Yang et al. Bioresources and Bioprocessing (2015) 2:1
DOI 10.1186/s40643-014-0030-8

Bioresources and Bioprocessing
a SpringerOpen Journal

RESEARCH Open Access

Genotyping of amino acid-producing *Corynebacterium glutamicum* strains based on multi-locus sequence typing (MLST) scheme

Junjie Yang[1,2,4], Yimeng Kong[1,3] and Sheng Yang[1,2,4*]

Abstract

Background: Confusing parental information may hinder to dissect mechanisms of amino acid hyper-producing *Corynebacterium glutamicum* strains. Thus, an efficient method for genotyping of the *C. glutamicum* is heavily called.

Results: Multi-locus sequence typing (MLST) is currently the most popular molecular typing technique. But currently this method is not available for *C. glutamicum*. In this study, a MLST scheme was established based on sequences of seven housekeeping genes, for genotyping of *C. glutamicum*. The MLST method performed an efficient discrimination of 17 strains and helps to understand the population structure of this bacterium.

Conclusions: This work has expanded the MLST method to *C. glutamicum* and developed an efficient technique to discriminate strains of uncertain origin.

Keywords: *Corynebacterium glutamicum*; Multi-locus sequence typing; Amino acid producing; *Corynebacterium crenatum*; 16S rDNA

Background

Non-spore-forming Gram-positive bacteria *Corynebacterium glutamicum* are widely used in amino acid production industry with numerous ideal attributes [1,2]. Due to roughly identical 16S rDNA sequences, many independent strains previously regarded as *Brevibacterium* and *Corynebacterium* species in the Corynebacteriaceae family have been categorized into the *C. glutamicum* species, especially *Brevibacterium lactofermentum*, *Brevibacterium flavum*, and *Corynebacterium acetoacidophilum* [3-5]. Consequently, the *C. glutamicum* species dropped into a mixture of strains from diversified resources.

Clear genetic or phylogenetic information is occasionally missing after a long-term utilization and preservation. Confusing parental information may hinder to dissect mechanisms of amino acid hyper-production in these strains. For example, although genome sequences are now available for several glutamate-hyper-producing strains, e.g., SCgG1, SCgG2, and Z188, detail typing and parental information of these strains were still absent. To our knowledge, only an identification method by phage sensitivity has been reported, but this method is not very convenient because of the experimental procedure for handling phages [6]. Thus, an efficient method for genotyping of the *C. glutamicum* is heavily called for in this field.

Compared with other molecular typing techniques, multi-locus sequence typing (MLST) is currently the most popular one [7,8]. MLST relies on allelic variants in conserved genes to calculate phylogenetic relationship of strains. Multiple housekeeping genes (usually seven genes) are examined in the analysis as their sequences are constrained and variations are nearly neutral. Each strain is then assigned with an allelic profile or sequence type (ST). The MLST method excels other molecular

* Correspondence: syang@sibs.ac.cn
[1]Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China
[2]Shanghai Research Center of Industrial Biotechnology, Shanghai 201201, China
Full list of author information is available at the end of the article

typing methods of being unambiguous and easily portable between laboratories.

To our knowledge, in the family of Corynebacteriaceae, MLST scheme is only developed in *Corynebacterium diphtheriae*. More than 70 STs were identified and fall into at least 11 groups of *C. diphtheriae*. Meanwhile, no ST is currently available for *C. glutamicum*. In order to have accurate typing techniques for *C. glutamicum*, we developed a MLST scheme, which would promote applied research of this amino acid-producing species.

Methods

The strains used in this study are listed in Table 1 [3-6,9-13]. Genome sequences of the strains used for MLST analysis include ATCC13032 [5,2,1], MB1 [14], R [15], ScCG1, ScCG2, ATCC14067 [16], S9114 [17], Z188, '*C. crenatum*' MT. Primers 27F 5′-AGAGTTTGATCMTGGCTCAG-3′ and 1492R 5′-TACGGYTACCTTGTTACGACTT-3′ were used to identify 16S rDNA sequences of the strains [3,4]. The retrieved 16S rDNA sequences are almost identical, with a similarity >99% compared with the sequence of the *C. glutamicum* ATCC 13032, indicating that all of the strains belong to *C. glutamicum* [3,4].

Referring to the genotyping scheme in *C. diphtheriae* [7], another species belonging to the same genus, seven housekeeping genes of the *C. glutamicum* strains were selected for analysis, including *atpA*, *dnaE*, *dnaK*, *fusA*, *rpoB*, *leuA*, and *odhA* (Table 2). All the loci are single copy in the genome sequences, according to the BLAST analysis. Considering the possibility to meet the case when only crude amino acid fermentation samples are available, which would limit the efficiency of PCR reactions, approximately 300~400 bp length of the PCR amplicon in housekeeping genes was designed for downstream experiments and analysis (Table 2).

Results and discussion

The PCR amplicon size ranges from 318 bp (*leuA*) to 402 bp (*fusA*), with a mean size of 369 bp (Table 2, Additional file 1: Figure S1 and Dataset S1). Each locus of housekeeping gene alleles for comparison and identification has a length ranging from 150 bp (*leuA*) to 243 bp (*dnaK*), with a mean length of 206 bp. All alleles for a given locus are of equal lengths and in a correct reading frame. The proportion of variation sites in each locus varies from 3% (*odhA*) to 17% (*leuA*) (Table 2). Primer and allele sequences are listed in Table 2 and Additional file 1: Dataset S2.

The dendrogram was drawn from allelic profile data using the UPGMA (unweighted pair group method with arithmetic mean) method. The online tool (http://pubmlst.org/) of the PHYLIP suite programs and Phylodendron [18] were implemented for tree generation, output, and display. The dendrogram demonstrates that eight groups of the strains could be classified according to their genetic distinctions (Figure 1). Except ST4/ST5 and ST9/ST10, which share six alleles respectively, most of STs share no more than three alleles with each other (Table 1), indicating that the strains could be well separated. The next section briefly describes the eight classified groups.

Group 1 (ST#1): The *C. glutamicum* type strain ATCC13032 is a member of this group. MB001 is a prophage-free variant of ATCC13032 with a 6% reduced genome [14]. Consistently, it has the same ST with ATCC13032.

Group 2 (ST#2): '*B. lactofermentum*' ATCC13869 represents an independent ST, as expected.

Group 3 (ST#3): '*C. acetoacidophilum*' ATCC13870 also represents an independent ST.

Group 4 (ST#4 and #5): Although '*B. flavum*' ATCC14067 and its derived strain ATCC21493 [13] contain different STs, in fact, there is only 1-bp difference (in *odhA*, resulting in Ala701Thr) among the seven analyzed genes. So, the classification is correct.

Group 5 (ST#6): *C. glutamicum* R [15] represents an independent ST.

Group 6 (ST#7): '*C. pekinense*' AS1.299 and its derived strain AS1.563 [9,10] contain the same ST and were consequently classified into the same group.

Group 7 (ST#8): *C. glutamicum* 617 (SIIM B1) also represents an independent ST.

Group 8 (ST#9 and #10): Several '*B. tianjinese*' and '*C. crenatum*' strains, including both 'wild-type' strains [11,12,19] and derived strains [12,17,20], represent the same ST (ST#9), so they are categorized in this group. The strain '*C. crenatum*' MT also fall into the group. It has a different ST (ST#10) from 'wild-type' strain AS1.542 with only one different locus.

Taken together, the MLST results reveal a correct typing of the related strains. We also tested a commercially obtained crude lysine sample and several amino acid-producing strains preserved in our laboratory. The crude lysine sample represented a MLST result same to 13032, so it would be produced by an ATCC 13032-derived strain. Three lysine-producing strains fall into group 1 (ST#1), group 2 (ST#2), and group 6 (ST#7), separately. A proline-producing strain falls into group 3 (ST#3). These results suggest that the eight groups defined in this study could encompass most of the amino acid-producing strains for laboratory use.

The aim of this work was to expand the MLST method to *C. glutamicum* and develop an efficient technique to discriminate strains of uncertain origin. After a long-term preservation, activation, and fermentation, parental information of the amino acid-producing strains is easily confused. The MLST scheme developed in our study addresses the problem

Table 1 Details and allelic profile of the strains used in this study

ST	Group	Strains	Synonym	Descriptions	Genome sequence[a]	Allelic profile						
						atpA	*dnaE*	*dnaK*	*fusA*	*leuA*	*odhA*	*rpoB*
		ATCC13032		*C. glutamicum* type strain	NC_003450, NC_006958							
1	1	MB001		Prophage-free variant of ATCC 13032 with a 6% reduced genome	NC_022040	1	1	1	1	1	1	1
		Crude sample		A crude lysine sample, commercially obtained								
2	2	ATCC13869	*B. lactofermentum*	'Wild-type *B. lactofermentum*'		1	2	2	4	2	4	4
3	3	ATCC13870	*C. acetoacidophilum*	'Wild-type *C. acetoacidophilum*'		4	6	5	5	6	1	1
4	4	ATCC14067	*B. flavum*	J. Bacteriol. 194 (3), 742-743 (2012)	AGQQ01000000	3	2	4	6	2	2	2
5	4	ATCC21493	*B. flavum*	Producing arginine, derived from ATCC 14067 (SIIM B234)		3	2	4	6	2	5	2
6	5	R		*C. glutamicum* isolated in Japan from a meadow soil sample	NC_009342	5	3	7	3	3	2	1
7	6	AS1.299	*C. pekinense*	'Wild-type *C. pekinense*,' producing glutamate (=CICC 10119, SIIM B3)		2	5	3	5	4	3	3
		AS1.563	*C. pekinense*	(=CICC 10178, SIIM B165) producing lysine, derived from AS1.299								
8	7	617		A glutamate-producing strain previously used in China (=CICC 10117, SIIM B1)		1	2	4	7	7	3	2
		T6-13	*B. tianjinese*	'Wild-type *B. tianjinese*' (=CICC 20182, SIIM B226)								
		SCgG1		Hyper-producing glutamate	NC_021351							
9	8	SCgG2		Hyper-producing glutamate	NC_021352	5	4	6	2	5	3	1
		S9114		A strain for industrial production of glutamate	AFYA01000000							
		Z188		Hyper-producing glutamate	AKXP01000000							
		AS1.542	*C. crenatum*	'Wild-type *C. crenatum*' (=CICC10124, SIIM B6)								
10	8	MT	*C. crenatum*	A mutant of AS1.542	AQPS01000000	6	4	6	2	5	3	1

[a]DDBJ/EMBL/GenBank accession number.

Table 2 Primer sequence and characteristics of the seven loci used in MLST

Gene	primer	Sequence (5′-3′)	Amplicon size (bp)	Amplicon bps (from-to) in NC_006958.1	Allele size (bp)	Number of distinct alleles	Number of polymorphic sites
atpA	atpA_F	ATGTACCAGGGCAACCAC	390	1277256-1277645	240	6	13
	atpA_R	GGCAACCTTCTTCATACC					
dnaE	dnaE_F	GGCGAAGGATATTCCGTTG	366	2220485-2220120	225	6	23
	dnaE_R	ACTTCCCCATCGCGGTTG					
dnaK	dnaK_F	CACCTCACAGGAAATCTC	397	2959446-2959050	243	7	17
	dnaK_R	GGACTGGAACTTCTCTAC					
fusA	fusA_F	CAAGGCAGCTATCCGTAAGATG	402	524614-525015	237	7	11
	fusA_R	ATGTTACCAGCGTGTGCAACC					
leuA	leuA_F	CTGGTTCAGGCTCGTGAGCA	318	267591-267274	150	7	26
	leuA_R	GATCATTGGGTTCTCAGGA					
odhA	odhA_F	CTTTGGCTGGGTCATGGAAGG	327	1175355-1175681	186	5	5
	odhA_R	CTGGGCATCGTGCCAGAAAC					
rpoB	rpoB_F	TATGTGACCGCGGAGTTC	357	513278-513634	165	4	5
	rpoB_R	GAAACGCTCGGTGATCTG					

and proved to be an efficient way to distinguish them. According to our results, three glutamate-producing strains mentioned above (SCgG1, SCgG2, and Z188) would derive from T6-13 or AS1.542. Their common genetic origin serves a firm premise for downstream comparative genomic analysis to compare parental strains and their offsprings. According to our MLST result, a lysine-producing strain CICC20042 was found to show ST#9 (including T6-13, S9114, AS1.542, etc.) (detailed data not shown). But according to the CICC database, CICC20042 is derived from AS1.563 (ST#7, including AS1.299 and AS1.563). This suggests that confusion about the parental information might have happened.

Several strains previously regarded as *Brevibacterium* and *Corynebacterium* species had been classified to *C. glutamicum* [3,4]. Our analysis extends the conclusion and proposes that *C. glutamicum* should also contain several *C. pekinense*, *B. tianjinese*, and *C. crenatum* strains. The 'wild-type' strains of *B. tianjinese* and *C. crenatum* exhibit the same ST. Both of them are highly sensitive to specific phages [6], supporting a close relationship of these strains. A brief comparative genomic analysis of S9114 ([GenBank:AFYA01000000], derived from *B. tianjinese* T6-13) and '*C. crenatum*' MT [GenBank:AQPS01000000] discovered only hundreds of SNPs and InDels (unpublished data). Some of these SNPs and InDels may be responsible for the high glutamate yield. Further genomic analyses should be conducted in a broad range of *C. glutamicum* strains, especially the parental wild-type strains, to reveal details of their relationships as well as mechanisms for amino acid hyper-production.

Conclusions

This work has expanded the MLST method to C. glutamicum and developed an efficient technique to discriminate strains of uncertain origin.

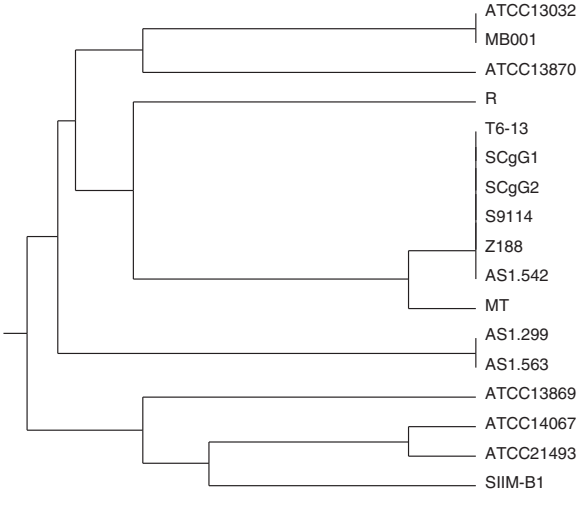

Figure 1 Dendrogram analysis of *Corynebacterium glutamicum* from allelic profile data. The dendrogram was generated with the UPGMA (unweighted pair group method with arithmetic mean) method based on the internal fragments of 7 housekeeping genes of the 17 strains. ATCC13032 and MB001, ATCC14067 and ATCC21493, and AS1.299 and AS1.563 were classified into the same groups, respectively. Several '*B. tianjinese*' and '*C. crenatum*' strains, including AS1.542, T6-13, S9114, etc., were also classified into the same group. Each of the other strains falls into an independent group separately.

Additional file

Additional file 1: Supplementary material. Figure S1. PCR fragments of seven genes: *atpA* 390 bp, *dnaE* 366 bp, *dnaK* 397 bp, *fusA* 402 bp, *rpoB* 357 bp, *leuA* 318 bp, and *odhA* 327 bp. Dataset S1. PCR fragment sequences of seven genes. Dataset S2. Allele sequences of seven genes.

Abbreviations
MLST: multi-locus sequence typing; ST: sequence type.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JY and SY designed the research. JY carried out the experiments. JY and YK carried out the computational experiments. JY and YK wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We thank Prof. Ji-Bin Sun (Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences) for providing the draft genome sequences of several strains. This work was supported by a grant from the National Basic Research Program (973 Program) of China (2014CB745100) and the National Key Technologies R&D Program of China (2012AA022101).

Author details
[1]Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China. [2]Shanghai Research Center of Industrial Biotechnology, Shanghai 201201, China. [3]University of the Chinese Academy of Sciences, Beijing 100080, China. [4]Shanghai Collaborative Innovation Center for Biomanufacturing Technology, Shanghai 200237, China.

Received: 17 October 2014 Accepted: 12 December 2014
Published online: 23 January 2015

References
1. Ikeda M, Nakagawa S. The Corynebacterium glutamicum genome: features and impacts on biotechnological processes. Appl Microbiol Biotechnol. 2003;62(2–3):99–109. doi:10.1007/s00253-003-1328-1.
2. Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, 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(1–3):5–25.
3. Liebl W, Ehrmann M, Ludwig W, Schleifer KH. Transfer of Brevibacterium divaricatum DSM 20297T, "Brevibacterium flavum" DSM 20411, "Brevibacterium lactofermentum" DSM 20412 and DSM 1412, and Corynebacterium glutamicum and their distinction by rRNA gene restriction patterns. Int J Syst Bacteriol. 1991;41(2):255–60.
4. Pascual C, Lawson PA, Farrow JA, Gimenez MN, Collins MD. Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences. Int J Syst Bacteriol. 1995;45(4):724–8.
5. Kinoshita S, Nakayama K, Akita S. Taxonomical study of glutamic acid accumulating bacteria, Micrococcus glutamicus nov. sp. Bull Agric Chem Soc Jpn. 1958;22:176–85.
6. Ge B-Z, Wang J-X, Zhu S-J, Si X-D. Identification of glutamic acid producing strains by phages (in Chinese with English abstract). Virol Sin. 1991;6(3):256–9.
7. Bolt F, Cassiday P, Tondella ML, Dezoysa A, Efstratiou A, Sing A, et al. Multilocus sequence typing identifies evidence for recombination and two distinct lineages of Corynebacterium diphtheriae. J Clin Microbiol. 2010;48(11):4177–85. doi:10.1128/JCM. 00274-10.
8. Maiden MC, van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol. 2013;11(10):728–36. doi:10.1038/nrmicro3093.
9. Chen Q, Zhang Z-Y, Li L-G. A new L-glutamic acid-producing species of Corynebacterium (in Chinese with English abstract). Acta Microbiol Sinica (Wei Sheng Wu Xue Bao). 1973;13(1):1–6.
10. Institute_of_Microbiology_Chinese_Academy_of_Sciences, Hangzhou_Glutamate_Factory. Study on the production of lysine by auxotrophic mutant of Corynebacterium pekinense AS1.299 (in Chinese). Microbiol China (Wei Sheng Wu Xue Tong Bao). 1974;1(1):7–11.
11. Chen Q, Li L-G. Studies on L-glutamic acid producing bacteria AS 1.542. I. Identification of strain AS 1.542 (in Chinese with English abstract). Acta Microbiol Sinica (Wei Sheng Wu Xue Bao). 1975;15(2):119–24.
12. Zhang K, Liu Y. Studies on glutamate dehydrogenase from Brevibacterium tianjinese T6-13 (in Chinese with English abstract). Acta Microbiol Sinica (Wei Sheng Wu Xue Bao). 1991;31(4):281–6.
13. Onoda T, Yoshinaga F, Kubota K, Kamijo H, Okumura S (1973) Method of producing l-arginine by microorganism. US Patent 3723249, patent date 27/03/1973
14. Baumgart M, Unthan S, Ruckert C, Sivalingam J, Grunberger A, Kalinowski J, et al. Construction of a prophage-free variant of Corynebacterium glutamicum ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. Appl Environ Microbiol. 2013;79(19):6006–15. doi:10.1128/AEM. 01634-13.
15. Yukawa H, Omumasaba CA, Nonaka H, Kos P, Okai N, Suzuki N, et al. Comparative analysis of the Corynebacterium glutamicum group and complete genome sequence of strain R. Microbiology. 2007;153(Pt 4):1042–58. doi:10.1099/mic.0.2006/003657-0.
16. Lv Y, Liao J, Wu Z, Han S, Lin Y, Zheng S. Genome sequence of Corynebacterium glutamicum ATCC 14067, which provides insight into amino acid biosynthesis in coryneform bacteria. J Bacteriol. 2012;194(3):742–3. doi:10.1128/JB.06514-11.
17. Lv Y, Wu Z, Han S, Lin Y, Zheng S. Genome sequence of Corynebacterium glutamicum S9114, a strain for industrial production of glutamate. J Bacteriol. 2011;193(21):6096–7. doi:10.1128/JB.06074-11.
18. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595. doi:10.1186/1471-2105-11-595.
19. Xu M, Rao Z, Dou W, Xu Z. The role of ARGR repressor regulation on L-arginine production in Corynebacterium crenatum. Appl Biochem Biotechnol. 2013;170(3):587–97. doi:10.1007/s12010-013-0212-4.
20. Yun F, Zhou W (1994) Breeding and application of a strain of high glutamic acid yielding bacterium S9114 (in Chinese with English abstract). J South China University of Technol (Natural Science). 1994;22(1):56-62.