5. 16S Illumina Library Preparation

Dave Baker

5.1. Purpose

Historically in the early days of DNA sequencing technologies, sequencing whole genomes was expensive and impractical. Also, to enable determination of bacterial species of unknown origin, or where a mixture of different species is anticipated, a targeted PCR (Polymerase Chain Reaction) is needed. The approach also needs to use a region of the genome that is common or conserved in most if not all species. The region must also contain enough diversity to distinguish one species or genus from another. With the development of NGS (Next Generation Sequencing) in the late 2000s, and superseding single-read Sanger sequencing, came Illumina technology which can generate millions of different reads from a single sample. Also, over time more individual species have been isolated, cultured and sequenced individually, improving the size and accuracy of reference databases. Here we describe how to take metagenomic DNA (a single sample although a mixture of species) and process it into a viable Illumina library for sequencing. In principle, this method can be applied to any conserved target gene e.g. yeast ITS, animal and fish Co1, or plant RbcL. Here we present the targeting of the prokaryotic 16S ribosomal RNA gene (16S rRNA). This target can be present in multiple copies across the genome and is another consideration when trying to quantitatively determine the proportion of different species present.

5.2. Introduction

As discussed above metagenomic studies commonly analyse the prokaryotic 16S ribosomal RNA gene (16S rRNA), which is approximately 1,500 bp long and contains nine variable regions surrounded by more highly conserved regions. Primers that anneal to the more conserved regions (enabling amplification by PCR) are used to amplify the variable regions of 16S rRNA (between a forward and reverse primer) enabling identification of genera or species within microbial populations by comparison with well characterised databases. At the Quadram Institute we currently use primer sets for the V1-2, V3-4 and V4 variable

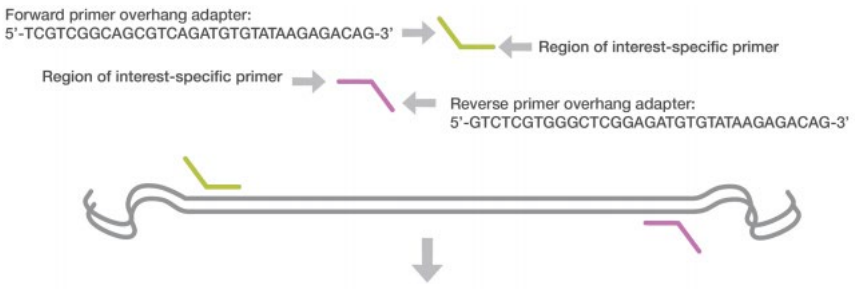
regions of the 16S rRNA gene [1]. The general principle of amplification by PCR is shown in Fig. 1; also shown is secondary PCR which enables the addition of short sequences (i.e. indexes or barcodes) that enable different samples to be separated, bioinformatically, after a sequencing run.

5.2.1. Concepts

- ***PCR (Polymerase Chain Reaction)*** – A method to enable a targeted approach in looking at any genome. Small synthetically synthesised single-stranded DNA bases are used to specifically anneal to the complementary sequence within a genome. In successive cycles of heating and cooling, a brand new, synthetic amplicon or PCR product is formed. By using two primers on opposite strands a defined region can be amplified exponentially.

- ***NGS (Next Generation Sequencing)*** – Original Sanger sequencing returned a single read and up to 96 of them. NGS is able to sequence millions of different DNA molecules simultaneously. Illumina NGS technology does this on a flowcell…

- ***Flowcell*** – imagine a flowcell to be a channel for fluid to pass through. That channel has a top and bottom surface onto which very small single-stranded DNA is bound. These allow the DNA suspension to be pumped through at a relatively high temperature and cooled to hybridise to the 'primer lawn' in the channel. In a process called Bridge Amplification, each DNA molecule forms a group of identical molecules in a very small space - this is called a cluster. Each cluster represents a single read.



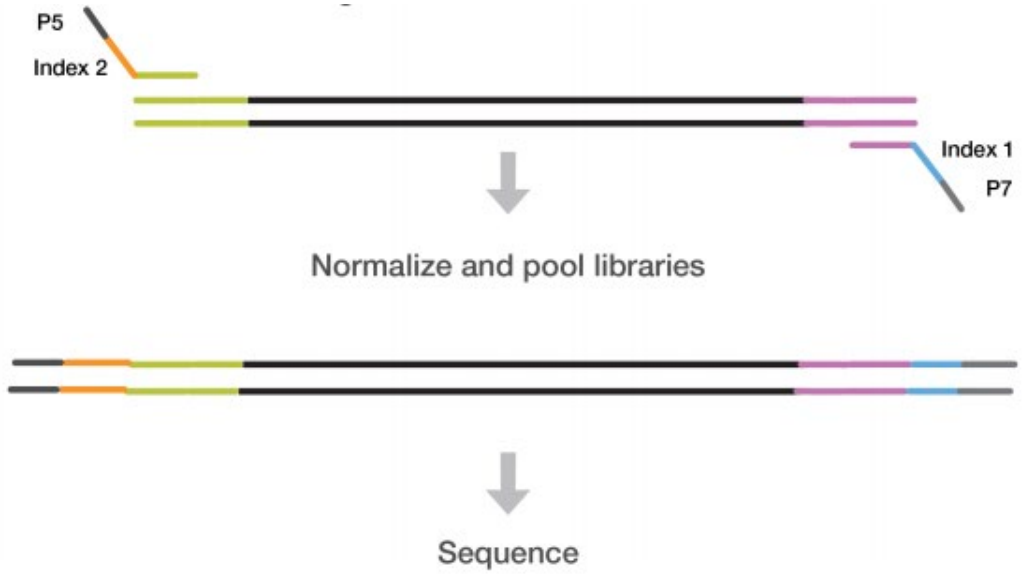


Figure 1: Schematic of two step PCR for amplification and barcoding of samples

To enable the PCR products to be manipulated and attached to an Illumina flowcell, specific sequences (known as 'P5' and 'P7') must be added to the tails of the index barcoding primers. The Illumina overhang adapter sequences to be added to locus-specific sequences in the first PCR are: Forward overhang: 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[locus- specific sequence] Reverse overhang: 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[locus- specific sequence]. As mentioned previously this method can be applied to any target you wish.

5.3. Materials (including products, kits, solutions etc.)

5.3.1. Equipment

- Semi-skirted 96-well plate
- Single channel pipettes 0.1-2.5 µl, 1-10 µl, 5-20 µl, 20-200 µl, 100-1000 µl
- Multi-channel pipette 2-10 µl
- PCR machine

5.3.2. Reagents

- 10mM Tris-HCl buffer (EB)
- 80% ethanol (Freshly made on the day)
- Kapa 2G PCR kit
- Ampure beads (Kapa or Beckman)
- Gene specific primer working stock at 10uM
- Multiple barcoding (UDI's) primers at 10uM (usually in a 96 well plate)
- RNAse/DNAse free dH_2O

5.4. Methods (step by step procedure)

- Gene Specific PCR (PCR 1)
- PCR bead clean
- Barcoding PCR (PCR 2)
- Quantify Library and equimolar pool
- Single and final bead clean-up
- Final Molarity Quantification

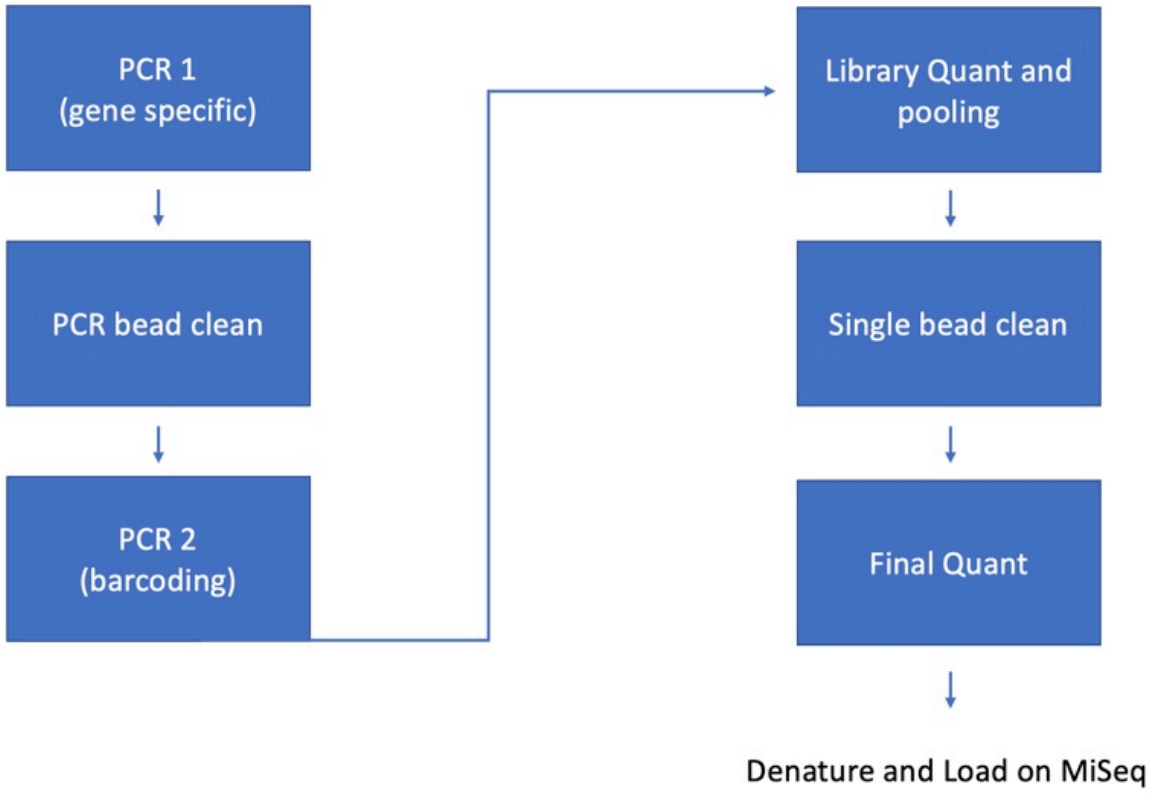


Figure 2: Workflow for generating an Illumina Library

5.4.1. Dilute DNA to approximately 5ng/ul

1. For optimum reproducibility across samples in an experiment we recommend normalising the DNA to the same concentration. **Note**: for samples with a low biomass and/or a large quantity of non-bacterial host DNA present, this should be increased to increase the number of viable targets.
2. Use Qubit or a plate reader to quantify DNA and take 5ul and add to the amount of water n-5 (where n is the concentration in ng/ul).
3. Seal and gently vortex the diluted DNA plate and pulse centrifuge for 5-10 seconds to bring the liquid to the bottom of the plate wells.

5.4.2. PCR 1 (Gene-specific PCR)

1. Before beginning, place the PCR enzyme on ice (unnecessary with 'Hot Start' polymerase); everything else can defrost at room temperature.
2. Generate a 'master mix' of PCR reagents and primers according to the following recipe, ensuring there is enough to fill all the sample wells (plus extra for pipetting error and dead volume - usually up to 10%):
    - 4µl 2G GC buffer
    - 0.4µl DNTPs
    - 0.08µl KAPPA2G polymerase
    - 11.52µl nuclease free dH_2O
    - 2ul @ 10uM primer mix
3. Add 18µl of master mix to each reaction well in a PCR plate. You can do this with stepper or multichannel pipette from a trough
4. Add 2µl of metagenomic DNA from a different sample into an individual well in the PCR plate (negative controls [of your choice] recommended)
5. Seal the reaction plate with a PCR-grade plate seal; be sure each well is carefully sealed between adjacent wells
6. Lightly vortex, and spin down gently
7. Run the first gene specific PCR (PCR 1) plate on the thermocycler with 95ºC for 3 minutes, 30 cycles of 95ºC for 10s, 55ºC for 20s and 72ºC for 3 minutes.

5.4.3. Clean up following gene specific PCR (PCR 1)

Note: Although it is recommended to clean up PCR 1 by removing PCR 1 gene-specific primers before moving to PCR 2, you do have the option to go straight into PCR 2 without clean-up. You could just add the barcoding primers (see 5.4.4 below) to PCR-1 (and reduce primer concentration in PCR 1 to exhaust the primers) to the first PCR plate. The downside of this is that it could increase variation in read count from sample to sample; overestimate the final library concentration; and underload the sequencing run.

1. Before beginning, remove beads from the fridge, thoroughly vortex, and allow to come to room temperature
2. Using a sterile trough and multichannel pipette, transfer 14ul of beads (0.7X the volume of PCR 1) to each well of the PCR 1 plate (**Note**: with the typical slight evaporation that occurs, this will be more like a 0.8X but that is not an issue; primers will still be removed in this range).

3. Seal (carefully) and lightly place on vortex for one or two seconds and just enough to homogenise (if you use a pipette to mix the sample there is no need to vortex) and pulse spin down. Leave at room temperature for 10 minutes off the magnet.
4. Add to 96-well magnetic block that contains magnets to enable the magnetic beads to accumulate on one side and wait 5 minutes or until the solution looks clear.
5. Remove supernatant from each well using a multichannel pipette set to just above the expected volume (14+20 = 34ul).
6. Do 2 X 80% ethanol washes using a convenient volume (circa 100ul) and a P100/200 pipette
7. After the second wash use a P10 pipette to remove the final drops of ethanol from each well.
8. Air dry for 1 minute and then resuspend the pellet in each well in 20ul of water or EB (10mM Tris-HCl). EB is more appropriate for long-term storage.
9. At QI we leave the beads in the original PCR plate but feel free to transfer to a new clean plate. Incubate at room temperature for 2 minutes prior to freezing the PCR 1 plate (with or without the beads present) at -20°C. This is now a safe stopping point.

5.4.4. Barcoding (PCR 2)

Alongside best practice as described here, using barcoding (unique dual indexing; UDI) will ensure that libraries sequence and demultiplex with the highest accuracy across all Illumina sequencing platforms. When preparing libraries for multiplexing, Illumina encourages customers to use unique dual indexing (UDI) whenever possible to ensure the most accurate demultiplexing. This will mean every sample will have a unique barcode at both ends. This also mitigates against 'index hopping' which is more prevalent on patterned flowcells (such as the Novaseq 6000) to which the DNA is attached for amplification. For longer reads MiSeq PE300 and PE250, and Novaseq 6000 Sp PE250 flowcells are the only options and cannot be barcoded.

1. Before beginning, collect the UDI primers from the freezer and allow them to thaw during master mix preparation (below). Use a different indexed primer plate for every 96 samples.
2. Generate a master mix of buffer, DNTPs and polymerase according to the following recipe ensuring there is enough to fill all sample wells (plus extra for pipetting error up to 10%):
    - 4µl 2G GC buffer
    - 0.4µl DNTPs
    - 0.08µl KAPPA2G polymerase

- 5.52µl nuclease free dH_2O

3. Add 5µl of UDI primers to each well in the new barcoding plate (PCR 2)
4. Add 5µl of clean material from each PCR 1 well to the corresponding well in the new, PCR 2, plate

Note: For KAPA2G Fast HotStart Readymix master mix, add 10µl to each well + 3µl of UDI primers per well (instead of 8µl of master mix + 5µl of indexed primers).

5. Add a PCR-grade plate seal to the PCR 2 reaction plate
6. Lightly vortex, and gently centrifuge
   Run the barcoding plate (PCR 2) on the thermocycler at 72ºC for 3 minutes, 95ºC for 1 minute, ten cycles of 95ºC for 10s, 55ºC for 20s and 72ºC for 3 minutes.

5.4.5. Quality Control of Multiplex Barcoding

1. Select a couple of samples per batch in addition to the blanks from each PCR 2 plate and run these on the 'Tapestation' as described above with a **D5000 tape** with enough spare gel lanes
2. Label the wells on the 'Tapestation' software to ensure ease of analysis

5.4.6. Pooling of Barcoded Multiplexes

1. Quantify your libraries using Qubit or a plate reader. Good libraries are usually between 10-50ng/ul; if yours are less then inspect them on 'Tapestation'.
2. In Excel calculate the number of ul needed for pooling to ensure that you have enough to pool the same quantity for each library, i.e. 50ng each. For example, if sample 1 is 15 ng/ul, then pool 50/15 = 3.33ul of this library and so on. Pool into a single tube/Eppendorf.
3. Wash the pooled sample with purification beads (Single SPRI 0.7X) as follows:
   - Add 100µl of pooled sample to a new Eppendorf tube (you can clean the entire volume, but it will probably be too high a concentration for loading).
   - Add 70µl of fresh beads (this is your 0.7X clean up).
   - Vortex to mix thoroughly
   - Incubate at room temperature for 10 minutes on a Hula mixer. In the absence of a Hula mixer, flick the tube at intervals to mix (e.g. after every minute of incubation)

- At the end of the 10 min incubation, place the tube on a magnetic tube rack to sediment the beads. Wait for the beads to sediment well (this can take several minutes).
- Remove the supernatant
- Double wash with freshly prepared 80% ethanol. Wait at least 30s after adding ethanol before removing the supernatant.
- Centrifuge to dry
- Remove excess ethanol with a P10 pipette set to 10ul.
- Add 30 µl EB to the sediment, close the lid and mix thoroughly by vortexing
- Incubate at room temperature for 5 minutes on a Hula mixer as above
- Place the tube back on a magnetic tube rack to sediment the purification beads
- Transfer the supernatant to a new Eppendorf tube, being careful not to transfer the beads (it doesn't matter about transferring the entire volume here as its more important to not transfer the beads)

Figure 3: What the beads look like when it is ready for you to remove the supernatant

- Clearly label the transferred eluate as 'clean library pool' and date appropriately. This is a safe stopping point, and the pool can be stored at -20 °C

5.4.7. Quantification of the Illumina Library

1. Quantify the pooled library using Qubit, **this time with 195µl of buffer and 5µl of the pooled library**
2. Note the concentration on the pooled library tube
3. Dilute the pooled library appropriately (**using EB**) to bring the concentration of pooled library down to ~2ng/µl AROUND 10nM) and repeat Qubit (**again with 195µl of buffer and 5µl of the diluted library**)
4. Once this concentration is achieved, run the diluted pooled sample on 'Tapestation' using a D5000 HS reagent and screen tape as follows:

5. On the Tapestation, note the concentration of the final library (this should be approximately the same as the concentration derived by Qubit), the insert size and the molarity. The insert size will vary depending on which region is amplified.
6. Calculate the molarity using the **V3 molarity calculator spreadsheet** as follows:
7. In column C (insert size), enter the insert size as estimated by 'Tapestation'
8. Enter the Qubit concentration in Column D (Qubit Stock Concentration)
9. This should display the Qubit molarity in Column E
10. Note the Qubit molarity for the next steps below

Note: ***If running locally at the Quadram Institute then follow the next steps otherwise ship to whoever is doing the sequencing***

5.4.8. Library denaturation and final dilutions

1. Determine the volume of diluted library necessary to take forward for denaturation and final library dilutions as follows.
2. Divide 20 by the molarity in nM, this should give the volume of library needed for the denaturation step to achieve 20 pM
3. To this volume, add 0.5µl freshly prepared 2N NaOH and make up the volume to 10µl with EB (check that this volume is 10ul in case an error has been made).
4. Incubate this denaturation mix at room temperature for 5 min, then add 990µl of thawed and chilled HT1. (HT1 is a buffer supplied with every MiSeq sequencing kit)
5. This brings the library to a concentration of 20 pM

Example:

For a library of Qubit concentration 2.03ng/µl and insert size of 465 bp, the molarity calculator yields a molarity of 6.73 nM. Treat this as 7 nM

20/7 yields 2.86, thus, to 2.86µl of library add 0.5µl 2N NaOH. This makes 3.36µl; thus 10-3.36 = 6.64µl of EB needs to be added to make up the volume to 10µl. To this, add 990µl HT1

Note:

- Make sure you 'eyeball' the 0.5ul in the tip and measure the final volume to make sure it is 10ul in total. If not, start again.
- Keep fresh aliquots of the 2NNaOH

As a result of the way the instrument determines the positions of the clusters in the first four cycles of the run, it is necessary for you to have at least 20% of non-amplicon diverse library as a spike-in when you are amplifying the same amplicon in all samples. This can be your

own shotgun library or the PhiX control sold by Illumina. Illumina recommends using ≥ 20% PhiX control spike-in.

5.4.9. Final dilutions

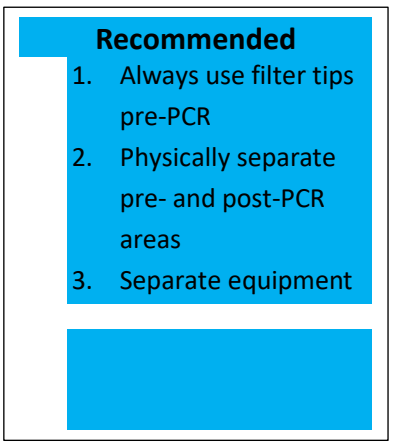
For a **MiSeq**, load the final library at **12pM** as follows:

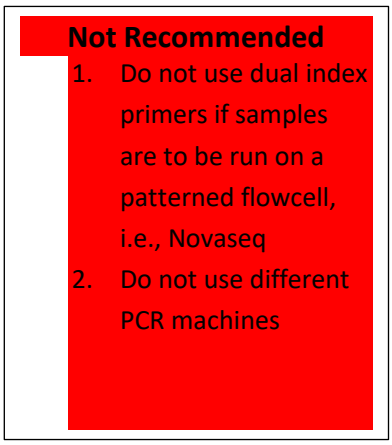
1. To a fresh 1.5ml tube add 300µl of library (20pM), 60 µl PhiX (20pM) + 240 µl HT1
2. Load into the correct well in the cartridge. Pierce first with a different P1000 tip.

Note: As described above for 20% phiX or any WGS (Whole Genome Sequencing) library. However, if you are running multiple amplicons you can reduce this down to 10%.

5.4.10. Loading the run on the sequencer

1. Always power cycle the sequencer before each run (this option can be found in 'Manage Instruments' on the sequencer)
2. Check you have enough local memory (usually at least 200GB is required)
3. After the power cycle, choose 'Sequence'
4. Create a sample sheet if you haven't already done so. At the Quadram Institute we save Binary Base Call (BCL's) files locally, and base call using local high-performance computing (HPC) after the run has finished. However, the MiSeq is currently able to BCL-FASTQ and also run a metagenomic analysis after the run using a modified 16S Greengenes database. Also, using the current MiSeq Reporter analysis on the instrument software you can actually put in your own custom databases including ITS/Co1/RbCL etc.
5. Follow the on-screen instructions; clean and load the flowcell and load the loaded cartridge. (You also have the option to use Basespace and monitor from your phone or laptop; the data will also be uploaded to Basespace as well as existing locally on the MiSeq.





5.5. Tips and Troubleshooting

Physically separate pre- and post-PCR areas. Also have separate equipment for working with pre- and post-PCR samples. At the Quadram Institute we have a sequencing instrument room which is a separate 'dirty' room and the only place where post-PCR libraries are handled, pooled and loaded. This reduces the risk of contamination in the pre-PCR set-up area. With the two step PCR it may require three areas: pre-PCR 1, pre-PCR 2 and post PCR. This will minimise risk of cross contamination between steps.

5.5.1. Avoiding cross-contamination

- When adding or transferring samples or reagent master mixes, change tips between each sample.
- Be extra careful during PCR 1 clean-up as this is a high-risk point for cross contamination.
- When adding index adapters with a multichannel pipette, change the tips between each row or each column. If using a single channel pipette, change tips between each sample.

5.5.2. Sealing the Plate

Always seal the 96-well plate with the adhesive seal using a rubber roller or pen to cover the plate and seal between the wells before the following steps in the protocol:

- Shaking steps (especially at the bead clean-up step)
- Thermal cycling steps
- Centrifuge steps

References

1. https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf