



## OPEN ACCESS

## EDITED BY

David W. Ussery,  
University of Arkansas for Medical Sciences,  
United States

## REVIEWED BY

Lun Cui,  
Changzhou University, China

## \*CORRESPONDENCE

Ernesto Borrayo  
✉ ernesto.borrayo@academicos.udg.mx

RECEIVED 14 December 2023

ACCEPTED 02 February 2024

PUBLISHED 27 February 2024

## CITATION

Gómez-Márquez C, Morales JA,  
Romero-Gutiérrez T, Paredes O and  
Borrayo E (2024) Decoding semiotic minimal  
genome: a non-genocentric approach.  
*Front. Microbiol.* 15:1356050.  
doi: 10.3389/fmicb.2024.1356050

## COPYRIGHT

© 2024 Gómez-Márquez, Morales,  
Romero-Gutiérrez, Paredes and Borrayo. This  
is an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Decoding semiotic minimal genome: a non-genocentric approach

Carolina Gómez-Márquez<sup>1</sup>, J. Alejandro Morales<sup>1</sup>,  
Teresa Romero-Gutiérrez<sup>1,2</sup>, Omar Paredes<sup>1</sup> and  
Ernesto Borrayo<sup>1\*</sup>

<sup>1</sup>Biodigital Innovation Lab, Translational Bioengineering Department, Exact Sciences and Engineering University Center, Universidad de Guadalajara, Guadalajara, Mexico, <sup>2</sup>Technological Innovation Department, Tlajomulco University Center, Universidad de Guadalajara, Guadalajara, Mexico

The search for the minimum information required for an organism to sustain a cellular system network has rendered both the identification of a fixed number of known genes and those genes whose function remains to be identified. The approaches used in such search generally focus their analysis on coding genomic regions, based on the genome to proteic-product perspective. Such approaches leave other fundamental processes aside, mainly those that include higher-level information management. To cope with this limitation, a non-genocentric approach based on genomic sequence analysis using language processing tools and gene ontology may prove an effective strategy for the identification of those fundamental genomic elements for life autonomy. Additionally, this approach will provide us with an integrative analysis of the information value present in all genomic elements, regardless of their coding status.

## KEYWORDS

minimum genome, non-genocentric, biological information flow, genomic semiotics, language processing tools

## 1 Introduction

Conceptually, *minimal genome* refers to the collection of the genomic information that encodes the fundamental semantic management required to sustain life, whose functional unit, according to general perception, is the cell.

Cell systems are networks of information exchange between genome and metabolism in complex feedback loops. These cell systems support several features that ensure proper cell persistence—e.g., redundancies, convergence, divergence, and other information-handling mechanisms— in different environments, giving the organism robustness and adaptability (Kurasawa et al., 2020). Such information exchange has so far been directly correlated with genome length and complexity (Andrews et al., 2023), which has narrowed the perception of genes as the mechanistic template for molecular –RNA and protein– products and has limited further research to the association between gene quantity and organism complexity.

In their search for the informational network essential core, scientists commenced under the assumption that the minimum number of functions would be equivalent to the minimum number of genes a cell required to sustain life. This genocentric paradigm initially allowed researchers to focus on most of the cellular functions. Nevertheless, this approach has become limited as it considers whole genes as single functional units and disregards mechanisms for gene-expression triggering and other putative functions present in non-translated regions of the genome (Baby et al., 2018; Wang et al., 2018).

Non-translated regions may have a meaningful role in information management. Some have already been described as critical regulatory elements (Andrews et al., 2023) and should, therefore, be included in minimal genome research, which has proved effective in understanding various biological mechanisms, such as the putative bases for biological autonomy, metabolism, adaptation, and evolution. It has also become an important starting point for bioengineering applications –bioproduct synthesis, for example– holding immediate applications in medicine, industry, and environmental sanitation (Lee and Kim, 2020).

In this study, we discuss an alternative perspective to the genocentric paradigm, focusing on DNA as genomic quanta rather than on genes. Therefore, we propose that the minimal genome be the set with the least genomic quanta—regardless of its functional nature—required for an organism to thrive.

## 2 Minimal genome

### 2.1 Minimum biological information flow

The search for the minimal genome became more prominent after JCVI-syn3.0 (Hutchison et al., 2016). Hutchinson, et al. achieved reproduction with a significantly smaller genome than any other autonomous organism. As a result, they obtained the minimum number of genes required for cell reproduction, which included known genes and many others whose functions were unknown. The JCVI-syn3.0 model and other efforts performed on natural organisms (Gil and Peret, 2015; Baby et al., 2018) with only essential functions that confer reproductive capability in highly controlled conditions (Baby et al., 2018; Wang et al., 2018) are focused only on the genes involved. Given this scenario, differential gene expression tends to be overlooked, and, consequently, expression regulation elements are usually ignored in minimal genome research. Beyond coding regions, few studies have expanded the minimal genome analysis to non-coding sequences, encompassing life-sustaining essential genomic motifs (Luo et al., 2014; Bartha et al., 2018). These regions include non-translated RNAs as promoters, enhancers, and replication origins (Luo et al., 2014).

Therefore, minimal life-sustaining processes should not only be considered as the genetic expression *per se* but also *when*, *where*, and *which* genes are to be expressed at any given cellular scenario. It is not likely that genes –as information packages alone– could explain such phenomena. Although survival and reproduction are the main success attributes in terms of minimal genome research, a minimal genome should also consider genomic requirements for cell autonomy, reproduction, adaptation, and evolution. Such processes involve complex interactions among genomic products, sequences, arrangements, and organization, all of which determine *information management*. Furthermore, the challenge of genome minimization resides in furnishing the organism with the “genomic chasis” that allows for its survival in as wide a range of environmental scenarios as possible, enhancing its odds for adaptation and evolution (Moger-Reischer et al., 2023).

### 2.2 The “essentiality” trait

Determining which genes are essential for cellular autonomy and reproduction departed from an intuitive perception of sequential genomic reduction with different analytical approaches [i.e., comparative genomic analyses, gene inactivation studies, and progressive genome reduction (Hutchison et al., 2016; Baby et al., 2018; Dong et al., 2018; Breuer et al., 2019)] that proved less straightforward than initially expected: Some genes that were dispensable when others were present became indispensable when the latter were amiss (Baby et al., 2018; Leeuwen et al., 2020). A further discussion over this has classified gene essentiality into four categories: *no essentiality*, *low essentiality*, *high essentiality*, and *complete essentiality*; in accordance with their indispensability for cellular autonomy in different contexts, such as metabolite disposition on media, protein complex redundancy, and alternative metabolic pathways (Rees-Garbutt et al., 2020). Therefore, cellular context highly determines *gene essentiality*, mainly provided by environmental conditions. To adequately categorize genes into such classification, their functionality regarding cellular autonomy should be considered. The analysis of an obligate parasite or a synthetic organism that has specific environmental condition requirements for its autonomy establishes the challenge to determine where to draw the line to consider the organism as *autonomous* (Wang et al., 2018; Breuer et al., 2019). This classification should consider its biological state: whether its single-celled or multi-cellular, and what its environmental and habitat conditions are. Each of these biological states will present different complexity levels in terms of adaptability, *quorum* sensing, and specialization, involving not only genetic expression regulation but also complex information management (Bartha et al., 2018; Rosconi et al., 2022). We consider important to highlight that minimum genome research has been mainly performed on obligate parasites (Gil and Peret, 2015; Baby et al., 2018) and/or in controlled laboratory conditions (Baby et al., 2018; Wang et al., 2018) which do not necessarily represent their natural biological environment and definitively do not involve complex information management.

However, biological information flow analysis should go beyond the binary functionality described by Hutchison et al. (2016) in other biological processes –genome preservation, gene expression, membrane function/structure, and cytosolic metabolism– and be analyzed from a continuous perspective. A well-described example of such continuity is the ribosome competition in regulation control, which implies the importance of a more profound understanding of generally unacquainted information management elements such as riboswitches (Brewer et al., 2021) and other putative non-coding regions.

At a higher level of autonomy, the concept of continuity expands further to address the understanding of adaptation adequately. This approach should include membrane receptors, genetic regulation elements, and related mechanisms that provide energetic optimization. Here, different metabolic pathways are regulated according to a dynamic environment, considering that the cell manages a dynamic information flow within itself (Malerba et al., 2020). We can highlight two evolutionary mechanisms that have been described to expand metabolic capabilities: on the one hand, genetic mutation driven by genetic expansion and selective

pressure loss, which provides the opportunity for functionality gain and protein specialization (Noack and Baumgart, 2019; Malerba et al., 2020); on the other hand, genomic reduction, which renders protein multifunctionality (Gil and Peret, 2015). These two different evolutionary processes address information management with different solutions, all of which endorses the concept that information management is an important point of attention in addition to the conventional genomic product analysis.

Following this, the next level would be how a cell interacts with other living organisms present in the environment, either by *Quorum sensing* (Boo et al., 2021), genetic material exchange, or even newly described Borgs (Al-Shayeb et al., 2022). Here, information management acquires a new dimension. Complex interactions now establish an information flow where products from one cell influence the genetic expression of another one, establishing unicellular communities that act as a consortium (Deter and Lu, 2022).

Specialization would be the last level, where cellular compartmentalized organization is now extended to other cells to fully cooperate in a way that each cell plays a determined and committed role as members of a cellular community (Bartha et al., 2018; Kurasawa et al., 2020). At this level, information management involves significantly more complex mechanisms in terms of expression, regulation, and inactivation, all of which modifies cellular behavior (Rosconi et al., 2022), giving rise to multi-cellular organisms.

### 2.3 The non-genocentric search for minimal genome

In Paredes et al. (2018), we explored the way biological phenomena are encoded not only within the genome's coding regions but also in their interactions with non-coding ones. Therefore, we stated the possibility to evidence that the genomic syntax's behavior includes those non-coding genomic regions. In that exploration, the non-coding regulatory genomic regions were disassembled into different size *k-mers*, and the regulatory function's vocabulary diversity among different cellular lines was analyzed. We provided evidence that the lexicon complexity linked to singular vocabularies occurs at a *k-mer* length around 10 nucleotides and that it is at this length where DNA sequences have the degeneracy degree and potential to encode regulatory functions.

In further exploration of genomic lexicon features, we broadened our analysis to all genomic regions, regardless of their coding or non-coding status in bacterial genomes. We screened the lexicons of three bacterial families at a fixed *k-mer* size to evaluate whether genomic organization among bacterial families diverged (Borrayo et al., 2020). We applied topic modeling methods to obtain representative *k-mer* sets for each family and discussed how such divergence in genomic organization could be related to their phenotypical traits.

When COVID strike presented an opportunity to explore not only family lexicon divergence but also mutational phenomena, we replicated the previous approach to hCoV-19 virus sequences (Aguilar-Valdez et al., 2021). We confirmed that *k-mer* sets—or informational compartments—can be linked to particular taxa. We

provided evidence that *k-mer* profiles and their organization—*i.e.*, pragmatics—closely reflected phenotypical traits or biological semantics, since a portion of *k-mer* in the virus did not follow expected behavior in a vocabulary arrangement, which in turn corresponded with coronavirus mutational hotspots.

## 3 Discussion

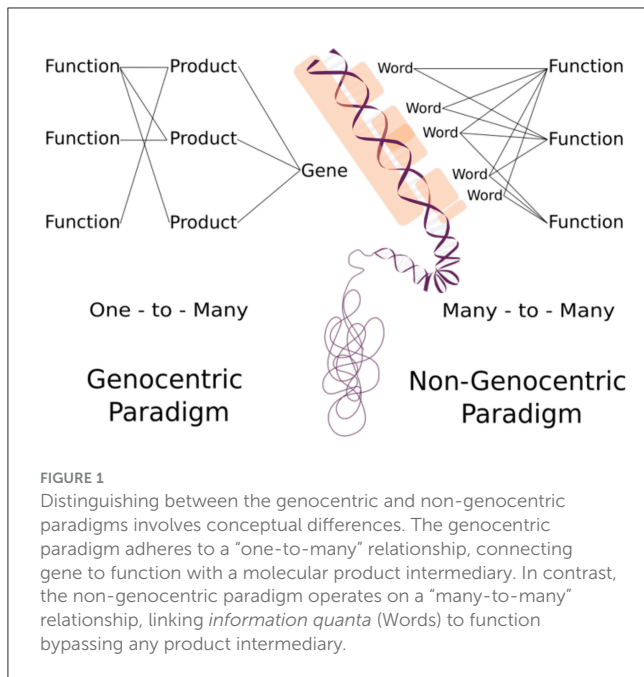
These results led our group to propose a plausible approach to minimum genome identification from a non-genocentric context through determining the minimum *corpora*—the information-quanta aggregate—to suffice the basic biological needs to be considered autonomous.

Under this perspective, the main challenge lies on identifying essential elements on the complete genome, regardless of their coding or non-coding traditional classification status. This new paradigm allows for both Top-bottom and Bottom-top approximations to a minimum genome (Wang et al., 2018; Rees-Garbutt et al., 2020) can benefit from a shift in focus on how information flow is perceived. Instead of a *gene-to-functional-product* perspective, artificial intelligence tools can be implemented using an *information-based genomic-element analysis* approach to support both types of laboratory procedures.

A non-genocentric approach provides both an integrative analysis of the informational value rendered by the genomic elements involved, and the implementation of information-analysis tools, which have traditionally been applied to other areas, such as natural language processing techniques (Gonzlez-Blas et al., 2019; Kherwa and Bansal, 2020). As genomic information has substantial similarities to natural languages, biological functions can be considered as semantic elements given by *genomic constructions* and not only by *genes*. In turn, these semantic items are driven by an analyzable syntax and vary pragmatically as some genomic constructs can be preferred under determined conditions or be preserved as a remnant of an evolutionary process (Caetano-Anolls, 2021). Therefore, a plausible strategy to address the minimum genome from an information management approach under a non-genocentric paradigm is an adequate use of language processing tools, which have been successfully applied to genomic sequence analysis in different contexts (Gonzlez-Blas et al., 2019; Kherwa and Bansal, 2020).

This strategic use of tools aligns with the broader goals of understanding genomic complexity and optimizing analytical procedures. By implementing these tools, it is possible not only to classify, align, or cluster sequences but also to extract and analyze large amounts of sequence data. The information obtained may in turn establish relationships between mentioned sequences and their related cellular processes, which, along with gene ontology (Carbon et al., 2021; Good et al., 2021), may prove effective in identifying the *corpora* that provide the minimum elements to sustain organism functions at a given autonomy level.

Altogether, genomic elements can be analyzed from a comprehensive point of view, considering their ontological function as their primary attribute and, therefore, disregarding their coding or non-coding status. This consideration could lead to a deeper understanding of the elemental genomic sequences for adequate biological function and regulation. Furthermore, such



analysis would render a “minimum genome” with the necessary *corpora* for performing autonomous biological functions in a given environment.

As stated, understanding the genomic minimum requirement for cell autonomy, reproduction, adaptation, and evolution is a key element for applied biotechnology, gene manipulation, protein design, and medical procedures. In addition, this understanding opens new perspectives regarding how life works, how it came to be, and how it copes with information management (Bartha et al., 2018; Hausser et al., 2019; Kurasawa et al., 2020).

The non-genocentric approach is not intended to substitute the current paradigm in terms of disregarding DNA molecules as the mechanistic template for molecular products. Instead, this approach aims to focus attention on which genomic regions are related to cellular functions. As we shift focus from products to *k-mers*, we are expanding our perception regarding the sequences' role in a particular biological process (Figure 1). The association of functionality between *information quanta* and a specific process potentially expands our understanding and manipulation of genomic information management. This expansion reaches a point where putative functions can be established even if they have not been found in living organisms.

The identification of the particular role of non-coding regions in biological information flow remains as the main challenge to achieve these goals. Potentially, *corpora*-based genomic reduction will establish the core mechanisms in cell systems and provide the scaffold for genetic circuit design implementation, as stated by McBride et al. (2021). Nevertheless, bioengineering has so far focused on genetic edition, neglecting equivalent processes on other types of genomic sequences (Kemp et al., 2020). A plausible solution to address this issue is to further explore how non-coding sequences have influenced the differential in biological complexity among different unicellular kingdoms, which would allow to

adequately assign a biological function to generated *corpora* derived from non-coding regions.

Currently, as this approach for a minimal genome develops, we suggest that it be implemented at a single cell autonomy/reproductive level, where all metabolic, regulatory, and information management processes can be supported by experimental data. We look forward to its implementation at higher autonomy levels.

The search for a minimum genome using our approach may render more than one solution, which is congruent with the different paths life has taken to adapt to specific conditions. In addition, it can provide significant insights into the information management basis, which in turn will prove useful in understanding biological functions, evolution, and subsequent applications in medical, environmental, and industrial biotechnologies. It is highly probable that guided *corpora* assembly will become a common bioengineering practice in the near future, either by means of cellular or cell-free system implementation, and through the design of information-flow solutions that have not necessarily been observed in natural systems.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

CG-M: Conceptualization, Writing—original draft, Writing—review & editing. JAM: Conceptualization, Writing—review & editing. T-RG: Conceptualization, Writing—review & editing. OP: Conceptualization, Writing—review & editing. EB: Conceptualization, Funding acquisition, Supervision, Writing—review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Consejo Nacional de Humanidades Ciencias y Tecnologías–CONAHCYT [Grant Number 320353].

## Acknowledgments

We would like to thank Gustavo Guzmán for the digitalization of Figure 1.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Aguilar-Valdez, S., Morales, J. A., and Paredes, O. (2021). "Unraveling the hcov-19 informational architecture," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (México: IEEE), 2392–2395. doi: 10.1109/EMBC46164.2021.9630954. Available online at: <https://ieeexplore.ieee.org/document/9630954>
- Al-Shayeb, B., Schoelmerich, M. C., West-Roberts, J., Valentin-Alvarado, L. E., Sachdeva, R., Mullen, S., et al. (2022). Borgs are giant genetic elements with potential to expand metabolic capacity. *Nature*. 610, 731–736. doi: 10.1038/s41586-022-05256-1
- Andrews, G., Fan, K., Pratt, H. E., Phalke, N., Consortium, Z., Karlsson, E. K., et al. (2023). Mammalian evolution of human cis-regulatory elements and transcription factor binding sites. *Science (New York, N.Y.)* 380, eabn7930. doi: 10.1126/science.abn7930
- Baby, V., Lachance, J.-C., Gagnon, J., Lucier, J.-F., Matteau, D., Knight, T., et al. (2018). Inferring the minimal genome of mesoplasma florum by comparative genomics and transposon mutagenesis. *mSystems*. 3, e00198–17. doi: 10.1128/mSystems.00198-17
- Bartha, I., Iulio, J. D., Venter, J. C., and Telenti, A. (2018). Human gene essentiality. *Nat. Rev. Genet.* 19, 51–62. doi: 10.1038/nrg.2017.75
- Boo, A., Amaro, R. L., and Stan, G. B. (2021). Quorum sensing in synthetic biology: A review. *Curr. Opin. Syst. Biol.* 28, 100378. doi: 10.1016/j.coisb.2021.100378
- Borrayo, E., May-Canche, I., Paredes, O., Morales, J. A., Romo-Vzquez, R., and Vlez-Prez, H. (2020). Whole-genome k-mer topic modeling associates bacterial families. *Genes* 11, 197. doi: 10.3390/genes11020197
- Breuer, M., Earnest, T. M., Merryman, C., Wise, K. S., Sun, L., Lynott, M. R., et al. (2019). Essential metabolism for a minimal cell. *Elife* 8, 1–75. doi: 10.7554/eLife.36842
- Brewer, K. I., Greenlee, E. B., Higgs, G., Yu, D., Arachchilage, G. M., Chen, X., et al. (2021). Comprehensive discovery of novel structured noncoding rnas in 26 bacterial genomes. *RNA Biol.* 18, 2417–2432. doi: 10.1080/15476286.2021.1917891
- Caetano-Anolls, G. (2021). The compressed vocabulary of microbial life. *Front. Microbiol.* 12, 655990. doi: 10.3389/fmicb.2021.655990
- Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., et al. (2021). The gene ontology resource: enriching a gold mine. *Nucleic Acids Res.* 49, D325–D334. doi: 10.1093/nar/gkaa1113
- Deter, H. S., and Lu, T. (2022). Engineering microbial consortia with rationally designed cellular interactions. *Curr. Opin. Biotechnol.* 76:102730. doi: 10.1016/j.copbio.2022.102730
- Dong, C., Jin, Y. T., Hua, H. L., Wen, Q. F., Luo, S., Zheng, W. X., et al. (2018). Comprehensive review of the identification of essential genes using computational methods: Focusing on feature implementation and assessment. *Brief. Bioinformatics* 21, 171–181. doi: 10.1093/bib/bby116
- Gil, R., and Peret, J. (2015). Small genomes and the difficulty to define minimal translation and metabolic machineries. *Front. Ecol. Evol.* 3, 123. doi: 10.3389/fevo.2015.00123
- Gonzalez-Blas, C. B., Minnoye, L., Papisokrati, D., Aibar, S., Hulselmans, G., Christiaens, V., et al. (2019). cistopic: cis-regulatory topic modeling on single-cell atac-seq data. *Nat. Methods* 16, 397–400. doi: 10.1038/s41592-019-0367-1
- Good, B. M., Auken, K. V., Hill, D. P., Mi, H., Carbon, S., Balhoff, J. P., et al. (2021). Reactome and the gene ontology: digital convergence of data resources. *Bioinformatics* 37, 3343–3348. doi: 10.1093/bioinformatics/btab325
- Hausser, J., Mayo, A., Keren, L., and Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. *Nat. Commun.* 10. doi: 10.1038/s41467-018-07391-8
- Hutchison, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351. doi: 10.1126/science.aad6253
- Kemp, L., Adam, L., Boehm, C. R., Breiting, R., Casagrande, R., Dando, M., et al. (2020). Bioengineering horizon scan 2020. *Elife* 9, 1–20. doi: 10.7554/eLife.54489
- Kherwa, P., and Bansal, P. (2020). Topic modeling: a comprehensive review. *EAI Endorsed Transactions on Scalable Information Systems* 7:1–16.
- Kurasawa, H., Ohno, T., Arai, R., and Aizawa, Y. (2020). A guideline and challenges toward the minimization of bacterial and eukaryotic genomes. *Curr. Opin. Syst. Biol.* 24, 127–134. doi: 10.1016/j.coisb.2020.10.012
- Lee, S. R., and Kim, P. (2020). Current status and applications of adaptive laboratory evolution in industrial microorganisms. *J. Microbiol. Biotechnol.* 30, 793–803. doi: 10.4014/jmb.2003.03072
- Leeuwen, J., Pons, C., Tan, G., Wang, J. Z., Hou, J., Weile, J., et al. (2020). Systematic analysis of bypass suppression of essential genes. *Mol. Syst. Biol.* 16, 9828. doi: 10.15252/msb.2020.9828
- Luo, H., Lin, Y., Gao, F., Zhang, C. T., and Zhang, R. (2014). Deg 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* 42, gkt1131. doi: 10.1093/nar/gkt1131
- Malerba, M. E., Ghedini, G., and Marshall, D. J. (2020). Genome size affects fitness in the eukaryotic alga *dunaliella tertiolecta*. *Curr. Biol.* 30, 3450–3456.e3. doi: 10.1016/j.cub.2020.06.033
- McBride, C. D., Grunberg, T. W., and Vecchio, D. D. (2021). Design of genetic circuits that are robust to resource competition. *Curr. Opin. Syst. Biol.* 28, 100357. doi: 10.1016/j.coisb.2021.100357
- Moger-Reischer, R. Z., Glass, J. L., Wise, K. S., Sun, L., Bittencourt, D. M. C., Lehmkühl, B. K. Jr., et al. (2023). Evolution of a minimal cell. *Nature* 37, 1–6. doi: 10.1038/s41586-023-06288-x
- Noack, S., and Baumgart, M. (2019). Communities of niche-optimized strains: Small-genome organism consortia in bioproduction. *Trends Biotechnol.* 37, 126–139. doi: 10.1016/j.tibtech.2018.07.011
- Paredes, O., May-Canche, I., and Fimmel, E. (2018). A code biology analysis of the regulatory regions in cell lines. *Revista Mexicana de Ingeniería Biomedica* 40, 1–18. doi: 10.17488/rmb.40.1.11
- Rees-Garbutt, J., Chalkley, O., Landon, S., Purcell, O., Marucci, L., and Grierson, C. (2020). Designing minimal genomes using whole-cell models. *Nat. Commun.* 11, 836. doi: 10.1038/s41467-020-14545-0
- Rosconi, F., Rudmann, E., Li, J., Surujon, D., Anthony, J., Frank, M., et al. (2022). A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat. Microbiol.* 7, 1580–1592. doi: 10.1038/s41564-022-01208-7
- Wang, L., Jiang, S., Chen, C., He, W., Wu, X., Wang, F., et al. (2018). Synthetische genomik: von der dna-synthese zu designer-genomen. *Angewandte Chemie* 130, 1764–1773. doi: 10.1002/ange.201708741