**Output a database of alleles out of a VCF file**

*Script name:* do\_database.pl

What it does: This script takes an input VCF file (diploid, triploid accessions etc.. from one subspecies e.g. AA, AAA, AAAA…) and makes a flat file/database

***Synopsis: perl do\_database.pl -i VCF\_FILENAME\_IN -o OUT\_DATABASE -r RNA\_acc\_names***

*Parameters:*

-i (string): VCF file

-o (string): Database name

-r (string): accession names (present in the VCF) coming from RNA sequencing (one per line)

*All the steps in detail*

Output is a tab delimited file like this:

chr11 34387 G A MISSING:0.12 G/G:0.69:RNA1 A/A:0.03 G/A:0.17

chr11 34397 T G MISSING:0.11 T/T:0.87 G/G:0.01 T/G:0.01

chr11 34414 T A MISSING:0.21 T/T:0.79 A/A:0.01

chr11 34475 A AC MISSING:0.09 A/A:0.86 AC/AC:0.03 A/AC:0.02

chr11 34496 A G MISSING:0.10 A/A:0.85 G/G:0.03 A/G:0.02

In each line, you’ll fine chr, position, alleles, genotypes and their proportions. Also, the number of time a genotype was found in RNA accession. RNA1 means 1 accession from RNAseq had the genotype G/G.