



Additional Figure A2 – Visual representation of M-GCAT’s multiple alignment of four *Haemophilus influenzae* genomes. hainA: *H. influenzae* strain 86-028NP (CP000057.1); **hainB:** *H. influenzae* Rd KW20 (L42023.1); **hainC:** *H. influenzae* PittEE (CP000671.1); **hainD:** *H. influenzae* PittGG (CP000672.1). There were 53 M-GCAT Clusters of average length 32,060 nt. In total, 91.3% of the original genome sequences are covered by the multiple alignment, and 8.7% was left unaligned.