de twee grote vragen van dag 1 waren als volgt:

* na het PCR voor MinION, heb je veel kleine sequenties. deze worden aan elkaar geplakt om zo door de nanopore sequencing geanalyseerd te worden. voordat de data door de pipeline gaat, moet deze geknipt worden waar de reverse en forward primer elkaar ontmoeten om zo de originele sequenties te krijgen.

Deze parkeren we even omdat we dit op het lab niet goed voorelkaar krijgen

* de originele pipeline werkte met clustering om zo één sequentie eruit te krijgen, bij de MinION data willen we kijken of een chimera sequentie ook zou werken met BLAST’en

Eerst is het handig om onderzoek te doen naar MinION en welke data deze gebruikt. Vervolgens moet er gekeken worden of er al tools beschikbaar zijn om deze data aan te passen voor de al bestaande pipeline in galaxy. Hiernaast is er een gesprek gepland met een hoogleraar die al een MinION draaiende heeft. Hiermee kunnen we veel informatie krijgen over de data editing die nodig is voor dit project. Het is ook belangrijk dat ik de huidige pipeline begrijp, hiervoor heb ik het eindverslag van de vorige student gekregen. Hierin staat een uitleg in over de pipeline samen met een tutorial om de pipeline te testen. Om de MinION data te processen, is er een laptop aangeschaft waar de bijgeleverde software opgezet moet worden. Zodra dit gedaan is, is er accurate test data beschikbaar waarmee getest kan worden en gekeken kan worden hoe de pipeline aangepast moet worden. Als de pipeline op de juiste manier is aangepast en geïmplementeerd, is het van belang om dit alles goed te documenteren.

Traject 1

Vind een clustering tool waar de consensus wordt gepakt ipv de seed of maak deze.

Doe de tutorial voor de galaxy pipeline, let op!! de data en tools kunnen wat afwijken. Kijk vooral naar het clusteren!!

Traject 2

kijk of je de data uit wageningen in onze pijplijn kan analyseren. Hierop is de basecalling al gedaan. Hiervoor moet je volgens mij demultiplexen middels de (illumina) labels. Daarna moeten de samples in fastQ of fasta format in een zip komen.

Dan clusteren en identificeren.

Traject 3

Wat is de optimale manier van basecalling. Hogeschool gebruikt Guppy. De vraag is of dit goed gaat werken op complexe mixed amplicons. Mijn idee is dus een binning tool te ontwikkelen waarin de basecalling en clustering in 1 keer wordt gedaan op basis van x% identity en x% consensus zodat we hier al een sequenties krijgen waaronder ook chimeren maar die we dus wel nog kunnen demultiplexen. Lastig is dus dat ik niet weet of dit gaat werken omdat je natuurlijk geen chimeren wilt van de uiteinden waarop we willen demultiplexen. Daarvoor denk ik dat Traject 1 nog steeds van toepassing zal zijn na Guppy.

Guppy werk het best met GPU. Die heb ik niet zomaar liggen. Ik moet weven vragen aan martin Rucklin hoe de GPU’s op zijn 3D computers beschikbaar zijn.

We hebben ook een 7 jaar oude T7500, even aan Frank Stokvis vragen wat de specs zijn. In ieder geval genoeg cores maar de GPU weet ik even niet.

Na gesprek met Michael (IBL) hebben we advies om na de basecalling (FASTQ komt hieruit) BWA te gebruiken voor alignments te maken. Hier komt ook een BAM file uit. Daarna gebruik je RECON voor de consensus te maken.