

TARDiS

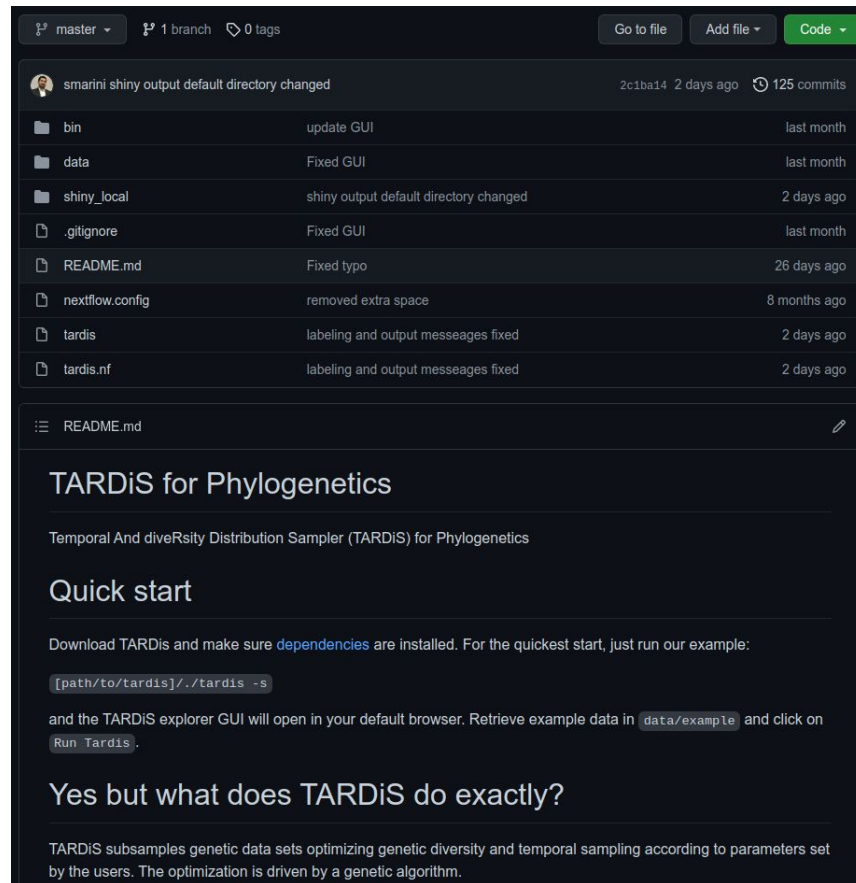
Temporal And diveRsity Distribution
Sampler for Phylogenetics

Simone Marini, University of Florida

TARDiS

- Method to **subsample** biosequences
 - In: many sequences
 - Out: fewer sequences
- Optimizing:
 - Genetic diversity
 - Temporal distribution
- Based on genetic algorithm
 - R
 - Python
 - Nextflow
- Usage
 - GUI
 - Command line, job allocation on HPC

<https://github.com/smarini/tardis-phylogenetics>



The screenshot shows the GitHub repository for TARDiS. The top bar indicates the 'master' branch with 1 branch and 0 tags. The repository name is 'smarini shiny output default directory changed' with a commit hash of '2c1ba14' from 2 days ago and 125 commits. The file list includes:

File	Commit Message	Time
bin	update GUI	last month
data	Fixed GUI	last month
shiny_local	shiny output default directory changed	2 days ago
.gitignore	Fixed GUI	last month
README.md	Fixed typo	26 days ago
nextflow.config	removed extra space	8 months ago
tardis	labeling and output messages fixed	2 days ago
tardis.nf	labeling and output messages fixed	2 days ago

The README.md file is open, showing the title 'TARDiS for Phylogenetics' and the subtitle 'Temporal And diveRsity Distribution Sampler (TARDiS) for Phylogenetics'. It includes a 'Quick start' section with instructions to download TARDiS and install dependencies, followed by a code block for running the command line example: `[path/to/tardis]/. /tardis -s`. It also mentions that the TARDiS explorer GUI will open in the default browser and provides a link to the example data in `data/example`. The README concludes with the question 'Yes but what does TARDiS do exactly?' and a brief description of its function: 'TARDiS subsamples genetic data sets optimizing genetic diversity and temporal sampling according to parameters set by the users. The optimization is driven by a genetic algorithm.'

WHY TARDiS

- Too much information
 - GISAID 3+ million viruses in late Aug 2021
- Bias! Genomes are not collected uniformly (of course)
 - Geographical
 - Temporal
 - ...



	Virus name	Passage date	Accession ID	Collection date	Submission date
<input type="checkbox"/>	hCoV-19/USA/TN-VUMC-000014/2020	Original	EPI_ISL_3730707	2020-07-12	2021-08-27
<input type="checkbox"/>	hCoV-19/USA/TN-VUMC-000034/2020	Original	EPI_ISL_3730706	2020-07-12	2021-08-27
<input type="checkbox"/>	hCoV-19/USA/PR-CDC-S548/2021	Original	EPI_ISL_3730705	2021-06-19	2021-08-27
<input type="checkbox"/>	hCoV-19/USA/PR-CDC-S547/2021	Original	EPI_ISL_3730704	2021-06-18	2021-08-27
<input type="checkbox"/>	hCoV-19/Angola/CERI-KRISP-K021670/2020	Original	EPI_ISL_3730486	2021-07-06	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021564	Original	EPI_ISL_3730485	2021-07-26	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021563	Original	EPI_ISL_3730484	2021-07-26	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021560	Original	EPI_ISL_3730483	2021-07-26	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021556	Original	EPI_ISL_3730482	2021-07-05	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021554	Original	EPI_ISL_3730481	2021-06-28	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021553	Original	EPI_ISL_3730480	2021-06-28	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021551	Original	EPI_ISL_3730479	2021-06-28	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021545	Original	EPI_ISL_3730478	2021-07-25	2021-08-27
<input type="checkbox"/>	hCoV-19/Zimbabwe/CERI-KRISP-K021500	Original	EPI_ISL_3730477	2021-07-15	2021-08-27
<input type="checkbox"/>	hCoV-19/Angola/CERI-KRISP-K021636/2020	Original	EPI_ISL_3730476	2021-06-16	2021-08-27
<input type="checkbox"/>	hCoV-19/Angola/CERI-KRISP-K021622/2020	Original	EPI_ISL_3730475	2021-06-15	2021-08-27
<input type="checkbox"/>	hCoV-19/Angola/CERI-KRISP-K021621/2020	Original	EPI_ISL_3730474	2021-06-15	2021-08-27
<input type="checkbox"/>	hCoV-19/Angola/CERI-KRISP-K021631/2020	Original	EPI_ISL_3730473	2021-06-15	2021-08-27

Total: 3,087,220 viruses

WAIT WHY DON'T WE BRUTE FORCE THIS?

- Select 5 sequences from 100 $\rightarrow 75,287,520$
- Select 10 sequences from 1000 $\rightarrow 2.6 \times 10^{23}$
- Select 100 sequences from 1000 $\rightarrow 6.4 \times 10^{139}$
- Select 150 sequences from 5000 $\rightarrow 1.3 \times 10^{292}$
- Recent TARDiS runs
 - Selected 150 sequences from 50000 $\rightarrow 5.61 \times 10^{448}$
 - Selected 1000 sequences from 15000 $\rightarrow 4.9 \times 10^{1593}$

Atoms in the universe

$\sim 10^{80}$



NGC4651, Jay GaBany, cc-by-sa-3.0

./tardis -s

Tardis

Sequences per individual <input type="text" value="5"/>	Fraction of random individuals <input type="text" value="0.05"/>	Genetic diversity weight <input type="text" value="0.5"/>
Individuals per generation <input type="text" value="100"/>	Fraction of evolved individuals <input type="text" value="0.9"/>	Temporal diversity weight: <input type="text" value="0.5"/>
Number of cores <input type="text" value="1"/>	Fraction of elite individuals <input type="text" value="0.05"/>	Optimize genetic diversity <div>Max diversity ▼</div>
	Generations <input type="text" value="10"/>	<div>Jukes-Cantor distance</div>

Distance file (csv, rds) <div>Browse... <input type="text" value="..."/></div>	Sequence file (fasta) <div>Browse... <input type="text" value="..."/></div>	Metadata file (csv) <div>Browse... <input type="text" value="..."/></div>
--	---	---

Dataset name <input type="text" value="example.dataset"/>	Select output directory <div><input type="text" value="..shiny_output"/> ...</div>
---	--

Tardis will run on individuals per generation, handled by 1 core(s), for 10 generations, counted as 0 to 9
The total combination of possible subsamples is not yet calculated (Please load metadata file)

Run Tardis

./tardis -s

Tardis

Sequences per individual	Fraction of random individuals	Genetic diversity weight
<input type="text" value="5"/>	<input type="text" value="0.05"/>	<input type="text" value="0.5"/>
Individuals per generation	Fraction of evolved individuals	Temporal diversity weight:
<input type="text" value="100"/>	<input type="text" value="0.9"/>	<input type="text" value="0.5"/>
Number of cores	Fraction of elite individuals	Optimize genetic diversity
<input type="text" value="1"/>	<input type="text" value="0.05"/>	<div>Max diversity ▼</div>
Generations		<div>Jukes-Cantor distance</div>
<input type="text" value="10"/>		

Distance file (csv, rds)

Sequence file (fasta)

Metadata file (csv)

Dataset name

Select output directory

../shiny_output

Tardis will run on individuals per generation, handled by 1 core(s), for 10 generations, counted as 0 to 9
The total combination of possible subsamples is not yet calculated (Please load metadata file)

Run Tardis

INPUT

Sequences data set

- fasta
- DNA
- **Aligned**

```
1 >Q1
2 GAAAAAAAAAAAAAAAAAACTTTAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGACTCACGCAGTATAATTA
3 ATAATAATTACTGTCGTTGACAGGACACGAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTGTTGCA
4 GCCGATCATCAGCACATCTAGGTTTGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAG
5 AAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGA
6 GGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAAC
7 TTGAACAGCCCTATGTGTTTCATCAAACGTTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCA
8 GAACTCGAAGGCATTTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGTGGGCGAAATACCAAGTGGC
9 TTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTG
10 CCCCCCGCGACGAGCTTGGCAGTGATCGAGGGGAAGATTTTCAAGATTATGGAACACTAAACATAGCAGTGGTGTTACC
11 >Q2
12 GAAATGTTCTCTAAACGAACTTTAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGACTCACGCAGTATAATTA
13 ATAATAATTACTGTCGTTGACAGGACACGAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTGTTGCA
14 GCCGATCATCAGCACATCTAGGTTTGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAG
15 AAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGA
16 GGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAAC
17 TTGAACAGCCCTATGTGTTTCATCAAACGTTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCA
18 GAACTCGAAGGCATTTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGTGGGCGAAATACCAAGTGGC
19 TTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTG
20 TTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTG
21 >Q3
22 GATCTGTTCTCTAAACGAACTTTAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGACTCACGCAGTATAATTA
23 ATAATAATTACTGTCGTTGACAGGACACGAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTGTTGCA
24 GCCGATCATCAGCACATCTAGGTTTGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAG
25 AAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGA
26 GGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAAC
27 TTGAACAGCCCTATGTGTTTCATCAAACGTTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCA
28 GAACTCGAAGGCATTTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGTGGGCGAAATACCAAGTGGC
```


INPUT

Distance matrix

- csv | rds (R)
- colnames = rownames = aligned sequence fasta header
- Distance
 - measure of similarity $[0, \infty]$
 - identical seqs \rightarrow dist 0
 - Hamming, Chebyshev, Patristic, Blast E-value, ...
 - Default: Jukes-Cantor

```

1  Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9,Q10,Q11,Q12,Q13,Q14,Q15,Q16,Q17,Q
2  Q1,0,0.110537629074506,0.025426163756761,0.148499527134888,0.
3  Q2,0,0.110537629074506,0,0.100942615479573,0.0355516791709379,0
4  Q3,0.025426163756761,0.100942615479573,0,0.128453790455527,0.
5  Q4,0.148499527134888,0.0355516791709379,0.128453790455527,0,0
6  Q5,0.0759403002788797,0.138409688862049,0.0487743527031409,0.
7  Q6,0.0399333858891092,0.0993553132997904,0.0140190997591144,0
8  Q7,0.0713437588228497,0.133415219794567,0.044340842255848,0.0
9  Q8,0.0577207808520962,0.136741167595466,0.0311954229741278,0.
10 Q9,0.0355516791709379,0.110537629074506,0.00978578654435664,0
11 Q10,0.0547267648999432,0.14512114304406,0.0399333858891092,0.
12 Q11,0.094613463995175,0.158726958347047,0.0667752167777981,0.
13 Q12,0.016854641889044,0.115381564104064,0.0283052459871352,0.
14 Q13,0.0297489422943236,0.0977713633974327,0.012605338737286,0
15 Q14,0.118628322421254,0.204731943796564,0.112148800478223,0.2
16 Q15,0.0899014066105536,0.15359580948451,0.0622343350962685,0.
17 Q16,0.0759403002788797,0.138409688862049,0.0487743527031409,0
18 Q17,0.0399333858891092,0.0993553132997904,0.0140190997591144,
19 Q18,0.0713437588228497,0.133415219794567,0.044340842255848,0.
20 Q19,0.074404987562586,0.136741167595466,0.0502580327122202,0.
21 Q20,0.0487743527031409,0.108929911417067,0.0225580914279585,0
22 Q21,0.0682949741367225,0.143437640653692,0.05323402270939357,0
23 Q22,0.14512114304406,0.221337788204474,0.15381564104064,0.18

```


INPUT

Metadata

- csv, 1 seq = 1 line
- Mandatory columns
 - “Accession.ID” = fasta header
 - “Collection.date” = dd/mm/yyyy
 - Any other column is ignored

```
1 Accession.ID,Collection.date,Country|
2 Q1,27/02/2020,Norway
3 Q2,27/02/2020,Norway
4 Q3,01/03/2020,Norway
5 Q4,01/03/2020,Norway
6 Q5,02/03/2020,Norway
7 Q6,29/02/2020,Norway
8 Q7,29/02/2020,Norway
9 Q8,10/03/2020,Norway
10 Q9,18/03/2020,Norway
11 Q10,28/02/2020,Norway
12 Q11,28/02/2020,Norway
13 Q12,17/03/2020,Norway
14 Q13,09/03/2020,Norway
15 Q14,16/03/2020,Norway
16 Q15,17/03/2020,Norway
17 Q16,14/03/2020,Norway
18 Q17,10/03/2020,Norway
19 Q18,10/03/2020,Norway
20 Q19,11/03/2020,Norway
21 Q20,09/03/2020,Norway
22 Q21,10/03/2020,Norway
23 Q22,17/04/2020,Norway
```

How many cores
do we use?

`./tardis -s`

Load metadata file

Load sequence file

Load distance file

How do we call the
output?

Tardis

Sequences per individual

5

Fraction of random individuals

0.05

Genetic diversity weight

0.5

Individuals per generation

100

Fraction of evolved individuals

0.9

Temporal diversity weight:

0.5

Number of cores

1

Fraction of elite individuals

0.05

Optimize genetic diversity

Max diversity

Generations

10

Jukes-Cantor
distance

Distance file (csv, rds)

Browse...

Sequence file (fasta)

Browse...

Metadata file (csv)

Browse...

Dataset name

example.dataset

Select output directory

...

../shiny_output

Tardis will run on individuals per generation, handled by 1 core(s), for 10 generations, counted as 0 to 9
The total combination of possible subsamples is not yet calculated (Please load metadata file)

Run Tardis

Calculate the
distance

Where do the
output goes?

PRINCIPLES OF GENETIC ALGORITHMS

- Artificial intelligence
(Biomimetic systems)
- Solve a problem with a Darwinian evolutionary approach
- Literally evolve the best solution
 - is the **most fit** to solve the problem
 - evolves from a population of solutions

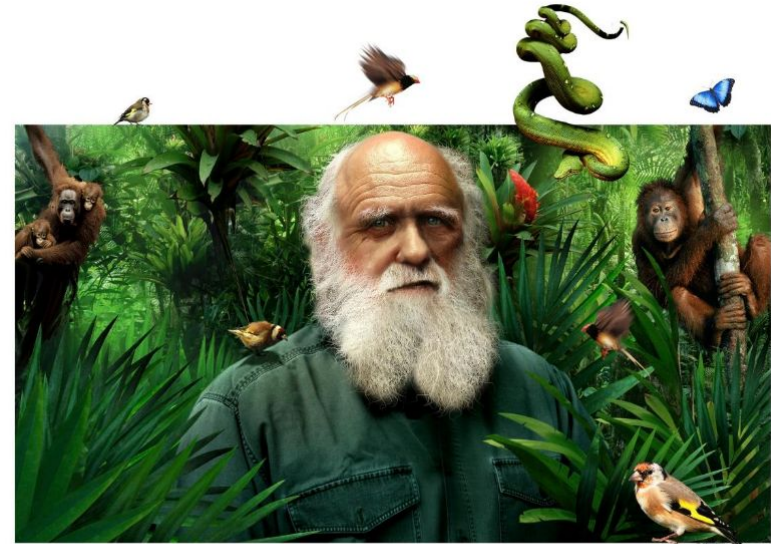


NTB, Science Photo Library

[Bruk bildet](#)

PRINCIPLES OF GENETIC ALGORITHMS

- Set a problem
 - Find the best sequence subsample in terms of **genetic diversity** and **temporal distribution**
- Set a fitness function
 - subset $a >$ subset b ?
 - Measure average genetic distance
 - Measure how sequences are distributed over the timeline (we want a uniform distribution)
 - Fitness is **number**
- Generate a random pop of **solutions**
 - Each solution = a subset
 - Think of solutions as animals in an environment (our problem)
 - Some are fitter than other (fitness function)
 - We can rank solutions based on fitness



NTB, Science Photo Library

[Bruk bildet](#)

PRINCIPLES OF GENETIC ALGORITHMS

- Select solutions for mating
 - Solutions are merged to generate new ones
- Selection depends of fitness
 - Fitter solution, higher the probability of being selected for mating
- Mating example
 - Subset of 5 sequences out of 100 total sequences $\rightarrow Q1, \dots, Q100$
 - $A = \text{seq}(Q1, Q2, \text{Q5, Q50, Q81})$
 - $B = \text{seq}(Q1, Q2, \text{Q9, Q8, Q15})$
 - $A \text{ mates } B \rightarrow C = \text{seq}(Q1, Q2, \text{Q8, Q10, Q81})$



PRINCIPLES OF GENETIC ALGORITHMS

- Selection depends of fitness
 - Fitter solution, higher the probability of being selected for mating
- Start with random generated pop
- Iteratively, solutions are selected to fill a new generation
- We expect that, based on Darwinian principles, our populations will get better at solving the problem

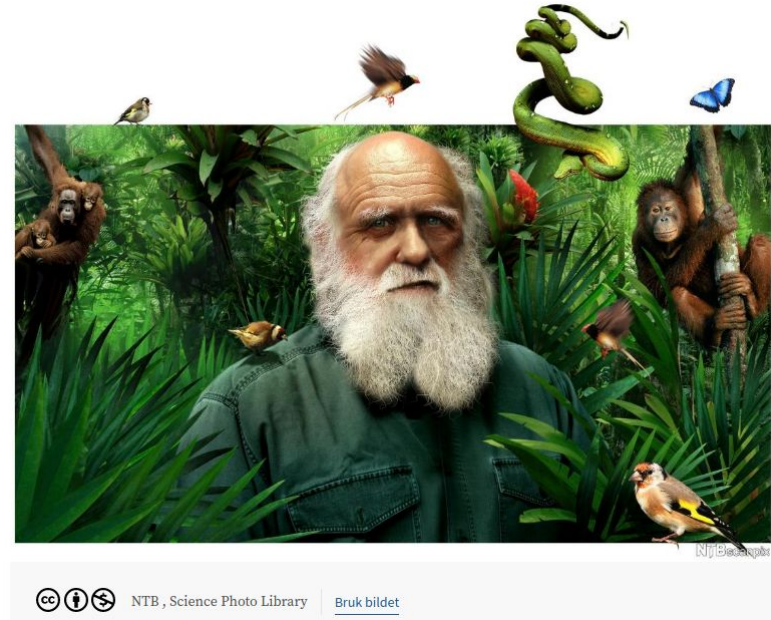


NTB, Science Photo Library

[Bruk bildet](#)

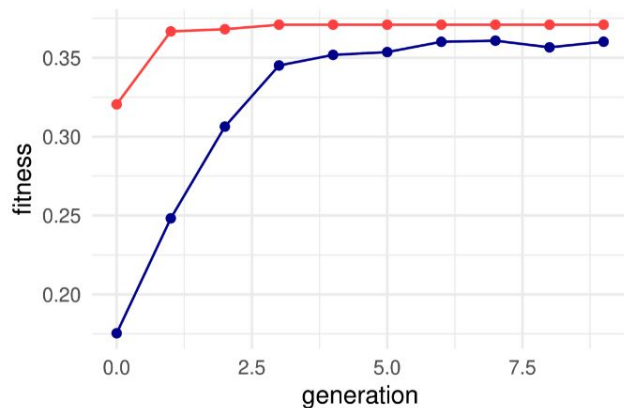
PRINCIPLES OF GENETIC ALGORITHMS

- Streetlight regulation
- Airlines Revenue Management
- Computer-automated design
- Trade & finance strategies
- Bioinformatics (e.g. motif discovery)
- Code-breaking and cryptography
- Container loading optimization
- Design of water distribution systems
- Logistics
- ...

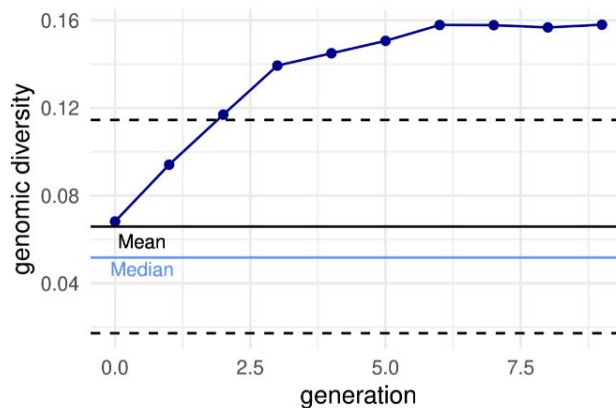


OUTPUT

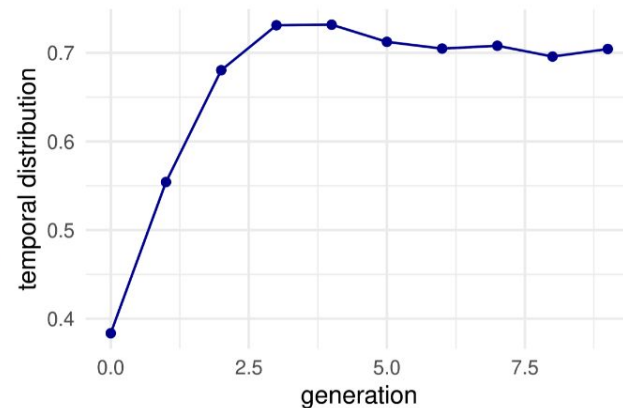
- Subsample fasta
- Per-generation fitness stats
 - General
 - Temporal distribution
 - Genetic diversity



—●— best fitness —●— mean fitness



—●— mean genetic diversity



—●— mean temporal spread

./tardis -s

Parameters of the
genetic algorithm

Tardis

Sequences per individual <input type="text" value="5"/>	Fraction of random individuals <input type="text" value="0.05"/>	Genetic diversity weight <input type="text" value="0.5"/>
Individuals per generation <input type="text" value="100"/>	Fraction of evolved individuals <input type="text" value="0.9"/>	Temporal diversity weight: <input type="text" value="0.5"/>
Number of cores <input type="text" value="1"/>	Fraction of elite individuals <input type="text" value="0.05"/>	Optimize genetic diversity <div>Max diversity ▼</div>
	Generations <input type="text" value="10"/>	<div>Jukes-Cantor distance</div>

Distance file (csv, rds) <div>Browse... <input type="text" value="..."/></div>	Sequence file (fasta) <div>Browse... <input type="text" value="..."/></div>	Metadata file (csv) <div>Browse... <input type="text" value="..."/></div>
---	--	--

Dataset name <input type="text" value="example.dataset"/>	Select output directory <div><div>Ⓜ</div> <input type="text" value="../shiny_output"/> <div>...</div></div>
--	--

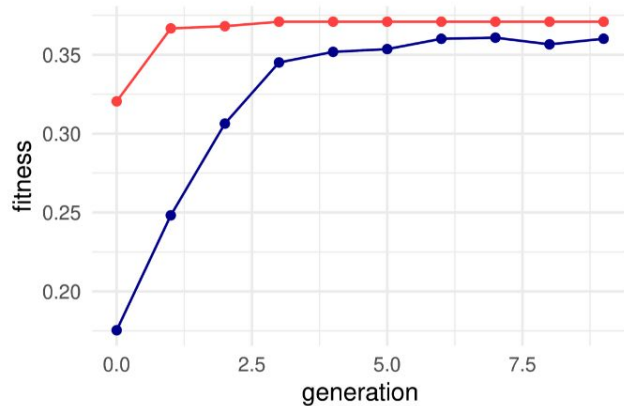
Tardis will run on individuals per generation, handled by 1 core(s), for 10 generations, counted as 0 to 9
The total combination of possible subsamples is not yet calculated (Please load metadata file)

Run Tardis

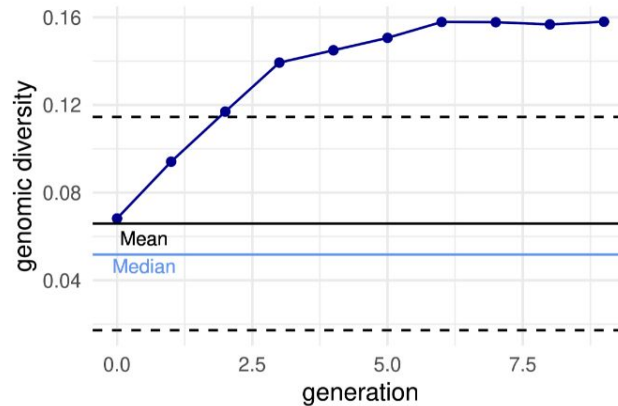
HOW TO DECIDE PARAMETERS?

Empirically

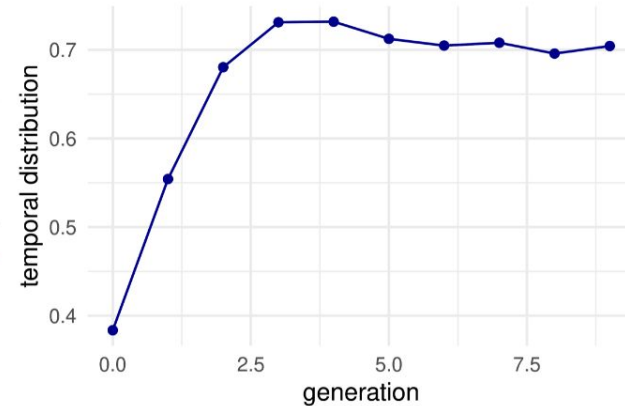
- Check the graphical output, is a plateau reached?
- What are the limits of the machine I am using?
- “Batches” to split the GA population and process one piece at the time



—●— best fitness —●— mean fitness



—●— mean genetic diversity

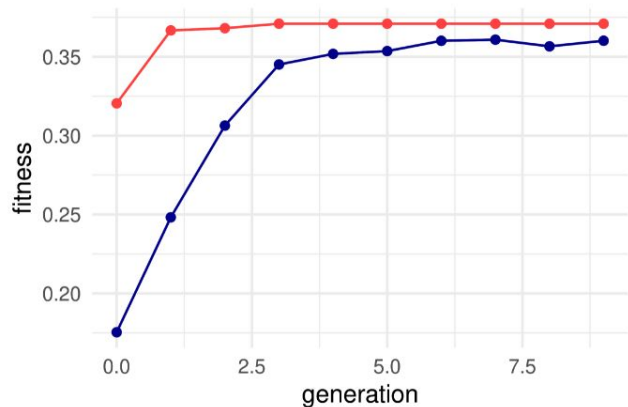


—●— mean temporal spread

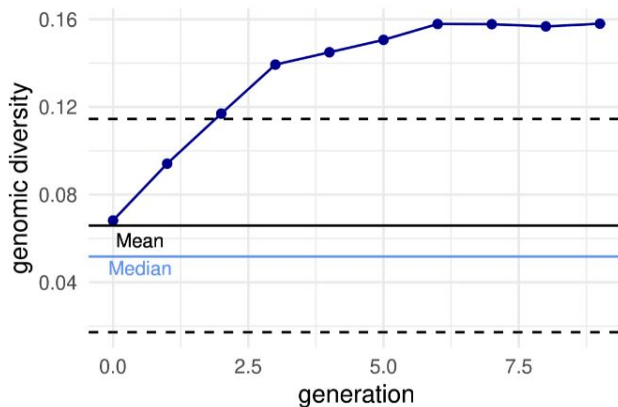
HOW ABOUT REPRODUCIBILITY?

Seed set by default

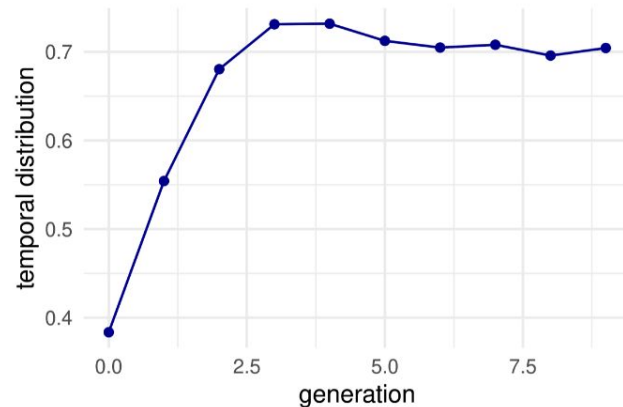
- seeds.txt file in /data
- Will ensure two runs on the same data set provide the same results
- User can change the file with own seeds



—●— best fitness —●— mean fitness



—●— mean genetic diversity



—●— mean temporal spread

COMMAND LINE

- GUI designed for small datasets/exploration
- For large datasets/memory problems → command line
- Command line
 - Local machine
 - HPC (used on Hipergator with SLURM)
 - Can handle large data splitting the data into batches
- Calculate distance from command line

```
Rscript bin/JC.pairwise.dist.R -i <alignement_file> \  
-d <output_distance_file> -c <number_of_cores>
```

- Launch TARDiS from command line

```
./tardis example/example.config
```


EXAMPLE.CONFIG

- Configuration file where parameters are specified

```
params.data_set = "example"  
params.nsamples = 4  
params.gensize = 100  
params.nbatches = 1  
params.ncores = 2  
params.ngenerations = 10  
params.fracnew = 0.14  
params.fracevolved = 0.85  
params.fracelite = 0.01  
params.wdiv = 1  
params.wtem = 1
```

NEXTFLOW.CONFIG

- Different NextFlow profiles
- Default is local
- Other can be defined based on need (e.g. memory)

```
profiles {  
  local {  
    executor.name = "local"  
  }  
  
  small {  
    executor.name = "slurm"  
    process.memory = "10G"  
    process.time = "96h"  
    process.clusterOptions="--qos=salemi-b"  
  }  
  
  medium {  
    executor.name = "slurm"  
    process.memory = "30G"  
  }  
  
  large {  
    executor.name = "slurm"  
    executor.memory = "128G"  
  }  
}
```

`./tardis -p large example/example.config`

TARDiS OPTIONS

`./tardis -h`

Temporal And diveRsity Diistribution Sampler - TARDiS

Usage: tardis [options] CONF

where CONF is the path to the configuration file. The configuration file should be named DATASET.config, where DATASET is the dataset name. Use -H to see the list of parameters that can be placed in the configuration file.

Options:

-s	Enable Shiny GUI (all other options ignored)
-g GROUPS	Enable groups mode
-b NBATCH	Number of batches (default:)
-n NGENS	Set number of generations for GA (default: 10)
-a AFIL	Name of alignment file (default: aln.fa) [*]
-d DFIL	Name of distances file (default: jc.distance.precalc.rds) [*]
-m MFIL	Name of metadata file (default: metadata.csv) [*]
-o ODIR	Name of output directory (default: output/dataset_name)
-t DISTOPT	Genetic diversity optimization, possible values: max, mean, median (default: max)
-p PROF	Profile name, possible values: local, small, medium, large (default: local)
-x	Enable debug mode (will not delete intermediate files)
-v	Display version number and copyright notice
-H	Describe format of configuration file

[*] If the default filename is used, the program assumes that the file is in the same directory as the configuration file.

GROUP OPTION

- Multiple aligned sequences with a group label
- Could be a per-country split, e.g. a single fasta file with
 - 1000 sequences from USA
 - 500 sequences from UK
 - 100 sequences from Italy
- We want an ad-hoc, per group sampling
 - 100 sequences from USA
 - 100 sequences from UK
 - 100 sequences from Italy
- The GA parameters for each group are different (diff search space)

```
./tardis -g <group_parameter_file> -m <group_metadata_file> \  
-a <group alignment file>
```

GROUP OPTION

group	params.ncores	params.nsamples	params.gensize	params.nbatches	params.ngenerations	params.fracnew	params.fracevolved	params.fracelite	profile
USA	8	100	1000	1	50	0.3	0.69	0.01	local
UK	4	100	250	1	25				local

- One line per group
- Needs column “group”
- Non-specified parameters are kept at default
- Non-specified groups not subsampled and retained in full for the output
- General output (full set) + per-group output

TARDiS TEAM

Feel free to contact me for help/bugs

Email simone.marini@ufl.edu

Twitter @simone__marini



Carla Mavian



Alberto Riva



Brittany Rife Magalis



Marco Salemi

