labibi Documentation

Release 1.0

C. Titus Brown

December 12, 2014

Contents

CHAPTER 1

Practical bioinformatics topics / NGS

I'm developing a survey instrument that I can use to assess bioinformatics training needs at UC Davis, with a particular emphasis on practical sequencing data analysis. (Please see my blog post on training for more information and background.)

A few notes –

1. I intend this survey to be for biologists to fill out. So, I'm avoiding technical and foundational skills (cloud computing, Linux/UNIX, R, Python, managing large data sets).
2. I'm also avoiding sequence analysis approaches for which there are no established pipelines.

Below is my list so far. I welcome comments, additions, and critiques! The live site is at http://ngs-training-needs-survey.rtfd.org/.

Please feel free to copy, fork, and modify freely - the source for this is on github at https://github.com/ngs-docs/ngs-training-needs-survey.

1.1 Genome assembly and annotation:

- Assembling and annotating bacterial and archaeal genomes (w/Illumina, PacBio)
- Assembling and annotating non-plant/animal eukaryotic genomes (w/Illumina)
- Assembling animal genomes (w/Illumina)
- Annotating animal genomes
- Assembling plant genomes (w/Illumina)
- Annotating plant genomes
- Annotating bacterial genomes
- Annotating fungal genomes
- Long-read technologies for large genomes (PacBio, Moleculo)
- Emerging technologies for genome sequencing, assembly, and annotation in plants
- Emerging technologies for genome sequencing, assembly, and annotation in bacteria and archaea
- Emerging technologies for genome sequencing, assembly, and annotation in animals

1.2 Resequencing and variant calling:

- Variant calling on bacterial, archaeal, and fungal genomes
- Variant calling on plant and animal genomes
- Genotyping by sequencing

1.3 Transcriptomics:

- mRNAseq expression analysis in major model organisms (human, mouse, zebrafish, Arabidopsis, yeast, worm, Drosophila)
- ab initio transcriptome assembly, annotation, and expression analysis (semi-model animals, plants, and fungi)
- de novo transcriptome assembly, annotation, and expression analysis (non-model eukaryotes)
- Reference-genome-based bacterial and archaeal transcriptomics
- De novo mRNAseq in bacteria and archaea (no reference genome)

1.4 Metagenomics and microbial ecology:

- Amplicon analysis of populations and population structure
- Reference-based metagenomics (e.g. human microbiome)
- De novo shotgun metagenome and metatranscriptome assembly and analysis

1.5 Other:

- ChIP-seq analysis
- Reduced representation analysis of genomes and populations
- Marker development

1.6 More open-ended questions:

What bioinformatics software/programs are you using right now?

- CLC Workbench;
- Galaxy;
- Other (pls specify)

What compute resources are you using, if any?

- Laptop or lab computer;
- iPlant;
- XSEDE;
- DIAG;

- Amazon cloud;
- Davis Genome Center;
- Other cloud (specify)
- Other (pls specify)

What scripting or programming languages are you using, if any?

- MATLAB
- R
- Python
- Perl
- SAS
- Other (pls specify)

What do you feel is your major bioinformatics or sequence analysis-related obstacle, i.e. what is getting in the way of doing your data analysis?

CHAPTER 2

Indices and tables

- *genindex*
- *modindex*
- *search*