Power Point 01 - Bioinformatics Topics - Part 00 -Informatics

The main purpose of this talk is to consider the major components of Bioinformatics.

The intention here is not to be comprehensive, but to suggest a frame work in which to fit the topics covered in your course.

<Click>

Here, I have divided the Bioinformatic topics to be considered into those that are primarily "informatics" issues

<Click>

and those that are directly concerned with the management of Biological data.

<Click>

The most fundamental Informatics issue is the Operating system.

One must be able to control the computer before one can expect it to do anything useful!

<Click>

Currently, and many would argue unfortunately, Windows is still the most common operating system, followed ever more closely by the Macintosh Operating system.

Both Windows and the Macintosh offer very sophisticated and intuitive Graphical User Interfaces (or GUIs) with which most people under 35 have been familiar since birth!

<Click>

The Linux operating system is becoming increasingly popular.

These days, Linux comes with very good graphical user Interfaces that, distressingly to some, ever more closely mimic Windows.

Linux, with a good Graphical User Interface should seem very familiar to all of you.

I suggest it is safe to assume familiarity with any of these systems with a good GUI.

<Click>

The real issue comes when it is necessary to use Linux from the command line.

That is to say ... no graphics, no mouse, just a keyboard!

This can be a bit of a culture shock for some, but it really is not as difficult as it seems.

Especially, after a gentle introduction and a bit of practice such as will be delivered as part of your current training programme.

<Click>

If you find command line Linux a struggle, it should be comforting to appreciate that it is conceptually identical to any GUI based Operating System.

Any conversation you might have with any operating system must exclusively concern the management of files with a hierarchy of directories (or folders) or which program you want to run next.

You can do this using a fancy GUI, or by typing instructions with a keyboard.

It is just the same.

<Click>

Command line Linux really is the only option for some tasks in Bioinformatics.

Specifically, it would be crazy to write programs to process high volume sequence datasets for a clumsy bloated OS such as Windows!

Compute and memory intensive software, required to run in a liveable time period, are Linux only.

Generally, it is not considered a worthwhile use of programming resources to provide friendly interfaces for such software.

Its command line or nothing I fear.

<Click>

A basic understanding of the principles of programming has become ever more important in recent years.

<Click>

It is generally not necessary for a Bioinformatics "user" to become a proficient programmer as the need for an individual researcher to write software from scratch is rare.

<Click>

However, it is of great benefit for many users to be able to construct simple scripts (or programs) to manage the large datasets now so common.

<Click>

Also, the capacity to write scripts to customise the order and fashion in which a series of programs might be invoked (that is, to construct simple pipelines) is commonly of value.

<Click>

Currently, the most popular programming language for Bioinformatics is Python.

The basics of Python can be mastered by most in just a few days of training and a little practice.

Learning Python is outside the scope of this course, but would be well worth the effort at some future date.

<Click>

Minimally, an understanding of programming that would enable a user to construct small programs from scratch, understand and adapt slightly larger programs and communicate meaningfully with a specialist programmer would be a reasonable objective.

<Click>

As more biological research becomes focused around the analysis of huge datasets, a good grasp of Statistics becomes ever more essential.

Like programming, this is beyond the remit of your current training program, but also might be worth some effort at a later date.

<Click>

Statistical evalution of an experiment is vital from the design stage.

It is too easy to invest resources in experiments that have no hope of generating results that are statistically meaningful.

It is always a very good plan to take careful note of the oft quoted advice of Sir Ronald Fisher, first offered way back in 1938!

<Click>

Having conducted a well designed experiment, generating vast quantities of raw data, Statistical analysis, implemented with quality Statistical software (that is a Statistical Package), is unavoidable.

The Statistical Package R is widely used in Bioinformatics. It is comprehensive, excellent quality, includes libraries specific to Bioinformatics ... and is free!

<Click>

Effectively all evaluations generated by Bioinformatics software are statistical judgements.

That is, the selection of the most probable answer from a set of many possible answers to a given question.

This point will be reinforced as we consider various individual analytical possibilities.

<Click>

And now, once more, it is your turn! To think though a few simple issues.

There is no intention here to be tricky. In fact, I have embedded some links to what I hope will be "helpful hints" for you.

The intention is to ensure that all the simple issues of these presentations are understood and to invite you all to be more than passive receptors of the wise thoughts paraded before thee!