

Introduction To Bash

Using the Terminal For Bioinformatics

Steve Pederson
Bioinformatics Hub
Level 4, Santos Petroleum Engineering Building

University of Adelaide

19th September, 2017

Today's Tutors

- Dr Hien To (Bioinformatics Hub, Adelaide University)
- Dr Jimmy Breen (Robinson Research Institute / Bioinformatics Hub)
- Dr Terry Bertozzi (SA Museum)
- Dr Aaron Casey (SAHMRI)
- Alastair Ludington (Bioinformatics Hub, Adelaide University)
- Robert Qiao & Exe Escobedo (Phoenix Support Team)

Why use the command line?

- We can work with huge files
- All HPC systems require basic scripting skills
- We retain more control over data and processes
- You better understand every process you've done
- It's easy & fun!

Why use the command line?

- Science is littered with retractions as a result of GUI mistakes
 - Google “Keith Baggerly”
- Reproducible Research \implies Script Everything!
- Nothing is forgotten & you can find any errors
- Common workflows can be easily automated

Today's Topics

1. We'll start very simple with copying files, making directories etc, how do we access internal help pages...
2. Important techniques for working with files
3. *Regular Expressions* for text searching
4. Two useful tools: `sed` & `awk`
5. Writing *Shell Scripts*

Today's Schedule

- Four 90 minute sessions
- Today is very much self-guided, working at your own pace
- Use the Post-it Notes
 - Yellow/Red: Help!
- Ask lots of questions!

Setup

- Setup is often a bit bumpy
- Using local installations for the first time
- We have 4 emergency USB drives only

Today's Notes

- Today's notes are available at

`https://uofabioinformaticshub.github.io/`

`Intro-Bash-Sept-2017/`

Most common problems

- You can't tell the letter 'l' from the number '1'
- Always, always, always use "Tab auto-complete"

After Today

- Today is *An Introduction!*
- Don't be afraid to ask for help (Bioinformatics Hub)
- Phoenix support in the Bioinformatics Hub Thurs 11am-1pm
- An excellent resource is www.linuxcommand.org/tlcl.php