Final Project Report

Name: Kate Howell

Title: Using the pan-genome of *Haemophilus parasuis* (HPS) to design new molecular diagnostics for serotyping and pathotyping

Aims of the project:

1. Investigation of genetic determinants of serotype
2. Identification of putative virulence factors comparing whole genome sequence data of disease-associated and non- associated strains.
3. Development of molecular tests, e.g. PCR, that rapidly detect HPS in field samples, identify the serotype and virulence type of the isolate, and indicate the likely importance of the isolate for the disease
4. Validation of the newly developed PCR tests using field samples.

Haemophilus parasuis is a Gram-negative bacterium that inhabits the upper respiratory tract of the pig. It is responsible for the systemic disease of pigs known as Glässer's disease and can cause or contribute to pneumonia. The current diagnostic or subtyping methods for this bacterium have shown limited correlation between the genetics of the organism and virulence. The current serotyping scheme is a serological assay that can identify fifteen serovars but shows only a partial association with virulence. Serotyping is important because currently available vaccines are serovar-specific, but only 80% of isolates can be typed by this method. The collection of *H. parasuis* isolates used in my work originates from twelve countries, collected over a sixty year period with detailed clinical meta-data. Based on a subset of UK isolates were separated isolates into clinical, which were identified from the tissues of pigs confirmed as having *H. parasuis* related disease*,* and non-clinical categories, which were taken from the upper respiratory tract of pigs with no evidence of *H. parasuis* related disease elsewhere in the animal*.* The clinical isolates were further separated into respiratory isolates or systemic isolates based on the tissue of origin and post-mortem symptoms.

The bacterial genome can be very variable within the same species, as recombination of genomic DNA, plasmids, phage and transposons all affect the genomic composition of a strain. Therefore the bacterial genome can be separated

into genes that are found in every single strain, which is considered to be the core genome; and those, that vary between strains, which are called the accessory genome. The pan-genome of the *H. parasuis* isolate collection was analysed using OrthoMCL based on 209 draft genomes and three published sequences. From this analysis, the pan-genome was found to contain over 7,000 genes, while the genome from a single isolate contains 2,000 genes. The core genome makes up approximately half of the genome and was made up of 1,049 genes found in every isolate, and within this core genome we identified 65,137 single nucleotide polymorphisms. The remainder of the genome varies greatly between isolates and so the isolate collection is very diverse and termed the accessory genome. The accessory genome was made up of 6,382 genes that varied between isolates. Important features of the genomes were also analysed including the presence of phage, putative virulence factors, antibiotic resistance genes, the capsule loci as well as the G+C content of the genomes. Detailed recombination analyses were performed using two methods (BratNextGen and gubbins), which identified very high rates of recombination, higher than that reported for any bacterial species in the literature to date. Furthermore, a statistical method called discriminant analysis of principal components (DAPC) was used as a genome wide association method to identify genes that were associated with the clinical meta-data for the isolates.

From the bioinformatic and statistical analyses it was possible to identify genes that were associated with two important phenotypes for this bacterium, serovar and virulence. Study of the capsule loci between known serovars identified that the majority of isolates (87%) had the same capsule loci as the reference strain for that serovar. The remainder of the isolates had the capsule locus of one of the serovars identified in the cross-reactions of their serotyping result, or had been designated non-typeable by traditional serotyping. Therefore it was possible to identify potential serovar-specific genes and these were used these to develop and validate a molecular serotyping PCR for fourteen of the fifteen serovars. No products were amplified from the closely related commensals or other pig pathogens when the multiplex PCR was tested on a negative control panel. The limit of detection of the multiplex PCR was 1 ng/ul and so HPS can be detected at low quantities in a sample. Both of these multiplex PCRs were validated on a subset of the isolates that had been sequenced as well as a new collection of isolates collected from the UK during

2012-2014, most of which had been previously serotyped. Overall the results of this multiplex PCR matched to the *in silico* predictions and had 100% accuracy. This multiplex PCR is faster, cheaper and more reliable than the current serotyping method, which has a non-typeable rate of 20%.

From the DAPC analysis and proportion of genes found in virulence categories, we also identified potential virulence markers and developed a general linear model for the prediction of virulence of isolates based on a subset of ten genes. This was used to develop a second multiplex PCR that can predict virulence of an isolate, which is both sensitive (75%) and specific (93%) and was termed a “pathotyping” multiplex PCR. This assay was also tested on the sequenced isolates with predicted band patterns, and the additional new clinical set from the UK and was 90% accurate. The limit of detection for this multiplex PCR was 0.1 ng/ul for the pathotyping assay.

Finally, these multiplex PCRs were also shown to work when tested on genomic DNA extractions of post-mortem diagnostic material (joint, lung and liver), which were collected from pigs during the same period in the UK. Both multiplex PCRs were able to detect *H. parasuis* from known cases without culture and purification of isolates. Overall I feel that my PhD has covered and investigated the original aims of the project and the diagnostic PCRs are being considered for further development.