Description of four new *Salinibacter* species, two cultivated and named following the rules of the bacteriological code: *Salinibacter pepae* sp. nov., *Salinibacter grassmerensis* sp. nov.; and two uncultivated and named following the rules of the SeqCode: *Salinibacter abyssi* sp. nov., and *Salinibacter pampae* sp. nov.

Tomeu Viver[1,2], Roth E. Conrad[3,4], Marianna Lucio[5], Mourad Harir[5,6], Mercedes Urdiain[1], Juan F. Gago[1], Ana Suárez-Suárez[1], Esteban Bustos-Caparros[1], Rodrigo Sanchez-Martinez[7], Eva Mayol[7], Federico Fassetta[8], Jinfeng Pang[9], Ionuț Mădălin Gridan[10], Stephanus Venter[11], Fernando Santos[7], Bonnie Baxter[12], María E. Llames[8], Adorján Cristea[13], Horia L. Banciu[14,15], Brian P. Hedlund[9], Matthew B. Stott[16], Peter Kämpfer[17], Rudolf Amann[2], Philippe Schmitt-Kopplin[5,6], Konstantinos T. Konstantinidis[3,4], Ramon Rossello-Mora[1]

Affiliations

[1] Marine Microbiology Group, Department of Animal and Microbial Biodiversity, Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB), Esporles, Spain.

[2] Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Bremen, Germany.

[3] Ocean Science & Engineering, School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA.

[4] School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

[5] Research Unit Analytical BioGeoChemistry, Helmholtz Munich, 85764 Neuherberg, Germany

[6] Chair of Analytical Food Chemistry, Technical University Munich, Maximus-von-Imhof-Forum 2, 85354 Freising, Germany

[7] Department of Physiology, Genetics and Microbiology, University of Alicante, 03690, San Vicent del Raspeig, Alicante, Spain.

[8] Laboratorio de Ecología Acuática, Instituto Tecnológico Chascomús (INTECH)-CONICET-UNSAM; Escuela de Bio y Nanotecnologías -UNSAM, Buenos Aires, Argentina

[9] School of Life Sciences, University of Nevada, Las Vegas, Nevada 89154-4004, USA

[10] Doctoral School of Integrative Biology, Faculty of Biology and Geology, Babeș-Bolyai University, Cluj-Napoca, Romania

[11] Department of Biochemistry, Genetics and Microbiology, and Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa.

[12] Great Salt Lake Institute, Westminster College, Salt Lake City, UT, 84105, U.S.A

[13] Department of Taxonomy and Ecology, Faculty of Biology and Geology, Babeș-Bolyai University, Cluj-Napoca, Romania

[14] Department of Molecular Biology and Biotechnology, Faculty of Biology and Geology, Babeș-Bolyai University, Cluj-Napoca, Romania

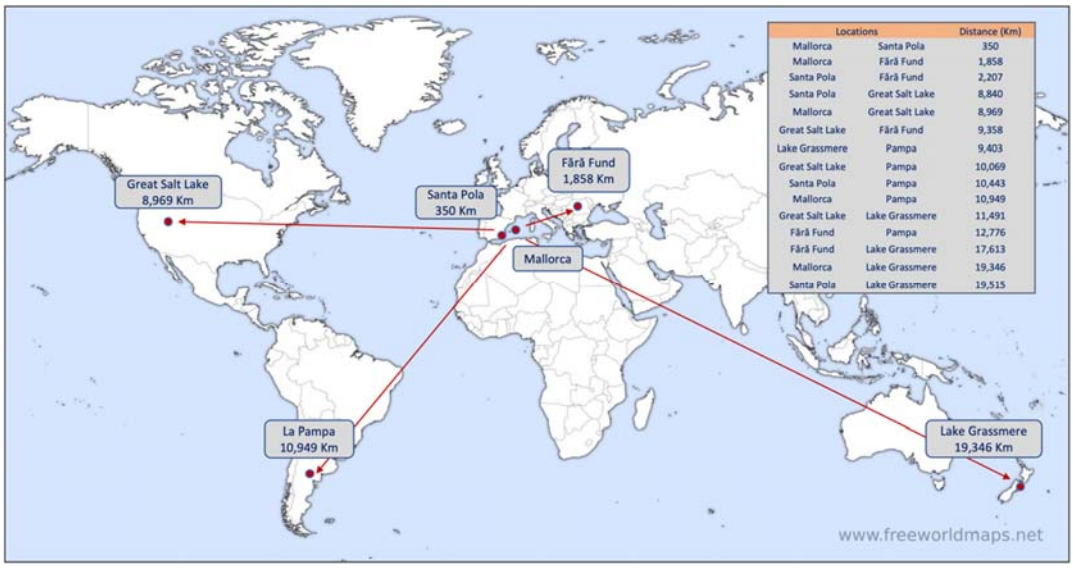
[15] Emil G. Racoviță Institute, Babeș-Bolyai University, Cluj-Napoca, Romania

[16] School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

[17] Institute of Applied Microbiology (IFZ), Justus Liebig Universität Giessen, Giessen, Germany.

Corresponding authors: Ramon Rosselló-Móra (ramon@imedea.uib-csic.es) and Tomeu Viver (tviver@imedea.uib-csic.es).

Figure S1. Geographical distribution of the six locations studied here and their inter-location distances (given in the table). The distances given in each location refer to Mallorca. The hypersaline here are: Mallorca (Es Trenc and S'Avall and separated by 3.5 km) and Santa Pola in Spain, Fără Fund in Romania, Great Salt Lake (USA), Pampa (Laguna Colorada Chica and Laguna Colorada Grande, separated by 23 Km), and Lake Grassmere (New Zealand).



Locations		Distance (Km)
Mallorca	Santa Pola	350
Mallorca	Fără Fund	1,858
Santa Pola	Fără Fund	2,207
Santa Pola	Great Salt Lake	8,840
Mallorca	Great Salt Lake	8,969
Great Salt Lake	Fără Fund	9,358
Lake Grassmere	Pampa	9,403
Great Salt Lake	Pampa	10,069
Santa Pola	Pampa	10,443
Mallorca	Pampa	10,949
Great Salt Lake	Lake Grassmere	11,491
Fără Fund	Pampa	12,776
Fără Fund	Lake Grassmere	17,613
Mallorca	Lake Grassmere	19,346
Santa Pola	Lake Grassmere	19,515

Table S1. Statistics of *Salinibacter* MAGs recovered using different bioinformatic tools. MAGs selected for further analysis and taxonomic classification are marked in red.

Sample (metagenome)	Assembler tool	Binning tool	Minimum contig length	Size (bp)	Num. contigs	N50	Min. contig length	1st Quartile	Media	Mean	3rd Quartile	Maximum contig length	Completeness (%)	Contamination (%)	rRNA
Romania, Fără Fund (FF)	SPADES	MetaBAT	**2,000**	3,018,448	279	16,630	2,041	4,059	6,986	10,819	15,225	60,507	90.2	2.9	23S/5S/16S
			5,000	2,700,571	185	17,946	5,019	7,228	11,277	14,598	18,243	60,507	82.4	2.0	23S/5S/16S
	MegaHIT	MetaBAT	**2,000**	3,513,196	420	11,218	2,000	3,326	5,917	8,364	9,774	64,855	87.3	4.9	23S/5S/16S
			5,000	2,964,537	248	13,591	5,046	6,519	8,659	11,953	13,695	64,855	80.2	3.7	23S/5S/16S
Argentina, Colorada Chica (CCH)	SPADES	MetaBAT	**2,000**	2,890,006	246	19,624	2,003	3,647	6,539	11,748	14,001	150,509	88.2	2.9	NO
			5,000	2,602,835	155	21,011	5,054	7,682	11,853	16,792	20,080	150,509	81.4	2.9	NO
	MegaHIT	MetaBAT	**2,000**	3,088,460	224	20,253	2,522	4,892	9,674	13,788	19,190	60,535	98	2.9	NO
			5,000	2,872,401	168	20,972	5,012	8,002	13,924	17,097	21,029	60,535	64.7	2.9	NO
Argentina, Colorarda Grande (CG)	SPADES	MetaBAT	**2,000**	2,809,353	321	13,734	2,013	3,328	5,580	8,751	11,225	61,085	84	2.9	NO
			5,000	2,225,774	164	16,282	5,084	6,779	10,857	13,571	16,683	61,085	50	1	NO
	MegaHIT	MetaBAT	**2,000**	3,147,671	278	16,602	2,014	4,352	7,612	11,322	14,982	65,533	96	3	NO
			5,000	2,834,870	193	18,634	5,012	7,273	11,659	14,688	18,634	65,533	83	2.9	NO

Text S1: In addition to the taxonomic study, here we assessed the different tools SPADES and MegaHIT assemblers, and MetaBAT binning tools selecting contigs with length > 2,000 pbs or > 5,000 pbs to retrieve the best quality MAG (Sup. Table S1). We observed that different tools and combinations rendered different qualities of the MAGs. The selection of contigs with length > 5,000 pbs binning rendered MAGs with a lower genome size, lower completeness and similar contamination. From the FF metagenome (Romania), the highest quality MAG was retrieved using the assembler SPADES and MetaBAT binning tool using a minimum contig length of 2,000 pbs. From CCH and CG metagenomes (Argentina), the highest quality MAGs were retrieved using the assembler MegaHIT and MetaBAT binning tool using a minimum contig length of 2,000 pbs (Sup. Table S1).

Figure S2. Genomic clustering based on Average Amino-acid Identity (AAI) of genomes included in the study.

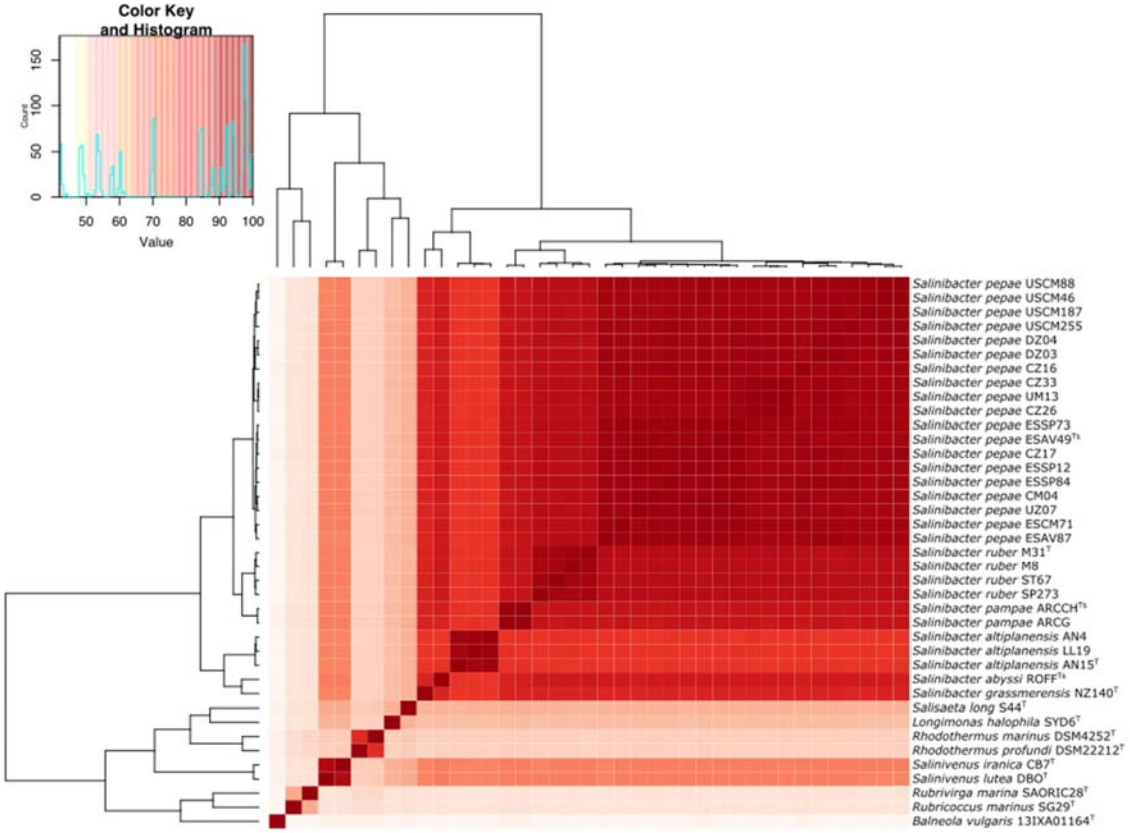
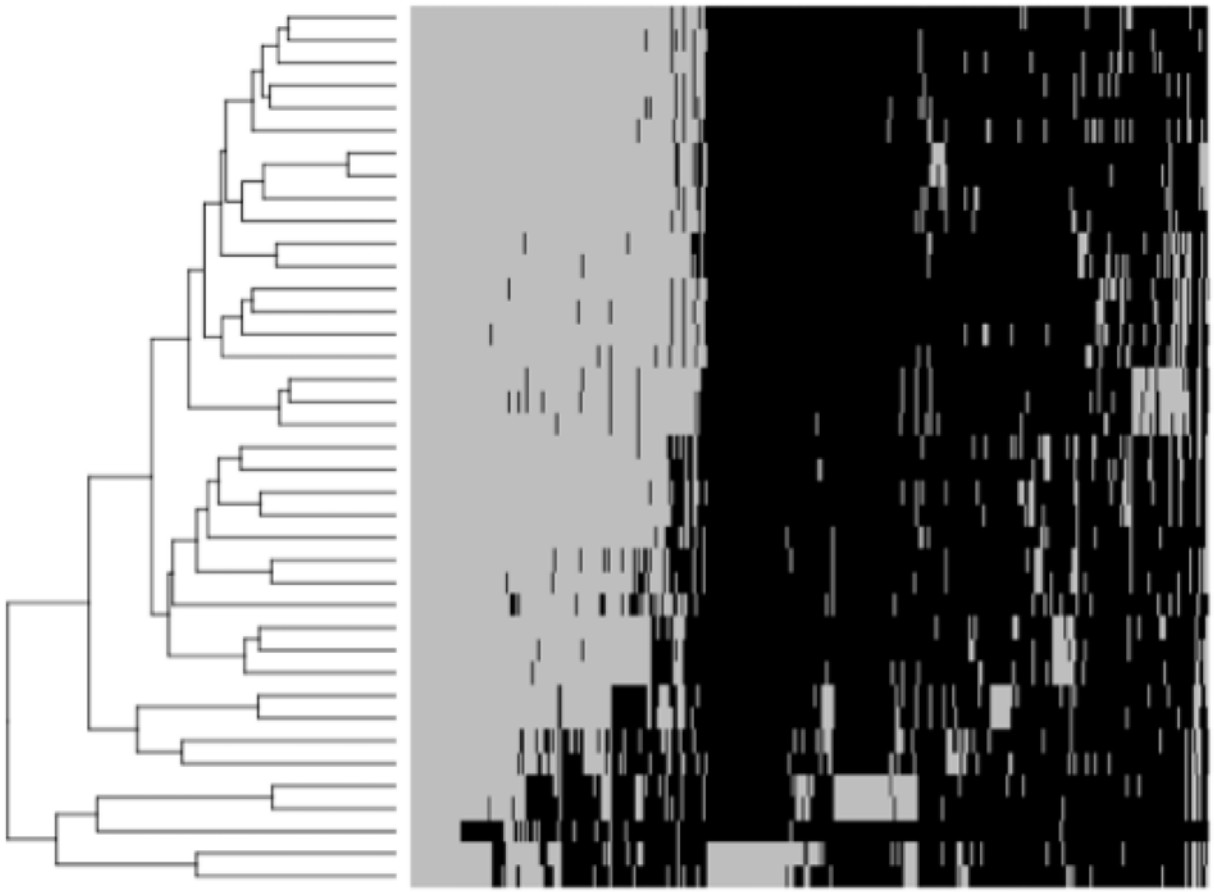


Figure S3. Pangenome hierarchical clustering based on the presence (grey) or absence (black) of orthologous genes.



Salinibacter pepae ESAV49^{Ts}
Salinibacter pepae ESSP73
Salinibacter pepae ESSP12
Salinibacter pepae ESCM71
Salinibacter pepae ESSP84
Salinibacter pepae ESAV87
Salinibacter pepae USCM88
Salinibacter pepae ESCM46
Salinibacter pepae ESCM187
Salinibacter pepae ESCM25
Salinibacter pepae DZ03
Salinibacter pepae DZ04
Salinibacter pepae CM04
Salinibacter pepae CZ17
Salinibacter pepae UZ07
Salinibacter pepae CZ16
Salinibacter pepae UM13
Salinibacter pepae CZ33
Salinibacter pepae CZ26
Salinibacter ruber SP273
Salinibacter ruber ST67
Salinibacter ruber M8
Salinibacter ruber M31^{T}
Salinibacter grassmerensis NZ140^{T}
Salinibacter pampae ARCG
Salinibacter pampae ARCCH^{Ts}
Salinibacter abyssi ROFF^{Ts}
Salinibacter altiplanensis AN4
Salinibacter altiplanensis AN15^{T}
Salinibacter altiplanensis LL19
Salinivenus lutea DGO^{T}
Salinivenus lutea CB7^{T}
Longimonas halophila SYD6^{T}
Salisaeta longa S44^{T}
Rhodothermus marinus DSM4252^{T}
Rhodothermus profundi DSM22212^{T}
Balneola vulgaris 13IXA01164^{T}
Rubrivirga marina SAORIC28^{T}
Rubricoccus marinus SG29^{T}

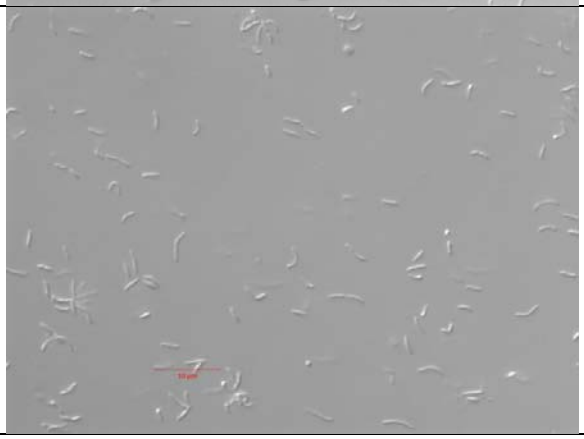
Table S2. Pangenome statistics between genomes belonging to the same species.

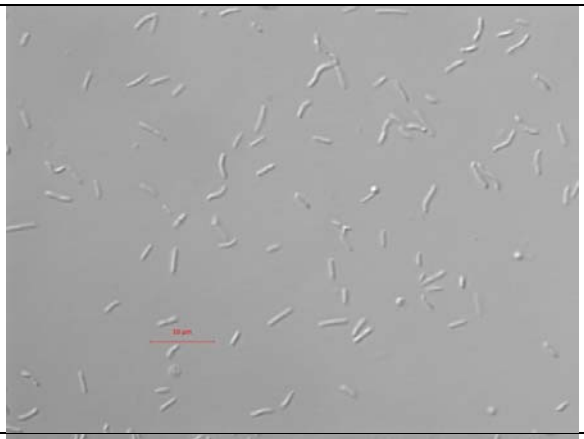
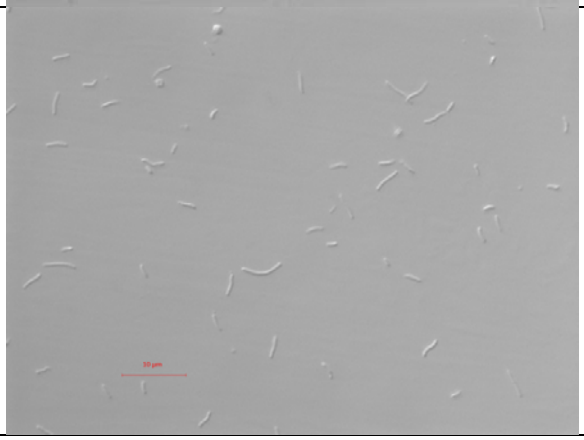
Species	Total genomes	Pan-genome	Core-genome	Specific genes	Pangenome Model γ
Sal. pepae	19	6,052	1,979	2,133	0.3
***Sal. ruber* subsampling 1**	19	7,101	1,903	2,743	0.36
***Sal. ruber* subsampling 2**	19	7,189	1,923	3,411	0.38
***Sal. ruber* subsampling 3**	19	7,687	1,967	3,145	0.37
***Sal. ruber* subsampling 4**	19	7,723	1,955	3,298	0.37
***Sal. ruber* subsampling 5**	19	7,390	1,973	3,176	0.36
***Sal. ruber* subsampling 6**	19	7,401	1,948	3,156	0.36
***Sal. ruber* subsampling 7**	19	7,943	1,967	3,744	0.4
***Sal. ruber* subsampling 8**	19	7,390	1,913	3,451	0.38
***Sal. ruber* subsampling 9**	19	7,549	1,943	3,573	0.39
***Sal. ruber* subsampling 10**	19	7,723	1,943	3,465	0.39
	MEAN	**7,509.60**	**1,943.50**	**3,316.20**	**0.38**
	Stdev	**262.90**	**23.84**	**279.24**	**0.014**

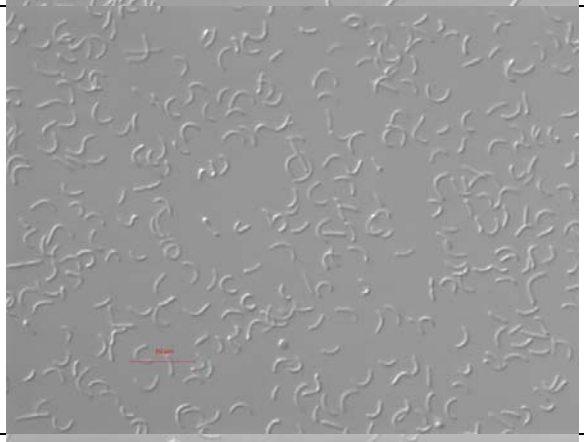
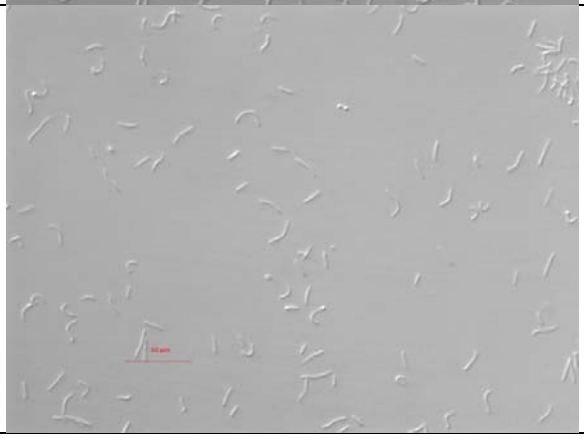
Table S3: CRISPR-Cas systems found in the *Salinibacter pepae* and *Salinibacter grassmerensis* genomes.

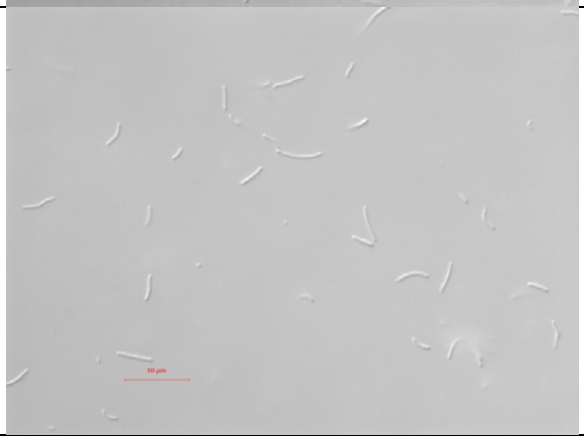
	Strain	Cas type	Gene	Cas genes
Sal. pepae	ESAV87	CAS Type I-E	6	Cas3_I, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas2_IE
	ESAV49^{Ts}	CAS-Type I-E	8	Cas2_IE, Cas1_IE, Cas6_IE, Cas5_IE, Cas7_IE, Cse2_IE, Cse1_IE, Cas3_I
	ESCM71	CAS-Type I-E	6	Cas3_I, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas2_IE
	ESSP12	CAS-Type I-E	8	Cas2_IE, Cas1_IE, Cas6_IE, Cas5_IE, Cas7_IE, Cse2_IE, Cse1_IE, Cas3_I
	ESSP73	CAS-Type I-E	8	Cas2_IE, Cas1_IE, Cas6_IE, Cas5_IE, Cas7_IE, Cse2_IE, Cse1_IE, Cas3_I
	ESSP87	CAS-Type I-E	8	Cas2_IE, Cas1_IE, Cas6_IE, Cas5_IE, Cas7_IE, Cse2_IE, Cse1_IE, Cas3_I
	USCM25	CAS-Type I-E	8	Cas3_I, Cse1_IE, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas1_IE, Cas2_IE
	USCM46	CAS-Type I-E	8	Cas3_I, Cse1_IE, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas1_IE, Cas2_IE
	USCM88	CAS-Type I-E	7	Cas1_IE, Cas2_IE, Cas5_IE, Cas6_IE, Cas7_IE, Cse1_IE, Cse2_IE
	USCM187	CAS-Type I-E	8	Cas2_IE, Cas1_IE, Cas6_IE, Cas5_IE, Cas7_IE, Cse2_IE, Cse1_IE, Cas3_I
		CAS-Type III-B	9	Cas2_I-II-III-V, Cas1_I-II-III-V, Csm3_1_IIIAD, Cas10_IIIB, Cmr3_1_IIIB, Cmr4_IIIB, Cmr5_IIIB, Cmr6_IIIB, Csm3_1_IIIAD
	CM04	CAS-Type I-E	8	Cas2_IE, Cas1_IE, Cas6_IE, Cas5_IE, Cas7_IE, Cse2_IE, Cse1_IE, Cas3_I
	CZ16	CAS-Type I-E	6	Cas3_I, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas2_IE
	CZ17	CAS-Type I-E	8	Cas2_IE, Cas1_IE, Cas6_IE, Cas5_IE, Cas7_IE, Cse2_IE, Cse1_IE, Cas3_I
	CZ26	CAS-Type I-E	6	Cas3_I, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas2_IE
	CZ33	CAS-Type I-E	6	Cas3_I, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas2_IE
	DZ03	CAS-Type I-E	6	Cas3_I, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas2_IE
	DZ04	CAS-Type I-E	6	Cas3_I, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas2_IE
	UM13	CAS-Type I-E	6	Cas3_I, Cse2_IE, Cas7_IE, Cas5_IE, Cas6_IE, Cas2_IE
	UZ07	CAS-Type I-E	8	Cas2_IE, Cas1_IE, Cas6_IE, Cas5_IE, Cas7_IE, Cse2_IE, Cse1_IE, Cas3_I
Sal. grassmerensis	NZ140^{T}	CAS-Type I-C	5	Cas3_I, Cas7c2_IC, cas8c2_IC, Cas5c2_IC, Cas6_I-III
		CAS-Type I-E	6	Cas3_I, Cas3_I, Cas5_IE, Cas6_IE, Cas7_IE, Cse2_IE

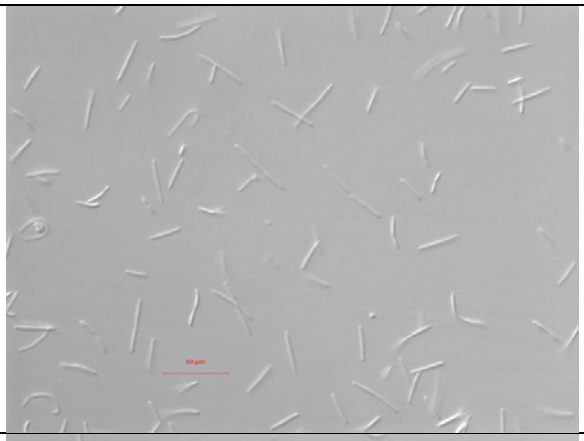
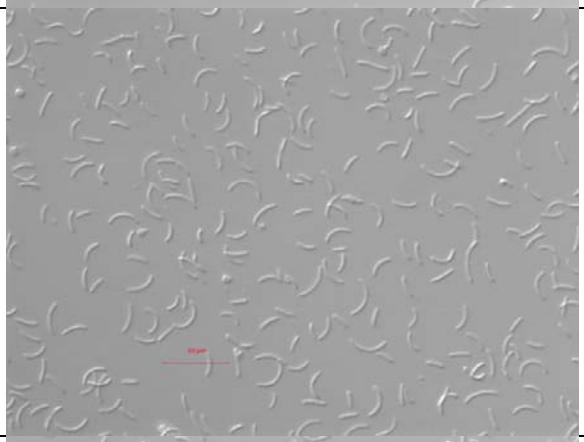
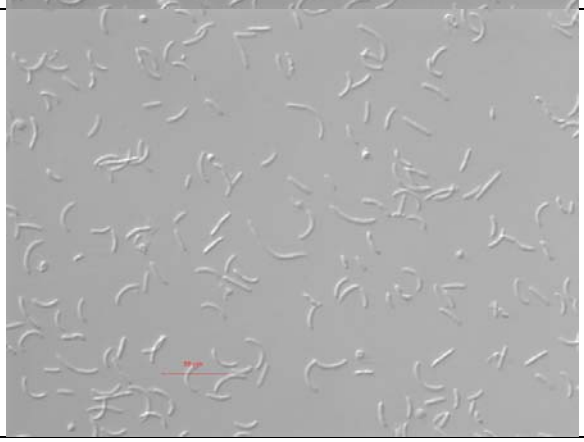
Table S4. Cell morphologies of the new isolates. All cells have been cultivated onto MA agar. for 10 days at 12ºC. The morphology was observed under an optical Microsope (Zeiss Axio Imager 100X). Photomicrograph showing cells of strain observed by oil-immersion differential interference contrast (DIC) microscopy (Nomarski).

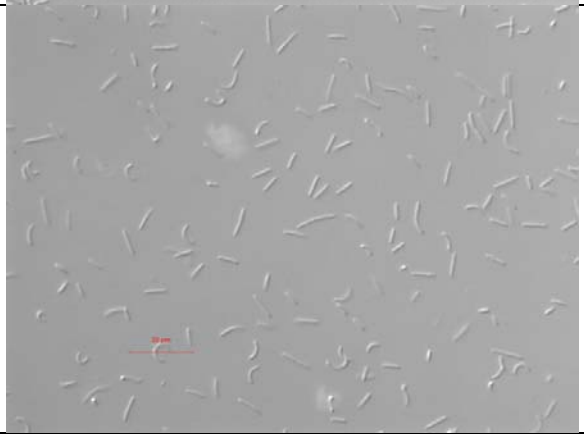
Sal. pepae – ESAV49^{Ts}		
Sal. pepae – ESAV87		
Sal. pepae – ESCM71		

Sal. pepae – ESSP12		
Sal. pepae – ESSP13		
Sal. pepae – ESSP87		
Sal. pepae – CM04		

Sal. pepae – CZ16		
Sal. pepae – CZ17		
Sal. pepae – CZ26		
Sal. pepae – CZ33		

Sal. pepae – DZ03		
Sal. pepae – DZ04		
Sal. pepae – UM13		
Sal. pepae – UZ07		

Sal. pepae – USCM187		
Sal. pepae – USCM25		
Sal. pepae – USCM46		
Sal. pepae – USCM88		

Sal. altiplanensis AN15^{T}		
Sal. altiplanensis AN4		
Sal. altiplanensis LL19		
Sal. ruber M31^{T}		

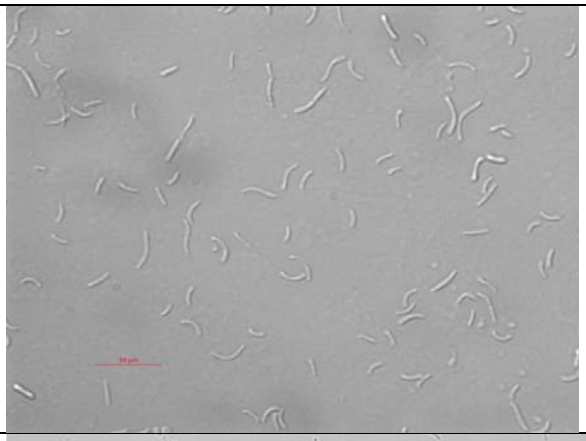
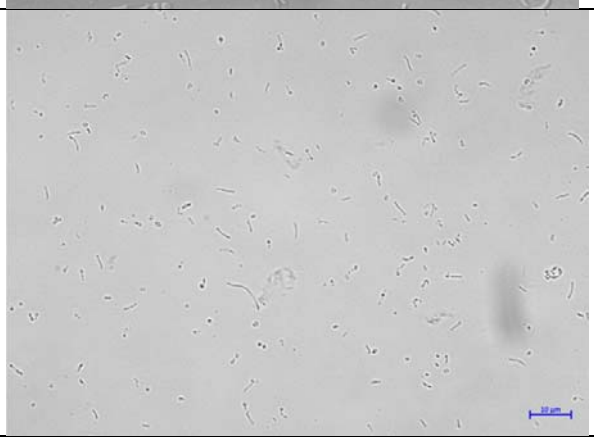
Sal. ruber M8		
Sal. grassmerensis – NZ140^{T}		

Figure S4. Hierarchical clustering of all cellular and supernatant metabolomes analyzed with electrospray-negative mode ICR-FT/MS (upper panel). Loading plots in where single molecules are colored based on the specific classes that they belong to: Grey: Core Metabolome. Yellow: discriminative for *Sal. ruber.* Red: discriminative for *Sal. pepae*. Orange: discriminative for *Sal. altiplanensis* (middle panel). Core metabolomes of the genus (lower panel). (A) water-soluble (left figures) and (B) water-insoluble (but methanol soluble; right figures) fractions.

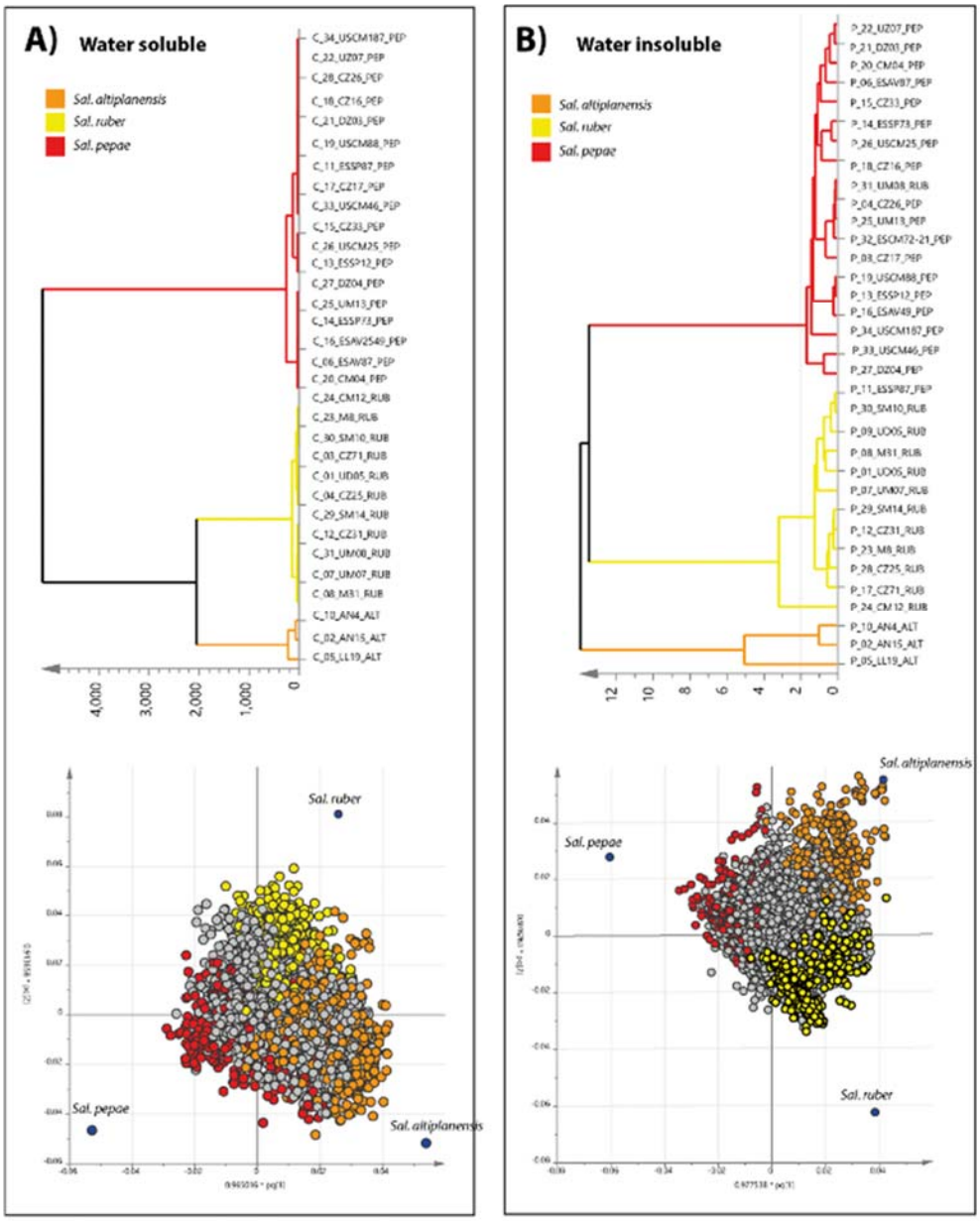


Figure S5. Count of saturated, mono-, di-, tri-, tetra-, and more unsaturated fatty acids classified as core metabolites in water-soluble (orange color) and water-insoluble (blue color) fractions versus their degree of unsaturation. Insert Venn diagram show count of unique and common fatty acids in each case.

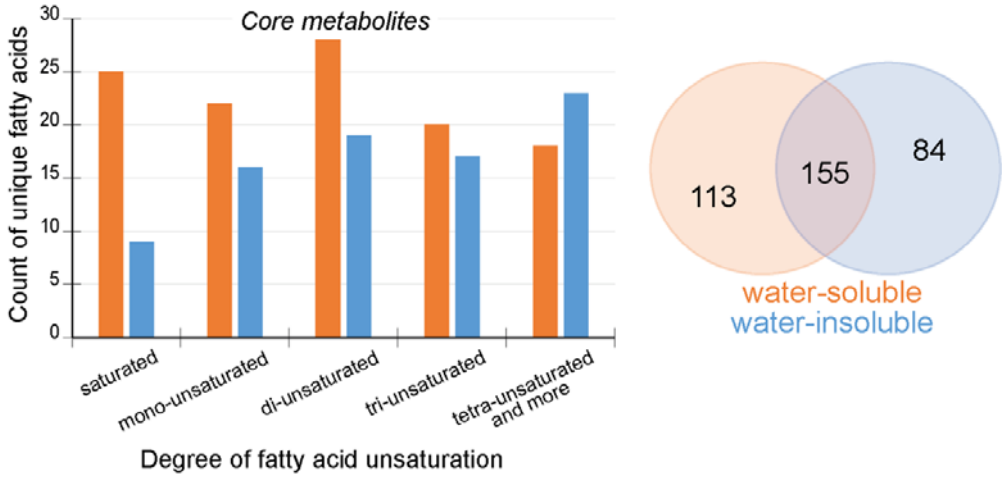


Figure S6. Count of discriminating fatty acids classified as saturated, mono-, di-, tri-, tetra-, and more unsaturated compounds in Sal. pepae, Sal. ruber and Sal. altiplanensis according to their degree of unsaturation in the water-soluble fractions or in the water-insoluble fractions.

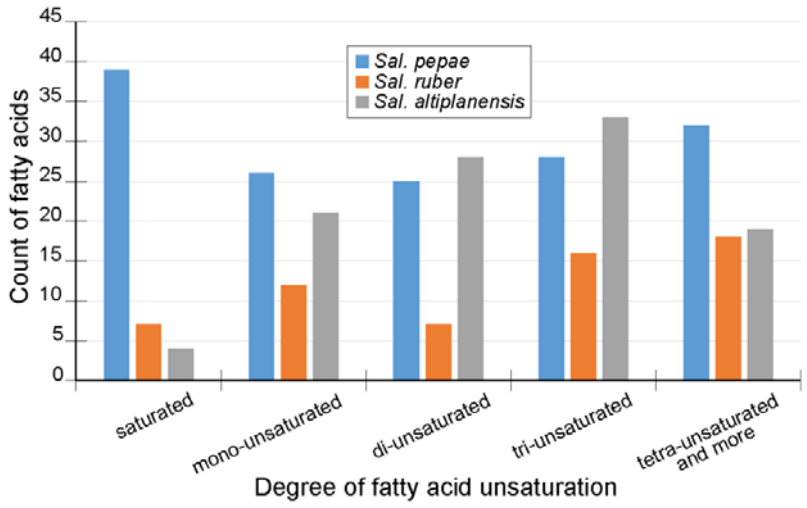
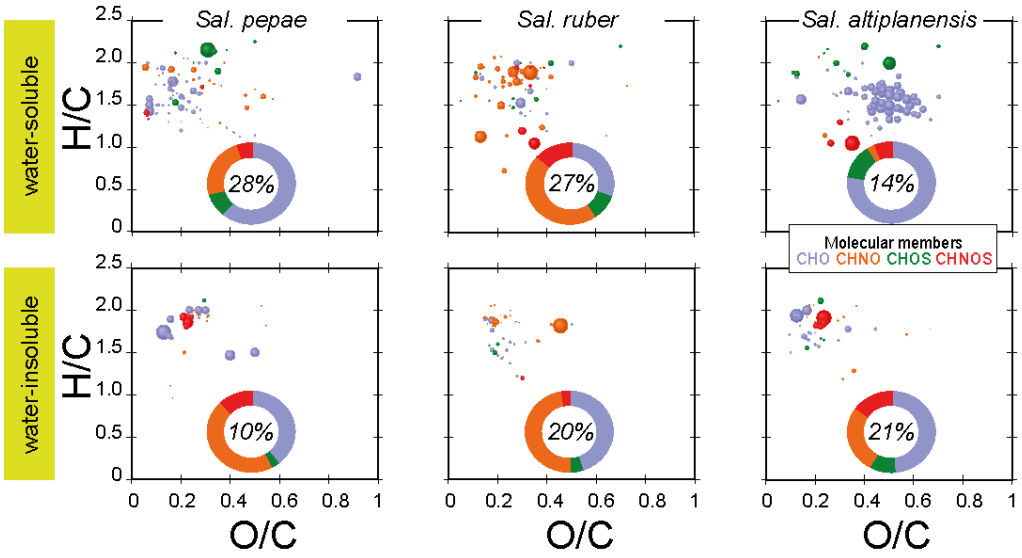


Figure S7. Van Krevelen plots showing shared compounds assigned in *Sal. grassmerensis* NZ140^T with the discriminating molecular compositions of *Sal. ruber*, *Sal. pepae* and *Sal. altiplanensis* when comparing sample pairs (see also Figure 5B, 5C and 5D). Here we looked at the molecular compositions assigned in the case of *Sal. grassmerensis* NZ140^T and checked whether they are present in the discriminating molecular compositions of *Sal. ruber*, *Sal. pepae* and *Sal. altiplanensis*. Insert histograms represent the molecular series based on CHO (blue), CHOS (green), CHNO (orange), and CHNOS (red) atom combinations. Insert percentages represent the numbers (in %) of shared molecular compositions of *Sal. grassmerensis* NZ140^T with respect to the total discriminant compounds of *Sal. ruber*, *Sal. pepae* as well as *Sal. altiplanensis* (see also Figure 5B, 5C and 5D).



Text S2: From the Laguna Colorada Chica, we detected 38 contigs encoding for a 16S rRNA gene and 3 affiliated with the *Salinibacteraceae* family (Sup. Figure S6). We identified one almost complete 16S rRNA sequence (1,555 bp) in a contig with a sequencing depth of 77.3X, and 2 partial sequences (< 945 bp) showing a sequencing depth < 12.7X. The largest 16S rRNA gene sequence with the highest coverage affiliated with the *Salinibacter* genus with an identity of 97.3% with *Sal. pepae* ESAV49^{Ts}, 96.9% with *Sal. ruber* M31^{T}, 96.3% with *Sal. altiplanensis* AN15^{T}, 96% with *Sal. grassmerensis* NZ140^{T} and 96.2% with *Sal. abyssi* ROFF^{Ts}, respectively (Sup. Figure S6 and Sup. Spreadsheet S2). The shorter assembled 16S rRNA sequences showed a percentage of similarity with any of the *Salinibacter* sequences < 95.4% (Sup. Figure S6). ARCCH^{Ts} represented the most abundant *Salinibacter* population in the metagenome of origin (with 1.9% relative abundance and 56X coverage), we confidently assigned the contig encoding the largest *Salinibacter* 16S rRNA gene to ARCCH^{Ts}. The sequence was deposited under the accession number GCA_947077715^{Ts}. Supporting the assignation, from the Laguna Colorada Grande we detected 28 contigs encoding a 16S rRNA gene, 3 affiliated with the *Salinibacteraceae* family. The single complete 16S rRNA gene sequence (accession number GCA_947077705) showed 100% identity with the one assigned to ARCCH^{Ts} (Sup. Figure S3). In agreement, ARCG coverage 12X and the 16S rRNA encoding contig 17X. The agreement between the coverages and both affiliations based on 16S rRNA gene sequence reconstruction (Figure 1A) and the core-genome reconstruction (Figure 1B) confidently assigned the 16S rRNA genes to their respective MAGs in the samples of origin.

Figure S8. Phylogenetic reconstruction based on the 16S rRNA gene sequence analysis of all *Salinibacter* species available in the LTP_01_2022, the MAGs recovered from metagenomes and the *Salinibacteraceae* 16S rRNA sequences recovered from assembled metagenomes from Colorada Chica (CCH) and Colorada Grande (CG), both located in Argentina. The tree was reconstructed using the maximum likelihood algorithm and is the result of the consensus of different approaches using distinct filters and datasets. The multifurcations indicate a branching order that could not be resolved. Bar indicates 10% sequence divergence. In brackets the accession number of each sequence is given.

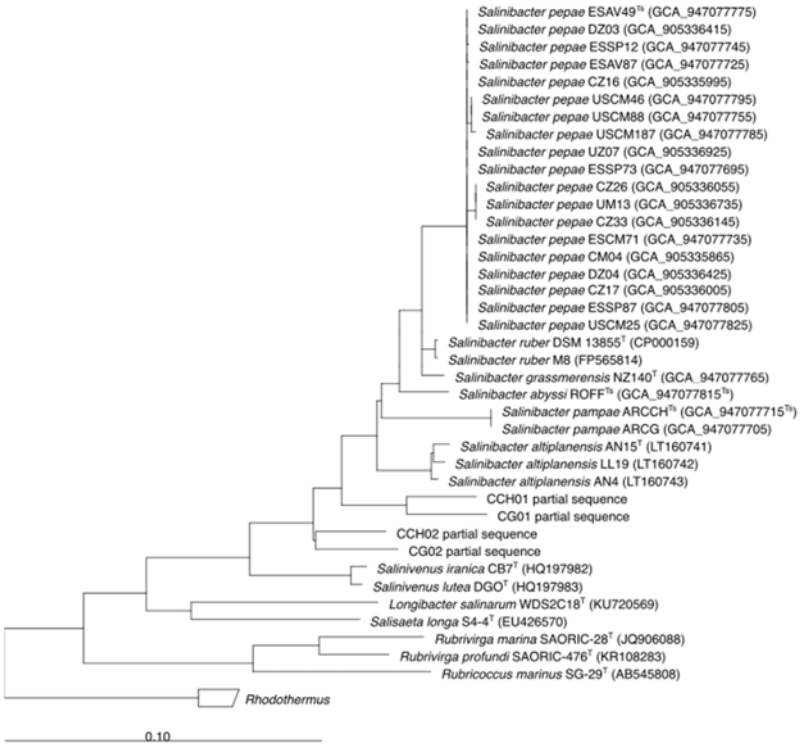


Table S5. MAGs metabolic reconstruction based on KEGG database annotations.

KEGG Mapper Reconstruction	***Sal. abyssi* ROFF^{Ts}**	***Sal. pampae* ARCCH^{Ts}**	***Sal. pampae* ARCG**
Oxidative phosphorylation			
- NADH dehydrogenase B/A	Complete	Complete	Complete
- Succinate dehydrogenase (*sdhC*, *sdhD*, *sdhA* and sdhB)	Complete	Complete	Complete
- Cytochrome C oxidase	Complete	Complete	Partial (lack *coxD* and *coxC*)
- F-type ATPase	Complete	Complete	Complete
Amino acid metabolism			
- Alanine	Complete	Complete	Complete
- Glycine	Complete	Complete	Complete
- Cysteine	Complete	Complete	Complete
- Valine	Complete	Complete	Complete
- Leucine	Complete	Complete	Complete
- Isoleucine	Complete	Complete	Complete
- Lysine	Complete	Complete	Complete
- Arginine	Complete	Complete	Complete
- Histidine	Complete	Complete	Complete
- Tyrosine	Complete	Complete	Complete
- Phenylalanine	Complete	Complete	Complete
- Tryptophan	Complete	Complete	Complete
- Phenylalanine	Complete	Complete	Complete
- Tyrosine	Complete	Complete	Complete
Carbohydrate metabolism	Complete	Complete	Complete
- Glycolysis / Gluconeogenesis	Complete	Complete	Complete
- Citrate cycle (TCA cycle)	Complete	Complete	Complete
- Pentose phosphate pathway	Complete	Complete	Complete
- Starch and sucrose metabolism	Complete	Complete	Complete
- Pyruvate metabolism	Complete	Complete	Complete
Number of ribosomal proteins	50	51	49
Number genes involved in Aminoacyl-tRNA biosynthesis	23	24	24
Flagellar assembly			
- Basalbody / Hook	FliG, FliI, FliJ, FlhA, FlgA, FlgD, FlgE, FlgF, FlgG, FlgH, FlgI FlgK and FlgL	FlgF and FlgG	FliM, FliN, FliR, FlhA, FlhB, FlgA, FlgF, FlgG, FlgH, FlgI
- Filament	FliD		
- Stator	MotB	MotB	MotB