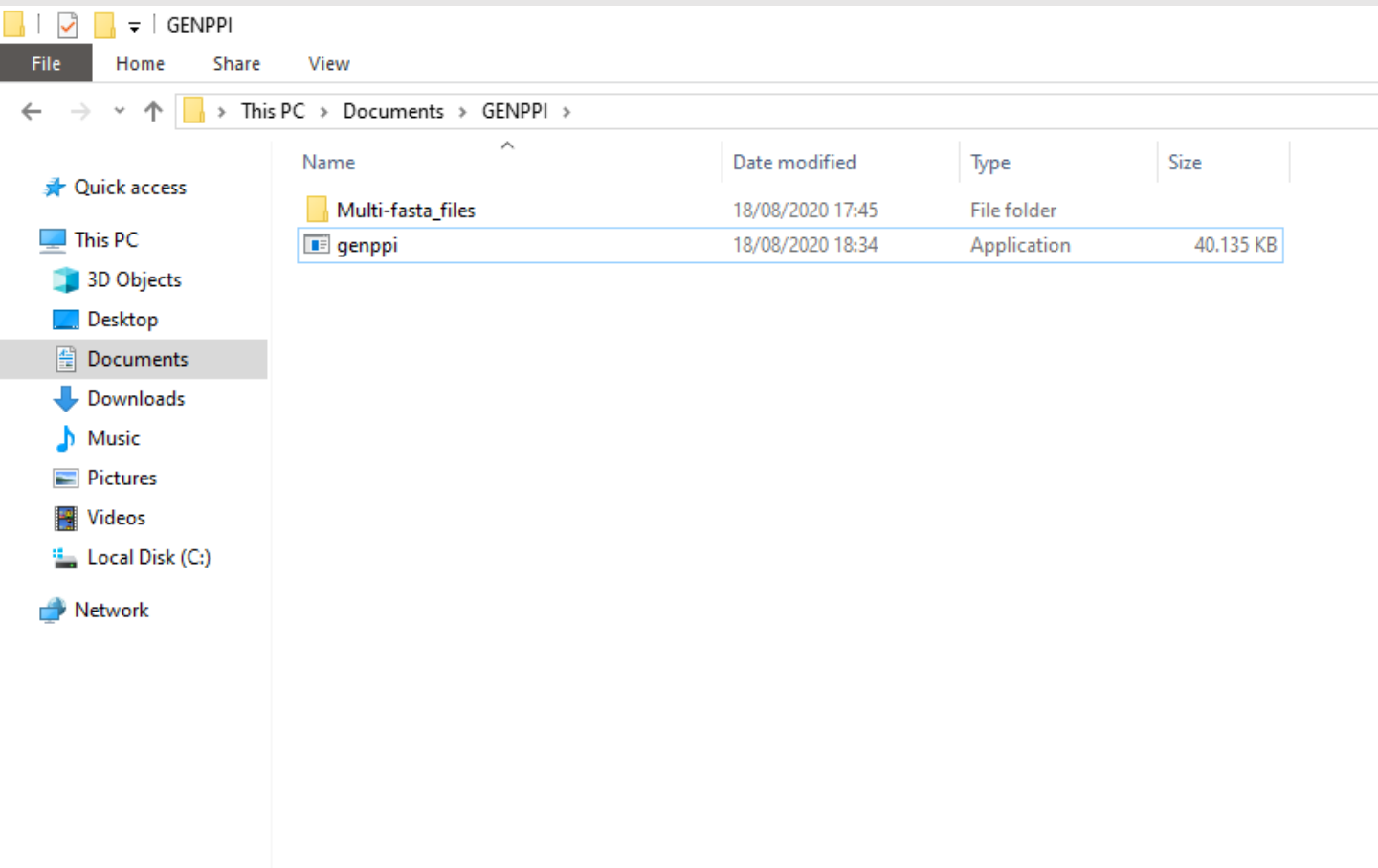


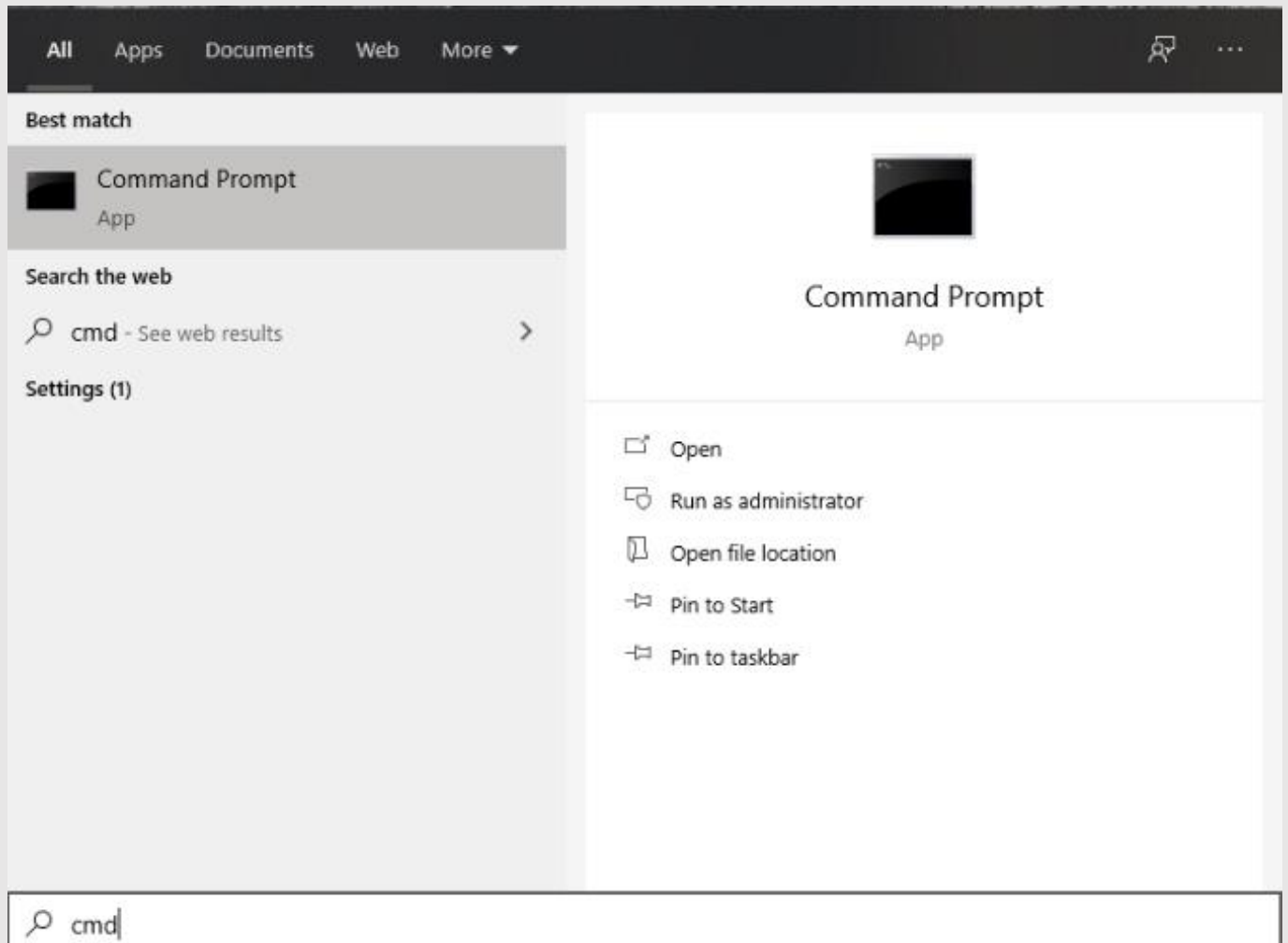
# GENPPI

Usage guide on Windows

- After downloading the executable, it is recommended to create a folder with the downloaded file and the multi-fasta files of predicted proteins for your genomes.
- If you prefer, rename the executable to an easier name like "genppi".



- Later, open the Windows command prompt.
  - This can be done by clicking on the Windows key and typing “cmd”.



- Once opened, the user can choose to run the program in their windows user's directory, as well as run the program after navigating the command prompt to the directory created for GENPPI

### ➤ User's directory path

```
Command Prompt
C:\Users\Lanes>dir
Volume in drive C has no label.
Volume Serial Number is 3978-2696

Directory of C:\Users\Lanes

19/08/2020  11:28    <DIR>        .
19/08/2020  11:28    <DIR>        ..
19/08/2020  11:29    <DIR>        .ssh
12/08/2020  23:43    <DIR>        3D Objects
12/08/2020  23:43    <DIR>        Contacts
18/08/2020  22:21    <DIR>        Desktop
19/08/2020  11:25    <DIR>        Documents
19/08/2020  15:06    <DIR>        Downloads
12/08/2020  23:43    <DIR>        Favorites
12/08/2020  23:43    <DIR>        Links
12/08/2020  23:43    <DIR>        Music
08/08/2020  22:13    <DIR>        OneDrive
19/08/2020  13:04    <DIR>        Pictures
13/08/2020  15:10    <DIR>        Roaming
12/08/2020  23:43    <DIR>        Saved Games
12/08/2020  23:43    <DIR>        Searches
12/08/2020  23:43    <DIR>        Videos
               0 File(s)              0 bytes
               17 Dir(s)  26.352.844.800 bytes free

C:\Users\Lanes>
```

### ➤ GENPPI's directory path

```
Command Prompt
C:\Users\Lanes\Documents\GENPPI>dir
Volume in drive C has no label.
Volume Serial Number is 3978-2696

Directory of C:\Users\Lanes\Documents\GENPPI

18/08/2020  17:47    <DIR>        .
18/08/2020  17:47    <DIR>        ..
18/08/2020  18:34    <DIR>        41.097.904 genppi.exe
19/08/2020  12:19    <DIR>        Multi-fasta_files
               1 File(s)      41.097.904 bytes
               3 Dir(s)  26.350.071.808 bytes free

C:\Users\Lanes\Documents\GENPPI>
```

- When executing the downloaded file, you have a list of all possible usage parameters

➤ Type the name of the file (e.g. genppi.exe)

```

C:\Users\Lanes\Documents\GENPPI>genppi.exe

Mandatory parameter

Directory parameter:
-dir <workingdir> directory

Optional parameters

Parameters of ppi by conserved neighbourhood:
-cnp <conserved neighbourhood score percentage> (65-default)
-expt <type of expansion of the conserved gene neighborhood> 'fixed' or 'dynamic' (fixed-default)

Parameters of conserved neighbourhood by fixed expansion:
-w1 <window size 1> (10-default)
-cw1 <gene conservation required for window 1> (4-default)
-w2 <window size 2> (7-default)
-cw2 <gene conservation required for window 2> (3-default)
-w3 <window size 3> (5-default)
-cw3 <gene conservation required for window 3> (2-default)
-w4 <window size 4> (3-default)
-cw4 <gene conservation required for window 4> (1-default)

Parameters of conserved neighbourhood by dynamic expansion:
-ws <dynamic expansion window size> (1-default)

Parameters of ppi by phylogenetic profile:
-ppp <phylogenetic profiles score percentage> (30-default)
-ppdifftolerated <difference in phylogenetic profiles tolerated to infer ppi> (0-default)

Amino acid histogram parameter settings for the phylogenetic profile:
-pphistofilter <build the phylogenetic profile of genes with a higher percentage of identity>
-ppaaddiffmlimit <amino acid difference limit> (0-default)
-ppaachckminlimit <minimum amount of amino acids to check> (26-default)

Methods of ppi prediction by phylogenetic profile and its parameters:

Method 1 - Predict ppi by phylogenetic profile only for interactions predicted by conserved neighborhood
-ppcn

Method 2 - PPI prediction by phylogenetic profile without filters 2
-ppcomplete

Method 3 - Prediction of ppi by phylogenetic profile with a limit of interactions
-ppiterlimit <maximum number of interactions desired> (500000-default)

Method 4 - Prediction of ppi by phylogenetic profile with interactions limit by weight
-trim <maximum number of interactions by weight> (45000-default)

Method 5 - Prediction of ppi by the phylogenetic profile only for genes with profiles that cover a greater or lesser number of genomes than an informed threshold
-threshold <phylogenetic profiles threshold>
-plusminus <parameter that receives the greater than or less than sign to apply the ppthreshod filter> '<' or '>' (signs greater than and less than, must be enclosed in single or double quotes)

Method 6 - Delete groups of ppi predicted by phylogenetic profile that exceed a limit of interactions by weight
-grouplimit <limit of tolerated interactions to maintain a group of ppi with the same weight> (45000-default)

Method 7 - To exclude genes with unwanted profiles in predicting ppi by phylogenetic profile
-profiles <number of genomes in the unwanted profiles> Entry example: 7 (genes that co-occur in a total of 7 genomes will be excluded). To insert more than one profile, the entry must be enclosed in single or double quotes, and the values separated by semicolons. Example: "7; 15; 21"

```

Obs: In this example, we use the GENPPI's directory path.

- Thus, the execution of the set of parameters present in the README occurs as follows:

➤ After navigating to the GENPPI's directory, type or paste the set of parameters "genppi.exe -expt fixed -w1 7 -cw1 4 -ppcomplete -ppdiff tolerated 1 -pphistofilter -dir Multi-fasta\_files\"

```

C:\Users\Lanes\Documents\GENPPI>genppi.exe -expt fixed -w1 7 -cw1 4 -ppcomplete -ppdiff tolerated 1 -pphistofilter -dir Multi-fasta_files\

GENPPI VERSION: 1.0
RELEASE NUMBER: 4354980a4dfaede2695e82e525914cead0f2c9ee
REPOSITORY: https://github.com/santosardr/genppi

Directory parameter:
-dir <workingdir> = Multi-fasta_files\

Parameters of ppi by conserved neighbourhood:
-cnp <conserved neighbourhood score percentage> = 70%
-expt <type of expansion of the conserved gene neighborhood> = fixed (default)
-w1 <window size 1> = 7
-cw1 <gene conservation required for window 1> = 4 (default)
-w2 <window size 2> = 7 (default)
-cw2 <gene conservation required for window 2> = 3 (default)
-w3 <window size 3> = 5 (default)
-cw3 <gene conservation required for window 3> = 2 (default)
-w4 <window size 4> = 3 (default)
-cw4 <gene conservation required for window 4> = 1 (default)

Parameters of ppi by phylogenetic profile:
Method 2 - PPI prediction by phylogenetic profile without filters
-ppp <phylogenetic profiles score percentage> = 30% (default)
-ppdiff tolerated <difference in phylogenetic profiles tolerated to infer ppi> = 1
-pphistofilter <build the phylogenetic profile of genes with a higher percentage of identity>
-ppaaddiff limit <amino acid difference limit> = 0 (default)
-ppaachemin limit <minimum amount of amino acids to check> = 26 (default)

Amino acid histogram parameters:
-aaddiff limit <amino acid difference limit> 1 (default)
-aachemin limit <minimum amount of amino acids to check> 25 (default)

Making ppi prediction, please wait.

Generating amino acids histogram;
[05 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100]%
[=====]

Generating pan-genome;
[05 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100]%
[=====]

Predicting ppi by conserved gene neighborhood;
[05 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100]%
[=====]

Predicting ppi by phylogenetic profiles;
[05 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100]%
[=====]

Writing pan-genome format 1;
[05 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100]%
[=====]

Writing pan-genome format 2;
[05 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100]%
[=====]

Writing Gene neighborhood conservation report;
[05 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100]%

```

Obs: Depending on the number of files, the analysis can take hours. Don't close the command prompt before the analysis is finished, as you can interrupt the process.

- In the end, the results will be inside the directory with the analyzed files.

Name	Type	Size
gene-neighborhood-conservation-report	File folder	
pan-genome	File folder	
phylogenetic-profiles-report	File folder	
ppi-files	File folder	
ppi-report	File folder	
slCacc.faa	FAA File	732 KB
slCammo.faa	FAA File	790 KB
slCatyp.faa	FAA File	514 KB
slCauri.faa	FAA File	840 KB
slCauris.faa	FAA File	473 KB
slCcallu.faa	FAA File	857 KB
slCcampo.faa	FAA File	750 KB
slCcasei.faa	FAA File	917 KB
slCcoyle.faa	FAA File	788 KB
slCcrudi.faa	FAA File	934 KB
slCcyst.faa	FAA File	889 KB
slCcysti.faa	FAA File	889 KB
slCdeser.faa	FAA File	884 KB
slCdip.faa	FAA File	739 KB
slCdip01.faa	FAA File	723 KB
slCdip02.faa	FAA File	734 KB
slCdip03.faa	FAA File	704 KB
slCdip04.faa	FAA File	759 KB