**Supplementary Materials and Methods**

***Thioalkalivibrio* sp. 10fs10 genome sequencing**

*Thioalkalivibrio* sp. 10fs10 genomic DNA was extracted by GenElute™ Bacterial Genomic DNA, following producer's instructions. DNA quantification was performed using Qubit® 3.0 (Invitrogen, Thermo Fisher Life Technologies) following producer's instructions. *Thioalkalivibrio* sp.10fs10 genome was sequenced in St. Petersburg state University, St. Petersburg (Russia). One Mate-Pair (MP) libraries was prepared using FC-132-1001 Nextera® Mate Pair Sample Prep Kit for 1μg DNA input with the following modifications: DNA was size-selected after the tagmentation using 0.8 % agarose gel and the selected region was 2.5 - 6 kb. The libraries were sequenced in paired-end 2x100 mode on Illumina HiSeq 2500 using PE-402-4002 HiSeq® Rapid PE Cluster Kit v2 and FC-402- 4021 HiSeq® Rapid SBS Kit v2 (200 Cycle). Two Pair-End (PE) libraries were prepared using FC 121-1031 Nextera® DNA library preparation kit on the same amount of DNA and sequenced on MiSeq PE sequencing 2x300 using MS-102-3003 MiSeq® Reagent Kit v3 (600 cycle). Demultiplexed raw data were trimmed and quality filtered using Trimmomatic v0.36 [1] on paired-ends libraries and NxTrim v0.4.2 [2] on mate-paired library. Genome assembly was performed using the approach previously described [3], with some modifications. The draft genome of *Thioalkalivibrio* sp. isolate 10fs10 was generated performing multiple steps of assembly, using two assemblers: Abyss 2.0.1 [4] and SPAdes 3.10.1 [5]. A total of two fake reads libraries were generated using BBmap (BBMap - Bushnell B. - sourceforge.net/projects/bbmap/) from the first assembly (using minum lenght option of 1.5 Kbp and 10 Kbp respectively). The overlapping fake reads libraries were used to harmonize the two different libraries obtained with different sequencing platforms, to improve de novo assembly performance, as suggested by [6]). To resolve the gaps generated during scaffolding step, GapFiller v1.10 [7] was used. Finally, consensus sequences were processed using Prodigal 2.50 [8] for coding sequence (CDS) prediction. CDS aligned against different databases, such as Swissprot, TrEMBL, Pfam, Tigrfam, using Blast 2.5.0 [9] and HMMER [10]. Blast2go 4.1 [11] was used to align sequences against InterPro and Gene Ontology, BlastKoala [12] was used to annotate proteins in Kegg Orthology (KO) and tRNAscaSE 1.3 [13] was used to find tRNA sequences. Assembled DNA automatic annotations were also performed using RAST platform (http://rast.nmpdr.org) and IMG-ER platform [14]. Genome assembly and annotation project are available at IMG, under IMG-ER submission ID 183849.

Genes involved in bacterial homeostasis in saline and alkaline environments are reported in Table S1

**Table S1:** Genes involved in bacterial homeostasis in saline and alkaline environments annotated in *Thioalkalivibrio* sp. 10fs10 genome. IMG-ER and Rast identifiers are reported

| | | IMG identifier | Rast identifier | Product name |
|---|---|---|---|---|
| **osmolyte synthesis** | **Glycine/betaine synthesis** | Ga0265412_1001641 | fig\|6666666.276697.peg.648 | Sarcosine/ dimethylglycine N-methyltransferase |
| | | Ga0265412_1001642 | fig\|6666666.276697.peg.649 | Glycine/ sarcosine N-methyltransferase |
| | | Ga0265412_1001186 | fig\|6666666.276697.peg.199 | Sucrose-phosphate synthase. |

| | | | | |
|---|---|---|---|---|
| | **Sucrose/sucrose-phosphate synthesis** | Ga0265412_10012349 | fig\|6666666.276697.peg.2296 | Sucrose synthase |
| **cell wall reinforcement** | **Squalene synthesis** | Ga0265412_1001303 | fig\|6666666.276697.peg.316 | Dxs |
| | | Ga0265412_1001626 | fig\|6666666.276697.peg.632 | IspC/Dxr |
| | | Ga0265412_1001647 | fig\|6666666.276697.peg.655 | IspD |
| | | Ga0265412_100157 | fig\|6666666.276697.peg.71 | IspE |
| | | Ga0265412_1001648 | fig\|6666666.276697.peg.656 | IspF |
| | | Ga0265412_1001803 | fig\|6666666.276697.peg.802 | IspG |
| | | Ga0265412_10012613 | fig\|6666666.276697.peg.2563 | IspH |
| | | Ga0265412_1001304 | fig\|6666666.276697.peg.317 | Farnesyl diphosphate synthase (GTT) |
| | | Ga0265412_10011010 | fig\|6666666.276697.peg.1007 | HpnC |
| | | Ga0265412_10011009 | fig\|6666666.276697.peg.1006 | HpnD |
| | | Ga0265412_10011008 | fig\|6666666.276697.peg.1005 | HpnE |
| | **Cardiolipin synthesis** | Ga0265412_10011218 | fig\|6666666.276697.peg.1200 | Phosphatidyl-glycerolphosphate synthase |
| | | Ga0265412_10011563 | fig\|6666666.276697.peg.1543 | Phosphatidyl-glycerolphosphate synthase |
| | | Ga0265412_10011050 | fig\|6666666.276697.peg.1048 | Phosphatidyl-glycerophosphatase A |
| | | Ga0265412_10011279 | fig\|6666666.276697.peg.1261 | Cardiolipin synthase |
| | | Ga0265412_10011784 | fig\|6666666.276697.peg.1748 | Cardiolipin synthase |
| | | Ga0265412_10011846 | fig\|6666666.276697.peg.1810 | Cardiolipin synthase |
| **primary pumps** | **$Na^+$-dependent NADH: quinone oxidoreductase (2 complete operons)** | Ga0265412_10011966 | fig\|6666666.276697.peg.1927 | RnfA |
| | | Ga0265412_10011967 | fig\|6666666.276697.peg.1928 | RnfB |
| | | Ga0265412_10011968 | fig\|6666666.276697.peg.1929 | RnfC |
| | | Ga0265412_10011969 | fig\|6666666.276697.peg.1930 | RnfD |
| | | Ga0265412_10011970 | fig\|6666666.276697.peg.1931 | RnfG |
| | | Ga0265412_10011971 | fig\|6666666.276697.peg.1932 | RnfE |
| | | Ga0265412_10012911 | fig\|6666666.276697.peg.2848 | RnfA |
| | | Ga0265412_10012912 | fig\|6666666.276697.peg.2849 | RnfB |
| | | Ga0265412_10012913 | fig\|6666666.276697.peg.2850 | RnfC |
| | | Ga0265412_10012914 | fig\|6666666.276697.peg.2851 | RnfD |
| | | Ga0265412_10012915 | fig\|6666666.276697.peg.2852 | RnfG |
| | | Ga0265412_10012916 | fig\|6666666.276697.peg.2853 | RnfE |
| | **NADH dehydrogenase** | Ga0265412_1001451 | fig\|6666666.276697.peg.460 | NuoA |
| | | Ga0265412_1001452 | fig\|6666666.276697.peg.461 | NuoB |
| | | Ga0265412_1001453 | fig\|6666666.276697.peg.462 | NuoC |

| | | | | |
|---|---|---|---|---|
| | **(NDH-1) primary proton pump** | Ga0265412_1001454 | fig\|6666666.276697.peg.463 | NuoD |
| | | Ga0265412_1001455 | fig\|6666666.276697.peg.464 | NuoE |
| | | Ga0265412_1001456 | fig\|6666666.276697.peg.465 | NuoF |
| | | Ga0265412_1001457 | fig\|6666666.276697.peg.466 | NuoG |
| | | Ga0265412_1001458 | fig\|6666666.276697.peg.467 | NuoH |
| | | Ga0265412_1001459 | fig\|6666666.276697.peg.468 | NuoI |
| | | Ga0265412_1001460 | fig\|6666666.276697.peg.469 | NuoJ |
| | | Ga0265412_1001461 | fig\|6666666.276697.peg.470 | NuoK |
| | | Ga0265412_1001462 | fig\|6666666.276697.peg.471 | NuoL |
| | | Ga0265412_1001463 | fig\|6666666.276697.peg.472 | NuoM |
| | | Ga0265412_1001464 | fig\|6666666.276697.peg.473 | NuoN |
| **secondary pumps** | **Mrp sodium /proton antiporter complex** | Ga0265412_10011858 | fig\|6666666.276697.peg.1826 | MrpG |
| | | Ga0265412_10011859 | fig\|6666666.276697.peg.1825 | MrpF |
| | | Ga0265412_10011860 | fig\|6666666.276697.peg.1824 | MrpE |
| | | Ga0265412_10011861 | fig\|6666666.276697.peg.1823 | MrpD |
| | | Ga0265412_10011862 | fig\|6666666.276697.peg.1822 | MrpC |
| | | Ga0265412_10011863 | fig\|6666666.276697.peg.1821 | MrpB |
| | | Ga0265412_10011864 | fig\|6666666.276697.peg.1820 | MrpA |
| | **sodium/proton antiporter NhaD** | Ga0265412_10011856 | fig\|6666666.276697.peg.1818 | ArsB/NhaD |
| | **sulphate-dependent bicarbonate antiporter** | Ga0265412_10045 | fig\|6666666.276697.peg.3021 | SulP |

**Taxonomic characterisation of the *Thioalkalivibrio* sp. 10fs10**

The full-length gene coding for the 16S rRNA of *Thioalkalivibrio* sp. 10fs10 isolate was extracted from draft genome. The sequences of the genes coding for the 16S rRNA of 77 different *Thioalkalivibrio* sp. strains were downloaded from Silva SSU database and from Joint Genome Institute (JGI) using IMG. All sequences were aligned using MUSCLE algorithm (gap open penalty -400.00, gap extension 0.0, UPGMA clustering method, 16 iterations), and trimmed to match the length of the shortest deposited sequence (1374 nucleotides). Maximum likelihood tree was computed using Tamura-Nei DNA evolutionary model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, +G = 0.4053 with evolutionary invariable 150 sites (+I =58.8102 %). Initial tree(s) for the heuristic search were obtained by applying NeighborIn review Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. Highest log likelihood obtained was -5299.2315 and the accuracy of calculated tree was assessed by bootstrapping 1000 replicates.

**Maximum likelihood tree based on Multi-Locus sequence (MLS) analysis**

Maximum Likelihood MLS tree was computed using concatenated amino acid sequences of ClpA-DnaJ-GyrA-RpoH-RpoS-SecF proteins, codified by respective housekeeping genes obtained from genomic annotation of

*Thioalkalivibrio* sp. 10fs10 and from available complete genomes of all *Thioalkalivibrio* spp. on IMG/M. DNA sequences of the aforementioned housekeeping genes were translated using genetic codes of table 11 (NCBI). The corresponding sequences of amino acids were aligned with MUSCLE algorithm (gap open penalty -400.00, gap extension 0.0, UPGMA clustering method, 16 iterations), equally trimmed to a final total length of 3158 amino acids and concatenated in the reported order. Maximum likelihood tree was computed using a method based on the Whelan and Goldman + Freq. model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.5336). The rate variation model allowed for some sites to be evolutionarily invariable sites (+I = 7.08 %). Initial tree(s) for the heuristic search were obtained by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. Highest log likelihood obtained was -98651.57 and the accuracy of calculated tree was assessed by bootstrapping of 1000 replicates.

**Classification of *Thioalkalivibrio* sp. 10fs10 by, Genome Taxonomy Database**

Computation of Maximum likelihood tree, used for taxonomical classification as Genome Taxonomy Database protocol, was performed using a method based on the Whelan and Goldman + Freq. model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.7814). The rate variation model allowed for some sites to be evolutionarily invariable sites (+I = 9.02 %). Initial tree(s) for the heuristic search were obtained by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, successively selecting the topology with superior log likelihood value. Highest log likelihood obtained was -68671.60 and the accuracy of calculated tree was assessed by bootstrapping of 500 replicates.

**References of supplementary Materials and Methods**

[1] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20. doi:10.1093/bioinformatics/btu170.

[2] O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics, 2015;31:2035-2037. doi: 10.1093/bioinformatics/btv057.

[3] Berben T, Sorokin DY, Ivanova N, Pati A, Kyrpides N, Goodwin LA, Woyke T, Muyzer G. Partial genome sequence of *Thioalkalivibrio* thiocyanodenitrificans ARhD 1T, a chemolithoautotrophic haloalkaliphilic sulfur-oxidizing bacterium capable of complete denitrification. Stand Genomic Sci 2015;10:84. doi:10.1186/s40793-015-0080-3.

[4] Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res 2017;27:768–77. doi:10.1101/gr.214346.116.

[5] Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 2013;20:714–37. doi:10.1089/cmb.2013.0084.

[6] Utturkar SM, Klingeman DM, Land ML, Schadt CW, Doktycz MJ, Pelletier DA, Brown SD. Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. Bioinformatics 2014;30:2709–16. doi:10.1093/bioinformatics/btu391.

[7] Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol 2012;13:R56. doi:10.1186/gb-2012-13-6-r56.

[8] Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119. doi:10.1186/1471-2105-11-119.

[9] McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 2004;32:W20–5. doi:10.1093/nar/gkh435.

[10] Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res 2013;41:e121–e121. doi:10.1093/nar/gkt263.

[11] Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005;21:3674–6. doi:10.1093/bioinformatics/bti610.

[12] Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 2016;428:726–31. doi:10.1016/j.jmb.2015.11.006.

[13] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.

[14] Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics 2009;25:2271–8. doi:10.1093/bioinformatics/btp393.