

The diversity of unique 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid coding common genes and novel Universal stress protein in Ectoine TRAP cluster (UspA) in 32 *Halomonas* species

Bhagwan Rekadwad (✉ rekadwad@gmail.com)

Yenepoya Research Centre, Yenepoya (Deemed to be University), University Road, Deralakatte, Mangalore 575018, Karnataka, India <https://orcid.org/0000-0002-0954-0180>

Wen-Jun Li

School of Life Sciences, Sun Yat-Sen University, Guangzhou, 510275, China

Rekha PD

Yenepoya Research Centre, Yenepoya (Deemed to be University), University Road, Deralakatte, Mangalore 575018, Karnataka, India

Research Article

Keywords: Ectoine, Life under extreme conditions, Saline environments, Single-copy genes, Ancient bacteria and Archaea, Bioactive compounds

Posted Date: May 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-537023/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at BMC Research Notes on August 3rd, 2021. See the published version at <https://doi.org/10.1186/s13104-021-05689-3>.

Abstract

Objectives To decipher the diversity of unique ectoine-coding housekeeping genes in the genus *Halomonas*.

Results In *Halomonas*, 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid has a crucial role as a stress-tolerant chaperone, a compatible solute, a cell membrane stabilizer, and a reduction in cell damage under stressful conditions. Apart from the current 16S rRNA biomarker, it serves as a blueprint for identifying *Halomonas* species. *Halomonas* elongate 1H9 was found to have 11 ectoine-coding genes. The presence of a superfamily of conserved ectoine-coding among members of the genus *Halomonas* was discovered after genome annotations of 93 *Halomonas* spp. As a result of the inclusion of 11 single copy ectoine coding genes in 32 *Halomonas* spp., genome-wide evaluations of ectoine coding genes indicate that 32 *Halomonas* spp. have a very strong association with *Halomonas elongata* 1H9, which has been proven evidence-based approach to elucidate phylogenetic relatedness of ectoine-coding child taxa in the genus *Halomonas*. Total 32 *Halomonas* species have a single copy number of 11 distinct ectoine-coding genes that help *Halomonas* spp. produce ectoine under stressful conditions. Furthermore, the existence of the Universal stress protein (UspA) gene suggests that *Halomonas* species developed directly from primitive bacteria, highlighting the role of *Halomonas* species in evolutionary terms.

Introduction

1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid is a natural pigment accumulates within the cytoplasm of salt-loving bacteria (e.g. genus *Ectothiorhodospira*) having osmoregulatory function termed as 'ectoine' (Galinski et al., 1985). The moderately halophilic members of family Halomonadaceae displays osmoadaptation facilitated by betaine, ectoine (Cánovas et al., 1996) and hydroxyectoine (Vargas et al., 2008). Family Halomonadaceae possess total 18 child taxa. Of these, 14 child taxa are with validly published and correct name, while 16 child taxa are with validly published name including synonyms under the International Code of Nomenclature of Prokaryotes (ICNP). On similar note, Genus *Halomonas* represented by 114 type strains with 112 candidates with validity published name and correct name and 10 with synonyms. Also, three species are orthographic misspelled variant and 18 invalidated species were not validated by ICNP (LPSN, 2021). *Halomonas* species are producer of, biotechnologically important biomolecules, ectoine. Being suspended in the cytoplasm, ectoine and hydroxyectoine coded by *Halomonas* species has benefits to cell. It acts as stress tolerant chaperones, as a compatible solute, stabilize of cell membrane and reduce cell damage. Moreover, ectoine and hydroxyectoines are high-value chemicals and exploited for cosmetics, immune protection, stabilization of antibodies, anti-inflammatory and tissue protective agent, for co-production of bioplastic polyhydroxybutyrate (Melmer and Schwarz, 2009), skin raging and protectant (wind, humidity, radiation and extreme temperatures), whole cell and macromolecule protection (freezing, drying, high salinity, heat stress, oxygen radicals, radiation and denaturing agents), protein protection and biomedical applications (Pastor et al., 2010). Various applications of *Halomonas* species reflect presence of diverse gene profiles in their genomes. It is therefore important to evaluate indicative signatures of ectoine coding genes among the genus *Halomonas* that governs them to perform vital biological function under extreme environmental conditions.

In current study, a blue print of ectoine coding genes was identified from *Halomonas elongata*. Genome annotations of existing *Halomonas* spp., uncovered existence of some common genes that codes ectoine (s) among members of the genus *Halomonas*. Thus, genome-wide evaluations of ectoine coding genes were also examined. We also analyzed highly close 32 *Halomonas* spp., with *Halomonas elongata* 1H9, which has phylogenetic related ectoine coding child taxa inferred using identified single copy genes.

Methods

128 type strains 16S rRNA genes and ninety-four *Halomonas* spp., genomes

One hundred twenty-eight 16S rRNA genes of type strains and ninety-four complete genomes and reference sequences of *Halomonas* spp., were obtained from LPSN and NCBI genome database.

Radar chart

Halomonas spp., possesses multiple quantitative variables (species in particular) i.e. variable genome length/data points for visualization. Radar chart makes the way easy to compare the intra-species variable length to see similar values and find high or low scoring within outliers in the genus.

RAST Genome analysis

Complete genome sequences of *Ectothiorhodospira haloalkaliphila* ATCC 51935 (CP007268), *Halomonas elongata* 1H9 (NC_014532), *Halorhodospira halochloris* DSM 1059 (AP017372) and *Halorhodospira halophila* SL1 (CP000544) analyses done using RAST v2.0 (<https://rast.nmpdr.org/>). RAST server is a SEED-based National Microbial Pathogen Database Resource (NMPDR), prokaryotic genome annotation service, to predict system coverage, subsystem category distribution and subsystem feature count (Brettin et al., 2015).

Identification of protein families and single copy genes

Protein families and single-copy genes in 93 *Halomonas* spp., were identified using PATRIC 3.6.9 (<https://www.patricbrc.org/>). PLfams within the genus were computed with MCL inflation = 3.0 to obtain higher sequence similarity and better specificity for intra-genus/species close comparisons.

Selecting single copy number genes

PLfams of 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid coding genes among 93 *Halomonas* spp., were extracted. Common genes coded by *Halomonas* species were selected for analysis. The topology of the phylogenetic tree generated using concatenated sequences was compared with the topology of 16S rRNA based *Halomonas* spp., child taxa tree.

Phylogeny reconstruction and topology analysis

The evolutionary history of one hundred twenty-eight 16S rRNA and 33 *Halomonas* single-copy genes were inferred using standalone tool MEGA X with 1000 bootstrap analysis followed by best scoring ML, NJ and ME tree. The Jukes-Cantor method and are in the units of the number of base substitutions per site. The closest child taxa of biotechnological important ectoine producing *Halomonas elongata* 1H9 were deciphered. It helps for phylogenetic analysis and topology comparison to delineate nearest species and 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid gene coding species.

Results

Phylogenetic analysis of 16S rRNA genes in the genus *Halomonas*

Halomonas elongata 1H9 is a saline bacterium known for 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid (ectoine) producing abilities under extreme environmental condition.

RAST genome analysis of the *Halomonas elongata* 1H9 shows that various subsystem feature counts consists of different pathways (Figure 3). Similarly, other members of the genus *Halomonas* are known to code and produce molecular variants 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid. Therefore, the diversity of ectoine coding *Halomonas* might form a cluster with a similar kind of *Halomonas* species. Hence, Phylogenetic analysis of 16S rRNA sequences of type strain amongst genus *Halomonas* revealed that type strains AJ261, 1H9, M8, 5-3, RS-16, AAD6, SS20, 11S, NTU-107, TBZ21, 5CR, F8-11, SL014B-69, TBZ202, KCTC 42685, Z-7009, SL014B-85, CIP 105456, 204, KMM 1376, 10-C-3, Hwa etc., (Figure 4) were clustered together. It suggests that those species have a similar gene pool regardless of their genome length forms cluster. Variation in some clusters may occur due to the use of single 16S rRNA genes for phylogenetic analysis. Hence, members of the genus *Halomonas* might possess similar single-copy ectoine coding genes reveals that apart from the 16S RNA gene.

Identification of protein families, single copy genes and pearson correlation

Whole-genome analyses and annotation have resolved the misery of uniqueness amongst *Halomonas* spp. The radar chart shows that existing genomic data of *Halomonas* spp., possesses complete genome sequences, reference genomes and some scaffolds (Figure 5). Available genomic sequence data shows a similar gene pool and all ectoine-coding sequences from 93 type strains not having sets of genes. To resolve this issue and find relevant species in the genus *Halomonas*, we, therefore, annotated all genomes and identified the single-copy gene that codes ectoine. It was noticed that few *Halomonas* species that more than 11 single copy ectoine-coding genes. Therefore, inferred ML tree (Figure 6) some type strains shows that ectoine biomarker (in 1H9, F9-6, AJ261, SP4, ACAM 71, 62, Hb3, DSM 15911, N12, NTU-107, G-16.1, ZJ2214, TBZ3, M29, 79, BJGMM-B45, LCB169, CFH 9008, AIR-2, DQD2-30, 4A, SL014B-69, TBZ202, DX6, 9-2 and MC28) possessed by species were more or less similar kind of representative species similar to concatenated sequence of 32 *Halomonas* species (Figure 1). It was observed that of the 93 annotated genome sequences, 31 + 1 (32) species have 11 ectoine coding genes (DoeA-DoeC-DoeX-EctC-EctD-EutB-EutC-TeaA-TeaB-TeaC-UspA) as single copy number genes (Figure 7; Table 1). Heatmap of 11 ectoine coding genes shows high degree of pearson correlation (Figure 2) value lies between 0.50 and ± 1 (0 = no correlation, 1 = high degree correlation).

Novel Universal stress protein in Ectoine TRAP cluster (UspA) and resistance mediated by UspA gene

Studies on genome sequence analyses and analysis of various ectoine coding in *Halomonas* spp., uncovered that type strains viz. *H. aestuarii* Hb2 (NZ_CP018139), *H. anticariensis* DSM 16096 (GCF_000409775), *H. azerbaijanica* TBZ202 (GCF_004551485), *H. bachuensis* DX6 (GCA_011742165), *H. beimenensis* NTU-111 (NZ_CP021435), *H. campisalis* SS10-MC5 (NZ_CP065435), *H. caseinilytica* DSM 18067 (GCF_001662285), *H. cerina* CECT 7282 (GCF_014192215), *H. cupida* (GCF_900142755), *H. daqingensis* CGMCC 1.6443 (GCF_900108215), *H. denitrificans* DSM 18045 (GCF_003056305), *H. endophytica* MC28 (GCF_002879615), *H. eurihalina* MS1 (GCF_008274785), *H. gudaonensis* (GCF_900100195), *H. halmophila* NBRC 15537 (GCF_006540005), *H. heilongjiangensis* 9-2 (GCF_003202165), *H. huangheensis* BJGMM-B45 (NZ_CP013106), *H. kenyensis* DSM 17331 (GCF_013697085), *H. korlensis* CGMCC 1.6981 (GCF_900116705), *H. lactosivorans* KCTC 52281 (GCF_003254665), *H. litopenaei* SYSU ZJ2214 (GCF_003045775), *H. niordiana* ATF 5.4 (GCF_004798965), *H. organivorans* CECT 5995 (GCF_014192055), *H. pacifica* (GCF_007989625), *H. qijiaojingensis* KCTC 22228 (GCF_014651875), *H. saliphila* LCB169 (GCF_002930105), *H. stenophila* CECT 7744 (GCF_014192275), *H. taeanensis* (GCF_900100755), *H. urmiana* TBZ3 (GCF_005780185), *H. ventosae* (GCF_004363555), *H. xinjiangensis* TRM 0175 (GCF_000759345) and *H. zincidurans* B6 (GCF_000731955) possess superfamily of conserved gene – UspA - suggests that the UspA gene/domain has been inherited from ancient protein family found in primitive bacteria. UspA protein helps *Halomonas* species provide support and assist *Halomonas* to function and produce ectoine in the saline environment under stressful conditions like high salt, low water activity and low temperature etc. Hence, UspA – stress protein – found in 32 species is a new report in the genus *Halomonas*.

Conclusion

Ectoine signatures can be found in 93 *Halomonas* genome sequences that are publicly available. 32 *Halomonas* species have 11 separate ectoine genes in a single copy number in their genomes, which help *Halomonas* spp. produce ectoine under stressful conditions. Based on existing genomic data, it was discovered that *Halomonas elongata* 1H9 has distinct ectoine-producing machinery from other *Halomonas* species. The existence of 11 distinct genes in 32 species, including the UspA gene, suggests that *Halomonas* species evolved directly from their primitive ancestor, shedding light on their evolutionary significance.

Limitations

A possible restriction would be the presence of biomarkers other than existing ectoine-coding genes responsible for *Halomonas* spp. producing 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid.

Abbreviations

CDS = Coding Sequence/the coding region of the gene

DoeA = Ectoine hydrolase

DoeC = Aspartate-semialdehyde dehydrogenase DoeC in ectoine degradation (EC 1.2.1.11)

DoeX = DNA-binding protein DoeX, ectoine utilization regulator

EctC = L-ectoine synthase (EC 4.2.1.108)

EctD = Ectoine hydroxylase

EutB = Ectoine utilization protein EutB, threonine dehydratase-like

EutC = Ectoine utilization protein EutC, similar to ornithine cyclodeaminase

ICNP = the International Code of Nomenclature of Prokaryotes

LPSN = List of Prokaryotic names with Standing in Nomenclature

MCL = The Markov Cluster Algorithm

ME = Minimum Evolution method

MEGA X = Molecular Evolutionary Genetics Analysis across computing platforms version 10.0

MEGA X = Molecular Evolutionary Genetics Analysis, Version 10

ML = Maximum-Likelihood method

NCBI = National Center for Biotechnology Information

NJ = Neighbor-Joining method

NMPDR = National Microbial Pathogen Database Resource

PLfams = Genus-specific families

RAST = Rapid Annotation using Subsystem Technology

rRNA = ribosomal RNA

spp. = species

TeaA = Ectoine/hydroxyectoine TRAP transporter substrate-binding periplasmic protein TeaA

TeaB = Ectoine/hydroxyectoine TRAP transporter small permease protein TeaB

TeaC = Ectoine/hydroxyectoine TRAP transporter large permease protein TeaC

UspA = Universal stress protein UspA in Ectoine TRAP cluster

Declarations

Ethics approval and consent to participate

Not applicable

Consent for Publication

Not applicable

Availability of data and material

Data is available within this manuscript.

Competing interests

The author declares no competing interests.

Funding

BR has used personal funds for this work.

Authors' contributions

BR conceived the original idea, collected the data, analyses, prepared figures, writing of entire manuscript. BR, WJL and RPD have revised and critically evaluated the drafts. All authors read and approved the final manuscript.

Acknowledgements

BR is thankful to editors and the reviewers for their comment to improve manuscript in present form.

References

1. Galinski, E.A., Pfeiffer, H.P., & Truper, H.G. 1985. 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid. *European J. Biochem.* 149: 135-139. <https://doi.org/10.1111/j.1432-1033.1985.tb08903.x>
2. Cánovas, D., Vargas, C., Csonka, L.N., Ventosa, A., & Nieto, J.J. 1996. Osmoprotectants in *Halomonas elongata*: high-affinity betaine transport system and choline-betaine pathway. *J. Bacteriol.* 178: 7221-7226. doi: 10.1128/jb.178.24.7221-7226
3. LPSN-List of Prokaryotic names with Standing in Nomenclature. <https://lpsn.dsmz.de/> Assessed on 21 April 2021
4. Melmer, G., & Schwarz, T. (2009). Ectoines: a new type of compatible solutes with great commercial potential. *Extremophiles – Vol. II, Encyclopedia of Life Support Systems (EOLSS)*, pp382. ISBN: 978-1-905839-93-3 (eBook), ISBN: 978-1-84826-993-4 (Print)
5. Pastor, J.M., Salvador, M., Argandoña, M., Bernal, V., Reina-Bueno, M., Csonka, L. N., Iborra, J. L., Vargas, C., Nieto, J. J., & Cánovas, M. 2010. Ectoines in cell stress protection: Uses and biotechnological production. *Biotech. Adv.* 28: 782-801. <https://doi.org/10.1016/j.biotechadv.2010.06.005>
6. Vargas, C., Argandoña, M., Reina-Bueno, M., Rodríguez-Moya, J., Fernández-Aunión, C., & Nieto, J.J. 2008. Unravelling the adaptation responses to osmotic and temperature stress in *Chromohalobacter salexigens*, a bacterium with broad salinity tolerance. *Aqua Biosyst.* 4: 14. <https://doi.org/10.1186/1746-1448-4-14>
7. RAST-Rapid Annotation using Subsystem Technology, v2.0. (<https://rast.nmpdr.org/>) Assessed on 15 April 2021
8. NCBI-National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/> Assessed on 21 April 2021
9. Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., 3rd, Stevens, R., Vonstein, V., Wattam, A. R., & Xia, F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5: 8365. <https://doi.org/10.1038/srep08365>

Table

Table 1. Functions of ectoine-coding genes in the Genus *Halomonas* under different scenarios

Gene Symbol	Gene Description/ Accepted name	EC Number	KEGG Id (KO) & Name	Gene name in Patric Server	Comment/ Involved in Pathways & KEGG KO Id
DoeA	ectoine hydrolase DoeA	3.5.4.44	—	Ectoine hydrolase	<ul style="list-style-type: none"> · The enzyme, found in some halophilic bacteria, is involved in the degradation of the compatible solute ectoine. · The enzyme, which belongs to peptidase family M24, only acts in the direction of ectoine hydrolysis. · It also produces smaller amounts of (2S)-4-acetamido-2-aminobutanoate, which is recycled back to ectoine by EC 4.2.1.108
DoeC	Aspartate-semialdehyde dehydrogenase	1.2.1.11	—	Aspartate-semialdehyde dehydrogenase DoeC in ectoine degradation	<ul style="list-style-type: none"> · Threonine and Homoserine Biosynthesis · Lysine Biosynthesis DAP Pathway, GJO scratch · Lysine Biosynthesis DAP Pathway
DoeX	DNA-binding protein DoeX, ectoine utilization regulator	No EC recorded	K15782 & Lrp/AsnC family transcriptional regulator, regulator of ectoine-degradation genes	DNA-binding protein DoeX, ectoine utilization regulator	<ul style="list-style-type: none"> · Glycine, serine and threonine metabolism · Monobactam biosynthesis · Cysteine and methionine metabolism · Lysine biosynthesis · Metabolic pathways · Biosynthesis of secondary metabolites · Microbial metabolism in diverse environments
EctC	Ectoine synthase	4.2.1.108	—	L-ectoine synthase (EC 4.2.1.108)	<ul style="list-style-type: none"> · Ectoine is an osmoprotectant that is found in halophilic eubacteria.

					<ul style="list-style-type: none"> · This enzyme is part of the ectoine biosynthesis pathway and only acts in the direction of ectoine formation.
EctD	Ectoine hydroxylase	1.14.11.55	—	Ectoine hydroxylase	<ul style="list-style-type: none"> · The enzyme, found in bacteria, is specific for ectoine · Glycine, serine and threonine metabolism · Metabolic pathways
EutB	Ethanolamine ammonia-lyase large subunit	4.3.1.7	—	Ectoine utilization protein EutB, threonine dehydratase-like	<ul style="list-style-type: none"> · Glycerophospholipid metabolism · Metabolic pathways
EutC	Ethanolamine ammonia-lyase small subunit	4.3.1.7	—	Ectoine utilization protein EutC, similar to ornithine cyclodeaminase	<ul style="list-style-type: none"> · Glycerophospholipid metabolism · Metabolic pathways
TeaA	TRAP transporter substrate-binding protein	No EC recorded	K11688 & C4-dicarboxylate-binding protein (DctP)	Ectoine/hydroxyectoine TRAP transporter substrate-binding periplasmic protein TeaA	—
TeaB	Ectoine/hydroxyectoine TRAP transporter small permease protein	No EC recorded	—	Ectoine/hydroxyectoine TRAP transporter small permease protein TeaB	—
TeaC	Ectoine/hydroxyectoine TRAP transporter large permease protein	No EC recorded	K11690 C4-dicarboxylate transporter, DctM subunit	Ectoine/hydroxyectoine TRAP transporter large permease protein TeaC	—
UspA	Universal stress protein A	No EC recorded		Universal stress protein UspA in Ectoine TRAP cluster	—

Figures

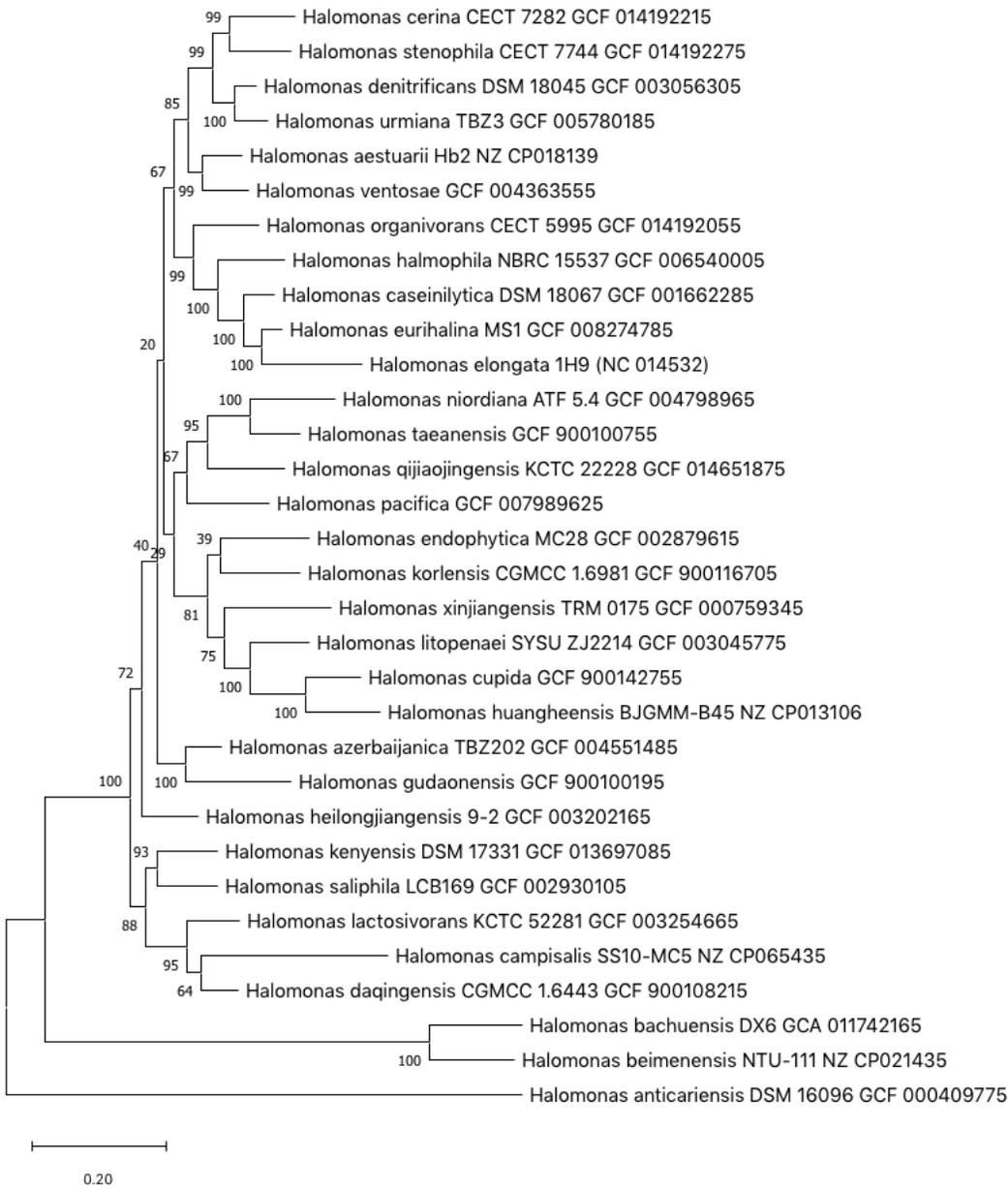


Figure 1

Maximum-likelihood (ML) analysis of concatenated sequences of 11 genes (DoeA-DoeC-DoeX-EctC-EctD-EutB-EutC-TeaA-TeaB-TeaC-UspA) from 32 *Halomonas* species in MEGA X. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates).

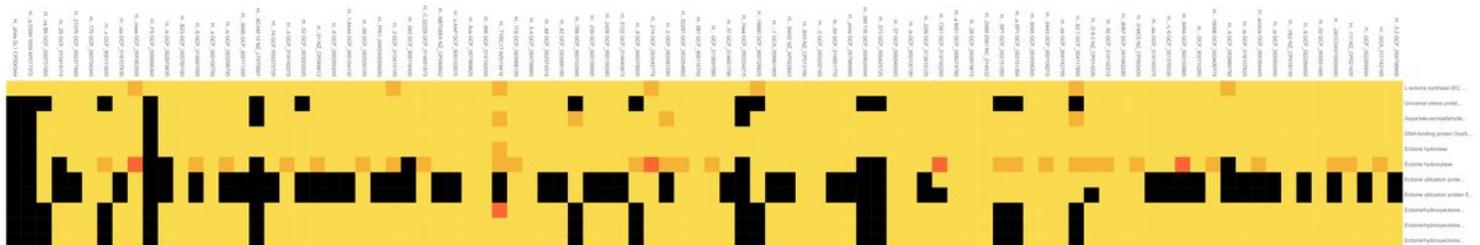


Figure 2

Heatmap of 11 ectoine coding genes in *Halomonas* spp., showing genome and protein-pairwise average linkage using pearson correlation

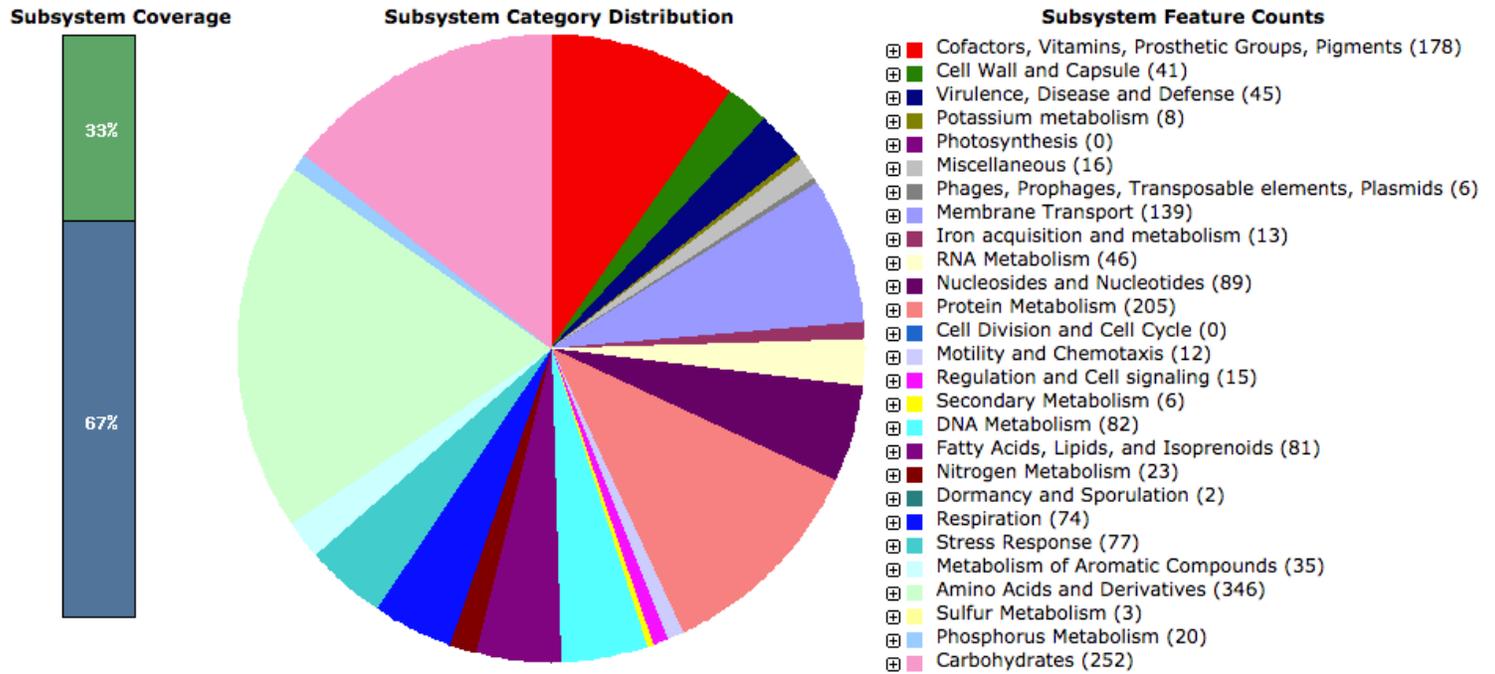


Figure 3

RAST genome analysis of *Halomonas elongata* 1H9

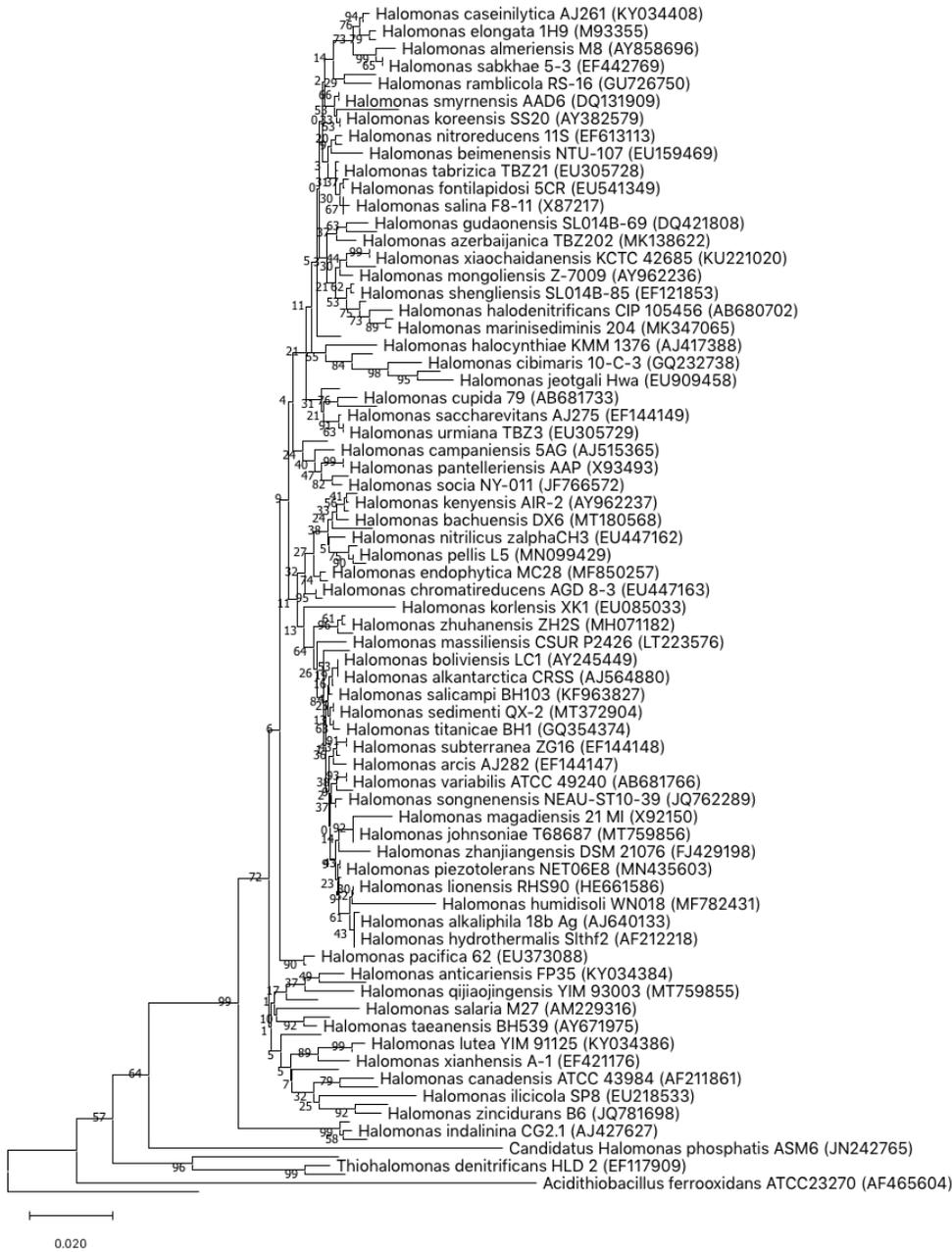


Figure 4

The evolutionary history of *Halomonas* species was inferred using the Neighbor-Joining method. Analysis using 16S rRNA gene sequences were conducted in MEGA X. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates).

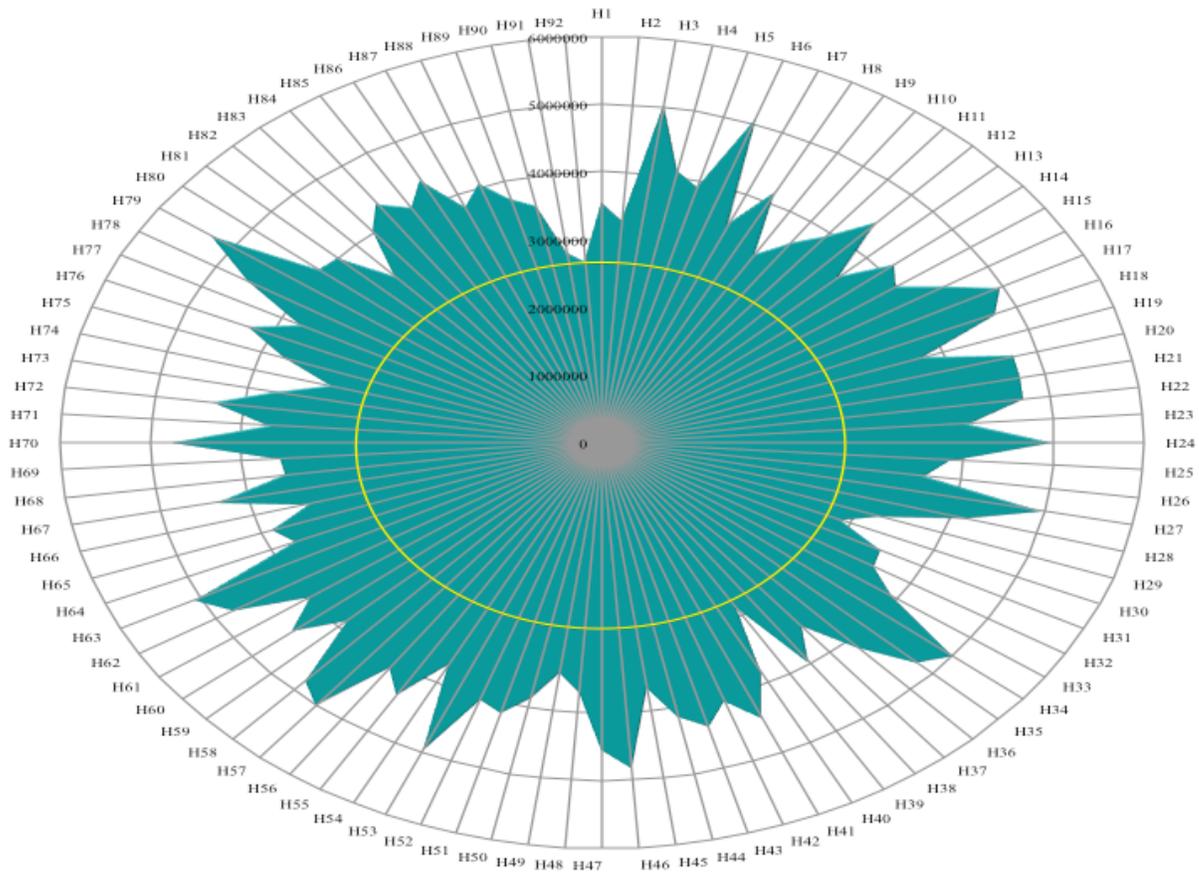


Figure 5

RADAR Chart of Genus *Halomonas* spp. (see supplementary table F1 for names of the species). Yellow circle indicates average genome length of each species.

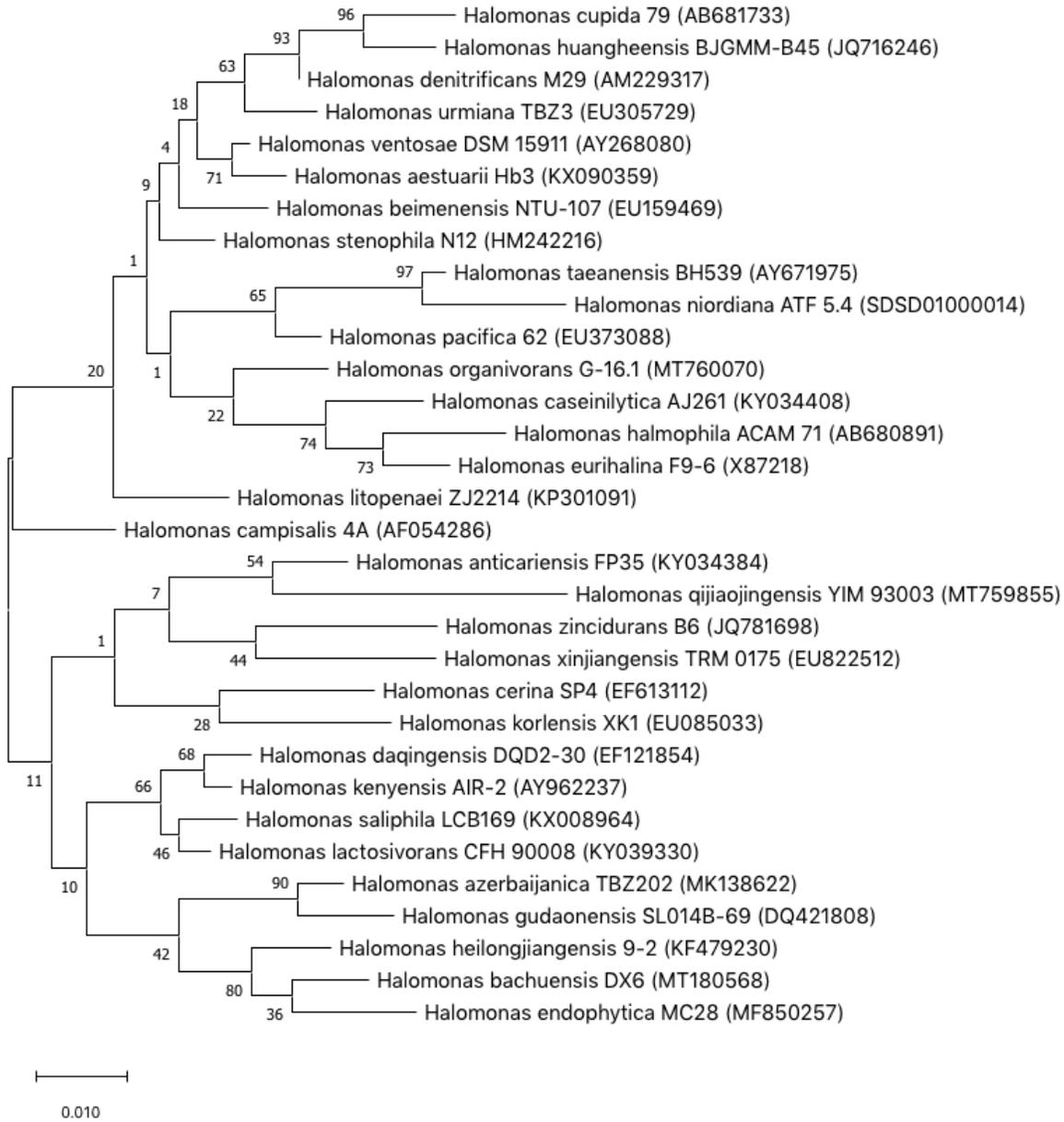


Figure 6

Maximum-likelihood analysis among *Halomonas* species was inferred from 16S rRNA gene sequences in MEGA X. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates).

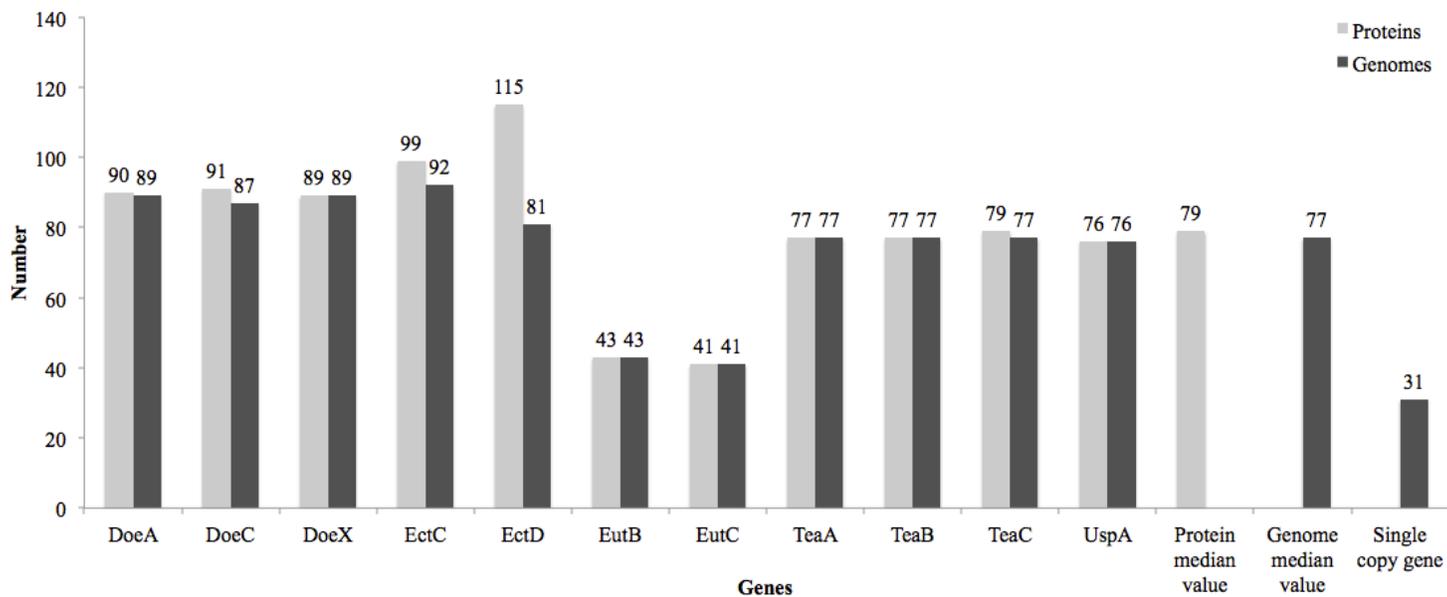


Figure 7

Single copy ectoine coding genes in the genus *Halomonas*

Supplementary Files

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

- [TableS1.csv](#)