# PREDICTIVE WEBSERVER T1

## BACKGROUND & STRATEGY

CREATING THE NEXT®

# OVERVIEW

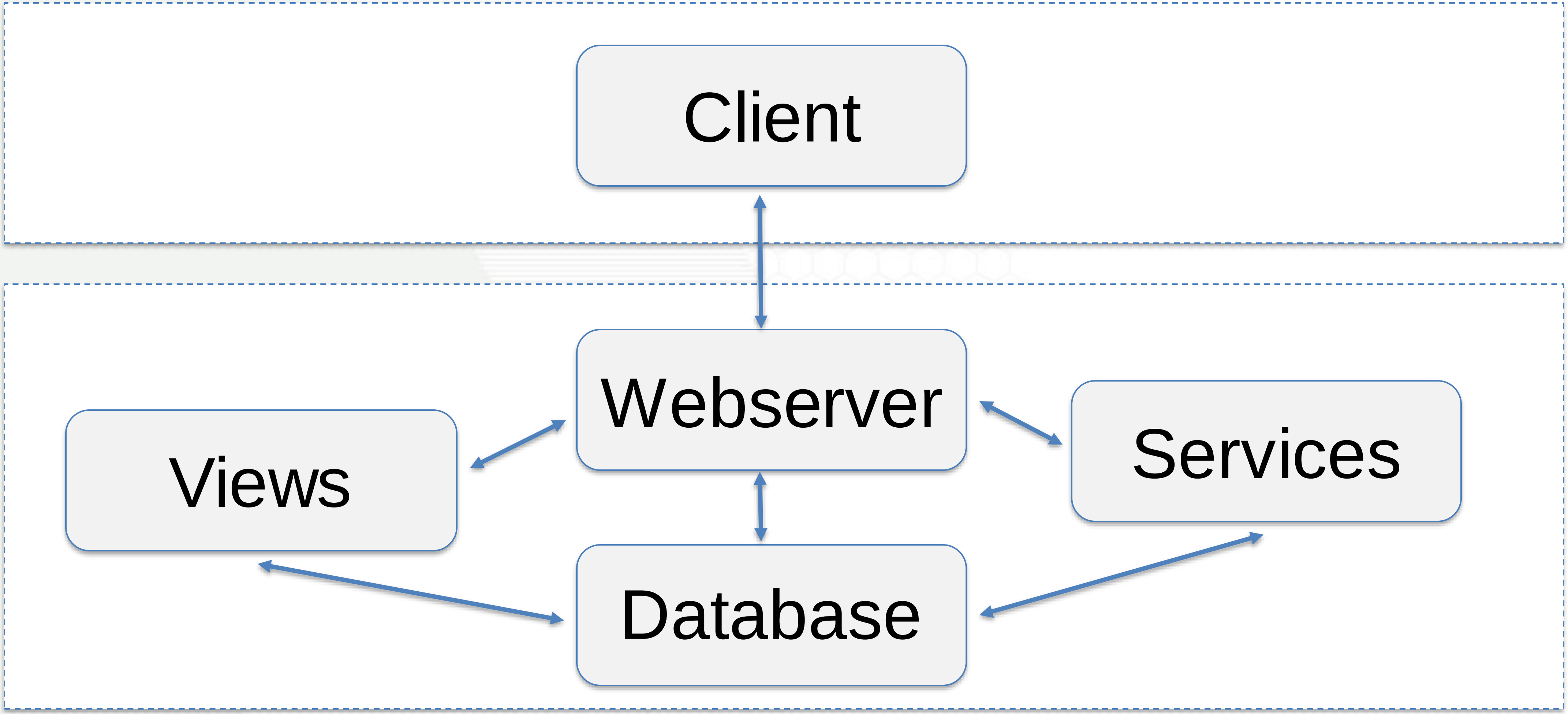
- Predictive Webserver
- Goal/Expectation
- Functionalities
- Technologies Used
- Future Plans

# Predictive Webserver

# PREDICTIVE WEBSERVER

- Deliver content or services to users via internet
  - Process biological data
  - Output the results of analysis

# PREDICTIVE WEBSERVER

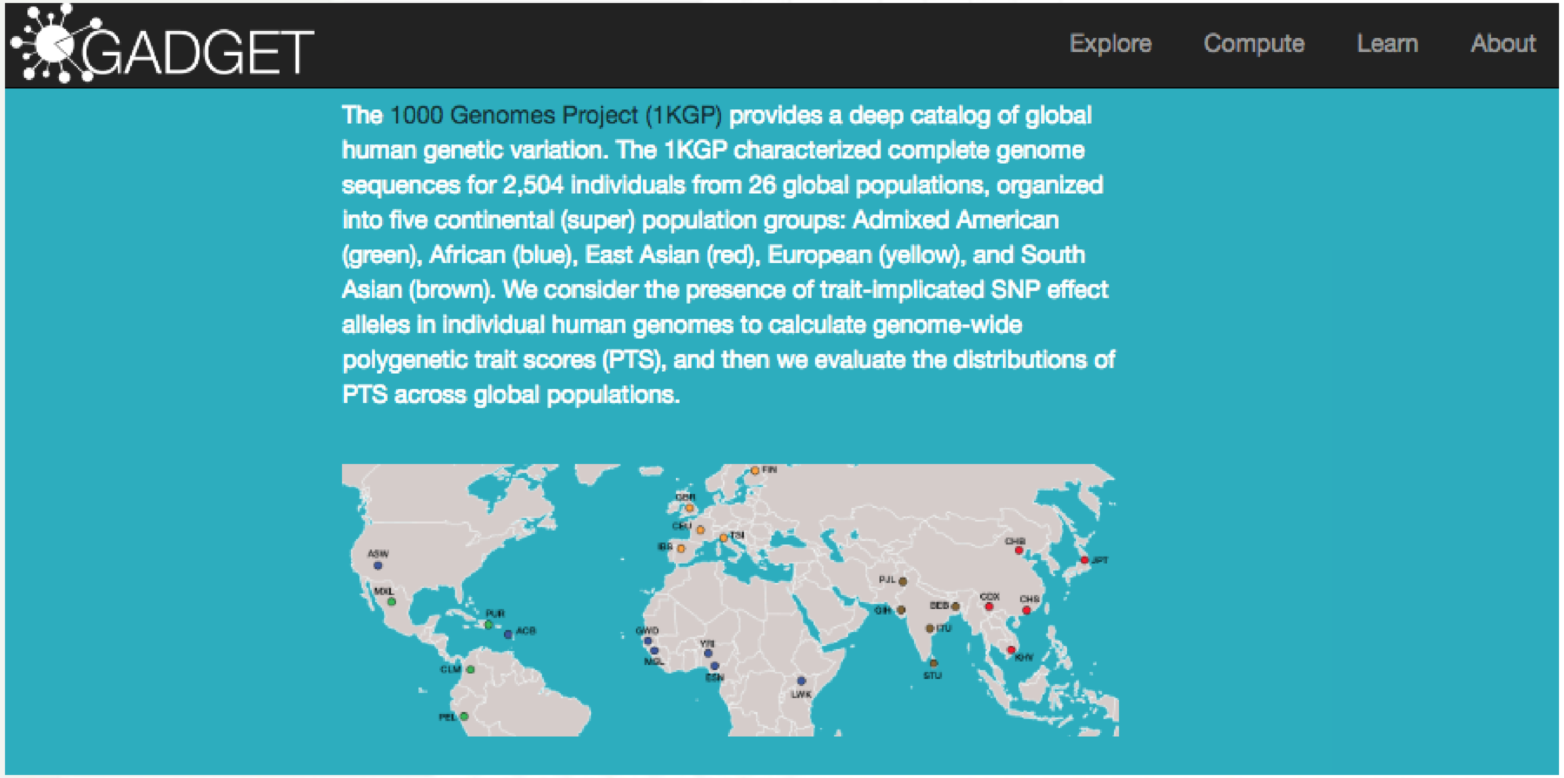


# PREDICTIVE WEBSERVER

GADGET - Global Distribution of Genetic Traits

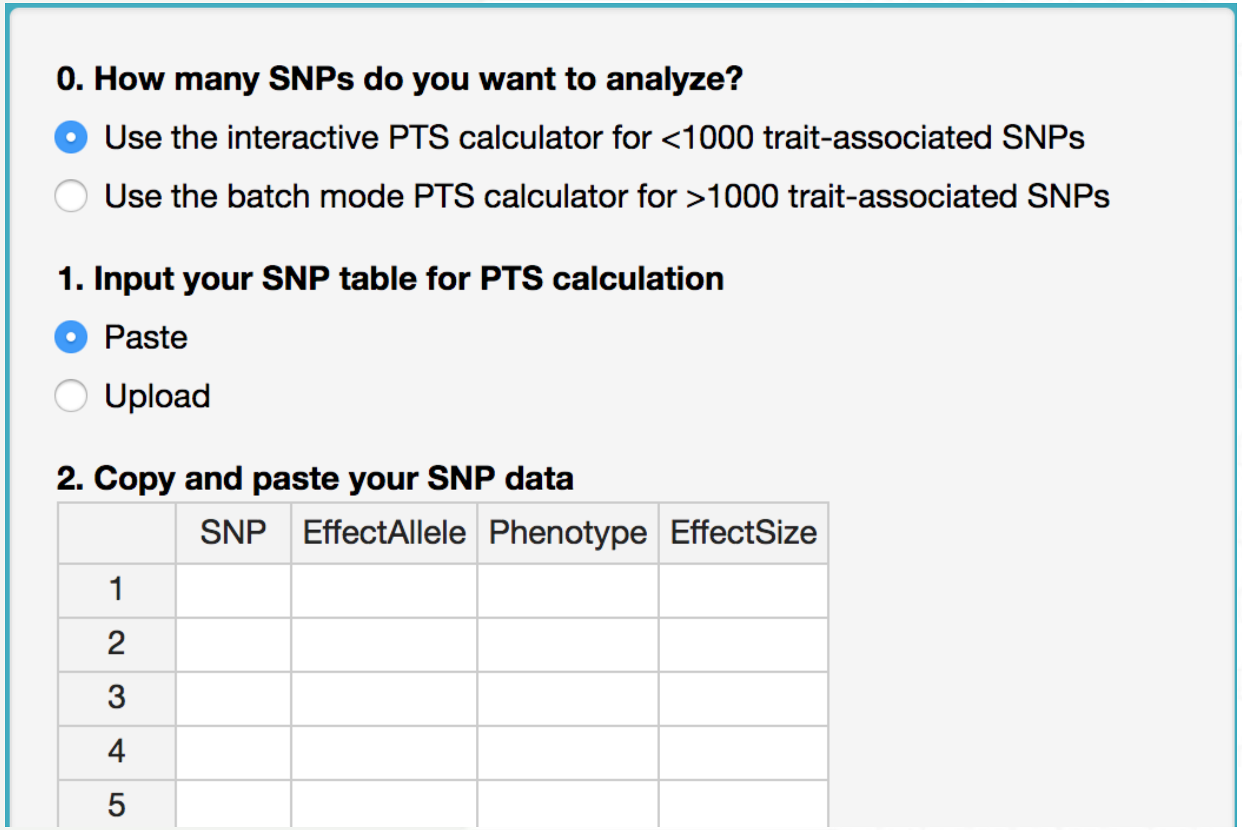
https://gadget.biosci.gatech.edu/

Visualize global distribution of traits of interest



The 1000 Genomes Project (1KGP) provides a deep catalog of global human genetic variation. The 1KGP characterized complete genome sequences for 2,504 individuals from 26 global populations, organized into five continental (super) population groups: Admixed American (green), African (blue), East Asian (red), European (yellow), and South Asian (brown). We consider the presence of trait-implicated SNP effect alleles in individual human genomes to calculate genome-wide polygenetic trait scores (PTS), and then we evaluate the distributions of PTS across global populations.

# PREDICTIVE WEBSERVER



**0. How many SNPs do you want to analyze?**

● Use the interactive PTS calculator for <1000 trait-associated SNPs

○ Use the batch mode PTS calculator for >1000 trait-associated SNPs

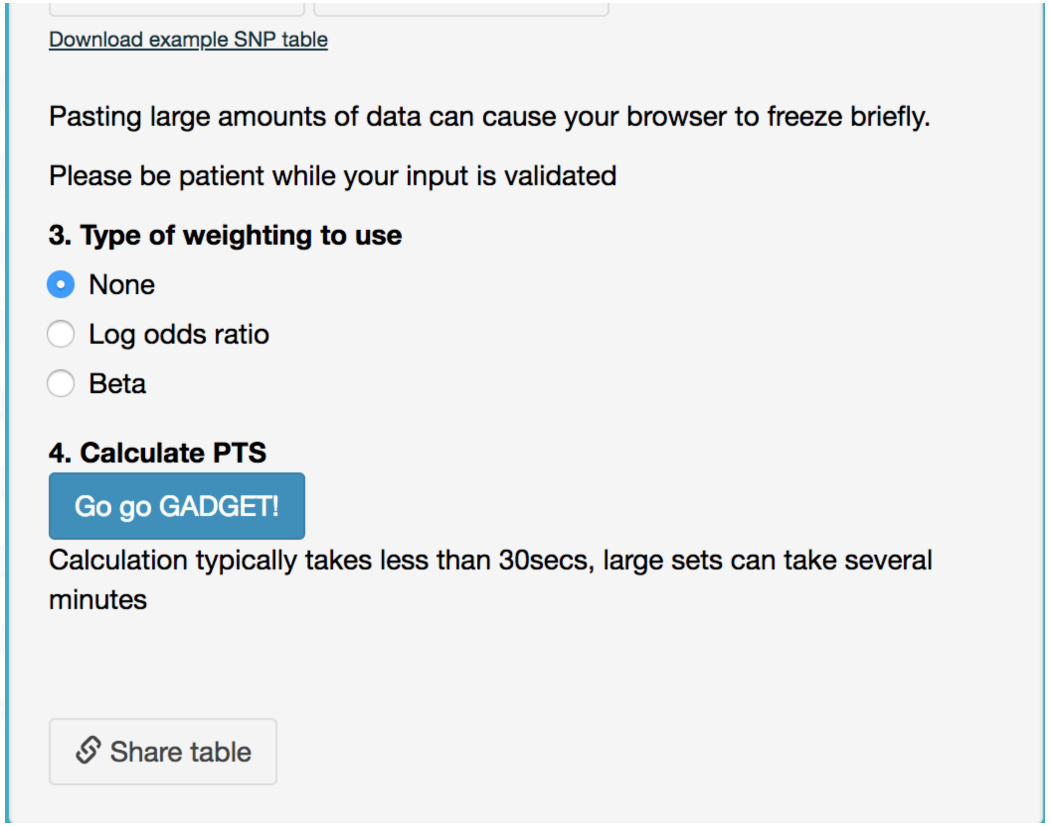
**1. Input your SNP table for PTS calculation**

● Paste

○ Upload

**2. Copy and paste your SNP data**

| | SNP | EffectAllele | Phenotype | EffectSize |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |



Download example SNP table

Pasting large amounts of data can cause your browser to freeze briefly.

Please be patient while your input is validated

**3. Type of weighting to use**

● None

○ Log odds ratio

○ Beta

**4. Calculate PTS**

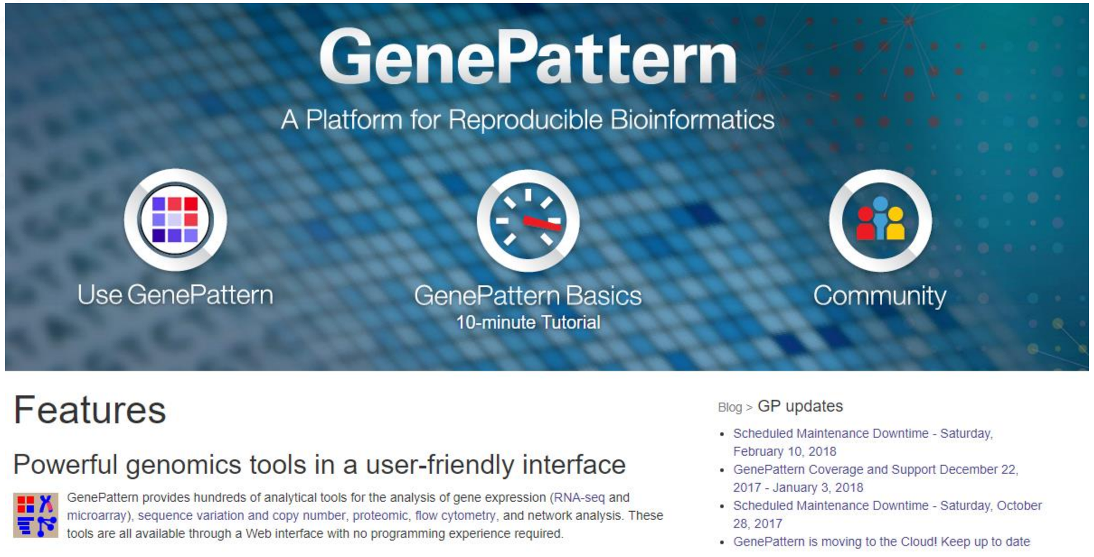
Calculation typically takes less than 30secs, large sets can take several minutes

# PREDICTIVE WEBSERVER

- Hosts hundreds of genomic analysis tools
  - Gene expression analysis, short-read sequencing, ...

- Users can create pipelines and perform analysis

- Open-source
  - Can be downloaded locally



Features

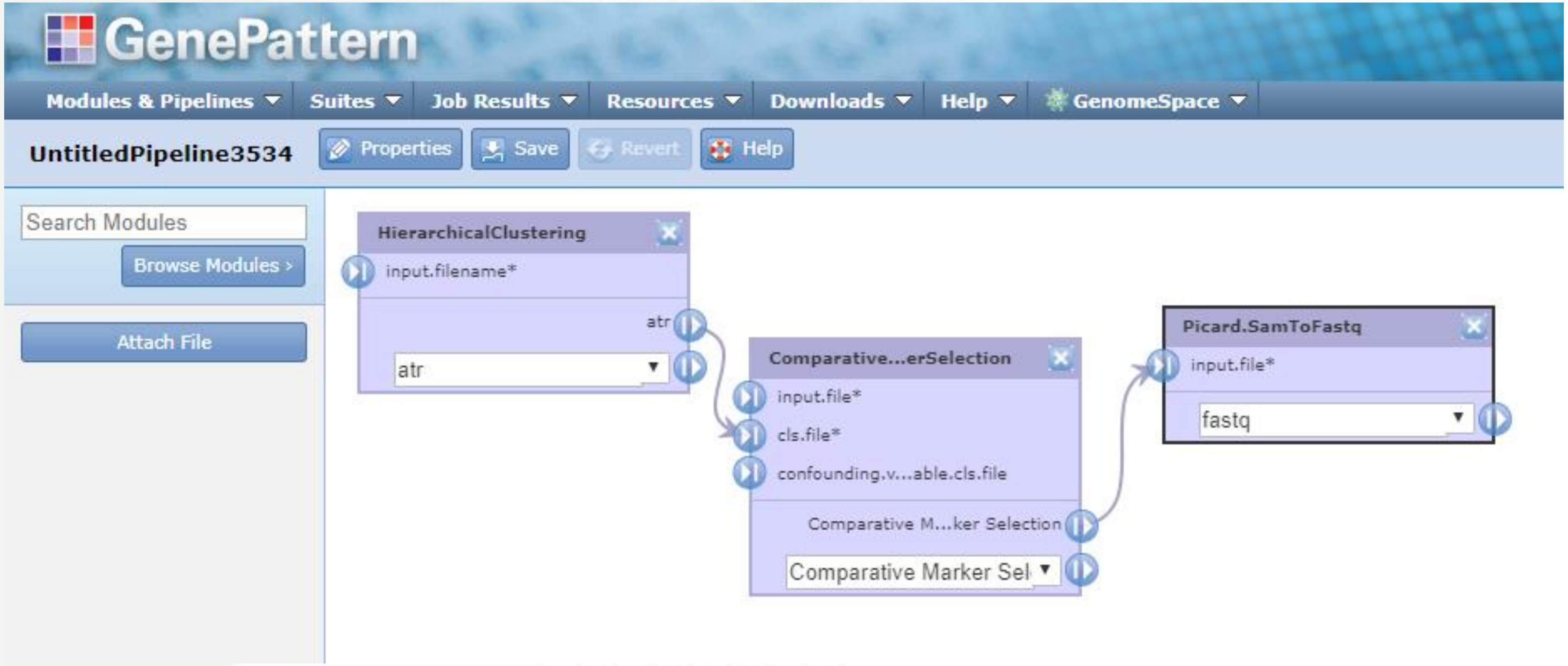
Powerful genomics tools in a user-friendly interface

GenePattern provides hundreds of analytical tools for the analysis of gene expression (RNA-seq and microarray), sequence variation and copy number, proteomic, flow cytometry, and network analysis. These tools are all available through a Web interface with no programming experience required.

Blog > GP updates

- Scheduled Maintenance Downtime - Saturday, February 10, 2018
- GenePattern Coverage and Support December 22, 2017 - January 3, 2018
- Scheduled Maintenance Downtime - Saturday, October 28, 2017
- GenePattern is moving to the Cloud! Keep up to date

# PREDICTIVE WEBSERVER



UntitledPipeline3534

HierarchicalClustering: atr

Comparative...erSelection: Comparative Marker Sel

Picard.SamToFastq: fastq

# Goals & Expectations

# GOALS/EXPECTATIONS

**GOALS:** Provide information about input sample's antibiotic resistance and other biological traits based on comparative genomics group's findings

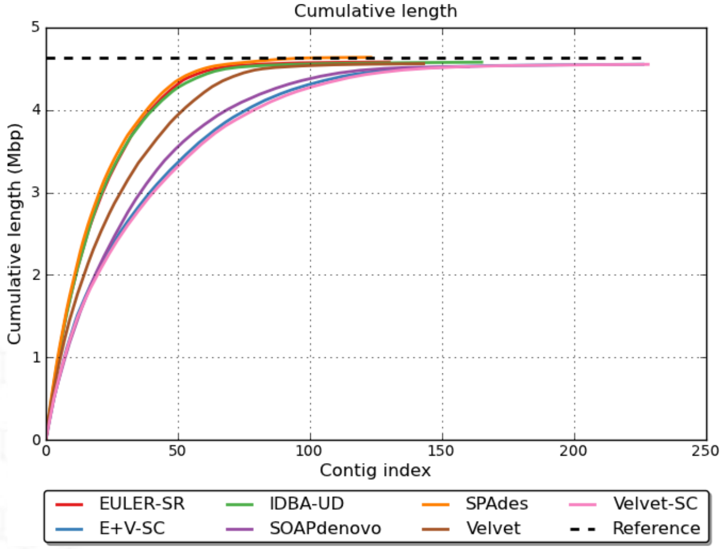
**EXPECTATIONS:**

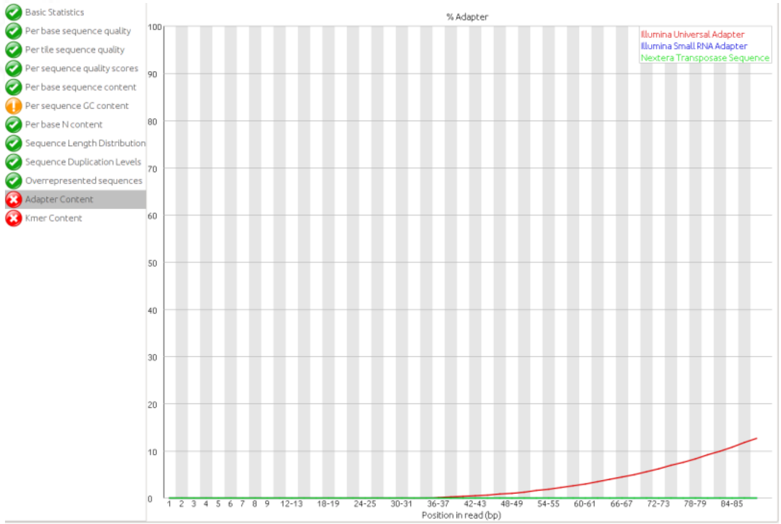
- Assemble reads
- Analyze assemblies
- Visualization of results
- Downloadable results (e.g. tabular, visualization, ...)

# Functionalities

# FUNCTIONALITIES

- Jobs
  - Notify user regarding the Job status of the process
  - Quality Checks
  - SKESA - Assembly/Sub-assembly, SSPACE – Scaffolding





Gurevich et al. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 29:1072-1075

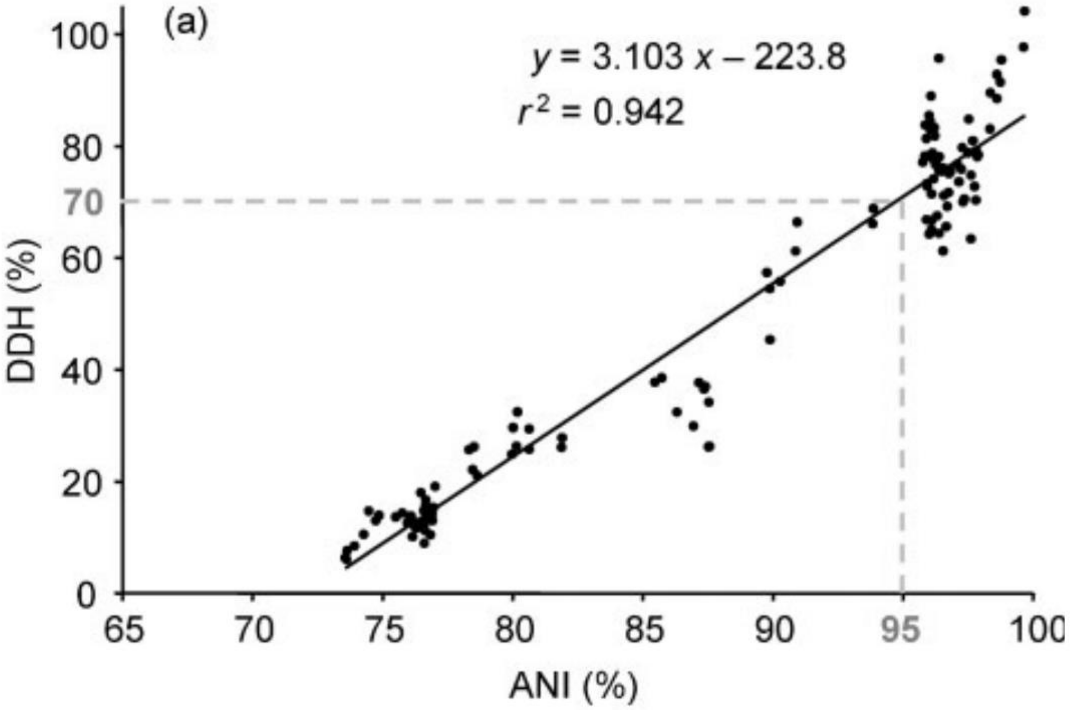
# FUNCTIONALITIES

What features will our website have?

- De novo assembly (SKESA)
  - Subassemblies
- Typing Tool
- Genome database
  - D/B to store critical information regarding the genomes of *Kleibsiella spp.*
  - Species Identification
  - Antibiotic Resistance Genes
  - Hetero-resistant and Hetero-susceptible genes

# FUNCTIONALITIES

- Comparative Genomics group will inform us on the best type of heteroresistant typing
- Currently, we are looking into using ANI (Average Nucleotide Identity) to compare the assembled genomes to the heteroresistant marker sequences



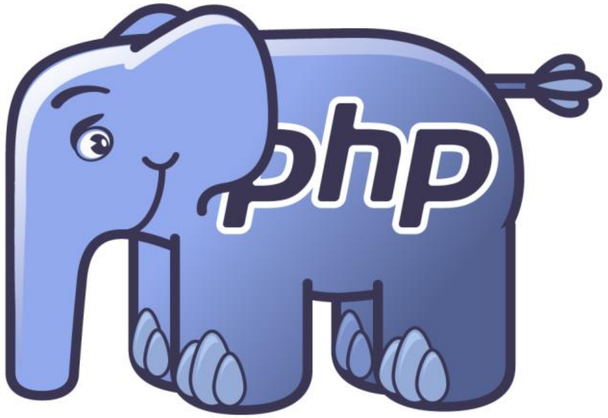
Goris, Johan, et al. "DNA–DNA hybridization values and their relationship to whole-genome sequence similarities." *International journal of systematic and evolutionary microbiology* 57.1 (2007): 81-91.

# Technologies Used

# TECHNOLOGIES USED

## Why PHP?

- Universal language for the web
  - 80% of the websites
  - Facebook, Wikipedia, ...
- Many useful libraries
- Syntax similar to Perl ($)
- Easy to integrate with HTML
- Extensive online documentation

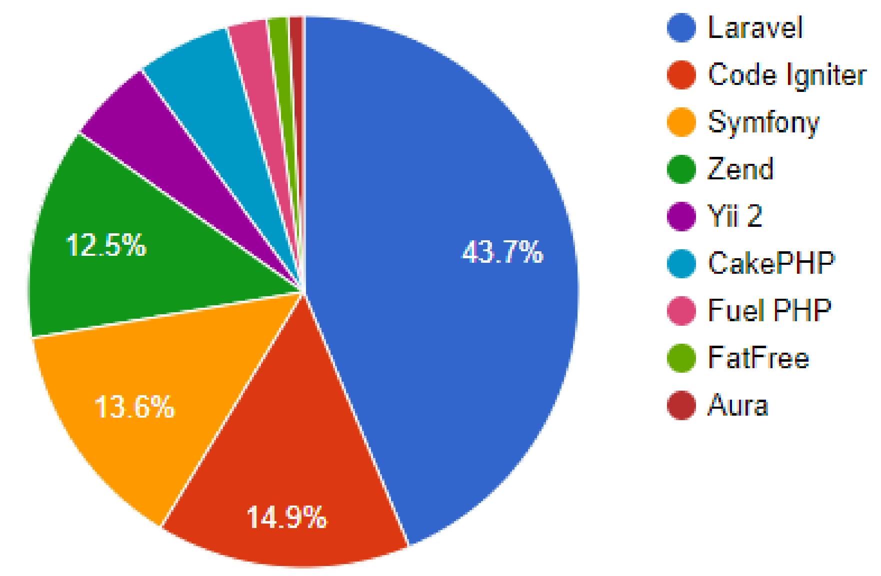


# TECHNOLOGIES USED

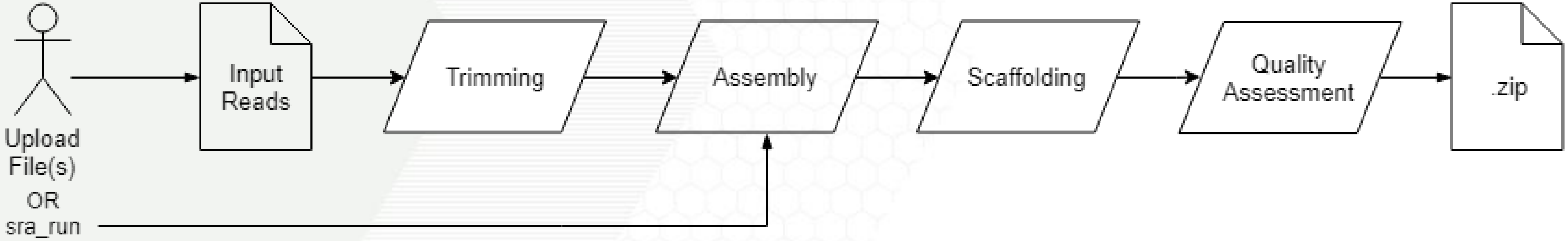
## Why Laravel?

- Well-structured documentation
- Many functionalities + partially built using Symfony components
- Allow rapid application development

PHP Framework Used for Project Use

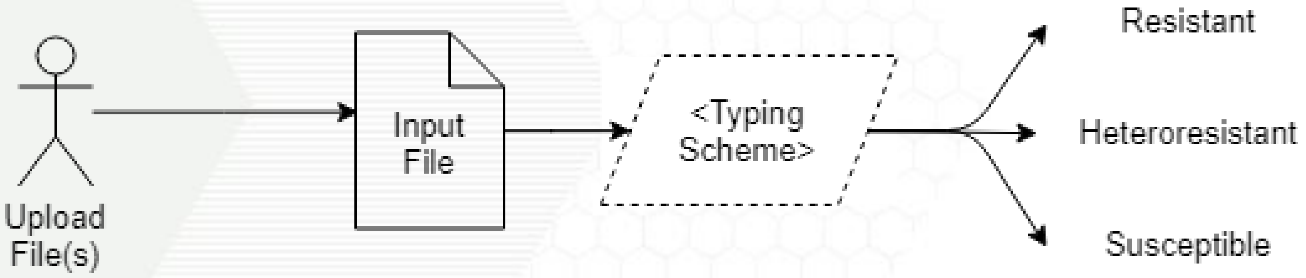


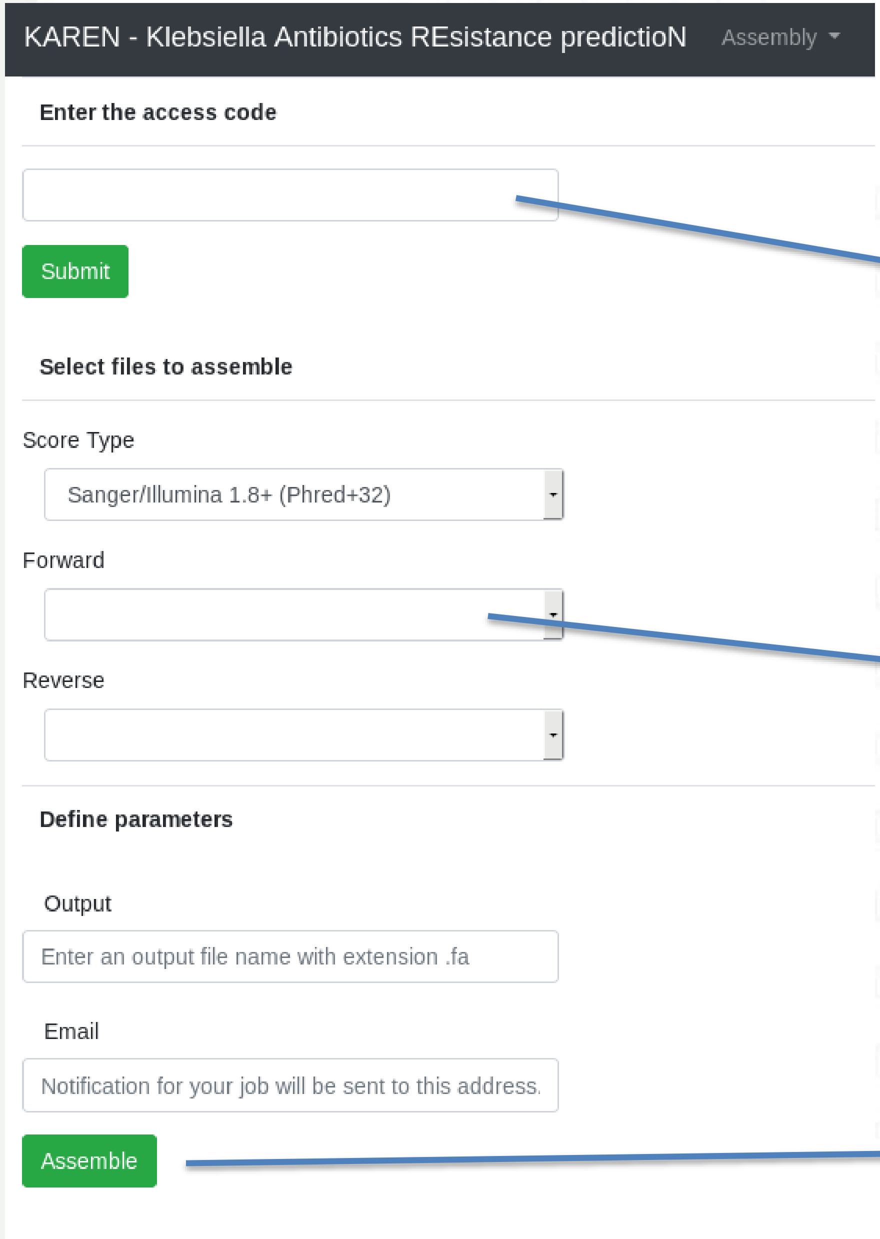
# Assembly


Upload
File(s)
OR
sra_run
Input
Reads
Trimming
Assembly
Scaffolding
Quality
Assessment
.zip

# TECHNOLOGIES USED

## Typing Tool (Genome DB)





KAREN - Klebsiella Antibiotics REsistance predictioN

**Enter the access code**

**Select files to assemble**

Score Type: Sanger/Illumina 1.8+ (Phred+32)

**Define parameters**

Output: Enter an output file name with extension .fa

Email: Notification for your job will be sent to this address.

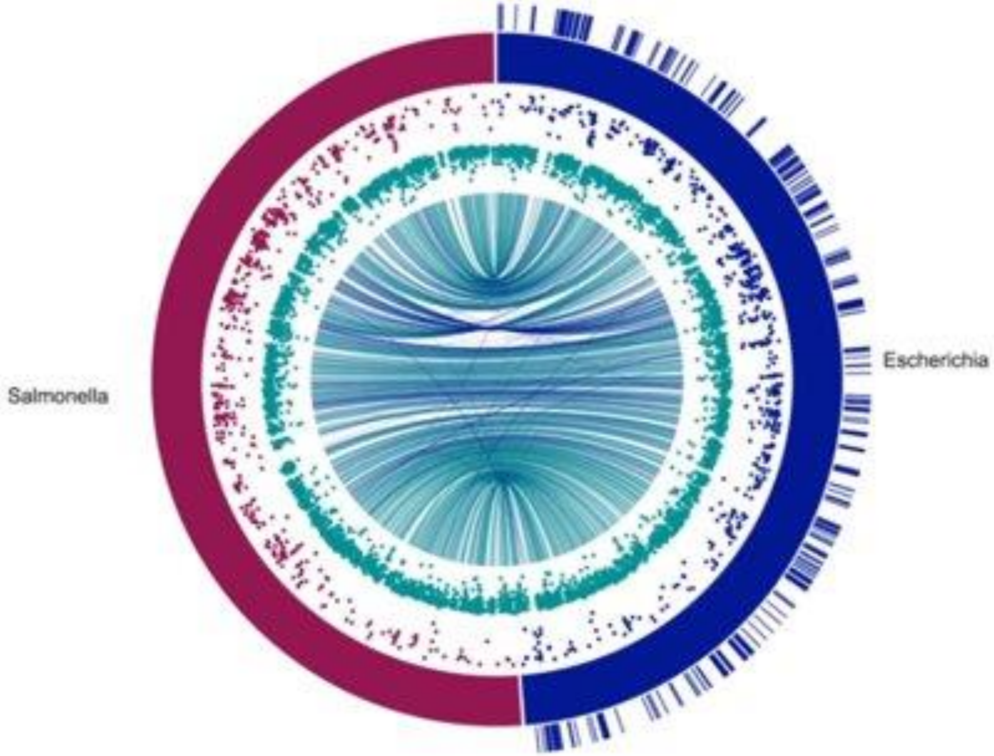
A token to access uploaded files for a user will be generated when files are uploaded

Specifying forward & reverse reads

runAssembly.sh script executes upon clicking

# FUTURE PLANS

- Finalize assembly
- Prioritizing tasks
- Prepare for production
- Prototype genome DB
- Job Status
- UI improvement
- Circos visualization?



# Thank you