Bionix Onboarding Documentation (Aurelia and Janya)

> 💡 Not sure what a word you read here means? If it's *italicized* make sure to check the "Key Terms" section!

What is Bionix?

BioNix is a software tool that helps researchers manage the software and computing environments they use for data analysis. BioNix aims to simplify the process of setting up and running complex data analysis *workflows* by providing a standard *framework* that can be used consistently across different computing environments. This allows researchers to get the same results no matter where they run their analyses.

As a result, this improves the *reproducibility* and *transparency* of research, making it easier for other researchers to validate and build upon previous work. Additionally, it saves time and improves the reproducibility of a researcher's analyses.

```
+---------------------+
| Data Analysis Script |
+---------+-----------+
          |
          v
+---------------------+
| BioNix Pure Functions|
+---------+-----------+
          |
          v
+---------------------+
|  Nix Package Manager |
+---------+-----------+
          |
          v
+---------------------+
```

https://github.com/PapenfussLab/bionix

```
| Computing Environment|
+----------------------+
```

https://github.com/victorwkb/BioNix-Doc

How does it work?

BioNix does this by organizing the different parts of a research workflow into small, self-contained units called "pure functions." These functions can be combined to create larger workflows that are easier to understand and reproduce. Moreover, Bionix is implemented as a lightweight library on top of the _Nix_ deployment system.

Setting Up Nix on Windows

1. Open up Windows Powershell / Command Prompt

    a. You can do so by clicking on the Windows Start button

    b. Search up "Windows Powershell" or "Command Prompt"

2. Type in `wsl --install`

    a. You may be prompted to enter a UNIX username and password

3. After installation is successful, type in

    ```
    curl -L https://nixos.org/nix/install | sh
    ```

4. Installation for Nix is now complete!

Setting Up Nix on Milton

https://github.com/WEHI-ResearchComputing/BioNix-GettingStarted-Milton
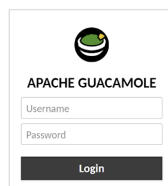
Milton is the Research Computing mascot and is visually represented by the blue monster. Milton represents WEHI's research computing *high-performance computer (HPC)* facilities and therefore using Milton simply means using WEHI's HPC.
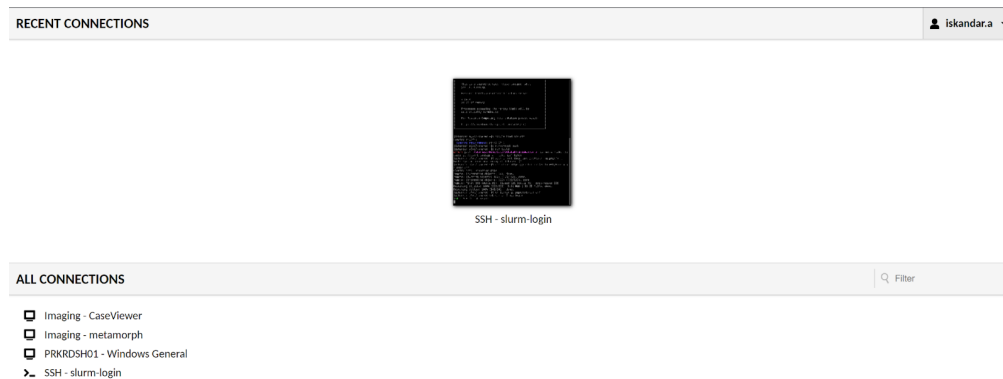
1. Go to RAP (https://rap.wehi.edu.au/global-protect/portal/portal.esp)
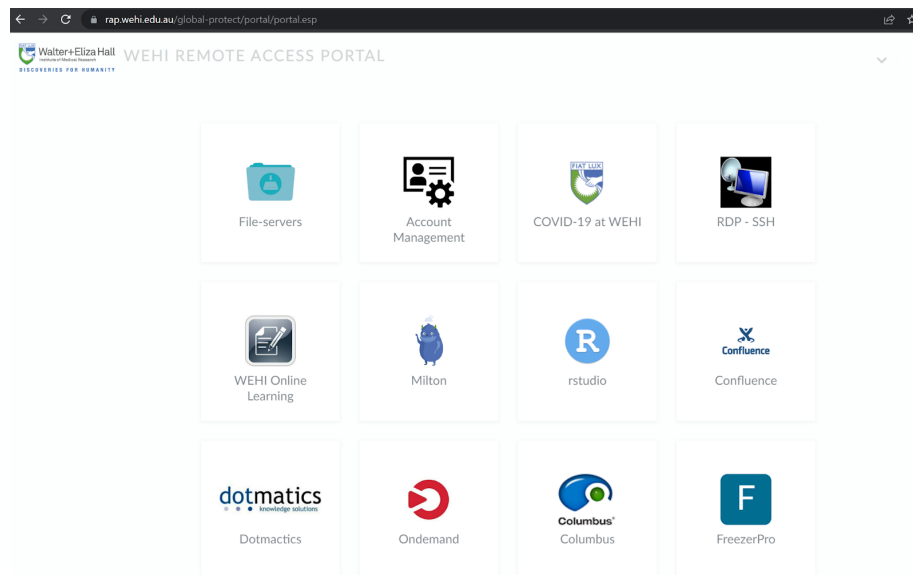
2. Click on "RDP - SSH"

3. This screen may show up:

    a. fill in your username that you received when setting up your WEHI email account

    b. your password would be the same one for your WEHI email

4. Click on `SSH - slurm-login`

5. Type `ssh vc7-shared`

6. Type `module load nix`

7. Type `nix-chroot bash`

8. Nix is set up!

Bionix Training

https://github.com/WEHI-ResearchComputing/BioNix-Training

https://github.com/WEHI-ResearchComputing/BioNix-qc-pipe

You can also look through this github to get a better understanding of how BioNix is used

https://github.com/jbedo/malaria-variant-calling

Nix Cheat Sheet

https://github.com/tazjin/nix-1p

https://learnxinyminutes.com/docs/nix/

VirusBreakEnd

!   This part is only if you are doing the VirusBreakEnd part of BioNix

- VirusBreakEnd article

  https://academic.oup.com/bioinformatics/article/37/19/3115/6273577

- VirusBreakEnd documentation can be found here

- VirusBreakEnd code implementation can be found <u>here</u>

Useful Links for VirusBreakEnd

- Bash Script

gridss/virusbreakend at master · PapenfussLab/gridss

GRIDSS: the Genomic Rearrangement IDentification Software Suite - gridss/virusbreakend at master · PapenfussLab/gridss

https://github.com/PapenfussLab/gridss/blob/master/scripts/virusbreakend

PapenfussLab/
gridss

GRIDSS: the Genomic Rearrangement IDentification Software Suite

15 Contributors   54 Issues   218 Stars   69 Forks

- VirusBreakEnd tools

gridss/VIRUSBreakend_Readme.md at master · PapenfussLab/gridss

GRIDSS: the Genomic Rearrangement IDentification Software Suite - gridss/VIRUSBreakend_Readme.md at master · PapenfussLab/gridss

https://github.com/PapenfussLab/gridss/blob/master/VIRUSBreakend_Readme.md

PapenfussLab/
gridss

GRIDSS: the Genomic Rearrangement IDentification Software Suite

15 Contributors   54 Issues   218 Stars   69 Forks

Glossary

Terms	Definitions
Workflow	A sequence of tasks or stages that are performed in a specific order to process data, run analyses, or generate output.
Framework	A structured way of organizing code or software components that provides a set of tools and guidelines for building applications or solving problems in a specific domain. (can be thought of as a blueprint that helps developers create software in a more efficient and consistent way)
Reproducibility	The ability to recreate the results of a scientific study or experiment using the same methods and data.

Terms	Definitions
Transparency	The act of being clear and upfront about the methods, data, and results of a study or experiment.
Nix	A package manager and build system that is used primarily in the UNIX and Linux operating systems.
High Performance Computing (HPC)	The use of advanced computer systems and software to solve complex problems that require significant computational resources.

References (links)

- What is Bionix? / Bionix Article

  - https://academic.oup.com/gigascience/article/9/11/giaa121/5987272?login=false

- Setting Up Nix on Windows

  - https://github.com/PapenfussLab/bionix

- Setting Up Nix on Milton

  - https://wehieduau.sharepoint.com/sites/rc2/SitePages/using-milton.aspx

  - https://github.com/WEHI-ResearchComputing/BioNix-GettingStarted-Milton