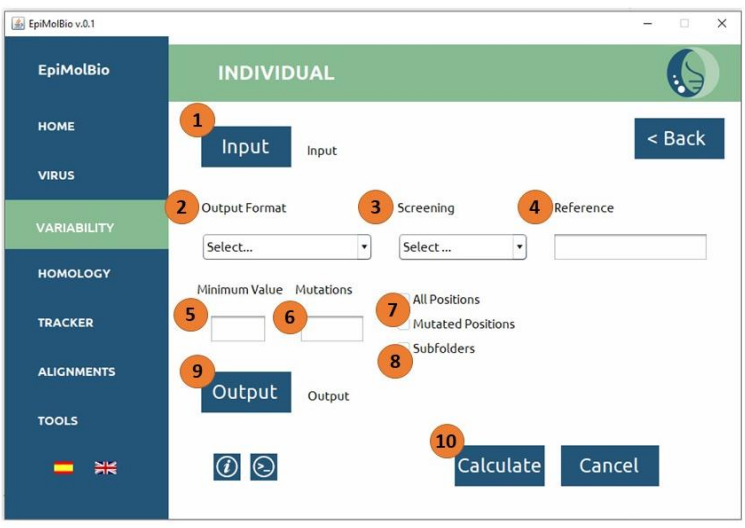
Variability Individual Polimorphisms

‘Mutated Positions’ and ‘Mutations Table’ enable the detection of polymorphisms, providing information about their location and frequency of occurrence using any sequence introduced by the user as a reference.

‘Markers’ allows the detection of mutations exclusive to each file compared to the rest of the input files.

‘Multiple Mutations’ enables the detection of mutation combinations providing their frequency of occurrence.

‘Mutations by Position’ allows the detection of residues at one position or several combined positions providing their frequency of occurrence.



① Select the input folder. It should only contain the .fasta files of the sequences to be analyzed, in either NT (nucleotide) or AA (amino acid) format, except for Markers and Multiple Mutations (only in AA). To perform the analysis on nucleotide sequences, you'll need to use the Find and Replace tool in File Editing to replace "N" with "?" to exclude them from the analysis.

⑧ The output formats **Mutations Table, Multiple Mutations**, and **Mutations by Position** allow the input file to contain subfolders. Check Subfolders if you want to analyze an input file with subfolders.

② Choose the output format, selecting from **Mutated Positions, Mutations Table, Markers, Multiple Mutations**, or **Mutations by Position**.

For **Mutated Positions:**

Variability Individual Polymorphisms All Positions

PR_01_AE.fasta

Position	Residues	Total Positions
P1	P(99.896%) S(0.078%) A(0.004%) L(0.007%) T(0.007%) H(0.004%) V(0.004%)	26838
Q2	Q(99.782%) E(0.071%) S(0.019%) H(0.056%) D(0.004%) K(0.023%) L(0.015%) P(0.008%) R(0.011%) T(0.004%) *(0.008%)	26649
V3	I(99.858%) V(0.078%) N(0.015%) L(0.041%) T(0.007%)	26831
T4	T(99.858%) M(0.004%) I(0.048%) N(0.019%) P(0.022%) S(0.034%) F(0.004%) A(0.007%) H(0.004%)	26816
L5	L(99.888%) F(0.075%) V(0.015%) S(0.004%) R(0.007%) I(0.007%) T(0.004%)	26780
W6	W(99.929%) G(0.030%) R(0.022%) *(0.007%) C(0.011%)	26836

The output is an .html table that displays: the reference amino acid or nucleotide and its position in the sequence, the found residue along with its percentage of occurrence colored according to the color code, and the total number of valid sequences for that position.

④ Enter the reference sequence without line breaks.

⑤ Enter a minimum value to filter mutations based on a specific frequency (e.g., 90.0). If left empty, it will display all mutations.

⑥ To search for specific mutations: enter the positions you want to search for separated by “,” without spaces (e.g., 63,69). To see the complete list of polymorphisms, leave the field empty.

⑦ Select ‘Mutated Positions’ to view only mutations or ‘All Positions’ to also see reference positions.

For **Mutations Table**:

	A	B	C	D	E	F	G
1	File	Number of Sequences	Reference	Position	Change	Percentage	Mutated Sequences
2	PR_01_AE.fasta	26838	P	1	P	99.90%	26810
3	PR_01_AE.fasta	26838	P	1	S	0.08%	21
4	PR_01_AE.fasta	26838	P	1	A	0.00%	1
5	PR_01_AE.fasta	26838	P	1	L	0.01%	2
6	PR_01_AE.fasta	26838	P	1	T	0.01%	2
7	PR_01_AE.fasta	26838	P	1	H	0.00%	1
8	PR_01_AE.fasta	26838	P	1	V	0.00%	1
9	PR_01_AE.fasta	26649	Q	2	Q	99.78%	26591
10	PR_01_AE.fasta	26649	Q	2	E	0.07%	19
11	PR_01_AE.fasta	26649	Q	2	S	0.02%	5
12	PR_01_AE.fasta	26649	Q	2	H	0.06%	15
13	PR_01_AE.fasta	26649	Q	2	D	0.00%	1
14	PR_01_AE.fasta	26649	Q	2	K	0.02%	6
15	PR_01_AE.fasta	26649	Q	2	L	0.02%	4
16	PR_01_AE.fasta	26649	Q	2	P	0.01%	2
17	PR_01_AE.fasta	26649	Q	2	R	0.01%	3
18	PR_01_AE.fasta	26649	Q	2	T	0.00%	1
19	PR_01_AE.fasta	26649	Q	2	*	0.01%	2
20	PR_01_AE.fasta	26831	V	3	I	99.86%	26793

The output is a .csv table that includes: the name of the input file, the total number of sequences for each file, the reference residue, the positions where mutations have been detected, the detected change, the frequency of occurrence of the detected change, and the total number of mutated sequences.

④ Enter the reference sequence without line breaks.

⑤ Enter a minimum value to filter mutations based on a specific frequency (e.g., 90.0). If left empty, it will display all mutations.

⑥ To search for specific mutations: enter the mutations you want to search for separated by “,” without spaces (e.g., M63I,H69K). To see the complete list of polymorphisms, leave the field empty.

⑦ Select ‘Mutated Positions’ to view only mutations or ‘All Positions’ to also see reference positions.

For **Markers**:

Variability Polymorphisms Individual Markers >= 90%		
File	Markers	Total Sequences
PR_107_01B.fasta	Q92K	4
PR_108_BC.fasta	T74S	15
PR_112_01B.fasta	L63M	5

The output is an .html table that includes: the name of the input file, the detected markers colored according to their frequency of occurrence, and the total number of analyzed sequences.

③ Choose the frequency of occurrence for the markers: >75% or ≥90%.

④ Enter the reference sequence without line breaks.

For **Multiple Mutations**:

	A	B	C	D
1	File	Number of Sequences	M46I/V32I	Frequency
2	PR_01_AE.fasta	26504	2	0.008
3	PR_02_AG.fasta	9418	0	0
4	PR_03_A6B.fasta	300	1	0.333
5	PR_04_cpx.fasta	15	0	0
6	PR_05_DF.fasta	24	0	0
7	PR_06_cpx.fasta	732	0	0
8	PR_07_BC.fasta	10819	0	0
9	PR_08_BC.fasta	2326	0	0
10	PR_09_cpx.fasta	91	0	0
11	PR_100_01C.fasta	5	0	0
12	PR_101_01B.fasta	4	0	0

The output is a .csv table that includes the name of the analyzed file, the number of valid sequences per file, the count of occurrences of combined mutations, and their frequency.

⑥ Enter the combined mutations you want to search for, separated by "," without spaces (e.g., D614G,A222V).

For **Mutations by Position**:

	A	B	C	D	E
1	File	Residues (12,15,17)	Number of Mutations	Frequency	Number of Sequences
2	PR_01_AE.fasta	AIG	289	1.134	25494
3	PR_01_AE.fasta	AVG	48	0.188	25494
4	PR_01_AE.fasta	HIG	1	0.004	25494
5	PR_01_AE.fasta	IIG	70	0.275	25494
6	PR_01_AE.fasta	ILG	1	0.004	25494
7	PR_01_AE.fasta	IVG	12	0.047	25494
8	PR_01_AE.fasta	KIG	13	0.051	25494

The output is a .csv table that includes the name of the input file, the detected combination of residues at the entered positions, the count of occurrences for each combination, the frequency of occurrence of the combination, and the number of valid sequences for those positions.

⑥ Enter the positions you want to analyze, separated by "," without spaces (e.g., 9,22,30).

⑨ Select the output folder and name the file according to the chosen Output Format: **Mutated Positions** and **Markers** are saved in an .html file. **Mutations Table, Multiple Mutations,** and **Mutations by Position** are saved in a .csv file.

⑩ Click Calculate.

More information about this function can be found in the User Manual, section:

II. VARIABILITY II.1 POLYMORPHISMS II.1.A) INDIVIDUAL