Whole Genome Sequencing

Joergen Schlundt
School of Chemical and Biomedical Engineering,
Nanyang Technological University (NTU),
Singapore, Singapore
Nanyang Technological University Food
Technology Centre (NAFTEC), Nanyang
Technological University (NTU), Singapore,
Singapore

Synonyms

High-throughput DNA sequencing; Next-generation DNA sequencing; Next-generation sequencing (NGS)

Definition

Whole genome sequencing (WGS) refers to the process of presenting the full sequence of DNA, referring to the sequence of the four letters (A, C, G, and T) which constitute the DNA molecule, containing the hereditary material of all types of life. WGS is typically based on the use of new, rapid methodologies to sequence DNA, mostly referred to as “Next-generation sequencing” (NGS) or high-throughput DNA sequencing.

Overview

WGS is done by using NGS methodology, maybe more correctly referred to as next-generation DNA sequencing methodology. DNA (deoxyribonucleic acid) provides the language of our genes, and is basically the same for all life on earth. DNA consists of a double-helix strand of four nucleic acid molecules: adenine, cytosine, guanine, and thymine, as first described by Watson and Crick (1953). There is a minor deviation from this generic life language, in that some virus use RNA (ribonucleic acid) instead, which leaves them with a single-strand molecule with the same sort of backbone structure but with uracil instead of thymine. So the genetic code of all humans, animals, plants, and microorganisms is basically written in the same four letters (four nucleotides): ACGT for DNA or, only for certain virus: ACGU for RNA. This is the explanation why all life is built the same way (as per genetic vocabulary) and also the basic explanation why all life forms basically are within the same family tree and have developed from the same primordial organisms, as originally suggested by Darwin (for more on Darwin’s work and theory of evolution, see Ayala 2007).

Apart from many other interesting, philosophical implications of the detection of the structure and function of DNA, it is important to note, that the construction of certain, specifically important enzymes or other molecules in human or animal bodies are actually the same down (or up) through

the family tree. For instance, tryptophan is one of the amino acids that are essential to humans, because the human body cannot produce it; however, most plants and many microorganisms can actually produce tryptophan. But there is basically only one way to produce tryptophan, meaning this construction has been created in DNA only once, and has then been inherited down through bacteria (very simple organisms with no cell nucleus) to plants (eukaryotic organisms with a cell nucleus). Because all life (including bacteria and plants) are basically one great family, as seen from the genetic make-up, the tryptophan genes have simply been inherited down multiple (tens of thousands) species, basically unchanged (Gutiérrez-Preciado et al. 2010). However, some would say unfortunately, it did not make it into the animal trunk of the life tree, which is why animals, including humans, have to get tryptophan from plants or microorganisms.

The Darwinist explanation of development of species through time has often been referred to as “survival of the fittest.” One very important, present development in dangerous bacteria surrounding us is the creation of antibiotic resistant bacteria, i.e., bacteria that can withstand the antibiotics that kill “normal” bacteria. The development of these resistant bacteria can very well be described as a demonstration of “survival of the fittest.” When we use antibiotics in humans, animals, or the environment, we create a pressure, meaning we kill all the billions of target microorganisms in the habitat we treat. But if a few of these have a mutation causing them to become resistant to the antibiotic, these will of course have an advantage over – i.e., be more fit than – the normal, sensitive microorganisms. A fact that lead Richard Dawkins to write in the foreword to “The Blind Watchmaker” 2006 edition: “If more Doctors (Veterinarians and Farmers – *my addition, cleared with Dawkins in 2016*) understood Darwinism, humanity would not now be facing a crisis of antibiotic resistance” (Dawkins 2006). I added veterinarians and farmers because the majority of antibiotics are now used to just make animals grow faster, not to treat sick humans (or animals).

So all life forms use the same alphabet of four letters: ACGT (with a U replacing T for RNA), and researchers could start reading this alphabet in the mid-1960s. This major life-science revolution started with the published sequence of some small RNA molecules from the common yeast, *Saccharomyces cerevisiae*. The sequencing of the more complex DNA molecule was more difficult, and only in 1975 a group led by F. Sanger published the first DNA sequence from the bacterium *Escherichia coli* (*E. coli*). Although the method was cumbersome, the group was able to sequence the full genome of a DNA virus, constituting 5386 nucleotides (Sanger et al. 1977a). The same year, the group perfected the sequencing method into what is now known to be the Sanger method (Sanger et al. 1977b). This was a major breakthrough and allowed long stretches of DNA to be accurately sequenced, and actually earned Sanger his second Nobel prize in Chemistry in 1980. The Sanger method was eventually used to sequence the entire human genome in the Human Genome Project.

The Human Genome Project (HGP) was an international research project aiming at determining the sequence of nucleotides in human DNA. The planning started in 1984 but the project was only formally launched in 1990 mainly with funding from the US Government, but with support from a big international consortium of many international partners. The project was finalized in 2003 and basically presented a generic human genome of more than three billion nucleotides (IHGSC 2001).

It carries some, maybe bittersweet, irony that the year after this major multi-year, multi-institutional international project was finalized, next-generation sequencing was invented, providing a much quicker, much more accurate, and soon also much cheaper methodology for DNA sequencing.

Next-Generation Sequencing (NGS) and Whole Genome Sequencing (WGS)

Sanger sequencing (“First generation” technology) was used for DNA sequencing from 1977; although the whole genome of humans was only

sequenced by 2003, in the period 1977–2003, many microbial whole genome sequences (WGS) were actually published (Besser et al. 2018). Sanger sequencing produces long, up to 500–1000 nucleotides (the four letters: AGCT), and the reads of these letters are typically high-quality, meaning very few errors. Thus, Sanger sequencing was long regarded the gold standard for sequencing DNA and was used to sequence the first bacterial genome, *Haemophilus influenza* in 1995 (Fleischmann et al. 1995). Many other bacterial genomes were sequenced over the next few years (Blattner et al. 1997; Cole et al. 1998; Fraser et al. 1998; Kunst et al. 1997; Parkhill et al. 2001). The reason why whole bacterial genomes were sequenced before the human genome is basically that while a bacterial genome is typically within the region of four million letters (size of a standard bible or ten qurans or ten torahs) a human genome is close to 1000-fold larger, so around four billion letters or roughly 500 times larger than the longest novel ever written (Marcel Proust: In search of lost time).

In 2005, the NGS (“next generation” technology) era began and this high-throughput (fast sequencing) technology in the beginning produced short sequencing reads (50–400 letters), but soon it also produced long sequencing reads (1000–100,000 letters) which could be generated in a single machine run. A “read” represents an unbroken piece of DNA read all the way through by the machine. Since the process can cause the long DNA molecule (four million letters for a bacterium) to break, the real outcome is actually a number of strands of letter sequences or “reads.” The short-read technologies, provided by companies such as Illumina (e.g., MiSeq and HiSeq machines) and Life Technologies (e.g., Ion Torrent Personal Genome Machine) now produce read lengths ranged from 100 to 600 letters with low error rate (usually less than 1% error) (Goodwin et al. 2016).

To generate a good-quality sequence of a full bacterial genome: the Whole Genome Sequence (WGS), the longer read technologies (“third generation” technology) that are incorporated into platforms by Pacific Biosystems (e.g., PacBio RSII) and Oxford Nanopore (e.g., MinION),

together with the above-mentioned short read technologies are commonly used. When both technologies are used in combination, the short reads generate good quality while the long reads can close the gaps for a bacterial genome. When sequencing a full human genome, in principle the same methodologies are in use, although some parts of the human genome are often excluded for technical reasons. But more importantly: what can this new knowledge be used for, and why is NGS by some referred to as a revolutionary technology.

Key Research Findings

The Use of Next-Generation Sequencing in Health Promotion and Disease Prevention for Older Adults

Since the NGS technology is used for humans, as well as any other form of life, the importance of whole genome sequencing (WGS) as related to human health can roughly be divided in three categories:

- (a) Analysis of human DNA to characterize health status and risk of disease as well as the potential for future specific treatment of diseases linked to certain genes
- (b) Analysis of microbiological DNA to characterize and compare microorganisms causing disease, thereby helping prevent and solve disease outbreaks
- (c) Analysis of microbial DNA in the near-environment, including our gut, enabling a new understanding of the importance of good gut microorganisms for human health status

In this somewhat restricted description of the usefulness of NGS relative to human health, it is important to remember that important uses of NGS relative to animal production and health, plant production and health, as well as a new understanding of environmental biotopes and interactions have been excluded. Also excluded are all the interesting areas related to forensic medicine (“who was the criminal perpetrator?”) and the exiting area of ancient human genetics,

understanding the evolutionary history of humans through DNA sequencing of (sometimes very old) fossils.

NGS and Human DNA: Prevention and Treatment of Diseases with a Focus on Older Adults

One of the important areas where NGS technology provides new knowledge is the characterization and diagnosis of genetic diseases. Even before NGS, it was possible to diagnose major genetic diseases/abnormalities, such as Down's syndrome, since these types of diseases stem from alterations at chromosome level, which means it can be diagnosed using simple light microscopy of basically any (dividing) human cell. Human cells carry the DNA in 23 pairs of rather large chromosomes, which are thus visible under the microscope when the cell is dividing. There are, however, a much larger number of genetic disorders that are not visible at chromosome level, and therefore, NGS technology will bring significant improvement in diagnostics. In addition, a totally new level of understanding of the functionality of human genes will be available, including whether specific DNA sequences can be the cause of disease. This is especially interesting in relation to elderly populations because the frequency of such diseases increases dramatically with age (e.g., cancer, cardiovascular diseases, etc.).

The passive knowledge of presence of a “disease gene” would not necessarily in all cases be considered a benefit for a human being. Some people might rather want to remain uninformed about potential gene defects. Therefore, it is very important to have clear rules for the testing – as well as for the fate of test results – and ongoing discussions, for instance about employers' right to ask for genetic testing is part of this debate. The WHO states: “Genetic tests give an assessment of an individual's inherent risk for disease and disability. This predictive power makes genetic testing particularly liable for misuse. Employers and insurance companies have been known to deny individuals essential health care or employment based on knowledge of genetic disposition” (WHO 2019).

It should be noted, however, that the diagnosis of certain genetic disorders through DNA sequencing can be considered a benefit for many if it can lead to considered medical advice about potential treatment and prevention options, including lifestyle changes such as diet and exercise that can reduce the risk of disease. Such advice now seems to be increasingly interesting in the elderly population as many of these diseases increase dramatically as a result of – or at least correlated to – the process of aging. These diseases include many cancers, cardiovascular disease, arthritis, and the whole complex of geriatric diseases now referred to as “frailty” (Berrut et al. 2013).

Because of our now available NGS-based methods, which provide continuously increasing knowledge about the functioning of our genes and the health effect of specific DNA sequences within genes, it is now also possible to utilize specific genes in the treatment of diseases; this is typically referred to as gene therapy. In the future, gene therapy may allow doctors to treat a disorder by inserting a gene into a patient's cells either to replace a mutated gene that causes disease with a healthy copy of the gene, or by inactivating a mutated gene that is functioning improperly (“knock-out”), or by introducing a new gene into the body to help fight a disease. Gene therapy could present a promising treatment for a number of diseases that are especially interesting in aging populations (including certain cancers and viral infections), but gene therapy is currently being tested only for diseases that have no other cures, such as sickle cell anemia (Hoban et al. 2016).

In the other end of the human lifespan, genetic testing can also be used by couples to investigate the genetic characteristics of their unborn child. Stemming from the informative potential of especially NGS-based, very accurate genetic testing, some critical ethical, legal, and social issues come to the fore. Notably, it is now under discussion whether parents should be allowed in the future to de-select certain specific traits in relation to the DNA composition of their baby through discontinuation of the pregnancy if such traits are

present, which in most societies at this stage is acceptable only for certain, very serious disorders.

Based on our knowledge of the DNA code from NGS and because of the development of a new, very precise gene-editing technology: CRISPR (clustered regularly interspaced short palindromic repeat), it is now theoretically possible to edit genetic errors out of the DNA of very early stages of fertilized human eggs (i.e., very early human embryos) (Tang et al. 2017a). And this is no longer only a theoretical potential; such editing has actually resulted in gene-edited babies born in China in November 2018. A young scientist from Southern University of Science and Technology in Guangdong, China, has through the use of CRISPR genome editing modified a gene to a form that supposedly would make the baby resistant to the HIV virus and included this gene in early stage human embryos in the end resulting in two babies harbouring the resistance gene. The work has not yet been published in a scientific journal (and might not be!), but has caused very significant uproar and debate in scientific - and other - circles as an example of gene-editing humans (Chemistry World 2019).

NGS and Prevention of Communicable Diseases in Aging Communities

Pathogens (microorganisms causing disease), including the now rapidly emerging antibiotic-resistant bacteria, are still a major concern as causes of communicable diseases (diseases caused by microorganisms and therefore communicable between humans) in all regions of the world. As recently exemplified by the spread of SARS and Avian influenza, it is clear that pathogens can be transmitted globally via human movement or via the food chain due to global trading (Holmes et al. 2016). The world needs surveillance and response systems that are capable of detecting these microorganisms rapidly, understanding them and responding to them (WHO 2018).

A number of communicable diseases are most dangerous for older adults (as well as for the immunocompromised). This is true to some degree for almost any pathogens, but especially *Listeria monocytogenes* (foodborne), and the

common seasonal flu virus results in significant excess mortality among older adults.

NGS and WGS provide a great potential to revolutionize how surveillance of communicable diseases and microorganisms are presently managed. “When used as part of a surveillance and response system, WGS has the power to increase the speed with which threats are detected and the detail in which the threats are understood, and ultimately lead to quicker and more targeted interventions” (WHO 2018, p. viii). WGS is already at this stage used in disease surveillance and outbreak investigations in a number of countries.

Likewise, and with special importance to aging populations, the isolation and characterization of the relevant pathogen will be possible much faster and more correctly with the use of WGS. Notably, it will be possible to select and produce the seasonal flu vaccine much faster than previously, resulting in a significant reduction in the flu mortality rate among older adults. Worldwide, seasonal flu is estimated to cause 300,000–600,000 deaths per year, with the majority of deaths in industrialized countries occurring among people aged 65 or older (Thompson et al. 2009).

Through WGS of microbial isolates, both pathogen identification and characterization, and detection of virulence factors and resistance genes can be directly obtained from the sequence data rapidly and at a level of precision that was not previously possible. In addition to better surveillance, these precise and rapid results leads to better, in some cases life-saving, treatment of serious disease. As opposed to traditional, diverse microbiological lab technology, WGS is exactly the same methodology irrespective of microorganism tested, enabling simpler, faster, and cheaper laboratory operations and opening the potential for one type of lab for the characterization and typing of all microorganisms (good/bad, human/animal, plant/environment, virus/bacteria, etc.) WGS also provides a common standardized language (the four letters!) that can be easily shared with and deposited to online international public databases, enabling future real-time global surveillance of pathogens and resistance.

In any communicable disease outbreak investigation, the pathogens must be linked to each

other through accurate typing – as well as to occurrence in the food, if the outbreak is foodborne. The ability of WGS technology to resolve an outbreak source was well demonstrated already during the 2010 Haiti cholera outbreak (Taboada et al. 2017), which is the most serious, recorded cholera epidemic in recent history. This outbreak is responsible for killing at least 8000 people and sickening over 600,000 individuals (CDC 2014). The successful WGS application was enabled through the rapid public release of sequenced genomes by researchers with historic *Vibrio cholerae* collections (Chin et al. 2011; Hendriksen et al. 2011; Reimer et al. 2011) and by US CDC from the Haitian outbreak (Taboada et al. 2017). This led to the documentation of a single-source introduction of the cholera strain from Nepal (Nepalese UN contingent in Haiti) into Haiti (Eppinger et al. 2014). In the area of foodborne outbreaks, most recently the largest ever *Listeria* outbreak, happening in South Africa over a period of more than a year, was basically only solved with the intervention of WGS typing of the strains involved (Allam et al. 2018). These outbreaks are in need of rapid solutions, and the use of WGS has clearly provided us with a much better tool, enabling much faster – and more correct – response (Ronholm et al. 2016; Sekse et al. 2017).

Globally, the most extensive and well-known WGS-based application for food safety management is the GenomeTrakr Network (Allard et al. 2016). Once genomic data of bacteria (e.g., *Salmonella*, *Listeria*, *E. coli*, *Campylobacter*), parasites, and viruses from US surveillance efforts are available, they are uploaded by GenomeTrakr into NCBI (National Center for Biological Information) (FDA 2018). There is also a global effort to create active DNA databases: the Global Microbial Identifier (GMI) suggests utilizing an international interactive system of DNA databases containing the full genomes of all investigated microorganisms in the world (GMI 2018). Notably, the GMI idea represents the notion of global inclusiveness to harness benefit from this novel technology for all mankind, society, and the environment. Since its inception in 2011, GMI has garnered increasing support to advance the debate

concerning the social, political, economic, ethical, and technological barriers to realizing the GMI vision. The GMI initiative now involves more than 300 scientists from more than 50 countries.

Antibiotic resistance in microorganisms is estimated to cause 700,000 deaths annually in the world (Rand Report 2014). The same report estimates that this figure will increase to ten million by 2050 (making it a larger killer than cancer, presently causing seven to eight million deaths annually), again with a skewed distribution towards older adults. When you characterize the WGS of any microorganism, you also get the resistance profile, since the relevant resistance genes are now basically known. This means that WGS surveillance in the future will enable oversight of communicable diseases but also of the rapidly increasing problem of antibiotic resistance (McDermott et al. 2018), hopefully leading to sensible, preventive action.

NGS and Positive Health Effects of Microorganisms: Focus on Changes in Gut Microorganisms in Older Adults

We seem to “always” (at least the last 150 years) have known that the microorganisms residing in our gut can be important for our health. Part of our general knowledge about the benefit of eating yogurt has historically been that it might help keep pathogens from invading the gut – a phenomenon also sometimes referred to as “colonization resistance” of the gut. However, there are very few (if any) scientific studies documenting this specific effect of beneficial food microorganisms. With the occurrence of NGS, we now have a tool to investigate all the microorganisms in our gut. It is estimated that there are more than 500 microorganism species in the gut, many of these we could actually not grow in the lab before and therefore not identify. Utilizing NGS we can now detect the DNA of all the gut microbial species, enabling a totally new line of research and understanding of this gut “microflora,” now referred to as the gut microbiota.

We therefore now know that the microorganisms in our gut have a number of previously unknown, but very specific effects on several parameters important for health and wellbeing.

For instance, the composition of the gut microbiota has been shown to play a pivotal role in the initiation and progression of a number of diseases associated with aging, including cardiovascular diseases (Tang et al. 2017b), atherosclerosis, and heart failure (Kitai et al. 2016; Organ et al. 2016).

Investigations with a specific focus on older adults have found that the diversity of microorganisms in the gut declines with age. This has led to mounting evidence showing that our gut microbiome influences our state of health and more specifically the inflammatory status of our body. Inflammatory factors are plausible causal factors for frailty in elderly people. This has prompted the hypothesis that gut microbiome manipulation could alter the inflammatory state of the individual and directly affect (meaning suppress) the development of frailty in older adults, leading to improved health. The detection of inflammatory markers in the gut might ultimately be used as intervention targets for encouraging healthy aging (Steves et al. 2016; Thomas et al. 2017; Vaiserman et al. 2017). This also most likely means that with a deeper understanding of these correlations, we might be able to construct a diet with “good” microorganisms preventing diseases of older adults.

An interesting example of how microorganisms in the gut can actually affect cardiovascular diseases seems to be in some way linking up with mechanisms for how exercise prevents heart disease. Both the composition of our gut microbiome and our level of physical exercise have been implicated in heart disease. A potential mechanism by which exercise might influence the gut microbiota involves changes in the microenvironment in which gut bacteria reside, following from changes in the intestinal blood perfusion affected by the level of exercise. These bacteria subsequently influence the immune system that again affects cardiac function (Cronin et al. 2018). In a way, you could say that exercise benefits the cardiovascular system through changes of the gut microbiome. So, the question becomes: Could we change our cardiovascular disease risk simply by changing our gut microbiota?

But it does not stop there: After initial studies found that you can change the average weight of

groups of mice by changing their gut microorganisms, a number of studies have looked further into this area. Both studies in animals and in humans have documented weight loss based on the formulation of bacteria used to populate the gut – typically through dietary interventions. In these studies, the utilization of dietary agents modulating the composition of gut microorganisms resulted in significant decreases in BMI (body mass index), weight, and fat mass (for a review, see John et al. 2018). These investigations have been rooted in the use of NGS to investigate the gut microbiome (microbiome = the collective genomes of all microorganisms in the environment in question). Basically, these types of investigations are now performed using metagenomic testing, i.e., separating all DNA in the sample and then sequencing everything, leaving the separation of all DNA pieces from the different microbial species to the computerized bioinformatics analysis (Cronin et al. 2018).

Future Directions of Research

There are many unsolved questions as per how the sequence in our DNA (the genotype) results in the way humans look (the phenotype). For things like eye color, hair color, height, weight, skin color the sequencing of DNA has now resulted in the description of specific genes (in most cases actually numerous different genes) that affect these traits. This somewhat confusing picture, even for what could be considered simple traits, also means that there is a long way to go before the book of life can be read in our DNA code. There is simply not enough knowledge yet to enable direct translation from genotype to phenotype. But this will be possible at some stage – and the start of this is the investigation of the DNA sequences of genomes – both for humans and for all other forms of life on this planet. Our future is in the genes (sic!).

The rapid development of technologies to sequence DNA and RNA is bringing down the cost of genomic sequencing and at the same time increasing the accuracy of the outcome. It is therefore likely that DNA sequencing capacity will

spread efficiently to generic routine labs, enabling a dramatic increase in the use of this technology for human genome analysis and for microbiological whole genome sequencing.

At the same time, some new developments in the field have resulted in the occurrence of DNA sequencing machines that are basically the size of mobile phones and could therefore be used even outside laboratory setting, potentially enabling sequencing investigations (especially of microorganisms) directly in the relevant environment, avoiding time-consuming transport of specimens to nearest lab (which can be very far away, e.g. during Ebola or other dangerous disease outbreaks).

The use of metagenomic testing, i.e., separating all DNA in the sample and then sequencing everything, leaving the separation of all DNA pieces from different species (human, animal, plant, microorganisms) to the computerized bioinformatics analysis is opening even more avenues of usefulness of this technology, both in relation to human DNA analysis, to food analysis (including food fraud analysis), and to microbiological analysis, relative to human, animal, and plant health and microbiology in general, including analysis of positive microorganisms in foods and/or in the human (or animal) gut and the environment.

The further development of active DNA databases as suggested by the Global Microbial Identifier (GMI) initiative represents a novel, open-source based global “machine” to harness benefit from this novel technology for all mankind, society, and the environment. The GMI vision of an open, active database system to characterize and type all microorganisms will become reality in the future, the real question is how long we have to wait, for the international community to debate the potential setup and initiate the work.

Summary

Next-generation DNA sequencing (NGS) technology is used to detect the sequence of the four

nucleotides constructing the genetic material of all life (A, C, G, and T), thus presenting the whole genome sequence (WGS) of any life form on Earth. The importance of achieving WGS results as it relates to human health can roughly be divided in three categories: Analysis of human DNA to characterize health status and risk of disease, analysis of microbiological DNA to characterize and combat diseases, and analysis of microbial DNA in the near-environment, including human and animal gut, enabling characterization of good gut microorganisms essential for human and animal health. The usefulness of NGS relative to, for instance, forensic medicine (“who was the criminal perpetrator?”), ancient evolutionary human genetics, or animal or plant genetic improvements is not included in this overview.

NGS technology provides new knowledge in the diagnosis of genetic diseases, which is especially interesting in relation to elderly populations because the frequency of such diseases increases with age. The detection of genetic disorders can lead to considered medical advice about potential treatment and prevention options. Gene therapy could be used to treat these types of diseases, but is currently being tested only for diseases that have no other cures. NGS and WGS provide a great potential to revolutionize the surveillance of communicable diseases and have the power to increase the speed with which threats are detected and combated. With NGS information, it will be possible to select and produce vaccines, such as the seasonal flu vaccine, much faster and more correctly than in the past. The use of NGS methodology has provided evidence that the microorganisms in human and animal gut systems have a number of previously unknown, but very specific, effects on health and wellbeing, including a number of diseases associated with aging, such as cardiovascular diseases, atherosclerosis, and heart failure.

NGS technologies are rapidly becoming cheaper, bringing these methods to generic routine labs, enabling a significant increase in human genome analysis and microbiological whole

genome sequencing. At the same time, novel DNA sequencing systems have dramatically reduced the size of sequencers down to the size of mobile phones, enabling analysis even outside laboratory setting.

Cross-References

- ▶ Chromosome
- ▶ DNA Sequencing
- ▶ Genotype
- ▶ Mutation

References

- Allam M et al (2018) Whole-genome sequences of *Listeria monocytogenes* sequence type 6 isolates associated with a large foodborne outbreak in South Africa, 2017 to 2018. *Genome Announc* 6(25):e-00538–e-00518. <https://doi.org/10.1128/genomeA.00538-18>
- Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R (2016) Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J Clin Microbiol* 54:1975–1983. <https://doi.org/10.1128/jcm.00081-16>
- Ayala FJ (2007) Darwin's greatest discovery: design without designer. *Proc Natl Acad Sci USA* 104(Suppl 1):8567–8573
- Berrut G et al (2013) Promoting access to innovation for frail old persons. *J Nutr Health Aging* 17(8):688–693
- Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E (2018) Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin Microbiol Infect* 24:335–341. <https://doi.org/10.1016/j.cmi.2017.10.013>
- Blattner FR et al (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1462
- CDC (2014) Cholera in Haiti. <https://www.cdc.gov/cholera/haiti/>. Accessed 8 Dec 2018
- Chemistry World (2019) News: Rogue Chinese scientist to be punished for creation of first gene-edited humans. <https://www.chemistryworld.com/news/rogue-chinese-scientist-to-be-punished-for-creation-of-first-gene-edited-humans/3010026.article>
- Chin CS et al (2011) The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364:33–42. <https://doi.org/10.1056/NEJMoa1012928>
- Cole ST et al (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537–544. <https://doi.org/10.1038/311159>
- Cronin O et al (2018) A prospective metagenomic and metabolomic analysis of the impact of exercise and/or whey protein supplementation on the gut microbiome of sedentary adults. *mSystems* 3:e00044–e00018. <https://doi.org/10.1128/mSystems.00044-18>
- Dawkins R (2006 (1986)) The blind watchmaker. Penguin Random House, Oxford. ISBN: 978-0-141-02616-9
- Eppinger M et al (2014) Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *MBio* 5:e01721. <https://doi.org/10.1128/mBio.01721-14>
- FDA (2018) GenomeTrakr network. <https://www.fda.gov/food/foodscienceresearch/wholegenomesequencingprogramwgs/ucm363134.htm>. Accessed 19 Mar 2019
- Fleischmann RD et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science* 269(5223):496–512
- Fraser CM et al (1998) Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* 281:375–388
- GMI (2018) Global microbial identifier. <http://www.globalmicrobialidentifier.org/>. Accessed 9 Dec 2018
- Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17:333–351. <https://doi.org/10.1038/nrg.2016.49>
- Gutiérrez-Preciado A, Romero H, Peimbert M (2010) An evolutionary perspective on amino acids. *Nat Educ* 3(9):29
- Hendriksen RS et al (2011) Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *MBio* 2:e00157–e00111. <https://doi.org/10.1128/mBio.00157-11>
- Hoban MD, Orkin SH, Bauer DE (2016) Genetic treatment of a molecular disorder: gene therapy approaches to sickle cell disease. *Blood* 127:839–848. <https://doi.org/10.1182/blood-2015-09-618587>
- Holmes AH et al (2016) Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* 387:176–187. [https://doi.org/10.1016/s0140-6736\(15\)00473-0](https://doi.org/10.1016/s0140-6736(15)00473-0)
- IHGSC (International Human Genome Sequencing Consortium) (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- John GK, Wang L, Nanavati J, Twose C, Singh R, Mullin G (2018) Dietary alteration of the gut microbiome and its impact on weight and fat mass: a systematic review and meta-analysis. *Genes* 9:167. <https://doi.org/10.3390/genes9030167>
- Kitai T, Kirsop J, Tang WH (2016) Exploring the microbiome in heart failure. *Curr Heart Fail Rep* 13:103–109. <https://doi.org/10.1007/s11897-016-0285-9>
- Kunst F et al (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249–256. <https://doi.org/10.1038/36786>
- McDermott PF, Zhao S, Tate H (2018) Antimicrobial resistance in nontyphoidal *Salmonella*. *Microbiol Spectr* 6.

- <https://doi.org/10.1128/microbiolspec.ARBA-0014-2017>
- Organ CL, Otsuka H, Bhushan S, Wang Z, Bradley J, Trivedi R et al (2016) Choline diet and its gut microbe-derived metabolite, trimethylamine N-oxide, exacerbate pressure overload-induced heart failure. *Circ Heart Fail* 9:e2314. <https://doi.org/10.1161/CIRCHEARTFAILURE.115.002314>
- Parkhill J et al (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413:523–527. <https://doi.org/10.1038/35097083>
- Rand Report (2014) The Wellcome Trust: a study, commissioned in the framework of the independent review led by Jim O'Neill, Tackling drug-resistant infections globally: final report and recommendations—the review on antimicrobial resistance. 2016. Wellcome Trust, UK Government. https://amr-review.org/sites/default/files/160525_Final%20paper_with%20cover.pdf Accessed 31 Oct 2019
- Reimer AR et al (2011) Comparative genomics of *Vibrio cholerae* from Haiti, Asia, and Africa. *Emerg Infect Dis* 17:2113–2121. <https://doi.org/10.3201/eid1711.110794>
- Ronholm J, Nasheri N, Petronella N, Pagotto F (2016) Navigating microbiological food safety in the era of whole-genome sequencing. *Clin Microbiol Rev* 29:837–857. <https://doi.org/10.1128/cmr.00056-16>
- Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchinson CA, Slocombe PM, Smith M (1977a) Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature* 265(5596):687–695
- Sanger F, Nicklen S, Coulson AR (1977b) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74(12):5463–5467
- Sekse C, Holst-Jensen A, Dobrindt U, Johannessen GS, Li W, Spilberg B, Shi J (2017) High throughput sequencing for detection of foodborne pathogens. *Front Microbiol* 8:2029–2029. <https://doi.org/10.3389/fmicb.2017.02029>
- Steves CJ et al (2016) The microbiome and musculoskeletal conditions of aging: a review of evidence for impact and potential therapeutics. *J Bone Miner Res* 31(2):261–269
- Taboada EN, Graham MR, Carriço JA, Van Domselaar G (2017) Food safety in the age of next generation sequencing, bioinformatics, and open data access. *Front Microbiol* 8:909–909. <https://doi.org/10.3389/fmicb.2017.00909>
- Tang L et al (2017a) CRISPR/Cas9-mediated gene editing in human zygotes using Cas9 protein. *Mol Gen Genomics* 292(3):525–533. <https://doi.org/10.1007/s00438-017-1299-z>
- Tang WH, Kitai T, Hazen SL (2017b) Gut microbiota in cardiovascular health and disease. *Circ Res* 120:1183–1196. <https://doi.org/10.1161/CIRCRESAHA.117.309715>
- Thomas S et al (2017) The host microbiome regulates and maintains human health: a primer and perspective for non-microbiologists. *Cancer Res* 77(8):1783–1812
- Thompson WW et al (2009) Estimates of US influenza-associated deaths made using four different methods. *Influenza Other Respir Viruses* 3(1):37–49
- Vaiserman AM, Koliada AK, Marotta F (2017) Gut microbiota: a player in aging and a target for anti-aging intervention. *Ageing Res Rev* 35:36–45
- Watson JD, Crick FHC (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171:737–738
- WHO (2018) Whole genome sequencing for foodborne disease surveillance: landscape paper. WHO, Geneva. <http://www.who.int/iris/handle/10665/272430>. License: CC BY-NC-SA 3.0 IGO
- WHO (2019) Human genomics in global health. <https://www.who.int/genomics/elsi/gentesting/en/>. Accessed 18 Mar 2019