

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



All content for Phytophosphorescence Alliance webinar purposes only



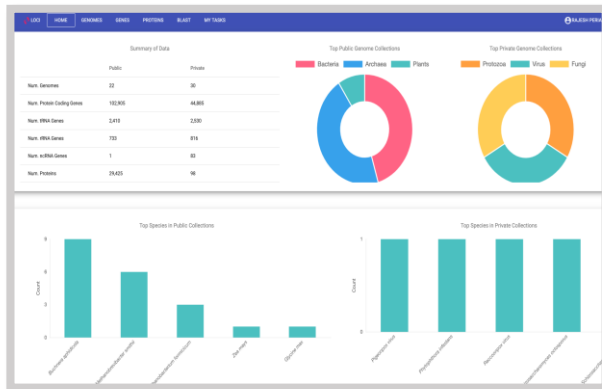
Introduction

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

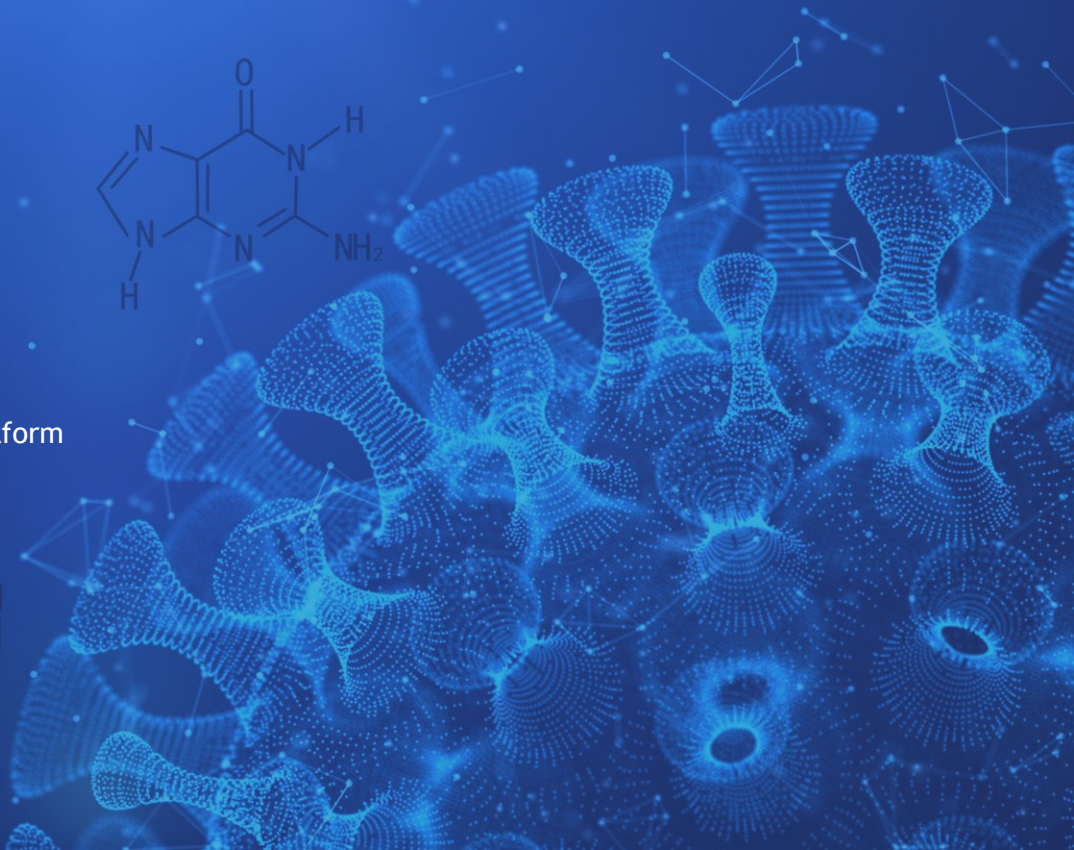
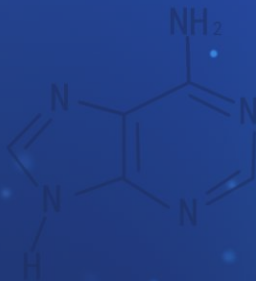
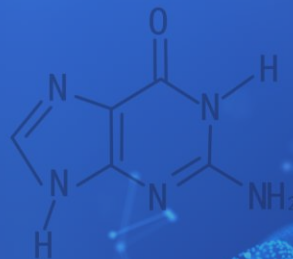
Loci Overview

Karyosoft's genomics data analytics and management platform

Benefits & Value

For Scientists; For Organizations

Summary



Case Study 1

Gene Discovery



Goal

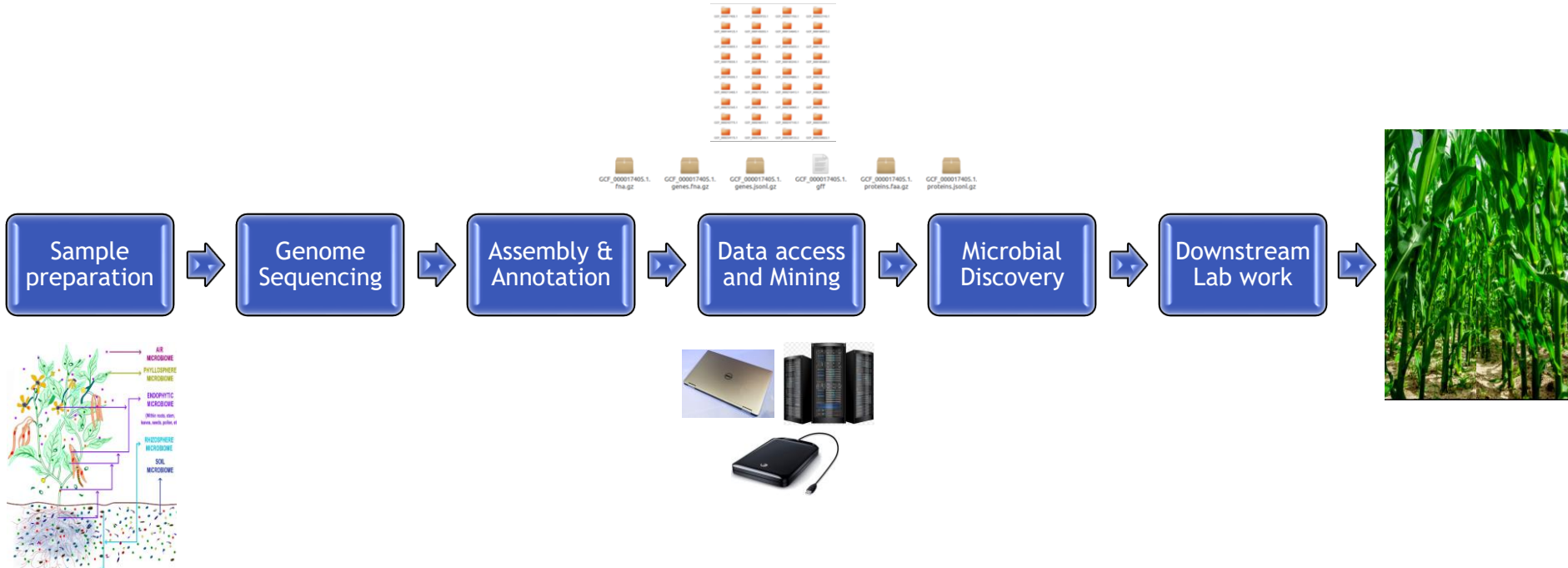
Discovery of a patentable pesticidal gene(s) from private bacterial collections and their novelty

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

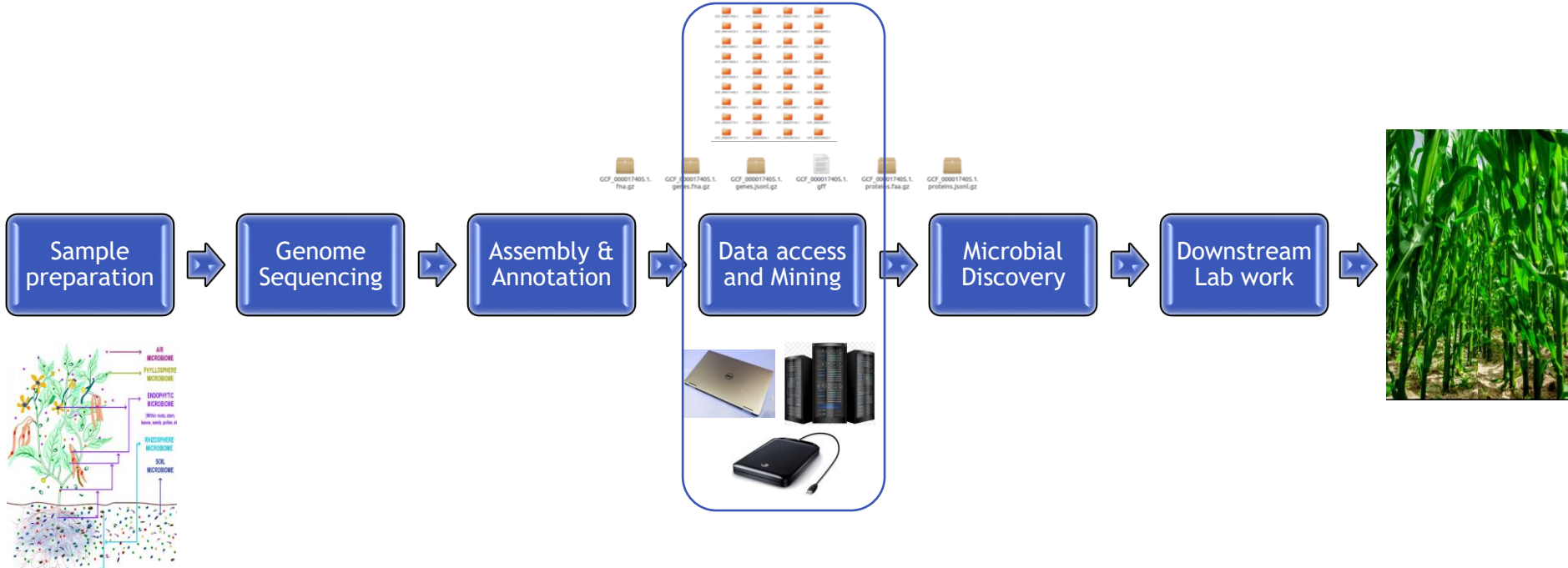
Typical Workflow

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



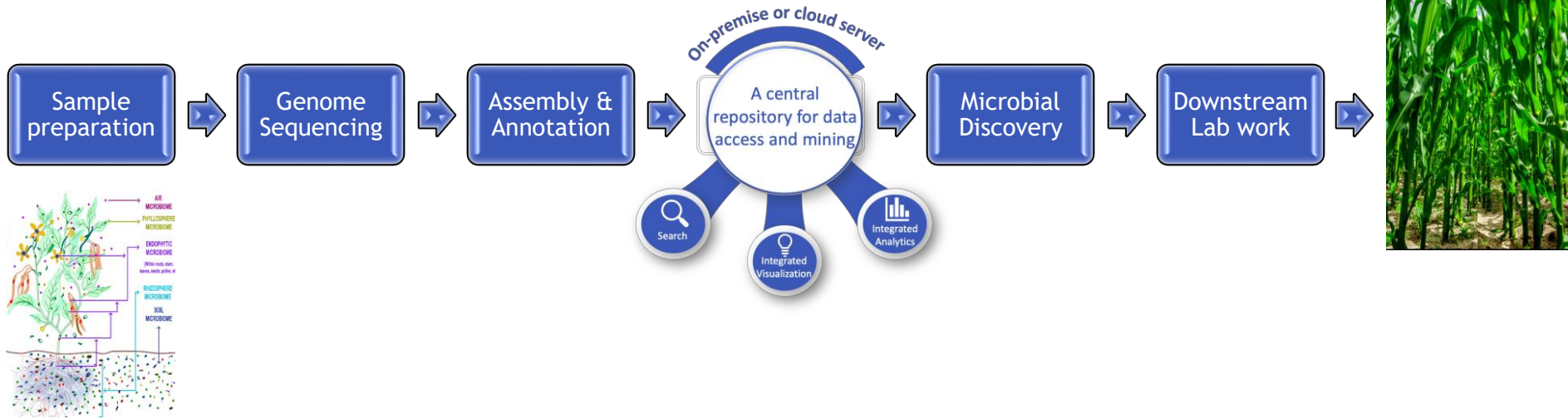
Typical Workflow

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




Loci in your internal workflow

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



Login to Loci within your secured environment

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

 [HOME](#) [GENOMES](#) [GENES](#) [PROTEINS](#) [BLAST](#) [MY TASKS](#) [LOGIN](#) [HELP](#)

Login

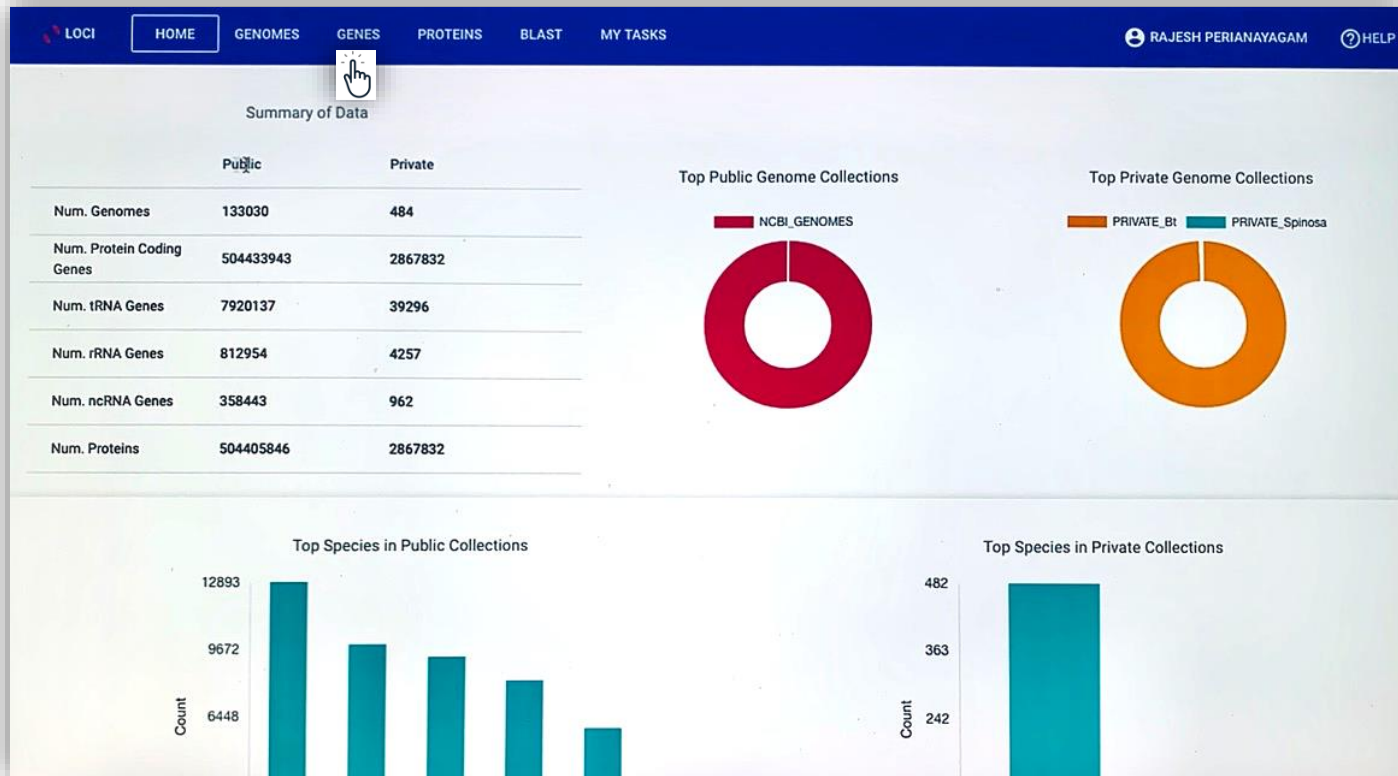
Email

Password

Login

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



i) How many pesticidal genes are in private collections?

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

The screenshot shows the LOCI database search interface. The top navigation bar includes links for HOME, GENOMES, GENES (selected), PROTEINS, BLAST, and MY TASKS. The user's name, RAJESH PERIANAYAGAM, and a HELP link are in the top right corner.

The search input field contains the text "pesti". Below the input field, a list of suggestions is displayed:

- pesticidal crystal protein cry1Ba (description)
- pesticin immunity protein (description)
- pesticidal protein Cry5Aa (description)
- pesticin (description)
- pesticidal protein Cry26Aa (description)

Each suggestion is accompanied by a hand icon pointing to it. Below the suggestions, there are three buttons: Search (blue), Reset (red), and Advanced (grey).

Below the buttons, there is a section titled "Select collections to search in:". It contains two columns of checkboxes:

- ☐ All Public Collections
- ☐ Public_1
- ☐ Public_2
- ☒ All Private Collections
- ☒ Private_1
- ☒ Private_2

Search for the genes by description

i) How many pesticidal genes are in private collections?

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

The screenshot shows the LOCI Genes search interface. The top navigation bar includes links for HOME, GENOMES, GENES (selected), PROTEINS, BLAST, and MY TASKS. The user is logged in as RAJESH PERIANAYAGAM. A search bar at the top contains the text "Enter gene name, ID, species, genus, class, etc." Below the search bar, three gene names are entered in separate boxes: "pesticidal crystal protein cry1Ba", "pesticidal protein Cry5Aa", and "pesticidal protein Cry26Aa". The "Search" button is highlighted. Below the search bar, the "Select collections to search in:" section shows two columns of checkboxes. The first column has "All Public Collections" (unchecked), "Public_1" (unchecked), and "Public_2" (unchecked). The second column has "All Private Collections" (checked), "Private_1" (checked), and "Private_2" (checked). The "Num. results: 2" is displayed. The "Filter Results" section shows a table with 9 columns: Selected, ID, Type, Description, Species, Assembly, Collection, Length (bp), and Analysis. Two results are shown, both with the "Selected" checkbox checked. The first result is for "GCF_000161515.1_gene6250" (protein_coding, pesticidal crystal protein cry1Ba, Bacillus thuringiensis, GCF_000161515.1, Private_1, 3687 bp). The second result is for "GCF_000177735.1_gene3770" (protein_coding, pesticidal protein Cry5Aa, Acinetobacter baumannii, GCF_000177735.1, Private_1, 276 bp). The "Analysis" column for both results contains links for "Browse", "BLAST", and "Download". The bottom right of the table shows "Items per page: 10" and "1 - 2 of 2".

Download FASTA - All Genes
Download FASTA - Selected Genes
Download CSV - All Genes
Download CSV - Selected Genes

Num. results: 2

Selected	ID	Type	Description	Species	Assembly	Collection	Length (bp)	Analysis
<input checked="" type="checkbox"/>	GCF_000161515.1_gene6250	protein_coding	pesticidal crystal protein cry1Ba	Bacillus thuringiensis	GCF_000161515.1	Private_1	3687	Browse BLAST Download
<input checked="" type="checkbox"/>	GCF_000177735.1_gene3770	protein_coding	pesticidal protein Cry5Aa	Acinetobacter baumannii	GCF_000177735.1	Private_1	276	Browse BLAST Download

Items per page: 10 1 - 2 of 2

Discovery of 2 pesticidal genes in their private collections

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

LOCI

HOME

GENOMES

GENES

PROTEINS

BLAST

MY TASKS

RAJESH PERIANAYAGAM

HELP

>GCF_000161515.1_gene6250 pesticidal crystal protein cry1Ba MyLociAttributes=[mylocidb=GCF_000161515.1_mylocigene;]=EndMyLociAttributes
TTGACTTCAAATAGGAAATGAGAATGAAATTATAAATGCTGTATCGAATCATTCGACAAATGGATCTATTACCAGA
TGCTCGTATTGAGGATAGCTTGTGTATAGCCGAGGGGAACAATATCGATCCATTTGTTAGCGCATCAACAGTCCAAACGG
GTATTAACATAGCTGGTAGAATCTAGCGCTATTGGCGCTACCGTTTGTCTGGACAACTAGCTAGTTTTTATAGTTTTCTT
CTTAATGCGATGCCGCTTTTCGCAATTAGAAACCAAGAGTTCCATTATTGATGGTATATGCTCAAGCTGCAAAATTACA
GAAACAGATACAGTATGGATTGAGATTGGAGAAACGGAAGGAAGTTTATTGTAGTAGCGGTGAATTACTCCTCATGGA
AGAAATAG
>GCF_000177735.1_gene3770 pesticidal protein Cry5Aa MyLociAttributes=[mylocidb=GCF_000177735.1_mylocigene;]=EndMyLociAttributes
ATGACCATCCTACCTTATGCTCTTGGCATTAGCTTGATTGCTATCACAATCTCTATGTTGCTTTGCGCTTTTCGAATGGT
GATGGGACCATTATTGTGCGACCGCCTACTAGCGCTTGATACCCCTATTTTAAATGCGACCTGTCTCATTATTGTGTTAG

Select Pre-formatted Databases - OR - Custom Databases

Pre-formatted Databases

Custom Databases

Nucleotide Databases

☐ Genes - All Public Collections

☐ Genes - All Private Collections

☐ Genomes - All Public Collections

☐ Genomes - All Private Collections

☐ NCBI nt

☐ NCBI env_nt - Nucleotide Sequences from Metagenomes

☒ NCBI Patent Nucleotide Sequences

☐ NCBI Bacterial and Archaeal 16S rRNA Sequences

Protein Databases

☐ Proteins - All Public Collections

☐ Proteins - All Private Collections

☐ NCBI nr

☐ NCBI env_nr - Protein Sequences from Metagenomes

☐ NCBI Patent Protein Sequences

Select a Program

☒ blastn (nucleotide query, nucleotide database)

☐ tblastn (protein query, nucleotide database)

☐ blastp (protein query, protein database)

☐ blastx (nucleotide query, protein database)

Optional

Task Name

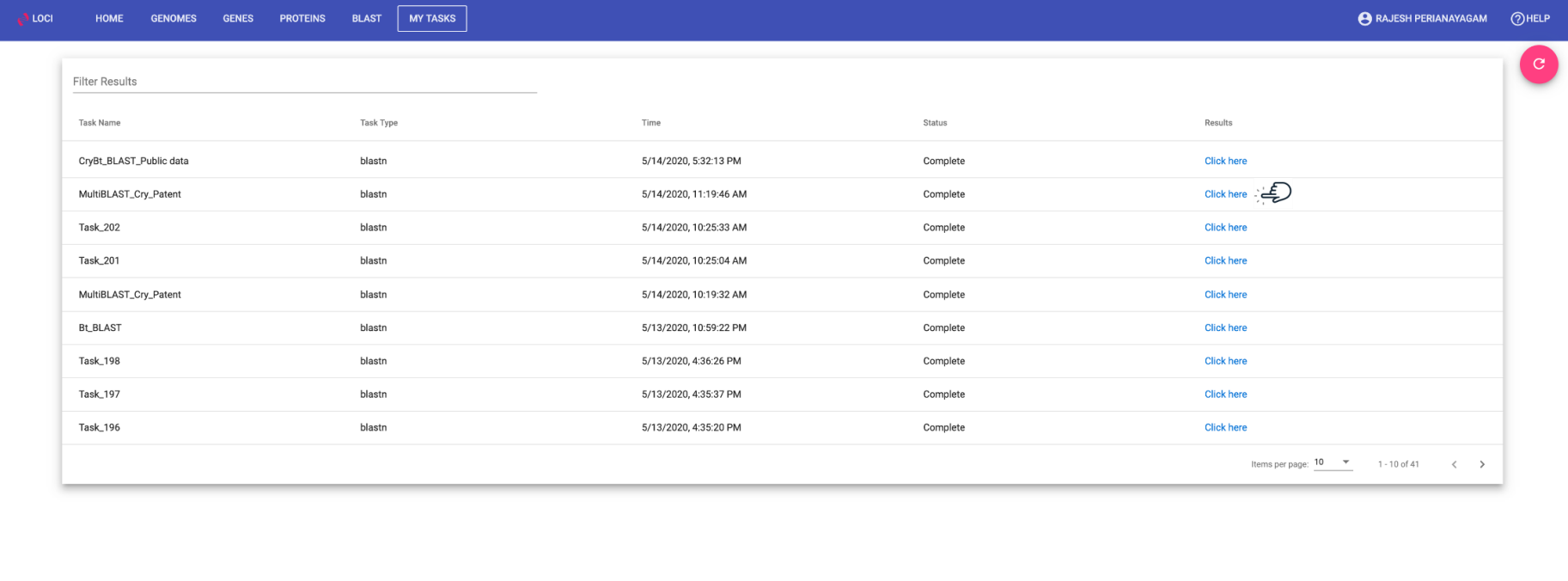
MultiBLAST_Cry

BLAST

RESET

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



The screenshot shows the 'MY TASKS' section of the LOCI web interface. The table lists 10 completed tasks, including public data and private patent collections. Each task has a 'Click here' link to view results. A red circular button with a white 'C' is visible in the top right corner of the interface.

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

Items per page: 10 1 - 10 of 41

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

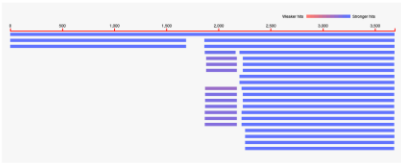
ii) Are *cry1Ba* in private collections patentable?

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

LOC HOME GENOMES GENES PROTEINS **BLAST** MY TASKS

Default is query sequence
GCF_000161515.1_gene6250 pesticidal cryst... -

Graphical Overview
Query: GCF_000161515.1_gene6250 pesticidal crystal prot... (Length: 3687)



Filter Results

Selected	Sequence ID	Description	Max Score	E-value	Query Coverage	% Identity	Download
<input type="checkbox"/>	gb 42103.1	Sequence 1 from patent US 5628995	3681.0	0.0	100	99	Download
<input type="checkbox"/>	gb 70138.1	Sequence 3 from patent US 5679343	1477.0	0.0	49	93	Download
<input type="checkbox"/>	gb 38760.1	Sequence 3 from patent US 5616319	1477.0	0.0	49	93	Download
<input type="checkbox"/>	gb U31527.1	Bacillus thuringiensis delta endotoxin CryET1 (cryET1) gene, complete cds	972.0	0.0	40	88	Download
<input type="checkbox"/>	gb 73040.1	Sequence 1 from patent US 5686069	962.0	0.0	39	88	Download
<input type="checkbox"/>	gb 11609.1	Sequence 4 from Patent US 5407825	962.0	0.0	39	88	Download
<input type="checkbox"/>	emb 427529.1	B.thuringiensis P581A2 endotoxin gene	959.0	0.0	39	88	Download
<input type="checkbox"/>	gb 70137.1	Sequence 1 from patent US 5679343	884.0	0.0	40	86	Download
<input type="checkbox"/>	gb 38759.1	Sequence 1 from patent US 5616319	884.0	0.0	40	86	Download
<input type="checkbox"/>	daj E00881.1	DNA encoding insecticidal protein	709.0	0.0	39	82	Download

Items per page: 10 1 - 10 of 100 < >

>42103.1 Sequence 1 from patent US 5628995 MyLocIAttributes=[mylocid=ncbi_patent;]=EndMyLocIAttributes
Length=3687

Score = 6798 bits (3681), Expect = 0.0
Identities = 3685/3687 (99%), Gaps = 0/3687 (0%)
Strand=Plus/Plus

Query 1 TTGACTTCAATGAGAAATGAGAAATGAAATATATAATCTGTGTATGATCGATATCCGCA 68
|||||
Sbjct 186 TTGACTTCAATGAGAAATGAGAAATGAAATATATAATCTGTGTATGATCGATATCCGCA 245

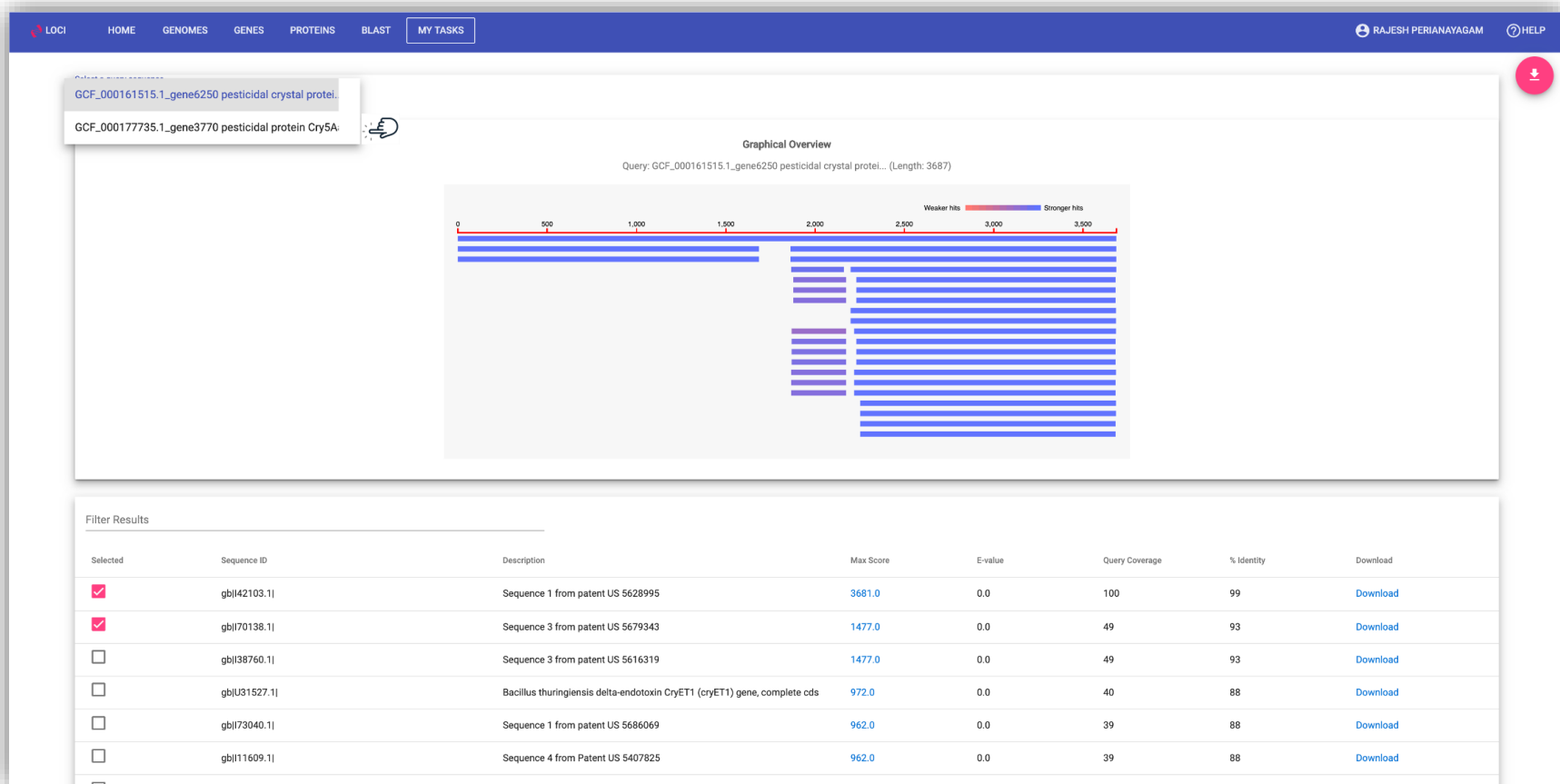
Query 61 CAATGATCTATTACAGATGCTGCTGATTAGGATAGCTGTGTATAGCCAGGGGAAC 128
|||||
Sbjct 246 CAATGATCTATTACAGATGCTGCTGATTAGGATAGCTGTGTATAGCCAGGGGAAC 385

Query 121 AATATCGATCATTGTTAGCGCATCAACAGTCCAAACGGGTATTACATAGCTGTAGA 188
|||||
Sbjct 386 AATATGATCATTGTTAGCGCATCAACAGTCCAAACGGGTATTACATAGCTGTAGA 365

Download FASTA - All Hits
Download FASTA - Selected ...
Download CSV - All Hits
Download CSV - Selected Hits


ii) Are *cry5Aa* in private collections patentable?

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



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


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

 LOC | [HOME](#) | [GENOMES](#) | [GENES](#) | [PROTEINS](#) | [BLAST](#) | **MY TASKS** | [RAJESH PERIANAYAGAM](#) | [HELP](#)

Select a query sequence
GCF_000177735.1_gene3770 pesticidal protein Cry... ▾

Graphical Overview

Query: GCF_000177735.1_gene3770 pesticidal protein Cry5Aa... (Length: 276)



Filter Results

Selected	Sequence ID	Description	Max Score	E-value	Query Coverage	% Identity	Download
Items per page: 5 ▾ 0 of 0 < >							

***** No hits found *****

Lambda	K	H
1.33	0.621	1.12

Gapped

Lambda	K	H
1.28	0.460	0.850

Effective search space used: 100528182454

Database: e607f30e90ea42c1a6a7c93970a7d802
Posted date: Apr 23, 2018 10:58 PM
Number of letters in database: 30,036,802
Number of sequences in database: 99,138

Matrix: blastn matrix 1 -2
Gap Penalties: Existence: 0, Extension: 2.5

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

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

LOCI HOME GENOMES GENES PROTEINS BLAST MY TASKS RAJESH PERIANAYAGAM HELP

```
>GCF_000161515.1_gene6250    pesticidal crystal protein cry1Ba    MyLocidAttributes=[mylocidb=GCF_000161515.1_mylocigene;]=EndMyLocidAttributes
TTGACTTCAAATAGGAAAAATGAGAATGAAATTATAAATGCTGTATCGAATCATTCGCCACAAATGGATCTATTACCAGA
TGCTCGTATTGAGGATAGCTTGTGTATAGCCGAGGGGAACAATCGATCCATTGTTAGCGCATCAACAGTCCAAACGG
GTATTAAACATAGCTGGTAGAATACTAGGCGTATTGGGCGTACCGTTTGTGGACAACACTAGCTAGTTTTATAGTTTTCTT
GAAACAGATACAGTATGGATTGAGATTGGAGAAACGGAAGGAAGTTATTGTAGATAGCTGGAATTACTCTCATGGA
AGAATAG
>GCF_000177735.1_gene3770    pesticidal protein Cry5Aa    MyLocidAttributes=[mylocidb=GCF_000177735.1_mylocigene;]=EndMyLocidAttributes
ATGACCATCCTACCTTATGCTCTTGGCATTAGCTTGATGCTATCACAATCTCTATGTTGCTTTTGCCTTTTTCGAATGGT
GATGGGACCATCTATTGTGACCGCCTACGCGCTTGATACCTATTTTAAATGCGACCTGCTCATTTATTGTGTGTTAG
GCATTTTATTGGACAACCACTCTTTATTGAAGGGCGCTGTTAGTCGCTATGCTAGGGTTTGTTCACCGGCTGCTTTA
```

Select Pre-formatted Databases - OR - Custom Databases

Pre-formatted Databases

Custom Databases

Nucleotide Databases

☒ Genes - All Public Collections

☐ Genes - All Private Collections

☒ Genomes - All Public Collections

☐ Genomes - All Private Collections

☒ NCBI nt

☒ NCBI env_nt - Nucleotide Sequences from Metagenomes

☐ NCBI Patent Nucleotide Sequences

☒ NCBI Bacterial and Archaeal 16S rRNA Sequences

Protein Databases

☐ Proteins - All Public Collections

☐ Proteins - All Private Collections

☐ NCBI nr

☐ NCBI env_nr - Protein Sequences from Metagenomes

☐ NCBI Patent Protein Sequences

Select a Program

☒ blastn (nucleotide query, nucleotide database)

☐ tblastn (protein query, nucleotide database)

☐ blastp (protein query, protein database)

☐ blastx (nucleotide query, protein database)

Optional

Task Name

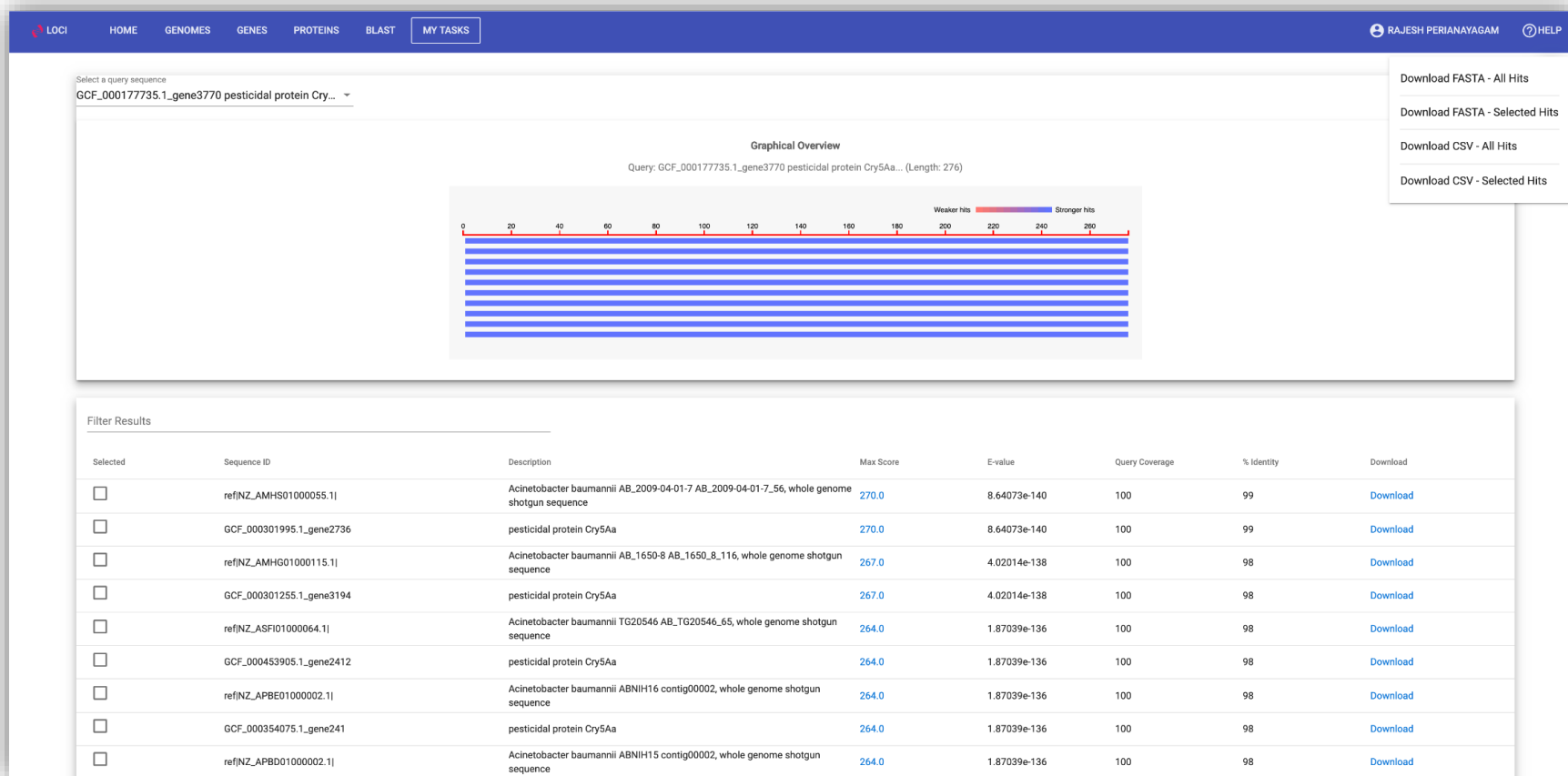
CryBt_BLAST_Public data

BLAST

RESET

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

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

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

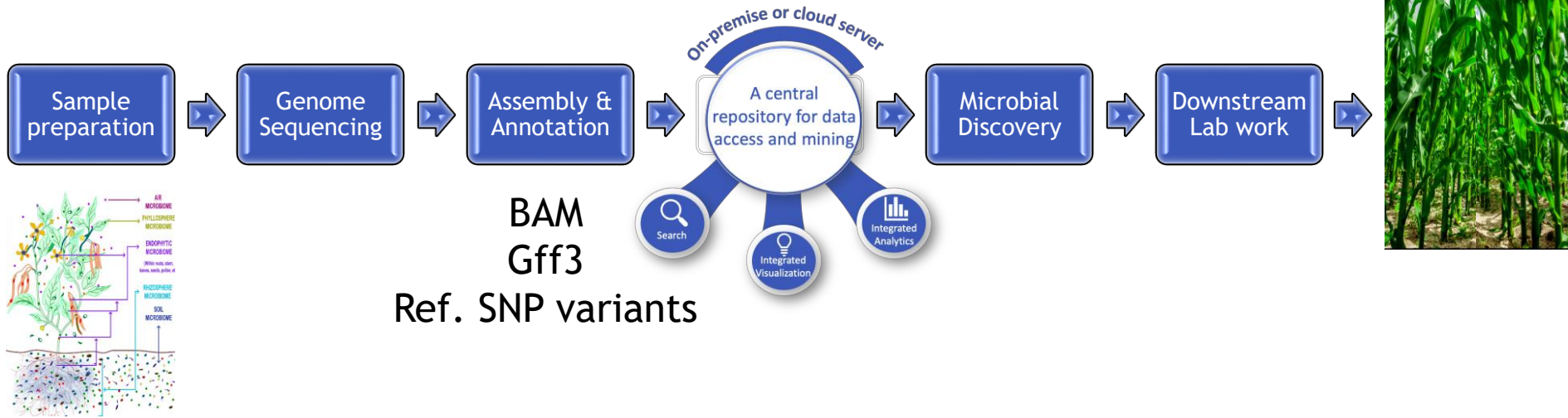
Case Study 2

Strain Engineering

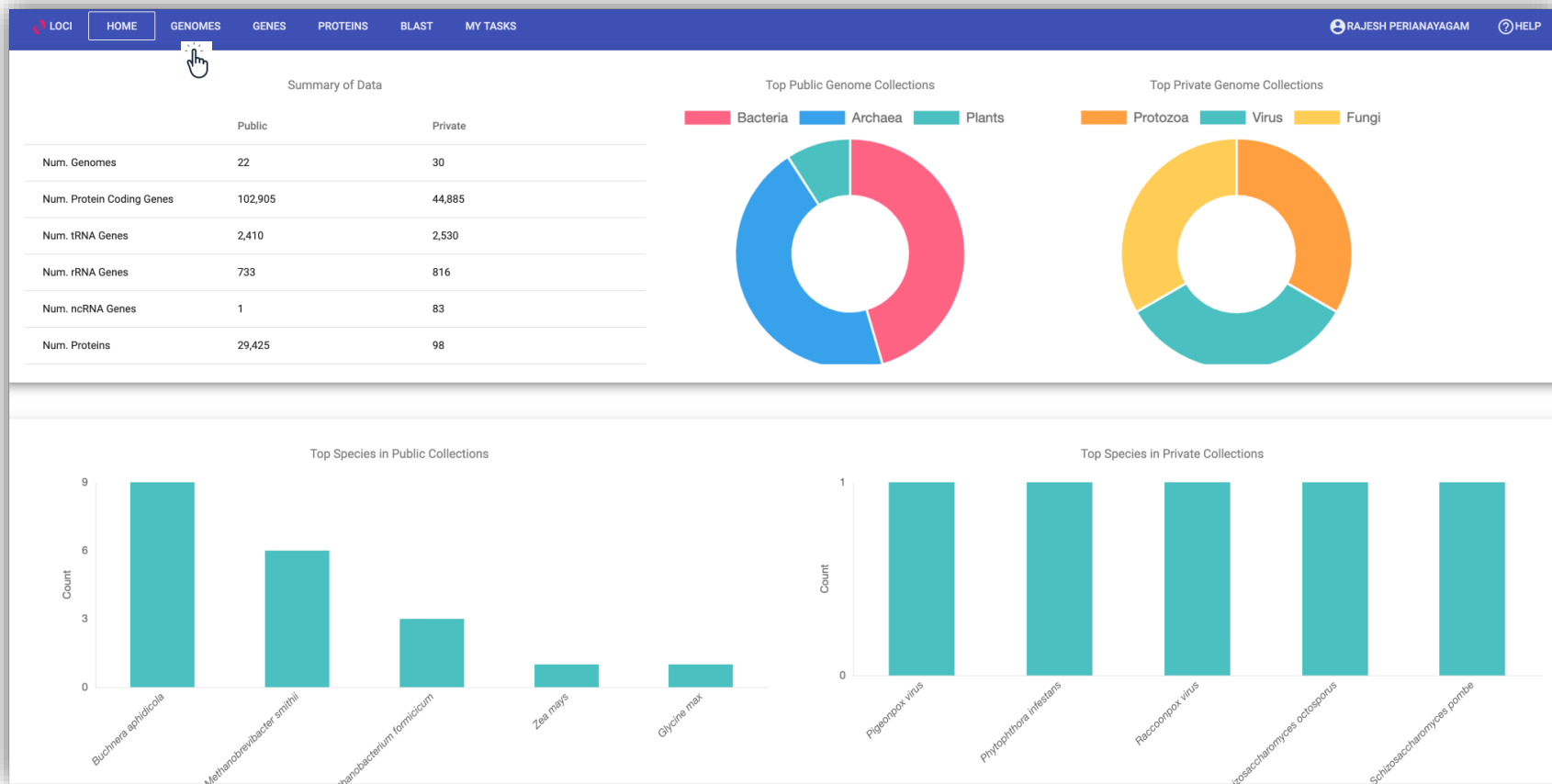


Loci in your workflow

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



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

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

LOCI

HOME

GENOMES

GENES

PROTEINS

BLAST

MY TASKS

RAJESH PERIANAYAGAM

HELP

Enter strain name, species, genus, class, etc.

esc

Escherichia coli (species)

Escherichia (genus)

Escherichia (genus)

Escherichia coli (species)

Search

Reset

Advanced

i) Searching for the *E.coli* reference genomeGoal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains

LOCI

HOME

GENOMES

GENES

PROTEINS

BLAST

MY TASKS

RAJESH PERIANAYAGAM

HELP

Enter strain name, species, genus, class, etc.

Escherichia coli

Search

Reset

Advanced

Download CSV - All Genomes

Download CSV - Selected Genom...

Num. results: 2

Filter Results

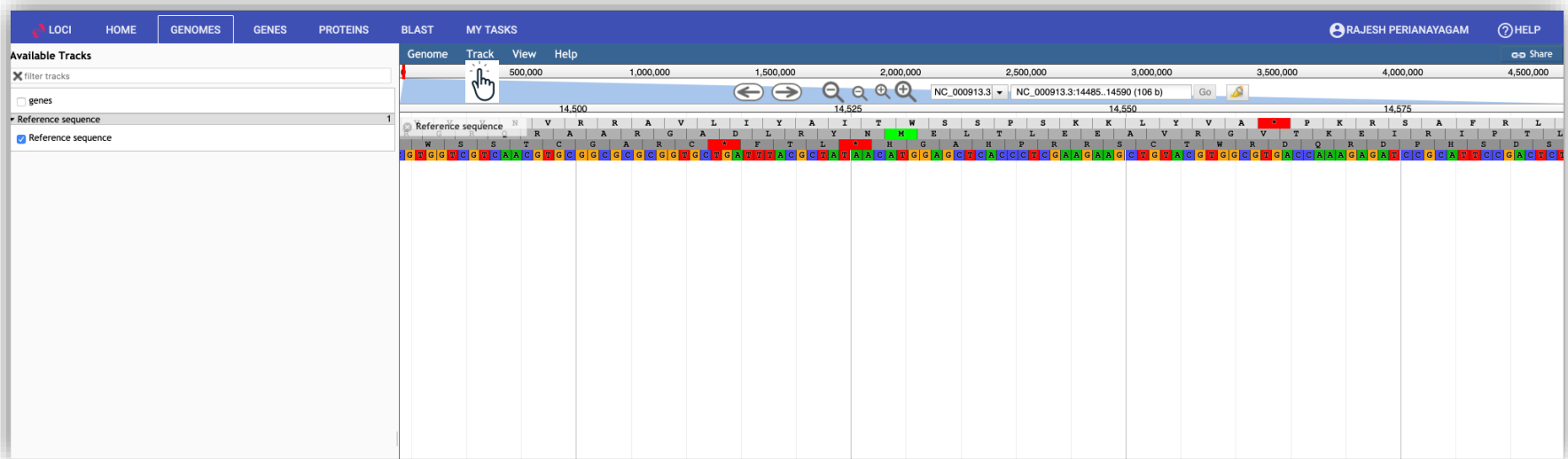
Species, Intraspecies Name, Collection Name, Genome Size, Number of Genes, ...

	Assembly	Species	Intraspecies Name	Collection	Size(Mb)	Num. Genes	Num. Proteins	Analysis
<input type="checkbox"/>	GCF_000005845.2	Escherichia coli	K-12 substr. MG1655	NCBI_GENOMES	4.64	4419	4240	Browse BLAST Download
<input type="checkbox"/>	GCF_000005845.2	Escherichia coli	K-12 substr. MG1655	NCBI_GENOMES	4.64	4419	4240	Browse BLAST Download

Items per page: 10 1 - 2 of 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



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

The screenshot displays the JBrowse genome browser interface. The top navigation bar includes links for LOCI, HOME, GENOMES (selected), GENES, PROTEINS, BLAST, and MY TASKS. The user's name, RAJESH PERIANAYAGAM, and a HELP icon are on the right. The left sidebar shows 'Available Tracks' with options to filter tracks, show genes, and reference sequences. The main panel shows a genomic track for NC_000913.3 (7.65 Kb) with a zoomed-in view of the *ribF* gene region (approximately 18,750 to 25,000 coordinates). A modal window titled 'Open files' is open, allowing users to add local files or remote URLs. Below this, a 'Files and URLs' section lists various files, including BAM files for four strains and GFF3 files for E. coli genes and variants. At the bottom, a 'New Tracks' section shows the configuration for each track, including the track name, display type, and an 'Edit Configuration' link. The 'Open immediately' radio button is selected, and the 'Open' button is visible at the bottom right of the modal.

Open files

Add any combination of data files and URLs, and JBrowse will automatically suggest tracks to display their contents.

Local files

Select Files...

Select or drag files here.

Remote URLs - one per line

<http://paste.urls.here/example.bam>

Files and URLs

File Type	File Name	Status
BAM	Strain1.bam	X
BAM index	Strain1.bam.bai	X
BAM	Strain2.bam	X
BAM index	Strain2.bam.bai	X
BAM	Strain3.bam	X
BAM index	Strain3.bam.bai	X
BAM	Strain4.bam	X
BAM index	Strain4.bam.bai	X
GFF3	E.Coli_genes.gff3	X
GFF3	E.coli_Variants.gff3	X

New Tracks

Name	Display	Edit Configuration
BAM Strain1	Alignments2	X
BAM Strain2	Alignments2	X
BAM Strain3	Alignments2	X
BAM Strain4	Alignments2	X
GFF3 E.Coli genes	CanvasFeatures	X
GFF3 E.coli Variants	CanvasFeatures	X

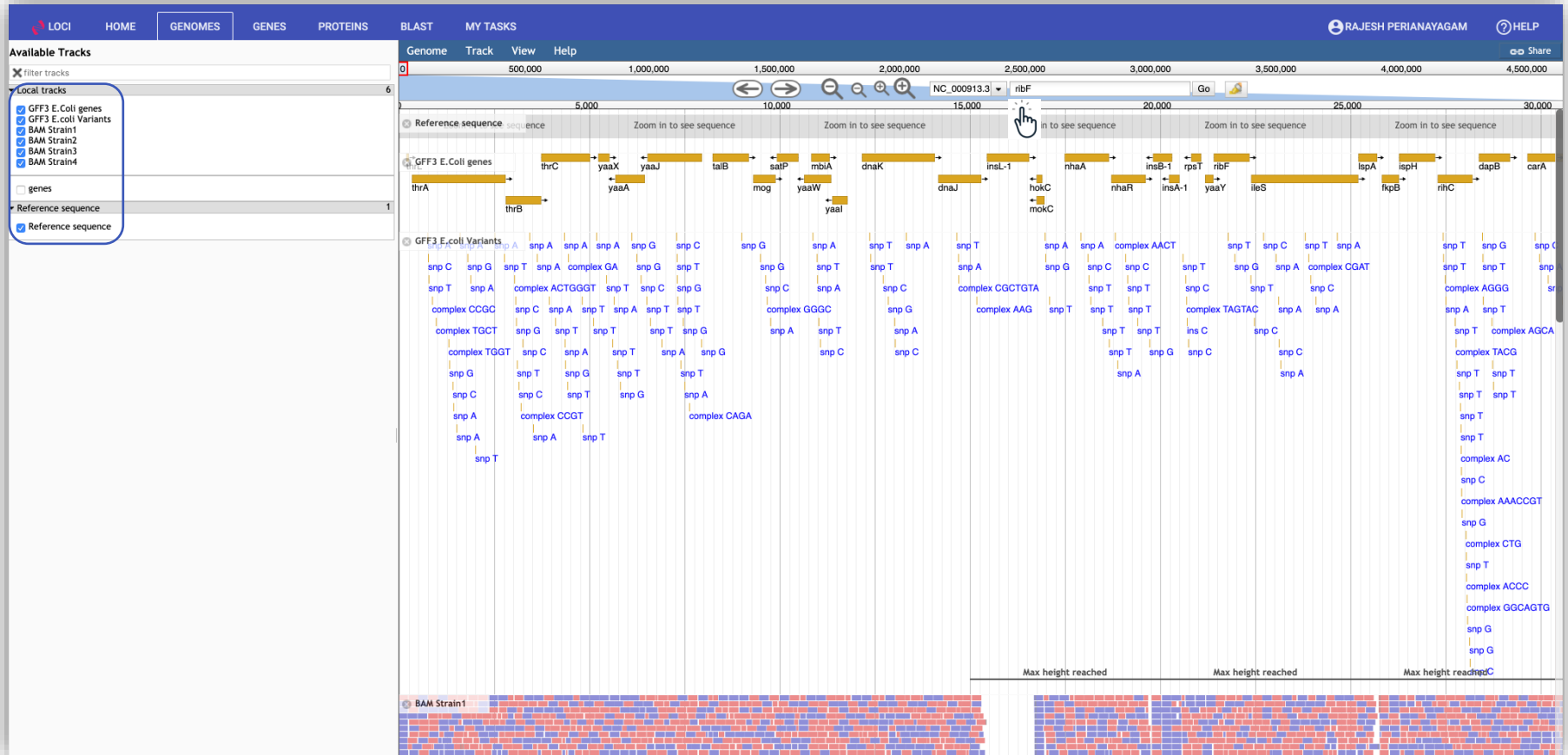
☒ Open immediately ☐ Add to tracks

Cancel Open

Case Study 2

iii) Search for *ribF* gene

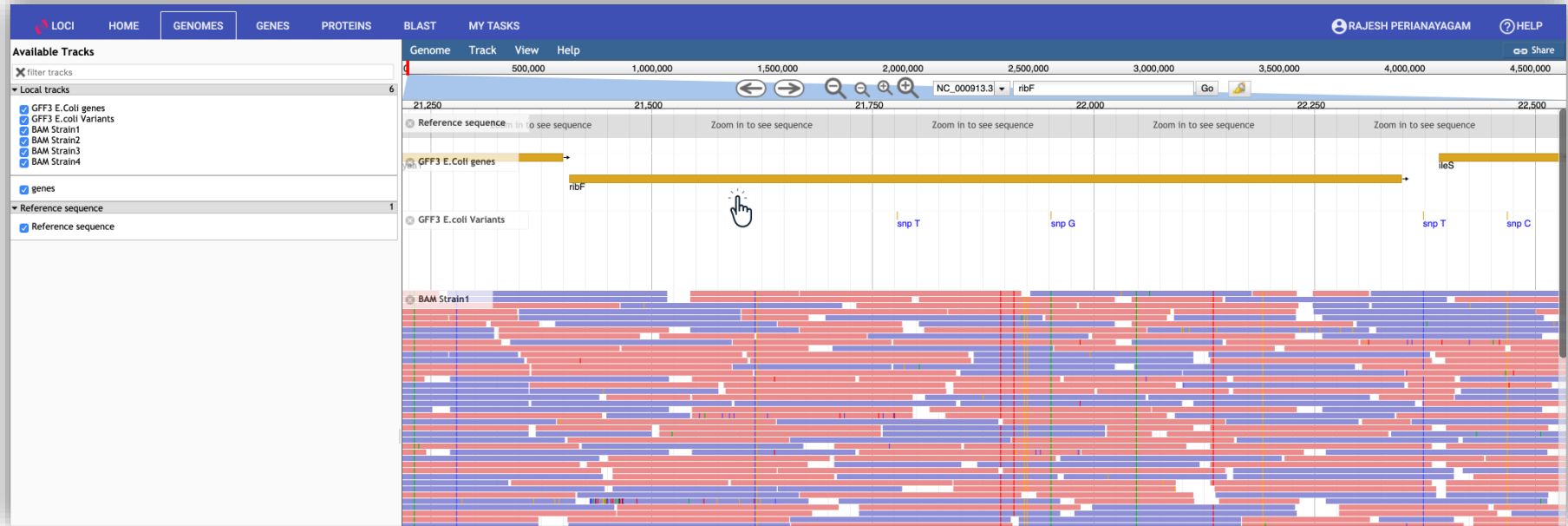
Goal: Discover mutations in *ribF* gene of *E. coli* strain 1 and compare it with 3 other strains

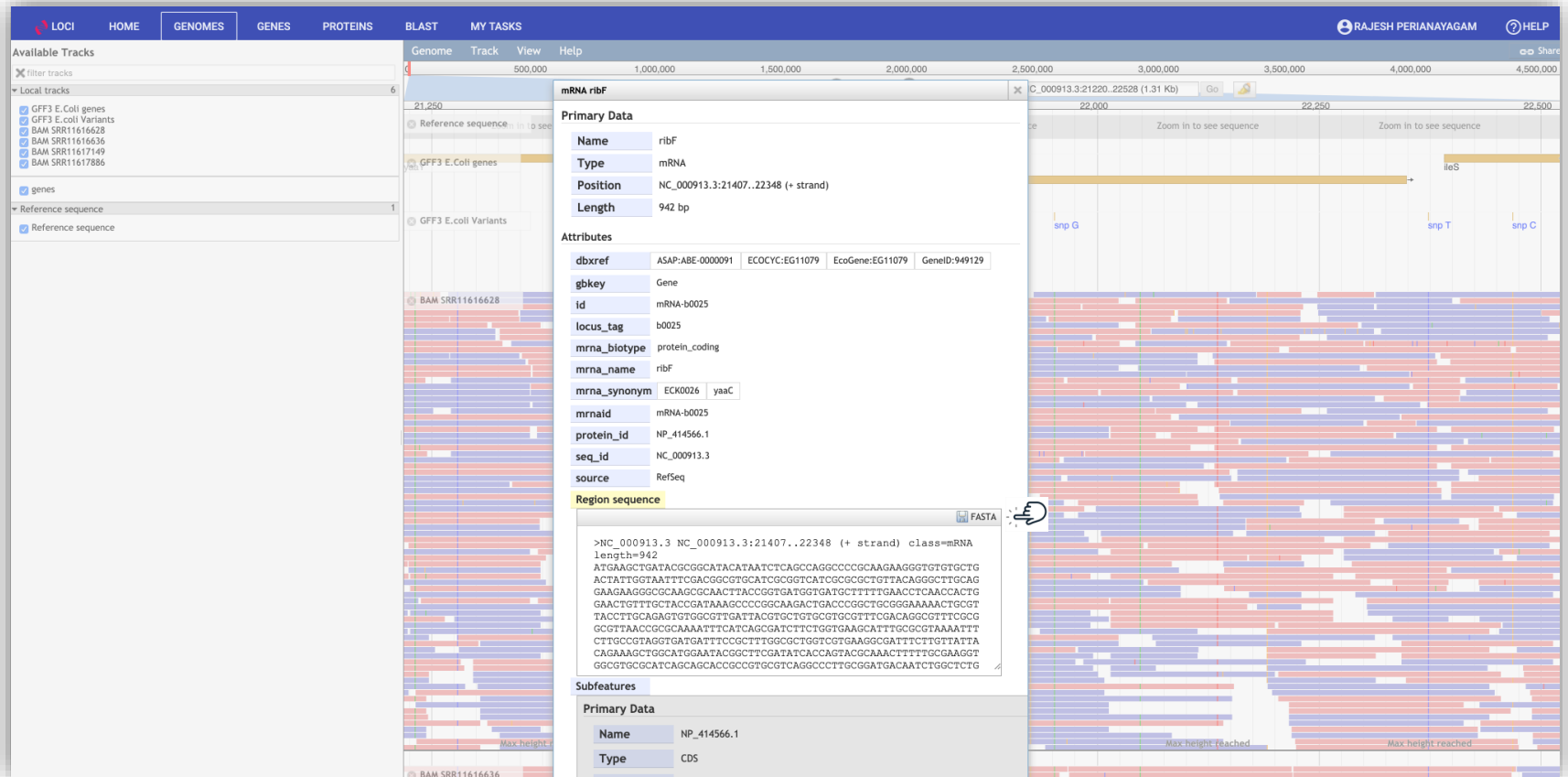


Case Study 2

iii) *ribF* gene

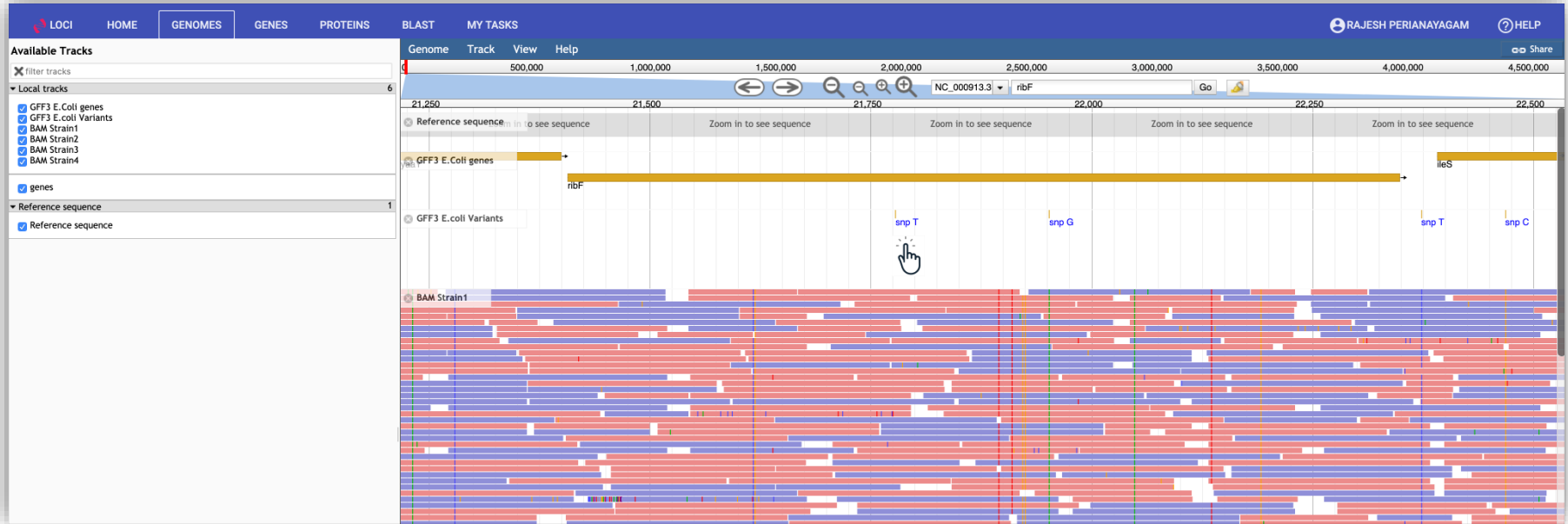
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



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

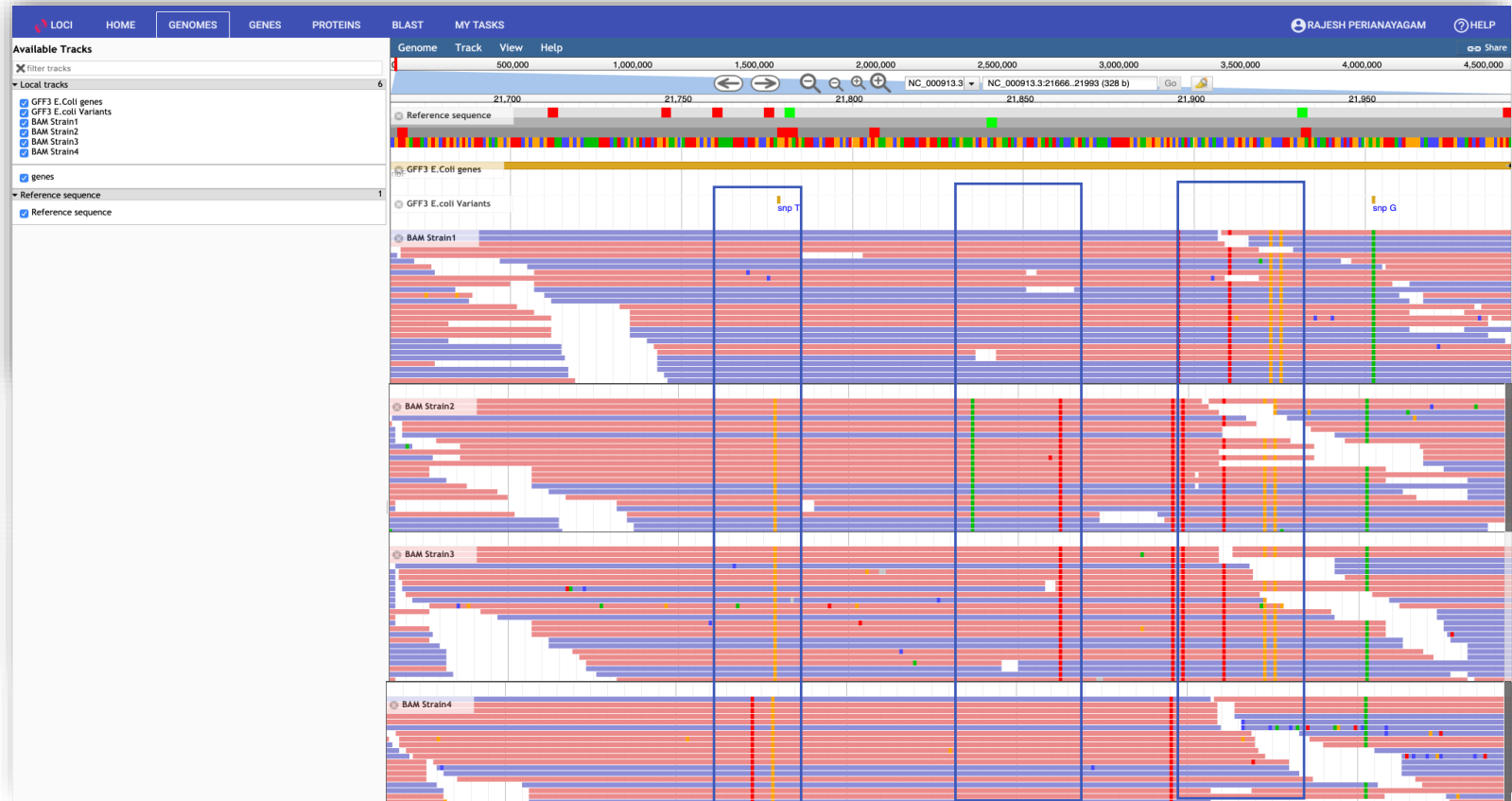
v) Reference SNP details

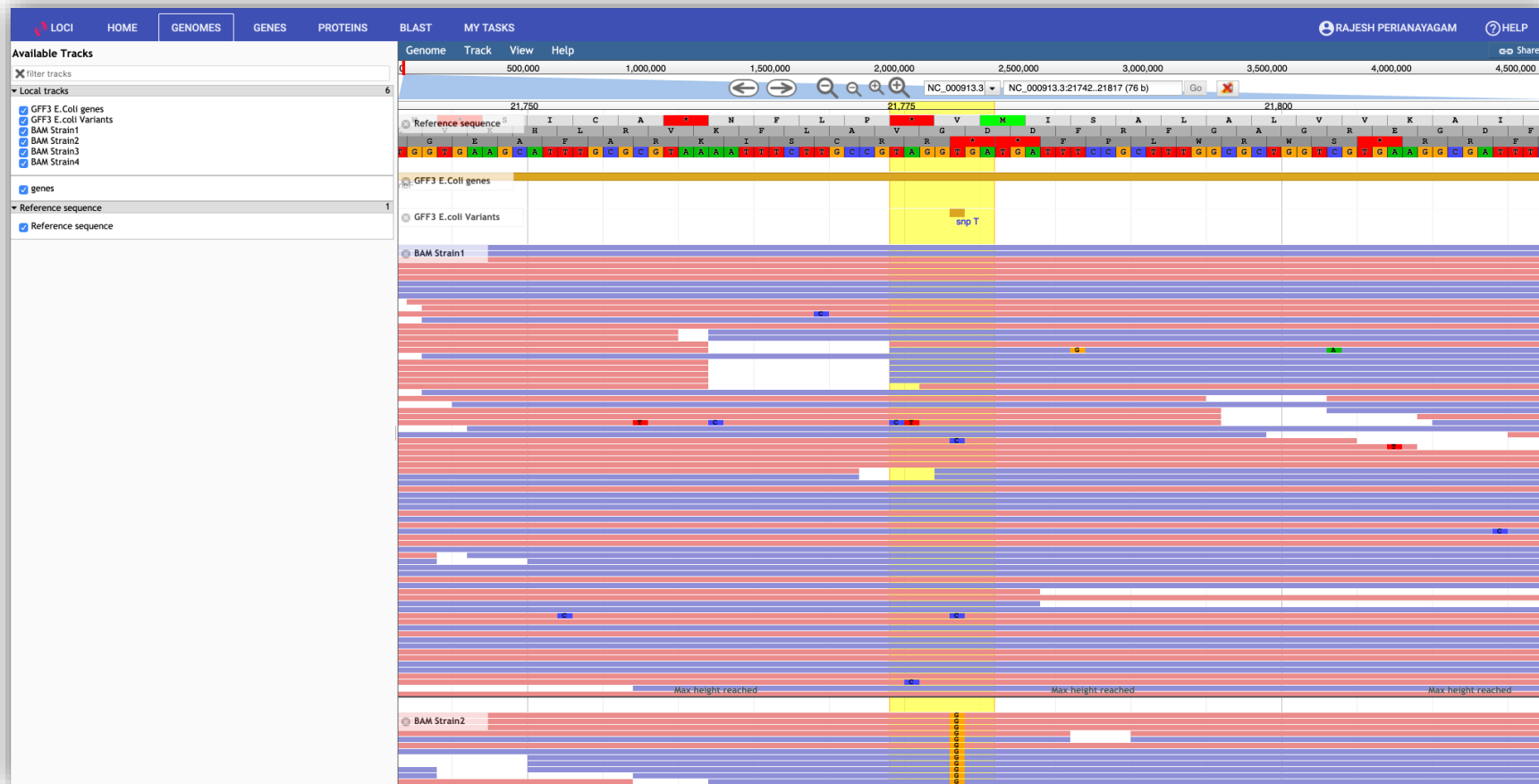
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains



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

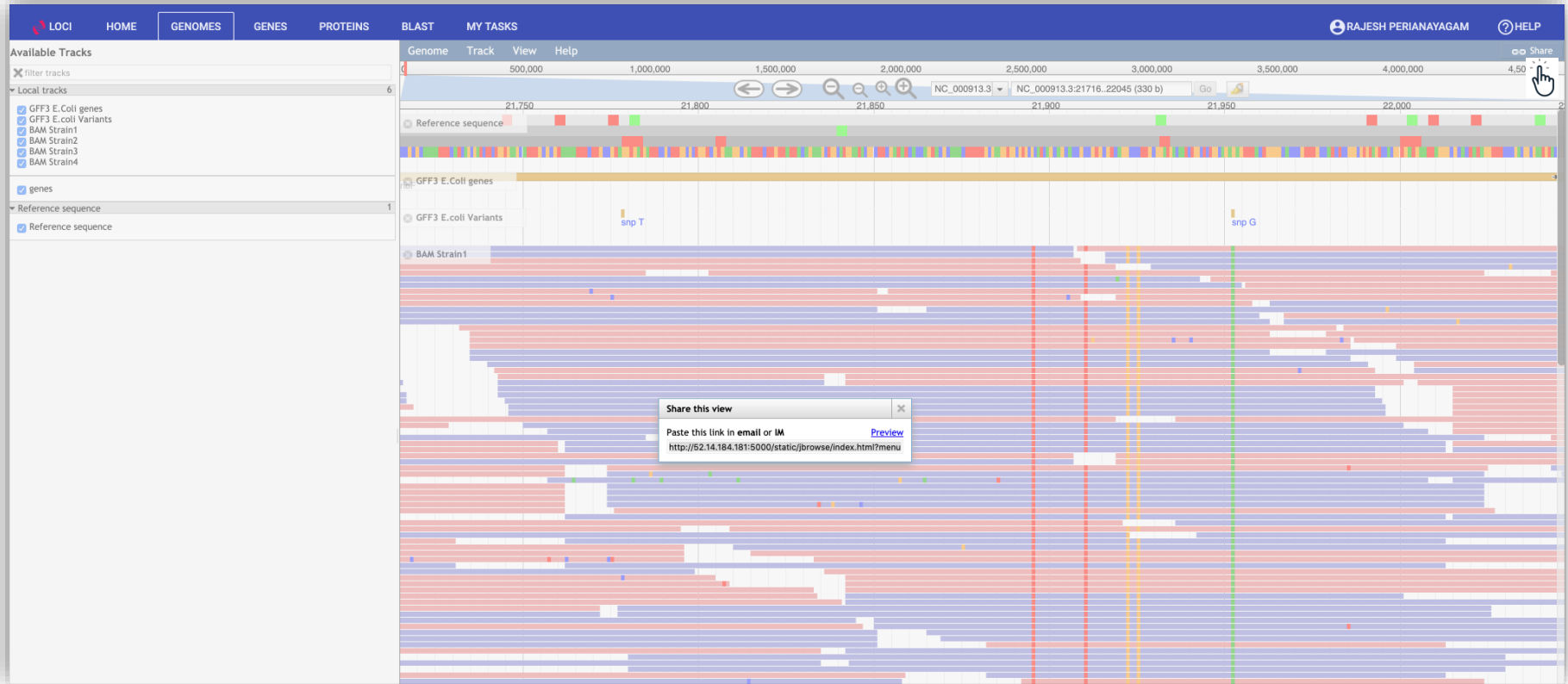
Goal: Discover mutations in *ribF* gene of *E.coli* strain 1 and compare it with 3 other strains





viii) Sharing the results among colleagues

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



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

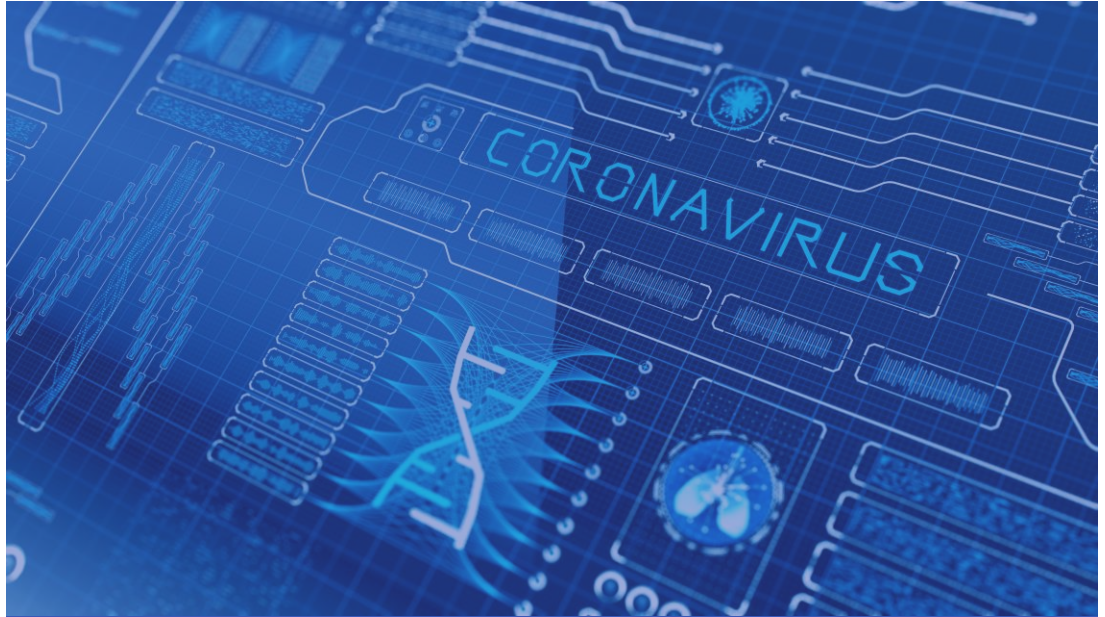
Case Study 2

Conclusions



Case Study 3

Microbial Discovery

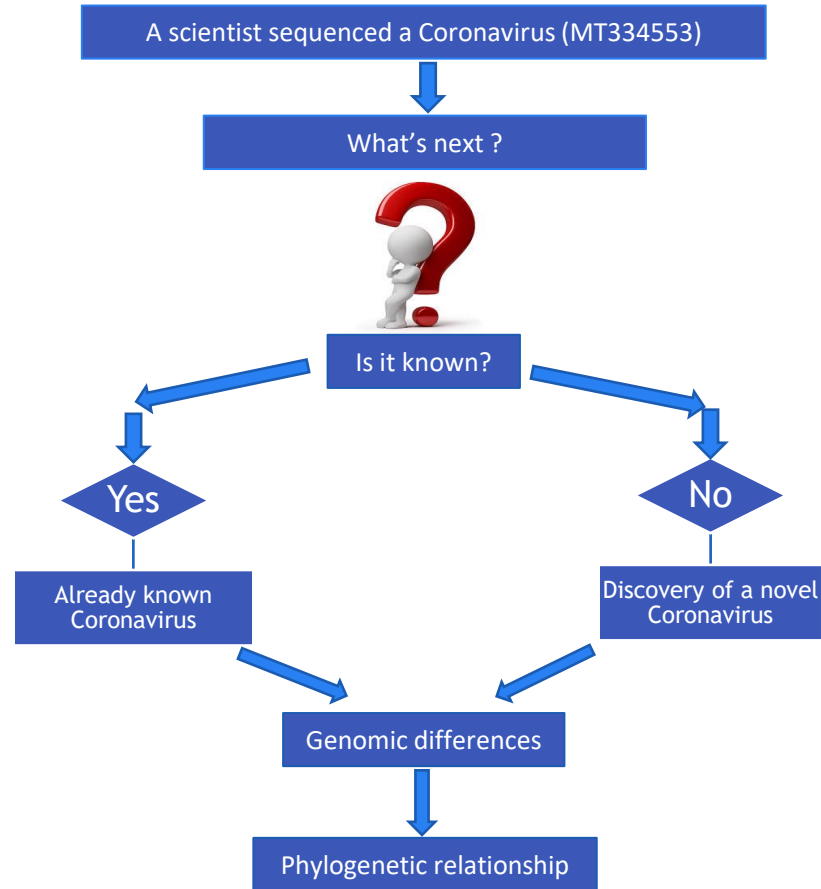


Goal

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

Project Plan

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



Project Plan

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

A scientist sequenced a Coronavirus (MT334553)



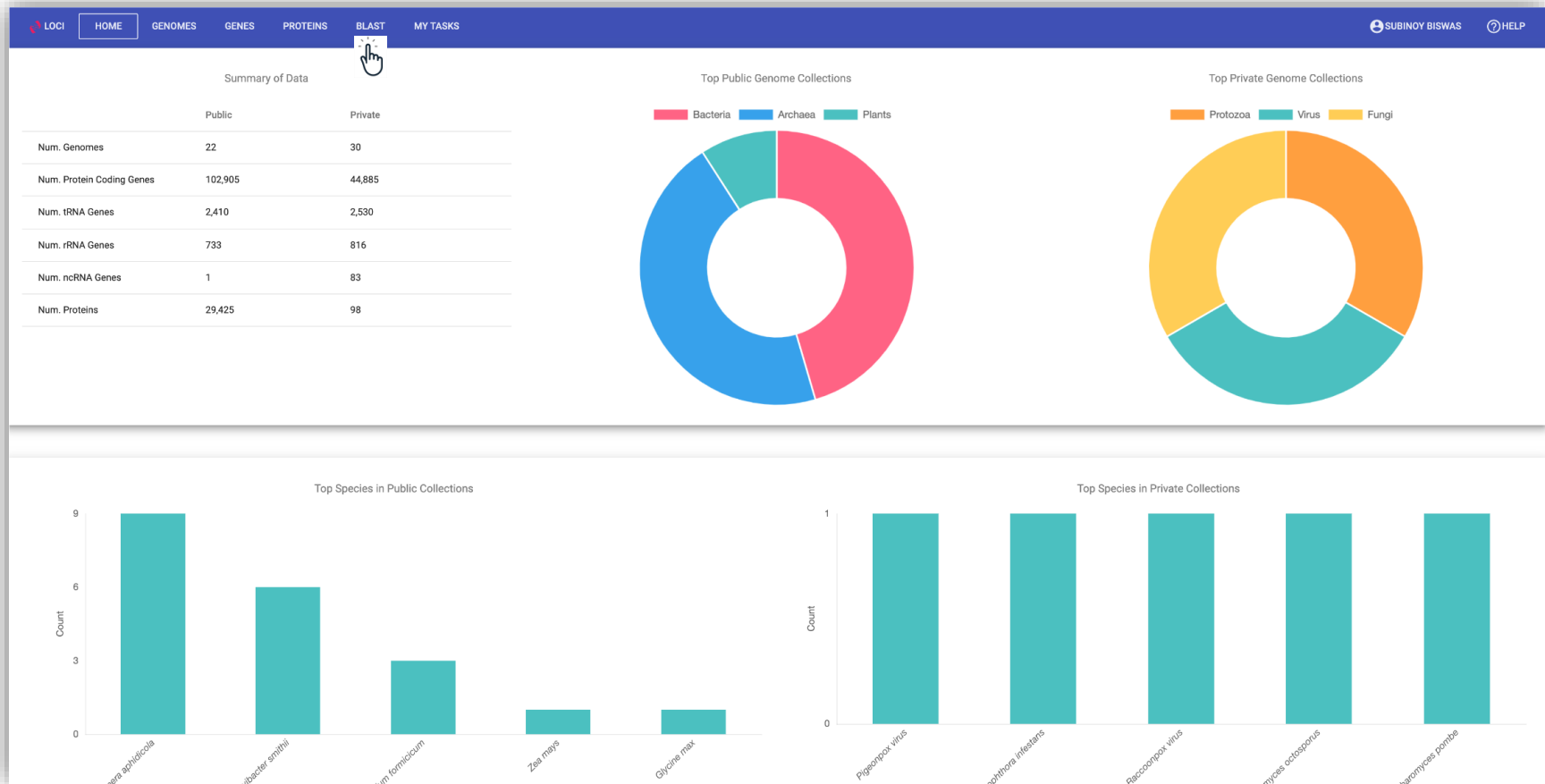
What next ?



Case Study 3

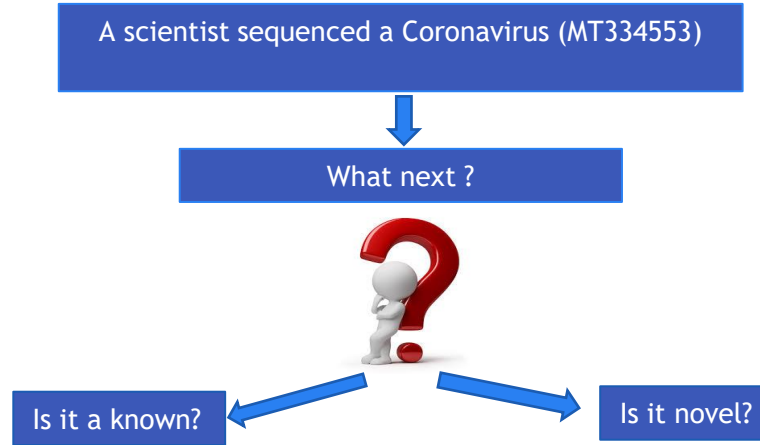
i) Data management of WGS data of multiple organisms within your secured environment

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



Project Plan

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



ii) Homology analysis using Coronavirus database

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

LOCI

HOME

GENOMES

GENES

PROTEINS

BLAST

MY TASKS

SUBINJOY BISWAS

HELP

>MT334553.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-00284/2020
TTTGGATCTCTGTAGATCTGTTCTCTAAACGAACCTTAAAATCTGTGTGGCTGTCACCTGGCTCATGC
TTAGTGCACCTCACGAGTAAATAAATACTAATTAATCTGCTGAGGACACGAGTAACCTGCTATC
TTCTGCAGGCTGCTTACGGTTTCGCTCCGTGTCAGCGATCATCAGCATCTAGGTTTCGCTCCGGGTG
TGACGAAAGGTAAAGTGAAGGCTTGCCTGGTTTCAACGAGAAACACACGTCCAACCTCAGTTTGC
CTGTTTACAGGTTCCGACGTGCTGCTACGTGGCTTTGGAGTCCGTGGAGGAGTCTTATCAGAGGC
ACGTCACATCTTAAAGATGGCACTTGTGGCTAGTAGAGTTGAAAGGCGTTTGGCTCAACTGAA
CAGCCCTATGTTTCATCAACGTTCCGATGCTGAACTGCACCTCATGTGCTATGTTGAGTGG
TAGCAGAACTCGAAGGCTTACGTACGGTGGTGGTGGAGACCTGGTGGCTTGTCCCTCATGTGGG
CGAAATACAGTGGCTTACGCAAGGTTCTTCTCGTAAGAACGTAATAAGGAGCTGGTGGCCATAGT

Select Pre-formatted Databases - OR - Custom Databases

Pre-formatted Databases

Custom Databases

Nucleotide Databases

Protein Databases

☐ Genes - All Public Collections ⓘ

☐ Genes - All Private Collections ⓘ

☐ Genomes - All Public Collections ⓘ

☐ Genomes - All Private Collections ⓘ

☒ NCBI SARS-CoV-2 Genomes Databases ⓘ

☐ NCBI 18S_fungal Databases ⓘ

☐ NCBI 28S_fungal Databases ⓘ

☐ NCBI ITS_Refseq Databases ⓘ

☐ Proteins - All Public Collections ⓘ

☐ Proteins - All Private Collections ⓘ

Select a Program

☒ blastn ⓘ (nucleotide query, nucleotide database)

☐ tblastn ⓘ (protein query, nucleotide database)

☐ blastp ⓘ (protein query, protein database)

☐ blastx ⓘ (nucleotide query, protein database)

BLAST

RESET

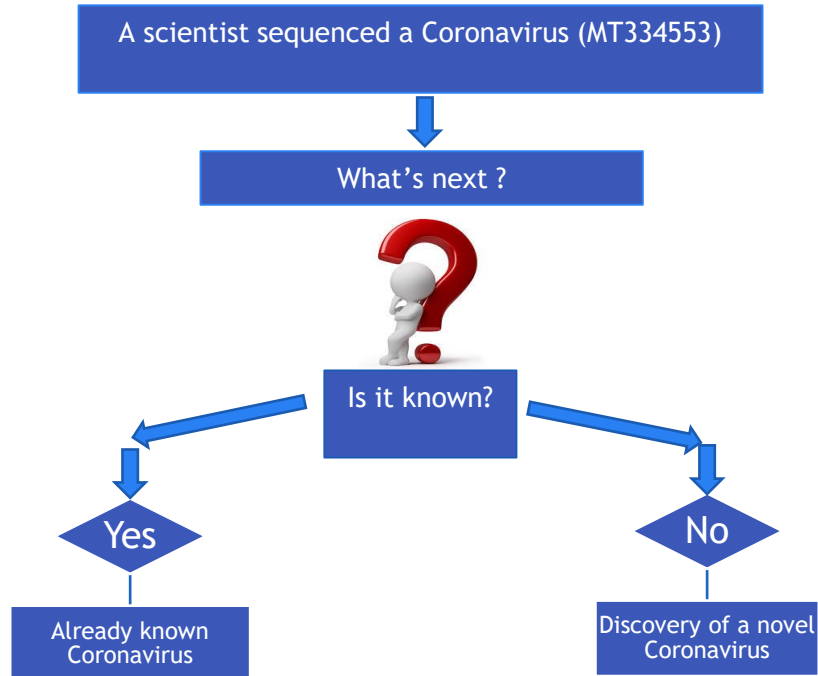
Optional

Task Name

MT334553_SARS-CoV-2

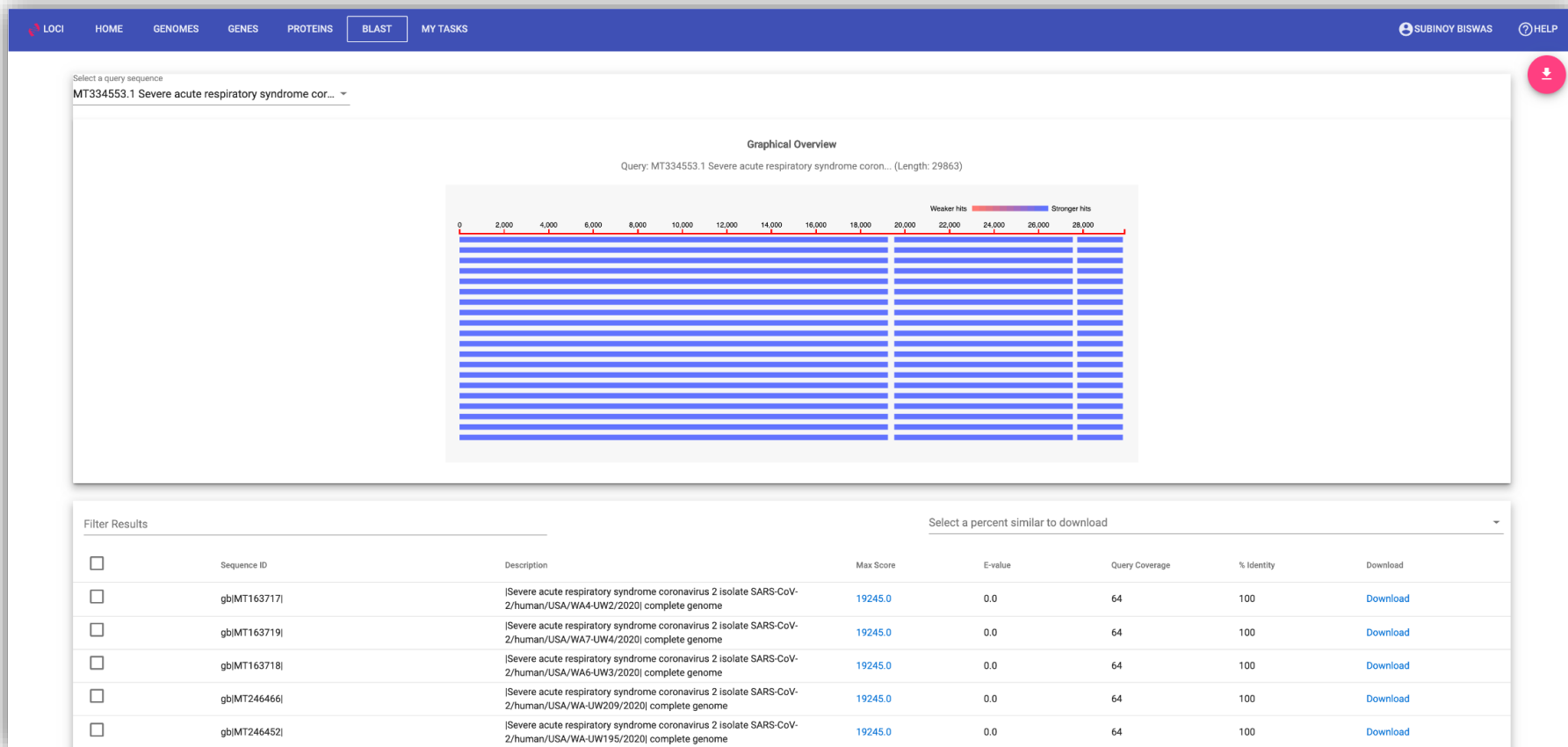
Project Plan

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



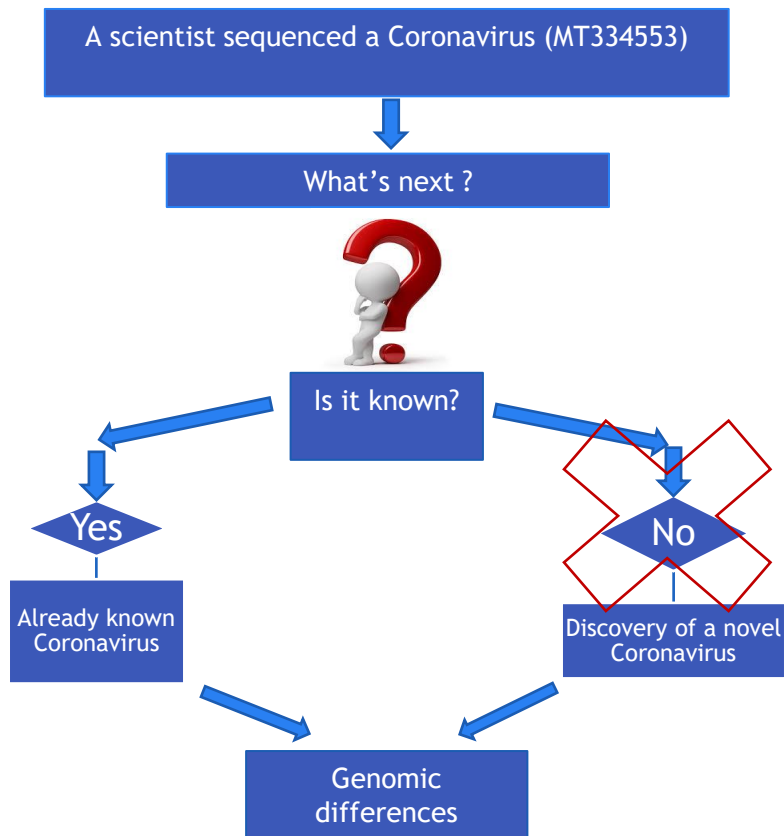
iii) High homology (SARS-CoV-2)

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



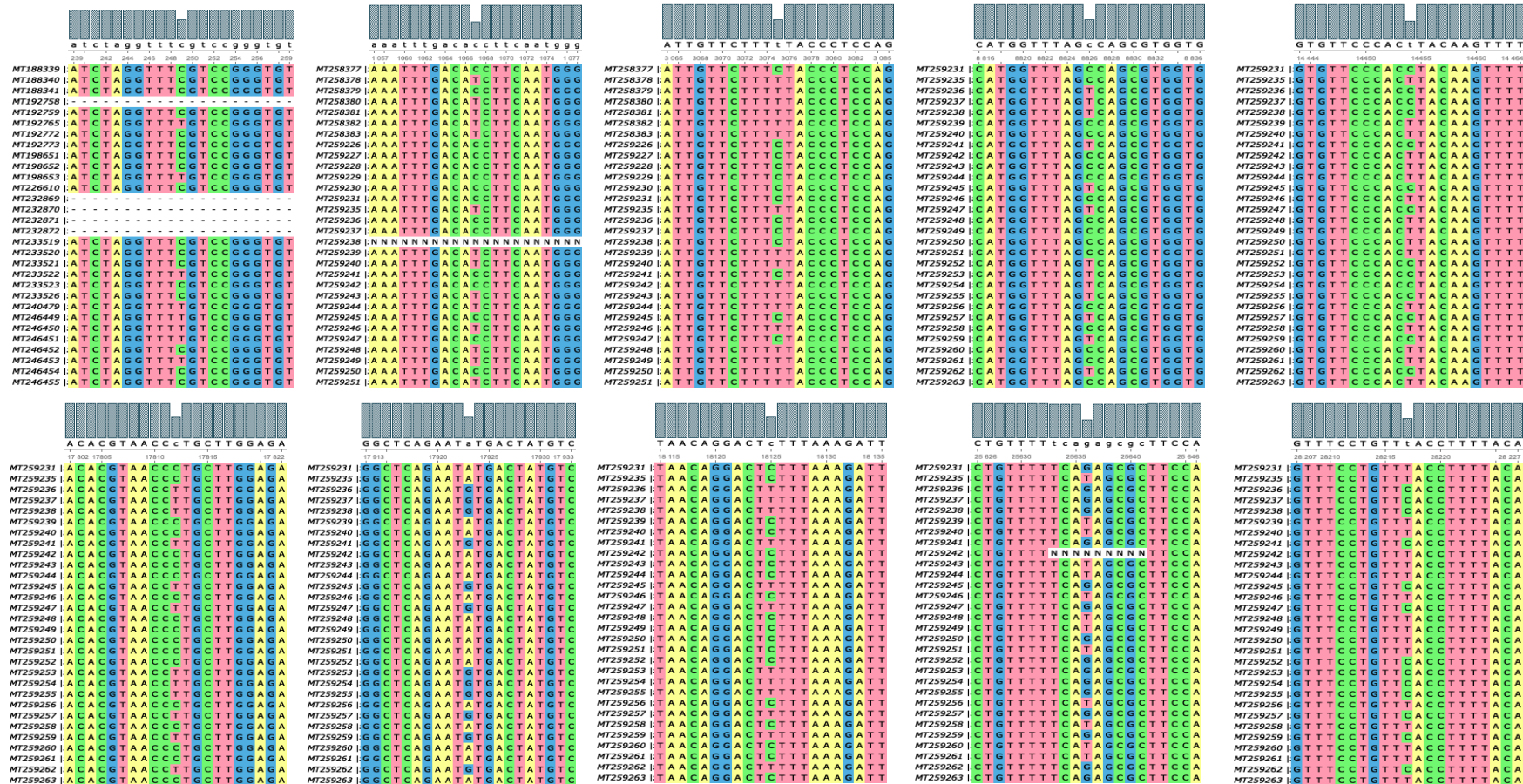
Project Plan

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



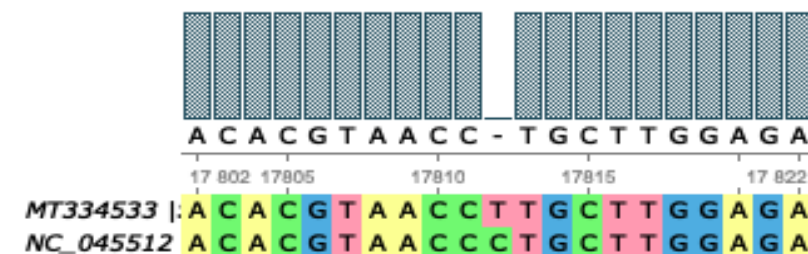
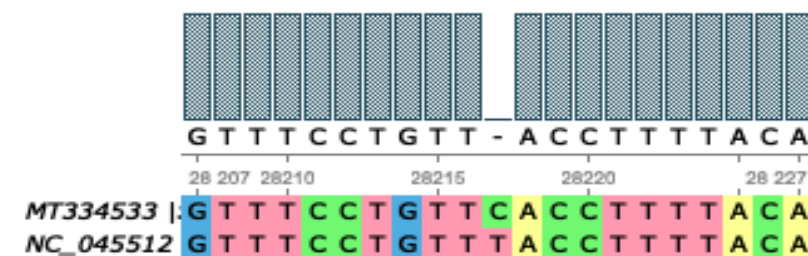
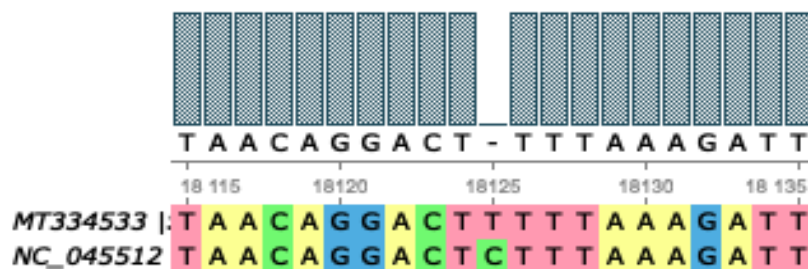
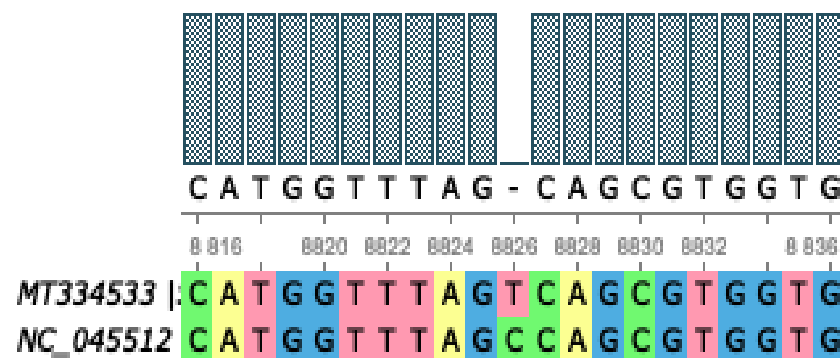
iv) Multiple Sequence Alignment to discover the differences in various genomic regions

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



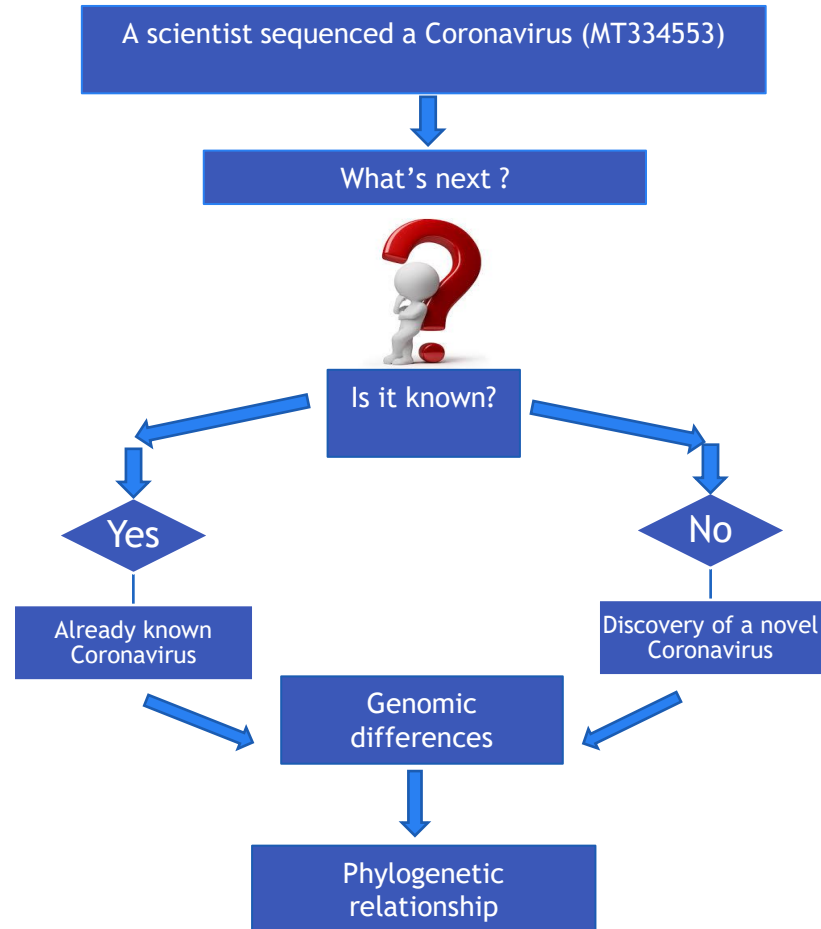
v) Sequence alignment of MT334553 with NC_045512.2 (Reference SARS_CoV-2)

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



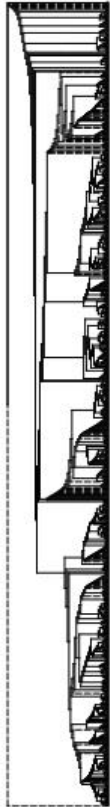
Project Plan

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



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



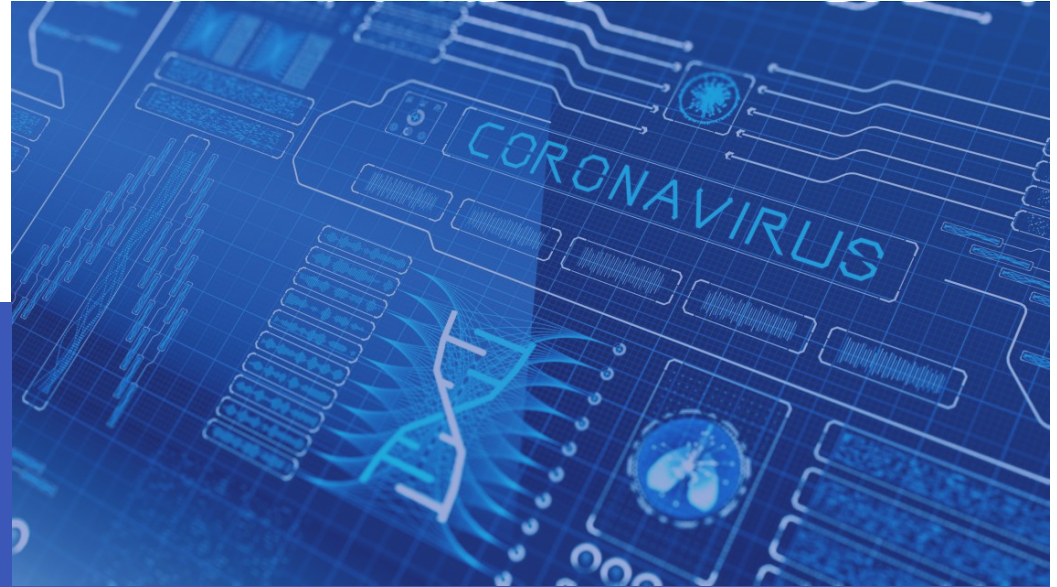
MT259266 clustered together with USA/WA and USA/MI sample

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

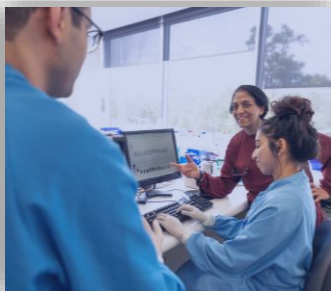
Data Integrity = No data prep, central repository, no data loss

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

