Subinoy Biswas, Ph.D.
Lead Genomics Data Scientist

Nassir Patel, MBA
Chief Marketing Officer

Rajesh Perianayagam, Ph.D.
Chief Executive Officer

Unleashing your Innovation Potential

Data management and analytics platform in microbial genomics

PHYTOBIOMES ALLIANCE

Karyosoft
Augmented Innovations

All content for Phytobiomes Alliance webinar purposes only

Introduction

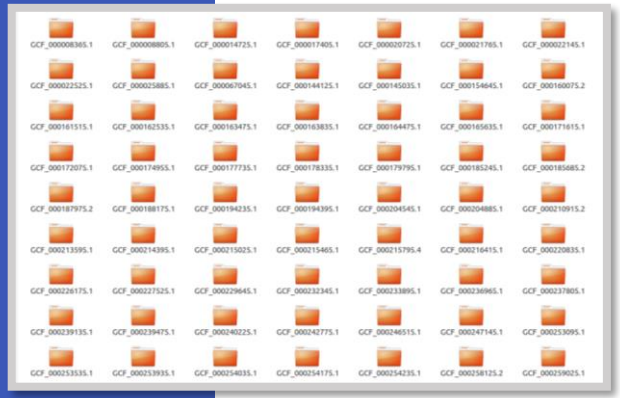
Introduction

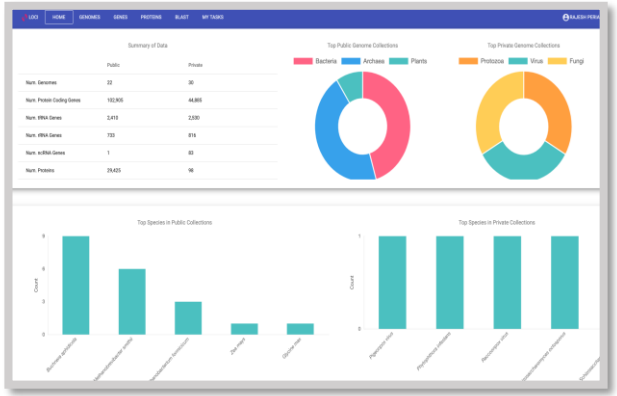
Typical applications within agricultural microbial innovations

Introduction

Challenges

1) Data repositories
2) Data formats
3) Private & public data collections
4) Data mining/visualization tools
5) Bioinformatician’s time





Case Study 1 | Gene Discovery
Patentable gene for pesticidal resistance through data mining

Case Study 2 | Strain Engineering
Discover mutation in a gene through SNP variant

Case Study 3 | Microbial Discovery
Understanding of Coronavirus of interest

Loci Overview
Karyosoft’s genomics data analytics and management platform

Benefits & Value
For Scientists; For Organizations

Summary

Case Study 1

Gene Discovery

Goal

Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty

Challenges

1) Querying public data because of data privacy
2) Mine the patent database
3) Data dispersion over many repositories both within and outside of organizations
4) No immediate availability bioinformaticians due to a lot of backlog work supporting various projects

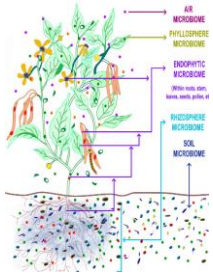
Note: data source from NCBI

Case Study 1

Typical Workflow

Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty

Sample preparation → Genome Sequencing → Assembly & Annotation → Data access and Mining → Microbial Discovery → Downstream Lab work

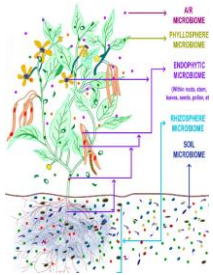


Case Study 1

Typical Workflow

Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty

Sample preparation → Genome Sequencing → Assembly & Annotation → Data access and Mining → Microbial Discovery → Downstream Lab work

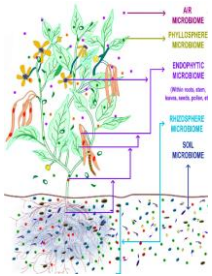


Case Study 1

Loci in your internal workflow

Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty

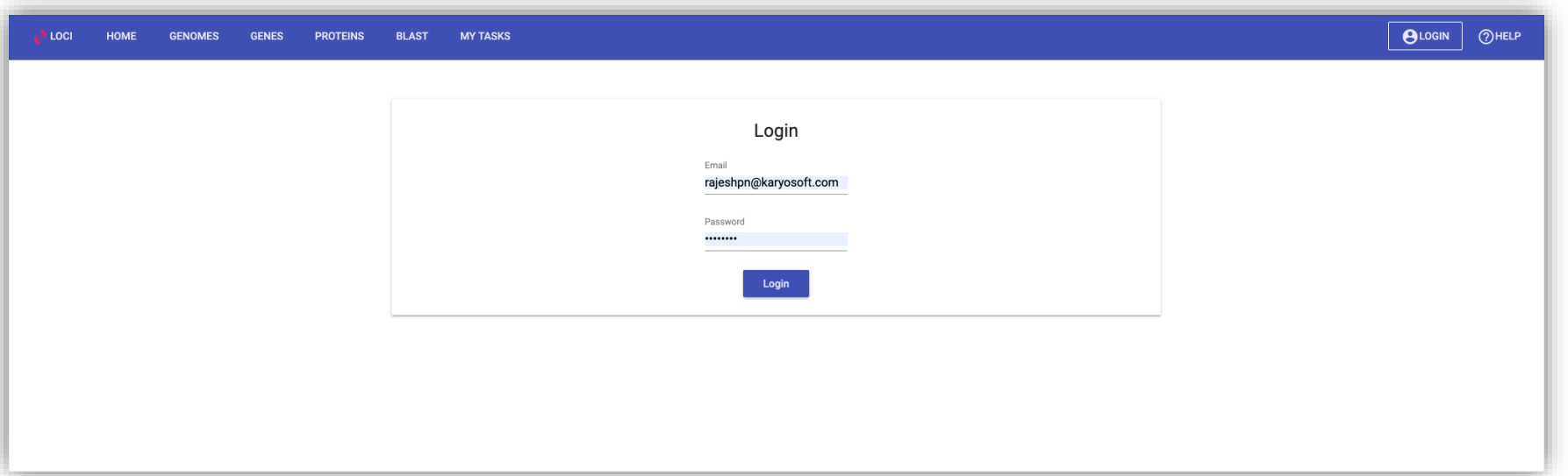
Sample preparation → Genome Sequencing → Assembly & Annotation → On-premise or cloud server: A central repository for data access and mining (Search, Integrated Visualization, Integrated Analytics) → Microbial Discovery → Downstream Lab work



Case Study 1

Login to Loci within your secured environment

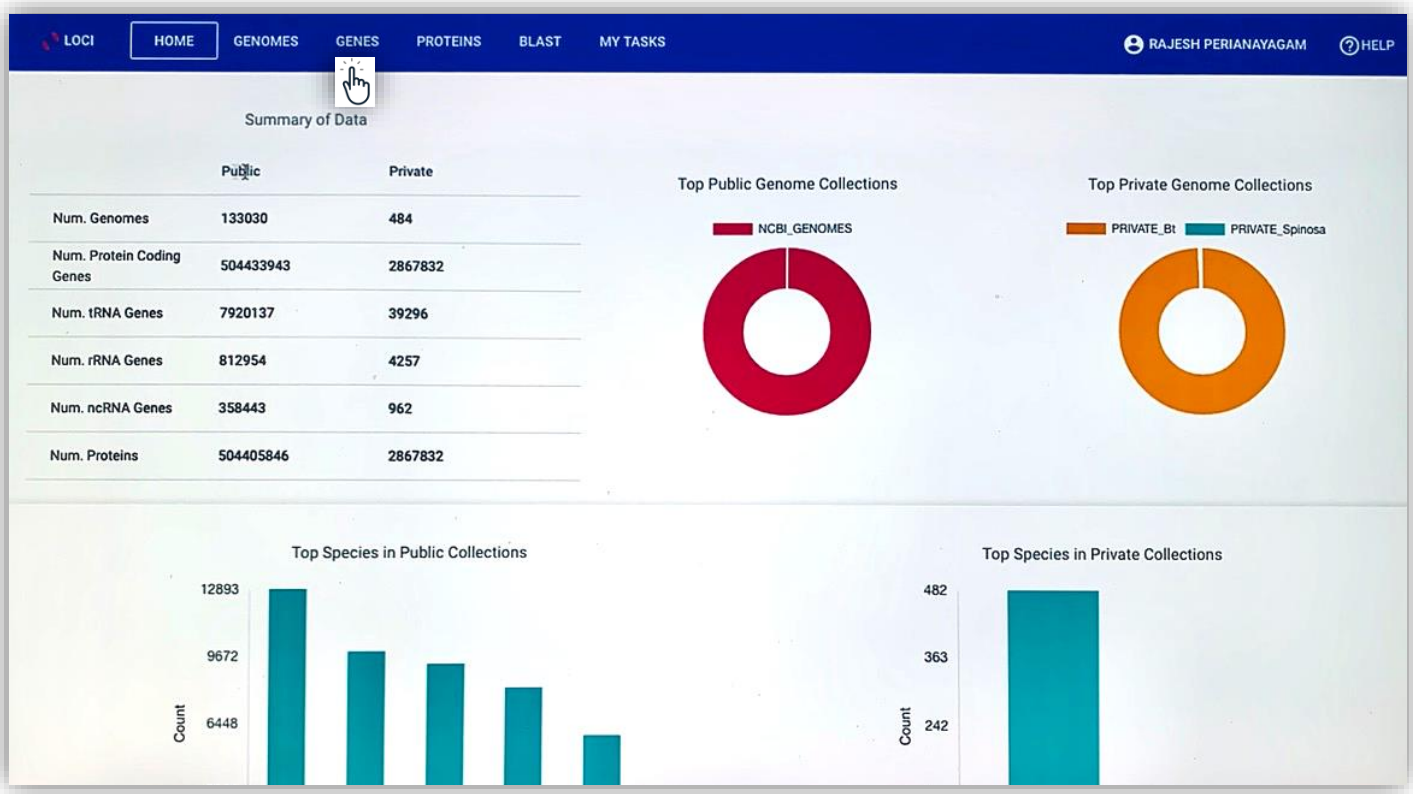
Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



Case Study 1

Data management of WGS data of all bacteria within your secured environment

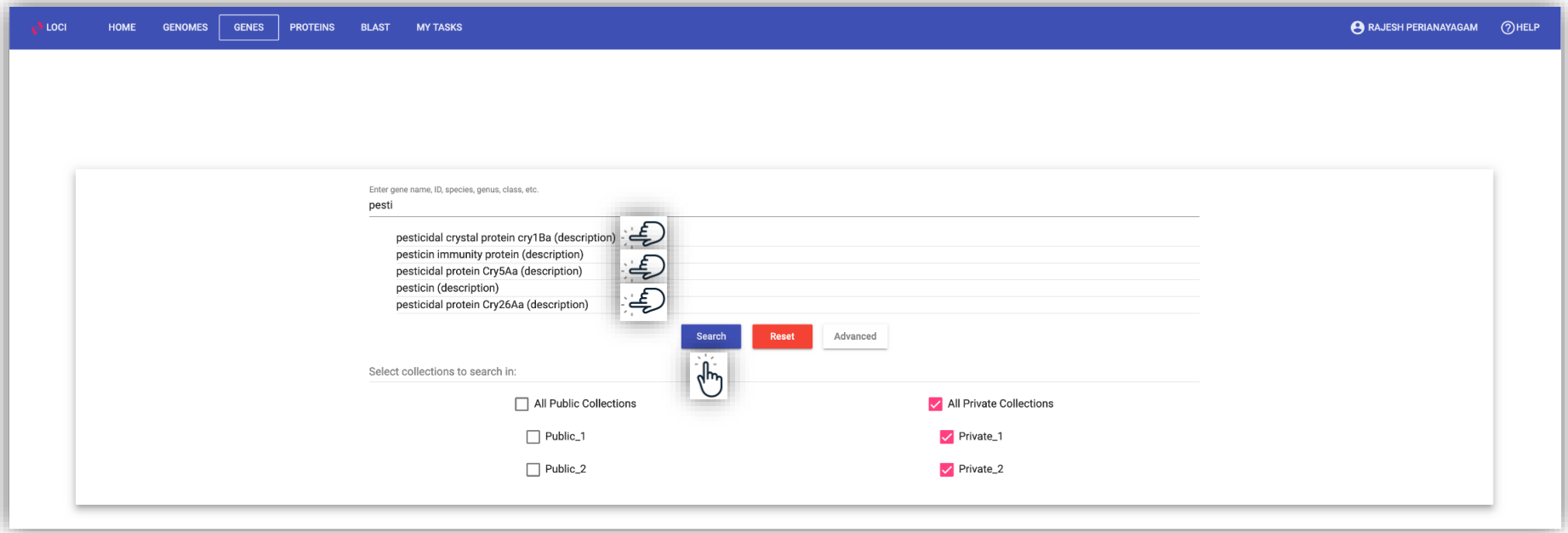
Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



Case Study 1

i) How many pesticidal genes are in private collections?

Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty

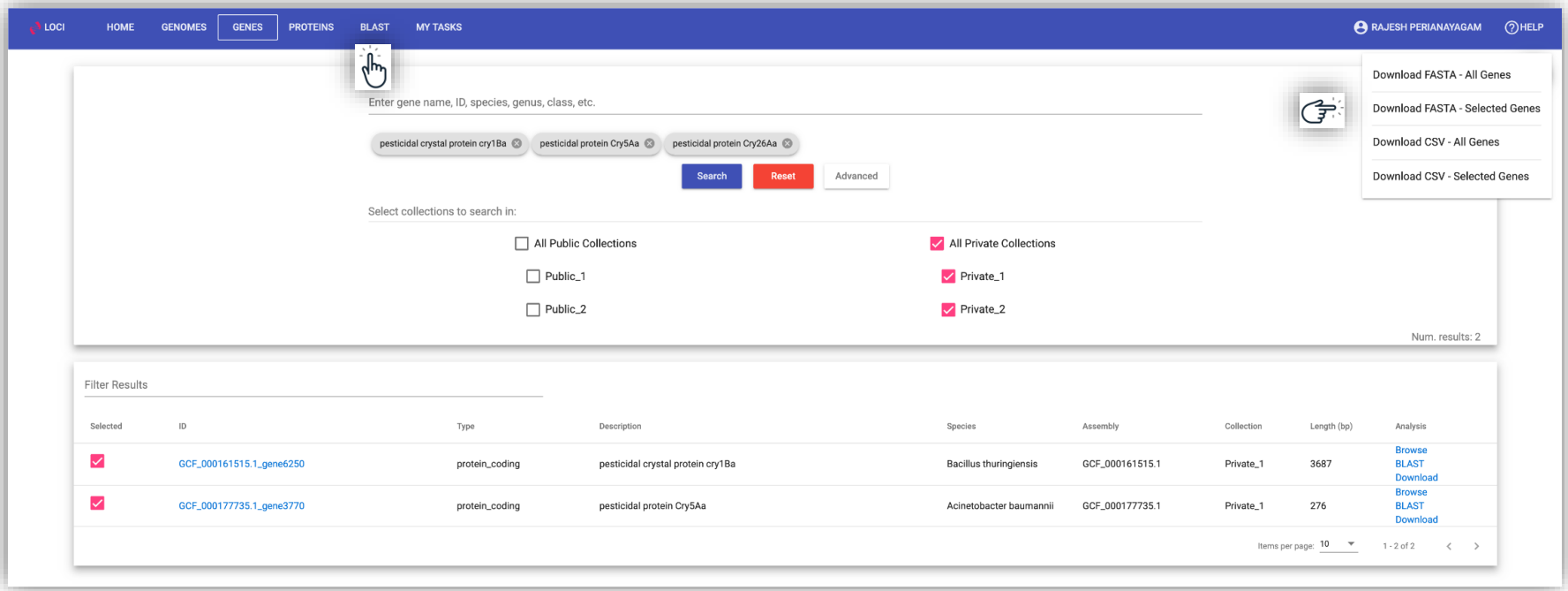


Search for the genes by description

Case Study 1

i) How many pesticidal genes are in private collections?

Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty

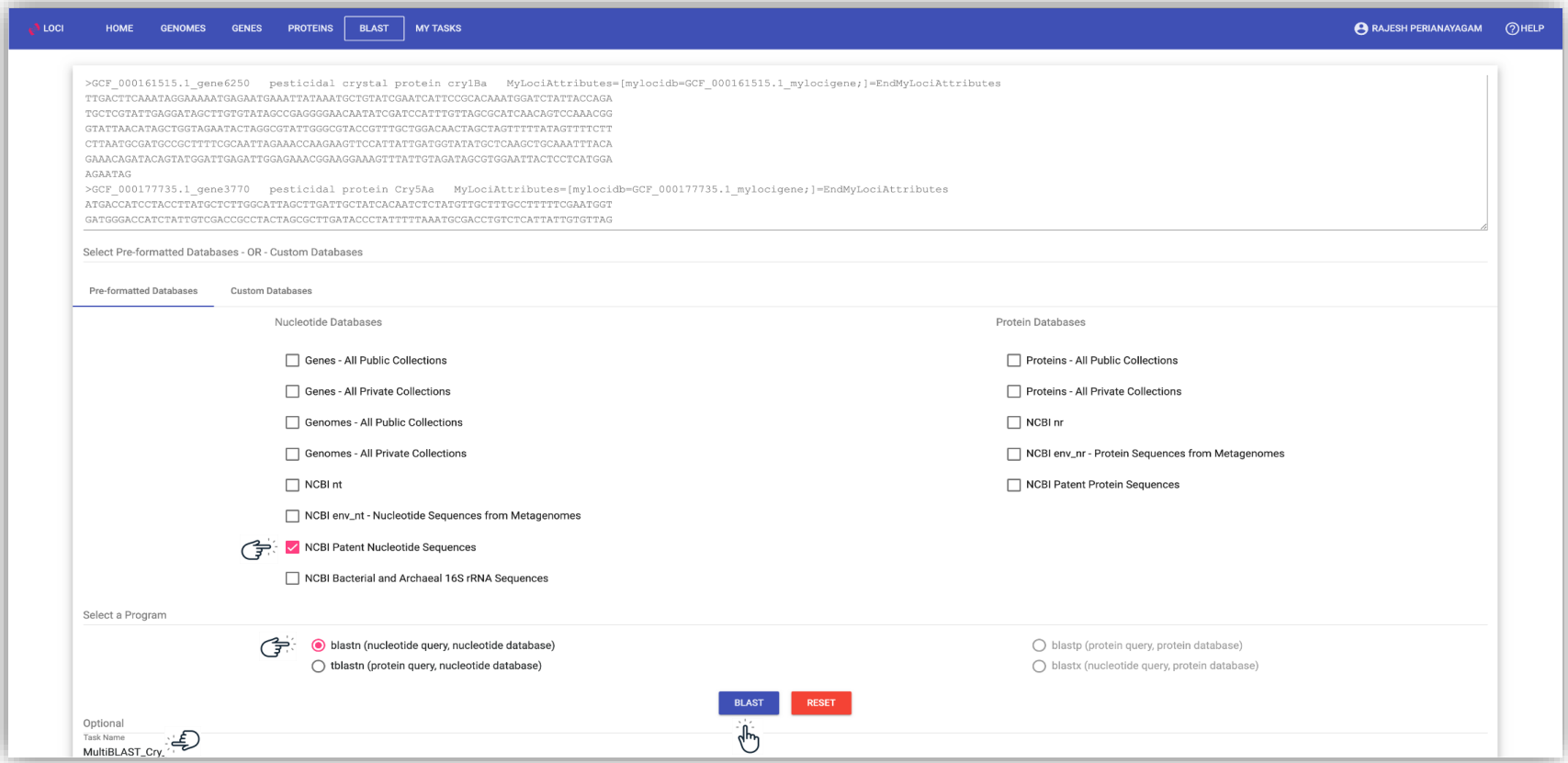


Discovery of 2 pesticidal genes in their private collections

Case Study 1

ii) Are *cry1Ba* and *cry5Aa* genes in private collections patentable?

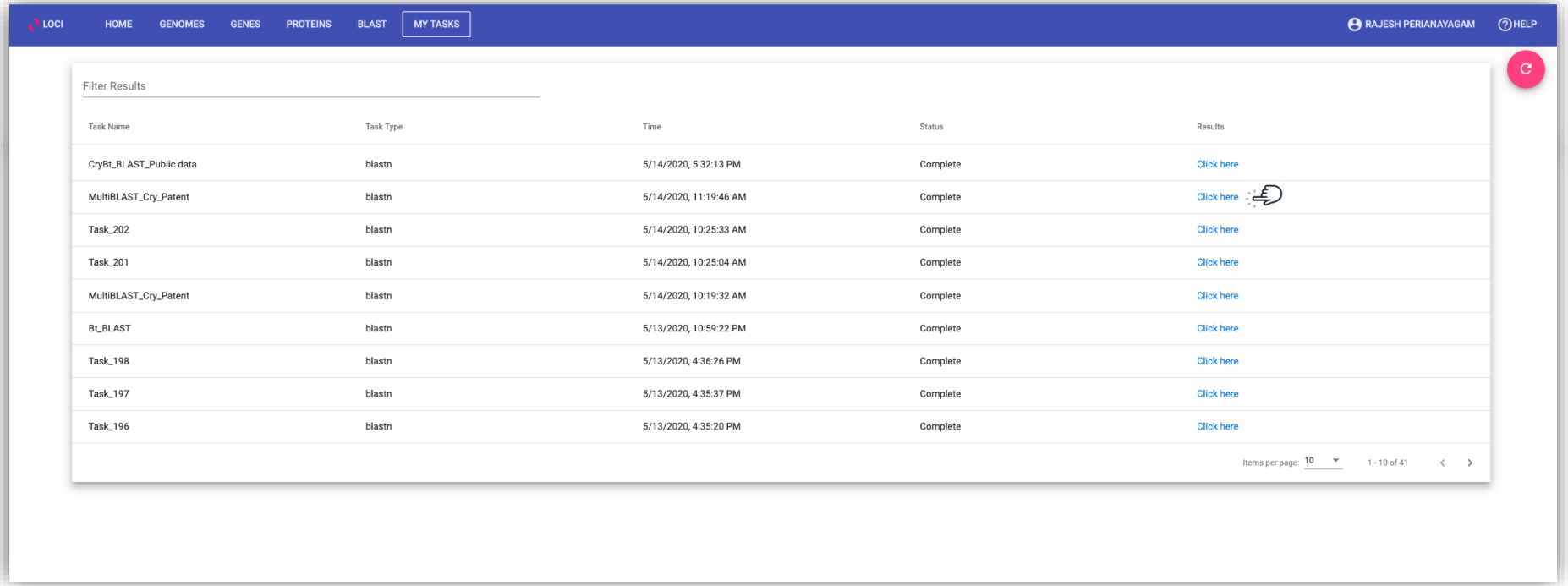
Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



Case Study 1

ii) Are *cry1Ba* and *cry5Aa* genes in private collections patentable?

Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



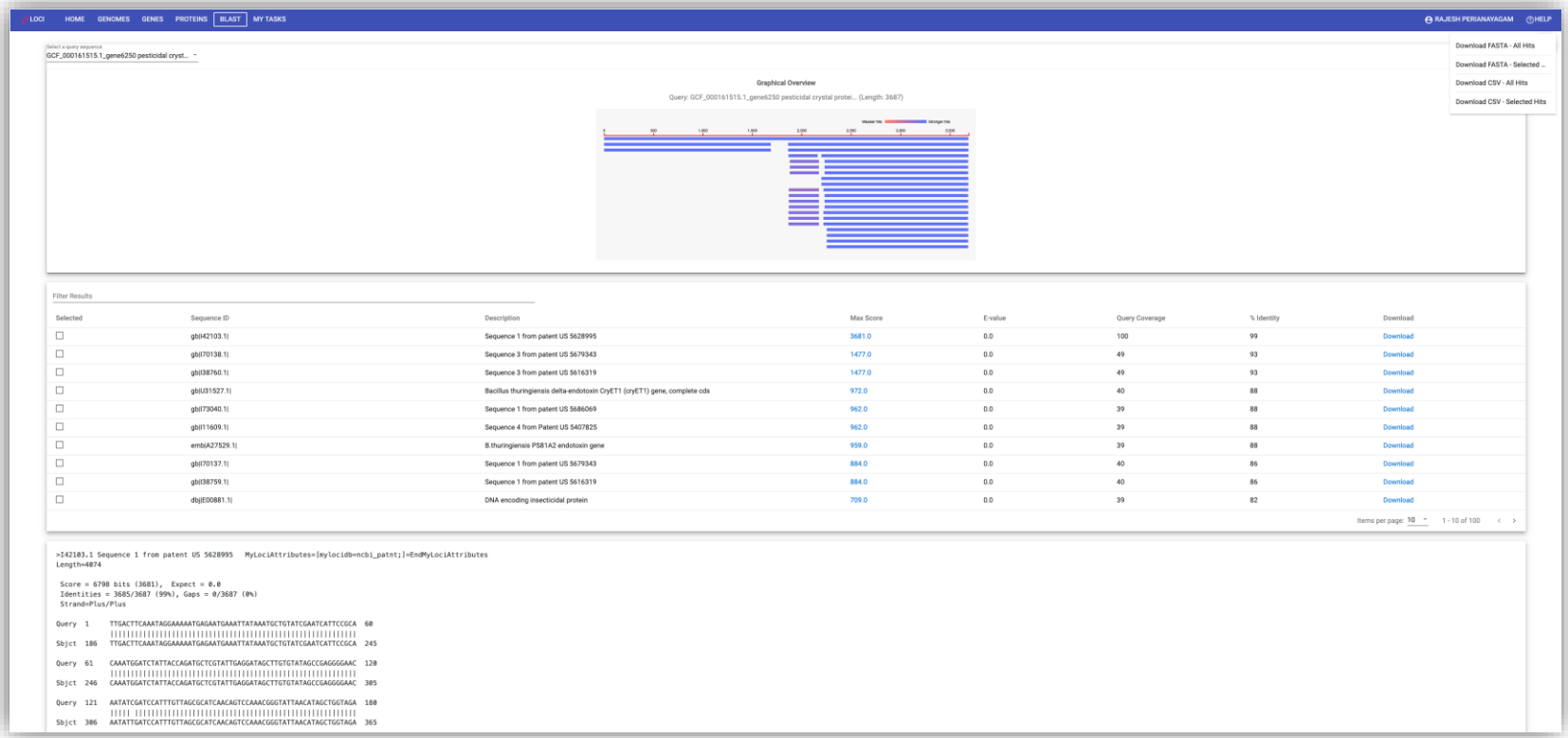
Task Name	Task Type	Time	Status	Results
CryBt_BLAST_Public data	blastn	5/14/2020, 5:32:13 PM	Complete	Click here
MultiBLAST_Cry_Patent	blastn	5/14/2020, 11:19:46 AM	Complete	Click here
Task_202	blastn	5/14/2020, 10:25:33 AM	Complete	Click here
Task_201	blastn	5/14/2020, 10:25:04 AM	Complete	Click here
MultiBLAST_Cry_Patent	blastn	5/14/2020, 10:19:32 AM	Complete	Click here
Bt_BLAST	blastn	5/13/2020, 10:59:22 PM	Complete	Click here
Task_198	blastn	5/13/2020, 4:36:26 PM	Complete	Click here
Task_197	blastn	5/13/2020, 4:35:37 PM	Complete	Click here
Task_196	blastn	5/13/2020, 4:35:20 PM	Complete	Click here

A user-specific library of all previous jobs for future reference

Case Study 1

ii) Are *cry1Ba* in private collections patentable?

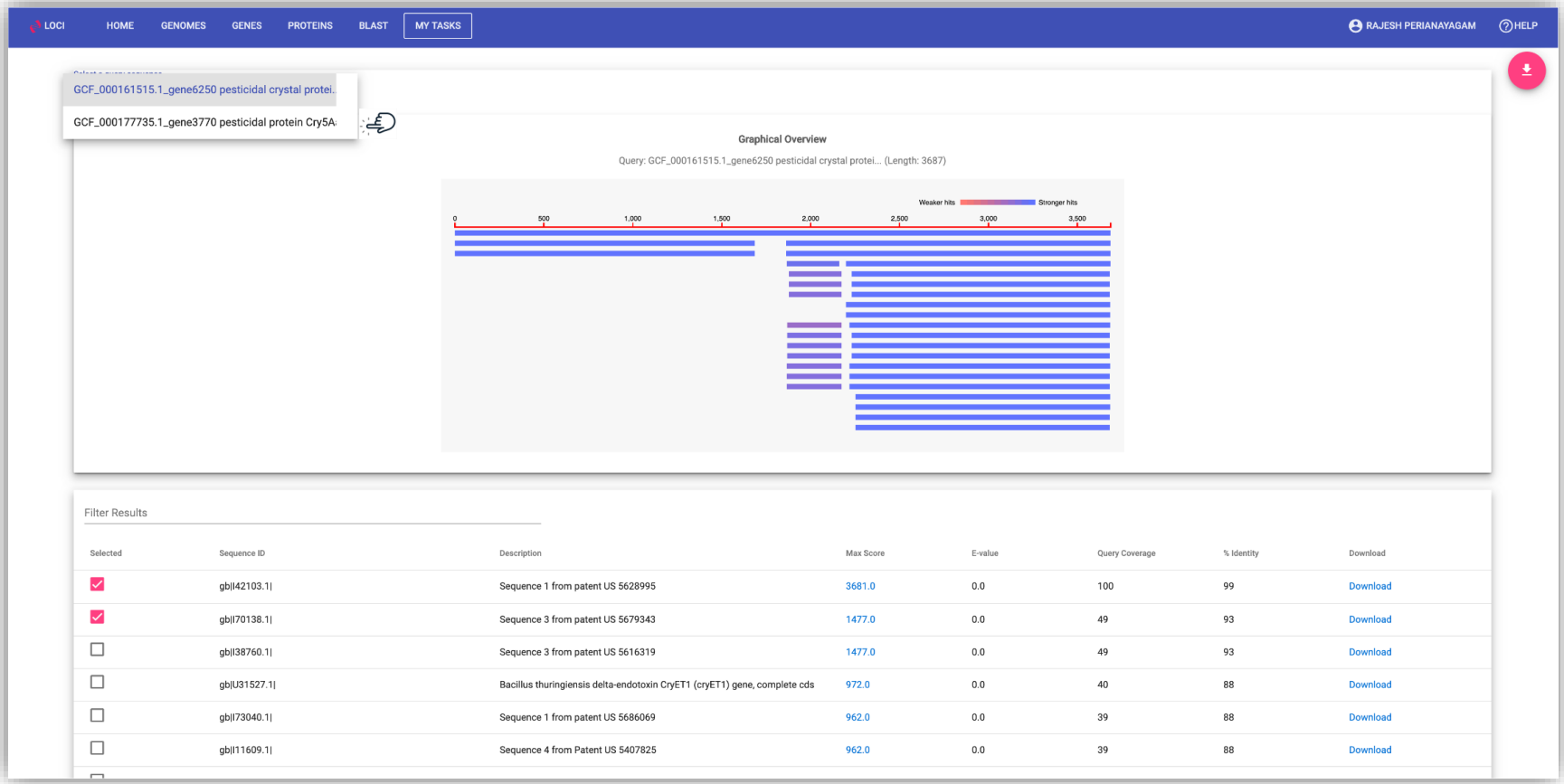
Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



Case Study 1

ii) Are *cry5Aa* in private collections patentable?

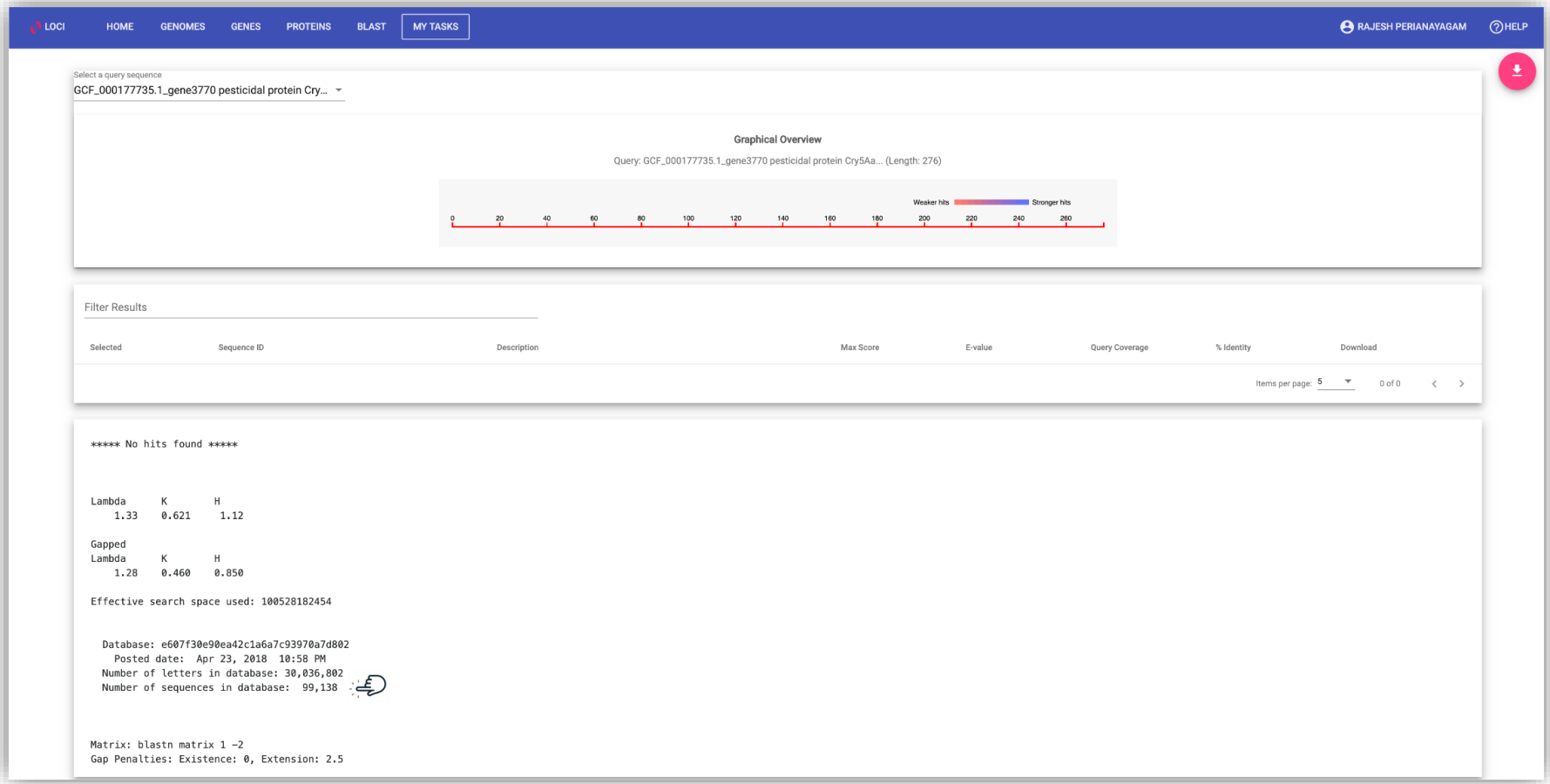
Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



Case Study 1

iii) Is *cry5Aa* gene in private collections novel?

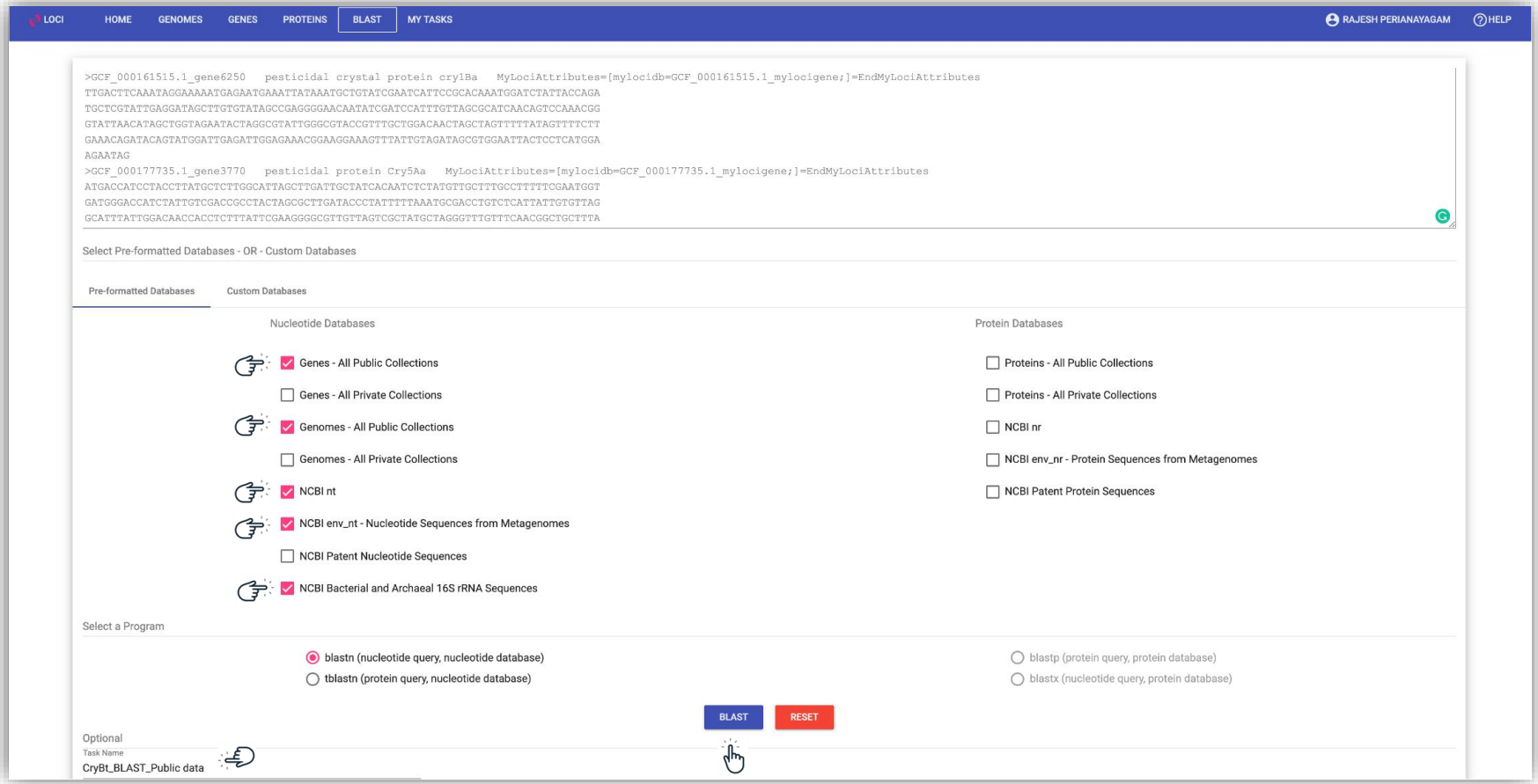
Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



Case Study 1

iii) Is *cry5Aa* gene in private collections novel?

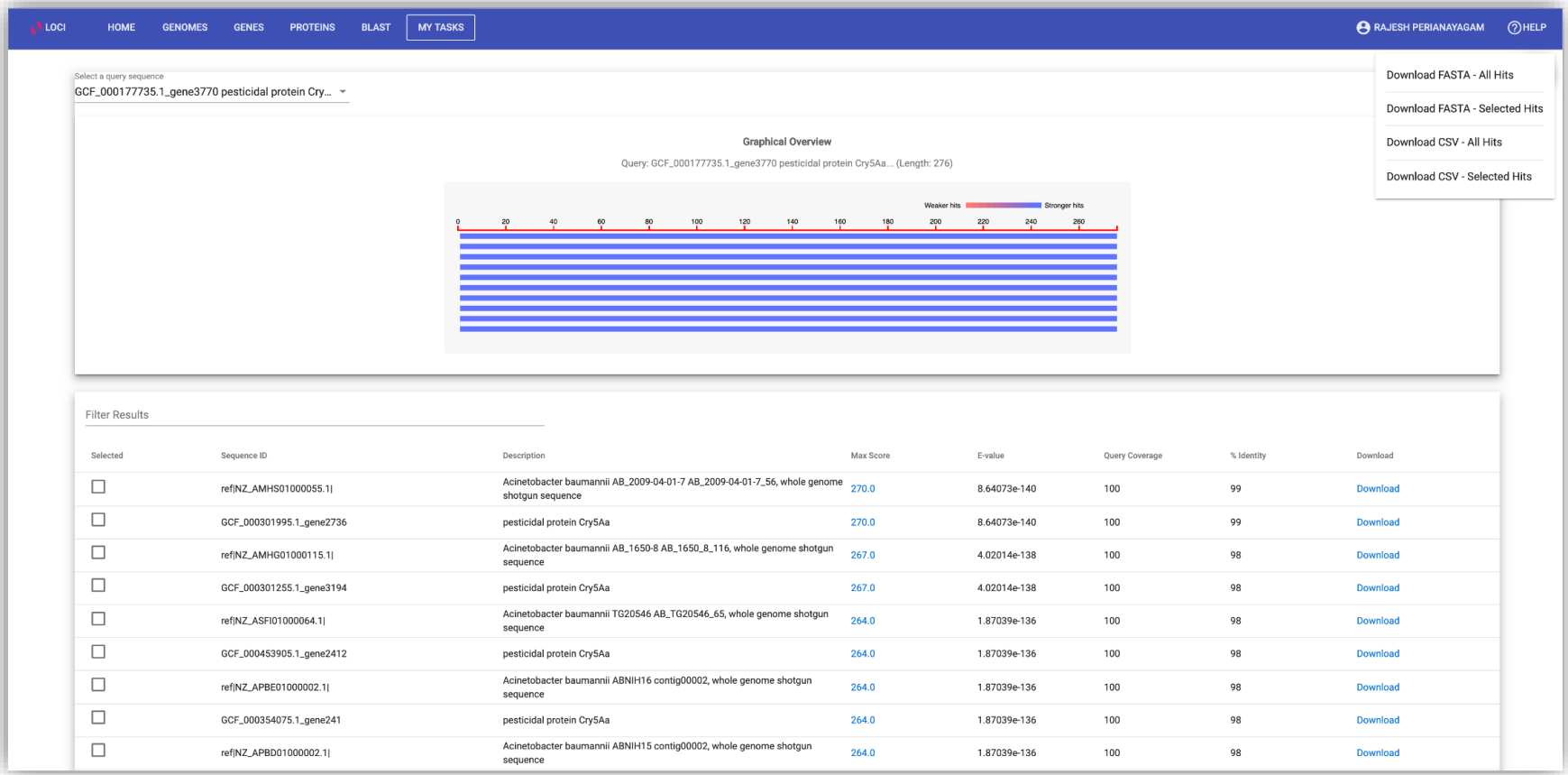
Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



Case Study 1

iii) Is *cry5Aa* gene in private collections novel?

Goal: Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty



Selected	Sequence ID	Description	Max Score	E-value	Query Coverage	% Identity	Download
☐	ref|NZ_AMHS01000055.1|	Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole genome shotgun sequence	270.0	8.64073e-140	100	99	Download
☐	GCF_000301995.1_gene2736	pesticidal protein Cry5Aa	270.0	8.64073e-140	100	99	Download
☐	ref|NZ_AMHG01000115.1|	Acinetobacter baumannii AB_1650-8 AB_1650_8_116, whole genome shotgun sequence	267.0	4.02014e-138	100	98	Download
☐	GCF_000301255.1_gene3194	pesticidal protein Cry5Aa	267.0	4.02014e-138	100	98	Download
☐	ref|NZ_ASFI01000064.1|	Acinetobacter baumannii TG20546 AB_TG20546_65, whole genome shotgun sequence	264.0	1.87039e-136	100	98	Download
☐	GCF_000453905.1_gene2412	pesticidal protein Cry5Aa	264.0	1.87039e-136	100	98	Download
☐	ref|NZ_APBE01000002.1|	Acinetobacter baumannii ABNIH16 contig00002, whole genome shotgun sequence	264.0	1.87039e-136	100	98	Download
☐	GCF_000354075.1_gene241	pesticidal protein Cry5Aa	264.0	1.87039e-136	100	98	Download
☐	ref|NZ_APBD01000002.1|	Acinetobacter baumannii ABNIH15 contig00002, whole genome shotgun sequence	264.0	1.87039e-136	100	98	Download

Case Study 1

Conclusions

Goal

Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty

1) There are 2 pesticidal genes (*Cry1Ba* and *Cry5Aa*) in their private collections

2) Out of these 2 genes, *Cry5Aa* is not patented and a novel gene - a valuable information

3) Scientists can decide to allocate and prioritize lab resources for both genes or one

Loci

Helps the scientists by accelerating their valuable gene discoveries

Innovation in <15 clicks

Note: data source from NCBI

Case Study 2

Strain Engineering

Goal

Discover mutations *in ribF** gene of *E.coli* strain 1 and compare it with 3 other strains

*Bifunctional riboflavin kinase/FMN adenylyltransferase; Catalyzes the phosphorylation of riboflavin to FMN followed by the adenylation of FMN to FAD

Challenges

1) Lack of user-friendly visualization tools or availability of stand-alone tools

2) Too many files to mine to discover a mutation; Labor intensive Process

3) No immediate availability bioinformaticians due to a lot of backlog work supporting various projects

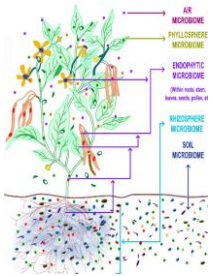
Note: data source from NCBI

Case Study 2

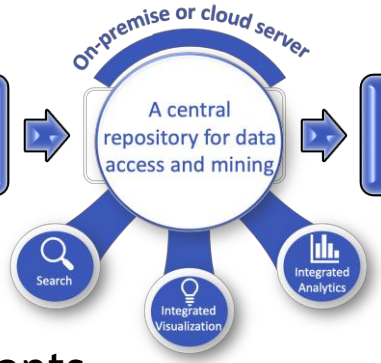
Loci in your workflow

Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains

Sample preparation → Genome Sequencing → Assembly & Annotation → Microbial Discovery → Downstream Lab work



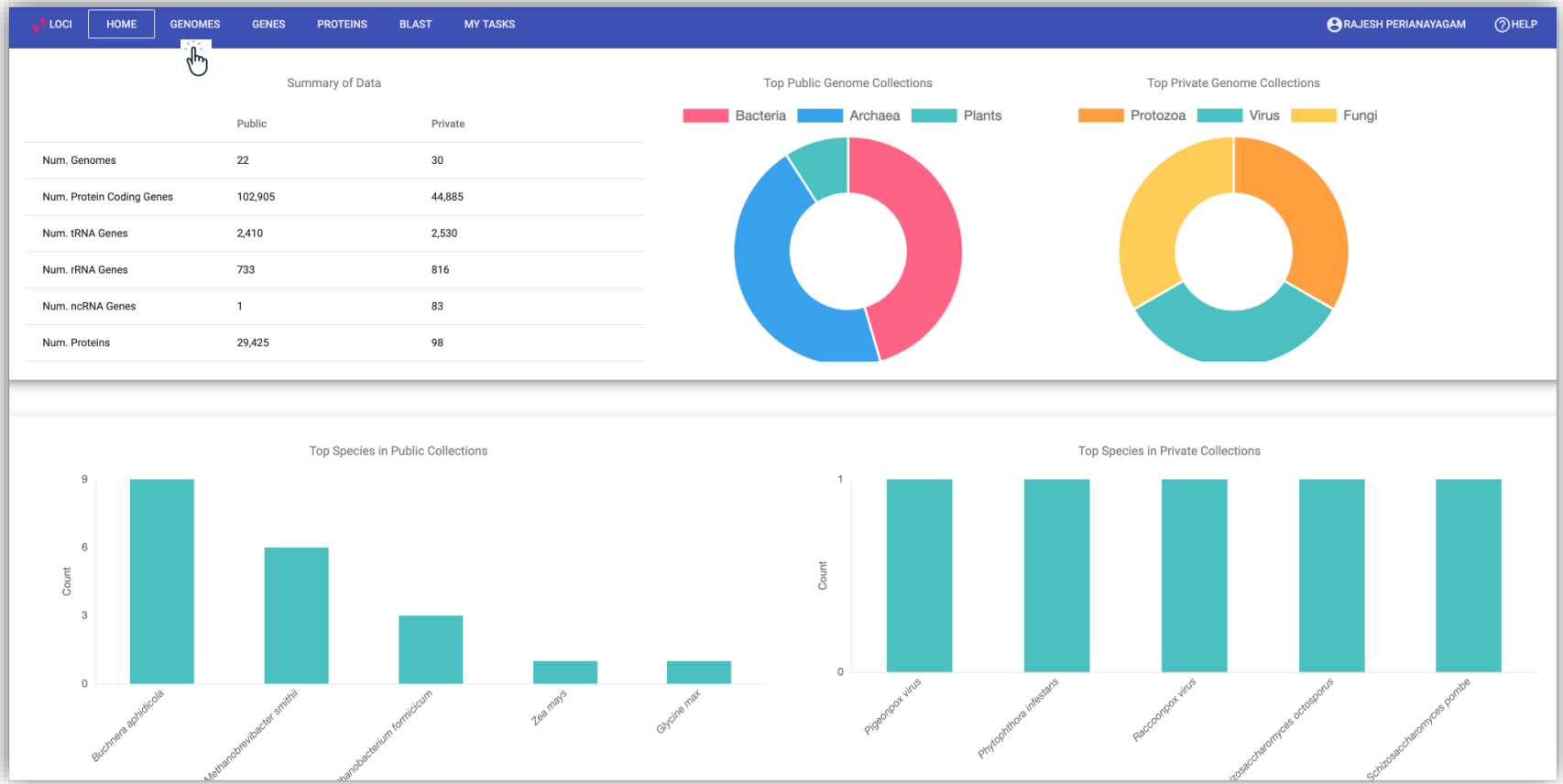
BAM
Gff3
Ref. SNP variants



Case Study 2

Data management of WGS data of multiple organisms within your secured environment

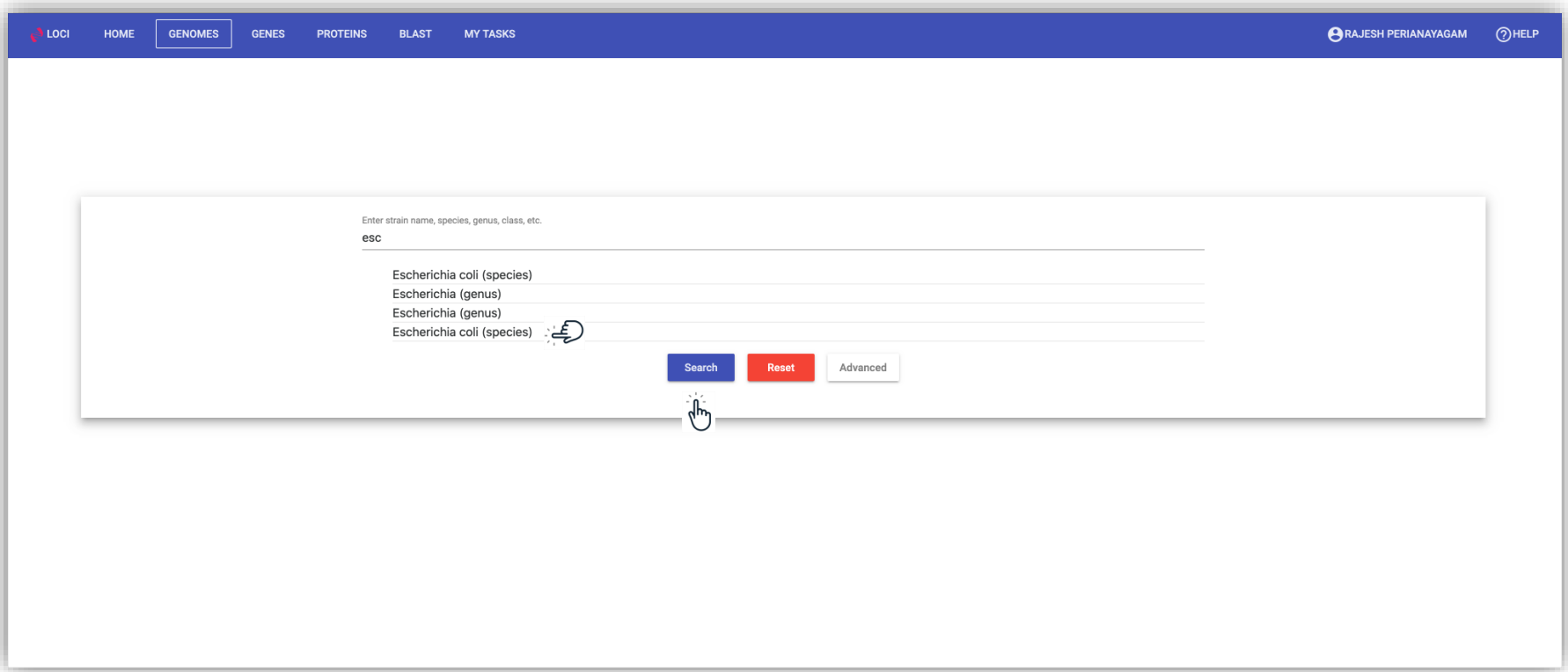
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

i) Searching for the *E.coli* reference genome

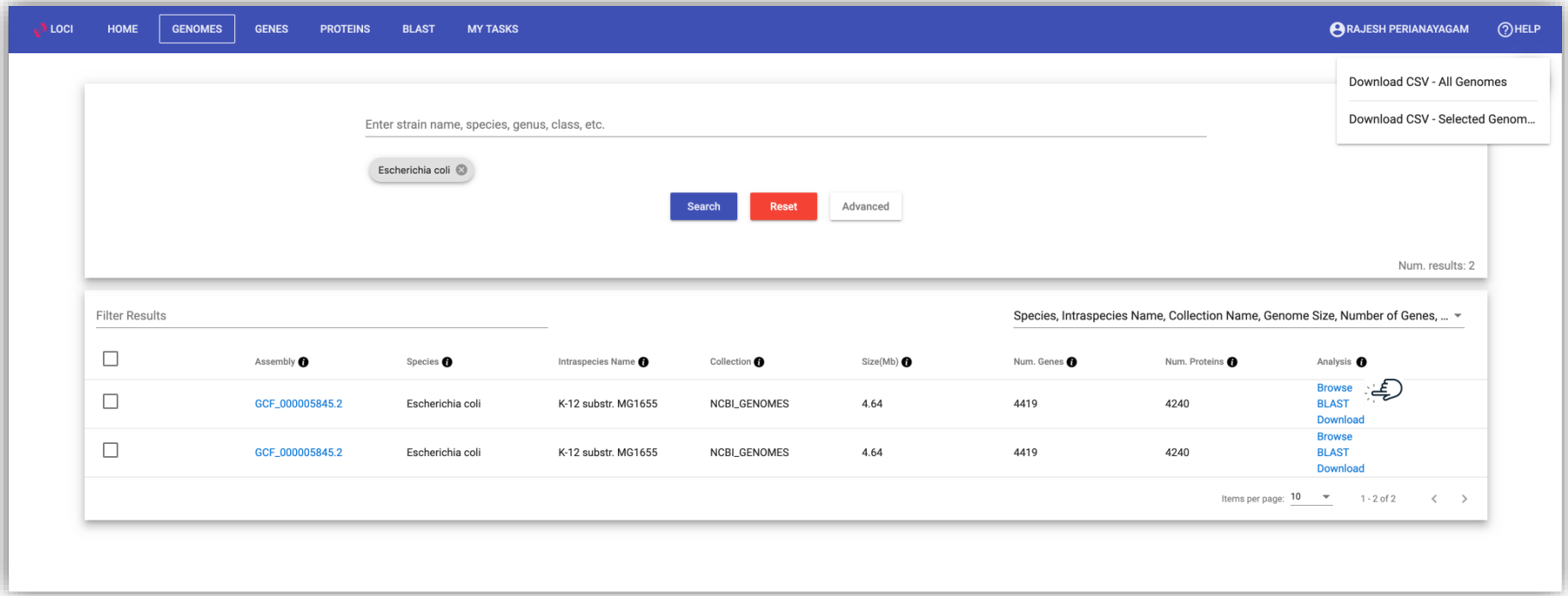
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

i) Searching for the *E.coli* reference genome

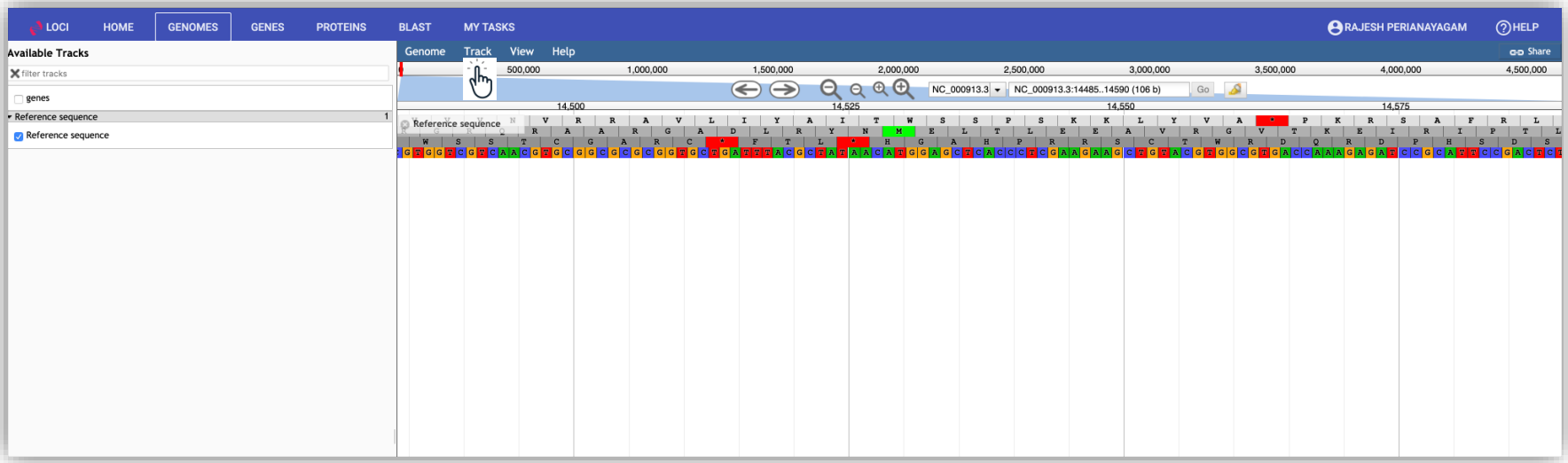
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

ii) *E.coli* Reference sequence track

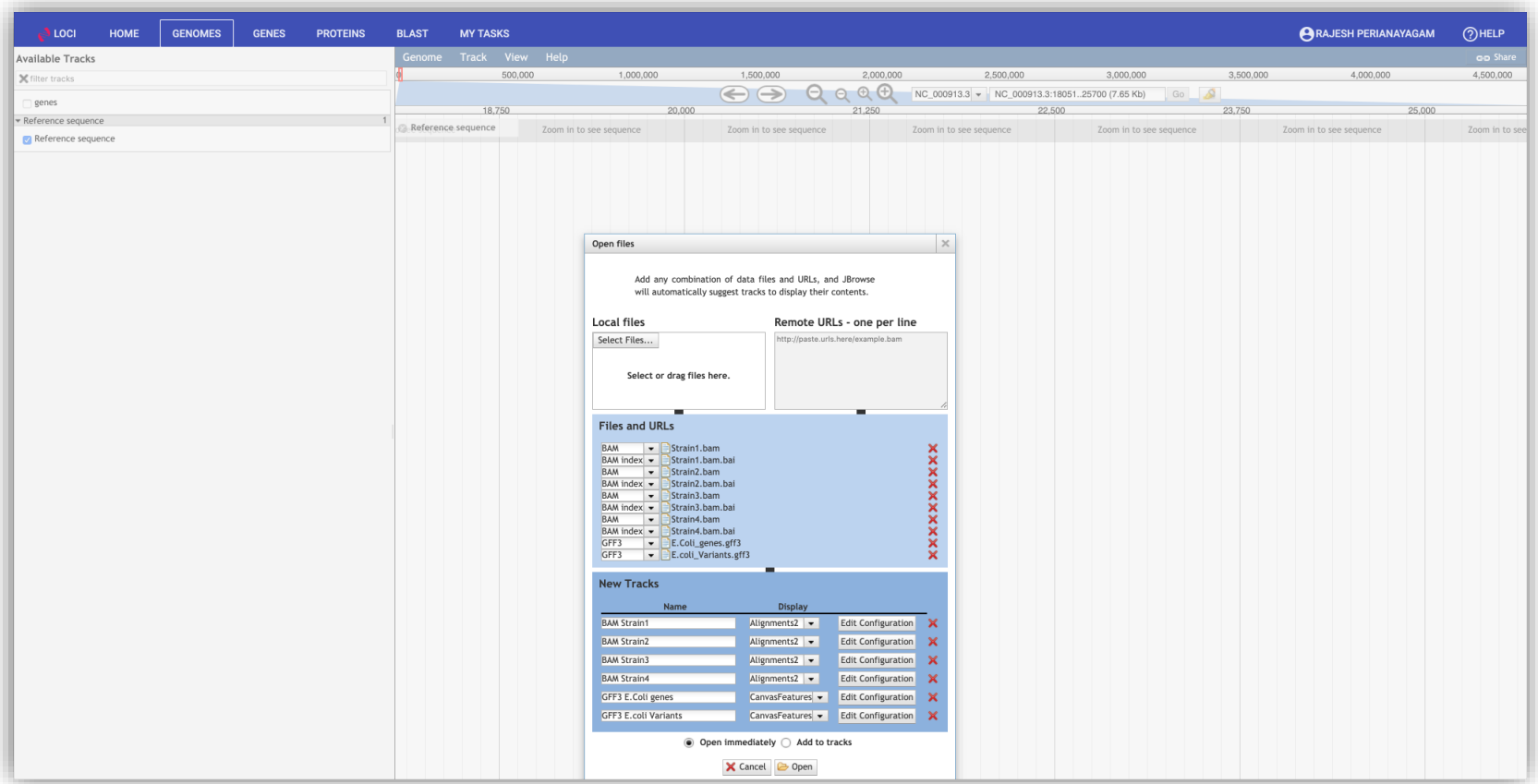
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

ii) Adding Gff3, Vcf and BAM files

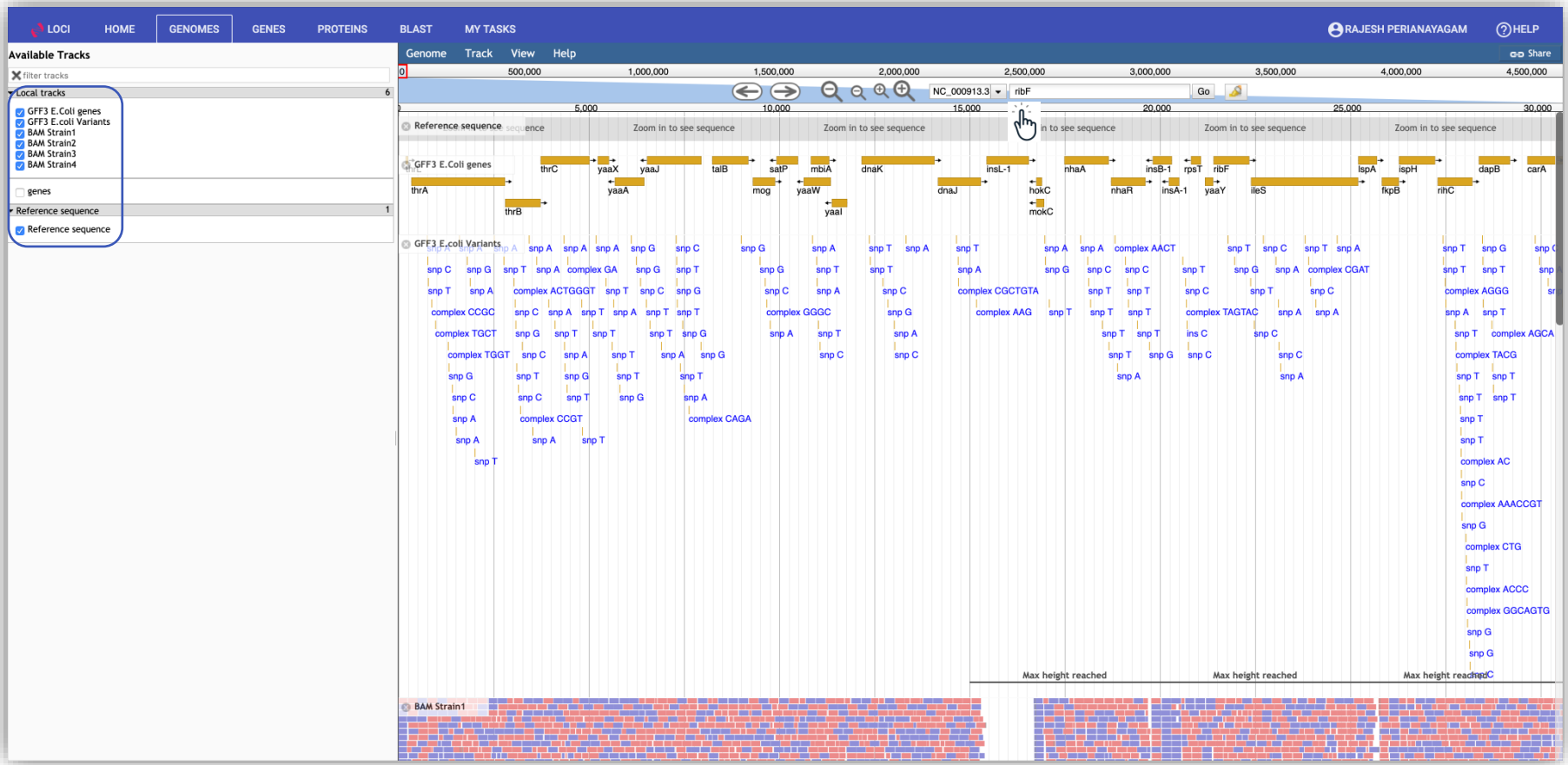
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

iii) Search for *ribF* gene

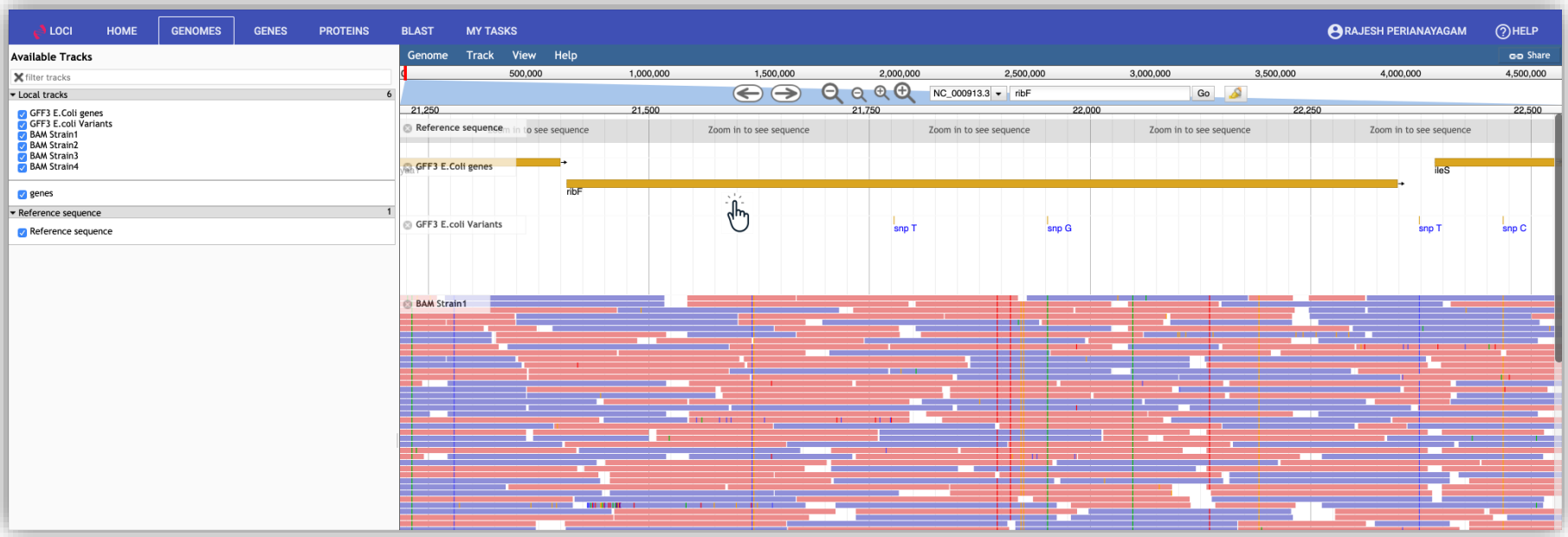
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

iii) *ribF* gene

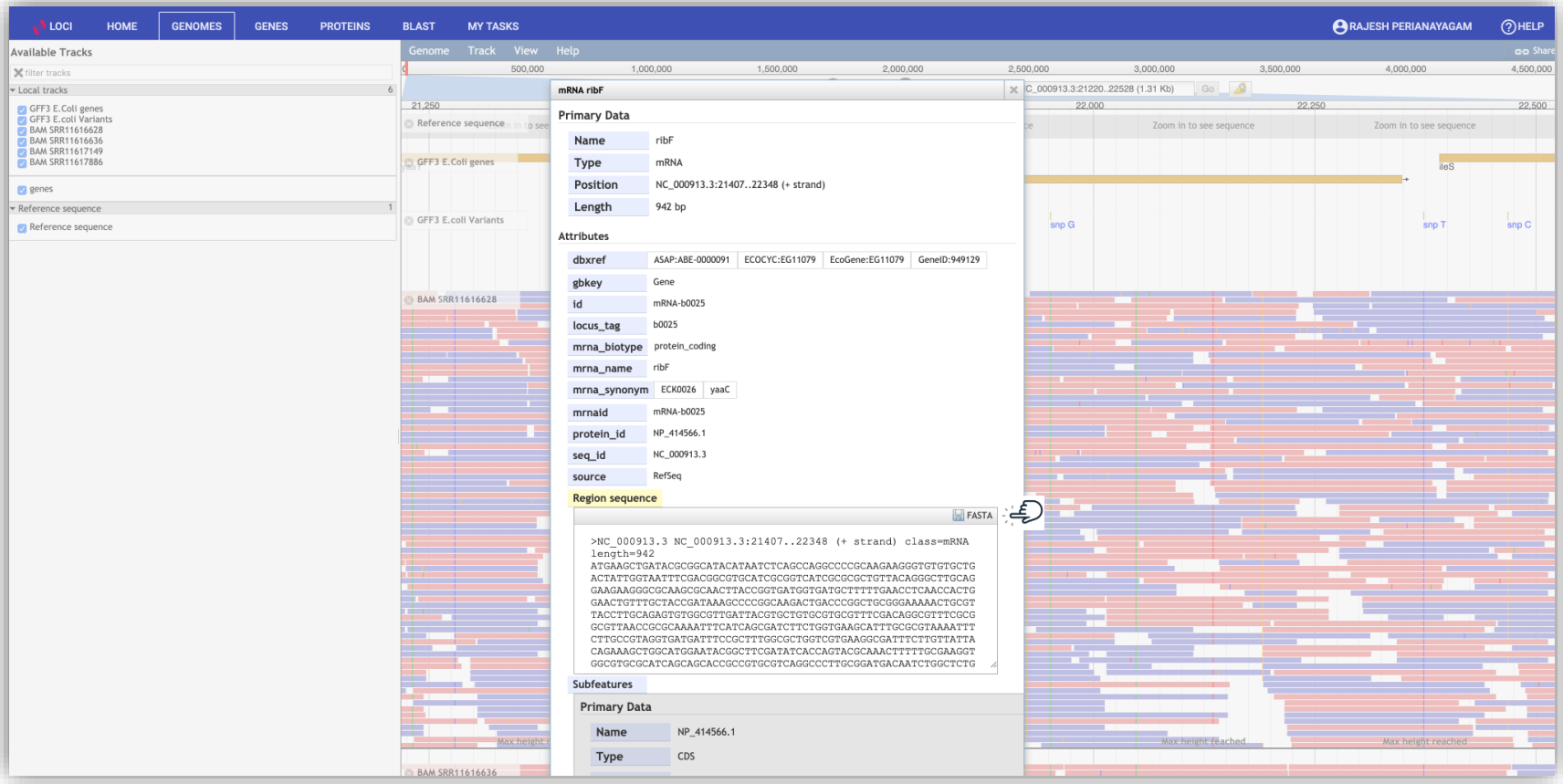
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

iv) Details of *ribF* gene

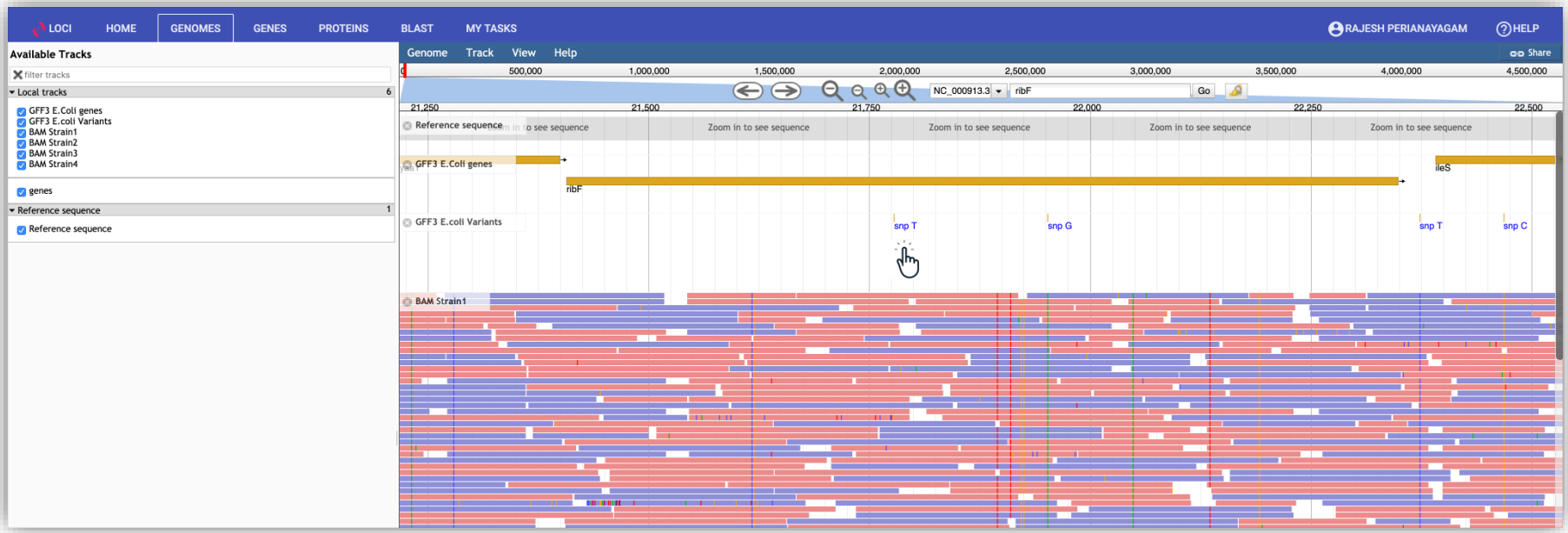
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

v) Reference SNP details

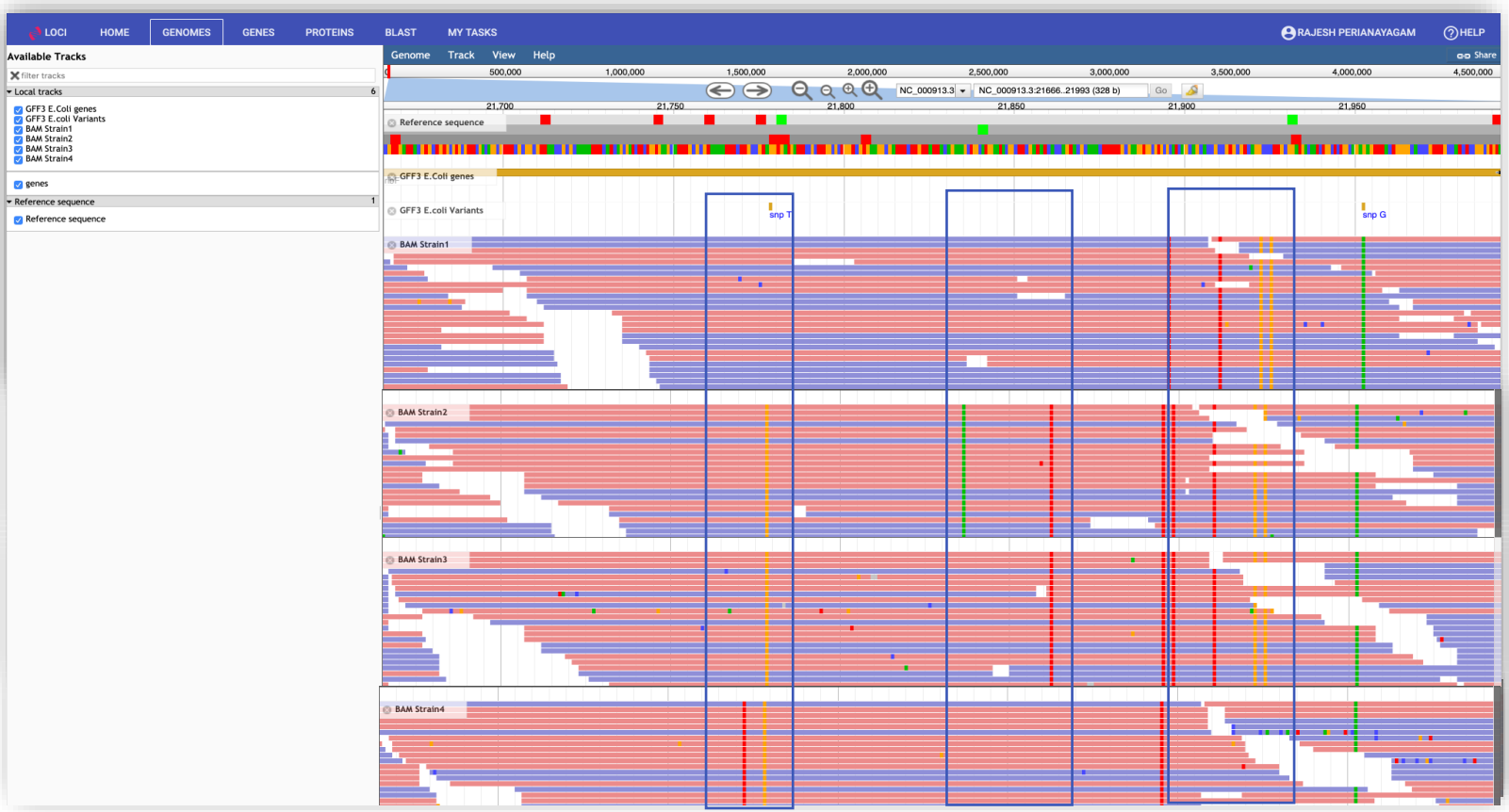
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

vi) *ribF* gene variation among 4 *E.coli* strains

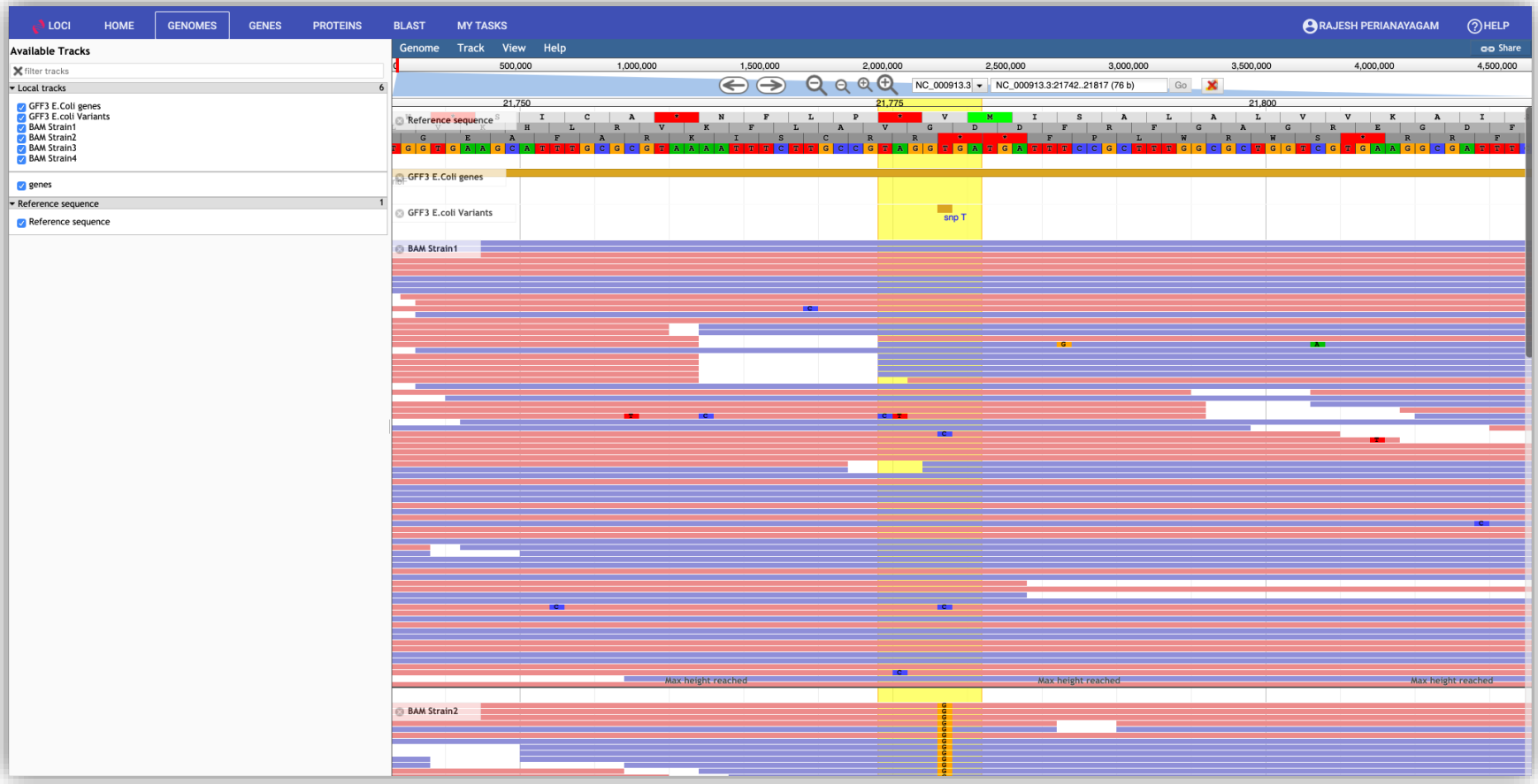
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

vii) Correlating the mutation to amino acid change instantly

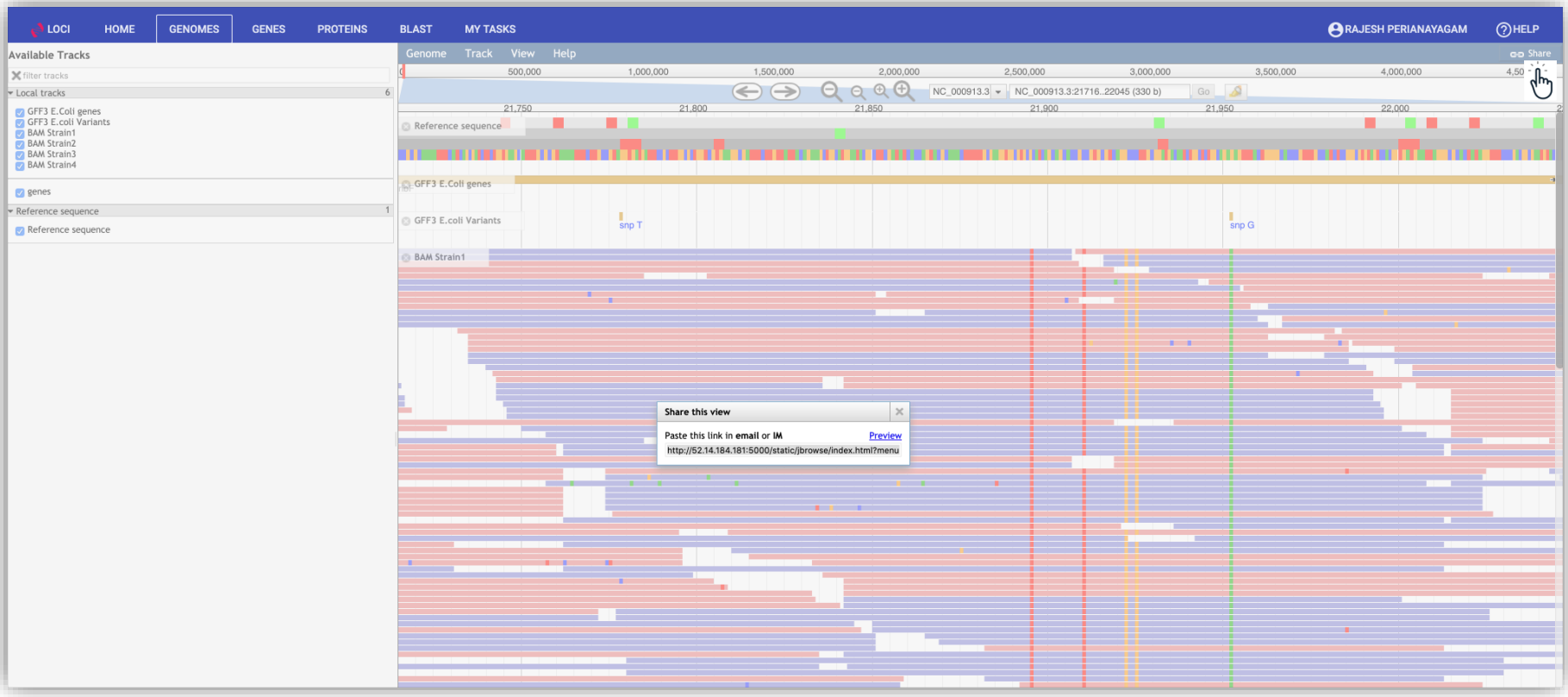
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

viii) Sharing the results among colleagues

Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



Case Study 2

Conclusions

Goal

Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains

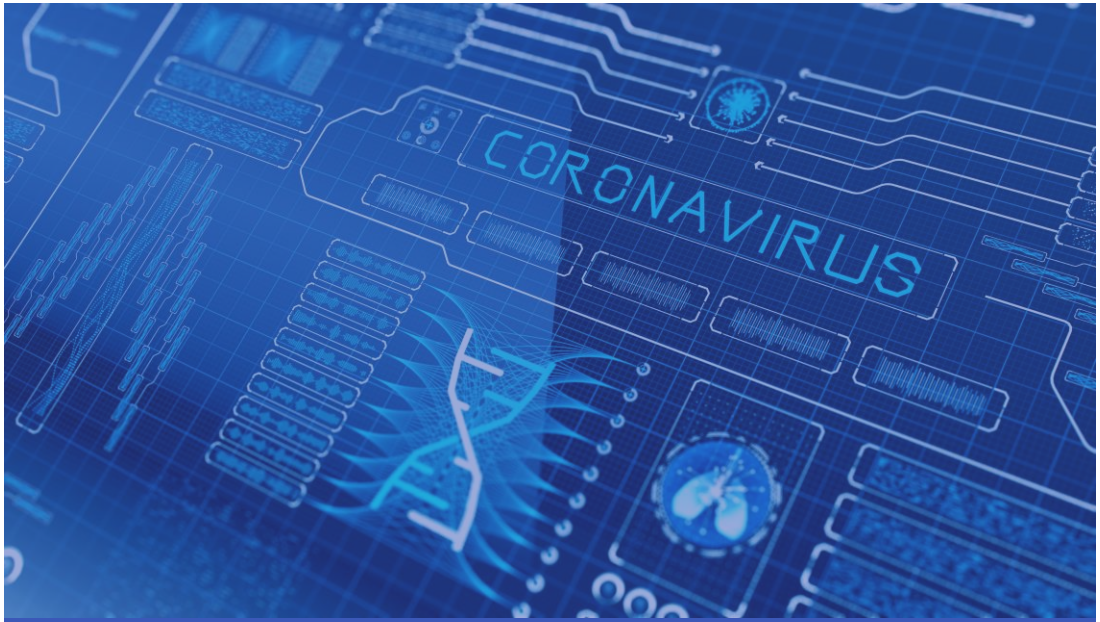
1) There are 8 mutations in the *ribF* gene by comparing multiple strains
2) One of the mutations clearly differentiates strain 1 from other strains
3) Able to correlate the mutation at the nucleotide level to amino acids and thus protein level
4) The scientist can easily share the discovery with other colleagues
5) In addition to FASTA, GFF3, BAM, scientists can analyze BigBed and BigWig data directly without any conversion

Loci

Makes complex mutation discovery easy for scientists in under 5 minutes

Case Study 3

Microbial Discovery



Goal

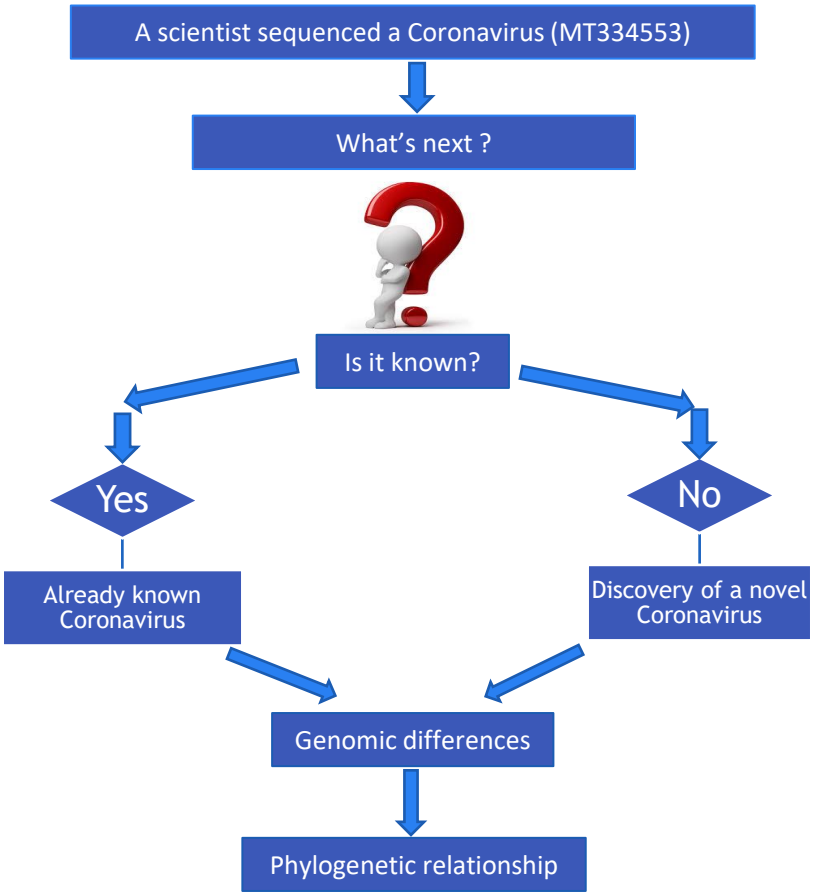
Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI

Note: data source from NCBI

Case Study 3

Project Plan

Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI



Case Study 3

Project Plan

Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI

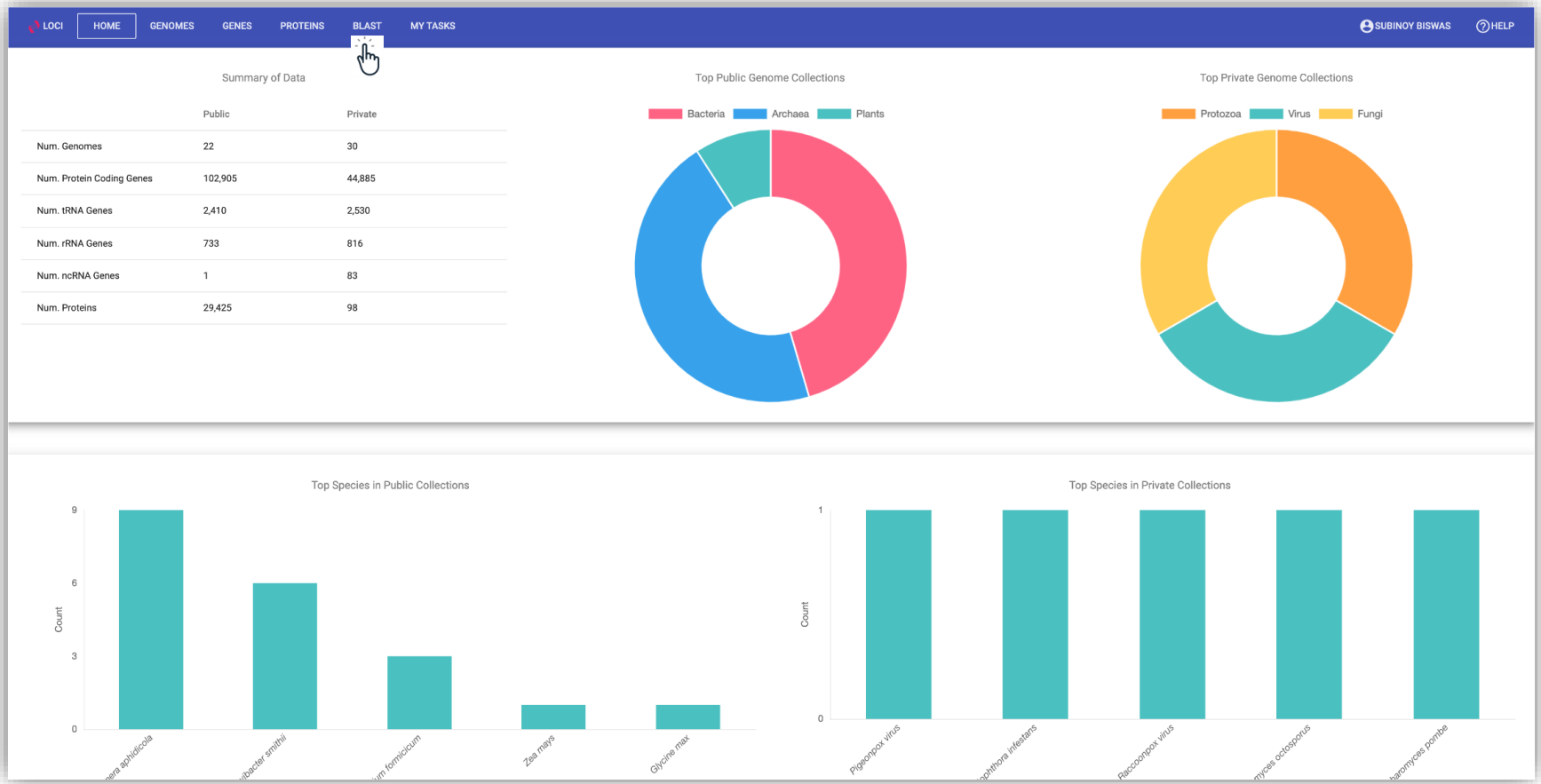
A scientist sequenced a Coronavirus (MT334553)

What next ?

Case Study 3

i) Data management of WGS data of multiple organisms within your secured environment

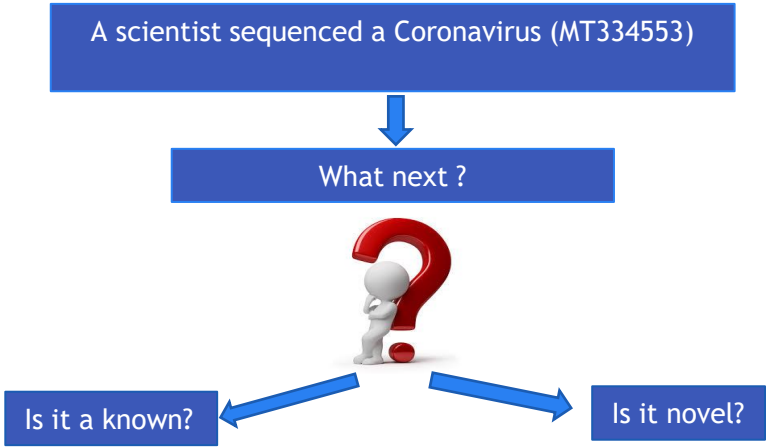
Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI



Case Study 3

Project Plan

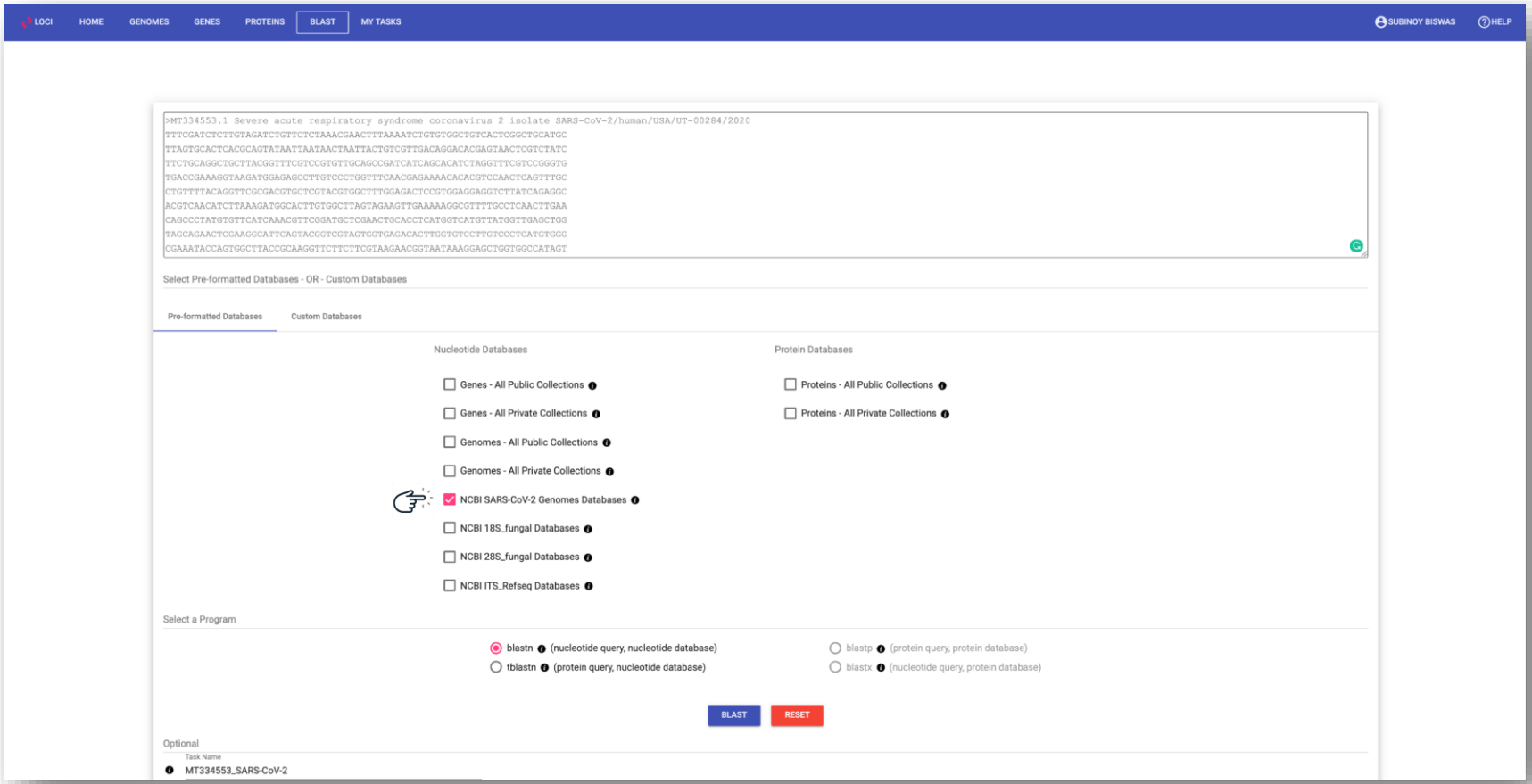
Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI



Case Study 3

ii) Homology analysis using Coronavirus database

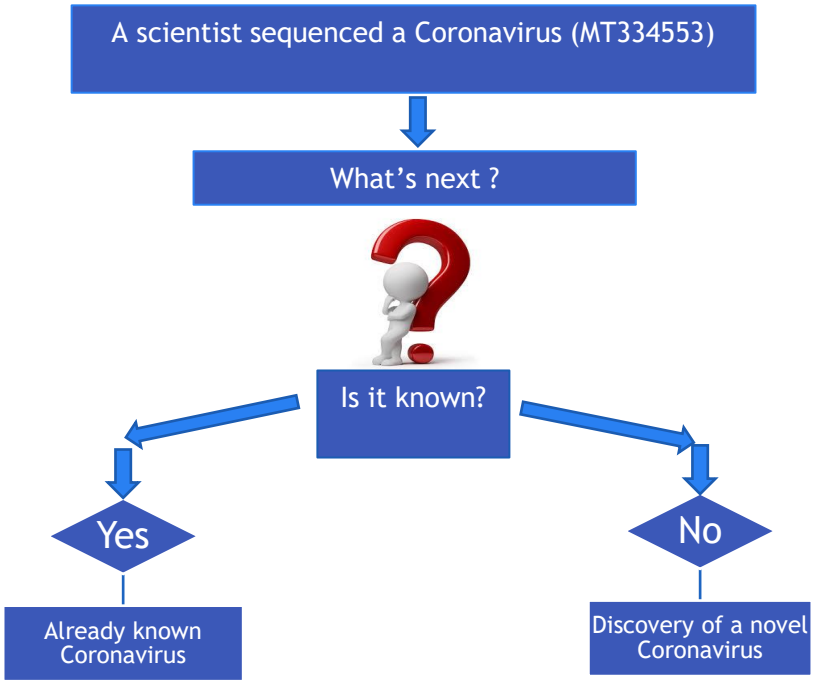
Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI



Case Study 3

Project Plan

Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI



Case Study 3

iii) High homology (SARS-CoV-2)

Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI

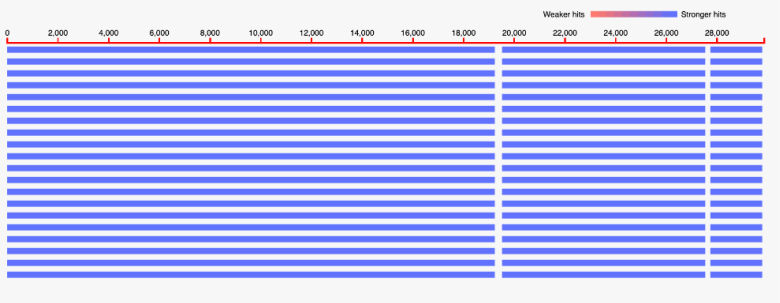
LOCI HOME GENOMES GENES PROTEINS BLAST MY TASKS SUBINOY BISWAS HELP

Select a query sequence

MT334553.1 Severe acute respiratory syndrome cor...

Graphical Overview

Query: MT334553.1 Severe acute respiratory syndrome coron... (Length: 29863)



Filter Results

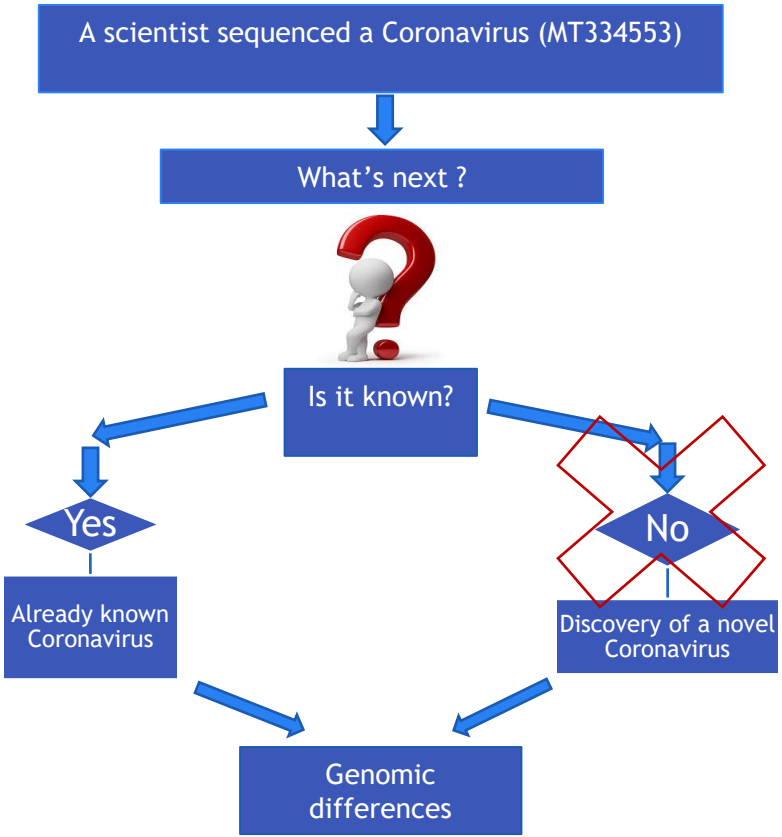
Select a percent similar to download

Sequence ID	Description	Max Score	E-value	Query Coverage	% Identity	Download
gb|MT163717|	|Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA4-UW2/2020| complete genome	19245.0	0.0	64	100	Download
gb|MT163719|	|Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA7-UW4/2020| complete genome	19245.0	0.0	64	100	Download
gb|MT163718|	|Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA6-UW3/2020| complete genome	19245.0	0.0	64	100	Download
gb|MT246466|	|Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW209/2020| complete genome	19245.0	0.0	64	100	Download
gb|MT246452|	|Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW195/2020| complete genome	19245.0	0.0	64	100	Download

Case Study 3

Project Plan

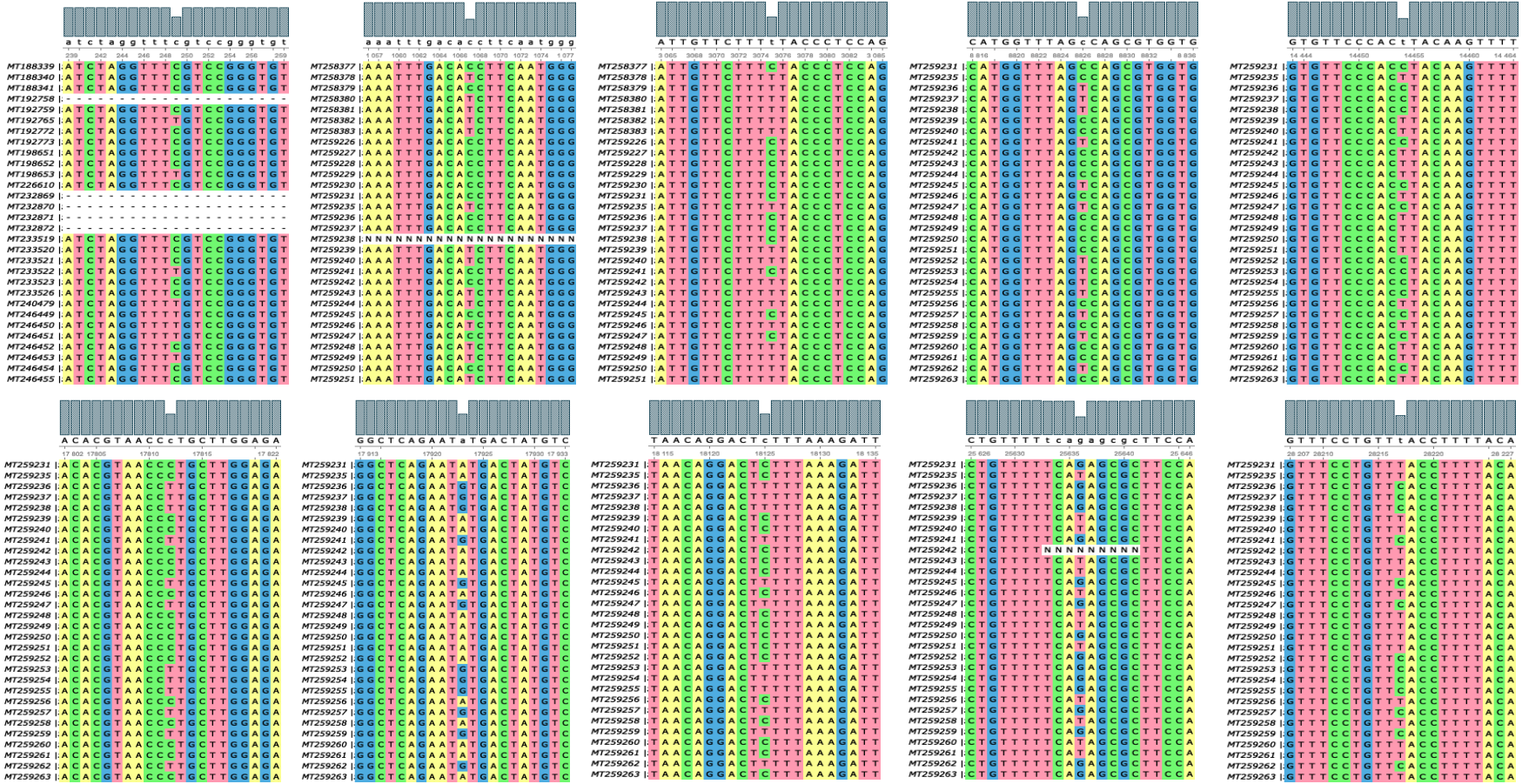
Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI



Case Study 3

iv) Multiple Sequence Alignment to discover the differences in various genomic regions

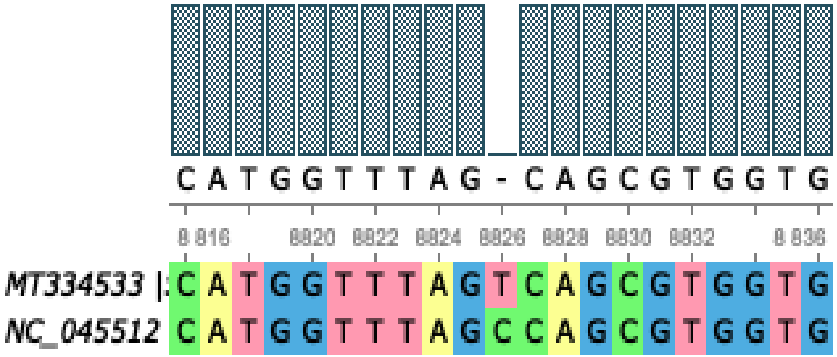
Goal: Understanding of Coronavirus (SARS-CoV-2) viruses using whole genome sequence data from NCBI

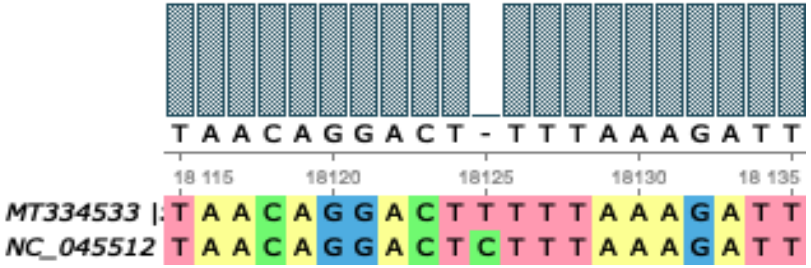


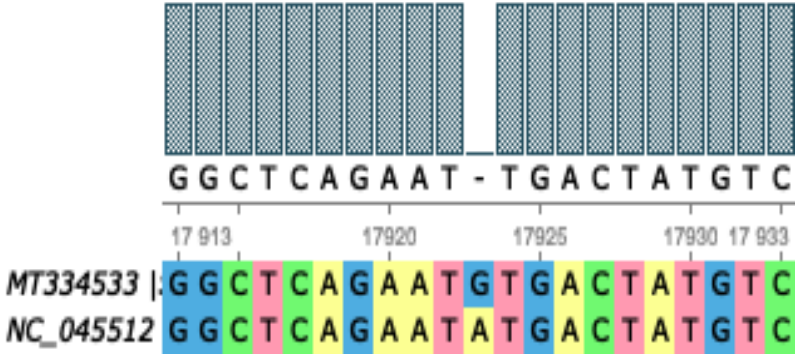
Case Study 3

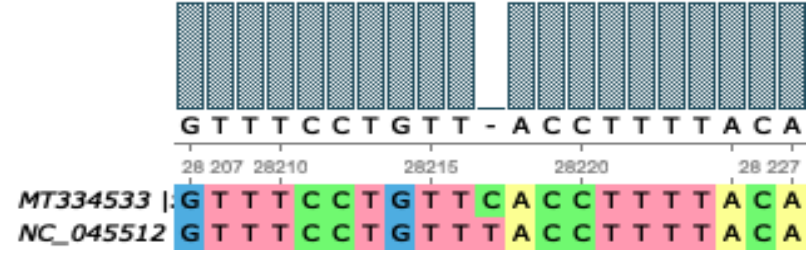
v) Sequence alignment of MT334553 with NC_045512.2 (Reference SARS_CoV-2)

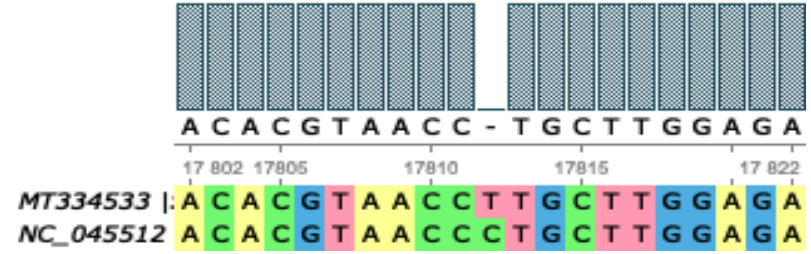
Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI







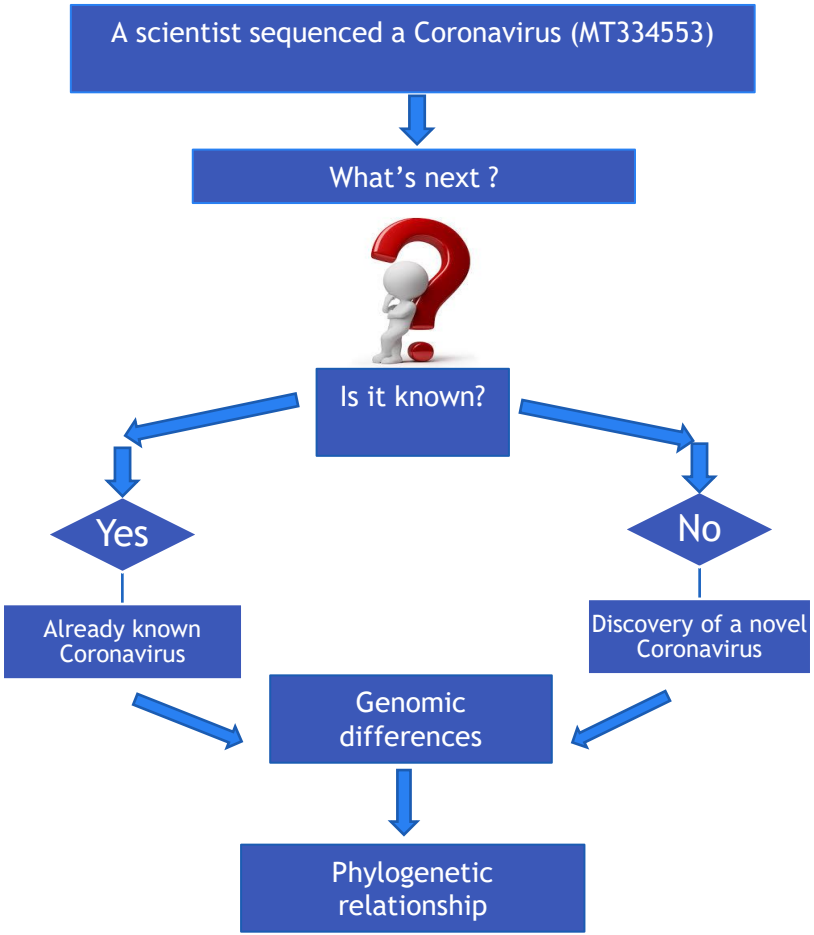




Case Study 3

Project Plan

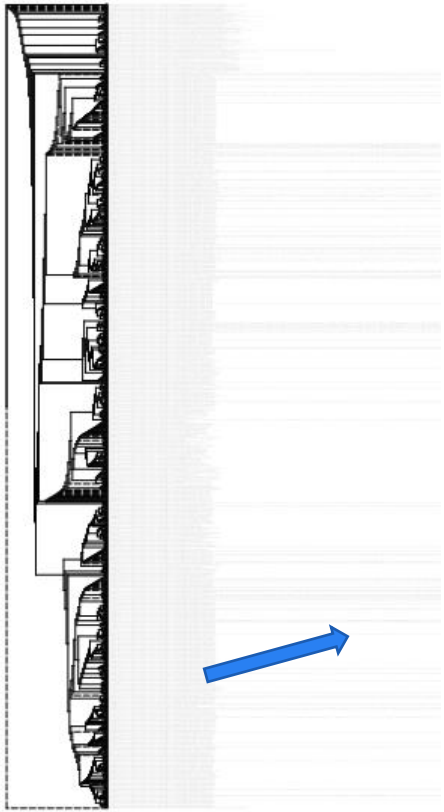
Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI



Case Study 3

vi) Phylogenetic relationship among all 928 coronavirus sequences

Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI

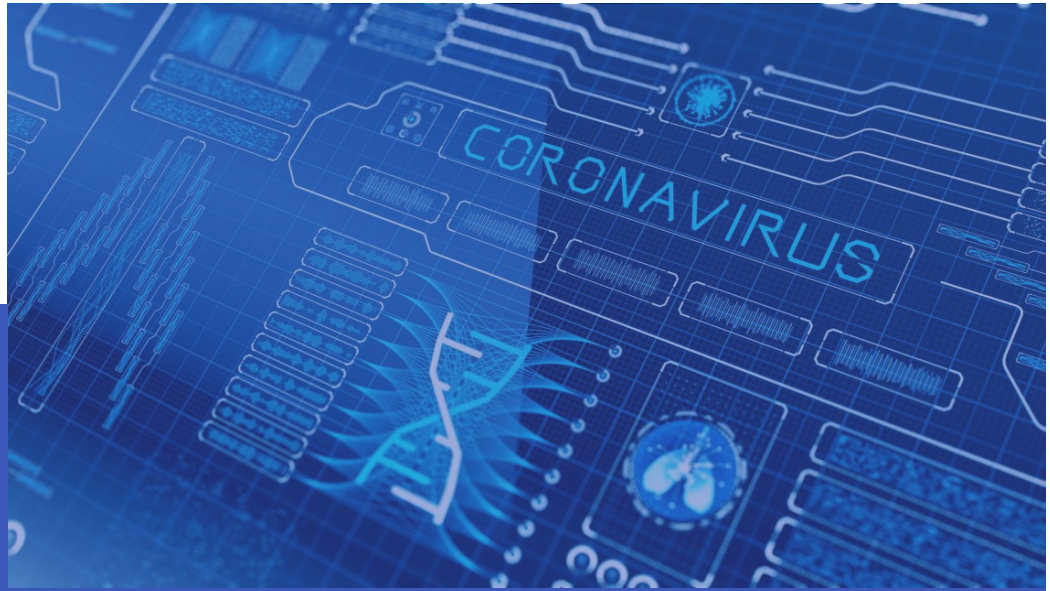


Complete phylogenetic tree of 928 SARS-Cov-2

MT259266 clustered together with USA/WA and USA/MI sample

Case Study 3

Conclusions



Goal

Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI

1) High homology of MT334553 with 928 Coronavirus (SARS-CoV-2) from NCBI.
2) Mutations only 10 genomic regions based on multiple sequence alignment.
3) MT334553 shows mutation with NC_045512.2 (Reference SARS_CoV-2) only in 5 conserved genomic regions.
4) Discovery of MT334553 in USA/WA and USA/MI variant group based on phylogenetic analysis.

Typical applications within agricultural microbial innovations

Loci Key Features

- High data security
- Interactive Search
- Data management
- Web based platform
- Integrated Analytics
- Visualization

An enterprise/cloud genomics data management and analytics platform

Loci
for Scientists

Data Integrity = No data prep, central repository, no data loss

Quick simple data mining, analysis, and visualization

Smooth integral workflow

Loci for Organizations

Team Efficiency ↑

R&D Costs ↓

Product Discovery ↑

Cost Savings

$ millions/year

Product Discovery

$100’s millions

Summary

1) exponential increase in genomics data is becoming a challenge

2) in-house data management and analytics is a necessity

3) Loci features:
   - high data security and structured data management
   - quick simple interactive search; integrated analytics and visualization
   - eliminates the need to install multiple stand-alone analytics tools
   - eliminates need to compromise security from accessing public data and tools

4) Loci can be deployed in various applications including orthology studies to seamlessly analyze genes, proteins and genomes from multiple organisms

5) Loci saves scientists time and may eliminate bottlenecks on constrained resources

6) Loci improves workflow efficiency, reduces R&D costs, and can enable increased rate of innovation

What's Next

Unleash your innovation potential.

Let's solve the world's most challenging problems. Together.

Loci, by Karyosoft.

Acknowledgements

Webinar participants

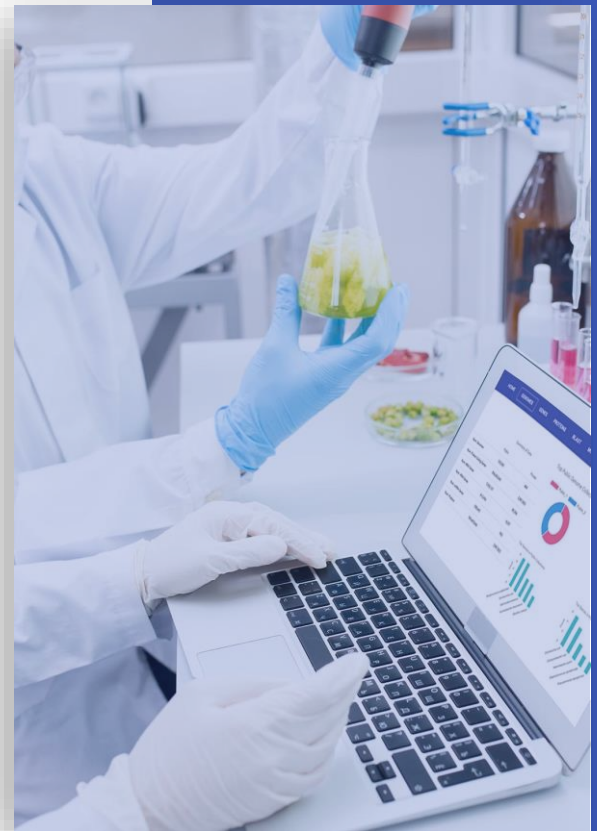
Phytobiomes Alliance Members
Isabelle Caugant
Kellye Eversole

Our clients
Our team members

Discussion, Q&A, and Collaboration

Empowering biologists to quickly transform and democratize vast amounts of genomics data into meaningful insights and discovery

Microbiome | Plant/Crop Sciences | Animal Health | Human Health

info@karyosoft.com
317-689-8493

Karyosoft
Augmented Innovations