International Journal of Advanced Research and Development
ISSN: 2455-4030, Impact Factor: RJIF 5.24
www.newresearchjournal.com/advanced
Volume 1; Issue 5; May 2016; Page No. 20-23

Application of knowledge of bioinformatics for food safety and production

[1] Dr. Sukumar Roy, [2] Dr. Partha Majumder

[1] Professor & Head, Department of Biomedical Engineering, Netaji Subhas Engineering College, Garia, Kolkata, India.
[2] Biomedical Scientist & Systems Biologist, Former Principal Scientist (Helixinfosystems, Chennai), Former Head of The Department of Department of Applied Biotechnology & Bioinformatics, Sikkim Manipal University, Kolkata, India.

Abstract
In the production of fermented foods, microbes play an important role. Optimization of fermentation processes or starter culture production traditionally was a trial-and-error approach inspired by expert knowledge of the fermentation process. Current developments in high-throughput Omics technologies allow developing more rational approaches to improve fermentation processes both from the food functionality as well as from the food safety perspective. Here, the authors thematically review typical bioinformatics techniques and approaches to improve various aspects of the microbial production of fermented food products and food safety.

Keywords: Bioinformatics, Microorganisms, food technology, genomics, predictive models

Introduction

Food is an indispensable part of our daily life. Many food products undergo some form of processing before they reach the consumer, ranging from fermentation to packaging. In many of these processes, microorganisms play important roles, either in transforming the food into the desired end product (fermentation of olives, rice, bread, alcoholic beverages such as beer and wine, fermented meat, kimchee and various fermented dairy products such as cheese and yogurt) or in spoiling or contaminating the food.

The type of microorganisms used in a fermentation process greatly influences the properties of the fermented product [1]. For example, yeasts produce ethanol as the main fermentation product, whereas the main fermentation product of lactic acid bacteria is lactic acid. The food industry is very active in optimizing strain performance with respect to diversification of product properties such as flavor and texture and with respect to controlling fermentation, by using defined starter cultures to initiate the fermentation process [1].

Strain optimization is an expert-knowledge-guided process involving trial-and-error approaches that are nowadays increasingly backed up by recent high-throughput omics developments to improve fermentation processes [2]. And to assess safety of food products [3]. Bioinformatics plays an increasing role in predicting and assessing the desired and undesired effects of microorganisms on food [4]. A combination of bioinformatics with laboratory verification of selected findings is particularly powerful. In this review, we focus on bioinformatics methods that can be used to improve the microbial production of fermented food products. These include genomics-based functional predictions, the creation of genome-scale metabolic models and prediction of complex food properties, such as taste and texture, and properties of complex fermentations. All application areas (outlined in the paragraphs below) and their relation to data streams and bioinformatics are described in Figure 1.

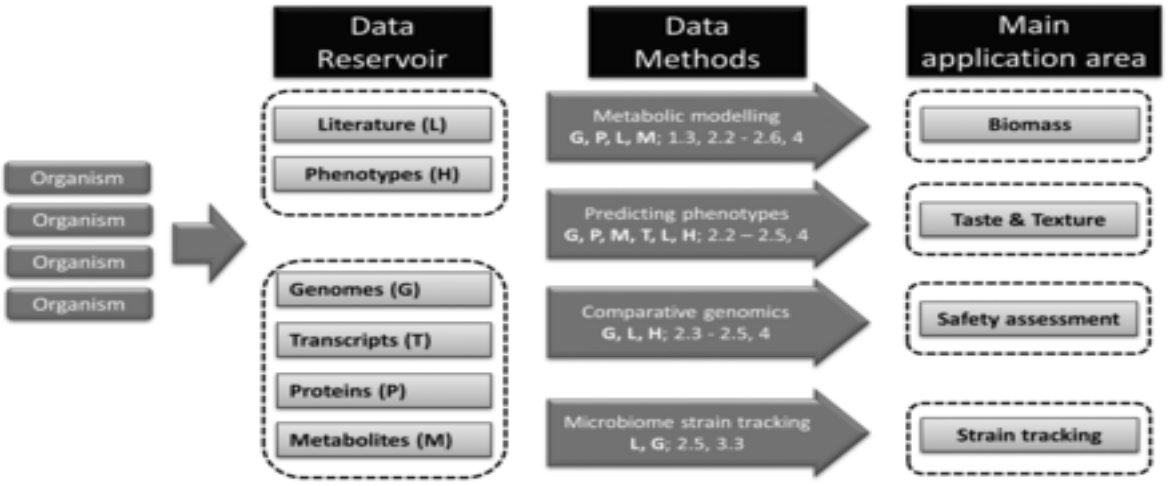

Fig 1: Data and bioinformatics applied in food application areas. Central in this figure are the food application areas (right panel). From organisms, different data sets can be obtained (data reservoir); their abbreviation is given within parentheses. Middle panel: one (of many important) methods and other methods/data sources (see Table 1 for an explanation) relevant for a main application area shown. Interpretation example: for safety assessment, genomes (G), literature (L) and phenotypes (H) are used with the gene function annotation (2.3), Orthology (2.4), comparative genomics (2.5) and predicting phenotypes (4) techniques

Translating genome information into functional predictions

The prediction of function from sequence information is one of the fundamental roles of bioinformatics. The large variety of sequencing techniques generates a large amount of genomics data. Harnessing the power of these data requires careful identification of functional elements in these data and associating the sequence information with function, for example by comparing predicted protein sequences to sequences with known functions. This type of analysis can identify functions for genes (crucial information for metabolic modelling; see below), e.g. prediction of laccases; predict functions for most genes in a bacterial genome; and suggest properties for specific strains of bacteria by projecting the predicted functions of all its genes on pathway databases, predicting properties of, e.g., Bifidobacteria in the gut environment or even predict functionalities of complex microbial communities. For genes where a sequence similarity search does not yield a good prediction, their function may be deduced by correlating the presence and absence of the gene in organisms with the presence and absence of a certain phenotypic trait in the same set of organisms (also referred to as gene–trait matching; GTM). For example, a set of proteins was predicted to be involved in the degradation of plant (oligo-) saccharides by linking isolation source of bacteria to gene presence/absence. Comparative analysis of the genome sequences of a species where some strains have a positive impact (e.g. flavour enhancement) while others are detrimental (e.g. spoilage) can be used to identify genetic elements potentially underlying these differences, as was done for the yeast *Brettanomyces bruxellensis*. Tools that can be used to link -omics data to phenotypes are PhenoLink and DuctApe. These approaches require a genome sequence, which might be relatively difficult to obtain for microbes that are difficult to grow in culture. Techniques like multiple displacement amplification can be used to amplify DNA from a single cell, and a range of genome assembly tools can be used to assemble the reads obtained from single-cell sequencing.

Mobile elements such as transposons, plasmids or phages can carry functionality from one bacterial strain to another. An example is the galactose utilization operon transfer between *Lactococcus lactis* strains studied by next-generation sequencing and bioinformatics techniques. Identifying potential transposon insertion sites is crucial to this end and can be facilitated by bioinformatics tools such as transposon insertion finder.

Improving metabolite production and biomass

Improvement of the food production process by optimizing biomass yield is a topic of continuous attention. A technique to rationally improve fermentation yield is genome-scale metabolic modelling. In this process, the genome sequence of the organism is used as an inventory of the metabolic potential of the strain of interest. Metabolic models have been made for many microbes, including several of food-relevant microorganisms. Although the quality of a genome sequence can be a limiting factor (e.g. missed gene due to low sequencing coverage), the metabolic model can be completed by identifying metabolic reactions that are missing in the model, but likely present due to the fact that they are part of metabolic reaction cascades or 'pathways'. Complete genome-scale metabolic models together with algorithms such as flux balance analysis allow the *in silico* simulation of growth of the organism under the (metabolic) restrictions provided by the substrate availability in the medium. These growth simulations can then be used to optimize medium composition to better fit the organism requirements. In addition, the models can suggest alternative or cheaper substrates for fermentation, and improve the production of compounds such as amino acids or succinic acid, taking into account possible changes in activity with respect to flavour or texture activity of the strain. These models have also been implemented in complex (multistrain) fermentation processes, providing insight in the interactions between different species/strains in a complex fermentation.

A second factor that improves the overall yield is the robustness of strains after harvesting. Also, this factor can significantly be influenced by changing fermentation conditions under which starter cultures are prepared. By correlating gene expression levels to the survival of *L. lactis*, an application of transcriptome–trait matching (TTM), a number of genes that were potentially causative related to survival were identified. Subsequent knock-out of the genes proved that these genes were indeed important for the strains' phenotype. This shows that not only gene content but also expression of genes is important for a given phenotype. In other words, preconditioning *L. lactis* strains, followed by GTM and TTM, allows improving their survival to heat and oxidative stresses, typically encountered during spray drying.

Risk assessment

Rather than predicting functions for all genes in a bacterial genome, selectively screening microbial genome sequences for genes with specific functionalities can be a highly sensitive and computationally efficient way of identifying potential health or safety risks of microbial strains present in a sample. The potential of a specific bacterium for antibiotic resistance or virulence can be investigated by comparing its genome sequence to a reference database containing known resistance genes and virulence factors. Similar approaches have been described for the identification of persistence of bacteria in food products, anaerobic spore-forming organisms in food and potential pathogens in metagenomics data. This (meta) genomics-based methodology can be extended to a wide range of functionalities, e.g. production of antimicrobial peptides and resistance to cleaning procedures commonly used in food production settings. A requirement for getting useful results out of metagenomics experiments is a dedicated database with gene–function relations and access to domain knowledge on the specific functionality to specify gene functions.

Mixed culture fermentations characterization

Complex fermentations involve an (un) defined (wild) starter culture with different microbes (bacteria, yeasts and fungi) that together ferment a substrate to the product. Examples are cheese, malolactic wine, soy and seafood fermentations. In these fermentations, strong succession of microbes can occur, for instance in wine fermentation, the microbes *Saccharomyces cerevisiae* and *Oenococcus oenii*. Similar to the above-described GTM and TTM approaches to associate (transcription of) genes to phenotypes, presence and absence of (combinations of) microorganisms (or their functionality) can be associated to fermentation product characteristics.

The first step in characterizing a fermentation is to determine what microorganisms are present at the different stages of the

fermentation and to correlate these to other measurements such as metabolomics or the presence of phages. The properties of microbial consortia are determined by the functional potential encoded in all microbial genomes. Metagenomics has an advantage over conventional sequencing of single isolates from consortia because it also reveals DNA of otherwise unculturable organisms. Based on the sequences found in a consortium, functionalities of the microbes can be predicted. Due to the succession of microbes in a fermentation, it is important to omit DNA from dead microbes before building predictive models based on sequences. One way to sequester 'dead' DNA, and therefore not sequencing it, is the use of propidium mono azide. Next-generation sequencing techniques that profile, e.g., the 16S gene present in all bacteria are increasingly used over molecular biology techniques, e.g. gel-based methods. The bioinformatics analysis of 16S data from food fermentations is quite well-established, resulting in descriptions of the taxa present in a particular fermentation at best at the species level, but for some taxa, the genus level is challenging to obtain.

There is a large biodiversity beyond the species level that is not taken into account with, e.g., 16S sequencing. Even within a bacterial species, there is considerable biodiversity. For example, all genes present in strains of the *Lactobacillus* genus (its pan-genome) comprise over 14 000 gene families, with a single genome encoding ~3 000 proteins. A gene family typically consists of genes that are evolutionary conserved, but that might have different enzymatic functions depending on the specific protein sequence. Comparative genomics, in combination with molecular strain typing, techniques have been used to uncover strain-level diversity in complex, yet relatively defined, fermentations in general and specifically for *L. lactis* and *Leuconostoc mesenteroides* from cheese, *Lactobacillus sakei* from meat fermentations, *Lactobacillus sanfranciscensis* in sourdough fermentations and wine yeasts.

With shotgun metagenomics, the DNA in the mixed-culture fermentation is profiled, but strain-level diversity is extremely difficult to deduce from shotgun metagenomics sequence fragments. On the other hand, due to the enormous biodiversity, the actual presence of any strain isolate thought to be of importance in a particular mixed-culture fermentation should be established. The combination of shotgun metagenomics and comparative genomics could prove to be particularly powerful, as the shotgun metagenome DNA sequences can be aligned to the genomes of isolates in order to prove that the functionality present in the isolates covers that of the metagenome.

Metatranscriptomics approaches allow profiling the mRNA-derived sequences of a complex fermentation. An advantage of metatranscriptomics over metagenomics approaches is that the gene expression measurement allows determining what genes are actually expressed in a mixed culture. Application of 'metatranscriptomics' using microarrays with the genomes of several species to determine global gene expression across species has been reported for Kimchi. Only recently, metagenome and metatranscriptome sequencing of bacterial communities involved in cheese rind fermentations has been reported. The strength of this study is that the metagenomics and metatranscriptomics profiles were traced to their likely sources (genome sequences of isolates from the rind cheese fermentation). Using experimental setups like the latter in combination with metabolomics measurements and appropriate follow-up studies should strengthen the point to use metagenomics / metatranscriptomics techniques to characterize and potentially optimize fermentations.

Bacteriophages play an important role in industrial fermentations due to the phenomenon of maintaining biodiversity through phage predation, but also because phage sweeps disrupt fermentation processes. Currently, however, predicting the specificity of bacteriophages and the interactions between microbes in mixed-culture fermentation are time-consuming tasks.

Bioinformatics techniques that analyse the interaction of microbes and bacteriophages, and in-depth knowledge of the metabolic requirements of the microbial consortia present during fermentation could in the future lead to knowledge-based improvements of fermentation stability. This could be achieved by performing experiments with synthetic microbial consortia. The design of these consortia is currently being developed, and cross-kingdom interactions are being studied. In a study where cheese rind bacterial communities were created based on various -omics data, knowledge of the fermentation and dedicated follow-up experiments, the potential of predicting properties of complex fermentations was demonstrated. This study did not explicitly describe whether the selected strains (or close relatives) were actually present in a real fermentation. This has been described for representative *L. lactis* and *L. mesenteroides* strains of a complex cheese fermentation and an *L. lactis* strain from a defined consortium.

Branding, tracing and detection

Food production and food consumption take place in complex environments in which next to the microorganisms present in the natural environment, many other sources of proteins, fat and carbohydrates are present. The presence of the endogenous flora as well as the macromolecular structures of the food can cause a lot of difficulty in detection and tracing of specific microorganisms, such as potential food pathogens or probiotic strains added to the food product for enhanced functionality.

Next to classical detection DNA-based techniques such as (q) PCR, new methods based on genomic data have been developed that allow for a fast and accurate tracking or detection of specific species or even strains among the natural microflora. By specific amplification and sequencing of a locus that was identified to be discriminatory between different *L. plantarum* strains, it was shown that one could quantify the relative presence of different strains through the passage of the gastrointestinal tract. This same approach can also be followed to design specific primers to discriminate between pathogenic and non-pathogenic populations of specific species and to detect a strain of interest in food products, allowing dedicated branding of a specific product.

Next to dedicated tracing of a single strain, metagenome approaches as described for studying complex fermented products, for example in cheese and fermented foods of plant origin, will also have their benefit in the detection of spoilage bacteria. Especially as these methods allow for direct profiling of the product, and do not require a culture step that could create bias in the results, they could very well prove to be more specific to detect spoilage bacteria from a product. Culturing steps will always have their merit due to limited

costs and requirement of limited amounts of material. Especially in fermented products, 16s community profiling approaches will allow detecting low abundant microbes that might be overgrown in culture-dependent detection methods.

Perspectives

Bioinformatics is increasingly applied in food fermentation and safety. Below we describe some new and exciting developments in this field.

Sequence-based prediction of microbial functionality is just starting. An inventory is needed of which functionality for which bacteria can reliably be determined using sequence data. New publicly available data sets with genotype/phenotype/transcriptome such as those available for *L. lactis* and *L. plantarum* could help to develop new sequence-based functional prediction strategies such as further specified protein domains to more specifically screen for, e.g., carbohydrate active enzymes and relating promoters or regulatory binding sites to phenotype. By consolidating the above information, a knowledge-based *in silico* screening of culture collections for desired traits can be established. This would require databases that use controlled vocabularies to integrate data from genomics, systems biology, phenotypes, ingredient information, properties of batches of foods, on-line measuring of parameters during the food making process and 'biomarkers' for functionality in specific taxa (based on, e.g., GTM). Specific emphasis should be put in propagating the FAIR (findable, accessible, interoperable, re-usable; http://datafairport.org/) principle in storing data. Given that analyses will become more standardized and computer resource-intensive, the software and databases could be set up in a virtual machine that can subsequently be run on computer clusters or in the cloud. First steps towards data consolidation are being made in the EU-funded project GenoBox (www.genobox.eu) that aims to create a database that consolidates genotype and phenotype data that allow screening microbial genomes for functionality and safety risk factors.

Similarly, IBM and MARS have established a consortium that aims to sequence the food supply chain (http://www.research.ibm.com/client-programs/foodsafety/).

Their aim is to determine nominal levels of microbial components in many food products across the globe. The resulting database can be used to assess risks of the presence of certain microbes/functionality in a given food product. Given that sufficient biodiversity has been recorded into this database, it could also be used for branding products based on unique microbiota signatures present in fermented products or foods that contain a microbiome. Another important factor to consider in steering the performance of fermentations is the interactions between microbes and their environment. This new layer of complexity has been studied, for instance, for microbe–plant interactions for rice or coconut and the use of systems biology beyond genome-scale metabolic models by using kinetic models to describe interactions between microbes and their matrix. These studies require a substantial knowledge base on both the properties of the microorganisms and the physical properties of the matrix in which the organism operate.

Conclution

In conclusion, the increasing amount of data on food fermentation and safety encourages consolidating this information in databases that with the right experimental design, algorithms, expertise and follow-up experiments should allow enhancing the prediction of fermentation performance and safety.

Acknowledgements

The entire matter has been achieved under extreme guidance favoring the in depth cultivation with a positive output from Dr. Sukumar Roy, Professor and Head, Department of Biomedical Engineering, Netaji Subhas Engineering College, Garia, Kolkata, India.. Dr. Sukumar Roy contributed a pioneer role to the design of the study, data analysis, and revision of the manuscript. And for that, Dr. Majumder is grateful for his extreme support to make the endeavor successful. Dr. Partha Majumder, being a Human Physiologist and Systems Biologist, contributed major role in order to establish correlation between different Systems biology and Bioinformatics tools & cascades.

References

1. Smid EJ, Kleerebezem M Production of aroma compounds in lactic fermentations. Annu Rev Food Sci Technol 2014;5:313-26
2. Cifuentes A Foodomics: the necessary route to boost quality, safety and bioactivity of foods. Electrophoresis 2014; 35:1517-18.
3. Bergholz TM, Moreno Switt AI, Wiedmann M Omics approaches in food safety: fulfilling the promise? Trends Microbiol 2014; 22:275-81.
4. Garrigues C, Johansen E, Crittenden R Pangenomics-an avenue to improved industrial starter cultures and probiotics. Curr Opin Biotechnol 2013; 24:187-91.
5. Talukdar V, Konar A, Datta A, Changing from computing grid to knowledge grid in life-science grid. Biotechnol J 2009; 4:1244-52.
6. Goecks J, Nekrutenko A, Taylor J, Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 2010; 11:R86.
7. Tiwari A, Sekhar AK Workflow based framework for life science informatics. Comput Biol Chem 2007; 31:305-19.
8. O'Driscoll A, Belogrudov V, Carroll J, HBLAST: parallelised sequence similarity-a Hadoop MapReducable basic local alignment search tool. J Biomed Inform 2015.
9. Deutsch EW, Mendoza L, Shteynberg D, Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics Cline Apple 2015.
10. Hwang Y, Lin C, Valladares O, HIPPIE: a high-throughput identification pipeline for promoter interacting enhancer elements. Bioinformatics 2014.
11. Tobes R, Pareja-Tobes P, Manrique M, Gene calling and bacterial genome annotation with BG7. Methods Mol Biol 2015; 1231:177-89.
12. Cortes C, Vapnik V Support-vector networks. Mach Learn 1995; 20:273-97.
13. Breiman L Random forests. Mach Learn 2001; 45:5-32.
14. Field D, Tiwari B, booth T. Open software for biologists: from famine to feast. Nat Bioethanol 2006; 24:801-3.
15. Koren S, Phillippy AM one chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. Curr Opin Microbiol 2015; 23:110-20.