



**HAL**  
open science

# Next-generation sequencing: a revolution in the field of fish diseases

Jean-Christophe Avarre

► **To cite this version:**

Jean-Christophe Avarre. Next-generation sequencing: a revolution in the field of fish diseases. Bulletin of the European Association of Fish Pathologists, 2020, 40 (2), pp.62-69. hal-02922349

**HAL Id: hal-02922349**

**<https://hal.science/hal-02922349>**

Submitted on 26 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Next-generation sequencing: a revolution in the field of fish diseases

Jean-Christophe Avarre<sup>1</sup>

<sup>1</sup> *ISEM, CNRS, IRD, EPHE, University of Montpellier, Montpellier, France*

Institut des Sciences de l'Evolution de Montpellier  
University of Montpellier, Place Eugène Bataillon, cc065  
34095 Montpellier cedex 5, France  
[Jean-christophe.avarre@ird.fr](mailto:Jean-christophe.avarre@ird.fr)

### Abstract

The recent technological advances in nucleic acid sequencing, called next-generation sequencing (NGS), have revolutionized the field of genomics. Now available in almost all molecular biology labs, NGS provides a huge amount of sequence data at relatively low cost. Research on fish pathogens has greatly benefited from these new technologies and NGS is now increasingly used to trace aquatic viruses, study virulence and evolution of pathogens or discover novel etiological agents that cause mortalities. This short review presents a brief history of the different sequencing technologies and illustrates their applications and limitations in the field of aquatic diseases.

### 1. Introduction

Limited to less than 4 million tons in 1970, global aquaculture production has grown exponentially since the early 80s, to reach 80 million tons in 2016 (FAO, 2018). It now accounts for more than half of the aquatic production strictly dedicated to human consumption, with a value estimated at \$230 billion. Given the increase in demand linked to population growth, a 37% increase in aquaculture production is expected by 2030 (FAO, 2018). Currently, the typical response to this increased food demand is the intensification of production, which is likely to favour disease development (Jansen et al., 2012; Rodger, 2016). For instance, intensification of aquaculture associated with increased global movements of live aquatic animals has been a major driver of the emergence of many viral

diseases (Walker & Winton, 2010). As a result, diseases now constitute a major threat for the sustainability of this ever-growing global industry.

The extent of disease spread and impacts is greatly affected by the ability to quickly identify the causative agents and understand critical epidemiological factors such as their mode of transmission and virulence mechanisms etc. (Walker & Winton, 2010). Next-generation sequencing (NGS) has opened many new avenues of research in the field of fish and shellfish diseases. It does not require *a priori* knowledge of pathogen sequences, thereby unlocking access to a wider world of microorganisms, including those that cannot be cultured, or those that persist in their environment before (re)emerging. However, although it is now well acknowledged that pathogen surveillance will greatly benefit from genomics, it is not without its challenges (Avarre, 2017; Stärk et al., 2019).

This short review will present a brief overview of the different sequencing approaches and their technological evolution, and illustrate their applications and limitations in the field of aquatic diseases, *e.g.* for identifying pathogens, studying microorganism communities or deciphering virulence mechanisms.

## **2. Evolution of sequencing technologies**

The first two DNA sequencing technologies were simultaneously described in 1977 by Maxam and Gilbert (Maxam & Gilbert, 1977) and Sanger et al. (Sanger et al., 1977). The method described by Maxam and Gilbert prevailed until the early 1980s but Sanger et al.'s method quickly replaced it, especially after the description of fluorescent dyes and the commercial launch of automated capillary sequencers (Ansorge et al., 1986; Smith et al., 1986). Today, these two methods belong to the first-generation of sequencing, whereas next-generation sequencing (NGS) refers to all other sequencing technologies, including second, third and even fourth generations (Ke et al., 2016). Though much less acknowledged, another first-generation technology, based on the luminometric detection of inorganic pyrophosphate (Nyren et al., 1993), later called pyrosequencing (Ronaghi et al., 1998), paved the way for the first second-generation sequencing technology (Margulies et al., 2005).

In addition to unprecedented parallelization, the second-generation sequencing methods introduced two major technical modifications: (i) instead of requiring bacterial cloning of DNA fragments, they rely on the *in vitro* preparation of DNA libraries through the addition of

adaptor sequences before or during clonal amplification; (ii) base interrogation is achieved through a sequential “wash-and-scan” procedure (Figure 1a,b). Varying platforms have been commercialized by several manufacturers, all with different characteristics providing different read lengths (50-700) and numbers (millions to billions) (reviewed in Buermans & den Dunnen (2014); Goodwin et al. (2016)). These improvements have also led to impressive gains in cost, speed and throughput. For example, the latest Illumina sequencer (NovaSeq 6000) can sequence up to 6000 gigabases (Gb) in less than 2 days.

Third-generation sequencing instruments brought two major additional improvements: the ability to sequence single molecules and the use of real time (Schadt et al., 2010; Heather & Chain, 2016; van Dijk et al., 2018). The most notable consequences are that i) they do not rely on a clonal amplification of DNA fragments to reach a detectable signal, therefore circumventing all known PCR biases (Krebschull & Zador, 2015); ii) they enable sequencing of very long DNA fragments (Figure 1c). The major drawback, however, is the high sequencing error rate (3-15%) compared to the second generation sequencing platforms (0.1-1%) (Goodwin et al., 2016; Ameer et al., 2019).

### **3. Successful use of second-generation sequencing for pathogen detection and characterisation**

Several heart diseases encountered in farmed Atlantic salmon, such as cardiomyopathy syndrome (CMS) or heart and skeletal muscle inflammation (HSMI), have long remained difficult to diagnose due to unknown aetiology. For example, CMS was first described in Norway in 1990 and spread to neighbouring countries in subsequent years with undetermined aetiology. A *Totivirus*, the piscine myocarditis virus (PMCV) was eventually identified by NGS in 2010 and confirmed as the infectious agent one year later (Løvoll et al., 2010; Haugland et al., 2011). HSMI was first described in Norway in 1999. It remained unexplained but spread to hundreds of farms in Norway and in the United Kingdom. Despite numerous studies, its causative agent was only identified in 2010 as a novel reovirus called piscine reovirus (PRV) with the use of NGS (Palacios et al., 2010). More recently, full genome sequencing of additional PRV strains enabled i) to unequivocally demonstrate the causal relationship of this *Piscine Orthoreovirus-1* (PRV-1) with HSMI (Wessel et al., 2017), and ii) to gain insights into the evolution of this virus by showing two main monophylogenetic clusters, with only one associated with HSMI (Dhamotharan et al., 2019).

Although there are many other examples in which NGS contributed to the identification of causative agents, there have also been cases where they failed to identify the right agent. An example of this is proliferative darkening syndrome (PDS), of unknown origin, which has been causing high mortalities in brown trout populations from pre-alpine rivers in Austria, Southern Germany and Switzerland for nearly 15 years (Lahnsteiner et al., 2011). In an attempt to identify a possible etiological agent with NGS, Kuehn et al. reported a piscine reovirus as the likely causative organism of PDS (Kuehn et al., 2018). However, they could not fulfil Koch's postulate and a same high throughput RNA sequencing approach applied on additional samples collected from both naïve and affected trout refuted the implication of PRV in the aetiology of PDS (Fux et al., 2019). Consequently, at the time of writing, the causative agent of PDS is still unknown. This case highlights the necessity of verifying sequencing results with other more conventional and/or quantitative methods. In a recent study on the *in vitro* evolution of *Cyprinid herpesvirus 3* (CyHV-3), Klafak et al. reported a 1.3-kb deletion in the genome of an isolate passaged 78 times compared to the wild type. This was evidenced by the absence of sequencing reads in the region concerned, suggesting that 100% of haplotypes in the passaged sample carried this deletion. Surprisingly, this deletion was not retrieved in the same isolate after 21 additional passages. Since a 'reverse mutation' of the same length and at the same location is unlikely, NGS results were verified by digital PCR, a highly sensitive PCR technology that enables quantitative measurement of low-copy variants. Results indicated that the non-deleted haplotype was in fact present after 78 passages, but in too low a proportion to be detected by sequencing, as it only represented 0.054% of the viral population (Klafack et al., 2019). This example again illustrates that second-generation sequencing technologies, which rely on clonal amplification (Figure 1), are prone to sequence representation biases, especially for sequences that are present at very low copy numbers (Krebschull & Zador, 2015; Illingworth et al., 2017).

Since first being reported in China in 2009, acute hepatopancreatic necrosis disease (AHPND), initially called early mortality syndrome (EMS), has spread through many countries. A great deal of research has been dedicated to understanding its origin due to its high societal and economic impacts. In 2013, the causative agent of AHPND was identified as a bacterial strain belonging to the *Vibrio parahaemolyticus* species, by fulfilling Koch's postulate (Tran et al., 2013). The mechanism by which this strain had become virulent was

resolved two years later, following the full genome comparison of several AHPND-causing and non-AHPND *V. parahaemolyticus* strains: the virulent strain had acquired a 70-kb plasmid that encodes homologs of the *Photobacterium* insect-related (Pir) toxins PirA and PirB, which induce host cell death (Lee et al., 2015).

Finally, second-generation sequencing is increasingly proposed as a routine tool for studying microorganism communities from environmental DNA (eDNA), directly collected from the water rather than from aquatic animals. eDNA enables the full diversity of microorganisms present in the aquatic environment to be characterised, including uncultivable pathogens (Nkili-Meyong et al., 2016). It is thus a promising tool for disease control, *e.g.* for the early identification of emerging pathogens or for studying the response of microbial communities following anthropogenic pressures or environmental changes (Munang'andu, 2016; Peters et al., 2018). Metabarcoding, which consists of sequencing a target amplicon common to many microorganism taxa (such as 16S ribosomal DNA), allows quick access to the taxonomic diversity of several samples simultaneously, thanks to the wealth of available databases. For example, this approach was recently used to assess the effect of probiotics on relative abundance of potential pathogenic bacteria in an oyster hatchery (Stevick et al., 2019). Although it is suitable for analysing bacterial communities, metabarcoding cannot be applied to viral communities because i) viruses lack a gene (or any other marker sequence) common to most viral taxa and ii) there are currently no suitable databases that reflect the actual viral diversity (Munang'andu, 2016; Nkili-Meyong et al., 2016). For this reason, metagenomics is now increasingly used in the molecular epidemiology of viral diseases (Munang'andu, 2016). Though more challenging than metabarcoding, since it sequences all DNA present in a sample, it allows both taxonomic and functional diversities to be evaluated (Nkili-Meyong et al., 2016; Nho et al., 2018).

#### **4- Conclusions and future trends in third generation sequencing**

If second-generation sequencing has now become a common - though still expensive - tool in the molecular epidemiology of aquatic pathogens (Bayliss et al., 2017), the third generation is gaining increasing interest. Indeed, in addition to its benefits described in section 2 and figure 1, it also offers practical advantages over second-generation sequencing, such as rapidity and portability. Combining second-generation (Ion Torrent) and third-generation (Oxford Nanopore) technologies, Beaton et al. succeeded in quickly

assembling the nearly full-length genome of an aquaculture strain of *Pseudomonas baetica*, resulting in a limited number of contigs (Beaton et al., 2018). Combined with phenotypic measures, analysis of its genome revealed a high level of plasticity, such as the ability to secrete proteases, higher tolerance to oxytetracycline or better adaptation to marine environments compared to other pseudomonads like *P. fluorescens*. In another attempt to deploy third-generation sequencing for rapid diagnostics of aquatic pathogens, Gallagher et al. showed it was possible to recover full genome sequences of two salmonid viruses from multiple samples within a couple of hours (Gallagher et al., 2018). In 2020, even though this appears very promising in terms of pathogen diagnostics, the high error rate associated with these technologies continues to make them less appropriate for characterising population variation within samples or understanding the complex evolution of rapidly mutating viruses.

## References

- Ameur, A, Kloosterman, WP and Hestand, MS (2019). Single-Molecule Sequencing: Towards Clinical Applications. *Trends in Biotechnology* **37**, 72-85.
- Anson, W, Sproat, BS, Stegemann, J and Schwager, C (1986). A non-radioactive automated method for DNA sequence determination. *Journal of Biochemical and Biophysical Methods* **13**, 315-323.
- Avarre, J-C (2017). Editorial: Molecular Tracing of Aquatic Viruses: Where Epidemiology Needs to Meet Genomics. *Frontiers in Microbiology* **8**, 1498.
- Bayliss, SC, Verner-Jeffreys, DW, Bartie, KL, Aanensen, DM, Sheppard, SK, Adams, A and Feil, EJ (2017). The Promise of Whole Genome Pathogen Sequencing for the Molecular Epidemiology of Emerging Aquaculture Pathogens. *Frontiers in Microbiology* **8**.
- Beaton, A, Lood, C, Cunningham-Oakes, E, Macfadyen, A, Mullins, AJ, Bestawy, WE, Botelho, J, Chevalier, S, Coleman, S, Dalzell, C, Dolan, SK, Faccenda, A, Ghequire, MGK, Higgins, S, Kutschera, A, Murray, J, Redway, M, Salih, T, Da Silva, AC, Smith, BA, Smits, N, Thomson, R, Woodcock, S, Welch, M, Cornelis, P, Lavigne, R, Van Noort, V and Tucker, NP (2018). Community-led comparative genomic and phenotypic analysis of the aquaculture pathogen *Pseudomonas baetica* a390T sequenced by Ion semiconductor and Nanopore technologies. *FEMS Microbiology Letters* **365**.
- Buermans, HPJ and Den Dunnen, JT (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1842**, 1932-1941.
- Dhamotharan, K, Tengs, T, Wessel, Ø, Braaen, S, Nyman, IB, Hansen, EF, Christiansen, DH, Dahle, MK, Rimstad, E and Markussen, T (2019). Evolution of the Piscine orthoreovirus Genome Linked to Emergence of Heart and Skeletal Muscle Inflammation in Farmed Atlantic Salmon (*Salmo salar*). *Viruses* **11**, 465.

- Fao (2018). "The state of world fisheries and aquaculture - Meeting the sustainable development goals". (Rome).
- Fux, R, Arndt, D, Langenmayer, MC, Schwaiger, J, Ferling, H, Fischer, N, Indenbirken, D, Grundhoff, A, Dölken, L, Adamek, M, Steinhagen, D and Sutter, G (2019). Piscine Orthoreovirus 3 Is Not the Causative Pathogen of Proliferative Darkening Syndrome (PDS) of Brown Trout (*Salmo trutta fario*). *Viruses* **11**, 112.
- Gallagher, MD, Matejusova, I, Nguyen, L, Ruane, NM, Falk, K and Macqueen, DJ (2018). Nanopore sequencing for rapid diagnostics of salmonid RNA viruses. *Scientific Reports* **8**, 16307.
- Goodwin, S, Mcpherson, JD and McCombie, WR (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**, 333.
- Haugland, O, Mikalsen, AB, Nilsen, P, Lindmo, K, Thu, BJ, Eliassen, TM, Roos, N, Rode, M and Evensen, O (2011). Cardiomyopathy syndrome of atlantic salmon (*Salmo salar* L.) is caused by a double-stranded RNA virus of the Totiviridae family. *Journal of virology* **85**, 5275-5286.
- Heather, JM and Chain, B (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8.
- Illingworth, CJR, Roy, S, Beale, MA, Tutill, H, Williams, R and Breuer, J (2017). On the effective depth of viral sequence data. *Virus Evolution* **3**, vex030.
- Jansen, PA, Kristoffersen, AB, Viljugrein, H, Jimenez, D, Aldrin, M and Stien, A (2012). Sea lice as a density-dependent constraint to salmonid farming. *Proceedings. Biological sciences* **279**, 2330-2338.
- Ke, R, Mignardi, M, Hauling, T and Nilsson, M (2016). Fourth Generation of Next-Generation Sequencing Technologies: Promise and Consequences. *Human mutation* **37**, 1363-1367.
- Kebschull, JM and Zador, AM (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research* **43**, e143-e143.
- Klafack, S, Fiston-Lavier, A-S, Bergmann, SM, Hammoumi, S, Schröder, L, Fuchs, W, Lusastuti, A, Lee, P-Y, Heredia, SV, Consortium, MS, Gosselin-Grenet, A-S and Avarre, J-C (2019). Cyprinid herpesvirus 3 Evolves In Vitro through an Assemblage of Haplotypes that Alternatively Become Dominant or Under-Represented. *Viruses* **11**, 754.
- Kuehn, R, Stoeckle, BC, Young, M, Popp, L, Taeubert, J-E, Pfaffl, MW and Geist, J (2018). Identification of a piscine reovirus-related pathogen in proliferative darkening syndrome (PDS) infected brown trout (*Salmo trutta fario*) using a next-generation technology detection pipeline. *PLOS ONE* **13**, e0206164.
- Lahnsteiner, F, Haunschmid, R and Mansour, N (2011). Possible reasons for late summer brown trout (*Salmo trutta* Linnaeus 1758) mortality in Austrian prealpine river systems. *Journal of Applied Ichthyology* **27**, 83-93.
- Lee, C-T, Chen, IT, Yang, Y-T, Ko, T-P, Huang, Y-T, Huang, J-Y, Huang, M-F, Lin, S-J, Chen, C-Y, Lin, S-S, Lightner, DV, Wang, H-C, Wang, AHJ, Wang, H-C, Hor, L-I and Lo, C-F (2015). The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 10798-10803.
- Løvoll, M, Wiik-Nielsen, J, Grove, S, Wiik-Nielsen, CR, Kristoffersen, AB, Faller, R, Poppe, T, Jung, J, Pedamallu, CS, Nederbragt, AJ, Meyerson, M, Rimstad, E and Tengs, T (2010).



- A novel totivirus and piscine reovirus (PRV) in Atlantic salmon (*Salmo salar*) with cardiomyopathy syndrome (CMS). *Virology journal* **7**, 309-309.
- Margulies, M, Egholm, M, Altman, WE, Attiya, S, Bader, JS, Bemben, LA, Berka, J, Braverman, MS, Chen, Y-J, Chen, Z, Dewell, SB, Du, L, Fierro, JM, Gomes, XV, Godwin, BC, He, W, Helgesen, S, Ho, CH, Irzyk, GP, Jando, SC, Alenquer, MLI, Jarvie, TP, Jirage, KB, Kim, J-B, Knight, JR, Lanza, JR, Leamon, JH, Lefkowitz, SM, Lei, M, Li, J, Lohman, KL, Lu, H, Makhijani, VB, Mcdade, KE, Mckenna, MP, Myers, EW, Nickerson, E, Nobile, JR, Plant, R, Puc, BP, Ronan, MT, Roth, GT, Sarkis, GJ, Simons, JF, Simpson, JW, Srinivasan, M, Tartaro, KR, Tomasz, A, Vogt, KA, Volkmer, GA, Wang, SH, Wang, Y, Weiner, MP, Yu, P, Begley, RF and Rothberg, JM (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.
- Maxam, AM and Gilbert, W (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 560-564.
- Munang'andu, HM (2016). Environmental Viral Metagenomics Analyses in Aquaculture: Applications in Epidemiology and Disease Control. *Frontiers in Microbiology* **7**.
- Nho, SW, Abdelhamed, H, Paul, D, Park, S, Mauel, MJ, Karsi, A and Lawrence, ML (2018). Taxonomic and Functional Metagenomic Profile of Sediment From a Commercial Catfish Pond in Mississippi. *Frontiers in Microbiology* **9**.
- Nkili-Meyong, AA, Bigarré, L, Labouba, I, Vallaeys, T, Avarre, JC and Berthet, N (2016). Contribution of Next-Generation Sequencing to Aquatic and Fish Virology. *Intervirology* **59**, 285-300.
- Nyren, P, Pettersson, B and Uhlen, M (1993). Solid Phase DNA Minisequencing by an Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay. *Analytical Biochemistry* **208**, 171-175.
- Palacios, G, Lovoll, M, Tengs, T, Hornig, M, Hutchison, S, Hui, J, Kongtorp, RT, Savji, N, Bussetti, AV, Solovyov, A, Kristoffersen, AB, Celone, C, Street, C, Trifonov, V, Hirschberg, DL, Rabadan, R, Egholm, M, Rimstad, E and Lipkin, WI (2010). Heart and skeletal muscle inflammation of farmed salmon is associated with infection with a novel reovirus. *PLoS One* **5**, e11487.
- Peters, L, Spatharis, S, Dario, MA, Dwyer, T, Roca, IJT, Kintner, A, Kanstad-Hanssen, Ø, Llewellyn, MS and Praebel, K (2018). Environmental DNA: A New Low-Cost Monitoring Tool for Pathogens in Salmonid Aquaculture. *Frontiers in Microbiology* **9**.
- Rodger, HD (2016). "Fish Disease Causing Economic Impact in Global Aquaculture," in *Fish Vaccines*, ed. A. Adams, 1-34. (Basel: Springer Basel).
- Ronaghi, M, Uhlén, M and Nyren, P (1998). A Sequencing Method Based on Real-Time Pyrophosphate. *Science* **281**, 363-365.
- Sanger, F, Nicklen, S and Coulson, AR (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**, 5463-5467.
- Schadt, EE, Turner, S and Kasarskis, A (2010). A window into third-generation sequencing. *Human Molecular Genetics* **19**, R227-R240.
- Shendure, J and Ji, H (2008). Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135-1145.
- Smith, LM, Sanders, JZ, Kaiser, RJ, Hughes, P, Dodd, C, Connell, CR, Heiner, C, Kent, SBH and Hood, LE (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-679.

- Stärk, KDC, Pękala, A and Muellner, P (2019). Use of molecular and genomic data for disease surveillance in aquaculture: Towards improved evidence for decision making. *Preventive Veterinary Medicine* **167**, 190-195.
- Stevick, RJ, Sohn, S, Modak, TH, Nelson, DR, Rowley, DC, Tammi, K, Smolowitz, R, Markey Lundgren, K, Post, AF and Gómez-Chiarri, M (2019). Bacterial Community Dynamics in an Oyster Hatchery in Response to Probiotic Treatment. *Frontiers in Microbiology* **10**.
- Tran, L, Nunan, L, Redman, RM, Mohney, LL, Pantoja, CR, Fitzsimmons, K and Lightner, DV (2013). Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome affecting penaeid shrimp. *Diseases of Aquatic Organisms* **105**, 45-55.
- Van Dijk, EL, Jaszczyszyn, Y, Naquin, D and Thermes, C (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics* **34**, 666-681.
- Walker, PJ and Winton, JR (2010). Emerging viral diseases of fish and shrimp. *Veterinary Research* **41**, 51.
- Wessel, Ø, Braaen, S, Alarcon, M, Haatveit, H, Roos, N, Markussen, T, Tengs, T, Dahle, MK and Rimstad, E (2017). Infection with purified Piscine orthoreovirus demonstrates a causal relationship with heart and skeletal muscle inflammation in Atlantic salmon. *PLOS ONE* **12**, e0183781.

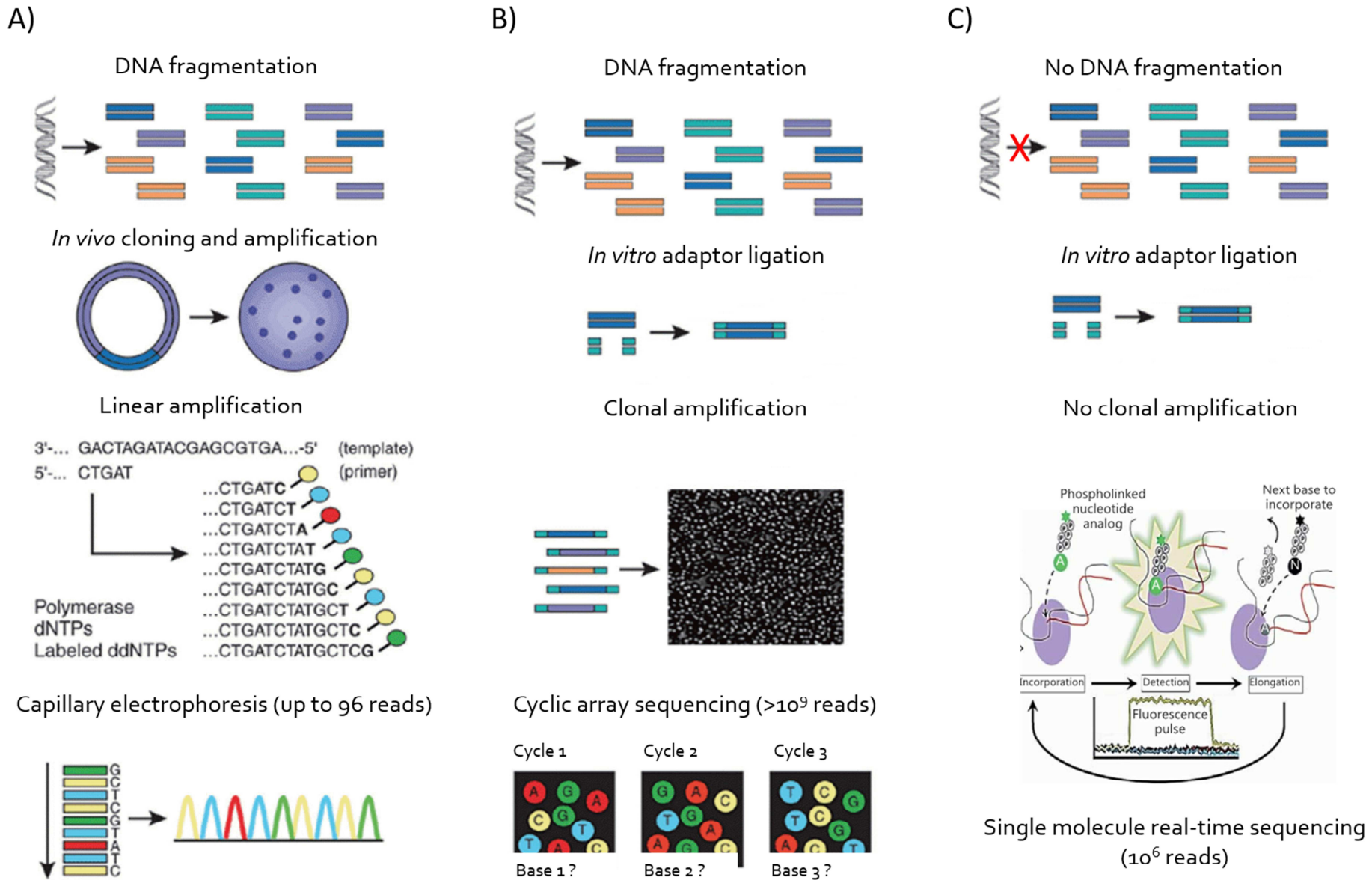


Figure 1: Illustration of the major differences between first (A), second (B) and third (C) generations of sequencing. Modified from Shendure & Ji (2008).