dbcAmplicons: A modular, highly multiplexed design for Illumina amplicon sequencing

Matthew L. Settles, Dan New, Sam Hunter, Larry J. Forney, Alida T. Gerritsen

Abstract

Background

DNA sequencing technology has allowed for a surge in new research in high throughput community ecology of bacterial, fungal, and other community types. In particular, both the Roche 454 platform and Illumina platforms have enabled such projects as the Human Microbiome Project (HMP), the International Census of Marine Microbes (ICOMM), and the Earth Microbiome Project (EMP). These projects seek to characterize the microbial and taxonomic diversity found at multiple body sites and geographical sites around the world. Similar projects in fungal communities have also been started showing the richness in fungal communities [fungal biogeography] around the world. [Bik] show similar projects in microscopic eukaryotes is both possible and needed, and the advent of ‘DNA metabarcoding’ to assay presence of DNA from higher order taxa in environmental DNA (eDNA) [Taberlet]. With the Illumina platform continuing to increase in both throughput and read length, the ability to simultaneously assay multiple communites across hundreds (or even thousands) of samples is now possible. We introduce dbcAmplicons a DNA to data workflow for assaying the community structure using a module highly multiplexed scheme for Illumina sequencing.

Polymerase chain reaction (PCR) amplicon sequencing is a common and important tool used to query genetic variation and structure in ecological communities. Applications range from determining the taxon community structure in microbial, fungal and other community types to determining mutation frequencies in a set of genes across many individuals. Common practice is to add a barcoded DNA sequencing adapter to the template specific primer, generating a PCR product that can be sequenced on an Illumina machine. As sequencing throughput continuing to increase the need for flexible methods that can maximize the sequencing effort are needed. For amplicons this implies sequencing more individual and amplicons in the same sequencing run.

An amplicon is an amplified molecule, usually generated via PCR, of a single type and is an exact replicate of the original DNA template. A pair of PCR primers is designed to uniquely target a particular region of DNA. In order to sequence a DNA fragment using the Illumina platform, certain sequences are necessary in order for the fragment to bind to the Illumina flowcell, amplify and then initiate sequencing. Typically, paired target specific PCR primers are designed to include the extra oligonucleotides necessary for sequencing (~60bp to each primer). PCR amplicon sequencing in this manner then requires the researcher to purchase a unique pair of ~80bp primer for every target region and sample (including barcode) in the experiment. A technique that is neither modular nor cost effective.

The increase in sequencing density offers an opportunity to sequence many loci across hundreds or even thousands of samples at significant depth of coverage. New techniques are needed however to both multiplex amplicons and samples in the same sequencing reaction.

**Other Names**

**Tag-encoded amplicons**

**DNA barcoding**

Methods

**Overview of analysis**

This protocol is written for users who wish to analyze sequence data with the *dbcAmplicons* analysis pipeline. Raw sequence files from the MiSeq will need to be configured by the miseq\_pipeline.py script to fit the expectations of *dbcAmplicons* before the analysis can begin. The sample sheet from the run is reconfigured to support the output of four individual reads (Read 1, Index 1, Index 2, Read 2) and CASAVA (the Illumina-specific post-run processing software) is then run using the reconfigured sample sheet. The output from this step is four individual reads ready for analysis using *dbcAmplicons*.

**Configuring raw MiSeq data for input into the pipeline**

The following items are necessary for the proper execution of the pipeline:

* Python version 2.X (this protocol written for 2.7.5)
* A properly formatted sample sheet named SampleSheet.csv and the RunInfo.xml sheet
* Raw data generated from sequencing, before CASAVA is run
* Access and read permissions to the raw data
* Installed and executable CASAVA (this protocol written for 1.8)

Navigate to the desired directory and execute the miseq\_pipeline.py script, giving the first parameter as the raw data from the run and the second parameter as the name of the desired output directory. The sample sheet from the run must be named SampleSheet.csv and it must be in the run output folder, along with the RunInfo.xml sheet.

An example command (written in the shell terminal) run in the desired directory is as follows:

**python miseq\_pipeline.py 150325\_M01380\_0037\_000000000-AEMV5 Output\_directory**

where 'python' is the call to execute the script, '150325\_M01380\_0037\_000000000-AEMV5' is the name of the directory that holds the MiSeq run information, and 'Output\_directory' is the prefix for the desired output. The script will generate and run CASAVA within the output directory until completion. Once the process is complete, the reads are ready for analysis.

**Analyzing dual-barcoded amplicon sequence data using *dbcAmplicons***

There are 4 levels of analysis within the *dbcAmplicons* package. The package was developed for command line usage to analyze high volumes of microbial 16S, LSU, and ITS sequences. All components of the package are designed to flow directly into the next with no intermediate formatting steps required. For a detailed vignette using sample data provided with the *dbcAmplicons* package, please refer to the github page within the man folder.

*Installation*

Minimum requirements for the *dbcAmplicons* package: Python 2.X is required to run any and all of the

elements in the package. Some applications within the package rely on **flash2** or the **RDP classifier** to complete, so the user must download and install the external tools.

All executable scripts are located in the github repository <https://github.com/msettles/dbcAmplicons.git>

The **flash2** executable and installation instructions can be found at <https://github.com/dstreett/FLASH2.git>. The **RDP** **classifier** may be found at <http://sourceforge.net/projects/rdp-classifier/>. The user must install the classifier.jar and its dependencies in the RDPtools package.

*Mac OSX*

Mac users should first download and install **Xcode** and then download the app via a terminal by:

git clone https://github.com/msettles/dbcAmplicons.git

cd dbcAmplicons

sudo python setup.py install

dbcAmplicons

If installation was successful, this will bring up the usage arguments.

*Linux*

For Linux users, some python dependencies may be required before the installation is successful. Root privileges are also required to install these dependencies. To install, open a terminal and type:

sudo apt-get install python-dev

And type the sudo password at the prompt. Then, download the app via a terminal by:

git clone https://github.com/msettles/dbcAmplicons.git

cd dbcAmplicons

sudo python setup.py install

dbcAmplicons

If installation was successful, this will bring up the usage arguments.

***dbcAmplicons******preprocess***

*dbcAmplicons* *preprocess* takes four fastq files (Read 1, Index 1, Read 2, and Index 2), the assigned barcode indices for each sample, the sequences of the template-specific primers, and the individual sample names along with an assigned project folder. *dbcAmplicons preprocess* will then assign each read in the fastq file to a specific project according to both barcode and primer sequences, and return 2 fastq files with all of the identifying information in the read header. If a project directory is specified, *dbcAmplicons preprocess* will create a directory with the indicated output prefix.

At a minimum, *dbcAmplicons* *preprocess* requires a properly formatted reference barcode sequence file to demultiplex raw sequences. For maximum separation and identification of the reads, the following three inputs should be used, all in a tab-separated text format:

1. Reference barcode sequences file.

2. Reference primer sequences file.

3. Sample metadata file with sample IDs, primer IDs, and barcode IDs.

Examples of each of these sheets may be found in Appendix A. *dbcAmplicons preprocess* will automatically run a validation of the inputs before any reads are separated. The barcode and primer sheets will be checked against the sample metadata file to make sure that any assigned barcodes and primers exist within the reference files.

The syntax for usage is as follows:

dbcAmplicons preprocess [-h] -B barcodesFile [-d BARCODEDIFF]

[-P primerFile] [-D PRIMERDIFF] [-e PRIMEREND]

[-S sampleFile] [-q MINQ] [-l MINL]

[-b BATCHSIZE] [-O PREFIX] [-U] [-u] [-v]

[-1 read1 [read1 ...]] [-2 read2 [read2 ...]]

[-3 read3 [read3 ...]][-4 read4 [read4 ...]] [--test]

[--keepPrimers] [--debug]

The user will give the paths to each of the input files. Default settings for primer and barcode distances may be changed, but it is not recommended to set them to be less stringent unless the user has good reason to do so. The final outputs of *preprocess* include a text file listing the barcodes that were identified and two fastq files for each project that was specified.

***dbcAmplicons join***

The next step in analysis is to overlap the reads to create longer sequences. *dbcAmplicons join* is the step that requires the **flash2** executable in the user path for overlapping. The process will check for the executable and if it is not found the user will be prompted to download and install. When the reads are overlapped *dbcAmplicons join* will automatically discard reads that have extremely low quality scores, which indicates the presence of primer dimers that are not biologically relevant. Usage is as follows:

dbcAmplicons join [-h] [-O PREFIX] [-u] [-x NUM] [-t NTHREADS]

[-v] [-1 read1] [-2 read2]

The minimum required inputs are the paths to the preprocessed read 1 and read 2 that the user wants to overlap. Default settings are recommended. There may be cases where the overlapping criteria are too stringent, especially if there are large amounts of errors on the ends that prevent overlaps. The user must determine which settings would be appropriate in that case.

The outputs from the *join* process include 3 fastq files, one of the overlapped sequences (named extendedFrags.fastq) and two of the unpaired sequences (named notCombined\_1/2). The process will also generate histograms showing the size and frequency of the overlapped sequences. If the user finds large quantities of unexpected sequence lengths there may be contamination issues within the sample, and classifying the reads will likely give an erroneous result. However, if the amplicons are within the expected size range the user can proceed to the next step in the pipeline.

***dbcAmplicons classify***

Once the reads have been overlapped, they can be classified by 16S, ITS, or fungal LSU sequence using the RDP classifier. The *classify* process will check for the RDP classifier similarly to how the *join* process will check for the flash2 executable. Organisms are classified to their closest taxonomic rank that meets a 50% bootstrap match. Usage is as follows:

dbcAmplicons classify [-h] [-b BATCHSIZE] [-q MINQ] [-l MINL]

[-p PROCS] --rdpPath PATH [-g <arg>] [-t TRAIN]

[-O outputPrefix] [-1 read1 [read1 ...]]

[-2 read2 [read2 ...]] [-U single [single ...]][-v]

[--debug]

The output from the *classify* process is a single fixrank file that can then be processed with *dbcAmplicons abundance.*

***dbcAmplicons abundance***

Within the *abundance* process, the fixrank file is reformatted to produce three outputs that parse the classified data according to user-specified criteria. Usage is as follows:

dbcAmplicons abundance [-h] [-r <arg>] [-t VALUE] [-m VALUE]

[-M VALUE][-S FILE] [-O FILE\_PREFIX] -F FILE [FILE ...][-b][-v] [--debug]

The user will determine the lowest classification levels that are determined, the largest and smallest amplicon sizes to be considered within the classification, and the threshold bootstrap value for a classification cutoff. These values are dependent on the quality and diversity of the samples as well as the user's own criteria. The minimum input would be the fixrank file if the user does not want to change any of the default settings.

The outputs include the following three text files. The abundance.txt file lists the taxon abun-

dance with the read counts. The proportions.txt lists the proportions of reads per sample, and the

taxa counts.txt lists a summary of the lowest level classification and the number of counts for that level.

Dual PCR

Results and Discussion

Appendix

Designing primers:

Excel sheet of primers

**Overview of full protocol**

This protocol describes the method for dual-barcoding a tageted amplicon via a two-step PCR process and then sequencing on the Illumina MiSeq. The first PCR step isolates the target-specific region whereas the second PCR step adds on Illumina sequencing adapters and indexing barcodes. Both PCR primers of the first PCR include universal complementary sequence tags (CS1 and CS2) that amplify the product in the second PCR step. These allow for maximum flexibility in target specific primer usage and the ability to separately barcode or include multiple targets in the same sequencing reaction without needing to purchase a large number of barcoded primers. Barcodes are included in both adapters, giving 24x24 adapter combinations to uniquely identify 576 samples.

The protocol assumes that starting DNA samples are clean and of relatively high quality, with no amount of additional contaminants or salts.

**Ordering FL1 and FL2 sequencing primers for the Illumina MiSeq**

The protocol for sequencing dual-barcoded amplicons is dependent upon the FL1 and FL2 primers as designed and explained in the Fluidigm Access ArrayTM System for Illumina Sequencing Systems User Guide (pp.126-136). The following details the steps for ordering and preparing the primers up to the sequencing step.

|  |  |  |
| --- | --- | --- |
| **Name** | **Oligo Name** | **Sequence (5'-3')** |
| FL1 | CS1 | A+CA+CTG+ACGACATGGTTCTACA |
| CS2 | T+AC+GGT+AGCAGAGACTTGGTCT |
| FL2 | CS1rc | T+GT+AG+AACCATGTCGTCAGTGT |
| CS2rc | A+GAC+CA+AGTCTCTGCTACCGTA |

The “+” indicates the following nuleotide is LNA.

Go to Exiqon’s website order page for Custom LNATM Oligos (<http://www.exiqon.com/order-lna-oligos>).

Under **Enter an oligonucleotide sequence below**, copy Oligo Name “CS1” from Table 1 for Oligo name.

* Select **DNA oligo** from Oligo type drop-down menu.
* Select **100 nmole** from Synthesis scale drop-down menu.
* Select **HPLC** **Purification** from Purification drop-down menu.
* Copy the sequence for the oligo CS1 from the above table for Sequence (5'-3').
* Repeat steps as above to order CS2, CS1rc, and CS2rc oligos.
* Complete checkout process.

**Preparation of FL1 and FL2 primer mixes for sequencing**

1. Spin tubes down before opening.

2. Calculate the resuspension volume required for each oligo using the following formula:

(X nmol oligo)\* 10 = μL volume needed to resuspend oligo.

3. Add correct volume of low EDTA TE (10 mM Tris pH 8. 0.1mM EDTA) buffer calculated from step 2 to each primer tube such that the final concentration of each primer is 100 μM.

4. Vortex and spin down briefly all components after resuspension in low EDTA TE buffer.

5. Prepare FL1 primer mix. Mix CS1 oligo and CS2 oligo as described below to a final concentration of 50 μM for each oligo. Vortex after mixing to ensure complete mixing. Spin down.

|  |  |  |
| --- | --- | --- |
| **Oligo Name** | **Volume** | **Final Concentration** |
| CS1 (100 μM) | 30 μL | 50 μM |
| CS2 (100 μM) | 30 μL | 50 μM |
| Total | 60 μL |  |

6. Prepare FL2 primer mix. Mix CS1rc oligo and CS2rc oligo as described below to a final concentration of 50 μM for each oligo. Vortex after mixing to ensure complete mixing. Spin down.

|  |  |  |
| --- | --- | --- |
| **Oligo Name** | **Volume** | **Final Concentration** |
| CS1rc (100 μM) | 30 μL | 50 μM |
| CS2rc (100 μM) | 30 μL | 50 μM |
| Total | 60 μL |  |

Store FL1 and FL2 stocks at -20C.

**Amplification Protocol**

Materials required:

- Fragment Analyzer 12- capillary array (Advanced Analytical Technologies)

- High sensitivity next-generation DNA Kit (Advanced Analytical Technologies DNF-474)

- PCR grade water, nuclease-free water

- Qubit 2.0 dsDNA Broad Range Assay kit (Life Technologies Q32850)

- Qubit 2.0 Fluorometer (Life Technologies)

- 96-well plate reader (Molecular Devices SPECTRA max GEMINI XPS)

- 96-well plates and adhesive sealing for 96-well plates

- Thermocycler (e.g. Applied Biosystems 4314445 and N8050200)

- PCR reagent kit (New England Biolabs E5000S)

- Magnetic beads (Ampure XP, MagBio)

- Magnetic stand or plate (Life Technologies 12027)

**PCR 1**

1. Quantify template DNA, preferably using fluorometry.

2. Use 100ng of DNA for the PCR 1 reaction, adding more template when necessary. Avoid excessive cycling, which may bias the downstream products with a false signal from PCR duplicates.

3. Prepare a reaction mix for PCR 1 using the following components:

|  |  |
| --- | --- |
| **Reagent** | **Volume** |
| 10X PCR Buffer | 5.0 |
| 25mM MgCl2 | 6.0 |
| 10mM dNTP mix | 1.0 |
| 10uM TS-CS1 forward primer | 0.25 |
| 10uM TS-CS2 forward primer | 0.25 |
| *Taq* DNA polymerase 5000U/mL | 0.25 |
| PCR-grade water | 36.25 |
| Template DNA | 1.0 |

Total reaction volume is 50 uL. Cycle the reaction with the following parameters:

Denaturation: 95C - 2 minutes

Cycle 20 times: 95C – 1 minute

51C – 1 minute

68C – 1 minute

Extention: 68C – 10 minutes

Hold at 4C until ready to proceed.

4. Run 2uL of the product through the Fragment Analyzer to check amplification of the target region. It is not necessary to check each product if the number of samples is too large, but a representative sample should be taken to ensure the amplicon amplified properly without excessive primer-dimers or non-specific amplification. Band sizes should incorporate the length of the CS-tags.

– Stopping point: PCR 1 products may be stored for up to 3 days at 4C, or at -20C for longer-term storage.

**PCR 2**

5. Prepare a 15-fold dilution of the PCR product from PCR 1 as follows: add 2µl of product from first PCR to 28µl of PCR grade water. Vortex well to mix, then centrifuge briefly for 30 seconds to spin down all components. The 1:15 dilute PCR 1 product is now ready to be used as template for PCR 2 below.

6. Prepare the reaction mix for PCR 2 using the following components:

|  |  |
| --- | --- |
| **Reagent** | **Volume** |
| 10X PCR Buffer | 2.0 |
| 25mM MgCl2 | 3.6 |
| 10mM dNTP mix | 0.40 |
| 2uM Barcoded Illumina primer | 0.75 |
| *Taq* DNA polymerase 5000U/mL | 0.20 |
| PCR-grade water | 36.25 |
| Template DNA (PCR 1 product) | 1.0 |

Total reaction volume is 20uL. Cycle the reaction with the following parameters:

Denaturation: 95C – 1 minute

Cycle 10 times: 95C – 30 seconds

60C – 30 seconds

68C – 1 minute

Extention: 68C – 5 minutes

Hold at 4C until ready to proceed.

– Stopping point: PCR 2 products may be stored for up to 3 days at 4C, or at -20C for longer.

**QC, pooling, and cleaning**

1. Quantify all PCR 2 products using fluorometry. Check the size and amplification of a representative sample of products by running on the Fragment Analyzer. If the amplicon is of the expected size (~69bp increased from PCR 1 products if barcoded adapters were successfully added) the sample is ready to be cleaned, pooled, and prepared for sequencing. If necessary, adjust DNA amounts in the previous PCR to optimize DNA yield.

2. Pool amplicon samples together based on DNA mass from fluorometry. Accurate quantification is crucial in this step to correctly measure the amount of double-stranded DNA in each sample without bias from unincorporated adapters. If the samples span a large range of mass, a qPCR quantification step of similarly massed pools may be necessary to determine absolute library masses and to normalize before all samples can be pooled.

3. Clean the amplicon pool(s) of small fragments with magnetic bead-based chemistry using a 0.8X volume of beads to the total volume of the pool (e.g. a 100uL pool will require 80uL of beads for cleaning). Confirm cleaning by running samples on the Fragment Analyzer and determine the average fragment size within the targeted peak.

4. Quantify DNA pool(s) with the KAPA Library Quantification DNA Kit for Illumina (KAPA Biosystems KK4824) according to instructions from the KAPA Library Quantification Technical Guide (v1.14). Serial dilutions are recommended as per protocol specifications.

5. Calculate library concentrations of each pool and normalize all libraries to 10nM. Pooling will be dependent on the concentration and volume of the smallest sample, so this protocol recommends normalizing using the following procedure:

- Transfer 10uL from each pool that was quantified using qPCR into a clean 1.5mL microcentrifuge tube. Calculate the nanomolar concentration of each pool using the masses quantified from the qPCR and the average fragment size from the Fragment Analyzer, and adjust volumes to obtain 10nM.

- If **dilution** is required:

* Determine the appropriate amount of buffer EBT to add for 10nM.
* Vortex 2X and centrifuge tube. Keep on ice

- If **concentration** is required:

* Determine the total volume of sample that is required for 10nM.
* Prepare a vacuum centrifuge to dry samples at 30C.
* Pipette measure sample frequently to check proper volume. Add additional EBT if necessary

6. Pool all normalized libraries together to make a working stock of 10nM. Pool according to the lowest sample volume and desired proportion of reads to come from each sample (e.g. if the smallest sample volume is 5uL and equimolar pooling is desired, pool 5uL from each sample to make the working stock).

Repeat KAPA qPCR quantification as above on the pooled samples.

**Preparing for MiSeq sequencing**

Please refer to the Illumina MiSeqTM User Guide for proper maintenance and operation of the MiSeq.

**\*Note: This protocol recommends at least 15% of the sequencing run to consist of diverse library. Amplicon sequences are not as diverse as general shotgun libraries, and too many clusters with the same sequence may cause the machine to have increased error rates and may even cause a run failure.**

Before beginning:

* Complete all necessary wash steps for the Miseq
* Prepare a sample sheet for the run
* Set a heat block at 96C for the last step heat shock
* Remove the frozen tray component of the 600 cycle MiSeq kit from -20C and allow to thaw at room temperature. Protect from light until ready to sequence.
* Thaw HT1 on ice and keep chilled for the duration of the setup.
* Prepare a fresh dilution of 0.1N NaOH
* Thaw FL1 and FL2 sequencing primers on ice.

1. Combine the following volumes of DNA and freshly diluted 0.1N NaOH in a 1.5μL tube:

* 5 μL 10 nM pooled sample DNA
* 5 μL 0.1N NaOH

Vortex briefly to mix and spin down.

2. Incubate library for 5 minutes at room temperature to denature DNA.

3. Add 990 μL of pre-chilled HT1 to the tube. This results in a 50pM denatured library in 1 mM NaOH. Vortex to mix and keep tube on ice.

4. Determine appropriate library concentration for optimal cluster generation using the below table as a guide. Low to mid-range clustering within the abilities of the MiSeq kit is recommended for low diversity dual-barcoded amplicons so that the sequencer can differentiate individual clusters. 18-20pM concentration for V3 chemistry is the most commonly used library concentration with the technique outlined in this protocol. Add appropriate amounts of denatured library and pre-chilled HT1 in a separate 1.5mL tube.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Final library concentration (pM) | 15pM | 18pM | 20pM | 25pM |
| Volume of 50 pM denatured DNA | 180 | 216 | 240 | 300 |
| **Volume of pre-chilled HT1** | 420 | 384 | 360 | 300 |

5. Vortex the mixture and spin down. Place on ice until ready for the final heat shock and loading. The final input volume is 600μL. If a PhiX control is added it must be in the same final concentration as the DNA and must be added within the final volume of 600μL for loading.

6. Prepare the Read 1 sequencing primer FL1:

* Ensure the reagent tray has fully thawed.
* Dilute 7 uL of sequencing reagent FL1 to a final concentration of 500 nM with 693 uL HT1 buffer. Vortex to mix and spin down.
* Pierce the foil seal covering reservoir 18 with a long pipette tip and load 680uL of diluted FL1 for Read 1.

7. Prepare the Read 2 sequencing primer FL2:

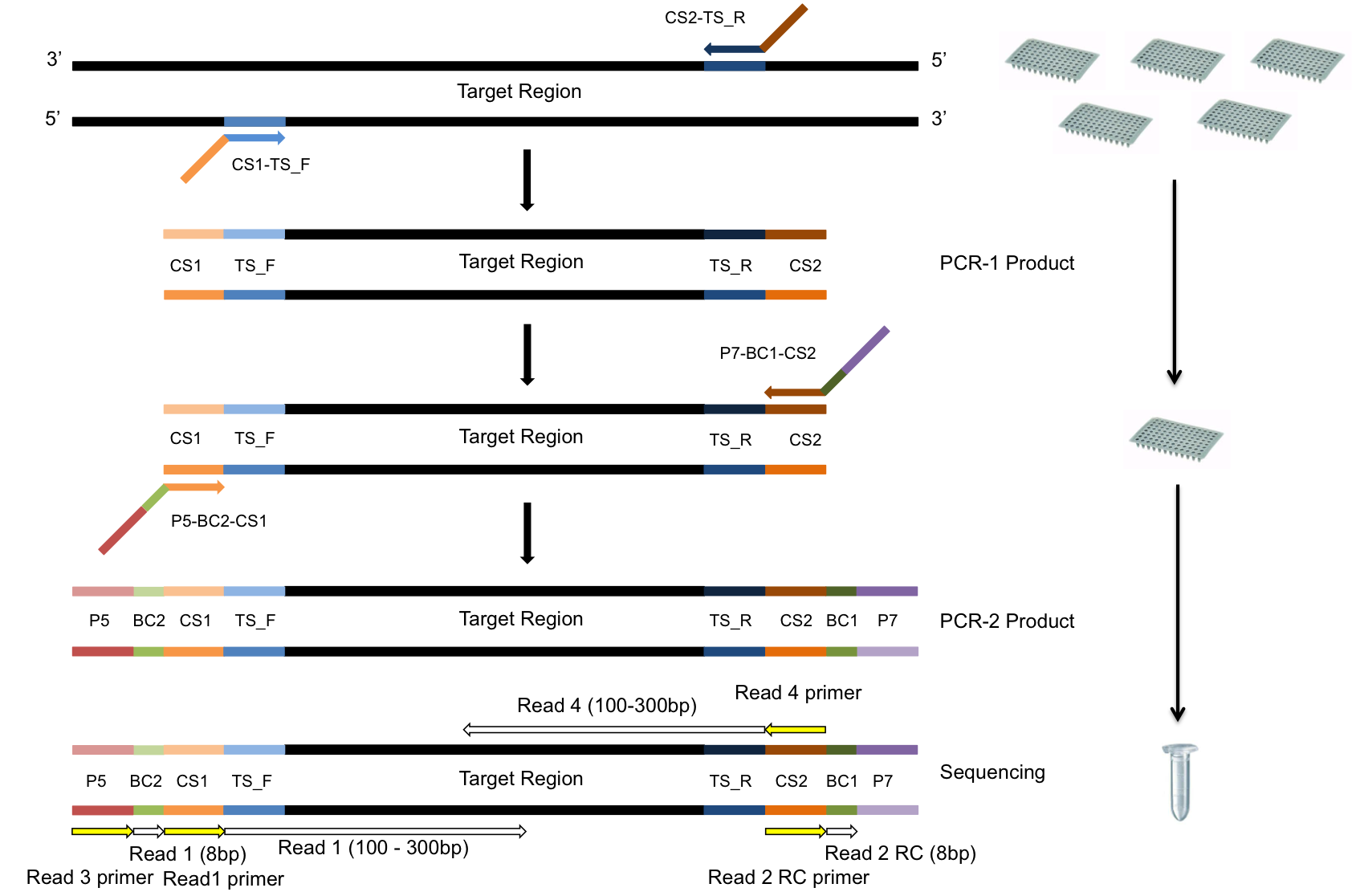
* Dilute 7 uL of sequencing reagent FL2 to a final concentration of 500 nM with 693 uL HT1 buffer. Vortex to mix and spin down.
* Pierce the foil seal covering reservoir 19 with a long pipette tip and load 680uL of diluted FL2 for Read 2.

8. Prepare the Read 3 sequencing primer FL1:

* Dilute 7 uL of sequencing reagent FL1 to a final concentration of 500 nM with 693 uL HT1 buffer. Vortex to mix and spin down.
* Pierce the foil seal covering reservoir 20 with a long pipette tip and load 680uL of diluted FL1 for Read 3.

9. Place the tube containing 600uL of denatured DNA in a 96C heat block for 2 minutes. Remove immediately and chill on ice until ready to load. Pierce foil covering sample reservoir with a long pipette tip and load 600uL into the well. Load reagent tray into MiSeq.

10. Load all necessary reagents into the MiSeq, including flow cell. Start the run and monitor on BaseSpace until it is completed.



|  |  |
| --- | --- |
| **FIRST PCR PRIMERS** | |
| CS1-TF\_F | 5’ – **ACACTGACGACATGGTTCTACA***[NNNNN]*TEMPLATE-SPECIFIC-SEQUENCE - 3’ |
| CS2-TF\_R | 5’ – **TACGGTAGCAGAGACTTGGTCT***[NNNNN]*TEMPLATE-SPECIFIC-SEQUENCE - 3’ |
|  |  |
| **SECOND PCR PRIMERS** | |
| P5-BC2-CS1 | 5’ - AATGATACGGCGACCACCGAGATCTACAC*NNNNNNNN***ACACTGACGACATGGTTCTACA**- 3’ |
| P7-BC1-CS2 | 5’ - CAAGCAGAAGACGGCATACGAGAT*NNNNNNNN***TACGGTAGCAGAGACTTGGTCT**- 3’ |
|  |  |
| **SEQUENCING PRIMERS** | |
| Read 1 Primer | 5’ - ACACTGACGACATGGTTCTACA- 3’ |
| Read 2 Primer | 5’ - AGACCAAGTCTCTGCTACCGTA- 3’ |
| Read 3 Primer | Uses the P5 amplification primer |
| Read 4 Primer | 5’ - TACGGTAGCAGAGACTTGGTCT- 3’ |