

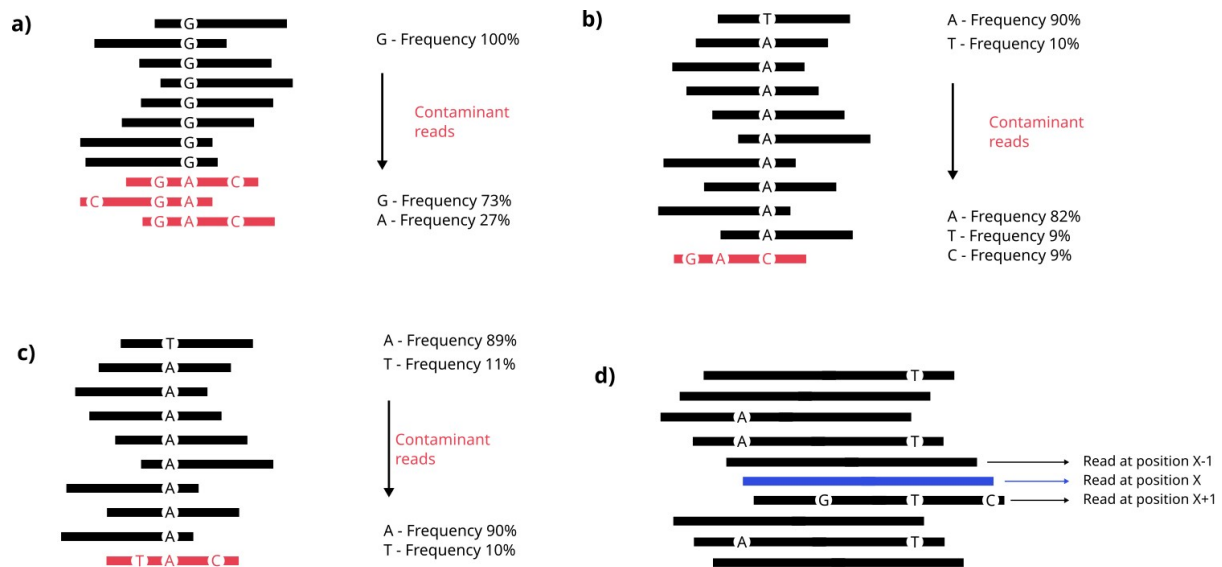


Figure S1. Effects of contaminating reads and taxonomic filtering in variant calling. In this figure we exemplified possible effects of contaminant reads (in red) in variant calling (**a-c**) and the inability of Kraken classifying some reads (**d**). **a)** Contaminant reads make G to be called as vSNP instead of fSNP and introduce a false positive vSNP (A) **b)** Contaminant reads make A to be called as vSNP instead of a fSNP and make one vSNP (T) to be lost due to SNP calling cutoffs (10% frequency for vSNPs). **c)** Contaminant reads make A to be called as fSNP instead of a vSNP. **d)** The read coloured in blue cannot be classified by Kraken up to the level of genus. This occurs despite the fact that this read is 100% identical to the reference and that surrounding reads, with several SNPs, can be classified even to the level of species.