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 Phytobiomes Alliance webinar purposes only

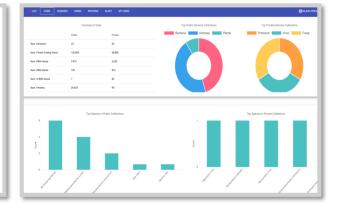




Introduction

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

GCF_00000836 GCF 000025885.1 CCF 000067045 1 GCF 000144125.1 GCE 000145035 1 CCF 0001546451 GCF 000160075 GCF 000022525 GCF_000161515 GCF_000172075.1 C/CE 000174955 1 GCF 000177735.1 GCF 000178335.1 GCF 000179795.1 GCF 000185245.1 CCE 000185685 GCF 000187975. GCF_000213595.1 GCF 000214395.1 GCF 000215025.1 GCF 000215465.1 GCF 000215795.4 GCF 000216415.1 CCE 000220835 GCF 000237805.1 GCF 000226175.1 GCF 000227525.1 GCF 000229645.1 GEF 000232345.1 GCF 000233895 GCF 000236965.1 GCF_000239135.1 GCF 000239475.1 GCF_000240225.1 GCF_000242775.1 GCF 000246515.1 GCF_000247145.1 GCF 000253095.1 GCF_000253535.1 GCF_000253935.1 GCF_000254035.1 GCF_000254175.1 GCF_000254235.1 GCF_000258125.2 GCF_000259025.1







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

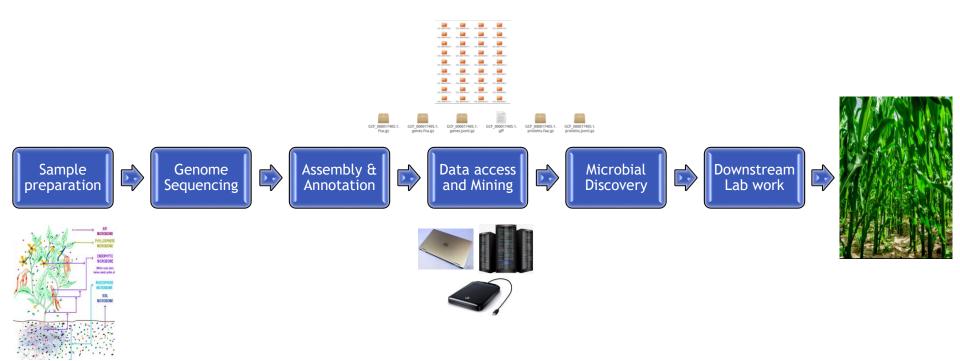
Challenges

- 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

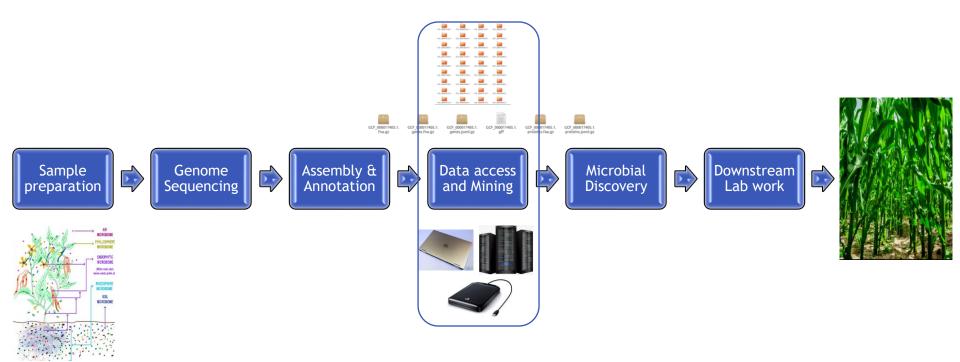
Typical Workflow

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



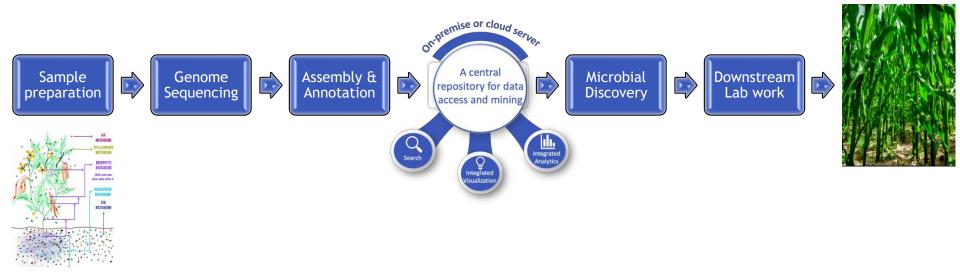
Case Study 1

Typical Workflow



Loci in your internal workflow

Case Study 1

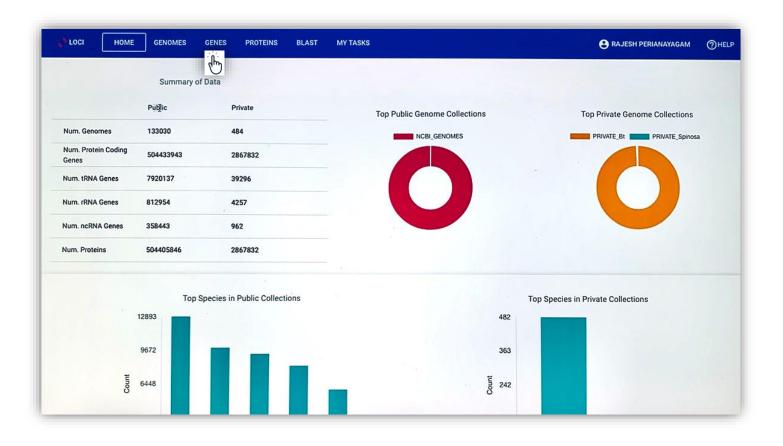


Login to Loci within your secured environment

Case Study 1

¢ ³ LOCI	HOME	GENOMES	GENES	PROTEINS	BLAST	MY TASKS		() HELP
							Login	
							rnaeshpn@karyosoft.com	
							Password	
							Login	

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



i) How many pesticidal genes are in private collections?

Case

Study 1

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
		Enter gene name, ID. species, genus, class, etc. pesti pesticidal crystal protein cry1Ba (description) pesticidal protein Cry5Aa (description) pesticin (description) pesticidal protein Cry26Aa (description)	Search Reset Advanced	-	
		Select collections to search in:	All Private Collections		
		Public_1	✓ Private_1		
		Public_2	Vivate_2		

Search for the genes by description

i) How many pesticidal genes are in private collections?

Case

Study 1

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

		Enter gene name, ID, species, genus, clas pesticidal crystal protein cry1Ba 🔘 pestic	dal protein CrySAa 🛞 pesticidal protein CryZ6Aa 🛞	Ivanced		-		Download FASTA - All Ge Download FASTA - Select Download CSV - All Gen Download CSV - Selecte
		Select collections to search in:	lic Collections	All Private Collections				
		—						
		Pub	ic_1	Private_1				
		Pub	ic_2	Private_2				
								Num. results: 2
Filter Results								
Selected	ID	Туре	Description	Species	Assembly	Collection	Length (bp)	Analysis
	GCF_000161515.1_gene6250	protein_coding	pesticidal crystal protein cry1Ba	Bacillus thuringiensis	GCF_000161515.1	Private_1	3687	Browse BLAST Download
	GCF_000177735.1_gene3770	protein_coding	pesticidal protein Cry5Aa	Acinetobacter baumannii	GCF_000177735.1	Private_1	276	Browse BLAST Download

Discovery of 2 pesticidal genes in their private collections

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

LOCI HOME GENOMES GENES PROTEINS BLAST MY TASKS	😝 RAJESH PERIANAYAGAM 🖉 ()HELP
<pre>>GCF_000161515.1_gene6250 pesticidal crystal protein crylBa MyLociAttributes=[mylocidb=GCF_000161515.1_my TGCACTCOAATAGGAAAAATGAAATGAAATGAAATGCGTATGCGAATCGAATCCACACAAATGGAACCAATTACCAGA GGTATTAACATAGCGATAGCTGGTGAAAATGAGAGCGAAGGGAACAATATGCGACCAATTGTAGCCATCCAAACGG GTATTAACATAGCTGGTGAGAATACTAGGCGTATTGGGCGACCGTTGGTGGACAACTAGCTAG</pre>	
Pre-formatted Databases Custom Databases	
Nucleotide Databases	Protein Databases
Genes - All Public Collections	Proteins - All Public Collections
Genes - All Private Collections	Proteins - All Private Collections
Genomes - All Public Collections	NCBI nr
Genomes - All Private Collections	NCBI env_nr - Protein Sequences from Metagenomes
NCBI nt	NCBI Patent Protein Sequences
NCBI env_nt - Nucleotide Sequences from Metagenomes	
🕝 र 🔽 NCBI Patent Nucleotide Sequences	
NCBI Bacterial and Archaeal 16S rRNA 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 Control of the second	RESET
Task Name MultiBLAST_Cry,	

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

Filter Results				
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	
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: 1 - 10 of 41 <

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

ii) Are *cry1Ba* in private collections patentable?

I HOME GENOMES GENES F	PROTEINS BLAST MY TASKS						😝 RAJESH PERIANAYAGAM
Select a query sequence GCF_000161515.1_gene6250 pesticidal cry	ygt						Download FASTA - Download FASTA -
			raphical Overview				Download CSV - All
		Query: GCF_000161515.1_gene	e6250 pesticidal crystal protei (Length: 3687)				Download CSV - Se
		<u>t m im im</u>					
Filter Results Selected	Sequence ID	Description	Max Score	E-value	Query Coverage	% Identity	Download
	gb(H42103.1)	Sequence 1 from patent US 5628995	3681.0	0.0	100	99	Download
	gb((70138.1)	Sequence 3 from patent US 5679343	1477.0	0.0	49	93	Download
	gb(I38760.1)	Sequence 3 from patent US 5616319	1477.0	0.0	49	93	Download
	gb(U31527.1)	Bacillus thuringiensis delta-endotoxin CryET1 (cryET1) gene, complete cds	972.0	0.0	40	88	Download
	gb(173040.1)	Sequence 1 from patent US 5686069	962.0	0.0	39	88	Download
	gb(I11609.1)	Sequence 4 from Patent US 5407825	962.0	0.0	39	88	Download
	emb(A27529.1)	B.thuringiensis PS81A2 endotaxin gene	959.0	0.0	39	88	Download
	gb(/70137.1)	Sequence 1 from patent US 5679343	884.0	0.0	40	86	Download
	gb(I38759.1)	Sequence 1 from patent US 5616319	884.0	0.0	40	86	Download
	dbj[E00881.1]	DNA encoding insecticidal protein	709.0	0.0	39	82	Download
							Items per page: 10 * 1 - 10 of 100 <
Length-#274 Score = 6798 bits (3681), E Jentities = 3685,3687 (994), 51randr-Vus/Plus Ourry 1 TTACTTCARTAGAN Ourry 1 TTACTTCARTAGAN Ourry 61 CAATGATCTATTACC IIIIIIIIIIIIIII Sbjct 186 CAATGATCTATTACC Ourry 121 ANTACGATCCATTGO Ourry 121 ANTACGATCCATTGO		LATT / Ibutes					

ii) Are *cry5Aa* in private collections patentable?

Case Study 1

HOME GEN	ENOMES GENES PROTEINS	BLAST	MY TASKS						S RAJESH PERIANAYAGAM	
	gene6250 pesticidal crystal protei gene3770 pesticidal protein Cry5/	-		Graphical Overvi Query: GCF_000161515.1_gene6250 pesticida)				
			2	500 1.000 1.000 2.00		his Stronge 2,00	his 3.000			
				≡						
Filter Results										
Selected	Sequence ID			Description	Max Score	E-value	Query Coverage	% identity	Download	
Selected	gb 42103.1			Sequence 1 from patent US 5628995	3681.0	0.0	100	99	Download	
Selected										
Selected	gb I42103.1 gb I70138.1			Sequence 1 from patent US 5628995 Sequence 3 from patent US 5679343	3681.0 1477.0 1477.0	0.0	100 49	99 93	Download Download	
Selected	gb (42103.1 gb (70138.1 gb (38760.1)			Sequence 1 from patent US 5628995 Sequence 3 from patent US 5679343 Sequence 3 from patent US 5616319	3681.0 1477.0 1477.0	0.0 0.0 0.0	100 49 49	99 93 93	Download Download Download	

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

Case

Study 1

CI	HOME	GENOMES	GENES	PROTEINS	BLAST	MY TASI	ĸs																				😫 RAJESH	PERIANAYAGAM	() не
	lect a query sequence of the s		0 pesticidal	orotein Cry 👻	• 																								6
										Que	ery: GCF_	_0001777		aphical O 3770 pest		/ rotein Cry5/	Aa (Len	igth: 276)											
							0	20	40	60	80	100	120	140	0	160	180	Weaker hits 200	220	240	onger hits 260								
ĩ	Filter Results																												1
L	Selected		Sequence ID					Description								Max Sc	ore		E-value			Query Cove	erage	% Identi	ity		Download		
L																									Items per j	page: 5	▼ 0 of	0 < >	
ĺ	Lambda		н																										1
	Gapped Lambda	0.621 K 0.460	н																										
	Effective	search spa	ce used: 10	0528182454																									
	Poster Number	d date: Ap	r 23, 2018 in database	7c93970a7d80 10:58 PM : 30,036,802 sse: 99,138	,																								
		lastn matri ties: Exist		tension: 2.5	5																								

iii) Is cry5Aa gene in private collections novel?

HOME GENOME	S GENES PROTEINS BLAST MY TASKS	🕒 RAJESH PERIANAYAGAM ⊘
TTGACTTCAAATAGGAAF TGCTCGTATTGAGGATAG GTATTAACATAGCTGGT GAAACAGATACAGTACG AGAATAG >GCP_000177735.1_C ATGACCATCCTACTTAT GATGGGACCATCTATTG7	gene6250 pesticidal crystal protein crylBa MyLociAttributes=[mylocidb=GCF_000161515.1_m; NATGAGAATGAATTATAATGCTGTATCGAATCATTCCGACGAATGGAATGCAATGCAATGCAAAGG AGAATACTAGGCGTATGGGGGTACCGTTTGCTGGACAACTAGCTAG	
Select Pre-formatted Da	tabases - OR - Custom Databases	
Pre-formatted Databases	Custom Databases	
	Nucleotide Databases	Protein Databases
	Genes - All Public Collections	Proteins - All Public Collections
	Genes - All Private Collections	Proteins - All Private Collections
	Genomes - All Public Collections	NCBI nr
	Genomes - All Private Collections	NCBI env_nr - Protein Sequences from Metagenomes
	NCBI nt	NCBI Patent Protein Sequences
	NCBI env_nt - Nucleotide Sequences from Metagenomes	
	NCBI Patent Nucleotide Sequences	
	C KINA Sequences	
Select a Program		
	blastn (nucleotide query, nucleotide database)	O blastp (protein query, protein database)
	tblastn (protein query, nucleotide database)	 blastx (nucleotide query, protein database)
Optional	BLAST	RESET
	a E	

iii) Is cry5Aa gene in private collections novel?

	ENOMES GENES PROTEINS	BLAST MY TASKS						RAJESH PERIANAYAGAM
elect a query sequence CF_000177735.1_	.gene3770 pesticidal protein Cry 👻							Download FASTA - All
			Graphical Ovi Query: GCF_000177735.1_gene3770 pestic		76)			Download CSV - All Hit
		2	<u>20 40 60 60 120 140</u>	Weak	or his Stranger	ns 90 		Download CSV - Select
Filter Results	Sequence ID		Description	Max Score	E-value	Query Coverage	% Identity	Download
	Sequence ID ref[NZ_AMHS01000055.1]		Description Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole shotgun sequence		E-value 8.64073e-140	Query Coverage	% Identity 99	Download Download
Selected			Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole					
Selected	ref NZ_AMHS01000055.1		Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole shotgun sequence	e genome 270.0 270.0	8.64073e-140	100	99	Download
Selected	ref[NZ_AMHS01000055.1] GCF_000301995.1_gene2736		Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole shotgun sequence pesticidal protein Cry5Aa Acinetobacter baumannii AB_1650-8 AB_1650_8_116, whole genome s	e genome 270.0 270.0	8.64073e-140 8.64073e-140	100	99	Download Download
	refINZ_AMHS01000055.1 GCF_000301995.1_gene2736 refINZ_AMHG01000115.1		Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole shotgun sequence pesticidal protein Cry5Aa Acinetobacter baumannii AB_1650-8 AB_1650_8_116, whole genome s sequence	e genome 270.0 270.0 267.0 267.0	8.64073e-140 8.64073e-140 4.02014e-138	100 100 100	99 99 98	Download Download Download
Selected	refINZ_AMHS01000055.1 GCF_000301995.1_gene2736 refINZ_AMHG01000115.1 GCF_000301255.1_gene3194		Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole shotgun sequence pesticidal protein Cry5Aa Acinetobacter baumannii AB_1650-8 AB_1650_8_116, whole genome s sequence pesticidal protein Cry5Aa Acinetobacter baumannii TG20546 AB_TG20546_65, whole genome sh	e genome 270.0 267.0 267.0	8.64073e-140 8.64073e-140 4.02014e-138 4.02014e-138	100 100 100 100	99 99 98 98	Download Download Download Download
Selected	refINZ_AMHS01000055.1 GCF_000301995.1_gene2736 refINZ_AMH601000115.1 GCF_000301255.1_gene3194 refINZ_ASFI01000064.1		Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole shotgun sequence pesticidal protein CrySAa Acinetobacter baumannii AB_1650-8 AB_1650_8_116, whole genome s sequence pesticidal protein CrySAa Acinetobacter baumannii TG20546 AB_TG20546_65, whole genome sh sequence	e genome 270.0 270.0 267.0 267.0 267.0 264.0 264.0	8.64073e-140 8.64073e-140 4.02014e-138 4.02014e-138 1.87039e-136	100 100 100 100 100	99 99 98 98 98	Download Download Download Download Download
Selected	refINZ_AMHS01000055.1 GCF_000301995.1_gene2736 refINZ_AMH601000115.1 GCF_000301255.1_gene3194 refINZ_ASFI01000064.1 GCF_000453905.1_gene2412		Acinetobacter baumannii AB_2009-04-01-7 AB_2009-04-01-7_56, whole shotgun sequence pesticidal protein Cry5Aa Acinetobacter baumannii AB_1650-8 AB_1650_8_116, whole genome s sequence pesticidal protein Cry5Aa Acinetobacter baumannii TG20546 AB_TG20546_65, whole genome sh sequence pesticidal protein Cry5Aa Acinetobacter baumannii ABNIH16 contig00002, whole genome shotgu	e genome 270.0 270.0 267.0 267.0 264.0 264.0	8.64073e-140 8.64073e-140 4.02014e-138 4.02014e-138 1.87039e-136 1.87039e-136	100 100 100 100 100 100	99 99 98 98 98 98	Download Download Download Download Download 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

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

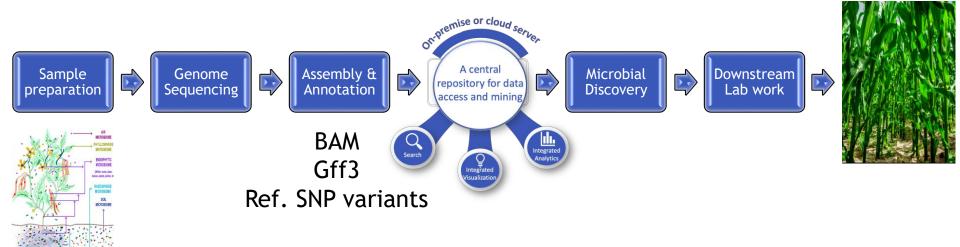
- 1) Lack of user-friendly visualization tools or availability of stand-alone tools
- 2) Too many files to mine to discover a mutation; Labor intensive Process
- 3) No immediate availability bioinformaticians due to a lot of backlog work supporting various projects

Case Study 2 Strain Engineering



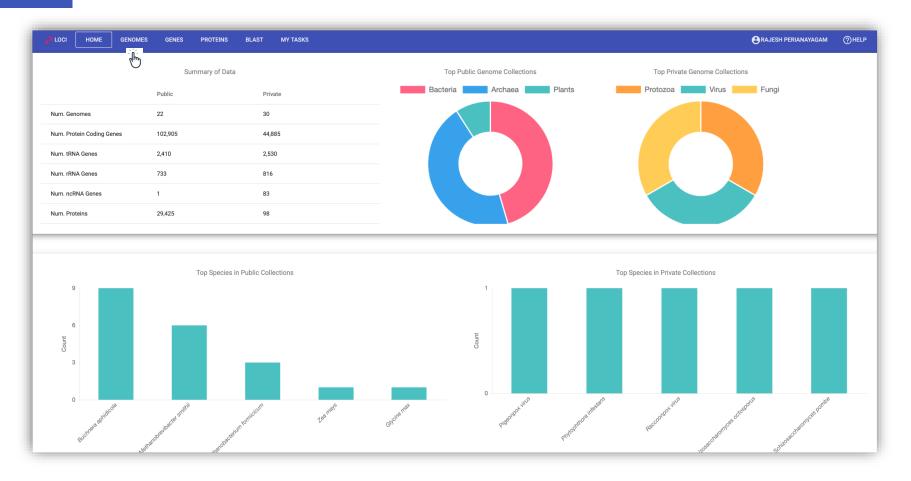
Loci in your workflow

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



Case Study 2

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



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

LOCI HOME GENOMES GENES PROTEINS BLAST MY TASKS	😁 RAJESH PERIANAYAGAM	() HELP
Enter strain name, species, genus, class, etc.		1
esc Escherichia coli (species) Escherichia (genus)		
Escherichia (genus) Escherichia coli (species)		
Search Reset Advanced		

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

Case Study 2

								Download CSV - All Ger
	E	Enter strain name, species, g	jenus, class, etc.					Download CSV - Selecte
		Escherichia coli 🔇		Search Reset	Advanced			
Filter Results						Species. Intraspecie	es Name. Collection Name. Ge	Num. results: 2
Filter Results	Assembly 💽	Species 🕕	Intraspecies Name 🕕	Collection ()	Size(Mb) 🚺	Species, Intraspecie	es Name, Collection Name, Ge	Num. results: 2
	Assembly C GCF_000005845.2	Species 💽 Escherichia coli	Intraspecies Name 🕐 K-12 substr. MG1655		Size(Mb) 🕖 4.64			enome Size, Number of Genes, 👻
				-		Num. Genes 🚯	Num. Proteins 👔	Analysis O Browse BLAST

ii) *E.coli* Reference sequence track

Case Study 2

LOCI HOME GENOMES GENES PROTEINS	BLAST MY TASKS	BRAJESH PERIANAYAGAM	() HELP
Available Tracks	Genome <u>Track</u> View Help		c-o Share
X filter tracks	500,000 1,000,000 1,500,000 2,000,000 2,500,000 3,000,000 3,500,000	4,000,000	4,500,000
genes	C C C C C C C C C C C C C C C C C C C		
Reference sequence	14,500 14,525 14,550	14,575 K R S A F	R L
Reference sequence	Parference sequence V R R V L I Y A I T W S S P S K K L Y A I T W S S P S K K L Y A I T W S S P S K K L Y A I T W S S P S K K L Y A I T W S S P S K K L Y A I T W S S P S K K L Y A I T W S S C T L L L L L L L L L L L L L L L L L L		

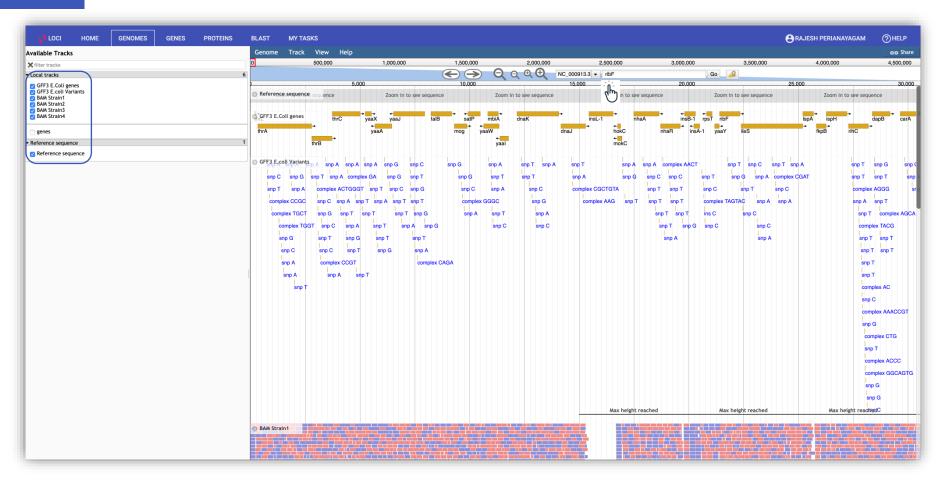
ii) Adding Gff3, Vcf and BAM files

								A	O
LOCI HOME GENOMES GENES PROTEINS BLA								RAJESH PERIANAYAGAM	() HELP
1	enome Track View 500,000	Help 1,000,000	1,500,000	2,000,000	2,500,000	3,000,000	3,500,000	4,000,000	4,500,000
X filter tracks	500,000	1,000,000		-	NC_000913.3 - NC_0009			4,000,000	4,000,000
genes	18,750	20.000		21.250	NO_000010.0 + NO_0000	22,500	23.750	25,000	
* Reference sequence 1	Reference sequence	Zoom in to see sequence	Zoom in to see sequenc	e	Zoom in to see sequence	Zoom in to see sequence		Zoom in to see sequence	Zoom in to see
Reference sequence									
		0			×				
		Open files			~				
			ation of data files and URLs,						
		will automatical	ly suggest tracks to display the	eir contents.					
		Local files	Remote l	JRLs - one per lir	ie i				
		Select Files	http://paste.	urls.here/example.bam					
		Select or drag file	es here.						
					11				
		Files and URLs		_					
		BAM - Strain1			×				
		BAM index Strain1 BAM Strain2			****				
		BAM index Strain2 BAM Strain3	.bam.bai		- X				
		BAM index - Strain3	.bam.bai		<u> </u>				
		BAM	.bam.bai		×				
		GFF3 - E.Coli_ GFF3 - E.coli_	genes.gff3		× I				
			—		~				
		New Tracks							
		Name	Display						
		BAM Strain1 BAM Strain2	Alignments2 -						
		BAM Strain2 BAM Strain3	Alignments2 - Alignments2 -						
		BAM Strain4	Alignments2 •						
		GFF3 E.Coli genes	CanvasFeatures •						
		GFF3 E.coli Variants	CanvasFeatures •	Edit Configuratio	n 🗙				
		۲	Open immediately 🔵 Add to	o tracks					
			🗙 Cancel 🗁 Open						

iii) Search for *ribF* gene

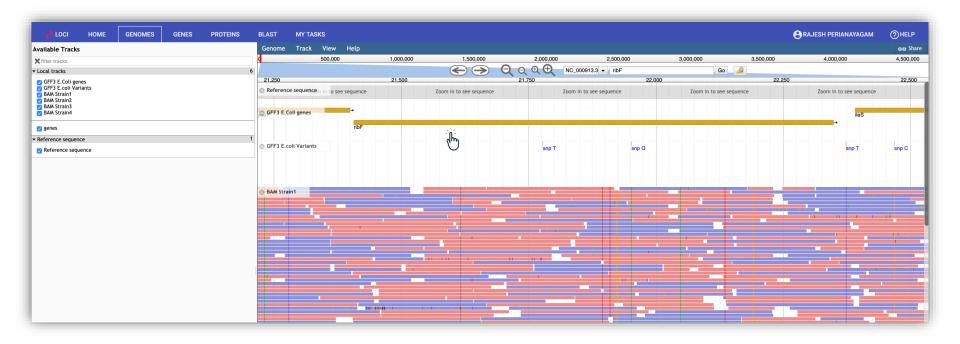
Case

Study 2



iii) ribF gene

Case Study 2

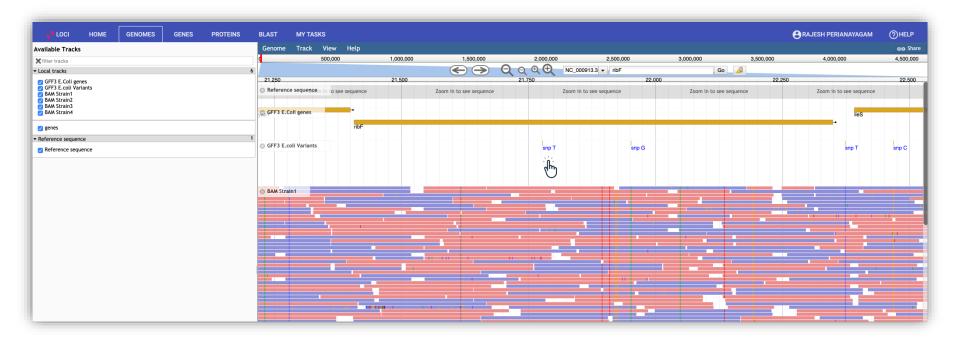


iv) Details of *ribF* gene

COLOCI HOME GENOMES GENES PROTEINS	BLAST	MY TASKS								BRA	JESH PERIANAYAGAM	?HELP
Available Tracks	Genome	Track View										œ⊕ Share
X filter tracks		500,00		1,000,000	1,500,000	2,000,000	2,500,000	3,000,000	3,500,0	00	4,000,000	4,500,000
▼ Local tracks	21,250		mRNA ribF					122022528 (1.31 Kb)	Go 🔬	22,250		22,500
GFF3 E.Coli genes GFF3 E.coli Variants		e sequence in to se	Primary Data	I			ce		see sequence	Andrey In 10710	Zoom in to see sequence	
BAM SRR11616628 BAM SRR11616636 BAM SRR11617149			Name	ribF								
Z BAM SRK1161/149 Z BAM SRR11617886	GFF3 E.C	oli genes	Туре	mRNA							ileS	
genes			Position	NC_000913.3:21	40722348 (+ strand)						→	
Reference sequence			Length	942 bp								
Z Reference sequence	GFF3 E.co	oli Variants	Attributes				snp G				snp T	snp C
			dbxref	ASAP: ABE-0000091	ECOCYC:EG11079	EcoGene:EG11079 GeneID:949129	2					
			gbkey	Gene								
	BAM SRR1	1616628	id	mRNA-b0025								
			locus_tag	b0025								
				pe protein_coding								
			mrna_name									
			mrna_syno	nym ECK0026 yaaC								
			mrnaid	mRNA-b0025								
			protein_id	NP_414566.1								
			seq_id	NC_000913.3								
			source	RefSeq								
			Region seq	uence								
						FA:	STA CE					
			length ATGAAG ACTATT GAAGAA GAACTT GCGTTA CTTGCC CAGAAA GGCGTG Subfeature	=942 CTGATACGCGGCATAC GGTAATTTCGACGGCCAACC GGCACCATACGATAAAG GCAGAGTGGCGCATA ACCGCGCATAAATTTC GTAGGGATAGATTCC GCTGGCATGGAATACC CCCATCAGCAGCACCC S	CATAATCTCAGCCAGG TTGCATCGCGGTCATCG TTACCGGTGATGGTGAT CCCCCGGCAAGACTGAC CCCCGGCCAGGCTGGTGCG TCAGCGATCTTCTGG GCTTTGGGCTGGTC GCTTTGGTATCACCAC	(+ strand) class=mRNA CCCCGCAAGAAGGGTGTGTGCTG SGCGCGTGTAAGGGCTTGCAG CCGGCTGCGGAAAAACTGCAT CCGGTTGCGAGGCGCTTGCGG TGAAGCATTTGCGCGGTAAAATTT SGCAGGCGATTTGCGCGGTAAAATTT SGCAGGCGATTTGCGCGGT TTGCGGATGACAATCTGGCTCTG				-1		
			Primary									
		Max heigh	Name	NP_414566.	1			Max hei	ght reached		Max height reached	
	BAM SRR1	4444424	Туре	CDS								
	DAM SRR1	1010030									T	_

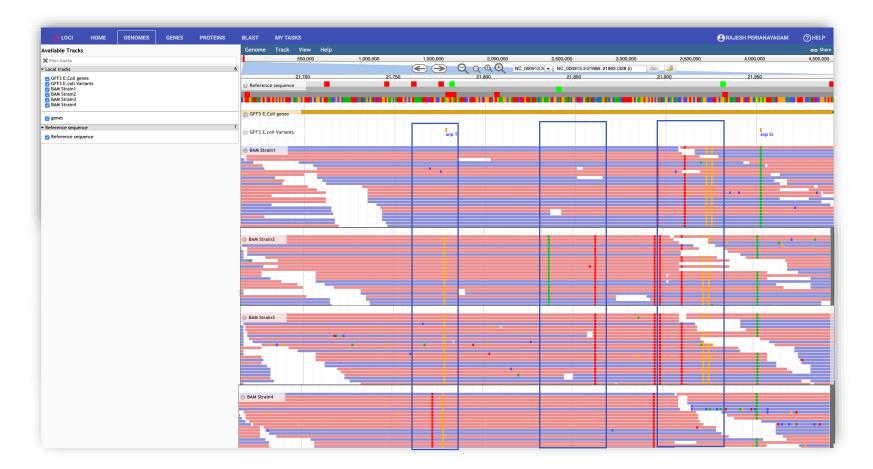
v) Reference SNP details

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

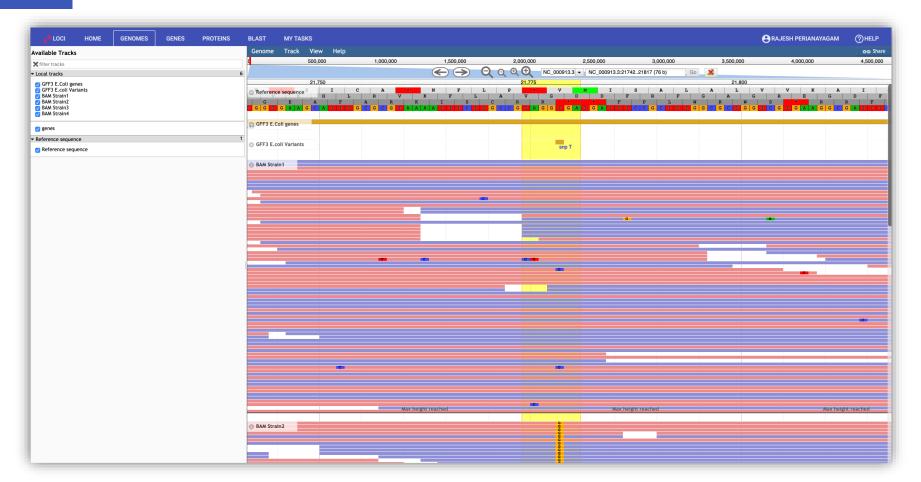


Case Study 2

vi) ribF gene variation among 4 E.coli strains



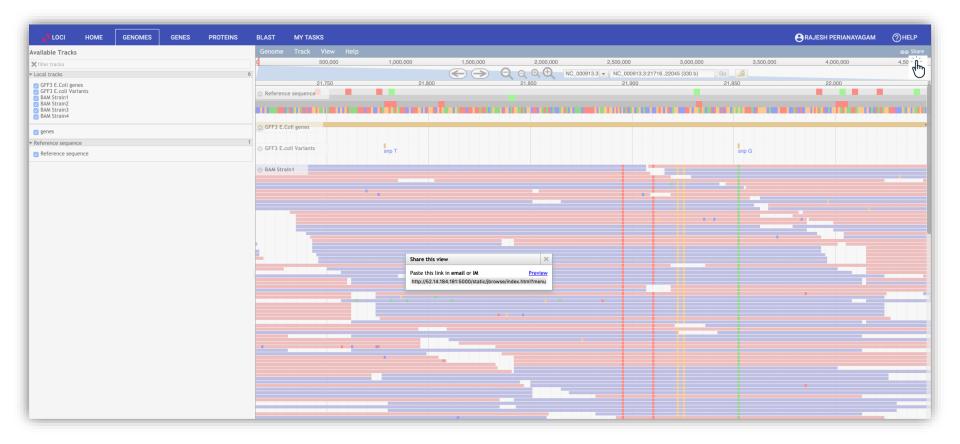
vii) Correlating the mutation to amino acid change instantly



viii) Sharing the results among colleagues

Case

Study 2



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
- 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

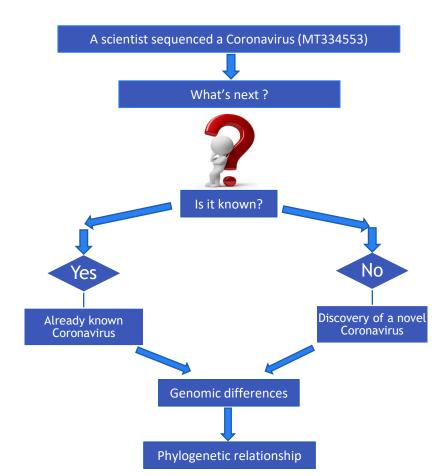


Case Study 3 Microbial Discovery



Goal

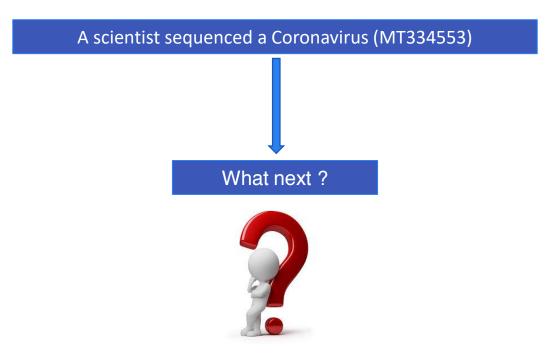
Goal: Understanding of Coronavirus (SARS-CoV-2) of interest using whole genome sequence data from NCBI



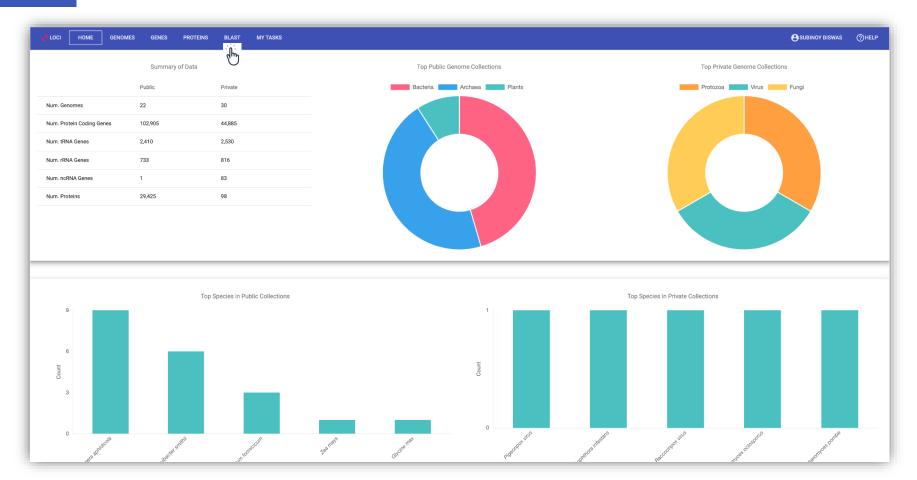
Case Study 3

Case

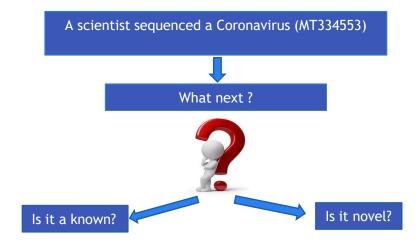
Study 3



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



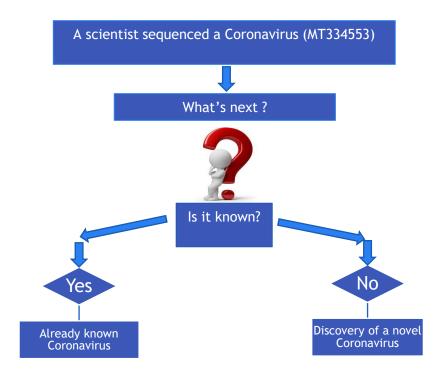
Case



ii) Homology analysis using Coronavirus database

LOCI HOME GENO	MES GENES PROTEINS BLAST MY TASKS			SUBINOY BISWAS	() HELP				
	MT334553.1 Severe acute respiratory syndrome coronavirus 2 isolate SABS-CoV-2/human/USA/UT-00284/2020 TTICGATCTCTTGTACATCTGTTCTCTAAAACGAACTTTAAAATCTGTGTGGGCTGTCATCGGCTGCATCGC								
	TTAGTGCACTCAOGCAGTATAATTAAACAATTAACTGACTGGGACAGGACA								
	TGACCCRAMGGTAGCGGGGCCTTGTCCCGGGGTTCGACGGGAAAACCACACGTCAGTTGC CTGTTTACAGGTTGCGGACGTGCTGGTAGGGTTCGGAGGAGGGGGTCTATACAGAGGC ACGTCAACATCTTAMAGATGCACTTGTGGCTATGTGGAAGTGGCATTGCGCCTCAACTGAA								
	CAGCCCTATGTGTTCATCAAACGTTCGGATGCTCGAACTGCACC TAGCAGAACTCGAAGGCATTCAGTACGGTCGTAGTGGTGAGACA	CTTGGTGTCCTTGTCCCTCATGTGGG							
	CGAAATACCAGTGGCTTACCGCAAGGTTCTTCTTCGTAAGAACG	STAATAAAGGAGCTGGTGGCCATAGT		3					
	Select Pre-formatted Databases - OR - Custom Databases								
	Pre-formatted Databases Custom Databases								
		Nucleotide Databases	Protein Databases						
		Genes - All Public Collections	Proteins - All Public Collections Proteins - All Private Collections						
		Generes - All Private Collections	Proteins - All Private Collections						
		Genomes - All Private Collections							
		CEI SARS-CoV-2 Genomes Databases O							
		🗌 NCBI 18S_fungal Databases 🕒							
		NCBI 28S_fungal Databases							
		NCBI ITS_Refseq Databases							
	Select a Program								
		 blastn (nucleotide query, nucleotide database) (the state of the s	 ○ blastp ● (protein query, protein database) ○ blastx ● (nucleotide query, protein database) 						
		BLAST	RESET						
	Optional Task Name								
	• MT334553_SARS-CoV-2			_	_				

Case

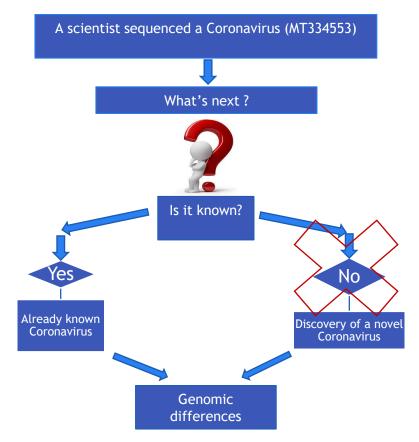


iii) High homology (SARS-CoV-2)

	OMES GENES	PROTEINS	BLAST	MY TASKS						SUBINOY BISWAS	?
elect a query sequence	acute respiratory syn	drome cor 👻									1
					Graphical Overview Query: MT334553.1 Severe acute respiratory sync		9863)				
				°	2.000 4.000 8.000 8.000 18.000 12.000 14.000 16.000		sakar his 500 22,000 24,000 28,000	anger His 28,000			
Filter Results						S	elect a percent similar to	download			
Filter Results	Sequence ID				Description	St	elect a percent similar to E-value	download Query Coverage	% identity	Download	
	Sequence ID gb]MT16371	71			Description Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV- 2/human/USA/WA4-UW2/2020[complete genome				% Identity 100		-
					Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-	Max Score	E-value	Query Coverage			-
	gb MT16371	9			Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV- 2/human/USA/WA4-UW2/2020] complete genome [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-	Max Score	E-value	Query Coverage	100	Download	-
	gb MT16371 gb MT16371	9) 8)			Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV- 2/human/USA/WA4-UW2/2020] complete genome [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV- 2/human/USA/WA7-UW4/2020] complete genome [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-	Max Score 19245.0 19245.0	E-value 0.0 0.0	Query Coverage 64 64	100	Download Download	

Case

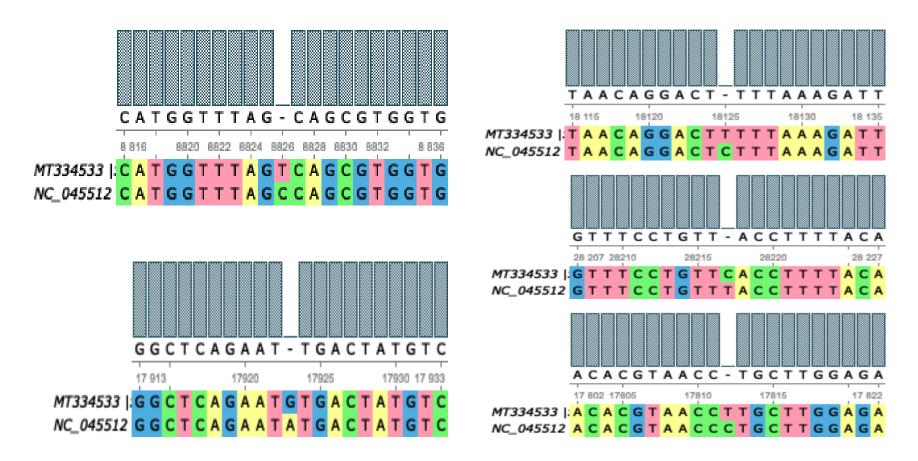
Study 3



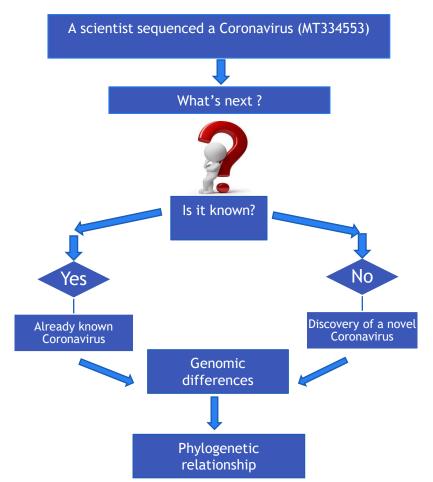
iv) Multiple Sequence Alignment to discover the differences in various genomic regions Goal: Understanding of Coronavirus (SARS-CoV-2) viruses using whole genome sequence data from NCBI

atctaggttttcgtccgggtgt		ATTGTTCTTTLTACCCTCCAG	CATGGTTTAGCCAGCGTGGTG	GTGTTCCCACTTACAAGTTTT
200 242 244 246 246 246 250 252 254 256 257 MT188339 I.A T C T A G G T T T C G T C C G G G T G T MT188340 I.A T C T A G G T T T C G T C C G G G T G T MT188341 I.A T C T A G G T T T C G T C C C G G G T G T MT192758 I	107 1060 1092 1094 1098 1094 1097 1072 1074 1077 MT258777 A A A T T T G A C A C C T T C A A T G G G MT258779 A A A T T T G A C A T C T T C A A T G G G MT258779 A A A T T T G A C A T C T T C A A T G G G MT258879 A A A T T T G A C A T C T T C A A T G G G	3 0 65 20 68 20 79 20 72 20 74 20 78 20 79 20 79 20 79 20 78	4 1 = 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	i 4 44 4 440 14400 14455 14400 14455 M7259231 i G T G T T C C C A C C T A C A A G T T T T M7259235 i G T G T T C C C A C C T A C A A G T T T T M7259235 i G T G T T C C C A C C T A C A A G T T T T M7259237 i G T G T T C C C A C C T A C A A G T T T T
MT192759 A T C T A G G T T T C G T C C G G G T G T MT192755 A T C T A G G T T T T G T C C G G G T G T MT192775 A T C T A G G T T T C G T C C G G G G T G T MT192773 A T C T A G G T T T C G T C C G G G T G T MT19555 A T C T A G G T T T C G T C C G G G G T G T	MT258381 ; A A A T T T G A C A T C T T C A A T G G G MT258382 ; A A A T T T G A C A T C T T C A A T G G G MT258383 ; A A A T T T G A C A T C T T C A A T G G G MT259326 ; A A A T T T G A C A C C T T C A A T G G G MT259227 ; A A A T T T G A C A C C T T C A A T G G G	MT258381 A T T G T T C T T T T T T A C C C T C C A G MT258382 A T T G T T C T T T T T A C C C T C C A G MT258382 A T T G T T C T T T T T A C C C T C C A G MT259225 A T T G T T C T T T T T A C C C T C C A G MT259222 A T T G T T C T T T C T A C C C T C C A G	M1259238 C A G T T A G C A G G T G C A G G T G C A G G T G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G G T A G T A G T G T A G T A G T G T A G T A G T A G T A G T<	M7259238 G T G T T C C C A C C T A C A A G T T T M7259238 G T G T T C C C A C T T A C A A G T T T T M7259240 G T G T T C C C A C T T A C A A G T T T T M7259241 G T G T T C C C A C T T A C A A G T T T T M7259242 G T G T T C C C A C T T A C A A G T T T T
MT198652 A T C T A G G T T T C G T C C G G G T G T MT198653 A T C T A G G T T T T G T C C G G G T G T MT22661 A T C T A G G T T T T G T C C G G G G T G T MT232869 - MT32870 -	M7259228 : A A A T T T G A C A C C T T C A A T G G G M7259229 : A A A T T T G A C A C C T T C A A T G G G M7259229 : A A A T T T G A C A C C T T C A A T G G G M7259231 : A A A T T T G A C A C C T T C A A T G G G M7259231 : A A A T T T G A C A C C T T C A A T G G G	M7259228 A T T G T T C T T T C T A C C C T C C A G M7259229 A T T G T T C T T T C T A C C C T C C A G M7259230 A T T G T T C T T T C T A C C C T C C A G M7259231 A T T G T C T T T C T A C C C T C C A G M7259235 A T T G T T C T T T C T A C C C T C C A G	M7259243 I.C A G G T T A G C A G G T G T G T G G T G G T G G T G G T G G G T G G G T G G G G T G G G T G G G T G	MT259243 : G T G T T C C C A C T T A C A A G T T T T MT259244 : G T G T T C C C A C T T A C A A G T T T T MT259245 : G T G T T C C C A C C T A C A A G T T T T MT259245 : G T G T T C C C A C C T A C A A G T T T T MT259247 : G T G T T C C C A C C T A C A A G T T T T
MT2328271 - MT2328272 - MT2328272 - MT2332520 - A T C T A G G T T T C G T C C G G G T G T MT235520 - A T C T A G G T T T C G T C C G G G G T G T MT235520 - A T C T A G G T T T C G T C C G G G G G T G T	MT259225 A A A T T T G A C A C C T T C A A T G G G MT259237 A A A T T T G A C A C C T T C A A T G G G MT259238 N N N N N N N N N N N N N N N N N N	MT252233 A T T G T T C T T T C T A C C C T C C A G MT252233 A T T G T T C T T T C T A C C C T C C A G MT252233 A T T G T T C T T T C T A C C C T C C A G MT25233 A T T G T C T T T T T A C C C T C C A G MT25233 A T T G T T C T T T T T A C C C T C C A G	M1259249 : C A T G G T T T A G C C A G C G T G G T G M7259249 : C A T G G T T T A G C C A G C G T G G T G M7259250 : C A T G G T T T A G C C A G C G T G G T G M7259251 : C A T G G T T T A G C C A G C G T G G T G M7259252 : C A T G G T T T A G C C A G C G T G G T G	MI252247 G T G T T C C C A C T T A C A A G T T T T MI252249 G T G T T C C C A C T T A C A A G T T T T MI252249 G T G T T C C C A C T T A C A A G T T T T MI252525 G T G T T C C C A C T T A C A A G T T T T MI252525 G T G T T C C C A C C T A C A A G T T T T
MT233322 IATCTAGGTTTGGTCCGGGTGT MT233322 IATCTAGGTTCGGTCGGGGGGGGGGGGGGGGGGGGGGGGG	11239246 A A A T T T G A C A C C T T C A A T G G G 11239242 A A A T T T G A C A C C T T C A A T G G G 11239242 A A A T T T G A C A C C T T C A A T G G G 11259243 A A A T T T G A C A T C T T C A A T G G G 11259244 A A A T T T G A C A T C T T C A A T G G G	MT259244 A T T G T T C T T T C T A C C C T C C A G MT259244 A T T G T T C T T T T T A C C C T C C A G MT259243 A T T G T T C T T T T T A C C C T C C A G MT259244 A T T G T T C T T T T T A C C C T C C A G MT259245 A T T G T T C T T T T T A C C C T C C A G	M1259252, I, C A T G G T T T A G T C A G C G T G G T G M1259254, I, C A T G G T T T A G T C A G C G T G G T G M1259254, I, C A T G G T T T A G T C A G C G T G G T G M1259255, I C A T G G T T T A G T C A G C G T G G T G M1259252, I C A T G G T T T A G C C A G C G T G G T G M1259252, I C A T G G T T T A G C C A G C G T G G T G	MI252525 G T G T T C C C A C C T A C A A G T T T T MI252525 G T G T T C C C A C C T A C A A G T T T T MI252525 G T G T T C C C A C C T A C A A G T T T T MI252525 G T G T T C C C A C C T A C A A G T T T T MI252525 G T G T T C C C A C C T A C A A G T T T T
M7246495), A T C T A G G T T T T G T C C G G G T G T M7246451), A T C T A G G T T T T G T C C G G G T G T M7246452), A T C T A G G T T T C G T C C G G G G T G T M7246453), A T C T A G G T T T C G T C C G G G T G T M7246453), A T C T A G G T T T C G T C C G G G G T G T	MT259245 A A A T T T G A C A T C T T C A A T G G G MT259247 A A A T T T G A C A T C T T C A A T G G G MT259248 A A A T T T G A C A T C T T C A A T G G G MT259249 A A A T T T G A C A T C T T C A A T G G G MT259249 A A A T T T G A C A T C T T C A A T G G G	MT259246 A T T G T T C T T T T T T A C C C T C C A G MT259247 A T T G T T C T T T T T A C C C T C C A G MT259248 A T T G T T C T T T T T A C C C T C C A G MT259249 A T T G T T C T T T T T A C C C T C C A G MT259250 A T T G T T C T T T T T A C C C T C C A G	M1259229 : C A T G G T T T A G C C A G C G T G G T G M1259229 : C A T G G T T T A G T C A G C G T G G T G M1259259 : C A T G G T T T A G C C A G C G T G G T G M1259260 : C A T G G T T T A G C C A G C G T G G T G M1259252 : C A T G G T T T A G C C A G C G T G G T G M1259252 : C A T G G T T T A G C C A G C G T G G T G	MT252525 G T G T T C C C A C T T A C A A G T T T T MT252525 G T G T T C C C A C T T A C A A G T T T T MT252526 G T G T T C C C A C T T A C A A G T T T T MT252526 G T G T T C C C A C T T A C A A G T T T T MT252526 G T G T T C C C A C T T A C A A G T T T T
MT246455 I.A T C T A G G T T T C G T C C G G G T G T	M7259251 : A A A T T T G A C A T C T T C A A T G G G	M7259251 A TT G TT C TT TT T A C C C T C C A G	M7259263 LC A T G G T T T A G C C A G C G T G G T G	MT259263 G T G T T C C C A C T T A C A A G T T T T
A C A C G T A A C C C T G C T T G G A G A 17 802 17805 17815 17815 1782	GGCTCAGAATSTGACTATGTC 17913 1700 1705 1706 1703	TAACAGGACTCTTTAAAGATT 19 115 19 20 19 25 19 190 19 135	CTGTTTTtcagagcgcTTCCA 25 629 2000 2003 2004 25 649	
MT259231 A C A C G T A A C C C T G G T T G G A G A MT259235 A C A C G T A A C C C T G C T T G G A G A MT259235 A C A C G T A A C C T T G C T T G G A G A MT259237 A C A C G T A A C C T T G C T T G G A G A	MT259231 G G C T C A G A A T A T G A C T A T G T C MT259235 G G C T C A G A A T A T G A C T A T G T C MT259236 G G C T C A G A A T A T G A C T A T G T C MT259237 G G C T C A G A A T G T G A C T A T G T C	M7259231 T A A C A G G A C T C T T T A A A G A T T M7259235 T A A C A G G A C T C T T T A A A G A T T M7259236 T A A C A G G A C T C T T T T A A A G A T T M7259237 T A A C A G G A C T T T T T A A A G A T T	M7259231 Ċ T G T T T T T C À G A G C G C T T C Ċ À M7259235 C T G T T T T T C A T A G C G C T T C C A M7259236 C T G T T T T T C A G A G C G C T T C C A M7259237 C T G T T T T T C A G A G C G C T T C C A	MT259231 G T T T C C T G T T T A C C T T T T A C A MT259235 G T T T C C T G T T T A C C T T T T A C A MT259236 G T T T C C T G T T C A C C T T T T A C A MT259237 G T T T C C T G T T C A C C T T T T A C A
MT259238 A C A C G T A A C C T T G C T T G G A G A MT259239 A C A C G T A A C C C T G C T T G G A G A MT259240 A C A C G T A A C C C T G C T T G G A G A MT259241 A C A C G T A A C C C T G C T T G G A G A	MT259238 G G C T C A G A A T G T G A C T A T G T C MT259239 G G C T C A G A A T A T G A C T A T G T C MT259240 G G C T C A G A A T A T G A C T A T G T C MT259241 G G C T C A G A A T A T G T G A C T A T G T C	M7259228 I.T.A.A.C.A.G.G.A.C.T.T.T.T.T.A.A.A.G.A.T.T. M7259239 I.T.A.A.C.A.G.G.A.C.T.C.T.T.T.A.A.A.G.A.T.T. M7259240 I.T.A.A.C.A.G.G.A.C.T.C.T.T.T.A.A.A.G.A.T.T. M7259241 I.T.A.A.C.A.G.G.A.C.T.T.T.T.T.A.A.A.G.A.T.T. M7259241 I.T.A.C.A.G.G.A.C.T.T.T.T.T.T.A.A.A.G.A.T.T. M7259242 I.T.A.C.A.G.G.A.C.T.T.T.T.T.T.A.A.G.G.A.T.T. M7259242 I.T.A.C.G.G.G.A.C.T.T.T.T.T.A.A.G.G.A.T.T.	M7259238 C T G T T T T C A G A G C G C T T C C A M7259239 C T G T T T T T C A T A G C G C T T C C A M7259240 C T G T T T T T C A T A G C G C T T C C A M7259240 C T G T T T T C A G A G C G C T T C C A M7259242 C T G T T T T T N N N N N N N N T C C A	M7259238 G T T T C C T G T T C A C C T T T T A C A M7259239 G T T T C C T G T T T A C C T T T T A C A M7259240 G T T T C C T G T T T A C C T T T T A C A M7259241 G T T T C C T G T T C A C C T T T T A C A
M7259242 A C A C G T A A C C C T G C T T G G A G A M7259243 A C A C G T A A C C C T G C T T G G A G A M7259244 A C A C G T A A C C C T G C T T G G A G A M7259245 A C A C G T A A C C T T G C T T G G A G A M7259246 A C A C G T A A C C C T G C T T G G A G A	MT259242 : G G C T C A G A A T A T G A C T A T G T C MT259243 : G G C T C A G A A T A T G A C T A T G T C MT259244 : G G C T C A G A A T A T G A C T A T G T C MT259245 : G G C T C A G A A T A T G T G A C T A T G T C	M7259243 T A A C A G G A C T C T T T A A A G A T T M7259244 T A A C A G G A C T C T T T A A A G A T T M7259245 T A A C A G G A C T C T T T T A A A G A T T M7259245 T A A C A G G A C T C T T T T A A A G A T T	M7259243 : C T G T T T T T C A T A G C G C T T C C A M7259244 : C T G T T T T T C A T A G C G C T T C C A M7259245 : C T G T T T T T C A T A G C G C T T C C A M7259246 : C T G T T T T C C A T A G C G C T T C C A	MT259242 LG T T T C C T G T T T A C C T T T T A C A MT259243 LG T T T C C T G T T T A C C T T T A C A MT259244 LG T T T C C T G T T T A C C T T T T A C A MT259245 LG T T T C C T G T T C A C C T T T T A C A MT259246 LG T T C C T G T T C A C C T T T T A C A MT259246 LG T T C C T G T T C A C C T T T T A C A
M7259247 A C A C G T A A C C T T G C T T G G A G A M7259248 A C A C G T A A C C T T G C T T G G A G A M7259259 A C A C G T A A C C C T G C T T G G A G A M7259259 A C A C G T A A C C C T G C T T G G A G A M7259251 A C A C G T A A C C C T G C T T G G A G A	MT259247]; G G C T C A G A A T A T G A C T A T G T C MT259248]; G G C T C A G A A T A T G A C T A T G T C MT259249]; G G C T C A G A A T A T G A C T A T G T C MT259250]; G G C T C A G A A T A T G A C T A T G T C MT259251]; G G C T C A G A A T A T G A C T A T G T C	M7259247 T A A C A G G A C T T T T T A A A G A T T M7259248 T A A C A G G A C T C T T T T A A A G A T T M7259249 T A C A G G A C T C T T T A A A G A T T M7259250 T A A C A G G A C T C T T T A A A G A T T M7259251 T A C A G G A C T C T T T A A A G A T T	M7259247 C T G T T T T C A G A G C G C T T C C A M7259248 : C T G T T T T T C A T A G C G C T T C C A M7259249 : C T G T T T T T C A T A G C G C T T C C A M7259250 : C T G T T T T C A G A G C G C T T C C A M7259251 : C T G T T T T T C A T A G C G C T T C C A	MT259247 G T T T C C T G T T C A C C T T T T A C A MT259248 G T T T C C T G T T T A C C T T T T A C A MT259249 G T T T C C T G T T T A C C T T T T A C A MT259250 G T T T C C T G T T T A C C T T T T A C A MT259251 G T T T C C T G T T T A C C T T T T A C A
M7259252 A C A C G T A A C C C T G C T T G G A G A M7259253 A C A C G T A A C C T T G C T T G G A G A M7259254 A C A C G T A A C C T T G C T T G G A G A M7259255 A C A C G T A A C C T T G C T T G G A G A	M1259252 6 G C T C A G A A T A T G A C T A T G T C M7259252 6 G C T C A G A A T A T G A C T A T G T C M7259253 6 G C T C A G A A T G T G A C T A T G T C M7259254 6 G C T C A G A A T G T G A C T A T G T C	M7259252 T A A C A G G A C T C T T T A A A G A T T M7259252 T A A C A G G A C T T T T T A A A G A T T M7259254 T A A C A G G A C T T T T T A A A G A T T M7259255 T A A C A G G A C T T T T T A A A G A T T	MT259252 -C T G T T T T T C A G A G C G C T T C C A MT259253 -C T G T T T T T C A G A G C G C T T C C A MT259254 -C T G T T T T T C A G A G C G C T T C C A MT259255 -C T G T T T T C A G A G C G C T T C C A	MT259223 I,G T T T C C T G T T C A C C T T T T A C A MT259253 I,G T T T C C T G T T C A C C T T T T A C A MT259254 I,G T T T C C T G T T C A C C T T T T A C A MT259255 I,G T T T C C T G T T C A C C T T T T A C A
MT259256 : A C A C G T A A C C C T G C T T G G A G A MT259257 : A C A C G T A A C C T T G C T T G G A G A				
M1259258 A C A C G T A A C C C T G C T T G G A G A M1259258 A C A C G T A A C C C T G C T T G G A G A M1259259 A C A C G T A A C C C T G C T T G G A G A	MT259256 I.G. G. C. T. C. A. G. A. T. A. T. G. A. C. T. A. T. G. T. C. MT259256 I.G. G. C. T. C. A. G. A. A. T. A. T. G. A. C. T. A. T. G. T. C. MT259257 I.G. G. C. T. C. A. G. A. A. T. A. T. G. A. C. T. A. T. G. T. C. MT259258 I.G. G. C. T. C. A. G. A. A. T. A. T. G. A. C. T. A. T. G. T. C. MT259259 I.G. G. C. T. C. A. G. A. A. T. A. T. G. A. C. T. A. T. G. T. C. MT259259 I.G. G. C. T. C. A. G. A. A. T. A. T. G. A. C. T. A. T. G. T. C. MT259259 I.G. G. C. T. C. A. G. A. A. T. A. T. G. A. C. T. A. T. G. T. C. MT259259 I.G. G. C. T. C. A. G. A. A. T. A. T. G. A. C. T. A. T. G. T. C.	M7259226 T A A C A G G A C T C T T T A A A G A T T M7259257 T A A C A G G A C T C T T T T A A A G A T T M7259258 T A A C A G G A C T C T T T A A A G A T T M7259259 T A A C A G G A C T C T T T T A A G A T T M7259250 T A A C A G G A C T C T T T T A A G A T T	M7259256 , C T G T T T T C A T A G C G C T T C C A M7259257 ; C T G T T T T T C A G A G C G C T T C C A M7259258 ; C T G T T T T T C A T A G C G C T T C C A M7259259 ; C T G T T T T C A G A G C G C T T C C A M7259250 ; C T G T T T T T C A T A G C G C T T C C A	MT259256 G T T T C C T G T T T A C C T T T T A C A MT259257 G T T T C C T G T T C A C C T T T T A C A MT259258 G T T T C C T G T T T A C C T T T T A C A MT259259 G T T T C C T G T T C A C C T T T T A C A MT259259 G T T T C C T G T T C A C C T T T T A C A

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

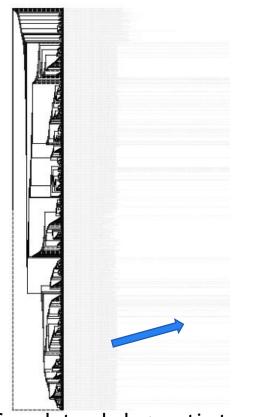


Case



vi) Phylogenetic relationship among all 928 coronavirus sequences

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



Case

Study 3

Complete phylogenetic tree of 928 SARS-Cov-2

MT163717 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA4-UW2/2020] complete genome /T246469 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW212/2020] complete genome [263463 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CT-UW329/2020| complete genome 263425 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW345/2020| complete genome 322409 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-0019/2020| complete genome [322407 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-0017/2020| complete genome 68 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW-1756/2020 ORF1ab polyprotein (OR 322396 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-0003/2020| complete genome MT246485 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW228/2020] complete genome MT326034 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW-1850/2020 ORF1ab polyprotein (OR MT188339 [Severe acute respiratory syndrome coronavirus 2 isolate USA/MN3-MDH3/2020] complete genome Severe acute respiratory syndrome coronavirus 2 isolate USA/MN1-MDH1/2020 complete genome Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/MI-SC2-0003/2020 ORF1ab polyprotein (OI evere acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW274/2020| complete genome Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-00284/2020 ORF1ab polyprotein (ORF1a I3203 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW-1426/2020 ORF1ab polyprotein (O) MT322404 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/V T322408 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-0018/2020] complete genome 263455 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW375/2020] complete genome 3396 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW315/2020| complete genome 322412 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/VA-DCLS-0024/2020| complete genome MT326040 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/OR-UW-1851/2020| complete genome /T263462 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW383/2020| complete genome MT293158 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW-1271/2020| complete genome Interpretation of the second s econd second sec evere acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW234/2020| complete genome MT259269 [Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW-1304/2020| complete genome Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USAWA-UW391/2020| complete genome

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

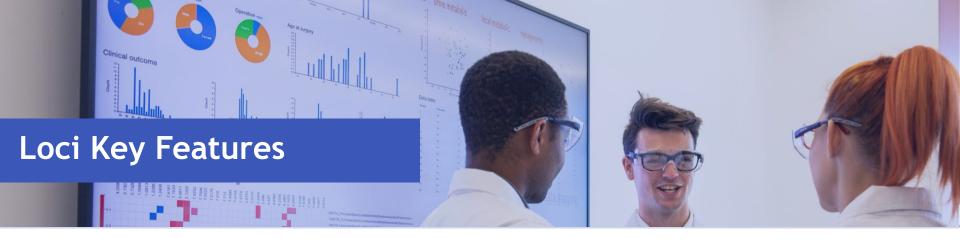
Case Study 3 Conclusions

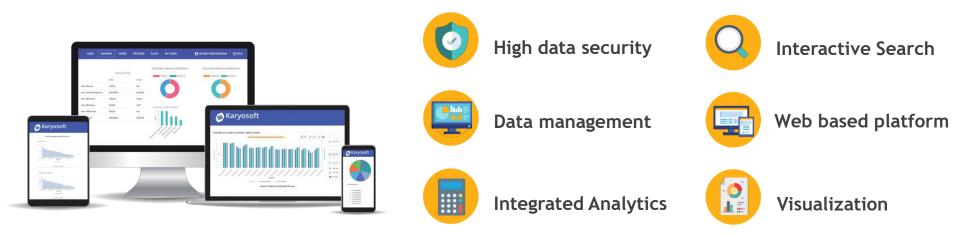
Goal



- 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.







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



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



