Genomic Assembly and Investigation of Isolates from Terrestrial Analogue Sites

Abstract

Contents

Introduction

The aim of the Centre for ExoLife Sciences (CELS)[1] is to improve our understanding of the physical, chemical, and biological conditions of the terrestrial planets of our solar system and exoplanets. The project "Effects of bacteria on atmospheres of Earth, Mars and exoplanets" is an interdisciplinary collaboration between the Department of Biology, Department of Chemistry, and the Niels Bohr institute at University of Copenhagen. The project is comprised by a research pipeline: Field collection of microorganisms from terrestrial analogue (TA) sites; extraction and mass spectroscopy of exhaust gasses from live microorganisms in The Jens-Martin-Knudsen Mars simulation chamber; chemical reaction mechanism and kinetics modelling of said exhaust gasses and finally modelling the life climate interactions with supercomputer simulations. The overarching goal of the project is to provide details on detectable atmospheric signs of life by comparing these modelled spectra with observations in the spectra of exoplanets with the Extremely Large Telescope ELT, currently being built.

Intending to contribute to the project, the aim of this paper is to provide annotated genome assemblies of the organisms extracted from the permafrost of Citron Fjord, North Greenland and lithic samples from the Atacama Desert, Chile. In the lab organisms were cultivated in a range of media to select for either halotolerance or psychro and halotolerance combined. Isolates from these media were cultured to yield cell densities suitable for sequencing as well as future experiments. Using polymerase chain reaction (PCR) of the 16s rRNA gene we first verified that the isolates were prokaryotic. Then by Sanger sequencing the 16s rRNA gene, we were able to identify close relatives to our isolated stains and use their reference genome as templates to map Illumine sequence data from the isolates to construct genome assemblies. These genome assemblies were then annotated so that the identified genes can be used in transcription analysis as well as serve as potential targets for future genetic modification to optimize the organisms' ability to survive and grow in the Mars simulation chamber.

Theory

Terrestrial Analogue Sites

Environments on earth with physical, chemical, geological or biological conditions similar to conditions on planets in our solar system, as well as planets outside of our solar system, exoplanets, are considered terrestrial analogue sites (TAS). Many currently studied TAS, including those utilized for this paper, are cold or dry with very low liquid water content, and reflect the conditions of Mars in particular.

The Atacama Desert qualifies as a high-fidelity terrestrial analogue of Mars in three respects. (Navarro-González, Rainey, Molina, Bagaley, & McKay, 2003) Martian soil can be characterized firstly by a complete lack of life, an absence of organic material at the level of parts-per-billion (ppb), (Navarro-Gonzalez, et al., 2006)[i] and highly oxidizing environment (Oyama & Bjerdahl, 1997) due to perchlorate among other oxidizing agents. (Hecht et al., 2009) The Atacama Desert is extremely arid (driest non polar desert on earth) and precipitation and temperature are inversely related along a latitudinal gradient. Along this gradient, the occurrence of organic molecules and culturable heterotrophic bacteria decreases with proximity to the extreme arid core of the Atacama. Meanwhile, the presence of oxidants, produced by photochemical reactions and preserved in the arid environment, increases moving toward the arid core, largely due to the lack of biosynthetically produced reducing agents. (Navarro-González, Rainey, Molina,

[i] The validity of this claim is disputed by K. Bienmann who claims lower levels of organic molecules were detectable:
Biemann, K. (2007) "On the Ability of the Viking Gas Chromatograph-Mass Spectrometer to Detect Organic Matter. *Chemistry 104(25) 10310-10313*

Bagaley, & McKay, 2003) The Martian soil perchlorate content, present at roughly 0.6%, which is toxic to plants and animals, may not be as much of a hindrance to bacteria. Additionally, perchlorate salts readily attract atmospheric water and dissolve therein forming saturated brines through deliquescence. (Al Soudi, Ferhat, Chen, Clark, & Schneegurt, 2017)

Citronen Fjord

Kap København

Cape Morris

Extremophiles and Extremotrophs

Some microorganisms thrive in extreme environments. They represent literal extremes as they are capable of surviving at the known limit for life, for any given environmental condition. These environmental conditions exist on a continuous scale and the organisms living within this scale are flanked by polar limits at which life is either physically impossible or selectively futile. Just below or at these limits exist the organisms called extemotrophs[ii] (Mueller, Vincent, Bonilla, & Laurion, 2005). The term extremophiles, coined by R.D MacElRoy (MacElroy, 1974) –which more closely, from an etymological standpoint, relates to classifications such as thermophiles, halophiles, and other such terms that define organisms, with *optimal* growth in a certain physical or chemical environment—by convention describes organisms that thrive or tolerate extreme environments. The two terms are used interchangeably—most often extremophile is used as a blanked term for extremotroph—but an important distinction is that for extremotrophic or extremotolerant organisms, "*a much larger diversity of organisms are known that can tolerate extreme conditions and grow, but not necessarily optimally in extreme habitats"* (Hirokoshi & Bull, 2011). Applicable to both extremophiles and extremotrophs is that from the perspective of the organism, the habitat to which they are well adapted, is only extreme from an anthropocentric, mesophilic point of view. Many psychrophiles would for instance be killed in the optimal temperature range of mesophiles as their proteins become denatured; An extremotrophic or extremophilic adaptive trait does not necessarily suggest homology either. Extremophile (and arguably extremotroph) is thus "*an artificial classification*" as it encompasses a host of different organisms across phylogenetic kingdoms with "*no unique unity"*. (MacElroy, 1974) MacElroy also commented on the increasing line of evidence, that mesophiles may have evolved from extremophiles and not vice versa. Indeed, hydrothermal vents, hosting thermophilic organisms are commonly suggested as a likely starting point for the beginning of life on earth by abiogenesis. (Martin, Baross, Kelley, & et al., 2008) This project seeks to use extreme microorganisms as tools to distance ourselves from our understanding of life as we know it on earth and generate qualitative data to predict and describe signs of life on planets with physical and chemical environments unlike our own.

Psychrophiles, Cryophiles and Halophiles

16S ribosomal subunit

Phylogenetic relationships within the prokaryotes and eukaryotes can reliably be established through comparison of the small subunit ribosomal RNA (SSU rRNA) via the 16s(prokaryotes) and 18s

[ii] Extremotrophs: An alternative definition exists pertaining to organisms which sustain themselves on substances not normally considered edible, relevant to the topic of bioremediation.

(eukaryotes) ribosomal RNA gene respectively. (Woese & Fox, 1977) The SSU rRNA gene is highly conserved and found in all self-replicating organisms. 16s rRNA gene sequencing can be used for prokaryotic identification, as the highly conserved sequence is subject to slow rates of evolution overall. The sequence can therefore function as a molecular clock, with distantly related prokaryotes retaining similar functionality in binding to the Shine-Dalgarno sequence ribosomal binding site. (Malys, 2012) In addition the 16s gene is comprised of nine hypervariable regions. The rate of mutation varies between these hypervariable regions (V1-9). Therefore, the less conserved regions can be used to distinguish between organisms on lower taxonomic levels such as species or even strain. (Pereira, et al., 2010) The 16srRNA gene sequence can then be compared to a database such as the National Center for Biotechnology Information (NCBI)[2]

PCR and Gel Electrophoresis

16s sanger sequencing will not function as intended for most eukaryotes as well as archaea from the order Thermoproteales. (Jay, 2015) Therefore, the presence of the 16s gene can be validated by sequence amplification followed by qualitative gel electrophoresis. Small amounts of DNA can be amplified by polymerase chain reaction (PCR). PCR relies on thermal cycling to repeatedly initiate and terminate the temperature-dependent reaction of DNA polymerization by the heat stable enzyme taq polymerase in a solution of forward and reverse primers as well as free deoxyribose nucleotide triphosphates (dNTPs). The basic stages of a single cycle of PCR are comprised of:

Denaturation: The temperature of the solution is raised to 94-98 °C for 20-30 seconds. At this temperature, the hydrogen bonds between DNA base pairs break causing the DNA to separate into two complimentary single strand DNA (ssDNA).

Annealing: Lowering the temperature just below T_m – the melting temperature by convention, but also the equilibrium temperature for primer annealing and denaturation, often between 50-65 °C—constructed primers, which are short synthetic single stranded RNA molecules, (re)anneal to complimentary sections of ssDNA.

Elongation: The temperature is raised to the optimal temperature, where the polymerase has highest rate of enzymatic activity. At roughly 75-80 °C, taq polymerase ads dNTPs in solution to the 3' end of the ssDNA-primer-complex by a dehydration-condensation reaction of the 5'-phosphate group of the dNTPs to the 3'-hydroxy group of the growing strand. This step proceeds depending on the length of the sequence. As the 16s rRNA gene is roughly 1500bp long and taq polymerase polymerizes at Vmax roughly 1kb per minute at optimal temperature, the 16s copy can be synthesized in 90 seconds.

The three stages of thermal cycling are repeated 20-40 times with each cycle theoretically doubling the concentration of DNA. The concentrated DNA can then be analyzed by gel electrophoresis. In gel electrophoresis a porous gel—agarose gel for nucleic acid solutions—allows DNA fragments to migrate through the pores of the gel, at a rate determined by the size of the molecule, as well as charge by applying an electric field which attracts negatively charged DNA to the positively charged anode. Several samples can be compared by by pipetting each sample into wells at equal distance to the anode, as well as a solution containing DNA of different known lengths called a DNA ladder. By the addition of ethidium bromide, a fluorescent intercalating agent, the results of a run can be visually compared under an ultraviolet lamp and camera setup.

16s Sanger Sequencing

Classical Sanger sequencing utilizes the segregation and visualization of 3'- terminally fluorescent ssDNA by electrophoreses. The ssDNA are in vitro synthesized in a continuum of lengths, due to the random incorporation of chain terminating dideoxy nucleotide triphosphates (ddNTPs) in a solution of dNTPs by DNA polymerase (DNA-pol). Each of the four ddNTP—corresponding to each of the four normal substrate dNTPs in DNA polymerization: dATP, dTTP, dGTP and dCTP—carry a fluorescent label and lack the ribose 3'-hydroxy group. When a ddNTP is added to a growing strand of DNA by DNA-pol the continued addition dNTPs is inhibited as the ddNTP lacks the 3'-OH group normally used in nucleophilic attack of triphosphate group, making the reaction energetically unfavorable.

In traditional Sanger sequencing four separate reactions are run, each with the same query sequence, but loaded with one of the ddNTPs and the four normal dNTPs substrates in a one-hundred-fold higher concentration as well as primers to initiate the polymerization. The four reaction solutions can then be loaded into four separate gel electrophoresis wells corresponding to nucleotides the ddNTP used. Running the gel and separating the replicated DNA by size and length the sequence of the unknown DNA can be manually read based the distance traveled in the gel. (Sanger, Nicklen, & Coulson, 1977)

In practice a method called dye-terminator sequencing is used to reduce cost and processing time. In dye-terminator sequencing each of the ddNTPs is differentially labelled with four fluorescent dyes corresponding to the four ddNTPs. Thereby, the sequencing reaction can be run in a single reaction and can be rapidly sequenced by running the solution by capillary electrophoreses, where a laser causes the fluorescent tags to emit light at different wavelengths. These emissions can be read by a digital fluorescence detector to produce a spectrum graph of fluorescence over time which can be used to determine the order of nucleotides in the sequence.

Illumina MiSeq

Illumina offers a small range of high output sequencers based on flow cell technology. Herein the MiSeq instrument produces a relatively smaller output of roughly 25 million reads per run when compared to the HiSeq or NovaSeq instruments, generating up to 4 billion and twenty billion respectively.[34] Due to the smaller output the MiSeq is relevant for sequencing prokaryotic and yeast genomes within the maximum output of 15 billion base pairs (Gbp).

Overall MiSeq sequencing—as well as the HiSeq and NovaSeq and other similar platforms, only on much larger scales— can be split into the following steps: Generating a sequencing library; Flow cell clustering; and finally simultaneous image-based sequencing of clusters on the flow cell.[5]

To generate a sequencing library, extracted DNA is fragmented either by mechanical stimulation or by restriction enzymes to yield linear, dsDNA fragments with a length of roughly 250bp with blunt ends. Adaptor sequences are added by ligation to both ends of the fragments. The adaptor sequences contain primer binding sites which are initially used to allow polymerization to occur when generating clusters of identical fragments on the flow cell, as well as in a final florescent sequencing reaction. Flanking the primer binding sites of each fragment are two different capture sequences, which can bind to complimentary oligonucleotides that coat the flow cell. The flow cell is a hollow glass side with one or more channels lanes.

The insert-fragment molecules (templates) are denatured to form single stranded molecules which are then loaded onto the flow cell, whereby one of the two different capture sequences bind to a complimentary anchored oligonucleotide. Roughly half of the sequences will bind with the other capture sequence and will therefore be in the opposite orientation producing paired reads, although in many cases only a single template strand will be captured, producing single end reads. DNA-pol and dNTPs are added to the flow cell and a polymerization reaction is induced by temperature regulation. The newly synthesized strand, which is a copy of the template is now anchored to the flow cell. The template-copy molecule is denatured, and the template strand is washed away leaving only the anchored strand, which now presents the capture sequence for the other flow cell oligonucleotide anchor on the unanchored free 3’-end. The strand is allowed to bind to another anchor and the polymerization followed by denaturing step is repeated yielding two anchored complimentary strands. The process is repeated until roughly 1000 copies are generated in a cluster. After using one of two oligonucleotide anchors as selective restriction target the flow cell is washed to remove the cleaved sequences. A sequencing primer is added and binds to the sequences still anchored, which are all identical strands in the same orientation. Generating the many clusters on the flow cell is automated after loading the DNA samples and reagents.

Finally, a sequencing by synthesis reaction is run under a microscope using Fluorescent Reversible Terminator Chemistry in a process comparable to Sanger Sequencing. 3′-O-allyl-dNTP-allyl-fluorophores use 4 different fluorescent allyl-groups corresponding to the four dNTPs as in sanger sequencing. However, the fluorescent allyl-group as well as the chain terminating 3’-allyl group can be removed after incorporation, allowing synthesis to begin again. 3′-O-allyl-dNTP-allyl-fluorophores are added to the flow cell and DNA-pol initiates polymerization, adding a single 3′-O-allyl-dNTP-allyl-flourophore to the primed and clustered DNA molecules anchored to the flow cell. Each cluster fluoresces one of four different colors. The fluorescent signal of all clusters is simultaneously registered by a camera through a microscope, and PHRED software produces a sequence trace file(chromatogram) to automatically call bases in the sequence as well as assign a quality value to each base. The allyl groups of the 3′-O-allyl-dNTP-allyl-fluorophores are cleaved leaving a 3’-OH group ribose on the dNTPs. 3′-O-allyl-dNTP-allyl-flourophores are added to the flow cell again and the single polymerization reaction is repeated until all roughly 250bp of each cluster have been read.

The most prominent limitation in of the simultaneous sequencing by synthesis of roughly a thousand strands is lagging and premature polymerization of some strands compared to the sequencing step of a cluster. As the sequencing process progresses, an increasing number of strands either add an additional 3′-O-allyl-dNTP-allyl-flourophore ahead of the other strands, or the addition of the 3′-O-allyl-dNTP-allyl-flourophore does not occur, causing the strand to lag. These strands may display the wrong fluorescent signal according to the template sequence. As these types of mistakes accumulate, the signal read by the camera becomes increasingly indistinguishable. This leads to the limit of fragment sequencing length of roughly 250bp before the read quality falls below an acceptable threshold which necessitates quality control and removal of the low-quality bases as well as the leftover adaptor sequences, before the sequence data can be used for genome assembly.

FASTQ

FASTQ format is the *de facto* standard text-based format for storing the output from Illumina sequencers. FASTQ format is based on the combination of FASTA file format for biological sequence data—most often nucleotide sequences but can also be applied to amino acid sequences—as well as the American Standard Code for Information Interchange (ASCII), which is used to encode a single quality score character to each nucleotide based on the PHRED quality score (Q_{PHRED}). The Q_{PHRED} value based on the probability of error (P_e) for the base called on the sequence trace file (Cock, Fields, Goto, Heuer, & Rice, 2010):

$$Q_{PHRED} = -10 \times log_{10}(P_e)$$

The FASTQ format is useful as it can compressed as a gzip file and directly piped into a quality control program such as FASTQC to generate an overview of the more than one million reads each of roughly 300 bases for each sequenced organism on the MiSeq platform.

FastQC Version 0.11.9

Trimmomatic Version 0.39

At the date of processing, these were the most up to date versions of the programs. Occasionally newer versions have unintended bugs and may no longer be viable for a given set of data.

Method

Isolation (Credit to Miguel)

Approximately 1g of sample taken from either lithic, crust or permafrost from terrestrial analogue (TA) sites were mixed with 1.00 ml sterile PBS + NaCl in 2 ml Eppendorf tubes. Tubes were gently inverted repeatedly for 5 minutes to separate cells from soil particles. The tubes were then centrifuged at 5000rpm for 5 minutes to precipitate soil particles. 100 µL supernatant was then plated and left to grow at in the medium and temperature indicated in supplemental data table X. Plates were controlled biweekly for new colonies. Colonies were streaked in the same medium as the original plate at least twice to obtain pure isolates. Isolates were then inoculated into new liquid media in order to optimize for rate of growth. Once optimal media was selected, the isolates were grown until late exponential phase and prepared for short term or long-term storage before future DNA extraction. For short-term storage 1.00mL of liquid culture was centrifuged to form a pellet, which was frozen at -20°C. For long-term storage 0.50mL liquid culture was mixed with 0.50mL 40% sterile glycerol solution (final concentration 20% glycerol) and stored at -80°C. Alternatively for agar plate cultures the gel is covered with 4-5ml of 20% solution. Colonies are scraped off the gel and resuspended in the liquid glycerol phase and extracted for centrifugation followed by storage at -80°C.

Extraction

DNA was extracted from each isolate by using the FastDNA^{TM} SPIN Kit for Soil (MP Biomedicals, LLC., Irvine, USA). The following modification to the protocol was conducted. Instead of using up to 500 mg of soil sample, the isolated samples frozen for storage were first gently thawed and resuspended with a tabletop vortex. Besides this first step, the product instruction manual protocol was followed.

Polymerase Chain Reaction and Gel Electrophoresis

The potential 16s sequences of the extracted isolate DNA were amplified by Polymerase Chain reaction (PCR) with PCRBIO HIFI Polymerase kit (PCR Biosystems Ltd., London, UK). The product manual was followed using 27f and 149r 16s universal primers. To confirm the presence of the 16s gene (and thereby validate the presence of prokaryotic DNA in the sample: Do mitochondrial 16s sequences within eukaryotic cells not give bands when sequencing? Can’t find any information on this) before sending isolate DNA for Sanger sequencing.

Preparation for Sanger Sequencing

Preparation for MiSeq

DNA concentration was calculated from the average bp output of a Fragment Analyzer (Results: table X) to ensure sufficient concentration for Sanger 16s, and MiSeq sequencing. The DNA concentration was calculated with the following formula:

$$Desired\ Library\ molarity = 4nM$$

$$DNA[ng * \mu L^{-1}] = Molarity[nM] \times 660\frac{g}{mol} \times mean\ size(pb) * 10^{-6}\frac{ng}{g},$$

(This equation doesn't make sense: by adding the desired library molarity of 4nM to the equation the output is not a concentration but a ratio from the desired molarity. I cannot find Chiara who made the table in the system to ask her how she arrived at this equation. I also don't find many sources recommending a DNA molecular mass of 660.[iii])

Based on the derived DNA concentrations (Results: Table 2) the isolates CF4.2, A.5 and A.10 were selected for whole genome sequencing

Data Processing

Quality Control and Trimming

Running every part of the bioinformatic analysis on the Galaxy scientific workflow system was technically possible, finding optimal trimming parameters for each of the three isolate samples chosen for sequencing required many iterations of slightly modified trimming criteria. Frequent crashes and slow run times on the Galaxy servers necessitated abandoning the platform for quality control and trimming of the sequence data. These operations were instead conducted in the University of Copenhagen Electronic Research Data Archive (ERDA UCPH). In addition to functioning as a centralized storage space, ERDA allows access to Jupyter data analysis services by a Data Analysis Gateway (DAG) with access to 8 threads and 16GB memory on remote server. From here a Linux Terminal could be launched, allowing data processing after initiating the conda package manager for bioinformatics software via Bioconda. (See Terminal Procedure)

Having configured the terminal, FastQC and Trimmomatic was installed, and initial quality control reports were output to html. Below is the are the general considerations and modification that need to be made based in the FastQC Reports. For all FastQC reports and specific modification see Supplementary Data: Initial FastQC Reports.

Basic Statistics

Each basic statistics table briefly indicates whether the file upon which the FastQC Report is made, is a valid file type and identifies the type sequencing performed, shown on the Encoding row. The basic statistics also show the total number and average length of sequences, as well as the number of sequences flagged as poor quality and total GC content. All runs passed the basic statistics. However, stark differences in GC content are considered in the discussion section.

Quality Control Report table

Per base sequence quality

Following an initial lower quality, all sequences show a relatively high quality per position in read, with the blue mean line falling under 28 (quality score) after 250bp. The initial lower quality was ignored as a head crop was applied to all sequences

[iii] https://www.thermofisher.com/dk/en/home/references/ambion-tech-support/rna-tools-and-calculators/dna-and-rna-molecular-weights-and-conversions.html

Per tile sequence quality

The per tile sequence quality (PTSQ) is a quality score heatmap of the flow cell tiles. A failure (red tile) is issued when any tile has a mean Phred score less than 5 compared to the mean of the base across all tiles. The PTSQ was largely disregarded, as most of the runs passed (blue tiles) or contained isolated failed tiles, indicating that the flow cells were not contaminated by smudges, debris, or bubbles. A.10 however, showed an unusual pattern of two parallel blotches in the rough position range of 6-9 bp and again in 45-50.

Per sequence quality score

Only A.10 Forward quality control issued a warning. (See Supplementary data: Initial FastQC Reports: A.10 Forward) In an optimal distribution of scores, the vast majority of sequences should have a high mean Phred score. In the case of A.10, the distribution of follows a normal distribution. This indicates that applying a high minimum threshold quality when trimming the A.10 forward sequence may remove many sequences and result in low depth when mapping.

Per base sequence content

All runs for all samples failed, due to low quality in the first 10-15 bases. This is expected from MiSeq sequencing and was resolved by applying a head crop optional argument when trimming.

Per sequence GC content

CF4.4 forward run failed the per sequence GC content quality control. This is likely because the isolate belongs to the phylum Actinomycetota which are high G+C content gram positive bacteria, based on the closest 16s BLAST hit *Nesterenkonia sandarika*. (See Supplementary data: Sequences Samples; 16s sequence data)

Per base N content

N base call substitutions occur when the sequencer is unable to call the base with sufficient confidence. A.10 showed significant per base N content in the 6-45 posion of reads, corresponding to the failed per tile sequence quality of that same region.w

Sequence length distribution

All sequences had an exact length of 301bp, indicating that no lagging or leading occurred during the synthesis reactions.

Sequence duplication levels

High levels of duplication may indicate enrichment bias by for instance PCR over amplification. This However was not seen in any of the run reports and all runs passed the quality control

Overrepresented sequences

No overrepresented sequences were found indicating diverse libraries and no contamination.

Adapter content

As expected, increasing adapter content was found for all runs. The Nextera Transposase Sequence was identified and was removed by adding an adapter argument to the trim function. The Nextera Transposase Sequence[6] was found and saved to a text file to be accessed by Trimmomatic (See Supplementary Data: Terminal Procedure)

Galaxy Workflow: Mapping to Reference and Calling Consensus sequence for genome assembly

Running minimum six slightly different Trimmomatic commands on the raw sequences produced a total of 25 different trimmed sequence files that were all valid for assembly. These renditions were generated in order to select an optimal Trimmomatic command for each raw sequence. (See Supplementary Data: Optimizing Trim Settings) The trimmed sequences were processed with the same following galaxy workflow:

NCBI's nucleotide BLAST is run for each of samples 16s sequences. The closest matching organism with an entire genome was selected to serve as a reference genome for mapping. (See Supplementary data: Sequences Samles; 16s sequence data) The selected genomes are show below in

Table 1

Sample ID	**Reference organism**	**Link to Genome assembly**
CF4.4	*Nesterenkonia sandarakina*	https://www.ncbi.nlm.nih.gov/assembly/GCF_013410215.1
A.6	*Bacillus cereus*	https://www.ncbi.nlm.nih.gov/assembly/GCF_018309165.1
A.10	*Bacillus halotolerans*	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/004/006/435/GCF_004006435.1_ASM400643v1/

Table 1: reference genome. Links to assemblies accessed and validated 6th of June 2022.

Each sample's reference organisms genomic fasta file was downloaded and uploaded to a galaxy history containing the first and second, paired and unpaired trimmed sequences (See Supplementary Data: Terminal Procedure). Each read is mapped to the reference genome using BWA-MEM (Galaxy version 0.7.17.2) (Li & Durbin, 2010) with Simple Illumina mode and sorting by chromosomal coordinates. The Map function produces three mapped BAM files for each sample: 1 with both paired reads and 1 for each of the unpaired files. The 3 mapped files are sorted with Samtools Sort (Galaxy version 2.0.3) (Li, et al., 2009) using coordinate as primary sort key, producing three sorted files. All three sorted files were then merged using Samtools merge (Galaxy version 1.13) The merge file contains all the reads overlayed the reference genome with varying depth and due to some incorrectly called bases different bases indicated for the same position. Ivar consensus was run to call the consensus sequence from the merge file alignment. (Grubaugh, et al., 2019) The consensus call was run with minimum quality score threshold to count base at 1, as the sequences were already trimmed with a quality threshold. Minimum frequency threshold set to 0 for majority or most common base at a given position. Minimum depth to call consensus was set to 1, as there should not be a reason not to trust a base call for a single position due to the length and high quality of the sequencing data. Finally, Prokka prokaryotic genome annotation was run on the consensus sequences to yield annotated genomes, (Cuccuru, et al., 2014) (Seemann, 2014) The histories containing all of the mentioned files were then shared and can be accessed via the link in Results: Table 3 Galaxy Histories

Results

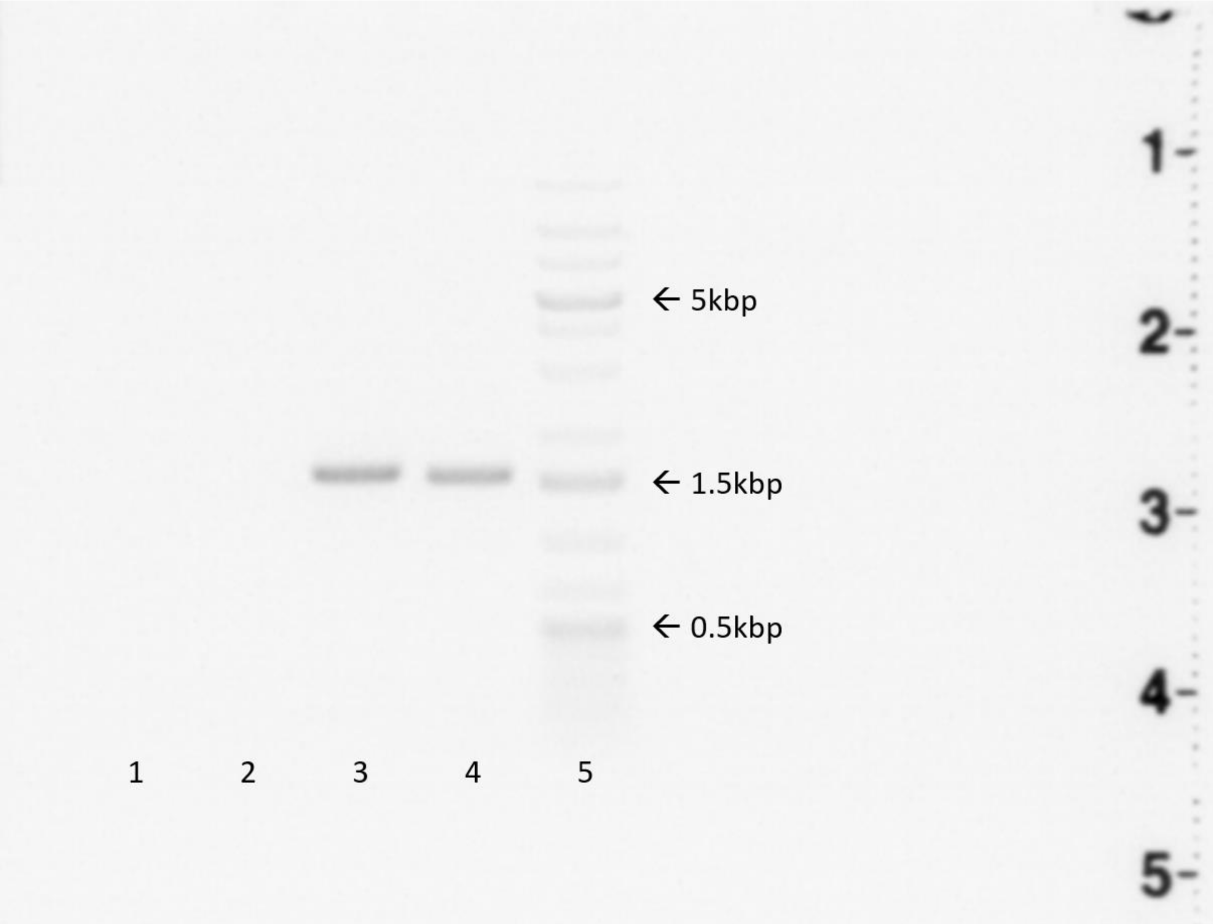
Table 2 Fragment Analyzer Data

Sample	CF4.1	CF4.2	CF4.4	A.5	A.6	A.9	A.10	A.11

Average size (bp)	-	672	1027	1099	668	883	642	-
ng/µL	-	1.77	2.71	2.90	1.76	2.33	1.69	-
note	Error	Insufficient purification	Ok	Ok	Ok	Ok	Ok	Error

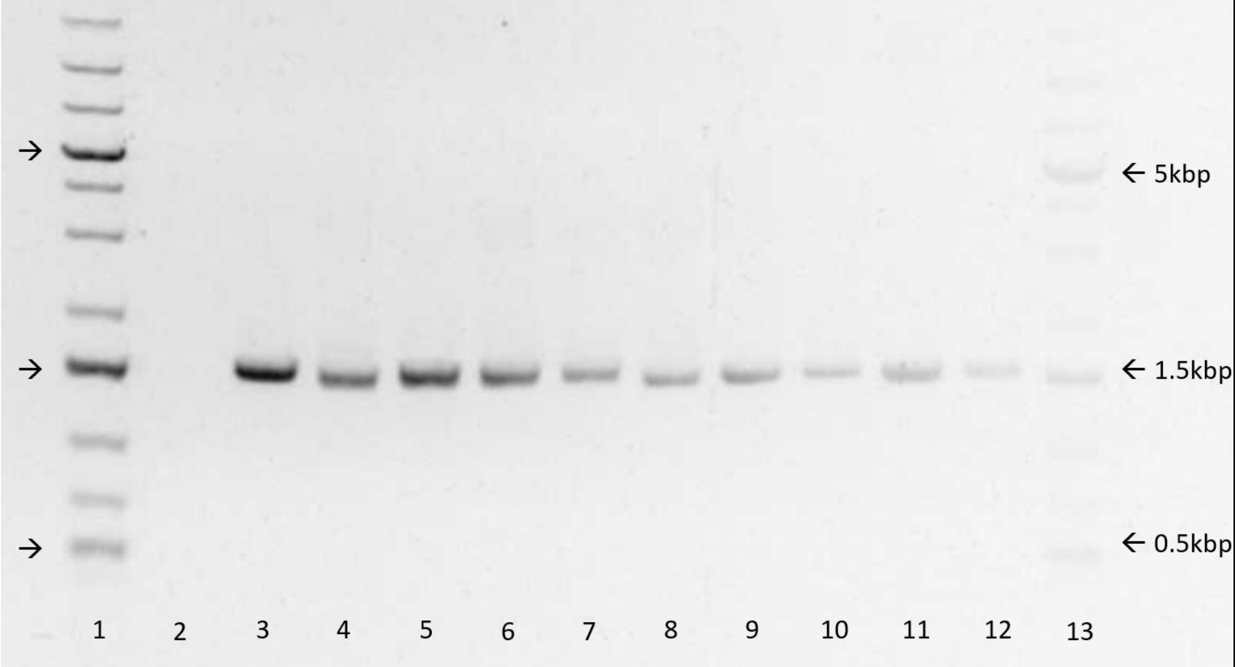
Table 1: Measured Fragment Analyzer DNA extraction concentrations of first sample cohort

Figure 1: First cohort gel electrophoresis of PCR product of 16s amplification



1: empty **2**: blank (water) **3**: CF4.2 **4**: CF4.4 **5**: ladder: GeneRuler: 1kb. **Arrows**: 5kbp, 1.5kbp, 0.5kbp descending

Figure 2 second cohort electrophoresis of PCR product of 16s amplification



1: ladder **2**: blank(water) **3**: CF1.1 **4**: CF3.3 **5**: CF4.1 **6**: CF4.5 **7**: CF4.6 **8**: CF4.9 **9**: CMS3.1 **10**: KK4.2, **11**: KK5.1 **12**: KK6.1 **13**: ladder GeneRuler: 1kb. **Right and left arrows**: 5kbp, 1.5kbp, 0.5kbp descending order.

Table 3 Galaxy Histories

Galaxy history description	Link
CF4.4 mapped to *Nesterenkonia aurantiaca*	https://usegalaxy.org/u/hankculp/h/224-n-content-nesterenkonia-aurantiaca-strain-dsm-27373
CF4.4 mapped to *Nesterenkonia sandarakina*	https://usegalaxy.org/u/hankculp/h/cf44-s328h15q20d1
A.6 mapped to Bacillus cereus	https://usegalaxy.org/u/hankculp/h/a6-s328h15
A.10 mapped to Bacillus halotolerans	https://usegalaxy.org/u/hankculp/h/a10-s320h15
Generalized workflow to be applied on future sequenced isolates	https://usegalaxy.org/u/hankculp/w/workflow-constructed-from-history-a10-s320h15

Discussion

Optimizing Assembly

References

Al Soudi, A. F., Ferhat, O., Chen, F., Clark, B. C., & Schneegurt, M. A. (2017). Bacterial Growth Tolerance to Concentrations of Chlorate and Perchlorate Salts Relevant to Mars. *International Journal of Astrobiology*, 16, (3), 229-235 .

Cock, P., Fields, C., Goto, N., Heuer, M., & Rice, P. (2010). The Sanger FASTQ file format for sequences with wuality scores, and the Solexa/Illumina Fastq variants. *Nucleic acids research, 39(6)*, 1767-1771.

Cuccuru, G., Orsini, M., Pinna, A., Sbardellati, A., Soranzo, N., Travaglione, A., . . . Fotia, G. (2014). Orione, a web-based framework for NGS analysis in microbiology. *Bioinformatics*, 1928-1929.

Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., . . . Andersen, K. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology*, 8-8.

Hecht et al. (2009). Detection of Perchlorate and the Soluble Chemistry of Martian Soil at the Pheonix Lander Site. *Science*, 325,5936, 64-67.

Hirokoshi, K., & Bull, A. (2011). Prologue: Definition, Categories, Distribution, Origin and Evolution, Pioneering Studies, and Emerging Fields of Extremophiles. In H. K. (eds), *Extremophiles Handbook* (pp. 3-15). Tokyo: Springer.

Jay, Z. I. (2015). Thr distribution, diversity, and importance of 16s rRNA gene introns in the order Thermoproteales. *Biol Direct 10, 35*.

Lagerström, M., Parik, J., Malmgren, H., Stewart, J., Pettersson, U., & Landegren, U. (1991). Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. *Genome Res*, 1: 111-119.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 2078-2079.

MacElroy, R. (1974). Some Comments on the Evolution of Extremophiles. *Biosystems*, 74-74.

Madsen, P. K. (2020). *Subjecting Bacteria to the Extremes of Mars.* CPH.: University of Copenhagen.

Malys, N. (2012). Shine-dalgarno sequence of bacteriophage T4: GAGG prevauls in early genes. *Molecular Biology Reports*, 39 (1), 33-39.

Martin, W., Baross, J., Kelley, D., & et al. (2008). Hydrothermal vents and the origin of life. *Nat Rev Microbiol 6*, 805-814.

Mueller, D., Vincent, W., Bonilla, S., & Laurion, I. (2005). Extremotrophs, Extremophiles and Broadband Pigmentation Strategies in High Arctic Ice Shelf Ecosystems. *FEMS Microbiology Ecology*, 73-87.

Navarro-Gonzalez, awda, aaesfa, aesfaesf, asefasef, asefasf, . . . asfaafaesfw. (2006). The Limitations on Organic Detection in Mars-Like Soils by Thermal Volatilization-Gas Chromatography-MS and

Their Implications for the Viking Results. *Proceedings of the National Academy of Sciences - PNAS*, 103, (44) 16089-16094.

Navarro-González, R., Rainey, F., Molina, P., Bagaley, D., & McKay, C. (2003). Mars-Like Soils in the Atacama Desert, Chile, and the Dry Limit of Microbial Life. *Science*, 1018-1021.

Oyama, V., & Bjerdahl, B. (1997). The Viking Gas Exchange Experiment results from Chryse and Utopia surface samples. *Journal of Geophysical Research*, 82, 28 4669-4676.

Pereira, F., J, C., R, M., B, v. A., N, P., L, G., & A, A. (2010). Identification of species by multiplex analysis of variable-length sequences. *Nucleic Acids Research*, 38 (22).

Sanger, F., Nicklen, S., & Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74 (12): 5463–5467.

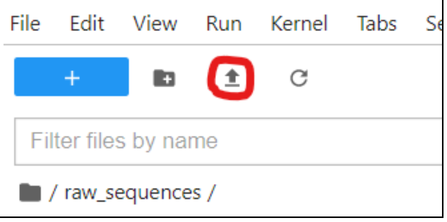
Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 2068-2069.

Woese, G. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America 74 (11)*, 5088-5090.

Supplementary Data

Terminal Procedure

Table 4: Terminal Commands and Comments

Commands	**Comments**
conda config --add channels defaults conda config --add channels bioconda conda config --add channels conda-forge	Initiate conda
mkdir raw_seq cd raw_seq	Create a directory to store the raw sequences (raw_seq) as well as a directory for the initial quality control (initial_qual), navigate to the raw_seq directory and upload MiSeq raw data to the directory via the directory browser upload icon (image below, red circle)  File Edit View Run Kernel Tabs Se + Filter files by name / raw_sequences /

conda install fastqc y conda install trimmomatic y	While the sequences are being uploaded install FastQC and Trimmomatic, accept prompt to continue #download (y)
fastqc ~/raw_seq* -o initial_qual	Once the the programs are installed, first ensure that all sequence files have been uploaded, then navigate to initial_qual directory and execute FastQC on all raw sequence files and save the output to the initial_qual directory
mkdir trimmed_cf4_4 cd trimmed_cf4_4 mkdir S3:28 cd mkdir trimmed_a_6 cd trimmed_a:6 mkdir S3:28 cd mkdir trimmed_A_10 cd trimmed_a_10 mkdir S3:20 cd	create directory to store the corresponding trimmed sequences each with a subdirectory specifying the trim settings
cd raw_sequences nano	Navigate to the raw sequence file and create a text file containing the adapter sequence by pasting the following into the nano file and saving as “adapters.fasta”: >Nextera_XT CTGTCTCTTATACACATCT
trimmomatic PE -threads 8 ~/work/raw_sequences/Run20211209Order375Sample003_S188_L001_R1_001.fastq.gz ~/work/raw_sequences/Run20211209Order375Sample003_S188_L001_R2_001.fastq.gz ~/work/trimmed_cf4_4/S3\:28_H15/cf4_4r1_paired.fastq.gz ~/work/trimmed_cf4_4/S3\:28_H15/cf4_4r1_unpaired.fastq.gz ~/work/trimmed_cf4_4/S3\:28_H15/cf4_4r2_paired.fastq.gz ~/work/trimmed_cf4_4/S3\:28_H15/cf4_4r2_unpaired.fastq.gz ILLUMINACLIP:adapters.fasta:2:30:10 SLIDINGWINDOW:3:28 HEADCROP:15 trimmomatic PE -threads 8 ~/work/raw_sequences/Run20211209Order375Sample005_S191_L001_R1_001.fastq.gz ~/work/raw_sequences/Run20211209Order375Sample005_S191_L001_R2_001.fastq.gz ~/work/trimmed_a_6/S3\:28_H15/a_6r1_paired.fastq.gz ~/work/trimmed_a_6/S3\:28_H15/a_6r1_unpaired.fastq.gz ~/work/trimmed_a_6/S3\:28_H15/a_6r2_paired.fastq.gz	Run the trim commands. Commands are each one line Once Trimmomatic has completed the command the paired and unpaired, first and second runs are available to download in the from the file explorer for each respective sample. These are then loaded into a galaxy history

~/work/trimmed_a_6/S3\:28_H15/a_6r2_unpaired.fastq.gz ILLUMINACLIP:adapters.fasta:2:30:10 SLIDINGWINDOW:3:28 HEADCROP:15 trimmomatic PE -threads 8 ~/work/raw_sequences/Run20211215Order375Sample007_S193_L001_R1_001.fastq.gz ~/work/raw_sequences/Run20211215Order375Sample007_S193_L001_R2_001.fastq.gz ~/work/trimmed_a_10/S3\:20_H15/a_10r1_paired.fastq.gz ~/work/trimmed_a_10/S3\:20_H15/a_10r1_unpaired.fastq.gz ~/work/trimmed_a_10/S3\:20_H15/a_10r2_paired.fastq.gz ~/work/trimmed_a_10/S3\:20_H15/a_10r2_unpaired.fastq.gz ILLUMINACLIP:adapters.fasta:2:30:10 SLIDINGWINDOW:3:20 HEADCROP:15	

Optimizing Trim settings

Table 5: Optimal Headcrop in CF4.4

Trimmomatic Settings			**Ivar Settings**		**Resutls**		
SLIDINGWINDOW	HEADCROP	MINLEN	Min Quality	Min Depth	Positions with 0 Depth	Length	N%
1:28	0	-	20	1	0	3009916	28.25
1:28	10	-	20	1	0	3009929	28.44
1:28	13	-	20	1	0	3009886	28.49
1:28	14	-	20	1	0	3009928	28.55
1:28	15	-	20	1	0	3009983	28.54
1:28	16	-	20	1	0	3009983	28.55
1:28	17	-	20	1	0	3009969	28.58

Figure 3 relationship between head crop length and genome N% in CF4.4

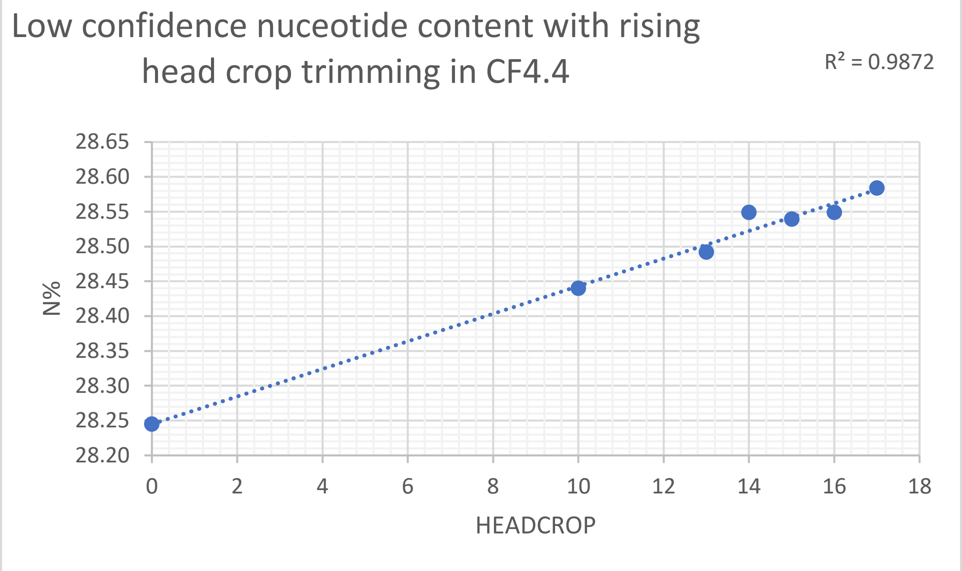


Figure3: Increasing head crop length correlates to a linear increase (y = 0.0198x + 28.245) in the percentage of genomic unknown base calls N%

Table 6 Optimal SLIDINGWINDOW Frame in CF4.4

N% with rising sliding window frame in CF4.4			
SLIDINGWINDOW	Sequence length	Number of N indicated base calls	N%
1	3009928	859024	28.54
2	3009872	847279	28.15
3	3009889	843226	28.02
4	3009889	839938	27.91
5	3009934	837557	27.83
6	3009941	835781	27.77

Figure 4 Optimal SLIDINGWINDOW Frame in CF4.4

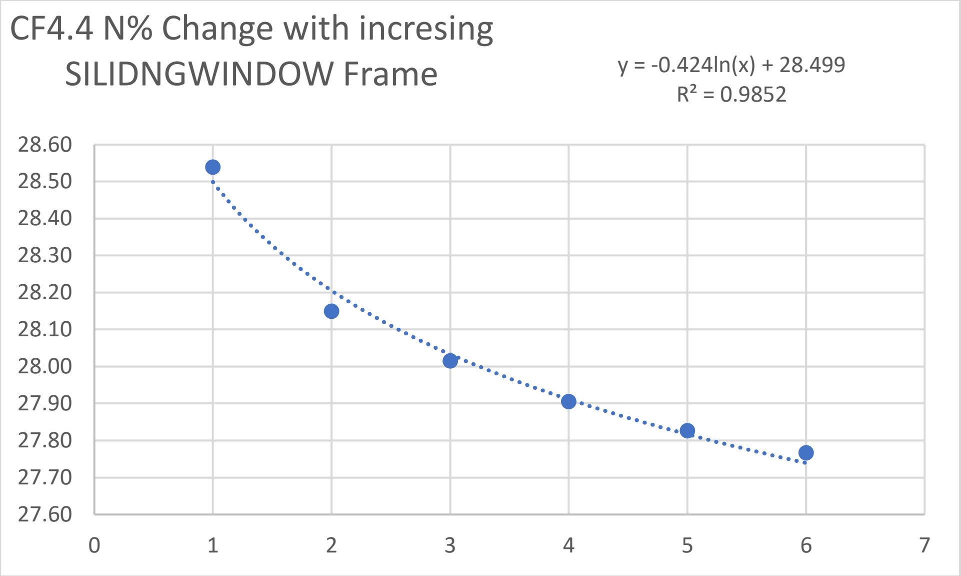


Figure 4: exponentially decreasing N% with increasing frame setting for CF4.4 at threshold quality: 28

Table 7 Optimal SLIDINGWINFOW Frame in A.6

N% with rising sliding window frame in A.6			
SLIDINGWINDOW	Sequence length	Number of N indicated base calls	N%
1	5413962	551670	10.19
2	5413938	549349	10.15
3	5413876	548618	10.13
4	5414604	548641	10.13
5	5414566	548391	10.13
6	5414504	547836	10.12

Figure 5 Optimal SLIDINGWINDOW Frame in A.6

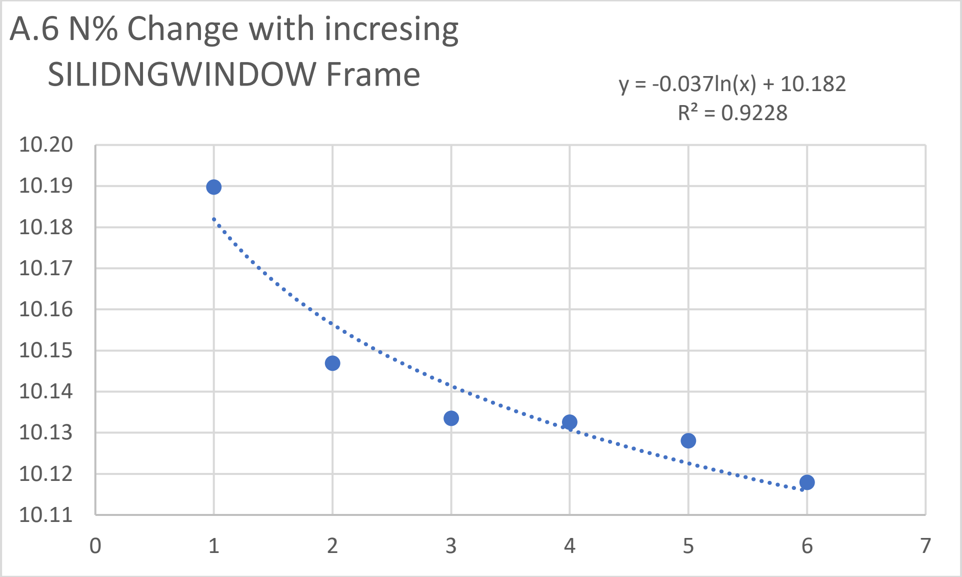


Figure 5: exponentially decreasing N% with increasing frame setting for A.6 at threshold quality: 28

Table 8 Optimal SLIDINGWINDOW Frame in A.10

N% with rising sliding window frame in A.10			
SLIDINGWINDOW	Sequence length	Number of N indicated base calls	N%
1	4152840	549347	13.23
2	4152771	546519	13.16
3	4152944	544732	13.12
4	4152889	544732	13.12
5	4152912	542549	13.06
6	4152945	541981	13.05

Figure 6 Optimal SLIDINGWINDOW Frame in A.10

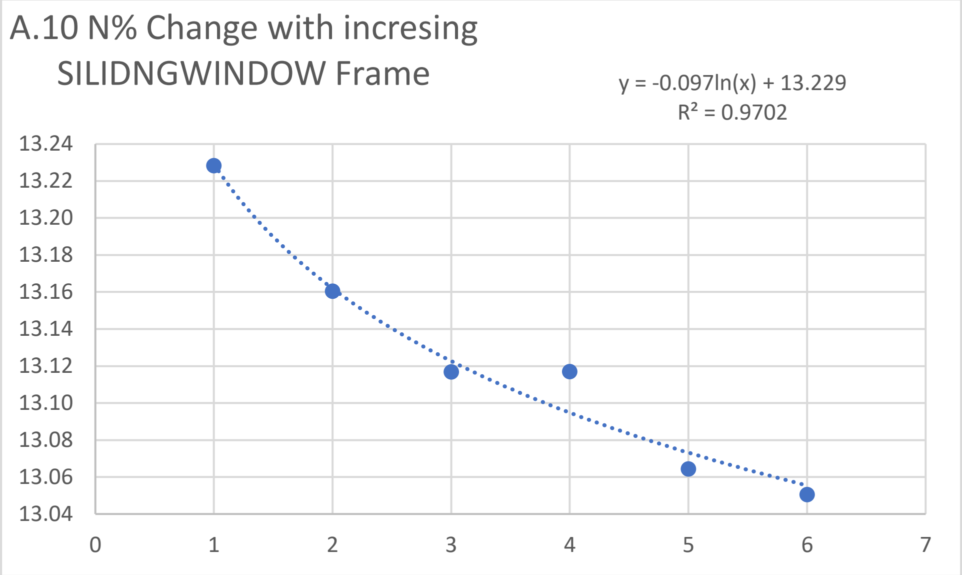


Figure 7: exponentially decreasing N% with increasing frame setting for A.10 at threshold quality: 20

Initial FastQC Reports

The Following subsections are reports of the raw sequence data. Each section features a basic statistics table with information about the file. Filenames correspond to the sample ID and run listed below in Table 9: Sequence file name key. Each run has report table indicating each property controlled as well as the pass, fail or warning grade under Qualitry Control Measure. Properties tagged as [Fail] or [Warning] are followed by additional comments with Trimmomatic trimming criteria to improve quality and pass the quality control.

Table 9: Sequence file name key

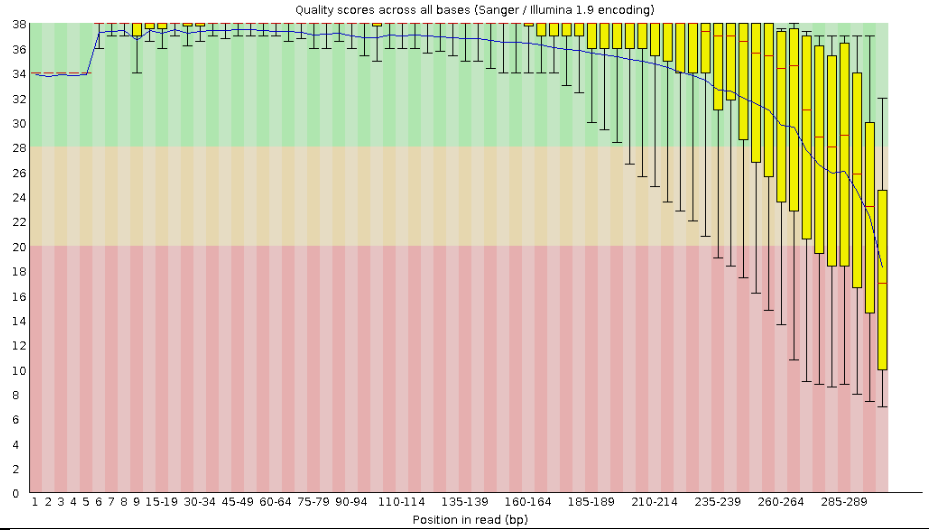
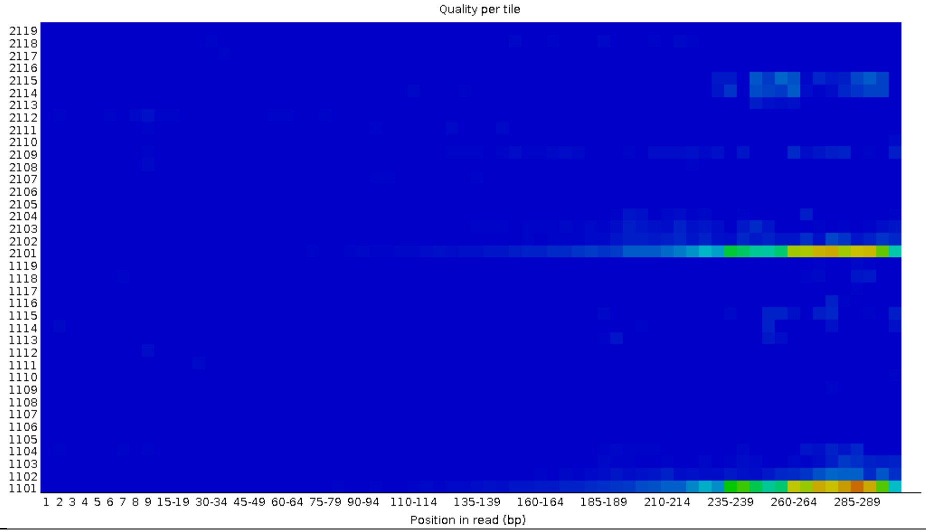
Sequence file name	Sample ID and Flow Cell read
Run20211209Order375Sample003_S188_L001_R1_001.fastq.gz	CF4.4: Forward
Run20211209Order375Sample003_S188_L001_R2_001.fastq.gz	CF4.4: Reverse
Run20211209Order375Sample005_S191_L001_R1_001.fastq.gz	A.5: Forward
Run20211209Order375Sample005_S191_L001_R2_001.fastq.gz	A.5: Reverse
Run20211215Order375Sample007_S193_L001_R1_001.fastq.gz	A.10: Forward
Run20211215Order375Sample007_S193_L001_R2_001.fastq.gz	A.10: Reverse

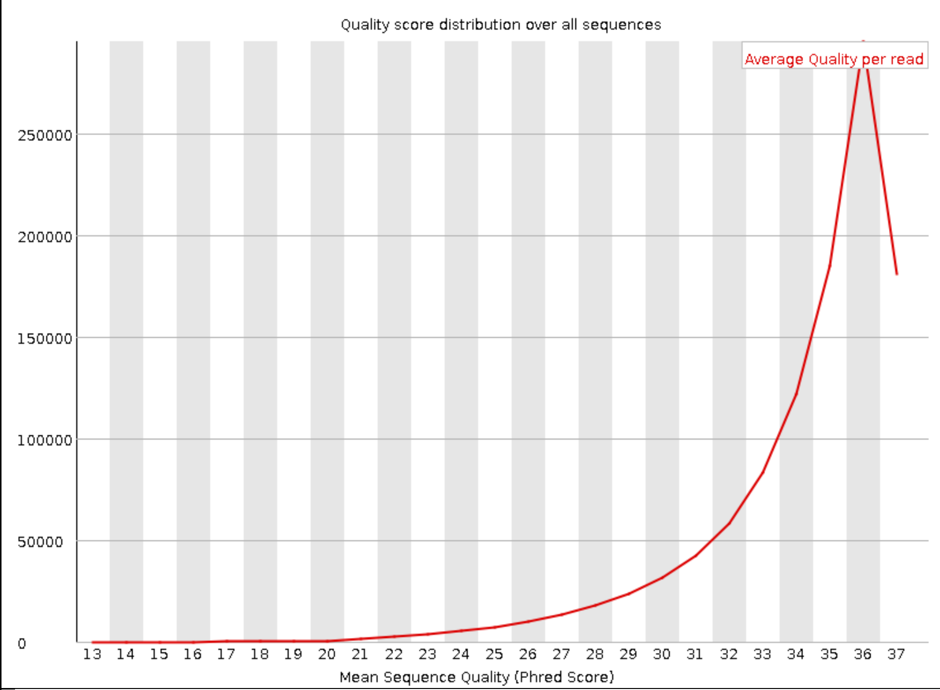
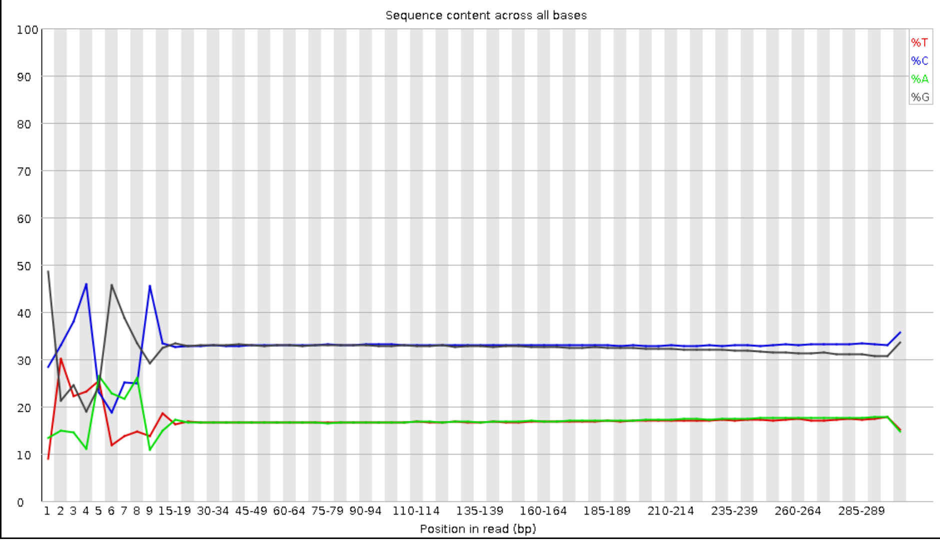
CF4.4: Forward

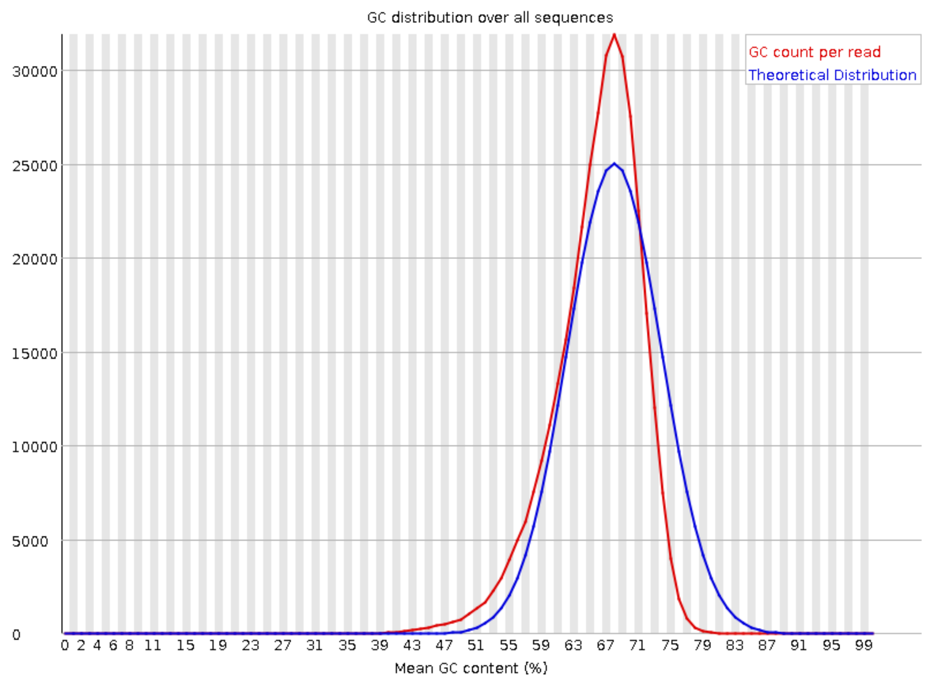
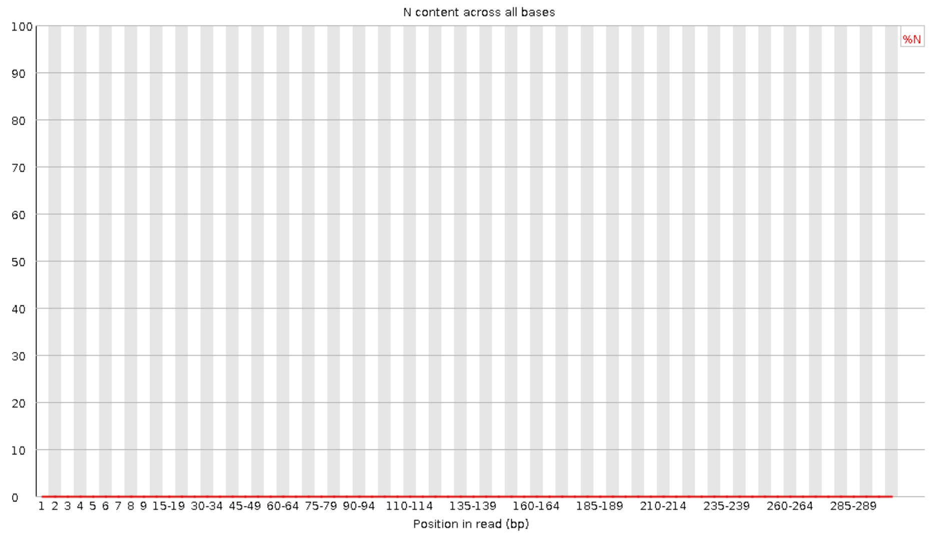
Basic Statistics

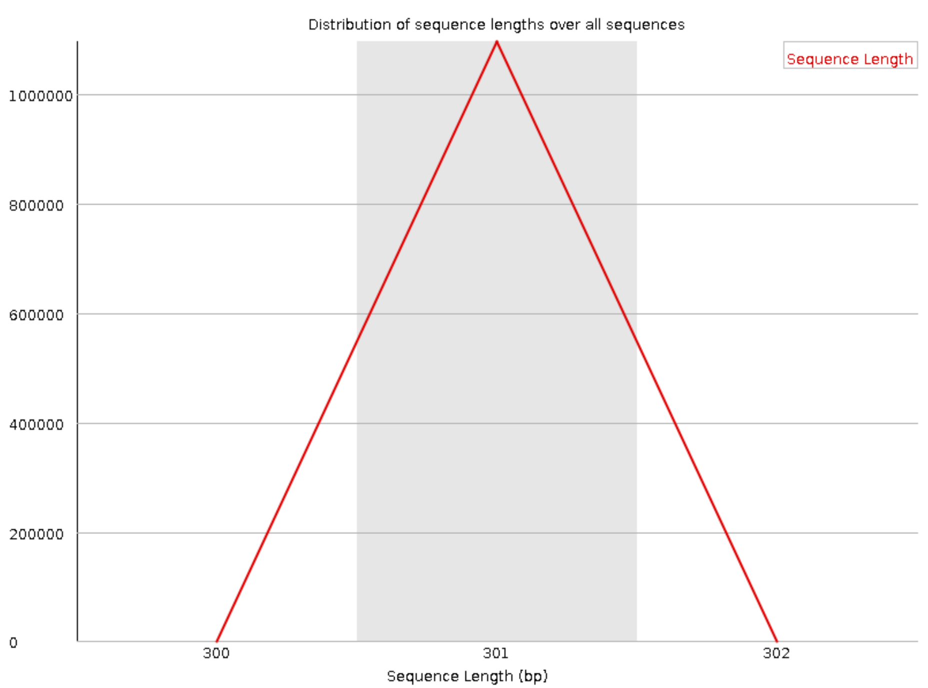
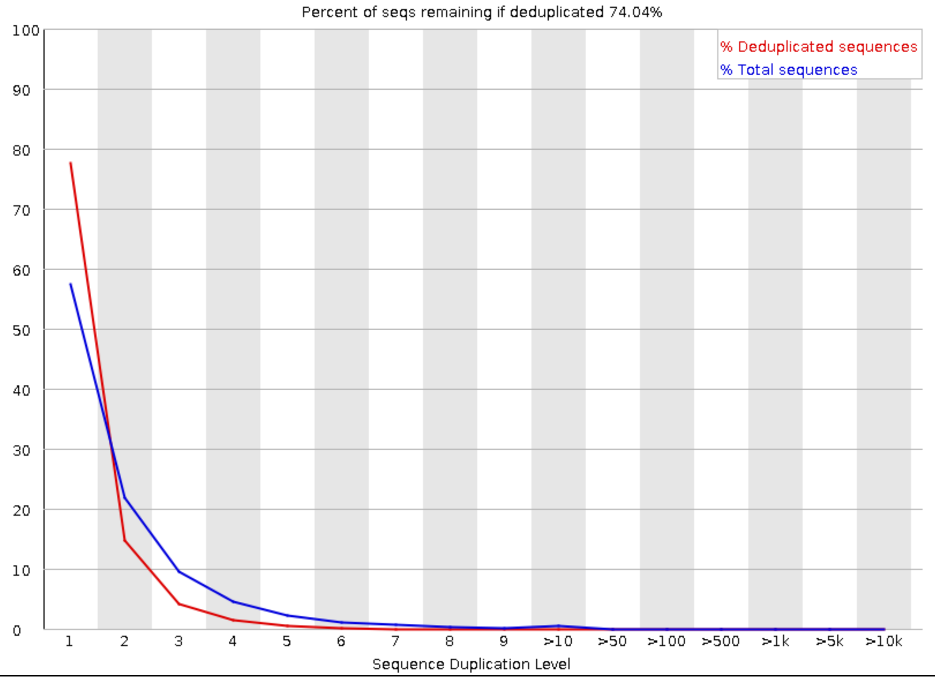
Measure	Value
Filename	Run20211209Order375Sample003_S188_L001_R1_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9

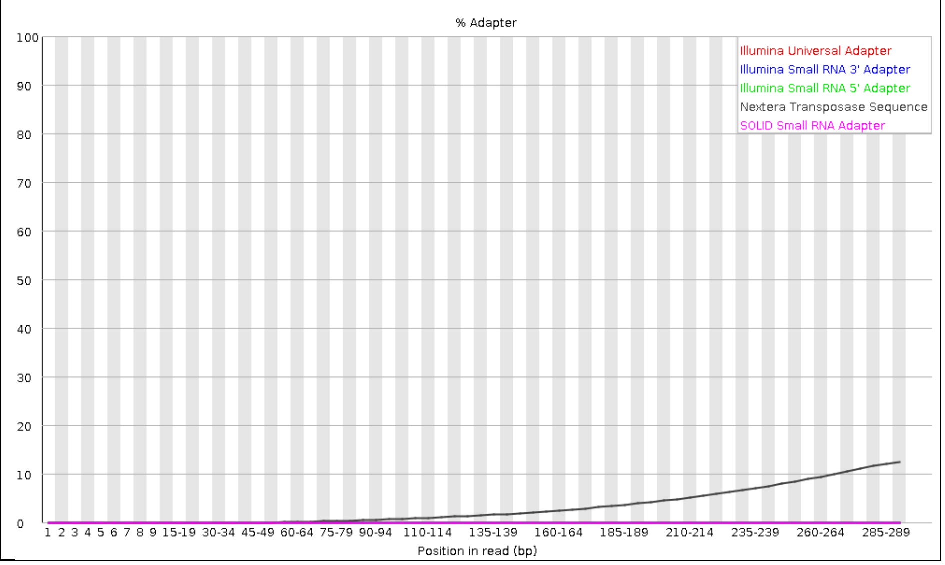
Total Sequences	1095679
Sequences flagged as poor quality	0
Sequence length	301
%GC	65

Quality Control Measure	**Figure**
[FAIL]Per base sequence quality Threshold quality: 28	
[WARNING]Per tile sequence quality Isolated lane warning issues may be ignored	

[PASS]Per sequence quality scores	
[FAIL]Per base sequence content Head crop: 15	

[FAIL]Per sequence GC content Not likely due to read error	
[PASS]Per base N content	

[PASS]Sequence Length Distribution	
[PASS]Sequence Duplication Levels	
[PASS]Overrepresented sequences	No overrepresented sequences

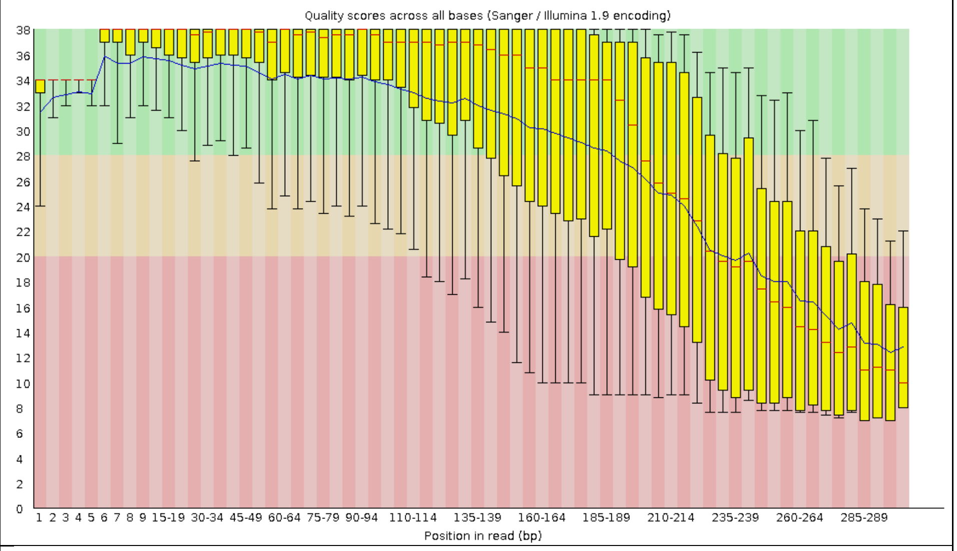
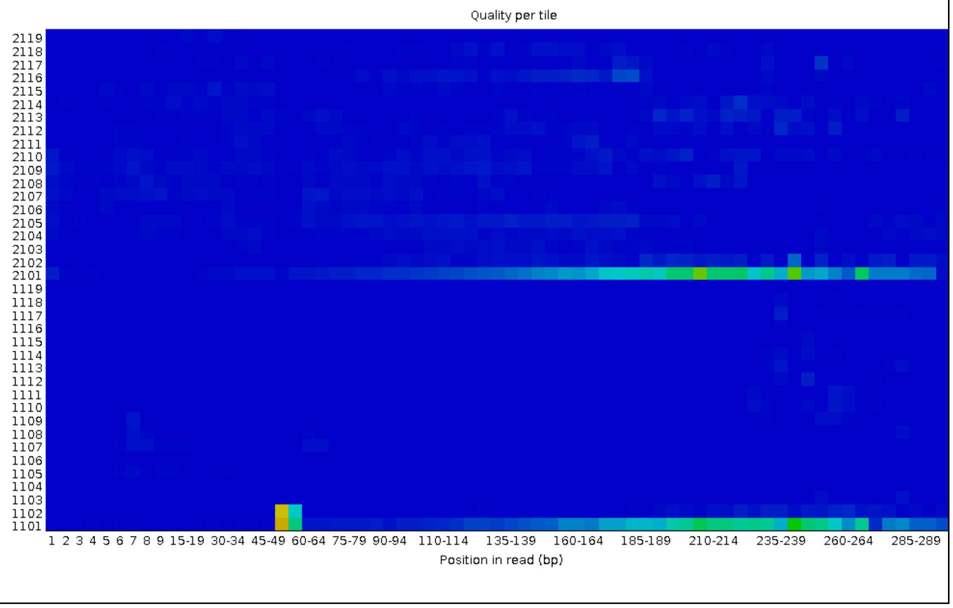
[FAIL]Adapter Content Trim Nextera adaptor	

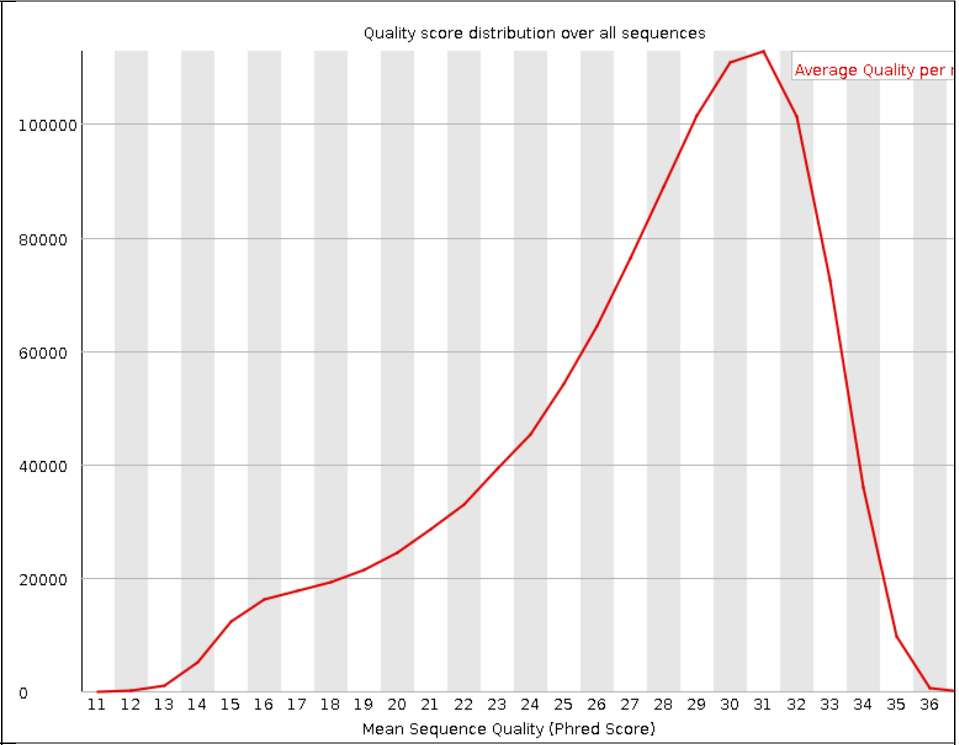
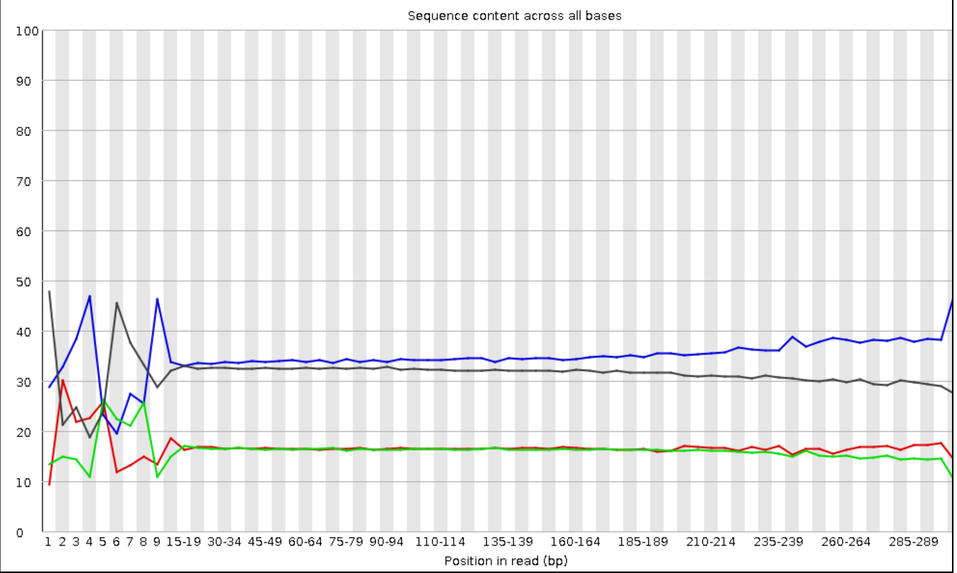
CF4.4: Reverse

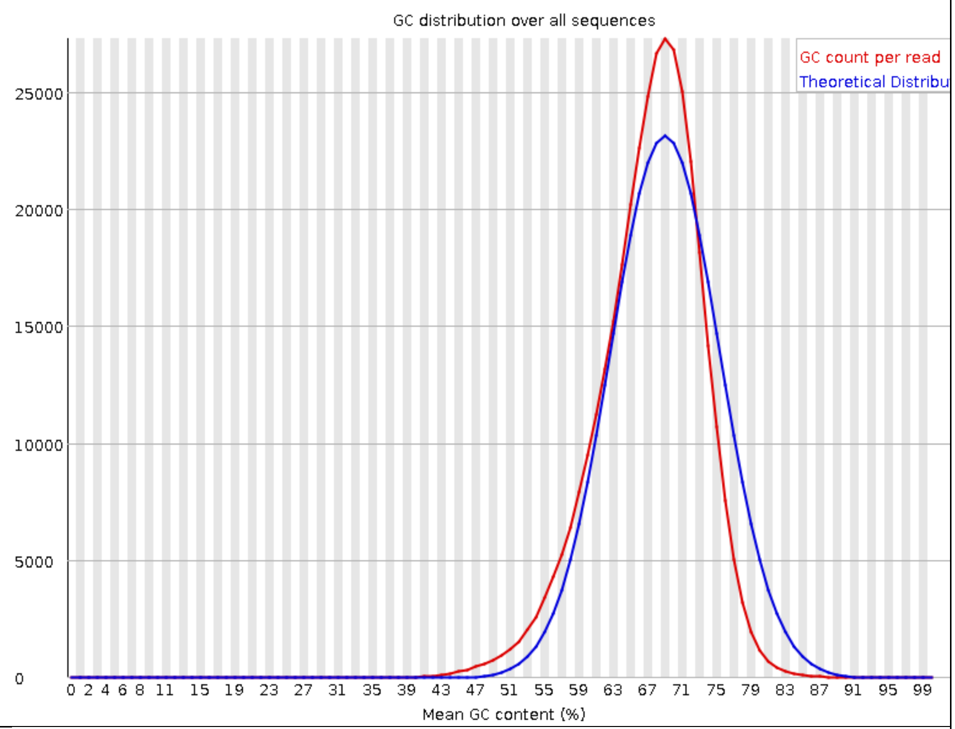
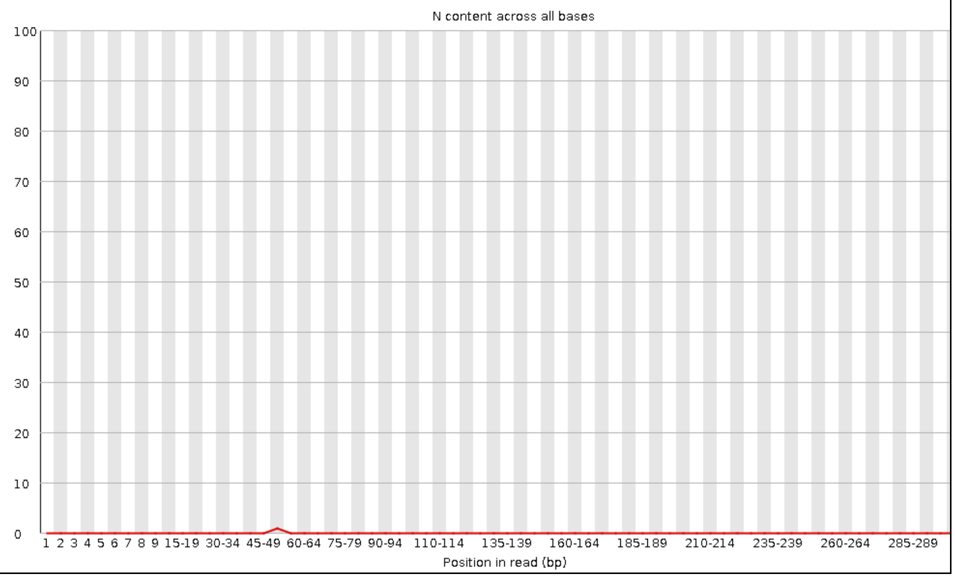
Basic Statistics

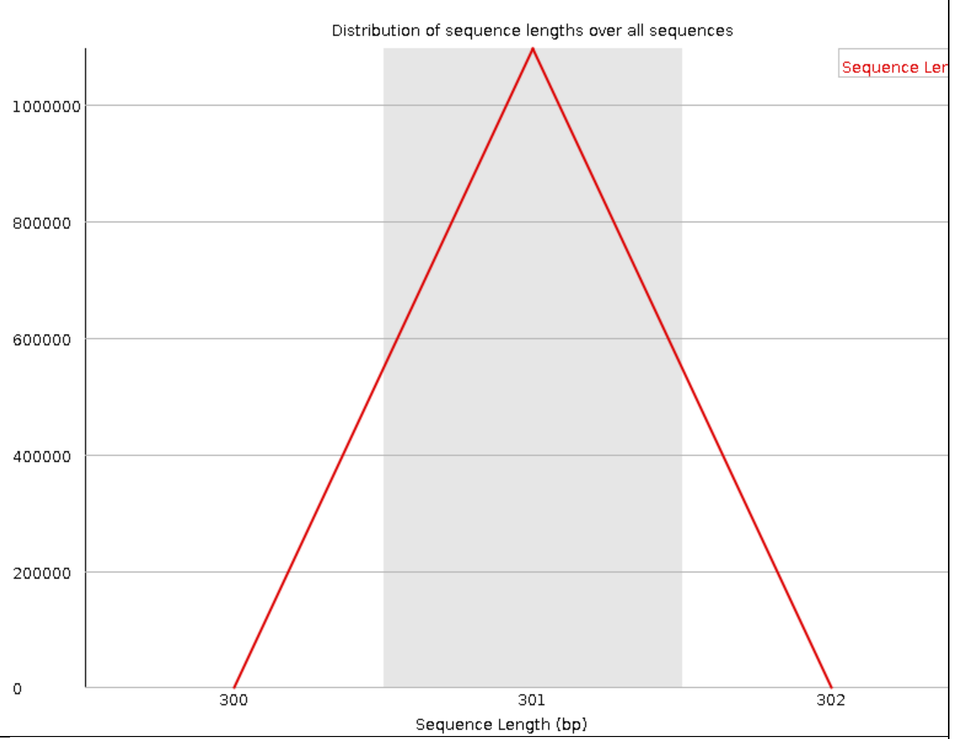
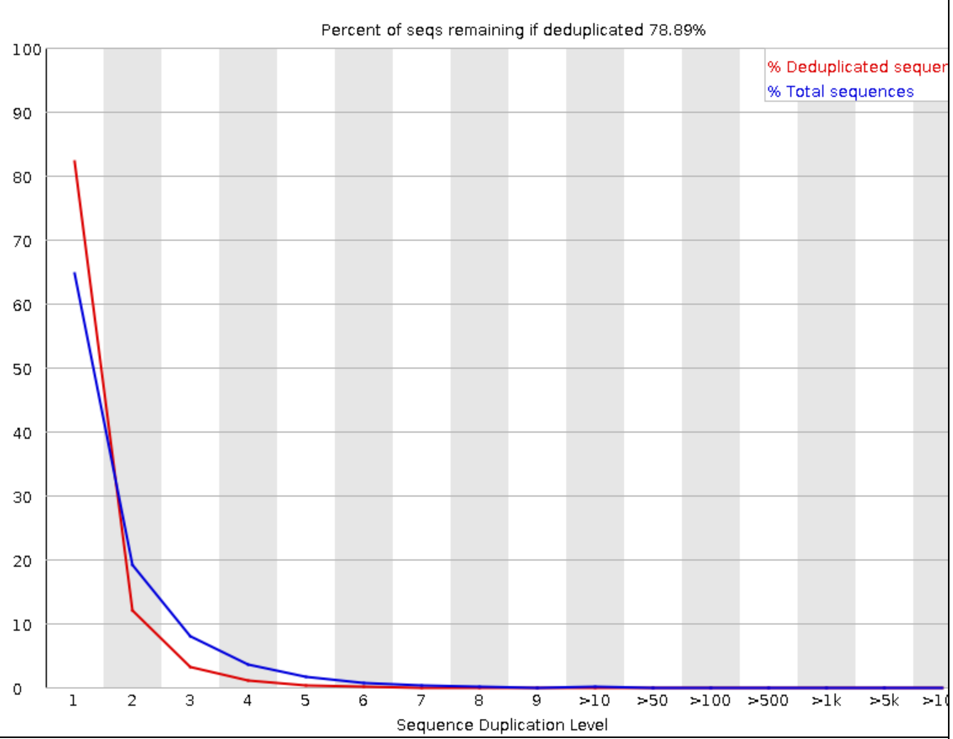
Measure	Value
Filename	Run20211209Order375Sample003_S188_L001_R2_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1095679
Sequences flagged as poor quality	0
Sequence length	301
%GC	67

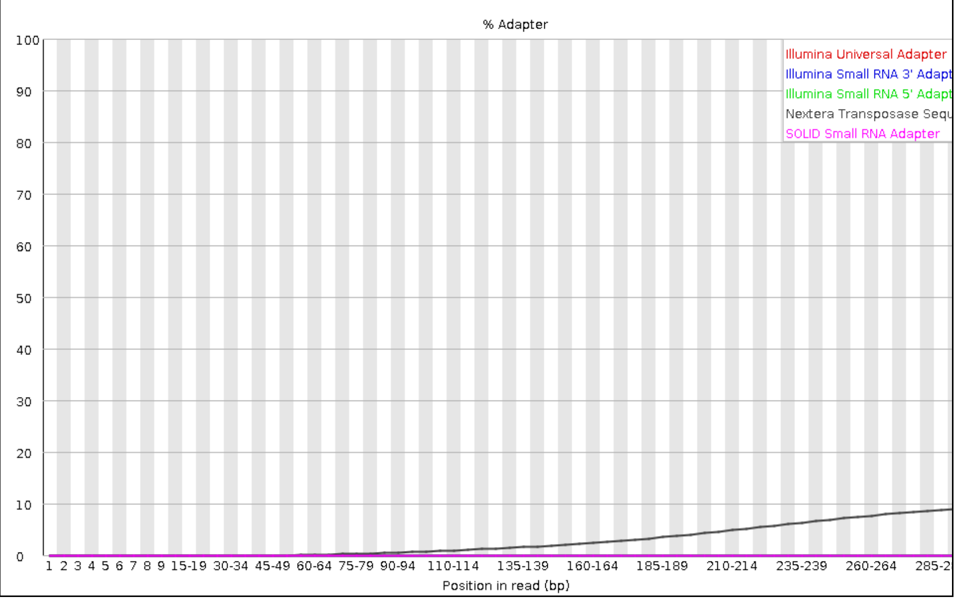
Quality Control Measure	Figure

[FAIL]Per base sequence quality Retain quality threshold: 28	
[WARNING]Per tile sequence quality Isolated warning issues may be ignored	

[PASS]Per sequence quality scores	
[FAIL]Per base sequence content Retain head crop: 15, Deviation at sequence end will be removed by SLIDINGWINDOW	

[WARNING]Per sequence GC content	
[PASS]Per base N content	

[PASS]Sequence Length Distribution	
[PASS]Sequence Duplication Levels	
[PASS]Overrepresented sequences	No overrepresented sequences

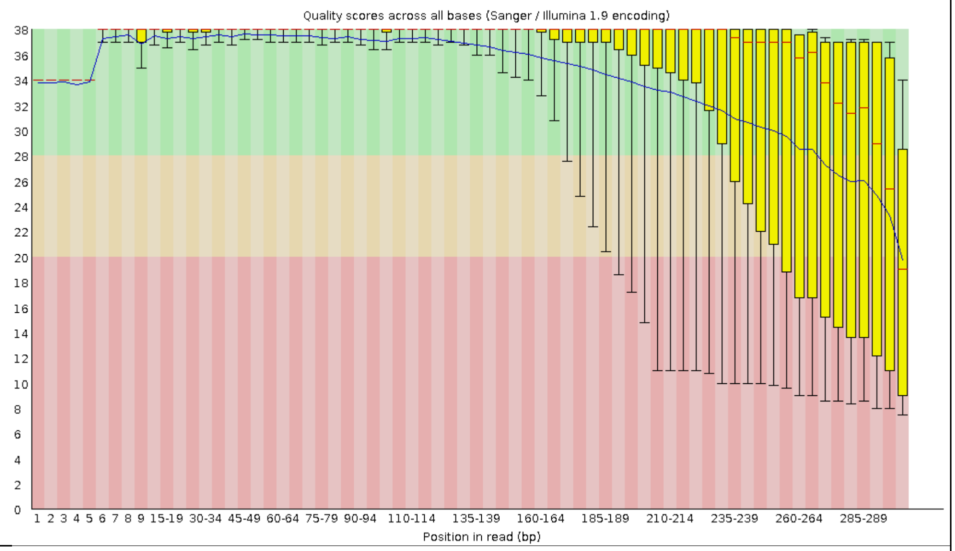
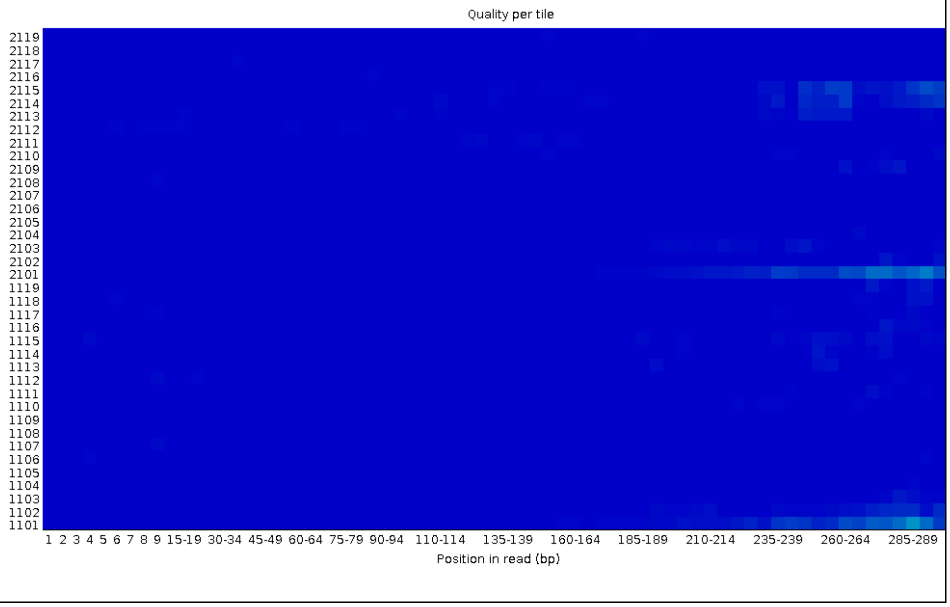
[WARNING]Adapter Content Trim Nextera adaptor	

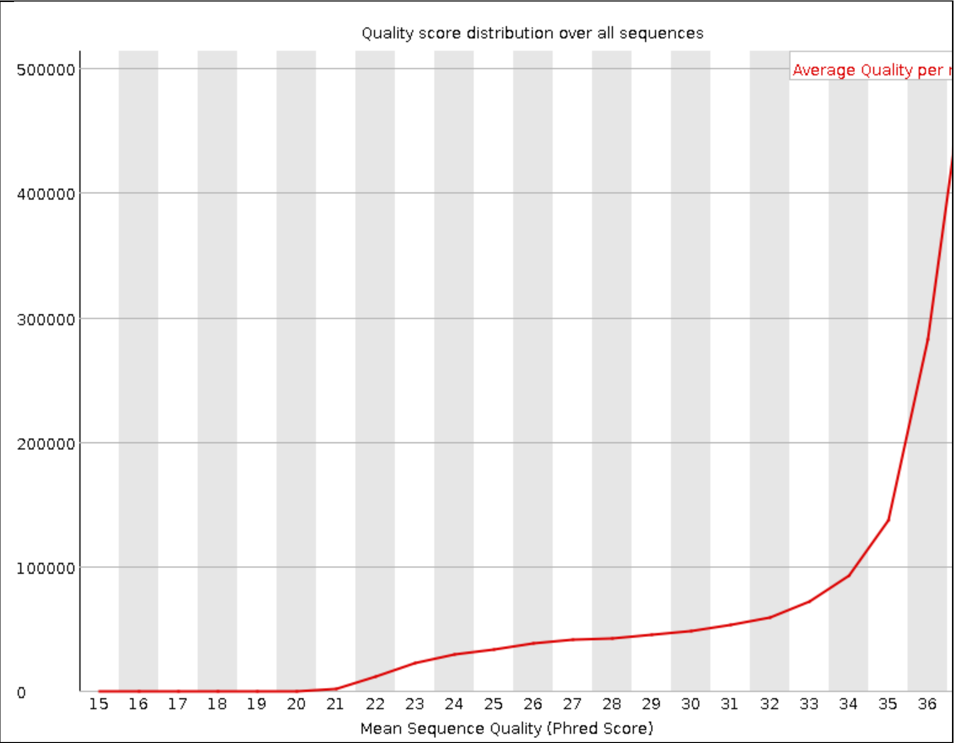
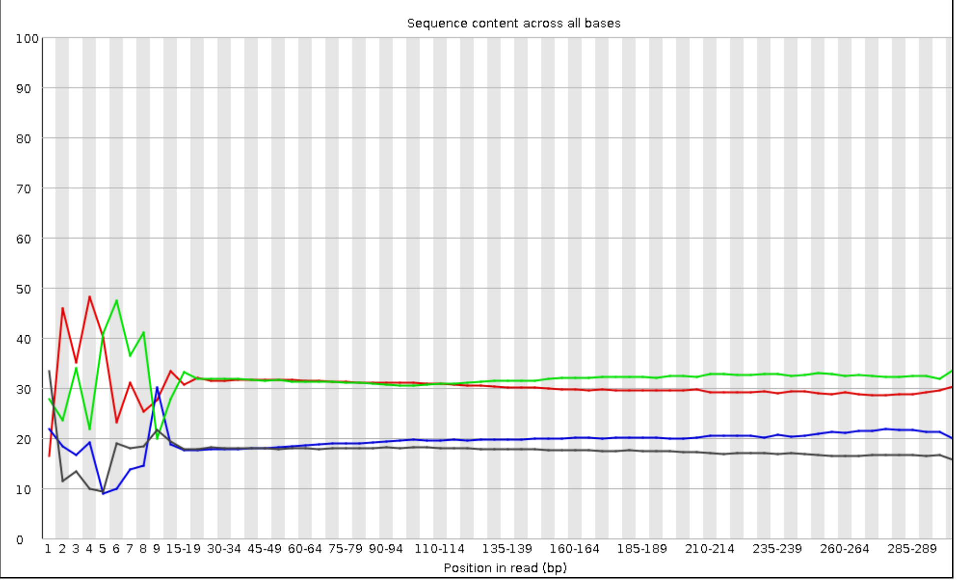
A.5: Forward

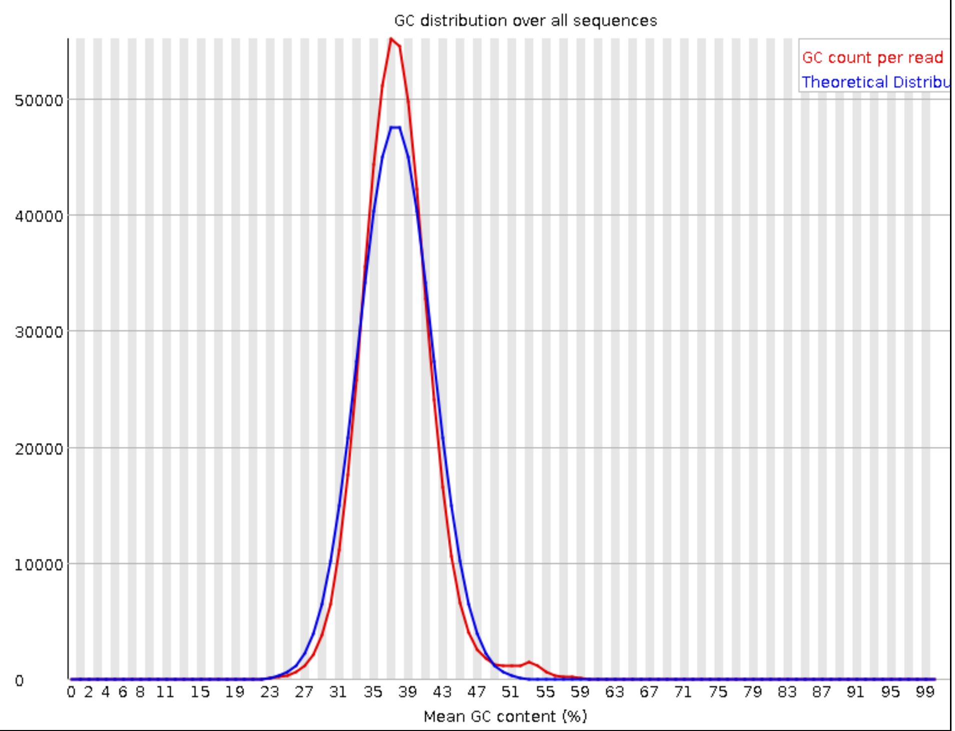
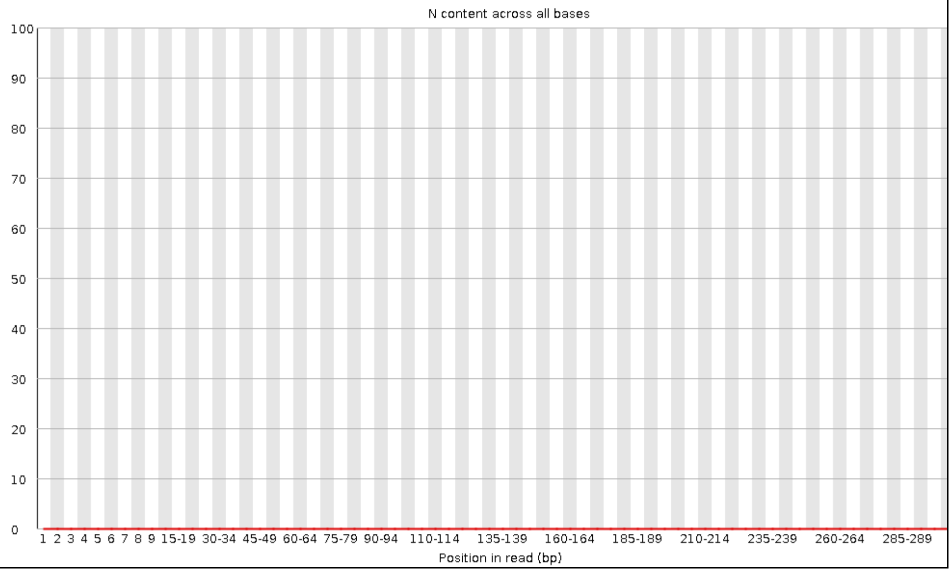
Basic Statistics

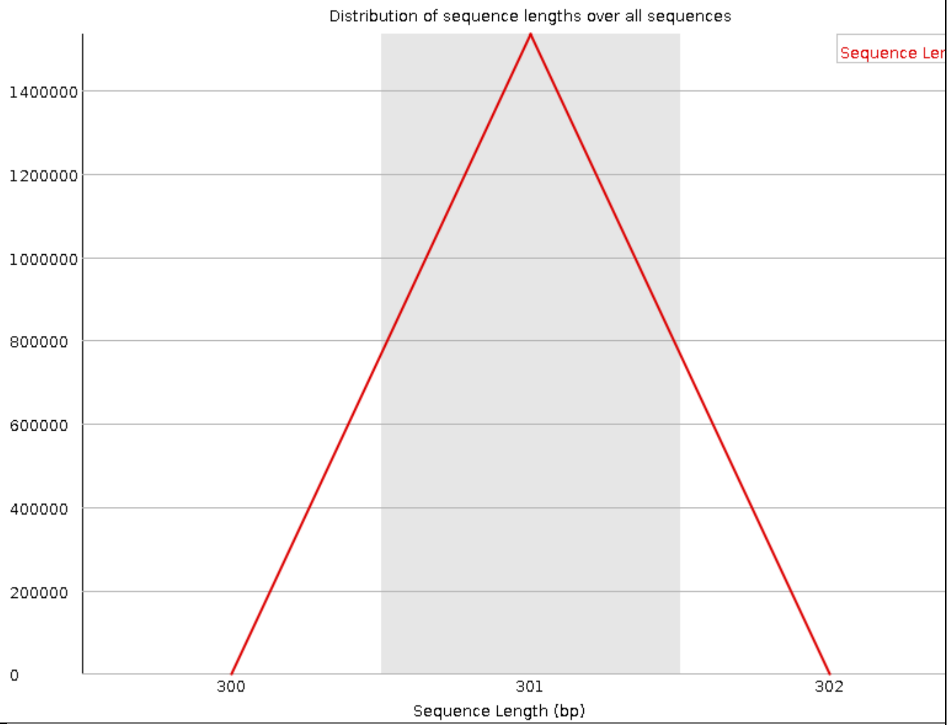
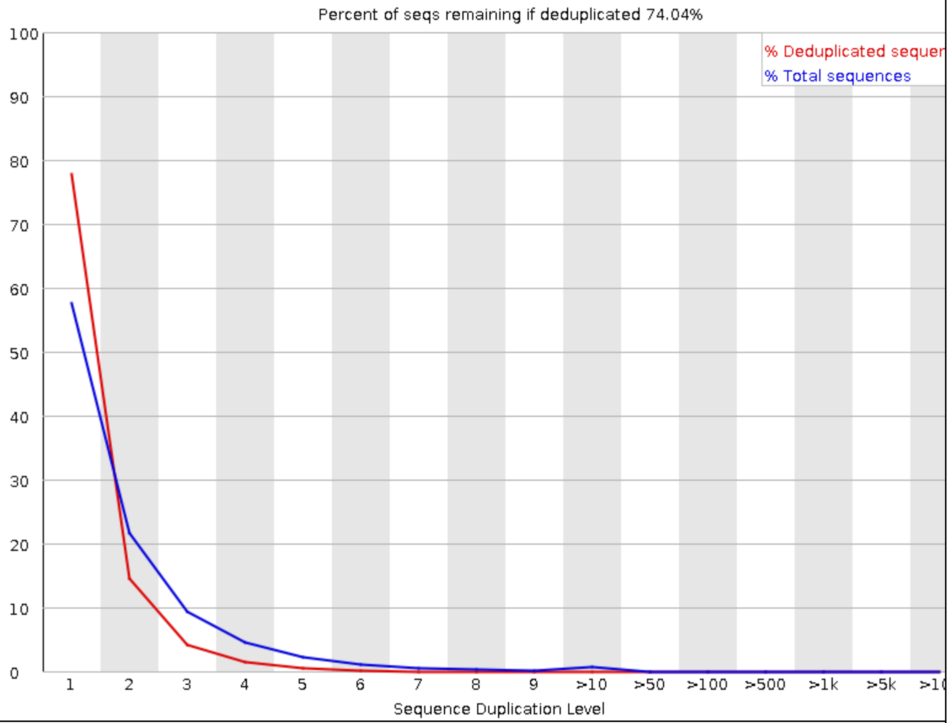
Measure	Value
Filename	Run20211209Order375Sample005_S191_L001_R1_001.fastq(1).gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1534464
Sequences flagged as poor quality	0
Sequence length	301
%GC	37

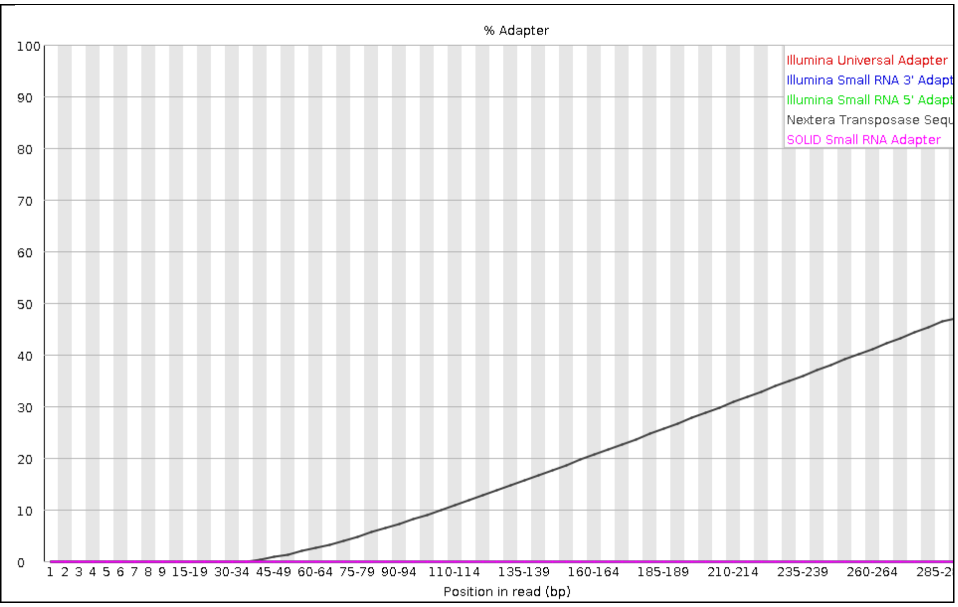
Quality Control Measure	Figure

[FAIL]Per base sequence quality Threshold quality 28	
[PASS]Per tile sequence quality	

[PASS]Per sequence quality scores	
[FAIL]Per base sequence content Head crop 15	

[WARNING]Per sequence GC content	
[PASS]Per base N content	

<table>
<tr>
<td>[PASS]Sequence Length Distribution</td>
<td>
</td>
</tr>
<tr>
<td>[PASS]Sequence Duplication Levels</td>
<td>
</td>
</tr>
<tr>
<td>[PASS]Overrepresented sequences</td>
<td>No overrepresented sequences</td>
</tr>
</table>

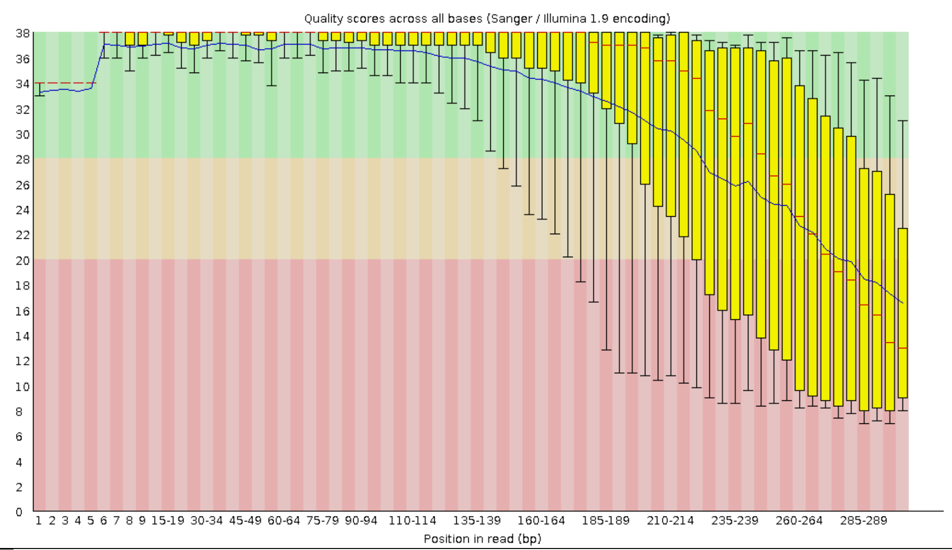
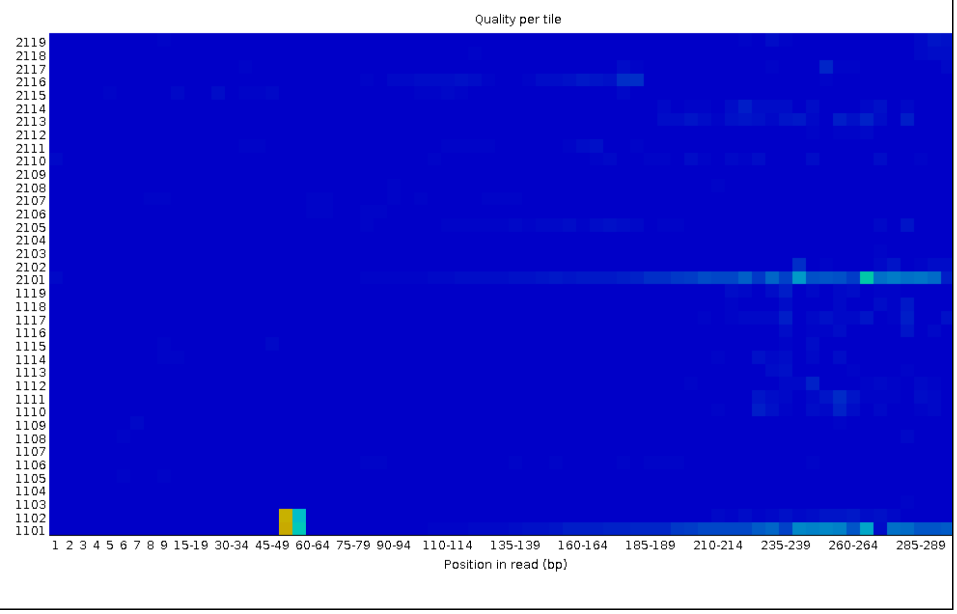
[FAIL]Adapter Content Trim Nextera adaptor	

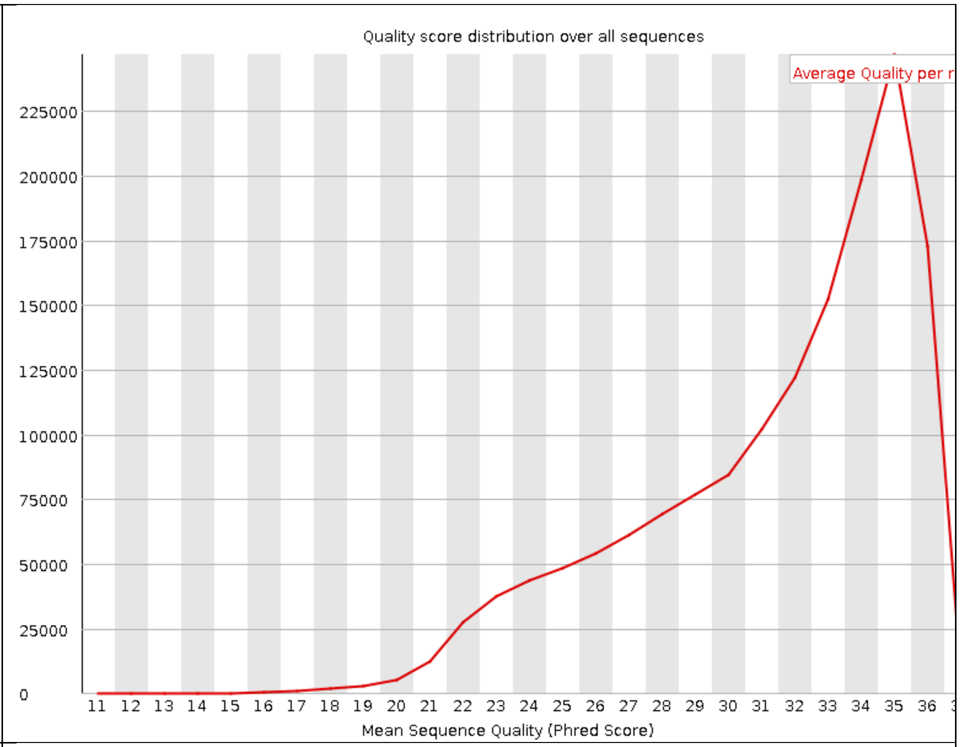
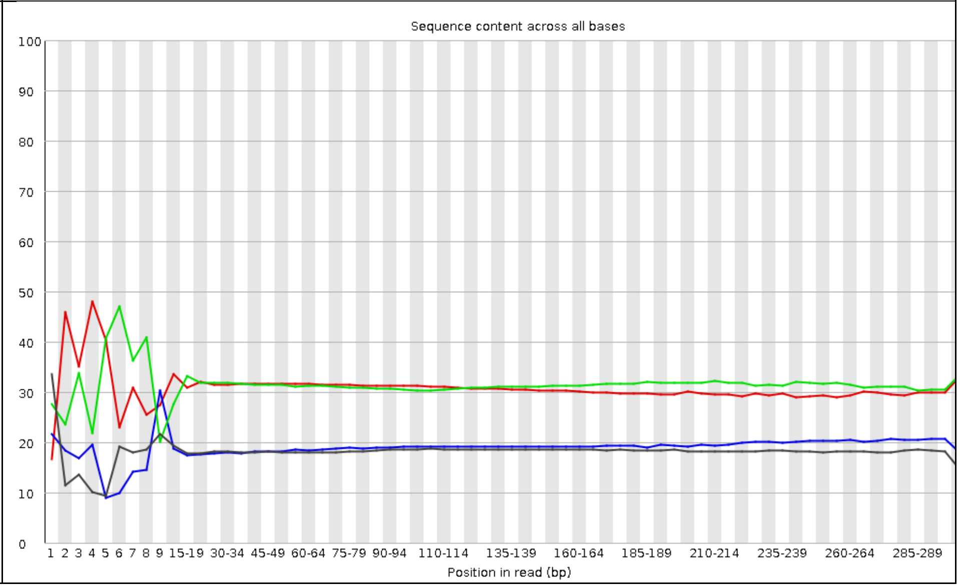
A.6: Reverse

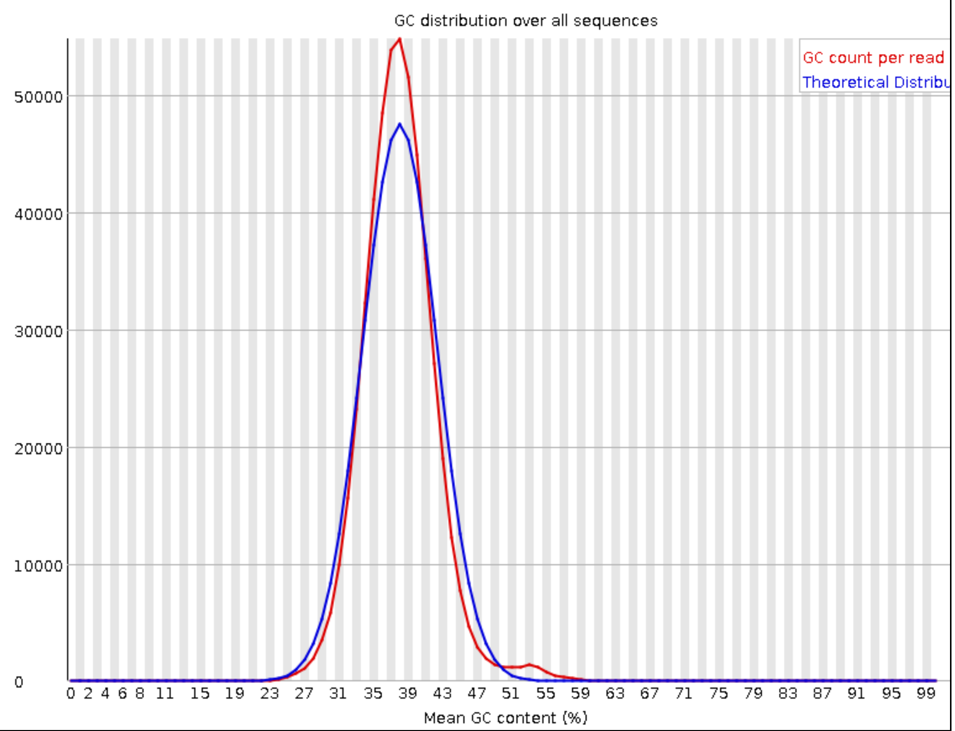
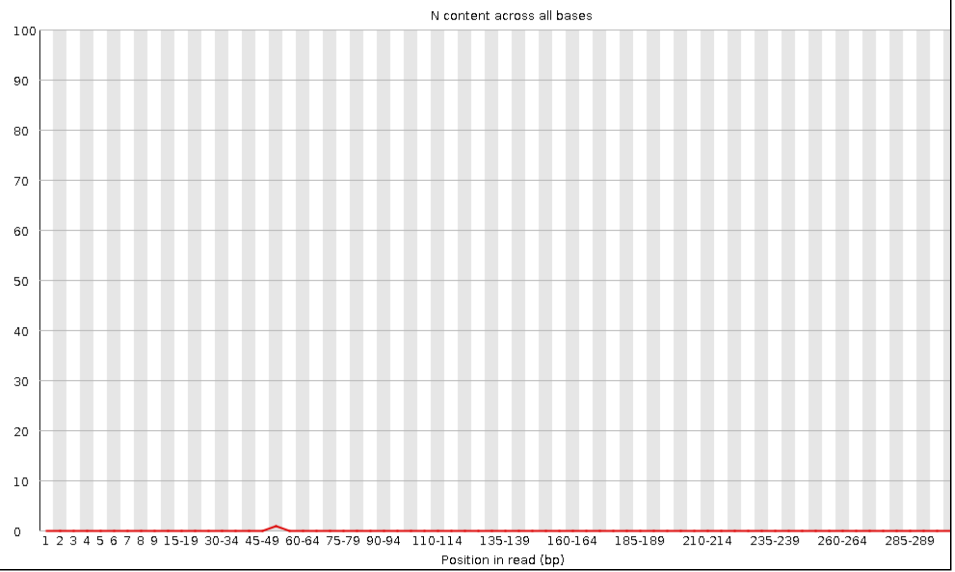
Basic Statistics

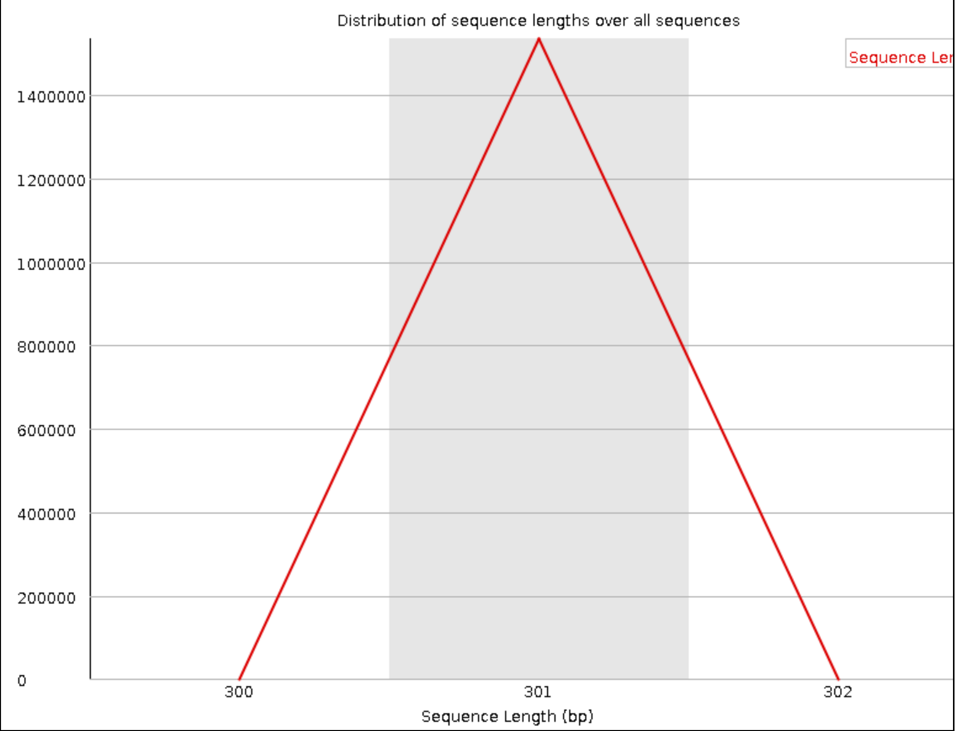
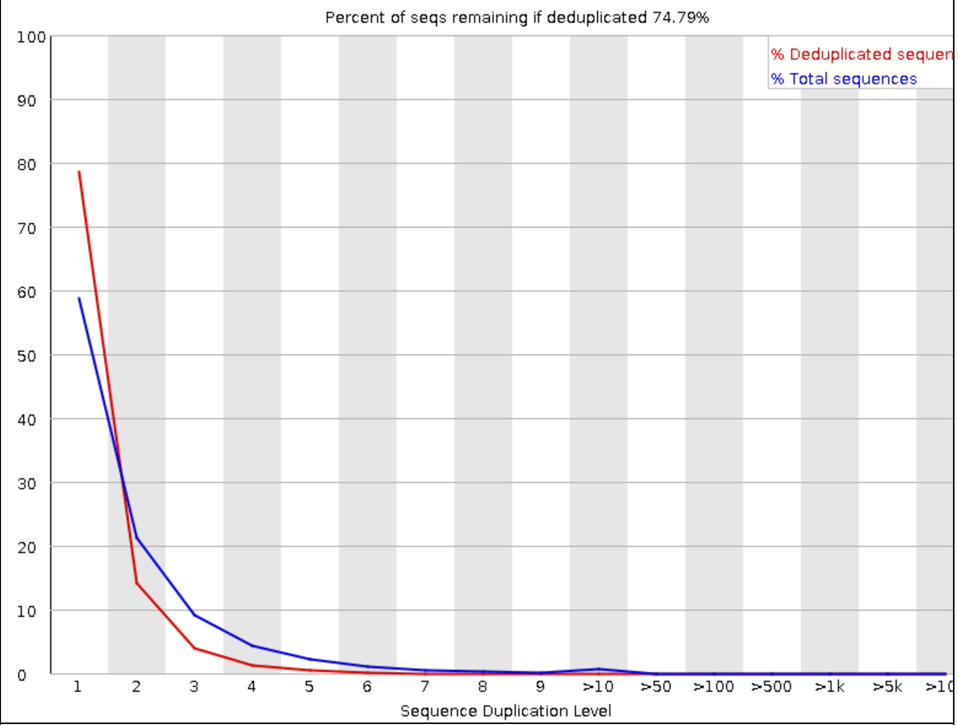
Measure	Value
Filename	Run20211209Order375Sample005_S191_L001_R2_001.fastq(2).gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1534464
Sequences flagged as poor quality	0
Sequence length	301
%GC	37

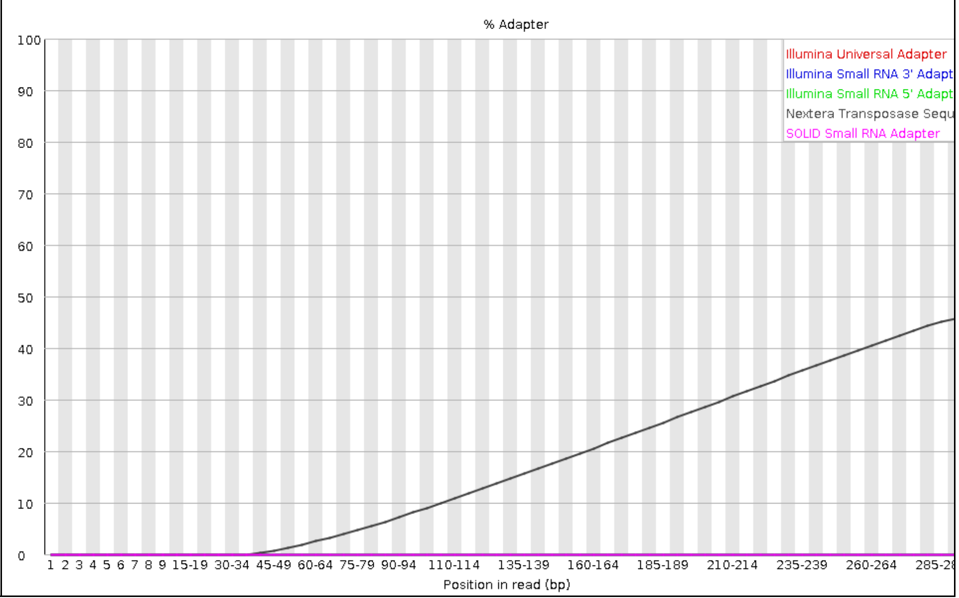
Quality Control Measure	Figure

[FAIL]Per base sequence quality Threshold quality 28	
[WARNING]Per tile sequence quality Isolated warning issues may be ignored	

[PASS]Per sequence quality scores	
[FAIL]Per base sequence content Head crop 15	

[WARNING]Per sequence GC content	
[PASS]Per base N content	

[PASS]Sequence Length Distribution	
[PASS]Sequence Duplication Levels	
[PASS]Overrepresented sequences	No overrepresented sequences

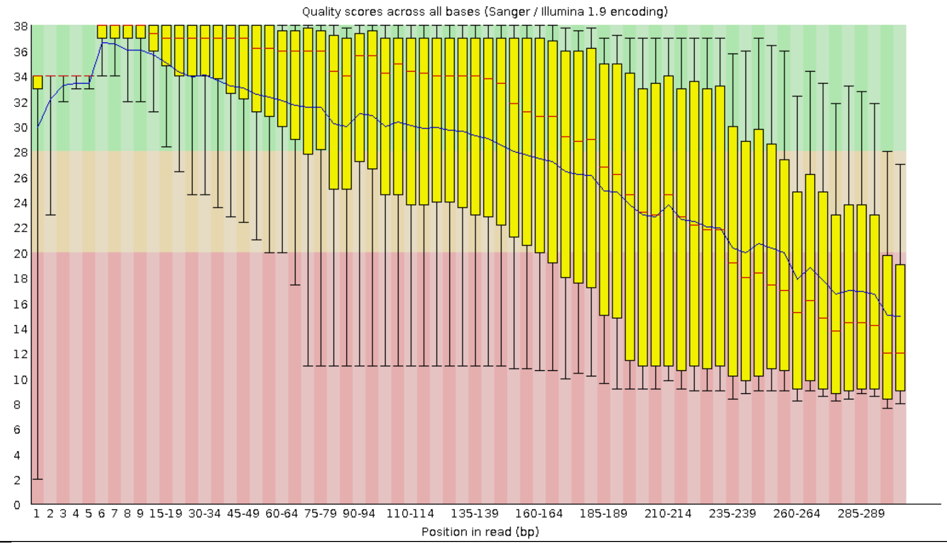
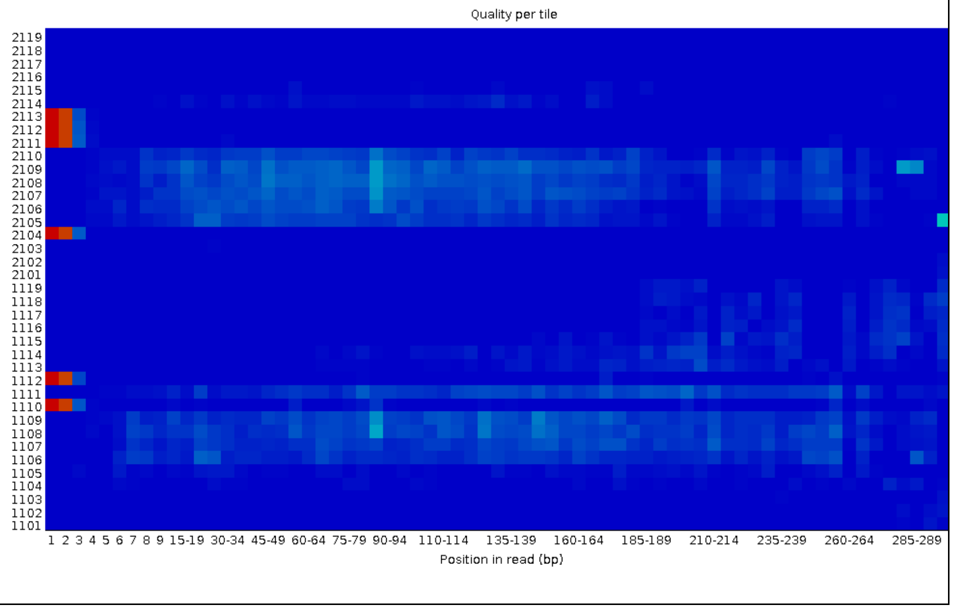
[FAIL]Adapter Content Trim Nextera adaptor	

A.10: Forward

Basic Statistics

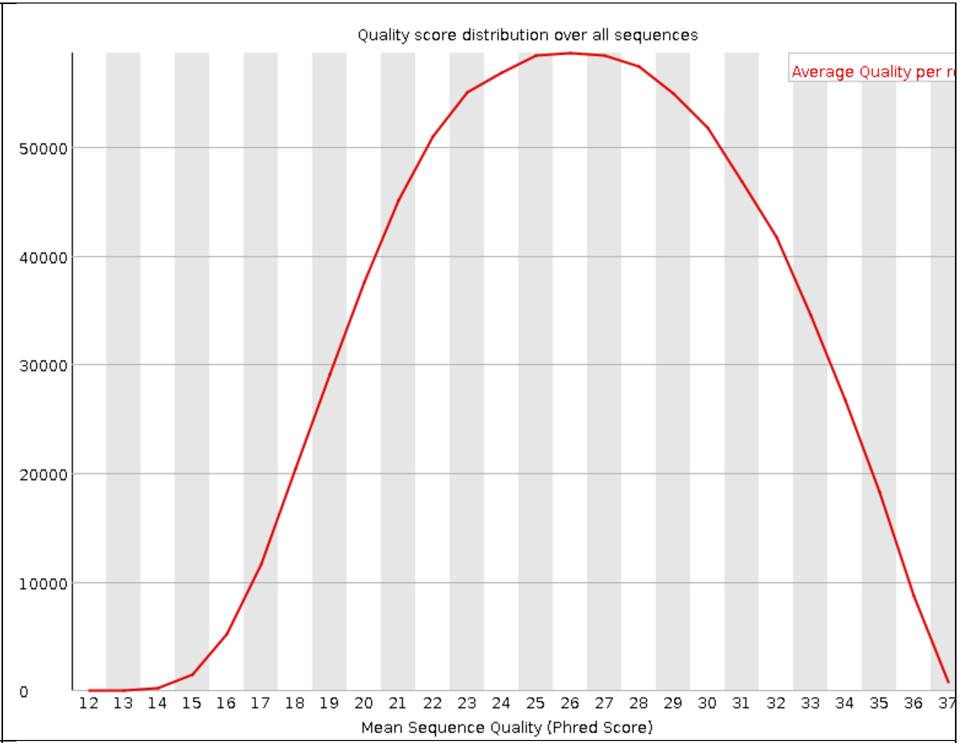
Measure	Value
Filename	Run20211215Order375Sample007_S193_L001_R1_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	831042
Sequences flagged as poor quality	0
Sequence length	301
%GC	44

Quality Control Measure	Figure

[FAIL]Per base sequence quality Conservative threshold quality of 20 to retain sequence length	
[FAIL]Per tile sequence quality Isolated failures will be removed by quality threshold	

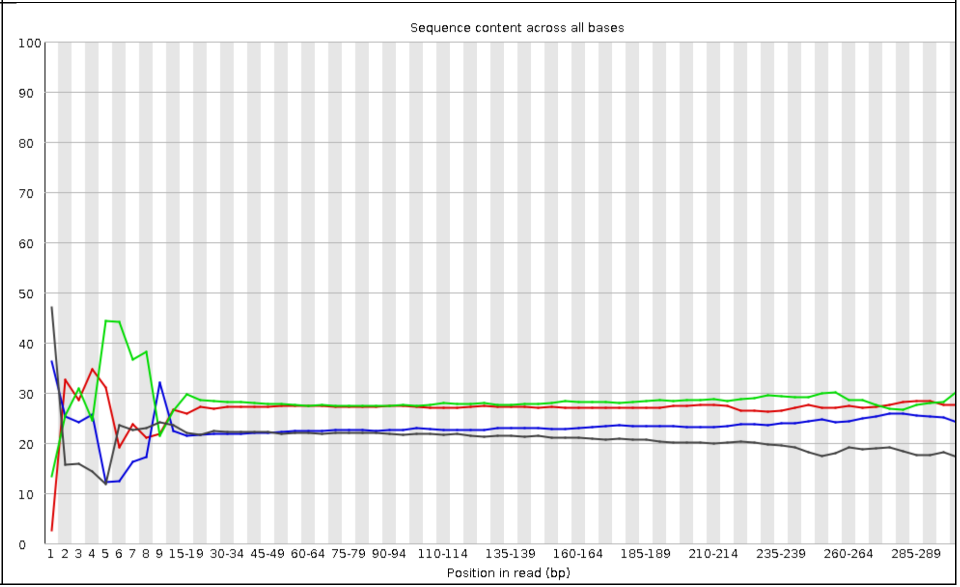
<table>
<tr>
<td>[WARNING]Per sequence quality scores

Unusual distribution</td>
<td>

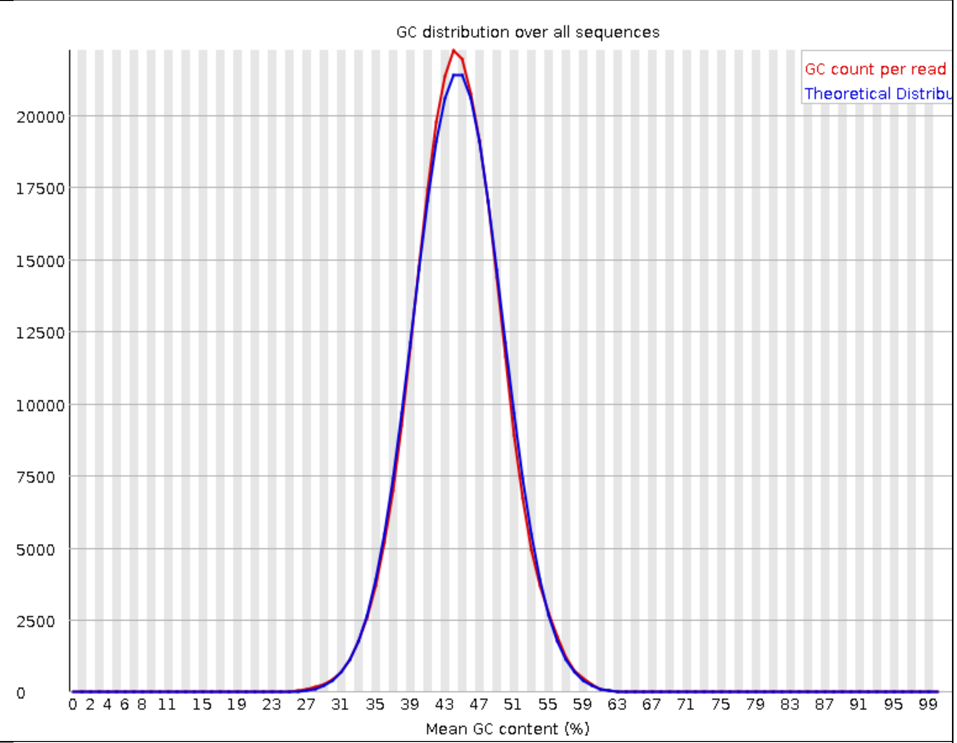
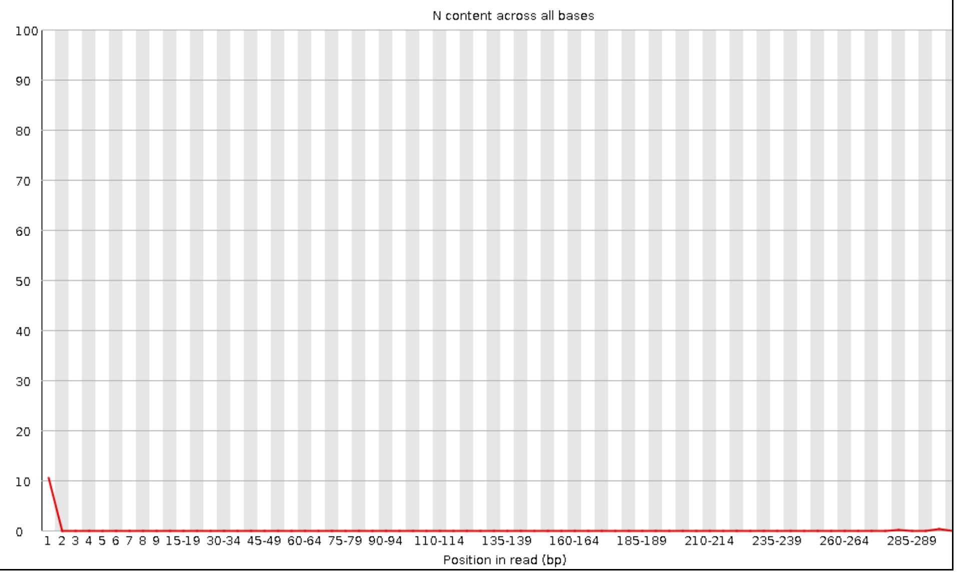


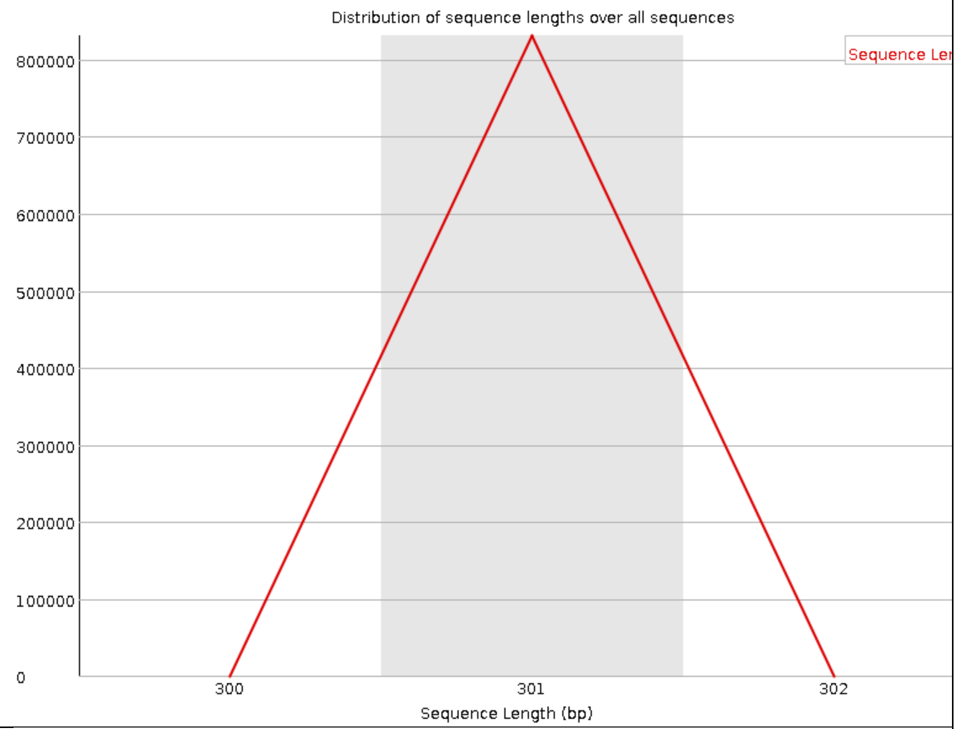
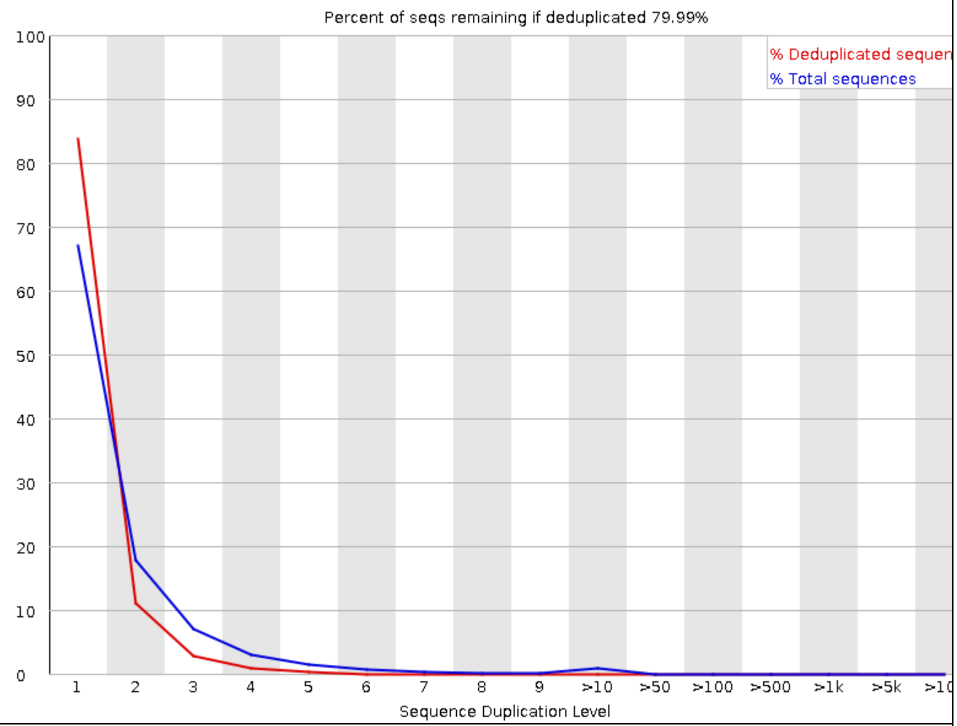
</td>
</tr>
<tr>
<td>[FAIL]Per base sequence content

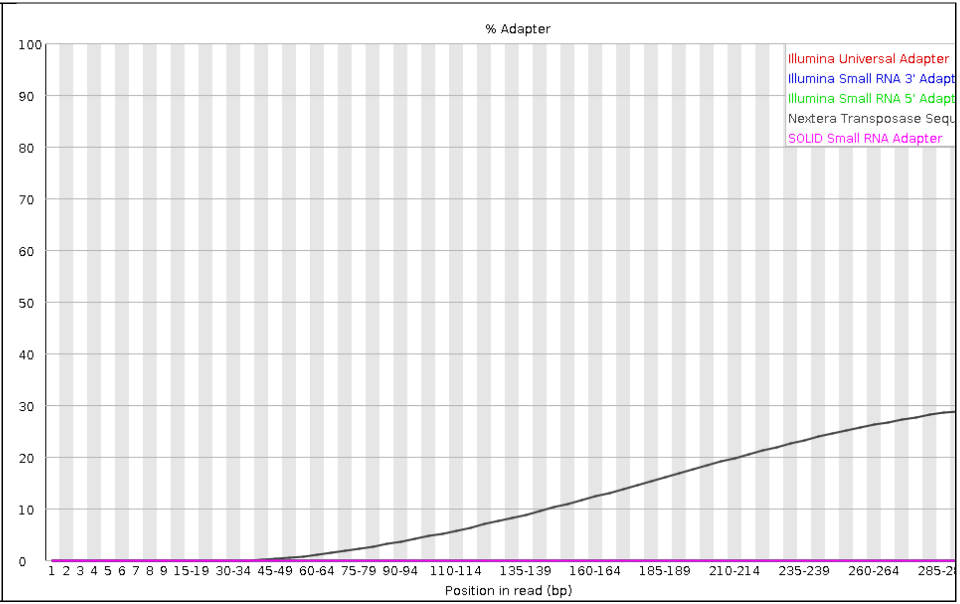
Head crop 15</td>
<td>



</td>
</tr>
</table>

[PASS]Per sequence GC content	
[WARNING]Per base N content Mirrors per tile sequence quality	

[PASS]Sequence Length Distribution	
[PASS]Sequence Duplication Levels	
[PASS]Overrepresented sequences	No overrepresented sequences

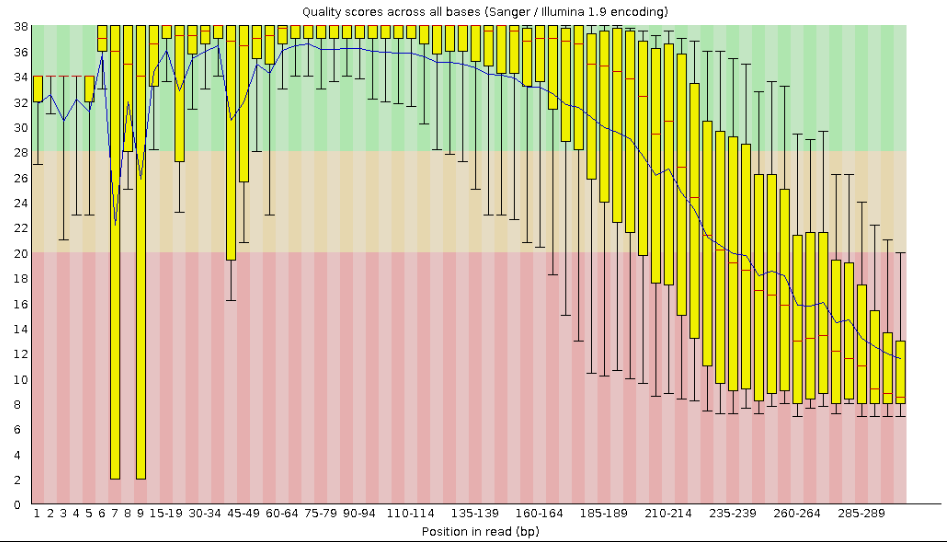
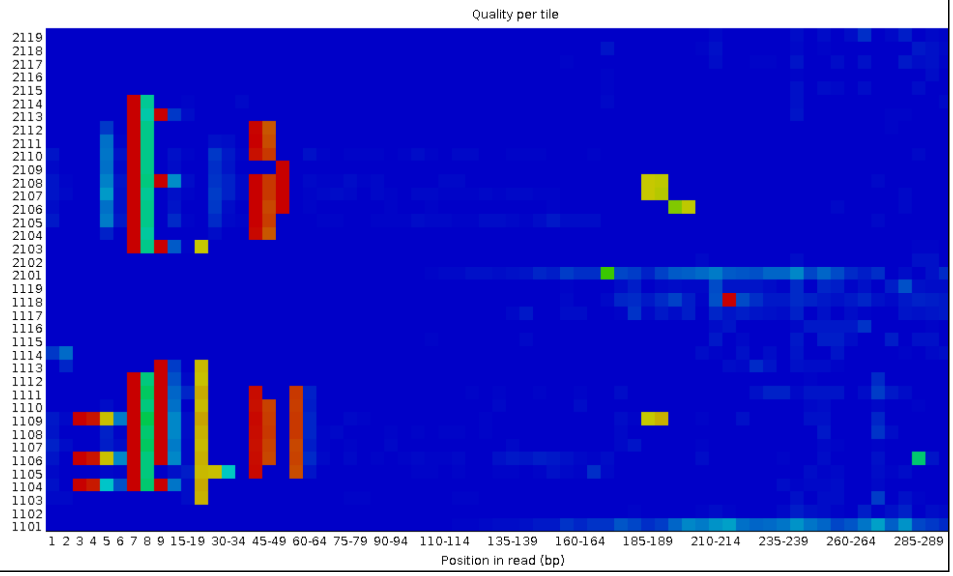
[FAIL]Adapter Content Trim Nextera adaptor	

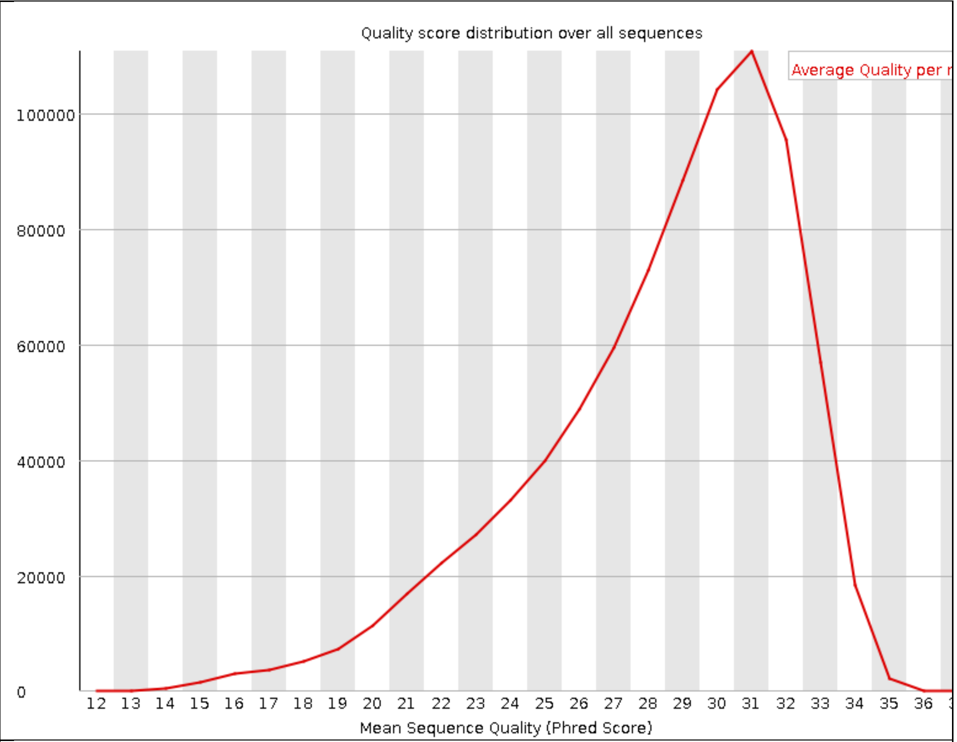
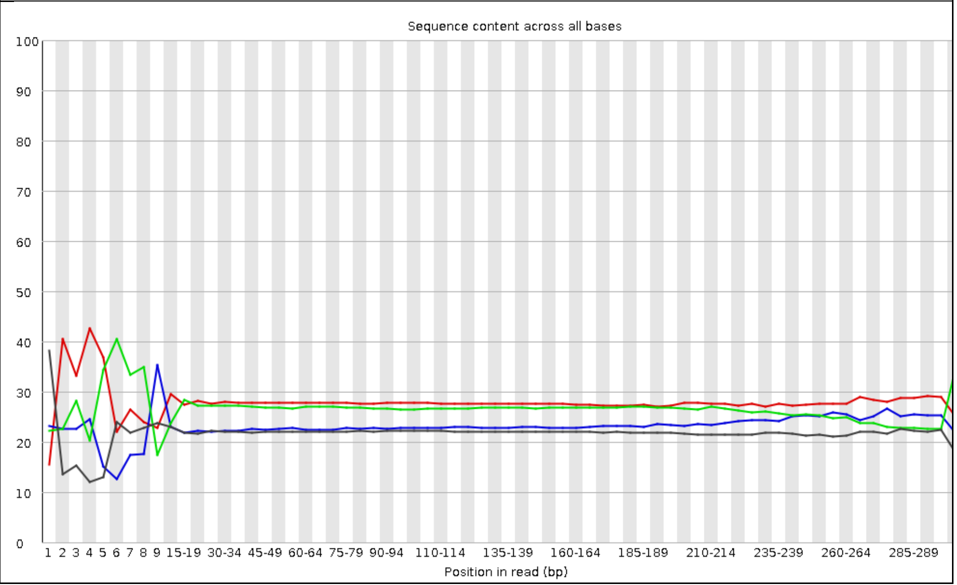
A.10: Reverse

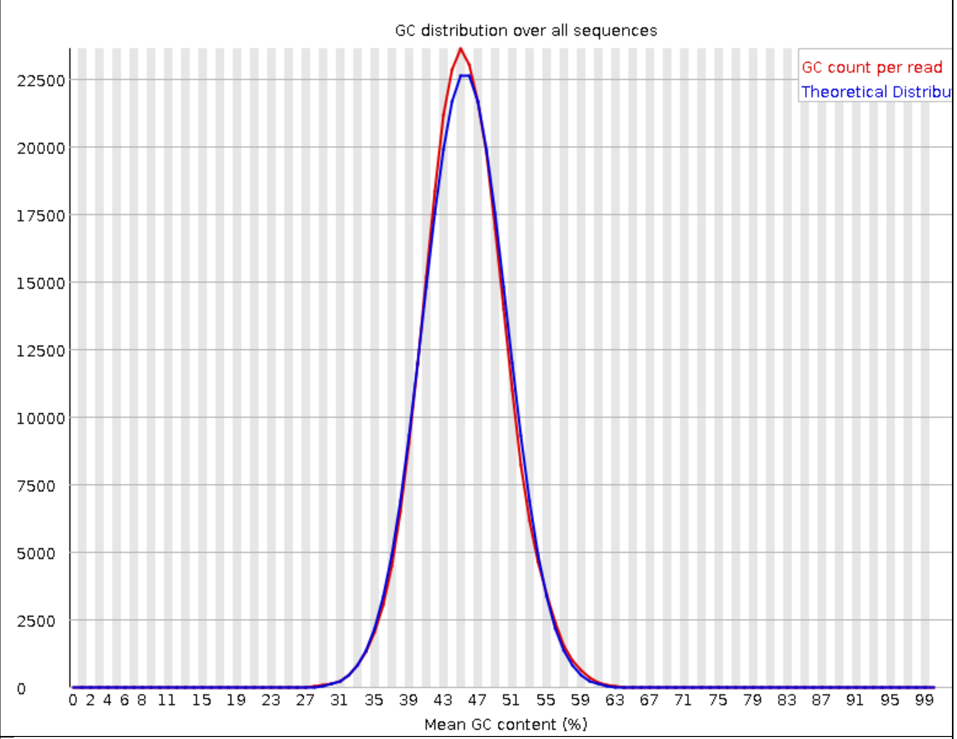
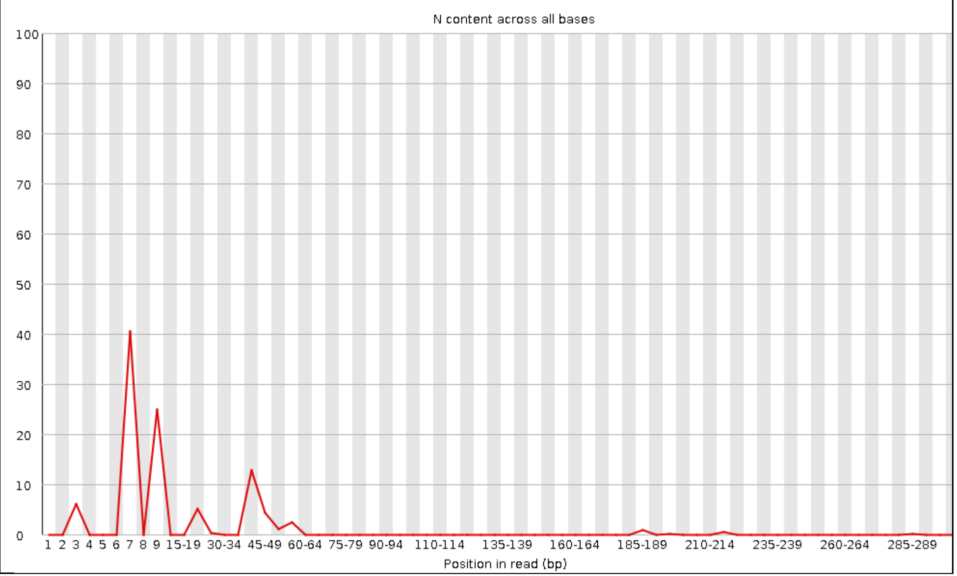
Basic Statistics

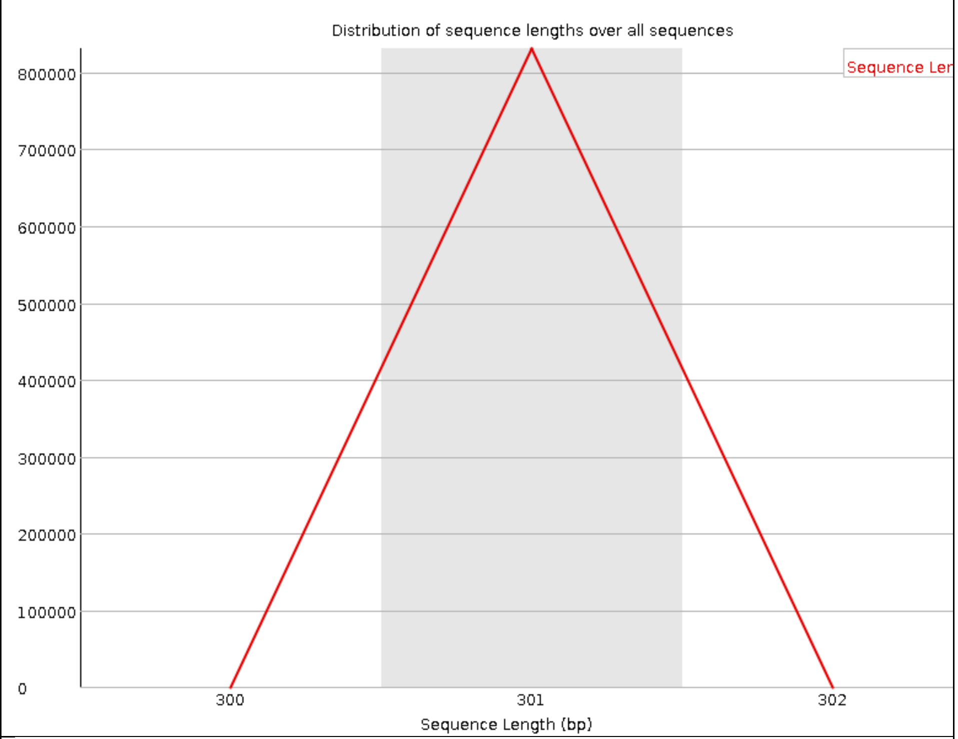
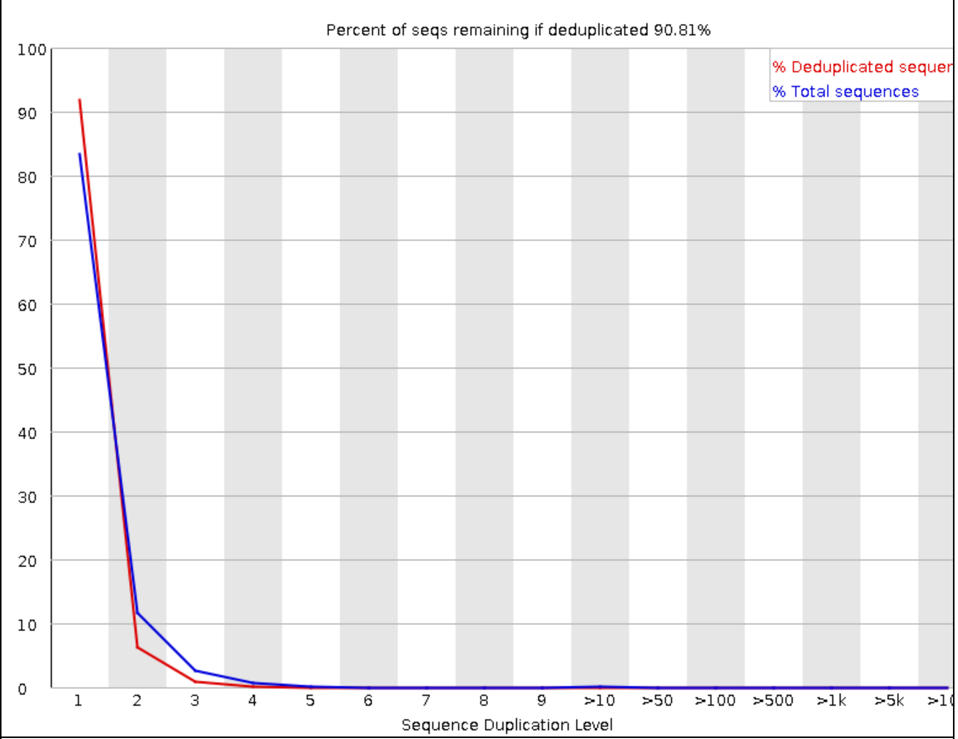
Measure	Value
Filename	Run20211215Order375Sample007_S193_L001_R2_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	831042
Sequences flagged as poor quality	0
Sequence length	301
%GC	45

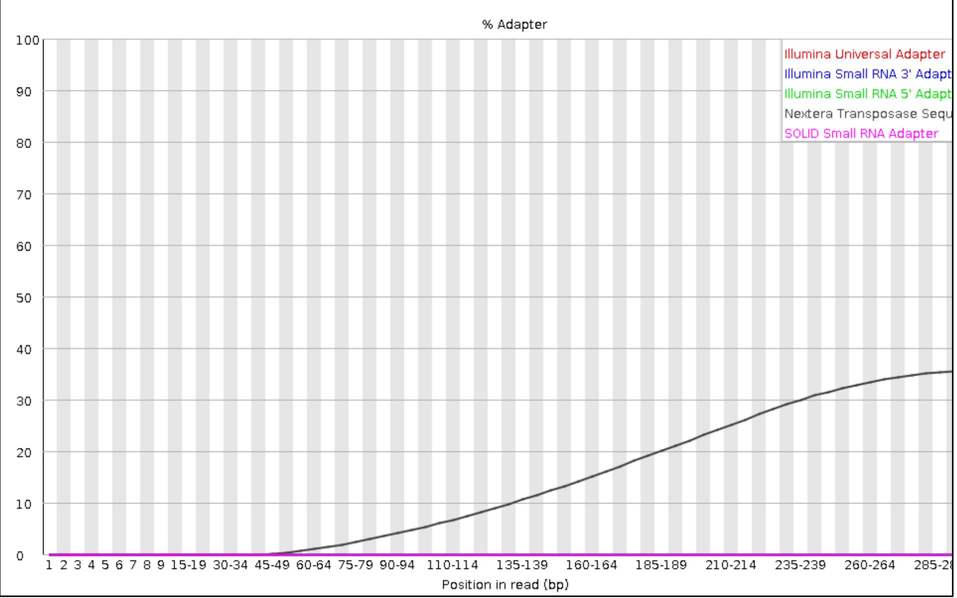
Quality Control Measure	Figure

[FAIL]Per base sequence quality Conservative quality threshold of 20 to preserve the length of sequences	
[FAIL]Per tile sequence quality Possible blotches may be fixed by quality threshold	

[PASS]Per sequence quality scores	
[FAIL]Per base sequence content Head crop 15	

[PASS]Per sequence GC content	
[FAIL]Per base N content Mirrors blotches from per tile sequence ccontent	

[PASS]Sequence Length Distribution	
[PASS]Sequence Duplication Levels	
[PASS]Overrepresented sequences	No overrepresented sequences

[FAIL]Adapter Content Trim Nextera adaptor	

Sequenced Samples

the first 16s sequencing cohort indicated in **bold.** Links to the BLAST alignment embedded in match name

Read Sample ID	Forward match	Forward Statistics: E-value/ Percent identity/ Accession length (bp)			Reverse Match	Reverse Statistics: E-value/ Percent identity/ Accession length		
A5*	Bacillus licheniformis strain LB 102-1	0.0	96.66	1319	Bacillus licheniformis strain QT338	0.0	97.74	1453
A.6*	Bacillus cereus strain D85	0.0	96.16	1363	Bacillus cereus strain MSM	0.0	07.18	1499
A.9*	Pseudomonas stutzeri strain DBNSCF2	0.0	97.65	1429	Pseudomonas stutzeri strain B13	0.0	97.40	1445
A.10*	Bacillus velezensis strain UOH-45	0.0	95.02	1483	Bacillus mojavensis strain UCMB5075	0.0	97.47	4031121
A.11*	Bacillus aquimaris strain DL36	0.0	93.24	1425	Bacillus sp. HMD3161	0.0	96.01	1430
CB4.7B	Streptomyces sp. strain EIIIA	0.0	99.69	1362	Streptomyces pratensis	0.0	99.90	1427
CF1.1	Jeotgalibacillus marinus strain 1019_C3F	0.0	99.26	871	Jeotgalibacillus campisalis strain CW126-A17	4E-175	97.30	1514
CF3.3	Nesterenkonia halotolerans strain YIM70084	0.0	99.55	1483	Nesterenkonia halotolerans strain YIM70084	0.0	99.55	1483

CF4.1	Bacillus neizhouensis	0.0	99.74	1520	Bacillus neizhouensis	0.0	99.56	1520
CF4.2	Oceanobacillus massiliensis	0.0	98.29	1506	Ornithinibacillus sp. SCULCB	0.0	99.04	1584
CF4.4	Nesterenkonia sandarakina strain SCA-110	0.0	99.30	1421	Nesterenkonia sandarakina strain SCA-110	0.0	00.63	1421
CF4.5	Oceanobacillus caeni strain M111p1-10	0.0	98.56	940	Oceanobacillus sp. strain APA_H-1(4)	0.0	98.73	1591
CF4.6	Oceanobacillus longus strain T9B	0.0	97.86	1484	Oceanobacillus sp. strain APA_H-1(4)	0.0	98.91	1591
CF4.7	Streptomyces griseus strain S10-TSA-15	0.0	99.82	1458	Streptomyces microflavus strain NA06532	0.0	99.90	1435
CF4.8	Streptomyces finlayi strain IHBA	0.0	100.00	1470	Streptomyces sp. QLS30	0.0	99.81	1459
CF4.9	Pseudarthrobacter siccitolerans strain 192-LR4	0.0	99.22	1428	Pseudarthrobacter sp. strain D18-7	0.0	99.26	1501
CMS3.1	Uncultured bacterium clone XZ7_G7	0.0	99.30	1496	No sequence data	-	-	-
KK2.1	Uncultured Alcanivorax sp. clone C114Chl091	0.0	7E-22	1444	Arthrobacter sp. strain 810	0.0	99.30	1419
KK3.1	Micrococcus yunnanensis strain G1-7-20	0.0	99.28	1485	Micrococcus luteus strain HKG359	0.0	99.73	1315
KK4.1	Arthrobacter sulfonivorans	0.0	99.81	1515	Arthrobacter sp.MT-A-S7	0.0	99.45	1414
KK4.2	Arthrobacter sp. strain 20TX0003	0.0	98.65	1443	Arthrobacter sp. R33S	0.0	99.28	1520
KK4.3	Arthrobacter alpinus strain S6-3	0.0	99.45	1530	Arthrobacter sp. Ia1	0.0	99.43	1445
KK5.1	Arthrobacter sulfonivorans	0.0	99.82	1515	Arthrobacter sp. UYEF18	0.0	99.72	1094
KK5.2	Arthrobacter sulfonivorans	0.0	99.28	1515	Arthrobacter sp.MT-A-S7	0.0	99.64	1414
KK6.1	Pseudomonas lini strain BS3782	0.0	99.25	1517	Pseudomonas fluorescens strain hp13	0.0	99.56	1428
KK7.1	Arthrobacter siccitolerans strain 24	0.0	99.91	1490	Arthrobacter sp. strain Ni723	0.0	100.00	1389
KK9.2	Pseudomonas sp. strain PAMC 27329	0.0	99.32	1468	Pseudomonas sp. strain PAMC 27304	0.0	99.81	1467

*Sequenced by Poul K. Madsen as part of master’s thesis. (Madsen, 2020)

Cultured Samples

Label	**Origin**	**Taxonomic classification**	**Medium**	**Temp**	**Comments**
CF1.1	Citronen Fjord	Jeotgalibacillus marinus	HM + 10%, Marine agar	4C	Irrelevant growth at 5% salt, but grows on Marine agar.
CF1.4	Citronen Fjord		LB	10C, 25C	Filamentous bacterium
CF1.4	Citronen Fjord		TSA	10C	Strep. Taints TSA with a dark orange/brown compounds
CF3.1	Citronen Fjord		TSA	10C	
CF3.2	Citronen Fjord		Marine Agar	4C	Cocci or bacilli
CF3.3	Citronen Fjord	Nesterenkonia halotolerans	HM + 10%, R2 + 10%, Marine Agar	4C, 10C, 25C	Pale red. Tolerates 15% salt much better than CF4.4
CF4.1	Citronen Fjord	Salipaludibacillus neizhouensis	HM + 5%. LB + 5%	5C, 25C	Up to 10% NaCl
CF4.11	Citronen Fjord		R2A + 10%, HM +5%, MarA	10C, 25C	Probably Nesterenkonia
CF4.12	Citronen Fjord		R2A + 10%, HM +5%, MarA	10C, 25C	Probably Nesterenkonia
CF4.12	Citronen Fjord		R2A + 10%, HM +5%, MarA	10C, 25C	Probably Nesterenkonia
CF4.2	Citronen Fjord	Oceanobacillus massiliensis	Marine Broth/Agar	5C, 25C	Up to 10% NaCl
CF4.3	Citronen Fjord		HM + 5%. LB + 5%	25C	Fungus
CF4.4	Citronen Fjord	Nesterenkonia sandarakina	Marine Broth/Agar	5C, 25C	Red. Very slow growth at 15% (weeks or months). Has a lots of genes to make proline, which can be used as a compatible solute. After antiSMASH analysis: Has clusters for carotenoid (hence the color) and ectoine (compatible solute). Also has a NAPAA cluster that seems related to either a blue, antioxidant pigment, a compound used in sunscreen and an aspergillus antibiotic (which also promotes nerve growth). Lastly, it has a polyketide cluster apparently related to hierridin B and C... I found some papers testing this compound against cancer cells.
CF4.5	Citronen Fjord	Oceanobacillus caeni / sp.	HM + 10%	5C, 10C	
CF4.7	Citronen Fjord	Streptomyces sp.	TSA	10C	Strep. Taints TSA with a dark orange/brown compounds
CF4.6	Citronen Fjord	Oceanobacillus longus / sp.	HM + 10%	5C, 10C	
CF4.7 B	Citronen Fjord	Streptomyces sp.	TSA	10C	Supposedly CF4.7, but did not stain TSA. I don't know if it is a different microbe or a different phenotype.
CF4.8	Citronen Fjord	Streptomyces sp.	LB/TSA	10C, 25C	F match is from a lake in the Himalayas, R match is from rhizosphere of Qilian Mountain.
CF4.9	Citronen Fjord	Pseudarthrobacter sp.	LB, Marine Agar	10C	Probably has PHA degrading genes, based on source of matched BLASTN sequences

CMC2.1	Cape Moris Jesup		Marine Broth/Agar, TSA	10C, 25C	Some sort of filamentous bacterium
CMC2.1 (12/4/16)	Cape Moris Jesup		Marine Agar	4C	Not filamentous. Name mix up from different isolation batches, probably.
CMS1.2	Cape Moris Jesup		TSA	4C	Bacillus, a little few were motile.
CMS1.3	Cape Moris Jesup		R2A + 10%, HM +5%, MarA	10C, 25C	Grows best in Marine agar/broth. Light brownish orange, they are bacilli. It must be closely related to CF1.1.
CMS2.1	Cape Moris Jesup		LB	10C, 25C	White fungus
CMS2.1(12/4/22)	Cape Moris Jesup		Marine Agar	4C	Not filamentous. Name mix up from different isolation batches, probably.
CMS3.1	Cape Moris Jesup	Massilia sp.	GM1	10C	Red colony. Not halophyle. Closest BLASTN match is an uncultured bacterium from biological ice nuclei from Tibet!
CMS3.3	Cape Moris Jesup		HM + 10% (can probably grow with less salt)	10C, 25C	White fungus
KK1.1	Kap København PF		Marine Agar	4C	
KK2.1	Kap København PF	Arthrobacter sp.	Marine Agar, GM1	4C	Extracellular amylases (as seen by balding of surrounding GM1 medium). Quality of 16S is not the best, but R is enough for blastn. Best match is from polar soil bacteria.
KK3.1	Kap København PF		Marine Broth/Agar	10C, 4C	Black fungus
KK3.1 (23/03/22)	Kap København PF	Micrococcus sp.	LB	10C, 25C	Showed up in liquid culture of KK3.1, which is supposed to be a black fungus… Beige phase-bright cocci that excrete a red compound after the exponential phase in LB. Best match for F 16S is a M. yunannensis sequence from desert soil from East Antarctica.
KK3.3	Kap København PF		Marine Agar	4C	
KK3.4	Kap København PF		TSA	10C	
KK4.1	Kap København PF	Arthrobacter sp.	TSA	10C	Best matches are: A. sulfonivorans (Arctic glacier) for F, A. sp (Arctic) for R
KK4.2	Kap København PF	Arthrobacter sp.	Marine Broth/Agar	10C, 4C	This genus has been used for bioremediation.
KK4.3	Kap København PF	Arthrobacter sp.	TSA, GM1 (seems to taint it red after some time?)	4C	Best match of F is Arthrobacter alpinus, from alpine soil. R is A. sp., isolated from amphibians.

KK5.1	Kap København PF	Arthrobacter sp.	TSA	10C	F match is same one as KK4.1. Best R match is A. sp., from "endolythic bacteria from Antarctica".
KK5.2	Kap København PF	Pseudarthrobacter sp. / Arthrobacter sp.	LB, Marine Agar, GM1	10C	Yellow. BLASTN returns something about a sulfonivorans species, so it could be cool to see if this has genes related to sulphur metabolism
KK5.3	Kap København PF		HM + 10%	10C, 25C	Black fungus
KK6.1	Kap København PF	Pseudomonas lini	LB, Marine Agar	10C	Isolated from soil first… But may be contaminant, since my medium was supposedly selective for actinobacteria
KK7.1	Kap København PF	Arthrobacter sp.	Marine Agar	4C	F match is A. siccitolerans from Tibet plateau.
KK8.1	Kap København PF		Marine Agar	4C	
KK9.1	Kap København PF		Marine Agar	4C	
KK9.2	Kap København PF	Pseudomonas sp.	GM1	4C	Matches are sequences from "Involvement of a laccase-like enzyme in humic substances degradation by diverse polar soil bacteria"

16s sequence data

>A.5_F

AGGAGGCGGGTGCCTATACATGCAAGTCGAGCGGACCGACGGGAGCTTGCTCCCTTAGGTCAGCGGCGGACGGGTGAGTAACAC
GTGGGTAACCTGCCTGTAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATGCTTGATTGAACCGCATGGTTCAATTA
TAAAAGGTGGCTTTTAGCTACCACTTACAGATGGACCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGA
TGCGTAGCCAACCTGAAAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATC
TTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAACTCTGTTGTTAGGGAAGA
ACAAGTACCGTTCGAATAGGGCGGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAAT
ACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCCCGGCT
CAACCGGGGAGGGTCATTGGAAACTGGGGAACTTGAGTGCAGAAAAGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAA
AGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTCTGGTCTGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCGAACAGGA
TTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGCAGCAAACGCATTA
AGCACTCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTT
TAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGACAACCCTAGAGATAGGGCTTCCCCTTCGGGGGCAAAG
TGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTATTTG
CCAGCATTCATTTGGGCACTCTAAGGTGACTGCCGTGAAAACCGAGGAAAAGGGGGGGGGATAACGTCAAATCTCATGCCCCTTT
GACTGGGCTACACCTGCTAAATGGGCAAAAAAAGGGAGGGAAACCCGAGGTAGCAATCCCAAATTTTTTCTTCTTTGGATACAAT
CTGACATCCCACGCCGGGAATAGGGAGTACATAAAATCCGAAAAAAACACCGCGGGGAGAAAATTTCGGGCTTGACACCCGCCC
CCCCACAAAGATATATAACAAAAAAAAGAAGGAACTTAGCACCCAGGGGAGATAGGGGGGAGGGGGGGGGAGGAGAGAAAT
AGGTCCCTGTGCCTTCTCGTCCTCGCCCGCCCGCAAAACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGCGCTCGCC
TGTGGGCGTTTGTTTTTTTTTTTTTTGGGGTGGCGCCGCCACCACCCC

>A.5_R

CCCATTTCTGTCACCTTCGGCGGCTGGCTCCAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTCTCGTGGTGTGACGGGCGG
TGTGTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTAGCGATTCCAGCTTCACGCAGTCGAGTTGCAG
ACTGCGATCCGAACTGAGAACAGATTTGTGGGATTGGCTTAGCCTCGCGGCTTCGCTGCCCTTTGTTCTGCCCATTGTAGCACGTG
TGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCACCTTAGAGTGCCC
AACTGAATGCTGGCAACTAAGATCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
TGCACCACCTGTCACTCTGCCCCCGAAGGGGAAGCCCTATCTCTAGGGTTGTCAGAGGATGTCAAGACCTGGTAAGGTTCTTCGC
GTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCA
GGCGGAGTGCTTAATGCGTTTGCTGCAGCACTAAAGGGCGGAAACCCTCTAACACTTAGCACTCATCGTTTACGGCGTGGACTAC
CAGGGTATCTAATCCTGTTCGCTCCCCACGCTTTCGCGCCTCAGCGTCAGTTACAGACCAGAGAGTCGCCTTCGCCACTGGTGTTC
CTCCACATCTCTACGCATTTCACCGCTACACGTGGAATTCCACTCTCCTCTTCTGCACTCAAGTTCCCCAGTTTCCAATGACCCTCC
CCGGTTGAGCCGGGGGCTTTCACATCAGACTTAAGAAACCGCCTGCGCGCGCTTTACGCCCAATAATTCCGGACAACGCTTGCCA
CCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTGGCTTTCTGGTTAGGTACCGTCAAGGTACCGCCCTATTCGAACGATA

CTTGTTCTTCCCTAACAACAGAGTTTTACGATCCGAAAACCTTCATCACTCACGCGGCGTTGCTCCGTCAGACTTTCGTCCATTGCG
GAAGATTCCTACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTACGCA
TCGTTGCCTTGGTGAACCGTTACCTCCCAACTAGCTAAGGCCCGCGGGTCCATTTGTAATGGGAACTAAAACCACCTTTTAAATTG
AACCTGGGGTTAATAAACATCCGGATTACCCCGGTTTCCGGAATTTCCAATTTTAAGGGGGGTTCCCCCGGTTTTCCCCTCCCCCG
AAAAAGAAAACTTGTCCTCATTCGTATGCCGGCCCGCCCCCCGCCCCCAGAACAACAAAAAAAAAAAAAAAAAAAAAATAAAAAA
GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAGGGGGGTTTTTTTCTTTTTTTTTTTTT

>A.6_F

AGGACTGCGGTGCCTATACATGCAAGTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACA
CGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATAACATTTTGAACTGCATGGTTCGAAA
TTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGA
TGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATC
TTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGA
ACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATA
CGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTC
AACCGTGGAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGA
GATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGCAAACAGGAT
TAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAA
GCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTT
AATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGT
GACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTATTGCC
ATCATTAAGTGGGCACTCTAAGTGACTGCCGTGACAACCGAGAGAGGGGGGGGAGAAGTCAAACATCATGCCCCTTATAACTGG
GGTACACCTGCTACAAGGAGGGTCAAAAAATGCAGAACCCCGGGGGGGGTAATTCTAAAACGTTTCCTTTCGGAATGAAGGGCC
ACCCCCCCCTAAAAGGGAAACCCTAAATCCGAAAAAAACCCCGGGGAAAAATTCCCGGCTGTACCCCCCCCCCCCCAAGATTAAC
CCAAACGGGGGACTTTGGCCCAGGGAAAAGAGTGGGGGGGGGGGGGGAGGAAAAAAAAGACAAGAAAAAAAAAAAAATAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAAGAA

>A.6_R

CCATCTCTGTCACCTTAGGCGGCTGGCTCCAAAAAGGTTACCCCACCGACTTCGGGTGTTACAAACTCTCGTGGTGTGACGGGCGG
TGTGTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTAGCGATTCCAGCTTCATGTAGGCGAGTTGCAG
CCTACAATCCGAACTGAGAACGGTTTTATGAGATTAGCTCCACCTCGCGGTCTTGCAGCTCTTTGTACCGTCCATTGTAGCACGTG
TGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCACCTTAGAGTGCCC
AACTTAATGATGGCAACTAAGATCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
TGCACCACCTGTCACTCTGCTCCCGAAGGAGAAGCCCTATCTCTAGGGTTTTCAGAGGATGTCAAGACCTGGTAAGGTTCTTCGCG
TTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAG
GCGGAGTGCTTAATGCGTTAACTTCAGCACTAAAGGGCGGAAACCCTCTAACACTTAGCACTCATCGTTTACGGCGTGGACTACC
AGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCGCCTCAGTGTCAGTTACAGACCAGAAAGTCGCCTTCGCCACTGGTGTTCC
TCCATATCTCTACGCATTTCACCGCTACACATGGAATTCCACTTTCCTCTTCTGCACTCAAGTCTCCCAGTTTCCAATGACCCTCCA
CGGTTGAGCCGTGGGCTTTCACATCAGACTTAAGAAACCACCTGCGCGCGCTTTACGCCCAATAATTCCGGATAACGCTTGCCACC
TACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTGGCTTTCTGGTTAGGTACCGTCAAGGTGCCAGCTTATTCAACTAGCACT
TGTTCTTCCCTAACAACAGAGTTTTACGACCCGAAAGCCTTCATCACTCACGCGGCGTTGCTCCGTCAGACTTTCGTCCATTGCGG
AAGATTCCTACTGCTGCCTCCCGTAGGAATCTGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTACCCAT
CGTTGCCTTGGTGAACCGTTACCTCCCACTAGCTAAGGGAACCGGGTCCTCCTAAATGAAACGCAAACCGCCTTTATTTTAAACAT
GCGGTTAAAAGGTTTCCGGGTTAACCCCGGTTTCGGAATTCCACCTTTTTGGGGGGTACCCCGTTTACCCCCCCCCCCCCAATTAAA
GAAACAATCATTTCCCATGAGTAAGCCCGGGGGCCGCGCCCGCCCACGCAACAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAGGTGTGTTTGTTTTATAAGAAAAAAAAAAAAAAAAAAATGTGGTCGTTGCGTGGCGGCTGGCGGTCCTGTCGGTCCTCCGGCC
ACCAGACTTCTTGTGTTTGTGTCTGTTGTT

>A.9_F

AGGGCTGTGGCGGCGGCTACACATGCAGTCGAGCGGATGAGGGGGGCTTGCTCCCTGATTCAGCGGCGGACGGGTGAGTAATGC
CTAGGAATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGGAACGCTAATACCGCATACGTCCTACGGGAGAAAGCAGGGGACC
TTCGGGCCTTGCGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACTG
GTCTGAGAGGATGATCAGTCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATG
GGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTA
AGTTAATACCTTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGC
AAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGGATGTGAAAGCCCCGGGCTCAACCTGGGAA
CTGCATCCAAAACTGGCAAGCTAGAGTATGGCAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGG
AACACCAGTGGCGAAGGCGACCACCTGGGCTAATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT
GGTAGTCCACGCCGTAAACGATGTCGACTAGCCGTTGGGATCCTTGAGATCTTAGTGGCGCAGCTAACGCATTAAGTCGACCGCC
TGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGC
AACGCGAAGAACCTTACCAGGCCTTGACATGCAGAGAACTTTCCAGAGATGGATGGGTGCCTTCGGGAACTCTGACACAGGTGCT
GCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCCTTAGTTACAGCACGTGAT
GGTGGGCACTCTAAGGAGACTGCCGTGACAACCGGAGAGAGAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGGCTGGGC

```
TACACGCTGCTACAAGGGGCGGGAAAAAGGGTGCCAACGCGCGAGGTGGACTATCCCTAAAACGAACGAAATCCGGATCCAGTT
GCACCCGCCGGGGTGAGTCGAAACCCTAAAAATGGAAAAAAAAAAAACGGGGAAAAAATTCCGGCTTTCCGCCCCCCCCCCCCCCGG
GGGGGGGTGGCCAAAAATAAAACCTCCGGGAGGGGCCGAGATACTGAGGGGGGGGGGGGGAAGGAAGAAAAAAGGAGTCGGT
CTCTGTCTTAAAAAAATAAAAATAAAAAAAAAAAAAAAAAAATAAAAATTAAAAAATACTCCCCCCCCCACCCCCGAGGGGGGG
GGGGAGGGGGGGGGGGGGGGGGGGGGGGGCAAAAAACATCTTTTTTGTGTTCTTTGTGTTG
```

>A.9_R

```
CAATCTGATCCTCCGTGGTACCGTCCCCCCGAAGGTTAGACTAGCTACTTCTGGAGCAACCCACTCCCATGGTGTGACGGGCGGTG
TGTACAAGGCCCGGGAACGTATTCACCGTGACATTCTGATTCACGATTACTAGCGATTCCGACTTCACGCAGTCGAGTTGCAGACT
GCGATCCGGACTACGATCGGTTTTATGGGATTAGCTCCACCTCGCGGCTTGGCAACCCTTTGTACCGACCATTGTAGCACGTGTGT
AGCCCAGGCCGTAAGGGCCATGATGACTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCTCCTTAGAGTGCCCACC
ATCACGTGCTGGTAACTAAGGACAAGGGTTGCGCTCGTTACGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCAT
GCAGCACCTGTGTCAGAGCTCCCGAAGGCACCCATCCATCTCTGGAAAGTTCTCTGCATGTCAAGGCCTGGTAAGGTTCTTCGCGT
TGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCGTACTCCCCAG
GCGGTCGACTTAATGCGTTAGCTGCGCCACTAAGATCTCAAGGATCCCAACGGCTAGTCGACATCGTTTACGGCGTGGACTACCA
GGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCACCTCAGTGTCAGTATTAGCCCAGGTGGTCGCCTTCGCCACTGGTGTTCCTT
CCTATATCTACGCATTTCACCGCTACACAGGAAATTCCACCACCCTCTGCCATACTCTAGCTTGCCAGTTTTGGATGCAGTTCCCA
GGTTGAGCCCGGGGCTTTCACATCCAACTTAACAAACCACCTACGCGCGCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCT
TCGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGTGCTTATTCTGTCGGTAACGTCAAAACAGCAAGGTATTAACTTACTGCCC
TTCCTCCCAACTTAAAGTGCTTTACAATCCGAAGACCTTCTTCCACACGCGGCATGGCTGGATCAGGCTTTCGCCCATTGTCCATA
TTCCCAACTGCTGCCTCCCGTAGGAATCTGGACCGTGTCTCAGTTCCAGTGTGACTGATCATCCTCTCAGACCGTTACGGATCGTC
GCCTTGGTGAACCATTACCTCCCAACTAGCTAATCCGACCTAGGCTCACTTGAAAGCCAAGGCCGAAGGCCCTGTTTTTCCCCAAG
AAAGAAGGCGGATAAGGTTCTTTCAACGTTTCCCCACAACAGGGAATCTAGGTTTTCCCGTCCCCCGAAAGGGAAGCCCTACCCC
ATGAGAGGCGGGGGGGGGGGGGGGAAAAACAAAAAAAAAAAAATAAAAAAAAAATAAAAAAAACAAAATCAATAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAATGATGAAAAAAATCACCCCCCCCCCCGCCCCCCCCCGCGCCCCCCCCCTCCTTTTTATA
TTTAAGAGAGGGGAGA
```

>A.10_F

```
AGGAGGGCGGCGTGCCTATACATGCAAGTCGAGCGGACAGATGGGAGCTTGCTCCCCTGATGTTAGCGGCGGACGGGTGAGTAA
CACGTGGGTAACCTGCCTGTAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATGCTTGTTTGAACCGCAGGTTCAAA
CATAAAAGGTGGCTTCGGCTACCACTTACAGATGGACCCGCGGCGCATAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACG
ATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAAT
CTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTAGGGAAG
AACAAGTACCGTTCGAATAGGGCGGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAA
TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGCAGGCGGTTCCTTAAGTCTGATGTGAAAGCCCCCGGC
TCAACCGGGGAGGGTCATTGGAAACTGGGGAACTTGAGTGCAGAAGAGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTA
GAGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTCTGGTCTGTAACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGG
ATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGGGGGTTTCCGCCCCTTAGTGCTGCAGCTAACGCATT
AAGCACTCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGT
TTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGACAATCCTAGAGATAGGACGTCCCCTTCGGGGGCAGA
GTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGTG
CCAGCATTCATTGGGCACTCTAAGTGACTGCCGTGACAACCGAGAAGAAGGGGGGGGATGACGTCAAATCTCATGCCCCTATGAC
TGGGCTACCCCTGCTACATGGAAGAAAAAAGGGAGCGAAAACCGCAGGTAACCATCCCCAAATTGTTTTTTTTTGGAACCAATTG
GACTCCACGGGTGAAAGGGAAACCCTTAATCCGGACAAAACCCCCCGGGGAAAAATTCCCGGCCTTTACCCCCGCCCCCCCCAAA
TTTTAACCCCACACCGGGGAACTTGCCGCCCGGGAATGGGGGGGGTGGGGGGAGGGGAAAAAAAAGAGTCTTTTTGCGCTAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCTGTCCTTTCTCGGCGGGGGTGGGGTGGGGTTGGGGGGGGGG
GTGAA
```

>A.10_R

```
CCAATCTCTGTCCACCTTCGGCGGCTGGCTCCATAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTCTCGTGGTGTGACGGGC
GGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTAGCGATTCCAGCTTCACGCAGTCGAGTTGC
AGACTGCGATCCGAACTGAGAACAGATTTGTGGGATTGGCTTAACCTCGCGGTTTCGCTGCCCTTTGTTCTGTCCATTGTAGCACG
TGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCACCTTAGAGTGC
CCAACTGAATGCTGGCAACTAAGATCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAAC
CATGCACCACCTGTCACTCTGCCCCCGAAGGGGACGTCCTATCTCTAGGATTGTCAGAGGATGTCAAGACCTGGTAAGGTTCTTCG
CGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCC
AGGCGGAGTGCTTAATGCGTTAGCTGCAGCACTAAGGGGCGGAAACCCCCTAACACTTAGCACTCATCGTTTACGGCGTGGACTA
CCAGGGTATCTAATCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTTACAGACCAGAGAGTCGCCTTCGCCACTGGTGTT
CCTCCACATCTCTACGCATTTCACCGCTACACGTGGAATTCCACTCTCCTCTTCTGCACTCAAGTTCCCCAGTTTCCAATGACCCTC
CCCGGTTGAGCCGGGGGCTTTCACATCAGACTTAAGGAACCGCCTGCGAGCCCTTTACGCCCAATAATTCCGGACAACGCTGGCC
ACCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTGGCTTTCTGGTTAGGTACCGTCAAGGTACCGCCCTATTCGAACGGT
ACTTGTTCTTCCCTAACAACAGAGCTTTACGATCCGAAAACCTTCATCACTCACGCGGCGTTGCTCCGTCAGACTTTCGTCCATTGC
```

GGAAGATTCCTACTGCTGCCTCCCGTAGGAATCTGGGCCGGGTCTCAATCCCAGGGTGGCCAATACCCTCTCAGGTCGGCTACC
ACCCGTTGCCCTGGTGAACCGTTACCTCCCAACAACTAAAGGGCCCGGGGCCCTCTGTTAATGGAGCAGAACCCCCCTTTTTTTGT
TAGAAAGGGGTTAAAAAAAAACCGATTAAACCACCGGTTTTCGGTGATAACAACCTTCAGGGAGGGAACCCCCGGTTTTTCCCTC
CCCCCCCAAACGGAAAAACCCCGCCCCCATCGTAGCCGGGGGGGGGGGGGGGGAATACATAAAAAAAAAAAAAATAATATAAAT
ACTAAAACACAAACAAAAAAAAGGCAAAAAAAAAAAAAAATAAAAAAAAAAAAAAACAAAAAGAAGACGTAAGAGCCTGGGGTC
GGGCTCTCGTGCTGCGGGGCGTCGCGCCGCGTGCCCCCCTGGT

>A.11_F

CGCATGCGCATCCTATACATGCAGTCGAGCGGACTGATGGGAGCTTGCTCCCTGAAGTCTCGGCGGACGGGTGAGTAACACGTGG
GTAACCTGCCTGTAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGAAGTTCTTTTCCCCGCACGAGGAAAAGTGGAAA
GGTGGCTTTTAGCTACCACTTACAGATGGACCCGCGGCGCAATAACTAGTTGGTGAGGTAACGGCTCACCAAGGCGACCATTCGT
AGCCCACCTGAGAGGGTGATCGGCCACACTGGGACTGAAACACGGCCCAAACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCG
CAATGGACGAAAGTCTGACGGAACAACGCCCCGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTTTTGTTAGGGAAGAACAAG
TGCCGTTCGAATAGGGCGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAA
GTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTCTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCG
TGGAGGGTCATTGGAAACTGGGGGACTTGAGTACAGAAGAAGAAAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGATATGT
GGAGGAACACCAGTGGCGAAAGCGACTTTCTGGTCTGTAACTGACGCTGAAGCGCGAAAGCGTGGGGAGCAAACCAGGAATAGA
TACCCTGGTAGTCCACAGCCGTAAACGATGAGATGGCTAAGTGTTTAGAGGGTTTCCGCCCCTTTAGTGCTGCAGCTAACGCAATT
AAGCACTTCCGGCCTGGGGGAGTACGGTCCGCAAGACTGAAACTCAAAAGGAATTGACCGGGGCCCGCCACAAGCGGTGGAGCA
TGTTGGTTTAATTCGAAGCAACGCGAGAACCTTACCAAGTCTTGACATCTCTGACAACCCTAGAGATAGGCTTTCTCCTTCGGAGA
CAGAGTGACAGTGGTGCATGTGTCGTCAGCTCGTGTCGTGAGAATGTTTGGGTTAGTCCCGCCACGAGCGCACCTTGAATCTAATG
CAGCATCAGTGGCACTCTAAGATGACTGCGGTACAACGAGAAAGGTGGGATAACTTCAATCATCTATGCTTAGACTGGTCAACGT
CTCAATGACGGTACAAGTGCGAA

>A.11_R

ACACTTCTGTCACTTAGGCGGCTGGCTCCAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTCTCGTGGTGTGACGGGCGGTGT
GTACCAGGGCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACCAGCGATTCCCGCTTCATGCAGGCGAATTGCAGCCT
GCAATCCGAACTGAGAACGGTTTTATGGGATTTGCTAAACCTCGCGGTCTTGCTGCCCTTTGTACCGTCCATTGTAGCACGTGTGT
AGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCATCTTAGAGTGCCCAAC
TGAATGCTGGCAACTAAGATCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACCACCATGC
ACCACCTGTCACTCTGTCCCCCGAAGGGGAAAGCCCTATCTCTAGGGTTGTCAGAGGATGTCAAGACCTGGTAAGGTTCTTCGCGT
TGCTTCAAATTAAACCACATGCTCCACCGCTGGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAG
GCGGAGTGCTTAATGCGTTAGCTGCAGCACTAAAGGGCGGAAACCCTCTAACACTTAGCACTCATCGTTTACGGCGTGGACTACC
AGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCGCCTCAGCGTCAGTTACAGACCAAAAAGTCGCCTTCGCCACTGGTGTTCC
TCCACATATCTACGCATTTCACCGCTACACGTGGAATTCCACTTTCCTCTTCTGTACTCAAGTCCCCCAGTTTCCAATGACCCTCCA
CGGTTGAGCCGTGGGCTTTCACATCACACTTAAGAGACCGCCTGCGCGCGCTTTACGCCCAATAATTCCGGACAACGCTGGCCAC
CTACGTATTACCGCGGCTGCTGGCACGTATTTAGCCGTGGCTTTCTGGTTAGGTACCGTCAAGGTGCCGCCCTATTCGAACGGCAC
TTGTTCTTCCCTAACAACAGATCTTTACAATCCGAAAACCTTCATCACTCACGCGGCGTTGCTCCGTCAAATTTCGTCCATTGCCGA
AAAATCCCTACTGCTGCCTCCCGTAAGAATCTGGGCCGGGTCCAATCCCAGTGGGCCAAACCCCTTCCAGGTCGCTACCATCGTCC
CCTTGGTAACCGTACCTCCCAATAGCTAAGGCCCCGGGTCCCTCTTAAAGGGAACCAAAACCCCCTTTCCCCTAACTCGGGCGGA
AAAAAAATTTCCGGGTTATAACCGCGGTTTTGGGAATTCACACATTATAGGGAAGATCACCGTTATACCCCCCCCCCCTTAATAAG
GAAAACTCAACTTCCCCCCATTGGTAAGCCCCGTCCGCCTGCCCGGACTCAAAACAATAATTAAAAAAAAAGGGCGGGTGGGAA
AAAAAAAAAAAAAAAGTTTGTGGGTGGCGTTTCTGTGCTCCGCTGTTGCTGGCGGGCGTGTTTGTGGGCGGTGTTGGGTGTCGGT
CGGCGGGCTGTGCGCGGGGGCGTGTGCTGTGTGGCGGTGCTCGTGCGGCCGCTCTTCTTGGCGCTTGTCTTCTGTGTCCGGCGGGT
TGTGTGTCCGTGGGTCGAGGGTGGTCGCTCTGGGCCGCCGTCTGTGTGTTCT

>CF4.2_F

AAACAAGCAGATCCCTTCGGGGTGACACTTGTGGAACGAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGATT
GGGATAACTCGCGGAAACGTGAGCTAATACCGAATAATACTTTTTGCCTCCTGGCAAGAAGATGAAAGGCGGCTTCGGCTGTCAC
TTACAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGGGGTAATGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAGGGTG
ATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGA
CGGAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAACTCTGTTGTCAGGGAAGAACAAGTGCGAGAGTAACTGCTCG
CACCTTGACGGTACCTGACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGA
ATTATTGGGCGTAAAGCGCTCGCAGGCGGTCTTTTAAGTCTGATGTGAAAGCCCACGGCTTAACCGTGGAGGGTCATTGGAAACT
GGAGGACTTGAGTACAGAAGAGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGA
AGGCGACTCTCTGGTCTGTAACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGT
AAACGATGAGTGCTAGGTGTTAAGGGGGTTTCCGCCCCTTAGTGCTGAAGTTAACGCATTAAGCACTCCGCCTGGGGAGTACGGC
CGCAAGGCTGAAACTCAAAAGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACGCAACGCGAAGAACC
TTACCAGGTCTTGACATCCTCTGACATCCTAGAGATAGGACGTTCCCTTCGGGGACAGAGTGA

>CF4.2_R

CATCGACTTCGGGTGTTACCAACTCTCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGAT
CCGCGATTACTAGCGATTCCGGCTTCATGTAGGCGAGTTGCAGCCTACAATCCGAACTGAGAATGGTTTTATGGGATTTGCTTGAC
CTCGCGGTTTTGCTTCCCTTTGTTCCATCCATTGTAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCC
ACCTTCCTCCGGTTTGTCACCGGCAGTCACCTTAGAGTGCCCAACTGAATGCTGGCAACTAAGATCAAGGGTTGCGCTCGTTGCGG
GACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCACTCTGTCCCCGAAGGGAACGCCCTATCTCT
AGGATTGTCAGAGGATGTCAAGACCTGGTAAGGTTCTTCGCGTTGCGTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCC
CGTCAATTCTTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGCGGAGTGCTTAATGCGTTAACTTCAGCACTAAGGGGCGGAAA
CCCCCTAACACCTAGCACTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGC
GTCAGTTACAGACCAGAGAGTCGCCTTCGCCACTGGTGTTCCTCCACATCTCTACGCATTTCACCGCTACACGTGGAATTCCACTC
TCCTCTTCTGTACTCAAGTCCTCCAGTTTCCAATGACCCTCCACGGTTAAGCCGTGGGCTTTCACATCAGACTTAAAAGACCGCCT
GCGAGCGCTTTACGCCCAATAATTCCGGACAACGCTTGCCACCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTGGCTTT
CTGGTCAGGTACCGTCAAGGTGCGAGCAGTACTCTCGCACTTGTTCTTCCTGACAACAGAGTTTTACGATCCGAAAACTTCATCAT
TCAAGC

>CF4.4_F

TCTAACGATGGAAGACCGTGACTTGCACGGTCGGATTAGTGGCGAACGGGTGAGTATCACGTGAGTAACCTTCCCTTGACTCTGG
GATAAGCCCGGGAAACTGGGTCTAATACCGGATACGACCAGTCCTCGCATGGGGTGCTGGTGGAAAGATTTATCGGTCTTGGATG
GACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGATGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGAC
GCCGCGTGCGGGATGACGGCCTTCGGGTTGTAAACCGCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACCTGCAGAAGAAGCG
CCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCGAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCG
GTTTGTCACGTCTGCTGTGAAAGCCCGAGGCTCAACCTCGGGTGTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACT
GGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCTGTTACTGACGCT
GAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGAACA
TTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGAC
GGGGCCCCGCACAAACGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATGGACCGGATCG
CTGCAGAGATGCAGTTTCCCTTCGGGGTCGGTTCACAGGTGGGGCATGGTTGTCGTCAGCTCGTGTCGGGA

>CF4.4_R

GGTTAGGCCACCGGCTTCGGGTGTTACCCACTTTCGTGACTTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGC
GTTGCTGATCTGCGATTACTAGCGACTCCAACTTCACGAAGTCGAGTTGCAGACTTCGATCCGAACTGAGACCGGCTTTTAGGGAT
TAGCTCCACCTCACAGTATCGCAACCCATTGTACCGGCCATTGTAGCATGCGTGAAGCCCAAGACATAAGGGGCATGATGATTTG
ACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCCATGAGTCCCCACCATAACGTGCTGGCAACATAGGATAGGGGTT
GCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTGAACCGACCCCGAAGGGA
AACTGCATCTCTGCAGCGATCCGGTCCATGTCAAGCCTTGGTAAGGTTCTTCGCGTTGCATCGAATTAATCCGCATGCTCCGCCGC
TTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAGGCGGGGCACTTAATGCGTTAGCTACGGCGC
GGAAAACGTGGAATGTTCCCCACACCTAGTGCCCAACGTTTACGGCATGGACTACCAGGGTATCTAATCCTGTTCGCTCCCCATGC
TTTCGCTCCTCAGCGTCAGTAACAGCCCAGAGACCTGCCTTCGCCATCGGTGTTCCTCCTGATATCTGCGCATTTCACCGCTACACC
AGGAATTCCAGTCTCCCCTACTGCACTCTAGTCTGCCCGTACCCACTGCACACCCGAGGTTGAGCCTCGGGCTTTCACAGCAGACG
TGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTCGCGCCCTACGTATTACCGCGGCTGCTGGCACGTAGT
TAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTTCCTGCTGAAAGCGGTTACAACCCGAAGGCCGTCATCCCGCAAGC
GGCGTCGCTGCATCAGGCTTTCGCCCATTGTGCAATATTCCCCATGCT

>CF1.1_F

CTGCTAATACATGCAAGTCGAGCGGAGTTAGAGAAGCTTGCTTCTCTAACTTAGCGGCGGACGGGTGAGTAACACGTGGGCAATC
TGCCCGTAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATAATCCTGACTCTCTCCTGAGAGTCAGTTGAAAGATGG
TTTCGGCTATCACTTACGGATGAGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATGCGTAGCCGA
CCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGG
ACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAACTCTGTTGTTAGGGAAGAACACGTACGA
GAGTAACTGCTCGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCA
AGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTCTTTAAGTCTGATGTGAAAGCCCCCGGCTCAACCGGGGAGG
GTCATTGGAAACTGGGGAACTTGAGTGCAGGAGAGGAAAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAATTGTGGAGGA
ACACCAGTGGCGAAGGCGACTTTCTGGCCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTG
GTAGTCCACGCCGTAAACGATGAGTGCTAGTGGTTGGGGGGGTTTCCGCCCCTTCATT

>CF1.1_R

TTCTGTCCCCTTAGGCGGCTGGCTCCATAAATGGGTTACCCCACCGACTTCGGGTGTTACAAACTCTCGTGGTGTGACGGGCGGTG
TGTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTAGCGATTCCAGCTTCATGTAGGCGAGTTGCAGCC
TACAATCCGAACTGAGAACGATTTTATGGGATTGGCTCCACCTCGCGGTCTTGCTGCCCTTTGTATCGTCCATTGTAGCACGTGTGT
AGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCCCCTTCCTCCGGTTTGTCGCCGGCAGTGACCTTAGAGTGCCCAAC
TAAATGCTGGCGACTAAGAGCAGGGG

>CF3.3_F

```
GCGGGTGCTTACACATGCAAGTCGAACGATGAAGACCGTGCTTGCACGGTTGGATTAGTGGCGAACGGGTGAGTATCACGTGAGT
AACCTTCCCTTAACTCTGGGATAAGCCCGGGAAACTGGGTCTAATACCGGATACGACCAGTCCTCGCATGGGGTGCTGGTGGAAA
GATTTATCGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAAAGGCTCACCAAGGCGATGACGGGTAGCCGGCCT
GAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCG
AAAGCCTGATGCAGCGACGCCGCGTGCGGGATGACGGCCTTCGGGTTGTAAACCGCTTTCAGCAGGGAAGAAGCGCAAGTGACG
GTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCGAGCGTTATCCGGAATTATTGGG
CGTAAAGAGCTCGTAGGCGGTTTGTCACGTCTGCTGTGAAAGCCCGGGGCTCAACCCCGGGTGTGCAGTGGGTACGGGCAGACTA
GAGTGCAGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTC
TCTGGGCTGTTACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGG
GCACTAGGTGTGGGGAACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCT
AAAACTCACAGGAATTGACGGCGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGG
CTTGACATGGACCGGACCGCTGCAGAGATGCAGTTTCCCTTCGGGGTCGGTTCACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTC
GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCACCCCTATCCTATGTTGCCAGCACGTTATGGTGGGGACTCATGGGAGACTGC
CGGGGTCAC

>CF3.3_R

GGAGTCACCTTCGACGGCTCCCCCACAAGGGTTAGGCACCGGCTTCGGGTGTTACCCACTTTCGTGACTTGACGGGCGGTGTGTAC
AAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCAACTTCACGAAGTCGAGTTGCAGACTTCG
ATCCGAACTGAGACCGGCTTTTAGGGATTAGCTCCACCTCACAGTATCGCAACCCATTGTACCGGCCATTGTAGCATGCGTGAAG
CCCAAGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCCATGAGTCCCCACCAT
AACGTGCTGGCAACATAGGATAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGC
ACCACCTGTGAACCGACCCCGAAGGGAAACTGCATCTCTGCAGCGGTCCGGTCCATGTCAAGCCTTGGTAAGGTTCTTCGCGTTG
CATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAGGC
GGGGCACTTAATGCGTTAGCTACGGCGCGGAAAACGTGGAATGTTCCCCACACCTAGTGCCCAACGTTTACGGCATGGACTACCA
GGGTATCTAATCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTAACAGCCCAGAGACCTGCCTTCGCCATCGGTGTTCCTC
CTGATATCTGCGCATTTCACCGCTACACCAGGAATTCCAGTCTCCCCTACTGCACTCTAGTCTGCCCGTACCCACTGCACACCCCG
GGGTTGAGCCCCGGGCTTTCACAGCAGACGTGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTCGCGCC
CTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTGCGCTTCTTCCCTGCTGAAAGC
GGTTTACAACCCGAAGGCCGTCATCCCGCACGCGGCGTCGCTGCATCAGGCTTTCGCCCATTGTGCAATATTCCCACTGCTGCCTC
CCGTAAGAA

>CF4.1_F

GTGCTAATACATGCAAGTCGAGCGCAGGAAACAAGTTGATCCCTTCGGGGTGACGCTTGTGGAATGAGCGGCGGACGGGTGAGT
AACACGTGGGCAACCTGCCTTGTAGACTGGGATAACTCCGGGAAACCGGAGCTAATACCGGATGACCAACGGAATCGCATGATT
CTGTTGTAAAAGTGGGGATTTATCCTCACACTACGAGATGGGCCCGCGGCGCATTAGCTAGTTGGTAAGGTAATGGCTTACCAAG
GCAACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTA
GGGAATCATCCGCAATGGGCGAAAGCCTGACGGTGCAACGCCGCGTGAACGATGAAGGTTTTCGGATCGTAAAGTTCTGTTATGA
GGGAAGAACAAGTGCCGTTCGAATAGGGCGGCACCTTGACGGTACCTCACGAGAAAGCCCCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTCTCTTAAGTCCGATGTGAAAGCC
CACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGGGACTTGAGTGTAGGAGAGGAAAGTGGAATTCCACGTGTAGCGGTGAAA
TGCGTAGATATGTGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGCCTACAACTGACGCTGAGGTGCGAAAGCGTGGGGAGCA
AACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAGGTGTTAGGGGTTTCGATGCCCTTAGTGCCGAAGTTA
ACACATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAGACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGC
ATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGACCACCCAAGAGATTGGGATTTCCCCTTCGG
GGGACAGAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
ATCTTAATTGCCAGCATTCAGTTGGGCACTCTAAGGTGACTGCCGGTGATAAACCGGAGGAA

>CF4.1_R

ACTCTGTCCACCTTCGGCGGCTGGCTCCAAAAGGTTACCTCACCGACTTCGGGTGTTACAAACTCTCGTGGTGTGACGGGCGGTGT
GTACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTAGCAATTCCGGCTTCATGCAGGCGAGTTGCAGCCT
GCAATCCGAACTGAGAATGGCTTTATGGGATTCGCTTGACCTCGCGGTCTTGCAGCCCTTTGTACCATCCATTGTAGCACGTGTGT
AGCCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGTTTATCACCGGCAGTCACCTTAGAGTGCCCAAC
TGAATGCTGGCAACTAAGATCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGC
ACCACCTGTCACTCTGTCCCCCGAAGGGGAAATCCCAATCTCTTGGGTGGTCAGAGGATGTCAAGACCTGGTAAGGTTCTTCGCGT
TGCTTCGAATTAAACCACATGCTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGG
CGGAGTGCTTAATGTGTTAACTTCGGCACTAAGGGCATCGAAACCCCTAACACCTAGCACTCATCGTTTACGGCGTGGACTACCA
GGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCACCTCAGCGTCAGTTGTAGGCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTC
CACATATCTACGCATTTCACCGCTACACGTGGAATTCCACTTTCCTCTCCTACACTCAAGTCCCCCAGTTTCCAATGACCCTCCACG
GTTGAGCCGTGGGCTTTCACATCGGACTTAAGAGACCGCCTGCGCGCGCTTTACGCCCAATAATTCCGGACAACGCTTGCCCCCTA
CGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGGGCTTTCTCGTGAGGTACCGTCAAGGTGCCGCCCTATTCGAACGGCACTTG
TTCTTCCCTCATAACAGAACTTTACGATCCGAAAACCTTCATCGTTCACGCGGCGTTGCACCGTCAGGCTTCGCCCATTGCGGATG
ATTCCTACTGCTGCCTCCCGAAG
```

```
>CF4.5_F

AGTGAACTAGTGGAACGAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTACCTGTAAGATTGGGATAACTCGCGGAAACGTG
AGCTAATACCGAATAATACTTTTTATCTCCTGATAAGAAGATGAAAGGCGGCTTATAGCTGTCACTTACAGATGGGCCCGCGGCG
CATTAGCTAGTTGGTGGGGTAATGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGA
GACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGT
GATGAAGGTTTTCGGATCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCGAGAGTAACTGCTCGCACCTTGACGGTACCTAACC
AGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCT
CGCAGGCGGTCTTTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGAGGACTTGAGTACAGAAG
AGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTCTGGTCTGTA
ACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAGGTG
TT

>CF4.5_R

TGGAGTCACCCTTCGGCGGCTGGCTCCAAAAGGTTACCTCACCGACTTCGGGTGTTACCAACTCTCGTGGTGTGACGGGCGGTGTG
TACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTG
CAATCCGAACTGAGAATGGTTTTATGGGATTTGCTTGACCTCGCGGTTTTGCTTCCCTTTGTTCCATCCATTGTAGCACGTGTGTAG
CCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCACCTTAGAGTGCCCAACTA
AATGCTGGCAACTAAGATCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCAC
CACCTGTCACTCTGTCCCCGAAGGGAACGTCCTATCTCTAGGATTGTCAGAGGATGTCAAGACCTGGTAAGGTTCTTCGCGTTGCG
TCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCTTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGCGG
AGTGCTTAATGCGTTAACTTCAGCACTAAGGGGCGGAAACCCCCTAACACCTAGCACTCATCGTTTACGGCGTGGACTACCAGGG
TATCTAATCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTTACAGACCAGAGAGTCGCCTTCGCCACTGGTGTTCCTCCAC
ATCTCTACGCATTTCACCGCTACACGTGGAATTCCACTCTCCTCTTCTGTACTCAAGTCCTCCAGTTTCCAATGACCCTCCACGGTT
GAGCCGTGGGCTTTCACATCAGACTTAAAAGACCGCCTGCGAGCGCTTTACGCCCAATAATTCCGGACAACGCTTGCCACCTACG
TATTACCGCGGCTGCTGGCACGTAATTAGCCGTGGCTTTTCTGGTTAGGTACCGTCAAGGTGCGAGCAGTTACTCTCGCACTTGTT
CTTCCTAACAACAGAGTTTTACGATCCGAAAACCTTCATCACTCACGCGGCGTTGCTCCGTCAGACTT

>CF4.6_F

TCAGCACCTCTTCGGAGAGTGAACTAGTGGAACGAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTACCTGTAAGATTGGGA
TAACTCGCGGAAACGTGAGCTAATACCGAATAATACTTTTTATCTCCTGATAAGAAGATGAAAGGCGGCTTATAGCTGTCACTTAC
AGATGGGCCCGCGGCGCATTAGCTAGTTGGTGGGGTAATGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAGGGTGATCG
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGA
GCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCGAGAGTAACTGCTCGCACC
TTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTA
TTGGGCGTAAAGCGCTCGCAGGCGGTCTTTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGAG
GACTTGAGTACAGAAGAGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGC
GACTCTCTGGTCTGTAACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAAC
GATGAGTGCTAGGTGTTTAGGGGGTTTCCGCCCCTTAGTGCTGAAGTTAACGCATTAAGCACTCCGCCTGGGGAGTACGGCCGCA
AGGCTGAAACTCAAAAGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACGCAACGCGAAGAACCTTAC
CAGGTCTTGACATCCTCTGACACTCCTAGAAGATAGGACGTTCCCTTCGGGGACAGAGTGACAGGTGGGTGCATGGTTGTCGTCA
GCTCGTGTCGTGAGAA

>CF4.6_R

TCAGTCCACCTTTCGGCGGCTGGCTCCAAAAGGTTACCTCACCGACTTCGGGTGTTACCAACTCTCGTGGTGTGACGGGCGGTGTG
TACAAGGCCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTG
CAATCCGAACTGAGAATGGTTTTATGGGATTTGCTTGACCTCGCGGTTTTGCTTCCCTTTGTTCCATCCATTGTAGCACGTGTGTAG
CCCAGGTCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCACCTTAGAGTGCCCAACTA
AATGCTGGCAACTAAGATCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCAC
CACCTGTCACTCTGTCCCCGAAGGGAACGTCCTATCTCTAGGATTGTCAGAGGATGTCAAGACCTGGTAAGGTTCTTCGCGTTGCG
TCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCTTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGCGG
AGTGCTTAATGCGTTAACTTCAGCACTAAGGGGCGGAAACCCCCTAACACCTAGCACTCATCGTTTACGGCGTGGACTACCAGGG
TATCTAATCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTTACAGACCAGAGAGTCGCCTTCGCCACTGGTGTTCCTCCAC
ATCTCTACGCATTTCACCGCTACACGTGGAATTCCACTCTCCTCTTCTGTACTCAAGTCCTCCAGTTTCCAATGACCCTCCACGGTT
GAGCCGTGGGCTTTCACATCAGACTTAAAAGACCGCCTGCGAGCGCTTTACGCCCAATAATTCCGGACAACGCTTGCCACCTACG
TATTACCGCGGCTGCTGGCACGTAGTTAGCCGTGGCTTTCTGGTTAGGTACCGTCAAGGTGCGAGCAGTTACTCTCGCACTTGTTC
TTCCTAACAACAGAGTTTTACGATCCGAAAACCTTCATCACTCACGCGGCGTTGCTCCGTCAGACTT

>CF4_9_F

TGCGGCGCTTACACATGCAAGTCGAACGATGAACCTCACTTGTGGGGGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACC
TGCCCTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACTCCTCATCGCATGGTGGGGGGTGGAAAGCTTT
ATTGTGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAG
AGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAA
```

GCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTA
CCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGT
AAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGGATCTGCGGTGGGTACGGGCAGACTAGAG
TGATGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCT
GGGCATTAACTGACGCTGAGGAGCGAAAGCATGCGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTACACGTTGGGC
ACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGAGAGTACGGCCGCAAGGCTAA
AACTCAGAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCT
TGACATGAACCGGAAAGACCTGGAAACAGGTGCCCCGCTTGCGGTCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCG
TGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCGCGTTATGGCGGGGACTCATAGGAGACTGC
CGGGTCACTCGGAGGAAGGTGGGGACGACGTCAAATCTTCATGCCCCTTATGT

>CF4.9_R

TTGCGTCACCTTCGAACAGCTCCCTCCCACAAGGGGTTAGTGACCACCGGCTTCGGGTGTTACCAACTTTCGTGACTTGACGGGCG
GTGTGTACAAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGC
AGACCCCAATCCGAACTGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCGCAACCCTTTGTACCGGCCATTGTAGCATG
CGTGAAGCCCAAGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTATGAGTCC
CCGCCATAACGCGCTGGCAACATAGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACA
ACCATGCACCACCTGTAAACCGACCGCAAGCGGGGCACCTGTTTCCAGGTCTTTCCGGTTCATGTCAAGCCTTGGTAAGGTTCTTC
GCGTTGCATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCC
CAGGCGGGGCACTTAATGCGTTAGCTACGGCGCGGAAAACGTGGAATGTCCCCCACACCTAGTGCCCAACGTTTACGGCATGGAC
TACCAGGGTATCTAATCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTTAATGCCCAGAGACCTGCCTTCGCCATCGGTGT
TCCTCCTGATATCTGCGCATTTCACCGCTACACCAGGAATTCCAGTCTCCCCTACATCACTCTAGTCTGCCCGTACCCACCGCAGA
ATCCGGAGTTGAGCCCCGGACTTTCACGGCAGACGCGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTT
GCGCCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTCCCTACCTGA
AAGAGGTTTACAACCCGAAGGCCGTCATCCCTCAGGCGGCGTCGCTGCATCAGGCTTGCGC

>KK4.2_F

CGGGTCTTACACATGCAAGTCGAACGATGATCTCCAGCTTGCTGGGGGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACC
TGCCCTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACTCCTCATCGCATGGTGGGGGGTGGAAAGCTTT
TTGTGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGA
GGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAG
CCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTACC
TGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAA
AGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTCAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTG
ATGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGG
GCATTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCAC
TAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGGAGTACGGCCGCAAGGCTAAA
ACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTT
GACATGGACCGGACCGGGCTGGAAACAGTCCTTCCCCTTTGGGGTCGGTTCACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGT
GAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCGCACGTGATGGTGGGGACTCATAGGAGACTGCCG
GGTCACTCGGAGGAAGGTGGGGACGACGTCAATCATCATGCCCCTTATGTCTTGGGCTTCCCGCTGCTACATGGCCGGT

>KK4.2_R

TGCGTCCACCTTCGAACAGCTCCCTCCCACAAGGGGTTAGGCCACCGGCTTCGGGTGTTACCAACTTTCGTGACTTGACGGGCGGT
GTGTACAAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAG
ACCCCAATCCGAACTGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCGCAACCCTTTGTACCGGCCATTGTAGCATGCG
TGAAGCCCAAGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTATGAGTCCCC
ACCATCACGTGCTGGCAACATAGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAAC
CATGCACCACCTGTGAACCGACCCCAAAGGGGAAGGACTGTTTCCAGCCCGGTCCGGTCCATGTCAAGCCTTGGTAAGGTTCTTC
GCGTTGCATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCC
CAGGCGGGGCACTTAATGCGTTAGCTACGGCGCGGAAAACGTGGAATGTCCCCCACACCTAGTGCCCAACGTTTACGGCATGGAC
TACCAGGGTATCTAATCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTTAATGCCCAGAGACCTGCCTTCGCCATCGGTGT
TCCTCCTGATATCTGCGCATTTCACCGCTACACCAGGAATTCCAGTCTCCCCTACATCACTCTAGTCTGCCCGTACCCACTGCAGA
AACCGGAGTTGAGCCCCGGTCTTTCACAGCAGACGCGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTT
GCGCCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTCCCTAACTG
AAAGAGGTTACAACCCGAAGGCCGTCATCCCTCACGCGGCGTCGCTGCATCAGGCTTTCGCCCATTGTGCATATTCCCCCTGCTGC

>KK5.2_F

TGCGGCGCTTACACATGCAAGTCGAACGATGATCCCAGCTTGCTGGGGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACC
TGCCCTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACTCCTCATCGCATGGTGGGGGGTGGAAAGCTTT
TTTGTGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAG
AGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAA

```
GCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTA
CCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGT
AAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGGATCTGCGGTGGGTACGGGCAGACTAGAG
TGATGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCT
GGGCATTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGC
ACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAA
AACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCT
TGACATGAACCGGAAATACCTGGAAACAGGTGCCCCGCTTGCGGTCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCG
TGAGATGGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCGCGTTATGGCGGGGACTCCTAGGAGACTG
CCGGGTCACTCCGGA

>KK5.2_R

ATGCGTCCACCTTTCGAACAGCTCCCTCCCACAAGGGGTTAGGCCACCGGCTTCGGGTGTTACCAACTTTCGTGACTTGACGGGCG
GTGTGTACAAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGC
AGACCCCAATCCGAACTGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCGCAACCCTTTGTACCGGCCATTGTAGCATG
CGTGAAGCCCAAGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTATGAGTCC
CCGCCATAACGCGCTGGCAACATAGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACA
ACCATGCACCACCTGTAAACCGACCGCAAGCGGGGCACCTGTTTCCAGGTATTTCCGGTTCATGTCAAGCCTTGGTAAGGTTCTTC
GCGTTGCATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCC
CAGGCGGGGCACTTAATGCGTTAGCTACGGCGCGGAAAACGTGGAATGTCCCCCACACCTAGTGCCCAACGTTTACGGCATGGAC
TACCAGGGTATCTAATCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTTAATGCCCAGAGACCTGCCTTCGCCATCGGTGT
TCCTCCTGATATCTGCGCATTTCACCGCTACACCAGGAATTCCAGTCTCCCCTACATCACTCTAGTCTGCCCGTACCCACCGCAGAT
CCGGAGTTGAGCCCCGGACTTTCACGGCAGACGCGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTTGC
GCCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTTCCTACTGAAA
GAGGTTACAACCCGAAGGCCGTCATCCCTCACGCGGCGTCGCTGCATCAGGCTTTCGCCCATTGTGCAA

>KK6.1_F

GCGCAGGCCTAACACATGCAAGTCGAGCGGTAGAGAGGTGCTTGCACCTCTTGAGAGCGGCGGACGGGTGAGTAATGCCTAGGA
ATCTGCCTGGTAGTGGGGGATAACGCTCGGAAACGGACGCTAATACCGCATACGTCCTACGGGAGAAAGCAGGGGACCTTCGGG
CCTTGCGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACTGGTCTGA
GAGGATGATCAGTCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGAA
AGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTTACCTAAT
ACGTATCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTT
AATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCGTTAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATT
CAAAACTGACGAGCTAGAGTATGGTAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCA
GTGGCGAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTC
CACGCCGTAAACGATGTCAACTAGCCGTTGGGAGCCTTGAGCTCTTAGTGGCGCAGCTAACGCATTAAGTTGACCGCCTGGGGAG
TACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGA
AGAACCTTACCAGGCCTTGACATCCAATGAACTTTCCAGAGATGGATTGGTGCCTTCGGGAACATTGAGACAGGTGCTGCATGGC
TGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTGTCCTTAGTTACCAGCACGTTAAGGTGGG
CACTCTAAGGAGACTGCCGTGACAACCGGAGGAAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGGCTGGGCTACCACG
TGCTACATGGTCGGG

>KK6.1_R

TGAACACCCGTGGGTAACCGTCCTCCCGAAGGTTAGACTAGCTACTTCTGGTGCAACCCACTCCCATGGTGTGACGGGCGGTGTG
TACAAGGCCCGGGAACGTATTCACCGCGACATTCTGATTCGCGATTACTAGCGATTCCGACTTCACGCAGTCGAGTTGCAGACTG
CGATCCGGACTACGATCGGTTTTCTGGGATTAGCTCCACCTCGCGGCTTGGCAACCCTCTGTACCGACCATTGTAGCACGTGTGTA
GCCCAGGCCGTAAGGGCCATGATGACTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCTCCTTAGAGTGCCCACCT
TAACGTGCTGGTAACTAAGGACAAGGGTTGCGCTCGTTACGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATG
CAGCACCTGTCTCAATGTTCCCGAAGGCACCAATCCATCTCTGGAAAGTTCATTGGATGTCAAGGCCTGGTAAGGTTCTTCGCGTT
GCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCGTACTCCCCAGG
CGGTCAACTTAATGCGTTAGCTGCGCCACTAAGAGCTCAAGGCTCCCAACGGCTAGTTGACATCGTTTACGGCGTGGACTACCAG
GGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCACCTCAGTGTCAGTATCAGTCCAGGTGGTCGCCTTCGCCACTGGTGTTCCTTC
CTATATCTACGCATTTCACCGCTACACAGGAAATTCCACCACCCTCTACCATACTCTAGCTCGTCAGTTTTGAATGCAGTTCCCAG
GTTGAGCCCGGGGATTTCACATCCAACTTAACGAACCACCTACGCGCGCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTC
TGTATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTATTCTGTCGGTAACGTCAAAACAGATACGTATTAGGTAACTGCCCT
TCCTCCCAACTTAAAGTGCTTTACAATCCGAAGACCTTCTTCACACACGCGGCATGGCTGGATCAGGCTTTCGCCCATTGTCCAAT
ATTCCCACTGCTGCCTCCCGAAG

>CMS3.1_R

GAACCTAACCGTGGGTAAGCGCCCCCCTTACGGTTAAGCTACCTACTTCTGGTAAAACCCGCTCCCATGGTGTGACGGGCGGTGT
GTACAAGACCCGGGAACGTATTCACCGCGACATGCTGATCCGCGATTACTAGCGATTCCAACTTCACGCAGTCGAGTTGCAGACT
```

GCGATCCGGACTACGATACACTTTCTGGGATTAGCTCCCCCTCGCGGGTTGGCGGCCCTCTGTATGTACCATTGTATGACGTGTGA
AGCCCTACCCATAAGGGCCATGAGGACTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCTCATTAGAGTGCTCAAC
TAAATGTAGCAACTAATGACAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCA
GCACCTGTGTTACGGTTCTCTTTCGAGCACACCTCGATCTCTCGAAGCTTCCGTACATGTCAAGGGTAGGTAAGGTTTTTCGCGTTG
CATCGAATTAATCCACATCATCCACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTTAATCTTGCGACCGTACTCCCCAGGCG
GTCTACTTCACGCGTTAGCTGCGTTACCAAGTTAATTAAAACCCGACAACTAGTAGACATCGTTTAGGGCGTGGACTACCAGGGT
ATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCATGAGCGTCAATCTTGACCCAGGGGGCTGCCTTCGCCATCGGTGTTCCTCCACA
TCTCTACGCATTTCACTGCTACACGTGGAATTCTACCCCCCTCTGCCAGATTCAAGCCTTGCAGTCTCCATCGCAATTCCCAGGTTG
AGCCCGGGGCTTTCACGACAGACTTACAAAACCGCCTGCGCACGCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTACGTA
TTACCGCGGCTGCTGGCACGTAGTTAGCCGGTGCTTATTCTTCAGGTACCGTCATTAGTAGAAGGTATTAACCTCCACCGTTTCTTC
CCTGACAAAAGAGCTTTACAACCCGAAGGCCTTCTTCACTCACGCGGCATTGCTGGATCAGGCTTGCGCCCATTGTCCAAAATTCC
CACTGCTGCCTCCCGA

>KK2.1_F

TTCTTGTGCTTTTGGATGAACTCACGTCCTATTAGCTTGTTGGTGAGGTAATGGCTCACCAAGGGCGACGATCCGTAACT

GGTCTGAAAA

>KK2.1_R

TCGACTCCTACTTCATGGGGTCGAGTTGCAGACCCCAATCCGAACTGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCG
CAACCCTTTGTACCGGCGATTGTAGCATGCGTGAAGCCCAAGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCG
AGTTGACCCCGGCAGTCTCCTATGAGTCCCCGCCATAACGCGCTGGCAACATAGAAGGAGGGTTGCGCTCGTTGCGGGACTTAAC
CCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTAAACCGACCGCAAGCGGGGCACCTGTTTCCAGGTATTTCC
GGTTCATGTCAAGCCTTGGTAAGGTTCTTCGCGTTGCATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTC

>KK3.1_F

CGGCTGCTTACACATGCAAGTCGAACGATGAAGCCCAGCTTGCTGGGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAAC
CTGCCCTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATAGGAGCGTCCACCGCATGGTGGGTGTTGGAAAGATT
TATCGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGA
GGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAG
CCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTACC
TGCAGAAGAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTATCCGGAATTATTGGGCGTAA
AGAGCTCGTAGGCGGTTTGTCGCGTCTGTCGTGAAAGTCCGGGGCTTAACCCCGGATCTGCGGTGGGTACGGGCAGACTAGAGTG
CAGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGG
GCTGTAACTGACGCTGAAGAGCGAAAGCATGGCGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTACACGTTGGGCACT
AGGTGTGGTGACCATTCCACGGTTTCCGCGCCGCAGCTAACGCATTAAGTGCCCCGCCTGTGGAGTACGGCCGCA

>KK3.1_R

TGGTCACCTTCGACGGCTCCCCCACAAGGGTTAGGCCACCGGCTTCGGGTGTTACCGACTTTCGTGACTTGACGGGCGGTGTGTAC
AAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAGACCCCA
ATCCGAACTGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCGCAACCCATTGTACCGGCCATTGTAGCATGCGTGAAGC
CCAAGACATAAGGGGCATGATGATTTGACGTCGTCCTCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCCATGAGTCCCCACCACG
ACGTGCTGGCAACATGGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCA
CCACCTGTGAACCCGCCCCAAAGGGGAAACCGTATCTCTACGGCGATCGAGAACATGTCAAGCCTTGGTAAGGTTCTTCGCGTTG
CATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAGGC
GGGGCACTTAATGCGTTAGCTGCGGCGCGGAAACCGTGGAATGGTCCCCACACCTAGTGCCCAACGTTTACGGCATGGACTACCA
GGGTATCTAATCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTTACAGCCCAGAGACCTGCCTTCGCCATCGGTGTTCCTC
CTGATATCTGCGCATTCCACCGCTACACCAGGAATTCCAGTCTCCCCTACTGCACTCTAGTCTGCCCGTACCCACCGCAGATCCGG
GGTTAAGCCCCGGACTTTCACGACAGACGCGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTCGCACCC
TACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGTGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTTCCTACTGAAAGAGGT
TACAACCCGAAGGCCGTCATCCCTCAAGCGGCGTCGCTGCATCAGGCTTGCGCCCATTGTGCA

>KK4.1_F

GCGCTTACACATGCAAGTCGAACGATGATCCCAGCTTGCTGGGGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCC
CTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACTCCTCATCGCATGGTGGGGGTGGAAAGCTTTTTTGT
GGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGT
GACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTG
ATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTACCTGC
AGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGA
GCTCGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGGATCTGCGGTGGGTACGGGCAGACTAGAGTGATG
TAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCA
TTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCACTAG

GTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTC
AAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACA
TGAACCGGAAATACCTGGAAACAGGTGCCCCGCTTGCGGTCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGA
TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCACGCGT

>KK4.1_R

ACCCGTCCCCTTCGAACAGCTCCCTCCCACAAGGGGTTAGGCCACCGGCTTCGGGTGTTACCAACTTTCGTGACTTGACGGGCGGT
GTGTACAAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAG
ACCCCAATCCGAACTGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCGCAACCCTTTGTACCGGCCATTGTAGCATGCG
TGAAGCCCAAGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTATGAGTCCCC
GCCATAACGCGCTGGCAACATAGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAA
CCATGCACCACCTGTAAACCGACCGCAAGCGGGGCACCTGTTTCCAGGTATTTCCGGTTCATGTCAAGCCTTGGTAAGGTTCTTCG
CGTTGCATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCC
AGGCGGGGCACTTAATGCGTTAGCTACGGCGCGGAAAACGTGGAATGTCCCCCACACCTAGTGCCCAACGTTTACGGCATGGACT
ACCAGGGTATCTAATCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTTAATGCCCAGAGACCTGCCTTCGCCATCGGTGTT
CCTCCTGATATCTGCGCATTTCACCGCTACACCAGGAATTCCAGTCTCCCCTACATCACTCTAGTCTGCCCGTACCCACCGCAGAT
CCGGAGTTGAGCCCCGGACTTTCACGGCAGACGCGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTTGC
GCCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTTCCTACTGAAA
GAGGTTACAACCCGAAGGCCGTCATCCCTCAAGCGGCGTCGCTGCATCAGGGTTTC

>KK4.3_F

TGCTTAACCATGCAAGTCGAACGATGAACCGGTGCTTGCACTGGGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGC
CCTTAACTCTGGGATAAGCCTTGGAAACGGGGTCTAATACTGGATATTGACTTTTCCTCGCATGGGGATTGGTTGAAAGATTTATT
GGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGT
GACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTG
ATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAACAAGGCCAGGCATTGTCTGGTTGA
GGGTACTTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGG
GCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACCCCGGATCTGCGGTGGGTACGGGCAGACT
AGAGTGATGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGT
CTCTGGGCATTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTG
GGCACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGC
TAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAG
GCTTGACATGAACTGGAAACGCTTGGAAACAAGTGCCCCGCTTGCGGTCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGT
CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCGCACGTAATGGTGGGGACTCTAA

>KK4.3_R

ATGCGTCCCCTTCGAAGCTCCCTCCCACAAGGGGGTTAGGCCACCGGCTTCGGGTGTTACCAACTTTCGTGACTTGACGGGCGGTG
TGTACAAGGCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAGA
CCCCAATCCGAACTGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCGCAACCCATTGTACCGGCCATTGTAGCATGCGT
GAAGCCCAAGACATAAGGGGCATGATGATTTGACGTCGTCCTCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTATGAGTCCCCA
CCATTACGTGCTGGCAACATAGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACC
ATGCACCACCTGTAAACCGACCGCAAGCGGGGCACTTGTTTCCAAGCGTTTCCAGTTCATGTCAAGCCTTGGTAAGGTTCTTCGCG
TTGCATCGAATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAG
GCGGGGCACTTAATGCGTTAGCTACGGCGCGGAAAACGTGGAATGTCCCCCACACCTAGTGCCCAACGTTTACGGCATGGACTAC
CAGGGTATCTAATCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTTAATGCCCAGAGACCTGCCTTCGCCATCGGTGTTCC
TCCTGATATCTGCGCATTTCACCGCTACACCAGGAATTCCAGTCTCCCCTACATCACTCTAGTCTGCCCGTACCCACCGCAGATCC
GGGGTTGAGCCCCGGACTTTCACGGCAGACGCGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTTGCGC
CCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAAGTACCCTCAACCAGACAATGCCTGGCCTTGTTC
CCTACTGAAAGAGGTT

>KK5.1_F

GGTGCTTACACATGCAAGTCGAACGATGATCCCAGCTTGCTGGGGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGC
CCTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACTCCTCATCGCATGGTGGGGGGTGGAAAGCTTTTTTG
TGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGG
TGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCT
GATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTACCTG
CAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAG
AGCTCGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGGATCTGCGGTGGGTACGGGCAGACTAGAGTGAT
GTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGC
ATTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCACTA
GGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACT
CAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGAC

ATGAACCGGAAATACCTGGAAACAGGTGCGCCGCTTGCGGTCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAG
ATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCGCGTTATGGCGGGGACTCCT

>KK5.1_R

TTCGGAAAGCTCCCTCCCACAAGGGGTTAGGCCACCGGCTTCGGGTGTTACCAACTTTCGTGACTTGACGGGCGGTGTGTACAAG
GCCCGGGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAGACCCCAATC
CGAACTGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCGCAACCCTTTGTACCGGCCATTGTAGCATGCGTGAAGCCCA
AGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTATGAGTCCCCGCCATAACG
CGCTGGCAACATAGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCA
CCTGTAAACCGACCGCAAGCGGGGCACCTGTTTCCAGGTATTTCCGGTTCATGTCAAGCCTTGGTAAGGTTCTTCGCGTTGCATCG
AATTAATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAGGCGGGGC
ACTTAATGCGTTAGCTACGGCGCGGAAAACGTGGAATGTCCCCCACACCTAGTGCCCAACGTTTACGGCATGGACTACCAGGGTA
TCTAATCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTTAATGCCCAGAGACCTGCCTTCGCCATCGGTGTTCCTCCTGAT
ATCTGCGCATTTCACCGCTACACCAGGAATTCCAGTCTCCCCTACATCACTCTAGTCTGCCCGTACCCACCGCAGATCCGGAGTTG
AGCCCCGGACTTTCACGGCAGACGCGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTTGCGCCCTACGT
ATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTTCCTACTGGAAAGAGGTTAC
AACCCGAAGGCCGTCATCCCTCACGCGGCGTCGCTGCATCAGGTTTT

>KK7.1_F

GTGCTTACACATGCAAGTCGAACGATGATGCCCACTTGTGGGTGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCC
CTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACTCCTCATCGCATGGTGGGGGGTGGAAAGCTTTATTG
TGGTTTTGGATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGG
TGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCT
GATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTACCTG
CAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAG
AGCTCGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGGATCTGCGGTGGGTACGGGCAGACTAGAGTGAT
GTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGC
ATTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCACTA
GGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACT
CAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGAC
ATGAACCGGAAATACCTGGAAACAGGTGCCCCGCTTGCGGTCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAG
ATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCGCGTTATGGCGGGGACTCTA

>KK7.1_R

AAAATGCCCTCCCACAAGGGGTTAGGCCACCGGCTTCGGGTGTTACCAACTTTCGTGACTTGACGGGCGGTGTGTACAAGGCCCG
GGAACGTATTCACCGCAGCGTTGCTGATCTGCGATTACTAGCGACTCCGACTTCATGGGGTCGAGTTGCAGACCCCAATCCGAAC
TGAGACCGGCTTTTTGGGATTAGCTCCACCTCACAGTATCGCAACCCTTTGTACCGGCCATTGTAGCATGCGTGAAGCCCAAGACA
TAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTATGAGTCCCCGCCATAACGCGCTG
GCAACATAGAACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGT
AAACCGACCGCAAGCGGGGCACCTGTTTCCAGGTATTTCCGGTTCATGTCAAGCCTTGGTAAGGTTCTTCGCGTTGCATCGAATTA
ATCCGCATGCTCCGCCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAGGCGGGGCACTTA
ATGCGTTAGCTACGGCGCGGAAAACGTGGAATGTCCCCCACACCTAGTGCCCAACGTTTACGGCATGGACTACCAGGGTATCTAA
TCCTGTTCGCTCCCCATGCTTTCGCTCCTCAGCGTCAGTTAATGCCCAGAGACCTGCCTTCGCCATCGGTGTTCCTCCTGATATCTG
CGCATTTCACCGCTACACCAGGAATTCCAGTCTCCCCTACATCACTCTAGTCTGCCCGTACCCACCGCAGATCCGGAGTTGAGCCC
CGGACTTTCACGGCAGACGCGACAAACCGCCTACGAGCTCTTTACGCCCAATAATTCCGGATAACGCTTGCGCCCTACGTATTACC
GCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTTCCCA

>KK9.2_F

TCTGCGCGGCCTAACCACATGCAAGTCGAGCGGTAGAGAGAAGCTTGCTTCTCTTGAGAGCGGCGGACGGGTGAGTAATGCCTAG
GAATCTGCCTGGTAGTGGGGGATAACGCTCGGAAACGGACGCTAATACCGCATACGTCCTACGGGAGAAAGCAGGGGACCTTCG
GGCCTTGCGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCGTAACTGGTCT
GAGAGGATGATCAGTCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCG
AAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAATTA
ATACTTTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCG
TTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCGTTAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCA
TTCAAAACTGTCGAGCTAGAGTATGGTAGAGGGTGGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACAC
CAGTGGCGAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAG
TCCACGCCGTAAACGATGTCAACTAGCCGTTGGGAGCCTTGAGCTCTTAGTGGCGCAGCTAACGCATTAAGTTGACCGCCTGGGG
AGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGC
GAAGAACCTTACCAGGCCTTGACATCCAATGAACTTTCAGAGAATGGATTGGTGCCTTCGGGAACATTGAGACAGGTGCTGCATG
GCTGTCGTCACCTCCTGTCCTGAGA

>KK9.2_R

ATCACCGTGGGTAACCGTCCTCCCGAAGGTTAGACTAGCTACTTCTGGTGCAACCCACTCCCATGGTGTGACGGGCGGTGTGTAC
AAGGCCCGGGAACGTATTCACCGCGACATTCTGATTCGCGATTACTAGCGATTCCGACTTCACGCAGTCGAGTTGCAGACTGCGA
TCCGGACTACGATCGGTTTTCTGGGATTAGCTCCACCTCGCGGCTTGGCAACCCTCTGTACCGACCATTGTAGCACGTGTGTAGCC
CAGGCCGTAAGGGCCATGATGACTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCTCCTTAGAGTGCCCACCATAA
CGTGCTGGTAACTAAGGACAAGGGTTGCGCTCGTTACGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCAG
CACCTGTCTCAATGTTCCCGAAGGCACCAATCCATCTCTGGAAAGTTCATTGGATGTCAAGGCCTGGTAAGGTTCTTCGCGTTGCT
TCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCGTACTCCCCAGGCGG
TCAACTTAATGCGTTAGCTGCGCCACTAAGAGCTCAAGGCTCCCAACGGCTAGTTGACATCGTTTACGGCGTGGACTACCAGGGT
ATCTAATCCTGTTTGCTCCCCACGCTTTCGCACCTCAGTGTCAGTATCAGTCCAGGTGGTCGCCTTCGCCACTGGTGTTCCTTCCTA
TATCTACGCATTTCACCGCTACACAGGAAATTCCACCACCCTCTACCATACTCTAGCTCGACAGTTTTGAATGCAGTTCCCAGGTT
GAGCCCGGGGATTTCACATCCAACTTAACGAACCACCTACGCGCGCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTCTGT
ATTACCGCGGCTGCTGGCACAGAGTTAGCCGGTGCTTATTCTGTCGGTAACGTCAAAACAGCAAAGTATTAATTTACTGCCCTTCC
TCCCAACTTAAAGTGCTTTACAATCCGAAGACTTCTTCCA

>CF4.7_F

CCTTAACCATGCAAGTCGAACGATGAAGCCCTTCGGGGTGGATTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTC
ACTCTGGGACAAGCCCTGGAAACGGGGTCTAATACCGGATAACACTCTGTCCCGCATGGGACGGGGTTAAAAGCTCCGGCGGTG
AAGGATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGAC
CGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATG
CAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACCTGCAGA
AGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCT
CGTAGGCGGCTTGTCACGTCGGATGTGAAAGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAG
GGGAGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATCTCTGGGCCATT
ACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTG
TTGGCGACATTCCACGTCGTCGGTGCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAA
GGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGCTTAATTCGACGCAACGCGAAGAACCTTACCAAGGCTTGACATAT
ACCGGAAAGCATCAGAGATGGTGCCCCCCTTGTGGTCGGTATACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTT
GGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTCTGTGTTGCCACATGCCCTTCGGGGTGATGGGGACTCCC

>CF4.7_R

AATCGACAGCTCCCTCCCACAAGGGGTTGGGCCACCGGCTTCGGGTGTTACCGACTTTCGTGACGTGACGGGCGGTGTGTACAAG
GCCCGGGAACGTATTCACCGCAGCAATGCTGATCTGCGATTACTAGCAACTCCGACTTCATGGGGTCGAGTTGCAGACCCCAATC
CGAACTGAGACCGGCTTTTTGAGATTCGCTCCGCCTCGCGGCATCGCAGCTCATTGTACCGGCCATTGTAGCACGTGTGCAGCCCA
AGACATAAGGGGCATGATGACTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTGTGAGTCCCCATCACCCCG
AAGGGCATGCTGGCAACACAGAACAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCC
ATGCACCACCTGTATACCGACCACAAGGGGGGCACCATCTCTGATGCTTTCCGGTATATGTCAAGCCTTGGTAAGGTTCTTCGCGT
TGCGTCGAATTAAGCCACATGCTCCGCTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAG
GCGGGGAACTTAATGCGTTAGCTGCGGCACCGACGACGTGGAATGTCGCCAACACCTAGTTCCCAACGTTTACGGCGTGGACTAC
CAGGGTATCTAATCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAATGGCCCAGAGATCCGCCTTCGCCACCGGTGTTC
CTCCTGATATCTGCGCATTTCACCGCTACACCAGGAATTCCGATCTCCCCTACCACACTCTAGCTAGCCCGTATCGAATGCAGACC
CGGGGTTAAGCCCCGGGCTTTCACATCCGACGTGACAAGCCGCCTACGAGCTCTTTACGCCCAATAATTCCGGACAACGCTTGCG
CCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTCGCTTCTTCCCTGCTGAAAG
AGGTTACAACCCGAAGG

>CF4.7B_F

AATCCAAGCCCTTCGGGGTGGGATTAGTGGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTG
GAAACGGGGTCTAATACCGGATAACACTCTGTCCCGCATGGGACGGGGTTAAAAGCTCCGGCGGTGAAGGATGAGCCCGCGGCC
TATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACAGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGA
GACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGG
GATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACCTGCAGAAGAAGCGCCGGCTAACTA
CGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGT
CGGATGTGAAAGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAGGGGAGATCGGAATTCCTGG
TGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAA
AGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTTGGCGACATTCCACGTCGT
CGGTGCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGC
ACAAGCAGCGGAGCATGTGGCTTAATTCGACGCAACGCGAAGAACCTTACCAAGGCTTGACATATACCGGAAAGCATCAGAGAT
GGTGCCCCCCTTGTGGTCGGTATACAGGTGGTGCATGGCTGTCGTCACCT

>CF7.7B_R

TACTTTCGACAGCTCCCTCCCACAAGGGGTTGGGCCACCGGCTTCGGGTGTTACCGACTTTCGTGACGTGACGGGCGGTGTGTACA
AGGCCCGGGAACGTATTCACCGCAGCAATGCTGATCTGCGATTACTAGCAACTCCGACTTCATGGGGTCGAGTTGCAGACCCCAA
TCCGAACTGAGACCGGCTTTTTGAGATTCGCTCCGCCTCGCGGCATCGCAGCTCATTGTACCGGCCATTGTAGCACGTGTGCAGCC

```
CAAGACATAAGGGGCATGATGACTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTGTGAGTCCCCATCACCC
CGAAGGGCATGCTGGCAACACAGAACAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAG
CCATGCACCACCTGTATACCGACCACAAGGGGGGCACCATCTCTGATGCTTTCCGGTATATGTCAAGCCTTGGTAAGGTTCTTCGC
GTTGCGTCGAATTAAGCCACATGCTCCGCTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCA
GGCGGGGAACTTAATGCGTTAGCTGCGGCACCGACGACGTGGAATGTCGCCAACACCTAGTTCCCAACGTTTACGGCGTGGACTA
CCAGGGTATCTAATCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAATGGCCCAGAGATCCGCCTTCGCCACCGGTGTT
CCTCCTGATATCTGCGCATTTCACCGCTACACCAGGAATTCCGATCTCCCCTACCACACTCTAGCTAGCCCGTATCGAATGCAGAC
CCGGGGTTAAGCCCCGGGCTTTCACATCCGACGTGACAAGCCGCCTACGAGCTCTTTACGCCCAATAATTCCGGACAACGCTTGC
GCCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGCGCTTCTTCTGCAGGTACCGTCACTTTTGCTTCTTC

>CF4.8_F

AAGGATGCAAGTCGAACGATGAAGCCCTTCGGGGTGGATTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTC
TGGGACAAGCCCTGGAAACGGGGTCTAATACCGGATAACACTCTGTCCTGCATGGGACGGGGTTAAAAGCTCCGGCGGTGAAGG
ATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGCGACCGGC
CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGC
GACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGCAAGTGACGGTACCTGCAGAAGAA
GCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAG
GCGGCTTGTCACGTCGGATGTGAAAGCTCGGGGCTTAACCCCGAGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAGGGGAG
ATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATCTCTGGGCCATTACTGAC
GCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTTGGCG
ACATTCCACGTCGTCGGTGCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATT
GACGGGGGCCCGCACAAGCAGCGGAGCATGTGGCTTAATTCGACGCAACGCGAAGAACCTTACCAAGGCTTGACATATACCGGA
AAGCATCAGAGATGGTGCCCCCCTTGTGGTCGGTATACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGAT

>CF4.8_R

TGATGGTCCTCCCCGTAAGGGGTTGGGCCACCGGCTTCGGGTGTTACCGACTTTCGTGACGTGACGGGCGGTGTGTACAAGGCCC
GGGAACGTATTCACCGCAGCAATGCTGATCTGCGATTACTAGCAACTCCGACTTCATGGGGTCGAGTTGCAGACCCCAATCCGAA
CTGAGACCGGCTTTTTGAGATTCGCTCCGCCTCGCGGCATCGCAGCTCATTGTACCGGCCATTGTAGCACGTGTGCAGCCCAAGAC
ATAAGGGGCATGATGACTTGACGTCGTCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTGTGAGTCCCCATCACCCCGAAGG
GCATGCTGGCAACACAGAACAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGC
ACCACCTGTATACCGACCACAAGGGGGGCACCATCTCTGATGCTTTCCGGTATATGTCAAGCCTTGGTAAGGTTCTTCGCGTTGCG
TCGAATTAAGCCACATGCTCCGCTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTCCCCAGGCGG
GGAACTTAATGCGTTAGCTGCGGCACCGACGACGTGGAATGTCGCCAACACCTAGTTCCCAACGTTTACGGCGTGGACTACCAGG
GTATCTAATCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAATGGCCCAGAGATCCGCCTTCGCCACCGGTGTTCCTCCT
GATATCTGCGCATTTCACCGCTACACCAGGAATTCCGATCTCCCCTACCACACTCTAGCTAGCCCGTATCGAATGCAGACTCGGGG
TTAAGCCCCGAGCTTTCACATCCGACGTGACAAGCCGCCTACGAGCTCTTTACGCCCAATAATTCCGGACAACGCTCGCACCCTAC
GTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGTGCTTCTTCTGCAGGTACCGTCACTTGCGCTTCTTCCCTGCTGAAAAAAGGT
TTACAACCCGAAGGGC
```

Endnote: Noteworthy Organizations, Technologies, and Databases

[1] Blaauw, S. (2021, December 8). Center for Extraterrestrisk Liv (CELS). Retrieved May 6, 2022, from https://cels.nbi.ku.dk/english

[2] U.S. National Library of Medicine. (n.d.). National Center for Biotechnology Information. Retrieved May 6, 2022, from https://www.ncbi.nlm.nih.gov/

[3] Illumina. (2015). *Hiseq 2500 Sequencing System - Illumina, Inc.*. HiSeq® 2500 Sequencing System. Retrieved May 6, 2022, from https://www.illumina.com/Documents/products/datasheets/datasheet_hiseq2500.pdf

[4] Illumina. (2022). *Sequencing platforms*. Compare NGS platform applications & specifications. Retrieved May 6, 2022, from https://www.illumina.com/systems/sequencing-platforms.html

[5] Illumina. (2017). *An introduction to Next-generation sequencing technology*. Retrieved May 6, 2022, from https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf

[6] Illumina. (2019). *Illumina adapter sequences (1000000002694) - clark science center*. Retrieved May 29, 2022, from https://www.science.smith.edu/cmbs/wp-content/uploads/sites/36/2020/01/illumina-adapter-sequences-1000000002694-11.pdf