ViralWasm: a client-side user-friendly web application suite for viral genomics

Daniel Ji, Robert Aboukhalil, Niema Moshiri

BioSys Workshop 2024

Computational Tools in Epidemiology: Problem Statement

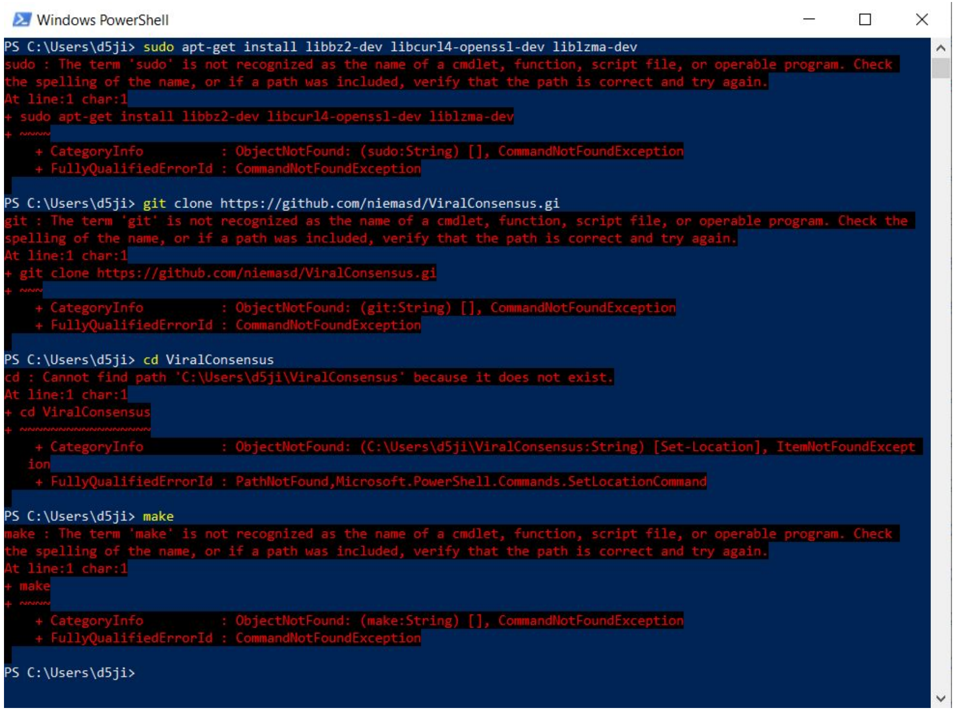
Computational Tools in Epidemiology: Problem Statement

- Goal: Fast and accurate diagnosis of patient viral sequences
    - Take raw sequence data and perform viral diagnosis and mutation evolution

Computational Tools in Epidemiology: Problem Statement

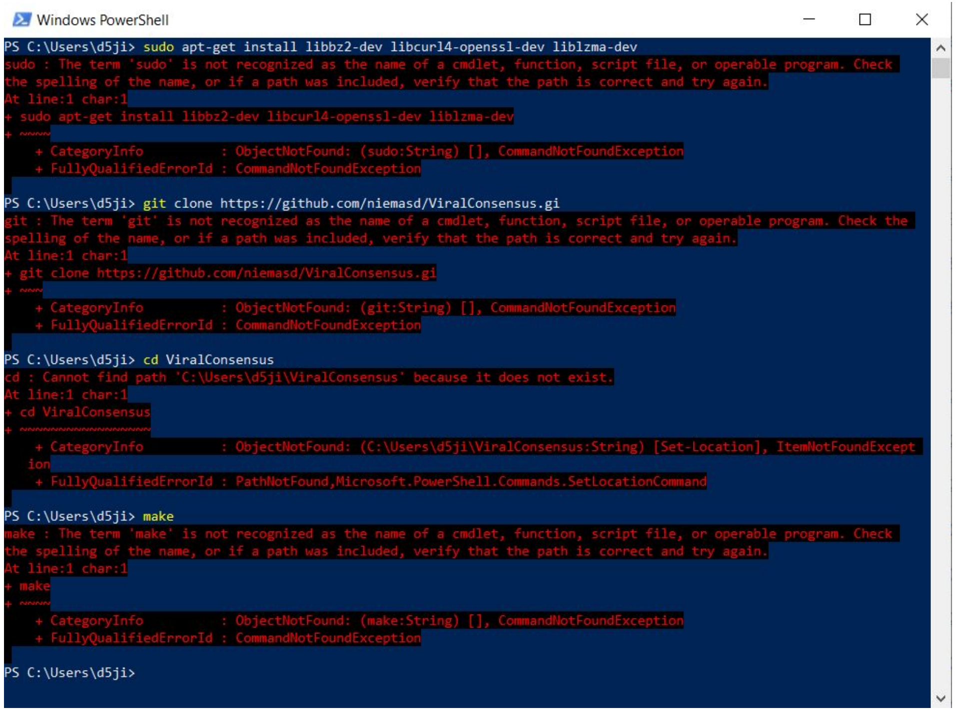
- Goal: Fast and accurate diagnosis of patient viral sequences
  - Take raw sequence data and perform viral diagnosis and mutation evolution
- Learning Curve
  - Requirement: Computer-science background
  - Technical know-how: Software installation, terminal usage

Computational Tools in Epidemiology: Problem Statement



```
PS C:\Users\d5ji> sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
sudo : The term 'sudo' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (sudo:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> git clone https://github.com/niemasd/ViralConsensus.gi
git : The term 'git' is not recognized as the name of a cmdlet, function, script file, or operable program. Check the
spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ git clone https://github.com/niemasd/ViralConsensus.gi
+ ~~~
    + CategoryInfo          : ObjectNotFound: (git:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> cd ViralConsensus
cd : Cannot find path 'C:\Users\d5ji\ViralConsensus' because it does not exist.
At line:1 char:1
+ cd ViralConsensus
+ ~~~~~~~~~~~~~~~~~
    + CategoryInfo          : ObjectNotFound: (C:\Users\d5ji\ViralConsensus:String) [Set-Location], ItemNotFoundExcept
   ion
    + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\d5ji> make
make : The term 'make' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ make
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (make:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji>
```

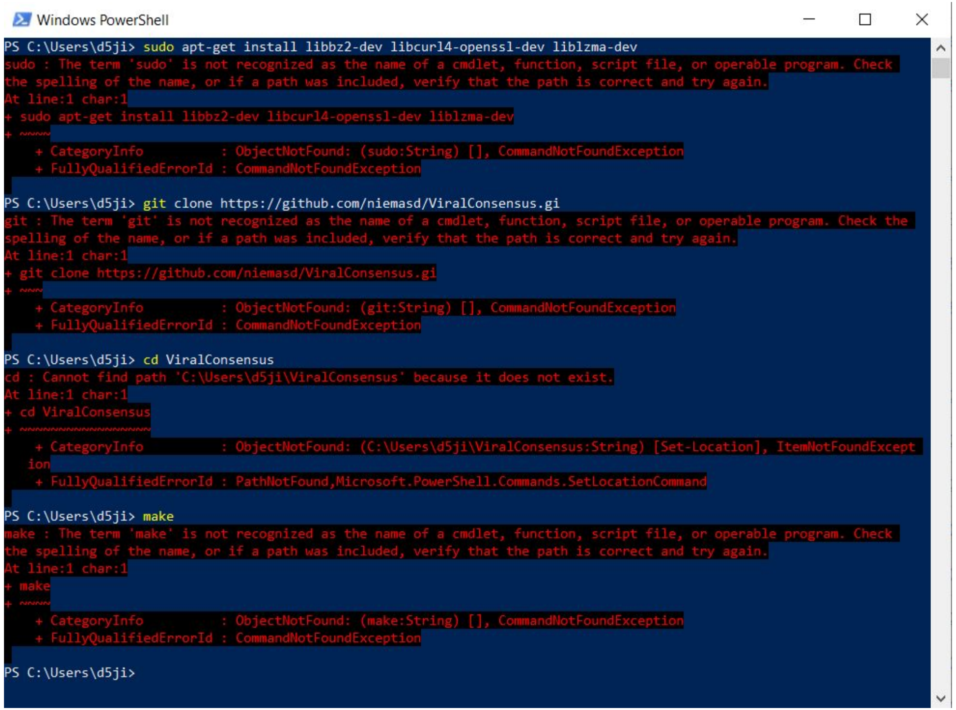
Computational Tools in Epidemiology: Problem Statement



```
PS C:\Users\d5ji> sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
sudo : The term 'sudo' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (sudo:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> git clone https://github.com/niemasd/ViralConsensus.gi
git : The term 'git' is not recognized as the name of a cmdlet, function, script file, or operable program. Check the
spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ git clone https://github.com/niemasd/ViralConsensus.gi
+ ~~~
    + CategoryInfo          : ObjectNotFound: (git:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> cd ViralConsensus
cd : Cannot find path 'C:\Users\d5ji\ViralConsensus' because it does not exist.
At line:1 char:1
+ cd ViralConsensus
+ ~~~~~~~~~~~~~~~~~
    + CategoryInfo          : ObjectNotFound: (C:\Users\d5ji\ViralConsensus:String) [Set-Location], ItemNotFoundExcept
   ion
    + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\d5ji> make
make : The term 'make' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ make
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (make:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji>
```

- Typically requires 3-4 molecular sequencing programs per job

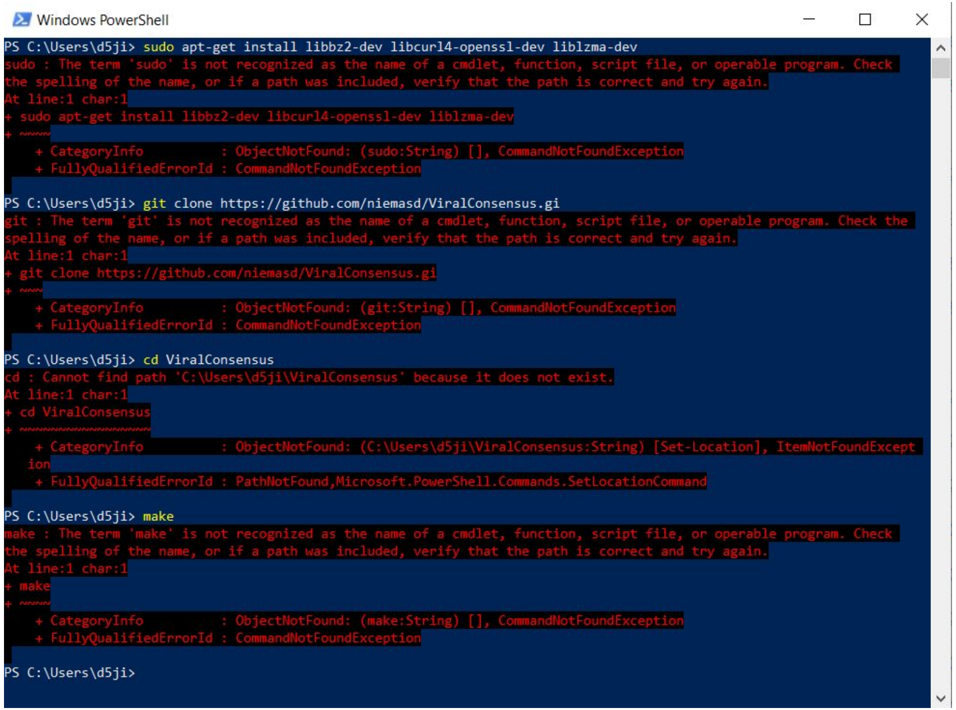
Computational Tools in Epidemiology: Problem Statement



```
PS C:\Users\d5ji> sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
sudo : The term 'sudo' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (sudo:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> git clone https://github.com/niemasd/ViralConsensus.gi
git : The term 'git' is not recognized as the name of a cmdlet, function, script file, or operable program. Check the
spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ git clone https://github.com/niemasd/ViralConsensus.gi
+ ~~~
    + CategoryInfo          : ObjectNotFound: (git:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> cd ViralConsensus
cd : Cannot find path 'C:\Users\d5ji\ViralConsensus' because it does not exist.
At line:1 char:1
+ cd ViralConsensus
+ ~~~~~~~~~~~~~~~~~
    + CategoryInfo          : ObjectNotFound: (C:\Users\d5ji\ViralConsensus:String) [Set-Location], ItemNotFoundExcept
   ion
    + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\d5ji> make
make : The term 'make' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ make
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (make:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji>
```

- Typically requires 3-4 molecular sequencing programs per job
- Tool selection: navigating choices is already difficult

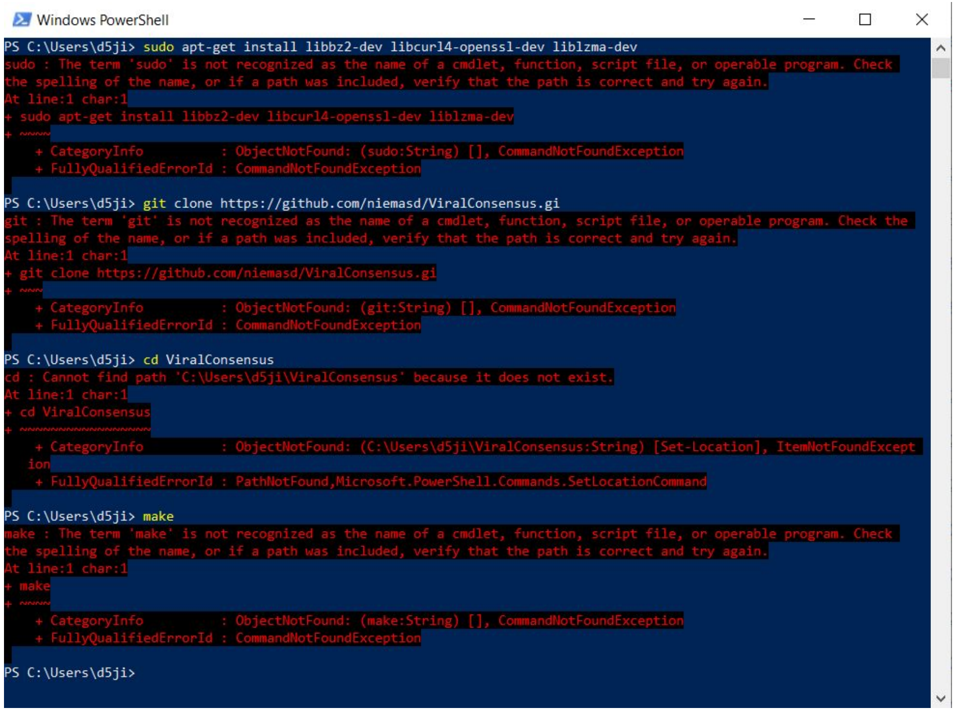
Computational Tools in Epidemiology: Problem Statement



```
PS C:\Users\d5ji> sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
sudo : The term 'sudo' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (sudo:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> git clone https://github.com/niemasd/ViralConsensus.gi
git : The term 'git' is not recognized as the name of a cmdlet, function, script file, or operable program. Check the
spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ git clone https://github.com/niemasd/ViralConsensus.gi
+ ~~~
    + CategoryInfo          : ObjectNotFound: (git:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> cd ViralConsensus
cd : Cannot find path 'C:\Users\d5ji\ViralConsensus' because it does not exist.
At line:1 char:1
+ cd ViralConsensus
+ ~~~~~~~~~~~~~~~~~
    + CategoryInfo          : ObjectNotFound: (C:\Users\d5ji\ViralConsensus:String) [Set-Location], ItemNotFoundExcept
   ion
    + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\d5ji> make
make : The term 'make' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ make
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (make:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji>
```

- Typically requires 3-4 molecular sequencing programs per job
- Tool selection: navigating choices is already difficult
- Tools requires dozens of dependencies and prerequisites

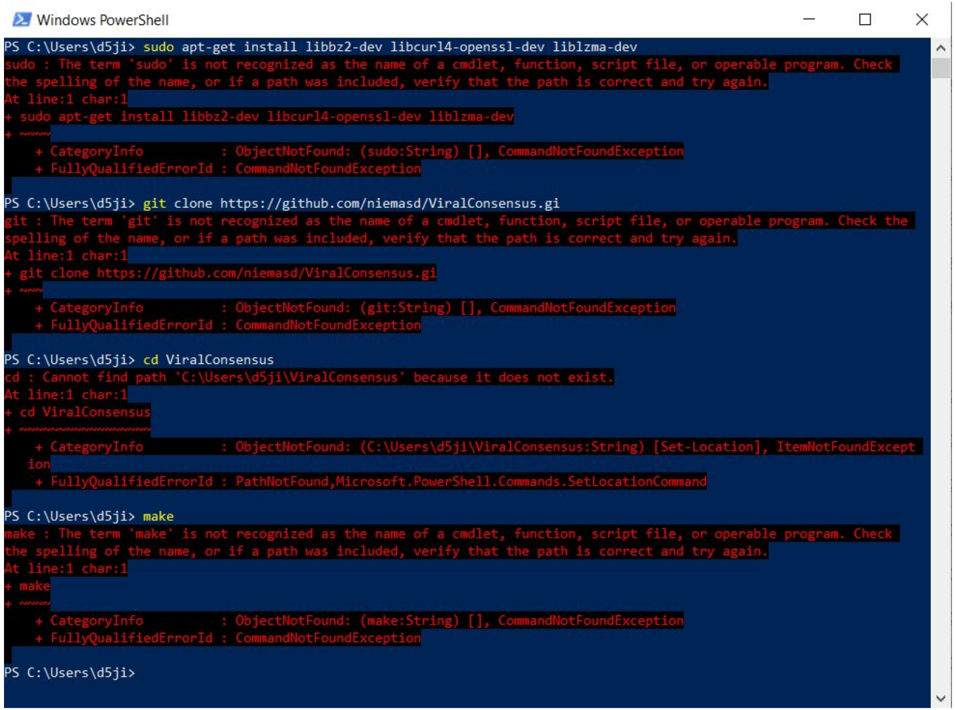
Computational Tools in Epidemiology: Problem Statement



```
PS C:\Users\d5ji> sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
sudo : The term 'sudo' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (sudo:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> git clone https://github.com/niemasd/ViralConsensus.gi
git : The term 'git' is not recognized as the name of a cmdlet, function, script file, or operable program. Check the
spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ git clone https://github.com/niemasd/ViralConsensus.gi
+ ~~~
    + CategoryInfo          : ObjectNotFound: (git:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> cd ViralConsensus
cd : Cannot find path 'C:\Users\d5ji\ViralConsensus' because it does not exist.
At line:1 char:1
+ cd ViralConsensus
+ ~~~~~~~~~~~~~~~~~
    + CategoryInfo          : ObjectNotFound: (C:\Users\d5ji\ViralConsensus:String) [Set-Location], ItemNotFoundExcept
   ion
    + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\d5ji> make
make : The term 'make' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ make
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (make:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji>
```

- Typically requires 3-4 molecular sequencing programs per job
- Tool selection: navigating choices is already difficult
- Tools requires dozens of dependencies and prerequisites
- Regular users: without admin access, installations can be nearly impossible

Computational Tools in Epidemiology: Problem Statement



```
PS C:\Users\d5ji> sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
sudo : The term 'sudo' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (sudo:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> git clone https://github.com/niemasd/ViralConsensus.gi
git : The term 'git' is not recognized as the name of a cmdlet, function, script file, or operable program. Check the
spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ git clone https://github.com/niemasd/ViralConsensus.gi
+ ~~~
    + CategoryInfo          : ObjectNotFound: (git:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji> cd ViralConsensus
cd : Cannot find path 'C:\Users\d5ji\ViralConsensus' because it does not exist.
At line:1 char:1
+ cd ViralConsensus
+ ~~~~~~~~~~~~~~~~~~
    + CategoryInfo          : ObjectNotFound: (C:\Users\d5ji\ViralConsensus:String) [Set-Location], ItemNotFoundExcept
   ion
    + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\d5ji> make
make : The term 'make' is not recognized as the name of a cmdlet, function, script file, or operable program. Check
the spelling of the name, or if a path was included, verify that the path is correct and try again.
At line:1 char:1
+ make
+ ~~~~
    + CategoryInfo          : ObjectNotFound: (make:String) [], CommandNotFoundException
    + FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\d5ji>
```

- Typically requires 3-4 molecular sequencing programs per job
- Tool selection: navigating choices is already difficult
- Tools requires dozens of dependencies and prerequisites
- Regular users: without admin access, installations can be nearly impossible
- Users have to worry more about successfully installing tools than using the tool itself

Problem Statement (Cont.)

- Data and Server Concerns

Problem Statement (Cont.)

- Data and Server Concerns
    - Data Security: Sending data to third-party servers

Problem Statement (Cont.)

- Data and Server Concerns
  - Data Security: Sending data to third-party servers
  - Hosting Costs:
    - Always-on servers: High cost
    - Infrequent-use (cold-boot) servers: Long boot times

Problem Statement (Cont.)

- Data and Server Concerns
  - Data Security: Sending data to third-party servers
  - Hosting Costs:
    - Always-on servers: High cost
    - Infrequent-use (cold-boot) servers: Long boot times
- Remote Area Limitations

Our Solution: ViralWasm

Our Solution: ViralWasm

ViralWasm:

- Web-based end-to-end molecular epidemiology pipelines and standalone tools.
- Compile from source code: no need to rewrite tools in web-based languages.

Our Solution: ViralWasm

ViralWasm:

- Web-based end-to-end molecular epidemiology pipelines and standalone tools.
- Compile from source code: no need to rewrite tools in web-based languages.

Complete Client-Side Functionality

- Ports tools to the web for client-side, serverless operation.
- Instead of C/C++/Python to machine code: compiles source to WebAssembly.
- Instant tool operation in the browser once deployed.

Our Solution: ViralWasm

ViralWasm:

- Web-based end-to-end molecular epidemiology pipelines and standalone tools.
- Compile from source code: no need to rewrite tools in web-based languages.

Complete Client-Side Functionality

- Ports tools to the web for client-side, serverless operation.
- Instead of C/C++/Python to machine code: compiles source to WebAssembly.
- Instant tool operation in the browser once deployed.

User Experience

- Clean, intuitive interfaces for data upload and parameter adjustments.

Our Solution: ViralWasm

ViralWasm:
- Web-based end-to-end molecular epidemiology pipelines and standalone tools.
- Compile from source code: no need to rewrite tools in web-based languages.

Complete Client-Side Functionality
- Ports tools to the web for client-side, serverless operation.
- Instead of C/C++/Python to machine code: compiles source to WebAssembly.
- Instant tool operation in the browser once deployed.

User Experience
- Clean, intuitive interfaces for data upload and parameter adjustments.

Offline Capabilities
- Entire website + tool downloadable for offline use, just through a few clicks.

Technical Approach

Technical Approach

Two primary pipelines

- ViralWasm-Epi: multi-sequence alignment, molecular clustering, phylogenetic tree inference, and tree rooting and dating
    - Pipeline software: ViralMSA, minimap2, tn93, FastTree 2, Seqtk, LSD2

Technical Approach

Two primary pipelines

- ViralWasm-Epi: multi-sequence alignment, molecular clustering, phylogenetic tree inference, and tree rooting and dating
    - Pipeline software: ViralMSA, minimap2, tn93, FastTree 2, Seqtk, LSD2
- ViralWasm-Consensus: sequence trimming, consensus genome generation
    - Pipeline software: minimap2, fastp, ViralConsensus

Technical Approach

Two primary pipelines

- ViralWasm-Epi: multi-sequence alignment, molecular clustering, phylogenetic tree inference, and tree rooting and dating
  - Pipeline software: ViralMSA, minimap2, tn93, FastTree 2, Seqtk, LSD2
- ViralWasm-Consensus: sequence trimming, consensus genome generation
  - Pipeline software: minimap2, fastp, ViralConsensus

Web technologies used: Emscripten, Biowasm, Pyodide, React.js, and Bootstrap

Benchmarks

Benchmarks

- Used Playwright testing framework for accuracy and performance benchmarking of ViralWasm web app
  - Simulated a page browser and automated user input and tool run

Benchmarks

- Used Playwright testing framework for accuracy and performance benchmarking of ViralWasm web app
    - Simulated a page browser and automated user input and tool run
- Benchmarks themselves ran automatically via Github Actions CI/CD workflow

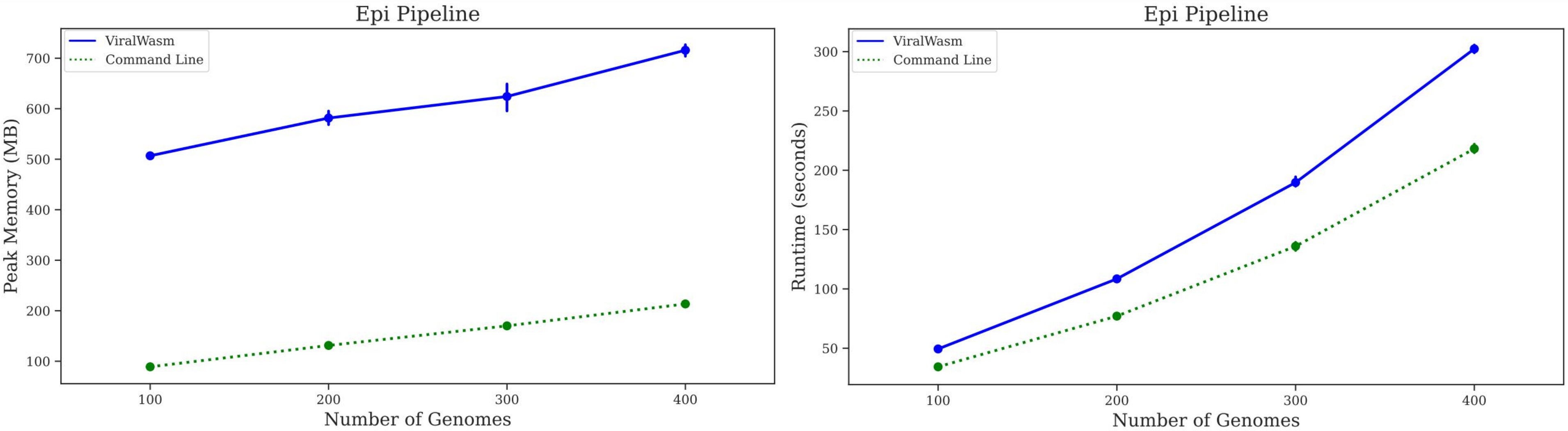
Benchmarks

- Used Playwright testing framework for accuracy and performance benchmarking of ViralWasm web app
  - Simulated a page browser and automated user input and tool run
- Benchmarks themselves ran automatically via Github Actions CI/CD workflow
- Ran on Ubuntu 22.04 machine, 2 processor cores on Intel Xeon CPU E5-2673 v3 @ 2.40GHz with 7GiB RAM

Benchmarks

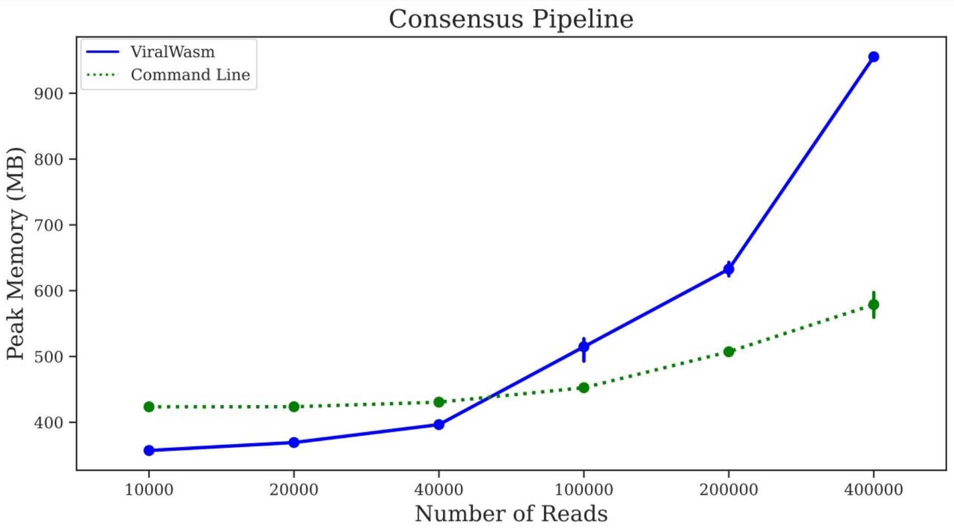
- Used Playwright testing framework for accuracy and performance benchmarking of ViralWasm web app
    - Simulated a page browser and automated user input and tool run
- Benchmarks themselves ran automatically via Github Actions CI/CD workflow
- Ran on Ubuntu 22.04 machine, 2 processor cores on Intel Xeon CPU E5-2673 v3 @ 2.40GHz with 7GiB RAM
- Web vs. CLI benchmarks (full-pipeline runs): web scale reasonably in terms of memory and runtime

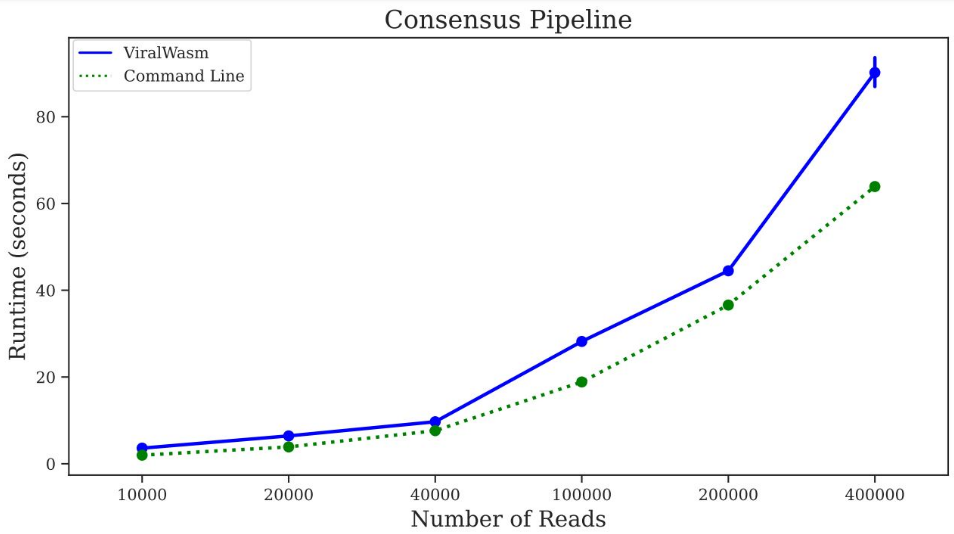
Benchmarks: ViralWasm-Epi



MSA, molecular clustering, phylogenetic tree inference, and tree rooting and dating

Benchmarks: ViralWasm-Consensus





sequence trimming, consensus genome generation

Results & Moving Forward

Results & Moving Forward

- A relative slowdown compared to CLI tool counterparts, by 2-3x

Results & Moving Forward

- A relative slowdown compared to CLI tool counterparts, by 2-3x
    - Tested datasets are large for individual public health departments

Results & Moving Forward

- A relative slowdown compared to CLI tool counterparts, by 2-3x
    - Tested datasets are large for individual public health departments
    - Job times are still reasonable for laptops and even tablets
        - Often less than a minute, at worst ~5 min.

Results & Moving Forward

- A relative slowdown compared to CLI tool counterparts, by 2-3x
    - Tested datasets are large for individual public health departments
    - Job times are still reasonable for laptops and even tablets
        - Often less than a minute, at worst ~5 min.
    - A worthwhile trade-off

Results & Moving Forward

- A relative slowdown compared to CLI tool counterparts, by 2-3x
    - Tested datasets are large for individual public health departments
    - Job times are still reasonable for laptops and even tablets
        - Often less than a minute, at worst ~5 min.
    - A worthwhile trade-off

- Moving Forward

Results & Moving Forward

- A relative slowdown compared to CLI tool counterparts, by 2-3x
    - Tested datasets are large for individual public health departments
    - Job times are still reasonable for laptops and even tablets
        - Often less than a minute, at worst ~5 min.
    - A worthwhile trade-off

- Moving Forward
    - Scaling to larger datasets

Results & Moving Forward

- A relative slowdown compared to CLI tool counterparts, by 2-3x
  - Tested datasets are large for individual public health departments
  - Job times are still reasonable for laptops and even tablets
    - Often less than a minute, at worst ~5 min.
  - A worthwhile trade-off

- Moving Forward
  - Scaling to larger datasets
  - Reducing web app / WebAssembly overhead
    - Utilizing WebGPU

Results & Moving Forward

- A relative slowdown compared to CLI tool counterparts, by 2-3x
  - Tested datasets are large for individual public health departments
  - Job times are still reasonable for laptops and even tablets
    - Often less than a minute, at worst ~5 min.
  - A worthwhile trade-off

- Moving Forward
  - Scaling to larger datasets
  - Reducing web app / WebAssembly overhead
    - Utilizing WebGPU
  - Automating custom tool Wasm web app creation

Live Demo

niema-lab.github.io/ViralWasm

Thank You! Questions?

daji@ucsd.edu | robert.aboukhalil@gmail.com | niema@ucsd.edu