

Hands on Workshop: Introduction to Bioinformatics

Monash University
School of Biological Sciences

https://khodgins.github.io/Bioinformatics_Introduction/

Instructors



Dr. Sonika Tyagi
Sonika.Tyagi@monash.edu



Dr. Matt McGee
Matt.McGee@monash.edu



Dr. Greg Owens,
UC Berkeley

gregory.owens@alumni.ubc.ca

Greg's seminar!
Tuesday, 3.30 pm
Sanson Room

*Widespread structural
variation underlies
ecotypic differentiation
in sunflowers*



Dr. Kay Hodgins
kathryn.hodgins@monash.edu

Support from eResearch



Damien Leong
Senior HPC Consultant



Research DevOps Systems Engineer

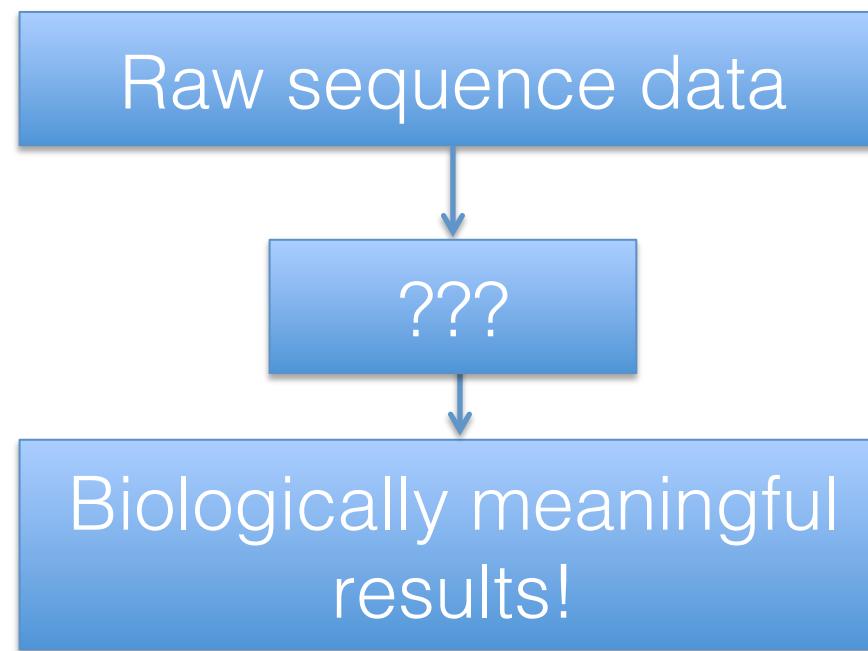
Topics

1. Workshop overview & sequencing technology and approaches
(Matt Tinning, NGS manager of AGRF) & R intro (Sonika)
2. Command line introduction (Day 1 PM) (Damien)
3. Sequence file formats & quality control/trimming (Day 2 AM) (Kay)
4. Sequence alignment (Day 2 PM) (Greg)
5. Variant calling (Day 3 AM) (Greg)
- 6 & 7. Population genomics & plotting in R (Day 4 AM) (Greg)
Thursday 9:30-1:30
8. Phylogenetics (Day 3 PM) (Matt)
Due to a scheduling conflict this topic will be Day 3 PM
9. Genome assembly (Day 5 AM) (Sonika)
10. RNAseq + differential expression analysis (Day 5 PM) (Sonika)

https://khodgins.github.io/Bioinformatics_Introduction/

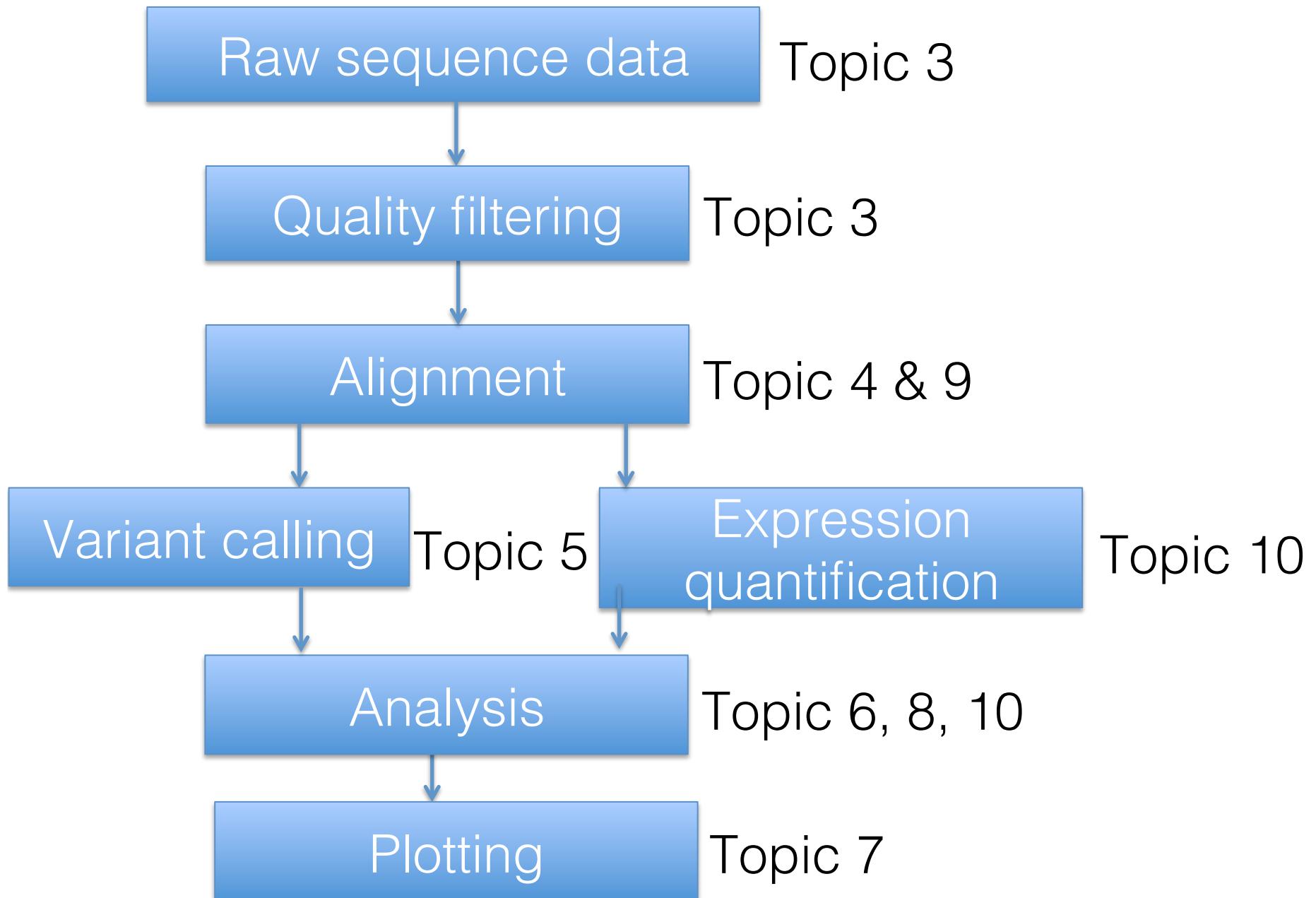
Workshop goals

- Provided theoretical knowledge and practical skills for the analysis of high throughput sequence data



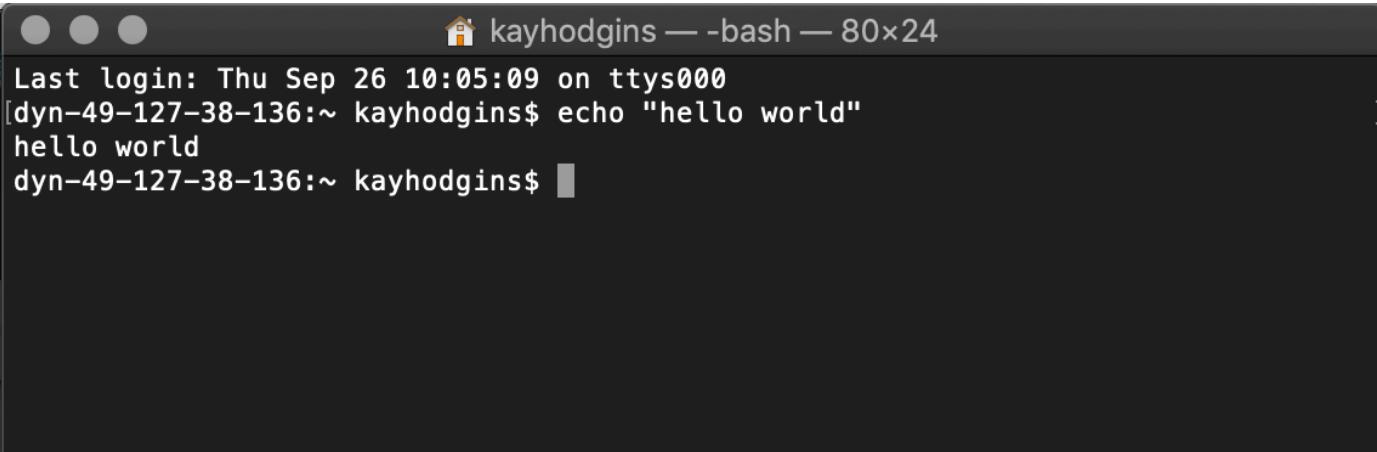
https://khodgins.github.io/Bioinformatics_Introduction/

Workshop goals



Workshop goals

- Learn command line – interact with computer without a GUI (graphical user interface) and a mouse etc
- Type commands in a terminal window



A screenshot of a macOS terminal window titled "kayhodgins — -bash — 80x24". The window shows a command-line session:

```
Last login: Thu Sep 26 10:05:09 on ttys000
[dyn-49-127-38-136:~ kayhodgins$ echo "hello world"
hello world
dyn-49-127-38-136:~ kayhodgins$ ]
```

- Topic 2 is an intro but all topics will involve command line
- Some non-command line options for bioinformatics (e.g. Galaxy, CLC workbench, geneious)

Why learn command line

- Extract subsets of a large file
- Powerful text editing tools
- Steps recorded and repeatable (easily reproduce the analysis)
- Generally faster than GUI based methods (automate file manipulation and program use)
- Arbitrarily combine programs and form a “pipeline”
- Many bioinformatics programs run only through command line
- Clusters/cloud computing is often only accessed through a command line interface
- Free!

https://khodgins.github.io/Bioinformatics_Introduction/

Shell scripting

- *the shell* is a program that takes commands from the keyboard and gives them to the operating system to perform
- *shell scripts* take command line arguments and put them into a reusable script
- easy to write
- provides the “glue” between programs written in different languages
- connects with external programs easily (awk, grep, sed)

https://khodgins.github.io/Bioinformatics_Introduction/

Workshop goals

- Lectures and tutorials designed to provide a solid foundation for your own genomics projects
- Assume limited bioinformatics/command line experience
- All lectures and tutorials as well as helpful links and background reading can be found at:

https://khodgins.github.io/Bioinformatics_Introduction/

Meet your neighbour

- Introduce yourself
- What is your background in genomics?
- What do you want to learn from this workshop?

https://khodgins.github.io/Bioinformatics_Introduction/

Schedule for today

10:15-11am sequencing technology and approaches

11-11:30am morning tea (thank you AGRF)

11:30-12 Introduction to R (Sonika)

Lunch 12-1

1-3pm command line tutorial (Damian)

https://khodgins.github.io/Bioinformatics_Introduction/



Australian not-for-profit organisation providing genomic services and expertise

Matthew Tinning:
Next Generation
Sequencing Manager

