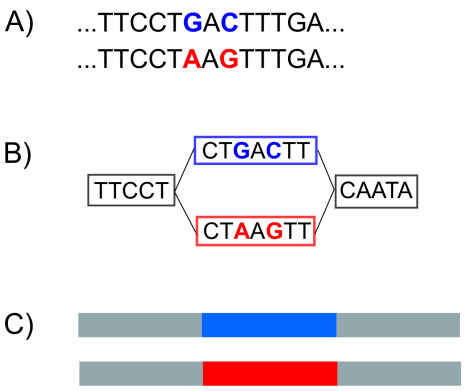
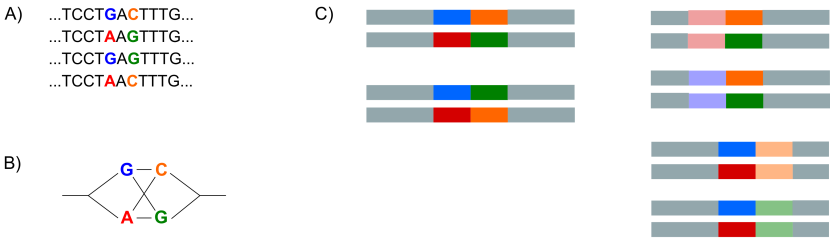
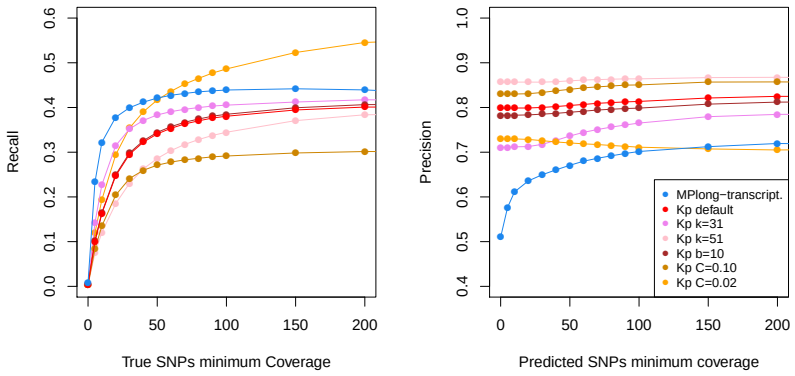
Supplementary Figures File



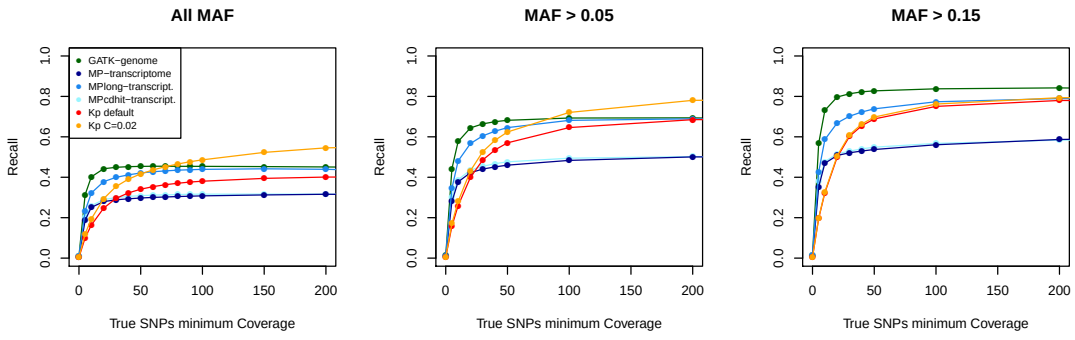
Supplementary Figure 1: Two SNPs separated by less than k nucleotides will be reported in the same bubble. If the SNPs are linked, only one bubble is reported.



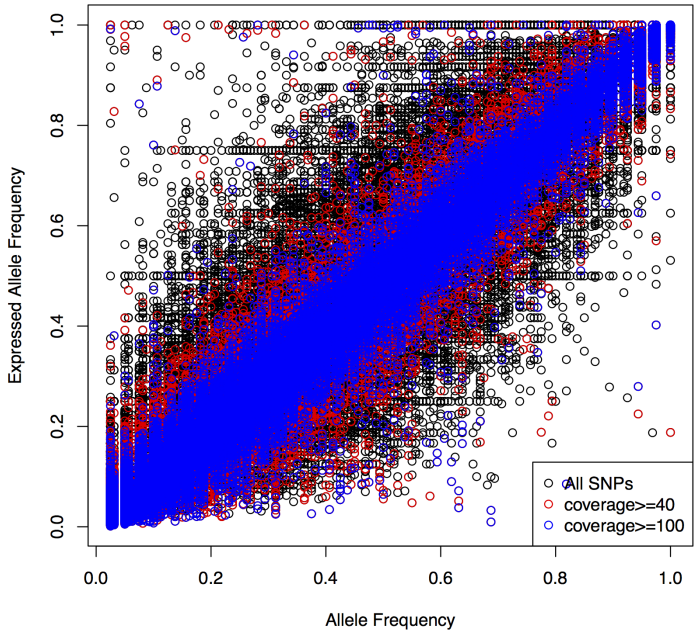
Supplementary Figure 2: Two SNPs separated by less than k nucleotides, but with no linkage, can correspond to 4 haplotypes. They will generate 6 bubbles.



Supplementary Figure 3: Influence of the parameters k, C, and b on the recall and precision of KisSplice. k is the kmer size. C is the relative coverage cutoff. b is the maximum number of branches allowed in a bubble. The default values are k=41, C=0.05 and b=5. Increasing k, increasing C or decreasing b results in a better precision but a worse recall. We also indicate the recall and precision of mp-long. The best recall is reached for C=0.02. The best precision is reached for k=51.

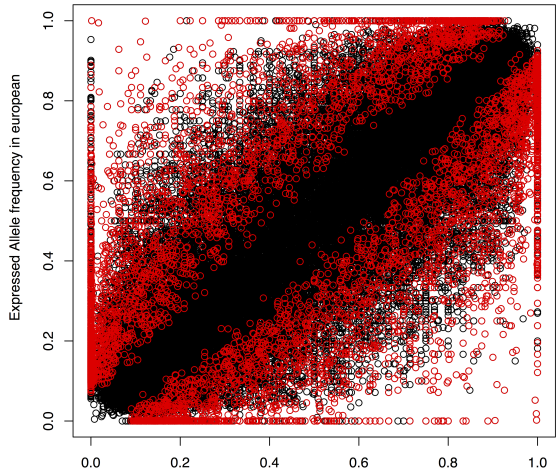


Supplementary Figure 4: Impact of minor allele frequency cut-off on the recall of all methods. The larger the MAF, the easier it is to detect the SNP.

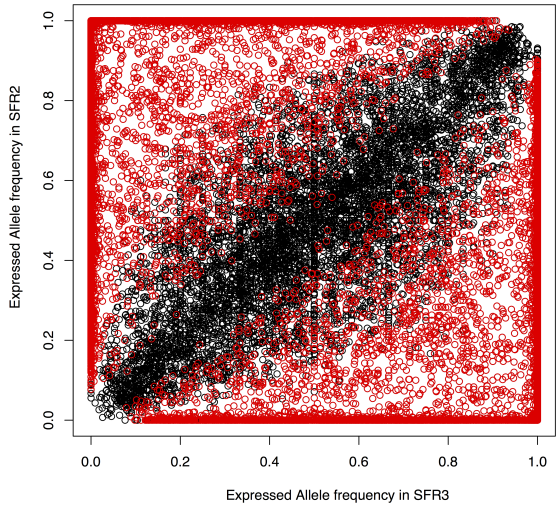


Supplementary Figure 5: Allele frequency estimated using RNA-seq data Vs the true allele frequency. The higher the expression, the higher the correlation.

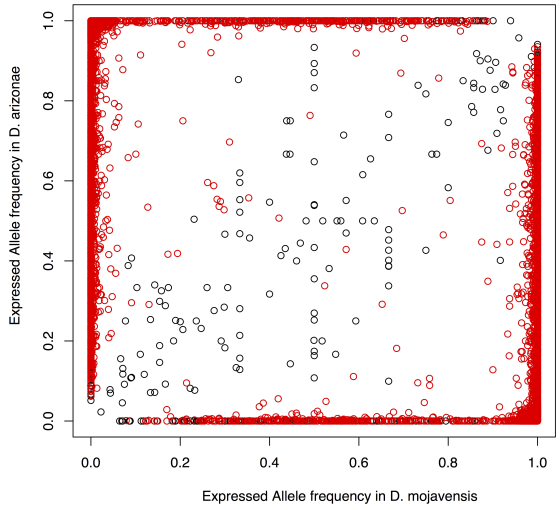
A)



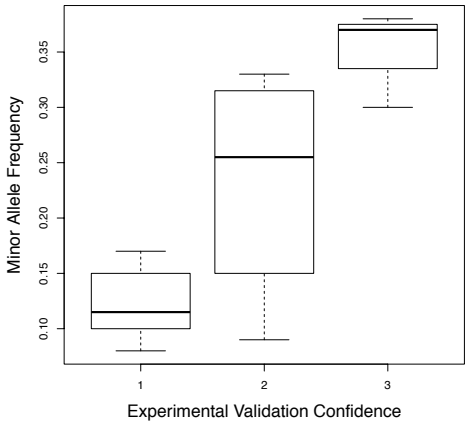
B)



C)



Supplementary Figure 6: 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 *Drosophila arizonae*



Supplementary Figure 7: 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.

Supplementary Table 1: List of SNPs predicted by KisSplice in *Asobara tabida* SFR2 and SFR3 lines. The 27 first cases were chosen for experimental validation because they covered a wide range of MAF and they fell in genes whose function was related to the constrasted phenotypes. The last 7 cases were chosen because they were found by KisSplice only.

Supplementary Table 2: List of divergent sites, SNPs and inexact repeats predicted by KisSplice in *Drosophila mojavensis* and *Drosophila arizonae*. They were chosen for experimental validation because they covered a wide range of MAF and were located in sufficiently expressed loci (at least 100 reads).