**Bacillus subtilis** and **Escherichia coli** essential genes and minimal cell factories after one decade of genome engineering

Mario Juhas,1 Daniel R. Reuß,2 Bingyao Zhu2 and Fabian M. Commichau2

1Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
2Department of General Microbiology, Georg-August-University Göttingen, Grisebachstr. 8, 37077 Göttingen, Germany

Investigation of essential genes, besides contributing to understanding the fundamental principles of life, has numerous practical applications. Essential genes can be exploited as building blocks of a tightly controlled cell ‘chassis’. **Bacillus subtilis** and **Escherichia coli** K-12 are both well-characterized model bacteria used as hosts for a plethora of biotechnological applications. Determination of the essential genes that constitute the **B. subtilis** and **E. coli** minimal genomes is therefore of the highest importance. Recent advances have led to the modification of the original **B. subtilis** and **E. coli** essential gene sets identified 10 years ago. Furthermore, significant progress has been made in the area of genome minimization of both model bacteria. This review provides an update, with particular emphasis on the current essential gene sets and their comparison with the original gene sets identified 10 years ago. Special attention is focused on the genome reduction analyses in **B. subtilis** and **E. coli** and the construction of minimal cell factories for industrial applications.

**Introduction**

Essential genes are absolutely vital for an organism’s survival. In general, genes can be considered essential if they cannot be individually knocked out under conditions where most of the required nutrients are provided in the growth medium and at a temperature which allows the best growth of the organism. In the case of two bacteria described in this review, namely **Bacillus subtilis** and **Escherichia coli**, these conditions were Luria–Bertani (LB) medium supplemented with glucose and cultivation at 37 °C (Baba et al., 2006; Kobayashi et al., 2003). The identification of essential genes will help to advance our understanding of the fundamental principles of life (Commichau et al., 2013; Jewett & Forster, 2010; Juhas et al., 2011; Moya et al., 2009). Moreover, bacterium-specific essential genes are considered to be suitable targets for the development of novel antimicrobials (Bumann, 2008; Juhas et al., 2012a, b). Finally, essential genes are important for the emerging field of synthetic biology as they form the building blocks of future custom-made minimal genomes of minimal cell factories (Juhas et al., 2013; Nandagopal & Elowitz, 2011; Seo et al., 2013). A wide variety of experimental approaches, such as targeted gene deletions, generation of conditional knockout mutants, genome-wide RNA interference screens and saturation transposon mutagenesis, were used to identify and study essential genes (Baba et al., 2006; Boutros & Ahringer, 2008; Christen et al., 2011; de Berardinis et al., 2008; French et al., 2008; Kato & Hashimoto, 2007; Kobayashi et al., 2003; Langridge et al., 2009; Yu et al., 2008). Furthermore, a number of computational approaches were employed to define the core essential gene set needed to sustain life (Juhas et al., 2011, 2012b; McCutcheon & Moran, 2010; Moya et al., 2009). However, the complete set of essential genes that is universally required by all organisms is difficult to determine experimentally and hard to predict. For instance, survival of the cell in the absence of a particular enzyme does not always mean that the enzymic function is not needed, as there might be other enzymes with the same function present in the cell. Multiple rRNA operons were shown to be required for efficient growth of **B. subtilis** (Yano et al., 2013). Moreover, some proteins can be considered essential because they protect the cell against harmful metabolites and proteins, such as phage-encoded toxins. Hence, these genes can lose their essentiality if the corresponding harmful gene/pathway is deleted (Durand et al., 2012).

The rod-shaped Gram-positive and Gram-negative bacteria **B. subtilis** and **E. coli**, respectively, are both well-characterized and easily amendable to genetic engineering. Furthermore, the comprehensive community-curated databases (SubtiWiki, EcoliWiki, BsubCyc and EcoCyc, with SubtiWiki and EcoCyc being the most frequently used) provide up-to-date information about these two model organisms (Caspers et al., 2014; Keseler et al., 2013; McIntosh et al., 2012; Michna et al., 2012).
et al., 2014). B. subtilis and E. coli are both endowed with complex regulatory and metabolic networks allowing them to thrive in a broad spectrum of environments (Buescher et al., 2012; Kohlstedt et al., 2014; Nicolas et al., 2012; Wang et al., 2010). B. subtilis is a commensal bacterium able to form metabolically inactive dehydrated endospores allowing survival in nutrient-free environments (McKenney et al., 2013). In contrast to B. subtilis, E. coli does not form endospores and, depending on the genome configuration, its lifestyle might vary from commensalism to pathogenicity (Clements et al., 2012; Leimbach et al., 2013). Some E. coli strains are important enteric and extra-intestinal pathogens (Leimbach et al., 2013). The E. coli strain O104:H4 has been associated with one of the largest recent infection outbreaks that led to more than 50 deaths and 4000 serious infections (Bielaszewska et al., 2011; Mellmann et al., 2011; Rasko et al., 2011). A number of genes encoding pathogenicity traits have been acquired by horizontal gene transfer (Johas, 2013). This may cause a significant ‘enlargement’ of the pathogenic E. coli genomes (Rasko et al., 2008; Touchon et al., 2009). Derivatives of the B. subtilis laboratory strain 168 and the commensal E. coli strain K-12 are used as production platforms in biotechnology (Chen et al., 2013; Schallmey et al., 2004). B. subtilis is an attractive host for industrial applications due to its ability to secrete proteins into the medium (Manabe et al., 2011, 2013; Simonen & Palva, 1993; Zweers et al., 2008). B. subtilis has been successfully engineered for the production of riboflavin (Hao et al., 2013; Shi et al., 2009). E. coli has been engineered for the production of organic acids and solvents, isoprenoids, amino acids and biofuels (Ajikumar et al., 2010; Chen et al., 2013; Clomburg & Gonzalez, 2010, 2011; Park et al., 2012; Yin et al., 2011; Zhou et al., 2012a, b). In many cases the development of a new bioprocess or improvement of existing biosynthesis strategies is hampered by lack of knowledge about each component of the host cell chassis (Lam et al., 2012). The introduction of a novel metabolic pathway into an existing metabolic network can lead to unpredictable interactions (Commichau et al., 2014; Kim & Copley, 2012). As these interactions can impair the fitness of the engineered strain, an ideal chassis would be a well-characterized cell harbouring only the minimal set of functions required to synthesize a product of interest. Systematic genome reductions in B. subtilis and E. coli suggest that the streamlined strains are more suitable for the production of industrially relevant compounds (Fehér et al., 2007).

In this review we give an overview of the current knowledge about the B. subtilis and E. coli essential genes. Special focus is paid to the current essential gene sets and their comparison with the original gene sets identified 10 years ago. Furthermore, we discuss recent advances in large-scale genome engineering and genome reduction approaches in both bacteria that aim to generate minimal cell factories for industrial applications.

**The current B. subtilis essential gene set**

In 2003, a genome-scale study described 271 ORFs as essential in B. subtilis (Kobayashi et al., 2003). This number was decreased to 261 as the result of experimental re-evaluation over the subsequent 10 years (Akanuma et al., 2012; Hunt et al., 2006; Meine et al., 2013; Tanaka et al., 2013). Of the re-evaluated set, 259 genes encoded essential proteins, among them six of unknown function, and two genes encoded essential RNAs (Commichau et al., 2013). In the meantime it has been shown that four genes from the six ORFs encoding proteins of unknown function and two other genes are dispensable for B. subtilis growth on complex medium (Figaro et al., 2013; Michna et al., 2014). Recently, we found that, of the remaining two genes ylaN and ycgG, the latter is not essential for growth of B. subtilis. By contrast, no transformants with inactivated ylaN were obtained (unpublished results). Moreover, our re-evaluation of the essentiality of topA encoding DNA topoisomerase I revealed that cells with inactivated topA are not viable in LB medium (unpublished results). Thus, as of July 2014, 253 genes can be regarded as being essential in B. subtilis. The largest group from the B. subtilis essential gene set plays a role in protein synthesis, quality control and secretion (Fig. 1a). Moreover, many essential genes are involved in basic metabolism, cell wall metabolism, cell division and DNA metabolism. A few genes encode essential proteins with protective functions, such as RNAS III, which is required to silence phase toxins and enzymes involved in RNA turnover. As shown above, the set of genes considered to be essential in B. subtilis has changed substantially over the past years. Therefore, it is plausible to assume that with the increasing number of analyses, the current set will be further modified in the future. The dynamics of the B. subtilis essential gene set has been reviewed recently (Commichau et al., 2013). For up-to-date information about B. subtilis essential genes we refer to the ‘Essential genes’ page on SubtiWiki (http://subtiwiki.uni-goettingen.de/wiki/index.php/Essential_genes).

**The current E. coli essential gene set**

The first E. coli minimal genome study employing genetic footprinting led to the identification of 620 putative essential genes in E. coli strain K-12 MG1655 (Gerdes et al., 2003). The following systematic inactivation of non-essential ORFs by precise, single-gene deletions led to the identification of 303 candidate essential genes in E. coli strain K-12 BW25113 (Baba et al., 2006). Non-essential genes in the latter study were inactivated by replacement with the kanamycin cassette flanked by FLP recognition target sites, thus generating in-frame deletions upon excision of the resistance cassette (Baba et al., 2006). The 303 candidate essential genes identified correspond to 315 clusters of orthologous groups. The most abundant were genes involved in information storage and processing, particularly translation, ribosome structure and biogenesis (and, to a lesser extent, transcription, DNA replication, recombination and repair). Many genes were involved in cell division and chromosome partitioning, cell envelope biogenesis, post-translational modification, and lipid and coenzyme metabolism. The function of 37 essential genes was unknown at the time of their identification (Baba et al.,
**B. subtilis** gene set is very similar to that of **E. coli** (Kato & Hashimoto, 2007). A later study by Kato & Hashimoto (2007) employing large-scale deletions to identify essential cis-acting regions of the **E. coli** chromosome showed that 35 of the 303 essential genes identified in 2006 were non-essential. The differences between these two essential gene sets were probably due to strain variations (BW25141 versus MG1655), growth conditions (glucose-containing versus glucose-free medium) and methodology (single-gene knockout versus large-scale deletions) (Kato & Hashimoto, 2007). This study also revealed that the origin of replication (oriC) is the only essential cis-acting DNA sequence in the **E. coli** chromosome, while the terminus of replication only affects growth, but is not essential (Kato & Hashimoto, 2007).

Recent research advancements led to significant modifications of the original **E. coli** essential gene set identified 10 years ago. The current list of genes considered to be essential in **E. coli**, consisting of 295 ORFs, can be found in the **EcoliWiki** (http://ecoliwiki.net/colipedia/index.php/Essential_genes). The largest proportion of the current **E. coli** essential gene set (109 ORFs) is involved in protein synthesis and quality control (Fig. 1b). Other large groups of essential genes play roles in cell wall biosynthesis and cell division (68 genes) and metabolism (59 genes). Furthermore, a number of essential genes are involved in DNA replication and chromosome maintenance and in protective functions. A small set (10 genes) is required for RNA synthesis and degradation, while the remaining 13 ORFs are hypotheticals with unknown function (Fig. 1b). The functional distribution of the current **E. coli** essential gene set is very similar to that of **B. subtilis** (Fig. 1a).

Inactivation of *rluD* encoding pseudouridine synthase was initially shown to have a negative effect on **E. coli** growth and ribosome function; however, a recent study demonstrated that *rluD* mutants are unstable and quickly give rise to suppressors (O’Connor & Gregory, 2011). The *rluD* suppressor mutations were found in *pfrB* and *pfc* encoding release factor 2 (RF2) and release factor 3 (RF3), respectively (O’Connor & Gregory, 2011). The investigation of unknown essential genes, particularly assigning function to the short ORFs encoding essential proteins of 50 or fewer amino acids, is challenging. Recent analysis suggests that a number of the short essential genes are involved in the cell’s response to a variety of stress conditions, such as heat shock, cold shock, cell envelope stress, oxidative stress, thiol stress and acid stress (Hemm et al., 2010; Hobbs et al., 2010). The function of 13 ORFs of the current **E. coli** essential gene set is entirely unknown (Fig. 1b). Most of these uncharacterized ORFs are either small RNAs (six genes) or small membrane protein-encoding genes (four genes) (Hemm et al., 2008). The remaining three ORFs encode hypothetical proteins.

**The orthologous essential genes of **B. subtilis** and **E. coli**

To identify the set of genes essential in both **B. subtilis** and **E. coli** we compared the corresponding protein sequences of analysed genes from the BsubCyc and the EcoCyc databases (Caspi et al., 2014; Keseler et al., 2013) with the help of the BLASTP tool (Hung & Hua, 2014). The direct orthologues of the **B. subtilis** and **E. coli** essential genes (135 ORFs) are shown in Tables S1 and S2 (available in the
online Supplementary Material), and their assignment to functional categories is shown in Table S3 and Fig. 1(c). Of the *B. subtilis* essential genes, 58% have direct orthologues among *E. coli* essential genes, and 46% of genes indispensable for *E. coli* are essential in *B. subtilis* (Tables S1–S3, Fig. 1c). The largest proportion of the orthologous essential gene set (approx. 50%) is involved in protein synthesis and quality control. Another large group of direct essential orthologues is involved in cell division and envelope biogenesis (19 genes) (Fig. 1c). This is not surprising as genes involved in the synthesis of proteins and cell membrane biogenesis are considered to be part of the ‘core’ essential gene set that is absolutely necessary for survival of any living cell (Juhas et al., 2011). For the same reason, *B. subtilis* and *E. coli* also share a number of genes involved in other aspects of ‘information storage and processing’, such as DNA replication and RNA synthesis. Twelve of the 17 *E. coli* genes involved in DNA replication, modification and chromosome maintenance are among *B. subtilis* essential genes and almost all (five of six) *B. subtilis* genes involved in RNA synthesis and degradation are indispensable for *E. coli* (Fig. 1c). A large proportion of shared metabolic genes could be attributed to similarities in core metabolism and to comparable environmental conditions (LB supplemented with glucose and growth at 37 °C), which were used to identify essential genes of both bacteria (Baba et al., 2006; Kobayashi et al., 2003). Variations in the sets of essential genes analysed could be caused by the fact that different cell types might have developed distinct solutions to the same biological problem (Juhas et al., 2011). For example, the variations in gene sets essential for cell membrane biosynthesis are probably due to different architecture of the cell envelope in the Gram-positive *B. subtilis* and Gram-negative *E. coli*. Genetic background might be also the reason why *B. subtilis* and *E. coli* do not share any of the unknown and protective function-encoding essential genes. It remains to be elucidated why there are so many essential genes with protective functions, particularly in *E. coli*. Recent analysis suggests that a number of the short essential genes of *E. coli* (*ythA*, *yobI*, *yqeL*, *yohP*, *ykgR*, *yoaJ*, *ypdK*, *yoaK*, *yqcG*) are involved in the cell’s response to stress conditions, such as heat shock, cold shock, cell envelope stress, oxidative stress, thiol stress and acid stress (Hemm et al., 2010; Hobbs et al., 2010). Cell envelope stress and acid stress are most relevant for the lifestyle of the pathogenic *E. coli* in the human stomach and intestine (Hobbs et al., 2010). However, it is as yet unclear why these genes are essential for *E. coli* K-12 under laboratory conditions. The above comparison shows that while *B. subtilis* and *E. coli* share approximately 50% of their essential genes, there are also a number of essential genes unique for each bacterium. This shows that the outcome of an essential genes identification study is strongly dependent on the bacterium cell type analysed. It is also important to specify the strain, environmental conditions and methodology used as this might influence the outcome of the study. For *B. subtilis* and *E. coli* essential gene analyses, strains 168 and K-12, respectively, are routinely grown at 37 °C in LB medium supplemented with glucose (Baba et al., 2006; Kobayashi et al., 2003).

**B. subtilis** and **E. coli** genome minimization

Attempts to minimize the genomes of model bacteria such as *B. subtilis* and *E. coli* address two major issues: (i) the identification of the minimal gene set and the encoded essential functions that are needed to sustain life, and (ii) the rational design and construction of the tightly controlled cell factories for industrially relevant products (Juhas et al., 2012a; Nandagopal & Elowitz, 2011). Engineering bacteria for the efficient production of a desired compound often fails due to inhibitory interactions caused by the introduction of the heterologous pathway into an existing metabolic network (Commichau et al., 2014; Kim & Copley, 2012; Pitera et al., 2007). Moreover, low production levels can result from the insufficient supply of the heterologous enzymes with precursors (Sandmann, 2002). Thus, decreasing the number of interactions between the heterologous pathway and the host’s metabolic network via the host’s genome reduction could lead to more efficient bioproduction. Two approaches, bottom-up and top-down, are employed to generate tailor-made minimal cell factories. It is as yet unclear which of these approaches will lead to construction of the first ‘real’ minimal cell consisting solely of essential genes; however, based on current knowledge, the top-down approach seems to be more straightforward.

**Bottom-up genome minimization**

The bottom-up approach is used to create a minimal cell from its basic building blocks: the protein-coding essential genes (Jewett & Forster, 2010; Juhas et al., 2012a). Inefficient DNA synthesis and assembly methods were among the hurdles of the bottom-up approach; however, this has changed over recent years (Esvelt & Wang, 2013; Fehér et al., 2012). The Gibson assembly (Gibson et al., 2009), CPEC (circular polymerase extension cloning; Quan & Tian, 2009; Quan & Tian, 2011), synthesis from oligonucleotide chips (Kosuri et al., 2010; Matzas et al., 2010), SLIC (sequence and ligation independent cloning; Li & Elledge, 2007), SLICE (seamless ligation cloning extract; Zhang et al., 2012), the ligase cycling reaction (de Kok et al., 2014) and other DNA synthesis and assembly methods have been developed. Another hurdle was overcome when the intact *Mycoplasma mycoides* genome transplanted into *Mycoplasma capricolum* generated cells phenotypically identical to the donor bacterium (Lartigue et al., 2007, 2009). This is important because genome transplantation is required to ‘boot-up’ the synthetic genome in the genome-free recipient cell’s chassis. The synthetic 1.08 Mbp *M. mycoides* genome dubbed ‘JC1-V-syn1.0’ assembled in yeast was successfully transplanted into *M. capricolum*, thus generating cells with the *M. mycoides* phenotype (Gibson et al., 2009, 2010; Itaya, 2010). Recent synthesis of a whole functional yeast chromosome (Annaluru et al., 2014) suggests that synthetic biology has reached the stage where
tailor-made prokaryotic and eukaryotic organisms can be constructed from standardized biological parts. However, this is still far from the case. The minimal set of genes required to sustain life is not yet known. Even the essential gene sets of the two well-studied model organisms B. subtilis and E. coli, modified significantly over the last 10 years, are still undergoing alterations. Thus, based on current knowledge, it is hard to imagine that essential genes from different sources could be put together and transplanted into a cell chassis to obtain a functional minimal cell factory. A conservative approach, including more genes than those constituting an absolutely minimal gene set, could make this goal more realistic. The genome of such a cell could be subsequently further minimized by the top-down approach. Alternatively, a straightaway top-down approach followed by the genome engineering-based augmentation of the reduced cell with desired properties could lead to the construction of the minimal cell factory.

**Top-down genome minimization**

The top-down approach is used to streamline existing cells by deleting large non-essential parts of their genomes. As the B. subtilis and E. coli genome sequences are available (Blattner et al., 1997; Kunst et al., 1997) they were often targets of top-down minimization. Significant progress has been made in the genome reduction of both model bacteria in recent years (Fehér et al., 2007) (Fig. 2). The first streamlined B. subtilis strain lacking 7.7% of the genome was generated by deleting the prophage PBSX, the three prophage-like elements skin, proφ1 and proφ3 and the pks gene cluster, in a strain that was already cured from the cryptic prophage SPβ (Dorenbos et al., 2002; Westers et al., 2003). Interestingly, the loss of 332 genes (about 325 kbp) did not cause a growth defect, and the flux through central carbon metabolism was identical to that of the parental strain. Moreover, although the reduced-genome strain was more motile than the parent strain, their competence and sporulation phenotypes were comparable. This showed that regions of the B. subtilis chromosome can be deleted without affecting viability under laboratory growth conditions. It is unclear why these DNA regions reside in the bacterial genome; their selective advantage has yet to be identified. In another work, based on the genome-scale identification of essential genes (Kobayashi et al., 2003), B. subtilis lacking 17 dispensable genomic regions, in addition to the prophage-like elements skin, proφ1–6, the cryptic prophage SPβ, the prophage PBSX and the pks gene cluster, was constructed (Ara et al., 2007). The resulting strain with a 3.2 Mbp genome was viable but showed unstable growth rate, cell morphology and protein production after extensive culturing (Ara et al., 2007). In 2008, the construction of the reduced-genome B. subtilis strain MGB874 lacking 874 kbp (20.7 %) of genomic DNA was reported (Morimoto et al., 2008). In this study, the impact of each large genomic deletion on growth was tested prior to combining all the deletions in a single strain. In total, 11 dispensable regions with lengths between 11 and 195 kbp were sequentially deleted from the parent strain MGB469. Compared with the parent strain, the constructed B. subtilis MGB874 had a similar growth rate and yield in LB medium, while showing only a slightly reduced growth in minimal medium (Morimoto et al., 2008). Transcriptome analyses revealed that the transcriptional regulatory network in the reduced-genome strain was disturbed during exponential growth and that it was reorganized after entry into the stationary growth phase (Morimoto et al., 2008). Although the molecular basis for this phenomenon remains to be elucidated, strain MGB874 serves as a cell factory for the production of proteins and other valuable compounds and as the promising basis for further genome reduction (see below). Recently, a B. subtilis strain lacking 35% of the genome was constructed by sequentially combining deletions from the repertoire of the chromosome regions dispensable for growth (Tanaka et al., 2013).

Like B. subtilis, E. coli has been also subjected to systematic genome reduction. In 2002, the first genome-reduced E. coli strain, MDS12 lacking 8.1 % of the genome, was constructed by deleting 12 strain-specific genomic islands (K-islands) from E. coli K-12 without leaving scars or markers (Kolisnychenko et al., 2002) (Fig. 2). The doubling time of MGB469 in LB and minimal medium with and without Casamino acids was similar to that of the parent strain. Surprisingly, the reduced-genome strain reached about 10% higher optical density in all tested media, thus showing that deletion of K-islands provides bacteria with a selective growth advantage under the tested conditions. In the same year, Yu et al. (2002) reported E. coli genome reduction by 6.7 % using the Tn5 transposon-targeted Cre/loxP excision system. The Cre/loxP excision system method is very elegant because it allows deletion of almost any non-essential E. coli DNA (Yu et al., 2002). First, two independent transposon libraries harbouring loxP recombination sites are generated and the transposon insertion sites are identified by sequencing. Next, selected loxP sites are combined in a single strain by transduction. Finally, the genomic locations between the two loxP sites are excised from the chromosome by activation of the Cre recombinase. As in the previous study, the genome-reduced E. coli with four large-scale deletions exhibited normal growth in LB medium. A few years later, Hashimoto et al. (2005) reported the construction of a series of E. coli mutants with marker-less deletions between 2.4 and 29 % of the genome. Although the strain with the largest deletion was viable in the rich medium, its nucleoid and cell morphology were different. Several small nucleoids were localized peripherally close to the cell envelope. Moreover, the cells of the deletion mutants were wider and longer than those of the parent strains. This study revealed that large parts of the E. coli genome can be deleted without a negative effect on growth. To obtain a stable and robust growing reduced-genome E. coli strain, Posfai et al. (2006) first identified non-essential genome parts, such as mobile and cryptic virulence genes. Next, individual large-scale, scarless deletions were made and growth of the resulting strains was tested on minimal medium. Finally, deletions not affecting
growth were serially combined in a single strain resulting in the removal of 15.3% of the chromosome. Besides confirming that bacteria with a considerably reduced genome can be constructed in the laboratory, this study also showed that the reduced-genome strains are endowed with beneficial properties such as increased electroporation efficiencies and decreased mutation rates (see below) (Pósfai et al., 2006). Two studies by Mizoguchi et al. (2007, 2008) and Kato & Hashimoto (2007, 2008) removed 22 and 30% of the E. coli genome, respectively. The regions targeted for deletion were selected by comparative genomics between E. coli and the natural reduced-genome symbiotic bacteria Buchnera spp. Before combination of deletions in a single strain each intermediate deletion strain was characterized to maintain robust growth in minimal medium. Similar to the work of Pósfai et al. (2006) the resulting streamlined E. coli strain MGF-01 lacking 22% of the genome was endowed with emerging properties. Furthermore, strain MGF-01 did not exhibit any growth defects in the exponential phase and grew to a 1.5 times higher final cell density than the wild-type in M9 minimal medium, which is a trait beneficial for biotechnology applications (see below) (Mizoguchi et al., 2008). Recently, two reduced-genome E. coli K-12 strains DGF-327 and DGF-298 lacking 29 and 35.2% of the chromosome, respectively, were constructed (Hirokawa et al., 2013). Interestingly, during the construction of strains DGF-327 and DGF-298 bacteria accumulated mutations in the rph and ilvG1 genes encoding RNase PH and the large subunit of the acetohydroxybutanoate synthase, respectively. These mutations are beneficial because they enable sufficient production of pyrimidine and isoleucine needed for growth in minimal medium (Jensen, 1993; Lawther et al., 1981). It is tempting to assume that the frameshift mutations in rph and ilvG1 had accumulated during the ‘domestication’ of E. coli K-12 (Blattner et al., 1997). The ilvG1 mutation disrupts one of the three biosynthetic pathways for isoleucine and valine while the mutation at the end of rph reduces expression of the downstream pyrE, which in turn increases the demand.

Fig. 2. Genome-scale engineering in B. subtilis and E. coli. Timeline showing the major achievements in the large-scale engineering of the B. subtilis and E. coli genomes.
for pyrimidine. Thus, the cryptic genetic information can be recovered in the reduced-genome bacteria by the accumulation of spontaneous mutations and selection for faster growing isolates. The step-wise genome reduction and evolution of the reduced-genome bacteria for robust growth might be an elegant strategy to promote genome minimization (see below).

Taken together, several studies showed that large regions of the *E. coli* and *B. subtilis* genomes are dispensable under standard laboratory growth conditions. However, to further reduce the genomes of these two model organisms it is crucial that we understand the functions of uncharacterized genes and the functional and genetic interactions between essential genes (both known and unknown) in the cell. This knowledge could prevent genetic and phenotypic instabilities caused by severe genome surgeries during the genome streamlining process. Partially this lack of knowledge can be replenished in the process of genome minimization. Moreover, better understanding of cellular processes is also necessary for advancements in the bottom-up construction of *E. coli*- and *B. subtilis*-based minimal cells.

### Combining genome reduction and evolution for the construction of fast-growing minimal cells

Besides numerous beneficial traits (Manabe et al., 2013; Pósfai et al., 2006), genome reductions in *B. subtilis* and *E. coli* often led to unwanted phenotypes, such as growth defects, and aberrant cell and nucleoid morphologies (Hashimoto et al., 2005). Thus, a knowledge-driven approach should be used to minimize the risk that *B. subtilis* and *E. coli* genome reductions end up in a ‘blind alley’. The genomic regions to be deleted should be carefully selected based on current information about the target genes stored in the community-curated databases SubtiWiki and EcoliWiki (Mcintosh et al., 2012; Michna et al., 2014). Even if the target gene was not designated essential, its inactivation can cause a severe growth defect. For instance, deletion of *rocG* encoding the glutamate dehydrogenase RocG severely affects growth of *B. subtilis* in complex medium (Gunka et al., 2013). This is not surprising as RocG is a bifunctional enzyme active both in glutamate catabolism and in controlling DNA-binding activity of the transcription factor GltC (Commachau et al., 2007). In the absence of RocG, GltC activates the expression of gltA and gltB involved in glutamate biosynthesis, which is not required on complex medium. Simultaneous uptake of glutamate from the medium and blocking its catabolism might increase glutamate concentration to toxic levels. Furthermore, the *B. subtilis* and *E. coli* genomes may contain genes encoding enzymes with redundant activities. For instance, the reduced-genome *B. subtilis* should synthesize at least one enzyme producing the recently discovered essential second messenger c-di-AMP (Corrigan & Gründling, 2013; Mehne et al., 2013). Moreover, multiple rRNA operons were shown to be essential for the efficient growth of *B. subtilis* (Yano et al., 2013). Thus, each genomic region targeted for deletion and the order of deletions has to be considered carefully as strains with severe growth defects do not provide a suitable basis for further genome reduction. Once the reduced-genome strain retaining robust growth is obtained it can be used for the next sequential deletions. In the ideal case, the deletions can be made until construction of the ultimate reduced-genome strain (Fig. 3). However, it is plausible to assume that several deletions will lead to growth defects. Therefore, the isolation of robustly growing strains by continuous selection under the desired conditions might identify suppressor mutants with enhanced growth rates. For instance, mutations in *rpoC*, an essential *E. coli* gene encoding the β′ subunit of RNA polymerase, were shown to increase the growth rate of *E. coli* K-12 MG1655 in glycerol M9 minimal medium by 60 %. Furthermore, *rpoC* mutants converted the carbon source to biomass up to 35 % more efficiently than the wild-type strain (Conrad et al., 2010). These phenotypes resulted from the massive adaptive rewiring of the *E. coli* transcription network via extensive changes in the gene expression profile and reprogramming of the kinetic parameters of RNA polymerase (Conrad et al., 2010). The identification of growth defect-relieving mutations by whole-genome sequencing might uncover new insights into gene functions, interactions and moonlighting (Copley, 2012) (Fig. 3). It was shown that the ‘decryption’ of *B. subtilis* and *E. coli* genetic information by spontaneous mutagenesis during growth under selection can recover growth defects of reduced-genome strains (Hirokawa et al., 2013). Moreover, overexpression of non-essential genes can completely compensate for the loss of essential genes (Bergmüller et al., 2012). Thus, each deletion strain has to be carefully analysed and tested for the presence of growth defect-restoring suppressor mutations. Once an ‘extremely’ reduced-genome strain has been isolated, the remaining non-essential genes can be identified by saturated transposon mutagenesis. Finally, the identified dispensable genomic regions can be removed employing the tools for marker-free gene deletions (Fig. 3).

### Minimal cell factories

*B. subtilis* and *E. coli* genome reductions often led to beneficial traits (Manabe et al., 2011, 2013; Mizoguchi et al., 2008). *B. subtilis* strain MGB874 lacking 20.7 % of the genome produced significantly higher levels of the heterologous enzymes alkaline cellulase and protease than the parent strain (Morimoto et al., 2008). The production of extracellular enzymes by *B. subtilis* MGB874 was further enhanced by deleting genes from the arginine degradation pathway (Manabe et al., 2011, 2013). Reductions of the *E. coli* genome also led to beneficial properties. *E. coli* strains without redundant metabolic pathways produced higher quantities of ethanol from hexose and pentose sugars (Trinh et al., 2008). *E. coli* lacking 22 % of the genomic DNA produced 2.4-fold more threonine that the wild-type strain (Mizoguchi et al., 2008). The *E. coli* strains MDS41–43 generated by deleting up to 15.3 % of the dispensable genomic DNA, such as insertion sequences, exhibited numerous beneficial traits, for example a decreased mutation rate and increased electroporation efficiency (Pósfai
The outcome of an essential genes identification study is strongly influenced by growth conditions and the genetic background of the investigated cell. As described above, B. subtilis and E. coli share approximately 50% of their essential genes. Currently, there is only one essential gene of unknown function in B. subtilis, while the functions of 13 essential ORFs are uncharacterized in E. coli. In the past years, B. subtilis and E. coli were subjected to genome reductions that often led to unexpected beneficial traits, while not having negative impacts on fitness. On the basis of our current knowledge about essential genes we are yet unable to create tailor-made minimal cells for industrial applications encoded by genomes composed of nothing else but essential genes. However, advances highlighted in this review suggest that the minimized cell generated from B. subtilis and E. coli could be used as a suitable chassis for biotechnology applications. It will be very interesting to see which organism provides the first chassis for the true minimal cell factory.

Acknowledgements

Cordial thanks to all researchers who contributed to advancements in the analysis of E. coli and B. subtilis essential genes and to investigation of minimal genomes in general. Work in the authors’ laboratories is supported by the DFG (grant CO 1139/1-1), the Fonds der Chemischen Industrie (http://www.vci.de/fonds) and the Göttingen Centre for Molecular Biology (GZMB). D.R.R. was supported by the Göttingen Graduate School for Neuroscience and Molecular Biosciences (DFG grant GSC 226/2). The authors declare no conflicts of interest associated with this paper.

References


Jensen, K. F. (1993). The *Escherichia coli* K-12 “wild types” W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. *J Bacteriol* 175, 3401–3407.


