

# Characterization of Saccharimonadia bacteria in Lakes Baikal and Khubsugul: taxonomic and functional diversity

Original Article

LIMNOLOGY  
FRESHWATER  
BIOLOGY

Gutnik D.I.<sup>1\*</sup>, Potapov S.A.<sup>1</sup>, Krasnopeev A.Y.<sup>1</sup>, Tikhonova I.V.<sup>1</sup>, Bukin Yu.S.<sup>1</sup>, Gladkikh A.S.<sup>2</sup>, Klyuchnikova E.O.<sup>2</sup>, Sbarzaglia V.A.<sup>2</sup>, Aslamov I.A.<sup>1</sup>, Belykh O.I.<sup>1</sup>

<sup>1</sup>Limnological Institute Siberian Branch of the Russian Academy of Sciences, Ulan-Batorskaya Str., 3, Irkutsk, 664033, Russia

<sup>2</sup>Saint Petersburg Pasteur Institute, 14 Mira Street, Saint Petersburg, 197101, Russia

**ABSTRACT.** Ultra-small bacteria are ubiquitous in the environment, but for some time they have remained out of scientific focus. One of the reasons is their small size. Ultramicrobacteria, which are defined as bacteria with a cell size of less than 0.1  $\mu\text{m}^3$ , can pass through a filter with a pore size of 0.2  $\mu\text{m}$ . Moreover, there are some difficulties associated with their cultivation. Currently, relatively pure cultures have been obtained, but mostly for dominant phyla. Rare phyla are represented only by single strains. At the same time, ultramicrobacteria are found in a wide range of environments, and their high-ordered, small genomes and unique physiological features attract the increasing attention of researchers. The development of metagenomics allows us to partially overcome cultivation limitations, and in this article, we present metagenomic assemblies derived from water samples of the largest oligotrophic lakes on the Earth, Baikal (Russian Federation) and Khubsugul (Mongolia). In this study, we carried out the taxonomic and functional analysis of reconstructed genomes of Saccharimonadia class bacteria (phylum Patescibacteria), which is intended to expand our knowledge of rare representatives of microbial communities and supplement the already available information regarding the inhabitants of the oldest ecosystems.

**Keywords:** Saccharimonadia, ultramicrobacteria, Baikal, Khubsugul, metagenomics, metabolism

**For citation:** Gutnik D.I., Potapov S.A., Krasnopeev A.Y., Tikhonova I.V., Bukin Yu.S., Gladkikh A.S., Klyuchnikova E.O., Sbarzaglia V.A., Aslamov I.A., Belykh O.I. Characterization of Saccharimonadia bacteria in Lakes Baikal and Khubsugul: taxonomic and functional diversity // Limnology and Freshwater Biology. 2024. - № 4. - P. 908-924. DOI: 10.31951/2658-3518-2024-A-4-908

## 1. Introduction

The superphylum Patescibacteria, also known as the Candidate Phyla Radiation (CPR) (Rinke et al., 2013), according to The Genome Taxonomy Database (GTDB) (Parks et al., 2022), includes 30 classes that comprise more than 15% of all bacterial phyla (Brown et al., 2015; Fujii et al., 2022). Patescibacteria inhabit a wide variety of environments: aquifer sediments and reservoirs, water treatment plants, activated sludge, gold mines, polar and rainforest soils, human saliva, in association with sponges and cockroaches, predominating in many of them (Brown et al., 2015; Hug et al., 2016; Fujii et al., 2022).

Currently, little is known about the functional characteristics of these bacteria, and there is no consensus on whether all CPR representatives are endosymbionts that rely on other bacteria for their vital activity, or whether some of them are capable of free-living.

Knowledge about this group of bacteria is mainly limited because currently only a few cases of cultured Patescibacteria have been described; this fact is largely due to their symbiotic/parasitic nature (Soro et al., 2014; He et al., 2021; Ibrahim et al., 2021; Yakimov et al., 2022). Some of the cultured species (e.g., TM7x) (He et al., 2015) belong to the Saccharimonadia class. Representatives of this class were first discovered by analyzing partial 16S rDNA sequences from a peat bog in Germany (Rheims et al., 1996), and they are among the best known. Interest in studying Saccharimonadia is driven mainly by their putative role in the development of oral diseases. Saccharimonadia (previously described as TM7) cultured under laboratory conditions clearly demonstrated a symbiotic relationship with actinobacteria, and a similar pattern was observed for other few cultured CPRs. Indeed, known Patescibacteria typically possess a small genome size (up to 1.0 Mb) (Lemos et al., 2019; Nakai, 2020) and incomplete biosynthe-

\*Corresponding author.

E-mail address: [gutnik@lin.irk.ru](mailto:gutnik@lin.irk.ru) (D.I. Gutnik)

**Received:** August 08, 2024; **Accepted:** August 21, 2024;

**Available online:** August 30, 2024

© Author(s) 2024. This work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License.



sis pathways (Castelle et al., 2018). A combination of these characteristics suggests a host-associated, symbiotic or parasitic lifestyle.

However, the unique habitat conditions in large deep-water oligotrophic lakes, such as Baikal and Khubsugul, may explain the decrease in genome size and the presence of functional specificities. These ecosystems are also of interest in terms of bacterial evolution, taking into account the age and common origin of the two tectonic lakes in the Baikal rift zone. In addition, samples from Lake Baikal were collected at different depths: in the upper water layers, i.e. in the photic layer (up to 25 m) and at depths of 1250-1350 m, which also potentially affects bacterial metabolism.

Therefore, this study aims to analyze and compare the functional features and lifestyle of *Saccharimonadia* obtained from Baikal and Khubsugul water samples using metagenomic assemblies and genome reconstruction.

## 2. Materials and methods

Water samples were collected in Lake Baikal at the central station of the Maloye More strait, at the central stations of the transects Listvyanka settlement-Tankhoi settlement, Ukhan Cape-Tonky Cape, Elokhin Cape-Davsha settlement, 7 km and 3 km from the Listvyanka settlement, 3 km from the Turka settlement, and 3 km from Elokhin Cape in March-September 2018-2022. During open water periods, sampling was conducted at the depths of 0, 5, 10, 15, 20, 25, 50 m by the SBE-32 bathometer system (Carousel Water Sampler, Sea Bird Electronics Inc., USA). During the freeze-up period, samples were taken under the ice using Niskin bathometers. In Lake Khubsugul, water samples were obtained 3 km from the Khankha settlement on October 13, 2023. Samples were collected from the water layers of 0, 10, 25, 40, 50 and 70 m using the bathometer from the boat.

Water samples were sequentially filtered through polycarbonate filters with pore diameters of 0.4  $\mu\text{m}$  and 0.2  $\mu\text{m}$  (Millipore, USA) to remove large forms of plankton. The filtrates were concentrated using a VivaFlow 200 tangential flow ultrafiltration system (Sartorius, Germany) to a final volume of  $\sim 20$  ml. Samples were further concentrated using a Vivaspin Turbo 15 (50 kDa, Sartorius, Germany) tangential flow ultrafiltration system to a volume of  $\sim 100$   $\mu\text{L}$ .

DNA was extracted using the standard phenol-chloroform method. DNA libraries including A\_63, A\_68, A\_9 were sequenced using the Illumina MiSeq platform. Metagenomic data was deposited in bioproject PRJNA1006167. DNA sequencing for I\_78 was performed on the DNBSEQ-G50 platform (MGISEQ-200).

Additionally, we used eight sets of raw metagenomic reads generated by Shotgun sequencing and placed for public access in bioprojects (PRJNA396997, PRJNA521725, PRJNA615165). The sequencing material was taken from Lake Baikal water samples collected in summer, on July 12-13, 2018, 8 km from the Listvyanka settlement at four depths (5, 20, 1250, and 1350 m). Winter water samples were taken from 5 and 20 m depths on March 14, 2016 and from 1250 and 1350 m depths on March 29, 2018, 7 km from the Listvyanka settlement (Cabello-Yeves et al., 2018; Cabello-Yeves et al., 2020; Coutinho et al., 2020). The information about the relationship between the names of the reconstructed genomes and the materials for their assemblies is provided in Table 1.

The reconstruction process of the A\_63, A\_68, and A\_9 genomes included sequencing quality assessment using the MultiQC v. 1.12 program (Ewels et al., 2016), removal of adaptors with the Trim Galore v. 0.6.5 program (Krueger et al., 2023), assembly of reads into contigs using SPAdes v. 3.15.5 (Bankevich et al., 2012), and alignment of reads to contigs via BWA v. 0.7.17 (Li and Durbin, 2010) and the Samtools package v. 1.18 (Danecek et al., 2021). Metagenomic binning and isolation of genomes from the assembly were performed using the MetaWRAP v. 1.3 software package (Uritskiy et al., 2018). Assembly and analysis of P\_794 and I\_78 genomes comprised sequencing quality assessment by the FastQC v. 0.12.1 program (Andrews et al., 2010), removal of adapters and short reads (less than 141 for P\_794 and 90 for I\_78) with the application of Trimmomatic v. 0.39 (Bolger et al., 2014), and purification from eukaryotic DNA by the Kraken 2 v. 2.1.3 software package (Wood et al., 2019; Lu et al., 2022).

To assemble contigs containing P\_794, an external assembly was performed by merging the reads of the corresponding direction from eight samples; thereby, the result was one merged set of paired-end reads, which were then assembled by the MEGAHIT assembler (Li et al., 2015). For external assembly, alignment of reads to contigs was performed using the Bowtie 2 v. 2.5.4 (Langmead et al., 2019) and the Samtools

**Table 1.** Sampling conditions and data sources

	P_794	A_63	A_68	A_9	I_78
Location	Baikal	Baikal			Khubsugul
Depths, m	5, 10, 1250, 1350	0, 5, 10, 15, 20, 25, 50			0, 10, 25, 40, 50, 70
Time Period	July-March, 2016-2018	March-September, 2018-2022			October, 2023
Source of Data	Data obtained from bioprojects: PRJNA396997, PRJNA521725, PRJNA615165 (Cabello-Yeves et al., 2018; Cabello-Yeves et al., 2020; Coutinho et al., 2020)	Data were collected and processed independently. Row reads deposited in the bioproject PRJNA1006167			Data were collected and processed independently

v. 1.20 package (Danecek et al., 2021). Based on the obtained coverage statistics, filtering and removal of short and low-coverage contigs were performed using the R script. Final contig assemblies were performed using the SqueezeMeta pipeline (Tamames and Puente-Sánchez, 2019), applying coassembly mode for P\_794, including all eight sets of reads and external assembly, and sequential mode for I\_78. The assembler used for P\_794 was MEGAHIT (Li et al., 2015), and for I\_78–SPAdes v. 3.15.5 (Bankevich et al., 2012). Binning was performed using CONCO v. 1.1.0 (Alneberg et al., 2014), MaxBin v. 2.2.7 (Wu et al., 2016), and MetaBAT2 v. 2.12.1 (Kang, 2015). The integration of the binning results was performed in DAS Tool v. 1.1.7 (Sieber, 2018).

Completeness of the final genomes were assessed using CheckM2 v. 1.0.2 (Chklovski et al., 2023). Open reading frames (ORFs) were detected using Prodigal v.2.6.3 (Hyatt et al., 2010). Gene annotation was performed using Bakta v1.9.2 (Schwengers et al., 2021) with default settings. KEGG annotation and assignment of KO numbers to proteins was performed by the KofamKOALA service (Aramaki et al., 2020). Genomes were taxonomically classified using the Genome Taxonomy Database Toolkit (GTDB-Tk) version 2.4.0 (Chaumeil et al., 2020). MacSyFinder v. 2.1.4 (Néron et al., 2023) and TXSScan v. 1.1.3 models (Abby et al., 2016) were used to search for protein secretion systems.

### 3. Results and discussion

#### 3.1. Genome description and taxonomic classification

Five *Saccharimonadia* genomes were assembled based on three metagenomic datasets. All of them show

a relatively high level of assembly completeness of 84.9-97.6%, with contamination below 2.6%, except for genome I\_78 where it was 7.0% (Table 2). Their coding density (median 90.5%; range 89.6%-91.8%) and GC content (median 42.2%; range 40,7-45,0%) are consistent with free-living organisms or facultative intracellular or extracellular symbionts/parasites (Chiriack et al., 2022).

P\_794 represents the largest genome with the maximum number of CDSs (1074) and a rather extensive range of functions for a member of the class. Furthermore, based on the GTDB taxonomic database search, it can be identified only up to genus, and the similarity with the nearest reference organism according to ANI is only 78.94. 93% of TPM (transcripts per kilobase million, the number of reads normalized by the length of each read multiplied by a million) belonging to this bin corresponds to the summer samples of Baikal water from 1250 and 1350 m depths.

Genomes A\_63, A\_68, and A\_9 were also obtained from Baikal samples at the upper water layers. The size of their genomes, as well as the number of CDSs, is approximately the same order of magnitude. Taxonomic identification by GTDB revealed a fairly close similarity in ANI for A\_63 and A\_9, identifying the genus of all genomes and, in the case of A\_63, the species as well.

I\_78 is a genome assembled from sequencing data of surface water samples from Lake Khubsugul. It is the second largest of those presented in the paper and the most contaminated. At the same time, the number of CDS in it amounts to 675. Interestingly, according to the taxonomic analysis, the most similar *Saccharimonadia* belonging to the same genus, according to the preliminary classification, are A\_68 from Baikal and I\_78 from Khubsugul samples.

**Table 2.** Description and taxonomic assignment of reconstructed genomes

Name	P_794	A_63	A_68	A_9	I_78
Genome Size, bp	984 478	655 626	594 118	612 023	718 727
Completeness, %	91.6	97.6	91.2	84.9	93.9
Contamination, %	2.08	-	0.29	2.56	7.00
GC Content, %	45	41	42	41	42
Contig N50, bp	10 865	57 148	35 054	12 346	3 970
Coding Density, %	89.6	90.7	89.6	91.8	90.5
CDS	1 016	686	594	645	675
Taxonomy					
Kingdom	Bacteria	Bacteria	Bacteria	Bacteria	Bacteria
Phylum	Patescibacteria	Patescibacteria	Patescibacteria	Patescibacteria	Patescibacteria
Class	Saccharimonadia	Saccharimonadia	Saccharimonadia	Saccharimonadia	Saccharimonadia
Order	Saccharimonadales	CAILAD01	CAILAD01	Saccharimonadales	CAILAD01
Family	UBA1547_A	BJGX01	CAILAD01	AWTP1-31	CAILAD01
Genus	2011-GWC2-44-17	CALRAY01	CAILAD01	CAJAXB01	CAILAD01
Species	-	CALRAY01	-	-	-
Closest reference	GCA_016199685.1	GCA_943353705.1	GCA_013816445.1	GCA_943353855.1	GCA_903832035.1
Closest ANI	78.94	97.66	78.36	92.13	77.02

Classification to species level using GTDB is realized only for A\_63. The two genomes, P\_794 and A\_9, belong to the order Saccharimonadales and to uncultivated families and genera. A\_63, A\_68 and I\_78 are assigned to uncultivated taxa at all levels.

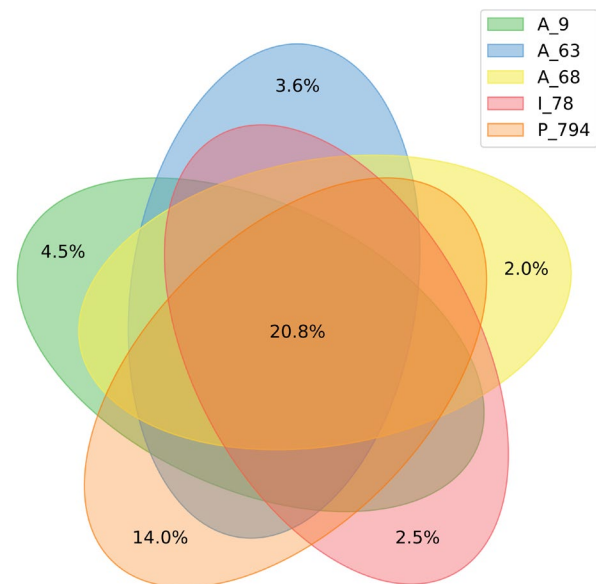
### 3.2. Annotation and functional analysis of genomes

Comparison of the annotated genomes by functional orthologs of the KEGG database showed that the number of common genes is 20.8% (Fig. 1). At the same time, the most unique gene set for this sample is possessed by P\_794, which has 14.0% of predicted genes that do not overlap with genes in the other studied genomes.

Functional analysis established a common mechanism, the most plausible for all genomes, for obtaining energy in the form of ATP—catabolism of carbohydrates with pyruvate synthesis. In this case, the production of pyruvate follows the de Vries-Stouthamer hexose monophosphate pathway or, in other words, the bifido-shunt (Pinevich, 2007). All the genomes have practically the entire set of enzymes necessary for its implementation: fructo-bisphosphate aldolase, transaldolase, and transketolase. It is remarkable that even in the absence of enzymes, a unified strategy is traced. The genes responsible for the production of phosphoglyceromutase, which is involved in the transformation of 3-phosphoglycerate into 2-phosphoglycerate and is necessary for the further pyruvate synthesis, are not present in all Saccharimonadia studied here.

From the stage of functional category analysis (Fig. 2), a distinctive feature of P\_794 was already apparent, namely, the presence of a cluster of genes responsible for the cleavage of sucrose and starch oligo- and polysaccharides to D-glucose, in particular through the production of  $\alpha$ -glucosidase, alpha-amylase, maltase-glucoamylase, etc. The second genome, in terms of the number of predicted genes, involved in carbohydrate metabolism is A\_63. However, its hexose monophosphate pathway begins with the cleavage of glucose-6-phosphate or fructose-6-phosphate. In other reviewed genomes, the functional set of carbohydrate metabolism is even smaller: A\_9, A\_68, and I\_78 are only capable of fructose-6-phosphate conversion.

Notably, in all studied Saccharimonadia, fermentation is not performed; the process is terminated by the formation of pyruvate and one ATP molecule per molecule of oxidized hexose. Normally, pyruvate, which is the most important intermediate of fermentation, undergoes oxidative carboxylation via one of the four pathways (Pinevich, 2007). In the case of the investigated genomes, the enzymes required for fermentation are absent in all genomes, except for A\_63. For this organism, lactate dehydrogenase is predicted to enable pyruvate-based lactic acid fermentation, which, however, does not find confirmation in the annotation by the KEGG. Moreover, other types of fermentation require the engagement of vitamins, the production of which is absent in all five microorganisms. At the same time, this practice is widespread in microbial commu-



**Fig.1.** Venn diagram showing overlaps in Saccharimonadia genes annotated using the KEGG Orthology database

nities where there are major producers that supply the demand of other bacteria for essential vitamins for metabolic reactions.

Another distinguishing characteristic of P\_794 is the increased number of genes involved in oxidative phosphorylation in comparison with others. Thus, a complete oxidase complex consisting of five subunits (CyoA, CyoB, CyoC, CyoD, and CyoE) was found in it, which is an aerobic respiratory chain reacting with oxygen, reducing it to water with simultaneous transfer of four protons across the membrane (Chepuri et al., 1990). The presence of the enzyme FAD-dependent oxidoreductase also confirms the respiration ability. The A\_9 genome also contains genes that are part of the photosynthesis and oxidative phosphorylation complexes, but all are associated with the F-type ATPase, a multi-subunit enzyme complex whose role changes under aerobic and anaerobic conditions (Walker, 2013). Under anaerobic conditions, enzymes reverse the reaction and generate proton motive force by utilizing the energy released during hydrolysis of ATP (Walker and Dickson, 2006).

In general, the investigated Saccharimonadia do not possess synthesis of all essential nucleotides, amino acids, lipids, vitamins, cofactors, or the complete tricarboxylic acid (TCA) cycle, which is consistent with the results obtained previously. Biosynthesis of secondary metabolites is also barely represented. Although there are individual variations in the presence of certain genes, which, nevertheless, do not provide a complete cycle for the production of compounds.

According to studies some Saccharimonadia demonstrate the ability to produce rhodopsin (Jaffe et al., 2022), but in the current samples, the genes responsible for its synthesis are absent. Meanwhile, regions encoding the production of heliorhodopsin, a new family of rhodopsins, are predicted in four genomes (A\_9, A\_63, A\_68, and I\_78). Heliorhodopsin has been recently discovered, and is still poorly studied (Pushkarev et al., 2018). At the same time, there is a current assumption

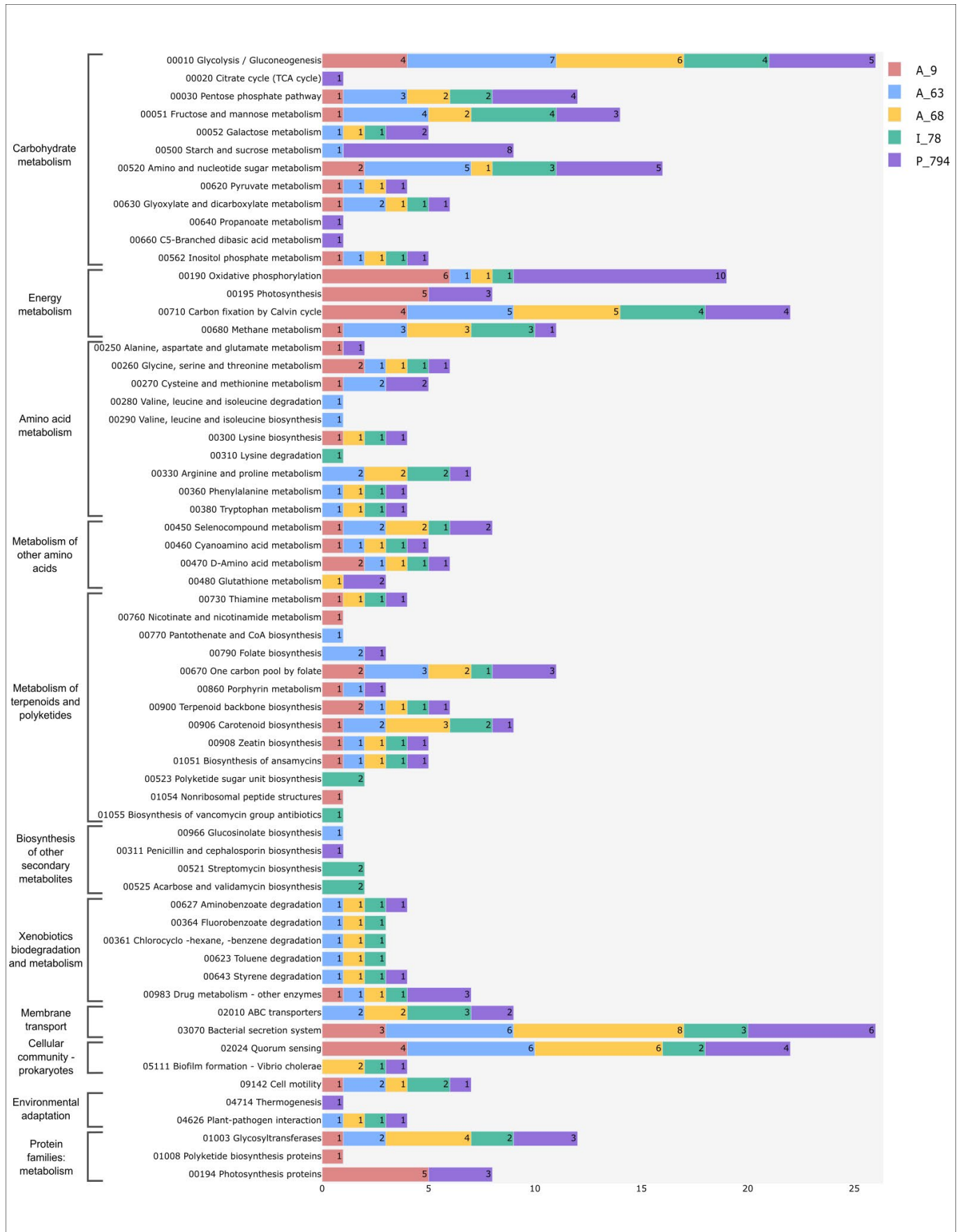


Fig.2. Number of genes belonging to Saccharimonadia in the categories defined according to the KEGG PATHWAY classification

that the presence of heliorhodopsin is not possible in organisms with double membranes and requires direct interaction with the environment for its functioning (Flores-Urbe et al., 2019).

Saccharimonadia also show differences in the mechanisms of adaptation to the environment; A\_68 contains the maximum set of genes for the synthesis of bacterioruberin, which is responsible for membrane adaptation to low temperatures (Flegler and Lipski, 2022). A\_9, A\_68, and P\_794 have genes of superoxide dismutase. A\_63, A\_68 and I\_78 have a glutathione-S-transferase gene. Both enzymes are required to cope with oxygen stress, which is relevant for lakes with high oxygen content, such as Baikal and Khubsugul (Scott et al., 1987; Röth et al., 2011).

Being auxotrophs in many categories, Saccharimonadia require alternative ways to obtain the compounds necessary for life activity, suggesting their involvement in symbiotic or parasitic relationships with other organisms. When such relationships occur, they are thoroughly regulated through protein secretion; therefore, secretion systems must include mechanisms for translocation of secreted proteins (effectors and toxins) across the host plasma membrane.

The type 2 secretion system (T2SS) is not fully represented in genomes. The T2SS is Sec-dependent and includes pseudopiles, outer and inner membrane complexes (Korotkov et al., 2012; Lomovatskaya and Romanenko, 2020). Only A\_9 has one of the three proteins of the inner membrane platform of the T2SS—GspF. The remaining Saccharimonadia encode only GspG, which belongs to the pseudopiles. Furthermore, none of the outer membrane complex proteins, such as GspD and GspS, were identified, raising doubts about the functional ability of the T2SS. At the same time, the SecA gene (A\_9, A\_63, A\_68, and P\_794), responsible for ATP hydrolysis to trigger protein translocation, as well as SecD-G and SecY genes encoding the complete membrane complex, are present in almost all genomes (Schneewind and Missiakas, 2012; Chiriac et al., 2022). Additionally, A\_63, A\_68, and P\_794 contain the YidC protein, which mediates the incorporation of Sec proteins into the membrane (Kuhn et al., 2003).

As a plausible mechanism for enabling symbiotic or parasitic lifestyles, Saccharimonadia may utilize a type IV protein (T4aP) secretion system that forms a pili attached to the host outer membrane (Ibrahim et al., 2021; Hendrickson et al., 2022). Type IV pili are surface appendages involved in many processes, including adhesion to biotic and abiotic surfaces, aggregation, DNA and phage uptake, specific motility and electron transfer (Tammam et al., 2013; Berry and Pelicic, 2015). Search with MacSyFinder and the TXSScan model set revealed the diversity of pili in the genomes. All Saccharimonadia, except for I\_78, contain clusters of major and minor type IV pili: pilABCDE, piMNO, pilT, and pilVWX. Another argument in favor of a Saccharimonadia dependent lifestyle is the presence of ABC transporters, including those with specific functions, namely transport of ATP, cobalt, iron, and sugar as well as those with unknown functional properties (Fujii et al., 2022).

## 4. Conclusions

The development of metagenomic analysis techniques has greatly expanded our ability to study uncultured bacteria. Although the completeness and contamination of metagenomic data are limiting factors to take into account, especially in the functional analysis of poorly characterized organisms. Nevertheless, even a first approximation, representing the study of five metagenomic assemblies, provided previously inaccessible information on ultramicrobacteria inhabiting the unique ecosystems of oligotrophic lakes Baikal and Khubsugul.

Preliminary taxonomic analysis showed that the database contains a sufficiently close reference organism only for one of the studied genomes; for all the others, the similarity by ANI is very distant. Moreover, the investigated genomes also do not demonstrate taxonomic similarity among themselves, except for A\_68 and I\_78 presumably belonging to the same genus. Functional comparison also showed differences in genes. Only 20.8% of them are shared, while P\_794 contains 14% of unique genes different from the class members.

Functional comparison indicates rather extensive, especially in the context of a small genome, differences in genome metabolism. Thus, P\_794 possesses abilities to degrade sucrose and starch, and is also suspected to be a facultative anaerobe capable of oxygen respiration due to the presence of a complete oxidase complex. At the same time, all studied Saccharimonadia probably carry out energy metabolism by glycolysis through a metabolic loop, involving the de Vries-Stouthamer hexose monophosphate pathway. The fate of pyruvate in this pathway remains unknown because enzymes necessary for fermentation are missing.

Genomes also exhibit different ways of adaptation to the environment and protein secretion systems. The protein secretion systems, in particular, the wide representation of genes encoding T4aP, along with the limited metabolic potential and the presence of ABC transporters with unknown functions, suggest a symbiotic or parasitic lifestyle of the Saccharimonadia. This is supplemented by the observed coding density, which, although quite high, is still below the coding densities expected for free-living microorganisms, where values > 95% are common (Chiriac et al., 2022). Nevertheless, although the studied genomes possess incomplete pathways for biosynthesis of compounds required for cell construction and life activity, we cannot exclude the ability of Saccharimonadia to synthesize them de novo or obtain them from the environment via unknown pathways, and that such rationalized genome reduction did not arise as an adaptation to low or nutrient-limited environments.

## Acknowledgements

This study was supported by the State Project No. 0279-2021-0015. The authors thank the crews of the R/V “Vereshchagin” and “Titov” of the LIN SB RAS fleet, as well as I.V. Khanaev and M.M. Makarov for

their assistance in sampling. Sample collection from Lake Khubsugul was carried out according to the State Project No. 0279-2021-0004. Bioinformatic calculations were performed on the computational cluster at LIN SB RAS.

## Conflict of interest

The authors declare no conflict of interest.

## References

- Abby S.S., Cury J., Guglielmini J. et al. 2016. Identification of protein secretion systems in bacterial genomes. *Scientific Reports* 6: 23080. DOI: [10.1038/srep23080](https://doi.org/10.1038/srep23080)
- Alneberg J., Bjarnason B.S., De Bruijn I. et al. 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods* 11: 1144–1146. DOI: [10.1038/nmeth.3103](https://doi.org/10.1038/nmeth.3103)
- Andrews S., Krueger F., Segonds-Pichon A. et al. 2010. FastQC: a quality control tool for high throughput sequence data
- Aramaki T., Blanc-Mathieu R., Endo H. et al. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36: 2251–2252. DOI: [10.1093/bioinformatics/btz859](https://doi.org/10.1093/bioinformatics/btz859)
- Bankevich A., Nurk S., Antipov D. et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* 19: 455–77. DOI: [10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021)
- Berry J.-L., Pelicic V. 2015. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives FEMS. *Microbiology Reviews* 39: 134–154. DOI: [10.1093/femsre/fuu001](https://doi.org/10.1093/femsre/fuu001)
- Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. DOI: [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170)
- Brown C.T., Hug L.A., Thomas B.C. et al. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523: 208–211. DOI: [10.1038/nature14486](https://doi.org/10.1038/nature14486)
- Cabello-Yeves P.J., Zemskaya T.I., Rosselli R. et al. 2018. Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal. *Applied and Environmental Microbiology* 84: e02132-17. DOI: [10.1128/AEM.02132-17](https://doi.org/10.1128/AEM.02132-17)
- Cabello-Yeves P.J., Zemskaya T.I., Zakharenko A.S. et al. 2020. Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. *Limnology and Oceanography* 65: 1471–1488. DOI: [10.1002/lno.11401](https://doi.org/10.1002/lno.11401)
- Castelle C.J., Brown C.T., Anantharaman K. et al. 2018. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nature Reviews Microbiology* 16: 629–645. DOI: [10.1038/s41579-018-0076-20](https://doi.org/10.1038/s41579-018-0076-20)
- Chaumeil P.-A., Mussig A.J., Hugenholtz P. et al. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36: 1925–1927. DOI: [10.1093/bioinformatics/btz848](https://doi.org/10.1093/bioinformatics/btz848)
- Chepuri V., Lemieux L., Au D.C. et al. 1990. The sequence of the cyo operon indicates substantial structural similarities between the cytochrome o ubiquinol oxidase of *Escherichia coli* and the aa3-type family of cytochrome c oxidases. *Journal of Biological Chemistry* 265: 11185–11192. DOI: [10.1016/S0021-9258\(19\)38574-6](https://doi.org/10.1016/S0021-9258(19)38574-6)
- Chiriac M.-C., Bulzu P.-A., Andrei A.-S. et al. 2022. Ecogenomics sheds light on diverse lifestyle strategies in freshwater CPR. *Microbiome* 10: 84. DOI: [10.1186/s40168-022-01274-3](https://doi.org/10.1186/s40168-022-01274-3)
- Chklovski A., Parks D.H., Woodcroft B.J. et al. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods* 20: 1203–1212. DOI: [10.1038/s41592-023-01940-w](https://doi.org/10.1038/s41592-023-01940-w)
- Coutinho F. H., Cabello-Yeves P.J., Gonzalez-Serrano R. et al. 2020. New viral biogeochemical roles revealed through metagenomic analysis of Lake Baikal. *Microbiome* 8: 163. DOI: [10.1186/s40168-020-00936-4](https://doi.org/10.1186/s40168-020-00936-4)
- Danecek P., Bonfield J.K., Liddle J. et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10: giab008. DOI: [10.1093/gigascience/giab008](https://doi.org/10.1093/gigascience/giab008)
- Ewels P., Magnusson M., Lundin S. et al. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32: 3047–3048. DOI: [10.1093/bioinformatics/btw354](https://doi.org/10.1093/bioinformatics/btw354)
- Flegler A., Lipski A. 2022. The C50 carotenoid bacterioruberin regulates membrane fluidity in pink-pigmented *Arthrobacter* species. *Archives of Microbiology* 204: 70. DOI: [10.1007/s00203-021-02719-3](https://doi.org/10.1007/s00203-021-02719-3)
- Flores-Urbe J., Hevroni G., Ghai R. et al. 2019. Heliorhodopsins are absent in diderm (Gram-negative) bacteria: Some thoughts and possible implications for activity. *Environmental Microbiology Reports* 11: 419–424. DOI: [10.1111/1758-2229.12730](https://doi.org/10.1111/1758-2229.12730)
- Fujii N., Kuroda K., Narihiro T. et al. 2022. Metabolic potential of the superphylum Patescibacteria reconstructed from activated sludge samples from a municipal wastewater treatment plant. *Microbes and Environments* 37: n/a. DOI: [10.1264/jsme2.ME22012](https://doi.org/10.1264/jsme2.ME22012)
- He C., Keren R., Whittaker M.L. et al. 2021. Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nature Microbiology* 6: 354–365. DOI: [10.1038/s41564-020-00840-5](https://doi.org/10.1038/s41564-020-00840-5)
- He X., McLean J.S., Edlund A. et al. 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proceedings of the National Academy of Sciences* 112: 244–249. DOI: [10.1073/pnas.1419038112](https://doi.org/10.1073/pnas.1419038112)
- Hendrickson E.L., Bor B., Kerns K.A. et al. 2022. Transcriptome of epibiont *Saccharibacteria* *Nanosynbacter lyticus* strain TM7x during the establishment of symbiosis. *Journal of Bacteriology* 204: e00112-22. DOI: [10.1128/jb.00112-22](https://doi.org/10.1128/jb.00112-22)
- Hug L.A., Baker B.J., Anantharaman K. et al. 2016. A new view of the tree of life. *Nature Microbiology* 1: 16048. DOI: [10.1038/nmicrobiol.2016.48](https://doi.org/10.1038/nmicrobiol.2016.48)
- Hyatt D., Chen G.-L., LoCasio P.F. et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification *BMC Bioinformatics* 11: 119. DOI: [10.1186/1471-2105-11-119](https://doi.org/10.1186/1471-2105-11-119)
- Ibrahim A., Maatouk M., Rajaonison A. et al. 2021. Adapted protocol for *Saccharibacteria* cocultivation: two new members join the club of candidate phyla radiation. *Microbiology Spectrum* 9: e01069-21. DOI: [10.1128/spectrum.01069-21](https://doi.org/10.1128/spectrum.01069-21)
- Jaffe A.L., Konno M., Kawasaki Y. et al. 2022. *Saccharibacteria* harness light energy using type-1 rhodopsins that may rely on retinal sourced from microbial hosts. *The ISME Journal* 16: 2056-2059. DOI: [10.1038/s41396-022-01231-w](https://doi.org/10.1038/s41396-022-01231-w)
- Kang D.D. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities.
- Korotkov K.V., Sandkvist M., Hol W.G.J. 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology* 10: 336–351. DOI: [10.1038/nrmicro2762](https://doi.org/10.1038/nrmicro2762)
- Krueger F., James F., Ewels P. et al. 2023. TrimGalore. DOI: [10.5281/zenodo.7598955](https://doi.org/10.5281/zenodo.7598955)
- Kuhn A., Stuart R., Henry R. et al. 2003. The Alb3/Oxa1/YidC protein family: membrane-localized chaperones facili-

tating membrane protein insertion? *Trends in Cell Biology* 13: 510–516. DOI: [10.1016/j.tcb.2003.08.005](https://doi.org/10.1016/j.tcb.2003.08.005)

Langmead B., Wilks C., Antonescu V. et al. 2019. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 35: 421–432. DOI: [10.1093/bioinformatics/bty648](https://doi.org/10.1093/bioinformatics/bty648)

Lemos L.N., Medeiros J.D., Dini-Andreote F. et al. 2019. Genomic signatures and co-occurrence patterns of the ultra-small *Saccharimonadia* (phylum CPR/Patescibacteria) suggest a symbiotic lifestyle. *Molecular Ecology* 28: 4259–4271. DOI: [10.1111/mec.15208](https://doi.org/10.1111/mec.15208)

Li H., Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26: 589–595. DOI: [10.1093/bioinformatics/btp698](https://doi.org/10.1093/bioinformatics/btp698)

Li D., Liu C.-M., Luo R. et al. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31: 1674–1676. DOI: [10.1093/bioinformatics/btv033](https://doi.org/10.1093/bioinformatics/btv033)

Lomovatskaya L.A., Romanenko A.S. 2020. Secret Systems of Bacterial Phytopathogens and Mutualists (Review). *Applied Biochemistry and Microbiology* 56: 107–122. DOI: [10.31857/S0555109920020105](https://doi.org/10.31857/S0555109920020105) (in Russian).

Lu J., Rincon N., Wood D.E. et al. 2022. Metagenome analysis using the Kraken software suite. *Nature Protocols* 17: 2815–2839. DOI: [10.1038/s41596-022-00738-y](https://doi.org/10.1038/s41596-022-00738-y)

Nakai R. 2020. Size matters: ultra-small and filterable microorganisms in the environment. *microbes and environments* 35: n/a. DOI: [10.1264/jsme2.ME20025](https://doi.org/10.1264/jsme2.ME20025)

Néron B., Denise R., Coluzzi C. et al. 2023. MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes. *Peer Community Journal* 3: e28. DOI: [10.24072/pcjournal.250](https://doi.org/10.24072/pcjournal.250)

Parks D.H., Chuvochina M., Rinke C. et al. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research* 50: D785–D794. DOI: [10.1093/nar/gkab776](https://doi.org/10.1093/nar/gkab776)

Pinevich A.V. 2007. Microbiology. Biology of prokaryotes (in three volumes, 2nd edition). In: Harazova A.D., Gamalei Yu.V (Eds.), St. Petersburg: St. Petersburg University Press (in Russian)

Pushkarev A., Inoue K., Larom S. et al. 2018. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* 558: 595–599. DOI: [10.1038/s41586-018-0225-9](https://doi.org/10.1038/s41586-018-0225-9)

Rheims H., Rainey F.A., Stackebrandt E. 1996. A molecular approach to search for diversity among bacteria in the environment. *Journal of Industrial Microbiology and Biotechnology* 17: 159–169. DOI: [10.1007/BF01574689](https://doi.org/10.1007/BF01574689)

Rinke C., Schwientek P., Sczyrba A. et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499: 431–437. DOI: [10.1038/nature12352](https://doi.org/10.1038/nature12352)

Röth E., Marczin N., Balatonyi B. et al. 2011. Effect of a glutathione S-transferase inhibitor on oxidative stress and ischemia-reperfusion-induced apoptotic signalling of cultured

cardiomyocytes. *Experimental and clinical cardiology* 16: 92–96.

Schneewind O., Missiakas D.M. 2012. Protein secretion and surface display in Gram-positive bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 1123–1139. DOI: [10.1098/rstb.2011.0210](https://doi.org/10.1098/rstb.2011.0210)

Schwengers O., Jelonek L., Dieckmann M.A. et al. 2021. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics* 7: 000685. DOI: [10.1099/mgen.0.000685](https://doi.org/10.1099/mgen.0.000685)

Scott M.D., Meshnick S.R., Eaton J.W. 1987. Superoxide dismutase-rich bacteria. Paradoxical increase in oxidant toxicity. *Journal of Biological Chemistry* 262: 3640–3645. DOI: [10.1016/S0021-9258\(18\)61401-2](https://doi.org/10.1016/S0021-9258(18)61401-2)

Sieber C.M.K. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* 3: 836–843. DOI: [10.1038/s41564-018-0171-1](https://doi.org/10.1038/s41564-018-0171-1)

Soro V., Dutton L.C., Sprague S.V. et al. 2014. Axenic culture of a candidate division TM7 bacterium from the human oral cavity and biofilm interactions with other oral bacteria. *Applied and Environmental Microbiology* 80: 6480–6489. DOI: [10.1128/AEM.01827-14](https://doi.org/10.1128/AEM.01827-14)

Tamames J., Puente-Sánchez F. 2019. SqueezeMeta, A highly portable, fully automatic metagenomic analysis pipeline. *Frontiers in Microbiology* 9: 3349. DOI: [10.3389/fmicb.2018.03349](https://doi.org/10.3389/fmicb.2018.03349)

Tammam S., Sampaleanu L.M., Koo J. et al. 2013. PilMNOPQ from the *Pseudomonas aeruginosa* type IV pilus system form a transenvelope protein interaction network that interacts with PilA. *Journal of Bacteriology* 195: 2126–2135. DOI: [10.1128/JB.00032-13](https://doi.org/10.1128/JB.00032-13)

Uritskiy G.V., DiRuggiero J., Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6: 158. DOI: [10.1186/s40168-018-0541-1](https://doi.org/10.1186/s40168-018-0541-1)

Walker J.E. 2013. F-ATPases. In: Lennarz W.J., Lane M.D. (Eds.), *Encyclopedia of Biological Chemistry*. Waltham: Academic Press, pp. 269–274. DOI: [10.1016/B978-0-12-378630-2.00294-2](https://doi.org/10.1016/B978-0-12-378630-2.00294-2)

Walker J.E., Dickson V.K. 2006. The peripheral stalk of the mitochondrial ATP synthase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1757: 286–296. DOI: [10.1016/j.bbabi.2006.01.001](https://doi.org/10.1016/j.bbabi.2006.01.001)

Wood D.E., Lu J., Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology* 20: 257. DOI: [10.1186/s13059-019-1891-0](https://doi.org/10.1186/s13059-019-1891-0)

Wu Y.-W., Simmons B.A., Singer S.W. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32: 605–607. DOI: [10.1093/bioinformatics/btv638](https://doi.org/10.1093/bioinformatics/btv638)

Yakimov M.M., Merkel A.Y., Gaisin V.A. et al. 2022. Cultivation of a vampire: “Candidatus *Absconditococcus praedator*”. *Environmental Microbiology* 24: 30–49. DOI: [10.1111/1462-2920.15823](https://doi.org/10.1111/1462-2920.15823)



# Характеристика бактерий класса *Saccharimonadia* в озерах Байкал и Хубсугул: таксономическое и функциональное разнообразие

Оригинальная статья  
LIMNOLOGY  
FRESHWATER  
BIOLOGY

Гутник Д.И.<sup>1\*</sup>, Потапов С.А.<sup>1</sup>, Краснопеев А.Ю.<sup>1</sup>, Тихонова И.В.<sup>1</sup>, Букин Ю.С.<sup>1</sup>, Гладких А.С.<sup>2</sup>, Ключникова Е.О.<sup>2</sup>, Сбарцалья В.А.<sup>2</sup>, Асламов И.А.<sup>1</sup>, Белых О.И.<sup>1</sup>

<sup>1</sup>Лимнологический институт Сибирского отделения Российской Академии наук, ул. Улан-Баторская, 3, г. Иркутск, 664033, Россия

<sup>2</sup>Федеральное бюджетное учреждение науки «Санкт-Петербургский научно-исследовательский институт эпидемиологии и микробиологии им. Пастера», ул. Мира, д. 14, г. Санкт-Петербург, 197101, Россия

**АННОТАЦИЯ.** Ультрамелкие бактерии повсеместно распространены в окружающей среде, однако некоторое время они оставались в стороне от научного фокуса. Среди причин можно выделить их малый размер. Ультрамикробактерии, которые определяются как бактерии с объемом клетки менее 0,1 мкм<sup>3</sup>, могут проходить через фильтр с размером пор 0,2 мкм. Кроме того, есть ряд сложностей, сопряжённых с их культивированием, на данный момент удалось получить чистые культуры, но, в основном, доминирующих филумов, редкие представлены лишь единичными штаммами. При этом ультрамикробактерии обнаружены в широком спектре сред, а их высокоупорядоченные, небольшие геномы и уникальные экофизиологические особенности привлекают все большее внимание исследователей. Развитие метагеномики позволяет частично преодолеть барьер, связанный с ограничениями в культивировании, и рассмотреть в данной статье метагеномные сборки, полученные из образцов воды крупнейших олиготрофных озер Земли – Байкала и Хубсугула. В данном исследовании проведен таксономический и функциональный анализ реконструированных геномов бактерий класса *Saccharimonadia* (филум *Patescibacteria*), который призван расширить наши знания о редких представителях микробных сообществ и дополнить уже имеющиеся сведения информацией об обитателях древнейших экосистем.

**Ключевые слова:** *Saccharimonadia*, ультрамикробактерии, Байкал, Хубсугул, метагеномика, метаболизм

**Для цитирования:** Гутник Д.И., Потапов С.А., Краснопеев А.Ю., Тихонова И.В., Букин Ю.С., Гладких А.С., Ключникова Е.О., Сбарцалья В.А., Асламов И.А., Белых О.И. Характеристика бактерий класса *Saccharimonadia* в озерах Байкал и Хубсугул: таксономическое и функциональное разнообразие // *Limnology and Freshwater Biology*. 2024. - № 4. - С. 908-924. DOI: 10.31951/2658-3518-2024-A-4-908

## 1. Введение

Суперфилум *Patescibacteria*, также известный как Candidate Phyla Radiation (CPR) (Rinke et al., 2013), согласно базе данных The Genome Taxonomy Database (GTDB) (Parks et al., 2022), включает 30 классов, которые составляют больше 15% всех бактериальных филумов (Brown et al., 2015; Fujii et al., 2022). *Patescibacteria* обитают в самых разнообразных средах: осадки водоносных горизонтов и водоемы, водоочистные сооружения, активный ил, золотые прииски, полярные почвы и почвы тропических лесов, слюна человека, в ассоциации с губками и тараканами, и, более того, во многих из них они доминируют (Brown et al., 2015; Hug et al., 2016; Fujii et al., 2022).

На данный момент не так много известно о функциональных особенностях этих бактерий, а также не существует единого мнения о том, являются ли все представители CPR симбионтами, которые полагаются в своей жизнедеятельности на другие бактерии, или же некоторые из них способны к свободному существованию. Знания об этой группе бактерий ограничены в основном тем, что в настоящее время описаны только единичные случаи культивирования *Patescibacteria*; данный факт частично обусловлен их симбиотической/паразитической природой (Soro et al., 2014; He et al., 2021; Ibrahim et al., 2021; Yakimov et al., 2022). Одни из немногих культивированных видов (например, TM7х) (He et al., 2015) принадлежит классу *Saccharimonadia*. Представители данного класса впервые были обна-

\*Автор для переписки.

Адрес e-mail: [gutnik@lin.irk.ru](mailto:gutnik@lin.irk.ru) (Д.И. Гутник)

**Поступила:** 08 августа 2024; **Принята:** 21 августа 2024;

**Опубликована online:** 30 августа 2024

© Автор(ы) 2024. Эта работа распространяется под международной лицензией Creative Commons Attribution-NonCommercial 4.0.



ружены при анализе частичных последовательностей 16S рДНК микробных сообществ торфяного болота в Германии (Rheims et al., 1996), и являются одними из самых известных. Интерес к изучению *Saccharimonadia* во многом вызван их предполагаемой ролью в развитии заболеваний полости рта. Те *Saccharimonadia*, которые удалось культивировать в лабораторных условиях, однозначно демонстрировали симбиотическую связь с актинобактериями, схожая картина наблюдалась и для других немногочисленных культивированных CPR. И действительно, известные *Patescibacteria* обычно обладают небольшим размером генома (до 1,0 Мб) (Lemos et al., 2019; Nakai, 2020) и неполными путями биосинтеза (Castelle et al., 2018), сочетание этих характеристик указывает на ассоциированный с хозяином, симбиотический или паразитический образ жизни.

Однако уникальные условия обитания в крупных глубоководных олиготрофных озерах, каковыми являются Байкал и Хубсугул, могут являться причиной уменьшения размеров геномов и отражаться в функциональных особенностях. Эти экосистемы также представляют интерес с точки зрения эволюции бактерий, принимая во внимания древность и общность происхождения двух тектонических озер Байкальской рифтовой зоны. Кроме того, образцы из оз. Байкал отобраны на разных глубинах: в верхних слоях воды, т.е. в фотическом слое (до 25 м) и на глубинах 1250-1350 м, что также потенциально влияет на метаболизм бактерий. Таким образом, целью работы является изучение и сравнение функциональных особенностей и образа жизни *Saccharimonadia*, полученных из проб воды Байкала и Хубсугула, с помощью метагеномных сборок и реконструкции геномов.

## 2. Материалы и методы

Пробы воды отобраны в оз. Байкал на центральной станции в проливе Малое Море, на центральных станциях разрезов пос. Листвянка – пос. Танхой, мыс Ухан – мыс Тонкий, мыс Елохин – пос. Давша, в 7 км и в 3 км от пос. Листвянка, в 3 км от пос. Турка, в 3 км от мыса Елохин в марте-сентябре 2018-2022 гг. В периоды открытой воды отбор проб проводили на глубинах 0, 5, 10, 15, 20, 25, 50 м, используя систему батометров SBE-32 (Carousel Water Sampler, Sea Bird Electronics Inc., США). В под-

лёдный период пробы отбирали со льда с помощью батометров Нискина. В оз. Хубсугул (Монголия) отбор проб воды произведён в 3 км от пос. Ханха в октябре 2023 г. Пробы отобраны на глубинах 0, 10, 25, 40, 50 и 70 м батометром с лодки.

Пробы воды последовательно фильтровали через поликарбонатные фильтры с диаметром пор 0,4 и 0,2 мкм (Millipore, США) для удаления крупных форм планктона. Фильтраты концентрировали с помощью системы ультрафильтрации в тангенциальном потоке VivaFlow 200 (Sartorius, ФРГ) до конечного объема ~20 мл. Далее образцы концентрировали, применяя Vivaspin Turbo 15 (50 кДа, Sartorius, ФРГ) до объема ~100 мкл.

ДНК экстрагировали стандартным фенол-хлороформным методом. ДНК библиотеки, включающие A\_63, A\_68, A\_9 секвенировали с использованием платформы Illumina MiSeq, метагеномные данные размещены в биопроекте PRJNA1006167. Секвенирование ДНК для I\_78 происходило на платформе DNBSEQ-G50 (MGISEQ-200).

Кроме того, были использованы 8 наборов необработанных метагеномных прочтений, полученных с помощью Shotgun-секвенирования и размещенных для публичного доступа в биопроектах (PRJNA396997, PRJNA521725, PRJNA615165). Материалом для секвенирования послужили пробы воды оз. Байкал, взятые летом 12-13 июля 2018 г. в 8 км от пос. Листвянка на 4 глубинах (5, 20, 1250 и 1350 м), а также зимние пробы воды, отобранные с глубин 5 и 20 м – 14 марта 2016 г., и с глубин 1250 и 1350 м – 29 марта 2018 г. в 7 км от пос. Листвянка (Cabello-Yeves et al., 2018; Cabello-Yeves et al., 2020; Coutinho et al., 2020). Информация о соответствии наименований реконструированных геномов и материалов, послуживших основой для их сборок, представлена в Таблице 1.

Процесс реконструкции геномов A\_63, A\_68, A\_9 включал оценку качества метагеномного секвенирования с помощью программы MultiQC v. 1.12 (Ewels et al., 2016), удаление адаптеров с помощью программы Trim Galore v. 0.6.5 (Krueger et al., 2023), сборку ридов в контиги с помощью SPAdes v. 3.15.5 (Bankevich et al., 2012) и выравнивание ридов на контиги посредством BWA v. 0.7.17 (Li and Durbin, 2010) и пакета SAMtools v. 1.18 (Danecsek et al., 2021). Метагеномный биннинг и выделение геномов из сборки проведены с использованием программного пакета MetaWRAP v. 1.3 (Uritskiy et al., 2018).

Таблица 1. Условия отбора проб и источники данных

	P_794	A_63	A_68	A_9	I_78
Водоем	Байкал	Байкал			Хубсугул
Глубины, м	5, 10, 1250, 1350	0, 5, 10, 15, 20, 25, 50			0, 10, 25, 40, 50, 70
Период	Июль-Март, 2016-2018	Март-Сентябрь, 2018-2022			Октябрь, 2023
Источник	Публичные данные: PRJNA396997, PRJNA521725, PRJNA615165 (Cabello-Yeves et al., 2018; Cabello-Yeves et al., 2020; Coutinho et al., 2020)	Собственные данные: PRJNA1006167			Собственные данные

Подготовка прочтений для сборок, относящихся к P\_794 и I\_78, состояла из оценки качества секвенирования программой FastQC v. 0.12.1 (Andrews et al., 2010), удаления адаптеров и коротких ридов (меньше 141 оснований для P\_794 и 90 для I\_78) с применением Trimmomatic v. 0.39 (Bolger et al., 2014) и очистки от эукариотической ДНК программным комплексом Kraken 2 v. 2.1.3 (Wood et al., 2019; Lu et al., 2022). Для сборки контигов, содержащих P\_794, создана внешняя сборка путем слияния прочтений соответствующего направления из 8 образцов, отличающихся по глубине и сезону, в итоге получен один объединенный набор парных прочтений, которые затем собраны ассемблером MEGANIT (Li et al., 2015). Для внешней сборки произведено выравнивание ридов на контиги при помощи Bowtie 2 v. 2.5.4 (Langmead et al., 2019) и пакета Samtools v. 1.20 (Danecsek et al., 2021). Основываясь на полученной статистике по покрытию, осуществлены фильтрация и удаление коротких и контигов с низким покрытием с помощью скрипта R. Финальные сборки контигов для выделения геномов P\_794 и I\_78 выполнены с использованием конвейера SqueezeMeta (Tamames and Puente-Sánchez, 2019), для P\_794 применялся режим coassembly с включением всех 8 наборов прочтений и внешней сборки, а для I\_78 – sequential. В качестве ассемблера для P\_794 использован MEGANIT (Li et al., 2015), а для I\_78 – SPAdes v. 3.15.5 (Bankevich et al., 2012). Биннинг проведен с использованием CONCOAT v. 1.1.0 (Alneberg et al., 2014), MaxBin v. 2.2.7 (Wu et al., 2016) и MetaBAT2 v. 2.12.1 (Kang,

2015). Интеграция результатов биннинга осуществлена в DAS Tool v. 1.1.7 (Sieber, 2018).

Полнота и завершенность окончательных геномов оценена при помощи CheckM2 v. 1.0.2 (Chklovski et al., 2023). Открытые рамки считывания (open reading frame, ORF) определяли, используя Prodigal v.2.6.3 (Hyatt et al., 2010). Аннотация генов осуществлена с применением Bakta v.1.9.2 (Schwengers et al., 2021), с настройками по умолчанию. Аннотацию KEGG и присвоение номеров КО белкам проводили сервисом KofamKOALA (Aramaki et al., 2020). Геномы таксономически классифицированы с использованием набора инструментов базы данных таксономии генома (GTDB-Tk) v. 2.4.0 (Chaumeil et al., 2020). MacSyFinder v. 2.1.4 (Néron et al., 2023) и модели TXSScan v. 1.1.3 (Abby et al., 2016) применяли для поиска систем секреции белков.

### 3. Результаты и обсуждение

#### 3.1. Описание геномов и таксономическая классификация

На основе трех наборов метагеномных данных собрано пять геномов *Saccharimonadia*. Все они демонстрируют достаточно высокий уровень полноты сборки 84,9-97,6% с контаминацией ниже 2,6%, за исключением генома I\_78, где она составила 7,0% (Таблица 2). Плотность кодирования (медиана 90,5%; диапазон 89,6%-91,8%) и содержание GC (медиана 42,2%; диапазон 40,7-45,0%) у них соответствуют свободноживущим организмам или

Таблица 2. Описание и таксономическая принадлежность реконструированных геномов

	P_794	A_63	A_68	A_9	I_78
Описание					
Размер, н.п.	984 478	655 626	594 118	612 023	718 727
Полнота, %	91.6	97.6	91.2	84.9	93.9
Контаминация, %	2.08	-	0.29	2.56	7.00
Г + Ц, %	45.0	41.3	42.3	40.7	42.2
Контиг N50, н.п.	10 865	57 148	35 054	12 346	3 970
Плотность кодирования, %	89.6	90.7	89.6	91.8	90.5
Кол-во CDS	1 016	686	594	645	675
Таксономия					
Царство	Bacteria	Bacteria	Bacteria	Bacteria	Bacteria
Филум	Patescibacteria	Patescibacteria	Patescibacteria	Patescibacteria	Patescibacteria
Класс	Saccharimonadia	Saccharimonadia	Saccharimonadia	Saccharimonadia	Saccharimonadia
Порядок	Saccharimonadales	CAILAD01	CAILAD01	Saccharimonadales	CAILAD01
Семейство	UBA1547_A	BJGX01	CAILAD01	AWTP1-31	CAILAD01
Род	2011-GWC2-44-17	CALRAY01	CAILAD01	CAJAXB01	CAILAD01
Вид	-	CALRAY01	-	-	-
Ближайший геном	GCA_016199685.1	GCA_943353705.1	GCA_013816445.1	GCA_943353855.1	GCA_903832035.1
ANI для ближайшего генома, %	78.94	97.66	78.36	92.13	77.02

факультативным внутриклеточным или внеклеточным симбионтам/паразитам (Chiriac et al., 2022).

Из исследуемых геномов P\_794 представляет собой самый крупный геном с максимальным числом CDS (1074) и достаточно обширным для представителя класса набором функций. При этом, согласно поиску по таксономической базе данных GTDB, определить его принадлежность возможно только до рода, причем сходство с ближайшим референсным организмом по ANI составляет всего 78,9%. 93% TPM (transcripts per kilobase million, число прочтений, нормализованных на длину каждого прочтения, умноженное на миллион), принадлежащих этому бину, соответствует июльским образцам воды Байкала с глубин 1250 и 1350 м.

Геномы A\_63, A\_68 и A\_9 также получены из байкальских проб с поверхностных горизонтов. Размер их геномов, как и число CDS, примерно одного порядка. Таксономическое определение с помощью GTDB позволило обнаружить достаточно близкое сходство по ANI для A\_63 и A\_9, определив род всех геномов, а в случае с A\_63 и вид.

I\_78 – геном, собранный из данных секвенирования поверхностных проб воды оз. Хубсугул. Он второй по величине из представленных в работе и наиболее контаминирован. При этом количество CDS у него составляет 675. Интересно, что согласно таксономическому анализу, наиболее близкими принадлежащими одному роду по предварительной классификации, оказались *Saccharimonadia* из поверхностных образцов Байкала A\_68 и Хубсугула I\_78.

Совпадения до вида с помощью классификации GTDB выявлено только для A\_63. Два генома P\_794 и A\_9 принадлежат порядку *Saccharimonadales* и некультивируемым семействам и родам. A\_63, A\_68 и I\_78 на всех уровнях отнесены к некультивируемым представителям.

### 3.2. Аннотация и функциональный анализ геномов

Сравнение аннотированных геномов по функциональным ортологам базы данных KEGG показало, что число общих генов составляет 20,8% (Рис. 1). При этом наиболее уникальным для данной выборки набором генов обладает P\_794, у которого 14,0% предсказанных генов не совпадают с генами других рассматриваемых геномов.

Функциональный анализ установил наиболее вероятную общую для всех геномов схему получения свободной энергии в форме АТФ – катаболизм углеводов с образованием пирувата. Причем в данном случае образование пирувата идет по гексозомонофосфатному пути де Фриза-Стаутамера или бифидо-шунту (Pinevich, 2007). Так, у всех геномов присутствует практически весь набор ферментов необходимых для его реализации: фруктозобисфосфат-альдолаза, трансальдолаза и транскетолаза. Примечательно, что и в отсутствие ферментов прослеживается единая стратегия. Гены, отвечающие за выработку фосфоглицеромутаза, участвующей в

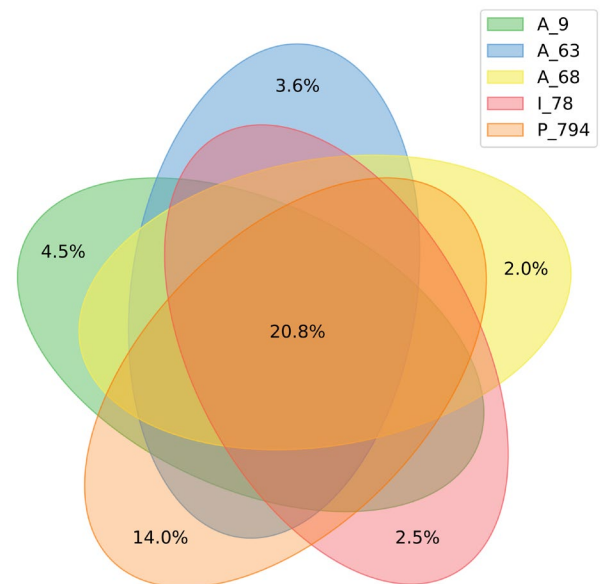


Рис.1. Диаграмма Венна, демонстрирующая пересечения в составе генов *Saccharimonadia*, аннотированных с помощью базы данных KEGG Orthology

реакции превращения 3-фосфоглицерата в 2-фосфоглицерат и необходимой для дальнейшего образования пирувата, отсутствуют у всех исследуемых в работе *Saccharimonadia*.

Уже на этапе анализа функциональных категорий (Рис. 2) можно отметить отличительную черту P\_794, а именно, присутствие кластера генов, отвечающих за расщепление олиго- и полисахаридов сахарозы и крахмала до D-глюкозы, в частности, за счет продукции  $\alpha$ -глюкозидазы, альфа-амилазы, мальтазы-глюкоамилазы и т.д. Вторым геномом по количеству предсказанных генов, задействованных в углеводном обмене, является A\_63. Однако, для него гексозомонофосфатный путь начинается с расщепления глюкозы-6-фосфат или фруктозы-6-фосфат. У остальных рассматриваемых геномов функциональный набор углеводного обмена еще меньше, A\_9, A\_68, I\_78 способны только к преобразованию фруктозы-6-фосфат.

Важно отметить, что у всех изучаемых *Saccharimonadia* отсутствует брожение, процесс обрывается на образовании пирувата и одной молекулы АТФ на молекулу окисленной гексозы. Обычно пируват, являющийся важнейшим промежуточным продуктом брожения, подвергается окислительному карбоксилированию по одному из 4 путей. В случае с изучаемыми геномами, необходимые для брожения ферменты отсутствуют у всех, кроме A\_63. Для этого организма предсказано наличие лактатдегидрогеназы, которая позволяет осуществлять молочнокислое брожение на основе пирувата, что, тем не менее, не находит подтверждение в аннотации с помощью KEGG. Кроме того, для осуществления других типов брожения необходимо участие витаминов, продукция которых отсутствует у всех пяти микроорганизмов. Вместе с тем, такая практика широко распространена в сообществах микроорга-

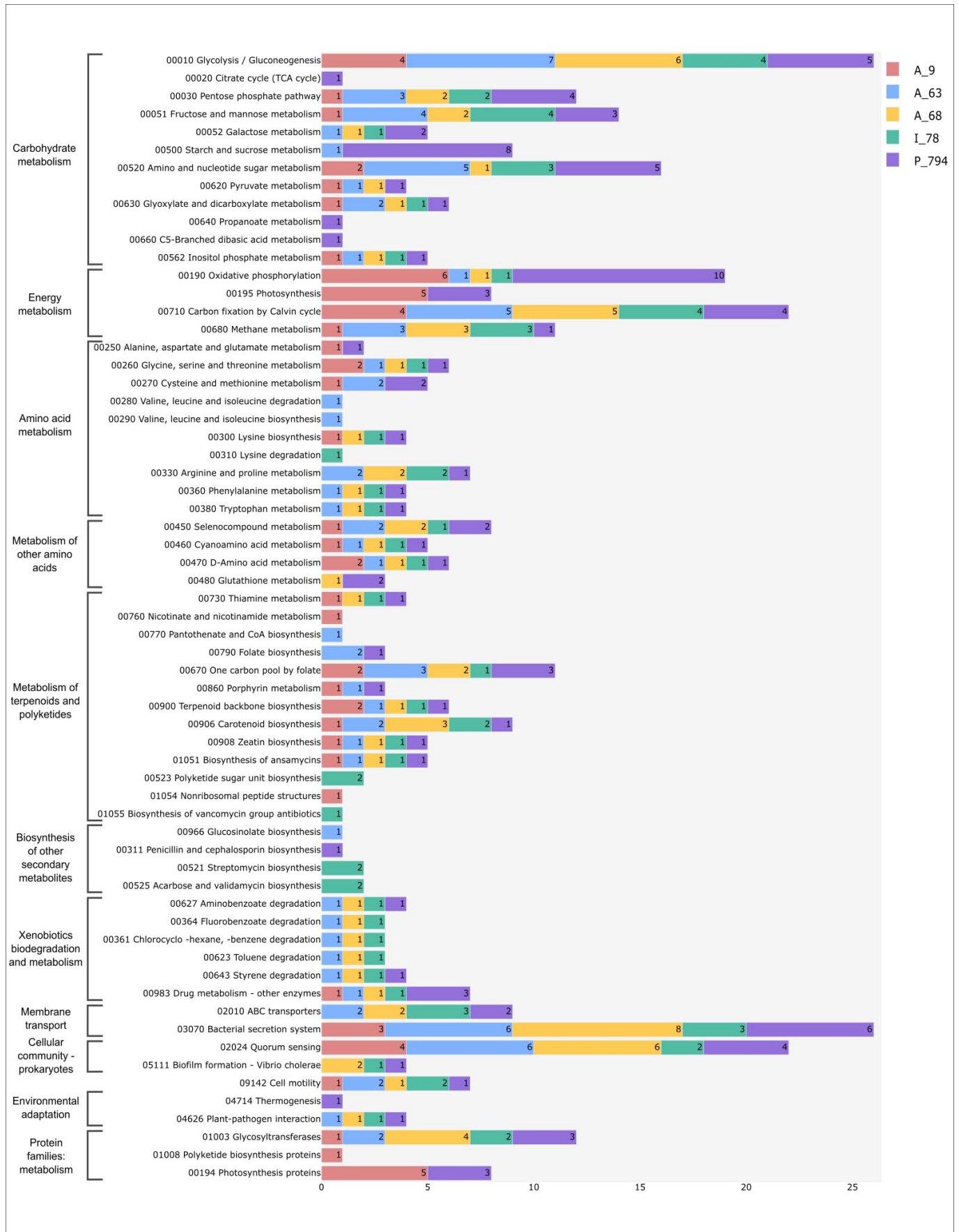


Рис.2. Количество генов у *Saccharimonadia* в категориях, определенных согласно классификации KEGG PATHWAY

низмов, когда существуют основные продуценты, обеспечивающие потребность других бактерий в незаменимых для реакций метаболизма витаминах.

Еще одной отличительной чертой P\_794 является увеличенное по сравнению с другими число генов, задействованных в окислительном фосфорилировании. Так, у него найден полный оксидазный комплекс, состоящий из пяти субъединиц (СуоА, СуоВ, СуоС, СуоD, СуоЕ), представляющий собой аэробную дыхательную цепь, реагирующую с кислородом, восстанавливая его до воды с одновременным переносом 4 протонов через мембрану (Cheruri et al., 1990). Способность к дыханию подтверждается и наличием фермента FAD-зависимой оксидоредуктазы. Геном A\_9 также содержит гены, входящие в состав комплексов фотосинтеза и окислительного фосфорилирования, но все они связаны с АТФазой F-типа – многосубъединичным ферментным комплексом, роль которого изменяется в аэробных и анаэробных условиях (Walker, 2013). В анаэробных условиях ферменты осуществляют обратную реакцию и генерируют протонную движущую силу, используя энергию, высвобождаемую при гидролизе АТФ (Walker and Dickson, 2006).

В целом, изучаемые *Saccharimonadia* не обладают синтезом всех необходимых нуклеотидов, аминокислот, липидов, витаминов, кофакторов или полным циклом трикарбоновых кислот (ТКК), что соответствует результатам, полученным в ранее опубликованных работах. Биосинтез вторичных метаболитов также практически не представлен, хотя и существуют индивидуальные особенности в наличии тех или иных генов, которые тем не менее не обеспечивают полного цикла производства веществ.

Согласно исследованиям, у некоторых *Saccharimonadia* обнаружена способность к продукции родопсина (Jaffe et al., 2022), однако в текущих образцах гены, отвечающие за его синтез, отсутствуют. При этом у четырех геномов (A\_9, A\_63, A\_68, I\_78) предсказаны области, кодирующие выработку гелиородопсина – нового семейства родопсинов. Гелиородопсин открыт совсем недавно, в 2018 году, и еще слабо изучен (Pushkarev et al., 2018). Вместе с тем на данный момент существует предположение, что наличие гелиородопсина невозможно у организмов с двойной мембраной и для его функционирования необходимо прямое взаимодействие с окружающей средой (Flores-Uribe et al., 2019).

*Saccharimonadia* демонстрируют отличия и в механизмах адаптации к окружающей среде, A\_68 обладает пусть и не полным, но максимальным среди рассматриваемых *Saccharimonadia* набором генов для синтеза бактериоруберина, отвечающего за адаптацию мембран к низким температурам (Flegler and Lipski, 2022). A\_9, A\_68 и P\_794 содержат гены супероксиддисмутазы. У A\_63, A\_68 и I\_78 обнаружена глутатион-S-трансфераза. Оба фермента необходимы для борьбы с кислородным стрессом, что актуально для озер с высоким содержанием кислорода, таких как Байкал и Хубсугул (Scott et al., 1987; Röth et al., 2011).

Будучи ауксотрофами по многим категориям, *Saccharimonadia* нуждаются в альтернативных путях получения необходимых для жизнедеятельности соединений, что позволяет предположить их вовлеченность в симбиотическое или паразитические отношения с другими организмами. При возникновении такого рода связей происходит тщательная их регуляция посредством секреции белков, так, системы секреции должны включать механизмы для транслокации секретируемых белков (эффекторов и токсинов) через плазматическую мембрану хозяина.

Система секреции второго типа (T2SS) представлена в геномах неполно. T2SS является Sec-зависимой и включает в себя: псевдопили, внешний и внутренний мембранные комплексы (Korotkov et al., 2012; Lomovatskaya and Romanenko, 2020). Лишь у A\_9 обнаружен один из трех белков внутренней мембранной платформы T2SS – GspF. Остальные *Saccharimonadia* кодируют только GspG, относящейся к псевдопилям. Кроме того, не был определен ни один из белков комплекса внешней мембраны, таких как GspD и GspS, что ставит под сомнение функциональную способность T2SS. При этом почти у всех присутствуют ген SecA (A\_9, A\_63, A\_68 и P\_794), ответственный за гидролиз АТФ для запуска транслокации белков, а также SecD-G и SecY гены, кодирующие полный мембранный комплекс (Schneewind and Missiakas, 2012; Chiriac et al., 2022). Кроме того, A\_63, A\_68 и P\_794 содержат белок YidC, опосредующий встраивание в мембрану белков Sec (Kuhn et al., 2003).

В качестве вероятного механизма для обеспечения симбиотического или паразитического образа жизни *Saccharimonadia* могут использовать систему секреции белков IV типа (T4aP), на основе которых формируются пили, прикрепляющиеся к внешней мембране хозяина (Ibrahim et al., 2021; Hendrickson et al., 2022). Пили типа IV – это поверхностные придатки, участвующие во многих процессах, включая адгезию к биотическим и абиотическим поверхностям, агрегацию, поглощение ДНК и фагов, специфичную подвижность и перенос электронов (Tammam et al., 2013; Berry and Pelicic, 2015). Поиск с помощью MacSyFinder и набора моделей TXSScan показал разнообразие пилей в рассматриваемых геномах. Все *Saccharimonadia*, за исключением I\_78, содержали кластеры мажорных и минорных пилей IV типа: pilABCDE, piMNO, pilT и pilVWX. Еще одним аргументом в пользу зависимого образа жизни *Saccharimonadia* является наличие ABC-транспортеров, в том числе и с определенными функциями, а именно, перенос АТФ, кобальта, железа, сахара, а также и с неизвестными функциональными свойствами (Fujii et al., 2022).

#### 4. Выводы

Развитие методик метагеномного анализа существенно расширило наши возможности исследования некультивируемых бактерий. Хотя полнота и контаминация метагеномных данных являются

ограничивающими факторами, которые следует учитывать, особенно при функциональном анализе мало описанных организмов. Тем не менее, даже первое приближение, представляющее собой изучение пяти метагеномных сборок, позволило получить ранее недоступную информацию об ультрамикробактериях, обитающих в уникальных экосистемах олиготрофных озер Байкала и Хубсугула.

Предварительный таксономический анализ показал, что только для одного из изучаемых геномов в базе данных содержится достаточно близкий референсный организм, для всех остальных сходство по ANI очень отдаленное. Причем, между собой рассматриваемые геномы тоже не демонстрируют таксономического родства. За исключением A\_68 и I\_78, предположительно относящихся к одному роду. Функциональное сравнение также показало различия в генах, всего 20,8% среди них являются общими, в то время как P\_794 содержит 14,0% уникальных отличных от собратьев по классу генов.

Предметное функциональное сравнение свидетельствует о достаточно обширных, особенно в условиях небольшого генома, отличиях в метаболизме геномов. Так, P\_794 обладает способностями к деградации сахарозы и крахмала, а также предположительно является факультативным анаэробом, способным к кислородному дыханию благодаря наличию полного оксидазного комплекса. При этом все рассматриваемые *Saccharimonadia*, вероятно, осуществляют энергетический метаболизм за счет процесса гликолиза через метаболическую петлю, включающую гексозомонофосфатный путь де Фриза-Стаутамера. Открытым остается судьба пирувата в этом пути, поскольку ферментов, необходимых для брожения, не было обнаружено.

Геномы также демонстрируют различные способы адаптации к окружающей среде и системы секреции белков. Именно системы секреции белков, в частности, обширная представленность генов, кодирующих T4aP, наряду с ограниченным метаболическим потенциалом и наличием ABC-транспортеров с неизвестными функциями, свидетельствуют о симбиотическом или паразитическом способе существования рассматриваемых *Saccharimonadia*. Дополняет это предположение наблюдаемая плотность кодирования, которая хоть и является достаточно высокой, но все же ниже наблюдаемых плотностей кодирования у свободноживущих микробов, где значения > 95% являются обычными (Chiriac et al., 2022). Тем не менее, хотя изученные нами геномы и обладают неполными путями биосинтеза соединений, необходимых для построения клеток и жизнедеятельности, нельзя и исключать способность *Saccharimonadia* синтезировать их *de novo* или получать из окружающей среды по неизвестным путям, и того, что такое рационализированное сокращение генома не возникло как адаптация к условиям с низким или ограниченным содержанием питательных веществ.

## Благодарности

Работа выполнена при финансовой поддержке Государственного задания № 0279-2021-0015. Авторы выражают благодарность экипажам НИС «Верещагин», «Титов» флота ЛИН СО РАН, а также Ханаеву И.В., Макарову М.М. за помощь в отборе проб. Сбор проб планктона озера Хубсугул выполнен по теме Государственного задания 0279-2021-0004. Биоинформатические расчеты выполнены на вычислительном кластере ЛИН СО РАН.

## Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

## Список литературы

- Abby S.S., Cury J., Guglielmini J. et al. 2016. Identification of protein secretion systems in bacterial genomes. *Scientific Reports* 6: 23080. DOI: [10.1038/srep23080](https://doi.org/10.1038/srep23080)
- Alneberg J., Bjarnason B.S., De Bruijn I. et al. 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods* 11: 1144–1146. DOI: [10.1038/nmeth.3103](https://doi.org/10.1038/nmeth.3103)
- Andrews S., Krueger F., Segonds-Pichon A. et al. 2010. FastQC: a quality control tool for high throughput sequence data
- Aramaki T., Blanc-Mathieu R., Endo H. et al. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36: 2251–2252. DOI: [10.1093/bioinformatics/btz859](https://doi.org/10.1093/bioinformatics/btz859)
- Bankevich A., Nurk S., Antipov D. et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* 19: 455-77. DOI: [10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021)
- Berry J.-L., Pelicic V. 2015. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives FEMS. *Microbiology Reviews* 39: 134–154. DOI: [10.1093/femsre/fuu001](https://doi.org/10.1093/femsre/fuu001)
- Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. DOI: [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170)
- Brown C.T., Hug L.A., Thomas B.C. et al. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523: 208–211. DOI: [10.1038/nature14486](https://doi.org/10.1038/nature14486)
- Cabello-Yeves P.J., Zemsкая T.I., Rosselli R. et al. 2018. Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal. *Applied and Environmental Microbiology* 84: e02132-17. DOI: [10.1128/AEM.02132-17](https://doi.org/10.1128/AEM.02132-17)
- Cabello-Yeves P.J., Zemsкая T.I., Zakharenko A.S. et al. 2020. Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. *Limnology and Oceanography* 65: 1471–1488. DOI: [10.1002/lno.11401](https://doi.org/10.1002/lno.11401)
- Castelle C.J., Brown C.T., Anantharaman K. et al. 2018. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nature Reviews Microbiology* 16: 629–645. DOI: [10.1038/s41579-018-0076-20](https://doi.org/10.1038/s41579-018-0076-20)
- Chaumeil P.-A., Mussig A.J., Hugenholtz P. et al. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36: 1925–1927. DOI: [10.1093/bioinformatics/btz848](https://doi.org/10.1093/bioinformatics/btz848)
- Chepuri V., Lemieux L., Au D.C. et al. 1990. The sequence of the cyo operon indicates substantial structural similarities

between the cytochrome o ubiquinol oxidase of *Escherichia coli* and the aa3-type family of cytochrome c oxidases. *Journal of Biological Chemistry* 265: 11185–11192. DOI: [10.1016/S0021-9258\(19\)38574-6](https://doi.org/10.1016/S0021-9258(19)38574-6)

Chiriac M.-C., Bulzu P.-A., Andrei A.-S. et al. 2022. Ecogenomics sheds light on diverse lifestyle strategies in freshwater CPR. *Microbiome* 10: 84. DOI: [10.1186/s40168-022-01274-3](https://doi.org/10.1186/s40168-022-01274-3)

Chklovski A., Parks D.H., Woodcroft B.J. et al. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods* 20: 1203–1212. DOI: [10.1038/s41592-023-01940-w](https://doi.org/10.1038/s41592-023-01940-w)

Coutinho F. H., Cabello-Yeves P.J., Gonzalez-Serrano R. et al. 2020. New viral biogeochemical roles revealed through metagenomic analysis of Lake Baikal. *Microbiome* 8: 163. DOI: [10.1186/s40168-020-00936-4](https://doi.org/10.1186/s40168-020-00936-4)

Danecek P., Bonfield J.K., Liddle J. et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10: giab008. DOI: [10.1093/gigascience/giab008](https://doi.org/10.1093/gigascience/giab008)

Ewels P., Magnusson M., Lundin S. et al. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32: 3047–3048. DOI: [10.1093/bioinformatics/btw354](https://doi.org/10.1093/bioinformatics/btw354)

Flegler A., Lipski A. 2022. The C50 carotenoid bacterioruberin regulates membrane fluidity in pink-pigmented *Arthrobacter* species. *Archives of Microbiology* 204: 70. DOI: [10.1007/s00203-021-02719-3](https://doi.org/10.1007/s00203-021-02719-3)

Flores-Uribe J., Hevroni G., Ghai R. et al. 2019. Heliorhodopsins are absent in diderm (Gram-negative) bacteria: Some thoughts and possible implications for activity. *Environmental Microbiology Reports* 11: 419–424. DOI: [10.1111/1758-2229.12730](https://doi.org/10.1111/1758-2229.12730)

Fujii N., Kuroda K., Narihiro T. et al. 2022. Metabolic potential of the superphylum Patescibacteria reconstructed from activated sludge samples from a municipal wastewater treatment plant. *Microbes and Environments* 37: n/a. DOI: [10.1264/jsme2.ME22012](https://doi.org/10.1264/jsme2.ME22012)

He C., Keren R., Whittaker M.L. et al. 2021. Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nature Microbiology* 6: 354–365. DOI: [10.1038/s41564-020-00840-5](https://doi.org/10.1038/s41564-020-00840-5)

He X., McLean J.S., Edlund A. et al. 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proceedings of the National Academy of Sciences* 112: 244–249. DOI: [10.1073/pnas.1419038112](https://doi.org/10.1073/pnas.1419038112)

Hendrickson E.L., Bor B., Kerns K.A. et al. 2022. Transcriptome of epibiont *Saccharibacteria Nanosynbacter lyticus* strain TM7x during the establishment of symbiosis. *Journal of Bacteriology* 204: e00112-22. DOI: [10.1128/jb.00112-22](https://doi.org/10.1128/jb.00112-22)

Hug L.A., Baker B.J., Anantharaman K. et al. 2016. A new view of the tree of life. *Nature Microbiology* 1: 16048. DOI: [10.1038/nmicrobiol.2016.48](https://doi.org/10.1038/nmicrobiol.2016.48)

Hyatt D., Chen G.-L., LoCasio P.F. et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification *BMC Bioinformatics* 11: 119. DOI: [10.1186/1471-2105-11-119](https://doi.org/10.1186/1471-2105-11-119)

Ibrahim A., Maatouk M., Rajaonison A. et al. 2021. Adapted protocol for *Saccharibacteria* cocultivation: two new members join the club of candidate phyla radiation. *Microbiology Spectrum* 9: e01069-21. DOI: [10.1128/spectrum.01069-21](https://doi.org/10.1128/spectrum.01069-21)

Jaffe A.L., Konno M., Kawasaki Y. et al. 2022. *Saccharibacteria* harness light energy using type-1 rhodopsins that may rely on retinal sourced from microbial hosts. *The ISME Journal* 16: 2056-2059. DOI: [10.1038/s41396-022-01231-w](https://doi.org/10.1038/s41396-022-01231-w)

Kang D.D. 2015. MetaBAT, an efficient tool for accurately

reconstructing single genomes from complex microbial communities.

Korotkov K.V., Sandkvist M., Hol W.G.J. 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology* 10: 336–351. DOI: [10.1038/nrmicro2762](https://doi.org/10.1038/nrmicro2762)

Krueger F., James F., Ewels P. et al. 2023. TrimGalore. DOI: [10.5281/zenodo.7598955](https://doi.org/10.5281/zenodo.7598955)

Kuhn A., Stuart R., Henry R. et al. 2003. The Alb3/Oxa1/YidC protein family: membrane-localized chaperones facilitating membrane protein insertion? *Trends in Cell Biology* 13: 510–516. DOI: [10.1016/j.tcb.2003.08.005](https://doi.org/10.1016/j.tcb.2003.08.005)

Langmead B., Wilks C., Antonescu V. et al. 2019. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 35: 421–432. DOI: [10.1093/bioinformatics/bty648](https://doi.org/10.1093/bioinformatics/bty648)

Lemos L.N., Medeiros J.D., Dini-Andreote F. et al. 2019. Genomic signatures and co-occurrence patterns of the ultra-small *Saccharimonadia* (phylum CPR/Patescibacteria) suggest a symbiotic lifestyle. *Molecular Ecology* 28: 4259–4271. DOI: [10.1111/mec.15208](https://doi.org/10.1111/mec.15208)

Li H., Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26: 589–595. DOI: [10.1093/bioinformatics/btp698](https://doi.org/10.1093/bioinformatics/btp698)

Li D., Liu C.-M., Luo R. et al. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31: 1674–1676. DOI: [10.1093/bioinformatics/btv033](https://doi.org/10.1093/bioinformatics/btv033)

Lomovatskaya L.A., Romanenko A.S. 2020. Secret Systems of Bacterial Phytopathogens and Mutualists (Review). *Applied Biochemistry and Microbiology* 56: 107–122. DOI: [10.31857/S0555109920020105](https://doi.org/10.31857/S0555109920020105) (in Russian).

Lu J., Rincon N., Wood D.E. et al. 2022. Metagenome analysis using the Kraken software suite. *Nature Protocols* 17: 2815–2839. DOI: [10.1038/s41596-022-00738-y](https://doi.org/10.1038/s41596-022-00738-y)

Nakai R. 2020. Size matters: ultra-small and filterable microorganisms in the environment. *microbes and environments* 35: n/a. DOI: [10.1264/jsme2.ME20025](https://doi.org/10.1264/jsme2.ME20025)

Néron B., Denise R., Coluzzi C. et al. 2023. MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes. *Peer Community Journal* 3: e28. DOI: [10.24072/pcjournal.250](https://doi.org/10.24072/pcjournal.250)

Parks D.H., Chuvochina M., Rinke C. et al. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research* 50: D785–D794. DOI: [10.1093/nar/gkab776](https://doi.org/10.1093/nar/gkab776)

Pinevich A.V. 2007. *Microbiology. Biology of prokaryotes* (in three volumes, 2nd edition). In: Harazova A.D., Gamalei Yu.V. (Eds.), St. Petersburg: St. Petersburg University Press (in Russian)

Pushkarev A., Inoue K., Larom S. et al. 2018. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* 558: 595–599. DOI: [10.1038/s41586-018-0225-9](https://doi.org/10.1038/s41586-018-0225-9)

Rheims H., Rainey F.A., Stackebrandt E. 1996. A molecular approach to search for diversity among bacteria in the environment. *Journal of Industrial Microbiology and Biotechnology* 17: 159–169. DOI: [10.1007/BF01574689](https://doi.org/10.1007/BF01574689)

Rinke C., Schwientek P., Sczyrba A. et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499: 431–437. DOI: [10.1038/nature12352](https://doi.org/10.1038/nature12352)

Röth E., Marczin N., Balatonyi B. et al. 2011. Effect of a glutathione S-transferase inhibitor on oxidative stress and ischemia-reperfusion-induced apoptotic signalling of cultured cardiomyocytes. *Experimental and clinical cardiology* 16: 92–96.

Schneewind O., Missiakas D.M. 2012. Protein secretion and surface display in Gram-positive bacteria. *Philosophical*



Transactions of the Royal Society B: Biological Sciences 367: 1123–1139. DOI: [10.1098/rstb.2011.0210](https://doi.org/10.1098/rstb.2011.0210)

Schwengers O., Jelonek L., Dieckmann M.A. et al. 2021. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics* 7: 000685. DOI: [10.1099/mgen.0.000685](https://doi.org/10.1099/mgen.0.000685)

Scott M.D., Meshnick S.R., Eaton J.W. 1987. Superoxide dismutase-rich bacteria. Paradoxical increase in oxidant toxicity. *Journal of Biological Chemistry* 262: 3640–3645. DOI: [10.1016/S0021-9258\(18\)61401-2](https://doi.org/10.1016/S0021-9258(18)61401-2)

Sieber C.M.K. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* 3: 836–843. DOI: [10.1038/s41564-018-0171-1](https://doi.org/10.1038/s41564-018-0171-1)

Soro V., Dutton L.C., Sprague S.V. et al. 2014. Axenic culture of a candidate division TM7 bacterium from the human oral cavity and biofilm interactions with other oral bacteria. *Applied and Environmental Microbiology* 80: 6480–6489. DOI: [10.1128/AEM.01827-14](https://doi.org/10.1128/AEM.01827-14)

Tamames J., Puente-Sánchez F. 2019. SqueezeMeta, A highly portable, fully automatic metagenomic analysis pipeline. *Frontiers in Microbiology* 9: 3349. DOI: [10.3389/fmicb.2018.03349](https://doi.org/10.3389/fmicb.2018.03349)

Tammam S., Sampaleanu L.M., Koo J. et al. 2013. PilMNO PQ from the *Pseudomonas aeruginosa* type IV pilus system form a transenvelope protein interaction network that

interacts with PilA. *Journal of Bacteriology* 195: 2126–2135. DOI: [10.1128/JB.00032-13](https://doi.org/10.1128/JB.00032-13)

Uritskiy G.V., DiRuggiero J., Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6: 158. DOI: [10.1186/s40168-018-0541-1](https://doi.org/10.1186/s40168-018-0541-1)

Walker J.E. 2013. F-ATPases. In: Lennarz W.J., Lane M.D. (Eds.), *Encyclopedia of Biological Chemistry*. Waltham: Academic Press, pp. 269–274. DOI: [10.1016/B978-0-12-378630-2.00294-2](https://doi.org/10.1016/B978-0-12-378630-2.00294-2)

Walker J.E., Dickson V.K. 2006. The peripheral stalk of the mitochondrial ATP synthase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1757: 286–296. DOI: [10.1016/j.bbabi.2006.01.001](https://doi.org/10.1016/j.bbabi.2006.01.001)

Wood D.E., Lu J., Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology* 20: 257. DOI: [10.1186/s13059-019-1891-0](https://doi.org/10.1186/s13059-019-1891-0)

Wu Y.-W., Simmons B.A., Singer S.W. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32: 605–607. DOI: [10.1093/bioinformatics/btv638](https://doi.org/10.1093/bioinformatics/btv638)

Yakimov M.M., Merkel A.Y., Gaisin V.A. et al. 2022. Cultivation of a vampire: “*Candidatus Absconditicoccus praedator*”. *Environmental Microbiology* 24: 30–49. DOI: [10.1111/1462-2920.15823](https://doi.org/10.1111/1462-2920.15823)