**Figure 9 – Saturation of short-read datasets.**

Short-read datasets used by Bemm *et al* for k-mer selection were assessed for saturation using bbmap’s bbcountunique.sh script (Version 35.82) with default settings. Raw saturation data was input into Prism (V 6.0g) and fit for visualization using nonlinear regression. The graph shows the percent of non-redundant reads sampled (Y-axis) as a function of total reads sampled (X-axis). Even in the best cases (500bp insert set 2 and 800bp set 1) the number of non-redundant reads samples is above 75%. A fully (or nearly) saturated dataset should approach 0% non-redundant reads.