Comparative Genome Analysis of Carbohydrate-Active Enzymes and Virulence Factors in Lichen-Associated Variovorax Sp. PAMC28711

Prasansah Shrestha
Sun Moon University
So-Ra Han
Sun Moon University
Jun Hyuck Lee
Korea Polar Research Institute
Hyun Park
Korea University
Tae-Jin Oh (tjoh3782@sunmoon.ac.kr)
Sun Moon University

Research Article

Keywords: Carbohydrate active enzyme, comparative genomics, lichen, Variovorax sp. PAMC28711, virulence factors

Posted Date: January 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-137024/v1

Abstract

Background: The genus *Variovorax* sp. PAMC28711 is a cold-adapted microorganism, isolated from Antarctica lichen *Himantormia*. The complete genomes of six *Variovorax* species were analyzed and compared along with the strain PAMC28711. The genomic information was collected from NCBI as well as PATRIC databases. Likewise, CAZyme annotation (dbCAN2 Meta server) was performed in order to predict the CAZyme family responsible for trehalose synthesis and trehalose degradation enzymes. The trehalose metabolic pathway was analyzed via the KEGG database. Bioinformatics tools such as OrthoANI software were used to analyze similar genes in different strains under the same genus. Likewise, MEGA X was used for evolutionary and conserved genes.

Results: The complete genome of genus *V.* sp. PAMC28711 was found to comprise CAZyme families GH (10), GT (9), CB (1), AA5 (1), and CE (1). The three trehalose synthetic pathways (OtsA/OtsB, TS, and TreY/TreZ) and trehalose degradation pathway (TreF) were identified only in *V.* sp. PAMC28711 among the different strains of *Variovorax* studied, whereas one to two pathways of trehalose biosynthesis, but not trehalose degradation pathways are involved in other Variovorax strains. The strain PAMC28711 comprises of cytoplasmic trehalase (TreF) as a trehalose degrading enzyme that belongs to the CAZyme family GH37, which is not identified in other strains of *Variovorax*.

Conclusions: To date, although the genus *V.* sp. PAMC28711 has not been reported to exhibit CAZyme activities such as trehalase, and no microorganism expressed different virulence factors, the results based on PATRIC database showed that the strain carried a few virulence genes. Further, this study provides additional information regarding trehalase as one of the factors facilitating bacterial survival under extreme environments and this enzyme has showed potential application in biotechnology fields.

Background

The genus *Variovorax* is a Gram-negative and motile bacterium belonging to the family Comamonadaceae [1] that is found in straight to slightly curved or rod-shaped form. Due to the presence of carotenoid pigments, the genus *Variovorax* exhibits yellow, slimy and shiny colonies. Many strains belonging to family Comamonadaceae thrive in polluted environments and degrade complex organic compounds [2]. The genus *Variovorax* generally inhabits soil and water [3]. *Variovorax* sp. PAMC28711 was isolated from *Himantormia* sp., Antarctic lichen collected from the Barton Peninsula, King George Island, Antarctica [4].

Antarctica is the largest continent, which is approximately the size of Europe. Therefore, several additional and extreme locales such as regions of volcanic activity, hypersaline lakes, subglacial lakes, and even the ice itself harbor specific extremophiles [5]. Therefore, numerous microorganisms have specifically adapted to a wide range of extreme environments to survive under novel biodiversity, much of which has yet to be elucidated [3]. Another key feature of the Antarctica ecosystem is the extreme variation in the physical conditions ranging from freshwater lakes (some of the most oligotrophic environments on Earth) to hypersaline lakes [6]. Microorganisms found under extreme environmental conditions like Antarctica are ideal candidates for the study of eco-physiological and biochemical adaptations of such life forms [5]. Antarctica is one of the most physically and chemically challenging terrestrial environments for habitation [7]. Habitats with permanently low temperature dominate the temperate biosphere and have been successfully colonized by a wide variety of organisms that are collectively termed psychrophiles or cold-adapted organisms [8]. In particular, the lichens are generally defined by mutualistic symbiosis between fungi and algae (Chlorophyta or Cyanobacteria); however, they also contain internal bacterial communities [9]. Bacteria associated with lichens were initially reported in the first half of the 20th century [10]. The lichen-associated microorganism was reported to carry genes involved in the degradation of polymers [11].

Carbohydrate-active enzymes (CAZymes) belong to a large class of enzymes that are involved in the breakdown of complex carbohydrates in the cell. Based on their amino acid sequences, they are classified into families with conserved catalytic mechanism, structure, and active site residues, but differing in substrate specificity [12]. They are responsible for carbohydrate synthesis through glycosyltransferases (GTs), degradation of complex carbohydrates via glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and enzymes for auxiliary activities (AAs) and recognition (carbohydrate-binding module, CBM) [13]. The CAZymes represent a continuously updated list of GH families [12]. The GHs are the largest family of CAZymes that hydrolyze the glycosidic bond between two or more carbohydrates, or between carbohydrate and non-carbohydrate moieties, via overall inversion of anomeric carbon [14].

Virulence is a microbial property that is observed only in susceptible hosts. Virulence is not absolute, and is always measured relative to a standard, usually another microbe or host [15]. Virulence factors are microbial gene products with potential to cause disease within the parasite as well as the host. Virulence factors are bacterial toxins, cell surface proteins that mediate bacterial attachment, cell surface carbohydrates, proteins that protect bacteria, and hydrolytic enzymes that may contribute to bacterial pathogenicity (VFanalyzer; http://www.mgc.ac.cn/VFs/main.htm) [16]. Some of the identified virulence factors facilitate physiological and metabolic adaptation of the bacteria in adverse environments [17].

In this study, the lichen-associated cold-adapted aerobic bacterium *V.* sp. PAMC28711 belonging to family Comamonadaceae was selected. Most of the *Variovorax* species are flagellated and motile. Other species of *Variovorax* were isolated from the soil, with optimum growth under mesophilic temperature. By contrast, *V.* sp. PAMC28711, isolated from the Antarctica, can tolerate temperature variation. Although the CAZyme has been studied in various microorganisms, it has yet to be reported in *Variovorax*. This study was thus carried out to compare the role of CAZyme families in the complete genome of *Variovorax* species. In addition to the comparative study, a virulence factor in the six species of genus *Variovorax* was compared and analyzed in silico to identify the enzymes involved in bacterial virulence.

Methods

General information

Information pertaining to the six complete genomes of the genus *Variovorax* was obtained from the National Center for Biotechnology Information (NCBI) database for comparative analysis. The gene bank accession numbers of the six *Variovorax* species are NZ_CP014517 (*V.* sp. PAMC28711), NZ_CP023284 (*V. boronicumulans* J1), NZ_022247 (*V. paradoxus* B4), NC_014931 (*V. paradoxux* EPS), NC_012791 (*V. paradoxus* S110), and NZ_CP027773 (*V.* sp. PMC12).

Phylogenetic tree and average nucleotide identity (ANI)

Phylogenetic tree of the 16S rRNA gene sequences in the complete genome of genus *Variovorax* was constructed using the neighbor-joining method, and the bootstrap values at the branch points are shown. Average nucleotide identity in the complete nucleotide sequences of genus *Variovorax* was analyzed using the OrthoANI software [34].

Comparative genome analysis

All six strains of the complete *Variovorax* genome were analyzed using bioinformatics tools, such as CAZyme annotation (dbCAN2 meta server; http://bcb.unl.edu/dbCAN2/). Each genome was annotated using DIAMOND, HMMER, and Hotpep via CAZy, dbCAN, and PPR databases [35]. The dbCAN2 meta server allows submission of nucleotide sequences for prokaryotic and eukaryotic genomes, although protein sequences are preferred. This server uses three tools including DIAMOND (for fast blast hits in the CAZy database), HMMER (for annotated CAZyme domain boundaries according to dbCAN CAZyme domain HMM database), and Hotpep (for conserved short motifs in the PPR library). In order to analyze the trehalose metabolic pathway of *V.* sp. PAMC28711, the Kyoto Encyclopedia of Genes and Genomics (KEGG) pathway database was used [36, 37]. In addition, the PATRIC database (https://patricbrc.org/) was used for genomic information, as well as for the number of virulence factors involved in the respective strains. PATRIC database generally uses the virulence factor database (VFDB), Victors, and PATRIC_VF. The VFDB is an integrated and comprehensive online resource for curating information about the virulence factors of bacterial pathogens [32].

Determination of polysaccharide degradation activity in V. sp. PAMC28711

Azurine cross-linked (AZCL) assay was performed in seven AZCL substrates (AZCL-amylose, AZCL-barley β-glucan, AZCL-arabinoxylan, AZCL-HE-cellulose, AZCL-xylan (beech wood), AZCL-xylan (birch wood), and AZCL-xyloglucan, to determine the enzyme activity of the polysaccharide degradation in *V.* sp. PAMC28711. A medium containing 1.0% agarose and 23 mM phosphoric buffer in addition to polysaccharide substrate powder was prepared. The media containing Bennett's agar (B's), Marine agar (MA), Malt Yeast media (MY), and Reasoner's 2A agar (R2A) were used, along with AZCL substrates. It was sterilized, and left to cool at 60–70 °C, and poured into Petri dishes. The activity was measured using a commercial kit from Megazyme© (Bray, Ireland) at different temperatures of (4, 15, 25, and 37)°C, and expressed as the area (cm^2) with a blue halo in the AZCL assays.

Results And Discussion

General information of the complete genome of genus Variovorax

V. sp. PAMC28711 accounts for 4,316,152 bp and a GC content of 65.97% that is less than that of the other complete genomes of *Variovorax* strains. Table 1 summarizes the general genomic information of all six *Variovorax* strains including GC content (percentage), chromosome number, contigs, and genome length. Likewise, Table 2 provides a comparative summary of isolation source, isolated information, host information, and phenotype information of the six complete genomes of *Variovorax* species.

Table 1
Comparison of genomic features among the six complete genomes of *Variovorax* species.

Genome name	Genome ID	Chromosome	Contig	Genome length	GC (%)	Gene	Protein
V. sp. PAMC28711	1,795,631.3	1	1	4,316,152	65.97	4,170	3,979
V. paradoxus EPS	595,537.3	1	1	6,550,056	67.50	6,055	5,917
V. paradoxus S110	543,728.3	2	2	6,754,997	67.50	6,424	6,247
V. paradoxus B4	1,246,301.3	2	2	7,148,516	67.14	6,778	6,565
V. boronicumulans J1	436,515.5	1	1	7,137,898	68.01	6,607	6,237
V. sp. PMC12	2,126,319.3	2	2	7,015,237	67.61	6,530	6,382

Table 2
Comparison of isolation source, isolated information, host information, and phenotype information of six complete genomes belonging to *Variovorax* species. Six complete genomes of *Variovorax* species and their phenotype information. PAMC 28711 = *V.* sp., EPS = *V. paradoxus*, S110 = *V. paradoxus*, B4 = *V. paradoxus*, J1 = *V. boronicumulans*, PMC = *V.* sp., and N/A = Non-available.

		PAMC28711	EPS	S110	B4	J1	PMC12
Isolation source		Isolated from Antarctica lichen species	Rhizosphere community of the sunflower	Interior of a potato plant	Polluted soil near a production plant of the chemical industry.	Soil	Tomato rhizosphere
Isolation information		Antarctica	N/A	N/A	Hamburg	China	South Korea
Isolation country		Barton Peninsula, King George Island	N/A	N/A	Germany	China	South Korea
Host information		*Himantormia* sp.	*Helianthus annuus*	N/A	N/A	N/A	N/A
Phenotype information	**a. Gram strain**	negative	negative	negative	negative	negative	negative
	b. Cell shape	N/A	Rod	Rod	Rod	N/A	N/A
	c. Motility	Yes	Yes	Yes	Yes	Yes	Yes
	d. Sporulation	N/A	No	No	N/A	No	N/A
	e. Temperature range	Psychrophilic to mesophilic (variable)	Mesophilic	Mesophilic	Mesophilic	N/A	N/A
	f. Optimal temperature	°C	Aerobic	Aerobic	N/A	N/A	N/A
	g. Oxygen requirement		Multiple	Multiple	N/A	N/A	N/A

Phylogenetic and ANI analysis within the genus Variovorax

A phylogenetic tree was constructed using the 16S rRNA gene sequence of *Variovorax* strains via neighbor-joining method [18]. The branches show the relationship of the species in the genus *Variovorax* (Fig. 1A). The percentage of replicate trees in which the associated taxa cluster together in the bootstrap test (1,000 replicates) are shown next to the branches [19]. The evolutionary distances were computed using the Maximum Composite Likelihood method [20], and are expressed by the number of base substitutions per site. This analysis involved six nucleotide sequences of *Variovorax* strains. Evolutionary analyses were conducted in MEGA-X [21]. ANI analysis was conducted with the complete genome sequence of six strains of *Variovorax* (Fig. 1B). ANI varied in the range of 81.10–93.35% among the six strains of *Variovorax*. The strain PAMC28711 is identical with other strains in the range of 81.10–81.68%, which is less than in other strains. Likewise, *V. paradoxus* B4 and *V. paradoxus* S110 show high identity, i.e., 95.35% (Table S1 of the Supplementary Information).

Analysis of trehalose-producing and -degrading CAZyme families and subfamilies in genus Variovorax

Variovorax can degrade complex organic compounds, as well as complex carbohydrates. Likewise, *V.* sp. PAMC28711 comprises CAZyme that is responsible for complex carbohydrate synthesis and degradation. The CAZyme family includes GHs, GTs, CEs, AAs, and CBMs, all of which occur in the genus *Variovorax* as well. Based on the CAZyme annotation outcomes, *V.* sp. PMC12 contains the highest number of GHs, CEs, and CBMs, when compared with other complete genomes of *Variovorax* (Tables S2 and S3 of the Supplementary Information). Among all six strains analyzed, almost similar number of GHs and GTs (10 and 9, respectively) occur in *V.* sp. PAMC28711 (Fig. 2). *V.* sp. PAMC28711 carries three pathways of trehalose synthesis (TPS/TPP, TS, and TreY/TreZ) (Table 3). The enzymes involved in TPS/TPP pathways include trehalose 6-phosphate and trehalose 6-phosphate phosphatase constituting GT20 CAZyme subfamilies (Table 4). Additionally, the TS pathway comprises a trehalose synthase enzyme that belongs to the GH13 CAZyme subfamily. This TS pathway is reversible and includes both the biosynthesis and degradation of trehalose from maltose (Table 4). The CAZyme subfamily GH37 (trehalase, EC 3.2.1.28) (Figs. 3A and C) was present in *V.* sp. PAMC28711, but not in the other *Variovorax* strains used for comparison. In addition, trehalase GH37 in PAMC28711 predicted using the CAZyme database was found to be a cytoplasmic trehalase based on the results of Rapid Annotations using Subsystems Technology (RAST) annotation [22]. Figure 3B and Table S2 (Supplementary Information) show the overall CAZyme subfamilies present in the six different strains of *Variovorax*. There are several alternative pathways for the degradation of trehalose [23]. Interestingly, bacterial trehalases are not as widely distributed as the trehalose biosynthetic pathway, since trehalose-6-phosphate synthases/phosphatases (TPSs/TPPs) occur in diverse living forms, ranging from micro- to macro-organisms [24]. The enzymes involved in trehalose degradation include alpha,alpha-trehalose phosphorylase (EC 2.4.1.64) and alpha,alpha-trehalase (EC 3.2.1.28). *E. coli* strain K12 contains two trehalases (cytoplasmic trehalase TreF and periplasmic trehalase TreA) [25]. TreF was predicted via the KEGG pathway map in *V.* sp. PAMC28711. TreF is the enzyme responsible for the degradation of the disaccharide alpha,alpha-trehalose yielding two glucose subunits [26]. The enzyme exists in a wide variety of organisms, and its sequence is highly conserved throughout evolution [27].

Table 3
Variovorax sp. PAMC28711 CAZyme families/subfamilies associated with trehalose metabolic pathway and their respective EC numbers.

Pathway	Gene	Gene products	E.C	CAZyme families/subfamilies
Biosynthesis	TPS	Trehalose 6-phosphate synthase	2.4.1.15	GT20
	TPP	Trehalose 6-phosphate phosphatase	3.1.3.12	
	TS	Trehalose synthase	5.499.16	GH13
	TreY	Maltooligosyl-trehalose synthase	5.499.15	CBM48 & GH13
	TreZ	Maltooligosyl-trehalose trealdohydrolase	3.2.1.141	
Degradation	TreF	Trehalase	3.2.1.28	GH37

Table 4
Comparison of genes involved in the complete genome of *Vairovorax* and its species. "√" denotes the respective gene, while "No" indicates the absence of the respective gene in the particular strains.

	Trehalose synthesis pathways				**Trehalose degradation pathway**	**Enzyme missing**
Strains	OtsA-OtsB	TreS	TreY	Tre Z	TreF	
V. sp. PAMC28711	√	√	√	√	√	TreY
V. paradoxus B4	√	√	√	√	√	TreF
V. boronicumulans J1	√	No	No	No	No	TreS, TreY, TreZ, TreF
V. paradoxus EPS	√	√	√	√	No	TreF
V. paradoxus S110	√	√	√	√	No	TreF
V. sp. PMC12	√	√	√	√	No	TreF

Generally, five distinct trehalose synthetic pathways exist in bacteria. They include TPS/TPP pathway (enzymes such as trehalose phosphate synthase and trehalose 6-phosphate phosphatase), TS pathway (trehalose synthase), TreY/TreZ pathway (enzymes such as maltooligosyl-trehalose synthase and malto-oligosyl-trehalose trehaldohydrolase), TreP pathway (trehalose phosphorylase), and TreT pathway (trehalose glycosyltransferring synthase) [28, 29]. However, according to KEGG database results, bacterial strain *V.* sp. PAMC28711 contains only three trehalose synthetic pathways (TPS/TPP, TS, and TreY/TreZ pathway). Figures 4 and 5 represent the estimated pathways in trehalose metabolism by *Variovorax*. In addition, RAST annotation provided information about enzymes, which are involved in trehalose metabolism of respective strains. Nevertheless, the KEGG database showed the absence of one of the enzymes (maltooligosyl-trehalose synthase) in the trehalose synthesis (TreY/TreZ pathway) in *V.* sp. PAMC28711, *V. paradoxus* B4, and *V.* sp. PMC12. Interestingly, the missed enzymes (maltooligosyl-trehalose synthase) were detected in the RAST annotation database, suggesting a limitation in the KEGG database. The results displayed in Fig. 5 suggest that only *Variovorax* strain PAMC28711 carries all enzymes involved in trehalose metabolism (i.e., trehalose synthesis and trehalose degradation).

AZCL screening of polysaccharide degradation potential of an Antarctic bacterium

The polysaccharide degradation activity was determined via AZCL screening of *V.* sp. PAMC28711. It was found that the strain PAMC28711 degraded various AZCL substrates, such as AZCL-amylose, AZCL-barley β-glucan, AZCL-arabinoxylan, AZCL-HE-cellulose, AZCL-xylan (beech wood), AZCL-xylan (birch wood), and AZCL-xyloglucan, as shown in Table 5. The present authors found that the synthesis of the related polysaccharide enzymes is indicated by blue staining of the small-hydrolyzed compounds and diffuse staining in the plate, forming blue circular zones. In addition, strain PAMC28711 showed the ability to degrade AZCL substrates mainly in the mesophilic temperature i.e., at 25 °C. This screening method has been utilized in several studies [30, 31], which somewhat confirms its reliability.

Table 5
AZCL screening of the polysaccharide degradation potential of *Variovorax* sp. PAMC28711 using AZCL-amylose, AZCL-barley β-glucan, AZCL-arabinoxylan, AZ AZCL-xylan (beech wood), AZCL-xylan (birch wood), and AZCL-xyloglucan ("+" = degradation of the respective AZCL substrate, and "-" = no degradation of the substrate).

	AZCL substrates																								
	AZCL-amylose				AZCL-barley β-glucan				AZCL-arabinoxylan				AZCL-HE-cellulose				AZCL-xylan (beech wood)				AZCL-xylan (birchwood)				AZC
Media/Temp (°C)	4	15	25	37	4	15	25	37	4	15	25	37	4	15	25	37	4	15	25	37	4	15	25	37	4
B's	-	-	+	-	+	+	+	+	+	+	+	+	-	-	+	-	-	-	+	+	-	-	+	+	-
MA	+	+	+	-	-	+	+	-	+	+	+	+	-	-	+	-	+	+	+	+	+	+	+	+	-
MY	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+	-	-	+	+	-	-
R2A	+	+	+	-	-	+	+	-	+	+	+	+	-	+	+	-	+	+	+	-	+	+	+	-	-

Virulence factor in genus Variovorax

Virulence factors are gene products that let bacteria colonize on or within a host organism, resulting in enhanced risk of disease [32]. Based on the PATRIC results (Fig. 6), all the different strains of genus *Variovorax* carry virulence factors. Among six *Variovorax* strains, *V. paradoxus* S110 and *V. boronicumulans* J1 carry a higher number of virulence factors, when compared with the other strains. Lee et al. [33] reported a similar result for *V. paradoxus* S110 (GenBank CP001635.1), which is consistent with the study of virulence genes in oil-contaminated seawater. *V.* sp. PAMC28711, *V. paradoxus* B4, and *V.* sp. PMC12, which was ranked second in the highest number of virulence factors (Table 6). The PATRIC database shows virulence factors by integrating three virulence factor databases: VFDB, PATRIC VF (PATRIC virulence factor), and Victors virulence factors. Based on the results of PATRIC database (Table 6) and Victors, the virulence factor database showed a higher number of virulence genes compared with the PATRIC VF and VFDB databases. Based on the results obtained from the PATRIC database, all six strains share the virulence factor, RNA-binding protein Hfq gene.

Table 6
Genes coding for virulence factors in the complete genome of *Variovorax* strains.

Strains	Gene No.	Source	Ref. sEq. locus	Products
Variovorax paradoxus **EPS**	3	Victors	Varpa_4360	Argininosuccinate synthase (EC 6.3.4.5)
			Varpa_3698	RNA-binding protein Hfq
			Varpa_5921	RecA protein
	1	PATRIC_VF	Varpa_3698	RNA-binding protein Hfq
Variovorax paradoxus **B4**	4	Victors	VAPA_1c54030	RecA protein
			VAPA_1c38960	Argininosuccinate synthase (EC 6.3.4.5)
			VAPA_1c23720	RNA-binding protein Hfq
			VAPA_2c09130	Catalase KatE (EC 1.11.1.6)
	2	VFDB	VAPA_2c09130	Catalase KatE (EC 1.11.1.6)
			VAPA_1c06250	T6SS component TssC (ImpC/VipB)
	1	PATRIC_VF	VAPA_1c23720	RNA-binding protein Hfq
Variovorax paradoxus **S110**	5	Victors	Vapar_3766	Argininosuccinate synthase (EC 6.3.4.5)
			Vapar_5195	RecA protein
			Vapar_5439	Catalase KatE (EC 1.11.1.6)
			Vapar_2192	RNA-binding protein Hfq
			Vapar_5049	LSU ribosomal protein L36p at LSU ribosomal protein L36p,zinc-dependent
	2	VFDB	Vapar_0549	T6SS component TssC (ImpC/VipB)
			Vapar_5439	Catalase KatE (EC 1.11.1.6)
	1	PATRIC_VF	Vapar_2192	RNA-binding protein Hfq
Variovorax boronicumulans **J1**	5	Victors	CKY39_32515	RecA protein
			CKY39_20035	RNA-binding protein Hfq
			CKY39_24660	Argininosuccinate synthase (EC 6.3.4.5)
			CKY39_31455	LSU ribosomal protein L36p at LSU ribosomal protein L36p,zinc-dependent
			N/A	Chorismate synthase (EC 4.2.3.5)
	1	PATRIC_VF	CKY39_20035	RNA-binding protein Hfq
	1	VFDB	CKY39_03085	T6SS component TssC (ImpC/VipB)
Variovorax **sp. PMC 12**	4	Victors	C4F17_20450	LSU ribosomal protein L36p at LSU ribosomal protein L36p,zinc-dependent
			C4F17_03070	RNA-binding protein Hfq
			C4F17_26520	Argininosuccinate synthase (EC 6.3.4.5)
			C4F17_19600	RecA protein
	1	PATRIC_VF	C4F17_03070	RNA-binding protein Hfq
	1	VFDB	C4F17_15975	T6SS component TssC (ImpC/VipB)
Variovorax **sp. PAMC28711**	4	Victors	AX767_11010	RecA protein
			N/A	LSU ribosomal protein L36p at LSU ribosomal protein L36p,zinc-dependent
			AX767_00715	RNA-binding protein Hfq
			AX767_05975	Argininosuccinate synthase (EC 6.3.4.5)
	1	PATRIC_VF	AX767_00715	RNA-binding protein Hfq
	1	VFDB	AX767_12640	T6SS component TssC (ImpC/VipB)

Conclusions

In this study, the complete genome of genus *Variovorax*, strain PAMC28711 was compared with that of five other strains: EPS, S110, B4, J1, and PMC12. A comparative analysis of the obtained genome showed that only strain PAMC28711 carries three metabolic pathways of trehalose (trehalose biosynthesis

pathways; TPS/TPP, TS, and TreY/TreZ) as well as the trehalose degradation pathway, TreF. The trehalose degradation pathway includes a trehalase enzyme, which belongs to the CAZyme subfamily GH37, and is only involved in strain PAMC28711. Based on the results of AZCL screening, the strain PAMC28711 thrived at 25 °C, even though it was isolated from cold-adapted lichen. Significantly, the strain PAMC28711 has the potential to survive in diverse temperatures ranging from psychrophilic to mesophilic habitats, which explains the role of trehalose metabolism in this strain. In addition, this finding suggests that even when it was isolated from a polar region, the strain PAMC28711 survived temperature variation, which explains the existence of different pathways for trehalose synthesis in this strain. Among the six strains, the strain PAMC28711 shows one of the highest numbers of virulence proteins involved. The results show the limitations of bioinformatics tools used in this study for genome analysis, even though they are popular databases available online. The finding indicates that even though bioinformatics tools are essential for prediction or prognosis, they are not completely reliable. The predicted results can only be validated through experimental approach. The preliminary comparative study of *Variovorax* suggests the need for additional investigations into *V.* sp. PAMC28711 in the future.

Abbreviations

AA: Auxiliary Activity; ANI: Average Nucleotide Identity; AZCL: Azurine Cross-Linked; B's: Bennett's Agar; CAZyme: Carbohydrate-Active Enzyme; CBM: Carbohydrate-Binding Module; CE: Carbohydrate Esterase; GH: Glycoside Hydrolase; GT: Glycosyltransferase; KEGG: Kyoto Encyclopedia of Genes and Genomics; MY: Malt Yeast Media; MA: Marine Agar; NCBI: National Center for Biotechnology Information; PL: Polysaccharide Lyase; R2A: Reasoner's 2A Agar; SI: Supplementary Information; VFDB: Virulence Factor Database.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

1. Park and T.-J. Oh designed and supervised the project. P. Shrestha and S.-R. Han performed the experiments; P. Shrestha, S.-R. Han, J.H. Lee, H. Park, and T.-J. Oh wrote the manuscript. All authors discussed the results, commented on the manuscript, and approved the manuscript.

Funding

This research was a part of the project titled "Development of potential antibiotic compounds using polar organism resources (15250103, KOPRI Grant PM20030)", funded by the Ministry of Oceans and Fisheries, Korea.

Availability of data and materials

All data of this article can be found in the article itself and the data supplement available with the online version of this article (word files with Table S1, Table S2, and Table S3).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have declared that no conflict of interest exists.

Author details

[1] Department of Life Science and Biochemical Engineering, Graduate School, SunMoon University, Asan 31460, Korea. [2] Unit of Research for Practical Application, Korea Polar Research Institute, Incheon 21990, Korea. [3] Department of Polar Sciences, University of Science and Technology, Incheon 21990, Korea. [4] Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Korea. [5] Genome-based BioIT Convergence Institute, Asan 31460, Korea. [6] Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan 31460, Korea

*Co-corresponding authors: H. Park & T.-J. Oh

1. Park, Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Korea. Tel: +82 2 3290 3051; E-mail: hpark@korea.ac.kr

T.-J. Oh, Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan 31460, Korea. Tel.: +82 41 530 2677; E-mail: tjoh3782@sunmoon.ac.kr

References

1. Willems A, Gillis M. Comamonadaceae. Bergey's Manual of Systematics of Archaea and Bacteria. John Wiley & Sons, Inc. 2015.
2. Satola B, Wübbeler JH, Steinbüchel A. Metabolic characteristics of the species *Variovorax paradoxus*. Appl Microbiol Biotechnol. 2013;97:541–60.
3. Stan-Lotter H, Fendrihan S. Adaption of Microbial Life to Environmental Extremes. Novel Research Results and Application. 2nd, Springer Verlag. 2017.
4. Han SR, Lee JH, Kang S, Park H, Oh TJ. Complete genome sequence of opine-utilizing *Variovorax* strain PAMC28711 isolated from an Antarctic lichen. J Biotechnol. 2016;225:46–7.
5. Pearce DA. Extremophiles in Antarctica: Life at low temperatures. In: Stan-Lotter H, Fendrihan S. (eds) Adaption of Microbial Life to Environmental Extremes. Springer, Vienna. 2012.
6. Pearce DA. Extremophiles in Antarctica: Life at low temperatures, Adaption of Microbial Life to Environmental Extremes, Novel Research Results and Application. 2nd, Springer Verlag. 2017.
7. Cary SC, McDonald IR, Barrett JE, Cowan DA. On the rocks: the microbiology of Antarctic Dry Valley soils. Nat Rev Microbiol. 2010;8:129–38.
8. Collins T, Margesin R. Psychrophilic lifestyles: mechanisms of adaptation and biotechnological tools. Appl Microbiol Biotechnol. 2019;103:2857–
9. Bates ST, Cropsey GW, Caporaso JG, Knight R, Fierer N. Bacterial communities associated with the lichen symbiosis. Appl Environ Microbiol. 2011;77:1309–14.
10. Aschenbrenner IA, Cernava T, Berg G, Grube M. Understanding microbial multi-species symbioses. Front Microbiol. 2016;7:180.
11. Berlemont R. Distribution and diversity of enzymes for polysaccharide degradation in fungi. Sci Rep. 2017;7:222.
12. Park SH, Lee B, Park KH. Extremophilic carbohydrate active enzymes (CAZymes). J Nutr Health Food Eng. 2017;7:230–7.
13. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490–
14. Imran M, Ghadi SC. Role of carbohydrate-active enzymes (CAZymes) in production of marine bioactive oligosaccharides and their pharmacological applications. Enzymatic Technologies for Marine Polysaccharides, CRC Press, Taylor & Francis Group. 2019;357.
15. Casadevall A. The Pathogenic Potential of a Microbe. mSphere. 2017;2:e00015-17.
16. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: A comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019.47:D687–92.
17. Forrellad MA, Klepp LI, Gioffré A, Sabio y García J, Morbidoni HR, de la Paz Santangelo M, Cataldi AA, Bigi F. Virulence factors of the *Mycobacterium tuberculosis* Virulence 2013;4:3–66.
18. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees.Mol Biol Evol. 1987;4:406–
19. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 1985;39:783–91.
20. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method.Proc Natl Acad Sci. USA 2004;101:11030–
21. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
22. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics 2008;9:75.
23. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, Paley S, Subhraveti P, Karp PD. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 2018;46:D633–9.
24. Barraza A, Sánchez F. Trehalases: A neglected carbon metabolism regulator? Plant Signal Behav. 2013;8:e24778.
25. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, Latendresse M, Muñiz-Rascado L, Ong Q, Paley S, Peralta-Gil M, Subhraveti P, Velázquez-Ramírez DA, Weaver D, Collado-Vides J, Paulsen I, Karp PD. The EcoCyc Database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res. 2017;45:D543–50.
26. Horlacher R, Uhland K, Klein W, Ehrmann M, Boos W. Characterization of a cytoplasmic trehalase of *Escherichia coli*. J Bacteriol. 1996;178:6250–7.
27. Kopp M, Müller H, Holzer H. Molecular analysis of the neutral trehalase gene from *Saccharomyces cerevisiae*. J Biol Chem. 1993;268:4766–74.
28. Schwarz S, Van Dijck P. Trehalose metabolism: A sweet spot for *Burkholderia pseudomallei* Virulence 2017;8:5–7.
29. Avonce N, Mendoza-Vargas A, Morett E, Iturriag G. Insights on the evolution of trehalose biosynthesis. BMC Evol Biol. 2006;6:109.
30. Brunecky R, Chung D, Sarai NS, Hengge N, Russell JF, Young J, Mittal A, Pason P, Wall TV, Michener W, Shollenberger T, Westpheling J, Himmel ME, Bomble YJ. High activity CAZyme cassette for improving biomass degradation in thermophiles. Biotechnol Biofuels 2018;11:22.
31. Wilkens C, Busk PK, Pilgaard B, Zhang WJ, Nielsen KL, Nielsen PH, Lange L. Diversity of microbial carbohydrate-active enzymes in Danish anaerobic digesters fed with wastewater treatment sludge. Biotechnol Biofuels 2017;10:158.
32. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2017;45:D535–42.
33. Lee YJ, van Nostrand JD, Tu Q, Lu Z, Cheng L, Yuan T, Deng Y, Carter MQ, He Z, Wu L, Yang F, Xu J, Zhou J. The PathoChip, a functional gene array for assessing pathogenic properties of diverse microbial communities. ISME J. 2013;7:1974–84.

34. Lee IC, Kim YO, Park SC, Chun JS. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016;66:1100–3.
35. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46:W95–101.
36. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.
37. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019;47:D590–5.

Figures

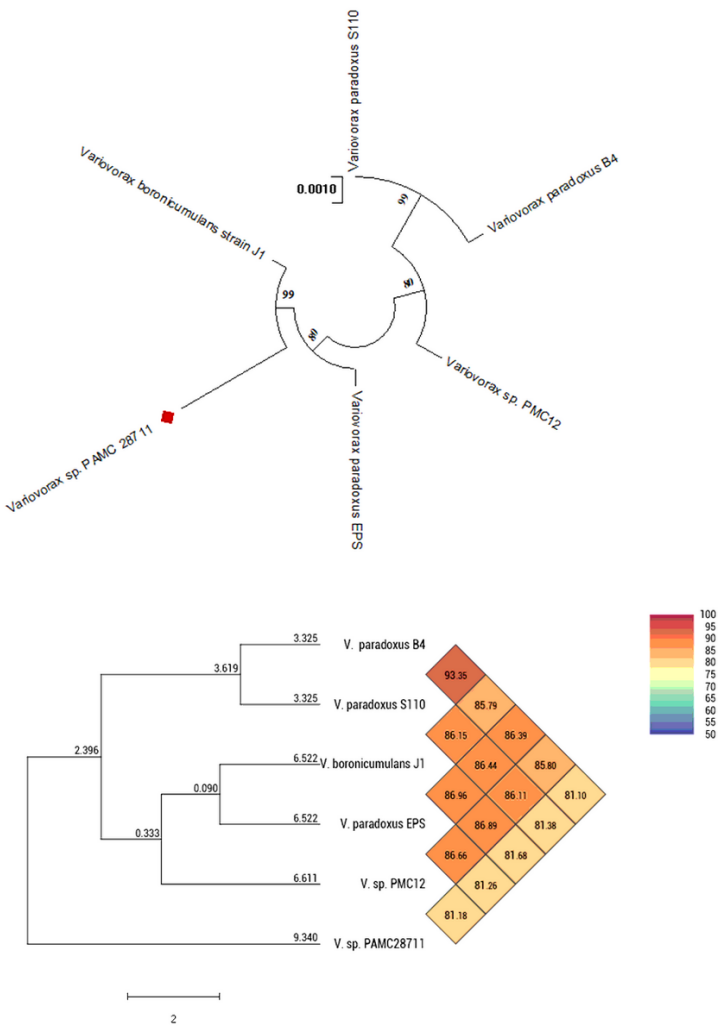
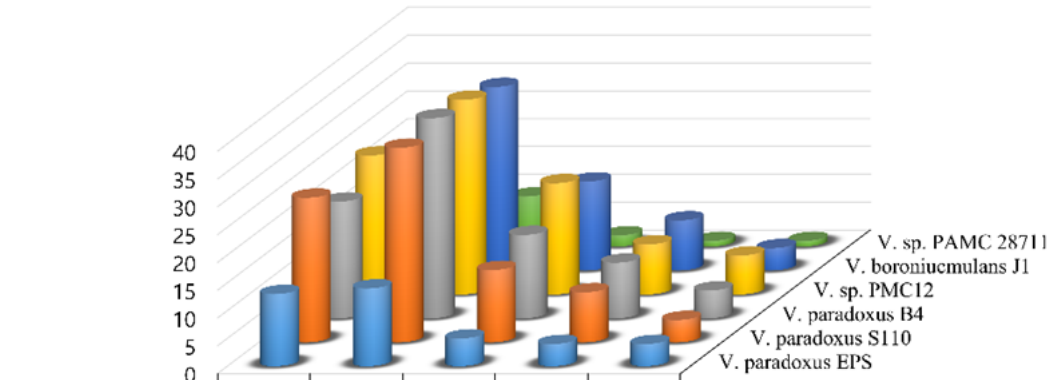


Figure 1

A. Phylogenetic tree analysis of 16S rRNA gene sequences of the complete Variovorax genome. The bootstrap value represents the confidentiality of the generated branches. The branch was unrooted. The marker represents the strain under study. B. Heat map of average nucleotide identity (ANI) of a complete genome sequence in the six strains of Variovorax. ANI values (%) of each strain in two-genome sequence comparison are shown.



	GH	GT	CE	AA	CBM
V. paradoxus EPS	13	14	5	4	4
V. paradoxus S110	26	35	13	9	4
V. paradoxus B4	21	36	15	10	5
V. sp. PMC12	25	35	20	9	7
V. boroniucmulans J1	19	33	16	9	4
V. sp. PAMC 28711	10	9	2	1	1

Figure 2

Comparison of CAZyme families in the complete genome of Variovorax species. GH, Glycoside Hydrolases; GT, Glycosyl Transferases; CE, Carbohydrate Esterases; AA, Auxiliary Activities; and CBM, Carbohydrate-Binding Modules.

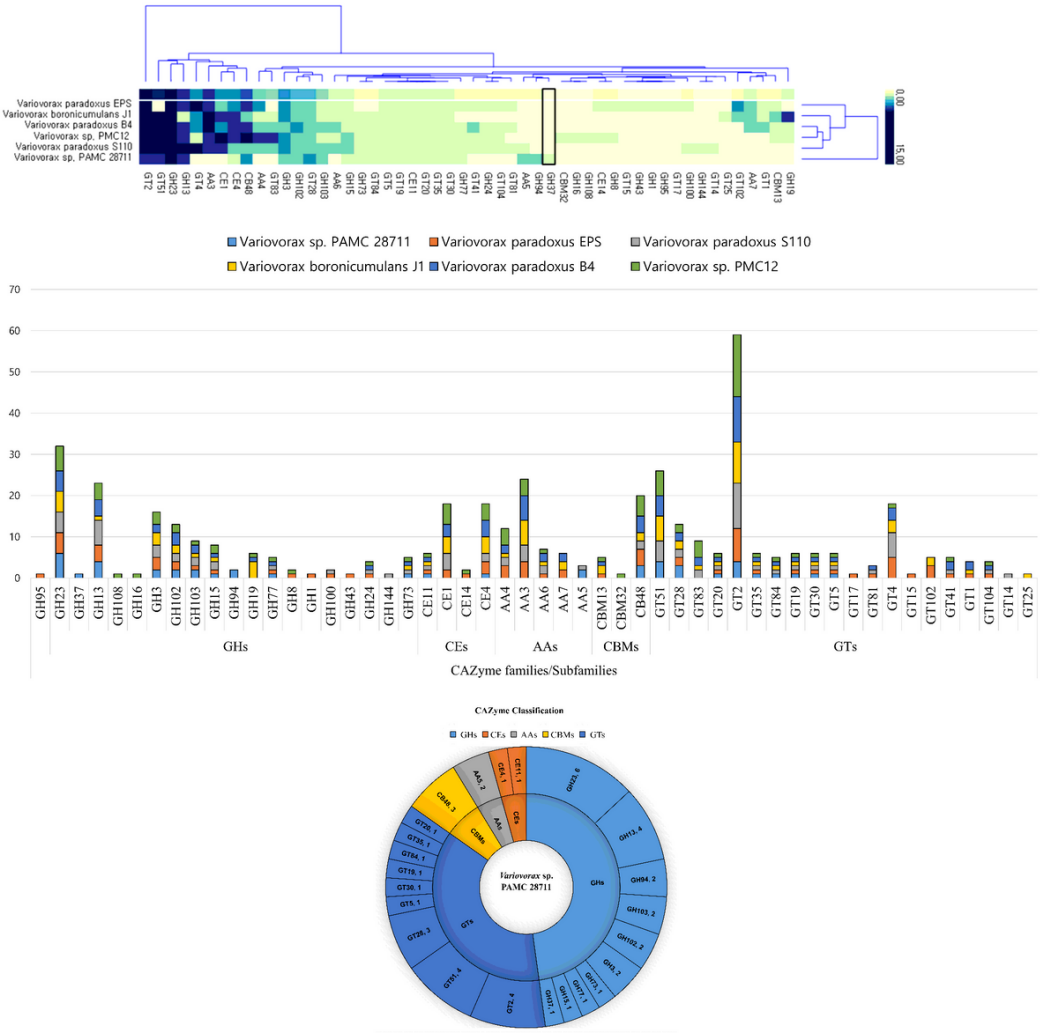


Figure 3

A. Heat map of CAZyme families and subfamilies distributed in the complete genome of various strains of Variovorax. Dark blue color represents the abundance of each family, whereas the light color indicates the specific subfamilies that are fewer in number or absent in a particular strain. B. Comparison of CAZyme subfamilies among the complete genome of six Variovorax strains. C. Sunburst chart shows CAZyme families and subfamilies involved in Variovorax sp. PAMC28711. The classification of CAZyme families is represented by light blue, GHs; dark blue, GTs; orange, CEs; yellow, CBMs; and grey, AAs.

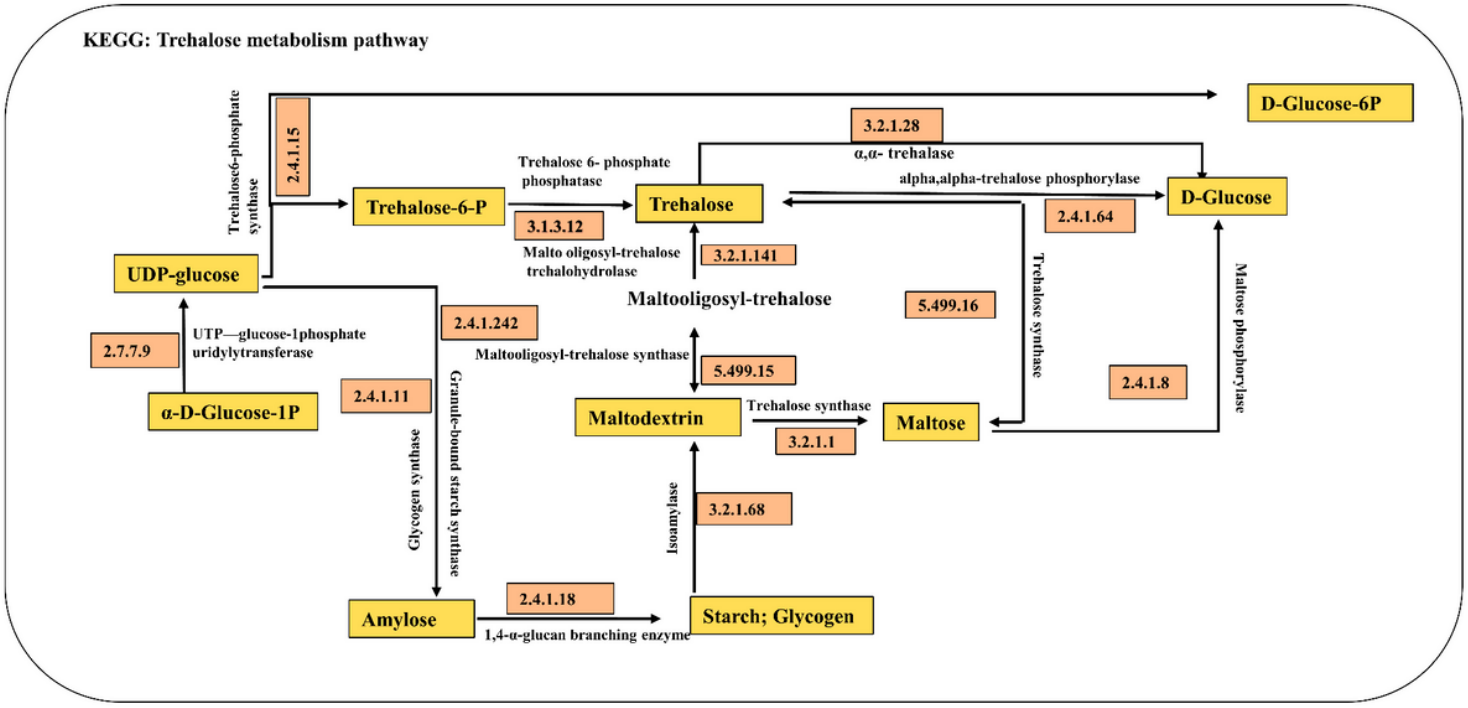


Figure 4

The trehalose metabolic pathway in bacteria (KEGG Map).

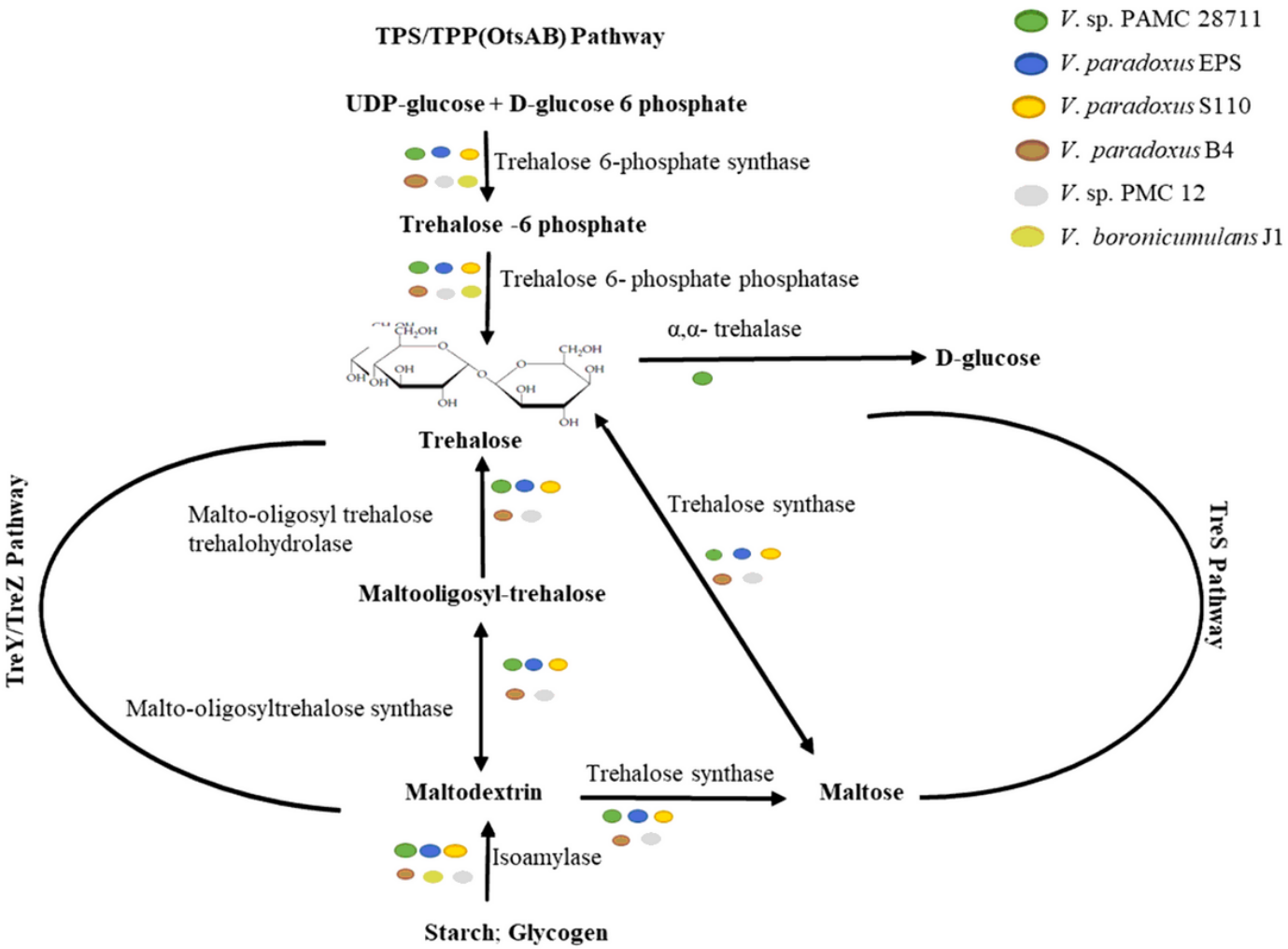


Figure 5

The three metabolic pathways of trehalose identified in different strains of the complete Variovorax genome.

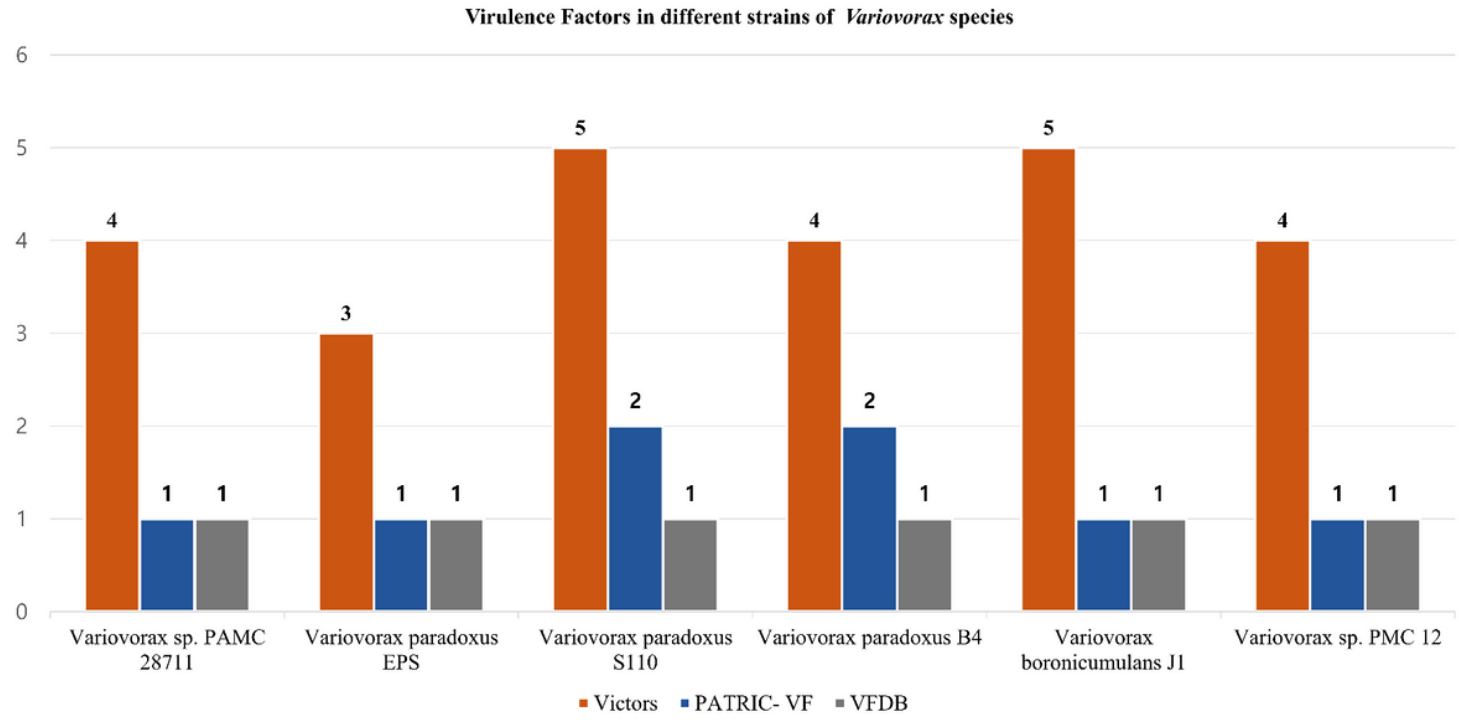


Figure 6

Virulence factors present in various strains of the complete genome of Variovorax species based on the PATRIC database.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytables.docx