation of SARS-CoV-2 genomes and tracking mutations and variants of concern

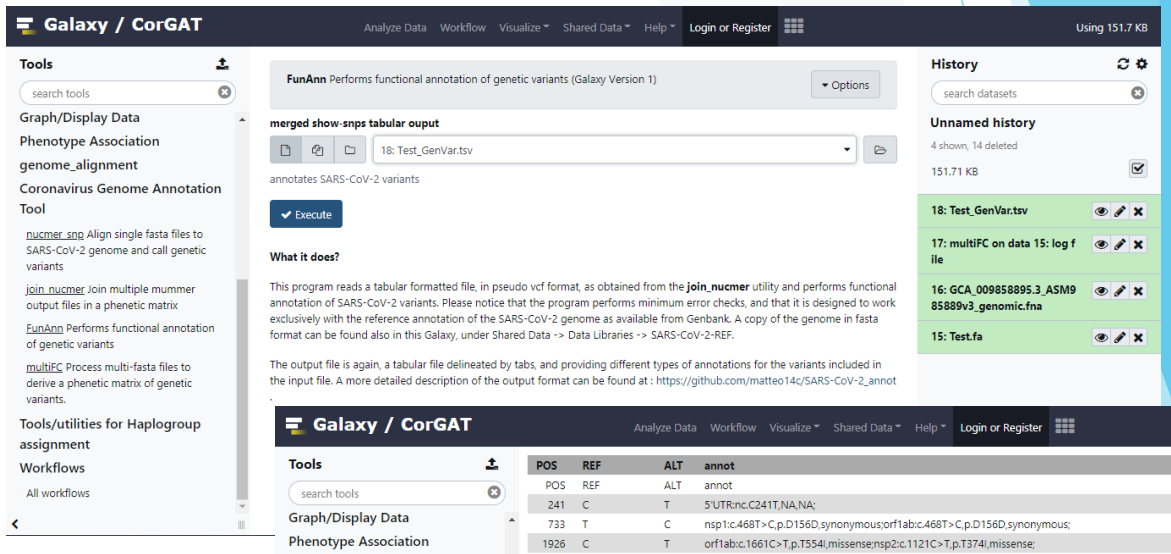
CorGAT

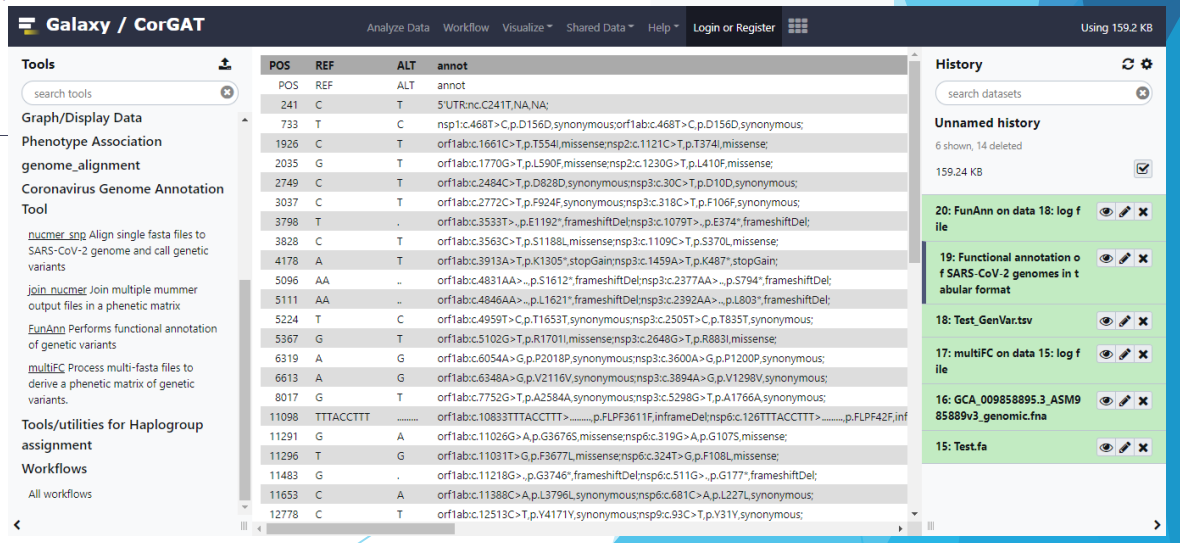
- Coronavirus Genome Analysis Tool or CorGAT (DOI:10.1093/bioinformatics/btaa1047).
- Collection of Perl utilities and annotation files.
- Performs the functional annotation of SARS-CoV-2 genetic variants.



- Brief explanation of functioning:
  1. Alignment of complete assemblies of SARS-CoV-2 genomes to the reference sequence.
  2. Obtain a list of polymorphic regions.
  3. Functional annotation of the identified variants.







- The output is a simple table containing:

1. Genomic position.
2. Reference allele.
3. Alternative allele.
4. Functional annotation.
5. Allele frequency
6. Epitopes annotation.
7. Annotation of sites under selective pressure.
8. MFE annotation.

POS	REF	ALT	annot
POS	REF	ALT	annot
241	C	T	5'UTR:nc.C241T,NA,NA;
733	T	C	nsp1:c.468T>C,p.D156D,synonymous;orf1ab:c.468T>C,p.D156D,synonymous;
1926	C	T	orf1ab:c.1661C>T,p.T554I,missense;nsp2:c.1121C>T,p.T374I,missense;
2035	G	T	orf1ab:c.1770G>T,p.L590F,missense;nsp2:c.1230G>T,p.L410F,missense;
2749	C	T	orf1ab:c.2484C>T,p.D828D,synonymous;nsp3:c.30C>T,p.D10D,synonymous;
3037	C	T	orf1ab:c.2772C>T,p.F924F,synonymous;nsp3:c.318C>T,p.F106F,synonymous;
3798	T	.	orf1ab:c.3533T>.,p.E1192*,frameshiftDel;nsp3:c.1079T>.,p.E374*,frameshiftDel;

AF	Epitopes
AF	
94.382501	
0.213728	
0.031701	
0.110145	410:LATNNLVVM,7,HLA-B*35:01;HLA-B*46:01;HLA-C*01:0
0.214636	7:FGDDTVIEV,1,HLA-C*08:01;5
96.462537	98:LASHMYCSF,5,HLA-B*15
0	

Hyphy	MFE
Hyphy	MFE
NA	NA
NA	NA
fel:true;meme:false;kind:negative;	NA
fel:true;meme:true;kind:positive;	NA
NA	NA
fel:true;meme:false;kind:negative;	NA
NA	NA

CorGAT-tracker

- Shiny based dashboard for the visualization of the prevalence of SARS-CoV-2 lineages and mutations of concern.
- Based on CorGAT derived annotations.
- A Galaxy release is under development.
- Data are represented in an interactive way.
- Users can personalize data visualization through a series of widgets that allow to modify as many parameters, among which:

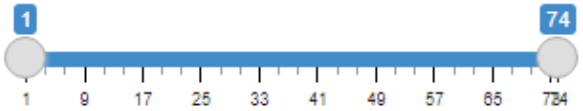
1. The country of origin of the data.
2. The interval of time to be displayed.
3. The minimum number of sequenced genomes.
4. A mutation of interest to be visualized.
5. A lineage of interest to be represented.

Country

Italy

Visualize data for the selected country

Weeks range



Time lapse of interest (number of weeks from a fixed date)

Min number of genomes (Lineages)

- 1
- 25
- 50
- (•) 100
- 500
- 1000

Minimum number of sequenced genomes required to display a Lineage

Min number of genomes (Lineages+)

- 1
- 5
- 10
- (•) 15
- 25
- 50

Minimum number of sequenced genomes required to display a Lineage+

Lineage

B.1.1.7

Produce a scatterplot for the selected Lineage

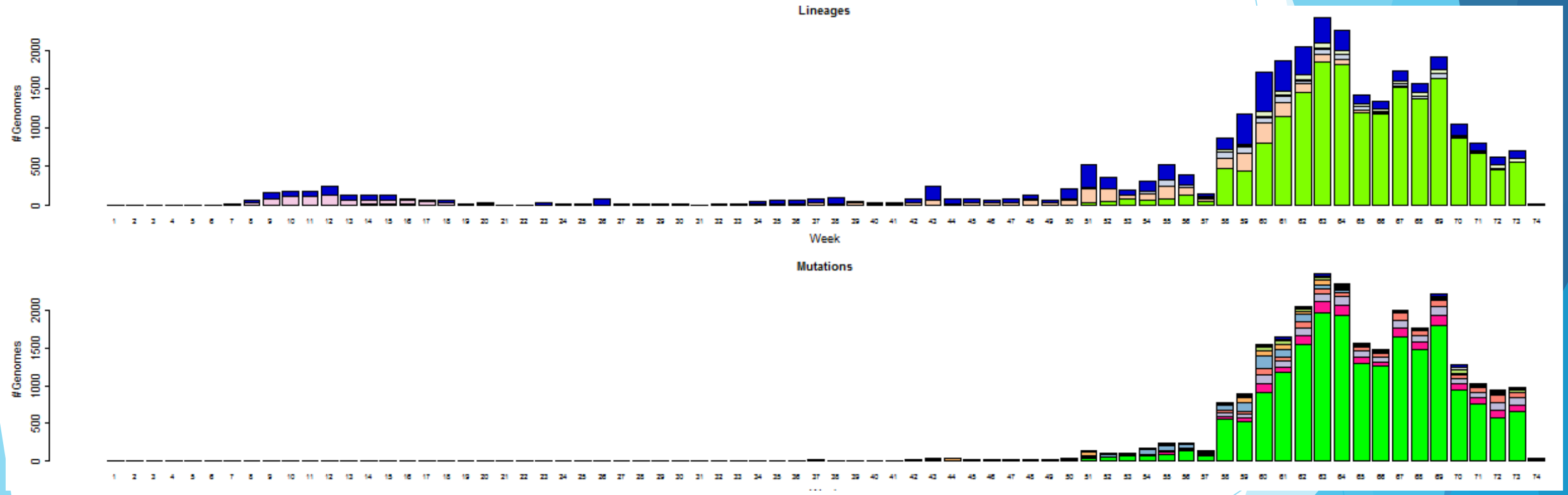
Mutation

N501Y

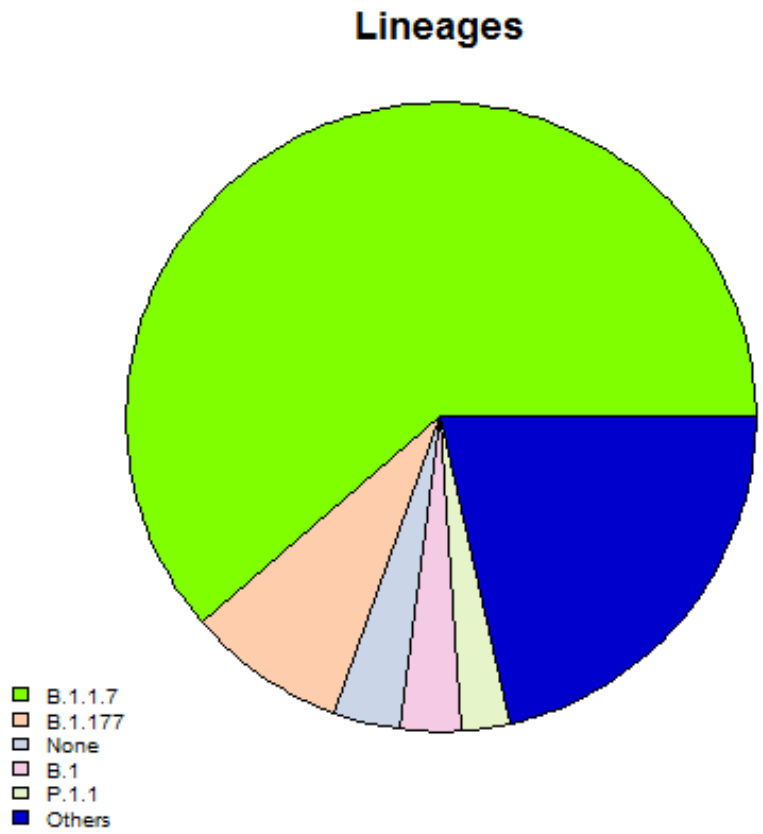
Produce a scatterplot for the selected Mutation

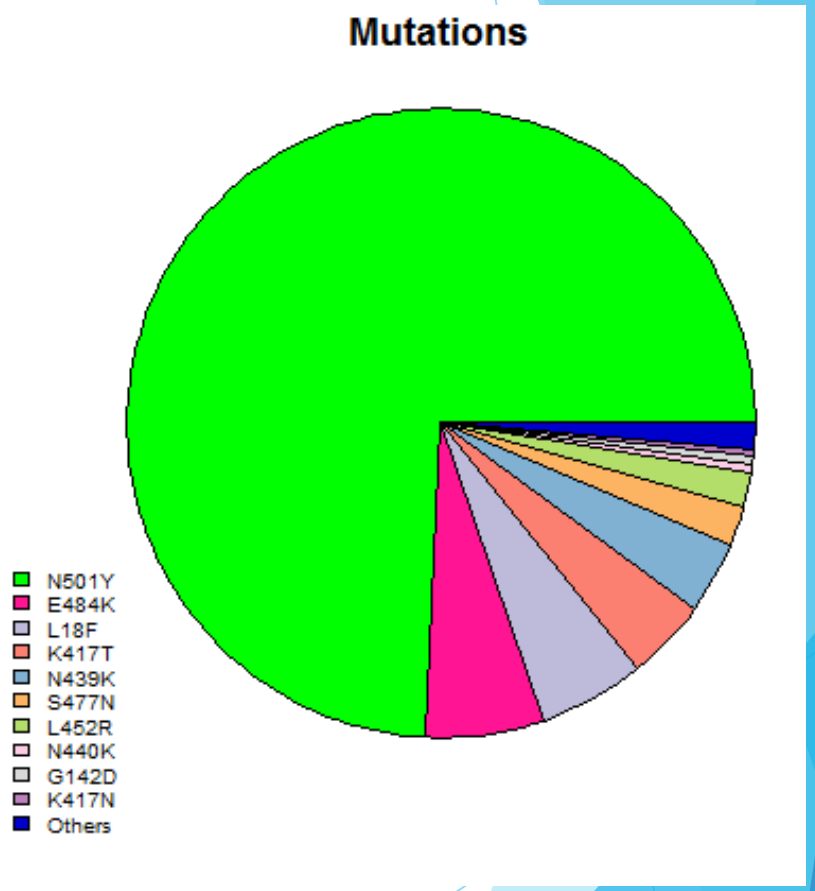
- CorGAT-tracker produces 3 different kinds of plot:

  1. Barplots

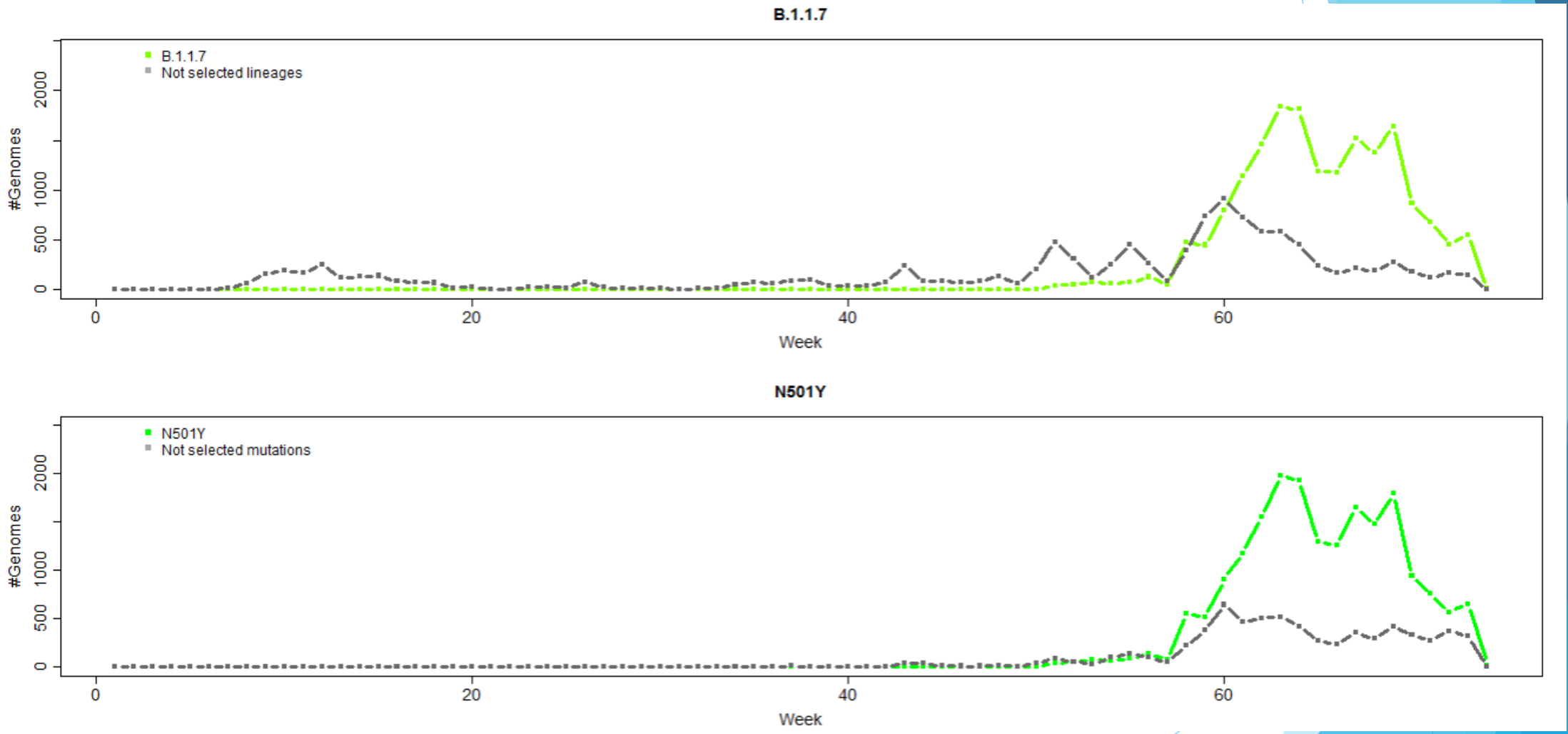


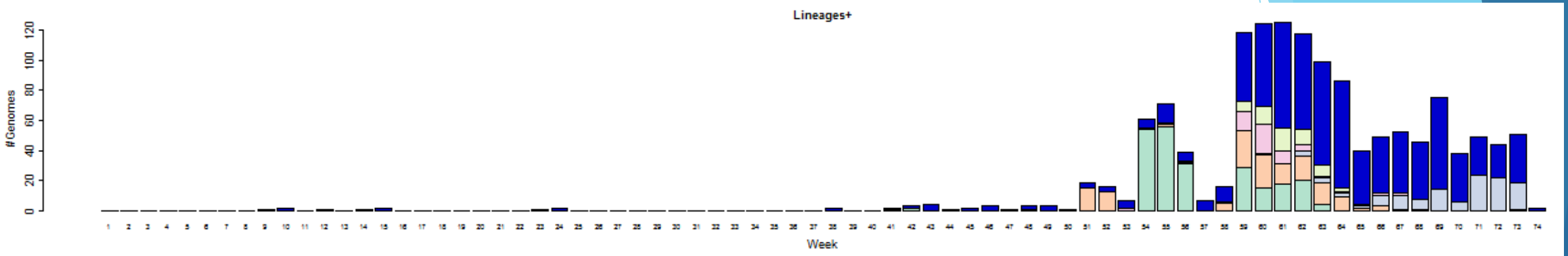
2. Pie-charts


Lineages
B.1.1.7
B.1.177
None
B.1
P.1.1
Others

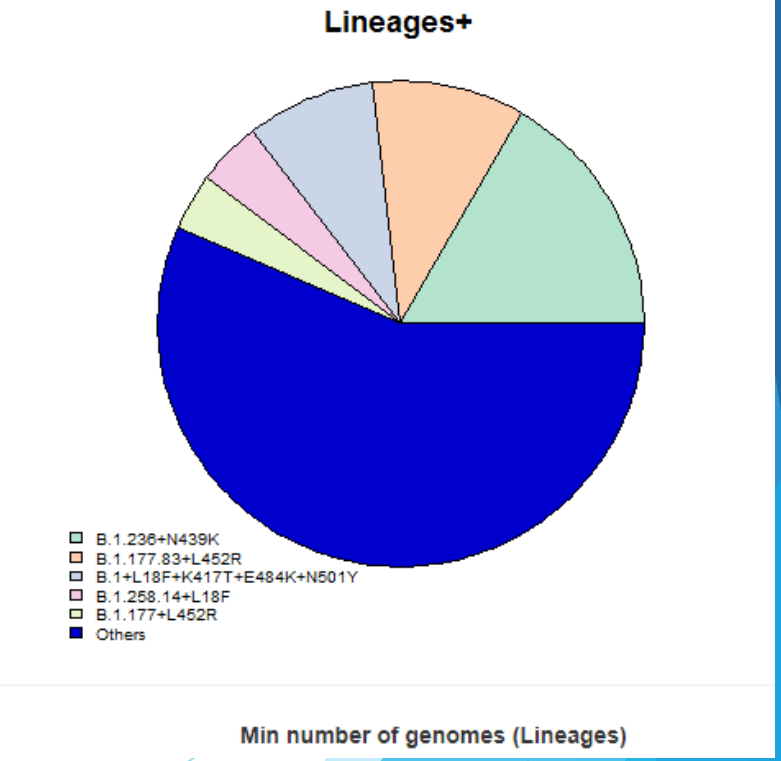

Mutations
N501Y
E484K
L18F
K417T
N439K
S477N
L452R
N440K
G142D
K417N
Others

3. Scatterplots





- In CorGAT-tracker lineage annotations can be “augmented” by reporting the list of MOC that are observed in a genome, but are not specific to its assigned lineage.
- Augmented annotations are called Lineages+ in-app.
- Lineages+ prevalence in time is represented using a barplot and a pie-chart



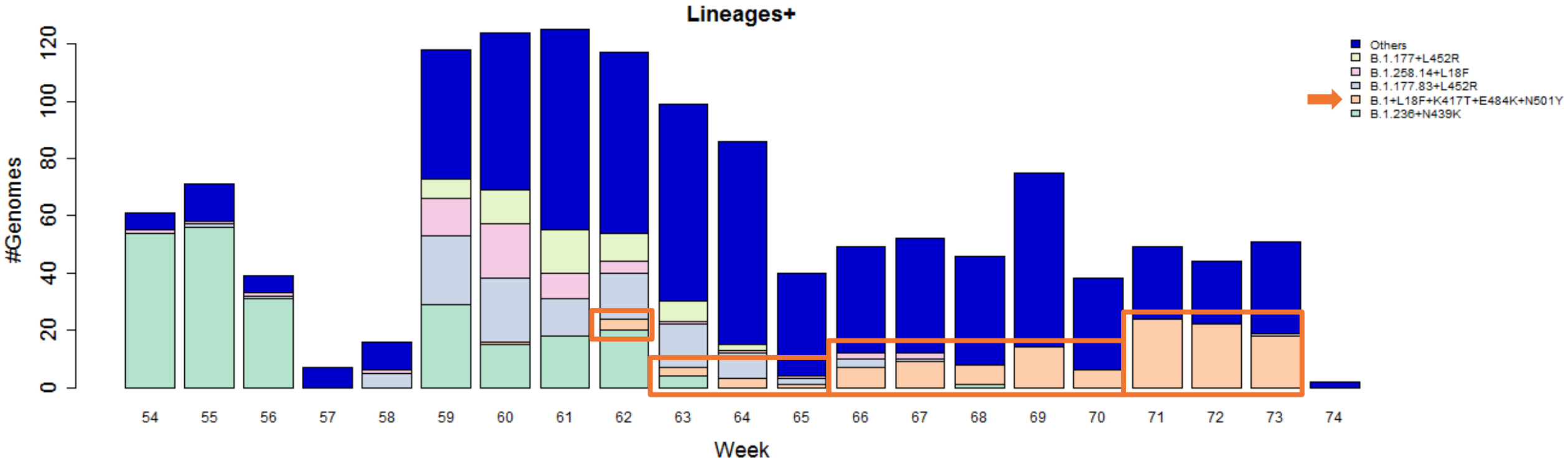
Final considerations

- CorGAT and CorGAT-tracker will provide a useful addition to the currently available "arsenal" of bioinformatics methods for the genomic surveillance of SARS-CoV-2.

- CorGAT has a sensitivity comparable to other similar tools, but also provides additional layers of annotation.

- Example: Identification of misclassified SARS-CoV-2 genomes in Italy.

- Genomes classified as B.1 but presented additional mutations on the spike protein.



- In depth studies highlighted that:
  1. The majority of the additional spike mutations in the misclassified B.1 were in common with P.1.
  2. A “group specific” mutation, P681H, can be identified in the spike protein of the misclassified genomes.

- It is possible to speculate that these genomes represent a newly emerged lineage, however further investigations are required.

- P.1+P681H was recently added to the ECDC list of Variants Under Monitoring

Availability

- CorGAT is already available through Galaxy at the following link:
  - http://corgat.cloud.ba.infn.it/galaxy
- A Galaxy release for CorGAT-tracker is under development.
- Further information about the tools can be found in the respective GitHub repositories:
  - https://github.com/matteo14c/CorGAT (CorGAT)
  - https://github.com/F3rika/CorGAT-tracker (CorGAT-tracker)

IBIOM
Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari

DBS
Dipartimento di Bioscienze

UNIVERSITÀ DEGLI STUDI DI MILANO

Thank you!

Special thanks to:

- Matteo Chiara
- Federico Zambelli
- Marco Antonio Tangaro
- Pietro Mandreoli
- David S. Horner
- Graziano Pesole