Janelle Mande

Assemble your genome using ABySS and SPAdes.

1) From the ABySS output, create a table for the unitigs, contigs, and scaffolds with the number of each, N50 for each, and predicted genome length.

n	n:500	L50	min	N75	N50	N25	E-size	max	sum	name
623	184	27	526	39285	70306	100050	75774	220508	5488095	microbe-unitigs.fa
463	83	13	539	94509	159508	221038	165860	394829	5543362	microbe-contigs.fa
443	77	12	541	94728	160024	260956	189933	467386	5545050	microbe-scaffolds.fa

2) https://github.com/bcgsc/abyssLinks to an external site. This is the link to the documentation for ABySS. In your own words, please summarize the function of each of the commands (e.g., abyss-pe, k, B, etc) that you included in your code.

Abyss on paired end (pe) reads, name the files 'microbe', Kamer 96 base pair no bigger than 2G on sample genome files

abyss-pe name=microbe k=96 B=2G in='Reads1.fastq.gz Reads2.fastq.gz'

3) Using either output, perform a BLAST search to identify your species. Write your species name here: *Klebsiella pneumoniae*

BLAST Results: Klebsiella pneumoniae strain 31285 chromosome, complete genome

4) Perform quality assessment using QUAST. You need find a reference genome and reference annotation to upload to QUAST for the best quality check. Which assembler gave you the higher quality output? How do you know?

N50 is indicative of the assembly's quality; the higher the value, the more quality the assembly. SPAdes has the higher quality output because the N50 output is 193741; whereas ABySS had at most 160025.

5) Describe what BUSCO is used for. What were the BUSCO values for your assembly?

BUSCO is used to quantify the genome assembly completeness (%) and partial completeness (%). Within SPAdes, the complete BUSCO was 97.97%, while the ABySS complete BUSCO was 97.30%. On the other hand, both the SPAdes and ABySS partial BUSCO was 0.00%.

6) Perform a genome annotation using Prokka. Find 3 of the 5 genes/features in your results file and create a table of those results: recA, gyrA, 16S rRNA, rpsB, dnaA.

locus_tag	ftype	length_bp	gene	EC_numbe	COG	product
GCJMANH	CDS	1059	recA		COG0468	Protein RecA
GCJMANH	rRNA	448				16S ribosomal RNA (partial)
GCJMANH	CDS	726	rpsB		COG0052	30S ribosomal protein S2

7) https://github.com/tseemann/prokkaLinks to an external site. Here is the documentation for prokka. In your own words, what is the function of each of the commands in your line of code?

run prokka - output directory name of file 'prokkaanotation' –-look for prefix 'microbe' in the folder AbyssOutput, the name of the file is microbe-8.fa

prokka -outdir prokkaannotation --prefix microbe AbyssOutput/microbe-8.fa

8) What is the function of the genes/features you chose?

The recA gene expresses the recA protein, which is involved in the maintenance and repair of DNA.

16S rRNA is involved in the synthesis of the 16S Subunit of the ribosomal protein

The rpsB gene expresses the 30S ribosomal protein S2, which discriminates against aminoacyl tRNAs that do not match the codon of mRNA, ensuring accuracy during translation.

9) Find those same genes/features in your RAST annotation. What information did you learn about them from RAST?

RecA – To prevent the deleterious effects of recA overexpression, recX is expressed to regulate the protein.

Subsystem: RecA and RecX

Curator RossO

This subsystem's description is:

RecA and RecX are clearly closely functionally linked. This little subsystem is used to make sure that we get annotations for RecX correct

For more information, please check out the description and the additional notes tabs, below

Diagram | Functional Roles | Subsystem Spreadsheet | Description | Additional Notes

1: Proc Natl Acad Sci U S A. 2002 Sep 17;99(19):12091-6. Epub 2002 Sep 6.Click here to read
RecX protein abrogates ATP hydrolysis and strand exchange promoted by RecA: insights into negative regulation of homologous recombination.

Venkatesh R, Ganesh N, Guhan N, Reddy MS, Chandrasekhar T, Muniyappa K.

In many eubacteria, coexpression of recX with recA is essential for attenuation of the deleterious effects of recA overexpression; however, the molecular mechanism has remained enigmatic. Here, we show that Mycobacterium tuberculosis RecX binds directly to M. tuberculosis RecA as well as M. smegmatis and E. coli RecA proteins in vivo and in vitro, but not single-stranded DNA binding protein. The direct association of RecX with RecA failed to regulate the specificity or extent of binding of RecA either to DNA or ATP, ligands that are central to activation of its functions. Significantly, RecX severely impeded ATP hydrolysis and the generation of heteroduplex DNA promoted by homologous, as well as heterologous, RecA proteins. These findings reveal a mode of negative regulation of RecA, and imply that RecX might act as an anti-recombinase to quell inappropriate recombinational repair during normal DNA metabolism.

PMID: 12218174

Proteins. 2009 Feb 1;74(2):530-7.
Crystal structure of RecX: a potent regulatory protein of RecA from Xanthomonas campestris.
Yang CY, Chin KH, Yang MT, Wang AH, Chou SH.

PMID: 19003992

16S rRNA – there's an enzyme responsible for transferring a methyl group in the presence of cytosine.

RsmB	16S rRNA (cytosine(967)-C(5))-methyltransferase (EC 2.1.1.176)	-	-	-	none

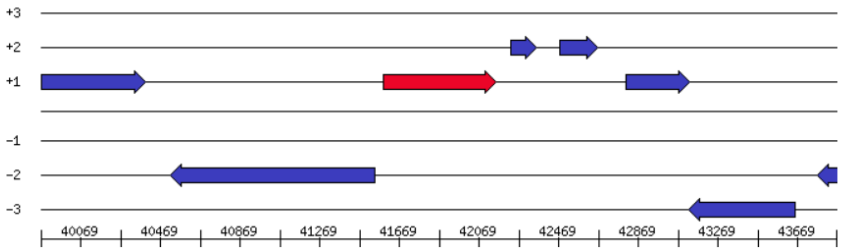
rpsB – From the Prokka annotation, we saw that the rpsB gene is 726 bp long; however, on RAST, I was unable to find a corresponding gene. However, I did find a gene that was <726 that is involved in Translation such as rpsB is.

Browse Genome: Klebsiella pneumoniae (573.68517)

Location | Focus | Upload List

ID	fig|573.68517.peg.329
Contig	2162
Type	CDS
Function	Translation elongation factor P
Subsystem	Translation elongation factors bacterial
Start	41587
Stop	42153
Length	567

zoom to sequence | details page | evidence page



export table | clear all filters

display 15 items per page

displaying 1 - 1 of 1

Feature ID	Type	Contig	Start	Stop	Length (bp)	Function	Subsystems	Region
	CDS	2162	<	<	< 726	translation		
fig|573.68517.peg.329	CDS	2162	41587	42153	567	Translation elongation factor P	Translation elongation factors bacterial	show

displaying 1 - 1 of 1

10) Upload the folder of this information to your GitHub in your Bioinfomatics Repository. Please share the link to your repository.

https://github.com/jmande1/bioinformatics