Viral Beacon Extended metadata on sample and variants and UI graphs

Contents

1 New annotations of query by feature

Use Uniprot domains of viral proteins: have to be implemented as Aminoacid queries
Include as nt queries as well, converter:
Check to include these, present in Chinese beacon: TRS:orf1ab TRS:S TRS:ORF3a TRS:E TRS:M TRS:ORF6 TRS:ORF7a TRS:ORF8 TRS:N TRS:ORF10 Furin:S

2 Metadata tables

To facilitate user to explore data: //
Expand Sample Matadata table to include/ do database query on the fly?:

- number of variants per sample (by type SNV: INDELS:)

- number of nonsynonymous variants per sample

- number of aminoacid changes per sample

so, user can filter by samples having so many variants.

Expand Variant Matadata table to include/ do database query on the fly?:

- number of samples harboring the variant (by country, month -boxplot)

- AF mean +- SD (by country, month -boxplot)

3 Summary at database level - Split by dataset

Graphs of summary statistics of samples and variants at whole database

3.1 Samples

Static distribution of samples per country, date, age and sample type and sex in dataset (Figure 1)
Number of samples: - by country (col) - by age and sex (histogram) - by sample source (col) - with variants by genomic position
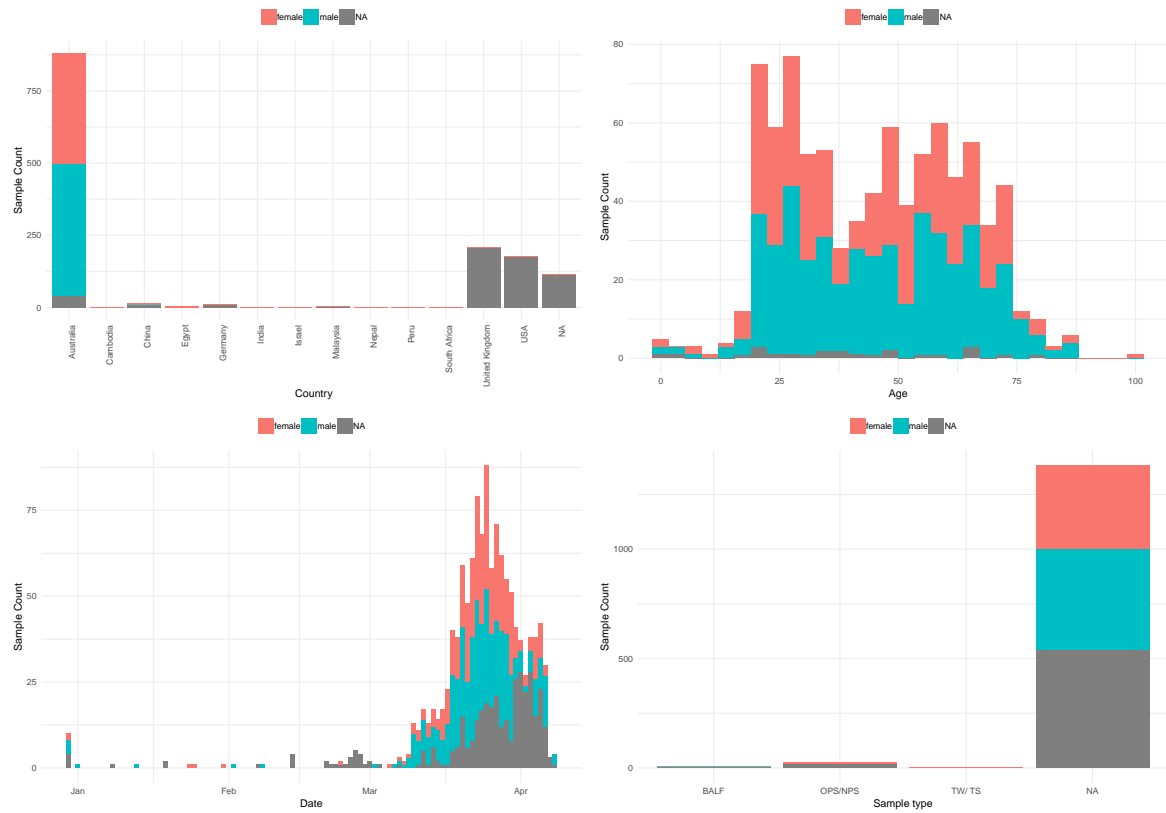
Figure 1: Distribution of samples by country, age, sex, sample type
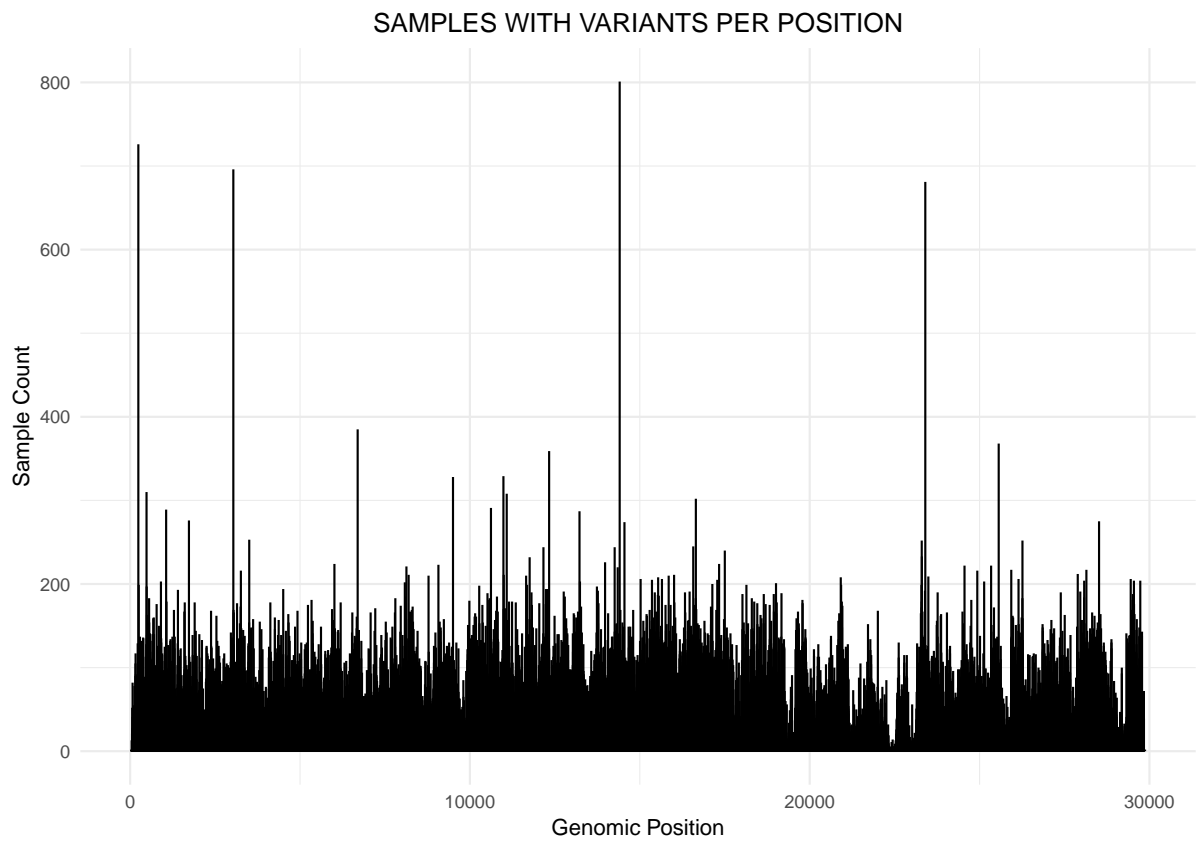
SAMPLES WITH VARIANTS PER POSITION

Figure 2:

4 Variants

Number of variants by population frequency, by country, age, date, sex, popuation frequency range, AF range.
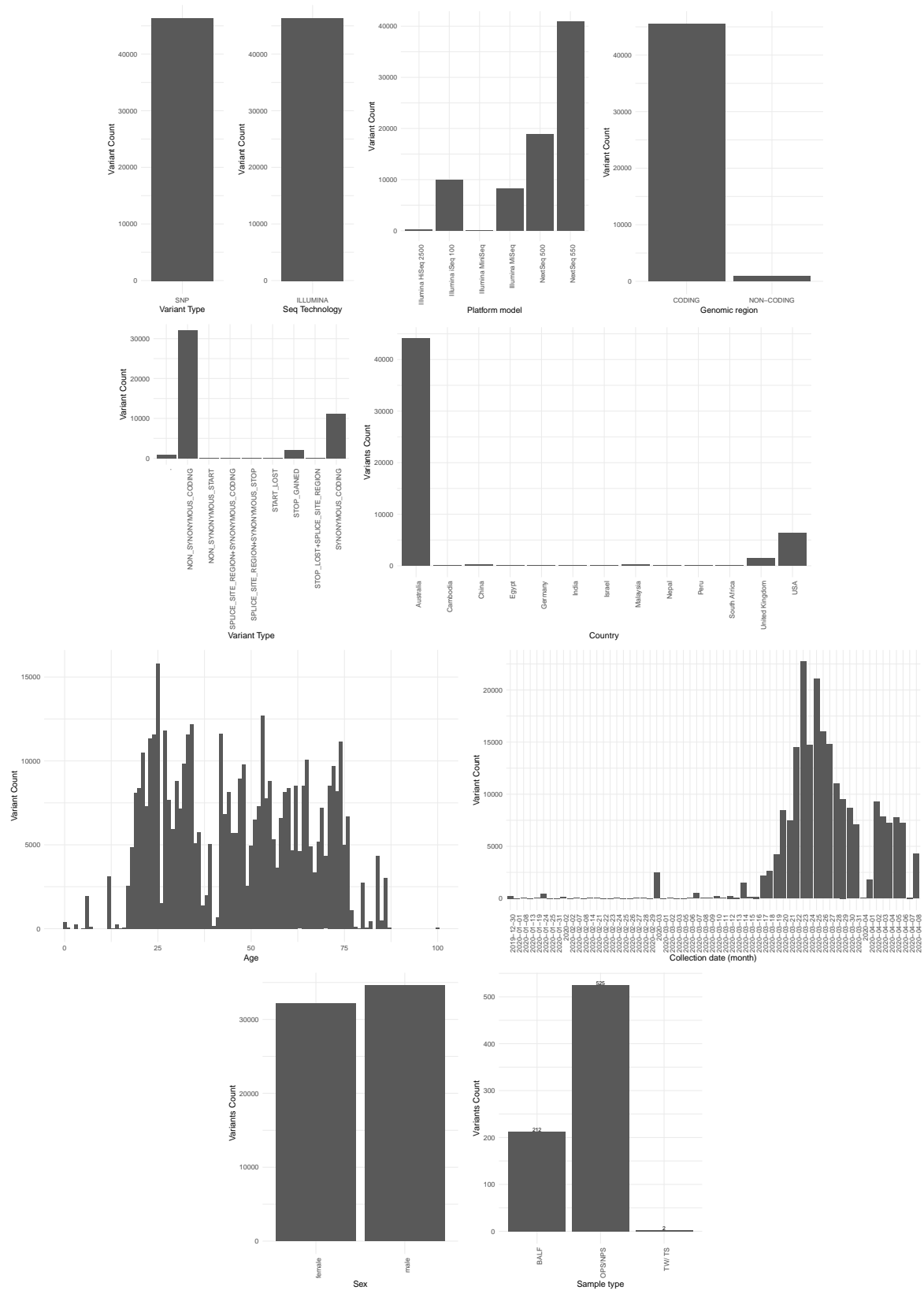

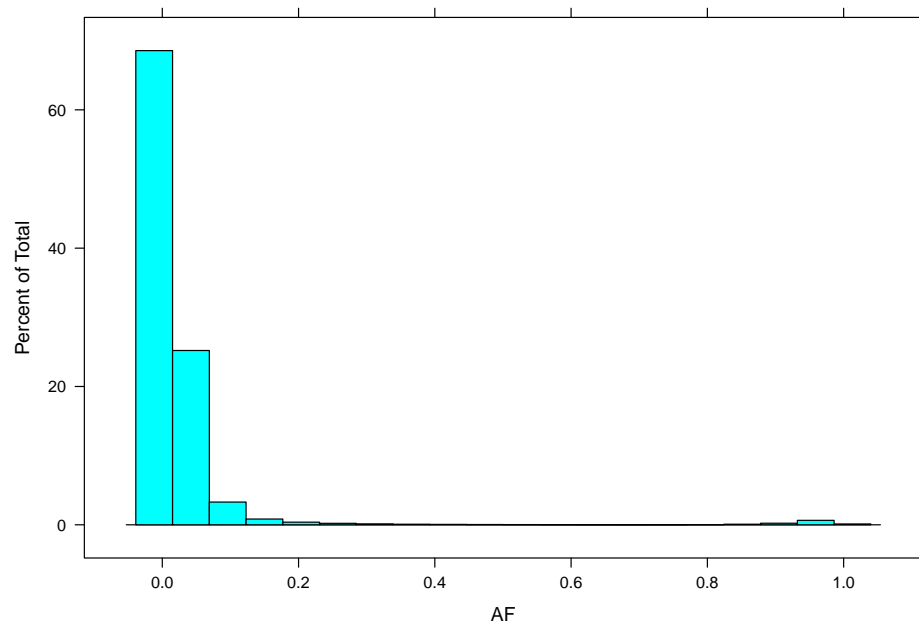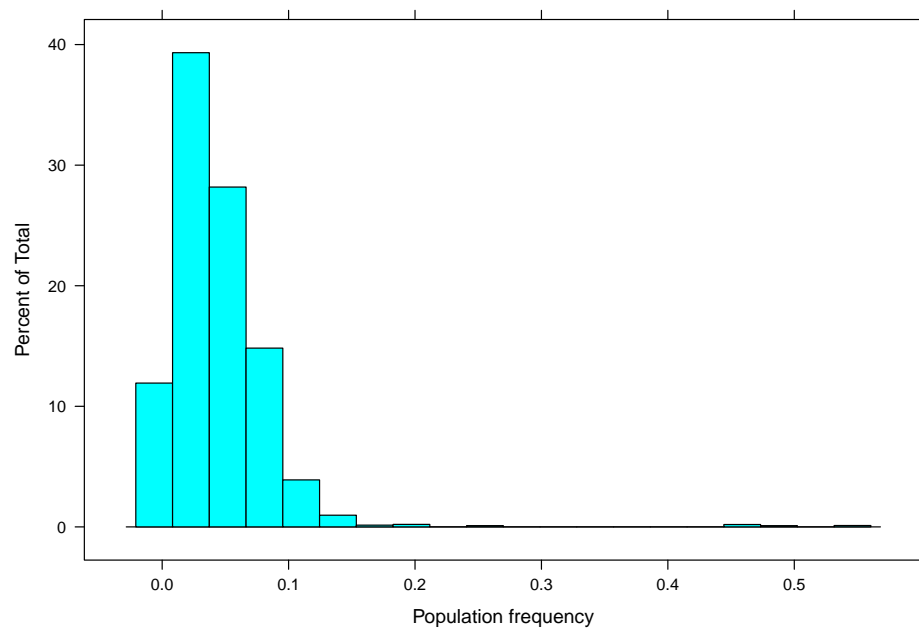Figure 3: Distribution of variants by country, age, sex

Figure 4: Number of variants per population frequency, AF group

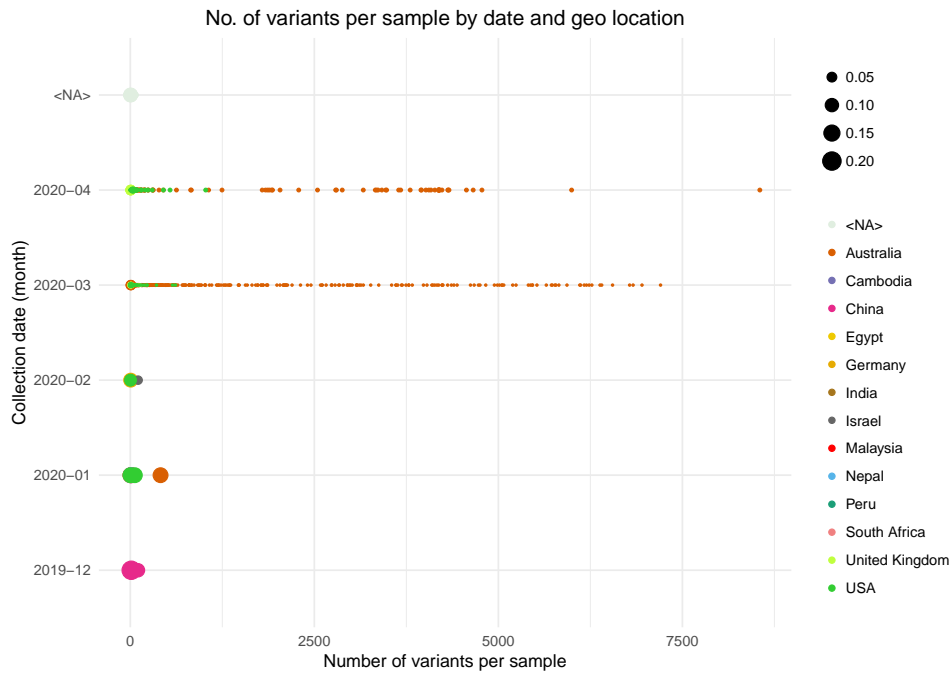Number of variants per sample, average

Figure 5: Number of variants in sample

5 Use case: Variants unique to some groups

Does number and id of variants associated with sample characteristics available?

5.1 unique/correlated with sample type

Curiously, one of the two mutations found in tracheal samples ("27653¿T¿C") is not found in any other sample type. It has been found in both tracheal samples in database (1 from Germany -February and 1 from Australia-March. It is a minor (mean AF 0.004970, 0.055556) missense variant in ORF7a)

5.2 unique/correlated with sex

5.3 unique/correlated with geo

6 Filter level

Dynamic histograms at filter level using Variants stats above, to see what is left of others field while filtering.
    Also:

In SNP query and region query pages, needle plot of positions and population frequency.
In region feature query page, number of variants per gene/mature peptide.

7 Hit table level

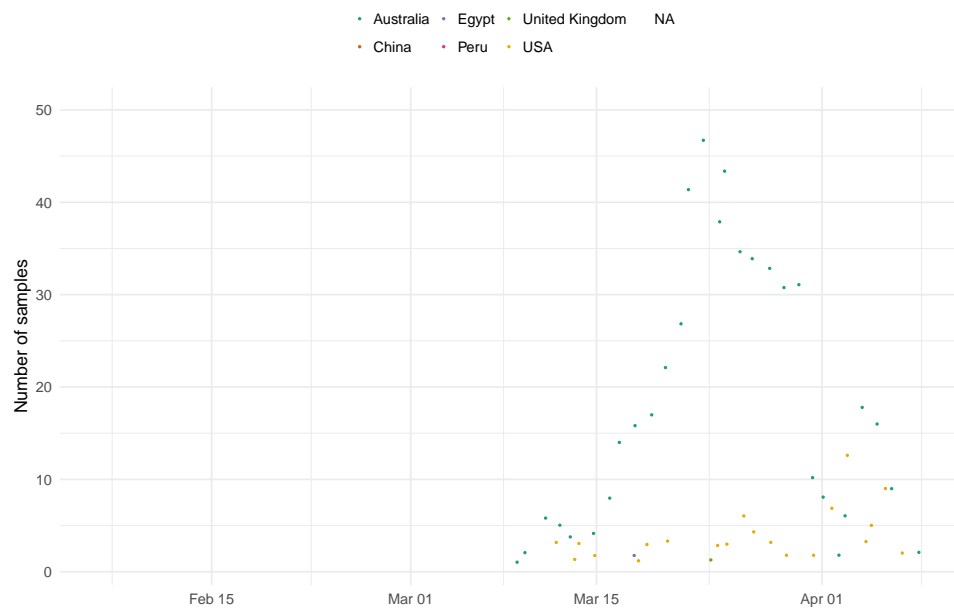Statistics on result

7.1 Variant query

For results of variant query, geographic and temporal, sample types distributions, other variants in same position, other variants covering same position, like indels?

7.2   Region query

Number of positions with variants/invariant positions over time positions changing dataset frequency and AF over time - heatmaps average number of variant positions in sample over time
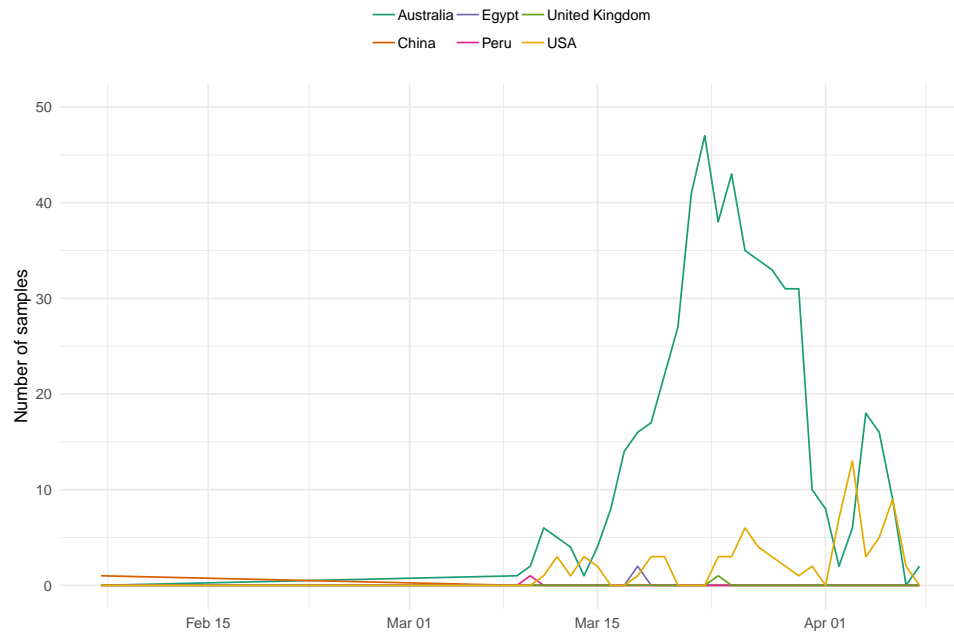
Population frequency of variant over time

AF of variant over time

7