GitHub - genomebits
User's Guide **version 1.0.x**; September 2021

Tool for Uncovering Characteristics Signals from Genome Sequences

genomebits *is a Graphics User Interface (GUI) to the signal analysis of complete genome sequences according to its progression along the nucleotide bases position. This method for uncovering distinctive patterns in the intrinsic data organization of genome sequences is based on a finite alternating sum series having independently distributed terms associated with binary (0,1) indicators for the nucleotide bases A,C,T,G. The GUI runs under Linux Ubuntu O.S., and it can be useful to study the dynamics of Human CoV-2 genome variants using available GISAID FASTA data.*

Genes 2021; 12(7):973 - doi.org/10.3390/genes12070973

Table of Contents

1 About genomebits 1

1.1 genomebits GUI Features 1

2 Requirements genomebits 4

2.1 Hardware & Software 4

2.1.1 Copyright 4

3 Install 5

3.1 Manual Install 5

3.2 The simple way 5

3.3 The hard way 5

3.4 Uninstall 6

3.5 Updates 6

4 How To use? 7

4.1 Run Alternating Sums 7

4.1.1 FASTA Outputs 8

4.1.2 Get/Save Single FASTA Files 8

4.1.3 Associated Binaries for A, C, T, G 9

4.1.4 Alternating Sums for A, C, T, G 10

4.1.5 Alternating Sums Plots for A, T and C, G 11

4.1.6 Alternating Sums Plots for A, C, T, G 12

4.2 Run Multiple 13

4.2.1 Alternating Sums Plots for A, C, T, G for each Country 13

4.2.2 Alternating Sums Plots for A, C, T, G Both Countries 15

4.2.3 Plot All genomebits GUI Results 15

4.2.4 Compare genomebits GUI and Article Results 16

1 About genomebits

genomebits is the Graphics User Interface (GUI) to the signal analysis method published in: Genes 2021; 12(7):973 –doi.org/10.3390/genes12070973, to uncover distinctive patterns regarding the intrinsic data organization of complete genomics sequences of (A)denine, (C)ytosine, (G)uanine and (T)hymine according to its progression along the nucleotide bases position (bp). It is based on a finite alternating sum having independently distributed terms mapped into four binary projections for the A,C,T,G nucleotide bases.

The present Graphics User Interface (GUI) for genomebits could be useful for the quantitative examination of distinctive patterns of complete genome data, such as the ongoing Human CoV-2 genome variants by using, for example, available GISAID FASTA sequences from www.gisaid.org

We deal here with the simplest alternating sum series

$$E_{\alpha,N}(X) = \sum_{k=1}^{N} (-1)^{k-1} X_{\alpha,k} \quad , \qquad (1)$$

satisfying the following relation

$$X_{\alpha,N} = |E_{\alpha,N}(X) - E_{\alpha,N-1}(X)| \quad ,$$

where the variable $\alpha = A, C, T, G$ is in correspondence with one of the four nucleotide bases, and the individual terms X_k are associated with binary 0 or 1 values according to its presence along the complete genome sequences of length N. In this mapping, the arithmetic progression carries positive and negative signs $(-1)^{k-1}$ and a finite non-zero first moment of the independently distributed variables $X_{k,\alpha}$. Analyzing genomics sequencing via this type of finite alternating sums allow to extract unique features at each bp with a small degree of noise variations. This mapping into binary (0,1) indicators for the genome sequences is motivated by previous studies on the three-base periodicity characteristic of protein-coding DNA sequences.

1.1 genomebits GUI Features

The genomebits GUI integrates different technologies under Linux O.S. **Ubuntu 21.04** to facilitate the calculations of, and the plotting of results obtained by, Eqn.(1) with just one click!

genomebits allows to save all single, or paired, outputs entering a (given set of) complete genomics sequences in FASTA format as the only input.

The FASTA format is a text-based format for representing either nucleotide sequences or amino acid (protein) sequences, in which nucleotides or amino acids are represented using single-letter codes: e.g., A,C,T,G. The format also allows for sequence names and comments to precede the sequences. The format originates from the FASTA software package, but has now become a near universal standard in the field of bioinformatics. -See Wikipedia: https://en.wikipedia.org/wiki/FASTA_format

Example of FASTA data file containing more than one (concatenated) genome sequence separated by the math symbol '>':

```
>hCoV-19/Italy/LAZ-AMC-202105034047-DS/2021|EPI_ISL_1970570|2021-05-03
TAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTTAAAA
TCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGAC
ACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTTTTGCAGCCGATCATCAGCACATCTAGGTTTTG
TCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACGTCCAACTCAGTTTGCCT
GTTTTACAGGTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCT
  ...
    ...
>hCoV-19/Italy/TAA-IGA-1900588747/2021|EPI_ISL_2318924|2021-05-12
AAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTTAAAATC
TGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACAC
GAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTTTTGCAGCCGATCATCAGCACATCTAGGTTTTGTC
CGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACGTCCAACTCAGTTTGCCTGT
TTTACAGGTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTA
  ...
    ...
```

Note: genomebits considers samples with A,C,T,G sequences ONLY! for (up to two) given Countries corresponding to genomics sequence data from (up to six) given variants/species and it discards uncompleted sequences containing codification errors (usually denoted with '$NNNNN$s' and other letters).

In brief, genomebits from (up to two) selected Countries for (up to six) given variants/species allows to

- **Run alternating sums in Eqn.(1)** for up to six-*times*-two inputs of FASTA files containing (*i.e.*, concatenating) more than one genome sequence each.

- **Separate concatenated genome sequences and save in single FASTA files** (for each Country), which may be containing in a single FASTA input file including more than one genome sequence.

- **Get single files for each of the four nucleotide bases** A, C, T, G associated with binary 0 or 1 values according to its presence along the complete single genome sequences.

- **Get the alternating sums results in single files for the four nucleotide bases** A, C, T, G associated with binary 0 or 1 values according to its presence along the genome sequences.

- **Plot the alternating sums of the binary data files to compare behaviour of the pairs** A, T **and** C, G **nucleotide bases versus bases position (bp)**.

- **Compare in one plot the alternating sums results versus bases position (bp)** for all four nucleotide bases A, C, T, G associated with binary 0 or 1 values according to its presence along the genome sequences for each Country.

- **Plot the alternating sums results versus bases position (bp)** for the four nucleotide bases A, C, T, G associated with binary 0 or 1 values according to its presence along the genome sequences *for each* Country.

- **Plot the alternating sums results versus bases position (bp)** for the four nucleotide bases A, C, T, G associated with binary 0 or 1 values according to its presence along the genome sequences *for both* Countries (if given as input).

- **Plot in one image all the `genomebits` GUI results versus bases position (bp)** for each nucleotide base A, C, T or G associated with binary 0 or 1 values according to its presence along the genome sequences for up to six variants/species and up to 4 FASTA files by Country.

- **Compare visualmente the `genomebits` GUI results versus bases position (bp)** for (up to six) given variants/species and (up to two) selected Countries, with the results in the paper "*Uncovering Signals from the Coronavirus Genome*" -see: Genes 2021; 12(7):973. https://doi.org/10.3390/genes12070973

2 Requirements `genomebits`

2.1 Hardware & Software

As of this writting September 2021, `genomebits`

- runs on any Laptop PC 64 bits running Ubuntu Linux O.S.

- The '`genomebits`' software is 6.4 MB in size, and it can be downloaded in the form of a Debian package '`genomebit-1.0.x-Linux.deb`', from GitHub - `genomebits`.

`genomebits` can be used under the License below:

2.1.1 Copyright

Copyright © `genomebits`

Permission to use, copy, and distribute the `genomebits` software and its documentation for educational purposes without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

Permission to modify the software is granted, but not the right to distribute the complete modified source code. Modifications are to be distributed as patches to the released version. Permission to distribute binaries produced by compiling modified sources is granted, provided you

1. distribute the corresponding source modifications from the released version in the form of a patch file along with the binaries,
2. add special version identification to distinguish your version in addition to the base release version number,
3. provide your name and address as the primary contact for the support of your modified version, and
4. retain our contact information (`GitHub - genomebits`) in regard to use of the `genomebits` base software.

Permission to distribute the released version of any `genomebits` code along with corresponding source modifications in the form of a patch file is granted with same provisions 2 through 4 for binary distributions.

This software/application is provided "*as is*" without any express or implied warranty.

3 Install

genomebits is an analyzing tool for uncovering characteristics signals from genome (FASTA) sequences.

To install/uninstall genomebits you can simply mouse click on the icon for the genomebits '.deb' binary file, or by means of command line in which you need to install first some extra packages and their dependencies.

3.1 Manual Install

3.2 The simple way

In order to install the package you can use the following command as super user (root):

```
sudo apt install ./genomebits-1.0.x-Linux.deb
```

The apt command will install genomebits and some necessary dependencies.

3.3 The hard way

If your apt system does not support your local file installing, you need to install first some extra packages and their dependencies.

You can check the list of needed packages by using the following command:

```
dpkg -I genomebits-1.0.x-Linux.deb
```

In 'Depends:' it is possibile to find the list of the required packages. These are: perl, python-tk, python-numpy, python2, and gnuplot.

To install the required packages (listed with the command above) issue the command: 'sudo apt-get install <pkg1> <pkg2>' and so on. For example,

```
sudo apt-get install python-tk, python-numpy  ...
```

Then, to install the genomebits (.deb) package anew type the command:

```
sudo dpkg -i genomebits-1.0.x-Linux.deb
```

An genomebits launcher icon will appear in your Show Applications of Linux Ubuntu Desktop as shown in Fig.1.

Fig.1: *Launcher icon for* genomebits *listed in Ubuntu Desktop's* Show Applications.

NOTE: The `genomebits` running shell scripts and data files can be found at:

```
/opt/genomebits/bin
/opt/genomebits/doc
/opt/genomebits/icons
/opt/genomebits/scripts
```

3.4 Uninstall

In order to uninstall the `openEyA (.deb)` package type the following command (and check that the `/opt/openeya-yt` directory is now empty!):

```
sudo dpkg -r genomebits
```

3.5 Updates

Check for updates of `genomebits` at GitHub: `https://github.com` A connection to the Internet is needed to download any future releases.

4 How To use?

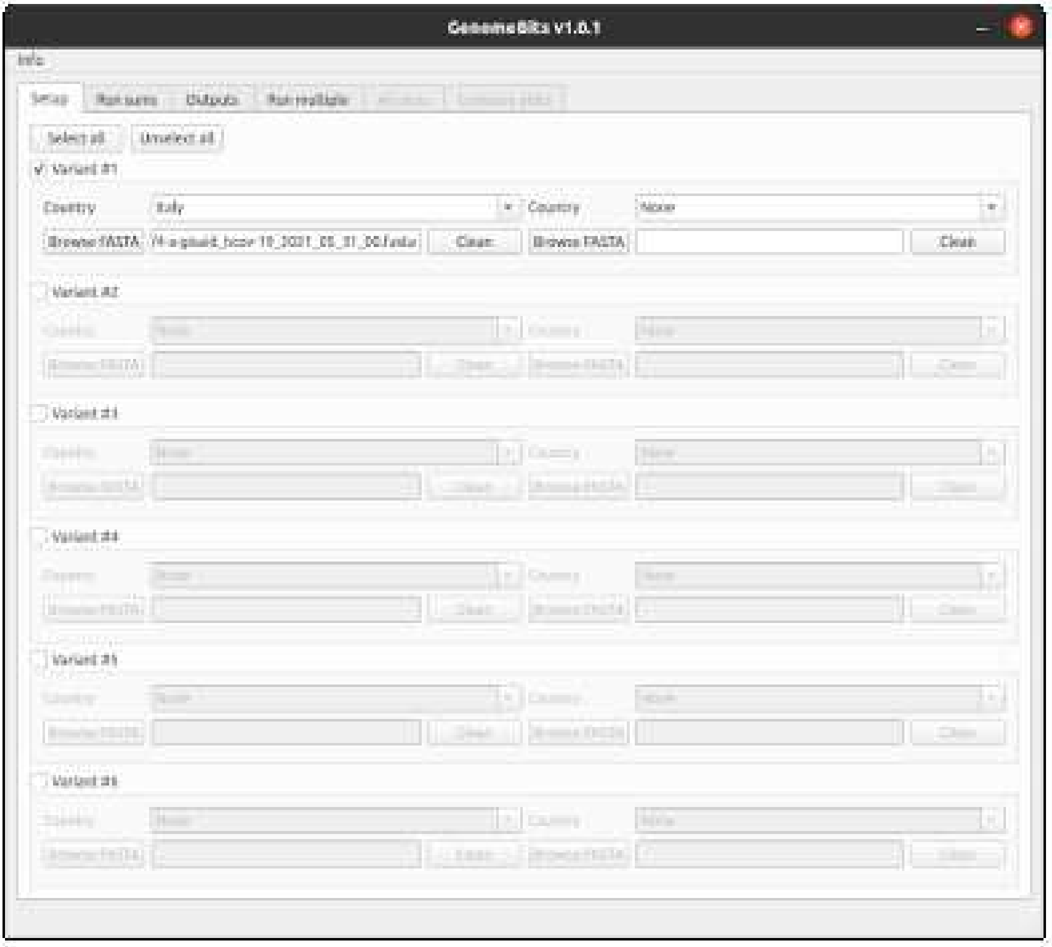
... It is very, very simple!

4.1 Run Alternating Sums

To start using the `genomebits` GUI for a quantitative examination of distinctive patterns of complete genome data, you need to input first (up to six-*times*-two) the paths to the FASTA files for the variants/species containing (*i.e.*, concatenating) more than one genome sequence each. As shown in figure 2, it also necessary to select in the '`Setup Tab`' (up to two) Countries from which the sequences were submitted by the laboratories.

Several thousands of complete FASTA sequences are available from the GISAID Initiative, such as those for the ongoing Human CoV-2 genome variants from all Continents. See: www.gisaid.org

By default, as example, the GUI links to a set of FASTA genome sequences obtained from Italy.



GenomeBits v1.0.1

Info

Setup | Run sums | Outputs | Run multiple

Select all | Unselect all

☑ Variant #1

Country: Italy | Country: None

Browse FASTA: /4-a-gisaid_hcov-19_2021_05_31_00.fasta | Clean | Browse FASTA: | Clean

☐ Variant #2

☐ Variant #3

☐ Variant #4

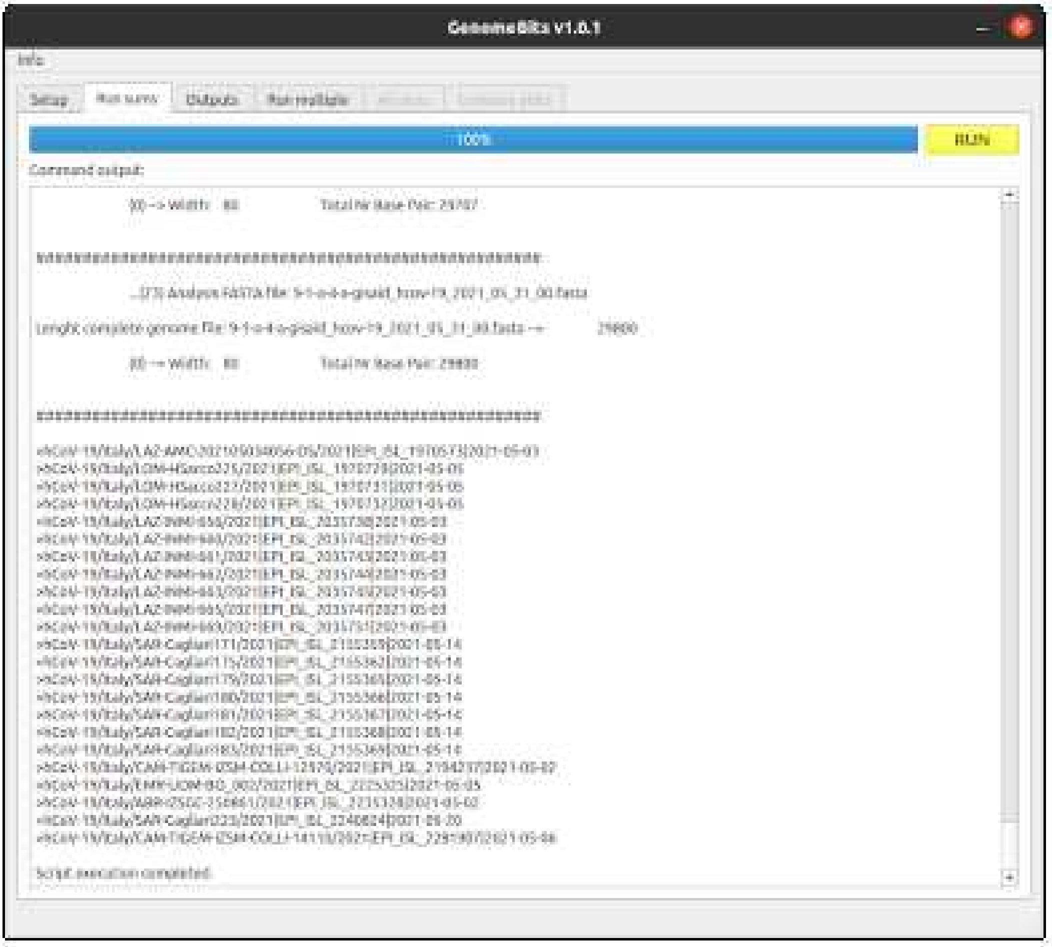
☐ Variant #5

☐ Variant #6

Fig.2: *Default* `genomebits` *GUI.*

4.1.1 FASTA Outputs

To calculate the alternating sums in Eqn.(1) for up to six-*times*-two inputs of FASTA files containing (*i.e.*, concatenating) more than one genome sequence each (submitted via the `genomebits` inputs in figure 2), press the RUN button in the '`Run sums`' Tab as in figure 3 to get the clickable Menu Tree tab:



```
                (0) --> Width:  80              Total Nr Base Pair: 29707


############################################################

                ...(73) Analysis FASTA file: 9-1-o-4-a-gisaid_hcov-19_2021_05_31_00.fasta

Lenght complete genome File: 9-1-o-4-a-gisaid_hcov-19_2021_05_31_00.fasta -->          29800

                (0) --> Width:  80              Total Nr Base Pair: 29800


############################################################

>hCoV-19/Italy/LAZ-AMC-202105034056-05/2021|EPI_ISL_1970573|2021-05-03
>hCoV-19/Italy/LOM-HSacco225/2021|EPI_ISL_1970729|2021-05-05
>hCoV-19/Italy/LOM-HSacco227/2021|EPI_ISL_1970731|2021-05-05
>hCoV-19/Italy/LOM-HSacco228/2021|EPI_ISL_1970732|2021-05-05
>hCoV-19/Italy/LAZ-INMI-656/2021|EPI_ISL_2035738|2021-05-03
>hCoV-19/Italy/LAZ-INMI-660/2021|EPI_ISL_2035742|2021-05-03
>hCoV-19/Italy/LAZ-INMI-661/2021|EPI_ISL_2035743|2021-05-03
>hCoV-19/Italy/LAZ-INMI-662/2021|EPI_ISL_2035744|2021-05-03
>hCoV-19/Italy/LAZ-INMI-663/2021|EPI_ISL_2035745|2021-05-03
>hCoV-19/Italy/LAZ-INMI-665/2021|EPI_ISL_2035747|2021-05-03
>hCoV-19/Italy/LAZ-INMI-669/2021|EPI_ISL_2035751|2021-05-03
>hCoV-19/Italy/SAR-Cagliari171/2021|EPI_ISL_2155359|2021-05-14
>hCoV-19/Italy/SAR-Cagliari175/2021|EPI_ISL_2155362|2021-05-14
>hCoV-19/Italy/SAR-Cagliari179/2021|EPI_ISL_2155365|2021-05-14
>hCoV-19/Italy/SAR-Cagliari180/2021|EPI_ISL_2155366|2021-05-14
>hCoV-19/Italy/SAR-Cagliari181/2021|EPI_ISL_2155367|2021-05-14
>hCoV-19/Italy/SAR-Cagliari182/2021|EPI_ISL_2155368|2021-05-14
>hCoV-19/Italy/SAR-Cagliari183/2021|EPI_ISL_2155369|2021-05-14
>hCoV-19/Italy/CAM-TIGEM-IZSM-COLLI-12570/2021|EPI_ISL_2194217|2021-05-02
>hCoV-19/Italy/EMR-UOM-BO_002/2021|EPI_ISL_2225325|2021-05-05
>hCoV-19/Italy/ABR-IZSGC-356461/2021|EPI_ISL_2235328|2021-05-02
>hCoV-19/Italy/SAR-Cagliari223/2021|EPI_ISL_2240624|2021-05-20
>hCoV-19/Italy/CAM-TIGEM-IZSM-COLLI-14110/2021|EPI_ISL_2291807|2021-05-06

Script execution completed.
```

Fig.3: *Computation of the alternating sums in Eqn.(1).*

4.1.2 Get/Save Single FASTA Files

After running the alternating sums in Eqn.(1), one can separate and get/save the single FASTA files (for each Country), which may be contained in the single FASTA input file including more than one genome sequence. As shown below, just select a FASTA file with the mouse within the '`Outputs Tab`'.

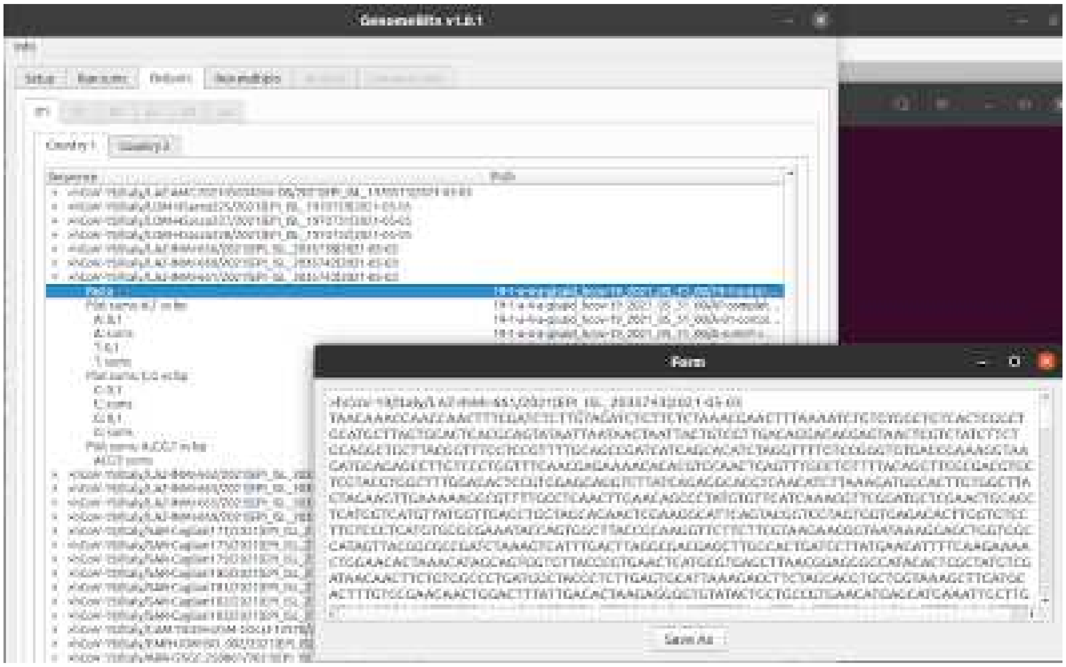
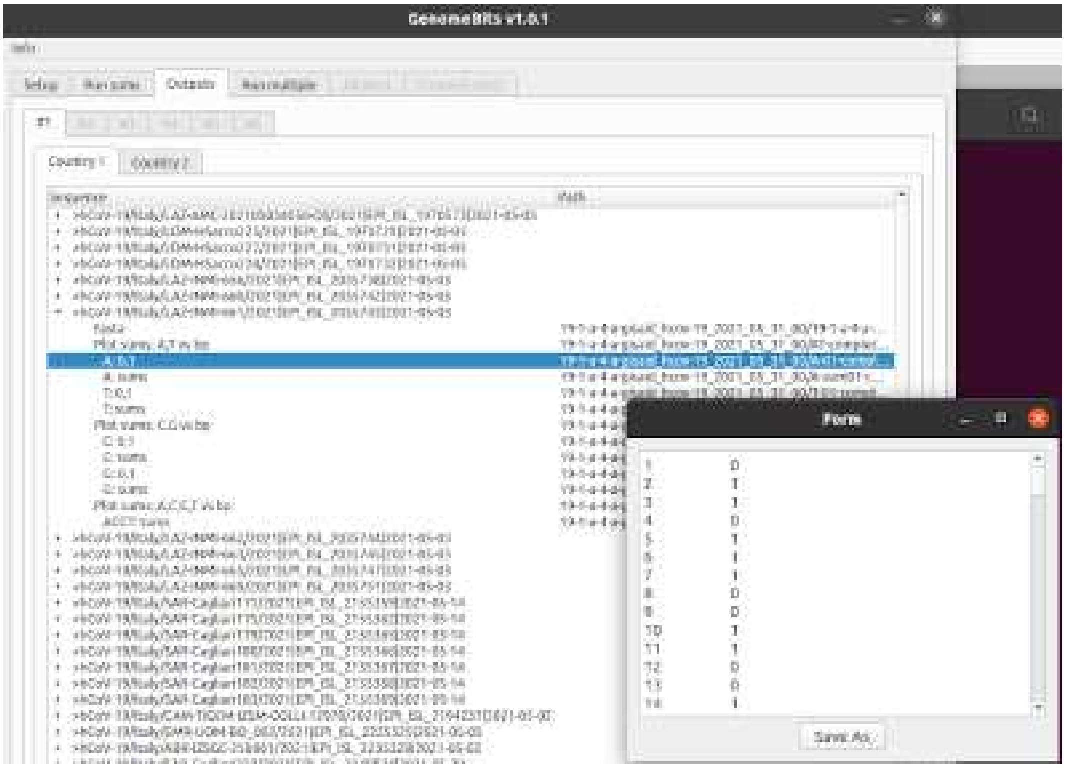


Fig.4: *Visualize and save single FASTA files for each Country.*

4.1.3 Associated Binaries for A, C, T, G

From the '`Outputs Tab`' you can also get single files for each of the four nucleotide bases A, C, T, G associated with binary 0 or 1 values according to its presence along the complete single genome sequences.



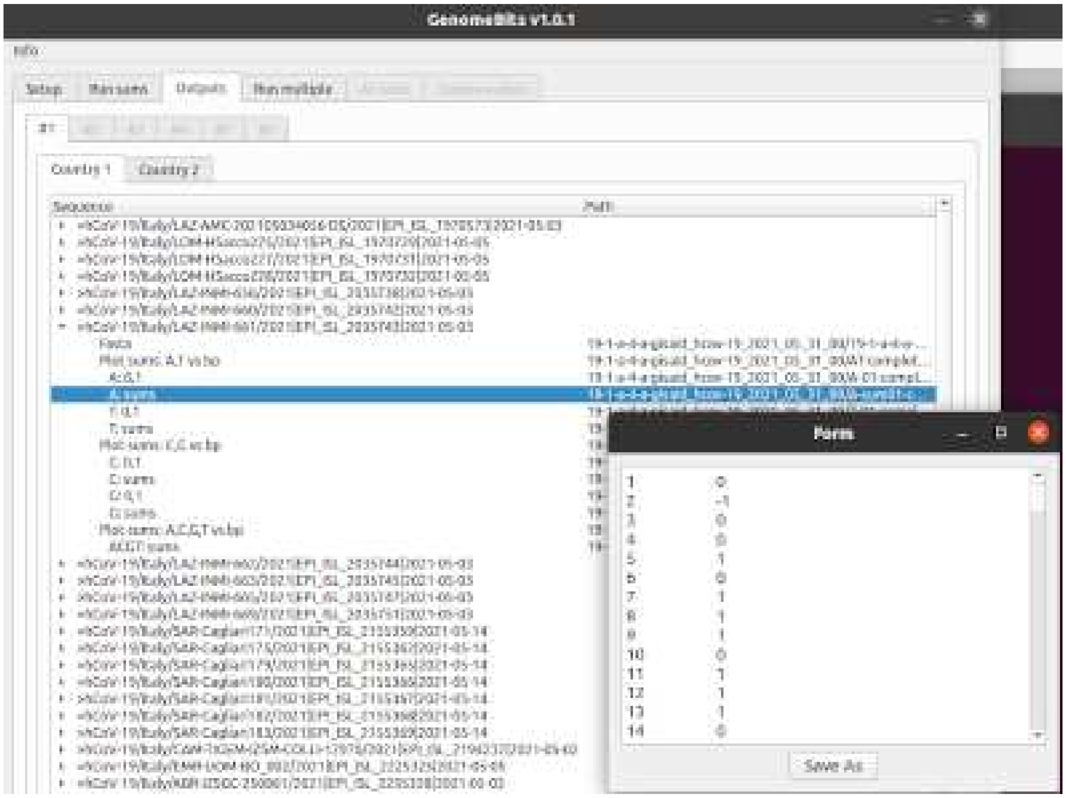
Fasta
Plot sums: A,T vs bp
A: 0,1
A: sums
T: 0,1
T: sums
Plot sums: C,G vs bp
C: 0,1
C: sums
G: 0,1
G: sums
Plot sums: A,C,G,T vs bp
ACGT sums

```
1     0
2     1
3     1
4     0
5     1
6     1
7     1
8     0
9     0
10    1
11    1
12    0
13    0
14    1
```

Fig.5: *Visualize and save associated binaries for each nucleotide bases A, C, T, G from a single genome sequence.*

4.1.4 Alternating Sums for A, C, T, G

You can also get the alternating sums results in single files for the four nucleotide bases A, C, T, G associated with binary 0 or 1 values according to its presence along the genome sequences.



```
▸ >hCoV-19/Italy/LAZ-AMC-20210503406-DS/2021|EPI_ISL_1970573|2021-05-03
▸ >hCoV-19/Italy/LOM-HSacco275/2021|EPI_ISL_1970729|2021-05-05
▸ >hCoV-19/Italy/LOM-HSacco277/2021|EPI_ISL_1970731|2021-05-05
▸ >hCoV-19/Italy/LOM-HSacco278/2021|EPI_ISL_1970732|2021-05-05
▸ >hCoV-19/Italy/LAZ-INMI-636/2021|EPI_ISL_2035738|2021-05-03
▸ >hCoV-19/Italy/LAZ-INMI-660/2021|EPI_ISL_2035742|2021-05-03
▾ >hCoV-19/Italy/LAZ-INMI-661/2021|EPI_ISL_2035743|2021-05-03
      Fasta                      19-1-a-4-a-gisaid_hcov-19_2021_05_31_00/19-1-a-4-a-...
      Plot sums: A,T vs bp       19-1-a-4-a-gisaid_hcov-19_2021_05_31_00/A1 complet...
        A: 0,1                   19-1-a-4-a-gisaid_hcov-19_2021_05_31_00/A-01 compl...
        A: sums                  19-1-a-4-a-gisaid_hcov-19_2021_05_31_00/A-complet...
        T: 0,1                   19-
        T: sums                  19-
      Plot sums: C,G vs bp       19-
        C: 0,1                   19-
        C: sums                  19-
        G: 0,1                   19-
        G: sums                  19-
      Plot sums: A,C,G,T vs bp   19-
        A,C,G,T sums             19-
▸ >hCoV-19/Italy/LAZ-INMI-662/2021|EPI_ISL_2035744|2021-05-03
▸ >hCoV-19/Italy/LAZ-INMI-663/2021|EPI_ISL_2035745|2021-05-03
▸ >hCoV-19/Italy/LAZ-INMI-665/2021|EPI_ISL_2035747|2021-05-03
▸ >hCoV-19/Italy/LAZ-INMI-669/2021|EPI_ISL_2035751|2021-05-03
▸ >hCoV-19/Italy/SAR-Cagliari171/2021|EPI_ISL_2155350|2021-05-14
▸ >hCoV-19/Italy/SAR-Cagliari175/2021|EPI_ISL_2155362|2021-05-14
▸ >hCoV-19/Italy/SAR-Cagliari179/2021|EPI_ISL_2155365|2021-05-14
▸ >hCoV-19/Italy/SAR-Cagliari180/2021|EPI_ISL_2155366|2021-05-14
▸ >hCoV-19/Italy/SAR-Cagliari181/2021|EPI_ISL_2155367|2021-05-14
▸ >hCoV-19/Italy/SAR-Cagliari182/2021|EPI_ISL_2155368|2021-05-14
▸ >hCoV-19/Italy/SAR-Cagliari183/2021|EPI_ISL_2155369|2021-05-14
▸ >hCoV-19/Italy/CAM-TIGEM-IZSM-COLLI-17976/2021|EPI_ISL_2196237|2021-05-03
▸ >hCoV-19/Italy/EMR-UOM-BO_892/2021|EPI_ISL_2225325|2021-05-05
▸ >hCoV-19/Italy/ABR-IZSGC-250061/2021|EPI_ISL_2225338|2021-05-03
```

```
1           0
2           -1
3           0
4           0
5           1
6           0
7           1
8           1
9           1
10          0
11          1
12          1
13          1
14          0
```

Fig.6: *Visualize and save the alternating sums results for each nucleotide bases A, C, T, G from a single genome sequence.*

4.1.5 Alternating Sums Plots for A, T and C, G

From within the '`Outputs Tab`' you can see Plots of the alternating sums of the binary data files to compare behaviour of the pairs A, T and C, G nucleotide bases versus bases position (bp) from a single genome sequence.

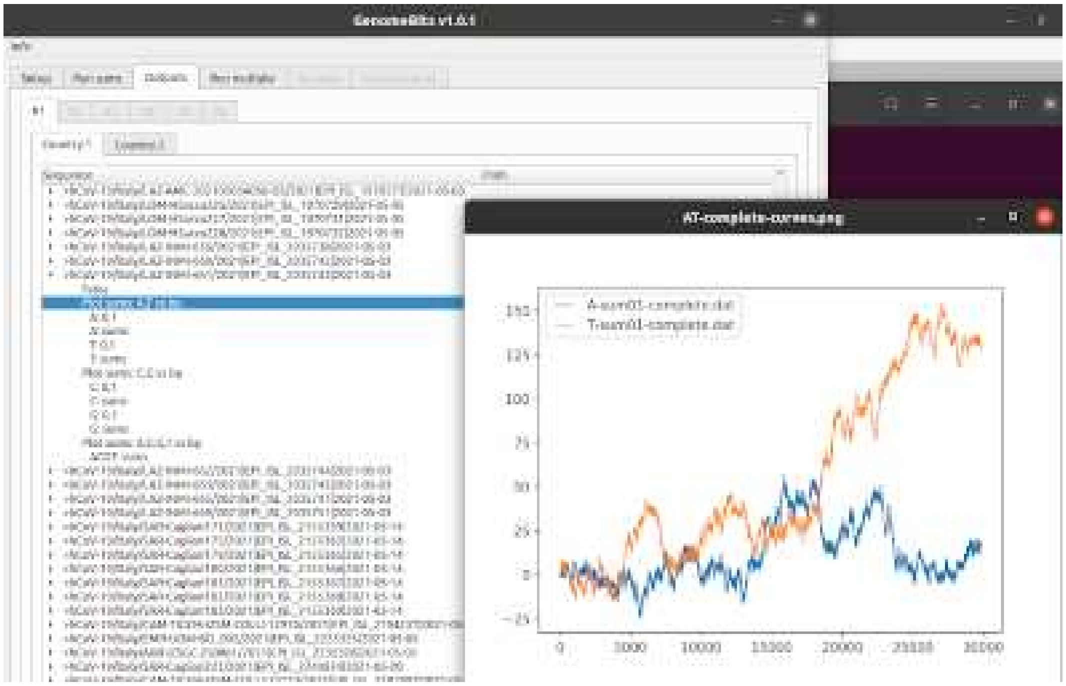


Fig.7: *Plots of alternating sums of binary data files for the pairs A, T and C, G.*

4.1.6 Alternating Sums Plots for A, C, T, G

As in figure 8, you can get a comparison in one plot the alternating sums results versus bases position (bp) for all four nucleotide bases A, C, T, G associated with binary 0 or 1 values according to its presence along the genome sequences for each Country.

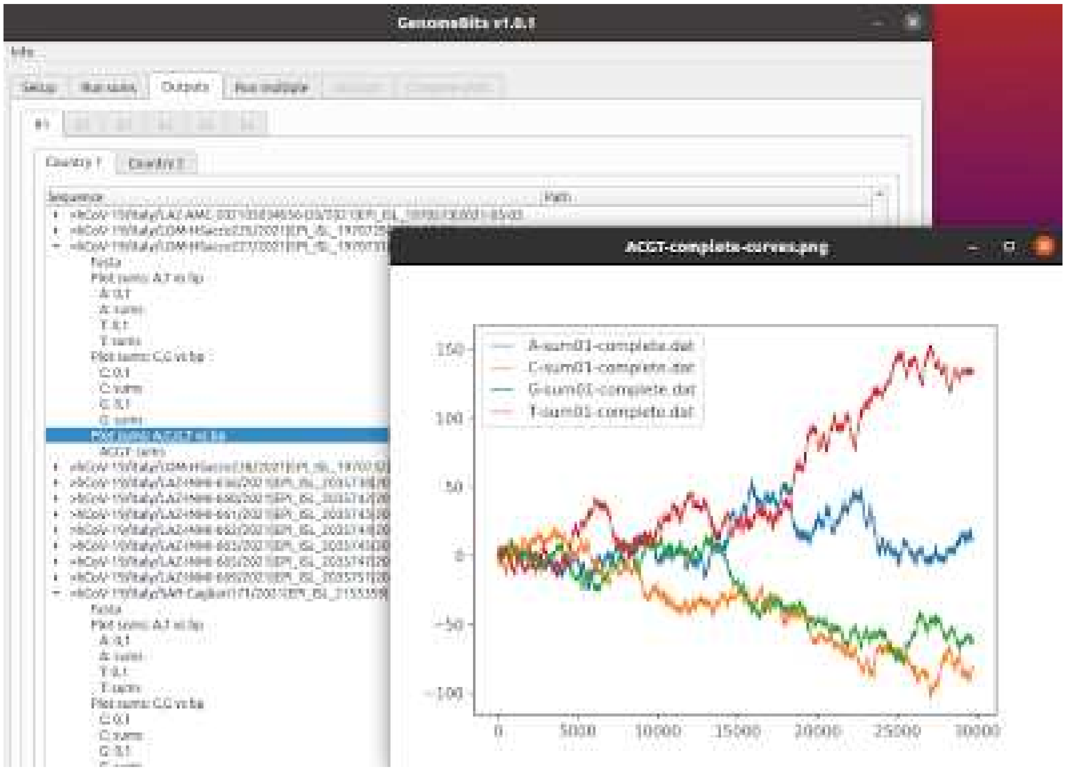


Fig.8: *Plots of alternating sums of binary data files for all A, C, T, G.*

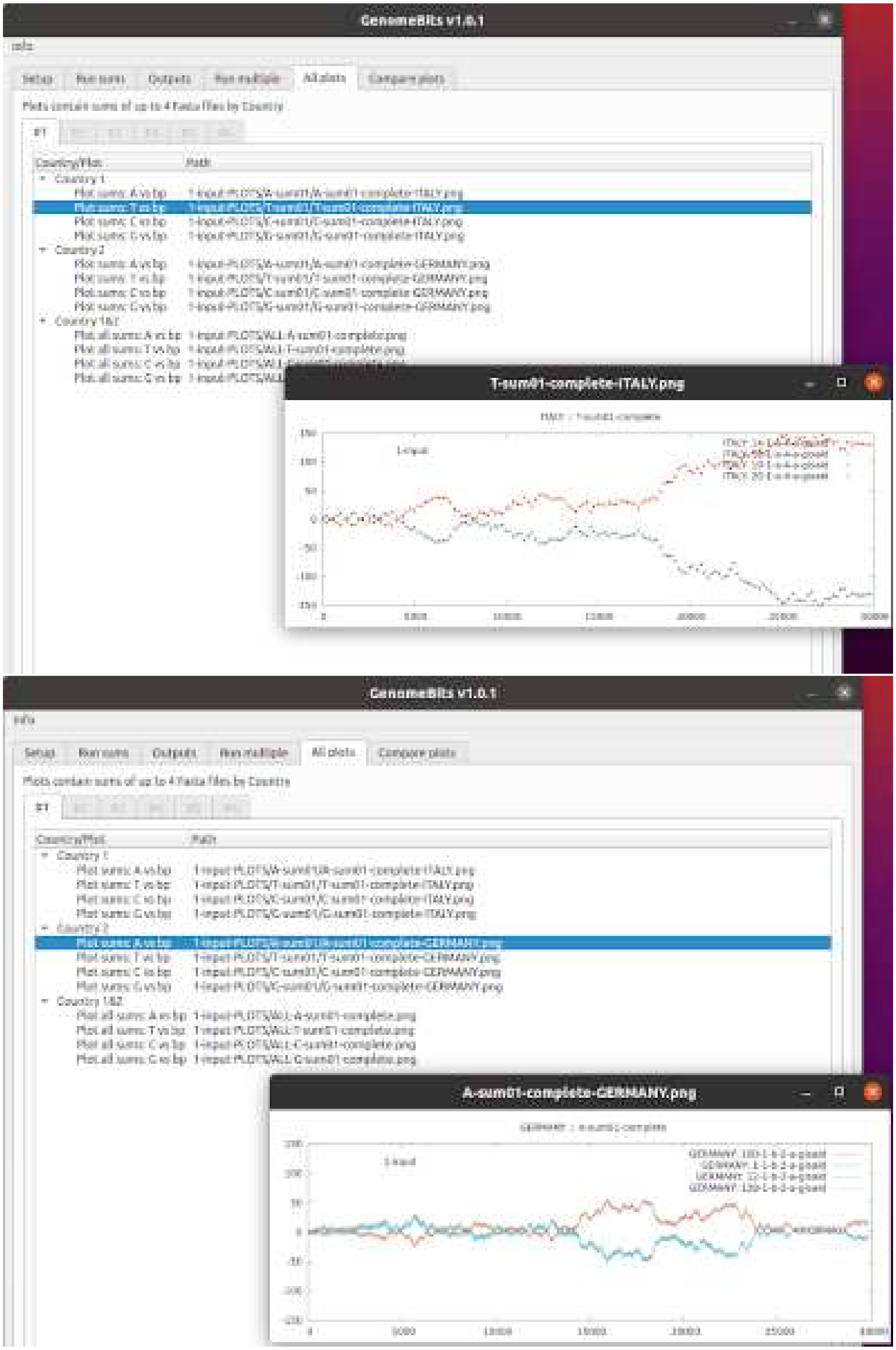
4.2 Run Multiple

The `genomebits` GUI also allows you to generated several other signal plots obtained from Eqn.(1) for each of the four nucleotide bases A, C, T, G to compare results from (up to six) given variants/species and (up to two) selected Countries in each case.

Finally, it is also possible to compare visualmente the `genomebits` GUI results with the results in the original paper "*Uncovering Signals from the Coronavirus Genome*".

4.2.1 Alternating Sums Plots for A, C, T, G for each Country

As shown below, you can plot the alternating sums results versus bases position (bp) for the four nucleotide bases A, C, T, G associated with binary 0 or 1 values according to its presence along the genome sequences *for each* single Country and up to 4 FASTA files for each variant/specie.



Plots contain sums of up to 4 Fasta files by Country

Country/Plot	Path
Country 1	
Plot sums: A vs bp	1-input-PLOTS/A-sum01/A-sum01-complete-ITALY.png
Plot sums: T vs bp	1-input-PLOTS/T-sum01/T-sum01-complete-ITALY.png
Plot sums: C vs bp	1-input-PLOTS/C-sum01/C-sum01-complete-ITALY.png
Plot sums: G vs bp	1-input-PLOTS/G-sum01/G-sum01-complete-ITALY.png
Country 2	
Plot sums: A vs bp	1-input-PLOTS/A-sum01/A-sum01-complete-GERMANY.png
Plot sums: T vs bp	1-input-PLOTS/T-sum01/T-sum01-complete-GERMANY.png
Plot sums: C vs bp	1-input-PLOTS/C-sum01/C-sum01-complete-GERMANY.png
Plot sums: G vs bp	1-input-PLOTS/G-sum01/G-sum01-complete-GERMANY.png
Country 1&2	
Plot all sums: A vs bp	1-input-PLOTS/ALL-A-sum01-complete.png
Plot all sums: T vs bp	1-input-PLOTS/ALL-T-sum01-complete.png
Plot all sums: C vs bp	1-input-PLOTS/ALL-C-sum01-complete.png
Plot all sums: G vs bp	1-input-PLOTS/ALL

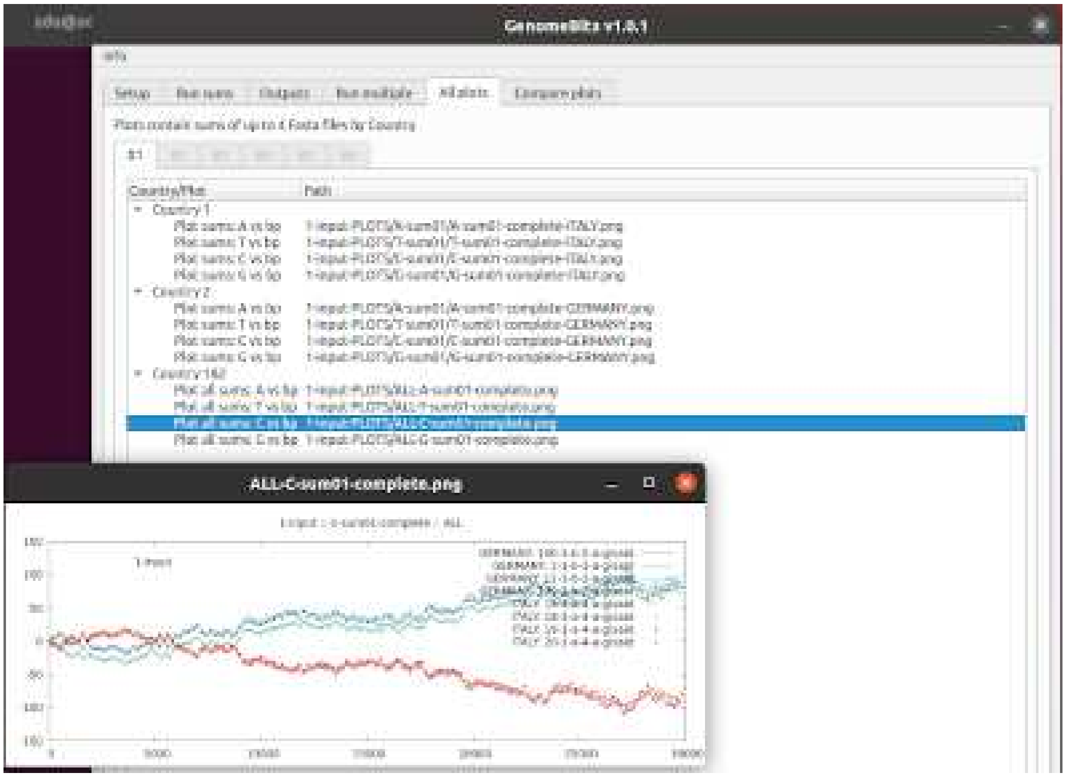
Plots contain sums of up to 4 Fasta files by Country

Country/Plot	Path
Country 1	
Plot sums: A vs bp	1-input-PLOTS/A-sum01/A-sum01-complete-ITALY.png
Plot sums: T vs bp	1-input-PLOTS/T-sum01/T-sum01-complete-ITALY.png
Plot sums: C vs bp	1-input-PLOTS/C-sum01/C-sum01-complete-ITALY.png
Plot sums: G vs bp	1-input-PLOTS/G-sum01/G-sum01-complete-ITALY.png
Country 2	
Plot sums: A vs bp	1-input-PLOTS/A-sum01/A-sum01-complete-GERMANY.png
Plot sums: T vs bp	1-input-PLOTS/T-sum01/T-sum01-complete-GERMANY.png
Plot sums: C vs bp	1-input-PLOTS/C-sum01/C-sum01-complete-GERMANY.png
Plot sums: G vs bp	1-input-PLOTS/G-sum01/G-sum01-complete-GERMANY.png
Country 1&2	
Plot all sums: A vs bp	1-input-PLOTS/ALL-A-sum01-complete.png
Plot all sums: T vs bp	1-input-PLOTS/ALL-T-sum01-complete.png
Plot all sums: C vs bp	1-input-PLOTS/ALL-C-sum01-complete.png
Plot all sums: G vs bp	1-input-PLOTS/ALL-G-sum01-complete.png

Figs.9: *Plots for a Country and up to 4 FASTA files the alternating sums of binary data files for* A, C, T, G.

4.2.2 Alternating Sums Plots for A, C, T, G Both Countries

Alternatively, as illustrated in figure 10, you can also plot the alternating sums results versus bases position (bp) for the four nucleotide bases A, C, T, G associated with binary 0 or 1 values according to its presence along the genome sequences *for both* Countries (if given so as input) and up to 4 FASTA files for each variant/specie.



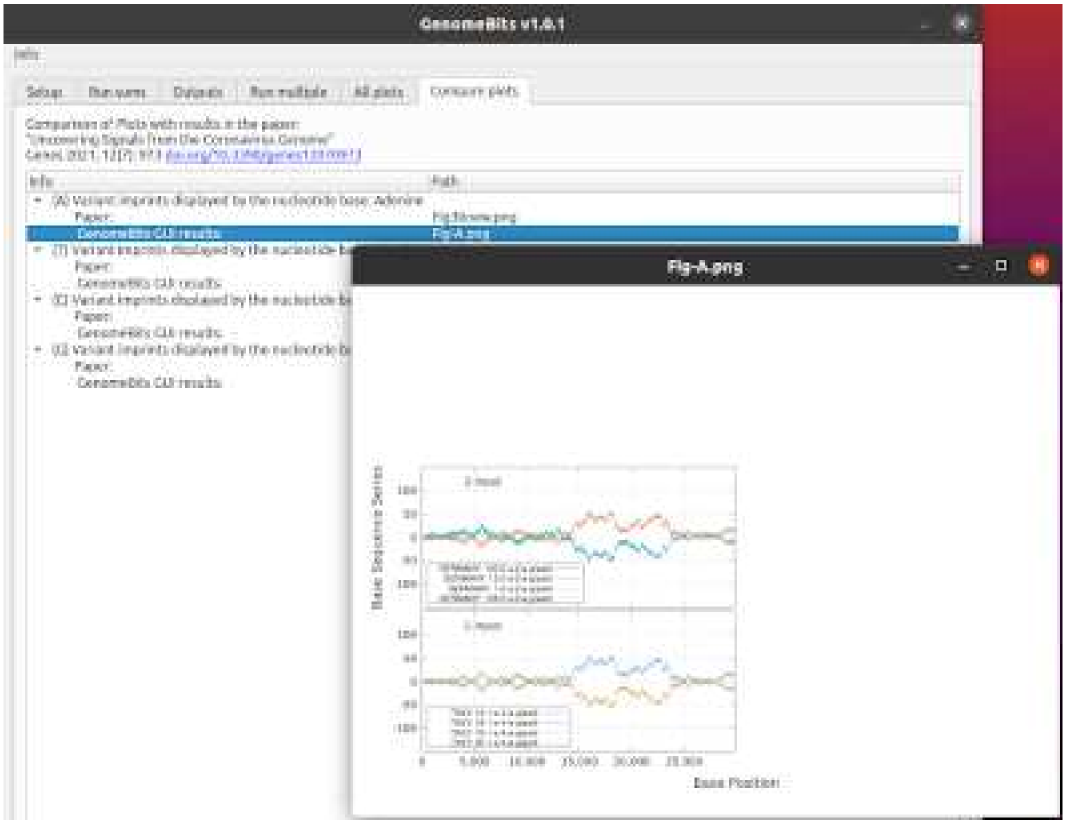
Plots contain sums of up to 4 Fasta files by Country

Country/Plot	Path
▾ Country 1	
Plot sums: A vs bp	1-input-PLOTS/A-sum01/A-sum01-complete-ITALY.png
Plot sums: T vs bp	1-input-PLOTS/T-sum01/T-sum01-complete-ITALY.png
Plot sums: C vs bp	1-input-PLOTS/C-sum01/C-sum01-complete-ITALY.png
Plot sums: G vs bp	1-input-PLOTS/G-sum01/G-sum01-complete-ITALY.png
▾ Country 2	
Plot sums: A vs bp	1-input-PLOTS/A-sum01/A-sum01-complete-GERMANY.png
Plot sums: T vs bp	1-input-PLOTS/T-sum01/T-sum01-complete-GERMANY.png
Plot sums: C vs bp	1-input-PLOTS/C-sum01/C-sum01-complete-GERMANY.png
Plot sums: G vs bp	1-input-PLOTS/G-sum01/G-sum01-complete-GERMANY.png
▾ Country 1&2	
Plot all sums: A vs bp	1-input-PLOTS/ALL-A-sum01-complete.png
Plot all sums: T vs bp	1-input-PLOTS/ALL-T-sum01-complete.png
Plot all sums: C vs bp	1-input-PLOTS/ALL-C-sum01-complete.png
Plot all sums: G vs bp	1-input-PLOTS/ALL-G-sum01-complete.png

Fig.10: *Plots for both Countries and up to 4 FASTA files the alternating sums of binary data files for A, C, T, G.*

4.2.3 Plot All `genomebits` GUI Results

As in the figure, one can plot in one image all the `genomebits` GUI results versus bases position (bp) for each nucleotide base A, C, T or G associated with binary 0 or 1 values according to its presence along the genome sequences for up to six variants/species and up to 4 FASTA files by Country.



Comparison of Plots with results in the paper:
"Uncovering Signals from the Coronavirus Genome"
Genes 2021, 12(7), 973 doi.org/10.3390/genes12070973

Info	Path
(A) Variant imprints displayed by the nucleotide base: Adenine	
Paper:	Fig3Alone.png
GenomeBits GUI results:	Fig-A.png
(T) Variant imprints displayed by the nucleotide ba	
Paper:	
GenomeBits GUI results:	
(C) Variant imprints displayed by the nucleotide ba	
Paper:	
GenomeBits GUI results:	
(G) Variant imprints displayed by the nucleotide ba	
Paper:	
GenomeBits GUI results:	

Fig.11: *Plot of all the* `genomebits` *GUI results.*

4.2.4 Compare `genomebits` GUI and Article Results

It is also possible to compare visualmente the `genomebits` GUI results versus bases position (bp) for (up to six) given variants/species and (up to two) selected Countries, with the results in the paper "*Uncovering Signals from the Coronavirus Genome*" -see: Genes 2021; 12(7):973. https://doi.org/10.3390/genes12070973 Complete coronavirus sequences with N nucleotides are of the order of 30,000 bp in length.

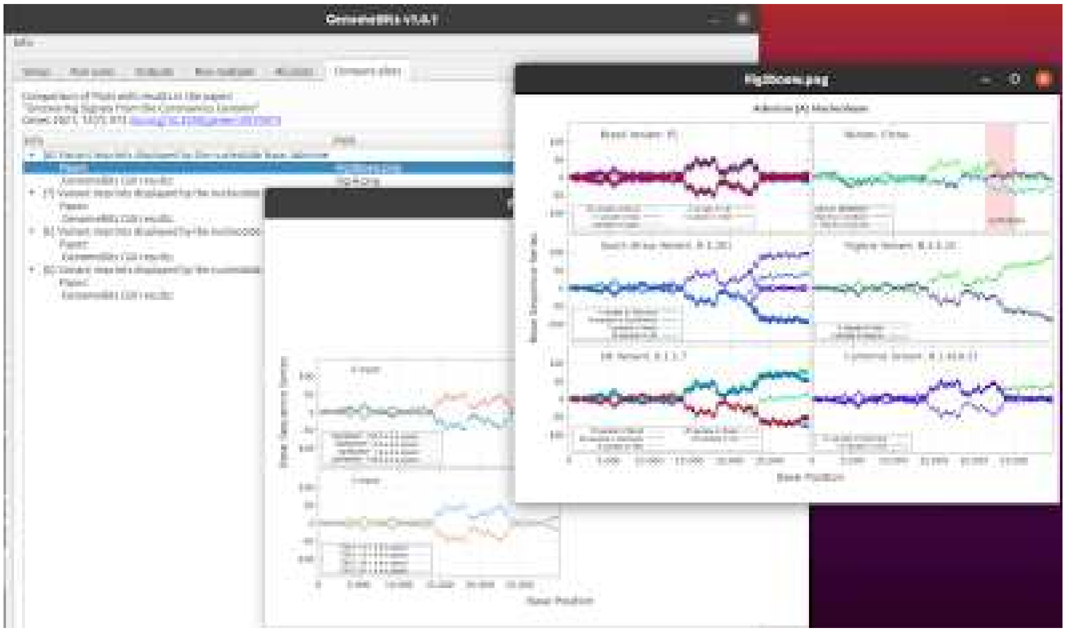


Fig.12: *Compare* `genomebits` *GUI results with Main Article Results for coronavirus genome.*

In summary, the `genomebots` GUI is based on the published quantitative method for the examination of distinctive patterns of complete genome data. One deals with a certain type of alternating finite series having terms converted to binary values (0,1) for the nucleotide bases (A)denine, (C)ytosine, (G)uanine and (T)hymine according to their progression along the genome sequences. This mapping into four binary projections of the sequences follows previous studies on the three-base periodicity characteristic of protein-coding DNA sequences. For example, by this method we uncover distinctive signals of the intrinsic gene organization revealed by the genome sequences of the single-stranded RNA coronaviruses.

The present `genomebots` tool is effective and easier to apply in protein sequence comparison. It is motivated by the need to identify genetic mechanisms involved in coronavirus spreading. The added value of the alternating sums of the type in Eqn.(1) is to have a distinctive function representation of naturally occurring genome sequences of variants. Plus and minus signs are chosen sequentially starting with `+1` by default. From the view of statistics, such a sequence is equivalent to a discrete-valued time series for statistical identification and characterization of data sets as studied in financial series.