

# 1 Are EnTrI results biased?

Since transposon insertion biases can affect the essentiality level inferred from transposon mutagenesis experiments, the dataset has been tested for two types of biases: the distance from the origin and GC content. As Figure 1 shows, although the regression line is flat, the insertion index decreases while increasing the distance from the origin. Therefore, there is a bias towards the distance from the origin. However, the threshold that has been selected for essential genes (0.2) separates the two groups well.

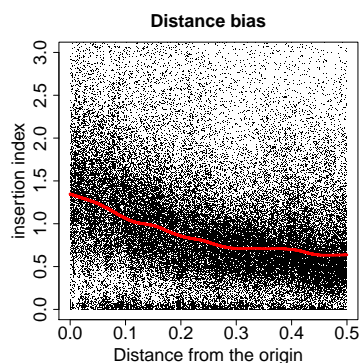


Figure 1: The insertion index value depends on the distance from the origin

As depicted in Figure 2, there is no bias towards the GC content of the genes and the regression line is flat. The nucleotides around the insertion sites (the insertion site and 10 nucleotides on each side) are stacked on top of each other and a sequence logo is generated from these sequences to see if there is any positional bias for any nucleotide. It can be inferred from Figure 3 that there is no bias in any position except a negligible bias in the insertion site towards GC pairs.

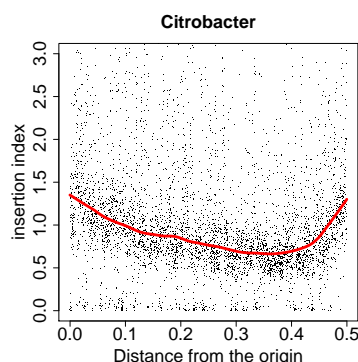


Figure 2: The insertion index value is not biased to GC content

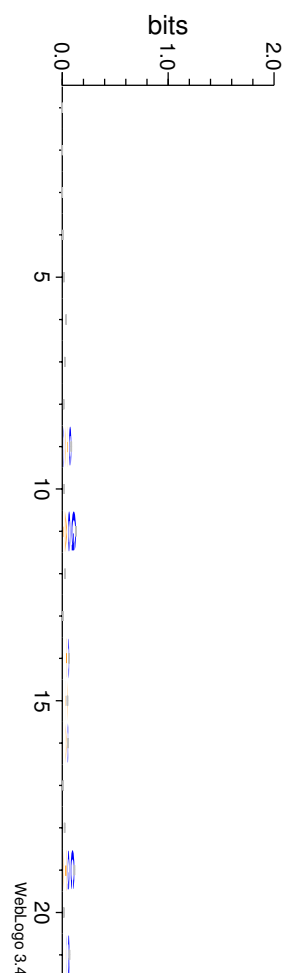


Figure 3: There is no bias in the nucleotides around the insertion site