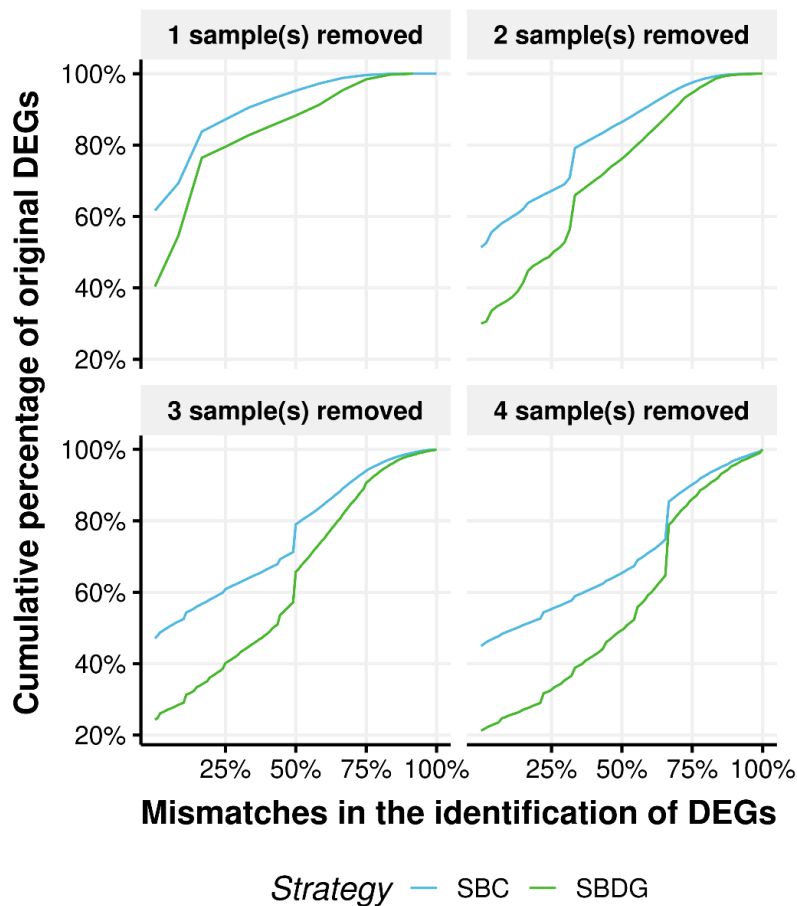


**Supplementary Material to “Sampling strategies for sugarcane using either clonal replicates or diverse genotypes can bias the conclusions of RNA-seq studies”**



**Figure S2** - Cumulative distribution of identified mismatches. The total number of combinations of subsets of samples are:  $n_1 = 12$ ,  $n_2 = 54$ ,  $n_3 = 108$ , and  $n_4 = 81$ , in which the index  $i$  represents the number of removed samples. Thus, the original DEGs and the subsamples may match from 0 to  $n_i$  times. We consider a match when a given iteration has the same differential expression result for a gene. Hence, the curves indicate the percentage of DEGs which showed the minimum mismatch rate in the x-axis. For example, with a single sample removed for the SBDG (top left panel), 40% of genes did not have a single mismatch for all twelve combinations, and 80% of genes presented a maximum mismatch rate of 25%.