Chapter 4
Whole-Genome Sequencing for Bacterial Virulence Assessment

Florian Tagini, Trestan Pillonel, and Gilbert Greub

4.1 Introduction

In recent years, whole-genome sequencing (WGS) is increasingly being considered a technique that could change clinical microbiology [1, 2]. In addition to microbial typing and prediction of antibiotic susceptibility, one of the major clinical application of bacterial genomics is the detection of clinically relevant virulence factors and virulence prediction. In this chapter, we will thus explore what this technique really implies [3]. Before discussing virulence factors, the terms "virulence" and "pathogenicity" need to be defined.

For virulence, the definition used in this chapter is "the relative capacity of a micro-organism to cause damage to a host" as proposed by Casadevall & Pirofski [4]. Conversely, the pathogenicity is the general capacity of a microorganism to cause damage to a host, and depends on both the pathogen and the host-response [4]. Pathogenicity is to be opposed to commensalism, where the interaction results in no clear benefits or damages for any of the involved organism. Of notes, the damage-response framework of pathogens is not restricted to the direct effects of a micro-organism on a host [4]. For instance, immunological molecular mimicry or oncogenesis can cause damage to a host and are not directly related to the entry of

F. Tagini · T. Pillonel
Institute of Microbiology, Department of Laboratory Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

G. Greub (✉)
Institute of Microbiology, Department of Laboratory Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Division of Infectious Diseases, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland
e-mail: gilbert.greub@chuv.ch

45
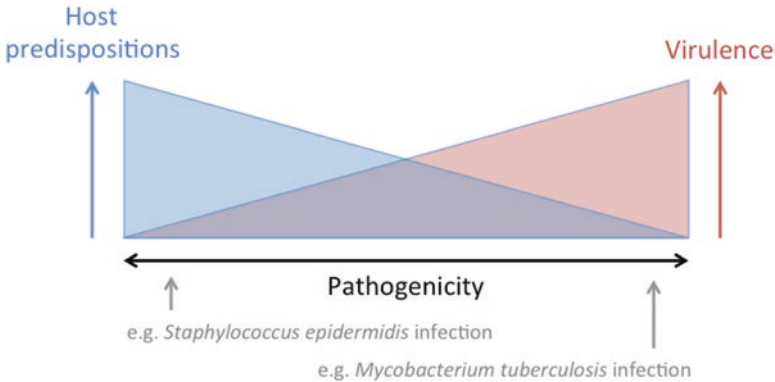
Fig. 4.1 Pathogenicity as the result of the host-pathogen interaction. In this model, highly virulent bacteria can be pathogenic regardless of the host status (e.g. *Mycobacterium tuberculosis*), while other bacteria are generally considered as less virulent than most pathogen and would be pathogenic only in specific situations (e.g. *Staphylococcus epidermidis* is pathogenic only when an intravenous catheter is in place or when patients are immunosuppressed)

a given bacterial isolate [4]. Thus, bacterial proteins implicated in such pathogenesis represent virulence factors. In this chapter, we will mainly focus on the direct damages that can be caused by a pathogen to a host and we will use a simplified model to define pathogenicity and virulence (Fig. 4.1), where a given bacteria, upon the presence and expression of virulence factors and according to host's susceptibility (e.g. immune status, epithelial breach, genetic predisposition), can be pathogenic, i.e. causing tissue lesions or organ damage.

The virulence of a bacterial strain depends on the presence and expression of virulence factors and is solely dependent on the strain characteristics. A virulence factor can be defined as a determining factor (i.e. gene product) that would help improve the survival of a bacterium within the host or by causing more cellular and tissue damage. Several classes of virulence factors should be recognised, including (1) toxins, (2) effectors of secretion systems, (3) adhesive factors, (4) invasive factors, (5) resistance to reactive oxygen and nitrogen species, (6) immune system escape, and (7) nutrient uptake. Antibiotic resistance determinants, although they may contribute to the pathogenesis (e.g. in a patient treated with antibiotics), form a special class of genes and are not discussed in this chapter.

4.1.1 Toxins

Bacterial toxin is a general term to describe a diverse set of virulence factors that are generally secreted by the bacterium and cause damage to host cells. It consists of several subcategories with different modes of actions: (i) pore-forming toxins, (ii) adenylate or guanylate cyclase-affecting toxins, (iii) protein synthesis-inhibiting toxins, (iv) surfactant-like toxins, (v) superantigens, and (vi) neurotoxins [5].

(i) Pore-forming toxins have one of the most straightforward mechanisms. Indeed, these molecules are able to form pores in host cells, which causes influx and efflux of ions, small molecules and proteins and eventually leading to cell death [6]. For instance, bacterial-mediated haemolysis, unraveled early on in the history of microbiology (19th and 20th centuries), was shown to be due to pore-forming toxins, such as the listeriolysin O of *Listeria monocytogenes*, streptolysins O and S of *Streptococcus pyogenes*, or the staphylococcal alpha- and gamma-toxin [7–12]. Another example of pore-forming toxins is the *Staphylococcus aureus* Panton-Valentine Leucocidin (PVL, LukSF) (or other leucocidins such as LukGH or LukDE), which can directly lyse human leuko- cytes [10].

(ii) Adenylate and guanylate cyclase–affecting toxins are a particular class of tox- ins, found for instance in enteropathogens such as *Escherichia coli* and *Vibrio cholerae,* or in respiratory pathogens such as *Bordetella pertussis*, the caus- ative agent of pertussis [5]. These toxins penetrate the host cell and lead to an increased production of cyclic AMP or cyclic GMP, and eventually to the up- regulation of ion channels and to an increased volume of mucosal secretion (water follows the osmolality and is attracted into the lumen) [13–20]. Diarrhoea, emesis and cough are, respectively, the resulting symptoms, which are thought to favor the bacterial spread to other hosts. Interestingly, these toxins may have a broader spectrum of action. For instance, *B. pertussis* toxins could also inhibit phagocytosis, cytokine production and oxidative burst [5].

(iii) Protein synthesis–inhibiting toxins dramatically contribute to the pathogenic- ity of several bacteria, eventually leading to host cell death. For instance, the diphtheria toxin, encoded by a lysogenic bacteriophage of *Corynebacterium diphtheriae*, ADP-ribosylates the elongation factor 2 of the host cell and thus prevents protein synthesis [21]. This toxin causes local damages at the site of infection, the respiratory tract, with the formation of characteristic pseudo- membranes and also systemic damages, such as heart and other end-organ injuries [22, 23]. Another classical example is the Shiga toxin producing *Escherichia coli* (STEC) and the role of Shiga toxins in the pathogenesis of haemolytic and uremic syndromes [24].

(iv) Surfactant-like toxins constitute a particular class of toxins with amphipathic properties capable to disrupt lipid bilayers of the host membrane. Best exem- plified by phenol-soluble modulins of staphylococci, they have a broad spec- trum of actions, such as host cell lysis, pro-inflammatory stimulation, and contribution to biofilm formation [25, 26]

(v) Superantigens, produced mainly by *S. aureus* or *Streptococcus pyogenes,* are a specific class of toxins that can bind both the lymphocytic T-cell receptors and the Major Histocompatibility Complex (MHC). Thus, it activates up to 20% of lymphocytes in a non-specific manner, ultimately leading to an inflammatory cytokine storm in the host and potentially to cardiovascular collapse due to an increased vascular permeability and death due to shock and multi-organ fail- ure [27].

(vi) Neurotoxins, such as the botulinum or tetanus toxins produced by *Clostridium botulinum* and *Clostridium tetani*, respectively, are a separate category of toxin, able to modulate the transmissions of nerves impulses [28, 29].

4.1.2 Secretion Systems and Their Effectors

Although most toxins are secreted by various bacterial secretion systems, some specific secretion systems have an important role in secreting the so called "effectors" that are able to induce damage in the target cell, and could be considered a distinct class of virulence factors. First, the type III secretion system (T3SS) is found in several Gram-negative pathogens, such as *Salmonella* spp*., Shigella, Pseudomonas* spp. and *Yersinia* spp. [30]. In addition, it is also encoded by intracellular pathogens, such as *Chlamydia trachomatis, Waddlia chondrophila* and plays a central role in their pathogenesis [31]. T3SS assemble into a needle-like apparatus conserved across distant bacterial species. It is able to secrete its effectors into the target cell, which may affect a broad range of cellular functions, such actin cytoskeletal dynamics, gene expression and post-translational modifications, signal transduction pathways, and vesicle transport and endocytic trafficking [32]. Second, the type-four secretion system is an important virulence factor of Gram negative and Gram positive bacteria involved in various cellular processes including conjugative horizontal gene transfers and contact-independent DNA uptake, as well as secretion of toxins or effector proteins in the target cell [33]. Third, the type VI secretion system is also particularly interesting for bacterial virulence: in the context of polymicrobial infection, it helps pathogens to compete with other bacteria and to colonize a niche; for intracellular bacteria such as *Burkholderia* spp. and *Francisella tularensis*, it can also specifically mediate virulence (e.g. phagosomal escape for *F. tularensis*) [34, 35]. Finally, the type VII secretion systems (T7SS) is a key virulence factor of *M. tuberculosis* and other mycobacteria [36]. It can also be found in many *Actinobacteria* and in *Firmicutes* (with a type-VII-like secretion system) [37]. The most famous *M. tuberculosis* effectors are Esx-A (ESAT-6) and Esx-B (CFP-10). They are involved in modulation of the T-cell response, phagosomal escape and exhibit some direct effect on the membrane of host cells [38].

4.1.3 Adhesive Factors

Adherence to various surfaces (e.g. the endothelium or any prosthetic device), can be an important determinant of bacterial pathogenesis [39, 40]. A broad range of proteins or protein assemblies can promote bacterial attachment and are classically called adhesins (one protein) or pili (large protein assemblies). For instance, several pathogens associated with endocarditis, such as *Bartonella henselae*, *Eikenella corrodens* and *Cardiobacterium hominis* have been shown to encode adhesins [41–43].

Regarding pili, an example is the type IV pili of *Neisseria meningitidis* that allows attachment to the epithelium, invasion into the bloodstream, attachment to the brain microvascular endothelium and crossing of the blood-brain barrier to cause meningitis [44, 45]. Furthermore many pathogens are able to produce biofilms, which are matrices of hydrated extracellular polymeric substances, composed mainly of polysaccharides, proteins, nucleic acids and lipids formation [46]. It promotes the mechanical attachment of the micro-organisms and the large three-dimensional structure of some adhesins, reduces the susceptibility of bacteria to various stresses and to antibiotic therapies [47] as well as their engulfment by phagocytic cells.

4.1.4 Invasive Factors

Some virulence factors can help the bacteria invade tissues and promote their systemic dissemination. For instance, streptokinase and staphylokinase activate plasminogen into plasmin, which can then break down fibrin clots. These proteases are involved in tissue spreading by destroying the extracellular matrix and fibrin fibers that holds cells together [48, 49]. Many other bacterial proteases can degrade the extracellular matrix or even surprisingly the DNA nets of neutrophils and participate in bacterial invasion [50–53].

4.1.5 Resistance to Reactive Oxygen and Nitrogen Species

Many genes are involved in the resistance to stresses that bacteria encounter within the host [54]. For instance, genes involved in resistance to reactive oxygen or nitrogen species can affect bacterial survival. The *S. aureus* catalase enables the breakdown of hydrogen peroxide and thus was thought to improve the survival of bacteria to the killing by neutrophils [55]. However, it was later shown that catalase-negative *S. aureus* infection can retain virulence, highlighting the plethora of bacterial compensatory "virulence" mechanisms [56, 57].

4.1.6 Immune System Escape

To increase their survival, bacteria have developed a broad range of molecular means to subvert both the innate and adaptive immune systems of the host. First, many bacteria are able to prevent phagocytosis, the most straightforward way to clear bacteria. For instance, the production of a polysaccharide capsule (e.g. for *Streptococcus pneumoniae*) can prevent bacterial opsonisation by the complement system or by immunoglobulins [58]. Furthermore, bacteria, and particularly intracellular bacteria, have developed many different ways to escape the endosome−/

phagosome- lysosome pathway, by manipulating the host cellular pathways [59–61]. Some pathogen strategies aim at degrading host chemokines involved in the inflammatory response to attract neutrophiles, like interleukin-8, with specific proteases (e.g. SpyCEP of *S. pyogenes*) [62]. Bacterial proteases can also cleave immunoglobulins, such as IgA1 or IgG, which promote bacterial attachment to mucosal surfaces and survival, respectively [63, 64]. Conversely, bacteria can also recruit regulatory molecules. *Neisseria meningitidis* recruits factor H, which prevents the activation of the complement [65]. Lipopolysaccharide, besides being an important immune system stimulating factor (that can eventually lead to septic shock), is also known to contribute to the resistance to complement of *K. pneumoniae* [66].

4.1.7   Nutrient Uptake

The fight for nutrients is a complex interplay between the host and the pathogen. Iron, an inorganic ion, is required for many eukaryotic and bacterial processes and is involved in virulence and pathogenicity [67]. Bacteria have developed many ways to circumvent every iron-sequestration strategies of the host. For instance, bacteria can acquire iron, which is bound to transferrin, lactoferrin or hemoglobin. Furthermore, iron tightly regulates many virulence factors (e.g. the diphtheria toxin) [68]. For intracellular bacteria, the acquisition of nutrients is also critical for their survival [61, 69]. In addition to iron and other nutrients, the acquisition of cholesterol, for instance, is made after manipulation of the host cell machinery [70].

From a genomic perspective, bacteria – including pathogenic ones – generally present highly plastic genomes. We should differentiate the core-genome, consisting of conserved genes between all bacteria of a species or any other clade (core genes), from the accessory genome, which includes all the variable genetic elements of a given species or clade. Horizontal gene transfers, mediated by bacteriophages, plasmids or recombination, are important drivers of bacterial evolution and pathogen adaptation [71]. Virulence determinants can thus be either encoded in all bacterial strains of a given species or sporadically occur in some virulent strains. For instance, several *S. aureus* toxins, such as the alpha-toxin and some phenol-soluble proteins [10], are present in every *S. aureus* isolates whereas the PVL toxin is only encoded by the genome of some more virulent *S. aureus* strains. Core-genome sizes vary a lot across different bacterial species [72–74]. Good knowledge of the genomics of the pathogen is thus required for a successful identification and interpretation of virulence markers [3]. If a virulence factor is encoded by a core-gene, it generally means that the species identification itself implies the presence of that trait (if strictly present in the core-genome). However, when assessing virulence factors we should not focus only on the accessory genes since variants (such as single-nucleotide polymorphisms (SNPs), deletion or insertion events) may occur in core genes, leading potentially to loss or gain of function that may increase or decrease the virulence of a given strain.

Most of the knowledge on virulence factors has been gathered thanks to *in vitro* experiments or animal models of infection. However, the overall contribution to pathogenicity of virulence factors is often less clear in humans. Indeed, significant differences between humans and mice may impact the interpretations of animal experiments and extrapolation of obtained results to humans. Furthermore, the presence of genes encoding virulence factors is usually not sufficient by itself to increase the virulence of a bacteria. Indeed, transcriptional modulation and expression of the protein, which depend on complex regulatory networks, are truly determining the virulence end-phenotype. These regulators depend on various mechanisms, such as the activity of two-component systems [75], expression of regulatory non-coding RNAs [76] or sigma factors [77].

By providing a detailed map of the virulence factors encoded in a bacterial strain, WGS could bring new useful predicting tools for clinical microbiologists. In this chapter, we will review the main current and foreseen clinical applications where WGS has or could have an added value. In addition, we will discuss the main technical approaches and limitations of WGS as well as the validation and interpretation of the results.

4.2 Clinical Applications

For WGS analyses, requests to identify known virulence factors may be driven by various reasons. A critical assessment of the benefits for the patient or public health is thus important. Therefore, one should always wonder: will the analysis have any impact on patient's treatment or on any other management aspects (for instance, by undertaking specific infection control measures)? If the answer is negative, then the utility of the analysis is probably limited to the research field. Thus, there are two main motives for virulence determinants detection: (a) VFs detection for personalised treatment and/or clinical management, and (b) epidemiological surveillance of virulent clones.

In conventional clinical microbiology workflows, the virulence properties of bacterial isolates are rarely characterised [1]. Indeed, the identification of the species brings up usually the possibility to infer the general pathogenic potential of an isolate. For instance, the identification of *Listeria monocytogenes* in the cerebrospinal fluid of a patient proves meningitis. Based soley on the identification of the bacterium, we assume that the strain is pathogenic in that situation, and that a set of genes involved in virulence and pathogenesis is present. Knowing whether the strain encodes some accessory virulence genes is unlikely to change patient management. However, in this particular example, WGS could still be useful for typing and public health surveillance. Furthermore, as the ultimate typing method, WGS can also provide a rapid taxonomic assignment of an isolate by classifying it to a particularly virulent clade (e.g. species, subspecies or serovar). Indeed, for many pathogens, several clonal complexes have been associated to increased virulence or poorer prognosis (e.g. for *S. aureus*, *E. coli*, and many more) [78, 79]. As many virulence

genes are acquired through horizontal gene transfer, typing analyses based on whole genomes data can add additional assessment of virulence factor content, allowing to monitor the spread of established virulence factors.

E. coli, is another well-illustrating example. Due to its highly plastic genome and in the context of the host-pathogen interaction, *Escherichia coli* presents variable phenotypes ranging from commensal interactions to very invasive diseases. For instance, upon the acquisition of specific virulence genes (Table 4.1), *E. coli* can present very specific pathogenic features, classified into pathotypes [80]. However, the pathotype definition seems to lose relevance with the rising number of virulence factor combinations and virulence phenotypes [80]. For instance, the recently described Shiga-toxin producing enteroaggregative *E.coli* (STEAEC) is a hybrid between STEC and EAEC [81]. In addition, the Shigella B13 carrying the locus enterocyte effacement (LEE) pathogenicity island is more closely related to *E. alberti* than to *E. coli* [82]. Therefore, there is a clear added value to detect virulence factors in order (a) to monitor and predict the emergence of new pathotype combinations, and (b) to identify horizontal transfers events of known virulence factors.

In addition to *E. coli* and *S. aureus*, the detection of virulence factors can also be recommended in some other specific cases. This concerns mostly bacterial species that are known to exhibit a high virulence variability. For instance, *C. diphtheriae* can cause the clinical disease diphtheria when encoding *tox* gene and expressing the diphtheria toxin (cf. section 1). Specific PCRs are available to detect the *tox* gene but are usually only available at national reference centers (together with toxin production assays). Toxigenic *C. diphtheriae* infections require specific isolation strategies and specific patient's follow-up to monitor potential cardiac toxicity. Ruling out the presence of the toxin may take time because of referral to a reference center for testing [83]. Local WGS analysis in this context can be more time- and cost-effective than the standard procedures [83].

Another example of extensively studied toxin is the Panton-Valentin leucocidin (PVL) of *S. aureus*. For recurrent skin and soft tissue infections, the detection of the toxin was shown to be useful. Indeed, specific decolonisation strategies may be introduced [84]. Although considered to be a marker of invasiveness, its association with pneumonia, bacteraemia and musculoskeletal infections was shown to be questionable [85]. Conversely, PVL-positive strains are associated with skin and soft tissue infections and are more likely to be treated surgically. A large set of virulence factors is encoded in *S. aureus* genomes but their identification is currently unlikely to lead to the modification of the antibiotic therapy (Table 4.1) [86]. Indeed, the choice of therapy is currently driven mainly by the antibiotic susceptibility of the strain and by clinical presentation rather than by the presence of the PVL or other virulence factors [85]. For instance, in the presence of severe necrotising pneumonia, clindamycin or linezolid, both inhibitors of the ribosomes and protein synthesis, can be introduced [87]. Interestingly, a genome-wide association study of methicillin-resistant *S. aureus* could predict isolates toxicity from their genomic sequences, looking at SNPs, insertion and deletions events [88]. However, there is a need for more genotype-outcome associations, before virulence determinants can be integrated into the clinical management of *S. aureus* infections [89].

Table 4.1 Selected virulence factors

Pathogen	Virulence determinants	Utility / potential role	Comments
Escherichia coli/ Shigella spp.	LEE PAI; *eae, bfpA*	EPEC defining region.	Part of the accessory genome; the search of these genes allows pathotyping of the isolates.
	Plasmid pINV; *ipaH* and *ipa* genes	EIEC/*Shigella* defining plasmid.	
	ST and LT toxins; *est, elt*	ETEC defining genes.	
	Shiga toxins; *stx1, stx2*	EHEC defining genes.	
	Plasmid pAA; *aggR, aatA, aaiC*	Associated with EAEC phenotype.	
	Afa/Dr: adhesins encoding genes	DAEC defining genes.	
Klebsiella pneumoniae	*rmpA/A2* (regulator of mucoid phenotype)	Association with community-acquired pyogenic liver abscesses and other hypervirulent phenotypes.	Accessory genes; some of them are present on plasmids.
	Yersiniabactin and aerobactin loci		
	Colibactin toxin locus		
	pld1 (phospholipase D family protein)		
Staphylococcus aureus	Panton–Valentine toxin; *lukSF*	Mainly useful for recurrent SSTIs; various associations with outcome.	Accessory gene (cf. text).
	Exfoliative toxin A, B and D; *eta, etb, etd*	Association with the pathogenesis of bullous impetigo and staphylococcal scaled-skin syndrome.	Although part of the accessory genome, the pathogenicity depends also on the gene expression.
	Staphylococcal enterotoxins and enterotoxin-like toxins; *se(a-e), se(g-j), se(r-t), sel(k-q), sel(u-w), tst* (TSST-1)	Associated with infective endocarditis, food poisoning and toxic shock syndrome.	
	Alpha-hemolysin; *hla/hly*	Its expression was associated with the outcome of bacteraemia and its activity with ventilator-associated pneumonia.	Core gene.
	Phenol soluble modulins; *psm* genes	Expression associated with SSTI.	Core genes (usually). Their expression influence virulence.
	Fibronectin-binding protein; *fnbA* and *fnbB*	Associated with cardiac device infections.	*fnbA* is a core-gene and *fnbB* is part of the accessory genome.
	Clumping-factor A and B (fibrinogen binding, platelet activation); *clfA, clfB*	Association with infective endocarditis.	Accessory genes.
	Collagen adhesin; *cna*	Associated with invasive infections and infective endocarditis.	Accessory genes.

Table 4.1 (continued)

Pathogen	Virulence determinants	Utility / potential role	Comments
	Biofilm formation; *ica* genes	Associated with invasive infections.	Accessory genes.
	Staphylokinase, chemotaxis inhibitory protein, staphylococcal complement inhibitor; *sak*, *chp*, *scn*	Associated with bacteremia (all) and with mortality (staphylokinase).	Accessory genes.
	Accessory gene regulator; *agr*	Gene polymorphisms and expression were associated with persistent bacteremia and mortality.	Core genes; involved in the regulation of many virulence genes.
Streptococcus pyogenes	Superantigens; *speA*, *speC*, *speH*, *speI*, *speJ*, *speLM*, *ssa*, SMEZ	Associated with the pathogenesis of toxic shock syndrome (and scarlet fever).	Accessory genes.
	Cysteine protease; *speB*	Its expression was associated with tissue invasion.	Core gene.
	Streptolysin S and O; *sagA*, *sagB*, *sagC*, *slo*	Expression associated with invasive diseases.	Core genes.
	Antiphagocytic protein (M protein); *emm*	Required for full virulence in animal models; Utility for reverse-compatibility with *emm*-typing.	Core gene.
	Hyaluronic acid capsule synthesis; *hasA*, *hasB*, *hasC*	The production of the capsule was associated with more invasive phenotypes.	Accessory genes.
	Interleukin-8 protease; *spyCEP*	Expression associated with invasive diseases.	Core gene.
	Streptodornase D; *sdaI*	Expression associated with invasive diseases.	Accessory gene.
Corynebacterium diphtheriae	Diphtheria toxin; *tox*	Clear role in the pathogenesis of diphtheria.	The absence of the gene rules out a toxigenic strain. If detected, the occurrence of a non-toxigenig *tox*-bearing strain cannot be excluded.

Data gathered and adapted from Robins-Browne et al., Tagini et al., Giulieri et al., Ye et al., Catalán-Nájera et al., Cole et al. and Turner et al. [3, 80, 89, 124–126]. EPEC, enteropathogenic *E. coli*; EIEC, enteroinvasive *E. coli*; ETEC, enterotoxigenic E. coli; EHEC, enterohaemorrhagic *E. coli*; EAEC, enteroaggregative *E. coli*; DAEC, diffusely adhering *E. coli*; LEE PAI, Locus of enterocyte effacement pathogenicity island; pINV and pAA, virulence plasmids of EIEC and EAEC, respectively; SSTI, skin and soft tissue infection

The characterisation of a strain in the presence of a severe clinical presentation could be beneficial for the patient or for the public health measures to prevent transmission that may occur. In the context of toxic shock syndrome, sequencing of the *S. aureus* or *S. pyogenes* strain can potentially bring useful information. For instance, when dealing with clustered cases of severe infection, WGS may be not only useful to investigate the genetic distance between organisms but also to characterise the hypervirulent emerging clone [72]. However, the management of toxic shock syndrome should be independent of WGS results. Indeed, clinical recognition of the syndrome, cardiovascular resuscitation, removal of the source (foreign body removal or surgery), antibiotic treatment and adjunctive therapies, such as clindamycin or intravenous immunoglobulins, should be introduced in every patient presenting these syndromes without delay [90].

Overall, when facing an outbreak or for epidemiological surveillance, WGS can be used to quickly characterise VFs encoded in the genome of epidemic clones or the emergence of a new virulent strain, which is particularly relevant from a public health or hospital hygiene perspectives. Furthermore, the characterisation of putative bioterrorist agents can be relevant for reference centers [91]. WGS-based clinical management will probably rise in the coming years.

4.3 Methods and Procedures

To implement WGS in clinical microbiology or public health laboratories, the genomic platform needs to fulfill national and international standards in laboratory medicine (for instance, the ISO 15189:2012 certification) [92]. Therefore, each part of the workflow should benefit from standard operating procedures (SOPs), each lab member from competency assessments and a strict monitoring of laboratory supplies should be done [93]. To ensure good quality results, several specific quality control checkpoints as well as the use of internal and external quality controls are required [94] (Table 4.2). External quality control programmes (proficiency testing) are currently developed for WGS in microbiology, but most of them have initially been aimed at assessing the performance of WGS in outbreak investigation, SNP calling and antibiotic resistance gene detection [95]. Furthermore, each proposed analysis should undergo validation studies, through research and development projects, to assess how the technique performs in the laboratory setting. Such accreditation has been completed by our bacterial genomics diagnostic laboratory in 2018 for typing and resistance genes prediction of any bacterial strain as well as for a few virulence factors of *S. aureus*, since analysis of "virulome" information is still in the process of routine implementation.

Table 4.2 General quality metrics and controls for an Illumina-based workflow

Category	Metrics	Comments
DNA extraction	DNA concentration	Hard criteria (e.g. >1 ng/ μl).
	DNA purity	Hard criteria (e.g. A260/280 ratio ≥ 0.7).
Library preparation	Library size concentration	Hard criteria (e.g. ≥1 nM).
	Library size distribution	Hard criteria (e.g. fragments sizes should be comprised between <0.3 kb and > 3 kb).
Sequencing	PhiX control	Hard criteria; assessment of error rates (e.g. <4.9%).
	Per cent of bases with quality score > Q30 FOR the run (%Q30)	Hard criteria; it should be defined according to the platform characteristics.
Reads QC	Number of reads passing the QC	This metric should provide sufficient theoretical coverage (number of reads * mean read length / expected size of the genome).
Assemblies	Assembly length	Soft criteria; should correspond to the expected genome size.
	Number of contigs	Soft criteria; it should correspond to expected fragmentation level.
	GC content	Soft criteria; it should correspond to the expected GC content.
	N50	Soft criteria; it is a metric of the assembly fragmentation and should be a warning sign when lower than an expected cutoff.
	L50	Soft criteria; number of contigs larger or equal than N50. It should be a warning when higher than an expected cutoff.
	Sequencing depth	Mean sequencing depth should be above a cutoff (e.g. hard cutoff of 60x; soft cutoff of 100x).
Detection of the VFs	Percentage of detected VF in a control strain	Hard criteria; the reason for the false negative should be determined before any interpretation of the run.
Detection of the variants	Number of called mutations of variant in a control strain	Hard criteria; if above a defined cutoff, the reason should be determined.

As for all WGS analysis, it is also recommended to use one control strain (e.g. for each run) that should pass each of the QC checkpoints and would be a positive control for the complete workflow [94]. In addition, external quality controls should also be performed on a regular basis. The use of a no-template control can also be added to control for de-multiplexing errors of Illumina. Hard criteria mean that if these steps do not pass the cutoff, the analysis should be repeated. When a value does not reach a soft criterion, this needs to be highlighted and should be critically interpreted. N50, size of the contig lying at the 50% of the total assembly length when contigs are ordered by sizes; L50, the smallest number of contigs to reach the N50

4.3.1 Sequencing

Whole-genome sequencing is usually performed from pure bacterial culture. Standard procedures are increasingly available for this technique [96]. The choice of sequencer should be done to match the laboratory settings, budget and desired

output [92, 97]. The main technical features of available sequencers was already discussed elsewhere [98]. In recent years, the Illumina company (San Diego, USA) took a large proportion of the sequencing market share [99]. Short read technologies can answer many clinical questions, including in terms of virulence assessment. Long read technologies, such as Nanopore Sequencing (Oxford Nanopore, GB) or Pacific Biosciences (Menlo Park, USA), have the advantage of improving the genome assembly by facilitating the resolution of repeats. Solving the genome structure provides additional useful information, such as the genomic localisation of virulence factors (ex: on a plasmid or a chromosomal location), which could indicate the potential for transmissibility of virulence factors between strains. Furthermore, the Nanopore sequencing technology can acquire data in real-time, which can speed up the time-to-diagnosis; as soon as a defined number of reads has been acquired, a preliminary interpretation could be theoretically done, potentially reducing the turnaround time.

4.3.2 Software and Pipelines

Raw sequencing reads should first be quality controlled and trimmed for remaining adapters (when using an Illumina sequencer) or low-quality nucleotides (for instance, using Trimmomatic [100]). Then, three main approaches can be used (Fig. 4.2):

(i) Virulence factors can be searched in the assembled genome (e.g. assembled using SPAdes [101]). A reference database of known virulence factors is used to identify homologs in the newly sequenced genome. The detection of virulence factors can be made with any sequence similarity search tool such as BLAST (NCBI). Searches can be performed using nucleotide or amino acid sequences. The use of amino acid sequences is less stringent than nucleotide sequences and allows the detection of more distantly related homologs. The presence of repeats in bacterial genomes impairs the assembly of the complete genome from short reads. Most assemblies are split in multiple fragments (what is commonly referred as "contigs") If a gene is split over two different contigs, it might be missed by homology search tools; particularly if results are filtered based on query coverage, which is actually recommended to avoid detecting only fragments sharing high sequence identity. An example of a specific tool is Kleborate, which is dedicated to the assessment of *Klebsiella spp*. virulence factors directly from genome assemblies [102].

(ii) The second approach bypass the assembly step and search for known virulence factors in reads, reducing the probability of false negative results associated with gapped assemblies. Various tools can be used to map reads (e.g BWA, Samtools, Diamond) on a chosen reference genome or a set of virulence factors [103–105]. This allows the detection of specific genetic variants of virulence genes. Alternatively, specialised homology search tools for big datasets
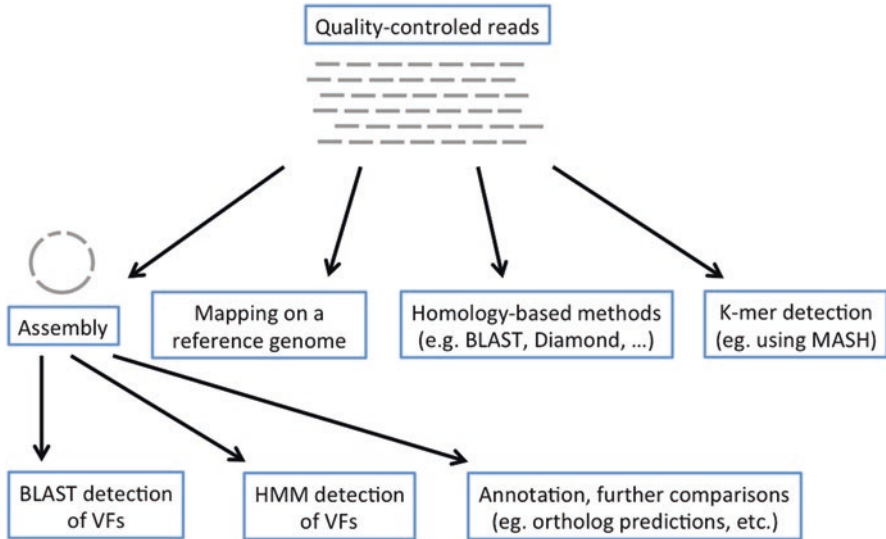
Fig. 4.2 General overview of the possible bioinformatic alternatives to detect virulence factors (VFs). Several analyses are proposed by online pipelines. This is a simplified view; approaches can be combined. BLAST analyses can also be performed after gene annotation (for instance, pipelines can BLAST predicted amino acid sequences on the reference database of virulence factors). HMM, hidden Markov model

    (e.g Diamond) can be used to perform direct searches of virulence factors in reads.

(iii) Methods based on the detection of *k*-mer associated with virulence, combine both the advantage of detecting variants and genes [106]. As a reminder, *k*-mers are *in silico* fragments of *k* size of a DNA sequence (i.e. reads). The general principle of this technique is to look for exact matches, allowing to check for the presence of specific genetic variants and of any sequence of interest. Nevertheless, this approach will fail to identify virulence factor genes that diverge too much from the reference sequence. When dealing with poorly characterised pathogens, Hidden Markov Models (HMM) can be used to detect amino acid sequences sharing low sequence identity but sharing likely a similar functional domain and by extension, functions. For instance, a curated database of protein domains associated to antibiotic resistance, called ResFAM, was recently developed [107]. Similarly, dedicated curated databases for virulence domain, when created, would be useful in the same manner. For secretion systems, a tool using HMM called Macsyfinder was developed [108]. However, the use of HMM may be limited to the research field, since its principal advantage is to identify domains encoded in distant bacterial genomes.

    To ensure reproducibility and to fulfill laboratory medicine standards, the use of robust and versioned bioinformatics pipeline is required. Depending on the bioinformatics workforce available at a given setting, clinical microbiologists may prefer

commercial, *in-house* software or web-based pipelines. Commercial software has the advantage of being developed and validated by the company. For instance, the Ridom SeqSphere and software, which can perform a broad range of WGS analyses, can also detect a set of virulence genes and variants developed for a DNA-microarray detecting *Staphylococcus aureus* virulence and resistance genes [109]. However, commercial software tools are usually black boxes, which prevent the understanding and correct interpretation of inherent technical limits. Conversely, *in-house* software or open-source software allow continuous developments and provide a larger flexibility when the analysis needs to be adapted to a certain case, to specific set of virulence factors. As many software tools have dependencies, dedicated tools allow the creation of stable informatic environments. Furthermore, the whole pipeline can be contained in virtual machines. One example of developed pipeline is MetaGenLab pipeline (docker container available https://hub.docker.com/r/metagenlab/diag_pipelines), which is a versioned snakemake pipeline, calling software using a conda environment to perform typing, antibiotic resistance and virulence analyses (development open source version available on GitHub (https://github.com/metagenlab). Finally, online resources such as VFanalyzer and PATRIC are briefly discussed below. Concerns about online platforms include data protection of the patients and, traceability, versioning, and reproducibility of the results.

4.3.3 Databases

A good reference is necessary to make reliable identifications of virulence factors. Several databases have been designed specifically for virulence factors or contain specific sections associated to virulence. The most widely used database of virulence factors for human pathogens is the curated Virulence Factor Database (VFDB) [110, 111]. VFDB is associated with a web-resource (named VFanalyzer) to submit assembled sequences for online analysis [112]. Victor is another manually curated database of virulence factors integrating data from bacterial, viral and eukaryotic pathogens [113]. The Pathosystems Resource Integration Center (PATRIC) is a large database integrating various data, including genomics, transcriptomics, protein–protein interactions, three-dimensional protein structures and sequence typing data as well as associated metadata [114]. PATRIC integrates data from both VFDB and Victors databases as well as additional manually curated virulence factors [115]. Sequencing reads can be submitted to the PATRIC platform for analysis [116]. Finally, PHI-base is a database containing curated genes involved in host-pathogen interactions [117]. Initially focusing on plant pathogens, it also contains approximately 30% of data on bacteria of medical and environmental importance [117].

Surprisingly, there is very little overlap between the VFs indexed in those four databases (Fig. 4.3). It means that each dataset presents very specific VFs. Focusing on *S. aureus* VFs, we observed that only VFDB (as well as PATRIC since it integrates VFDB data) presented the classical and main VFs that could be expected for this pathogen (Table 4.1). For Victors and PHI-base, many of listed VFs were
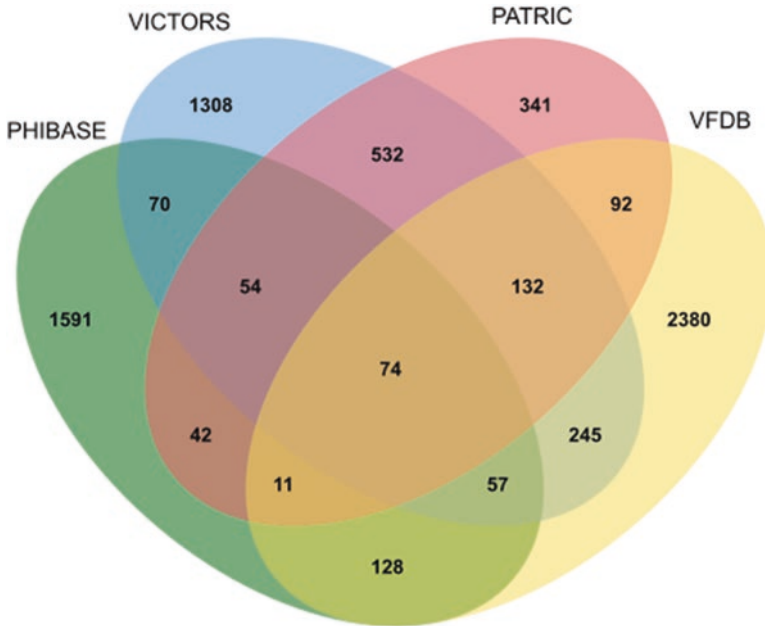
Fig. 4.3 Comparison of the content of VFDB, VICTORS, PHI-Base and PATRIC_VF (February 2019). Protein sequences were clustered at 90% of amino acid sequence identity with silix (https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-116). PATRIC also integrate VFDB and VICTORS VFs and those were discarded from the analysis

identified from large screens for mutants presenting attenuated in virulence as compared to wild-type strains (as reported for instance by Mei et al. [118]). As of 2019, VFDB or PATRIC seem to be the most suited database to investigate the virulence of human pathogens.

4.3.4   Ontologies

There is a need for standardised vocabularies (or ontologies) to properly describe virulence factors and their interactions with their host. Ontologies provide a denomination reference and should be resistant to homonymy. VFDB uses a standardised classification system and PHI-base uses standardised terms to annotate VFs, but there are currently no standards that are used by those web resources. A lot of work is still needed to setup an ontology for pathogen-host interactions and virulence factors that could be used to harmonise the information stored in curated VF databases [119].

4.3.5   Further Developments

In the previous section, we discussed methods focusing on WGS from pure bacterial cultures. However, the development of culture-independent approaches, such as shotgun metagenomics, is very promising. Indeed, the detection of specific reads associated with virulence factors in various specimen types, could be sufficient in order to take measures to improve patient management. This approach has already provided promising results for pathogen detection in the context of antibiotic treatments or for the diagnosis of fastidious bacteria [120, 121]. Going further than assessing the presence and absence of genes associated to virulence, the development of dedicated diagnostic tools and databases allowing to effectively characterise SNPs (in coding and non-coding sequences) as well as insertion / deletions events, would be a major advance in the understanding of virulence. In addition, conditional gene annotation, taking into account the presence of other genetic features (mutations or other genes), could help determine the virulence of a gene. Combination with other omics technologies, such as transcriptomics and proteomics could be the next revolution in clinical microbiology. For instance, the development of SWATH-MS for proteomics analyses showed promising results for *M. tuberculosis* infection [122]. It is also likely that particular VF could also be detected and integrated in the clinical interpretation.

WGS opens up the possibility to perform large-scale studies (correlation studies) in order to identify putative variant or genes with prognosis value [106]. Furthermore, a progressive integration of WGS data in clinical score or even genomic status of the host could be the next step needed to reach a good predictive score, always aiming for more personalised medicine. Predictive data could set for instance the indication for a dedicated treatment, follow-up or management.

4.4   Interpretation, Validation and Impact

As for any microbiological analysis, the interpretation of the results should take into account pre-analytical and analytical variables. Several quality scores should be assessed in order to validate the analysis (Table 4.2). Hard criteria (e.g. if below, the analysis should be repeated) and soft criteria may be used (e.g. if borderline with a quality metrics, results can still be interpreted depending on the rest of the metrics and on the performed analysis). Once the analysis has been validated technically, a critical interpretation must be done by the clinical microbiologist, who should be able to integrate (i) technical aspects and understand the limits of the test, and (ii) the clinical and microbiological aspects, such as the isolation site, the suspected disease, the biology of the bacterial species, etc. All these data should finally be transmitted to the clinician requesting the analysis using a standardised report. The format of these reports can be complex to design, as it is required to present in a highly summarised way the main patient's data, the main genomic findings and the

interpretation. To help design such a report, back-and-forth discussions between the technical team, the clinical microbiologists and the physicians should be done [123]. As WGS data also includes typing analysis and the identification of antibiotic resistant determinant, generic reports must have the possibility to integrate all these data in a comprehensive manner.

The training in the interpretation of WGS analyses will be a challenge in the coming years for clinical microbiologists, particularly in the context of a rapidly evolving technology. Clinical microbiologists will also have to teach this to a variety of medical personal such as medical students, infectious diseases specialists, epidemiologists and any person involved in the management of patients with severe infectious disease (e.g. intensive-care specialists, …).

4.5   Future Perspectives

As of 2019, WGS appears to be a game-changing technology for clinical microbiology because it allows the broad detection of any DNA sequence, regardless of the availability of a specific diagnostic test (e.g. PCRs). As such, sequencing platforms definitely open up the possibility to detect specific virulence factors in a competitive turnaround time if the strain needs to be sent to a national reference center. Furthermore, WGS allows the epidemiological surveillance of the emergence of virulent clones, therefore possibly preventing the spread of these at early stages.

However, virulence assessment using WGS has not yet revealed its full potential. Indeed, for many situations, the technique is limited by its poor predictive value in terms of patient outcome, which depends on the expression level of virulence factors as well as on the host's susceptibility. Many developments are foreseen thanks to various methods, including the identification of new targets, their implication in clinical scores, and the combinations with other omics techniques, such as transcriptomics and proteomics, which could be the next developmental steps. On the genomic side, not only the detection and characterisation of virulence factors will be further developed, but also the detection of specific variants in regulatory mechanisms associated to increased virulence.

Conflict of Interest   The authors declare no conflict of interest.

References

1. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW (2012) Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13(9):601–612
2. Bertelli C, Greub G (2013) Rapid bacterial genome sequencing: methods and applications in clinical microbiology. Clin Microbiol Infect 19(9):803–813
3. Tagini F, Greub G (2017) Bacterial genome sequencing in clinical microbiology: a pathogen-oriented review. Eur J Clin Microbiol Infect Dis 36(11):2007–2020

4. Casadevall A, Pirofski L (2003) The damage-response framework of microbial pathogenesis. Nat Rev Microbiol 1(1):17–24

5. Rudkin JK, McLoughlin RM, Preston A, Massey RC (2017) Bacterial toxins: offensive, defensive, or something else altogether? PLoS Pathog 13(9):e1006452

6. Peraro MD, van der Goot FG (2015) Pore-forming toxins: ancient, but never really out of fashion. Nat Rev Microbiol 14:77

7. Macfarlane MG (1950) The biochemistry of bacterial toxins. 5. Variation in haemolytic activity of immunologically distinct lecithinases towards erythrocytes from different species. Biochem J 47(3):270–279

8. Henkel JS, Baldwin MR, Barbieri JT (2010) Toxins from Bacteria. EXS 100:1–29

9. Berube BJ, Bubeck Wardenburg J (2013) Staphylococcus aureus α-toxin: nearly a century of intrigue. Toxins 5(6):1140–1166

10. Otto M (2014) Staphylococcus aureus toxins. Curr Opin Microbiol 0:32–37

11. Kayal S, Charbit A (2006) Listeriolysin O: a key protein of Listeria monocytogenes with multiple functions. FEMS Microbiol Rev 30(4):514–529

12. Alouf JE (1980) Streptococcal toxins (streptolysin O, streptolysin S, erythrogenic toxin). Pharmacol Ther 11(3):661–717

13. Sharp GW, Hynie S (1971) Stimulation of intestinal adenyl cyclase by cholera toxin. Nature 229(5282):266–269

14. Chen LC, Rohde JE, Sharp GW (1971) Intestinal adenyl-cyclase activity in human cholera. Lancet Lond Engl 1(7706):939–941

15. Cohen MS, Chang P (2018) Insights into the biogenesis, function, and regulation of ADP-ribosylation. Nat Chem Biol 14(3):236–243

16. Currie MG, Fok KF, Kato J, Moore RJ, Hamra FK, Duffin KL et al (1992) Guanylin: an endogenous activator of intestinal guanylate cyclase. Proc Natl Acad Sci U S A 89(3):947–951

17. Kaper JB, Nataro JP, Mobley HL (2004) Pathogenic Escherichia coli. Nat Rev Microbiol 2(2):123–140

18. Schulz S, Green CK, Yuen PS, Garbers DL (1990) Guanylyl cyclase is a heat-stable enterotoxin receptor. Cell 63(5):941–948

19. Melvin JA, Scheller EV, Miller JF, Cotter PA (2014) Bordetella pertussis pathogenesis: current and future challenges. Nat Rev Microbiol 12(4):274–288

20. Graf R, Codina J, Birnbaumer L (1992) Peptide inhibitors of ADP-ribosylation by pertussis toxin are substrates with affinities comparable to those of the trimeric GTP-binding proteins. Mol Pharmacol 42(5):760–764

21. Simon NC, Aktories K, Barbieri JT (2014) Novel bacterial ADP-ribosylating toxins: structure and function. Nat Rev Microbiol 12(9):599–611

22. Boyer NH, Weinstein L (1948) Diphtheritic myocarditis. N Engl J Med 239(24):913–919

23. Kneen R, Nguyen MD, Solomon T, Pham NG, Parry CM, Nguyen TTH et al (2004) Clinical features and predictors of diphtheritic cardiomyopathy in Vietnamese children. Clin Infect Dis Off Publ Infect Dis Soc Am 39(11):1591–1598

24. Melton-Celsa AR. Shiga Toxin (Stx) classification, structure, and function. Microbiol Spectr [Internet]. 2014 [cited 2019 Feb 5];2(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4270005/

25. Li S, Huang H, Rao X, Chen W, Wang Z, Hu X (2014) Phenol-soluble modulins: novel virulence-associated peptides of staphylococci. Future Microbiol 9(2):203–216

26. Cheung GYC, Joo H-S, Chatterjee SS, Otto M (2014) Phenol-soluble modulins – critical determinants of staphylococcal virulence. FEMS Microbiol Rev 38(4):698–719

27. Spaulding AR, Salgado-Pabón W, Kohler PL, Horswill AR, Leung DYM, Schlievert PM (2013) Staphylococcal and streptococcal superantigen exotoxins. Clin Microbiol Rev 26(3):422–447

28. Rossetto O, Scorzeto M, Megighian A, Montecucco C (2013) Tetanus neurotoxin. Toxicon 66:59–63

29. Rossetto O, Pirazzini M, Montecucco C (2014) Botulinum neurotoxins: genetic, structural and mechanistic insights. Nat Rev Microbiol 12(8):535–549

30. Galán JE, Collmer A (1999) Type III secretion machines: bacterial devices for protein delivery into host cells. Science 284(5418):1322–1328
31. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L et al (1998) Genome sequence of an obligate intracellular pathogen of humans: chlamydia trachomatis. Science 282(5389):754–759
32. Deng W, Marshall NC, Rowland JL, McCoy JM, Worrall LJ, Santos AS et al (2017) Assembly, structure, function and regulation of type III secretion systems. Nat Rev Microbiol 15(6):323–337
33. Grohmann E, Christie PJ, Waksman G, Backert S (2018) Type IV secretion in gram-negative and gram-positive bacteria. Mol Microbiol 107(4):455–471
34. Russell AB, Peterson SB, Mougous JD (2014) Type VI secretion system effectors: poisons with a purpose. Nat Rev Microbiol 12(2):137–148
35. Brodmann M, Dreier RF, Broz P, Basler M (2017) Francisella requires dynamic type VI secretion system and ClpB to deliver effectors for phagosomal escape. Nat Commun 16(8):15853
36. Simeone R, Bottai D, Frigui W, Majlessi L, Brosch R (2015) ESX/type VII secretion systems of mycobacteria: Insights into evolution, pathogenicity and protection. Tuberculosis 95(Supplement 1):S150–S154
37. Bottai D, Gröschel MI, Brosch R (2017) Type VII secretion systems in gram-positive bacteria. Curr Top Microbiol Immunol 404:235–265
38. Renshaw PS, Panagiotidou P, Whelan A, Gordon SV, Hewinson RG, Williamson RA et al (2002) Conclusive evidence that the major T-cell antigens of the Mycobacterium tuberculosis complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence. J Biol Chem 277(24):21598–21603
39. Okaro U, Addisu A, Casanas B, Anderson B (2017) Bartonella species, an emerging cause of blood-culture-negative endocarditis. Clin Microbiol Rev 30(3):709–746
40. Oliveira WF, Silva PMS, Silva RCS, Silva GMM, Machado G, Coelho LCBB et al (2018) Staphylococcus aureus and Staphylococcus epidermidis infections on implants. J Hosp Infect 98(2):111–117
41. Tagini F, Pillonel T, Asner S, Prod'hom G, Greub G (2016) Draft genome sequence of a cardiobacterium hominis strain isolated from blood cultures of a patient with infective endocarditis. Genome Announc 4(5)
42. Riess T, Raddatz G, Linke D, Schäfer A, Kempf VAJ (2007) Analysis of Bartonella adhesin a expression reveals differences between various B. henselae strains. Infect Immun 75(1):35–43
43. Yumoto H, Azakami H, Nakae H, Matsuo T, Ebisu S (1996) Cloning, sequencing and expression of an Eikenella corrodens gene encoding a component protein of the lectin-like adhesin complex. Gene 183(1–2):115–121
44. Merz AJ, Enns CA, So M (1999) Type IV pili of pathogenic Neisseriae elicit cortical plaque formation in epithelial cells. Mol Microbiol 32(6):1316–1332
45. Kolappan S, Coureuil M, Yu X, Nassif X, Egelman EH, Craig L (2016) Structure of the Neisseria meningitidis type IV pilus. Nat Commun 7:13015
46. Flemming H-C, Wingender J (2010) The biofilm matrix. Nat Rev Microbiol 8(9):623–633
47. Flemming H-C, Wingender J, Szewzyk U, Steinberg P, Rice SA, Kjelleberg S (2016) Biofilms: an emergent form of bacterial life. Nat Rev Microbiol 14(9):563–575
48. Peetermans M, Vanassche T, Liesenborghs L, Claes J, Vande Velde G, Kwiecinksi J et al (2014) Plasminogen activation by staphylokinase enhances local spreading of S. aureus in skin infections. BMC Microbiol 14:310
49. McArthur JD, Cook SM, Venturini C, Walker MJ (2012) The role of streptokinase as a virulence determinant of Streptococcus pyogenes--potential for therapeutic targeting. Curr Drug Targets 13(3):297–307
50. Elgaml A, Miyoshi S-I (2017) Regulation systems of protease and hemolysin production in Vibrio vulnificus. Microbiol Immunol 61(1):1–11

51. Hyde JA (2017) Borrelia burgdorferi keeps moving and carries on: a review of Borrelial dissemination and invasion. Front Immunol [Internet]. [cited 2019 Feb 11];8. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2017.00114/full

52. Boehm M, Simson D, Escher U, Schmidt A-M, Bereswill S, Tegtmeyer N et al (2018) Function of serine protease HtrA in the lifecycle of the foodborne pathogen campylobacter jejuni. Eur J Microbiol Immunol 8(3):70–77

53. Storisteanu DML, Pocock JM, Cowburn AS, Juss JK, Nadesalingam A, Nizet V et al (2017) Evasion of neutrophil extracellular traps by respiratory pathogens. Am J Respir Cell Mol Biol 56(4):423–431

54. Delany I, Seib KL (2012) Stress response in microbiology. Horizon Scientific Press, Requena JM. 450 p

55. Mandell GL (1975) Catalase, superoxide dismutase, and virulence of Staphylococcus aureus. In vitro and in vivo studies with emphasis on staphylococcal--leukocyte interaction. J Clin Invest 55(3):561–566

56. Messina CGM, Reeves EP, Roes J, Segal AW (2002) Catalase negative Staphylococcus aureus retain virulence in mouse model of chronic granulomatous disease. FEBS Lett 518(1–3):107–110

57. Berenger B, Chen J, Bernier A-M, Bernard K (2016) Draft whole-genome sequence of a catalase-negative Staphylococcus aureus subsp. aureus (sequence type 25) Strain isolated from a patient with endocarditis and septic arthritis. Genome Announc 4(6)

58. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS (2010) The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infect Immun 78(2):704

59. Weiss G, Schaible UE (2015) Macrophage defense mechanisms against intracellular bacteria. Immunol Rev 264(1):182–203

60. Ray K, Marteyn B, Sansonetti PJ, Tang CM (2009) Life on the inside: the intracellular lifestyle of cytosolic bacteria. Nat Rev Microbiol 7(5):333–340

61. Mitchell G, Chen C, Portnoy DA (2016) Strategies used by bacteria to grow in macrophages. Microbiol Spectr 4(3)

62. Edwards RJ, Taylor GW, Ferguson M, Murray S, Rendell N, Wrigley A et al (2005) Specific C-terminal cleavage and inactivation of interleukin-8 by invasive disease isolates of Streptococcus pyogenes. J Infect Dis 192(5):783–790

63. Weiser JN, Bae D, Fasching C, Scamurra RW, Ratner AJ, Janoff EN (2003) Antibody-enhanced pneumococcal adherence requires IgA1 protease. Proc Natl Acad Sci U S A 100(7):4215–4220

64. von Pawel-Rammingen U, Johansson BP, Björck L (2002) IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. EMBO J 21(7):1607–1615

65. McNeil LK, Zagursky RJ, Lin SL, Murphy E, Zlotnick GW, Hoiseth SK et al (2013) Role of factor H binding protein in Neisseria meningitidis virulence and its potential as a vaccine candidate to broadly protect against meningococcal disease. Microbiol Mol Biol Rev MMBR 77(2):234–252

66. Merino S, Camprubí S, Albertí S, Benedí VJ, Tomás JM (1992) Mechanisms of Klebsiella pneumoniae resistance to complement-mediated killing. Infect Immun 60(6):2529–2535

67. Sheldon JR, Laakso HA, Heinrichs DE (2016) Iron acquisition strategies of bacterial pathogens. Microbiol Spectr. 4(2)

68. Tao X, Schiering N, Zeng HY, Ringe D, Murphy JR (1994) Iron, DtxR, and the regulation of diphtheria toxin expression. Mol Microbiol 14(2):191–197

69. Cornejo E, Schlaermann P, Mukherjee S (2017) How to rewire the host cell: a home improvement guide for intracellular bacteria. J Cell Biol 216(12):3931–3948

70. Samanta D, Mulye M, Clemente TM, Justis AV, Gilk SD (2017) Manipulation of host cholesterol by obligate intracellular bacteria. Front Cell Infect Microbiol 7:165

71. Juhas M (2015) Horizontal gene transfer in human pathogens. Crit Rev Microbiol 41(1):101–108

72. Tagini F, Aubert B, Troillet N, Pillonel T, Praz G, Crisinel PA et al (2017) Importance of whole genome sequencing for the assessment of outbreaks in diagnostic laboratories: analysis of a case series of invasive Streptococcus pyogenes infections. Eur J Clin Microbiol Infect Dis 36(7):1173–1180

73. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P et al (2008) The Pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol 190(20):6881–6893

74. Namouchi A, Didelot X, Schöck U, Gicquel B, Rocha EPC (2012) After the bottleneck: genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. Genome Res 22(4):721–734

75. Churchward G (2007) The two faces of Janus: virulence gene regulation by CovR/S in group A streptococci. Mol Microbiol 64(1):34–41

76. Quereda JJ, Cossart P (2017) Regulating bacterial virulence with RNA. Ann Rev Microbiol 71:263–280

77. Kazmierczak MJ, Wiedmann M, Boor KJ (2005) Alternative sigma factors and their roles in bacterial virulence. Microbiol Mol Biol Rev MMBR 69(4):527–543

78. Xiong YQ, Fowler VG, Yeaman MR, Perdreau-Remington F, Kreiswirth BN, Bayer AS (2009) Phenotypic and genotypic characteristics of persistent methicillin-resistant Staphylococcus aureus bacteremia in vitro and in an experimental endocarditis model. J Infect Dis 199(2):201–208

79. Chattaway MA, Day M, Mtwale J, White E, Rogers J, Day M et al (2017) Clonality, virulence and antimicrobial resistance of enteroaggregative Escherichia coli from Mirzapur, Bangladesh. J Med Microbiol 66(10):1429–1435

80. Robins-Browne RM, Holt KE, Ingle DJ, Hocking DM, Yang J, Tauschek M (2016) Are Escherichia coli Pathotypes still relevant in the era of whole-genome sequencing? Front Cell Infect Microbiol 6:141

81. Clements A, Young JC, Constantinou N, Frankel G (2012) Infection strategies of enteric pathogenic Escherichia coli. Gut Microbes 3(2):71–87

82. Walters LL, Raterman EL, Grys TE, Welch RA (2012) Atypical Shigella boydii 13 encodes virulence factors seen in attaching and effacing Escherichia coli. FEMS Microbiol Lett 328(1):20–25

83. Tagini F, Pillonel T, Croxatto A, Bertelli C, Koutsokera A, Lovis A et al (2018) Distinct genomic features characterize two clades of Corynebacterium diphtheriae: proposal of Corynebacterium diphtheriae Subsp. diphtheriae Subsp. nov. and Corynebacterium diphtheriae Subsp. lausannense Subsp. nov. Front Microbiol 9:1743

84. Saeed K, Gould I, Esposito S, Ahmad-Saeed N, Ahmed SS, Alp E et al (2018) Panton–valentine leukocidin-positive Staphylococcus aureus: a position statement from the International Society of Chemotherapy. Int J Antimicrob Agents 51(1):16–25

85. Shallcross LJ, Fragaszy E, Johnson AM, Hayward AC (2013) The role of the Panton-valentine leucocidin toxin in staphylococcal disease: a systematic review and meta-analysis. Lancet Infect Dis 13(1):43–54

86. Grumann D, Nübel U, Bröker BM (2014) Staphylococcus aureus toxins--their functions and genetics. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 21:583–592

87. Micek ST, Dunne M, Kollef MH (2005) Pleuropulmonary complications of Panton-valentine Leukocidin-positive community-acquired methicillin-resistant Staphylococcus aureus: importance of treatment with antimicrobials inhibiting exotoxin production. Chest 128(4):2732–2738

88. Laabei M, Recker M, Rudkin JK, Aldeljawi M, Gulay Z, Sloan TJ et al (2014) Predicting the virulence of MRSA from its genome sequence. Genome Res 24(5):839–849

89. Giulieri SG, Holmes NE, Stinear TP, Howden BP (2016) Use of bacterial whole-genome sequencing to understand and improve the management of invasive Staphylococcus aureus infections. Expert Rev Anti-Infect Ther 14(11):1023–1036

90. Wilkins AL, Steer AC, Smeesters PR, Curtis N (2017) Toxic shock syndrome – the seven Rs of management and treatment. J Infect 74:S147–S152

91. Gilchrist CA, Turner SD, Riley MF, Petri WA, Hewlett EL (2015) Whole-genome sequencing in outbreak analysis. Clin Microbiol Rev 28(3):541–563

92. Rossen JWA, Friedrich AW, Moran-Gilad J (2018) Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. Clin Microbiol Infect 24(4):355–360

93. Carey RB, Bhattacharyya S, Kehl SC, Matukas LM, Pentella MA, Salfinger M et al (2018) Implementing a quality management system in the medical microbiology laboratory. Clin Microbiol Rev 31(3)

94. Kozyreva VK, Truong C-L, Greninger AL, Crandall J, Mukhopadhyay R, Chaturvedi V (2017) Validation and implementation of clinical laboratory improvements act-compliant whole-genome sequencing in the public health microbiology laboratory. J Clin Microbiol 55(8):2502–2520

95. Moran-Gilad J, Sintchenko V, Pedersen SK, Wolfgang WJ, Pettengill J, Strain E et al (2015) Proficiency testing for bacterial whole genome sequencing: an end-user survey of current capabilities, requirements and priorities. BMC Infect Dis 15:174

96. Fricke WF, Rasko DA (2014) Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. Nat Rev Genet 15(1):49–55

97. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W et al (2017) Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin Microbiol Rev 30(4):1015–1063

98. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E (2018) Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin Microbiol Infect 24(4):335–341

99. CNBC.com AP special to. Illumina manufactures a genetically-designed market disaster [Internet]. 2016 [cited 2017 Oct 20]. Available from: https://www.cnbc.com/2016/10/12/illumina-manufactures-a-genetically-designed-market-disaster.html

100. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120

101. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477

102. Lam MMC, Wick RR, Wyres KL, Gorrie CL, Judd LM, Jenney AWJ, et al (2018) Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICEKp in Klebsiella pneumoniae populations. Microb Genomics [Internet]. [cited 2019 Feb 13] 4(9). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6202445/

103. Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59

104. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al (2009) The sequence alignment/map format and SAMtools. Bioinforma Oxf Engl 25(16):2078–2079

105. Li H, Durbin R (2010) Fast and accurate long-read alignment with burrows-wheeler transform. Bioinforma Oxf Engl 26(5):589–595

106. Méric G, Mageiros L, Pensar J, Laabei M, Yahara K, Pascoe B et al (2018) Disease-associated genotypes of the commensal skin bacterium Staphylococcus epidermidis. Nat Commun 9(1):5034

107. Gibson MK, Forsberg KJ, Dantas G (2015) Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J 9(1):207–216

108. Abby SS, Rocha EPC (1615) Identification of protein secretion systems in bacterial genomes using MacSyFinder. Methods Mol Biol Clifton NJ 2017:1–21

109. Strauß L, Ruffing U, Abdulla S, Alabi A, Akulenko R, Garrine M et al (2016) Detecting Staphylococcus aureus virulence and resistance genes: a comparison of whole-genome sequencing and DNA microarray technology. J Clin Microbiol 54(4):1008–1016

110. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y et al (2005) VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33(Database issue):D325–D328

111. Chen L, Zheng D, Liu B, Yang J, Jin Q (2016) VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. Nucleic Acids Res 44(D1):D694–D697

112. Liu B, Zheng D, Jin Q, Chen L, Yang J (2019) VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res 47(D1):D687–D692

113. Sayers S, Li L, Ong E, Deng S, Fu G, Lin Y et al (2019) Victors: a web-based knowledge base of virulence factors in human and animal pathogens. Nucleic Acids Res 47(D1):D693–D700

114. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL et al (2014) PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42(Database issue):D581–D591

115. Mao C, Abraham D, Wattam AR, Wilson MJC, Shukla M, Yoo HS et al (2015) Curation, integration and visualization of bacterial virulence factors in PATRIC. Bioinforma Oxf Engl 31(2):252–258

116. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2017;45(D1):D535–42

117. Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R et al (2017) PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database. Nucleic Acids Res 45(Database issue):D604–D610

118. Mei J-M, Nourbakhsh F, Ford CW, Holden DW (1997) Identification of Staphylococcus aureus virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. Mol Microbiol 26(2):399–407

119. Korves T, Colosimo ME (2009) Controlled vocabularies for microbial virulence factors. Trends Microbiol 17(7):279–285

120. Lazarevic V, Gaïa N, Girard M, Leo S, Cherkaoui A, Renzi G et al (2018) When bacterial culture fails, metagenomics can help: a case of chronic hepatic Brucelloma assessed by next-generation sequencing. Front Microbiol 9:1566

121. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G et al (2014) Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med 370(25):2408–2417

122. Schubert OT, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE et al (2015) Building high-quality assay libraries for targeted analysis of SWATH MS data. Nat Protoc 10(3):426–441

123. Crisan A, McKee G, Munzner T, Gardy JL (2018) Evidence-based design and evaluation of a whole genome sequencing clinical report for the reference microbiology laboratory. PeerJ 6:e4218

124. Catalán-Nájera JC, Garza-Ramos U, Barrios-Camacho H (2017) Hypervirulence and hyper-mucoviscosity: two different but complementary Klebsiella spp. phenotypes? Virulence 8(7):1111–1123

125. Ye M, Tu J, Jiang J, Bi Y, You W, Zhang Y et al (2016) Clinical and genomic analysis of liver abscess-causing Klebsiella pneumoniae identifies new liver abscess-associated virulence genes. Front Cell Infect Microbiol 6:165

126. Cole JN, Barnett TC, Nizet V, Walker MJ (2011) Molecular insight into invasive group a streptococcal disease. Nat Rev Microbiol 9(10):724–736