**From VCF to Allele count (for RAD or RNA)**

*Script name:* from\_vcf\_to\_allele\_count.pl

What it does: takes a VCF, predict SNP effect with snpeff, remove SNP with missing data, remove A variants (based on autopolyploids) and print allele counts.

***perl from\_vcf\_to\_allele\_count.pl -r musa.fasta -x vcf –n 3 –m 10***

/!\ The jobarray was done with SGE on a cluster

/!\ VCF files have to be present in the current directory

/!\ Change the parameters –t (5th line of the script) to the number of accessions you have to treat in the current directory

*Parameters:*

-r (string): reference fasta filename

-x (string): file extension (fastq)

-n (int): number of individuals

-m (int): min read depth

*All the steps in detail*

* Annotate SNPs with their effect on genes

<http://snpeff.sourceforge.net/>

* Remove SNPs with missing data with SelectVariants from GATK <https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-0/org_broadinstitute_gatk_tools_walkers_variantutils_SelectVariants.php>
* Remove potential A variants based on autoalloploids individuals and the reference fasta file with SnpSift

<http://snpeff.sourceforge.net/SnpSift.html>

* Output allele count for each accession in a tab delimited file with a custom perl script (output\_allele\_count\_from\_vcf.pl)

Remarks: All the fixed parameters can be changed directly in the script.