Recent advances in genome annotation and synthetic biology for the development of microbial chassis

Saltiel Hamese[1,2], Kanganwiro Mugwanda[1,3], Mutsa Takundwa[1], Earl Prinsloo[2] and Deepak B. Thimiri Govinda Raj[1*]

Abstract

This article provides an overview of microbial host selection, synthetic biology, genome annotation, metabolic modeling, and computational methods for predicting gene essentiality for developing a microbial chassis. This article focuses on lactic acid bacteria (LAB) as a microbial chassis and strategies for genome annotation of the LAB genome. As a case study, *Lactococcus lactis* is chosen based on its well-established therapeutic applications such as probiotics and oral vaccine development. In this article, we have delineated the strategies for genome annotations of lactic acid bacteria. These strategies also provide insights into streamlining genome reduction without compromising the functionality of the chassis and the potential for minimal genome chassis development. These insights underscore the potential for the development of efficient and sustainable synthetic biology systems using streamlined microbial chassis with minimal genomes.

Keywords Microbial chassis, Synthetic biology, Genome annotation, Gene essentiality, Genome reduction, Lactic acid bacteria (LAB), Lactococcus lactis

Background

Synthetic biology, precision medicine, and nanotechnology are the three emerging research areas that can be applied as converging fields across various industrial sectors. Synthetic biology is described as the design of new biological parts and the (re-)design of existing biological systems for functional applications. Some synthetic biology applications include the development of synthetic microbes as chassis for recombinant therapeutic production and vaccine development. Microbial chassis are versatile platforms where various bacteria are engineered with genetic components for specific functionalities and address unmet application needs. Synthetic biology, entailing the design and manipulation of biological systems, assumes paramount importance in bioengineering and in silico biology. Computational tools for predicting essential genes and facilitating genome reduction are crucial, offering advantages such as simplified metabolism, improved production, and ease of manipulation. Genome annotation is discussed, focusing on identifying and labeling functional elements in a genome sequence. The generation of synthetic microbes or otherwise called microbial chassis requires the design of minimal genomes that are facilitated through genome-scale metabolic (GSM) models and are critical for chassis development [70]. Furthermore, genome-scale metabolic (GSM) models play a vital role

*Correspondence:
Deepak B. Thimiri Govinda Raj
dgovindaraj@csir.co.za
[1] Synthetic Nanobiotechnology and Biomachines Group, Centre for Synthetic Biology and Precision Medicine, Next Generation Health Cluster, CSIR Pretoria, South Africa
[2] Biotechnology Innovation Centre, Rhodes University, PO Box 94, Makhanda 6140, South Africa
[3] Department of Microbiology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

Hamese *et al. Journal of Genetic Engineering and Biotechnology* (2023) 21:156

Page 2 of 9

in understanding metabolic capabilities, resource allocation, and adaptation in microbial chassis.

The advantages of chassis with minimal genome have been reported to reduce organism's complexity by allowing metabolic modeling and functional predictions with higher agility [38]. Improved genome stability has been demonstrated in genome-reduced *Streptomyces chattanoogesis* and *E. coli* strains by deleting biosynthetic clusters and error-prone DNA polymerase [12, 18]. Another major advantage is that microbes with reduced genomes require lower bioenergy and this has been demonstrated with the 6.9% reduction of the genome of *Lactococcus lactis* N8 by deleting prophages and genomic islands, resulting in a shortened generation time by 17% [55]. Other benefits of genome-reduced strains include increased production of desired products, improved transformation efficiency, and ease of genetic manipulation [12]. Finally, genome-reduced strains have the potential to be used for downstream applications such as expressing heterologous genes and producing biomolecules using tailored metabolic pathways [38] due to improved growth characteristics, more straightforward metabolism, and fewer functions being performed within the cell of genome reduced strains. This study outlines computational tools for predicting essential genes and designing genomic deletions to facilitate genome reduction. This study has demonstrated the application of computational synthetic biology using *L. lactis* as an example of microbial chassis with potential applications in vaccine development.

Microbial chassis
Choosing the right microbe as a microbial chassis to re-engineer is critical for synthetic biology-driven applications. Engineering of bacterial chassis is considered the most sought-after versatile platform due to robustness, smaller genome size, and simple transcriptional and translational control. Several microbes like *Mollicutes*, *Pseudomonas*, *Escherichia coli (E. coli)*, *Comamonas testosteroni*, and *Bacillus subtilis (B. subtilis)* have been tailored as microbial chassis. *Mollicutes* chassis which are characterized by their absence of cell walls offer insights into the fundamental boundaries of cell survival and division [23]. *Pseudomonas* chassis excels in metabolizing aromatic compounds, enhancing heterologous gene expression. Large-scale genomic deletions in *Pseudomonas putida* chassis yield cells with robust growth [39, 40]. Similarly, *E. coli* chassis with deleted insertion sequences and auxotrophic phenotypes exhibit improved growth fitness [27]. *Comamonas testosteroni* harnesses its natural pollutant-degrading capabilities, making it a promising bioremediation chassis [1]. *B. subtilis* chassis, including delta6, MG1M, and MGB874, are known for their capacity to enhance extracellular protein productivity. Additionally, gram-positive bacteria, like *B. subtilis*, are favored enzyme producers due to their low immunogenicity and limited extracellular protease production [4, 44, 72]. Furthermore, yeast chassis cells display temperature-sensitive attributes, influencing ethanol and glycerol yields [45]. The choice of microbial chassis depends on specific applications targeted and also requires full genome annotation of the chassis in order to effectively engineer thereby highlighting the significance of host genome annotation.

Genome annotation
Genome annotation identifies functional elements of a genome sequence, indicating its significance. Annotating a genome entails following these steps: identifying genes (including protein-encoding genes and some RNA-encoding genes), predicting the functions of the identified genes, creating metabolic reconstructions and connecting them to genes, labeling phage insertion sequences and transposons, predicting frameshifts and pseudogenes, and identifying regulatory sites and operons, ultimately creating a list of regulons [51]. Regulons are a group of genes or operons that are upregulated or downregulated as a unit by the same protein in response to the same signal. Several genome annotation tools have been developed. These annotation tools may be automated or manual. Automated gene-annotation tools are often used because of the faster annotation and ease of use. However, it is highly recommended that beginners select automatic and semi-automatic annotation methods [31]. Moreover, automatic annotation algorithms, frequently based on orthologs from distantly related model organisms, cannot yet correctly identify all genes within a genome due to confidence and reliability of outcomes as results from different servers or databases are often dissimilar; obtaining accurate gene sets and model manual annotation is often required [21]. Several pipelines for the annotation of genomes have been developed; examples are in Table 1. The gene or protein sequences identified by structural annotation describing the gene structure (e.g., introns, exons, coding sequences, and start and end coordinates) are linked to biological data in a process known as functional annotation, which usually begins with gene identification or gene calling. The different tools for functional annotation are summarized in Table 2. With many genomes sequenced, computational annotation approaches to characterize genes and proteins from their sequences are essential for designing genome deletions.

Hamese *et al. Journal of Genetic Engineering and Biotechnology*   (2023) 21:156

Page 3 of 9

Table 1 Genome annotation pipelines

Pipeline name	Details	Source	URL
PGAP	An automatic prokaryotic genome annotation pipeline, which is a set of computational tools and algorithms that are used to predict the presence and location of genes in the genome of a prokaryotic organism	[63]	https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
Prokka v1.14.6	A software tool for rapidly annotating genomes, which uses external feature prediction tools like RNAmmer and Prodigal to identify the coordinates of genomic features within contigs	[61]	https://github.com/tseemann/prokka
RAST	A fully automated pipeline for annotating bacterial and archaeal genomes	[10]	https://rast.nmpdr.org/
MicrobeAnnotator	A tool that predicts all types of prokaryotic genes from a single or a set of anonymous genomic sequences of varying lengths. This tool is commonly used in the analysis of prokaryotic genomes	[58]	https://github.com/cruizperez/MicrobeAnnotator
EggNOG-mapper v2.1.9	A tool for automatically annotating the function of genes based on precomputed orthology assignments	[13]	http://eggnog-mapper.embl.de/
DRAM	A tool that uses databases like KEGG and PFAM to organize microbial genomic information into a catalog of microbial traits	[62]	https://github.com/WrightonLabCSU/DRAM
DFAST	A pipeline that supports genome submission to the public database DNA Data Bank of Japan (DDBJ) using the GHOSTX algorithm	[67]	https://dfast.ddbj.nig.ac.jp/
GenSAS	An online pipeline that provides structural and functional annotations of genomic sequences	[28]	https://www.gensas.org/
BlastKOALA	A tool that carries out automated annotation of fully sequenced genomes by using KEGG's internal annotation tool, KOALA, to assign K numbers to KEGG genes using SSEARCH computation	[33]	https://www.kegg.jp/blastkoala/

Table 2 Functional annotation tools that can be used in microbial genome annotation

Tool	Description	Source
(Meta)GeneMark	A webpage that provides access to gene prediction in metagenomes	[24]
(Meta)Prodigal	Gene prediction software that is used to identify protein-coding genes in prokaryotic genomes	[29]
MetaGeneAnnotator	Predicts all kinds of prokaryotic genes from a single or a set of anonymous genomic sequences having a variety of lengths	[47]
CDD	Comprises protein domains conserved throughout molecular evolution	[41]
ChEBI	The CHEBI database contains manually curated data on chemical entities	[26]
GO FEAT	A web-based functional annotation tool for genomic and transcriptomic data	[5]
GO	The Gene Ontology database provides a standardized vocabulary for describing gene function	[2, 9]
KEGG	Links genomic data and higher-order functional information	[32]
Rhea	Rhea is a database that contains functional annotations for enzymes and descriptions of metabolic pathways. The annotations in Rhea are based on expert-curated, non-redundant information on biochemical reactions	[43]
InterPro	InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites	[53]
NCBI Blast	BLAST is a program that identifies regions of similarity between sequences	[30]
SEED	The SEED database provides genome annotations across thousands of genomes	[52]
UniProt	A freely accessible resource on protein sequences and functional annotation	[11]
RefSeq	RefSeq provides a well-annotated set of sequences, including genomic DNA, transcripts, and proteins	[48]
FunMappOne	FunMappOne is a tool for functional analysis and visualization of gene lists	[60]
g:Profiler	Web server for functional enrichment analysis and additional information mining	[37]
GAEV	A tool that was developed for the construction of complete sets of molecular pathways for non-model organisms using KEGG gene function annotations	[73]
GOPlot	GOPlot is an R package for the functional analysis of gene lists	[69]

Hamese *et al. Journal of Genetic Engineering and Biotechnology*   (2023) 21:156

Page 4 of 9

Metabolic modeling

The development of microbial chassis, mainly focusing on LAB (lactic acid bacteria), is significantly propelled by genome-scale metabolic (GSM) models and system biology methodologies. GSM models employ constraints-based modeling, a widely adopted computational method, to map the metabolic pathways and predict phenotypic behavior. Initially applied in the food industry to enhance target product production, GSM models have expanded their utility to system-wide therapeutic targeting for infectious microorganisms and malignancies [3, 15]. Recent advancements, exemplified by creating the iCN1361 GSM model for *Cupriavidus necator H16*, demonstrate the integration of omics data and network visualization to improve model applications [54]. Evaluating how well GSM models predict metabolic phenotypes involves contrasting model results with experimental data and subjecting models to in silico simulations under various growth conditions [42]. These GSM models are crucial in understanding a microbial chassis's metabolic capabilities, predicting metabolic fluxes, and providing insights into resource allocations and adaptation to changing conditions [59]. Moreover, in genome reduction efforts, the models may serve as input alongside essentiality and gene location data [70]. Finally, Fig. 1 illustrates the model-guided approach for designing microbial chassis integrated into the synthetic biology Design-Build-Test-Learn (DBTL) cycle. This approach utilizes metabolic models and a minimal synthetic genome to develop a microbial chassis.

Lactic acid bacteria (LAB) as a chosen chassis

Lactic acid bacteria (LAB) have been investigated for their potential use in vaccine development due to their ability to induce a strong immune response. For example, *Lactococcus lactis*, has been modified to deliver antigens and stimulate an immune response in animal models. A recent study explored the expression and secretion of human interleukin-22 (hIL-22) by *Lactobacillus reuteri (L. reuteri)*. The results showed that hIL-22 expression and secretion resulted in a growth defect in *L. reuteri* and cleavage of most of the secreted hIL-22, although the reason for this is unclear. The study found that changing the signal peptide improved hIL-22 secretion and showed promise for the active hIL-22 on the human intestinal epithelium in vivo, as it was able to stimulate the production of the antimicrobial peptide Reg3α in human intestinal enteroids. LAB have the potential as a vaccine delivery vehicle due to their ability to induce a strong immune response [50]. Synthetic biology tools can be utilized to enhance the properties of LAB for vaccine use, but challenges such as antigen stability and elicitation of an unwarranted immune response must be addressed. The recent study of hIL-22 expression and secretion by *L. reuteri* showed promising results, but further research is needed to fully understand the implications and potential limitations.
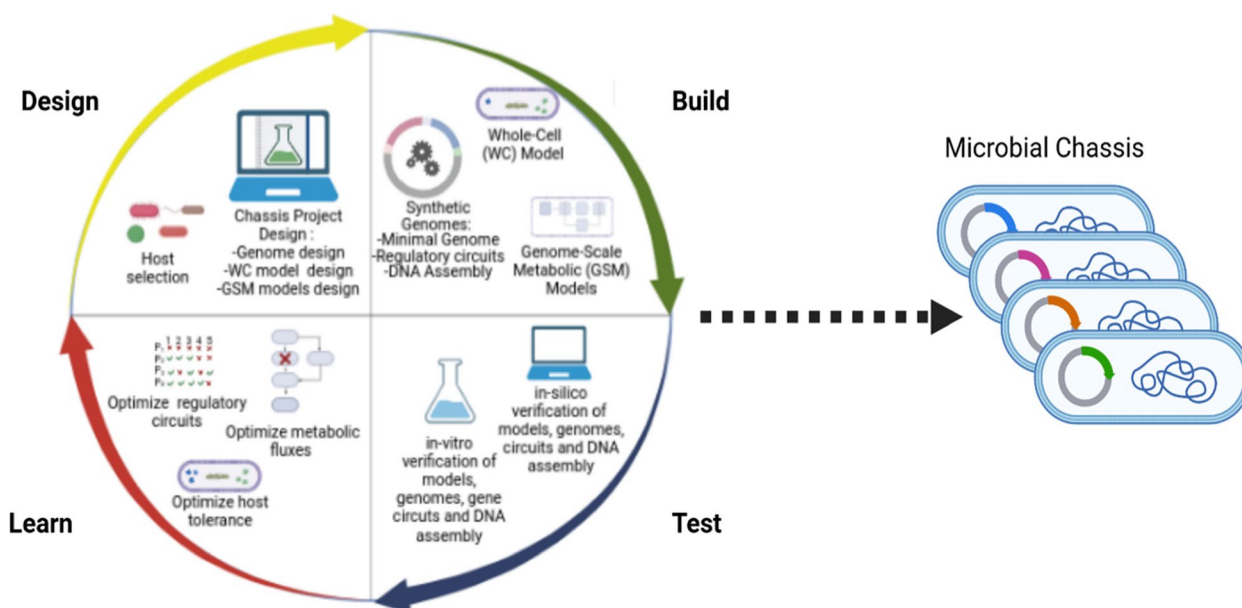

Fig. 1 Illustration of the model-guided approach for designing microbial chassis integrated into the synthetic biology design-build-test-learn (DBTL) cycle. This approach requires and utilizes metabolic models, and a minimal synthetic genome to develop a microbial chassis. Illustration created with BioRender

Workflow for the design to reduce microbial genome as a chassis

Step 1: Choosing lactic acid bacteria (LAB) as host chassis

Lactic acid bacteria (LAB), including genera like *Bifidobacterium*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, and *Streptococcus*, play a crucial role as microbial chassis hosts. Lactic acid bacteria (LAB) are considered safe and versatile microbial chassis hosts and are widely used in ingredient production. In recent years, LAB have gained prominence as live delivery vehicles for therapeutic agents, including vaccines, cytokines, enzymes, and allergens. They possess unique attributes such as safety, non-colonizing behavior, and easy elimination from the human body, making them valuable in therapeutic applications [22]. LAB's potential in vaccine development is notable, given their ability to induce a robust immune response. Synthetic biology tools optimize LAB's ability to produce, deliver, and express antigens, enhancing their potential as vaccine vectors. However, antigen stability and immune response elicitation must be addressed [50, 57]. Their safety profile, versatility, and potential for immune response induction make them invaluable in developing therapeutic agents and vaccine delivery systems.

Step 2: Testing the fitness of Lactococcus lactis as hosts

Lactococcus *lactis* is a mesophilic, Gram-positive, non-motile, non-spore-forming, facultative anaerobe, previously *Streptococcus lactis*. It has been used for centuries in producing fermented food products, including cheese and yogurt. It is considered heterofermentative because it produces (S)-lactate as its primary fermentation product and contains genes for enzyme 6-phosphofructokinase (pfkA and pfkB). However, it can have heterofermentative metabolism due to its ability to produce diacetyl, (S)-acetoin, and acetaldehyde, as well as (S)-lactate. Such characteristics made *L. lactis* a microorganism of industrial importance. Metabolic efforts of this bacterium have also led to the production of B vitamins (folate and riboflavin), biofuels (ethanol), and therapeutics [65]. Due to its industrial importance, *L. lactis* has been categorized as GRAS (generally recognized as safe) by the Food and Drug Administration (FDA).

Step 3: Predicting gene essentiality

Gene essentiality studies are often performed to determine which genes are essential before reducing an organism's genome. Previous gene essentiality studies involved comparative genomics in search of homologs and paralogs among closely related species [46] or systemic inactivation of single individual genes [8, 36]. Experimentally or computationally determined essential gene sets may be deposited into available databases of essential genomic regions. Experimentally determined essential gene sets may be deposited into the following databases: DEG (Database of Essential Genes) 15, OGEE (Online GEne Essentiality), and EGGS (Essential Genes on Genome-Scale) whereas pDEG, NetGenes, and ePath are predicted essential gene set databases. The advantages of incorporating computational tools to predict essential genes include low cost and time efficiency. A few algorithms (a series of steps that attempt to solve a problem) have been developed to identify those regions in the genome that may be eliminated. Algorithms that have been developed to identify essential genes include DELEAT (DELetion design by Essentiality Analysis Tool) and Geptop 2.0 [64, 71]. Geptop 2.0 is simple to use, with an interface to input DNA or protein sequences and receive the predicted essentiality with probabilities of genes or proteins. However, it can only be used with fully sequenced organisms. Essential gene databases and computational programs will continue to be utilized to predict essential genes, facilitating the design of genomic deletions [6, 7, 14, 17, 32, 34, 66].

Step 4: Performing enrichment analysis

Once potential genes of interest, including gene essentiality predictions, are identified through a large-scale screening, the subsequent challenge is discerning false positives and negatives within these predictions. Integrating gene annotations with the genes of interest is vital to uncovering and evaluating enriched functions of interest. Gene set enrichment analysis is a valuable method for identifying functional classes overrepresented within sets of genes or proteins. Tools such as STRING-db [66] and FUNAGE-Pro [19] play crucial roles in annotating biological functions from gene sets generated through analyses of differential gene or protein expression. The primary data sources for these tools are the complete bacterial genomes housed in the NCBI RefSeq and Genbank databases [16]. The identified protein sequences are mapped against the reviewed and manually curated prokaryote database embedded in UniProt [11]. Functional classes like GO, KEGG, InterPro, and COG can be assigned to each protein, utilizing the UniProt protein annotation. The statistical method for the gene set enrichment analysis is "hypergeometric testing," employed to identify overrepresented class IDs [20]. This statistical test relies on four key parameters: population size (total annotated genes in the genome), population identified as successful (genes with significant differential expression), sample size (genes in a class-ID), and sample identified as successful (significant values in the class-ID). Additionally, we apply a Benjamini–Hochberg multiple-testing correction to compute the final *P* value, which facilitates the development of ranking scores for

Hamese *et al. Journal of Genetic Engineering and Biotechnology* (2023) 21:156

Page 6 of 9

visualization purposes, revealing enrichment patterns within the gene sets under investigation.

Step 5: Computational design of genome reduction

As more is learned about bacterial genomes, deciding which genes to remove and how to remove those genes becomes increasingly complex. A few computational programs have been developed to assist in the deletion selection and genome design. Moreover, there needs to be more ability to analyze and evaluate genomic designs and an overwhelming number of genome configurations, even for bacteria with small genomes. In genome minimization, two main approaches are used: the top-down approach and the bottom-up approach. The top-down approach involves deleting non-essential genomic regions from an existing genome until the reduced genome supports desired growth yield and rate [68, 70].

On the other hand, the bottom-up approach entails designing and building an artificially synthesized genome from scratch using enzymatic assembly [25],K. [35]. Moreover, Fig. 2 compares the two approaches. The top-down approach is primarily used compared to the bottom-up approach due to the cheaper cost and relative ease of the underlying procedures associated with the top-down genome reduction strategy (K. [35]. Both approaches are essential for advancing our understanding of the genetic basis of life and for developing efficient and sustainable biotechnological systems such as microbial chassis.

Step 6: Gene circuit design

The availability of gene essentiality data makes it plausible to achieve genome minimization using the bottom-up or top-down approaches. In addition to making gene essentiality predictions, MinGenome and DELEAT computer programs may further be utilized for the in silico top-down reduction of bacterial genomes, with the ability to design large genomic deletions to minimize the organism's genome [64, 70]. In chassis development, gene circuits are pivotal in controlling gene expression levels and implementing feedback mechanisms to enhance yields and optimize cell populations. The construction of genetic circuits involves assembling well-characterized biological parts essential for achieving the desired expression levels within a cellular chassis. Fundamental biological parts used in genetic circuit
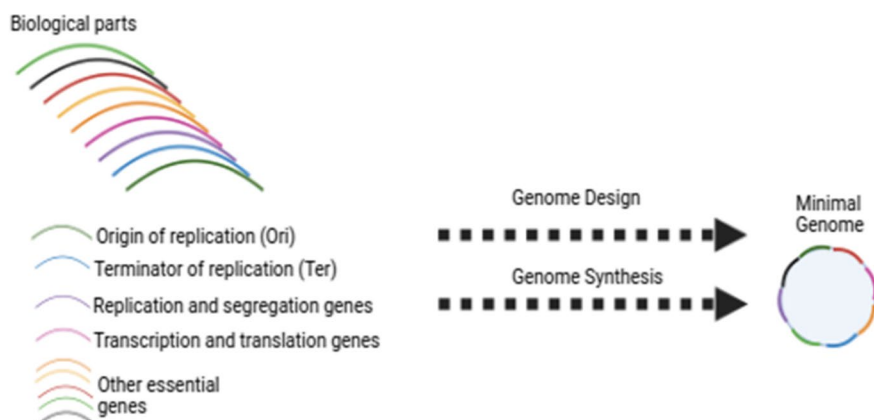

Fig. 2 Illustration of the two different genome minimization strategies. **A** The top-down genome minimization approach. DELEAT-v0.1 and MinGenome are examples of tools to design minimal genomes using the top-down strategy. **B** The bottom-up genome minimization approach, where well-characterized, reliable, and context-independent biological parts are constructed into a minimal genome

Hamese *et al. Journal of Genetic Engineering and Biotechnology* (2023) 21:156

Page 7 of 9

design include transcriptional switches, functional non-coding RNAs like riboswitches, ribozymes, and aptamers, as well as CRISPR-based genetic switches and toggle switches. Promoters, critical in controlling gene expression, can be combined and regulated to create internal logic circuits, enabling the engineering of complex microbial behaviors. Additionally, promoters can be combined with ribosome binding sites (RBS) to fine-tune gene expression levels [49]. Toggle switches, acting as memory devices, determine when the chassis will produce specific molecules, such as therapeutic compounds. Secretion tags are often added to the polypeptide chains to ensure that the therapeutic molecules produced do not harm the producing cells. CRISPR-based switches, which can repress gene expression, have been developed, although they may impact the growth of the microbial chassis [56].

Thus, gene circuit design is a crucial aspect of chassis development, leveraging well-characterized biological parts and sophisticated tools to engineer microbial behavior and optimize gene expression within a biological chassis for various applications.

Conclusions

Herein, we reviewed the critical role of computational methods in obtaining a genome-reduced bacterial strain, focusing on the versatile and safe microbial chassis hosts, lactic acid bacteria (LAB), particularly *L. lactis*. LAB, due to their safety profile, non-colonizing behavior, and ease of elimination from the human body, are versatile chassis hosts extensively utilized in ingredient production and emerging as live delivery vehicles for therapeutic agents, including vaccines. Computational tools play a pivotal role in predicting gene essentiality, aiding in the design of a streamlined genome. Machine learning techniques, particularly deep neural networks, have shown promise in predicting essential genes, which may guide downstream genome reduction strategies. Furthermore, advancements in gene circuit design and metabolic modeling significantly contribute to the engineering of microbial behavior, optimizing gene expression for diverse applications.

Authors' contributions
Saltiel Hamese wrote the paper together with the contributions from all the authors as follows Kanganwiro Mugwanda, Mutsa Takundwa, Earl Prinscloo, and Deepak B. Thimiri Govinda Raj. All authors read and approved the manuscript.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
We have secured ethics clearance from CSIR for this project.

Consent for publication
All authors have given consent for publication.

Competing interests
The authors declare that they have no competing interests.

References

1. Aksu D, Diallo MM, Şahar U, Uyaniker TA, Ozdemir G (2021) High expression of ring-hydroxylating dioxygenase genes ensure efficient degradation of p-toluate, phthalate, and terephthalate by Comamonas testosteroni strain 3a2. Arch Microbiol 203(7):4101–4112
2. Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Van Auken, K., … Westerfield, M. (2023). The gene ontology knowledgebase in 2023. GENETICS, 224(1). https://doi.org/10.1093/genetics/iyad031
3. Alper H, Jin Y-S, Moxley JF, Stephanopoulos G (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. Metab Eng 7(3):155–164
4. Ara K, Ozaki K, Nakamura K, Yamane K, Sekiguchi J, Ogasawara N (2007) Bacillus minimum genome factory: effective utilization of microbial genome information. Biotechnol Appl Biochem 46(Pt 3):169–178. https://doi.org/10.1042/BA20060111
5. Araujo FA, Barh D, Silva A, Guimarães L, Ramos RTJ (2018) GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. Sci Rep 8(1):1–4. https://doi.org/10.1038/s41598-018-20211-9
6. Aromolaran, O., Aromolaran, D., … I. I.-B. in, & 2021, undefined. (n.d.). Machine learning approach to gene essentiality prediction: a review. Academic.Oup.ComO Aromolaran, D Aromolaran, I Isewon, J OyeladeBriefings in Bioinformatics, 2021•academic.Oup.Com. Retrieved September 21, 2023, from https://academic.oup.com/bib/article-abstract/22/5/bbab128/6219158
7. Aromolaran, O., Oyelade, J., & Adebiyi, E. (2021). Performance evaluation of features for gene essentiality prediction. IOP Conference Series: Earth and Environmental Science, 655(1)
8. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2(1):8–2006. https://doi.org/10.1038/1755-1315/655/1/012019
9. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. In Nature Genetics (Vol. 25, Issue 1, pp. 25–29). https://doi.org/10.1038/75556
10. Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson,

R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., … Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics, 9. https://doi.org/10.1186/1471-2164-9-75

11. Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bye-A-Jee H, Cukura A, Denny P, Dogan T, Ebenezer TG, Fan J, Garmiri P, da Costa Gonzales LJ, Hatton-Ellis E, Hussein A, Ignatchenko A, Zhang J (2023) UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res 51(D1):D523. https://doi.org/10.1093/NAR/GKAC1052

12. Bu, Q. T., Yu, P., Wang, J., Li, Z. Y., Chen, X. A., Mao, X. M., & Li, Y. Q. (2019). Rational construction of genome-reduced and high-efficient industrial Streptomyces chassis based on multiple comparative genomic approaches. Microbial Cell Factories, 18(1). https://doi.org/10.1186/S12934-019-1055-7

13. Cantalapiedra CP, Hernandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38(12):5825–5829. https://doi.org/10.1093/MOLBEV/MSAB293

14. Cheng, J., Wu, W., Zhang, Y., Li, X., Jiang, X., Wei, G., & Tao, S. (2013). A new computational strategy for predicting essential genes. BMC Genomics, 14(1). https://doi.org/10.1186/1471-2164-14-910

15. Choi HS, Lee SY, Kim TY, Woo HM (2010) In silico identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76(10):3097–3105

16. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. Nucleic Acids Res 44(D1):D67–D72. https://doi.org/10.1093/nar/gkv1276

17. Clough E, Barrett T (2016) The Gene Expression Omnibus database. Methods Mol Biol 1418:93. https://doi.org/10.1007/978-1-4939-3578-9_5

18. Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R., & Pósfai, G. (2012). Low-mutation-rate, reduced-genome Escherichia coli: an improved host for faithful maintenance of engineered genetic constructs. Microbial Cell Factories, 11. https://doi.org/10.1186/1475-2859-11-11

19. De Jong, A., Kuipers, O. P., & Kok, J. (2022). FUNAGE-Pro: comprehensive web server for gene set enrichment analysis of prokaryotes. Nucleic Acids Research, 50. https://doi.org/10.1093/nar/gkac441

20. De Jong A, Kuipers OP, Kok J (2022) FUNAGE-Pro: comprehensive web server for gene set enrichment analysis of prokaryotes. Nucleic Acids Res 50(W1):W330–W336. https://doi.org/10.1093/NAR/GKAC441

21. Ejigu GF, Jung J (2020) Review on the computational genome annotation of sequences obtained by next-generation sequencing. Biology 9(9):295. https://doi.org/10.3390/BIOLOGY9090295

22. Fong, F. L. Y., Lam, K. Y., Lau, C. S., Ho, K. H., Kan, Y. H., Poon, M. Y., El-Nezami, H., & Sze, E. T. P. (2020). Reduction in biogenic amines in douchi fermented by probiotic bacteria. PLoS ONE, 15(3). https://doi.org/10.1371/journal.pone.0230916

23. Garcia-Morales L, Ruiz E, Gourgues G, Rideau F, Piñero-Lambea C, Lluch-Senar M, Blanchard A, Lartigue C (2020) A RAGE based strategy for the genome engineering of the human respiratory pathogen Mycoplasma pneumoniae. ACS Synth Biol 9(10):2737–2748

24. Gemayel, K., Lomsadze, A., & Borodovsky, M. (2022). MetaGeneMark-2: improved gene prediction in metagenomes. BioRxiv, 2022.07.25.500264. https://doi.org/10.1101/2022.07.25.500264

25. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6(5):343–345

26. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res 44(D1):D1214–D1219. https://doi.org/10.1093/NAR/GKV1031

27. Hirokawa Y, Kawano H, Tanaka-Masuda K, Nakamura N, Nakagawa A, Ito M, Mori H, Oshima T, Ogasawara N (2013) Genetic manipulations restored the growth fitness of reduced-genome Escherichia coli. J Biosci Bioeng 116(1):52–58. https://doi.org/10.1016/j.jbiosc.2013.01.010

28. Humann JL, Lee T, Ficklin S, Main D (2019) Structural and functional annotation of eukaryotic genomes with GenSAS. Methods Mol Biol 1962:29–51. https://doi.org/10.1007/978-1-4939-9173-0_3

29. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11(1):1–11. https://doi.org/10.1186/1471-2105-11-119/TABLES/5

30. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008) NCBI BLAST: a better web interface. Nucleic Acids Res 36(suppl_2):W5–W9. https://doi.org/10.1093/NAR/GKN201

31. Jung H, Ventura T, Sook Chung J, Kim WJ, Nam BH, Kong HJ, Kim YO, Jeon MS, Eyun SI (2020) Twelve quick steps for genome assembly and annotation in the classroom. PLoS Comput Biol 16(11):e1008325. https://doi.org/10.1371/JOURNAL.PCBI.1008325

32. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28(1):27. https://doi.org/10.1093/NAR/28.1.27

33. Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428(4):726–731. https://doi.org/10.1016/J.JMB.2015.11.006

34. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, Ong WK, Paley SM, Subhraveti P (2019) The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform 20(4):1085. https://doi.org/10.1093/BIB/BBX085

35. Kim, K., Choe, D., Lee, D.-H., & Cho, B.-K. (2020). Engineering biology to construct microbial chassis for the production of difficult-to-express proteins. International Journal of Molecular Sciences, 21(3). https://doi.org/10.3390/ijms21030990

36. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P (2003) Essential Bacillus subtilis genes. Proc Natl Acad Sci 100(8):4678–4683

37. Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H (2023) g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res 51(W1):W207–W212. https://doi.org/10.1093/NAR/GKAD347

38. LeBlanc, N., & Charles, T. C. (2022). Bacterial genome reductions: tools, applications, and challenges. Frontiers in Genome Editing, 4. https://doi.org/10.3389/FGEED.2022.957289

39. Leprince A, de Lorenzo V, Völler P, van Passel MWJ, Martins dos Santos VAP (2012) Random and cyclical deletion of large DNA segments in the genome of Pseudomonas putida. Environ Microbiol 14(6):1444–1453

40. Lieder S, Nikel PI, de Lorenzo V, Takors R (2015) Genome reduction boosts heterologous gene expression in Pseudomonas putida. Microb Cell Fact 14(1):1–14

41. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer R, He C, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res 43(Database issue):D222–D22. https://doi.org/10.1093/NAR/GKU1221

42. Montagud A, Navarro E, Fernandez de Cordoba P, Urchueguía JF, Patil KR (2010) Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. BMC Syst Biol 4(1):1–16

43. Morgat A, Lombardot T, Axelsen KB, Aimo L, Niknejad A, Hyka-Nouspikel N, Coudert E, Pozzato M, Pagni M, Moretti S, Rosanoff S, Onwubiko J, Bougueleret L, Xenarios I, Redaschi N, Bridge A (2017) Updates in Rhea – an expert curated resource of biochemical reactions. Nucleic Acids Res 45(D1):D415–D418. https://doi.org/10.1093/NAR/GKW990

44. Morimoto T, Kadoya R, Endo K, Tohata M, Sawada K, Liu S, Ozawa T, Kodama T, Kakeshita H, Kageyama Y (2008) Enhanced recombinant protein productivity by genome reduction in Bacillus subtilis. DNA Res 15(2):73–81

45. Murakami K, Tao E, Ito Y, Sugiyama M, Kaneko Y, Harashima S, Sumiya T, Nakamura A, Nishizawa M (2007) Large scale deletions in the Saccharomyces cerevisiae genome create strains with altered regulation of carbon metabolism. Appl Microbiol Biotechnol 75(3):589–597

46. Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci 93(19):10268–10273

47. Noguchi H, Taniguchi T, Itoh T (2008) MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes 15(6):387. https://doi.org/10.1093/DNARES/DSN027

48. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Pruitt

KD (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44(D1):D733–D745. https://doi.org/10.1093/NAR/GKV1189

49. Oesterle, S., Gerngross, D., Schmitt, S., Roberts, T. M., & Panke, S. (2017). Efficient engineering of chromosomal ribosome binding site libraries in mismatch repair proficient Escherichia coli. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-12395-3

50. Ortiz-Velez L, Goodwin A, Schaefer L, Britton RA (2020) Challenges and pitfalls in the engineering of human interleukin 22 (hIL-22) secreting Lactobacillus reuteri. Frontiers in Bioengineering and Biotechnology 8:543. https://doi.org/10.3389/FBIOE.2020.00543/FULL

51. Overbeek R, Bartels D, Vonstein V, Meyer F (2007) Annotation of bacterial and archaeal genomes: improving accuracy and consistency. Chem Rev 107(8):3431–3447. https://doi.org/10.1021/cr068308h

52. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42(Database issue):D206. https://doi.org/10.1093/NAR/GKT1226

53. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Bateman A (2023) InterPro in 2022. Nucleic Acids Research 51(D1):D418–D427. https://doi.org/10.1093/NAR/GKAC993

54. Pearcy N, Garavaglia M, Millat T, Gilbert JP, Song Y, Hartman H, Woods C, Tomi-Andrino C, Reddy Bommareddy R, Cho B-K (2022) A genome-scale metabolic model of Cupriavidus necator H16 integrated with TraDIS and transcriptomic data reveals metabolic insights for biotechnological applications. PLoS Comput Biol 18(5):e1010106

55. Qiao, W., Liu, F., Wan, X., Qiao, Y., Li, R., Wu, Z., Saris, P. E. J., Xu, H., & Qiao, M. (2022). Genomic features and construction of streamlined genome chassis of nisin z producer lactococcus lactis n8. Microorganisms, 10(1). https://doi.org/10.3390/microorganisms10010047

56. Pedrolli, D. B., Ribeiro, N. V., Squizato, P. N., de Jesus, V. N., Cozetto, D. A., Tuma, R. B., Gracindo, A., Cesar, M. B., Freire, P. J. C., da Costa, A. F. M., Lins, M. R. C. R., Correa, G. G., & Cerri, M. O. (2019). Engineering microbial living therapeutics: the synthetic biology toolbox. In Trends in Biotechnology (Vol. 37, Issue 1, pp. 100–115). Elsevier Ltd. https://doi.org/10.1016/j.tibtech.2018.09.005

57. Quintana, I., Espariz, M., Villar, S. R., González, F. B., Pacini, M. F., Cabrera, G., Bontempi, I., Prochetto, E., Stülke, J., Perez, A. R., Marcipar, I., Blancato, V., & Magni, C. (2018). Genetic engineering of Lactococcus lactis co-producing antigen and the mucosal adjuvant 3′ 5′- cyclic di adenosine monophosphate (c-di-AMP) as a design strategy to develop a mucosal vaccine prototype. Frontiers in Microbiology, 9(SEP), 2100. https://doi.org/10.3389/FMICB.2018.02100/BIBTEX

58. Ruiz-Perez CA, Conrad RE, Konstantinidis KT (2021) MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. BMC Bioinformatics 22(1):1–16. https://doi.org/10.1186/S12859-020-03940-5/FIGURES/4

59. Sarkar D, Maranas CD (2019) Engineering microbial chemical factories using metabolic models. BMC Chemical Engineering 1(1):1–11. https://doi.org/10.1186/S42480-019-0021-9

60. Scala G, Serra A, Marwah VS, Saarimäki LA, Greco D (2019) FunMappOne: a tool to hierarchically organize and visually navigate functional gene annotations in multiple experiments. BMC Bioinformatics 20(1):1–7. https://doi.org/10.1186/S12859-019-2639-2/TABLES/1

61. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069. https://doi.org/10.1093/BIOINFORMATICS/BTU153

62. Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L. 0003 3527 8101, Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Research, 48(16), 8883–8900. https://doi.org/10.1093/NAR/GKAA621

63. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res 50(W1):W216–W221. https://doi.org/10.1093/NAR/GKAC194

64. Solana J, Garrote-Sánchez E, Gil R (2021) DELEAT: gene essentiality prediction and deletion design for bacterial genome reduction. BMC Bioinformatics 22(1):1–17. https://doi.org/10.1186/S12859-021-04348-5/FIGURES/3

65. Song AAL, In LLA, Lim SHE, Rahim RA (2017) A review on Lactococcus lactis: from food to factory. Microb Cell Fact 16(1):1–15

66. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, Von Mering C (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 51(D1):D638–D646. https://doi.org/10.1093/NAR/GKAC1000

67. Tanizawa Y, Fujisawa T, Nakamura Y (2018) DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics (Oxford, England) 34(6):1037–1039

68. Vickers CE, Blank LM, Krömer JO (2010) Grand challenge commentary: chassis cells for industrial biochemical production. Nat Chem Biol 6(12):875–877

69. Walter W, Sánchez-Cabo F, Ricote M (2015) GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics 31(17):2912–2914. https://doi.org/10.1093/BIOINFORMATICS/BTV300

70. Wang L, Maranas CD (2018) MinGenome: an in silico top-down approach for the synthesis of minimized genomes. ACS Synth Biol 7(2):462–473. https://doi.org/10.1021/acssynbio.7b00296

71. Wen, Q. F., Wei, W., & Guo, F. B. (2022). Geptop 2.0: accurately select essential genes from the list of protein-coding genes in prokaryotic genomes. In Methods in Molecular Biology (Vol. 2377, pp. 423–430). Humana Press Inc. https://doi.org/10.1007/978-1-0716-1720-5_23

72. Westers H, Dorenbos R, Van Dijl JM, Kabel J, Flanagan T, Devine KM, Jude F, Séror SJ, Beekman AC, Darmon E (2003) Genome engineering reveals large dispensable regions in Bacillus subtilis. Mol Biol Evol 20(12):2076–2090

73. Xu, S., & Huynh, T. (2019). Gene Annotation Easy Viewer (GAEV): integrating KEGG's gene function annotations and associated molecular pathways. F1000Research, 7.

Publisher's Note