

Rapid Ligation

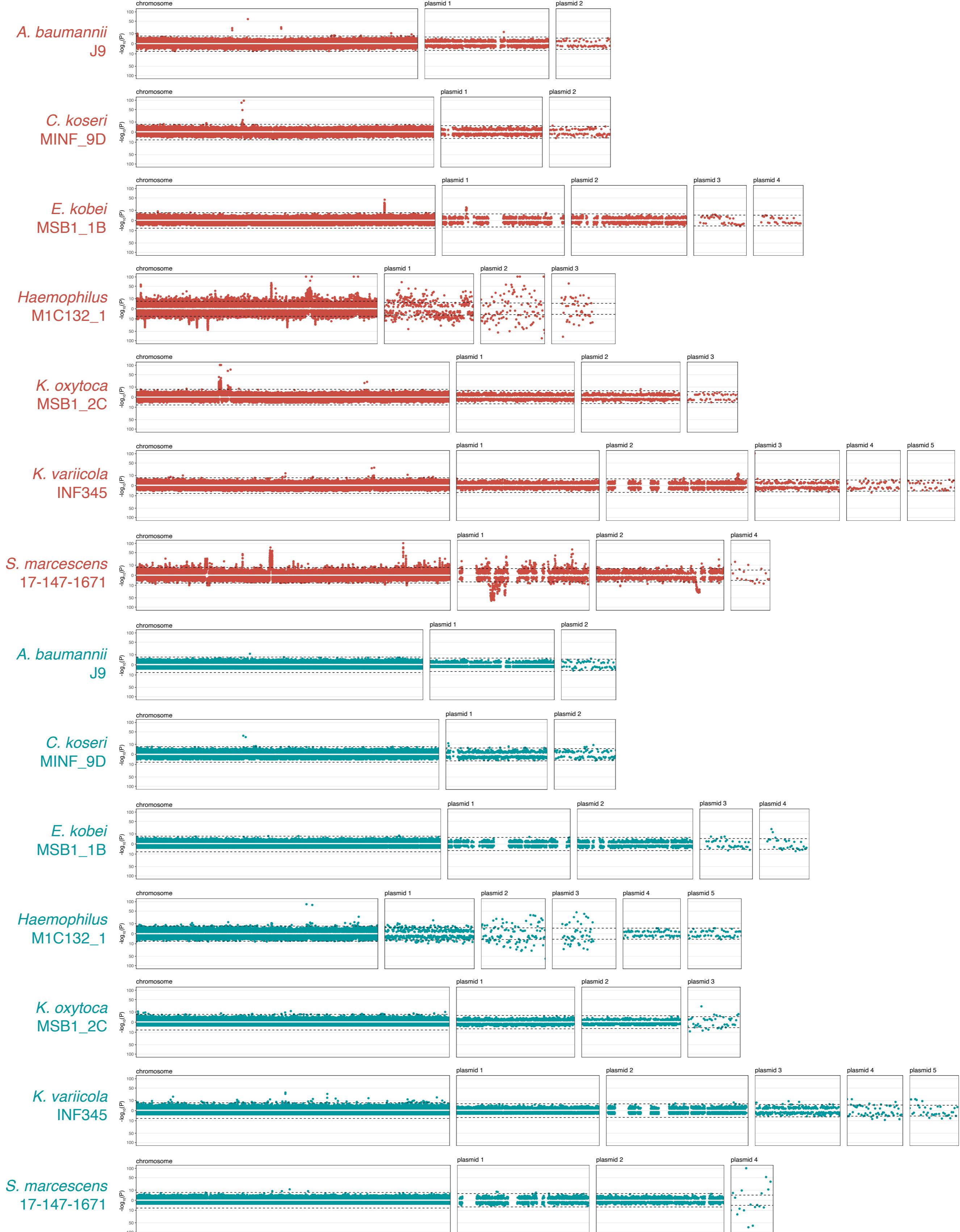


Figure S4: two-sided Manhattan plots for read-start density in each of the replicons from both ONT library preparation types. Each point represents a 100-bp window in which the number of read-start events was quantified. A p -value was determined for each read-start count by comparing against a Poisson distribution with a mean equal to the replicon's mean read-starts per window: for a window with k read-starts and a Poisson with mean λ , if $k > \lambda$ then $p = P(X \geq k)$ and if $k < \lambda$ then $p = P(X \leq k)$. Points above the zero-line indicate more read-starts than expected from the Poisson distribution, and points below the zero-line indicate fewer read-starts than expected from the Poisson distribution. Dashed lines indicate Bonferroni-corrected $p=0.05$ levels, so points outside the dashed lines are statistically significant deviations from a Poisson distribution.

Ligation read sets showed more localised deviations from the Poisson distribution than rapid read sets, suggesting some parts of the sequence are more prone to breakage than others. This was particularly evident for the *Haemophilus M1C132_1* and *S. marcescens 17-147-1671* genomes.

Repetitive sequence regions were excluded from this analysis because they contained unreliable read alignments. *E. kobei* plasmid 5 and *S. marcescens* plasmid 3 were excluded because they could contain multiple variants with different sequence lengths.