Francisella tularensis: persistence, dissemination and source attribution

A theoretical and computational approach

Chinmay Dwibedi
The bacterium *Francisella tularensis* causing tularemia in humans and other mammals displays little genetic diversity among genomes across temporal and spatial scales. *F. tularensis* infects humans with an extremely low infectious dose and causes natural seasonal tularemia outbreaks. During the Cold War, this bacterium was developed as a biological weapon.

In paper I, we aimed at investigating the genetic diversity of *F. tularensis* over space and time and were especially interested in the influence of spatial dispersal on the genetic diversity. By analyses of single-nucleotide polymorphisms (SNPs) among 205 *F. tularensis* genomes, we found that tularemia had moved from East to West over the European continent by dispersal patterns characterized by multiple long-range dispersal events. Evolutionary rate estimates based on the year of bacterial isolation from 1947 to 2012 indicated non-measurable rates. In outbreak areas with multiple recent outbreaks, however, there was a measurable rate of 0.4 SNPs/genome/year indicating that in areas with more intense disease activity, there is a detectable evolutionary rate. The findings suggest that long-range geographical dispersal events and mostly very low evolutionary rates are important factors contributing to a very low genetic diversity of *F. tularensis* populations.

In paper II, we focused on a geographically restricted area with a history of frequent tularemia outbreaks to study *F. tularensis* persistence. By analyzing *F. tularensis* genomes from 138 individuals infected from 1994 to 2010 in Örebro County in Sweden and performing a long-term laboratory storage experiment, we explored the microbial population concept of a pathogen seed-bank. We found that eight indistinguishable genomes – each of them defined by no SNPs across 1.65 million whole-genome nucleotides – locally persisted over 2-9 years. We found unmeasurable SNP accumulation rates and overlapping bacterial generations among the outbreak genomes and that *F. tularensis* survived in saline for four years without nutrients. By these findings, and analyses of nucleotide substitution patterns, we suggest that a pathogen seed-bank effect is an important feature of *F. tularensis* ecology influencing genetic diversity.

In paper III, we developed a new concept for source attribution of a *F. tularensis* sample. We aimed to identify genetic variation that is characteristic to laboratory culturing and we used culture amplification to identify genetic variation present at exceedingly low frequencies in a sample. Based on a biological enrichment scheme followed by high-throughput sequencing, we could track genetic variation back to a source sample. These results suggest that the concept has potential for linking a *F. tularensis* sample to its laboratory source sample.

Taken together, the results presented in this thesis provide new understanding of the dissemination patterns and local persistence of tularemia. This is important for the interpretation of molecular epidemiology investigations of the disease. In a wider context, the results demonstrate how spatial dispersal and a microbial seed-bank effect may contribute to the diversity of a disease-causing agent. Finally, we have described a promising concept for source attribution of *F. tularensis* samples.

**Keywords**

*Francisella tularensis*, persistence, dispersal, source attribution, clonal bacteria, epidemiology, microbial genomics, population genetics, seed bank, microbial evolution, mutation rate