Diagnostic metagenomics: potential applications to bacterial, viral and parasitic infections

M. J. PALLEN*

Division of Microbiology and Infection, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK

SUMMARY

The term 'shotgun metagenomics' is applied to the direct sequencing of DNA extracted from a sample without culture or target-specific amplification or capture. In diagnostic metagenomics, this approach is applied to clinical samples in the hope of detecting and characterizing pathogens. Here, I provide a conceptual overview, before reviewing several recent promising proof-of-principle applications of metagenomics in virus discovery, analysis of outbreaks and detection of pathogens in contemporary and historical samples. I also evaluate future prospects for diagnostic metagenomics in the light of relentless improvements in sequencing technologies.

Key words: metagenomics, high-throughput sequencing, diagnosis.

INTRODUCTION

The word *metagenome* was first used in its modern sense in 1998 to refer to the collection of all microbial genomes found in a sample of soil, including sequences from organisms that could not cultured (Handelsman *et al*. 1998). The term has since been generalized to cover any set of multiple genomes found in an environmental or clinical sample.

In its strictest sense, the term *metagenomics* is used to describe the recovery of information from meta-genomes via the creation of shotgun sequence libraries. Such libraries can then be sequenced or, if cloned in an expression system, screened for functional activities of interest. In 2004, Venter and his colleagues described a landmark study on the metagenomics of the Sargasso Sea that established the utility of this approach in exploring not just microbial communities but also taxonomic and sequence space (Venter *et al*. 2004).

Unfortunately, the term metagenomics is often misleadingly used loosely to cover any culture-independent sequence-based profiling of microbial communities, particularly amplification and sequencing of molecular barcodes, such as sequences from rRNA genes. To make it clear that one is using the term in its stricter sense, the phrase *shotgun metagenomics* is often used.

Here, I review the progress of and prospects for the use of shotgun metagenomics in the discovery and detection of microbial pathogens in clinical samples: an approach I call *diagnostic metagenomics*. Due to

* Corresponding author: Division of Microbiology and Infection, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK. E-mail: m.pallen@warwick.ac.uk

constraints of space, I cannot provide a detailed description of all the relevant sequencing and bioinformatics protocols. These are summarized in Fig. 1 and are described in more depth in other recent reviews (Wooley *et al*. 2010; Loman *et al*. 2012; Morgan and Huttenhower, 2012; Weinstock, 2012).

PROBLEMS WITH CURRENT APPROACHES

Before discussing diagnostic metagenomics, it is worth briefly reviewing the problems with existing approaches. Even though we are now well into the 21st century, diagnostic bacteriology is still largely reliant on techniques developed over a century ago in the 1880s: in particular, the detection and characterization of bacteria under the microscope using the staining technique developed by Hans Christian Gram and their propagation as colonies on solid growth media, as pioneered by Robert Koch (Koch, 1881; Gram, 1884). As different bacteria have different growth requirements, this leads to a complex set of workflows for handling samples in the clinical microbiology laboratory, which in turn requires input from a skilled workforce. However, one advantage of the Gram–Koch paradigm in diagnosis is that it is to a large degree open-ended, in that it will detect a wide range of pathogens, including those that are unsuspected. For this reason, microscopy and/or culture have also been used widely in other branches of diagnostic microbiology, including virology and parasitology. For example, microscopy of blood films is still used to diagnose malaria.

In situations where microscopy is cumbersome or unrewarding and culture proves difficult or even impossible, culture-independent approaches to
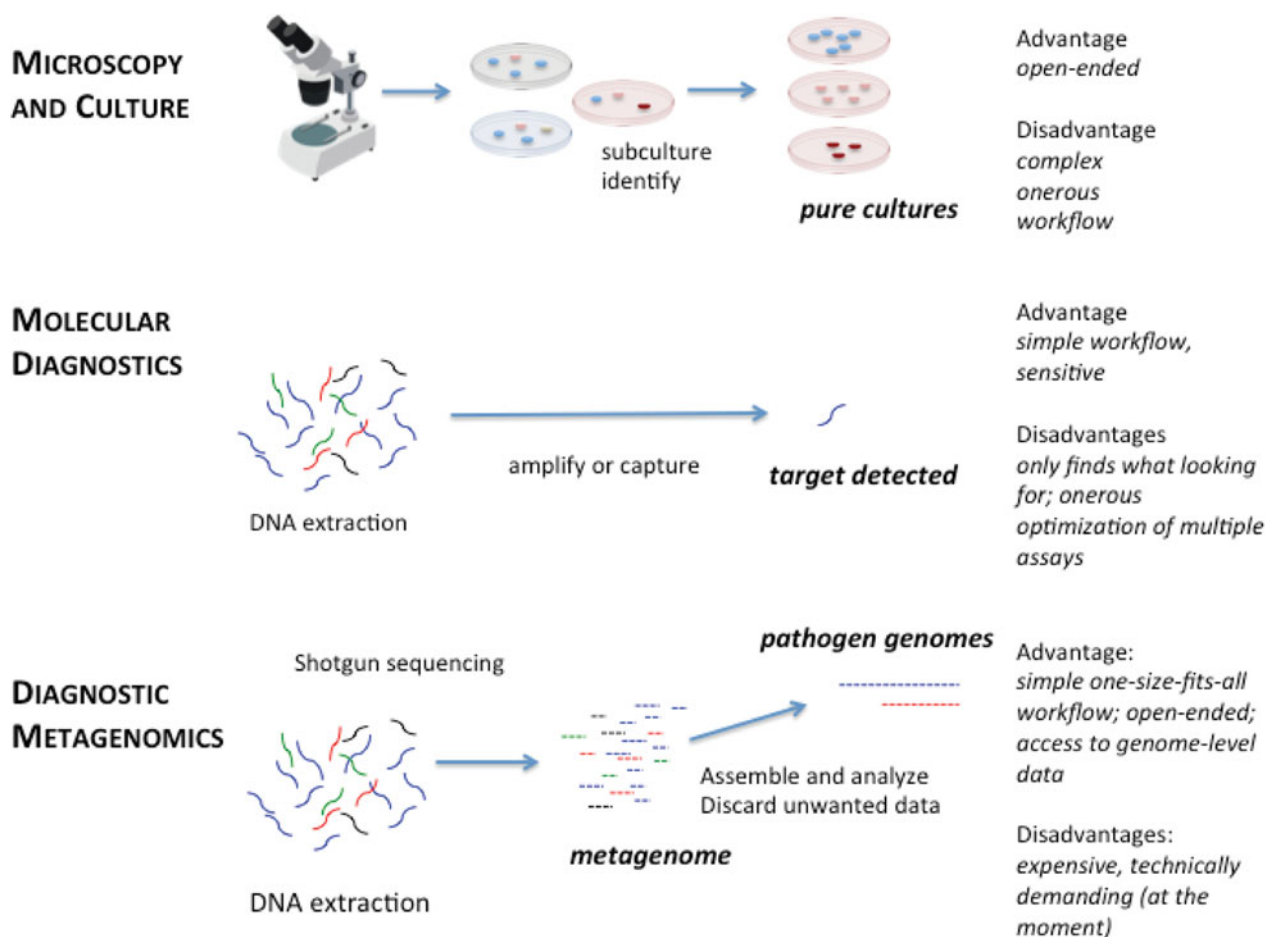
Fig. 1. Comparisons between culture, targeted molecular assays and diagnostic metagenomics.

pathogen detection have been developed, including immunoassays and detection of nucleic acid sequences. However, these approaches are generally target-specific and thus lack the ability to detect unsuspected pathogens. This means a battery of tests may have to be applied to each sample, each of which requires onerous optimization and standardization.

One attempt to combine the open-endedness of culture with the ease of culture-independent molecular methods is the use of a molecular bar code, such as rRNA gene sequences, that can be amplified en masse from a sample using primers targeting conserved sequences, with subsequent analyses focused on taxon-specific sequences flanked by the conserved regions. However, there are several problems with this approach, which have led to it being dubbed 'the one-eyed king' or 'scratched lens' (Forney *et al.* 2004; Temperton and Giovannoni, 2012). For example, so-called 'universal primers' may not in fact detect all organisms, taxon counts may be inflated with some sequencing technologies and taxonomic resolution is generally poor. Thus, a 16S rRNA gene sequence may tell you that *Escherichia coli* is present in a sample, but will give no clue as to a strain's potential for virulence or antibiotic resistance. In addition, an entirely different set of primers will be needed to amplify homologous 18S rRNA barcodes from eukaryotes and there are no broad-range barcodes for viruses.

MICROBIOMES: A DISTRACTION FROM DIAGNOSTIC METAGENOMICS

In recent years, concepts and techniques from environmental microbial ecology have entered clinical microbiology, largely due to recognition of the importance of the *human microbiome*, the complex community of microbes and their genomes associated with the human body. Culture-independent sequence-based approaches to the detection and enumeration of the components of the human microbiome have been widely adopted because they provide greater ease of use and higher throughput than culture-based approaches. These efforts have been energized by steady improvements in high-throughput sequencing, with the result that large sums of money have been spent on high-profile projects to characterize the human microbiome through whole-genome sequencing of cultured isolates and through culture-independent sequence-based profiling of host-associated microbial

Table 1. Examples of diagnostic metagenomics

Pathogen	Context	Reference
Lymphocytic choriomeningitis virus	Fatal infections in transplant recipients	(Palacios *et al.* 2008)
Lujo virus	Hospital outbreak of haemorrhagic fever	(Briese *et al.* 2009)
Bundibugyo ebolavirus	Outbreak of haemorrhagic fever in Uganda	(Towner *et al.* 2008)
Norovirus	Fecal samples	(Nakamura *et al.* 2009; Batty *et al.* 2013; Wong *et al.* 2013)
Adenovirus, respiratory syncytial virus, rhinovirus, influenza and parainfluenza bocavirus, human herpesvirus 5	Nasopharyngeal aspirates from respiratory tract infections	(Yang *et al.* 2011)
Pandemic influenza	Nasopharyngeal swabs	(Greninger *et al.* 2010)
Rotavirus, adenovirus, picobirnavirus, enterovirus	Fecal samples from children with acute flaccid paralysis	(Victoria *et al.* 2009)
Chlamydia trachomatis, Prevotella, Gardnerella, Clostridiales genomosp. BVAB3, *Mycoplasma hominis*	Vaginal swabs	(Andersson *et al.* 2013; Seth-Smith *et al.* 2013)
Escherichia coli, Enterococcus, Gardnerella, Proteus	Urine samples	(Hasman *et al.* 2014)
Borrelia burgdorferi, Mycobacterium tuberculosis, Mycobacterium leprae, Yersinia pestis, Plasmodium falciparum, Toxoplasma gondii	Historic and prehistoric material	(Bos *et al.* 2011; Bouwman *et al.* 2012; Keller *et al.* 2012; Chan *et al.* 2013; Khairat *et al.* 2013; Schuenemann *et al.* 2013)

communities (Turnbaugh *et al.* 2007; Peterson *et al.* 2009; Arumugam *et al.* 2011; Le Chatelier *et al.* 2013).

It is worth highlighting three themes that have emerged from the application of microbial ecology to clinical microbiology. First is the implicit assumption of the 'uncultured microbial majority' (Rappe and Giovannoni, 2003) – i.e. that most bacteria cannot be isolated in the laboratory in pure culture and so molecular methods will report a wider range of organisms than culture and are generally more sensitive. While most researchers accept these assumptions, in a provocative counter-blast, Raoult and his colleagues have claimed that adoption of a wide range of cultural approaches (which they term 'culturomics') is actually more sensitive than sequence-based approaches (Lagier *et al.* 2012). By contrast, Dowd and his colleagues claim that management of infection is improved when culture-independent approaches are adopted (Dowd *et al.* 2011).

A second theme is the recognition that differences in host-associated microbial communities can influence the balance between health and disease in conditions not normally thought of as microbial or infectious in origin: for example, inflammatory bowel disease, cancer or obesity (Kinross *et al.* 2011; Lozupone *et al.* 2012).

The third emerging theme is that it may not be sufficient to focus diagnostic efforts on single 'headline pathogens' in clinical samples that are thought single-handedly to cause disease. Instead, it is now recognized that interactions between organisms in a community can influence disease outcome and in some cases it might even be appropriate to treat a whole microbial community as a pathogenic entity (Rogers *et al.* 2010).

Although community profiling using molecular barcodes still predominates in the field of microbiome studies, there have now been several shotgun metagenomic studies of human-derived samples, chiefly focused on the characterization of such communities and their genes in healthy volunteers and non-infectious diseases (Qin *et al.* 2010, 2012; Arumugam *et al.* 2011; Karlsson *et al.* 2013). Curiously, comparatively little attention has focused on the question of whether metagenomics can be used to discover, detect and characterize pathogens in samples from diseased individuals.

VIRUS DISCOVERY

Not surprisingly, given the difficulty or impossibility of culturing most viruses, virologists were the first to explore the potential of open-ended, shotgun sequencing to identify and detect human-associated viruses. The genomes of DNA viruses can be recovered through shotgun sequencing of DNA directly extracted from a sample. To detect RNA viruses, RNA extracted from a sample has to be converted to cDNA (Batty *et al.* 2013). In such cases, one is searching for a viral genome in the midst of a sample-derived metatranscriptome.

Diagnostic viral metagenomics has now been used to detect an unknown pathogen in a number of high-profile cases or outbreaks of disease (Table 1). In an

early example, sequencing of cDNA from three transplant recipients with fatal infections yielded 14 sequences resembling segments of the genome of lymphocytic choriomeningitis virus (Palacios *et al*. 2008). Similarly diagnostic metagenomics uncovered a novel arenavirus responsible for a hospital outbreak of haemorrhagic fever in South Africa (Briese *et al*. 2009) and identified a novel species of Ebola virus (*Bundibugyo ebolavirus*) from Uganda (Towner *et al*. 2008).

Metagenomics has now been used widely in virus discovery (Capobianchi *et al*. 2013; Smits and Osterhaus, 2013), revealing the existence of numerous 'orphan viruses' that have not been associated with any disease and thereby establishing the existence of a normal human virus microbiome or 'virome' (Li and Delwart, 2011; Lecuit and Eloit, 2013). Diagnostic metagenomics has also been used for the detection of established viral pathogens (e.g. influenza and norovirus) in clinical samples (Nakamura *et al*. 2009).

DETECTION OF BACTERIAL PATHOGENS

In a pioneering 2008 metagenomics survey of fecal samples from a single individual, Nakamura and colleagues showed the feasibility of detecting bacterial pathogens such as *Campylobacter* through unbiased sequencing of DNA extracted from stool samples (Nakamura *et al*. 2008). In this case, 156 *Campylobacter* sequences were found in a sample taken during a bout of illness, but were absent from a convalescent sample from the same individual.

More recently, Loman, myself and others, in collaboration with clinical microbiologists from Hamburg, explored the potential of diagnostic metagenomics on stool samples collected during the outbreak of Shiga-toxigenic *E. coli* O104:H4 that struck Germany in May–June 2011 (Loman *et al*. 2013). Contrary to expectations, we obtained deep coverage of the outbreak strain genome from several stool metagenomes, even using a benchtop-sequencing platform (the Illumina MiSeq). We subsequently sequenced a larger set of stool metagenomes on a higher-throughput instrument (the HiSeq2500) and in some cases obtained much deeper coverage of a pathogen genome. We also recovered genome-level coverage of other pathogens (*Campylobacter jejuni, Clostridium difficile, Salmonella enterica*) that had been detected by routine microbiological investigation in several STEC-negative samples.

In this study, we established proof-of-principle that metagenomics could be used not only to detect, but also to characterize, bacterial pathogens within a sample. For example, we were able to obtain typing data and phylogenetic profiles for strains of use in epidemiology and population genetics directly from the metagenomes without culture. Another study has recently confirmed the utility of metagenomics in recovering single-nucleotide polymorphisms from different strains of gut bacteria (Schloissnig *et al*. 2013).

In our study, we also demonstrated the open-endedness of the approach by detecting organisms that we did not expect to find. In one sample positive for *C. difficile*, we also discovered an abundance of sequences from *Campylobacter concisus*, which has been described as an emerging diarrhoeal pathogen. In several STEC-positive samples, we discovered sequences from *C. difficile*. These findings not only illustrate the ability of diagnostic metagenomics to find 'unknown unknowns' but also question the assumption that a single headline pathogen causes disease on its own: are the second-line pathogens in these samples merely colonizing the altered microenvironment caused by disease or are they also contributing to pathology?

In the initial phase of our study, we aligned reads from the metagenomes against the known outbreak strain genome. This might be seen as 'cheating', as we were in effect using a crib sheet to answer the question of whether the strain was present or not. In a subsequent phase, we used the MetaPhlAn pipeline to identify reads that were likely to originate from pathogens (Segata *et al*. 2012). It was this more open-ended approach that allowed us to discover the unknown unknowns. In the final phase of the project, we devised an entirely *de novo* approach to identify the outbreak strain and recover its genome. We started with two simplifying assumptions: that strain-specific reads from the outbreak strain would be present in at least half of the samples and that they would be absent from the stool metagenomes of healthy individuals. Applying these criteria to the collection of fecal metagenomes, we obtained a strain-specific set of reads. By clustering these reads with other reads that showed the same patterns of abundance in different samples, we were able to obtain the entire outbreak strain genome.

Several recent studies have shown the ability of diagnostic metagenomics to make a diagnosis of infection long after the individual is dead. In 2011, a study of the genome of the Tyrolean ice mummy Ötzi revealed sequences from the bacterium *Borrelia burgdorferi*, identifying the first known case of Lyme disease (Keller *et al*. 2012). Krause and colleagues subsequently recovered a genome from the leprosy bacillus, *Mycobacterium leprae*, from the metagenome obtained from a historical dental sample (Schuenemann *et al*. 2013). Around the same time, we described the first example of post-mortem metagenomic diagnosis of tuberculosis, from mummified lung tissue from a young woman who died in 1797 (Chan *et al*. 2013). We found evidence of mixed infection with two distinct genotypes of *Mycobacterium tuberculosis*, both related to strains circulating in Europe and North America.

Recently, diagnostic metagenomics has also been applied to urine samples, allowing the recovery of uropathogen genome sequences directly from clinical material without culture, while also providing taxonomic, epidemiological and phylogenetic data (Hasman *et al*. 2014). In low-biomass samples, it may still be possible to obtain enough pathogen DNA for sequencing thanks to target-independent whole-genome amplification. A recent metagenomics study using this approach on a vaginal swab reported recovery not just of a chlamydial genome, but also high genome-wide coverage for *Prevotella melaninogenica*, *Gardnerella vaginalis* and *Mycoplasma hominis* (Andersson *et al*. 2013).

RELEVANCE TO PARASITOLOGY

Diagnostic metagenomics has so far seen little use in the detection of parasitic infection, although there is no reason to expect that parasitology will not soon catch up with other branches of microbiology in this regard. In a single unreplicated study, it has been claimed that *Plasmodium* and *Toxoplasma* sequences can be found in the metagenomes of Egyptian mummies (Khairat *et al*. 2013). However, no sequence data were presented, so it is hard to assess the veracity of this claim. Shotgun metagenomics of cDNA has been used to analyse the taxa associated with *Lutzomyia longipalpis*, the vector of visceral leishmaniasis (McCarthy *et al*. 2011). Potential future applications of metagenomics in parasitology include recovery of genomic-epidemiological data (Griffing *et al*. 2011; Khaireh *et al*. 2013) and determining the influence of parasites on the microbial ecology of the gut. It remains to be seen how much parasite DNA can be detected in samples without amplification or capture: this might present a challenge for bloodstream infections. However, in samples such as feces, where parasite biomass is likely to be relatively high, this will be much less of an issue. Also of interest might be the recovery of parasite sequences from historical material; museum samples might provide insights into extinct parasite lineages.

FUTURE DIRECTIONS

Diagnostic metagenomics brings the promise of an open-ended, assumption-free one-size-fits-all workflow that could be applied to any specimen to detect any kind of pathogen (viruses, bacteria, fungi and parasites). Nonetheless, the sceptic will dismiss diagnostic metagenomics as 'a sledgehammer to crack a nut' in that it is orders of magnitude too expensive for routine use and requires significant expert technical and bioinformatics input. Plus, why bother when there are easier methods? However, with likely future improvements in the ease, throughput and cost-effectiveness of sequencing, twinned with commoditization of laboratory and informatics

workflows, one can foresee a tipping point when a unified automated metagenomics-based workflow will start to compete with the plethora of methods currently in use in the diagnostic laboratory, while also delivering additional useful information (e.g. on genomic epidemiology, antimicrobial resistance, virulence). In the case of an outbreak of life-threatening infection with an unknown cause, we may already have reached that point.

However, we are faced with a number of challenges before diagnostic metagenomics can be used routinely. One is the broad range of abundances of different taxa within a specimen. Metagenomics does appear to be able to deliver whole-genome coverage of the most abundant organisms, but it is probably not safe to assume that the pathogenic species in a sample are always the most abundant. In addition, assembly of a metagenome presents considerable computational challenges, although there have been recent methodological improvements in this area.

A number of potential approaches to simplifying the problem can be envisaged. It is possible to capture or enrich for cells or sequences from a pathogen of interest. This approach has been applied to clinical samples to obtain genomes from *Chlamydia* genomes without culture (Seth-Smith *et al*. 2013) and has been used to recover leprosy, tuberculosis and plague genomes from historical and ancient material (Bos *et al*. 2011; Bouwman *et al*. 2012; Schuenemann *et al*. 2013). If used alone, this risks compromising the open-endedness of diagnostic metagenomics, but such approaches could be used as adjuncts to an open-ended approach.

It is possible to simplify the analysis of the metagenome by sorting cells or DNA fragments into sub-metagenomic bins: for example, a recent study reported use of cell sorting to recover the genome of an unculturable bacterium from a hospital sink biofilm (McLean *et al*. 2013). Another attractive option would be to normalize a sample, so that all taxa and their sequences appear equally abundant. Experimental approaches to a similar problem for cDNA libraries have proven successful using nuclease cleavage of abundant re-annealed DNA duplexes (Christodoulou *et al*. 2011) and this might also work for metagenomes. The imminent launch of cheap and accessible long-read high-throughput nanopore sequencing also promises to improve the prospects for recovering pathogen genomes from metagenomes (Maitra *et al*. 2012). If these technologies live up to the hype, it will soon be possible to pipette a sample into a highly portable USB-stick-sequencer and see sequences stream off on to a computer in real time, which could transform field work and make inroads into the diagnostic laboratory.

In conclusion, it is clear that diagnostic metagenomics works in a research setting and already has a role to play in identifying the causes of unknown

illnesses and outbreaks. A number of hurdles need to be overcome before it can be integrated into routine practice. Nonetheless, it is possible to envisage a tipping point, sometime in the next decade or two, when the old Gram–Koch paradigm gives way to a diagnostic metagenomics approach to the detection and characterization of pathogens in clinical samples fit for the 21st century.

REFERENCES

Andersson, P., Klein, M., Lilliebridge, R. A. and Giffard, P. M. (2013). Sequences of multiple bacterial genomes and a *Chlamydia trachomatis* genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen. *Clinical Microbiology and Infection* **19**, E405–E408.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Dore, J., Antolin, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K. U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S. D. and Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature* **473**, 174–180.

Batty, E. M., Wong, T. H., Trebes, A., Argoud, K., Attar, M., Buck, D., Ip, C. L., Golubchik, T., Cule, M., Bowden, R., Manganis, C., Klenerman, P., Barnes, E., Walker, A. S., Wyllie, D. H., Wilson, D. J., Dingle, K. E., Peto, T. E., Crook, D. W. and Piazza, P. (2013). A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. *PLoS One* **8**, e66129.

Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., McPhee, J. B., DeWitte, S. N., Meyer, M., Schmedes, S., Wood, J., Earn, D. J., Herring, D. A., Bauer, P., Poinar, H. N. and Krause, J. (2011). A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510.

Bouwman, A. S., Kennedy, S. L., Muller, R., Stephens, R. H., Holst, M., Caffell, A. C., Roberts, C. A. and Brown, T. A. (2012). Genotype of a historic strain of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences USA* **109**, 18511–18516.

Briese, T., Paweska, J. T., McMullan, L. K., Hutchison, S. K., Street, C., Palacios, G., Khristova, M. L., Weyer, J., Swanepoel, R., Egholm, M., Nichol, S. T. and Lipkin, W. I. (2009). Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathogens* **5**, e1000455.

Capobianchi, M. R., Giombini, E. and Rozera, G. (2013). Next-generation sequencing technology in clinical virology. *Clinical Microbiology and Infection* **19**, 15–22.

Chan, J. Z., Sergeant, M. J., Lee, O. Y., Minnikin, D. E., Besra, G. S., Pap, I., Spigelman, M., Donoghue, H. D. and Pallen, M. J. (2013). Metagenomic analysis of tuberculosis in a mummy. *New England Journal of Medicine* **369**, 289–290.

Christodoulou, D. C., Gorham, J. M., Herman, D. S. and Seidman, J. G. (2011). Construction of normalized RNA-seq libraries for next-generation sequencing using the crab duplex-specific nuclease. *Current Protocols in Molecular Biology* **94**, 4.12.1–4.12.11.

Dowd, S. E., Wolcott, R. D., Kennedy, J., Jones, C. and Cox, S. B. (2011). Molecular diagnostics and personalised medicine in wound care: assessment of outcomes. *Journal of Wound Care* **20**, 232, 234–9.

Forney, L. J., Zhou, X. and Brown, C. J. (2004). Molecular microbial ecology: land of the one-eyed king. *Current Opinion in Microbiology* **7**, 210–220.

Gram, H. C. (1884). Über die isolierte Färbung der Schizomyceten in Schnitt- und Trockenpräparaten. *Fortschritte der Medizin* **2**, 185–189.

Greninger, A. L., Chen, E. C., Sittler, T., Scheinerman, A., Roubinian, N., Yu, G., Kim, E., Pillai, D. R., Guyard, C., Mazzulli, T., Isa, P., Arias, C. F., Hackett, J., Schochetman, G., Miller, S., Tang, P. and Chiu, C. Y. (2010). A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One* **5**, e13381.

Griffing, S. M., Mixson-Hayden, T., Sridaran, S., Alam, M. T., McCollum, A. M., Cabezas, C., Marquino Quezada, W., Barnwell, J. W., De Oliveira, A. M., Lucas, C., Arrospide, N., Escalante, A. A., Bacon, D. J. and Udhayakumar, V. (2011). South American *Plasmodium falciparum* after the malaria eradication era: clonal population expansion and survival of the fittest hybrids. *PLoS One* **6**, e23486.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemical Biology* **5**, R245–R249.

Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C. A., Frimodt-Moller, N. and Aarestrup, F. M. (2014). Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *Journal of Clinical Microbiology* **52**, 139–146.

Karlsson, F. H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C. J., Fagerberg, B., Nielsen, J. and Backhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103.

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., Stade, B., Franke, A., Mayer, J., Spangler, J., McLaughlin, S., Shah, M., Lee, C., Harkins, T. T., Sartori, A., Moreno-Estrada, A., Henn, B., Sikora, M., Semino, O., Chiaroni, J., Rootsi, S., Myres, N. M., Cabrera, V. M., Underhill, P. A., Bustamante, C. D., Vigl, E. E., Samadelli, M., Cipollini, G., Haas, J., Katus, H., O'Connor, B. D., Carlson, M. R., Meder, B., Blin, N., Meese, E., Pusch, C. M. and Zink, A. (2012). New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Communications* **3**, 698.

Khairat, R., Ball, M., Chang, C. C., Bianucci, R., Nerlich, A. G., Trautmann, M., Ismail, S., Shanab, G. M., Karim, A. M., Gad, Y. Z. and Pusch, C. M. (2013). First insights into the metagenome of Egyptian mummies using next-generation sequencing. *Journal of Applied Genetics* **54**, 309–325.

Khaireh, B. A., Assefa, A., Guessod, H. H., Basco, L. K., Khaireh, M. A., Pascual, A., Briolant, S., Bouh, S. M., Farah, I. H., Ali, H. M., Abdi, A. I., Aden, M. O., Abdillahi, Z., Ayeh, S. N., Darar, H. Y., Koeck, J. L., Rogier, C., Pradines, B. and Bogreau, H. (2013). Population genetics analysis during the elimination process of *Plasmodium falciparum* in Djibouti. *Malaria Journal* **12**, 201.

Kinross, J. M., Darzi, A. W. and Nicholson, J. K. (2011). Gut microbiome-host interactions in health and disease. *Genome Medicine* **3**, 14.

Koch, R. (1881). Zur Untersuchung von pathogenen Organismen. *Mitteilungen aus dem Kaiserliche Gesundheitsamte* **1**, 1–48.

Lagier, J. C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G., Maraninchi, M., Trape, J. F., Koonin, E. V., La Scola, B. and Raoult, D. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection* **18**, 1185–1193.

**Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J. M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jorgensen, T., Brandslund, I., Nielsen, H. B., Juncker, A. S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E. G., Brunak, S., Clement, K., Dore, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W. M., Zucker, J. D., Raes, J., Hansen, T., Bork, P., Wang, J., Ehrlich, S. D., Pedersen, O., Guedon, E., Delorme, C., Layec, S., Khaci, G., van de Guchte, M., Vandemeulebrouck, G., Jamet, A., Dervyn, R., Sanchez, N., Maguin, E., Haimet, F., Winogradsky, Y., Cultrone, A., Leclerc, M., Juste, C., Blottiere, H., Pelletier, E., LePaslier, D., Artiguenave, F., Bruls, T., Weissenbach, J., Turner, K., Parkhill, J., Antolin, M., Manichanh, C., Casellas, F., Boruel, N., Varela, E., Torrejon, A., Guarner, F., Denariaz, G., Derrien, M., van Hylckama Vlieg, J. E.,

Veiga, P., Oozeer, R., Knol, J., Rescigno, M., Brechot, C., M'Rini, C., Merieux, A. and Yamada, T. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546.

Lecuit, M. and Eloit, M. (2013). The human virome: new tools and concepts. *Trends in Microbiology* **21**, 510–515.

Li, L. and Delwart, E. (2011). From orphan virus to pathogen: the path to the clinical lab. *Current Opinion in Virology* **1**, 282–288.

Loman, N. J., Constantinidou, C., Chan, J. Z., Halachev, M., Sergeant, M., Penn, C. W., Robinson, E. R. and Pallen, M. J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology* **10**, 599–606.

Loman, N. J., Constantinidou, C., Christner, M., Rohde, H., Chan, J. Z., Quick, J., Weir, J. C., Quince, C., Smith, G. P., Betley, J. R., Aepfelbacher, M. and Pallen, M. J. (2013). A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *Journal of the American Medical Association* **309**, 1502–1510.

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230.

Maitra, R. D., Kim, J. and Dunbar, W. B. (2012). Recent advances in nanopore sequencing. *Electrophoresis* **33**, 3418–3428.

McCarthy, C. B., Diambra, L. A. and Rivera Pomar, R. V. (2011). Metagenomic analysis of taxa associated with *Lutzomyia longipalpis*, vector of visceral leishmaniasis, using an unbiased high-throughput approach. *PLoS Neglected Tropical Diseases* **5**, e1304.

McLean, J. S., Lombardo, M. J., Badger, J. H., Edlund, A., Novotny, M., Yee-Greenbaum, J., Vyahhi, N., Hall, A. P., Yang, Y., Dupont, C. L., Ziegler, M. G., Chitsaz, H., Allen, A. E., Yooseph, S., Tesler, G., Pevzner, P. A., Friedman, R. M., Nealson, K. H., Venter, J. C. and Lasken, R. S. (2013). Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proceedings of the National Academy of Sciences USA* **110**, E2390–E2399.

Morgan, X. C. and Huttenhower, C. (2012). Chapter 12: human microbiome analysis. *PLoS Computational Biology* **8**, e1002808.

Nakamura, S., Maeda, N., Miron, I. M., Yoh, M., Izutsu, K., Kataoka, C., Honda, T., Yasunaga, T., Nakaya, T., Kawai, J., Hayashizaki, Y., Horii, T. and Iida, T. (2008). Metagenomic diagnosis of bacterial infections. *Emerging Infectious Diseases* **14**, 1784–1786.

Nakamura, S., Yang, C. S., Sakon, N., Ueda, M., Tougan, T., Yamashita, A., Goto, N., Takahashi, K., Yasunaga, T., Ikuta, K., Mizutani, T., Okamoto, Y., Tagami, M., Morita, R., Maeda, N., Kawai, J., Hayashizaki, Y., Nagai, J., Horii, T., Iida, T. and Nakaya, T. (2009). Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* **4**, e4219.

Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., Conlan, S., Quan, P. L., Hui, J., Marshall, J., Simons, J. F., Egholm, M., Paddock, C. D., Shieh, W. J., Goldsmith, C. S., Zaki, S. R., Catton, M. and Lipkin, W. I. (2008). A new arenavirus in a cluster of fatal transplant-associated diseases. *New England Journal of Medicine* **358**, 991–998.

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., Bonazzi, V., McEwen, J. E., Wetterstrand, K. A., Deal, C., Baker, C. C., Di Francesco, V., Howcroft, T. K., Karp, R. W., Lunsford, R. D., Wellington, C. R., Belachew, T., Wright, M., Giblin, C., David, H., Mills, M., Salomon, R., Mullins, C., Akolkar, B., Begg, L., Davis, C., Grandison, L., Humble, M., Khalsa, J., Little, A. R., Peavy, H., Pontzer, C., Portnoy, M., Sayre, M. H., Starke-Reed, P., Zakhari, S., Read, J., Watson, B. and Guyer, M. (2009). The NIH Human Microbiome Project. *Genome Research* **19**, 2317–2323.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D. and Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y.,

Xue, W., Li, J., Han, L., Lu, D., Wu, P., Dai, Y., Sun, X., Li, Z., Tang, A., Zhong, S., Li, X., Chen, W., Xu, R., Wang, M., Feng, Q., Gong, M., Yu, J., Zhang, Y., Zhang, M., Hansen, T., Sanchez, G., Raes, J., Falony, G., Okuda, S., Almeida, M., LeChatelier, E., Renault, P., Pons, N., Batto, J. M., Zhang, Z., Chen, H., Yang, R., Zheng, W., Li, S., Yang, H., Wang, J., Ehrlich, S. D., Nielsen, R., Pedersen, O., Kristiansen, K. and Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60.

Rappe, M. S. and Giovannoni, S. J. (2003). The uncultured microbial majority. *Annual Reviews in Microbiology* **57**, 369–394.

Rogers, G. B., Stressmann, F. A., Walker, A. W., Carroll, M. P. and Bruce, K. D. (2010). Lung infections in cystic fibrosis: deriving clinical insight from microbial complexity. *Expert Reviews in Molecular Diagnostics* **10**, 187–196.

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D. R., Kultima, J. R., Martin, J., Kota, K., Sunyaev, S. R., Weinstock, G. M. and Bork, P. (2013). Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45–50.

Schuenemann, V. J., Singh, P., Mendum, T. A., Krause-Kyora, B., Jager, G., Bos, K. I., Herbig, A., Economou, C., Benjak, A., Busso, P., Nebel, A., Boldsen, J. L., Kjellstrom, A., Wu, H., Stewart, G. R., Taylor, G. M., Bauer, P., Lee, O. Y., Wu, H. H., Minnikin, D. E., Besra, G. S., Tucker, K., Roffey, R., Sow, S. O., Cole, S. T., Nieselt, K. and Krause, J. (2013). Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* **341**, 179–183.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O. and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* **9**, 811–814.

Seth-Smith, H. M., Harris, S. R., Skilton, R. J., Radebe, F. M., Golparian, D., Shipitsyna, E., Duy, P. T., Scott, P., Cutcliffe, L. T., O'Neill, C., Parmar, S., Pitt, R., Baker, S., Ison, C. A., Marsh, P., Jalal, H., Lewis, D. A., Unemo, M., Clarke, I. N., Parkhill, J. and Thomson, N. R. (2013). Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. *Genome Research* **23**, 855–866.

Smits, S. L. and Osterhaus, A. D. (2013). Virus discovery: one step beyond. *Current Opinion in Virology* doi: 10.1016/j.coviro.2013.03.007. [Epub ahead of print].

Temperton, B. and Giovannoni, S. J. (2012). Metagenomics: microbial diversity through a scratched lens. *Current Opinion in Microbiology* **15**, 605–612.

Towner, J. S., Sealy, T. K., Khristova, M. L., Albarino, C. G., Conlan, S., Reeder, S. A., Quan, P. L., Lipkin, W. I., Downing, R., Tappero, J. W., Okware, S., Lutwama, J., Bakamutumaho, B., Kayiwa, J., Comer, J. A., Rollin, P. E., Ksiazek, T. G. and Nichol, S. T. (2008). Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathogens* **4**, e1000212.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R. and Gordon, J. I. (2007). The human microbiome project. *Nature* **449**, 804–810.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H. and Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74.

Victoria, J. G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S. and Delwart, E. (2009). Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *Journal of Virology* **83**, 4642–4651.

Weinstock, G. M. (2012). Genomic approaches to studying the human microbiota. *Nature* **489**, 250–256.

Wong, T. H., Dearlove, B. L., Hedge, J., Giess, A. P., Piazza, P., Trebes, A., Paul, J., Smit, E., Smith, E. G., Sutton, J. K., Wilcox, M. H., Dingle, K. E., Peto, T. E., Crook, D. W., Wilson, D. J. and Wyllie, D. H. (2013). Whole genome sequencing and *de novo* assembly identifies Sydney-like variant noroviruses and recombinants during the winter 2012/2013 outbreak in England. *Virology Journal* **10**, 335.

Wooley, J. C., Godzik, A. and Friedberg, I. (2010). A primer on metagenomics. *PLoS Computational Biology* **6**, e1000667.

Yang, J., Yang, F., Ren, L., Xiong, Z., Wu, Z., Dong, J., Sun, L., Zhang, T., Hu, Y., Du, J., Wang, J. and Jin, Q. (2011). Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *Journal of Clinical Microbiology* **49**, 3463–3469.