

Genome sequence of the halotolerant bacterium *Corynebacterium halotolerans* type strain YIM 70093^T (= DSM 44683^T)

Christian Rückert¹, Andreas Albersmeier¹, Arwa Al-Dilaimi¹, Karsten Niehaus², Rafael Szczepanowski¹, Jörn Kalinowski^{1*}

¹ Technology Platform Genomics, CeBiTec, Bielefeld University, Bielefeld, Germany

² Proteome and Metabolome Research, Bielefeld University, Bielefeld, Germany

*Corresponding author: Jörn Kalinowski (Joern@CeBiTec.Uni-Bielefeld.DE)

Keywords: aerobic, non-motile, Gram-positive, mesophilic, halotolerant

Corynebacterium halotolerans Chen *et al.* 2004 is a member of the genus *Corynebacterium* which contains Gram-positive bacteria with a high G+C content. *C. halotolerans*, isolated from a saline soil, belongs to the non-lipophilic, non-pathogenic corynebacteria. It displays a high tolerance to salts (up to 25%) and is related to the pathogenic corynebacteria *C. freneyi* and *C. xerosis*. As this is a type strain in a subgroup of *Corynebacterium* without complete genome sequences, this project describing the 3.14 Mbp long chromosome and the 86.2 kbp plasmid pCha1 with their 2,865 protein-coding and 65 RNA genes will aid the *Genomic Encyclopedia of Bacteria and Archaea* project.

Introduction

Strain YIM 70093^T (= DSM 44683^T) is the type strain of the species *Corynebacterium halotolerans* [1] and was originally isolated from saline soil in Xinjiang Province in western China. The genus *Corynebacterium* is comprised of Gram-positive bacteria with a high G+C content. It currently contains over 80 members [2] isolated from diverse backgrounds like human clinical samples [3] and animals [4], but also from soil [5] and ripening cheese [6].

Within this diverse genus, *C. halotolerans* has been proposed to form a subclade together with *C. freneyi* and *C. xerosis* [1]. Data concerning salt tolerance is not available for most corynebacteria, but *C. halotolerans* YIM 70093^T displays the highest resistance to salt (up to 25%) described for *Corynebacterium* so far. Here we present a summary classification and a set of features for *C. halotolerans* YIM 70093^T, together with the description of the genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of *C. halotolerans* YIM 70093^T was compared to the Ribosomal Database Project database [7], confirming the initial taxonomic classification. Addition of the recently published species *C. maris* Coryn-1^T [8], *C. marinum* 7015^T [9] and *C. humireducens* MFC-5^T

[10] as well as *C. diphtheriae* NCTC 11397^T [11] indicates that *C. halotolerans* YIM 70093^T, together with *C. maris*, *C. marinum*, and *C. humireducens*, form a distinct subclade within the genus *Corynebacterium*. Interestingly, *C. xerosis* and *C. freneyi* do not group closely with this subclade when *C. diphtheriae* is added to the comparison.

Figure 1 shows the phylogenetic neighborhood of *C. halotolerans* in a 16S rRNA based tree. The sequences of the four identical 16S rRNA gene copies in the genome differ by eight nucleotides from the previously published 16S rRNA sequence (AY226509), which contains two ambiguous bases.

C. halotolerans YIM 70093^T is Gram-positive and cells are rod-shaped, 0.5-1 μm long and 0.25-0.5 μm wide (Table 1 and Figure 2). It is described to be non-motile [1], which coincides with a complete lack of genes associated with 'cell motility' (functional category N). Optimal growth of YIM 70093^T was shown to occur at 28°C, pH 7.2 and 100 g/l KCl, albeit the strain tolerates a wide range of salinity, between 0-250 g/l, NaCl, and MgCl₂ [1]. Carbon sources utilized by strain YIM 70093^T include glucose, galactose, sucrose, arabinose, mannose, mannitol, maltose, xylose, ribose, salicin, dextrin, and starch [1], although the latter is doubtful as *C. halotolerans* cannot hydrolyze starch [1].

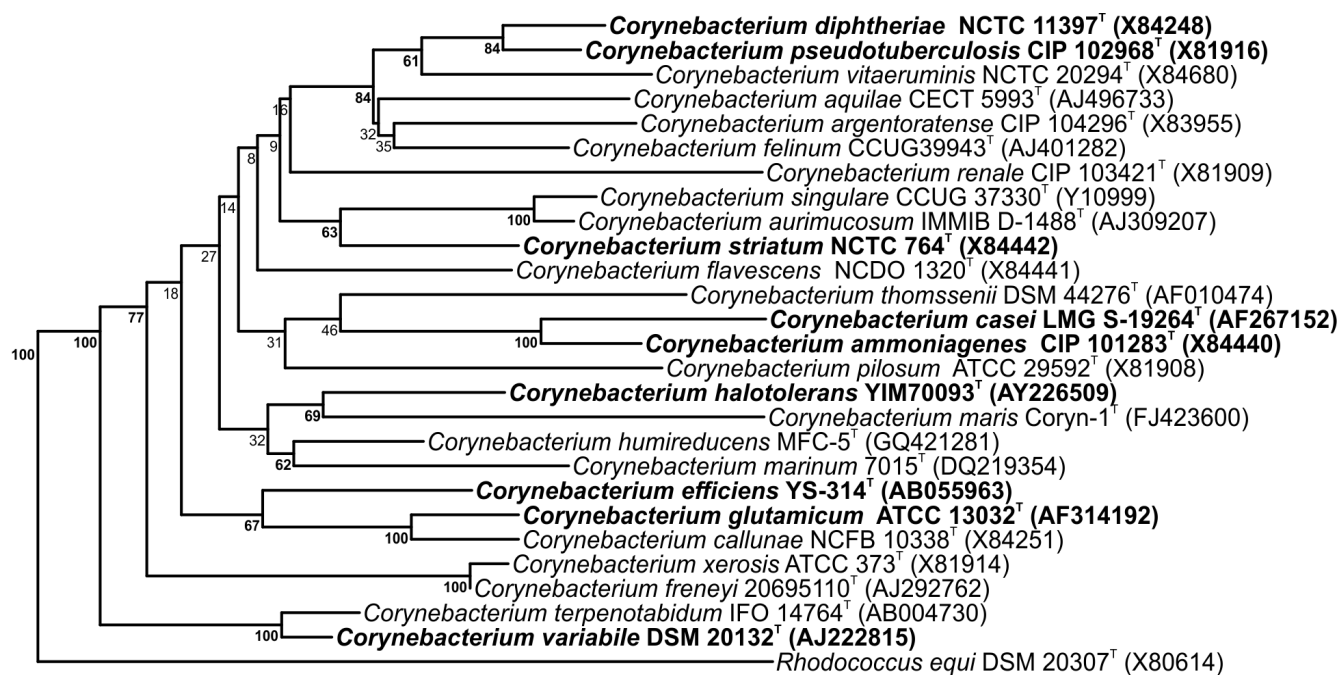


Figure 1. Phylogenetic tree highlighting the position of *C. halotolerans* relative to type strains of other species within the genus *Corynebacterium* as selected by Chen et al. [1]. In addition, the recently described *C. maris*, *C. marinum*, and *C. humireducens* were added, as they were shown to be closely related. Furthermore, the type strain of the genus, *C. diphtheriae* [11], was included. Species with at least one publicly available genome sequence (not necessarily the type strain) are highlighted in **bold face**. The tree is based on sequences aligned by the RDP aligner, utilizes the Jukes-Cantor corrected distance model to construct a distance matrix based on alignment model positions without the use of alignment inserts, and uses a minimum comparable position of 200. The tree is built with RDP Tree Builder, which uses Weighbor [12] with an alphabet size of 4 and length size of 1,000. The building of the tree also involves a bootstrapping process repeated 100 times to generate a majority consensus tree [13]. *Rhodococcus equi* (X80614) was used as an outgroup.

Chemotaxonomy

The peptidoglycan of strain YIM 70093^T contains *meso*-diaminopimelic acid, galactose, and arabinose [1], therefore it belongs to cell wall type IV, sugar type A. The menaquinones detected in the cell membrane of YIM 70093^T are MK-8(H₂) (35.5%) and MK-9(H₂) (64.5%) [1]. Cellular fatty acids are predominantly saturated straight chain acids, C_{16:0} (42.1%), C_{14:0} (7.3%); and C_{18:0} (4.5%), and unsaturated acids, *cis*-9-C_{18:1} (28.9%) and *cis*-9-C_{16:1} (9.8%), in addition to 10-methyl C_{18:0} (7.4%) [1]. Like many, but not all corynebacteria, *C. halotolerans* also contains mycolic acids, predominantly of the short chain type (C₃₂-C₃₆): C_{32:0} (36.0%), C_{34:0} (20.8%), C_{34:1} (25.1%), C_{36:0} (3.6%), C_{36:1} (8.4%), and C_{36:2} (5.1%) [1]. The reported major polar lipids consist of diphosphatidylglycerol (DPG), phosphatidylglycerol (PG),

phosphatidylinositol (PI), glycolipid and phosphatidylinositol mannosides (PIM) [1].

Genome sequencing and annotation

Genome project history

C. halotolerans YIM 70093^T was selected for sequencing as part of a project to define the core genome and pan genome of the non-pathogenic corynebacteria due to its phylogenetic position and interesting capabilities, i.e. high salt tolerance. While not being a part of the *Genomic Encyclopedia of Bacteria and Archaea* (GEBA) project [26], sequencing of the type strain will nonetheless aid the GEBA effort. The genome project is deposited in the Genomes On Line Database [27] and the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the Center of Biotechnology (CeBiTec). A summary of the project information is shown in Table 2.

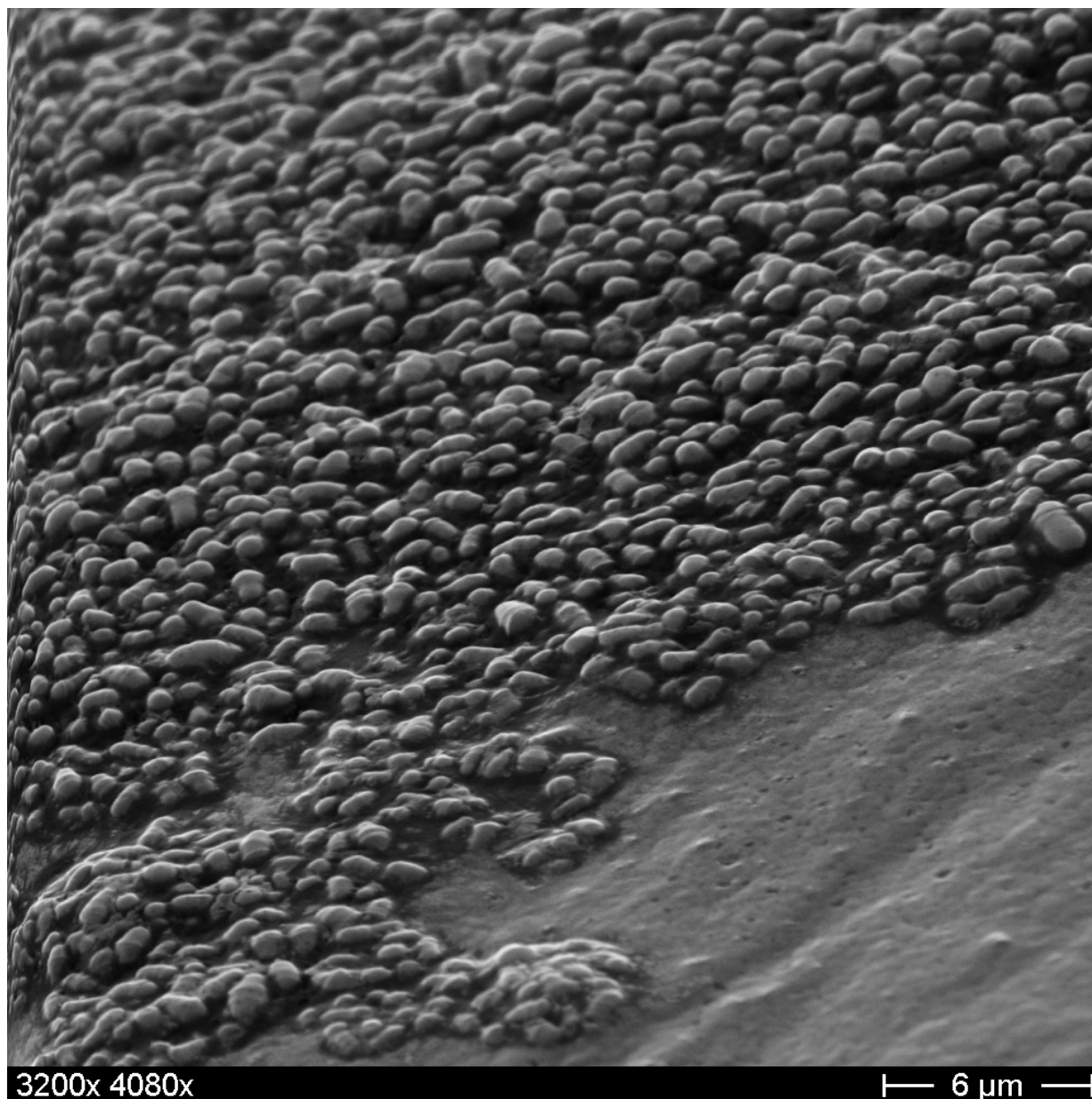


Figure 2. Scanning electron micrograph of *C. halotolerans* YIM 70093^T.

Growth conditions and DNA isolation

C. halotolerans strain YIM 70093^T, DSM 44683, was grown aerobically in CASO broth (Carl Roth GmbH, Karlsruhe, Germany) at 30°C. DNA was isolated from $\sim 10^8$ cells using the protocol described by Tauch *et al.* 1995 [28].

Genome sequencing and assembly

The genome was sequenced using a 454 sequencing platform. A standard 3k paired end sequencing library was prepared according to the manufacturer's protocol (Roche). Pyrosequencing reads were assembled using the Newbler assembler v2.3

(Roche). The initial Newbler assembly consisted of 81 contigs in six scaffolds with an additional 26 lone contigs. Analysis of the six scaffolds revealed one to be an extrachromosomal element (plasmid pCha1), four to make up the chromosome with the remaining one to contain the four copies of the RRN operon which caused the scaffold breaks. The scaffolds were ordered based on alignments to the complete genomes of *C. glutamicum* [29] and *C. efficiens* [30] and subsequent verification by restriction digestion, Southern blotting and hybridization with a 16S rDNA specific probe.

Table 1. Classification and general features of *C. halotolerans* YIM 70093^T according to the MIGS recommendations [14].

MIGS ID	Property	Term	Evidence code ^a
		Domain <i>Bacteria</i>	TAS [15]
		Phylum <i>Actinobacteria</i>	TAS [16]
		Class <i>Actinobacteria</i>	TAS [17]
	Current classification	Order <i>Actinomycetales</i>	TAS [17-20]
		Family <i>Corynebacteriaceae</i>	TAS [17,18,20,21]
		Genus <i>Corynebacterium</i>	TAS [18,22,23]
		Species <i>Corynebacterium halotolerans</i>	TAS [1]
		Type-strain YIM 70093 (=DSM 44683)	TAS [1]
	Gram stain	Positive	TAS [1]
	Cell shape	diphtheroid, irregular rods	TAS [1]
	Motility	non-motile	TAS [1]
	Sporulation	non-sporulating	TAS [1]
	Temperature range	Mesophile	NAS
	Optimum temperature	28°C	TAS [1]
	Salinity	0-250 g/l KCl/NaCl/MgCl ₂	TAS [1]
MIGS-22	Oxygen requirement	Aerobe	TAS [1]
	Carbon source	glucose, galactose, sucrose, arabinose, mannose, mannitol, maltose, starch, xylose, ribose, salicin, dextrin	TAS [1]
	Energy metabolism	Chemoorganoheterotroph	TAS [1]
	Terminal electron acceptor	Oxygen	NAS
MIGS-6	Habitat	saline soil	TAS [1]
MIGS-15	Biotic relationship	free living	NAS
MIGS-14	Pathogenicity	non-pathogenic	NAS
	Biosafety level	1	TAS [24]
MIGS-23.1	Isolation	saline soil	TAS [1]
MIGS-4	Geographic location	Xinjiang Province, China	TAS [1]
MIGS-5	Sample collection time	Not reported	
MIGS-4.1	Latitude	Not reported	
MIGS-4.2	Longitude	Not reported	
MIGS-4.3	Depth	Not reported	
MIGS-4.4	Altitude	Not reported	

a) Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [25].

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Two genomic libraries: one 454 pyrosequencing PE library (3.2 kb insert sizes), one Illumina library
MIGS-29	Sequencing platforms	454 GS FLX Titanium, Illumina GA IIx
MIGS-31.2	Sequencing coverage	22.5 × Pyrosequencing; 23.5 × SBS
MIGS-30	Assemblers	Newbler version 2.3
MIGS-32	Gene calling method	GeneMark, Glimmer
	INSDC ID	CP003697, CP003698
	GenBank Date of Release	July 1, 2013 / after publication
	GOLD ID	Gi19308
	NCBI project ID	168616
MIGS-13	Source material identifier	DSM 44683
	Project relevance	Industrial, GEBA

The Phred/Phrap/Consed software package [31-34] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, gaps between contigs were closed by editing in Consed (for repetitive elements) and by PCR with subsequent Sanger sequencing (IIT Biotech GmbH, Bielefeld, Germany). A total of 61 additional reactions were necessary to close gaps not caused by repetitive elements. To raise the quality of the assembled sequence, Illumina reads were used to correct potential base errors and increase consensus quality. A WGS library was prepared using the Illumina-Compatible Nextera DNA Sample Prep Kit (Epicentre, WI, U.S.A) according to the manufacturer's protocol. The library was sequenced in an 80 bp single read *GAIIx* run, yielding 1,497,321 total reads. Together, the combination of the Illumina and 454 sequencing platforms provided 46.0× coverage of the genome.

Genome annotation

Gene prediction and annotation were done using the PGAAP pipeline [35]. Genes were identified using GeneMark [36], GLIMMER [37], and Prodigal

[38]. For annotation, BLAST searches against the NCBI Protein Clusters Database [39] were performed and the annotation was enriched by searches against the Conserved Domain Database [40] and subsequent assignment of coding sequences to COGs. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE [41], Infernal [42], RNAMMer [43], Rfam [44], TMHMM [45], and SignalP [46].

Genome properties

The genome includes one plasmid, for a total size of 3,222,008 bp, with one circular chromosome of 3,135,752 bp (68.44% G+C content) and one plasmid of 86,256 bp (63.20% G+C content) [Figure 3 and Figure 4]. For the main chromosome, 2,856 genes were predicted, 2,791 of which are protein-coding genes. 1,632 (57%) of the protein-coding genes were assigned to a putative function with the remaining annotated as hypothetical proteins. 1,914 protein coding genes belong to 396 paralogous families in this genome corresponding to a gene content redundancy of 66.8%. The properties and the statistics of the genome are summarized in Table 3, Tables 4 and 5.

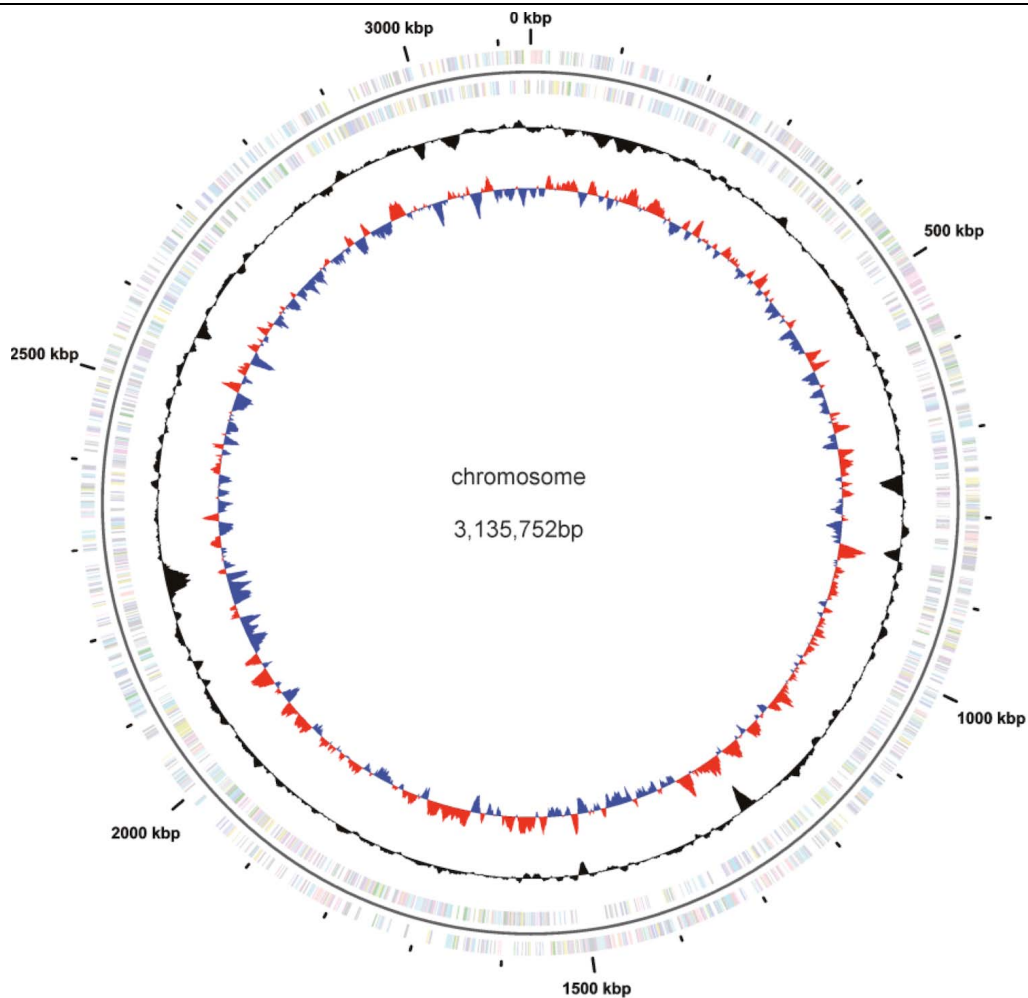


Figure 3. Graphical map of the chromosome (not drawn to scale with plasmid). From the outside in: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), GC content, GC skew.

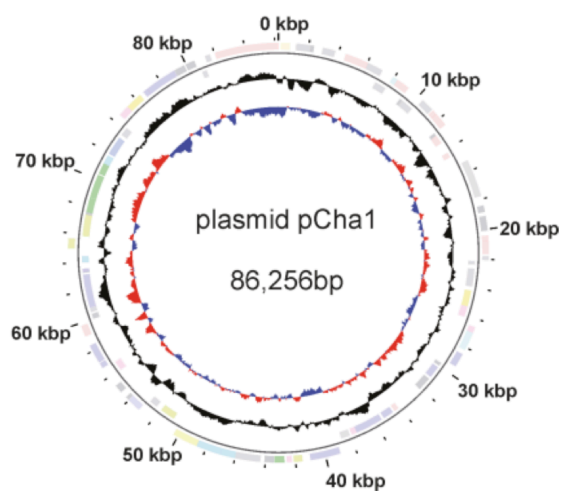


Figure 4. Graphical map of the plasmid pCha1 (not drawn to scale with chromosome). From the outside in: Genes on forward strand (color by COG categories), genes on reverse strand (color by COG categories), GC content, GC skew.

Table 3. Summary of genome: one chromosome and one plasmid

Label	Size (Mb)	Topology	INSDC identifier
Chromosome	3.136	circular	CP003697.1
Plasmid pCha1	0.086	circular	CP003698.1

Table 4. Genome Statistics

Attribute	Value	% of total ^a
Genome size (bp)	3,222,008	100.00%
DNA coding region (bp)	2,791,134	86.63%
DNA G+C content (bp)	2,200,760	68.30
Total genes ^b	2,930	100.00%
RNA genes	65	2.22%
rRNA operons	4	
tRNA genes	53	1.81%
Protein-coding genes	2,865	97.78%
Genes with function prediction (protein)	1,632	56.96%
Genes assigned to COGs	2,234	77.98%
Gene in paralog clusters	1,914	66.81%
Genes with signal peptides	251	8.76%
Genes with transmembrane helices	686	23.94%

a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

Table 5. Number of genes associated with the general COG functional categories

Code	Value	%age	Description
J	155	5.41	Translation, ribosomal structure and biogenesis
A	1	0.03	RNA processing and modification
K	185	6.46	Transcription
L	141	4.92	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	20	0.70	Cell cycle control, cell division, chromosome partitioning
Y	0	0.00	Nuclear structure
V	44	1.54	Defense mechanisms
T	81	2.83	Signal transduction mechanisms
M	126	4.40	Cell wall/membrane biogenesis
N	0	0.00	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	25	0.87	Intracellular trafficking and secretion, and vesicular transport
O	88	3.07	Posttranslational modification, protein turnover, chaperones
C	176	6.14	Energy production and conversion
G	183	6.39	Carbohydrate transport and metabolism
E	262	9.14	Amino acid transport and metabolism
F	68	2.37	Nucleotide transport and metabolism
H	122	4.26	Coenzyme transport and metabolism
I	88	3.07	Lipid transport and metabolism
P	196	6.84	Inorganic ion transport and metabolism
Q	85	2.97	Secondary metabolites biosynthesis, transport and catabolism
R	360	12.57	General function prediction only
S	214	7.47	Function unknown
-	631	22.02	Not in COGs

Acknowledgements

Christian Rückert acknowledges funding through a grant by the Federal Ministry for Education and Research (0316017) within the BioIndustry2021 initiative.

References

- Chen HH, Li WJ, Tang SK, Kroppenstedt RM, Stackebrandt E, Xu LH, Jiang CL. *Corynebacterium halotolerans* sp. nov., isolated from saline soil in the west of China. *Int J Syst Evol Microbiol* 2004; **54**:779-782. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.02919-0>
- Euzéby 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>
- Renaud FNR, Aubel D, Riegel P, Meugnier H, Bollet C. *Corynebacterium freneyi* sp. nov., alpha-glucosidase-positive strains related to *Corynebacterium xerosis*. *Int J Syst Evol Microbiol* 2001; **51**:1723-1728. [PubMed](#) <http://dx.doi.org/10.1099/00207713-51-5-1723>
- Collins MD, Hoyles L, Foster G, Falsen E. *Corynebacterium caspium* sp. nov., from a Caspian seal (*Phoca caspica*). *Int J Syst Evol Microbiol* 2004; **54**:925-928. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.02950-0>
- Zhou Z, Yuan M, Tang R, Chen M, Lin M, Zhang W. *Corynebacterium deserti* sp. nov., isolated from desert sand. *Int J Syst Evol Microbiol* 2012; **62**:791-794. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.030429-0>
- Brennan NM, Brown R, Goodfellow M, Ward AC, Beresford TP, Simpson PJ, Fox PF, Cogan TM. *Corynebacterium mooreparkense* sp. nov. and *Corynebacterium casei* sp. nov., isolated from the surface of a smear-ripened cheese. *Int J Syst Evol Microbiol* 2001; **51**:843-852. [PubMed](#) <http://dx.doi.org/10.1099/00207713-51-3-843>
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 2009; **37**(Database issue):D141-D145. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkn879>
- Ben-Dov E, Ben Yosef DZ, Pavlov V, Kushmaro A. *Corynebacterium maris* sp. nov., a marine bacterium isolated from the mucus of the coral *Fungia granulosa*. *Int J Syst Evol Microbiol* 2009; **59**:2458-2463. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.007468-0>
- Du ZJ, Jordan EM, Rooney AP, Chen GJ, Austin B. *Corynebacterium marinum* sp. nov. isolated from coastal sediment. *Int J Syst Evol Microbiol* 2010; **60**:1944-1947. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.018523-0>
- Wu CY, Zhuang L, Zhou SG, Li FB, He J. *Corynebacterium humireducens* sp. nov., an alkaliphilic, humic acid-reducing bacterium isolated from a microbial fuel cell. *Int J Syst Evol Microbiol* 2011; **61**:882-887. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.020909-0>
- Cerdeño-Tárraga AM, Efstratiou A, Dover LG, Holden MT, Pallen M, Bentley SD, Besra GS, Churcher C, James KD, De Zoysa A, et al. The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res* 2003; **31**:6516-6523. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkg874>
- Bruno WJ, Socci ND, Halpern AL. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol* 2000; **17**:189-197. [PubMed](#) <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026231>
- Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* 2007; **35**(Database issue):D169-D172. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkl889>
- 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>
- 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>
- 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.
17. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a New Hierarchic Classification System, *Actinobacteria* classis nov. *Int J Syst Bacteriol* 1997; **47**:479-491. <http://dx.doi.org/10.1099/00207713-47-2-479>
 18. Skerman VBD, McGowan V, 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>
 19. Buchanan RE. Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the *Schizomycetes*. *J Bacteriol* 1917; **2**:155-164. [PubMed](#)
 20. Lehmann KB, Neumann R. Lehmann's Medizin, Handatlanten. X Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik., Fourth Edition, Volume 2, J.F. Lehmann, München, 1907, p. 270.
 21. Zhi XY, Li WJ, Stackebrandt E. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *Actinobacteria*, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int J Syst Evol Microbiol* 2009; **59**:589-608. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.65780-0>
 22. Lehmann KB, Neumann R. Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik, First Edition, J.F. Lehmann, München, 1896, p. 1-448.
 23. Bernard KA, Wiebe D, Burdz T, Reimer A, Ng B, Singh C, Schindle S, Pacheco AL. Assignment of *Brevibacterium stationis* (ZoBell and Upham 1944) Breed 1953 to the genus *Corynebacterium*, as *Corynebacterium stationis* comb. nov., and emended description of the genus *Corynebacterium* to include isolates that can alkalize citrate. *Int J Syst Evol Microbiol* 2010; **60**:874-879. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.012641-0>
 24. BAuA. 2010 Classification of bacteria and archaea in risk groups, TRBA 466. <http://www.baua.de/de/Themen-von-A-Z/Biologische-Arbeitsstoffe/TRBA/TRBA-466.html>.
 25. 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. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](#) <http://dx.doi.org/10.1038/75556>
 26. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, et al. A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea*. *Nature* 2009; **462**:1056-1060. [PubMed](#) <http://dx.doi.org/10.1038/nature08656>
 27. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes OnLine 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>
 28. Tauch A, Kassing F, Kalinowski J, Pühler A. The *Corynebacterium xerosis* composite transposon Tn5432 consists of two identical insertion sequences, designated IS1249, flanking the erythromycin resistance gene ermCX. *Plasmid* 1995; **34**:119-131. [PubMed](#) <http://dx.doi.org/10.1006/plas.1995.9995>
 29. Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, et al. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol* 2003; **104**:5-25. [PubMed](#) [http://dx.doi.org/10.1016/S0168-1656\(03\)00154-8](http://dx.doi.org/10.1016/S0168-1656(03)00154-8)
 30. Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, Sugimoto S, Matsui K, Yamagishi A, Kikuchi H, Ikeo K, et al. Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. *Genome Res* 2003; **13**:1572-1579. [PubMed](#) <http://dx.doi.org/10.1101/gr.1285603>
 31. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; **8**:186-194. [PubMed](#)
 32. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**:195-202. [PubMed](#)
 33. Gordon D. Viewing and editing assembled sequences using Consed. *Curr Protoc Bioinformatics* 2003;Chapter 11:Unit11 2.
 34. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; **8**:175-185. [PubMed](#)

35. NCBI. 2010 NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP). <http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html>.
36. Borodovsky M, Mills R, Besemer J, Lomsadze A. Prokaryotic gene prediction using GeneMark and GeneMark.hmm. *Curr Protoc Bioinformatics* 2003;Chapter 4:Unit4 5.
37. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 1999; **27**:4636-4641. [PubMed](http://dx.doi.org/10.1093/nar/27.23.4636) <http://dx.doi.org/10.1093/nar/27.23.4636>
38. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [PubMed](http://dx.doi.org/10.1186/1471-2105-11-119) <http://dx.doi.org/10.1186/1471-2105-11-119>
39. Klimke W, Agarwala R, Badretdin A, Chetvernin S, Ciufu S, Fedorov B, Kiryutin B, O'Neill K, Resch W, Resenchuk S, *et al.* The National Center for Biotechnology Information's Protein Clusters Database. *Nucleic Acids Res* 2009; **37**(Database issue):D216-D223. [PubMed](http://dx.doi.org/10.1093/nar/gkn734) <http://dx.doi.org/10.1093/nar/gkn734>
40. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, *et al.* CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 2009; **37**(Database issue):D205-D210. [PubMed](http://dx.doi.org/10.1093/nar/gkn845) <http://dx.doi.org/10.1093/nar/gkn845>
41. 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/gkn845)
42. Eddy SR. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. *BMC Bioinformatics* 2002; **3**:18. [PubMed](http://dx.doi.org/10.1186/1471-2105-3-18) <http://dx.doi.org/10.1186/1471-2105-3-18>
43. 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>
44. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*. 2005;33 Database Issue:D121-124.
45. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; **305**:567-580. [PubMed](http://dx.doi.org/10.1006/jmbi.2000.4315) <http://dx.doi.org/10.1006/jmbi.2000.4315>
46. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795. [PubMed](http://dx.doi.org/10.1016/j.jmb.2004.05.028) <http://dx.doi.org/10.1016/j.jmb.2004.05.028>