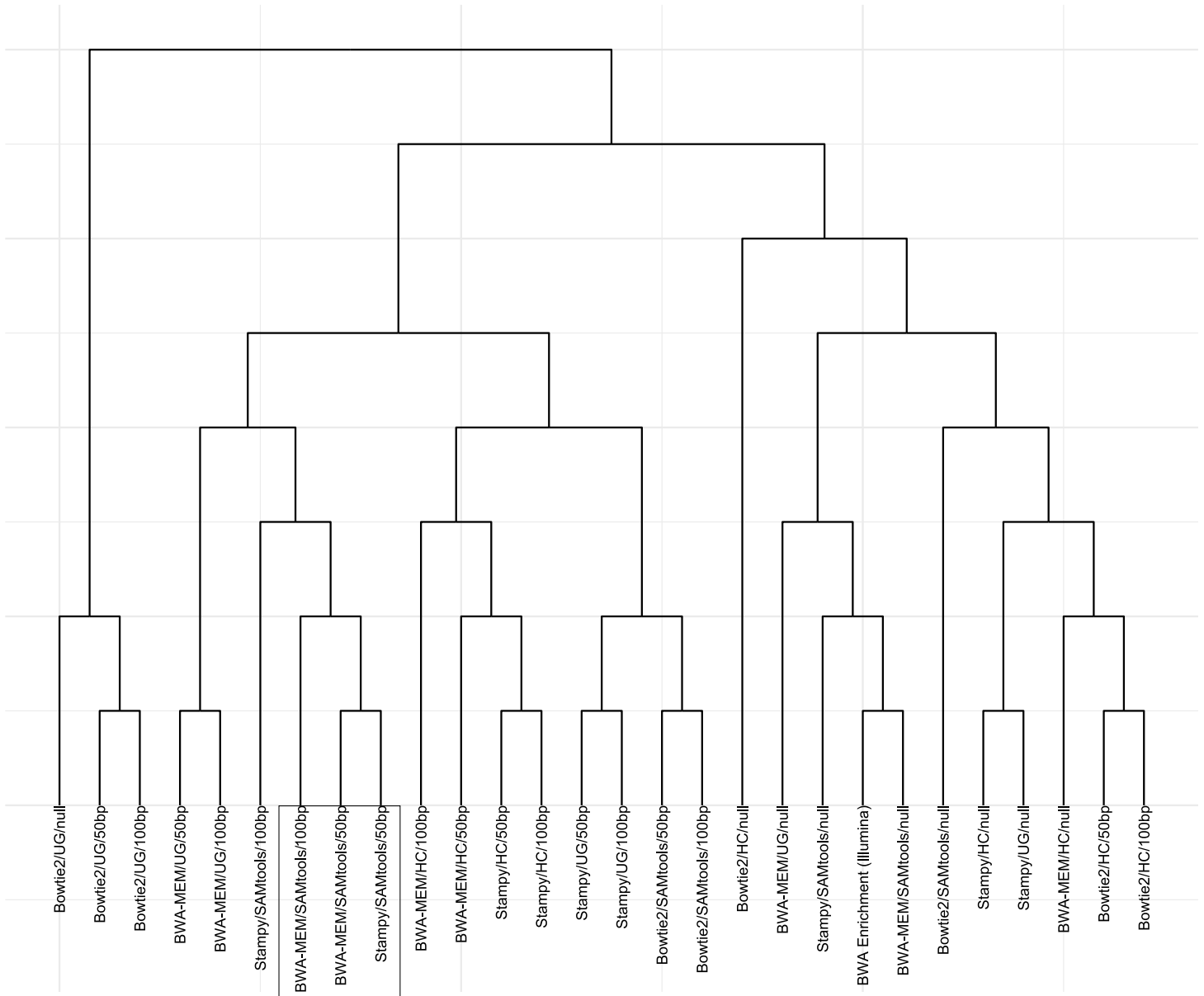


Supplementary Figure 2: Hierarchical clustering of the tools

The d, MCC, p, r and F1 values were used to perform hierarchical clustering analysis based on the Lance–Williams agglomerative hierarchical clustering algorithm, which at each stage recomputes dissimilarities between clusters.



HC, GATK HaplotypeCaller; UG, GATK UnifiedGenotyper; null, zero interval padding; 50bp, 50bp interval padding; 100bp, 100bp interval padding. Denoted with box, top performing pipelines based on the perpendicular distance from the random-guess line.