# *Corynebacterium glutamicum* chassis C1*: Building and testing a novel platform host for synthetic biology and industrial biotechnology

Meike Baumgart*[1], Simon Unthan[2], Ramona Kloß[2], Andreas Radek[2], Tino Polen[1], Niklas Tenhaef[1], Moritz Fabian Müller[2], Andreas Küberl[1], Daniel Siebert[3], Natalie Brühl[4], Kay Marin[5], Stephan Hans[5], Reinhard Krämer[4], Michael Bott[1], Jörn Kalinowski[6], Wolfgang Wiechert[2,7], Gerd Seibold[3], Julia Frunzke[1], Christian Rückert[6], Volker F. Wendisch[8], Stephan Noack*[2]

Institute of Bio- and Geosciences, IBG-1: Biotechnology, Systemic Microbiology[1] and Systems Biotechnology[2], Forschungszentrum Jülich, Jülich, Germany

[3]Institute for Microbiology and Biotechnology, Ulm University, Ulm, Germany

[4]Institute of Biochemistry, University of Cologne, Cologne, Germany

[5]Evonik Nutrition & Care GmbH, Halle/Westphalia, Germany

[6]Microbial Genomics and Biotechnology, Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

[7]Computational Systems Biotechnology, RWTH Aachen University, Aachen, Germany

[8]Chair of Genetics of Prokaryotes, Faculty of Biology & CeBiTec, Bielefeld University, Bielefeld, Germany

*Correspondence:

Dr. Meike Baumgart and Dr.-Ing. Stephan Noack, Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, 52428 Jülich, Germany

E-mail: m.baumgart@fz-juelich.de; s.noack@fz-juelich.de

## Abstract

Targeted top-down strategies for genome reduction are considered to have a high potential for providing robust basic strains for synthetic biology and industrial biotechnology. Recently, we created a library of 26 genome-reduced strains of *Corynebacterium glutamicum* carrying broad deletions in single gene clusters and showing wild-type-like biological fitness. Here, we proceeded with combinatorial deletions of these irrelevant gene clusters in two parallel orders and the resulting library of 28 strains was characterized under various environmental conditions. The final chassis strain C1* carries a genome reduction of 13.4% (412 deleted genes) and shows wild-type-like growth behavior in defined medium with D-glucose as carbon and energy source. Moreover, C1* proves to be robust against several stresses (including oxygen limitation) and shows long-term growth stability under defined and complex medium conditions. In addition to providing a novel prokaryotic chassis strain, our results comprise a large strain library and a revised genome annotation list, which will be valuable sources for future systemic studies of *C. glutamicum*.



**Keywords:** *C. glutamicum*; chassis; genome reduction; synthetic biology; systems biology; RamA

## Introduction

The vast majority of microbial production processes with industrial relevance are currently being developed in three mutually dependent tasks. First, an appropriate producer organism is selected from a list of well-established workhorses, such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Bacillus subtilis* or *Corynebacterium glutamicum*, for which advanced tools for strain engineering are readily available[1-4]. Second, based upon the whole, native genome of the selected organism the desired metabolic properties for target compound production are systematically introduced and the resulting candidate strains are characterized by combining tools for fast microbial phenotyping[5, 6] with detailed omics analytics[7-9] and integrative modelling approaches[10, 11]. Third, a production process is set up preferring well-defined cultivation conditions to enable further process optimization towards stable and long-term production[12, 13].

However, this three-step workflow often results in unsatisfactory production hosts since the biological input is always suboptimal. This is simply due to the fact that the native metabolism of microorganisms (including the producer strains) has evolved towards a complexity that is required to ensure the host's survival in nature. Hence, a high number of metabolic, regulatory and stress response properties are encoded in the microbial genomes that are unnecessary or even undesired in the context of microbial production in artificial bioreactor environments.

To overcome this limitation, two different strategies for reducing bacterial genomes, namely the *de novo* synthesis of minimal genomes (bottom-up) versus the targeted deletion of dispensable genes (top-down), were proposed and their opportunities and risk critically discussed in several reviews[14-16]. It is common sense that the bottom-up strategy is preferred to gain a fundamental understanding of the basic principles of autonomous cellular life. The Craig Venter group followed this strategy and created a minimal cell on the basis of symbiotic prokaryotes such as *Mycoplasma sp.*[17]. By contrast, top-down approaches are the method of choice for the construction of chassis hosts that are genetically stable, robust against environmental stresses and maintain at least the biological fitness (e.g. growth performance) of their ancestor strain under certain defined conditions. Clearly, these features of potential chassis are mandatory for their application in industrial biotechnology. However, they inevitably lead to a high number of so-called "relevant" genes that need to be retained in the final

strain to maintain wild-type-like performance, but are not necessarily essential for its survival[18]. Unfortunately, due to the still limited knowledge (including *E. coli* as the best studied prokaryotic model organism for biotechnology), such relevant gene sets include several genes with unknown biological function. Noteworthy, the first *de novo* creation of a minimal *Mycoplasma mycoides* cell (JCVI-syn3.0) is also based on a set of 149 (31.5% of the genome) "quasi-essential" genes with undiscovered functions essential for life[17].

Almost six years ago we initiated the top-down construction of a genome reduced chassis organism based on *Corynebacterium glutamicum* ATCC 13032 and in the meantime derived 26 genome reduced strains carrying broad deletions of gene clusters irrelevant for growth on defined medium[18]. We suggested that up to 722 kbp (22%) of the wild-type genome (3.28 Mb) can be eliminated in a straightforward manner by further combinatorial deletions of the identified irrelevant gene clusters. By successfully deleting five different pairs of gene clusters by double homologous recombination we demonstrated a proof-of-concept. However, some combinatorial gene cluster deletions were found to be incompatible, likely due to an accidental elimination of all genes exhibiting the same essential function, which have not been characterized to date. Interestingly, apart from genes encoding relevant metabolic functions, Martínez-García et al.[14] also pointed out that there might be a level of tridimensional cellular organization mediated by specific DNA sequences and the deletion (or non-synthesis) of such genetic pieces could have a similar negative effect on the biological fitness of the chassis host (or minimal cell).

In this study, we continued our work towards a *C. glutamicum* chassis by constructing genome-reduced strains with various combinatorial deletions of irrelevant gene clusters. In contrast to other minimal genome projects aiming at a minimal cell, we want to construct chassis strains for basic research and biotechnological application with a reduced complexity while maintaining the same growth behavior and robustness like the parental strain under defined conditions. A detailed characterization of intermediate pre-chassis under different environmental conditions revealed some genes whose deletion led to altered cell morphology, but the majority of genes was found to be dispensable under all test conditions. Finally, we demonstrate the successful construction of the novel chassis *C. glutamicum* C1*, which carries a genome reductions of 13.4% (412 deleted genes). C1*

shows a biological fitness comparable to the wild-type strain, making it highly interesting as future platform strain for basic research and industrial bioprocess development.

# Results and Discussion

## Combinatorial deletion of irrelevant gene clusters

In our previous study we had identified 26 gene clusters irrelevant for growth under standard conditions[18], which we wanted to combine to continue our work towards a *C. glutamicum* chassis. For the combinatorial deletions we started with different strains carrying single irrelevant cluster deletions distributed along the whole genome and introduced additional cluster deletions sequentially in the order as they appeared in the genome. All deletions were performed by homologous recombination. After each deletion cycle, the strains were immediately analyzed using established fast phenotyping methods[18] to detect combinatorial dead ends and guide the ongoing deletion process. With this strategy, we had several lines of combinatorial deletions. In case we reached a point where an additional cluster deletion was not possible or the performance of the resulting strain was worse compared to the wild type, we could simply go one step back or continue with another line.

Figure 1 shows a selection of eight genome-reduced strains (GRS) with partially overlapping combinatorial deletions. The overall genome reduction of the tested strains ranged between 7.7% and 12.6% with respect to the wild-type genome (Figure 1A), including 6% resulting from the prophage deletions[19]. The biomass yield and maximum specific growth rate of all strains was determined on CGXII medium with D-glucose as sole carbon and energy source and compared to the wild type (Figure 1B and C).

One successful line of combinatorial deletions was started from cluster cg2312-cg2322 and finally resulted in the first pre-chassis PC1. During characterization, none of the tested GRS showed an impaired biological fitness but, in fact, the maximum growth rate was even slightly higher compared to the wild type. In particular, strain PC1 with a total genome reduction of 8.8% grew at $\mu_{max} = 0.48 \pm 0.01\ h^{-1}$ in the BioLector® (a microtiter plate-based cultivation device), whereas the wild type grew at $\mu_{max} = 0.44 \pm 0.02\ h^{-1}$. This finding supports the general applicability of the chosen approach to combine deletions of individually irrelevant gene clusters and, thus, enable the stepwise construction of a chassis.

By contrast, the strain GRS48_52_53, which was constructed by deletion of three irrelevant clusters, of which cluster cg3102-cg3111 had already been deleted in PC1, showed a significantly lower growth rate ($\mu_{max}$ = 0.36 ± 0.02 $h^{-1}$). At least two of the three tested clusters seem to be interdependent and cannot be deleted simultaneously without affecting the biological fitness of *C. glutamicum*. Consequently, the two clusters cg3263-cg3301 and cg3324-cg3345 were excluded from the further chassis construction.

A further line of combinatorial deletions was started along the genome from cluster cg0414-cg0440 and finally resulted in the second pre-chassis PC2. None of the tested GRS showed an impaired biological fitness. Most importantly, the strain PC2, which displayed the hitherto largest genome reduction of all constructed GRS (12.6%), grew at $\mu_{max}$ = 0.47 ± 0.02 $h^{-1}$, which is as fast as the wild-type strain.

At this stage, the pre-chassis PC1 and PC2 were chosen for further in-depth characterizations under different cultivation conditions to check to what extent both strains maintained the properties of the originating wild type.

## Phenotypic characterization of pre-chassis (PC)

With the aim to evaluate the robustness of PC1 and PC2 both strains were subjected to nitrogen and phosphate limitation as well as osmotic stress conditions in a series of BioLector experiments. Nitrogen and phosphate limitation are typically applied in fed-batch production processes with *C. glutamicum* to decouple growth from production. Osmotic stress is induced by high concentrations of nutrients and products in the culture broth and can negatively affect product yield and productivity. In particular, *C. glutamicum* is known to effectively counterbalance hyperosmotic stresses by various mechanisms[20, 21] and this beneficial property should be retained in a final chassis. PC1 and PC2 were cultivated over a broad range of ammonium sulfate, dipotassium phosphate and sodium chloride concentrations, and no altered responses to the applied stress conditions compared to the wild type were observed (Figure S1).

Next, both pre-chassis and the wild type were grown with carbon sources other than D-glucose, including hexoses, disaccharides, sugar alcohols as well as organic acids (Figure 2). Interestingly, strain PC1 grew significantly faster with D-glucose, acetate and D-maltose compared to the wild type,

while no altered phenotype was found during growth with the remaining C-sources (Figures 2A and B). By contrast, strain PC2 grew significantly slower with acetate as sole carbon source. As the two known genes for acetate metabolism to acetyl-CoA via acetyl-phosphate (acetate kinase, cg3047 and phosphotransacetylase, cg3048) are not deleted in this strain, this effect might be due to the deletion of so far unknown genes relevant for the utilization of this C-source. As observed before, the maximum growth rate of PC2 with D-glucose was slightly higher compared to the wild type. However, the final biomass (determined as cell dry weight (CDW)) of PC2, was significantly lower, which was less obvious according to the backscatter signal (Figure 1). The latter result suggests an altered cell morphology of PC2, which was further investigated during the subsequent lab-scale bioreactor cultivations.

PC1 and PC2 were cultivated in lab-scale bioreactors to evaluate their growth properties under controlled conditions and by making use of different biomass measurements (Table 1). While the pre-chassis PC1 grew comparable to the wild type, the maximum growth rate of PC2 based on cell number measurements increased by 26% (Table 1). Moreover, in all bioreactor experiments with PC2 occurred an increased foaming and a thick film of cells was attached to the inner reactor wall after the exponential growth phase (Figure S2). The removal of cells from the liquid phase during foaming also contributed to the significantly lower estimates for the biomass yield of PC2 (Table 1). To narrow down the effect on the morphology of PC2 to one or just a few gene clusters, another series of bioreactor experiments with different ancestor strains of PC2 was carried out (see Supporting Information 1 for more details). As a result, the deletion of gene cluster cg0414-cg0440 was identified as cause for the observed effect on cell morphology.

The slightly larger gene cluster cg0414-cg0443 has a significantly lower GC-content than the median of the whole genome and it is absent from the genome of the closely related *C. glutamicum* R strain, suggesting that it was acquired by a recent horizontal gene transfer[22, 23]. Many genes of this cluster are annotated as being involved in cell wall and lipopolysaccharide biosynthesis but none has been characterized in detail to date. Therefore, one can only speculate about the consequences of their deletion. Several genes of this cluster are also predicted to have redundant functions, for example cg0415 (*ptpA2*, putative protein-tyrosine-phosphatase) and cg2459, cg0422 (*murA*, UDP-N-

acetylglucosamine 1-carboxyvinyltransferase) and cg2829, cg0423 (*murB*, UDP-N-acetylenolpyruvoylglucosamine reductase) and cg0476 as well as cg0435 (*udgA1*, UDP-glucose 6-dehydrogenase) and cg3154. *Bacillus anthracis capD*, which is homologous to cg0417, is required for the anchoring of the capsule to peptidoglycan[24], but *C. glutamicum* does not have a capsule. Homologs of the genes *wzy* (cg0437), *wzx* (cg0421) and *wzz* (cg0414) in other bacteria are known to be involved in the synthesis of cell surface polysaccharides[25], but their role in *C. glutamicum* remains unknown. Although the altered phenotype of PC2 could not be traced back to specific single or multiple genes of the cluster cg0414-0440, there are several interesting candidates, but additional in-depth biochemistry studies are required to characterize their functions. Nevertheless, these results nicely demonstrate how libraries of genome reduced strains can be used to quickly assign an altered phenotype to a comparably small genomic region.

## Building and testing of *C. glutamicum* chassis (C)

As described above, the combinatorial deletion of several gene clusters led to the construction of PC1 and PC2. As a next step, we decided to combine all cluster deletions within one strain starting from both pre-chassis in parallel in case any interdependency occurs (Figure 3A). The construction of C1 was stopped after deletion of cg1291-cg1305 because the gene cluster cg1340-cg1352 could subsequently not be deleted in several attempts. The reason for this could not be clarified yet, but maybe the order of cluster deletions is critical.

The resulting GRS, including the potential chassis C1 and C2, share several deletions and were characterized following our standard phenotyping procedure (Figure 3B and C). Again, some of the GRS originating from PC1 showed slightly higher maximum growth rates on D-glucose medium compared to the wild type. This beneficial effect was not observed for C1, which showed the wild-type growth rate of $\mu_{max} = 0.45 \pm 0.01\ h^{-1}$. By contrast, all GRS originating from PC2 showed significantly lower biomass yields and growth rates. In particular, the biomass yield of C2 was reduced by 13%, which is most likely caused by an altered cell morphology as already discussed above. Moreover, the growth rate of C2 was reduced by 8.9% to $\mu_{max} = 0.41 \pm 0.01\ h^{-1}$. From these results, we can deduce that C2 is not an optimal basic strain for industrial application and therefore we decided to focus on C1 for a more detailed characterization of its genetic and metabolic properties.

The biological fitness of chassis strain C1 was tested on the same set of alternative carbon and energy sources as for the characterization of the pre-chassis (Figure S3). While the phenotype on D-maltose was unaltered compared to the wild type and pre-chassis PC1, significant delays in the initiation of exponential growth were observed for C1 on acetate (90 h), pyruvate (135 h) and D-arabitol (30 h). These effects might either result from the deletion of one or more gene clusters associated to PC2 (carry-over effect, Figure 2) or they were accidentally introduced during the construction of C1 following combinatorial gene cluster deletions. Subsequently, this question was addressed by comparative genome re-sequencing.

## Genome re-sequencing of PC1, PC2 and C1

The genomes of the strains PC2, PC1 and C1 were sequenced not only to confirm the deletion of all target gene clusters, but also to determine if, or how many, additional undesired mutations occurred during the deletion process. The sequences were mapped against the theoretical sequences based on MB001 (GenBank CP005959). Large deletions or insertions were absent, confirming the reliability of the targeted deletion process. Nine (PC2), three (PC1) and eleven (C1) partially overlapping SNPs were found in the respective strains (Table S3). Several of these were silent mutations within genes that do not lead to any change in protein sequence and are therefore not further discussed. Table 2 shows a selection covering only those mutations which may influence physiology.

The mutation A468T in Cg1720 was present in all three strains and it was located outside of the regions selected for deletion. It was presumably already present in CR099 as this is the last common ancestor of PC2 and PC1. This mutation did not appear in MB001[19] so it was likely introduced during the construction of CR099 based on MB001. Cg1720 encodes the ATPase component of an uncharacterized ABC transporter. Therefore, it is difficult to predict the impact of this mutation.

The genome of C1 lacked a single nucleotide in the coding region of the 6C-RNA gene which might be involved in DNA damage response and control of cell division[26]. This deletion is within the second conserved loop resulting in 7 instead of 8 consecutive cytosines. As only 6 of these cytosines are widely conserved and the deletion has likely no impact on structure or folding of this RNA, we do not expect a major effect of this deletion.

Furthermore, the genome of C1 contained a mutation in the promoter-region of regulatory gene *ramA* (cg2831) (Figure 4). The second position of the -10-region (TACACT) was changed to G. As this is one of the two most conserved positions with respect to the consensus sequence of the -10 region (TAnnnT)[27], one would expect a significantly lower transcription of *ramA*. Moreover, this mutation is located in the SugR binding sites overlapping the -10 region (Figure 4). The respective position belongs to the most conserved bases within the SugR binding motif[28], but its relevance has not been studied in the *ramA* promoter. However, it was shown that the *ramA* regulation by SugR is not very strong[29] and is presumably overruled by the effect on the -10 region. As this mutation did not appear in the parental strain PC1, it must have occurred in one of the six subsequent deletion rounds. The effect of this mutation on RamA is discussed further in the transcriptome analysis section. For all other mutations the consequences cannot be predicted as the related genes or promoter sequences have not been studied in detail. The sequence of the final strain C1 including the mutations detected by genome sequencing was deposited in GenBank as CP017995.

## Transcriptome analysis of C1 against MB001

To assess the changes of the C1 transcriptome upon deletion of the several genomic regions, DNA microarray analysis was performed in comparison to MB001. The latter was chosen as reference, because earlier studies had revealed no significant differences between MB001 and ATCC 13032 other than within the prophage regions[19]. Samples were taken in the exponential growth phase, processed and analyzed as described in the methods section. The results are presented in Table 3 and deposited in the GEO database by series entry GSE88717.

Neglecting the deleted clusters, in total 22 genes were found more than three-fold down regulated and 15 genes were more than three-fold up regulated. Four genes appeared more than 10-fold down regulated and three more than 10-fold up regulated. The most strongly up-regulated genes are cg0693 (*groEL*), cg1623 (putative manganese transporter) and cg1696 (putative antibiotic efflux permease of the major facilitator superfamily).

The explanation for the increase of cg1623 is obvious, since we could recently show that this gene is repressed by MntR (Cg0741) which is located within one of the deleted regions[30]. In the case of cg0693 the increased expression can most likely be explained by the deletion of ISCg1c (cg0692),

which was integrated into the *groEL* gene (cg0691 & cg0693), presumably blocking transcription. After deletion, the gene can now be fully transcribed and appears up-regulated in the transcriptome analysis. The up-regulation of cg1696 can be explained by comparison of the genomic context with related species (Figure S5). Next to the homologous gene in *C. efficiens* lays a MarR-type regulator which presumably regulates this gene. The homolog to this regulator in *C. glutamicum* is Cg1211, which has also been deleted in C1 and, thus, could have resulted in the stronger transcription of cg1696. This hypothesis is further supported by the fact that a homologous gene of cg1696 as well as ISCg1a are located next to cg1211. The two genes cg1211 and cg1212 possibly have been transferred to their current location by the action of the IS element.

The genes cg2601and 2602 are located next to the ISCg1d deletion site, which might be the reason for the altered transcription, either by read through from cg2599 or by changes in the promoter due to the amplification of the deletion region from *B. lactofermentum* (For details please see section "Deletion of the IS-elements ISCg1 and ISCg2" in Supporting information 1).

As anticipated from the genome sequencing results, the transcription of *ramA* is more than six-fold reduced in C1, presumably due to the mutation in the promoter region (Figure 4). The LuxR-type transcriptional regulator RamA was originally identified by affinity chromatography with the promoter region of *aceA* and *aceB* (divergent genes) and a *ramA* deletion mutant turned out to be unable to grow on acetate as sole carbon source[31]. RamA is a master regulator of carbon metabolism with currently more than 40 experimentally proven direct target genes[32-34]. The transcriptome analysis presented here was performed with D-glucose as carbon source and already under these conditions a significant influence on genes belonging to the RamA regulon, such as *aceA* (cg2560 encoding isocitrate lyase), *aceB* (cg2559 encoding malate synthase) and *ald* (cg3069 encoding aldehyde dehydrogenase), was obvious (2-4-fold reduction). As RamA is required for the activation of genes involved in acetate metabolism, the effect of the reduced RamA levels on the transcriptome of the C1 strain are likely much more severe with acetate as carbon source, which could explain the poor growth of this strain on acetate.

## Reversion of the mutation in the *ramA* promoter in chassis C1

Genome sequencing as well as transcriptome studies revealed an undesired mutation in the *ramA* promoter region that is likely responsible for the phenotype of chassis C1 on defined acetate medium (Figure S3). To test this hypothesis the mutation was reverted to the wild-type sequence using the same double recombination strategy as for the chromosomal deletions (Supporting Information 1, Strain construction). After the second recombination, several clones could grow on defined CGXII medium agar plates with acetate as sole carbon and energy source and contained the corrected *ramA* promoter. This was the first indication that the mutation on the *ramA* promoter contributed significantly to the inhibited growth of C1 on acetate.

Subsequently, we repeated the growth experiments on the different carbon sources with one single clone (denoted here as C1*) carrying the native *ramA* promoter region (Figure S4). The resulting phenotypes of C1* still showed significant delays in the initiation of exponential growth when compared to the wild-type strain. However, compared to the mutated strain C1, the specific delay times are greatly reduced for nearly all tested carbon sources (acetate: -60%; pyruvate: -95%; D-arabitol: -90%). Taking together, these findings point to hitherto unknown systemic effects resulting from the intensive combinatory deletion of gene clusters during the construction of chassis C1.

Finally, we checked the stability and robustness of chassis C1* during growth in media relevant for bioprocess development (i.e. defined media exemplified by CGXII) and industrial production (i.e. complex media exemplified by BHI). Long-term cultivation experiments with C1* and wild-type ATCC 13032 were performed, utilizing a recently introduced workflow that enables automated repetitive batch operation at microscale[35]. For the CGXII condition, we used 40 g $L^{-1}$ D-glucose and this high amount of carbon introduced a prolonged phase of oxygen limitation and, thus, the specific growth rate significantly decreased along each single batch (Figure 5A). Following model-based data evaluation, comparable average specific growth rates of $\mu_{avg}$ = 0.48 ± 0.01 $h^{-1}$ and $\mu_{avg}$ = 0.49 ± 0.01 $h^{-1}$ were estimated for C1* and the wild type, respectively.

The BHI medium contained smaller amounts of various carbon sources (including amino acids from peptone) and only 2 g $L^{-1}$ D-glucose. Therefore, this media should enable unbalanced growth conditions throughout each batch cycle. Indeed, both strains showed a nearly constant average specific

growth rate of $\mu_{avg}$ = 0.78 ± 0.01 $h^{-1}$ (Figure 5B). Moreover, both strains showed comparably small lag times under both conditions (Supporting Information 1). From these results we can conclude that the chassis C1* can be cultivated stably on defined CGXII and complex BHI medium as well as under oxygen stress conditions without any loss in growth performance.

## Revised *C. glutamicum* genome annotation linked to literature

During our studies we realized that the available genome annotation list of *C. glutamicum* ATCC 13032 is not up to date and lacks some valuable information. Furthermore, the different numbering systems (cg-numbers, NCgl-numbers, Cgl-numbers, R-strain locus tags) complicate the finding of published information related to a certain gene, which is often required for the interpretation of omics data (e.g. from transcriptomics) or to predict the consequence of a specific mutation detected by genome sequencing. In order to simplify these processes a new type of data collection was required. Therefore, we went through the whole annotation list and collected digital object identifier (DOI) codes of articles related to a specific gene (Supporting Information 2). The annotation of genes that were recently characterized was updated according to available literature.

Furthermore, we introduced a functional categorization for all genes that allows for a quick overview of the result of a transcriptome analysis without the need to look at single genes. The current genome annotation of *C. glutamicum* wild type still contains a high number of genes with "unknown function" or "general function prediction only" (in total 1161, Supporting Information 2). As expected, several of these poorly characterized genes are located within the deleted gene clusters. Based on our experimental data we can conclude for 269 of these genes that they are not relevant under the tested conditions. Consequently, to realize a complete functional annotation of relevant genes in *C. glutamicum*, the number of genes that need to be studied is now reduced by 23% to 892.

Moreover, the table contains two columns containing regulation data, namely "regulated by" and "regulates following genes" while discriminating between predicted and experimental proven data. Here the CoryneRegNet database[36] served as starting point and the available information was revised and complemented according to recent publications. Our data could now be used to easily update CoryneRegNet.

## Conclusions

In this study, we successfully applied a targeted top-down approach to construct the novel chassis strain *C. glutamicum* C1*. Starting from wild type ATCC 13032, 412 genes (13.4% of the genome) were deleted without any negative effect on growth of the strain under the primary screening conditions, i.e. batch cultivation on defined D-glucose medium. In comparison with genome reduction projects of other organisms, 13.4% does not appear to be so much, but one has to consider the different strategies, aims and testing conditions used in the different projects that have been reviewed recently[14]. The so far most prominent genome reduction project is presumably the work with *Mycoplasma mycoides* reaching a genome reduction of more than 50%[17]. This represents an outstanding achievement and is highly interesting from a scientific point of view, but the organism is not relevant for biotechnological applications. For *B. subtilis*, an organism well used in industry, the ultimate aim is the construction of a "Minibacillus" where all remaining genes are essential and the function for each of them is known[37]. The current genome reduction is at 42.3% (www.minibacillus.org, July 17th, 2017). In contrast to our approach, cultivation of genome reduced *B. subtilis* strains is performed in complex media and negative effects on growth are accepted. In 2008 an *E. coli* strain with a 1.03 Mb genome reduction (22.2%) was described which growths even better than the parental strain in M9 minimal medium[38]. This strain was further manipulated resulting in strain DGF-298, an *E. coli* strain with a 2.98 Mb genome (35.8% genome reduction compared to MG1655) with close to wild-type-like growth in minimal medium[39]. A further genome reduction project focusing on the biotechnological application of the resulting strains is currently ongoing with *Pseudomonas putida*. A strain with a 4.3% reduced genome has recently been extensively characterized with respect to factors of industrial relevance such as growth and heterologous protein production[40]. Taken together, there are currently a few projects aiming at the construction of genome reduced chassis organisms applicable for biotechnological processes. Nevertheless, considering the different starting genome sizes (*E. coli*: 4.64 Mb, *C. glutamicum*: 3.28), our approach with a genome size of strain C1* of 2.84 Mb is still well comparable with other projects focusing on the construction of robust chassis hosts.

Figure 6 shows a plot of the whole ATCC 13032 genome from our initial chassis blueprint[18] with an additional color-coding of those genomic regions where a deletion was either successful, not successful or had a negative effect on growth, respectively. It becomes obvious that we already targeted most of the large regions without essential genes and the remaining regions are too small to allow for a significant further genome reduction within a reasonable time frame. Therefore, alternative strategies need to be employed for further genome reduction and chassis optimization. The arrows in Figure 6 indicate exemplarily the location of all glycolysis genes, demonstrating that these are widely scattered along the genome. One possibility would be to join all these genes at one genomic region, thereby allowing the deletion of larger clusters in other regions. Furthermore, one could target clusters with mostly deletable genes containing only a few essential ones by designing synthetic gene clusters which can be introduce into the genome in direct exchange with the respective larger region. This genome redesign leads to a more structured genome, which will likely prove beneficial for future bottom-up approaches.

Alongside this project, a large strain library and a revised genome annotation list were generated (Supporting Information 2). The library not only offers the opportunity to screen for irrelevant gene cluster under different growth conditions but also provides, together with the revised genome annotation list, a very valuable tool for further genome reduction and systematic studies of *C. glutamicum* to unravel the function of the still numerous uncharacterized genes and to unscramble additional layers of microbial architecture.

The slightly impaired growth of C1* with some alternative carbon sources such as acetate might be attributed primarily to the law: “You get what you screen for”. This would imply that during the genome reduction process genes were hit (by random mutation or targeted deletion) with already known or still unknown functions for the assimilation of carbon sources other than D-glucose. All target sites for deletion were carefully selected through *a priori* classification of gene essentiality. We therefore hypothesize that the construction of C1*, including extensive deletions of large gene clusters, was inherently accompanied by changes in genotype to phenotype relationships. This effect could also be a result from changes in the tridimensional genome organization as suggested by Martínez-García et al.[14]. However, with respect to our original aim, C1* is a chassis with a

significantly reduced genome (13.4%) that maintains the growth performance and long-term stability of the wild type under our defined conditions. With its reduced complexity, C1* has the potential for a broad range of applications in basic research and bioprocess development, similar to the prophage-free progenitor strain MB001, which has proven to be useful for a wide range of projects[41-43].

# Methods

## Strains, plasmids and deletion of gene clusters

All strains used in this study including precursor strains are listed in Supporting Information 2 (List of GRS), a selection of the most important strains can be found in Table 4. All oligonucleotides used in this study are listed in Table S1 and all plasmids in Table S2. The deletion of IS elements is described in detail in Supporting Information 1, Strain construction. Gene clusters were deleted from the *C. glutamicum* genome by double crossover as described previously[44, 45]. Deletions were carried out successively starting from different strains with single cluster deletions which showed no negative effect under standard conditions[18]. All strains are based on CR099, which is the prophage free strain MB001[19] with additional deletions of all copies of ISCg1 and ISCg2.

## Growth experiments

For cultivations in defined CGXII medium[46] all chemicals were purchased from SIGMA Aldrich. During medium preparation, some substances were added sterile after autoclaving (D-glucose, PCA, biotin, trace elements, AF204, vitamins) and pH 7.0 was adjusted with 4 M NaOH. In experiments with C-source variation 55.5 mM D-glucose was exchanged while keeping carbon moles constant by either of the following: 166.5 mM acetate, 66.6 mM D-arabitol, 55.5 mM D-fructose, 55.5 mM D-gluconate, 27.8 mM D-maltose, 111 mM pyruvate or 27.8 mM D-sucrose.

For cultivations on complex medium, brain heart infusion (BHI) broth was purchased from Carl Roth GmbH containing 7.5 g $L^{-1}$ pig brain infusion, 10 g $L^{-1}$ pig heart infusion, 10 g $L^{-1}$ peptone, 2 g $L^{-1}$ D-glucose, 5 g $L^{-1}$ NaCl and 2.5 g $L^{-1}$ disodium phosphate.

Microtiter plate cultivations were carried out in 48-well Flowerplates® (m2p-labs GmbH, Baesweiler) with DO and pH optodes in a BioLector® (m2p-labs GmbH, Baesweiler) at 1,000 rpm, 95% humidity and 30 °C. Inoculated Flowerplates were covered with a sterile gas permeable foil (m2p-labs GmbH, Baesweiler) and gain for backscatter measurements was set to 20. Cultures were started at an optical

density of $OD_{600} \approx 0.2$ by inoculating 990 µL medium with 10 µL of a working cell bank (WCB), which had already been grown on the same medium beforehand. For further details regarding the preparation of WCBs, strain handling and data analysis procedures, the reader is referred to our previous study[18].

Lab-scale bioreactor batch cultivations were carried with 1 L medium in four parallel vessels (Eppendorf AG, Jülich, Germany), each of which was equipped with a stirrer (two rushton turbines), sample port, off gas condenser and probes as denoted in the following. All cultivations were inoculated directly from washed cryo culture aliquots, which were derived from exponentially growing cultures on CGXII medium. During cryo preparation, the cells were harvested at $OD_{600} = 10$, washed with 0.9% (w $v^{-1}$) NaCl and stored at -80 °C in 0.9% (w $v^{-1}$) NaCl and 20% (v $v^{-1}$) glycerol. For later inoculation, one of these aliquots was washed with 0.9% (w $v^{-1}$) NaCl and added to the bioreactor. Dissolved oxygen concentration (DO) was kept at 30% under constant air flow (1 vvm) by adjustment of the stirrer speed (200 – 1200 rpm, p = 0.1, Ti = 300 s). The pH of each culture was regulated by addition of 4 M HCl and 4 M NaOH to pH 7.0 (p = 10, Ti = 2400 s, no auto-reset, deadband = 0.02 pH). Online measurements were taken for DO (Visiferm DO 225, Hamilton), pH (405-DPAS-SC-K80/225, Mettler Toledo) and exhaust gas composition (GA4, Eppendorf AG, Jülich, Germany). Offline monitoring of growth was performed by $OD_{600}$, cell dry weight (CDW) and Coulter Counter measurements. For $OD_{600}$ determination samples were diluted with 0.9% NaCl and measured in a photometer (UV PharmaSpec 1700, Shimadzu) at 1 cm path length. For CDW, 2 mL samples were centrifuged 10 min at 13,000 rpm (Biofuge pico, Heraeus instruments) in pre-weighted reaction tubes. The pellet was washed once with 1 mL 0.9% NaCl solution and finally dried for 48 hours at 80 °C (Kelvitron T, Heraeus Instruments) for 48 hours before the dry weight was determined. For morphology studies Coulter Counter measurements were performed using a Multisizer 3 (Beckmann Coulter, Inc.), equipped with a 30 µM capillary. Culture samples were diluted with CasyTon (Roche Diagnostics GmbH) and analyzed for cell number, median single cell volume and overall cell volume (biovolume)[47].

Automated repetitive batch experiments were carried out by utilizing the Mini Pilot Plant (MPP) technology as introduced recently[35]. In short, the MPP integrates a BioLector with liquid handling

robotics (Tecan, Freedom Evo 200) to allow for automated preparation, running, monitoring and controlling of microtiter plate experiments. For repetitive batch cultivations, biomass growth in each well of a FlowerPlate was monitored and the automated inoculation of the next batch was triggered when the backscatter (BS) signal reached a fixed threshold, i.e. BS > 400 (for CGXII condition) and BS > 68 (for BHI condition). A predefined volume of 250 mL of each running culture was then transferred to inoculate the wells of the next culture. Subsequently, these wells were filled with 750 mL fresh CGXII or BHI media, stored at 4 °C on the robotic platform. This procedure was repeated until all 48-wells of one MTP were used (i.e. allowing for a maximum number of ten (CGXII) and eleven (BHI) repetitive batches, respectively). The resulting backscatter measurements were fitted to a process model to enable model-based estimation of specific growth rates and lag times (see Supporting Information 1 for more details).

## Genome resequencing

Genome resequencing was performed as described previously[48]. Genomic DNA of *C. glutamicum* samples was isolated as described[49] and 4 µg was fragmented to an average size of 600 bp by sonication (Bioruptor® Pico, diagenode). 2 µg fragmented DNA was used for library preparation and indexing using the TruSeq DNA PCR-free sample preparation kit according to the manufacturer's instruction (Illumina). The resulting libraries were quantified using the KAPA library quant kit (Peqlab) and normalized for pooling. Sequencing of pooled libraries was performed on a MiSeq system (Illumina) using paired-end sequencing with a read-length of 2 x 150 bases. Data analysis and base calling were accomplished with the Illumina instrument software and stored as fastq output files. The obtained sequencing data of each sample were imported into CLC Genomics Workbench (Qiagen Aarhus A/S) for trimming, quality filtering, mapping to the reference genome and further downstream analysis. Genome sequences for mapping of the genome-reduced strains were created by introducing all deletions *in silico* into the genome sequence of MB001 (GenBank CP005959). The genome sequence of C1 was deposited in GenBank under accession number CP017995. All mutations detected by sequencing (Tables 2 and S3) were introduced into this sequence. Due to the one nucleotide deletion in the 6C RNA, all bases after position 313097 are shifted -1 in the deposited sequence compared to the theoretical sequence of C1.

### Transcriptome Analysis

Comparative transcriptome analysis was performed as described previously[50]. Briefly, *C. glutamicum* MB001 and chassis strain C1 cells were grown in 5 ml BHI (Brain Heart Infusion, Difco) with 2% (w $v^{-1}$) D-glucose for about 6 hours at 30°C. The cells were washed once with sterile saline (0.9% (w $v^{-1}$) NaCl) and used to inoculate the second precultures which were performed overnight in defined CGXII medium containing 2% (w $v^{-1}$) D-glucose as carbon source. The main cultures were inoculated to an $OD_{600}$ of 1 in defined CGXII medium with 2% (w $v^{-1}$) D-glucose. At an $OD_{600}$ of 5 the cells were harvested by centrifugation (4120 x *g*, 10 min and 4 °C). The cell pellet was subsequently frozen in liquid nitrogen and stored at -70°C. The preparation of total RNA was performed using the RNeasy Kit from Qiagen (Hilden, Germany). Synthesis of fluorescently-labeled cDNA was carried out using SuperScript III reverse transcriptase (Life Technologies, Darmstadt, Germany). Purified cDNA samples of MB001 and C1 strain were pooled and the prepared two-color samples were hybridized at 65°C while rotating for 17 hours using Agilent's Gene Expression Hybridization Kit, hybridization oven and hybridization chamber. After hybridization the arrays were washed using Agilent's Wash Buffer Kit according to the manufacturer's instructions. Fluorescence of hybridized DNA microarrays was determined at 532 nm (Cy3) and 635 nm (Cy5) at 5 µm resolution with a GenePix 4000B laser scanner and GenePix Pro 7.0 software (Molecular Devices, Sunnyvale, CA, USA). Fluorescence images were saved to raw data files in TIFF format (GenePix Pro 7.0). Quantitative TIFF image analysis was carried out using GenePix image analysis software and results were saved as GPR-file (GenePix Pro 7.0). For background correction of spot intensities, ratio calculation and ratio normalization, GPR-files were processed using the BioConductor R-packages limma and marray (http://www.bioconductor.org).

## Declarations

### Availability of data and material

All data generated or analyzed during this study are included in this published article and in Supporting Information 1 and 2 as well as in the GEO database by series entry GSE88717, www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=cjavskagdnopfmx&acc=GSE88717.

## Competing interest

The authors declare no financial or commercial conflict of interest.


## Funding

This work was funded by the German Federal Ministry of Education and Research (BMBF, grant. no. 0316017). Further funding was received from the Innovation Lab initiative of the German Helmholtz Association to support the "Microbial Bioprocess Lab - A Helmholtz Innovation Lab."


## Authors' contributions

MBa: coordination, planned and supervised experiments, strain construction, genetic characterization, data analysis, genome annotation list, wrote the paper; SU: developed and supervised physiological strain characterization, data analysis, wrote the paper; RK, AR, NT, MFM: physiological strain characterization; TP: genome re-sequencing; AK: genome annotation list DS, NB: strain construction; KM, SH, RK, MBo, JK, WW, GS, JF: coordination, helped to finalize the manuscript; CR: coordination, strain construction, genome database deposition VFW: project coordination, helped to finalize the manuscript; SN: coordination, planned and supervised experiments, data analysis, genome annotation list, wrote the paper.


## Acknowledgements

We thank Cornelia Gätgens and Christina Mack for excellent technical assistance.

# Tables

**Table 1:** Phenotypic data from lab-scale bioreactor cultivations of *C. glutamicum* wild type and pre-chassis PC1 and PC2 in defined CGXII medium with 10 g $L^{-1}$ D-glucose. Cell number and biovolume were measured using a Coulter Counter.

| Parameter | | Wild type | PC1 | PC2 |
|---|---|---|---|---|
| $\mu_{max}$ [$h^{-1}$] according to: | $OD_{600}$ | 0.43 ± 0.02 | 0.42 ± 0.01 | 0.43 ± 0.02 |
| | cell number | 0.46 ± 0.02 | 0.48 ± 0.01 | 0.54 ± 0.03 |
| | biovolume | 0.43 ± 0.01 | 0.44 ± 0.01 | 0.46 ± 0.01 |
| $Y_{X/S}$ [g $g^{-1}$] | | 0.60 ± 0.04 | 0.56 ± 0.01 | 0.49 ± 0.05 |

**Table 2:** SNPs in the strains PC2, PC1 and C1 identified by genome re-sequencing and occurring with a frequency of more than 90% that may have an influence on physiology. The frequency refers to the share of mapped reads at this position showing the respective SNP in percent. A full list including also silent mutations is provided as Table S3.

| Strain and position of SNP | Frequency | Within gene or intergenic region | Comment |
|---|---|---|---|
| **Mutations occurring in all three strains** | | | |
| 1397928_SNV_C_T (PC2) | 99.55 | cg1720, ABC-transporter (Ala475Thr or Ala468Thr, N-Terminus unclear) | ABC-transporter of unknown function, consequence of mutation not predictable. |
| 1591757_SNV_C_T (PC1) | 100 | | |
| 1439962_SNV_C_T (C1) | 100 | | |
| **Mutations occurring in PC1 and C1** | | | |
| 2026166_SNV_G_A (PC1) | 99.07 | intergenic region of cg2359-cg2360 | in between the transcriptional and translational start of isoleucyl-tRNA ligase (Cg2359), no regulators known |
| 1874371_SNV_G_A (C1) | 98.92 | | |
| **Mutations occurring only in PC2** | | | |
| 2019162^2019163_ Insertion_-_C | 96.52 | cg2543 (*glcD*) Asp188fs, putative glycolate oxidase, FAD-linked subunit oxidoreductase | insertion, frameshift, consequence of mutation not predictable |
| 2755744_SNV_A_G | 100 | cg3322 Val598Ala, uncharacterized putative secreted membrane-fusion protein | consequence of mutation not predictable |
| **Mutations occurring only in C1** | | | |
| 313097_Deletion_C_- | 91.26 | cg4100 (6C-RNA) | in the coding region of the 6C-RNA, effect unclear |
| 856702_SNV_T_G | 100 | intergenic region of cg1017-cg1035 | within the region of the deletion primers, about 100 bp upstream of the first transcriptional start of cg1035, presumably no effect |
| 2155026_SNV_T_G | 100 | intergenic region of cgt5048-cg2701 | downstream of both His-tRNA and cg2701 (putative membrane protein), presumably no effect |
| 2266772_SNV_T_C | 100 | intergenic region of cg2831-cg2833 | within the promoter region of *ramA*, leads to reduced transcription, see main text for further details |

**Table 3:** Transcriptome analysis of *C. glutamicum* C1 in comparison to MB001. The table presents all genes that are regulated more than 3-fold in at least three out of four biological replicates with a *p*-value of ≤ 0.05. The cells were cultivated in defined CGXII medium with 2% (w $v^{-1}$) D-glucose as carbon source and harvested in the exponential growth phase.

| Locus Tag | Gene | Annotated function | Ratio | *p*-value |
|---|---|---|---|---|
| cg3226 | | putative L-lactate permease, in operon with *lldD*, MFS-type, repressed or activated by RamA depending on the growth medium | **0.05** | <0.001 |
| cg3267 | | putative membrane protein, putative pseudogene, C-terminal fragment | **0.05** | <0.001 |
| cg0228 | *hkm* | putative sensor histidine kinase fragment of two-component system, putative pseudogene, located next to ISCg2f | **0.09** | <0.001 |
| cg3266 | *tnp5c* | transposase | **0.09** | 0.001 |
| cg0953 | *mctC* | monocarboxylic acid transporter, activated by RamA | **0.12** | <0.001 |
| cg0952 | | putative integral membrane protein, activated by RamA | **0.12** | <0.001 |
| cg2610 | | putative ABC-type dipeptide/oligopeptide/nickel transport system, secreted component | **0.13** | 0.001 |
| cg3195 | | putative flavin-containing monooxygenase FMO | **0.14** | <0.001 |
| cg2831 | *ramA* | transcriptional regulator, acetate metabolism, LuxR-family | **0.15** | <0.001 |
| cg2490 | | putative secreted guanine-specific ribonuclease | **0.16** | 0.005 |
| cg1808 | | hypothetical protein | **0.21** | 0.030 |
| cg2139 | *gluD* | glutamate uptake system, ABC-type, permease subunit 2 | **0.22** | 0.001 |
| cg3227 | *lldD* | menaquinone-dependent L-lactate dehydrogenase, operon with cg3226, repressed by RamA | **0.22** | 0.010 |
| cg2136 | *gluA* | glutamate uptake system, ABC-type, ATP-binding protein | **0.23** | 0.002 |
| cg2137 | *gluB* | glutamate uptake system, secreted glutamate-binding protein | **0.23** | <0.001 |
| cg2138 | *gluC* | glutamate uptake system, ABC-type, permease subunit 1 | **0.24** | 0.002 |
| cg3096 | *ald* | aldehyde dehydrogenase, activated by RamA | **0.27** | 0.003 |
| cg3215 | *glpQ1* | putative glycerophosphoryl diester phosphodiesterase 1 | **0.27** | 0.002 |
| cg3298 | *tnp19a* | transposase fragment, putative pseudogene | **0.27** | 0.004 |
| cg1373 | | putative glyoxalase/bleomycin resistance/dioxygenase superfamily protein | **0.28** | <0.001 |
| cg2491 | | hypothetical protein, conserved | **0.29** | 0.006 |
| cg2560 | *aceA* | isocitrate lyase, part of glyoxylate shunt, activated by RamA | **0.33** | <0.001 |
| cg2559 | *aceB* | malate synthase, part of glyoxylate shunt, activated by RamA | **0.42*** | 0.013 |
| cg1227 | *ykoE* | thiamin-regulated ECF transporter for hydroxymethylpyrimidine, substrate-specific component | **3.02** | 0.002 |
| cg2430 | | hypothetical protein | **3.05** | 0.013 |
| cg1287 | | hypothetical protein, conserved | **3.13** | 0.001 |
| cg1782 | *tnp13b* | transposase | **3.15** | 0.003 |
| cg1567 | | hypothetical protein | **3.21** | 0.004 |
| cg2381 | | hypothetical protein, conserved | **3.46** | 0.016 |
| cg2950 | *radA* | putative ATP-dependent protease, DNA repair | **3.66** | 0.003 |
| cg1231 | *chaA* | Na+(K+)/H+ antiporter | **3.83** | <0.001 |
| cg2602 | | hypothetical protein, conserved, might be influenced by deletion of ISCg1d (cg2600) | **4.03** | 0.001 |
| cg2747 | *mepA* | putative cell wall peptidase, M23/M37-family | **4.36** | 0.025 |

| | | | | |
|---|---|---|---|---|
| cg2601 | | putative pirin-related protein-fragment, putative pseudogene, might be influenced by deletion of ISCg1d (cg2600) | **8.32** | <0.001 |
| cg0611 | | putative secreted protein | **9.71** | 0.003 |
| cg1696 | | putative antibiotic efflux permease of the major facilitator superfamily | **12.15** | 0.009 |
| cg1623 | | putative manganese transporter, repressed by MntR (Cg0741) | **61.69** | <0.001 |
| cg0693 | *groEL* | 60 kDa chaperon, (protein CPN60 groEL) C-terminal fragment, gene is disrupted by ISCg1c in ATCC 13032 | **345.21** | 0.004 |

*This gene was included as it is also a *ramA* target and transcribed divergently to cg2560.

**Table 4:** Genotypes of selected strains used in this work. A full list is provided in Supporting Information 2.

| **Strain name** | **Genotype (all based on CR099, which is ATCC 13032ΔCGP123ΔISCg12)** |
|---|---|
| PC1 | Δcg2312-cg2322 Δcg2621-cg2643 Δcg2663-cg2686 Δcg2755-cg2760 Δcg3102-cg3111 |
| PC2 | Δcg0414-cg0440 Δcg0635-cg0646 Δcg0704-cg0748 Δcg0822-cg0845 Δcg1018-cg1033 Δcg1172-cg1213 Δcg1291-cg1305 Δcg1340-cg1352 |
| C1 | PC1 Δcg0635-cg0646 Δcg0704-cg0748 Δcg0822-cg0845 Δcg1018-cg1033 Δcg1172-cg1213 Δcg1291-cg1305 |
| C2 | PC2 Δcg2312-cg2322 Δcg2621-cg2643 Δcg2663-cg2686 Δcg2755-cg2760 Δcg3102-cg3111 Δcg3072-cg3091 |
| C1* | C1 with the corrected SNP in the *ramA* promoter sequence. |

# Figures

A

| | WT | GRS 33_37 | GRS 33_37_40 | GRS 33_37_38_40 | PC1 | GRS 48_52_53 | GRS 16_17_19 | GRS 17_18_19_22_23 | PC2 |
|---|---|---|---|---|---|---|---|---|---|
| ΔCGP123 | | X | X | X | X | X | X | X | X |
| ΔISCg12 | | X | X | X | X | X | X | X | X |
| Δcg0414-cg0440 | | | | | | | X | | X |
| Δcg0635-cg0646 | | | | | | | X | X | X |
| Δcg0704-cg0748 | | | | | | | | X | X |
| Δcg0822-cg0845 | | | | | | | X | X | X |
| Δcg1018-cg1033 | | | | | | | | X | X |
| Δcg1172-cg1213 | | | | | | | | X | X |
| Δcg1291-cg1305 | | | | | | | | | X |
| Δcg1340-cg1352 | | | | | | | | | X |
| Δcg2312-cg2322 | | X | X | X | X | | | | |
| Δcg2621-cg2643 | | X | X | X | X | | | | |
| Δcg2663-cg2686 | | | | X | X | | | | |
| Δcg2755-cg2760 | | | X | X | X | | | | |
| Δcg3102-cg3111 | | | | | X | X | | | |
| ΔrrnC-cg3298 | | | | | | | | | |
| Δcg3263-cg3301 | | | | | | X | | | |
| Δcg3324-cg3345 | | | | | | X | | | |
| **Genome reduction [%]** | 0.0 | 7.7 | 7.8 | 8.5 | 8.8 | 8.5 | 9.0 | 10.7 | 12.6 |

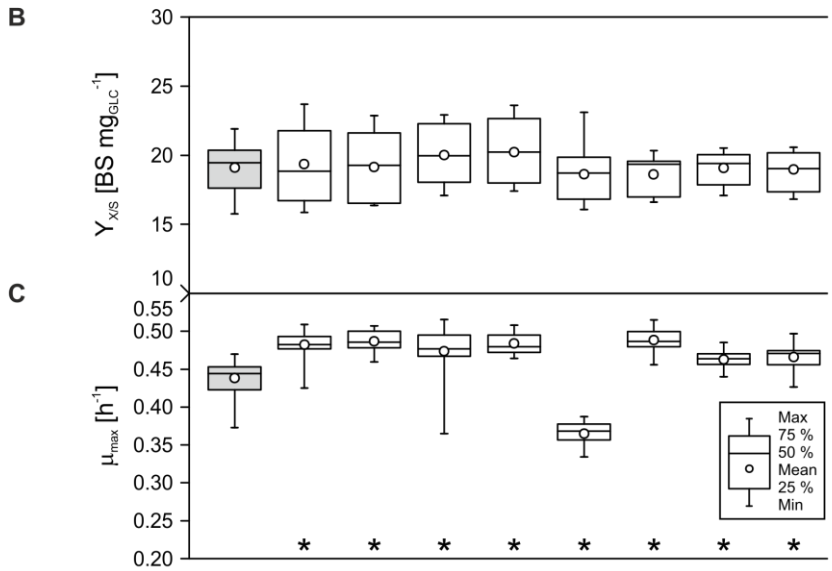
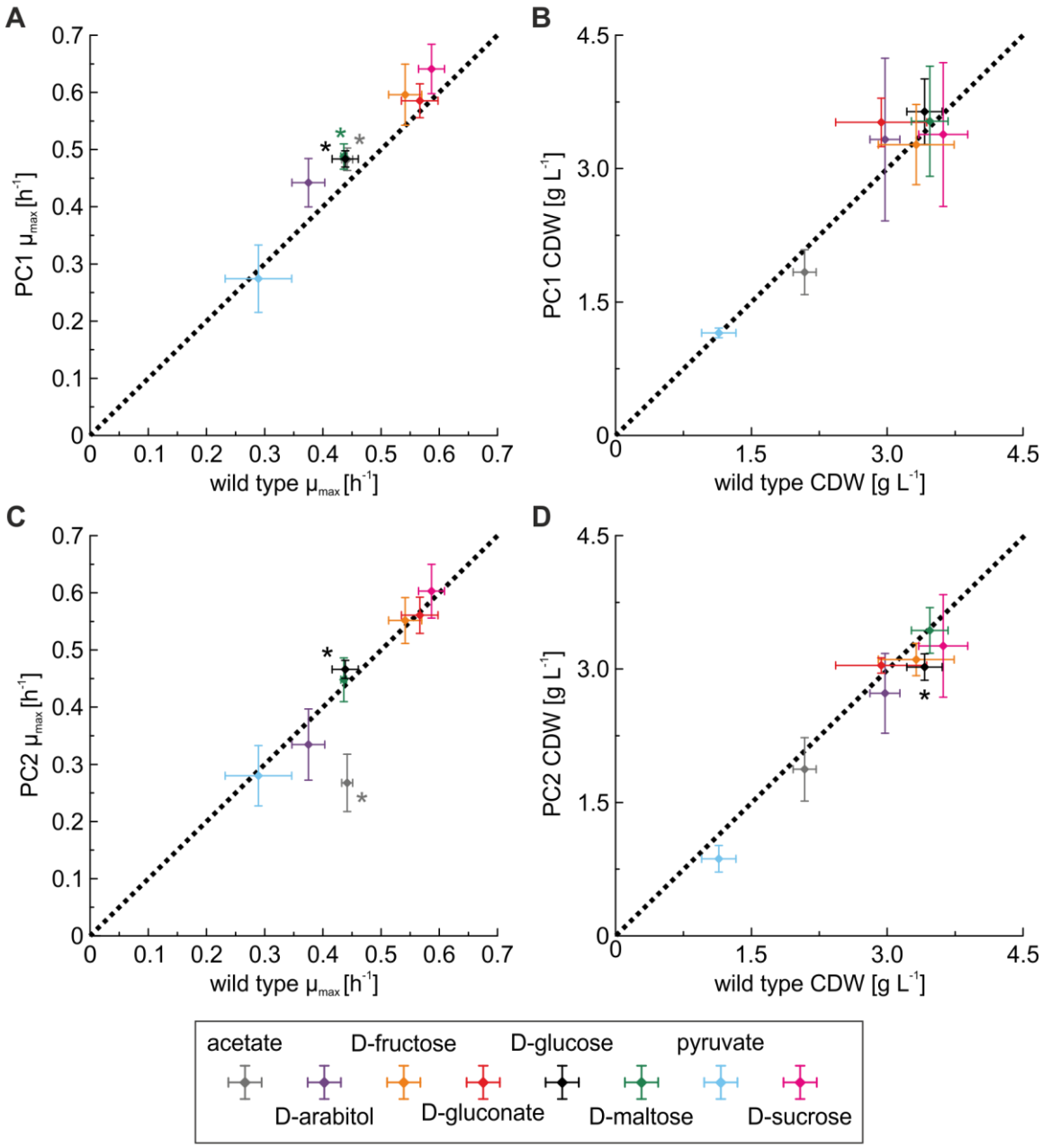



**Figure 1:** Combinatorial deletions and interdependence of selected gene clusters in *C. glutamicum*. (A) Strains with combinatorial deletions of gene clusters that had individually been classified as irrelevant for growth on CGXII medium with D-glucose as carbon source. The overall genome reduction of the combinatory strains is given as percentage with respect to the wild-type strain. (B+C) Biomass yields and maximum specific growth rates are derived from BioLector cultivations in CGXII medium ($n \geq 4$). Significantly altered values compared to the wild type are marked by an asterisk (two-sided t-test, $p$-value < 0.01).




555

556 **Figure 2:** Growth phenotypes of pre-chassis PC1 (A, B) and PC2 (C, D) in comparison to the wild
557 type. Cultivations were performed in CGXII medium with the indicated C-sources ($n \geq 3$). During
558 medium preparation, the listed C-sources were added to reach the same total concentration of carbon
559 compared to the standard D-glucose concentration of 10 g $L^{-1}$. Significantly changed $\mu_{max}$ or CDW
560 values compared to the wild type were marked by an asterisk in the color of the respective C-source
561 (two-sided t-test, $p$-value < 0.01).

A

| | WT | PC1_1 | PC1_2 | PC1_3 | PC1_4 | PC1_5 | C1 | PC2_1 | PC2_2 | PC2_3 | PC2_4 | PC2_5 | C2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ΔCGP123 | | x | x | x | x | x | x | x | x | x | x | x | x |
| ΔISCg12 | | x | x | x | x | x | x | x | x | x | x | x | x |
| Δcg0414-cg0440 | | | | | | | | x | x | x | x | x | x |
| Δcg0635-cg0646 | | x | x | x | x | x | x | x | x | x | x | x | x |
| Δcg0704-cg0748 | | | x | x | x | x | x | x | x | x | x | x | x |
| Δcg0822-cg0845 | | | | x | x | x | x | x | x | x | x | x | x |
| Δcg1018-cg1033 | | | | | x | x | x | x | x | x | x | x | x |
| Δcg1172-cg1213 | | | | | | x | x | x | x | x | x | x | x |
| Δcg1291-cg1305 | | | | | | | x | x | x | x | x | x | x |
| Δcg1340-cg1352 | | | | | | | | x | x | x | x | x | x |
| Δcg2312-cg2322 | | x | x | x | x | x | x | x | x | x | x | x | x |
| Δcg2621-cg2643 | | x | x | x | x | x | x | | x | x | x | x | x |
| Δcg2663-cg2686 | | x | x | x | x | x | x | | | x | x | x | x |
| Δcg2755-cg2760 | | x | x | x | x | x | x | | | | x | x | x |
| Δcg3072-cg3091 | | | | | | | | | | | | | x |
| Δcg3102-cg3111 | | x | x | x | x | x | x | | | | | x | x |
| **Genome reduction [%]** | 0.0 | 9.1 | 10.3 | 11.5 | 11.9 | 12.8 | 13.4 | 12.8 | 13.6 | 14.2 | 14.4 | 14.7 | 15.2 |

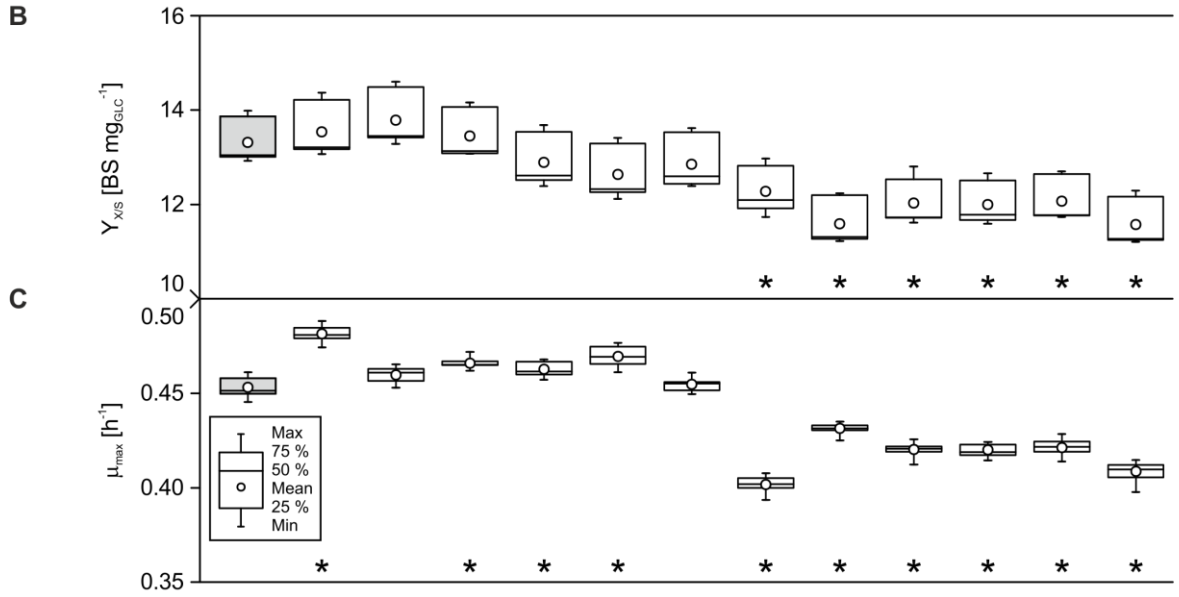



**Figure 3:** Phenotyping results of pre-chassis derived from PC1 and PC2 and chassis C1 and C2. The overall genome reduction of the combinatory strains is given as percentage with respect to the wild-type genome. (B+C) Biomass yields and maximum specific growth rates in CGXII medium in BioLector cultivations ($n$ = 9). Significantly changed values compared to the wild type are marked by an asterisk (two-sided t-test, $p$-value < 0.01).

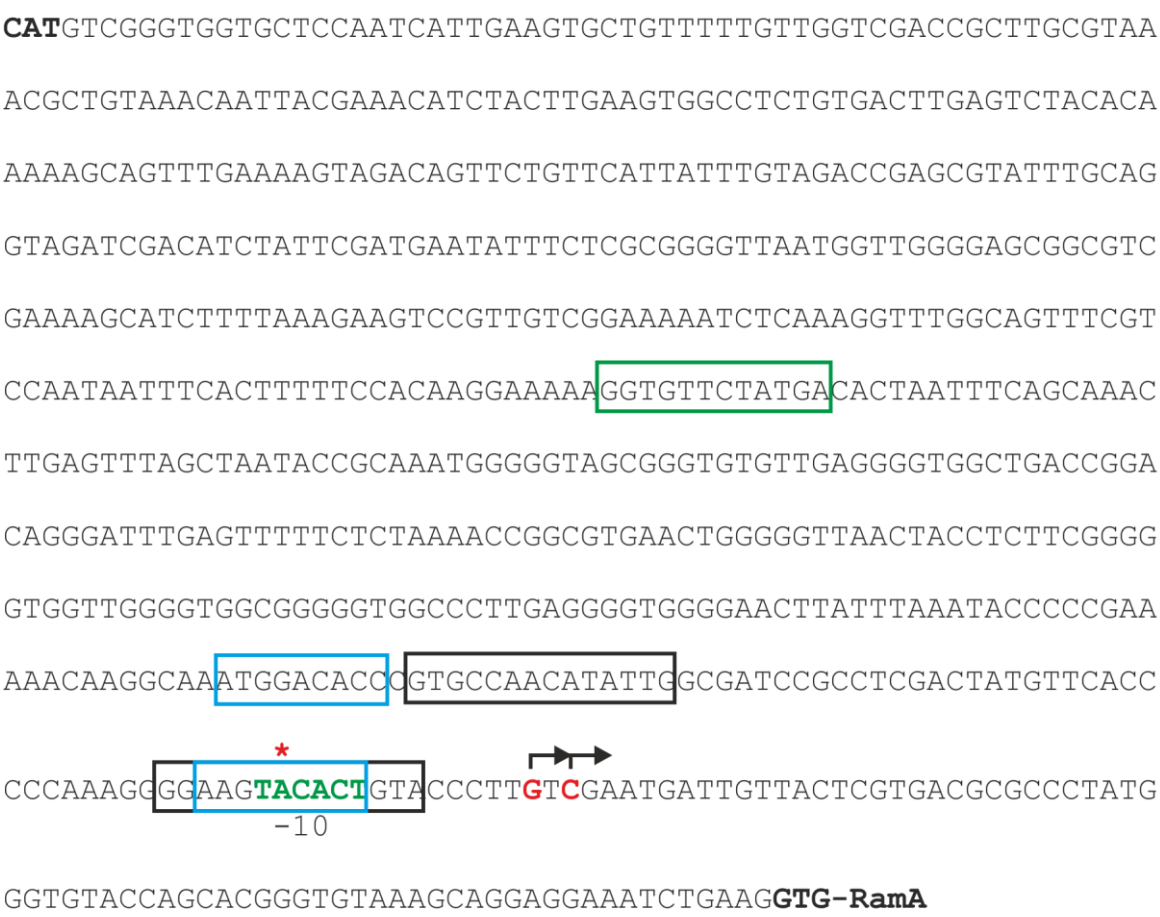



**Figure 4:** Promoter region of *C. glutamicum ramA*. The transcriptional start sites of *ramA* (red) were determined by Cramer *et al.*[51] (“C”, -71 with respect to the translational start) and Pfeifer-Sancar *et al.*[27] (“G”, -73 with respect to the translational start). The -10 region (green letters) was annotated by Pfeifer-Sancar *et al.*[27]. The start codons of *ramA* and *cysK* are printed in bold, black letters. SugR binding sites were differentially described by Engels *et al.*[28] (black boxes) and Toyoda *et al.*[29] (cyan boxes). The green box indicates the GlxR binding site identified by[29, 52]. *ramA* is autoregulated and several putative RamA binding sites have been described[29, 51], but where excluded for clarity. The red asterisk indicates A to G mutation in C1. Please note that due to the orientation of *ramA* the minus-strand of the genome is shown.

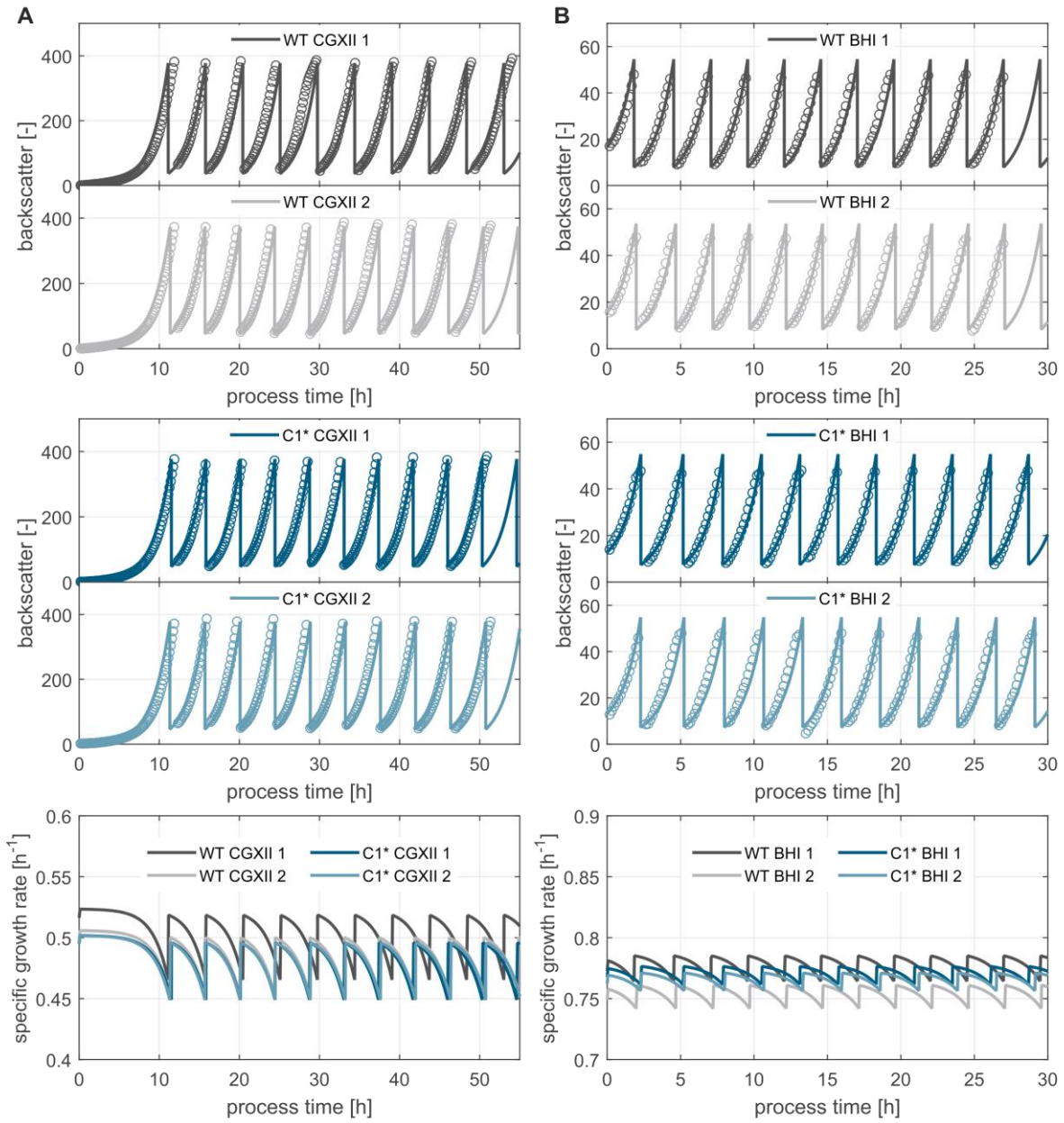



**Figure 5:** Long-term cultivation of *C. glutamicum* C1* and wild-type ATCC 13032 in defined CGXII medium with 40 g $L^{-1}$ D-glucose (A) and brain heart infusion (BHI) medium without additional D-glucose (B). Automated repetitive batch cultivations were performed as introduced recently[35]. Both strains were cultivated in two independent series under each condition as annotated. Backscatter data (symbols) was fitted to a process model (see Supporting Information 1 for more details) to simulate biomass growth during repetitive batch operation (solid lines).

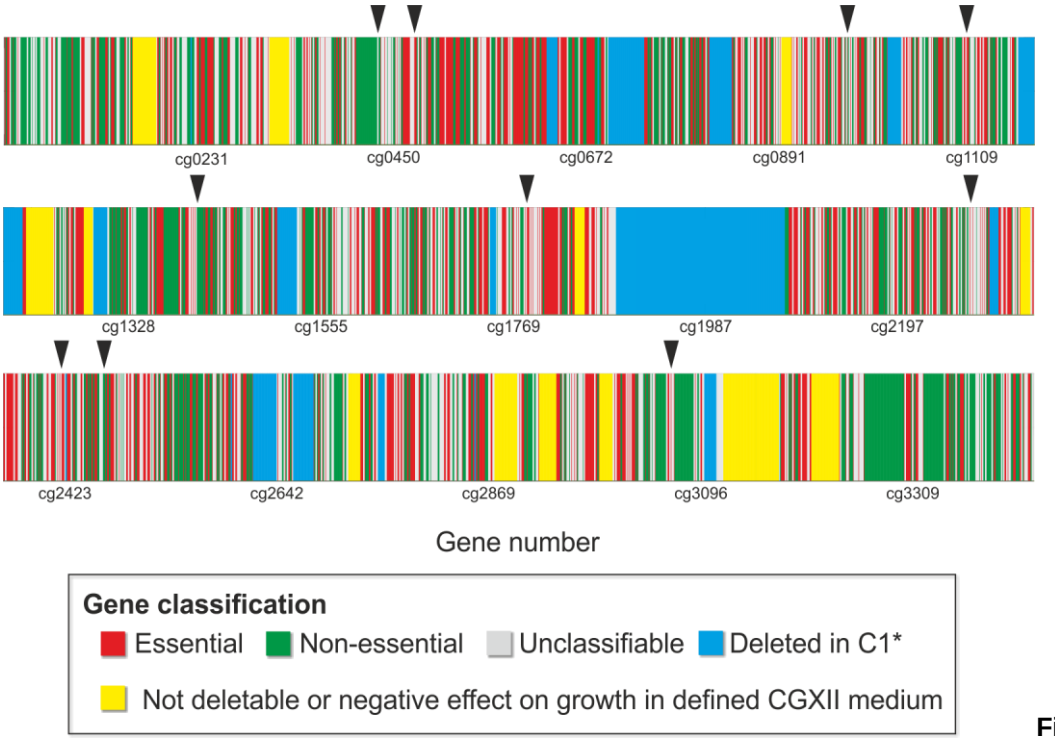



**Figure 6:** *C. glutamicum* ATCC 13032 genome map with classification results of essential, non-essential and unclassifiable genes according to Unthan et al. (2015)[18]. All clusters deleted in C1* are shown in yellow. Clusters that could not be deleted or deletions leading to impaired growth in defined CGXII medium are shown in yellow. Black arrows are pointing towards glycolysis genes *pgi* (cg0973), *pfkA* (cg1409), *fda* (cg3068), *tpi* (cg1789), *gap* (cg1791), *pgk* (cg1790), *gpmA* (cg0482), *eno* (cg1111), *pyk* (cg2291), *aceA* (cg2466), *lpd* (cg0441), and *sucB* (cg2421).

# References


1. Li, S., Jendresen, C. B., Grünberger, A., Ronda, C., Jensen, S. I., Noack, S., and Nielsen, A. T. (2016) Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metab. Eng. 38*, 274-284.
2. Cleto, S., Jensen, J. V. K., Wendisch, V. F., and Lu, T. K. (2016) *Corynebacterium glutamicum* Metabolic Engineering with CRISPR Interference (CRISPRi). *ACS Synth. Biol. 5*, 375-385.
3. Shi, S., Liang, Y., Zhang, M. M., Ang, E. L., and Zhao, H. (2016) A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. *Metab. Eng. 33*, 19-27.
4. Altenbuchner, J. (2016) Editing of the *Bacillus subtilis* Genome by the CRISPR-Cas9 System. *Appl. Environ. Microbiol. 82*, 5421-5427.
5. Unthan, S., Radek, A., Wiechert, W., Oldiges, M., and Noack, S. (2015) Bioprocess Automation on a Mini Pilot Plant enables Fast Quantitative Microbial Phenotyping. *Microb. Cell Fact. 14*.
6. Cruz Bournazou, M. N., Barz, T., Nickel, D., Lopez Cárdenas, D., Glauche, F., Knepper, A., and Neubauer, P. (2016) Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities for validation of macro-kinetic growth models using *E. coli* as an example. *Biotechnol. Bioeng. 114*, 610-619.
7. Kappelmann, J., Klein, B., Geilenkirchen, P., and Noack, S. (2016) Comprehensive and Accurate Tracking of Carbon Origin of LC-Tandem Mass Spectrometry Collisional Fragments for $^{13}$C-MFA. *Anal. Bioanal. Chem. 409*, 2309-2326.
8. Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrne, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016) The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol. 34*, 104-110.
9. Teleki, A., Sanchez-Kopper, A., and Takors, R. (2015) Alkaline conditions in hydrophilic interaction liquid chromatography for intracellular metabolite quantification using tandem mass spectrometry. *Anal. Biochem. 475*, 4-13.
10. Monk, J. M., Koza, A., Campodonico, M. A., Machado, D., Seoane, J. M., Palsson, B. O., Herrgard, M. J., and Feist, A. M. (2016) Multi-omics Quantification of Species Variation of *Escherichia coli* Links Molecular Features with Strain Phenotypes. *Cell Syst. 3*, 238-251.
11. Vasilakou, E., Machado, D., Theorell, A., Rocha, I., Nöh, K., Oldiges, M., and Wahl, S. A. (2016) Current state and challenges for dynamic metabolic modeling. *Curr. Opin. Microbiol. 33*, 97-104.
12. Burgard, A., Burk, M. J., Osterhout, R., Van Dien, S., and Yim, H. (2016) Development of a commercial scale process for production of 1,4-butanediol from sugar. *Curr. Opin. Biotechnol. 42*, 118-125.
13. Limberg, M. H., Schulte, J., Aryani, T., Mahr, R., Baumgart, M., Bott, M., Wiechert, W., and Oldiges, M. (2017) Metabolic profile of 1,5-diaminopentane producing *Corynebacterium glutamicum* under scale-down conditions: Blueprint for robustness to bioreactor inhomogeneities. *Biotechnol. Bioeng. 114*, 560-575.
14. Martínez-García, E., and de Lorenzo, V. (2016) The quest for the minimal bacterial genome. *Curr. Opin. Biotechnol. 42*, 216-224.
15. Choe, D., Cho, S., Kim, S. C., and Cho, B.-K. (2016) Minimal genome: Worthwhile or worthless efforts toward being smaller? *Biotechnol. J. 11*, 199-211.
16. Beites, T., and Mendes, M. V. (2015) Chassis optimization as a cornerstone for the application of synthetic biology based strategies in microbial secondary metabolism. *Front. Microbiol. 6*.
17. Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., and Ma, L. (2016) Design and synthesis of a minimal bacterial genome. *Science 351*, aad6253.
18. Unthan, S., Baumgart, M., Radek, A., Herbst, M., Siebert, D., Brühl, N., Bartsch, A., Bott, M., Wiechert, W., Marin, K., Hans, S., Krämer, R., Seibold, G., Frunzke, J., Kalinowski, J., Rückert, C., Wendisch, V. F., and Noack, S. (2015) Chassis organism from *Corynebacterium glutamicum*-a top-down approach to identify and delete irrelevant gene clusters. *Biotechnol. J. 10*, 290-301.

19. Baumgart, M., Unthan, S., Rückert, C., Sivalingam, J., Grünberger, A., Kalinowski, J., Bott, M., Noack, S., and Frunzke, J. (2013) Construction of a prophage-free variant of *Corynebacterium glutamicum* ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl. Environ. Microbiol. 79*, 6006-6015.
20. Fränzel, B., Trötschel, C., Rückert, C., Kalinowski, J., Poetsch, A., and Wolters, D. A. (2010) Adaptation of *Corynebacterium glutamicum* to salt-stress conditions. *Proteomics 10*, 445-457.
21. Ochrombel, I., Becker, M., Krämer, R., and Marin, K. (2011) Osmotic stress response in *C. glutamicum*: impact of channel- and transporter-mediated potassium accumulation. *Arch. Microbiol. 193*, 787-796.
22. Yukawa, H., Omumasaba, C. A., Nonaka, H., Kos, P., Okai, N., Suzuki, N., Suda, M., Tsuge, Y., Watanabe, J., Ikeda, Y., Vertes, A. A., and Inui, M. (2007) Comparative analysis of the *Corynebacterium glutamicum* group and complete genome sequence of strain R. *Microbiology 153*, 1042-1058.
23. Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Krämer, R., Linke, B., McHardy, A. C., Meyer, F., Möckel, B., Pfefferle, W., Pühler, A., Rey, D. A., Rückert, C., Rupp, O., Sahm, H., Wendisch, V. F., Wiegräbe, I., and Tauch, A. (2003) The complete Corynebacterium glutamicum ATCC 13032 genome sequence and its impact on the production of l-aspartate-derived amino acids and vitamins. *J. Biotechnol. 104*, 5-25.
24. Candela, T., and Fouet, A. (2005) Bacillus anthracis CapD, belonging to the γ-glutamyltranspeptidase family, is required for the covalent anchoring of capsule to peptidoglycan. *Mol. Microbiol. 57*, 717-726.
25. Islam, S. T., and Lam, J. S. (2014) Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. *Can. J. Microbiol. 60*, 697-716.
26. Pahlke, J., Dostalova, H., Holatko, J., Degner, U., Bott, M., Patek, M., and Polen, T. (2016) The small 6C RNA of Corynebacterium glutamicum is involved in the SOS response. *RNA Biol. 13*, 848-860.
27. Pfeifer-Sancar, K., Mentz, A., Rückert, C., and Kalinowski, J. (2013) Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. *BMC Genomics 14*, 888.
28. Engels, V., Lindner, S. N., and Wendisch, V. F. (2008) The global repressor SugR controls expression of genes of glycolysis and of the L-lactate dehydrogenase LdhA in Corynebacterium glutamicum. *J. Bacteriol. 190*, 8033-8044.
29. Toyoda, K., Teramoto, H., Gunji, W., Inui, M., and Yukawa, H. (2013) Involvement of regulatory interactions among global regulators GlxR, SugR, and RamA in expression of ramA in Corynebacterium glutamicum. *J. Bacteriol. 195*, 1718-1726.
30. Baumgart, M., and Frunzke, J. (2015) The manganese-responsive regulator MntR represses transcription of a predicted ZIP family metal ion transporter in *Corynebacterium glutamicum*. *FEMS Microbiol. Lett. 362*, 1-10.
31. Cramer, A., Gerstmeir, R., Schaffer, S., Bott, M., and Eikmanns, B. J. (2006) Identification of RamA, a novel LuxR-type transcriptional regulator of genes involved in acetate metabolism of *Corynebacterium glutamicum*. *J. Bacteriol. 188*, 2554-2567.
32. Schröder, J., and Tauch, A. (2010) 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. 34*, 685-737.
33. Auchter, M., Cramer, A., Hüser, A., Rückert, C., Emer, D., Schwarz, P., Arndt, A., Lange, C., Kalinowski, J., Wendisch, V. F., and Eikmanns, B. J. (2011) RamA and RamB are global transcriptional regulators in *Corynebacterium glutamicum* and control genes for enzymes of the central metabolism. *J. Biotechnol. 154*, 126-139.
34. Abreu, V. A., Almeida, S., Tiwari, S., Hassan, S. S., Mariano, D., Silva, A., Baumbach, J., Azevedo, V., and Rottger, R. (2015) CMRegNet-An interspecies reference database for corynebacterial and mycobacterial regulatory networks. *BMC Genomics 16*, 452.
35. Radek, A., Tenhaef, N., Müller, M. F., Brüsseler, C., Wiechert, W., Marienhagen, J., Polen, T., and Noack, S. (2017) Miniaturized and automated adaptive laboratory evolution: Evolving Corynebacterium glutamicum towards an improved d-xylose utilization. *Bioresour. Technol.*

36. Pauling, J., Röttger, R., Tauch, A., Azevedo, V., and Baumbach, J. (2012) CoryneRegNet 6.0--Updated database content, new analysis methods and novel features focusing on community demands. *Nucleic Acids Res. 40*, D610-614.
37. Reuss, D. R., Commichau, F. M., Gundlach, J., Zhu, B., and Stülke, J. (2016) The Blueprint of a Minimal Cell: MiniBacillus. *Microbiol. Mol. Biol. Rev. 80*, 955-987.
38. Mizoguchi, H., Sawano, Y., Kato, J., and Mori, H. (2008) Superpositioning of deletions promotes growth of Escherichia coli with a reduced genome. *DNA Res. 15*, 277-284.
39. Hirokawa, Y., Kawano, H., Tanaka-Masuda, K., Nakamura, N., Nakagawa, A., Ito, M., Mori, H., Oshima, T., and Ogasawara, N. (2013) Genetic manipulations restored the growth fitness of reduced-genome Escherichia coli. *J. Biosci. Bioeng. 116*, 52-58.
40. Lieder, S., Nikel, P. I., de Lorenzo, V., and Takors, R. (2015) Genome reduction boosts heterologous gene expression in Pseudomonas putida. *Microb. Cell Fact. 14*, 23.
41. Heider, S. A., Wolf, N., Hofemeier, A., Peters-Wendisch, P., and Wendisch, V. F. (2014) Optimization of the IPP Precursor Supply for the Production of Lycopene, Decaprenoxanthin and Astaxanthin by *Corynebacterium glutamicum*. *Front Bioeng Biotechnol 2*, 28.
42. Chen, Z., Huang, J., Wu, Y., and Liu, D. (2016) Metabolic engineering of *Corynebacterium glutamicum* for the de novo production of ethylene glycol from glucose. *Metab. Eng. 33*, 12-18.
43. Kortmann, M., Kuhl, V., Klaffl, S., and Bott, M. (2015) A chromosomally encoded T7 RNA polymerase-dependent gene expression system for *Corynebacterium glutamicum*: construction and comparative evaluation at the single-cell level. *Microb. Biotechnol. 8*, 253-265.
44. Niebisch, A., and Bott, M. (2001) Molecular analysis of the cytochrome $bc_1$-$aa_3$ branch of the *Corynebacterium glutamicum* respiratory chain containing an unusual diheme cytochrome $c_1$. *Arch. Microbiol. 175*, 282-294.
45. Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., and Pühler, A. (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* Plasmids pK18 and pK19: Selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene 145*, 69-73.
46. Keilhauer, C., Eggeling, L., and Sahm, H. (1993) Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon. *J. Bacteriol. 175*, 5595-5603.
47. Unthan, S., Grunberger, A., van Ooyen, J., Gatgens, J., Heinrich, J., Paczia, N., Wiechert, W., Kohlheyer, D., and Noack, S. (2014) Beyond growth rate 0.6: What drives Corynebacterium glutamicum to higher growth rates in defined medium. *Biotechnol. Bioeng. 111*, 359-371.
48. Mahr, R., Gätgens, C., Gätgens, J., Polen, T., Kalinowski, J., and Frunzke, J. (2015) Biosensor-driven adaptive laboratory evolution of L-valine production in *Corynebacterium glutamicum*. *Metab. Eng. 32*, 184-194.
49. Eikmanns, B. J., Thum-Schmitz, N., Eggeling, L., Lüdtke, K. U., and Sahm, H. (1994) Nucleotide sequence, expression and transcriptional analysis of the *Corynebacterium glutamicum gltA* gene encoding citrate synthase. *Microbiology 140*, 1817-1828.
50. Vogt, M., Haas, S., Klaffl, S., Polen, T., Eggeling, L., van Ooyen, J., and Bott, M. (2014) Pushing product formation to its limit: metabolic engineering of Corynebacterium glutamicum for L-leucine overproduction. *Metab. Eng. 22*, 40-52.
51. Cramer, A., and Eikmanns, B. J. (2007) RamA, the transcriptional regulator of acetate metabolism in Corynebacterium glutamicum, is subject to negative autoregulation. *J. Mol. Microbiol. Biotechnol. 12*, 51-59.
52. Toyoda, K., Teramoto, H., Inui, M., and Yukawa, H. (2011) Genome-wide identification of in vivo binding sites of GlxR, a cyclic AMP receptor protein-type regulator in Corynebacterium glutamicum. *J. Bacteriol. 193*, 4123-4133.