CHAPTER 1

Pan-omics focused to Crick's central dogma

Arun Kumar Jaiswal*[ab], Sandeep Tiwari*[a], Guilherme Campos Tavares[h], Wanderson Marques da Silva[d], Letícia de Castro Oliveira[b], Izabela Coimbra Ibraim[a], Luis Carlos Guimarães[e], Anne Cybelle Pinto Gomide[a], Syed Babar Jamal[c], Yan Pantoja[e], Basant K. Tiwary[i], Andreas Burkovski[j], Faiza Munir[k], Hai Ha Pham Thi[l], Nimat Ullah[k], Amjad Ali[k], Marta Giovanetti[a,m], Luiz Carlos Junior Alcantara[a,m], Jaspreet Kaur[n], Dipali Dhawan[o], Madangchanok Imchen[p], Ravali Krishna Vennapu[p], Ranjith Kumavath[p], Mauricio Corredor[q], Henrique César Pereira Figueiredo[g], Debmalya Barh[f], Vasco Azevedo[a], Siomar de Castro Soares[b]

[a]PG Program in Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil
[b]Department of Immunology, Microbiology and Parasitology, Institute of Biological Science and Natural Sciences, Federal University of Triângulo Mineiro (UFTM), Uberaba, Brazil
[c]Department of Biological Sciences, National University of Medical Sciences, Rawalpindi, Pakistan
[d]Institute of Agrobiotechnology and Molecular Biology, INTA-CONICET, Buenos Aires, Argentina
[e]Institute of Biological Sciences, Federal University of Pará (UFPA), Belém, Brazil
[f]Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Purba Medinipur, India
[g]AQUACEN, National Reference Laboratory for Aquatic Animal Diseases, Ministry of Fisheries and Aquaculture, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[h]Universidade Nilton Lins, Manaus, Brazil
[i]Centre for Bioinformatics, Pondicherry University, Pondicherry, India
[j]Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[k]Department of Plant Biotechnology, Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan
[l]Faculty of Biotechnology and Environmental Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
[m]Laboratório de Flavivírus, IOC, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil,
[n]University Institute of Engineering and Technology (UIET), Department of Biotechnology, Panjab University, Chandigarh, India
[o]Baylor Genetics, Houston, TX, United States
[p]Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Kasaragod, India
[q]GEBIOMIC Group, FCEN, University of Antioquia, Medellin, Colombia

1 Introduction

Since the development of the first DNA sequencing technologies, many organisms had their complete DNA repertoire sequenced by Sanger and next-generation sequencing (NGS) technologies, creating the area of genomics, which was originated by the fusion of the words gene and chromosome [1]. In this scenario, a genome is the complete dataset of genes of a given organism. Nowadays, there are more than 200,000 genome projects registered at the Genome Online Database (GOLD), whereas more than 120,000 are

* These authors contributed equally to this work.

Pan-genomics: Applications, Challenges, and Future Prospects
https://doi.org/10.1016/B978-0-12-817076-2.00001-9

genomes isolated from bacteria (https://gold.jgi.doe.gov/statistics). Bacteria are widely distributed all over the world and have implications in health, agriculture, industry, and others. Besides, their genomes are small, highly compact, and do not present many repetitions, making them good targets for genome sequencing, once their genomes are easier to sequence than the ones from other organisms. Also, from the genome sequence of bacteria, it is possible to find virulence factors, antibiotic resistance genes, new therapeutic targets for vaccine and drug development, and industrially important genes [2, 3].

Another important point of the development of NGS technologies was the genome sequencing process that has become cheaper and faster, making it possible for small laboratories to use the technology in daily routine. NGS made possible the comparison of several genomes in a multipronged strategy, where phylogenomics, genome plasticity, and whole genome synteny analyses are easier to perform nowadays (Fig. 1). Also, RNA sequencing (RNA-seq) by these platforms and the development of new technologies for sequencing the complete dataset of proteins of an organism created the areas of transcriptomics and proteomics, respectively [4, 5]. Altogether, genomics is responsible for the identification of the complete dataset of genes of a given organism, whereas transcriptomics and proteomics are important for the identification of genes that are differentially expressed

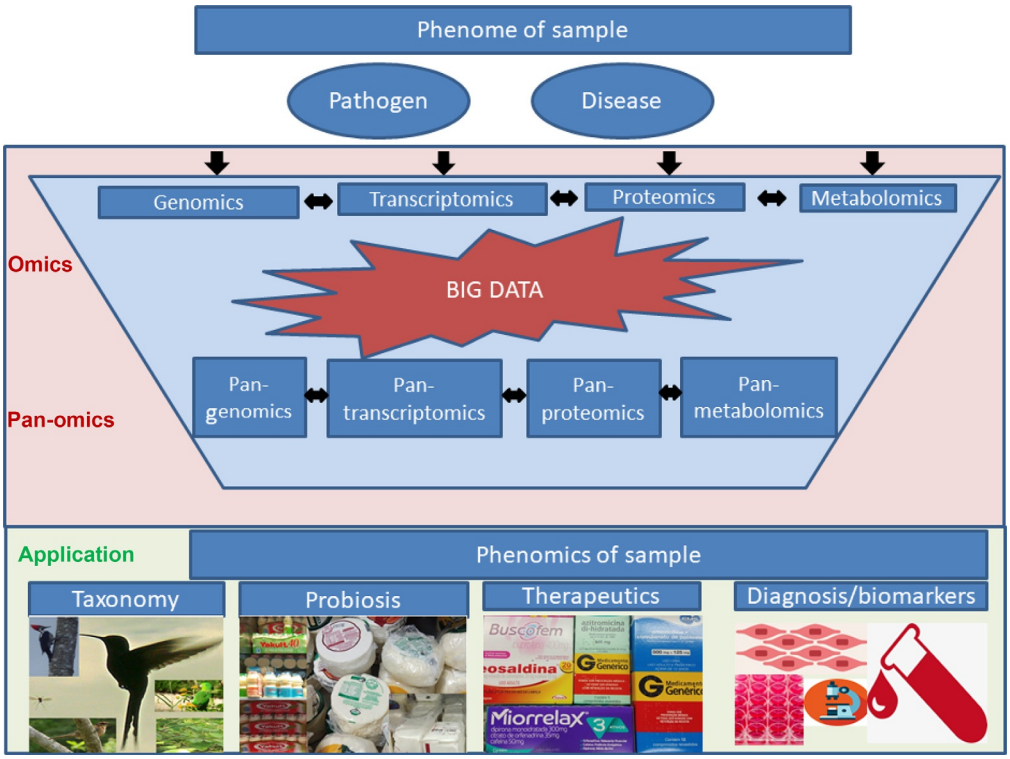

Fig. 1 Pan-omics and its applications.

between strains or species. Finally, the efforts to compare several genomes at once created the area of pan-genomics, which will be further discussed in this book.

1.1 Brief overview of pan-genomics

The term pan-genomics was created by Tettelin and collaborators, in 2005 [6], to describe the complete dataset of genes of a given species through the sequencing of several strains of this species. The pan-genome is composed of the core genome, shared genome, and singletons subsets, whereas the core genome is composed of all the commonly shared genes by all strains of the species; the shared genome contains genes that are present in two or more, but not all strains from a species; and the singletons are strain-specific genes (Fig. 2). From these subsets, one can extrapolate the data to find vaccines and drug targets from the core genome, whereas the shared genes and singletons are responsible for differences between the strains that are normally responsible for the emergence of new pathogens and the adaptation to new traits [6–10].

Normally, the core genome is composed of housekeeping genes and other genes important for metabolism and other important functions of the organism, whereas the shared genes and singletons are the result of genome plasticity. Genome plasticity is the dynamic property of DNA which involves the gain, loss, and rearrangement of genes through plasmids, phages, and genomic islands (GEIs). GEIs are huge blocks of genes acquired through horizontal gene transfer (HGT) that normally share a function in common. They are classified according to the functions of the genes into: pathogenicity islands, harboring virulence factors; metabolic islands, composed of metabolism-related genes; resistance islands, with antibiotic resistant genes; and symbiotic islands, which share in common the presence of symbiotic-related genes [11, 12].

Normally, the subsets of the pan-genome are identified by the use of orthology analyses, which first identify all orthologous genes from the complete dataset using all-vs-all blasts or other alignment search tools. Next, the datasets are classified according to their homology to genes from other strains in the subsets. After the classification, the data is plotted in a chart and mathematical formulas are used to fit the specific curves. Two such

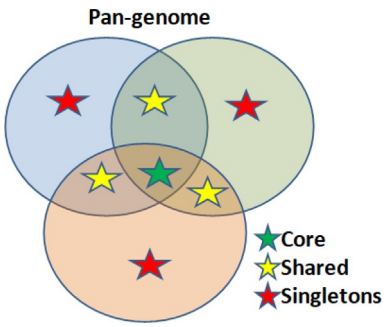

Fig. 2 Schematic representation of the core genome, shared genome, and singleton subsets of pan-genome analysis.

formulas are Heaps' law for the pan-genome development and least-squares fit of the exponential regression decay for the core genome and singleton subsets, which are described respectively as: $n = k \cdot N^{-\alpha}$, where n is the number of genes, N is the number of genomes, and k and α are constants defined by the formula; and $n = k \cdot e^{-x/\tau} + tg\theta$, where n is the number of genes, x is the number of genomes, e is Euler's number, and k, τ, and $tg\theta$ are constants defined by the formula [6, 9].

1.2 Open and closed pan-genomes

According to Heap's law, the α value is representative of the current dynamics of the pan-genome, where an α higher than 1 is representative of a closed pan-genome and an α lower than 1 represents an open pan-genome. A closed pan-genome has all possible genes represented and only few genes will be added to the pan-genome if more genomes are to be sequenced, whereas an open pan-genome is still not fully represented and the sequencing of new genomes will add many genes to the analyses [6, 9]. This definition is controversial, however, once the incorporation of GEIs may change the composition of the pan-genome drastically, even for closed pan-genomes, taking it to be open again. Most important, environmental bacteria and extracellular pathogens normally have open pan-genomes, once they still need to adapt to new traits, whereas obligate intracellular pathogens tend to have closed pan-genomes once they are not in constant contact with other bacteria. Also, intracellular pathogens have lost many genes during evolution, completely adapting to the host organism and, thus, present very compact genomes with a high percentage of essential genes [13].

According to least-squares fit of the exponential regression decay, the $tg\theta$ is representative of the number of genes present in the core genome after stabilization of the core genome curve and, also, of the number of genes that will be added to the analyses after a new genome is sequenced from the singleton development curve. Based on that, researchers may choose the species that need more strains to be sequenced and which do not. Finally, the highest the $tg\theta$ on the singleton development, the lower the α, once a high number of genes will be added to the analyses taking the pan-genome to be more open and the α to be lower (Fig. 3). The opposite is also true, the lower the $tg\theta$, the higher is the α value [6, 7, 10].

1.3 Computational methods used in pan-genomics

Computational methods to find more efficient data structures, algorithms, and statistical methods to perform bioinformatics analyses of pan-genomes have been studied because it is known that in a pan-genome analysis the greater the number of genomes taken to the analysis the greater will be the computational costs, that is, the discovery of a pan-genome content is an NP-hard problem because comparisons between all sets of genes are necessary to solve the task. Furthermore, in an effort to compute standardized pan-genome analysis and minimize computational challenges, several online tools and software suites

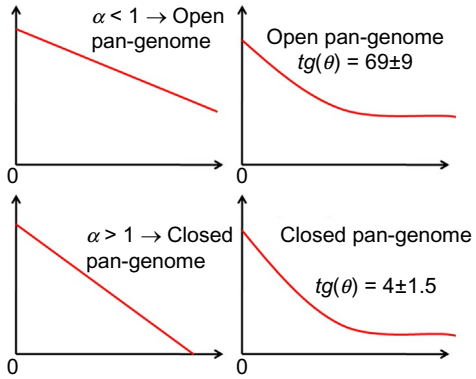

Fig. 3 The concept of open and close pan-genome.

have been developed. Examples of such applications are: PGAP [14], one of the most complete profile available for performing five analysis modules, but the runtime of the analysis grow approximately quadratically with the size of input data and are computationally infeasible with large datasets. The software Roary [15] and BPGA [16] was created to address the computational issues related to performance and execution time. Roary performs a rapid clustering of highly similar sequences, which can reduce the runtime of BLAST. BPGA is an ultrafast computational pipeline with seven functional modules for comprehensive pan-genome studies and downstream analyses. Pan-genome analysis can be applied in many different application domains, such as microbes, metagenomics, viruses, plants, cancer, and others [17]. Nowadays, the processes of similarity search and pan-genome visualization are two of the wide variety of particular computational challenges that need to be considered. For this, novel different computational methods and paradigms are needed over the years, making the computational pan-genomics a subarea of research in rapid extension. Furthermore, new technologies that are emerging in rapid development allow to infer the pan-genome with three-dimensional conformation, which means that possibly in the future three-dimensional pan-genomes will not only represent all sequence variation of the species or genus, but will also encode their spatial organization, as well as their mutual relationships in this regard.

1.4 Applications of pan-genomics in evolutionary studies

The manifestation of rich genetic diversity in the form of a pan-genome in a species is an evolutionary puzzle. These three distinct parts of a pan-genome (core, shared, and singletons) of a particular species may undergo different evolutionary trajectories under the differential influence of evolutionary forces. An ideal pan-genome is expected to be very complete, comprehensive, efficient, and stable [18]. The pan-genome of a species has some evolutionary signatures in the form of gene content and single nucleotide

polymorphism (SNP). These evolutionary signatures are useful in inferring the phylogenetic relationship among different strains of a species based on the pan-genome.

An evolutionary pan-genomic study of microbes provides a holistic picture of all the genomic variations of a species. These genomic variations endow the bacteria with their unique pathogenic properties and subsequent development of resistance to various antibiotics. Thus, a complete mechanistic detail of the processes involved in the pathogenesis and frequent antibiotic resistance in a bacterium will further pave the way for better detection methods and effective control strategies for the pathogen. In addition, evolutionary pan-genomics of a useful bacterium will help us in exploiting maximally the full potential of the microbe in enhancing industrial productivity. In fact, it will be a boom for the industries actively involved in the production of pharmaceuticals and dairy products using microbial cultures. Eukaryotes including crop plants and farm animals have abundant genomic variations in the form of SNP, copy number variants (CNVs), and presence/absence variants (PAVs). The discovery of SNPs associated with productivity or disease resistance in a crop or a farm animal will be much more efficient with the availability of a complete pan-genome of the species [19].

In a recent past, a work published by Benevides et al. [20] utilized 16S rRNA gene phylogeny, whole-genome multilocus sequence typing (wgMLST), phylogenomics, gene synteny, average nucleotide identity (ANI), and pan-genome to explain the phylogenetic relationships in a better way among strains of *Faecalibacterium*. For this, they used 12 newly sequenced, assembled, and curated genomes of *Faecalibacterium prausnitzii*, which were isolated from the feces of healthy volunteers from France and Australia, and combined these with five strains already published, which were downloaded from public databases. The phylogenetic analysis of the 16S rRNA along with the wgMLST profile and the phylogenetic tree based on the comparison of the similarity of genome supports the grouping of *Faecalibacterium* strains in different genospecies [20].

In another work published by Chen et al. [21], the comparison of whole genome and core genome multilocus sequence typing (MLST) and SNP analyses were carried out to show the maximum biased power achieved by using multiple analyses. It was required to differentiate isolates associated with outbreak from a pulsed-field gel electrophoresis (PFGE)-indistinguishable isolate collected in 2012 from a nonimplicated food source. Whole genome sequencing (WGS) has been proven as a powerful subtyping tool for bacteria like *L. monocytogenes*, a foodborne pathogen [21]. A company produced an environmental isolate that was highly similar to all outbreak isolates. The difference observed between unrelated isolates and outbreak isolates was only 7–14 SNPs; consequently, the minimum spanning tree from the analyses of whole genome, phylogenetic algorithm, and usual variant calling approach for core genome-based analyses could not offer the difference between unrelated isolates. This also suggested that the SNP/allele counts should always be pooled with WGS clustering analysis produced by phylogenetically meaningful algorithms on an adequate number of isolates, and the SNP/allele

onset alone does not provide enough evidence to demarcate an outbreak [21]. Hence, it was proposed that the comparison of pan-genome subcategories and their related α value may be utilized as an alternate approach, along with ANI, in the in silico cataloging of new species [20, 22]. We hope that the ever-expanding pan-genome across different species and genera will give impetus to a better data structure of the pan-genome and novel computational methods for a robust evolutionary pan-genomic analysis in near future.

2 Applications of Pan-genomics in Bacteria

2.1 Applications of pan-genomics in model bacteria

Advancement in sequencing technologies and development in sophisticated bioinformatics tools created an overwhelming number of microbial genomic data and allowed the scientific community to estimate the pan-genome of a species. Identification of novel dispensable genes has applications in characterizing novel metabolic pathways, virulence determinants, and molecular fingerprinting targets for epidemiological studies and core genes can be used to predict the evolutionary history of the organism [9]. Therefore, pan-genome analyses are now considered the indispensable and gold standard for bacterial genome comparisons, evolution, and diversity. It is also useful to develop a vaccine against the pathogens of epidemic diseases by filtering different functional genes in the core genome using reverse vaccinology approaches [23].

There are a number of freely accessible tools, pipelines, and web-servers available to estimate the microbial pan-genome including Roary, BPGA, PGAP, PGAPx, Panseq, PanOCT, etc. [16]. A number of model bacterial species pan-genome is determined by researchers and a vast majority of those human pathogens exhibit an open pan-genome, as they colonize multiple environments that facilitate them to exchange genetic materials. These organisms include *Escherichia coli*, *Meningococci*, *Streptococci*, *Salmonellae*, *Helicobacter pylori*, etc. [24]. Therefore, in dealing with such species a reasonable number of genomes is usually required to define the complete gene repertoire of these species. On the other hand, species living in isolated (close) habitats having less possibility to exchange genetic material tend to have closed pan-genome, for example, *Mycobacterium tuberculosis*, *B. anthracis*, and *Chlamydia trachomatis* [25]. Hence, pan-genome analyses serve as a framework to determine and understand the genomic diversity in bacterial species. In Chapter 17, we have discussed the bacterial pan-genome analysis performed till date with specific examples from model organisms along with studying approaches, technical implementations, and their outcome.

2.2 Applications of pan-genomics in *Corynebacterium diphtheriae* and *Corynebacterium ulcerans*

The development of diphtheria toxoid vaccines in the 1920s, the start of mass immunization in the 1940s, and the global introduction of the Expanded Program on

Immunization (EPI) by the World Health Organization (WHO) in 1974 led to a dramatic decrease of diphtheria cases, both in industrialized and developing countries [26]. However, despite this tremendous success story, diphtheria has not been eradicated yet. This has been illustrated dramatically by a diphtheria pandemic connected to the breakdown of the former Union of Socialist Soviet Republics with more than 157,000 cases and more than 5000 deaths reported between 1990 and 1998. Even after the pandemic has finally stopped, local breakouts have been observed constantly during the last years and the reported global cases increased from about 7000 in 2016 to almost 9000 in 2017 with a focus on countries with limited or lacking public health systems, for example India, Indonesia, Nepal, Pakistan, Venezuela, and Yemen. Consequently, *Corynebacterium diphtheriae*, the etiological agent of respiratory and cutaneous diphtheria, is still present on the list of the most important global pathogens [27]. Furthermore, the frequency of human diphtheria-like infections associated with *Corynebacterium ulcerans* appears to be increasing [28]. This species, which was recognized before as a commensal of a large number of animal species, is closely related to *C. diphtheriae* and recognized as an emerging pathogen today [28, 29].

The need of fast and unequivocal identification of especially pathogenic *C. diphtheriae* led to the early development of a number of different methods such as biovar discrimination based on different biochemical reactions, Elek's test to immunologically distinguish between toxigenic and nontoxigenic strains, restriction fragment length polymorphism (RFLP), single-strand conformation polymorphism (SSCP), phagetyping, spoligotyping, ribotyping, MLST and others. This plethora of methods was significantly improved when next-generation sequencing was introduced. The first genome sequence of *C. diphtheriae* was published in 2003 and showed the presence of the *tox* gene on a bacteriophage in addition to a number of other horizontally acquired virulence-associated genes [30]. Subsequent pan-genome studies allowed unraveling the extent of genomic diversity within *C. diphtheriae* and the role of HGT as a source of variation between strains. Furthermore, pan-genomics of *C. ulcerans* helped to estimate the virulence potential of different strains and to verify zoonotic transmission from animals to patients. Today, pan-genomics of *C. diphtheriae* and *C. ulcerans* allow elucidating global transmission traits and local adaptations of pathogenic corynebacteria and, hopefully, a better understanding of population dynamics and strain evolution will help combat diphtheria and other *Corynebacterium*-associated diseases in future.

2.3 Applications of pan-genomics in multidrug-resistant human pathogenic bacteria and pan-resistome

The pan-genome will probably be the largest molecular evolutionary history of the organism ever written. This will integrate all the pan-phenotypes existing on Earth, such as the pan-proteome, the pan-transcriptome, and especially, a portion of pan-genome that has made the organisms successful on Earth: the pan-resistome. The pan-genome

represents the set of all current genes in the genomes of a group of organisms. The basic genome common to all bacteria contains about 250 gene families in the extended core, the specific niche adaptive genome of about 8000 gene families in the character gene pool, and the pan-genomic diversity (accessory genes) of more than 139,000 rare gene families scattered throughout the bacterial genomes [31]. The pan-genome analysis, whereby the size of the gene repertoire accessible to any given species is characterized along with an estimate of the number of whole genome sequences required to proper analysis, and currently it is increasing 10 years after Tettelin et al. [6] publication. Different current models for the pan-genome analysis, accuracy, and applicability depend on the case at hand [32]. The NCBI, EMBL, KEEG, PATRIC, MBGD, ENSEMBL, and JGI-IMG/M databases provide complete downloadable genomics information, which can be analyzed for intraspecies diversity, and determine the pan-genome using software tools, currently developed to perform via a personal server [32], or even online resources. The pan-genomics is now a cutting edge of computational genomics field. Pan-genomics is a subarea of computational biology [17]. Therefore, the notion of computational pan-genomics intentionally passes through many other bioinformatics-related disciplines.

The resistome, a term coined by Wright [33], comprises all the genes and their products that contribute to resist whatever environment, substance, or some extreme grow factor. Updated data will close to the metadata available for establishing what part of resistome traits belong both to core-genome as accessory genome inside all bacterial species as well as will offer a broader perspective of bacterial antibiotic resistance. The WHO summarizes antimicrobial resistance (AMR) as the resistance of a microorganism to an antimicrobial drug that was originally effective for the treatment of infections caused by themselves. An adequate approach to solving major questions about the resistome inside of the bacterial genome [34] is to perform a pan-genomics analysis. The updated pan-genome data will be close to the metadata available for establishing the part of resistome traits that belong both to core-genome as accessory genome in bacterial species; as well as a broader perspective of antibiotic resistance in bacteria. The emergent antibiotic-resistant pathogenic bacteria are a current menacing concern. *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, and coliform bacteria are the new emergent antibiotic-resistant bacteria according to the WHO. Pan-genomics has tackled some important concerns, which would be impossible to solve using classical molecular biology or descriptive genomics: it is very important to define the core and accessory genome for establishing the plasticity of resistome. Thousands of unknown bacteria and microorganisms are exposed to manufactured antibiotics, leading us to assume that there are no means to prevent this catastrophe. In opposition, pan-genomics is a powerful approach to prevent such disaster. We must move toward sequencing of known and unknown species, classify them, and establishing its antibiotic-resistant status, their pan-genome, and come out with new alternatives for reducing antibiotic consumption nowadays.

2.4 Applications of pan-genomics in veterinary pathogens

Following the development of NGS, the number of sequenced genomes filed exponentially [35]. Thus, projects aimed at studying groups of organisms became viable, and thus, several studies appeared that are called Omics studies. The studies involving pan-genomes are exposing important information on the differences and similarity between organisms of the same or between species. For concept purposes, we have the Pan-genome as a set of genes in a given group of individuals [10]. This information is being explored and applied by several scientific fronts, for example, in bacteria that infect animals and humans. The main applications of these studies are in the development of prophylactic and diagnostic methods in less time and with less cost, more precise taxonomic studies, studies on genetic variations, and pathogenesis [17]. In this chapter, we describe more recent research involving pan-genomics of the pathogenic bacteria that cause veterinary diseases, including some responsible for zoonoses, they are: *Corynebacterium pseudotuberculosis*; *Corynebacterium ulcerans*; *Streptococcus suis*; *Brachyspira hyodysenteriae*; *Moraxella bovoculi*; *Pasteurella multocida*; *Mannheimia haemolytica*; *Clostridium botulinum*; *Campylobacter*; *Streptococcus agalactiae*; *Francisella tularensis*; *Corynebacterium diphtheriae*; *Brucella* spp. Finally, it is worth highlighting that the influence of the approaches with big data and artificial intelligence are increasing and the influences of these in Pan-genomic studies will bring a new era of studies and discoveries.

2.5 Applications of pan-genomics in aquatic pathogenic bacteria

The sustainability of aquaculture industry is critical both for global food security and economic welfare. However, the massive wealth of pathogenic bacteria poses a key challenge to the development of a sustainable biocontrol method. Recent advances in genome sequencing study combined with pan-genome analysis can be an efficacious management applied to numerous aquatic pathogens [36]. Thus, routine pan genome analyses of genomic-derived aquatic pathogens will deduce the phylogenomic diversity and possible evolutionary trends of aquatic bacterial pathogen strains, elucidate the mechanisms of pathogenesis, as well as estimate patterns of pathogen transmission across epidemiological scales. The whole genome sequencing data is the opportunity to revolutionize the molecular epidemiology of aquaculture pathogens as it has for those pathogens of relevance to public health [37]. Challenges of aquaculture disease management are the biological diversity of pathogens, host-pathogen interactions (e.g., different modes of adaptation and transmission), and shifting environmental pressures, in particular climate change. Hence, analysis of pathogenic phenotype combined with genotype derived from the full potential of genome sequencing data is critical to reconstruct pathogen transmission routes on local and global scales, as well as mitigate disease emergence and spread.

Comparative pan-genome analyses are an effective tool which could possibly be extended to the analysis of aquatic microorganisms and to dynamic characteristics and adaptation to a broad range of their hosts and environmental niches. Conspicuously, our previous pan-genome analysis [38] showed that strain WFLU12 isolated from marine fish exhibited niche-specific characteristics of energy production and conversion, and carbohydrate transport and metabolism by exploring genes in the gene repertoire of strains. Based on the pan-genome categories, the functional annotations of selected genes can be reanalyzed with the Virulence Factors Database (VFDB), Clusters of Orthologous Groups (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Antibiotic Resistance Genes Database (ARDB). Also, comparative pan-genome has advanced to the point when genes are predicted as belonging to cell surface-exposed proteins (SEPs) from important pathogens, including outer membrane proteins, and extracellular proteins. These predicted genes are serving as vaccine candidates in an animal model called Reversed Vaccinology (RV) [39]. In aquaculture, SEPs from pathogens include several important virulence factors that play key roles in bacterial pathogenesis and host immune responses. For example, the expression of *esa1* from *Edwardsiella tarda*, a D15-like surface antigen, in the Japanese flounder model induced the expression of a broad spectrum of genes possibly involved in both innate and adaptive immunity, as well as a high level of fish survival and produced specific serum antibodies [40]. Vaccination using SEPs results in the development of protective effects against *Aeromonas hydrophila* infection, *Flavobacterium columnare* infection, *Pseudomonas putida* infection, and Edwardsiellosis [as in the review of Abdelgayed [41]]. A recent study [42] has successfully implemented a pan-genome analysis to screen SEPs from 17 representative *Leptospira interrogans* strains covering multiepidemic serovars from around the world, and 118 new candidate antigens were identified in addition to several known outer membrane proteins and lipoproteins. We highly consider that the rapid increase in the number of genome sequencing of aquatic pathogens will allow us to develop a rapid-response infection control protocols, but also be a potential trend for studying aquatic pathogenic bacteria to improve the cross-serotype efficacy of vaccines in farmed fish and stem the disease outbreak when implementing pan-genome analysis (using RV strategy). In the chapter "Pan-genomics of aquatic animal pathogens and its applications," we reviewed comparative pan-genome analysis with a particular focus on controlling aquatic diseases and give real-world examples by analyzing genome sequencing data derived from aquatic bacterial isolates.

2.6 Pan-genomics applications for therapeutics

The emergence of bacterial resistance is occurring, threatening the ability of antibiotics that have transformed medicine and saved millions of lives around the globe [43, 44]. The occurrence of bacterial resistance has been identified since the beginning of the antibiotic era but the emergence of most dangerous and easily communicated strains has been

reported in past two decades [45, 46]. After several years of the first patient treated with antibiotics, bacterial infections became a threat for society once again. This situation is mainly because of the misuse and/or overuse of antibiotics as well as the inefficiency of pharmaceutical companies for not producing advanced drugs, once economic investments have been reduced [44]. The Centers for Disease Control and Prevention (CDC) has categorized several bacterial strains as an alarming threat that need serious consideration for proper treatment and are already responsible for putting significant burden on the health-care system in the United States (US), ultimately, affecting patients and their families [43, 47, 48]. The infections caused by antibiotic-resistant strains of bacteria are pervasive worldwide [43, 44]. A national survey of infectious-disease specialists led by the IDSA Emerging Infections Network in 2011 found that about two-third (2/3) of the participants had seen a pan-resistant and deadly bacterial infection within the past few years [49]. The rapid emergence of resistant bacteria has been described as a nightmare by several public health organizations that could have disastrous results [50]. The WHO cautioned in 2014 that the disaster of antibiotic resistance is becoming dreadful [51]. Among Gram-positive pathogens, a universal endemic of resistant *S. aureus* and *Enterococcus* species are presently the biggest intimidation [48]. Vancomycin-resistant enterococci (VRE) and additional emergent pathogens are evolving resistance to numerous antibiotics used commonly [43]. The worldwide distribution of common respiratory pathogens includes *Streptococcus pneumoniae* and *Mycobacterium tuberculosis*, which are reported as epidemic [48]. Gram-negative pathogens are in general more troublesome because of the fact that they are becoming more resistant to almost all the available therapeutics, making the conditions evocative to the preantibiotic era [44]. The occurrence of multidrug resistant (MDR) Gram-negative bacilli has outdated all the practice in field of medicine [43]. The most common infections caused by Gram-negative bacteria in health-care settings are usually by *Enterobacteriaceae* (mostly *Klebsiella pneumoniae*), *Acinetobacter*, and *Pseudomonas aeruginosa* [43, 44]. The evolution of bacterial strains and development of antibiotic-resistant genes through HGT make it necessary to look for novel and advanced strategies to cope with the infections [52].

The in silico approaches like pan-genome, pan-modelome, subtractive genomics, and reverse vaccinology are playing vital roles in rapid identification of new therapeutic targets in the postgenomic era [53–55]. Comparative microbial genomics approach along with statistical analysis are useful tools for the identification of essential genetic contents commonly present in all pathogenic isolates, based on sequence similarity. In addition to essential genetic contents, it also helps to identify subset of genes encoding virulence and novel functions as the variable genome [56]. A pan-genome is usually divided into three parts, that is, core genes, accessory genes, and strain-specific genes. In the drug and vaccine discovery process, the very first step is always the identification of a suitable target. Subtractive genomics is a widely used process in this regard. In recent past, working with pathogenic bacteria, using computational approaches, a large number of novel

therapeutic targets has been identified, which are either resistant to drugs or no appropriate vaccine is available for these targets [54, 57]. The most popular approach for rapid identification of novel vaccine targets in postgenomic era is reverse vaccinology [54]. Strategies such as comparative genomics, subtractive genomics, and differential genome analyses are being broadly utilized for the identification of targets in several human and animal pathogens (Table 1), that includes *Mycobacterium tuberculosis* [62], *Treponema pallidum* [54], *Corynebacterium diphtheriae* [53, 64], *Hemophilus ducreyi* [52], *Neisseria gonorrhoeae* [59], and *Salmonella typhi* [63]. The basic principle of these approaches is the identification of genes/proteins that are not homologous to gene/protein of the host but are essential for the survival of the pathogen. However, the identified targets might be slightly homologous to host gene/protein but still can be selected for structure-based selective inhibitor development as a supplementary molecular target [54, 64–66].

2.7 Pan-genomics applications for probiotics

The term probiotic has become highlighted in the last few years, but few know that its use is already registered as fermented foods in books such as: the Holy Bible and sacred books of Hinduism [67, 68]. Probiotics are live microorganisms that may provide health to the host [69].

Table 1 Pan-genome studies in bacterial pathogens

Name	Strain/no of strains	No of genes/proteins	Host	Therapeutic drug/vaccine targets	References
Treponema pallidum	13	837	Human	15 vaccine/6 drug	[54]
Haemophilus ducreyi	28	1257	Human	13 vaccine/3 drug	[52]
Chlamydia trachomatis	NC_010287.1	934	Human	63 drug	[58]
Neisseria gonorrhoeae	FA 1090		Human	67 drug	[59]
Ureaplasma urealyticum	ATCC 33699	646	Human	2 drug	[60]
Corynebacterium diphtheriae	13	Not mentioned	Animal	8 drug	[53]
Helicobacter pylori	39	59,958	Human	28 vaccine	[61]
Mycobacterium tuberculosis	H37Rv genome	3989	Human	135 drug	[62]
Salmonella typhi		4718	Human	149	[63]

Its importance gained pace in the medical and biotechnological fields with the results found not only related with inflammatory bowel diseases (IBDs) [70, 71], but also with diabetes [72], multiple sclerosis [73], dermatitis [74], and in the production of heterologous proteins [75]. Many species play a role as probiotic and much more are in the process of testing (Table 2).

Table 2 Probiotics and their effects

Name	Strain	Status	Effect	References
Acinetobacter sp.	BR-12	R	Plant phosphate supply	[76]
Acinetobacter sp.	BR-12	R	Plant phosphate supply	[77]
Acinetobacter sp.	WR922	R	Plant growth	[78]
Bacillus amyloliquefaciens	G1	R	Bacterial infections in animals	[79]
Bacillus amyloliquefaciens	SC06	R	Bacterial infections in animals	[80]
Bacillus clausii	UBBC 07	C	Acute diarrhea	[81]
Bacillus coagulans	–	M	Irritable bowel syndrome (IBS)	[82]
Bacillus coagulans	–	C	Antibiotic-induced diarrhea	[83]
Bacillus licheniformis	2336	M	Acute enteric infections	[84]
Bacillus licheniformis	26L-10/3RA	M	Bacterial infections in animals	[85]
Bacillus licheniformis	8-37-0-1	M	Maintenance of aquatic conditions for animals; Heavy metal accumulation	[86]
Bacillus subtilis	E20	M	Immuno-protection for animals	[87]
Bacteroides fragilis	–	R	Autism spectrum disorders (ASD)	[88]
Bifidobacterium animalis subsp. *lactis*	BB-12	M	Reduces the risk of infections in early childhood	[89]
Bifidobacterium animalis subsp. *lactis*	Bb-12	M	*H. pylori* related	[90]
Bifidobacterium animalis subsp. *lactis*	Bb-12	C	Atopic dermatitis	[91]
Enterococcus faecalis (*Streptococcus faecalis*)	SL-5	C	Acne vulgaris	[92]
Enterococcus faecium (*Streptococcus faecium*)	CTC492	R	Antilisteral effect	[93]

Table 2 Probiotics and their effects—cont'd

Name	Strain	Status	Effect	References
Escherichia coli	M-17	R	Pouchitis	[94]
Escherichia coli	Nissle 1917	C	Ulcerative colitis; Crohn's disease; Inflammatory bowel disease (IBD)	[95–97]
Lactobacillus acidophilus	L-92	C	Atopic dermatitis	[98]
Lactobacillus acidophilus	LA-02 (DSM 21717)	C	Vulvovaginal candidiasis	[99]
Lactobacillus brevis	D7	M	Antioxidation process in animals	[100, 101]
Lactobacillus buchneri	P2	R	Cholesterol removal	[102]
Lactobacillus casei	DN-114001	C	Immune modulation	[103]
Lactobacillus casei	F-19	M	Food digestion	[104]
Lactobacillus crispatus	CTV-05	C	Urinary tract infection	[105]
Lactobacillus delbrueckii subsp. *bulgaricus*	OLL1073R-1	C	Reduces the risk of infection in the elderly	[106]
Lactobacillus rhamnosus	CGMCC 1.3724	C	Obesity	[107]
Lactobacillus rhamnosus	JCM1136	M	Immuno-protection for animals	[108]
Lactococcus lactis subsp. *cremoris*	IBB SC1	R	Immunomodulation	[109]
Oxalobacter formigenes	OxCC13	R	Calcium oxalate stone disease	[110]
Propionibacterium freudenreichii subsp. *shermanii*	–	C	Liver cancer	[111]
Streptococcus salivarius	K12	R	Halitosis	[112]
Weissella koreensis	OK1–6	R	Antiobesity	[113, 114]

R = research; C = Clinical trial; M = Marketed.

The Omics studies allowed an advance in the elucidation and characterization of the properties of these organisms, opening a vast field of application, besides providing new ways to access the information about their genomes. Following the pan-genomic approach, the pan-probiosis analysis consists in comparison of two or more strains, aiming to identify some points in the organism genome that differs or presents similarities related with probiotic characteristics, such as genes coding for adhesion.

In comparative genomics, for example, it is possible to retrieve a high number of genome information in silico—an attractive and cheap way [115]. There are some requirements that are important for an organism to be considered as probiotic which is determined through some mechanisms of action, like surviving to gastric acidity and bile salts [116], competing with other organisms via exclusion mechanisms and antimicrobial activity [117], and modulating the immune system [118], and these features may be used to gather the genome information in silico.

A comparative analysis with *L. lactis* subsp. *lactis* NCDO 2118 was performed aiming to find the potential probiotic characteristics of this strain. The authors found, through comparative genomics, phage regions, GEIs (metabolic and symbiotic), bacteriocins of three different classes, bile salts, and acid stress resistance genes found in other *L. lactis*, adhesion-related, and antibiotic-resistant genes. Besides that, comparing in vitro data of the aforementioned strain with another species, already described as nonprobiotic, they could identify genes encoding proteins (secreted and expressed) that are exclusive of NCDO 2118 [119].

Using a pan-genome microarray with probiotic *E. coli* isolates, Willenbrock and coauthors could characterize the pan-genome of 32 species based in two-control strain: *E. coli* K-12 and O157:H7. Despite they observed different sizes of genomes within the species, they believe they achieved the expected results, one of them being the characterization of the core genome with around 1560 essential genes [120].

Pan-genome approach was also used to discover probiotic characteristics of *L. lactis* WFLU12 [38] that showed resistance against streptococcal infection and improved the growth in olive flounder [121]. They identified some data that supported their previous work, like the identification of bacteriocins and genes involved in stress response. Comparing WFLU12 with other *L. lactis*, there are genes and gene clusters for specific niches based on carbohydrate metabolism, defense mechanisms, and envelope biogenesis [38].

Following the idea about niche-specific, Kant and coauthors worked with 13 *Lactobacillus rhamnosus* from different origins with the pan-genomic analysis. They used *L. rhamnosus* GG as reference, focusing in SEPs that may play a role in niche adaptability. The interesting thing was, they could find uncommon information in lactic acid bacteria, a *spaCBA* operon. This operon may be related with the origin of these strains, maybe of a similar microhabitat, for example [122].

Another species used as probiotic was analyzed via pan-genomics in the study by Smokvina and coauthors, in which 34 different *Lactobacillus paracasei* strains were studied using comparative genomics and pan-genomics. They identified 1800 orthologous groups representing the core genome and these genes were related with cell envelope, pili, hydrolases, or the production of branched short-chain fatty acid (SCFAs). About this, they found genes that encode these SCFAs: *bdkABCD*, only found in *Lactobacillus* until this date [123].

Nowadays, we have a lot of information about potential probiotic organisms, beyond those whose are commonly known in the market, but there is no database concentrating all the information about them, like genes related with bile juice and gastric acid resistance, genes coding adhesion, or secret proteins. A database with those information about known probiotic organism could help in future analysis be them in silico, in vivo, and in vitro. Finally, the comparative and pan-genomic analyses have an important role in the most diverse organism analyses and in the case of probiotic ones, it could be very helpful and elucidating in the precision to characterize new potential probiotics. The diversity inside the genomes may be observed and with this information it is possible to have a better idea of how many genomes will be necessary to characterize fully the organisms in these studies.

3 Pan-genomics of virus and its applications

Advances in DNA sequencing technology have ushered in a new era of pan-genomics and genomic surveillance, in which traditional molecular diagnostics and genotyping methods are being enhanced and even replaced by genomic-based methods to aid epidemiologic investigations of communicable diseases [124]. The ability to compare and analyze entire pathogen's genomes has allowed unprecedented resolution into how and why infectious diseases spread. The rapid development of sequencing technologies has made sequencing routine of viral genomes possible [125]. As these genomic-based methods continue to improve regarding speed, costs, and accuracy, they will increasingly be used to inform and guide infection control and public health practices [125a].

There are currently two major ways in which high-throughput sequencing technologies are used in public health and diagnostic applications: (i) to track outbreaks and epidemics in order to call public health responses and (ii) to characterize individual infections to tailor treatment decisions [126, 127].

Focusing on these aims, genome sequencing has been successfully used to describe unique and detailed insights into the transmission, biology, and epidemiology of many health-care-associated viral pathogens. Considering the improvements on portability and quality of sequencing, and the acceleration and standardization of analytical pipelines, the applicable routine of genome sequencing may soon become the common de facto method for infectious diseases control. Using genomic analysis tools to complement existing genotyping and epidemiologic methods, the future of infection control and prevention will lead to more targeted and successful interventions for outbreaks, which will ultimately result in the reduction of infectious diseases impact.

Next-generation sequencing techniques have transformed genomic studies from the analysis of single or few genomes to an ever-increasing amount of genomic data, bringing with it the need to develop novel techniques to treat efficiently, novel tools to assemble, analyze, and derive useful information from overwhelmingly large datasets. The analysis

of pan-genomes can uncover significant information regarding the genomes of interest. According to Guimaraes et al. [128], pan-genomic studies can help understand pathogen evolution, niche adaptation, population structure, and host interaction. Furthermore, it can help in vaccine and drug design, as well as in the identification of virulence genes.

In the context of virus investigations, pan-genomics and bioinformatics in general face great challenges. Rapid extraction of genomic features with an evolutionary signal will facilitate evolutionary analyses ranging from the reconstruction of species phylogenies to tracing epidemic outbreaks. Improvements on genome assembly using machine learning techniques are proposed by Padovani De Souza et al. [129]. Finally, in order to better use all the information acquired by high-throughput real-time sequencing and its analysis, text mining and knowledge discovery techniques, integrated with medical and scientific literature and gene family and metabolic pathway databases, could help generate new insights and speed up discoveries. High-throughput real-time next-generation sequencing projects have transformed the field of bioinformatics from single-genome studies to pan-genome analyses. The limiting factor now is no longer data rarity, but immense data availability and dimensionality. In this new context, bottom-up analyses stemming from big data provide great challenges and also great rewards.

4 Pan-genomics of plants and its applications

The plants genomes are highly dynamic as compared to many higher eukaryotes due to the presence of transposable elements and frequent genome duplication events [130]. Thus, the identification of such structural variations and dynamics in plant genomes is a prerequisite for subsequent understanding and their applications based on the sequence-trait associations. Several plant genomes were sequenced during the sequencing initiative in 2000 allowing an assembly of their reference genomes [131]. These reference genomes were mainly used to compare genomes of different plant species and to identify the SNPs across populations [132]. These studies increased our understanding regarding the allelic variations associated with phenotypic outcomes in general. However, such studies were not able to capture fully the diversity of sequence variations in plant genomes being themselves dependent on large genetic variations within strains/species. To this end, the advent of high throughput sequencing has played a major role in examining the genetic variations including SNPs, CNV, and presence/absence variations (PAV) comprehensively. The reduced costs of high-throughput sequencing methods have now revolutionized the ways being used for the analyses of plant genomes previously and for asking relevant biological questions. It has made it possible to easily sequence and compare the whole genomes of many individuals of same plants species and thus capturing the interspecies genetic diversity. Accordingly, the full genome content capturing the interspecies genomic diversity is termed as pan-genome [133]. The pan-genome approach allows to predict the number of additional genome sequences

that are necessary to characterize fully the genomic diversity of a species [133]. Analyses of pangenomes of several plants have now revealed the role of structural variations in different plant phenotypes such as flowering times, different stress-resistant mechanisms, etc. [134]. These studies have enhanced our understanding of the diverse applications of these genotypic to phenotypic association such as for increasing the crop production of better varieties in terms of size and flavors, increasing the abiotic stress and pathogens/disease resistances among many others reviewed in this chapter. The pan-genome approach is especially suitable for plant-breeding applications in contrast to the single liner reference genomes because of reduced sampling biases along with the comprehensive representation of genetic diversity [133]. The field of pan-genomics is rapidly evolving based on the underlying sequencing paradigms and the analytical pipelines, tools, and algorithms for sequencing data. The current pangenome assembly approaches can be categorized into a k-mer-based approach, comparative de-novo assembly approach, and iterative assembly approach. One of the challenges associated with the analysis of pan-genome data is related to requiring the increase in precision of the underlying genome assembly approaches. This review chapter aims to describe comprehensively the structural variations in plants genomes, explain the concept of pangenome, and its characterization along with the applications, methods, and approaches to conduct pan-genome analyses for a wide range of plant species.

4.1 Applications of pan-genomics in plant pathogens

The knowledge of plant diseases and host-pathogen interactions is one of the fundamental and active areas of genetic research with a wide array of applications [135]. Previously, linear reference genomes have been widely used for the subsequent analyses of phylogenetic relationships, identification of casual agents, virulence factors, host specificity associations, and pathogenic mechanisms [136]. These studies aided better disease management for economically important crops and plants by counteracting the stress-based resistance factors and better vaccine development. However, there is increasing evidence that the single reference genomes are insufficient in capturing the entire genetic diversity of the strains and subsequent delineation of principles governing the adaptive success of plant pathogens along with the determination of pathogenicity factors [137]. Accordingly, the concept of pan-genome emerged to cater to the interstrain genetic diversity based on different structural variations including CNV, presence/absence variations (PAV), and other allelic transformations. Pan-genome approach is now emerging as an analytical approach for analyzing the genetic diversity of genomes at an unprecedented level of details in contrast to the single reference genome. The strain-specific genome content is especially beneficial for gaining insights into the pathogenic mechanisms of plant pathogens as most of the pathogenic determinants are often strain specific and highly variable. Moreover, the pan-genome analysis allows determining the

genome plasticity through studying the evolutionary impact of HGT. As of yet, pan-genome analyses have already been used to identify and detect new strains along with development of vaccines against many plant pathogens [138]. Several computational pipelines based on tools and software especially designed to conduct a pan-genome analysis are available now. These tools can perform several functions including homologous gene clustering, SNPs identification, pan-genomic profiles visualization, phylogenetic analysis based on orthologous genes or gene families based information, pan-genome visualization, curation, and function-based searching. Most of the established pan-genome analysis methods were initially developed to deal with smaller prokaryotic genomes and thus are beneficial in analyzing most of the plant pathogens including bacteria and fungi. However, there are still certain challenges in assembling and analyzing the pangenomes of the species with complex genome structures [32]. Despite this, the pan-genome analyses is emerging as an important research tool to enhance our understanding about host-pathogen interactions and to develop universal vaccines. Since this approach has a potential for organizing pathogenic diversity, integrating pan-genomics with phylogeny and phylogenomics will be an interesting viewpoint for the future. Overall, we have comprehensively reviewed the studies conducted to assemble the pan-genomes of plant pathogens, its applications, available methods, and tools to conduct a pan-genome analysis in our chapter.

5 Genomics of algae and its applications

Genome sequencing unveils the basis of various fundamental processes and origin as well as the evolution of the organism. Advancement in whole-genome sequencing in the field of algal biomass has answered our queries of ecological and economic importance extending from the adaptation of organisms in diverse environments to synthesizing abundant metabolites of vast economical future. WGS of diverse algal genome has been performed using sequencing approaches ranging from shortgun to high throughput. Shortgun approach includes cloning 1–10 kb g-DNA fragments into pUC18 or pBluescript II KS (Stratagene). Plasmids have been sequenced using PE BigDye Terminator/ET DYEnamic terminator kit. Sequences have been resolved using PE 377 Automated DNA Sequencers and assembled from end sequences using PHRAP (P. Green) and Consed. Primer walking has been used for gap filling. Glimmer, GeneMarks, and Critica have been used to identify ORFs in the genome. High-throughput sequencing technologies include Illumina HiSeq 2000 technology, Illumina GA II x and Solexa Genome Analyzer (Illumina) and paired reads have been assembled using a DeBruijn method or CLC Genomics Workbench tools.

This development has also initiated metagenomics and metatranscriptomics, maneuvering the expression analysis and functional assays to study intraspecies and interspecies variability among nonmodel and complex biological communities of worth.

Comparative genomics is another approach to identify the essential mechanisms of origin and evolution. Genome analysis showed that a cyanobacterium *Synechococcus* sp. strain WH8102 is nutritionally more adaptable as it has acquired more sodium-dependent transporters for the uptake of organic nitrogen and phosphorus. Reduced gene complement in marine cyanobacterium *P. marinus* SS120 is consistent with the fact that the oligotrophic marine environment where it preferentially thrives is much more stable than freshwaters [139]. There are also examples from other algal genome analysis that unveiled the adaptation strategies to thrive under harsh conditions such as *Ostreococcus tauri* that has adapted costly C4 photosynthetic pathway to acquire critical ecological advantage in the CO_2-limiting conditions of phytoplankton blooms, green alga *Chloroidium* sp. UTEX 3007 is able to survive high temperatures in deserts by accumulation of thermostable palmitic acid [140]. Also, an acidophilic green alga *Chlamydomonas eustigma* NIES-2499 has acquired phytochelatin synthase genes providing it tolerance to toxic metal ions such as cadmium [141]. *Galdieria sulphuraria* and *C. merolae* belong to the Cyanidiophyceae group but at the same time possess many contrasting features. The foremost is the ability of *G. sulphuraria* to adapt to extreme acidic thermophilic environments. It is the only alga in this group with an adaptation of the heterotrophic mode of nutrition with multiple substrates, which indicates how it survives in harsh environments [142]. In the process of evolution of ancestral lineages of red algae, the role of HGT is undeniable. This was indicated in the genome of other red algae, *Porphyridium purpureum*. Along with that, several light-harvesting complexes (LHC) were identified. Genomic analysis revealed evidence for sexual reproduction [143]. To cope with ecological stress, the genome of *P. umbilicalis* reveals the presence of genes coding for high-affinity iron transport complex necessary for the iron uptake processes to obtain nutrients during stressful high tides [144]. The study of gene sequences has also thrown light on the conservation of certain key enzymes such as GDP-mannose 6-dehydrogenase (GMD) required in the process of synthesis of alginates in brown algae *Cladosiphon okamuranus*. Also, *C. okamuranus* holds significant commercial importance as it is cultivated for fucoidan, which is a sulfated polysaccharide, a kind of Japanese seaweed [145]. The information on genomics has opened doors to various other research fields like proteomics, expression analysis, structural biology, metabolomics, etc.

6 Pan-metagenomics and human microbiome

Pan-metagenome is the collective study of all or several metagenomes from all possible units belonging to a particular type of ecosystem or host.

In the past decade, most of the metagenomic studies have aimed at understanding the microbial community from a relatively small set of samples. Such studies could miss out important rare taxa. However, the reduction in cost of gigabyte of NGS data has made the NGS application affordable and widespread [146]. This has given rise to an enormous

amount of publicly accessible data from various types of samples. The application of pan-metagenome ranges from the mosquito gut microbiome [147] to human gut microbiome [148], including various ecosystems [149, 150]. Pan-metagenome primarily aims to explore and redefine the microbial community at a global scale. This will help to capture all the taxonomical variations between samples and understand the shifts in microbial community on a larger scale.

A pan-metagenome comprising thousands of samples pertaining to an ecosystem or host from multiple locations and studies at global level collaborations could be used as a standard reference. Such a reference-based pan-metagenome could serve as a guideline to answer several questions: What types of ecosystems are most vulnerable to global warming? Are rare taxa distributed based on geography?

7 Pan-proteomics and its applications

In the proteomic approach it is possible to identify and quantify a set of proteins synthetized by a determined cell, tissue, or microorganism [151] when exposed to different experimental conditions (such as temperature, osmolarity, antibiotics, nitric oxide, and others), or different steps of the cell growth, or during infection process [151–153]. At a specific condition, the identified proteins from the complex protein mixtures may be characterized in relation to their expression, cellular localization, structure, biological functions, and interactions with other proteins, posttranslation modifications, and metabolic pathways. In this way, proteomic studies contribute to understanding about cellular adaptation in response to external changes, metabolic stresses, or infection, and this response can vary according to time and environment [154], The proteomic analysis have been considered the most relevant approach to describe a biological system [151].

Proteomic approach in eukaryotic cells is relatively complex due to posttranslational modification, like phosphorylation of proteins, which is involved in protein signaling in different cellular pathways [155]. In humans, datasets from proteome studies have allowed to evaluate the potential methods in diagnosis, prognosis, and treatment for some diseases, including cancer [156]. On the other hand, in prokaryotes the proteomic assays have enabled the investigation of physiological behaviors, mutations, adaptability to different environmental conditions, presence of proteins involved in virulence, and the identification of putative immunogenic proteins [157].

The protein synthesis in eukaryotic and prokaryotic calls can be evaluated by different technologies, such as chromatography-based methods, enzyme-linked immunosorbent assay (ELISA), Western blotting, protein separation using gel-based approaches, especially two-dimensional (2D) polyacrylamide gel electrophoresis, or through the identification and sequencing of polypeptides through mass spectrometry technologies [151]. In chromatography-based techniques, the proteins can be obtained from separation based

on their charge nature and charge strength (ion exchange chromatography), molecular size (size exclusion chromatography), or specificity (affinity chromatography) [158]. On the other hand, ELISA uses antibodies or antigens on the solid surface to detect specific peptides or enzymes from the biological sample, forming enzyme-conjugated antibodies which allow to measure the enzyme activity or protein concentration [159]. Last, Western blotting enables the identification of low abundance proteins after electrophoresis separation, transfer onto nitrocellulose membrane, and detection by enzyme-conjugated antibodies [160]. Nevertheless, these three methodologies allow to evaluate few proteins, and they are unable to determine protein expression level [151]. 2D gel electrophoresis is an efficient and widely used technique in proteomic studies to analyze complex protein mixtures extracted especially from bacterial cells. This methodology involves separation of proteins by isoelectric focusing (proteins with different isoelectric points) and by molecular weight (in polyacrylamide gel electrophoresis). Each spot in a 2D matrix corresponds to a single protein in the sample evaluated. In this way, 2D gel electrophoresis allows to obtain information of several proteins simultaneously as apparent molecular weight, isoelectric point, and quantity of each one [161]. And, mass spectrometry can be defined as the study of matter through the formation of ions in the gas phase and their characterization by mass, charge, structure, or physicochemical properties, using mass spectrometer that measures m/z values and abundance of ions [162].

The association between 2D gel electrophoresis and mass spectrometry was already considered the most appropriate method to recognize and identify proteins from pathogenic microorganisms [163] for being a methodology used for the construction of proteomic databases, due to its greater efficiency and high resolution to investigate the complex mixtures of proteins present in cell or tissues [164]. Nevertheless, with the technical advances achieved in recent years, such as solubilization of complex samples, pH gradient, and detection of proteins present in small quantities, the technique of liquid chromatography associated with mass spectrometry (LC-MS) started to be used and allowed the analysis of complex mixtures of proteins by tryptic digestion without prior gel separation [165]. This technique had the advantage of having a low detection limit for peptides and proteins, capability to identify hundreds to thousands of proteins in a simple experiment as well as allowing the study of membrane proteins, poorly accessible by other methods [166].

LC-MS is divided into two approaches: stable isotopic labeling [167] and label-free quantification [168]. In the first, two solutions containing the proteins to be analyzed are labeled with different molecular mass isotopes, and are mixed, trypsin-digested to obtain peptides and submitted to the LC-MS system [169]. The molecular weight difference allows the identification and quantification of peptides of both samples tested [170], but the labeling occurs after the extraction step, which can lead to a reduction in the precision of the quantification method [171]. Alternatively, label-free quantification allows the evaluation of numerous samples at the same time within the LC-MS system, with

data-independent acquisition, and the concentration of a given peptide is proportional to its chromatographic area [172].

Among the strategies used in proteomic studies in prokaryotic cells surfome and secretome analyses stand out. The bacterial surface has been considered of great importance for understanding the pathogenesis of an infectious disease. On the surface, it can be found that proteins are associated with mechanisms of defense and virulence factors, which can promote adhesion and cellular invasion, culminating consequently in the appearance of clinical signs in an infected host [173]. Therefore, surfome is a proteome-based method, in which allows the identification of bacterial surface proteins [174]. Apart from surface proteins, extracellular and secreted proteins are important in bacterial pathogenesis, since they also mediate the interaction of the bacterium with the host and by stimulating the immune response. Therefore, the secretome has been associated with adhesion, invasion, immune evasion, and spread of bacterium in host tissues. In addition, these proteins can also be used for the development of antibiotics and vaccines [175]. Besides these two methods, comparative proteome analysis has been used for both prokaryotic and eukaryotic cells. This method has also been used to identify virulence factors and to obtain information on physiological and environmental adaptations in different pathogens [176], as well as to compare cells, tissues, and organs from the eukaryotic host in normal and pathological (inflammation, infection, and cancer) conditions [156].

In this context, pan-proteomics is also an approach with characterizes and compares the qualitative and quantitative proteome; however, the comparison occurs across organisms inside a species, with genetic variation and phenotype [177]. Pan-proteomics can be performed using 2D gel electrophoresis or LC-MS; nevertheless, LC-MS by bottom-up/shotgun techniques, from our expertise, is recommended for this type of study, otherwise, we will always have only part of the proteome and not the whole proteome.

Conceptually, pan-proteome refers to the proteins identified from a whole set of samples/strains tested, which are usually more than two samples, under the same experimental conditions. The analysis of two samples is equivalent to comparative proteomic methodology. Pan-proteome can be divided into core proteome, accessory proteome, and orphan (or unique) proteome [177]. The core proteome represents the subset of identified proteins simultaneously in all samples, whereas accessory proteome represents the detected proteins shared by at least two samples, and orphan proteome represents proteins identified exclusively in a single sample.

In the microbiological field, the genetic variation among isolates has been implicated with virulence factors, drug resistance, and environmental adaptation [178]. In this way, understanding about these mechanisms needs the evaluation of several proteomes and not from single proteome analysis [177]. Thus, pan-proteomic analysis may increase knowledge about the adaptation and pathogenicity of a given microorganism, independent of

the genotype. Besides that, this approach can be used to classify bacterial strains in types [179], identify putative vaccine targets from conserved proteins among isolates [178], as well as, to determine drug targets and drug mode of action in analysis with multiple strains [177].

The term pan-proteome and core proteome have been used in different studies of protein identification and quantitation. In this type of study, pan-proteome and core proteome were referenced in the first time from analysis of four epidemic *Salmonella* Paratyphi A strains, with different PFGE types, using 2D gel electrophoresis [180]. From this analysis, the authors verified a high covered (over than 81%) of core proteome among the isolates tested, regardless of the range of pH applied, suggesting a high similarity in protein expression. Proteins involved in metabolic pathways and survival of the bacterium were the most identified within the core proteome. Moreover, the proteome comparison among isolates suggest a geospatial and temporal differentiation of expressed protein profile (spots).

The conserved core proteome was also observed in other works, where this category represented approximately 92% of pan-proteome of five fish-adapted *Streptococcus agalactiae* strains, which belonged to three MLST profiles. This study was performed using a label-free proteomic analysis [178]. The authors suggest that the identified proteins reflect an adaptation to an aquatic environment and fish-pathogen interaction. In addition, in the same study, conserved antigenic proteins were identified and suggested as targets in vaccine design, seeing that the high degree of conservation of these proteins among the isolates would suggest the production of a monovalent vaccine effective against all genetic variants tested.

Another study, despite the conservation of proteins identified simultaneously in avirulent, virulent, and two clinical strains of *Mycobacterium tuberculosis*, the quantitative protein expression profiling revealed a strain-specific variation in proteome patterns of isolates [181]. This study was also performed using label-free analysis, being identified 257 differentially expressed proteins. The differences in virulence among four isolates were suggested to a two-component system, oxidative stress, ribosome biogenesis, energy generation, and transcriptional regulator proteins.

The pan-proteomic analysis of four biotechnological *Lactococcus lactis* strains was performed using label-free analysis and showed a conservation of 52% of core proteome. The identified proteins contribute to physiological adaptation of bacteria, metabolic pathways, microbial metabolism in diverse environment, and proteins involved in post-translational modification, which enable maintenance of cellular integrity and physiological process bacterial during adverse environmental conditions, like temperature and oxidative stress. In this way, the authors suggested that with the results found it would be possible to increase the biotechnological potential of *L. lactis* [182].

On the other hand, in eukaryotic cells, the term pan and core proteome was used in a comparative proteomic analysis of *Gammarus* female reproductive systems (ovaries).

Nevertheless, in this study the authors verified a core proteome relatively low among the three amphipods belonging to *Gammarus* genus [183], identifying proteins involved in cellular process, localization, catalytic activity, and binding. Nevertheless, proteins involved in reproductive process were little found due to the absence of their sequences in the database used.

For the success of pan-proteomic experiments, it is necessary to be attentive as to: sample preparation, being important an optimization of the protocols of protein extraction from the multistrain or multiclinical samples; types of data acquisition from gel-based or gel-free methods; construction of pan-proteome database containing all possible proteins, including the same protein but with sequence variation, to use during searching for peptide identification; and better understand the biological functions of the identified proteins through bioinformatics analysis. All these points were extensively revised in a previous study [177].

8 Pan-transcriptomics and its applications

Transcriptome profiling is a powerful approach to identify and quantify the entire repertoire of transcripts in a cell, including mRNAs, noncoding RNAs, and small RNAs, during specific developmental stages or conditions [184]. Transcriptome analysis has enabled the study of the functional elements of the genome, increasing our understanding of the transcriptional dynamics of biological processes and disease development [185]. Among the various technologies that have been developed for high-throughput transcriptome analyses, microarray and RNA-seq are at the forefront of large-scale genome transcriptome profiling [186]. Microarray is a hybridization-based approach developed in the mid-1990s that measure the abundance of a known set of genes using an array of complementary probes. Microarray is a cost-effective, easy to analyze approach that remains the most extensively used methodology in the scientific community. RNA microarrays are generated using complementary DNA (cDNA) immobilized on a glass slide, where each cDNA fragment represents an individual gene of interest. RNA arrays have been used to identify regulated genes, pathways, networks, biological mechanisms, and processes in a variety of biological conditions [187].

However, since its commercial availability, RNA-seq has been widely applied to identify genes within a genome or to measure the expression of transcripts in an organism in different tissues, conditions, and time points [188]. RNA-seq has many advantages over array-based technology, including a high level of data reproducibility, detection of low abundant transcripts, and identification of isoforms over a wider dynamic range. Moreover, the technology does not depend on existing genome data or annotation, allowing the identification and quantification of novel transcripts [189]. Generating data on RNA transcripts require RNA to be first isolated from the experimental organism, following synthesis of cDNA, PCR amplification of cDNA transcripts, and deep sequencing [188].

Following the increased number of high-throughput RNA data, a wide range of strategies for transcriptome analysis has emerged, ranging from single cell to comparative pan-transcriptomic analysis. The pan-transcriptomics analysis consists of a comparison between complete sets of RNA transcripts, under specific circumstances, aiming to identify genes that are differentially expressed in distinct or related populations, or in response to different treatments to better understand the functional and structural aspect of genes. The integration and collective analysis of transcriptome data has enabled the identification of core and distinct molecular responses that functionally reflect the phenotypical diversity of a specific group or condition including patterns of expression associated with parasitism [190], construction of co-expression networks of differentially expressed genes encoding virulence factors [191], the identification of universal biomarkers of cellular senescence [192], comprehensive analysis of molecular alterations across multiple cancer types [193], and the characterization of tissue-specific expression of long noncoding RNAs (lncRNAs) [194]. Pan-transcriptome analysis is particularly applicable in prokaryotes and has been proven valuable in shedding light on gene expression and transcriptome organization among bacterial groups where the difference in phenotypes cannot be explained by the genome sequences alone [195] (Table 3). Moreover, a comparative approach using high-throughput studies can also show the molecular basis of pathogenicity, orthologous biological features, virulence factors, and signaling pathways responsible for stress tolerance and pathogen resistance of related surrogate bacterial species as well as within larger groups of the bacterial domain (Table 3). In addition, integrated analysis can aid the search for potential targets that can be used in the development of therapeutic strategies against relevant pathogens.

Table 3 Pan-transcriptome studies in prokaryotes

Species	Strains/isolates	Approach	Conditions/remarks	References
Mycobacterium tuberculosis and *Mycobacterium bovis*	Mtb H37Rv Mtb H37Rv Mtb H37Rv Mbovis AF2122/97 Mbovis AF2122/97 Mtb H37Rv	Microarray	Bacterial response to aerobic chemostat, low oxygen chemostat—0.2% DOT, aerobic rolling, batch culture, aerobic chemostat, aerobic rolling batch culture, harvested from macrophages	[196]
Bacillus subtilis	BR16 BR17 16BCE	Microarray	Bacterial stringent response by mimicking isoleucine and leucine starvation	[197]

Continued

Table 3 Pan-transcriptome studies in prokaryotes—cont'd

Species	Strains/isolates	Approach	Conditions/remarks	References
Acinetobacter baumannii		RNA-seq	Dynamics of gene expression in the transcriptomic response of drug resistance multidrug-resistant strains and sensitive strains	[198]
Campylobacter jejuni	NCTC11168 81–176 81,116 RM1221	RNA-seq	Comparative analysis of regulatory elements between four isolates	[195]
Pseudomonas aeruginosa	PA14	RNA-seq	Identification of phenotypic variability among bacteria dependent on gene expression in response to different environments including growth within biofilms, at various temperatures, growth phases, osmolarities, phosphate, and iron concentrations, under anaerobic conditions, attached to a surface, and conditions encountered within the eukaryotic host	[199]
Mycobacterium tuberculosis	TKK-01-0084 TKK-01-0025 TKK-01-0033 TKK-01-0040	RNA-seq	Identification of novel transcriptional mechanisms of drug resistance in Mtb strains	[200]
Escherichia coli	EPEC1 EPEC5 EPEC7	RNA-seq	Investigate the global transcriptional responses of the enteropathogenic *E. coli* (EPEC) and enterotoxigenic *E. coli* (ETEC) using 7 isolates	[201]

9 Pan-cancer analysis and its applications

Pan-cancer analysis has enabled in identifying the molecular aspects underlying cancer thereby benefiting diagnosis, prevention, and therapy for patients. One of the major applications of the pan-cancer data is for drug development by ranking drug targets that can be further exploited to develop targeted therapies for cancer. Further analysis of the data is needed for understanding gene-gene interactions and roles of genetic variants affecting pathways. Extensive research has been done to elucidate the underlying mechanisms of cancer occurrence and progression [202–204]. However, most of the studies are conducted independently on smaller sample sizes, thereby limiting the essence of information that needs to come out of such studies. The numerous projects involved in pan-cancer analysis generated huge volumes of data using various technologies including high-end molecular genetics and cytogenetics techniques. Various web tools have been developed and used to interpret the large amount of data generated by the pan-cancer projects [205]. The International Cancer Genome Consortium hence made a group of researchers conducting such cancer analysis across various tumor types in order to generate a pan-cancer atlas [206]. Data generated through these projects will enable in understanding the molecular aspects of cancer occurrence and further help in cancer prevention and designing cancer therapeutics. There are certain challenges that need to be overcome for the development of clinical trial strategies to connect tumor subsets from diverse tissue types [207].

10 Conclusions

The emergence of NGS technologies and the use of the data generated by these technologies for comparative genomics is a major advancement in understanding the diversity of genomes. There are effective examples of pan-genomic studies in various fields of research. The concept of pan-genomics is so deep that it has been perfectly applied in the studies of several organisms and diseases, for example, in the study of dynamics of biological processes and disease development, identification of therapeutic targets against deadly and emerging pathogens, and in the development of new probiotics. It has great potential, which may bring a closer understanding and help combat prokaryotic and eukaryotic diseases in a better way. Finally, several other fields of research that use pan-genomic idea exist, such as pan-cancer, pan-genomics of plants, virus and fungi, pan-metabolomics, and others. All those fields will be further discussed in the following chapters.

References

[1] J.M. Heather, B. Chain, The sequence of sequencers: the history of sequencing DNA, Genomics 107 (2016) 1–8.
[2] E.S. Donkor, Sequencing of bacterial genomes: principles and insights into pathogenesis and development of antibiotics, Genes (Basel) 4 (2013) 556–572.

[3] M. Land, L. Hauser, S.R. Jun, I. Nookaew, M.R. Leuze, T.H. Ahn, T. Karpinets, O. Lund, G. Kora, T. Wassenaar, S. Poudel, D.W. Ussery, Insights from 20 years of bacterial genome sequencing, Funct. Integr. Genom. 15 (2015) 141–161.
[4] J.W. Prokop, T. May, K. Strong, S.M. Bilinovich, C. Bupp, S. Rajasekaran, E.A. Worthey, J. Lazar, Genome sequencing in the clinic: the past, present, and future of genomic medicine, Physiol. Genom. 50 (2018) 563–579.
[5] J. Zhang, R. Chiodini, A. Badr, G. Zhang, The impact of next-generation sequencing on genomics, J. Genet. Genom. 38 (2011) 95–109.
[6] H. Tettelin, D. Riley, C. Cattuto, D. Medini, Comparative genomics: the bacterial pan-genome, Curr. Opin. Microbiol. 11 (2008) 472–477.
[7] D. Medini, C. Donati, H. Tettelin, V. Masignani, R. Rappuoli, The microbial pan-genome, Curr. Opin. Genet. Dev. 15 (2005) 589–594.
[8] A.J. Van Tonder, S. Mistry, J.E. Bray, D.M.C. Hill, A.J. Cody, C.L. Farmer, K.P. Klugman, A. Von Gottberg, S.D. Bentley, J. Parkhill, K.A. Jolley, M.C.J. Maiden, A.B. Brueggemann, Defining the estimated core genome of bacterial populations using a Bayesian decision model. PLoS Comput. Biol. 10 (8) (2014) e1003788 https://doi.org/10.1371/journal.pcbi.1003788.
[9] L. Rouli, V. Merhej, P.E. Fournier, D. Raoult, The bacterial pangenome as a new tool for analysing pathogenic bacteria, New Microbes New Infect. 7 (2015) 72–85.
[10] H. Tettelin, V. Masignani, M.J. Cieslewicz, C. Donati, D. Medini, N.L. Ward, S.V. Angiuoli, J. Crabtree, A.L. Jones, A.S. Durkin, R.T. Deboy, T.M. Davidsen, M. Mora, M. Scarselli, Y. Margarit, I. Ros, J.D. Peterson, C.R. Hauser, J.P. Sundaram, W.C. Nelson, R. Madupu, L.M. Brinkac, R.J. Dodson, M.J. Rosovitz, S.A. Sullivan, S.C. Daugherty, D.H. Haft, J. Selengut, M.L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K.J. O'connor, S. Smith, T.R. Utterback, O. White, C.E. Rubens, G. Grandi, L.C. Madoff, D.L. Kasper, J.L. Telford, M.R. Wessels, R. Rappuoli, C.M. Fraser, Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome" Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 13950–13955.
[11] S.C. Soares, V.A. Abreu, R.T. Ramos, L. Cerdeira, A. Silva, J. Baumbach, E. Trost, A. Tauch, R. Hirata Jr., A.L. Mattos-Guaraldi, A. Miyoshi, V. Azevedo, PIPS: pathogenicity island prediction software, PLoS ONE 7 (2012) e30848.
[12] S.C. Soares, H. Geyik, R.T. Ramos, P.H. De Sa, E.G. Barbosa, J. Baumbach, H.C. Figueiredo, A. Miyoshi, A. Tauch, A. Silva, V. Azevedo, GIPSy: genomic island prediction software, J. Biotechnol. 232 (2016) 2–11.
[13] M. De Barsy, A. Frandi, G. Panis, L. Theraulaz, T. Pillonel, G. Greub, P.H. Viollier, Regulatory (pan-) genome of an obligate intracellular pathogen in the PVC superphylum, ISME J. 10 (2016) 2129–2144.
[14] Y. Zhao, J. Wu, J. Yang, S. Sun, J. Xiao, J. Yu, PGAP: pan-genomes analysis pipeline, Bioinformatics 28 (2012) 416–418.
[15] A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T. Holden, M. Fookes, D. Falush, J.A. Keane, J. Parkhill, Roary: rapid large-scale prokaryote pan genome analysis, Bioinformatics 31 (2015) 3691–3693.
[16] N.M. Chaudhari, V.K. Gupta, C. Dutta, BPGA- an ultra-fast pan-genome analysis pipeline, Sci. Rep. 6 (2016) 24373.
[17] C. Computational Pan-Genomics, Computational pan-genomics: status, promises and challenges, Brief. Bioinform. 19 (2018) 118–135.
[18] Computational Pan-Genomics Consortium, Computational pan-genomics: status, promises and challenges. Brief. Bioinform. 19 (1) (2018) 118–135, https://doi.org/10.1093/bib/bbw089.
[19] B. Hurgobin, D. Edwards, SNP discovery using a pangenome: has the single reference approach become obsolete? Biology (Basel) 6 (1) (2017) pii: E21.
[20] L. Benevides, S. Burman, R. Martin, V. Robert, M. Thomas, S. Miquel, F. Chain, H. Sokol, L.G. Bermudez-Humaran, M. Morrison, P. Langella, V.A. Azevedo, J.M. Chatel, S. Soares, New insights into the diversity of the genus *Faecalibacterium*, Front. Microbiol. 8 (2017) 1790.
[21] Y. Chen, Y. Luo, H. Carleton, R. Timme, D. Melka, T. Muruvanda, C. Wang, G. Kastanis, L.S. Katz, L. Turner, A. Fritzinger, T. Moore, R. Stones, J. Blankenship, M. Salter, M. Parish,

T.S. Hammack, P.S. Evans, C.L. Tarr, M.W. Allard, E.A. Strain, E.W. Brown, Whole genome and core genome multilocus sequence typing and single nucleotide polymorphism analyses of *Listeria monocytogenes* associated with an outbreak linked to cheese, United States, 2013. Appl. Environ. Microbiol. 83 (15) (2017) e00633–17 https://doi.org/10.1128/AEM.00633-17.

[22] G. Vernikos, D. Medini, D.R. Riley, H. Tettelin, Ten years of pan-genome analyses, Curr. Opin. Microbiol. 23 (2015) 148–154.

[23] H. Tettelin, The bacterial pan-genome and reverse vaccinology, Genome Dyn. 6 (2009) 35–47.

[24] O. Lukjancenko, T.M. Wassenaar, D.W. Ussery, Comparison of 61 sequenced *Escherichia coli* genomes, Microb. Ecol. 60 (2010) 708–720.

[25] V. Periwal, A. Patowary, S.K. Vellarikkal, A. Gupta, M. Singh, A. Mittal, S. Jeyapaul, R.K. Chauhan, A.V. Singh, P.K. Singh, P. Garg, V.M. Katoch, K. Katoch, D.S. Chauhan, S. Sivasubbu, V. Scaria, Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of *Mycobacterium tuberculosis* pangenome, PLoS ONE 10 (2015) e0122979.

[26] M.W. Tiwari, Diphtheria toxoid, in: Plotkin's Vaccines, seventh ed., Elsevier, 2017.

[27] M. Hessling, J. Feiertag, K. Hoenes, Pathogens provoking most deaths worldwide: A review, Biosci. Biotechnol. Res. Commun. 10 (2017) 1–7.

[28] E. Hacker, C.A. Antunes, A.L. Mattos-Guaraldi, A. Burkovski, A. Tauch, Corynebacterium ulcerans, an emerging human pathogen, Future Microbiol. 11 (2016) 1191–1208.

[29] A. Burkovski, Pathogenesis of *Corynebacterium diphtheriae* and *Corynebacterium ulcerans*, in: Human Emerging and Re-emerging Infections, Wiley, 2015, pp. 699–709 Print ISBN: 9781118644713, Online ISBN: 9781118644843.

[30] A.M. Cerdeno-Tarraga, A. Efstratiou, L.G. Dover, M.T. Holden, M. Pallen, S.D. Bentley, G.S. Besra, C. Churcher, K.D. James, A. De Zoysa, T. Chillingworth, A. Cronin, L. Dowd, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Moule, M.A. Quail, E. Rabbinowitsch, K.M. Rutherford, N.R. Thomson, L. Unwin, S. Whitehead, B.G. Barrell, J. Parkhill, The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129, Nucleic Acids Res. 31 (2003) 6516–6523.

[31] P. Lapierre, J.P. Gogarten, Estimating the size of the bacterial pan-genome, Trends Genet. 25 (2009) 107–110.

[32] J. Xiao, Z. Zhang, J. Wu, J. Yu, A brief review of software tools for pangenomics, Genom. Proteom. Bioinform. 13 (2015) 73–76.

[33] G.D. Wright, The antibiotic resistome: the nexus of chemical and genetic diversity, Nat. Rev. Microbiol. 5 (2007) 175–186.

[34] M.R. Gillings, Evolutionary consequences of antibiotic use for the resistome, mobilome and microbial pangenome, Front. Microbiol. 4 (2013) 4.

[35] M.L. Metzker, Sequencing technologies—the next generation, Nat. Rev. Genet. 11 (2010) 31–46.

[36] S. Ghatak, J. Blom, S. Das, R. Sanjukta, K. Puro, M. Mawlong, I. Shakuntala, A. Sen, A. Goesmann, A. Kumar, S.V. Ngachan, Pan-genome analysis of Aeromonas hydrophila, Aeromonas veronii and Aeromonas caviae indicates phylogenomic diversity and greater pathogenic potential for Aeromonas hydrophila, Antonie Van Leeuwenhoek 109 (2016) 945–956.

[37] S.C. Bayliss, D.W. Verner-Jeffreys, K.L. Bartie, D.M. Aanensen, S.K. Sheppard, A. Adams, E.J. Feil, The promise of whole genome pathogen sequencing for the molecular epidemiology of emerging aquaculture pathogens, Front. Microbiol. 8 (2017) 121.

[38] T.L. Nguyen, D.-H. Kim, Genome-wide comparison reveals a probiotic strain *Lactococcus lactis* WFLU12 isolated from the gastrointestinal tract of olive flounder (*Paralichthys olivaceus*) harboring genes supporting probiotic action, Mar. Drugs 16 (5) (2018) pii: E140.

[39] M. Dalsass, A. Brozzi, D. Medini, R. Rappuoli, Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery, Front. Immunol. 10 (2019) 113.

[40] Y. Sun, C.S. Liu, L. Sun, Construction and analysis of the immune effect of an Edwardsiella tarda DNA vaccine encoding a D15-like surface antigen, Fish Shellfish Immunol 30 (2011) 273–279.

[41] M.Y. Abdelgayed, Y.G. Alkhateib, A.M. Laila, S.Z. Mona, DNA-based vaccines against bacterial fish diseases: trials and prospective, Rep. Opinion 9 (2017) 1–16.

[42] L. Zeng, D. Wang, N. Hu, Q. Zhu, K. Chen, K. Dong, Y. Zhang, Y. Yao, X. Guo, Y.F. Chang, Y. Zhu, A novel pan-genome reverse vaccinology approach employing a negative-selection strategy for screening surface-exposed antigens against leptospirosis, Front. Microbiol. 8 (2017) 396.

[43] Z. Golkar, O. Bagasra, D.G. Pace, Bacteriophage therapy: a potential solution for the antibiotic resistance crisis, J. Infect. Dev. Ctries. 8 (2014) 129–136.
[44] C.L. Ventola, The antibiotic resistance crisis: part 1: causes and threats, P T 40 (2015) 277–283.
[45] P.C. Appelbaum, 2012 and beyond: potential for the start of a second pre-antibiotic era? J. Antimicrob. Chemother. 67 (2012) 2062–2068.
[46] R.J. Fair, Y. Tor, Antibiotics and bacterial resistance in the 21st century, Perspect. Medicin. Chem. 6 (2014) 25–64.
[47] B.D. Lushniak, Antibiotic resistance: a public health crisis, Public Health Rep. 129 (2014) 314–316.
[48] G.M. Rossolini, F. Arena, P. Pecile, S. Pollini, Update on the antibiotic resistance crisis, Curr. Opin. Pharmacol. 18 (2014) 56–60.
[49] B. Spellberg, D.N. Gilbert, The future of antibiotics and resistance: a tribute to a career of leadership by John Bartlett, Clin. Infect. Dis. 59 (Suppl 2) (2014) S71–S75.
[50] V.K. Viswanathan, Off-label abuse of antibiotics by bacteria, Gut Microbes 5 (2014) 3–4.
[51] C.A. Michael, D. Dominey-Howes, M. Labbate, The antimicrobial resistance crisis: causes, consequences, and management, Front. Public Health 2 (2014) 145.
[52] A. De Sarom, A. Kumar Jaiswal, S. Tiwari, L. De Castro Oliveira, D. Barh, V. Azevedo, C. Jose Oliveira, S. De Castro Soares, Putative vaccine candidates and drug targets identified by reverse vaccinology and subtractive genomics approaches to control Haemophilus ducreyi, the causative agent of chancroid, J. R. Soc. Interface 15 (142) (2018) 20180032.
[53] S.B. Jamal, S.S. Hassan, S. Tiwari, M.V. Viana, L.J. Benevides, A. Ullah, A.G. Turjanski, D. Barh, P. Ghosh, D.A. Costa, A. Silva, R. Rottger, J. Baumbach, Azevedo, V.a.C., An integrative in-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheriae, PLoS ONE 12 (2017) e0186401.
[54] A. Kumar Jaiswal, S. Tiwari, S.B. Jamal, D. Barh, V. Azevedo, S.C. Soares, An in silico identification of common putative vaccine candidates against *Treponema pallidum*: a reverse vaccinology and subtractive genomics based approach, Int. J. Mol. Sci. (2017) 18.
[55] C.D. Rinaudo, J.L. Telford, R. Rappuoli, K.L. Seib, Vaccinology in the genome era, J. Clin. Invest. 119 (2009) 2515–2525.
[56] T. Bhardwaj, P. Somvanshi, Pan-genome analysis of *Clostridium botulinum* reveals unique targets for drug development, Gene 623 (2017) 48–62.
[57] D. Barh, S. Tiwari, N. Jain, A. Ali, A.R. Santos, A.N. Misra, V. Azevedo, A. Kumar, In silico subtractive genomics for target identification in human bacterial pathogens, Drug Dev. Res. 72 (2011) 162–177.
[58] A. Praveena, R. Sindhuja, V. Anuradha, S.K.M. Habeeb, Putative drug target identification for *Chlamydia trachomatis*: an insilico proteome analysis, Int. J. Biomed. Res. 2 (2011) 151–160.
[59] D. Barh, A. Kumar, In silico identification of candidate drug and vaccine targets from various pathways in *Neisseria gonorrhoeae*, In Silico Biol. 9 (2009) 225–231.
[60] S. Madagi, V. Malipatil, Putative drug targets in Ureaplasma urealyticum serovar 10 str. ATCC 33699 by insilico genomics approach and virtual screening, Int. J. Pharma Bio Sci. 4 (2013) 8.
[61] A. Ali, A. Naz, S.C. Soares, M. Bakhtiar, S. Tiwari, S.S. Hassan, F. Hanan, R. Ramos, U. Pereira, D. Barh, H.C.P. Figueiredo, D.W. Ussery, A. Miyoshi, A. Silva, V. Azevedo, Pan-genome analysis of human gastric pathogen *H. pylori*: comparative genomics and pathogenomics approaches to identify regions associated with pathogenicity and prediction of potential core therapeutic targets, Biomed. Res. Int. 2015 (2015) 1–17.
[62] S.M. Asif, A. Asad, A. Faizan, M.S. Anjali, A. Arvind, K. Neelesh, K. Hirdesh, K. Sanjay, Dataset of potential targets for Mycobacterium tuberculosis H37Rv through comparative genome analysis, Bioinformation 4 (2009) 245–248.
[63] B. Rathi, A.N. Sarangi, N. Trivedi, Genome subtraction for novel target definition in *Salmonella typhi*, Bioinformation 4 (2009) 143–150.
[64] S.S. Hassan, S.B. Jamal, L.G. Radusky, S. Tiwari, A. Ullah, J. Ali, Behramand, P. De Carvalho, R. Shams, S. Khan, H.C.P. Figueiredo, D. Barh, P. Ghosh, A. Silva, J. Baumbach, R. Rottger, A.G. Turjanski, V.A.C. Azevedo, The druggable pocketome of *Corynebacterium diphtheriae*: a new approach for in silico putative druggable targets, Front. Genet. 9 (2018) 44.

[65] D. Barh, N. Jain, S. Tiwari, B.P. Parida, V. D'afonseca, L. Li, A. Ali, A.R. Santos, L.C. Guimaraes, S. De Castro Soares, A. Miyoshi, A. Bhattacharjee, A.N. Misra, A. Silva, A. Kumar, V. Azevedo, A novel comparative genomics analysis for common drug and vaccine targets in Corynebacterium pseudotuberculosis and other CMN group of human pathogens, Chem. Biol. Drug Des. 78 (2011) 73–84.
[66] S.S. Hassan, S. Tiwari, L.C. Guimaraes, S.B. Jamal, E. Folador, N.B. Sharma, S. De Castro Soares, S. Almeida, A. Ali, A. Islam, F.D. Povoa, V.A. De Abreu, N. Jain, A. Bhattacharya, L. Juneja, A. Miyoshi, A. Silva, D. Barh, A. Turjanski, V. Azevedo, R.S. Ferreira, Proteome scale comparative modeling for conserved drug and vaccine targets identification in Corynebacterium pseudotuberculosis, BMC Genom. 15 (Suppl 7) (2014) S3.
[67] D.J. Bibel, Elie Metchnikoff's Bacillus of Long Life, ASM News (1988) 661–665.
[68] A. Hosono, Fermented milk in the orient, in: Y. Naga Sawa, A. Hosono (Eds.), Functions of fermented milk. Challenges for the health sciences, 1992. Elsevier Applied Science.
[69] A.W. FAO, Guidelines for the Evaluation of Probiotics in Food, Food and Agriculture Organization of the United Nations, 2002.
[70] R. Bibiloni, R.N. Fedorak, G.W. Tannock, K.L. Madsen, P. Gionchetti, M. Campieri, C. De Simone, R.B. Sartor, VSL#3 probiotic-mixture induces remission in patients with active ulcerative colitis, Am. J. Gastroenterol. 100 (2005) 1539–1546.
[71] A. Tursi, G. Brandimarte, A. Papa, A. Giglio, W. Elisei, G.M. Giorgetti, G. Forti, S. Morini, C. Hassan, M.A. Pistoia, M.E. Modeo, S. Rodino, T. D'amico, L. Sebkova, N. Sacca, E. Di Giulio, F. Luzza, M. Imeneo, T. Larussa, S. Di Rosa, V. Annese, S. Danese, A. Gasbarrini, Treatment of relapsing mild-to-moderate Ulcerative Colitis with the probiotic VSL#3 as adjunctive to a standard pharmaceutical treatment: a double-blind, randomized, Placebo-Controlled Study, Am. J. Gastroenterol. 105 (2010) 2218–2227.
[72] F. Calcinaro, S. Dionisi, M. Marinaro, P. Candeloro, V. Bonato, S. Marzotti, R.B. Corneli, E. Ferretti, A. Gulino, F. Grasso, C. De Simone, U. Di Mario, A. Falorni, M. Boirivant, F. Dotta, Oral probiotic administration induces interleukin-10 production and prevents spontaneous autoimmune diabetes in the non-obese diabetic mouse, Diabetologia 48 (2005) 1565–1575.
[73] D. Unutmaz, S. Lavasani, B. Dzhambazov, M. Nouri, F. Fåk, S. Buske, G. Molin, H. Thorlacius, J. Alenfall, B. Jeppsson, B. Weström, A novel probiotic mixture exerts a therapeutic effect on experimental autoimmune encephalomyelitis mediated by IL-10 producing regulatory T cells, PLoS ONE 5 (2) (2010) e9009.
[74] M. Viljanen, E. Pohjavuori, T. Haahtela, R. Korpela, M. Kuitunen, A. Sarnesto, O. Vaarala, E. Savilahti, Induction of inflammation as a possible mechanism of probiotic effect in atopic eczema–dermatitis syndrome, J. Allergy Clin. Immunol. 115 (2005) 1254–1259.
[75] A. Miyoshi, E. Jamet, J. Commissaire, P. Renault, P. Langella, V. Azevedo, A xylose-inducible expression system for *Lactococcus lactis*, FEMS Microbiol. Lett. 239 (2004) 205–212.
[76] M.T. Islam, A. Deora, Y. Hashidoko, A. Rahman, T. Ito, S. Tahara, Isolation and identification of potential phosphate solubilizing bacteria from the rhizoplane of *Oryza sativa* L. cv. BR29 of Bangladesh, Z. Naturforsch. C 62 (2007) 103–110.
[77] D. Thakuria, N.C. Talukdar, C. Goswami, S. Hazarika, R.C. Boro, M.R. Khan, Characterization and screening of bacteria from rhizosphere of rice grown in acidic soils of Assam, Curr. Sci. 86 (7) (2004) 978–985.
[78] M. Ogut, F. Er, N. Kandemir, Phosphate solubilization potentials of soil *Acinetobacter strains*, Biol. Fertil. Soils 46 (2010) 707–715.
[79] H. Cao, S. He, R. Wei, M. Diong, L. Lu, *Bacillus amyloliquefaciens* G1: a potential antagonistic bacterium against eel-pathogenic *Aeromonas hydrophila*, Evid. Based Complement. Alternat. Med. 2011 (2011) 1–7.
[80] J. Ji, S. Hu, W. Li, Probiotic *Bacillus amyloliquefaciens* SC06 prevents bacterial translocation in weaned mice, Indian J. Microbiol. 53 (2013) 323–328.
[81] M.R. Sudha, S. Bhonagiri, M.A. Kumar, Efficacy of *Bacillus clausii* strain UBBC-07 in the treatment of patients suffering from acute diarrhoea, Benefic. Microbes 4 (2013) 211–216.
[82] H.A. Hong, L.H. Duc, S.M. Cutting, The use of bacterial spore formers as probiotics: Table 1, FEMS Microbiol. Rev. 29 (2005) 813–835.

[83] M. La Rosa, G. Bottaro, N. Gulino, F. Gambuzza, F. Di Forti, G. Ini, E. Tornambe, Prevention of antibiotic-associated diarrhea with Lactobacillus sporogens and fructo-oligosaccharides in children. A multicentric double-blind vs placebo study, Minerva Pediatr. 55 (2003) 447–452.
[84] N.M. Gracheva, A.F. Gavrilov, A.I. Solov'eva, V.V. Smirnov, I.B. Sorokulova, S.R. Reznik, N.V. Chudnovskaia, The efficacy of the new bacterial preparation biosporin in treating acute intestinal infections, Zh. Mikrobiol. Epidemiol. Immunobiol. 1 (1996) 75–77.
[85] P. Pattnaik, S. Grover, V.K. Batish, Effect of environmental factors on production of lichenin, a chromosomally encoded bacteriocin-like compound produced by *Bacillus licheniformis* 26L-10/3RA, Microbiol. Res. 160 (2005) 213–218.
[86] C. Liu, J. Lu, L. Lu, Y. Liu, F. Wang, M. Xiao, Isolation, structural characterization and immunological activity of an exopolysaccharide produced by *Bacillus licheniformis* 8-37-0-1, Bioresour. Technol. 101 (2010) 5528–5533.
[87] D.-Y. Tseng, P.-L. Ho, S.-Y. Huang, S.-C. Cheng, Y.-L. Shiu, C.-S. Chiu, C.-H. Liu, Enhancement of immunity and disease resistance in the white shrimp, *Litopenaeus vannamei*, by the probiotic, *Bacillus subtilis* E20, Fish Shellfish Immunol. 26 (2009) 339–344.
[88] J.A. Gilbert, R. Krajmalnik-Brown, D.L. Porazinska, S.J. Weiss, R. Knight, Toward effective probiotics for autism and other neurodevelopmental disorders, Cell 155 (2013) 1446–1448.
[89] M. Saxelin, S. Tynkkynen, T. Mattila-Sandholm, W.M. De Vos, Probiotic and other functional microbes: from markets to mechanisms, Curr. Opin. Biotechnol. 16 (2005) 204–211.
[90] K.Y. Wang, S.N. Li, C.S. Liu, D.S. Perng, Y.C. Su, D.C. Wu, C.M. Jan, C.H. Lai, T.N. Wang, W.M. Wang, Effects of ingesting Lactobacillus- and Bifidobacterium-containing yogurt in subjects with colonized *Helicobacter pylori*, Am. J. Clin. Nutr. 80 (2004) 737–741.
[91] C.K. Dotterud, O. Storrø, R. Johnsen, T. Øien, Probiotics in pregnant women to prevent allergic disease: a randomized, double-blind trial, Br. J. Dermatol. 163 (2010) 616–623.
[92] B.S. Kang, J.-G. Seo, G.-S. Lee, J.-H. Kim, S.Y. Kim, Y.W. Han, H. Kang, H.O. Kim, J.H. Rhee, M.-J. Chung, Y.M. Park, Antimicrobial activity of enterocins from Enterococcus faecalis SL-5 against Propionibacterium acnes, the causative agent in acne vulgaris, and its therapeutic effect, J. Microbiol. 47 (2009) 101–109.
[93] T. Aymerich, M.G. Artigas, M. Garriga, J.M. Monfort, M. Hugas, Effect of sausage ingredients and additives on the production of enterocin A and B by *Enterococcus faecium* CTC492. Optimization of in vitro production and anti-listerial effect in dry fermented sausages, J. Appl. Microbiol. 88 (2000) 686–694.
[94] B. Olle, Medicines from microbiota, Nat. Biotechnol. 31 (2013) 309–315.
[95] W. Kruis, Maintaining remission of ulcerative colitis with the probiotic *Escherichia coli* Nissle 1917 is as effective as with standard mesalazine, Gut 53 (2004) 1617–1623.
[96] H.A. Malchow, Crohn's disease and *Escherichia coli*. A new approach in therapy to maintain remission of colonic Crohn's disease? J. Clin. Gastroenterol. 25 (1997) 653–658.
[97] A. Sturm, K. Rilling, D.C. Baumgart, K. Gargas, T. Abou-Ghazale, B. Raupach, J. Eckert, R.R. Schumann, C. Enders, U. Sonnenborn, B. Wiedenmann, A.U. Dignass, *Escherichia coli* Nissle 1917 distinctively modulates T-cell cycling and expansion via toll-like receptor 2 signaling, Infect. Immun. 73 (2005) 1452–1465.
[98] Y. Inoue, T. Kambara, N. Murata, J. Komori-Yamaguchi, S. Matsukura, Y. Takahashi, Z. Ikezawa, M. Aihara, Effects of oral administration of *Lactobacillus acidophilus* L-92 on the symptoms and serum cytokines of atopic dermatitis in Japanese adults: a double-blind, randomized, clinical trial, Int. Arch. Allergy Immunol. 165 (2014) 247–254.
[99] F. Murina, A. Graziottin, F. Vicariotto, F. De Seta, Can *Lactobacillus fermentum* LF10 and *Lactobacillus acidophilus* LA02 in a slow-release vaginal product be useful for prevention of recurrent vulvovaginal candidiasis? J. Clin. Gastroenterol. 48 (2014) S102–S105.
[100] Y.-J. Lai, S.-H. Tsai, M.-Y. Lee, Isolation of exopolysaccharide producing Lactobacillus strains from sorghum distillery residues pickled cabbage and their antioxidant properties, Food Sci. Biotechnol. 23 (2014) 1231–1236.
[101] N. Waki, N. Yajima, H. Suganuma, B.M. Buddle, D. Luo, A. Heiser, T. Zheng, Oral administration ofLactobacillus brevisKB290 to mice alleviates clinical symptoms following influenza virus infection, Lett. Appl. Microbiol. 58 (2014) 87–93.

[102] X.Q. Zeng, D.D. Pan, Y.X. Guo, The probiotic properties of *Lactobacillus buchneri* P2. J. Appl. Microbiol. 108 (6) (2010) 2059–2066, https://doi.org/10.1111/j.1365-2672.2009.04608.x.
[103] A. Marcos, J. Wärnberg, E. Nova, S. Gómez, A. Alvarez, R. Alvarez, J.A. Mateos, J.M. Cobo, The effect of milk fermented by yogurt cultures plus *Lactobacillus casei* DN-114001 on the immune response of subjects under academic examination stress, Eur. J. Nutr. 43 (2004) 381–389.
[104] R.J. Siezen, G. Wilson, Probiotics genomics, Microb. Biotechnol. 3 (2010) 1–9.
[105] A.E. Stapleton, M. Au-Yeung, T.M. Hooton, D.N. Fredricks, P.L. Roberts, C.A. Czaja, Y. Yarova-Yarovaya, T. Fiedler, M. Cox, W.E. Stamm, Randomized, placebo-controlled phase 2 trial of a *Lactobacillus crispatus* probiotic given intravaginally for prevention of recurrent urinary tract infection, Clin. Infect. Dis. 52 (2011) 1212–1217.
[106] S. Makino, S. Ikegami, A. Kume, H. Horiuchi, H. Sasaki, N. Orii, Reducing the risk of infection in the elderly by dietary intake of yoghurt fermented with Lactobacillus delbrueckii ssp. bulgaricus OLL1073R-1, Br. J. Nutr. 104 (2010) 998–1006.
[107] M. Sanchez, C. Darimont, V. Drapeau, S. Emady-Azar, M. Lepage, E. Rezzonico, C. Ngom-Bru, B. Berger, L. Philippe, C. Ammon-Zuffrey, P. Leone, G. Chevrier, E. St-Amand, A. Marette, J. Doré, A. Tremblay, Effect of Lactobacillus rhamnosus CGMCC1.3724 supplementation on weight loss and maintenance in obese men and women, Br. J. Nutr. 111 (2013) 1507–1519.
[108] S. Chabot, H.-L. Yu, L. De Léséleuc, D. Cloutier, M.-R. Van Calsteren, M. Lessard, D. Roy, M. Lacroix, D. Oth, Exopolysaccharides from Lactobacillus rhamnosus RW-9595M stimulate TNF, IL-6 and IL-12 in human and mouse cultured immunocompetent cells, and IFN-γ in mouse splenocytes, Lait 81 (2001) 683–697.
[109] J.P. Madej, T. Stefaniak, M. Bednarczyk, Effect ofin ovo-delivered prebiotics and synbiotics on lymphoid-organs' morphology in chickens, Poult. Sci. 94 (2015) 1209–1219.
[110] M.L. Ellis, A.E. Dowell, X. Li, J. Knight, Probiotic properties of Oxalobacter formigenes: an in vitro examination, Arch. Microbiol. 198 (2016) 1019–1026.
[111] H.S. El-Nezami, N.N. Polychronaki, J. Ma, H. Zhu, W. Ling, E.K. Salminen, R.O. Juvonen, S.J. Salminen, T. Poussa, H.M. Mykkänen, Probiotic supplementation reduces a biomarker for increased risk of liver cancer in young men from Southern China, Am. J. Clin. Nutr. 83 (2006) 1199–1203.
[112] J.P. Burton, C.N. Chilcott, J.R. Tagg, The rationale and potential for the reduction of oral malodour using *Streptococcus salivarius* probiotics, Oral Dis. 11 (2005) 29–31.
[113] Y.J. Moon, J.R. Soh, J.J. Yu, H.S. Sohn, Y.S. Cha, S.H. Oh, Intracellular lipid accumulation inhibitory effect of *Weissella koreensis* OK1-6 isolated from Kimchi on differentiating adipocyte, J. Appl. Microbiol. 113 (2012) 652–658.
[114] J.A. Park, P.B. Tirupathi Pichiah, J.J. Yu, S.H. Oh, J.W. Daily, Y.S. Cha, Anti-obesity effect of kimchi fermented withWeissella koreensisOK1-6 as starter in high-fat diet-induced obese C57BL/6J mice, J. Appl. Microbiol. 113 (2012) 1507–1516.
[115] J. Touchman, *Comparative Genomics* [Online], in: Nature Education Knowledge, 2010. Available: https://www.nature.com/scitable/knowledge/library/comparative-genomics-13239404. (Accessed 14 January 2019).
[116] A. Bezkorovainy, Probiotics: determinants of survival and growth in the gut, Am. J. Clin. Nutr. 73 (2001) 399S–405S.
[117] G. Konuray, Z. Erginkaya, Potential use of *Bacillus coagulans* in the food industry, Foods 7 (2018).
[118] B.R. Johnson, T.R. Klaenhammer, Impact of genomics on the field of probiotic research: historical perspectives to modern paradigms, Antonie Van Leeuwenhoek 106 (2014) 141–156.
[119] L.C. Oliveira, T.D. Saraiva, W.M. Silva, U.P. Pereira, B.C. Campos, L.J. Benevides, F.S. Rocha, H. C.P. Figueiredo, V. Azevedo, S.C. Soares, Analyses of the probiotic property and stress resistance-related genes of Lactococcus lactis subsp. lactis NCDO 2118 through comparative genomics and in vitro assays. PLoS ONE 12 (4) (2017) e0175116. https://doi.org/10.1371/journal.pone.0175116.
[120] H. Willenbrock, P.F. Hallin, T.M. Wassenaar, D.W. Ussery, Characterization of probiotic Escherichia coli isolates with a novel pan-genome microarray, Genome Biol. 8 (2007).
[121] T.L. Nguyen, C.-I. Park, D.-H. Kim, Improved growth rate and disease resistance in olive flounder, Paralichthys olivaceus, by probiotic Lactococcus lactis WFLU12 isolated from wild marine fish, Aquaculture 471 (2017) 113–120.

[122] R. Kant, J. Rintahaka, X. Yu, P. Sigvart-Mattila, L. Paulin, J.-P. Mecklin, M. Saarela, A. Palva, I. Von Ossowski, A comparative pan-genome perspective of niche-adaptable cell-surface protein phenotypes in *Lactobacillus rhamnosus*. PLoS ONE 9 (7) (2014) e102762. https://doi.org/10.1371/journal.pone.0102762.
[123] T. Smokvina, M. Wels, J. Polka, C. Chervaux, S. Brisse, J. Boekhorst, J.E. Van Hylckama Vlieg, R.J. Siezen, Lactobacillus paracasei comparative genomics: towards species pan-genome definition and exploitation of diversity, PLoS ONE 8 (2013) e68731.
[124] J.L. Gardy, N.J. Loman, Towards a genomics-informed, real-time, global pathogen surveillance system, Nat. Rev. Genet. 19 (2018) 9–20.
[125] J. Shendure, H. Ji, Next-generation DNA sequencing, Nat. Biotechnol. 26 (2008) 1135–1145.
[125a] J. Quick, N.D. Grubaugh, S.T. Pullan, et al., Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. 12 (6) (2017) 1261–1276, https://doi.org/10.1038/nprot.2017.066.
[126] N.R. Faria, J. Quick, I.M. Claro, J. Theze, J.G. De Jesus, M. Giovanetti, M.U.G. Kraemer, S.C. Hill, A. Black, A.C. Da Costa, L.C. Franco, S.P. Silva, C.H. Wu, J. Raghwani, S. Cauchemez, L. Du Plessis, M.P. Verotti, W.K. De Oliveira, E.H. Carmo, G.E. Coelho, A. Santelli, L.C. Vinhal, C. M. Henriques, J.T. Simpson, M. Loose, K.G. Andersen, N.D. Grubaugh, S. Somasekar, C. Y. Chiu, J.E. Munoz-Medina, C.R. Gonzalez-Bonilla, C.F. Arias, L.L. Lewis-Ximenez, S. A. Baylis, A.O. Chieppe, S.F. Aguiar, C.A. Fernandes, P.S. Lemos, B.L.S. Nascimento, H.A. O. Monteiro, I.C. Siqueira, M.G. De Queiroz, T.R. De Souza, J.F. Bezerra, M.R. Lemos, G. F. Pereira, D. Loudal, L.C. Moura, R. Dhalia, R.F. Franca, T. Magalhaes, E.T. Marques Jr., T. Jaenisch, G.L. Wallau, M.C. De Lima, V. Nascimento, E.M. De Cerqueira, M.M. De Lima, D.L. Mascarenhas, J.P.M. Neto, A.S. Levin, T.R. Tozetto-Mendoza, S.N. Fonseca, M. C. Mendes-Correa, F.P. Milagres, A. Segurado, E.C. Holmes, A. Rambaut, T. Bedford, M.R. T. Nunes, E.C. Sabino, L.C.J. Alcantara, N.J. Loman, O.G. Pybus, Establishment and cryptic transmission of Zika virus in Brazil and the Americas, Nature 546 (2017) 406–410.
[127] J. Theze, T. Li, L. Du Plessis, J. Bouquet, M.U.G. Kraemer, S. Somasekar, G. Yu, M. De Cesare, A. Balmaseda, G. Kuan, E. Harris, C.H. Wu, M.A. Ansari, R. Bowden, N.R. Faria, S. Yagi, S. Messenger, T. Brooks, M. Stone, E.M. Bloch, M. Busch, J.E. Munoz-Medina, C.R. Gonzalez-Bonilla, S. Wolinsky, S. Lopez, C.F. Arias, D. Bonsall, C.Y. Chiu, O.G. Pybus, Genomic epidemiology reconstructs the introduction and spread of zika virus in Central America and Mexico, Cell Host Microbe 23 (855-864) (2018).
[128] L.C. Guimaraes, J. Florczak-Wyspianska, L.B. De Jesus, M.V. Viana, A. Silva, R.T. Ramos, C. Soares Sde, C. Soares Sde, Inside the pan-genome—methods and software overview, Curr. Genom. 16 (2015) 245–252.
[129] K. Padovani De Souza, J.C. Setubal, F. Ponce De Leon, A.C. De Carvalho, G. Oliveira, A. Chateau, R. Alves, Machine learning meets genome assembly, Brief Bioinform. (2018) 1–14.
[130] S.I. Lee, N.S. Kim, Transposable elements and genome size variations in plants, Genom. Inform. 12 (2014) 87–97.
[131] I. Arabidopsis Genome, Analysis of the genome sequence of the flowering plant Arabidopsis thaliana, Nature 408 (2000) 796–815.
[132] K.L. McNally, K.L. Childs, R. Bohnert, R.M. Davidson, K. Zhao, V.J. Ulat, G. Zeller, R.M. Clark, D.R. Hoen, T.E. Bureau, R. Stokowski, D.G. Ballinger, K.A. Frazer, D.R. Cox, B. Padhukasahasram, C.D. Bustamante, D. Weigel, D.J. Mackill, R.M. Bruskiewich, G. Ratsch, C.R. Buell, H. Leung, J.E. Leach, Genomewide SNP variation reveals relationships among landraces and modern varieties of rice, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 12273–12278.
[133] A.A. Golicz, J. Batley, D. Edwards, Towards plant pangenomics, Plant Biotechnol. J. 14 (2016) 1099–1105.
[134] J.D. Montenegro, A.A. Golicz, P.E. Bayer, B. Hurgobin, H. Lee, C.K. Chan, P. Visendi, K. Lai, J. Dolezel, J. Batley, D. Edwards, The pangenome of hexaploid bread wheat, Plant J. 90 (2017) 1007–1013.
[135] M.G. Milgroom, T.L. Peever, Population biology of plant pathogens: the synthesis of plant disease epidemiology and population genetics, Plant Dis. 87 (2003) 608–617.

[136] B.M. Tyler, S. Tripathy, X. Zhang, P. Dehal, R.H. Jiang, A. Aerts, F.D. Arredondo, L. Baxter, D. Bensasson, J.L. Beynon, J. Chapman, C.M. Damasceno, A.E. Dorrance, D. Dou, A. W. Dickerman, I.L. Dubchak, M. Garbelotto, M. Gijzen, S.G. Gordon, F. Govers, N. J. Grunwald, W. Huang, K.L. Ivors, R.W. Jones, S. Kamoun, K. Krampis, K.H. Lamour, M.K. Lee, W.H. Mcdonald, M. Medina, H.J. Meijer, E.K. Nordberg, D.J. Maclean, M. D. Ospina-Giraldo, P.F. Morris, V. Phuntumart, N.H. Putnam, S. Rash, J.K. Rose, Y. Sakihama, A.A. Salamov, A. Savidor, C.F. Scheuring, B.M. Smith, B.W. Sobral, A. Terry, T.A. Torto-Alalibo, J. Win, Z. Xu, H. Zhang, I.V. Grigoriev, D.S. Rokhsar, J.L. Boore, Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis, Science 313 (2006) 1261–1266.

[137] J.O. McInerney, A. McNally, M.J. O'Connell, Why prokaryotes have pangenomes, Nat. Microbiol. 2 (2017) 17040.

[138] C. Plissonneau, F.E. Hartmann, D. Croll, Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome, BMC Biol. 16 (2018) 5.

[139] J.C. Meeks, E.L. Campbell, M.L. Summers, F.C. Wong, Cellular differentiation in the cyanobacterium Nostoc punctiforme, Arch. Microbiol. 178 (2002) 395–403.

[140] D.R. Nelson, B. Khraiwesh, W. Fu, S. Alseekh, A. Jaiswal, A. Chaiboonchoe, K.M. Hazzouri, M. J. O'connor, G.L. Butterfoss, N. Drou, J.D. Rowe, J. Harb, A.R. Fernie, K.C. Gunsalus, K. Salehi-Ashtiani, The genome and phenome of the green alga Chloroidium sp. UTEX 3007 reveal adaptive traits for desert acclimatization. Elife 6 (2017) e25783 https://doi.org/10.7554/eLife.25783.

[141] S. Hirooka, Y. Hirose, Y. Kanesaki, S. Higuchi, T. Fujiwara, R. Onuma, A. Era, R. Ohbayashi, A. Uzuka, H. Nozaki, H. Yoshikawa, S.Y. Miyagishima, Acidophilic green algal genome provides insights into adaptation to an acidic environment, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) E8304–E8313.

[142] G. Barbier, C. Oesterhelt, M.D. Larson, R.G. Halgren, C. Wilkerson, R.M. Garavito, C. Benning, A.P. Weber, Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and Cyanidioschyzon merolae, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae, Plant Physiol. 137 (2005) 460–474.

[143] D. Bhattacharya, D.C. Price, C.X. Chan, H. Qiu, N. Rose, S. Ball, A.P. Weber, M.C. Arias, B. Henrissat, P.M. Coutinho, A. Krishnan, S. Zauner, S. Morath, F. Hilliou, A. Egizi, M.M. Perrineau, H.S. Yoon, Genome of the red alga Porphyridium purpureum, Nat. Commun. 4 (2013) 1941.

[144] S. Bose, S.K. Herbert, D.C. Fork, Fluorescence characteristics of photoinhibition and recovery in a sun and a shade species of the red algal genus porphyra, Plant Physiol. 86 (1988) 946–950.

[145] K. Nishitsuji, A. Arimoto, K. Iwai, Y. Sudo, K. Hisata, M. Fujie, N. Arakaki, T. Kushiro, T. Konishi, C. Shinzato, N. Satoh, E. Shoguchi, A draft genome of the brown alga, *Cladosiphon okamuranus*, S-strain: a platform for future studies of 'mozuku' biology, DNA Res. 23 (2016) 561–570.

[146] A. Sboner, X.J. Mu, D. Greenbaum, R.K. Auerbach, M.B. Gerstein, The real cost of sequencing: higher than you think!, Genome Biol. 12 (2011) 125.

[147] M. Guegan, K. Zouache, C. Demichel, G. Minard, V. Tran Van, P. Potier, P. Mavingui, C. Valiente Moro, The mosquito holobiont: fresh insight into mosquito-microbiota interactions, Microbiome 6 (2018) 49.

[148] D. Aguirre De Carcer, The human gut pan-microbiome presents a compositional core formed by discrete phylogenetic units, Sci. Rep. 8 (2018) 14069.

[149] M.H. Leung, P.K. Lee, The roles of the outdoors and occupants in contributing to a potential pan-microbiome of the built environment: a review, Microbiome 4 (2016) 21.

[150] P. Vandenkoornhuyse, A. Quaiser, M. Duhamel, A. Le Van, A. Dufresne, The importance of the microbiome of the plant holobiont, New Phytol. 206 (2015) 1196–1206.

[151] B. Aslam, M. Basit, M.A. Nisar, M.H. Rasool, M. Khurshid, Proteomics: technologies and their applications, J. Chromatogr. Sci. 55 (2017) 182–196.

[152] W.M. Silva, R.D. Carvalho, S.C. Soares, I.F.S. Bastos, E.L. Folador, G.H.M.F. Souza, Y. Le Loir, A. Miyoshi, A. Silva, V. Azevedo, Label-free proteomic analysis to confirm the predicted proteome of

Corynebacterium pseudotuberculosis under nitrosative stress mediated by nitric oxide, BMC Genom. 15 (2014) 1065.
[153] W.M. Silva, R.D.O. Carvalho, F.A. Dorella, E.L. Folador, G.H.M.F. Souza, A.M.C. Pimenta, H.C. P. Figueiredo, Y. Le Loir, A. Silva, V. Azevedo, Quantitative proteomic analysis reveals changes in the benchmark *Corynebacterium pseudotuberculosis* biovar equi exoproteome after passage in a murine host. Front. Cell. Infect. Microbiol. 7 (2017) 325, https://doi.org/10.3389/fcimb.2017.00325.
[154] T.-C. Chao, N. Hansmeier, The current state of microbial proteomics: where we are and where we want to go, Proteomics 12 (2012) 638–650.
[155] M.A. Moseley, Quantitative proteomics in genomic medicine, in: G.S. Ginsburg, H.F. Willard (Eds.), Genomic and Personalized Medicine, second ed., Academic Press, 2013, pp. 155–165 (Chapter 13).
[156] M.A. Reymond, W. Schlegel, Proteomics in cancer, Adv. Clin. Chem. 44 (2007) 103–142.
[157] M.A. Hussain, F. Huygens, Proteomic and bioinformatics tools to understand virulence mechanisms in *Staphylococcus aureus*, Curr. Proteom. 9 (2012) 2–8.
[158] O. Coskun, Separation techniques: Chromatography, North. Clin. Istanb. 3 (2016) 156–160.
[159] R.M. Lequin, Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA), Clin. Chem. 51 (2005) 2415–2418.
[160] B.T. Kurien, R.H. Scofield, Western blotting: an introduction, in: B.T. Kurien, R.H. Scofield (Eds.), Western Blotting: Methods and Protocols, Springer New York, New York, NY, 2015, pp. 17–30.
[161] M. D'Innocenzo, Identificação das proteínas por meio da eletroforese 2D, in: R. Verlengia, R. Curi, E. Bevilacqua, P. Newsholme (Eds.), Análises de RNA, proteínas e metabólitos: metodologia e procedimentos técnicos, Santos Editora, São Paulo, 2013, pp. 261–280.
[162] R. Vessecchi, N.P. Lopes, F.C. Gozzo, F.A. Dörr, M. Murgu, D.T. Lebre, R. Abreu, O.V. Bustillos, J.M. Riveros, Nomenclaturas de espectrometria de massas em língua portuguesa, Quím. Nova 34 (2011) 1875–1887.
[163] S.J. Cordwell, A.S. Nouwens, B.J. Walsh, Comparative proteomics of bacterial pathogens, Proteomics 1 (2001) 461–472.
[164] P.M. Bisch, Genômica funcional: proteômica, in: L. Mir (Ed.), Genômica, Atheneu, São Paulo, 2004, pp. 139–162.
[165] E.-H. Jeong, B. Vaidya, S.-Y. Cho, M.-A. Park, K. Kaewintajuk, S.R. Kim, M.-J. Oh, J.-S. Choi, J. Kwon, D. Kim, Identification of regulators of the early stage of viral hemorrhagic septicemia virus infection during curcumin treatment, Fish Shellfish Immunol. 45 (2015) 184–193.
[166] N. Solis, S.J. Cordwell, Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins, Proteomics 11 (2011) 3169–3189.
[167] S. Hölper, A. Ruhs, M. Krüger, Stable isotope labeling for proteomic analysis of tissues in mouse, in: B. Warscheid (Ed.), Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC): Methods and Protocols, Springer New York, New York, NY, 2014, pp. 95–106.
[168] K. Cheng, A. Sloan, S. Mccorrister, L. Peterson, H. Chui, M. Drebot, C. Nadon, J.D. Knox, G. Wang, Quality evaluation of LC-MS/MS-based E. coli H antigen typing (MS-H) through label-free quantitative data analysis in a clinical sample setup, Proteom. Clin. Appl. 8 (2014) 963–970.
[169] S. Kosono, M. Tamura, S. Suzuki, Y. Kawamura, A. Yoshida, M. Nishiyama, M. Yoshida, Changes in the acetylome and succinylome of *Bacillus subtilis* in response to carbon source, PLoS ONE 10 (2015) e0131169.
[170] S.P. Gygi, B. Rist, T.J. Griffin, J. Eng, R. Aebersold, Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags, J. Proteome Res. 1 (2002) 47–54.
[171] V.J. Patel, K. Thalassinos, S.E. Slade, J.B. Connolly, A. Crombie, J.C. Murrell, J.H. Scrivens, A comparison of labeling and label-free mass spectrometry-based proteomics approaches, J. Proteome Res. 8 (2009) 3752–3759.
[172] D. Chelius, P.V. Bondarenko, Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry, J. Proteome Res. 1 (2002) 317–323.
[173] M.J.G. Hughes, J.C. Moore, J.D. Lane, R. Wilson, P.K. Pribul, Z.N. Younes, R.J. Dobson, P. Everest, A.J. Reason, J.M. Redfern, F.M. Greer, T. Paxton, M. Panico, H.R. Morris, R.

G. Feldman, J.D. Santangelo, Identification of major outer surface proteins of *Streptococcus agalactiae*, Infect. Immun. 70 (2002) 1254–1259.
[174] F. Doro, S. Liberatori, M.J. Rodríguez-Ortega, C.D. Rinaudo, R. Rosini, M. Mora, M. Scarselli, E. Altindis, R. D'aurizio, M. Stella, I. Margarit, D. Maione, J.L. Telford, N. Norais, G. Grandi, Surfome analysis as a fast track to vaccine discovery: identification of a novel protective antigen for group B *Streptococcus hypervirulent* strain COH1, Mol. Cell. Proteom. 8 (2009) 1728–1737.
[175] W.M. Silva, N. Seyffert, A.V. Santos, T.L.P. Castro, L.G.C. Pacheco, A.R. Santos, A. Ciprandi, F.A. Dorella, H.M. Andrade, D. Barh, A.M.C. Pimenta, A. Silva, A. Miyoshi, V. Azevedo, Identification of 11 new exoproteins in *Corynebacterium pseudotuberculosis* by comparative analysis of the exoproteome, Microb. Pathog. 61–62 (2013) 37–42.
[176] W.M. Silva, N. Seyffert, A. Ciprandi, A.V. Santos, T.L.P. Castro, L.G.C. Pacheco, D. Barh, Y. Le Loir, A.M.C. Pimenta, A. Miyoshi, A. Silva, V. Azevedo, Differential exoproteome analysis of two *Corynebacterium pseudotuberculosis* biovar ovis strains isolated from goat (1002) and sheep (C231), Curr. Microbiol. 67 (2013) 460–465.
[177] J.A. Broadbent, D.A. Broszczak, I.U.K. Tennakoon, F. Huygens, Pan-proteomics, a concept for unifying quantitative proteome measurements when comparing closely-related bacterial strains, Expert Rev. Proteom. 13 (2016) 355–365.
[178] G.C. Tavares, F.L. Pereira, G.M. Barony, C.P. Rezende, W.M. Da Silva, G.H.M.F. De Souza, T. Verano-Braga, V.A. De Carvalho Azevedo, Leal, C.a.G., and Figueiredo, H.C.P., Delineation of the pan-proteome of fish-pathogenic Streptococcus agalactiae strains using a label-free shotgun approach, BMC Genom. 20 (2019) 11.
[179] J. Rothen, J.F. Pothier, F. Foucault, J. Blom, D. Nanayakkara, C. Li, M. Ip, M. Tanner, G. Vogel, V. Pflüger, C.A. Daubenberger, Subspecies typing of *Streptococcus agalactiae* based on ribosomal subunit protein mass variation by MALDI-TOF MS, Front. Microbiol. 10 (2019) 471.
[180] L. Zhang, D. Xiao, B. Pang, Q. Zhang, H. Zhou, L. Zhang, J. Zhang, B. Kan, The core proteome and pan proteome of *Salmonella* Paratyphi A epidemic strains, PLoS ONE 9 (2014) e89197.
[181] G.D. Jhingan, S. Kumari, S.V. Jamwal, H. Kalam, D. Arora, N. Jain, L.K. Kumaar, A. Samal, K.V.S. Rao, D. Kumar, V.K. Nandicoori, Comparative proteomic analyses of avirulent, virulent, and clinical strains of *Mycobacterium tuberculosis* identify strain-specific patterns, J. Biol. Chem. 291 (2016) 14257–14273.
[182] W.M. Silva, C.S. Sousa, L.C. Oliveira, S.C. Soares, G. Souza, G.C. Tavares, C.P. Resende, E.L. Folador, F.L. Pereira, H. Figueiredo, V. Azevedo, Comparative proteomic analysis of four biotechnological strains *Lactococcus lactis* through label-free quantitative proteomics, Microb. Biotechnol. 12 (2019) 265–274.
[183] J. Trapp, C. Almunia, J.-C. Gaillard, O. Pible, A. Chaumot, O. Geffard, J. Armengaud, Proteogenomic insights into the core-proteome of female reproductive tissues from crustacean amphipods, J. Proteome 135 (2016) 51–61.
[184] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (2009) 57–63.
[185] I. Korf, Genomics: the state of the art in RNA-seq analysis, Nat. Methods 10 (2013) 1165–1166.
[186] M.F. Rai, E.D. Tycksen, L.J. Sandell, R.H. Brophy, Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears, J. Orthop. Res. 36 (2018) 484–497.
[187] S.C. Sealfon, T.T. Chu, RNA and DNA microarrays, Methods Mol. Biol. 671 (2011) 3–34.
[188] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, T. Shafee, Transcriptomics technologies, PLoS Comput. Biol. 13 (2017) e1005457.
[189] S. Zhao, W.P. Fung-Leung, A. Bittner, K. Ngo, X. Liu, Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells, PLoS ONE 9 (2014) e78644.
[190] M. Blaxter, S. Kumar, G. Kaur, G. Koutsovoulos, B. Elsworth, Genomics and transcriptomics across the diversity of the Nematoda, Parasite Immunol. 34 (2012) 108–120.
[191] M.S. Kim, H. Zhang, H. Yan, B.J. Yoon, W.B. Shim, Characterizing co-expression networks underpinning maize stalk rot virulence in Fusarium verticillioides through computational subnetwork module analyses, Sci. Rep. 8 (2018) 8310.

[192] Z. Wei, H. Guo, J. Qin, S. Lu, Q. Liu, X. Zhang, Y. Zou, Y. Gong, C. Shao, Pan-senescence transcriptome analysis identified RRAD as a marker and negative regulator of cellular senescence, Free Radic. Biol. Med. 130 (2019) 267–277.
[193] X. Ma, Y. Liu, Y. Liu, L.B. Alexandrov, M.N. Edmonson, C. Gawad, X. Zhou, Y. Li, M.C. Rusch, J. Easton, R. Huether, V. Gonzalez-Pena, M.R. Wilkinson, L.C. Hermida, S. Davis, E. Sioson, S. Pounds, X. Cao, R.E. Ries, Z. Wang, X. Chen, L. Dong, S.J. Diskin, M.A. Smith, J.M. Guidry Auvil, P.S. Meltzer, C.C. Lau, E.J. Perlman, J.M. Maris, S. Meshinchi, S.P. Hunger, D.S. Gerhard, J. Zhang, Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours, Nature 555 (2018) 371–376.
[194] C.R. Cabanski, N.M. White, H.X. Dang, J.M. Silva-Fisher, C.E. Rauck, D. Cicka, C.A. Maher, Pan-cancer transcriptome analysis reveals long noncoding RNAs with conserved function, RNA Biol. 12 (2015) 628–642.
[195] G. Dugar, A. Herbig, K.U. Forstner, N. Heidrich, R. Reinhardt, K. Nieselt, C.M. Sharma, High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates, PLoS Genet. 9 (2013) e1003495.
[196] B. Sidders, M. Withers, S.L. Kendall, J. Bacon, S.J. Waddell, J. Hinds, P. Golby, F. Movahedzadeh, R. A. Cox, R. Frita, A.M. Ten Bokum, L. Wernisch, N.G. Stoker, Quantification of global transcription patterns in prokaryotes using spotted microarrays, Genome Biol. 8 (2007) R265.
[197] C. Eymann, G. Homuth, C. Scharf, M. Hecker, Bacillus subtilis functional genomics: global characterization of the stringent response by proteome and transcriptome analysis, J. Bacteriol. 184 (2002) 2500–2520.
[198] H. Qin, N.W. Lo, J.F. Loo, X. Lin, A.K. Yim, S.K. Tsui, T.C. Lau, M. Ip, T.F. Chan, Comparative transcriptomics of multidrug-resistant *Acinetobacter baumannii* in response to antibiotic treatments, Sci. Rep. 8 (2018) 3515.
[199] A. Dotsch, M. Schniederjans, A. Khaledi, K. Hornischer, S. Schulz, A. Bielecka, D. Eckweiler, S. Pohl, S. Haussler, The *Pseudomonas aeruginosa* transcriptional landscape is shaped by environmental heterogeneity and genetic variation, MBio 6 (2015) e00749.
[200] L. De Welzen, V. Eldholm, K. Maharaj, A.L. Manson, A.M. Earl, A.S. Pym, Whole-transcriptome and -genome analysis of extensively drug-resistant *Mycobacterium tuberculosis* clinical isolates identifies downregulation of etha as a mechanism of ethionamide resistance, Antimicrob. Agents Chemother. 61 (2017).
[201] T.H. Hazen, J. Michalski, Q. Luo, A.C. Shetty, S.C. Daugherty, J.M. Fleckenstein, D.A. Rasko, Comparative genomics and transcriptomics of *Escherichia coli* isolates carrying virulence factors of both enteropathogenic and enterotoxigenic *E. coli*, Sci. Rep. 7 (2017) 3513.
[202] P.A. Northcott, C. Lee, T. Zichner, A.M. Stutz, S. Erkek, D. Kawauchi, D.J. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D.T. Jones, M. Kool, M. Remke, F.M. Cavalli, S. Zuyderduyn, G.D. Bader, S. Vandenberg, L.A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jager, I. Buchhalter, C.D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H.J. Warnatz, T. Risch, M.L. Yaspo, U.D. Weber, C.C. Bartholomae, C. Von Kalle, E. Turanyi, P. Hauser, E. Sanden, A. Darabi, P. Siesjo, J. Sterba, K. Zitterbart, D. Sumerauer, P. Van Sluis, R. Versteeg, R. Volckmann, J. Koster, M.U. Schuhmann, M. Ebinger, H.L. Grimes, G.W. Robinson, A. Gajjar, M. Mynarek, K. Von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A.E. Kulozik, A. Von Deimling, O. Witt, R. Eils, R.J. Gilbertson, A. Korshunov, M.D. Taylor, P. Lichter, J.O. Korbel, R.J. Wechsler-Reya, S.M. Pfister, Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma, Nature 511 (2014) 428–434.
[203] E. Papaemmanuil, M. Cazzola, J. Boultwood, L. Malcovati, P. Vyas, D. Bowen, A. Pellagatti, J. S. Wainscoat, E. Hellstrom-Lindberg, C. Gambacorti-Passerini, A.L. Godfrey, I. Rapado, A. Cvejic, R. Rance, C. Mcgee, P. Ellis, L.J. Mudie, P.J. Stephens, S. Mclaren, C.E. Massie, P. S. Tarpey, I. Varela, S. Nik-Zainal, H.R. Davies, A. Shlien, D. Jones, K. Raine, J. Hinton, A. P. Butler, J.W. Teague, E.J. Baxter, J. Score, A. Galli, M.G. Della Porta, E. Travaglino, M. Groves, S. Tauro, N.C. Munshi, K.C. Anderson, A. El-Naggar, A. Fischer, V. Mustonen, A.

J. Warren, N.C. Cross, A.R. Green, P.A. Futreal, M.R. Stratton, P.J. Campbell, Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium, Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts, N. Engl. J. Med. 365 (2011) 1384–1395.

[204] X.S. Puente, M. Pinyol, V. Quesada, L. Conde, G.R. Ordonez, N. Villamor, G. Escaramis, P. Jares, S. Bea, M. Gonzalez-Diaz, L. Bassaganyas, T. Baumann, M. Juan, M. Lopez-Guerra, D. Colomer, J. M. Tubio, C. Lopez, A. Navarro, C. Tornador, M. Aymerich, M. Rozman, J.M. Hernandez, D. A. Puente, J.M. Freije, G. Velasco, A. Gutierrez-Fernandez, D. Costa, A. Carrio, S. Guijarro, A. Enjuanes, L. Hernandez, J. Yague, P. Nicolas, C.M. Romeo-Casabona, H. Himmelbauer, E. Castillo, J.C. Dohm, S. De Sanjose, M.A. Piris, E. De Alava, J. San Miguel, R. Royo, J. L. Gelpi, D. Torrents, M. Orozco, D.G. Pisano, A. Valencia, R. Guigo, M. Bayes, S. Heath, M. Gut, P. Klatt, J. Marshall, K. Raine, L.A. Stebbings, P.A. Futreal, M.R. Stratton, P. J. Campbell, I. Gut, A. Lopez-Guillermo, X. Estivill, E. Montserrat, C. Lopez-Otin, E. Campo, Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia, Nature 475 (2011) 101–105.

[205] Z. Liu, S. Zhang, Toward a systematic understanding of cancers: a survey of the pan-cancer study, Front. Genet. 5 (2014) 194.

[206] T.J. Hudson, W. Anderson, A. Artez, A.D. Barker, C. Bell, R.R. Bernabe, M.K. Bhan, F. Calvo, I. Eerola, D.S. Gerhard, A. Guttmacher, M. Guyer, F.M. Hemsley, J.L. Jennings, D. Kerr, P. Klatt, P. Kolar, J. Kusada, D.P. Lane, F. Laplace, L. Youyong, G. Nettekoven, B. Ozenberger, J. Peterson, T.S. Rao, J. Remacle, A.J. Schafer, T. Shibata, M.R. Stratton, J.G. Vockley, K. Watanabe, H. Yang, M.M. Yuen, B.M. Knoppers, M. Bobrow, A. Cambon-Thomsen, L. G. Dressler, S.O. Dyke, Y. Joly, K. Kato, K.L. Kennedy, P. Nicolas, M.J. Parker, E. Rial-Sebbag, C.M. Romeo-Casabona, K.M. Shaw, S. Wallace, G.L. Wiesner, N. Zeps, P. Lichter, A. V. Biankin, C. Chabannon, L. Chin, B. Clement, E. De Alava, F. Degos, M.L. Ferguson, P. Geary, D.N. Hayes, T.J. Hudson, A.L. Johns, A. Kasprzyk, H. Nakagawa, R. Penny, M. A. Piris, R. Sarin, A. Scarpa, T. Shibata, M. Van De Vijver, P.A. Futreal, H. Aburatani, M. Bayes, D.D. Botwell, P.J. Campbell, X. Estivill, D.S. Gerhard, S.M. Grimmond, I. Gut, M. Hirst, C. Lopez-Otin, P. Majumder, M. Marra, J.D. Mcpherson, H. Nakagawa, Z. Ning, X. S. Puente, Y. Ruan, T. Shibata, M.R. Stratton, H.G. Stunnenberg, H. Swerdlow, V. E. Velculescu, R.K. Wilson, H.H. Xue, L. Yang, P.T. Spellman, G.D. Bader, P.C. Boutros, P. J. Campbell, P. Flicek, et al., International network of cancer genome projects, Nature 464 (2010) 993–998.

[207] D.A. Levine, Integrated genomic characterization of endometrial carcinoma, Nature 497 (2013) 67–73.