UNIVERSITÀ DEGLI STUDI DI MILANO

Doctoral Program in Nutritional Sciences

Department of Health, Animal Science and Food Safety "Carlo Cantoni" (VESPA)

Investigation on the role of nutrition and bacteria in human health using Next Generation Sequencing (NGS): from gut microbiota in immune-related disorders to genomic epidemiology

PhD Candidate
Matteo Perini

Tutor
Prof. Gian Vincenzo Zuccotti

Chair of the Doctoral Program
Prof. Luciano Pinotti

2021 - XXXIV

Table of contents

Abstract

Nutrition and the immune system have a main role in shaping the gut microbial community; on its side, variation in microbiota composition can greatly affect health status. Some bacterial lineages are generally associated with beneficial effects (e.g. *Lactobacillus* and Bifidobacteria in the human gut), while other members of the microbiota hold the potential to cause pathological alterations, and also systemic infections. During my PhD period, I used Next Generation Sequencing (NGS) technologies to investigate these bacteria-human relationships: the influence of gut microbiota on human health and the genetic characterization of bacterial pathogens, mainly opportunistic, that are frequently found in the human gut (pathobionts), or associated with food. More in detail, I studied the influence and the composition of the gut bacterial community in the pediatric age and in immune-related disorders (atopic dermatitis and inflammatory bowel disease). Furthermore, I performed NGS-based genomic epidemiology investigations to reconstruct the transmission routes of bacterial pathogens, such as *Escherichia coli* and *Staphylococcus aureus*. Lastly, I was involved to the development of a novel approach to High Resolution Melting-based subspecies typing (called Hypervariable Locus Melting Typing, HLMT), and I developed a specific HLMT protocol for *Klebsiella pneumoniae*, one of the most important opportunistic pathogens associated with the human gut.

Riassunto

La nutrizione e il sistema immunitario hanno un ruolo fondamentale nel plasmare la comunità microbica intestinale; da parte sua, la variazione nella composizione del microbiota può influenzare notevolmente lo stato di salute. Alcuni ceppi batterici sono generalmente associati ad effetti benefici (ad esempio i Lattobacilli ed i Bifidobatteri nell'intestino umano), mentre altri membri del microbiota possono causare alterazioni patologiche ed infezioni sistemiche. Durante il mio periodo di dottorato, ho utilizzato le tecnologie Next Generation Sequencing (NGS) per indagare queste relazioni batteri-uomo: l'influenza del microbiota intestinale sulla salute umana e la caratterizzazione genetica dei patogeni batterici, principalmente opportunistici, che si trovano frequentemente nell'intestino umano (patobionti) o associati al cibo. Più in dettaglio, ho studiato l'influenza e la composizione della comunità batterica intestinale in età pediatrica e nei disturbi immuno-correlati (dermatite atopica e malattie infiammatorie croniche intestinali). Inoltre, ho eseguito indagini di epidemiologia genomica basate su NGS per ricostruire le vie di trasmissione di agenti patogeni batterici, come *Escherichia coli* e *Staphylococcus aureus*. Infine, sono stato coinvolto nello sviluppo di un nuovo approccio alla tipizzazione delle sottospecie batteriche basata su High Resolution Melting (chiamato Hypervariable Locus Melting Typing, HLMT), e ho sviluppato un protocollo HLMT specifico per *Klebsiella pneumoniae*, uno dei più importanti patogeni opportunisti associati all'intestino umano.

Introduction

Gut microbiota, Nutrition and Immune System

The last decade has been characterized by the ever increasing interest in microbial communities. This can be due to two main reasons: the great revolutions occurred in DNA sequencing technologies, and by the establishment of the concept that resident microbial communities are essential and sometimes vital to their host [1]. In particular, gut microbiota accounts for approximately 100 trillion microorganisms (mostly bacteria) and the gut microbiome (i.e. the collective genomes of the gut microorganisms, or the "metagenome") encodes for over three million genes that are able to substitute and/or implement the functions of the host [2], thus influencing most areas of human health like immunity, appetite and energy metabolism [3]. Indeed, the bond between gut microbiota and the human host is so tight that it can also be seen through an ecological and evolutionary lens: the beneficial human gut microbiota evolves its functions to ensure its persistence in the gut, while the human host attempts to control the composition of microbiota to gain the maximum benefit. This coevolution mechanism was called "evolution on a leash" by Foster et al. 2017 [4], indicating the strong interplay in the evolution of the microbiota together with its human host. In this scenario, as food sources have largely guided the evolution of our species, our microbiota has evolved alongside, being strongly influenced by the diet of its host. Unsurprisingly, even short term dietary interventions are able to significantly change the composition of gut microbiota. On the other hand, the diet of the host directly influences the gut microbiota even at the genomic level: these changes, in turn, lead to alterations in the production of metabolites that directly affect host physiology [5] (Figure 1). Given the benefit of maintaining a healthy gut microbiota, there is an increasing interest in developing and studying dietary interventions and nutritional strategies to modulate its composition. Among these strategies, probiotics, prebiotics and postbiotic compounds play a prominent role. Probiotics are defined as "live microorganisms that, when administered in adequate amounts, confer a health benefit on the host" [6].

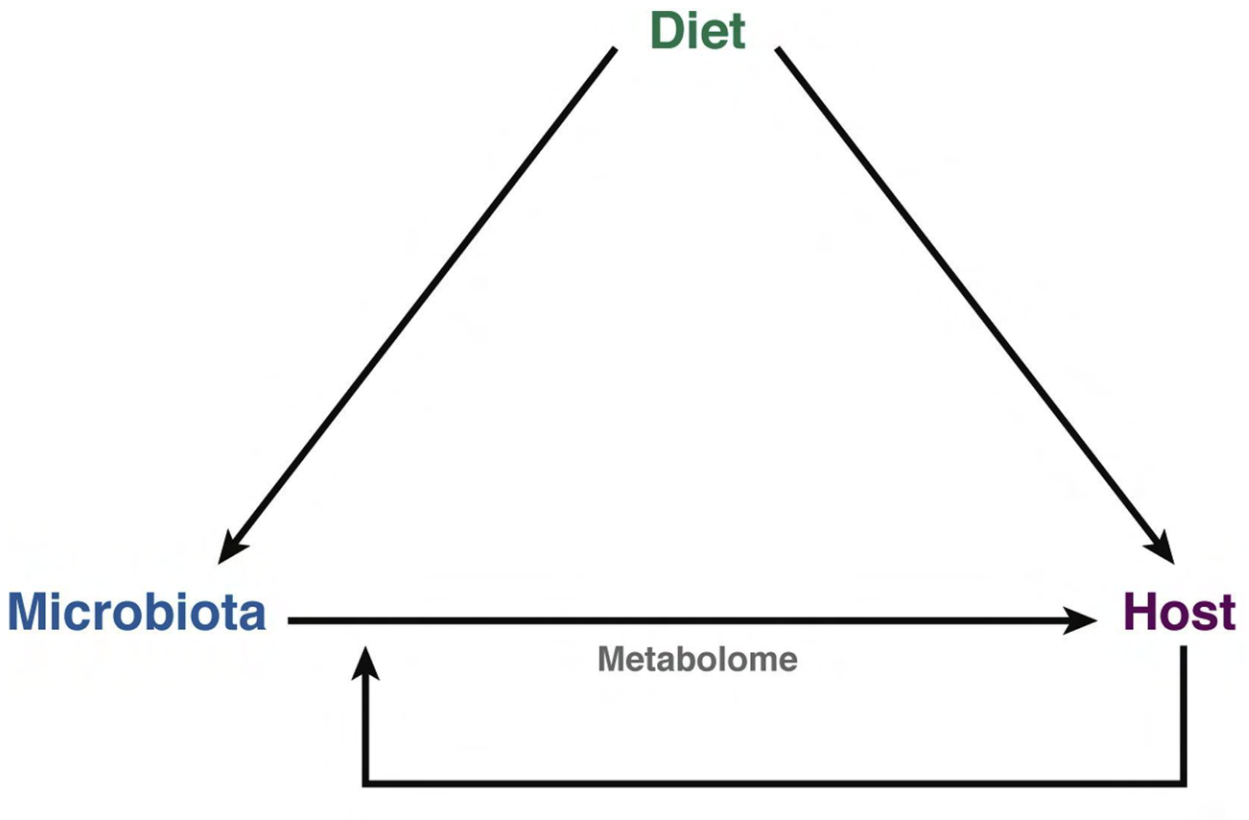


Figure 1: Diet-Microbiota-Host interactions (adapted from Albenberg and Wu 2014 [5])
Diet and nutrition affect directly both the human host and its microbiota, but the effect on the host can be also mediated by the microbiota itself: its metabolome shapes the health status of the host by variation in metabolite production. On the other hand, the human host indirectly shapes the microbiota composition with its diet (among other factors).

These preparations contain a limited list of bacterial genera such as *Lactobacillus* spp. and *Bifidobacterium* spp. that are usually found in higher abundance in the microbiota of healthy individuals [7]. Even if they are shown to have benefit in a number of diseases, it is still unclear if they are able to induce stable alterations in gut microbiota composition that co-occur with health-promoting effects [8]. Conversely, with the term prebiotics we refer to the substances "that are selectively utilized by host microorganisms conferring a health benefit" [9]. Therefore, prebiotic supplementation is expected to stimulate the selective growth of the gut microbiota species that promote health benefits to the human host. Most of these compounds are fermentable carbohydrates that are digested by specific gut microbiota species able to

produce the metabolites needed to maintain or restore the physiological equilibrium (i.e Short Chain Fatty Acids, see below). Lastly, postbiotic is defined by the International Scientific Association of Probiotics and Prebiotics (ISAPP) as "preparation of inanimate microorganisms and/or their components that confers a health benefit on the host" [10]. These compounds were recently developed as supplements and their final definition is still debated [11]. The issue is mostly due to the use of confusing terms like "inanimate" and "and/or": the presence of inactivated or killed cells itself can have an effect in combination with the metabolites produced by those bacteria, hence this definition can again be a bit confusing. Nevertheless, the great advantage of postbiotic preparations is that they are stable at room temperature with a long shelf life, they can be mixed in hot liquids and they have a better safety profile [7]. Regardless, the most studied and active components of postbiotic preparations are Short Chain Fatty Acids (SCFA). SCFA are the major metabolites produced by the gut microbiota [12]. In particular, acetate, propionate, and butyrate are the three main SCFA produced in the anaerobic fermentation of dietary fibers and starch from the microbiota [13]. Even if the role of these metabolites has not been fully elucidated, they have a main role in the so-called "Microbiota-Gut-Brain Axis" [14]. SCFAs directly maintain the equilibrium of the immune system showing effects as immuno-modulator molecules. In detail, SCFAs regulate the activation, differentiation and recruitment of neutrophils, Dendritic Cells (DCs), T cells, macrophages and monocytes [12]. Lower amounts of SCFA, and lower abundance of SCFAs producers in the gut microbiota, was found in patients with Inflammatory Bowel Diseases (IBD), Irritable Bowel Syndrome (IBS), Type 2 Diabetes (T2D), obesity, autoimmune disorders and in cancer [15]. Indeed, these disorders and conditions are also characterized by a particularly altered gut microbiota composition. Nevertheless, gut microbiota varies largely also among healthy individuals [16] even if some strong patterns were found in these variations. In 2010 Arumugam and colleagues described that individuals with similar abundance of some specific bacterial genera in the gut can be clustered together: the variation in gut microbiota composition is generally stratified, not continuous [17]. They called these clusters "enterotypes", and found them to be characterized by the increased relative abundance of *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2) and *Ruminococcus* (enterotype 3) [17]. The

enterotype of an individual was found to not depend on gender, body weight, or national divisions, but rather on dietary patterns. The original enterotypes classification has been discussed [18] and the consensus on enterotypes shifted to only two community types: one dominated by *Bacteroides* and/or by other *Firmicutes* and *Bifidobactreria* (enterotype 1) and the other dominated by *Prevotella* (enterotype 2) [12]. The link between enterotypes and long-term diet has been assessed [19,20]. In detail, individuals with diets rich in protein and animal fat were classified as enterotype 1 (*Bacteroides*-rich microbiota) while individuals that follow a more carbohydrates-rich diet are much more likely to belong to the enterotype 2 (*Prevotella*-rich microbiota) [19]. *Bacteroides* is exclusively found in the gut of mammals. It is generally a commensal, mutualistic bacterial genus that confers numerous health benefits to the human host [21], but it can also be the cause of endogenous infections that lead to acute appendicitis, bacteremia, endocarditis, and intra-abdominal abscesses [22]. *Prevotella* spp. is abundant in healthy microbiota; some studied showed the effects of the abundance of some *Prevotella* strains in the gut microbiota: its beneficial role has been seen in cardiovascular diseases and glucose metabolism, while its pathogenic potential has been demonstrated in metabolic syndrome, obesity, IBD, and other inflammatory diseases such as rheumatoid arthritis, asthma, bacterial vaginitis and HIV infection [23]. *Bacteroides* and *Prevotella* are just two examples that show how the interactions of gut bacteria with their host cannot simply be classified as ‘good' or 'bad': they might provide necessary functions in some circumstances, but they can be detrimental to health in other scenarios [24]. Such bacteria are defined as “pathobionts”: symbiotic or commensal bacteria that have the potential to be environmentally induced to cause a dysregulation of the immune system, leading to a disease state [25]. In addition, the gut is the main site where microorganisms interact with the immune system. For this reason, immune-mediated diseases (autoimmune diseases, allergies, cancer, etc...) are influenced by bacterial colonization and by gut microbiota composition [25]. In this setting, healthy microbiota composition is maintained in an equilibrium that is able to influence the immune system. Environmental factors, such as diet, have the potential to alter the microbiota composition in favor of the pathobionts, a condition known as dysbiosis. This imbalance triggers a pro-inflammatory reaction that can lead to

inflammatory diseases and pathogen infections (Figure 2) [25]. A balanced healthy microbiota acts as a barrier in the gut environment by supporting the physiological activity of the organ [26]. This barrier function can be lost during illness, leading to an increased permeability of the gastrointestinal wall. In this scenario, pathobionts and pathogens might increase their virulent expression, resulting in dysbiosis (alteration in microbiota composition) and immunological dysequilibrium that are the building blocks of inflammation [27]. Some bacterial species of the gut microbiota can be considered pathobionts, but they are often referred to as opportunistic pathogens. Despite being harmless to their host with a healthy immune system, these bacteria represent a serious burden as they cause infection in immunocompromised individuals, in particular to nosocomial patients. *Klebsiella pneumoniae* is commonly found in the gut microbiota, and it has been also categorized as a pathobiont [28]. However, *K. pneumoniae* is mostly known as an opportunistic pathogen able to cause severe multi-organs infections involving lungs, urinary tract and surgical wounds. Another important opportunistic pathogen is *Escherichia coli*: the bacterium is often present in the gut of healthy individuals, but some specific strains can cause serious harm. In particular, the bacterium has the potential to become pathogenic in certain circumstances, like in patients affected by IBD [29].

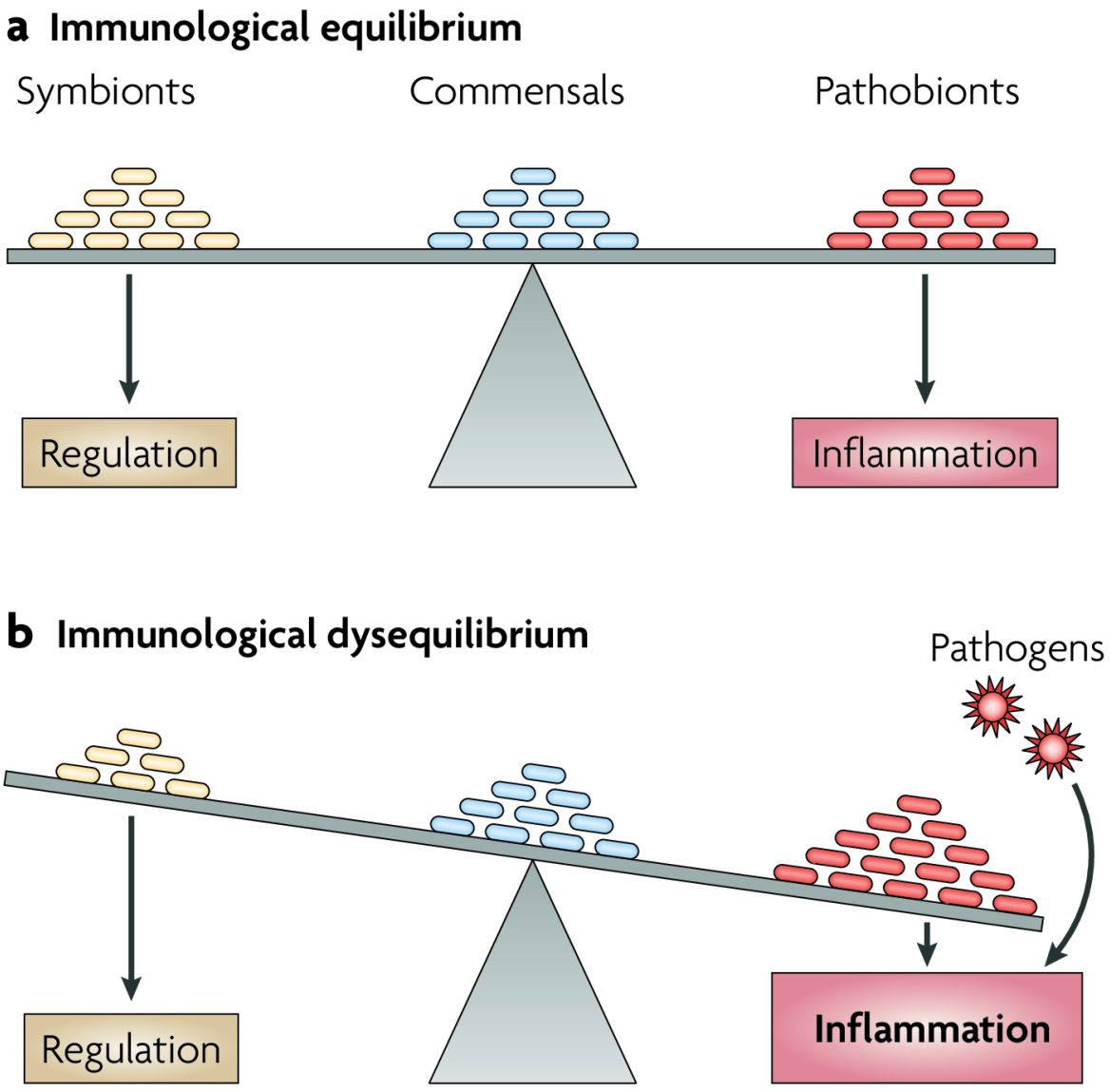


Figure 2: Immunological equilibrium and microbiota dysbiosis [25]
a) the main feature of an healthy microbiota is its balanced composition of symbiotic bacteria (bacteria with beneficial function), commensals (permanent resident of the microbiota without a known beneficial or detrimental function) and pathobionts (permanent resident of the human gut) with a potential to become pathogens . b) the microbiota in dysbiosis loses its equilibrium. A shift toward the pathobionts causes nonspecific inflammation: in this case a pathobiont can become harmful for the host, and the inflammatory state promotes the development of many other immune-related diseases.

Immune System and T-helper cell differentiation

T lymphocytes are major actors of the human immune system, particularly of the adaptive response. They are involved in the immune recognition of foreign pathogens, but they can also induce an immune response against normal tissue in autoimmune disease episodes [30]. The two main subsets of T cells are characterized by the presence of specific cell surface proteins, used as markers: CD8 and CD4. CD8+ T cells are called "cytotoxic" or "killer" T cells because they are able to actively kill autologous cells infected by viruses or other intracellular pathogens, or that have undergone a tumorigenic transformation. Conversely, CD4+ T cells are called "helper" T cells (Th cells): they are the most prolific cytokine producers [30] and their function is fundamental to sustain the response of CD8+ cells and to activate the memory B cell. They are required for almost every adaptive immune response, hence Th cells are considered the most important cells of adaptive immunity [31]. Th cells can be further subdivided in various subpopulations based on the patterns of cytokines they produce: the main ones are called Th1 and Th2, as firstly described by Mosmann and Coffman in 1989 [32]. In particular, the proinflammatory cytokines produced by Th1 cells (especially interferon gamma, IFN-γ) are responsible for killing intracellular pathogens but they also perpetuate autoimmune reactions. On the other hand, Th2 cytokines (like Interleukins, IL, 4 and 5) promote IgE production mainly against parasitic (e.g. helminth) infestations and stimulate an antiinflammatory response that counterbalances the proinflammatory capabilities of Th1 cytokines (e.g. IL-10) [30]. Uncommitted CD4+ cells are differentiated to Th1 or Th2 cells on the basis of the concentration of IFN-γ and IL-12 [33] (Figure 3) and in physiological conditions, there is a precise balance between Th1 and Th2 responses that is suited to (and shaped by) each specific immune challenge [30]. Immune response dominated by Th1 cytokines is more protective against intracellular pathogens and usually leads to their clearance. If that is not the case, persisting Th1 response may cause inflammatory tissue damage. Instead, extracellular parasites trigger a Th2 immune response that is more efficient compared to Th1 response against intestinal nematodes, among many other helminths [34].

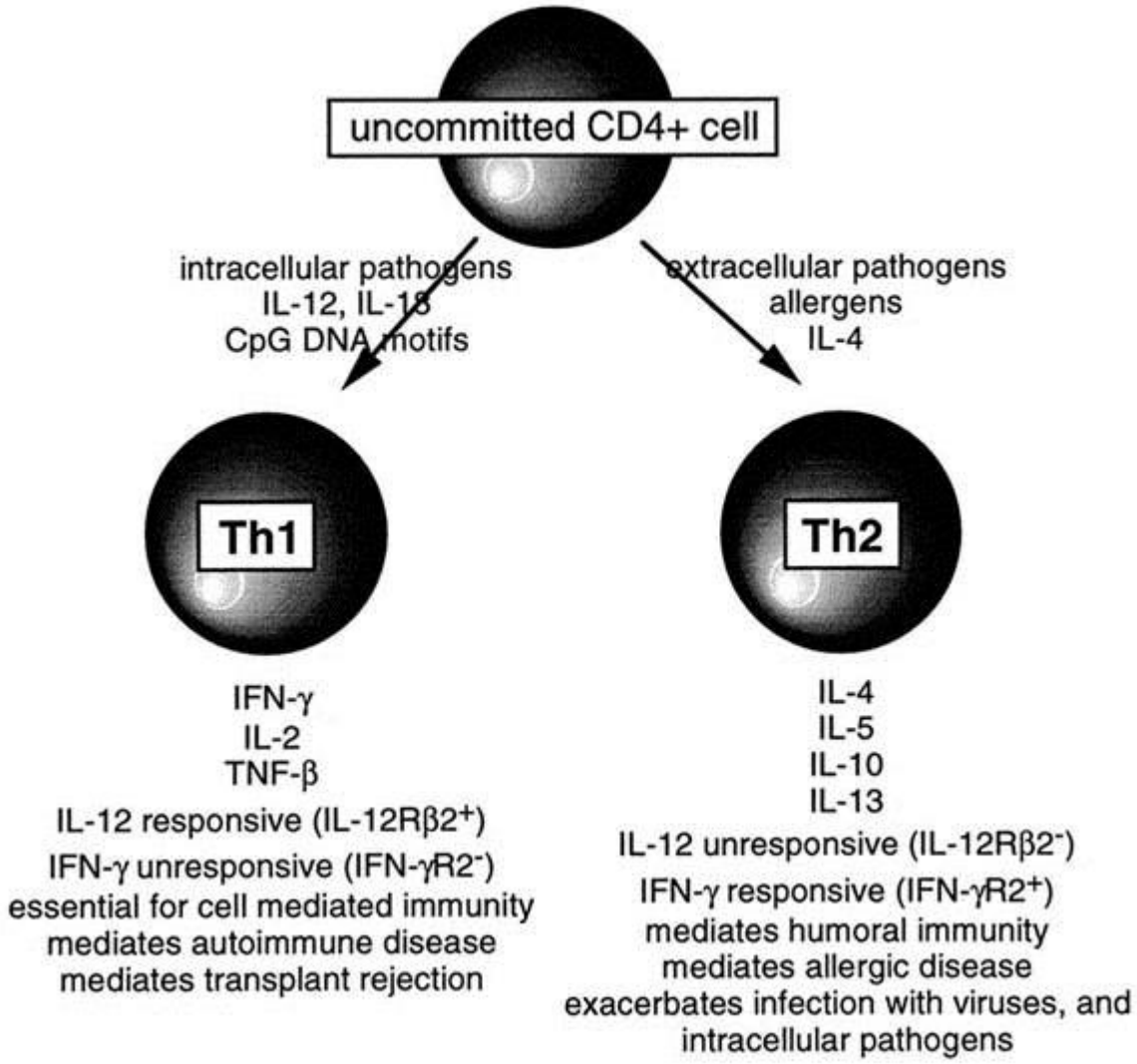


Figure 3: Helper T-Cell (Th) Subgroups [35]
Th1 and Th2 cells are characterized by their cytokine profiles and by their specific response to IL-12 and to IFN-γ. A Correct immune response depends also on the balance between the two subgroups of CD4+ cells.

Inflammatory Bowel Diseases (IBDs) is multifactorial disease characterized by the chronic inflammation of the intestines; clinically, the most common presentations of IBDs are: Crohn's Disease (CD) and Ulcerative Colitis (UC) [36]. In IBD patients a persistent imbalanced Th1-dominant immune response causes damage of the mucosal barrier of the gut, which in turn causes impaired absorption and increases the risk of bacterial pathogen and pathobiont invasion. The onset and progression of the disease

is due to the interplay of genetic predisposition, environmental factors that directly influence the inflammatory pathways, composition of the gut microbiota [37]. Indeed, IBD has been strongly correlated to intestinal dysbiosis and impaired immune homeostasis [38]. In particular, IBD patients show a significant decrease in the levels of SCFA [39]. On the other hand, imbalances towards Th2 differentiation happen in allergic, asthmatic and atopic individuals. [35,36]. Atopic Dermatitis (AD) is a chronic, relapsing disease of the skin. It usually emerges in the first year of life and it affects up to 20% of children, most of which stop to show symptoms in a few years [40]. The immunopathogenesis of AD is characterized by a strong immune reaction sustained by the disruption of the skin barrier and by the resulting inflammation reaction [41]. The skin disruption has been demonstrated, among other factors, to be due to the pathogenic role of Th2 cytokines [42]. Altered gut microbiota composition has been observed in AD patients, and some evidence shows the effect of prebiotic and probiotic supplementation used to alleviate the onset of the disease [43].

In a broader view, the coevolution of gut microbiota with his host built a strong relationship that is tightly linked to the human immune system. Microbiota actively takes part in the balance of the immune system between activation and tolerance, and the absence of this balance contributes to disease development [44]. With this in mind, diet and supplementation can be used to affect host immunity both directly and through the microbiota mediation [45].

Next Generation Sequencing

The first Next Generation Sequencing (NGS) technology was introduced in the market in 2005 [46]. The throughput of DNA sequencing and improvements in processing power allowed the sequencing capability to increase exponentially. As a result, the price for sequencing dropped significantly making high throughput sequencing available to individual laboratories [47]. Roche 454 (454; Branford, CT, USA; now Roche, Basel) was the first NGS platform introduced to the market [46]. It requires a library preparation step, where DNA molecules are fragmented and attached to microbeads. The beads individually fall into the micro wells of a matrix allowing the simultaneous pyrosequencing of thousands of fragments: nucleotide (A, T, G or C) specific colored light is released when a new nucleotide binds to the fragment we are sequencing; this light is detected simultaneously for each bead. With this technology it was possible to get quite long reads (~700 bp) at a reasonable price (10$/Mb) that is some order of magnitude less than Sanger sequencing [48].

Nevertheless, this technology was slowly replaced by Ion Torrent instruments (Thermo Fisher Scientific, Waltham, MA, USA). The sequencing approach in this instrument is similar to the 454 system but with a main difference: H^+ ions are released during DNA sequencing reactions. This causes a change in pH of the solution which, in turns, is translated to an electric signal detected by the instrument [49]. Ion Torrent sequencing technology is cheaper than 454 (~$1.10/MB) producing ~400 bp long reads [50] and applied a policy that focuses on ease of use: there are sequencing kit and automatic analysis softwares that makes the analysis more standard and easy to be performed.

Nonetheless, the current market leader in NGS is Illumina (San Diego, CA, USA), the first sequencer using their technology was introduced in 2006 (Solexa Genome Analyser). Illumina sequencers rely on their Sequencing By Synthesis technology. With this technology the DNA fragments are attached to a glass flow-cell and clonally amplified to obtain clusters of identical fragments that will amplify the signal detectable by the instrument. Then, a modified DNA polymerase starts to add modified fluorescent nucleotides to synthesize the complementary strand of the DNA fragments. Every addition of a nucleotide causes the emission of a light beam with a color that is

specific to each nucleotide (A, C, T, G). A high sensor camera detects the color of the emitted light for each fragment in each position. This process is repeated a number of times that determine the final read length (max 300bp for Illumina MiSeq v3) (Figure 4). The advantage of Illumina sequencing systems is the very high yield per run, which greatly decreases the cost of sequencing (~$0.10 MB) [51,52]. Even if it produces shorter reads, their higher quality ensures better performance than Ion Torrent sequencers. Moreover, Illumina sequencing technology can be highly parallelized. For instance, HiSeq 4000 platforms are able to have a maximum yield of 360GB, sequencing 1.2 billion reads per run.

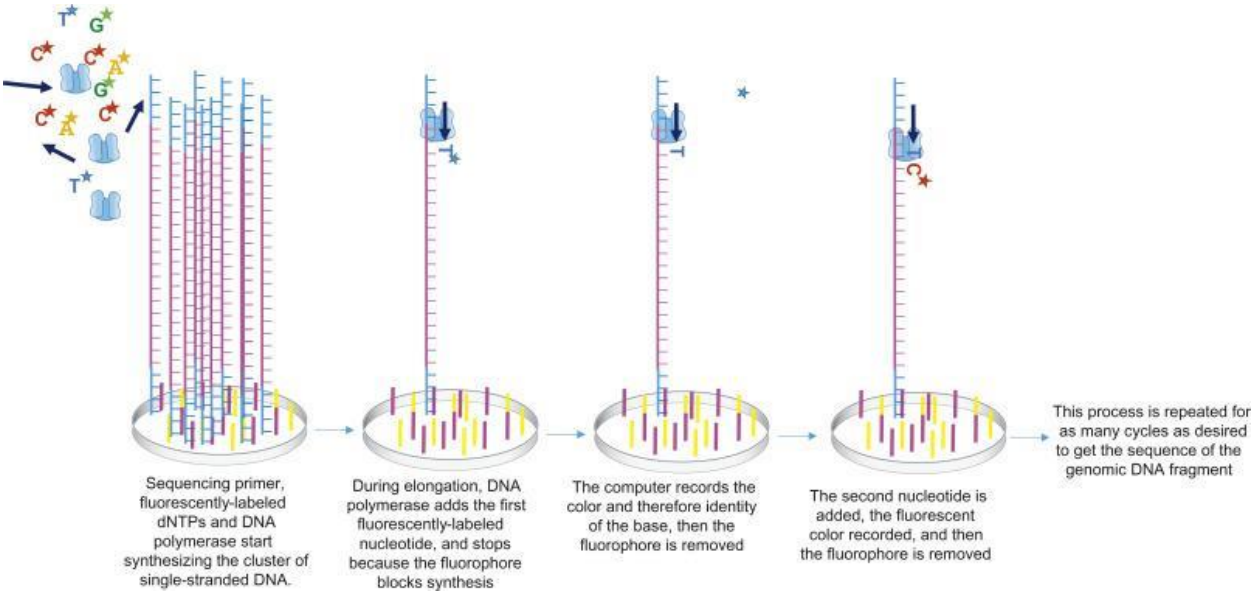


Figure 4: Illumina Sequencing by synthesis technology [53]

Fragments of DNA are attached to the adapter that are physically bound to the flow-cell. Each fragment is clonally amplified to form clusters then the flow-cell is supplied with sequencing primers, fluorescently-labeled dNTPs, and DNA polymerase. Through temperature adjustments, primers anneal to the DNA in the clusters so that DNA polymerase can start its synthesis (the picture depicts only a DNA strand per cluster, but the process occurs in every DNA strand simultaneously). After the addition of the first nucleotide, the synthesis is temporarily blocked to allow the computer to record a high resolution image and detect the type of fluorescence light emitted by each cluster. Then the fluorophore is washed away and the DNA polymerase can add another complementary nucleotide. The cycle is repeated to synthesize the entire fragments; the result is the read of the cluster.

Third Generation Sequencing

The cutting edge technology in the field of DNA sequencing is the so-called "third generation sequencing" technology. With this term we mainly refer to sequencing technology that produces longer reads (10K - 1M base pair in length) compared to Next Generation Sequencing. Even if the reads of Next Generation Sequencing are exceptionally reliable, their short length might prevent the assembler software to resolve genomic regions where tandem repeats or any other larger structural variations are present. This is often a cause for misassemblies and gaps in the final genome assembly that would be heavily fragmented in many contigs [54]. The first Third Generation Sequencing platform available in the market was launched in 2011 by Pacific Biosciences (PacBio, Menlo Park, CA, United States) and it is based on the Single-Molecule Real-Time" (SMRT) sequencing. Indeed, PacBio sequencers capture sequence information during the replication process of the single molecule of target DNA. The DNA molecule flows inside a sequencing unit called a Zero-Mode Waveguide (ZMW). A modified DNA polymerase is immobilized at the bottom of each ZMW well. A light beam is emitted anytime a fluorescent nucleotide is incorporated into the template DNA, each nucleotide (A, T, G or C) emits a specific wavelength. The light pulses are captured by the instrument for each ZMW. This system allows to obtain much longer reads: the maximum length of the reads depends mainly on the life-time of the polymerase [55]. One drawback of PacBio sequencing is the lower accuracy (error rate of 10% - 15% [55]). Nevertheless, the errors are distributed randomly so the error can be reduced by sequencing multiple times the same DNA molecule: hairpin adaptors are inserted during library preparation forming a loop with the target DNA. In this way the polymerase will sequence the target DNA multiple times reducing its error rate at the expense of the maximum read length that is greatly reduced (Figure 5). In 2014 Oxford Nanopore Technologies (Oxford, United Kingdom) introduced a new instrument that adopted a new sequencing method called nanopore sequencing [56]. This sequencing method uses electrophoresis to transport DNA through a membrane covered in porin proteins that have a diameter of 1 nanometer. The negatively-charged DNA molecules are pushed through the porin proteins thanks to the current flow.

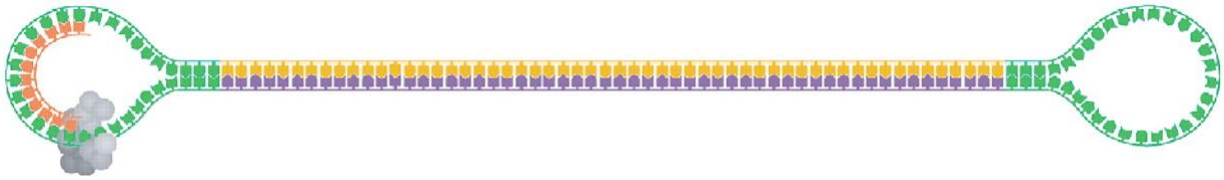


Figure 5: SMRTbell technology [57]
The insertion of two hairpin loops (green) flanking the template DNA insert (purple) lead to the formation of a SMRTbell template. The polymerase molecule (bottom left) can start to repeatedly sequence the belt thanks to specific primers (orange) that binds to the hairpin sequence.

The porin proteins slow down the flow of the DNA molecule through the pore: this DNA flow fluctuation causes the temporary interruption of the current flux that can be detected by the instrument [58]. The amplitude and duration of this current fluctuation is determined by the chemical and physical properties of the nucleotide that is crossing the membrane (Figure 6). Statistical analysis on the current amplitude over time allows to determine the sequence of the DNA [59]. Nanopore sequencing usually provides the longest reads. With the development of molecular and bioinformatic tools it is possible to sequence "ultra-long" reads, the longest sequenced read reported is 2,272,580 bases in length [60]. A drawback of this method is its high error rate (about 15% [61]) as it does not allow the sequencing of the same DNA molecule multiple times. New methods are under development to increase the accuracy of the Nanopore sequencing.
Nevertheless, in most cases Third Generation Sequencing is still coupled with Next Generation Sequencing to get complete genomes and to get accurate single variant calling. This procedure consists of sequencing the same DNA sample with short reads (e.g Illumina) and long reads (e.g PacBio). In this way it is possible to have the high reliability of Next Generation Sequencing (thanks to its high throughput) and to obtain non fragmented long genomic stretches that act as a backbone during genome assembly. This can eventually result in the assembly of complete genomes.

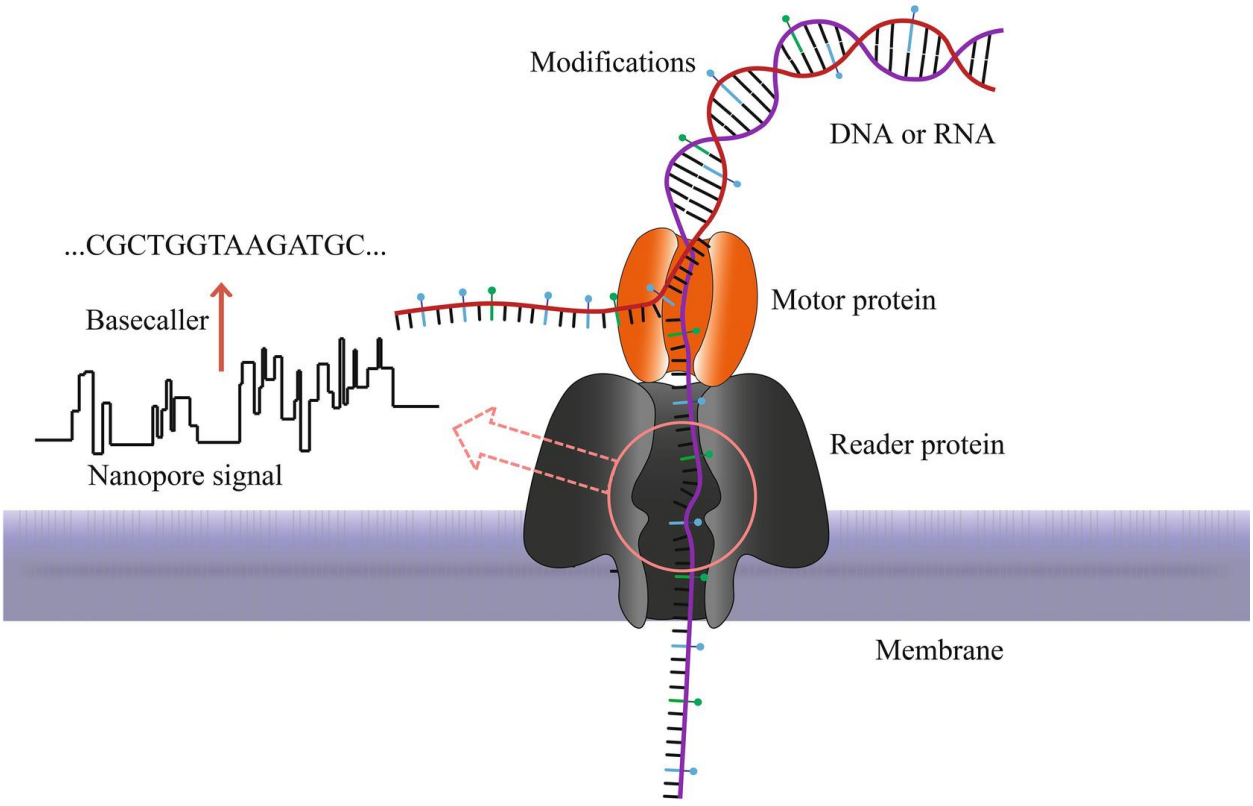


Figure 6: Schematic structure of a Nanopore sequencer [62]

The DNA molecule is guided into the reader protein by the motor proteins. The porin protein across the membrane detects current fluctuation due to the different nucleotides and their modifications. This signal is converted in basecalles by the instrument.

Metagenomics

Metagenomics is defined as the direct (i.e., culture-independent) characterization of the genetic material present in a sample. The term derives from "metagenome" which is the "collective genome" of a sample or of an environment [63]. Metagenomics aims to analyse the genomes of, at least, the most abundant members of a bacterial (or microbial) community, without the need to isolate or cultivate them [64]. Thus, it is very useful in many research fields, considering that the large majority of the microbial species cannot be cultivated in laboratory conditions. These unknown microbes are often referred to as "microbial dark matter" [65]. The first metagenomic approaches were based on the cloning of large pieces of microbial DNA into easy-to-cultivate *Escherichia coli* cells that were screened for biological activity and for the production of novel natural products [63]. Two main revolutions in the field are represented by the development of DNA sequencing by Sanger and colleagues [66] and by the idea of using genes coding for ribosomal RNAs as molecular markers for life classification, as proposed by Woese and Fox [67]. But only some decades later these technologies and ideas were successfully applied to the study of bacterial communities (Figure 7), mostly thanks to major advances in molecular biology, like the introduction of DNA amplification through PCR and, subsequently, to the use of computers for data analysis [68]. In recent years, bacterial communities were studied using amplicon metagenomics techniques based on the gene coding for the 16S ribosomal RNA (16S rRNA), i.e., by amplifying and NGS-sequencing the whole 16S ribosomal RNA genes or some of its nine hypervariable regions to assess microbial diversity [69]. The gene is characterised by nine variable regions interlocked by conserved ones (Figure 8). The composition of this gene makes it exceptionally suitable to design PCR primers, in conserved regions, to amplify the hypervariable ones. Stable and universal PCR primers can thus be designed in the conserved regions of the gene to amplify the variable parts, whose sequence allows to discriminate bacterial genera and, sometimes, species. Next Generation Sequencing technologies have been largely applied to perform amplicon metagenomic analysis. In detail, one or more variable regions of the 16S gene are amplified by PCR.

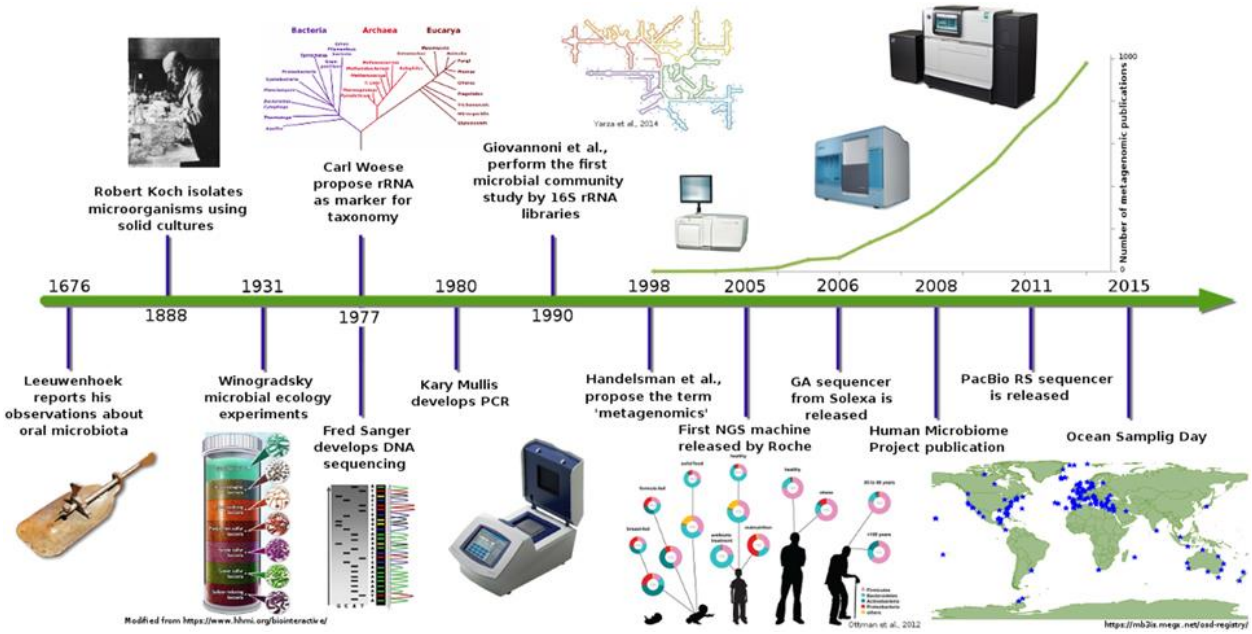


Figure 7: the evolution of metagenomics studies [68]

This timeline reports the major milestones occurred in the development of metagenomic studies.

Given the universal nature of the primers, all the bacteria present in the sample should be ideally amplified. The resulting fragments are used to prepare libraries for the sequencing step. The analytical procedure to analyse the raw reads of amplicon metagenomics is not trivial. Therefore, several bioinformatics pipelines have been developed. The most common pipelines used to analyze amplicon metagenomics data based on the 16S gene are: Mothur [70], QIIME2 [71], Bioconductor [72] and USEARCH [73]. To have great benefits in terms of performance, these pipelines cluster highly similar sequences (typically 97%) into Operational Taxonomic Units (OTUs) that were proven to be ecologically homogeneous on a broad ecological scale [74]. A representative sequence is then selected for each OTU and is used for the taxonomic annotation with one of the available databases collecting the 16s rRNA gene sequences of millions of bacterial species; some examples are: Silva [75], Greengenes [76] and the Ribosomal Database Project [77]. The number of OTUs assigned to a species is related to the frequency of that species in the sample.

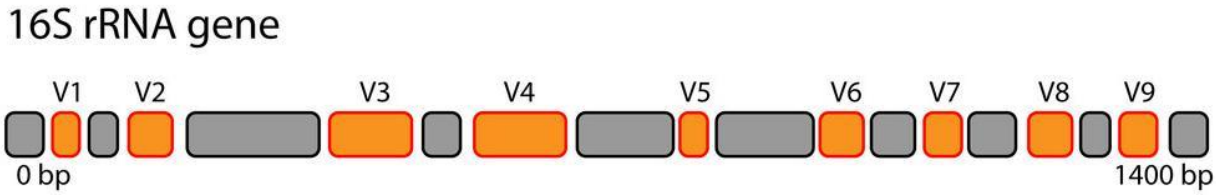


Figure 8: Graphical representation of the regions of a 16S rRNA gene The nine orange hypervariable regions are interrupted by the conserved grey regions along the 16S ribosomal RNA gene. This feature is due to the peculiar secondary structure of the RNA molecule that must be conserved to function properly. Some parts of the molecule have only a structural function and as such they are less affected by selective pressure. In this way, any nucleotide change in these regions would be passed to the next generation. This characteristic makes these regions particularly variable among bacterial species.

Mathematical analyses performed on the species frequencies want to answer questions regarding population diversity of the samples. In particular, alpha diversity, or within-sample diversity, refers to the number of species present in a sample (species richness) and/or the relative abundances of these species (i.e. species evenness). Many indexes are used to measure alpha diversity: among them Chao 1 index, Shannon's index and Simpson's index are the most used. In general, the reduction of diversity in gut microbiota has been associated with a large number of diseases. Nevertheless, it is difficult to assess if the decrease in diversity of gut microbiota is a consequence rather than the cause of the pathological condition [78]. Conversely beta diversity, or between-sample diversity, compares the absolute or relative species abundance among different samples. It can be measured by simply counting the overlapping species, or by calculating the Bray-Curtis dissimilarity index that takes into account the abundance of each species (Figure 9) [79,80].

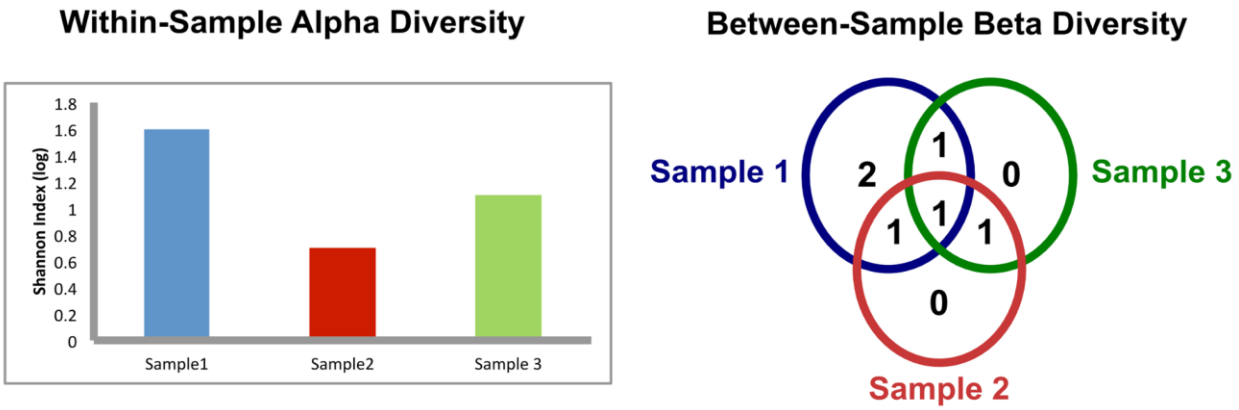


Figure 9: Representation of the alpha and beta diversity of three samples (adapted from Morgan and Huttenhower 2012[80]**)**

Alpha diversity (i.e the species richness and evenness of the three samples) is evaluated by the Shannon's index (left panel). The presence/absence of each species in the sample is an evaluation of the beta diversity (right panel).

Molecular typing of bacteria

Gut microbiota is a complex and dynamic community composed of a great variety of bacterial species. Among them, some are also known to be sources of food contamination: the main example is *Escherichia coli*. One of the main habitats of this bacterium is the human gut [81]; however, it is also one of the most common pathogens that causes foodborne outbreaks (as reported by the Centers for Disease Control and Prevention [82]). There are some specific *E. coli* strains that are a hazard for human health and that can contaminate food. The transmission of this pathogen usually occurs by consumption of contaminated foods, such as raw or undercooked ground meat products, raw milk, and contaminated raw vegetables and sprouts [83]. One of the biggest *E. coli* outbreaks worldwide occurred between May and July 2011. The outbreak was caused by a point source: contaminated sprouts from a horticultural farm in Germany [84]. The total cases of bloody diarrhea due to the bacteria was almost 3000 including 53 deaths in Germany, less cases were reported in France and other European countries but also in Canada, and in the United States [84]. Another opportunistic pathogen that can be associated with foodborne outbreaks is *Klebsiella pneumoniae.* Food can be a vector for its transmission [85], even though this bacteria is most commonly known to cause bacteremia, pneumonia, and urinary tract infection in nosocomial settings [86,87]. *K. pneumoniae* has been isolated from raw meat, raw vegetables, fruit juice, and ready-to-eat food [88]. A nine-month *K. pneumoniae* outbreak occurred in a Spanish hospital, involving 156 patients. The peculiarity of this outbreak was that 35% of the samples collected from surfaces of the hospital kitchen were colonized and that 14% of food handlers were carriers of the *K. pneumoniae* strain that sustained the outbreak. Thanks to this epidemiological event, the authors reported the first dissemination of an epidemic clone of *K. pneumoniae* through the food chain in a hospital [85]. In these two examples of foodborne bacterial outbreaks, the reconstruction of the epidemiological events was carried out with molecular typing methods able to identify very specific genomic features of the strain that was spreading. Molecular typing, in addition to strains metadata, helps to reconstruct the transmission path that took place during the outbreak. Many molecular tools have been

developed to investigate and pinpoint the source of a foodborne outbreak or any other epidemiological event. Such molecular typing tools often need to discriminate strains below the species level (like identifying specific antibiotic resistances, virulence factors, plasmids presence/absence..). Thus, most of the sub-typing bacterial methods interrogate DNA to obtain the adequate amount of information. Among the most used molecular typing techniques we have Pulsed-Field Gel Electrophoresis (PFGE), MultiLocus Sequence Typing (MLST) and Whole Genome Sequencing (WGS). PFGE consists of an agarose gel electrophoresis experiment that allows the separation of large pieces of genomic DNA after the digestion with a restriction endonuclease [89] (Figure 10). The peculiarity of the technique when compared to standard gel electrophoresis is the use of electrical current flows that periodically change direction. This change in the current flow allows larger fragments of DNA to disentangle and be set apart in the gel matrix. The PFGE banding patterns of genetically related strains will be highly comparable (Figure 10). PFGE has been considered the 'gold standard' typing for foodborne pathogen outbreak investigations and other epidemiological studies [90] and even if many laboratories are transitioning to WGS of the strains, PFGE will probably remain a standard typing procedure for years to come [91]. Nonetheless, this method requires 4-6 days and it is quite laborious and expensive (~70€ per bacterial sample). Unfortunately, MLST can lacks the discrimination power to discriminate closely related bacterial strains because of the use of housekeeping genes that are inherently genetically stable [92]. In addition, the sequencing of seven fragments per strain increases, in part, the total cost (~50€ per sample) and the time required to get the results (~3 days).

Thanks to the availability of next- and third-generation sequencing technology, it is now possible to perform the sequencing of the entire genome of bacterial strains (Whole Genome Sequencing, WGS) in a few days. WGS makes it possible to perform bacterial subspecies typing exploiting the entire genetic variability, obtaining the best discriminatory power.

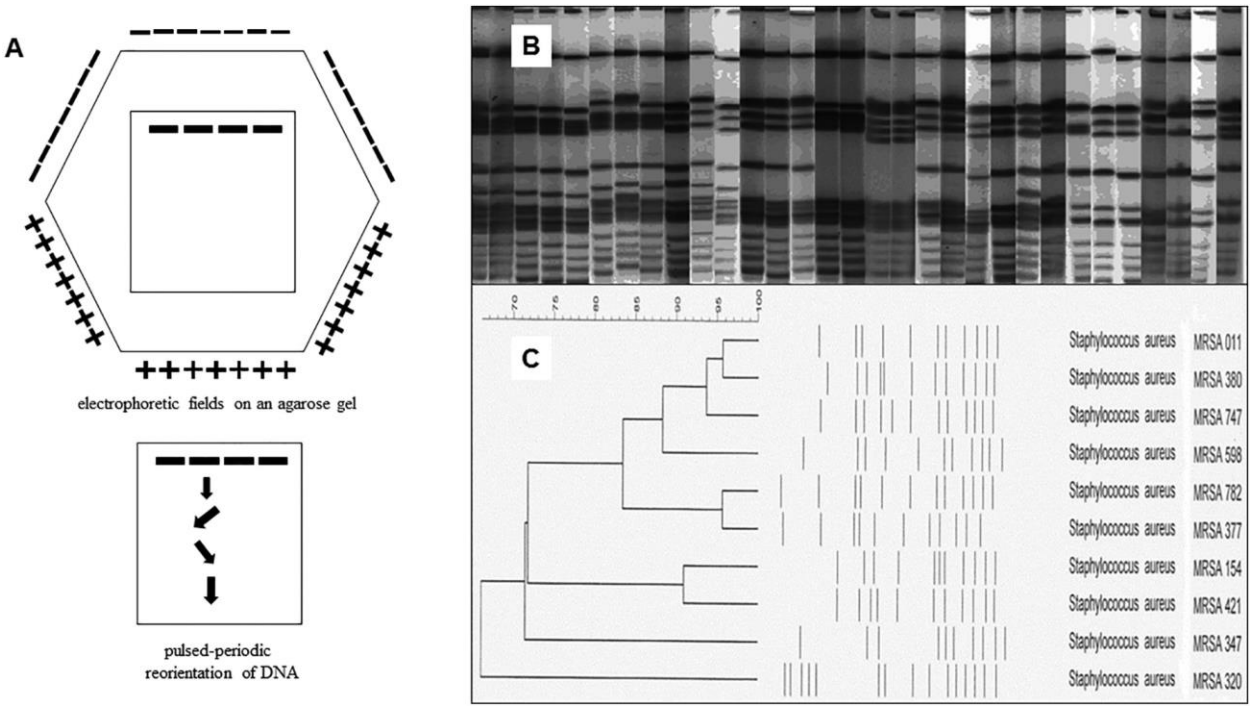


Figure 10: Schematic representation of PFGE and its results [91].

A) The current flow switches periodically its direction, this allows the larger DNA molecules to run more easily in the agarose gel. B) Example of the typical banding pattern of PFGE. C) Hierarchical clustering analysis of the strains based on the banding pattern. Stains with less than three band differences are considered closely-related while strains with more than seven differences are considered unrelated.

This is why WGS will likely become the new "gold standard" for surveillance studies, including bacterial typing, phylogenetic reconstructions and outbreak investigations [94]. The great advantage of WGS is the additional information that we can obtain from the genomes of the pathogens (virulence factors, antibiotic resistance genes, plasmids). This information can help physicians to choose the right therapy for infected patients, but it also helps epidemiologists to better characterize the epidemiological event. The problem in the widespread implementation of WGS in bacterial typing laboratories are: i) the very high cost of the instruments, ii) the lengthy and laborious library preparation steps, iii) the bioinformatic analysis obtain the genome assemblies, iv) data handling, v) the time required to get the final results (around 5 days), and vi) the cost per sample (~100€ per bacterial genome).

Along with these widely used methods, High Resolution Melting (HRM) has been proposed as a suitable fast and inexpensive bacterial typing method [95]. HRM is a rapid single-step real-time PCR based method that can be performed in most generic real time PCR Platforms. The experiment starts after the DNA extraction and consists in the real-time PCR amplification followed by a temperature ramping able to incrementally melt the DNA double strand. While the DNA melts, the intercalating dye used in the real time PCR experiment leaves the DNA molecules causing a drop in fluorescence that is detected by the instrument that draws the melting curve. The inflection point of the melting curve is the melting temperature (Figure 11). This temperature characterizes the DNA fragment and it is strictly related to the nucleotide composition of the amplicon, in particular to its guanine and cytosine content [95]. This feature allows the discrimination of DNA fragments on the basis of the sequence avoiding the lengthy and expensive sequencing step. Despite having a lower discrimination power than WGS, HRM typing is very fast (~26 hours), can be performed in most generic real-time PCR platforms with a cost of ~5€ per sample making it very useful for large microbiological screenings.

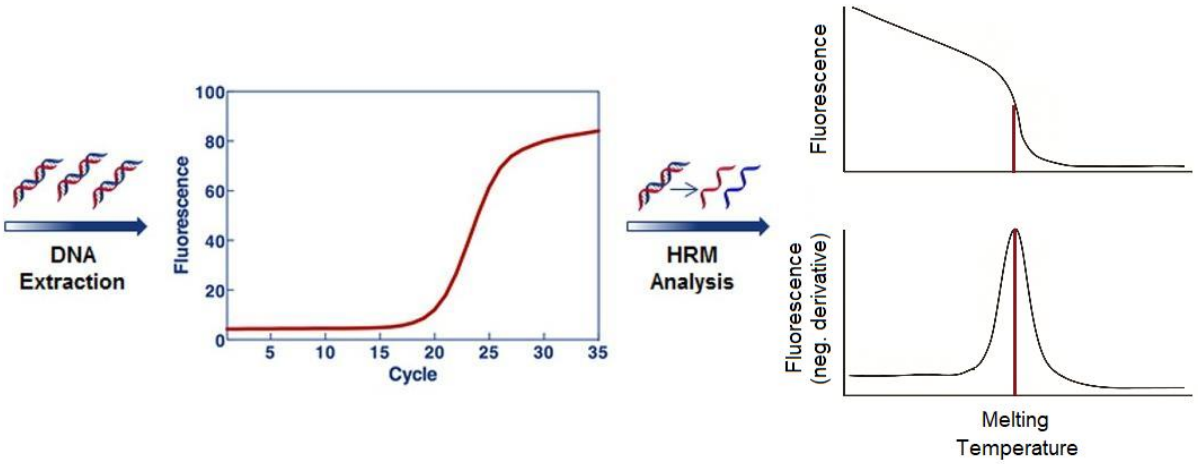


Figure 11: Graphical representation of a High Resolution Melting Experiment (adapted from Andini et al. 2017 [96])

The fragment of genomic DNA is amplified through real-time PCR, then the amplicons are subjected to a gradual increase in temperature while the instrument detects the amount of fluorescence at fixed intervals. In this way the instrument detects the gradual drop in fluorescence caused by the release of the intercalating dye used in the real-time PCR experiment (it is released when the double stranded DNA melts). The negative derivative of the resulting curve can be calculated and the peak in this new curve indicates the melting temperature of the amplified DNA fragment.

Related articles

Section A: Influence of nutrition and gut microbiota on human health

In this section, I investigated the effect of gut microbiota and food supplementation on atopic dermatitis, inflammatory bowel diseases and in pediatric age. In the first work (Appendix 1) we assessed the effects of a postbiotic supplementation on atopic dermatitis severity in children. The second work (Appendix 2) is a review regarding the connections between the hygiene hypothesis and inflammatory bowel disease, with a focus on the parasitic and fungal components of the gut microbiota. The third paper (Appendix 3) is a review that aims to help pediatricians to set up, analyze and interpret gut microbiota studies.

Appendix 1
D'Auria E, Panelli S, Lunardon L, Pajoro M, Paradiso L, Beretta S, Loretelli C, Tosi D, Perini M, Bedogni G, Abdelsalam A, Fiorina P, Bandi C, Zuccotti GV. **Rice flour fermented with *Lactobacillus paracasei* CBA L74 in the treatment of atopic dermatitis in infants: A randomized, double- blind, placebo-controlled trial.** Pharmacol Res. 2021 Jan;163:105284. doi: 10.1016/j.phrs.2020.105284 [97]

Appendix 2
Panelli S, Epis S, Cococcioni L, Perini M, Paroni M, Bandi C, Drago L, Zuccotti GV. **Inflammatory bowel diseases, the hygiene hypothesis and the other side of the microbiota: Parasites and fungi.** Pharmacol Res. 2020 Sep;159:104962. doi:10.1016/j.phrs.2020.104962 [98]

Appendix 3
Drago L, Panelli S, Bandi C, Zuccotti G, Perini M, D'Auria E. **What Pediatricians Should Know Before Studying Gut Microbiota**. J Clin Med. 2019 Aug 12;8(8):1206. doi: 10.3390/jcm8081206 [99]

Section B: Genetic characterization of gut bacterial opportunistic pathogens

In the second section I used Next Generation Sequencing (NGS) technology to characterize two clinical cases caused by two opportunistic bacterial pathogens. The first (Appendix 4), regards a particular *Staphylococcus aureus* strain isolated from an Italian man who likely become infected in Egypt. The second paper (Appendix 5) describes a reconstruction of a large nosocomial outbreak sustained by an *Escherichia coli* strain and that involved 106 patients in 25 wards of a hospital in Italy.

Appendix 4
Rimoldi SG, Comandatore F, Longhi E, Romeri F, Piazza A, Pagani C, Tamoni A, Longobardi C, Negri C, Bestetti G, Gervasoni C, Perini M, Antinori S, Bandi C, Gismondo MR. **Genomic Characterization of an ST1153 PVL-producing Methicillin Resistant *Staphylococcus aureus* Clinical Isolate in Italy**. New Microbiol. 2019 Apr;42(2):129-131 [100]

Appendix 5
Piazza A, Principe L, Comandatore F, Perini M, Meroni E, Mattioni Marchetti V, Migliavacca R, Luzzaro F. **Whole-Genome Sequencing Investigation of a Large Nosocomial Outbreak Caused by ST131 H30Rx KPC-Producing *Escherichia coli* in Italy**. Antibiotics (Basel). 2021 Jun 15;10(6):718. doi: 10.3390/antibiotics10060718 [101]

Section C: Development of a sequencing-free bacterial subspecies typing approach

This last section shows the development of an innovative, fast and scalable approach to pathogen typing suitable for large screening and real-time microbiological surveillance, that we named Hypervariable-Locus Melting Typing (HLMT). In the first publication (Appendix 6) we described EasyPrimer, a web tool we developed to help PCR primer design in difficult contexts and the use of the tool to develop an HLMT protocol for *Klebsiella pneumoniae* typing. In the second publication (Appendix 7) we tested the repeatability and reproducibility of the protocol. In the third work of this section (Appendix 8) we describe MeltingPlot, the web tool that implements the innovative graph-based algorithm we developed for strain clustering and strain typing on the basis of HRM data. In the last article (Appendix 9) we provided a comprehensive description of the HLMT approach, and we applied the *Klebsiella pneumoniae* protocol in four nosocomial epidemiological investigations, comparing the results with the standard typing methods.

Appendix 6
Perini M, Piazza A, Panelli S, Di Carlo D, Corbella M, Gona F, Vailati F, Marone P, Cirillo DM, Farina C, Zuccotti G, Comandatore F. **EasyPrimer: user- friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on *wzi* gene for fast *Klebsiella pneumoniae* typing.** Sci Rep. 2020 Jan 28;10(1):1307. doi: 10.1038/s41598-020-57742-z [102]

Appendix 7
Pasala AR*, Perini M*, Piazza A, Panelli S, Di Carlo D, Loretelli C, Cafiso A, Inglese S, Gona F, Cirillo DM, Zuccotti GV, Comandatore F. **Repeatability and reproducibility of the *wzi* high resolution melting-based clustering analysis for *Klebsiella pneumoniae* typing.** AMB Express. 2020 Dec 14;10(1):217. doi 10.1186/s13568-020-01164-7 [103]

Appendix 8
Perini M, Batisti Biffignandi G, Di Carlo D, Pasala AR, Piazza A, Panelli S, Zuccotti GV, Comandatore F. **MeltingPlot, a user-friendly online tool for epidemiological investigation using High Resolution Melting data.** BMC Bioinformatics. 2021 Feb 18 ;22(1):76. doi: 10.1186/s12859-021-04020-y [104]

Appendix 9
Perini M, Piazza A, Panelli S, Papaleo S, Alvaro S, Vailati F,Corbella M, Saluzzo F, Gona F, Castelli F, Farina C, Marone P, Cirillo DM, Cavallero A, Zuccotti GV, Comandatore F. **Hypervariable-Locus Melting Typing (HLMT): a novel, fast and inexpensive sequencing-**

free approach to pathogen typing based on High Resolution Melting (HRM) analysis. bioRxiv. 2021 Nov 08. doi: 10.1101/2021.07.01.450706 [105]

* These authors contributed equally

Other Articles

Comandatore F*, Chiodi A*, Gabrieli P, Batisti Biffignandi G, Perini M, Ricagno S, Mascolo E, Petazzoni G, Ramazzotti M, Rimoldi SG, Gismondo MR, Micheli V, Sassera D, Gaiarsa S, Bandi C, Brilli M. **Insurgence and worldwide diffusion of genomic variants in SARS-CoV-2 genomes.** bioRxiv. 28 May 2020 doi: 10.1101/2020.04.30.071027 [106]

Micheli V, Rimoldi SG, Romeri F, Comandatore F, Mancon A, Gigantiello A, Perini M, Mileto D, Pagani C, Lombardi A, Gismondo MR. **Geographical reconstruction of the SARS-CoV-2 outbreak in Lombardy (Italy) during the early phase.** J Med Virol. 2021 Mar;93(3):1752-1757. doi: 10.1002/jmv.26447 [107]

* These authors contributed equally

Conclusions

There is a strong bond and interplay between gut microbiota, nutrition and immune system, and they all contribute in maintaining or restoring the health status. Dysbiosis of the gut microbiota alters the immunological equilibrium, promoting pathogen infections and non-specific systemic inflammation. Healthy gut microbiota is generally characterized by the higher abundance of bacteria producing Short Chain Fatty Acids (SCFA). These metabolites directly help to maintain the immune system in equilibrium and they are found in lower amounts in the gut of patients that show various immune-related disorders. I used Next Generation Sequencing (NGS) technologies to investigate the role of postbiotic supplementation and gut microbiota composition in immune-related disorders and in the pediatric age. Some bacterial components of the gut microbiota have the potential to cause dysregulation of the immune system that leads to disease state, these bacteria are called “pathobionts”. I used NGS-based genomic epidemiology to characterize and reconstruct two clinical epidemiological events sustained by two gut pathobionts known to cause major epidemiological cases, particularly in hospitals. To control the spread of nosocomial pathogens, hospitals implement and perform microbiological surveillance programs. We developed a novel approach to bacterial typing that we called Hypervariable-Locus Melting Typing (HLMT) that is particularly fast and inexpensive because it is based on High Resolution Melting (HRM) analysis. Thanks to the two bioinformatic web tools that we published, we developed a new HLMT protocol for *Klebsiella pneumoniae* and we applied it to reconstruct various epidemiological events. We also showed that HLMT is suitable for fast epidemiological investigation and that HLMT protocols can be implemented even for continuous microbiological surveillance and for large screenings programs.

References

1. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014;157: 121–141. doi:10.1016/j.cell.2014.03.011.

2. Bull MJ, Plummer NT. Part 1: The Human Gut Microbiome in Health and Disease. Integr Med . 2014;13: 17–22. Available: https://www.ncbi.nlm.nih.gov/pubmed/26770121.

3. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and health. BMJ. 2018;361: k2179. doi:10.1136/bmj.k2179.

4. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. Nature. 2017;548: 43–51. doi:10.1038/nature23292.

5. Albenberg LG, Wu GD. Diet and the intestinal microbiome: associations, functions, and implications for health and disease. Gastroenterology. 2014;146: 1564–1572. doi:10.1053/j.gastro.2014.01.058.

6. O'Toole PW, Marchesi JR, Hill C. Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. Nat Microbiol. 2017;2: 17057. doi:10.1038/nmicrobiol.2017.57.

7. Wegh CAM, Geerlings SY, Knol J, Roeselers G, Belzer C. Postbiotics and Their Potential Applications in Early Life Nutrition and Beyond. Int J Mol Sci. 2019;20. doi:10.3390/ijms20194673.

8. Kim S-K, Guevarra RB, Kim Y-T, Kwon J, Kim H, Cho JH, et al. Role of Probiotics in Human Gut Microbiome-Associated Diseases. J Microbiol Biotechnol. 2019;29: 1335–1340. doi:10.4014/jmb.1906.06064.

9. Gibson GR, Hutkins R, Sanders ME, Prescott SL, Reimer RA, Salminen SJ, et al. Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. Nat Rev Gastroenterol Hepatol. 2017;14: 491–502. doi:10.1038/nrgastro.2017.75.

10. Salminen S, Collado MC, Endo A, Hill C, Lebeer S, Quigley EMM, et al. The International Scientific Association of Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of postbiotics. Nat Rev Gastroenterol Hepatol. 2021;18: 649–667. doi:10.1038/s41575-021-00440-6.

11. Aguilar-Toalá JE, Arioli S, Behare P, Belzer C, Berni Canani R, Chatel J-M, et al. Postbiotics - when simplification fails to clarify. Nature reviews. Gastroenterology & hepatology. 2021. pp. 825–826. doi:10.1038/s41575-021-00521-6.

12. Dalile B, Van Oudenhove L, Vervliet B, Verbeke K. The role of short-chain fatty acids in microbiota-gut-brain communication. Nat Rev Gastroenterol Hepatol. 2019;16: 461–478. doi:10.1038/s41575-019-0157-3.

13. Pascale A, Marchesi N, Marelli C, Coppola A, Luzi L, Govoni S, et al. Microbiota and metabolic diseases. Endocrine. 2018;61: 357–371. doi:10.1007/s12020-018-1605-5.

14. Silva YP, Bernardi A, Frozza RL. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. Front Endocrinol . 2020;11: 25.

doi:10.3389/fendo.2020.00025.

15. Markowiak-Kopeć P, Śliżewska K. The Effect of Probiotics on the Production of Short-Chain Fatty Acids by Human Intestinal Microbiome. Nutrients. 2020;12. doi:10.3390/nu12041107.

16. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, et al. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms. 2019;7. doi:10.3390/microorganisms7010014.

17. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473: 174–180. doi:10.1038/nature09944.

18. Knights D, Ward TL, McKinlay CE, Miller H, Gonzalez A, McDonald D, et al. Rethinking "enterotypes." Cell Host Microbe. 2014;16: 433–437. doi:10.1016/j.chom.2014.09.013.

19. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334: 105–108. doi:10.1126/science.1208344.

20. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A. 2010;107: 14691–14696. doi:10.1073/pnas.1005963107.

21. Wexler AG, Goodman AL. An insider's perspective: Bacteroides as a window into the microbiome. Nat Microbiol. 2017;2: 17026. doi:10.1038/nmicrobiol.2017.26.

22. Wanger A, Chavez V, Huang R, Wahed A, Dasgupta A, Actor JK. Microbiology and Molecular Diagnosis in Pathology: A Comprehensive Review for Board Preparation, Certification and Clinical Practice. Elsevier; 2017. Available: https://play.google.com/store/books/details?id=QpLUDQAAQBAJ.

23. Precup G, Vodnar D-C. Gut *Prevotella* as a possible biomarker of diet and its eubiotic versus dysbiotic roles: a comprehensive literature review. British Journal of Nutrition. 2019. pp. 131–140. doi:10.1017/s0007114519000680.

24. Jochum L, Stecher B. Label or Concept - What Is a Pathobiont? Trends Microbiol. 2020;28: 789–792. doi:10.1016/j.tim.2020.04.011.

25. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol. 2009;9: 313–323. doi:10.1038/nri2515.

26. Kamada N, Chen GY, Inohara N, Núñez G. Control of pathogens and pathobionts by the gut microbiota. Nat Immunol. 2013;14: 685–690. doi:10.1038/ni.2608.

27. McClave SA, Lowen CC, Martindale RG. The 2016 ESPEN Arvid Wretlind lecture: The gut in stress. Clin Nutr. 2018;37: 19–36. doi:10.1016/j.clnu.2017.07.015.

28. Chow J, Tang H, Mazmanian SK. Pathobionts of the gastrointestinal microbiota and inflammatory disease. Curr Opin Immunol. 2011;23: 473–480. doi:10.1016/j.coi.2011.07.010.

29. Mirsepasi-Lauridsen HC, Vallance BA, Krogfelt KA, Petersen AM. *Escherichia coli* pathobionts associated with inflammatory bowel disease. Clin Microbiol Rev. 2019;32. doi:10.1128/cmr.00060-18.

30. Berger A. Th1 and Th2 responses: what are they? BMJ. 2000;321: 424. doi:10.1136/bmj.321.7258.424.

31. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Helper T Cells and Lymphocyte Activation. Molecular Biology of the Cell 4th edition. Garland Science; 2002. Available: https://www.ncbi.nlm.nih.gov/books/NBK26827/.

32. Mosmann TR, Coffman RL. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annu Rev Immunol. 1989;7: 145–173. doi:10.1146/annurev.iy.07.040189.001045.

33. Luckheeram RV, Zhou R, Verma AD, Xia B. $CD4^+$T cells: differentiation and functions. Clin Dev Immunol. 2012;2012: 925135. doi:10.1155/2012/925135

34. D'Elios M, Del Prete G. Th1/Th2 balance in human disease. Transplant Proc. 1998;30: 2373–2377. doi:10.1016/s0041-1345(98)00659-9.

35. Umetsu DT, DeKruyff RH. TH1 and TH2 CD4+ cells in human allergic diseases. J Allergy Clin Immunol. 1997;100: 1–6. doi:10.1016/s0091-6749(97)70186-6.

36. Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. Lancet. 2007;369: 1641–1657. doi:10.1016/S0140-6736(07)60751-X.

37. Silva FAR, Rodrigues BL, Ayrizono M de LS, Leal RF. The Immunological Basis of Inflammatory Bowel Disease. Gastroenterol Res Pract. 2016;2016: 2097274. doi:10.1155/2016/2097274.

38. Miyoshi J, Chang EB. The gut microbiota and inflammatory bowel diseases. Transl Res. 2017;179: 38–48. doi:10.1016/j.trsl.2016.06.002.

39. Machiels K, Joossens M, Sabino J, De Preter V, Arijs I, Eeckhaut V, et al. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. Gut. 2014;63: 1275–1283. doi:10.1136/gutjnl-2013-304833.

40. Brandt EB, Sivaprasad U. Th2 Cytokines and Atopic Dermatitis. J Clin Cell Immunol. 2011;2. doi:10.4172/2155-9899.1000110.

41. Yang G, Seok JK, Kang HC, Cho Y-Y, Lee HS, Lee JY. Skin Barrier Abnormalities and Immune Dysfunction in Atopic Dermatitis. International Journal of Molecular Sciences. 2020. p. 2867. doi:10.3390/ijms21082867.

42. Gandhi NA, Bennett BL, Graham NMH, Pirozzi G, Stahl N, Yancopoulos GD. Targeting key proximal drivers of type 2 inflammation in disease. Nat Rev Drug Discov. 2016;15: 35–50. doi:10.1038/nrd4624.

43. van der Aa LB, Heymans HSA, van Aalderen WMC, Sprikkelman AB. Probiotics and prebiotics in atopic dermatitis: review of the theoretical background and clinical evidence. Pediatr Allergy Immunol. 2010;21: e355–67. doi:10.1111/j.1399-3038.2009.00915.x.

44. Lee N, Kim W-U. Microbiota in T-cell homeostasis and inflammatory diseases. Exp Mol Med. 2017;49: e340. doi:10.1038/emm.2017.36.

45. Burr AHP, Bhattacharjee A, Hand TW. Nutritional Modulation of the Microbiome and Immune Response. The Journal of Immunology. 2020;205: 1479–1487. doi:10.4049/jimmunol.2000419.

46. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437: 376–380. doi:10.1038/nature03959.

47. Rothberg JM, Leamon JH. The development and impact of 454 sequencing. Nat Biotechnol. 2008;26: 1117–1124. doi:10.1038/nbt1485.

48. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012: 251364. doi:10.1155/2012/251364.

49. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475: 348–352. doi:10.1038/nature10242.

50. Pereira FL, Soares SC, Dorella FA, Leal CAG, Figueiredo HCP. Evaluating the efficacy of the new Ion PGM Hi-Q Sequencing Kit applied to bacterial genomes. Genomics. 2016;107: 189–198. doi:10.1016/j.ygeno.2016.03.004.

51. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13: 341. doi:10.1186/1471-2164-13-341.

52. Glenn TC. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011;11: 759–769. doi:10.1111/j.1755-0998.2011.03024.x.

53. Clark DP, Pazdernik NJ, McGehee MR. DNA Sequencing. Molecular Biology. Elsevier; 2019. pp. 240–269. doi:10.1016/b978-0-12-813288-3.00008-2.

54. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. Trends Genet. 2018;34: 666–681. doi:10.1016/j.tig.2018.05.008.

55. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015;13: 278–289. doi:10.1016/j.gpb.2015.08.002.

56. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. Nat Methods. 2015;12: 351–356. doi:10.1038/nmeth.3290.

57. Travers KJ, Chin C-S, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010;38: e159–e159. doi:10.1093/nar/gkq543.

58. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. Genomics Proteomics Bioinformatics. 2015;13: 4–16. doi:10.1016/j.gpb.2015.01.009.

59. Venkatesan BM, Dorvel B, Yemenicioglu S, Watkins N, Petrov I, Bashir R. Highly Sensitive, Mechanically Stable Nanopore Sensors for DNA Analysis. Adv Mater. 2009;21: 2771. doi:10.1002/adma.200803786.

60. Payne A, Holmes N, Rakyan V, Loose M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics. 2019;35: 2193–2198. doi:10.1093/bioinformatics/bty841.

61. Jain M, Tyson JR, Loose M, Ip CLC, Eccles DA, O'Grady J, et al. MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. F1000Res. 2017;6: 760. doi:10.12688/f1000research.11354.1.

62. He M, Chi X, Ren J. Applications of Oxford Nanopore Sequencing in *Schizosaccharomyces pombe*. Methods Mol Biol. 2021;2196: 97–116. doi:10.1007/978-1-0716-0868-5_9.

63. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 1998;5: R245–9. doi:10.1016/s1074-5521(98)90108-9.

64. Carabeo-Pérez A, Guerra-Rivera G, Ramos-Leal M, Jiménez-Hernández J. Metagenomic approaches: effective tools for monitoring the structure and functionality of microbiomes in anaerobic digestion systems. Appl Microbiol Biotechnol. 2019;103: 9379–9390. doi:10.1007/s00253-019-10052-5.

65. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. mSystems. 2018;3. doi:10.1128/mSystems.00055-18.

66. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74: 5463–5467. doi:10.1073/pnas.74.12.5463.

67. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A. 1977;74: 5088–5090. doi:10.1073/pnas.74.11.5088.

68. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. Front Genet. 2015;6: 348. doi:10.3389/fgene.2015.00348.

69. Neefs JM, Van de Peer Y, De Rijk P, Chapelle S, De Wachter R. Compilation of small ribosomal subunit RNA structures. Nucleic Acids Res. 1993;21: 3025–3049. doi:10.1093/nar/21.13.3025.

70. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75: 7537–7541. doi:10.1128/AEM.01541-09.

71. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37: 852–857. doi:10.1038/s41587-019-0209-9.

72. Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. F1000Res. 2016;5: 1492. doi:10.12688/f1000research.8986.2.

73. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26: 2460–2461. doi:10.1093/bioinformatics/btq461.

74. Schmidt TSB, Matias Rodrigues JF, von Mering C. Ecological consistency of SSU rRNA-based operational taxonomic units at a global scale. PLoS Comput Biol. 2014;10: e1003594.

doi:10.1371/journal.pcbi.1003594.

75. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41: D590–6. doi:10.1093/nar/gks1219.

76. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72: 5069–5072. doi:10.1128/AEM.03006-05.

77. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73: 5261–5267. doi:10.1128/AEM.00062-07

78. Mosca A, Leclerc M, Hugot JP. Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? Front Microbiol. 2016;7: 455. doi:10.3389/fmicb.2016.00455.

79. Lozupone CA, Knight R. Species divergence and the measurement of microbial diversity. FEMS Microbiol Rev. 2008;32: 557–578. doi:10.1111/j.1574-6976.2008.00111.x.

80. Morgan XC, Huttenhower C. Chapter 12: Human microbiome analysis. PLoS Comput Biol. 2012;8: e1002808. doi:10.1371/journal.pcbi.1002808.

81. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol. 2010;8: 207–217. doi:10.1038/nrmicro2298.

82. Centers for Disease Control and Prevention. Foodborne Outbreaks. In: www.cdc.gov [Internet]. 8 Dec 2021 [cited 17 Dec 2021]. Available: https://www.cdc.gov/foodsafety/outbreaks/index.html.

83. World Health Organization. *E. coli*. In: www.who.int [Internet]. [cited 19 Dec 2021]. Available: https://www.who.int/news-room/fact-sheets/detail/e-coli.

84. Sarno E, Pezzutto D, Rossi M, Liebana E, Rizzi V. A Review of Significant European Foodborne Outbreaks in the Last Decade. J Food Prot. 2021;84: 2059–2070. doi:10.4315/JFP-21-096.

85. Calbo E, Freixas N, Xercavins M, Riera M, Nicolás C, Monistrol O, et al. Foodborne nosocomial outbreak of SHV1 and CTX-M-15-producing *Klebsiella pneumoniae*: epidemiology and control. Clin Infect Dis. 2011;52: 743–749. doi:10.1093/cid/ciq238.

86. Karlowsky JA, Jones ME, Thornsberry C, Friedland IR, Sahm DF. Trends in antimicrobial susceptibilities among *Enterobacteriaceae* isolated from hospitalized patients in the United States from 1998 to 2001. Antimicrob Agents Chemother. 2003;47: 1672–1680. doi:10.1128/AAC.47.5.1672-1680.2003.

87. Siu LK, Kristopher Siu L, Fung C-P, Chang F-Y, Lee N, Yeh K-M, et al. Molecular Typing and Virulence Analysis of Serotype K1 *Klebsiella pneumoniae* Strains Isolated from Liver Abscess Patients and Stool Samples from Noninfectious Subjects in Hong Kong, Singapore, and Taiwan. Journal of Clinical Microbiology. 2011. pp. 3761–3765. doi:10.1128/jcm.00977-11.

88. Hartantyo SHP, Chau ML, Koh TH, Yap M, Yi T, Cao DYH, et al. Foodborne *Klebsiella*

pneumoniae: Virulence Potential, Antibiotic Resistance, and Risks to Food Safety. J Food Prot. 2020;83: 1096–1103. doi:10.4315/JFP-19-520.

89. Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell. 1984;37: 67–75. doi:10.1016/0092-8674(84)90301-5.

90. Alonso R, Martín A, Peláez T, Marín M, Rodríguez-Creixéms M, Bouza E. An improved protocol for pulsed-field gel electrophoresis typing of *Clostridium difficile*. J Med Microbiol. 2005;54: 155–157. doi:10.1099/jmm.0.45808-0.

91. Neoh H-M, Tan X-E, Sapri HF, Tan TL. Pulsed-field gel electrophoresis (PFGE): A review of the "gold standard" for bacteria typing and current alternatives. Infect Genet Evol. 2019;74: 103935. doi:10.1016/j.meegid.2019.103935.

92. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 1998;95: 3140–3145. doi:10.1073/pnas.95.6.3140.

93. Adzitey F, Huda N, Ali GRR. Molecular techniques for detecting and typing of bacteria, advantages and application to foodborne pathogens isolated from ducks. 3 Biotech. 2013;3: 97–107. doi:10.1007/s13205-012-0074-4.

94. Uelze L, Grützke J, Borowiak M, Hammerl JA, Juraschek K, Deneke C, et al. Typing methods based on whole genome sequencing data. One Health Outlook. 2020;2: 3. doi:10.1186/s42522-020-0010-1.

95. Tamburro M, Ripabelli G. High Resolution Melting as a rapid, reliable, accurate and cost-effective emerging tool for genotyping pathogenic bacteria and enhancing molecular epidemiological surveillance: a comprehensive review of the literature. Ann Ig. 2017;29: 293–316. doi:10.7416/ai.2017.2153.

96. Andini N, Wang B, Athamanolap P, Hardick J, Masek BJ, Thair S, et al. Microbial Typing by Machine Learned DNA Melt Signatures. Sci Rep. 2017;7: 42097. doi:10.1038/srep42097.

97. D'Auria E, Panelli S, Lunardon L, Pajoro M, Paradiso L, Beretta S, et al. Rice flour fermented with *Lactobacillus paracasei* CBA L74 in the treatment of atopic dermatitis in infants: A randomized, double- blind, placebo- controlled trial. Pharmacol Res. 2021;163: 105284. doi:10.1016/j.phrs.2020.105284.

98. Panelli S, Epis S, Cococcioni L, Perini M, Paroni M, Bandi C, et al. Inflammatory bowel diseases, the hygiene hypothesis and the other side of the microbiota: Parasites and fungi. Pharmacol Res. 2020;159: 104962. doi:10.1016/j.phrs.2020.104962.

99. Drago L, Panelli S, Bandi C, Zuccotti G, Perini M, D'Auria E. What Pediatricians Should Know Before Studying Gut Microbiota. J Clin Med Res. 2019;8. doi:10.3390/jcm8081206.

100. Rimoldi SG, Comandatore F, Longhi E, Romeri F, Piazza A, Pagani C, et al. Genomic Characterization of an ST1153 PVL-producing Methicillin Resistant *Staphylococcus aureus* Clinical Isolate in Italy. New Microbiol. 2019;42: 129–131. Available: https://www.ncbi.nlm.nih.gov/pubmed/31034079.

101. Piazza A, Principe L, Comandatore F, Perini M, Meroni E, Mattioni Marchetti V, et al. Whole-Genome Sequencing Investigation of a Large Nosocomial Outbreak Caused by ST131

H30Rx KPC-Producing *Escherichia coli* in Italy. Antibiotics (Basel). 2021;10. doi:10.3390/antibiotics10060718.

102. Perini M, Piazza A, Panelli S, Di Carlo D, Corbella M, Gona F, et al. EasyPrimer: user-friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on *wzi* gene for fast *Klebsiella pneumoniae* typing. Sci Rep. 2020;10: 1307. doi:10.1038/s41598-020-57742-z.

103. Pasala AR, Perini M, Piazza A, Panelli S, Di Carlo D, Loretelli C, et al. Repeatability and reproducibility of the *wzi* high resolution melting-based clustering analysis for *Klebsiella pneumoniae* typing. AMB Express. 2020;10: 217. doi:10.1186/s13568-020-01164-7.

104. Perini M, Batisti Biffignandi G, Di Carlo D, Pasala AR, Piazza A, Panelli S, et al. MeltingPlot, a user-friendly online tool for epidemiological investigation using High Resolution Melting data. BMC Bioinformatics. 2021;22: 76. doi:10.1186/s12859-021-04020-y.

105. Perini M, Piazza A, Panelli S, Papaleo S, Alvaro A, Vailati F, et al. Hypervariable-Locus Melting Typing (HLMT): a novel, fast and inexpensive sequencing-free approach to pathogen typing based on High Resolution Melting (HRM) analysis. bioRxiv. bioRxiv; 2021. doi:10.1101/2021.07.01.450706.

106. Comandatore F, Chiodi A, Gabrieli P, Biffignandi GB, Perini M, Ricagno S, et al. Insurgence and worldwide diffusion of genomic variants in SARS-CoV-2 genomes. bioRxiv. bioRxiv; 2020. doi:10.1101/2020.04.30.071027.

107. Micheli V, Rimoldi SG, Romeri F, Comandatore F, Mancon A, Gigantiello A, et al. Geographical reconstruction of the SARS-CoV-2 outbreak in Lombardy (Italy) during the early phase. J Med Virol. 2021;93: 1752–1757. doi:10.1002/jmv.26447.

Appendices

Section A:

Appendix 1
Rice flour fermented with *Lactobacillus paracasei* CBA L74 in the treatment of atopic dermatitis in infants: A randomized, double- blind, placebo-controlled trial.

Appendix 2
Inflammatory bowel diseases, the hygiene hypothesis and the other side of the microbiota: Parasites and fungi

Appendix 3
What Pediatricians Should Know Before Studying Gut Microbiota.

Section B:

Appendix 4
Genomic Characterization of an ST1153 PVL-producing Methicillin Resistant *Staphylococcus aureus* Clinical Isolate in Italy.

Appendix 5
Whole-Genome Sequencing Investigation of a Large Nosocomial Outbreak Caused by ST131 H30Rx KPC-Producing *Escherichia coli* in Italy.

Section C:

Appendix 6
EasyPrimer: user- friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on *wzi* gene for fast *Klebsiella pneumoniae* typing.

Appendix 7
Repeatability and reproducibility of the *wzi* high resolution melting-based clustering analysis for *Klebsiella pneumoniae* typing.

Appendix 8
MeltingPlot, a user-friendly online tool for epidemiological investigation using High Resolution Melting data.

Appendix 9
Hypervariable-Locus Melting Typing (HLMT): a novel, fast and inexpensive sequencing-free approach to pathogen typing based on High Resolution Melting (HRM) analysis.

Section A

Appendix 1

Rice flour fermented with Lactobacillus paracasei CBA L74 in the treatment of atopic dermatitis in infants: A randomized, double- blind, placebo-controlled trial.

D'Auria E, Panelli S, Lunardon L, Pajoro M, Paradiso L, Beretta S, Loretelli C, Tosi D, Perini M, Bedogni G, Abdelsalam A, Fiorina P, Bandi C, Zuccotti GV.

Pharmacol Res. 2021 Jan;163:105284. doi: 10.1016/j.phrs.2020.105284

Pharmacological Research 163 (2021) 105284

Contents lists available at ScienceDirect

Pharmacological Research

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

Rice flour fermented with *Lactobacillus paracasei* CBA L74 in the treatment of atopic dermatitis in infants: A randomized, double- blind, placebo-controlled trial

Enza D'Auria [a,*], Simona Panelli [b], Luisa Lunardon [a], Massimo Pajoro [b], Laura Paradiso [a], Silvia Beretta [a], Cristian Loretelli [b], Diego Tosi [a], Matteo Perini [b], Giorgio Bedogni [c], Ahmed Abdelsalam [d], Paolo Fiorina [d,e], Claudio Bandi [d], Gian Vincenzo Zuccotti [a]

[a] Department of Paediatrics, Vittore Buzzi Children's Hospital, University of Milan, Milan, Italy
[b] Pediatric Clinical Research Center "Invernizzi", University of Milan, Milan, Italy
[c] Clinical Epidemiology Unit, Liver Research Center, Basovizza, Trieste, Italy
[d] Pediatric Clinical Research Center Fondazione Romeo e Enrica Invernizzi, Department of Biomedical and Clinical Science L. Sacco, University of Milan, Milan
[e] Nephrology Division, Boston Children's Hospital, Harvard Medical School, Enders Building 5th Floor Room EN511, 300 Longwood Ave, Boston, MA, USA

ARTICLE INFO

Keywords:
Atopic dermatitis
SCORAD
Microbiota
Postbiotics
Topical corticosteroids
Sparing-effect

ABSTRACT

To assess the effect of a fermented rice-flour obtained from *Lactobacillus paracasei* CBA L74 in managing infants with moderate to severe atopic dermatitis. Infants with moderate to severe atopic dermatitis, aged 6–36 months, were randomly assigned to receive once-daily consumption of rice flour containing heat-killed probiotic *Lactobacillus paracasei* CBA L74 or placebo for 12 weeks as supplementary approach to topical treatment. Primary outcome was SCORAD index change from baseline to 12 weeks; secondary outcomes were gut microbiota composition, as evaluated by the analysis of fecal samples, and serum cytokines at baseline and at the end of the intervention period in both groups, and steroid usage over the treatment period and one month after stopping it. V3–V4 region of the 16S ribosomal RNA gene was sequenced to evaluate changes in the gut microbiota. SCORAD index decreased over the treatment period in both groups. The difference in the SCORAD change was -2.1 (-5.5 to 1.3; p = 0.223) for the experimental vs. the placebo group, not reaching the minimal clinical difference of 8.7 units. The use of topical steroids, measured as finger tips units, decreased from 4 to 16 weeks, in both groups; the reduction was significantly higher in experimental than in placebo group (p value from Wilcoxon rank sum test = 0.031). No significant differences were observed for cytokines levels between groups. The composition of gut microbiota at the phylum and class taxonomic levels resulted very similar, at baseline and after intervention, in both groups. Similarly, no significant differences were observed in the relative abundance of bacterial genera between groups. In conclusion, though the heat-killed *Lactobacillus paracasei*was not proved to be effective in reducing the severity of atopic dermatitis, it showed a steroid sparing effect the value of which needs to be further investigated.

1. Introduction

Atopic Dermatitis (AD), an itchy eczema with a chronic relapsing course, is a common clinical manifestation of atopy in the first years of life and the most common chronic inflammatory skin disease, with a prevalence of 20 % in children [1]. AD, as a multidimensional disease, has a great impact on patients and their families' quality of life, comparable with epilepsy or type 1 diabetes. It also deeply affects social activities and patient well-being [2,3].

Based on the available knowledge, AD immunopathogenesis is characterized by an inflammatory reaction, characterized by a complex interplay between skin barrier disruption and immune inflammation [4,5].

Nowadays, it is recognized that type 2 cytokines play a common pathogenetic role in atopic diseases [6], leading to epidermal barrier dysfunction and skin inflammation [7].

* Corresponding author at: Department of Paediatrics, Vittore Buzzi Children's Hospital, University of Milan, Via Lodovico Castelvetro, 32, 20154, Milano, Italy.
E-mail address: enza.dauria@unimi.it (E. D'Auria).

https://doi.org/10.1016/j.phrs.2020.105284
Received 20 September 2020; Received in revised form 27 October 2020; Accepted 27 October 2020
Available online 4 November 2020

Imbalances in gut microbial composition and function, termed dysbiosis, have been linked to the pathophysiology of many disorders, including allergic diseases and AD, especially during early childhood [8,9].

Accordingly, lists of taxa specifically under- or over-represented in AD have been produced by a number of recent papers, with conflicting results. Reddel et al. (2019) reported a dysbiotic status of the gut microbiota, characterized by the reduction of *Bifidobacterium*, *Blautia*, *Coprococcus*, *Eubacterium*, *Propionibacterium* and an increase in *Faecalibacterium*, *Oscillospira*, *Bacteroides*, *Parabacteroides in children with AD* [1]. Zheng et al. (2016), in a case-control study of infants with eczema and controls, identified five genera significantly enriched in the former (*Escherichia/Shigella*, *Veillonella*, *Faecalibacterium*, *Lachnospiraceae incertae sedis* and *Clostridium XIVa*) and four in the latter (*Bifidobacterium*, *Megasphaera*, *Haemophilus* and *Streptococcus*) [10].

The conventional management of pediatric AD targets the restoration of the skin barrier using moisturizers and the prevention of flare-ups via topical corticosteroids (TCS). Calcineurin inhibitors are used as second-line effective agents [11]. Moderate-to-severe AD affected patients need to be treated with long-term applications of topical corticosteroids; long-term TCS therapy has been associated with adverse local effects, including skin atrophy and flares rebounds [12,13]. The AD skin barrier impairment can also increase TCS percutaneous absorption that leads to rare but severe adverse systemic effects, such as the hypothalamic pituitary-adrenal (HPA) axis suppression,poor growth, hypertension, hyperglycemia [14]. These safety concerns are increased in pediatric patients, whose greater body surface area-to-weight ratio is thought to cause increased percutaneous absorption.

Furthermore, parents' TCS-phobia often causes poor adherence to an appropriate treatment which may have a dramatic effect on diseases outcomes [15]. In addition, long-term safety data of mid- to high-potency TCS and calcineurin inhibitors are lacking for pediatric patients [16,17]. These considerations, joint to the observed gut dysbiosis, raised the interest for complementary treatment strategies, directed at immunomodulation, possibly targeting gut microbiota via dietary supplements [18–20].

Available data regarding the effects of probiotics and other functional food (prebiotics, e.g., non-digestible oligosaccharides or dietary substrates, stimulating the growth or activities of specific taxa within the gut microflora) are not conclusive yet and conflicting, as evidenced from recent meta-analyses [21,22].

Thus, one different possibility could be exploiting the probiotic-like effects of inactivated bacteria, with all their metabolic products (e.g., SCFA), of microbial fractions or isolated bacterial components, known as postbiotics [23].

The present randomized, double-blind, placebo- controlled study aimed to evaluate the effects of a fermented rice flour obtained from *Lactobacillus paracasei* CBA L74 in infants with moderate to severe atopic dermatitis.

2. Methods

2.1. Study design and study population

A prospective, randomized, double-blind, 2-arm placebo-controlled trial was conducted at Department of Pediatrics, Vittore Buzzi Children's Hospital in Milan (Italy). The study protocol was approved by the local ethics committee and was conducted in accordance with the principles of the amended Declaration of Helsinki and Ethical Guideline for Epidemiological Research. Written informed consent was obtained from parents or legal representative prior to enrollment in the study. Trial was registered in the Registry Clinical Trial (www.clinicaltrial.govI.D NCT 03157284).

Inclusion criteria were as follows: infants and young children age 6 months-36months old at the enrolment, with a diagnosis of AD according to Hanifin and Rajka criteria and moderate to severe atopic dermatitis according to the SCORAD (index)system [24]

SCORAD (SCORing Atopic Dermatitis) is a reliable and validated tool to assess the extent and severity of eczema in clinical trials [25]. It was developed by the European Task Force on Atopic Dermatitis in 1993 [24]. This index is a composite score, which combines objective criteria, such as extent and intensity of lesions and subjective ones, namely daily pruritus and sleeplessness. It ranges between 0 (lowest possible score) and 103 (highest possible score) and it is interpreted as follows: < 25: mild AD; 25–50 moderate AD; > 60: severe AD.

The major exclusion criteria were: mild dermatitis (SCORAD < 25), children with comorbid chronic diseases and/or dermatologic conditions other than AD, children using antihistamines, oral corticosteroids, pre and/or probiotics during the 4 weeks before enrollment, treatments with topical calcineurin inhibitors in the previous three months and recognized allergy or intolerance to probiotics or excipients of the tested product.

Records and diaries were kept of concomitant medications and adverse effects. Topical use of corticosteroids was allowed and all subjects were asked not to change the kind of topical therapeutics, as far as possible, through the intervention period.

Fifty-eight infants and young children (age 6–36 months), fulfilling the inclusion criteria, were enrolled in the study (Fig. 1). They were randomly assigned to receive rice fermented with *Lactobacillus paracasei* CBA L74 or placebo. All enrolled subjects were assessed on day 0 (baseline) and underwent medical examination at week 4, 8 and 12. At each visit, the Medical Investigators examined the number of days and the finger tips units (FTU) of TCS usages, the consumption of allotted study diets (experimental product and placebo) and the occurrence of adverse events based on the individual's study diary.

A flare was defined as a worsening of the disease leading to use of topical corticosteroids for at least 3 consecutive days.

2.2. Recruitment and randomization of participants

Children aged 6 months to 3 years were recruited over two years, from January 2017 to March 2017 and from October 2018 to February 2019, among patients admitted to the Allergy Unit or Dermatology Unit for AD evaluation. Two dermatologists with expertise in pediatric dermatology evaluated their AD and SCORAD index.

Participants were then randomized, using computer-generated 4-block design lists, drawn up by a statistician. Both the investigators and the subjects were blinded throughout the study regarding the assignment of the experimental product vs placebo.

2.3. Sample size

The minimum clinically important difference (MCID) for SCORAD is 8.7 units [24]. Using an analysis of covariance for repeated measures (RM-ANCOVA) [26], we calculated that a sample size of 25 subjects per group (Experimental group, EXP *vs*. Placebo group, PLA) ensures 80 % power to detect a difference of 8.7 as statistically significant at an alpha level of 0.05 assuming a mean difference (standard deviation) of -41.6 (17.0) between the first and the last measurement for EXP [27], three repeated measurements beyond baseline [27], a correlation of 0.5 between repeated measures [27], and the same standard deviation of the difference (17.0) for the EXP and PLA groups. Assuming an estimated dropout rate of 15 %, we decided to enroll 29 subjects per group.

2.4. Interventions

Subjects were treated for a period of 12 weeks. During the study period all patients received treatment or placebo, moisturizers, and topical corticosteroids in case of disease flares, according to the guidelines for the management of AD [13]. Patients in the experimental group received a daily 8 g of a rice-dried powder containing heat-killed *Lactobacillus paracasei* CBA L74; placebo rice-powder was matched for

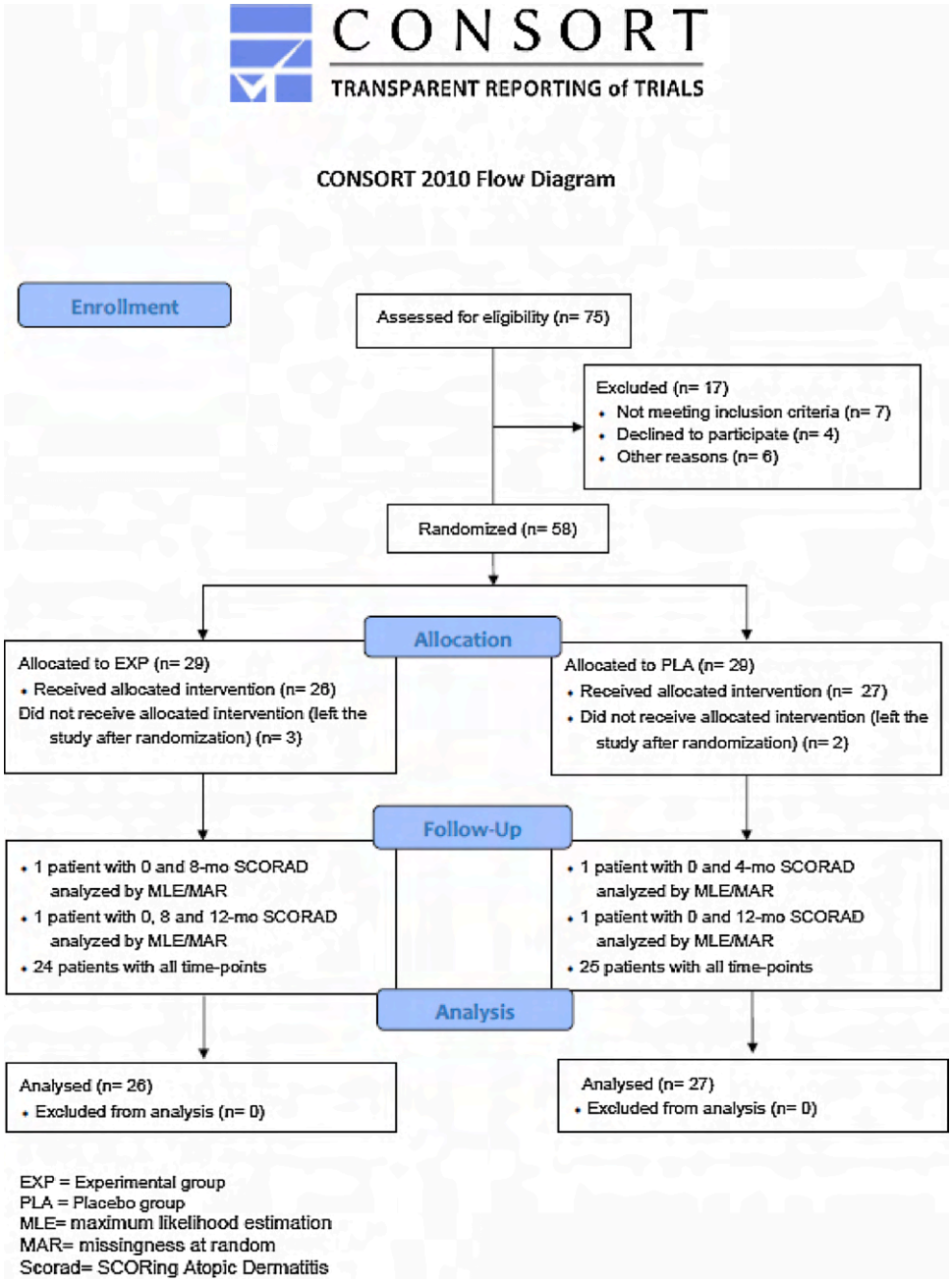


Fig. 1. The flow diagram of the study.

size, shape, and volume of contents. Both the fermented matrix and the placebo were dispensed by the protocol dedicated staff. Once daily each subject was fed with 8 g of *Lactobacillus paracasei* CBA L74 powder or placebo diluted in beverage or liquid food.

2.5. Outcomes evaluation

The primary outcome was the change in the severity of AD after 12 weeks of intervention, compared to baseline. The standardized SCORAD scoring system was used to evaluate AD severity [24]. Assessments and SCORAD Index were made at baseline and after 4, 8 and 12 weeks of treatment.

Secondary outcomes were the evaluation of the steroid-usage in both group at baseline, at 4, 8, 12 weeks after the start of treatment, and one month after stopping it; taxonomic composition and diversity indices of the gut microbiota, as evaluated by the analysis of fecal samples and serum cytokines levels at baseline and after 12 weeks intervention period.

The overall consumption of topical corticosteroids, measured as number of finger tips units (FTU) over the overall study period, from 4 to 16 weeks, was recorded by the parents and reported to the investigators.

Taxonomic composition and diversity indices were available for 21 PLA children out of 27 and 22 EXP children out of 26. Reasons were the following: one of the fecal samples at T0 or T12 (generally the latter) was not conferred in eight patients; two patients dropped-out from the study.

2.6. Serum cytokine assessment

Serum samples of PLA and EXP patients were collected both before (T0) and after 12 weeks (T12) from the treatment start, and then frozen at -80 °C. On the day of assessment, samples were allowed to gradually

thaw at +4 °C, spun at 10,000xg for 10 min and supernatants were then used for the evaluation of G-CSF, GM-CSF, IFN-γ, IL-1β, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12, IL-13, IL-17A, MCP-1, MIP-1β and TNF-α levels by a magnetic microsphere-based Bio-Plex Pro Human Cytokine 17-plex immunoassay on a Bio-Plex 200 system (both from Bio-Rad), following manufacturer's instructions. Briefly, 50 µl of 4-fold diluted serum samples were sequentially mixed with immunomagnetic beads, detection antibodies and with PE-labeled streptavidin, with appropriate washing between each step. Finally, hybridized beads were loaded on a Bio-Plex 200 reader (Bio-Rad) and fluorescence results were analyzed using the in-house Bio-Plex Manager software, v.6.0 (Bio-Rad).

2.7. Production of 16S rRNA amplicons (V3-V4 regions) and sequencing

Amplicon metagenomic analysis was performed in each subject enrolled in the study, at the initial (T0) and final (T12 timepoints). DNA was extracted from 200 mg of each faecal sample by the QIAamp DNA Stool Mini kit (Qiagen; Hilden, DE), following the manufacturer's protocol to assure an unbiased representation of the various bacterial populations. The DNA concentration of each sample was assessed fluorometrically. For amplicon production, the V3-V4 hypervariable regions of the prokariotic 16S rRNA gene were targeted [28]. PCR was set up in a 50-µl volume with template DNA, 1x HiFi Hot Start Ready Mix (Kapa Biosystems, Wilmington, MA), 0.5µM of each primer. The amplification was carried out on a Bio-Rad T100 thermal cycler (Bio-Rad, Hercules, CA) and included: initial denaturation (95 °C for 3 min); 30 cycles at 94 °C for 30 s (s), 55 °C for 30 s, 72 °C for 30 s; final extension (72 °C for 5 min). Clean-up of amplicons was performed using Agencourt AMPure XP SPRI magnetic beads (Thermo Fisher Scientific). Illumina sequencing libraries were finally prepared through the link of indexes (Nextera XT Index Kit, Illumina, San Diego, CA), quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA), normalized and pooled. Libraries were subjected to paired-end sequencing (2 × 300 bp format) on an Illumina MiSeq platform at BMR Genomics (Padova, Italy). Two amplicons were produced and sequenced for each subject enrolled in the study at the two timepoints.

2.8. Bioinformatics and community analyses

The bioinformatics analysis of sequencing data was based on the Mothur pipeline [29]. Briefly, raw FASTQ files were quality-filtered using Trimmomatic [30] and high-quality reads analysed following the SOP Mothur procedure. Chimeric sequences were identified and removed using UCHIME [31]. The remaining sequences were clustered into OTUs at the 97 % homology level using VSEARCH [32]. OTUs were finally annotated, and taxonomy assigned against the reference database SILVA [33]. The main ecological indexes of within-sample, alfa diversity (Shannon, Chao, inverse Simpson, Observed Richness) were computed using Mothur. Diversity in composition among samples (□-diversity) was evaluated at all taxonomic ranks (from phyla to genera) byplotting the relative heatmap using the function heatmap.2 of the Gplots R library [34], and the relative Principal Component Analysis (PCA) using the R library Ade4 [35].

2.9. Statistical analysis

Continuous variables are reported as median (50^{th} percentile) and interquartile range (IQR, 25^{th} and 75^{th} percentiles). Discrete variables are reported as the number and proportion of subjects with the characteristic of interest. Besides analyzing the change in SCORAD using the originally planned RM-ANCOVA [26], we used random effects linear regression (RER) to handle missing data using maximum likelihood estimation (MLE) [36]. Theoretically, multiple imputation (MI) could be applied to RM-ANCOVA to handle missing data, but we preferred MLE over MI because it is simpler to implement and more efficient [36]. MLE assume missingness at random (MAR), which was highly plausible on the basis of the missingness pattern (reported under Results). The response variable of the RER model was SCORAD; the predictors were baseline SCORAD, treatment (discrete, 0 = Placebo group; 1 = Experimental group), time (discrete, 0 = baseline; 1 = other times) and a treatment*time interaction (discrete); and the random effect was the child. It should be noted that the final time was coded as originally implied by the RM-ANCOVA design, i.e. as a weighted mean of SCORAD at 4, 8 and 12 weeks. RER was also used to analyze the other outcomes. All analyses were performed applying the intention to treat principal. Missing data were handled using MLE under the MAR assumption [36]. RER models similar to the above were used for the study of the other continuous outcomes. The only exception is that we did not include the baseline value of the outcome among the RER predictors. Changes in topical FTU of topical steroids from week 4 to week 16 were compared between groups using the Wilcoxon rank sum test.

Statistical analysis was performed using Stata 16.0 (Stata Corporation, College Station, TX, USA).

3. Results

75 children were assessed for eligibility and 17 were excluded for the following reasons: 7 not satisfying the inclusion criteria, 4 declining to participate, and 6 for other reasons.

58 children were randomized into the EXP ($n = 29$) and PLA ($n = 29$) groups.

3 children in the EXP and 2 in the PLA group decided not to proceed with the study shortly after randomization, leaving 26 children in the EXP and 27 children in the PLA group. 24 of the EXP children had all the planned SCORAD data points (0, 4, 8 and 12 months) available, one had just the baseline and 8-mo data points, and one had just the baseline, 8-mo and 12-mo data points.25 of the EXP children had all the planned data points available, one had just the baseline and 4-mo data points, and 1 had just the baseline and 12-mo data points.

3.1. Baseline characteristics of the study groups

No differences between groups at baseline (Table 1).

Even if sex distribution is likely due to chance, we took it into account by performing the RER analysis of the main outcome using sex as further covariable of the regression model.

3.2. Main outcome

The main outcome was the change of SCORAD in the EXP *vs.* the PLA group. As estimated by RER, the mean (95 %CI) SCORAD was 41.5 (37.5–45.5) at baseline and 21.8 (19.2–24.4) at the end of the treatment period in the PLA group, with corresponding values of 42.5 (38.4–46.6) and 18.7 (16.0–21.3) in the EXP group. The difference in the SCORAD change was -2.1 (-5.5 to 1.3) for the EXP *vs.* the PLA group ($p = 0.223$) (Fig. 2).

When sex was added as covariable to the RER model, the SCORAD

Table 1
The baseline characteristics of patients in both groups.

	PLA N = 29	EXP N = 29
Sex (female)	14 (48 %)	20 (69 %)
Age at enrollment (months)	12 (8; 23)	12 (6; 16)
Birth weight (g)	3250 (3000; 3600)	3230 (3040; 3645)
Delivery		
vaginal	22 (76 %)	21 (72 %)
cesarean	7 (24 %)	8 (28 %)
Formula milk (any)	6 (21 %)	3 (10 %)
Family - Atopy	23 (79 %)	18 (62 %)
Family - smoke	13 (45 %)	10 (34 %)
Family - pets	7 (24 %)	7 (24 %)
SCORAD	41 (32; 48)	44 (35; 46)

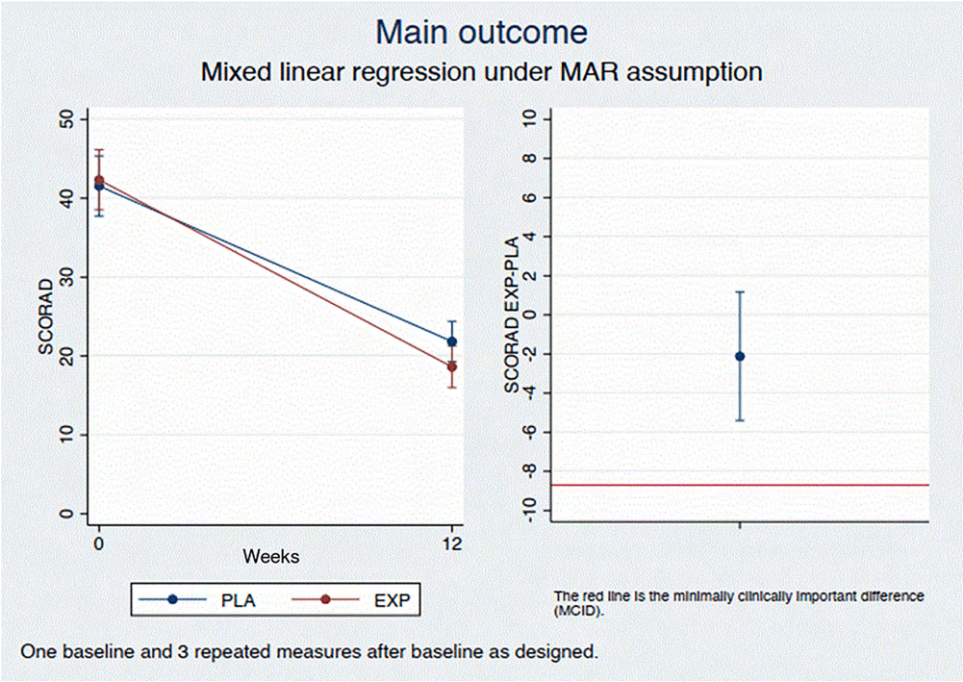


Fig. 2. SCORAD change from baseline to 12 weeks in Experimental (Exp) vs Placebo (PLA) group.

change was -2.7 (-6.1 to 0.7) in both boys and girls ($p = 0.126$ for the comparison). Thus, the different sex distribution of the EXP and PLA groups (Table 1) did not appear to have impacted the SCORAD change even if, of course, the 95 %CI of the within-sex difference are wide because of the reduced effective sample size.

It should be noted that the change in SCORAD calculated by the originally planned RM-ANCOVA was virtually identical that that obtained by MLR, i.e. -2.2 (95 %CI -5.6 to 1.3, p = 0.221). As explained under Statistical analysis, the reason why we preferred RER over RM-ANCOVA is its ability to take into account missing data using (MLE) [37].

3.3. Secondary outcomes

Evaluation of taxonomic composition and diversity indices of the gut microbiota as evaluated by the analysis of fecal samples and cytokines pattern:

The change in the relative abundance of nine preselected fecal bacterial genera for the EXP vs. the PLA groups (Akkermansia, Bacteroides, Bifidobacterium, Clostridium sensu stricto, Faecalibacterium, Lachnospira, Lactococcus, Rothia and Veillonella) was assessed. These 9 genera were selected on the basis of the available literature pointing to their possible association, either positive and negative with atopy. Instead of Lachnospira, the genus annotated as "Lachnospiraceae_unidentified" was considered, in agreement with the output of the analytical pipeline.

To this aim, two 16S rRNA amplicons covering the V3-V4 regions were obtained, sequenced and analysed at T0 and T12 timepoints for PLA and EXP groups. After quality filtering, a total of 40,495,212 high-quality reads were obtained (20,247,606 for the second repeat) and clustered into 12096 operational taxonomic units (OTUs) at 97 % similarity level. The average number of OTUs was 356.86 per sample (standard deviation: 151.94; min: 89; max: 826). These OTUs were classified into 19 phyla, 40 classes, 74 orders, 140 families and 330 genera.

The composition at the phylum and class taxonomic levels resulted very similar for PLA and EXP, at both timepoints, despite minor differences in relative rankings. Some differences were observed between the two groups for the Bacteroidetes and Firmicutes phyla. These differences, however, were limited to the T0 timepoint, and then tended to disappear at 12 weeks (T1) (Fig. 3). The same results were observed for classes distribution between PLA vs EXP (Fig. 4a-b)

As for classes, *Bacteroidia* show the behaviour described above for the corresponding phylum (*Bacteroidetes*); *Clostridia* (phylum: *Firmicutes*) display minor differences in the PLA and EXP groups that remain constant at both timepoints.

In most if not all subjects, the majority of preselected genera had low relative abundances, below the 0.05, which we considered the limit to give reliable results. On the basis of this criterion, we did not perform a formal analysis for the following genera: *Akkermansia*, *Clostridium*, *Faecalibacterium*, *Lactococcus*, *Rothia* and *Veilonella*. As determined by RER, the changes in prevalence for the 3 remaining genera in the EXP *vs.* PLA group were: 0.00 (95 %CI -0.06 to 0.05, $p = 0.877$) for *Bacteroides*, -0.02 (-0.04 to 0.01, $p = 0.186$) for *Lachnospira* and 0.00 (-0.02 to 0.02, $p = 0.903$) for *Veilonella*.

As shown in Fig. 4c, *Bacteroides* spp and "*Lachnospiraceae*_unidentified were the most abundant. The other 7 ones, together, made up <10 % of all genera in all cases and < 5 % in 6 out of 7 cases. *Rothia* and *Veillonella* resulted virtually indetectable. The very small baseline values possibly make it difficult to appreciate any change of these genera between EXP and PLA groups (Fig. 5). *The* change in the main ecological indexes of within-sample (a) diversity (Chao1, Shannon, Observed Richness, Inverse Simpson) were evaluated using RER. The treatment*time interaction was not significant for all cases (data not shown). As for between-sample (b) diversity, the principal component analysis on the Bray-Curtis dissimilarity matrix confirmed that the bacterial consortia had similar structures in PLA and EXP at both time points, taking into account the taxonomic level of family. In detail, children did not seem cluster on the basis of the postbiotic (not shown). As well as for gut microbiota composition, no significant differences were observed for the cytokines pattern between groups at baseline and after 12 weeks (data not shown).

We also evaluated if any treatment-dependent change occurred in the serum cytokine profile between the EXP and PLA. We thus assessed the levels of several relevant immunological factors – including the Type 2-related cytokines IL-4, IL-5 and IL-13 – in the serum of EXP and PLA patients, before and after treatment. Similar to the gut microbiota

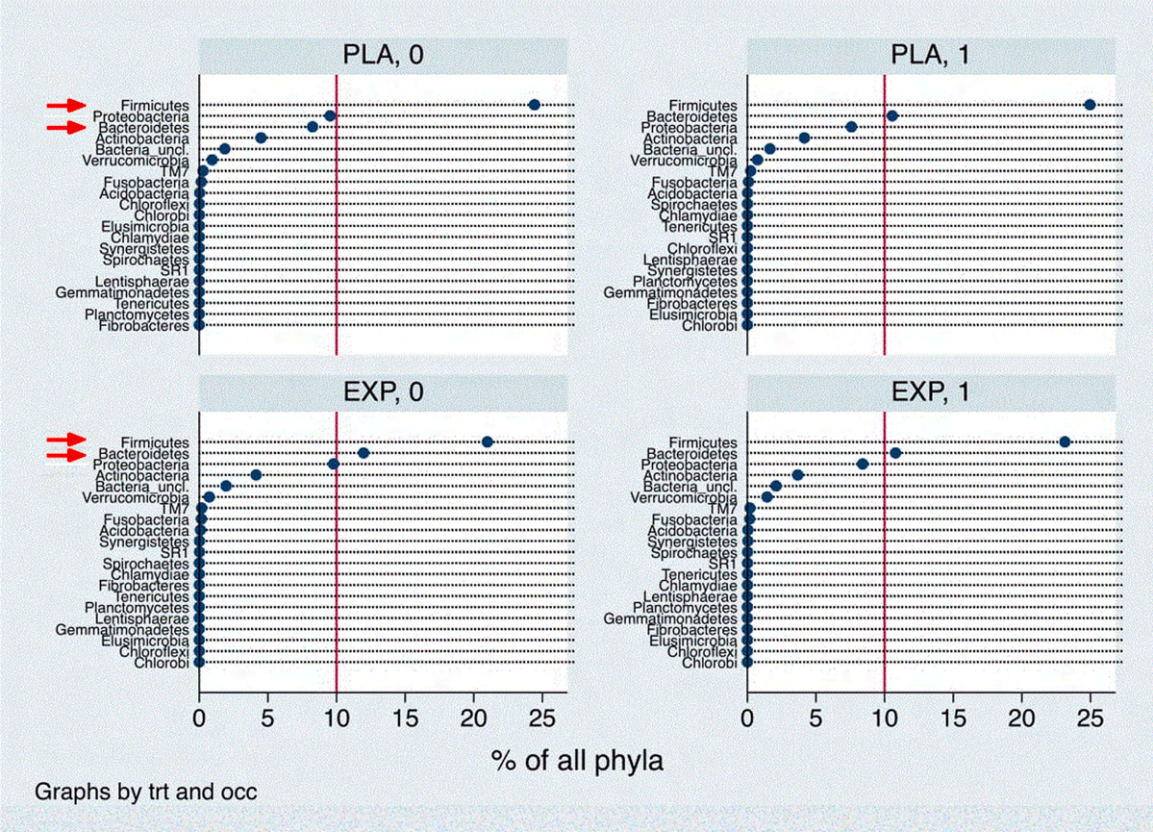


Fig. 3. Phyla distribution Experimental (EXP) vs Placebo (PLA) group at T0 and T1.

composition, no relevant treatment-associated changes in cytokine levels were noted between the two groups of patients after 12 weeks of intervention.

Steroid-sparing effect. Eight patients in the experimental group and 6 in the placebo group did not use any topical corticosteroids during the overall study period. Among patients reporting some corticosteroids use, FTU of topical steroids decreased from 4 to 12 weeks in both groups: EXP (median decrease: -10; interquartile range, I.Q.R -20; 0), PLA (median =-3.5; I.Q.R -10, 3; p value from Wilcoxon rank sum test = 0.197; at 16 weeks a significantly higher decrease was observed in experimental (median decrease: -14, I.Q.R: -25, -8) vs placebo group (median decrease: -4, I.Q.R: -10, 0); p value from Wilcoxon rank sum test = 0.031 (Fig. 6).

4. Discussion

Beside the well-known pre- and probiotics, the –biotic family includes postbiotics, bioactive compounds produced by food-grade microorganisms during a fermentation process. Postbiotics include microbial cells, cell's constituents and metabolites [23]. Probiotics have been deeply investigated as gut microbiota modulators; nevertheless, their therapeutic effect on AD is still controversial [38,21,39,40]. Reports on postbiotics products have been emerging for less than a decade and, more specifically, *in vivo* studies are a still relatively unexplored field in AD affected patients. *Lactobacillus paracasei* has been diffusely investigated and showed beneficial effects *in vitro* and *in vivo* as pro-inflammatory cytokines inhibitor ad T-regulatory cell-like response inductor [41].

Preclinical data show that *Lactobacilli*-originated postbiotics may be effective in reducing contact hypersensitivity reaction and development of atopic skin lesion [42]. Their biological responses have also been observed in human trials [43–45]. Postbiotics use could be an advantage in terms of safety profile, longer shelf life, and resistance to mammalian enzymes [46,47]. Furthermore, fermented matrix no need for refrigerated storage,they are stable to digestive system conditions and there is the chance to mix them in warm to hot liquids (47). To our knowdledge, *in vivo* studies using postbiotics in AD are limited to three, all of which using heat-killed *Lactobacillus paracasei,* and two performed in pediatric patients. The former, by Moroi et al., evaluated the effect of *Lactobacillus paracasei* K71 in a randomized, double-blind, placebo-controlled study on adult AD patients, who were treated with conventional topical corticosteroid and tacrolimus [48]. The second study, performed on pediatric subjects, considered only 12 children with moderate to severe AD in a repeated-measure cohort design [27]. The most recent one, by Yan et al., compared the effects of consumption of heat-killed *Lactobacillus paracasei* GM-080 to placebo, as a supplementary approach in AD affected infants, aged 4–30 months [49].

The present RTC primarily tested the effect of a fermented rice flour, containing heat killed *Lactobacillus paracasei* CBA L74 (Heinz Italia SpA, Latina, Italy) on AD severity as complementary approach to the conventional, topical treatment. To our best knowledge, this is the first study evaluating the effect of a postbiotic on clinical outcome, e.g AD severity, and gut microbiota composition. Over the period treatment, as determined by three repeated measures, SCORAD decreased of -3.1 (-5.5 to 2.3) in the EXP *vs.* the PLA group, the observed difference was below the minimal clinical difference of 8.7 units. Different sex distribution in the EXP and PLA groups had no impact on SCORAD change. Yan et al. obtained the same results while administering daily heat-treated probiotic *Lactobacillus paracasei* GM-080 for 16 weeks, as supplementary approach, in AD infants aged 4–30 months. Moroi et al. found a scant benefit in adult with AD who received a supplementary diet containing *Lactobacillus paracasei* K71 vs placebo. Even if Δ values of scores deviated more from baseline in the intervention group at weeks 8 and 12 (i.e., the differences resulted greater than for the placebo group), none of the differences reached a significant level. It should be pointed out that in this study skin severity scores were measured using the criteria of the

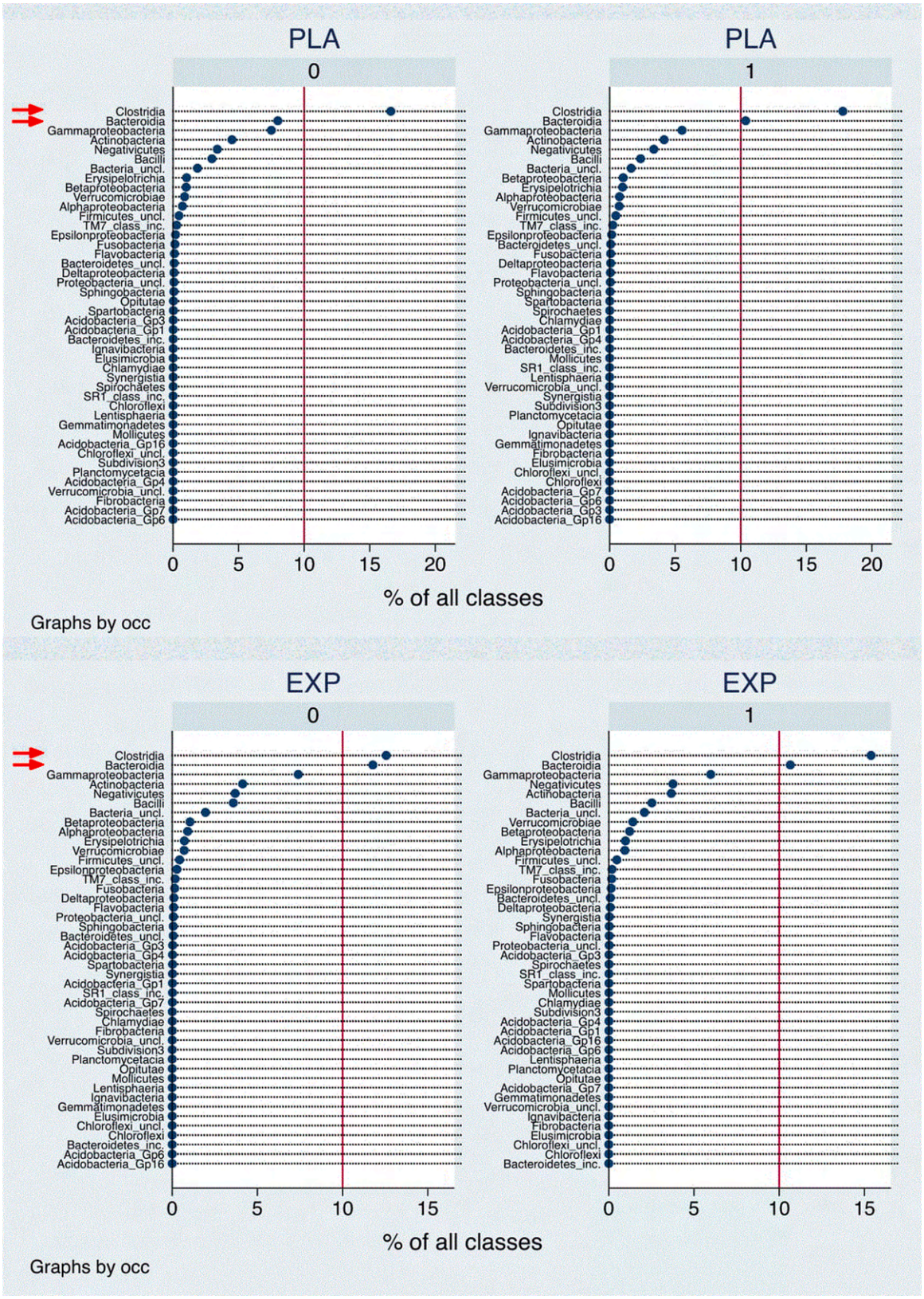


Fig. 4. Class distribution Experimental (EXP) vs Placebo (PLA) group at T0 and T1.

Japanese Dermatological Association, which is not considered among the instruments valid to measure the AD severity, according to the International Guidelines [50]. Promising results in Beretta et al.'s pilot study, might have been affected by the upcoming summer season, which could have led to AD remission, and were limited by the lack of a placebo control group.

So far, heat-inactivated *Lactobacillus paracasei* CBA L74 was not proved to be effective in reducing AD severity in infants with moderate

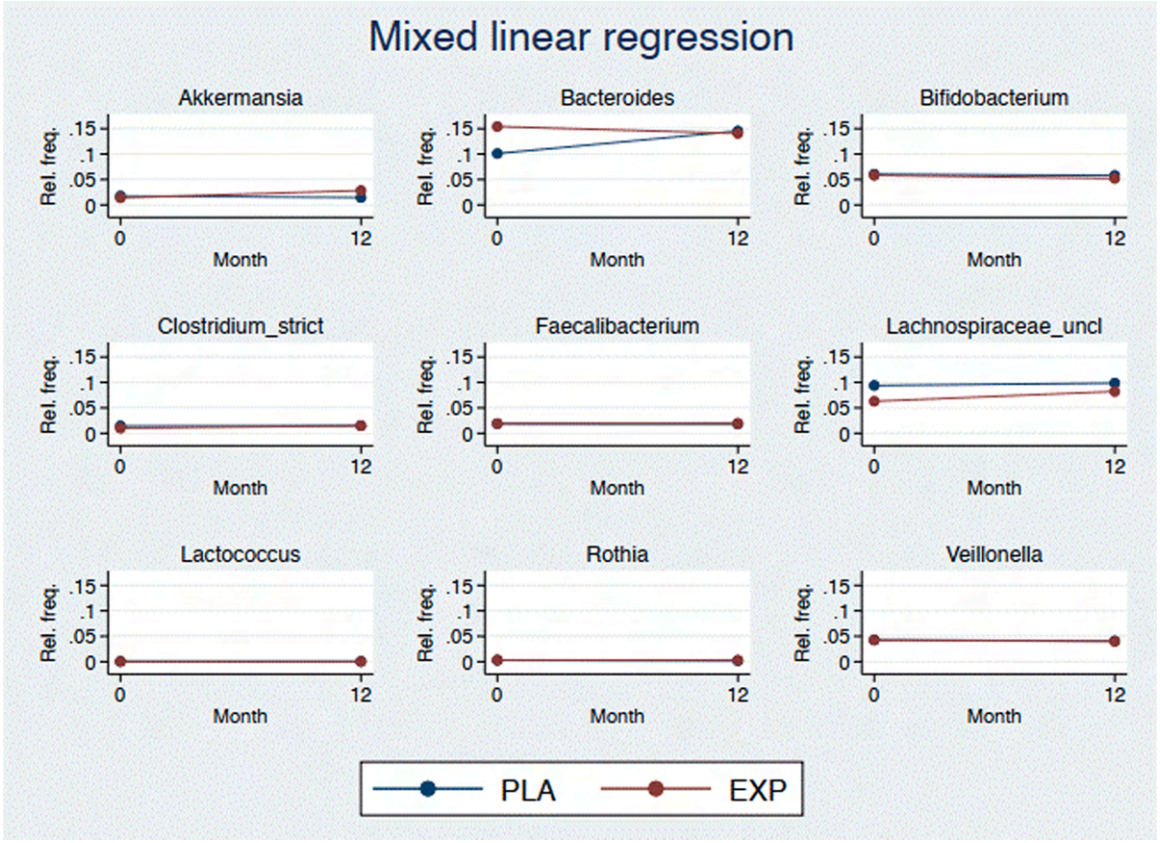


Fig. 5. Genera distribution Experimental (EXP) vs Placebo (PLA) group at T0 and T1.

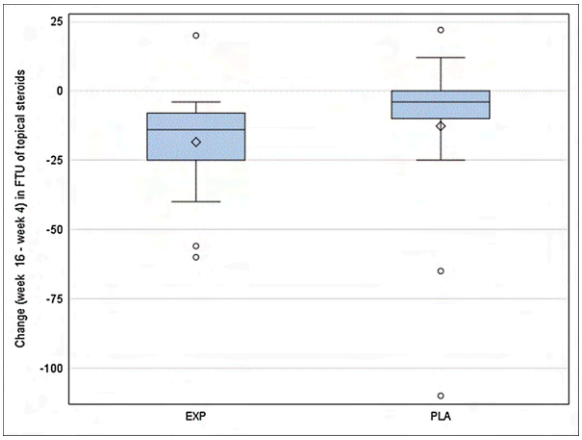


Fig. 6. Box plot for the change (week 16 – week 4) in FTU of topical steroids in Exp vs PLA groups.

to severe atopic dermatitis. Although the consumption of corticosteroids decreased at the end of the intervention period in both groups, the decrease was higher in experimental group, reaching a significant difference one month after stopping treatment. The steroid sparing effect observed in the group treated with heat-inactivated *Lactobacillus paracasei* CBA L74 might be explained by its anti-inflammatory properties, as proven by preclinical data [42] and in human trials [43–45]. In our setting, no relevant treatment-related change in the level of serum cytokines – including IL-4, IL-5 and IL-13 – was observed, suggesting that such anti-inflammatory effect might be potentially mediated by cell and/or metabolites, other than cytokines.

Regarding the analysis of the 9 pre-specified bacteria genera, there are differences between the results here reported and literature data; these may be due to demographic differences in the examined cohorts, different geographical locations, environment and diets, methodological and experimental differences.

In our hands, and for our population, slightly differential trends among EXP and PLA limited to *Akkermansia* (trend to increase at T1 in exp, in agreement with the results reported by Fujimura et al., that described the depletion in the gut microbiota associated with atopy and asthma). However, the caveats linked to the scarce representation of these genera remain [51].

Data for *Bacteroides* differ to those reported in Reddel et al. (2019). They reported an increase of the genus in atopic subjects while in our cohort we see it constant in the EXP group, considering the timepoints T0 and T12. On the contrary, an increase is appreciated in T1 for PLA. One important point is the different age of the enrolled subjects: 1–6 years for Reddel; 6–36 months for us. Another issue is that data on *Bacteroides* are not congruent in the various papers.

In our cohort, *Bacteroides* presents a behaviour comparable to what described above for its phylum, *Bacteroidetes,* and class, *Bacteroidia*. The same applies, to a minor extent, to *Lachnospiraceae_unclassified* compared to *Firmicutes* (phylum) and *Clostridia* (class). In summary, with regard to gut microbiota no significant differences were observed between the groups.

In conclusion, the present study did not prove the efficacy of a fermented rice flour obtained from heat treated *Lactobacillus paracasei* CBA L74 as a complementary approach in significantly reducing AD severity. However, heated killed *Lactobacillus paracasei* CBA L74 showed a corticosteroid sparing effect beyond the treatment period. This issue deserves further and more specific investigations in the light of the growing interest for steroid-sparing strategies. An effective adjunctive treatment in AD management could be able to minimize corticosteroid adverse effects and improve the patients and parents' adherence to the

therapy.

Funding

None to declare for all the Authors

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

The study was supported by a grant from Heinz Italia SpA, Latina, Italy to the Department of Pediatrics of Vittore Buzzi Children's Hospital, University of Milan. The sponsor had no influence on the study design; the collection, analysis, and interpretation of data; the writing of the manuscript; or the decision to submit the manuscript for publication.

References

[1] S. Reddel, F. Del chierico, A. Quagliariello, S. Giancristoforo, P. Vernocchi, A. Russo, et al., Gut microbiota profile in children affected by atopic dermatitis and evaluation of intestinal persistence of a probiotic mixture, Sci. Rep. 9 (2019) 4996.
[2] A.C. Bridgman, J.K. Block, A.M. Drucker, The multidimensional burden of atopic dermatitis: an update, Ann. Allergy Asthma Immunol. 120 (2018) 603–606.
[3] D. Boccardi, E. D'Auria, F. Turati, M. DI Vito, S. Sortino, E. Riva, A. Cerri, Disease severity and quality of life in children with atopic dermatitis: PO-SCORAD in clinical practice, Minerva Pediatr. 69 (5) (2017) 373–380.
[4] G. Yang, J.K. Seok, H.C. Kang, Y.Y. Cho, H.S. Lee, J.Y. Lee, Skin barrier abnormalities and immune dysfunction in atopic dermatitis, Int. J. Mater. Sci. 20: 21 (8) (2020) 2867.
[5] S. Tasneem, B. Liu, B. Li, M.I. Choudhary, W. Wang, Molecular pharmacology of inflammation: medicinal plants as anti-inflammatory agents, Pharmacol. Res. 139 (2019) 126–140.
[6] N.A. Gandhi, B.L. Bennett, N.M. Graham, G. Pirozzi, N. Stahl, G.D. Yancopoulos, Targeting key proximal drivers of type 2 inflammation in disease, Nat. Rev. Drug Discov. 15 (2016), 35-35.
[7] D.Y. Leung, E. Guttman-yassky, Deciphering the complexities of atopic dermatitis: shifting paradigms in treatment approaches, J. Allergy Clin. Immunol. 134 (2014) 769–779.
[8] L.B. van der Aa, H.S. Heymans, W.M. van Aalderen, A.B. Sprikkelman, Probiotics and prebiotics in atopic dermatitis: review of the theoretical background and clinical evidence, Pediatr. Allergy Immunol. 21 (2010) e355–e367.
[9] M. Nermes, J.M. Kantele, T.J. Atosuo, S. Salminen, E. Isolauri, Interaction of orally administered Lactobacillus rhamnosus GG with skin and gut microbiota and humoral immunity in infants with atopic dermatitis, Clin. Exp. Allergy 41 (2011) 370–377.
[10] H. Zheng, H. Liang, Y. Wang, M. Miao, T. Shi, F. Yang, et al., Altered gut microbiota composition associated with eczema in infants, PLoS One 11 (2016), e0166026.
[11] E. Galli, I. Neri, G. Ricci, E. Baldo, M. Barone, A. Belloni Fortina, et al., Consensus conference on clinical management of pediatric atopic dermatitis, Ital. J. Pediatr. 42 (2016) 26.
[12] U.R. Hengge, T. Ruzicka, R.A. Schwartz, M.J. Cork, Adverse effects of topical glucocorticosteroids, J. Am. Acad. Dermatol. 54 (2006) 1–15.
[13] L.F. Eichenfield, W.L. Tom, T.G. Berger, A. Krol, A.S. Paller, K. Schwarzenberger, et al., Guidelines of care for the management of atopic dermatitis: section 2. Management and treatment of atopic dermatitis with topical therapies, J. Am. Acad. Dermatol. 71 (2014) 116–132.
[14] J. Del Rosso, S.F. Friedlander, Corticosteroids: options in the era of steroid-sparing therapy, J. Am. Acad. Dermatol. 53 (Suppl. 1) (2005) S50–S58.
[15] A. Chiricozzi, P. Comberiati, E. D'auria, G. Zuccotti, D.G. Peroni, Topical corticosteroids for pediatric atopic dermatitis: thoughtful tips for practice, Pharmacol. Res. 158 (2020), 104878.
[16] E.C. Siegfried, J.C. Jaworski, J.D. Kaiser, A.A. Hebert, Systematic review of published trials: long-term safety of topical corticosteroids and topical calcineurin inhibitors in pediatric patients with atopic dermatitis, BMC Pediatr. 16 (2016) 75.
[17] T. Luger, M. Boguniewicz, Cork M. CarrW, M. Deleuran, L. Eichenfield, et al., Pimecrolimus in atopic dermatitis: consensus on safety and the need to allow use in infants, Pediatr. Allergy Immunol. 26 (4) (2015) 306–315.
[18] M. Sepidarkish, F. Farsi, M. Akbari-fakhrabadi, N. Namazi, A. Almasi-Hashiani, A. Maleki Hagiagha, J. Heshmati, The effect of vitamin D supplementation on oxidative stress parameters: a systematic review and meta-analysis of clinical trials, Pharmacol. Res. 139 (2019) 141–152.
[19] E. D'Auria, S. Barberi, A. Cerri, D. Boccardi, F. Turati, S. Sortino, et al., Vitamin D status and body mass index in children with atopic dermatitis: a pilot study in Italian children, Immunol. Lett. 181 (2017) 31–35.
[20] J. Kim, H. Kim, Microbiome of the skin and gut in atopic dermatitis (AD): understanding the pathophysiology and finding novel management strategies, J. Clin. Med. 8 (4) (2019) 444.
[21] A. Makrgeorgou, J. Leonardi-bee, F.J. Bath-hextall, D.F. Murrell, M.L. Tang, A. Roberts, R.J. Boyle, Probiotics for treating eczema, Cochrane Database Syst. Rev. 11 (2018), CD006135.
[22] Y.S. Chang, M.K. Trivedi, A. Jha, Y.F. Lin, L. Dimaano, M.T. García-romero, Synbiotics for prevention and treatment of atopic dermatitis: a meta-analysis of randomized clinical trials, JAMA Pediatr. 170 (2016) 236–242.
[23] C.A.M. Wegh, S.Y. Geerlings, J. Knol, G. Roeselers, C. Belzer, Postbiotics and their potential applications in early life nutrition and beyond, Int. J. Mol. Sci. 20 (2019) E4673.
[24] Severity scoring of atopic dermatitis: the SCORAD index. Consensus report of the European task force on atopic dermatitis, Dermatology (Basel) 186 (1993) 23–31.
[25] J. Schmitt, S. Langan, H.C. Williams, European Dermato-Epidemiology Network, What are the best outcome measurements for atopic eczema? A systematic review, J. Allergy Clin. Immunol. 120 (6) (2007) 1389–1398.
[26] J.R. Rausch, S.E. Maxwell, K. Kelley, Analytic methods for questions pertaining to a randomized pretest, posttest, follow-up design, J. Clin. Child Adolesc. Psychol. 32 (2003) 467–486.
[27] S. Beretta, V. Fabiano, M. Petruzzi, A. Budelli, G.V. Zuccotti, Fermented rice flour in pediatric atopic dermatitis.[letter], Dermatitis 26 (2015) 104–106.
[28] S. Takahashi, J. Tomita, K. Nishioka, T. Hisada, M. Nishijima, Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using Next-Generation Sequencing, PLoS One 9 (2014), e105592.
[29] P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, et al., Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, Appl. Environ. Microbiol. 75 (2009) 7537–7541.
[30] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[31] R.C. Edgar, B.J. Haas, J.C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection, Bioinformatics 27 (2011) 2194–2200.
[32] T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: a versatile open source tool for metagenomics, PeerJ 4 (2016) e2584.
[33] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, et al., The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic Acids Res. 41 (2012) D590–D596.
[34] H. Wickham, ggplot2: Elegant Graphics for Data Analysis. s.l., Springer Nature Editor, Heidelberg, 2016.
[35] S. Dray, A.B. Dufour, The ade4 package: implementing the duality diagram for ecologists, J. Stat. Softw. 22 (2007) 1–20.
[36] S. Rabe-Hesketh, Multilevel and Longitudinal Modeling Using Stata. Volume I: Continuous Responses, Stata Press Publication, College Station TX, 2012.
[37] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, 3rd edition, Wiley, 2019 s.l.
[38] E.B.M. Petersen, L. Skov, J.P. Thyssen, P. Jensen, Role of the gut microbiota in atopic dermatitis: a systematic review, Acta Derm. Venereol. 99 (2019) 5–11.
[39] M. Zhao, C. Shen, L. Ma, Treatment efficacy of probiotics on atopic dermatitis, zooming in on infants: a systematic review and meta-analysis, Int. J. Dermatol. 57 (2018) 635–641.
[40] R. Huang, H. Ning, M. Shen, J. Li, J. Zhang, X. Chen, Probiotics for the treatment of atopic dermatitis in children: a systematic review and meta-analysis of randomized controlled trials, Front. Cell. Infect. Microbiol. 7 (2017) 392.
[41] T. Von der weid, C. Bulliard, E.J. Schiffrin, Induction by a lactic acid bacterium of a population of CD4(+) T cells with low proliferative capacity that produce transforming growth factor beta and interleukin-10, Clin. Diagn. Lab. Immunol. 8 (2001) 695–701.
[42] R.D. Wagner, C. Pierson, T. Warner, M. Dohnalek, M. Hilty, E. Balish, Probiotic effects of feeding heat-killed Lactobacillus acidophilus and Lactobacillus casei to Candida albicans-colonized immunodeficient mice, J. Food Prot. 63 (2000) 638–644.
[43] C. Agostoni, O. Goulet, S. Kolacek, B. Koletzko, L. Moreno, J. Puntis, et al., Fermented infant formulae without live bacteria, J. Pediatr. Gastroenterol. Nutr. 44 (2007) 392–397.
[44] F. Indrio, G. Ladisa, A. Mautone, O. Montagna, Effect of a fermented formula on thymus size and stool pH in healthy term infants, Pediatr. Res. 62 (2007) 98–100.
[45] C.N. de Almada, C.N. Almada, R.C.R. Martinez, A.S. Sant'Ana, Paraprobiotics: evidences on their ability to modify biological responses, inactivation methods and perspectives on their application in foods, Trends Food Sci. Technol. 58 (2016) 96–114.
[46] K. Tsilingiri, M. Rescigno, Postbiotics: what else? Benef. Microbes 4 (2013) 101–107.
[47] J.E. Aguilar-Toalá, R. Garcia-Varela, H.S. Garcia, V. Mata-Haro, A.F. González-Córdova, B. Vallejo-Cordoba, et al., Postbiotics: an evolving term within the functional foods field, Trends Food Sci. Technol. 75 (2018) 105–114.
[48] M. Moroi, S. Uchi, K. Nakamura, S. Sato, N. Shimizu, M. Fujii, et al., Beneficial effect of a diet containing heat-killed Lactobacillus paracasei K71 on adult type atopic dermatitis, J. Dermatol. 38 (2011) 131–139.
[49] D.C. Yan, C.H. Hung, L.B. Sy, K.H. Lue, I.H. Shih, C.Y. Yang, et al., A randomized, double-blind, placebo-controlled trial assessing the oral administration of a heat-treated Lactobacillus paracasei supplement in infants with atopic dermatitis

receiving topical corticosteroid therapy, Skin Pharmacol. Physiol. 32 (4) (2019) 201–211.
[50] J. Schmitt, S. Langan, S. Deckert, A. Svensson, L. Von Kobyletzki, K. Thomas, et al., Assessment of clinical signs of atopic dermatitis: a systematic review and recommendation, J. Allergy Clin. Immunol. 132 (2013) 1337–1347.
[51] K.E. Fujimura, A.R. Sitarik, S. Havstad, D.L. Lin, S. Levan, D. Fadrosh, et al., Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation, Nat. Med. 22 (2016) 1187–1191.

Section A

Appendix 2

Inflammatory bowel diseases, the hygiene hypothesis and the other side of the microbiota: Parasites and fungi.

Panelli S, Epis S, Cococcioni L, Perini M, Paroni M, Bandi C, Drago L, Zuccotti GV.

Pharmacol Res. 2020 Sep;159:104962. doi:10.1016/j.phrs.2020.104962

Inflammatory bowel diseases, the hygiene hypothesis and the other side of the microbiota: parasites and fungi

Simona Panelli[a], Sara Epis[b], Lucia Cococcioni[c], Matteo Perini[a], Moira Paroni[d], Claudio Bandi[b*], Lorenzo Drago[e], Gian Vincenzo Zuccotti[c]

[a] Department of Biomedical and Clinical Sciences “L. Sacco” and Pediatric Clinical Research Center “Romeo ed Enrica Invernizzi”, University of Milano, Italy

[b] Department of Biosciences and Pediatric Clinical Research Center “Romeo ed Enrica Invernizzi”, University of Milano, Italy

[c] Department of Pediatrics, Children's Hospital Vittore Buzzi, University of Milano, Italy

[d] Department of Biosciences, University of Milano, Italy

[e]Department of Biomedical Sciences for Health, Università di Milano

*Corresponding Author. Email: claudio.bandi@unimi.it

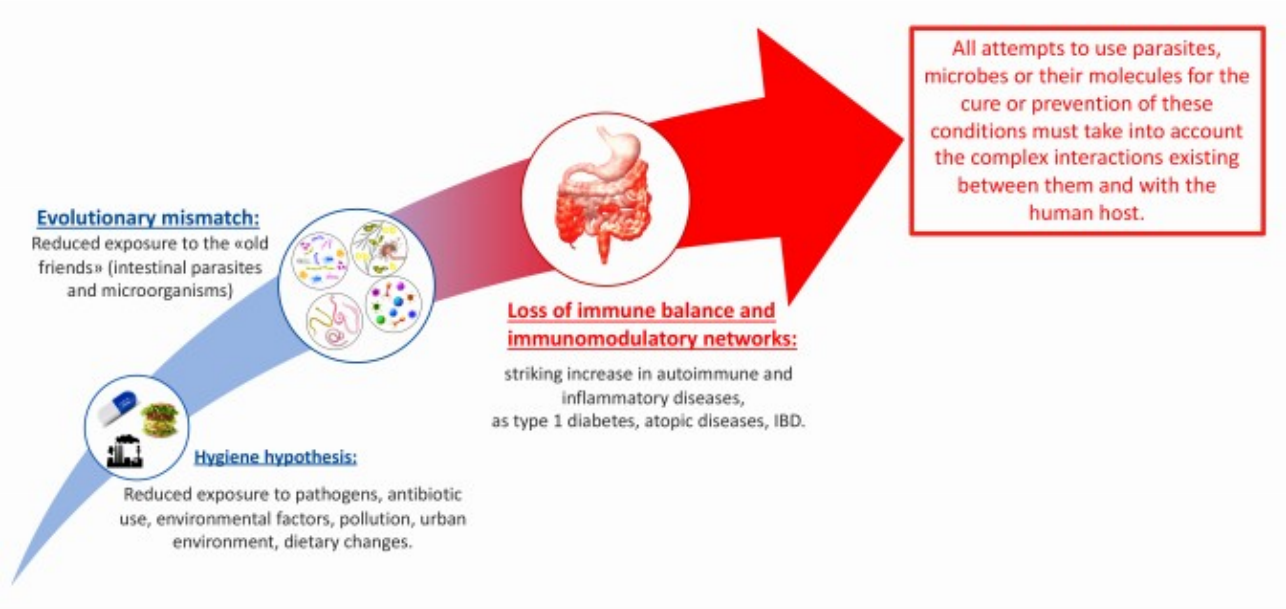
Keywords: microbiota; IBD; Evolutionary mismatch; Darwinian medicine; mycobiota, hygiene hypothesis

Declaration of interest: none

ABSTRACT

This review tackles the concept of the evolutionary mismatch, in relation with the reduction of the prevalence of the so-called “dirty old friends”. These formed the variegated community of parasites and microorganisms, either prokaryotic or eukaryotic, that, over long evolutionary times, co-evolved with humans and their ancestors, inhabiting their digestive tracts, and other body districts. This community of microbial symbionts and metazoan parasites is thought to have evolved a complex network of inter-independence with the host, in particular in relation with their immune stimulating capacity, and with the consequent adaptation of the host immune response to this chronic stimulation. Strictly related to this evolutionary mismatch, the hygiene hypothesis, proposed by David Strachan in 1989, foresees that the increase in the incidence of inflammatory and autoimmune disorders during the twentieth century has been caused by the reduced exposure to parasites and microorganisms, especially in industrialized countries. Among these pathologies, inflammatory bowel diseases (IBDs) occupy a prominent role. From these premises, this review summarizes current knowledge on how variations in the composition of the gut bacterial microbiota, as well as its interactions with fungal communities, influence the overall immune balance, favoring or counteracting gut inflammation in IBDs. Additionally, the effect of worm parasites, either directly on the immune balance, or indirectly, through the modulation of bacterial and fungal microbiota, will be addressed. Finally, we will review a series of studies related to the use of molecules derived from parasitic worms and fungi, which hold the potential to be developed as postbiotics for the treatment of IBDs.

Graphical abstract



Darwinian medicine: evolutionary mismatch and the hygiene hypothesis

A key concept in Darwinian medicine is that of evolutionary mismatch [1]. According to this concept, an organism that evolved in a given environment might be maladapted to a novel, different environment. In other words, a mismatch in the conditions in which we live today, compared to the conditions in which our genes and bodies had been shaped by natural selection, might imply that several hereditary traits are maladapted; this, in turn, could lead to overall perturbances in homeostatic equilibria. In this context, we should first consider the scenario in which *Homo sapiens* is supposed to have evolved. Whether we were mainly hunters or gatherers, the hygienic conditions of our ancestors in the pre-agricultural era were likely similar to those of wild animals. And the common condition for a wild mammal is to be repeatedly or chronically infected by parasites [2]. There is no reason to think that this was not true for our ancestors, who lived in small nomadic communities, formed by 150-250 individuals, for tens of thousands of years [3]. The history of humankind (or part of it) then experienced a dramatic change with the shift to agriculture and husbandry, approximately 10.000 – 5000 years ago - depending on the locations [4,5]. With this passage, the co-existence with parasites and pathogens in general not only did not diminish, but it is even likely to have increased, at least for some pathogens. The almost co-habitation with animals and the sedentary lifestyle facilitated the transmission of disease agents from livestock to humans, and the fecal-oral transmission of parasites and microbes. Co-existence with parasites, and with a variety of other “almost-pathogenic” microorganisms, is thus to be regarded as the normal condition for our ancestors, in the pre-agricultural era. This likely led to the establishment of a complex network of interdependence, among us, our parasites, and all the microbes that we harbored. Then something happened. With the transition to industrial society, and later with the adoption of more hygienic living conditions, the human environment underwent a rapid and important change [6]. Even the diet had changed over time, with the transition to agriculture, then with the progresses in food production and distribution, in the nineteenth and twentieth centuries [7,8]. In conclusion, we, humans, evolved in a type of environment that no longer exists, and are now living in a radically different condition: we are experiencing an evolutionary mismatch. In summary, our immune system evolved to cope with a variety of parasites and a multitude of quasi-parasitic (or quasi-mutualistic) microorganisms, the so called “dirty old friends”, that almost disappeared from our bodies, in several countries, during the last century.

The concept of the evolutionary mismatch, related to the reduction in the prevalence of the dirty old friends, also forms a theoretical background for the hygiene hypothesis. Proposed by Strachan in 1989, the hygiene hypothesis foresees that the increase in the incidence of inflammatory and autoimmune disorders during the twentieth century has been caused by the reduced exposure to parasites and microorganisms in industrialized countries [9,10]. The above picture is presented schematically in Figure 1. Among such pathologies, inflammatory bowel diseases (IBDs) occupy a prominent position.

The following paragraphs summarize current knowledge on how variations of the gut microbiota influence the overall immune balance, and on the effects of parasitic worms and fungi on gut immunity, either direct and indirect (i.e., exerted through the modulation of the microbiota). Finally, we will explore whether parasitic worms and fungi could be exploited to derive novel products for the cure or prevention of IBDs.

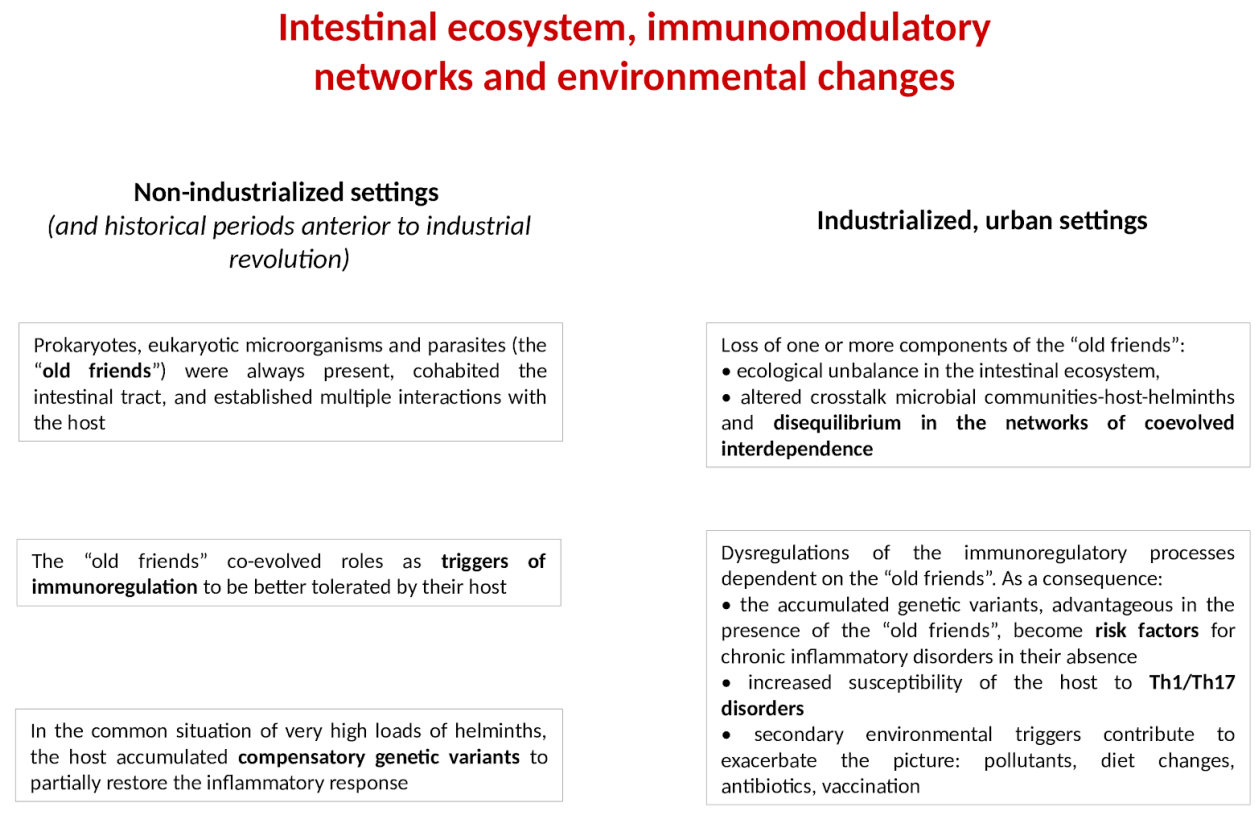


Figure 1. *The hygiene hypothesis and the "dirty old friends": the "old friends" co-evolved as "interactors" of immunoregulatory networks to be better tolerated by their hosts. Loss of one or more components, or unbalances among them, lead to perturbations in immunomodulatory cross-talks and can act as triggers of inflammatory, autoimmune disorders.*

2. Alterations of gut microbiota in IBDs

In this review, the term gut microbiota is used in its broadest sense to comprehend either its prokaryotic and eukaryotic component, and to refer to both microorganisms (bacteria, protozoans, fungi) and metazoans (tapeworms and roundworms). The review is focused on two eukaryotic components of the microbiota: fungi and parasitic worms. Concerning bacteria, we give a short introduction on their role in IBDs in this paragraph (for comprehensive reviews see: Goodrich et al., 2014 [11]; Kostic et al., 2014 [12]; Imhann et al., 2018 [13]; Ni et al., 2017 [14]; Andersen et al., 2019 [15]). We will not deal with protozoa, in relation to the great taxonomic diversity and phylogenetic depth of these organisms, from which an equally large immunological diversity is likely to be derived. Indeed, despite their phylogenetic diversity, protozoans remained largely unexplored in relation to IBDs and to the concept of the hygiene hypothesis (for a recent review on protozoans as a component of the gut microbiota, see Matijašić et al. 2020 [16]).

During the last 20 years, the bacterial component of the gut microbiota emerged for its prominent role in the development and preservation of a proper balance of the immune system. While individual variability among the bacterial taxa composing the gut microbiota is wide, and no unique formula exists for a “healthy microbiota” thanks to a remarkable functional redundancy [17], the maintaining of the relative balance among different microbial groups is pivotal for the host’s health. Alterations of this balance are referred to as dysbiosis and have been correlated with a plethora of pathological conditions [18–20]. IBDs are among the first pathologies that have been associated with dysbiosis in the gut microbiota [21]. For example, even before the recent development of metagenomics, an altered composition of the gut microflora in IBD patients had been described, with a significantly increased representation of anaerobic, sulfate-reducing bacteria (e.g., *Desulfovibrio* spp.) and a consequent increased production of molecules, as hydrogen sulphide, known to exert toxic, apoptotic effects on intestinal epithelial cells [22,23]. In general, many concordant data suggest that IBDs may be triggered by an aberrant inflammatory response of the mucosal immune system towards intestinal bacteria, in which the commensal bacteria themselves play a key role. The fact that IBDs respond to short-term antibiotic treatment [24] supports the idea that intestinal bacteria contribute to the inflammatory picture.

The pathogenesis of IBDs is multifactorial (Figure 2). Important factors at work, interacting with each other and with the gut microbiota, are related to host genetics and

to epigenetic modifications [13,25–27]. Individual genetic makeup seems necessary, even if not sufficient, for IBD pathogenesis, as witnessed by the familial occurrence of the disease [28]. To date, more than 200 susceptibility loci have been identified that contribute to the multifactorial onset of the disease [25]. The corresponding genes are mainly involved in host-microbe interactions, T cell signalling, innate immunity and epithelial barrier function [25]. One gene that has been the subject of extensive work is *NOD2*, whose mutations have been associated with Crohn's diseases (CD) by several authors (Yamamoto and Ma, 2009 [29] for a review). *NOD2* codes for an intracellular bacterial receptor.

There is now strong evidence indicating that the composition of the bacterial component of the gut microbiota is influenced by the composition of the eukaryotic component: protozoans, fungi and metazoan parasites (see sections 3 and 4). On its turn, the bacterial component is likely to influence the eukaryotic microbiota. In this tripartite interplay, of prokaryotes, eukaryotes, and host, the further multi-faced factor to be considered is the environment in its broadest sense, which include the diet. As already emphasized, in this complex interaction the immune system of the host is thought to play a major role, being one of the effector arms that participate to the modulation of the microbiota composition, and a major trigger of gut inflammation and disease in IBD patients.

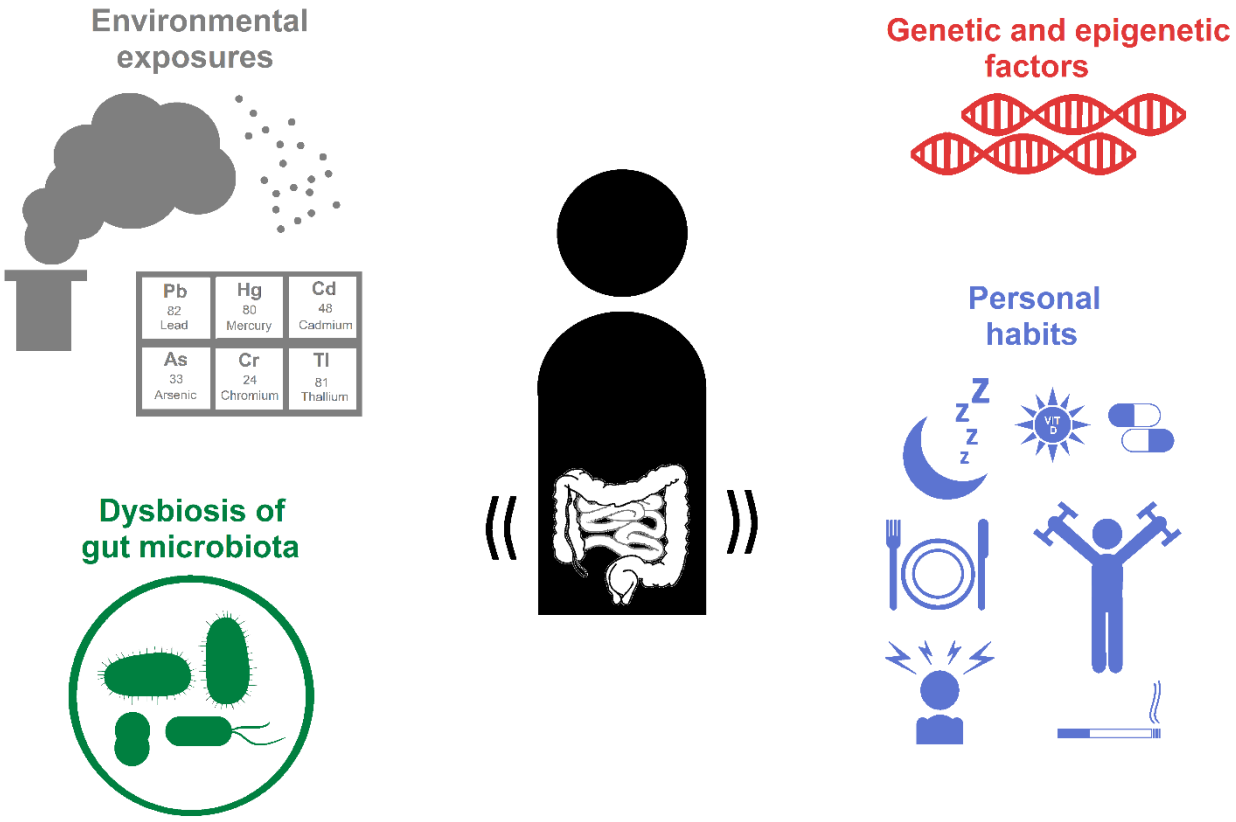


Figure 2. *The multifactorial origins of IBDs: genetics and epigenetics, environmental factors, gut microbiota and individual habits (clockwise from above: sleep, use of drugs, physical activity, smoking, stress, diet).*

3. Gut-inhabiting fungi and other eukaryotic microbes: implications for IBDs

Most of the studies thus far published on the role of the microbiota in IBDs, including those quoted in section 2, are focussed on the bacterial component. Microbiologists are now facing with another taxonomic community in human gut, represented by fungal microbiota (or mycobiota), less abundant than the bacterial microbiota, and still less explored in its impacts on the host's health and immunity. Since the first studies on this component of the gut microbiota, scientists are discovering year by year new species and taxa of fungi in some ways associated with faeces/gut content, nearby 300 so far [30]. The main challenge remains whether this fungal taxonomic diversity associated with the gut can be considered persistent/symbiotic, or instead transient/contaminant. Hoffman et al. by sequencing Internal Transcribed Spacer 1 (ITS1) regions of 98 individuals have retrieved many fungal taxa, with the top three being *Saccharomyces*, *Candida*, and *Cladosporium* [31]. The authors then identified a 'core mycobiome', with

only 29 fungal genera shared among these individuals. A more recent study reported a high prevalence of the genera *Saccharomyces*, *Malassezia* and *Candida*, with the most abundant species represented by *Saccharomyces cerevisiae*, *Malassezia restricta* and *Candida albicans [32]*. This study was conducted by collecting faecal samples from 317 healthy donors from the Human Microbiome Project. These and other studies established that *Candida* species, including *C. albicans*, might represent true gut symbiotic fungi, while the symbiotic status of *S. cerevisiae* and *M. restricta* is still debated [33,34].

Recently, dysbiosis of the human mycobiota, typically accompanied by the expansion of *Candida* spp., has been correlated to signatures tracts observed in IBDs as well as in several other diseases, as peptic ulcers, antibiotic-associated diarrhoea, hepatitis, chemotherapy-induced enteric disorders, graft-versus-host disease [35]. A very comprehensive study of the human mycobiota in IBD patients, recently conducted by Sokol et al., demonstrated an increase in *C. albicans* during the inflammatory flares. Additionally, the bacterial microbiota in IBD patients presented disease-specific correlations with the fungal microbiota in comparison to healthy controls, suggesting a bacterial dysbiotic state causing fungal expansion [36]. For example, these authors reported positive correlations, in IBD patients, between the abundance of protective fungi (*Saccharomyces cerevisiae*) and bacteria (*Bifidobacterium, Blautia, Roseburia, Ruminococcus*). These taxa, able to promote an anti-inflammatory environment, are all decreased in IBD patients, especially during flares. A negative correlation, instead, linked the same bacterial genera with fungi from the *Basidiomycota* phylum (class: *Malasseziales*). Other researchers reported a positive correlation between the abundance of *Candida tropicalis*, significantly increased in IBDs, and the potentially pathogenic bacterial species *Serratia marcescens* and *E. coli*, and observed the formation of a triple-species biofilm *in vitro* [37]. Linked to these observations, Standaert-Vitse et al. previously demonstrated that serum antibodies against *S. cerevisiae* mannan (ASCA), now considered pan-fungal antibodies, are elevated in several inflammatory diseases, including CD [37,38].

Even if the links mycobiota-gut immunity are still partially unknown, Leonardi et al. (2018) [39] recently identified the CX3CR1^{+} subset of mononuclear phagocytes (MNPs) as being essential for the initiation of innate and adaptive immune responses to intestinal fungi. CX3CR1^{+} MNPs may express antifungal receptors and activate antifungal responses. Indeed, the authors found that genetic ablation of CX3CR1^{+}

MNPs in mice led to changes in gut fungal communities (principally an increase in α-diversity, mainly ascribable to an increased abundance and diversity among the *Ascomycota* phylum) and to severe colitis, rescued by antifungal treatment. In some CD patients, a missense mutation in the gene encoding CX3CR1 was identified and found to be associated with impaired antifungal responses.

A further recent study [40] showed that *M. restricta* is another example of fungal species related to an IBD-associated polymorphism, in this case in the gene coding for CARD9, a signalling adaptor important for anti-fungal defence. *M. restricta* indeed elicits innate inflammatory responses largely through CARD9 and is recognized by specific antibodies in CD patients. This yeast strongly stimulates inflammatory cytokine production from innate immune cells bearing the IBD-linked mutation in the *CARD9* gene, and also exacerbates colitis via CARD9 in mouse models of disease. In summary, these studies suggest that targeting specific commensal fungi could provide novel strategies toward the development of IBD therapies.

Fungi are part of the gut microbiome, and trans-kingdom interactions should be expected, e.g. the already mentioned bacterial modulation of fungi. Indeed, bacterial pathobionts of the gut have recently been shown to produce both fungal modifying compounds as well as secreted enzymes with effects on the fungal component of the gut ecosystem [37]. Trunk et al. have identified two types of secretion system proteins produced by *Serratia marcescens*, called Tfe1 and Tfe2, with antifungal properties, especially against *C. albicans*, *Candida glabrata* and *S. cerevisiae*. This study indicates a dynamic *in vivo* relationship between *Serratia* and *Candida*, with possible involvement in the disease-induced inflammation in IBDs [41]. Overall, these data are consistent with resource competition models. Other concordant evidences show that short chain fatty acids (SCFAs) produced by bacteria can impair *C. albicans* colonization and translocation across the intestinal barrier, and that there is a consequent fungal overgrowth following antibiotic treatment [42]. Another recent study demonstrated that fungal colonization of the gut is also affected by *Enterobacteriaceae*, which can indirectly modify the effects of mycobiota on the gut. In this study, the bacterial and fungal relative abundances resulted dramatically decreased in colistin-treated mice, suggesting that colistin-sensitive bacteria as *Enterobacteriaceae* are involved in interactions with fungi [43]. In paediatric patients several authors have shown the existence of peculiar dynamics of host–mycobiome interactions.

There is growing evidence that fungal opportunistic pathogens are associated with IBD pathogenesis and chronicity. This concept is further supported by the observations that CD patients display elevated antibodies against fungal targets, even before disease diagnosis [44]. From an immunological point of view, it is well demonstrated that fungal species in the gut interact with the host immune system via the innate immune receptor Dectin-1, which recognizes β-1,3-glucans of the fungal cell wall. Dectin-1 and its isoforms activate intracellular signals through CARD9, leading to polarizing cytokine production and, consequently, promote T-helper 1 and T-helper 17 (Th1 and Th17) differentiation [45,46]. As detailed above, CARD9 genetic variants are known to be associated with IBDs [47]. The same pathway, dependent on the glucan components of the fungal wall, is also able to induce the so-called immune training, consisting in an immune memory of innate immune cells, which are influenced by previous contacts with pathogens or their products, and respond more effectively in the case of a second infection, or in some cases even in the case of unrelated infections [48].

Current treatments of IBDs rely on the early introduction of immune-targeted therapies to reduce chronic inflammation. The most used therapeutic antibodies for IBDs target tumour necrosis factor α (TNF-α), interleukins (IL) 12 and 23, and leukocyte adhesion molecules. A potential drawback of these therapies, looking at the mycobiota, could be the risk of fungal overgrowth. In fact, immune responses to fungal pathogens like *C. albicans* involve TNF-α, IL-17, and IL-23 [49,50], and notably some therapeutic strategies are now associated with increased incidence of fungal infection, such as histoplasmosis, blastomycosis, and coccidioidomycosis [51].

As discussed in section 2, current knowledge on the aetiology of IBDs strongly suggests that the gut bacterial microbiota has an important role in triggering the inflammation. Literature also reports that mycobiota unbalances (fungal dysbiosis) may occur in parallel, suggesting complex inter-kingdom ecological interactions. The study of gut mycobiota may thus reveal the etiological role of certain fungal taxa in IBDs, and probably suggest how to obtain the best outcomes for patients during treatments. In the light of these data and considered the heterogeneity in responsiveness to treatments across patients, future research on IBD therapy should carefully consider the potential role of fungal involvement in IBD pathogenesis.

4. Parasitic worms and their direct and indirect implications in IBDs

As discussed in section 1, parasitic worms have been colonizing the intestinal tract of humans and human ancestors for millions of years [52]. WHO estimated that more than 1.5 billion people are infected by soil-transmitted helminths [53]. Among soil-transmitted helminths, the most common species that infect humans are the roundworm (*Ascaris lumbricoides*), the whipworm (*Trichuris trichiura*) and hookworms ***(****Necator americanus* and *Ancylostoma duodenale*). For instance, still nowadays, an estimated 1 billion people in the world are infected by *Ascaris lumbricoides* [54] and over 600 million people by *Strongyloides stercoralis*. All these parasitic nematodes release factors, such as immunomodulatory proteins, glycoproteins, micro RNAs (miRNAs), which modulate the activity of the immune system of the human host.

During a helminth infection, the host rises a strong type 2 immune response, to provide protection against excessive worm colonization [55]. Type 2 immune axis is characterized by a polarization of macrophages on the M2 phenotype and by the expansion of the T-helper 2 (Th2) population of CD4+ lymphocytes. This is accompanied by the production of a specific pattern of cytokines (e.g. IL-4, IL-5, IL-9, IL-10 and IL-13) and immunoglobulins (Ig) classes and subclasses (IgG1, IgG4 and IgE). Type 2 responses promote rapid intestinal epithelial cell turnover, mucus production and increased gut motility to encourage helminth expulsion. This can be paired with an expanded Treg population, and production of wound healing molecules to limit inflammation and promote intestinal repair. The global effect of this immune response is to create an anti-inflammatory environment, favourable to parasite survival and to the maintenance of a state of balance with the host: worms are tolerated, and host tissue damage is minimized.

At the end of the twentieth century, Hagel and colleagues correlated the absence of intestinal worms with the emergence of asthma and allergies in Venezuelans [56], not surprisingly in the light of the already discussed “dirty old friends” hypothesis and of its prediction that lack of exposure to these organisms could have some detrimental effects on the human health. Indeed, the results of this epidemiological study, showed that 90% of Venezuelan Indians living in the rainforest had worms and no allergies; on the other hand, only 10% of the Venezuelans living in the cities had worm infections, generally light, and 43% had allergies. In the following years, it was reported that increased urbanization, as well as medical and public health improvements, could be connected to the observed increase of immune-mediated diseases, including IBDs. On the contrary, it is known that the prevalence of IBDs in the poorest areas of the world is

relatively low, and this evidence is highly correlated with the prevalence of nematode infections in these areas [57,58].

Numerous studies support the hypothesis that helminths are able to induce beneficial effects to modulate IBDs, both directly and/or indirectly. Directly, by suppressing the damaging Th1 and Th17 effector cells, responsible for maintaining inflammation in IBDs [59,60]. Indirectly, by an action on gut microbiota to promote the expansion of protective bacterial taxa, able to reduce inflammatory responses [61].

Exploiting the immunomodulating "capabilities" of helminths to treat inflammatory human diseases is hence considered a promising field, and therapies using live parasitic worms, worm secretions, and worm-derived synthetic molecules to treat autoimmune and inflammatory diseases have also been attempted, or have been suggested (Wu et al., 2017 [62] for a review).

Considering specifically IBDs, whipworms of the genus *Trichuris* and hookworms of the genus *Necator* have been investigated as potential candidates for a helminth-exploiting therapy. As the pig whipworm *Trichuris suis* met the safety requirements, its use in clinical trials has been approved. Pig is the natural host for *T. suis*, but this worm can also determine short-term colonization in human gut, not causing disease [63]. The parasite can infect humans through ingestion of the eggs. Since several years, numerous studies reported the positive effects of *T. suis* viable egg ingestion in patients suffering from both CD and ulcerative colitis (UC) (see Huang et al., 2018 [64] for a recent meta-analysis). In most treated patients there was evidence for improvements in symptoms, without side-effects [65,66]. Furthermore, Sandborn and colleagues (2013) demonstrated the safety and tolerability, in clinical trials, of different doses of *T. suis* ova in patients suffering from CD, without any short or long-term adverse reactions [67]. More recently, the review of Smallwood and colleagues (2017) reports the results of several clinical trials using *T. suis* ova in oral administration; part of these studies highlights the efficacy of helminthic therapy for the treatment of UC and CD [68].

Another worm proposed as an alternative to *T. suis* is the hookworm *Necator americanus*; the immune response to hookworm infection is similar to that against whipworm, with a strong systemic and mucosal Th2 and regulatory responses [69]. The hallmarks of such strong Th2 response are the enhanced production of IL-4, IL-5, IL-9 and IL-13, while the regulatory response is due to IL-10 and transforming growth factor β (TGF-β). Unlike the whipworms, humans can get infected with *N. americanus*

larvae through skin contact with contaminated soil. A study reporting an experimental infection of humans suffering from asthma showed that small numbers of *N. americanus* worms are safe and well-tolerated by patients [70]. In CD patients, the treatment with infective larvae of *N. americanus* determined remission effects, even if associated with several side effects such as itching [71]. Recently, two human hookworm clinical trials tested the efficacy of this therapy in patients affected by celiac disease [72,73], with conflicting results. While the former paper concluded that the treatment promoted gluten tolerance, the latter reported no benefits on the pathology.

Recent investigations have reported that the immunomodulating properties of helminths can also influence the intestinal microbiota of the hosts. As discussed above, the microbiota is among the factors implied in the development of IBDs: mutualistic bacteria, along with the pathogenic ones, can induce macrophages to produce pro-inflammatory cytokines such as IL-12/23, TNF-α, IL-6, IL-1β. Activated T helper cells are then primed by these cytokines, which are considered the stimulus driving to chronic inflammation in IBD patients. In 2016, Ramanan and colleagues reported that helminths can reduce intestinal inflammatory responses by promoting the expansion of protective bacteria that inhibit the growth of proinflammatory taxa [61]. These researchers found that *Trichuris muris* infection protected mice deficient in the CD susceptibility gene *NOD2* by inhibiting the colonization of inflammatory *Bacteroides* species; this inhibition led to a shift towards a Th2 response, which in turn promoted the establishment of a protective microbiota, mainly composed by *Clostridiales*. The impact of worm infection on human gut microbiota was also investigated by Cantacessi et al. (2014) [74] who reported an increased richness and diversity of the gut microbiota in humans infected by helminths and living in endemic areas. Several other investigations linking the increase in microbial species richness to human nematode infections are available [75–77].

However, the treatments with living worms might have several disadvantages; several papers highlighted nematodes' ability to modify the homeostasis of host cells, and to increase susceptibility to oncogenic transformation by secreting molecules that interact with host cells [78,79]. For example, intestinal nematodes produce excretory-secretory (E/S) molecules that might have the capacity of inducing over-proliferation in normal intestinal epithelium [80]. Finally, Donskow-Łysoniewska and colleagues (2013) reported that, during a helminth therapy treatment of colitis, the changes in the intestinal *niche* favored the adaptation and the growth of larvae by changing their

proteome and, consequently, their immunogenicity. This was associated with a weaker immune recognition, and increased larval adaptation and worm growth, as well as with an altered localization in the intestine [81].

5. Deciphering the mechanisms: cells, molecules, receptors, and signal cascades in helminth-mediated immunomodulation

As discussed above, live nematodes and their molecules (e.g., E/S molecules) can modulate the innate and adaptive immune response of their hosts [55]. Numerous studies reported that nematodes interact and prime the activation of dendritic cells (DCs), and are also capable of influencing other immune subsets, including T cells, B cells and macrophages [82].

In IBD patients, activated DCs expressing toll-like receptors (TLRs), essential for the recognition of bacteria, as well as co-stimulatory molecules, play a key role in the induction of Th1 and Th17 responses. As reviewed in McSorley and colleagues (2013), these activated DCs might become anergic or tolerogenic following infections with nematodes [83]. For example, the parasitic nematode *Heligmosomoides polygyrus bakeri* can alter the function of these DC subsets [84], suppressing the inflammation in colitis.

Not surprisingly, different nematodes present different mechanisms involved in immune modulation. For example, during a *T. muris* infection, the suppression of IL-12/23p40 expression is dependent on the cytokine TSLP (Thymic Stromal Lymphopoietin), while during an *H. polygyrus* infection, TSLP is not necessary as there is instead a direct suppression of the expression of pro-inflammatory cytokines by DCs through E/S molecules [85].

In other cases, it has been shown that infection with nematodes such as *Trichinella spiralis* redirects the mucosal immune system to a protective Th2 response and induces the production of the anti-inflammatory cytokine IL-10 in a dose-dependent manner [86].

Concerning the induction of type 2 responses during helminth infection, a key role is exerted by tuft cells, taste-chemosensory intestinal epithelial cells that, in response to helminths, secrete IL-25. On its turn, IL-25 initiates the recruitment of Th2 lymphocytes and of group 2 innate lymphoid cells (ILC2s), and activates the cascades leading to production of the Th2 cytokines IL-4 and/or IL-13 [87]. Macrophages are then activated by these cytokines to assume a M2 phenotype with a reduced production of the

inducible nitric oxide synthase (iNOS) and of pro-inflammatory cytokines. Mediators released by ILC2s, granulocytes and natural killer T (NKT) cells also contribute to skew macrophage differentiation towards M2. Remarkably, the activation of this type 2 immune axis is also directly responsible for remodeling of the intestinal epithelium and hypercontractability of smooth muscle cells to favor worm expulsion and the maintenance of a state of balance with the host.

Concerning the Treg suppressive T cell populations, their role for the control of immune disorders and maintenance of a systemic homeostasis under inflammatory diseases is essential, and exerted through cell-cell contact and through the production of immunosuppressive cytokines (e.g. IL-10 and TGF-β) [88,89]. The transcription factor Foxp3 is fundamental for the differentiation of Tregs: only CD4+CD25+ Treg cells expressing Foxp3 exert immunomodulatory effects. Treg cells are essential for helminths to survive in hosts; indeed, it is reported that their depletion in mice lead to clearance of the infection, whereas the expansion of Tregs makes mice more susceptible [90]. Figure 3 summarizes these networks of immunomodulation dependent on helminths and their products.

Concordant data have shown that a major role in the immunomodulation induced by worms is played by their products, such as E/S molecules [91–94]. Other helminth products relevant for host immunomodulation are extracellular vesicles normally released by many parasites, with reported roles in both parasite–parasite and parasite–host interactions [95]. For example, Eichenberger et al. (2018) report that extracellular vesicles from *Nippostrongylus brasiliensis,* after intraperitoneal injection in the gut of mice, protected against colitic inflammation, with associated suppression of IL-6, IL-1β, IFNγ, and IL-17a cytokines [93].

Further nematode-derived molecules with key immunomodulatory roles are the cysteine protease inhibitors (cystatins), part of a large superfamily with diverse biological activity. Nematode cystatins inhibit the proteases involved in the antigen processing by antigen presenting cells (APCs) as macrophages and DCs, that normally would end up with presentation to (and activation of) T-helper cells. Further, cystatins modulate the immune response upregulating IL-10. For example, increased IL-10 production was observed in macrophages treated with recombinant cystatin of liver fluke *Clonorchis sinensis*, with the remission of chemically induced colitis [96]. In 2019, Xu and colleagues investigated the effect of the nematode *T. spiralis* cysteine protease inhibitors on a colitis mouse model. The authors reported a significant increase in Tregs and highlighted the mitigating role of cystatin to relieve IBDs [97].

As reported in section 4, a helminth infection, and the immune response induced by it, can alter the composition of the bacterial microbiota [98], as expected for an ecosystem, as the intestinal community, that is dynamic, responsive and characterized by interactions between distinct biological entities. For example, it is reported that the infection with the nematodes *T. muris* or *H. polygyrus* favors the expansion of protective *Clostridiales* in deference to harmful *Bacteroides;* this change in the intestinal microbiota protects the host from disease. As shown in Atarashi and colleagues (2011), the induction of a TGF-β-rich environment by a defined mix of *Clostridium* strains promotes Treg cells accumulation in the colon and the resistance to colitis [99].

Finally, other subsets, as B cells, are involved in the control of the immune response during helminth infection. Unfortunately, the role of B cells has been poorly investigated so far, as well as their potential use in a helminth-based therapy in IBDs. *Schistosoma* is an example of a parasite whose action on the B cell-mediated immunity has been studied, trying to exploit it to exert a protective role in allergic airway diseases. Results, in mouse models, showed that allergy inflammation can be suppressed by B cells isolated from *Schistosoma mansoni*–infected mice, that both produce the anti-inflammatory cytokine IL-10 and enhance Treg cells activity. Finally, in humans infected by trematode worms of the genus *Schistosoma haematobium*, the subset of IL-10 producing B cells was expanded and able to downregulate inflammatory responses mediated by T cells; besides, more Treg cells were also found associated with B cells [100].

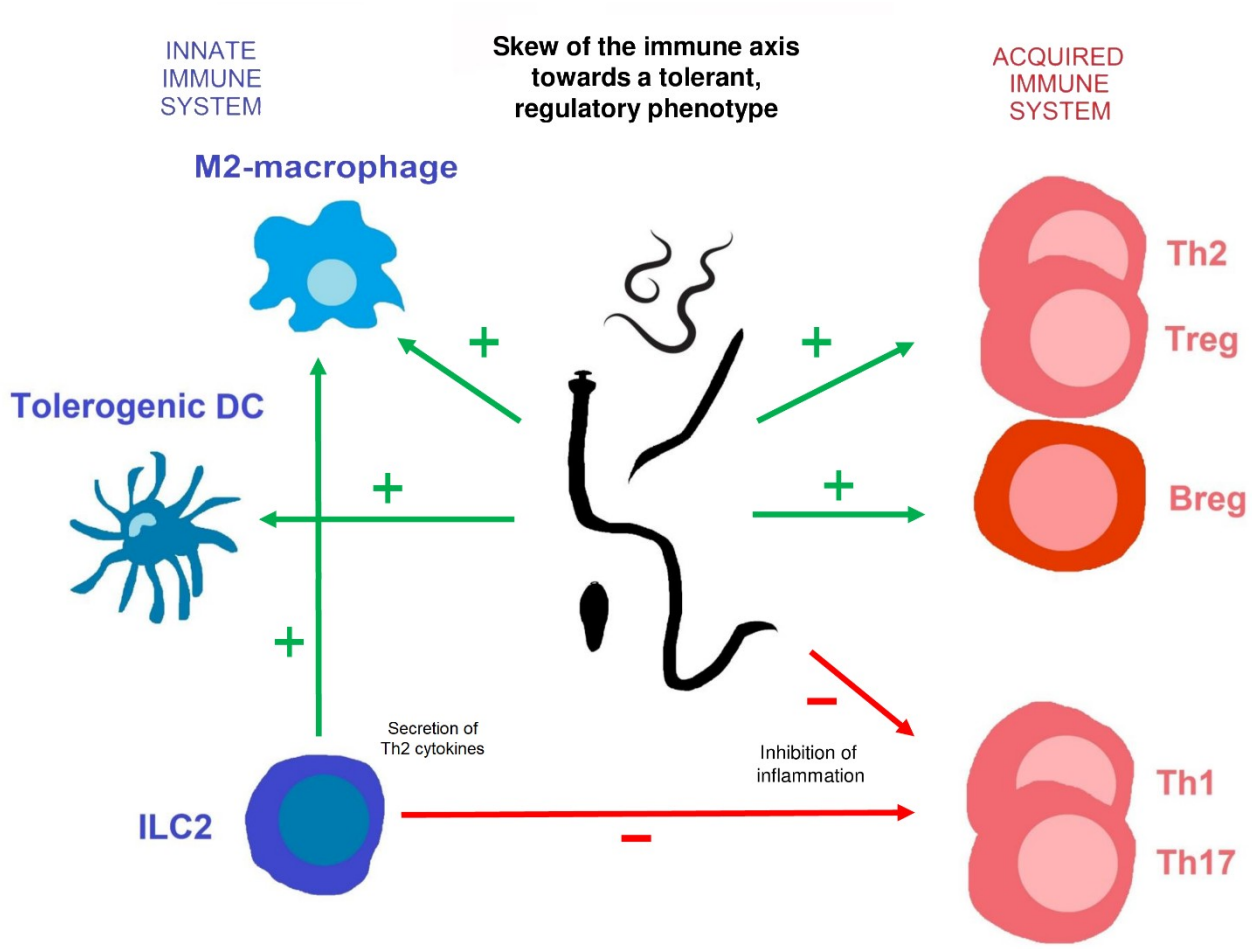


Figure 3. *Immunomodulatory effects of helminths and their products on innate and adaptive responses. Helminths promote a type 2 response through: induction of the "M2" differentiation of macrophages; induction of a tolerigenic phenotype of dendritic cells; promotion of the Th2 subset of CD3+ T-cells; promotion of the regulatory phenotypes of T- and B-cells. On the contrary, the pro-inflammatory Th1/Th17 axis is down-regulated. Type 2 Innate lymphoid cells, on their hand, secrete Th2 cytokines in response to helminth infection and promote the proliferation of M2 macrophages. These cells are also responsible for mucus accumulation and smooth muscle hypercontractility, that favour the establishment of a status of host-helminthic parasite equilibrium. List of abbreviations. DC: dendritic cells; ILC2: type 2 innate lymphoid cells; Th1: T helper 1 cells; Th2: T helper 2 cells; Th17: T helper 17 cells; Treg: regulatory T cells; Breg: regulatory B cells.*

6. Prospects: pre-biotics, post-biotics and/or novel drugs from gut inhabiting worms and fungi?

In this review, we have emphasized that complex interactions and coevolved interdependence exist among the human host, the gastrointestinal helminths, and the resident microbial communities (microbiota and mycobiota). The implications of such networks are only now starting to be elucidated in an emergent, interdisciplinary area at

the edge between microbiology and parasitology, that considers the ecosystem of the mammalian intestine on its whole [101]. A better understanding of these networks is indeed important, as all the three parties (host, worms, microbes) are directly involved in the modulation of inflammatory disorders in general, and of IBDs in particular, as widely discussed above. For example, as already discussed, the ameliorating effect of helminths and their products on inflammation is often mediated by effects on the microbiota [102]. Consequently, every attempt to improve symptoms of inflammatory disorders administering pro-, pre-, or post-biotics, or parasite-derived molecules, must carefully consider these biotic interactions and cross-talks, up to now only partially understood, and difficult to predict. For example, most studies on probiotic strains of *Lactobacillus* and *Bifidobacterium* suggest a decrease in helminths following administration; on the other hand, a number of intestinal helminth species, as the already cited *N. americanus* and *H. polygyrus*, have been shown to promote the expansion of species belonging to the genus *Lactobacillus* and of SCFA-producing taxa, (reviewed in Leung et al., 2018 [102]). To make the picture even more complex and interconnected, there are other important pieces of the puzzle that need to be considered, as host genotype and diet. In the light of these considerations, it is perhaps not surprising that the vast majority of clinical trials attempted to induce or maintain remission in IBDs, or to prevent complications, using pro-biotics have proven a limited efficacy, often with conflicting results (Damaskos and Kolios, 2008 [103]; Scaldaferri et al., 2013 [104] for reviews). Results are scarcer for pre-biotics, whose efficacy in IBDs is mostly confined to *in vitro* and animal models, with only a few human trials available (e.g., with fructooligosaccharides, FOS or inulin) often on small cohorts and again with contrasting conclusions on the clinical improvements [104]. A possibility for future studies is certainly linked to the exploration of post-biotics. Their use is still in its infancy, but deserves attention for their capacity to offer physiological benefits to the host [105].

Parasite-derived molecules stand as promising therapeutics in the treatment of several autoimmune disorders, among which IBDs, for their capacity to create a tolerogenic environment, to inhibit the progression of inflammation, and consequently alleviate symptoms, by acting on the immune subsets and through the molecular mechanisms discussed above. Several parasite-derived antigens have been tested in animal models, with promising outcomes (reviewed in Wu et al., 2017 [62]). One remarkable example is ES-62, a major E/S product of the rodent filarial nematode *Acanthocheilonema vitae*, known to induce a strong Th2 response via interaction with

TLR4, to block mast cell degranulation inhibiting the release of prostaglandins and leukotrienes, and to down-regulate complement activation. Animal experiments have shown protective effects exerted in various autoimmune diseases (rheumatoid arthritis, lupus erythematosus, airway hyperresponsiveness). Recently, small-molecule analogues of ES-62 were identified that presented similar modulatory effects, without the immunogenicity that instead characterizes ES-62 [62]. Another example, with promising implications for IBDs is represented by the antigen HdAg from *Hymenolepis diminuita*, an intestinal tapeworm. There is evidence that this antigen ameliorates the symptoms of dextrane sodium sulfate (DSS)-induced colitis in mice, a model for human IBDs. HdAg increases the levels of two key cytokines with anti-colitic activity (IL-10 and IL-4) and inhibits the progression of colitis [106]. Other parasite-derived compounds that ameliorated symptoms in animal models of colitis comprehend the E/S L1 product from *Trichinella spiralis* and the already mentioned cystatins, interfering with antigen processing and produced by many parasites, among which the filariae *A. vitae* and *Onchocerca volvulus* [62].

Another source of anti-inflammatory molecules is represented by fungal metabolites such as polysaccharides (glucans *in primis*), phenolic and indolic compounds, proteins, mycosteroids, fatty acids, and carotenoids. Since centuries, mushrooms are used for their medicinal properties, as in the case of *Ganoderma lucidum* in traditional Chinese medicine. Modern research has corroborated the therapeutic effects of traditionally used species. Several fungal molecules can influence the proliferation and differentiation of lymphocyte subsets, regulate cytokine release, suppress pro-inflammatory proteins as cycloossigenase 2, iNOS and IL-1 [107]. In some cases, molecular pathways are beginning to be elucidated and, as above reported for parasitic molecules, data from animal models are accumulating that point to anti-inflammatory activities of fungal (and also plants) metabolites in several contexts (e.g., in cases of oedema, models of arthritis, heart disorders caused by chronic inflammation; reviewed in Toghueo, 2019 [108]). Still, there are scarce data on the administration of these molecules to humans: from parasites and fungi to pills there is still a long way to go.

7. Concluding remarks: knowledge, or translation knowledge?

There is now convincing evidence that a correct shaping of the immune balance requires some type of contact with parasites, and with a varied microcosm of gut-dwelling microorganisms, or perhaps even with just transient, but frequent, “gut-passengers”. The so called dirty old friends. It is otherwise clear that we cannot restore

a “non-hygienic” lifestyle, and it is also difficult to imagine an extensive use of worm therapies, in the form of true infections or administration of crude preparations of entire worms, or their eggs. However, the development of the hygiene hypothesis, in the context of the evolutionary mismatch concept, revitalized the study of parasite and fungal immunology, not only with the scope to reconstruct the pathophysiology of parasitic and mycotic diseases, but also to uncover how these organisms modulate the host immunity, and how this can fit with the idea that these organisms once protected humans from chronic inflammation. Molecular mechanisms in parasite-host interactions, and the cascades in immune signal transduction, are now being unraveled; similarly, molecules with therapeutic potential are being discovered from parasites and fungi, with identification of receptor cells, and a description of their effects in animal models. Thus, the prospect is not to develop novel ways to administer living worms to cure IBDs, but instead to derive novel drugs from worms, once the molecular mechanisms underlying the anti-inflammatory effects of these parasites are elucidated. As discussed, a few molecules, which could be classified as post-biotics, have already been tested in mouse models, for the cure of chemically-induced colitis. The novel discoveries on the role of tuft cells as receptors for nematode presence, and their role in inducing a Th2 response and a reshaping of the gut bacterial microbiota, open a new perspective toward the possibility to derive compound from worms, that could prevent or cure IBDs. We are still far from applications, but the prospects are encouraging, for a future in which molecules derived from parasites and fungi will vicariate the disappearance of our dirty old friends, helping us in the maintenance of a healthy gut.

References

[1] P.D. Gluckman, M.A. Hanson, F.M. Low, Evolutionary and developmental mismatches are consequences of adaptive developmental plasticity in humans and have implications for later disease risk, Philos. Trans. R. Soc. Lond. B Biol. Sci. 374 (2019) 20180109. https://doi.org/10.1098/rstb.2018.0109.

[2] F. Bordes, S. Morand, The impact of multiple infections on wild animal hosts: a review, Infect. Ecol. Epidemiol. 1 (2011). https://doi.org/10.3402/iee.v1i0.7346.

[3] D. Lieberman, The Story of the Human Body: Evolution, Health, and Disease, Pantheon Books, 2013. https://books.google.it/books?id=Ht_zugAACAAJ.

[4] J.M. Diamond, Guns, Germs, and Steel: The Fates of Human Societies, W.W. Norton, 1997.

[5] I. Kuijt, N. Goring-Morris, Foraging, Farming, and Social Complexity in the Pre-Pottery Neolithic of the Southern Levant: A Review and Synthesis, Journal of World Prehistory. 16 (2002) 361–440. https://doi.org/10.1023/A:1022973114090.

[6] G. Fields, Urbanization and the Transition from Agrarian to Industrial Society, Berkeley Planning Journal. 13 (1999). https://doi.org/10.5070/BP313113032.

[7] A.N. Crittenden, S.L. Schnorr, Current views on hunter-gatherer nutrition and the evolution of the human diet, Am. J. Phys. Anthropol. 162 Suppl 63 (2017) 84–109. https://doi.org/10.1002/ajpa.23148.

[8] W.P.T. James, R.J. Johnson, J.R. Speakman, D.C. Wallace, G. Frühbeck, P.O. Iversen, P.J. Stover, Nutrition and its role in human evolution, J. Intern. Med. 285 (2019) 533–549. https://doi.org/10.1111/joim.12878.

[9] D.P. Strachan, Hay fever, hygiene, and household size, BMJ. 299 (1989) 1259–1260. https://doi.org/10.1136/bmj.299.6710.1259.

[10] J.-F. Bach, The hygiene hypothesis in autoimmunity: the role of pathogens and commensals, Nat. Rev. Immunol. 18 (2018) 105–120. https://doi.org/10.1038/nri.2017.111.

[11] J.K. Goodrich, J.L. Waters, A.C. Poole, J.L. Sutter, O. Koren, R. Blekhman, M. Beaumont, W. Van Treuren, R. Knight, J.T. Bell, T.D. Spector, A.G. Clark, R.E. Ley, Human genetics shape the gut microbiome, Cell. 159 (2014) 789–799. https://doi.org/10.1016/j.cell.2014.09.053.

[12] A.D. Kostic, R.J. Xavier, D. Gevers, The microbiome in inflammatory bowel disease: current status and the future ahead, Gastroenterology. 146 (2014) 1489–1499. https://doi.org/10.1053/j.gastro.2014.02.009.

[13] F. Imhann, A. Vich Vila, M.J. Bonder, J. Fu, D. Gevers, M.C. Visschedijk, L.M. Spekhorst, R. Alberts, L. Franke, H.M. van Dullemen, R.W.F. Ter Steege, C. Huttenhower, G. Dijkstra, R.J. Xavier, E.A.M. Festen, C. Wijmenga, A. Zhernakova, R.K. Weersma, Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease, Gut. 67 (2018) 108–119. https://doi.org/10.1136/gutjnl-2016-312135.

[14] J. Ni, G.D. Wu, L. Albenberg, V.T. Tomov, Gut microbiota and IBD: causation or correlation?, Nat. Rev. Gastroenterol. Hepatol. 14 (2017) 573–584. https://doi.org/10.1038/nrgastro.2017.88.

[15] S.B. Andersen, B.J. Shapiro, C. Vandenbroucke-Grauls, M.G.J. de Vos, Microbial evolutionary medicine: from theory to clinical practice, Lancet Infect. Dis. 19 (2019) e273–e283. https://doi.org/10.1016/S1473-3099(19)30045-3.

[16] M. Matijašić, T. Meštrović, H.Č. Paljetak, M. Perić, A. Barešić, D. Verbanac, Gut Microbiota beyond Bacteria-Mycobiome, Virome, Archaeome, and Eukaryotic Parasites in IBD, Int. J. Mol. Sci. 21 (2020). https://doi.org/10.3390/ijms21082668.

[17] E. Rinninella, P. Raoul, M. Cintoni, F. Franceschi, G.A.D. Miggiano, A. Gasbarrini, M.C. Mele, What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases, Microorganisms. 7 (2019). https://doi.org/10.3390/microorganisms7010014.

[18] C.A. Lozupone, J.I. Stombaugh, J.I. Gordon, J.K. Jansson, R. Knight, Diversity, stability and resilience of the human gut microbiota, Nature. 489 (2012) 220–230. https://doi.org/10.1038/nature11550.

[19] S.V. Lynch, O. Pedersen, The Human Intestinal Microbiome in Health and Disease, N. Engl. J. Med. 375 (2016) 2369–2379. https://doi.org/10.1056/NEJMra1600266.

[20] S. Selber-Hnatiw, B. Rukundo, M. Ahmadi, H. Akoubi, H. Al-Bizri, A.F. Aliu, T.U. Ambeaghen, L. Avetisyan, I. Bahar, A. Baird, F. Begum, H. Ben Soussan, V. Blondeau-Éthier, R. Bordaries, H. Bramwell, A. Briggs, R. Bui, M. Carnevale, M. Chancharoen, T. Chevassus, J.H. Choi, K. Coulombe, F. Couvrette, S. D'Abreau, M. Davies, M.-P. Desbiens, T. Di Maulo, S.-A. Di Paolo, S. Do Ponte, P. Dos Santos Ribeiro, L.-A. Dubuc-Kanary, P.K. Duncan, F. Dupuis, S. El-Nounou, C.N. Eyangos, N.K. Ferguson, N.R. Flores-Chinchilla, T. Fotakis, M. Gado Oumarou H D, M. Georgiev, S. Ghiassy, N. Glibetic, J. Grégoire Bouchard, T. Hassan, I. Huseen, M.-F. Ibuna Quilatan, T. Iozzo, S. Islam, D.B. Jaunky, A. Jeyasegaram, M.-A. Johnston, M.R. Kahler, K. Kaler, C. Kamani, H. Karimian Rad, E. Konidis, F. Konieczny, S. Kurianowicz, P. Lamothe, K. Legros, S. Leroux, J. Li, M.E. Lozano Rodriguez, S. Luponio-Yoffe, Y. Maalouf, J.

Mantha, M. McCormick, P. Mondragon, T. Narayana, E. Neretin, T.T.T. Nguyen, I. Niu, R.B. Nkemazem, M. O'Donovan, M. Oueis, S. Paquette, N. Patel, E. Pecsi, J. Peters, A. Pettorelli, C. Poirier, V.R. Pompa, H. Rajen, R.-O. Ralph, J. Rosales-Vasquez, D. Rubinshtein, S. Sakr, M.S. Sebai, L. Serravalle, F. Sidibe, A. Sinnathurai, D. Soho, A. Sundarakrishnan, V. Svistkova, T.E. Ugbeye, M.S. Vasconcelos, M. Vincelli, O. Voitovich, P. Vrabel, L. Wang, M. Wasfi, C.Y. Zha, C. Gamberi, Human Gut Microbiota: Toward an Ecology of Disease, Front. Microbiol. 8 (2017) 1265. https://doi.org/10.3389/fmicb.2017.01265.

[21] C.P. Tamboli, C. Neut, P. Desreumaux, J.F. Colombel, Dysbiosis in inflammatory bowel disease, Gut. 53 (2004) 1–4. https://doi.org/10.1136/gut.53.1.1.

[22] G.R. Gibson, J.H. Cummings, G.T. Macfarlane, Growth and activities of sulphate-reducing bacteria in gut contents of healthy subjects and patients with ulcerative colitis, FEMS Microbiology Letters. 86 (1991) 103–111. https://doi.org/10.1111/j.1574-6941.1991.tb01742.x.

[23] I. Kushkevych, D. Dordević, P. Kollar, M. Vítězová, L. Drago, Hydrogen Sulfide as a Toxic Product in the Small-Large Intestine Axis and its Role in IBD Development, J. Clin. Med. Res. 8 (2019). https://doi.org/10.3390/jcm8071054.

[24] S.-L. Wang, Z.-R. Wang, C.-Q. Yang, Meta-analysis of broad-spectrum antibiotic therapy in patients with active inflammatory bowel disease, Exp. Ther. Med. 4 (2012) 1051–1056. https://doi.org/10.3892/etm.2012.718.

[25] M. Zhou, J. He, Y. Shen, C. Zhang, J. Wang, Y. Chen, New Frontiers in Genetics, Gut Microbiota, and Immunity: A Rosetta Stone for the Pathogenesis of Inflammatory Bowel Disease, Biomed Res. Int. 2017 (2017) 8201672. https://doi.org/10.1155/2017/8201672.

[26] E.A. Franzosa, A. Sirota-Madi, J. Avila-Pacheco, N. Fornelos, H.J. Haiser, S. Reinker, T. Vatanen, A.B. Hall, H. Mallick, L.J. McIver, J.S. Sauk, R.G. Wilson, B.W. Stevens, J.M. Scott, K. Pierce, A.A. Deik, K. Bullock, F. Imhann, J.A. Porter, A. Zhernakova, J. Fu, R.K. Weersma, C. Wijmenga, C.B. Clish, H. Vlamakis, C. Huttenhower, R.J. Xavier, Gut microbiome structure and metabolic activity in inflammatory bowel disease, Nat Microbiol. 4 (2019) 293–305. https://doi.org/10.1038/s41564-018-0306-4.

[27] J. Lloyd-Price, C. Arze, A.N. Ananthakrishnan, M. Schirmer, J. Avila-Pacheco, T.W. Poon, E. Andrews, N.J. Ajami, K.S. Bonham, C.J. Brislawn, D. Casero, H. Courtney, A. Gonzalez, T.G. Graeber, A.B. Hall, K. Lake, C.J. Landers, H. Mallick, D.R. Plichta, M. Prasad, G. Rahnavard, J. Sauk, D. Shungin, Y. Vázquez-Baeza, R.A. White 3rd, IBDMDB Investigators, J. Braun, L.A. Denson, J.K. Jansson, R. Knight, S. Kugathasan,

D.P.B. McGovern, J.F. Petrosino, T.S. Stappenbeck, H.S. Winter, C.B. Clish, E.A. Franzosa, H. Vlamakis, R.J. Xavier, C. Huttenhower, Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases, Nature. 569 (2019) 655–662. https://doi.org/10.1038/s41586-019-1237-9.

[28] J.D. Doecke, L.A. Simms, Z.Z. Zhao, N. Huang, K. Hanigan, K. Krishnaprasad, R.L. Roberts, J.M. Andrews, G. Mahy, P. Bampton, P. Lewindon, T. Florin, I.C. Lawrance, R.B. Gearry, G.W. Montgomery, G.L. Radford-Smith, Genetic susceptibility in IBD: overlap between ulcerative colitis and Crohn's disease, Inflamm. Bowel Dis. 19 (2013) 240–245. https://doi.org/10.1097/MIB.0b013e3182810041.

[29] S. Yamamoto, X. Ma, Role of Nod2 in the development of Crohn's disease, Microbes Infect. 11 (2009) 912–918. https://doi.org/10.1016/j.micinf.2009.06.005.

[30] M.J. Suhr, H.E. Hallen-Adams, The human gut mycobiome: pitfalls and potentials--a mycologist's perspective, Mycologia. 107 (2015) 1057–1073. https://doi.org/10.3852/15-147.

[31] C. Hoffmann, S. Dollive, S. Grunberg, J. Chen, H. Li, G.D. Wu, J.D. Lewis, F.D. Bushman, Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents, PLoS One. 8 (2013) e66019. https://doi.org/10.1371/journal.pone.0066019.

[32] A.K. Nash, T.A. Auchtung, M.C. Wong, D.P. Smith, J.R. Gesell, M.C. Ross, C.J. Stewart, G.A. Metcalf, D.M. Muzny, R.A. Gibbs, N.J. Ajami, J.F. Petrosino, The gut mycobiome of the Human Microbiome Project healthy cohort, Microbiome. 5 (2017) 153. https://doi.org/10.1186/s40168-017-0373-4.

[33] T.A. Auchtung, T.Y. Fofanova, C.J. Stewart, A.K. Nash, M.C. Wong, J.R. Gesell, J.M. Auchtung, N.J. Ajami, J.F. Petrosino, Investigating Colonization of the Healthy Adult Gastrointestinal Tract by Fungi, mSphere. 3 (2018). https://doi.org/10.1128/mSphere.00092-18.

[34] J.N. Witchley, P. Penumetcha, N.V. Abon, C.A. Woolford, A.P. Mitchell, S.M. Noble, Candida albicans Morphogenesis Programs Control the Balance between Gut Commensalism and Invasive Infection, Cell Host Microbe. 25 (2019) 432–443.e6. https://doi.org/10.1016/j.chom.2019.02.008.

[35] M.L. Richard, H. Sokol, The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases, Nat. Rev. Gastroenterol. Hepatol. 16 (2019) 331–345. https://doi.org/10.1038/s41575-019-0121-2.

[36] H. Sokol, V. Leducq, H. Aschard, H.-P. Pham, S. Jegou, C. Landman, D. Cohen, G. Liguori, A. Bourrier, I. Nion-Larmurier, J. Cosnes, P. Seksik, P. Langella, D. Skurnik,

M.L. Richard, L. Beaugerie, Fungal microbiota dysbiosis in IBD, Gut. 66 (2017) 1039–1048. https://doi.org/10.1136/gutjnl-2015-310746.

[37] G. Hoarau, P.K. Mukherjee, C. Gower-Rousseau, C. Hager, J. Chandra, M.A. Retuerto, C. Neut, S. Vermeire, J. Clemente, J.F. Colombel, H. Fujioka, D. Poulain, B. Sendid, M.A. Ghannoum, Bacteriome and Mycobiome Interactions Underscore Microbial Dysbiosis in Familial Crohn's Disease, MBio. 7 (2016). https://doi.org/10.1128/mBio.01250-16.

[38] A. Standaert-Vitse, T. Jouault, P. Vandewalle, C. Mille, M. Seddik, B. Sendid, J.-M. Mallet, J.-F. Colombel, D. Poulain, Candida albicans is an immunogen for anti-Saccharomyces cerevisiae antibody markers of Crohn's disease, Gastroenterology. 130 (2006) 1764–1775. https://doi.org/10.1053/j.gastro.2006.02.009.

[39] I. Leonardi, X. Li, A. Semon, D. Li, I. Doron, G. Putzel, A. Bar, D. Prieto, M. Rescigno, D.P.B. McGovern, J. Pla, I.D. Iliev, CX3CR1 mononuclear phagocytes control immunity to intestinal fungi, Science. 359 (2018) 232–236. https://doi.org/10.1126/science.aao1503.

[40] J.J. Limon, J. Tang, D. Li, A.J. Wolf, K.S. Michelsen, V. Funari, M. Gargus, C. Nguyen, P. Sharma, V.I. Maymi, I.D. Iliev, J.H. Skalski, J. Brown, C. Landers, J. Borneman, J. Braun, S.R. Targan, D.P.B. McGovern, D.M. Underhill, Malassezia Is Associated with Crohn's Disease and Exacerbates Colitis in Mouse Models, Cell Host Microbe. 25 (2019) 377–388.e6. https://doi.org/10.1016/j.chom.2019.01.007.

[41] K. Trunk, J. Peltier, Y.-C. Liu, B.D. Dill, L. Walker, N.A.R. Gow, M.J.R. Stark, J. Quinn, H. Strahl, M. Trost, S.J. Coulthurst, The type VI secretion system deploys antifungal effectors against microbial competitors, Nat Microbiol. 3 (2018) 920–931. https://doi.org/10.1038/s41564-018-0191-x.

[42] H.A. Filyk, L.C. Osborne, The Multibiome: The Intestinal Ecosystem's Influence on Immune Homeostasis, Health, and Disease, EBioMedicine. 13 (2016) 46–54. https://doi.org/10.1016/j.ebiom.2016.10.007.

[43] B. Sovran, J. Planchais, S. Jegou, M. Straube, B. Lamas, J.M. Natividad, A. Agus, L. Dupraz, J. Glodt, G. Da Costa, M.-L. Michel, P. Langella, M.L. Richard, H. Sokol, Enterobacteriaceae are essential for the modulation of colitis severity by fungi, Microbiome. 6 (2018) 152. https://doi.org/10.1186/s40168-018-0538-9.

[44] J. Miyoshi, M.A. Sofia, J.F. Pierre, The evidence for fungus in Crohn's disease pathogenesis, Clin. J. Gastroenterol. 11 (2018) 449–456. https://doi.org/10.1007/s12328-018-0886-9.

[45] S.-C. Cheng, F.L. van de Veerdonk, M. Lenardon, M. Stoffels, T. Plantinga, S. Smeekens, L. Rizzetto, L. Mukaremera, K. Preechasuth, D. Cavalieri, T.D. Kanneganti, J.W.M. van der Meer, B.J. Kullberg, L.A.B. Joosten, N.A.R. Gow, M.G. Netea, The dectin-1/inflammasome pathway is responsible for the induction of protective T-helper 17 responses that discriminate between yeasts and hyphae of Candida albicans, J. Leukoc. Biol. 90 (2011) 357–366. https://doi.org/10.1189/jlb.1210702.

[46] S. LeibundGut-Landmann, O. Gross, M.J. Robinson, F. Osorio, E.C. Slack, S.V. Tsoni, E. Schweighoffer, V. Tybulewicz, G.D. Brown, J. Ruland, C. Reis e Sousa, Syk- and CARD9-dependent coupling of innate immunity to the induction of T helper cells that produce interleukin 17, Nat. Immunol. 8 (2007) 630–638. https://doi.org/10.1038/ni1460.

[47] M. Beaudoin, P. Goyette, G. Boucher, K.S. Lo, M.A. Rivas, C. Stevens, A. Alikashani, M. Ladouceur, D. Ellinghaus, L. Törkvist, G. Goel, C. Lagacé, V. Annese, A. Bitton, J. Begun, S.R. Brant, F. Bresso, J.H. Cho, R.H. Duerr, J. Halfvarson, D.P.B. McGovern, G. Radford-Smith, S. Schreiber, P.L. Schumm, Y. Sharma, M.S. Silverberg, R.K. Weersma, Quebec IBD Genetics Consortium, NIDDK IBD Genetics Consortium, International IBD Genetics Consortium, M. D'Amato, S. Vermeire, A. Franke, G. Lettre, R.J. Xavier, M.J. Daly, J.D. Rioux, Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis, PLoS Genet. 9 (2013) e1003723. https://doi.org/10.1371/journal.pgen.1003723.

[48] M.G. Netea, L.A.B. Joosten, E. Latz, K.H.G. Mills, G. Natoli, H.G. Stunnenberg, L.A.J. O'Neill, R.J. Xavier, Trained immunity: A program of innate immune memory in health and disease, Science. 352 (2016) aaf1098. https://doi.org/10.1126/science.aaf1098.

[49] N. Whibley, J.R. Jaycox, D. Reid, A.V. Garg, J.A. Taylor, C.J. Clancy, M.H. Nguyen, P.S. Biswas, M.J. McGeachy, G.D. Brown, S.L. Gaffen, Delinking CARD9 and IL-17: CARD9 Protects against Candida tropicalis Infection through a TNF-α-Dependent, IL-17-Independent Mechanism, J. Immunol. 195 (2015) 3781–3792. https://doi.org/10.4049/jimmunol.1500870.

[50] C.O. Maher, K. Dunne, R. Comerford, S. O'Dea, A. Loy, J. Woo, T.R. Rogers, F. Mulcahy, P.J. Dunne, D.G. Doherty, Candida albicans stimulates IL-23 release by human dendritic cells and downstream IL-17 secretion by Vδ1 T cells, J. Immunol. 194 (2015) 5953–5960. https://doi.org/10.4049/jimmunol.1403066.

[51] A.C. Ford, L. Peyrin-Biroulet, Opportunistic infections with anti-tumor necrosis factor-α therapy in inflammatory bowel disease: meta-analysis of randomized controlled trials, Am. J. Gastroenterol. 108 (2013) 1268–1276. https://doi.org/10.1038/ajg.2013.138.

[52] J.A. Jackson, I.M. Friberg, S. Little, J.E. Bradley, Review series on helminths, immune modulation and the hygiene hypothesis: immunity against helminths and immunological phenomena in modern human populations: coevolutionary legacies?, Immunology. 126 (2009) 18–27. https://doi.org/10.1111/j.1365-2567.2008.03010.x.

[53] World Health Organization, Soil-transmitted helminth infections, Https://www.who.int. (2020). https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections (accessed May 18, 2020).

[54] CDC-Centers for Disease Control, CDC - Ascariasis, Https://www.cdc.gov. (2010). https://www.cdc.gov/parasites/ascariasis/index.html (accessed May 18, 2020).

[55] R.M. Maizels, A. Balic, N. Gomez-Escobar, M. Nair, M.D. Taylor, J.E. Allen, Helminth parasites--masters of regulation, Immunol. Rev. 201 (2004) 89–116. https://doi.org/10.1111/j.0105-2896.2004.00191.x.

[56] I. Hagel, N.R. Lynch, M.C. DiPrisco, R.I. Lopez, N.M. Garcia, Allergic reactivity of children of different socioeconomic levels in tropical populations, Int. Arch. Allergy Immunol. 101 (1993) 209–214. https://doi.org/10.1159/000236521.

[57] D.E. Elliott, J.R. Urban JF, C.K. Argo, J.V. Weinstock, Does the failure to acquire helminthic parasites predispose to Crohn's disease?, FASEB J. 14 (2000) 1848–1855. https://doi.org/10.1096/fj.99-0885hyp.

[58] P. Zaccone, Z. Fehervari, J.M. Phillips, D.W. Dunne, A. Cooke, Parasitic worms and inflammatory diseases, Parasite Immunol. 28 (2006) 515–523. https://doi.org/10.1111/j.1365-3024.2006.00879.x.

[59] A.R. Khan, P.G. Fallon, Helminth therapies: translating the unknown unknowns to known knowns, Int. J. Parasitol. 43 (2013) 293–299. https://doi.org/10.1016/j.ijpara.2012.12.002.

[60] M. Heylen, N.E. Ruyssers, E.M. Gielis, E. Vanhomwegen, P.A. Pelckmans, T.G. Moreels, J.G. De Man, B.Y. De Winter, Of worms, mice and man: an overview of experimental and clinical helminth-based therapy for inflammatory bowel disease, Pharmacol. Ther. 143 (2014) 153–167. https://doi.org/10.1016/j.pharmthera.2014.02.011.

[61] D. Ramanan, R. Bowcutt, S.C. Lee, M.S. Tang, Z.D. Kurtz, Y. Ding, K. Honda, W.C. Gause, M.J. Blaser, R.A. Bonneau, Y.A.L. Lim, P. 'ng Loke, K. Cadwell, Helminth infection promotes colonization resistance via type 2 immunity, Science. 352 (2016) 608–612. https://doi.org/10.1126/science.aaf3229.

[62] Z. Wu, L. Wang, Y. Tang, X. Sun, Parasite-Derived Proteins for the Treatment of Allergies and Autoimmune Diseases, Front. Microbiol. 8 (2017) 2164. https://doi.org/10.3389/fmicb.2017.02164.

[63] H. Helmby, Human helminth therapy to treat inflammatory disorders - where do we stand?, BMC Immunol. 16 (2015) 12. https://doi.org/10.1186/s12865-015-0074-3.

[64] X. Huang, L.-R. Zeng, F.-S. Chen, J.-P. Zhu, M.-H. Zhu, Trichuris suis ova therapy in inflammatory bowel disease: A meta-analysis, Medicine . 97 (2018) e12087. https://doi.org/10.1097/MD.0000000000012087.

[65] R.W. Summers, D.E. Elliott, J.F. Urban Jr, R. Thompson, J.V. Weinstock, Trichuris suis therapy in Crohn's disease, Gut. 54 (2005) 87–90. https://doi.org/10.1136/gut.2004.041749.

[66] R.W. Summers, D.E. Elliott, J.V. Weinstock, Is there a role for helminths in the therapy of inflammatory bowel disease?, Nat. Clin. Pract. Gastroenterol. Hepatol. 2 (2005) 62–63. https://doi.org/10.1038/ncpgasthep0087.

[67] W.J. Sandborn, D.E. Elliott, J. Weinstock, R.W. Summers, A. Landry-Wheeler, N. Silver, M.D. Harnett, S.B. Hanauer, Randomised clinical trial: the safety and tolerability of Trichuris suis ova in patients with Crohn's disease, Aliment. Pharmacol. Ther. 38 (2013) 255–263. https://doi.org/10.1111/apt.12366.

[68] T.B. Smallwood, P.R. Giacomin, A. Loukas, J.P. Mulvenna, R.J. Clark, J.J. Miles, Helminth Immunomodulation in Autoimmune Disease, Front. Immunol. 8 (2017) 453. https://doi.org/10.3389/fimmu.2017.00453.

[69] S. Gaze, H.J. McSorley, J. Daveson, D. Jones, J.M. Bethony, L.M. Oliveira, R. Speare, J.S. McCarthy, C.R. Engwerda, J. Croese, A. Loukas, Characterising the mucosal and systemic immune responses to experimental human hookworm infection, PLoS Pathog. 8 (2012) e1002520. https://doi.org/10.1371/journal.ppat.1002520.

[70] K. Mortimer, A. Brown, J. Feary, C. Jagger, S. Lewis, M. Antoniak, D. Pritchard, J. Britton, Dose-ranging study for trials of therapeutic infection with Necator americanus in humans, Am. J. Trop. Med. Hyg. 75 (2006) 914–920. https://www.ncbi.nlm.nih.gov/pubmed/17123987.

[71] J. Croese, J. O'neil, J. Masson, S. Cooke, W. Melrose, D. Pritchard, R. Speare, A proof of concept study establishing Necator americanus in Crohn's patients and reservoir donors, Gut. 55 (2006) 136–137. https://doi.org/10.1136/gut.2005.079129.

[72] J. Croese, P. Giacomin, S. Navarro, A. Clouston, L. McCann, A. Dougall, I. Ferreira, A. Susianto, P. O'Rourke, M. Howlett, J. McCarthy, C. Engwerda, D. Jones, A. Loukas, Experimental hookworm infection and gluten microchallenge promote tolerance in

celiac disease, J. Allergy Clin. Immunol. 135 (2015) 508–516. https://doi.org/10.1016/j.jaci.2014.07.022.

[73] A.J. Daveson, D.M. Jones, S. Gaze, H. McSorley, A. Clouston, A. Pascoe, S. Cooke, R. Speare, G.A. Macdonald, R. Anderson, J.S. McCarthy, A. Loukas, J. Croese, Effect of hookworm infection on wheat challenge in celiac disease--a randomised double-blinded placebo controlled trial, PLoS One. 6 (2011) e17366. https://doi.org/10.1371/journal.pone.0017366.

[74] C. Cantacessi, P. Giacomin, J. Croese, M. Zakrzewski, J. Sotillo, L. McCann, M.J. Nolan, M. Mitreva, L. Krause, A. Loukas, Impact of experimental hookworm infection on the human gut microbiota, J. Infect. Dis. 210 (2014) 1431–1434. https://doi.org/10.1093/infdis/jiu256.

[75] R.W. Li, S. Wu, W. Li, K. Navarro, R.D. Couch, D. Hill, J.F. Urban Jr, Alterations in the porcine colon microbiota induced by the gastrointestinal nematode Trichuris suis, Infect. Immun. 80 (2012) 2150–2157. https://doi.org/10.1128/IAI.00141-12.

[76] S.C. Lee, M.S. Tang, Y.A.L. Lim, S.H. Choy, Z.D. Kurtz, L.M. Cox, U.M. Gundra, I. Cho, R. Bonneau, M.J. Blaser, K.H. Chua, P. 'ng Loke, Helminth colonization is associated with increased diversity of the gut microbiota, PLoS Negl. Trop. Dis. 8 (2014) e2880. https://doi.org/10.1371/journal.pntd.0002880.

[77] I. Martin, M.M.M. Kaisar, A.E. Wiria, F. Hamid, Y. Djuardi, E. Sartono, B.A. Rosa, M. Mitreva, T. Supali, J.J. Houwing-Duistermaat, M. Yazdanbakhsh, L.J. Wammes, The Effect of Gut Microbiome Composition on Human Immune Responses: An Exploration of Interference by Helminth Infections, Front. Genet. 10 (2019) 1028. https://doi.org/10.3389/fgene.2019.01028.

[78] K. Donskow, N. Drela, M. Doligalska, Heligmosomoides bakeri antigen rescues CD4-positive T cells from glucocorticoid-induced apoptosis by Bcl-2 protein expression, Parasite Immunol. 33 (2011) 158–169. https://doi.org/10.1111/j.1365-3024.2010.01262.x.

[79] K. Donskow-Łysoniewska, K. Brodaczewska, M. Doligalska, Heligmosomoides polygyrus antigens inhibit the intrinsic pathway of apoptosis by overexpression of survivin and Bcl-2 protein in CD4 T cells, Prion. 7 (2013) 319–327. https://doi.org/10.4161/pri.25008.

[80] F. Huby, H. Hoste, S. Mallet, S. Fournel, J.L. Nano, Effects of the excretory/secretory products of six nematode species, parasites of the digestive tract, on the proliferation of HT29-D4 and HGT-1 cell lines, Epithelial Cell Biol. 4 (1995) 156–162. https://www.ncbi.nlm.nih.gov/pubmed/9439903.

[81] K. Donskow-Łysoniewska, J. Bien, K. Brodaczewska, K. Krawczak, M. Doligalska, Colitis promotes adaptation of an intestinal nematode: a Heligmosomoides polygyrus mouse model system, PLoS One. 8 (2013) e78034. https://doi.org/10.1371/journal.pone.0078034.

[82] A.S. MacDonald, R.M. Maizels, Alarming dendritic cells for Th2 induction, J. Exp. Med. 205 (2008) 13–17. https://doi.org/10.1084/jem.20072665.

[83] H.J. McSorley, J.P. Hewitson, R.M. Maizels, Immunomodulation by helminth parasites: defining mechanisms and mediators, Int. J. Parasitol. 43 (2013) 301–310. https://doi.org/10.1016/j.ijpara.2012.11.011.

[84] A.M. Blum, L. Hang, T. Setiawan, J.P. Urban Jr, K.M. Stoyanoff, J. Leung, J.V. Weinstock, Heligmosomoides polygyrus bakeri induces tolerogenic dendritic cells that block colitis and prevent antigen-specific gut T cell responses, J. Immunol. 189 (2012) 2512–2520. https://doi.org/10.4049/jimmunol.1102892.

[85] J.C. Massacand, R.C. Stettler, R. Meier, N.E. Humphreys, R.K. Grencis, B.J. Marsland, N.L. Harris, Helminth products bypass the need for TSLP in Th2 immune responses by directly modulating dendritic cell function, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 13968–13973. https://doi.org/10.1073/pnas.0906367106.

[86] Y. Zhao, M.Y. Liu, X.L. Wang, X.L. Liu, Y. Yang, H.B. Zou, S.M. Sun, L. Yu, B. Rosenthal, H.N. Shi, P. Boireau, X.P. Wu, Modulation of inflammatory bowel disease in a mouse model following infection with Trichinella spiralis, Vet. Parasitol. 194 (2013) 211–216. https://doi.org/10.1016/j.vetpar.2013.01.058.

[87] M.R. Howitt, S. Lavoie, M. Michaud, A.M. Blum, S.V. Tran, J.V. Weinstock, C.A. Gallini, K. Redding, R.F. Margolskee, L.C. Osborne, D. Artis, W.S. Garrett, Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut, Science. 351 (2016) 1329–1333. https://doi.org/10.1126/science.aaf1648.

[88] Y. Belkaid, B.T. Rouse, Natural regulatory T cells in infectious disease, Nat. Immunol. 6 (2005) 353–360. https://doi.org/10.1038/ni1181.

[89] X. Li, Y. Zheng, Regulatory T cell identity: formation and maintenance, Trends Immunol. 36 (2015) 344–353. https://doi.org/10.1016/j.it.2015.04.006.

[90] R.M. Maizels, H.J. McSorley, Regulation of the host immune system by helminth parasites, J. Allergy Clin. Immunol. 138 (2016) 666–675. https://doi.org/10.1016/j.jaci.2016.07.007.

[91] I. Ferreira, D. Smyth, S. Gaze, A. Aziz, P. Giacomin, N. Ruyssers, D. Artis, T. Laha, S. Navarro, A. Loukas, H.J. McSorley, Hookworm Excretory/Secretory Products Induce Interleukin-4 (IL-4) + IL-10 + CD4 + T Cell Responses and Suppress Pathology in a

Mouse Model of Colitis, Infect. Immun. 81 (2013) 2104–2111. https://doi.org/10.1128/IAI.00563-12.

[92] S. Navarro, D.A. Pickering, I.B. Ferreira, L. Jones, S. Ryan, S. Troy, A. Leech, P.J. Hotez, B. Zhan, T. Laha, R. Prentice, T. Sparwasser, J. Croese, C.R. Engwerda, J.W. Upham, V. Julia, P.R. Giacomin, A. Loukas, Hookworm recombinant protein promotes regulatory T cell responses that suppress experimental asthma, Sci. Transl. Med. 8 (2016) 362ra143. https://doi.org/10.1126/scitranslmed.aaf8807.

[93] R.M. Eichenberger, S. Ryan, L. Jones, G. Buitrago, R. Polster, M. Montes de Oca, J. Zuvelek, P.R. Giacomin, L.A. Dent, C.R. Engwerda, M.A. Field, J. Sotillo, A. Loukas, Hookworm Secreted Extracellular Vesicles Interact With Host Cells and Prevent Inducible Colitis in Mice, Front. Immunol. 9 (2018) 850. https://doi.org/10.3389/fimmu.2018.00850.

[94] C. Han, J. Yu, Z. Zhang, P. Zhai, Y. Zhang, S. Meng, Y. Yu, X. Li, M. Song, Immunomodulatory effects of Trichinella spiralis excretory-secretory antigens on macrophages, Exp. Parasitol. 196 (2019) 68–72. https://doi.org/10.1016/j.exppara.2018.10.001.

[95] G. Coakley, R.M. Maizels, A.H. Buck, Exosomes and Other Extracellular Vesicles: The New Communicators in Parasite Infections, Trends Parasitol. 31 (2015) 477–489. https://doi.org/10.1016/j.pt.2015.06.009.

[96] S.W. Jang, M.K. Cho, M.K. Park, S.A. Kang, B.-K. Na, S.C. Ahn, D.-H. Kim, H.S. Yu, Parasitic helminth cystatin inhibits DSS-induced intestinal inflammation via IL-10(+)F4/80(+) macrophage recruitment, Korean J. Parasitol. 49 (2011) 245–254. https://doi.org/10.3347/kjp.2011.49.3.245.

[97] J. Xu, M. Liu, P. Yu, L. Wu, Y. Lu, Effect of recombinant Trichinella spiralis cysteine proteinase inhibitor on TNBS-induced experimental inflammatory bowel disease in mice, Int. Immunopharmacol. 66 (2019) 28–40. https://doi.org/10.1016/j.intimp.2018.10.043.

[98] M.M. Zaiss, A. Rapin, L. Lebon, L.K. Dubey, I. Mosconi, K. Sarter, A. Piersigilli, L. Menin, A.W. Walker, J. Rougemont, O. Paerewijck, P. Geldhof, K.D. McCoy, A.J. Macpherson, J. Croese, P.R. Giacomin, A. Loukas, T. Junt, B.J. Marsland, N.L. Harris, The Intestinal Microbiota Contributes to the Ability of Helminths to Modulate Allergic Inflammation, Immunity. 43 (2015) 998–1010. https://doi.org/10.1016/j.immuni.2015.09.012.

[99] K. Atarashi, T. Tanoue, T. Shima, A. Imaoka, T. Kuwahara, Y. Momose, G. Cheng, S. Yamasaki, T. Saito, Y. Ohba, T. Taniguchi, K. Takeda, S. Hori, I.I. Ivanov, Y. Umesaki,

K. Itoh, K. Honda, Induction of colonic regulatory T cells by indigenous Clostridium species, Science. 331 (2011) 337–341. https://doi.org/10.1126/science.1198469.

[100] L.E.P.M. van der Vlugt, J.F. Zinsou, A. Ozir-Fazalalikhan, P.G. Kremsner, M. Yazdanbakhsh, A.A. Adegnika, H.H. Smits, Interleukin 10 (IL-10)-producing CD1dhi regulatory B cells from Schistosoma haematobium-infected individuals induce IL-10-positive T cells and suppress effector T-cell cytokines, J. Infect. Dis. 210 (2014) 1207–1216. https://doi.org/10.1093/infdis/jiu257.

[101] R.E. Ley, C.A. Lozupone, M. Hamady, R. Knight, J.I. Gordon, Worlds within worlds: evolution of the vertebrate gut microbiota, Nat. Rev. Microbiol. 6 (2008) 776–788. https://doi.org/10.1038/nrmicro1978.

[102] J.M. Leung, A.L. Graham, S.C.L. Knowles, Parasite-Microbiota Interactions With the Vertebrate Gut: Synthesis Through an Ecological Lens, Front. Microbiol. 9 (2018) 843. https://doi.org/10.3389/fmicb.2018.00843.

[103] D. Damaskos, G. Kolios, Probiotics and prebiotics in inflammatory bowel disease: microflora “on the scope,” Br. J. Clin. Pharmacol. 65 (2008) 453–467. https://doi.org/10.1111/j.1365-2125.2008.03096.x.

[104] F. Scaldaferri, V. Gerardi, L.R. Lopetuso, F. Del Zompo, F. Mangiola, I. Boškoski, G. Bruno, V. Petito, L. Laterza, G. Cammarota, E. Gaetani, A. Sgambato, A. Gasbarrini, Gut microbial flora, prebiotics, and probiotics in IBD: their current usage and utility, Biomed Res. Int. 2013 (2013) 435268. https://doi.org/10.1155/2013/435268.

[105] J.E. Aguilar-Toalá, R. Garcia-Varela, H.S. Garcia, V. Mata-Haro, A.F. González-Córdova, B. Vallejo-Cordoba, A. Hernández-Mendoza, Postbiotics: An evolving term within the functional foods field, Trends in Food Science & Technology. 75 (2018) 105–114. https://doi.org/10.1016/j.tifs.2018.03.009.

[106] J.L. Reyes, F. Lopes, G. Leung, N.L. Mancini, C.E. Matisz, A. Wang, E.A. Thomson, N. Graves, J. Gilleard, D.M. McKay, Treatment with Cestode Parasite Antigens Results in Recruitment of CCR2+ Myeloid Cells, the Adoptive Transfer of Which Ameliorates Colitis, Infect. Immun. 84 (2016) 3471–3483. https://doi.org/10.1128/IAI.00681-16.

[107] B. Muszyńska, A. Grzywacz-Kisielewska, K. Kała, J. Gdula-Argasińska, Anti-inflammatory properties of edible mushrooms: A review, Food Chem. 243 (2018) 373–381. https://doi.org/10.1016/j.foodchem.2017.09.149.

[108] R.M.K. Toghueo, Anti-leishmanial and Anti-inflammatory Agents from Endophytes: A Review, Nat. Products Bioprospect. 9 (2019) 311–328. https://doi.org/10.1007/s13659-019-00220-5.

Box 1. Definition of key terms

α-diversity: within-sample microbiota diversity. It describes the key features of a microbial community in a certain niche (e.g., the gut) in terms of species **richness** (i.e., the number of different species present in that niche) and species **evenness** (i.e., the distribution and relative abundances of the various species). To compute and compare α-diversity values, various indexes have been developed (e.g., 34Shannon-Weaver, Chao1, Simpson). Each index attributes different weights to species richness and evenness, thus returning a slightly different picture of the community under study. A reduction in the ecological indexes of α-diversity has been repeatedly described to accompany a wide range of disorders, especially inflammatory and autoimmune diseases.

Pro-biotics: The World Health Organization, in 2001, defined probiotics as live microorganisms that, "when administered in adequate amounts, confer a health benefit on the host." These microorganisms must be able to survive stomach acid and bile, maintain viability for long periods, and must be safe for the host.

Pre-biotics: dietary substrates that stimulate growth or activities of specific taxa within the gut microbiota. Examples of pre-biotics are: 1) substrates that favour the growth of specific taxa/groups of taxa; 2) specific carbohydrate mixtures, that once ingested will be fermented by anaerobic bacteria to produce immunomodulatory short chain fatty acids (SCFA), among which the anti-inflammatory butyrate

Post-biotics: a new generation food supplements, constituted by lysed, inactivated microbes (with all beneficial their metabolic products), microbial fractions or isolated bacterial components, as well as microbial products.

Section A

Appendix 3

What Pediatricians Should Know Before Studying Gut Microbiota.

Drago L, Panelli S, Bandi C, Zuccotti G, Perini M, D'Auria E.

J Clin Med. 2019 Aug 12;8(8):1206. doi: 10.3390/jcm8081206

Journal of
Clinical Medicine

MDPI

Review

What Pediatricians Should Know before Studying Gut Microbiota

Lorenzo Drago [1,2,*], Simona Panelli [2], Claudio Bandi [2,3], Gianvincenzo Zuccotti [4], Matteo Perini [2] and Enza D'Auria [4]

1 Department of Biomedical Sciences for Health, Università di Milano, 20133 Milan, Italy
2 Department of Biomedical and Clinical Sciences "L. Sacco", Pediatric Clinical Research Center "Invernizzi", Università di Milano, 20157 Milan, Italy
3 Department of Biosciences, Università di Milano, 20133 Milan, Italy
4 Department of Pediatrics, Children's Hospital Vittore Buzzi, Università di Milan, 20141 Milan, Italy
* Correspondence: lorenzo.drago@unimi.it

Received: 2 July 2019; Accepted: 9 August 2019; Published: 12 August 2019

Abstract: Billions of microorganisms, or "microbiota", inhabit the gut and affect its homeostasis, influencing, and sometimes causing if altered, a multitude of diseases. The genomes of the microbes that form the gut ecosystem should be summed to the human genome to form the hologenome due to their influence on human physiology; hence the term "microbiome" is commonly used to refer to the genetic make-up and gene–gene interactions of microbes. This review attempts to provide insight into this recently discovered vital organ of the human body, which has yet to be fully explored. We herein discuss the rhythm and shaping of the microbiome at birth and during the first years leading up to adolescence. Furthermore, important issues to consider for conducting a reliable microbiome study including study design, inclusion/exclusion criteria, sample collection, storage, and variability of different sampling methods as well as the basic terminology of molecular approaches, data analysis, and clinical interpretation of results are addressed. This basic knowledge aims to provide the pediatricians with a key tool to avoid data dispersion and pitfalls during child microbiota study.

Keywords: gut microbiota; microbiome; maternal–fetal interface; newborn; child; pediatric disease; dysbiosis

1. Introduction

The field of microbiome research is quickly evolving and unravelling. Causal links between distinct microbial consortia, their collective functions, and host pathophysiology during the various stages of life are becoming increasingly clear. Studies of microbiome plasticity, composition, and function based on a distinction of the host phenotypes may lay the foundation for both therapeutic and preventive interventions [1]. Indeed, new practical aspects of microbiome studies will be focused on the personalization of actions as well as on an understanding of the inherent individual variability of microbiomes at different ages, stages of development, conditions, and internal or external influences. These studies will allow the comprehension of physiological features to explain, or predict, human health and disease states. Therefore, clinical studies need to be well designed and the subject/patient phenotype properly selected. Age and many other factors have the potential to strongly influence the results, thus clinical studies on microbiota in children should take into account the differences that naturally occur during growth. Other technical challenges that need to be addressed are linked to properly establishing, harmonizing, and standardizing clinical protocols for sample collection, processing, sequencing, and analysis that also takes into account the "microbiome's age". The issues of diet, environment, host immune system, and genetics as key factors for determining microbiome

and microbiota profiles have not been fully resolved yet. All of these influences can impact on the microbiota composition at any age and may sometimes be difficult to harmonize and standardize during clinical investigation.

Clinical and microbiological translation urgently needs to implement the main information on microbiota. This review aims to give a rapid overview of child microbiota in order to guide pediatricians to a better understanding of the field while trying to limit biases and intrinsic pitfalls before the study design and starting any clinical trials. Even if most of the reported literature and data specifically refer to the best studied community, in other words, the one inhabiting the gut, the knowledge discussed in the text, together with more practical aspects and recommendations, can also be adapted to the study of other medically-relevant communities (e.g., in nasal-oral cavities).

2. Basic Knowledge on Gut Microbes

The human body harbors trillions of microbial cells mainly represented by bacteria, but also includes archea, viruses, fungi, and parasites. These communities establish extensive networks of cross-feeding (trophic) interactions, consuming, producing, and exchanging hundreds of metabolites with each other and with their human host, with whom they constitute a unique ecological entity called "holobiont" [2,3]. Their highest density is reached in the intestinal compartment, particularly in the lower segments. Here, bacteria are estimated to reach a number of 10^{14} cells and their density in stool have been calculated in the order of 10^{11} per gram of dry material [4]. Although less-well studied, many other body habitats within healthy individuals are occupied by microbial communities such as the mouth and oral tract, nostrils, skin, vagina. The term 'microbiota' literally means all living organisms within a body-site habitat. More specifically, the term "gut microbiota" indicates the resident intestinal bacterial communities, and from a practical point of view, it is generally investigated, with obvious biases, through the analysis of fecal samples, which are easy and non-invasive to collect. The term 'microbiome' is used instead to refer to the genetic content of these microorganisms. Conventionally, research in the field is mainly focused on bacterial microbiome, but further fascinating results have come from the study of "virome", or the viruses inhabiting the gut, of "mycome", which reveals another intriguing world of gut fungi, and of "parasitome".

New genetic and sequencing technologies have opened the way to the 'metagenomic' approach, which directly analyzes the total microbial genomes contained in a sample, that in turn, allows information to be acquired on the genomic links between function and phylogenetic evolution. Other approaches faced in the field include 'metatranscriptomics', the study of the whole RNA repertoire from a microbial community; 'metaproteomics', the study of the entire protein content from the community; and 'meta-metabolomics', the study of small-molecule metabolites produced through the interaction of diet and microbiome [5–7].

The analysis of the gene coding for the ribosomal 16S rRNA is very useful for studying gut bacteria. 16S rRNA is a component of the prokaryotic ribosome and is coded by a gene spanning about 1500 bp. The 16S rRNA gene is highly conserved between different species of bacteria, but presents nine variable ("V") regions that allow identification at the genus or species level. After amplification of, typically, 2–3 V regions, the obtained sequences are clustered into nearly-identical tags called 'phylotypes' or 'operational taxonomic units' (OTUs). These terms refer to a group of microbes generally through the threshold of sequence homology between their 16S rRNA genes (e.g., ≥98% for a 'species'-level phylotype) [8].

Eukaryotic components of the microbiota (e.g., fungi and protozoans) can be analyzed through homologous ribosomal gene sequences (small-subunit rRNA, SSU rRNA), while viral communities that lack ribosomal genes are investigated through shotgun DNA sequencing, or via primers targeted on conserved sequences in viral families. The above approaches are referred to as culture-independent, while culturomics is a culturing approach that uses multiple culture conditions, combined with the MALDI-TOF mass spectrometry and/or the 16S rRNA sequencing, for the isolation and identification of the largest possible number of bacterial species [9].

The gut hosts taxonomically diverse archaea, bacteria, fungi, and viruses. Studies report at least 22 bacterial phyla in the body, mainly represented (>90%) by *Actinobacteria*, *Firmicutes*, *Proteobacteria*, and *Bacteroidetes*. In the gut, *Bacteroidetes* and *Firmicutes* represent the predominant phyla [10–12]. In addition to taxonomic composition, taxonomic diversity also needs to be considered in evaluating the homeostasis of microbiota. In particular, two parameters are routinely employed for this purpose: alpha diversity (within-sample diversity, how many taxa or lineages are present in a sample), and beta diversity (between-sample diversity, to which extent the guts of different subjects or patients share taxa or lineages). Parameters that need to be evaluated when computing these ecological indices are richness (i.e., how many bacterial taxa) and evenness, which also takes into account the relative abundance of taxa, in addition to presence/absence, and compares it between subjects or patients [13].

In this context, measures of species richness (for example, the number of observed species or the Chao1 index, which is an abundance-based estimator of species diversity) and phylogenetic measures (Faith's phylogenetic diversity) are sensitive to the number of sequences per sample, whereas this is true to a much minor extent for metrics that combine richness and evenness (Shannon index).

Statistical and computational analyses still remain the main challenge in microbiome research. Some methods currently used for their power and effect size analysis are based on PERMANOVA, Dirichlet Multinomial, or random forest analysis [14]. Parametric statistical tests (for example, the Student's t-test and ANOVA) as well as measures of correlation including Spearman's rank correlation can be used on the basis of the phenotypes under study and the type of information the researcher wants to capture.

3. The Intestinal Microbiota from Birth Throughout Childhood

Addressing neonatal and early-life microbiota is pivotal as many of the events capable of shaping microbial communities even in adults take place during this phase of life: gestational age at birth, type of delivery, breast vs. formula feeding, weaning, use of antibiotics, etc. [15,16]. When neonatal microbiota begins is still a subject of great debate. The "sterile womb paradigm", in other words, the notion that, under physiological conditions, the human fetal environment is sterile and microbial colonization begins with birth, has been accepted for decades. Recently, with the burst of metagenomic studies, there has been a group of papers that have found traces of a lowly abundant bacterial colonization in the placenta, endometrium, amniotic fluid, and meconium in healthy, full-term pregnancies (see Nature Editorial by C. Willyard, 2018, [17] and references therein). This has led some researchers to date back the seeding of the microbiota to before birth ("in utero colonization hypothesis"). The field is still the subject of much debate, and the results appear in general to be controversial. Recently, several scientists have underlined that, even if it is possible that not all healthy babies are born sterile as previously thought, particular caution is necessary when working on samples bearing a low microbial biomass due to the heavy contamination issues notoriously connaturated with such samples when using molecular approaches based on next-generation sequencing [17]. Other important points that have been raised are the difficulty of maintaining a strict sterility when collecting samples related to the in utero environment within a clinical setting, and the impossibility of using NGS-based techniques to discriminate DNA from viable cells and DNA belonging to dead organisms or derived from translocation from the blood stream [15,17].

The human intestine at birth is an aerobic environment, as such, while the adult gut microbiota is dominated by obligate anaerobes belonging to the *Firmicutes* and *Bacteroidetes* phyla, the neonatal pioneer flora is composed by aerotolerant taxa, mainly belonging to the *Enterobacteriaceae* family (phylum: *Proteobacteria*). In a matter of days, however, these microorganisms will reduce oxygen levels, and the intestinal lumen becomes anaerobic. This allows the colonization by strict anaerobes, dominated by *Bifidobacterium* (phylum: *Actinobacteria*); *Clostridium* (phylum: *Firmicutes*); and *Bacteroides* (phylum: *Bacteroidetes*) [18,19]. During the first months, the diet of the infant is almost exclusively milk, favoring milk oligosaccharide fermenters as the already cited *Bifidobacterium*, represented, at this stage,

by many species. Other predominant bacterial taxa are represented by *Enterococcaceae*, *Streptococcaceae*, and *Lactobacillaceae* [15].

A very recent paper [20] addressed the development of gut microbiota in a large cohort of children, comprising cases who seroconverted to islet cell autoantibody positivity, children who developed type 1 diabetes (T1D), and matched controls (healthy). This interesting analysis followed the longitudinal maturation of the microbiome from 3 to 46 months of age and determined the covariates that significantly affected its development. Globally, this study harmonized data by collecting 12,500 stool samples from 903 children in three different European countries and three US states. Breastfeeding and birth mode resulted in being the main factors able to drive gut microbiome during the developmental phase by changing some relevant bacterial clusters. The authors proposed three distinct phases of microbiome progression: a developmental phase (months 3–14), a transitional phase (months 15–30), and a stable phase (≥31 months). The Shannon diversity index changed significantly during the first two phases, unchanging only during the stable phase. This study represents a very nice model of how to harmonize the age of the children with other covariate factors. Figure 1 presents a proposal for pediatricians to use a personalized staging of the enrolled individuals to differentiate relevant microbial clusters and dominating phyla.

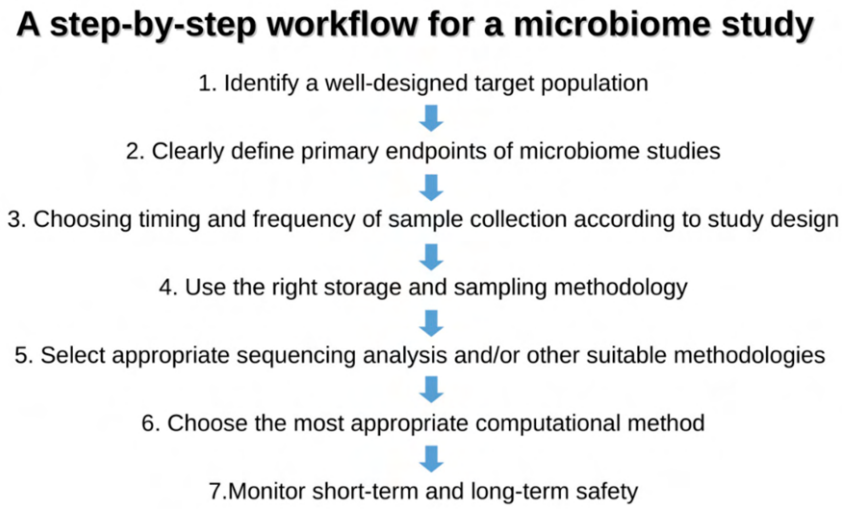


Figure 1. The figure represents the seven golden steps that the pediatrician should follow before the enrollment of individuals/patients in the microbiota study.

4. Issues to be Considered for Studying Microbiome in Clinical Studies

Study Design and Patient Selections

Pediatricians should select children cohorts by trying to limit the confounding factors that have the potential of diluting the statistical estimates of the effect sizes of the microbiome. Thus, as an example, when defining disease-specific signatures, the diseased population should be recruited with particular care in choosing patients who display a relatively homogeneous clinical phenotype. The choice of controls is also a challenging question: a good control population includes patients with a clinical phenotype that is a clear contrast from the one under study, while matching other relevant criteria. To reduce the heterogeneity of the cohort, it is indeed mandatory to clearly define inclusion and exclusion criteria by considering the factors affecting microbiota analysis (see below) and matching, accordingly, cases and controls. In this regard, it is crucially important to collect information about potential confounding factors, among which age group, for moderating influences that can artifactually alter results and the outcomes of interest. This is important in order to decrease co-variability and heterogeneity during the enrollment, by increasing the power of the analysis in parallel. The collected information will form part of the "metadata" (covariates) surrounding the sample and will later be used

in analyzing the data. To ensure consistency, recording the maximum information about the subjects, sample, and experimental procedures is recommended. Finally, before starting the study protocol, a sample size should be estimated on the basis of the expected effect size, and evaluated by means of a pilot study or based on similar previous studies. Other recent approaches rely on computing the estimated sample sizes by calculating the independent effect sizes on microbiota variation of other factors (covariates) relevant to the phenomenon under study [21].

Table 1 summarizes the key aspects to consider when designing and conducting a microbiome study, lists the possible confounders and pitfalls, and presents practical solutions for risk mitigation.

Table 1. Practical aspects to follow when drawing and studying a Microbiome.

Stages and Pitfalls	Considerations and Practical Solutions
Study question	• Clearly define the aim(s) of the study and the relevant biological question(s) before setting up the study design.
Statistically underpowered studies	• Correctly determine the sample size: consider that enrolling enough participants is important to ensure that the expected effect will be detected. • The sample size can be estimated by means of pilot studies, or from previous similar studies, or alternatively from computational approaches that consider the effect of covariates on the total microbiota variation (see main text).
Selection of subjects: avoiding heterogeneity of the population	• Clearly define inclusion and exclusion criteria: consider that an initial heterogeneity of the population will then dilute the statistical estimates of effect sizes on the microbiome. • The list of exclusion criteria from the National Institutes of Health (NIH) Human Microbiome Project can be relied on with regard to the above-mentioned. • In a "cases vs. controls" study, aimed at detecting microbiota-based markers of a disease, choose "cases" with a care in maintaining a relatively homogeneous clinical phenotype. "Controls", in turn, must have a clinical phenotype in clear contrast, while matching other relevant criteria to avoid confounding factors. • Consider that multiple controls groups that are selected based on various criteria may provide more insights. • Additionally consider that for more generalizable results, independent cohorts may be selected to identify the microbiota signatures ("discovery cohort") and test the results ("validation cohort"). • In longitudinal studies, individuals can be treated as their own controls, by collecting baseline samples before and during/after a treatment.
Confounding factors (lifestyle and clinical factors)	• Be exhaustive in the collection of "metadata" (covariates) surrounding the sample: this will be pivotal later, when analyzing the data. Collect information on possible confounding, mediating, and moderating factors that can either influence the microbiome composition or the outcome of interest.
Timing and frequency of sample collection	• Cross-sectional sampling from patients is appropriate to discover and validate diagnostic microbiome signatures. • Repeated samplings of the same subject (time series or longitudinal sampling) ensure more insights into temporal dynamics and community changes. • Longitudinal sampling should be chosen for monitoring disease severity or response to a treatment. Frequency should be similar between subjects.
Sample collection and storage	• Storage and transit conditions are important variables in microbiome study outcomes as they impact DNA yields and quality. • After collecting samples, freeze immediately. When immediate freezing is not possible, short-term refrigeration (+4 °C) is helpful. An alternative is to use stabilizing solutions. • Long-term storage: currently the norm is −80 °C. • Minimize freezing-thawing cycles. To this aim, it is helpful to aliquot samples before freezing.

Table 1. *Cont.*

Stages and Pitfalls	Considerations and Practical Solutions
Experimental Lab procedures	• Use the same procedures and reagents throughout the study. Document everything and be consistent. If, for example, different batches of an enzyme are used, document it among the metadata. • DNA extraction: This is an important source of variation and bias because of the differential resistance to lysis of microbial cells. Combine chemical and mechanical lysing procedures to capture the most accurate community composition. • Contamination may significantly impact results, especially if working on low-biomass samples. It may derive from laboratory contaminants (e.g., previously produced amplicons), from reagents and commercial kits ("kitome"). It is recommendable to separate pre- and post-PCR areas and to introduce appropriate negative controls in different sample processing steps (e.g., blank extraction control: DNA-free water undergoes DNA extraction and all subsequent experimental procedures; blank PCR control: DNA-free water undergoes PCR and all subsequent procedures). • Selection of 16S primers: Rely on previous studies and consider that different couples of universal 16S primers may be biased toward (or against) certain bacterial taxa, thus giving artefactual over- (or under-representations) of them. For example, the 27F/338R primer sets (targeting the V1–V3 regions) is biased against the amplification of *Bifidobacteria*. Another possible pitfall is given by primer sets poorly resolving specific taxa. • PCR amplification: Low DNA template concentration and high number of PCR cycles introduce biases. To reduce their effects, minimize PCR cycles, use a standard (and relatively high) DNA template concentration, and pool multiple PCR (e.g., triplicates) for each sample. The use of proof-reading DNA polymerases and longer annealing times (to reduce chimera formation) is also recommended.
Sequencing	• Use positive controls to calibrate the sequencing method: (i) pure strains of, e.g., *Escherichia coli* that produce strong PCR bands of a known size; and (ii) a synthetic mock microbial community to ensure that amplification, sequencing, and taxonomic classification workflows have not introduced substantial bias or distortions in the expected microbiome profiles. Consider that, in addition to the DNA extraction and PCR steps, errors can be introduced during library preparation, sequencing, imaging, and data analysis.
Data analysis	• The design and choice of the analyses is strictly connected with the research objectives of the study. • Be consistent with the procedures and software used for analyzing data. Consider that different software versions can behave differently. • Integrate non-microbiome sources of data (e.g., clinical parameters) with microbiome data to answer the biological questions that primed the study. • Consider that microbiota data are high-dimensional in nature, with the total number of variable measurements far exceeding the number of samples. • Incorporate the patient and experimental covariates collected in the "metadata" file of the analysis. Evaluate if some of them act as confounding factors. • Repeat the analyses introducing some changes (e.g., change some parameters or algorithms, include or exclude metadata) and the evaluate reproducibility of results. • The complexity of questions in a translational study makes its useful to test multiple statistical models using several combinations of independent-dependent variables. • If a variable is continuous, using it directly in the model is substantially more informative than using a categorical or binary encoding. • Remember that DNA-based techniques are not able to reveal if the microbes under study are alive or dead. If precise information on this is needed, consider performing meta-transcriptomics.
Risk-benefit assessment	• Studies need to be designed to ensure that short term and long-term reliable data are collected.

5. Major Pre-, Peri-, and Post-Natal Factors Affecting the Child Gut Microbiota

A schematic representation of the factors that are able to affect the dynamics and composition of the intestinal microbiota is given in Figure 2.

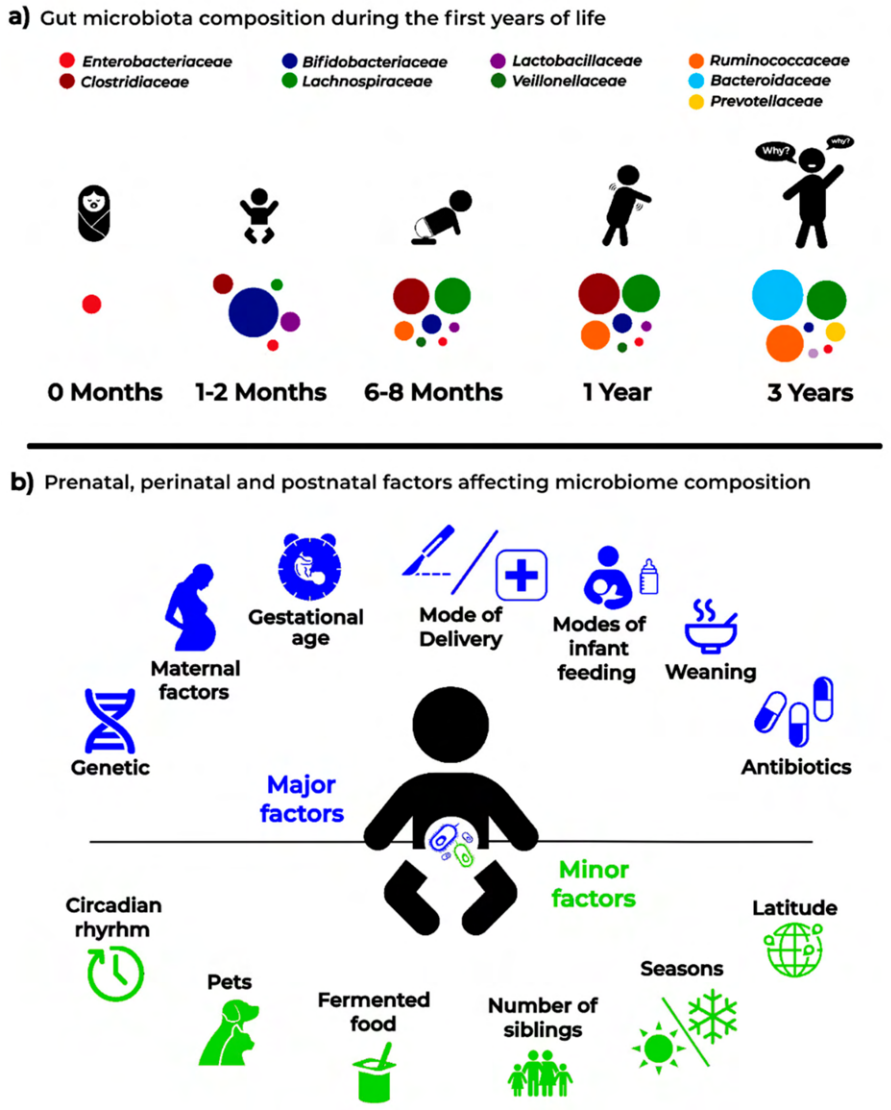


Figure 2. Infant microbiota composition (**a**) and the main "major" and "minor" factors affecting analysis and results in microbiota studies (**b**).

5.1. Maternal Factors Influencing Infant Microbiota

5.1.1. Changes Related to Vertical Transmission of Maternal Metabolites

During gestation, bacteria in the mother's intestine have been shown to drive the future immune maturation of the neonatal gut through the passage of soluble molecules from the placenta in the absence of direct colonization and of the vertical transmission of viable bacterial cells [22,23]. These bacteria are able to induce specific changes in the gut of newborns, creating new microbiota profiles.

5.1.2. Changes Related to Dietary Patterns and Lifestyle

The intestinal microbiota is strongly personalized and influenced by a plethora of environmental and inter-individual variables including body mass index (BMI), exercise frequency, and dietary patterns and habits (which in turn, are strongly related with cultural factors and lifestyle). It has

been reported that the infant's fecal microbiota composition is influenced by the BMI and weight gain of the mother during pregnancy [24,25]. In general, the maternal microbial reservoir plays a crucial role in the acquisition and development of early infant microbiota, which in turn is the key to establishing a healthy host–microbiome symbiosis with long-lasting health effects. Therefore, it can be easily understood as to why maternal diet and lifestyle should be monitored and categorized as relevant metadata in infant microbiota studies. In an early phase, after the huge microbial "inoculum" at birth, the infant continues to directly acquire maternal gut strains from different sources (e.g., from skin, mouth, milk) and these are likely to become stable colonizers of the infant gut. Later in life, increasingly important roles are also played by other factors such as shared diet and lifestyle.

5.2. Genetic Factors

There is growing evidence that geographical origin and host genetic makeup influence the acquisition and development of the gut microbiota, with clear associations reported between the host genotype and the relative abundances of different bacterial taxa. For example, Bonder et al. [26] described a single nucleotide polymorphism (SNP) in the LCT locus (coding for human lactase) that is related to varying abundances of *Bifidobacterium*. Goodrich et al. [27], by comparing microbiota across samples belonging to either monozygotic and dizygotic twin pairs, reported a number of microbial taxa whose abundances were strongly influenced by host genetics. Among such taxa, the *Christensellaceae*, considered a microbiome-based marker of obesity and is significantly enriched in individuals with low BMI, resulted in the most highly heritable taxon. Any data related to the genetic hardware of the child should then be noticed.

5.3. Mode of Delivery

At birth, the infant gut communities tend to resemble the maternal vagina or skin microbiota in cases of vaginal or cesarean section (C-section) delivery, respectively [19,28]. Even later, when these "pioneer" foundation populations have been replaced, the birth mode seems to exert significant long-term effects on the structure of the gut microbiota. At 24 months of age, the gut microbial communities of cesarean delivered infants still appear to be less diverse [15]. Even in children as old as seven years, some authors have reported the enduring influence of the mode of delivery, but data are somewhat contrasting regarding this point [19]. Vaginally delivered infants tend to be colonized by *Lactobacillus* and *Prevotella*, while C-section neonates are preferentially colonized by microorganisms from maternal skin, and the hospital staff or environment.

5.4. Mode of Infant Feeding

Breastfed infants receive, from their mothers' milk, a complex mix that will affect the milieu within which their own microbiota will develop. This mix is made up of nutrients, antimicrobial proteins, short chain fatty acids (SCFA), secretory IgA, non-digestible oligosaccharides (HMOs, human milk oligosaccharides, that promote the proliferation of specific gut bacterial taxa in the neonate), and live bacteria, even if previously considered germ-free [15]. The source of the "milk microbiota", which has a transient nature and declines rapidly at weaning, has recently been another subject of debate. At least some of the bacteria is thought to reach the mammary gland through an endogenous route called the enteromammary pathway, which has not been fully elucidated yet. It has also been suggested that mammary skin microbiota can travel via the lymphatic and vascular circulations to the breast ([15,16] and references therein). Gut microbiota differences between breastfed and formula-fed infants are indeed well documented. The former exhibit lower diversity indexes, indicative of a more uniform population where *Bifidobacterium* and *Lactobacillus* dominate. The latter are characterized by more diverse communities, with higher proportions of *Bacteroides*, *Clostridium*, *Streptococcus*, *Veillonella*, *Atopobium*, and *Enterobacteriaceae* [29]. Finally, compositional differences in microbial communities in human milk sampled from different geographical locations have been studied and reported to create strong variability between newborn microbiota [30].

5.5. Gestational Age

While in full-term infants, delivery and feeding mode are reported to represent the major drivers of microbiota development, in preterm (PT) infants (<37 weeks of gestation), the gestational age seems to have the biggest impact on the assembly of gut communities [19,31,32]. PT neonates experience a number of unique challenges in the establishment of their microbiota. Their colonization patterns are characterized by the involvement of peculiar microbial sources, mainly bacteria deriving from the neonatal intensive care unit (NICU) environment [33]. Not rarely, these are strains implicated in nosocomial infections such as *Enterococcus* spp., *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter* spp., *Pesudomonas aeruginosa*, and other *Enterobacteriaceae* [34] with their burden of antibiotic resistance genes. Other relevant features of this peculiar colonization trajectory are its extreme inter-individual variability, and the fact that, across studies, it does not appear to be univocally linked to health outcomes as necrotizing enterocolitis and late-onset sepsis. Instead, the colonization process seems to reflect the co-occurrence of a variety of nosocomial "variables" [35], among which are parenteral nutrition and antibiotic usage (see below). Antibiotics, normally administered to these patients, in turn perturbate the colonization process by killing bacteria acquired during birth and promoting the growth of taxa significantly different from those found in more physiological situations [31]. In conclusion, the PT microbiota appears to be more unstable than that of full-term equivalents and is believed to be associated with a delay in the establishment of an adult-type signature microbiota [16]. All these individuals should be carefully selected and clearly categorized by the clinician before enrollment into the microbiota study.

5.6. Antibiotics

Specific properties of antibiotics, as a mode of action and antimicrobial spectrum, might act as powerful forces for the selection of intestinal bacterial populations, especially if the infant is exposed to antibiotics too early and/or for long periods of time [3,15]. Antibiotics are able to alter the abundances of resident bacteria, significantly impact the growth of otherwise dominant bacterial phyla, and lead to an overall decrease in microbial diversity. A study by Fouhy and colleagues [36] showed that infants exposed to ampicillin and gentamicin shortly after birth harbored higher proportions of *Proteobacteria* and *Actinobacteria*, and the genus *Lactobacillus* for up to four weeks after concluding treatment. Another study reported an attenuation in colonization with *Bifidobacterium* and an increase of *Enterococcus* in subjects receiving oral or intravenous antibiotics during the first four days of life [37].

This variability among individuals suggests caution when including subjects who have been treated with antibiotics [38]. Indeed, the exclusion criteria from the NIH Human Microbiome Project (HMP, dbGAP, see the url https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000228.v4.p1) include the use of systemic antibiotics, antifungals, antivirals, or antiparasitics within six months of sampling. However, this criterion, although optimal, may not be easily applicable with subjects in the pediatric age. For this reason, shorter time windows are often considered. In any case, it is mandatory to accurately document, within the metadata file, any history of antibiotics as well as other medication use.

5.7. Weaning

The transition to more varied, solid food is an important step in the development of the early-life gut microbiota; infants begin to be exposed to a much larger array of substrates and non-digestible carbohydrates that promote the survival and proliferation of more various bacterial taxa. As a consequence, the alpha diversity increases; moreover, *Proteobacteria* and *Actinobacteria* are replaced by *Firmicutes* and *Bacteroidetes* as the dominant phyla, in a more adult-like compositional structure. The cessation of exclusive milk feeding correlates with the decrease of saccharolytic bacteria as *Bifidobacteriaceae* (phylum: *Actinobacteria*). The increased protein intake is thought to be associated with

an increase of *Lachnospiraceae* (phylum: *Firmicutes*), while the ingestion of fibers with that of higher levels of *Prevotellaceae* (phylum: *Bacteroidetes*) [39].

In general, the relative abundance of our intestinal microbes is highly influenced by dietary patterns and habits [11], that should therefore be taken into account in clinical studies targeting microbiota.

6. Minor Factors Affecting Gut Microbiota

Various minor factors can affect and modify the gut microbiota, which can occur at any stage of life. Insomnia and circadian rhythm disruption, latitude with time zone shift and intercontinental flights (with the consequent jet lag), household siblings, and companion animals as well as seasonal changes can modify gut microbiota and determine different microbiota profiles with high inter-individual variability to responses to the different factors [40–42]. All of these factors can influence the results and should be carefully considered before starting a clinical study and accurately reported in the metadata to then be considered later in the downstream bioinformatics and statistical analyses. Other similar confounder factors such as bowel movement preparations, evacuants or laxatives, or any microorganism-supplemented food (such as probiotics) can act as deep and long-time gut modifiers, thus a plot-to-plot variation needs to be addressed with nested statistical tests.

7. Sample Collection

Donors/patients to enroll, their genetic or disease phenotypes as well as the expertise of the clinician in methodology used for collecting samples are very relevant in designing a correct study. The number of samples and patients to be enrolled is an intriguing and still hotly debated topic. Sample stability as well as shipping and storage requirements need to be more appropriate and will surely be improved and standardized in the future. Researchers may find some procedures at http://www.microbiome-standards.org or at https://www.hmpdacc.org/resources/metagenomics_sequencing_analysis.php and other papers [43–45].

Concerning the practical aspects, an important question is how often to collect samples because the microbiome ecology is intrinsically dynamic. This largely depends on what question one is trying to answer. If, for gastrointestinal disorders, remarkable changes can be observed between one day and the next (e.g., in times surrounding surgery or in correspondence with periods of activity or remission of the pathology), changes induced by other factors (e.g., diet) often take place on a longer timescale. Collection of multiple samples from the same patient is preferred to allow for better standardization on the basis of the type of patients, centers involved, and statistical power. Whether or not samples collected from the same individual can be pooled before analysis is another topic to be standardized. An important point is that sampling and storage do affect microbiota composition in healthy as well as in diseased subjects. The most widely accepted protocols include immediate homogenization and freezing either with dry ice or in liquid nitrogen, followed by storage at −80 °C. However, this approach is not always practical, particularly for stool samples, or in the case of stool collection from a large scale cohort or remote/rural areas. Whether samples must be immediately frozen (and at what temperature) or whether they can withstand a period of room temperature remains controversial. The above-mentioned studies showed that the effects of short-term storage conditions on the structure and diversity of communities are quite small in general. In particular, storage at −80 °C, −20 °C for a week, or 4 °C for 24 h were found to not significantly affect the ecological indexes of between-sample diversity or the abundance of major taxa [45]. In contrast, the number of freeze–thaw cycles seems to have an effect on the composition of the microbial community, thus it is strongly recommended to aliquot samples at the beginning. Of course, some DNA stabilizers can be used to prolong the stability of samples. In the study of Choo et al. [46] Omnigene Gut and Tris EDTA appeared to show the same performance as storage in an ultrafreezer (−80 °C). In addition to feces, swabs can be an alternative starting material for DNA extractions, especially within hospital settings, even if some studies have shown that the stool swabs of some subjects had limited and not detectable bacterial DNA. A recent study by Christine M. Bassis [47], by comparing stool versus rectal swab samples and their storage

conditions, demonstrated minor differences in the bacterial community profiles between the stool and swab from the same subject as well as when samples were stored up to 27 h at +4 °C before freezing at −80 °C. Interestingly, this study also concluded that it was possible to thaw and refreeze samples a limited number of times under particular conditions (i.e., immediately frozen at −20 °C, first thaw cycle, refrozen at −80 °C; immediately frozen at −20 °C, first thaw cycle, refrozen at −20 °C, second thaw cycle and frozen at −80 °C) without strong effects on the community composition. A word of caution is, however, due on this point, as the consensus recommendations are different, as detailed above. Finally, it is to be underlined that as the collection of stool can be difficult from some subjects under certain experimental conditions, swab collection may be useful in such cases, which also has the advantage that they are more easily shipped and handled. A further recent study confirmed that swab samples reliably replicate the stool microbiota bacterial composition when swabs are processed quickly (≤2 days) [48].

Finally, special considerations are needed if addressing peculiar samples such as the newborn's first intestinal discharge (meconium). The debate about "when" the neonatal microbiota begins has been previously mentioned. Recently, several scientists have underlined that, even if it is possible that not all healthy babies are born sterile as previously thought, particular caution is due when working on samples bearing low microbial biomass such as meconia because of the contamination issues connaturated with molecular approaches based on PCR amplification and next-generation sequencing [17,49,50]. The presence of contaminating DNA in laboratory reagents (so-called "kitome") is a serious challenge in these cases; low levels of target bacterial DNA in a sample have been reported to correlate with a high proportion of sequences being attributable to contamination [51,52].

8. Discussion

The Anna Karenina principle, based on Leo Tolstoy's great book and cited in 1878 (*All happy families are alike: each unhappy family is unhappy in its own way*), has been recently translated by Zaneveld et al. [53] as the response to stress against the stability of animal microbiomes. These authors discussed how healthy microbiomes may be quite similar between individuals, but each dysbiotic microbiota is dysbiotic in its own way. The associations between microbiome instability/variability and many confounding factors as well as with diseases, suggest that microbiome may have many and simultaneous multiple faces.

This "stochastic" drift, occurring at any stage of life under stress conditions, can create several phenotypes that need to be known and harmonized when planning a study on microbiota.

Early childhood possesses distinct microbiota tracts compared with later ones, where different clusters and phyla may be differently represented. One common characteristic during this early stage of life is that bacterial richness and diversity increase during growth. Therefore, pediatricians should know that there are several age-related microbiota profiles, and should also be aware of the need to categorize each individual in a defined, monthly range by carefully considering the above-mentioned interference factors.

Several specialties need to be involved in this aim as well as the combination of different knowledge. The "Clinical Microbiota Expert" is not only a new job, but represents a step forward to create competence in this field where clinical microbiologists, clinicians, and bioinformaticians are merged into one. This new job-role will have to create awareness on the study of the "dynamic body" such as the gut microbiota during early age by creating novel models and approaches as well as solutions to solve and interpret the clinical microbiology results. Therefore, translational methodologies to approach a new way of designing clinical trials need to use feasibility and efficacy tools, and a deeper preparation in the field to avoid uncontrolled errors, unsubstantiated results, and excessive costs.

9. Conclusions and Future Perspectives

Next-generation sequencing methodologies still remain expensive and the diagnostic market is offering different solutions, thus a proper, and especially judicious, use of these methods is definitively mandatory. The clinical microbiota expert and pediatricians involved in the field will also have to guide through this jungle by trying to avoid false myths and promises that could be difficult to realize. In the near future, all of these studies and experiences will necessarily lead to a better understanding of the real key phases of microbiome progression from birth throughout childhood.

A final consideration to underline is that the metagenomics community still needs to fully converge toward standardized methods and procedures, leading to an investigation of the sources of variability and bias at each step of the workflow, and to an improved reproducibility and comparability between studies. This is a necessary premise for moving from correlation studies to causation investigations and to answer complex questions in a translational setting.

Author Contributions: L.D. designed and conceived the study; S.P. revised the paper and the technical aspects and checked the literature; C.B. and G.Z. revised the manuscript; M.P. conceived the figures and revised the manuscript; E.D. revised the clinical aspects and the manuscript.

Acknowledgments: The authors wish to thank the Fondazione "Romeo ed Enrica Invernizzi".

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zmora, N.; Zilberman-Schapira, G.; Suez, J.; Mor, U.; Dori-Bachash, M.; Bashiardes, S.; Kotler, E.; Zur, M.; Regev-Lehavi, D.; Brik, R.B.Z.; et al. Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. *Cell* **2018**, *174*, 1388–1405. [CrossRef] [PubMed]
2. Yatsunenko, T.; Rey, F.E.; Manary, M.J.; Trehan, I.; Dominguez-Bello, M.G.; Contreras, M.; Magris, M.; Hidalgo, G.; Baldassano, R.N.; Anokhin, A.P.; et al. Human gut microbiome viewed across age and geography. *Nature* **2012**, *486*, 222–227. [CrossRef] [PubMed]
3. Rinninella, E.; Raoul, P.; Cintoni, M.; Franceschi, F.; Miggiano, G.A.D.; Gasbarrini, A.; Mele, M.C. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* **2019**, *7*, 14. [CrossRef] [PubMed]
4. Lozupone, C.A.; Stombaugh, J.I.; Gordon, J.I.; Jansson, J.K.; Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **2012**, *489*, 220–230. [CrossRef] [PubMed]
5. Williams, S.C. The other microbiome. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2682–2684. [CrossRef] [PubMed]
6. Thomas, T.; Gilbert, J.; Meyer, F. Metagenomics—A guide from sampling to data analysis. *Microb. Inform. Exp.* **2012**, *2*, 3. [CrossRef] [PubMed]
7. Turnbaugh, P.J.; Gordon, J.I. An invitation to the marriage of metagenomics and metabolomics. *Cell* **2006**, *134*, 708–713. [CrossRef]
8. Trivedi, B. Microbiome: The surface brigade. *Nature* **2012**, *492*, S60–S61. [CrossRef]
9. Lagier, J.C.; Dubourg, G.; Million, M.; Cadoret, F.; Bilen, M.; Fenollar, F.; Levasseur, A.; Rolain, J.M.; Fournier, P.E.; Raoult, D. Culturing the human microbiota and culturomics. *Nat. Rev. Genet.* **2018**, *16*, 540–550. [CrossRef]
10. Peterson, J.; Garges, S.; Giovanni, M.; McInnes, P.; Wang, L.; Schloss, J.A.; Bonazzi, V.; McEwen, J.E.; Wetterstrand, K.A.; Deal, C.; et al. The NIH human microbiome project. *Genome Res.* **2009**, *19*, 2317–2323.
11. Costello, E.K.; Lauber, C.L.; Hamady, M.; Fierer, N.; Gordon, J.I.; Knight, R. Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science* **2009**, *326*, 1694–1697. [CrossRef] [PubMed]
12. Qin, J.; Li, R.; Raes, J.; Arumugam, M.; Burgdorf, K.S.; Manichanh, C.; Nielsen, T.; Pons, N.; Levenez, F.; Yamada, T.; et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **2010**, *464*, 59–65. [CrossRef] [PubMed]
13. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **2012**, *486*, 207–214. [CrossRef] [PubMed]

14. Kelly, B.J.; Gross, R.; Bittinger, K.; Sherrill-Mix, S.; Lewis, J.D.; Collman, R.G.; Bushman, F.D.; Li, H. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* **2015**, *31*, 2461–2468. [CrossRef]
15. Arrieta, M.C.; Stiemsma, L.T.; Amenyogbe, N.; Brown, E.M.; Finlay, B. The Intestinal Microbiome in Early Life: Health and Disease. *Front. Immunol.* **2014**, *5*, 427. [CrossRef]
16. Milani, C.; Duranti, S.; Bottacini, F.; Casey, E.; Turroni, F.; Mahony, J.; Belzer, C.; Palacio, S.D.; Montes, S.A.; Mancabelli, L.; et al. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. Microbiol. *Mol. Biol. Rev.* **2017**, *81*, e00036-17. [CrossRef]
17. Willyard, C. Could baby's first bacteria take root before birth? *Nature* **2018**, *553*, 264–266. [CrossRef]
18. Matamoros, S.; Guen, C.G.L.; Le Vacon, F.; Potel, G.; De La Cochetière, M.F. Development of intestinal microbiota in infants and its impact on health. *Trends Microbiol.* **2013**, *21*, 167–173. [CrossRef]
19. Hill, C.J.; Lynch, D.B.; Murphy, K.; Ulaszewska, M.; Jeffery, I.B.; O'Shea, C.A.; Watkins, C.; Dempsey, E.; Mattivi, F.; Touhy, K.; et al. Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome* **2017**, *5*, 4. [CrossRef]
20. Stewart, C.J.; Ajami, N.J.; O'Brien, J.L.; Hutchinson, D.S.; Smith, D.P.; Wong, M.C.; Ross, M.C.; Lloyd, R.E.; Doddapaneni, H.; Metcalf, G.A.; et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **2018**, *562*, 583. [CrossRef]
21. Falony, G.; Joonssens, M.; Vieine-Silva, S.; Wang, J.; Darzi, Y.; Faust, K.; Kurilshikov, A.; Bonder, M.J.; Valles-Colomer, M.; Vandeputte, P.; et al. Population-level analysis of gut microbiota variation. *Science* **2016**, *352*, 560–564. [CrossRef] [PubMed]
22. De Agüero, M.G.; Ganal-Vonarburg, S.C.; Fuhrer, T.; Rupp, S.; Uchimura, Y.; Li, H.; Steinert, A.; Heikenwalder, M.; Hapfelmeier, S.; Sauer, U.; et al. The maternal microbiota drives early postnatal innate immune development. *Science* **2016**, *351*, 1296–1302. [CrossRef] [PubMed]
23. Rakoff-Nahoum, S. Another reason to thank mum: Gestational effects of microbiota metabolites. *Cell Host Microbe* **2016**, *19*, 425–427. [CrossRef] [PubMed]
24. MacPherson, A.J.; De Agüero, M.G.; Ganal-Vonarburg, S.C. How nutrition and the maternal microbiota shape the neonatal immune system. *Nat. Rev. Immunol.* **2017**, *17*, 508–517. [CrossRef] [PubMed]
25. Kearney, J.F.; Patel, P.; Stefanov, E.K.; King, R.G. Natural Antibody Repertoires: Development and Functional Role in Inhibiting Allergic Airway Disease. *Annu. Rev. Immunol.* **2015**, *33*, 475–504. [CrossRef] [PubMed]
26. Bonder, M.J.; Kurilshikov, A.; Tigchelaar, E.F.; Mujagic, Z.; Imhann, F.; Vila, A.V.; Deelen, P.; Vatanen, T.; Schirmer, M.; Smeekens, S.P.; et al. The effect of host genetics on the gut microbiome. *Nat. Genet.* **2016**, *48*, 1407–1412. [CrossRef] [PubMed]
27. Goodrich, J.K.; Waters, J.L.; Poole, A.C.; Sutter, J.L.; Koren, O.; Blekhman, R.; Beaumont, M.; Van Treuren, W.; Knight, R.; Bell, J.T.; et al. Human genetics shape the gut microbiome. *Cell* **2014**, *159*, 789–799. [CrossRef] [PubMed]
28. Dominguez-Bello, M.G.; Costello, E.K.; Contreras, M.; Magris, M.; Hidalgo, G.; Fierer, N.; Knight, R. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11971–11975. [CrossRef] [PubMed]
29. Fallani, M.; Young, D.; Scott, J.; Norin, E.; Amarri, S.; Adam, R.; Aguilera, M.; Khanna, S.; Gil, A.; Edwards, A.C.; et al. Intestinal Microbiota of 6-week-old Infants Across Europe: Geographic Influence Beyond Delivery Mode, Breast-feeding, and Antibiotics. *J. Pediatr. Gastroenterol. Nutr.* **2010**, *51*, 77–84. [CrossRef]
30. Drago, L.; Toscano, M.; De Grandi, R.; Grossi, E.; Padovani, E.M.; Peroni, D.G. Microbiota network and mathematic microbe mutualism in colostrum and mature milk collected in two different geographic areas: Italy versus Burundi. *ISME J.* **2017**, *11*, 875–884. [CrossRef]
31. Groer, M.W.; Luciano, A.A.; Dishaw, L.J.; Ashmeade, T.L.; Miller, E.; Gilbert, A.J. Development of the preterm infant gut microbiome: A research priority. *Microbiome* **2014**, *2*, 38. [CrossRef] [PubMed]
32. Stewart, C.J.; Embleton, N.D.; Clements, E.; Luna, P.N.; Smith, D.P.; Fofanova, T.Y.; Nelson, A.; Taylor, G.; Orr, C.H.; Petrosino, J.F.; et al. Cesarean or Vaginal Birth Does Not Impact the Longitudinal Development of the Gut Microbiome in a Cohort of Exclusively Preterm Infants. *Front. Microbiol.* **2017**, *8*, 1008. [CrossRef] [PubMed]

33. Brooks, B.; Olm, M.R.; Firek, B.A.; Baker, R.; Thomas, B.C.; Morowitz, M.J.; Banfield, J.F. Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. *Nat. Commun.* **2017**, *8*, 1814. [CrossRef] [PubMed]
34. Hu, H.; Johani, K.; Gosbell, I.; Jacombs, A.; Almatroudi, A.; Whiteley, G.; Deva, A.; Jensen, S.; Vickery, K. Intensive care unit environmental surfaces are contaminated by multidrug-resistant bacteria in biofilms: Combined results of conventional culture, pyrosequencing, scanning electron microscopy, and confocal laser microscopy. *J. Hosp. Infect.* **2015**, *91*, 35–44. [CrossRef] [PubMed]
35. Wandro, S.; Osborne, S.; Enriquez, C.; Bixby, C.; Arrieta, A.; Whiteson, K. The microbiome and metabolome of preterm onfant stool are personalized and not driven by health outcomes, including necrotizing enterocolitis and late-onset sepsis. *Msphere* **2018**, *3*, e00104–e00118. [CrossRef] [PubMed]
36. Fouhy, F.; Guinane, C.M.; Hussey, S.; Wall, R.; Ryan, C.A.; Dempsey, E.M.; Murphy, B.; Ross, R.P.; Fitzgerald, G.F.; Stanton, C.; et al. High-Throughput Sequencing Reveals the Incomplete, Short-Term Recovery of Infant Gut Microbiota following Parenteral Antibiotic Treatment with Ampicillin and Gentamicin. *Antimicrob. Agents Chemother.* **2012**, *56*, 5811–5820. [CrossRef] [PubMed]
37. Tanaka, S.; Kobayashi, T.; Songjinda, P.; Tatejama, A.; Tsubouchi, M.; Kiyohara, C.; Shirakawa, T.; Sonomoto, J.; Nakayama, J. Influence of antibiotic exposure in the early postnatal period in the development of intestinal microbiota. *FEMS Immunol. Med. Microbiol.* **2009**, *56*, 80–87. [CrossRef]
38. Dethlefsen, L.; Relman, D.A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4554–4561. [CrossRef]
39. Koenig, J.E.; Spor, A.; Scalfone, N.; Fricker, A.D.; Stombaugh, J.; Knight, R.; Angenent, L.T.; Ley, R.E. Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, S4578–S4585. [CrossRef]
40. Thaiss, C.A.; Levy, M.; Korem, T.; Dohnalová, L.; Shapiro, H.; Jaitin, D.A.; David, E.; Winter, D.R.; Gury-BenAri, M.; Tatirovsky, E.; et al. Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. *Cell* **2016**, *167*, 1495–1510. [CrossRef]
41. Misic, A.M.; Davis, M.F.; Tyldsley, A.S.; Hodkinson, B.P.; Tolomeo, P.; Hu, B.; Nachamkin, I.; Lautenbach, E.; Morris, O.D.; Grice, A.E. The shared microbiota of humans and companion animals as evaluated from Staphylococcus carriage sites. *Microbiome* **2015**, *3*, 2. [CrossRef] [PubMed]
42. Davenport, E.R.; Mizrahi-Man, O.; Michelini, K.; Barreiro, L.B.; Ober, C.; Gilad, Y. Seasonal Variation in Human Gut Microbiome Composition. *PLoS ONE* **2014**, *9*, e90731. [CrossRef] [PubMed]
43. Costea, P.I.; Zeller, G.; Sunagawa, S.; Pelletier, E.; Alberti, A.; Levenez, F.; Tramontano, M.; Driessen, M.; Hercog, R.; Jung, F.E.; et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* **2017**, *35*, 1069–1076. [CrossRef] [PubMed]
44. Jordan, S.; Baker, B.; Dunn, A.; Edwards, S.; Ferranti, E.; Mutic, A.D.; Yang, I.; Rodriguez, J. Maternal–Child Microbiome: Specimen Collection, Storage and Implications for Research and Practice. *Nurs. Res.* **2017**, *66*, 175–183. [CrossRef] [PubMed]
45. Tedjo, D.I.; Jonkers, D.M.A.E.; Savelkoul, P.H.; Masclee, A.A.; Van Best, N.; Pierik, M.J.; Penders, J. The Effect of Sampling and Storage on the Fecal Microbiota Composition in Healthy and Diseased Subjects. *PLoS ONE* **2015**, *10*, 0126685. [CrossRef] [PubMed]
46. Choo, J.M.; Leong, L.E.; Rogers, G.B. Sample storage conditions significantly influence faecal microbiome profiles. *Sci. Rep.* **2015**, *5*, 16350. [CrossRef]
47. Bassis, C.M.; Moore, N.M.; Lolans, K.; Seekatz, A.M.; Weinstein, R.A.; Young, V.B.; Hayden, M.K. CDC Prevention Epicenters Program. Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles. *BMC Microbiol.* **2017**, *17*, 78–85. [CrossRef]
48. Bokulich, N.A.; Maldonado, J.; Kang, D.W.; Krajmalnik-Brown, R.; Caporaso, J.G. Rapidly Processed Stool Swabs Approximate Stool Microbiota Profiles. *mSphere* **2019**, *4*, e00208–e00219. [CrossRef]
49. Lauder, A.P.; Roche, A.M.; Sherrill-Mix, S.; Bailey, A.; Laughlin, A.L.; Bittinger, K.; Leite, R.; Elovitz, M.A.; Parry, S.; Bushman, F.D. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* **2016**, *4*, 29. [CrossRef]
50. Perez-Munoz, M.E.; Arrieta, M.C.; Ramer-Tai, A.E.; Walter, J. A critical assessment of the "sterile womb" and "in utero colonization" hypothesis: Implications for research of the pioneer infant microbiome. *Microbiome* **2017**, *5*, 48–67. [CrossRef]

51. Panelli, S.; Schneider, L.; Comandatore, F.; Bandi, C.; Zuccotti, G.V.; D'Auria, E.; Dauria, E. Is there life in the meconium? A challenging, burning question. *Pharmacol. Res.* **2018**, *137*, 148–149. [CrossRef] [PubMed]
52. Ferri, E.; Genco, F.; Gulminetti, R.; Bandi, C.; Novati, S.; Casiraghi, M.; Sambri, V. Plasma Levels of Bacterial DNA in HIV Infection: The Limits of Quantitative Polymerase Chain Reaction. *J. Infect. Dis.* **2010**, *202*, 176–177. [CrossRef] [PubMed]
53. Zaneveld, J.R.; McMinds, R.; Thurber, R.V. Stress and stability: Applying the Anna Karenina principle to animal microbiomes. *Nat. Microbiol.* **2017**, *2*, 17121. [CrossRef] [PubMed]

Section B

Appendix 4

Genomic Characterization of an ST1153 PVL-producing Methicillin Resistant *Staphylococcus aureus* Clinical Isolate in Italy.

Rimoldi SG, Comandatore F, Longhi E, Romeri F, Piazza A, Pagani C, Tamoni A, Longobardi C, Negri C, Bestetti G, Gervasoni C, Perini M, Antinori S, Bandi C, Gismondo MR.

New Microbiol. 2019 Apr;42(2):129-131

New Microbiologica, 42, 2, 129-131, 2019, ISN 1121-7138

CASE REPORT

Genomic Characterization of an ST1153 PVL-producing Methicillin Resistant *Staphylococcus aureus* Clinical Isolate in Italy

Sara Giordana Rimoldi[1], Francesco Comandatore[2], Erika Longhi[3], Francesca Romeri[1], Aurora Piazza[2], Cristina Pagani[1], Alessandro Tamoni[1], Concetta Longobardi[1], Cristina Negri[4], Giovanna Bestetti[4], Cristina Gervasoni[5], Matteo Perini[2], Spinello Antinori [4,5], Claudio Bandi[2,5], Maria Rita Gismondo[1,5]

[1]*Laboratorio di Microbiologia Clinica, Virologia e Diagnostica delle Bioemergenze, ASST Fatebenefratelli Sacco, Milano, Italia;*
[2]*Romeo ed Enrica Invernizzi Pediatric Research Center, Department of Biomedical and Clinical Sciences L. Sacco, University of Milan, Italy;*
[3]*Anatomia Patologica, ASST Fatebenefratelli Sacco, Milano, Italia;*
[4] *III Divisione di Malattie Infettive, ASST Fatebenefratelli Sacco, Milano, Italia;*
[5]*Dipartimento di Scienze Biomediche e Cliniche "L Sacco", Università degli Studi di Milano, Italia*

SUMMARY

Methicillin-resistant *Staphylococcus aureus* (MRSA) clones are rapidly increasing beyond the hospital into the community, livestock farming and environmental settings. An Italian man, a professional diver working in Egypt, was admitted to Infectious Diseases Clinic-ASST Fatebenefratelli Sacco for ulcerative skin lesions. An MRSA strain was isolated from the lesions' purulent exudate and the nasal colonization was also ascertained. The strain, characterized by whole genome sequencing, resulted to be Panton-Valentine Leukocidin (PVL) positive, SCC*mec*I - *spa*-type t504, and belonging to the sequence type 1153, sporadically described worldwide.

Received October 10, 2018 *Accepted* March 31, 2019

INTRODUCTION

In the last two decades, community-acquired methicillin-resistant *Staphylococcus aureus* (CA-MRSA) emerged worldwide, representing a public health major concern (Zetola *et al*., 2005). Moreover, MRSA strains have been isolated from environmental sources in different countries (Levin-Edens *et al*., 2012; Tice *et al*., 2010; Goodwin *et al*., 2009; El-Shenawy *et al*., 2005). The majority of *S. aureus* isolated from skin and soft tissue infections harbors the Panton-Valentine Leukocidin (PVL) genes, encoding for a cytotoxin causing leukocyte destruction and tissue necrosis (Niemann *et al*., 2018).
The molecular epidemiology and the disease burden of CA staphylococcal disease is highly variable in Europe and is clonally diverse (DeLeo, 2011). CA-MRSA clones harboring PVL genes have already been reported in Italy (Vignaroli *et al*., 2009; Bassetti *et al*., 2010).

CASE REPORT

In early April 2018, an Italian man in his 60s, working as professional diver in Egypt, was examined for multiple itching skin lesions, that evolved in a short time into cutaneous abscesses, by a dermatologist, who recommended a cycle of steroid therapy for one month. His medical history was notable for sporadic recurrence of cutaneous rashes in the previous 10 years. The patient also referred a hospitalization more than ten years before in Belgium, where he lived for years, due to an acute myocardial infarction. Because of the partial resolution of the clinical picture, two months later (June 2018) he requested a second opinion at the Infectious Diseases Clinic-ASST Fatebenefratelli Sacco. Here for the first time, a microbiological investigation was carried out and an MRSA strain was detected both in ulcerated skin lesions and in nasal swab. Due to the presence of ulcerated lesions the molecular analysis for the PVL genes detection was performed. The patient was discharged with a 10-day treatment of linezolid 600 mg twice a day. Due to poor tolerance of the drug, an additional hospitalization was needed and the therapy was changed with intravenous daptomycin (500 mg every 24 h) associated with amoxicillin-clavulanate (1 g 3 times a day). Finally, the patient was discharged in good condition without the appearance of new lesions. Given the peculiar clinical history of the patient we decided to analyze study the strain with the Whole Genome Sequencing (WGS) approach, investigating its possible origin and epidemiological links.
Both *S. aureus* strains collected from nasal (Sa_04-18_NC) and cutaneous (Sa_04-18_SI) swabs of the patient and from the other members of his family (wife and daughter) were characterized by phenotypic and molecular methods. The species identification and antimicrobial susceptibility profiles were accomplished by Vitek 2 system

Key words:
MRSA, *Staphylococcus aureus*, ST1153, *spa*-type t504, Panton-Valentine Leukocidin, Italy.

Corresponding author:
Sara Giordana Rimoldi
E-mail: sara.rimoldi@asst-fbf-sacco.it

(BioMérieux) (EUCAST breakpoints, (http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_8.1_Breakp oint_Tables.pdf). Furthermore, the presence of *mecA* and PVL genes was investigated using the RealCycler SAMAPV real-time Polymerase Chain Reaction (Progenie Molecular, Valencia, Spain).

Both Sa_04-18_SI and Sa_04-18_NC strains resulted susceptible to all antibiotics tested except for oxacillin (>2 mg/L), fusidic acid (>16 mg/L) and benzilpenicillin (>0.25 mg/L); the cefoxitin test was positive. Molecular typing revealed that Sa_04-18_SI and Sa_04-18_NC strains belonged to *spa-type* t504 and harbored the SCCmec type I genetic element, typical of hospital-acquired MRSA. The nasal swab from the patient's daughter was positive for a different strain, methicillin susceptible SA PVL-negative; while no colonization was found for his wife.

The Sa_04-18_SI strain was subjected to WGS typing. After DNA extraction with the UltraClean microbial DNA isolation kit (Mo Bio Laboratories, Carlsbad, CA, USA), the genomic libraries were sequenced on a Pgm apparatus (Ion Torrent) with a 2 by 250 paired-end run. The genome assembly was obtained using SPAdes (Bankevich *et al.*, 2012) and in silico characterized. All the results of SCCmecFinder (Kaya *et al.*, 2018), spaTyper 1.0 (Bartels *et al.*,

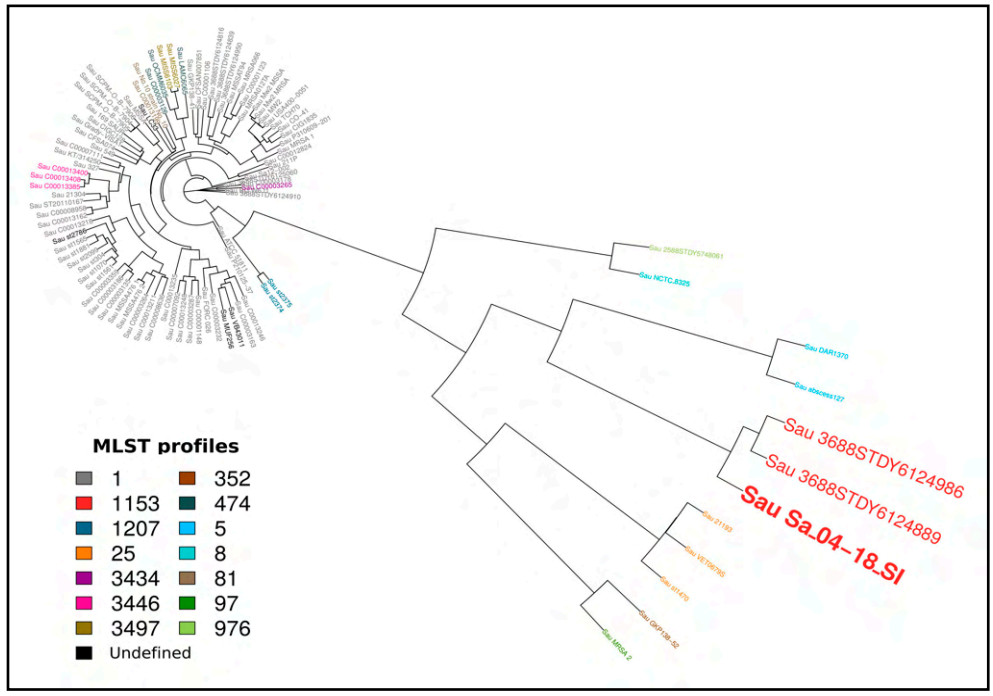


Figure 1 - *Phylogenetic reconstruction.*
Phylogenetic reconstruction of the relationships between the studied isolate (Sa_04-18_SI) and a dataset of 100 genomes selected to be the most similar in PATRIC database. MLST profiles of the isolates are mapped on the tree using the color code reported in the legend.

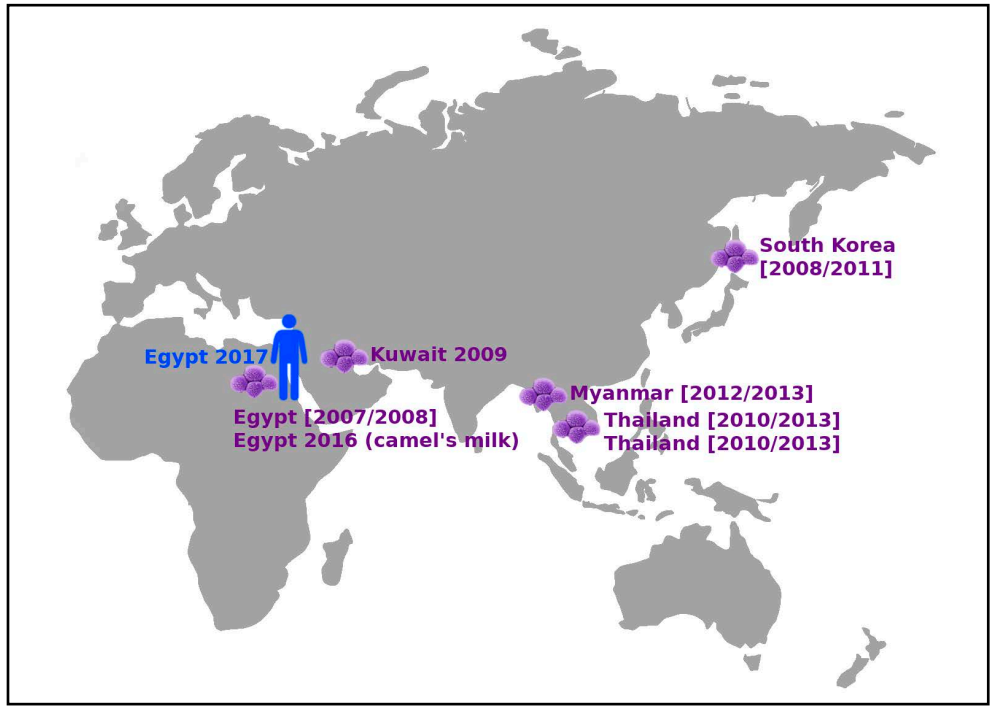


Figure 2 - *Geographical distribution of ST1153 Staphylococcus aureus.*
The map represents the sampling sites and the collection years of the ST1153 S. aureus isolates reported in literature and in this study. In details, the human icon refers to the studied isolate (Sa_04-18_SI), sampled in Egypt in 2017. The remaining isolates are represented with violet S. aureus shaped icons. When the exact sampling year was not available in literature, the sampling period was inferred from the paper and reported in brackets. All the strains derived from human samples, with the exception of one isolated from camel's milk.

2014) and PVL *(lukS* and *lukF*) genes BLAST (Altschul *et al*., 1990) searches confirmed the molecular data. Furthermore, the in silico MLST determination, accomplished by MLST 2.0 tool (Larsen *et al*., 2012), assigned the strain to the Sequence Type (ST) 1153, sporadically isolated in the Mediterranean area. Finally, a phylogenetic analysis was performed on a dataset including the Sa_04-18_SI genome, the *S. aureus* reference strain NCTC_8325 and 100 genomes from the PATRIC database (Wattam et al. 2017) selected to be the most genetically similar to Sa_04-18_SI, on the basis of the Mash distance (Ondov *et al*., 2016). Selected genomes were subjected to SNP calling analysis (Gaiarsa *et al*., 2015) using the NCTC_8325 strain as reference. Furthermore, an SNPs-based phylogenetic reconstruction was performed by RAxML software (Stamatakis 2014) with 100 pseudo bootstrap replicates, setting the best evolutionary model selected by ModelTest-NG (Darriba *et al*., 2017). SNP calling produced an alignment of 43,122 bases; the best phylogenetic model found was TVM and the obtained phylogenetic tree (*Figure 1*) clustered the Sa_04-18_SI strain to the only two ST1153 strains present in PATRIC database.

DISCUSSION

In this work we describe the first ST1153 PVL-producing MRSA strain, isolated from recurrent human skin infections, eventually resolved after an appropriate PVL diagnosis and the consequent antibiotic therapy.

Unfortunately, the patient's long (10-year) clinical history and the lack of previous microbiological investigations made it impossible to date the initial acquisition of the ST1153 MRSA strain. The main risk factors for the onset of the infection might be the hospitalization in Belgium and the patient's profession as a diver.

The ST1153 PVL-producing MRSA has already been sporadically described in Asia (Aung et al. 2016) and Africa (*Figure 2*), from both human samples and camel's milk in Egypt (Ali *et al*., 2017), but never in Europe. We had no evidences for the intrafamilial transmission of the strain, possibly due to the fact that all the members of the family were adults (Cocchi *et al*., 2013). All these data led us to hypothesize that the ST1153 MRSA strain might be acquired from the environment, possibly from the seawater in Egypt, considering the patient's job. Indeed, PVL-producing MRSA strains isolated from seawater have been already reported in the literature (Plano *et al*., 2013).

To obtain a more detailed epidemiological reconstruction, the Sa_04-18_SI strain was included in a global *S. aureus* phylogeny. The tree clustered Sa_04-18_SI only with two strains belonging to the ST1153, harboring PVL genes and sampled in Thailand. With such a limited ST1153 dataset, this result cannot be considered as strong evidence for an epidemiological link among the three strains. The difficulties that we met in this epidemiological reconstruction reveal that a greater effort on WGS surveillance programs is necessary, particularly for non-predominant clones.

Acknowledgments

We thank the patient and his family, who collaborated with clinicians to provide the material needed for microbiological investigations.

Thanks to the Romeo ed Enrica Invernizzi Foundation.

References

Ali A.O., Mahmoud HYAH. (2017). Epidemiological studies based on multi-locus sequence typing genotype of methicillin susceptible Staphylococcus aureus isolated from camel's milk. *Onderstepoort J Vet Res.* **84**, e1-e5.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). Basic local alignment search tool. *J Mol Biol.* **215**, 403-410.

Aung M.S, Zi H., Nwe K.M., Maw W.W., Aung M.T., Min W.W., et al. (2016). Drug resistance and genetic characteristics of clinical isolates of staphylococci in Myanmar: high prevalence of PVL among methicillin-susceptible Staphylococcus aureus belonging to various sequence types. *New Microbes New Infect.* **10**, 58-65.

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

Bartels M.D., Petersen A., Worning P., Nielsen J.B., Larner-Svensson H., Johansen H.K., et al. (2014). Comparing whole-genome sequencing with Sanger sequencing for spa typing of methicillin-resistant Staphylococcus aureus. *J Clin Microbiol.* **52**, 4305-4308.

Cocchi P., Taccetti G., Montagnani C., Campana S., Galli L., Braggion C., de Martino M. (2013). Evidence of transmission of a Panton-Valentine leukocidin-positive community-acquired methicillin-resistant Staphylococcus aureus clone: a family affair. *Clin Microbiol Infect.* **19**, 1158-1162.

Darriba D., Posada D., Stamatakis A. (2017). ModelTest-NG: Best-fit evolutionary model selection. https://github.com/ddarriba/modeltest. Website. Accessed November 20, 2017.

DeLeo F.R., Kennedy A.D., Chen L., Bubeck Wardenburg J., Kobayashi S.D., Mathema B., et al. (2011). Molecular differentiation of historic phage-type 80/81 and contemporary epidemic Staphylococcus aureus. *Proc Natl Acad Sci USA.* **108**, 18091-18096.

El-Shenawy, Mohamed. (2004). Staphylococcus aureus and fecal indicators in Egyptian coastal waters of Aqaba Gulf, Suez Gulf, and Red Sea. *Egypt J Aquatic Res.* **31**, 113e124.

Gaiarsa S., Comandatore F., Gaibani P., Corbella M., Dalla Valle C., Epis S., et al. (2015). Genomic epidemiology of Klebsiella pneumoniae in Italy and novel insights into the origin and global evolution of its resistance to carbapenem antibiotics. *Antimicrob Agents Chemother.* **59**, 389-396.

Goodwin K.D., Pobuda M. (2009). Performance of CHROMagar Staph aureus and CHROMagar MRSA for detection of Staphylococcus aureus in seawater and beach sand-comparison of culture, agglutination, and molecular analyses. *Water Res.* **43**, 4802-4811.

Kaya H., Hasman H., Larsen J., Stegger M., Johannesen T.B., Allesøe R.L., et al. (2018). SCCmecFinder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome mec in Staphylococcus aureus Using Whole-Genome Sequence Data. mSphere. **3**, pii: e00612-17.

Larsen M.V., Cosentino S., Rasmussen S., Friis C., Hasman H., Marvig R.L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol.* **50**, 1355-1361.

Levin-Edens E., Soge O.O., No D., Stiffarm A., Meschke J.S., Roberts M.C. (2012). Methicillin-resistant Staphylococcus aureus from Northwest marine and freshwater recreational beaches. *FEMS Microbiol Ecol.* **79**, 412-420.

Niemann S., Bertling A., Brodde M.F., Fender A.C., Van de Vyver H., Hussain M., et al. (2018). Panton-Valentine Leukocidin associated with S. aureus osteomyelitis activates platelets via neutrophil secretion products. *Sci Rep.* **8**, 2185.

Ondov B.D., Treangen T.J., Melsted P., Mallonee A.B., Bergman N.H., Koren S., Phillippy A.M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. **17**, 132.

Plano L.R., Shibata T., Garza A.C., Kish J., Fleisher J.M., Sinigalliano C.D., et al. (2013). Human-associated methicillin-resistant Staphylococcus aureus from a subtropical recreational marine beach. *Microb Ecol.* **65**, 1039-1051.

Stamatakis A. (2014). RAxML version 8. A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. **30**, 1312-1313.

Tice A.D., Pombo D., Hui J., Kurano M., Bankowski M.J., Seifried S.E. (2010). Quantitation of Staphylococcus aureus in seawater using CHROMagar SA. *Hawaii Med J.* **69**, 8-12.

Vignaroli C. (2009). Methicillin-resistant Staphylococcus aureus USA400 Clone, Italy. *Emerg Infect Dis.* **15**, 995-996.

Wattam A.R., Davis J.J., Assaf R., Boisvert S., Brettin T., Bun C., et al. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* **45**, D535-D542.

Zetola N., Francis J.S., Nuermberger E.L., Bishai W.R. (2005). Community-acquired meticillin-resistant Staphylococcus aureus: an emerging threat. *Lancet Infect Dis.* **5**, 275-286.

Section B

Appendix 5

Whole-Genome Sequencing Investigation of a Large Nosocomial Outbreak Caused by ST131 H30Rx KPC-Producing *Escherichia coli* in Italy.

Piazza A, Principe L, Comandatore F, Perini M, Meroni E, Mattioni Marchetti V, Migliavacca R, Luzzaro F.

Antibiotics (Basel). 2021 Jun 15;10(6):718. doi: 10.3390/antibiotics10060718

antibiotics

MDPI

Article

Whole-Genome Sequencing Investigation of a Large Nosocomial Outbreak Caused by ST131 H30Rx KPC-Producing *Escherichia coli* in Italy

Aurora Piazza [1,*,†], Luigi Principe [2,†], Francesco Comandatore [3], Matteo Perini [3], Elisa Meroni [4], Vittoria Mattioni Marchetti [5], Roberta Migliavacca [1] and Francesco Luzzaro [4]

1 Department of Clinical-Surgical, Diagnostic and Pediatric Sciences, Unit of Microbiology and Clinical Microbiology, University of Pavia, 27100 Pavia, Italy; roberta.migliavacca@unipv.it
2 Clinical Pathology and Microbiology Unit, S. Giovanni di Dio Hospital, 88900 Crotone, Italy; luigi.principe@gmail.com
3 Romeo and Enrica Invernizzi Pediatric Research Center, Department of Biomedical and Clinical Sciences L. Sacco, University of Milan, 20157 Milan, Italy; francesco.comandatore@unimi.it (F.C.); matteo.perini@unimi.it (M.P.)
4 Microbiology and Virology Unit, A. Manzoni Hospital, 23900 Lecco, Italy; el.meroni@asst-lecco.it (E.M.); f.luzzaro@asst-lecco.it (F.L.)
5 Biomedical Center, Faculty of Medicine in Pilsen, Charles University, 323 00 Pilsen, Czech Republic; vittoria.mattionimarche01@universitadipavia.it
* Correspondence: aurora.piazza@unipv.it
† These authors contributed equally to this work.

Citation: Piazza, A.; Principe, L.; Comandatore, F.; Perini, M.; Meroni, E.; Mattioni Marchetti, V.; Migliavacca, R.; Luzzaro, F. Whole-Genome Sequencing Investigation of a Large Nosocomial Outbreak Caused by ST131 H30Rx KPC-Producing *Escherichia coli* in Italy. *Antibiotics* **2021**, *10*, 718. https://doi.org/10.3390/antibiotics10060718

Academic Editor: Jesus Simal-Gandara

Received: 27 April 2021
Accepted: 11 June 2021
Published: 15 June 2021

Abstract: KPC-producing *Escherichia coli* (KPC-Ec) remains uncommon, being mainly reported as the cause of sporadic episodes of infection rather than outbreak events. Here we retrospectively describe the dynamics of a large hospital outbreak sustained by KPC-Ec, involving 106 patients and 25 hospital wards, during a six-month period. Twenty-nine representative KPC-Ec isolates (8/29 from rectal swabs; 21/29 from other clinical specimens) have been investigated by Whole-Genome Sequencing (WGS). Outbreak isolates showed a multidrug-resistant profile and harbored several resistance determinants, including $bla_{\text{CTX-M-27}}$, *aad*A5, *dfr*A17, *sul*I, *gyr*A1AB and *par*C1aAB. Phylogenomic analysis identified the ST131 cluster 1 (23/29 isolates), H30Rx clade C, as responsible for the epidemic event. A further two KPC-Ec ST131 clusters were identified: cluster 2 (n = 2/29) and cluster 3 (n = 1/29). The remaining KPC-Ec resulted in ST978 (n = 2/29) and ST1193 (n = 1/29), and were $bla_{\text{KPC-3}}$ associated. The KPC-Ec ST131 cluster 1, originated in a previous KPC-Kp endemic context probably by plasmid transfer, and showed a clonal dissemination strategy. Transmission of the bla_{KPC} gene to the globally disseminated high-risk ST131 clone represents a serious cause of concern. Application of WGS in outbreak investigations could be useful to better understand the evolution of epidemic events in order to address infection control and contrast interventions, especially when high-risk epidemic clones are involved.

Keywords: KPC-producing; *Escherichia coli*; ST131; WGS; outbreak; H30Rx

1. Introduction

The ongoing rise of carbapenemase-producing *Enterobacterales* (CPE) represents an important threat to public health worldwide, in both healthcare and community settings. The pervasive dissemination of CPE substantially impacts on patient safety since few therapeutic alternatives remain. Overall, the worldwide predominant carbapenemase is the Ambler class A *Klebsiella pneumoniae* carbapenemase (KPC) enzyme, encoded by alleles of the bla_{KPC} gene, with KPC-producing *K. pneumoniae* (KPC-Kp) being the most common among KPC-producing CPE. Although less prevalent than in *K. pneumoniae*, KPC production in other species belonging to *Enterobacterales* is increasingly reported [1].

The presence of this highly transferable carbapenemase in *Escherichia coli* is of particular concern, being the most common community- and hospital-acquired pathogen, that can be transmitted among humans, animals and in the environment [2]. Moreover, the recent detection of the bla_{KPC} gene in *E. coli* belonging to sequence type (ST) 131, a globally disseminated successful clone, represents a cause of serious concern [3]. In fact, the ST131 clone emerged as the most common extraintestinal pathogen that, in association with fluoroquinolone and extended-spectrum cephalosporin resistance, was responsible for the worldwide spread of the extended-spectrum beta-lactamase (ESBL) $bla_{CTX\text{-}M\text{-}15}$ gene. Due to its ability to asymptomatically colonize the gastrointestinal tract of both community and healthcare-associated infections, a stable association of ST131 lineage with bla_{KPC} could have relevant consequences for the management of *E. coli* infections [4].

KPC-producing *E. coli* (KPC-Ec) isolates were mainly reported in countries showing a high prevalence of KPC-Kp, probably reflecting a spill-over of resistance genes from the *K. pneumoniae* reservoir [5]. However, acquired carbapenem resistance in *E. coli* is still considered a rare and recent event. The first cases of KPC-Ec were observed in 2004–2005 in the USA (Cleveland, n = 1; New York City, n = 2; New Jersey, n = 1), and in Israel (Tel Aviv, n = 4) [3]. After the first detection in Europe—reported in 2008 in France from a patient initially hospitalized in Israel [6]—KPC-Ec has been sporadically reported in the USA [7], Israel [8] and some European countries [9,10].

In Italy, KPC-Kp has been endemic since 2013, but the presence of KPC-Ec isolates remains limited [11]. Results of the most recent Italian Nationwide survey highlighted that KPC was the most frequent carbapenemase from bloodstream infections, mainly in *K. pneumoniae* (95.2%). A recent epidemiological study showed that KPC-Ec were only 1.3% of KPC-producing invasive isolates, but they accounted for 81.4% of cabapenem-resistant *E. coli*, thus highlighting the propensity of these strains to cause invasive infections [12].

Outbreaks caused by KPC-Ec have been described in the USA, Greece, Canada, Israel, Italy and the UK [2,13–17]. Despite this, KPC-Ec has been more frequently reported in sporadic cases than as the cause of outbreaks [18]. In fact, unlike *K. pneumoniae*, in which the KPC determinant is often associated to predominant plasmids and epidemic clones (e.g., IncFII, -FIA, -I2 and ST258), the sporadic appearance of KPC-producing *E. coli* belonging to different genetic backgrounds is largely due to horizontal transfer of different bla_{KPC}-harboring plasmid groups, thus suggesting a "less successful" association enzyme-pathogen than in KPC-Kp [19].

It is of particular concern that ST131 KPC-Ec strains able to cause outbreaks are emerging, as suggested by the recent reports from Israel, Italy and UK [13,14].

The aim of our study was to report an epidemiological and genomic (Whole-Genome Sequencing, WGS) investigation of a large nosocomial outbreak caused by KPC-Ec, which occurred in Northern Italy in 2016.

2. Results

2.1. Bacterial Isolates and Phenotypic Characterization

From February to July 2016, 123 KPC-*Escherichia coli* (KPC-Ec) isolates were collected from 106 patients. Twenty-five hospital wards were involved in the outbreak, with those of Internal Medicine (n = 34 patients), Cardiology Rehabilitation (n = 17), Cardiology (n = 10) and Nephrology (n = 9) being the most represented. From August 2016 to June 2017, eight additional KPC-Ec were isolated from rectal swabs (n = 7) and respiratory secretions (n = 1) of seven patients. Four of the above isolates were included in the WGS analysis for comparison. Twenty out of 106 patients belonging to the outbreak period were previously colonized by KPC-*Klebsiella pneumoniae* (KPC-Kp), while in 23 cases the detection of both species was associated, indicating a co-presence at the intestinal level. KPC-Ec was isolated prior to KPC-Kp in seven patients. Six KPC-Kp included in the study were isolated from different wards. Among them, four were from rectal swabs whereas the remaining two were from urine samples. According to both the MicroScan autoSCAN-4 semi-automated system and Sensititre broth microdilution results, representative isolates were consistently

resistant to amoxicillin/clavulanate, piperacillin/tazobactam, cefotaxime, ceftazidime, aztreonam, ertapenem, and ceftolozane/tazobactam, but susceptible to colistin, tigecycline and ceftazidime-avibactam. Detailed data concerning the susceptibility results of the 29 KPC-Ec representative isolates are reported in Table 1. Of note, MIC values for imipenem ranged from 4 to >8 mg/L, whereas MICs of meropenem ranged from 4 to 8 mg/L.

Table 1. Antimicrobial susceptibility profiles of selected KPC-positive *E. coli* strains.

ID Strain	Antimicrobial Susceptibility (MIC, µg/mL)																
	AMC	PTZ	CTX	CAZ	CZA	C/T	MER	IMI	ERT	AMK	GNT	AZT	CIP	TBR	TIG	SXT	COL
sk46y46t	>32 (R)	>16 (R)	>32 (R)	16 (R)	<0,5/4 (S)	4/4 (R)	4 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	<0,25 (S)
sk36y36t	>32 (R)	>16 (R)	>32 (R)	32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	4 (I)	≤1 (S)	>4/76 (R)	<0,25 (S)
sk35y35t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	>4 (R)	>4 (R)	>1 (R)	>4 (R)	≤1 (S)	≤2/38 (S)	0,5 (S)
sk39y39t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk37y37t	>32 (R)	>16 (R)	>32 (R)	8 (R)	<0,5/4 (S)	2/4 (R)	4 (I)	4 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk38y38t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk42y42t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	16 (I)	≤2 (S)	>4 (R)	>1 (R)	4 (I)	≤1 (S)	>4/76 (R)	0,5 (S)
sk40y40t	>32 (R)	>16 (R)	>32 (R)	32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	4 (I)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk47y47t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	>8 (R)	>1 (R)	16 (I)	≤2 (S)	>4 (R)	>1 (R)	4 (I)	≤1 (S)	>4/76 (R)	0,5 (S)
sk41y41t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	16/4 (R)	8 (I)	8 (I)	>1 (R)	>16 (R)	4 (I)	>4 (R)	≤0,06 (S)	>4 (R)	≤1 (S)	≤2/38 (S)	<0,25 (S)
sk43y43t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	32/4 (R)	8 (I)	8 (I)	>1 (R)	>16 (R)	4 (I)	>4 (R)	≤0,06 (S)	>4 (R)	≤1 (S)	≤2/38 (S)	0,5 (S)
sk54y54t	>32 (R)	>16 (R)	>32 (R)	32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	>8 (R)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	4 (I)	≤1 (S)	≤2/38 (S)	0,5 (S)
sk44y44t	>32 (R)	>16 (R)	>32 (R)	16 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk45y45t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk48y48t	>32 (R)	>16 (R)	>32 (R)	16 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	4 (I)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	<0,25 (S)
sk49y49t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	>8 (R)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	≤2/38 (S)	<0,25 (S)
sk51y51t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	<0,25 (S)
sk50y50t	>32 (R)	>16 (R)	>32 (R)	16 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	4 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk52y52t	>32 (R)	>16 (R)	>32 (R)	16 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	4 (I)	>4 (R)	>1 (R)	4 (I)	≤1 (S)	>4/76 (R)	<0,25 (S)
sk53y53t	>32 (R)	>16 (R)	>32 (R)	32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk56y56t	>32 (R)	>16 (R)	>32 (R)	16 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk55y55t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	16/4 (R)	4 (I)	4 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	≤2/38 (S)	0,5 (S)
sk57y57t	>32 (R)	>16 (R)	>32 (R)	8 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	<0,25 (S)
sk58y58t	>32 (R)	>16 (R)	>32 (R)	8 (R)	<0,5/4 (S)	2/4 (R)	4 (I)	4 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk59y59t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk60y60t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	32/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	1 (S)
sk185y185t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk136y136t	>32 (R)	>16 (R)	>32 (R)	32 (R)	<0,5/4 (S)	2/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)
sk137y137t	>32 (R)	>16 (R)	>32 (R)	>32 (R)	<0,5/4 (S)	4/4 (R)	8 (I)	8 (I)	>1 (R)	<8 (S)	≤2 (S)	>4 (R)	>1 (R)	≤2 (S)	≤1 (S)	>4/76 (R)	0,5 (S)

MIC: minimum inhibitory concentration; AMC: amoxicillin/clavulanate; PTZ: piperacillin/tazobactam; CTX: cefotaxime; CAZ: ceftazidime; MER: meropenem; IMI: imipenem; ERT: ertapenem; AMK: amikacin; GNT: gentamicin; AZT: aztreonam; CIP: ciprofloxacin; TBR: tobramycin; TIG: tigecycline; SXT: trimethoprim-sulfamethoxazole; CZA: ceftazidime/avibactam; C/T: ceftolozane/tazobactam; COL: colistin; S: susceptible; I: intermediate; R: resistant. Susceptibility results were interpreted according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST, 2016) criteria. The "index strain" is indicated in bold.

2.2. Epidemiological Context

Epidemiological surveillance data showed that the epidemic event followed a previous KPC-Kp outbreak, starting in October 2015 and partially overlapping the KPC-Ec outbreak (Supplementary Figure S1). The prevalence of KPC-Ec increased from 0% in 2014 and 2015 to 2.8% in 2016. This trend seems to be related to the increase of KPC-Kp prevalence, from 12.4% in 2014, to 17.3% in 2015 and to 23.1% in 2016 (Supplementary Figure S1). Notably, 50 (47.2%) out of 106 patients' results were positive to both KPC-Ec and KPC-Kp during previous hospitalizations. In particular, 20 out of 50 patients (40%) were previously colonized by KPC-Kp, while in 23 cases (46%) the isolation of both species was concomitant, indicating the co-presence at the intestinal level. Only in seven patients (14%), however, was KPC-Ec isolated previously with respect to KPC-Kp.

2.3. Whole-Genome Sequencing Characterization

Twenty-nine KPC-Ec isolates were further investigated by whole-genome sequencing. They were from urine (n = 11), rectal swabs (n = 8), blood (n = 2), purulent exudate (n = 2), respiratory secretions (n = 3), drainage fluid (n = 1), peritoneal fluid (n = 1) and a surgical wound swab (n = 1) (Table 2).

Twenty-six out of 29 isolates belonged to ST131 while the remaining three isolates were of ST978 (n = 2) and ST1193 (n = 1). ST131 isolates carried the following resistance determinants: *aadA5*, *aph(6)-Id*, *aph(3")-Ib*, $bla_{\text{CTX-M-type}}$, $bla_{\text{KPC-type}}$, $bla_{\text{OXA-9}}$, $bla_{\text{TEM-1A}}$, *mph(A)*, *sul1*, *sul2*, *tet(A)* and *dfrA17* (Table 3). According to the observed quinolone resistance phenotype (see Table 3), ST131 isolates harbored the *parC*1aAB and *gyrA*1AB gene variants. Twenty-four ST131 strains harbored the $bla_{\text{KPC-2}}$ gene, while only two strains' results were $bla_{\text{KPC-3}}$ positive. As shown in Table 3, all but one ST131 strain harbored the $bla_{\text{CTX-M-27}}$, whereas the remaining one was positive for $bla_{\text{CTX-M-15}}$. All the ST131 strains showed the O25b:H4 serotype and the fimbrial variant *fimH*30. Both the isolates belonging to ST978 carried the $bla_{\text{KPC-3}}$ determinant, and showed the serotype O83:H27, and the *fimH*2 variant. The ST1193 isolate harbored the $bla_{\text{KPC-3}}$ gene, and showed the serotype O75:H5, and the *fimH64* variant.

The virulence genes *sat* (secret autotransporter toxin), *iss* (increased serum survival) and *gad* (glutamate decarboxylase) were found in all ST131 strains, while the *iha* (adherence protein) determinant was found in all isolates but one ST131. The *senB* (enterotoxin) gene was present in the majority of ST131 isolates, while the *cnf1* (cytotoxic necrotizing factor type 1) was only identified in the isolate sk35y35t, which was the first KPC-Ec that emerged in the outbreak period. A different pattern of virulence genes was detected in the two ST978 strains, consisting of *vat* (vacuolating autotransporter toxin), *pic* (serine protease) and *gad* determinants. Lastly, the ST1193 isolate harbored *iha*, *sat* and *vat* as virulence genes (Table 3).

The comparison of the 29 KPC-Ec isolates with other *E. coli* genomes deposited in the PATRIC database is shown in Figure 1. Notably, most of them showed a major similarity with KPC-Ec strains collected in the UK.

Table 2. Main metadata of the 35 clinical samples included in the study and subjected to whole-genome sequencing. The "index strain" is indicated in bold.

ID Sample	Microorganism	Ward	Isolation Date	Material
sk35y35t	*Escherichia coli*	Infectious Diseases	15 February 2016	Rectal swab
sk36y36t	*Escherichia coli*	Nephrology	16 February 2016	Rectal swab
sk37y37t	*Escherichia coli*	Medicine	26 February 2016	Urine
sk38y38t	*Escherichia coli*	Nephrology	29 February 2016	Rectal swab
sk39y39t	*Escherichia coli*	Cardiology Rehabilitation	1 March 2016	Urine
sk40y40t	*Escherichia coli*	Neurointensive Care	3 March 2016	Purulent exudate
sk42y42t	*Escherichia coli*	Dermatology	3 March 2016	Urine
sk41y41t	*Escherichia coli*	Medicine I Soap	8 March 2016	Blood
sk43y43t	*Escherichia coli*	Medicine	17 March 2016	Rectal swab
sk44y44t	*Escherichia coli*	Nephrology	23 March 2016	Blood
sk45y45t	*Escherichia coli*	Cardiology Rehabilitation	5 April 2016	Urine
sk46y46t	*Escherichia coli*	Cardiosurgery	18 April 2016	Bronchial aspirate
sk47y47t	*Escherichia coli*	Medicine I Soap	18 April 2016	Urine
sk48y48t	*Escherichia coli*	Nephrology	21 April 2016	Surgical wound swab
sk49y49t	*Escherichia coli*	Long-term Surgery II	24 April 2016	Peritoneal fluid
sk50y50t	*Escherichia coli*	Medicine II	10 May 2016	Respiratory secretion
sk51y51t	*Escherichia coli*	Cardiology Rehabilitation	13 May 2016	Urine
sk52y52t	*Escherichia coli*	Medicine I Soap	22 May 2016	Urine
sk53y53t	*Escherichia coli*	Medicine II men	11 June 2016	Urine
sk54y54t	*Escherichia coli*	Medicine II men	25 June 2016	Urine
sk55y55t	*Escherichia coli*	Surgery	29 June 2016	Drainage fluid
sk56y56t	*Escherichia coli*	Cardiology	30 June 2016	Urine
sk57y57t	*Escherichia coli*	Medicine soap	13 July 2016	Urine
sk58y58t	*Escherichia coli*	Medicine soap	13 July 2016	Rectal swab
sk59y59t	*Escherichia coli*	Medicine soap	14 July 2016	Purulent exudate
sk60y60t	*Escherichia coli*	Intensive care	30 September 2016	Rectal swab
sk185y185t	*Escherichia coli*	Infectious Diseases	14 November 2016	Respiratory secretion
sk136y136t	*Escherichia coli*	Nephrology	23 January 2017	Rectal swab
sk137y137t	*Escherichia coli*	Surgery	5 June 2017	Rectal swab
sk138y138t	*Klebsiella pneumoniae*	Nephrology	22 January 2016	Urine
sk139y139t	*Klebsiella pneumoniae*	Cardiology	16 February 2016	Rectal swab
sk140y140t	*Klebsiella pneumoniae*	Rehabilitation Medicine	25 February 2016	Rectal swab
sk141y141t	*Klebsiella pneumoniae*	Urology	25 February 2016	Urine
sk142y142t	*Klebsiella pneumoniae*	Infectious Diseases	22 March 2016	Rectal swab
sk143y143t	*Klebsiella pneumoniae*	Medicine	31 March 2016	Rectal swab

Table 3. Genomic characteristics of the sequenced isolates. The "index strain" is indicated in bold.

ID Sample	Isolation Date	MLST	Cluster	Serotype	*fimH*	Resistance Genes	Virulence Genes	Plasmid Incompatibility Groups
sk35y35t	15 February 2016	131	ST131 cluster 3	O25b:H4	30	*ampH*, bla_{TEM-90}, bla_{KPC-2},*blaOXA-9*, $bla_{CTX-M-15}$, *ampC2*, *catB4*, bla_{OXA-1}, *aac(3)-IIa*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *cnf1*, *gad*	IncFIB(pQil)_1_pQil_JN233705
sk36y36t	16 February 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-122}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk37y37t	26 February 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, bla_{TEM-90}, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk38y38t	29 February 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-122}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk39y39t	1 March 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-141}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk40y40t	3 March 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-122}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk42y42t	3 March 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, bla_{TEM-79}, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk41y41t	8 March 2016	978	ST978	O83:H27	2	*aac(6)-Ib*, *ampH*, *ampC2*, bla_{KPC-3}	*vat*, *pic*, *gad*	IncX3_1__JN247852
sk43y43t	17 March 2016	978	ST978	O83:H27	2	*aac(6)-Ib*, *ampH*, *ampC2*, bla_{KPC-3}	*vat*, *pic*, *gad*	IncX3_1__JN247852
sk44y44t	23 March 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-122}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk45y45t	5 April 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, bla_{TEM-79}, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk46y46t	18 April 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-156}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk47y47t	18 April 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, bla_{TEM-54}, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk48y48t	21 April 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-192}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk49y49t	24 April 2016	131	ST131 cluster 1	O25b:H4	30	*ampH*, $bla_{TEM-122}$, bla_{KPC-2}, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk50y50t	10 May 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-122}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781

Table 3. *Cont.*

ID Sample	Isolation Date	MLST	Cluster	Serotype	*fimH*	Resistance Genes	Virulence Genes	Plasmid Incompatibility Groups
sk51y51t	13 May 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, bla_{TEM-79}, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk52y52t	22 May 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-150}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk53y53t	11 June 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-168}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk54y54t	25 June 2016	131	ST131 cluster 1	O25b:H4	30	*ampH*, $bla_{TEM-122}$, bla_{KPC-2}, *bla*OXA-9, $bla_{CTX-M-27}$, *ampC2*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705
sk55y55t	29 June 2016	131	ST131 cluster 2	O25b:H4	30	*ampH*, bla_{TEM-79}, bla_{KPC-3}, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *parC1aAB*, *gyrA1AB*	*sat*, *iss*, *gad*	IncFIB(pQil)_1_pQil_JN233705
sk56y56t	30 June 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, bla_{TEM-79}, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781, Col(BS512)_1_NC_010656
sk57y57t	13 July 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-168}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk58y58t	13 July 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-141}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk59y59t	14 July 2016	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, bla_{TEM-79}, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk60y60t	30 September 2016	131	ST131 cluster 2	O25b:H4	30	*ampH*, bla_{TEM-79}, bla_{KPC-3}, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *parC1aAB*, *gyrA1AB*	*sat*, *iss*, *gad*	IncFIB(pQil)_1_pQil_JN233705
sk185y185t	14 November 2016	1193	ST1193	O75:H5	64	*qnrB1*, $bla_{TEM-198}$, bla_{KPC-3}, *ampC2*, *aac(3)-IId*, *dfrA14*	*iha*, *sat*, *vat*	IncFIB(pQil)_1_pQil_JN233705, Col(BS512)_1_NC_010656
sk136y136t	23 January 2017	131	ST131 cluster 1	O25b:H4	30	*strA*, *ampH*, $bla_{TEM-168}$, bla_{KPC-2}, *sulII*, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *aadA5*, *mphA*, *tetR*, *dfrA17*, *tetA*, *sulI*, *strB*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk137y137t	5 June 2017	131	ST131 cluster 1	O25b:H4	30	*ampH*, $bla_{TEM-150}$, bla_{KPC-2}, bla_{OXA-9}, $bla_{CTX-M-27}$, *ampC2*, *parC1aAB*, *gyrA1AB*	*iha*, *sat*, *iss*, *senB*, *gad*	IncFIB(pQil)_1_pQil_JN233705, Col156_1_NC_009781
sk138y138t	22 January 2016	35				*oqxBgb*, bla_{KPC-2}, bla_{TEM-79}, bla_{OXA-9}, *oqxA*, *tetD*, *ampH*		IncFIB(pQil)_1_pQil_JN233705, Inc-FIB(K)_1_Kpn3_JN233704
sk139y139t	16 February 2016	3033				*strA*, *strB*, *oqxBgb*, *catB4*, bla_{KPC-2}, $bla_{TEM-198}$, bla_{OXA-1}, bla_{OXA-9}, $bla_{CTX-M-15}$, *oqxA*, *dfrA14*, *sulII*, *ampH*		IncFIB(pQil)_1_pQil_JN233705, Inc-FIB(K)_1_Kpn3_JN233704
sk140y140t	25 February 2016	17				*oqxBgb*, bla_{KPC-2}, $bla_{TEM-122}$, bla_{OXA-9}, *oqxA*, *ampH*		IncFIB(pQil)_1_pQil_JN233705, Inc-FIB(K)_1_Kpn3_JN233704
sk141y141t	25 February 2016	35				*oqxBgb*, bla_{KPC-2}, $bla_{TEM-122}$, bla_{OXA-9}, *oqxA*, *tetD*, *ampH*		IncFIB(pQil)_1_pQil_JN233705, Inc-FIB(K)_1_Kpn3_JN233704

Table 3. *Cont.*

ID Sample	Isolation Date	MLST	Cluster	Serotype	*fimH*	Resistance Genes	Virulence Genes	Plasmid Incompatibility Groups
sk142y142t	22 March 2016	2279				*strA*, *strB*, *oqxBgb*, bla_{KPC-2}, $bla_{TEM-198}$, bla_{OXA-9}, *aac(3)-Iia*, $bla_{CTX-M-15}$, *oqxA*, *dfrA14*, *sulII*, *ampH*		IncFIB(pQil)_1_pQil_JN233705, Inc-FIB(K)_1_Kpn3_JN233704
sk143y143t	31 March 2016	17				*oqxBgb*, bla_{KPC-2}, $bla_{TEM-122}$, bla_{OXA-9}, *oqxA*, *ampH*		IncFIB(pQil)_1_pQil_JN233705, Inc-FIB(K)_1_Kpn3_JN233704

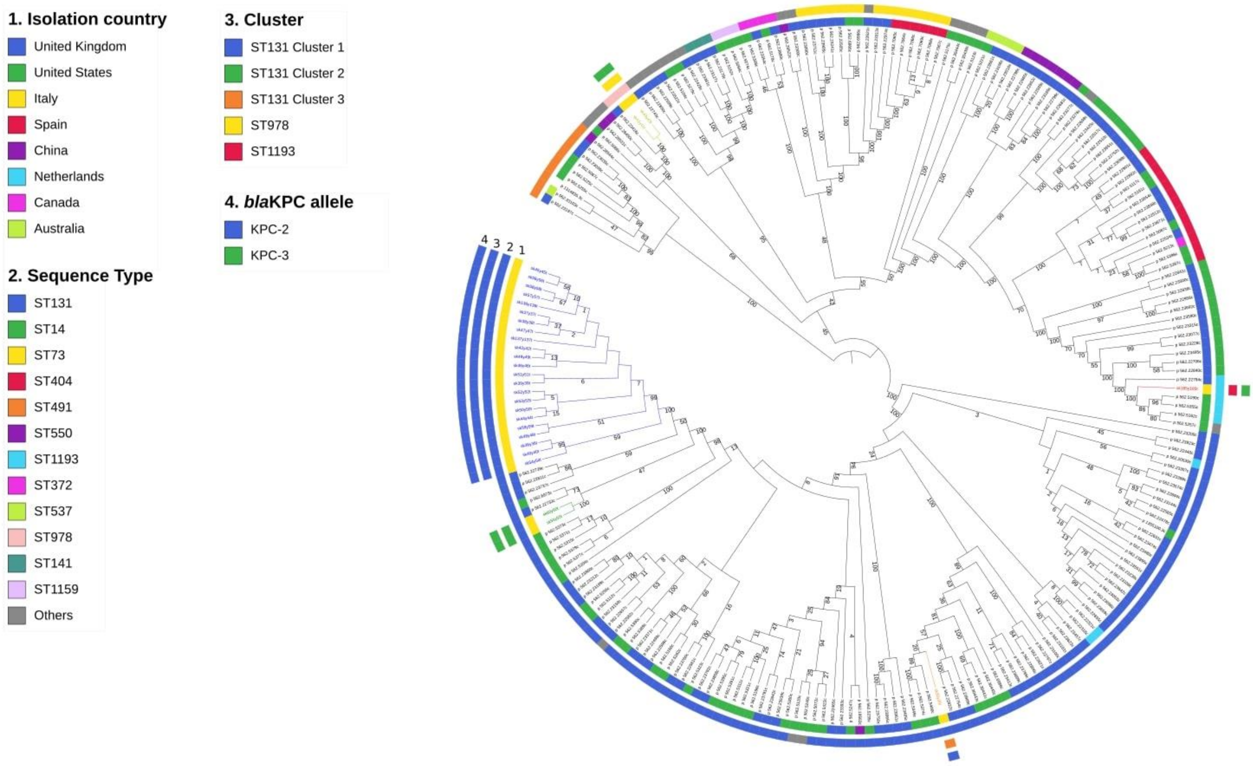


Figure 1. Maximum likelihood phylogenetic tree including the 29 KPC-producing *Escherichia coli* isolates and background strains retrieved from the PATRIC database. The clusters were identified as monophyletic highly supported groups and are reported on the colored inner ring. The geographic origin of all the strains is reported on the second level colored ring. The Sequence Type of the strains (Achtman scheme) is reported on the third colored ring, while the KPC variant is on the outer ring.

Figure 1. Maximum likelihood phylogenetic tree including the 29 KPC-producing *Escherichia coli* isolates and background strains retrieved from the PATRIC database. The clusters were identified as monophyletic highly supported groups and are reported on the colored inner ring. The geographic origin of all the strains is reported on the second level colored ring. The Sequence Type of the strains (Achtman scheme) is reported on the third colored ring, while the KPC variant is on the outer ring.

The phylogenetic reconstruction showed the presence of five distinct clusters, three of which belonged to ST131 ("ST131 cluster 1", "ST131 cluster 2" and "ST131 cluster 3"), one to ST978 and another to ST1193 (Figure 1). The ST131 cluster 1, characteristic of the majority of isolates (n = 26), was found in all wards with the exception of the Intensive Care and the Infectious Diseases Units (Figure 2), thus representing the outbreak epidemic clone.

Notably, sporadic isolates belonging to the ST131 cluster 1 were observed till June 2017. Based on phylogenetic data, the first isolate belonging to the ST131 cluster 1 (i.e., the outbreak index strain, coded sk36y36t) was isolated in the Nephrology Ward. The patient had been previously hospitalized for long periods in the past years in both Nephrology and ICU. Of note, the index patient was previously colonized by KPC-Kp (strain code, sk138y138t) (Table 2). The ST131 cluster 2 included two isolates, collected in Surgery and ICU in June and September 2016, respectively. The ST131 cluster 3 was represented by a single isolate, obtained in February 2016 from the Infectious Diseases Ward. ST978 strains were recovered in March 2016 from the blood culture and rectal swab of the same patient (Table 2 and Figure 2). Lastly, the ST1193 strain was isolated in November 2016 in the

of which belonged to ST131 ("ST131 cluster 1", "ST131 cluster 2" and "ST131 cluster 3"), one to ST978 and another to ST1193 (Figure 1). The ST131 cluster 1, characteristic of the majority of isolates (n = 26), was found in all wards with the exception of the Intensive Care and the Infectious Diseases Units (Figure 2), thus representing the outbreak epidemic clone.

Notably, sporadic isolates belonging to the ST131 cluster 1 were observed till June 2017. Based on phylogenetic data, the first isolate belonging to the ST131 cluster 1 (i.e., the outbreak index strain, coded sk36y36t) was isolated in the Nephrology Ward. The patient had been previously hospitalized for long periods in the past years in both Nephrology and ICU. Of note, the index patient was previously colonized by KPC-Kp (strain code, sk138y138t) (Table 2). The ST131 cluster 2 included two isolates, collected in Surgery and ICU in June and September 2016, respectively. The ST131 cluster 3 was represented by a single isolate, obtained in February 2016 from the Infectious Diseases Ward. ST978 strains were recovered in March 2016 from the blood culture and rectal swab of the same patient (Table 2 and Figure 2). Lastly, the ST1193 strain was isolated in November 2016 in the
Infectious Diseases Ward. Figure 2 shows the ward distribution of the KPC-Ec strains belonging to different clusters and STs.

Infectious Diseases Ward. Figure 2 shows the ward distribution of the KPC-Ec strains belonging to different clusters and STs.

Recent studies have focused on deciphering the genomic evolution and diversity within the ST131 lineage [4,20,21]. Since three different ST131 clusters were identified, we compared their genomes with those published by Petty and colleagues [20]. As shown in Figure 3, all the studied strains could be assigned to the ST131 clade C, characterized by the fimbrial variant *fimH*30Rx, and the *gyrA*1AB and *parC*1aAB alleles, associated with fluoroquinolone resistance. Phylogeny highlighted that the only difference was the presence of the $bla_{\text{CTX-M-27}}$ gene variant instead of $bla_{\text{CTX-M-15}}$ for ST131 clusters 1 and 3.

Antibiotics **2021**, *10*, x FOR PEER REVIEW 9 of 17

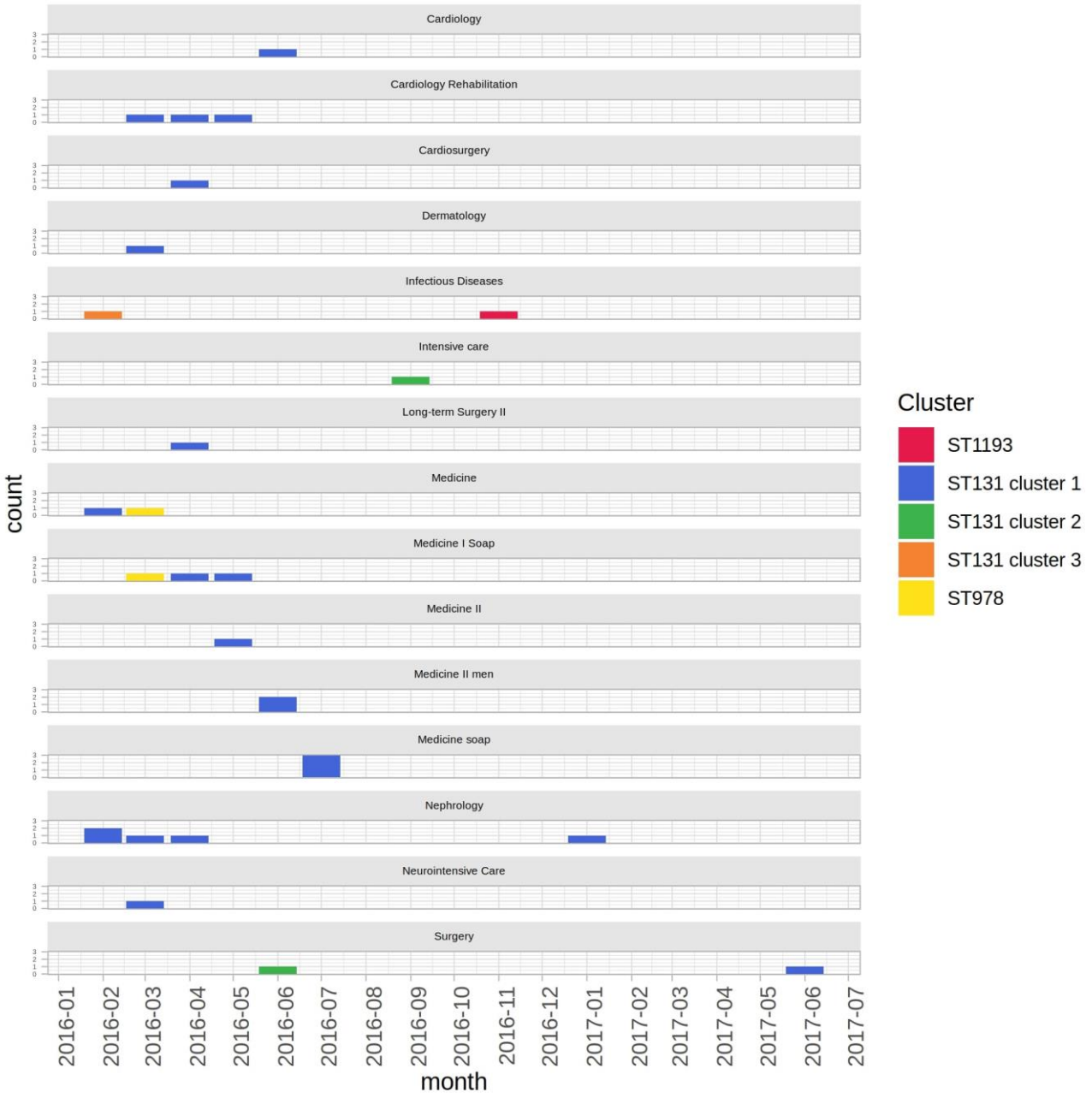


Figure 2. Barplot of the number of isolates per month per ward. Bars were colored on the basis of the phylogenetic clusters shown in Figure 1.

Recent studies have focused on deciphering the genomic evolution and diversity within the ST131 lineage [4,20,21]. Since three different ST131 clusters were identified, we compared their genomes with those published by Petty and colleagues [20]. As shown in Figure 3, all the studied strains could be assigned to the ST131 clade C, characterized by the fimbrial variant *fimH*30Rx, and the *gyrA*1AB and *parC*1aAB alleles, associated with fluoroquinolone resistance. Phylogeny highlighted that the only difference was the presence of the $bla_{\text{CTX-M-27}}$ gene variant instead of $bla_{\text{CTX-M-15}}$ for ST131 clusters 1 and 3.

Antibiotics **2021**, *10*, 718 10 of 17

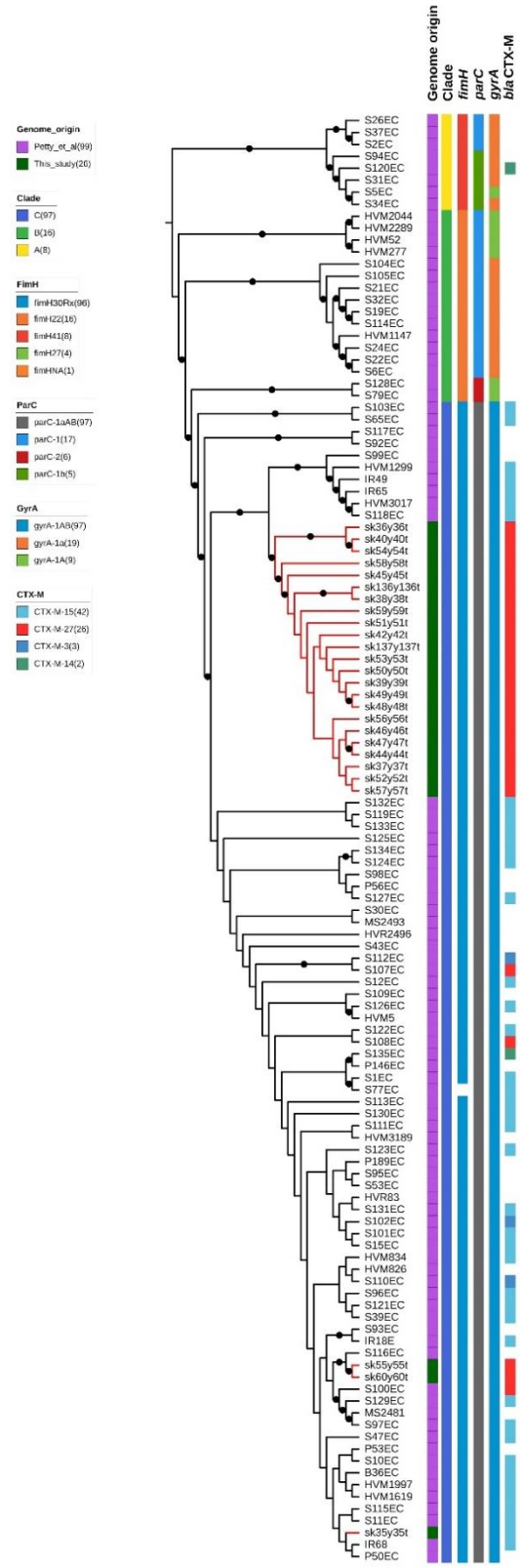


Figure 3. Maximum likelihood phylogenetic tree including the here studied ST131 strains and the ones from Petty *et al.* [20]. The branches corresponding to the strains of this study are highlighted in red. Metadata have been represented as follows: the first column indicates the origin of the strains; the second one the **ST131 clades** as called by Petty et al. [20]. The subsequent four columns report the allele variants of the *fimH*, *parC*, *gyrA* and *bla*CTX-M genes.

The six KPC-Kp strains, isolated during the early stages of the outbreak, were investigated for ST and plasmid incompatibility group arrangement. KPC-Kp strains belonged to ST35 (*n* = 2, from Nephrology and Urology), ST17 (*n* = 2, from the Rehabilitation Unit and Internal Medicine), ST3033 (*n* = 1, from Cardiology) and ST2279 (*n* = 1, from Infectious Diseases).

The positive colonization date (and co-isolation date, when it occurred) and related clinical samples of KPC-Ec and KPC-Kp isolates involved in WGS analyses are shown in **Supplemental Tables S1 and S2**.

Regarding plasmids, IncFIBpQil was found in all ST131 and ST1193 KPC-Ec strains, as well as in all the KPC-Kp strains. IncFIA was found in the ST131 cluster 1 and cluster 2 KPC-Ec strains, Col156 was found in the ST131 cluster 1 isolates, IncX3 was detected in both ST978 isolates. Lastly, all the KPC-Kp strains harbored IncFIB(K)_Kpn3. The KPC-Ec strains belonging to the epidemic ST131 of both cluster 1 and cluster 3 produced the KPC-2 enzyme, as well as the KPC-Kp isolates. The bla_{KPC-3} allele was instead found in the ST131 cluster 2, ST1193 and ST978 strains. Analysis of the bla_{KPC} surrounding genetic environment showed that all clusters, except ST978 and ST1193, were characterized by a conserved *blaKPC* scaffold (Figure 4). ST978 and ST1193 showed two different scaffolds. All the clusters harbored the Tn4401a, a structural variant of the Tn4401 transposon.

Species	ID	Cluster	blaKPC allele	Tn4401 variant
Kpn	sk142y142t	ST2279	bla_{KPC-2}	Tn4401a
Kpn	sk140y140t	ST17	bla_{KPC-2}	Tn4401a
Kpn	sk138y138t	ST35	bla_{KPC-2}	Tn4401a
Eco	sk36y36t	ST131 *cluster* 1	bla_{KPC-2}	Tn4401a
Eco	sk55y55t	ST131 *cluster* 2	bla_{KPC-3}	Tn4401a
Eco	sk35y35t	ST131 *cluster* 3	bla_{KPC-2}	Tn4401a
Eco	sk185y185t	ST1193	bla_{KPC-3}	Tn4401a
Eco	sk41y41t	ST978	bla_{KPC-3}	Tn4401a-unk

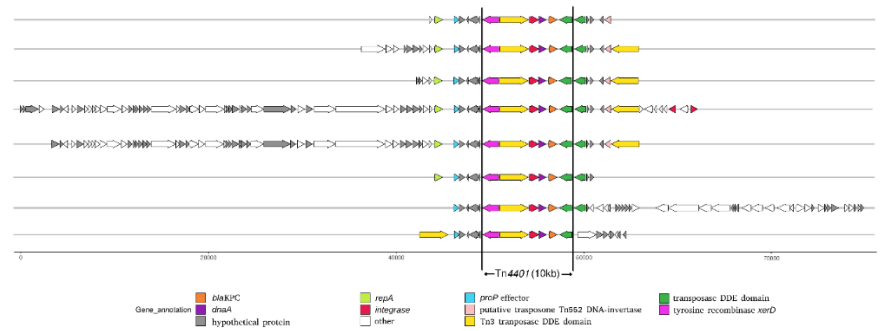


Figure 4. Arrow plot of the different bla_{KPC} gene contigs aligned on the Tn4401 transposon (shown inside the vertical lines). The plot includes one representative strain for each *Escherichia coli* cluster and one for each *Klebsiella pneumoniae* ST.

Given the presence of the IncFIBpQil, a well-known KPC-harboring plasmid, the detection of the KPC-2 enzyme, and evaluating the extreme similarity of the bla_{KPC} genomic environment in KPC-Kp and KPC-Ec isolates belonging to the ST131 cluster 1; a plasmid-mediated transmission of the KPC-2 determinant from KPC-Kp to KPC-Ec isolates could reasonably have happened.

Although different temperature conditions were tested, conjugation assay results were negative, thus indicating that the resistance determinants were not located on a conjugative plasmid.

3. Discussion

Our study describes a large intrahospital outbreak caused by KPC-Ec, involving a total of 106 patients in 25 wards. To the best of our knowledge, this is the largest outbreak caused by KPC-Ec reported worldwide. Genomic analyses allowed us to ascertain that five different KPC-Ec clusters were involved, three of which belonged to the high-risk ST131 clone. Notably, only the cluster 1 was responsible for the epidemic event and was associated to the H30Rx clade. Among patients in different wards, ST131 KPC-Ec showed a high propensity to cause infections. In fact, most samples were collected

specimens. Of note, only two isolates (1.6%) were from blood cultures. Overall, no deaths were attributable to infection caused by the ST131 KPC-Ec cluster 1 outbreak clone. Conversely, the treatment of infections caused by ST131 KPC-Ec was a challenge. In fact, isolates were mostly nonsusceptible to beta-lactams, including carbapenems and ceftolozane/tazobactam, due to the presence of the *bla*KPC and *bla*CTX-M determinants. KPC-Ec were also mostly resistant to ciprofloxacin and trimethoprim-sulphamethoxazole, another typical feature related to the pandemic ST131 lineage. Therapeutic options were aminoglycosides, tigecycline, colistin and the ceftazidime/avibactam combination.

The dissemination of the bla_{KPC} determinant in *E. coli*, largely due to horizontal transfer of plasmids or other mobile elements into diverse genetic backgrounds, has been previously described [19]. Since the acquisition of bla_{KPC} by *E. coli* is a very uncommon event, no data are to date available in the literature about a different prevalence between bla_{KPC-2} and bla_{KPC-3} variants. Nonetheless, in our experience, the bla_{KPC-2} gene seems to be the most represented in Italy, mainly associated with ST131 (unpublished data from a multicentric clinical study). The bla_{KPC} genetic background was conserved among both the KPC-Ec ST131 cluster 1 and the KPC-Kp strains of the same period, suggesting that transmission events of plasmid/mobile elements occurred. More importantly, the index patient was colonized during the same time period by both KPC-Kp (strain code sk138y138t) and KPC-Ec (strain code sk36y36t) strains (Table 2), the last one reasonably representing the origin of the outbreak. On the other hand, conjugation assays showed that the bla_{KPC} gene was not located on a conjugative plasmid, highlighting that the outbreak was caused by the spread of a dominant clone that had acquired the plasmid, rather than by the dissemination of a resistance plasmid to unrelated strains. Transmissions were drastically interrupted in May 2016, thanks to the adoption of a strict cohorting of both colonized and infected patients, that was assisted by dedicated healthcare staff. During the outbreak period, infection prevention and control measures were implemented. Training courses for the staff based on infection control, contact precautions and hand washing campaigns were promoted. Only sporadic cases related to KPC-Ec (ST131 cluster 1) were observed till June 2017, representing the tail of the outbreak. After this period, no other cases related to KPC-Ec were observed.

Taking together epidemiological and WGS data, we can speculate that the KPC-Ec outbreak clone developed in the previous context of the KPC-Kp outbreak. Furthermore, the outbreak appeared to be caused by the diffusion of a dominant clone (that probably acquired a bla_{KPC}-harboring plasmid), and not by the dissemination of a resistance plasmid among the strains of the different clusters.

Rapid application of WGS in outbreak investigations could be useful to better understand the dynamics of epidemic events in order to address infection control and contrast interventions.

4. Materials and Methods

4.1. Epidemiological Context and Characterization of Bacterial Isolates

We retrospectively studied an outbreak caused by KPC-Ec that occurred at the Hospital of Lecco (Northern Italy, close to Milan) across a six-month period (February to July 2016). The hospital accounts for about one thousand beds, and has a catchment area of about 340,000 inhabitants. During the outbreak period, as a part of the surveillance activity of the hospital team for infection control, 123 KPC-Ec nonrepetitive isolates were collected from 106 patients. A total of 103 isolates were from colonization surveillance rectal swabs. Rectal swabs have been used to screen intestinal colonization by KPC-Ec, as indicated by international guidelines (Centers for Diseases Control and Prevention, CDC, 2015; https://www.cdc.gov/hai/pdfs/cre/cre-guidance-508.pdf, accessed on 14 June 2021). Twenty strains from other sites were isolated from symptomatic patients with suspected infection. These isolates were from urine (n = 11), blood (n = 2), purulent exudate (n = 2), respiratory secretions (n = 2), drainage fluid (n = 1), peritoneal fluid (n = 1) and a surgical wound swab (n = 1). Strains from the same patients were included only when isolated from

different sites. A further four KPC-Ec isolates were sporadically collected after the epidemic event and until June 2017 and were also investigated to verify their clonal relationship to outbreak strains. To better clarify the origin of the bla_{KPC} gene in the KPC-Ec strains, six KPC-Kp isolates were collected from patients cocolonized by KPC-Kp and KPC-Ec and were included in the study. Of them, two were obtained at the beginning of the episode (including those isolated from the index patient), while the remaining were collected during the outbreak from patients admitted to those wards that were mainly involved in the episode. Isolates from rectal swabs were screened for carbapenemase production using chromogenic Brilliance CRE agar (Thermo Fisher Scientific). Bacterial isolates were identified to the species level using MALDI-TOF mass spectrometry (Vitek MS, bioMérieux), while susceptibility testing was routinely determined by the Vitek 2 system (bioMérieux). Isolates suspected of carbapenemase production (MIC values for ertapenem and/or meropenem >0.125 mg/L) were evaluated to assess the presence of specific carbapenem resistance determinants using the immunochromatographic technique (RESIST-4 O.K.N.V., Coris BioConcept) and/or a molecular dedicated assay (Xpert Carba-R, Cepheid).

To characterize epidemic isolates and better understand the dynamics of the outbreak, a total of 29 KPC-Ec strains (four of which were sporadically isolated in the *post*-outbreak period) were selected as representatives based on the site of infection or colonization, date and ward of admission (Table 2). MIC values of these isolates were determined by the MicroScan autoSCAN-4 system (NMDRM1 panel, Beckman Coulter). Selected antimicrobials (i.e., ceftazidime-avibactam, ceftolozane/tazobactam, and colistin) were evaluated by a broth microdilution Sensititre panel used for multidrug-resistant Gram-negative strains (DKMGN panel, Thermo Fisher Scientific). EUCAST criteria were used for determining susceptibility categories [22]. Finally, these isolates were analyzed by WGS-based typing and SNP-based phylogenetic reconstruction.

4.2. High Resolution Melting Assay

The selection of KPC-producing *K. pneumoniae* strains chosen for WGS investigation was made on the basis of the High Resolution Melting (HRM) assay results. The HRM was performed on the *wzi* hypervariable capsular gene as described by Perini et al. [23] using MeltingPlot software [24].

4.3. Whole-Genome Sequencing

A total of 35 strains (29 KPC-Ec and 6 KPC-Kp), representative of the epidemic event and of the post outbreak period, were processed for WGS analysis. In detail, inclusion criteria were: (i) isolates from ascertained infections (other than from screening rectal swabs) were chosen preferentially; (ii) KPC-Ec from all hospital wards involved in the epidemic event; (iii) KPC-Ec isolated in different periods of the outbreak (at the beginning, medium period, tail of the outbreak); (iv) In addition, 8 KPC-Ec from rectal swabs were included in order to evaluate the presence of the outbreak clone in the patients' intestinal microbiota. Genomic DNA was extracted using a QIAamp DNA minikit (Qiagen) following the manufacturer's instructions and sequenced using the Illumina Miseq platform with a 2 × 250 paired-end run after Nextera XT library preparation (Illumina Inc., San Diego, CA, USA).

4.4. CoreSNP Calling and Phylogenetic Analyses

For each of the 35 strains included in the study, the reads quality was assessed using FastqC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/, accessed on 29 June 2020), and the low quality terminal bases were trimmed using Trimmomatic software [25]. Reads were assembled using SPAdes software [26].

All the genome assemblies were submitted to the European Nucleotide Archive (ENA) with the project code PRJEB40388. All the ID codes are listed in Table S3.

The genome distance of Each KPC-Ec genome assembly was estimated using Mash software [27] against a collection of 3325 *E. coli* genomes retrieved from the PATRIC database [28] and the 50 most similar genomes were selected for subsequent analyses. All

the selected genome assemblies (from this study and the PATRIC database) were aligned against the *E. coli* MG1655 reference genome using progressive Mauve and coreSNPs were called as described by Gona and colleagues [29]. Repeated regions in the reference genome assembly were detected using Blastn. Then coreSNPs localized within repeated regions were masked. From here, the obtained coreSNP alignment was called "Global coreSNP".

Global coreSNP alignments were subjected to phylogenetic analysis using RAxML software with a 100 pseudo-bootstrap, after best model selection using ModelTest-NG [30]. The strains were then clustered on the basis of the Global phylogenetic tree and SNP distance. At first, we identified on the ML phylogenetic tree the largest highly supported (>75 bootstrap) monophyletic groups including study strains only.

4.5. Whole-Genome Sequencing-Based Typing

Resistance genes of the 29 KPC-Ec strains were identified using the ResFinder online tool [31] and SRST2 software [32] with the ARG-ANNOT dataset [33]. Virulence genes were detected by the VirulenceFinder online tool [34]. Plasmid incompatibility groups were detected using PlasmidFinder [35]. The Multi Locus Sequence Typing profiles of the 29 *E. coli* and six *K. pneumoniae* strains were determined in silico according to the Achtman and Pasteur schemes, respectively, using an in-house Perl script.

4.6. KPC-Harboring Contigs Comparison

For each identified cluster (see above) one representative strain was selected and the genome assembly was analyzed as follows. The contig harboring bla_{KPC} gene was identified by Blastn search (E-value threshold: 0.00001). The extracted contigs were oriented on the basis of bla_{KPC} gene orientation, then annotated using Prokka [36] and aligned with progressive Mauve [37]. The Tn4401 transposon was annotated using TETyper [38]. Lastly, the gene composition and synteny of extracted contigs were graphically represented using the R library genoPlotR [39].

4.7. Conjugation Assay

To assess the possible transferability of resistance determinants identified, a conjugation assay was performed using the *E. coli* J53 Azide^R as the recipient strain at temperatures of both 25 °C and 37 °C, and with MER 0.5mg/L for the selection of transconjugants.

5. Conclusions

Although KPC-producing *Escherichia coli* (KPC-Ec) remains uncommon, and mainly reported as the cause of sporadic episodes of infection rather than outbreaks, the present work shows that the acquisition of bla_{KPC} gene by a high-risk successful clone, as the ST131, can lead to even large and potentially difficult to manage epidemic events. The attention on the presence and circulation of carbapenemase-producing enterobacteria (CPE) should be always kept high, especially in healthcare settings. In this context, the application of WGS could be useful to better understand the evolution and dynamic of outbreaks sustained by CPE in order to promptly address infection control and contrast interventions.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/antibiotics10060718/s1, Figure S1. Local prevalence of KPC-Kp and KPC-Ec in the period 2014–2016 in the hospital setting. Table S1. *Escherichia coli* (KPC-Ec) isolates selected for WGS analysis and co-isolation of *Klebsiella pneumoniae* (KPC-Kp) isolates from the same patients. Table S2. *Klebsiella pneumoniae* (KPC-Kp) isolates selected for WGS analysis and co-isolation of *Escherichia coli* (KPC-Ec) isolates from the same patients. Table S3. List of the ENA codes for the genome assemblies of the studied strains.

Author Contributions: A.P. conceived the study, performed microbiological experiments and sequencing, analyzed and interpreted the data and wrote the paper; L.P. conceived the study, performed microbiological experiments, analyzed and interpreted the data and drafted the paper; F.C. and M.P. performed bioinformatic analyses; E.M. performed microbiological experiments and provided the epidemiological data; V.M.M. performed microbiological experiments and the conjugation assay; R.M. conceived the study, supervised the activities and revised the manuscript; F.L. conceived the study, supervised the activities and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in the context of normal clinical routine. All information and metadata about patients had been anonymized. Samples were coded and analyses were performed with anonymized database.

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome assemblies of the sequenced strains are available at the European Nucleotide Archive (ENA). The ID code for each genome submitted is reported in Supplementary Table S3.

Acknowledgments: Thanks to the Romeo and Enrica Invernizzi Foundation for its general support.

Conflicts of Interest: All the authors declare no conflict of interest.

References

1. Brolund, A.; Lagerqvist, N.; Byfors, S.; Struelens, M.J.; Monnet, D.L.; Albiger, B.; Kohlenberg, A.; European Antimicrobial Resistance Genes Surveillance Network EURGen-Net Capacity Survey Group. Worsening epidemiological situation of carbapenemase-producing *Enterobacteriaceae* in Europe, assessment by national experts from 37 countries, July 2018. *Euro Surveill* **2019**, *24*. [CrossRef]
2. Decraene, V.; Phan, H.T.T.; George, R.; Wyllie, D.H.; Akinremi, O.; Aiken, Z.; Cleary, P.; Dodgson, A.; Pankhurst, L.; Crook, D.W.; et al. A Large, Refractory Nosocomial Outbreak of *Klebsiella pneumoniae* Carbapenemase-Producing *Escherichia coli* demonstrates carbapenemase gene outbreaks involving sink sites require novel approaches to infection control. *Antimicrob. Agents Chemother.* **2018**, *62*, e01689-18. [CrossRef] [PubMed]
3. Stoesser, N.; Sheppard, A.E.; Peirano, G.; Anson, L.W.; Pankhurst, L.; Sebra, R.; Phan, H.T.T.; Kasarskis, A.; Mathers, A.J.; Peto, T.E.A.; et al. Genomic epidemiology of global *Klebsiella pneumoniae* carbapenemase (KPC)-producing *Escherichia coli*. *Sci. Rep.* **2017**, *7*, 5917. [CrossRef]
4. Stoesser, N.; Sheppard, A.E.; Pankhurst, L.; De Maio, N.; Moore, C.E.; Sebra, R.; Turner, P.; Anson, L.W.; Kasarskis, A.; Batty, E.M.; et al. Evolutionary History of the Global Emergence of the *Escherichia coli* Epidemic Clone ST131. *MBio* **2016**, *7*, e02162. [CrossRef]
5. Grundmann, H.; Glasner, C.; Albiger, B.; Aanensen, D.M.; Tomlinson, C.T.; Andrasević, A.T.; Cantón, R.; Carmeli, Y.; Friedrich, A.W.; Giske, C.G.; et al. Occurrence of carbapenemase-producing *Klebsiella pneumoniae* and *Escherichia coli* in the European survey of carbapenemase-producing *Enterobacteriaceae* (EuSCAPE): A prospective, multinational study. *Lancet Infect Dis.* **2017**, *17*, 153–163. [CrossRef]
6. Petrella, S.; Ziental-Gelus, N.; Mayer, C.; Renard, M.; Jarlier, V.; Sougakoff, W. Genetic and structural insights into the dissemination potential of the extremely broad-spectrum class A beta-lactamase KPC-2 identified in an *Escherichia coli* strain and an *Enterobacter cloacae* strain isolated from the same patient in France. *Antimicrob. Agents Chemother.* **2008**, *52*, 3725–3736. [CrossRef] [PubMed]
7. Urban, C.; Bradford, P.A.; Tuckman, M.; Segal-Maurer, S.; Wehbeh, W.; Grenner, L.; Colon-Urban, R.; Mariano, N.; Rahal, J.J. Carbapenem-resistant *Escherichia coli* harboring *Klebsiella pneumoniae* carbapenemase beta-lactamases associated with long-term care facilities. *Clin Infect Dis.* **2008**, *46*, e127–e130. [CrossRef] [PubMed]
8. Goren, M.G.; Navon-Venezia, S.; Chmelnitsky, I.; Carmeli, Y. Carbapenem-resistant KPC-2-producing *Escherichia coli* in a Tel Aviv Medical Center, 2005 to 2008. *Antimicrob. Agents Chemother.* **2010**, *54*, 2687–2691. [CrossRef]
9. Naas, T.; Cuzon, G.; Gaillot, O.; Courcol, R.; Nordmann, P. When carbapenem-hydrolyzing beta-lactamase KPC meets *Escherichia coli* ST131 in France. *Antimicrob. Agents Chemother.* **2011**, *55*, 4933–4934. [CrossRef]
10. Morris, D.; Boyle, F.; Ludden, C.; Condon, I.; Hale, J.; O'Connell, N.; Power, L.; Boo, T.W.; Dhanji, H.; Lavallee, C.; et al. Production of KPC-2 carbapenemase by an *Escherichia coli* clinical isolate belonging to the international ST131 clone. *Antimicrob. Agents Chemother.* **2011**, *55*, 4935–4936. [CrossRef]
11. Giani, T.; Antonelli, A.; Caltagirone, M.; Mauri, C.; Nicchi, J.; Arena, F.; Nucleo, E.; Bracco, S.; Pantosti, A.; Luzzaro, F.; et al. Evolving beta-lactamase epidemiology in *Enterobacteriaceae* from Italian nationwide surveillance, October 2013: KPC-carbapenemase spreading among outpatients. *Euro Surveill.* **2017**, *22*, 30583. [CrossRef] [PubMed]

12. Iacchini, S.; Sabbatucci, M.; Gagliotti, C.; Rossolini, G.M.; Moro, M.L.; Iannazzo, S.; D'Ancona, F.; Pezzotti, P.; Pantosti, A. Bloodstream infections due to carbapenemase-producing *Enterobacteriaceae* in Italy: Results from nationwide surveillance, 2014 to 2017. *Euro Surveill.* **2019**, *24*. [CrossRef] [PubMed]
13. Piazza, A.; Caltagirone, M.; Bitar, I.; Nucleo, E.; Spalla, M.; Fogato, E.; D'Angelo, R.; Pagani, L.; Migliavacca, R. Emergence of *Escherichia coli* Sequence Type 131 (ST131) and ST3948 with KPC-2, KPC-3 and KPC-8 carbapenemases from a Long-Term Care and Rehabilitation Facility (LTCRF) in Northern Italy. *Adv. Exp. Med. Biol.* **2016**, *901*, 77–89. [CrossRef]
14. Adler, A.; Miller-Roll, T.; Assous, M.V.; Geffen, Y.; Paikin, S.; Schwartz, D.; Weiner-Well, Y.; Hussein, K.; Cohen, R.; Carmeli, Y. A multicenter study of the clonal structure and resistance mechanism of KPC-producing *Escherichia coli* isolates in Israel. *Clin. Microbiol. Infect.* **2015**, *21*, 230–235. [CrossRef] [PubMed]
15. Mavroidi, A.; Miriagou, V.; Malli, E.; Stefos, A.; Dalekos, G.N.; Tzouvelekis, L.S.; Petinaki, E. Emergence of *Escherichia coli* sequence type 410 (ST410) with KPC-2 β-lactamase. *Int. J. Antimicrob. Agents.* **2012**, *39*, 247–250. [CrossRef]
16. Landman, D.; Urban, C.; Bäcker, M.; Kelly, P.; Shah, N.; Babu, E.; Bratu, S.; Quale, J. Susceptibility profiles, molecular epidemiology, and detection of KPC-producing *Escherichia coli* isolates from the New York City vicinity. *J. Clin. Microbiol.* **2010**, *48*, 4604–4607. [CrossRef]
17. Leung, V.; Loo, V.G.; Frenette, C.; Domingo, M.C.; Bourgault, A.M.; Mulvey, M.R.; Robson, H.G. First Canadian outbreak of *Enterobacteriaceae*-expressing *Klebsiella pneumoniae* carbapenemase type 3. *Can. J. Infect. Dis. Med. Microbiol.* **2012**, *23*, 117–120. [CrossRef]
18. Accogli, M.; Giani, T.; Monaco, M.; Giufrè, M.; García-Fernández, A.; Conte, V.; D'Ancona, F.; Pantosti, A.; Rossolini, G.M.; Cerquetti, M. Emergence of *Escherichia coli* ST131 sub-clone H30 producing VIM-1 and KPC-3 carbapenemases, Italy. *J. Antimicrob. Chemother.* **2014**, *69*, 2293–2296. [CrossRef]
19. Chavda, K.D.; Chen, L.; Jacobs, M.R.; Bonomo, R.A.; Kreiswirth, B.N. Molecular Diversity and Plasmid Analysis of KPC-Producing *Escherichia coli*. *Antimicrob. Agents Chemother.* **2016**, *60*, 4073–4081. [CrossRef]
20. Petty, N.K.; Ben Zakour, N.L.; Stanton-Cook, M.; Skippington, E.; Totsika, M.; Forde, B.M.; Phan, M.D.; Gomes Moriel, D.; Peters, K.M.; Davies, M.; et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5694–5699. [CrossRef]
21. Ben Zakour, N.L.; Alsheikh-Hussain, A.S.; Ashcroft, M.M.; Khanh Nhu, N.T.; Roberts, L.W.; Stanton-Cook, M.; Schembri, M.A.; Beatson, S.A. Erratum for Ben Zakour et al., Sequential Acquisition of Virulence and Fluoroquinolone Resistance Has Shaped the Evolution of *Escherichia coli* ST131. *MBio* **2016**, *7*, e00958-16. [CrossRef] [PubMed]
22. European Committee on Antimicrobial Susceptibility Testing. *Breakpoint Tables for Interpretation of MICs and Zone Diameters, Version 6.0*; EUCAST: Basel, Switzerland, 2016; Available online: http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/Breakpoint_table_v_6.0.pdf (accessed on 14 June 2021).
23. Perini, M.; Piazza, A.; Panelli, S.; Di Carlo, D.; Corbella, M.; Gona, F.; Vailati, F.; Marone, P.; Cirillo, D.M.; Farina, C.; et al. EasyPrimer: User-friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on *wzi* gene for fast *Klebsiella pneumoniae* typing. *Sci. Rep.* **2020**, *10*, 1307. [CrossRef] [PubMed]
24. Perini, M.; Batisti Biffignandi, G.; Di Carlo, D.; Pasala, A.R.; Piazza, A.; Panelli, S.; Zuccotti, G.V.; Comandatore, F. MeltingPlot, a user-friendly online tool for epidemiological investigation using High Resolution Melting data. *BMC Bioinform.* **2021**, *22*, 76. [CrossRef] [PubMed]
25. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
26. Nurk, S.; Bankevich, A.; Antipov, D.; Gurevich, A.A.; Korobeynikov, A.; Lapidus, A.; Prjibelski, A.D.; Pyshkin, A.; Sirotkin, A.; Sirotkin, Y.; et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **2013**, *20*, 714–737. [CrossRef]
27. Ondov, B.D.; Treangen, T.J.; Melsted, P.; Mallonee, A.B.; Bergman, N.H.; Koren, S.; Phillippy, A.M. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **2016**, *17*, 132. [CrossRef]
28. Wattam, A.R.; Davis, J.J.; Assaf, R.; Boisvert, S.; Brettin, T.; Bun, C.; Conrad, N.; Dietrich, E.M.; Disz, T.; Gabbard, J.L.; et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* **2017**, *45*, D535–D542. [CrossRef]
29. Gona, F.; Comandatore, F.; Battaglia, S.; Piazza, A.; Trovato, A.; Lorenzin, G.; Cichero, P.; Biancardi, A.; Nizzero, P.; Moro, M.; et al. Comparison of core-genome MLST, coreSNP and PFGE methods for *Klebsiella pneumoniae* cluster analysis. *Microb. Genom.* **2020**, *6*, e000347. [CrossRef]
30. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]
31. Zankari, E.; Hasman, H.; Cosentino, S.; Vestergaard, M.; Rasmussen, S.; Lund, O.; Aarestrup, F.M.; Larsen, M.V. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **2012**, *67*, 2640–2644. [CrossRef]
32. Inouye, M.; Dashnow, H.; Raven, L.A.; Schultz, M.B.; Pope, B.J.; Tomita, T.; Zobel, J.; Holt, K.E. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **2014**, *6*, 90. [CrossRef]
33. Gupta, S.K.; Padmanabhan, B.R.; Diene, S.M.; Lopez-Rojas, R.; Kempf, M.; Landraud, L.; Rolain, J.M. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **2014**, *58*, 212–220. [CrossRef] [PubMed]

34. Joensen, K.G.; Scheutz, F.; Lund, O.; Hasman, H.; Kaas, R.S.; Nielsen, E.M.; Aarestrup, F.M. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Micobiol.* **2014**, *52*, 1501–1510. [CrossRef] [PubMed]
35. Carattoli, A.; Zankari, E.; García-Fernández, A.; Voldby Larsen, M.; Lund, O.; Villa, L.; Møller Aarestrup, F.; Hasman, H. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **2014**, *58*, 3895–3903. [CrossRef] [PubMed]
36. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]
37. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **2010**, *5*, e11147. [CrossRef] [PubMed]
38. Sheppard, A.E.; Stoesser, N.; German-Mesner, I.; Vegesana, K.; Sarah Walker, A.; Crook, D.W.; Mathers, A.J. TETyper: A bioinformatic pipeline for classifying variation and genetic contexts of transposable elements from short-read whole-genome sequencing data. *Microb. Genom.* **2018**, *4*, e000232. [CrossRef]
39. Guy, L.; Kultima, J.R.; Andersson Siv, G.E. genoPlotR: Comparative gene and genome visualization in R. *Bioinformatics* **2010**, *26*, 2334–2335. [CrossRef]

Section C

Appendix 6

EasyPrimer: user- friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on wzi gene for fast *Klebsiella pneumoniae* typing.

Perini M, Piazza A, Panelli S, Di Carlo D, Corbella M, Gona F, Vailati F, Marone P, Cirillo DM, Farina C, Zuccotti G, Comandatore F

Sci Rep. 2020 Jan 28;10(1):1307. doi: 10.1038/s41598-020-57742-z

SCIENTIFIC REPORTS
nature research

OPEN

EasyPrimer: user-friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on *wzi* gene for fast *Klebsiella pneumoniae* typing

Matteo Perini[1], Aurora Piazza[1], Simona Panelli[1], Domenico Di Carlo[1], Marta Corbella[2], Floriana Gona[3], Francesca Vailati[4], Piero Marone[2], Daniela Maria Cirillo[3], Claudio Farina[4], Gianvincenzo Zuccotti[1,5] & Francesco Comandatore[1*]

In this work we present EasyPrimer, a user-friendly online tool developed to assist pan-PCR and High Resolution Melting (HRM) primer design. The tool finds the most suitable regions for primer design in a gene alignment and returns a clear graphical representation of their positions on the consensus sequence. EasyPrimer is particularly useful in difficult contexts, e.g. on gene alignments of hundreds of sequences and/or on highly variable genes. HRM analysis is an emerging method for fast and cost saving bacterial typing and an HRM scheme of six primer pairs on five Multi-Locus Sequence Type (MLST) genes is already available for *Klebsiella pneumoniae*. We validated the tool designing a scheme of two HRM primer pairs on the hypervariable gene *wzi* of *Klebsiella pneumoniae* and compared the two schemes. The *wzi* scheme resulted to have a discriminatory power comparable to the HRM MLST scheme, using only one third of primer pairs. Then we successfully used the *wzi* HRM primer scheme to reconstruct a *Klebsiella pneumoniae* nosocomial outbreak in few hours. The use of hypervariable genes reduces the number of HRM primer pairs required for bacterial typing allowing to perform cost saving, large-scale surveillance programs.

Most methods used for the identification and typing of prokaryotes are based on DNA amplification and sequencing. Indeed, the sequence of specific genes can harbour enough information to classify bacteria at species, subspecies or also to a clonal level. For instance, Multi-Locus Sequence Typing (MLST) is one of the most used methods for bacterial typing and it is based on the amplification and sequencing of few housekeeping genes[1]. During the last ten years, the analysis of the entire bacterial genome by Whole Genome Sequencing (WGS) approach revolutionized the field, drastically increasing the typing precision[1].

The reconstruction of nosocomial outbreaks is one of the most important clinical applications of bacterial typing. A nosocomial outbreak occurs when the number of patients infected by a pathogen increases above the expected in a limited time[2]. In these situations, it is fundamental to determine the clonality of bacteria causing disease in the patients to define the proper strategy to handle the emergency. Pulsed-Field Gel Electrophoresis (PFGE), MLST and WGS are the most frequently applied molecular methods in outbreak investigation[1].

During a nosocomial outbreak, clinicians need bacterial typing information in the shortest time possible. Despite the high potential of WGS in outbreak reconstruction, the sequencing of a complete genome requires two to four working days, introducing an important time lag. Similarly, PFGE typing requires five days and also MLST needs few days. During the last years, the High Resolution Melting (HRM) assay has emerged as a low-cost and fast method for bacterial typing, particularly promising for epidemiological applications[3–6]. HRM is a single-step

[1]Department of Biomedical and Clinical Sciences "L. Sacco", Università di Milano, Pediatric Clinical Research Center "Romeo and Enrica Invernizzi", Milan, 20157, Italy. [2]S.C. Microbiologia e Virologia, Fondazione IRCCS Policlinico San Matteo, Pavia, 27100, Italy. [3]Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan, 20132, Italy. [4]Microbiology Institute, A.S.S.T. "Papa Giovanni XXIII", Bergamo, 24127, Italy. [5]Department of Pediatrics, V. Buzzi Childrens' Hospital, Università di Milano, Milan, Italy. *email: francesco.comandatore@unimi.it

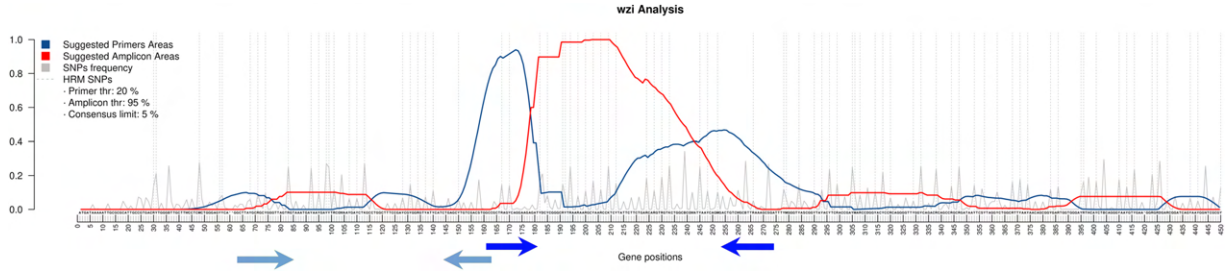


Figure 1. EasyPrimer output for the *wzi gene* (563 alleles on https://bigsdb.pasteur.fr). The consensus sequence calculated from the gene alignment is reported on the x-axis. Residues under the peaks of the blue curve are highly conserved and thus suitable for primer design. Conversely the red curve increases over the highly variable regions suggested to be amplified. The grey peaks represent all the Single Nucleotide Polymorphisms (SNPs) with their own frequency. The dotted lines are used to highlight the "HRM-detectable" SNPs, i.e. the ones causing a change in the GC content. The blue arrows were manually added to show the positions of the two primer pairs designed on *wzi* in this work.

procedure for the discrimination of sequence variants on the basis of their melting temperature. This method allows to perform bacterial typing in less than five hours[7].

To develop a novel HRM-based typing procedure, it is necessary to: i) select one or more core genes; ii) design a primer pair in conserved regions flanking a gene portion where the melting temperature varies among the strains.

Andersson and colleagues have developed the "MinimumSNP" tool[8], which identifies, in a gene alignment, the variable positions that can lead to a melting temperature change (called informative SNPs). MinimumSNP identifies single informative positions, not necessarily grouped in a single portion of the gene. In other words, it does not indicate which regions are more suitable for primer design: two low-variable regions neighbouring a SNP-rich informative stretch. Thus, the user of MinimumSNP has to choose one (or few) SNPs and then design primers around it (or around them).

Herein, we present EasyPrimer, a web-based tool for the identification of the gene regions suitable for primer design to perform HRM studies and any kind of pan-PCR experiments. Moreover, we validated EasyPrimer by designing HRM primers for the discrimination of clinical isolates of *Klebsiella pneumoniae*, an important opportunistic pathogen frequently cause of infections in humans and animals[9].

This article was submitted to an online preprint archive[10].

Results

EasyPrimer: a tool for primers design. EasyPrimer is a user-friendly open-source tool developed to assists primer design in difficult contexts, e.g. on an alignment of hundreds of sequences and/or on hypervariable genes. The tool uses as input a nucleotide multi-fasta file and identifies the best regions for primer design: two low variable regions flanking a highly variable one. The on-line and the stand-alone versions of the tool are freely available at https://skynet.unimi.it/index.php/tools/easyprimer.

Primers design. We downloaded *pgi*, *gapA* and *wzi* gene sequences from BigsDB database (https://bigsdb.pasteur.fr) and we run EasyPrimer to identify the best regions for primer design. The EasyPrimer output for the *wzi* gene is reported in Fig. 1, while the outputs relative to *pgi* and *gapA* genes are reported in Supplementary Figs. S1 and S2, respectively. Then we designed a total of four novel primer pairs: one for *pgi*, one for *gapA* and two for *wzi* (reported in Table 1).

High-resolution melting analysis. In this work we considered three clinical strains collections:

- the "background" collection, which includes 17 *K. pneumoniae* strains belonging to 17 different Sequence Types (STs);
- the "outbreak" collection, which includes 11 *K. pneumoniae* strains isolated during a nosocomial outbreak;
- the "validation" collection that includes 54 *K. pneumoniae* strains belonging to six of the most epidemiologically relevant STs (i.e. ST258, ST512, ST11, ST101, ST15 and ST307)[11,12].

The strains of the background and outbreak collections were analysed using all the ten primer pairs listed in Table 1. The strains of the validation collection were subjected to HRM experiments using only the two primer pairs designed on *wzi* gene (*wzi-3* and *wzi-4* primer pairs). Four out of the ten primer pairs were newly designed in this work (see above), while the remaining six were already available in literature[7]. The obtained melting temperatures ("Tm") of the three HRM replicates and their relative average temperature ("aTm") values are reported in Supplementary Table S1.

Primer pairs and schemes comparison. For each of the ten primer pairs we calculated the strain distance matrix among the background collection strains based on the aTm values (see Methods). The calculated aTm distances ranged from zero to three degrees, and the median distances varied among the genes (as shown in Fig. 2). In particular, the two *wzi* primer pairs showed median distance values significantly higher than those obtained for many of the other primer pairs (see Supplementary Table S2 for details).

Gene name	Primer name	Primer pair name - in this work	Primer scheme name	Sequence (5′ - 3′)	Amplicon length (bp)	Reference
infB	*infB*729-F	*infB-1*	MLST6, MLST8	CCTGCCGGAAGAGTGG	50	Andersson *et al.* 2012
	*infB*729-R		MLST6, MLST8	TCGCGGAAACGTGGAC		Andersson *et al.* 2012
mdh	*mdh*1197-F	*mdh-1*	MLST6, MLST8	ATTGCCGACCTGACTAAACG	58	Andersson *et al.* 2012
	*mdh*1197 R		MLST6, MLST8	CTTTCGCTTCCACGACTTC		Andersson *et al.* 2012
phoE	*phoE*2013-F	*phoE-1*	MLST6, MLST8	GAAGGGGTGGGGAGTGA	78	Andersson *et al.* 2012
	*phoE*2013-R		MLST6, MLST8	GGCGTTCATGTTTTTGTTGA		Andersson *et al.* 2012
rpoB	*rpoB*2227-F	*rpoB-1*	MLST6, MLST8	TGATTAACTCCCTGTCCGTGT	132	Andersson *et al.* 2012
	*rpoB*2227-R		MLST6, MLST8	CGTAGTTGCCTTCTTCGATAGC		Andersson *et al.* 2012
tonB	*tonB*2693-F	*tonB-1*	MLST6, MLST8	GTTGAACCCGAACCTGAGC	101	Andersson *et al.* 2012
	*tonB*2693-R		MLST6, MLST8	GGTTTGGGCTTCGGCTTA		Andersson *et al.* 2012
tonB	*tonB*2886-F	*tonB-2*	MLST6, MLST8	AAAAGGTTGAACAGCCGAAG	120	Andersson *et al.* 2012
	*tonB*2886-R		MLST6, MLST8	CCGCTGCTGTCGAGGT		Andersson *et al.* 2012
gapA	*gapA*_F	*gapA-1*	MLST8	AAAGTCGTTCTGACTGGC	95	This work
	*gapA*_R		MLST8	TTRAAACGATGTCCTGGC		This work
pgi	*pgi*_F	*pgi-1*	MLST8	CCAAAATGGTACCGTGCGATT	156	This work
	*pgi*_R		MLST8	CCTGATCGCGRTATTCCTGCT		This work
wzi	*wzi*3_F	*wzi-3*	wzi	GCTTAYGCRGCYGGGTTAGTRGT	114	This work
	*wzi*3_R		wzi	GGCCASGTCGACARGCTCAG		This work
wzi	*wzi*4_F	*wzi-4*	wzi	GCCGCTRAGYCAGGAAGAGAT	101	This work
	*wzi*4_R		wzi	GACTGTCWGCBTTRAAAGCSGA		This work

Table 1. Primer pairs used in this work.

We also compared the aTm distance matrices of the following schemes:

- "MLST6": which includes the HRM-MLST primer pairs already present in literature;
- "MLST8": which includes the MLST6 primer pairs and the two newly designed;
- "wzi": which includes the two primer pairs designed for the *wzi* gene.

The median pairwise distance did not significantly change among the three schemes (Wilcoxon test with Holm post-hoc correction, p-value > 0.05) and the relative boxplot graphs are reported in Supplementary Fig. S3.

Furthermore, we compared the aTm distance matrices of wzi and MLST8 schemes for each strain pair of the background collection, subtracting the two matrices (see Fig. 3). We found that, among all the 136 possible strain pairs, 66 (48.5%) showed a higher distance for wzi scheme than MLST8 scheme. More in detail, the ST258 *wzi_29* (also known as ST258 Clade 1[13]) was better discriminated from the ST512 *wzi_154* (which is part of the ST258 Clade 2[14]) by wzi scheme than MLST8 (see Fig. 3). ST307 was better or equally discriminated from all the other strains by wzi scheme than MLST8 (except for the ST512 and ST147 strains). Similarly, ST15 (*wzi_89*) strain was better discriminated from all the background strains by wzi scheme than MLST8, apart from the ST147 strain (see Fig. 3). Furthermore, the ST101 strain was better discriminated by wzi than MLST8 scheme from ST258 (*wzi_29*), ST307, ST512 and ST15 (*wzi_89*). ST11 (*wzi_75*) was better discriminated by wzi scheme from ST15 (*wzi_89*), ST307, ST258 and ST512. Conversely, ST11 (*wzi_75*) and ST101 were better discriminated by MLST8 than wzi. Lastly, ST147 was the only ST better discriminated by MLST8 scheme than wzi for all the strains pairs (see Fig. 3).

Whole genome sequencing-based strain typing. A total of 82 *K. pneumoniae* strains have been subjected to WGS-based typing in this work.

24 out of 82 strains have been previously subjected to NGS sequencing as part of two published works:

- 12 from Gaiarsa and colleagues[14] (10/12 for the background and 2/12 for the validation collection).
- 12 from Gona and colleagues[15] (8/12 included in the background collection and 4/12 included in the validation collection).

For these strains, the NGS sequences were retrieved from public database.

For the remaining 58 out of 82 *K. pneumoniae* strains, the reads and genomic sequences were obtained in this work (see Supplementary Table S3):

- 11/58 strains of the outbreak collection (all from Papa Giovanni XXIII hospital).
- 47/58 strains of the validation collection (45/47 from San Raffaele hospital and 2/47 from Papa Giovanni XXIII hospital).

The *wzi* allele, the ST, the K-type as well as the accession numbers of these 82 *K. pneumoniae* strains are reported in Supplementary Table S3.

WGS-based outbreak reconstruction. Ten out of the 11 outbreak isolates belonged to the ST512 while the isolate "BG-Kpn-22–18" belonged to the ST307 (see Supplementary Table S3). An alignment of 66 core-SNPs

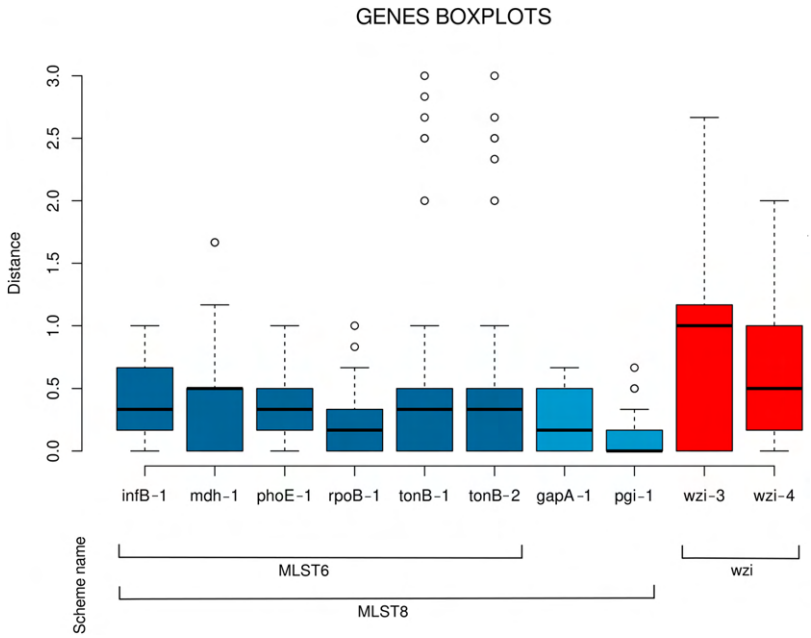


Figure 2. Distribution of the average melting temperature differences among the 17 *Klebsiella pneumoniae* strains for the ten primer pairs. Boxes are the 25th and 75th quartiles divided by the medians, whiskers are 1.5x the interquartile ranges and dots are outliers. The lines in the bottom show the composition of the three primer schemes used in this work.

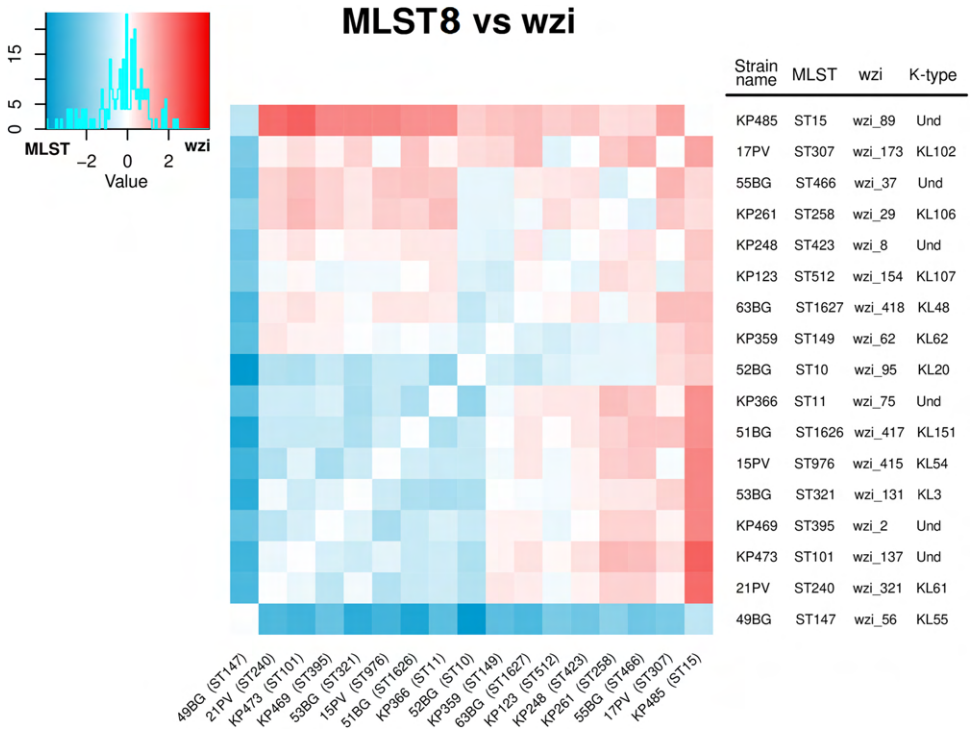


Figure 3. Arithmetic difference between the average melting temperature distance matrices computed among the 17 *Klebsiella pneumoniae* strains (selected to belong to 17 different STs) using the MLST8 scheme (eight primer pairs on seven genes) and wzi scheme (two primer pairon one gene). The heatmap colours range from blue to white to red: if the temperature distance between two strains is greater for the MLST8 than the wzi scheme the relative position on the heatmap is coloured in blue, otherwise in red.

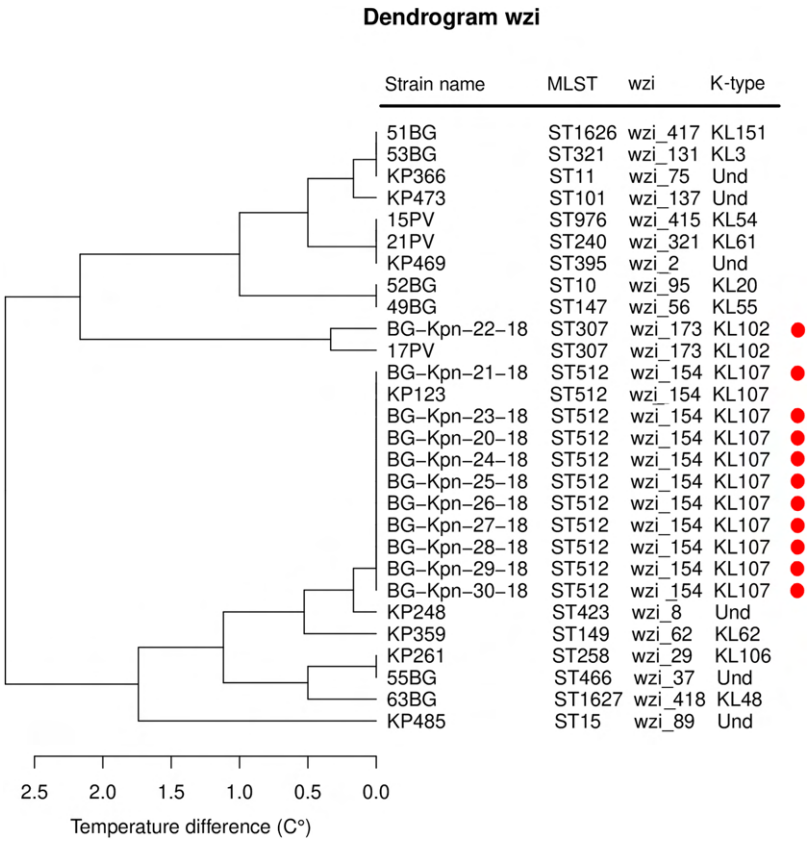


Figure 4. Dendrogram of the hierarchical clustering analysis on the average temperature distance matrix obtained using the wzi primer scheme. The 17 "background" strains belonging to 17 different MLSTs are written in black, while the 11 strains isolated during the nosocomial outbreak are highlighted by a red dot.

was obtained from the 11 outbreak strains. The relative Maximum Likelihood phylogenetic tree is reported in Supplementary Fig. S4. The ST512 strains have SNP distances ranging from zero to four SNPs. Conversely, the SNP distances among the ST512 strains and the ST307 strain ranged from 63 to 66 SNPs.

HRM-based outbreak reconstruction. The three dendrograms obtained by hierarchical clustering on the aTms strain distances for the schemes MLST6, MLST8 and wzi are reported in Supplementary Figs. S5 and S6 and Fig. 4, respectively. All the schemes correctly discriminated the outbreak ST512 strains from the ST307 one. Notably, only the wzi scheme correctly clustered the outbreak strains with the background strain of the same ST.

Repeatability of the *wzi* HRM protocol. To validate the repeatability of the *wzi* HRM typing protocol, we included in the analysis 54 additional strains belonging to the most epidemiologically relevant clades (ST258 Clade 1, ST258 Clade 2, ST307, ST101, ST11 and ST15)[11,12], for a total of 82 strains (see Supplementary Table S3).

For each of these clades, the aTms obtained for *wzi-3* and *wzi-4* primer pairs varied in a range $<=0.5\,°C$, corresponding to the sensitivity of the machine used for the HRM experiments (see Table 2).

Clustering analysis on *wzi-3* and *wzi-4* melting temperatures grouped the strains in seven clusters (see Figs. 5 and 6). Among the most epidemiologically relevant clades, all the strains of the ST258 Clade 1 (ST258 *wzi_29*), ST258 Clade 2 (ST258 *wzi_154* and ST512 *wzi_154*) and ST307 were correctly clustered; while ST11 and ST101 strains fell in the same cluster. The three ST15 strains fell in two different clusters coherently to their *wzi* alleles: the one harbouring *wzi_89* clustered alone, while the other two strains, both harbouring *wzi_24*, fell in the ST11/ST101 cluster (see Figs. 5 and 6).

Discussion

High Resolution Melting (HRM) is a real-time PCR analysis for the detection of mutations and polymorphisms[3,4], also applicable for fast bacterial typing in hospital surveillance and real-time nosocomial outbreak detection[5]. Several works applied HRM to bacterial typing, exploiting Multi Locus Sequence Type (MLST) genes[7,16,17] which have been considered the gold standard genes for bacterial typing for almost 20 years[18]. These genes have been selected to be housekeeping therefore they display low variability. In this work we show that it is possible to increase HRM discriminatory power using hypervariable genes.

On the other hand, the identification of the regions suitable for primer design can be challenging when the number of aligned sequences is high or when the gene is hypervariable. Thus, we developed EasyPrimer, a tool for the identification of the best regions for primer design for HRM analysis and, more in general, for any kind of

Clade	MLST profile	*wzi* allele	Min-Max *wzi-3* T° (°C)	Min-Max *wzi-4* T° (°C)	Range *wzi-3* T° (°C)	Range *wzi-4* T° (°C)	# of strains
ST258 Clade 2	ST512	*wzi_154*	83.50–84.00	83.33–83.50	0.50	0.17	23
	ST258	*wzi_154*	83.83–83.83	83.50–83.50	0	0	3
ST258 Clade 1	ST258	*wzi_29*	83.50–84.00	84.33–84.67	0.50	0.34	8
ST307	ST307	*wzi_173*	84.00–84.50	82.33–82.83	0.50	0.50	19
ST101	ST101	*wzi_137*	84.83–85.00	83.33–83.83	0.17	0.50	12
ST11	ST11	*wzi_24*	85.00–85.00	83.50–83.50	0	0	2
ST15	ST15	*wzi_24*	85.00–85.17	83.83–83.83	0.17	0	2
	ST15	*wzi_89*	82.33–82.33	84.00–84.00	0	0	1

Table 2. Ranges of *wzi-3* and *wzi-4* melting temperatures of the most epidemiologically relevant clades and Sequence Types.

pan-PCR study. EasyPrimer shows, with an easy-to-read graphical output, which are the best regions for primer design: two conserved regions flanking a highly variable one. The on-line and the stand-alone versions of the tool are freely available at https://skynet.unimi.it/index.php/tools/.

We validated the tool designing HRM primers for the nosocomial pathogen *Klebsiella pneumoniae*. A scheme including six HRM primer pairs for five out of the seven *K. pneumoniae* MLST genes was already available in literature[7] (MLST6 scheme). Thus, we used EasyPrimer to design the primers for the remaining two MLST genes (*pgi* and *gapA*), obtaining a larger scheme with eight primer pairs (MLST8 scheme). Furthermore, we designed two HRM primer pairs for the hypervariable capsular gene *wzi* (wzi scheme). We tested the discriminatory power of these schemes on 17 *K. pneumoniae* strains belonging to 17 different STs (background collection) and we used the HRM approach to study an outbreak occurred in an Italian hospital.

Notably, most of the epidemiologically relevant *K. pneumoniae* clades (and/or STs) emerged after large recombination events that involved the capsule locus (which includes *wzi* gene), often leading to K-type change. For this reason, the emergence of a novel clade/ST is often associated to *wzi* allele change[19]. This makes *wzi* gene particularly suitable for *K. pneumoniae* typing. Our analyses on the background collection showed a good discriminatory power for both the MLST-based and *wzi*-based HRM assays: both schemes successfully discriminated most of the analysed strains. Wzi scheme discriminated better than MLST8 scheme all the highly epidemiologically relevant clades (ST258 Clade 1, ST258 Clade 2, ST307, ST11, ST101 and ST15), except for the pairs ST258 Clade 2 - ST307 and ST11 - ST101 (see Fig. 3). Nonetheless, wzi scheme is able to discriminate the ST258 Clade 2 from ST307 as well (see Fig. 4). Conversely, *wzi* scheme does not discriminate ST11 from ST101 (see Figs. 5 and 6). The latter may represent a minor flaw of the *wzi* HRM protocol as the two STs are mostly isolated in geographically distant areas of the globe (namely: ST11 in Asia[20], ST101 in Europe-Africa[21]).

Clustering analysis on the 82 strains, including multiple strains from the same clade (see Materials), allowed to evaluate with more precision the discriminatory power of the *wzi* HRM protocol. The analysis clearly showed that the protocol is able to discriminate five of the six most epidemiologically relevant *K. pneumoniae* clades, discriminating ST258 Clade 1, ST258 Clade 2, ST307, ST11/ST101 and ST15 (see Figs. 5 and 6). In our dataset we found strains of the ST15 and ST11 harbouring different *wzi* alleles. This is not surprising, considering that the capsule locus (which contain the *wzi* gene) is a recombinational hotspot in ST11[19]. Wzi scheme discriminated among the different ST15 strains present in the collections, according to the different *wzi* alleles they harbour (see Figs. 5 and 6). This highlights the benefits of using hypervariable genes instead of MLST genes in typing methods: e.g. *wzi* HRM protocol can rule out an ST15 outbreak of strains harbouring different *wzi* alleles, any MLST-based protocol cannot.

As stated above, *K. pneumoniae* clades are often associated to specific *wzi* alleles and K-types. Despite the *wzi* HRM protocol was designed on *wzi* gene, it correctly discriminates most of the epidemiological relevant clades. Furthermore, we found that every K-types correspond to a specific *wzi* allele (see Supplementary Table S3).

Moreover, the analysis of the 82 strains clearly showed that the *wzi* HRM protocol is highly repeatable: *wzi-3* and *wzi-4* aTms ranged within 0.5 °C (the machine sensitivity) among the strains of the same clade (see Table 2). For this study, dozens of independent HRM experiments have been performed in different months by two different operators (M.P. and A.P.). The observed stability of HRM aTms for each clade clearly shows that the results of *wzi* HRM protocol are portable. This makes the method suitable for studies involving several isolates, such as large hospital surveillance programs.

Additionally, we want to highlight that the observed HRM discriminatory power was obtained using a BioRad CFX Connect real-time PCR instrument (BioRad, Hercules, California): a machine not specifically designed for HRM experiments but for real-time PCR, with a melting temperature sensitivity of 0.5 °C (i.e. a lower sensitivity compared to HRM machines).

We applied the wzi scheme to the reconstruction of a nosocomial outbreak occurred in an Italian hospital. During the outbreak, 11 patients resulted colonized or infected by *K. pneumoniae* and the WGS typing revealed that the isolates belonged to two different clones. These clones were identified on the basis of core SNP distance (SNP distance < 5) and MLST profile (one isolate belongs to the ST307 and ten isolates to the ST512). As shown in Fig. 4, the wzi scheme not only correctly discriminated the outbreak isolates of the two clones but it also clustered them with the background isolates of the corresponding ST profile.

During the last years, WGS has revolutionized clinical microbiology, allowing the precise description of bacterial genomic features in few days (including the presence of resistance and/or virulence factors). Despite this, its application during real-time outbreak reconstructions still shows some limits: the time required to be

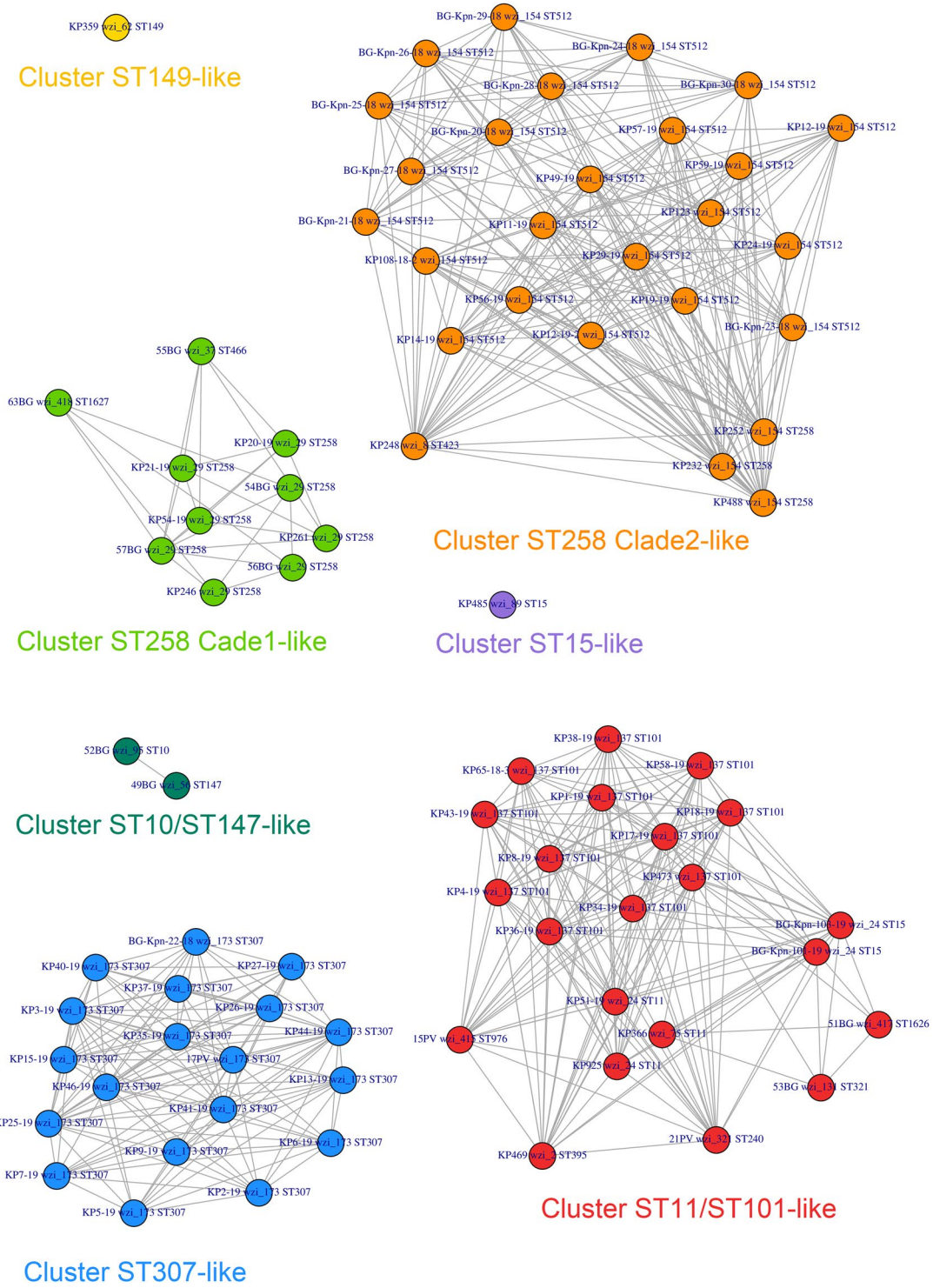


Figure 5. Strain-to-strain network of the 82 isolates analysed, generated on the basis of *wzi-3* and *wzi-4* HRM melting temperatures. Two strains were connected if both *wzi-3* and *wzi-4* gave difference in melting temperature < 0.5 °C. Clusters were identified as separated sub-networks on the strain-to-strain network and they were named from the major Sequence Type they include.

completed, the cost and the necessity of qualified personnel for library preparation, bioinformatic analyses and results interpretation. Indeed, the complete sequencing of a bacterial strain genome costs at least ~100 euros (using an Illumina MiSeq machine) and requires one or two days for library preparation and 5–36 hours for sequencing. During the first days of a nosocomial outbreak the number of cases still increases slowly. In this time frame, it is crucial to quickly understand if the bacterial strains are genetically related, and if the clone is spreading in the nosocomial environment. In this situation HRM is a "first-line" typing technology to figure out when an

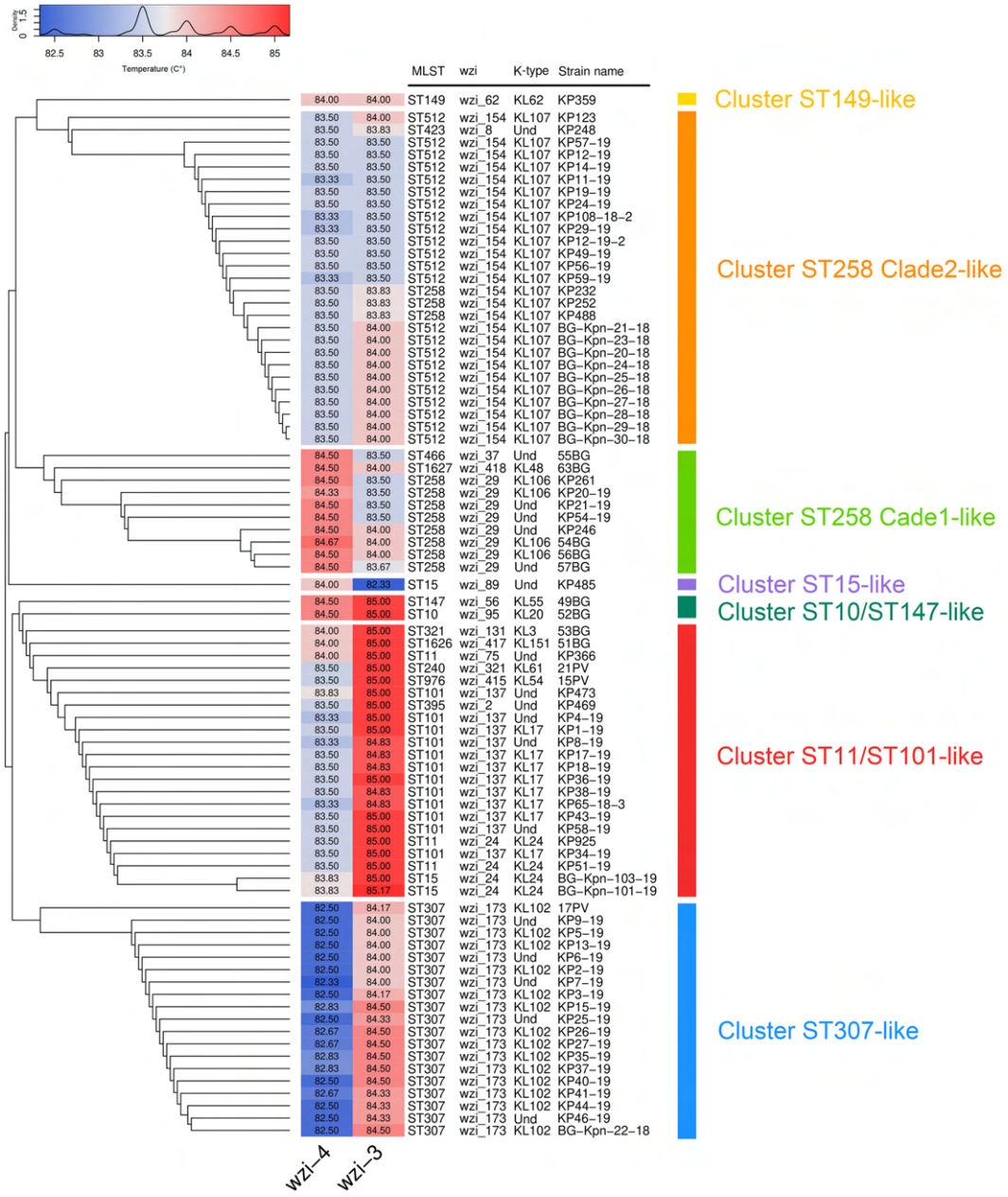


Figure 6. On the left the dendrogram obtained from the strain-to-strain network reported in Fig. 5. In the middle, the average melting temperatures of *wzi-3* and *wzi-4* primer pairs, the Sequence Type, the *wzi* allele and the K-type of the 82 isolates analysed. On the right, the names of the clusters identified by network analysis, corresponding to the clusters in Fig. 5.

outbreak is starting. Indeed, HRM is less precise than WGS but it can be reliable for a fast, preliminary bacterial typing, fundamental in the first days of a nosocomial outbreak. If the outbreak is identified, WGS could be used to further investigate the transmission dynamics. HRM assay represents a fast, simple and time/cost saving approach for bacterial typing, allowing to analyse several bacterial samples per days. Furthermore, this technique does not require advanced skills in molecular biology and the results can be analysed without the use of any specific software. This method can be useful also in veterinary and dairy farming settings: *K. pneumoniae* is a relevant veterinary pathogen and one of the most frequent cause of mastitis in dairy cattle[9].

We found that the discriminatory power of an HRM scheme does not strictly depend on the number of genes but also on their genetic variability. Indeed, comparing the MLST6 and MLST8 schemes, we found that the median distance among the strains did not change significantly. Wzi scheme contains two primer pairs and this reduces drastically the amount of time and costs required for typing. For instance, using only two primer pairs on a 96-well PCR plate, it is possible to type 15 isolates per run (five hours, including DNA extraction, HRM run and analysis of results) with a cost of ~5 euros each. This makes the HRM a feasible method for real-time surveillance and for a preliminary typing step in large epidemiological studies. Lately, Multi-Drug Resistance (MDR) *K. pneumoniae* strains have become a major burden for public health worldwide. Despite WGS represents an important

tool for precise bacterial typing, it remains too demanding for developing countries healthcare systems. Low cost and simple protocols, as the *wzi* HRM typing proposed here, represent a real opportunity for surveillance programs.

The use of hypervariable genes in HRM-based bacterial typing, such as *wzi* in *K. pneumoniae*, can drastically increase the discriminatory power of the method. With the large number of genomes available in databases, it is now possible to find the most variable genes for a species. Unfortunately, it is not easy to identify the best regions to design primers in such hypervariable genes, particularly when hundreds of different alleles are available. EasyPrimer can represent a useful tool to overcome this limit.

Methods

Isolates collections. We considered three strain collections: the background, the outbreak and the validation collections. The background collection includes 17 strains belonging to 17 different STs retrieved from two previously WGS typed bacterial collections: nine strains from Gaiarsa and colleagues[14] and eight strains form Gona and colleagues[15] (for details see Supplementary Table S3). The outbreak collection includes 11 *K. pneumoniae* isolates gathered during a 16 days nosocomial outbreak occurred in April 2018, in the Papa Giovanni XXIII hospital (Bergamo) (For details see Supplementary Tables S3 and S4). The validation collection includes 54 *K. pneumoniae* isolates belonging to six of the most epidemiologically relevant clades[11,12]:

- 17 strains belonging to ST307.
- 15 strains belonging to ST258 Clade 2, including ST258 *wzi_154* and ST512 *wzi_154*.
- seven strains belonging to ST258 Clade 1 (ST258 *wzi_29*).
- 11 strains belonging to ST101.
- two strains belonging to ST11.
- two strains belonging to ST15.

45/54 strains were isolated at San Raffaele hospital (Milan), 2/54 strains were isolated at Papa Giovanni XXIII hospital (Bergamo), 4/54 retrieved from the *K. pneumoniae* collection of Gona and colleagues[15] 3/54 retrieved from the *K. pneumoniae* collection of Gaiarsa and colleagues[14].

Neither ethics committee approval, nor informed consent were required as all collected data are fully anonymized, there was no contact with patients and/or their families and no interventions or changes to treatment and management were made, in accordance with institutional guidelines.

DNA extraction and whole-genome sequencing. The genomic DNA of the 45 strains isolated from San Raffaele hospital (Milan) were extracted using Maxwell 16 Cell DNA purification kit. The extracted DNA was sequenced using the NextSeq. 500 platform with 2×150 bp paired-ends runs, after Nextera XT library preparation.

The genomic DNA of the 13 strains isolated from Papa Giovanni XXIII hospital (Bergamo) was extracted using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The extracted DNA was sequenced using the Illumina Miseq platform with a 2×250 bp paired-end run, after Nextera XT library preparation.

The genomic DNA of the 12 strains previously sequenced by Gaiarsa and colleagues[14] were extracted using QIAsymphony Virus/Pathogen minikit, version 1 (Qiagen, Hilden, Germany) with the automated instrument QIAsimphony (Qiagen, Hilden, Germany) according to manufacturer's instructions.

The genomic DNA of the 12 strains previously sequenced by Gona and colleagues[15] were extracted using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) following the manufacturer's instructions.

Details about strains hospital isolation are reported in Supplementary Table S3.

High resolution melting primer design using EasyPrimer. The EasyPrimer tool was developed for the identification of the most suitable regions for primer design in HRM and, more in general, in pan-PCR experiments. Briefly, the tool starts from gene sequences in multi-fasta format. The sequences are considered as not aligned by default and they are aligned by Muscle software[22] as the first step of the pipeline (see Supplementary Fig. S7 and Supplementary Note S1). The user can also decide to submit aligned sequences (in multi-fasta format) and skip the alignment step. EasyPrimer evaluates the amount of genetic variation for each position of the alignment and identifies the most reliable regions for primer design. EasyPrimer flags as good candidates for primer design two conserved regions flanking a highly variable one (taking into consideration, in advance, the optimal lengths of primers and amplicon). The user can decide either to evaluate the variability of the amplicon considering HRM-detectable SNPs only (the best option for HRM primer design) or all the SNPs (the best setting for pan-PCR experiments). A detailed description of the algorithm is reported in the Supplementary Note S1.

To develop an HRM-based protocol for *K. pneumoniae* typing, we focused on the seven MLST genes and on the hypervariable capsular gene *wzi*[23]. The HRM primer pairs for five out of the seven *K. pneumoniae* MLST genes were already available in literature[7] (*infB, mdh, phoE, rpoB* and two pairs on *tonB*). For the remaining two MLST genes (*pgi* and *gapA*) and for the *wzi* capsular gene (two primer pairs) the primers were designed using EasyPrimer. For each gene, the sequences were downloaded from the BigsDB database (https://bigsdb.pasteur.fr, 218 alleles for *pgi*, 183 for *gapA* and 563 for *wzi*), EasyPrimer was run and primer pairs were designed on the basis of its output.

High-resolution melting assays. We performed HRM assays using the genomic DNA extracted from each of the 82 *K. pneumoniae* strains included in this work. On the strains of the background and outbreak collections we used each of the ten primer pairs mentioned above. On the validation collection strains we used *wzi-3*

and *wzi-4* primer pairs only. HRM analyses were performed on the BioRad CFX Connect real-time PCR System (BioRad, Hercules, California). Each 10 µl reaction contained: 5 µl of 2x SsoAdvanced Universal SYBR® Green Supermix (BioRad, Hercules, California), 0.4 µl of each primer (0.4 µM) and 1 µl of template DNA (25–50 ng/µl). The thermal profile was as follows: 98 °C for 2 min, 40 cycles of [95 °C for 7 s, 61 °C for 7 s, and 72 °C for 15 s], 95 °C for 2 min, followed by HRM ramping from 70–95 °C with fluorescence data acquisition at 0.5 °C increments. Three technical replicates were performed for each strain and for each gene analysed. Negative controls were added in every run and for each gene.

Comparison of the HRM primer pairs and schemes. We compared the discriminatory power of the ten HRM primer pairs on the 17 background collection strains. For each primer pair we calculated the average melting temperatures (aTms) of the three replicates for each strain and the relative strain distance matrix based on the obtained aTms. Thus, we compared the discriminatory power of the different primer pairs by comparing the relative distance matrix values using Wilcoxon test with Holm post-hoc correction.

Furthermore, we grouped the primer pairs in three schemes (MLST6, MLST8 and wzi) and we compared the relative strain distance matrices using Wilcoxon test with Holm post-hoc correction. The scheme compositions were as follows: the MLST6 scheme included the six primer pairs proposed by Andersson and colleagues[7] for five MLST genes (with two primer pairs for *tonB*); the MLST8 included all the MLST6 primer pairs, the primers for *pgi* and *gapA* (newly designed in this work using the EasyPrimer tool); the wzi scheme included the two primer pairs for the *wzi* gene (newly designed in this work). For details see Table 1 and Fig. 2.

Then, we compared the discriminatory power of MLST8 and wzi schemes by subtracting the relative distance matrices (wzi – MLST8) and studying the obtained matrix with a heatmap.

All these analyses were performed using R (https://www.r-project.org/) and the R libraries Ape and Gplots.

HRM-based outbreak reconstruction. From the aTms of the outbreak and background collections we calculated the distance matrices for MLST6, MLST8 and wzi primer schemes (for more details see above) and clustered the strains using the hierarchical clustering method implemented in the Hclust function in R.

Repeatability of the *wzi* HRM protocol. To test the repeatability of the *wzi* HRM typing protocol we analysed the wzi scheme aTms for all the strains of the three collections (17 background strains, 11 outbreak and 54 validation, see above). This allowed to compare the *wzi*-3 and *wzi*-4 aTms of multiple strains for each of the most epidemiologically relevant clades[11,12] (eight ST258 clade 1, 26 ST258 Clade 2, 19 ST307, 12 ST101, two ST11 and three ST15 strains) (see Table 2 and Supplementary Table S3).

We clustered the strains on the basis of wzi scheme aTms. Given the 0.5 °C sensitivity of the machine, we considered the strains with differences both in *wzi*-3 and *wzi*-4 aTMs < 0.5 as indistinguishable. Thus, we built a strain-to-strain network, in which the indistinguishable strains pairs were connected. Clusters were extracted from the network using the decompose igraph R function (https://www.r-project.org/). Lastly, the strain-to-strain network was converted to a dendrogram and merged to *wzi* aTms in a heatmap plot, using R (https://www.r-project.org/).

WGS-based strain typing. We retrieved the genome assemblies of the 12 *K. pneumoniae* strains previously WGS-typed by Gaiarsa and colleagues[14] from NCBI database and the reads files of the 12 strains by Gona and colleagues[15] using fastq-dump tool (accession numbers are reported in Supplementary Table S3).

We performed *de novo* assembly on the reads obtained from the 58 strains sequenced in this work and on the reads of the 12 strains retrived from Gona and colleagues[15] using SPAdes software[24].

We retrieved the sequences of the *K. pneumoniae* MLST gene alleles and the relative scheme tables from the BigsDB database. Thus, we determined the MLST profiles using an in-house Blastn-based Perl script.

We retrieved the *wzi* allele sequences from BigsDB database and we annotated the *wzi* allele present in each of the 82 genome assemblies included in the study by Blastn search and manual curation of the results.

We annotated the K-type of the 82 strains using Kaptive[23] on the genome assemblies.

Core-SNP-based phylogenetic reconstruction on outbreak strains. We aligned the reads obtained from the 11 outbreak strains against the NTUH_K2044 reference genome (accession number NC_016845.1), and performed the SNPs calling following the GATK best practice procedure. We masked SNPs localized within repeated regions, identified using MUMmer[25], or prophages, identified using PhiSpy[26], and we called the core-SNPs among the strains using an in-house Python script. Thus, we subjected the core-SNPs alignment to phylogenetic analysis as follows: the best evolutionary model was assessed by ModelTest-NG and phylogenetic reconstruction was performed using the selected best model, with RAxML8 software[27]. We evaluated the core-SNPs distances among the strains using the R Ape library (https://www.r-project.org/).

Data availability

We deposited all Illumina sequence data from the 58 strains in NCBI's Short Read Archive under BioProject ID ERP119329 and all Illumina data were deposited under BioProject ID PRJEB36171.

Received: 19 July 2019; Accepted: 30 December 2019;
Published online: 28 January 2020

References

1. Nutman, A. & Marchaim, D. How to: molecular investigation of a hospital outbreak. *Clinical Microbiology and Infection* **25**, 688–695 (2019).
2. Mehta, Y. *et al*. Guidelines for prevention of hospital acquired infections. *Indian J. Crit. Care Med.* **18**, 149–163 (2014).
3. Naze, F. *et al*. *Pseudomonas aeruginosa* Outbreak Linked to Mineral Water Bottles in a Neonatal Intensive Care Unit: Fast Typing by Use of High-Resolution Melting Analysis of a Variable-Number Tandem-Repeat Locus. *Journal of Clinical Microbiology* **48**, 3146–3152 (2010).
4. Mayerhofer, B. *et al*. Improved protocol for rapid identification of certain spa types using high resolution melting curve analysis. *PLoS One* **10**, e0116713 (2015).
5. Tamburro, M. & Ripabelli, G. High Resolution Melting as a rapid, reliable, accurate and cost-effective emerging tool for genotyping pathogenic bacteria and enhancing molecular epidemiological surveillance: a comprehensive review of the literature. *Ann. Ig.* **29**, 293–316 (2017).
6. Wong, Y. P., Chua, K. H. & Thong, K. L. One-step species-specific high resolution melting analysis for nosocomial bacteria detection. *J. Microbiol. Methods* **107**, 133–137 (2014).
7. Andersson, P., Tong, S. Y. C., Bell, J. M., Turnidge, J. D. & Giffard, P. M. Minim Typing – A Rapid and Low Cost MLST Based Typing Tool for *Klebsiella pneumoniae*. *PLoS ONE* **7**, e33530 (2012).
8. Price, E. P., Inman-Bamber, J., Thiruvenkataswamy, V., Huygens, F. & Giffard, P. M. Computer-aided identification of polymorphism sets diagnostic for groups of bacterial and viral genetic variants. *BMC Bioinformatics* **8**, 278 (2007).
9. Klaas, I. C. & Zadoks, R. N. An update on environmental mastitis: Challenging perceptions. *Transbound. Emerg. Dis.* **65**(Suppl 1), 166–185 (2018).
10. Perini, M. *et al*. Fast Klebsiella pneumoniae typing for outbreak reconstruction: an highly discriminatory HRM protocol on *wzi* capsular gene developed using EasyPrimer tool. *bioRxiv* (2019).
11. the EuSCAPE Working Group. *et al*. Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nat Microbiol* **4**, 1919–1929 (2019).
12. Wyres, K. L. *et al*. *Emergence and rapid global dissemination of CTX-M-15-associated Klebsiella pneumoniae* strain ST307. *J. Antimicrob. Chemother.* **74**, 577–581 (2019).
13. Deleo, F. R. *et al*. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. *Proc. Natl. Acad. Sci. USA* **111**, 4988–4993 (2014).
14. Gaiarsa, S. *et al*. Genomic epidemiology of *Klebsiella pneumoniae* in Italy and novel insights into the origin and global evolution of its resistance to carbapenem antibiotics. *Antimicrob. Agents Chemother.* **59**, 389–396 (2015).
15. Gona, F. *et al.* Comparison of core genome MLST, coreSNP and PFGE methods for *Klebsiella pneumoniae* cluster analysis. (Paper under revision on Microbial Genomics)
16. Jeremiah, C. J. *et al*. Differing epidemiology of two major healthcare-associated meticillin-resistant *Staphylococcus aureus* clones. *J. Hosp. Infect.* **92**, 183–190 (2016).
17. Larssen, K. W., Nor, A. & Bergh, K. Rapid discrimination of *Staphylococcus epidermidis* genotypes in a routine clinical microbiological laboratory using single nucleotide polymorphisms in housekeeping genes. *Journal of Medical Microbiology* **67**, 169–182 (2018).
18. Chan, M. S., Maiden, M. C. & Spratt, B. G. Database-driven multi locus sequence typing (MLST) of bacterial pathogens. *Bioinformatics* **17**, 1077–1083 (2001).
19. Comandatore, F. *et al*. Gene Composition as a Potential Barrier to Large Recombinations in the Bacterial Pathogen *Klebsiella pneumoniae*. *Genome Biol. Evol.* **11**, 3240–3251 (2019).
20. Qi, Y. *et al*. ST11, the dominant clone of KPC-producing *Klebsiella pneumoniae* in China. *J. Antimicrob. Chemother.* **66**, 307–312 (2011).
21. Roe, C. C., Vazquez, A. J., Esposito, E. P., Zarrilli, R. & Sahl, J. W. Diversity, Virulence, and Antimicrobial Resistance in Isolates From the Newly Emerging ST101 Lineage. *Front. Microbiol.* **10**, 542 (2019).
22. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797 (2004).
23. Wyres, K. L. *et al*. Identification of capsule synthesis loci from whole genome data. *Microb Genom* **2**, e000102 (2016).
24. Bankevich, A. *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
25. Delcher, A. L. *et al*. Alignment of whole genomes. *Nucleic Acids Research* **27**, 2369–2376 (1999).
26. Akhter, S., Aziz, R. K. & Edwards, R. A. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* **40**, e126 (2012).
27. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

Acknowledgements

Thanks to the Romeo ed Enrica Invernizzi Foundation and to professor Claudio Bandi for supporting and revising the project.

Author contributions

M.P. developed the tool, performed the H.R.M. experiments and drafted the paper. A.P. performed the H.R.M. experiments and revised the manuscript. S.P. designed the primers and revised the manuscript. D.D.C. implemented the tool online. M.C., F.G., F.V. collected the samples and extracted the D.N.A. P.M., D.M.C., C.F. collected the samples. G.V.Z. wrote the paper. F.C. conceived and designed the experiments and wrote the paper. All authors read, revised and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-57742-z.

Correspondence and requests for materials should be addressed to F.C.

Reprints and permissions information is available at www.nature.com/reprints.

Section C

Appendix 7

Repeatability and reproducibility of the wzi high resolution melting-based clustering analysis for *Klebsiella pneumoniae* typing.

Pasala AR*, Perini M*, Piazza A, Panelli S, Di Carlo D, Loretelli C, Cafiso A, Inglese S, Gona F, Cirillo DM, Zuccotti GV, Comandatore F

**contributed equally*

AMB Express. 2020 Dec 14;10(1):217. doi 10.1186/s13568-020-01164-7

Pasala *et al. AMB Expr* (2020) 10:217
https://doi.org/10.1186/s13568-020-01164-7

AMB Express

ORIGINAL ARTICLE

Open Access

Repeatability and reproducibility of the *wzi* high resolution melting-based clustering analysis for *Klebsiella pneumoniae* typing

Ajay Ratan Pasala[1], Matteo Perini[1], Aurora Piazza[2], Simona Panelli[1], Domenico Di Carlo[1], Cristian Loretelli[3], Alessandra Cafiso[4], Sonia Inglese[5], Floriana Gona[6], Daniela Maria Cirillo[7], Gian Vincenzo Zuccotti[1,8] and Francesco Comandatore[1*]

Abstract

High resolution melting (HRM) is a fast closed-tube method for nucleotide variant scanning applicable for bacterial species identification or molecular typing. Recently a novel HRM-based method for *Klebsiella pneumoniae* typing has been proposed: it consists of an HRM protocol designed on the capsular *wzi* gene and an HRM-based algorithm of strains clustering. In this study, we evaluated the repeatability and reproducibility of this method by performing the HRM typing of a set of *K. pneumoniae* strains, on three different instruments and by two different operators. The results showed that operators do not affect melting temperatures while different instruments can. Despite this, we found that strain clustering analysis, performed using MeltingPlot separately on the data from the three instruments, remains almost perfectly consistent. The HRM method under study resulted highly repeatable and thus reliable for large studies, even when several operators are involved. Furthermore, the HRM clusters obtained from the three different instruments were highly conserved, suggesting that this method could be applied in multicenter studies, even if different instruments are used.

Introduction

Klebsiella pneumoniae is a Gram-negative opportunistic pathogen often present in the gut of healthy individuals but also able to cause severe infections. Furthermore, the bacterium is one of the most important nosocomial pathogens, causing healthcare-acquired infections worldwide with a mortality rate ranging from 20 to 70% (Angus et al. 2001; Mayr et al. 2014). Indeed, *K. pneumoniae* has been described as an "urgent threat to human health" by the United States Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) (Munoz-Price et al. 2013). Genomic studies revealed that, despite the high genetic variability of the bacterium (Gaiarsa et al. 2015; Holt et al. 2015), most of the nosocomial outbreaks are caused by only few Multi Drug Resistant (MDR) clones, in particular ST258, ST512, ST307, ST11, ST101 and ST15 (David et al. 2019; Wyres et al. 2019). Thus, a genetic-based nosocomial surveillance can represent an important tool to promptly detect *K. pneumoniae* high risk clones in the hospital setting.

High Resolution Melting (HRM)-based typing is a promising tool for clinical and epidemiological applications (Tamburro and Ripabelli 2017). HRM is a fast closed-tube method to discriminate nucleotide variants on the basis of PCR amplicon melting temperature. The method is particularly reliable for nosocomial surveillance: it can be performed on most real-time PCR instruments; the entire protocol takes ~ 5 h and it is inexpensive (~ 5$ per sample).

Perini and colleagues (Perini et al. 2020a) proposed an HRM-based method for *K. pneumoniae* typing. The

*Correspondence: francesco.comandatore@unimi.it
[1] Department of Biomedical and Clinical Sciences "L. Sacco", Università di Milano, Pediatric Clinical Research Center "Romeo and Enrica Invernizzi", Milan, Italy
Full list of author information is available at the end of the article

method consists of an HRM protocol designed on the hypervariable capsular gene *wzi* and followed by a strains clustering analysis based on the melting temperatures. The method was able to discriminate most of the *K. pneumoniae* Sequence Types (STs) known as "high risk" (Perini et al. 2020b).

Different real-time PCR/HRM instruments can vary in thermal precision and melting temperature acquisition rate (Wittwer 2009; Li et al. 2014). In literature, studies on HRM protocols designed for human samples revealed that the measured melting temperature can vary among the instruments (Wittwer 2009; Li et al. 2014). In this study, we evaluated the repeatability and reproducibility (Schulten et al. 2000; Bustin et al. 2009) of the HRM method described by Perini and colleagues (Perini et al. 2020b) repeating HRM typing on three different instruments by two operators.

Materials and methods

Dataset selection

The dataset for the analyses was a subset of the 82-strains collection analyzed by Perini and colleagues (Perini et al. 2020b). Perini and colleagues grouped the 82 strains in a total of seven clusters labelled as: "Cluster ST258 Clade1-like" (including 10 strains), "Cluster ST258 Clade2-like" (27 strains), "Cluster ST11/ST101-like" (22 strains), "Cluster ST307-like" (19 strains), "Cluster ST10/ST147-like" (including two strains), "Cluster ST15-like" (one strain), "Cluster ST149-like" (one strain). To test the repeatability of the protocol we selected a subset of 43 out of the 82 strains, considering the strains from the same cluster as biological replicates. More in detail, we selected 10 strains from the "Cluster ST258 Clade1-like", 10 from the "Cluster ST258 Clade2-like", 10 from the "Cluster ST11/ST101-like", 10 from the "Cluster ST307-like", one from the "Cluster ST10/ST147-like", one from the "Cluster ST15-like" and one from the "Cluster ST149-like" cluster.

All the isolates were retrieved from clinical collections and they were isolated from hospital patients (see Additional file 1: Table S1). The 43 *K. pneumoniae* strains belong to eleven different STs, including the highly epidemiologically relevant ST258, ST512, ST307, ST11, ST101 and ST15 (David et al. 2019; Wyres et al. 2019) (see Additional file 1: Table S1 for details).

Real-time PCR/HRM instruments

HRM analysis (see below) was performed on three different real-time PCR/HRM instruments:

- Bio-Rad CFX96 real-time PCR machine (Bio-Rad Laboratories), from here "CFX96".
- Eco Real-Time PCR System (Illumina), from here "Eco_RT".
- QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems), from here "QS_6Flex".

The three instruments were placed in three different laboratories in two cities, for details see Additional file 2: Table S2.

DNA extraction

Bacterial strains were freshly streaked on MacConkey agar plate and incubated overnight at 37 °C; then a single colony was inoculated into 5 mL of LB broth (DifcoTM) and incubated overnight at 37 °C with vigorous shaking. For each strain, 1×10^9 cells have been used as starting material for total DNA extraction using the DNeasy blood and tissue kit (Qiagen) following the manufacturer's instructions.

High resolution melting analysis

For each strain, the extracted DNA was subjected to six HRM analyses: two operators (AP and MP) independently performed the HRM analyses on the three real-time PCR/HRM instruments listed above. In each of the six HRM analysis, three technical replicates were performed for each strain, amplified with the two primer pairs in the Perini and colleagues (Perini et al. 2020b) HRM protocol (*wzi-3* and *wzi-4*). Negative controls were added in every HRM analysis for each primer pair.

The HRM reaction mix (10 µl) contained: 5 µl of 2x SsoAdvanced Universal SYBR® Green Supermix (BioRad, Hercules, California), 0.4 µl of each primer (0.4 µM) and 1 µl of template DNA (25–50 ng/µl). The thermal profile was as follows: 98 °C for 2 min, 40 cycles of [95 °C for 7 s, 61 °C for 7 s, and 72 °C for 15 s], 95 °C for 2 min, followed by HRM ramping from 70 to 95 °C. Fluorescence data were acquired at increments of 0.5 °C for CFX96, 0.3 °C QuantStudio 6 Flex, and 0.1 °C for Eco Real-Time PCR System. Each CFX96 and QuantStudio 6 Flex HRM analysis was performed in a single 96-well optical plate, while for the Eco Real-Time PCR System each HRM analysis required three 48-well optical plates.

DNA and reagents aliquots for all the experiments were prepared in advance to reduce the risk of contamination. In each experiment, the two operators independently prepared the HRM mixes in a pre-PCR 'clean' room using the same pipettes each day for each individual experiment.

Statistical analysis

For each strain, the average of the melting temperatures (aTm) obtained from the three technical replicates were computed for *wzi-3* and *wzi-4* primer sets. A preliminary

qualitative comparison of the aTms obtained by the different instruments and operators was performed reporting the median, minimum and maximum temperature differences for *wzi-3* and *wzi-4* aTms and plotting the aTm distributions by boxplots.

Then, the effects of operators or instruments on *wzi-3* and *wzi-4* aTms were investigated as independent and as combined factors. The statistical analyses were performed on 1,000,000 bootstrapping strain subsets, randomly selected with replacement. For each subset, aTms were analyzed using R v.3.6.1 (https://www.r-project.org/) as follows:

1. Independent factors:
   - Normality distribution of aTm values was tested by the Shapiro–Wilk test.
   - Homoskedasticity variances of aTm values between operators and among the three instruments were compared using F test and Bartlett test, respectively.
   - If aTms were normally distributed, the operators were compared using *t* test (applying the Welch approximation in case of heteroskedasticity variance), otherwise using Mann–Whitney test.
   - If aTms were normally distributed, the instruments were compared using one-way ANOVA (ANalysis Of VAriance, or Welch one-way ANOVA in case of heteroskedasticity variance), otherwise Kruskal–Wallis test.
2. Combined factors:
   - The effects of operator and instrument, and their interactions, were evaluated using the non-parametric analysis of variance implemented in the art function of ARTool R 3.6.1 package (Kay 2020).

Then, we evaluated the percentage of subsets for which the effect of operator, instrument or their interaction were significant (p-value < 0.05).

HRM clustering analysis

For each instrument (CFX96, Eco_RT and QS_6Flex), the obtained melting temperatures were subjected to HRM clustering analysis using the MeltingPlot tool (Perini et al. 2020b) (available at https://skynet.unimi.it/index.php/tools/meltingplot/). More in detail, for each strain the tool computes the replicates average melting temperatures (aTms) for each primer set. Using the igraph R library (http://igraph.org/), the tool builds a graph connecting the strains with aTm distance ≤ 0.5 °C for every primer set, and it clusters the strains on the basis of their betweenness (Perini et al. 2020b). Since the clustering algorithm is based on melting temperature only, the melting curves were not subjected to normalization, smoothing or background adjustment.

Results

High resolution melting analysis

Forty-three: *Klebsiella pneumoniae* strains were subjected to the HRM protocol proposed by Perini and colleagues (Perini et al. 2020b) using *wzi-3* and *wzi-4* primer sets. The analysis was repeated by two operators on three different instruments, for a total of six experiments per strain. The resulting average melting temperatures (aTms) are reported in Additional file 3: Table S3.

Statistical analysis

Boxplots of the aTms from the six HRM experiments are reported in Fig. 1. The median, minimum and maximum aTm differences among operators/instruments are reported in Additional file 4: Table S4, for a total of 15 operator/instrument combinations per primer set.

For either primer sets, the combinations of different operators on the same machine gave a maximum difference below or equal to 0.5 °C, the threshold set by Perini and colleagues (Perini et al. 2020b) for clustering analysis. Conversely, all the combinations among different machines gave maximum differences above 0.5 °C.

The results of the statistical analysis are summarized in Additional file 5: Table S5. For both *wzi-3* and *wzi-4*, T-test/Wilcox tests on the operators resulted non-significant (p-value ≥ 0.05) for the 100% of the 1,000,000 bootstrap replicates, while the instrument resulted significant (p-value < 0.05) for all of them. For *wzi-3*, the non-parametric analysis of the variance found the operator to be significant for 13.37% of the bootstraps replicates, the instrument for 100% of replicates, and the interaction among the two factors (operator/instrument) was never significant among the bootstrap replicates. For *wzi-4*, the instrument was found significant for 100% of the bootstrap replicates, while the operator and the operator-instrument interaction were never significant.

HRM clustering analysis

For each of the three instruments (CFX96, Eco_RT and QS_6Flex), the 43 strains were clustered on the basis of *wzi-3* and *wzi-4* aTms using the MeltingPlot tool (Perini et al. 2020b). The graph clustering obtained by MeltingPlot is reported in Fig. 2. All the obtained melting curves coloured following the obtained clusters are reported in Fig. 3. The clusters obtained from MeltingPlot analysis on the basis of the melting temperatures obtained from each instrument are reported in Additional file 1: Table S1. For each instrument, the clustering analysis grouped the 43 strains in five main clusters. The strains clusters resulted almost exactly conserved among the instruments. Each

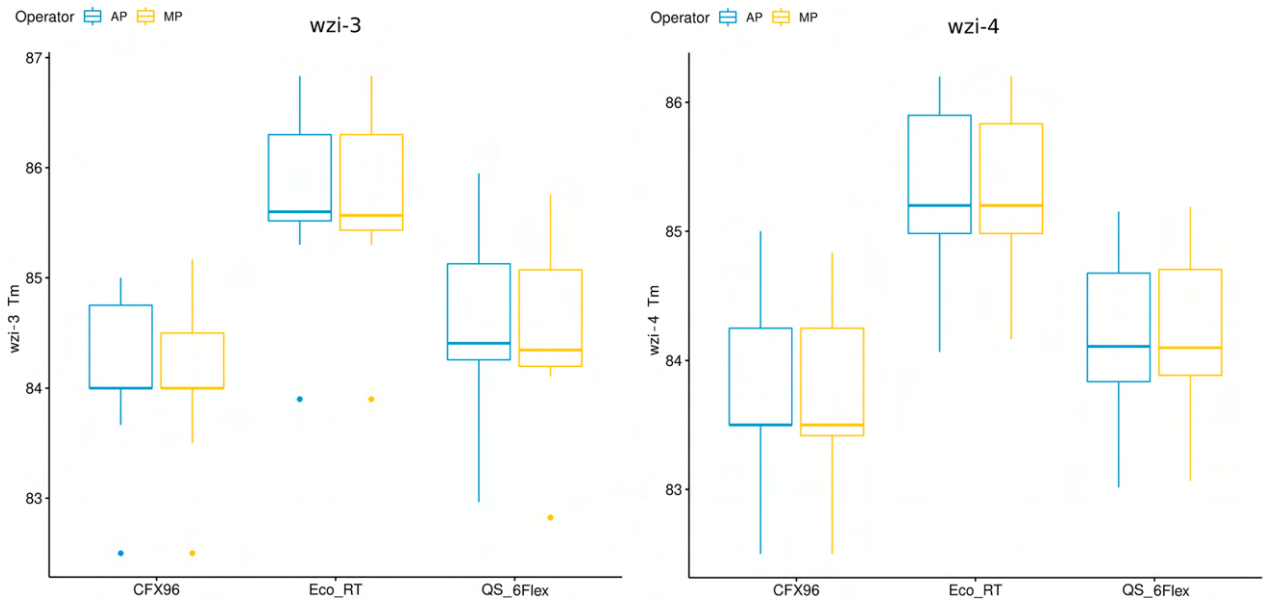


Fig. 1 Boxplot of the wzi-3 and wzi-4 average melting temperatures. The distributions of *wzi-3* and *wzi-4* average melting temperatures (aTm) obtained by each instrument ("CFX96" for Biorad CFX96; "Eco_RT" for Illumina Eco Real-Time; "QS_6Flex" for Applied Biosystem QuantStudio 6 Flex) and operator (AP and MP; colored in blue and yellow, respectively) are shown. Boxes range between the 25th and the 75th quartiles, and bold horizontal lines represent the median values

of the 43 strains in this work was subjected to HRM analysis twice, one per operator, for a total of 86 instances (each instance is represented by a node in the graph in Fig. 2). The clustering analysis performed using MeltingPlot on the melting temperatures obtained using Illumina Eco Real-Time (ECO_RT) and Applied Biosystem QuantStudio 6 Flex (QS_6Flex), gave the same result for 85 out of 86 instances (98.8%). Indeed, KP13-19 strain for AP operator was classified as Undetermined in the Illumina Eco Real-Time (ECO_RT). On the other hand, the Biorad CFX96 clusters show a total of 10 instances out of 86 (11.6%) that clustered differently from the other two instruments, and two instances (2.3%) classified as Undetermined. More in details, the difference are the following: (i) KP27-19 (for MP operator) and KP27-19 (AP) were assigned to a separate cluster in CFX96 dataset, while to the cluster 04 for the other instruments; (ii) 52BG (MP and AP) were assigned to the cluster 06 in CFX96 dataset, while to cluster 01 for the other instruments. (iii) KP232 (MP), KP252 (MP) were assigned to cluster 09 in CFX96, while to cluster 04 in the other instruments. (iv) KP359 (MP and AP) were assigned to cluster 10 and 11 respectively in CFX96, while they were assigned to cluster 04 in the other instruments. (v) 57BG (MP) and KP246 (MP) were assigned to cluster 07 and 8 respectively in CFX96, while they were assigned to cluster 03 in the other instruments. (vi) KP18-19 (MP) and KP4-19 (MP) were classified as undetermined in the CFX96 while they were classified as cluster 01 in the other instruments.

We compared the clusters obtained by the three instruments to the Perini et al. 2020b clusters, obtained using a CFX96 instrument. As shown in Additional file 6: Table S6, CFX96 experiments from this work clustered 78 instances out of 86 (90.6%) in accordance with Perini et al. 2020b. Among the eight instances differently clustered included, two were classified as Undetermined. The experiments performed using ECO_RT and QS_6Flex grouped the instances in less clusters than Perini et al. 2020b. More in details, both for ECO_RT and QS_6Flex, cluster 01 contains all the instances previously grouped in the Cluster ST10/ST147-like and Cluster ST11/ST101-like clusters, and cluster 04 contains the instances grouped in the Cluster ST149-like and Cluster ST258 Clade2-like. ECO_RT experiments classified one instance as Undetermined. Overall, ECO_RT experiments led to cluster 81 instances out of 86 (94.2%) in accordance with Perini et al. 2020b clusters, while QS_6Flex 84 out of 86 (97.7%).

Discussion

In the present study we evaluated the repeatability and reproducibility of the *wzi* HRM protocol for *Klebsiella pneumoniae* typing (Perini et al. 2020b). For this validation study, we selected a subset of 43 strains representative of the entire collection of 82 *K. pneumoniae* isolates

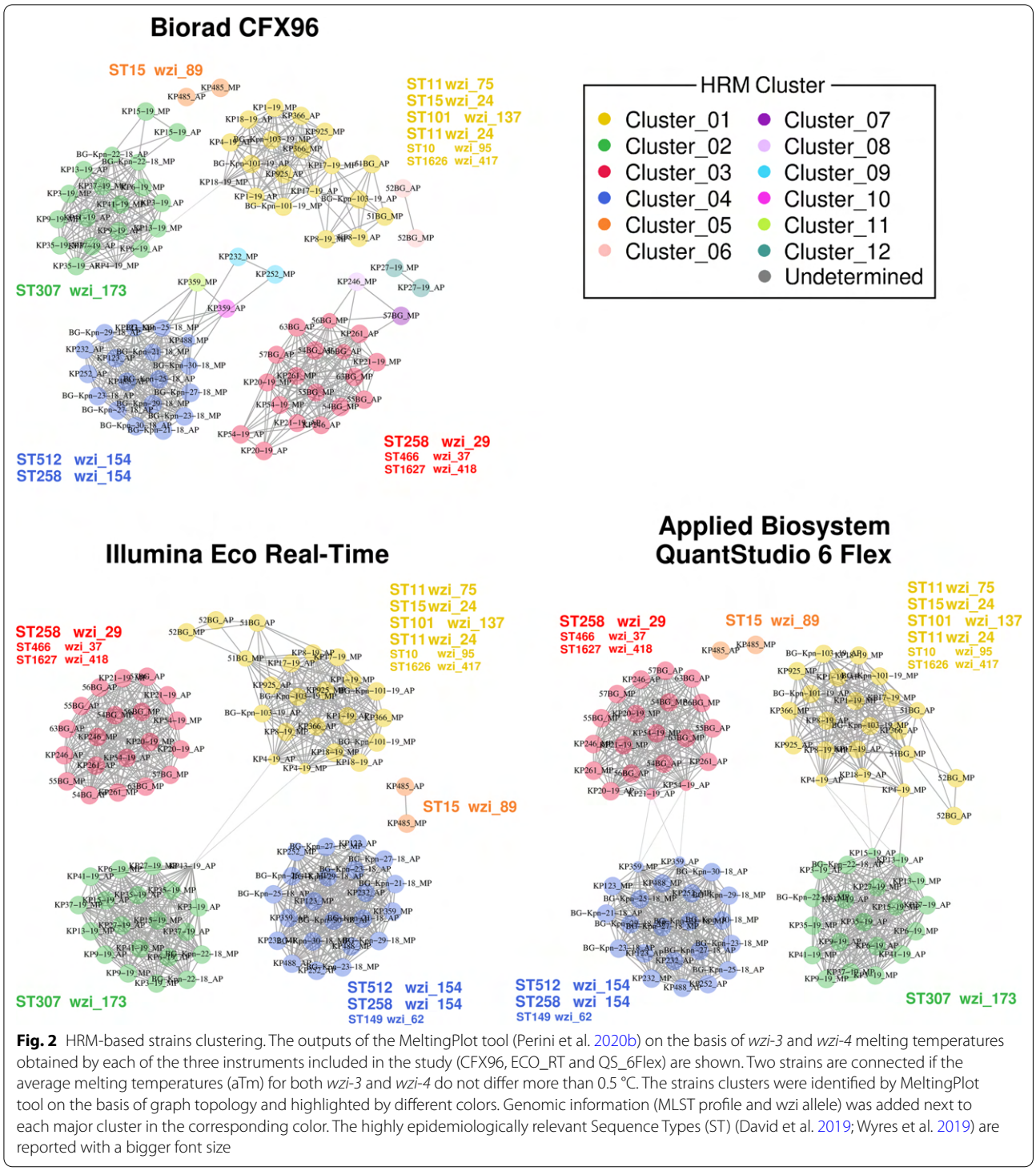


Fig. 2 HRM-based strains clustering. The outputs of the MeltingPlot tool (Perini et al. 2020b) on the basis of *wzi-3* and *wzi-4* melting temperatures obtained by each of the three instruments included in the study (CFX96, ECO_RT and QS_6Flex) are shown. Two strains are connected if the average melting temperatures (aTm) for both *wzi-3* and *wzi-4* do not differ more than 0.5 °C. The strains clusters were identified by MeltingPlot tool on the basis of graph topology and highlighted by different colors. Genomic information (MLST profile and wzi allele) was added next to each major cluster in the corresponding color. The highly epidemiologically relevant Sequence Types (ST) (David et al. 2019; Wyres et al. 2019) are reported with a bigger font size

typed by Perini and colleagues (Perini et al. 2020b). The dataset strains belong to eight different Multi Locus Sequence Typing profiles, including the most epidemiologically relevant clones (David et al. 2019; Wyres et al. 2019), i.e. ST258, ST512, ST307, ST11, ST101 and ST15. In this study we validated the protocol on three real-time PCR/HRM instruments: Biorad CFX96, Illumina Eco Real-Time and Applied Biosystem QuantStudio 6 Flex. The three instruments were placed in three different laboratories in two cities. We also studied the effect

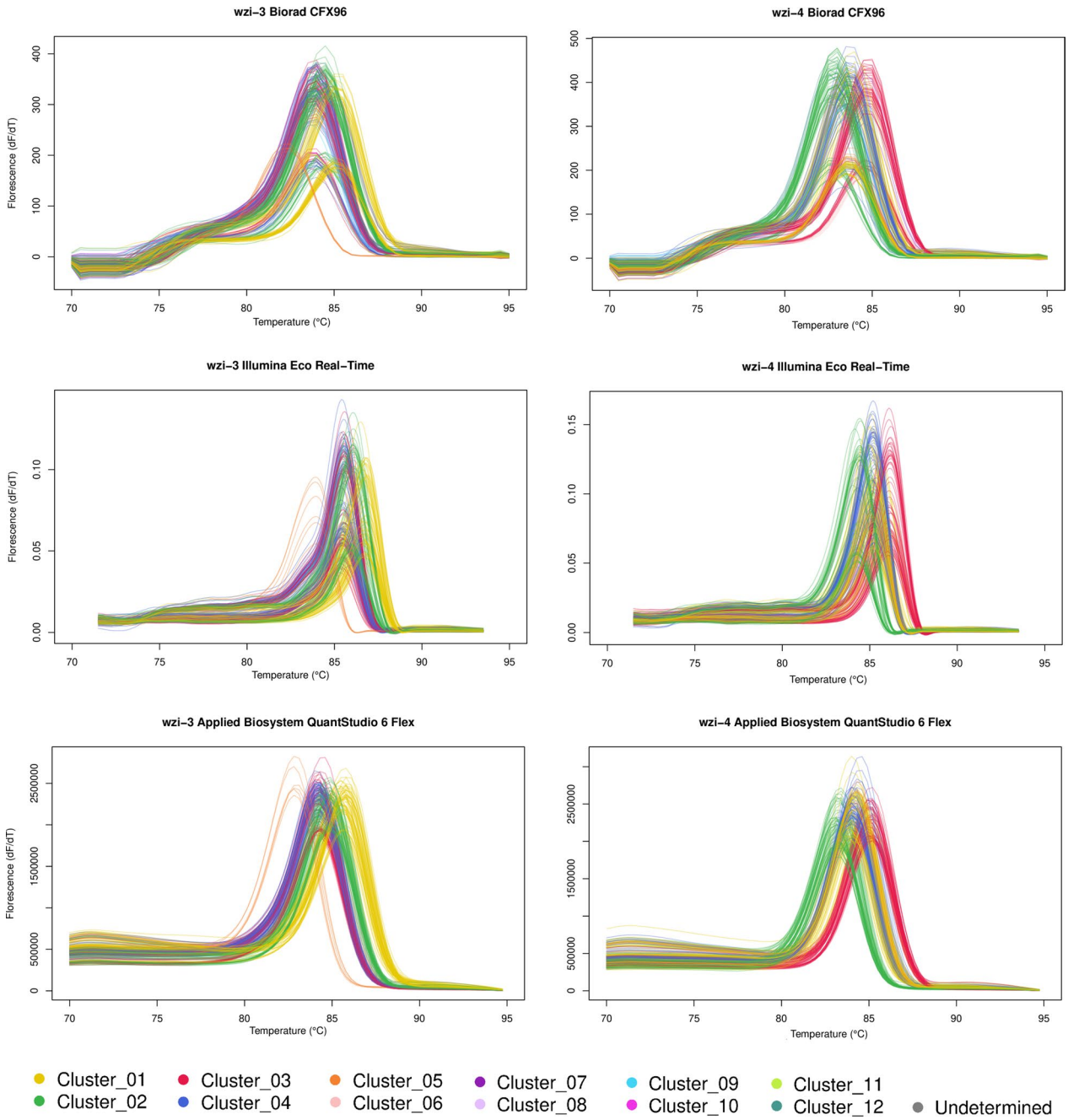


Fig. 3 HRM curves. All the raw derivative High Resolution Melting curves obtained for the experiments are reported. The colors represent the clusters found by MeltingPlot tool (Perini et al. 2020b) and correspond to those used in Fig. 2

of different operators on the results. Two operators (AP and MP) independently performed HRM analysis on the 43 *K. pneumoniae* strains on the three instruments.

Comparing the melting temperatures obtained by the two operators (AP and MP), statistical analyses revealed that the operator does not affect the measured melting temperatures. Furthermore, the operators did not affect the results of clustering analysis (see Fig. 2 and Additional file 1: Table S1): in almost all cases, the same strain analysed independently by the two operators is assigned to the same HRM cluster. This results shows that the HRM protocol proposed by Perini and colleagues (Perini et al. 2020b) is highly repeatable and thus reliable for large scale studies, even if several operators are involved.

Conversely, the instruments resulted to significantly affect the measured melting temperatures (Additional file 5: Table S5). The melting temperatures obtained from the same collection of strains by different instruments can significantly vary (see Fig. 1 and Additional file 4: Table S4). As shown in Additional file 4: Table S4, the differences among the melting temperatures obtained for the same strain by the three instruments often exceed 0.5 °C, the threshold used for the HRM clustering analysis (Perini et al. 2020b). For this reason, melting temperatures obtained using different instruments can not be included in the same HRM clustering analysis. Therefore, a clustering analysis can only be performed using melting temperatures obtained from the same instrument. Nevertheless, the results of the HRM clustering analyses performed on the melting temperatures obtained from three different instruments are almost perfectly conserved (Fig. 2, Fig. 3 and Additional file 1: Table S1). The repeatability of the protocol is also evident comparing the clusters obtained by Perini et al. 2020b with those obtained in this work (Additional file 6: Table S6). The few discrepancies regarded a limited number of strains, while almost all the strains were coherently classified in each of the four independent experiments (Perini et al. 2020b and each of the three instruments used in this work).

The clustering analysing applied by MeltingPlot is based only on the melting temperatures measured by the real time PCR instrument. For this reason the melting curves are not normalized, smoothed or background adjusted as these procedures change the overall shape of the curve but not the melting temperature, i.e. the peak of the derivative melt curve (as in Fig. 3). Despite this approach may cause the loss of information present in the curve, it reduces the influence of experimental noise, thus increasing the repeatability of the obtained results. To compensate this information loss, we developed the HRM protocol on an highly variable gene (*wzi*) and, in particular, on two gene regions rich in HRM detectable SNPs. In this way it is possible to have a wider range of melting temperatures than in HRM typing protocols developed around one or a few specific SNPs.

The main limit of this typing protocol is the low sensitivity of the HRM assay. In particular, using two targets (*wzi-3* and *wzi-4*) for the HRM assay, the protocol is able to discriminate only strains for which the melting temperatures differ more than 0.5 °C for at least one of the targets. For this reason, strains harbouring different wzi alleles can cluster together (Fig. 2). On the other hand, this limited sensitivity gives robustness to the final results: whenever two strains are clustered separately, they harbour different *wzi* alleles and, therefore, they likely belong to different clones.

In this work we can conclude that the *wzi* HRM protocol is highly repeatable on the same instrument, without significant effect of the operator. Considering the low cost per sample of this protocol (~ 5$ per sample) and short time required to accomplish the analysis (~ 5 h), it is a typing method suitable for the real time monitoring of the epidemiological scenario in a hospital setting. A day by day monitoring of the *K. pneumoniae* clones circulating in a hospital can allow the prompt detection of the emergence of a nosocomial outbreak and to follow the spreading of the outbreak clone among the patients/wards.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13568-020-01164-7.

Additional file 1: Table S1. Strains genomic information and Results of clustering analysis. For each strain included in the study the cluster assigned by MeltingPlot tool on the basis of *wzi-3* and *wzi-4* melting temperatures obtained by each instrument and operator ("CFX96" for Biorad CFX96; "Eco_RT" for Illumina Eco Real-Time; "QS_6Flex" for Applied Biosystem QuantStudio 6 Flex) are reported. The relative hospital of isolation (Hospital), Multi Locus Sequence Typing (MLST) profiles, the *wzi* alleles (*wzi*) and the clusters described in Perini et al. 2020b (Perini et al. 2020b) (Perini et al. 2020b Clusters)are also reported.

Additional file 2: Table S2. Instruments information. For each instrument used in this work, the model, the short name (used in manuscript, tables and figure), the sensitivity and the location are reported.

Additional file 3: Table S3. Strains melting temperatures. The instrument used for the HRM analysis, the operator who performed the HRM analysis, the primer set used, the melting temperature replicates (T1, T2 and T3) and their average melting temperature (aTm) are reported for each strain.

Additional file 4: Table S4. Melting temperature differences among instruments and operators. The median of the average melting temperatures differences for each strain among instruments and operators are reported. In brackets, minimum and maximum difference values are reported. The comparisons between the two operators on the same instrument are reported in bold.

Additional file 5: Table S5. Results of statistical analyses. The results of statistical analyses are reported.

Additional file 6: Table S6. Cross table of the clusters found in Perini et al. 2020b and the clusters of each instrument used in this work is reported. The discrepancies from the clusters found in Perini et al. 2020b are reported in bold.

Acknowledgements
Thanks to the Romeo ed Enrica Invernizzi Foundation.

Authors' contributions
ARP and MP performed the experiments and wrote manuscript; AP and SP extracted the DNA and drafted the manuscript; DDC performed the statistical analysis; CL, AC, SI, FG, DMC prepared the samples and revised the manuscript; GVZ revised the manuscript; FC conceived the work, performed the statistical and clustering analyses, wrote and revised the manuscript. All authors read and approved the final manuscript.

Funding
Not applicable.

Availability of data and materials
The isolates analysed in this work is a subset of an isolate collection analysed also in a published work (Perini et al. 2020b): Short Read Archive ERP119329; BioProject PRJEB36171.
MeltingPlot, the tool used for the clustering analysis, is available online at https://skynet.unimi.it/index.php/tools/meltingplot/.

Ethics approval and consent to participate
All the data were retrieved from a published work (Perini et al. 2020b), where it is stated: "Neither ethics committee approval, nor informed consent were required as all collected data are fully anonymized, there was no contact with patients and/or their families and no interventions or changes to treatment and management were made, in accordance with institutional guidelines."

Consent for publication
No individual person's data where used in this work.

Competing interests
No competing interests to declare.

Author details
[1] Department of Biomedical and Clinical Sciences "L. Sacco", Università di Milano, Pediatric Clinical Research Center "Romeo and Enrica Invernizzi", Milan, Italy. [2] Department of Clinical Surgical Diagnostic and Pediatric Sciences, Microbiology and Clinical Microbiology Unit, University of Pavia, Pavia, Italy. [3] International Center for T1D, Pediatric Clinical Research Center "Romeo Ed Enrica Invernizzi", Department of Biomedical and Clinical Science L. Sacco, Università di Milano, Milan, Italy. [4] Department of Veterinary Medicine, Università di Milano, Lodi, Italy. [5] Microbiology and Virology Unit, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy. [6] Laboratorio Microbiologia e Virologia-Ospedale San Raffaele Dibit, 2-San Gabriele 1, Milan, Italy. [7] Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan, Italy. [8] Department of Pediatrics, Children's Hospital Vittore Buzzi, Università di Milano, Milan, Italy.

Received: 3 July 2020 Accepted: 8 December 2020
Published online: 14 December 2020

References

Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR (2001) Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med 29(7):1303–1310

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55(4):611–622

David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, Abudahab K, Goater R, Giani T, Errico G, Aspbury M, Sjunnebo S, Feil EJ, Rossolini GM, Aanensen DM, Grundmann H (2019) Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. Nat Microbiol 4(11):1919–1929

Gaiarsa S, Comandatore F, Gaibani P, Corbella M, Valle CD, Epis S, Scaltriti E, Carretto E, Farina C, Labonia M, Landini MP, Pongolini S, Sambri V, Bandi C, Marone P, Sassera D (2015) Genomic Epidemiology of *Klebsiella pneumoniae* in Italy and Novel Insights into the Origin and Global Evolution of Its Resistance to Carbapenem Antibiotics. Antimicrob Agents Chemother 389:396. https://doi.org/10.1128/aac.04224-14

Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, Brisse S, Cao H, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Van Nguyen K, Nguyen TV, Dao TT, Mensink M, Minh VL, Nhu NTK, Schultsz C, Kuntaman K, Newton PN, Moore CE, Strugnell RA, Thomson NR (2015) Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci USA 112(27):E3574–E3581

Kay M (2020) *ARTool*. Github. https://github.com/mjskay/ARTool. Accessed: 18 April 2020

Li M, Zhou L, Palais RA, Wittwer CT (2014) Genotyping accuracy of high-resolution DNA melting instruments. Clin Chem 60(6):864–872

Mayr FB, Yende S, Angus DC (2014) Epidemiology of severe sepsis. Virulence 5(1):4–11

Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, Cornaglia G, Garau J, Gniadkowski M, Hayden MK, Kumarasamy K, Livermore DM, Maya JJ, Nordmann P, Patel JB, Paterson DL, Pitout J, Villegas MV, Wang H, Woodford N, Quinn JP (2013) Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. Lancet Infect Dis 13(9):785–796

Perini M, Batisti Biffignandi G, Di Carlo D, Pasala AR, Piazza A, Panelli S, Zuccotti GV, Comandatore F (2020) MeltingPlot: a user-friendly online tool for epidemiological investigation using High Resolution Melting data. https://www.medrxiv.org/. https://doi.org/10.1101/2020.06.16.20132142

Perini M, Piazza A, Panelli S, DiCarlo D, Corbella M, Gona F, Vailati F, Marone P, Cirillo DM, Farina C, Zuccotti G, Comandatore F (2020) EasyPrimer: user-friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on *wzi* gene for fast *Klebsiella pneumoniae* typing. Sci Rep 10(1):1307

Schulten SM, Veld PH, Nagelkerke NJD, Scotter S, de Buyser ML, Rollier P, Lahellec C (2000) Evaluation of the ISO 7932 standard for the enumeration of Bacillus cereus in foods. Int J Food Microbiol 57(1):53–61

Tamburro M, Ripabelli G (2017) High Resolution Melting as a rapid, reliable, accurate and cost-effective emerging tool for genotyping pathogenic bacteria and enhancing molecular epidemiological surveillance: a comprehensive review of the literature. Annali di igiene: medicina preventiva e di comunita 29(4):293–316

Wittwer CT (2009) High-resolution DNA melting analysis: advancements and limitations. Human mutation. Wiley. https://onlinelibrary.wiley.com/doi/abs/10.1002/humu.20951

Wyres KL, Hawkey J, Hetland MAK, Fostervold A, Wick RR, Judd LM, Hamidian M, Howden BP, Löhr IH, Holt KE (2019) Emergence and rapid global dissemination of CTX-M-15-associated *Klebsiella pneumoniae* strain ST307. J Antimicrob Chemother 74(3):577–581

Publisher's Note

Section C

Appendix 8

MeltingPlot, a user-friendly online tool for epidemiological investigation using High Resolution Melting data.

Perini M, Batisti Biffignandi G, Di Carlo D, Pasala AR, Piazza A, Panelli S, Zuccotti G V, Comandatore F

BMC Bioinformatics. 2021 Feb 18 ;22(1):76. doi: 10.1186/s12859-021-04020-y

BMC Bioinformatics

SOFTWARE

Open Access

MeltingPlot, a user-friendly online tool for epidemiological investigation using High Resolution Melting data

Matteo Perini[1], Gherard Batisti Biffignandi[2], Domenico Di Carlo[1], Ajay Ratan Pasala[1], Aurora Piazza[2], Simona Panelli[1], Gian Vincenzo Zuccotti[1,3] and Francesco Comandatore[1*]

*Correspondence: francesco.comandatore@unimi.it
[1] Department of Biomedical and Clinical Sciences "L. Sacco", Pediatric Clinical Research Center "Romeo and Enrica Invernizzi", Università Di Milano, 20157 Milan, Italy
Full list of author information is available at the end of the article

Abstract

Background: The rapid identification of pathogen clones is pivotal for effective epidemiological control strategies in hospital settings. High Resolution Melting (HRM) is a molecular biology technique suitable for fast and inexpensive pathogen typing protocols. Unfortunately, the mathematical/informatics skills required to analyse HRM data for pathogen typing likely limit the application of this promising technique in hospital settings.

Results: MeltingPlot is the first tool specifically designed for epidemiological investigations using HRM data, easing the application of HRM typing to large real-time surveillance and rapid outbreak reconstructions. MeltingPlot implements a graph-based algorithm designed to discriminate pathogen clones on the basis of HRM data, producing portable typing results. The tool also merges typing information with isolates and patients metadata to create graphical and tabular outputs useful in epidemiological investigations and it runs in a few seconds even with hundreds of isolates. Availability: https://skynet.unimi.it/index.php/tools/meltingplot/.

Conclusions: The analysis and result interpretation of HRM typing protocols can be not trivial and this likely limited its application in hospital settings. MeltingPlot is a web tool designed to help the user to reconstruct epidemiological events by combining HRM-based clustering methods and the isolate/patient metadata. The tool can be used for the implementation of HRM based real time large scale surveillance programs in hospital settings.

Keywords: High Resolution Melting, Epidemiology, Bacterial typing, Real time surveillance, Nosocomial infection, Outbreak reconstruction, Web interface

Background

The rapid typing of pathogens is pivotal to perform fast epidemiological investigations and to detect and block outbreaks. High-Resolution Melting (HRM) analysis is a single-step molecular biology technique able to determine the melting temperature of a PCR amplicon. The main output of an HRM assay is the melting curve, which indicates the denaturation level of the DNA amplicon in relation to temperature. The melting

temperature is defined as the temperature in which half of the DNA duplex is dissociated. Considering that the melting temperature depends on the nucleotide composition of the amplicon, melting temperature can be used to discriminate sequence alleles. For each isolate, HRM analysis interrogates *n* specific genomic regions returning *n* melting temperatures, where each genomic region is defined by a specific PCR primer set. As stated above, The melting temperatures of each interrogated genomic region depend on its nucleotide composition. Consequently, melting temperatures can be used to cluster the isolates in a *n*-dimensional space. Previous works suggested that HRM data can be used for fast pathogen typing [1–4]. Recently, we developed a graph-based algorithm for isolate clustering on the basis of HRM temperatures and we showed that this approach is able to discriminate the most epidemiologically relevant clones of *Klebsiella pneumoniae* [4], one of the most important nosocomial pathogens world-wide [5]. In the same work, we compared this HRM *K. pneumoniae* typing protocol with Multi Locus Sequence Typing (MLST) and Whole Genome Sequencing (WGS) approaches: HRM typing protocol showed a discrimination power comparable to MLST on clinical *K. pneumoniae* isolates [4]. Additionally, in another work we showed that the protocol is highly reproducible and repeatable among instruments and operators [6].

Here we present MeltingPlot, a tool for rapid epidemiological investigation using HRM data. The tool implements an evolution of the clustering algorithm we already published [4]. Moreover, MeltingPlot merges HRM typing information with metadata of isolates and patients to get a comprehensive epidemiological investigation. MeltingPlot has a user-friendly web interface (the standalone command line version is also available) and it creates easy to read graphical and tabular outputs. The tool runs in a few seconds even with hundreds of isolates.

Implementation

The flow of MeltingPlot can be divided in three main steps**:** HRM-based clustering/typing of isolates, prevalence analysis and transmission analysis. HRM-based clustering/typing is computed only on the basis of the High Resolution Melting (HRM) temperatures of the isolates amplicons, that are the only inputs needed for this step. Actually, the tool does not use any other information that can be derived from the melting curves, like the shape or the height of the curve. Indeed, in our experience, these features are more subjected to experimental noise than melting temperature so we are not using them to type the isolates. After the computation of the average HRM temperature of the technical replicates, the isolates are organized in a graph where the vertices represent the isolates and two vertices are connected if the difference of their average HRM temperatures is less or equal to 0.5 °C for each PCR primer set used in the HRM typing method. The graph is then decomposed into separate components (groups of connected vertices) and each one is then divided in clusters using the Edge Betweenness Clustering algorithm [7] implemented in the cluster_edge_betweenness function of the igraph R library [8]. Briefly, the betweenness centrality of each edge of the graph was computed as the number of shortest paths that go through the edge, and clusters were identified by gradually removing the edges with the highest betweenness centrality values. Therefore, high betweenness centrality values among two vertices indicates that the two vertices most probably do not belong to the same cluster and vice versa.

Furthermore, the betweenness centrality of a vertex was computed as the number of graph short paths passing that vertex. Hence, vertices with higher betweenness centrality values are those that connect two or more clusters. We used this parameter to identify vertices that were not strongly associated with a single cluster. Thus, vertices with normalized betweenness centrality values above a threshold were not assigned to any cluster and they were classified as "undetermined" by the tool (this threshold of normalized betweenness value can be set by the user, the default is 0.5).

Unfortunately, HRM-based clustering results obtained from different datasets are not directly comparable. To obtain comparable HRM typing results, the user can include in the analysis the HRM temperatures of a collection of reference strains: isolates previously analysed by the same HRM protocol and for which typing annotation is known (e.g. Sequence Type). When a reference collection is provided, MeltingPlot labels each cluster with the annotation of the reference isolates contained in it. For details see the Additional file 1.

Prevalence analysis and transmission analysis steps can be performed only when patients/isolates metadata is provided. In these steps the tool joins the HRM clustering results with the isolates' metadata to create various outputs that depict the spreading of pathogen clones among wards and patients over time. For more details see the output files section below or the Additional file 1. MeltingPlot was developed in R and its dependencies are the libraries igraph [8], gplots [9], xlsx [10], ggplot2 [11], scales [12]. The user interface on the website was developed in PHP.

Input file

Users are required to download and fill an xls template spreadsheet that contains four sheets: *HRM_temperatures*, *Isolates_metadata*, *Reference_isolates* and an *HELP_notes* sheet:

- *HRM_temperatures:* in this sheet the user has to report the high resolution melting (HRM) temperatures of the study isolates. This is the only mandatory data and it is used to perform the HRM-based clustering/typing analysis. If HRM experiments were performed using technical replicates, the users have to report all the replicate temperatures;
- *Isolates_metadata*: in this sheet the users can provide patients/isolates metadata, e.g. isolation date, isolation location (e.g. hospital ward) and an ID for the patients (e.g. Pz1, Pz2, …). This information is not mandatory for HRM isolates typing but it is required to perform the complete epidemiological investigation (i.e. prevalence analysis and transmission analysis);
- *Reference_isolates*: this sheet contains the HRM temperatures of the reference isolates and their annotation (e.g. Sequence Type). The reference isolates annotation will be used to label the clusters, making the obtained HRM typing results portable when the same reference collection is used.
- *HELP_notes*: this sheet contains important information about the rules for each column of the spreadsheet.

(See figure on next page.)

Fig. 1 MeltingPlot Output, example of an epidemiological investigation. Here we report the three most significant MeltingPlot output plots obtained on the simulated dataset. The plots were selected to show the power of the HRM-based epidemiological investigation performed by the tool. Higher resolution images are available in the Additional file 1. **a** Prevalence analysis: the plot shows the number of isolates collected from each hospital ward over time. Each HRM cluster is represented with a different color. This analysis allows the detection of the pathogen clones emergence in the hospital setting. **b** Patients' timeline: each row refers to a patient and the symbols represent isolates. The shape of the symbols report the location where the isolates were collected while the colors indicate the HRM cluster. **c** Patient-to-patient graph: each vertex represents a patient and two vertices are connected if isolates belonging to the same HRM cluster were collected from both patients. Vertices are reported as pie charts and colors show the locations (wards) where the isolates of the patients were collected. The edges of the graph are thicker if the isolates from the same HRM cluster were collected within seven days (this threshold can be defined by the user) from the same location. This plot can help to identify the transmission routes of the pathogen in the hospital setting

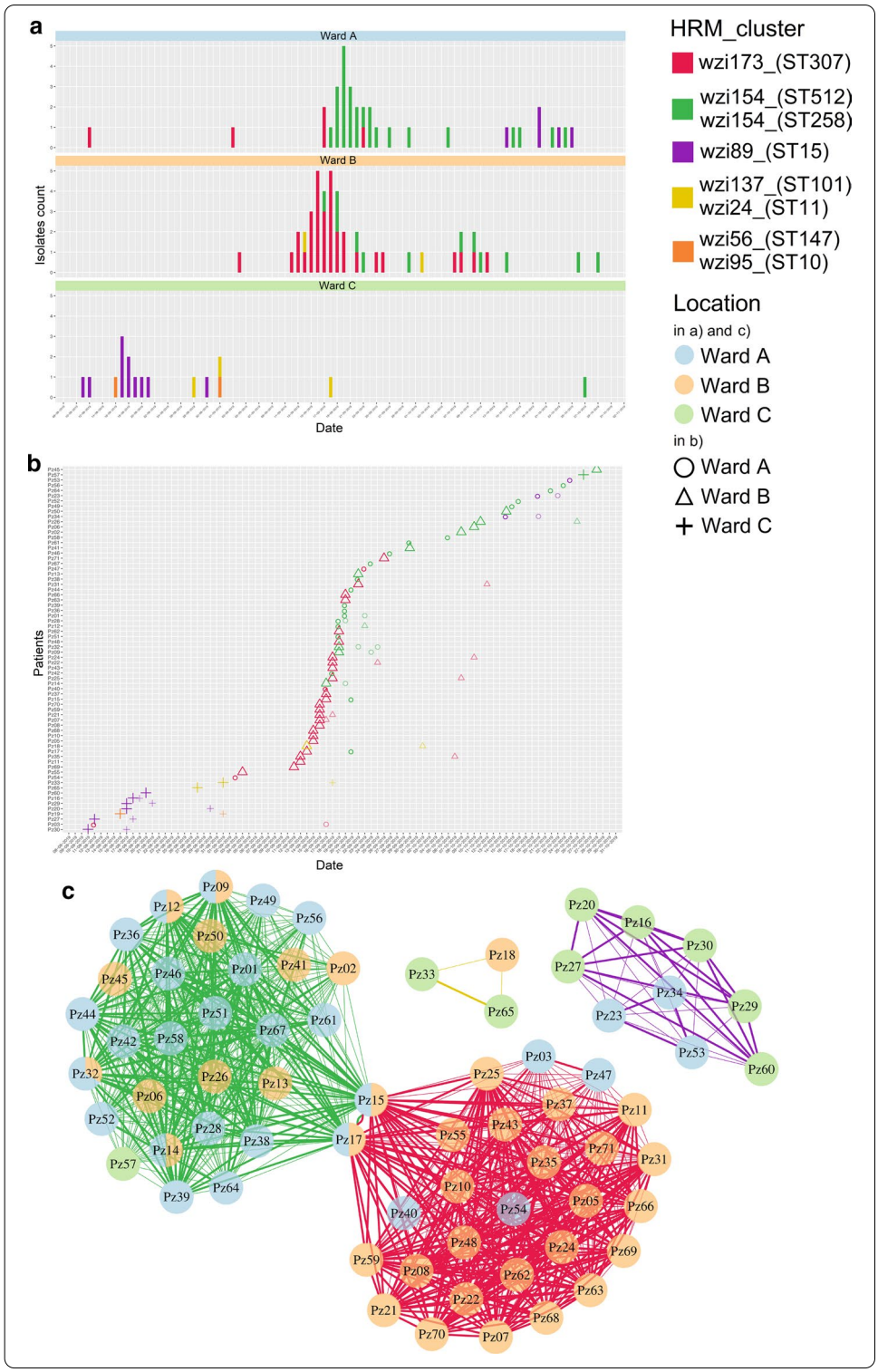
All the templates (the blank template, the templates with reference HRM temperature collections, and the example files) are available on the MeltingPlot webpage.

Output files

MeltingPlot creates three groups of plot files (in PDF and PNG format), one for each step of the analysis: HRM-based clustering/typing, prevalence analysis and transmission analysis. The HRM-based clustering/typing plot group includes the isolates graph (where each isolate is colored on the basis of its cluster) and a heatmap showing the HRM temperatures and the isolates clusters. In the isolates graph each vertex is an isolate and two isolates are connected as described above. The last two groups are created when metadata of the isolates is provided. The prevalence analysis plot group includes bar plots showing the distribution of the clusters over time in the different locations. The transmission analysis plot group contains a patient timeline and a patient-to-patient graph. In the latter, two patients are connected when two isolates belonging to the same HRM cluster were collected from both patients. The edge is thicker when the isolates were collected in the same location (e.g. ward) within a number of days set by the user (7 by default). Thus, thicker edges highlight most probable transmission events. MeltingPlot also produces xls spreadsheets containing the isolates HRM clusters and metadata. See Additional file 1 for details.

Results

To show the capability of MeltingPlot, we simulated a complex large *Klebsiella pneumoniae* nosocomial outbreak (100 outbreak isolates) sustained by multiple clones spread in different wards, a situation comparable to real large nosocomial outbreaks [13]. The isolates were simulated using HRM temperatures retrieved from a dataset of *K. pneumoniae* isolates previously analyzed in our laboratory. We simulated an outbreak scenario sustained by five different clones: three major clones that caused the outbreak and two sporadic clones. Thus we included in the dataset the temperature of previously HRM typed isolates belonging to five highly epidemiologically relevant clusters: *wzi173_(ST307)*, *wzi154_(ST512/ST258)*, *wzi89_(ST15)*, *wzi137_(ST101)/wzi24_(ST11)* and *wzi56_(ST147)/wzi95_(ST10)*; the first three causing the outbreaks and the last two being sporadic. We also simulated the metadata of the isolates to obtain a complex



outbreak with three wards showing different epidemiological scenarios (see Fig. 1 for more detail). We run MeltingPlot using the 100 outbreak isolates dataset and a collection of 18 reference isolates previously typed by HRM and WGS [4] (Reference_isolates

sheet in the input file, see above; the Reference_isolates temperature collection is also available on the tool web site). In our outbreak simulation, the HRM typing of the 100 isolates would be performed after every pathogen isolation during the entire outbreak period (~3 months). The entire epidemiological investigation using this HRM typing protocol would cost ~500 euros while it would cost ~5000€ using Multi Locus Sequence Typing (MLST) and ~10,000€ using Whole Genome Sequencing (WGS). As expected, Melting Plot results showed that the outbreak is sustained by three major isolate clusters (Fig. 1). MeltingPlot labelled these clusters as *wzi173_(ST307)* (in red), *wzi154_(ST512/ST258)* (in green) and *wzi89_(ST15)* (in violet) using the annotation given by the user for the reference isolates. Most of the infections are caused by two pathogen clusters (green and red), each one associated with a single ward: the green one with Ward A and the red one with Ward B. A smaller cluster (in violet) caused an outbreak in Ward C at the beginning of the investigated period. The patient's timeline and the patient-to-patient graph clearly show that two patients (Pz15 and Pz17) were infected by isolates of both the red and green clusters and they also crossed the wards A and B: this highlights two possible pathogen transmission routes among the wards. A complete description of each output file is available in the Additional file 1. MeltingPlot is available online at https://skynet.unimi.it/index.php/tools/meltingplot/. The source code for the stand alone version is available at https://github.com/MatteoPS/MeltingPlot.

Discussion

High Resolution Melting (HRM) is a fast and inexpensive molecular biology technique [2] applicable to pathogen typing and suitable for large scale surveillance programmes as well as for fast outbreak reconstruction [1, 3]. In this work we propose MeltingPlot, a tool that allows to perform epidemiological investigation and transmission analysis using HRM data.

The tool implements an algorithm for the HRM-based clustering that groups isolates on the basis of their melting temperatures. Unfortunately, the HRM-based clusters obtained from different collections of isolates are not directly comparable. To overcome this limitation we made MeltingPlot able to include in the clustering analysis the melting temperatures of a collection of reference isolates. MeltingPlot uses the reference isolates as a guide to label the obtained isolates clusters. On this way, MeltingPlot results obtained from different isolates collections become comparable.

Furthermore, MeltingPlot performs complete epidemiological investigations merging HRM clustering results with isolates/patients metadata. It produces easy-to-read graphical representations and tabular files (see Results section) useful to reconstruct epidemiological scenarios and to identify pathogen transmission routes.

The HRM typing is cost saving and it can be carried out using instruments usually present in hospital microbiology laboratories and by not highly specialized personnel. MeltingPlot also eases the HRM data analysis for epidemiological investigation. In our opinion, the implementation of HRM typing can improve nosocomial surveillance programs with a limited impact on the hospital policies in terms of costs, workload and personnel involved.

MeltingPlot is available in stand-alone and web versions. The web interface makes the tool user-friendly and the user has only to upload the data into an xls template spreadsheet. MeltingPlot analyses hundreds of isolates in a few seconds.

Conclusions

HRM technique allows pathogen typing in a few hours and ~ 5 euros per sample. Despite this, the mathematical/informatics skills required for the analysis and interpretation of HRM results limit the application of HRM typing protocols in hospital real time surveillance. MeltingPlot is a user-friendly tool that facilitates the application of HRM to real time large scale surveillance programs in hospital settings.

Availability and requirements

Project name: MeltingPlot.

Project home page: https://skynet.unimi.it/index.php/tools/meltingplot/

Operating system(s): Platform independent.

Programming language: R, PHP.

Other requirements: Any web browser.

License: GPL.

Any restrictions to use by non-academics: none.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-021-04020-y.

Additional file 1: This file contains a detailed description of the clustering algorithm, of the input file and of each output file created by the tool.

Abbreviations

HRM: High Resolution Melting; PCR: Polymerase Chain Reaction; MLST: Multi-Locus Sequence Typing; WGS: Whole Genome Sequencing.

Acknowledgements

Thanks to the Romeo ed Enrica Invernizzi Foundation.

Authors' contributions

MP wrote the code, implemented the web interface and drafted the paper; GBB wrote the code; DDC implemented the web interface; ARP, AP, SP, GVZ wrote the paper; FC conceived the tool, wrote the code and wrote the paper. All authors read and approved the final manuscript.

Funding

None.

Availability of data and materials

The tool described here, the template input file and the example files are freely available at https://skynet.unimi.it/index.php/tools/meltingplot/. The source code is available in the GitHub repository https://github.com/MatteoPS/MeltingPlot.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Department of Biomedical and Clinical Sciences "L. Sacco", Pediatric Clinical Research Center "Romeo and Enrica Invernizzi", Università Di Milano, 20157 Milan, Italy. [2] Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Pavia 27100, Italia. [3] Department of Pediatrics, Children's Hospital Vittore Buzzi, Università Di Milano, Milan, Italy.

Received: 11 September 2020 Accepted: 11 February 2021
Published online: 18 February 2021

References

1. Ruskova L, Raclavsky V. The potential of High Resolution Melting Analysis (HRMA) to streamline, facilitate and enrich routine diagnostics in medical microbiology. Biomed Pap. 2011;155:239–52. https://doi.org/10.5507/bp.2011.045.
2. Tamburro M, Ripabelli G. High Resolution Melting as a rapid, reliable, accurate and cost-effective emerging tool for genotyping pathogenic bacteria and enhancing molecular epidemiological surveillance: a comprehensive review of the literature. Ann Ig. 2017;29:293–316. https://doi.org/10.7416/ai.2017.2153.
3. Mongelli G, Bongiorno D, Agosta M, Benvenuto S, Stefani S, Campanile F. High resolution melting-typing (HRMT) of methicillin-resistant *Staphylococcus aureus* (MRSA): the new frontier to replace multi-locus sequence typing (MLST) for epidemiological surveillance studies. J Microbiol Methods. 2015;117:136–8. https://doi.org/10.1016/j.mimet.2015.08.001.
4. Perini M, Piazza A, Panelli S, Di Carlo D, Corbella M, Gona F, et al. EasyPrimer: user-friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on wzi gene for fast Klebsiella pneumoniae typing. Sci Rep. 2020;10:1307. https://doi.org/10.1038/s41598-020-57742-z.
5. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, et al. Epidemic of carbapenem-resistant Klebsiellapneumoniae in Europe is driven by nosocomial spread. Nat Microbiol. 2019;4:1919–29. https://doi.org/10.1038/s41564-019-0492-8.
6. Pasala AR, Perini M, Piazza A, Panelli S, Di Carlo D, Loretelli C, et al. Repeatability and reproducibility of the wzi high resolution melting-based clustering analysis for Klebsiella pneumoniae typing. bioRxiv. 2020. https://doi.org/10.1101/2020.06.19.161703.
7. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E. 2004. https://doi.org/10.1103/physreve.69.026113.
8. Csardi G, Nepusz T, et al. The igraph software package for complex network research. Int J Complex Syst. 2006;1695:1–9.
9. Warnes GR, Bolker B, Bonebakker L, Gentleman R. gplots: various R programming tools for plotting data. 2009. https://cran.r-project.org/package=gplots.
10. Dragulescu A, Arendt C. CRAN.R-project.org. https://CRAN.R-project.org/package=xlsx. 2020. xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files. R package version 0.6.3. Accessed 23 Jul 2020.
11. Wilkinson L. ggplot2: Elegant Graphics for Data Analysis by WICKHAM. H Biometrics. 2011;67:678–9. https://doi.org/10.1111/j.1541-0420.2011.01616.x.
12. Wickham H, Seidel D. CRAN.R-project.org. https://CRAN.R-project.org/package=scales. 2020. scales: Scale Functions for Visualization. R package version 1.1.1. Accessed 23 Jul 2020.
13. Ferrari C, Corbella M, Gaiarsa S, Comandatore F, Scaltriti E, Bandi C, et al. Multiple KPC clones contribute to an extended hospital outbreak. Front Microbiol. 2019;10:2767. https://doi.org/10.3389/fmicb.2019.02767.

Publisher's Note

Section C

Appendix 9

Hypervariable-Locus Melting Typing (HLMT): a novel, fast and inexpensive sequencing-free approach to pathogen typing based on High Resolution Melting (HRM) analysis

Perini M, Piazza A, Panelli S, Papaleo S, Alvaro S,Vailati F,Corbella M, Saluzzo F, Gona F, Castelli F, Farina C, Marone P, Cirillo DM, Cavallero A, Zuccotti GV, Comandatore F.

bioRxiv. 2021 Nov 08. doi: 10.1101/2021.07.01.450706

Hypervariable-Locus Melting Typing (HLMT): a novel, fast and inexpensive sequencing-free approach to pathogen typing based on High Resolution Melting (HRM) analysis

Matteo Perini 1; Aurora Piazza 2; Simona Panelli 1; Stella Papaleo 1; Alessandro Alvaro 1; Francesca Vailati 3; Marta Corbella 4; Francesca Saluzzo 5; Floriana Gona 6; Daniele Castelli 7, Claudio Farina 3; Piero Marone 4; Daniela Maria Cirillo 5; Annalisa Cavallero 7; Gian Vincenzo Zuccotti 1,8; Francesco Comandatore* 1.

1. Department of Biomedical and Clinical Sciences “L. Sacco”, Pediatric Clinical Research Center “Romeo and Enrica Invernizzi”, Università Di Milano, 20157 - Milan (Italy).
2. Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, 27100 - Pavia (Italy).
3. Microbiology Institute, A.S.S.T. “Papa Giovanni XXIII”, 24127 - Bergamo (Italy).
4. Microbiology and Virology Unit, Fondazione IRCCS Policlinico San Matteo, 27100 - Pavia (Italy).
5. Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, 20132 - Milan (Italy).
6. Laboratory Microbiology and Virology -Ospedale San Raffaele Dibit, 2-San Gabriele 1, 20132 - Milan (Italy).
7. Laboratory of Microbiology, ASST Monza, San Gerardo Hospital, 20900 - Monza (Italy).
8. Department of Pediatrics, Children’s Hospital Vittore Buzzi, Università Di Milano, 20154 - Milan (Italy).

*corresponding author
francesco.comandatore@unimi.it
Via Giovanni Battista Grassi, 74, 20157 Milano
tel. +39 0250319826

Abstract

Objectives

Subspecies pathogen typing is a pivotal tool to detect the emergence of high-risk clones in hospital settings and to limit their spreading among patients. Unfortunately, the most used subspecies typing methods (i.e. Pulsed-field Gel Electrophoresis - PFGE, Multi-Locus Sequence Typing - MLST and Whole Genome Sequencing - WGS) are too expensive and time consuming to be suitable for real-time surveillance. Here we present Hypervariable-Locus Melting Typing (HLMT), a novel subspecies typing approach based on High Resolution Melting (HRM) analysis, which allows pathogen typing in a few hours and with ~5 euros per sample.

Methods

HLMT types the strains by clustering them using melting temperatures (HLMT-clustering) and/or by assigning them to Melting Types (MTs) on the basis of a reference dataset (HLMT-assignment). We applied HLMT (clustering and typing) to 134 *Klebsiella pneumoniae* strains collected during outbreaks or surveillance programs in four hospitals. Then, we compared HLMT typing results to PFGE, MLST and WGS.

Results

HLMT-clustering distinguishes most of the *K. pneumoniae* high-risk clones with a sensitivity comparable to PFGE and MLST. It also drawed surveillance epidemiological curves comparable to those obtained by MLST, PFGE and WGS typing. Furthermore, the results obtained by HLMT-assignment were coherent to MLST for 96% of the typed strains with a Jaccard index of 0.912.

Conclusions

HLMT is a fast and scalable method for pathogen typing, suitable for real-time hospital microbiological surveillance. HLMT is also inexpensive and thus it is applicable to infection control programs in low-middle income countries.

Keywords:

Microbiological surveillance,

High Resolution Melting,
Outbreak reconstruction,
Low-middle income countries,
Real-time surveillance

Introduction

Healthcare-associated infections (HAIs) are a major burden for global public health [1]. The microbiological surveillance programs are pivotal to establish effective infection control strategies. In particular, subspecies typing is fundamental to detect the emergence of high-risk clones. The most used methods for subspecies bacterial typing are Pulsed Field Gel Electrophoresis (PFGE), Multi-Locus Sequence Typing (MLST) and Whole Genome Sequencing (WGS). All these methods require several hours (up to days) to be performed and this limits their application in nosocomial real-time surveillance programs.

High Resolution Melting (HRM) assay has been proposed as a suitable method for fast bacterial typing [2,3]. This technique measures the melting temperatures of qPCR amplicons, which depends on the GC content. HRM can even distinguish amplicons diverging for just one Single Nucleotide Polymorphism (SNP) and it is widely used to detect human allele variants [4]. HRM protocols designed on hypervariable genes are able to discriminate among bacterial clones within the same species [2,3], because their amplicons will melt at different temperatures (e.g. they differ in GC content). HRM is particularly promising for microbiological surveillance: it is fast (~5 hours to complete the analysis), discriminatory, inexpensive (~5 euros per sample) and it can be performed on the most common qPCR platforms [5].

Despite the numerous HRM protocols proposed so far for bacterial typing [2], the method is rarely applied in hospital settings for microbiological surveillance. Indeed, most protocols have been designed to distinguish only among the few clones used for protocol development, including only a fraction of the entire genetic variability of the pathogen. Moreover, only a few algorithms and software are available to analyse HRM data for epidemiological purposes.

Recently, we developed a novel approach for HRM-based subspecies typing. We focused on hypervariable genes and implemented a tool (i.e. EasyPrimer) to facilitate HRM primers designing in this difficult context [3]. Additionally, we developed an algorithm for pathogen typing using HRM data [6]. We used this approach to develop an

HRM-based typing protocol for *Klebsiella pneumoniae* [3] and then we validated the repeatability and portability of the method [5].

In this study we provide a comprehensive description of this novel approach that we named Hypervariable-Locus Melting Typing (HLMT). We applied HLMT in four nosocomial epidemiological investigations on *Klebsiella pneumoniae*, comparing its typing efficiency to PFGE, MLST and WGS.

Methods

Ethics Statement

This study uses bacterial isolates from human samples that were obtained as part of hospital routines. No extra human samples were obtained for this research. Therefore, informed consent (either written or verbal) was not required.

Isolates datasets

Four Italian hospitals were included in the study: “San Gerardo” Hospital in Monza (from here HSG), “IRCCS Fondazione Policlinico San Matteo” Hospital in Pavia (PSM), “ASST Papa Giovanni XXIII” Hospital in Bergamo (PG23) and “IRCCS San Raffaele” Hospital in Milan (OSR). The hospitals provided a total of 134 *Klebsiella pneumoniae* strains: i) HSG provided 10 strains isolated during an outbreak that involved 10 patients in the oncohematology ward between August 14, 2018 and September 24, 2018 (HSG dataset); ii) PSM provided 24 strains isolated during an outbreak already investigated with Whole Genome Sequencing (WGS) and described by Ferrari and colleagues [7] (PSM dataset) (the original dataset consisted of 32 strains, but only 24 were successfully revitalized in this work); iii) PG23 hospital provided 20 strains isolated during an outbreak that involved 16 patients and nine hospital wards, between May 07, 2019 and November 04, 2019 (PG23 dataset); iv) OSR hospital provided all the 80 strains isolated during a one-year-long WGS surveillance in 2017 and already typed with WGS and PFGE by Gona and colleagues [8] (OSR dataset).

DNA extraction

For each of the 134 strains, the bacterial culture was subjected to two consecutive single colony selections on MacConkey agar, incubated overnight at 37 °C (Becton Dickinson, Franklin Lakes, NJ, USA). A single bacterial colony was then suspended in liquid medium, incubated overnight and the DNA was extracted using the DNeasy blood and tissue kit, following the manufacturer's instructions (Qiagen, Hilden, Germany).

Whole Genome Sequencing

The 104 *K. pneumoniae* strains isolated from PSM (n=24) and OSR (n=80) have been already subjected to WGS in previous studies [7,8]. The remaining 30 isolates (HSG and PG23 datasets) were subjected to WGS on the Illumina MiSeq platform, (Illumina, San Diego, CA, USA), after Nextera XT 2×250 bp paired-end library preparation. The reads were quality-checked using FastQC and trimmed using Trimmomatic software [9]. SPAdes[10] was then used to assembly the pair-end reads.

WGS-based typing

The 15,699 public genome assemblies of *K. pneumoniae* present in the PATRIC database on February 8, 2021 for which the publication code was available (in accordance with Fort Lauderdale and Toronto agreements) were retrieved. Each of the four datasets was separately subjected to core SNPs calling as follows: i) each genome was compared to the retrieved PATRIC dataset using Mash [11] and the 50 most similar strains were included in the background dataset; ii) these selected PATRIC genomes were merged to the dataset genomes; iii) core SNP calling was performed on the merged genome dataset using Purple tool [8]. For each of the four datasets, the obtained core SNPs alignment was subjected to Maximum Likelihood (ML) phylogenetic analysis with 100 bootstraps using the software RAxML8 [12], after best model selection using ModelTest-NG [13] (GTR+G for HSG and PG23; TVM+G for PSM and OSR). For each of the four datasets, clusters were identified on the resulting trees as the largest monophyla of dataset strains (not from PATRIC) with a bootstrap support >= 75.

Multi-Locus Sequence Typing and *wzi* alleles

K. pneumoniae clones are often defined by combining the MLST profile and the *wzi* gene allele [14]. Multi-Locus Sequence Typing (MLST) profiles and the *wzi* alleles of the 134 genome assemblies were determined using Kleborate [15].

Pulsed Field Gel Electrophoresis Clustering

Pulsed Field Gel Electrophoresis (PFGE) clusters were described by Gona and colleagues [8] on the 80 strains of the OSR dataset following digestion with XbaI enzyme and separation into a CHEF-DRIII electrophoretic system (BioRad, Hercules, California).

Hypervariable-Locus Melting Typing

HLMT includes two different typing strategies: HLMT-clustering and HLMT-assignment. HLMT-clustering groups the strains on the basis of melting temperatures without a reference dataset. HLMT-assignment classifies each strain into a Melting Type (MT) by comparing the melting temperatures of the strains to a reference dataset. The reference dataset used in this work was reconstructed using the 43 *K. pneumoniae* strains previously typed by HRM by Pasala and colleagues [5]: HLMT-clustering grouped the strains in seven clusters that we used to define seven Melting Types (MTs). For each MT, the reference melting temperatures were computed as the mid-range melting temperatures (the arithmetic mean of the highest and the lowest temperature) of the strains in that cluster. The MTs were then labelled using the names of the most relevant lineages of the strains that they contain. The reference dataset is available at https://skynet.unimi.it/wp-content/uploads/MeltingPlot/TemplateHLMT_ref_KPN_03-15-2021.xls. All the *K. pneumoniae* strains included in this work were subjected to High Resolution Melting (HRM) assays using the protocol described in Perini et al., 2020 [3]. For each of the four datasets, the obtained melting temperatures and the above-mentioned reference dataset were used to perform HLMT-clustering and HLMT-assignment analyses with MeltingPlot v2.0 tool [6] (available online at https://skynet.unimi.it/index.php/tools/MeltingPlot/).

Comparison of the typing results

HLMT results were compared to PFGE, MLST+*wzi* and WGS by means of heatmaps and correlation plots, produced using the R libraries gplots [16] and corrplot [17], respectively. Furthermore, the HLMT-assignment results were compared to MLST+*wzi* by Jaccard similarity index, computed using the R package clustelval. For each dataset,

HLMT-clustering, MLST+*wzi* and WGS typing results were combined to the collection dates of the strains to obtain the epidemiological curves showing the prevalence of the clusters and groups over time [6]. Additionally, PFGE typing results were retrieved from Gona et al. 2020 [8] and used to obtain the relative epidemiological curves.

Data availability

Genome assemblies of the 30 strains sequenced in this work are available from the NCBI BioProject repository under project PRJEB44864 (ERP128959).

Results

Whole Genome Sequencing

The assembly statistics and the accession numbers of the 30 sequenced *K. pneumoniae* strains (10 of the HSG dataset and 20 of the PG23 dataset) are reported in the Supplementary Table S1.

Typing methods results

Strains from the four datasets (HSG, PSM, PG23 and OSR) were typed by HLMT (clustering and assignment), MLST+*wzi* and WGS. Furthermore, the PFGE typing results of the 80 OSR dataset strains were retrieved from Gona and colleagues 2020 [8]. The results of HLMT, MLST+*wzi*, WGS and PFGE typing are reported in the Supplementary Table S1 and the core SNP-based phylogenetic trees used for WGS typing are shown in the Supplementary Figure S1.

Comparison of the typing methods

For each dataset, the HLMT-clustering results were compared to MLST+*wzi* and WGS: the graphical representation of the relative contingency tables are reported in Supplementary Figure S2 and Supplementary Figure S3, respectively. HLMT-clustering results were compared to PFGE for the OSR dataset only (see Supplementary Figure S4). The HLMT-assignment algorithm classifies the strains into Melting Types. This analysis was able to classify 120 out of the 134 strains (~90%) included in this study. Seven of these 120 strains (~6%) belonged to MLST+*wzi* profiles not included in the HLMT reference strain dataset used for the analyses (see Methods). This made it impossible to assess, for these seven strains, if the HLMT-assignment and MLST+*wzi* results were coherent. For 108 out of the remaining 113 strains (96%) the HLMT-assignment and MLST+*wzi* were coherent, with a Jaccard similarity index of 0.912. The HLMT-assignment results are reported in the Supplementary Table S1 and the correlation matrix plot of the HLMT-assignment vs MLST+*wzi* is shown in Figure 1.

For each dataset, the epidemiological curves obtained combining typing information (HLMT-clustering, MLST+*wzi*, WGS and PFGE) and isolation dates of the strains are shown in Figure 2.

Discussion

Hypervariable-Locus Melting Typing (HLMT) is an innovative approach to High Resolution Melting (HRM)-based typing: it makes it easier to design highly discriminatory HRM protocols and to perform reliable, robust, repeatable and portable HRM-based typing analyses. In this work we show the application of HLMT to the typing of *Klebsiella pneumoniae* in four real hospital scenarios, comparing the results with those obtained by more established approaches, as Whole Genome Sequencing (WGS), Multi-Locus Sequence Typing (MLST) and Pulsed-Field Gel Electrophoresis (PFGE). As summarised in Figure 3, the workflow of HLMT consists of three main parts: i) HLMT protocol design on hypervariable genes; ii) HRM experiments; iii) HRM data analysis for HLMT-clustering and HLMT-assignment.

HRM primer design

HRM is less sensitive than sequencing to discriminate among DNA sequences. Thus, to obtain a stronger signal, the HRM protocol should include more than one primer pair and it should be designed on hypervariable genes. Primer design on hypervariable genes is challenging because it requires the analysis of up to thousands of different gene alleles to identify conserved regions suitable for primer design. We already developed the EasyPrimer tool to automatically identify these gene regions [3].

HRM experiments

HLMT analysis can be performed using HRM data obtained from any HRM-capable qPCR platform. As shown by Pasala et al. [5], HLMT-clustering is highly repeatable when the experiments are performed on the same model of instrument.

HLMT-clustering and HLMT-assignment

HLMT analysis includes two different strain typing methods: HLMT-clustering and HLMT-assignment. HLMT-clustering groups the strains on the basis of their melting temperatures using a graph-based clustering algorithm [6]. Strains with similar melting temperatures for all the primers included in the HLMT protocol are grouped together.

This clustering approach recalls the PFGE clustering, where a hierarchical clustering algorithm groups the strains on the basis of their restriction patterns.

On the other hand, in HLMT-assignment each strain is assigned to a Melting Type (MT) by comparing each strain to a reference dataset. The dataset consists of melting temperatures of previously typed strains selected to represent the genetic variability of the pathogen (see methods).

The strength of HLMT-clustering over HLMT-assignment is that the clustering works even on strains belonging to lineages absent in the HLMT strains dataset. On the other hand, the HLMT-assignment results are portable and they can be shared among laboratories and/or they can be compared to previous HLMT experiments.

The use of multiple primer pairs makes HLMT-clustering and HLMT-assignment more powerful but it also makes the analyses trickier. To tackle this issue, we have already developed MeltingPlot v2.0 tool [6] that automatically performs HLMT-clustering and HLMT-assignment analyses using HRM data.

HLMT protocol for K. pneumoniae

We already designed an HLMT protocol for *K. pneumoniae* typing [3]. In this work, we show the applicability of the method for hospital surveillance programs. More in detail, we followed the three steps of the workflow described in Figure 3: i) HRM protocol design has been already carried out by Perini et al. 2020 [3], selecting the hypervariable capsular gene *wzi* as target and designing two primer sets (called *wzi*-3 and *wzi*-4) analysing the hundreds *wzi* allele sequences available, using EasyPrimer tool; ii) HRM experiments were performed in this work; iii) HLMT-clustering and HLMT-assignment analyses were also performed in this work using the HLMT reference strains dataset (see methods) and the melting temperatures of the 134 strains of the four datasets, using MeltingPlot v2.0 tool.

We reconstructed the *K. pneumoniae* reference dataset naming the Melting Types as the most epidemiologically relevant clone(s) present in the MT, recalling the labelling approach used in MLST for Clonal Clusters (e.g. *K. pneumoniae* CC258).

HLMT-assignment analysis classified almost every strain in concordance with MLST+*wzi* (Figure 1), showing the portability and repeatability power of the method. As shown in Figure 2, the HLMT-clustering epidemiological scenarios were highly consistent with those from the other typing methods: the curves are very similar for HSG, PG23 and OSR datasets and HLMT-clustering was also able to detect the emergence of the outbreak in PSM.
Furthermore, despite HLMT-clustering is less sensitive than MLST+*wzi*, the method was able to discriminate the most epidemiologically relevant STs (i.e. ST258 *wzi*29, ST258 *wzi*154, ST307 *wzi*173 and ST11) and the only not distinguished group of high-risk clones were ST101-ST11-ST15 and ST307-ST147 (Supplementary Figure S2). Here we also want to highlight that HLMT was able to discriminate between the two major sub-clones of the pandemic ST258, the Clade1 (harbouring the *wzi*29) and Clade2 (*wzi*154). These two clones, harbouring the same MLST gene alleles, are not distinguishable by MLST typing without further sequencing the gene *wzi*.
HLMT is less discriminatory than WGS but it is fast, easy, inexpensive and repeatable. The lower sensitivity of HRM in comparison to sequencing, makes HLMT unable to discriminate clones harbouring target genes with similar melting temperatures (e.g. ST101, ST11 and ST15 that harbour *wzi* allele with similar melting temperatures). Anyway, as shown in this work, this low sensitivity does not affect the repeatability of the method: indeed it fails to distinguish only some specific clones. On the other hand, this low sensitivity drastically reduces the possibility of obtaining false negative results. Therefore, when two strains are assigned to two distinct Melting Types (MT) or two Melting Clusters we can safely exclude their genetic relatedness. Having this information quickly during the early stage of a nosocomial outbreak can be particularly useful to establish an effective infection control strategy.

Applicability

The only two instruments required to perform HLMT are a HRM-capable qPCR platform and a standard Personal Computer (PC); the molecular biology skills required are also minimal. No bioinformatic skills are necessary because both HLMT-clustering and

HLMT-assignment can be automatically performed online, using the free and user-friendly tool MeltingPlot [6].

When the metadata of the strains is available, MeltingPlot can be used to merge HLMT clusters and strains metadata (isolation date and location) to produce graphical descriptions of the epidemiological scenario under study. Lastly, HLMT allows to type large amounts of isolates in a few hours for a few euros per sample (Figure 4).

WGS is the most sensitive method for bacterial typing and, even if it is becoming the gold standard typing approach for several pathogen species, PFGE and MLST are applied more often in real-time hospital surveillance programs [18]. In this work we show that HLMT is four times faster than MLST and 33 times faster than PFGE, maintaining a sensitivity comparable to these typing methods. Moreover, the low cost of HLMT (~5 euros/isolate) makes it one of the most inexpensive sub-species pathogen typing methods available, being up to 20 times less expensive than MLST, PFGE and WGS (Figure 4). This makes HLMT suitable for epidemiological investigations in real-time and for the development of low cost surveillance programs even in low and middle income countries.

Acknowledgements

Thanks to the Romeo ed Enrica Invernizzi Foundation.

Funding

None.

References

1. Suetens C, Latour K, Kärki T, Ricchizzi E, Kinross P, Moro ML, et al. Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017. Euro Surveill. 2018;23. doi:10.2807/1560-7917.ES.2018.23.46.1800516

2. Tamburro M, Ripabelli G. High Resolution Melting as a rapid, reliable, accurate and cost-effective emerging tool for genotyping pathogenic bacteria and enhancing molecular epidemiological surveillance: a comprehensive review of the literature. Ann Ig. 2017;29: 293–316. doi:10.7416/ai.2017.2153

3. Perini M, Piazza A, Panelli S, Di Carlo D, Corbella M, Gona F, et al. EasyPrimer: user-friendly tool for pan-PCR/HRM primers design. Development of an HRM protocol on wzi gene for fast Klebsiella pneumoniae typing. Sci Rep. 2020;10: 1307. doi:10.1038/s41598-020-57742-z

4. Vossen RHAM, Aten E, Roos A, den Dunnen JT. High-resolution melting analysis (HRMA): more than just sequence variant screening. Hum Mutat. 2009;30: 860–866. doi:10.1002/humu.21019

5. Pasala AR, Perini M, Piazza A, Panelli S, Di Carlo D, Loretelli C, et al. Repeatability and reproducibility of the wzi high resolution melting-based clustering analysis for Klebsiella pneumoniae typing. AMB Express. 2020;10: 217. doi:10.1186/s13568-020-01164-7

6. Perini M, Biffignandi GB, Di Carlo D, Pasala AR, Piazza A, Panelli S, et al. MeltingPlot, a user-friendly online tool for epidemiological investigation using High Resolution Melting data. BMC Bioinformatics. 2021;22. doi:10.1186/s12859-021-04020-y

7. Ferrari C, Corbella M, Gaiarsa S, Comandatore F, Scaltriti E, Bandi C, et al. Multiple KPC Clones Contribute to an Extended Hospital Outbreak. Front Microbiol. 2019;10: 2767. doi:10.3389/fmicb.2019.02767

8. Gona F, Comandatore F, Battaglia S, Piazza A, Trovato A, Lorenzin G, et al. Comparison of core-genome MLST, coreSNP and PFGE methods for Klebsiella pneumoniae cluster analysis. Microbial Genomics. 2020. doi:10.1099/mgen.0.000347

9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170

10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19: 455–477. doi:10.1089/cmb.2012.0021

11. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17: 132. doi:10.1186/s13059-016-0997-x

12. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033

13. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol. 2020;37: 291–294. doi:10.1093/molbev/msz189

14. Brisse S, Passet V, Haugaard AB, Babosan A, Kassis-Chikhani N, Struve C, et al. wzi Gene sequencing, a rapid method for determination of capsular type for Klebsiella strains. J Clin Microbiol. 2013;51: 4073–4078. doi:10.1128/JCM.01924-13

15. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. Genomic surveillance framework and global population structure for Klebsiella pneumoniae. bioRxiv. 2021. p. 2020.12.14.422303. doi:10.1101/2020.12.14.422303

16. Warnes GR, Bolker B, Bonebakker L, Gentleman R. gplots: Various R programming tools for plotting data. R package version. 2009. Available: https://scholar.google.ca/scholar?cluster=16259089987736248658,2882899000591514106,10487119221144378713,14931491367744286770,15403587050690174877,2678983030000690315,7728989541145325132,6493232030609196922,15135737191892172790,3419986939540951798,14348715541294975909,2097443665141942139,16943805789325834065,17912403305300623252,5163193380339229439,519532344241888888,11350738125292490332,5524558152475975507,6737430022897577572,10553771455524588631,10567053186574042785&hl=en&as_sdt=0,5&sciodt=0,5

17. Wei T, Simko V. R package “corrplot”: Visualization of a Correlation Matrix. (Version 0.88). In: github [Internet]. [cited 21 May 2021]. Available: https://github.com/taiyun/corrplot

18. ECDC. ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations. In: www.ecdc.europa.eu [Internet]. 4 Apr 2019 [cited 15 Jun 2021]. Available: https://www.ecdc.europa.eu/en/publications-data/ecdc-strategic-framework-integration-molecular-and-genomic-typing-european

Figures

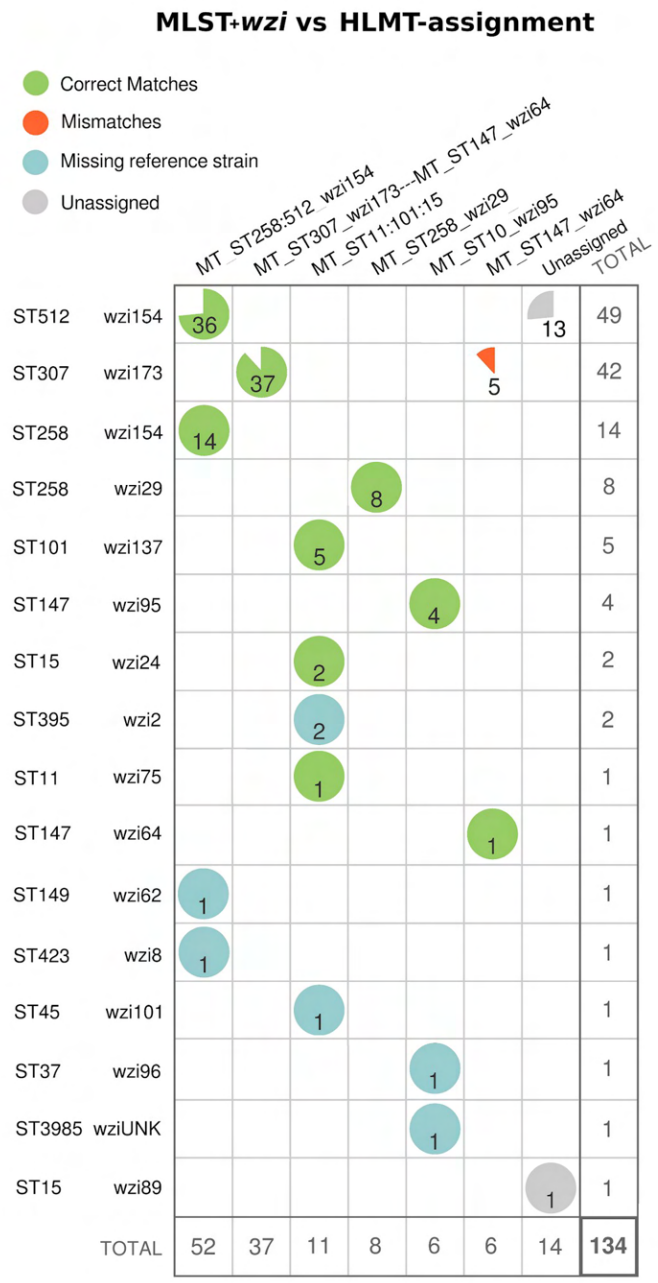


Figure1: MLST+*wzi* vs HLMT-assignment correlation matrix

Correlation matrix between Multi-Locus Sequence Typing (MLST)+*wzi* typing and Hypervariable-Locus Melting Typing (HLMT)-assignment on the 134 strains analysed in

the study. The pie charts indicate, for each MLST+*wzi* profile, the proportion of matches with the Melting Types (MT). Green pies show the correct matches and the red ones the mismatches. Light blue pies show the strains belonging to MLST+*wzi* profiles absent in the reference dataset used to perform HLMT-assignment analysis. Gray pies show the strains that were classified as “Unassigned” by HLMT-assignment analysis.

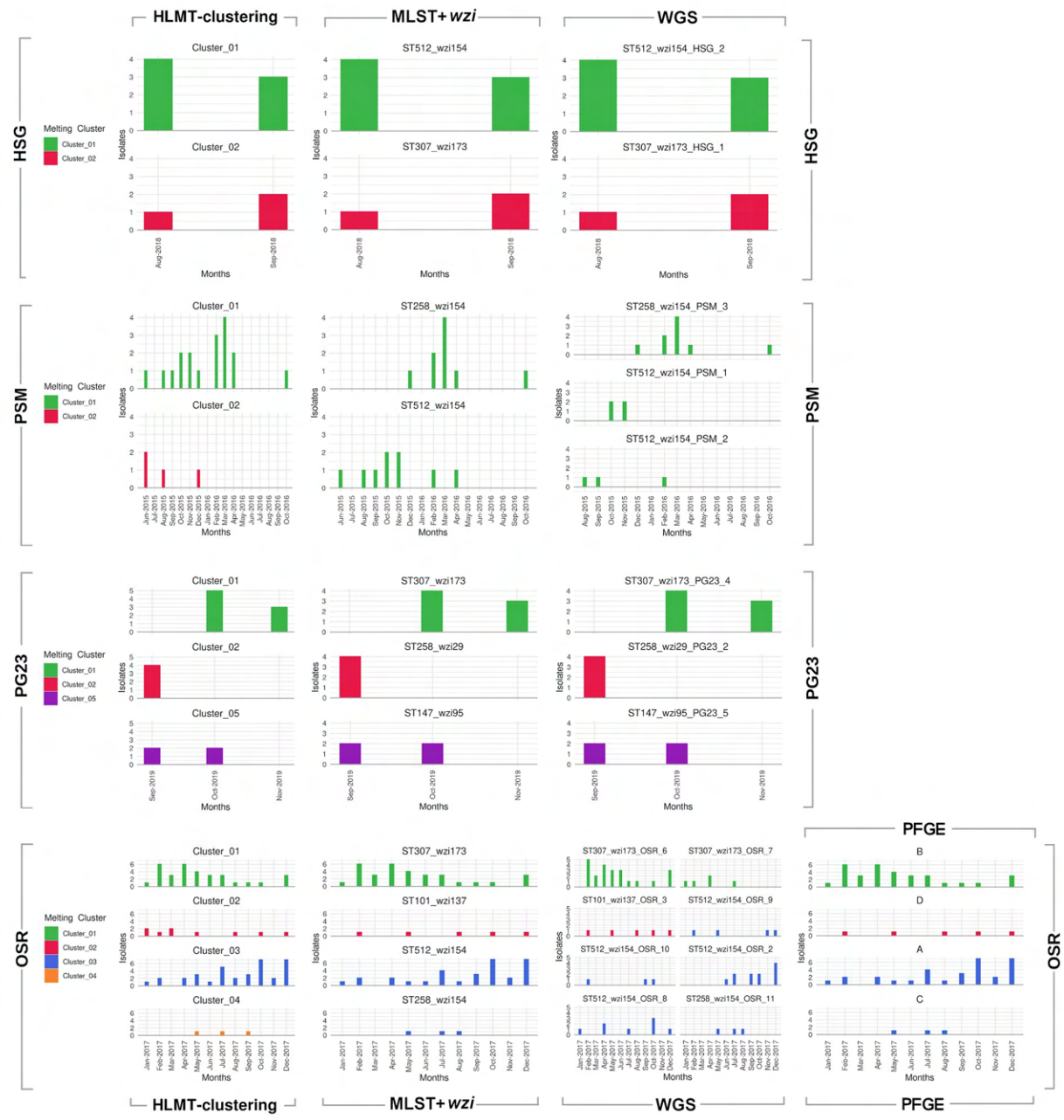


Figure2: Epidemiological curves comparison

Epidemiological curves reconstructed using isolation dates and typing information obtained by HLMT-clustering, MLST+*wzi*, WGS and PFGE on the four dataset analysed in this study. Each line of the plot refers to a dataset while each column to a typing method. The colors used in the barplots correspond to the HLMT clusters.

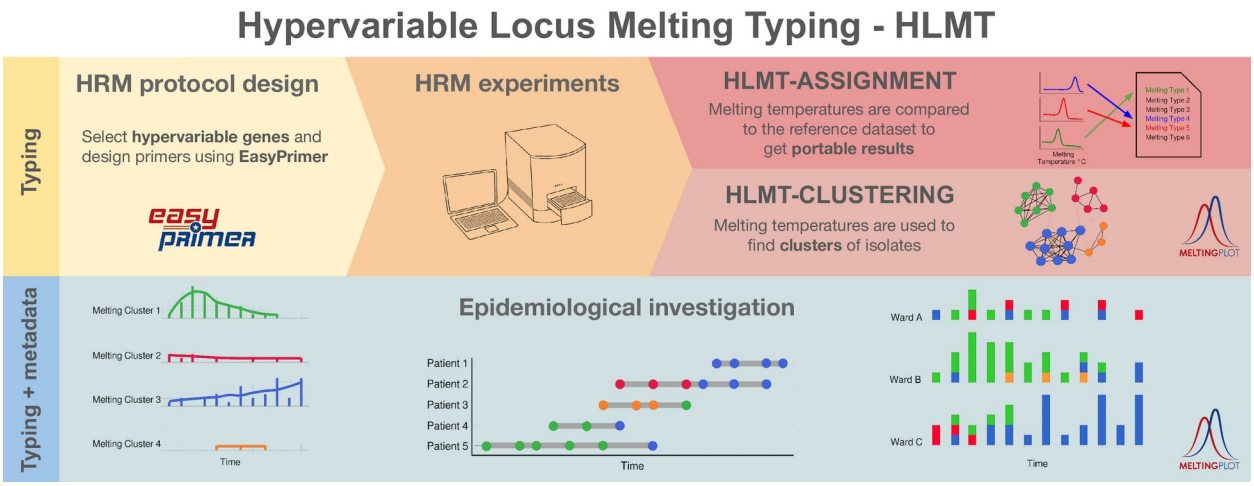
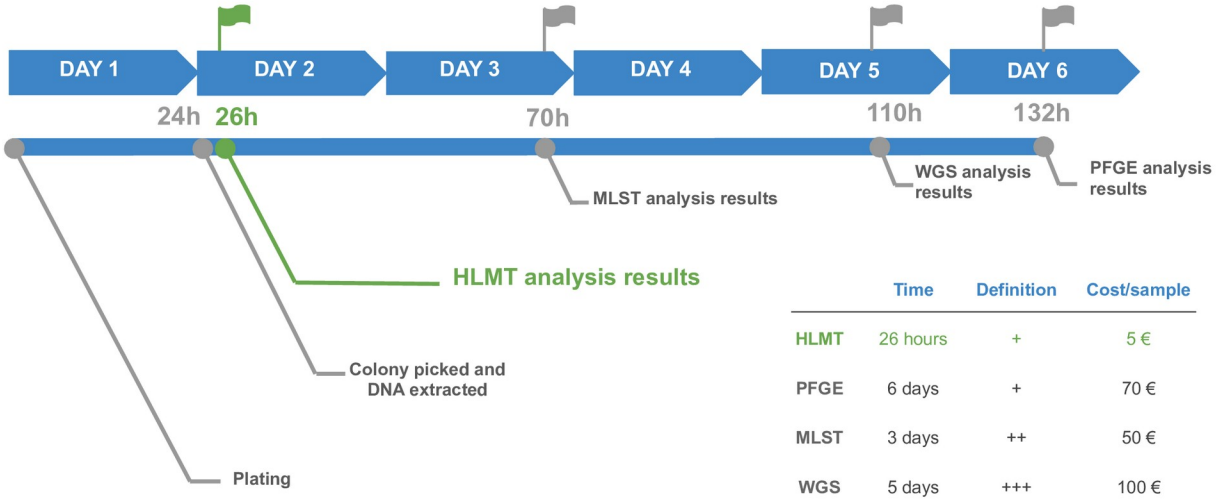


Figure3: Hypervariable-Locus Melting Typing flow

Graphical representation of the flow of Hypervariable-Locus Melting Typing (HLMT). On the top, the three main steps of the HLMT method: HRM protocol design on hypervariable genes, HRM experiments and HRM data analysis for typing (HLMT-clustering and HLMT-assignment). On the bottom, the combination of HLMT typing and isolates metadata allow to perform epidemiological investigations, e.g. epidemiological curves and patient timeline. Primer design, HLMT-clustering, HLMT-assignment and epidemiological curve production can be performed using on-line free user-friendly tools (namely: EasyPrimer for primer design and MeltingPlot v2.0 for the others).



	Time	Definition	Cost/sample
HLMT	26 hours	+	5 €
PFGE	6 days	+	70 €
MLST	3 days	++	50 €
WGS	5 days	+++	100 €

Figure4: Timeline and costs of the typing methods

The timeline summarises the time required and the prices of Hypervariable-Locus Melting Typing (HLMT) (in green), Hypervariable-Locus Melting Typing (MLST), Whole Genome Sequencing (WGS) and Pulsed-Field Gel Electrophoresis (PFGE). The time required to perform each typing analysis, the relative typing definition and cost per sample are also reported.

Acknowledgements

I would like to thank the coordinator of the Doctoral Program in Nutritional Sciences of the University of Milan, Prof. Luciano Pinotti, for the opportunity of joining the program.

I would also like to acknowledge the Romeo ed Enrica Invernizzi Foundation and the Pediatric Clinical Research Center for giving me the possibility to do research in the best conditions.

Thanks to my tutor Prof. Gian Vincenzo Zuccotti and to Prof. Claudio Bandi for the advice and for supporting me in the time I spent in Milano.

My biggest thanks go to the past and present members of the Sky Net lab, in particular to Komma (a.k.a. Francesco Comandatore): the Principal Investigator of the group and the one who taught me how to think as a researcher. In this lab I really felt part of a pleasant and prolific group of researchers that develop and implement ideas at the cutting edge of science. Here I grew not only as a researcher but also as a person.

Lastly, I want to thank my wonderful Family for being together by my side all the time.