Bacteria Genome Assembly

This workflow has a number of components to it, they are as follows

---

High MW gDNA extraction

This protocol uses the Qiagen G100 genomic tips to purify high MW DNA. I have found this to produce fragments > 60kb consistently.

It might be worth trying the G20 tips in the future. Should still be able to purify enough gDNA.

I recommend ordering the Genomic DNA Buffer Set, but the buffers could be made manually if need be.

I designed racks that are 3D printable and hold the Qiagen tips. They make the process much easier, especially when you have several samples.

Make sure that you have RNAaseA, lysozyme, and proteinase K before starting the purification. These do not come in any of the kits and must be ordered seperately.

Always use wide bore pipette tips when working with high MW DNA or it will shear into smaller fragments. Can either order these (prefered) or use a razor blade to cut off end of tip (pain in the ass)

I have personally tested the protocol using the following strains. Other strains will certainly work with it (perhaps not all), but I have not tested them myself yet. I think getting enough biomass is paramount (I have not been standardizing input across strains as suggested, I think it takes way too much time to do CFU to OD600 for each strain of interest.

Strain	input	Note
Streptococuss pnemonaie	cell pellet from 6mL turbid culture	
Bacteroides vulgatus	cell pellet from 6mL turbid culture	got out less concentrated DNA than from Streptococcus.

The protocol these instructions are adapted from can be found here: https://www.qiagen.com/us/resources/resourcedetail?id=d2b85b26-16dd-4259-a3a7-a08cbd2a08a3&lang=en

1. Pellet turbid culture via centrifugation (3000-5000 x g for 5 min.) 6mL of turbid culture seems to work pretty well but it will vary somewhat by strain and density.
   a. Cell pellets can then be frozen at -80C until they are ready to be used.
2. To process 1x sample, add 7 µl, of RNase A solution (100 mg/ml Qiagen Cat. No. / ID: **19101**) to a 3.5 ml aliquot of Buffer B1. For additional N samples, multiply by N.
3. Dissolve lysozyme in distilled water to a concentration of 100 mg/ml. We will only need 80uL for 1X sample. I just make and use a fresh stock each time.
4. Resuspend each pellet in 3.5mL of the B1 + RNAase A solution that you made in step 2. (I don't worry about the loss due to dead volume in the last sample). Make sure this is as homogenous as possible by a combination of vortexing and pipetting.
5. Add 80uL of lysozyme and 100uL of proteinase K (Qiagen Cat. No. / ID: **19131**) to each sample, and mix by inverting.
6. Incubate at 37C for at least 30 min.
   a. I am not usually very careful about timing this. Anecdotally, I think it might be possible to do this for too long. (I left it for 3 hrs once and ended up with very viscous lysate that didn't go through the column well and ultimately did not end up with a high DNA yield).
   b. The solution should become clear after this step (cells should lyse). If not clear, would be best to increase the length of incubation as needed.
7. Add 1.2 mL of buffer B2 to each sample, mix by inverting.
8. Incubate at 50C for 30 min
9. While the incubation is occurring, equilibrate a 100-G genomic tip by adding 4mL of buffer QBT

10. Vortex the sample (from step 8) for 10 s at maximum speed and apply it to the equilibrated QIAGEN Genomic-tip. Allow it to enter the resin by gravity flow.
11. Wash the tip with 2x 7.5 mL of buffer QC
12. Elute the gDNA with 5mL of QF buffer (preheat it at 50C to increase yield)
13. Precipitate the DNA by adding 3.5 ml (0.7 volumes) of room-temperature isopropanol to the eluted DNA.
    a. mix by inverting, at this point you should see long strands of DNA start to precipitate and stick to each other.
14. Recover the precipitated DNA by using a heat-bent glass Pasteur pipette to physically spool the DNA and transfer it to a microcentrifuge tube containing 0.1 - 0.2 ml of 10 mM Tris·Cl, pH 8.5 (I just use Qiagen EB buffer). Dissolve the DNA overnight on a shaker. Store at 4C.
15. Sometimes it can be hard to see the precipitated DNA if the yield is not fantastic. When this occurs you can use this alternative method:
    a. centrifuge immediately at >5000 x g (we don't have a centrifuge that can go that high so I just do max speed) for at least 15 min at 4°C. Carefully remove the supernatant.
    b. Wash the centrifuged DNA pellet with 2 ml of cold, freshly made, 70% ethanol. Vortex briefly and centrifuge at >5000 x g for 10 min at 4°C. Carefully remove the supernatant without disturbing the pellet. Air-dry for 5 10 min, and resuspend DNA in 0.1 - 0.2 ml of 10 mM Tris·Cl, pH 8.5 (I just use Qiagen EB buffer). Dissolve the DNA overnight on a shaker. Store at 4C.

At this point, you probably have a good bit of high mW DNA, time to prove exactly what you have!

DNA Quantification

Measure DNA purity via nanodrop

- Don't trust the concentration, it is not accurate enough for a good quant. This is just for general QC
- Nanopore recommends only using chemically pure samples that have 260/280 nm ratios between 1.8-2 and 260/230 nm ratios between 2.0-2.2.

Measure DNA Concentration via Qubit BR

1. Prepare the appropriate amount of working solution by diluting qubit reagent 1:200 in the BR buffer.
    a. 200 x (n of samples to quant + 2)
2. Add the appropriate amount of sample or standard to qubit assay tubes according to the table below (I usually only use 1uL of sample to be quantified unless it ends up being too low)

Component	1X standards	1X sample to be quantified
Working solution volume	190	180-199
Volume of standard to add	10	
Volume of user sample		1-20
Total volume	200	200

3. Add the appropriate amount of working solution according to the table above to your qubit assay tube.

4. Run on qubit and record concentration.

DNA Size Measurement

We can use the Tapestation at IGM to measure the size distribution of our DNA samples for ~$8 a sample. This needs to be done directly after DNA extraction, and once more after the final sequencing library has been prepared.

> If you have 16 or less samples, submit them in eppi lobind tubes.
>
> If you have more than 16 samples, submit them in Thermo-Fast 96 PCR Detection Plates CAT# AB1400L.

1. Fill out the sample manifest that can be found here, taking care to follow the instructions.
2. Aliquot so that there is 4 uL of your sample between 10 - 100ng/uL in concentration.
3. Email completed manifest to igmsamplesubmission@ucsd.edu and Kristen Jepsen @ kjepsen@ucsd.edu. Also, print it out to add to sample box.
4. Bring samples to IGM during the designated drop off hours
   Monday, Wednesday and Friday: 10:00 am to 12:00 pm
   Tuesday and Thursday: 2:00 pm to 4:00 pm.
5. You can expect the data to come back pretty quickly, previously they delivered the data overnight.

Library Preparation

There are many possible ways you could prepare DNA libraries, but so far I have used the Native Barcoding Kit (24).

Be sure to use a library prep chemistry that is compatible with the flow cell you want to use.

Also make sure that you have enough of the NEB enzymes required on hand! They do not come in the kit but are absolutely necessary for the kit to work.

- NEB Blunt/TA Ligase Master Mix (NEB, Cat # M0367)
- NEBNext FFPE Repair Mix (NEB, M6630)
- NEBNext Ultra II End repair/dA-tailing Module (NEB, E7546)
- NEBNext Quick Ligation Module (NEB, E6056)

Nanopore has a great protocol that they have already made, so I will not bother retyping it. Here it is attached. I would recommend printing this off and following it each time.

Stop following the protocol once you get to the section labeled Priming and loading the SpotON flow cell.

At this point, quantify the library again via qubit and send for Tapestation size measurement according to the sections above. The size distribution that you get is necessary for determining how much DNA is needed to achieve the number of fmols to load.

Use this attached excel template to calculate how much of the library to load on the cell.

Loading the Flow Cell

The following applies to the parts of the protocol after you reach the Priming and loading the SpotON flow cell section

The BSA is def not necessary (as the protocol indicates), I loaded a library without it and it was fine. It probably makes it better, but if you happen to not have any on hand don't worry too much about it.

It is normal for some bubbles to form as the run progresses (initally I thought this meant the flow cell was destroyed but that ended up not being the case at all - they are not close enough to the membrane to kill the pores.)

Starting the Run in Minknow

see https://nanoporetech.com/accuracy for up to date statistics about different options and their effects on accuracy

1. Open the Minknow software, make sure the account is logged in
2. Select the options that you want to use to
    a. I haven't been able to get the SUP model to run on the cp in real-time, just use the HAC model to do base calling. It is almost as good and much less intensive. If we really needed higher accuracy, we could always run standalone guppy after the files have finished.
    b. So far I have only used the 260 bps option, but I think the 420 bps is supposed to work almost just as well (and produce more data).

Once, real time base calling wasn't working and it ended up being fixed by doing a complete reinstall of minknow. When this wasn't working, we ended up with a bunch of .raw files that I had to manually convert to .fast5 using the recover_reads utility. See this link -https://community.nanoporetech.com/support/articles/134 - for instructions.

Bioinformatic Analysis

I have put together a pipeline designed to automate the genome assembly and variant calling for each nanopore run. I will continue to update this as need be. See the GitHub repo for up-to-date details.

A general overview of the pipelines is as follows:

De novo assembly

nanopore reads -> flye -> medaka polishing -> assembled genome -> prokka for annotation -> automated QC of nanopore reads and assembly via R script in Quattro

Variant calling relative to reference

nanopore reads -> `medaka_haploid_variant` -> VCF file