phylogeny free mutational spectrum reconstruction

1. control file generation: one species - one sequence (RefSeq parsing):

Species	Gene	Seq	Code	CDS
Mus_musculus	tRNALeu	AAT....GGC	2	0
SARS-CoV-2	ORF1ab	ATA....CGY	1	1
...				

2. within-species polymorphisms: one species - many sequences (biopython blast api):

if (CDS == 1) {tblastn with "Organism" limitation; => codon alignment}

if (CDS == o) {blastn with "Organism" limitation; nucleotide alignment}

QUERY-ANCHORED ALIGNMENT
>Mus_musculus.A1
ATCGCACGCGGCCATCGGGG
>Mus_musculus.A2
ATCGCACGCGGCCATCGGGG

3. deviations from consensus - observed & expected (python):

OBSERVED:

Pos	From	To	MAF
chrM:567	atGcc	atTcc	0.1
...			

EXPECTED:

Pos	From	To	Effect
chrM:1	atGcc	atTcc	syn
chrM:1	atGcc	atAcc	syn
chrM:1	atGcc	atCcc	nons
...			

4. mutational spectrum generation (python, R):