Genetica (2015) 143:129–132
DOI 10.1007/s10709-015-9831-8

Next-generation sequencing as a powerful motor for advances in the biological and environmental sciences

Denis Faure · Dominique Joly

Received: 20 February 2015 / Accepted: 23 February 2015 / Published online: 4 March 2015

Abstract Next-generation sequencing (NGS) provides unprecedented insight into (meta)genomes, (meta)transcriptomes (cDNA) and (meta)barcodes of individuals, populations and communities of Archaea, Bacteria and Eukarya, as well as viruses. This special issue combines reviews and original papers reporting technical and scientific advances in genomics and transcriptomics of non-model species, as well as quantification and functional analyses of biodiversity using NGS technologies of the second and third generations. In addition, certain papers also exemplify the transition from Sanger to NGS barcodes in molecular taxonomy.

Keywords Next-generation sequencing (NGS) · Rad-seq · Illumina · PacBio · 454-Pyrosequencing

Next-generation sequencing (NGS) provides unprecedented insight into (meta)genomes, (meta)transcriptomes (cDNA) and (meta)barcodes of individuals, populations and communities of Archaea, Bacteria and Eukarya, as well as viruses. For over a decade, NGS techniques have irreversibly changed the methods for DNA sequencing, as well as the exchange and storage of the enormous quantities of resulting sequence data. This has opened avenues for understanding life on Earth: especially concerning the exploration and description of the biodiversity and taxonomy of organisms and viruses, their evolution and adaptation under changing and challenging environments, as well as their ecology in present and past ecosystems.

In the late 1970's, the first generation of DNA sequencing was born. The more used method being that working with dideoxy chain-termination developed by Frederick Sanger (Sanger et al. 1977). This manual technology was improved until the beginning of the 2000's, when the second generation (now automatic) of DNA-sequencers (Solexa by Illumina, pyrosequencing by 454/Roche and Solid by Life-Technologies) came into use (see the commentary by Graveley 2008; and a comparison by Loman et al. 2012). More recently, other automated DNA sequencers, such as MiSeq by Illumina, Ion Torrent by Life Technology and Pacific Bioscience (PacBio) machines, have appeared in the competitive NGS arena. Noticeably, PacBio uses single molecule real time sequencing by synthesis. This technology belongs to the third generation of DNA-sequencers and promises DNA-reads of 20 kbp or more, with average read lengths of 5 kbp. Other technologies (Heliscope single molecule sequencing by Helicos Biosciences, DNA nanoball sequencing by Complete Genomics, Nanopore by Oxford Nanopore Technologies…) have been proposed and are under development.

NGS allows a massive production of DNA sequences with a cost which has decreased by five orders of magnitude in one decade (Fig. 1). Today, the production cost of DNA sequences has decreased faster than that of the storage or computing, which makes their systematic

D. Faure (✉) · D. Joly (✉)
GDR3692 Génomique Environnementale, CNRS, Université Paris-Sud, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France
e-mail: denis.faure@isv.cnrs-gif.fr

D. Joly
e-mail: dominique.joly@legs.cnrs-gif.fr

D. Faure
UMR9198, CNRS, CEA, Institut for Integrative Biology of the Cell, Saclay Plant Sciences, Université Paris-Sud, bat 23, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

D. Joly
Laboratoire Evolution, Génomes, Comportement, Ecologie, UMR9191, CNRS-IRD-Paris-Sud, bat 13, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

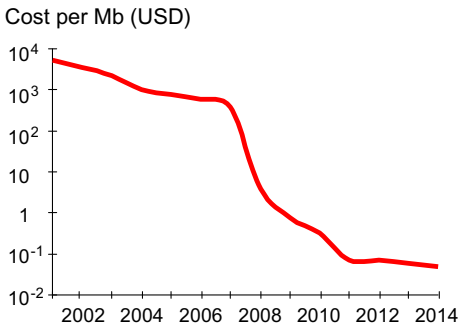

Fig. 1 Decrease of DNA-sequencing costs over the past decade. Data from 2001 through October 2007 represent the costs of generating DNA sequence using Sanger-based chemistries and capillary-based instruments (first generation sequencing platforms). Beginning in January 2008, the *data* represent the costs of generating DNA sequence using second-generation sequencing platforms. Cost calculation and updated excel table of the data are available at the National Human Genome Research Institute (NHGRI) of USA (www.genome.gov/sequencingcosts)

backup somewhat problematic and should stimulate the development of new processes (including cloud computing) for the storage, exchange and analysis of NGS data (Stein 2010; the commentary by Baker 2010). NGS tools were rapidly adopted by researchers, and exponentially propagated in scientific articles (Fig. 2). NGS development is also observed in the public databases of the International Nucleotide Sequence Database Collaboration (INSDC http://www.insdc.org/) which comprises the DNA DataBank of Japan (DDBJ http://www.ddbj.nig.ac.jp/), the European Molecular Biology Laboratory (EMBL http://www.embl.org/), and GenBank at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genbank/). In addition to the standard submission of nucleotide sequences, the abundance and format of the NGS-derived sequences provoked the creation of additional submission categories, such as whole genome shotgun (WGS) projects, high throughput genomic (HTG) and metagenomic projects. The number of the released base-pairs (bp) and sequences belonging to these novel categories is growing more rapidly than that of standard submissions (Fig. 3).

Massive production of NGS sequences is supported by the private and public sequencing centers and local platforms which constantly renew their DNA-sequencer facilities. Sequencing centers propose a large panel of NGS technologies, with a dominant position of the Illumina machines (Fig. 4). To fuel their increasing sequencing capacity, the sequencing centers have announced ambitious projects. The Beijing Institute of Genomics (BGI; http://

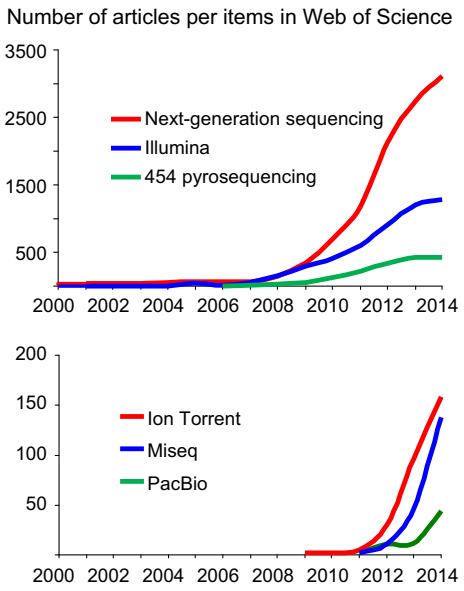

Fig. 2 Emergence of NGS tools in scientific production. Bibliometric data were collected at Web of Science (http://thomsonreuters.com/thomson-reuters-web-of-science/) using the topics Next-generation sequencing or Illumina or 454 pyrosequencing (*upper graph*) and PacBio or Ion Torrent or Miseq (*lower graph*) over the years 2000–2014. Note that this analysis underestimates the number of retrieved papers, as only the items in the selected topics were counted

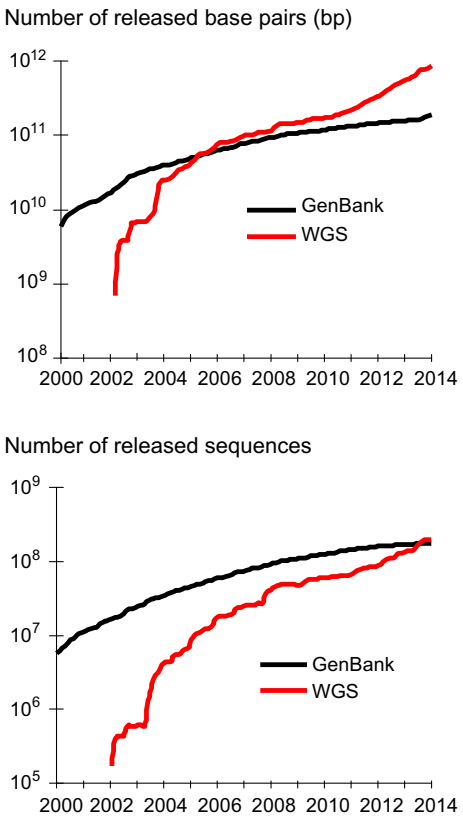

Fig. 3 Release of NGS sequences by public databases. The *graphs* indicate the number of released base pairs (*upper graph*) and sequences (*lower graph*) of the standard GenBank and WGS data over the years 2000–2014. The data are available at http://www.ncbi.nlm.nih.gov/genbank/statistics

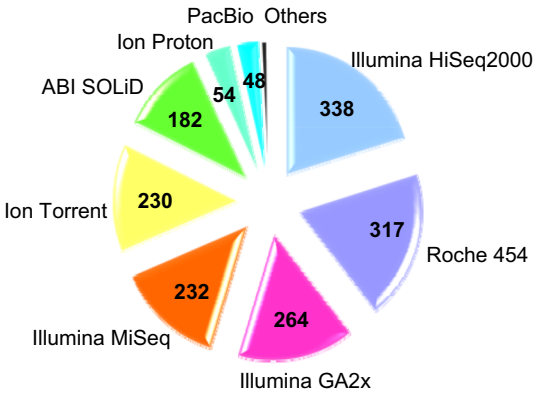

Fig. 4 Sequencing centers holding the different NGS machines in the world. The *figure* indicates the number of sequencing centers with the indicated NGS-machine types. There are 7,389 total machines listed in the databases, situated in 1,027 centres. The data are available at the site http://omicsmaps.com/stats

english.big.cas.cn) as one of the largest DNA-sequencing centers, launched sequencing programs of several thousand genomes and metagenomes (Normile 2012). Other centers, such as the Joint Genome Institute (JGI; http://jgi.doe.gov/) and France Génomique (https://www.france-genomique.org/spip/) also developed their own research programs and competitive calls. Aside from sequencing centers and platforms, many laboratories can now afford to buy and use bench sequencing machines, such as MiSeq, GS Junior and Ion Torrent. Producing sequences is now becoming easier (cost and time) than analyzing and publishing them.

Many highlighted works based on the analyses of (meta)genomes, (meta)transcriptomes and (meta)barcodes have explored and exemplified the strength of NGS in biological and environmental sciences. The question arose as to how to facilitate access to NGS for all community members. A biennial symposium concerned with Environmental Genomics was launched by a French national network on Environmental Genomics. This network was called "*RTP Génomique Environnementale*" between 2011 and 2013 and now is called "*GDR3692 Génomique Environnementale*". This meeting was held in Lyon (November 2011), Rennes (November 2013) and Montpellier (October 2015) and is accessible to scientists from all other countries. The Environmental Genomics symposium assembles communities which share NGS approaches in the fields of taxonomy, ecology, evolution, ecotoxicology and paleobiology to provide a better comprehension of socio-ecological systems. It constitutes a forum for updating NGS tools and use, for presenting NGS-based advances, as well as for stimulating interdisciplinary exchanges between the different NGS-using communities.

This special issue reflects some aspects discussed at the Environmental Genomics symposia. It combines reviews and original papers reporting technical and scientific advances in genomics and transcriptomics of non-model species, as well as quantification and functional analysis of biodiversity using second and third generations NGS technologies. Certain papers also exemplify the transition from Sanger to NGS barcodes in molecular taxonomy.

In the review entitled "Next Generation Sequencing for characterizing biodiversity: Promises and challenges", Pompanon and Samadi (2015) discuss the advantages, constraints and opportunities in the combination of barcoding NGS for the purposes of a deep, fine and quantitative characterization of the taxons in environmental samples.

RAD-seq is another approach which has been popularized by NGS for analyzing population genomics in taxonomy, evolution and biodiversity studies. This reduced representation genomics approach offers the possibility to produce genome wide data from potentially any species, without previous genomic information. In their review entitled "Cost and robustness optimization for highly multiplexed RADseq libraries", Henri et al. (2015) discussed how to reduce experimentation RAD-seq costs by multiplexing hundreds of specimens.

NGS allows the development of another approach in reduced representation genomics: the high-density, genome-wide micro-arrays for the interrogation of genetic variation. Gurgul et al. (2015) used the BovineSNP50 BeadChip which features thousands SNP probes, allowing comparative genetic studies in cattle.

An emerging NGS-based approach concerns metatranscriptomics, which is especially developed for the functional characterization of microbial communities. In a review entitled "Technical challenges in metatranscriptomic studies applied to the bacterial communities of freshwater ecosystems", Pascault et al. (2014) analyzed and discussed in detail each of the transcriptomics steps with regard to the choices and difficulties encountered and to the recent literature.

The three next experimental papers by Bellanger et al. (2015), Aznar-Cormano et al. (2015) and Robuchon et al. (2014) illustrate the efforts of deep completion and analyses of barcodes using Sanger-sequencing. The collected data will be helpful for proposing appropriate NGS-barcodes according to the recommendations of Pompanon and Samadi (2015). These publications deal with taxonomy and diversity analyses of the Basidiomycota *Lyophyllaceae* (Bellanger et al. 2015), Caridean decapods (Aznar-Cormano et al. 2015) and red seaweeds (Rhodophyta) of *Laminaria* (Robuchon et al. 2014).

One of the strengths of NGS is to facilitate comparative and functional genomics in non-model species. The following papers exemplifying three different approaches: *de novo* transcriptome assembly of the blood-sucking bug *Triatoma brasiliensis* by combining 454-Roche and

Illumina HiSeq use (Marchant et al. 2014), *de novo* genome assembly of the plant-pathogen bacterium *Pectobacterium wasabiae* by Illumina HiSeq sequencing of a long distance mate-pair library and a short fragment paired-end library (Khayi et al. 2014), and *de novo* genome assembly of a soil bacterium *Rhodococcus erythropolis* using the Pacific Biosciences (PacBio) DNA-sequencer belonging to the third generation NGS (Kwasiborski et al. 2015).

NGS may be associated with other techniques for developing novel tools, such as single cell genomics. In integrative approaches, NGS may be also combined with other high-throughput technologies, such a (meta)proteomics, metabolomics and screening of functional traits in organisms and holobionts for investigating the complexity and dynamics of living cells and organisms. The NGS era is now really taking off.

Acknowledgements The authors thank Pr. M. DuBow (Université Paris-Sud, Orsay) for critical reading of the manuscript, as well as CNRS-InEE for supporting the RTP Génomique Environnementale (2011–2013) and with INRA the GDR3692 Génomique Environnementale (2015–2018).

References

Aznar-Cormano L, Brisset J, Chan TY, Corbari L, Puillandre N, Utge J, Zbinden M, Zuccon D, Samadi S (2015) An improved taxonomic sampling is a necessary 1 but not sufficient condition for resolving inter-families relationships in Caridean decapods. Genetica. doi:10.1007/s10709-014-9807-0

Baker M (2010) Next-generation sequencing: adjusting to data overload. Nat Methods 7:495–499

Bellanger JM, Moreau PA, Corriol G, Bidaud A, Chalange R, Dudova Z, Richard F (2015) Plunging hands into the mushroom jar: a phylogenetic framework for *Lyophyllaceae* (Agaricales, *Basidiomycota*). Genetica. doi:10.1007/s10709-015-9823-8

Graveley BR (2008) Molecular biology: power sequencing. Nature 453:1197–1198

Gurgul A, Jasielczuk I, Szmatoła T, Pawlina K, Ząbek T, Żukowski K, Bugno-Poniewierska M (2015) Genome-wide characteristics of copy number variation in Polish Holstein and Polish Red cattle using SNP genotyping assay. Genetica. doi:10.1007/s10709-015-9822-9

Henri H, Cariou M, Terraz G, Martinez S, el Filali A, Veyssiere M, Duret L, Charlat S (2015) Optimization of multiplexed RADseq libraries using low-cost adaptors. Genetica. doi:10.1007/s10709-015-9828-3

Khayi S, Raoul des Essarts Y, Quêtu-Laurent A, Moumni M, Hélias V, Faure D (2014) Genomic overview of the phytopathogen *Pectobacterium wasabiae* strain RNS2 08.42.1A suggests horizontal acquisition of quorum-sensing genes. Genetica. doi:10.1007/s10709-014-9793-2

Kwasiborski A, Mondy S, Chong TM, Chan KG, Beury-Cirou A, Faure D (2015) Core genome and plasmidome of the quorum-quenching bacterium *Rhodococcus erythropolis*. Genetica. doi:10.1007/s10709-015-9827-4

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30:434–439

Marchant A, Mougel1 F, Almeida C, Jacquin-Joly E, Costa J, Harry M (2014) De novo transcriptome assembly for a non-model species, the blood-sucking bug *Triatoma brasiliensis*, a vector of Chagas disease. Genetica. doi:10.1007/s10709-014-9790-5

Normile D (2012) China's sequencing powerhouse comes of age. Science 335:516–519

Pascault N, Loux V, Derozier S, Martin V, Debroas D, Maloufi S, Humbert JF, Leloup J (2014) Technical challenges in metatranscriptomic studies applied to the bacterial communities of freshwater ecosystems. Genetica. doi:10.1007/s10709-014-9783-4

Pompanon F, Samadi S (2015) Next generation sequencing for characterizing biodiversity: promises and challenges. Genetica. doi:10.1007/s10709-015-9816-7

Robuchon M, Valero M, Gey D, Le Gall L (2014) How does molecular-assisted identification affect our estimation of α, β and γ biodiversity? An example from understory red seaweeds (Rhodophyta) of *Laminaria* kelp forests in Brittany, France. Genetica. doi:10.1007/s10709-014-9796-z

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74(12):5463–5467

Stein LD (2010) The case for cloud computing in genome informatics. Genome Biol 11:207