Rhodobacter sphaeroides is a gram-negative facultative photosynthetic bacterium and has many metabolic capabilities. It can perform photosynthesis, anaerobic and aerobic respiration, which are made possible by the production of the tetrapyrroles, chlorophyll, heme, and vitamin B12. Biosynthesis of all tetrapyrroles begins with the formation of the essential metabolite 5-aminolevulinic acid (ALA). Ironically the growth of certain strains of this bacterium, including wild type strain 2.4.1, is completely inhibited by the presence of exogeneous ALA. However, mutant strain AT1, which is a derivative of 2.4.1, is insensitive to exogenous ALA, as is the subsequent derivative strain of AT1, CH10. So, 2.4.1 is sensitive, AT1 and CH10 are insensitive. What we are attempting to do here is to compare the DNA of these bacteria in order to tell us the culprit sequence responsible for ALA sensitivity/insensitivity. DNA sequencing has been performed using two methodologies which both generate large sets of sequence data that need to be assembled properly to generate the entire genome sequence. However, the final products do not always match up because of the difference in sequencing and assembly processes between the two methodologies.

All *R. sphaeroides* bacteria are from the laboratory collection, and each contain two chromosomes and five plasmids. In order to align the assemblies, the sequences were concatenated into a single sequence, meaning that the separate chromosome and plasmid sequences were all joined together into one mega-sequence. These mega-sequences were then compared pairwise using the BLASTn program at the National Center for Biotechnology Information. They were also globally compared using the progressiveMauve program. The dot plots generated using BLASTn show that the plasmids are the most problematic with respect to the two genome assembly algorithms. The progressiveMauve alignments show the relative orders of the homologous regions in all four genomic sequences.

These results will be used to properly rearrange the sequence segments in order to make it possible to compare the bacterial genomes at the nucleotide level using a program called UGENE. This program allows us to visualize the alignments at the nucleotide level. By inspecting these nucleotide alignments, it will allow us to finally identify the actual sequence differences between the ALA sensitive strain 2.4.1 and the two ALA insensitive strains AT1 and CH10, any of which may be responsible for the difference in ALA sensitivity. Then, using molecular biology techniques, we will be able to confirm the sequence differences that account for ALA sensitivity.