BINF sights.

junk DNA

how remnants of ancient viruses raise exciting questions in the field of paleovirology.

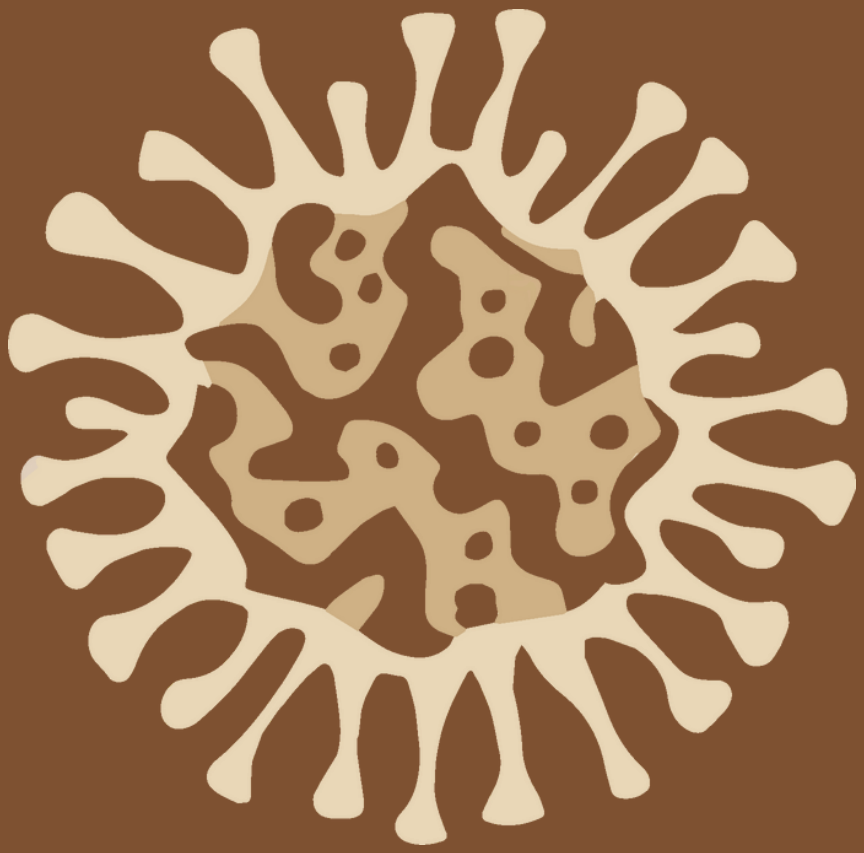


IMAGE SOURCE:
Image Creator
powered by DALL·E

Society News

BINFSOC

Recent Events.

-- Bioinformatics Debugged

-- Annual Networking Night

-- BINFSOC BBQ

-- BINFSOC Movie Night - Gattaca

-- Introduction to R Workshop

Annual Publication.

Do you any pressing you want answered about the field of bioinformatics? Worried about what comes after you graduate? Curious about job opportunities, journal articles, or academia?

Look no further!

For BINFSOC's annual publication, we are hoping to release a tell-all article featuring findings from professionals in the field of bioinformatics, recent graduates, and academics about their experiences.

If you have any questions you want answered, come to any of our BINFSOC events and ask an executive or sub-committee member (preferably Elisabeth or Vibhuti) to include your question in the publication and we'll make sure to ask it! Also feel free to email at **digitalbranding@unswbinfsoc.com**

Recent Event Roundup

Some of our most exciting recent events include some social activities! We sat down, got the popcorn out, and dimmed the lights as we watched *Gattaca (1997)* which was a great opportunity to see some (relatively) realistic science in entertainment! Tiger Woods has nothing on us as we practiced our golfing skills at Holy Moley in Newtown. Follow our Facebook page for any new events, and we are excited to see you at the next one.

Networking Night

BINFSOC's Networking Night was one of our biggest events this year, where we saw representatives from industry and aspiring bioinformaticians discuss what it's like to work in the field of bioinformatics, how to navigate and get started in the world of biotech and software, and any exciting projects or research that they are involved in. We would like to extend a huge thank-you to our representatives who attended; Laurence Wilson from CSIRO, Joseph Copty from the Garvan Institute, Mark Cowley and Antoine de Weck from the Children's Cancer Institute, Bruno Gaeta, Raymond Louie, Soleille Miller, Hamid Alinejad Rokny from UNSW and Margarita Psaras, Phuong Lam Phung and Daniel Cordeiro from Quantium. We hope to continue hosting events just like this one, so stay tuned!

viral fossils and 'junk DNA'

EMMA HARDING, current PhD candidate and a researcher working with Dr. Peter White, writes about how remnants of ancient viruses in marsupial cells pose very interesting questions for the future of paleovirology.

***Author**: Emma Harding **Compiled**: Vibhuti Nandel, Elisabeth Cola*

relics from the past

Viruses are everywhere: in the ocean, in our gut, on our shoes … but did you know they are in our DNA as well? Endogenous viral elements (EVEs) are fragments of viral sequences that are preserved in our DNA from past infections. Some viruses like retroviruses (eg. HIV) integrate into our DNA as part of their replication cycle, whilst others are accidentally integrated by chance. If this integration occurs in a germline cell, the DNA from the virus is then passed on to our offspring. Over many generations, these viral fragments become part of the normal chromosomes of a population. These EVEs act like viral "fossils" and can be used to trace the long-term evolutionary history of viruses over hundreds of millions of years.

In recent years, the availability and quality of DNA sequencing has drastically improved. As a result, animal genomes are being published at unprecedented rates – along with all of the EVEs inside them. My work in Professor Peter White's lab in BABS focuses on finding EVEs in a range of animals to better understand where modern-day viruses came from.

To do this, we use a range of bioinformatic techniques to identify sequences based on similarity, parse and filter them, conduct multiple sequence alignments and phylogenetic analysis. From this, we obtain a list of hundreds to thousands of EVEs from each animal comprising fragments from various ancient viruses.

Our most recent study, published in Virus Evolution, identified ancestors of Ebola virus (Filoviridae), Borna disease virus (Bornaviridae) and parvoviruses (Parvoviridae) from a range of Australian marsupials. No viruses like this have ever been isolated from marsupials, but this technique of studying EVEs allows us to trace these viral lineages regardless.

Analysis on marsupial cells (such as koala's) have shown fragments of ancient viruses. Image: Canva

In addition to finding and characterising the EVEs, we also performed analysis of the small RNA from marsupial cells. Small RNA is involved in a myriad of cellular functions including development, antiviral resistance and mRNA editing. We found that some of the EVEs were

being transcribed as small RNA molecules, indicating that rather than just being “junk fossil DNA”, EVEs could be playing important roles in our cells. Could these viral fossils be protecting us from future infections, like an inheritable vaccine? Could they be triggering diseases like cancer?

This field of paleovirology is very young and many questions have yet to be answered, but it is definitely an exciting area of research to watch!

Emma hopes to publish another paper this year on ancient marsupial viruses and we are very excited to see more excellent work and research coming …

Read more about this work here:
https://academic.oup.com/ve/article/7/2/veab076/6362200

Keep up to date with Emma and her research here:
https://www.unsw.edu.au/staff/emma-harding

Code Your Way

Welcome to "Code Your Way"

This edition onwards the BINFSOC team will put forward a section discussing some shell , python, perl or R commands. These will be further illustrated with examples that are relevant to Bioinformatics be it fetching data, manipulating it or visualizing. We hope you enjoy learning or refreshing your memory.
If you want us to cover something do write to us at the address shared on the last page.

Shell Scripting

Before we dive into the commands lets have a basic understanding of what is shell? What is shell scripting and why do we use it?

What is Shell ?

It is a type of a program which accepts the command entered by the user. It then analyses the command, interprets it for the computer, then executes it.

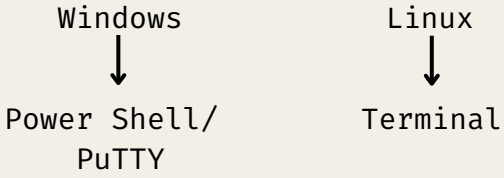
Imagine you want to fetch a couple of sequences from a database and dont want to type the command over and over again.

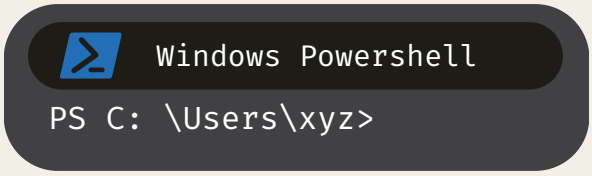
You can simply write a shell script that defines a command to fetch the sequences corresponding to the specific IDs and store it in a folder. So this shell script basically reduces your effort and time. Hence, shell scripting can be used to automate such repetitive tasks.

Now that we have an idea of what and why do we use shell scripting lets start with some commands.

Connecting Remotely to your Workspace

Imagine you are feeling super lazy today and don't want to go to UNSW and sit in front of the computer just to access your data on Katana or your lab workstation. But today you need to submit your assignment. So how do we access the data?





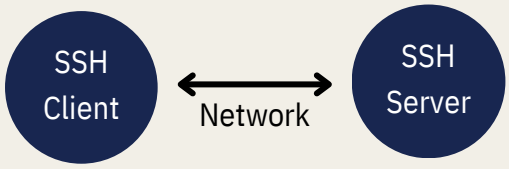
Terminal or terminal emulator is a program that allows the user to run the commands.

1 First, make sure you have an internet connection

```
> ping www.google.com
```

The ***ping*** command informs the user the amount of time taken by the message to go back and forth between the server and user.

2 Access your workspace via SSH



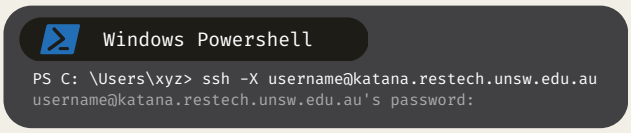
Now that you are sure that the internet connection is there. Let's access your workspace via SSH, that stands for Secure Shell. SSH allows the user (SSH client) to communicate with SSH server.

As an example, here I am connecting to my workspace on UNSW HPC, Katana.

```
>  ssh -X username@ip_address
```

-X option lets the user to remotely run graphical apps as it enables X11 forwarding.

3 Follow the prompts



You will be prompted to enter your password

4 Explore the manual

Other options of ssh can be accessed by typing

```
>  man SSH
```

This opens the manual and provides you the details of the options. Or you can access it here: https://man7.org/linux/man-pages/man1/ssh.1.html

ATTGCCTTAGATSHELL
ACTGAATGCACORNER_

Readings: https://goteleport.com/blog/x11-forwarding/

5 Example: grep command

```
> grep [option...] [pattern...]
          [file...]
```

grep or egrep are two commands that are used to match patterns find in inputted files. These can be used to find specific lines of data that may contain a phrase (e.g. username, subset) that you need to analyse or compile.

Normally they can be very complex but lets suppose a random file of data (e.g. integers between 1 and 1000) has 1000 entries but you only want to access lines that start in the 600's but doesn't contain any odd numbers...

randomFile.txt

```
129
426
794
303
495
180
926
509
367
626
907
...
```

Windows Powershell

```
PS C: \Users\xyz> egrep -x "^6[^13579]*" randomFile.txt
626
686
68
628
644
664
...
```

6 Other common commands

Below are examples of some other common commands, discover what they mean by using the man pages!

```
> man ls
> man rm
> man cd
> man mkdir
> man cp
> man cat
> man sed
> man wget
> man echo
> man chmod
> man diff
> man pwd
```

R

What is R?

Believe it or not, this isn't the set up of a really bad pirate joke. R is a programming language used for statistical computation and is very common in the world of data analytics, bioinformatics, and even risk and insurance analysis.

The R environment provides different facilities for:

- *performing simple and complex calculations*
- *handling, manipulating, analysing and graphically presenting data effectively.*

Set-up

You can access R simply by tying R in the terminal (if you already have R installed) or you can download an IDE e,g. **RStudio** via
https://posit.co/downloads/

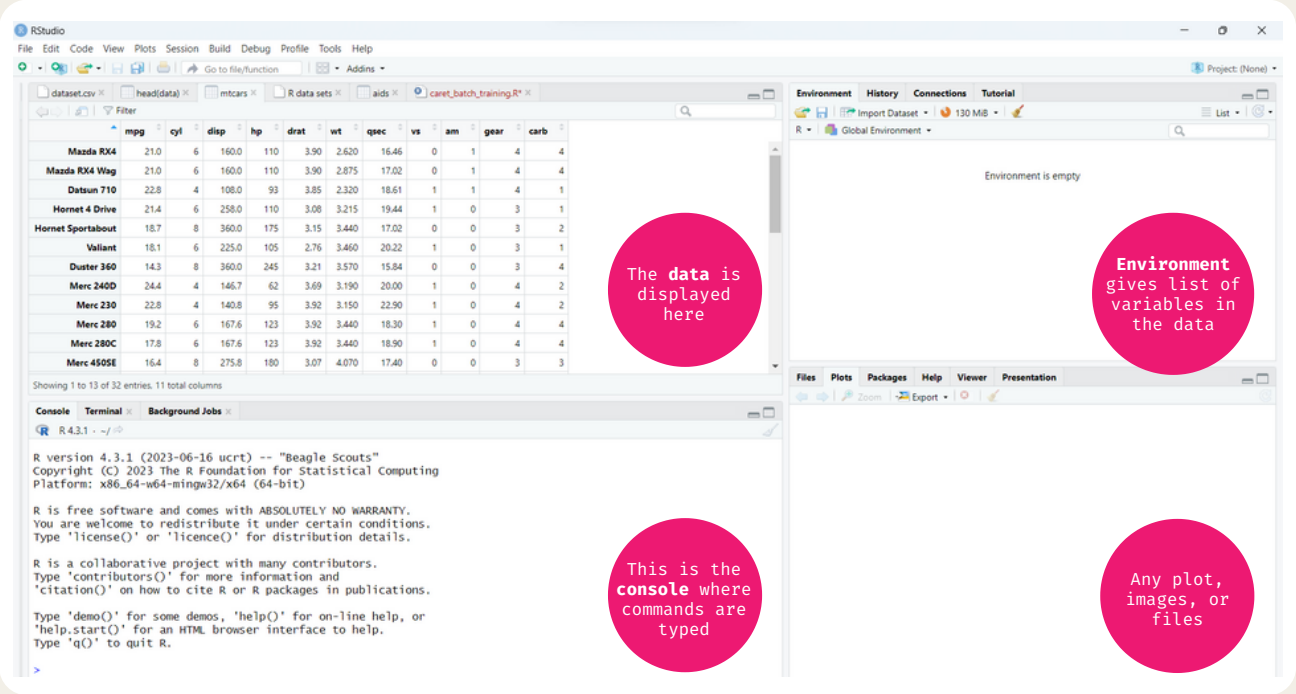
Different packages are available that help perform the above purposes seamlessly and also provide bespoke solutions to many other problems/data sets you may encounter.

In the example below, packages used will be mentioned alongside. For illustration we will use **RStudio**.

Reference:

Venables, William N.., Smith, David M.. An Introduction to R: A Programming Environment for Data Analysis and Graphics. United Kingdom: Network Theory, 2009.

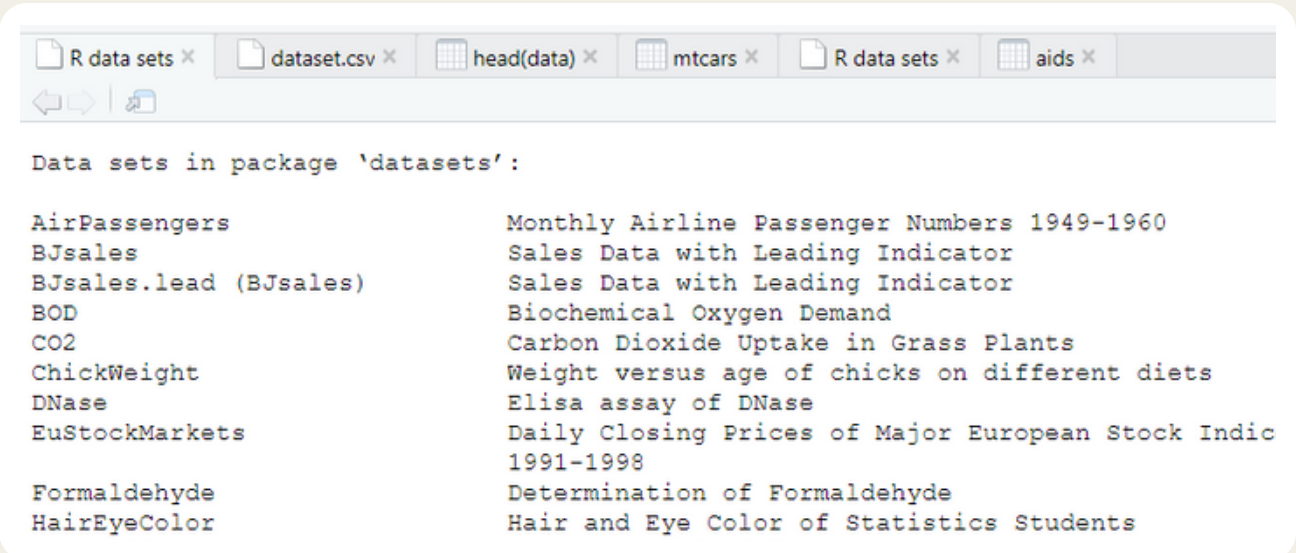
1 RStudio Interface



2 Data Viewing

Lets load some data! RStudio has many built-in datasets available, and see the list of datasets type the following command:

```
> data()
```



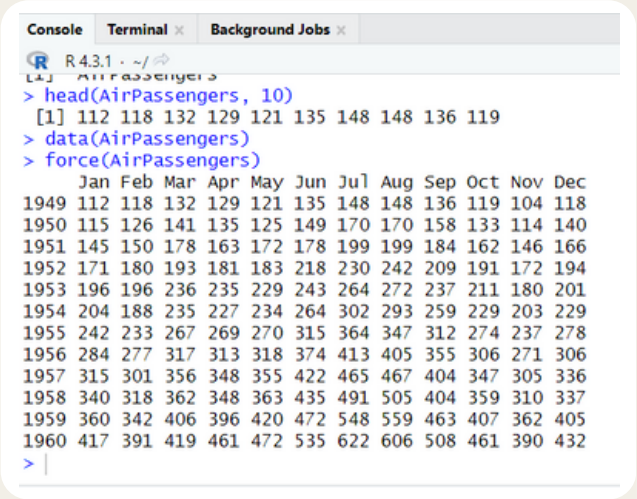
ATTGCCTTAGATR
ACTGATACORNER_

3 Select any dataset

```
> data(AirPassengers)
```

Now that you have loaded the AirPassengers dataset notice that the environment shows what type of data is there in the dataset.

Clicking on the dataset will expand it:



4 Try some commands yourself

Number of rows:

```
> nrows(AirPassengers)
```

Number of columns:

```
> ncol(AirPassengers)
```

Know more about the data:

```
> ?AirPassengers
```

Our Sponsors

BINFSOC

Thank you to all our partners and sponsors for their continuous support.

Quantium is a leading innovator in the world of advanced data analytics and artificial intelligence, where they use these tools to build powerful, transformative solutions for clients - individuals, organisations, and society.

Sponsor Us

If you would like to become a sponsor of BINFSOC, please reach out to sponsorships@unswbinfsoc.com

Contact us

IF YOU HAVE ANY COMMENTS or feedback regarding BINFsights, please write to us at binfsights@unswbinfsoc.com

We also encourage anyone to share with us anything you'd like us to take a look at, be it
a bioinformatics tool that you have made or find useful; or news in the bioinformatics world that you'd like to see written about in future issues.

TO VIEW PAST AND PRESENT issues of BINFsights, check out our website at unswbinfsoc.com/binfsights

Stay tuned on our Facebook page for updates regarding events and society news.

-- The BINFSOC Team

BINF sights.