Virulence genes detection

Briefly, virulence genes were detected by blasting reference sequences against a NCBI Blast+ protein database containing CDS from genomes of all strains included in this manuscript. Preparation of the strains database and parsing of the blastp outputs has been done with BioPython and is documented in a Jupyter Notebook available on gitHub (<https://github.com/soda460/finding_virulence_genes>). To detect enterohaemolysin (ehxA) and porcine attaching and effacing-associated protein (paa), the NCBI Reference sequence WP\_011310119 and GenBank sequence U82533.4 were used respectively. Similarly, CDS from bfp genes encoded on Escherichia coli B171 plasmid pB171 described under the Genbank Accession number AB024946 ( Abraham et al. Year) were used as distinct queries to screen for the bfp genes. Finally, the 143 complete eae sequences referenced in Xiong et al. 2016 were clustered with CD-HIT in 11 divergent sequences showing no more than 90% of similarity and the latter sequences were used as queries to detect intimin genes.