Comparative microbial genomics in *Pasteurellaceae* family using second generation sequencing

Isabelle da Piedade\(^1\), Timothy J Johnson\(^2\), Anders M Bojesen\(^1\)*

\(^1\) University of Copenhagen, Department for Veterinary Disease Biology, Faculty of Life Sciences, Copenhagen, Denmark
\(^2\) University of Minnesota, Department of Veterinary and Biomedical Sciences, Saint Paul, MN, USA
* corresponding author

**Background**

The development of next generating sequencing technologies has propelled the field of comparative microbial genomics such forward that is now providing an exciting cost-effective way to study related bacterial genomes with unprecedented resolution. *Pasteurellaceae* is an increasing and diverse family of Gram-negative bacteria that include primary pathogens, such as *Haemophilus influenza*, *Pasteurella multocida* and *Actinobacillus pleuropneumoniae*, as well as several potential pathogens. Members of the *Pasteurellaceae* family represent natural models to study pathogenesis and host-pathogen-interactions, as the different species have been tightly associated humans or other vertebrates over a considerable time span. Interestingly, mechanisms of virulence of most *Pasteurellaceae* are still relatively unknown or poorly understood.

**Materials and Methods**

We have performed *de novo* sequencing of several *Gallibacterium* genomes, a recently defined genus within the *Pasteurellaceae* family, with the aim of understanding the evolutionary history of the genus better and hereby get a deeper insight into mechanisms of pathogenicity. Initially, three *Gallibacterium anatis* genomes (strains UMN179 isolated from a Midwestern US commercial laying hen with peritonitis; strain12656-12, a septicemia-associated isolate from a chicken in Denmark and strain F149 [1-2], a non-lesion-associated isolate from a healthy duck in Denmark [3]) have been sequenced by pyrosequencing using Roche GS FLX machines (454 sequencing). Other *Gallibacterium* genomes were sequenced using the Illumina Genome Analyzer II system to 30X depth coverage using paired-ends method.

**Results**

Our preliminary results show that the availability of genomes for all major branches of *Pasteurellaceae* enables a more definitive analysis of their evolutionary relationships. The *Gallibacterium* sequenced genomes were each assembled into approximately 100-120 contiguous sequences, with 25-40X coverage. Using the paired-end reads, the UMN179 contigs were assembled into 6 scaffolds, ranging in size from 60 kb to 1.5 Mb. Using this scaffolding strategy, three *Gallibacterium anatis* genomes were aligned to one another based upon the largest scaffold sequence. In this manner, genomic islands (DNA regions in addition to backbone sequence) were identified that were strain-specific, pathogen-specific or shared by 2/3 genomes. Macrosynteny was observed for all three genomes, with very few genetic rearrangements observed between them. Nearly 64% of
the Gallibacterium anatis scaffold sequence examined here can be considered backbone sequence common to all of the strains studied.

**Conclusion**

In this study, we present an extensive comparative analysis of different clades of the Pasteurellaceae family. For example, the macrosynteny of sequenced Gallibacterium anatis and its apparent divergence from closely related genomes support its separation as a distinct genus with associated species. The pathogenic lifestyle for some members of the Gallibacterium genus may reflect acquisition of foreign genetic elements associated with antimicrobial resistance or virulence factors possibly from primary pathogens sharing the same environmental niches.

**References**