Setting up a portable metagenome assembly pipeline for CBIO - Support #27

What we already know and have setup

03/26/2018 03:36 PM - Gerrit Botha

Status:	New	Start date:	03/26/2018
Priority:	Normal	Due date:	
Assignee:	Katie Lennard	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:		Spent time:	0.00 hour

Description

Jairam Lingappa vissited Heather's group last year and was planning to setup an metagenome assembly pipeline on Hex. He worked with Ulas Karaoz on this. We need to follow up how far he got with this setup.

Katie has visited JCVI and were introduced to some tools for this type of analysis. She also started in setting up a pipeline on Hex. Not sure if the pipeline was intended for Mark Nicol's or for Heather Jaspan's group.

Gerrit was part of the Wits course early 2017. Fiona did a short tutorial on metagenome assembly. No slides were shared she how ever worked with the tools: spades, kraken, prodigal, bandage, prokka, https://huttenhower.sph.harvard.edu/galaxy/ , https://www.onecodex.com/ .

Erik Hjerde is currently working with the Nicol group. They will be running some sequence data through their pipelines at University of Tromso. He pointed us to their course material here

Chris has also shared that they have been using DIAMOND for raw analysis and annotation. We can also think of including DIAMOND with MEGAN as a way to detect composition from raw reads.

The Wits group are also aiming to setup a similar pipeline. Maybe we can collaborate with them on that.

All in all we need to see what we already have setup and also what material we have gathered thus for to setup a production pipeline. Most probably in Nextflow.

History

#1 - 03/26/2018 03:38 PM - Gerrit Botha

Hi Katie,

Can you please add any info you have on the setup you have at Hex?

Also for who did you plan to setup the pipeline? Mark Nicol's or Heather's lab?

Regards,
Gerrit

#2 - 04/11/2018 02:22 PM - Gerrit Botha

Hi Katie,

I picked up this message from you on the bst-redmine@cbio.uct.ac.za account.

"Gerrit this was with Heather's future data in mind. I've been working with Ulas (Heather's collaborator) on this using tools including cutadapt, megan and CONCOCT (test dataset = publically available infant stool dataset). I've had to pause the development though to prioritise active requests from IDM members, hope to pick it up soon.

Thanks. You have send it on 26/03/2018. The Redmine IMAP scraper is however not compatible with our current IMAP server so tickets are not being populated when you reply on an email.

I've now setup the bst-redmine account that it directly forwards any email send to the mail address to my UCT address (I had the bst-redmine account setup on my thunderbird client but did not monitor it regularly). What it means is if you now directly respond on an email I will get it and I will then manually populate the ticket. However to make is easier for me rather respond to ticket directly in Redmine. Sorry for the hassle. I'm waiting for the CBIO mail server to be updated and will then reconfigure the IMAP scraping. I'm sure it will work then.

Regarding this ticket. Jai has not left any of the work he has done on Hex for us to continue with?

Gerrit

#3 - 04/11/2018 02:25 PM - Katie Lennard

No problem Gerrit. No, Jai only got as far as asking Andrew to install the necessary software.

#4 - 04/11/2018 02:26 PM - Gerrit Botha

Ok

#5 - 04/25/2018 03:24 PM - Gerrit Botha

I met the developer of this workflow: https://github.com/alesssia/YAMP at the Nextflow workshop last year. I do not think it is exactly what we want to do but maybe we can use parts of it.

#6 - 04/26/2018 09:51 AM - Katie Lennard

Thanks Gerrit,

This workflow https://bitbucket.org/berkeleylab/metabat is also an option - it is dockerised, uses less memory than comparable tools and appears to outperform other tools for binning.
I've mailed Ulas to ask for input on the final steps of the test pipeline on hex but he's hard to get hold of. Have also asked him about this metaBAT pipeline so may be more time efficient to switch to metaBAT.

#7 - 04/26/2018 01:13 PM - Gerrit Botha

Hi Katie,

Regarding metabat.

1. Yes it is dockerised but Hex does not support that so we would need to convert it to a Singularity image. We would need to look at the Dockerfile to see if it will work with us. Sometimes tools are packed so tricky in Docker containers that it makes calling them through a Singularity image very complicated.

2. We can install it using conda, which can be our first option to try. Packing tools later into containers using conda makes things much easier.

3. I can talk to Suresh to setup one of the nodes to be Docker enabled. You would however then need to switch between resources if find the metabat is the answer. Should be last option.

I also find that it is sometimes difficult to get Ulas's attention. If you want I can also email him?

Gerrit