H. pylori was the first bacterium identified as a class I carcinogen in 1994. H. pylori infection is associated with stomach ulcers and stomach cancer. Stomach cancer is the fourth most common cancer and second leading cause of cancer-related deaths worldwide. The incidence of stomach cancer varies substantially around the world with high levels in parts of East Asia and Latin America and this can be partially explained by the presence of the H. pylori virulence factor, cagA, which is under positive selection. Given that pathogenic factors are often under positive selection and correlated with a bacterial phenotype that produces specific manifestations of disease, we wished to identify novel pathogenic factors using this evolutionary perspective.

We chose to analyze genomes from East Asia and Europe as these regions have high and low incidences of H. pylori associated pathogenicity, respectively. A genome wide Ka/Ks scan using the PAML package was conducted to detect genes that are under positive selection, doing pairwise comparisons between H. pylori strains from Europe (G27/Italy vs B38/France, 26695/UK vs HPAG1/Sweden) and Asia (51/Korea vs 52/Korea and 30/Japan vs 32/Japan). The analysis revealed a larger number of genes under positive selection in Asian genomes (Korea: 50 and Japan: 27, Northwest Europe: 9 and Southern Europe: 12). For Korean strains, several genes were identified as candidate virulence factors. These included outer membrane proteins (Omp3 and HopK), Lipoproteins (accession number: gb|ACX98011.1), carbonic anhydrase, labile enterotoxin outputA and hypothetical proteins that could be considered as secreted proteins. These genes are potentially related with a high incidence of H. pylori associated pathogenicity present in East Asia.