Differential Gene Expression Analysis of RNA-Seq Data

Elise Delaporte, written for BIOL-7263 taught by Dr. Toomey

Introduction

Campylobacteriosis is considered to be the leading cause of bacterial diarrhea globally. Infection is caused by *Campylobacter spp.* and it frequently occurs after contact with or ingestion of contaminated raw or undercooked poultry. Serious long-term effects can occur after infection such as IBS and reactive arthritis. Additionally, Campylobacter causes at least a third of cases of Guillain-Barré syndrome, which can result in progressive and occasionally permanent paralysis. *Campylobacter* bacteria such as *C. jejuni* and *C. coli* are microaerophilic, which means they prefer oxygen concentrations of 2-10%. Despite its microaerophilic nature, *Campylobacter* has shown itself to be capable of surviving oxygen exposure during meat processing. It is now known that some strains of *Campylobacter* are aerotolerant and it appears to be this trait that allows them to survive long enough to contaminate our meat supply. The precise genetic mechanisms underlying aerotolerance are still unknown. The objective of this study is to examine gene expression of aerotolerant *Campylobacter spp*. under both aerobic and microaerobic conditions for the purpose of determining which genes contribute most to aerotolerance. Differential gene expression analysis is one method that can be used to examine how gene expression changes in response to certain conditions, such as oxygen exposure. In the present study, RNA samples were previously collected from two strains of *Campylobacter*, one from *C. jejuni* and one from *C. coli*, from multiple timepoints. The control timepoint, Time 0, collected the samples from cultures which had been kept under microaerobic conditions. The

remaining samples were collected at various timepoints after exposure to atmospheric levels of oxygen. A cDNA library was created and the samples were sent for sequencing. The objective of this project is to perform data analysis on the reads from the collected samples. In a previous but similar project, I performed the RNA-Seq analysis using CLC Genomics 22, followed by a differential expression analysis using the same program. I would like to move away from performing the differential expression analysis as it uses a somewhat unsuitable normalization method. I am currently testing using DeSeq2 (via a visual interface, DEBrowser) for the differential gene expression analysis. I hope to learn more methods that may be appropriate for both the read alignment and for the differential gene expression analyses. My expected outcome is to discover genes that are highly upregulated or downregulated under aerobic conditions as compared to the microaerobic timepoint.