Predictive Webserver Strategy

Team 1

Bhavay Aggarwal, Joanne Boysen, Daniel Groves, Chloe Pryor, & Charlotte Royer

Predictive Webserver

Webservers as we know them

- **Provide functional interface for users to input data**
    - Server side handles all computation and data processing
    - Client side presents easy to use GUI or platform for user
    - Data is simply uploaded and results are returned automatically

- **Webservers should be:**
    - Easy for users to use
    - Built for efficiency not looks
    - Process and return data in a timely manner
    - Built on reliable established framework

Approach

Concepts

Design & Workflow

Outcomes

Webserver architecture

- Front/Back end
- Functionality
- Connectivity

Webserver platform

- Django
- Structural layout

Core Workflow

- Final tool choices and webserver integration

Webserver functions

- Comparative Genomics pipeline

- Current progress

Webserver Concepts

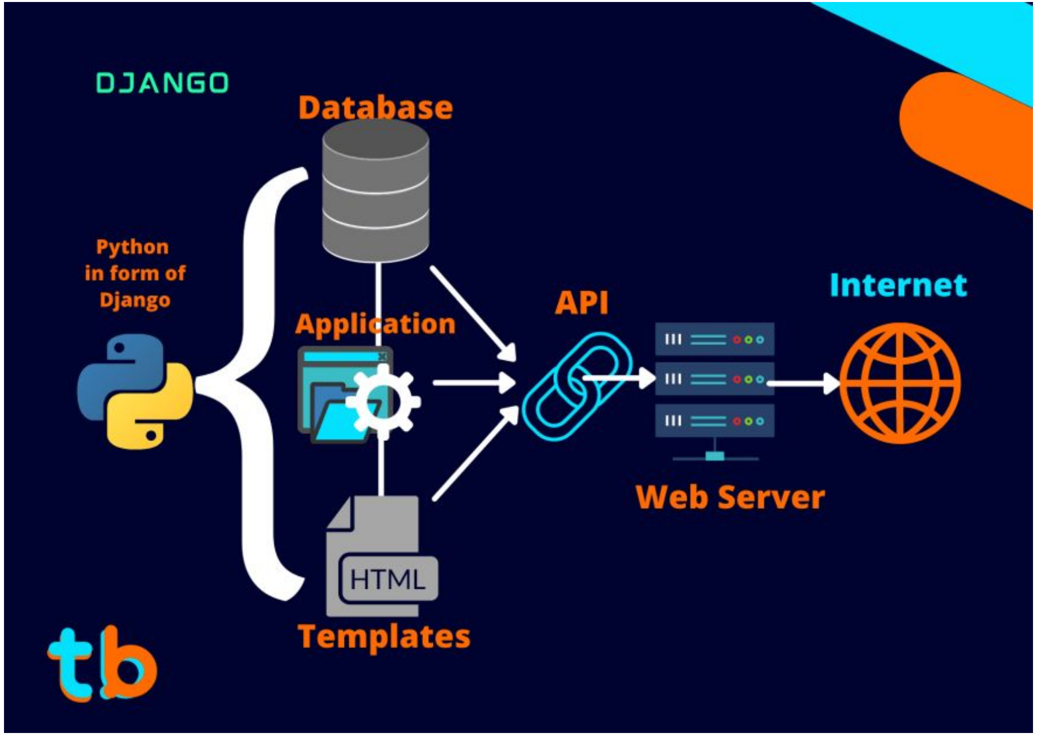
Architecture

Front end:

- Part of server visible to user
- GUI user can interact with

Back end:

- Encodes server structure and function
- Consists of
  - Database
  - application program interface (API)
  - server itself to connect everything

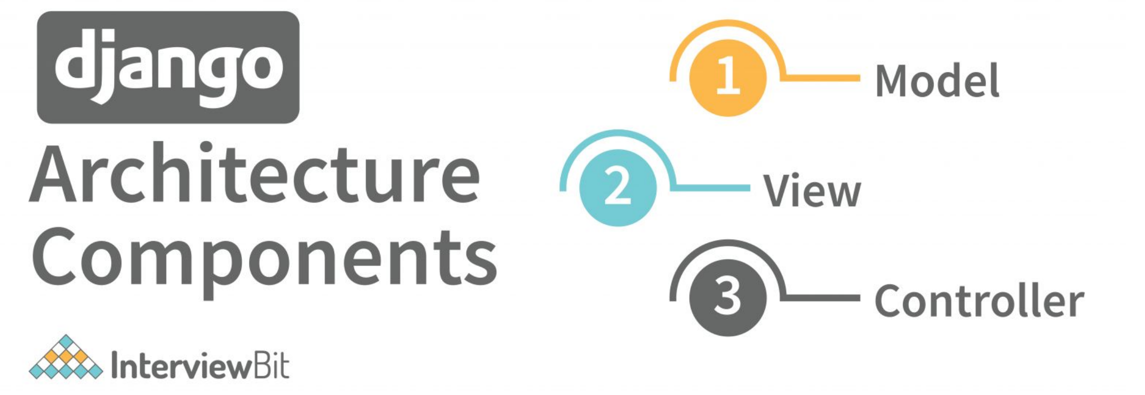


dev.to

Design and Workflow

Django

- Open source web application framework based in python
- Considered best-practice framework for web development
- Easy to use, flexible and has many third-party codes available for integration

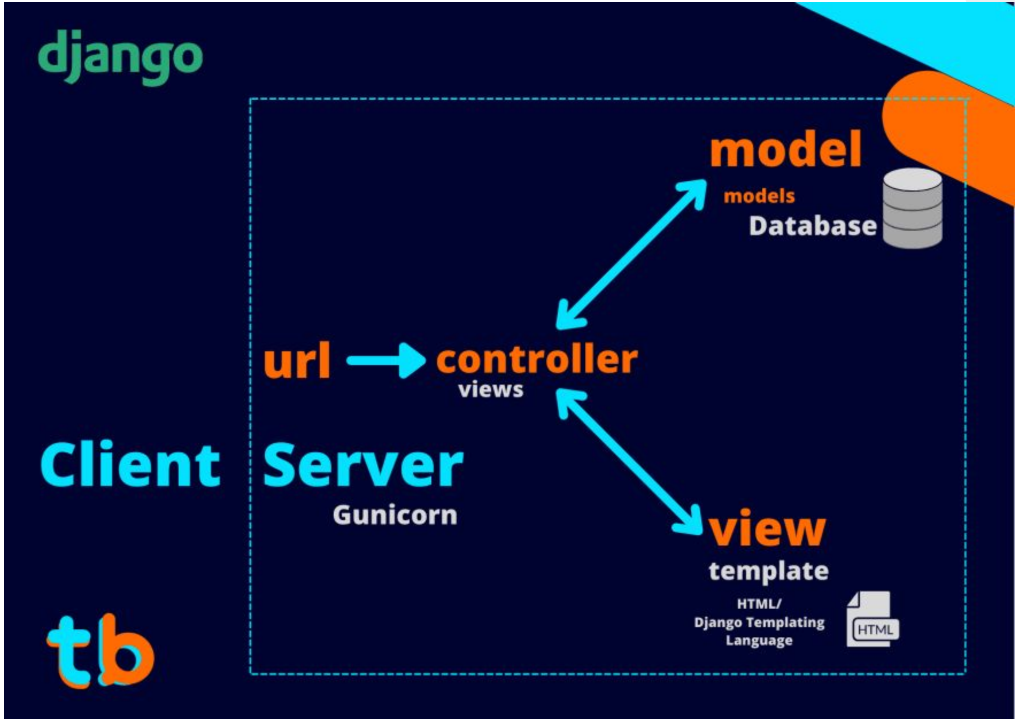


Design and Workflow

Django structure

- **Model**
  - Stores data and related logic
  - Represents and manipulates data
- **Template**
  - Controls how interface is presented to user
  - Processes user requests
- **View**
  - Receives and sends HTTP requests and responses. Interacts with both template and model to complete response

“MTV” structure

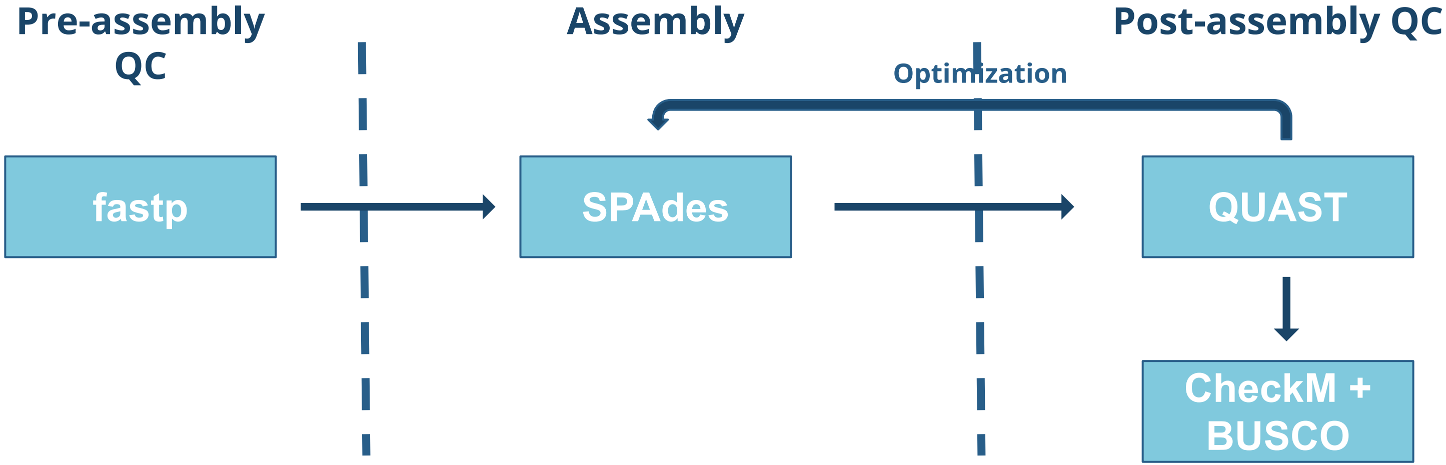


dev.to

Webserver Function 1

Genome Assembly:

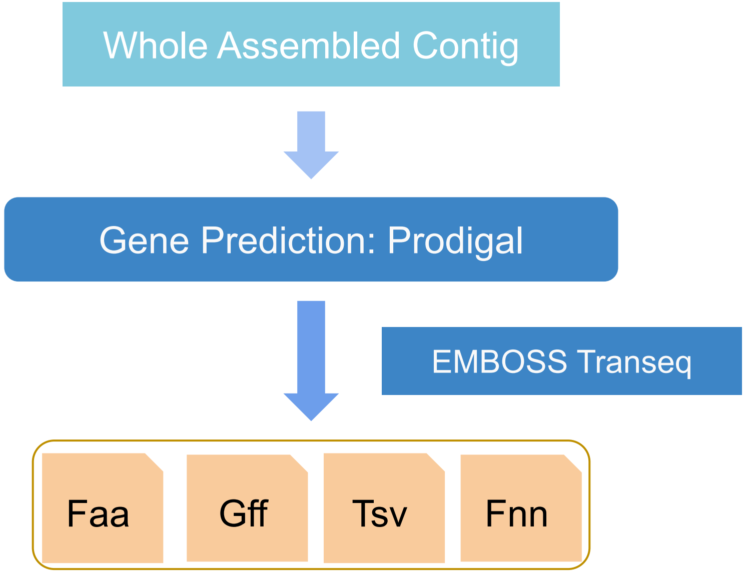
- QC, assembly, etc
- Job tracking and access
- Display outputs of QUAST, etc



Webserver Function 2

Gene Prediction

- Use Prodigal to predict genes
- Produce FastA files of gene nucleotides and protein sequences



Whole Assembled Contig

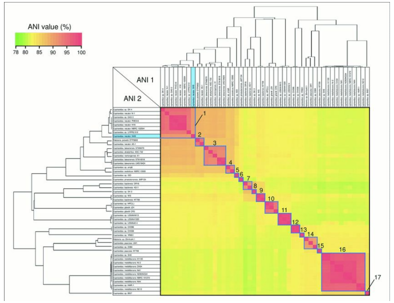
Gene Prediction: Prodigal

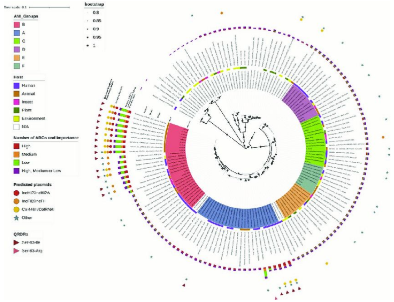
EMBOSS Transeq

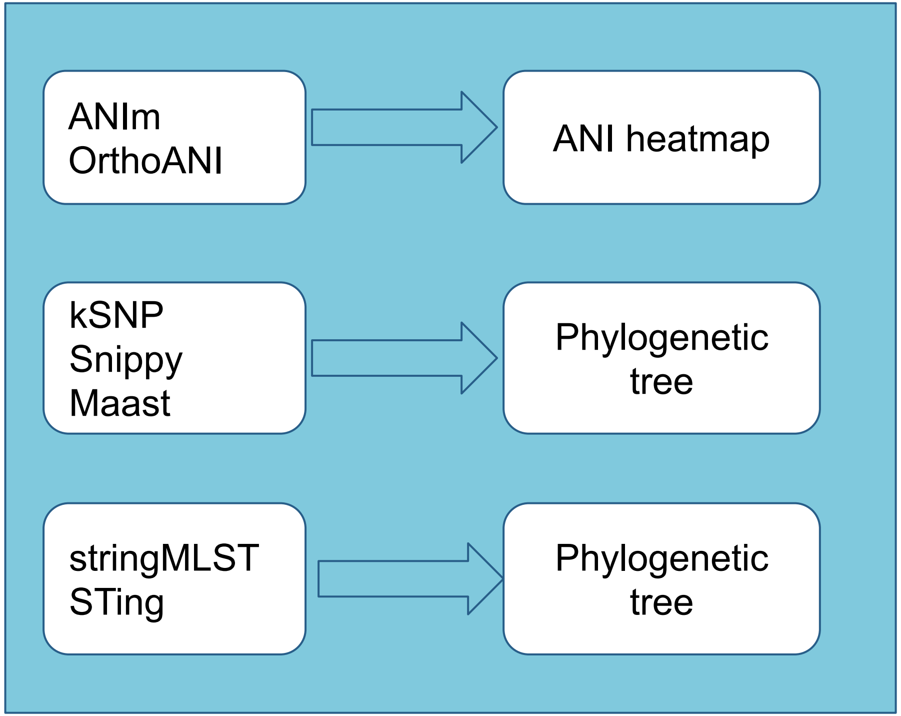
Webserver Function 3

Function 2: Comparative Genomics

- ANI calculation (all-against-all)
- MLST tree
- SNP analysis





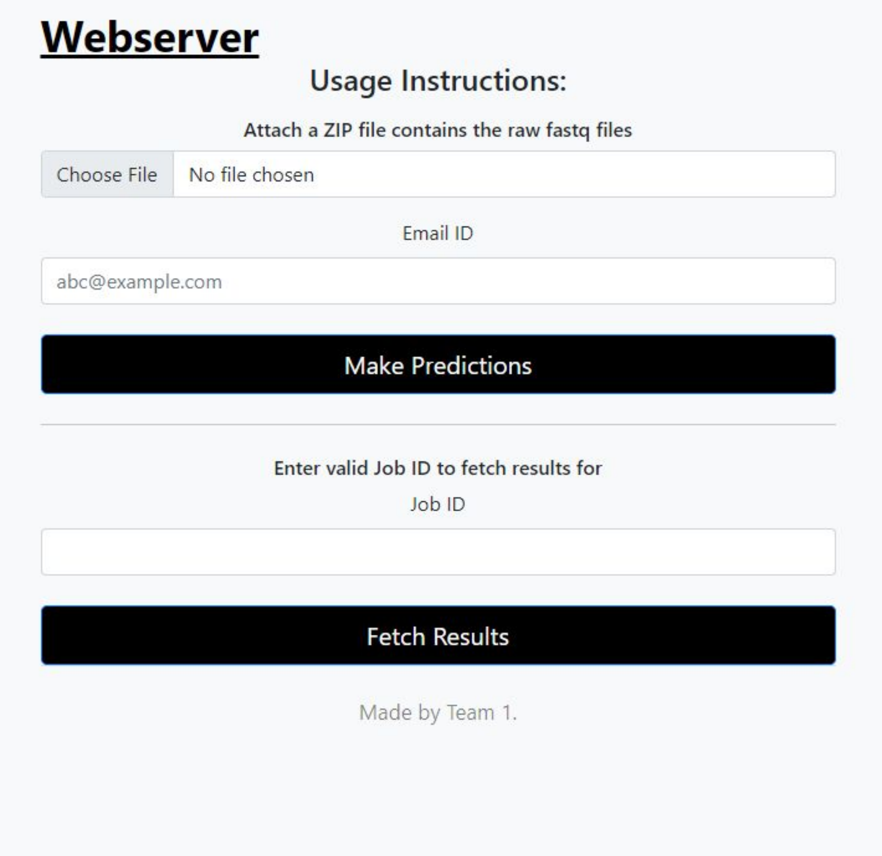


ANIm
OrthoANI
→ ANI heatmap

kSNP
Snippy
Maast
→ Phylogenetic tree

stringMLST
STing
→ Phylogenetic tree

Current Progress



Webserver

Usage Instructions:

Attach a ZIP file contains the raw fastq files

No file chosen

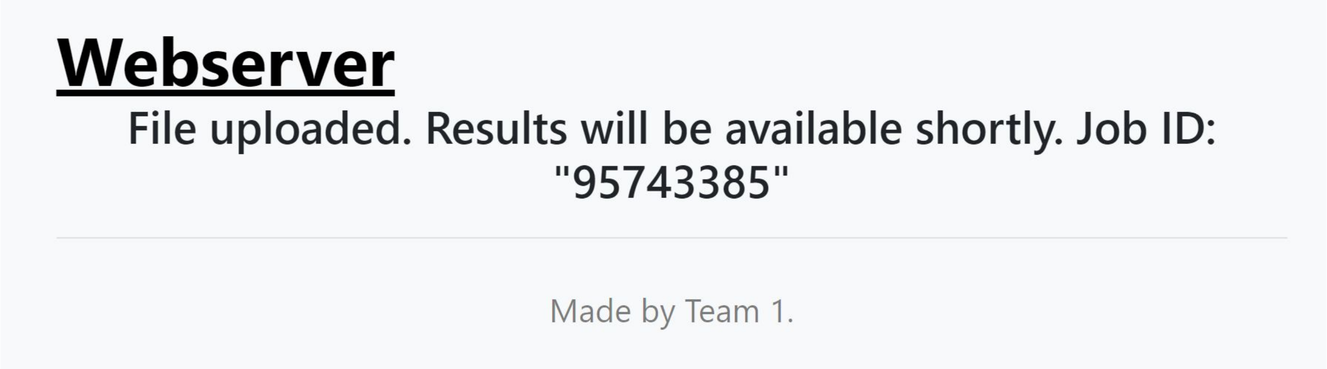
Email ID

abc@example.com

Enter valid Job ID to fetch results for

Job ID

Made by Team 1.

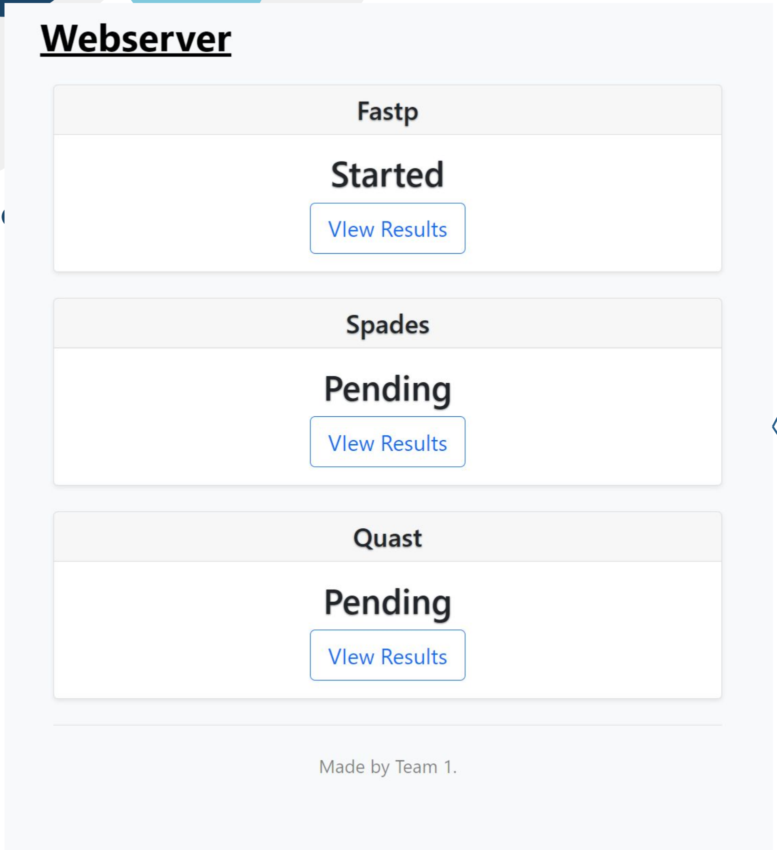


Webserver

File uploaded. Results will be available shortly. Job ID: "95743385"

Made by Team 1.

Webserver



Fastp

Started

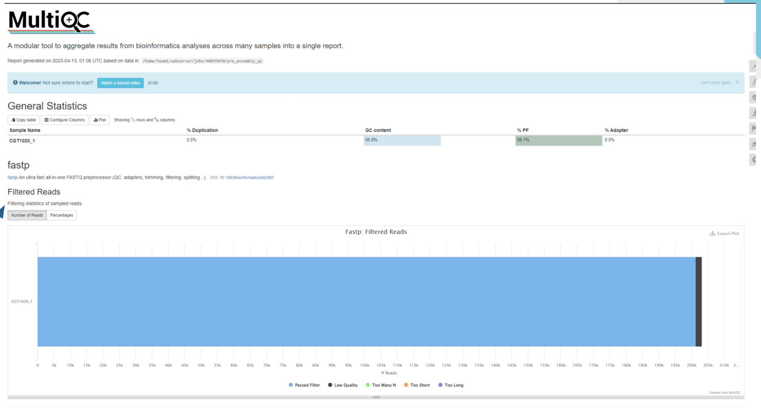
Spades

Pending

Quast

Pending

Made by Team 1.



MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2023-04-13, 01:08 UTC based on data in: `/home/team1/webserver/jobs/48035070/pre_assembly_qc`

Welcome! Not sure where to start? Watch a tutorial video (6:06) don't show again

General Statistics

Showing $^1/_1$ rows and $^4/_6$ columns.

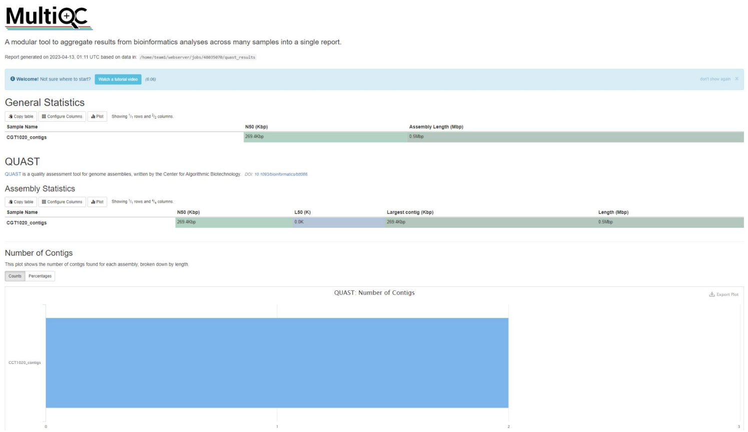
Sample Name	% Duplication	GC content	% PF	% Adapter
CGT1020_1	0.0%	50.8%	99.1%	0.0%

fastp

fastp An ultra-fast all-in-one FASTQ preprocessor (QC, adapters, trimming, filtering, splitting...). DOI: 10.1093/bioinformatics/bty560.

Filtered Reads

Filtering statistics of sampled reads.



MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2023-04-13, 01:11 UTC based on data in: `/home/team1/webserver/jobs/48035070/quast_results`

Welcome! Not sure where to start? Watch a tutorial video (6:06) don't show again

General Statistics

Showing $^1/_1$ rows and $^2/_2$ columns.

Sample Name	N50 (Kbp)	Assembly Length (Mbp)
CGT1020_contigs	269.4Kbp	0.5Mbp

QUAST

QUAST is a quality assessment tool for genome assemblies, written by the Center for Algorithmic Biotechnology. DOI: 10.1093/bioinformatics/btt086.

Assembly Statistics

Showing $^1/_1$ rows and $^4/_4$ columns.

Sample Name	N50 (Kbp)	L50 (K)	Largest contig (Kbp)	Length (Mbp)
CGT1020_contigs	269.4Kbp	0.0K	269.4Kbp	0.5Mbp

Number of Contigs

This plot shows the number of contigs found for each assembly, broken down by length.

End