Database Setup

sonar setup	
--db *FILE*	specify the name of the new database
--gbk *FILE*	set a custom reference based on the specified Genbank file *[opt.]*
-a --auto-create	add pre-defined sample properties to the new database *[opt.]*

Show Infos

sonar info	
--db *FILE*	specify name of the database

Sample Management

sonar import	*to import data*
--db *FILE*	specify name of the database
--fasta *FILE [FILE, ...]*	specify fasta file to import sequences [opt]
--tsv *FILE [FILE, ...]*	specify tab-delimited text file to import meta information [opt]
--cols *STR [STR, ...]*	link properties with tsv columns in the form of *prop_name=col_header* [opt]
--no-autodetect	do not auto-link properties with tsv columns based on identical names [opt]
--no-update	skip data of existing samples [opt]
--cache *DIR*	specify a folder to cache import data [opt]
-t *INT* --threads *INT*	specify number of threads / cpus to use [opt]
sonar delete	*to delete sample(s)*

Sample Management (cont)

--db *FILE*	specify name of the database
--sample *STR [STR, ...]*	specify names of samples to delete [opt]
--sample-file *FILE*	specify file containing names of samples to delete [opt]
--force *FILE*	force deletion and skip user confirmation [opt]
sonar optimize	}*to clean database*
--db *FILE*	specify name of the database
sonar db-upgrade	upgrade database to the latest version
--db *FILE*	specify name of the database

Property Management

sonar list-prop	*to list available properties*
--db *FILE*	specify name of the database
sonar add-prop	add new property
--db *FILE*	specify name of the database
--name *STR*	specify new property name matching to [A-Z][A-Z0-9_]+
--descr *STR*	provide a short property description
--dtype *STR*	specify data type
--qtype *STR*	specify query type [opt]
--default	specify default value [opt]
sonar delete-prop	delete property
--db *FILE*	specify name of the database
--name *STR*	specify new property name
--force	force deletion without user confirmation [opt]

Query Genomes

sonar match	
--sample *STR [STR, ...]*	consider given samples only [opt]
--sample-file *FILE	consider samples listed in the given file only [opt]
--profile *STR [STR, ...]*	match given mutation profile (use argument multiple times to define alternate profiles) [opt]
--*PROP VALUE [VALUE, ...]*	filter based on sample property (*PROP* has to be replaced by the actual property name) [opt]
--with-sublineage *PROP*	include pangolin sub-lineages for the given property [opt] ℹ Use *sonar update-lineage-info* to update lineage info before
--showNX	show uninformative variant allels for genome profiles [opt]

--out-cols *STR [STR, ...]*	specify columns to show (for tsv and csv only) [opt]
--count	show number of matching genomes only [opt]
--format *{csv, tsv, vcf}*	specify the output format [opt]
-o *FILE* --out *FILE*	write output to the given file [opt]

Mutation Notations

	Genome level	Protein level
Substitution	*REFposALT* e.g. A3451T	*PROT*:*REFposALT* e.g. A3451T
Insertion	*REFposALT* e.g. A4057TGAT	*PROT*:*REFposALT* e.g. N:A34AK
1-bp Deletion	del:*pos* e.g. del:11288	*PROT*:del:*pos* e.g. ORF1ab:del:300
multi-bp Deletion	del:*start*-*end* e.g. del:112-115	*PROT*:del:*start*-*end* e.g. S:del:68-69
Deletion with fixed coordinates	del:=*start*-=*end* e.g. del:=8-=11	*PROT*:del:=*start*-=*end* e.g. S:del:=68-=69

ℹ Any notation can be negated by a preceding caret (^), e.g. ^A1345G only matches target genomes not carrying this SNP.

❗ *REF*: reference allele; *ALT*: variant allele; *pos*: reference position (1-based); *start* and *end*: start and end reference coordinates (1-based); *PROT*: protein symbol

Data Types for Properties

integer	for storing integers
float	for storing decimal numbers
text	for storing strings
dates	for storing dates in the form YYYY-MM-DD
zip	for storing zip codes

Operators for Data Types

Operator	Description	Valid data type
>	larger than (e.g. >1)	integer, float, date
<	smaller than (e.g. <1)	integer, float, date
>=	larger than or equal to (e.g. >=1)	integer, float, date
<=	smaller than or equal to (e.g. <=1)	integer, float, date
!=	different than (e.g. !=1)	integer, float, date
:	*from*:*to* range (e.g. 2021-01-01:2021-12-31)	integer, float, date
^	not the same as (e.g. ^B.1.1.7)	text
%	wildcard standing for any character(s) (e.g. %human%)	text

⚠ Make sure that expressions with > or < are quoted to prevent misinterpretation by the shell