Leveraging Lactic Acid Bacteria for new sustainable processes

Laura Troiani (laura.troiani@unipr.it)

Dept. of Food and Drug, University of Parma, 43124, Italy

Tutor: Prof. Erasmo Neviani

This Ph.D. thesis dealt with the phenotypic characterization of a microbial core (SMC) of Lactic Acid Bacteria (LAB) selected from the “University of Parma Culture Collection”. The investigation of LAB functional potentialities aims to valorize the collection biodiversity and develop targeted fermentation processes for agri-food waste recovery. Results showed the aptitude of LAB to metabolize different compounds characterizing the composition of many agri-food waste and by-products, highlighting meanwhile their capability of growing far from the optimal conditions. Phenotyping microarrays represent a powerful approach to implement the genotypic knowledge on bacterial physiology and to move towards novel industrial applications of LAB.

Caratterizzazione funzionale di una collezione di batteri lattici per lo sviluppo di processi industriali sostenibili

Questa tesi di dottorato ha riguardato la caratterizzazione fenotipica di un core microbico (SMC) di batteri lattici (LAB) selezionati dalla collezione microbica di ateneo. Lo studio delle potenzialità funzionali dei LAB mira alla valorizzazione della biodiversità della collezione e allo sviluppo di processi specifici per il recupero di scarti agro-industriali. I risultati hanno mostrato l’attitudine dei LAB a metabolizzare sostanze caratterizzanti numerosi scarti agro-alimentari, evidenziando contemporaneamente la loro capacità di crescere in condizioni di stress. L’approccio fenotipico rappresenta un potente mezzo per implementare le conoscenze genotipiche sulla fisiologia batterica e per indirizzare la ricerca verso nuove applicazioni industriali dei LAB.

**Keywords**: phenotypic microarrays; biodiversity; lactic acid bacteria; waste recovery.

# 1. Introduction

In accordance with part of the Ph.D. dissertation project previously described (Troiani, 2021), this oral communication reports the core results of the following activities directed to:

A1) define the metabolic profile of the SMC’s strains;

A2) study the microbial flexibility of growing under stressful environmental conditions;

A3) investigate the technological potentialities of strains concerning the bioplastic industry;

A4) predict the microbial behaviour for targeted fermentation processes.

# 2. Exploitation of microbial biodiversity

Microorganisms represent the greatest biodiversity in every ecosystem. The University of Parma owns a wide microbial collection, denominated “University of Parma Culture Collection” (UPCC), mainly composed of Lactic Acid Bacteria (LAB) isolated from diverse food matrices. LAB are widespread in nature, especially in food and human, and they are employed in many industrial applications (Buron-Moles *et al.*, 2019). Over the time, they evolved the ability to metabolize several carbon sources and tolerate various environmental conditions, thus becoming able to colonize many different habitats.

Current technologies in genome sequencing, bioinformatics, and high-throughput screening techniques applied to large microbial strain collections give us new opportunities in natural product discovery (Steele *et al.*, 2019), such as precursors of bioplastics, to move towards a sustainable economic system (Bosco *et al.*, 2021). The implementation of a circular economy model allows to recover waste and by-products produced every day all around the world and to reduce the over-exploitation of natural resources and environment (Venkata Mohan *et al.*, 2020), by providing significant economic benefits (Socas-Rodríguez *et al.*, 2021) and promoting sustainable development (Picot-Allain *et al.*, 2020). Producing interesting molecules for the industry by biological methods is one of the strategies of the circular economy, especially in the agri-food sector where food waste and by-products can be used as alternative fermentation matrices (Hadj Saadoun *et al.*, 2022).

The term ‘phenotype’ does not only mean the set of observable traits of an organism. Phenotyping aims to explore strain’s niche-specific metabolic traits related to cell physiology and growth, such as substrate consumption, resistance to chemicals, and osmolyte tolerance or other stresses, thus establishing the potential adaptation under specific environmental conditions. Changes in phenotype depend on the interweaving interactions between environmental pressure and genotype (Acin-Albiac *et al.*, 2020). Studies of comparative genomics and metagenomics have been extensively employed to indagate the potential functionality of microbes, while (meta)transcriptomics especially focused on mechanisms for niche adaptation. Metabolic traits interest the phenotype expression but, whilst data analysis methodologies and technologies for sequence-based omics quickly evolved, the systematic study of phenotype profiling stayed somehow hampered by some limitations (Houle *et al.*, 2010). Techniques like Biolog System make possible to screen many phenotypes concurrently thus introducing phenomics as the last of the omics techniques, whose implementation still has wide margins (Acin-Albiac *et al.*, 2020).

However, experiments are expensive, time-consuming, and manually intensive to perform. Machine learning is a broad field of advanced computational and statistical methods whose models can help to describe complex data relationships and to facilitate the interpretation of large sets of data in all sectors, including microbiology. A combination of genotypic and phenotypic data provides an insight into the metabolic capabilities of industrially relevant strains, disclosing their potential application for the biotransformation of various substrates, including agro-industrial waste and by-products. Amplified Fragment Length Polymorphism (AFLP) is a well-known PCR-based technique that generates genomic fingerprints (Ramadan, 2022). This is largely shown by wet lab analyses, but no computational model is currently available to exploit its power.

# 3. Materials and Methods

The SMC consists of 150 strains of LAB of food origin belonging to the University of Parma Culture Collection (UPCC). The screening involved species belonging to 12 genera, previously isolated across a 10 years temporal range (2012-2022) from several food matrices, especially of dairy origin. Ecological diversity indices were used to describe the biodiversity of the microbial core (Table 1).

## 3.1 Genotyping

The DNA of each strain was extracted and used to perform the Amplified Fragment Lenght Polymorfism (AFLP) as reported by Bertani *et al.* (2019). Resulting electropherograms were analysed through Bionumerics Software v 8.0 (Applied Maths NV, Belgium) to have a clear fingerprint of the SMC.

## 3.2 Phenotyping

### 3.2.1 Metabolic activity on several carbon sources

Metabolic profiling of the SMC was evaluated through Biolog GENIII MicroPlates (BIOLOG Inc.©, USA) which incorporate a patented tetrazolium violet dye used as colorimetric indicator of the microbial activity on 71 different carbon sources. The final concentration of the cell suspension used for the inoculum ranged between 90-98% of transmittance. The microplates were incubated at the optimal temperature of strains for 72 h. The metabolic activity was detected by absorbance reading at 590nm at a single end-point (72 h). Raw data were referred against the negative control well and normalized by the Average Well Color Development (AWCD, average absorbance of all wells considered). Results were elaborated with Rstudio (R version 4.2.0).

Ecological diversity indices were adapted as reported in Table 1 to estimate also the metabolic functional diversity of the SMC (Daly *et al.*, 2018; Dubey *et al.*, 2022).

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1** Ecological diversity indices used to describe the biodiversity of the SMC and its metabolic variability. | | | |
| **Index** | **Formula** | **SMC biodiversity** | **Metabolic functional diversity** |
| Richness  (S) (SR) |  | Total number of species in the microbial community | Total number of substrates utilized by each strain (normalized value at 72h ≥ 1) |
| % SR |  | / | Percentage of substrates assimilated |
| Shannon Diversity Index (H’) |  | Entropy or disorder of the population  proportional abundance of species *i* where:  *ni* = number of strains of species *i*  *N* = total number of strains | Diversity of substrate utilization pattern where:  is the proportional color development of the well over the total color development of all wells of a plate, with *Rsi* = normalized OD value |
| Pielou’s Evenness Index (E) |  | Equitability of the population’s species abundance distribution | Equitability of C source assimilation across all utilized substrates |
| Simpson Dominance Index (D) |  | Probability that two strains randomly picked up from the community belong to the same species | / |
| Simpson Dominance Index (D’) |  | Weight the common species more than the rare ones | / |

### 3.2.2 Tolerance to extreme growing conditions

The aptitude of strains to grow and colonize habitats particularly different from the optimal laboratory cultivation was investigated by phenotyping microarrays. Strains were inoculated (3% v/v) in 96-well plates on their optimal medium properly modified as reported in Table 2. Microplates were incubated until 72 h in aerobic condition. Growth was indirectly measured through absorbance readings at 590 nm, at 9 different time point (0h, 2h, 4h, 6h, 8h, 16h, 24h, 48h, and 72h) by using a BIOLOG MicroReader Station (BIOLOG Inc.©, USA). Results were referred against the blank and the variability of tolerating non-physiological conditions was described by the areas under the various sections of the curve: 0-8 h, 8-16 h, 16-24 h, 24-48 h, 48-72 h. Not to lose important information at the beginning of the growth, intervals of 2 h from 0 to 8 h were evaluated. Thanks to a collaboration with the Department of Mathematical, Physical and Computer Sciences, raw data were also elaborated with Phyton to reduce the dimensionality of the dataset through t-SNE analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 2** Non-physiological conditions utilized to evaluate microbial environmental tolerance. | | | |
| **Condition** | **Temperature (°C)** | **pH** | **NaCl (%)** |
| Tmax | 50 | 6.5 | 0 |
| Tmin | 15 | 6.5 | 0 |
| pHmax | 30-37 | 9 | 0 |
| pHmin | 30-37 | 4 | 0 |
| salt | 30-37 | 6.5 | 6 |

### 3.2.3 Technological capabilities

The capability of LAB to produce key-molecules for the bioplastic industry was investigated.

Lactic acid (LA) is the building block of polylactic acids (PLAs), widely used as a sustainable alternative to oil-based plastics. From 24 h old cultures grown in optimal conditions, the two isoforms of LA were determined by enzymatic method (Megazyme®, Ireland) measuring the OD at 340nm. Based on the highest amount and purity (> 97%) of D-LA produced, two strains (UPPC-4516 and UPCC-2214) were selected. During the period abroad at the University of Natural Resources and Life Science of Vienna, the cultures were scaled in 1L bioreactors (DASGIP AG, Jülich, Germany) carrying out a batch process with 20 g/L initial glucose, followed by a fed-batch one after total glucose consumption (constant feeding rate = 2,5 ml/h of 50 % (w/v) glucose). The process was then optimized by keeping constant the pH = 6.2 and increasing the feeding rate (4 ml/h for UPCC-2214 and 6 ml/h for UPPC-4516).

The production of polyhydroxyalkanoates (PHAs), other precursors of bioplastics, was explored as suggested by Bosco *et al.* (2021) with some modifications to the protocol. Strains were firstly screened on Malt Extract Agar plates containing 0,5 μg/ml Nile Red in DSMO which binds to lipid molecules and emits fluorescence under UV light. Plates were incubated at the optimal temperature in the dark until 48 h. Strains that showed fluorescence were moved in liquid culture to boost the biomass production and extract as much lipids as possible to make their chemical characterization possible.

## 3.3 Machine learning for prediction

This part was focused on the exploitation of the fragment length distribution profile of bacterial genomes provided by the AFLP technology to obtain computational predictive models. Two different benchmark has been developed using genotypic and phenotypic information obtained from this work (141 strains) and extracted from publicly available database (in silico AFLP profiles and API-50 results from BacDive) (509 genomes). These datasets were used for training the machine learning models and classification algorithms to predict the metabolic traits encoded by bacteria according to their phylogenetic profiles were applied.

# 4. Results and Discussion

## 4.1 Biodiversity into the SMC

Indices calculated to describe the SMC’s biodiversity gave us information concerning the entropy of the population (H´), the homogeneity of the community (E), and the dominance of some species over others (D’). The higher the number of species (S = 29), the higher H’: H’ = 1.32 indicates that the selection criteria guaranteed a good level of diversity into the SMC. E = 0.90 means a comparable distribution among all species as it ranges between 0 and 1. Despite the diversity, D’ = 17.88 highlights that some species are more representative than others in terms of number of individuals. The AFLP clustering generally validated the biodiversity since it reflects the belonging to the species, as well as the metabolism characterizing microorganisms. Despite the overall evenness of clusters, the fingerprints of individual strains showed a certain level of intra-specific biodiversity and even more within the same genus, mainly due to the different source and moment of isolation. However, the distribution of metabolic profiles based on both the species and the metabolism type, was more disordered than the genotypic one indicating that the variability into a community is better represented by phenotyping than genotyping.

## 4.2 Phenotypic studies

### 4.2.1 Metabolic profiling

The majority of LAB mostly metabolized 25 of the 71 carbon sources supplied. No strain showed any activity on gelatine, p-hydroxy-phenylacetic acid, γ-amino-butryric acid, and D-lactic acid methyl-ester. All other substrates were used by less than 30 strains over 150 tested. Ecological diversity indices were used also to express the variability among metabolic profiles. Phenotype of homofermentative bacteria (HMF, 62 strains), facultatively heterofermentative bacteria (FHTF, 51 strains), and obligately heterofermentative bacteria (OHTF, 37 strains) was evaluated separately (Figure 1). HMF showed a wider phenotypic variability than FHTF and OHTF as they belong to genera phylogenetically more divergent compared to the latter. Within HMF, it is remarked a metabolic heterogeneity inside the *Streptococcus* genus and within the species *L. delbruecki* and *L. helveticus.* Metabolic profiles of FHTF were homogeneous within each species except for *L. curvatus* whose strains showed different substrates assimilation capability, while, among OHTF, strains of *L. fermentum* showed different metabolic patterns. Among the favourite C sources of the screening there are monosaccharides, disaccharides, sugar derivatives and a purine nucleoside. Glucose was the most frequently used substrate, however most of other substrates can be actively consumed by LAB to levels comparable with it. Differently, some compounds were only metabolized by specific strains to a much higher measure than glucose. HMF and FHTF resulted to be more flexible and adaptable than OHTF. This versatility allows a wider spectrum of applications, while metabolically limited microorganisms can be involved in adaptive evolutionary studies by betting on their edges and addressing towards targeted selection processes for industrial specific uses.

**Figure 1** Heatmaps of favourite C sources of the SMC’s strains divided by type of microbial metabolism.

### 4.2.2 Stress tolerance

The areas of the trapezes under the curve based on the different time intervals were used to describe the microbial growth and they were referred against the optimal condition. Only species composed of more than 5 individuals are here discussed. The further away the growth from the optimum and the longer the lag phase, the lower the slope of the curve and the lower the y-end. In the optimal condition, *Lacticaseibacillus* spp. and *L. plantarum* generally grew better than other species in terms of biomass, while *L. fermentum* was the fastest in the first 8 h. Applying an osmotic pressure, only *L. zeae* kept on growing until 72 h, although it reached less than half growth of the optimal one. *Lactobacillus* spp., *L. brevis* and *Streptococcus* spp. did not grow with 6% NaCl. Moving away from neutrality, *Lacticaseibacillus* spp. well tolerated both pH = 9 and pH = 4, with a remarkable preference for alkaline environments. However, *Lactobacillus* spp., *L. brevis,* and *L. fermentum* did not grow at pH = 9, as well as *Leu. mesenteroides* and *Streptococcus* spp. did not tolerate acidic environments. Finally, studying the growth at extreme temperatures, *L. rhamnosus* and *L. delbrueckii* well reacted to T = 50°C, with a short lag phase, contrarily to *L. curvatus*, *L. brevis*, *Leuconostoc* spp. and *Streptococcus* spp. At high temperature, all species generally reached the death phase before 72 h. On the other hand, growth at low temperature was characterized by a very long lag phase. This condition was well tolerated by *Lacticaseibacillus* spp., *L. plantarum,* and *L. curvatus*, but not by *L. fermentum*, *Streptococcus* spp. and *Lactobacillus* spp. Summarizing, in terms of adaptability, *Lacticaseibacillus* appears the most promising genus within the screening, while *Lactobacillus* spp., *Streptococcus* spp., and *L. brevis* struggle more than others to get out of their comfort zone. The t-SNE analysis clustered strains based on their similarities, confirming the behaviours cited above and highlighting differences among individuals belonging to the same species or genus.

### 4.2.3 Lactic acid production

Over 150 strains, 72 produced L-LA (> 80% of the total LA), 18 produced D-LA and the rest produced a racemic mixture of them. *L. delbrueckii* UPCC-2214 and *Leu. Citreum* UPCC-4516 resulted the main producers of D-LA with the highest isomeric purity. In small scale tests, they produced 20.40 g and 11.69 g in 72 h, respectively. They were scaled up in bioreactors and, after optimization, *Leu. Citreum* consumed the double amount of glucose (134.66 ± 14.21 g) compared to *L. delbrueckii* (71.77 ± 4.42 g) to finally produce about the same amount of lactate (58.33 ± 4.98 and 51.93 ± 1.22, respectively). In 58 h, *L. delbruecki* reached twice the yield (g/g) of *Leu. Citreum* (0.73 ± 0.03 and 0.43 ± 0.01, respectively), although presented a similar volumetric productivity (1.17 ± 0.03 and 1.02 ± 0.02, respectively). Cultivation in bioreactor allowed to further select UPCC-2214 as the most efficient strains of the SMC for the D-lactic acid production, and to increase the productivity.

### 4.2.4 PHA production

PHAs are produced and stored as lipidic granules inside the cell when the culture medium is strongly unbalanced for the C/N ratio. From the primary screening resulted 56 lipid-producing strains. This work is promising since the production of PHAs by LAB is still explorable, thus the next step is to chemically characterize the molecules produced and eventually scale the production up.

## 4.3 Prediction

As a result, we could correctly classify a good portion of phenotypic traits with an average accuracy of 0.8 and by using 70% of the data set for the training. Regarding the BacDive benchmark, all the computational models showed an accuracy greater than 0.6 (average 0.85) for every API-50 activity. Therefore, Random-Forest was the best machine-learning model for BacDive dataset. Similar results were obtained for the other benchmark with an average accuracy of 0.72. Feature importance techniques were also applied to produce a selection of fragment lengths associated with the predictive power within a given group of genomes and/or for a specific metabolic activity. On top of this selection, the idea is to perform functional enrichment of the selected fragments by comparing their position within the genome with the coordinates of genes for which functional annotation is already available.

# 5. Conclusions and Future Perspectives

The combination of genotypic and phenotypic data provides an insight into the functionalities of industrially relevant strains, displaying their potential leverage for the recovery of food waste and by-products and the production of added-value compounds, with the advantage of reducing costs and increasing sustainability. Results of this primary screening notice that LAB can metabolize substrates not traditionally associated with the matrices of isolation. Moreover, the variability on metabolic profiles and tolerance to stressful environmental conditions can allow us to better understand niche adaptations. Finally, besides the enrichment of the UPCC database, all data obtained from this study will be useful to develop an Artificial Intelligence (AI) algorithm to predict the microbial behaviour in terms of what LAB metabolize, where and how they grow, what they produce and in which conditions. The predictive approach will allow to design ad-hoc experimentation and have a preliminary idea of the feasibility and microbial performance to avoid waste of material, work, and time in the laboratory. The application of AI to investigate the biodiversity of microorganisms represents a powerful approach to expand the existing knowledge on bacterial physiology and can set a step in developing novel industrial applications of LAB.

# 6. References

Acin-Albiac M, Filannino P, Gobbetti M, Di Cagno R (2020) Microbial high throughput phenomics: The potential of an irreplaceable omics. *CSBJ* **18**, 2290–2299.

Bertani G, Savo Sardaro ML, Neviani E, Lazzi C (2019) AFLP protocol comparison for microbial diversity fingerprinting. *J. Appl. Genet.* **60**(2), 217–223.

Bosco F, Cirrincione S, Carletto R, Marmo L, Chiesa F, Mazzoli R, Pessione E (2021) PHA Production from Cheese Whey and “Scotta”: Comparison between a Consortium and a Pure Culture of Leuconostoc mesenteroides. *Microorganisms* **9**(12), 2426.

Buron-Moles G, Chailyan A, Dolejs I, Forster J, Mikš MH (2019) Uncovering carbohydrate metabolism through a genotype-phenotype association study of 56 lactic acid bacteria genomes. *Appl. Microbiol. Biotechnol.* **103**(7), 3135–3152.

Daly AJ, Baetens JM, De Baets B (2018) Ecological Diversity: Measuring the Unmeasurable. *Mathematics* **6**(7), Article 7.

Dubey A, Malla MA, Kumar A (2022) Taxonomical and functional bacterial community profiling in disease-resistant and disease-susceptible soybean cultivars. *Braz. J. Microbiol*. https://doi.org/10.1007/s42770-022-00746-w.

Hadj Saadoun J, Levante A, Ferrillo A, Trapani F, Bernini V, Galaverna G, Neviani E, Lazzi C (2022) Exploring the Potential of Lactic Acid Fermentation for the Recovery of Exhausted Vanilla Beans. *FRONT NUTR* **9**, 858716.

Houle D, Govindaraju DR, Omholt S (2010) Phenomics: The next challenge. *Nat. Rev. Genet* **11**(12), Article 12.

Picot-Allain MCN, Ramasawmy B, Emmambux MN (2022) Extraction, Characterisation, and Application of Pectin from Tropical and Sub-Tropical Fruits: A Review. *Food Rev. Int.* **38**(3), 282–312.

Ramadan AA (2022) Bacterial typing methods from past to present: A comprehensive overview. *Gene Rep.* **29**, 101675.

Socas-Rodríguez B, Álvarez-Rivera G, Valdés A, Ibáñez E, Cifuentes A (2021) Food by-products and food wastes: Are they safe enough for their valorization? *Trends Food Sci Technol* **114**, 133–147.

Steele AD, Teijaro CN, Yang D, Shen B (2019) Leveraging a large microbial strain collection for natural product discovery. *JBC* **294**(45), 16567–16576.

Troiani L (2021) Leveraging lactic acid bacteria collection for new sustainable processes. In Proc.s of the 1st Virtual (XXV) Workshop on the *Developments in the Italian PhD Research on Food Science and Technology*, Palermo (Italy), 14-15 September, 2021, pp. 140-141.

Venkata Mohan S, Varjani S, Pant D, Sauer M, Chang JS (2020) Circular bioeconomy approaches for sustainability. *Bioresour. Technol.* **318**, 124084.