Interest of bacterial pangenome analyses in clinical microbiology

Hussein Anani, Rita Zgheib, Issam Hasni, Didier Raoult, Pierre-Edouard Fournier

Microbial Pathogenesis 149 (2020) 104275

Contents lists available at ScienceDirect

Microbial Pathogenesis

journal homepage: www.elsevier.com/locate/micpath

Interest of bacterial pangenome analyses in clinical microbiology

Hussein Anani[a,b], Rita Zgheib[a,b], Issam Hasni[b,c], Didier Raoult[b,c,d], Pierre-Edouard Fournier[a,b,*]

[a] Aix Marseille Univ, Institut de Recherche pour le Développement (IRD), Service de Santé des Armées, AP-HM, UMR Vecteurs Infections Tropicales et Méditerranéennes (VITROME), Institut Hospitalo-Universitaire Méditerranée Infection, Marseille, France
[b] Institut Hospitalo-Universitaire Méditerranée Infection, Marseille, France
[c] Aix-Marseille Université, Institut de Recherche pour le Développement (IRD), UMR Microbes Evolution Phylogeny and Infections (MEPHI), Institut Hospitalo-Universitaire Méditerranée-Infection, Marseille, France
[d] Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Keywords:
Bacteria
Pangenome
Genomics
Genetic diversity
Clinical microbiology

ABSTRACT

Thanks to the progress and decreasing costs in genome sequencing technologies, more than 250,000 bacterial genomes are currently available in public databases, covering most, if not all, of the major human-associated phylogenetic groups of these microorganisms, pathogenic or not. In addition, for many of them, sequences from several strains of a given species are available, thus enabling to evaluate their genetic diversity and study their evolution. In addition, the significant cost reduction of bacterial whole genome sequencing as well as the rapid increase in the number of available bacterial genomes have prompted the development of pangenomic software tools. The study of bacterial pangenome has many applications in clinical microbiology. It can unveil the pathogenic potential and ability of bacteria to resist antimicrobials as well identify specific sequences and predict antigenic epitopes that allow molecular or serologic assays and vaccines to be designed. Bacterial pangenome constitutes a powerful method for understanding the history of human bacteria and relating these findings to diagnosis in clinical microbiology laboratories in order to optimize patient management.

1. Background

Bacteria are single-cell microorganisms that live in a wide array of environments [1]. Their adaptation to these various ecosystems paralleled their genetic evolution [2], leading to an extraordinary chromosomic and phenotypic diversification, including at the species level [3]. The development of complete genome sequencing as early as 1995, followed ten years later by the development of high-throughput whole genome sequencing, unprecedented access to the genetic information and bacterial evolution has been made possible [4–6]. In parallel, a renewed interest in culture methods, notably the culturomics strategy, has enabled a significant increase in the number of new human-associated bacterial species [7–9]. These progresses resulted in the availability of an increasingly high number of bacterial genome sequences. As 1st of April 2020, more than 250,000 bacterial genomes, encompassing most human pathogens and many commensals, are available in public databases. However, it is clear that the sequence of a single genome does not reflect the whole genetic variability within a bacterial species. Among the progresses permitted by genomic analyses, pangenome studies have enabled investigating the genetic diversity of bacteria at the species level [6,10,11]. Pangenomics was first developed in 2005 by Tettelin et al. to study multiple pathogenic isolates of Streptococcus agalactiae [12]. These authors were the first to describe a core genome made of genes common to all strains within a species, and a dispensable genome that was comprised of genes diversely present in some strains. Later, the pangenome of a species was defined as the sum of the core genome (all conserved genes including essential gene families), the dispensable genome (accessory genes) and unique genes specific of a given strain [11,13]. It has also been demonstrated that the pangenome size may significantly vary according to the bacterial species considered [14–16]. Sympatric bacteria that live collectively in contact with other microorganisms in the same ecological niche typically exhibit open pangenomes, i. e., pangenomes in which the number of genes constantly increases with the inclusion of genomes from new isolates of the same species. In contrast, allopatric species, living in isolated and restricted environments with limited access to external genetic resources, exhibit a closed pangenome, in which a limited number of strains is sufficient to complete the pangenomic analysis [10,11,17,18]. As examples, Escherichia coli and Bacillus anthracis exhibit open and closed pangenomes, respectively [11,13]. This genetic variation represents the main key to study and understand the structure and evolution of the bacterial pangenome. In recent years, the number

* Corresponding author. Institut Hospitalo-Universitaire Méditerranée-Infection, 19-21 Boulevard Jean Moulin, 13385, Marseille, cedex 05, France.
E-mail address: pierre-edouard.fournier@univ-amu.fr (P.-E. Fournier).

https://doi.org/10.1016/j.micpath.2020.104275
Received 1 April 2020; Received in revised form 22 May 2020; Accepted 25 May 2020
Available online 17 June 2020

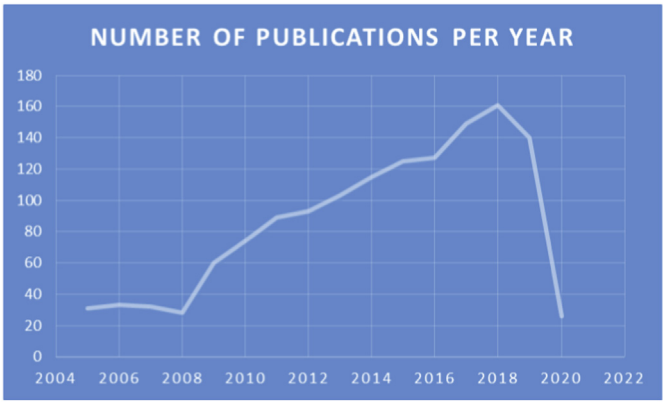

Fig. 1. Evolution of the annual number of publications with 'Bacteria', 'pangenome', 'pan genome' and 'pan-genome' as keywords in Pubmed (https://www.ncbi.nlm.nih.gov/pubmed/).

of pangenomic studies has significantly increased (Fig. 1). Such a progress was permitted by two main factors, the reducing cost of genome sequencing [19] and the development of improved pangenome analysis tools [20]. In 1995, when Fleishman et al. sequenced the first complete bacterial genome (*Haemophilus influenzae*), the sequencing cost was estimated around 878,000 dollars (48 cents/base pair for 1,830,137 bp) [4]. Currently, complete bacterial genomes can be sequenced for less than 600 dollars (https://www.illumina.com/science/customer-stories/icommunity-customer-interviews-case-studies/johnson-agrilife-interview-ag-novaseq.html). In parallel, more than 40 pangenomic tools are currently available in the form of local applications or online platforms.

Pangenomic studies have concrete applications in the clinical field, by allowing not only the characterization of human-associated pathogens, but also by offering a better understanding of their dynamics, transmission, virulence and resistance [16]. These roles were investigated in several types of infections, including skin and pulmonary infections, brain abscess, tuberculosis and other human diseases [21–24].

In this review, we examine the importance of pangenomics in clinical microbiology.

2. Pangenomic analysis tools

From 1995 to 2019, the cost of whole genome sequencing for a single bacterium has decreased around 1,464 times [4,25]. This significant cost reduction as well as the rapid increase of available bacterial genomes in databases have prompted the development of pangenomic software tools [26] (See Table 1). Panseq [27] and PanCGHweb [28] were among the first softwares that have been developed in 2010 [20]. Computational pangenome analysis tools can be categorized into online platforms (websites) and applications that are downloadable on personal computers. Online platforms are user-friendly but may be limited to a certain number of genomes, whereas downloadable softwares may require informatics skills for installation and powerful computers to analyze large datasets [31]. In addition, some softwares may be available in both forms, like Roary [29] and PanX [30]. In 2019, the number of pangenomic tools has increased to more than 40 free-access applications available online (https://omictools.com/company and https://github.com/). Regarding online tools, few clicks may be sufficient to generate a bacterial pangenome within minutes, like with PanX [30]. However, other online softwares require more time, notably due to the necessity to upload sequences in specific formats, such as gene-finding format (GFF) as recommended by the Roary pipeline [29] in the online Galaxy server [31]. Regarding local applications, the upload of annotated genome files may also be mandatory, as is the case for Fasta amino acid (FAA) files required by the ProteinOrtho [32] and Get Homologues [33] softwares.

Table 1
Online bioinformatics platforms or downloadable softwares for pangenomic analysis.

Tool name	Online platform (O) or downloadable tool (D)	References
1. PanX	O, D	[30]
2. PanOCT	D	[124]
3. PanRGP	D	https://microscope.readthedocs.io/en/stable/content/compgenomics/panRGP.html
4. PGAweb	O	[125]
5. Roary	O, D	[29]
6. ProteinOrtho	D	[32]
7. Get Homologues	D	[33]
8. PGAP	D	[126]
9. Panseq	D	[27]
10. PanViz	D	[127]
11. MetaPGN	D	[128]
12. PanCGHweb	O	[28]
13. Pancake	D	[129]
14. PPanGGOLiN	D	[130]
15. PanGet	D	[131]
16. PanArray	D	https://omictools.com/panarray-tool
17. BGDMdocker	D	[132]
18. MSPminer	D	[133]
19. PanPhlAn	D	https://omictools.com/panphlan-tool
20. PIRATE	D	[134]
21. Piggy	D	[135]
22. BPGA-Pan	D	[136]
23. Pan4Draft	D	[137]
24. PanACEA	D	[138]
25. PanWeb	O	[139]
26. PanFP	D	[140]
27. Pan-tetris	D	[141]
28. NGSpanPipe	D	[142]
29. PanFunPro	D	[143]
30. PGAT	D	[144]
31. EDGAR	D	[145]
32. ITEP	D	[146]
33. PanGP	D	[147]
34. LS-BSR	D	[148]
35. microPan	D	[149]
36. PanCoreGen	D	[150]
37. Pan Delos	D	[151]
38. SaturnV	D	https://github.com/ejfresch/saturnV
39. PANINI	D	[152]
40. Parsnp	D	[153]

Following pangenome analysis by the previously cited softwares, other online applications can be used to visualize the output data, like Phandango [34] or Morpheus (https://software.broadinstitute.org/morpheus/). Currently, thanks to the availability of dedicated online softwares, generating a pangenome analysis is easy, provided that certain specific steps are followed. These latter include downloading sequences of interest in an annotated format from genome databases, uploading all files, including the genomes generated in-house, in an appropriate format in the online tool, submitting the job and then visualizing the results.

3. *Role of horizontal gene transfer in pangenome*

The bacterial pangenome may undergo many variations during evolution [35]. The genetic content variability of prokaryotes [36,37] can notably result from genetic exchanges with its environment [38]. The process of gene trafficking between organisms with no parental relationship is defined as horizontal gene transfer (HGT) [16]. HGT was demonstrated as the main way for bacteria to acquire new genes in comparison to gene duplication [35,38]. HGT can occur via several

mechanisms [17,39], including bacterial transformation (DNA uptake) [40], transduction (DNA transfer via bacteriophages) [41] and conjugation (DNA transfer via type IV secretion system) [42]. Moreover, genes can be relocated by selfish genetic elements [43]. The rate of HGT varies greatly between allopatric and sympatric bacteria, and may also vary for a given bacterium when its ecological niche is modified [17,44–46]. As example, Freschi et al. studied the contribution of HGT in antimicrobial resistance and virulence mechanisms from the human pathogen Pseudomonas aeruginosa. These authors performed a pangenomic analysis based on 1,311 high-quality genomes of P. aeruginosa and characterized accessory genes, estimated of the prevalence of phages and identified plasmidic genes. Furthermore, they identified the genera Sinorhizobium, Ralstonia, Escherichia and Klebsiella as having the greatest rate of HGT events with P. aeruginosa. 5% of detected HGT-acquired genes were antimicrobial-resistance genes and 12% included sequences matching known virulence genes [47]. Using pangenomics, Dumas et al. also demonstrated the role of HGT in the adaptation of mycobacteria to new environments and hosts during evolution [48]. In particular, this study of 41 genomes demonstrated that plasmids carrying ESAT-6 secretory (ESX) genes were a major driving force for the acquisition and the diversification of type VII systems in these microorganisms [48]. These examples highlight the interest of pangenomics to decipher HGT events between bacteria (Fig. 2).

4. Role of mobilome, virulome and resistome in the pangenome

4.1. Mobilome

The bacterial mobilome has been defined as the repertoire of all mobile genetic elements (MGEs), notably plasmids, bacteriophages and transposons (Fig. 2) [49–51]. These MGEs can circulate intra-genomically (transposons) or inter-genomically (plasmids, bacteriophages, transposons) and play a significant role in bacterial evolution. In particular, they can allow adaptation to specific environments and increased genetic variability through the acquisition or loss of DNA sequences [18,51]. The genetic material carried by MGEs can contribute to the emergence of new pathogens (*Bacillus anthracis*, *Vibrio cholerae*) and drug resistance markers (*Escherichia coli*) [49,52–54]. In Pseudomonas aeruginosa, Freschi et al. deciphered the HGT mechanisms that underlie its genetic diversity. MGEs were detected using a pangenome matrix based on 1,311 high-quality P. aeruginosa genomes [47]. A total of 15, 2,017 and 2,177 phage genes were identified in the core genome, dispensable genome and unique genes, respectively (2.3%, 7.6% and 8%, respectively). In parallel, 102, 4,410 and 5,411 plasmid genes were detected among core, accessory and unique genes, respectively (15%, 17% and 20%, respectively). A pangenome of 17 sequenced genomes of brain abscess-associated Streptococcus intermedius was generated, in order to decipher the genetic mechanisms involved in the onset of this disease [24]. From the pangenome matrix, 14 bacteriophages (Bacillus phage G, Streptococcus phage phiARI0131-1, Lactobacillus phage PLE2, Campylobacter phage PC14, Streptococcus phage phiARI0468-1 …) were detected in addition to genomic islands containing a type seven secretion system (T7SS) and transposons carrying hyaluronidase and virulence-inducing antigens [24,55,56]. In total, these studies showed that pangenomic analyses facilitated the detection of MGEs between different bacterial repertoires at the species level.

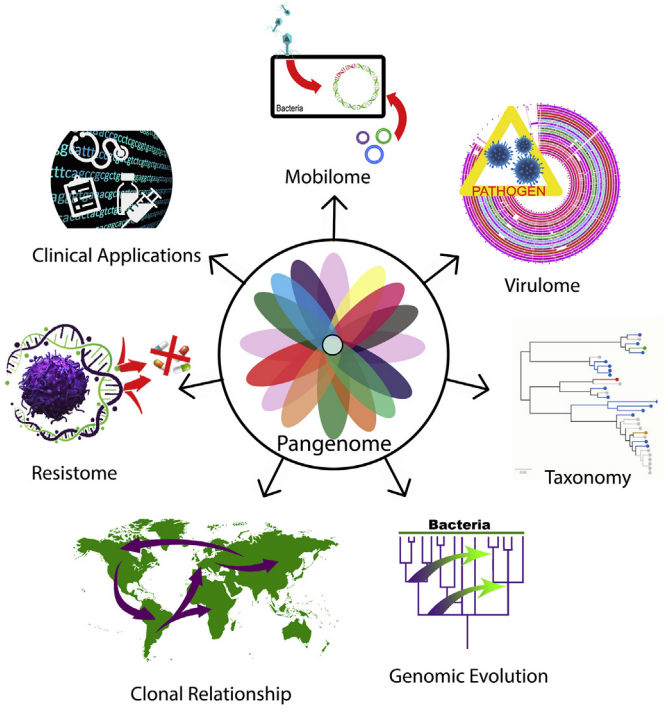

Fig. 2. Schematic representation of the bacterial pangenome applications.

4.2. Resistome

The emergence of antibiotic resistance in microorganisms is a global health threat. Resistance of microorganisms to antimicrobial agents can be caused by various mechanisms, mostly gene acquisition or mutation [57]. Wright et al. defined the resistome of a bacterial genome as all antibiotic resistance-related genes. These include those that encode resistance mechanisms, that may already be present in the genome and that are expressed or not, and those that are imported by HGT or mutated [58] (Fig. 2). Pangenomic analysis may also help in identifying resistance markers that play an important role in antimicrobial resistance by distinguishing between those present in all strains and those that are distributed differently [59–62]. Pangenomic results can also help designing antimicrobial combinations that are not classically used and subsequently improve the patient management. In a pangenomic analysis of 59 strains from the *Achromobacter* genus, Jeukens et al. were able, in comparison with the Comprehensive Antibiotic Resistance Database (CARD), to show that strains obtained from clinical specimens had more resistance genes than the others and that extra-genes were acquired by HGT [63]. In a large pangenomic study of *E. coli* comparing 4,022 genomes, Goldstone and Smith identified 2,172 antimicrobial-resistance genes that they classified into a core (50 genes) and accessory resistomes (2,122 genes) [64]. It was observed that the core resistome of *E. coli* consists essentially of multi-drug efflux pumps that can confer non-specific resistance [64]. Moreover time-dependent resistance transmission was noticed, when evaluating the isolation dates of implicated strains [64].

Recently, the detection of antimicrobial resistance in pangenomic analyses was improved by the use of machine learning. The machine learning strategy enables computers to improve automated biological predictions (prediction of biological activity, structure of interest, target sequence …) using data from previous experiences [65,66]. In 2018, an *E. coli* model was used by Moradigaravand et al. and Her et al. [67,68]. For the first example, a large pangenomic analysis of 1,936 *E. coli* strains was carried out to assess the ability of four predictive machine learning models (Random forest classifier, Gradient boosted decision trees, Deep neural networks and Rule-based baseline) to predict resistance to 11 antibiotics relying on different variables (population structure, isolation year and single nucleotide polymorphism data) [68]. The gradient boosted decision tree machine learning model was the best in comparison with other models tested with an average accuracy of 0.91 (average accuracy for predicting resistance and employing a number of possible combinations of gene presence, population structure and year of isolation). In a second example, Her et al. used a machine learning approach based on genetic algorithms from 59 *E. coli* strains [67]. They used four different predictive machine learning

models (Support Vector Machine, Naïve Bayes, Random Forest and Adaboost) to analyze genetic and phenotypic data related to 38 antimicrobial resistance genes. The authors demonstrated that only 68 of the total 13,076 accessory genes were annotated as having anantimicrobial activity with a high predictive accuracy [67]. This study referred for the first time to the pangenome as an antimicrobial resistance gene activity benchmark [67]. Another pangenome-based machine learning was generated from the genome sequences of 1,595 *Mycobacterium tuberculosis* strains [23]. One of the authors' objectives was to analyze and decipher the gene content of all sequences aiming to study the diverse signatures of antimicrobial resistance gene evolution to 13 antibiotics [23]. A high conservation of antimicrobial resistance genes to isoniazid, rifampicin, streptomycin and ethambutol was detected in the pangenome among 1,282 strains. In contrast, 946 strains were resistant to isoniazid and rifampicin only [23].

These studies, among many others [69–72], underscore the importance of pangenomic analyses to reach a better understanding of how antibiotic resistance is evolving among microbial populations. These data can also be used to optimize treatment regimens and antibiotic combinations.

4.3. Virulome

Virulence markers are specific genes conferring pathogenic bacteria the capacity to infect, adapt, replicat and persist in humans [73,74]. The set of genes contributing to the pathogenicity of a bacterium has been defined as its virulome [75]. Identifying the virulome of a bacterium is an important issue to improve the understanding of clinicians and researchers on pathogenic mechanisms and, subsequently enable the development of specific preventive and therapeutic approaches [76,77]. In recent years, bacterial pathogenicity studies have greatly benefited from the availability of complete bacterial genome sequences and virulence factor databases [78,79]. However, the study of a single bacterial genome can neither reveal all the virulence mechanisms possessed by this species nor explain how the virulome was acquired and evolved. Therefore, the comparative genomic analysis of multiple independent isolates is an indispensable issue to improve the knowledge of virulence within a bacterial species [11,48,80]. As described by Beye et al., the pangenome analysis of multiple *Raoultella ornytholica* isolates (13 genomes) has highlighted a specific genomic island composed of various genes encoding a type IVa secretion system (SS) possibly associated to a strain causing a chronic prosthetic joint infection in an immunocompetent patient [81]. More recently, the pangenomic analysis of 36 *Mycobacterium tuberculosis* (Mtb) isolates revealed 67 "super core genes" (SCG) present in all studied genomes [82]. Of these, 28 SCG genes were mostly related to pathogenesis, including *PPE19* (encoding proline-proline-glutamate) and *plcC* (encoding phospholipase C) genes that are involved in the survival of Mtb in macrophages and the bypass of phagosomal vacuoles, respectively [74,83]. In another study, the investigation of dispensable genes in the Mtb species revealed more genes associated with virulence in Mtb lineage 2, which could explain why this lineage is one of the most virulent [84]. Furthermore, pangenomic analyses of multiple pathovars were demonstrated to provide a better understanding of pathogenic mechanisms causing diverse clinical pathologies [85]. For instance, Rasko et al. demonstrated that only few specific *E. coli* virulence genes were conserved among distinct pathovar isolates [86], and that *E. coli* strains harboring prophages and phage elements exhibited a rapid evolution in comparison with other pathovar isolates [87,88]. A pangenomic study was used to compare the genomes from 13 *Clostridium botulinum* strains and characterize the virulent potential of the toxin/antitoxin modules in order to develop novel therapies against antibiotic resistance [89]. Bhardwaj et al. succeeded to unveil the presence of 120 genes encoding seven toxin/antitoxin families (*VapI*, *HipB*, *MazF*, *PemK*, *SpotVB_AbrB*, *phd/doc* and *ParE/D*) through the analysis of the 13 *Clostridium botulinum* strain pangenome [89]. Finally, in order to determine whether genetic variations between carriage and invasive strains could influence their virulence, Obolski et al. analyzed the whole genome sequences of 378 invasive *Streptococcus pneumoniae* isolates causing bacteremia [90]. The total pangenome yielded 9,032 genes including 43 (*hpp1* to *hpp37*, *ydcP_1*, *lytB*, *lox*, 2 copies of *cpsA* and *bgaA*) that were associated to virulence and especially to invasive pneumococcal disease. Of these, some genes were located into or around the capsule polysaccharide synthesis locus [90]. Therefore, pangenomic analyses are considered as efficient strategies to compare multiple bacterial isolates in order to screen the presence of novel virulence factors and better understand the pathogenesis at the species level. However, to validate the effect of virulence genes is necessary to complement the study by *in vitro* and *in vivo* experimentations [91].

5. Clinical applications of bacterial pangenomics

5.1. Impact of pangenome analysis on clonal relationship

The use of pangenomics offers the opportunity to highlight the adaptation of bacterial clones to various ecological niches in different geographical sites and their relationships. Hence, studying the pangenome of a species may provide information about its lifestyle and its ability to enclose clonal clusters or not, notably for human pathogenic microorganisms [92]. In addition, pangenome may serve to understand and study the clonal diversity of bacterial species [93] (Fig. 2). As an example, multi-locus sequence typing (MLST) based on the core genome of S. aureus demonstrated that various clones were able to spread in different geographic locations worldwide. In particular, this enabled to study the evolution of the methicilin-resistant clonal complex 5 of S. aureus, which includes many prevalent clones causing hospital-associated infections in the western hemisphere [94]. In 2014, Choo et al. investigated the clinical isolates of the emerging human pathogen Mycobacterium abscessus [22]. Using core genome single nucleotide polymorphism (SNP), they reported the high similarity between isolates of this virulent human bacterium that causes skin and soft tissue infections, in different geographical sites. In addition, the authors were able to detect the migration of Mycobacterium massiliense isolates from the Indian subcontinent to Southeast Asia then to Europe and the USA [22]. Pangenomic analysis was also applied to Propionibacterium acnes, another major human skin bacterium which has been implicated in acne [21]. A clonal expansion was demonstrated for the P. acnes species, even in individual or different microbiomes, which explains the ability of this species to exhibit a sustained pathogenic potential [21]. Pangenomic analyses may also be used for genotyping human bacterial pathogens in order to better understand their genetic diversity [93]. In a recent pangenomic study of 76 *C. burnetii* isolates, Hemsley et al. used multispacer sequence typing (MST) and single nucleotide polymorphism (SNP) to discriminate genotypes [95]. The authors could identify only one MST genotype, MST 20, in *C. burnetii* strains isolated from placental tissue (originating from abortions of ruminants in the United Kingdom), and were able to identify two new genotypes (MST32 and MST 33) for the first time in the UK among *C. burnetii* strains from goats. Using core genome SNP analysis, they were also able to demonstrate that *C. burnetii* strains were distributed into seven phylogenetic clades, and that cattle and sheep strains from the UK were clustered in the group of European and US *C. burnetii* strains [95]. In contrast, strains isolated from goats in the UK were grouped in a cluster containing human Q fever outbreak-causing strains [95]. Consequently, core/pangenome analyses help researchers understand and predict the ability of bacterial species to adapt to, and be pathogenic to various ecosystems.

5.2. New vaccine candidates

Faced with a pathogen, the human body reacts by activating its immune system. This immune reaction is not random, it is specific

against certain antigens covering the pathogen, and vaccines act by exposing the body to similar antigens. Conventional vaccines consist of injections of either attenuated or inactivated organisms that can induce an immune response without causing the disease [96]. Vaccines can also be an injection of only the antigenic components (subunit vaccine) of an organism and finally can be as toxoid vaccines [97]. However, the development of these vaccines is time-consuming, difficult to produce in large scale, the studied organism must be grown under laboratory conditions and the antigenic variability of some species may limit their efficiency [98]. In 2000, Rappuoli et al. proposed the concept of reverse vaccinology as an alternative to the usual vaccine design for serogoup B *Neisseria meningitidis* [98]. Reverse vaccinology is based on the computational analysis of a pathogen's genome in order to predict antigenic epitopes among which potential vaccine candidates may be selected. This screening step is followed by testing of the immunogenic properties of these antigens [98].

In 2005, Maione et al. were the first to use the pangenomic approach to identify putative antigens for the development of a vaccine against Group B *Streptococcus* (GBS) [99]. Given the high level of diversity between GBS strains, these authors first determined the core genome of eight strains representing the most important disease-causing genotypes. Then, they identified 589 genes coding surface-exposed proteins (396 core genes and 193 dipensable genes). Subsequently, they successfully expressed in mice 312 of these proteins which, following purification, were tested in a mouse model. Only four of the 312 proteins (three from the dispensable genome and one from the core genome) significantly increased the survival rate among challenged mice. As these four antigens were all variable between strains, the authors were unable to use a single antigen in the final vaccine formulation, but combined all four to produce a GBS vaccine that was protective against a wide range of strains [99]. Therefore, the pangenomic approach may solve the problem of antigenic variability of strains.

However, the first success of reverse vaccinology application was the development of a vaccine against group B *Neisseria meningitidis* (GBNM), a major human pathogen that had eluded vaccine development using traditional methods for decades [100,101]. The main reason for this delayed vaccine development is the antigenic variability among group B *N. meningitidis* strains [102]. Nearly 600 putative antigens were predicted from genomic analyses, among which 350 could be expressed in mice. Of these, 4 antigens were combined to produce the first GBNM vaccine that was commercialized in 2012.

Currently, other attempts to develop vaccines using reverse vaccinology are ongoing for several human pathogens including multidrug-resistant *Acinetobacter baumannii*, group A *Streptococcus*, *Streptococcus pneumoniae* and pathogenic *Escherichia coli* [103,104].

5.3. New therapeutic targets

Historically, the development of new drugs relied on two main strategies: whole-cell screening and target-based screening [105]. Whole-cell screening consists of testing, *in vitro*, the direct inhibitory effect of chemical or natural compounds on the growth of intact bacterial cells [106]. The target-based screening relies on the identification of potential molecular targets that have essential roles in the disease, and the design of a drug that interacts with these targets [107]. A valuable drug target should exhibit a low resistance rate and be testable *in vitro*, but most importantly, it should have an essential biological role or function (for example, a gene that maintains the viability of the organism is considered more essential than a gene necessary for growth because targeting the former will cause cell death, but for the latter it will only cause stasis). Besides, it should be conserved among strains from the species and not present in the human host in order to cause an inhibitory effect on the pathogen(s) with minimal side effects for humans [108]. Analyzing the pangenome is a good approach to determine the genes coding essential functions. The core genome of an organism contains genes that are conserved among strains and the majority of these genes are related to essential housekeeping functions, cell envelope, regulatory roles, transport and binding proteins [109]. Therefore, the core genome is a pool of potential drug targets, and comparing it to the human genome avoids potential targets that are present in the human host. Although there are no antimicrobial drugs on the market from a pangenomic analysis to date, there are various published works that have shown the potential of the pangenome to identify theoretical potential drug targets in pathogens such as *Clostridium botulinum* [89], *Helicobacter pylori* [110] and leptospirosis [111]. A recent study used the pangenome approach to identify putative new targets for developing future drugs against *Corynebacterium diphtheriae* [112]. The genomes from 13 *C. diphtheriae* strains were first aligned and the core genome was defined as containing 463 protein-coding sequences that share 95–100% sequence similarity between all strains. After a comparison with the Database of Essential Genes [113] and the elimination of genes encoding proteins homologous to human proteins, the authors identified 8 essential non-host proteins that could be used as potential drug targets. *In silico* screening and molecular docking was further performed on these eight proteins in order to determine the compounds that are active on them [112]. The results of all these studies may allow the development of therapeutic agents, but further *in vitro* and *in vivo* experimental studies are needed.

5.4. New diagnostic tools

Pangenomics can be used with the purpose of designing new diagnostic methods that can facilitate the screening and targeting of human pathogenic bacteria [114]. For instance, conserved sequences can be extracted from the core genome in order to design molecular systems (PCR or qPCR) enabling the identification of the target pathogen in clinical specimens. These molecular systems based on housekeeping genes as the *16S rRNA* and RNA polymerase beta subunit (*rpoB*) genes were expansively used to study the bacterial communities and explore their diversity and structure [115]. These molecular markers have been successfully used to identify bacterial isolates and admitted by taxonomists to classify all bacterial taxa [116]. However, recent studies have demonstrated that these genes cannot be highly accurate in term of distinction and differentiation between certain bacterial species [117]. For example, Ho et al. generated an automated pangenome analysis of *Salmonella enterica* serovar Typhi to identify and select specific PCR targets for directly detecting and identifying the pathogen without requiring gene sequencing [118]. By comparing the genomes from 11 *S. enterica* serovar Typhi strains, the authors identified 9 conserved intergenic region sequences specific of this serovar [118]. In 2017, Laing et al. performed a large pangenomic study based on 4,939 *Salmonella enterica* strains in order to identify species-specific and predictive markers for *Salmonella enterica* serovars [119]. The authors detected 207, 192, 135, 134, 93 and 9 subspecies-specific markers in *S. enterica* subsp *arizonae*, *indica*, *salamae*, *houtenae*, *diarizonae* and *enterica*, respectively. Furthermore, among 4870 *S. enterica* subsp. *enterica* genomes, many core-regions representing universal markers were identified in ten different serovars (Typhi, Typhimurium, Enteritidis, Heidelberg, Paratyphi, Kentucky, Agona, Weltevreden, Bareilly and Newport). The specificity of these markers was confirmed by comparison with 3,984 genomes in the EnteroBase database (https://enterobase.warwick.ac.uk/species/index/senterica) [119]. In addition, the pangenome matrix may be used to predict antigenic proteins that may serve to develop new serological assays. As an example, D'amato et al. identified putative antigenic epitopes in the core genome from *Coxiella burnetii* [93]. Then, a combination of functional annotation of these proteins and proteomic experiments has enabled the search for antigenic candidates suitable for serodiagnosis of Q fever [93]. Comparison of these proteins to the Clusters of Orthologous Groups (Cogs) database classified them within the DNA replication recombination and repair, translation and post-translational modification membrane

Table 2
Examples of published pangenomic studies.

Microorganism	Year	Bioinformatic tool	Pangenomic study
Streptococcus agalactiae	2005	–	[12]
Mycobacterium abscessus	2012	Panseq	[154]
Bifidobacterium animalis	2013	PGAP	[155]
Bacillus anthracis	2014	OrthoMCL	[92]
Streptococcus mutans	2014	Panseq	[156]
Escherichia and Shigella	2015	USEARCH	[11]
Mycobacterium abscessus	2015	PGAP	[22]
Bifidobacterium longum	2015	PGAP	[123]
Mycobacterium avium	2015	ProteinOrtho	[157]
Lactobacillus reuteri	2015	Panseq	[158]
Yersinia enterocolitica	2015	Roary	[159]
Mycobacterium africanum	2016	Roary	[160]
Lactobacillus kunkeei	2016	Roary	[161]
Staphylococcus epidermidis	2016	ProteinOrtho	[162]
Vibrio cholerae	2016	GET_HOMOLOGUES	[163]
Achromobacter spp.	2017	SaturnV	[63]
Corynebacterium diphtheriae	2017	EDGAR	[112]
Salmonella enterica	2017	Panseq	[119]
Helicobacter pylori	2017	Roary	[120]
Bacillus cereus	2017	Roary	[164]
Pseudomonas syringae	2017	GET_HOMOLOGUES	[165]
Mycobacterium tuberculosis	2018	CD-HIT	[23]
Escherichia coli	2018	CD-HIT	[67]
Escherichia coli	2018	Roary	[68]
Raoultella ornithinolytica	2018	Roary	[81]
Lactobacillus spp.	2018	GET_HOMOLOGUES	[121]
Vibrio diabolicus	2018	GET_HOMOLOGUES	[166]
Streptococcus sanguinis	2018	PGAP	[167]
Streptococcus intermedius	2019	Roary	[24]
Pseudomonas aeruginosa	2019	SaturnV	[47]
Coxiella burnetii	2019	BPGA	[95]
Akkermansia muciniphila	2019	Roary	[168]
Streptococcus pneumoniae	2019	Roary	[169]
Staphylococcus aureus	2019	Roary	[170]
Listeria monocytogenes	2019	Roary	[171]
Mycobacterium tuberculosis	2019	ProteinOrtho	[172]
Acinetobacter seifertii	2019	GET_HOMOLOGUES	[173]
Klebsiella pneumoniae	2020	Roary	[174]
Streptococcus pneumoniae	2020	Roary	[175]
Klebsiella pneumoniae	2020	EDGAR	[176]
Stenotrophomonas maltophilia	2020	Roary	[177]
Acinetobacter baumannii	2020	Roary	[178]
Salmonella infantis	2020	Roary	[179]
Pseudomonas aeruginosa	2020	Roary	[180]
Serratia marcescens	2020	Roary	[181]
Staphylococcus aureus	2020	Roary	[182]
Enterococcus faecium	2020	Roary	[183]

proteins, DNA denaturation in stress conditions and heat shock protein categories. These proteins were chosen to act as different potential Q fever markers [93].

6. Impact on taxonomy

Recent studies have shown that pangenome can be used successfully as a powerful bacterial taxonomy tool [120–122]. This argument was supported by the fact that pangenome analyses encompass the whole bacterial genetic information. Usually, the authors mainly used the core genome or the core/pangenome ratio [122] to identify and classify the species studied. In 2015, a pangenome analysis of 33 Bifidobacterium longum strains was generated in order to investigate the genomic diversity and phylogeny within this species [123]. In a core genome-based phylogenetic tree, two major clusters (clades 1 and 2) were identified. Clade 1 included 25 strains that were able to metabolize plant carbohydrates [123]. Clade 2 included eight strains and was divided into 3 phylogenetic groups (B, C and D), the latter only made of animal strains. In group C, the authors were able to identify a new B. longum subspecies originating from China [123]. In 2018, Inglin et al. used a pangenome analysis to study Lactobacillus species [121]. The study was based on 98 complete and 202 draft genomes from Lactobacillus species. The authors concluded that the L. delbrueckii subspecies could be re-classified according to their genomic content in environmental functions and their ecological niche, constituting a parapatric-like speciation model [121]. In a recent study, Caputo et al. investigated the pangenome from Klebsiella species in order to describe the taxonomic profile of K. pneumoniae. Using 12 genomes (8 genomes from K. pneumoniae, 2 from K. oxytoca, 1 from *K. variicola* and 1 from *K. mobilis*), the authors demonstrated that K. pneumoniae subsp. ozaenae and K. pneumoniae subsp. rhinoscleromatis exhibited a high genomic distance between them and with the remaining K. pneumoniae strains, suggesting that this discontinuous variation may classify these two subspecies into different species within Klebsiella genus [122]. These examples emphasize the importance of pangenome in bacterial taxonomy (Fig. 2).

7. Conclusion

In this review, we have demonstrated that studying the bacterial pangenome has a large spectrum of applications in clinical microbiology (Table 2). The pangenome is increasingly being investigated to explore the bacterial genetic content of pathogens and unveil their diversity in several clinical settings. Pangenome analyses can simultaneously provide a wealth of information about a human-associated bacterial species, including its ability to interfere with its environment and evolve by acquiring foreign genetic material, its pathogenic potential and its degree of resistance to antimicrobials. In addition, and although functional predictions should be confirmed using *in vitro/in vivo* inquiries, pangenomic analyses may allow designing vaccines and diagnostic tools. With the exponential increase in available genome sequences, pangenome analyses, which offer unprecedented access to the genetic diversity of human-associated bacteria, have the potential to be widely used in the coming years in clinical microbiology.

Authors' statements

Hussein Anani performed bibliographic search and drafted the manuscript, Rita Zgheib and Issam Hasni drafted the manuscript, Didier Raoult drafted and revised the manuscript, and Pierre-Edouard Fournier designed the study and drafted the manuscript.

Declaration of competing interest

The authors declare that they have no competing interest.

Acknowledgments

The study was supported by the Méditerranée Infection foundation, the National Research Agency under the program "Investissements d'avenir", reference ANR-10-IAHU-03 and by Région Provence-Alpes-Côte d'Azur and European funding FEDER PRIMI.

References

[1] R. Rosselló-Móra, R. Amann, Past and future species definitions for Bacteria and Archaea, Syst. Appl. Microbiol. 38 (2015) 209–216, https://doi.org/10.1016/j.syapm.2015.02.001.
[2] C.C. Boutte, S. Crosson, Bacterial lifestyle shapes stringent response activation, Trends Microbiol. 21 (2013) 174–180, https://doi.org/10.1016/j.tim.2013.01.002.
[3] G.F. Webb, M.J. Blaser, Dynamics of bacterial phenotype selection in a colonized host, Proc. Natl. Acad. Sci. Unit. States Am. 99 (2002) 3135–3140, https://doi.org/10.1073/pnas.042685799.
[4] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, Whole-genome random sequencing and assembly of Haemophilus influenzae Rd, Science 269 (1995) 496–512, https://doi.org/10.1126/science.7542800.
[5] P.-E. Fournier, M. Drancourt, D. Raoult, Bacterial genome sequencing and its use in infectious diseases, Lancet Infect. Dis. 7 (2007) 711–723, https://doi.org/10.

1016/S1473-3099(07)70260-8.
[6] H. Tettelin, D. Riley, C. Cattuto, D. Medini, Comparative genomics: the bacterial pan-genome, Curr. Opin. Microbiol. 11 (2008) 472–477, https://doi.org/10.1016/j.mib.2008.09.006.
[7] J.-C. Lagier, F. Armougom, M. Million, P. Hugon, I. Pagnier, C. Robert, F. Bittar, G. Fournous, G. Gimenez, M. Maraninchi, J.-F. Trape, E.V. Koonin, B. La Scola, D. Raoult, Microbial culturomics: paradigm shift in the human gut microbiome study, Clin. Microbiol. Infect. 18 (2012) 1185–1193, https://doi.org/10.1111/1469-0691.12023.
[8] J.-C. Lagier, P. Hugon, S. Khelaifia, P.-E. Fournier, B. La Scola, D. Raoult, The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota, Clin. Microbiol. Rev. 28 (2015) 237–264, https://doi.org/10.1128/CMR.00014-14.
[9] J.-C. Lagier, S. Khelaifia, M.T. Alou, S. Ndongo, N. Dione, P. Hugon, A. Caputo, F. Cadoret, S.I. Traore, E.H. Seck, G. Dubourg, G. Durand, G. Mourembou, E. Guilhot, A. Togo, S. Bellali, D. Bachar, N. Cassir, F. Bittar, J. Delerce, M. Mailhe, D. Ricaboni, M. Bilen, N.P.M. Dangui Nieko, N.M. Dia Badiane, C. Valles, D. Mouelhi, K. Diop, M. Million, D. Musso, J. Abrahão, E.I. Azhar, F. Bibi, M. Yasir, A. Diallo, C. Sokhna, F. Djossou, V. Vitton, C. Robert, J.M. Rolain, B. La Scola, P.-E. Fournier, A. Levasseur, D. Raoult, Culture of previously uncultured members of the human gut microbiota by culturomics, Nat. Microbiol. 1 (2016), https://doi.org/10.1038/nmicrobiol.2016.203.
[10] D. Medini, C. Donati, H. Tettelin, V. Masignani, R. Rappuoli, The microbial pan-genome, Curr. Opin. Genet. Dev. 15 (2005) 589–594, https://doi.org/10.1016/j.gde.2005.09.006.
[11] L. Rouli, V. Merhej, P.-E. Fournier, D. Raoult, The bacterial pangenome as a new tool for analysing pathogenic bacteria, New Microbes New Infections 7 (2015) 72–85, https://doi.org/10.1016/j.nmni.2015.06.005.
[12] 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, I. Margarit y 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.B. 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. Unit. States Am. 102 (2005) 13950–13955, https://doi.org/10.1073/pnas.0506758102.
[13] C. Ku, S. Nelson-Sathi, M. Roettger, S. Garg, E. Hazkani-Covo, W.F. Martin, Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 10139–10146, https://doi.org/10.1073/pnas.1421385112.
[14] L.A. David, A.C. Materna, J. Friedman, M.I. Campos-Baptista, M.C. Blackburn, A. Perrotta, S.E. Erdman, E.J. Alm, Host lifestyle affects human microbiota on daily timescales, Genome Biol. 15 (2014) R89, https://doi.org/10.1186/gb-2014-15-7-r89.
[15] M. Anwesh, K.V. Kumar, M. Nagarajan, M.P. Chander, C. Kartick, V. Paluru, Elucidating the richness of bacterial groups in the gut of Nicobarese tribal community – perspective on their lifestyle transition, Anaerobe 39 (2016) 68–76, https://doi.org/10.1016/j.anaerobe.2016.03.002.
[16] J.O. McInerney, A. McNally, M.J. O'Connell, Why prokaryotes have pangenomes, Nat. Microbiol. 2 (2017) 17040.
[17] K. Georgiades, D. Raoult, Defining pathogenic bacterial species in the genomic era, Front. Microbiol. 1 (2010) 151, https://doi.org/10.3389/fmicb.2010.00151.
[18] S.M. Diene, V. Merhej, M. Henry, A. El Filali, V. Roux, C. Robert, S. Azza, F. Gavory, V. Barbe, B. La Scola, D. Raoult, J.-M. Rolain, The rhizome of the multidrug-resistant Enterobacter aerogenes genome reveals how new "killer bugs" are created because of a sympatric lifestyle, Mol. Biol. Evol. 30 (2013) 369–383, https://doi.org/10.1093/molbev/mss236.
[19] J.Y. Taylor, M.L. Wright, K.T. Hickey, D.E. Housman, Genome sequencing technologies and nursing: what are the roles of nurses and nurse scientists? Nurs. Res. 66 (2017) 198–205, https://doi.org/10.1097/NNR.0000000000000211.
[20] J. Xiao, Z. Zhang, J. Wu, J. Yu, A brief review of software tools for pangenomics, genomics, Proteomics Bioinf. 13 (2015) 73–76, https://doi.org/10.1016/j.gpb.2015.01.007.
[21] S. Tomida, L. Nguyen, B.-H. Chiu, J. Liu, E. Sodergren, G.M. Weinstock, H. Li, Pan-genome and comparative genome analyses of Propionibacterium acnes reveal its genomic diversity in the healthy and diseased human skin microbiome, mBio 4 (2013) e00003–13, https://doi.org/10.1128/mBio.00003-13 mBio.00003-13.
[22] S.W. Choo, W.Y. Wee, Y.F. Ngeow, W. Mitchell, J.L. Tan, G.J. Wong, Y. Zhao, J. Xiao, Genomic reconnaissance of clinical isolates of emerging human pathogen Mycobacterium abscessus reveals high evolutionary potential, Sci. Rep. 4 (2015) 4061, https://doi.org/10.1038/srep04061.
[23] E.S. Kavvas, E. Catoiu, N. Mih, J.T. Yurkovich, Y. Seif, N. Dillon, D. Heckmann, A. Anand, L. Yang, V. Nizet, J.M. Monk, B.O. Palsson, Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance, Nat. Commun. 9 (2018) 4306, https://doi.org/10.1038/s41467-018-06634-y.
[24] E. Issa, T. Salloum, B. Panossian, D. Ayoub, E. Abboud, S. Tokajian, Genome mining and comparative analysis of Streptococcus intermedius causing brain abscess in a child, Pathogens 8 (2019) 22, https://doi.org/10.3390/pathogens8010022.
[25] I.F. Bronner, M.A. Quail, Best practices for illumina library preparation, Curr. Protoc. Hum. Genet. 102 (2019) e86, https://doi.org/10.1002/cphg.86.
[26] T. Zekic, G. Holley, J. Stoye, Pan-genome storage and analysis techniques, Methods Mol. Biol. 1704 (2018) 29–53, https://doi.org/10.1007/978-1-4939-7463-4_2.
[27] C. Laing, C. Buchanan, E.N. Taboada, Y. Zhang, A. Kropinski, A. Villegas, J.E. Thomas, V.P.J. Gannon, Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions, BMC Bioinf. 11 (2010) 461, https://doi.org/10.1186/1471-2105-11-461.
[28] J.R. Bayjanov, R.J. Siezen, S.A.F.T. van Hijum, PanCGHweb: a web tool for genotype calling in pangenome CGH data, Bioinformatics 26 (2010) 1256–1257, https://doi.org/10.1093/bioinformatics/btq103.
[29] A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T.G. Holden, M. Fookes, D. Falush, J.A. Keane, J. Parkhill, Roary: rapid large-scale prokaryote pan genome analysis, Bioinformatics 31 (2015) 3691–3693, https://doi.org/10.1093/bioinformatics/btv421.
[30] W. Ding, F. Baumdicker, R.A. Neher, panX: pan-genome analysis and exploration, Nucleic Acids Res. 46 (2018) e5, https://doi.org/10.1093/nar/gkx977.
[31] E. Afgan, D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, B.A. Grüning, A. Guerler, J. Hillman-Jackson, S. Hiltemann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, D. Blankenberg, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update, Nucleic Acids Res. 46 (2018) W537, https://doi.org/10.1093/nar/gky379 –W544.
[32] M. Lechner, S. Findeiss, L. Steiner, M. Marz, P.F. Stadler, S.J. Prohaska, Proteinortho: detection of (co-)orthologs in large-scale analysis, BMC Bioinf. 12 (2011) 124, https://doi.org/10.1186/1471-2105-12-124.
[33] B. Contreras-Moreira, P. Vinuesa, GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis, Appl. Environ. Microbiol. 79 (2013) 7696–7701, https://doi.org/10.1128/AEM.02411-13.
[34] J. Hadfield, N.J. Croucher, R.J. Goater, K. Abudahab, D.M. Aanensen, S.R. Harris, Phandango: an interactive viewer for bacterial population genomics, Bioinformatics 34 (2018) 292–293, https://doi.org/10.1093/bioinformatics/btx610.
[35] T.J. Treangen, E.P.C. Rocha, Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes, PLoS Genet. 7 (2011) e1001284, , https://doi.org/10.1371/journal.pgen.1001284.
[36] S. Schneiker, O. Perlova, O. Kaiser, K. Gerth, A. Alici, M.O. Altmeyer, D. Bartels, T. Bekel, S. Beyer, E. Bode, H.B. Bode, C.J. Bolten, J.V. Choudhuri, S. Doss, Y.A. Elnakady, B. Frank, L. Gaigalat, A. Goesmann, C. Groeger, F. Gross, L. Jelsbak, L. Jelsbak, J. Kalinowski, C. Kegler, T. Knauber, S. Konietzny, M. Kopp, L. Krause, D. Krug, B. Linke, T. Mahmud, R. Martinez-Arias, A.C. McHardy, M. Merai, F. Meyer, S. Mormann, J. Muñoz-Dorado, J. Perez, S. Pradella, S. Rachid, G. Raddatz, F. Rosenau, C. Rückert, F. Sasse, M. Scharfe, S.C. Schuster, G. Suen, A. Treuner-Lange, G.J. Velicer, F.-J. Vorhölter, K.J. Weissman, R.D. Welch, S.C. Wenzel, D.E. Whitworth, S. Wilhelm, C. Wittmann, H. Blöcker, A. Pühler, R. Müller, Complete genome sequence of the myxobacterium Sorangium cellulosum, Nat. Biotechnol. 25 (2007) 1281–1289, https://doi.org/10.1038/nbt1354.
[37] J.P. McCutcheon, B.R. McDonald, N.A. Moran, Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont, PLoS Genet. 5 (2009) e1000565, , https://doi.org/10.1371/journal.pgen.1000565.
[38] H. Ochman, E. Lerat, V. Daubin, Examining bacterial species under the specter of gene transfer and exchange, Proc. Natl. Acad. Sci. Unit. States Am. 102 (2005) 6595–6599, https://doi.org/10.1073/pnas.0502035102.
[39] M. Juhas, Horizontal gene transfer in human pathogens, Crit. Rev. Microbiol. 41 (2015) 101–108, https://doi.org/10.3109/1040841X.2013.804031.
[40] N.-J. Krüger, K. Stingl, Two steps away from novelty - principles of bacterial DNA uptake: principles of bacterial DNA uptake, Mol. Microbiol. 80 (2011) 860–867, https://doi.org/10.1111/j.1365-2958.2011.07647.x.
[41] R.P. Novick, G.E. Christie, J.R. Penadés, The phage-related chromosomal islands of Gram-positive bacteria, Nat. Rev. Microbiol. 8 (2010) 541–551, https://doi.org/10.1038/nrmicro2393.
[42] K. Wallden, A. Rivera-Calzada, G. Waksman, Microreview: type IV secretion systems: versatility and diversity in function: diversity and versatility of the type IV secretion system, Cell Microbiol. 12 (2010) 1203–1212, https://doi.org/10.1111/j.1462-5822.2010.01499.x.
[43] J.A. Ågren, A.G. Clark, Selfish genetic elements, PLoS Genet. 14 (2018) e1007700, , https://doi.org/10.1371/journal.pgen.1007700.
[44] W.B. Whitman, The modern concept of the procaryote, J. Bacteriol. 191 (2009), https://doi.org/10.1128/JB.00962-08 2000–2005; discussion 2006-2007.
[45] V. Merhej, M. Royer-Carenzi, P. Pontarotti, D. Raoult, Massive comparative genomic analysis reveals convergent evolution of specialized bacteria, Biol. Direct 4 (2009) 13, https://doi.org/10.1186/1745-6150-4-13.
[46] C. Moliner, P.-E. Fournier, D. Raoult, Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution, FEMS Microbiol. Rev. 34 (2010) 281–294, https://doi.org/10.1111/j.1574-6976.2009.00209.x.
[47] L. Freschi, A.T. Vincent, J. Jeukens, J.-G. Emond-Rheault, I. Kukavica-Ibrulj, M.-J. Dupont, S.J. Charette, B. Boyle, R.C. Levesque, The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity, Genome Biol. Evol. 11 (2019) 109–120, https://doi.org/10.1093/gbe/evy259.
[48] E. Dumas, E. Christina Boritsch, M. Vandenbogaert, R.C. Rodríguez de la Vega, J.-M. Thiberge, V. Caro, J.-L. Gaillard, B. Heym, F. Girard-Misguich, R. Brosch, G. Sapriel, Mycobacterial pan-genome analysis suggests important role of plasmids in the radiation of type VII secretion systems, Genome Biol. Evol. 8 (2016) 387–402, https://doi.org/10.1093/gbe/evw001.
[49] L.S. Frost, R. Leplae, A.O. Summers, A. Toussaint, Mobile genetic elements: the agents of open source evolution, Nat. Rev. Microbiol. 3 (2005) 722–732, https://doi.org/10.1038/nrmicro1235.

[50] J.L. Siefert, Defining the mobilome, in: M.B. Gogarten, J.P. Gogarten, L.C. Olendzenski (Eds.), Horizontal Gene Transfer, Humana Press, Totowa, NJ, 2009, pp. 13–27, , https://doi.org/10.1007/978-1-60327-853-9_2.

[51] T.M. Ghaly, M.R. Gillings, Mobile DNAs as ecologically and evolutionarily independent units of life, Trends Microbiol. 26 (2018) 904–912, https://doi.org/10.1016/j.tim.2018.05.008.

[52] T.M. Koehler, Bacillus anthracis physiology and genetics, Mol. Aspect. Med. 30 (2009) 386–396, https://doi.org/10.1016/j.mam.2009.07.004.

[53] J.F. Turton, M. Doumith, K.L. Hopkins, C. Perry, D. Meunier, N. Woodford, Clonal expansion of Escherichia coli ST38 carrying a chromosomally integrated OXA-48 carbapenemase gene, J. Med. Microbiol. 65 (2016) 538–546, https://doi.org/10.1099/jmm.0.000248.

[54] I.B. Naser, M.M. Hoque, A. Abdullah, S.M.N. Bari, A.N. Ghosh, S.M. Faruque, Environmental bacteriophages active on biofilms and planktonic forms of toxigenic Vibrio cholerae: potential relevance in cholera epidemiology, PloS One 12 (2017) e0180838, , https://doi.org/10.1371/journal.pone.0180838.

[55] D. Willner, M. Furlan, R. Schmieder, J.A. Grasis, D.T. Pride, D.A. Relman, F.E. Angly, T. McDole, R.P. Mariella, F. Rohwer, M. Haynes, Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity, Proc. Natl. Acad. Sci. Unit. States Am. 108 (2011) 4547–4553, https://doi.org/10.1073/pnas.1000089107.

[56] I. Sitkiewicz, N.M. Green, N. Guo, L. Mereghetti, J.M. Musser, Lateral gene transfer of streptococcal ICE element RD2 (region of difference 2) encoding secreted proteins, BMC Microbiol. 11 (2011) 65, https://doi.org/10.1186/1471-2180-11-65.

[57] J.M. Munita, C.A. Arias, Mechanisms of antibiotic resistance, Microbiol. Spectr. 4 (2016), https://doi.org/10.1128/microbiolspec.VMBF-0016-2015.

[58] G.D. Wright, The antibiotic resistome: the nexus of chemical and genetic diversity, Nat. Rev. Microbiol. 5 (2007) 175–186, https://doi.org/10.1038/nrmicro1614.

[59] E. Zankari, H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F.M. Aarestrup, M.V. Larsen, Identification of acquired antimicrobial resistance genes, J. Antimicrob. Chemother. 67 (2012) 2640–2644, https://doi.org/10.1093/jac/dks261.

[60] S.K. Gupta, B.R. Padmanabhan, S.M. Diene, R. Lopez-Rojas, M. Kempf, L. Landraud, J.-M. Rolain, ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes, Antimicrob. Agents Chemother. 58 (2014) 212–220, https://doi.org/10.1128/AAC.01310-13.

[61] B. Jia, A.R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K.K. Tsang, B.A. Lago, B.M. Dave, S. Pereira, A.N. Sharma, S. Doshi, M. Courtot, R. Lo, L.E. Williams, J.G. Frye, T. Elsayegh, D. Sardar, E.L. Westman, A.C. Pawlowski, T.A. Johnson, F.S.L. Brinkman, G.D. Wright, A.G. McArthur, CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database, Nucleic Acids Res. 45 (2017) D566–D573, https://doi.org/10.1093/nar/gkw1004.

[62] S.M. Lakin, C. Dean, N.R. Noyes, A. Dettenwanger, A.S. Ross, E. Doster, P. Rovira, Z. Abdo, K.L. Jones, J. Ruiz, K.E. Belk, P.S. Morley, C. Boucher, MEGARes: an antimicrobial resistance database for high throughput sequencing, Nucleic Acids Res. 45 (2017) D574–D580, https://doi.org/10.1093/nar/gkw1009.

[63] J. Jeukens, L. Freschi, A.T. Vincent, J.-G. Emond-Rheault, I. Kukavica-Ibrulj, S.J. Charette, R.C. Levesque, A pan-genomic approach to understand the basis of host adaptation in achromobacter, Genome Biol. Evol. 9 (2017) 1030–1046, https://doi.org/10.1093/gbe/evx061.

[64] R.J. Goldstone, D.G.E. Smith, A population genomics approach to exploiting the accessory "resistome" of Escherichia coli, Microb. Genom. 3 (2017), https://doi.org/10.1099/mgen.0.000108.

[65] Y. Baştanlar, M. Ozuysal, Introduction to machine learning, Methods Mol. Biol. 1107 (2014) 105–128, https://doi.org/10.1007/978-1-62703-748-8_7.

[66] A.N. Lima, E.A. Philot, G.H.G. Trossini, L.P.B. Scott, V.G. Maltarollo, K.M. Honorio, Use of machine learning approaches for novel drug discovery, Expet Opin. Drug Discov. 11 (2016) 225–239, https://doi.org/10.1517/17460441.2016.1146250.

[67] H.-L. Her, Y.-W. Wu, A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the Escherichia coli strains, Bioinformatics 34 (2018), https://doi.org/10.1093/bioinformatics/bty276 i89–i95.

[68] D. Moradigaravand, M. Palm, A. Farewell, V. Mustonen, J. Warringer, L. Parts, Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data, PLoS Comput. Biol. 14 (2018) e1006258, , https://doi.org/10.1371/journal.pcbi.1006258.

[69] M. Pérez-Vázquez, J. Oteo, S. García-Cobos, B. Aracil, S.R. Harris, A. Ortega, D. Fontanals, J.M. Hernández, S. Solís, J. Campos, G. Dougan, R.A. Kingsley, Phylogeny, resistome and mobile genetic elements of emergent OXA-48 and OXA-245 *Klebsiella pneumoniae* clones circulating in Spain, J. Antimicrob. Chemother. 71 (2016) 887–896, https://doi.org/10.1093/jac/dkv458.

[70] A. Johnning, N. Karami, E. Tång Hallbäck, V. Müller, L. Nyberg, M. Buongermino Pereira, C. Stewart, T. Ambjörnsson, F. Westerlund, I. Adlerberth, E. Kristiansson, The resistomes of six carbapenem-resistant pathogens – a critical genotype–phenotype analysis, Microb. Genom. 4 (2018), https://doi.org/10.1099/mgen.0.000233.

[71] T. Pillonel, P. Nordmann, C. Bertelli, G. Prod' hom, L. Poirel, G. Greub, Resistome analysis of a carbapenemase (OXA-48)-Producing and colistin-resistant *Klebsiella pneumoniae* strain, Antimicrob. Agents Chemother. 62 (2018), https://doi.org/10.1128/AAC.00076-18 e00076-18,/aac/62/5/e00076-18.atom.

[72] C. Ng, B. Tan, X.-T. Jiang, X. Gu, H. Chen, B.W. Schmitz, L. Haller, F.R. Charles, T. Zhang, K. Gin, Metagenomic and resistome analysis of a full-scale municipal wastewater treatment plant in Singapore containing membrane bioreactors, Front. Microbiol. 10 (2019) 172, https://doi.org/10.3389/fmicb.2019.00172.

[73] A.S. Cross, What is a virulence factor? Crit. Care 12 (2008) 196, https://doi.org/10.1186/cc7127.

[74] M.A. Forrellad, L.I. Klepp, A. Gioffré, J. Sabio y García, H.R. Morbidoni, M. de la Paz Santangelo, A.A. Cataldi, F. Bigi, Virulence factors of the Mycobacterium tuberculosis complex, Virulence 4 (2013) 3–66, https://doi.org/10.4161/viru.22329.

[75] S. Kohler, V. Foulongne, S. Ouahrani-Bettache, G. Bourg, J. Teyssier, M. Ramuz, J.-P. Liautard, Nonlinear partial differential equations and applications: the analysis of the intramacrophagic virulome of Brucella suis deciphers the environment encountered by the pathogen inside the macrophage host cell, Proc. Natl. Acad. Sci. Unit. States Am. 99 (2002) 15711–15716, https://doi.org/10.1073/pnas.232454299.

[76] T.M. Fuchs, Molecular mechanisms of bacterial pathogenicity, Naturwissenschaften 85 (1998) 99–108, https://doi.org/10.1007/s001140050463.

[77] L. Cegelski, G.R. Marshall, G.R. Eldridge, S.J. Hultgren, The biology and future prospects of antivirulence therapies, Nat. Rev. Microbiol. 6 (2008) 17–27, https://doi.org/10.1038/nrmicro1818.

[78] L. Chen, J. Yang, J. Yu, Z. Yao, L. Sun, Y. Shen, Q. Jin, VFDB: a reference database for bacterial virulence factors, Nucleic Acids Res. 33 (2005) D325–D328, https://doi.org/10.1093/nar/gki008.

[79] 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, https://doi.org/10.1007/s10142-015-0433-4.

[80] M.J. Pallen, B.W. Wren, Bacterial pathogenomics, Nature 449 (2007) 835–842, https://doi.org/10.1038/nature06248.

[81] M. Beye, I. Hasni, P. Seng, C. Michelle, B. La Scola, D. Raoult, P.-E. Fournier, Genomic analysis of a Raoultella ornithinolytica strain causing prosthetic joint infection in an immunocompetent patient, Sci. Rep. 8 (2018) 9462, https://doi.org/10.1038/s41598-018-27833-z.

[82] T. Yang, J. Zhong, J. Zhang, C. Li, X. Yu, J. Xiao, X. Jia, N. Ding, G. Ma, G. Wang, L. Yue, Q. Liang, Y. Sheng, Y. Sun, H. Huang, F. Chen, Pan-genomic study of Mycobacterium tuberculosis reflecting the primary/secondary genes, generality/individuality, and the interconversion through copy number variations, Front. Microbiol. 9 (2018) 1886, https://doi.org/10.3389/fmicb.2018.01886.

[83] E. Dubnau, I. Smith, Mycobacterium tuberculosis gene expression in macrophages, Microb. Infect. 5 (2003) 629–637, https://doi.org/10.1016/s1286-4579(03)00090-x.

[84] M. Coscolla, S. Gagneux, Consequences of genomic diversity in Mycobacterium tuberculosis, Semin. Immunol. 26 (2014) 431–444, https://doi.org/10.1016/j.smim.2014.09.012.

[85] J. Chun, C.J. Grim, N.A. Hasan, J.H. Lee, S.Y. Choi, B.J. Haley, E. Taviani, Y.-S. Jeon, D.W. Kim, J.-H. Lee, T.S. Brettin, D.C. Bruce, J.F. Challacombe, J.C. Detter, C.S. Han, A.C. Munk, O. Chertkov, L. Meincke, E. Saunders, R.A. Walters, A. Huq, G.B. Nair, R.R. Colwell, Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae, Proc. Natl. Acad. Sci. Unit. States Am. 106 (2009) 15442–15447, https://doi.org/10.1073/pnas.0907787106.

[86] D.A. Rasko, M.J. Rosovitz, G.S.A. Myers, E.F. Mongodin, W.F. Fricke, P. Gajer, J. Crabtree, M. Sebaihia, N.R. Thomson, R. Chaudhuri, I.R. Henderson, V. Sperandio, J. Ravel, The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates, J. Bacteriol. 190 (2008) 6881–6893, https://doi.org/10.1128/JB.00619-08.

[87] L.M. Wick, W. Qi, D.W. Lacher, T.S. Whittam, Evolution of genomic content in the stepwise emergence of Escherichia coli O157:H7, J. Bacteriol. 187 (2005) 1783–1791, https://doi.org/10.1128/JB.187.5.1783-1791.2005.

[88] Y. Zhang, C. Laing, M. Steele, K. Ziebell, R. Johnson, A.K. Benson, E. Taboada, V.P. Gannon, Genome evolution in major Escherichia coli O157:H7 lineages, BMC Genom. 8 (2007) 121, https://doi.org/10.1186/1471-2164-8-121.

[89] T. Bhardwaj, P. Somvanshi, Pan-genome analysis of Clostridium botulinum reveals unique targets for drug development, Gene 623 (2017) 48–62, https://doi.org/10.1016/j.gene.2017.04.019.

[90] U. Obolski, A. Gori, J. Lourenço, C. Thompson, R. Thompson, N. French, R.S. Heyderman, S. Gupta, Identifying genes associated with invasive disease in S. pneumoniae by applying a machine learning approach to whole genome sequence typing data, Sci. Rep. 9 (2019) 4049, https://doi.org/10.1038/s41598-019-40346-7.

[91] E. Potvin, D.E. Lehoux, I. Kukavica-Ibrulj, K.L. Richard, F. Sanschagrin, G.W. Lau, R.C. Levesque, In vivo functional genomics of Pseudomonas aeruginosa for high-throughput screening of new virulence factors and antibacterial targets: Pseudomonas in vivo functional genomics, Environ. Microbiol. 5 (2003) 1294–1308, https://doi.org/10.1046/j.1462-2920.2003.00542.x.

[92] L. Rouli, M. Mbengue, C. Robert, M. Ndiaye, B. La Scola, D. Raoult, Genomic analysis of three African strains of Bacillus anthracis demonstrates that they are part of the clonal expansion of an exclusively pathogenic bacterium, New Microbes New Infections 2 (2014) 161–169, https://doi.org/10.1002/nmi2.62.

[93] F. D'Amato, C. Eldin, D. Raoult, The contribution of genomics to the study of Q fever, Future Microbiol. 11 (2016) 253–272, https://doi.org/10.2217/fmb.15.137.

[94] L. Challagundla, J. Reyes, I. Rafiqullah, D.O. Sordelli, G. Echaniz-Aviles, M.E. Velazquez-Meza, S. Castillo-Ramírez, N. Fittipaldi, M. Feldgarden, S.B. Chapman, M.S. Calderwood, L.P. Carvajal, S. Rincon, B. Hanson, P.J. Planet, C.A. Arias, L. Diaz, D.A. Robinson, Phylogenomic classification and the evolution of clonal complex 5 methicillin-resistant Staphylococcus aureus in the western hemisphere, Front. Microbiol. 9 (2018) 1901, https://doi.org/10.3389/fmicb.2018.01901.

[95] C.M. Hemsley, P.A. O'Neill, A. Essex-Lopresti, I.H. Norville, T.P. Atkins, R.W. Titball, Extensive genome analysis of Coxiella burnetii reveals limited evolution within genomic groups, BMC Genom. 20 (2019) 441, https://doi.org/10.1186/s12864-019-5833-8.

[96] N.J. Willis, Edward Jenner and the eradication of smallpox, Scott Med. J. 42 (1997) 118–121, https://doi.org/10.1177/003693309704200407.

[97] P. Angsantikul, R.H. Fang, L. Zhang, Toxoid vaccination against bacterial infection using cell membrane-coated nanoparticles, Bioconjugate Chem. 29 (2018) 604–612, https://doi.org/10.1021/acs.bioconjchem.7b00692.

[98] R. Rappuoli, Reverse vaccinology, Curr. Opin. Microbiol. 3 (2000) 445–450, https://doi.org/10.1016/s1369-5274(00)00119-3.

[99] D. Maione, Identification of a universal group B Streptococcus vaccine by multiple genome screen, Science 309 (2005) 148–150, https://doi.org/10.1126/science.1109869.

[100] M.M. Giuliani, J. Adu-Bobie, M. Comanducci, B. Aricò, S. Savino, L. Santini, B. Brunelli, S. Bambini, A. Biolchi, B. Capecchi, E. Cartocci, L. Ciucchi, F. Di Marcello, F. Ferlicca, B. Galli, E. Luzzi, V. Masignani, D. Serruto, D. Veggi, M. Contorni, M. Morandi, A. Bartalesi, V. Cinotti, D. Mannucci, F. Titta, E. Ovidi, J.A. Welsch, D. Granoff, R. Rappuoli, M. Pizza, A universal vaccine for serogroup B meningococcus, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 10834–10839, https://doi.org/10.1073/pnas.0603940103.

[101] A. Sette, R. Rappuoli, Reverse vaccinology: developing vaccines in the era of genomics, Immunity 33 (2010) 530–541, https://doi.org/10.1016/j.immuni.2010.09.017.

[102] X. Bai, R. Borrow, Genetic shifts of *Neisseria meningitidis* serogroup B antigens and the quest for a broadly cross-protective vaccine, Expet Rev. Vaccine 9 (2010) 1203–1217, https://doi.org/10.1586/erv.10.116.

[103] K.L. Seib, X. Zhao, R. Rappuoli, Developing vaccines in the era of genomics: a decade of reverse vaccinology, Clin. Microbiol. Infect. 18 (Suppl 5) (2012) 109–116, https://doi.org/10.1111/j.1469-0691.2012.03939.x.

[104] F.A. Bidmos, S. Siris, C.A. Gladstone, P.R. Langford, Bacterial vaccine antigen discovery in the reverse vaccinology 2.0 era: progress and challenges, Front. Immunol. 9 (2018) 2315, https://doi.org/10.3389/fimmu.2018.02315.

[105] R.P. Hertzberg, Whole cell assays in screening for biologically active substances, Curr. Opin. Biotechnol. 4 (1993) 80–84, https://doi.org/10.1016/0958-1669(93)90037-w.

[106] A. Muzzi, V. Masignani, R. Rappuoli, The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials, Drug Discov. Today 12 (2007) 429–439, https://doi.org/10.1016/j.drudis.2007.04.008.

[107] M.K. Hellerstein, A critique of the molecular target-based drug discovery paradigm based on principles of metabolic control: advantages of pathway-based discovery, Metab. Eng. 10 (2008) 1–9, https://doi.org/10.1016/j.ymben.2007.09.003.

[108] L. Miesel, J. Greene, T.A. Black, Genetic strategies for antibacterial drug discovery, Nat. Rev. Genet. 4 (2003) 442–456, https://doi.org/10.1038/nrg1086.

[109] A. Mira, The bacterial pan-genome: a new paradigm in microbiology, Int. Microbiol. (2010) 45–57, https://doi.org/10.2436/20.1501.01.110.

[110] 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) 1–17, https://doi.org/10.1155/2015/139580.

[111] 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), https://doi.org/10.3389/fmicb.2017.00396.

[112] S.B. Jamal, S.S. Hassan, S. Tiwari, M.V. Viana, L. de J. Benevides, A. Ullah, A.G. Turjanski, D. Barh, P. Ghosh, D.A. Costa, A. Silva, R. Röttger, J. Baumbach, V.A.C. Azevedo, An integrative in-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheriae, PloS One 12 (2017) e0186401, , https://doi.org/10.1371/journal.pone.0186401.

[113] R. Zhang, H.-Y. Ou, C.-T. Zhang, DEG: a database of essential genes, Nucleic Acids Res. 32 (2004) D271–D272, https://doi.org/10.1093/nar/gkh024.

[114] S.J. Joseph, T.D. Read, Bacterial population genomics and infectious disease diagnostics, Trends Biotechnol. 28 (2010) 611–618, https://doi.org/10.1016/j.tibtech.2010.09.001.

[115] F.F. Mota, E.A. Gomes, E. Paiva, L. Seldin, Assessment of the diversity of Paenibacillus species in environmental samples by a novel rpoB-based PCR-DGGE method, FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol. 53 (2005) 317–328, https://doi.org/10.1016/j.femsec.2005.01.017.

[116] M. Drancourt, C. Bollet, A. Carlioz, R. Martelin, J.-P. Gayral, D. Raoult, 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates, J. Clin. Microbiol. 38 (2000) 8.

[117] A. Diop, K. El Karkouri, D. Raoult, P.-E. Fournier, Genome sequence-based criteria for demarcation and definition of species in the genus Rickettsia, Int. J. Syst. Evol. Microbiol. (2020), https://doi.org/10.1099/ijsem.0.003963.

[118] C.-C. Ho, A.K.L. Wu, C.W.S. Tse, K.-Y. Yuen, S.K.P. Lau, P.C.Y. Woo, Automated pangenomic analysis in target selection for PCR detection and identification of bacteria by use of ssGeneFinder webserver and its application to Salmonella enterica serovar Typhi, J. Clin. Microbiol. 50 (2012) 1905–1911, https://doi.org/10.1128/JCM.06843-11.

[119] C.R. Laing, M.D. Whiteside, V.P.J. Gannon, Pan-genome analyses of the species Salmonella enterica, and identification of genomic markers predictive for species, subspecies, and serovar, Front. Microbiol. 8 (2017) 1345, https://doi.org/10.3389/fmicb.2017.01345.

[120] A.H.M. van Vliet, Use of pan-genome analysis for the identification of lineage-specific genes of *Helicobacter pylori*, FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett. 364 (2017) fnw296, https://doi.org/10.1093/femsle/fnw296.

[121] R.C. Inglin, L. Meile, M.J.A. Stevens, Clustering of pan- and core-genome of Lactobacillus provides novel evolutionary insights for differentiation, BMC Genom. 19 (2018) 284, https://doi.org/10.1186/s12864-018-4601-5.

[122] A. Caputo, P.-E. Fournier, D. Raoult, Genome and pan-genome analysis to classify emerging bacteria, Biol. Direct 14 (2019) 5, https://doi.org/10.1186/s13062-019-0234-0.

[123] A. O'Callaghan, F. Bottacini, M. O'Connell Motherway, D. van Sinderen, Pangenome analysis of Bifidobacterium longum and site-directed mutagenesis through by-pass of restriction-modification systems, BMC Genom. 16 (2015) 832, https://doi.org/10.1186/s12864-015-1968-4.

[124] D.E. Fouts, L. Brinkac, E. Beck, J. Inman, G. Sutton, PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species, Nucleic Acids Res. 40 (2012) e172, https://doi.org/10.1093/nar/gks757.

[125] X. Chen, Y. Zhang, Z. Zhang, Y. Zhao, C. Sun, M. Yang, J. Wang, Q. Liu, B. Zhang, M. Chen, J. Yu, J. Wu, Z. Jin, J. Xiao, PGAweb: a web server for bacterial pan-genome analysis, Front. Microbiol. 9 (2018) 1910, https://doi.org/10.3389/fmicb.2018.01910.

[126] Y. Zhao, J. Wu, J. Yang, S. Sun, J. Xiao, J. Yu, PGAP: pan-genomes analysis pipeline, Bioinformatics 28 (2012) 416–418, https://doi.org/10.1093/bioinformatics/btr655.

[127] T.L. Pedersen, I. Nookaew, D. Wayne Ussery, M. Månsson, PanViz: interactive visualization of the structure of functionally annotated pangenomes, Bioinformatics (2017) btw761, https://doi.org/10.1093/bioinformatics/btw761.

[128] Y. Peng, S. Tang, D. Wang, H. Zhong, H. Jia, X. Cai, Z. Zhang, M. Xiao, H. Yang, J. Wang, K. Kristiansen, X. Xu, J. Li, MetaPGN: a Pipeline for Construction and Graphical Visualization of Annotated Pangenome Networks, GigaScience, 2018, https://doi.org/10.1093/gigascience/giy121.

[129] C. Ernst, S. Rahmann, PanCake: A Data Structure for Pangenomes, (2013), p. 11, https://doi.org/10.4230/OASICS.GCB.2013.35.

[130] T.K. Tan, K.Y. Tan, R. Hari, A. Mohamed Yusoff, G.J. Wong, C.C. Siow, N.V.R. Mutha, M. Rayko, A. Komissarov, P. Dobrynin, K. Krasheninnikova, G. Tamazian, I.C. Paterson, W.C. Warren, W.E. Johnson, S.J. O'Brien, S.W. Choo, PGD: a Pangolin Genome Hub for the Research Community, Database (Oxford), (2016), https://doi.org/10.1093/database/baw063.

[131] I. Yuvaraj, J. Sridhar, D. Michael, K. Sekar, PanGeT: pan-genomics tool, Gene 600 (2017) 77–84, https://doi.org/10.1016/j.gene.2016.11.025.

[132] G. Cheng, Q. Lu, L. Ma, G. Zhang, L. Xu, Z. Zhou, BGDMdocker: a Docker workflow for data mining and visualization of bacterial pan-genomes and biosynthetic gene clusters, Peer J. 5 (2017) e3948, https://doi.org/10.7717/peerj.3948.

[133] F. Plaza Oñate, E. Le Chatelier, M. Almeida, A.C.L. Cervino, F. Gauthier, F. Magoulès, S.D. Ehrlich, M. Pichaud, MSPminer: abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data, Bioinformatics 35 (2019) 1544–1552, https://doi.org/10.1093/bioinformatics/bty830.

[134] S.C. Bayliss, H.A. Thorpe, N.M. Coyle, S.K. Sheppard, E.J. Feil, PIRATE: a fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria, GigaScience 8 (2019), https://doi.org/10.1093/gigascience/giz119 giz119.

[135] H.A. Thorpe, S.C. Bayliss, S.K. Sheppard, E.J. Feil, Piggy: a rapid, large-scale pan-genome analysis tool for intergenic regions in bacteria, GigaScience 7 (2018), https://doi.org/10.1093/gigascience/giy015.

[136] N.M. Chaudhari, V.K. Gupta, C. Dutta, BPGA- an ultra-fast pan-genome analysis pipeline, Sci. Rep. 6 (2016) 24373, https://doi.org/10.1038/srep24373.

[137] A. Veras, F. Araujo, K. Pinheiro, L. Guimarães, V. Azevedo, S. Soares, A. da Costa da Silva, R. Ramos, Pan4Draft: a computational tool to improve the accuracy of Pan-genomic analysis using draft genomes, Sci. Rep. 8 (2018) 9670, https://doi.org/10.1038/s41598-018-27800-8.

[138] T.H. Clarke, L.M. Brinkac, J.M. Inman, G. Sutton, D.E. Fouts, PanACEA: a bioinformatics tool for the exploration and visualization of bacterial pan-chromosomes, BMC Bioinf. 19 (2018) 246, https://doi.org/10.1186/s12859-018-2250-y.

[139] Y. Pantoja, K. Pinheiro, A. Veras, F. Araújo, A. Lopes de Sousa, L.C. Guimarães, A. Silva, R.T.J. Ramos, PanWeb: a web interface for pan-genomic analysis, PloS One 12 (2017) e0178154, , https://doi.org/10.1371/journal.pone.0178154.

[140] S.-R. Jun, M.S. Robeson, L.J. Hauser, C.W. Schadt, A.A. Gorin, PanFP: pangenome-based functional profiles for microbial communities, BMC Res. Notes 8 (2015) 479, https://doi.org/10.1186/s13104-015-1462-8.

[141] A. Hennig, J. Bernhardt, K. Nieselt, Pan-Tetris: an interactive visualisation for Pan-genomes, BMC Bioinf. 16 (2015), https://doi.org/10.1186/1471-2105-16-S11-S3 S3.

[142] U. Kulsum, A. Kapil, H. Singh, P. Kaur, NGSPanPipe: a pipeline for pan-genome identification in microbial strains from experimental reads, in: R. Adhikari, S. Thapa (Eds.), Infectious Diseases and Nanomedicine III, Springer Singapore, Singapore, 2018, pp. 39–49, , https://doi.org/10.1007/978-981-10-7572-8_4.

[143] L. Carlos Guimaraes, L. Benevides de Jesus, M. Vinicius Canario Viana, A. Silva, R. Thiago Juca Ramos, S. de Castro Soares, V. Azevedo, Inside the pan-genome - methods and software overview 16 (2015), pp. 245–252, https://doi.org/10.2174/1389202916666150423002311 CG.

[144] M.J. Brittnacher, C. Fong, H.S. Hayden, M.A. Jacobs, M. Radey, L. Rohmer, PGAT: a multistrain analysis resource for microbial genomes, Bioinformatics 27 (2011) 2429–2430, https://doi.org/10.1093/bioinformatics/btr418.

[145] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.

[146] M.N. Benedict, J.R. Henriksen, W.W. Metcalf, R.J. Whitaker, N.D. Price, ITEP: an integrated toolkit for exploration of microbial pan-genomes, BMC Genom. 15

(2014) 8, https://doi.org/10.1186/1471-2164-15-8.
[147] Y. Zhao, X. Jia, J. Yang, Y. Ling, Z. Zhang, J. Yu, J. Wu, J. XiaoPanGP, A tool for quickly analyzing bacterial pan-genome profile, Bioinformatics 30 (2014) 1297–1299, https://doi.org/10.1093/bioinformatics/btu017.
[148] J.W. Sahl, C.J. Allender, R.E. Colman, K.J. Califf, J.M. Schupp, B.J. Currie, K.E. Van Zandt, H.C. Gelhaus, P. Keim, A. Tuanyok, Genomic characterization of burkholderia pseudomallei isolates selected for medical countermeasures testing: comparative genomics associated with differential virulence, PloS One 10 (2015) e0121052, , https://doi.org/10.1371/journal.pone.0121052.
[149] L. Snipen, K.H. Liland, micropan: an R-package for microbial pan-genomics, BMC Bioinf. 16 (2015) 79, https://doi.org/10.1186/s12859-015-0517-0.
[150] S. Paul, A. Bhardwaj, S.K. Bag, E.V. Sokurenko, S. Chattopadhyay, PanCoreGen — profiling, detecting, annotating protein-coding genes in microbial genomes, Genomics 106 (2015) 367–372, https://doi.org/10.1016/j.ygeno.2015.10.001.
[151] V. Bonnici, R. Giugno, V. Manca, PanDelos: a dictionary-based method for pan-genome content discovery, BMC Bioinf. 19 (2018) 437, https://doi.org/10.1186/s12859-018-2417-6.
[152] K. Abudahab, J.M. Prada, Z. Yang, S.D. Bentley, N.J. Croucher, J. Corander, D.M. Aanensen, PANINI: pangenome neighbour identification for bacterial populations, Microb. Genom. 5 (2019), https://doi.org/10.1099/mgen.0.000220.
[153] T.J. Treangen, B.D. Ondov, S. Koren, A.M. Phillippy, The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes, Genome Biol. 15 (2014) 524, https://doi.org/10.1186/s13059-014-0524-x.
[154] Y.F. Ngeow, W.Y. Wee, Y.L. Wong, J.L. Tan, C.S. Ongi, K.P. Ng, S.W. Choo, Genomic analysis of Mycobacterium abscessus strain M139, which has an ambiguous subspecies taxonomic position, J. Bacteriol. 194 (2012) 6002–6003, https://doi.org/10.1128/JB.01455-12.
[155] C. Milani, S. Duranti, G.A. Lugli, F. Bottacini, F. Strati, S. Arioli, E. Foroni, F. Turroni, D. van Sinderen, M. Ventura, Comparative genomics of Bifidobacterium animalis subsp. lactis reveals a strict monophyletic bifidobacterial taxon, Appl. Environ. Microbiol. 79 (2013) 4304–4315, https://doi.org/10.1128/AEM.00984-13.
[156] S. Argimón, K. Konganti, H. Chen, A.V. Alekseyenko, S. Brown, P.W. Caufield, Comparative genomics of oral isolates of Streptococcus mutans by in silico genome subtraction does not reveal accessory DNA associated with severe early childhood caries, Infect. Genet. Evol. 21 (2014) 269–278, https://doi.org/10.1016/j.meegid.2013.11.003.
[157] V.J. Timms, K.A. Hassan, H.M. Mitchell, B.A. Neilan, Comparative genomics between human and animal associated subspecies of the Mycobacterium avium complex: a basis for pathogenicity, BMC Genom. 16 (2015) 695, https://doi.org/10.1186/s12864-015-1889-2.
[158] U. Wegmann, D.A. MacKenzie, J. Zheng, A. Goesmann, S. Roos, D. Swarbreck, J. Walter, L.C. Crossman, N. Juge, The pan-genome of Lactobacillus reuteri strains originating from the pig gastrointestinal tract, BMC Genom. 16 (2015) 1023, https://doi.org/10.1186/s12864-015-2216-7.
[159] S. Reuter, J. Corander, M. de Been, S. Harris, L. Cheng, M. Hall, N.R. Thomson, A. McNally, Directional gene flow and ecological separation in Yersinia enterocolitica, Microb. Genom. 1 (2015), https://doi.org/10.1099/mgen.0.000030.
[160] I.D. Otchere, S.R. Harris, S.L. Busso, A. Asante-Poku, S. Osei-Wusu, K. Koram, J. Parkhill, S. Gagneux, D. Yeboah-Manu, The First population structure and comparative genomics analysis of Mycobacterium africanum strains from Ghana reveals higher diversity of Lineage 5, Int. J. Mycobacteriology 5 (2016) S80–S81, https://doi.org/10.1016/j.ijmyco.2016.09.051.
[161] F. Asenjo, A. Olmos, P. Henríquez-Piskulich, V. Polanco, P. Aldea, J.A. Ugalde, A.N. Trombert, Genome sequencing and analysis of the first complete genome of *Lactobacillus kunkeei* strain MP2, an *Apis mellifera* gut isolate, Peer J. 4 (2016) e1950, , https://doi.org/10.7717/peerj.1950.
[162] J.Y.H. Lee, I.R. Monk, S.J. Pidot, S. Singh, K.Y.L. Chua, T. Seemann, T.P. Stinear, B.P. Howden, Functional analysis of the first complete genome sequence of a multidrug resistant sequence type 2 Staphylococcus epidermidis, Microb. Genom. 2 (2016), https://doi.org/10.1099/mgen.0.000077.
[163] T. Azarian, A. Ali, J.A. Johnson, M. Jubair, E. Cella, M. Ciccozzi, D.J. Nolan, W. Farmerie, M.H. Rashid, S. Sinha-Ray, M.T. Alam, J.G. Morris, M. Salemi, Non-toxigenic environmental Vibrio cholerae O1 strain from Haiti provides evidence of pre-pandemic cholera in Hispaniola, Sci. Rep. 6 (2016) 36115, https://doi.org/10.1038/srep36115.
[164] A.L. Bazinet, Pan-genome and phylogeny of Bacillus cereus sensu lato, BMC Evol. Biol. 17 (2017) 176, https://doi.org/10.1186/s12862-017-1020-1.
[165] M. Gomila, A. Busquets, M. Mulet, E. García-Valdés, J. Lalucat, Clarification of taxonomic status within the Pseudomonas syringae species group based on a phylogenomic analysis, Front. Microbiol. 8 (2017) 2422, https://doi.org/10.3389/fmicb.2017.02422.
[166] J.W. Turner, J.J. Tallman, A. Macias, L.J. Pinnell, N.C. Elledge, D. Nasr Azadani, W.B. Nilsson, R.N. Paranjpye, E.V. Armbrust, M.S. Strom, Comparative genomic analysis of Vibrio diabolicus and six taxonomic synonyms: a first look at the distribution and diversity of the expanded species, Front. Microbiol. 9 (2018) 1893, https://doi.org/10.3389/fmicb.2018.01893.
[167] S.P. Baker, T.J. Nulton, T. Kitten, Genomic, phenotypic, and virulence analysis of *Streptococcus sanguinis* oral and infective-endocarditis isolates, Infect. Immun. 87 (2018) e00703–e00718, https://doi.org/10.1128/IAI.00703-18 /iai/87/1/IAI.00703-18.atom.
[168] J. Xing, X. Li, Y. Sun, J. Zhao, S. Miao, Q. Xiong, Y. Zhang, G. Zhang, Comparative genomic and functional analysis of Akkermansia muciniphila and closely related species, Genes Genomics 41 (2019) 1253–1264, https://doi.org/10.1007/s13258-019-00855-1.
[169] Sheppard, Groves, Andrews, Litt, Fry, Southern, Miller, The genomics of Streptococcus pneumoniae carriage isolates from UK children and their household contacts, pre-PCV7 to post-PCV13, Genes 10 (2019) 687, https://doi.org/10.3390/genes10090687.
[170] R.T. Rao, N. Sivakumar, K. Jayakumar, Analyses of livestock-associated *Staphylococcus aureus* pan-genomes suggest virulence is not primary interest in evolution of its genome, OMICS, J. Integr. Biol. 23 (2019) 224–236, https://doi.org/10.1089/omi.2019.0005.
[171] B. Ferwerda, M.M. Maury, M.C. Brouwer, L. Hafner, A. van der Ende, S. Bentley, M. Lecuit, D. van de Beek, Residual variation intolerance score detects loci under selection in neuroinvasive Listeria monocytogenes, Front. Microbiol. 10 (2019) 2702, https://doi.org/10.3389/fmicb.2019.02702.
[172] Á. Chiner-Oms, L. Sánchez-Busó, J. Corander, S. Gagneux, S.R. Harris, D. Young, F. González-Candelas, I. Comas, Genomic determinants of speciation and spread of the *Mycobacterium tuberculosis* complex, Sci. Adv. 5 (2019), https://doi.org/10.1126/sciadv.aaw3307 eaaw3307.
[173] J.P.R. Furlan, O.G.G. de Almeida, E.C.P. De Martinis, E.G. Stehling, Characterization of an environmental multidrug-resistant acinetobacter seifertii and comparative genomic analysis reveals Co-occurrence of antimicrobial resistance and metal tolerance determinants, Front. Microbiol. 10 (2019) 2151, https://doi.org/10.3389/fmicb.2019.02151.
[174] B. Lu, C. Lin, H. Liu, X. Zhang, Y. Tian, Y. Huang, H. Yan, M. Qu, L. Jia, Q. Wang, Molecular characteristics of Klebsiella pneumoniae isolates from outpatients in sentinel hospitals, beijing, China, 2010–2019, Front. Cell. Infect. Microbiol. 10 (2020) 85, https://doi.org/10.3389/fcimb.2020.00085.
[175] R.A. Gladstone, S.W. Lo, R. Goater, C. Yeats, B. Taylor, J. Hadfield, J.A. Lees, N.J. Croucher, A.J. van Tonder, L.J. Bentley, F.X. Quah, A.J. Blaschke, N.L. Pershing, C.L. Byington, V. Balaji, W. Hryniewicz, B. Sigauque, K.L. Ravikumar, S.C.G. Almeida, T.J. Ochoa, P.L. Ho, M. du Plessis, K.M. Ndlangisa, J.E. Cornick, B. Kwambana-Adams, R. Benisty, S.A. Nzenze, S.A. Madhi, P.A. Hawkins, A.J. Pollard, D.B. Everett, M. Antonio, R. Dagan, K.P. Klugman, A. von Gottberg, B.J. Metcalf, Y. Li, B.W. Beall, L. McGee, R.F. Breiman, D.M. Aanensen, S.D. Bentley, The global pneumococcal sequencing consortium, visualizing variation within global pneumococcal sequence clusters (GPSCs) and country population snapshots to contextualize pneumococcal isolates, Microb. Genom. (2020), https://doi.org/10.1099/mgen.0.000357.
[176] Department of Microbiology, Assam University, Silchar 788011, Assam, India, J. Rajkumari, Distinctive features gleaned from the comparative genomes analysis of clinical and non-clinical isolates of Klebsiella pneumoniae, Bioinformation 16 (2020) 256–266, https://doi.org/10.6026/97320630016256.
[177] M.I. Gröschel, C.J. Meehan, I. Barilar, M. Diricks, A. Gonzaga, M. Steglich, O. Conchillo-Solé, I.-C. Scherer, U. Mamat, C.F. Luz, K. De Bruyne, C. Utpatel, D. Yero, I. Gibert, X. Daura, S. Kampmeier, N.A. Rahman, M. Kresken, T.S. van der Werf, I. Alio, W.R. Streit, K. Zhou, T. Schwartz, J.W.A. Rossen, M.R. Farhat, U.E. Schaible, U. Nübel, J. Rupp, J. Steinmann, S. Niemann, T.A. Kohl, The phylogenetic landscape and nosocomial spread of the multidrug-resistant opportunist Stenotrophomonas maltophilia, Nat. Commun. 11 (2020) 2044, https://doi.org/10.1038/s41467-020-15123-0.
[178] J. Loraine, E. Heinz, R. Soontarach, G.A. Blackwell, R.A. Stabler, S.P. Voravuthikunchai, P. Srimanote, P. Kiratisin, N.R. Thomson, P.W. Taylor, Genomic and phenotypic analyses of acinetobacter baumannii isolates from three tertiary care hospitals in Thailand, Front. Microbiol. 11 (2020) 548, https://doi.org/10.3389/fmicb.2020.00548.
[179] T. Nagy, A. Szmolka, T. Wilk, J. Kiss, M. Szabó, J. Pászti, B. Nagy, F. Olasz, Comparative genome analysis of Hungarian and global strains of Salmonella infantis, Front. Microbiol. 11 (2020) 539, https://doi.org/10.3389/fmicb.2020.00539.
[180] A. Xu, D. Wang, Y. Ding, Y. Zheng, B. Wang, Q. Wei, S. Wang, L. Yang, L.Z. Ma, Integrated comparative genomic analysis and phenotypic profiling of Pseudomonas aeruginosa isolates from crude oil, Front. Microbiol. 11 (2020) 519, https://doi.org/10.3389/fmicb.2020.00519.
[181] C. Saralegui, M. Ponce-Alonso, B. Pérez-Viso, L. Moles Alegre, E. Escribano, F. Lázaro-Perona, V.F. Lanza, M.S. de Pipaón, J.M. Rodríguez, F. Baquero, R. del Campo, Genomics of Serratia marcescens isolates causing outbreaks in the same pediatric unit 47 Years apart: position in an updated phylogeny of the species, Front. Microbiol. 11 (2020) 451, https://doi.org/10.3389/fmicb.2020.00451.
[182] M. Su, J.T. Lyles, R.A. Petit III, J. Peterson, M. Hargita, H. Tang, C. Solis-Lemus, C.L. Quave, T.D. Read, Genomic analysis of variability in Delta-toxin levels between *Staphylococcus aureus* strains, Peer J. 8 (2020) e8717, , https://doi.org/10.7717/peerj.8717.
[183] R. Rios, J. Reyes, L.P. Carvajal, S. Rincon, D. Panesso, A.M. Echeverri, A. Dinh, S.-O. Kolokotronis, A. Narechania, T.T. Tran, J.M. Munita, B.E. Murray, P.J. Planet, C.A. Arias, L. Diaz, Genomic epidemiology of vancomycin-resistant Enterococcus faecium (VREfm) in Latin America: revisiting the global VRE population structure, Sci. Rep. 10 (2020) 5636, https://doi.org/10.1038/s41598-020-62371-7.