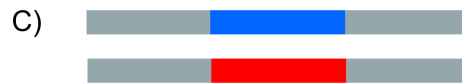
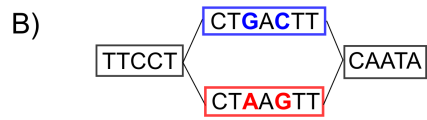
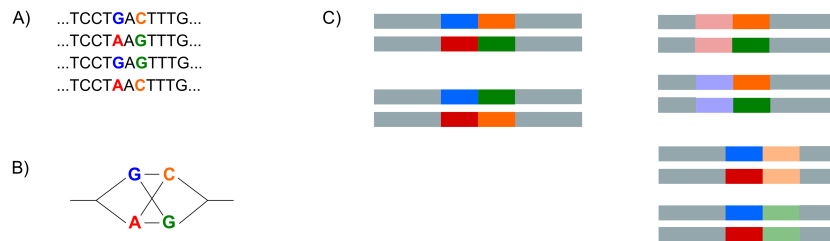


Supplementary files:
De novo identification, differential analysis and
functional annotation of SNPs from RNA-seq
data in non-model species

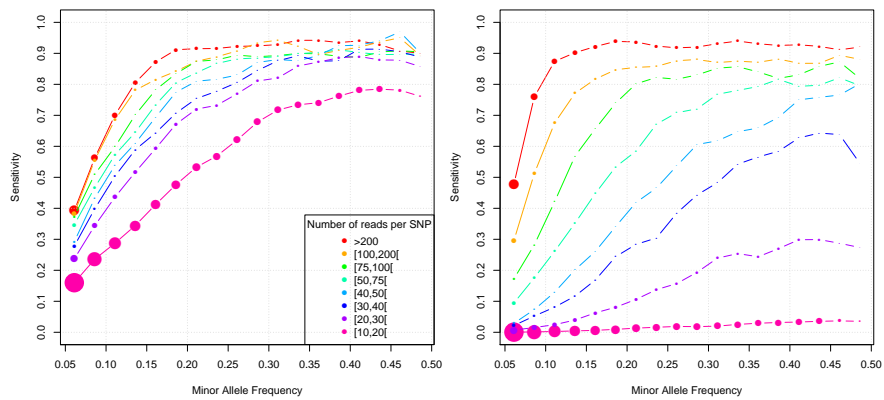
A) ...TTCCT**GAC**TTTGA...
...TTCCT**AAG**TTTGA...



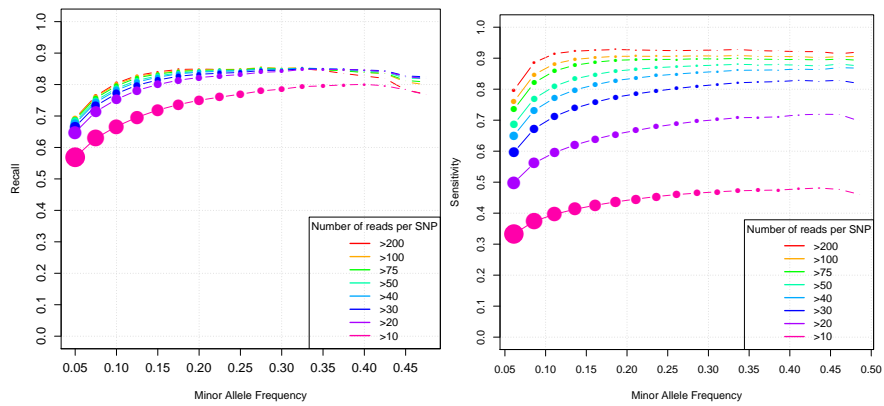
Supplementary Figure 1: Two SNPs separated by less than k nucleotides will be reported in the same bubble. If the SNPs are linked, only one bubble is reported.



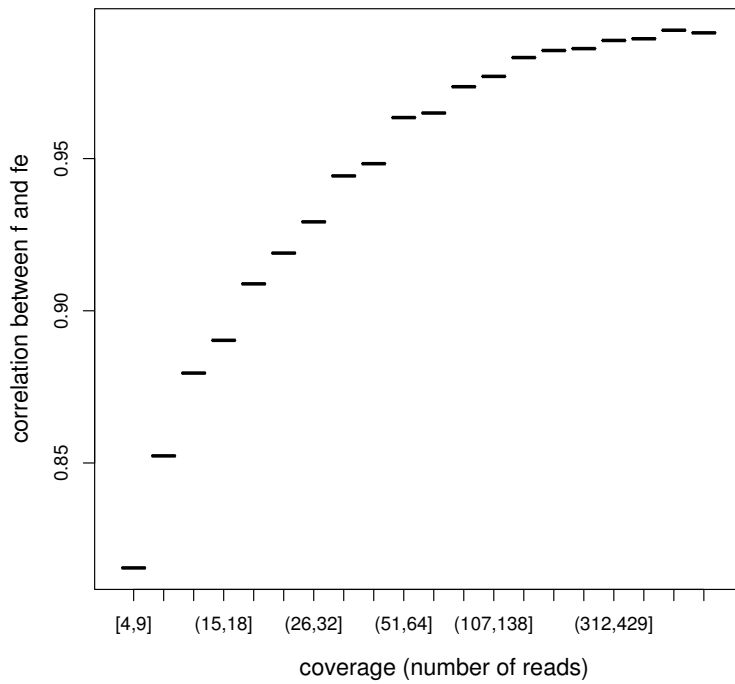
Supplementary Figure 2: Two SNPs separated by less than k nucleotides, but with no linkage, can correspond to 4 haplotypes. They will generate 6 bubbles.

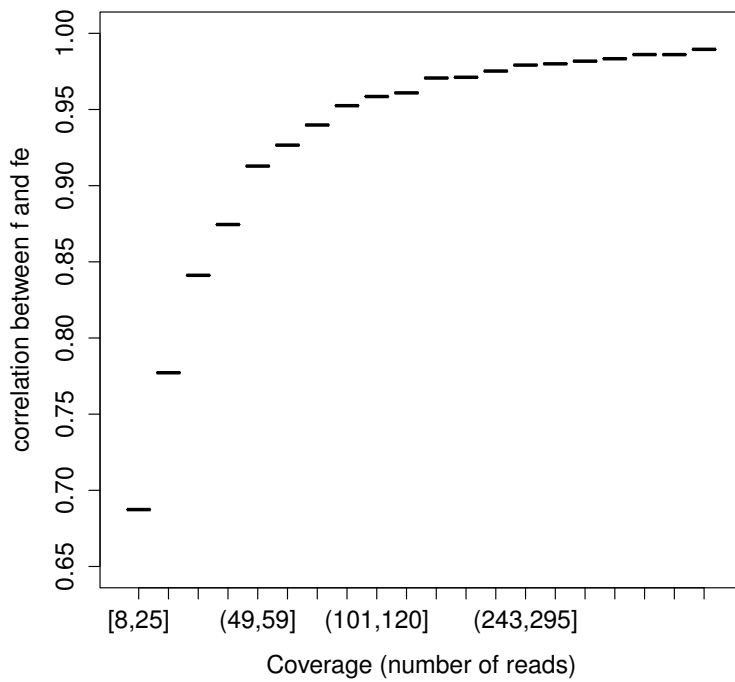


Supplementary Figure 3: GATK(left) and KISSPLICE (right) recall as a function of gene expression and Minor Allele Frequency. Counts are split by categories.

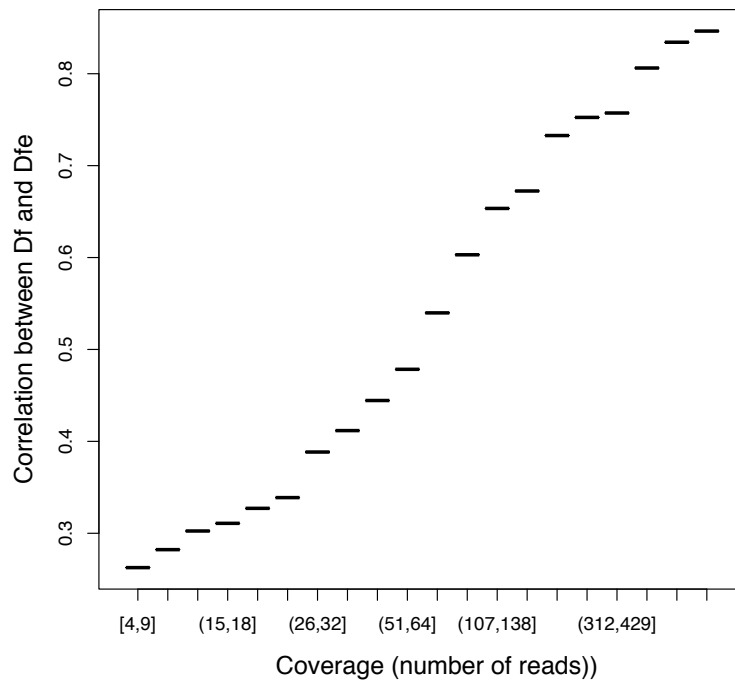


Supplementary Figure 4: GATK(left), and KISSPLICE (right) recall as a function of gene expression and Minor Allele Frequency

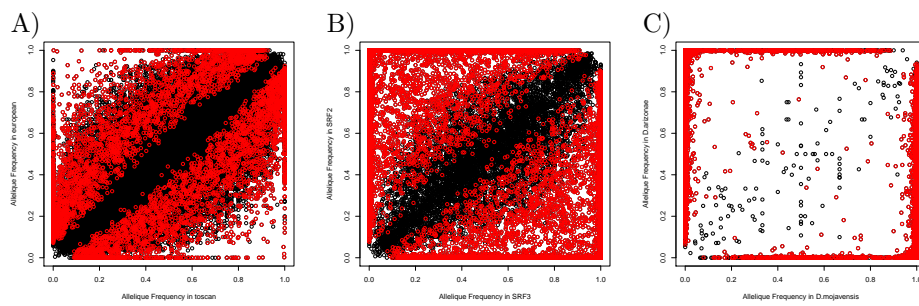




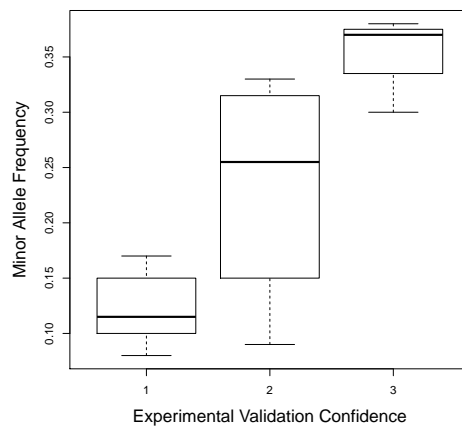
Supplementary Figure 6: Correlation between the allele frequency estimated using RNA-seq data and the true allele frequency as a function of locus expression.



Supplementary Figure 7: Correlation between the difference of allele frequency estimated using RNA-seq data and the true difference of allele frequency as a function of locus expression.



Supplementary Figure 8: Expressed allele frequencies of one lineage/population Vs expressed allele frequency of the other. Red dots are conditions specific SNPs. Black dots are SNPs whose allele frequency is not different across populations. A) Human TSI Vs CEU B) *Asobara tabida* SFR2 Vs SFR3 C) *Drosophila mojavensis* Vs *arizonae*



Supplementary Figure 9: Confidence in the experimental validation depends on the minor allele frequency. A scale ranging from 1 to 3 indicates the confidence degree of the experimental validation process; a number of 3 corresponding to the highest confidence.