Genome Wide Identification, Characterization and Evolutionary Analysis of T6SS in *Burkholderia cenocepacia* Strains

Sardarni K. Raza[1*], Muhammad Ibrahim[2], Kainaat[1], Afshan Farid[1], Sardar S. N. Khan[3], Aqsa Talat[3], Tayyaba Arzoo[4], Syeda I. Fatima[3], Abdul Waheed[1]

[1]*Department of Botany*
Pir Mehr Ali Shah Arid Agriculture University
Shamsabad, Murree Road, Rawalpindi, Punjab, Pakistan
E-mails: khadijaraza3@gmail.com, kainaatkarim@gmail.com, afshanfarid2017@gmail.com, abdul.waheed@uaar.edu.pk

[2]*Department of Biosciences*
Comsats University Islamabad, Sahiwal Campus, Sahiwal
Islamabad, Pakistan
E-mail: ibrahim@cuisahiwal.edu.pk

[3]*Department of Food Safety and Quality Management*
Pir Mehr Ali Shah Arid Agriculture University
Shamsabad, Murree Road, Rawalpindi, Punjab, Pakistan
E-mails: sardarshahnawazkhankhan@gmail.com, aqsatalat9099@gmail.com, izzatfatima3248@gmail.com

[4]*Department of Biosciences*
Comsats University Islamabad, Main Campus
Islamabad, Pakistan
E-mail: tayyabaarzo04@gmail.com

[*]*Corresponding author*

Received: October 16, 2022 **Accepted: August 24, 2023**

Published: March 31, 2024

Abstract: *Pathogens of the Burkholderia genus are causing diseases in a diverse variety of hosts. After the discovery of T6SS, it was found to play a pivotal role in virulence and other pathogenicity factors in different pathogenic Burkholderia species. For this study, three strains of Burkholderia cenocepacia were selected from different ecological niches; J2315 from humans, MC0-3 from the rhizosphere of maize, and YG-3 from the Populus tree. The sequenced genomes were retrieved from PATRIC. It was found that B. cenocepacia J2315 and MC0-3 strains had only 1 cluster of T6SS in their genomes while the YG-3 strain had 3 clusters. The circular genomic map and phylogenetic tree suggested major differences in T6SS clusters 2 and 3 of the YG-3 strain from other clusters. From the results obtained in the study and reviewing the literature, it was concluded that all 3 strains harbor T6SS-1 type cluster that is involved in causing virulence in eukaryotic organisms and several bacterial species. This factor of causing virulence in the bacteria species might be helpful for B. cenocepacia strains J2315, MC0-3 and YG-3 in survival and niche adaptation.*

Keywords: *Burkholderia cenocepacia, Cross-kingdom, J2315, MC0-3, T6SS, YG-3.*

Introduction

Genus *Burkholderia* of subphylum β-proteobacteria comprises several gram-negative bacteria species that are adapted to various ecological niches and can occupy various diverse kinds of ecosystems including soil, fungi, plant, animal and even humans. Many of them are found to be causing diseases in their respective hosts [25].

The very first member of this genus was isolated in 1942 from a carnation that was showing root-rot and wilt symptoms. Initially, it was given the name, *Phytomonas caryophylli*, which was later renamed as *Pseudomonas caryophylli*. A few years later a species *Pseudomonas cepacia* (now known as *Burkholderia cepacia*) was then isolated and described from onions displaying sour skin rot by W. H. Burkholder in 1950. Later on, many researchers found *Burkholderia cepacia* and *Burkholderia cenocepacia* causing skin rot disease in onions [17, 25].

Some closely related cosmopolitan human pathogenic species of *Burkholderia genus* that form *Burkholderia cepacia* complex (BCC) cause different types of pulmonary infections, especially in immune-compromised and cystic fibrosis (CF) sufferers. Owing to their importance as human pathogens, profound research efforts and discussions have been analyzing the causes, effects and patterns of diseases caused by *Burkholderia* species [25]. Strains of *B. cepacia* complex are found to cause diseases like pneumonia, urinary tract infection, bacteremia and septic arthritis in people suffering from cystic fibrosis [43] and these strains are usually present in the samples extracted from bodies of intensive care units (ICU) and cystic fibrosis patients [46].

B. cenocepacia is one of the genomovars of BCC. Even though BCC was initially recognized as a pathogen in onions but now it is commonly familiar to us as an important opportunistic human pathogen [24]. Certain strains of BCC species being found in both plants and humans are providing a lead on cross-kingdom pathogenicity. *B. cenocepacia* strain J2315 was found among cystic fibrosis patients and is considered a potential epidemic pathogen because of its unique genetic adaptations [16]. On the other hand, *B. cenocepacia* strain MCO-3 was isolated from maize rhizosphere [5] and the YG-3 sample was first isolated from a populous tree in China [42].

Cross-kingdom pathogens use various disease-causing strategies to infect different unrelated hosts and their disease-causing strategies are of particular interest to researchers. Although different host species have distinctive physical barriers and defence responses still, certain plant and human pathogens have evolved disease factors successfully to exploit their respective hosts [19]. This study involved genome wide from identification of T6SS in 3 strains, i.e., J2315, MCO-3 and YG-3 by using published sequenced data from different databases as comprehensive analysis from sequenced and published genomic data of an organism can provide information about functions and evolutionary changes in a specific gene or gene family [8].

A specialist pathogen may evolve certain factors that can overcome the physical barriers and innate defenses in their hosts but a cross-kingdom pathogen would require a diverse library of genes and disease strategies to do so [19]. Protein secretion by microorganisms is one way through which they regulate their extracellular environment. In the case of pathogenic microorganisms, these secretions help in targeting and affecting host cells by exporting certain toxins and enzymes (having the ability to alter cellular function) or by assembling certain extracellular structures like pili to promote adhesion to host cell surfaces [32].

To deliver proteins and enzymes in the environment or host cell an evolution in bacterial surface structures is witnessed. Pathogens of gram-negative bacteria have not only evolved their surface structures but also evolved their secretion systems as evident through the use of six different extracellular protein secreting systems. These secretion systems are referred to as type I-VI secretion systems, or T1SS-T6SS. Their main function is to export proteins via multilayer cell envelope into the host target cells [32]. To date, nine secretion systems referred as type I-IX have been identified in both gram-positive and negative bacterial species [8].

These secretion systems are distinguished in part by their conserved structural components, substrate characteristics and the path they follow during the export process [32]. This is able to either necessitate a set of pathogenicity factors to enable; attachment to host cells, disease development and a universal disease strategy during which an equivalent suite of pathogenicity factors is employed for all hosts [19]. A recent addition to this growing collection of secretion devices was the type VI secretion system (T6SS). The T6SS arose as a fascinating topic, as its impact on the interaction with the host is a determinant for a successful infection [11]. Phytopathogenic *Burkholderia* organisms have different groups of T6SS's. They are involved in the delivery of effectors to eukaryotic and prokaryotic cells. Thus, T6SS matters a lot for a pathogen in order to compete with other adjoining microbes and for virulence to host plants [25].

T6SS are effector translocation apparatuses composed of different protein types [10]. The structure of T6SS is thought to be like that of T-Phage Tail. The type VI secretion apparatus is formed by a double-membrane-spanning structure that, like many other secretion systems, works by the one-step mechanism where bacterial cytoplasmic substrates are conveyed directly into a target cell or to the extracellular space. The partially/totally folded substrates are transported as in the chaperone-usher pathway. The minimal apparatus components called core are thought to be formed by 13 subunits, and in numerous cases, the core components also contain additional proteins [27].

It has been noted the T6SS machinery can translocate effectors in the following ways: (i) bound to structural components as specialized VgrG effectors, or (ii) through non-covalent interaction with any of the components of the core (cargo effectors). Since in these cases, the effector being translocated has to be associated with Hcp-VgrG-PAAR which are components of the expelled structure, it is thought that several effectors associated with this puncturing protein are delivered at once in one lethal shot inside the target cell. It is thought this delivery in one lethal shot allows a single event of T6SS sheath contraction (one toxic payload of effectors) by these T6SS [27].

Genome wide identification, characterization and evolutionary analysis of T6SS; in *B. cenocepacia* strains J2315, MCO-3 and YG-3 have been addressed here. In the study, *B. cenocepacia* J2315 was used as a reference strain. This research involves the prediction and characterization of T6SS along with genome wide comparative analysis and evolutionary analysis of T6SS in those strains of *B. cenocepacia*.

Materials and methods

Acquisition of Burkholderia genomic data

The assembled genome sequence data of *B. cenocepacia* strains, i.e., *B. cenocepacia* J2315, *B. cenocepacia* MC0-3 and *B. cenocepacia* YG-3 were retrieved from Pathosystems Resource Integration Center (PATRIC) [44].

Genome annotation and analysis

The genome sequences retrieved from PATRIC [44] of *B. cenocepacia* strains were first refined and then annotated. The annotation was conducted using the Rapid Annotation using Subsystem Technology (RAST) Server Version 2.0 [30] with default parameters. After annotation was completed, it was then browsed in SEED Viewer [30]. The annotated genomes obtained with the aid of SEED Viewer of the RAST Server were used to analyze coding sequences (CDS), GC contents, genome sizes and general features of the strains.

Prediction of T6SS clusters

Annotated genomes of three strains were used to predict T6SS clusters. For this purpose, we used a reference strain that was *B. cenocepacia* J2315 to examine the T6SS in the genomes of, *B. cenocepacia* MC0-3 and *B. cenocepacia* YG-3 strains. The gene clusters were predicted in the genome browser of RAST Server [30] and were later on retrieved in table form.

Comparative analysis of T6SS gene cluster

For comparative analysis of the T6SS gene cluster Basic Local Alignment Search Tool (BLAST) [38] was done. Nowadays it is one of the most commonly used biotechnological tools for comparison of different sequence information. We took gene sequences of different components of T6SS from *B. cenocepacia* J2315 as a reference. We did a comparative analysis of T6SS gene clusters of the other two *B. cenocepacia* strains using CGView Server [14].

Protein family, function and domain search

Protein families are groups of proteins sharing common evolutionary origins reflected by their functions and similarities in their structures or sequences while protein domains are conserved parts of proteins that can evolve, function and exist independently. For comparing and associating protein domains and protein families, NCBI's Conserved Domain Database (CDD) [26] was used.

Phylogenetic study of T6SS components

To elucidate the position of T6SS clusters in *B. cenocepacia* strains MEGA X software (Version 10.0, Pennsylvania State University, Pennsylvania, United States) [20] was used to construct the phylogenetic tree. Firstly, all the DNA sequences of T6SS clusters were introduced in MEGA X software then they were aligned by ClustalW. Then those aligned sequences were used to construct a phylogenetic tree using Maximum Likelihood Method with default parameters.

Results and discussion

This study revolves around the genome wide identification and characterization of T6SS components in the *Burkholderia* species. The T6SS components were identified in the *B. cenocepacia* J2315, *B. cenocepacia* MC0-3 and *B. cenocepacia* YG-3 where *B. cenocepacia* J2315 was taken as the reference genome to make comparative analysis with the other two strains.

Genome and annotation analysis

The annotated genomes of all three strains gave some significant values like genome size, number of coding sequences, GC content, etc., that are mentioned in Table 1.

Table 1. Genome wide general features of *B. cenocepacia* strain J2315, MC0-3 and YG-3

Genome	*B. cenocepacia* J2315	*B. cenocepacia* MC0-3	*B. cenocepacia* YG-3
Genome size, (bp)	8055782	7971389	8036463
GC content, (%)	66.9	66.6	66.813736
No. of coding sequences	7534	7330	7966
No. of features	7116	7008	7421

bp – base pairs

It was found that *B. cenocepacia* J2315 has the largest genome with a genome size of 8055782 bp as compared to the other two. But despite of larger genome, the numbers of coding sequences in *B. cenocepacia* J2315 were lesser as compared to *B. cenocepacia* YG-3. A slight variation in the GC content of 3 strains was also observed.

The results obtained from the annotation of genomes provided significant information regarding many aspects of a genome. From that data, it was found that *B. cenocepacia* J2315 had the largest genome. Despite of larger genome, the coding sequences in *B. cenocepacia* J2315 were lesser as compared to *B. cenocepacia* YG-3. Genes are represented by coding sequences. Hundreds of bp form a CDS. These CDS represent specific genes that translate into mRNA and then into a specific protein that further performs the functions [36]. So, it would not be wrong if it is said that, the number of coding sequences has a positive correlation with the number of features as we can see in Table 1 that *B. cenocepacia* YG-3 has the largest number of coding sequences and so the features. Also, there was a slight variation in the GC content of 3 strains that might have occurred because of replication and repair mechanisms. Usually, replication and repair in genomes are influenced by environmental factors [47]. However, GC content in related bacterial strains is found to be almost the same [29] as in this case. Differences in genomic content in different genomes might be due to the effect of various factors; some of which involve an environment and diet [21].

Gene prediction analysis

In comparative genomics, the prediction of genes is considered a vital phenomenon. For T6SS gene prediction purposes a table containing information regarding all the genes of a particular strain was exported from the genome browser of the RAST Server for all 3 strains. Reviewing data from that table it was concluded that *B. cenocepacia* J2315 (Table 2) and MC0-3 (Table 3, coloured boxes indicate T6SS gene cluster) both had 1 cluster of T6SS genes, while *B. cenocepacia* YG-3 (Tables 4-6) had 3 clusters of T6SS. In the tables colored boxes indicate the T6SS gene cluster. Thirteen core components of T6SS [6] were used for the examination of all clusters. T6SS cluster of *B. cenocepacia* J2315 had 12 out of 13 core components.

The same was the case with the T6SS cluster of *B. cenocepacia* MC0-3. In both strains, the VgrG component was not present in the cluster. In the case of the *B. cenocepacia* YG-3, one cluster was found to have all 13 core components while the other two clusters, i.e., Clusters 2 and 3 contain 10 and 11 core components, respectively, as mentioned in Table 7. Though VgrG's are not present in T6SS clusters of *B. cenocepacia* J2315 and MC0-3 but they are found as orphan components scattered in the genome.

Table 2. *B. cenocepacia* J2315 gene cluster

Feature ID	Type	Start	Stop	Length, (bp)	RAST automatic annotation	T6SS COG's
fig|6666666.667837.peg.359	CDS	361885	361085	801	ABC transporter, substrate-binding protein	
fig|6666666.667837.peg.360	CDS	362167	362484	318	hypothetical protein	
fig|6666666.667837.peg.361	CDS	362566	362901	336	FIG00848710: hypothetical protein	
fig|6666666.667837.peg.362	CDS	363797	363015	783	T6SS outer membrane component TssL (ImpK/VasF)	COG3455
fig|6666666.667837.peg.363	CDS	365140	363794	1347	T6SS component TssK (ImpJ/VasE)	COG3522
fig|6666666.667837.peg.364	CDS	365845	365246	600	T6SS secretion lipoprotein TssJ (VasD)	COG3521
fig|6666666.667837.peg.365	CDS	366232	366858	627	FIG140336: TPR domain protein	
fig|6666666.667837.peg.366	CDS	366905	367420	516	T6SS component TssB (ImpB/VipA)	COG3516
fig|6666666.667837.peg.367	CDS	367436	368926	1491	T6SS component TssC (ImpC/VipB)	COG3517
fig|6666666.667837.peg.368	CDS	368997	369500	504	T6SS component Hcp	COG3157
fig|6666666.667837.peg.369	CDS	369563	370048	486	T6SS lysozyme-like component TssE	COG3518
fig|6666666.667837.peg.370	CDS	370125	371960	1836	T6SS component TssF (ImpG/VasA)	COG3519
fig|6666666.667837.peg.371	CDS	371924	373024	1101	T6SS component TssG (ImpH/VasB)	COG3520
fig|6666666.667837.peg.372	CDS	373066	375735	2670	T6SS AAA+ chaperone ClpV (TssH)	COG0542
fig|6666666.667837.peg.373	CDS	375780	376901	1122	T6SS component TssA (ImpA)	COG3515
fig|6666666.667837.peg.374	CDS	377945	376995	951	T6SS peptidoglycan-binding component TagN	
fig|6666666.667837.peg.375	CDS	378939	377950	990	T6SS associated component TagF (ImpM)	COG3913
fig|6666666.667837.peg.376	CDS	382880	378936	3945	T6SS component TssM (IcmF/VasK)	COG3523
fig|6666666.667837.peg.377	CDS	382940	383053	114	hypothetical protein	
fig|6666666.667837.peg.378	CDS	383793	383182	612	T6SS-associated peptidoglycan hydrolase TagX	
fig|6666666.667837.peg.379	CDS	384149	385141	993	Tail fiber protein	
fig|6666666.667837.peg.380	CDS	385353	386381	1029	Magnesium and cobalt transport protein CorA	
fig|6666666.667837.peg.381	CDS	386561	386409	153	FIG026426: Hypothetical protein	

Table 3. *B. cenocepacia* MC0-3 T6SS gene cluster

Feature ID	Type	Start	Stop	Length, (bp)	RAST automatic annotation	T6SS COG's
fig|6666666.667854.peg.472	CDS	477776	477916	141	hypothetical protein	
fig|6666666.667854.peg.473	CDS	477992	478438	447	Rhs-family protein	
fig|6666666.667854.peg.474	CDS	478662	478979	318	hypothetical protein	
fig|6666666.667854.peg.475	CDS	479850	479068	783	T6SS outer membrane component TssL (ImpK/VasF)	COG3455
fig|6666666.667854.peg.476	CDS	481193	479847	1347	T6SS component TssK (ImpJ/VasE)	COG3522
fig|6666666.667854.peg.477	CDS	481895	481296	600	T6SS secretion lipoprotein TssJ (VasD)	COG3521
fig|6666666.667854.peg.478	CDS	482280	482915	636	FIG140336: TPR domain protein	
fig|6666666.667854.peg.479	CDS	482962	483477	516	T6SS component TssB (ImpB/VipA)	COG3516
fig|6666666.667854.peg.480	CDS	483493	484983	1491	T6SS component TssC (ImpC/VipB)	COG3517
fig|6666666.667854.peg.481	CDS	485054	485557	504	T6SS component Hcp	COG3157
fig|6666666.667854.peg.482	CDS	485620	486105	486	T6SS lysozyme-like component TssE	COG3518
fig|6666666.667854.peg.483	CDS	486182	488017	1836	T6SS component TssF (ImpG/VasA)	COG3519
fig|6666666.667854.peg.484	CDS	487981	489081	1101	T6SS component TssG (ImpH/VasB)	COG3520
fig|6666666.667854.peg.485	CDS	489123	491792	2670	T6SS AAA+ chaperone ClpV (TssH)	COG0542
fig|6666666.667854.peg.486	CDS	491837	492958	1122	T6SS component TssA (ImpA)	COG3515
fig|6666666.667854.peg.487	CDS	492963	493235	273	FIG00457408: hypothetical protein	
fig|6666666.667854.peg.488	CDS	493275	494105	831	T6SS component Hcp	COG3157
fig|6666666.667854.peg.489	CDS	494174	494347	174	hypothetical protein	
fig|6666666.667854.peg.490	CDS	494472	495260	789	T6SS component Hcp	COG3157
fig|6666666.667854.peg.491	CDS	495247	495429	183	FIG00464863: hypothetical protein	
fig|6666666.667854.peg.492	CDS	496498	495548	951	T6SS peptidoglycan-binding component TagN	
fig|6666666.667854.peg.493	CDS	497492	496503	990	T6SS associated component TagF (ImpM)	COG3913
fig|6666666.667854.peg.494	CDS	501433	497489	3945	T6SS component TssM (IcmF/VasK)	COG3523
fig|6666666.667854.peg.495	CDS	501493	501606	114	hypothetical protein	
fig|6666666.667854.peg.496	CDS	502580	501735	846	T6SS-associated peptidoglycan hydrolase TagX	
fig|6666666.667854.peg.497	CDS	502702	503694	993	Tail fiber protein	
fig|6666666.667854.peg.498	CDS	503870	504934	1065	Magnesium and cobalt transport protein CorA	
fig|6666666.667854.peg.499	CDS	505114	504962	153	FIG026426: Hypothetical protein	

Table 4. *B. cenocepacia* YG-3 T6SS gene Cluster 1

Feature ID	Type	Start	Stop	Length, (bp)	RAST automatic annotation	T6SS COG's
fig|6666666.667850.peg.413	CDS	431335	431916	582	hypothetical protein	
fig|6666666.667850.peg.414	CDS	431930	432388	459	hypothetical protein	
fig|6666666.667850.peg.415	CDS	432411	432752	342	FIG00462163: hypothetical protein	
fig|6666666.667850.peg.416	CDS	433648	432866	783	T6SS outer membrane component TssL (ImpK/VasF)	COG3455
fig|6666666.667850.peg.417	CDS	434991	433645	1347	T6SS component TssK (ImpJ/VasE)	COG3522
fig|6666666.667850.peg.418	CDS	435693	435094	600	T6SS secretion lipoprotein TssJ (VasD)	COG3521
fig|6666666.667850.peg.419	CDS	436079	436726	648	FIG140336: TPR domain protein	
fig|6666666.667850.peg.420	CDS	436773	437288	516	T6SS component TssB (ImpB/VipA)	COG3516
fig|6666666.667850.peg.421	CDS	437304	438794	1491	T6SS component TssC (ImpC/VipB)	COG3517
fig|6666666.667850.peg.422	CDS	438865	439368	504	T6SS component Hcp	COG3157
fig|6666666.667850.peg.423	CDS	439431	439916	486	T6SS lysozyme-like component TssE	COG3518
fig|6666666.667850.peg.424	CDS	439993	441828	1836	T6SS component TssF (ImpG/VasA)	COG3519
fig|6666666.667850.peg.425	CDS	441792	442892	1101	T6SS component TssG (ImpH/VasB)	COG3520
fig|6666666.667850.peg.426	CDS	442932	445601	2670	T6SS AAA+ chaperone ClpV (TssH)	COG0542
fig|6666666.667850.peg.427	CDS	445657	446778	1122	T6SS component TssA (ImpA)	COG3515
fig|6666666.667850.peg.428	CDS	446845	449406	2562	VgrG protein	COG3501
fig|6666666.667850.peg.429	CDS	449406	450446	1041	hypothetical protein	
fig|6666666.667850.peg.430	CDS	450460	451197	738	T6SS component Hcp	COG3157
fig|6666666.667850.peg.431	CDS	451381	452226	846	T6SS component Hcp	COG3157
fig|6666666.667850.peg.432	CDS	452219	454519	2301	T6SS component Hcp	COG3157
fig|6666666.667850.peg.433	CDS	454543	455445	903	T6SS component Hcp	COG3157
fig|6666666.667850.peg.434	CDS	455807	455598	210	T6SS peptidoglycan-binding component TagN	
fig|6666666.667850.peg.435	CDS	455961	456434	474	hypothetical protein	
fig|6666666.667850.peg.436	CDS	456431	457087	657	hypothetical protein	
fig|6666666.667850.peg.437	CDS	457337	457218	120	hypothetical protein	
fig|6666666.667850.peg.438	CDS	458287	457334	954	T6SS peptidoglycan-binding component TagN	
fig|6666666.667850.peg.439	CDS	459281	458292	990	T6SS associated component TagF (ImpM)	COG3913
fig|6666666.667850.peg.440	CDS	463210	459278	3933	T6SS component TssM (IcmF/VasK)	COG3523
fig|6666666.667850.peg.441	CDS	463270	463383	114	hypothetical protein	
fig|6666666.667850.peg.442	CDS	464357	463512	846	T6SS-associated peptidoglycan hydrolase TagX	
fig|6666666.667850.peg.443	CDS	464479	465465	987	Tail fiber protein	
fig|6666666.667850.peg.444	CDS	465672	466700	1029	Magnesium and cobalt transport protein CorA	
fig|6666666.667850.peg.445	CDS	466919	466767	153	FIG026426: Hypothetical protein	

Table 5. *B. cenocepacia* YG-3 T6SS gene Cluster 2

Feature ID	Type	Start	Stop	Length, (bp)	RAST automatic annotation	T6SS COG's
fig|6666666.667850.peg.1971	CDS	2044549	2044007	543	hypothetical protein	
fig|6666666.667850.peg.1972	CDS	2045599	2044673	927	2-dehydro-3-deoxygluconokinase	
fig|6666666.667850.peg.1973	CDS	2046964	2045603	1362	Ribosomal protein S12p Asp88 methylthiotransferase	
fig|6666666.667850.peg.1974	CDS	2047643	2047386	258	T6SS PAAR-repeat protein	
fig|6666666.667850.peg.1975	CDS	2048582	2047677	906	hypothetical protein	
fig|6666666.667850.peg.1976	CDS	2049441	2048587	855	hypothetical protein	
fig|6666666.667850.peg.1977	CDS	2052445	2049446	3000	hypothetical protein	
fig|6666666.667850.peg.1978	CDS	2053388	2052438	951	hypothetical protein	
fig|6666666.667850.peg.1979	CDS	2055511	2053385	2127	VgrG protein	COG3501
fig|6666666.667850.peg.1980	CDS	2056428	2055532	897	hypothetical protein	
fig|6666666.667850.peg.1981	CDS	2057767	2056430	1338	T6SS component TssK (ImpJ/VasE)	COG3522
fig|6666666.667850.peg.1982	CDS	2058897	2057764	1134	hypothetical protein	
fig|6666666.667850.peg.1983	CDS	2060237	2058894	1344	T6SS forkhead associated domain protein ImpI/VasC	COG3456
fig|6666666.667850.peg.1984	CDS	2061253	2060234	1020	T6SS component TssG (ImpH/VasB)	COG3520
fig|6666666.667850.peg.1985	CDS	2062983	2061217	1767	T6SS component TssF (ImpG/VasA)	COG3519
fig|6666666.667850.peg.1986	CDS	2063431	2063003	429	Uncharacterized protein similar to VCA0109	
fig|6666666.667850.peg.1987	CDS	2064928	2063447	1482	T6SS component TssC (ImpC/VipB)	COG3517
fig|6666666.667850.peg.1988	CDS	2065472	2064969	504	T6SS component TssB (ImpB/VipA)	COG3516
fig|6666666.667850.peg.1989	CDS	2065950	2068517	2568	T6SS AAA+ chaperone ClpV (TssH)	COG0542
fig|6666666.667850.peg.1990	CDS	2068530	2070092	1563	hypothetical protein	
fig|6666666.667850.peg.1991	CDS	2070089	2070793	705	hypothetical protein	
fig|6666666.667850.peg.1992	CDS	2070790	2072220	1431	T6SS component TssA (ImpA)	COG3515
fig|6666666.667850.peg.1993	CDS	2072236	2075997	3762	T6SS component TssM (IcmF/VasK)	COG3523
fig|6666666.667850.peg.1994	CDS	2076083	2076835	753	hypothetical protein	
fig|6666666.667850.peg.1995	CDS	2077125	2078075	951	Mobile element protein	
fig|6666666.667850.peg.1996	CDS	2078884	2078183	702	hypothetical protein	
fig|6666666.667850.peg.1997	CDS	2081826	2079148	2679	hypothetical protein	
fig|6666666.667850.peg.1998	CDS	2083928	2081841	2088	VgrG protein	COG3501
fig|6666666.667850.peg.1999	CDS	2084538	2084020	519	T6SS component Hcp	COG3157
fig|6666666.667850.peg.2000	CDS	2084806	2084660	147	hypothetical protein	
fig|6666666.667850.peg.2001	CDS	2086387	2085104	1284	hypothetical protein	
fig|6666666.667850.peg.2002	CDS	2088119	2086377	1743	hypothetical protein	
fig|6666666.667850.peg.2003	CDS	2088146	2088382	237	hypothetical protein	

Table 6: *B. cenocepacia* YG-3 T6SS gene Cluster 3

Feature ID	Type	Start	Stop	Length, (bp)	RAST automatic annotation	T6SS COG's
fig|6666666.667850.peg.6046	CDS	1249496	1248666	831	ABC-type phosphate, periplasmic component	
fig|6666666.667850.peg.6047	CDS	1249741	1250766	1026	Fatty acid desaturase	
fig|6666666.667850.peg.6048	CDS	1250941	1252149	1209	Benzoate transport protein	
fig|6666666.667850.peg.6049	CDS	1252829	1252302	528	Type VI secretion lipoprotein/VasD	COG3521
fig|6666666.667850.peg.6050	CDS	1253952	1252843	1110	T6SS component TssG (ImpH/VasB)	COG3520
fig|6666666.667850.peg.6051	CDS	1255763	1253949	1815	T6SS component TssF (ImpG/VasA)	COG3519
fig|6666666.667850.peg.6052	CDS	1257353	1255773	1581	T6SS component TssA (ImpA)	COG3515
fig|6666666.667850.peg.6053	CDS	1261192	1257362	3831	hypothetical protein	
fig|6666666.667850.peg.6054	CDS	1261522	1261232	291	FIG00455915: hypothetical protein	
fig|6666666.667850.peg.6055	CDS	1262850	1261600	1251	hypothetical protein	
fig|6666666.667850.peg.6056	CDS	1263989	1263090	900	hypothetical protein	
fig|6666666.667850.peg.6057	CDS	1265097	1264195	903	hypothetical protein	
fig|6666666.667850.peg.6058	CDS	1265210	1265094	117	hypothetical protein	
fig|6666666.667850.peg.6059	CDS	1266202	1265303	900	hypothetical protein	
fig|6666666.667850.peg.6060	CDS	1267989	1266199	1791	Putative transmembrane protein	
fig|6666666.667850.peg.6061	CDS	1270488	1268020	2469	VgrG protein	COG3501
fig|6666666.667850.peg.6062	CDS	1272181	1270568	1614	T6SS component TssC (ImpC/VipB)	COG3517
fig|6666666.667850.peg.6063	CDS	1272686	1272174	513	T6SS component TssB (ImpB/VipA)	COG3516
fig|6666666.667850.peg.6064	CDS	1272840	1272700	141	hypothetical protein	
fig|6666666.667850.peg.6065	CDS	1273284	1274636	1353	T6SS component TssK (ImpJ/VasE)	COG3522
fig|6666666.667850.peg.6066	CDS	1274633	1275364	732	hypothetical protein	
fig|6666666.667850.peg.6067	CDS	1275448	1277082	1635	Outer membrane protein	
fig|6666666.667850.peg.6068	CDS	1277191	1277676	486	T6SS component Hcp	COG3157
fig|6666666.667850.peg.6069	CDS	1277813	1280536	2724	T6SS AAA + chaperone ClpV (TssH)	COG0542
fig|6666666.667850.peg.6070	CDS	1280583	1280957	375	hypothetical protein	
fig|6666666.667850.peg.6071	CDS	1280981	1281406	426	T6SS lysozyme-like component TssE	COG3518
fig|6666666.667850.peg.6072	CDS	1282383	1281490	894	hypothetical protein	
fig|6666666.667850.peg.6073	CDS	1283441	1282518	924	Transcriptional regulator, AraC family	

B. cenocepacia strain J2315 and MC0-3 were found to have 1 cluster of T6SS in their genomes while YG-3 had 3 clusters. In most of the cases, *B. cenocepacia* strains are found to contain only 1 T6SS cluster that is T6SS-1 [2, 35] as was observed in the case of *B. cenocepacia* J2315 and MC0-3 strains. T6SS-1 is found to be involved mostly in intercellular interactions within a biofilm [2] and is an antibacterial effector that helps in the survival of a bacterium possessing the secretion system in a competition by causing toxicity that kills the competitor [12]. T6SS-1 has also been found to have an anti-eukaryotic function so it can be said as a key virulence-causing cluster [40]. Across the genus *Burkholderia* eight distinct types of T6SS have been found. These types vary based on the presence or absence of certain core and accessory components [35]. As in the T6SS-1 type cluster, all other core components are present, exceptions are mostly seen in the case of ClpV and VgrG [12]. Along with serving the purpose of virulence, T6SS in *B. cenocepacia* is found to help it survive when competition is tough in the rhizosphere and outer soil [35].

Table 7. T6SS core components in *B. cenocepacia* strains J2315, MC0-3 and YG-3 clusters

T6SS core components	***B. cenocepacia* J2315**	***B. cenocepacia* MC0-3**	***B. cenocepacia* YG-3**		
			Cluster 1	**Cluster 2**	**Cluster 3**
ImpA/TssA	Present	Present	Present	Present	Present
ImpB/VipA/TssB	Present	Present	Present	Present	Present
ImpC/VipB/TssC	Present	Present	Present	Present	Present
Hcp/TssD	Present	Present	Present	Present	Present
TssE	Present	Present	Present	Absent	Present
ImpG/VasA/TssF	Present	Present	Present	Present	Present
ImpH/VasB/TssG	Present	Present	Present	Present	Present
ClpV/TssH	Present	Present	Present	Present	Present
VgrG/TssI	Absent	Absent	Present	Present	Present
VasD/TssJ	Present	Present	Present	Absent	Present
ImpJ/VasE/TssK	Present	Present	Present	Present	Present
ImpK/VasF/DotU/TssL	Present	Present	Present	Absent	Absent
IcmF/VasK/TssM	Present	Present	Present	Present	Absent

In Cluster 2 of *B. cenocepacia* YG-3 TssE, TssJ and TssL components were absent while in Cluster 3 TssL and TssM components were absent as can be seen in Table 2. Out of all TssH, TssL and TssM were found to be involved in causing virulence [34]. TssL, TssM along with TssJ were also found to be involved in the formation of membrane anchoring complex [50].

Evaluation of gene architecture and gene organization of T6SS clusters

The T6SS cluster sizes were calculated and it was found that *B. cenocepacia* J2315 had a T6SS cluster of 19.7 kilobase pairs (kb) size. T6SS cluster of *B. cenocepacia* MC0-3 was of 22.2 kb size. *B. cenocepacia* YG-3 was found to have three clusters of T6SS and these clusters were of sizes 29.8 kb, 35.9 kb and 28.4 kb respectively. Out of all, Cluster 2 of *B. cenocepacia* YG-3 was found to have a maximum size. The gene organization of T6SS clusters in *B. cenocepacia* J2315, MC0-3 and YG-3 is presented in Fig. 1.

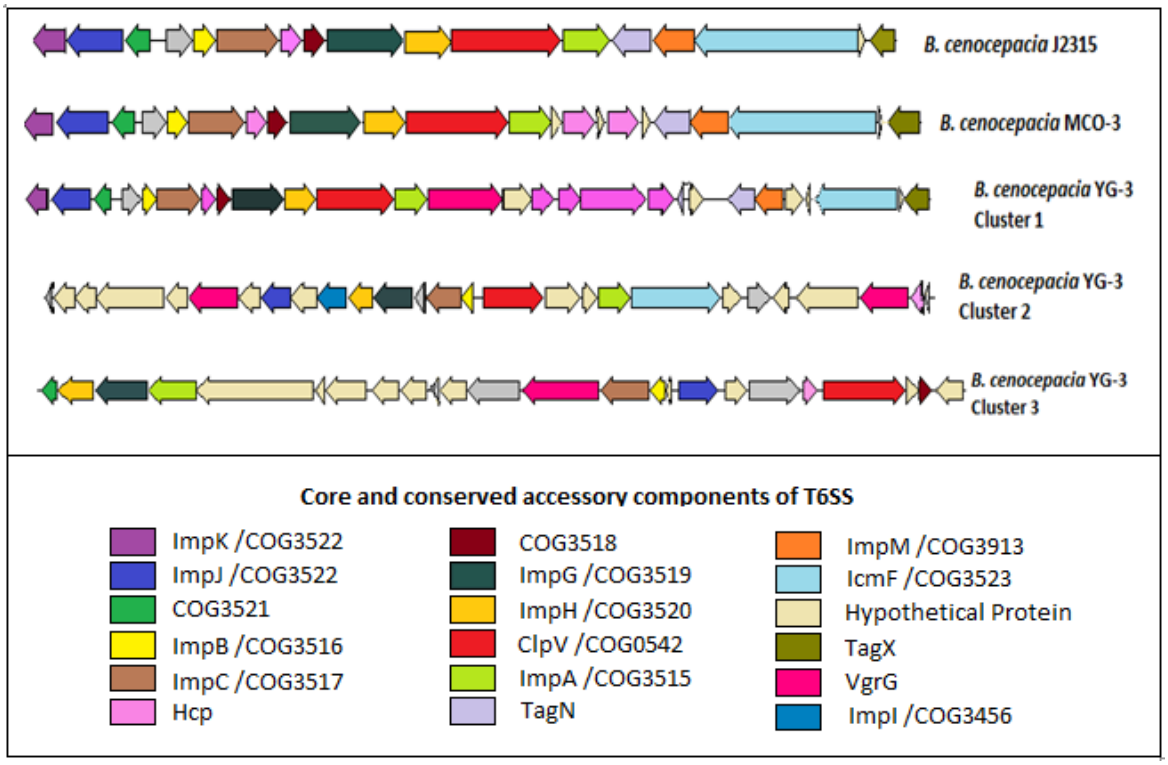

Fig. 1 Gene organization of T6SS gene clusters of *B. cenocepacia* strains J2315, MC0-3 and YG-3 obtained from SEED Viewer of RAST Server. Different colours represent different components. The size of the arrow represents the size of the T6SS component in base pairs. Arrows also represent the direction and location of gene transcription.

Out of all clusters Cluster 2 of the YG-3 strain was found to have the maximum size though it lacked a maximum number of core components out of all clusters. T6SS clusters with PAAR proteins were found to have extended cluster sizes [4] as can be seen in the case of Cluster 2 of the YG-3 strain as it was the only strain that possessed PAAR protein.

Circular genome study of B. cenocepacia strains and T6SS clusters

Annotated genome sequences of *B. cenocepacia* strains were downloaded in BioEdit Software [15], from RAST Server in Genbank format. Using these sequences, a circular genome of *B. cenocepacia* J2315 was generated in the CGView Server. The CGView Server is a tool for comparative genomic studies of circular genomes including bacterial genomes. After the generation of *B. cenocepacia* J2315 circular genome, annotated genome sequences of MC0-3 and YG-3 were uploaded on the server to BLAST them against the reference strain, i.e., J2315. A circular diagram representing genomes of *B. cenocepacia* J2315, MC0-3 and YG-3 along with contigs, GC content and GC skew is retrieved from the server (Fig. 2).

A circular genome map of 3 strains was retrieved from the CGView server. In Fig. 2 the circle representing GC content was observed to be showing discontinuous variation. The peaks heading outwards show GC content higher as compared to the genome average while peaks pointing toward the inside of the circle represent lower GC content in comparison to the genome average [48]. Below GC content GC skew is present. GC skew divides the genome into 2 regions one is termed GC skew+ which represents the excess of guanine over cytosine, while the other is termed as GC skew- which represents the excess of cytosine over guanine. The one with GC skew+ is termed as the leading strand and the other with GC skew- is termed as the lagging strand [3].

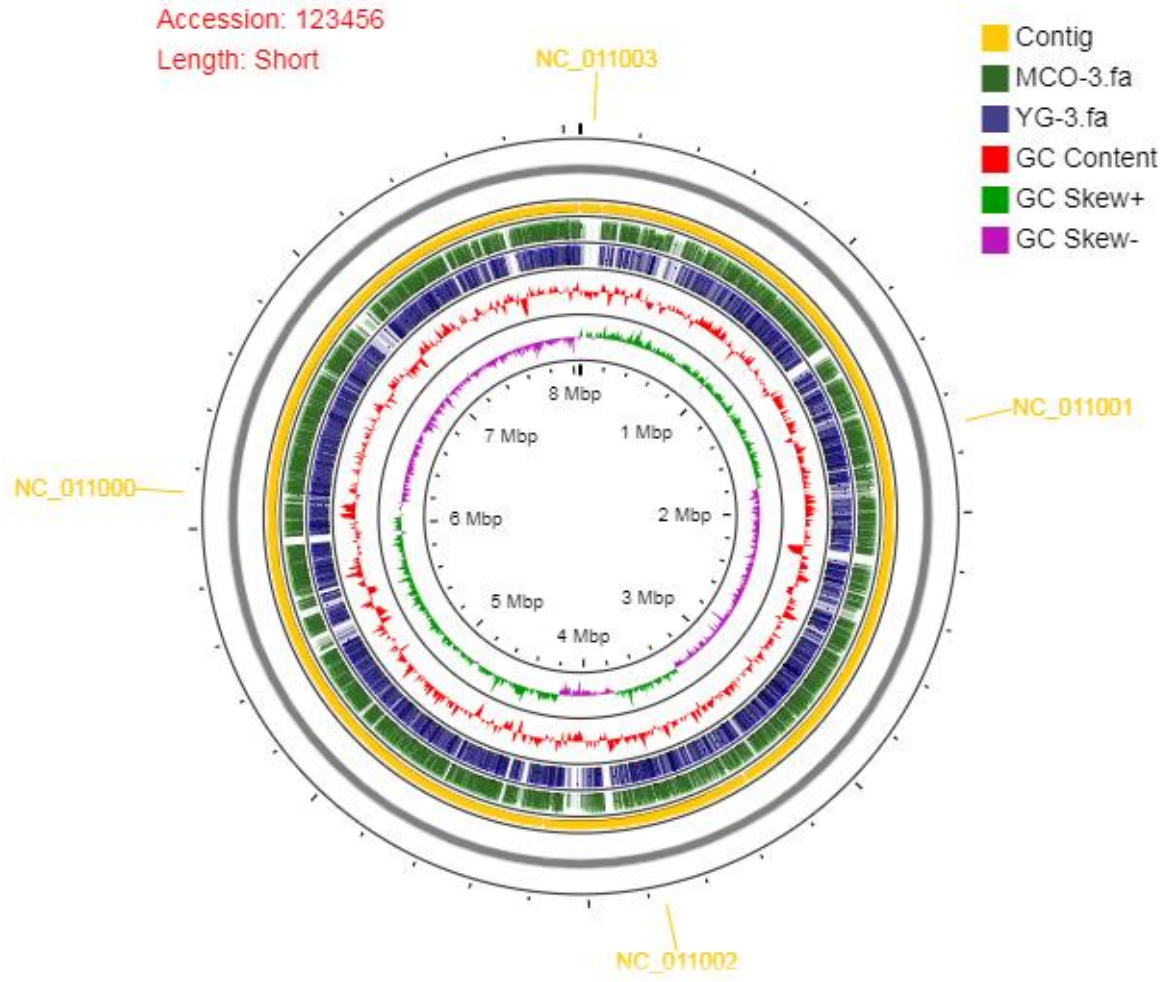

Fig. 2 Circular genome map of *B. cenocepacia* J2315 built-in CGView Server showing results of BLAST with MC0-3 and YG-3 genomes along with GC content and GC skew of the J2315 strain

Another circular genome map showing comparative results of J2315 with MC0-3 and YG-3 was obtained from RAST with the help of sequence-based comparison (Fig. 3). In this type of comparison, reference strain (J2315) was first selected and then comparison strains (MC0-3 and YG-3) were selected and comparison was done with default parameters. The map represents the percent protein identity in 3 strains.

DNA sequences of every component of each cluster of T6SS were retrieved separately from SEED Viewer of RAST Server in FASTA format. DNA sequences of *B. cenocepacia* J2315 were uploaded on CGView Server as reference sequences. After construction of the map from the reference sequence, sequences of the other 4 clusters of MC0-3 and YG-3 were BLAST on it. The BLAST results are shown in Fig. 4, where the outermost colored circle represents the T6SS cluster of the J2315 strain. Below that T6SS cluster of the MC0-3 strain is present. The inner 3 circles represent T6SS Clusters 1, 2 and 3 of strain YG-3.

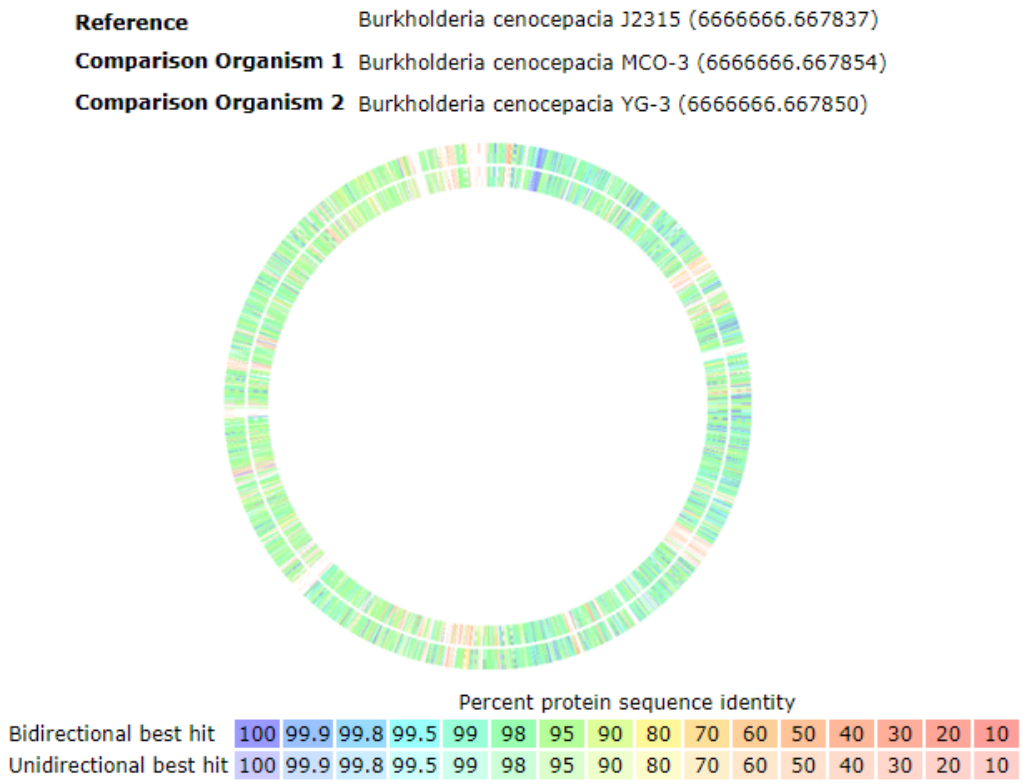

Fig. 3 Comparative genome map of *B. cenocepacia* J2315 from MC0-3 and YG-3 obtained from RAST Server representing percent protein identity of comparison strains, i.e., MC0-3 and YG-3. The map was generated in the RAST Server.

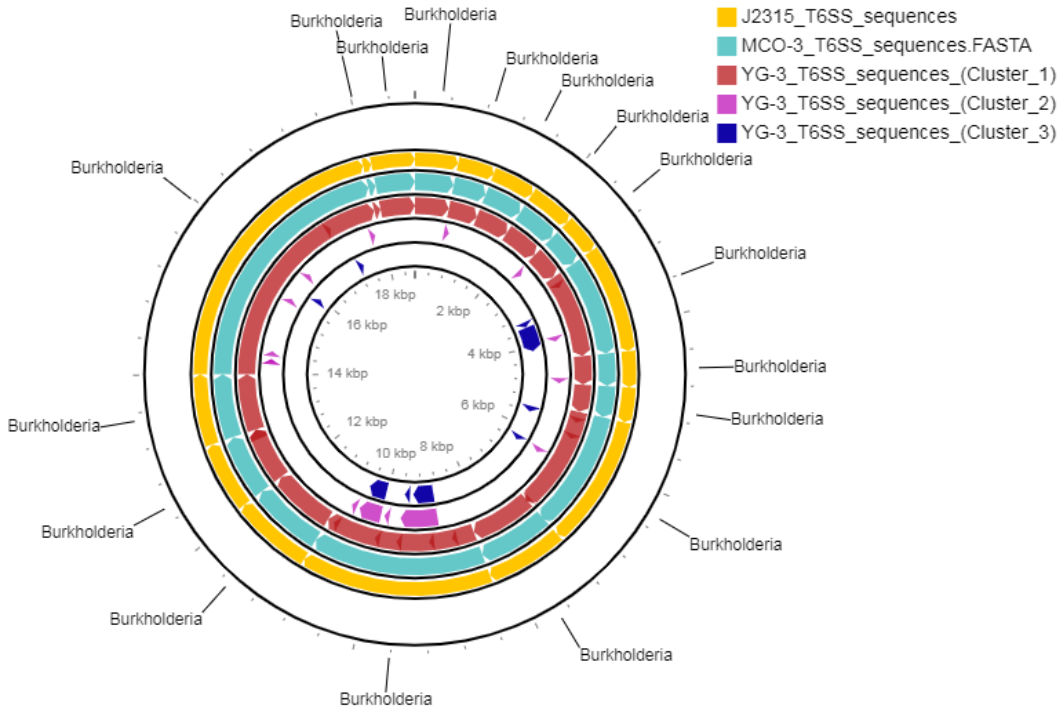

Fig. 4 Circular representation of 5 clusters of T6SS from *B. cenocepacia* strain J2315, MC0-3 and YG-3 built-in CGView Server

Protein family, domain and function analysis

Protein family and domain analysis of components of T6SS was done using the CDD of NCBI. For those purposes, protein sequences of every component of each cluster were retrieved from the RAST Server. Sequences are entered in CDD one by one and results are introduced in the table. Most of them were found to hit only one domain while some did not hit any. Only ImpA of YG-3 cluster 3 was found to hit 2 domains, i.e., T6SS_VasJ (pfam16989) and ImpA_N (pfam06812). Domain hits of ImpB and ImpG in all 5 clusters were the same that were T6SS_VipA (pfam05591) and VI_chp_6 (TIGR03359) respectively. All the results of *B. cenocepacia* J2315, MC0-3 and YG-3 Cluster 1 were the same. Exceptions were observed mostly in Cluster 2 and Cluster 3 (Table 8).

Table 8. Protein family and domains of T6SS components of all T6SS clusters of *B. cenocepacia* J2315, MC0-3 and YG-3

T6SS Comp		***B. cenocepacia* J2315**	***B. cenocepacia* MC0-3**	***B. cenocepacia* YG-3**		
				Cluster 1	**Cluster 2**	**Cluster 3**
ImpA	Domain name	VI_chp_8 (TIGR03363)	VI_chp_8 (TIGR03363)	VI_chp_8 (TIGR03363)	ImpA_N (pfam06812)	T6SS_VasJ (pfam16989), ImpA_N (pfam06812)
	Protein family	ImpA_N (cl19907)	ImpA_N (cl19907)	ImpA_N (cl19907)	ImpA_N (cl19907)	T6SS_VasJ (cl11880), ImpA_N (cl19907)
ImpB	Domain name	T6SS_VipA (pfam05591)	T6SS_VipA (pfam05591)	T6SS_VipA (pfam05591)	T6SS_VipA (pfam05591)	T6SS_VipA (pfam05591)
	Protein family	T6SS_VipA (cl01402)	T6SS_VipA (cl01402)	T6SS_VipA (cl01402)	T6SS_VipA (cl01402)	T6SS_VipA (cl01402)
ImpC	Domain name	VI_chp_2 (TIGR03355)	VI_chp_2 (TIGR03355)	VI_chp_2 (TIGR03355)	VI_chp_2 (TIGR03355)	Not found
	Protein family	VipB (cl05484)	VipB (cl05484)	VipB (cl05484)	VipB(cl05484)	VipB(cl05484)
ImpG	Domain name	VI_chp_6 (TIGR03359)	VI_chp_6 (TIGR03359)	VI_chp_6 (TIGR03359)	VI_chp_6 (TIGR03359)	VI_chp_6 (TIGR03359)
	Protein family	T6SS_TssF (cl15462)	T6SS_TssF (cl15462)	T6SS_TssF (cl15462)	T6SS_TssF (cl15462)	T6SS_TssF (cl15462)
ImpH	Domain name	VI_chp_1 (TIGR03347)	VI_chp_1 (TIGR03347)	VI_chp_1 (TIGR03347)	VI_chp_1 (TIGR03347)	T6SS_TssG (pfam06996)
	Protein family	T6SS_TssG (cl01404)	T6SS_TssG (cl01404)	T6SS_TssG (cl01404)	T6SS_TssG (cl01404)	T6SS_TssG (cl01404)
ImpI	Domain name	Not present	Not present	Not present	Not found	Not present
	Protein family				VI_FHA (cl37254)	
ImpJ	Domain name	T6SS_VasE (pfam05936)	T6SS_VasE (pfam05936)	T6SS_VasE (pfam05936)	COG3522 (COG3522)	T6SS_VasE (pfam05936)

	Protein family	T6SS_VasE (cl01406)	T6SS_VasE (cl01406)	T6SS_VasE (cl01406)	T6SS_VasE (cl01406)	T6SS_VasE (cl01406)
ImpK	Domain name	DotU (pfam09850)	DotU (pfam09850)	DotU (pfam09850)	Not present	Not present
	Protein Family	DotU (cl01370)	DotU (cl01370)	DotU (cl01370)		
ImpM	Domain name	Not found	Not found	Not found	Not present	Not present
	Protein family	DUF2094 (cl01611)	DUF2094 (cl01611)	DUF2094 (cl01611)		
TssE	Domain name	COG3518 (COG3518)	COG3518 (COG3518)	COG3518 (COG3518)	Not present	VI_zyme (TIGR03357)
	Protein family	GPW_gp25 (cl01403)	GPW_gp25 (cl01403)	GPW_gp25 (cl01403)		GPW_gp25 (cl01403)
TssH	Domain name	VI_ClpV1 (TIGR03345)	VI_ClpV1 (TIGR03345)	VI_ClpV1 (TIGR03345)	Not found	Not Found
	Protein family	VI_ClpV1 (cl37250)	VI_ClpV1 (cl37250)	VI_ClpV1 (cl37250)	VI_ClpV1 (cl37250)	VI_ClpV1 (cl37250)
TssJ	Domain name	Not found	Not found	Not found	Not present	T6SS-SciN (pfam12790)
	Protein family	T6SS-SciN (cl01405)	T6SS-SciN (cl01405)	T6SS-SciN (cl01405)		T6SS-SciN (cl01405)
TagN	Domain name	OmpA_C-like (cd07185)	OmpA_C-like (cd07185)	OmpA_C-like (cd07185)	Not present	Not present
	Protein family	OmpA_C-like (cl30079)	OmpA_C-like (cl30079)	OmpA_C-like (cl30079)		
TagX	Domain name	L-Ala-D-Glu_peptidase_like (cd14845)	L-Ala-D-Glu_peptidase_like (cd14845)	L-Ala-D-Glu_peptidase_like (cd14845)	Not present	Not present
	Protein family	Peptidase_M15 (cl38918)	Peptidase_M15 (cl38918)	Peptidase_M15 (cl38918)		
IcmF	Domain name	IcmF (COG3523)	IcmF (COG3523)	IcmF (COG3523)	Not found	Not present
	Protein family	IcmF (cl34628)	IcmF (cl34628)	IcmF (cl34628)	IcmF (cl34628)	

CDD of NCBI was helpful in protein family and domain analysis while functions of different components were analyzed by reviewing the literature. Out of 13 core components Hcp was the one that got translocated to the cytoplasm of the host cell where it causes apoptosis and prevents phagocytosis activity while VgrG is involved in the puncturing of the host cell membrane [33]. Other components are mainly involved in base plate formation, tail sheath formation, Hcp secretion and other related processes as discussed in Table 9.

Table 9. Functions of various T6SS components

T6SS components	Functions	References
ImpA/TssA	They are involved in the facilitation and priming of sheath tube polymerization	[37]
ImpB/VipA/TssB	One of the tail sheath subunits (smaller) involved in sheath polymerization and virulence factor secretion pathway	[10]
ImpC/VipB/TssC	Tail sheath lager subunit involved in sheath polymerization and virulence factor secretion pathway	[6]
TssE	Needed for TssB assembly with TssC	[27]
ImpG/VasA/TssF	Two molecules of TssF are involved in the production of the T6SS base plate by forming complexes with other components. It is recruited in the apparatus before sheath extension.	[7, 22]
ImpH/VasB/TssG	One molecule of TssG forms a complex with other components for base plate formation.	[22]
ClpV/TssH	Particular to T6SS, involved in providing energy to the protein secretion process	[6]
VasD/TssJ	It encodes the outer membrane lipoprotein.	[1]
ImpJ/VasE/TssK	TssK functions as an essential positional marker for the recruitment, nucleation and assembly of contractile apparatus	[22]
ImpK/VasF/DotU/TssL	It is involved in Hcp secretion and intracellular host response modulation	[41]
IcmF/VasK/TssM	It is involved in Hcp secretion and functions as an ATPase	[23]
TagX	Essential for Hcp secretion	[44]
TagN	No data available	[28]
ImpM/TagF	Involved in repression of T6SS and anti-bacterial activity post-translation	[18]
ImpI/VasC	It is involved in the post-translational regulation of T6SS	[9]

The literature review not only explained the individual roles of several components but also suggested links between different components of T6SS that link with each other by the formation of different types of complexes and their functions sometimes depend upon the presence or absence of a particular component. TssK, TssF and TssG are found to form TssKFG complex [22]. This complex is involved in base plate formation and interacts with TssBC for sheath polymerization [31]. ClpV is involved in the recycling process of TssB and TssC [22].

Phylogenetic tree buildup and analysis

Annotated DNA sequences of all T6SS clusters of 3 strains were downloaded from RAST SEED Viewer in FASTA format. Those sequences were aligned by ClustalW in MEGA X software. The evolutionary history was inferred by using the Maximum Likelihood method and the Tamura-Nei model [39]. The tree with the highest log likelihood (-123748.22) is shown in Fig. 5. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 117 nucleotide sequences. There was a total of 4322 positions in the final dataset.

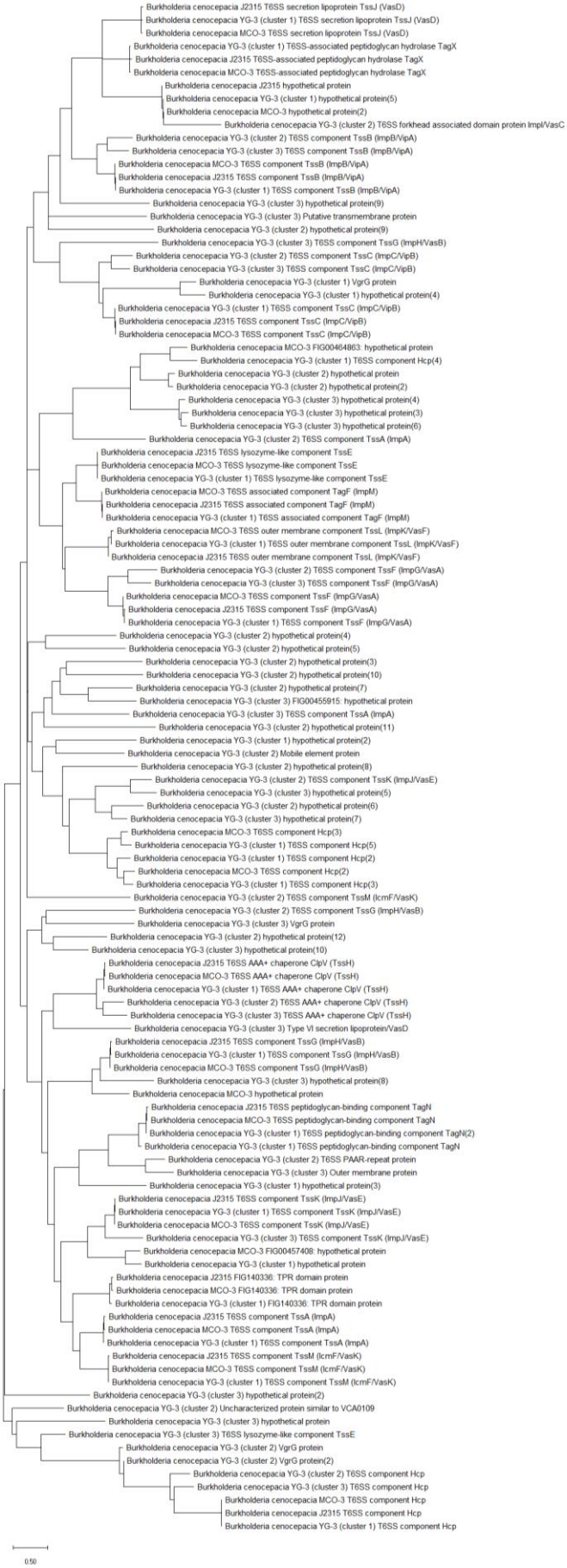

Fig. 5 Tree with the highest log likelihood

The phylogenetic tree of different clusters of T6SS found in *B. cenocepacia* J2315, MC0-3 and YG-3 (Fig. 5) represents evolutionary links between core components of J2315, MC0-3 and YG-3 Cluster 1 as most of the components of Cluster 3 can be seen sharing the same branch while core components of YG-3 Cluster 2 and Cluster 3 are found to be linked to each other more closely in comparison to other clusters. Exceptions can be seen mostly in the cases of Hcp and VgrG components along with hypothetical proteins.

B. cenocepacia MC0-3 was isolated from soil associated with maize roots and YG-3 was isolated from the roots of the Populus tree. The main objective of the study was to find if the T6SS clusters in *B. cenocepacia* MC0-3 and YG-3 were involved in niche adaptation and causing pathogenicity in their specific hosts. *B. cenocepacia* J2315 strain which is commonly found in causing Cystic Fibrosis in humans was used as a reference organism. The results I found and the literature reviewed suggested that strains J2315, and MC0-3 bear a T6SS-1 type cluster that is involved in causing pathogenicity and survival of a bacterium in competition with other bacteria. Also, the presence of TssH, TssL and TssM in these clusters suggests their role in virulence. Though it is still unknown if these 3 components work in collaboration or individually, but they were found to be involved in causing virulence in previous studies. This study predicts that these clusters are functional and are involved in niche adaptation and causing pathogenicity. YG-3 strain had 3 clusters out of which Cluster 1 is predicted to be T6SS-1 type so it is also performing the same functions as that of T6SS clusters of J2315 and MC0-3. But the other two clusters lack certain core components inside the cluster. In Cluster 2, TssE, TssJ and TssL components were absent, while in Cluster 3, TssL and TssM components were absent. The absence of these core components from the cluster suggests that these clusters might not be functional as T6SS assembly is not possible without membrane-forming components, i.e., TssJ, TssL and TssM.

Conclusion

The study provides insights into genome wide identification, characterization and evolutionary analysis of T6SS in *B. cenocepacia* strains. Five potentially functional T6SS gene clusters were found in 3 strains of *B. cenocepacia* with numerous components including core components, accessory components and effector proteins. Literature reveals the functionality of different T6SS components, both core and accessory, in virulence and other pathogenicity factors along with the roles like niche adaptation and bacterial competition. Evolutionary analysis of components of all clusters reveals links in T6SS clusters of J2315, MC0-3 and YG-3 Cluster 1. YG-3 Cluster 2 and 3 lack major components and their present components did not show evolutionary links with components of T6SS clusters of J2315, MC0-3 and YG-3 Cluster 1. Overall, the findings highlight the diversity and evolutionary dynamics of T6SS in studied strains providing a foundation for further research on the functional roles of various components in collaboration and exploiting T6SS components as potential drug targets in the context of *Barkholderia* infections.

Acknowledgements

We want to express our gratitude towards all the colleagues and faculty at the Department of Botany, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi and the Department of Biosciences, COMSATS University Islamabad, Sahiwal Campus, Sahiwal for their useful suggestions during the research.

References

1. Alteri C. J., H. L. T. Mobley (2016). The Versatile Type VI Secretion System, Virulence Mechanisms of Bacterial Pathogens, 337-356.
2. Angus A. A., C. M. Agapakis, S. Fong, S. Yerrapragada, et al. (2014). Plant-associated Symbiotic *Burkholderia* Species Lack Hallmark Strategies Required in Mammalian Pathogenesis, PloS One, 9, e83779.
3. Arakawa K., M. Tomita (2007). The GC Skew Index: A Measure of Genomic Compositional Asymmetry and the Degree of Replicational Selection, Evolutionary Bioinformatics, 3, 117693430700300006.
4. Bayer-Santos E., L. D. M. Ceseti, C. S. Farah, C. E. Alvarez-Martinez (2019). Distribution, Function and Regulation of Type 6 Secretion Systems of Xanthomonadales, Frontiers in Microbiology, 10, 1635.
5. Bodilis J., E. Denet, E. Brothier, A. Graindorge, et al. (2018). Comparative Genomics of Environmental and Clinical *Burkholderia cenocepacia* Strains Closely Related to the Highly Transmissible Epidemic ET12 Lineage, Frontiers in Microbiology, 9, 383.
6. Boyer F., G. Fichant, J. Berthod, Y. Vandenbrouck, et al. (2009). Dissecting the Bacterial Type VI Secretion System by a Genome Wide *in silico* Analysis: What Can be Learned from Available Microbial Genomic Resources?, BMC Genomics, 10, 104.
7. Brunet Y. R., A. Zoued, F. Boyer, B. Douzi, et al. (2015). The Type VI Secretion TssEFGK-VgrG Phage-like Baseplate is Recruited to the TssJLM Membrane Complex via Multiple Contacts and Serves as Assembly Platform for Tail Tube/Sheath Polymerization, PLOS Genetics, 11, e1005545.
8. Cuixia L. I., L. I. Zhifang, D. O. Lingling, Y. U. Yi, et al. (2020). A Genome-wide Identification of the BLH Gene Family Reveals BLH1 Involved in Cotton Fiber Development, Journal of Cotton Research, 3(26), https://doi.org/10.1186/s42397-020-00068-y.
9. De Maayer P., S. N. Venter, T. Kamber, B. Duffy, et al. (2011). Comparative Genomics of the Type VI Secretion Systems of *Pantoea* and *Erwinia* Species Reveals the Presence of Putative Effector Islands That May be Translocated by the VgrG and Hcp Proteins, BMC Genomics, 12, 1-15.
10. Douzi B., L. Logger, S. Spinelli, S. Blangy, et al. (2018). Structure-function Analysis of the C-terminal Domain of the Type VI Secretion TssB Tail Sheath Subunit, Journal of Molecular Biology, 430, 297-309.
11. Filloux A. (2009). The Type VI Secretion System: A Tubular Story, The EMBO Journal, 28, 309-310.
12. Fridman C. M., K. Keppel, M. Gerlic, E. Bosis, et al. (2020). A Comparative Genomics Methodology Reveals a Widespread Family of Membrane-disrupting T6SS Effectors, Nature Communications, 11(1), 1-14.
13. Gallique M., M. Bouteiller, A. Merieau (2017). The Type VI Secretion System: A Dynamic System for Bacterial Communication?, Frontiers in Microbiology, 8, 1454.
14. Grant J. R., P. Stothard (2008). The CGView Server: A Comparative Genomics Tool for Circular Genomes, Nucleic Acids Research, 36, W181-W184.
15. Hall T., I. Biosciences, C. Carlsbad (2011). BioEdit: An Important Software for Molecular Biology, GERF Bull Biosci, 2, 60-61.
16. Holden M. T., H. M. Seth-Smith, L. C. Crossman, M. Sebaihia, et al. (2009). The Genome of *Burkholderia cenocepacia* J2315, an Epidemic Pathogen of Cystic Fibrosis Patients, Journal of Bacteriology, 191, 261-277.

17. Jacobs J. L., A. C. Fasi, A. Ramette, J. J. Smith, et al. (2008). Identification and Onion Pathogenicity of *Burkholderia cepacia* Complex Isolates from the Onion Rhizosphere and Onion Field Soil, Applied and Environmental Microbiology, 74, 3121-3129.
18. Jiang X., A. Beust, P. K. Sappa, U. Völker, et al. (2019). Two Functionally Deviating Type 6 Secretion Systems Occur in the Nitrogen-fixing Endophyte *Azoarcus olearius* BH72, Frontiers in Microbiology, 10, 419168.
19. Kirzinger M. W., G. Nadarasah, J. Stavrinides (2011). Insights into Cross-kingdom Plant Pathogenic Bacteria, Genes, 2, 980-997.
20. Kumar S., G. Stecher, M. Li, C. Knyaz, et al. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, Molecular Biology and Evolution, 35, 1547-1549.
21. LaSalle J. M. (2011). A Genomic Point-of-view on Environmental Factors Influencing the Human Brain Methylome, Epigenetics, 6, 862-869.
22. Liebl D., M. Robert-Genthon, V. Job, V. Cogoni, et al. (2019). Baseplate Component TssK and Spatio-temporal Assembly of T6SS in *Pseudomonas aeruginosa*, Frontiers in Microbiology, 10, 1615.
23. Ma L. S., F. Narberhaus, E. M. Lai (2012). IcmF Family Protein TssM Exhibits ATPase Activity and Energizes Type VI Secretion, Journal of Biological Chemistry, 287, 15610-15621.
24. Mahenthiralingam E., T. A. Urban, J. B. Goldberg (2005). The Multifarious, Multireplicon *Burkholderia cepacia* Complex, Nature Reviews Microbiology, 3, 144-156.
25. Mannaa M., I. Park, Y. S. Seo (2019). Genomic Features and Insights into the Taxonomy, Virulence, and Benevolence of Plant-associated *Burkholderia* Species, International Journal of Molecular Sciences, 20, 121.
26. Marchler-Bauer A., C. Zheng, F. Chitsaz, M. K. Derbyshire, et al. (2012). CDD: Conserved Domains and Protein Three-dimensional Structure, Nucleic Acids Research, 41, D348-D352.
27. Navarro-Garcia F., F. Ruiz-Perez, Á. Cataldi, M. Larzábal (2019). Type VI Secretion System in Pathogenic *Escherichia coli*: Structure, Role in Virulence, and Acquisition, Frontiers in Microbiology, 10, 1965.
28. Nguyen V. S., S. Spinelli, É. Cascales, A. Roussel, et al. (2021). Anchoring the T6SS to the Cell Wall: Crystal Structure of the Peptidoglycan Binding Domain of the TagL Accessory Protein, PloS One, 16, e0254232.
29. Nishida H. (2012). Evolution of Genome Base Composition and Genome Size in Bacteria, Frontiers in Microbiology, 3, 420.
30. Overbeek R., R. Olson, G. D. Pusch, G. J. Olsen, et al. (2014). The SEED and the Rapid Annotation of Microbial Genomes Using Subsystems Technology (RAST), Nucleic Acids Research, 42, D206-D214.
31. Park Y. J., K. D. Lacourse, C. Cambillau, F. DiMaio, et al. (2018). Structure of the Type VI Secretion System TssK-TssF-TssG Baseplate Subcomplex Revealed by Cryo-electron Microscopy, Nature Communications, 9, 1-11.
32. Pukatzki S., A. T. Ma, A. T. Revel, D. Sturtevant, et al. (2007). Type VI Secretion System Translocate a Phage Tail Spike-like Protein into Target Cells where it Cross-links Actin, Proceedings of the National Academy of Sciences, 104, 15508-15513.

33. Sha J., J. A. Rosenzweig, E. V. Kozlova, S. Wang, et al. (2013). Evaluation of the Roles Played by Hcp and VgrG Type 6 Secretion System Effectors in *Aeromonas hydrophila* SSU Pathogenesis, Microbiology, 159, 1120.
34. Shalom G., J. G. Shaw, M. S. Thomas (2007). *In vivo* Expression Technology Identifies a Type VI Secretion System Locus in *Burkholderia pseudomallei* that is Induced upon Invasion of Macrophages, Microbiology, 153, 2689-2699.
35. Spiewak H. L., S. Shastri, L. Zhang, S. Schwager, et al. (2019). *Burkholderia cenocepacia* utilizes a Type VI Secretion System for Bacterial Competition, Microbiology Open, 8, e00774.
36. Stan C., M. T. Cristescu, B. I. Luiza, C. P. Cristescu (2013). Investigation on Series of Length of Coding and Non-coding DNA Sequences of Bacteria Using Multifractal Detrended Cross-correlation Analysis, Journal of Theoretical Biology, 321, 54-62.
37. Stietz M. S., X. Liang, H. Li, X. Zhang, T. G. Dong (2020). TssA-TssM-TagA Interaction Modulates Type VI Secretion System Sheath-tube Assembly in *Vibrio cholera*, Nature Communications, 11, 1-11.
38. Syngai G. G., P. Barman, R. Bharali, S. Dey (2013). BLAST: An Introductory Tool for Students to Bioinformatics Applications, Keanean Journal of Science, 2, 67-76.
39. Tamura K., M. Nei (1993) Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees, Molecular Biology and Evolution, 10, 512-526.
40. Tao Z., Y. Xu, S. Zhou, D. Qian, et al. (2020). Acquisition of a Type VI Secretion System is Critical for *Pseudomonas plecoglossicida* Induced Granulomas in Fish Internal Organs, Aquaculture, 516, 734629.
41. Wang S., J. Dai, Q. Meng, X. Han, et al. (2014). DotU Expression is Highly Induced during *in vivo* Infection and Responsible for Virulence and Hcp1 Secretion in Avian Pathogenic *Escherichia coli*, Frontiers in Microbiology, 5, 588.
42. Wang X., X. Zhang, X. Liu, Z. Huang, et al. (2019). Physiological, Biochemical and Proteomic Insight into Integrated Strategies of an Endophytic Bacterium *Burkholderia cenocepacia* Strain YG-3 Response to Cadmium Stress, Metallomics, 11, 1252-1264.
43. Wanger A., V. Chavez, R. S. P. Huang, A. Wahed, et al. (2017). Overview of Bacteria, Microbiology and Molecular Diagnosis in Pathology, 75-117.
44. Wattam A. R., J. J. Davis, R. Assaf, S. Boisvert, et al. (2017). Improvements to PATRIC, the All-bacterial Bioinformatics Database and Analysis Resource Center, Nucleic Acids Research, 45, D535-D542.
45. Weber B. S., S. W. Hennon, M. S. Wright, N. E. Scott, et al. (2016). Genetic Dissection of the Type VI Secretion System in Acinetobacter and Identification of a Novel Peptidoglycan Hydrolase, TagX, Required for Its Biogenesis, mBio, 7, e01253-16.
46. Wisplinghoff H. (2017). *Pseudomonas* spp., *Acinetobacter* spp. and Miscellaneous Gram-negative Bacilli, Infectious Diseases, 1579-1599.
47. Wu H., Z. Zhang, S. Hu, J. Yu (2012). On the Molecular Mechanism of GC Content Variation among Eubacterial Genomes, Biology Direct, 7, 1-16.
48. Yar A. M., G. Zaman, A. Hussain, Y. Changhui, et al. (2018). Comparative Genome Analysis of 2 *Mycobacterium Tuberculosis* Strains from Pakistan: Insights Globally into Drug Resistance, Virulence, and Niche Adaptation, Evolutionary Bioinformatics, 14, 1-9.

49. Yu K. W., P. Xue, Y. Fu, L. Yang (2021). T6SS Mediated Stress Responses for Bacterial Environmental Survival and Host Adaptation, International Journal of Molecular Sciences, 22, 478.
50. Zoued A., Y. R. Brunet, E. Durand, M. S. Aschtgen, et al. (2014). Architecture and Assembly of the Type VI Secretion System, Biochimica and Biophysica Acta (BBA) - Molecular Cell Research, 1843, 1664-1673.

Sardarni Khadija Raza, M.Sc., M.Phil.
E-mail: khadijaraza3@gmail.com

Sardarni Khadija Raza received her M.Sc. and M.Phil. Degree in Botany from Pir Mehr Ali Shah (PMAS) Arid Agriculture University Rawalpindi, Punjab, Pakistan. She served at Visiting Faculty at the Department of Botany, Rawalpindi Women's University, Rawalpindi, Punjab, Pakistan from October 2022 till February 2023. Currently, she is working on different research manuscripts for publication in various journals. Those include the Gibberellic acid effect on Chrysanthemum cut flowers and heavy metal detection in drinking water and fish meat from District Rawalpindi, Pakistan.

Assoc. Prof. Muhammad Ibrahim, Ph.D.
E-mail: ibrahim@cuisahiwal.edu.pk

Muhammad Ibrahim is an Associate Professor at the Department of Biosciences and he also holds the responsibilities as a Principal Investigator in HEC funded project on Next Generation Sequencing and Mapping of Opportunistic Bacteria. he is a Post Doctorate from Human Genome Centre, Institute of Medical Sciences, University of Tokyo, Japan. His research interest include bioinformatics focusing on microbial genome information analysis such as genome assembly, genome annotation, comparative genomics, RNA sequencing and analysis using TopHat, Cufflink, Galaxy and Proteomics, parallel computing framework for genome analysis, cloud computing, molecular biology such as gene cloning, transformation, gene hybridization, LC-MS/MS, ICP-MS, RT-qPCR, BIOLOG, transmission and scanning electron microscopy.

Kainaat, M.Phil.
E-mail: kainaatkarim@gmail.com

Kainaat has completed her M.Phil. Degree in Botany from PMAS Arid Agriculture University, Rawalpindi, Pakistan. During her M.Phil., she worked on COBRA-like genes in the Poaceae family using the tools of bioinformatics. She is also a co-author of the book "Synthetic Bacteria: The Practical Application of Genomics and Genetic Engineering". Currently, she is writing the review papers and pursuing a teaching career in biology. The review papers that she is currently working on include the databases used for botanical research and studies and the work done on some particular genes in plants.

Afshan Farid, M.Phil.
E-mail: afshanfarid2017@gmail.com

Afshan Farid belongs to Dera Ghazi Khan (DGK), South Punjab, Pakistan. She got her basic education from DGK, South Punjab, Pakistan. She completed her M.Phil. Degree in Botany from PMAS Arid Agricultural University, Rawalpindi, Pakistan. Her research experience includes the response of wheat to proline concentration under drought stress and *in silico* gene-wide identification and characterization of salt-tolerant genes in plant growth-promoting bacteria. She has work experience at Ghazi University DGK, Pakistan (visiting lecturer) from February 2018 to 2019 and now. Extra qualification includes basic knowledge of computers and software.

Sardar Shah Nawaz Khan, M.Sc.
E-mail: sardarshahnawazkhankhan@gmail.com

Sardar Shah Nawaz Khan has B.Sc. Hons. and M.Sc. Hons. Degree in Agriculture (Food Technology) and Food Safety and Quality Management respectively from Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Punjab, Pakistan. During his M.Sc. Hons., he helped in writing and reviewing various research and review articles including genome-wide identification, characterization and evolutionary analysis of T6SS in *B. cenocepacia* strains and heavy metal detection in drinking water and fish meat from District Rawalpindi, Pakistan. His abstract on the postharvest storage process of fruits and vegetables by electrolyte water treatment and its solicitation was published in Food Safety and Nutritional Expo 2022.

Aqsa Talat, M.Sc. Student
E-mail: aqsatalat9099@gmail.com

The author has completed her B.Sc. (Hons.) Degree in Agriculture in Food Technology from PMAS Arid Agriculture University, Rawalpindi, Pakistan. During her B.Sc., she has worked on publishing different articles relating to agriculture using the tools of bioinformatics. She is also a co-author of a paper "Energy-providing protein: creating and analyzing the impact of healthy soy bars on school-aged kids". Currently, she is writing the review papers and doing a M.Sc. (Hons.) in Food Safety and Quality Management. The review papers that she is currently working on include the bioinformatics databases used for plant-based research and the work done on some particular genes in plants.

Tayyaba Arzoo, M.Sc. Student
E-mail: tayyabaarzo04@gmail.com

Tayyba is a dedicated bioinformatics researcher currently pursuing her M.Sc. Degree in Bioinformatics at COMSATS University Islamabad. Her research primarily focuses on proteogenomic studies in various plant species, utilizing mass spectrometry as a powerful tool to unravel biological complexities. Her academic journey commenced at Quaid-e-Azam University Islamabad, where she successfully completed her graduation and actively engaged in diverse bioinformatics projects, acquiring expertise in essential tools. Recently, she collaborated on the project titled "Genomewide Identification, Characterization, and Evolutionary Analysis of T6SS in *Burkholderia cenocepacia* Strains". Her contributions to this project involved providing valuable insights and analyses, contributing to the success of the study.

Syeda Izzat Fatima, M.Sc. Student
E-mail: izzatfatima3248@gmail.com

Syeda Izzat Fatima has completed her B.Sc. (Hons.) Degree in Agriculture (Food Technology) and currently enrolled in M.Sc. (Hons.) in Food Safety and Quality Management from PMAS Arid Agriculture University, Rawalpindi, Pakistan. With a profound interest in safeguarding public health, Izzat's academic research focuses on understanding the risks associated with foodborne illness, related to contaminants and pathogens. She is also writing review papers that are related to the antimicrobial resistance of pathogens. Through her coursework, she is gaining knowledge about food safety regulations, quality assurance and risk assessment that are necessary to ensure that people consume safe food of the highest standards.

Prof. Abdul Waheed, Ph.D.
E-mail: abdul.waheed@uaar.edu.pk

Abdul Waheed completed his earlier education in Multan and got his Ph.D. Degree in Botany from the Institute of Pure and Applied Biology, Bahauddin Zakaria University, Multan, Punjab, Pakistan. He served as a research officer at Bahauddin Zakaria University; Senior subject specialist in Biology; Assistant, Associate and Professor of Botany at PMAS Arid Agriculture University, Rawalpindi; Director of Advance Studies at PMAS Arid Agriculture University, Rawalpindi; Campus Director at Comsats University Islamabad, Sahiwal Campus; Director of University Institute of Management Sciences, PMAS Arid Agriculture University, Rawalpindi.