Bioinformatics:

The Power of Computers in Biology

Lesson Plan (100 mins)

|  |
| --- |
| Version 3.0 (extended)  4273pi Bioinformatics Education Project, <https://4273pi.org> |
|  |
| Copyright and related rights waived via CC0 1.0 Public Domain Dedication |
| (<https://creativecommons.org/publicdomain/zero/1.0>). |

***IMPORTANT:*** *Task C depends on our Linux computer server,* ***which is private****. If this server has not been made available to you: either switch to the short (50 minute) version; or equivalently, remove Task C from the slides, worksheet and answers and extend the “Own animals” section.*

Timings are approximate and assume a total of **100** minutes.

If more time is available, longer can be spent on “Own animals intro” and “Students search animal genomes of their choice”.

Students should work in **pairs** where possible.

The general structure is:

* Intro presentation/interaction.
* Task A: introduction to bioinformatics online.
* Task B: in-depth look substitution and particularly frameshift.
* Task C: introduction to bioinformatics in a scientific computing environment.
* “Own animals” (off worksheet, if time permits): students apply learning from Task B to look at animals of their choice.

## BEFORE THE WORKSHOP – 15 minutes

1. **Off-site:** Log in to the Linux server as tutor and enable user accounts.
2. **On-site:** Check the NCBI BLAST site is working (Tasks A and B). If not, be prepared to demonstrate EBI and/or Ensembl BLAST sites.
3. **On-site:** Check Linux server is accessible via a school Web browser (Task C).

## INTRODUCTION – 15 minutes – 0:00 to 0:15

**Who we are**

Briefly.

**What is Bioinformatics?**

Bioinformatics is on the syllabus. It is not very clear from the textbook what it actually involves. Bioinformatics involves computer analysis of biological data, especially DNA sequence. We will show you some practical bioinformatics.

**Double helix**

DNA sequencing generates large files of As, Cs, Ts and Gs, from one strand of the double helix. This generates too much information interpret without a computer.

**Mutations**

***Interact with class: who can tell us one of the types of mutation?***

Terminology will vary depending on current SQA requirements and teaching so far (point mutation vs single gene mutation).

Answers may include valid mutations not covered in our workshop (e.g. duplication).

This is all fine.

Guide students towards substitutions, insertions and deletions, which will be seen in our workshop.

***Interact with class: who can tell us what a frameshift mutation is?***

**Today’s practical exercise / sequence**

We cannot tell what this sequence does, just by looking at it.

**Today’s practical exercise**

Pause presentation here.

## TASK A – 10 minutes – 15:00 to 25:00

Hand out pens as required. Everyone should write answers on their own worksheet.

**Lock-step**

After students have finished, ask them what they found. Emphasise how quickly this conclusion was obtained (using bioinformatics).

## TASK B – 20 minutes – 25:00 to 45:00

Make sure everyone sees and understands a frameshift mutation.

## RECAP AND IMPLICATIONS – 10 minutes – 45:00 to 55:00

**Practical: Results**

We found mouse have a functional gene coding for the enzyme that makes vitamin C, humans don’t.

***Interact with class: how do we get our vitamin C?***

***Interact with class: what happens if we don’t get enough vitamin C?***  
Answer: scurvy – gums bleed, our teeth fall out, bleeding sores all over the body, death from blood loss. Vitamin C is essential for synthesis of collagen. (NB only if it comes up: vitamin C does not prevent colds and flu.)

***Interact with class: who often suffered from scurvy in the past?***

Before people knew about vitamin C, sailors tried various treatments for scurvy that did not work, e.g. drinking seawater. Eating the rats on the boat helped to some extent.

***Interact with class: why might eating rats help?***

Answer: rats are close relatives of mice and can also synthesize vitamin C. They won’t have a lot in their bodies, but it could still be helpful.

NB do not try this at home – eat an orange instead.

Vitamin C is essential. The frameshift mutation occurred at random, in the genome of the ancestor of humans and most other primates.

***Interact with class: why wasn’t this mutation fatal?***

Answer: enough vitamin C in the ancestor’s diet – it must have been living in a place with a good year-round supply.

The frameshift either was a neutral mutation, or possibly (we don’t know) a slight advantage since it saves some energy if we do not synthesize vitamin C. Anyway, the mutation spread to fixation and we are stuck with it now. We have to eat vitamin C.

Bioinformatics has allowed an investigation of mutations and also we can begin thinking about diet and evolution.

## LINUX INTRODUCTION – 5 minutes – 55:00 to 60:00

**Slides “Your investigation” to the photo of a hacking screen**

Details are flexible.

***Interact with class: does anyone use the cloud? What is the cloud?***

We will use something approaching the cloud – a server provided by our contractor in Aberdeenshire.

Linux command-line as an alternative way to run BLAST and scientific analyses.

## TASK C – 15 minutes – 60:00 to 75:00

Provide usernames on paper. Provide URL and password on-screen.

Make sure all groups get a BLAST result and find the frameshift mutation.

## TASK C RECAP – 5 minutes – 75:00 to 80:00

Command-line has advantages for repeatability, as well as giving the ability to parallelise our work. This is highly relevant when we are analysing very large datasets, such as whole genomes.

***Interact with class: who preferred Web, who preferred the Linux server?***

Any answer is OK. We use both.

The same biological conclusions are drawn, as we would expect.

**Linux rules supercomputers**

Used on the world’s fastest supercomputers. If you ever use such a machine, you will now have an advantage over most people in the world.

## OWN ANIMALS INTRO – 5 minutes – 80:00 to 85:00

Note: this can easily be contracted or expanded to suit the time available.

**Animals slide**

Pause presentation here.

Demonstrate BLAST of guinea pig genome

Same as Task B, except:

1. Type name of animal in box, don’t click human
2. Select animal name from drop-down menu

Still remember to click “Somewhat similar sequences” on the BLAST form page.

**Interact with class: does/did anyone own a guinea pig or know someone who did? What did it eat? Can we make a hypothesis about whether it had a gene (like mice and rats) or a pseudogene (like ourselves)?**

Scroll down through BLAST results until a frameshift mutation is seen.

Other worthwhile demonstrations, if time and mood permit: rat (very similar to mouse), gorilla (very similar to human).

## STUDENTS SEARCH ANIMAL GENOMES OF THEIR CHOICE – 10 minutes – 85:00 to 95:00

Ask students to choose their own animals, search their genomic DNA with BLAST. Either ask for conclusions or – if there will be time – ask students wto write on a board or in a shared Excel spreadsheet:

ANIMAL: GENE/PSEUDOGENE/UNCERTAIN

Encourage an “uncertain” conclusion where appropriate – mainly if BLAST failed to find the right part of the genome, e.g. due to little DNA being in the database so far.

**NB intriguing results:**

1. Many mammalian carnivores e.g. dog, cat, bear, have the pseudogene AND a gene. Show students that the top two matches both have low E-values. They have to look at both.
2. Emperor penguin and some other birds have a frameshift which is almost immediately followed by another frameshift restoring the original reading frame. These are genes not pseudogenes, though an “uncertain” conclusion would be fine.

Students tend to over-predict pseudogenes, due to misinterpretation of statistically nonsignificant alignments and also due to limitations in the workshop’s methodology. But don’t dwell on mistakes unless there is time to explain more fully (e.g. the pseudogene and gene in mammalian carnivores). It is interesting to look back on results anyway.

## WRAP-UP – max. 5 minutes – 95:00 to max. 100:00

**Be sure to say**: If the hand-out seemed intense, this is because **it is based on a university practical class**. This ran at the University of St Andrews for many years. Students all got through it fine – should congratulate themselves.

**Hand out answer sheets to students**. They can keep the worksheet and answers.

Perhaps photograph the board with student results on it.

## AFTER WORKSHOP – 15 minutes

**On-site:**

Check all connections to Linux server are logged out or closed.

Leave room tidy.

**Off-site:**

Use admin rights to disable user accounts on the Linux server.

Delete files students have created.

Change user passwords.