Communities of niche-optimized strains: small-genome organism consortia in bioproduction

Stephan Noack[*#], Meike Baumgart[*#]

Institut für Bio- und Geowissenschaften, IBG-1: Biotechnologie, Forschungszentrum Jülich, Jülich, Germany

#Address correspondence to Stephan Noack, s.noack@fz-juelich.de or Meike Baumgart, m.baumgart@fz-juelich.de

*authors contributed equally

Lab websites:

Stephan Noack: http://www.fz-juelich.de/ibg/ibg-1/EN/Research/SystemsBiotechnology/bioproc/micropheno_research.html?nn=981220

Meike Baumgart: http://www.fz-juelich.de/ibg/ibg-1/EN/Research/SystemicMicrobiology/regulation/regulation_node.html

Keywords: Genome reduction, streamlining, minimal genome, relevant genes, CoNoS

Abstract

Bacterial genome reduction is a common phenomenon in nature that has also inspired several projects with biotechnologically relevant organisms. In many cases, however, no performance improvements with the streamlined chassis were obtained. A comparative analysis of natural and synthetic reduced genomes and their corresponding ecological niches reveals that *i*) only deleting expressed genes can lead to noticeable energy conservation usable for improving biomass or product synthesis and *ii*) the simplified strains depend on the environment, *e.g.*, another community member, to complement deleted functions. Based on this analysis, we introduce the concept of Community of Niche-optimized Strains (CoNoS) as an alternative strategy to generate superior biotechnological production processes.

Can nature teach us how to design superior genome reduced strains for biotechnological applications?

Genome reduction is currently a popular topic both in fundamental research and applied biotechnology. On the one hand, researchers are studying how species have evolved towards smaller genomes in their natural environment and how the resulting pathogenic or symbiotic relationships are beneficial for the survival of single species or whole communities. On the other hand, there is an increasing interest for genome-reduced strains with superior performance, *e.g.*, for bio-based production of chemical compounds. Despite the great efforts to construct such strains, there are only a few examples where genome reduction actually led to improved production performance. While discussing possible reasons for the comparably low success rate in biotechnology, one hypothesis was that previous projects often focused too much on the deletion targets, while other potentially relevant parameters, such as the surrounding conditions including, *e.g.*, growth medium and carbon source, were neglected. As nature has often proven to be a good model for all kinds of developments, we were curious whether we could also learn from nature how to generate superior genome reduced strains.

In this opinion, we contrast the current perception of genome reduction as a natural adaptation process with that of a tool for synthetic biology. Based on this analysis we developed the concept of Community of Niche-optimized Strains (CoNoS) as an alternative strategy to generate superior biotechnological production processes. CoNoS focuses strongly on the consideration of surrounding conditions (the niche) as key determinant for suitable gene targets whose deletion can lead to minimization of carbon and energy demands of individual cells in a population. Consequently, the

available feedstock could be utilized more efficiently for biocatalyst supply and product formation.

Genome reduction in nature

In nature, genome reduction is driven by evolution and enables microorganisms to adapt to specific **ecological niches** (see Glossary). Over time, organisms tend to lose genes that are not required for survival in the new niche [1]. Although the environments are rather diverse, the prerequisite for gene loss is usually the supply of the function (*e.g.*, metabolite, protein, etc.) by another organism, either the host or another community member (Figure 1a, Key figure). Typical examples are symbiotic or pathogenic relationships where the host provides a defined environment allowing the loss of certain gene functions [2, 3], or communities of free-living genome reduced strains in aquatic environments [4, 5] (Table 1).

Obligate endosymbionts usually have a long history of co-evolution with their host and are maternally transmitted from the host to the next generation, which is presumably the prerequisite for such extreme genome reduction [1] (Table 1). The population size of bacterial endosymbionts is comparably small, as they are often living in specialized structures such as **bacteriocytes** in an **bacteriome** and only a limited number of cells is transferred to the next host generation [3, 6]. This leads to low effective population sizes (small N_e), making genetic drift one of the main factors for reductive genome evolution in symbiotic relationships (Box 1) [4, 7, 8].

As one example, the obligate endosymbiont *Buchnera aphidicola* provides its aphid host with essential amino acids and some vitamins [3]. Upon additional association with the secondary endosymbiont *Serratia symbiotica, B. aphidicola* lost the ability to synthesize riboflavin and tryptophan, which is now provided by *S. symbiotica*.

Furthermore, the biotin biosynthesis pathway is split between both endosymbionts, leading to multiple **mutualism** [3]. In terms of potential for genome reduction, this means by enriching the ecological niche with an additional metabolite or gene function, the primary endosymbiont can lose further genome content without negative consequences for its own survival in the respective niche. Moreover, this example shows that evolution associated with genome reduction must result in a benefit for the whole consortium, which is likely due to a more efficient utilization of carbon and energy resources mediated through reduced protein synthesis costs. In organisms currently undergoing genome reduction this is often recognizable by a large number of predominantly non-expressed pseudogenes within the genome, *e.g.* in *Mycobacterium lepraemurium* with 1139 pseudogenes in 4.05 Mbp genome [2] (Table 1).

Interestingly, in recent years several examples for genome reduction of free living organisms have been described. Similar to the environments of symbionts, genome reduction also appears in stable environments such as the pelagic zone of freshwater lakes or the ocean [4, 5]. However, since these environments are very nutrient poor and the cell populations much larger (large N_e), the genome reduction follows a different path called “streamlining” (Box 1) [4].

Streamlined organisms have lost certain gene functions that are considered to be important, if not essential. One famous example is the enzyme catalase-peroxidase, which catalyzes the detoxification of hydrogen peroxide (H_2O_2) into O_2 and H_2O [9]. H_2O_2 is generated in the ocean by photo-oxidation of dissolved organic carbon and is able to cross biological membranes similarly to water. Catalase-peroxidase (encoded by *katG*) is missing in all *Prochlorococcus* isolates, but present in most strains of the closely related genus *Synechococcus* (Table 1). A common assumption is that *katG*

was present in their common ancestor and lost later on in the *Prochlorococcus* group. Therefore, it seems sufficient that only a part of the microbial community produces catalase (helpers) to protect all other bacteria (beneficiaries) by eliminating H_2O_2 from water (Figure 1a). How is this equilibrium maintained? The loss of catalase production provides a small selection benefit to the beneficiaries, *e.g.*, by saving iron and energy for protein production. This small benefit ensures that the beneficiaries are not overgrown by the helper cells. In contrast, the helpers are required by the whole community, so the share of helpers will only decrease to such an extent that the survival of the beneficiaries is still ensured. This concept was named “Black Queen Hypothesis” [9].

Man-made reduced genomes

While the study of genome reduction as a natural process has a long tradition in environmental microbiology, its application as a tool in biotechnology or basic research is rather new. According to their motivation, the corresponding genome reduction projects can be assigned to one of the two following aims.

First, to gain a more fundamental understanding of the basic principles of autonomous cellular life, the creation of **minimal cells** is sought by deletion of all **non-essential genes** in a given microbial host (Box 2). This approach was demonstrated for parasitic prokaryotes such as *Mycoplasma sp.* [10] and for industrial model organisms such as *Escherichia coli* and *Bacillus subtilis* [11, 12] (see Table 2 for recent developments). Currently, the quest for minimal bacterial genomes exclusively focuses on such cells that are capable of autonomous replication in undefined, nutrient rich environments. These studies are undoubtedly pioneering and important, but gained insights of minimal cell metabolism are significantly affected by the applied growth and test conditions [13].

Second, to improve the bio-based production of chemical compounds, the construction of microbial **chassis** strains is sought by deletion of **irrelevant genes** (Box 2) that are not required in the specific niches of artificial bioreactor environments (Figure 1b). This synthetic streamlining of bacterial genomes was carried out for a number of industrial workhorses including *E. coli*, *B. subtilis*, *Corynebacterium glutamicum*, *Lactococcus lactis*, and *Pseudomonas sp.* [14, 15] (see Table 2 for recent developments). In summary, there are some promising examples where streamlining led to genome-reduced strains with improved production properties, *e.g.* for heterologous protein production. Only in a few cases, however, the **biological fitness** of these chassis was superior to the parental strain.

What are the differences between natural and man-made reduced-genomes?

In genome reduction projects with *E. coli*, *B. subtilis* and *C. glutamicum* large genomic regions were deleted from the respective wild-type strains and, due to potential energy savings, a greater positive effect on biomass formation (growth rate and yield) was expected. Obviously, this expectation cannot be fulfilled based on the current selection strategies for deletion targets, consisting predominantly of non-expressed genes (see Box 3 for more details).

For all natural examples discussed here, the environment or the ecological niche was crucial for an improved fitness upon genome reduction. When comparing the niches of natural small genome organisms with the laboratory growth conditions of “man-made” genome reduced strains, the most apparent difference is the community structure on the one hand and the isolated culture of a single strain on the other hand (Fig. 1a and b). Although the use of a single strain reduces complexity in the context

of basic research and biotechnological production, it is a rather artificial situation. Therefore, certain niche characteristics that could make genome-reduced strains superior might be absent (see Outstanding Questions).

Conclusively, for chassis construction an advanced strategy would be to define the niche “bioreactor” and focus on the deletion of irrelevant genes that are highly expressed but whose gene products are not required under bioreactor conditions. The resulting **lean-proteome** strains are expected to show superior bioprocess performance under optimal environmental conditions [16]. Indeed, the deletion of flagellar genes in *P. putida* was shown to increase ATP content and, hence, growth and production properties of the genome-reduced variant on defined D-glucose media (Table 2). Moreover, recent studies with different model hosts showed that the cellular amounts of some central metabolic enzymes are higher than necessary for maintaining optimal growth in bioreactor environments [17-20]. Consequently, there is also a large potential for saving carbon and energy resources by targeted knockdown of relevant genes.

Community of Niche-optimized Strains (CoNoS)

Based on the comparative analysis of natural and man-made reduced genomes, we propose the concept for a “Community of Niche-optimized Strains” (CoNoS) as a possible guideline for future genome reduction projects for biotechnological application (Figure 1c).

In the first step, the niche needs to be described as precisely as possible. For a production process this relates to classical parameters, *e.g.*, growth medium, macro- and micronutrients, temperature, pH(-range), product(s) and possible stresses.

Moreover, the niche description should be as closely as possible to the later (potentially large-scale) production conditions [21].

The next important step is to define evaluation criteria (related to growth and production phenotypes), because these determine the relevant gene set for the final strains and, in the optimized versions, their expression levels. In most genome reduction projects, the specific growth rate is used as one independent criteria to judge on the biological fitness of a minimal cell or chassis (Table 2). Following the CoNoS concept, a clear focus is put on novel production processes and, thus, additional criteria such as product titer and productivity come into play. Since not all possible effects of gene deletion or knockdown can be predicted, these criteria are used later to test intermediate strains for their performance.

Afterwards, suitable community members have to be selected, which should enable stable cultivations and those properties are expected to be beneficial for the whole community. Depending on the niche and the target process, the community might be designed from equal or mixed species [22].

When all these surrounding conditions have been set, only then would it be possible to define the target genes to be deleted or knocked down. As learned from nature, these could include pathways for amino acid biosynthesis that are provided by another community member or certain stress response factors for stresses that do not appear under the defined conditions (Figure 1c).

Based on this theoretical concept, engineered strains should be built and tested separately to see whether the prediction is accurate. Most importantly, strain characterization should be as comprehensive and reproducible as possible not to overlook any negative side-effects of genetic engineering. At this stage, external supplementation according to the introduced auxotrophies is required, and the

resulting phenotypic data can be used as reference to check for altered properties of the same strain in the community. Novel technologies for automated and fast strain phenotyping are readily available to support this step [23].

As the final step, the CoNoS is created and, again, thoroughly tested regarding growth and production performance. At this point, further optimization of growth and production performances of the whole community or selected single strains, *e.g.*, with the help of automated laboratory evolution approaches [24], is highly promising since it simply copies what happens in nature.

Concluding Remarks and Future Perspectives

The main idea behind CoNoS in the context of biotechnological production is to optimize carbon and energy use within the whole cell culture to speed-up and increase biocatalyst supply for both growth-coupled and -decoupled production processes. This is especially important for bulk products, such as amino acids, where also small yield increases can make a big difference, and for products with high energy demands, such as proteins. In the following, we will disclose one possible implementation example for CoNoS and discuss potential challenges and how they might be overcome.

Let's assume we want to generate a superior community for the production of a certain amino acid *A*. The niche for this process is defined as following: large-scale bioreactor operated in fed-batch mode, media with known carbon and energy sources, controlled process parameters (*e.g.*, pH, p_{O2}, T) and possible bioreactor inhomogeneities. The latter represents a significant challenge [25], which has to be thoroughly addressed during strain design. The criteria for process evaluation are

amino acid titer, yield and productivity and the CoNoS should exhibit robust and stable production performance.

In our example, the CoNoS is composed of two genome-reduced strains of one species: the producer cell that is auxotrophic for another amino acid *B*, and the helper cell that is auxotrophic for the product amino acid *A*. In the community, they share the catalytic power of two biosynthesis pathways and cross-feed each other with the essential amino acids. In contrast to earlier genome reduction projects, primarily relevant (expressed) genes are targeted for deletion or knockdown that result in carbon and energy savings and thereby can improve production. Clearly, the depicted approach will require additional energy for the transport of amino acids in and out of the cells, but this demand should be negligible in comparison to the resources that can be saved by a more efficient usage of biosynthetic enzymes. Preferably, community members are created from genome-reduced chassis that have already been freed in large parts from irrelevant genes [26] thereby enabling further simplification of cellular metabolism and better predictability as a long term goal [27]. Most importantly, all genes required for the utilization of available carbon and energy resources, as well as genes coding for the regulation of stress responses induced by bioreactor inhomogeneities, should be maintained in each engineered strain.

Most of the relevant metabolic pathways do not operate isolated in the cell, but are highly interlinked with other reaction steps by sharing common intermediates, generating ATP and recycling cofactors. Consequently, metabolic engineering should be supported by metabolic modeling to design strains with a high potential for increasing the fitness of the bacterial community [28]. Clearly, not all model predictions can be realized in the chassis and some of them will lead to unwanted

side-effects under *in vivo* conditions. To quickly evaluate the suitability of certain amino acid biosynthesis pathways as targets for our community design, whole pathways can be silenced *e.g.* by CRISPR interference (CRISPRi), which also works for more than one gene simultaneously [29]. Although clean chromosomal deletions are likely preferred for final strain design, CRISPRi allows a much faster pathway evaluation, even in different strain backgrounds. The resulting strains will be thoroughly experimentally characterized with external supply of the respective amino acid and only pathway deletions with positive effects should be considered for further community design. A comparable approach with *E. coli* has already proven the feasibility of this procedure [20]. For growth-decoupled processes, growth rate and product formation are equally important to reach high overall productivities. Therefore, a targeted shut-off of growth-relevant genes at the transition to production phase might be another interesting approach. This would minimize energy demand for growth-associated maintenance and more carbon could be directed toward the desired products.

Finally, the CoNoS is established and optimized by applying adaptive laboratory evolution using miniaturized and automated repetitive batch approaches [24] and model-based experimental design strategies [30]. In this way, material and personal costs for long-term experimentation can be minimized. Moreover, the repetitive batch mode enables easier co-cultivation of species with significant differences in growth rate without facing the typical washout problem of continuous culture approaches.

CoNoS is a universal concept that can potentially be applied not only for communities of different strains of the same species, but also for consortia of different species or even more complex organism combinations and the outlined step-by-step procedure helps to consider many of the factors that later influence production.

Acknowledgements

Patrick Keeling and Vittorio Boscaro for additional information related to their publication and Julia Frunzke for helpful discussion. This work was partly funded by the Bioeconomy Science Center (BioSC, Grant. No. 325-40000213) as part of the Focus FUND project “HyImPAct – Hybrid processes for important precursor and active pharmaceutical ingredients”.

References

1 Moran, N.A. and Bennett, G.M. (2014) The tiniest tiny genomes. *Annu. Rev. Microbiol.* 68, 195-215

2 Benjak, A., *et al.* (2017) Insights from the genome sequence of *Mycobacterium lepraemurium*: massive gene decay and reductive evolution. *MBio* 8, e01283-01217

3 Manzano-Marin, A., *et al.* (2016) Reinventing the wheel and making it round again: evolutionary convergence in *Buchnera-Serratia* symbiotic consortia between the distantly related lachninae aphids *Tuberolachnus salignus* and *Cinara cedri*. *Genome Biol. Evol.* 8, 1440-1458

4 Giovannoni, S.J., *et al.* (2014) Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1553-1565

5 Neuenschwander, S.M., *et al.* (2017) Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* 12, 185-198

6 Salem, H., *et al.* (2017) Drastic genome reduction in an herbivore's pectinolytic symbiont. *Cell* 171, 1520-1531

7 Batut, B., *et al.* (2014) Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* 12, 841-850

8 Charlesworth, B. (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10, 195-205

9 Morris, J.J., *et al.* (2012) The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *MBio* 3, e00036-00012

10 Hutchison, C.A., 3rd, *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science* 351, aad6253

11 Reuß, D.R., *et al.* (2016) The blueprint of a minimal cell: MiniBacillus. *Microbiol. Mol. Biol. Rev.* 80, 955-987

12 Juhas, M., *et al.* (2014) *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiology* 160, 2341-2351

13 Juhas, M. (2016) On the road to synthetic life: the minimal cell and genome-scale engineering. *Crit. Rev. Biotechnol.* 36, 416-423

14 Sung, B.H., *et al.* (2016) Construction of a minimal genome as a chassis for synthetic biology. *Essays Biochem.* 60, 337-346

15 Choe, D., *et al.* (2016) Minimal genome: Worthwhile or worthless efforts toward being smaller? *Biotechnol. J.* 11, 199-211

16 Valgepea, K., *et al.* (2015) Lean-proteome strains - next step in metabolic engineering. *Front. Bioeng. Biotechnol.* 3, 11

17 Voges, R., *et al.* (2015) Absolute quantification of *Corynebacterium glutamicum* glycolytic and anaplerotic enzymes by QconCAT. *J. Proteomics* 113, 366-377

18 Noack, S., *et al.* (2017) The linkage between nutrient supply, intracellular enzyme abundances and bacterial growth: New evidences from the central carbon metabolism of *Corynebacterium glutamicum*. *J. Biotechnol.* 258, 13-24

19 Muntel, J., *et al.* (2014) Comprehensive absolute quantification of the cytosolic proteome of *Bacillus subtilis* by data independent, parallel fragmentation in liquid chromatography/mass spectrometry (LC/MS(E)). *Mol. Cell. Proteomics* 13, 1008-1019

20 D'Souza, G., *et al.* (2014) Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution* 68, 2559-2570

21 Löffler, M., *et al.* (2016) Engineering *E. coli* for large-scale production - Strategies considering ATP expenses and transcriptional responses. *Metab. Eng.* 38, 73-85

22 Sgobba, E., *et al.* (2018) Synthetic *Escherichia coli-Corynebacterium glutamicum* consortia for l-lysine production from starch and sucrose. *Bioresour. Technol.* 260, 302-310

23 Unthan, S., *et al.* (2015) Bioprocess automation on a Mini Pilot Plant enables fast quantitative microbial phenotyping. *Microb. Cell Fact.* 14, 32
24 Radek, A., *et al.* (2017) Miniaturized and automated adaptive laboratory evolution: Evolving *Corynebacterium glutamicum* towards an improved d-xylose utilization. *Bioresour. Technol.* 245, 1377-1385
25 Limberg, M.H., *et al.* (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
26 Baumgart, M., *et al.* (2018) *Corynebacterium glutamicum* chassis C1*: building and testing a novel platform host for synthetic biology and industrial biotechnology. *ACS Synth. Biol.* 7, 132-144
27 Cvijovic, M., *et al.* (2014) Bridging the gaps in systems biology. *Mol. Genet. Genomics* 289, 727-734
28 McNally, C.P. and Borenstein, E. (2018) Metabolic model-based analysis of the emergence of bacterial cross-feeding via extensive gene loss. *BMC Syst Biol* 12, 69
29 Peters, J.M., *et al.* (2016) A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* 165, 1493-1506
30 LaCroix, R.A., *et al.* (2017) A model for designing adaptive laboratory evolution experiments. *Appl. Environ. Microbiol.* 83
31 Petrovska, L., *et al.* (2017) Genome reduction for niche association in *Campylobacter hepaticus*, a cause of spotty liver disease in poultry. *Front. Cell. Infect. Microbiol.* 7, 354
32 Rocha, E.P. and Danchin, A. (2002) Base composition bias might result from competition for metabolic resources. *Trends Genet.* 18, 291-294
33 Grzymski, J.J. and Dussaq, A.M. (2012) The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J.* 6, 71-80
34 Orth, J.D., *et al.* (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* 7, 535
35 Henry, C.S., *et al.* (2009) *i*Bsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biol.* 10, R69
36 Kjeldsen, K.R. and Nielsen, J. (2009) In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol. Bioeng.* 102, 583-597
37 Boscaro, V., *et al.* (2017) Parallel genome reduction in symbionts descended from closely related free-living bacteria. *Nat. Ecol. Evol.* 1, 1160-1167
38 Manzano-Marin, A. and Latorre, A. (2016) Snapshots of a shrinking partner: Genome reduction in *Serratia symbiotica*. *Sci. Rep.* 6, 32590
39 Van, T.T., *et al.* (2016) *Campylobacter hepaticus* sp. nov., isolated from chickens with spotty liver disease. *Int. J. Syst. Evol. Microbiol.* 66, 4518-4524
40 Fraser, C.M., *et al.* (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397-403
41 Fookes, M.C., *et al.* (2017) *Mycoplasma genitalium*: whole genome sequence analysis, recombination and population structure. *BMC Genomics* 18, 993
42 Gibson, D.G., *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52-56
43 Manso-Silvan, L., *et al.* (2009) *Mycoplasma leachii* sp. nov. as a new species designation for *Mycoplasma* sp. bovine group 7 of Leach, and reclassification of *Mycoplasma mycoides* subsp. *mycoides* LC as a serovar of *Mycoplasma mycoides* subsp. *capri*. *Int. J. Syst. Evol. Microbiol.* 59, 1353-1358
44 Brewer, T.E., *et al.* (2016) Genome reduction in an abundant and ubiquitous soil bacterium '*Candidatus* Udaeobacter copiosus'. *Nat. Microbiol.* 2, 16198
45 Delmont, T.O. and Eren, A.M. (2018) Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* 6, e4320

382 46 Biller, S.J., *et al.* (2015) *Prochlorococcus*: the structure and function of collective
383 diversity. *Nat. Rev. Microbiol.* 13, 13-27
384 47 Dufresne, A., *et al.* (2003) Genome sequence of the cyanobacterium *Prochlorococcus*
385 *marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl. Acad. Sci. USA* 100,
386 10020-10025
387 48 Zhou, J., *et al.* (2016) CasHRA (Cas9-facilitated Homologous Recombination Assembly)
388 method of constructing megabase-sized DNA. *Nucleic Acids Res.* 44, e124
389 49 Reuß, D.R., *et al.* (2017) Large-scale reduction of the *Bacillus subtilis* genome:
390 consequences for the transcriptional network, resource allocation, and metabolism. *Genome*
391 *Res.* 27, 289-299
392 50 Yuan, X., *et al.* (2017) Single-cell microfluidics to study the effects of genome deletion on
393 bacterial growth behavior. *ACS Synth. Biol.* 6, 2219-2227
394 51 Karcagi, I., *et al.* (2016) Indispensability of horizontally transferred genes and its impact
395 on bacterial genome streamlining. *Mol. Biol. Evol.* 33, 1257-1269
396 52 Couto, J.M., *et al.* (2018) The effect of metabolic stress on genome stability of a synthetic
397 biology chassis *Escherichia coli* K12 strain. *Microb. Cell Fact.* 17, 8
398 53 Park, M.K., *et al.* (2014) Enhancing recombinant protein production with an *Escherichia*
399 *coli* host strain lacking insertion sequences. *Appl. Microbiol. Biotechnol.* 98, 6701-6713
400 54 Wenzel, M. and Altenbuchner, J. (2015) Development of a markerless gene deletion
401 system for *Bacillus subtilis* based on the mannose phosphoenolpyruvate-dependent
402 phosphotransferase system. *Microbiology* 161, 1942-1949
403 55 Zhu, D., *et al.* (2017) Enhanced heterologous protein productivity by genome reduction in
404 *Lactococcus lactis* NZ9000. *Microb. Cell Fact.* 16, 1
405 56 Shen, X., *et al.* (2017) Developing genome-reduced *Pseudomonas chlororaphis* strains for
406 the production of secondary metabolites. *BMC Genomics* 18, 715
407 57 Martinez-Garcia, E., *et al.* (2014) *Pseudomonas* 2.0: genetic upgrading of *P. putida*
408 KT2440 as an enhanced host for heterologous gene expression. *Microb. Cell Fact.* 13, 159
409 58 Lieder, S., *et al.* (2015) Genome reduction boosts heterologous gene expression in
410 *Pseudomonas putida*. *Microb. Cell Fact.* 14, 23
411
412

Glossary

Bacteriocytes: specialized host cells containing symbiotic bacteria.

Bacteriome: organ-like structure specialized for hosting of symbiotic bacteria, contains bacteriocytes.

Biological fitness: is a measure for the ability of a (micro)organism to survive in a specific environment. For a microbial strain, the number of new daughter cells that are formed in a certain period of time and from a fixed amount of substrate can be quantified as specific growth rate and yield, respectively.

Chassis: is characterized by a relevant gene set that enables autonomous cell replication under the constraint of at least one biological fitness criterion.

Ecological niche: All biotic and abiotic factors influencing the survival of a certain species.

Labile organic carbon: The portion of soil carbon that can readily be decomposed by soil organisms.

Lean-proteome strain: is created by targeted deletion of irrelevant genes whose protein products represent a significant translational burden.

Minimal cell: is characterized by the smallest possible genome consisting of those essential genes that enable autonomous cell replication in a specific environment.

Mutualism: Two organisms of different species in a symbiotic relationship where both rely on each other.

***N*$_e$:** Effective population size. Description of a natural population as a theoretical population with ideal characteristics and a similar level of diversity. N_e describes the level of genetic drift a species undergoes during evolution. Large N_e leads to low genetic drift, low N_e leads to large genetic drift. For further details see [7, 8].

Pseudogene: a non-functional member of a gene family that derived from a functional gene by mutation.

Text Boxes

Box 1. The environment determines the pathway for genome reduction

The analysis of several examples of natural small genomes reveals that genome reduction in nature is not entirely random. The conditions within the ecological niche determine the driving force and the outcome of the evolved genome. Here we briefly present the different concepts currently available to explain these processes.

Genetic drift is proposed to be the main driving force for genome reduction in symbionts and pathogens (Figure I) and describes random fluctuations in the number of gene variants in a population. It usually appears in small populations or after **population bottlenecks**, such as the transfer of only a few symbiotic bacteria from the host to the following generation. In case of *Stammera*, only 0.02 % of the parent's population is transferred to newly emerging adults [6]. This drastically reduces the genetic diversity in each generation and may allow a faster establishment of beneficial mutations, such as loss of certain genes. For symbionts with extracellular lifestyle, the mode of vertical transmission and the exposure to stresses during this process is also of major interest. In case of *Stammera*, the evolution of the egg caplet is one example how stress can be reduced during transmission [6].

Free living organisms face entirely different environmental conditions which influence the process toward smaller genomes. **Streamlining** describes the **selection** of cells with reduced cell size and complexity that can use the scarce resources more efficiently (Figure I). While the resource saving of a reduced genome size is negligible under nutrient rich conditions, it can become critical when essential

elements such as nitrogen and phosphorus are very rare, *e.g.*, in surface waters. Moreover, the decreased surface/volume ratio can be beneficial for nutrient uptake [7]. Under these conditions also the replication costs of the genome may become relevant, because only very few pseudogenes are found in these organisms.

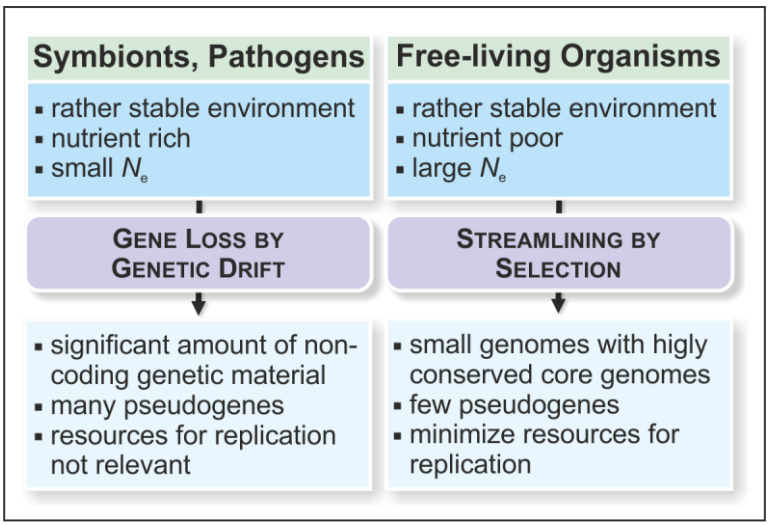

Figure I. The environment determines the path toward reduced genomes and the outcome thereof. Adapted from [4].

Box 2. Essentiality and relevance of genes

The definition of gene essentiality is currently quite diverse and no uniform concept exists that allows differentiation between the minimal cell and chassis approaches. In particular, terms such as “quasi-essential” or “co-essential” are misleading, since they are not applied to characterize cell viability (lethal or not) but biological fitness (*e.g.*, slow or fast growth). Therefore, we suggest strict differentiation between the essentiality and relevance of genes with the following general definitions:

Essential genes encode for the basic properties of any living organism, including encapsulation, information storage, gene expression, and cell replication. In microorganisms, the essentiality of single genes can be decided by blocking the

synthesis of the corresponding gene product (*e.g.* by gene knock-out), followed by testing the viability of the resulting mutant cell. Depending on the cellular environment (*e.g.*, poor or rich medium), the number of essential genes can be high or low. In contrast, **non-essential genes** encode for additional cellular properties that have no influence on cell viability, but can be beneficial for cell growth in a specific environment. Towards the construction of a minimal cell, all non-essential genes are potential deletion targets (Figure II).

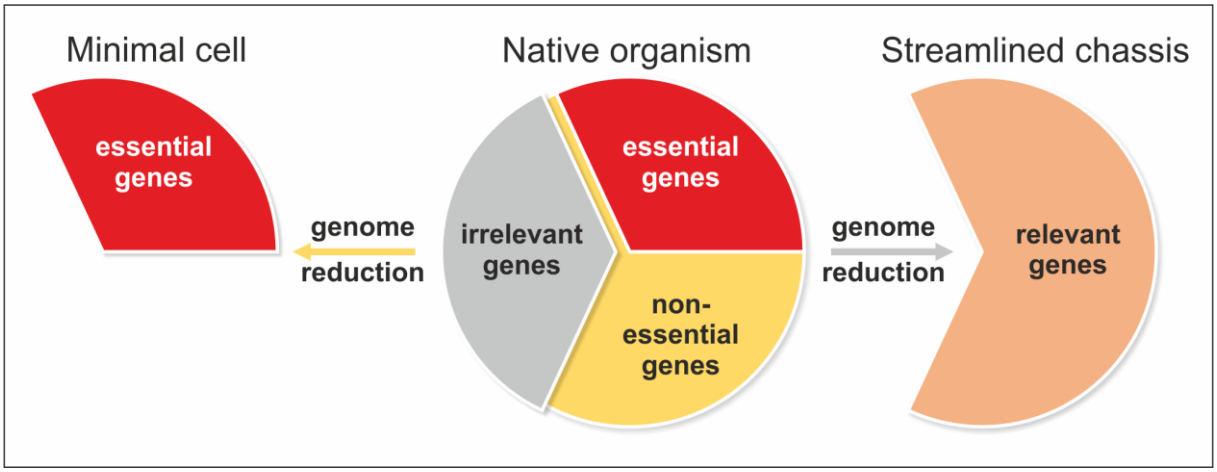

Figure II. Genome reduction in two directions: Depending on the aim, it leads to minimal cells or streamlined chassis. Please note that irrelevant genes represent a part of the non-essential genes.

Relevant genes cover all essential genes and a subset of non-essential genes that encode for cellular properties that are required to ensure the biological fitness of a microorganism strain at a specified level in a predefined environment. The relevance of single genes can usually be determined by blocking the synthesis of the corresponding gene product (*e.g.* by gene knock-out), followed by phenotyping of the resulting mutant cell under the desired environmental conditions. In contrast, **irrelevant genes** encode for cellular properties that are not required for maintaining the biological fitness of a microorganism strain at a specified level in a predefined

environment. Toward the construction of a streamlined chassis, all irrelevant genes are potential deletion targets (Figure II).

Box 3. Targets for synthetic and natural genome reduction

Targets of recent genome reduction projects were predominantly genes with low expression levels and/or unknown function when cells are exposed to the niche "bioreactor". By precisely defining this artificial environment, many genes are classified as irrelevant (Box 2) and can be deleted without any negative effect on the biological fitness of the host (*e.g.*, as shown for the model organism *C. glutamicum*, Table 2). The corresponding energy savings are, however, limited to the genome replication process. As shown in Figure III, these savings are comparably small and, thus, cannot be expected to result in an improved biological fitness of the genome-reduced chassis.

In contrast, natural reduced genomes not only have common deletion targets, but they also share some general characteristics that have evolved convergently. One of these is the tendency towards an increased AT content in the genome [3, 5, 7, 31-33]. A higher AT content in the genome seems advantageous as the synthesis of ATP and UTP is energetically less expensive than the building of GTP and CTP, and the ATP/UTP pools are higher in the cell compared to the GTP/CTP pools [32]. The AT content influences not only the codon usage of the cell, but also the amino acid composition of proteins. Especially in nutrient-poor environments there is a tendency toward amino acids with lower N-content (*e.g.* the exchange of arginine with lysine) [4, 33].

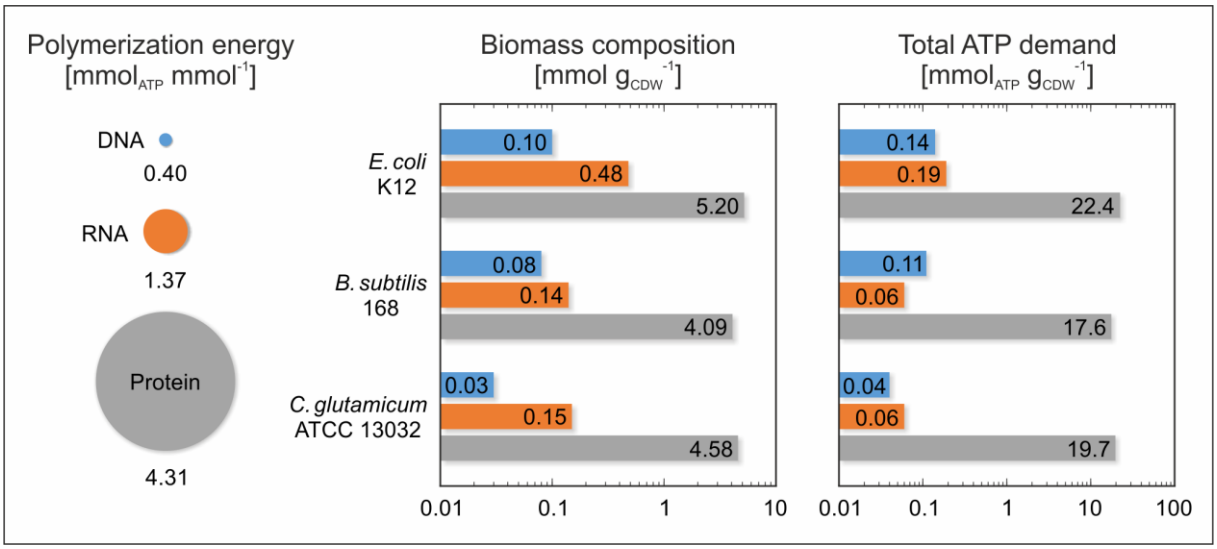

Figure III. Energy demand (ATP) for the synthesis of DNA, RNA and proteins in three different model hosts. Polymerization energies and biomass compositions were derived from [34-36]. The product of both quantities gives the total ATP demand for each macromolecule in one gram of cell dry weight of each species. The energy demand for protein synthesis is more than two orders of magnitude higher than that for DNA and RNA.

Figure Legends

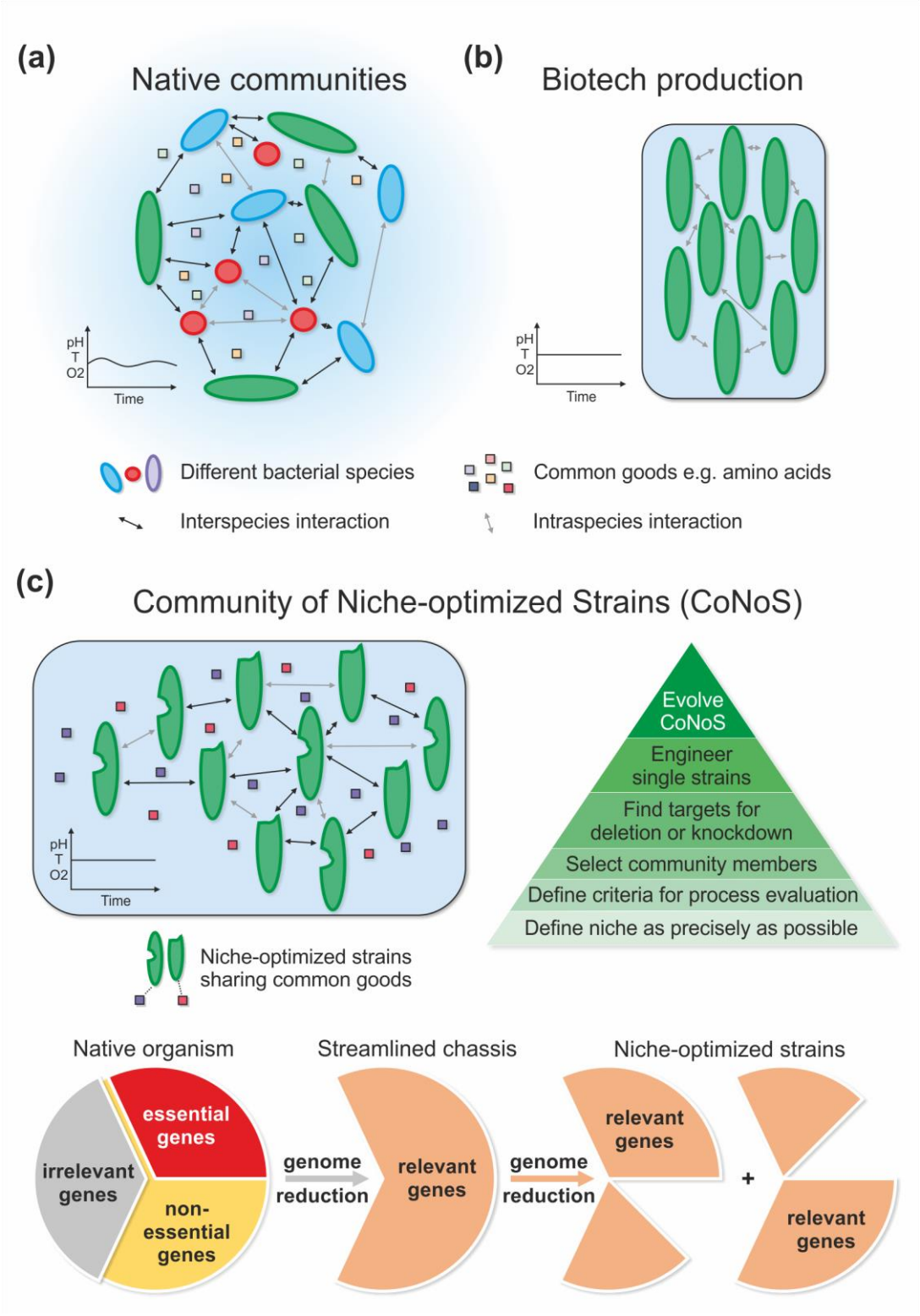

Figure 1. Genome reduction: From nature to communities of niche-optimized strains (CoNoS). (a) Native communities consist of different bacterial species with all kinds of interactions, *e.g.*, the exchange of common goods. (b) Biotechnological production processes are usually based on single engineered strains (including genome-

reduced variants) and, thus, interaction potential is limited. (c) CoNoS can be built from streamlined chassis by additional deletion of specific relevant genes. This will introduce artificial interspecies interactions and, by harnessing the resulting synergies, overall resource utilization in the production process can be optimized.

Table 1. Genome reduction occurring in nature.

Organism[a,b]	Niche description	Reference genome[b]			Reduced genome[c]				Refs
		Mbp	ORFs[d]	GC (%)	Mbp Red. (%)	ORFs[d] Red. (%)	GC (%)	Common targets[e]	
Symbiotic organisms									
Buchnera aphidicola Bcc (produces essential amino acids and some vitamins for host) [*Erwinia billingiae* Eb661]	aphids, *e.g. Acyrthosiphon pisum*, grow intracellularly in specialized cells called bacteriocytes, often together with *Serratia symbiotica*	5.37	4828 [77]	55.0	0.43 (-92.1)	367 [3] (-92.4)	20.1	AB[W,R], CVB [riboflavin, biotin], PB	[3]
Polynucleobacter (9 strains) [free living *Polynucleobacter duraquae* MWH-MoK4]	obligate intracellular endosymbionts of closely related ciliate hosts (*Euplotes)*	2.03	2023 [19]	45.2	1.55 (-23.6)	1172 [n.a.] (-42.1)	46.4	CVB[thiamine] RS	[37]
Serratia symbiotica STs [*Serratia marcescens* Db11]	obligate secondary endosymbiont in aphids with *Buchnera*	5.11	4678 [45]	59.5	0.65 (-87.3)	492 [7] (-89.5)	20.9	AB[all EAAs], CVB[thiamin]	[3, 38]
Stammera capleta provides pectin degradation genes [*Erwinia billingiae* Eb661]	tortoise leaf beetle *Cassida rubignosa*, extracellular symbiont, restricted localization, *e.g.* symbiotic organs	5.37	4828 [77]	55.0	0.27 (-95.0)	251 [n.a.] (-94.8)	15.4	AB CVB[B-vitamins], M PB	[6]
Pathogenic organisms									
Campylopacter hepaticus HV10 [*C. jejuni* ATCC 700819]	chicken liver, iron rich, microaerobic, 37-42°C	1.64	1572 [40]	30.5	1.48 (-9.7)	1380 [91] (-12.2)	27.9	AB	[31, 39]
Mycobacterium lepraemurium [*Mycobacterium smegmatis* MC2 155]	obligatory intracellular lifestyle, macrophage phagosome, inhibit phagosome maturation	6.99	6717 [167]	67.4	4.05 (-42.0)	2682 [1139] (-60.1)	69.0	T[sugars]	[2]
Mycoplasma genitalium G37 [*Lactococcus lactis* subsp. lactis Il1403]	human urogenital tract, causes urethritis and cervicitis	2.37	2277 [45]	35.3	0.58 (-75.5)	515 [9] (-77.4)	31.7	AB[almost all], CVB, RS, T,	[40, 41]
Mycoplasma mycoides subspecies *capri* GM12 [*Lactococcus lactis* subsp. lactis Il1403]	ruminant pathogen, causes pneumonia, mastitis, arthritis, anaerobic, similar to JCV-syn1.0	2.37	2277 [45]	35.3	1.09 (-54.0)	850 [59] (-62.7)	23.9	No detailed analysis of this specific strain available	[42, 43]
Free living examples									
Planctophila (6 Species) and *Nanopelagicus* (3 species) *species)* [*Streptomyces coelicolor* A3(2)]	pelagic zone of fresh waters, competition for limited resources, dominated by microbes with small cell volume (<0.1 µm^3), Lake Zurich	9.05	8152 [60]	72.0	1.27 (-86.0) 1.16 (-87.2)	1258 [n.a.] (-84.6) 1153 [7] (-85.9)	48.2 40.2	AB[C, S, H, ornithine, betaine] CVB[thiamine, panthothenate cobalamin, biotin], RS, T[carbohydrates, glutamate, organic acids][f]	[5]
Udaeobacter copiosus [*Chthoniobacter flavus* Ellin428]	soil, preferentially grassland	7.85	6716 [-]	61.1	2.66[g] (-66.2)	3042[g] [n.a.] (-54.7)	54	AB[I, L, V, W, Y, F, R, H] CVB[cobalamin,	[44]
Prochlorococcus marinus CCMP1375 [*Synechococcus elongatus* PCC 6301]	euphotic zone of the tropical and subtropical oligotrophic ocean, continuous macroscale gradients of light, temperature and nutrients, low phosphate	2.70	2602 [43]	55.5	1.75 (-35.1)	1882 [1] (-27.7)	36.4	RS, DR, T, M, CVB[thiamine] catalase	[45-47]

[a] In some cases a group of organisms.

[b] Square brackets indicate the reference organism. For symbionts and pathogens usually a free living related organism. For others related organisms with larger genomes.

[c] for some cases only a selected example strain is given, number of genes according to NCBI Genome database

[d] ORFs = protein coding genes, square brackets indicate number of pseudogenes, round brackets indicate reduction in comparison to reference genome. n.a., data not available. Data according to the NCBI genome database

[e] AB[X], amino acid biosynthesis, X stands for the respective amino acid in 1-letter-code; CVB, cofactor and vitamin biosynthesis; RS, Regulation and Signal transduction; PB, phospholipid biosynthesis; M, motility; Transporters T[X], X stands for the respective transporters; DR, DNA Repair

[f] as the strains differ a lot in their gene composition, only examples are given here

[g] draft sequence, full genome estimated ~2.81 Mbp

547 **Table 2.** Selected examples of man-made reduced genomes.

Organism[a]	Niche	Reference genome		Reduced genome				Refs
		Mbp	ORFs[b]	Mbp Red. (%)	ORFs[b] Red. (%)	Targets	Results	
Minimal cell approach								
Mycoplasma mycoides [JCVI-syn1.0]	Complex SP-4 media with 17 % fetal bovine serum	1.08	901	0.53 (-50.9)	473 (-47.5)	All non-essential genes, retention of selected genes relevant for robust growth	Strain JCVI-syn3.0 with smallest synthetic genome enabling autonomous cell replication, decreased growth rate, altered morphology	[10]
Escherichia coli [MDS42]	-	3.98	3436	1.03 (-74.1)	716 (-79.2)		DNA construct MGE-syn1.0 could not be functionally transplanted to *E. coli*	[48]
Bacillus subtilis [Δ6]	Complex Luria-Bertani media	3.88	3877	2.68 (-30.9)	2648 (-46.4)		Strain PS38 with unaltered growth rate	[49]
Chassis approach								
Escherichia coli [K-12 MG1655]	Defined M9 media with D-glucose	4.64	4140	3.98 (-14.2)	3436 (-17.0)	IS elements, transposases, defective phages, integrases, site-specific recombinases	Strain MDS42 with decreased growth rate and stress robustness, longer lag time, increased electroporation efficiency, decreased mutation rate	[50-52]
	Defined Riesenberg media with D-glucose			3.58 (-22.8)	3135 (-24.3)	Genes deleted in MDS42 and further non-essential genes	Strain MS56 with improved heterologous protein production	[53]
Bacillus subtilis [168]	Complex Luria-Bertani and defined MG1 media with D-glucose and casamino acids	4.22	4174	3.64 (-13.6)	-	9 prophages,7 antibiotic biosynthesis gene clusters, 2 sigma factors for sporulation	Strain IIG-Bs20-4 with unaltered growth rate and morphology	[54]
Corynebacterium glutamicum [ATCC 13032]	Defined CGXII media with different C-sources and complex Brain Heart Infusion media	3.28	2959	2.84 (-13.4)	2547 (-13.9)	3 Prophages, 2 IS elements, 10 irrelevant gene clusters	Strain C1* with unaltered growth rate, biomass yield and stress robustness	[26]
Lactococcus lactis [NZ9000]	Complex M17G media or defined SA media with D-glucose	2.53	2363	2.46 (-2.8)	2246 (-5.0)	4 large irrelevant gene clusters (including prophages and transposons)	Strain 9 k-4 with improved heterologous protein production, increased growth rate, enhanced biomass yield, increased ATP content, diminished maintenance demands	[55]
Pseudomonas chlororaphis [GP72]	Complex Luria-Bertani or King's media	6.63	5855	5.95 (-10.3)	5330 (-9.0)	5 secondary metabolite clusters, 17 non-essential gene clusters	Strain MDS22 with improved secondary metabolite production, decreased growth rate, altered morphology, absent swimming ability	[56]
Pseudomonas putida [KT2440]	Defined M12 media with D-glucose	6.18	5564	5.91 (-4.3)	-	Flagellar genes, 4 prophages, 2 dexoribonucleases, Type I restriction modification system, Transposases Tn*7* and Zn*4652*, Recombinase A	Strain EM383 with improved heterologous gene expression, increased growth rate, enhanced biomass yield, higher viability, increased plasmid stability, increased ATP content, diminished maintenance demands	[57, 58]

548 [a] Square brackets indicate the reference strain.
549 [b] ORFs = protein coding genes, square brackets indicate number of pseudogenes, round brackets indicate reduction in comparison to reference genome. n.a., data not available.
550 Data according to the NCBI genome database