Complete genome sequence of *Corynebacterium variabile* DSM 44702 isolated from the surface of smear-ripened cheeses and insights into cheese ripening and flavor generation

Jasmin Schröder¹, Irena Maus¹, Eva Trost¹² and Andreas Tauch¹*

Abstract

**Background:** *Corynebacterium variabile* is part of the complex microflora on the surface of smear-ripened cheeses and contributes to the development of flavor and textural properties during cheese ripening. Still little is known about the metabolic processes and microbial interactions during the production of smear-ripened cheeses. Therefore, the gene repertoire contributing to the lifestyle of the cheese isolate *C. variabile* DSM 44702 was deduced from the complete genome sequence to get a better understanding of this industrial process.

**Results:** The chromosome of *C. variabile* DSM 44702 is composed of 3,433,007 bp and contains 3,071 protein-coding regions. A comparative analysis of this gene repertoire with that of other corynebacteria detected 1,534 predicted genes to be specific for the cheese isolate. These genes might contribute to distinct metabolic capabilities of *C. variabile*, as several of them are associated with metabolic functions in cheese habitats by playing roles in the utilization of alternative carbon and sulphur sources, in amino acid metabolism, and fatty acid degradation. Relevant *C. variabile* genes confer the capability to catabolize gluconate, lactate, propionate, taurine, and gamma-aminobutyric acid and to utilize external caseins. In addition, *C. variabile* is equipped with several siderophore biosynthesis gene clusters for iron acquisition and an exceptional repertoire of AraC-regulated iron uptake systems. Moreover, *C. variabile* can produce acetoin, butanediol, and methanethiol, which are important flavor compounds in smear-ripened cheeses.

**Conclusions:** The genome sequence of *C. variabile* provides detailed insights into the distinct metabolic features of this bacterium, implying a strong adaption to the iron-depleted cheese surface habitat. By combining *in silico* data obtained from the genome annotation with previous experimental knowledge, occasional observations on genes that are involved in the complex metabolic capacity of *C. variabile* were integrated into a global view on the lifestyle of this species.

**Background**

Cheese is one of the oldest dairy products and generally based on the addition of rennet and lactic acid bacteria to milk of cows, sheep, goats, or buffalos. Nowadays, cheese is also the most diverse group of dairy products with hundreds of varieties that are capable of being differentiated mostly by their type of ripening [1]. Bacterial smear-ripened cheeses, such as Appenzeller, Gubbeen, Limburger, Livarot, Munster, and Tilsit, are characterized by the development of a viscous, red-orange smear on the surface of the cheese body during ripening. This smear is a microbial mat composed of various species of yeasts and bacteria. Their combined metabolic activities are mainly responsible for the typical flavor and textural properties of this type of cheese [2,3]. The microbiology of the red-orange smear is poorly understood so far. Generally, acid-tolerant yeasts grow during the first days of ripening of smear-ripened cheeses and metabolize the lactic acid produced by the lactic starter cultures to CO₂.
and H$_2$O$_2$, thereby increasing the pH and producing growth factors, such as pantothentic acid, that permit the growth of Gram-positive and Gram-negative bacteria [4,5]. Yeast species can also deaminate amino acids to ketoacids and NH$_3$, which leads to deacidification of the cheese curd in such a way that bacteria are able to grow [6,7]. The bacterial microflora dominating the later stages of ripening of smear-ripened cheeses is composed of salt-tolerant micrococci, staphylococci, and corynebacteria [2,8]. In the past, it was supposed that the coryneform species Brevibacterium linens represents the major organism on the surface of smear-ripened cheeses [9], but recent investigations emphasized the importance of other coryneform bacteria to the ripening process and identified Arthrobacter nicotianae, Arthrobacter arilatisensis, Brevibacterium ammonitigenes, Microbacterium gubbeenense, Rhodococcus fascians, and members of the genus Corynebacterium (Corynebacterium ammonitigenes, Corynebacterium casei, and Corynebacterium variabile) dominating the microflora of Gubbeen, Tilsit, and other smear-ripened cheeses [10-14].

The genus Corynebacterium includes (besides numerous human and animal pathogens) a diverse collection of non-pathogenic species that have been detected in a wide variety of habitats, such as soil, plant material, waste water, and dairy products [15]. C. casei, for instance, belongs to the main lineage of the genus Corynebacterium [4,16] and was identified as the dominant species on the surface of a Gubbeen-type Irish farmhouse smear-ripened cheese by taxonomic analyses [12,17]. The abundance of C. casei on the cheese surface is closely followed by C. variabile, which is taxonomically located in a distinct subline (cluster 3) of the genus Corynebacterium, with Corynebacterium jeikeium and Corynebacterium urealyticum as the most prominent phylogenetic relatives [18,19]. C. casei and C. variabile are salt-tolerant and able to grow in the presence of 8.0% NaCl and at pH values below 4.9 [17]. Both corynebacteria also metabolize the lactate produced by the starter bacteria, implying that they can grow from the beginning of cheese ripening and may therefore not depend on significant yeast growth during the initial stages of the ripening process [17].

In the present study, we characterize the complete genome sequence of C. variabile DSM 44702 (formerly Corynebacterium mooreparkense) that was originally isolated from the smear-ripened cheese Gubbeen and identified by 16S rDNA sequencing and DNA-DNA hybridization studies [4]. It was initially described as the type strain of a separate species named C. mooreparkense because the organism was isolated at the Dairy Products Research Centre in Moorepark, Ireland [4]. However, additional molecular taxonomic studies demonstrated considerable similarities of all C. mooreparkense isolates with the species C. variabile at the levels of 16S rDNA gene sequence and DNA-DNA relatedness [20]. It was therefore concluded that the name C. mooreparkense is a later heterotypic synonym of C. variabile [20]. In the present study, the annotation of the complete genome sequence and the deduced genetic repertoire of C. variabile DSM 44702 provide detailed insights into the lifestyle and metabolic features of this species, which is strongly adapted to the cheese surface habitat, where it contributes to flavor and texture of the final product.

**Results**

Pyrosequencing and annotation of the C. variabile DSM 44702 genome

The DNA sequence of the C. variabile DSM 44702 chromosome was determined by a whole-genome shotgun approach using pyrosequencing. A quarter of a sequencing run with the Genome Sequencer FLX Instrument and Titanium chemistry yielded 253,845 reads and 97,765,747 bases that were assembled into 93 large contigs (≥500 bases) and 41 small contigs, indicating numerous repetitive elements and insertion sequences in the C. variabile DSM 44702 genome. The remaining gaps in the chromosomal sequence were closed by PCR strategies, supported by the r2cat tool [21] and the Consed program [22]. The final assembly of the DNA sequences yielded a circular chromosome with a size of 3,433,007 bp (Figure 1). A 28-fold coverage was thus obtained by pyrosequencing when considering the final size of the C. variabile DSM 44702 chromosome. The mean G+C content of the C. variabile DSM 44702 genome is 67.15%, which is significantly higher than the value (60%) initially determined by a whole-genome shotgun approach using pyrosequencing. A quarter of a sequencing run with the Genome Sequencer FLX Instrument and Titanium chemistry yielded 253,845 reads and 97,765,747 bases that were assembled into 93 large contigs (≥500 bases) and 41 small contigs, indicating numerous repetitive elements and insertion sequences in the C. variabile DSM 44702 genome. The remaining gaps in the chromosomal sequence were closed by PCR strategies, supported by the r2cat tool [21] and the Consed program [22]. The final assembly of the DNA sequences yielded a circular chromosome with a size of 3,433,007 bp (Figure 1). A 28-fold coverage was thus obtained by pyrosequencing when considering the final size of the C. variabile DSM 44702 chromosome. The mean G+C content of the C. variabile DSM 44702 genome is 67.15%, which is significantly higher than the value (60%) initially determined by a reverse-phase HPLC method [4]. The annotation of the C. variabile genome sequence was performed with the GenDB software system [23] and resulted in the detection of 3,071 protein-coding regions. Furthermore, 59 tRNA genes were predicted by the tRNAscan-SE program [24] and six rrr operons were detected on the leading strands of the C. variabile chromosome (Figure 2).

General architecture of the C. variabile DSM 44702 genome

The plot of the calculated G/C skew [(G-C)/(G+C)] indicated a typical bi-directional replication mechanism of the C. variabile chromosome (Figure 1). According to the presence and distribution of six conserved DnaA boxes, the origin of chromosomal replication (oriC) is located downstream of the dnaA coding region [25]. Moreover, the biased distribution of architecture imparting sequences (AIMS) on the leading strands of the C. variabile chromosome indicated the presence of a dif region [26] at position 1,603 kb on the chromosomal map, dividing the chromosome of C. variabile into two replicores of nearly similar size (Figure 2). Due to the highly biased presence of AIMS on the leading strands of the C. variabile chromosome
**Figure 1** *Corynebacterium variabile* DSM 44702 genome plot. The circles represent from the outside: circle 1, DNA base position [Mb]; circle 2, protein-coding regions transcribed clockwise; circle 3, protein-coding regions transcribed anticlockwise; circle 4, G+C content plotted using a 10-kb window; circle 5, G/C skew plotted using a 10-kb window. The genome plot of *C. variabile* DSM 44702 was generated with the web version of the DNAPlotter tool.

**Figure 2** Distribution of architecture imparting sequences in the *C. variabile* DSM 44702 chromosome. The distribution of the octamer sequences G(A/T/C)GGGGGA and (T/C)GGGGGAG on the leading and lagging strands of the *C. variabile* chromosome is shown. The origin of chromosomal replication (*oriC*) is marked. The deduced *dif* site is located at 1.603 Mb of the chromosomal map. The location of six rRNA operons (*rmA*-*rmF*) on the leading strands of the *C. variabile* chromosome is indicated.
(Figure 2), we also examined the gene-strand bias in this strain. In total, 59.6% of the predicted gene repertoire is located on the leading strands of the *C. variabile* chromosome, revealing a moderate gene-strand bias in this species. However, it has been suggested previously that essentiality is the driving force for gene-strand bias in bacterial genomes [27]. We therefore calculated also the gene-strand bias for candidate essential genes in *C. variabile*. For this purpose, the set of 658 candidate essential genes detected in the genome of *Corynebacterium glutamicum* R by high-density transposon mutagenesis [28,29] was compared to the predicted gene repertoire of *C. variabile* by reciprocal BLASTP matches with the EDGAR software [30]. This comparative content analysis revealed 373 of the candidate essential genes from *C. glutamicum* R having orthologs in the *C. variabile* genome. The majority of these genes (80.4%) are located on the leading strands of the *C. variabile* chromosome, with 68% of all candidate essential genes being located on the left replichore.

Synteny analysis by reciprocal best BLASTP hits [30] revealed a highly conserved order of orthologous genes between the chromosomes of *C. variabile* DSM 44702 and *C. jeikeium* K411, with the exception of a translocated DNA region comprising a segment of approximately 125 kb in size (Figure 3). This result is consistent with the very close phylogenetic relationship of both species that belong to the same subline (cluster 3) of the genus *Corynebacterium* [18,19], and the general observation that genetic rearrangements are rare in corynebacterial genomes [15,31]. The genome of *C. variabile* DSM 44702 thus contains the previously predicted inversions around *dif* and *oriC*, which are both characteristic genomic features of cluster 3 species in the genus *Corynebacterium* [31]. However, the synteny between both genomes is interrupted

---

**Figure 3** Synteny between the chromosomes of *C. variabile* DSM 44702 and *C. jeikeium* K411. The X-Y plot is composed of dots forming syntenic regions between both chromosomes. Each dot represents a predicted *C. variabile* protein having an ortholog in the *C. jeikeium* chromosome, with co-ordinates corresponding to the position of the respective coding region in each genome. Orthologous proteins were identified by reciprocal best BLASTP matches with the EDGAR software. The genomic region of *C. variabile* DSM 44702 indicating a translocation event is marked by an arrow. Selected genomic regions characterized by breakpoints of synteny between *C. variabile* and *C. jeikeium* are labeled according to their predicted gene content.
due to the presence of additional genes in the C. variabile chromosome that were functionally assigned to a phage island, several transposon hotspots, and species-specific gene islands, the latter containing for instance large gene clusters involved in iron acquisition and siderophore biosynthesis (Figure 3). Some of these species-specific gene islands show a significant deviation from the mean G+C content of the C. variabile chromosome and are flanked by several insertion sequences, suggesting that horizontal gene transfer contributed to the evolution of the current gene repertoire of strain DSM 44702 (Figure 1).

The phage island of the C. variabile chromosome has a size of about 48.3 kb and comprises 60 genes (CVAR_0826-CVAR_0886). This genomic region is located between two tRNAGly genes that may represent the integration site of a putative phage. This view is supported by the presence of 50-bp direct repeats, which are part of the tRNA genes and flank the prophage-like region. Most genes (39 out of 60) of the phage island encode hypothetical proteins of unknown function, whereas others encode enzymes involved in phage DNA replication, recombination, and repair. The structural proteins of the putative phage revealed similarities to the tail structure of Rhodococcus phage ReqPine5 [32] and to the capsid structure of Listeria phage A006 [33].

The calculated reciprocal best BLASTP hits [30] were moreover used to compare the predicted proteome of C. variabile DSM 44702 with the complete set of proteins encoded in the genomes of C. jeikeium K411 [34] and C. urealyticum DSM 7109 [31], which are close taxonomic relatives with known genome sequences. This comparative content analysis at proteome level revealed that 1, 120 proteins (36.5%) of C. variabile share homologs in the genomes of C. jeikeium and C. urealyticum (Figure 4A). According to this comparative data, C. variabile contains 1, 699 proteins with no homologous counterparts in the proteomes of the taxonomic relatives C. jeikeium and C. urealyticum. However, it has to be considered that the three genomes of the taxonomically closely related species differ considerably in their sizes as well as in the numbers of protein-coding regions [31,34]. Thus, the 3.31 Mb sequence data of the soil bacterium C. glutamicum R [28] was incorporated into this comparative analysis, revealing a comparably extensive set of 1, 534 genes that are specific for C. variabile when compared to the selected corynebacteria (Figure 4B). This calculation indicates the presence of an extensive and still unexplored variability in the gene equipment of non-pathogenic corynebacteria. This set of genes might contribute to the characteristic features of C. variabile that define its distinct metabolic capabilities.

In the following sections, we present the detailed analysis of these metabolic features of C. variabile DSM 44702 and combine them with data deduced from the predicted transcriptional regulatory repertoire. This approach in evaluating the complete genome sequence

![Figure 4](http://www.biomedcentral.com/1471-2164/12/545) Comparative content analysis of the gene repertoire of C. variabile DSM 44702 and other corynebacteria. (A) Venn diagram showing the comparison between the genomes of C. variabile DSM 44702, C. jeikeium K411, and C. urealyticum DSM 7109, all belonging to the cluster 3 subline of the genus Corynebacterium. (B) Venn diagram showing the comparison of C. variabile DSM 44702, C. jeikeium K411, C. urealyticum DSM 7109, and C. glutamicum R, the latter being a member of the main lineage of the genus Corynebacterium. This comparative content analysis exploits all protein-coding regions of the selected genomes and detects orthologs by reciprocal best BLASTP matches with the EDGAR software.
of *C. variabile* DSM 44702 revealed a collection of relevant genes contributing to the lifestyle of this species and their integration into a transcriptional gene-regulatory network.

**General metabolic features of *C. variabile* DSM 44702 deduced from the complete genome sequence**

The experimental work on the taxonomic description of *C. variabile* DSM 44702 as the type strain of the former species *C. mooreparkense* revealed the ability of this cheese isolate to utilize glucose, fructose, mannose, and ribose as carbon and energy sources [4]. Bioinformatic analysis of the gene repertoire assigned to the central carbohydrate metabolism detected a complete set of genes involved in glycolysis, gluconeogenesis, and the pentose phosphate pathway, as well as the presence of glucose- and fructose-specific components (*ptsG* and *ptsH*) of the bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system (*ptsH* and *ptsI*) and at least one ABC-type transport system (*sugABCD*) for sugar uptake (Additional file 1). The *stcRS* genes of a two-component signal transduction system are located directly upstream of the putative *sugABCD* operon and might be involved in the transcriptional control of this sugar importer. Glucokinase and ribokinase genes (*glk* and *rbsK*) are present in *C. variabile* DSM 44702, allowing the conversion of “free” sugars into phosphorylated central pathway intermediates (Additional file 1).

According to the genome annotation, *C. variabile* DSM 44702 can utilize gluconate that is imported by gluconate permease (*gntP*) and converted to 6-phosphogluconate by gluconate kinase (*gntK*) (Additional file 1). *C. variabile* DSM 44702 can moreover channel propionate via its methylcitrate cycle genes [35] into the tricarboxylic acid cycle (Additional file 1). Calcium propionate occurs naturally in many dairy products including cheese, and some types of cheese contain as much as 1% of natural propio-

tality in many dairy products including cheese, and some types of cheese contain as much as 1% of natural propio-

of 128 families [43,46]. *C. variabile* DSM 44702 encodes 15 candidate glycoside hydrolases associated to eight families. The protein families GH13, GH23, and GH25 are represented by three proteins each, whereas the families GH32, GH44, GH51, and GH52 contain only a single protein member. Two proteins of *C. variabile* DSM 44702 were associated to the glycoside hydrolase family GH3. Glycosyl transferases (GT) catalyze the transfer of sugar moieties during polysaccharide biosynthesis and are classified into 94 families [43,47]. A total number of 28 candidate glycosyl transferases was detected in the genome of *C. variabile* DSM 44702 and grouped into 13 protein families. The largest families are GT2 and GT4, each containing six proteins, followed by family GT87 with three enzymes, and GT1, GT3, and GT51 with two members. One protein
was identified in each of the families GT5, GT20, GT28, GT35, GT39, GT53, and GT85. Carbohydrate esterases (CE) catalyze the de-\(O\)-acylation or de-\(N\)-acylation of substituted saccharides and are grouped into 16 families [43,48]. \textit{C. variabile} DSM 44702 encodes 14 candidate carbohydrate esterases, six of them grouped into family CE1 and six enzymes classified into family CE9. Family CE14 contains two proteins. A carbohydrate-binding module (CBM) is defined as a contiguous amino acid sequence within a carbohydrate-active enzyme with a discreet fold having carbohydrate-binding activity [49]. In \textit{C. variabile} DSM 44702, two proteins were assigned to family CBM48, but the respective proteins are simultaneously associated to family GH13. Altogether, 57 candidate proteins constitute the set of carbohydrate-active enzymes of \textit{C. variabile} DSM 44702. Moreover, the hitherto known biosynthesis pathways for vitamins and cofactors in corynebacteria were reconstructed and shown to be complete in \textit{C. variabile} DSM 44702, including the synthesis of molybdenum cofactor (Additional file 2). The anabolic pathways for purine and pyrimidine nucleotides are also complete (data not shown), as are all pathways for the \textit{de novo} biosynthesis of standard proteinogenic amino acids (Additional file 3). The \textit{C. variabile} DSM 44702 genome is equipped with several transport systems allowing the import of “free” amino acids (Additional file 3). The content of amino acids varies between different types of cheese mainly due to the duration and intensity of proteolysis during the ripening process [50]. Seven amino acids are often concentrated in cheese and contribute to the taste of the final product: leucine, lysine, and phenylalanine (bitter); proline and valine (bitter sweet); aspartate and glutamate (salty umami) [50]. According to the detailed knowledge of amino acid transporters in \textit{C. glutamicum}, putative importers for these amino acids are encoded in \textit{C. variabile} DSM 44702 by \textit{brnQ} (leucine, valine, and isoleucine) [51], \textit{lysI} (lysine) [52], \textit{pup} (phenylalanine) [53], \textit{aroP} (phenylalanine, tyrosine, and tryptophan) [54], \textit{proP} and \textit{putP} (proline) [55,56], and \textit{gluABCD} (glutamate) [57] (Additional file 3). In addition, the sulphur-containing amino acid methionine, which is the precursor of the flavor compound methanethiol, is imported into \textit{C. variabile} DSM 44702 by two transport systems, an ABC-type transporter encoded by the \textit{metQNI} genes and a two-subunit member of the neurotransmitter:sodium symporter family encoded by the \textit{metP-metPS} gene pair [58] (Additional file 3).

**The secretome of \textit{C. variabile} DSM 44702 deduced from the complete genome sequence**

Like other members of the genus \textit{Corynebacterium}, \textit{C. variabile} DSM 44702 encodes two types of protein secretion systems, the general secretory (Sec) pathway and the twin-arginine translocation (Tat) pathway. The Sec translocation pathway has two distinct functions in bacterial protein trafficking, the cotranslational integration of proteins into the cytoplasmic membrane and the posttranslational translocation of proteins across the membrane [59]. The posttranslational export of proteins is facilitated by the SecYEG translocon in conjunction with the multifunctional protein SecA. The latter protein binds to cytosolic precursor proteins destined for export and delivers them to the SecYEG translocon. As the SecA protein is also an ATPase, it provides energy for the stepwise translocation of the precursor protein through the channel of the SecYEG translocon. The precursor proteins have a typical signal sequence at the amino-terminal end, which is cleaved from the precursor by either signal peptidase I or signal peptidase II [59].

The genome sequence of \textit{C. variabile} DSM 44702 encodes all protein components of the basic Sec machinery necessary for protein secretion. The \textit{lepB1} and \textit{lepB2} genes encode homologs of signal petidase I (EC 3.4.21.89), whereas the \textit{lspA} gene codes for signal petidase II, also named lipoprotein signal peptidase (EC 3.4.23.36). In addition, the \textit{secA2} gene of \textit{C. variabile} DSM 44702 provides a homolog of the preprotein translocase subunit SecA. The presence of the accessory protein SecA2 was reported previously for other Gram-positive bacteria. In \textit{M. tuberculosis} and \textit{M. smegmatis}, the SecA2 protein is required for the export of a small subset of exoproteins, whereas SecA1 remains the essential housekeeping translocase [60]. On the other hand, both SecA homologs are probably essential for viability in \textit{C. glutamicum} [61]. \textit{C. variabile} DSM 44702 moreover encodes the signal recognition particle SRP, which is essential for targeting of almost all inner membrane proteins in \textit{E. coli} [62], as well as its receptor FtsY [59].

To estimate the number of secreted proteins encoded by \textit{C. variabile} DSM 44702, the first 70 amino acid residues of each protein were used to search for amino-terminal signal peptides with SignalP 4.0 [63]. In this way, a total number of 363 proteins were identified to be potentially secreted by \textit{C. variabile}. To furtheron deduce the number of putative lipoproteins, all secreted proteins were screened for the presence of a lipobox motif using the DOLOP database tools [64]. Altogether, 66 lipobox-containing proteins were identified in the proteome of \textit{C. variabile} DSM 44702. The signal peptides of these proteins are therefore most likely recognized by signal peptidase II. The signal peptides of the remaining 297 proteins are recognized by signal peptidase I and the respective proteins are exported by the machinery of the Sec pathway of \textit{C. variabile} DSM 44702. However, it has to be considered that proteins with membrane-spanning domains might also contain signal peptides that are recognized by type I signal peptidases [65]. Out of the 363 proteins predicted
to be secreted by *C. variabile* DSM 44702, a set of 100 proteins with membrane-spanning domains were detected by the TMHMM tool [66], which therefore may be destined for the integration into the cytoplasmic membrane. This sub-set of proteins comprises eleven membrane-associated enzymes with functions in the respiratory chain or in cell wall turnover of *C. variabile* DSM 44702 and 25 putative permeases with transport functions for diverse substrates. Two-thirds of the proteins were annotated as putative membrane proteins with unknown functions. After exclusion of proteins with membrane-spanning domains, the remaining 263 predicted extracellular proteins of *C. variabile* DSM 44702 were grouped into functional categories according to the classification scheme of the Clusters of Orthologous Groups of proteins (COG) [67]. A large portion of the predicted extracellular proteins was assigned to COG classes R (16%), S (6.4%), and X (39.8%), containing proteins that are currently not grouped into COGs or whose precise functions are hitherto unknown. Prominent functional classes of the extracellular proteins of *C. variabile* DSM 44702 comprise inorganic ion transport and metabolism (COG class P; 16%), cell wall and membrane biogenesis (COG class M; 7.6%), and signal transduction mechanisms (COG class T; 3.4%).

The Tat pathway of *C. variabile* DSM 44702 is represented by two members of the TatC protein family encoded by tatC1 and tatC2, as well as four members of the TatA family encoded by paralogous versions of *tata* and *tatb* [68]. Proteins are targeted to the Tat pathway by N-terminal signal peptides containing the twin-arginine motif that is followed by a hydrophobic stretch of amino acids [69]. In *C. variabile* DSM 44702, a total number of ten proteins containing signal peptides meeting these criteria were detected by the prediction tools TATP [70] and TATFIND [71]. This small number of proteins secreted by the Tat pathway is comparable to those predicted previously in *C. diphtheriae* [72], *C. glutamicum*, and *C. efficiens* [65,73]. The set of proteins secreted by the Tat pathway of *C. variabile* DSM 44702 includes for instance an alkaline phosphatase (EC 3.1.3.1), a ceramidase (EC 3.5.1.23), an esterase of the SGNH type, and a Dyp-type peroxidase.

The transcriptional regulatory repertoire of *C. variabile* DSM 44702 deduced from the complete genome sequence

The repertoire of candidate transcription regulators encoded in the *C. variabile* DSM 44702 genome was deduced from the functional genome annotation, taking into account the comprehensive knowledge of the gene-regulatory network of *C. glutamicum* ATCC 13032 [74]. A collection of 156 genes encoding DNA-binding transcription regulators (132 genes), sigma factors (8 genes), and response regulators of two-component signal transduction systems (16 genes) can be regarded as the minimal regulatory repertoire of *C. variabile* DSM 44702 (Figure 5). It is noteworthy that among the genes for response regulators, four are not associated with a corresponding sensor histidine kinase gene (CVAR_0325, CVAR_0577, CVAR_1009, and CVAR_2072). The deduced set of candidate transcription regulators represents 5.1% of the predicted protein-coding genes of the *C. variabile* genome. This value is in the range known from the 3.28 Mb genome of *C. glutamicum* ATCC 13032 (5.3%) [74] and is in agreement with previous observations that less than 10% of the total number of predicted proteins is associated with transcriptional regulatory processes in bacteria [75].

The collection of potential transcription regulators was grouped into regulatory protein families according to their amino acid sequence similarities and domain organizations [76]. This bioinformatic classification assigned the predicted transcription regulators of *C. variabile* to 37 regulatory protein families that vary significantly in their number of representatives (Figure 5). The largest family of DNA-binding transcription regulators is TetR with 30 members, followed by AraC with 13 proteins. The TetR family of transcription regulators is widely distributed among bacterial species [77] and is also the most prevalent group of regulatory proteins in other corynebacteria [78]. The AraC family of transcription regulators comprises a diverse group of proteins that are involved in the regulation of various biological processes, such as carbon and nitrogen metabolism, adaptive responses, stress responses, and virulence [79]. The number of AraC-type transcriptional regulators is remarkably high in *C. variabile*, as other non-pathogenic corynebacteria encode only three or four members of this protein family, whereas genes encoding AraC-type regulators are less prevalent or even absent in the genomes of pathogenic corynebacteria [78,80,81].

*C. variabile* encodes only 22 proteins of the previously detected collection of 24 transcription regulators that constitute the core set of DNA-binding transcription regulators in corynebacteria. The genome of *C. variabile* DSM 44702 lacks orthologs of the conserved regulators RamB and SugR that are involved in the transcriptional control of acetate metabolism and central sugar metabolism, respectively [74,78]. To detect regulatory interactions contributing to metabolic pathways relevant for cheese ripening and flavor generation by *C. variabile*, the deduced knowledge of the transcriptional regulatory repertoire was combined with bioinformatic motif searches for DNA-binding sites of prominent regulatory proteins [82]. In this way, relevant metabolic pathways of *C. variabile* DSM 44702 described below were linked with the transcriptional regulatory network of this cheese isolate.
Specific features of *C. variabile* DSM 44702 related to carbohydrate metabolism and its regulation in the cheese habitat

The characteristic features of the central carbohydrate metabolism in *C. variabile* DSM 44702 with respect to the ripening of smear-ripened cheeses are in part specified by genes providing the ability to metabolize citrate, L-lactate, taurine, and γ-amino butyric acid (Figure 6). These metabolites are well-known as prominent components on the surface of smear-ripened cheeses from several previous studies [1,83-86].

**Citrate uptake in *C. variabile* DSM 44702**

The metabolic pathway reconstruction of *C. variabile* DSM 44702 revealed the presence of the *citP* gene encoding a citrate transporter of the CitMHS protein family (Figure 6). Members of this transporter family import citrate in symport with Mg²⁺ or Ca²⁺ ions. Citrate is naturally occurring in milk, containing 8 to 9 mM of citrate during fermentation [87]. Therefore, *C. variabile* DSM 44702 can benefit from the natural composition of the cheese habitat by importing citrate as an additional carbon and energy source. Citrate metabolism is widespread among lactic starter bacteria and can contribute to the synthesis of flavor compounds, such as acetoin and butanediol [88-91].

The expression of the citrate transporter gene *citP* is most likely regulated by the two-component signal transduction system CitAB. The corresponding *citAB* genes are located upstream of *citP* in the *C. variabile* DSM 44702 genome. The orthologous two-component system of *C. glutamicum* ATCC 13032 was shown to regulate genes involved in citrate uptake and metabolism [92]. Target genes of the CitAB system are apparently coregulated by the cAMP-sensing transcription regulator GlxR in this species [74]. The GlxR protein is currently the only known global transcription regulator in corynebacteria, connecting different functional modules of the gene-regulatory network [93]. The *glxR* gene is conserved among all hitherto sequenced corynebacterial species and also present in *C. variabile* DSM 44702. A GlxR binding site was detected upstream of the *citP* coding region, indicating that the expression of this gene is under local control by CitAB and under global control by GlxR in *C. variabile* DSM 44702.

**Lactate metabolism in *C. variabile* DSM 44702**

Two main sources of carbon available to microorganisms at the beginning of cheese ripening are lactose and lactate [1]. Lactose is rapidly converted to lactate in the early days of cheese ripening by the large number of lactic starter bacteria present in the cheese. Lactate is thus the most important carbon source after a few days of cheese ripening [1,94]. According to the taxonomic description and genome sequence data, *C. variabile* DSM 44702 is unable to utilize lactose [4]. On the other
hand, the growth rate of \( C. \text{variabile} \) and lactate dehydrogenase activity were significantly increased in the presence of lactate [95]. The \( C. \text{variabile} \) DSM 44702 genome contains the \( lldT \) gene coding for L-lactate permease involved in L-lactate uptake, which is located upstream of the \( lldD1 \) gene. A paralog of \( lldD1 \), named \( lldD2 \), is located elsewhere on the chromosome. Both genes encode L-lactate dehydrogenase (EC 1.1.1.27) that facilitates the conversion of L-lactate to pyruvate. In addition, L-lactate can be converted to acetate by lactate 2-monoxygenase (EC 1.13.12.4) encoded by the \( lldA \) gene of \( C. \text{variabile} \) DSM 44702. D-Lactate is also used as substrate by three \( dld \) paralogs (\( dld1-dld3 \)) coding for D-lactate dehydrogenases (EC 1.1.1.28). D-Lactate is a natural product in smear-ripened cheeses formed by lactic acid bacteria that can produce both stereoisomers of lactic acid [94]. Due to efficient metabolism of L-lactate and D-lactate, \( C. \text{variabile} \) can probably grow from the beginning of the ripening process and is independent of significant yeast growth during the initial stages of cheese ripening [17]. GlxR binding sites were detected upstream of the putative \( lldT-lldD1 \) operon, upstream of the \( lldD2 \) coding region as well as in front of the paralogous \( dld \) genes (data not shown). Lactate metabolism of \( C. \text{variabile} \) DSM 44702, with the exception of \( lldA \) expression, is thus under global control by GlxR, as it was demonstrated previously in \( C. \text{glutamicum} \) ATCC 13032 [74,96].
Taurine utilization by C. variabile DSM 44702

Taurine (2-aminoethanesulfonic acid) is a sulfonated amino acid that is not incorporated into polypeptides. It derives from the metabolism of sulphur-containing amino acids and is the most abundant free amino acid in goat's milk, with a mean value of 66.2 mg/l [97]. The amount of taurine in cow's milk is considerably lower (1.6-10 mg/l) [98]. The C. variabile DSM 44702 genome contains a gene cluster involved in the assimilation of taurine (Figure 6). The respective genomic region comprises the tauCBA genes encoding a predicted taurine transporter of the ABC superfamily, the tauD1 gene encoding a putative taurine dioxygenase, and the tauR gene coding for a ROK-type transcription regulator. Moreover, C. variabile DSM 44702 contains a paralog of tauD1 not located in this gene cluster, tauD2, encoding a second taurine dioxygenase. This enzyme (EC 1.14.11.17) catalyzes the key step in taurine assimilation by generating sulfite, aminoacetaldehyde, and succinate (Figure 6). As the resulting sulfite can be converted to cysteine, the cheese component taurine can serve as a combined sulphur and carbon source for C. variabile DSM 44702.

The expression of the tauCBA and tauD1 genes is apparently regulated by TauR, a ROK-type transcription regulator, whose corresponding gene is located upstream of the tauD1CBA operon. ROK-type regulators are known from C. glutamicum ATCC 13032 to contribute to the transcriptional control of sulphur metabolism that is under hierarchical control by the TetR-type regulator McbR [99-101]. McbR belongs to the core set of corynebacterial transcription regulators [74] and acts as the master regulator of all aspects of sulphur metabolism in C. glutamicum and C. jeikeium [101,102]. The McbR protein is also encoded in C. variabile DSM 44702 by the mcbR gene that is located at the boundary of the small duplicated genomic region CVAR_2865-CVAR_2874 (mcbR). The second copy of this DNA segment is part of the iron acquisition island II (Figure 3), comprises the coding regions CVAR_2821-CVAR_2830, and ends with a S’ truncated version of the mcbR gene, named mcbR2. A genome-wide search for McbR binding sites in C. variabile DSM 44702 detected 23 operators that are probably controlling 41 genes involved in (S-adenosyl)methionine biosynthesis, methionine import, cysteine biosynthesis, and sulfate reduction (data not shown). In addition, McbR binding sites were detected in the tauR-tauD1 intergenic region and in front of the tauD2 coding region. This data strongly indicates that all genes involved in taurine assimilation are embedded in the sulphur regulatory network of C. variabile DSM 44702.

GABA utilization by C. variabile DSM 44702

GABA (γ-aminobutyric acid) is a four-carbon, non-proteinogenic amino acid known to occur in cheese with a varying concentration that is strongly affected by the protocol of cheese manufacture, the type of primary starter cultures, the cheese microbiota, and the ripening conditions [85,86]. Several species of lactic starter bacteria are able to synthesize GABA from L-glutamate by glutamate decarboxylase (EC 4.1.1.15) [85] that is widely distributed among eukaryotes and prokaryotes [103] and also encoded in the C. variabile DSM 44702 genome by the gadB gene. In addition, GABA is degraded to succinate by the enzymes γ-aminobutyrate aminotransferase (EC 2.6.1.19) and succinate semialdehyde dehydrogenase (EC 1.2.1.16) in C. variabile DSM 44702 (Figure 6). These three enzymatic reactions belong to a metabolic pathway known as GABA shunt [104]. Two genes encoding GABA transporters (gabP1 and gabP2) were detected in C. variabile DSM 44702. The first one is located in a cluster comprising the genes gabD1 encoding succinate semialdehyde dehydrogenase, gabT1 coding for γ-aminobutyrate aminotransferase, and gabR1 specifying a transcription regulator. The second gene cluster attributed to GABA utilization is composed of gabP2, gabT2, and gabR2, and is thus lacking a gabD component. In total, six potential gabD genes (gabD1-gabD6) are distributed across the C. variabile DSM 44702 chromosome.

According to this genome annotation, the uptake and catabolism of GABA in C. variabile DSM 44702 is represented by at least three gene regions (Figure 6). The first gene cluster (gabT1-gabD1-gabP1) is preceded by the regulatory gene gabR1 coding for a transcription regulator of the COG2508 class. Likewise, the second gene cluster (gabT2-gabP2) is preceded by the regulatory gene gabR2 also coding for a transcription regulator assigned to the COG2508 class. The presence of this type of regulatory genes upstream of gab gene clusters is conserved among GABA-utilizing corynebacteria [105], suggesting that the respective regulators of the COG2508 class are involved in the transcriptional control of the divergently arranged gabT(D)P genes. The succinate semialdehyde dehydrogenase gene gabD2 is located elsewhere on the chromosome and organized in an operon with the regulatory gene gabR3, which encodes a transcription regulator of the LysR protein family [106]. Moreover, GlxR binding sites were detected in the intergenic region of gabR1/gabT1 and gabR2/gabT2, as well as upstream of the gabR3 coding region, indicating that the respective GABA-related genes of C. variabile DSM 44702 are under hierarchical regulation by local GabR regulators and the global cAMP-sensing regulator GlxR. Integration of GABA utilization into the global GlxR network of C. variabile DSM 44702 is reasonable, because the central pathway intermediate succinate is finally generated by the GABA shunt (Figure 6). Furthermore, it is remarkable that GABA utilization as well as taurine and propionate metabolism channel
succinate into the tricarboxylic acid cycle, as the typical generation of this compound by succinyl-CoA synthetase is not encoded in C. variabile DSM 44702.

Nitrate, nitrite, and oxaloacetate metabolism of C. variabile DSM 44702 in the cheese habitat
Low amounts of nitrates and nitrites are natural constituents of many types of cheese [107]. Both compounds may exert important effects on human health when present in higher concentrations, as they may react with naturally present amines to form potent carcinogens. The content of nitrates and nitrites in cheese is therefore strictly regulated by law in many European countries. The annotation of the C. variabile DSM 44702 genome sequence revealed the presence of the nrtABC genes encoding an ABC-type transport system involved in nitrate uptake (Figure 6). Nitrate can be reduced by a putative nitrate reductase (EC 1.7.99.4) encoded by the narB gene of C. variabile DSM 44702, which is located upstream of nirB and nirD, both coding for subunits of nitrite reductase (EC 1.7.1.4). This enzyme converts nitrite to ammonium hydroxide, resulting in the detoxification of nitrite and the generation of a suitable nitrogen source for C. variabile DSM 44702. Alternatively, nitrite can be exported by the putative nitrite extrusion protein NarT, which is encoded in the same gene cluster by the narG gene.

Moreover, the oxalate:formate antiporter gene oxtT was identified in C. variabile DSM 44702 (Figure 6). The OxtT protein is a member of the major facilitator superfamily and exchanges formate for oxalate across the cytoplasmic membrane [108]. Oxalate is present in milk and cheese in relatively small amounts. Nevertheless, oxalic acid can react with calcium ions in the milk producing calcium oxalate, thereby preventing this vital nutrient from being absorbed in the human intestine [109]. In principle, oxalate can be converted to formate by oxalate decarboxylase (EC 4.1.1.2), but the respective oxD gene is absent in the genome of C. variabile DSM 44702. However, formate can be oxidized in C. variabile DSM 44702 by formate dehydrogenase (EC 1.2.1.2) encoded by the fdhFD genes (Additional file 1). Due to import of oxalic acid by C. variabile DSM 44702, the binding of calcium ions probably occurs to a lower degree in smear-ripened cheeses in the presence of this strain. Therefore, the gene equipment of C. variabile DSM 44702 provides a substantial benefit for the quality of smear-ripened cheeses.

Lipolysis and fatty acid metabolism by C. variabile DSM 44702 in the cheese habitat
Cow’s milk contains high amounts of lipids and triacylglycerides (about 35 g/l), which may serve as carbon and energy sources for the microflora of smear-ripened cheeses [110]. Free fatty acids derive from breakdown of triacylglycerides, and their further degradation can lead to volatile compounds having low olfactory thresholds in cheese flavor perception [111]. In smear-ripened cheeses, more than 4 g/kg of free fatty acids are available for the surface microbiota [1]. The genome of C. variabile DSM 44702 is equipped with a large set of genes encoding enzymes involved in lipolysis and the degradation of fatty acids (Figure 6). It contains five genes (estA1-estA5) encoding secreted esterases of the SGNH-hydrolase superfamily and the estB gene coding for an esterase of the AB hydrolase superfamily [112]. Secreted lipases are encoded by five genes in the C. variabile DSM 44702 genome. The paralogous lipA genes (lipA1-lipA3) encode class 2 lipases that hydrolyse ester bonds in triacylglycerol giving diacylglycerol, monoacylglycerol, glycerol, and free fatty acids [112]. The genes lipS1 and lipS2 belong to the LIP superfamily containing lipases with broad lipolytic activities [112]. All these enzymes may thus contribute to the generation of free fatty acids from precursor molecules present in the cheese habitat.

Furthermore, 13 fadD genes encoding acyl-CoA synthetases were identified in the genome of C. variabile DSM 44702, including the fadD1 gene that is involved in mycolic acid biosynthesis [113]. The amino acid sequences of the FadD enzymes in C. variabile DSM 44702 revealed low levels of overall similarity, but the characteristic AMP-binding motif of acyl-CoA synthetase is highly conserved among the twelve paralogs (FadD2-FadD13). In principle, acyl-CoA synthetases are involved in activating free fatty acids to form acyl-CoA of various chain lengths concomitant with the transport into the bacterial cell [114]. The amino acid sequence diversity of the predicted acyl-CoA synthetases might indicate different substrate specificities of these enzymes, which would enable C. variabile to utilize a broader range of fatty acid substrates in its natural cheese habitat, probably leading to a growth advantage as a well-adapted species.

In addition to the set of fadD genes, a complete β-oxidation pathway for fatty acid degradation is encoded in the C. variabile DSM 44702 genome (Figure 6). In this pathway, the acyl-CoA is converted into a product with a shortened acyl chain (n-2) and acetyl-CoA that can be used in the tricarboxylic acid cycle to generate energy, whereas the shortened acyl-CoA is used in a new round of β-oxidation [115]. In principle, C. variabile DSM 44702 can utilize exogenous fatty acids also as carbon sources, since the glyoxylate bypass genes aceA and aceB are present and the gluconeogenesis pathway is complete (Additional file 1). The predicted enzymes involved in β-oxidation of C. variabile DSM 44702 are encoded by ten paralogs of fadE (encoding acyl-CoA dehydrogenase), the bifunctional fadB1 gene (enoyl-CoA hydratase/hydroxyacyl-CoA dehydrogenase), three monofunctional fadB genes (hydroxyacyl-CoA dehydrogenase), three echA genes (enoyl-CoA reductase), three fadF genes (octanoyl-CoA dehydrogenase), and three fadK genes (acyl-CoA oxidase).
hydratase), and five paralogs of fadA (ketoacyl-CoA thiolase). The amino acid sequences of these paralogueous gene families vary substantially, again suggesting diverse substrate specificities of the respective enzymes, at least in the case of the ten acyl-CoA dehydrogenases. Interestingly, all genes coding for secreted esterases or lipases and almost all fad and echA genes of C. variabile DSM 44702 contain at least one GlxR binding site in their upstream regions (data not shown), indicating that the complete metabolic process of lipolysis and β-oxidation is under global control by the GlxR protein.

In contrast to the lipid- and nutrient auxotrophic relatives C. jeikeium and C. urealyticum, the metabolic pathway analysis of the C. variabile DSM 44702 genome revealed the presence of the fatty acid synthase gene fas, showing that growth of this strain is not dependent on the availability of exogenous fatty acids [31,34]. These compounds are, moreover, building blocks for the synthesis of corynomycolic acids, which are major constituents of the corynebacterial cell envelope [116]. It was shown previously that a polyketide synthase encoded by the pks13 gene represents a condensase that catalyzes the key step in the biosynthesis of corynomycolic acids, in conjunction with an acyl-CoA carboxylase and a distinct acyl-AMP ligase [117-119]. The pks13 gene is located in a conserved cluster in the C. variabile DSM 44702 genome, including genes coding for an acyl-CoA carboxylase (accD3), an acyl-CoA synthetase/acyl-AMP ligase (fadD1), the envelope lipids regulation factor ElrF (elrF), and trehalose corynomycol transferases (cmt). It is noteworthy that eight genes encoding trehalase corynomycolyl transferases (cmtA-cmtH) were identified in the C. variabile DSM 44702 genome, catalyzing the transfer of mycolic acids from trehalose monocorynomycolate to another molecule of trehalose monocorynomycolate or on the cell wall arabinogalactan [116]. Five of these genes (cmtC-cmtG) are clustered with the pks13 gene.

A second gene cluster involved in fatty acid metabolism of C. variabile DSM 44702 includes genes coding for putative acyl-CoA carboxylase subunits (CVAR_2092 and CVAR_2091), an acyl-CoA dehydrogenase domain-containing protein (CVAR_2090), an enoyl-CoA hydratase domain-containing protein (CVAR_2089), a citrate lyase β-subunit (citE), an acyl-CoA synthetase (fadD6), acyl-CoA:3-ketoacid-CoA transferase subunits (scoA and scoB), and a ketoacyl-CoA thiolase (fadA4). A similar gene cluster is present only in the genome of the lipophilic isolate C. jeikeium K411, whereas some genes were found in a similar genetic arrangement in the genome of Mycobacterium tuberculosis H37Rv [120]. As most of the genes present in these conserved clusters are linked to fatty acid degradation, they might be involved in the activation and subsequent degradation of a distinct type of fatty acid substrate, whose chemical composition is currently unknown.

Proteolysis and utilization of caseins by C. variabile DSM 44702 in the cheese habitat

Protein degradation contributes to cheese texture and the generation of cheese flavor in all varieties of smear-ripened cheeses, as many relevant flavor compounds derive directly from amino acids [1,87,91,121]. Milk and especially cheese have a high protein content, which is represented by a mixture of αS1-, αS2-, and β-caseins [1]. Extracellular caseins can be degraded by C. variabile DSM 44702 by means of secreted proteolytic enzymes, such as a secretory serine protease encoded by the sepP gene and a secretory aminopeptidase encoded by the pepC gene (Figure 6). In addition, amino-terminal prolines are released from peptides by the enzymatic action of proline iminopeptidase (EC 3.4.11.5) encoded by the pepl gene. The enzymatic activity of an extracellular proline iminopeptidase from C. variabile NCDO 2101 was demonstrated previously [122]. Since caseins are very rich in proline residues and proline is the most abundant amino acid in cheese, it may constitute a prominent substrate for C. variabile. The import of “free” or released proline is facilitated in C. variabile DSM 44702 by the amino acid transporters ProP and PutP (Additional file 3). The subsequent oxidation of proline to glutamate is catalyzed as a two-step reaction by a bifunctional proline dehydrogenase (EC 1.5.99.8 and EC 1.5.1.12) encoded by the putA gene [123].

Interestingly, McbR binding sites were detected in the upstream regions of sepP and pepC, suggesting the integration of these genes into the sulphur regulatory network of C. variabile DSM 44702. Therefore, proteolysis might be used by C. variabile DSM 44702 to cover the demand for sulphur-containing amino acids and is probably enhanced under sulphate starvation conditions. Other genes assigned to proteolysis by C. variabile DSM 44702 are also linked to unfavorable environmental conditions by regulatory features. The proline importer gene putP, for instance, is specified by the presence of a typical SOS box and seems to be under the direct transcriptional control by the SOS response regulator LexA, as previously demonstrated in C. glutamicum [124]. Moreover, the proline iminopeptidase gene pepl is organized in an operon and located downstream of sodC in the C. variabile DSM 44702 genome. The sodC gene encodes a secreted copper, zinc-dependent superoxide dismutase (EC 1.15.1.1) that is characterized by a lipobox motif to be anchored in the cell membrane [64]. The extracellular location of SodC indicates that it helps to protect the cell surface of C. variabile against superoxide generated externally by the microflora of smear-ripened cheeses. The gene arrangement of pepL and sodC implies that the proline iminopeptidase gene plays a role in the responses to external stresses, which is reasonable since proline can be used as compatible solute in corynebacteria under osmotic stress conditions [125].
Osmotic and oxidative stress responses of \textit{C. variabile} DSM 44702

Due to permanently changing environmental conditions during the ripening process of smear-ripened cheeses, the bacteria of the cheese microbiota are constantly exposed to several external stresses, such as pH stress and osmotic stress [110,126]. As the pH of the cheese curd can be as low as 4.5 during the early stage of cheese processing, \textit{C. variabile} has to cope with this external stress condition. It was demonstrated in \textit{C. glutamicum} that potassium transport is essential for corynebacterial pH homeostasis and growth at acidic pH [127]. The potassium channel CglK was found to be relevant for the maintenance of the internal pH and the adjustment of the membrane potential. The genome sequence of \textit{C. variabile} DSM 44702 was therefore screened for the presence of genes coding for actinobacterial potassium transporters, of which seven different types were described very recently [128]. \textit{C. variabile} DSM 44702 encodes a homolog of the potassium channel CglK (CVAR\_2147) and a Kef-type potassium transport system (CVAR\_0647). Moreover, a sodium or potassium/H\textsuperscript{+} antiporter of the NhaP-type (CVAR\_1814) might contribute to pH homeostasis in \textit{C. variabile} DSM 44702.

\textit{C. variabile} DSM 44702 is also well equipped with genes offering protection from osmotic stress that derives from high salt concentrations on the cheese surface during the ripening process (Figure 6). One mechanism to overcome this osmotic stress is the accumulation of osmoprotectants, such as ectoine, proline, and glycine betaine in the cytoplasm [125]. The genome of \textit{C. variabile} DSM 44702 contains the \textit{ectP} gene encoding an ectoine transporter, \textit{proP} coding for an osmoregulated proline transporter, and six genes (betP1-betP3 and betT1-betT3) encoding putative betaine/carnitine/choline transporters of the BCCT family. Betaine can be synthesized in two steps from the precursor choline by choline oxidase (EC 1.1.3.17) and betaine-aldehyde dehydrogenase (EC 1.2.1.8) mediated reactions. Milk and cheese contain high amounts of choline [129], making it available as precursor for the synthesis of betaine by \textit{C. variabile}. The \textit{C. variabile} DSM 44702 genome contains two \textit{betA} genes (\textit{betA1} and \textit{betA2}) encoding choline oxidases, of which the \textit{betA1} gene is clustered with the betaine-aldehyde dehydrogenase gene \textit{betB} and the putative choline transporter gene \textit{betT1}. The homologous uptake systems for osmoprotectants were extensively studied in \textit{C. glutamicum} ATCC 13032 [130], as was the two-component system MtrAB that is involved in the transcriptional control of the genes encoding osmoregulated compatible solute carriers [131]. The \textit{mtrAB} gene pair was also identified in the genome of \textit{C. variabile} DSM 44702, suggesting the presence of a similar MtrAB-controlled network of osmoprotection in this species.

The \textit{C. variabile} DSM 44702 genome is moreover equipped with several genes involved in oxidative stress responses, such as two genes (\textit{katA} and \textit{katG}) encoding catalases (EC 1.11.1.6), duplicated genes (\textit{sodA1} and \textit{sodA2}) coding for intracellular manganese-dependent superoxide dismutases (EC 1.15.1.1), and the aforementioned secreted copper, zinc-dependent superoxide dismutase gene \textit{sodC}. Additional functions involved in oxidative stress responses are provided by \textit{bcp} coding for a peroxiredoxin, \textit{tpx} encoding a thiol peroxidase, and the \textit{ahpCD} gene pair encoding the subunits of alkyl hydroperoxide reductase (EC 1.11.1.15). The latter protein is an antioxi- dant enzyme responsible for directly reducing organic hydroperoxides [132], whereas thiol peroxidase represents an oxidative stress defense system that uses reduction equivalents from thioredoxin and thioredoxin reductase to reduce alkyl hydroperoxides [133]. Thioredoxin (EC 1.8.1.8) and thioredoxin reductase (EC 1.8.1.9) are also encoded by the \textit{trxA} and \textit{trxB} genes in the genome of \textit{C. variabile} DSM 44702. The Bcp protein belongs to the bacterioferritin comigratory protein subfamily of the peroxiredoxin superfamily and represents also a thioredoxin-independent thiol peroxidase [112].

Furthermore, the \textit{C. variabile} DSM 44702 genome contains the regulatory gene \textit{oxyR} encoding a homolog of the central regulator of oxidative stress responses in bacteria [134,135]. The \textit{oxyR} gene belongs to the core set of actinobacterial transcription regulators [74] and is located adjacent to the \textit{ahpCD} gene pair of \textit{C. variabile} DSM 44702. This gene order is also conserved in several corynebacterial genomes and in genomes of slow-growing mycobacteria, where \textit{oxyR} is involved in the activation of \textit{ahpCD} gene expression [136,137]. According to this conserved gene arrangement in actinobacterial species and experimental data from mycobacteria, it is likely that OxyR controls at least the expression of the \textit{ahpCD} genes in \textit{C. variabile} DSM 44702. The \textit{katG} gene of \textit{C. variabile} DSM 44702 is probably under transcriptional control by LexA, as a DNA-binding site (SOS box) of this regulator was detected in front of the \textit{katG} coding region (data not shown). The LexA protein is the key component of the bacterial SOS response [138], and the \textit{lexA} gene is conserved in all corynebacterial genomes [74].

Iron acquisition by \textit{C. variabile} DSM 44702 in the iron-depleted cheese habitat

For most bacteria, iron is essential as a cofactor for proteins involved in important cellular functions, such as DNA biosynthesis and aerobic respiration [139]. Thus, iron acquisition is a vital function for bacterial survival. However, milk and cheese contain the iron-binding protein lactoferrin that can deny iron to microorganisms in such a way that smear-ripened cheeses contain only small amounts of readily available iron [140]. To circumvent this
limitation of iron in the cheese habitat, bacteria of the cheese microflora have evolved several molecular strategies, including the synthesis of siderophores during the ripening process [141]. Siderophores are high-affinity iron-chelating compounds, often synthesized by non-ribosomal peptide synthetases in response to low-iron availability in the environment [142]. The genome of *C. variabile* DSM 44702 contains five genes (nrpS1-nrpS5) encoding proteins with sequence similarities to multi-domain peptide synthetases (Figure 6). The peptide synthetase genes *nrpS2 to nrpS5* are clustered in iron-acquisition island II of the *C. variabile* genome (Figure 3). This genomic island contains also the *entC2* and *entE2* genes, coding for isochorismatase synthase (EC 5.4.4.2) and 2, 3-dihydroxybenzoate-AMP ligase (EC 2.7.7.58), respectively. These gene products revealed similarities to proteins involved in the biosynthesis of the catechol siderophores bacillibactin from *Bacillus subtilis* [143] and enterobactin from *Escherichia coli* [144]. Additional enzymatic functions involved in the biosynthesis of catechol siderophores are encoded in the *entA-entC1-entB-fes* gene cluster of *C. variabile* DSM 44702. The *entA* gene encodes 2, 3-dihydro-2, 3-dihydroxybenzoate dehydrogenase (EC 1.3.1.28) and the *entB* gene codes for isochorismatase (EC 3.2.3.21). The product of the *fes* gene is similar to enterobactin esterase that is probably required for the removal of iron from the siderophore. In conjunction with a non-ribosomal peptide synthetase, these genes provide a complete pathway for the synthesis of catechol siderophores with chorismate as the precursor [145]. Among the cluster 3 species of the genus *Corynebacterium*, similar siderophore biosynthesis genes are only present in the genome of *C. jeikeium* K411 [34].

Further analysis of the *C. variabile* DSM 44702 genome sequence revealed the presence of 29 genes with predicted functions in the utilization of iron-siderophores, named either *siu* (siderophore-iron utilization protein) or *sii* (siderophore-iron interacting protein) (Figure 7). *siu* gene clusters are represented by *siuS* genes encoding substrate-binding proteins (*siuS1-siuS17*), *siuU* genes encoding siderophore utilization proteins (*siuU1* and *siuU2*), and *sii* genes coding for siderophore-interacting proteins (*siiA-siiE*), which might be necessary for the removal of iron from iron-siderophore complexes [110]. In addition, five regulatory genes (*siuR1-siuR5*) encoding transcription regulators of the AraC protein family were assigned to these gene clusters. Interestingly, numerous genes involved in iron utilization are located in the iron acquisition islands I (CVAR_230-CVAR_0239) and II (CVAR_2806-CVAR_2820) of the *C. variabile* genome (Figure 3).

A further aspect of iron aquisition is the transport of this trace element into the bacterial cell. The *C. variabile* DSM 44702 genome contains 23 genes with assigned functions in ferric iron uptake (*fiu*) (Figure 7). The *fiu* gene clusters are represented by genes coding for substrate-binding proteins (*fiuS*), permease components (*fiuP* and *fiuQ*), ATP-binding proteins (*fiuA*), and an iron-utilization protein (*fiuU1*). In addition, three *fiu* gene clusters contain regulatory genes (*fiuR1-fiuR3*) encoding transcription regulators of the AraC family (Figure 7). The accumulation of genes coding for AraC-like proteins in the *siu* and *fiu* gene clusters explains the large number of transcription regulators assigned to this protein family in *C. variabile* DSM 44702 (Figure 5), when compared with the regulatory repertoires in other corynebacterial genomes [78,80,81].

As iron limitation poses a problem for the microflora of the cheese surface, the microorganisms have to compete for iron, but nonetheless, they must regulate iron metabolism to prevent excess iron that can initiate the generation of toxic oxygen radicals from normal products of metabolism by the Fenton reaction [146]. Bacteria have solved the problem of iron acquisition and homeostasis by encoding uptake systems that are tightly regulated at the transcriptional level [139]. DtxR is the key transcription regulator controlling complex gene-regulatory networks involved in iron homeostasis in corynebacteria [74,147]. The complete genome sequence of *C. variabile* DSM 44702 was screened for the presence of DtxR binding sites, using the experimentally validated binding motif from *C. glutamicum* ATCC 13032 as input [147]. This genome-wide motif search detected 26 operators that are probably controlling the expression of 62 target genes, including for instance *dps* (DNA protection during starvation protein), *ftn* (ferritin), *hmuO* (heme oxygenase), *sdhCAB* (succinate dehydrogenase), and the *suf* operon involved in Fe-S cluster assembly. Among the detected operators, 17 are located in the *siu* and *fiu* gene clusters of *C. variabile* DSM 44702, regulating in total 41 genes that are organized in 18 transcription units (Figure 7). This result indicates that almost all *siu* and *fiu* gene clusters are integrated into the iron-regulatory network of *C. variabile* DSM 44702 that obviously exerts a hierarchical control of gene expression by DtxR in concert with the AraC-type regulators of these clusters. Such a complex transcriptional regulation of iron homeostasis has not been detected in any sequenced corynebacterial genome before [74].

**Contribution of *C. variabile* DSM 44702 to flavor generation during cheese ripening**

The flavor compounds commonly found in smear-ripened cheeses are produced by the cheese microbiota as a diverse group of volatile substances, including volatile sulphur-containing compounds, esters, aldehydes, and ketones [148]. Half of these compounds derive from lactose fermentation, citrate degradation, and a few from lipolysis. The second half derives from amino acid catabolism by complex metabolic pathways [87]. Methanethiol, for instance, derives from methionine and is a prominent
Figure 7 Gene regions associated with iron acquisition in *C. variabile* DSM 44702. Gene clusters probably involved in the uptake of ferric iron (*fiu*) and iron-siderophores (*siu*) are presented. Detected DtxR binding sites are indicated as black boxes upstream of the coding regions. Genes encoding iron uptake and utilization systems are colored according to their functional assignments, whereas grey arrows indicate genes of unknown functions present in the *siu* clusters.
aroma compound in smear-ripened cheeses. Several cheese-related strains of \textit{B. linens} produce high amounts of methanethiol due to the presence of L-methionine-\(\gamma\)-lyase (EC 4.4.1.11) [149]. This enzyme converts methionine to methanethiol via a \(\gamma\)-elimination step. A corresponding \textit{mgl} gene was not detectable in the genome of \textit{C. variabile} DSM 44702, although this strain was shown previously to produce methanethiol [4]. This observation indicates that other enzymes must be responsible for the synthesis of methanethiol in \textit{C. variabile} DSM 44702. Cystathionine \(\beta\)-lyase (EC 4.4.1.8) and cystathionine \(\gamma\)-lyase (EC 4.4.1.1) catalyze similar reactions in lactococci [150,151], but only the former enzyme is encoded by the \textit{aecD} gene in \textit{C. variabile} DSM 44702.

Acetoin (3-hydroxybutanone) is another aroma compound of smear-ripened cheeses and gives butter its characteristic flavor [87,152]. \textit{C. variabile} DSM 44702 produces acetoin from pyruvate, with acetolactate being an intermediate. The conversion of acetolactate to acetoin is mediated by acetolactate decarboxylase (EC 4.1.1.5) encoded by the \textit{alsD} gene (Figure 6). Moreover, acetoin can be converted to 2, 3-butanediol by the enzymatic activity of acetoin reductase (EC 1.1.1.4) encoded by the \textit{butA} gene in \textit{C. variabile} DSM 44702. Butanediol is also an important flavor compound in some varieties of smear-ripened cheeses [153].

Discussion

Cheese is one of the oldest dairy products produced by humanity and its origin predates recorded history [1]. Nowadays, hundreds of cheese varieties are produced throughout the world in wide-ranging textures, forms, and flavors. The production processes of all cheese varieties share one common characteristic: the microflora present in each cheese contributes to the typical properties of the final product, such as taste, aroma, and color [1,94]. The complex interactions of these microorganisms and the various biochemical processes occurring during cheese ripening are the result of centuries of adaptation of yeasts, moulds, and bacteria to the cheese habitat [94]. During the traditional ripening of smear-ripened cheeses, the surface flora of old, ripened cheeses was transferred to young, unripened cheeses by a brining step in the early stage of ripening, a process named “old-young smearing” [154]. Therefore, the combined metabolic activities of a highly adapted bacterial community on the surface of smear-ripened cheeses have evolved over many years to determine all aspects of the mature product [94]. However, the process of “old-young smearing” involves the transfer of a complex, undefined, and variable microflora. Therefore, a significant improvement in the industrial production of smear-ripened cheeses is nowadays achieved by using defined surface starter cultures. \textit{C. variabile} was shown previously to be an important bacterial component on the surface of the smear-ripened cheeses Brick and Gubbeen [4,10,17,20]. Likewise, \textit{C. casei} is also a dominant component on the surface of Gubbeen, but is moreover found to be dominant on the surface of other smear-ripened cheeses, such as Ardrahan, Durrus, and Milleens [11]. Therefore, we reconstructed in this study the metabolic capabilities of \textit{C. variabile} DSM 44702 from the complete genome sequence to understand its adaptation to the cheese surface habitat and its physiological role in cheese ripening and flavor generation.

The \textit{C. variabile} genome (3.43 Mb) contains large regions with high levels of synteny with the chromosome of the close phylogenetic relative \textit{C. jeikeium} (2.46 Mb), a multi-drug resistant nosocomial pathogen [155]. However, the size of the \textit{C. variabile} genome is rather characteristic of environmental corynebacteria [15], such as \textit{C. glutamicum} (3.28-3.31 Mb), \textit{Corynebacterium efficiens} (3.14 Mb), and \textit{Corynebacterium nuruki} (3.11 Mb), than of pathogenic relatives of the cluster 3 subline [31,34,80]. In contrast to the genetic repertoires of the cluster 3 species \textit{C. jeikeium} and \textit{C. urealyticum} that both show a clear reduction of genes associated with catabolic activities of carbon sources, the metabolic reconstruction of \textit{C. variabile} revealed the presence of a complete set of enzymes required for central carbohydrate metabolism, vitamin and cofactor biosynthesis, and the synthesis of amino acids. \textit{C. variabile} is equipped for instance with several additional enzymes and proteins required for the utilization of substrates typically present on the cheese surface, including citrate, lactate, propionate, taurine, and \(\gamma\)-amino butyric acid (GABA). The utilization of the sulfonated amino acid taurine as a combined carbon and sulphur source by \textit{C. variabile} has not been reported so far in conjunction with other bacterial species associated with smear-ripened cheeses. GABA was shown to be synthesized by lactic starter bacteria within the body of 22 Italian cheese varieties, such as Gorgonzola, Mozzarella, and Pecorino [85], and is also utilized as carbon source by \textit{Arthrobacter arilaitensis} [110], which is one of the major bacterial species detected on the surface of smear-ripened cheeses. This bacterium is also able to utilize the stereoisomers L-lactate and D-lactate, but not citrate, presumably as a result of a mutation in the gene cluster encoding the two-component system CitAB necessary for activating the expression of the cognate citrate import system [110].

\textit{C. variabile} is also equipped with an extensive set of enzymes participating in \(\beta\)-oxidation or the degradation of lipids and with enzymes involved in proteolysis. Cheese has a high content of lipids and triacylglycerides [1] that are probably made available for the metabolism of \textit{C. variabile} by the enzymatic actions of secreted lipases and esterases. Likewise, caseins are the major protein sources in milk and cheese [1] and are thus prominent substrates

\(\text{http://www.biomedcentral.com/1471-2164/12/545}\)
for proteolysis by C. variabile, presumably carried out by secreted proteases. Moreover, proline iminopeptidase can contribute to proteolysis [122], as prolines are the major amino acid constituents of caseins [1]. It is obvious that these metabolic specificities of C. variabile are closely linked to its niche adaption. Another specific feature of C. variabile is its high salt tolerance [4]. The bodies of smear-ripened cheeses are brined with salt solutions prior to the start of the ripening period to control microbial growth and to extend shelf-life [94]. The efficient accumulation of osmoprotectants by C. variabile may contribute to its salt tolerance, as a large number of genes accounting for the accumulation of compatible solutes were detected in the genome sequence.

Moreover, the C. variabile genome is well equipped with genes involved in iron acquisition, presumably as a result of adaptation to the iron-depleted cheese habitat. At least one complete siderophore biosynthesis cluster was identified in this species, which shows similarity to gene clusters involved in the synthesis of catechol siderophores in B. subtilis [143] and E. coli [144]. The biosynthesis of siderophores was also observed in cheese-related strains of B. linens and A. arilaitensis that are frequently found on the surface of smear-ripened cheeses [110,156]. Moreover, about 50 additional genes were assigned to iron-siderophore and ferric iron uptake components, which have few orthologs in other corynebacteria, indicating that the acquisition of iron is a key factor for growth of C. variabile on the surface of smear-ripened cheeses. It would be interesting to investigate whether C. variabile is able to use siderophores produced by other species sharing its habitat, as it has been assumed for A. arilaitensis [110].

In principle, flavor compounds in cheese can derive from pathways of lipolysis or proteolysis and the metabolism of lactate and citrate [87]. Volatile sulphur compounds, such as methanethiol, are particularly important in the flavor of smear-ripened cheeses and produced by the bacterial microflora [94]. There is experimental evidence that C. variabile produces methanethiol [4], although no methionine-γ-lyase encoding gene (mgl) was detectable in this study. Another flavor compound produced by C. variabile from pyruvate is acetoin, which is also generated by lactococci in fresh unripened cheese and can be reduced to 2,3-butanediol by the action of the enzyme acetoin reductase [157]. The corresponding butA gene was detected in the C. variabile genome, indicating that the synthesis of 2,3-butanediol might contribute to flavor generation of smear-ripened cheeses. Moreover, lipolysis plays an important role in the formation of cheese flavor in several cheese varieties, such as Swiss cheese [158], Hispánico cheese [159], Domiati cheese [160], and smear-ripened cheeses [87]. The quantification of the production of flavor compounds from fat during cheese ripening by the action of lipases and esterases indicated that lipolysis can reach up to 20% in mould-ripened cheeses, such as Camembert [111]. Due to the presence of several lipases and esterases encoded by C. variabile, it is likely that this species may also contribute to cheese flavor by producing free fatty acids from lipids and triacylglycerides and by generating volatile compounds with low olfactory threshold values in cheese flavor perception.

Furthermore, the detected regulatory pattern for relevant metabolic pathways led to the conclusion that almost all major processes occurring in C. variabile are regulated in a hierarchical manner. We identified local regulators, whose corresponding genes are clustered with their proposed target genes, for instance the tauR gene that is located adjacent to the tauDCBA cluster involved in uptake and assimilation of taurine, gabR genes located upstream of gabT(D)P genes responsible for GABA utilization, as well as siuR and fiuR genes that are embedded in iron-siderophore and ferric iron uptake gene clusters and are all encoding members of the AraC family of transcription regulators. In addition, DNA-binding sites of master regulators known from C. glutamicum were detected in the C. variabile genome sequence. McbR binding sites, for instance, indicate the connection of the tauABCD genes to the sulphur regulon of C. variabile, whereas DtxR binding sites within the siu and fiu gene clusters indicate the integration of the respective genes into an iron-dependent gene-regulatory network. The presence of GlxR binding sites support the view of a hierarchical regulation of several prominent aspects of the C. variabile metabolism, including the utilization of γ-amino butyric acid, lipolysis, and β-oxidation.

**Conclusions**

The annotation of the C. variabile DSM 44702 genome in conjunction with comparative analysis provides evidence that a strong selective pressure exists on the surface of smear-ripened cheeses, concerning the ability to catabolize substrates, such as γ-amino butyric acid, lactate, lipids, and amino acids (mainly proline), and the presence of efficient iron acquisition and salt-tolerance systems. Our analysis of the genome data revealed that C. variabile is strongly adapted to this habitat and contributes to the main aspects of cheese quality, such as flavor formation. Future studies will certainly benefit from the comprehensive genomic information on C. variabile by integrating this knowledge into models describing the complex cheese ecosystem, which is composed of various bacterial species and yeasts and their microbial interactions.

**Methods**

**Bacterial strain and growth conditions**

C. variabile DSM 44702 (CIP 107183; JCM 12073; LMG S-19265; NCIMB 30131) was originally isolated from the surface of an Irish farmhouse smear-ripened cheese [4].
The strain used in this study was obtained as a lyophilized culture from the German Culture Collection DSMZ (Braunschweig, Germany). *C. variabile* DSM 44702 was formerly described as the type strain of the species *C. mooreparkense*, which turned out to be a later heterotypic synonym for *C. variabile* [4,20]. *C. variabile* DSM 44702 was routinely cultivated in BYT complex medium [161] at 30°C.

**Sequencing of the *C. variabile* DSM 44702 genome**
Genomic DNA was purified from *C. variabile* cells by an alkaline lysis procedure [162], using 20-ml aliquots of an overnight culture grown in liquid BYT complex medium that was supplemented with 1.25% (w/v) glycine. The alkaline lysis protocol applied in this study includes two minor modifications: (i) The *C. variabile* cells were incubated in a 30 mg/ml lysozyme solution at 37°C for 1 h. (ii) The harvested *C. variabile* cells were lysed in 0.7 ml 10% (w/v) SDS solution at 37°C for 15 min. A total amount of 5 μg of purified genomic DNA was used for constructing a single-stranded template DNA library. The DNA concentration of the library was measured by using the Agilent RNA 6000 Nano Kit. The preparation and sequencing of the DNA library were performed according to standard protocols from Roche Applied Science. Pyrosequencing of the genomic DNA sample was carried out with the Genome Sequencer FLX System and Titanium chemistry (Roche Applied Science). The sequence reads were assembled with the GS Assembler Software (version 2.3).

**Gap closure of the *C. variabile* DSM 44702 genome**
The gap closure was supported by the related reference contig arrangement tool r2cat, using the complete *C. jejenum* K411 genome sequence as a reference [21]. The remaining gaps in the genome sequence of *C. variabile* were closed by PCR with Phusion hot start high-fidelity DNA polymerase (Finnzymes) and genomic template DNA. All primers used in this study were synthesized by Metabion. The PCR assays were carried out with a TProfessional PCR thermocycler (Biometra) according to standard protocols (Finnzymes). Amplified DNA fragments linking individual contigs were sequenced by IIT Biotech. All DNA sequences were uploaded into the Consed program [22] and manually inspected to generate the complete genome sequence of *C. variabile* DSM 44702.

**Annotation and bioinformatic analysis of the *C. variabile* DSM 44702 genome**
The assembled genome sequence of *C. variabile* was uploaded into the bacterial genome annotation system GenDB [23]. The annotation of the *C. variabile* genome sequence was performed as described previously [34], followed by manual curation. A combined gene prediction strategy was executed by using GLIMMER 2.1 and the CRITICA program suite [163] in conjunction with post-processing by the RBSFinder tool [164]. Additional regional information on protein-coding regions was obtained with REGANOR [165] and integrated into the final genome annotation. The deduced proteins of *C. variabile* were functionally characterized by automated searches in public databases according to default search parameters of the GenDB system [23]. The origin of chromosomal replication of *C. variabile* was predicted with the web version of the Ori-Finder [25]. The genome plot of *C. variabile* was generated with the DNAPlotter tool [166]. Comparative analyses of the predicted gene content and reconstructions of metabolic properties of *C. variabile* were accomplished by the computer programs EDGAR [30] and CARMEN [167]. The software tool CARMEN supports the in silico reconstruction and visualization of automatically derived metabolic networks based on pathway information from the KEGG database and user-defined SBML templates. The visualization and manual curation of pathway maps was performed with the CellDesigner software (version 3.2) [168]. The synteny between the genomes of *C. variabile* DSM 44702 and *C. jejenum* K411 was calculated with reciprocal best BLASTP matches by the EDGAR software [30]. The detection and classification of the transcriptional regulatory repertoire of *C. variabile* was performed as described previously [76]. The DNA motif discovery for regulon predictions followed a combined workflow, using both position weight matrices and hidden Markov models [96]. The transcriptional regulatory network transfer between corynebacteria on the genome-scale was described in detail previously [82].

The annotated genome sequence of *C. variabile* DSM 44702 has been deposited in the GenBank database with accession number CP002917 and is also available from the RefSeq database with accession number NC_015859.

### Additional material

**Additional file 1: Pathways involved in the central metabolism of *C. variabile* DSM 44702**
The PDF contains a reconstructed pathway map of the central carbohydrate metabolism.

**Additional file 2: Pathways involved in the biosynthesis of vitamins and cofactors by *C. variabile* DSM 44702**
The PDF contains a reconstructed pathway map of vitamin and cofactor biosynthesis.

**Additional file 3: Pathways involved in amino acid metabolism of *C. variabile* DSM 44702**
The PDF contains a reconstructed pathway map of amino acid biosynthesis and amino acid transport.

### Acknowledgements

The valuable contribution of the CeBiTec genome sequencing team (A. Albersmeier, K. Brinkrolf, and R. Szczepanowski) to this genome project is greatly acknowledged. The authors thank S. Jaenicke and J. Schneider (IBF/ CeBiTec) for implementing and maintaining the *C. variabile* GenDB project and for additional bioinformatic support. The authors acknowledge support...
of the publication fee by Deutsche Forschungsgemeinschaft and the Open Access Publication Funds of Bielefeld University.

Author details
Institut für Genomforschung und Systembiologie, Centrum für Biotechnologie, Universität Bielefeld, Universitätsstraße 27, D-33615 Bielefeld, Germany. 
CLIB Graduate Cluster Industrial Biotechnology, Centrum für Biotechnologie, Universität Bielefeld, Universitätsstraße 27, D-33615 Bielefeld, Germany.

Authors’ contributions
JS sequenced and annotated the C. variabile genome and prepared the manuscript. IM participated in the gap closure process. ET and AT participated in data evaluation. AT supervised the project. All authors read and approved the final version of the manuscript.

Received: 24 August 2011 Accepted: 3 November 2011
Published: 3 November 2011

References
1. Fox PF, Guineau T, Cogan T, McGweeney P. Fundamentals of cheese science. Berlin: Springer; 2000.
2. Beresford T, Fitzsimons N, Brennan N, Cogan T. Recent advances in cheese microbiology. Int Dairy J 2001, 11:259-274.
3. Reps A. Bacteria surface-riched cheeses. In: Cheese. chemistry, physics and microbiology. Volume 2. Edited by: Fox P. London: Aspen Publishers; 1993:137-172.
4. Viljoen BC. Corynebacterium resistens sp. nov., isolated from the surface of a smear-riched cheese. Int J Food Microbiol 2001, 57:651-660.
5. Larpin-Laborde S, Imran M, Bonaiti C, Bora N, Gelsomino R, Goerges S, Larkin R, Farrell R, Irlinger F, Bimet F, Delettre J, Lefevre M, Grimont PA. Actinobacteria Genomics of coryneform bacteria from the surface microflora of brick cheeses. Int J Food Microbiol 2001, 69:3713-3717.
6. Reps A. Bacteria surface-riched cheeses. In: Cheese. chemistry, physics and microbiology. Volume 2. Edited by: Fox P. London: Aspen Publishers; 1993:137-172.
7. Fox PF, Guineau T, Cogan T, McGweeney P. Fundamentals of cheese science. Berlin: Springer; 2000.
8. Beresford T, Fitzsimons N, Brennan N, Cogan T. Recent advances in cheese microbiology. Int Dairy J 2001, 11:259-274.
9. Viljoen BC. Corynebacterium resistens sp. nov., isolated from the surface of a smear-riched cheese. Int J Food Microbiol 2001, 57:651-660.
10. 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-riched cheese. Int J Food Microbiol 2003, 71:843-852.
11. Larpin-Laborde S, Imran M, Bonaiti C, Bora N, Gelsomino R, Goerges S, Irlinger F, Goodfellow M, Ward AC, Vancanneyt M, et al. Surface microbial consortia from Livarot, a French smear-riched cheese. Can J Microbiol 2011, 57:651-660.
12. van de Perre I, De Smet M, Hoste B, Scherer S, Swings J, Fitzgerald GF, Cogan TM. Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat Genet 2003, 34:377-378.
13. Yawata H, Osumasaba CA, Nonaka H, Kos P, Okai N, Suzuki N, Suda M, Tsuge Y, Watanabe J, Ikeda Y, et al. Comparative analysis of the Corynebacterium glutamicum group and complete genome sequence of strain R. Microbiology 2007, 153:1042-1058.
14. Suzuki N, Okai N, Nonaka H, Tsuge Y, Inui M, Yawata H. High-throughput transposon mutagenesis of Corynebacterium glutamicum and construction of a single-gene disruptant mutant library. Appl Environ Microbiol 2006, 72:3750-3755.
15. Blom J, Albaum SP, Doppmeier D, Pühler A, Vorholter FJ, Zakrzewski M, Goessmann A. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC Bioinformatics 2005, 10:154.
16. Tauch A, Trent E, Tilker A, Ludewig U, Schreier S, Goessmann A, Arnold W, Binkel T, Brinkkof K, Brune L, et al. The lifestyle of Corynebacterium urealyticum derived from its complete genome sequence established by pyrosequencing. J Bacteriol 2008, 190:11-21.
17. Summer EI, Liu M, Gill JJ, Grant M, Chain-Coates TN, Ferguson L, Janes C, Lange K, Bortoli M, Moore C, et al. Genomic and functional analyses of Rhodococcus equi phages ReqIPepys, ReqIPoco6, ReqIPine5, and ReqIDocB7. Appl Environ Microbiol 2011, 77:669-683.
18. Dorsch J, Klumpp J, Beilmann R, Schmelcher M, Born Y, Zimmer M, Calendar R, Loessner MJ. Comparative genome analysis of Listeria bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site. J Bacteriol 2009, 191:2706-2715.
19. Tauch A, Kaiser O, Hain T, Goessmann A, Weisshaar B, Albersmeier A, Binkel T, Bischoff N, Brune L, Jaekraborty T, et al. Complete genome sequence and analysis of the multiresistant nosocomial pathogen Corynebacterium jeikeium K411, a lipid-requiring bacterium of the human skin flora. J Bacteriol 2005, 187:4671-4682.
20. Bott M. Offering surprises: TCA cycle regulation in Corynebacterium glutamicum. Trends Microbiol 2007, 15:417-425.
21. Mann DA, Beuchat LR. Combinations of antimycotics to inhibit the growth of moulds capable of producing 1, 3-pentadiene. Food Microbiol 2008, 25:144-153.
22. Jolker E, Eder D, Ballan S, Kramer R, Eikmanns BJ, Marin K. Identification and characterization of a bacterial transport system for the uptake of pyruvate, propionate, and acetate in Corynebacterium glutamicum. J Bacteriol 2009, 191:940-948.
23. Aykol I, Serdaroglu K, Gezgin Y, Dailoglu K, Delic M, Oziek E. Redirection of pyruvate pathway of lactic acid bacteria to improve cheese quality. Food Biotechnology 2009, 23:200-213.
24. Lowe TM, Eddy SR. rRNA classification and phylogenetic analysis of the anaerobic succinate degradation pathway in Clostridium kluvleri. J Bacteriol 1996, 178:871-880.
41. Bott M, Niebisch A: The respiratory chain of Corynebacterium glutamicum. J Biotechnol 2003, 104:129-153.
42. Kana BD, Weinstien EA, Avarkoch D, Dawes SS, Rubin H, Mizrahi V: Characterization of the cydA-encoded cytochrome bd oxidase from Mycobacterium smegmatis. J Bacteriol 2003, 185:7076-7086.
43. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Hennissen B: The Carbohydrate-Active EnZymes database (CAZY): an expert resource for glycoconjugates. Nucleic Acids Res 2009, 37:D233-238.
44. Eddy SR: Profile hidden Markov models. Bioinformatics 1998, 14:755-763.
45. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Ponnution J-E, Gavin OL, Gusnarekan P, Ceric G, Forslund K, et al: The Pfam protein families database. Nucleic Acids Res 2010, 38:D211-222.
46. Henrissat B, Davies GJ: The glycoside hydrolase family database. Bioinformatics 2004, 20:437-444.
47. Coutinho PM, Deleury E, Davies GJ, Henrissat B: A hierarchical classification of polysaccharide lyases for glycogenomics. Nucleic Acids Res 2004, 32:D211-222.
48. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Hennissen B: A hierarchical classification of polysaccharide lyases for glycanomics. Biochem J 2010, 432:437-444.
49. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ: A new twist on an old pathway for protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. Mol Microbiol 2002, 45:943-950.
50. Hansmeier N, Chao T-C, Kalinowski J, Pühler A: M. tuberculosis: A notation of free amino acids in cheeses from the Czech market. Czech J Food Sci 2009, 27:143-150.
51. Tauch A, Hermann T, Burkovski A, Krämer R, Pühler A, Kalinowski J: Isolucine uptake in Corynebacterium glutamicum ATCC 13032 is directed by the bmg gene product. Arch Microbiol 1998, 169:303-312.
52. Seep-Feldhaus AH, Kalinowski J, Pühler A: Molecular analysis of the Corynebacterium glutamicum bmg gene involved in lysine uptake. Mol Microbiol 1991, 5:2995-3005.
53. Zhao Z, Ding JY, Li T, Zhou NY, Liu SJ: The nagA1108 (PheP (Cg)) gene encodes a new L-Phe transporter in Corynebacterium glutamicum. Appl Microbiol Biotechnol 2011, 90:2005-2013.
54. Wehmann A, Morakabati S, Krämer R, Sahn H, Egeleng L: Functional analysis of sequences adjacent to dapE of Corynebacterium glutamicum reveals the presence of araoP, which encodes the aromatic amino acid transporter. J Bacteriol 1995, 177:5991-5993.
55. Pohlschröder M, Koonin EV, Lipman DJ: Application to complete genomes. J Mol Biol 2001, 305:567-580.
56. Rigel NW, Braunstein M: A neural network as a functional indicator: the case of the AraC/XylS family of transcriptional regulators. J Biotechnol 2009, 143-151.
57. Rodionov DA: Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. Chem Rev 2007, 107:3467-3497.
58. Brinkoff T, Brune I, Tauch A: The individual and common repertoire of DNA-binding transcriptional regulators of Corynebacterium glutamicum, Corynebacterium diphtheriae and Corynebacterium jeikeium deduced from the complete genome sequences. BMC Genomics 2005, 6:86.
59. Ramos JL, Martinez-Bueno M, Molina-Henares AJ, Teran W, Watanabe K, Zhang X, Gallegos MT, Brennan R, Tores B: The TetR family of transcriptional repressors. Microbiol Mol Biol Rev 2009, 69:526-556.
60. Brinkoff T, Brune I, Tauch A: The transcriptional regulatory network of the amino acid producer Corynebacterium glutamicum. J Bacteriol 2007, 129:191-211.
61. Ibarra JA, Perez-Rueda E, Segovia L, Puente JL: The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors. Genetica 2008, 133:65-78.
62. Tauch A, Schneider JB, Szczepanowski R, Tilker A, Viehoever P, Gartnermann KH, Arnold W, Blom J, Brinkoff T, Brune I, et al: Ultrafast phylosequencing of Corynebacterium kroppenstedtii DSM44385 revealed insights into the physiology of a lipophilic corynebacterium that lacks mycolic acids. J Biotechnol 2008, 136:22-30.
63. Foster E, Ott L, Schneider J, Schroder J, Jemnicke S, Goessmann A, Husmann P, Stoye J, Dorela F, Rocha FS, et al: The complete genome sequence of Corynebacterium pseudotuberculosis FRC41 isolated from a 12-year-old girl with necrotising lymphadenitis reveals insights into gene-regulatory networks contributing to virulence. BMC Genomics 2010, 11:728.
64. Schröder J, Tauch A: Transcriptional regulation of gene expression in Corynebacterium glutamicum: the role of global, master and local regulators in the modular and hierarchical gene regulatory network. FEMS Microbiol Rev 2010, 34:685-757.
65. Schröder J, Tauch A: Transcriptional regulation of gene expression in Corynebacterium glutamicum, Corynebacterium diphtheriae and Corynebacterium jeikeium deduced from the complete genome sequences. BMC Genomics 2005, 6:86.
86. Nomura M, Kimoto H, Someya Y, Funakawa S, Suzuki I: Production of gamma-aminobutyric acid by cheese starters during cheese ripening. J Dairy Sci 1998, 81:1486-1491.
87. McSweeney P, Sousa M: Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. Lett 2000, 80:293-324.
88. De Figueiredo R, Gerutti de Guglielmino G, Betino de Cardenas I, Oliver G: Flavour production and citrate metabolism in Lactococcus rhamnosus ATCC 7469. Milchwissenschaft 1998, 53:617-619.
89. Dinnan DF, Robin S, Cogan TM: Citric acid metabolism in hetero- and homofeitelomorphic lactic acid bacteria. Appl Environ Microbiol 1976, 31:481-486.
90. Hugenholtz J: Citrate metabolism in lactic acid bacteria. FEMS Microbiol Rev 1993, 12:165-178.
91. Manirley L, Casey MG: Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. Int J Food Microbiol 2004, 90:139-159.
92. Brocker M, Schaffer S, Mack C, Bött M: Citrate utilization by Corynebacterium glutamicum is controlled by the CitAB two-component system through positive regulation of the citrate transport genes citH and citCBA. J Bacteriol 2009, 191:3869-3880.
93. Kohl TA, Tauch A: The GlnR regulon of the amino acid producer Corynebacterium glutamicum: Detection of the coregulatory region and integration into the transcriptional regulatory network model. J Biotechnol 2009, 143:239-246.
94. Brennan NM, Cogan TM, Loessler N, Scherer S: Bacterial surface-reipened cheeses. In Cheese. Chemistry, Physics, and Microbiology. Volume 2. Edited by Fox PF, McSweeney PL, Cogan TM, Guinee T. Amsterdam: Elsevier Academic Press; 2004:199-222.
95. Mounier J, Rea MC, O’Connor PM, Fitzgerald GF, Cogan TM: Growth characteristics of Brevibacterium, Corynebacterium, Microbacterium, and Staphylococcus spp. isolated from surface-reipened cheese. Appl Environ Microbiol 2003, 70:7733-7740.
96. Kohl TA, Baumbach J, Jungwirth B, Pühler A, Tauch A: The GlnR regulon of the amino acid producer Corynebacterium glutamicum: in silico and in vitro detection of DNA binding sites of a global transcription regulator. J Biotechnol 2008, 135:340-350.
97. Tripaldi C, Martilli F, Terramoccia S: Content of taurine and other free amino acids in milk of goats bred in Italy. Small Ruminant Res 1998, 30:127-136.
98. Erbersdobler HF, Trautwein E, Greulich HG: Determinations of taurine in milk and infant formula diets. Eur J Pediatr 1984, 142:133-134.
99. Rückert C, Mike J, Albersmeier A, Koch DJ, Pühler A, Kalinowski J: The dual transcriptional regulator CysR in Corynebacterium glutamicum ATCC13032 controls a subset of genes of the McbR regulon in response to the availability of sulphide acceptor molecules. BMC Genomics 2008, 9:483.
100. Koch DJ, Rückert C, Albersmeier A, Huser AT, Tauch A, Pühler A, Kalinowski J: The transcriptional regulator SsuR modulates expression of the Corynebacterium glutamicum sulphite utilization genes in the absence of sulphate. Mol Microbiol 2005, 58:480-494.
101. Rey DA, Nentwich SS, Koch DJ, Rückert C, Pühler A, Tauch A, Kalinowski J: The McbR repressor modulated by the effector substance S-adenosylhomocysteine controls directly the transcription of a region involved in sulphur metabolism of Corynebacterium glutamicum ATCC 13032. Mol Microbiol 2005, 56:871-887.
102. Brune I, Baranthy N, Klotzeli M, Jones J, James G, Tauch A: Identification of McbR as transcription regulator of aceD and genes involved in methionine and cysteine biosynthesis in Corynebacterium jeikeium K411. J Biotechnol 2011, 151:22-29.
103. Ueno H: Enzymatic and structural aspects on glutamate decarboxylase. J Mol Catal B Enzym 2000, 16:707-79.
104. Bovin AW, Shelp BJ: The Metabolism and Functions of (gamma)-Aminobutyric Acid. Plant Physiol 1997, 115:1-5.
105. Mavromati K, Chu K, Ivanova N, Hooper SD, Markowitz VM, Kyrpides NC: Gene context analysis in the Integrated Microbial Genomes (IMG) data management system. PLoS One 2009, 14:e97997.
106. Middeldorp SE, Ozyurt PC: Structure and function of the LysR-type transcriptional regulator (LTTTR) family proteins. Microbiology 2008, 154:3609-3632.
107. Kyriakidis N, Tarantili-Georgiou K, Tsani-Batzaka E: Nitrate and nitrite content of greek cheeses. Journal of Food Composition and Analysis 1997, 10:343-349.
108. Hirai T, Subramaniam S: Structure and transport mechanism of the bacterial oxalate transporter OstX. Biophys J 2004, 87:3600-3607.
109. Center of Disease Control and Prevention: Effects of switching from whole to low-fat/fat-free milk in public schools – New York city, 2004-2009. Morb Mortal Wkly Rep 2010, 59:70-73.
110. Monnet C, Lou V, Gibrat F, Spinelli E, Barbe V, Vacherie B, Gavroy F, Gourbeyre E, Siguier P, Marrachi H, Daffe M, Queunard A: The Pks13/FadD32 crosstalk for the biosynthesis of mycolic acids in Mycobacterium tuberculosis. J Biol Chem 2009, 284:19255-19264.
111. Black PN, DiRusso CC: Transmembrane movement of exogenous long-chain fatty acids: proteins, enzymes, and vectorial esterification. Microbiol Mol Biol Rev 2003, 67:454-472.
112. Hiltunen JK, Oin Y: Beta-oxidation – strategies for the metabolism of a wide variety of acyl-CoA esters. Biochim Biophys Acta 2000, 1484:117-128.
113. Bayan N, Houssin C, Chami M, Leblon G: Mycomembrane and S-layer: two important structures of Corynebacterium glutamicum cell envelope with promising biotechnology applications. J Biotechnol 2003, 105:85-87.
114. Portevin D, de Sousa-D’Auria C, Houssin C, Grimaldi C, Chami M, Daffe M, Guilhot C: A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. Proc Natl Acad Sci USA 2004, 101:314-319.
115. Gande R, Gibson KJ, Brawn AK, Krumback K, Dover LG, Sahm H, Shirayama S, Oikawa T, Bens G, Egelberg L: Acyl-CoA carboxylases (accc2 and accc3), together with a unique polyketide synthase (Cg-pks), are key to mycolic acid biosynthesis in Corynebacteriaceae such as Corynebacterium glutamicum and Mycobacterium tuberculosis. J Biol Chem 2004, 279:44847-44857.
116. Portevin D, de Sousa-D’Auria C, Montraizer H, Houssin C, Stella A, Lanelle MA, Bardou F, Guilhot C, Daffe M: The acyl-AMP ligase FadD32 and AccD4-containing acyl-CoA carboxylase are required for the synthesis of mycolic acids and essential for mycobacterial growth: identification of the carboxylation product and determination of the acyl-CoA carboxylase components. J Biol Chem 2005, 280:8862-8874.
117. Baranthy N, Brune I, Tauch A: Molecular basis of human body odour formation: insights deduced from corynebacterial genome sequences. Int J Chem Ecol 2011.
118. Deetee P, Bonnarme P, Spinnler HE, Helinck S: Production of volatile aroma compounds by bacterial strains isolated from different surface-reipened French cheeses. Appl Microbiol Biotechnol 2007, 76:1161-1171.
119. Gobbetti M, Smacchi E, Semeraro M, Fox PF, Lanciotti R, Cogan T: Purification and characterization of an extracellular proline iminopeptidase from Corynebacterium variabilis NCD0 2101. J Appl Microbiol 2001, 90:649-656.
120. Ling M, Allen SW, Wood JM: Sequence analysis identifies the proline dehydrogenase and delta-1-pyrroline-5-carboxylate dehydrogenase domains of the multifunctional Escherichia coli polA protein. J Mol Biol 1994, 243:950-956.
121. Jochmann N, Kurze AK, Czaja LF, Brinkrolf K, Brune I, Huser AT, Hansmeier N, Pühler A, Borovok I, Tauch A: Genetic makeup of the Corynebacterium glutamicum LexA regulon deduced from comparative transcriptomics and in vitro DNA band shift assays. Microbiology 2009, 155:1459-1477.
122. Kramar R: Osmosensing and osmosignaling in Corynebacterium glutamicum. Amino Acids 2009, 37:487-497.
123. Klein J, Schwaar T, Lentzen G: Ectoine as a natural component of food: detection in red smerge cheeses. J Dairy Res 2007, 74:446-451.
124. Follmann M, Becker M, Ochrombel I, Oettl V, Kramer R, Mann K: Potassium transport in Corynebacterium glutamicum is facilitated by the putative channel protein CgKII, which is essential for pH homeostasis and growth at acid pH. J Bacteriol 2009, 191:2944-2952.
