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

Daniel Ji, Robert Aboukhalil, Niema Moshiri BioSys Workshop 2024

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

- 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

🔀 Windows PowerShell				_		×
sudo : The term 'sudo' i	t-get install libbz2-dev libcurl4- not recognized as the name of a c or if a path was included, verify	mdlet, function, script file,		ogram.	. Check	^
+ sudo apt-get install 1:	bbz2-dev libcurl4-openssl-dev libl					
+ CategoryInfo + FullyQualifiedErro	: ObjectNotFound: (sudo:String) Id : CommandNotFoundException	[], CommandNotFoundException				
PS C:\Users\d5ji> git clo	ne https://github.com/niemasd/Vira	lConsensus.gi				
spelling of the name, or At line:1 char:1	ot recognized as the name of a cmd if a path was included, verify tha	let, function, script file, or t the path is correct and try	operable progragain.	ram. (Check th	e
+ git clone https://gith	b.com/niemasd/ViralConsensus.gi					
+ CategoryInfo + FullyQualifiedErro	: ObjectNotFound: (git:String) Id : CommandNotFoundException	[], CommandNotFoundException				
PS C:\Users\d5ji> cd Vira						
cd : Cannot find path 'C	\Users\d5ji\ViralConsensus' becaus	e it does not exist.				
At line:1 char:1 + cd ViralConsensus						
+ CategoryInfo	: ObjectNotFound: (C:\Users\d5j	i\ViralConsensus:String) [Set-	Location], Iter	NotFo	oundExce	pt
	Id : PathNotFound,Microsoft.PowerS	hell.Commands.SetLocationComma	and			
the spelling of the name	not recognized as the name of a c or if a path was included, verify			ogram.	. Check	
At line:1 char:1 + make + www						
+ CategoryInfo + FullyQualifiedErro	: ObjectNotFound: (make:String) Id : CommandNotFoundException	[], CommandNotFoundException				
PS C:\Users\d5ji>						

2 Windows PowerShell	-		×
PS C:\Users\d5ji> sudo apt-get install libb22-dev libcurl4-openssl-dev liblzma-dev sudo : The term 'sudo' is not recognized as the name of a cmdlet, function, script file, or operable 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 libb22-dev libcurl4-openssl-dev liblzma-dev	program	. Check	
<pre>+ ///// * CategoryInfo : ObjectNotFound: (sudo:String) [], CommandNotFoundException + FullyQualifiedErrorId : CommandNotFoundException</pre>			
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 pr spelling of the name, or if a path was included, verify that the path is correct and try again. At line:1 char:1	ogram.	Check th	ie
<pre># ist clone https://github.com/niemasd/ViralConsensus.gi # use is a close is close is a close</pre>			
PS C:\Users\d5ji≻ cd ViralConsensus d : Cannot find path 'C:\Users\d5ji\ViralConsensus' because it does not exist. At line:1 char:1			
<pre>+ cd ViralConsensus + categoryInfo : ObjectNotFound: (C:\Users\d5ji\ViralConsensus:String) [Set-Location], I ion + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand</pre>	temNotF	oundExce	pt
PS C:\Users\d5ji> make make : The term "make' is not percentized as the name of a redlat function, script file, or openable	program	. Check	
<pre>wake + concerning in the construction of the construction of</pre>			
+ FullyQualifiedErrorId : CommandNotFoundException PS C:\Users\d5ji>			

 Typically requires 3-4 molecular sequencing programs per job

🔁 Windows PowerShell	-		×
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 operab 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		m. Check	^
+ sudo apt-get install 11052-dev 110curl4-openssi-dev 1101zma-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 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	program.	Check t	ne
 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] ion + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand	, ItemNot	FoundExce	ept
PS C:\Users\d5ji> make make : The term 'make' is not recognized as the name of a cmdlet, function, script file, or operab 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	le progra	m. Check	
<pre>+ categoryInfo : ObjectNotFound: (make:String) [], CommandNotFoundException + FullyQualifiedErrorId : CommandNotFoundException</pre>			
PS C:\Users\d5ji>			

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

Windows PowerShell	-		×
<pre>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 operabl 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</pre>		n. Check	^
<pre>+ CategoryInfo : ObjectNotFound: (sudo:String) [], CommandNotFoundException + FullyQualifiedErrorId : CommandNotFoundException 25 C:\Users\d5ji> git clone https://github.com/niemasd/ViralConsensus.gi</pre>			
<pre>S C: OSEPS(05)12 git Caome nttp://github.com/niemas0/viralconsensus.gi git : The term 'git' is not recognized as the name of a cmdlet, function, script file, or operable spelling of the name, or if a path was included, verify that the path is correct and try again. t line:1 char:1</pre>	program.	Check th	e
+ 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], ion + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand	ItemNot	FoundExce	ot
<pre>PS C:\Users\d5ji> make make : The term "make' is not recognized as the name of a cmdlet, function, script file, or operabl 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 + make</pre>	e progran	n. Check	
<pre>+ CategoryInfo : ObjectNotFound: (make:String) [], CommandNotFoundException + FullyQualifiedErrorId : CommandNotFoundException PS C:\Users\d5ji></pre>			

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

Windows PowerShell						-		×
PS C:\Users\d5ji> sudo apt sudo : The term 'sudo' is the spelling of the name,	-get install libb not recognized as or if a path was	<pre>z2-dev libcurl4-c the name of a cr included, verify</pre>	openssl-dev li mdlet, functio that the path	blzma-dev n, script file, o is correct and t	r operable ry again.	program	. Check	^
At line:1 char:1 + sudo apt-get install lib + ~~~~								
+ CategoryInfo + FullyQualifiedErrorI	: ObjectNotFoun d : CommandNotFou	d: (sudo:String) ndException	[], CommandNo	tFoundException				
PS C:\Users\d5ji> git clon	e https://github.	com/niemasd/Vira	1Consensus gi					
git : The term 'git' is no				script file, or	operable pr	ogram.	Check the	
spelling of the name, or i At line:1 char:1	f a path was incl	uded, verify that	t the path is	correct and try a	gain.			
+ git clone https://github	.com/niemasd/Vira	lConsensus.gi						
+ CategoryInfo + FullyQualifiedErrorI	: ObjectNotFound d : CommandNotFound	d: (git:String) ndException	[], CommandNot	FoundException				
PS C:\Users\d5ji> cd Viral	Consensus							
cd : Cannot find path 'C:\		onsensus' because	e it does not	exist.				
+ cd ViralConsensus								
+ CategoryInfo	: ObjectNotFoun	d: (C:\Users\d5ji	i\ViralConsens	us:String) [Set-L	ocation], I	temNotF	oundExce	9t
ion + FullyQualifiedErrorI								
+ FullyQualifiedErrori	d : PathNotFound,	Milcrosoft.PowerSr	nell.Commands.	SetLocationComman	a			
PS C:\Users\d5ji> make								
make : The term 'make' is	not recognized as	the name of a cr	mdlet, functio	n, script file, o	r operable	program	. Check	
the spelling of the name,	or if a path was	included, verify		is correct and t				
+ make								
+ CategoryInfo + FullyQualifiedErrorI	: ObjectNotFound d : CommandNotFou	d: (make:String) ndException	[], CommandNo	tFoundException				
PS C:\Users\d5ii>								

- 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

2 Windows PowerShell	-		×
<pre>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 p the spelling of the name, or if a path was included, verify that the path is correct and try again. At line:1 chan:1 + sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev + sudo apt-get install libbz2-dev libcurl4-openssl-dev liblzma-dev + CategoryInfo : ObjectNotFound: (sudo:String) [], CommandNotFoundException + FullyQualifiedErrorId : CommandNotFoundException</pre>	rogram	. Check	^
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 pro 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 + wave + CategoryInfo : ObjectNotFound: (git:String) [], CommandNotFoundException + FullyQualifiedErrorId : CommandNotFoundException	gram. I	Check the	
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], It ion + FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand	emNotF	oundExcer	t
PS C:\Users\d5ji> make make : The term 'make' is not recognized as the name of a cmdlet, function, script file, or operable p 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	rogram	. Check	
<pre>+ categoryInfo : ObjectNotFound: (make:String) [], CommandNotFoundException + FullyQualifiedErrorId : CommandNotFoundException</pre>			

- 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

• Data and Server Concerns

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

- 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

- 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

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.

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.

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.

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.

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

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

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

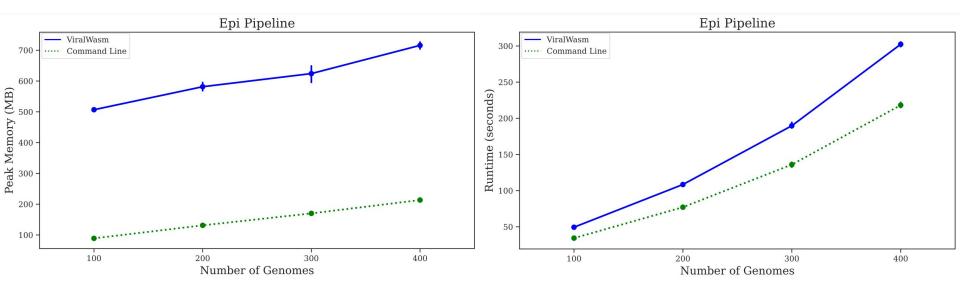
- Used Playwright testing framework for accuracy and performance benchmarking of ViralWasm web app
 - Simulated a page browser and automated user input and tool run

- 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

- 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

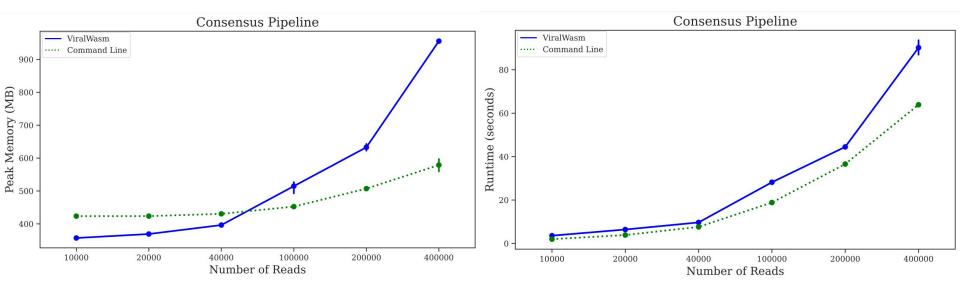
- Used Playwright testing framework for accuracy and performance benchmarking of ViralWasm web app
 - Simulated a page browser and automated user input and tool run
- Benchmarks themselves ran automatically via Github Actions CI/CD workflow
- Ran on Ubuntu 22.04 machine, 2 processor cores on Intel Xeon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

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

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

- 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 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

- 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

- 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

- 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

- 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