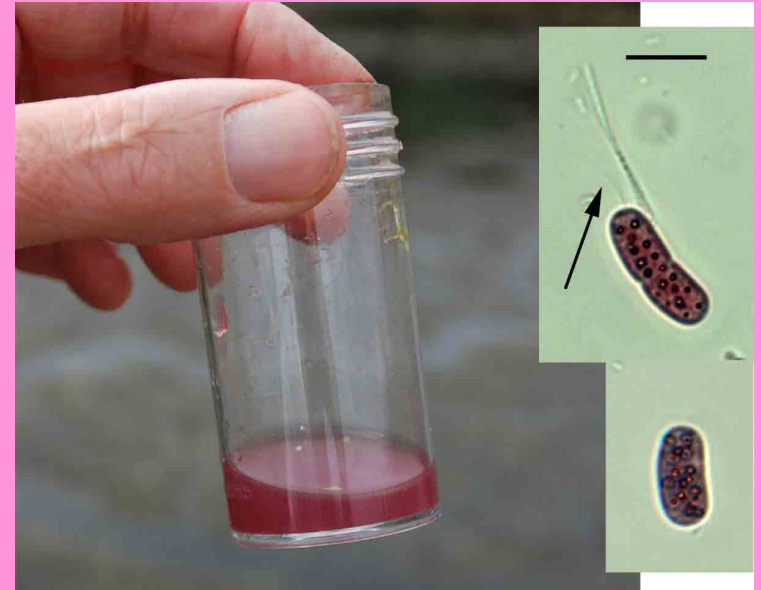




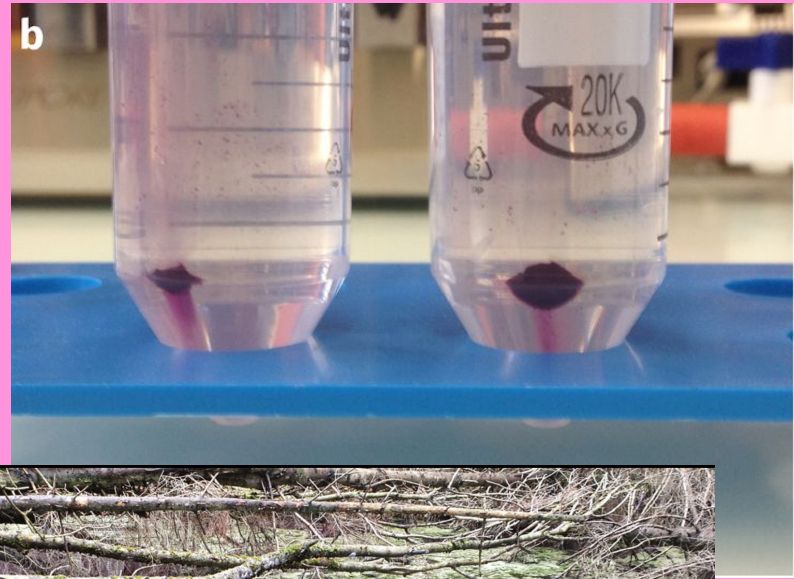
Chromatium okenii

Lillyanna Azevedo



Introduction

- To identify the products and pathways of genes responsible for producing its sparkly and slimy secretions.
- I chose this bacterium because of its unique ability to leave sparkly trails as it moves. It can also turn bodies of water pink or purple when its present in high concentrations.



Methods and Program Function

- Spades: takes short DNA fragments (sequencing reads) and assembles them into a complete genome
- Barnapp: Locates and identifies rRNA genes (16S, 23S, 5S), their direction, and can annotate rRNA regions in incomplete genomes.
- Bedtools: Extracts the sequences of the rRNA regions from the assembled genome and save them as a fasta file
- NCBI: Identified my bacterium and its neighbors by using its nucleotide sequence
- FastANI: Calculates how similar bacterial genomes are
- RAST: takes the assembled genome and identifies genes, predicts their function, and organizes them into biological systems (pathways and processes)

SPAdes has the better quality

output

report.txt

All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

```
Assembly
# contigs ( $\geq 0$  bp) 595
# contigs ( $\geq 1000$  bp) 103
# contigs ( $\geq 5000$  bp) 63
# contigs ( $\geq 10000$  bp) 53
# contigs ( $\geq 25000$  bp) 36
# contigs ( $\geq 50000$  bp) 23
Total length ( $\geq 0$  bp) 3381870
Total length ( $\geq 1000$  bp) 3164274
Total length ( $\geq 5000$  bp) 3085401
Total length ( $\geq 10000$  bp) 3012554
Total length ( $\geq 25000$  bp) 2729234
Total length ( $\geq 50000$  bp) 2218210
# contigs 267
Largest contig 191506
Total length 3266544
GC (%) 52.62
N50 74383
N90 15382
auN 84698.8
L50 14
L90 48
# N's per 100 kbp 45.92
```

SPAdes



report.txt

All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

```
Assembly
# contigs ( $\geq 0$  bp) 3545
# contigs ( $\geq 1000$  bp) 251
# contigs ( $\geq 5000$  bp) 171
# contigs ( $\geq 10000$  bp) 113
# contigs ( $\geq 25000$  bp) 28
# contigs ( $\geq 50000$  bp) 4
Total length ( $\geq 0$  bp) 3675575
Total length ( $\geq 1000$  bp) 3060268
Total length ( $\geq 5000$  bp) 2838377
Total length ( $\geq 10000$  bp) 2423970
Total length ( $\geq 25000$  bp) 1069809
Total length ( $\geq 50000$  bp) 245917
# contigs 329
Largest contig 74255
Total length 3116287
GC (%) 52.58
N50 18512
N90 5263
auN 22871.8
L50 52
L90 165
# N's per 100 kbp 292.50
```

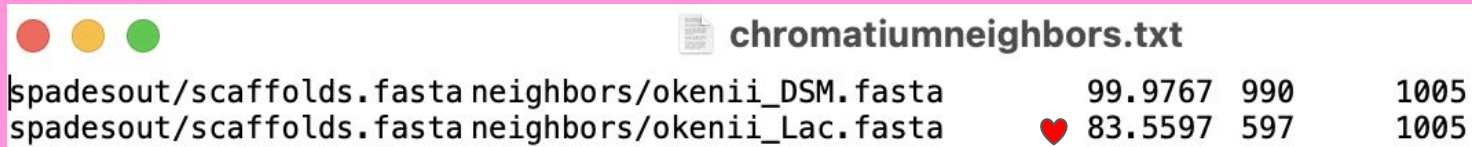
ABYSS



Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download Select columns Show 100								
select all 100 sequences selected								
GenBank Graphics Distance tree of results MSA Viewer								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Chromatium okenii 16S rRNA gene	Chromatium okenii	2604	2604	96%	0.0	98.83%	1455	Y12376.1
Chromatium okenii strain DSM 169 16S ribosomal RNA, partial sequence	Chromatium okenii	2558	2558	91%	0.0	99.86%	1391	NR_025315.1
Phototrophic bacteria 16S ribosomal RNA, clone 359	phototrophic bacterium	2523	2523	91%	0.0	99.43%	1390	AJ006221.1
Thiocystis violascens DSM 198, complete genome	Thiocystis violascens DSM 198	2471	7415	100%	0.0	95.93%	5017071	CP003154.1
Thiocystis violascens DSM 198 16S ribosomal RNA, partial sequence	Thiocystis violascens DSM 198	2459	2459	100%	0.0	95.92%	1517	NR_102951.1
Thiocystis sp. strain TopKyr17 16S ribosomal RNA gene, partial sequence	Thiocystis sp.	2364	2364	98%	0.0	95.37%	1489	MK278671.1
Chromatium weisselii partial 16S rRNA gene, strain 5910 = DSM 5161	Chromatium weisselii	2364	2364	96%	0.0	95.84%	1462	FN293053.1
Allochrochromatium vinosum DSM 180 chromosome, complete genome	Allochrochromatium vinosum DSM 180	2337	7011	100%	0.0	94.37%	3526903	CP001866.1
Candidatus Thiodictyon syntrophum strain Cae16T chromosome, complete genome	Candidatus Thiodictyon syntrophum	2324	4648	100%	0.0	94.17%	6837296	CP020370.1
Candidatus Thiodictyon intracellulare clone MG01 16S ribosomal RNA gene, partial sequence	Candidatus Thiodictyon intracellulare	2324	2324	100%	0.0	94.17%	1534	MW203126.1
Allochrochromatium vinosum DSM 180 16S ribosomal RNA, partial sequence	Allochrochromatium vinosum DSM 180	2324	2324	100%	0.0	94.35%	1516	NR_074584.1
Thiocystis gelatinosa partial 16S rRNA gene, type strain 4310 = DSM 215T	Thiocystis gelatinosa	2316	2316	96%	0.0	95.34%	1458	FN293058.1
Thiocystis gelatinosa strain N, Pflennig 2811 16S ribosomal RNA, partial sequence	Thiocystis gelatinosa	2303	2303	95%	0.0	95.32%	1451	NR_020334.1
Thiorhodovibrio frisiai strain 970 chromosome, complete genome	Thiorhodovibrio frisiai	2285	4570	100%	0.0	93.73%	5417120	CP121471.1
Thiorhodococcus sp. TROCPS10 16S ribosomal RNA gene, partial sequence	Thiorhodococcus sp. TROCPS10	2285	2285	97%	0.0	94.53%	1480	KC702859.1
Thiocystis minor strain 1211 16S ribosomal RNA, partial sequence	Thiocystis minor	2285	2285	96%	0.0	94.89%	1463	NR_036978.1
Thiodictyon bacillosum strain DSM 224 16S ribosomal RNA, partial sequence	Thiodictyon bacillosum	2276	2276	98%	0.0	94.07%	1500	NR_044364.1
Allochrochromatium tepidum NZ DNA, complete genome	Allochrochromatium tepidum	2272	6806	100%	0.0	93.58%	3208690	AP024563.1
Allochrochromatium vinosum partial 16S rRNA gene, strain JA169	Allochrochromatium vinosum	2261	2261	97%	0.0	94.33%	1479	AM690350.1

NCBI results

fastANI



A terminal window titled "chromatiumneighbors.txt" displays the output of a fastANI analysis. The window has a title bar with three colored buttons (red, yellow, green) on the left. The output shows two rows of data, each with four columns: the file paths, the ANI percentage, the number of aligned regions, and the total number of bases. The first row shows a high similarity of 99.9767% between two scaffolds. The second row shows a lower similarity of 83.5597% between a scaffold and a neighbor, with a red heart icon placed next to the percentage value.

File 1	File 2	ANI (%)	Regions	Bases
spadesout/scaffolds.fasta	neighbors/okenii_DSM.fasta	99.9767	990	1005
spadesout/scaffolds.fasta	neighbors/okenii_Lac.fasta	♥ 83.5597	597	1005

Due to the lack of research on this bacterium and its characterization, the neighboring genome with an 83% similarity might not be closely related. This data supports a distant relationship instead.

Results/Conclusion

Secretions are extracellular polymeric substances (EPS) excreted through its cell membrane

- Surface trails for possible protection, movement, or attachment

-used RAST to locate genes and their products for secretion

Special enzymes/protein /pathways for secretion	Main function
Glycosyltransferase	Modifies molecules that are secreted (functional, active, stable)
Type II secretory pathway	Transports proteins across membrane
Sec(A,B,C,D,etc) protein, unfolded Tat(A,B,C, D, etc.) protein, folded	Transports proteins across membrane
Lipase, Phospholipase, Amylase	Enzymes that are secreted
Hemolysin and other toxins	Toxins that are secreted
Sulfate reductase	Responsible for sparkly appearance
Bacteriochlorophyllide A dehydrogenase	Produces bacteriochlorophyll A, giving it its purple/pink pigment