SUMMER INTERNSHIP PROJECT REPORT

submitted by

Name: **S Varsha**

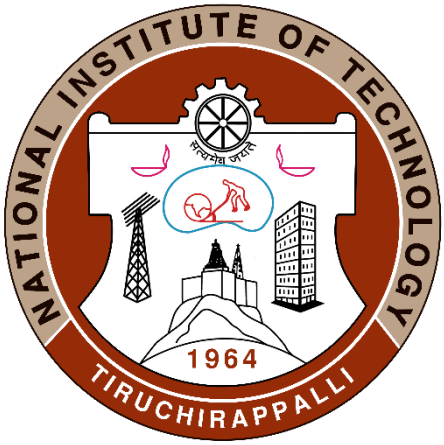
Roll no: **102118064**

in fulfilment of summer internship for the award of the degree of

BACHELOR OF TECHNOLOGY

in

CHEMICAL ENGINEERING



National Institute of Technology

Tiruchirappalli, Tamil Nadu

Contents

Acknowledgements

I want to thank Dr. Guhan Jayaraman, Professor at the Department of Biotechnology, Indian Institute of Technology, Madras, for giving me the opportunity to do an online internship with his team during the pandemic.

I would also like to thank Dr. Anantha Barathi Muthukrishnan, Principal Scientist, Department of Biotechnology, Indian Institute of Technology, Madras, for introducing me to the subject and patiently helping me do a background study to understand the research.

I would also like to thank Mr. Prashant Singh, Research Scholar, Department of Biotechnology, Indian Institute of Technology, Madras, for patiently helping me understand the steps involved in the process and guiding me through my work without any hassle.

Organisation Profile

The Department of Biotechnology at IIT Madras was founded in 2004 with a vision to be recognised as a department of international reputation with a strong interdisciplinary research and teaching base in biological sciences and engineering and active collaboration with industries and healthcare institutions. The department is housed in the 'Bhupat and Jyoti Mehta School of Biosciences'.

Modern biotechnology finds applications in a variety of domains, such as in the manufacture of high-technology pharmaceuticals and the making of enzymes for the detergents we use for washing our clothes. With expanding knowledge of the cell and its functioning at the molecular level, biotechnologists are now seeking answers to prevent life-threatening diseases, meet the energy and food requirements of an ever-growing population, and solve the environmental problems that threaten our very existence on this planet. The diversity of challenges that biotechnology tackle is reflected in the research activities of the faculty of Biotechnology at IIT Madras. Although there are many different research groups in the department, their research generally falls into one or more of the following categories: medical biotechnology related to cancer and cardiovascular aspects, chemical biology, computational biology and bioprocesses.

It carries out collaborative research with other departments within the institute and other institutes in India and abroad. In the past few years, their research contributions through journal publications have been significant, and they have also filed for several patents.

Background of Study

It is widely known that the population of the world multiplies exponentially with every passing generation. This results in the increased requirement of technology and power, which demands massive quantities of fuel. The primary fuels used are products of petroleum, which are non-renewable sources of energy. They also produce carbon dioxide after combustion, which results in what is known as the greenhouse effect. This has started to burn a hole in the ozone layer in the stratosphere and cause alarming climate change.

This calls for the need for renewable sources of energy, which also need to have little to no negative impact on the atmosphere. There are various options like solar power, wind power, hydroelectric power and tidal power. But it is important to study further and find more sources of such power in unlikely places and learn to harness and exploit these resources.

Acinetobacter is a group of bacteria commonly found in the environment, like in soil and water. It is a genus of gram-negative bacteria that are oxidase-negative, exhibit twitching motility and occurs in pairs under magnification. They are important soil organisms that contribute to the mineralisation of aromatic compounds.

Most strains of *Acinetobacter* grow well on MacConkey agar without salt. Bacteria of the genus *Acinetobacter* are known to form intracellular inclusions of polyhydroxyalkanoates under certain environmental conditions (e.g. lack of elements such as phosphorus, nitrogen, or oxygen combined with an excessive supply of carbon sources).

Different species of bacteria in this genus can be identified using fluorescence-lactose-denitrification to find the amount of acid produced by the metabolism of glucose. The other reliable identification test at the genus level is the chromosomal DNA transformation assay. In this assay, a naturally competent tryptophan auxotrophic mutant of *Acinetobacter baylyi* (BD4 trpE27) is transformed with the total DNA of a putative *Acinetobacter* isolate, and the transformation mixture is plated on a brain heart infusion agar. The growth is then harvested after incubation for 24 h at 30 °C, plating on an *Acinetobacter* minimal agar (AMA), and incubating at 30 °C for 108 h. Growth on the AMA indicates a positive transformation assay.

This *A. baylyi* is modified through a process called adaptive laboratory evolution by growing on a mixture of monomers derived from lignin. Adaptive laboratory evolution (ALE) is a frequent method in biological studies to gain insights into the basic mechanisms of molecular evolution and adaptive changes that accumulate in microbial populations during long term selection under specified growth conditions. It is used to break down lignin to produce an eco-friendly source of energy. This study involves looking for the mutations that may have a role to play in the evolved strain and uptake of the lignin monomers. This can be validated by performing experiments in the laboratory.

Whole-genome sequencing (WGS) is a comprehensive method for analyzing entire genomes. Genomic information has been instrumental in identifying inherited disorders, characterizing the mutations that drive cancer progression, and tracking disease outbreaks. This method is used to analyse mutations and conclude the effects of said mutations. In this study, whole

genome sequence analysis has been done to identify the mutations undergone by the strain through the process of ALE. The various steps performed in this analysis can be broadly listed as follows:

- Genome assembly
- Ordering and viewing of the assembled contigs
- Genome annotation

These mutations are then studied to gauge the effects they produce in the strain. These effects can be validated by testing the strains in the laboratory.

Work Done

Whole-genome sequence analysis is the process of determining the entirety, or nearly the entirety, of the DNA sequence of the genome of an organism at a single time. This entails sequencing all of the organism's chromosomal and cellular DNA.

Bacterial whole-genome sequencing (WGS) is becoming a widely-used technique in research, clinical diagnostic, and public health laboratories. It enables high-resolution characterisation of bacterial pathogens in terms of properties that include antibiotic resistance, molecular epidemiology, and virulence.

This computational analysis involved the whole genome sequence analysis of *Acinetobacter baylyi,* and it was performed on https://usegalaxy.org/. The process was done in a few steps, which have been put together below.

1. *Reference Genome*

   Downloaded the reference genome – Acinetobacter sp. ADP1 (GenBank: CR 543861.1)

2. *FASTQ Summary Statistics*

   Found the quality statistics of the wild type and evolved strain using FASTQ Summary Statistics tool [1] and obtained a tabular file containing the following as output:

   - column = column number (1 to 36 for a 36-cycles read Solexa file)
   - count = number of bases found in this column.
   - min = Lowest quality score value found in this column.
   - max = Highest quality score value found in this column.
   - sum = Sum of quality score values for this column.
   - mean = Mean quality score value for this column.
   - Q1 = 1st quartile quality score.
   - med = Median quality score.
   - Q3 = 3rd quartile quality score.
   - IQR = Inter-Quartile range (Q3-Q1).
   - lW = 'Left-Whisker' value (for boxplotting).
   - rW = 'Right-Whisker' value (for boxplotting).
   - outliers = Scores falling beyond the left and right whiskers (comma separated list).
   - A_Count = Count of 'A' nucleotides found in this column.
   - C_Count = Count of 'C' nucleotides found in this column.
   - G_Count = Count of 'G' nucleotides found in this column.
   - T_Count = Count of 'T' nucleotides found in this column.
   - N_Count = Count of 'N' nucleotides found in this column.
   - Other_Nucs = Comma separated list of other nucleotides found in this column.
   - Other_Count = Comma separated count of other nucleotides found in this column.

3. *Quality assessment of raw reads*

Performed quality assessment of raw reads using FASTQC tool [2] and obtained a basic text and a HTML file containing the following results:

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per base GC content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Kmer Content

All except the basic statistics and overrepresented sequences were plots.

4. *Trimming of raw reads*

Trimmed the raw reads using Trimmomatic flexible read trimming tool [3] for Illumina NGS data, based on the quality assessment to remove reads of low quality and adapter sequences and performed the following steps to obtain trimmed files:

- ILLUMINACLIP Step with TruSrq2 adapter sequences (Single-ended, for Illumina GAII)
- LEADING operation to keep bases of a minimum quality of 3
- TRAILING operation to keep bases of a minimum quality of 3
- SLIDINGWINDOW operation with average quality being 15 across 4 bases
- MINLEN operation to keep reads of minimum length 20

The following outputs were obtained:

- R1 and R2 paired reads of the wild type
- R1 and R2 unpaired reads of the wild type
- R1 and R2 paired reads of the evolved strain
- R1 and R2 unpaired reads of the evolved strain

5. *Mapping of genome to reference*

Mapped the sequence to reference genome after removing redundant data using Map with BWA tool. [4] [5] BWA is a software package for mapping low-divergent sequences against a large reference genome, such as the human genome. The bwa-aln algorithm is

designed for Illumina sequence reads up to 100bp. This tool wraps bwa-aln, bwa-samse and bwa-sampe modules of bwa read mapping tool:

- bwa aln - actual mapper placing reads onto the reference sequence
- bwa samse - post-processor converting suffix array co-ordinates into genome co-ordinates in SAM format for single reads
- bam sampe - post-processor for paired reads

The implementation takes fastq or BAM (unaligned BAM) datasets as input and produces output in BAM format, which can be further processed using various BAM utilities (BAMTools, SAMTools, Picard).

6. Sorting the BAM dataset

Sorted the BAM dataset using the SortSAM tool [6], which uses Picard documentation. Both input and output were in BAM format.

7. Marking and removing duplicate reads

Examined the aligned records in the BAM dataset given as input to locate duplicate molecules using the MarkDuplicates tool [6], which uses Picard documentation. All records are then written to the output file with duplicate records flagged. The outputs included the co-ordinate sorted BAM files and the metrics in a text file.

8. Variant Calling

Used FreeBayes,[7] a Bayesian genetic variant detector designed to find small polymorphisms, specifically single nucleotide polymorphisms (SNPs), insertions and deletions (indels), multiple nucleotide polymorphisms (MNPs) and complex events (composition insertion and substitution events) smaller than the length of a short-read sequencing alignment.

Provided some BAM dataset(s) and a reference sequence, FreeBayes will produce a VCF dataset describing SNPs, indels, and complex variants in samples in the input alignments.

By default, FreeBayes will consider variants supported by at least two observations in a single sample (-C) and also by at least 20% of the reads from a single sample (-F). These settings are suitable for low to high depth sequencing in haploid and diploid samples, but users working with polyploid or pooled samples may wish to adjust them depending on the characteristics of their sequencing data.

FreeBayes is capable of calling variant haplotypes shorter than a read length where multiple polymorphisms segregate on the same read. The maximum distance between polymorphisms phased in this way is determined by the --max-complex-gap, which defaults to 3bp. In practice, this can comfortably be set to half the read length.

Ploidy may be set to any level (-p), but by default, all samples are assumed to be diploid. FreeBayes can model per-sample and per-region variation in copy-number (-A) using a copy-number variation map.

FreeBayes can act as a frequency-based pooled caller and describe variants and haplotypes in terms of observation frequency rather than called genotypes. Allele observation counts will be described by AO and RO fields in the VCF output.

9. Building a database

Used the SnpEff build – genbank [8] command in the SnpEff build tool to create an SnpEff database. Output was in snpeffdb format.

10. Annotating variants

Annotated and predicted the effects of genetic variants using the SnpEff eff tool.[8]

A typical SnpEff use case would be:

- Input: The inputs are predicted variants (SNPs, insertions, deletions and MNPs). The input file is usually obtained as a result of a sequencing experiment, and it is usually in variant call format (VCF).
- Output: SnpEff analyses the input variants. It annotates the variants and calculates the effects they produce on known genes (e.g. amino acid changes). A list of effects and annotations that SnpEff can calculate can be found here.

In a typical sequencing experiment, there will be many places in the genome where your sample differs from the reference genome. These are called "genomic variants" or just "variants".

11. Compiling mutations

Manually checked the positions of the reference genome to find out the type of gene, the product it coded for, and the change that occurred using a genome browser tool. [9] The data was compiled in a spreadsheet.

Outputs

1. *FASTQ Summary Statistics*

#column	count	min	max	sum	mean	Q1	med	Q3	IQR	lW	rW	outliers	A_Count	C_Count	G_Count	T_Count	N_Count	other_bases	other_base_count
1	13712696	2	37	4.98E+08	36.31728	37	37	37	0	37	37	2,11,25	4334861	2719996	2336964	4320413	462		
2	13712696	2	37	4.96E+08	36.19089	37	37	37	0	37	37	2,11,25	4541143	2218276	2637831	4314839	607		
3	13712696	2	37	4.99E+08	36.36998	37	37	37	0	37	37	2,11,25	4448047	2481389	2511278	4270445	1537		
4	13712696	2	37	4.99E+08	36.39283	37	37	37	0	37	37	2,11,25	4206828	2626982	2665679	4213200	7		
5	13712696	11	37	5E+08	36.44283	37	37	37	0	37	37	11,25	4074397	2643747	2810982	4183570	0		
6	13712696	11	37	5E+08	36.46456	37	37	37	0	37	37	11,25	4081504	2790699	2800690	4039803	0		
7	13712696	11	37	4.99E+08	36.37169	37	37	37	0	37	37	11,25	4121926	2640062	2871671	4079037	0		
8	13712696	11	37	5E+08	36.42819	37	37	37	0	37	37	11,25	4112492	2722388	2778126	4099690	0		
9	13712696	11	37	4.99E+08	36.39369	37	37	37	0	37	37	11,25	4033670	2769365	2736562	4173099	0		
10	13712696	11	37	5E+08	36.44157	37	37	37	0	37	37	11,25	4116129	2738696	2740163	4117708	0		
11	13712696	11	37	5E+08	36.49084	37	37	37	0	37	37	11,25	4116111	2723874	2738538	4134173	0		
12	13712696	11	37	5E+08	36.44811	37	37	37	0	37	37	11,25	4124551	2744582	2720975	4122588	0		
13	13712696	11	37	5E+08	36.42867	37	37	37	0	37	37	11,25	4163528	2709971	2696840	4142357	0		
14	13712696	11	37	5E+08	36.44238	37	37	37	0	37	37	11,25	4157026	2701496	2722758	4131416	0		
15	13712696	11	37	5E+08	36.44634	37	37	37	0	37	37	11,25	4127569	2745424	2713185	4126518	0		
16	13712696	11	37	5E+08	36.42759	37	37	37	0	37	37	11,25	4143843	2717913	2707379	4143561	0		
17	13712696	11	37	5E+08	36.45301	37	37	37	0	37	37	11,25	4126293	2718984	2733829	4133590	0		
18	13712696	11	37	5E+08	36.42779	37	37	37	0	37	37	11,25	4099552	2766745	2727998	4118401	0		
19	13712696	11	37	4.99E+08	36.42479	37	37	37	0	37	37	11,25	4121105	2738222	2727270	4126099	0		
20	13712696	11	37	5E+08	36.42804	37	37	37	0	37	37	11,25	4111083	2732467	2756771	4112375	0		
21	13712696	11	37	5E+08	36.42971	37	37	37	0	37	37	11,25	4087331	2776148	2746481	4102736	0		
22	13712696	11	37	4.99E+08	36.42209	37	37	37	0	37	37	11,25	4118778	2751768	2730776	4111374	0		
23	13712696	11	37	4.99E+08	36.40719	37	37	37	0	37	37	11,25	4109418	2739252	2760955	4103071	0		
24	13712696	11	37	4.99E+08	36.40505	37	37	37	0	37	37	11,25	4081864	2779459	2758363	4093010	0		
25	13712696	11	37	4.99E+08	36.40533	37	37	37	0	37	37	11,25	4106750	2759028	2742245	4104673	0		
26	13712696	11	37	4.99E+08	36.38348	37	37	37	0	37	37	11,25	4095567	2750296	2767769	4099064	0		
27	13712696	11	37	4.99E+08	36.38491	37	37	37	0	37	37	11,25	4072720	2785262	2761323	4093391	0		
28	13712696	11	37	4.99E+08	36.37034	37	37	37	0	37	37	11,25	4102005	2762328	2750268	4098095	0		
29	13712696	11	37	4.99E+08	36.35681	37	37	37	0	37	37	11,25	4089621	2752819	2776152	4094104	0		
30	13712696	11	37	4.99E+08	36.37614	37	37	37	0	37	37	11,25	4069076	2788738	2771026	4083856	0		
31	13712696	11	37	4.98E+08	36.34038	37	37	37	0	37	37	11,25	4095307	2769170	2757253	4090966	0		
32	13712696	11	37	4.98E+08	36.33851	37	37	37	0	37	37	11,25	4087008	2759742	2781996	4083950	0		
33	13712696	11	37	4.98E+08	36.32083	37	37	37	0	37	37	11,25	4066469	2796340	2774588	4075299	0		
34	13712696	11	37	4.98E+08	36.34607	37	37	37	0	37	37	11,25	4093223	2772078	2763116	4084279	0		
35	13712696	11	37	4.98E+08	36.34402	37	37	37	0	37	37	11,25	4090632	2763661	2786941	4071462	0		
36	13712696	11	37	4.98E+08	36.34346	37	37	37	0	37	37	11,25	4066114	2799311	2776964	4070307	0		
37	13712696	11	37	4.98E+08	36.35165	37	37	37	0	37	37	11,25	4094361	2775737	2764267	4078331	0		
38	13712696	11	37	4.98E+08	36.32983	37	37	37	0	37	37	11,25	4089128	2762800	2785170	4075598	0		
39	13712696	11	37	4.98E+08	36.32732	37	37	37	0	37	37	11,25	4068434	2799738	2778075	4066449	0		
40	13712696	11	37	4.98E+08	36.34814	37	37	37	0	37	37	11,25	4088569	2778670	2766145	4079312	0		
41	13712696	11	37	4.98E+08	36.30365	37	37	37	0	37	37	11,25	4084585	2766983	2789183	4071945	0		
42	13712696	11	37	4.98E+08	36.32458	37	37	37	0	37	37	11,25	4059797	2802693	2783303	4066903	0		
43	13712696	11	37	4.98E+08	36.32835	37	37	37	0	37	37	11,25	4084479	2779423	2769138	4079656	0		
44	13712696	11	37	4.98E+08	36.29203	37	37	37	0	37	37	11,25	4079099	2769251	2793062	4071284	0		
45	13712696	11	37	4.98E+08	36.2909	37	37	37	0	37	37	11,25	4064032	2799159	2785098	4064407	0		
46	13712696	11	37	4.98E+08	36.28265	37	37	37	0	37	37	11,25	4087786	2778377	2768497	4078036	0		
47	13712696	11	37	4.98E+08	36.30317	37	37	37	0	37	37	11,25	4080095	2769178	2793637	4069786	0		
48	13712696	11	37	4.98E+08	36.28217	37	37	37	0	37	37	11,25	4060555	2804616	2787906	4059619	0		
49	13712696	11	37	4.98E+08	36.28496	37	37	37	0	37	37	11,25	4086389	2781313	2770609	4074385	0		
50	13712696	11	37	4.97E+08	36.26827	37	37	37	0	37	37	11,25	4082200	2773028	2793042	4064426	0		
51	13712696	11	37	4.97E+08	36.26567	37	37	37	0	37	37	11,25	4060305	2804566	2785204	4062621	0		
52	13712696	11	37	4.97E+08	36.27072	37	37	37	0	37	37	11,25	4089795	2779842	2770847	4072212	0		
53	13712696	11	37	4.97E+08	36.26994	37	37	37	0	37	37	11,25	4080765	2771508	2793752	4066671	0		
54	13712696	11	37	4.97E+08	36.25325	37	37	37	0	37	37	11,25	4062704	2802045	2787345	4060602	0		
55	13712696	11	37	4.97E+08	36.25709	37	37	37	0	37	37	11,25	4089169	2782495	2771752	4069280	0		
56	13712696	11	37	4.98E+08	36.29255	37	37	37	0	37	37	11,25	4080743	2773034	2792649	4066270	0		
57	13712696	11	37	4.97E+08	36.24859	37	37	37	0	37	37	11,25	4060841	2802277	2789616	4059962	0		
58	13712696	11	37	4.97E+08	36.22326	37	37	37	0	37	37	11,25	4087107	2780376	2773961	4071252	0		
59	13712696	11	37	4.97E+08	36.22833	37	37	37	0	37	37	11,25	4078018	2775288	2795282	4064108	0		
60	13712696	11	37	4.97E+08	36.23002	37	37	37	0	37	37	11,25	4061759	2805771	2786947	4058219	0		
61	13712696	11	37	4.97E+08	36.23216	37	37	37	0	37	37	11,25	4087024	2785265	2772017	4068390	0		
62	13712696	11	37	4.97E+08	36.21803	37	37	37	0	37	37	11,25	4083590	2772624	2793403	4063079	0		
63	13712696	11	37	4.96E+08	36.17339	37	37	37	0	37	37	11,25	4061598	2806695	2788439	4055964	0		
64	13712696	11	37	4.97E+08	36.21849	37	37	37	0	37	37	11,25	4085204	2786125	2771738	4069629	0		
65	13712696	11	37	4.97E+08	36.22676	37	37	37	0	37	37	11,25	4082902	2771610	2795158	4063026	0		
66	13712696	11	37	4.96E+08	36.20664	37	37	37	0	37	37	11,25	4062055	2804209	2786702	4059730	0		
67	13712696	11	37	4.97E+08	36.21256	37	37	37	0	37	37	11,25	4090248	2780596	2773835	4068017	0		
68	13712696	11	37	4.96E+08	36.16804	37	37	37	0	37	37	11,25	4080631	2774921	2792785	4064359	0		
69	13712696	11	37	4.96E+08	36.2033	37	37	37	0	37	37	11,25	4062779	2804112	2785630	4060175	0		
70	13712696	11	37	4.97E+08	36.21707	37	37	37	0	37	37	11,25	4086680	2784486	2770416	4071114	0		
71	13712696	11	37	4.96E+08	36.20623	37	37	37	0	37	37	11,25	4084078	2773943	2792707	4061968	0		
72	13712696	11	37	4.96E+08	36.16645	37	37	37	0	37	37	11,25	4058996	2805115	2787460	4061125	0		
73	13712696	11	37	4.96E+08	36.17802	37	37	37	0	37	37	11,25	4085921	2783097	2775540	4068138	0		
74	13712696	11	37	4.96E+08	36.1943	37	37	37	0	37	37	11,25	4079099	2774071	2795416	4064110	0		
75	13712696	11	37	4.96E+08	36.19133	37	37	37	0	37	37	11,25	4058894	2809742	2786120	4057940	0		
76	13712696	11	37	4.96E+08	36.19021	37	37	37	0	37	37	11,25	4084495	2785702	2770003	4072496	0		
77	13712696	11	37	4.96E+08	36.16532	37	37	37	0	37	37	11,25	4082639	2770800	2794124	4065133	0		
78	13712696	11	37	4.96E+08	36.17142	37	37	37	0	37	37	11,25	4063816	2804089	2787041	4057750	0		
79	13712696	11	37	4.96E+08	36.16883	37	37	37	0	37	37	11,25	4088675	2786472	2771451	4066098	0		
80	13712696	11	37	4.95E+08	36.10764	37	37	37	0	37	37	11,25	4085160	2772849	2791439	4063248	0		
81	13712696	11	37	4.95E+08	36.07591	37	37	37	0	37	37	11,25	4064702	2801595	2783779	4062620	0		
82	13712696	11	37	4.95E+08	36.07378	37	37	37	0	37	37	11,25	4088086	2783743	2770637	4070230	0		
83	13712696	11	37	4.95E+08	36.13165	37	37	37	0	37	37	11,25	4084073	2771188	2797562	4059873	0		
84	13712696	11	37	4.95E+08	36.13406	37	37	37	0	37	37	11,25	4058908	2807302	2786679	4059807	0		
85	13712696	11	37	4.95E+08	36.07694	37	37	37	0	37	37	11,25	4089549	2784651	2770957	4067539	0		
86	13712696	11	37	4.95E+08	36.10646	37	37	37	0	37	37	11,25	4083427	2771766	2794095	4063408	0		
87	13712696	2	37	4.95E+08	36.1175	37	37	37	0	37	37	2,11,25	4062470	2803921	2787436	4058865	4		
88	13712696	11	37	4.95E+08	36.09556	37	37	37	0	37	37	11,25	4089440	2783471	2773481	4066304	0		
89	13712696	11	37	4.95E+08	36.10175	37	37	37	0	37	37	11,25	4081245	2775197	2790519	4065735	0		
90	13712696	11	37	4.95E+08	36.06546	37	37	37	0	37	37	11,25	4064554	2803325	2784434	4060383	0		
91	13712696	11	37	4.95E+08	36.08541	37	37	37	0	37	37	11,25	4087636	2783654	2772253	4069153	0		
92	13712696	11	37	4.95E+08	36.09059	37	37	37	0	37	37	11,25	4081526	2774144	2794462	4062564	0		
93	13712696	11	37	4.94E+08	36.0182	37	37	37	0	37	37	11,25	4064511	2807088	2784730	4056367	0		
94	13712696	11	37	4.95E+08	36.07371	37	37	37	0	37	37	11,25	4088403	2783862	2772977	4067454	0		
95	13712696	11	37	4.95E+08	36.09075	37	37	37	0	37	37	11,25	4084729	2773109	2793946	4060912	0		
96	13712696	11	37	4.94E+08	36.03089	37	37	37	0	37	37	11,25	4063453	2804139	2784627	4060477	0		
97	13712696	11	37	4.94E+08	36.0169	37	37	37	0	37	37	11,25	4089496	2782599	2772077	4068524	0		
98	13712696	11	37	4.94E+08	36.00151	37	37	37	0	37	37	11,25	4079920	2772444	2794252	4066080	0		
99	13712696	11	37	4.94E+08	36.0363	37	37	37	0	37	37	11,25	4061960	2807263	2785056	4058417	0		
100	13712696	11	37	4.94E+08	36.0101	37	37	37	0	37	37	11,25	4087169	2783739	2773416	4068372	0		
101	13712696	11	37	4.94E+08	35.99663	37	37	37	0	37	37	11,25	4082274	2774475	2794196	4061751	0		

#column	count	min	max	sum	mean	Q1	med	Q3	IQR	lW	rW	outliers	A_Count	C_Count	G_Count	T_Count	N_Count	other_bases	other_base_count
1	13712696	2	37	4.78E+08	34.88032	37	37	37	0	37	37	2,11,25	4327526	2704307	2332035	4333343	15485		
2	13712696	2	37	4.77E+08	34.76612	37	37	37	0	37	37	2,11,25	4620088	2133611	2722317	4230688	5992		
3	13712696	2	37	4.81E+08	35.06586	37	37	37	0	37	37	2,11,25	4463674	2411276	2615241	4222491	14		
4	13712696	11	37	4.83E+08	35.19121	37	37	37	0	37	37	11,25	4164917	2598864	2771736	4177179	0		
5	13712696	11	37	4.84E+08	35.28636	37	37	37	0	37	37	11,25	4044945	2619915	2908593	4139243	0		
6	13712696	11	37	4.84E+08	35.30578	37	37	37	0	37	37	11,25	4050251	2784136	2904346	3973963	0		
7	13712696	11	37	4.83E+08	35.22262	37	37	37	0	37	37	11,25	4080477	2636169	2955168	4040882	0		
8	13712696	11	37	4.83E+08	35.22615	37	37	37	0	37	37	11,25	4099027	2691909	2856998	4064762	0		
9	13712696	11	37	4.83E+08	35.21868	37	37	37	0	37	37	11,25	4017050	2743239	2818114	4134293	0		
10	13712696	11	37	4.84E+08	35.28431	37	37	37	0	37	37	11,25	4103014	2707783	2813677	4088222	0		
11	13712696	11	37	4.84E+08	35.32254	37	37	37	0	37	37	11,25	4106600	2688211	2809774	4108111	0		
12	13712696	11	37	4.84E+08	35.32991	37	37	37	0	37	37	11,25	4108241	2718004	2793813	4092638	0		
13	13712696	11	37	4.84E+08	35.30304	37	37	37	0	37	37	11,25	4150627	2678962	2761644	4121463	0		
14	13712696	11	37	4.84E+08	35.28495	37	37	37	0	37	37	11,25	4150759	2663598	2788054	4110285	0		
15	13712696	11	37	4.83E+08	35.23449	37	37	37	0	37	37	11,25	4116318	2720075	2774760	4101543	0		
16	13712696	11	37	4.84E+08	35.26476	37	37	37	0	37	37	11,25	4131217	2689749	2765035	4126695	0		
17	13712696	11	37	4.83E+08	35.25716	37	37	37	0	37	37	11,25	4118199	2686055	2794073	4114369	0		
18	13712696	11	37	4.83E+08	35.25806	37	37	37	0	37	37	11,25	4085553	2737438	2788175	4101530	0		
19	13712696	11	37	4.82E+08	35.14295	37	37	37	0	37	37	11,25	4117932	2704792	2778690	4111282	0		
20	13712696	2	37	4.83E+08	35.22849	37	37	37	0	37	37	2,11,25	4111674	2694348	2812140	4094503	31		
21	13712696	11	37	4.83E+08	35.24118	37	37	37	0	37	37	11,25	4083539	2741641	2810532	4076984	0		
22	13712696	11	37	4.83E+08	35.19014	37	37	37	0	37	37	11,25	4112538	2722362	2779014	4098782	0		
23	13712696	11	37	4.83E+08	35.20218	37	37	37	0	37	37	11,25	4109961	2702127	2813668	4086940	0		
24	13712696	11	37	4.82E+08	35.14205	37	37	37	0	37	37	11,25	4086229	2745548	2815403	4065516	0		
25	13712696	11	37	4.82E+08	35.18521	37	37	37	0	37	37	11,25	4109814	2725050	2795828	4082004	0		
26	13712696	11	37	4.82E+08	35.11746	37	37	37	0	37	37	11,25	4098994	2713967	2822555	4077180	0		
27	13712696	11	37	4.82E+08	35.11418	37	37	37	0	37	37	11,25	4075717	2755334	2820601	4061044	0		
28	13712696	2	37	4.81E+08	35.09881	37	37	37	0	37	37	2,11,25	4103236	2734279	2797228	4077868	85		
29	13712696	11	37	4.81E+08	35.07686	37	37	37	0	37	37	11,25	4098215	2718646	2830450	4065385	0		
30	13712696	11	37	4.81E+08	35.09978	37	37	37	0	37	37	11,25	4071571	2762398	2827035	4051692	0		
31	13712696	11	37	4.81E+08	35.05525	37	37	37	0	37	37	11,25	4097834	2743948	2809165	4061749	0		
32	13712696	11	37	4.8E+08	35.02537	37	37	37	0	37	37	11,25	4098237	2726703	2834230	4053526	0		
33	13712696	11	37	4.8E+08	34.98056	37	37	37	0	37	37	11,25	4072759	2768368	2832162	4039407	0		
34	13712696	11	37	4.8E+08	35.02089	37	37	37	0	37	37	11,25	4103734	2743192	2804418	4061352	0		
35	13712696	11	37	4.81E+08	35.0538	37	37	37	0	37	37	11,25	4099473	2731062	2836212	4045949	0		
36	13712696	11	37	4.81E+08	35.06818	37	37	37	0	37	37	11,25	4076509	2769440	2832745	4034002	0		
37	13712696	11	37	4.81E+08	35.04127	37	37	37	0	37	37	11,25	4103527	2746246	2807880	4055043	0		
38	13712696	2	37	4.8E+08	34.97932	37	37	37	0	37	37	2,11,25	4094446	2735813	2835128	4046987	322		
39	13712696	11	37	4.79E+08	34.96345	37	37	37	0	37	37	11,25	4073735	2773807	2831635	4033519	0		
40	13712696	11	37	4.8E+08	35.02246	37	37	37	0	37	37	11,25	4102993	2749038	2807249	4053416	0		
41	13712696	11	37	4.79E+08	34.94893	37	37	37	0	37	37	11,25	4088052	2736683	2839535	4048426	0		
42	13712696	11	37	4.79E+08	34.95226	37	37	37	0	37	37	11,25	4069424	2774614	2839150	4029508	0		
43	13712696	11	37	4.8E+08	35.00188	37	37	37	0	37	37	11,25	4098820	2753747	2810967	4049162	0		
44	13712696	11	37	4.79E+08	34.91069	37	37	37	0	37	37	11,25	4096386	2734406	2835341	4046563	0		
45	13712696	11	37	4.79E+08	34.92247	37	37	37	0	37	37	11,25	4073343	2772032	2834785	4032536	0		
46	13712696	11	37	4.79E+08	34.94176	37	37	37	0	37	37	11,25	4096698	2752578	2808567	4054853	0		
47	13712696	11	37	4.79E+08	34.90024	37	37	37	0	37	37	11,25	4092763	2739724	2839497	4040712	0		
48	13712696	11	37	4.79E+08	34.91417	37	37	37	0	37	37	11,25	4071928	2778099	2832564	4030105	0		
49	13712696	11	37	4.79E+08	34.90825	37	37	37	0	37	37	11,25	4102271	2754421	2809923	4046081	0		
50	13712696	11	37	4.78E+08	34.86455	37	37	37	0	37	37	11,25	4098137	2738265	2832880	4043414	0		
51	13712696	2	37	4.78E+08	34.85432	37	37	37	0	37	37	2,11,25	4074688	2778101	2832272	4027614	21		
52	13712696	11	37	4.78E+08	34.82732	37	37	37	0	37	37	11,25	4102025	2755965	2808327	4046379	0		
53	13712696	11	37	4.78E+08	34.87626	37	37	37	0	37	37	11,25	4095887	2738116	2838683	4040010	0		
54	13712696	11	37	4.77E+08	34.80667	37	37	37	0	37	37	11,25	4074763	2775759	2833717	4028457	0		
55	13712696	11	37	4.77E+08	34.76744	37	37	37	0	37	37	11,25	4104293	2752065	2809537	4046801	0		
56	13712696	11	37	4.76E+08	34.71511	37	37	37	0	37	37	11,25	4096453	2736757	2837641	4041845	0		
57	13712696	11	37	4.77E+08	34.76605	37	37	37	0	37	37	11,25	4073367	2776839	2831266	4031224	0		
58	13712696	11	37	4.76E+08	34.74565	37	37	37	0	37	37	11,25	4107454	2751021	2807182	4047039	0		
59	13712696	2	37	4.77E+08	34.76139	37	37	37	0	37	37	2,11,25	4101454	2737788	2836939	4036495	20		
60	13712696	11	37	4.77E+08	34.78313	37	37	37	0	37	37	11,25	4071410	2778995	2834842	4027449	0		
61	13712696	11	37	4.76E+08	34.73416	37	37	37	0	37	37	11,25	4105094	2754263	2804668	4048671	0		
62	13712696	11	37	4.77E+08	34.76756	37	37	37	0	37	37	11,25	4098430	2741071	2833032	4040163	0		
63	13712696	11	37	4.76E+08	34.69583	37	37	37	0	37	37	11,25	4077548	2777771	2828746	4028631	0		
64	13712696	11	37	4.75E+08	34.66546	37	37	37	0	37	37	11,25	4100851	2755937	2810532	4045376	0		
65	13712696	11	37	4.76E+08	34.68046	37	37	37	0	37	37	11,25	4099147	2738773	2833633	4041143	0		
66	13712696	11	37	4.76E+08	34.69049	37	37	37	0	37	37	11,25	4076879	2775598	2829289	4030930	0		
67	13712696	11	37	4.76E+08	34.6859	37	37	37	0	37	37	11,25	4104939	2751990	2806714	4049053	0		
68	13712696	11	37	4.75E+08	34.64206	37	37	37	0	37	37	11,25	4102315	2735283	2831913	4043185	0		
69	13712696	11	37	4.75E+08	34.65422	37	37	37	0	37	37	11,25	4083700	2771735	2826827	4030434	0		
70	13712696	11	37	4.75E+08	34.6203	37	37	37	0	37	37	11,25	4107503	2753320	2805344	4046529	0		
71	13712696	11	37	4.75E+08	34.62925	37	37	37	0	37	37	11,25	4096567	2740613	2832289	4043227	0		
72	13712696	11	37	4.75E+08	34.67041	37	37	37	0	37	37	11,25	4074586	2778226	2830842	4029042	0		
73	13712696	11	37	4.75E+08	34.637	37	37	37	0	37	37	11,25	4102088	2755655	2808481	4046472	0		
74	13712696	11	37	4.74E+08	34.58817	37	37	37	0	37	37	11,25	4100354	2740036	2830718	4041588	0		
75	13712696	11	37	4.74E+08	34.56754	37	37	37	0	37	37	11,25	4081358	2773285	2828486	4029567	0		
76	13712696	11	37	4.74E+08	34.58641	37	37	37	0	37	37	11,25	4107015	2755545	2802567	4047569	0		
77	13712696	11	37	4.75E+08	34.62706	37	37	37	0	37	37	11,25	4100062	2736026	2829587	4047021	0		
78	13712696	11	37	4.74E+08	34.56019	37	37	37	0	37	37	11,25	4081203	2774344	2827384	4029765	0		
79	13712696	11	37	4.73E+08	34.5209	37	37	37	0	37	37	11,25	4108097	2752548	2804140	4047911	0		
80	13712696	2	37	4.73E+08	34.4692	37	37	37	0	37	37	2,11,25	4105713	2738133	2828665	4040177	8		
81	13712696	11	37	4.72E+08	34.4531	37	37	37	0	37	37	11,25	4080565	2775008	2824509	4032614	0		
82	13712696	11	37	4.74E+08	34.53383	37	37	37	0	37	37	11,25	4109728	2750885	2797851	4054232	0		
83	13712696	11	37	4.73E+08	34.46655	37	37	37	0	37	37	11,25	4104402	2735485	2833933	4038876	0		
84	13712696	11	37	4.72E+08	34.42137	37	37	37	0	37	37	11,25	4080246	2774304	2829889	4028257	0		
85	13712696	11	37	4.73E+08	34.47098	37	37	37	0	37	37	11,25	4109187	2754764	2804864	4043881	0		
86	13712696	11	37	4.72E+08	34.42913	37	37	37	0	37	37	11,25	4108442	2736927	2828795	4038532	0		
87	13712696	2	37	4.73E+08	34.47405	37	37	37	0	37	37	2,11,25	4081507	2776909	2825581	4028642	57		
88	13712696	11	37	4.74E+08	34.53854	37	37	37	0	37	37	11,25	4106106	2752480	2803986	4050124	0		
89	13712696	11	37	4.73E+08	34.46659	37	37	37	0	37	37	11,25	4106617	2734599	2831928	4039552	0		
90	13712696	11	37	4.71E+08	34.38094	37	37	37	0	37	37	11,25	4084661	2774182	2824543	4029310	0		
91	13712696	11	37	4.71E+08	34.33228	37	37	37	0	37	37	11,25	4112619	2751653	2802118	4046306	0		
92	13712696	11	37	4.72E+08	34.39884	37	37	37	0	37	37	11,25	4103161	2738438	2828087	4043010	0		
93	13712696	2	37	4.71E+08	34.37928	37	37	37	0	37	37	2,11,25	4083330	2774029	2822850	4032479	8		
94	13712696	11	37	4.7E+08	34.2575	37	37	37	0	37	37	11,25	4111865	2752787	2802759	4045285	0		
95	13712696	11	37	4.69E+08	34.22161	37	37	37	0	37	37	11,25	4103794	2736766	2830683	4041453	0		
96	13712696	11	37	4.71E+08	34.32647	37	37	37	0	37	37	11,25	4082424	2771283	2827889	4031100	0		
97	13712696	11	37	4.72E+08	34.40615	37	37	37	0	37	37	11,25	4109812	2749070	2805538	4048276	0		
98	13712696	11	37	4.71E+08	34.34929	37	37	37	0	37	37	11,25	4101845	2738492	2829070	4043289	0		
99	13712696	11	37	4.7E+08	34.28225	37	37	37	0	37	37	11,25	4081657	2775901	2827215	4027923	0		
100	13712696	11	37	4.7E+08	34.29488	37	37	37	0	37	37	11,25	4114950	2749022	2799555	4049169	0		
101	13712696	2	37	4.71E+08	34.36198	37	37	37	0	37	37	2,11,25	4104496	2736809	2831692	4039661	38		

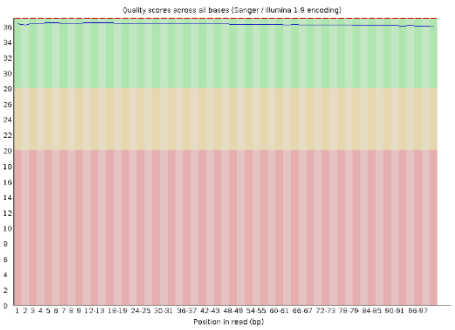
Thus, the quality characteristics of the evolved strain have been compiled.

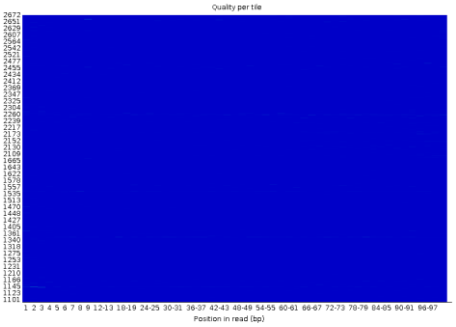
#column	count	min	max	sum	mean	Q1	med	Q3	IQR	lW	rW	outliers	A_Count	C_Count	G_Count	T_Count	N_Count	other_bases	other_base_count
1	15251046	2	37	5.41E+08	35.48442	37	37	37	0	37	37	2,11,25	4777058	2993809	2905471	4574183	525		
2	15251046	2	37	5.48E+08	35.90594	37	37	37	0	37	37	2,11,25	5004967	2516008	2898394	4813587	18090		
3	15251046	2	37	5.52E+08	36.20007	37	37	37	0	37	37	2,11,25	4820442	2775250	2797657	4857669	28		
4	15251046	11	37	5.54E+08	36.31029	37	37	37	0	37	37	11,25	4593562	2909478	3017520	4730486	0		
5	15251046	2	37	5.55E+08	36.40696	37	37	37	0	37	37	2,11,25	4539051	2952859	3160254	4598880	2		
6	15251046	11	37	5.56E+08	36.43477	37	37	37	0	37	37	11,25	4520503	3125605	3152568	4452370	0		
7	15251046	11	37	5.56E+08	36.43455	37	37	37	0	37	37	11,25	4537368	3014138	3177217	4522323	0		
8	15251046	11	37	5.56E+08	36.44148	37	37	37	0	37	37	11,25	4523366	3093799	3079755	4554126	0		
9	15251046	11	37	5.56E+08	36.44065	37	37	37	0	37	37	11,25	4453265	3114696	3064323	4618762	0		
10	15251046	11	37	5.56E+08	36.4469	37	37	37	0	37	37	11,25	4558511	3061572	3064143	4566820	0		
11	15251046	11	37	5.56E+08	36.46819	37	37	37	0	37	37	11,25	4560600	3036080	3055115	4599251	0		
12	15251046	11	37	5.56E+08	36.47541	37	37	37	0	37	37	11,25	4548828	3068085	3034191	4599942	0		
13	15251046	11	37	5.56E+08	36.46132	37	37	37	0	37	37	11,25	4587576	3032081	3018689	4612700	0		
14	15251046	11	37	5.56E+08	36.45678	37	37	37	0	37	37	11,25	4575374	3030259	3031424	4613989	0		
15	15251046	11	37	5.56E+08	36.44848	37	37	37	0	37	37	11,25	4546407	3074578	3024128	4605933	0		
16	15251046	11	37	5.56E+08	36.44907	37	37	37	0	37	37	11,25	4568912	3040122	3030406	4611606	0		
17	15251046	2	37	5.56E+08	36.46089	37	37	37	0	37	37	2,11,25	4560637	3039311	3045306	4605780	12		
18	15251046	11	37	5.56E+08	36.45963	37	37	37	0	37	37	11,25	4538308	3080053	3037851	4594834	0		
19	15251046	11	37	5.56E+08	36.45311	37	37	37	0	37	37	11,25	4568709	3045972	3042962	4593403	0		
20	15251046	11	37	5.56E+08	36.45394	37	37	37	0	37	37	11,25	4562187	3035419	3064656	4588784	0		
21	15251046	11	37	5.56E+08	36.44933	37	37	37	0	37	37	11,25	4540025	3076327	3059369	4575325	0		
22	15251046	11	37	5.56E+08	36.43303	37	37	37	0	37	37	11,25	4566015	3057550	3048289	4579192	0		
23	15251046	11	37	5.56E+08	36.4283	37	37	37	0	37	37	11,25	4556463	3048458	3071548	4574577	0		
24	15251046	11	37	5.56E+08	36.42757	37	37	37	0	37	37	11,25	4522623	3091054	3074643	4562726	0		
25	15251046	11	37	5.55E+08	36.41759	37	37	37	0	37	37	11,25	4545859	3073180	3062739	4569268	0		
26	15251046	11	37	5.55E+08	36.39802	37	37	37	0	37	37	11,25	4539780	3066401	3082740	4562125	0		
27	15251046	11	37	5.55E+08	36.39028	37	37	37	0	37	37	11,25	4509135	3100679	3082540	4558692	0		
28	15251046	11	37	5.55E+08	36.38156	37	37	37	0	37	37	11,25	4535582	3081474	3078151	4555839	0		
29	15251046	11	37	5.55E+08	36.38596	37	37	37	0	37	37	11,25	4529426	3068627	3097717	4555276	0		
30	15251046	11	37	5.55E+08	36.38284	37	37	37	0	37	37	11,25	4507061	3108634	3088200	4547151	0		
31	15251046	2	37	5.55E+08	36.3639	37	37	37	0	37	37	2,11,25	4534267	3088559	3084114	4544105	1		
32	15251046	11	37	5.55E+08	36.37126	37	37	37	0	37	37	11,25	4531059	3071760	3104678	4543549	0		
33	15251046	2	37	5.55E+08	36.37091	37	37	37	0	37	37	2,11,25	4500811	3115265	3099633	4535317	20		
34	15251046	11	37	5.55E+08	36.36997	37	37	37	0	37	37	11,25	4530975	3092721	3091649	4535701	0		
35	15251046	11	37	5.54E+08	36.34873	37	37	37	0	37	37	11,25	4526982	3082707	3106181	4535176	0		
36	15251046	11	37	5.54E+08	36.32999	37	37	37	0	37	37	11,25	4502951	3118574	3097842	4531679	0		
37	15251046	2	37	5.54E+08	36.34403	37	37	37	0	37	37	2,11,25	4533745	3092854	3094079	4530349	19		
38	15251046	2	37	5.54E+08	36.33422	37	37	37	0	37	37	2,11,25	4524635	3084990	3112092	4529319	10		
39	15251046	11	37	5.54E+08	36.32853	37	37	37	0	37	37	11,25	4501895	3119177	3102552	4527422	0		
40	15251046	2	37	5.54E+08	36.32941	37	37	37	0	37	37	2,11,25	4534150	3096160	3096681	4523629	426		
41	15251046	2	37	5.54E+08	36.32134	37	37	37	0	37	37	2,11,25	4519237	3087209	3115166	4529268	166		
42	15251046	2	37	5.54E+08	36.30701	37	37	37	0	37	37	2,11,25	4492844	3126213	3108747	4522330	912		
43	15251046	2	37	5.54E+08	36.3119	37	37	37	0	37	37	2,11,25	4522780	3101553	3098540	4528126	47		
44	15251046	2	37	5.54E+08	36.30276	37	37	37	0	37	37	2,11,25	4519291	3090112	3115311	4526320	12		
45	15251046	2	37	5.54E+08	36.29733	37	37	37	0	37	37	2,11,25	4490363	3125689	3107485	4527495	14		
46	15251046	2	37	5.54E+08	36.30919	37	37	37	0	37	37	2,11,25	4523672	3105444	3098562	4523236	132		
47	15251046	2	37	5.54E+08	36.30978	37	37	37	0	37	37	2,11,25	4517565	3089757	3118286	4525322	116		
48	15251046	2	37	5.53E+08	36.28773	37	37	37	0	37	37	2,11,25	4495557	3122962	3112041	4520326	160		
49	15251046	2	37	5.53E+08	36.28594	37	37	37	0	37	37	2,11,25	4526160	3104118	3099083	4521671	14		
50	15251046	2	37	5.53E+08	36.28618	37	37	37	0	37	37	2,11,25	4522294	3088997	3117244	4522467	44		
51	15251046	2	37	5.53E+08	36.28049	37	37	37	0	37	37	2,11,25	4494948	3125029	3113423	4517612	34		
52	15251046	2	37	5.53E+08	36.26683	37	37	37	0	37	37	2,11,25	4523896	3108450	3100892	4517709	99		
53	15251046	11	37	5.53E+08	36.24726	37	37	37	0	37	37	11,25	4518872	3092764	3115116	4524294	0		
54	15251046	11	37	5.53E+08	36.25355	37	37	37	0	37	37	11,25	4493430	3124531	3116740	4516345	0		
55	15251046	11	37	5.53E+08	36.25672	37	37	37	0	37	37	11,25	4522048	3104193	3102272	4522533	0		
56	15251046	2	37	5.53E+08	36.24486	37	37	37	0	37	37	2,11,25	4518901	3095022	3114971	4522150	2		
57	15251046	11	37	5.53E+08	36.23938	37	37	37	0	37	37	11,25	4495404	3123830	3110040	4521772	0		
58	15251046	11	37	5.52E+08	36.2206	37	37	37	0	37	37	11,25	4520614	3102144	3099360	4528928	0		
59	15251046	2	37	5.52E+08	36.20355	37	37	37	0	37	37	2,11,25	4518054	3096637	3118006	4518346	3		
60	15251046	11	37	5.52E+08	36.20078	37	37	37	0	37	37	11,25	4496634	3126363	3108500	4519549	0		
61	15251046	2	37	5.52E+08	36.20863	37	37	37	0	37	37	2,11,25	4523968	3106920	3100976	4519168	14		
62	15251046	11	37	5.52E+08	36.20066	37	37	37	0	37	37	11,25	4522314	3088985	3118243	4521504	0		
63	15251046	2	37	5.52E+08	36.20034	37	37	37	0	37	37	2,11,25	4493412	3124832	3112603	4520148	51		
64	15251046	2	37	5.52E+08	36.19177	37	37	37	0	37	37	2,11,25	4520814	3103894	3101301	4524959	78		
65	15251046	2	37	5.52E+08	36.17547	37	37	37	0	37	37	2,11,25	4517054	3089696	3120856	4523422	18		
66	15251046	2	37	5.52E+08	36.16409	37	37	37	0	37	37	2,11,25	4498276	3127454	3110307	4515003	6		
67	15251046	2	37	5.51E+08	36.15249	37	37	37	0	37	37	2,11,25	4529181	3101875	3100382	4519442	166		
68	15251046	2	37	5.51E+08	36.16087	37	37	37	0	37	37	2,11,25	4523590	3093792	3118544	4515024	96		
69	15251046	2	37	5.52E+08	36.16686	37	37	37	0	37	37	2,11,25	4501276	3124865	3112326	4512518	61		
70	15251046	11	37	5.51E+08	36.14533	37	37	37	0	37	37	11,25	4525677	3103590	3102857	4518922	0		
71	15251046	2	37	5.51E+08	36.1258	37	37	37	0	37	37	2,11,25	4517377	3092090	3117677	4523898	4		
72	15251046	2	37	5.51E+08	36.11368	37	37	37	0	37	37	2,11,25	4494808	3122586	3110216	4523373	63		
73	15251046	2	37	5.51E+08	36.10872	37	37	37	0	37	37	2,11,25	4525157	3101406	3104357	4520046	80		
74	15251046	2	37	5.51E+08	36.10089	37	37	37	0	37	37	2,11,25	4520720	3093649	3115617	4520961	99		
75	15251046	11	37	5.51E+08	36.10016	37	37	37	0	37	37	11,25	4494612	3124323	3111748	4520363	0		
76	15251046	11	37	5.51E+08	36.10094	37	37	37	0	37	37	11,25	4527406	3101244	3100382	4522014	0		
77	15251046	11	37	5.5E+08	36.07078	37	37	37	0	37	37	11,25	4519546	3087756	3118993	4524751	0		
78	15251046	2	37	5.5E+08	36.06984	37	37	37	0	37	37	2,11,25	4496356	3128569	3109641	4516434	46		
79	15251046	2	37	5.5E+08	36.05498	37	37	37	0	37	37	2,11,25	4528209	3100495	3102710	4519628	4		
80	15251046	11	37	5.5E+08	36.0565	37	37	37	0	37	37	11,25	4524245	3091144	3115090	4520567	0		
81	15251046	2	37	5.5E+08	36.06152	37	37	37	0	37	37	2,11,25	4497245	3122764	3112823	4518213	1		
82	15251046	2	37	5.5E+08	36.04321	37	37	37	0	37	37	2,11,25	4529019	3102538	3099577	4519907	5		
83	15251046	2	37	5.49E+08	36.01457	37	37	37	0	37	37	2,11,25	4524481	3092158	3115744	4518633	30		
84	15251046	2	37	5.49E+08	36.01392	37	37	37	0	37	37	2,11,25	4494767	3124904	3111342	4520007	26		
85	15251046	11	37	5.49E+08	36.01744	37	37	37	0	37	37	11,25	4529759	3101298	3102542	4517447	0		
86	15251046	2	37	5.49E+08	36.01403	37	37	37	0	37	37	2,11,25	4521588	3091512	3115917	4521973	56		
87	15251046	11	37	5.49E+08	35.99006	37	37	37	0	37	37	11,25	4497255	3124189	3114709	4514893	0		
88	15251046	2	37	5.49E+08	35.9786	37	37	37	0	37	37	2,11,25	4531647	3101781	3098123	4518480	1015		
89	15251046	2	37	5.48E+08	35.94898	37	37	37	0	37	37	2,11,25	4520217	3088856	3117608	4523031	1334		
90	15251046	2	37	5.48E+08	35.91629	37	37	37	0	37	37	2,11,25	4494509	3121548	3113056	4520045	1888		
91	15251046	2	37	5.48E+08	35.95182	37	37	37	0	37	37	2,11,25	4530182	3102129	3097521	4519626	1588		
92	15251046	2	37	5.48E+08	35.9397	37	37	37	0	37	37	2,11,25	4518404	3086821	3118007	4524432	3382		
93	15251046	2	37	5.48E+08	35.92545	37	37	37	0	37	37	2,11,25	4499679	3121580	3110004	4515681	4102		
94	15251046	2	37	5.48E+08	35.94454	37	37	37	0	37	37	2,11,25	4520678	3101613	3101263	4518230	9262		
95	15251046	2	37	5.48E+08	35.91845	37	37	37	0	37	37	2,11,25	4519317	3089674	3119769	4522284	2		
96	15251046	2	37	5.47E+08	35.87627	37	37	37	0	37	37	2,11,25	4493469	3122202	3113317	4519006	3052		
97	15251046	2	37	5.47E+08	35.87961	37	37	37	0	37	37	2,11,25	4521585	3101447	3102594	4525419	1		
98	15251046	2	37	5.47E+08	35.86335	37	37	37	0	37	37	2,11,25	4515862	3087176	3120832	4525940	1236		
99	15251046	2	37	5.47E+08	35.89072	37	37	37	0	37	37	2,11,25	4497098	3123263	3114554	4516068	63		
100	15251046	2	37	5.47E+08	35.86286	37	37	37	0	37	37	2,11,25	4522129	3103951	3102749	4522040	177		

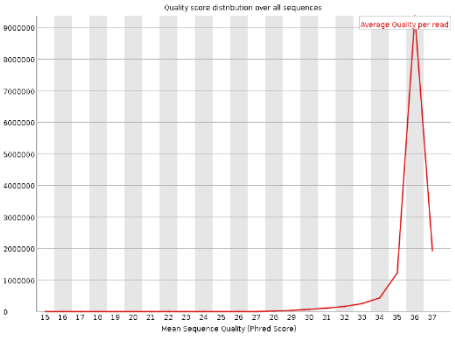
#column	count	min	max	sum	mean	Q1	med	Q3	IQR	lW	rW	outliers	A_Count	C_Count	G_Count	T_Count	N_Count	other_bases	other_base_count
1	15251046	2	37	5.3E+08	34.73391	37	37	37	0	37	37	2,11,25	4819150	3012516	2898010	4519413	1957		
2	15251046	2	37	5.28E+08	34.64803	37	37	37	0	37	37	2,11,25	5106803	2479887	3030194	4634116	46		
3	15251046	2	37	5.28E+08	34.63212	37	37	37	0	37	37	2,11,25	4851309	2758722	2944241	4696742	32		
4	15251046	11	37	5.27E+08	34.58245	37	37	37	0	37	37	11,25	4570250	2918534	3147423	4614839	0		
5	15251046	11	37	5.28E+08	34.64523	37	37	37	0	37	37	11,25	4515203	2943140	3257441	4535262	0		
6	15251046	11	37	5.28E+08	34.62806	37	37	37	0	37	37	11,25	4506850	3112643	3267580	4363973	0		
7	15251046	11	37	5.28E+08	34.5979	37	37	37	0	37	37	11,25	4503302	2997567	3292907	4457270	0		
8	15251046	2	37	5.28E+08	34.60085	37	37	37	0	37	37	2,11,25	4517409	3054990	3181851	4496794	2		
9	15251046	11	37	5.26E+08	34.49777	37	37	37	0	37	37	11,25	4442382	3092062	3161933	4554669	0		
10	15251046	2	37	5.26E+08	34.51553	37	37	37	0	37	37	2,11,25	4541020	3043955	3153763	4512293	15		
11	15251046	11	37	5.26E+08	34.50709	37	37	37	0	37	37	11,25	4547924	3018250	3140835	4544037	0		
12	15251046	2	37	5.26E+08	34.4925	37	37	37	0	37	37	2,11,25	4539393	3051545	3122232	4537515	361		
13	15251046	11	37	5.26E+08	34.46836	37	37	37	0	37	37	11,25	4581319	3009468	3097230	4563029	0		
14	15251046	11	37	5.25E+08	34.44525	37	37	37	0	37	37	11,25	4581588	2998259	3108097	4563102	0		
15	15251046	11	37	5.25E+08	34.43751	37	37	37	0	37	37	11,25	4548026	3052522	3097805	4552693	0		
16	15251046	2	37	5.25E+08	34.41756	37	37	37	0	37	37	2,11,25	4568983	3018792	3097636	4565593	42		
17	15251046	11	37	5.25E+08	34.42242	37	37	37	0	37	37	11,25	4572560	3006143	3113316	4559027	0		
18	15251046	11	37	5.25E+08	34.4208	37	37	37	0	37	37	11,25	4542603	3059736	3107731	4540976	0		
19	15251046	11	37	5.25E+08	34.39373	37	37	37	0	37	37	11,25	4572747	3026507	3102717	4549075	0		
20	15251046	11	37	5.24E+08	34.38405	37	37	37	0	37	37	11,25	4569102	3015858	3129660	4536426	0		
21	15251046	11	37	5.24E+08	34.35384	37	37	37	0	37	37	11,25	4542899	3057654	3128888	4521605	0		
22	15251046	11	37	5.24E+08	34.3789	37	37	37	0	37	37	11,25	4572301	3041434	3104331	4532980	0		
23	15251046	11	37	5.23E+08	34.30357	37	37	37	0	37	37	11,25	4566445	3025598	3129244	4529759	0		
24	15251046	11	37	5.23E+08	34.31668	37	37	37	0	37	37	11,25	4535263	3070154	3136269	4509360	0		
25	15251046	11	37	5.23E+08	34.30411	37	37	37	0	37	37	11,25	4560099	3054164	3119816	4516967	0		
26	15251046	11	37	5.23E+08	34.29478	37	37	37	0	37	37	11,25	4555415	3042262	3143037	4510332	0		
27	15251046	2	37	5.23E+08	34.29238	37	37	37	0	37	37	2,11,25	4526958	3080533	3143725	4499829	1		
28	15251046	11	37	5.23E+08	34.27615	37	37	37	0	37	37	11,25	4555730	3063508	3125274	4506534	0		
29	15251046	11	37	5.22E+08	34.24114	37	37	37	0	37	37	11,25	4544571	3050386	3152943	4503146	0		
30	15251046	2	37	5.22E+08	34.19731	37	37	37	0	37	37	2,11,25	4519316	3091215	3155714	4484800	1		
31	15251046	2	37	5.21E+08	34.17644	37	37	37	0	37	37	2,11,25	4547012	3075111	3142063	4486832	28		
32	15251046	2	37	5.21E+08	34.1757	37	37	37	0	37	37	2,11,25	4547225	3056522	3159732	4487516	51		
33	15251046	2	37	5.2E+08	34.11018	37	37	37	0	37	37	2,11,25	4515524	3096115	3163188	4476198	21		
34	15251046	2	37	5.2E+08	34.11027	37	37	37	0	37	37	2,11,25	4550941	3078505	3140583	4480993	24		
35	15251046	2	37	5.2E+08	34.10961	37	37	37	0	37	37	2,11,25	4547738	3060566	3163626	4479114	2		
36	15251046	11	37	5.2E+08	34.09325	37	37	37	0	37	37	11,25	4520326	3098310	3161397	4471013	0		
37	15251046	11	37	5.19E+08	34.05693	37	37	37	0	37	37	11,25	4552392	3078857	3147691	4472106	0		
38	15251046	11	37	5.19E+08	34.03821	37	37	37	0	37	37	11,25	4549010	3064739	3164260	4473037	0		
39	15251046	2	37	5.18E+08	33.99203	37	37	37	0	37	37	2,11,25	4518178	3102067	3166080	4464714	7		
40	15251046	11	37	5.18E+08	33.98083	37	37	37	0	37	37	11,25	4549088	3082819	3146873	4472266	0		
41	15251046	11	37	5.18E+08	33.99113	37	37	37	0	37	37	11,25	4545754	3060112	3165763	4479417	0		
42	15251046	11	37	5.18E+08	33.96223	37	37	37	0	37	37	11,25	4514352	3106076	3164746	4465872	0		
43	15251046	11	37	5.18E+08	33.95383	37	37	37	0	37	37	11,25	4547820	3085646	3147655	4469925	0		
44	15251046	11	37	5.17E+08	33.91709	37	37	37	0	37	37	11,25	4542680	3067260	3164715	4476391	0		
45	15251046	11	37	5.17E+08	33.89354	37	37	37	0	37	37	11,25	4512606	3103501	3165907	4469032	0		
46	15251046	11	37	5.17E+08	33.87176	37	37	37	0	37	37	11,25	4544949	3086263	3145600	4474234	0		
47	15251046	11	37	5.16E+08	33.86005	37	37	37	0	37	37	11,25	4545521	3068772	3165637	4471116	0		
48	15251046	11	37	5.16E+08	33.84306	37	37	37	0	37	37	11,25	4518724	3108651	3162148	4461523	0		
49	15251046	11	37	5.16E+08	33.82382	37	37	37	0	37	37	11,25	4544556	3088168	3146100	4472222	0		
50	15251046	11	37	5.16E+08	33.82081	37	37	37	0	37	37	11,25	4541307	3068991	3166060	4474688	0		
51	15251046	11	37	5.16E+08	33.80862	37	37	37	0	37	37	11,25	4517053	3107959	3163516	4462518	0		
52	15251046	11	37	5.16E+08	33.8041	37	37	37	0	37	37	11,25	4544922	3090223	3147653	4468248	0		
53	15251046	11	37	5.15E+08	33.76304	37	37	37	0	37	37	11,25	4542809	3069630	3167999	4470608	0		
54	15251046	11	37	5.15E+08	33.74222	37	37	37	0	37	37	11,25	4519857	3105219	3166194	4459776	0		
55	15251046	2	37	5.14E+08	33.71378	37	37	37	0	37	37	2,11,25	4548943	3085949	3149571	4466574	9		
56	15251046	11	37	5.14E+08	33.7331	37	37	37	0	37	37	11,25	4547957	3073935	3163417	4465737	0		
57	15251046	2	37	5.14E+08	33.67311	37	37	37	0	37	37	2,11,25	4524108	3100914	3163615	4462218	191		
58	15251046	11	37	5.14E+08	33.68426	37	37	37	0	37	37	11,25	4550529	3083314	3147762	4469441	0		
59	15251046	11	37	5.14E+08	33.6787	37	37	37	0	37	37	11,25	4546846	3069525	3164830	4469845	0		
60	15251046	11	37	5.14E+08	33.69077	37	37	37	0	37	37	11,25	4523812	3106469	3161006	4459759	0		
61	15251046	11	37	5.13E+08	33.61011	37	37	37	0	37	37	11,25	4552064	3087367	3145625	4465990	0		
62	15251046	11	37	5.12E+08	33.56611	37	37	37	0	37	37	11,25	4547747	3065154	3171006	4467139	0		
63	15251046	11	37	5.12E+08	33.57156	37	37	37	0	37	37	11,25	4516246	3106352	3167215	4461233	0		
64	15251046	11	37	5.12E+08	33.60099	37	37	37	0	37	37	11,25	4543320	3087143	3150261	4470322	0		
65	15251046	11	37	5.12E+08	33.54693	37	37	37	0	37	37	11,25	4545329	3068567	3167667	4469483	0		
66	15251046	11	37	5.11E+08	33.50588	37	37	37	0	37	37	11,25	4514703	3107983	3166859	4461501	0		
67	15251046	11	37	5.11E+08	33.52619	37	37	37	0	37	37	11,25	4547914	3084113	3150197	4468822	0		
68	15251046	11	37	5.11E+08	33.50834	37	37	37	0	37	37	11,25	4547702	3066932	3169075	4467337	0		
69	15251046	2	37	5.11E+08	33.53783	37	37	37	0	37	37	2,11,25	4519663	3103412	3165022	4462948	1		
70	15251046	11	37	5.12E+08	33.54426	37	37	37	0	37	37	11,25	4552248	3084865	3143194	4470739	0		
71	15251046	11	37	5.11E+08	33.47818	37	37	37	0	37	37	11,25	4550284	3071167	3163743	4465852	0		
72	15251046	11	37	5.1E+08	33.42501	37	37	37	0	37	37	11,25	4521506	3103420	3166810	4459310	0		
73	15251046	11	37	5.09E+08	33.40712	37	37	37	0	37	37	11,25	4552263	3083559	3149898	4465326	0		
74	15251046	11	37	5.1E+08	33.4225	37	37	37	0	37	37	11,25	4545142	3069482	3162728	4473694	0		
75	15251046	2	37	5.1E+08	33.43918	37	37	37	0	37	37	2,11,25	4518962	3104264	3165535	4462282	3		
76	15251046	11	37	5.09E+08	33.3657	37	37	37	0	37	37	11,25	4545643	3086844	3153007	4465552	0		
77	15251046	11	37	5.08E+08	33.30203	37	37	37	0	37	37	11,25	4542343	3068144	3167065	4473494	0		
78	15251046	11	37	5.08E+08	33.28596	37	37	37	0	37	37	11,25	4513516	3101226	3170130	4466174	0		
79	15251046	11	37	5.08E+08	33.30551	37	37	37	0	37	37	11,25	4542257	3089272	3149980	4469537	0		
80	15251046	11	37	5.08E+08	33.33795	37	37	37	0	37	37	11,25	4546698	3069970	3167110	4467268	0		
81	15251046	11	37	5.08E+08	33.32206	37	37	37	0	37	37	11,25	4515160	3101815	3169993	4464078	0		
82	15251046	2	37	5.07E+08	33.26979	37	37	37	0	37	37	2,11,25	4543609	3084044	3154538	4468327	528		
83	15251046	11	37	5.07E+08	33.26981	37	37	37	0	37	37	11,25	4543883	3069188	3168804	4469171	0		
84	15251046	11	37	5.07E+08	33.24845	37	37	37	0	37	37	11,25	4518961	3102277	3170287	4459521	0		
85	15251046	11	37	5.05E+08	33.14338	37	37	37	0	37	37	11,25	4545290	3080789	3157323	4467644	0		
86	15251046	11	37	5.06E+08	33.16781	37	37	37	0	37	37	11,25	4543070	3068614	3172016	4467346	0		
87	15251046	11	37	5.06E+08	33.18521	37	37	37	0	37	37	11,25	4516002	3099834	3172589	4462621	0		
88	15251046	11	37	5.06E+08	33.16849	37	37	37	0	37	37	11,25	4544916	3083029	3156209	4466892	0		
89	15251046	11	37	5.05E+08	33.12377	37	37	37	0	37	37	11,25	4540113	3065299	3174041	4471593	0		
90	15251046	11	37	5.04E+08	33.03052	37	37	37	0	37	37	11,25	4512811	3101693	3173623	4462919	0		
91	15251046	11	37	5.05E+08	33.12355	37	37	37	0	37	37	11,25	4547588	3082339	3154392	4466727	0		
92	15251046	2	37	5.05E+08	33.10491	37	37	37	0	37	37	2,11,25	4540753	3061617	3177372	4471171	133		
93	15251046	2	37	5.03E+08	32.9946	37	37	37	0	37	37	2,11,25	4515949	3100294	3174215	4460582	6		
94	15251046	2	37	5.04E+08	33.06997	37	37	37	0	37	37	2,11,25	4543357	3085193	3157854	4464557	85		
95	15251046	11	37	5.04E+08	33.02544	37	37	37	0	37	37	11,25	4537022	3067612	3176001	4470411	0		
96	15251046	2	37	5.02E+08	32.90911	37	37	37	0	37	37	2,11,25	4510435	3103034	3177399	4460151	27		
97	15251046	11	37	5.03E+08	32.97562	37	37	37	0	37	37	11,25	4540953	3082831	3156070	4471192	0		
98	15251046	2	37	5.04E+08	33.04537	37	37	37	0	37	37	2,11,25	4533192	3065275	3175799	4476777	3		
99	15251046	11	37	5.04E+08	33.0168	37	37	37	0	37	37	11,25	4510486	3101073	3175827	4463660	0		

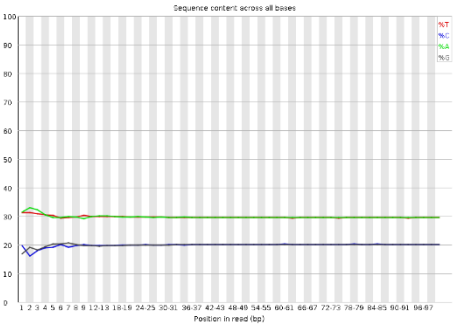
Thus, the quality statistics of the wild type have been compiled.

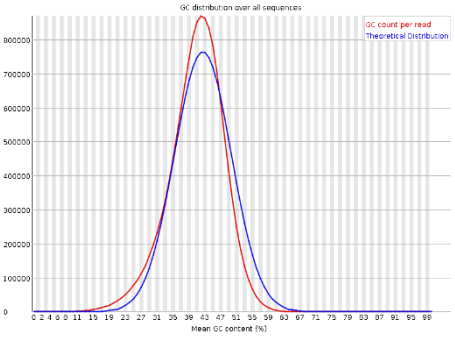
2. *Quality assessment of raw reads*

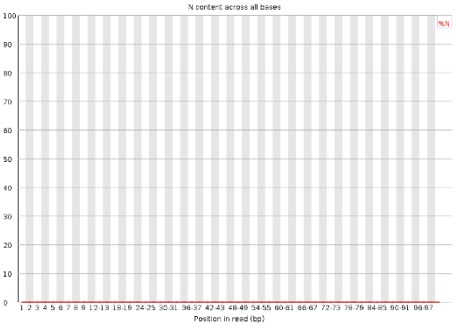


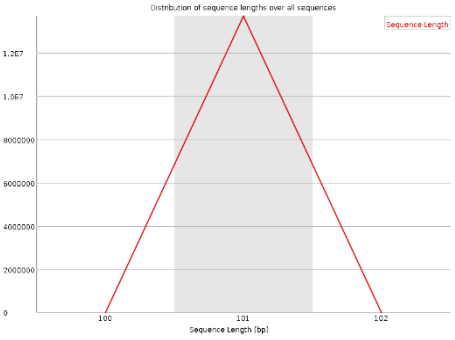


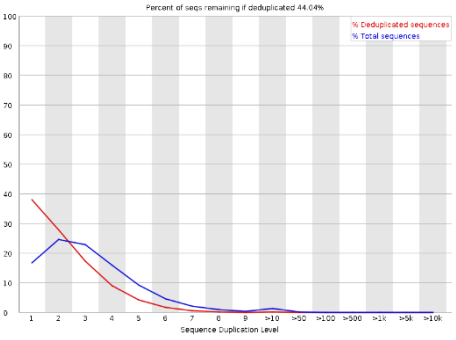


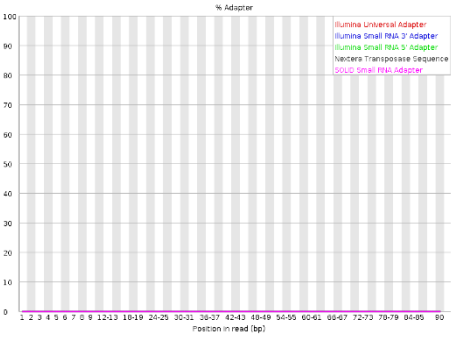


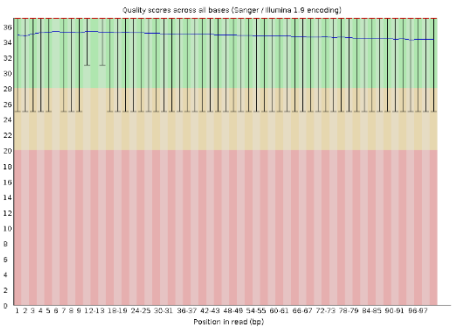


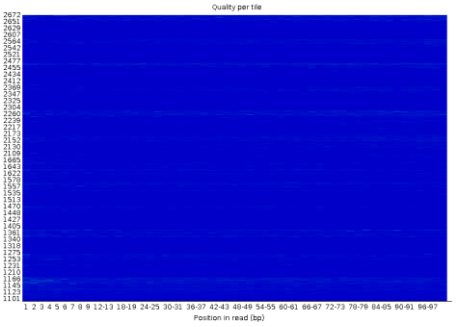


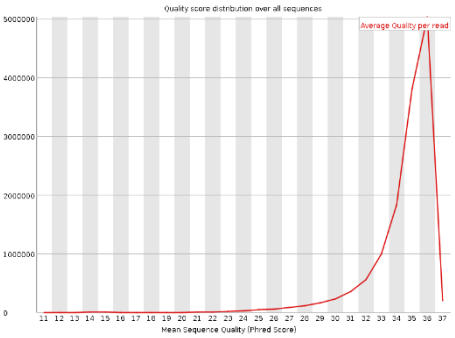


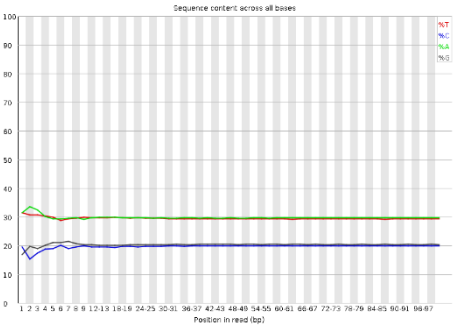


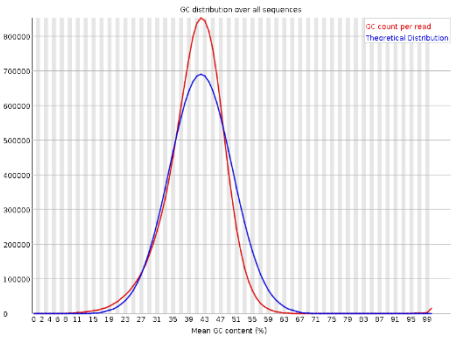


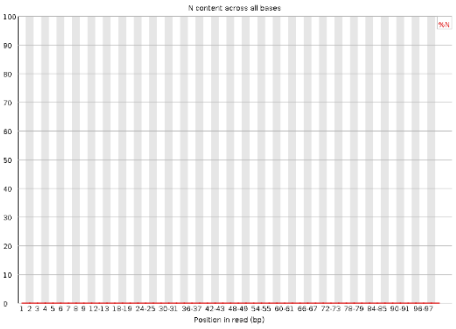


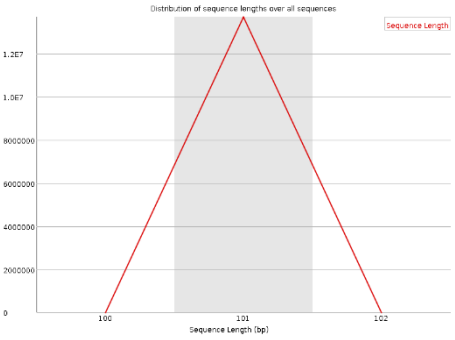


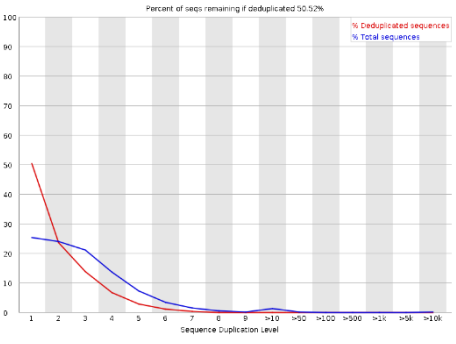


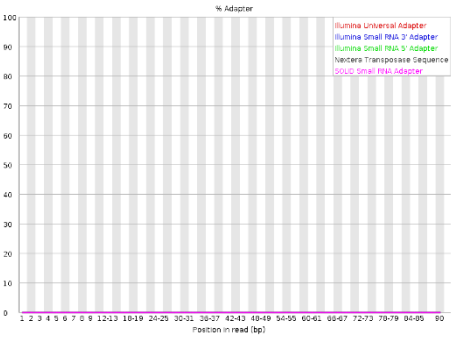




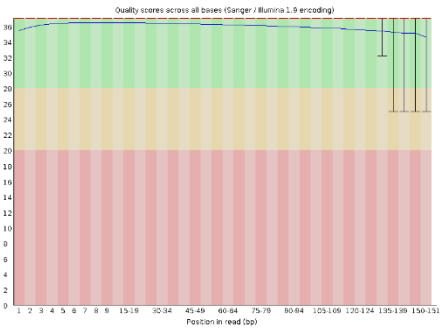


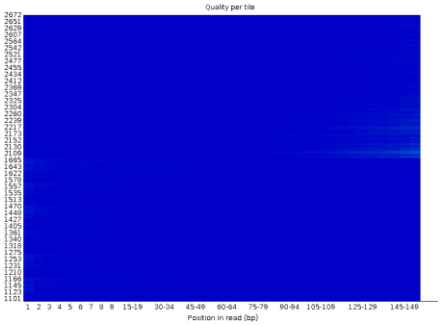


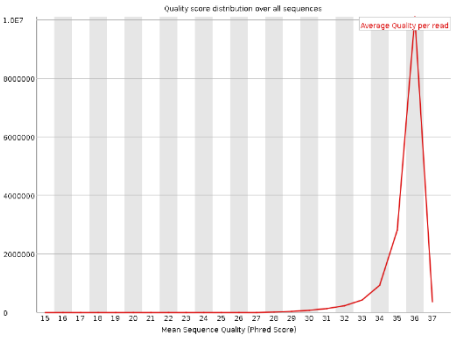


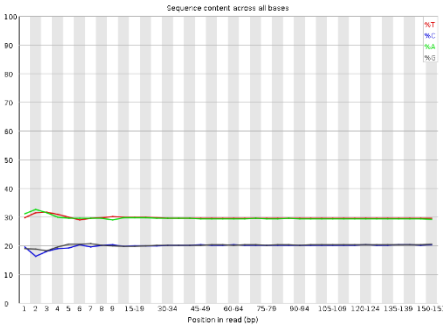


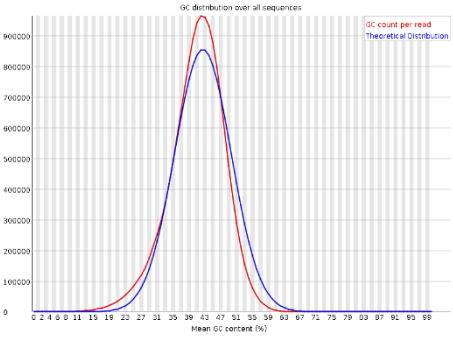
From the above graphs for the evolved strain, we can see that the while the sequence quality per base remains predominantly constant, the quality score per sequence peaks towards the end of the graph and then falls quickly. We can also see that the percentage of duplicates reduces as we move forward.

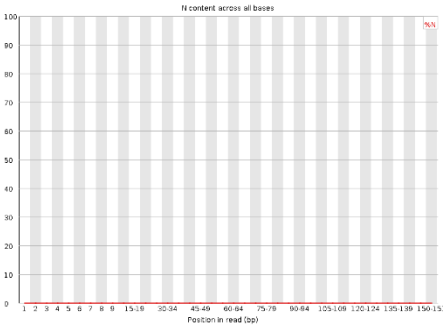

Quality scores across all bases (Sanger / Illumina 1.9 encoding)
Position in read (bp)

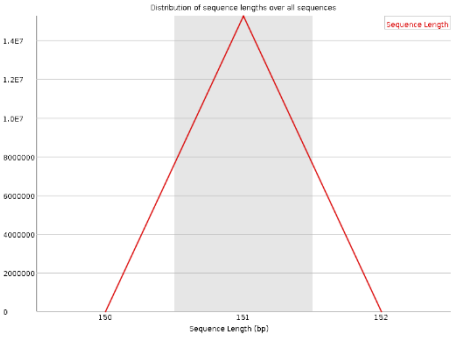

Quality per tile
Position in read (bp)

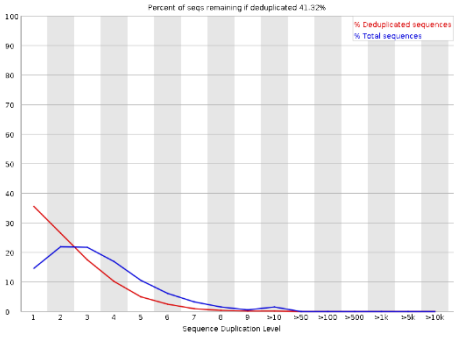

Quality score distribution over all sequences
Average Quality per read
Mean Sequence Quality (Phred Score)

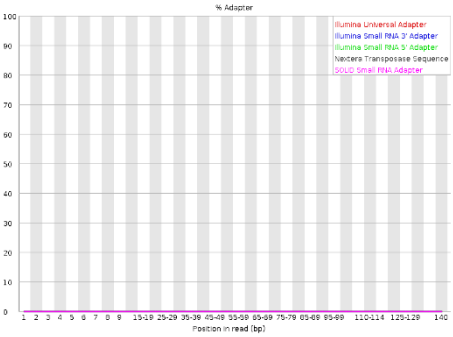

Sequence content across all bases
%T
%C
%A
%G
Position in read (bp)

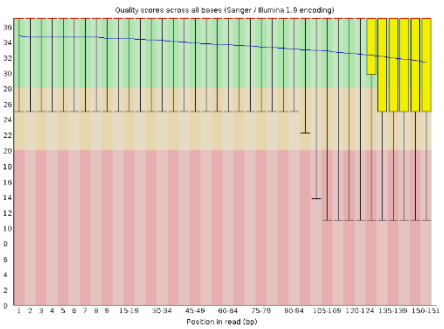

GC distribution over all sequences
GC count per read
Theoretical Distribution
Mean GC content (%)

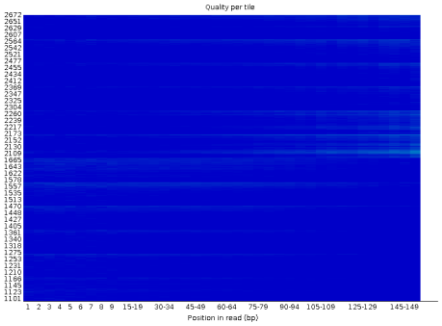

N content across all bases
%N
Position in read (bp)

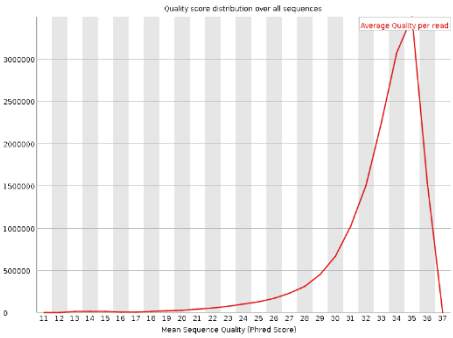

Distribution of sequence lengths over all sequences
Sequence Length
Sequence Length (bp)

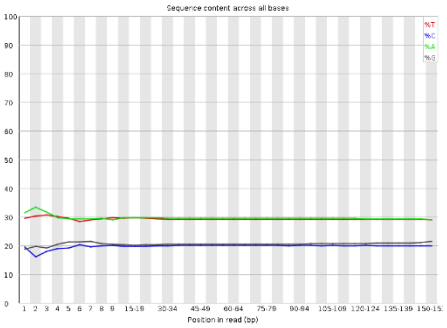

Percent of seqs remaining if deduplicated 41.92%
% Deduplicated sequences
% Total sequences
Sequence Duplication Level

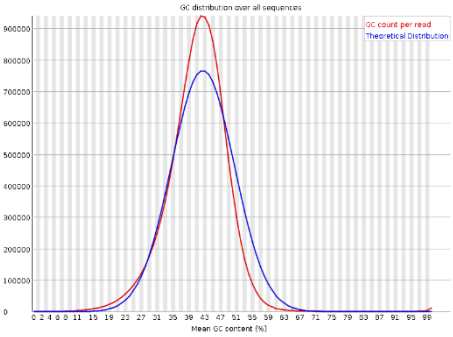

% Adapter
Illumina Universal Adapter
Illumina Small RNA 3' Adapter
Illumina Small RNA 5' Adapter
Nextera Transposase Sequence
SOLiD Small RNA Adapter
Position in read (bp)

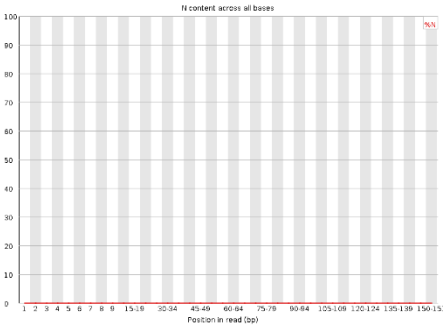


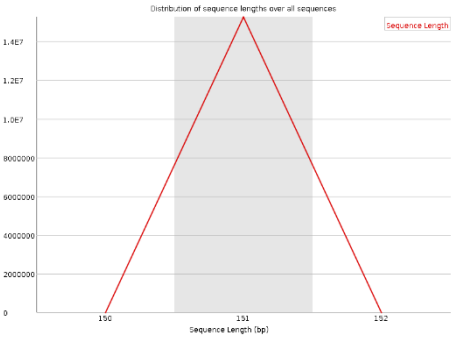


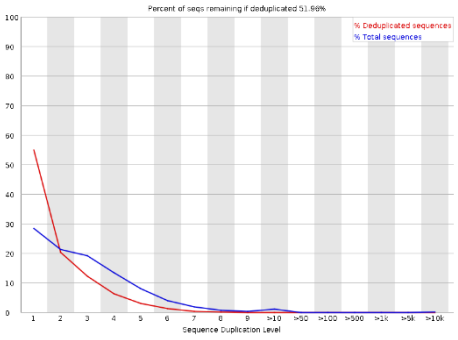


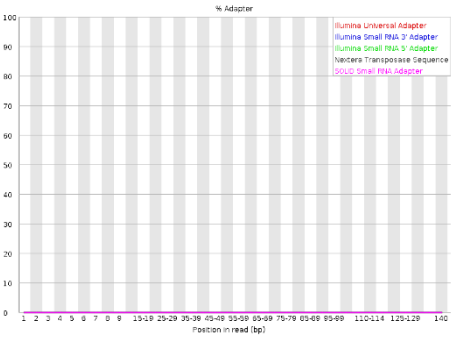












From the above graphs for the wild type, we can see that the while the sequence quality per base starts to decrease, the quality score per sequence peaks towards the end of the graph and falls again. We can also see that the percentage of duplicates reduces significantly as we move forward.

3. Mapping of genome to reference

https://usegalaxy.org/datasets/bbd44e69cb8906b590145d713b9c50ec/display/?preview=True

In the table that can be found above, the evolved strain has been mapped to the reference genome. We can see the differences between the reference and alternate genomes for each position which has a difference.

https://usegalaxy.org/datasets/bbd44e69cb8906b53e905a98c9039547/display/?preview=True

In the table that can be found above, the wild type has been mapped to the reference genome. We can see the differences between the reference and alternate genomes for each position which has a difference.

4. Marking and removing duplicate reads

https://usegalaxy.org/datasets/bbd44e69cb8906b5c7cb0d6e86abfd8d/display/?preview=True

In the table that can be found above, the duplicate reads have been marked and removed for the BAM file, which contains the mapping of the reference genome to the evolved strain.

https://usegalaxy.org/datasets/bbd44e69cb8906b59969c829e06f97bb/display/?preview=True

In the table that can be found above, the duplicate reads have been marked and removed for the BAM file, which contains the mapping of the reference genome to the wild type.

5. Variant calling

https://usegalaxy.org/datasets/bbd44e69cb8906b590145d713b9c50ec/display/?preview=True

In the table that can be found above, variants were called for the evolved strain with respect to the reference genome after duplicates had been removed. We can see that most of the mutations are single nucleotide polymorphisms.

https://usegalaxy.org/datasets/bbd44e69cb8906b53e905a98c9039547/display/?preview=True

In the table that can be found above, variants were called for the wild type with respect to the reference genome after duplicates had been removed. We can see that most of the mutations are single nucleotide polymorphisms.

6. Annotating variants

https://usegalaxy.org/datasets/bbd44e69cb8906b50943750f2d471b9a/display/?preview=True

In the list of tables and illustrations that can be found above, variants were annotated for the evolved strain with respect to the reference genome after variant calling. The 104 mutations have been summarised, and various graphs have been plotted. The ratio of transitions to transversions is found to be 3.2. The length of insertions and deletions has also been found.

https://usegalaxy.org/datasets/bbd44e69cb8906b5b575f55c1a2ffcca/display/?preview=True

In the list of tables and illustrations that can be found above, variants were annotated for the wild type with respect to the reference genome after variant calling. The 37 mutations have been summarised, and various graphs have been plotted. The ratio of transitions to transversions is found to be 8, a significant increase from the previous case. The length of insertions and deletions has also been found.

7. Compiling mutations

Wild type

Pos	Ref	Alt	Type of Mutation		Type	Gene	Product
592336	T	G	Substitution	Transversion	CDS	-	putative glycosyltransferase
592340	C	G,T	Substitution		CDS	-	putative glycosyltransferase
592343	G	T	Substitution	Transversion	CDS	-	putative glycosyltransferase
1179108	C	T	Substitution	Transition	CDS	phrB	deoxyribodipyrimidine photolyase (photoreactivation), FAD-binding
1636002	C	A	Substitution	Transversion	CDS	-	putative surface protein (partial adhesin)
1803843	A	G	Substitution	Transition	CDS	-	hypothetical protein; putative membrane protein
1803858	CGCT	TGCC	MNP		CDS	-	hypothetical protein; putative membrane protein
1803900	CTAT	TTAC	MNP		CDS	-	hypothetical protein; putative membrane protein
1803915	G	A	Substitution	Transition	CDS	-	hypothetical protein; putative membrane protein
1803921	A	G	Substitution	Transition	CDS	-	hypothetical protein; putative membrane protein
1803930	TTTA	CTTG	MNP		CDS	-	hypothetical protein; putative membrane protein
1803939	C	T	Substitution	Transition	CDS	-	hypothetical protein; putative membrane protein
1851676	T	G	Substitution	Transversion	CDS	-	uncharacterized prophage protein
2216887	C	A	Substitution	Transversion	CDS	-	putative outer membrane porin protein precursor
2372248	C	G	Substitution	Transversion	CDS	-	conserved hypothetical protein; putative chaperone
2420277	ACCACG	ACCCACG	Insertion		CDS	glnA	glutamine synthetase
2542652	G	A	Substitution	Transition	CDS	lepA	GTP-binding protein
2765010	A	C	Substitution	Transversion	CDS	-	putative periplasmic binding protein of transport/transglycosylase
2803531	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2804110	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2804983	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2805028	A	G	Substitution	Transition	CDS	bap	biofilm associated protein
2805037	A	G	Substitution	Transition	CDS	bap	biofilm associated protein
2805054	T	C	Substitution	Transition	CDS	bap	biofilm associated protein
2805067	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2805145	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2805196	A	G	Substitution	Transition	CDS	bap	biofilm associated protein
2843851	T	A	Substitution	Transversion	tRNA	-	Gln tRNA
2843852	A	T	Substitution	Transversion	tRNA	-	Gln tRNA
2844054	CAAAA	GATAT	MNP		tRNA	-	Gln tRNA
3357842	G	A	Substitution	Transition	CDS	-	putative DUF4357 domain-containing protein
3449576	C	A	Substitution	Transversion	CDS	-	conserved hypothetical protein
3449577	T	A	Substitution	Transversion	CDS	-	conserved hypothetical protein
3533222	C	A	Substitution	Transversion	CDS	-	conserved hypothetical protein; putative membrane protein
3533224	AG	AT	Substitution	Transition	CDS	-	conserved hypothetical protein; putative membrane protein
3538607	G	A	Substitution	Transition	CDS	ppc	phosphoenolpyruvate carboxylase

Evolved Strain

Pos	Ref	Alt	Type of Mutation		Type	Gene	Product
18404	G	T	Substitution	Transversion	rRNA	-	16S
82911	GA	TC	MNP		CDS	-	putative UDP-glucose/GDP-mannose dehydrogenase
82918	CAACGTT	GATCCTA	MNP		CDS	-	putative UDP-glucose/GDP-mannose dehydrogenase
173224	T	A	Substitution	Transversion	CDS	-	conserved hypothetical protein; putative dyp-type peroxidase
173228	GCATA	TGATC	MNP		CDS	-	conserved hypothetical protein; putative dyp-type peroxidase
173233	A	C	Substitution	Transversion	CDS	-	conserved hypothetical protein; putative dyp-type peroxidase
178458	AGGGGGGGGGCTCT	AGGGGGGGGCTCT	Deletion		CDS	atpI	ATP synthase protein I
449101	CAACGTGT	CGTAGATGA	Insertion		CDS	-	putative acyl-CoA dehydrogenase (AidB)
449276	TAT	CAA	MNP		CDS	-	putative acyl-CoA dehydrogenase (AidB)
449279	TTTTCAGCATA	ATTTTAGAATC,TTTTAGAATC	MNP		CDS	-	putative acyl-CoA dehydrogenase (AidB)
592336	T	G	Substitution	Transversion	CDS	-	putative glycosyltransferase
592338	TTC	TTG			CDS	-	putative glycosyltransferase
786523	A	T	Substitution	Transversion	tRNA	-	Asp tRNA
935746	AGTCA	TGATC			CDS	-	conserved hypothetical protein
941297	TCATTAAA	AAATATAAG	Insertion		IS	IS1236	Insertion Sequence IS1236, IS3 family
941305	G	A	Substitution	Transition	IS	IS1236	Insertion Sequence IS1236, IS3 family
942543	AAGTGATAAT	AAGGATTGATATT	Insertion		CDS	-	conserved hypothetical protein
942553	TA	TT			CDS	-	conserved hypothetical protein
942556	A	G	Substitution	Transition	CDS	-	conserved hypothetical protein
943612	G	A	Substitution	Transition	CDS	-	conserved hypothetical protein
943613	A	T	Substitution	Transversion	CDS	-	conserved hypothetical protein
943615	AATTAT	AAAGAT	MNP		CDS	-	conserved hypothetical protein
944841	TATTTTGG	GATACTGT	MNP		IS	IS1236	Insertion Sequence IS1236, IS3 family
944857	AATTG	ATAATC,ATCTTT,AGATGG,AATCTA	Insertion		CDS	-	conserved hypothetical protein; putative membrane protein
944863	C	G	Substitution	Transversion	CDS	-	conserved hypothetical protein; putative membrane protein
944865	GC	AG,GT	MNP		CDS	-	conserved hypothetical protein; putative membrane protein
944867	TT	TC	Substitution	Transition	CDS	-	conserved hypothetical protein; putative membrane protein
964810	CATCAGG	CTGATCCTA	Insertion		CDS	vanR	transcriptional regulator for ferulate or vanillate catabolism (GntR family)
968302	TGGATA	GGCAATTA	Insertion		CDS	-	hypothetical protein
968310	GATATG	CATAAT	MNP		CDS	-	hypothetical protein
1168348	TTGT	GATC	MNP		CDS	-	conserved hypothetical protein
1168357	TCG	GAT	MNP		CDS	-	conserved hypothetical protein
1179108	C	T	Substitution	Transition	CDS	phrB	deoxyribodipyrimidine photolyase (photoreactivation), FAD-binding
1202039	G	A	Substitution	Transition	CDS	hemL	glutamate-1-semialdehyde aminotransferase
1360293	TGGCT	TCT	Deletion		CDS	clpA	ATP-binding protease component
1501968	T	G	Substitution	Transversion	CDS	-	putative Dibenzothiophene desulfurization enzyme
1501969	GGTCC	GATCA	MNP		CDS	-	putative Dibenzothiophene desulfurization enzyme
1501978	TTAT	GATC	MNP		CDS	-	putative Dibenzothiophene desulfurization enzyme
1522654	TCATAT	TAGGATCA	Insertion		CDS	-	putative transport protein (ABC superfamily, ATP_bind)
1522664	ACG	GAT	MNP		CDS	-	putative transport protein (ABC superfamily, ATP_bind)
1534825	T	G	Substitution	Transversion	CDS	-	putative tonB-dependent receptor protein (outer membrane salicin receptor)
1534829	ACTTTT	AATGC	Deletion		CDS	-	putative tonB-dependent receptor protein (outer membrane salicin receptor)
1572873	TTTTG	GTCTA	MNP		CDS	-	putative transcriptional regulator (TetR family)
1572878	GGATCG	GGATCACG	Insertion		CDS	-	putative transcriptional regulator (TetR family)
1636002	C	A	Substitution	Transversion	CDS	-	putative surface protein (partial adhesin)
1708197	G	A	Substitution	Transition	CDS	pcaU	regulatory protein for pca operon (activator)
1723516	CAGCAG	GATCCTA,CAGCTG	Insertion		CDS	quiA	quinate/shikimate dehydrogenase [Pyrroloquinoline-quinone] (NAD(P)-independent quinate dehydrogenase)
1729396	AGAAGA	TGATCC			CDS	hcaE	porin
1729596	ATTGGT	CTAGGATC	Insertion		CDS	hcaE	porin
1789246	ACAT	GATC	MNP		CDS	ppk	polyphosphate kinase (Polyphosphoric acid kinase) (ATP- polyphosphate phosphotransferase)
1789408	TTTATT	GCTAGG	MNP		CDS	ppk	polyphosphate kinase (Polyphosphoric acid kinase) (ATP- polyphosphate phosphotransferase)
1803858	CGCT	TGCC	MNP		CDS	-	hypothetical protein; putative membrane protein
1803900	CTAT	TTAC	MNP		CDS	-	hypothetical protein; putative membrane protein

1803915	G	A	Substitution	Transition	CDS	-	hypothetical protein; putative membrane protein
1803921	A	G	Substitution	Transition	CDS	-	hypothetical protein; putative membrane protein
1803930	TTTA	CTTG	MNP		CDS	-	hypothetical protein; putative membrane protein
1803939	C	T	Substitution	Transition	CDS	-	hypothetical protein; putative membrane protein
1845223	TGCGG	GGATC	MNP		CDS	-	putative transcriptional regulator (AraC family)
1845232	AA	TG	MNP		CDS	-	putative transcriptional regulator (AraC family)
1845236	A	C	Substitution	Transversion	CDS	-	putative transcriptional regulator (AraC family)
1857795	A	C	Substitution	Transversion	CDS	-	putative repressor of phage gene expression protein (RstR-like)
1858199	C	G	Substitution	Transversion	CDS	-	conserved hypothetical protein
2164696	ATATAAGGTTATTGA	GGATCAGTTATTGA,ATATAAGGTGATCCT	Deletion/MNP		CDS	-	conserved hypothetical protein
2216887	C	A	Substitution	Transversion	CDS	-	putative outer membrane porin protein precursor
2372312	G	T	Substitution	Transversion	CDS	-	conserved hypothetical protein; putative chaperone
2420277	ACCACG	ACCCACG	Insertion		CDS	glnA	glutamine synthetase
2448584	AG	TC	MNP		CDS	-	putative two-component response regulator protein
2448591	AACGG	GATCC	MNP		CDS	-	putative two-component response regulator protein
2542652	G	A	Substitution	Transition	CDS	lepA	GTP-binding protein
2667494	T	G	Substitution	Transversion	fCDS	-	fragment of component of DNA polymerase V (UmuC) (part 1)
2753757	TC	CA	MNP		CDS	-	conserved hypothetical protein
2753764	TTTT	ATCC	MNP		CDS	-	conserved hypothetical protein
2754220	CGCATT	AGGATC	MNP		CDS	-	putative transcriptional regulator (IcIR family)
2754229	CTAGA	CAGA	Deletion		CDS	-	putative transcriptional regulator (IcIR family)
2754235	C	T	Substitution	Transition	CDS	-	putative transcriptional regulator (IcIR family)
2754236	T	A	Substitution	Transversion	CDS	-	putative transcriptional regulator (IcIR family)
2754238	A	G	Substitution	Transition	CDS	-	putative transcriptional regulator (IcIR family)
2765010	A	C	Substitution	Transversion	CDS	-	putative periplasmic binding protein of transport/transglycosylase
2803531	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2804110	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2804938	A	G	Substitution	Transition	CDS	bap	biofilm associated protein
2804956	T	C	Substitution	Transition	CDS	bap	biofilm associated protein
2804983	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2805028	A	G	Substitution	Transition	CDS	bap	biofilm associated protein
2805037	A	G	Substitution	Transition	CDS	bap	biofilm associated protein
2805054	T	C	Substitution	Transition	CDS	bap	biofilm associated protein
2805067	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2805145	G	A	Substitution	Transition	CDS	bap	biofilm associated protein
2805196	A	G	Substitution	Transition	CDS	bap	biofilm associated protein
2843816	G	C	Substitution	Transversion	tRNA	-	Gln tRNA
2843851	T	A	Substitution	Transversion	tRNA	-	Gln tRNA
2843852	A	T	Substitution	Transversion	tRNA	-	Gln tRNA
2843853	C	A	Substitution	Transversion	tRNA	-	Gln tRNA
2843856	C	A	Substitution	Transversion	tRNA	-	Gln tRNA
2843964	G	T	Substitution	Transversion	tRNA	-	Gln tRNA
2844056	A	T	Substitution	Transversion	tRNA	-	Gln tRNA
2937249	A	C	Substitution	Transversion	CDS	ksgA	S-adenosylmethionine-6-N',N'-adenosyl (rRNA) dimethyltransferase, kasugamycin resistance
2937252	A	C	Substitution	Transversion	CDS	ksgA	S-adenosylmethionine-6-N',N'-adenosyl (rRNA) dimethyltransferase, kasugamycin resistance
2937253	A	C	Substitution	Transversion	CDS	ksgA	S-adenosylmethionine-6-N',N'-adenosyl (rRNA) dimethyltransferase, kasugamycin resistance
3241277	G	A	Substitution	Transition	CDS	comP	competence factor involved in DNA binding and uptake
3241284	C	T	Substitution	Transition	CDS	comP	competence factor involved in DNA binding and uptake
3241285	AA	CC,TA			CDS	comP	competence factor involved in DNA binding and uptake
3241287	T	G	Substitution	Transversion	CDS	comP	competence factor involved in DNA binding and uptake
3329806	TATTG	GATCA	MNP		CDS	-	putative outer membrane protein precursor
3329814	AAG	TGA	MNP		CDS	-	putative outer membrane protein precursor
3357842	G	A	Substitution	Transition	CDS	-	putative outer membrane protein precursor

Learning Experience

As a chemical engineering student with little prior exposure to biotechnology, this internship and the processes that led up to the task gave me something new to learn.

I learnt about a new way to be able to make greener fuel with this study, something I had not known before. This was a unique learning experience.

I began by making a literature study where I learnt about how actually to perform a literature study. I read several papers about whole-genome sequence analysis and adaptive laboratory evolution, which taught me what these processes were and how they could be implemented. I learnt about the applications of these concepts and gave myself an introduction to the work I had to perform.

I learnt about different bioinformatics tools that can be used to perform various tasks in the pipeline for whole-genome sequencing. I learnt about what kind of codes these tools use, how they work, and what results they produce.

Subsequently, I learnt about mutations, their types, and their effects. I was able to use this knowledge when compiling the final spreadsheet.

This experience gave me an exciting introduction to this branch of science and gave me even more motivation to pursue biotechnology further. I am incredibly grateful for the people involved, who taught me with patience and understanding.

Conclusion

Climate change has become a growing concern that needs to be given serious consideration. Today's population and their demands result in the use of petroleum products which are non-renewable, non-eco-friendly sources of energy. This study takes us one step closer to realising the goal of production of renewable and harmless sources of energy.

A single mutation can have a significant effect, but in many cases, evolutionary change is based on the accumulation of many mutations with minor consequences. Mutational effects can be beneficial, harmful, or neutral, depending on their context or location. Most non-neutral mutations are deleterious. In general, the more base pairs that are affected by a mutation, the more significant is the effect of the mutation, and the larger is the mutation's probability of being deleterious.

Biofilm associated protein (BAP) [10] plays a vital role in biofilm formation and adhesion to host cells in *Acinetobacter*. A possible outcome is that the mutation of a bap gene of a coding sequence could result in the sequence of proteins changing. A likely outcome is a change in film properties and adhesive properties of the strain. This can be used to enhance the breakdown of lignin.

The mutations of tRNA [11] affect how the ribosomes read the mRNA. This affects the fidelity of the decoding process. The tRNA adopts a different conformation when it reaches a ribosome that correctly decodes it. This could affect the protein structure, which could result in a change in the conformation of these proteins.

Thus, the whole-genome sequence analysis of *Acinetobacter baylyi* has been performed using computational methods as explained previously. The predictions made have to be validated by conducting experiments. When the pattern of mutations and their effects are studied, the strain can be modified further to produce desirable outcomes. These mutations in the evolved strain would be used to uptake the lignin monomers and produce a fuel that can be used in place of petroleum products without damaging the environment.

Bibliography

1. Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., & Nekrutenko, A.: **Manipulation of FASTQ data with Galaxy,** *Bioinformatics (2010)* https://academic.oup.com/bioinformatics/article/26/14/1783/178709

2. Andrews, S (n.d.).: **FastQC A Quality Control tool for High Throughput Sequence Data** https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

3. Bolger, A. M., Lohse, M., & Usadel, B.: **Trimmomatic: a flexible trimmer for Illumina sequence data**, *Bioinformatics (2014)* https://academic.oup.com/bioinformatics/article/30/15/2114/2390096

4. Li, H., & Durbin, R.: **Fast and accurate short read alignment with Burrows-Wheeler transform**, *Bioinformatics (2009)* https://academic.oup.com/bioinformatics/article/25/14/1754/225615

5. Li, H., & Durbin, R.: **Fast and accurate long-read alignment with Burrows-Wheeler transform,** *Bioinformatics (2010)* https://academic.oup.com/bioinformatics/article/26/5/589/211735

6. Institute, B. (n.d.): Picard, Broad Institute, GitHub repository http://broadinstitute.github.io/picard/

7. O. Tange.: **GNU Parallel - The Command-Line Power Tool,** *Login: The USENIX Magazine (2011)* http://www.gnu.org/software/parallel/

8. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M.: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff,** *Fly (2012)* https://www.tandfonline.com/doi/full/10.4161/fly.19695

9. D. Vallenet, E. Belda, A. Calteau, S. Cruveiller, S. Engelen, A. Lajus, F. Le Fèvre, C. Longin, D. Mornico, D. Roche, Z. Rouy, G. Salvignol, C. Scarpelli, A.A.T. Smith, M. Weiman, C. Médigue: **MicroScope - An integrated microbial resource for the curation and comparative analysis of genomic and metabolic data**, *Nucleic Acids Res (2013)* https://academic.oup.com/nar/article/41/D1/D636/1068146

10. E. de Gregorio, M. del Franco, P. P. di Nocera: **Biofilm-associated proteins – News from Acinetobacter**, *BMC Genomics (2015)* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4647330/

11. T. M. Schmeing, R. M. Voorhees, V. Ramakrishnan: **How mutations in tRNA distant from the anticodon affect the fidelity of decoding**, *Nat Struct Mol Biol (2011)* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3072312/