Genomic Evolutionary Comparison of *Streptomyces* species Living in Different Geological Habitats

Matthew Hanks

Department of Biological Sciences

(Chemical Biology)

Dr. Naowarat (Ann) Cheeptham (Ph.D.) supervisor. Department of Biological Sciences

Abstract:

In bacteria, chemical communication involves producing, releasing, detecting, and responding to small signal molecules which are termed autoinducers. The way bacteria communicate with one another using chemical signal molecules is similar to how higher or more complex organisms communicate. *Streptomyces* is a genus of Gram-positive, spore forming, generally aerobic, biofilm forming bacteria. Many bacteria share conserved genes and gene clusters for essential functions; however, they can also adapt to the unique environments they live in. The goal of this study is to determine evolutionary similarity of *Streptomyces* that live in different habitats around the world and how this relates the conservation of specific genes that are responsible for biofilm formation. To accomplish this, a phylogenetic tree was created using the genomes (obtained from NCBI) of *Streptomyces* that live in unique geological habitats. Additionally, the acetate cation symporter was identified to be crucial in the biofilm formation process and the protein coding genes were compared for each *Streptomyces*. The phylogenetic trees were created using the software DIVERGE 3.0 and MEGA11 and compared to each other to see if the evolutionary similarity of the whole genome is consistent with the genes responsible for biofilm formation and how this relates to the habitat they live in. This project has given a better understanding of the relationship between the genome and the habitat of an organism and suggests that there is no direct correlation between the whole genome sequence, protein coding sequence, and habitat that the bacteria living.

Table of Contents

Introduction:

Bacterial Communication / Cognition

In bacteria, chemical communication involves producing, releasing, detecting, and responding to small signal molecules which are called autoinducers (Waters 2005). The way bacteria communicate with one another using chemical signal molecules is similar to how higher or more complex organisms communicate (Federle 2003). The information supplied by these molecules is critical for synchronizing and organizing the activity of a large group of cells. An example of bacterial communication is a process called quorum sensing (Waters 2005). Quorum sensing involves the bacteria monitoring their cell-population density by measuring the concentration of secreted signal molecules. As the population of bacteria grows, the individual organisms produce and secrete the signal molecules into the extracellular environment. The bacteria are therefore able to determine the amount of other bacterial cells and alter specific gene expression accordingly. Other examples of bacterial communication include bioluminescence and biofilm formation.

Streptomyces Bacteria

Streptomyces are Gram-positive, aerobic, filamentous, spore and biofilm forming bacteria within the class of Actinobacteria (Chater 2006). These bacteria are mainly found in the soil but are occasionally isolated from other environments such as manure and water. They reproduce by sending up specialized aerial branches called mycelium, which from and hold the spores. *Streptomyces* are important in the decomposition of organic matter such as decaying leaves which contributes to the fertility and earthy odour of soil. Some species are becoming commonly known for their ability to produce broad-spectrum antibiotics (Gosse 2019). These chemicals are naturally produced by the bacteria to kill or inhibit the growth of other microorganisms.

Biofilm Formation

Biofilms are a microbial survival strategy that provide stability to bacterial colonies (Kostakioti 2013). Biofilm formation is a process where microorganisms including bacteria attach and grow on a surface as well as produce extracellular polymers to facilitate the film formation. Bacterial biofilm formation can be described in three steps: attachment, maturation and dispersion. Bacteria form biofilms in response to many environmental stresses such as UV radiation, limited nutrients, extreme temperature, etc. Biofilms can be removed by strong detergents or viruses which infect the bacterial cells with high specificity (such as bacteriophages).

Acetic acid as a Volatile Signal for Biofrilm formation

Volatiles are small air-transmittable chemicals with diverse biological function. A specific volatile, acetic acid plays a central role in the formation of biofilms (Chen 2015). Acetic acid is used as a metabolic signal to coordinate the timing of biofilm formation in many bacteria. In *streptomyces*, the acetate, sodium-dependent symporter is a protein that acts as a transporter for the acetic acid volatile. From literature research it was identified that the gene coding for this protein also contained conserved and variable regions (Rorick 2010).

Goal of the study

There were two goals for this study. The first was to determine the evolutionary similarity of *Streptomyces* that live in different geological habitats around the world. The second was to compare the conservation of the gene that codes for the sodium-dependent symporter protein to the whole genome sequence of that species. This research project was an introduction into bacterial cognition and

the concept that cell communication is a critical factor in the diverse behaviors found in different bacteria. By identifying the genes that are involved and responsible for the formation of biofilms is a step towards better understanding the cognitive function of bacteria. In further research, identifying and labeling the genes that are responsible for different bacterial behaviors would allow us to learn more about the ecological function, adaptation, and evolution of bacteria.

Methods

Familiarization with BLAST, DIVERGE 3.0, and MEGA11

The first step in this study was to learn about the bioinformatic tools and software that would be used for analysis. This included NCBI Blast, DIVERGE 3.0, and MEGA11. All these tools require the genome and gene sequences in an alignment file format in order to run the analysis. New and separate alignments files needed to be created for the whole genome shotgun sequences and the protein coding gene sequence. This was done by obtaining each sequence separately and manually adding them to a new file in the correct format.

Identification of *Streptomyces* Species

The next step in this study was to identify different *Streptomyces* species that formed biofilms, were sampled from different geological habitats, and that contained the sodium-dependent symporter protein. This involved reviewing literature articles and *Streptomyces* species information in databases such as NCBI.

Compare the whole genome shotgun sequences

The whole genome shotgun sequences were compared using the DIVERGE 3.0 and MEGA11 software. Both software's gave the same phylogenetic tree result. The phylogenetic tree was created using the Kimura distance and the neighbor-joining (NJ) tree model. The comparison was attempted through the NCBI Blast software however, there was a nucleotide sequence limit which prohibited the whole genome shotgun sequence from being analyzed.

Compare the protein coding gene sequence

The protein coding gene sequences were all compared using the NCBI Blast, DIVERGE 3.0, and MEGA11 software's. The conserved regions of these genes were not analyzed since they are shared between the different *Streptomyces* species. However, the variable regions were compared and used to create a phylogenetic tree. The phylogenetic tree was created using the Kimura distance and the neighbor-joining (NJ) tree model. NCBI blast also provided numerical data about the similarity between each of the sequences.

Results

Table 1. *Streptomyces* species chosen as well as their NCBI accession number and the habitat they live in.

***Streptomyces* Species**	**NCBI Accession Number**	**Geological habitat and Location Found in**
Streptomyces sp. ICC1	CP030287	Cave Rock, Canada
Streptomyces sp. ICC4	CP030286	Cave Rock, Canada
Streptomyces tricolor	NZ_MUMF00000000	Soil, Germany
Streptomyces cocklensis	NZ_CAJSLV000000000	Hay Meadow Soil, UK
Streptomyces tirandamycinicus	CM001165	Marine Sponge, China
Streptomyces abyssalis	NZ_LJGS00000000	Deep Sea Sediment, China

From the whole genome shotgun sequences, a phylogenetic tree was created comparing the six chosen *Streptomyces* species. Figure 1 shows the layout of this tree and the similarity between the different species.

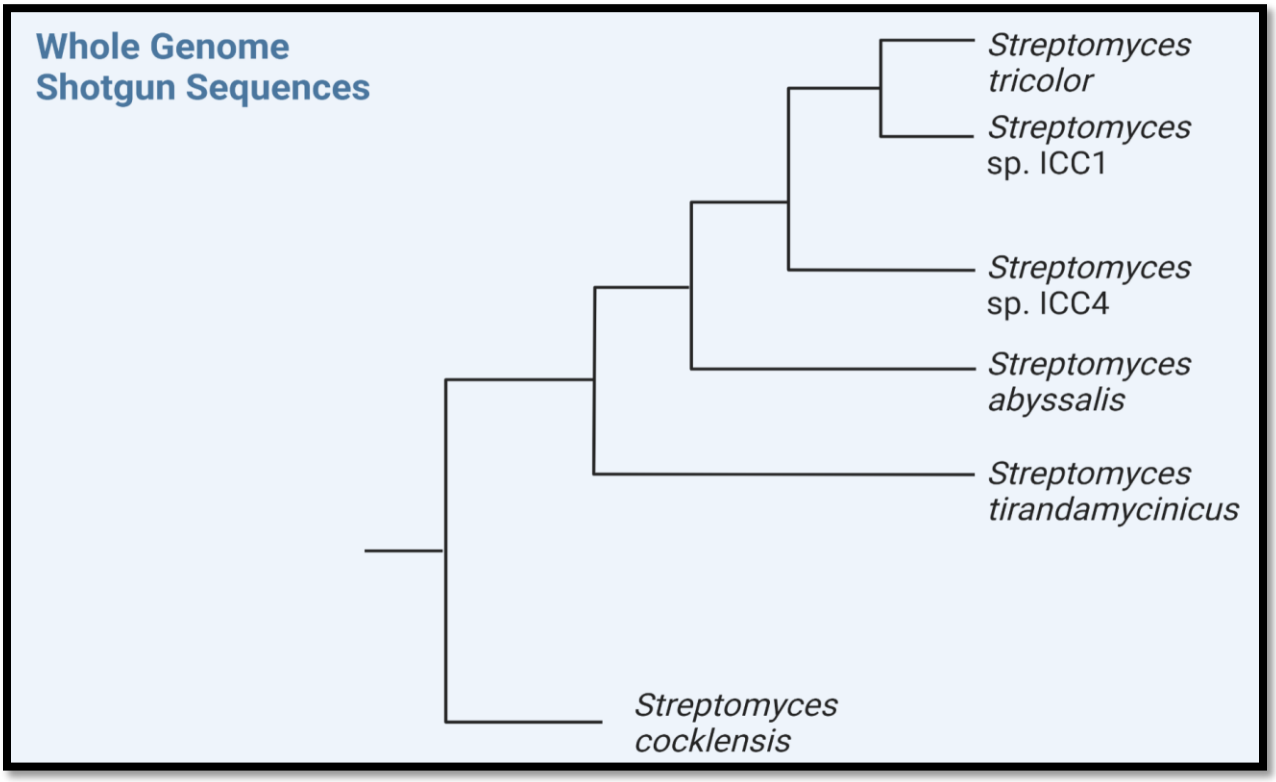

Figure 1. Phylogenetic tree showing full genome similarity of *Streptomyces* species.

The whole genome shotgun sequence phylogenetic tree (Figure 1) was used to compare the similarity in whole genome to the geological environment that the bacteria live in.

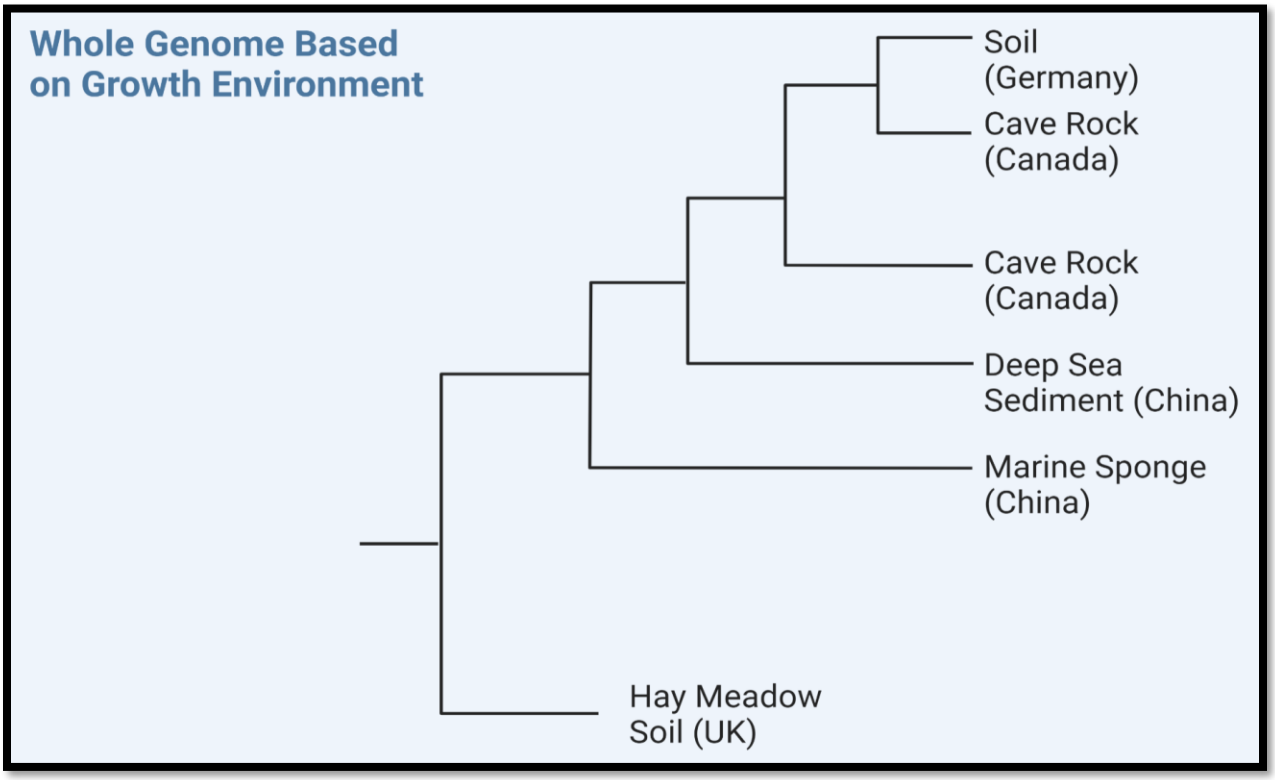

Figure 2. Phylogenetic tree of the species of Figure 1 showing geological habitats of each species.

From the variable sequences of the protein coding gene, a phylogenetic tree was created comparing the six chosen *Streptomyces* species. Figure 3 shows the layout of this tree and the most similarity between the different species. Figure 4 shows the whole protein coding genes including the conserved and variable regions.

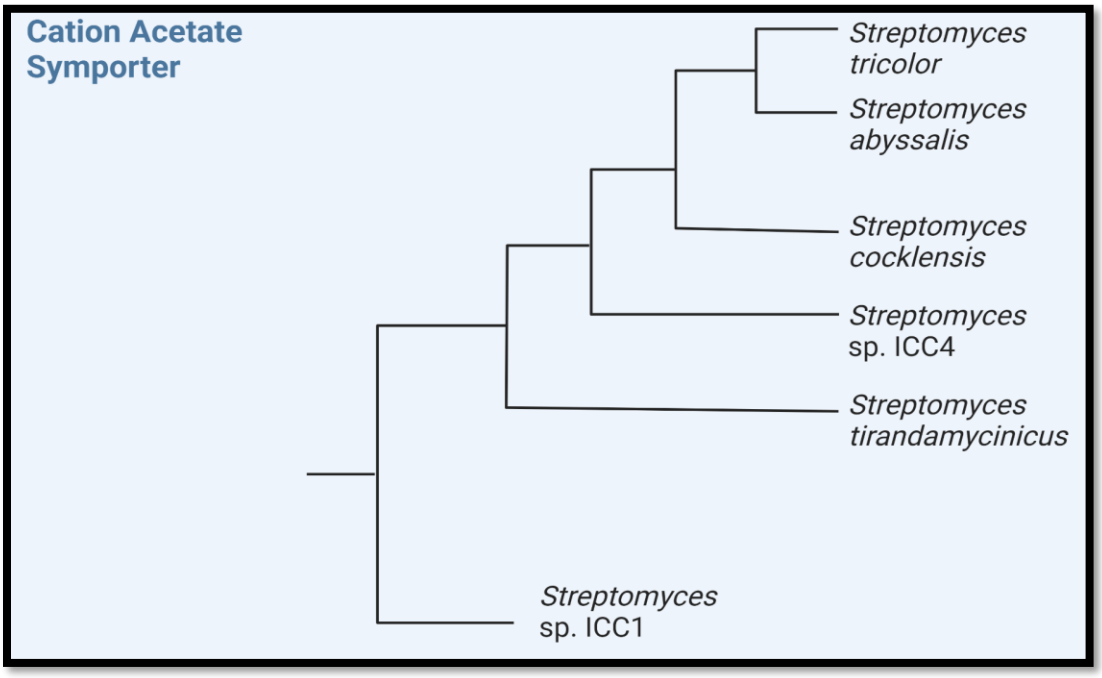

Figure 3. Phylogenetic tree of the targeted cation symporter of each *Streptomyces* species.

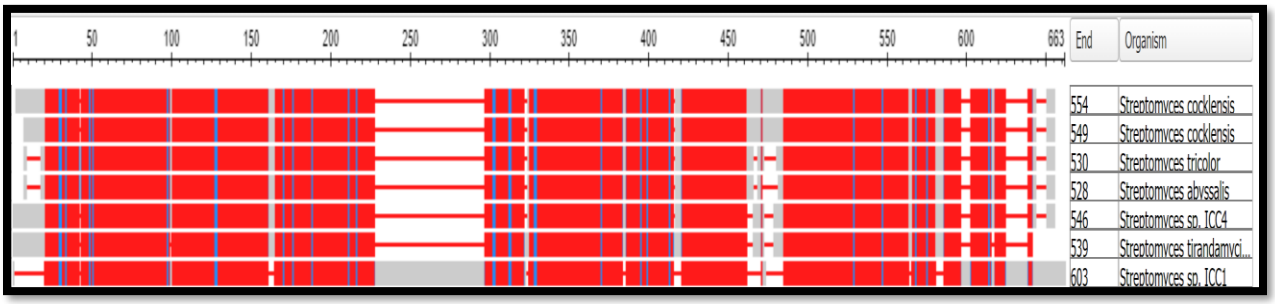

Figure 4. The genes coding for the targeted protein from different *Streptomyces* species in alignment view to show similarities and differences.

Discussion

By investigating figure 1, it can be observed that *Streptomyces* species ICC1 and ICC4 are not the most closely related. This is interesting since ICC1 and ICC4 were both isolated from the Iron Curtain Caves in Chilliwack, BC. Instead, *Streptomyces* species ICC1 and *tricolor* are the most similar. Next when you observe figure 2, this shows that the marine samples are not the closet related and soil samples are not the closest related. This is similar to the previous statement regarding ICC1 and ICC4. This suggests that the similarity in whole genome sequences is not directly correlated to the geological living environment of the bacteria.

By comparing figures 1 and 3, it can be observed that the phylogenetic tree showing the similarity in whole genome sequence is different than that of the protein coding gene sequence. Based on whole genome sequences, the species *tricolor* and ICC1 are most similarly related. However, based on the protein coding gene sequence, the species *tricolor* and *abyssalis* are most closely related.

This study only analyzed six different *Streptomyces* species which is a small sample number. Therefore, more research is needed to confirm these hypotheses. The purpose of this study was an introduction into bacterial cognition and the concept cell communication. A large amount time at the beginning of this study was spent learning how to use the different bioinformatic tools and software such as NCBI Blast, DIVERGE 3.0, and MEGA11 since they all worked vary differently. It was necessary to learn how to use these software's in order to conduct the analysis.

In future research, identifying and labeling the genes that are responsible for different bacterial behaviors would allow us to learn more about the ecological function, adaptation, and evolution of bacteria.

Conclusion

This study used literature research that showed the acetate sodium-dependent symporter protein is necessary for biofilm formation and that the gene coding for this protein contains conserved and variable regions. Furthermore, this study suggests that the similarity in the variable region of the protein coding gene does not appear to be correlated with geological habitat and shows large variation between species. Also, this study suggests the similarity of the full genome of different *Streptomyces* species also does not appear to be correlated to the geological habitat or location the sample was taken from.

Acknowledgments:

I would like to thank my supervisor Dr. Naowarat Cheeptham (Ann) as well as Dr. Jessica Thandara-Gosse from the University of Ottawa for supporting my research. I would like to acknowledge that the NCBI databases were consulted for all whole genome sequences as well as the protein sequences. I would also like acknowledge that the bioinformatic software NCBI Blast, DIVERGE 3.0, and MEGA11 were used in the analysis of this project

References

1. Chater K. F. (2006). Streptomyces inside-out: a new perspective on the bacteria that provide us with antibiotics. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 361(1469), 761–768. https://doi.org/10.1098/rstb.2005.1758
2. Chen, Y., Gozzi, K., & Chai, Y. (2015). A bacterial volatile signal for biofilm formation. Microbial cell (Graz, Austria), 2(10), 406–408. https://doi.org/10.15698/mic2015.10.233
3. Federle, M. J., & Bassler, B. L. (2003). Interspecies communication in bacteria. The Journal of clinical investigation, 112(9), 1291–1299. https://doi.org/10.1172/JCI20195
4. Gosse, J. T., Ghosh, S., Sproule, A., Overy, D., Cheeptham, N., & Boddy, C. N. (2019). Whole Genome Sequencing and Metabolomic Study of Cave Streptomyces Isolates ICC1 and ICC4. Frontiers in microbiology, 10, 1020. https://doi.org/10.3389/fmicb.2019.01020
5. Kostakioti, M., Hadjifrangiskou, M., & Hultgren, S. J. (2013). Bacterial biofilms: development, dispersal, and therapeutic strategies in the dawn of the postantibiotic era. Cold Spring Harbor perspectives in medicine, 3(4), a010306. https://doi.org/10.1101/cshperspect.a010306
6. Rorick, M. M., & Wagner, G. P. (2010). The origin of conserved protein domains and amino acid repeats via adaptive competition for control over amino acid residues. Journal of molecular evolution, 70(1), 29–43. https://doi.org/10.1007/s00239-009-9305-7
7. Valentini, M., & Filloux, A. (2016). Biofilms and Cyclic di-GMP (c-di-GMP) Signaling: Lessons from Pseudomonas aeruginosa and Other Bacteria. The Journal of biological chemistry, 291(24), 12547–12555. https://doi.org/10.1074/jbc.R115.711507
8. Waters, C. M., & Bassler, B. L. (2005). Quorum sensing: cell-to-cell communication in bacteria. Annual review of cell and developmental biology, 21, 319–346. https://doi.org/10.1146/annurev.cellbio.21.012704.131001

9. Wei, C., Jiang, W., Zhao, M. et al. A systematic analysis of the role of GGDEF-EAL domain proteins in virulence and motility in Xanthomonas oryzae pv. oryzicola. Sci Rep 6, 23769 (2016). https://doi.org/10.1038/srep23769