HealthFerm.eu

HealthFerm

Innovative pulse and cereal-based food fermentations for human health and sustainable diets

From SNVs to function: integrating genomic variation and regulation for the selection of optimal starter cultures in food fermentations

Gabriela Samaniego Vinachi

August 2026

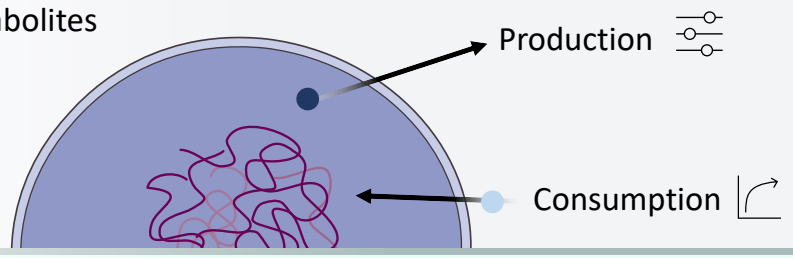
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Co-funded by the European Union

VUB INDUSTRIAL MICROBIOLOGY & FOOD BIOTECHNOLOGY RESEARCH GROUP

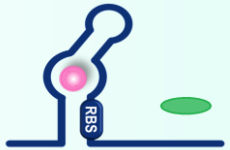
HealthFerm

Phenotype



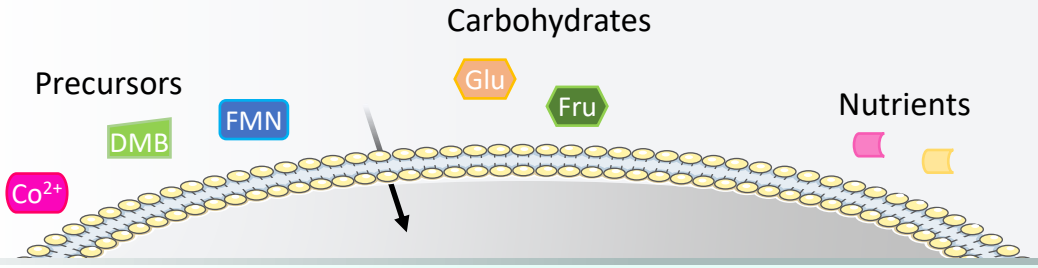
Regulation

Riboswitch specificity

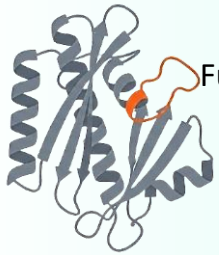


Carbon catabolite repression
CCR

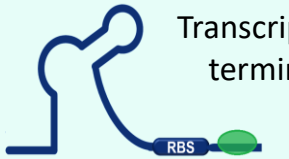
Environmental context



Molecular interpretation



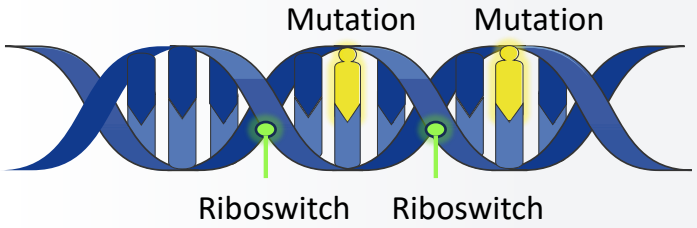
Functional domain



Transcription terminator

Functional domain
RNA transcription

Genomic variation



Copy number
Intergenic variant
Genic variant

Cellular heterogeneity

Stochastic noise

VUB INDUSTRIAL MICROBIOLOGY & FOOD BIOTECHNOLOGY RESEARCH GROUP

VLAAMS SUPERCOMPUTER CENTRUM Vlaanderen is supercomputing

HealthFerm

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Co-funded by the European Union

Integrating genomic variation and regulation for the selection of optimal starter cultures in food fermentations

Gabriela **Samaniego** and Stefan **Weckx**

Research Group of Transdisciplinary Agrifood Studies, Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium

Introduction

While high sequence identity (>99% ANI) suggests that microbial strains of the same species are very similar, variations across coding and non-coding sequences, regulatory elements, and different substrate availability can generate distinct phenotypic outcomes. This integrated multi-layer study focused on:

(i) contextualizing genomic variants (SNVs) beyond genomic annotations to assess functional impact
(ii) understanding the interplay between substrate availability and regulatory networks
(iii) assessing multi-factorial mechanisms (structural, environmental, regulatory, and stochastic) interactions that shape the final technological phenotype in microbial strains

Methods

1. Microbial isolates (n= 131)
- Sourdough (82 strains, different flour types)
- Other fermented foods (49 strains)

2. Obtention of genomes
- Whole genome sequencing (WGS)
- Genome assembly
- Genome annotation

3. Impact of coding variants (target genes)
- Variant calling
- Functional impact prediction
- Prediction of protein transmembrane positions, functional domains and binding sites

4. Impact of non-coding variants
- Riboswitches and regulatory regions prediction
- Sequence alignments
- Prediction of transcription terminators and effects on RNA secondary structures

Results

Final observed substrate consumption and metabolite production result from multi-layer integration.

Riboflavin overproduction in *Staphylococcus shinii* roseoflavin-adapted strains was explained by a non-coding regulatory deletion plus a coding deletion in the *ribU* transporter (reducing toxic analogue uptake).

Regulatory networks control metabolic flux and substrate utilization

Carbon Catabolite Repression (CCR) fine-tuned carbohydrate preferences and consumption rates in response to media composition, driving distinct substrate utilization profiles when comparing control media against pulse-based matrices across 76 microbial strains (58 bacteria, 18 yeast)
Riboswitch control: specificity and abundance of cobalamin (RF00174) and adenosylcobalamin (RF01482) riboswitches dictated final vitamin B12 production levels in *Propionibacterium freudenreichii* strains.

Genetics defines the potential of a strain, but substrate and precursor availability trigger specific pathway outcomes.

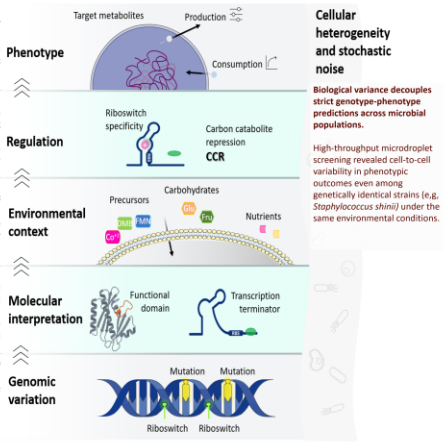
Vitamin B12 synthesis depended on the availability of precursors, as external cobalt (Co^{2+}) and the ligand DMB. Precursor availability determined whether cells synthesized canonical B12 or alternative corrinoids.

Variations must be evaluated within their protein-domain position or RNA secondary structure.

Coding mutation: Predicted "high-impact" stop codons occurring outside of protein functional domains left the metabolic capacity fully intact, as seen in a subset of microbial strains across a total set of 76 strains (58 bacteria, 18 yeast).
Non-Coding variation: A regulatory deletion in *Staphylococcus shinii* altered a transcription terminator without mutating the riboswitch of the *rib* operon, linking riboflavin hyperproduction in strains carrying this mutation.

Divergence driven by SNVs, indels, genic/intergenic variants, and copy-number variations (CNVs) of genes and regulatory elements.

Gene CNVs: Increased copy numbers of α-acetolactate decarboxylase (*aldC*) in *Companilactobacillus paralimentarius* strains marked potential high-acetoin producers.
Regulatory CNVs: Strains of *Propionibacterium freudenreichii* with two cobalamin riboswitches (vs. one) detected precursors faster, truncating pathway expression earlier.



Cellular heterogeneity and stochastic noise

Biological variance decouples strict genotype-phenotype predictions across microbial populations.

High-throughput microdroplet screening revealed cell-to-cell variability in phenotypic outcomes even among genetically identical strains (e.g., *Staphylococcus shinii*) under the same environmental conditions.

Figure 1. Multi-layered biological framework leading to phenotype changes in microbial strains for food fermentation. The final metabolite production and consumption is shaped by five interconnected layers: sequence-level diversity, molecular context, environmental inputs, and regulatory processing. Cellular heterogeneity and stochastic noise act across all layers.

Conclusion

The targeted selection of optimal microbial strains for food fermentations requires more than knowing the presence of a gene, a genomic variant, or a single phenotype. Phenotypic outcomes depend on multiple interacting layers: genomic variation, including SNVs, indels, and copy-number variation; the structural and functional context of these variants; regulatory mechanisms such as carbon catabolite repression and riboswitches; environmental context and precursor availability; and, ultimately, stochastic cellular heterogeneity. By this approach, it is possible to understand not only which strain performs best, but why it performs best under optimal matrix and environmental conditions, and whether that performance can be predicted and reproduced.

Acknowledgments:
Co-funded by the European Union
Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra
This work was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under Grant number 22.00159

www.healthferm.eu

Coordinator contacts:
Christophe Courtin (christophe.courtin@kuleuven.be)
Yamina De Bondt (yamina.debondt@kuleuven.be)