ISSN 2816-8089

RESEARCH ARTICLE

A Comparative Genomic Analysis of *Georgenia sps.* for Mining of LysR Transcriptional Regulator Sequences

Tejas Oza[1,2,3], Pooja Patel[1], Vrinda S Thaker[1,4*]

[1]*Plant Biotechnology and Genetic Engineering Laboratory, Department of Biosciences, Saurashtra University, Rajkot 360005, Gujarat, India.*
[2]*Department of Microbiology, Government Medical College, Bhavnagar 364001, Gujarat, India.*
[3]*Department of Microbiology, Marwadi University, Rajkot 360003, Gujarat, India.*
[4]*Vimal Research Society for Agrobiotech, Aji area, 80-feet Road, Rajkot 360002, Gujarat, India.*

Abstract

Georgenia is a genus belonging to the actinomycetes group. The genera comprise only 33 poorly characterized species with reference genomes of 10 distinct species. However, none of the species is well characterized by their genome characteristics. Our laboratory isolate from tomato plant leaf was identified and sequenced for identification and found to be *Georgenia sp.* Later genomic analysis revealed many functional genes having characteristic functions to be analyzed. The source of isolation raised the possibility of having functional genes to enhance senescence or having plant pathogenic activity by *Georgenia sp. SUBG003*. To explore *in silico* presence of these genes or gene pools genomic islands were identified and analyzed for our isolate and other 10 reference genomes of the Georgenia genus. Genomic islands were further explored for transcription regulators and finally, LysR transcriptional regulator sequences were extracted and a phylogeny among sequences was built from multiple sequence alignment.

Key Words: *Comparative genomics; Georgenia sp. SUBG003; Genomic islands; LysR transcriptional factors; Phylogeny; Pathogenicity*

***Corresponding Author**: Vrinda S. Thaker, Plant Biotechnology and Genetic Engineering Laboratory, Department of Biosciences, Saurashtra University, Rajkot 360 005, Gujarat, India; E-mail: thakervs@gmail.com*

***Received Date**: December 14, 2023, **Accepted Date**: January 03, 2024, **Published Date**: January 08, 2024*

***Citation:** Oza T, Patel P, Thaker VS. A Comparative Genomic Analysis of Georgenia sps. for Mining of LysR Transcriptional Regulator Sequences. Int J Bioinfor Intell Comput. 2024;3(1):01-12.*

1. Introduction

Bacteria and fungi are among the most frequent pathogens of plants [1]. Plants are considered to have the most complex metabolic and biochemical reactions to prevent pathogens and sustain differential outer environmental changes. Besides complexity, the plant systems are also susceptible to infections [2]. For example, *Erwinia amylovora* causing fire blight disease causing necrosis in apples, pears, and rosaceous plants was identified in 1880 for the first time [3]. Among bacteria the majority of the plant pathogens are gram-negative for example, *Pseudomonas syringae, Pseudomonas cichorii, Agrobacterium tumefaciens, Xanthomonas campestris, X. oryzae* [4].

Considering this fact, the underlying mechanisms and responsible genes are well characterized and extensively studied. For example, understanding plant pathogenicity mechanism to infect plants by bacteria is to invade the defence of plant delivery of effector protein by hypersensitive response pathogenicity (Hrp) Type III secretory system (T3SS) proteins of Group-2 genera of *Xanthomonas* species has been well studied [5].

Another example to consider *C. michiganensis* is a gram-positive bacterium that produces the virulence-related gene tomatinase (tomA), an antimicrobial saponin-degrading enzyme that invades tomato plants causing infection. However, on the other hand, many of the gram-positive bacteria have beneficial effects and are a part of natural biocontrol strategies. They are effective in producing phytohormones, antibiosis, mycoparasitism, endophytic colonization, Actinorhizal symbiosis, associative nitrogen fixation, and associative bioremediation [6].

The mechanisms are complex, and inculcated genes are dual-natured i.e. core and/or flexible components. Core components are essential for the survival of organisms and flexible are the ones that are active during adaptation to ecological and/or growth conditions. Among them, a few also remain as a part of their integral plasmids or phage-induced, transposons or insertion sequence elements [7].

To gain a focus on a specific character within multiple characters playing role in plant pathogenesis, the role of LysR transcriptional regulator is focused on due to its presence in the genome functions predicted. LysR transcriptional regulator is well defined in *Xanthomonas axonopodis* Pv. *glycines* wherein it is found to be involved in virulence, biofilm formation, swimming motility, siderophore secretion, and growth under lab conditions [8].

As per NCBI (National Centre for Biotechnology Information) taxonomy database *Georgenia* genus comprises about 33 established species names yet only 10 species have full-length quality genomes available on the genome database. The genus has a timeline of identifying species starting from 1998 to the present wherein a selective number of species are identified and reported for existence only. The phenotypic characteristics of *Georgenia* species are discussed around remediation biology and soil physiology roughly [9].

The current genomic era gives everyone an equal opportunity to explore interesting features of genomic content but correlating them to real-life or *In vivo* systems may require some time and effort. Still, it will be less as compared to studying an unknown organism for the first time [10].

Considering this flexibility into account and utilizing possible computational advancements the study aimed to identify unknown features of the genome of our laboratory isolate *Georgenia sp. SUBG003* and other genomes of *Georgenia* species, which are available in the resource database NCBI. *Georgenia* is among the poorly studied genera, it was accidental to find an actinomycete from the leaf of the tomato plant. Hence, we aimed to analyze the possible role of this organism and similar others within the same genus to find their possible role in plant pathogenicity and senescence.

2. Material and Methods

2.1. Isolation and identification of the organism

The organism was isolated from the leaves of tomato (*Solanum lycopersicum* L.) plant from the Botanical Garden, Saurashtra University, Rajkot, India. From the organism, DNA was isolated and identified using 16s rDNA sequence. The complete genome was sequenced and deposited to the NCBI database (GCA_000715355.1). The annotation data of the genome revealed a suite of genes responsible for pathogenicity [11].

2.2. Data retrieval of *Georgenia* species genomes

The Genome Database of NCBI was searched for available submissions for different species under the *Georgenia* genus. A total of 10 different genome datasets including our laboratory isolate *Georgenia sp. SUBG003* was retrieved and downloaded from the genome database in the format of '.fasta', '.gbff' file formats. These files were used as source material for further analysis [12,13].

2.3 Genome functional analysis

The genome comparison was done by uploading '.fasta' files into RAST auto pipeline for functional analysis. During the submission, a '.fasta' file was uploaded keeping all parameters default and selecting 11 generic codes under the archaea domain with the RASTtk v1.073 annotation scheme with auto fix errors option enabled. The submitted files were subjected to auto-pipeline analysis. After the analysis finished the genome functional analysis was retrieved in .csv file format. Necessary functional roles were analyzed and illustrated [14-16].

2.4 Pathogenicity/stress-related gene finding

The '.gbff' annotation filed from NCBI Genome database [17] retrieved was uploaded into Islandviewer 4.0 for finding genomic islands using *Georgenia Z294* sp genome as a reference and keeping the rest of the parameters as default to perform auto pipeline analysis designed in the tool. The resulting predicted islands, genes, predicted functions, and their sequences were retrieved and downloaded as '.csv', '.fasta' files for all integrated predictions made by IslandPick, SIGI-HMM, IslandPath-DIMOB, and Islander [18].

2.5 Phylogenetic analysis

The extracted DNA sequences from all 10 different genomes of *Georgenia* species for LysR Transcriptional regulators were compiled in a single '.fasta' file for phylogeny predictions (Figure 1). A single '.fasta' file was uploaded into NGPhylogen.fr one-click workflows for multiple sequence alignment using MAFFT v7.407_1, alignment curation using BMGE v1.12_1, Tree Inference using PhyML v3.3_1, and Tree rendering using Newick display v1.6 tools for construction of phylogeny. A Phylogram was visualized in PRESTO (Phylogenetic tree visualization tool) as a linear branch layout along with branch length displayed for each clad and branch within the tree [19].

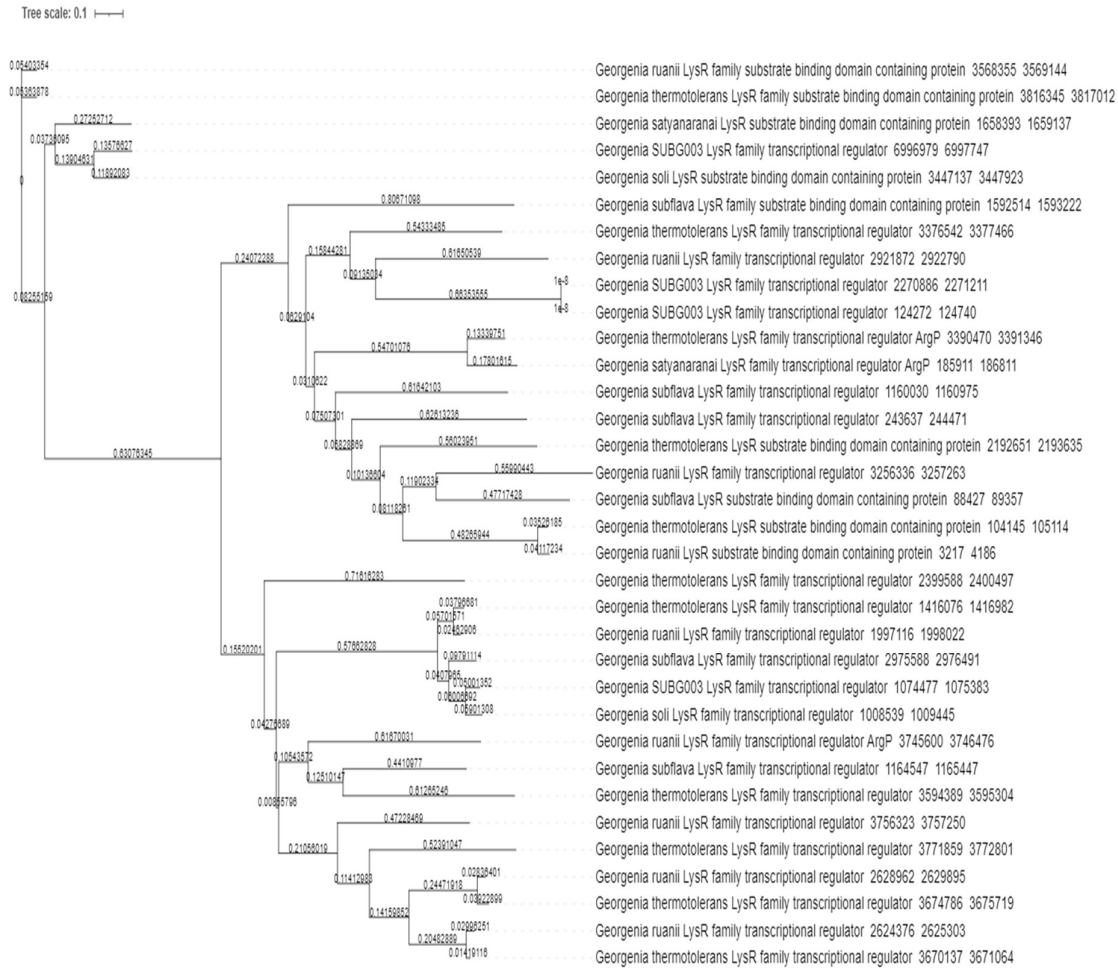


Figure 1: *Phylogenetic analysis of LysR Transcriptional Regulators from different Georgenia sps. (Sequence annotation includes name of species belonging to and location on the respective genome-Start and End positions).*

3. Results

3.1 Genome data and functional predictions

The taxon ID 154116 with genomes of different *Georgenia* species. named, *Georgenia yuyongxinii*, *Georgenia wutianyii*, *Georgenia faecalis*, *Georgenia satyanarayanai*, *Georgenia muralis*,

Georgenia soli, Georgenia subflava, Georgenia ruanii, Georgenia thermotolerans, Georgenia sp. SUBG003 and their respective assembly accession number GCA_006352065.1, GCA_006349365.1,GCA_003710105.1,GCA_009193135.1,GCA_003814705.1,GCA_002563695.1 ,GCF_009193155.1,GCF_009193175.1,GCF_009193185.1,GCF_000715355.1 were retrieved. These downloaded assembly files were analyzed through the RAST auto pipeline genome analyzer (data not shown).

The preliminary analysis for our laboratory isolates *Georgenia sp. SUBG003* suggested the presence of more than 900 figs with the class of functional involvement in Stress Response, Defence, and Virulence. These sets of genes or figs were broadly analyzed and found to have the presence of various heat shock proteins/cold shock proteins, ABC transporters, Antibiotic resistance, metal resistance, transcriptional regulators, DnaK operon genes, arsenate reductase, LysR transcriptional regulators, and many more roles (data not shown) indicating the involvement of this organism in pathogenesis and stress response.

3.2 Pathogenicity function prediction

The further analysis of the genome by Islandviewer 4.0 suggested the presence of several genomic islands in *Georgenia sp. SUBG003*-68, *Georgenia soli*-17, *Georgenia faecalis*-7, *Georgenia muralis*-27, *Georgenia ruanii*-30, *Georgenia satyanarayanai*-19, *Georgenia subflava*-21, *Georgenia thermotolerans*-39, *Georgenia wutianyii*-13, and *Georgenia yuyongxinii*-15 (Table 1). While analysing the individual sequences and predicted functions from genomic islands for pathogenicity role was found to be *Georgenia sp. SUBG003-354*, *Georgenia soli-357*, *Georgenia faecalis-97*, *Georgenia muralis-453*, *Georgenia ruanii*-300, *Georgenia satyanarayanai-269*, *Georgenia subflava*-430, *Georgenia thermotolerans-433*, *Georgenia wutianyii*-146, and *Georgenia yuyongxinii-205* among them 64, 151, 52, 227, 148, 111, 196, 174, 62, 121 are respective unique functions within genomic islands skipping hypothetical and duplicate names with distinctly located places within respective genomes (Table 2).

Comparative analysis for these predicted unique functions among different species of *Georgenia* demonstrated the 53, 59, 23, 111, 62, 31, 79, 68, 19, 60 number of unique sequences respectively without any overlap among or within the genomes of *Georgenia* species. The inter-species comparison has the varying number of common functions as shown in Table 3.

To highlight the distinction of *Georgenia sp. SUBG003* genome the number of sequences common to it are, vs *Georgenia soli* – 4, vs *Georgenia faecalis*-1, vs *Georgenia muralis*- 3, vs *Georgenia ruanii*- 4, vs *Georgenia satyanarayanai*- 2, vs *Georgenia subflava*- 5, vs *Georgenia thermotolerans*- 6, vs *Georgenia wutianyii*- 1, vs *Georgenia yoyongxii*- 4 only (Table 3).

Further from the genomic islands the number of genes for transcriptional regulators are found to be *Georgenia sp. SUBG003*-3, *Georgenia soli*-7, *Georgenia faecalis*-3, *Georgenia muralis*-10, *Georgenia ruanii*-9, *Georgenia satyanarayanai*-7, *Georgenia subflava*-9, *Georgenia thermotolerans*-11, *Georgenia wutianyii*-2, *Georgenia yoyongxii*-11 numbers (Table 4).

Table 1: *Island viewer predicted pathogenicity islands within genomes of 10 Georgenia species.*

Prediction Method	*Georgenia sp. SUBG003*	*Georgenia soli*	*Georgenia faecalis*	*Georgenia muralis*	*Georgenia ruanii*	*Georgenia satyanarayanai*	*Georgenia subflava*	*Georgenia thermotolerans*	*Georgenia wutianyii*	*Georgenia yoyongxii*
Integrated	68	17	7	27	30	19	21	39	13	15
InslandPath-DIMOB	4	6	3	17	3	7	8	5	5	6
SIGI-HMM	64	11	4	10	27	12	13	34	8	9

Table 2: *Island viewer predicted pathogenicity functions and sequences.*

Organism Name	Total Gene Sequences	Functional Sequences	Hypothetical	Unique
Georgenia sp. SUBG003	354	73	281	64
Georgenia soli	357	261	96	151
Georgenia faecalis	97	64	33	52
Georgenia muralis	453	362	91	227
Georgenia ruanii	300	226	74	148
Georgenia satyanarayanai	269	183	86	111
Georgenia subflava	430	314	116	196
Georgenia thermotolerans	433	272	161	174
Georgenia wutianyii	146	90	56	62
Georgenia yoyongxii	205	144	61	121

3.3 Phylogeny of LysR transcriptional regulator sequences

Considering the importance of transcriptional regulators in plant pathogenicity and induction of senescence, LysR transcriptional regulator sequences were retrieved from genomic islands and genes predicted by Island Viewer 4.0 within genomes of *Georgenia* species. The number of DNA sequences are *Georgenia sp. SUBG003*-4, *Georgenia soli*-2, *Georgenia ruanii*-9, *Georgenia satyanarayanai*-2, *Georgenia subflava*-6, *Georgenia thermotolerans*-11 while the remaining other species don't carry these specific sequences. The length of the sequences are also vary suggesting its expression regulation in the organism. (Table 5). Several studies have reported a correlation between gene length and biological timing for protein production [20-22]. According to these studies, the smaller the gene, the faster the protein synthesis process upon stimuli, and the production of the longer gene is often regulated by these smaller proteins and in turn, is expressed later in the response. Further connection of the gene length and expression for biological function is tried to decipher as (more evolutionarily conserved are

often associated with longer gene length and higher intronic burden [23-25]. On the other hand, a shorter gene length is correlated with smaller proteins, a high expression, and little intronic content [21]. Thus, it allows for regulatory mechanisms to be set up in the expression of important genes in response to biological stimuli [20].

Table 3: *Comparative analysis of pathogenicity functions within 10 Georgenia species.*

Organism Name	*Georgenia sp. SUBG003*	*Georgenia soli*	*Georgenia faecalis*	*Georgenia muralis*	*Georgenia ruanii*	*Georgenia satyanarayanai*	*Georgenia subflava*	*Georgenia thermotolerans*	*Georgenia wutianyii*	*Georgenia yoyongxii*	Unique
Georgenia sp. SUBG003	-	-	-	-	-	-	-	-	-	-	53
Georgenia soli	4	-	-	-	-	-	-	-	-	-	59
Georgenia faecalis	1	16	-	-	-	-	-	-	-	-	23
Georgenia muralis	3	29	15	-	-	-	-	-	-	-	111
Georgenia ruanii	4	32	11	38	-	-	-	-	-	-	62
Georgenia satyanarayanai	2	20	11	40	28	-	-	-	-	-	31
Georgenia subflava	5	38	13	58	28	32	-	-	-	-	79
Georgenia thermotolerans	6	42	17	46	31	38	43	-	-	-	68
Georgenia wutianyii	1	8	7	16	18	20	21	14	-	-	19
Georgenia yoyongxii	4	27	10	22	29	21	26	28	7	-	60

Table 4: *Number of sequences of transcriptional regulator genes.*

Organism Name	Transcriptional regulator
Georgenia sp. SUBG003	3
Georgenia soli	7
Georgenia faecalis	3
Georgenia muralis	10
Georgenia ruanii	9
Georgenia satyanarayanai	7
Georgenia subflava	9
Georgenia thermotolerans	11
Georgenia wutianyii	2
Georgenia yoyongxii	11

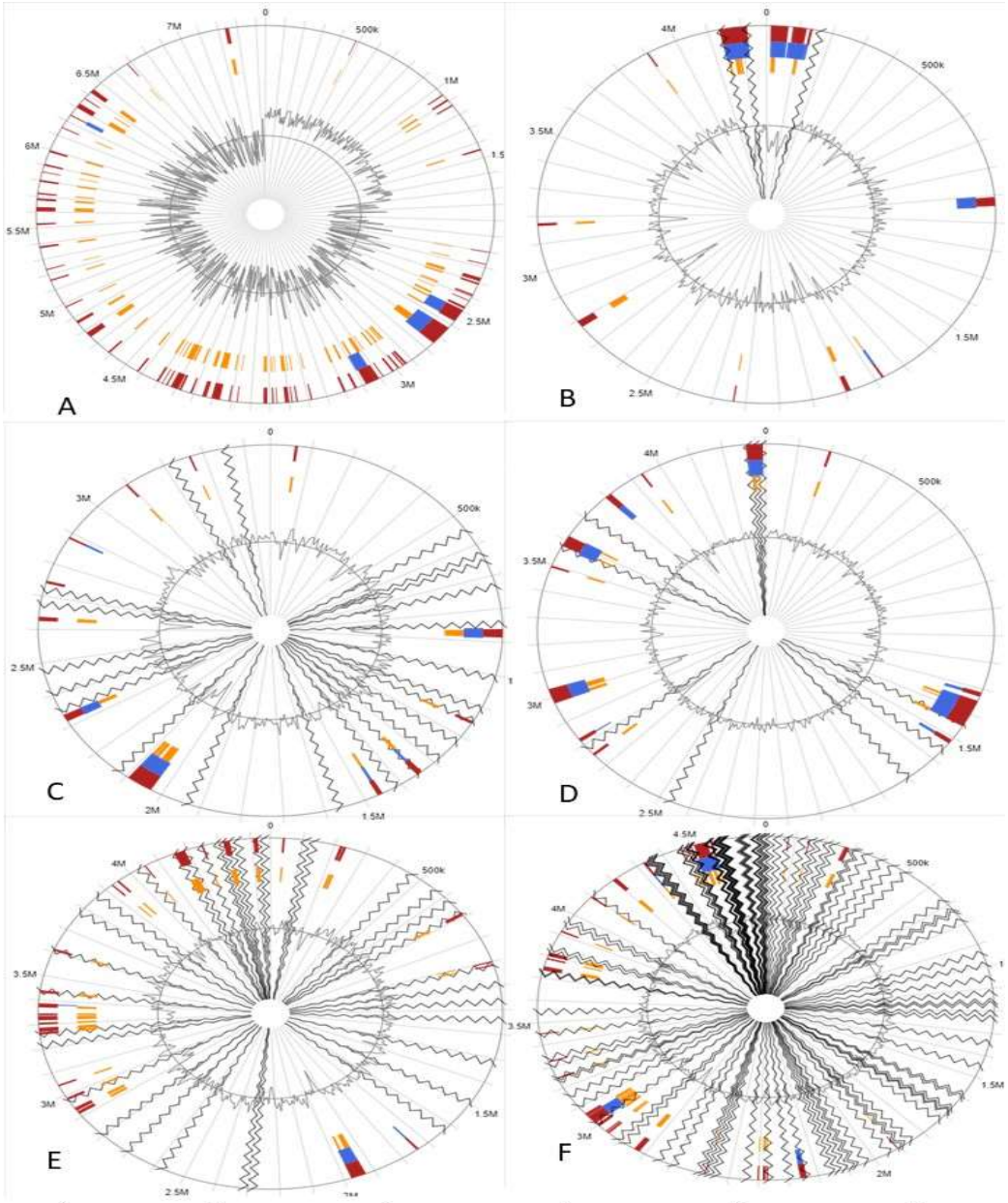
Table 5: *LysR transcriptional regulator sequence presence in Georgenia species.*

Organism Name	Gene Pool of LysR Transcriptional Regulator	
	No. of Sequences Extracted	Length
Georgenia sp. SUBG003	4	468, 906, 325, 768
Georgenia soli	2	786, 906
Georgenia faecalis	0	0
Georgenia muralis	0	0
Georgenia ruanii	9	969,906,927,933,927,918, 789, 876
Georgenia satyanarayanai	2	900, 744
Georgenia subflava	6	930, 834, 945, 900, 708, 903
Georgenia thermotolerans	11	667, 942, 933, 927, 915, 876, 924, 909, 984, 906, 969
Georgenia wutianyii	0	0
Georgenia yoyongxii	0	0

The phylogeny of LysR transcriptional regulators within *Georgenia sp. SUBG003*, *Georgenia soli*, *Georgenia ruanii*, *Georgenia satyanarayanai*, *Georgenia subflava*, *Georgenia thermotolerans* has displayed the identical sequences of *Georgenia sp. SUBG003* wherein 2 of the sequences are distinctly placed having similarity with *Georgenia ruanii* sharing a common origin. The other 2-sequences share a common origin with *Georgenia soli* which is the most similar genome of *Georgenia sp. SUBG003*. A few sequences are distinctly different among any of the sequences of the species are *Georgenia thermotolerans, Georgenia satyanaranai and Georgenia ruanii*.

4. Discussion

The *Georgenia* genus species belonging to the Actinomycetes group has inbuilt characteristics of surviving extreme environments considering their isolation sources [26]. The included species of *Georgenia* are representative to compare with our laboratory isolate *Georgenia sp. SUBG003* is also a reference genome species considered by the NCBI genome database. The isolation source of our isolation is a tomato plant leaf which is infected and showed the appearance of senescence. Further exploration of the genome using NGS analysis explored entire genome characteristics. Firstly, the annotation and functional analysis by RAST displayed the characteristic gene functions of metal resistance, antibiotic resistance, virulence, and stress response and resistance (Data not shown). These characteristics are briefly summarized to gain a specific focus on a set of genes and sequences. Further to confirm these characteristics specific approaches to finding Genomic islands representing pathogenicity functions are predicted using Island Viewer 4.0 (Figure 2). The filtration of information was done at multiple stages to get the most accurate results. Hence, only specific genomic islands with specific functional genes are filtered omitting hypothetical sequences. Further, to focus on the evolutionary aspect within species, one of the characteristic functional transcriptional regulator genes was assessed for its presence among all selected species. For the selected functional fragment to be analyzed for LysR transcriptional regulator sequences were identified and separated from genomic islands. Out of 10 selected species, only 6 of the species have got presence of LysR transcriptional regulator sequences with varying numbers and length.



A	B	C	D	E	F
KEP24129.1	WP_098484503.1	WP_193311309.1	WP_226920813.1	WP_152232420.1	WP_152203986.1
KEP23392.1	WP_098482715.1	WP_152189911.1	WP_152194106.1	WP_152229884.1	WP_152202193.1
KEP22414.1			WP_152193390.1	WP_152232257.1	WP_152203017.1
KEP22935.1			WP_152193385.1	WP_193314517.1	WP_152203097.1
			WP_152196532.1	WP_193314497.1	WP_152202282.1
			WP_152194868.1	WP_152233229.1	WP_152202273.1
				WP_152231676.1	WP_152200332.1
				WP_152231514.1	WP_152200204.1
				WP_152231510.1	WP_211489673.1
					WP_170297201.1
					WP_152204650.1

Figure 2: *Genomic islands predicted with various methods, red color-integrated methods, orange color-SIGI-HMM, blue color-IslandPath-DIMOB. A. Georgenia SUBG003, B. Georgenia soli, C. Georgenia satyanarayanai, D. Georgenia subflava, E.Georgenia ruanii, F.Georgenia thermotolerans. Along with their respective NCBI sequence IDs of LysR transcriptional regulator sequences used for phylogeny.*

Hence to get an idea about the inter-species evolutionary aspect of this specific function phylogeny was assessed. Phylogeny demonstrated the distinct sequences of *Georgenia sp. SUBG003* as compared to its closely related species within the genus *Georgenia soli*. The separate clades of *Georgenia thermotolerans, Georgenia satyanaranai, and Georgenia ruanii* may also indicate the possibility of a new function evolving. The presence of LysR transcriptional regulator and other similar functional sequences indicated the role of *Georgenia sp. SUBG003* and other studied species are to be active for their association with plants to perform various functionalities. These functions are to be confirmed using a wet lab but considering the isolation source on plant leaves it can be assessed that there is a strong possibility of *Georgenia sp. SUBG003* is a plant pathogen that performs senescence of leaves. Real-time studies like RT-PCR, Microarray, and Proteomic analysis may reveal the exact picture of the said function.

5. Conclusions

LysR transcriptional regulator is among the factors involved in the plant pathogenicity in gram-positive bacteria and in the senescence of leaves, while our isolate is of leaf lesion origin hence the phenotypic and genotypic character matches. The *Georgenia sp. SUBG003* has the presence of LysR transcriptional regulator sequences; however, literature on this is rather scanty. The comparison of 10 Georgenia genomes showed the presence of the same character in *Goergenia soli, Georgenia thermotolerans, Georgenia satyanarayanai, Georgenia subflava, and Goergenia ruanii*. The variations among sequences demonstrated the similarity of sequences with *Goergenia soli* and *Georgenia subflava* which may have similar activity or role. This knowledge is in addition to the profile of *Georgenia* species and genus.

Acknowledgments

The authors are thankful to the Department of Biosciences, Saurashtra University, Centre for Advanced Studies in Plant Biotechnology & Genetic Engineering, Department of Higher Education, Gujarat State Government and Financial Assistance Project (FAP-2010), GSBTM for providing the required facility.

Funding

Gujarat State Government, Financial Assistance Project (FAP-2010), GSBTM, Gandhinagar, Gujarat, India.

References

1. Nazarov PA, Baleev DN, Ivanova MI, et al. Infectious plant diseases: etiology, current status, problems and prospects in plant protection. Acta Naturae. 2020;12:46-59.

2. Kaur S, Samota MK, Choudhary M, et al. How do plants defend themselves against pathogens-biochemical mechanisms and genetic interventions. Physiol Mol Biol Plants. 2022;28:485-504.

3. Vanneste JL. Fire Blight: The Disease and its Causative Agent, *Erwinia amylovora*. CABI Publishing, New York, USA. 2000.

4. Sharma A, Gupta AK, Devi B. Current trends in management of bacterial pathogens infecting plants. Antonie Van Leeuwenhoek. 2023;116:303-26.

5. Meline V, Delage W, Brin C, et al. Role of the acquisition of a type 3 secretion system in the emergence of novel pathogenic strains of Xanthomonas. Mol Plant Path. 2019;20:33-50.

6. Francis I, Holsters M, Vereecke D. The gram-positive side of plant–microbe interactions. Env Microbiol. 2010;12:1-12.

7. Srivastava A, Al-Karablieh N, Khandekar S, et al. Genomic distribution and divergence of levansucrase-coding genes in *Pseudomonas syringae*. Genes (Basel). 2012;3:115-37.

8. Park H, Do E, Kim M, et al. A LysR-type transcriptional regulator LcrX is involved in virulence, biofilm formation, swimming motility, siderophore secretion, and growth in sugar sources in *Xanthomonas axonopodis* pv. glycines. Front Plant Sci. 2020;10:1657.

9. Chauhan J, Gohel S. Exploring plant growth-promoting, biocatalytic, and antimicrobial potential of salt tolerant rhizospheric *Georgenia soli* strain TSm39 for sustainable agriculture. Braz J Microbiol. 2022;53:1817-28.

10. Vinatzer BA, Yan S. Mining the genomes of plant pathogenic bacteria: how not to drown in gigabases of sequence. Mol Plant Pathol. 2008;9:105-18.

11. Patel, PP, Rakhashiya PM, Thaker VS. Genomic analysis of novel phytopathogenic *Georgenia sp.* strain SUB25. Genom Data. 2015;5:320-2.

12. Sayers EW, Beck J, Bolton EE, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2021;49:D10-7.

13. O'Leary NA, Wright MW, Brister JR, et al. Reference Sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44:D733-45.

14. Brettin T, Davis JJ, Disz T, et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 2015;5:8365.

15. Overbeek R, Olson R, Pusch GD, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42:206-14.

16. Aziz RK, Bartels D, Best AA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genom. 2008;9:75.

17. Tatusova T, DiCuccio M, Badretdin A, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44:6614-24.

18. Claire B, Matthew RL, Kelly PW, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 2017;45:W30-5.

19. Lemoine F, Correia D, Lefort V, et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. Nucleic Acids Res. 2019;47:W260-5.

20. Kirkconnell KS, Magnuson B, Paulsen MT, et al. Gene length as a biological timer to establish temporal transcriptional regulation. Cell Cycle. 2017;16:259-70.

21. Urrutia AO, Hurst LD. The signature of selection mediated by expression on human genes. Genome Res. 2013;13:2260-4.

22. Grishkevich V, Yanai I. Gene length and expression level shape genomic novelties. Genome Res. 2014;24:1497-503.

23. Wolf YI, Novichkov PS, Karev GP. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. Proc Natl Acad Sci. 2009;106:7273-80.

24. Vishnoi A, Kryazhimskiy S, Bazykin GA, et al. Young proteins experience more variable selection pressures than old proteins. Genome Res. 2010;20:1574-81.

25. Gorlova O, Fedorov A, Logothetis C, et al. Genes with a large intronic burden show greater evolutionary conservation on the protein level. BMC Evol Biol. 2014;14:50.

26. Nouioui I, Carro L, García-López M, et al. Genome-based taxonomic classification of the phylum *Actinobacteria*. Front Microbiol. 2018;9:2007.