



Public Health  
England

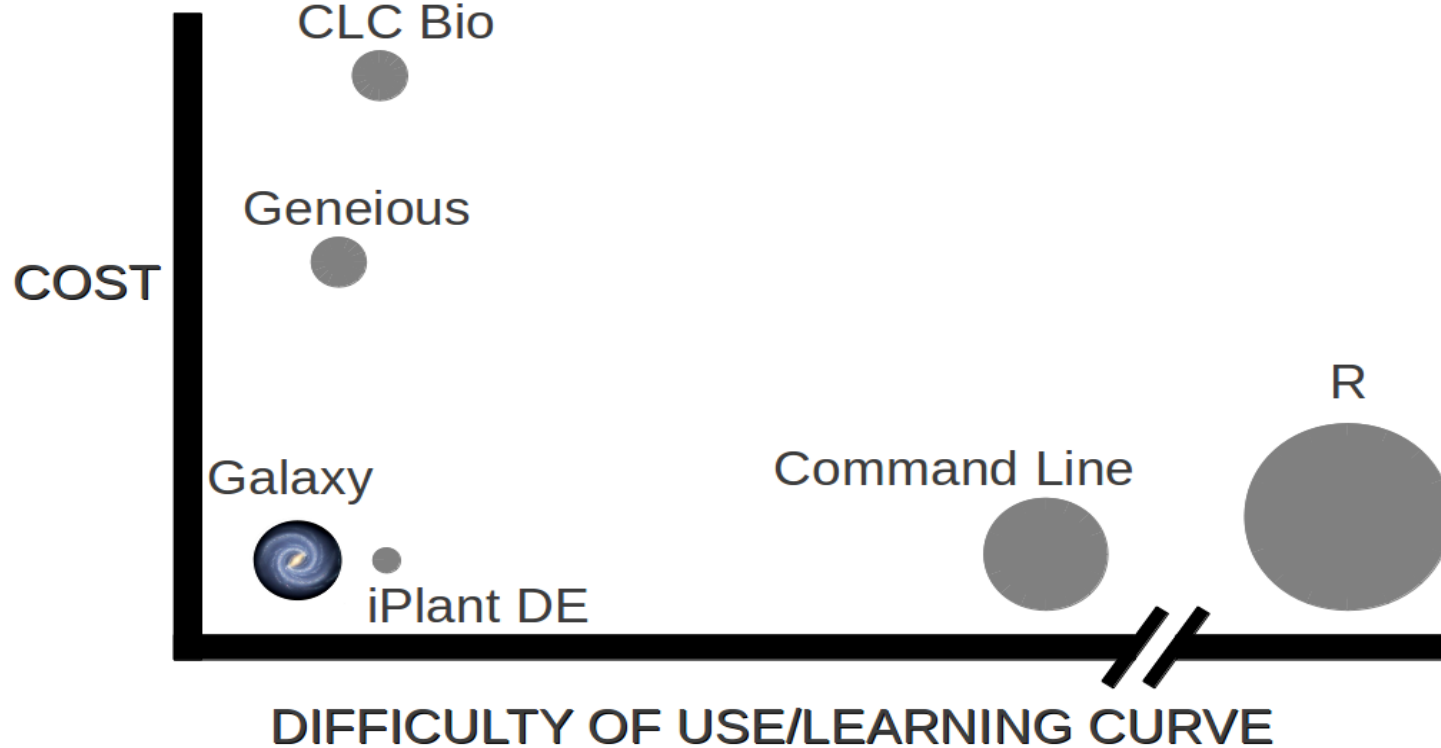
# Training Microbiologists in NGS Analysis: from CLI to Galaxy: Pain to Gain

Ali Al-Shahib  
Public Health England  
Bioinformatics Unit  
London



# Comparison of available Bioinformatics platforms

## RNA-Seq Analysis Tools



Size of dot indicates flexibility/power



# What we used to use before

```
ulf@depot:~  
HOSTNAME=/bin/hostname 2>/dev/null`  
HISTSZIE=1000  
if [ "$HISTCONTROL" = "ignorespace" ] ; then  
    export HISTCONTROL=ignoreboth  
else  
    export HISTCONTROL=ignoredups  
fi  
  
export PATH USER LOGNAME MAIL HOSTNAME HISTSZIE HISTCONTROL  
  
# By default, we want umask to get set. This sets it for login shell  
# Current threshold for system reserved uid/gids is 200  
# You could check uidgid reservation validity in  
# /usr/share/doc/setup-*/uidgid file  
if [ $UID -gt 199 ] && [ "`id -gn`" = "`id -un`" ]; then  
    umask 002  
else  
    umask 022  
fi  
  
for i in /etc/profile.d/*.sh ; do  
    if [ -r "$i" ]; then  
        if [ "${-#*i}" != "${-}" ]; then  
            . "$i"  
        else  
            . "$i" >/dev/null 2>&1  
        fi  
    fi  
done  
  
unset i  
unset pathmunge  
ulf@depot ~]$ ll  
total 332  
drwx----- 20 ulf ulf 4096 May 21 12:49 .  
drwxr-xr-x. 14 root root 4096 Nov 12 2013 ..  
-rw----- 1 ulf ulf 90515 Nov 5 11:39 .bash_history  
-rw-r--r-- 1 ulf ulf 18 May 10 2012 .bash_logout  
-rw-r--r-- 1 ulf ulf 176 May 10 2012 .bash_profile  
-rw-r--r-- 1 ulf ulf 1458 Jul 12 2013 .bashrc  
drwxrwxr-x 2 ulf ulf 4096 Nov 4 2013 bin  
drwxr-xr-x 3 ulf ulf 4096 Jul 25 2013 .cache  
drwx----- 4 ulf ulf 4096 Jul 29 2013 .config  
drwx----- 3 ulf ulf 4096 Jul 8 2013 .dbus  
drwxr-xr-x 2 ulf ulf 4096 Jul 25 2013 Desktop  
drwxrwxr-x 2 ulf ulf 4096 Jul 8 2013 Downloads  
drwx----- 3 ulf ulf 4096 Oct 22 2013 .gconf  
drwx----- 2 ulf ulf 4096 Oct 22 2013 .gconfd  
drwxr-xr-x 4 ulf ulf 4096 Jul 25 2013 .gnome2  
-rw----- 1 ulf ulf 1116 Jan 30 2014 .lesshst  
drwxrwxr-x 3 ulf ulf 4096 May 6 2014 .matplotlib  
drwxr-xr-x 4 ulf ulf 4096 Mar 15 2012 .mozilla  
drwxr-xr-x 2 ulf ulf 4096 Jul 25 2013 .nautilus  
drwxrw---- 3 ulf ulf 4096 Jul 8 2013 .pki  
drwxrwxr-x 4 ulf ulf 4096 May 6 2014 project  
-rw----- 1 ulf ulf 125171 Oct 22 2013 .recently-used.xbel  
drwx----- 2 ulf ulf 4096 Jul 16 2013 .ssh  
drwxrwxr-x 2 ulf ulf 4096 Oct 29 2013 temp  
drwxr-xr-x 3 ulf ulf 4096 May 1 2013 Trimmomatic-0.30  
drwxr-xr-x 2 ulf ulf 4096 Aug 12 2013 .vim  
-rw----- 1 ulf ulf 12308 May 6 2014 .viminfo  
-rw----- 1 ulf ulf 198 May 21 12:49 .Xauthority  
ulf@depot ~]$
```



# What we use now

The screenshot displays the Galaxy / PHE web interface. The browser address bar shows the URL 158.119.147.85. The interface includes a navigation menu with options like Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. A 'Using 49%' indicator is visible in the top right. The main content area features a green welcome message: 'Welcome to the PHE Galaxy environment' with a checkmark icon. Below this is the Public Health England logo and a paragraph of text: 'Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.' The left sidebar contains a 'Tools' section with a search bar and a list of tool categories: Get Data, Text Manipulation, Filter and Sort, Join, Subtract and Group, Operate on Genomic BED Intervals, Convert Formats, Genome Annotation, Extract Features, Picard tools, Statistics, Graph/Display Data, NGS: Simulation, Multiple Alignments, Phylogenetics, FASTA manipulation, NCBI BLAST+, NGS: PHE internal tools, NGS: QC and manipulation, NGS: Assembly, NGS: Mapping, NGS: SAM Tools, NGS: GATK Tools (beta), and Workflows. The right sidebar shows a 'History' section with 'Unnamed history' (0 bytes) and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.



Public Health  
England





Public Health  
England

# When using tools like Galaxy, you need to...





# Training Contents

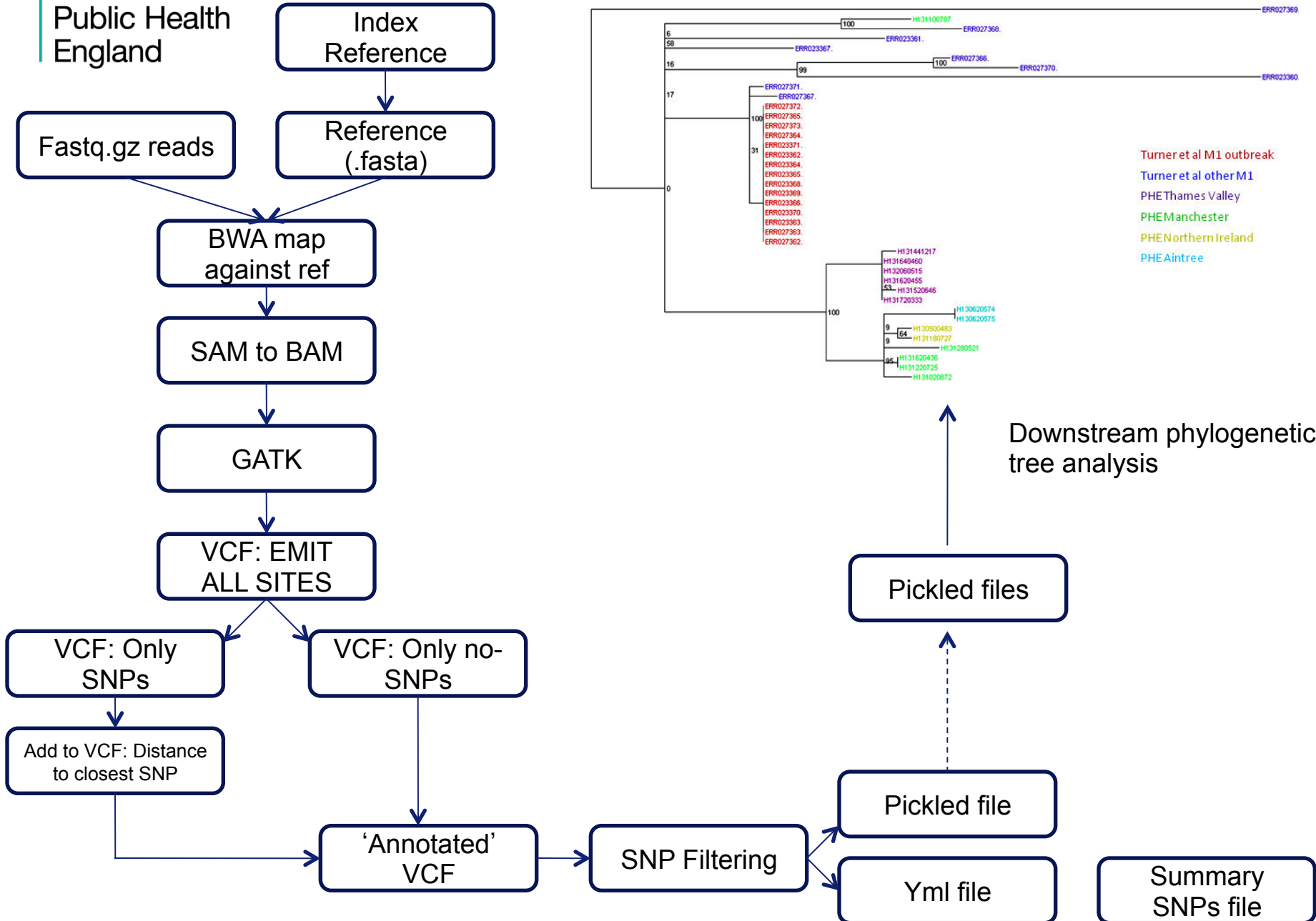
## Half a day:

- Note on disk space
- Upload Data into History
- Delete from History
- Computing Sequence length
- Creating Workflows
- Calculating the GC content and adding it to Workflow
- Compute sequence length + GC Contents Workflow
- Gene presence/absence exercise
- SNP pipeline



Public Health  
England

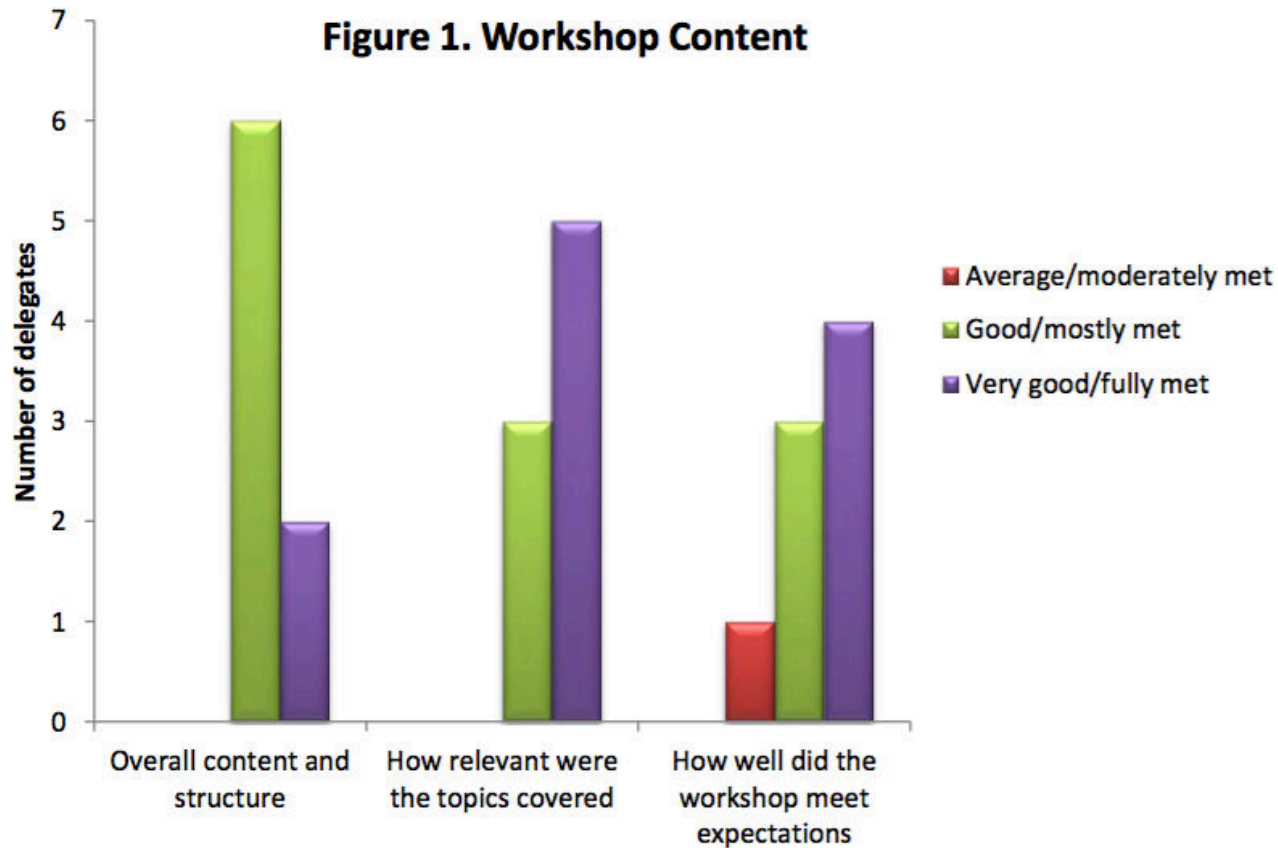
# SNP Pipeline





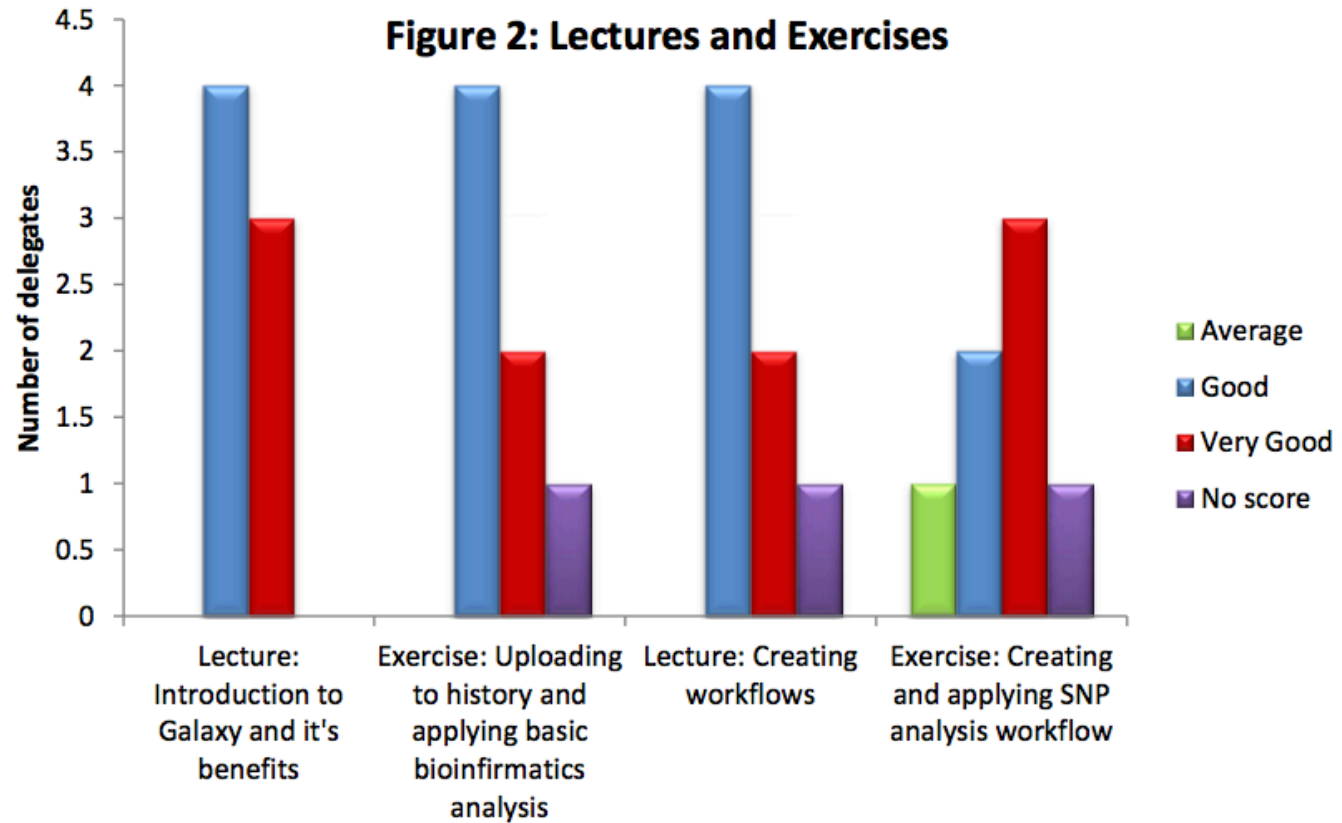


# Feedback





# Feedback





- *“Really enjoyed Galaxy”*
  - *“Entailed important material and content relevant to our day to day work.”*
  - *“Nice intro to the web app. Building workflows and creating histories”*
  - *“Need to be able to analyse NGS from lab. Feel that I can start Gaining an insight into the Galaxy platform”*
  - *“Enjoyed it very much.”*
- *“Perhaps one day on Galaxy”*
  - *“I didn’t get fair help to get through the course. It was very new to me.”*
  - *“Find Bioinformatics very difficult to understand”*
  - *“Got a bit confused Ali went into it too fast”*



Public Health  
England

And finally...

*Thank you to the Galaxy team  
For their hard work  
And keep up the good work!*

