

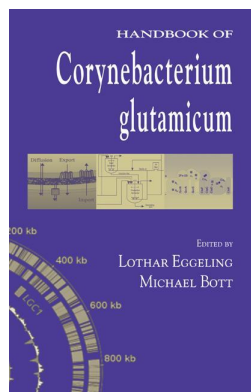
This article was downloaded by: 10.3.98.93

On: 19 Sep 2019

Access details: *subscription number*

Publisher: *CRC Press*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Corynebacterium glutamicum

Lothar Eggeling, Michael Bott

The Genomes of Amino Acid–Producing Corynebacteria

Publication details

<https://www.routledgehandbooks.com/doi/10.1201/9781420039696.pt3>

J. Kalinowski

Published online on: 30 Mar 2005

How to cite :- J. Kalinowski. 30 Mar 2005, *The Genomes of Amino Acid–Producing Corynebacteria* from: Handbook of Corynebacterium glutamicum CRC Press

Accessed on: 19 Sep 2019

<https://www.routledgehandbooks.com/doi/10.1201/9781420039696.pt3>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Part III

Genome, Plasmids, and Gene Expression

3 The Genomes of Amino Acid–Producing Corynebacteria

J. Kalinowski

CONTENTS

3.1	Introduction	37
3.2	Mapping and Sequencing of the <i>C. glutamicum</i> Genome	38
3.3	Sequencing Other Corynebacterial Genomes.....	39
3.4	Annotation of the <i>C. glutamicum</i> Genome	39
3.5	The Overall Structure of the <i>C. glutamicum</i> Genome	41
3.6	Prophages in the <i>C. glutamicum</i> Genome.....	44
3.7	The Gene Inventory of <i>C. glutamicum</i>	46
3.8	Comparative <i>Corynebacterium</i> Genome Analysis	47
3.9	Conclusions	51
	Acknowledgments.....	52
	References.....	53

3.1 INTRODUCTION

In the mid-1950s, Kinoshita and co-workers in Japan isolated a bacterium that was shown to excrete large quantities of L-glutamic acid into the culture medium [23]. This bacterium, now named *Corynebacterium glutamicum*, was described as a short, aerobic, Gram-positive rod capable of growing on a variety of sugars or organic acids. Following the first description as '*Micrococcus glutamicus*,' a number of different names were assigned to various isolates of *C. glutamicum*, e.g., '*Brevibacterium lactofermentum*,' '*Brevibacterium flavum*,' '*Brevibacterium divaricatum*,' and '*Corynebacterium lilium*.' This naming confusion was first clarified by the publication of Liebl et al. [26], who showed by modern taxonomic classification methods that the above-mentioned species designations were invalid and that all strains belong to the species *Corynebacterium glutamicum*. The type strain of this species is *C. glutamicum* ATCC 13032 (alternative designations: DSM 20300, IMET 10482, and NCIB 10025).

DNA sequencing of individual *C. glutamicum* genes started in the mid-1980s, when several genes from amino acid–biosynthetic pathways in *C. glutamicum* were

cloned and analyzed. These genes were mainly identified by heterologous complementation of *Escherichia coli* mutants and, occasionally, in the homologous system, e.g., by isolating genes of producer strains conferring resistance of a *C. glutamicum* strain against an amino acid analog [44] or by complementation of an export deficiency [46]. These studies already led to a general understanding of metabolic pathways, but a complete picture of the complex interactions could not be achieved owing to the lack of comprehensive genetic information. The sequencing of the complete corynebacterial genome turned out to represent the ideal method to obtain the missing genetic information and thus the basis for an efficient rational development of industrial producer strains.

This chapter is devoted to the description of the current knowledge on corynebacterial genomes. The genome sequences presented have been determined recently, and first analyses using bioinformatics have already provided interesting and valuable information for basic science as well as for industrial application. Because new genome information as well as novel bioinformatics tools are generated at high speed, a large number of fascinating discoveries on corynebacteria can be expected in the near future.

3.2 MAPPING AND SEQUENCING OF THE *C. GLUTAMICUM* GENOME

The first steps toward a complete genome sequence of *C. glutamicum* were made by determining the genome size and establishing a linked physical and genetic map [3]. The genome size was determined by pulsed-field gel electrophoresis of large DNA fragments generated by digestion with rare-cutting restriction enzymes. By this, the genome size was estimated to be 3.1 megabase pairs (Mbp). The large DNA fragments were arranged by Southern hybridizations using an established cosmid library of *C. glutamicum*. These studies revealed that the *C. glutamicum* ATCC 13032 genome consists of one circular chromosome.

Later, the cosmid library was used to generate an ordered minimal set of clones for the sequencing of the *C. glutamicum* genome by the same research team. In the study, 95 cosmid clones were arranged in a minimal set, but unfortunately the genome sequence coverage by cosmids was only 87% [42]. Therefore, an additional library in bacterial artificial chromosomes (BACs) was generated, and 21 of these BACs helped to cover the whole chromosome. With the higher resolution obtained in this mapping project, the estimated genome size was corrected to 3.28 Mbp.

In a collaborative research project led by the Degussa Company (Düsseldorf, Germany) and the Department of Genetics of Bielefeld University (Bielefeld, Germany), the 116 overlapping BAC and cosmid clones were sequenced individually by the shotgun method up to coverages of five- to ninefold and assembled by bioinformatics software. After finishing the cosmid and BAC sequences by PCR and primer-walking sequencing methods, contiguous high-quality sequences were available early in the project. The nucleotide sequences of the individual cosmid and BAC clones were finally assembled to a whole-genome sequence of 3,282,708 base pairs, harboring 3,002 potential genes [21].

The complete genome sequence of *C. glutamicum* ATCC13032 was determined in independent approaches by at least two other research teams. A Japanese team consisting of a collaboration between the Kyowa Hakko Company (Machida, Japan) and Kitasato University (Sagamihara, Japan) applied the whole-genome shotgun method based on plasmid and cosmid libraries [20]. These two libraries were sequenced to a five- to sixfold redundancy, and the sequences were assembled by bioinformatics software. The cosmid library was later used to close the sequence gaps resulting in a contiguous sequence of 3,309,401 base pairs and the identification of 3,099 genes. This genome sequence was the first that appeared in public databases.

A third project was carried out by the BASF company (Ludwigshafen, Germany) together with Integrated Genomics Inc. (Chicago, IL, U.S.A.). Information on how this project was carried out is scarce, but 2,900 genes were reported, of which 1,400 found entry into several patent applications. Because at the time of this writing only the first two complete genome sequences have found entry into public databases, the BASF/Integrated Genomics Project will not be discussed in this text.

3.3 SEQUENCING OTHER CORYNEBACTERIAL GENOMES

In parallel with the tight race between the different consortia that were sequencing *C. glutamicum* ATCC 13032 at almost the same time, the Ajinomoto Co. (Kawasaki, Japan) isolated a closely related species that was also sequenced. A special feature of this strain is its adaptation to higher temperatures around 40°C, a fact that led to its initial naming as '*Corynebacterium thermoaminogenes*.' A deeper taxonomic analysis of this soil-inhabiting and glutamate-producing *Corynebacterium* strain led to the proposal of a new species clearly distinct from *C. glutamicum*, named *C. efficiens* [13] (see Chapter 2).

The complete genome sequence of *C. efficiens* [33] was determined by the whole-genome shotgun method using two plasmid clone libraries with different insert size classes (0.8 to 1.2 kbp, 2.0 to 2.5 kbp). Sequences were assembled with standard bioinformatics software. The sequence was deposited in public databases and displays 3,147,090 bp for the main chromosome. In addition to the chromosome, the sequenced strain inherits two plasmids (Table 3.1). The GC content of the *C. efficiens* chromosome is unexpectedly high (63.4%), and 2,950 genes were predicted [33].

The other corynebacterial genome sequenced because of its great medical interest was that of *Corynebacterium diphtheriae*. This species is the causative agent of diphtheria, and the NCTC 13129 strain is a recent U.K. clinical isolate representative of an epidemic clone currently circulating in Eastern Europe. The genome is 2,488,635 bp long, with a GC content of 53.5% (Table 3.1), and was assembled from 66,099 sequencing reads [8].

3.4 ANNOTATION OF THE *C. GLUTAMICUM* GENOME

The initial step in genome annotation is gene finding, which can be carried out using a variety of bioinformatics tools that work reasonably well but at a certain level fail

TABLE 3.1
Features of *Corynebacterium* Genome Sequences

Strain	<i>C. glutamicum</i> ATCC13032 ^a	<i>C. glutamicum</i> ATCC13032 ^a	<i>C. efficiens</i> YS-314	<i>C. diphtheriae</i> NCTC13129
Accession number	BA000036/NC003450 ^b	BX927147	BA000035	BX248353
Data produced by	Kyowa Hakko and Kitasato University	Degussa AG and Bielefeld University	Ajinomoto	Sanger Institute
Size (bp)	3,309,401 ^a	3,282,708 ^a	3,147,090 23,743 (pCE1) 48,672 (pCE2)	2,488,635
Average G+C content (%)	53.8	53.8	63.4 54.4 (pCE1) 56.4 (pCE2)	53.5
Number of ORFs	3,099/2,993 ^b	3,002	2,950 15 (pCE1) 41 (pCE2)	2,320
Number of rRNA operons	6	6	5	5
Number of tRNAs	60	60	56	54
Coding regions (%)	86.8	88.3	90.4	89.6
Mean ORF length (bp)	933	916	979	964
Start codon usage (%)				
AUG	62.5	66.5	53.8	68.0
GUG	24.3	23.1	32.5	23.7
UUG	13.2	10.3	13.7	7.5
Other (pseudogenes)	—	0.1	—	0.8

^a The *C. glutamicum* strains are variants differing with respect to point mutations, insertion elements and a putative prophage (see text).

^b The submitted genome sequence (GenBank Acc. No. BA000036) was reannotated by the National Center for Biotechnology Information and deposited under Acc. No. NC003450.

to predict some genes or falsely predict nonexistent genes. The numbers of false-positives and false-negatives depend on the software tool used, its parameter settings, or training sets as well as on the composition of the genomic DNA to be analyzed. Therefore, it is preferable to use different tools for gene prediction in conjunction with visual inspection of each anticipated gene. This holds true also for the predicted gene starts, which in some instances have to be modified.

The *C. glutamicum* genome sequence established by the Degussa–Bielefeld University consortium was consistently annotated in the software GenDB [27]. Gene finding was performed by combining two bioinformatics tools: CRITICA [2] was used to define a gene set, which was subsequently used by GLIMMER [11] to construct a training model and to perform the final gene finding. This combination makes effective use of the selectivity of CRITICA (very few false-positives) and the sensitivity of GLIMMER (very few true-negatives). Genes were validated and coding sequence starts were checked by visual inspection after TBLASTN comparisons of the protein sequences deduced from all *C. glutamicum* ORFs against four other genome sequences from the *Actinomycetales* phylogenetic lineage, comprising *C. diphtheriae*, *C. efficiens*, *Mycobacterium tuberculosis*, and *Streptomyces coelicolor*.

The *C. glutamicum* genome sequence established by the Kyowa Hakko–Kitasato University consortium was first analyzed by automatic gene prediction using GLIMMER, and the results were intensively controlled and modified by manual intervention [20]. These researchers identified 3,099 putative protein-coding genes. The National Center for Biotechnology Information has independently searched the deposited sequence with another gene-finding tool, GeneMarkS [4], and has predicted 2,993 genes. It must be kept in mind that automatic prediction is by no means perfect and gene numbers in *C. glutamicum* may vary in the future between the reported extreme values of 2,900 and 3,100. Variations are also expected for other values deduced from the first annotation of a genome sequence, e.g., the genomic coverage by open reading frames, the mean ORF length, as well as the distribution of start codons depend heavily on the gene-finding strategies employed (Table 3.1).

For annotation, additional databases were used in each of the genome projects but only reported in detail for the Degussa–Bielefeld University project [21]. These authors used additional databases for gene function analysis, the nonredundant protein sequence database (nr), SWISSPROT, and INTERPRO, including several protein pattern databases. Additionally, SignalP [32] and TMHMM [24] were used to identify proteins that are potentially secreted or located in the cytoplasmic membrane, respectively.

3.5 THE OVERALL STRUCTURE OF THE *C. GLUTAMICUM* GENOME

As mentioned, two complete genomic sequences for the strain *C. glutamicum* ATCC 13032 are available to date. However, these genomes are not identical, a fact that is first reflected by the differing genome sizes, 3,282,708 bp and 3,309,401 bp. The difference of roughly 27 kbp in size is mainly due to additional copies of insertion elements and an additional putative prophage inserted in the larger genome (or deleted from the smaller one). It must be concluded that these mobile elements are capable of changing the *C. glutamicum* genome sequence by insertion or recombination in relatively short time periods, leading to two different genome sequences of the same ATCC 13032 strain.

The general features of the *C. glutamicum* genome sequences are shown in Table 3.1 and Figure 3.1. The *C. glutamicum* genome is represented by a circular

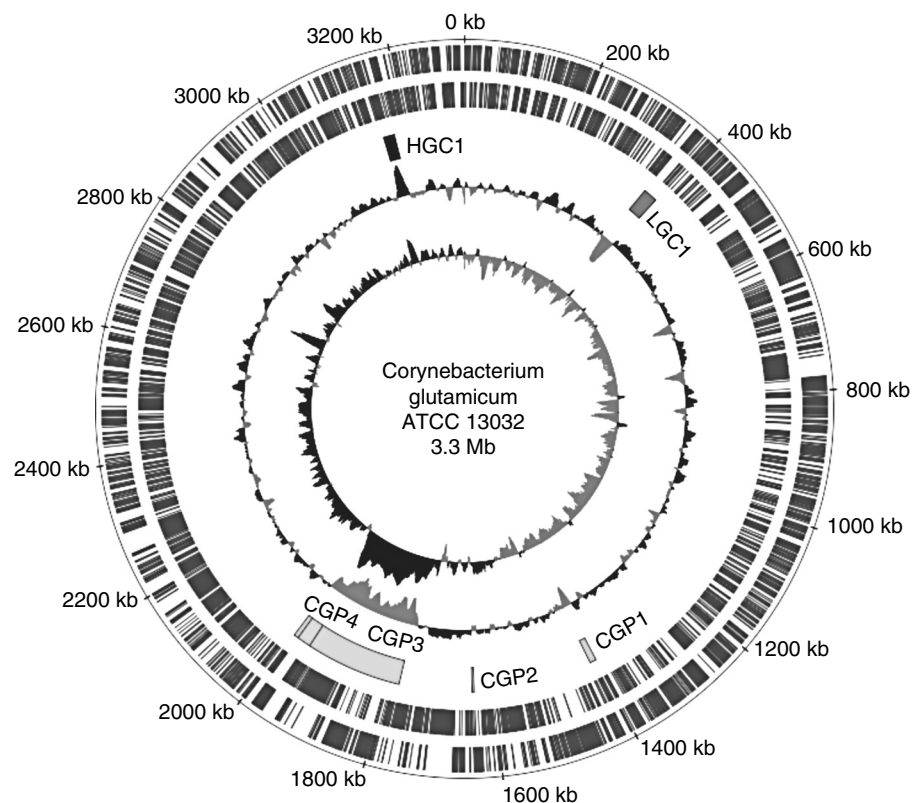


FIGURE 3.1 (Color insert follows page 208.) Circular representation of the *C. glutamicum* ATCC 13032 genome (GenBank Acc. No. NC003450). The plot was generated by GenDB version 2.0 [27]. Circles denote (outward to inward) the following: coding regions transcribed in clockwise and counterclockwise direction, respectively; GC content and GC skew. Bars pointing outward indicate values positively deviating from the median, and bars pointing inward indicate values negatively deviating from the median. The locations of low-GC and high-GC genomic regions as well as the prophages described in the text are represented by red, black, and green bars, respectively.

chromosome of 3.3 Mbp in size, which is a little larger than that of its close relative, *C. efficiens* (3.1 Mbp), and significantly larger than that of *C. diphtheriae* (2.5 Mbp). The G+C content of the genome is 53.8%, which is close to that of *E. coli* and at the lower boundary for the taxonomic class of the *Actinobacteria*, referred to as high-G+C Gram-positive bacteria. However, *C. diphtheriae* also has a similar G+C content of 53.5%, whereas the genome of *C. efficiens* has one of 63.4%. In contrast to the other corynebacterial genomes, *C. efficiens* inherits two plasmids, pCE1 and pCE2, which display G+C contents of 54.4 and 56.4%, respectively. These numbers are clearly different from that of the chromosome and might indicate a recent acquisition of these plasmids by *C. efficiens*.

The GC skew analysis [15], which is generally applicable to identify the leading and the lagging strand in DNA replication, indicated a bidirectional replication that

TABLE 3.2
Genomic Regions Differing in G+C Content and Prophages in *C. glutamicum*

Name	Location ^a	Size (kbp)	Genes ^b	Description
LGC1	NB 360.111-390.733	26.9	NCgl0335-0357 Cgl0342-0368	Low-G+C region containing genes involved in cell wall formation and lipopoly–saccharide synthesis
CGP1	CG 363.825-390.734 NB1.400.182-1.413.641	13.5	cg0414-0443 NCgl1281-1298 Cgl1336-1352	Putative incomplete prophage inserted at a tRNA-Leu carrying an integrase gene
CGP2	CG 1.401.510–1.415.069 NB 1.637.081-1.641.004	3.9	cg1507-1524 NCgl1490-1492 Cgl1548-1553	Putative prophage remnant carrying integrase and lysin genes
CGP3	CG 1.638.548–1.642.471 NB 1.776.613-1.995.294 ^c	(219.7) ^c	cg1746-1752 NCgl1611-1816 Cgl1675-1891	Putative prophage inserted at a tRNA-Val gene carrying integrase, primase, restriction/modification, and lysin genes
CGP4 ^c	CG 1.778.085–1.965.342 NB 1.963.136-1.986.590	187.3 23.5	cg1890-2071 NCgl1773-1806 Cgl1847-1881	Putative prophage inserted into CGP3 and carrying integrase, nuclease, single-strand–binding protein, and lysin genes. Terminal duplication of ca. 4.5 kb
HGC1	NB 3.156.304-3.176.905 CG 3.129.610-3.150.211	20.6	NCgl2851-2871 Cgl2952-2974 cg3267-3295	High-G+C region with a 7-kbp region highly conserved in <i>C. diphtheriae</i> encoding a putative copper transport and chaperone system

^a NB refers to GenBank Acc. Nos. NC003450 and BA000036, CG refers to BX927147

^b NCgl refers to NC003450, Cgl refers to BA000036, and cg refers to BX927147.

^c The CGP4 prophage is inserted into CGP3 and is only present in the NC003450/BA000036 genome sequence.

starts at the proposed *oriC* sequence near the *dnaA* gene and ends near the calculated replication terminus at around 1.6 Mbp (Figure 3.1). Note that several regions of the *C. glutamicum* genome deviate significantly in G+C content from the median (Figure 3.1, Table 3.2).

There is one region of 20 kbp in size located at around 3,150 kbp, which deviates significantly to a high-G+C content and was named HGC1. The genes of this region have G+C contents of up to 66% and are flanked by defective insertion sequences. The leftward 7 kbp of this region are more than 95% identical at the nucleotide level

to a segment from the *C. diphtheriae* genome and contain a putative copper transport system and a two-component sensor–regulator system. The extraordinarily high level of sequence conservation points to a recent horizontal gene transfer. This interpretation is supported by the fact that *C. efficiens*, which has a similar high mean G+C content, contains orthologs to the genes of this region that are conserved only on the protein sequence level. The puzzling question concerns the source organism in which the genes of the high-G+C region have evolved. Although *C. diphtheriae* contains a nearly identical gene region, the high-G+C content of this region is also exceptional for this organism. Therefore, parallel gene transfer events into both organisms, or two events with one of both being the source for the second gene transfer, must have occurred. This horizontal gene transfer might have been initiated by flanking insertion sequences and a possibly involved plasmid. Plasmids replicating in different *Corynebacterium* species have been reported [43] (see Chapter 4).

In contrast to the HGC1 region's being exceptional in having a higher G+C content, a number of genomic regions were identified as exceptionally deficient in GC (Figure 3.1, Table 3.2). One of these regions is located at around 380 kbp, has a size of approximately 27 kbp, and covers around 20 coding regions with G+C contents of 41 to 49% (LGC1). The genes that are located in this region are involved in some aspects of murein formation, e.g., *murA* and *murB*, and lipopolysaccharide synthesis. It is interesting to note that two genes encoding the enzymes for the initial steps in murein formation are duplicated in *C. glutamicum*. Whereas *murA* and *murB* located in this region have a low-G+C content of 44%, the second copy of these genes (*murA2*, *murB2*) exhibit a G+C content typical for *C. glutamicum* ORFs. Also in other organisms, exceptionally low-G+C regions carrying genes involved in cell wall and cellular surface formation have been found (e.g., *Bacillus subtilis*), indicating a preferred horizontal gene transfer or special selective benefit from receiving such functions by horizontal gene transfer. This is also the case in the so-called pathogenicity islands, which often confer functions involved in bacterial surface modifications [25].

Horizontal gene transfer is generally mediated by transposable elements, plasmids, and bacteriophages. In fact, at least 24 insertion elements are present in the *C. glutamicum* genome [21]. These elements are frequently found at the borders of regions of unusual G+C content, e.g., the HGC1 region (Table 3.2).

3.6 PROPHAGES IN THE *C. GLUTAMICUM* GENOME

Ubiquitous to bacterial genomes are bacteriophages, and their genomically integrated forms are referred to as *prophages* [6]. Integration of prophages into a bacterial genome is generally recognizable by a discontinuity in the DNA composition (mean G+C, GC skew). Prophages are diverse in size and are found in bacterial genomes in various stages of degeneration.

In the search for prophages in the *C. glutamicum* genome, the criteria for their identification were the above-mentioned irregularities in G+C content or GC skew, the presence of genes with bacteriophage homologs of known function (e.g., integrases), or the presence of gene regions lacking homologs, especially in the closely related strains (*C. efficiens*, *C. diphtheriae*).

Integrase genes (*int*) encoding the vital enzyme responsible for integration of the bacteriophage genome into the genome of its host are particularly easy to detect by similarity searches. Altogether, five hypothetical phage integrase genes (*'int*, *int1*, *int2'*, *'int2*, *int3*) were found in the Degussa–Bielefeld University sequence, with *'int* being partially deleted and *int2* appearing to be disrupted by a frame-shift mutation. A closer inspection revealed a large low-G+C region, including the *int2* gene fragments as the putative *C. glutamicum* prophage CGP3, whereas the *int1* gene is part of the CGP1 putative prophage and the *'int* gene might be part of another defective prophage named CGP2 (Table 3.2). The CGP3 element present in the larger genome sequence is extended by approximately 20 kbp and displays a 5-kbp duplicated gene region containing a second *'int2* gene at its end. The putative integrase gene *int3* is not flanked by genes fulfilling the above-mentioned criteria for a prophage origin.

The largest prophage region CGP3 spans more than 180 kbp. It covers approximately 200 coding regions, most of which lack any significant similarities to known bacterial genes. However, there are a few exceptions, including the three genes of the already known restriction-modification system (*cglIM*, *cglIR*, *cglIIR*) [39], genes encoding transposases, putative recombination enzymes, and a number of homologs to known bacteriophage proteins, especially a phage primase and the phage-type integrase *'int2*. It is interesting to note that the left border of this region is formed by a cluster of tRNA genes, whereas the phage-type integrase gene is located near the right border of the insertion. These observations might be explained by the integration of one or more prophage-like elements at a specific tRNA locus, a mechanism that is common for phages and integrative plasmids [47].

The putative prophage region CGP3 represents the major difference in the two genome sequences. In the left flank of this region, the Kyowa Hakko–Kitasato University sequence has a large insertion of another prophage (CGP4) and a duplicated gene region containing the committed phage-type integrase gene. The assumption that this *C. glutamicum* ATCC 13032 genome inherits an additional prophage is supported by the fact that the integrated region carries another serine protease as a putative cellulytic enzyme. An alternative explanation is that the unique fragment of the original prophage has been deleted in the Degussa–Bielefeld University sequence.

The potential prophages found in *C. glutamicum* are diverse in size. Whereas the CGP3 element is larger than most known prophages, the smaller putative prophages CGP1 and CGP2 are presumably highly degenerated remnants of former functional bacteriophages. In the case of CGP1 and CGP3, the place of insertion is a tRNA gene, and the insertion site sequence is detectable as a direct repeat flanking the element. A flanking tRNA gene and direct repeats are missing in case of the apparently highly degenerated CGP2 element.

The origin of these prophages is not known. Isolations of bacteriophages infecting different *C. glutamicum* strains have been described several times [29,35,40,45], but no report on the successful induction of a prophage from strain ATCC 13032 is available. Also, nucleotide sequences of genes from these corynephages were obtained only in a few cases, thus leaving the relationship between the presumed prophages and known infectious corynebacterial phages unclear.

3.7 THE GENE INVENTORY OF *C. GLUTAMICUM*

From a soil bacterium, it can be expected that its genome has to encode all necessary functions for primary metabolism, for catabolism of a wide variety of nutrients, and for optimal adaptation to changes in the environment. *Corynebacterium glutamicum* was the first completely sequenced Gram-positive soil bacterium from the *Corynebacteriaceae*. The other members of this group whose genomes are known are *C. efficiens*, *C. diphtheriae*, *Mycobacterium tuberculosis*, *M. leprae*, *M. bovis*, and *M. marinum* — most of them important pathogens. Since nonpathogenic model systems are necessary, *C. glutamicum* may serve as an ideal system for studying the cell wall structure and synthesis and, especially, mycolic acid synthesis.

Corynebacterium glutamicum is capable of growing in a simple mineral salts medium; i.e., it is able to synthesize from simple precursors all cell constituents, including metabolites, cofactors, and vitamins, except for D-biotin. This defect is most probably due to the fact that the gene *bioF*, encoding the biotin biosynthetic enzyme 7-keto-8-aminopelargonic acid synthetase, is missing in *C. glutamicum* [17,18].

All of the genes already described for various *C. glutamicum* strains and represented as nucleotide sequences in public databases were found also in the genome sequences, with one important exception. The gene for the para-crystalline surface-layer protein *cspB* [36] from *C. glutamicum* ATCC 17965, which is synthesized in extremely large amounts and has a possible function in protecting the bacterium in soil against rough conditions, is missing in both *C. glutamicum* ATCC 13032 sequences. It is not clear why this gene is absent, but it can be speculated that bacterial strains in laboratories adapt to the specific growth conditions by losing functions that provide a heavy metabolic load and do not confer any advantage under optimal growth conditions [14]. Because also the immediate *cspB* flanking regions are not represented in the *C. glutamicum* ATCC 13032 sequence, it is impossible to determine its original place in the genome.

The conjunction of automated and manual annotation of coding regions by similarities to known genes in public databases helped to annotate 82% of the *C. glutamicum* ORFs. However, annotation sometimes consisted only of a global functional characterization based on minor similarities with other ill-characterized genes or proteins. By accepting the huge bandwidth in functional assignments of *C. glutamicum* genes, only a small fraction of genes remained hypothetical (9%), meaning not similar to a database entry, or conserved hypothetical (9%), meaning similar to a hypothetical protein from another organism. In the case of a conserved hypothetical protein, the assumption is that it represents a real gene, whereas the hypothetical genes must be confirmed by further studies, such as by proteome analysis.

Annotated genes can be assigned to functional classes with the widely used COG (cluster of orthologous groups) system [41]. Figure 3.2 shows that about 500 genes of *C. glutamicum* fall into the COG categories S (function unknown) and R (general function prediction only). About 900 genes cannot be classified by the COG system. However, the numbers of the members within the other classes are more or

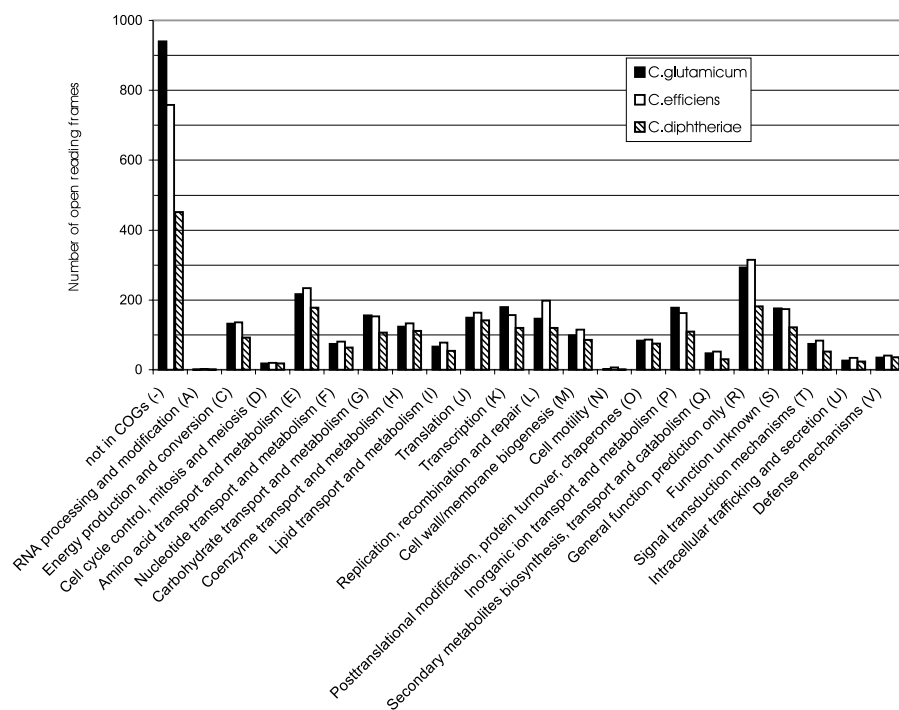


FIGURE 3.2 Functional classification of proteins from *C. glutamicum*, *C. efficiens*, and *C. diphtheriae*. The assignments have been made by using the COG functional classification scheme. The bars represent the numbers of genes of a functional class in a genome.

less the same in all three corynebacterial genomes, differing only in the three categories mentioned. The ill-characterized genes, which cannot be classified, might be interpreted as additional ones that carry probably nonessential genetic information whereas a common core of around 1,600 genes is found in all three genomes.

Expert-manual annotation already provided a deeper understanding of gene function and helped to reconstruct most parts of the central metabolism, starting from sugar consumption and ending with produced amino acids [21]. Other functional complexes have also been analyzed and reconstructed using the information provided by the genome projects, e.g., in the excellent review on the respiratory chain by Bott and Niebisch [5]. Further reconstructions of other parts of metabolism can be expected in the near future.

3.8 COMPARATIVE CORYNEBACTERIUM GENOME ANALYSIS

Up to now, four genome sequences from corynebacteria are available. These sequences represent the three species *C. glutamicum*, *C. efficiens*, and *C. diphtheriae*, of which *C. glutamicum* and *C. efficiens* represent both natural producers of

L-glutamic acid. Since *C. diphtheriae* is a human pathogen and more distantly related to the biotechnologically relevant corynebacteria, it is especially interesting to compare the *C. glutamicum* and *C. efficiens* genomes. This comparison is of even higher relevance because both species differ by approximately 10% in their overall G+C content. Irrespective of this, the amino acid sequences of the proteins are fairly well conserved, facilitating the discrimination between protein-coding and noncoding DNA.

Nishio and co-workers [33] have studied the connections among overall G+C content, codon preference, and phenotype in *C. efficiens* in comparison to *C. glutamicum*. These authors discovered that the differences in G+C content between both genomes were accompanied by a significant bias in amino acid substitutions. Three major substitutions were identified in *C. efficiens*, from lysine to arginine, serine to alanine, and serine to threonine. These substitutions are suggested to be the cause of the higher thermostability of *C. efficiens* proteins and the organism's attribute for growing at temperatures above 40°C.

Although some of the numbers deduced from the genome sequences are dependent on the gene-finding strategies, a rough comparison can be made. *Corynebacterium glutamicum* and *C. efficiens* have comparable genome sizes and gene numbers (Table 3.1). The numbers of six (*C. glutamicum*) and five rRNA operons (*C. efficiens*) of the order 16S-23S-5S and of 60 and 56 tRNA genes, respectively, are typical for fast-growing environmental bacteria. However, *C. diphtheriae* as a human pathogen displays similar figures.

Other genome features, such as the fractional genomic coverage by ORFs and the mean ORF length, are also rather similar and close to the values determined for many other bacterial genomes. In contrast to this, the usage of start codons is affected by the G+C content. Therefore, *C. efficiens* has a slightly higher fraction of ORFs starting with GUG in comparison to the two other corynebacterial species.

An evaluation of gene-order conservation revealed an astonishing degree of synteny between all three *Corynebacterium* species (Figure 3.3). A possible reason for this unusual genome stability is given by Nakamura and co-workers [30]. These researchers found out that corynebacteria did not contain *recBCD* genes, encoding the recombinational repair system. It was suggested that the absence of this system prevented gene shuffling and retained an ancestral gene order in corynebacteria.

The gene order analysis performed on the three genomes also clearly showed that the putative bacteriophage insertion regions are of alien origin. Additionally, there are several smaller regions carrying genes with no counterpart in one of the other species. A closer inspection of these regions might reveal genes that are also either horizontally transferred or are necessary only in a certain ecological niche. An example for this is the LGC1 region, which carries genes possibly involved in lipopolysaccharide synthesis, having their closest homologs in Gram-negative pathogens such as *Neisseria* and *Campylobacter*.

In addition to putative prophages and the large number of different insertion elements that are present in both genomes, further differences in gene content between the two soil-inhabiting corynebacteria were detected. Only slight differences are detectable between the *C. glutamicum* and *C. efficiens* genomes if COG classes are compared (Figure 3.3). The differences are found mainly in the unclassified

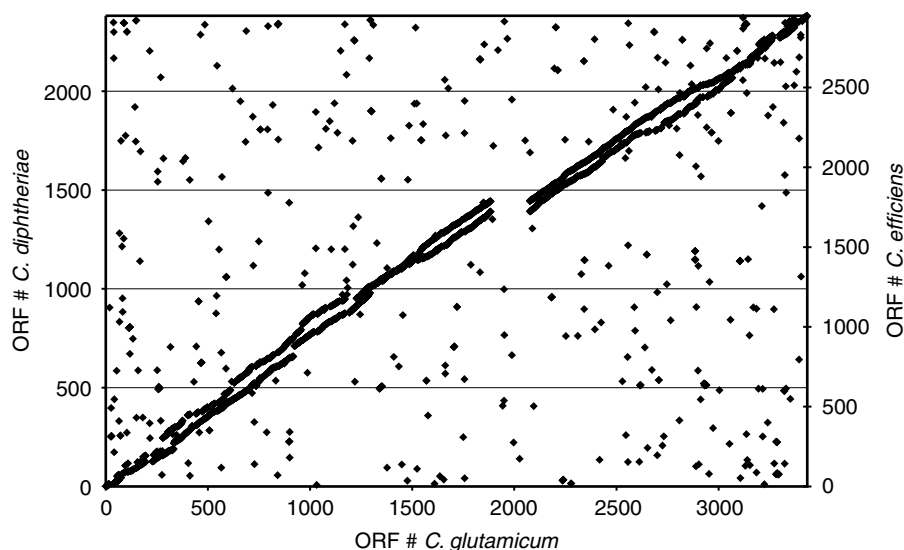


FIGURE 3.3 (Color insert follows page 208.) Gene order comparison for three corynebacterial genomes. The dots give the results of a bidirectional best hit generated by comparing the amino acid sequences with BLASTP [1]. The gene numbering refers to the BX927147 sequence.

genes, confirming the notion that they might carry nonessential information. However, it must be noted that *C. glutamicum* contains a larger number of transcriptional regulator genes (class K), whereas the *C. efficiens* genome is richer in transposases (class L).

A more detailed analysis revealed a number of genes from *C. glutamicum* ATCC 13032 where the functions are known to be due to detailed characterization of mutant strains and functional analyses of the encoded proteins and that are absent from the *C. efficiens* YS-314 genome (Table 3.3). Significant is the missing of the *brnEF*

TABLE 3.3
C. glutamicum* Genes Encoding Known Functions without Homologues in *C. efficiens

Gene	Function of Gene Product	Reference
<i>brnE</i>	Component of a branched-chain amino acid exporter	[22]
<i>brnF</i>	Component of a branched-chain amino acid exporter	
<i>lrp</i>	transcriptional Regulator, Lrp family	
<i>cma</i>	Cyclopropane-fatty-acyl-phospholipid synthase	[31]
<i>prpD2</i>	Component of the methylcitrate cycle for propionate degradation	[7]
<i>prpB2</i>	Component of the methylcitrate cycle for propionate degradation	
<i>prpC2</i>	Component of the methylcitrate cycle for propionate degradation	
<i>porA</i>	Porin	[9]

genes encoding the components of a branched-chain amino acid exporter [22]. This transporter was shown to export the amino acids L-isoleucine, L-leucine, and L-valine. In addition, the adjacent *lrp* gene encoding a transcriptional regulator, is missing also in *C. efficiens*. Whether this means that *C. efficiens* is not capable of excreting branched-chain amino acids should be clarified by experimental studies.

A second gene cluster missing in *C. efficiens* is the *prpDBC2* cluster, whereas the *prpDBC1* cluster is present. The function of the latter cluster is unknown, whereas the *prpDBC2* gene cluster has been shown to encode the enzymes of the methylcitrate cycle essential for propionate degradation in *C. glutamicum* [7]. In agreement with the absence of *prpDBC2* in *C. efficiens* is the characterization of this species by Fudou and co-workers [13] showing that *C. efficiens* is unable to degrade propionate as carbon source.

The *cma* gene possibly encoding a mycolic acid cyclopropane synthase [31] is also missing in *C. efficiens*. Cyclopropanation of mycolic acids is a general feature of mycobacteria. At present, it is not known whether *C. glutamicum* possesses cyclopropanated mycolic acids, but the difference with respect to the *cma* gene might result in a difference in mycolic acid structure between the two species.

Another structural difference with respect to the outer cell layers is also implicated by the missing of a clear *porA* gene homologue in *C. efficiens*. The *porA* gene encodes the major porin of *C. glutamicum* [9], which forms a channel necessary to allow the transport of hydrophilic substrates through the highly hydrophobic outer layer of *C. glutamicum*. Because it can be expected that the *C. efficiens* surface is hydrophobic enough to require the presence of a porin, an unrelated porin gene might be present in *C. efficiens*. Such a gene is probably not detected by automated gene-finding methods owing to the small size of the protein (PorA: 45 amino acids).

In comparison to *C. glutamicum*, much less is known on the biology and the biochemistry of *C. efficiens*. Therefore, the *C. efficiens* genome is annotated in a rather conservative way. However, the comparison of both genomes clearly identified a number of *C. efficiens* genes without homologs in *C. glutamicum* ATCC 13032. Among these are the examples displayed in Table 3.4.

It is striking that *C. efficiens* carries gene equipment for the degradation of a number of aromatic compounds from biological sources. One example is the gene cluster for the CoA-ligase and the monooxygenase enzymes degrading phenylacetic acid (CE0663-CE0672). In addition, the enzymes for the release of inorganic sulfate from aromatic sulfate esters (arylsulfatase, CE2198) as well as for the degradation of phenols (tyrosinase, CE1756) and amides (formamidase, CE2198) seem to be present. However, tests for tyrosinase enzyme activity proved negative for *C. efficiens* [13], indicating that all hypotheses deduced via bioinformatics have to be verified by additional experimental approaches. However, disagreements between predicted and observed phenotypes might result from incorrect annotations based on sequence similarity but also from possible difficulties in finding the right parameters for the expression of a regulated gene.

Beside degradative functions, interesting biosynthetic functions are predicted to exist in *C. efficiens*. By judging from sequence similarities, the genes CE1202 and CE1203 might encode the subunits of a cellulose synthase. Up to now, such enzymes

TABLE 3.4
Examples of Annotated *C. efficiens* Genes without Homologs in *C. glutamicum*

Gene Number	Annotated Function	Comment
CE0663-CE0672	Phenylacetic acid degradation gene cluster	Usage of phenylacetate as carbon source
CE1202, CE1203 CE1568	Subunits of cellulose synthase Arylsulfatase	Cellulose synthesis Usage of aromatic sulfate esters as sulfur source
CE1756	Tyrosinase	Degradation of monophenols
CE2198	Formamidase	Usage of amides as nitrogen source
CE2362, CE2363	Ribonucleotide reductase type III	Enzyme only active under anaerobic conditions
CE2454-CE2458 CE2737-CE2741	Fimbrial proteins and export system	Important for attachment to eukaryotic cells and biofilms

have been found only in some Gram-negative bacteria [37], and it will be fascinating to find out whether *C. efficiens* is able to produce cellulose.

Additional genes might give predictions on the habitat and lifestyle of *C. efficiens*. Examples of such genes are CE2454–CE2458 and CE2737–2741 predicted to encode fimbriae proteins as well as their specific transport mechanisms. Fimbriae are often found in bacteria of medical relevance and are important for attachment to eukaryotic cells. In addition, they might be needed for the buildup of bacterial biofilms that also occur in the environment [10]. Another example is the possession of a type III ribonucleotide reductase encoded by CE2362 and CE2363. Ribonucleotide reductases of type III are oxygen sensitive and found exclusively in anaerobic or facultatively anaerobic microorganisms [12]. *Corynebacterium efficiens* possesses an additional oxygen-insensitive ribonucleotide reductase of type 1b (encoded by *nrdF* and *nrdE*), very similar to that of *C. glutamicum*. This is in agreement with the strain description for *C. efficiens* as being capable of life under aerobic and facultatively anaerobic conditions [13].

The genes mentioned in the examples given above as well as other genes encoding interesting functions in *C. efficiens* now await experimental proof. Therefore, genetic engineering techniques for *C. efficiens* should be developed or *C. glutamicum* could be used as a model system in heterologous expression experiments. However, owing to their much higher GC-content and differences in codon usage [33], expression of *C. efficiens* genes in *C. glutamicum* is not guaranteed.

3.9 CONCLUSIONS

The establishment of complete annotated genome sequences of *Corynebacterium* strains is a big leap forward in the understanding of these organisms and will boost genetic analyses as well as metabolic engineering to overproduce compounds of commercial relevance.

The *C. glutamicum* genome sequence has already helped to directly identify missing genes along close biosynthetic pathways of interest or to provide a limited number of candidate genes for testing. An example of this is the work of Hartmann and co-workers [16], in which the genes *dapF* and *dapC* completing the lysine biosynthetic pathway have been successfully located. Another example is provided by Rückert and co-workers [38], who studied the biosynthesis of methionine. In this study, the *C. glutamicum* genome was first scanned for candidate genes involved in methionine biosynthesis. Then, by generating deletion mutant strains for all candidate genes and auxanography, a complete branched pathway for L-methionine synthesis was reconstructed.

Another issue for which a genome sequence is essential is the discovery of novel or unexpected functions. An example is the discovery of two genes encoding carbonic anhydrases in this organism [28] or the discovery of a second glyceraldehyde-3-dehydrogenase gene, which is expressed exclusively under gluconeogenic conditions [19]. As mentioned before, the *C. efficiens* genome awaits the development of genetic engineering techniques to prove some fascinating hypotheses deduced via bioinformatics.

The information generated by the joint application of high-quality bioinformatics and broad biochemical and biotechnological knowledge — as well as genetic engineering will allow the creation of highly efficient production strains for well-established and for novel products. Of particular importance with respect to production strain development is the genome analysis of the highly efficient production strains developed traditionally. Through sequence comparison of alleles relevant for the production process with their wild-type counterparts and subsequent introduction of the identified mutations into a wild-type strain, a minimally mutated production strain with improved properties can be derived. This technique of “genome breeding” was introduced by Ohnishi et al. [34] with the development of an efficient L-lysine-producing strain of *C. glutamicum* by the introduction of only three different mutations. The engineered strain reached a high production level but showed far better growth than the traditionally developed production strain, which allows fermentation times to be shortened by about 50%.

The complete genome sequence also forms the basis for most methods of global expression analyses, e.g., proteome and transcriptome studies. Such analyses will lead to a comprehensive systemic understanding of gene expression and regulatory networks in corynebacteria in the future.

ACKNOWLEDGMENTS

The help of Christian Rückert, Alexander Goesmann, Burkhard Linke, Oliver Rupp, Alice McHardy, and Daniela Bartels during various steps of bioinformatics analyses and generation of figures is thankfully acknowledged.

REFERENCES

1. Altshul SF, Gish W, Miller W, Myers EW, and Lipman D. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
2. Badger JH and Olsen GJ. (1999) CRITICA: Coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.* 16:512–524.
3. Bathe B, Kalinowski J, and Pühler A. (1996) A physical and genetic map of the *Corynebacterium glutamicum* ATCC 13032 chromosome. *Mol. Gen. Genet.* 252:255–265.
4. Besemer J, Lomsadze A, and Borodovsky M. (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29:2607–2618.
5. Bott M and Niebisch A. (2003) The respiratory chain of *Corynebacterium glutamicum*. *J. Biotechnol.* 104:129–153.
6. Casjens S. (2003) Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* 49:277–300.
7. Claes WA, Pühler A, and Kalinowski J. (2002) Identification of two *prpDBC* gene clusters in *Corynebacterium glutamicum* and their involvement in propionate degradation via the 2-methylcitrate cycle. *J. Bacteriol.* 184:2728–2739.
8. Cerdeno-Tarraga AM, Efstratiou A, Dover LG, Holden MT, Pallen M, Bentley SD, Besra GS, Churcher C, James KD, De Zoysa A, Chillingworth T, Cronin A, Dowd L, Feltwell T, Hamlin N, Holroyd S, Jagels K, Moule S, Quail MA, Rabinowitsch E, Rutherford KM, Thomson NR, Unwin L, Whitehead S, Barrell BG, and Parkhill J. (2003) The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res.* 31:6516–6523.
9. Costa-Riu N, Burkovski A, Krämer R, and Benz R. (2003) PorA represents the major cell wall channel of the gram-positive bacterium *Corynebacterium glutamicum*. *J. Bacteriol.* 185:4779–4786.
10. Dalton HM and March PE. (1998) Molecular genetics of bacterial attachment and biofouling. *Curr. Opin. Biotechnol.* 9:252–255.
11. Delcher AL, Harmon D, Kasif S, White O, and Salzberg SL. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27:4636–4641.
12. Fontecave M, Mulliez E, and Logan DT. (2002) Deoxyribonucleotide synthesis in anaerobic microorganisms: the class III ribonucleotide reductase. *Prog. Nucleic Acid Res. Mol. Biol.* 72:95–127.
13. Fudou R, Jojima Y, Seto A, Yamada K, Kimura E, Nakamatsu T, Hiraishi A, and Yamanaka S. (2002) *Corynebacterium efficiens* sp. nov., a glutamic-acid-producing species from soil and vegetables. *Int. J. Syst. Evol. Microbiol.* 52:1127–1131.
14. Fujita M, Moriya T, Fujimoto S, Hara N, and Amoko K. (1997) A deletion in the *sapA* homologue cluster is responsible for the loss of the S-layer in *Campylobacter fetus* strain TK. *Arch. Microbiol.* 167:196–201.
15. Grigoriev A. (1998) Analyzing genomes with cumulative skew diagrams. *Bioinformatics* 14:252–258.
16. Hartmann M, Tauch A, Eggeling L, Bathe B, Möckel B, Pühler A, and Kalinowski J. (2003) Identification and characterization of the last two unknown genes, *dapC* and *dapF*, in the succinylase branch of the L-lysine biosynthesis of *Corynebacterium glutamicum*. *J. Biotechnol.* 104:199–211.

17. Hatakeyama K, Kohama K, Vertes AA, Kobayashi M, Kurusu Y, and Yukawa H. (1993) Analysis of the biotin biosynthesis pathway in coryneform bacteria: cloning and sequencing of the *bioB* gene from *Brevibacterium flavum*. *DNA Seq.* 4:87–93.
18. Hatakeyama K, Kohama K, Vertes AA, Kobayashi M, Kurusu Y, and Yukawa H. (1993) Genomic organization of the biotin biosynthetic genes of coryneform bacteria: Cloning and sequencing of the *bioA-bioD* genes from *Brevibacterium flavum*. *DNA Seq.* 4:177–184.
19. Hayashi M, Mizoguchi H, Shiraishi N, Obayashi M, Nakagawa S, Imai J, Watanabe S, Ota T, and Ikeda M. (2002) Transcriptome analysis of acetate metabolism in *Corynebacterium glutamicum* using a newly developed metabolic array. *Biosci. Biotechnol. Biochem.* 66:1337–1344.
20. Ikeda M and Nakagawa S. (2003) The *Corynebacterium glutamicum* genome: Features and impacts on biotechnological processes. *Appl. Microbiol. Biotechnol.* 62:99–109.
21. Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Krämer R, Linke B, McHardy AC, Meyer F, Möckel B, Pfefferle W, Pühler A, Rey DA, Rückert C, Rupp O, Sahl H, Wendisch VF, Wiegräbe I, and Tauch A. (2003) The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J. Biotechnol.* 104:5–25.
22. Kennerknecht N, Sahl H, Yen MR, Patek M, Saier MH Jr, and Eggeling L. (2003) Export of L-isoleucine from *Corynebacterium glutamicum*: a two-gene-encoded member of a new translocator family. *J. Bacteriol.* 184:3947–3956.
23. Kinoshita S, Udaka S, and Shimono M. (1957) Studies on the amino acid fermentation. I. Production of L-glutamic acid by various microorganisms. *J. Gen. Appl. Microbiol.* 3:193–205.
24. Krogh A, Larsson B, von Heijne G, and Sonnhammer EL. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305:567–580.
25. Lee CA. (1996) Pathogenicity islands and the evolution of bacterial pathogens. *Infect. Agents Dis.* 5:1–7.
26. Liebl W, Ehrmann M, Ludwig W, and Schleifer KH. (1991) Transfer of *Brevibacterium divaricatum* DSM 20297^T, "*Brevibacterium flavum*" DSM 20411, "*Brevibacterium lactofermentum*" DSM 20412 and DSM 1412, and *Corynebacterium lilium* DSM 20137^T to *Corynebacterium glutamicum* and their distinction by rRNA gene restriction patterns. *Int. J. Syst. Bacteriol.* 41:255–260.
27. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, and Pühler A. (2003) GenDB: an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 31:2187–2195.
28. Mitsuhashi S, Ohnishi J, Hayashi M, and Ikeda M. (2004) A gene homologous to β -type carbonic anhydrase is essential for growth of *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 63:592–601.
29. Moreau S, Leret V, Le Marrec C, Varangot H, Ayache M, Bonnassie S, Blanco C, and Trautwetter A. (1995) Prophage distribution in coryneform bacteria. *Res. Microbiol.* 146:493–505.
30. Nakamura Y, Nishio Y, Ieko K, and Gojobori T. (2003) The genome stability in *Corynebacterium* species due to lack of the recombinational repair system. *Gene* 317:149–155.

31. Nampoothiri KM, Hoischen C, Bathe B, Möckel B, Pfefferle W, Krumbach K, Sahm H, and Eggeling L. (2002) Expression of genes of lipid synthesis and altered lipid composition modulates L-glutamate efflux of *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 58:89–96.
32. Nielsen H, Engelbrecht J, Brunak S, and von Heijne G. (1999) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int. J. Neural. Sys.* 8:581–599.
33. Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, Sugimoto S, Matsui K, Yamagishi A, Kikuchi H, Ikeo K, and Gojobori T. (2003) Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. *Genome Res.* 13:1572–1579.
34. Ohnishi J, Mitsuhashi S, Hayashi M, Ando S, Yokoi H, Ochiai K, and Ikeda M. (2002) A novel methodology employing *Corynebacterium glutamicum* genome information to generate a new L-lysine-producing mutant. *Appl. Microbiol. Biotechnol.* 58:217–223.
35. Patek M, Ludvik J, Benada O, Hochmannova J, Nesvera J, Krumphanz V, and Bucko M. (1985) New bacteriophage-like particles in *Corynebacterium glutamicum*. *Virology* 140:360–363.
36. Peyret JL, Bayan N, Joliff G, Gulik-Krzywicki T, Mathieu L, Shechter E, and Leblon G. (1993). Characterization of the *cspB* gene encoding PS2, an ordered surface-layer protein in *Corynebacterium glutamicum*. *Mol. Microbiol.* 9:97–109.
37. Ross P, Mayer R, and Benziman M. (1991) Cellulose biosynthesis and function in bacteria. *Microbiol Rev.* 55:35–58.
38. Rückert C, Pühler A, and Kalinowski J. (2003) Genome-wide analysis of the L-methionine biosynthetic pathway in *Corynebacterium glutamicum* by targeted gene deletion and homologous complementation. *J. Biotechnol.* 104:213–228.
39. Schäfer A, Tauch A, Droste N, Pühler A, and Kalinowski J. (1997) The *Corynebacterium glutamicum cglIM* gene encoding a 5-cytosine methyltransferase enzyme confers a specific DNA methylation pattern in an McrBC-deficient *Escherichia coli* strain. *Gene* 203:95–101.
40. Sonnen H, Schneider J, and Kutzner HJ. (1990) Coryneophage Cog, a virulent bacteriophage of *Corynebacterium glutamicum*, and its relationship to phi GA1, an inducible phage particle from *Brevibacterium flavum*. *J. Gen. Virol.* 71:1629–1633.
41. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, and Koonin EV. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29:22–28.
42. Tauch A, Homann I, Mormann S, Rüberg S, Billault A, Bathe B, Brand S, Brockmann-Gretza O, Rückert C, Schischka N, Wrenger C, Hoheisel J, Möckel B, Huthmacher K, Pfefferle W, Pühler A, and Kalinowski J. (2002a) Strategy to sequence the genome of *Corynebacterium glutamicum* ATCC 13032: use of a cosmid and a bacterial artificial chromosome library. *J. Biotechnol.* 95:25–38.
43. Tauch A, Kirchner O, Löffler B, Götter S, Pühler A, and Kalinowski J. (2002b) Efficient electrotransformation of *Corynebacterium diphtheriae* with a mini-replicon derived from the *Corynebacterium glutamicum* plasmid pGA1. *Curr. Microbiol.* 45:362–367.
44. Thierbach G, Kalinowski J, Bachmann B, and Pühler A. (1990) Cloning of a DNA fragment from *Corynebacterium glutamicum* conferring aminoethyl cysteine resistance and feedback resistance to aspartokinase. *Appl. Microbiol. Biotechnol.* 32:443–448.

45. Trautwetter A, Blanco C, and Sicard AM. (1987) Structural characteristics of the *Corynebacterium lilium* bacteriophage CL31. *J. Virol.* 61:1540–1545.
46. Vrljic M, Sahm H, and Eggeling L. (1996) A new type of transporter with a new type of cellular function: L-lysine export from *Corynebacterium glutamicum*. *Mol. Microbiol.* 22:815–826.
47. Williams KP. (2002) Integration sites for genetic elements in prokaryotic tRNA and TmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.* 30:866–875.