Biological Forum – An International Journal **15(6): 01-16(2023)**

ISSN No. (Print): 0975-1130
ISSN No. (Online): 2249-3239

Comparative Genomics Analyses of some selected *Pseudomonas* strains having Biocontrol, Plant Growth Promoting and Bioremediation activities using Bioinformatic Tools

Mehzabin Rehman[1*], Sabnoor Yeasrin Jyoti[2], Preetom Regon[3], Subrat Kakati[4], Piyush Kumar Mishra[5] and Bhaben Tanti[3]

[1]Department of Botany, University of Science & Technology, Meghalaya (Meghalaya), India.
[2]Department of Botany, PDUAM, Amjonga (Assam), India.
[3]Department of Botany, Gauhati University (Assam), India·
[4]Department of Botany, Nabajyoti College, Kalgachia, Barpeta (Assam), India.
[5]Department of Botany, B.N. College, Dhubri (Assam), India.

(Corresponding author: Mehzabin Rehman)*
(Received: 05 April 2023; Revised: 26 April 2023; Accepted: 20 May 2023; Published: 05 June 2023)
(Published by Research Trend)

ABSTRACT: *Pseudomonas* that are associated with plants, often found living as parasites or saprophytes on the surfaces or inside plant species. Such species of *Pseudomonas* associated with plants may promote growth of plants by eliminating pathogenic microbes thereby synthesizing plant growth stimulating hormones and enhancing disease resistance in plants, biological control of plant pathogens and bioremediation. The present investigation was conducted with an aim to study comparative genomic studies of 14 *Pseudomonas* strains having biocontrol, PGPR and bioremediation activities keeping *P.* fluorescens as the reference strain. The study revealed that these strains are somewhat nearly related strains based on the various parameters undertaken, and therefore can be used collectively. With the increasing availability of sequences, the complexity of genome alignment and analysis is growing drastically with which the computational requirements of the EDGAR 2.0 and Mauve 2.3.1 have risen considerably over the past decade which supports an easy, user-friendly interface of evolutionary relationships in terms of gene order thereby gaining new biological insights of differential gene content.

Keywords: *Pseudomonas*, biocontrol, PGPR, bioremediation, genomics, EDGAR 2.0, MAUVE 2.3.1.

INTRODUCTION

Pseudomonas, a rod-shaped, gram-negative, non-spore forming aerobic gamma proteobacteria, belonging to the family Pseudomonadaceae having one or more flagellum containing 313 members (Moore *et al.*, 2021; Qin *et al.*, 2022). They are also known as oxidase positive and catalase positive bacteria. The members show a vast range of metabolic diversity and therefore consequently able to colonize huge range of niches (Scales *et al.*, 2014; Kharte *et al.*, 2022). Due of its vastness in plants seeds and water, the *Pseudomonads* were observed earlier in the history of microbiology. They are known to show antimicrobial, biocontrol, plant growth promoting (PGP) and bioremediating activities. Since the mid-1980s, specific members have been treated with grain seeds or directly applied to soils with an aim to prevent the establishment of crop pathogens. This practice referred to as biocontrol includes all aspects of utilizing microbes or their by-products in controlling pests thereby preventing plant diseases (Bajpai *et al.*, 2021; Thakur *et al.*, 2022). Meanwhile, it also imparts methods that are compatible in order promote to sustainable agriculture. Understanding the underlying mechanisms of biocontrol through various host-pathogen interactions may help to select and create more effective biological control agents that can manipulate the environment of the soil to create a condition for successful biocontrol (Moore *et al.*, 2021; Sah *et al.*, 2021; Kumar and Pareek 2022).

Plant growth promoting rhizobacteria (PGPR) improves plant growth through addition of nitrogen and phosphorous, phytohormones like auxin (AAI), indoleacetic acid (IAA) to the soil or by decreasing the level of ethylene under stressed conditions (Backer *et al.*, 2018). Bacteria associated with plants can be beneficial, deleterious or neutral based onits effects on plant growth. Some of these beneficial free-living bacteria colonize roots of plant and promote plant growth (Koul *et al.*, 2019; Kaur *et al.*, 2022). It affects growth of plant in two different ways i.e. both indirectly or directly. The direct effect includes supplying plants with microbial phytohormones, to facilitate the uptake of specific compounds from the environment (Saeed *et al.*, 2021). On the other hand, the indirect method involves lessening or preventing the ill effects of one or more phytopathogenic organisms. This can occur by synthesizing combative compounds or by inducting tolerance to pathogens (Tsukanova *et al.*, 2017; Backer *et al.*, 2018).

Some members of *Pseudomonas* metabolize toxic pollutants in the environment as such can be used for bioremediation. Bioremediation may occur independently (intrinsic bioremediation or natural attenuation) or occur collectively through the addition

of fertilizers, oxygen, etc., that help in improving the growth of the pollution-consuming microbes within the medium known as biostimulation (Choudhury and Bordolui 2022; Kumar and Pareek 2022). Depleted soil nitrogen level may stimulate biodegradation of some nitrogenous organic compounds and soil materials with high potential to absorb pollutants may slow down the process of biodegradation owing to limited bioavailability of the compounds to microbes. Recent advances have showed successful incorporation of matched microbial strains to the medium to increase the resident microbial population's potentiality to catabolize contaminants. Thus, bioremediation can be termed as a waste management procedure that ensure the use of microorganisms to neutralize pollutants from a contaminated site (Ojuederie and Babalola 2017; Raklami *et al.,* 2022; Bhargavanandha *et al.,* 2021). Such microbes that have vital roles in the process of bioremediation are known as bioremediators.

In comparative genomics studies, genome sequences of various species are compared. Through this researchers can determine these species at their molecular level thereby distinguishing various life forms from each other (Das *et al.,* 2021). It also provides a powerful interface for analyzing evolutionary changes among organisms thereby paving the way to identify genes that are conserved among species, also genes that provide organisms with its unique features (Dieckmann *et al.,* 2021; Jayachandran *et al.,* 2022).

Research has been carried out so far into the insights of PGPR, biocontrol and bioremediation which provided a thorough understanding of the multiple aspects of disease suppression. Yet, most of the focus has been on free-living rhizobacterial strains, especially *Pseudomonas*. However, no previous works on Pseudomonas comparative genomics have been carried out sp far. Therefore, keeping in view the above facts, the present study was carried out with an aim to study comparative genomics of different strains of *Pseudomonas*, keeping *Pseudomonas fluorescens* as the reference strain in terms of biocontrol, plant growth promoting and bioremediation aspects. *P. fluorescens* belonging to PGPR, have vital roles in enhanced plant growth, induced systemic resistance and biocontrol. *P. fluorescens* grows rapidly *in vitro*, has the ability to rapidly utilize seeds and roots exudates, can colonize and multiply itself in the rhizosphere and also in the interior of the plants, has a wide range of bioactive metabolites, can compete assertively with other microbes, inexpensive and can adapt to various environmental conditions.

MATERIALS AND METHODS

14 *Pseudomonas* strains were considered for the present investigation. The names of the strains, their source of isolation and functions are highlighted in Table 1.

Table 1: Pseudomonas strains considered during study.

Sr. No.	Name of the strains	Isolated from	Function	References
1.	*Pseudomonas_aeruginosa*_M18	Rhizosphere of sweet melon	Effective against Mycosphaerellamel on is mycelium	(Zhang *et al.,* 2020)
2.	*Pseudomonas_aeruginosa*_PAO1	Burn wound in Melbourne, Australia especially in patients with cystic fibrosis	Effective against Caenorhabditis elegans	(Chandler *et al.,* 2019)
3.	*Pseudomonas_brassicacearum*_subsp_bra ssicacearum_NFM421	Arabidopsis thaliana	Growth and Mn tolerance of the Mn-stressed plants	(Franzino *et al.,* 2021)
4.	*Pseudomonas_denitrificans*_ATCC_1386 7	Soil after enrichment with succinate-nitrate medium	Used as a nitrogen fixing strain	(Ainala *et al.,* 2013)
5.	*Pseudomonas_ fluorescens*_F113 (reference strain)	Sugar-beet rhizophere	Effective against Pythium ultimum, Phytophthora cactorum, Fusarium oxysporum	(Patel *et al.,* 2013)
6.	*Pseudomonas_protegens*_CHA0	Roots of tobacco in Swiss soil	Curb plant diseases and to partly replace synthetic chemical pesticides that are harmful to humans	(Flury *et al.,* 2019)
7.	*Pseudomonas_protegens*_Pf_5	Mushroom tissue	Nitrogen fixing strain	(Henkels *et al.,* 2014)
8.	*Pseudomonas_putida*_BIRD_1	Rhizosphere soil	Plant growth promoting rhizobacteria	(Roca *et al.,* 2013)

9.	*Pseudomonas_putida*_DOT_T1E	A wastewater treatment plant	Toluene degradation strain	(Weimer *et al.*, 2020)
10.	*Pseudomonas_putida*_F1	Polluted soil	Toluene degradation strain	(Dangi *et al.*, 2021)
11.	*Pseudomonas_putida*_GB_1	Soil and water	Manganese oxidizer	(Zheng *et al.*, 2018)
12.	*Pseudomonas_putida*_S16	Soil samples obtained from a field under continuous tobacco cropping in Shandong, China	Nicotine degrading strain	(Maity *et al.* (2023)
13.	*Pseudomonas_putida*_W619	*Populus trichocarpa* x deltoides cv. "Hoogvorst"	Endophyte of poplar	(Wu *et al.*, 2011)
14.	*Pseudomonas_stutzeri*_A1501	Soilborne	Nitrogen-fixing bacterium	(Sah *et al.*, 2021)

The study of comparative genomics of all the above mentioned strains were analyzed using two most important bioinformatics tools – namely, Edgar 2.0 and Mauve 2.3.1.

EDGAR 2.0. The advent of Next Generation Sequencing methods led to the rapid expansion of complete sequenced genomes which made it is easier to evaluate large datasets in a comparative approach. This includes identification and classification of orthologous genes in different genomes as core genes and singletons. To conduct these analyses a software was developed known as Efficient Database Framework for comparative Genome Analyses using BLAST scores Ratios (EDGAR). It enables to perform comparative analysis of genomes in a high-throughput manner. 582 genomes across 75 genus groups from the NCBI database were comparatively analyzed and the outputs were integrated into an underlying database (Blom *et al.*, 2016; Dieckmann *et al.*, 2021).

The biocontrol/plant growth promoting bacteria and bioremediation strains of *Pseudomonas* were studied. The strains were then searched on NCBI. Their accession numbers were noted down. Using the website - https://edgar.computational.bio, Edgar application was run. Then, comparing the accession numbers, we select our required strains for analyses using one as a reference strain. Out of the 14 *Pseudomonas* strains taken for study, 9 were biocontrol and plant growth promoting bacterial strains and 5 were bioremediation agents. The analyses were carried out keeping *Pseudomonas fluorescens* F113, as the reference strain.

The following parameters were considered during the study period. All the data were calculated keeping *Pseudomonas fluorescens_F113* as the reference strain.

Genomic subsets. It includes core and pan genomes; and singletons. The analysis requires selection of a reference strain to comparatively study the set of genomes undertaken for evaluation. The output is presented in a tabular form (Li *et al*., 2018).

Core genome. The core genome is the set of homologous genes shared by all the strains of the same bacterial species. Most of these genes are associated in vital roles for the survival of bacteria (Park *et al.*, 2019).

Pan genome. Pan genome also known as supra genomere presents full complement of genes in a clade especially in bacteria and archaea that have larger variation of gene content among its closely associated strains (Inglin *et al.*, 2018).

Singletons. A singleton is a read with a sequence that is present exactly once, i.e. is unique among the reads. They are however removed at later stages to reduce sequencing errors (Cubry *et al.*, 2017).

Geneset. It is the calculation of all the genes found in the genome of all the 14 biocontrol Pseudomonas strains taken for analysis. A table depicting all genomes present in EDGAR is seen on the upper part of the feature. It also contains a set of options such as "INCLUDE" and "EXCLUDE" for each genome. The geneset is evaluated in such a way that there has to be a set of orthologous genes in "INCLUDE" and no such genes "EXCLUDE" genomes. Genomes not belonging to either of these categories are ignored. (Cubry *et al.*, 2017).

Venn diagrams. It shows all possible combination of the number of genes of the selected genomes. It allowseasy visual interpretation of genome size of the core genes number of genes in each subset of the dispensable genome. EDGAR creates Venn diagrams with an upper limit of 5 genomes for an informative graphical representation (Blom *et al.*, 2009; Blom *et al.*, 2016).

Set size statistics

Core and singleton development plot. To prepare a core and singleton development plot of genome size for increasing genome numbers a curve fitting approach is made with an exponential decay function (Cubry *et al.*, 2017; Park *et al.*, 2019).

Pan development plot. In order to create a pan development plot, genome sizes can be evaluated through Heaps' law function. It an empirical law used in linguistics to describe the number of distinct words in a single or set of documents as a function of the length of the document. When a large text is analyzed, the different number increases according to a sub-linear power law of the total number of scanned words (Inglin *et al.*, 2018).

Synteny plots. Synteny refers to the physical co-localization of gene locus on the same chromosome within a species or individual. However, during the present day, a term referred to as shared synteny allows researchers to compare conservation of blocks of order within two sets of chromosomes. In order to monitor the conservation of gene order among the *Pseudomonas* chromosomes, pairwise synteny plots were constructed in which the position of each CDS of the chromosome (X-axis) is plotted against the position of its homologous chromosome(Y-axis). Similar chromosomes shows a diagonal plot. In our study, synteny plots were analysed using one strain each time with the reference strain (Blom *et al.,* 2016).

Phylogeny

Create AAI/ANI matrix. While the computation of a phylogenetic tree based on the complete core genome shows good results, it is still a computationally intensive task. Two different approaches toward a phylogenetic evaluation based on the increasing availability of whole-genome sequences were proposed by Konstantinidis and Tiedje (2005), i.e. the average amino acid identity (AAI) and the average nucleotide identity (ANI). For the AAI method, the average AAIs of all conserved genes in the core genome as computed by the BLAST algorithm are collected. The results can be easily extracted from the EDGAR database. For both methods, the resulting phylogenetic distance values are arranged in an AAI/ANI matrix, clustered according to their distance patterns and visualized as heatmaps. The blast hits between the orthologous genes of the core of the selected genome will be analyses for their mean/median percent identity values. For clustering, the centroids were used and for distance measure the euclidean distance between the contigs was calculated (Blom *et al.,* 2016; Hugenholtz *et al.,* 2021).

Phylogenetic tree. For comparison of different genomes, a phylogenetic tree was constructed using a slightly modified version proposed by Zdobnov and Bork (2007). The core genome is calculated as described above. In the next step multiple alignments for all core genes are created using MUSCLE. Non matching parts of the alignments are masked using GBLOCKS and then removed. The matching parts are concatenated to one big multiple alignment of more than 1 Mb length. Finally, a phylogenetic tree is generated from this long alignment using PHYLIP (Blom *et al.,* 2016; Young and Gillung 2019; Zhao *et al.,* 2021).

Genome to genome distance. The genome to genome distance calculation is based on Stefan Auchs' "Genome BLAST Distance Phylogeny", which is used to calculate the dissimilarity of given genomes by calculating high-scoring segment pairs. The current web interface serves as a user friendly wrapper around the original tool. By default, the genomic distance calculation is achieved by leveraging a formula recommended by Meier-Kolthoff (Blom *et al.,* 2016).

Mauve 2.3.1 - Genome Alignment Visualization. Mauve constructs multiple genome alignments in presence of large-scale evolutionary events such as rearrangement and inversion. It uses algorithmic techniques that scale well in the lengths of sequences being aligned. Because recombination can cause genome rearrangements, orthologous regions of one genome may be reordered or inverted relative to another genome. During the alignment process, Mauve identifies conserved segments that appear to be internally free from genome rearrangements. Such regions are referred to as Locally Collinear Blocks (LCBs) (Marcais *et al.,* 2018).

The Mauve 2.3.1 software was installed and all the 14 *Pseudomonas* strains (constituting all the biocontrol, PGPR and bioremediation strains including the reference strain) that were previously used in Edgar 2.0 were now aligned keeping *Pseudomonas_fluorescens*_F113 each time.

RESULTS AND DISCUSSION

EDGAR 2.0

Genomic subsets. These were calculated for all Pseudomonas strain considered for the study keeping *Pseudomonas fluorescens*_F113 used as a reference strain.

Core genome. The core genome consists of 1928 CDS (Cell Data Set of an Open Reading Frame).

Pan genome. The pan genome consists of 16942 CDS.

Singletons. There are 372 singletons for the reference strain *Pseudomonas_fluorescens*_F113 vs. all other strains taken for analyses.

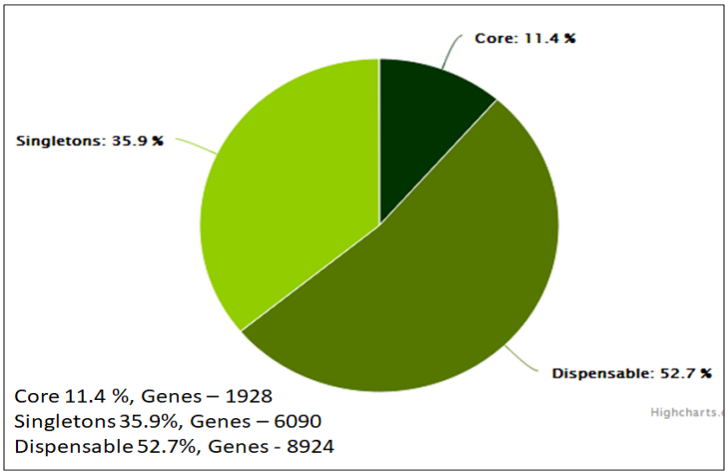

Fig. 1. Graphical representation of Genomic subset distribution.

Set size statistics:

Core development plot. There is an exponential decrease in the number of core genes with the increase in the number of genomes, as because with the addition of each new strains the number of genes those are common to all decreases.

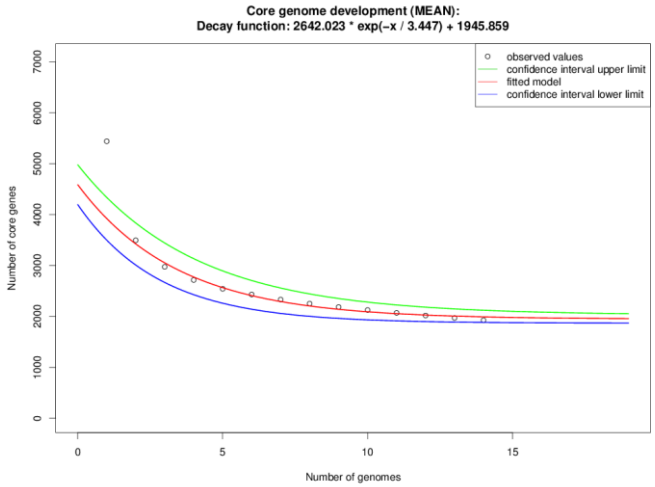

Fig. 2. An exponential decay function fitted to the calculated numbers of core genomes for genome counts from 2-12 (1 is ignored as it is just the number of genes per genome). Mean values are used for each genome count. The approximated core genome size is 1945 in this example. The number of genomes of all the *Pseudomonas* strains was placed on the x-axis to the number of core genes on the y-axis.

Singleton development plot. There is an exponential decrease in the number of singletons with the increase in the number of genomes the singletons are seen to decrease.

Table 2: The core development plot showed the following mean number of core genes of the contigs.

Sr. No.	Number of contigs	Mean no. of core genes
1.	*Pseudomonas_fluorescens_F113*	
2.	*Pseudomonas_protegens_CHA0*	3495.736
3.	*Pseudomonas_putida_W619*	2973.379
4.	*Pseudomonas_putida_DOT_T1E*	2716.968
5.	*Pseudomonas_denitrificans_ATCC_13867*	2542.265
6.	*Pseudomonas_stutzeri_A1501*	2431.231
7.	*Pseudomonas brassicacearum subsp. Brassicacearum_NFM421*	2330.364
8.	*Pseudomonas_putida_GB_1*	2251.284
9.	*Pseudomonas_S16*	2184.747
10.	*Pseudomonas_putida_BIRD_1*	2125.199
11.	*Pseudomonas_aeruginosa_PAO1*	2069.739
12.	*Pseudomonas_aeruginosa_M18*	2018.593
13.	*Pseudomonas_protegens_Pf_5*	1970.571
14.	*Pseudomonas_putida_F1*	1925.000

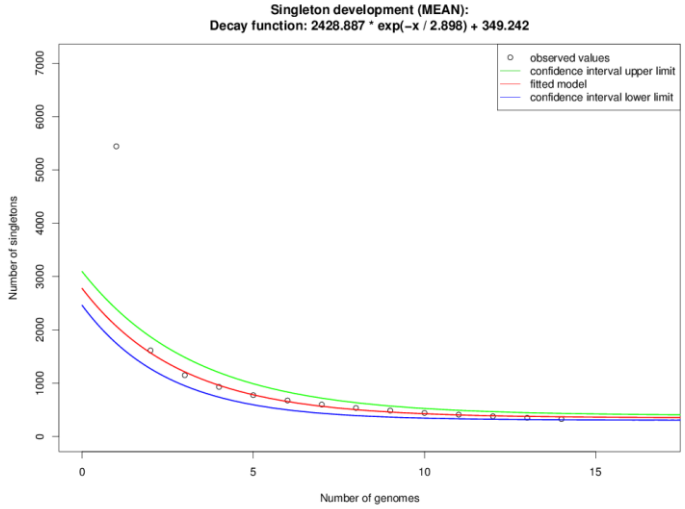

Fig. 3. An exponential decay function fitted to the calculated numbers of singletons for genome counts from 2-12 (1 is ignored as it is just the number of genes per genome). Mean values are used for each genome count. The approximated singleton size is 349. The total genomes of all the *Pseudomonas* strains were taken and the plot was observed showing the number of genomes on the x-axis and the number of singletons on the y-axis.

Table 3: The singleton development plot showed the following mean number of singleton genes of the contigs.

Sr. No.	No. of contigs	Mean no. of singleton genes
1.	*Pseudomonas_fluorescens_F113*	
2.	*Pseudomonas_protegens_CHA0*	1613.346
3.	*Pseudomonas_putida_W619*	1148.213
4.	*Pseudomonas_putida_DOT_T1E*	929.529
5.	*Pseudomonas_denitrificans_ATCC_13867*	774.953
6.	*Pseudomonas_stutzeri_A1501*	675.914
7.	*Pseudomonas brassicacearum* sub sp. *Brassicacearum_NFM421*	598.445
8.	*Pseudomonas_putida_GB_1*	534.854
9.	*Pseudomonas_S16*	485.928
10.	*Pseudomonas_putida_BIRD_1*	441.989
11.	*Pseudomonas_aeruginosa_PAO1*	409.400
12.	*Pseudomonas_aeruginosa_M18*	378.461
13.	*Pseudomonas_protegens_Pf_5*	350.846
14.	*Pseudomonas_putida_F1*	325.714

Pan development plot. The number of pan genes increased with the increase in the total number of genomes. Since pan genome considers all the full complement of genes in a clade, therefore with the addition of new genomes in the plot the pan genome increases.

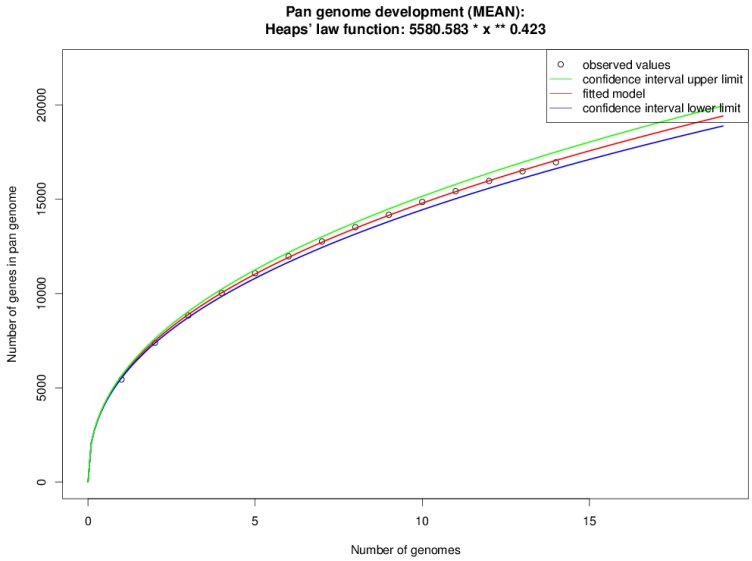

Fig. 4. Pan genome development plot for 14 *Pseudomonas* strains. The red curve shows the fitted exponential Heaps' law function. Based on these results the pan genome is considered to be open with a growth exponent of 0.423. The total genomes of all the *Pseudomonas* strains were taken and the plot was observed showing the number of genomes on the x-axis and the number of singletons on the y-axis.

Table 4: The pan development plot showed the following mean number of pan genes of the contigs.

Sr. No.	No. of contigs	Mean no. of pan genes
1.	*Pseudomonas_fluorescens_F113*	
2.	*Pseudomonas_protegens_CHA0*	7387.945
3.	*Pseudomonas_putida_W619*	8836.082
4.	*Pseudomonas_putida_DOT_T1E*	10020.448
5.	*Pseudomonas_denitrificans_ATCC_13867*	11079.376
6.	*Pseudomonas_stutzeri_A1501*	11972.704
7.	*Pseudomonas brassicacearum* subsp. *Brassicacearum_NFM421*	12762.436
8.	*Pseudomonas_putida_GB_1*	13521.946
9.	*Pseudomonas_S16*	14176.922
10.	*Pseudomonas_putida_BIRD_1*	14848.886
11.	*Pseudomonas_aeruginosa_PAO1*	15427.157
12.	*Pseudomonas_aeruginosa_M18*	15973.242
13.	*Pseudomonas_protegens_Pf_5*	16483.571
14.	*Pseudomonas_putida_F1*	16961.000

Synteny plots. In order to monitor the conservation of gene order among the *Pseudomonas* chromosomes, pairwise synteny plots were generated with EDGAR 2.0, where the position of each CDS of the chromosome given on the X axis is plotted against the position of its homologue in the second chromosome given on Y-axis.

Identical plots result in a diagonal plot. In the 1st plot, i.e. of *Pseudomonas fluorescens* F113 and *Pseudomonas brassicacearum* subsp. *brassicacearum* NFM421, there are very few chromosomal rearrangements which signify close similarity between their gene orders. The number of rearrangements seems to increase substantially from 4th to the 12th plot, and the plot takes somewhat a diagonal shape which indicates that the species may have similar gene order to their reference strain. But in *Pseudomonas putida* DOT_T1E the gene order seems to be completely disintegrated. Moreover, in the 2nd and 3rd plots, the graph is seen to be almost similar to each other, from which we can conclude that the strains *Pseudomonas_protegens*_CHA0 and *Pseudomonas_protegens*_Pf5 are very similar to each other in terms of their gene order; and they further takes quite a diagonal plot; which is an indication that they have similar gene order to the reference strain i.e. *Pseudomonas_fluorescens*_F113 (Fig. 5-7).

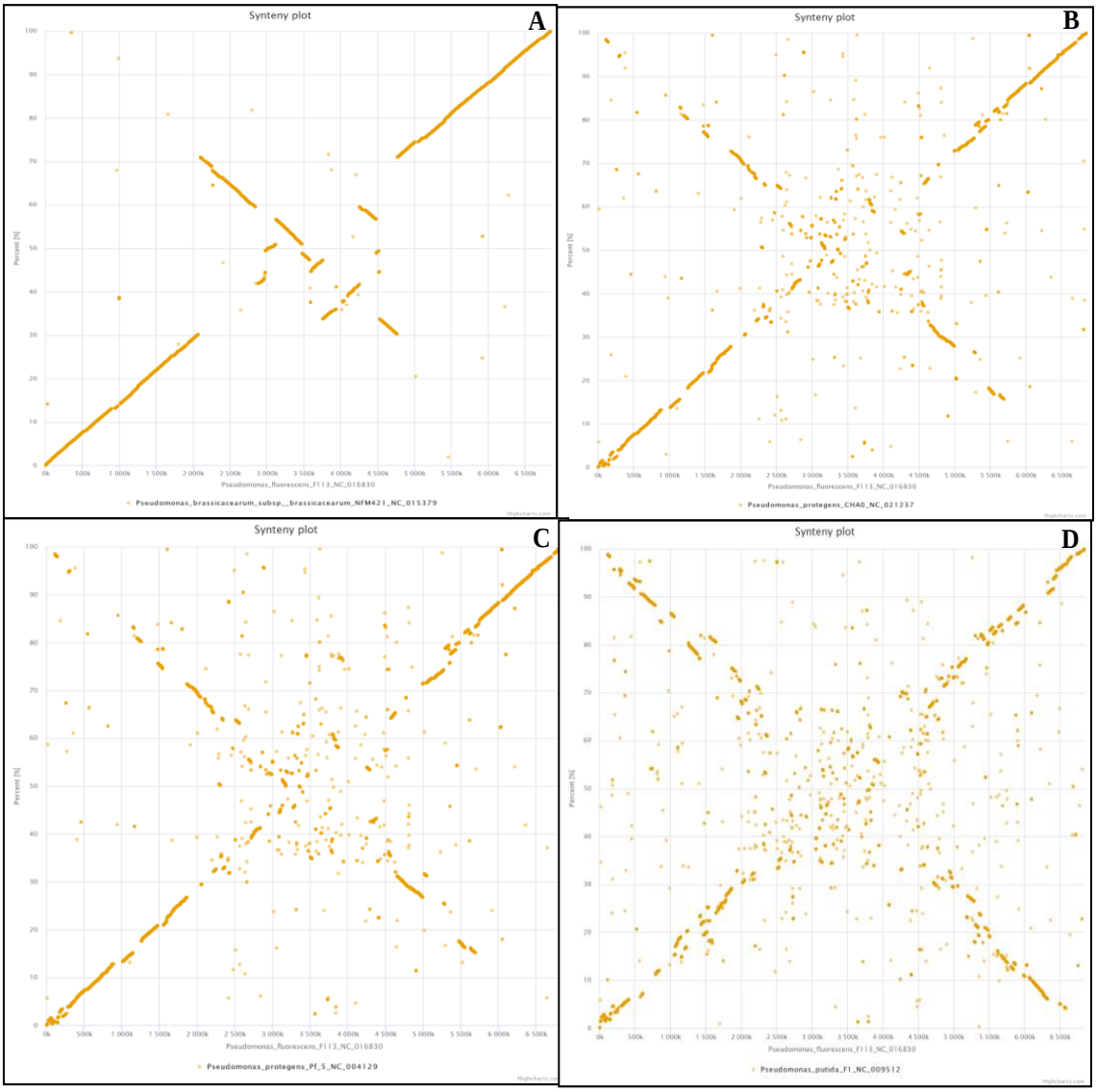

Fig. 5. The synteny plots for contigs **A.** *Pseudomonas brassicacearum_subsp._brassicacearum*_NFM42 **B.** *Pseudomonas_protegens*_CHAO **C.** *Pseudomonas_protegens*_Pf5 **D.** *Pseudomonas_putida*_F1 comparing each time with the reference strain (*Pseudomonas_fluorescens*_F113).

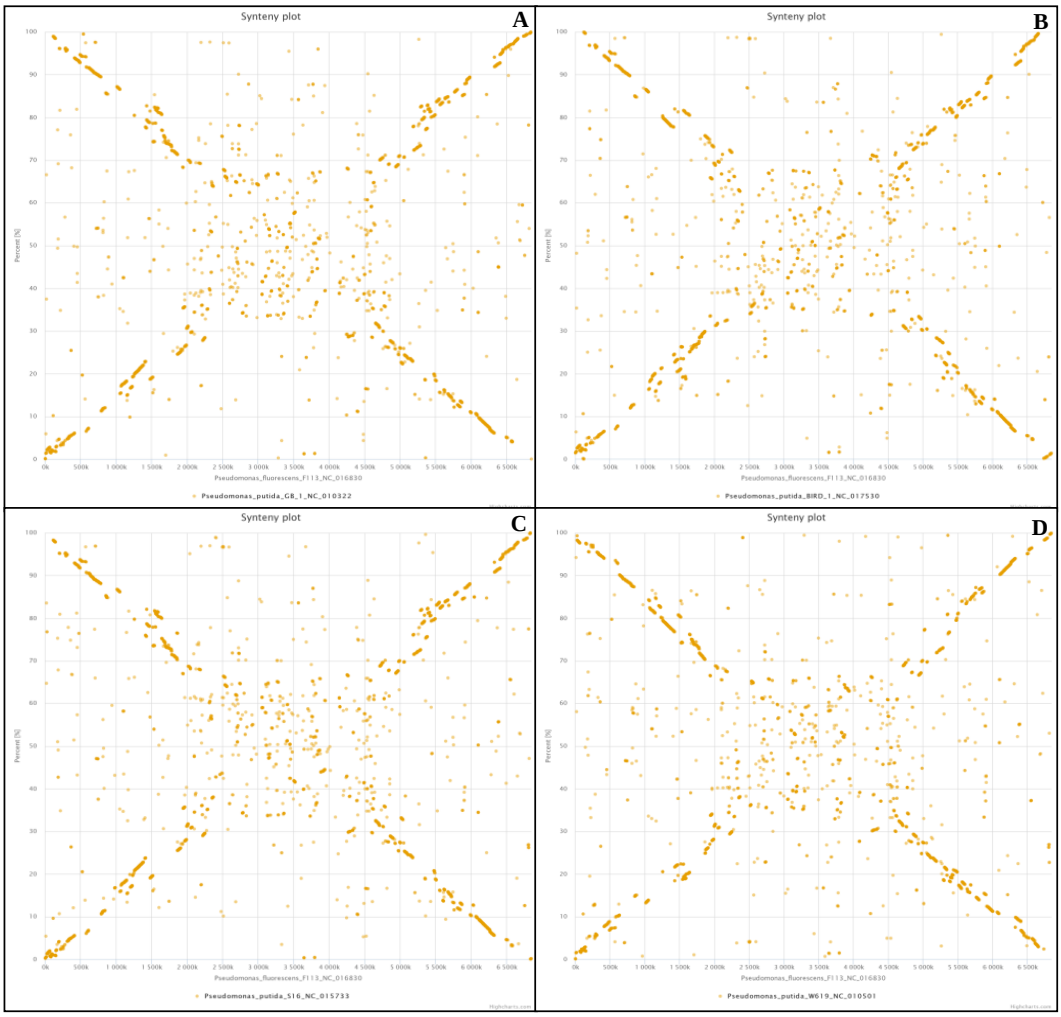

Fig. 6. The synteny plot for contigs **A.** *Pseudomonas_putida*_GB1 **B.** *Pseudomonas_putida*_BIRD **C.** *Pseudomonas putida*_S16 **D.** *Pseudomonas_putida*_W619 comparing each time with the reference strain (*Pseudomonas_fluorescens*_F113)

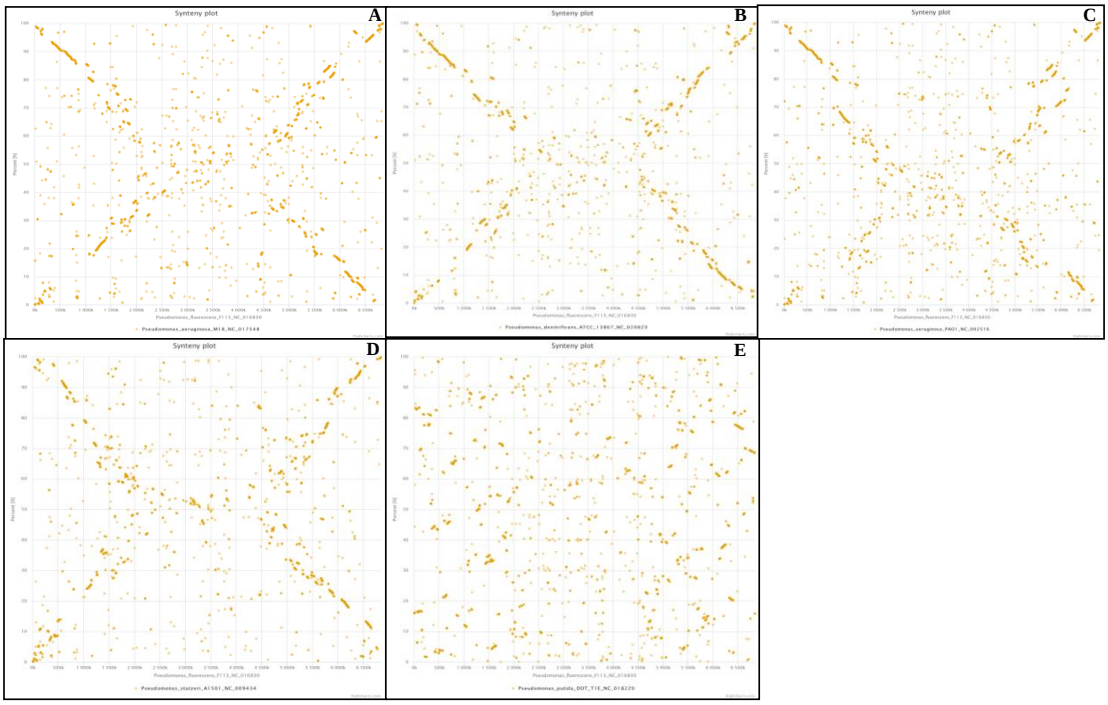

Fig. 7. The synteny plot for contigs **A.** *Pseudomonas_denitrificans*_ATCC_13867, **B.** *Pseudomonas_aeruginosa*_M1 **C.** *Pseudomonas_aeruginosa_PAO* **D.** *Pseudomonas_stutzeri*_A1501 **E.** *Pseudomonas_putida*_DOT_T1 Ecomparing each time with the reference strain (*Pseudomonas_fluorescens_F113)*

Genesets

Venn diagrams. The venn diagrams for 14 contigs were analyzed comparing each time with the reference strain. Since 5 contigs can be selected at a time, the analysis was carried out in 4 different sets.

The first set showed that the venn diagram generated shows that the chromosomes shared 2832 orthologous CDS exclusively among all the four strains including the reference strain *Pseudomonas fluorescens F113*. While *Pseudomonas fluorescens F113* (reference) shared 1501 orthologous CDS with *Pseudomonas brassicacearum* sub sp. *brassicacearum NFM421* which shows it shares a great similarity in their gene order. On the other hand, Pseudomonas *aeruginosa PAO1* and *Pseudomonas aeruginosa M18* shares 1154 orthologous CDS exclusively indicating somewhat high similarity in their gene order. The second set showed that the venn diagram generated shows, that the chromosomes of all the four strains shared 3029 orthologous CDS. It is also seen that *Pseudomonas protegens CHAO* and *Pseudomonas protegens Pf5* shares 1362 orthologous CDS which indicates close similarity of their gene orders between these two species. The third set showed that the venn diagram generated shows that the chromosomes of all the three Pseudomonas strains shares 3210 orthologous CDS. While the fourth set showed that the venn diagram generated shows that the chromosomes of the two Pseudomonas shares 2286 orthologous CDS exclusively (Fig. 8).

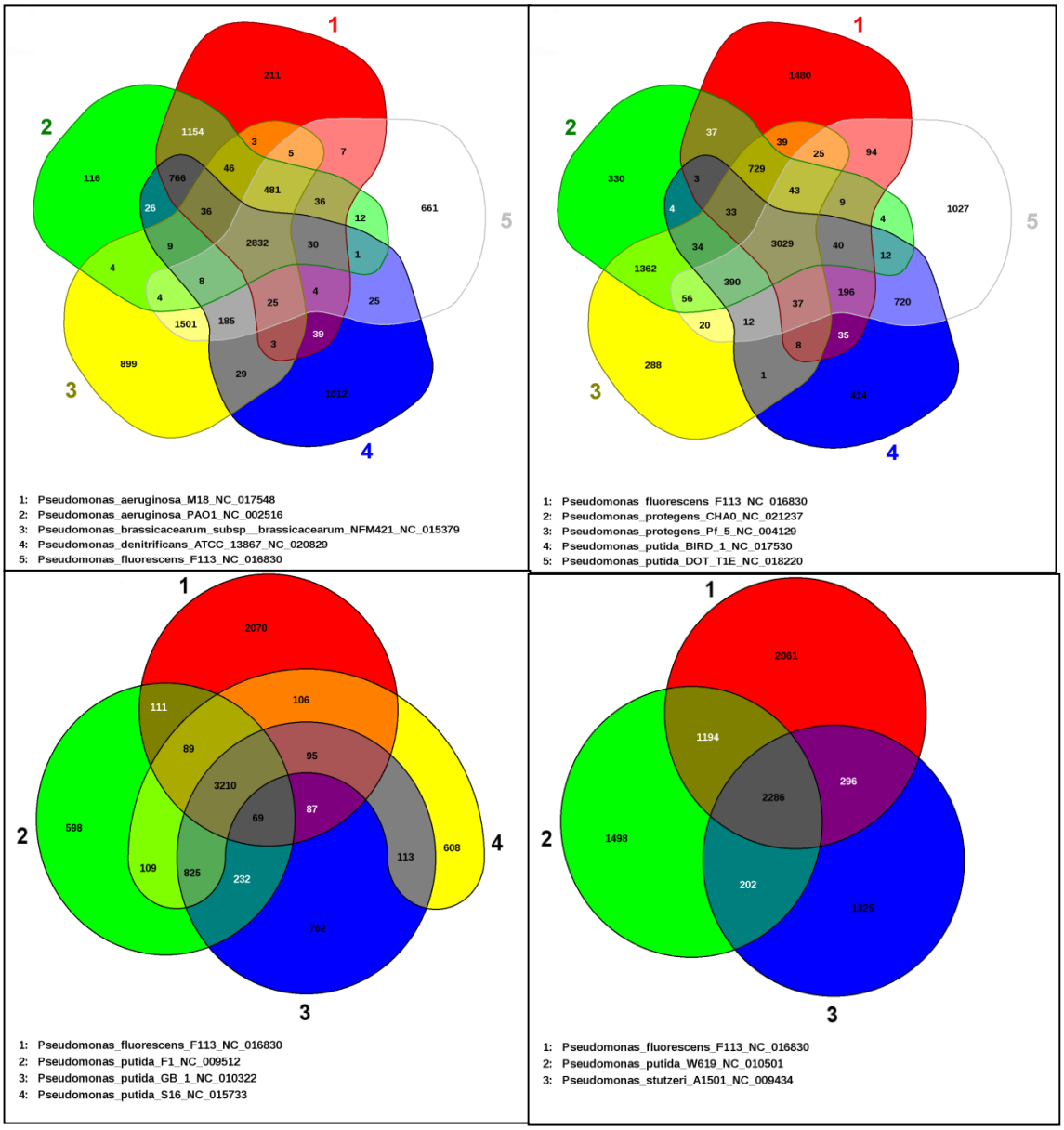

Fig. 8. Venn diagrams generated for the 14 contigs analyzed in 4 different data sets

AAI and ANI matrix. The AAI/ANI mean and median matrix thus generated showing the average AAIs/ANIs of all conserved genes in the core genome. The resulting phylogenetic distance values are arranged in an AAI/ANI matrix, clustered according to their distance patterns and visualized as heatmaps. Based on the colours of the markers given on each left corner of the heatmaps, the average distance patterns between the different strains are known (Fig. 9).

Pseudomonas putida_GB, with *Pseudomonas putida_BIRD1* shows 37495 pairwise LCBs, 366 Mb working set size, 345 Mb Pagefile usage, root

alignment having 593 superintervals, 8852132 root alignment length and the organisms have 61.2% of GC content. While with *Pseudomonas putida_S16* shows 35988 pairwise LCBs, 344Mb working set size, 257 Mb Pagefile usage, root alignment having 577 superintervals, 8980641 root alignment length and the organisms have 61.5% of GC content. On the other hand, with *Pseudomonas putida_W619* shows 35049 pairwise LCBs, 358 Mb working set size, 313 Mb Pagefile usage, root alignment having 362 superintervals, 8664271 root alignment length and the organisms have 61% of GC content (Fig. 11).

Pseudomonas_fluorescens F113 and *Pseudomonas denitificans_ATCC_13867* showed 45240 pairwise LCBs, 370 Mb working set size, 313 Mb Pagefile usage, root alignment having 603 superintervals, 9103152 root alignment length and the organisms have 62.8% of GC content. While with *Pseudomonas aeruginosa M18* shows 45399 pairwise LCBs, 359 Mb working set size, 324 Mb Pagefile usage, root alignment having 703 superintervals, 9317418 root alignment length and the organisms have 63.5% of GC content. On the other hand, *Pseudomonas aeruginosa PAO1*shows 44954 pairwise LCBs, 376 Mb working set size, 299 Mb Pagefile usage, root alignment having 703 superintervals, 9257021 root alignment length and the organisms have 63.5% of GC content, with *Pseudomonas stutzeri A1501* showed 41959 pairwise LCBs, 367 Mb working set size, 373 Mb Pagefile usage, root alignment having 312 superintervals, 8894155 root alignment length and the organisms have 62% of GC content. While *Pseudomonas putida DOT_T1E* showed 42125 pairwise LCBs, 333 Mb working set size, 274 Mb Pagefile usage, root alignment having 687 superintervals, 9257906 root alignment length and the organisms have 61.1% of GC content (Fig. 12).

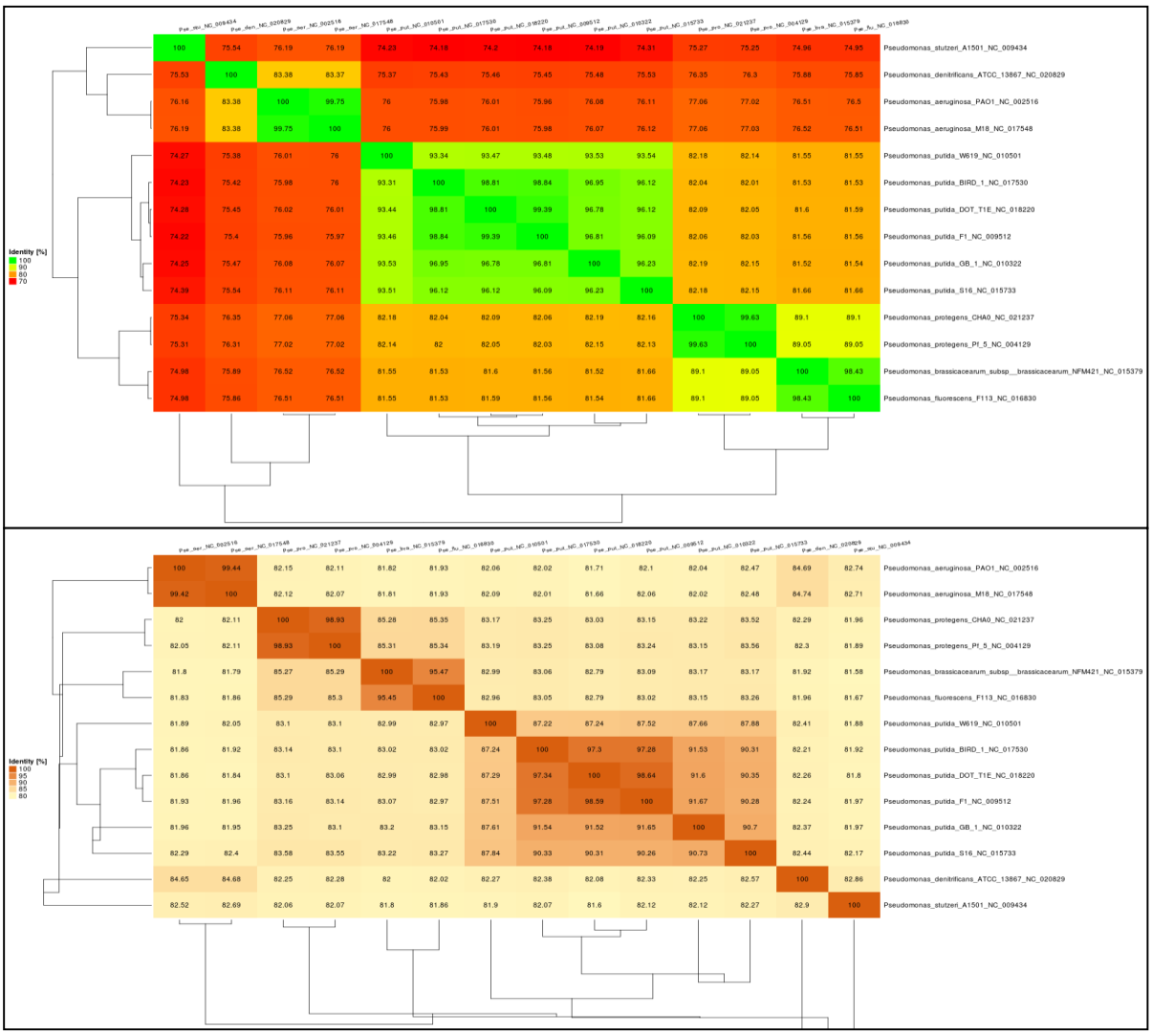

Fig. 9. Heat maps generated for AAI and ANI matrix for all 14 Pseudomonas strains considered for the study.

MAUVE 2.3.1. The display layout of the various alignments, shows the alignments between *Pseudomonas_fluorescens*_F113 (reference strain) with all the other 13 *Pseudomonas* strains taken for study. The figures shows colored blocks in the first genome that are connected by lines to similarly colored blocks in the second genomes. These lines indicate which regions in each genome are homologous. Each contiguously colored region is a locally collinear block, a region without rearrangement of homologous backbone sequence. LCBs below a genome's center line are in the reverse complement orientation relative

to the reference genome. Lines between genomes trace each orthologous LCB through every genome. The images were generated by the Mauve rearrangement viewer. The Mauve rearrangement viewer enables users to interactively zoom in on regions of interest and examine the local rearrangement structure.

The results revealed that *Pseudomonas_fluorescens* F113 (reference strain) with *Pseudomonas brassicacearum subsp. brassicacearum* NFH421 showed 4659 pairwise LCBs, 239Mb working set size, 172 Mb Pagefile usage, root alignment having 9 superintervals, 7521541 root alignment length and the organisms have 60.7% of GC content. While with *Pseudomonas protegens*_CHAO shows 7269 pairwise LCBs, 250Mb working set size, 223 Mb Pagefile usage, root alignment having 28 superintervals, 7514346 root alignment length and the organisms have 61.3% of GC content, with *Pseudomonas_protegens*_Pf5 showed 40832 pairwise LCBs, 340 Mb working set size, 263 Mb Pagefile usage, root alignment having 343 superintervals, 9158705 root alignment length and the organisms have 62% of GC content, with *Pseudomonas putida_F1* shows 38870 pairwise LCBs, 348 Mb working set size, 270 Mb Pagefile usage, root alignment having 623 superintervals, 8932757 root alignment length and the organisms have 61.3% of GC content (Fig. 10).

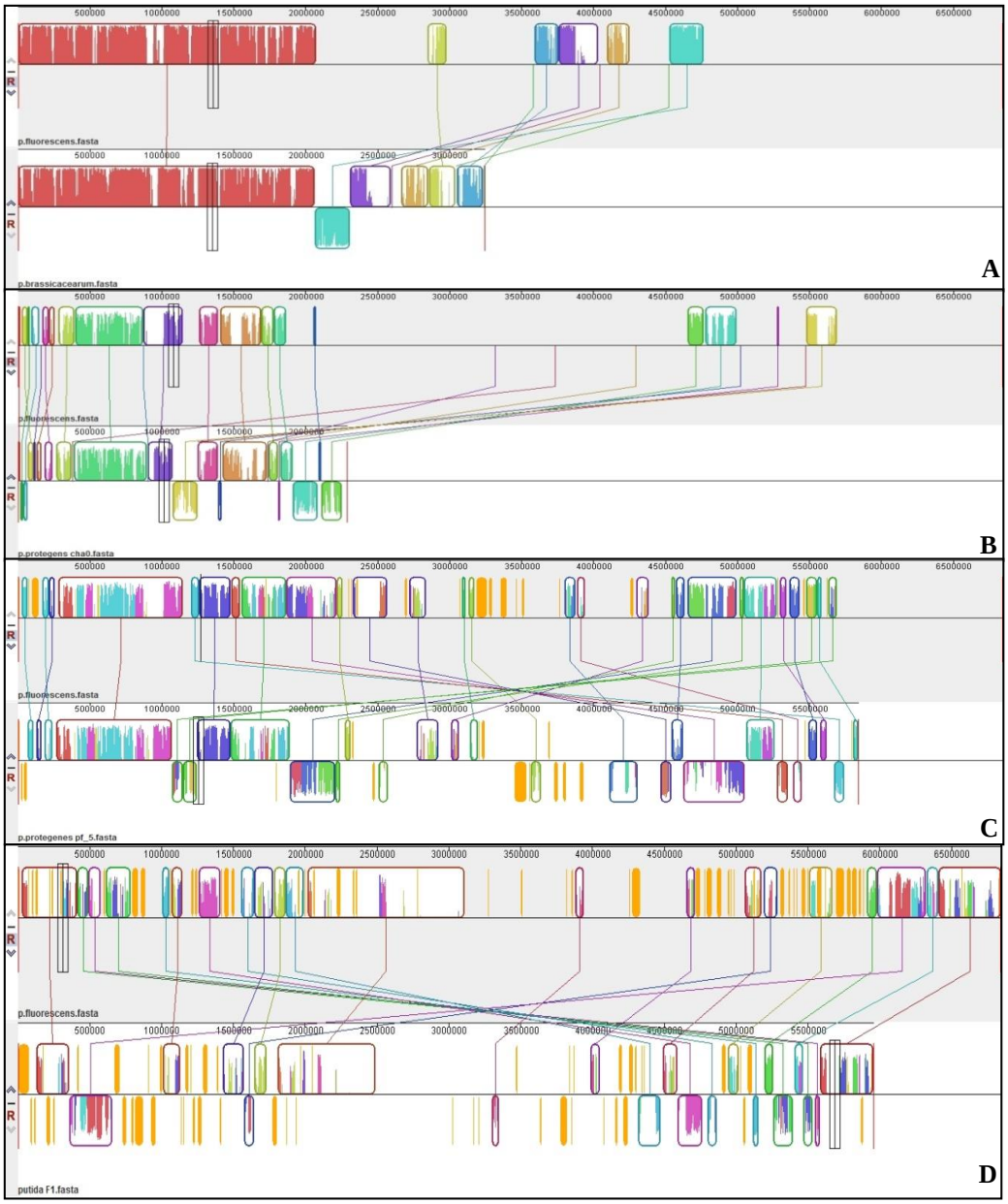

Fig. 10. The Mauve output file format *Pseudomonas_fluorescens*_F113 (reference strain) with **A.** *Pseudomonas brassicacearum* sub sp. *Brassicacearum,* **B.** *Pseudomonas protegens_CHAO,* **C.** *Pseudomonas_protegens*_Pf5, **D.** *Pseudomonas putida_F1*

On the other handrefrence strain showed 39556 pairwise LCBs, 352 Mb working set size, 284 Mb Pagefile usage, root alignment having 603 superintervals, 9149257 root alignment length and the organisms have 61.3% of GC contentwith *Pseudomonas putida_GB,* with *Pseudomonas putida_BIRD1* shows 37495 pairwise LCBs, 366 Mb working set size, 345 Mb Pagefile usage, root alignment having 593 superintervals, 8852132 root alignment length and the organisms have 61.2% of GC

content. While with *Pseudomonas putida_S16* shows 35988 pairwise LCBs, 344Mb working set size, 257 Mb Pagefile usage, root alignment having 577 superintervals, 8980641 root alignment length and the organisms have 61.5% of GC content. On the other hand, with *Pseudomonas putida_W619* shows 35049 pairwise LCBs, 358 Mb working set size, 313 Mb Pagefile usage, root alignment having 362 superintervals, 8664271 root alignment length and the organisms have 61% of GC content (Fig. 11).

Pseudomonas_fluorescens F113 and *Pseudomonas denitificans_ATCC_13867* showed 45240 pairwise LCBs, 370 Mb working set size, 313 Mb Pagefile usage, root alignment having 603 superintervals, 9103152 root alignment length and the organisms have 62.8% of GC content. While with *Pseudomonas aeruginosa M18* shows 45399 pairwise LCBs, 359 Mb working set size, 324 Mb Pagefile usage, root alignment having 703 superintervals, 9317418 root alignment length and the organisms have 63.5% of GC content. On the other hand, *Pseudomonas aeruginosa PAO1* shows 44954 pairwise LCBs, 376 Mb working set size, 299 Mb Pagefile usage, root alignment having 703 superintervals, 9257021 root alignment length and the organisms have 63.5% of GC content, with *Pseudomonas stutzeri A1501* showed 41959 pairwise LCBs, 367 Mb working set size, 373 Mb Pagefile usage, root alignment having 312 superintervals, 8894155 root alignment length and the organisms have 62% of GC content. While *Pseudomonas putida DOT_T1E* showed 42125 pairwise LCBs, 333 Mb working set size, 274 Mb Pagefile usage, root alignment having 687 superintervals, 9257906 root alignment length and the organisms have 61.1% of GC content (Fig. 12).

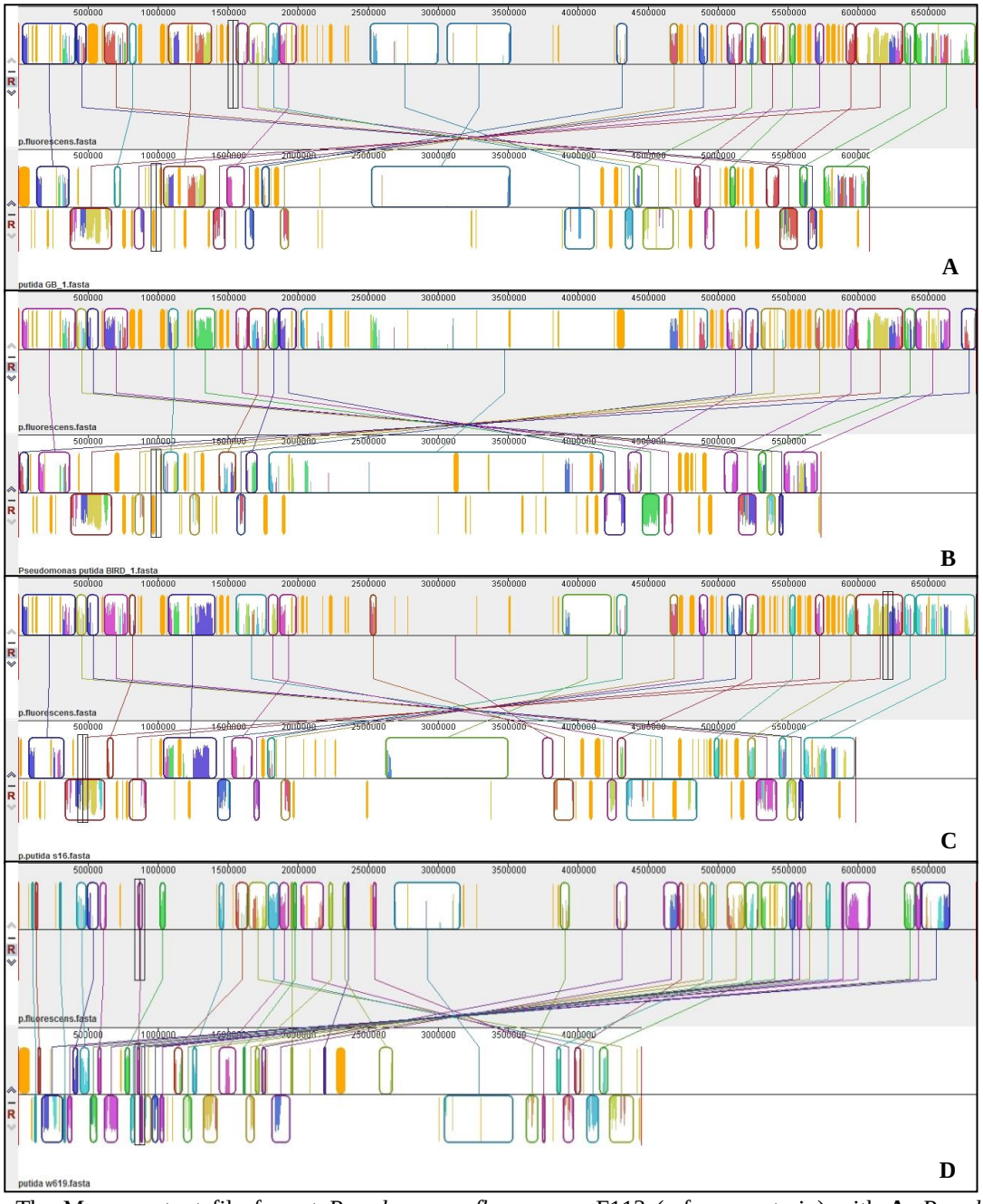

Fig. 11. The Mauve output file format *Pseudomonas_fluorescens*_F113 (reference strain) with **A.** *Pseudomonas putida_GB1*, **B.** *Pseudomonas putida_BIRD1*, **C.** *Pseudomonas putida_S16*, **D.** *Pseudomonas putida_W619.*

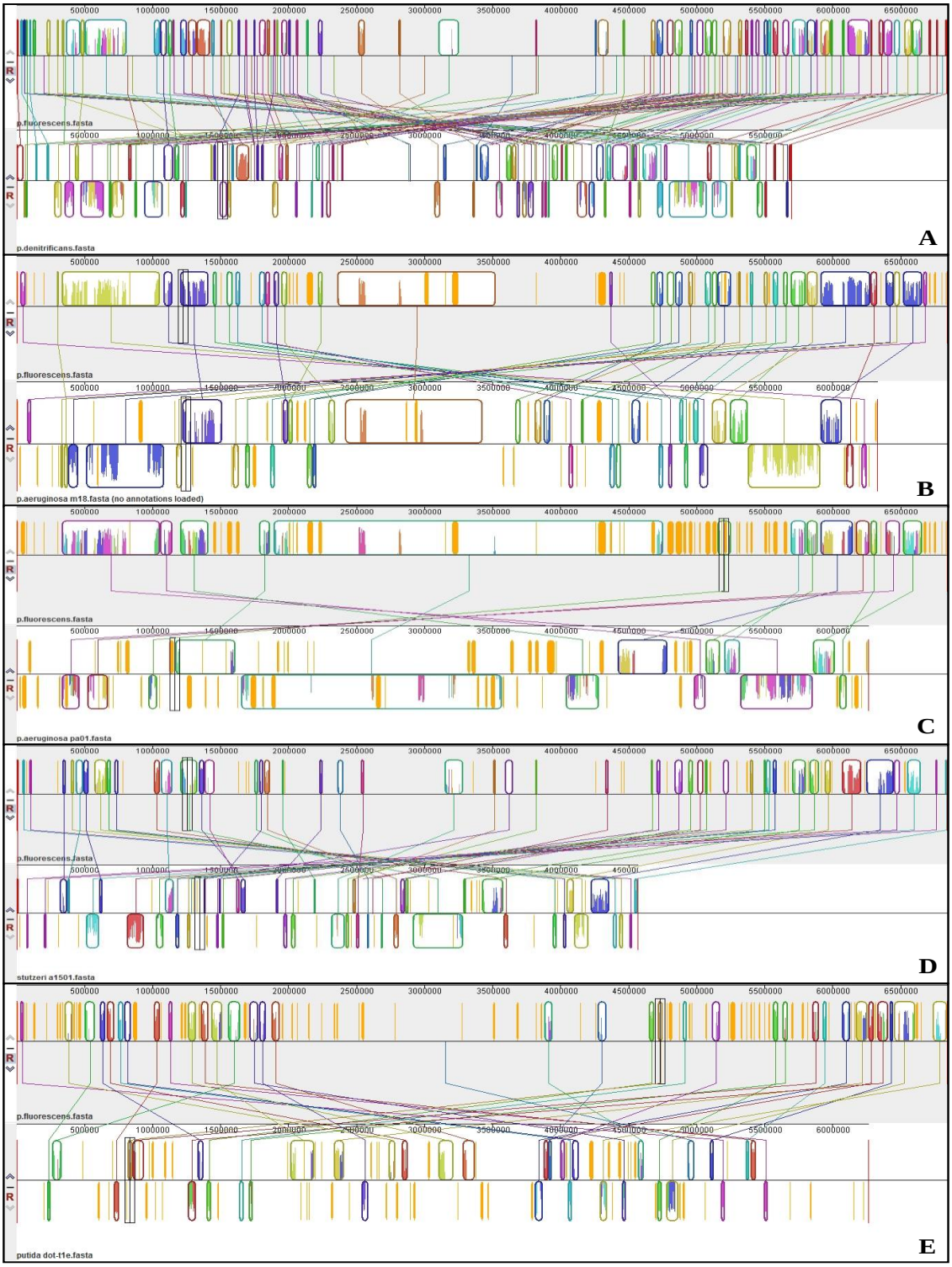

Fig. 12. The Mauve output file format *Pseudomonas_fluorescens*_F113 (reference strain) with **A.** *Pseudomonas denitificans_ATCC_13867*, **B.** *Pseudomonas aeruginosa M18*, **C.** *Pseudomonas aeruginosa PAO1*, **D.** *Pseudomonas stutzeri A1501*, **E.** *Pseudomonas putida DOT_T1E*

From the above observations, it is seen that the lowest GC content was found in aligments made between *Pseudomonas fluorescens F113* (reference strain) and *Pseudomonas brassicacearum subsp. brassicacearum NFH421*. On the other hand the pairwise LCBs were seen to be less i.e. 4659, indicating that there were very fewer rearrangement in these two species depicting it to have close order of their genes. Thus, the results revealed that although differential gene content exist, however comparative studies of biocontrol and plant growth promoting bacteria shows closely related gene order and gene content which makes it a potent source in the application of sustainable agriculture. Previous studies showed that phytohormones synthesized by some species of endophytic bacteria such as *Burkholderia* and *Paraburkholderia* lead to enhanced growth of plants which on the other hand makes it vital for use as biocontrol agents (Dias *et al.*, 2019; Chen *et al.*, 2023). Similar works were also carried out in some strains of *Bacillus* and *Pseudomonas* (Wang *et al.*,

2023). Moreover, genome of *Stropharia rugosoannulata* showed its strong potentiality to be used to bioremediate and degrade lignin alongwith with ability to function as biocontrol agents against nematodes (Yang *et al.,* 2022).

CONCLUSIONS

The study undertaken elucidated the comparative genomics between different *Pseudomonas* strains that have biocontrol, plant-growth promoting and bioremediation activities. The study revealed how closely or distantly the species are related in terms of their gene orders that enumerated genome rearrangements, orthologous regions of one genome to that of another in course of their evolution. The observations also revealed that some of the biocontrol and plant-growth promoting *Pseudomonas* are somewhat closely related in terms of their gene orders, and hence, can be used in either way i.e. in pathogenic and remediation projects. Thereby, the softwares used supports a quick and user-friendly survey of evolutionary relationships between microbial genomes and simplifies the process of obtaining new biological insights into their differential gene content.

FUTURE SCOPE

The *Pseudomonas* strains were identified as having biocontrol, bioremediation, plant growth promoting activity altogether therefore can be used reciprocally possibly replacing the use of harmful pesticides for sustainable agriculture.

Acknowledgement. Sincere thanks to Botany Department, Gauhati University for providing the necessary resources and guidance in completing the research.

Conflict of Interest. None.

REFERENCES

Ainala, S. K., Somasundar, A. and Park, S. (2013). Complete genome sequence of *Pseudomonas denitrificans* ATCC 13867. *Genome announcements*, *1*(3), e00257-13.

Backer, R., Rokem, J. S., Ilangumaran, G., Lamont, J., Praslickova, D., Ricci, E., Subramanian, S. and Smith, D. L. (2018). Plant growth-promoting rhizobacteria: context, mechanisms of action, and roadmap to commercialization of biostimulants for sustainable agriculture. *Frontiers in Plant Science*, *9*, 1473.

Bajpai, R., Basavaraj Teli, M., Rashid, M., Nanda, S., Yadav, S. K. and Kumar, G. (2021). Biocontrol of *Fusarium* Wilt in Tomato: An Eco-friendly and Cost Effective Approach. *Biological Forum – An International Journal 13*(2), 62-69.

Bhargavanandha, K. S., Kshirsagar, R. B., Patil, B. M. and Bachate, A. H. (2021). Effect of addition of partially hydrolyzed guar gum on the quality attributes of tomato soup powder.

Blom, J., Albaum, S. P., Doppmeier, D., Pühler, A., Vorhölter, F. J., Zakrzewski, M. and Goesmann, A. (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC bioinformatics*, *10*, 1-14.

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C. and Goesmann, A. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic acids research*, *44*(W1), pp.W22-W28.

Chandler, C. E., Horspool, A. M., Hill, P. J., Wozniak, D. J., Schertzer, J. W., Rasko, D. A. and Ernst, R. K. (2019). Genomic and phenotypic diversity among ten laboratory isolates of *Pseudomonas aeruginosa* PAO1. *Journal of bacteriology*, *201*(5), e00595-18.

Chen, Y., Yang, L., Zhang, L., Li, J., Zheng, Y., Yang, W., Deng, L., Gao, Q., Mi, Q., Li, X. and Zeng, W. (2023). Autotoxins in continuous tobacco cropping soils and their management. *Frontiers in Plant Science*, *14.*

Choudhury, A. and Bordolui, S. K. (2022). Inducement of Seed Priming with Potassium Nitrate on quality Performance of Chickpea (*Cicer arietinum* L.). *Biological Forum–An International Journal 14*(4), 779-783.

Cubry, P., Vigouroux, Y. and François, O. (2017). The empirical distribution of singletons for geographic samples of DNA sequences. *Frontiers in Genetics*, *8*, 139.

Dangi, M. B., Urynowicz, M. A., Schultz, C. L., Budhathoki, S. and Dangi, S. R. (2021). Analysis of the effects of in-situ chemical oxidation on microbial activity using *Pseudomonas putida* F1. *Heliyon*, *7*(12), e08665.

Das, H., Saha, O., Rahman, N., Phukon, M. and Kalita, R. (2021). Genomics-led Insight into Potyvirus Family, Prevalence and Management. *In Biological Forum–An International Journal 13*(4), 254-262).

Dias, G. M., de Sousa Pires, A., Grilo, V. S., Castro, M. R., de Figueiredo Vilela, L. and Neves, B. C. (2019). Comparative genomics of *Paraburkholderia kururiensis* and its potential in bioremediation, biofertilization, and biocontrol of plant pathogens. *Microbiology Open*, *8*(8), e00801.

Dieckmann, M. A., Beyvers, S., Nkouamedjo-Fankep, R. C., Hanel, P. H. G., Jelonek, L., Blom, J. and Goesmann, A. (2021). EDGAR3. 0: comparative genomics and phylogenomics on a scalable infrastructure. *Nucleic Acids Research*, *49*(W1), W185-W192.

Flury, P., Vesga, P., Dominguez-Ferreras, A., Tinguely, C., Ullrich, C. I., Kleespies, R. G., Keel, C. and Maurhofer, M. (2019). Persistence of root-colonizing *Pseudomonas protegens* in herbivorous insects throughout different developmental stages and dispersal to new host plants. *The ISME journal*, *13*(4), 860-872.

Franzino, T., Boubakri, H., Cernava, T., Abrouk, D., Achouak, W., Reverchon, S., Nasser, W. and elZahar Haichar, F. (2021). Implications of carbon catabolite repression for plant-microbe interactions. *Plant Communications*, p.100272.

Henkels, M. D., Kidarsa, T. A., Shaffer, B. T., Goebel, N. C., Burlinson, P., Mavrodi, D. V., Bentley, M. A., Rangel, L. I., Davis, E. W., Thomashow, L. S. and Zabriskie, T. M. (2014). *Pseudomonas protegens* Pf-5 causes discoloration and pitting of mushroom caps due to the production of antifungal metabolites. *Molecular Plant-Microbe Interactions*, *27*(7), 733-746.

Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D. H. and Soo, R. M. (2021). Prokaryotic taxonomy and nomenclature in the age of big sequence data. *the ISME Journal*, *15*(7), 1879-1892.

Inglin, R. C., Meile, L. and Stevens, M. J. (2018). Clustering of pan-and core-genome of *Lactobacillus* provides novel evolutionary insights for differentiation. *BMC genomics*, *19*(1), 1-15.

Jayachandran, A., Jain, S., Saini, S., Maurya, P., Jagga, S., Shukla, K.K. and Kiran, B. (2018). Genome Editing in Fruit Crops – A Review. *Biological Forum – An International Journal*, *14*(4), 01-06.

Kaur, T., Kour, D., Pericak, O., Olson, C., Mohan, R., Yadav, A., Mishra, S., Kumar, M., Rai, A. K. and Yadav, A. N. (2022). Structural and functional diversity of plant growth promoting microbiomes for agricultural sustainability. *Journal of Applied Biology and Biotechnology*, *10*(1), 70-89.

Kharte, S., Kumar, A., Sharma, S., Ramakrishnan, R.S., Kumar, S., Malvi, S., Singh, Y. and Kurmi, S. (2022). In vitro Evaluation of Fungicides and Bio-agents for the Management of Lentil Wilt caused by *Fusarium oxysporum* f. sp. lentis. *In Biological Forum–An International Journal*, *14*(4), 489-495.

Konstantinidis, K. T. and Tiedje, J. M. (2005). Towards a genome-based taxonomy for prokaryotes. *Journal of bacteriology*, *187*(18), 6258-6264.

Koul, B., Singh, S., Dhanjal, D. S. and Singh, J. (2019). Plant growth-promoting rhizobacteria (PGPRs): a fruitful resource. *Microbial Interventions in Agriculture and Environment: Volume 3: Soil and Crop Health Management*, 83-127.

Kumar, R. and Pareek, N. K. (2022). Influence of Different Irrigation and Nitrogen Levels on Water and Nitrogen Use Efficiency of Cluster bean (*Cyamopsis tetragonoloba*). *Biological Forum–An International Journal*, *14*(1), 944-949.

Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A. and Li, Y. (2018). Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Frontiers in genetics*, *9*, 237.

Maity, J. P., Samal, A. C., Rajnish, K., Singha, S., Sahoo, T. R., Chakraborty, S., Bhattacharya, P., Chakraborty, S., Sarangi, B. S., Dey, G. and Banerjee, P. (2023). Furfural removal from water by bioremediation process by indigenous *Pseudomonas putida* (OSBH3) and *Pseudomonas aeruginosa* (OSBH4) using novel suphala media: An optimization for field application. *Groundwater for Sustainable Development*, *20*, 100895.

Marcais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L. and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS computational biology*, *14*(1), e1005944.

Moore, E. R., Tindall, B. J., Martins Dos Santos, V. A., Pieper, D. H., Ramos, J. L. and Palleroni, N. J. (2006). Nonmedical: *Pseudomonas*. *The prokaryotes*, *6*, 646-703.

Ojuederie, O. B. and Babalola, O. O. (2017). Microbial and plant-assisted bioremediation of heavy metal polluted environments: a review. *International journal of environmental research and public health*, *14*(12), 1504.

Park, S. C., Lee, K., Kim, Y. O., Won, S. and Chun, J. (2019). Large-scale genomics reveals the genetic characteristics of seven species and importance of phylogenetic distance for estimating pan-genome size. *Frontiers in microbiology*, *10*, 834.

Patel, S. K., Pratap, C. B., Verma, A. K., Jain, A. K., Dixit, V. K. and Nath, G. (2013). *Pseudomonas fluorescens*-like bacteria from the stomach: a microbiological and molecular study. *World Journal of Gastroenterology: WJG*, *19*(7), 1056.

Qin, S., Xiao, W., Zhou, C., Pu, Q., Deng, X., Lan, L., Liang, H., Song, X. and Wu, M. (2022). *Pseudomonas aeruginosa*: pathogenesis, virulence factors, antibiotic resistance, interaction with host, technology advances and emerging therapeutics. *Signal Transduction and Targeted Therapy*, *7*(1), p.199.

Raklami, A., Meddich, A., Oufdou, K., Baslam, M. (2022). Plants—microorganisms-based bioremediation for heavy metal cleanup: recent developments, phytoremediation techniques, regulation mechanisms, and molecular responses. *International Journal of Molecular Sciences*, *23*(9), 5031.

Roca, A., Pizarro-Tobías, P., Udaondo, Z., Fernández, M., Matilla, M. A., Molina-Henares, M. A., Molina, L., Segura, A., Duque, E. and Ramos, J. L. (2013). Analysis of the plant growth-promoting properties encoded by the genome of the rhizobacterium *Pseudomonas putidaBIRD-1*. *Environmental microbiology*, *15*(3), 780-794.

Saeed, Q., Xiukang, W., Haider, F.U., Kučerik, J., Mumtaz, M.Z., Holatko, J., Naseem, M., Kintl, A., Ejaz, M., Naveed, M. and Brtnicky, M. (2021). Rhizosphere bacteria in plant growth promotion, biocontrol, and bioremediation of contaminated sites: A comprehensive review of effects and mechanisms. *International Journal of Molecular Sciences*, *22*(19), 10529.

Sah, S., Krishnani, S. and Singh, R. (2021). *Pseudomonas* mediated nutritional and growth promotional activities for sustainable food security. *Current Research in Microbial Sciences*, *2*, 100084.

Scales, B. S., Dickson, R. P., LiPuma, J. J. and Huffnagle, G. B. (2014). Microbiology, genomics, and clinical significance of the *Pseudomonas fluorescens* species complex, an unappreciated colonizer of humans. *Clinical microbiology reviews*, *27*(4), 927-948.

Thakur, P., Tiwari, R. K. S., Nirmalkar, V. K., Verma, S. K. and Chaure, N. K. (2022). Antagonistic activity and Bioefficacy of *Bacillus subtilis* against Sheath Blight of Rice (*Oryza sativa* L.) caused by *Rhizoctonia solani*. *In Biological Forum-An International Journal* *14*(3), 1305-1311.

Tsukanova, K. A., Meyer, J. J. M. and Bibikova, T. N. (2017). Effect of plant growth-promoting Rhizobacteria on plant hormone homeostasis. *South African Journal of Botany*, *113*, 91-102.

Wang, Z., Lu, K., Liu, X., Zhu, Y. and Liu, C. (2023). Comparative Functional Genome Analysis Reveals the Habitat Adaptation and Biocontrol Characteristics of Plant Growth-Promoting Bacteria in NCBI Databases. *Microbiology Spectrum*, e05007-22.

Weimer, A., Kohlstedt, M., Volke, D. C., Nikel, P. I. and Wittmann, C. (2020). Industrial biotechnology of *Pseudomonas putida*: advances and prospects. *Applied Microbiology and Biotechnology*, *104*, 7745-7766.

Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J. and van der Lelie, D. (2011). Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS microbiology reviews*, *35*(2), 299-323.

Yang, Y., Meng, G., Ni, S., Zhang, H. and Dong, C. (2022). Genomic analysis of *Stropharia rugosoannulata* reveals its nutritional strategy and application potential in bioremediation. *Journal of Fungi*, *8*(2), 162.

Young, A. D. and Gillung, J. P. (2020). Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, *45*(2), 225-247.

Zdobnov, E. M. and Bork, P. (2007). Quantification of insect genome divergence. *Trends in genetics*, *23*(1), 16-20.

Zhang, L., Chen, W., Jiang, Q., Fei, Z. and Xiao, M. (2020). Genome analysis of plant growth-promoting rhizobacterium *Pseudomonas chlororaphis subsp. aurantiaca* JD37 and insights from comparison of genomics with three *Pseudomonas* strains. *Microbiological research*, *237*, 126483.

Zhao, T., Zwaenepoel, A., Xue, J. Y., Kao, S. M., Li, Z., Schranz, M. E. and Van de Peer, Y. (2021). Whole-genome microsynteny-based phylogeny of angiosperms. *Nature Communications*, *12*(1), 3498.

Zheng, Y., Li, Y., Long, H., Zhao, X., Jia, K., Li, J., Wang, L., Wang, R., Lu, X. and Zhang, D. (2018). bifA regulates biofilm development of *Pseudomonas putida* MnB1 as a primary response to H_2O_2 and Mn^{2+}. *Frontiers in microbiology*, *9*, 1490.

How to cite this article: Mehzabin Rehman, Sabnoor Yeasrin Jyoti, Preetom Regon, Subrat Kakati, Piyush Kumar Mishra and Bhaben Tanti (2023). Comparative Genomics Analyses of some Selected *Pseudomonas* strains having Biocontrol, Plant Growth Promoting and Bioremediation activities using Bioinformatic Tools. *Biological Forum – An International Journal*, *15*(6): 01-16.