

# Genome sequence of the moderately halophilic bacterium *Salinicoccus carniancri* type strain Crm<sup>T</sup> (= DSM 23852<sup>T</sup>)

Dong-Wook Hyun<sup>1</sup>, Tae Woong Whon<sup>1</sup>, Yong-Joon Cho<sup>2</sup>, Jongsik Chun<sup>2</sup>, Min-Soo Kim<sup>1</sup>, Mi-Ja Jung<sup>1</sup>, Na-Ri Shin<sup>1</sup>, Joon-Yong Kim<sup>1</sup>, Pil Soo Kim<sup>1</sup>, Ji-Hyun Yun<sup>1</sup>, Jina Lee<sup>1</sup>, Sei Joon Oh<sup>1</sup>, and Jin-Woo Bae<sup>1\*</sup>

<sup>1</sup>Department of Life and Nanopharmaceutical Sciences and Department of Biology, Kyung Hee University, Republic of Korea

<sup>2</sup>ChunLab, Inc., Seoul National University, Seoul, Republic of Korea

\*Corresponding author: Jin-Woo Bae ([baejw@khu.ac.kr](mailto:baejw@khu.ac.kr))

Keywords: moderately halophilic, *Salinicoccus carniancri*, *Staphylococcaceae*

*Salinicoccus carniancri* Jung *et al.* 2010 belongs to the genus *Salinicoccus* in the family *Staphylococcaceae*. Members of the *Salinicoccus* are moderately halophilic and originate from various salty environments. The halophilic features of the *Salinicoccus* suggest their possible uses in biotechnological applications, such as biodegradation and fermented food production. However, the genus *Salinicoccus* is poorly characterized at the genome level, despite its potential importance. This study presents the draft genome sequence of *S. carniancri* strain Crm<sup>T</sup> and its annotation. The 2,673,309 base pair genome contained 2,700 protein-coding genes and 78 RNA genes with an average G+C content of 47.93 mol%. It was notable that the strain carried 72 predicted genes associated with osmoregulation, which suggests the presence of beneficial functions that facilitate growth in high-salt environments.

## Introduction

The genus *Salinicoccus* in the family *Staphylococcaceae* was first proposed by Ventosa *et al.* (1990) and is defined as moderately halophilic, aerobic, Gram-positive, non-motile, non-sporulating, and heterotrophic cocci [1]. The genus name is derived from the Latin adjective *salinus*, saline, and the Greek masculine noun *kokkos*, meaning a grain or berry, i.e., saline coccus [2]. Most species in the genus *Salinicoccus* have been found in salty environments, such as fermented foods [3-5], solar salterns [1,6], salt mines [7,8], a salt lake [9], and saline soils [10,11]. All type strains of *Salinicoccus* species were characterized as halotolerant organisms, where NaCl concentrations of 2–20% (wt/vol) were suitable for growth [12-14].

These moderately halophilic bacteria can survive in salt-rich environments and grow optimally at 5–20% (wt/vol) NaCl [15]. These bacteria can utilize compatible solutes or osmolytes, such as carbohydrates, amino acid, polyols, betaines, and ectoines, by regulating their osmotic concentrations in high-salt content environmental

conditions [16,17]. Therefore, these organisms may have biotechnological importance with possible applications in food biotechnology for the production of fermented food [18], in environmental biotechnology for the biodegradation of organic pollutants and the production of alternative energy [19].

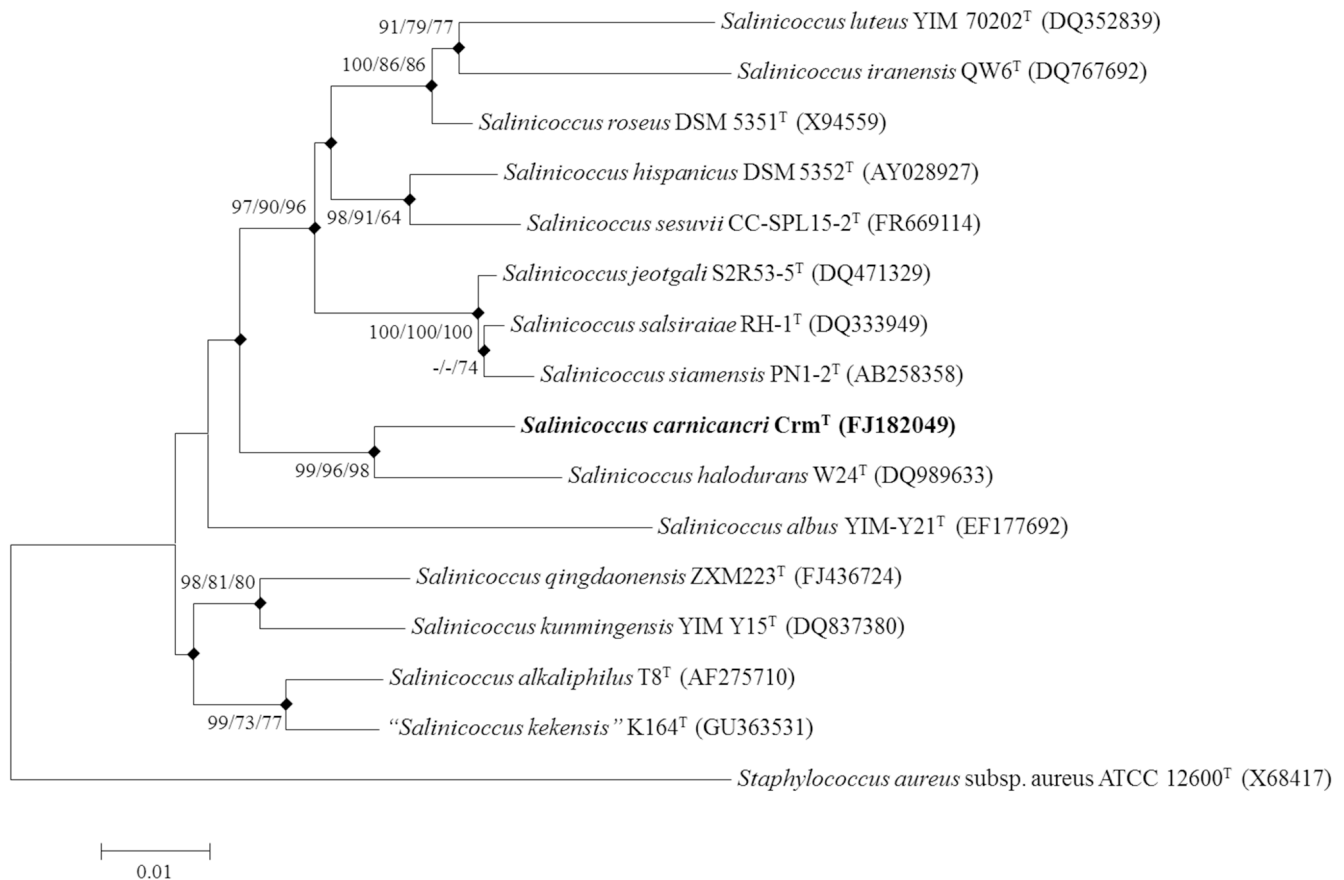
Strain Crm<sup>T</sup> (= DSM 23852 = JCM 15796 = KCTC 13301) is the type strain of the species *Salinicoccus carniancri*. This strain was isolated from a traditional Korean fermented seafood, known as ‘ganjang-gejang,’ which is made from raw crabs preserved in soy sauce [20]. The species name was derived from the Latin nouns *caro carnis*, flesh, and *cancer -cri*, a crab, i.e., the flesh of a crab [2]. The strain can grow in 0–20% (wt/vol) NaCl with optimal growth at 12% (wt/vol) NaCl [20]. The present study summarizes the features of *S. carniancri* strain Crm<sup>T</sup> and provides an analysis of its draft genome sequence, which is the first reported genome sequence of a species in the genus *Salinicoccus*.

### Classification and features

A taxonomic analysis was conducted based on the 16S rRNA gene sequence. The representative 16S rRNA gene sequence of strain *S. carniancri* CrmT was compared with the most recent release of the EzTaxon-e database [21]. The multiple sequence alignment program CLUSTAL W [22] was used to generate alignments with other gene sequences collected from databases. The alignments were trimmed and converted to the MEGA format before phylogenetic analysis. Phylogenetic consensus trees were constructed based on the aligned gene sequences using the neighbor-joining [23], maximum-parsimony [24], and maximum-likelihood [25] methods with 1,000 randomly selected bootstrap replicates using MEGA version 5

[26]. The phylogenetic analysis based on the 16S rRNA gene sequence showed that strain CrmT was most closely related to *Salinicoccus halodurans* W24T with 96.99% similarity. The phylogenetic consensus tree based on the 16S rRNA gene sequences indicated that strain CrmT was clustered within a branch containing other species in the genus *Salinicoccus* (Figure 1).

Strain CrmT (Table 1) was isolated from the fermented seafood ganjang-gejang during a project that investigated microbial communities in fermented foods, i.e., the Next-Generation BioGreen 21 Program (No. PJ008208) in Korea. Ganjang-gejang, with a NaCl (w/v) concentration of 24.5%, was produced by preserving scabbard crabs in soy sauce, garlic, and onions at -5°C for 4–5 days.



**Figure 1.** Phylogenetic consensus tree based on 16S rRNA gene sequences showing the relationship between *Salinicoccus carniancri* strain CrmT and the type strains of other species in the genus *Salinicoccus*. The type strain of *Staphylococcus aureus* was used as an outgroup. The GenBank accession numbers for the 16S rRNA genes of each strain are shown in parentheses. Filled diamonds indicate identical branches present in the phylogenetic consensus trees constructed using the neighbor-joining (NJ), maximum-parsimony (MP), and maximum-likelihood (ML) algorithms. The numbers at the nodes represent the bootstrap values as percentages of 1,000 replicates and values <70% are not shown at the branch points. The scale bar represents 0.01 nucleotide change per nucleotide position.

*S. carniancricri* strain Crm<sup>T</sup> is a Gram-positive, moderately halophilic, non-motile, non-sporulating, and aerobic heterotrophic coccus with a diameter of 1.0–2.5  $\mu\text{m}$  [20]. Figure 2 shows the morphological features of strain Crm<sup>T</sup>, which were obtained by scanning electron microscopy (SEM). Colonies were ivory-colored [20]. Growth occurred at 4–45°C, with an optimum of 30–37°C, and at pH values of 6.0–11.0, with an optimum of 7.0–8.0. The salinity range suitable for growth was 0–20% (w/v) NaCl, with an optimum of 12% (w/v) NaCl [20]. Strain Crm<sup>T</sup> contains menaquinone MK-6 as the predominant respiratory quinone [20]. The major fatty acids (>10% of total fatty acid) are anteiso-C<sub>15:0</sub> (40.61%), iso-C<sub>15:0</sub> (22.0%), and anteiso-C<sub>17:0</sub> (12.12%) [20]. The major cellular polar lipids are phosphatidylglycerol and diphosphatidylglycerol [20]. Glycine and lysine are the major amino acid constituents of the cell-wall hydrolysate [20].

## Genome sequencing and annotation

### Genome project history

*S. carniancricri* strain Crm<sup>T</sup> was selected for sequencing because of its environmental potential as part of the Next-Generation BioGreen 21 Program (No.PJ008208). The genome project is deposited in the Genomes OnLine Database [40] and the genome sequence is deposited in GenBank. Sequencing and annotation were performed by ChunLab Inc., South Korea. A summary of the project information is shown in Table 2.

### Growth conditions and DNA isolation

*S. carniancricri* strain Crm<sup>T</sup> was grown aerobically in marine 2216 (Marine medium, BBL), supplemented with 10% (w/v) NaCl at 30°C. Genomic DNA was extracted using a Wizard Genomic DNA

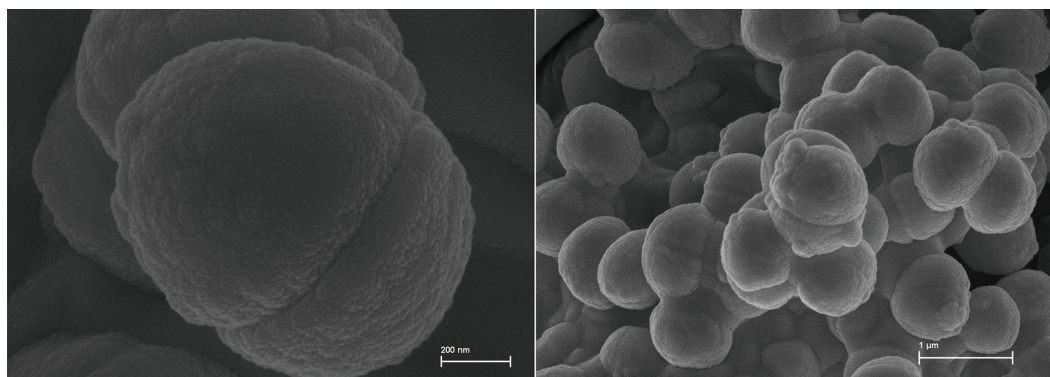
Purification Kit (Promega A1120), according to the manufacturer's instructions.

### Genome sequencing and assembly

The genome of *S. carniancricri* Crm<sup>T</sup> was sequenced using a combination of a 454 Genome Sequencer FLX Titanium system (Roche Diagnostics) with an 8 kb paired end library, an Illumina Hiseq system with a 150 base pair (bp) paired end library, and a PacBio RS system (Pacific Biosciences). A total of 7,434,400 sequencing reads (443.6-fold genome coverage) were obtained using the Roche 454 system (187,030 reads; 12.1-fold coverage), Illumina Hiseq system (7,219,019 reads; 408.4-fold coverage), and PacBio RS system (28,351 reads; 23.1-fold coverage) combined. The Roche 454 pyrosequencing and Illumina sequencing reads were assembled using Roche gsAssembler 2.6 (Roche Diagnostics) and CLCbio CLC Genomics Workbench 5.0 (CLCbio), respectively. Table 2 shows the project information and its associated MGS version 2.0 compliance levels [27].

### Genome annotation

The open reading frames (ORFs) of the assembled genome were predicted using a combination of the Rapid Annotation using Subsystem Technology (RAST) pipeline [41] and the GLIMMER 3.02 modeling software package [42]. Comparisons of the predicted ORFs using the SEED [43], NCBI COG [44], NCBI Refseq [45], CatFam [46], Ez-Taxon-e [21], and Pfam [47] databases were conducted during gene annotation. RNAmmer 1.2 [48] and tRNAscan-SE 1.23 [49] were used to find rRNA genes and tRNA genes, respectively. Additional gene prediction analyses and functional annotation were performed using the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [50].



**Figure 2.** Scanning electron microscopy images of *S. carniancricri* Crm<sup>T</sup> obtained using a SUPRA VP55 (Carl Zeiss) at an operating voltage of 15kV. The scale bars represent 200 nm (left) and 1  $\mu\text{m}$  (right), respectively.

**Table 1.** Classification and general features of *Salinicoccus carniancri* strain Crm<sup>T</sup> according to the MIGS recommendations [27].

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [28]
		Phylum <i>Firmicutes</i>	TAS [29-31]
		Class <i>Bacilli</i>	TAS [32,33]
	Current classification	Order <i>Bacillales</i>	TAS [34,35]
		Family <i>Staphylococcaceae</i>	TAS [36,37]
		Genus <i>Salinicoccus</i>	TAS [1,38]
		Species <i>Salinicoccus carniancri</i>	TAS [20]
		Type strain Crm <sup>T</sup>	TAS [20]
	Gram stain	Positive	TAS [20]
	Cell shape	Cocci	TAS [20]
	Motility	Non-motile	TAS [20]
	Sporulation	Non-sporulating	TAS [20]
	Temperature range	4–45°C	TAS [20]
	Optimum temperature	30–37°C	TAS [20]
	Salinity range	0–20% (w/v)	TAS [20]
	Optimum salinity	12% (w/v)	TAS [20]
	pH range	6–11	TAS [20]
	Optimum pH	7–8	TAS [20]
	Carbon source	Heterotroph	TAS [20]
	Energy source	Not reported	
MIGS-6	Habitat	Fermented seafood (marinated crab)	TAS [20]
MIGS-6.1	Temperature	–5 to 5°C	IDA
MIGS-6.3	Salinity	20%	IDA
MIGS-22	Oxygen	Aerobic	TAS [20]
MIGS-15	Biotic relationship	Free-living	TAS [20]
MIGS-14	Pathogenicity	Unknown	
	Biosafety level	1	
MIGS-23.1	Isolation	The traditional Korean fermented seafood ‘ganjang-gejang’ (Crabs preserved in soy sauce)	TAS [20]
MIGS-4	Geographic location	Republic of Korea	TAS [20]
MIGS-5	Sample collection time	August, 2010	NAS
MIGS-4.1	Latitude	Not reported	
MIGS-4.2	Longitude	Not reported	
MIGS-4.3	Depth	Not reported	
MIGS-4.4	Altitude	Not reported	

Evidence codes, as follows: IDA: inferred from direct assay; TAS: traceable author statement (i.e., a direct report exists in the literature); NAS: non-traceable author statement (i.e., not observed directly in a living, isolated sample, but based on a generally accepted property of the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [39].

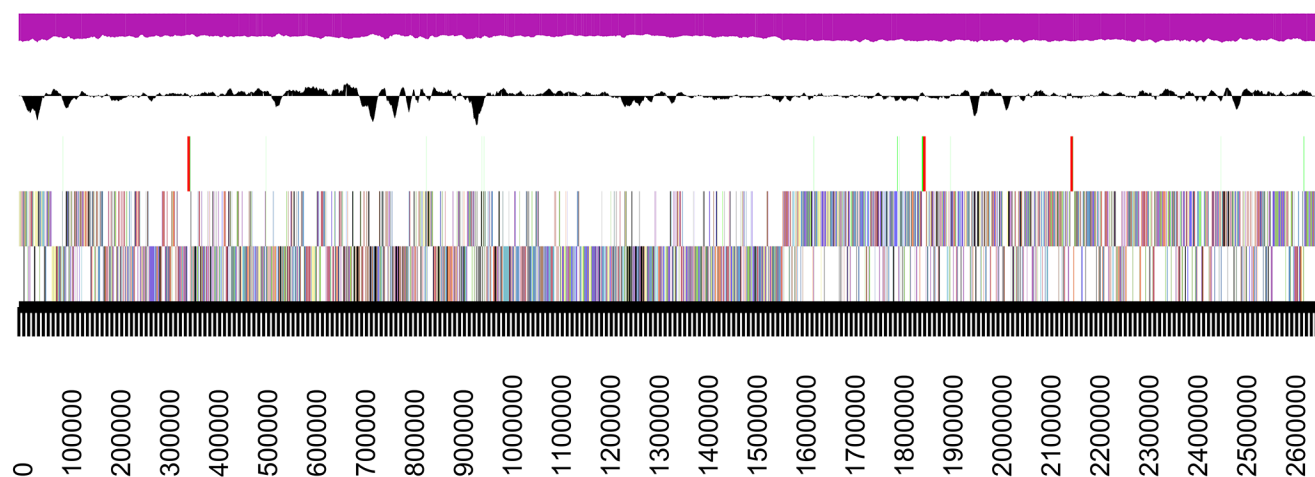
**Table 2.** Genome sequencing project information.

MIGS ID	Property	Term
MIGS-31	Finishing quality	Improved high-quality draft
MIGS-28	Libraries used	454 PE library (8 kb insert size) and Illumina library
MIGS-28.2	Number of reads	7,434,400 sequencing reads
MIGS-29	Sequencing platforms	454 GS FLX Titanium, Illumina Hiseq, and PacBio RS system
MIGS-31.2	Sequencing coverage	443.60-fold coverage (12.1 × 454 pyrosequencing, 408.4 × Illumina, and 23.1 × PacBio)
MIGS-30	Assemblers	gsAssembler 2.6, CLC Genomics Workbench 5.0
MIGS-32	Gene calling method	GLIMMER 3.02
	Genbank ID	ANAM01000000
	Genbank Date of Release	January 2, 2013
	GOLD ID	Gi21266
	NCBI project ID	175941
	Database: IMG-ER	2521172676
	Source material identifier	DSM 23852, JCM 15796, KCTC 13301
	Project relevance	Environmental and biotechnological

## Genome properties

The draft genome sequence of *S. carnicancri* Crm<sup>T</sup> was 2,673,309 bp, which comprised three scaffolds that included 12 contigs. The G+C content was 47.93 mol% (Figure 3 and Table 3). RAST and GLIMMER predicted 2,778 coding sequences (CDSs) in the genome. Of the predicted ORFs, 2,700 ORFs were assigned to protein-coding genes. A total of 2,298 genes (82.72%) were

assigned putative functions, whereas the remaining genes were annotated as hypothetical proteins. The genome contained 78 ORFs assigned to RNA genes, including 61 predicted tRNA genes, nine rRNA genes (three 5S rRNA, three 16S rRNA, and three 23S rRNA genes), and eight other RNA genes. The distributions of genes in the COG functional categories are presented in Table 4.



**Figure 3.** Graphical map of the largest scaffold, C792\_Scaffold00001.1, which represented >99.6% of the chromosome. The smaller scaffolds of the chromosome are not shown. From bottom to top: genes on the forward strand (colored according to COG categories), genes on the reverse strand (colored according to COG categories), RNA genes (tRNAs = green, rRNAs = red, and other RNAs = black), GC content, and GC skew.

**Table 3.** Genome statistics.

Attribute	Value	% of total <sup>a</sup>
Genome size (bp)	2,673,309	100.00%
DNA coding region (bp)	2,420,461	90.54%
DNA G+C content (bp)	1,279,282	47.93%
Total genes	2,778	100%
RNA genes	78	2.81%
rRNA operons	9	0.32%
Protein-coding genes	2,700	97.19%
Genes with predicted functions	2,298	82.72%
Genes in paralog clusters	1,850	66.59%
Genes assigned to COGs	2,255	81.17%
Genes assigned Pfam domains	2,333	83.98%
Genes with signal peptides	437	15.73%
Genes with transmembrane helices	679	24.44%
CRISPR repeats	1	

<sup>a</sup>The total is based on either the size of the genome (bp) or the total number of protein-coding genes in the annotated genome.

**Table 4.** Numbers of genes associated with the 25 general COG functional categories.

Code	Value	%age <sup>a</sup>	Description
J	152	5.6	Translation
A	0	0.0	RNA processing and modification
K	194	7.2	Transcription
L	120	4.4	Replication, recombination, and repair
B	1	0.0	Chromatin structure and dynamics
D	27	1.0	Cell cycle control, mitosis, and meiosis
Y	0	0.0	Nuclear structure
V	38	1.4	Defense mechanisms
T	75	2.8	Signal transduction mechanisms
M	126	4.7	Cell-wall/membrane biogenesis
N	6	0.2	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	35	1.3	Intracellular trafficking and secretion
O	71	2.6	Posttranslational modification, protein turnover, and chaperones
C	148	5.5	Energy production and conversion
G	189	7.0	Carbohydrate transport and metabolism
E	241	8.9	Amino acid transport and metabolism
F	77	2.9	Nucleotide transport and metabolism
H	120	4.4	Coenzyme transport and metabolism
I	83	3.1	Lipid transport and metabolism
P	152	5.6	Inorganic ion transport and metabolism
Q	53	2.0	Secondary metabolites biosynthesis, transport, and catabolism
R	346	12.8	General function prediction only
S	215	8.0	Function unknown
-	231	8.6	Not in COGs

<sup>a</sup>The total is based on the total number of protein-coding genes in the annotated genome.

## Insights from the genome sequence

*S. carnicancri* Crm<sup>T</sup> encoded 72 predicted genes associated with the biosynthesis of compatible solutes and the transport of osmolytes, such as choline-glycine betaine transporter (BetT) and periplasmic glycine betaine/choline-binding lipoprotein of an ABC-type transport system (OpuBC). Potentially, these genes are key factors that allow

*S. carnicancri* to adapt to high-salt environments (e.g., salt-fermented food) by regulating the osmotic concentration. Further studies are required to elucidate the osmoregulation mechanism, which could facilitate biotechnological applications of this halophilic bacterium.

## Acknowledgements

We gratefully acknowledge the help of Dr. Seong Woon Roh and Mr. Hae-Won Lee during SEM analysis (Jeju Center, Korea Basic Science Institute, Korea). This work

was supported by a grant from the Next-Generation BioGreen 21 Program (No.PJ008208), Rural Development Administration, Republic of Korea.

## References

- Ventosa AM, Ruizberraquero MC, Kocur F. M. *Salinicoccus roseus* gen. nov, sp. nov, a new moderately halophilic gram-positive coccus. *Syst Appl Microbiol* 1990; **13**:29-33. [http://dx.doi.org/10.1016/S0723-2020\(11\)80177-3](http://dx.doi.org/10.1016/S0723-2020(11)80177-3)
- Euzeby JP. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* 1997; **47**:590-592. [PubMed http://dx.doi.org/10.1099/00207713-47-2-590](http://dx.doi.org/10.1099/00207713-47-2-590)
- França L, Rainey FA, Nobre MF, da Costa MS. *Salinicoccus salsiraiiae* sp. nov.: a new moderately halophilic gram-positive bacterium isolated from salted skate. *Extremophiles* 2006; **10**:531-536. [PubMed http://dx.doi.org/10.1007/s00792-006-0532-1](http://dx.doi.org/10.1007/s00792-006-0532-1)
- Aslam Z, Lim JH, Im WT, Yasir M, Chung YR, Lee ST. *Salinicoccus jeotgali* sp. nov., isolated from jeotgal, a traditional Korean fermented seafood. *Int J Syst Evol Microbiol* 2007; **57**:633-638. [PubMed http://dx.doi.org/10.1099/ijs.0.64586-0](http://dx.doi.org/10.1099/ijs.0.64586-0)
- Pakdeeto A, Tanasupawat S, Thawai C, Moonmangmee S, Kudo T, Itoh T. *Salinicoccus siamensis* sp. nov., isolated from fermented shrimp paste in Thailand. *Int J Syst Evol Microbiol* 2007; **57**:2004-2008. [PubMed http://dx.doi.org/10.1099/ijs.0.64876-0](http://dx.doi.org/10.1099/ijs.0.64876-0)
- Ventosa A, Marquez MC, Weiss N, Tindall BJ. Transfer of *Marinococcus hispanicus* to the genus *Salinicoccus* as *Salinicoccus hispanicus* comb. Nov. *Syst Appl Microbiol* 1992; **15**:530-534. [http://dx.doi.org/10.1016/S0723-2020\(11\)80112-8](http://dx.doi.org/10.1016/S0723-2020(11)80112-8)
- Chen YG, Cui XL, Pukall R, Li HM, Yang YL, Xu LH, Wen ML, Peng Q, Jiang CL. *Salinicoccus kunmingensis* sp. nov., a moderately halophilic bacterium isolated from a salt mine in Yunnan, south-west China. *Int J Syst Evol Microbiol* 2007; **57**:2327-2332. [PubMed http://dx.doi.org/10.1099/ijs.0.64783-0](http://dx.doi.org/10.1099/ijs.0.64783-0)
- Chen YG, Cui XL, Wang YX, Zhang YQ, Li QY, Liu ZX, Wen ML, Peng Q, Li WJ. *Salinicoccus albus* sp. nov., a halophilic bacterium from a salt mine. *Int J Syst Evol Microbiol* 2009; **59**:874-879. [PubMed http://dx.doi.org/10.1099/ijs.0.003251-0](http://dx.doi.org/10.1099/ijs.0.003251-0)
- Gao M, Wang L, Chen SF, Zhou YG, Liu HC. *Salinicoccus kekensis* sp. nov., a novel alkaliphile and moderate halophile isolated from Keke Salt Lake in Qinghai, China. *Anton Leeuw Int J G* 2010; **98**:351-357. [PubMed http://dx.doi.org/10.1007/s10482-010-9449-x](http://dx.doi.org/10.1007/s10482-010-9449-x)
- Wang X, Xue Y, Yuan S, Zhou C, Ma Y. *Salinicoccus halodurans* sp. nov., a moderate halophile from saline soil in China. *Int J Syst Evol Microbiol* 2008; **58**:1537-1541. [PubMed http://dx.doi.org/10.1099/ijs.0.65467-0](http://dx.doi.org/10.1099/ijs.0.65467-0)
- Chen YG, Cui XL, Li WJ, Xu LH, Wen ML, Peng Q, Jiang CL. *Salinicoccus salitudinis* sp. nov., a new moderately halophilic bacterium isolated from a saline soil sample. *Extremophiles* 2008; **12**:197-203. [PubMed http://dx.doi.org/10.1007/s00792-007-0116-8](http://dx.doi.org/10.1007/s00792-007-0116-8)
- Kampfer P, Arun AB, Busse HJ, Young CC, Lai WA, Rekha PD, Chen WM. *Salinicoccus sesuvii* sp. nov., isolated from the rhizosphere of *Sesuvium portulacastrum*. *Int J Syst Evol Microbiol* 2011; **61**:2348-2352. [PubMed http://dx.doi.org/10.1099/ijs.0.027524-0](http://dx.doi.org/10.1099/ijs.0.027524-0)
- Qu Z, Li Z, Zhang X, Zhang XH. *Salinicoccus qingdaonensis* sp. nov., isolated from coastal seawater during a bloom of green algae. *Int J Syst*

- Evol Microbiol* 2012; **62**:545-549. [PubMed](#)  
<http://dx.doi.org/10.1099/ijs.0.030551-0>
14. Ramana CV, Srinivas A, Subhash Y, Tushar L, Mukherjee T, Kiran PU, Sasikala C. *Salinicoccus halitifaciens* sp. nov., a novel bacterium participating in halite formation. *Anton Leeuw Int J G* 2013.
  15. DasSarma SAP. Halophiles. In *Encyclopedia of Life Sciences*, Nature Publishing Group 2002; Volume 8:458-466.
  16. Galinski EA. Compatible Solutes of Halophilic Eubacteria - Molecular Principles, Water-Solute Interaction, Stress Protection. *Experientia* 1993; **49**:487-496.  
<http://dx.doi.org/10.1007/BF01955150>
  17. Roberts MF. Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Syst* 2005; **1**:5. [PubMed](#)  
<http://dx.doi.org/10.1186/1746-1448-1-5>
  18. Margesin R, Schinner F. Potential of halotolerant and halophilic microorganisms for biotechnology. *Extremophiles* 2001; **5**:73-83. [PubMed](#)  
<http://dx.doi.org/10.1007/s007920100184>
  19. Le Borgne S, Paniagua D, Vazquez-Duhalt R. Biodegradation of organic pollutants by halophilic bacteria and archaea. *J Mol Microbiol Biotechnol* 2008; **15**:74-92. [PubMed](#)  
<http://dx.doi.org/10.1159/000121323>
  20. Jung MJ, Kim MS, Roh SW, Shin KS, Bae JW. *Salinicoccus carniancri* sp. nov., a halophilic bacterium isolated from a Korean fermented seafood. *Int J Syst Evol Microbiol* 2010; **60**:653-658. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.012047-0>
  21. Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, Park SC, Jeon YS, Lee JH, Yi H, et al. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol* 2012; **62**:716-721. [PubMed](#)  
<http://dx.doi.org/10.1099/ijs.0.038075-0>
  22. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**:4673-4680. [PubMed](#)  
<http://dx.doi.org/10.1093/nar/22.22.4673>
  23. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**:406-425. [PubMed](#)
  24. Kluge AGFF. Quantitative phyletics and the evolution of anurans. *Syst Zool* 1969; **18**:1-32.  
<http://dx.doi.org/10.2307/2412407>
  25. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 1981; **17**:368-376. [PubMed](#)  
<http://dx.doi.org/10.1007/BF01734359>
  26. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011; **28**:2731-2739. [PubMed](#)  
<http://dx.doi.org/10.1093/molbev/msr121>
  27. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](#)  
<http://dx.doi.org/10.1038/nbt1360>
  28. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](#)  
<http://dx.doi.org/10.1073/pnas.87.12.4576>
  29. Gibbons NEMR. Proposals concerning the higher taxa of bacteria. *Int J Syst Bacteriol* 1978; **28**:1-6.  
<http://dx.doi.org/10.1099/00207713-28-1-1>
  30. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.
  31. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), *Bergey's Manual of Systematic Bacteriology*, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.
  32. Ludwig WSK, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds). *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 3, Springer-Verlag, New York 2009:p. 19-20.
  33. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; **60**:469.  
<http://dx.doi.org/10.1099/ijs.0.022855-0>
  34. Prévot AR. Dictionnaire des Bactéries Pathogènes. In Hauduroy, Ehringer, Guillot, Magrou, Prévot,



- Rossetti and Urbain (eds) 2nd edition. Masson, Paris, 1953:1-692.
35. Skerman VBDMV. Sneath PHA Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. <http://dx.doi.org/10.1099/00207713-30-1-225>
  36. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; **60**:469-472. <http://dx.doi.org/10.1099/ijs.0.022855-0>
  37. Schleifer KH, Bell JA. Family VIII. *Staphylococcaceae* fam. nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 392.
  38. Validation List no. 34. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. *Int J Syst Bacteriol* 1990; **40**:320-321. <http://dx.doi.org/10.1099/00207713-40-3-320>
  39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25. [PubMed http://dx.doi.org/10.1038/75556](http://dx.doi.org/10.1038/75556)
  40. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010; **38**:D346-D354. [PubMed http://dx.doi.org/10.1093/nar/gkp848](http://dx.doi.org/10.1093/nar/gkp848)
  41. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008; **9**:75. [PubMed http://dx.doi.org/10.1186/1471-2164-9-75](http://dx.doi.org/10.1186/1471-2164-9-75)
  42. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 2007; **23**:673-679. [PubMed http://dx.doi.org/10.1093/bioinformatics/btm009](http://dx.doi.org/10.1093/bioinformatics/btm009)
  43. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 2005; **33**:5691-5702. [PubMed http://dx.doi.org/10.1093/nar/gki866](http://dx.doi.org/10.1093/nar/gki866)
  44. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 2000; **28**:33-36. [PubMed http://dx.doi.org/10.1093/nar/28.1.33](http://dx.doi.org/10.1093/nar/28.1.33)
  45. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* 2012; **40**:D130-D135. [PubMed http://dx.doi.org/10.1093/nar/gkr1079](http://dx.doi.org/10.1093/nar/gkr1079)
  46. Yu C, Zavaljevski N, Desai V, Reifman J. Genome-wide enzyme annotation with precision control: catalytic families (CatFam) databases. *Proteins* 2009; **74**:449-460. [PubMed http://dx.doi.org/10.1002/prot.22167](http://dx.doi.org/10.1002/prot.22167)
  47. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al. The Pfam protein families database. *Nucleic Acids Res* 2010; **38**:D211-D222. [PubMed http://dx.doi.org/10.1093/nar/gkp985](http://dx.doi.org/10.1093/nar/gkp985)
  48. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. [PubMed http://dx.doi.org/10.1093/nar/gkm160](http://dx.doi.org/10.1093/nar/gkm160)
  49. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. [PubMed](http://dx.doi.org/10.1093/nar/gkp985)
  50. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed http://dx.doi.org/10.1093/bioinformatics/btp393](http://dx.doi.org/10.1093/bioinformatics/btp393)