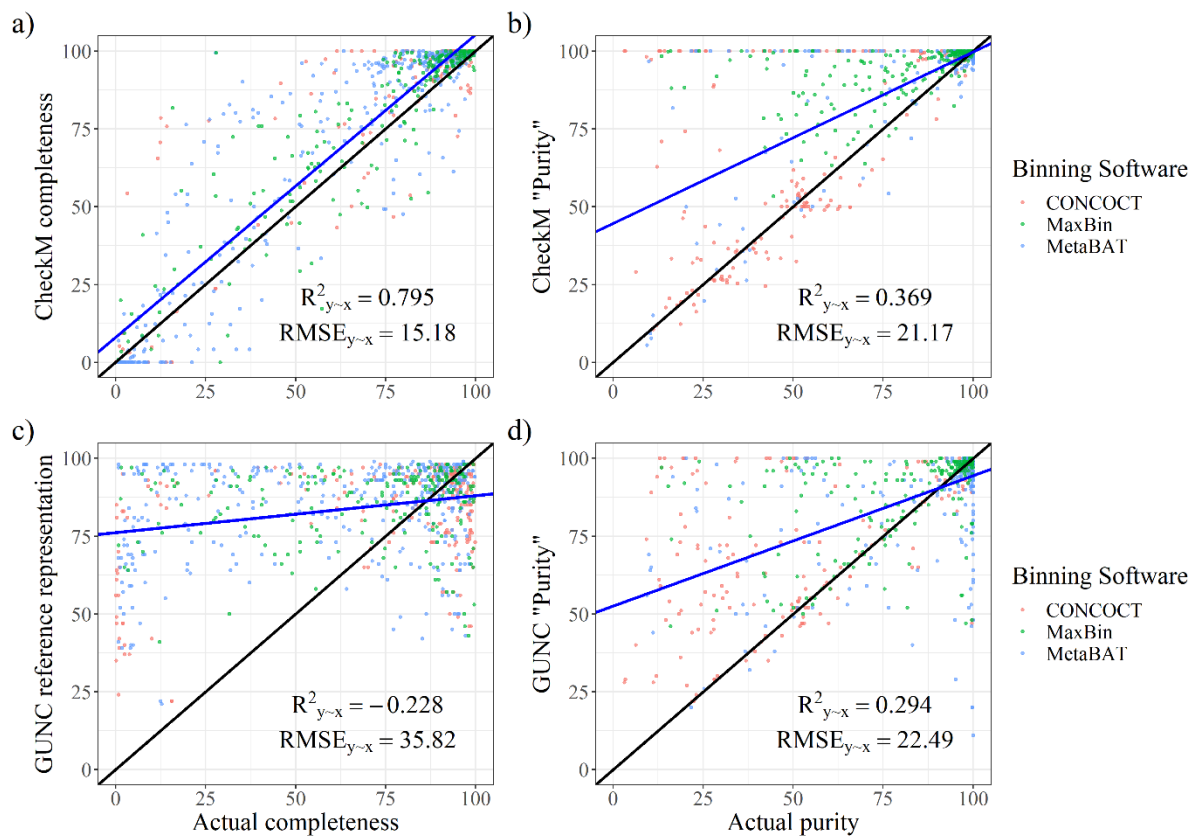




Closest analogues for completeness and purity obtained from the output of CheckM (a,b) and GUNC (c,d) as a function of the actual values in the training dataset bins. Data points are coloured according to the binner used. The blue line is the best linear fit.