EEB-528

Cal State San Bernardino
(909) 537-5000
christopher.neville@csusb.edu

Three High G+C soil proteobacterial genomes assembled using a hybrid approach

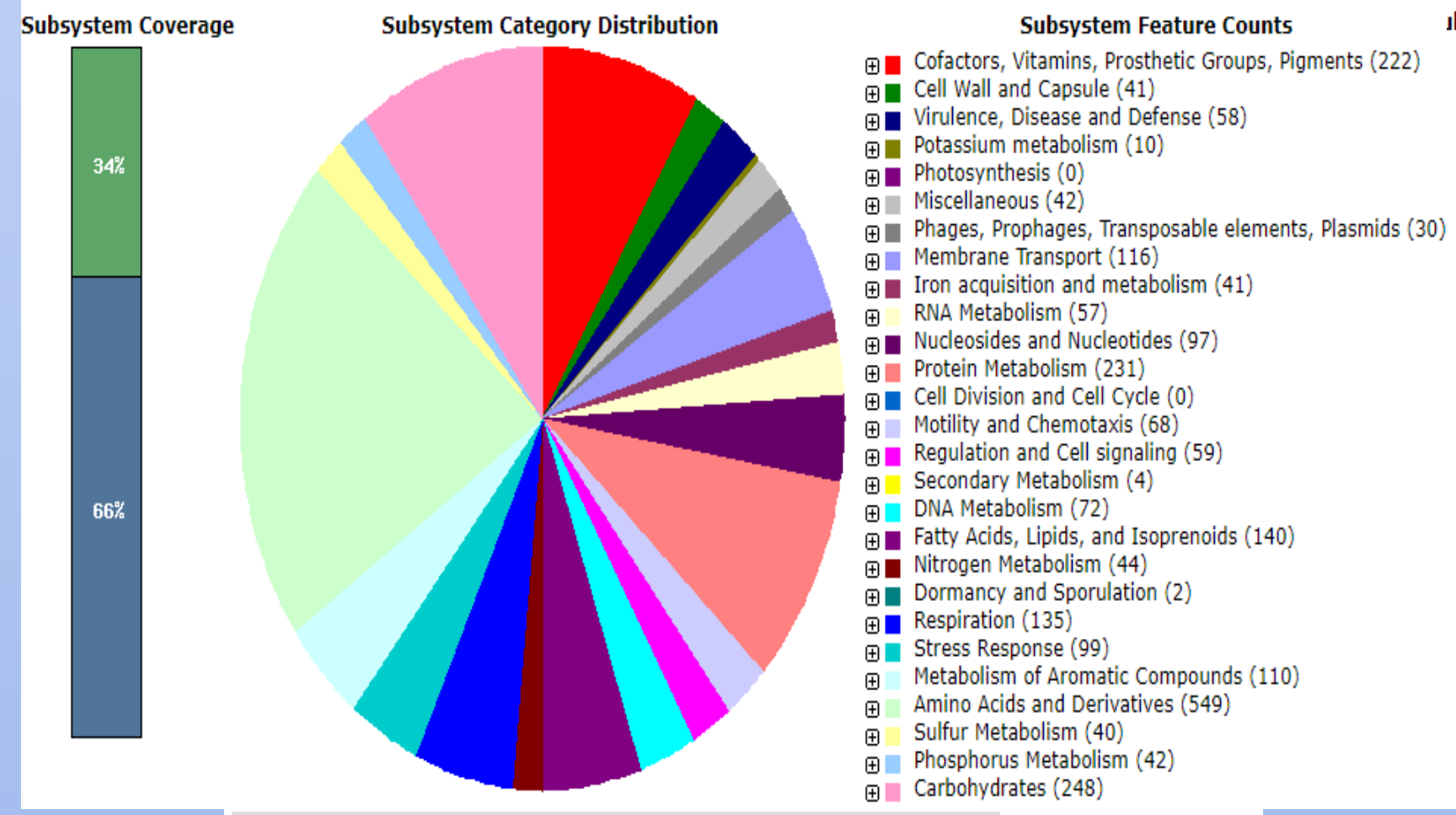
by Christopher Ne Ville, Jeremy Dodsworth, and Paul Orwin
Dept of Biology, California State University, San Bernardino

Abstract

Three high G+C bacterial genomes were sequenced and assembled using a hybrid approach. Long scaffolding reads were obtained for all three organisms using the Oxford Nanopore MinION system, and short accurate reads obtained through Illumina sequencing were used to alleviate non-random mutation in the final assemblies. The three soil bacteria sequenced were *Variovorax paradoxus* strain CSUSB, *V. paradoxus* strain VAI-C (a quorum sensing signal degrader), and a soil isolate of *Pseudomonas alkylphenolica* that was identified as tolerant to high levels of imidacloprid. Genomic DNA was isolated using the Quick and Loman ultra-long read protocol followed by shearing using a 24-gauge needle. The DNA was tagged and sequenced using Oxford Nanopores 3rd generation Sequencer (Flowcell MIN-106, Barcoding kit RBK-004). The long-read sequences were basecalled in Guppy demultiplexed with Deepbinner and adaptors removed with Porechop (v0.2.4). The strains were assembled using the long reads from the Oxford Nanopore in Unicycler without references. Short reads for polishing the *V. paradoxus* VAI-C and *P. alkylphenolica* strain were generated using the Illumina I-Seq 100 platform, starting with the same genomic DNA preparations and generating a 250-300 nt fragment size library using Nextera DNA Flex LPK. Sequencing on the iSEQ was performed using 2x150 paired end read generation from these libraries. The CSUSB strain was assembled with short reads obtained from the Illumina MiSeq platform (500-600 Library fragment size using Nextera XT kit, 2x250 reads). The genome assemblies were annotated in RAST, the VAI-C strain assembled into a three contig with 9,8,15,064 bp (68.9% GC) while the CSUSB strain gave a single assembly with 5,574,400 bp (65.7% GC). The P. alkylphenolica strain assembly generated a single contig with 5,612,010 (61.2% GC). Two of the three assemblies suggest a single chromosome structure for each of these isolates, in contrast with VAI-C with a secondary replicon & plasmids and other Variovorax paradoxus strains that carry as second chromosomal element. These finished genomes will form the basis for more substantial exploration of the evolution of genome architecture in these species.

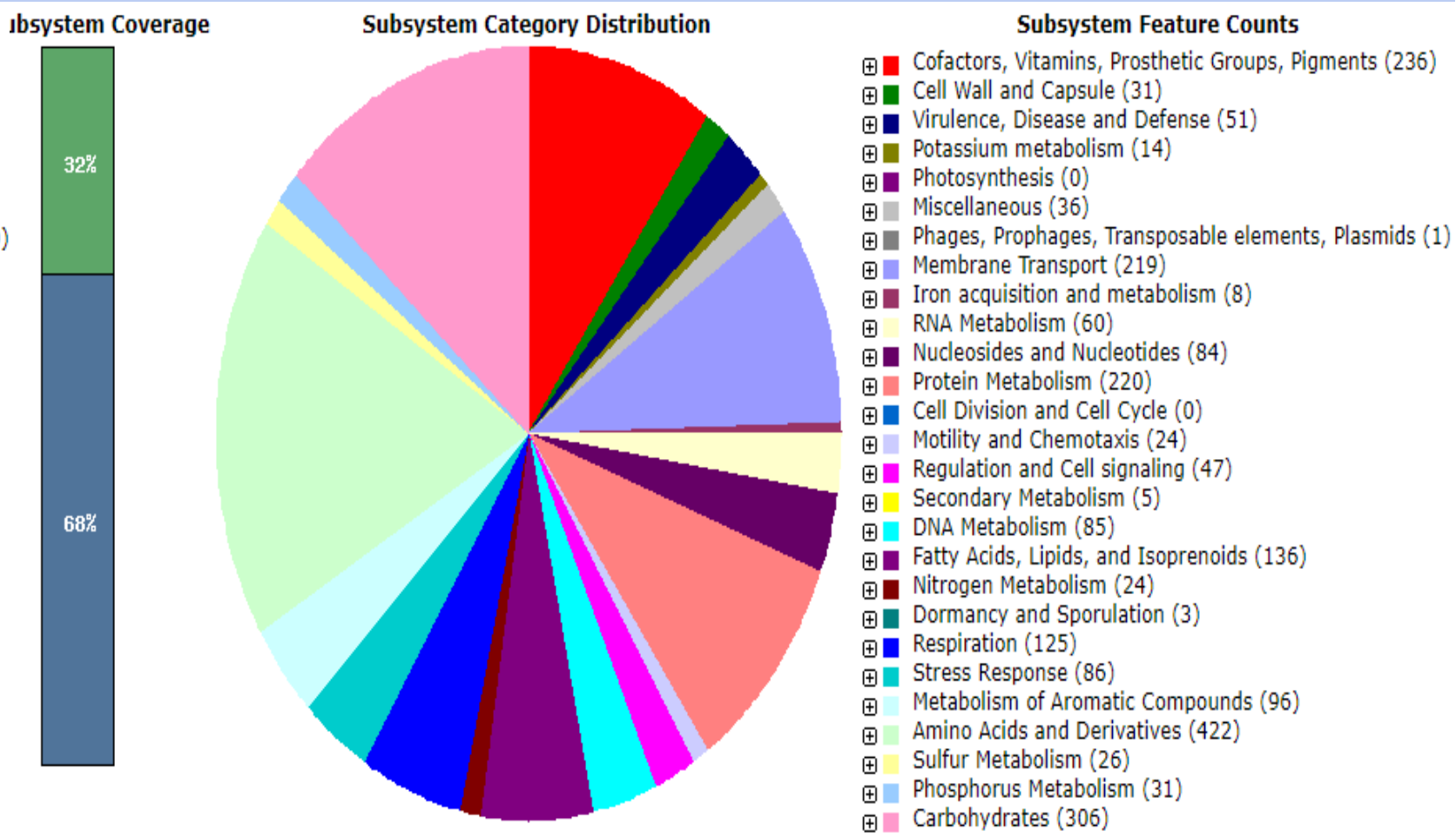
Genome Statistics, RAST annotation

Pseudomonas alkylphenolica



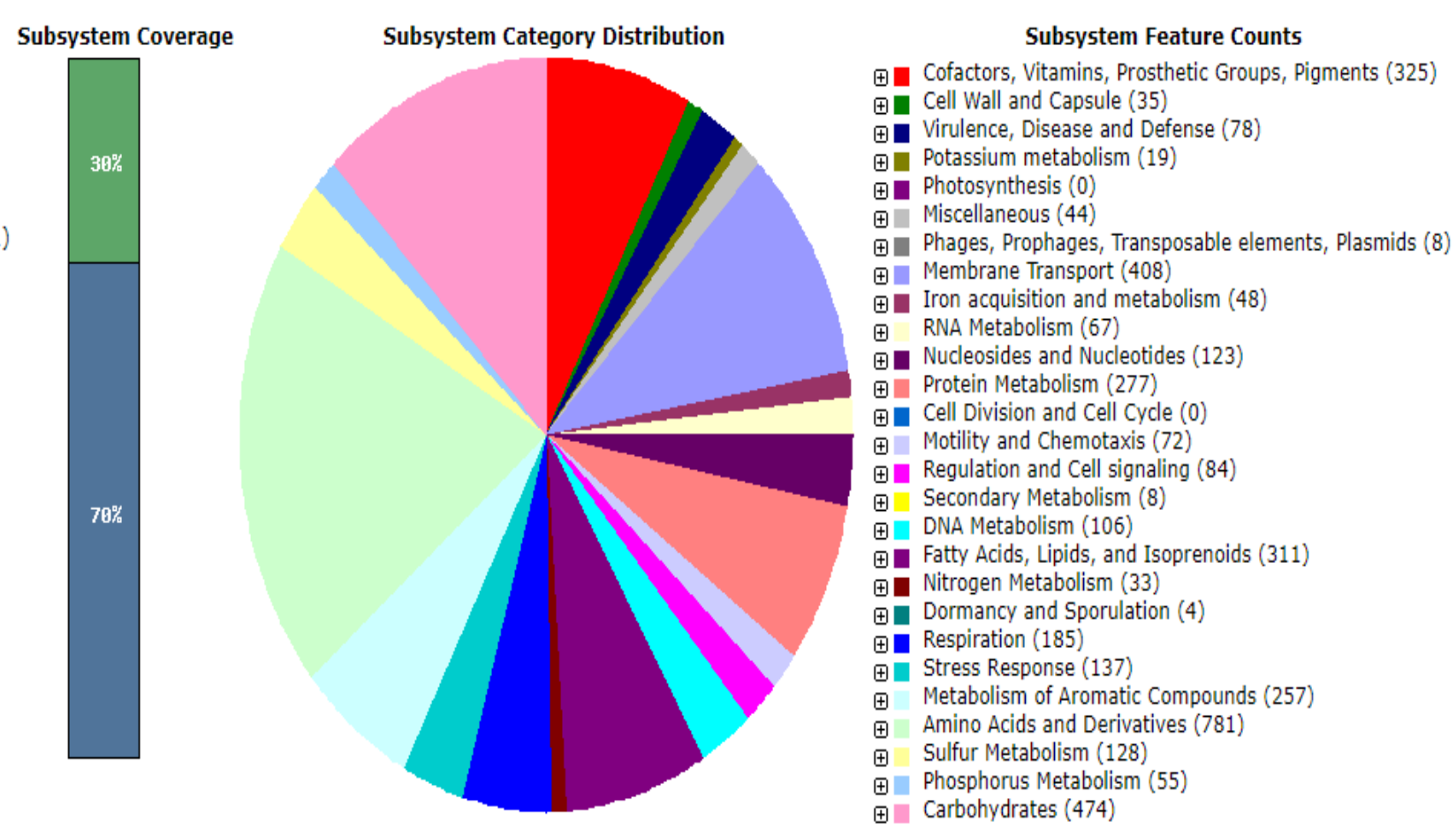
Size	5,612,010
GC Content	61.2
L50	1
Number of Contigs (with PEGs)	1
Number of Subsystems	398
Number of Coding Sequences	5079
Number of RNAs	89

Variovorax paradoxus CSUSB



Size	5,574,400
GC Content	65.7
L50	1
Number of Contigs (with PEGs)	1
Number of Subsystems	360
Number of Coding Sequences	5228
Number of RNAs	58

Variovorax paradoxus VAI-C

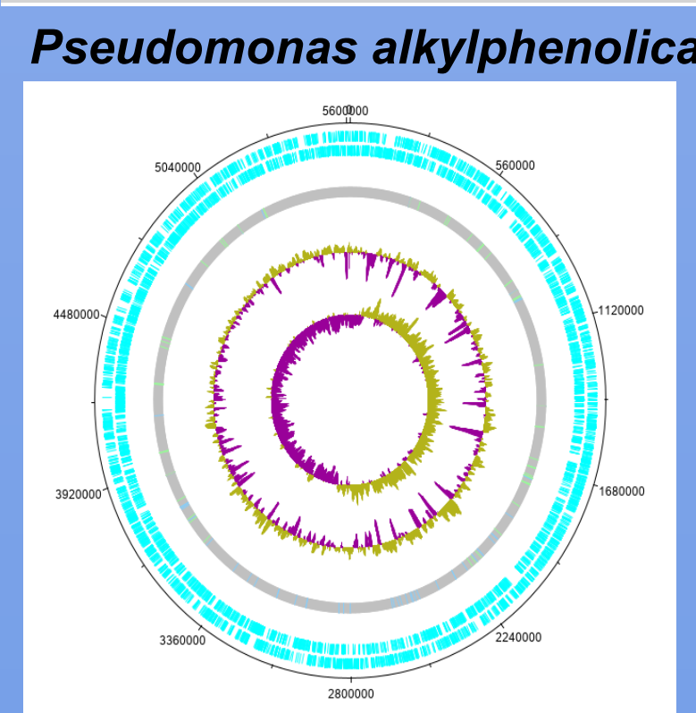


Size	9,815,064
GC Content	68.9
Number of Contigs (with PEGs)	3
Number of Subsystems	399
Number of Coding Sequences	9473
Number of RNAs	67

Read Stats & Genome Maps & Assembly

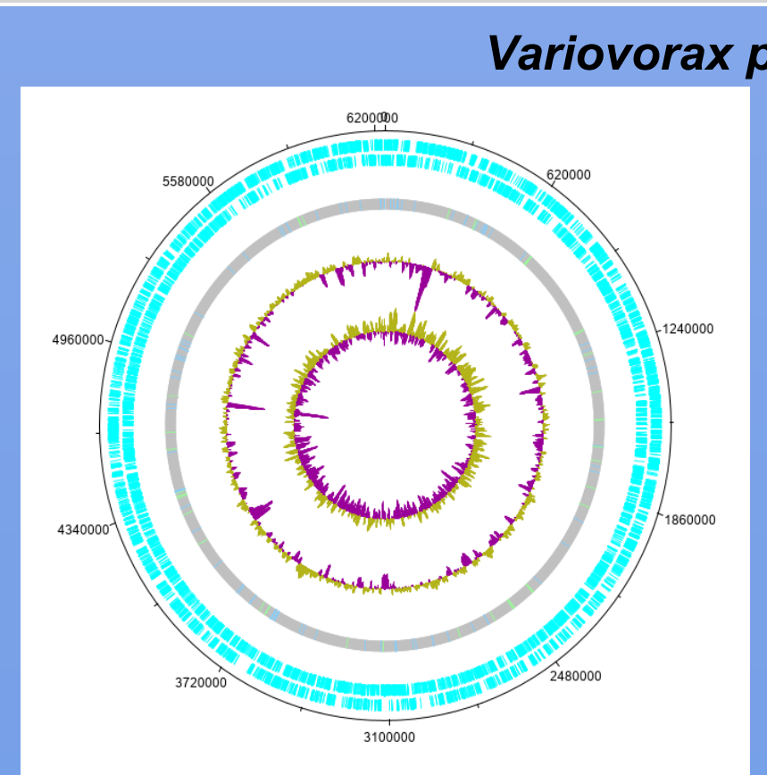
Organism-Long Read Stats	Total Number	Total Bases	Mean Length	Stddev	Fold Coverage
Variovorax sp. VAI-C	225,505	1,136,202,649	5038.480961	6,321.42	115.81x
Variovorax paradoxus CSUSB	657,284	4,969,806,827	7561.125521	9,721.00	885.56x
Pseudomonas alkylphenolica	35,176	336,763,152	9573.662497	10,371.88	60.00x
Organism-Short Read Stats	Total Number	Total Bases	Mean Length	Stddev	Fold Coverage
Variovorax sp. VAI-C	3,041,704	381,205,515	125.326302	19.440934	38.85x
Variovorax paradoxus CSUSB	1,582,102	309,602,593	195.690646	40.606	55.54x
Pseudomonas alkylphenolica	7,601,102	995,744,362	131	0	177.43x

Pseudomonas alkylphenolica

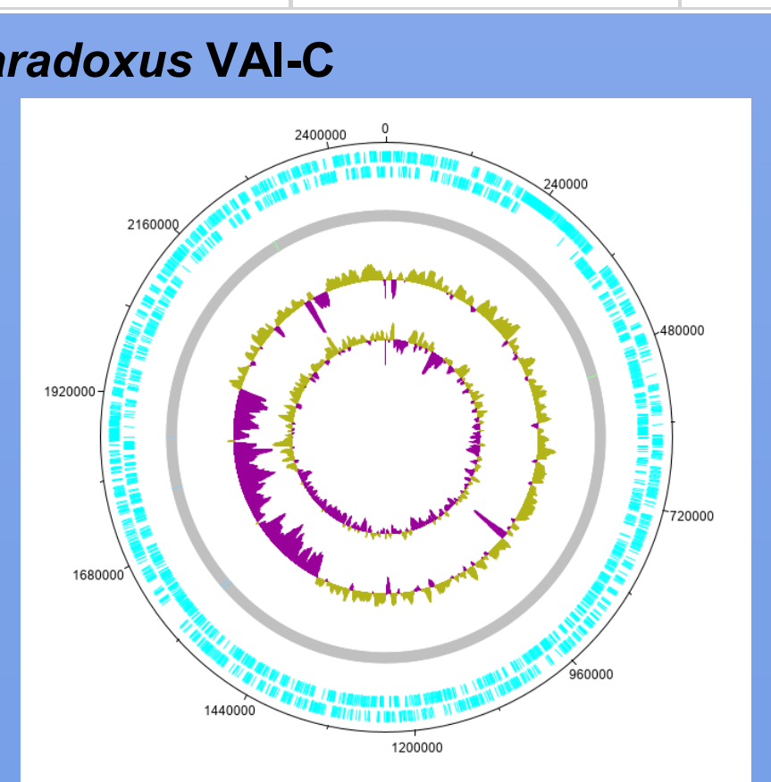


Circular Chromosome: 5,612,010 Bp

***Variovorax paradoxus* VAI-C**

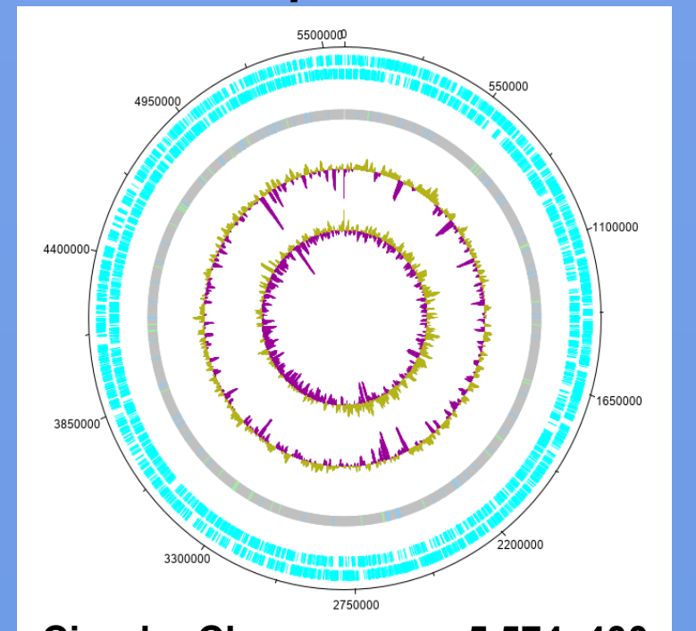


Primary Chromosome: 6,921,119 BP

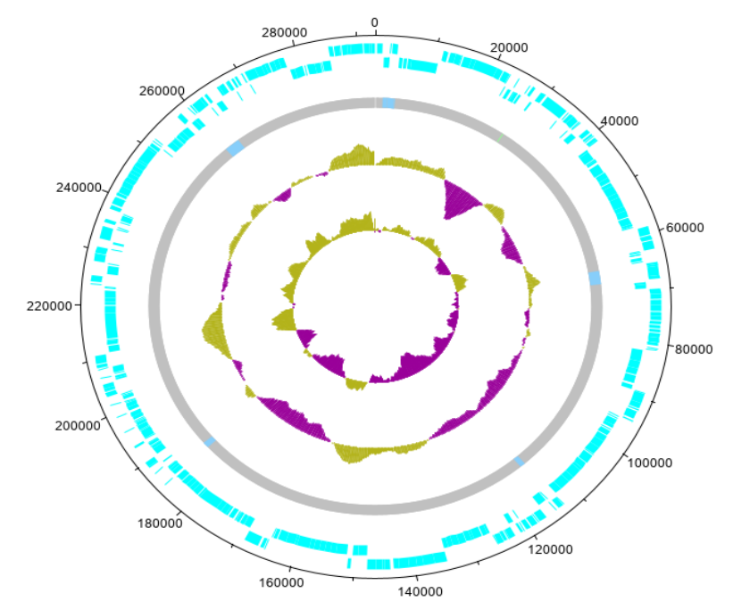


Secondary Replicon: 2,556,463 BP

***Variovorax paradoxus* CSUSB**

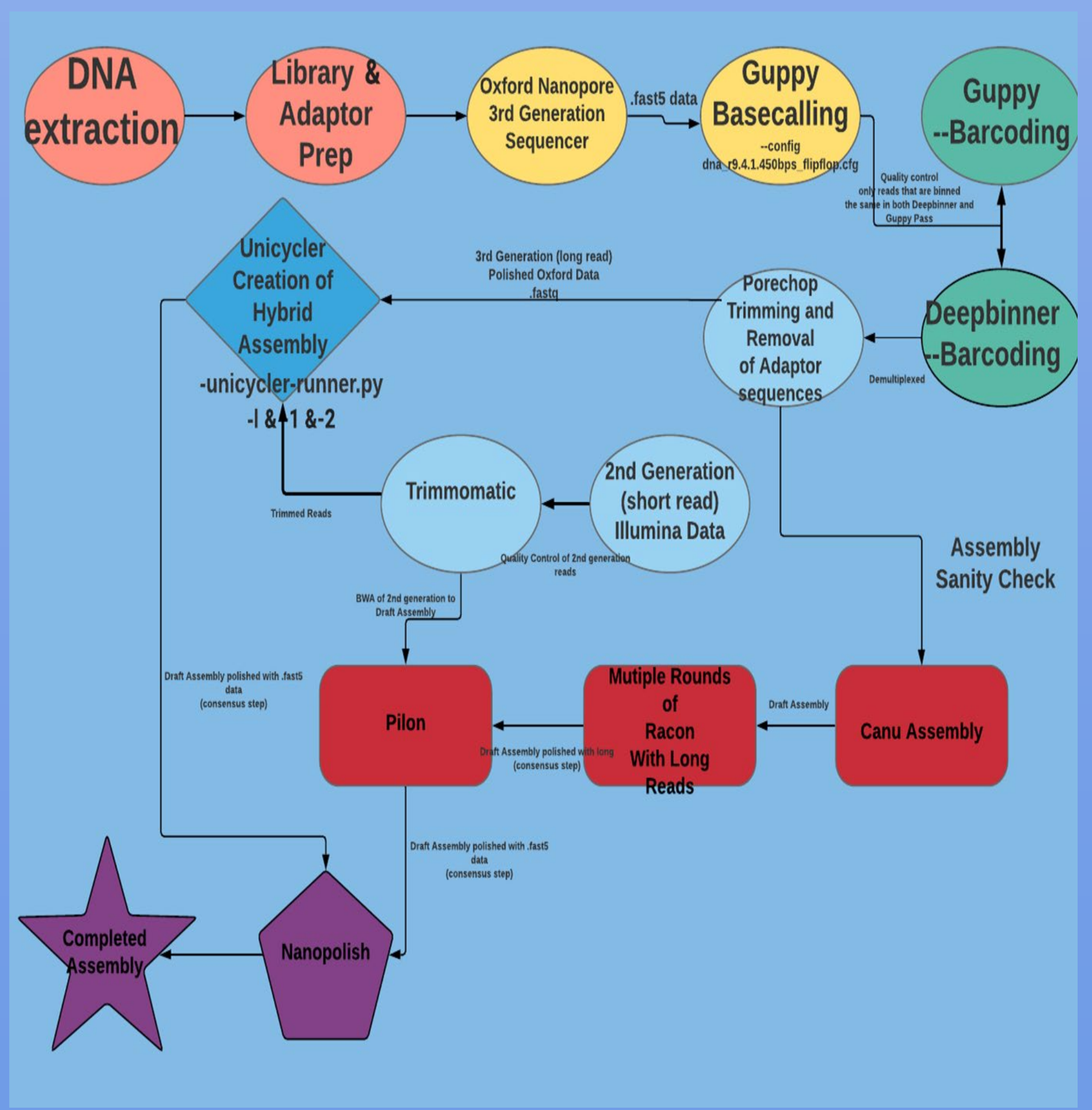


Circular Chromosome: 5,574, 400



Plasmid: 337,482 BP

Bioinformatics Pipeline



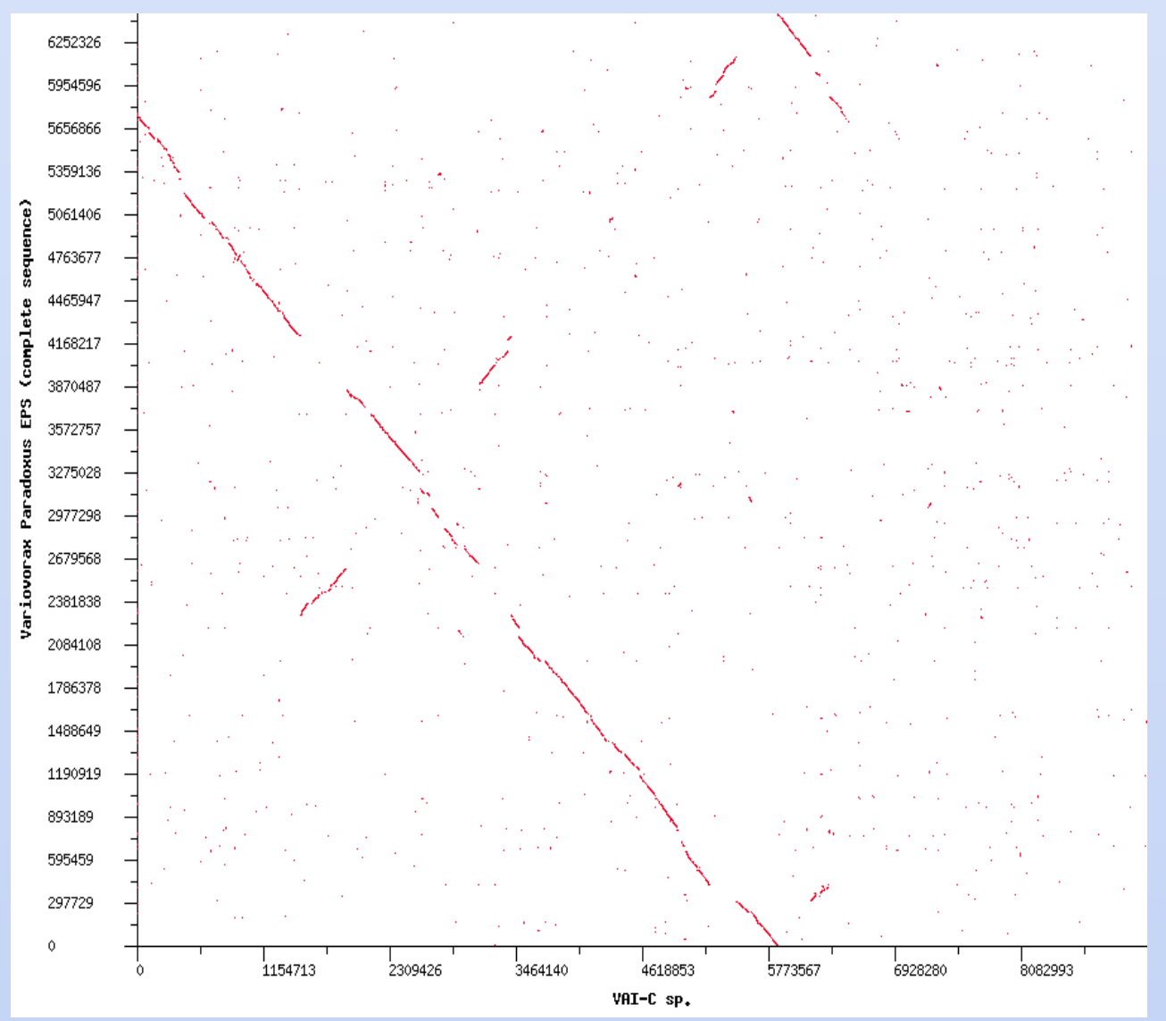
Analysis of VAI-C's Secondary Replicon

One Chromosome

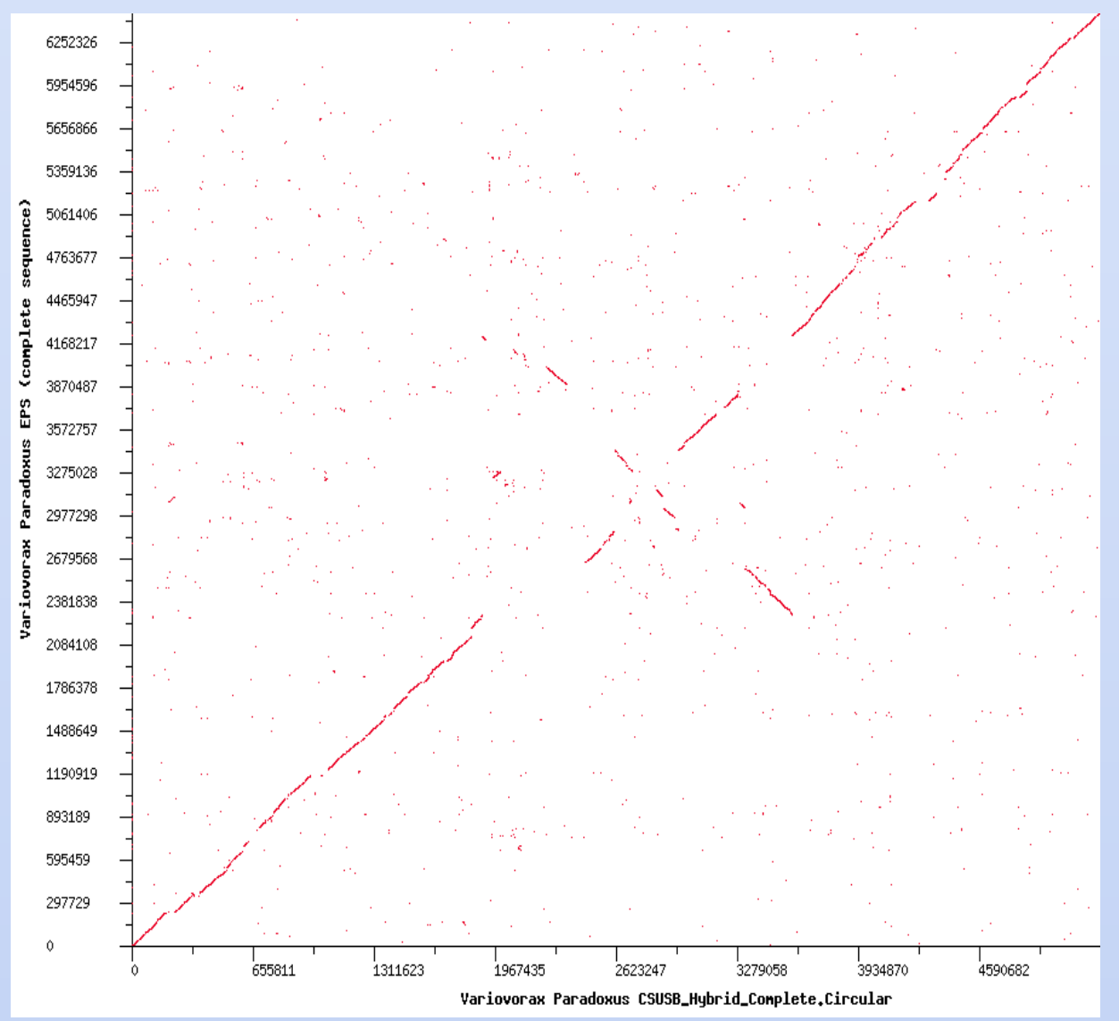
V. Paradoxus EPS

(Han, JI et al 2013)

Variovorax paradoxus VAI-C



Variovorax paradoxus CSUSB



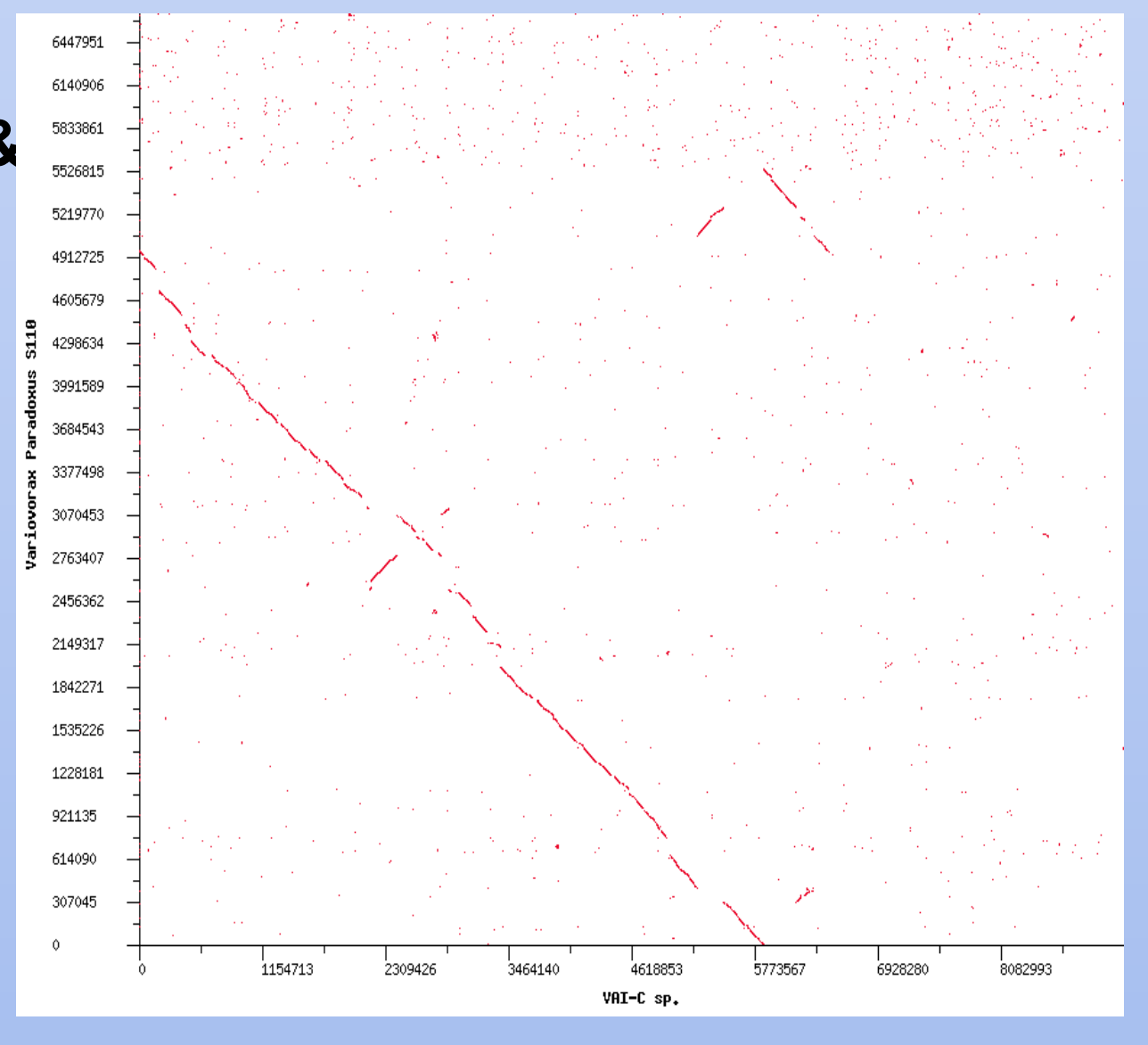
Two Chromosomes

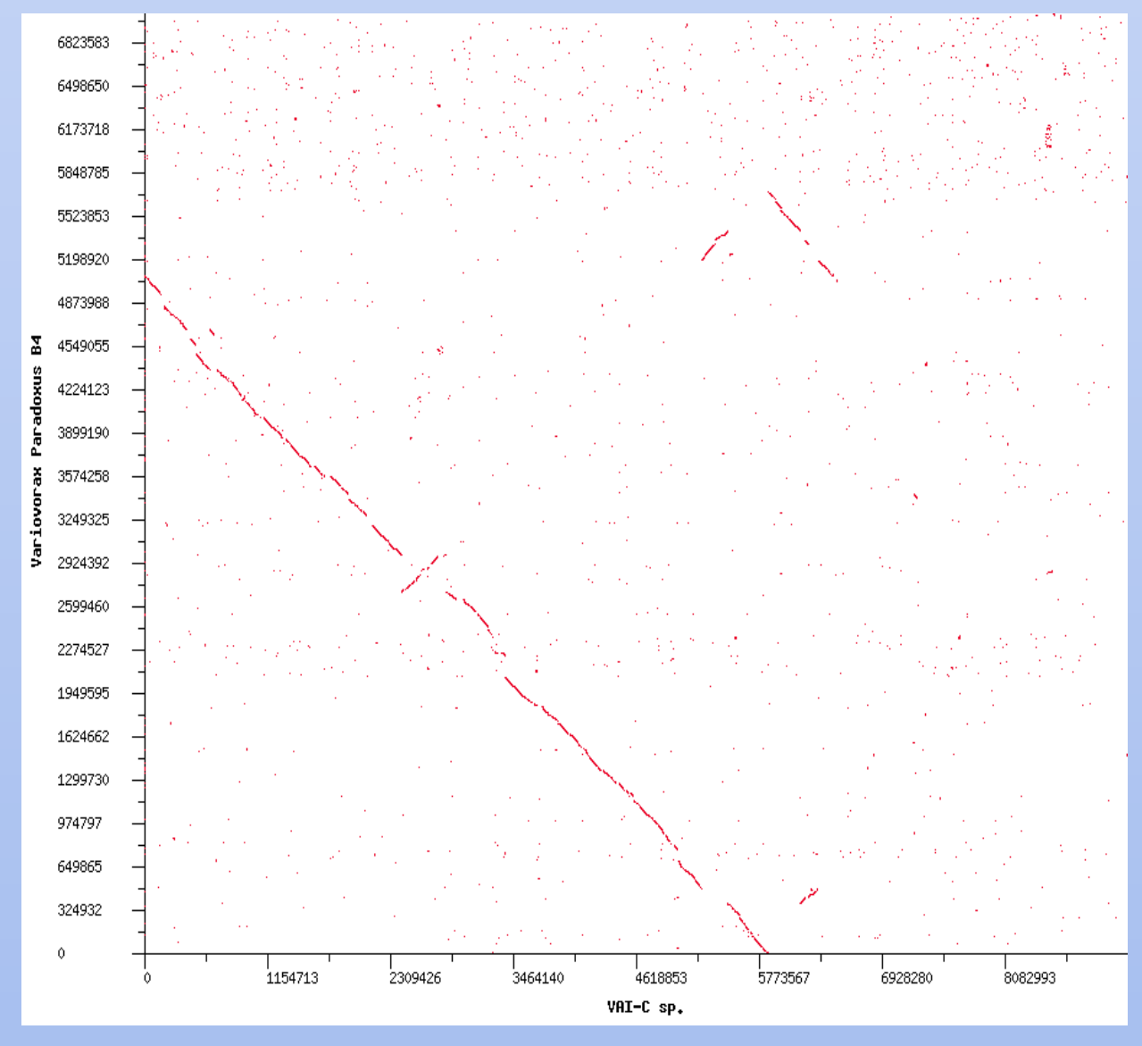
V. Paradoxus S110 &

V. Paradoxus B4

(Tringe,S et al. 2013)

(Irma Carbajal-Rodríguez et al 2011)





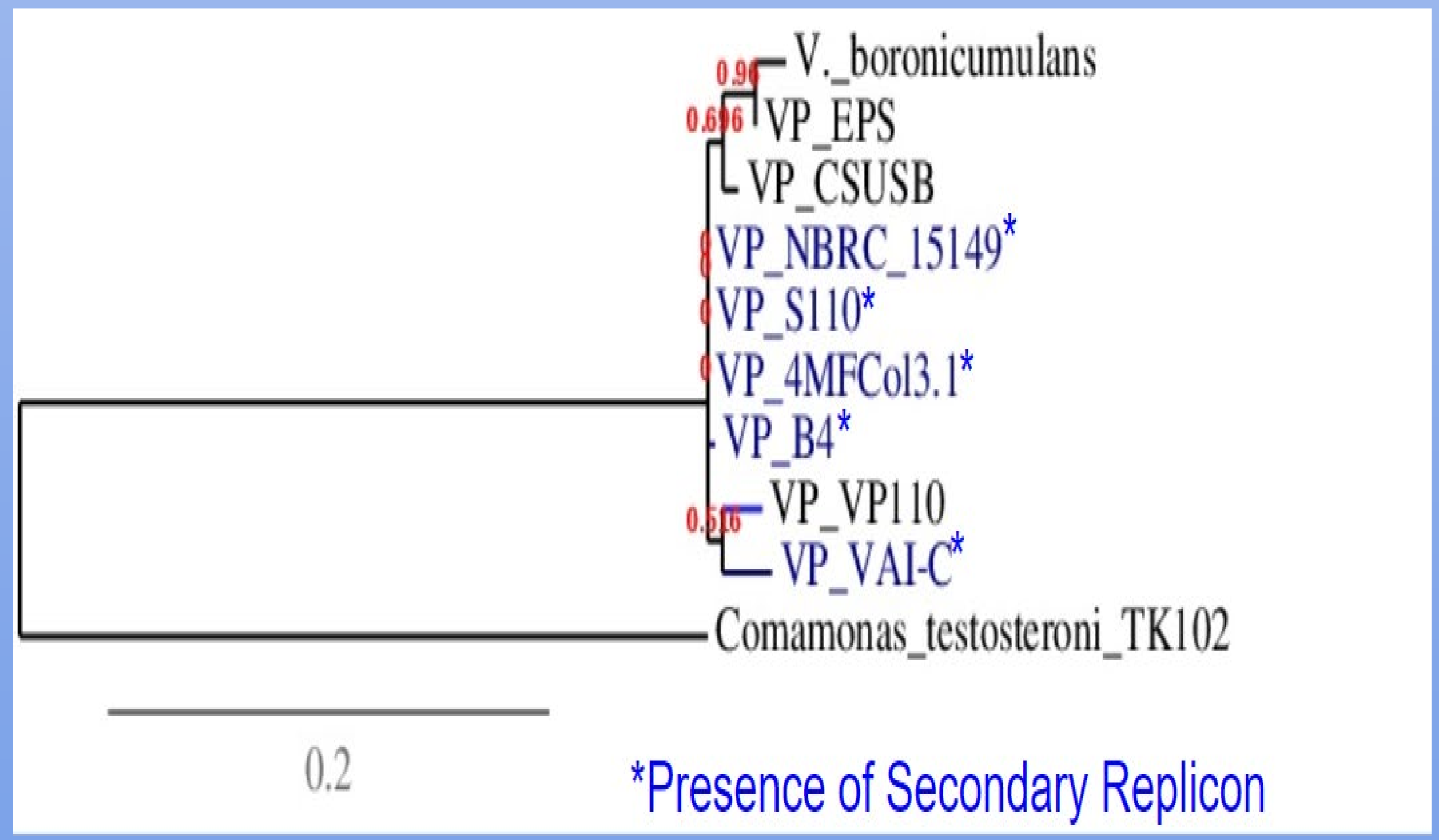
isDDH Results

Querygenome	Reference genome	DDH	Model C.I.	Distance	Prob. DDH >= 70%	G+C differ
VAI-C_Polished_Complete_Genome	4MFCol3.1	29.1	[26.7 - 31.6%]	0.147	0.07	1.39
VAI-C_Polished_Complete_Genome	B4	28.5	[26.1 - 31%]	0.1506	0.05	1.76
VAI-C_Polished_Complete_Genome	Comamonas_testosteroni_TK102	20.7	[18.5 - 23.1%]	0.2122	0	7.04
VAI-C_Polished_Complete_Genome	EPS	28.2	[25.8 - 30.7%]	0.1526	0.05	2.43
VAI-C_Polished_Complete_Genome	NBRC_15149	29.1	[26.7 - 31.6%]	0.1471	0.07	1.17
VAI-C_Polished_Complete_Genome	S110	29.3	[26.9 - 31.8%]	0.1461	0.08	1.38
VAI-C_Polished_Complete_Genome	Variovorax_boronicumulans_J1	29.4	[27 - 31.9%]	0.1453	0.08	0.9
VAI-C_Polished_Complete_Genome	VP_110	29	[26.7 - 31.5%]	0.1476	0.07	1.18
VAI-C_Polished_Complete_Genome	VP_EPS	28.2	[25.8 - 30.7%]	0.1526	0.05	2.43
VAI-C_Polished_Complete_Genome	VP_CSUSB_Final_Assembly	27.3	[24.9 - 29.8%]	0.1584	0.03	3.17
VAI-C_Polished_Complete_Genome	ZNC0006	65.7	[62.8 - 68.5%]	0.0423	69.14	0.55
VAI-C_Polished_PrimaryChromosome	4MFCol3.1	29.5	[27.1 - 32%]	0.1451	0.08	1.94
VAI-C_Polished_PrimaryChromosome	B4_PrimaryChromosome	29	[26.6 - 31.5%]	0.1476	0.07	2.25
VAI-C_Polished_PrimaryChromosome	Comamonas_testosteroni_TK102	20.7	[18.4 - 23.1%]	0.2125	0	7.59
VAI-C_Polished_PrimaryChromosome	EPS	28.4	[26 - 30.9%]	0.1516	0.05	2.98
VAI-C_Polished_PrimaryChromosome	NBRC_15149	29.5	[27.1 - 32%]	0.145	0.08	1.72
VAI-C_Polished_PrimaryChromosome	S110_PrimaryChromosome	29.8	[27.4 - 32.3%]	0.1435	0.09	1.83
VAI-C_Polished_PrimaryChromosome	Variovorax_boronicumulans_J1	29.7	[27.3 - 32.2%]	0.1439	0.09	1.45
VAI-C_Polished_PrimaryChromosome	VP_110	29.4	[27 - 31.9%]	0.1457	0.08	1.73
VAI-C_Polished_PrimaryChromosome	VP_EPS	28.4	[26 - 30.9%]	0.1516	0.05	2.98
VAI-C_Polished_PrimaryChromosome	ZNC0006	67.2	[64.2 - 70%]	0.0401	72.61	1.1
VAI-C_Polished_SecondaryChromosom	ZNC0006	56.5	[53.7 - 59.2%]	0.058	40.19	0.17
VAI-C_Polished_SecondaryReplicon	B4_SecondaryChromosome	23.4	[21.1 - 25.9%]	0.1865	0	1.61
VAI-C_Polished_SecondaryReplicon	S110_SecondaryChromosome	24	[21.7 - 26.5%]	0.1818	0.01	1.48

The 16s data coupled with the dot plots supports the phylogenetic relationship between these species. While the isDDH & difference in G+C content suggest that there is a species level difference between the known *Variovorax* and VAI-C . The secondary replicon is not related to any known secondary replicons in *Variovorax* suggesting that this might be a second instance of second replicon acquisition.

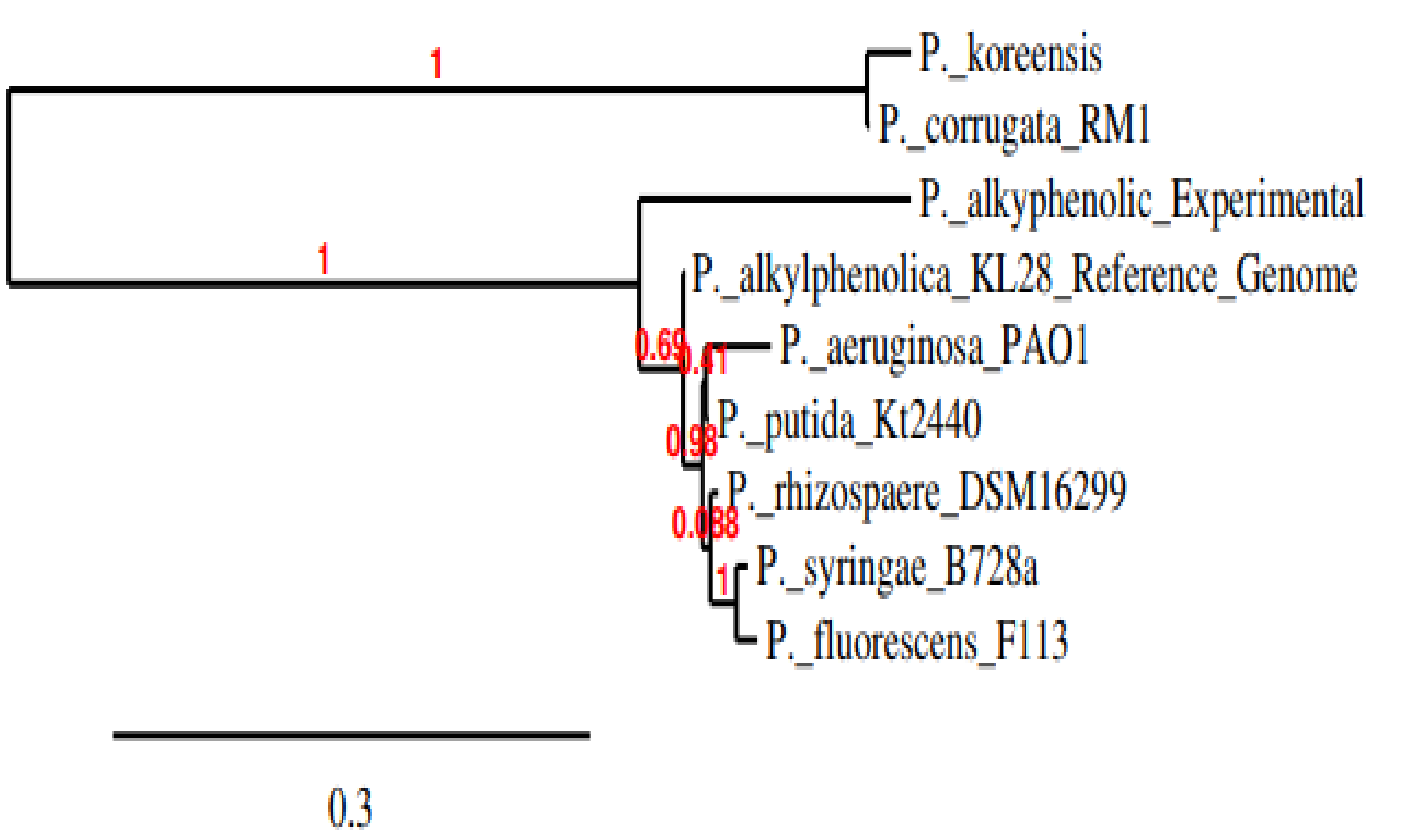
Phylogenetic Trees

Variovorax



*Presence of Secondary Replicon

Pseudomonas



Future Directions

- Sequence new and older organisms that have been left in the permanent draft state using ANI to see if they suggest a two replicon genome architecture
- Look in detail into the genes that are present on the secondary replicons to determine evolutionary history and origin, particularly into uptake and fixation into the population

Acknowledgements

The authors would like to thank Jared Leadbetter for the VAI-C isolate, Dylan Enright for the original culture of *Pseudomonas alkylphenolica*, and Ivan Hernandez for the *V. paradoxus* CSUSB culture. We would also like to thank Rast.org, Usegalaxy.com, LIRMM, GGDC.dsmz. I would also like to thank the entire Orwin Lab, Dr. Orwin, Dr. Nickerson and Monique Quinn.