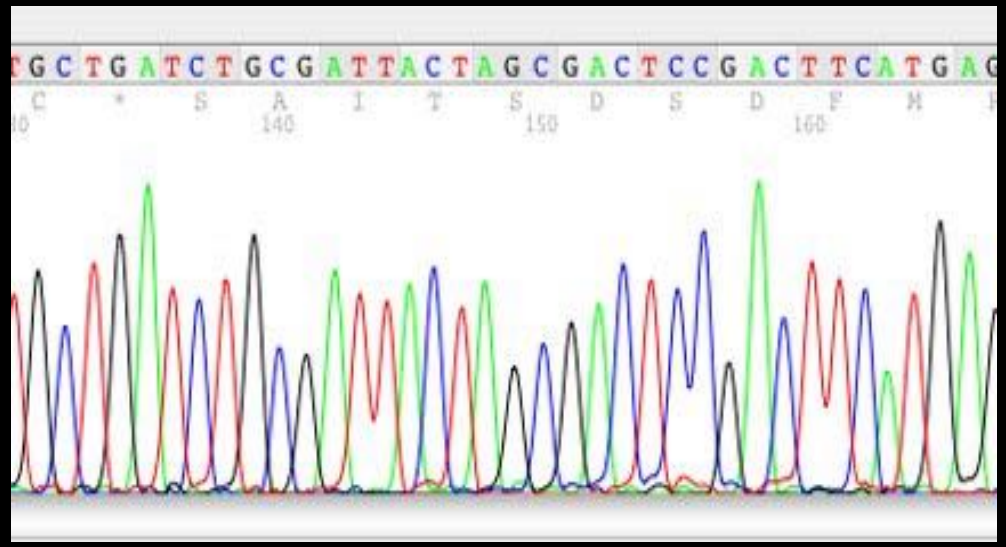
Introduction to DNA Sequencing Technology

Hendra Wibawa

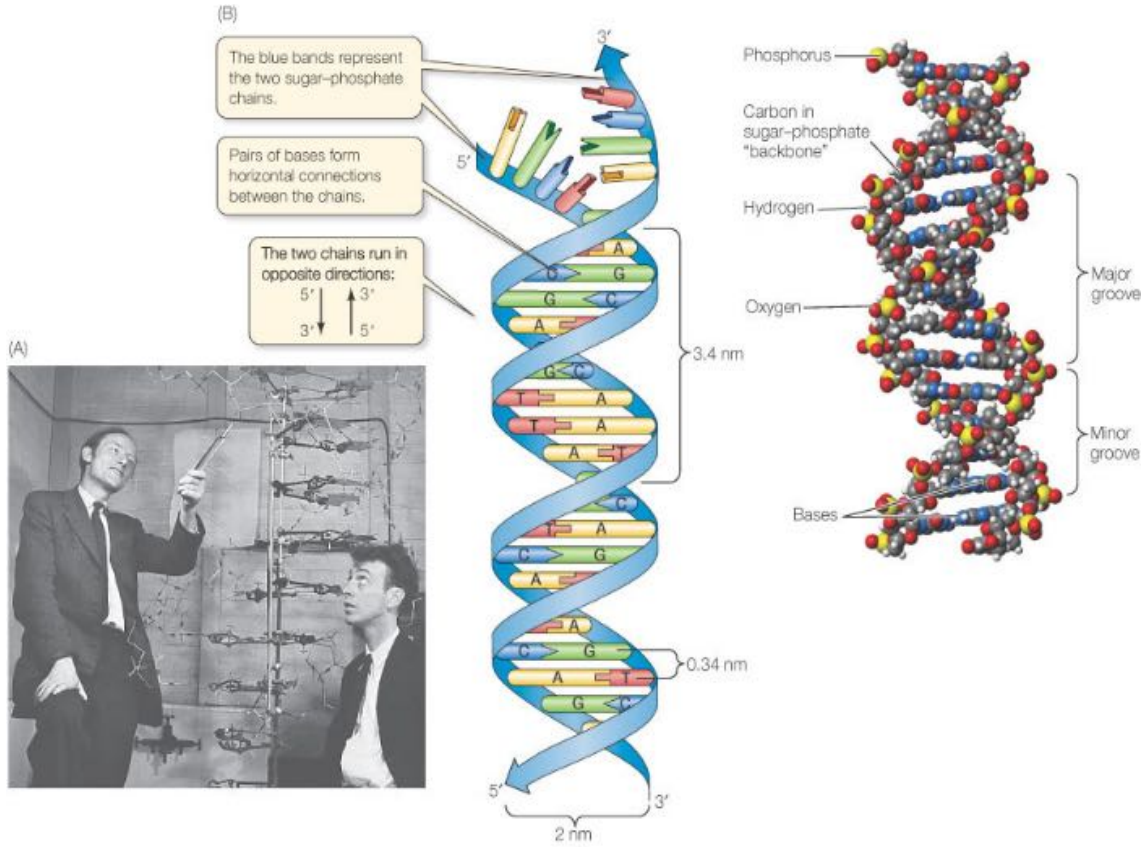
WHAT IS DNA SEQUENCING?



The process for the determining the right and precise order of nucleotide in a DNA molecule

From DNA to Sequencing

DNA Stucture Discovery



(Watson and Crick, 1953)

'First generation' sequencing

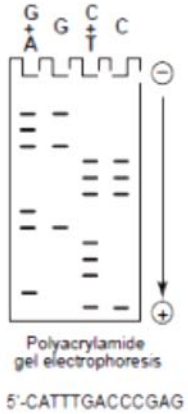
A Tale of Two Cambridges

A Maxam-Gilbert method

Based on chemical degradation of end-labeled DNA (one strand is labeled at 5′ end).

G+A: DMS, piperidine
G: HCl, DMS, piperidine
C+T: hydrazine, piperidine
C: NaCl, hydrazine, piperidine

Degradation products are separated by slab gel polyacrylamide gelelectrophoresis.

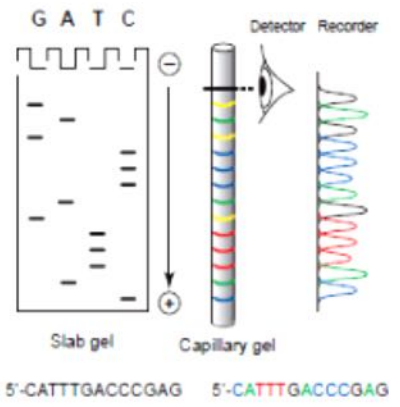


B Sanger method

Based on DNA synthesis from a single-strand template with DNA polymerase and ddNTPs.

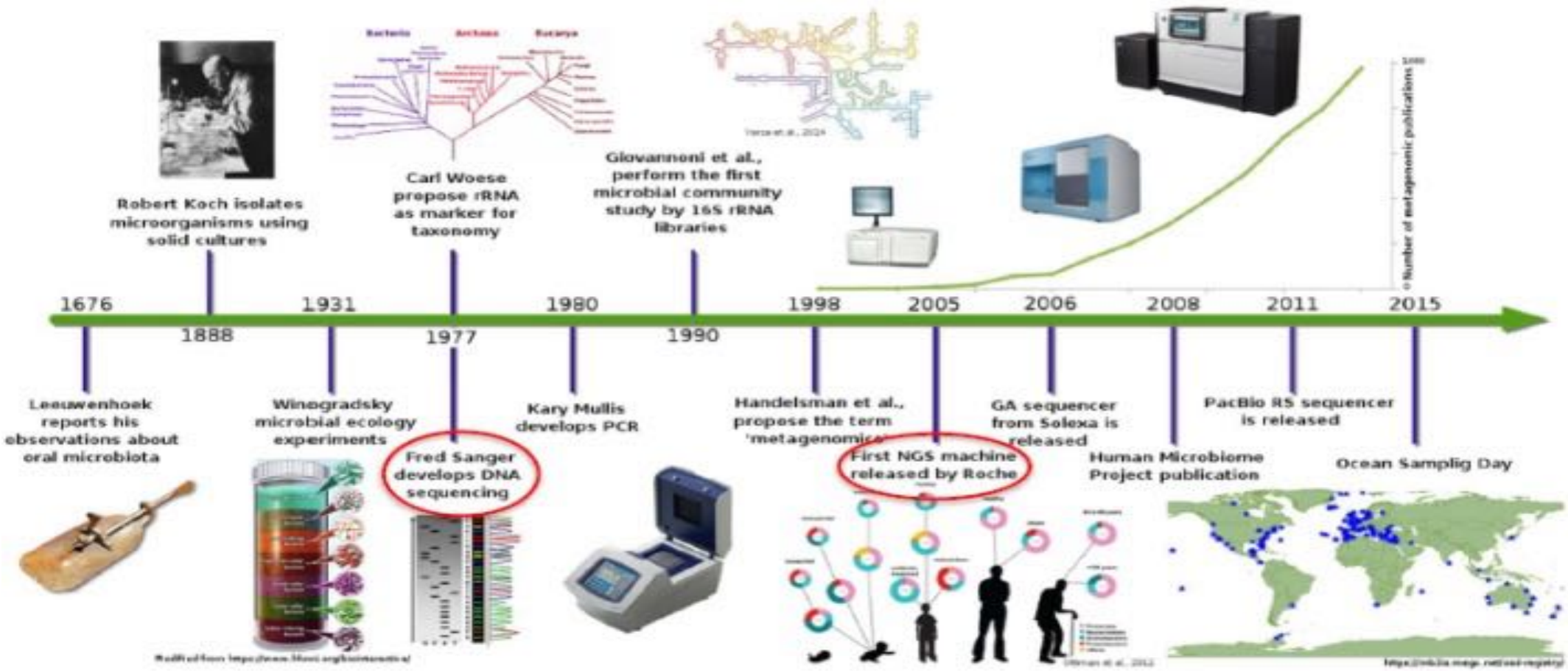
G: reaction with ddGTP
A: reaction with ddATP
T: reaction with ddTTP
C: reaction with ddCTP

Labeled products are separated by slab gel polyacrylamide gelelectrophoresis (left) or by column gelelectrophoresis (right).



(Maxam-Gilbert and Sanger, 1977)

Historical timeline for metagenomics analysis



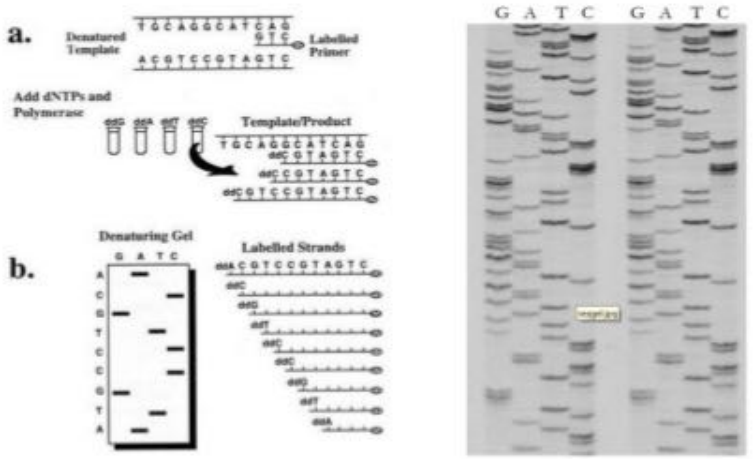
Escobar-Zepeda et al. (2015) The Road to Metagenomics: from microbiology to DNA sequencing Technologies and bioinformatics. *Frontiers in Genetics* 6: 348

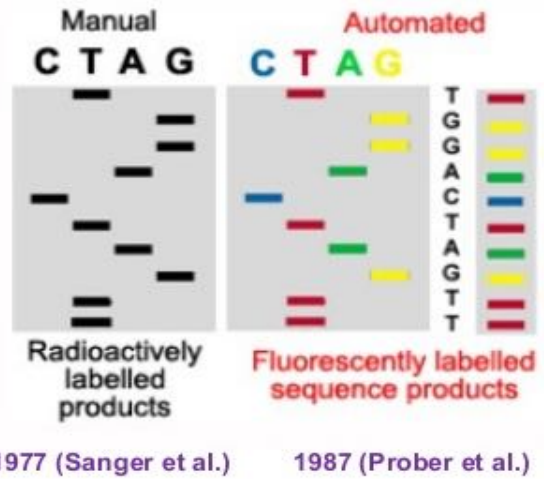
SANGER SEQUENCING - 1st generation sequencing

For the past 30 years until now, the Sanger method has been the gold standard for DNA sequencing

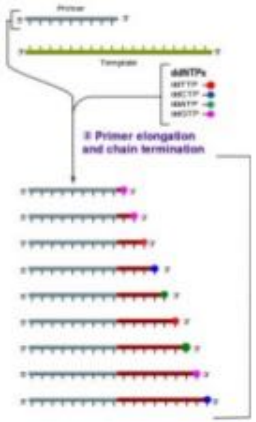
The Methods

Sanger sequencing: the dideoxy method - Chain termination sequencing



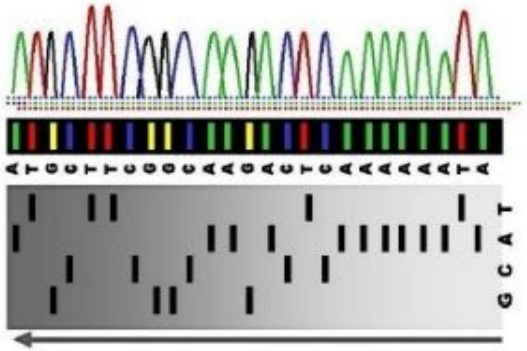


1977 (Sanger et al.) 1987 (Prober et al.)



Sanger Sequencing Workflow:

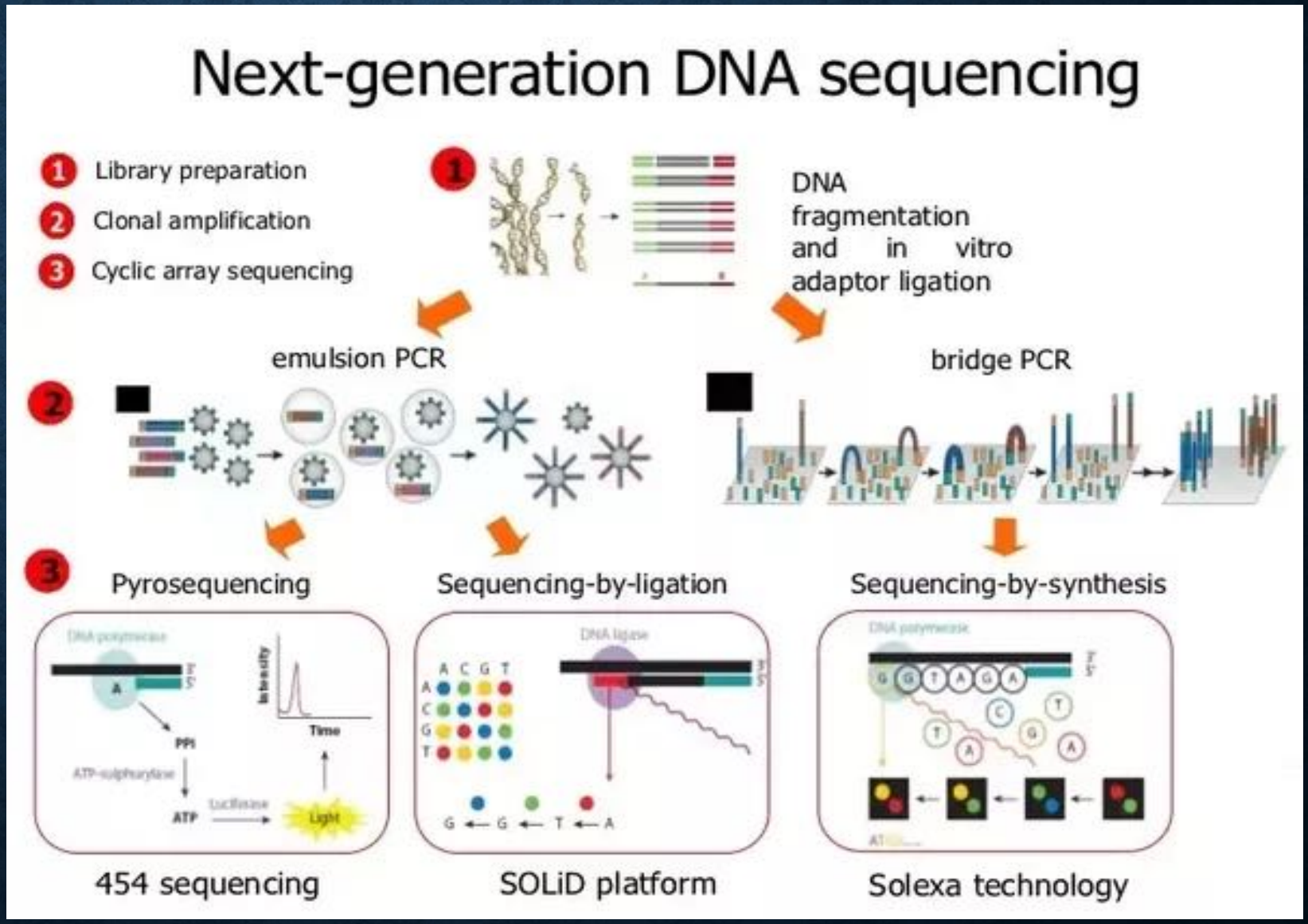
- ✓ PCR amplification (target enrichment)
- ✓ PCR purification (primer, dNTPs)
- ✓ **Sequencing reaction (bi-directional)**
- ✓ Sequencing purification (primer, dNTPs, ddNTPs)
- ✓ Electrophoretic run on sequencer
- ✓ Sequencing lecture
- ✓ Alignment to reference



Toma

The Methods

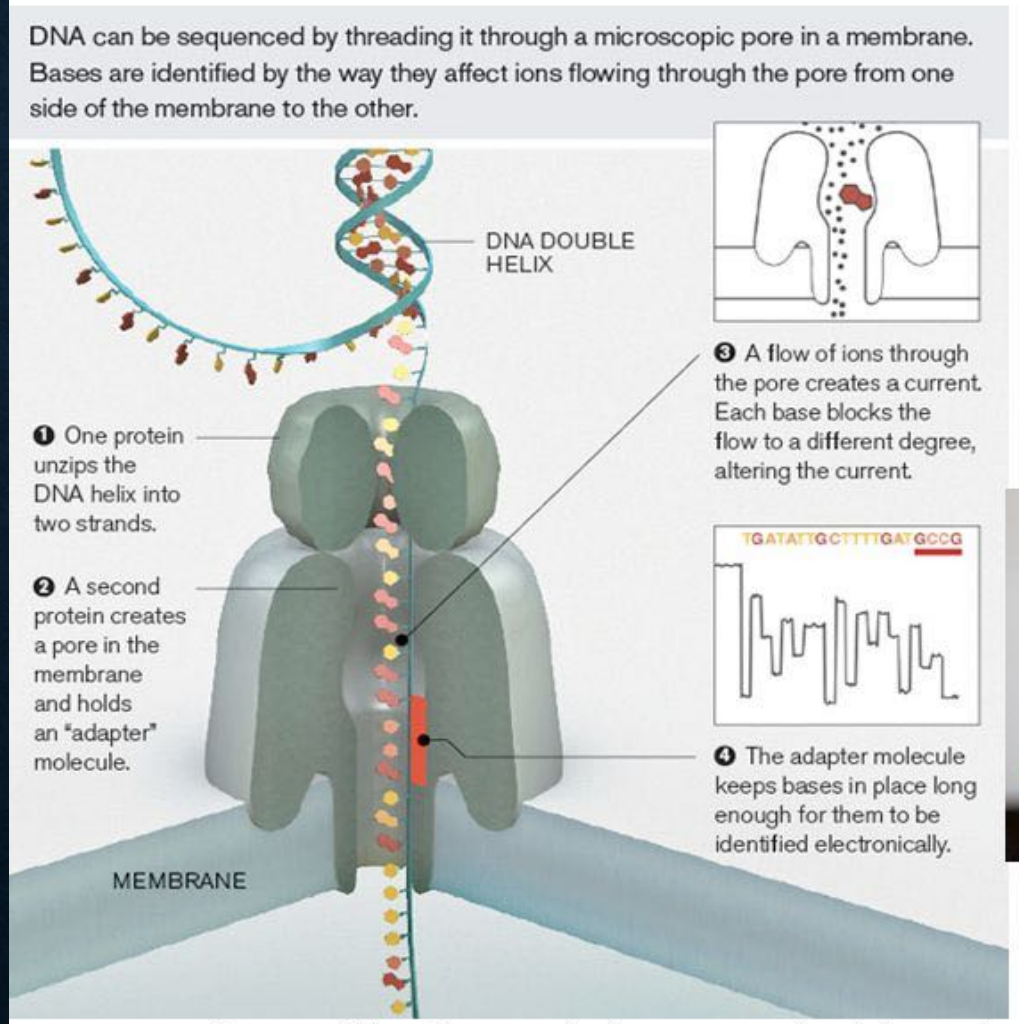
2nd Generation Sequencing



The Methods

3nd Generation Sequencing

Nanopore Sequencing



- In development since 1995
- Company: Oxford Nanopore
- First working 'development stage' devices (MinION) released to testing groups

Image source: John MacNeill, http://www2.technologyreview.com/article/427677/nanopore-sequencing

COMPARISON OF NGS SYSTEMS

Quail et al. BMC Genomics 2012, 13:341
http://www.biomedcentral.com/1471-2164/13/341

BMC Genomics

RESEARCH ARTICLE Open Access

A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers

2012

Review Article

Comparison of Next-Generation Sequencing Systems

Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law

Miyamoto et al. BMC Genomics 2014, 15:699
http://www.biomedcentral.com/1471-2164/15/699

2014

BMC Genomics

RESEARCH ARTICLE Open Access

Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes

Each sequencing platform has advantages and disadvantages

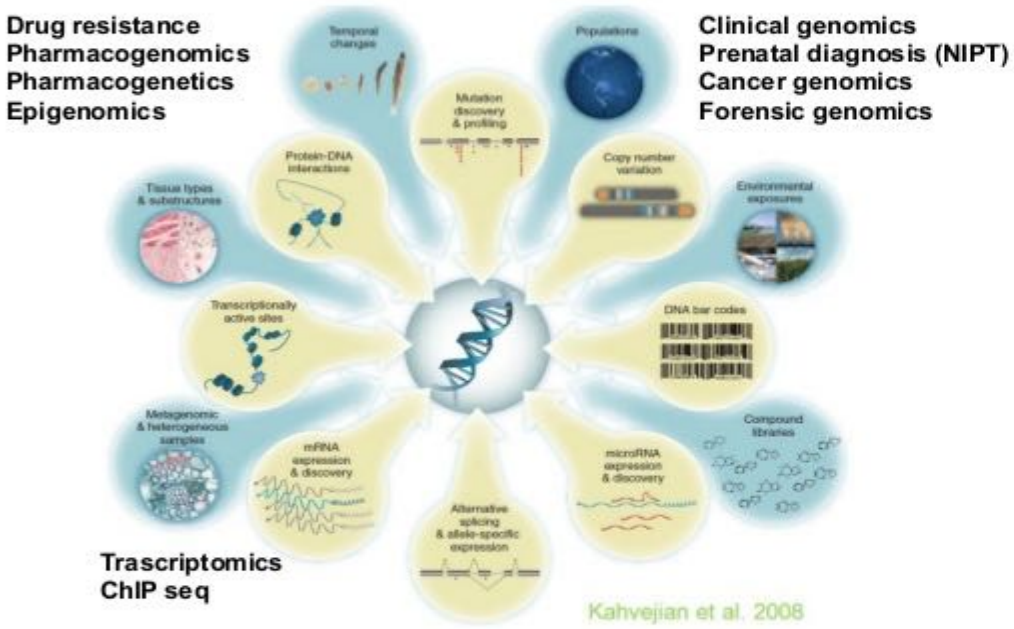
Illumina	Ion Torrent (Life Technologies)	PacBio RS
➢ Lower error rate ➢ Lowest cost per base ➢ Wide range of applications	➢ Low error rate ➢ Medium/low cost per base ➢ Fast run (hours) ➢ Low startup costs	➢ No amplification required ➢ Extremely long read lengths (max 15000bp) ➢ de-novo assembly
➢ Short read length (50-150bp) ➢ Runs take multiple days ➢ No de-novo assembly	➢ Homopolymers reads problem ➢ Read lengths only 100-200bp ➢ coverage bias with GC-rich regions ➢ New, developing technology	➢ High error rates (5-15%) ➢ Medium/high cost per base ➢ High startup costs

no mutation detection (diagnostic)

Toma

Next Generation SEQUENCING: Applications

What would you do if you could sequence everything?



Metagenomics
(microbiome – infectious agents)

Agrigenomics

- **Whole Genome Sequencing (WGS):** characterize entire genomes of any size and complexity
- **Exome Sequencing :** sequence protein coding regions, as cost-effectivealternative to WGS
- **Targeted Resequencing:** sequence specific genes or other regions of interest
- **De novo Sequencing:** sequence and assemble novel genomes

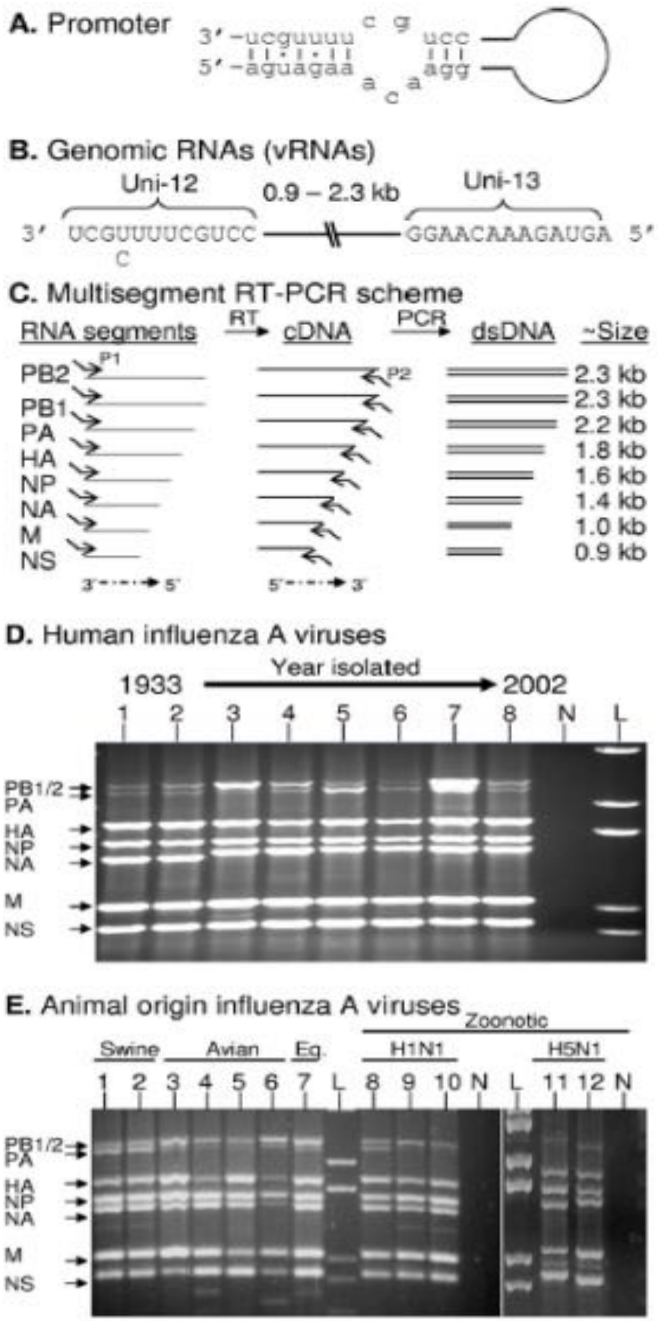
Kahvejian et al. Nature Biotechnology 2008

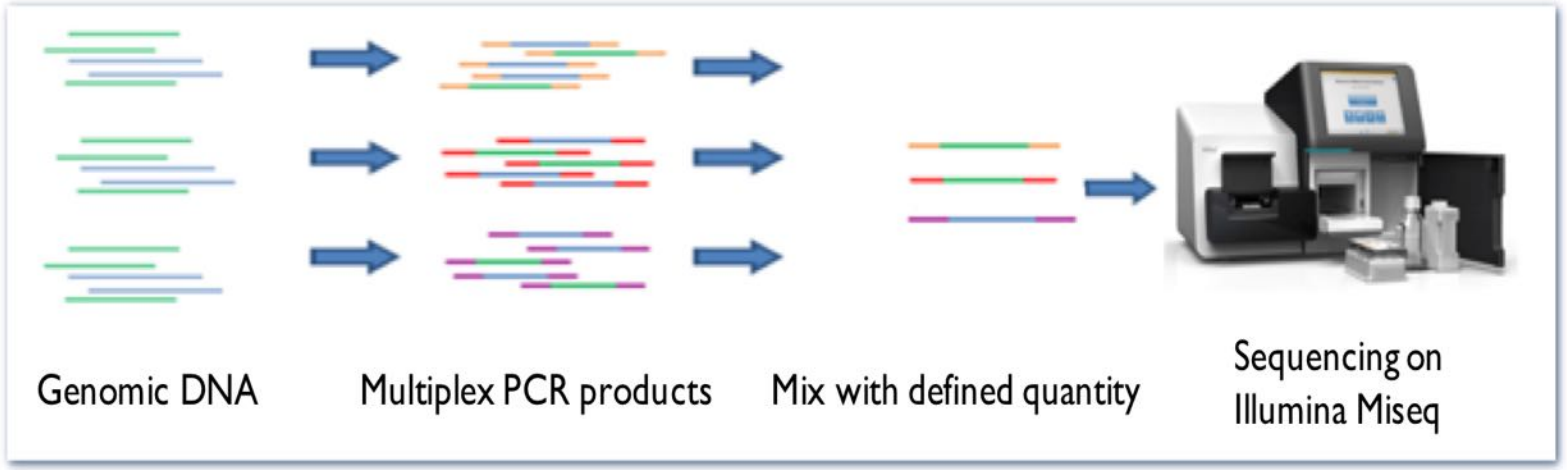
Toma

COMPARISON SANGER SEQUENCING & NEXT GENERATION SEQUENCING

	Sanger Sequencing	NGS
Number of reactions	Separate reactions for different genes	One single reaction for different genes
Starting material	Greater amount of genomic DNA	<10 ng genomic DNA
Output	Low throughput, one sample needs two separate reactions for forward and reverse primers	High throughput, allows massively parallel and millions of fragments can be sequenced simultaneously
Sequencing cost for full genomes and manpower	Less cost-effective and more labor-intensive	More cost-effective and less labor-intensive
Read length	Longer reads	Shorter reads
Raw data storage	Easy data storage	Requires powerful data storage (e.g. a full influenza genome of 1.4kb will result in approx. 2Gb/sample)

WHOLE GENOME SEQUENCING AI VIRUS DISEASE INVESTIGATION CENTER WATES





Schematic Workflow for FastTarget

Multisegments RT-PCR (Zhou et al., 2009)

NGS Applications for Pathogen Characterization in Disease Investigation Center Wates

- Avian Influenza
- African Swine Fever
- Bovine Viral Diarrhea
- SARS-CoV-2 B. anthracis
- E. coli (AMR)

WGS Publications

Veterinary World, EISSN: 2231-0916
Available at www.veterinaryworld.org/Vol.12/July-2019/27.pdf

RESEARCH ARTICLE
Open Access

Genetic analysis of NS5B gene from bovine viral diarrhea virus-infected cattle in Central and East Java, Indonesia

S. H. Irianingsih[1,2], H. Wuryastuty[3], R. Wasito[4], H. Wibawa[2], F. S. Tjatur Rasa[5] and B. Poermadjaja[2]

1. Doctoral Study Program, Faculty of Veterinary Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia; 2. Disease Investigation Centre Wates, Yogyakarta, Indonesia; 3. Department of Veterinary Internal Medicine, Faculty of Veterinary Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia; 4. Department of Veterinary Pathology, Faculty of Veterinary Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia; 5. Directorate of Animal Health, Directorate General of Livestock Services and Animal Health, Ministry of Agriculture, The Republic of Indonesia, Jakarta, Indonesia.

Corresponding author: H. Wuryastuty, e-mail: hastari@ugm.ac.id
Co-authors: SHI: yanibiotech@gmail.com, RW: prof_wst@yahoo.com, HWi: hi.wibawa@gmail.com, FSTR: fadjarstr@yahoo.com, BP: bagoespoermadjaja@yahoo.co.id
Received: 26-03-2019, **Accepted:** 13-06-2019, **Published online:** 25-07-2019

doi: 10.14202/vetworld.2019.1108-1115 **How to cite this article:** Irianingsih SH, Wuryastuty H, Wasito R, Wibawa H, Tjatur Rasa FS, Poermadjaja B (2019) Genetic analysis of NS5B gene from bovine viral diarrhea virus-infected cattle in Central and East Java, Indonesia, *Veterinary World*, 12(7): 1108-1115.

Abstract

Background and Aim: A previous study divided Indonesian bovine viral diarrhea virus (BVDV)-1 into subgenotypes BVDV-1a to BVDV-1d based on the partial NS5B gene using strain Bega as reference for BVDV-1a. In fact, it is clustered into BVDV-1c with strain Bega-like Australia. BVDV genotyping has been done on isolates from Jakarta, West and Central Java, but East Java isolates have not been genotyped. This study aimed to analyze genetic variability and amino acid residues in the nucleotide-binding pocket of the NS5B gene from infected cattle.

Materials and Methods: Samples were obtained from the Sera Bank originating from active and passive surveillance of cattle that had been tested for BVDV antigen from 2013 to 2017. Detection of the p80 antibody and BVDV genotyping was carried out using ELISA and nested-multiplex-polymerase chain reaction (PCR), respectively. We defined 15 nested PCR products for partial sequencing of NS5B. Those field samples were selected from each location and year using proportional calculation as a representative sample. Homological and phylogenetic analyses of the partial NS5B gene were performed using BLAST and MEGA version 6.

Results: Based on the phylogenetic tree analysis using 360 nucleotides as the partial NS5B gene, Indonesian BVDV-1 isolates from Central and East Java were subdivided to BVDV-1a (n=9), BVDV-1b (n=1), and BVDV-1c (n=5). In the present study, the homology of BVDV subgenotype -1a, -1b, and -1c was compared to the BVDV GenBank data and found 90-93%, 93%, and 92-95% respectively with the average pairwise distance of 0.207. A point mutation was shown at R283K of all BVDV isolates based on the sequence of three amino acid residues R283, R285, and I287 in the nucleotide-binding pocket as a part of the encoded RNA-dependent RNA polymerase.

Conclusion: This study revealed the genetic variability of BVDV infecting cattle in Central Java and East Java, Indonesia, the subtypes BVDV-1a, BVDV-1b, BVDV-1c, and a point mutation at the R283K residue.

Keywords: bovine viral diarrhea virus, NS5B gene, phylogenetic analysis, point mutation, subgenotype.

Introduction

Bovine viral diarrhea virus (BVDV) is an important viral pathogen of cattle that has spread globally and that causes significant economic loss to both dairy and beef cattle [1]. BVDV causes thousands and up to tens of millions of dollars of loss per calving interval [2] due to productivity and reproductive disorders in the herd [3]. Around 70-90% of infected cattle show no clinical signs [4-6]. The immunosuppressive condition may increase both the risk of secondary infection and inefficient reproduction and productivity. The BVDV genome is a single-stranded positive-sense ribonucleic acid (RNA) belonging to the genus *Pestivirus* and the family Flaviviridae [7]. The BVDV genome is about 12.3 kb long, which organized as an open reading frame flanked by 5'- and 3'-untranslated regions (UTR) [8-10]. It encodes a single polyprotein of about 4000 amino acids consisting of proteins in the order of NH2-Npro-C-Erns-E1-E2-P7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH. The BVDV can be categorized into two genotypes or species: BVDV-1 and BVDV-2 [11]. Based on the nucleotide sequence variation in the 5' UTR [12] and four other regions including Npro, E2, NS3, and NS5B–3'UTR [13], the genotypes BVDV-1 and BVDV-2 can be divided into numerous subgenotypes. Nonstructural NS5B was classified as a highly conserved gene [14] with a nucleotide length of 2.156

ORIGINAL ARTICLE

WILEY

Co-circulation and characterization of HPAI-H5N1 and LPAI-H9N2 recovered from a duck farm, Yogyakarta, Indonesia

Lestari[1,2] | Hendra Wibawa[2] | Elly Puspasari Lubis[2] | Rama Dharmawan[2] | Rina Astuti Rahayu[2] | Herdiyanto Mulyawan[2] | Kamonpan Charoenkul[1] | Chanakarn Nasamran[1] | Bagoes Poermadjaja[2] | Alongkorn Amonsin[1]

[1]Department of Veterinary Public Health, Center of Excellence for Emerging and Re-emerging Infectious Diseases in Animals, Faculty of Veterinary Science, Chulalongkorn University, Bangkok, Thailand

[2]Disease Investigation Center Wates Yogyakarta, Directorate General of Livestock and Animal Health Services, Ministry of Agriculture Indonesia, Yogyakarta, Indonesia

Correspondence
Alongkorn Amonsin, Department of Veterinary Public Health, Center of Excellence for Emerging and Re-emerging Infectious Diseases in Animals, Faculty of Veterinary Science, Chulalongkorn University, Bangkok, 10330, Thailand.
Email: alongkorn.a@chula.ac.th

Funding information
Thailand Research Fund, Grant/Award Number: RTA6080012; Chulalongkorn University

Abstract

In July 2016, an avian influenza outbreak in duck farms in Yogyakarta province was reported to Disease Investigation Center (DIC), Wates, Indonesia, with approximately 1,000 ducks died or culled. In this study, two avian influenza (AI) virus subtypes, A/duck/Bantul/04161291-OR/2016 (H5N1) and A/duck/Bantul/04161291-OP/2016 (H9N2) isolated from ducks in the same farm during an AI outbreak in Bantul district, Yogyakarta province, were sequenced and characterized. Our results showed that H5N1 virus was closely related to the highly pathogenic AI (HPAI) H5N1 of clade 2.3.2.1c, while the H9N2 virus was clustered with LPAI viruses from China, Vietnam and Indonesia H9N2 (CVI lineage). Genetic analysis revealed virulence characteristics for both in avian and in mammalian species. In summary, co-circulation of HPAI-H5N1 of clade 2.3.2.1c and LPAI-H9N2 was identified in a duck farm during an AI outbreak in Yogyakarta province, Indonesia. Our findings raise a concern of the potential risk of the viruses, which could increase viral transmission and/or threat to human health. Routine surveillance of avian influenza viruses should be continuously conducted to understand the dynamic and diversity of the viruses for influenza prevention and control in Indonesia and SEA region.

KEYWORDS
co-circulation, H5N1, H9N2, Indonesia, Influenza

1 | INTRODUCTION

Highly pathogenic avian influenza subtype H5N1 (HPAI-H5N1) is a highly contagious virus causing high morbidity and mortality in avian and mammal species. HPAI-H5N1 became internationally of concern due to its serious impact on animal and human health. The HPAI-H5N1 has been reported worldwide including Asia, Africa and Europe since the first reported in China in 1996 (Webster & Govorkova, 2006). As of April 2019, WHO has reported a total of 860 human cases of HPAI-H5N1 in 16 countries with 454 death (WHO, 2019b). Currently, the HPAI-H5N1 virus continues to cause influenza outbreaks in poultry and sporadic human cases in Asia and Africa. In addition, outbreaks of reassortant H5NX were reported in poultry and wild birds in Europe and North America (OIE, 2018).

Low pathogenic avian influenza subtype H9N2 (LPAI-H9N2) was first isolated from turkeys in the United States in 1966 (Homme & Easterday, 1970). The virus did spread by waterfowl and shorebirds in North America (Jackwood & Stallknecht, 2007) and has become endemic in poultry across East Asia and Middle East with some sporadic infections in Europe (Aamir, Wernery, Ilyushina, & Webster, 2007; Guan et al., 2000; Werner, 1998). As of July 2019, 26 confirmed human cases of H9N2 have been reported in China (Butt et al., 2005; Pan et al., 2018; Peiris et al., 1999; WHO, 2019a). Recurrence of H9N2 human cases has raised a potential risk of

PeerJ

Full-length genome characterization and phylogenetic analysis of SARS-CoV-2 virus strains from Yogyakarta and Central Java, Indonesia

Gunadi[1], Hendra Wibawa[2], Marcellus[1], Mohamad Saifudin Hakim[3], Edwin Widyanto Daniwijaya[4], Ludhang Pradipta Rizki[3], Endah Supriyati[5], Dwi Aris Agung Nugrahaningsih[6], Afiahayati[7], Siswanto[8], Kristy Iskandar[9], Nungki Anggorowati[10], Alvin Santoso Kalim[1], Dyah Ayu Puspitarani[1], Kemala Athollah[1], Eggi Arguni[11], Titik Nuryastuti[3] and Tri Wibawa[3]

[1] Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2] Disease Investigation Center Wates, Yogyakarta, Ministry of Agriculture, Indonesia
[3] Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
[4] Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/UGM Academic Hospital, Yogyakarta, Indonesia
[5] Centre of Tropical Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
[6] Department of Pharmacology and Therapy, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
[7] Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
[8] Department of Physiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/UGM Academic Hospital, Yogyakarta, Indonesia
[9] Department of Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/UGM Academic Hospital, Yogyakarta, Indonesia
[10] Department of Anatomical Pathology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
[11] Department of Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

Submitted 22 September 2020
Accepted 24 November 2020
Published 21 December 2020

Corresponding authors
Gunadi, drgunadi@ugm.ac.id
Hendra Wibawa, hendra.wibawa@pertanian.go.id

Academic editor
Yuriy Orlov

Additional Information and Declarations can be found on page 12

DOI 10.7717/peerj.10575

ABSTRACT

Background: Recently, SARS-CoV-2 virus with the D614G mutation has become a public concern due to rapid dissemination of this variant across many countries. Our study aims were (1) to report full-length genome sequences of SARS-CoV-2 collected from four COVID-19 patients in the Special Region of Yogyakarta and Central Java provinces, Indonesia; (2) to compare the clade distribution of full-length genome sequences from Indonesia ($n = 60$) from March to September 2020 and (3) to perform phylogenetic analysis of SARS-CoV-2 complete genomes from different countries, including Indonesia.

Methods: Whole genome sequencing (WGS) was performed using next-generation sequencing (NGS) applied in the Illumina MiSeq instrument. Full-length virus genomes were annotated using the reference genome of hCoV-19/Wuhan/Hu-1/2019 (NC_045512.2) and then visualized in UGENE v. 1.30. For phylogenetic

DATA ANALYSIS

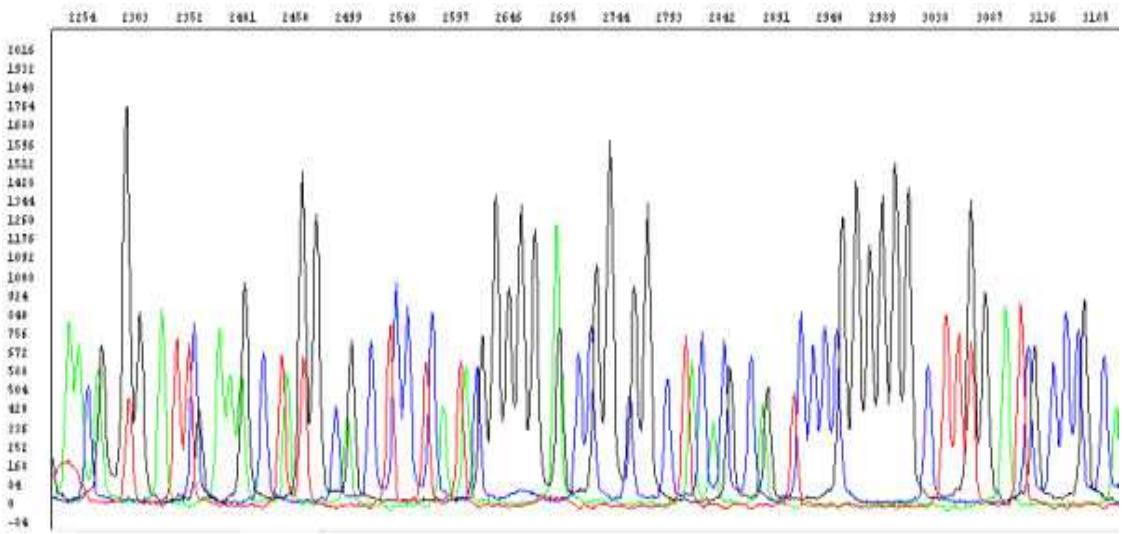
DATA (SANGER SEQUENCING OUTPUT)

Results of DNA sequencing are provided in three data files – .ab1 file, .seq file and .phd.1 file.

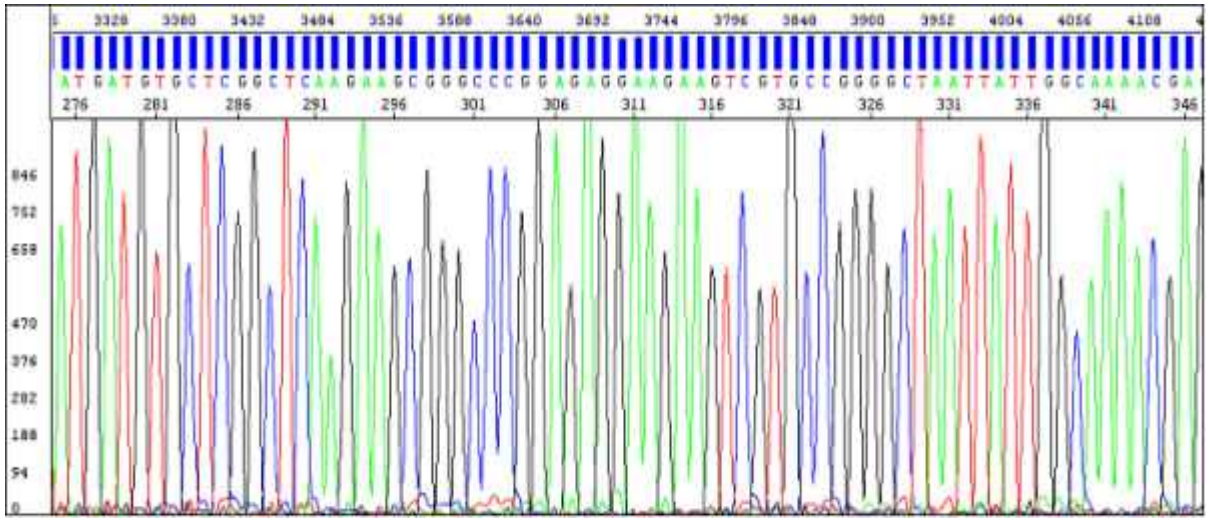
- *.ab1 file contains the DNA sequence electropherogram as well as raw data and some other information.
- *.seq file is a simple sequence text file in FASTA format.
- *.phd.1 file (Phred file) is a simple text file containing bases with quality values for each base.

DATA (SANGER SEQUENCING OUTPUT)

Raw data (data before analysis by the base caller algorithm) are data as they are recorded by the sequencer:



Electropherogram (data after analysis) shows a sequence of peaks in four colors, each color represents the base called for that peak and there is a textual version of recorded sequence visible:



DATA (SANGER SEQUENCING OUTPUT)

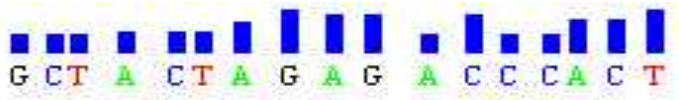
Data analysis

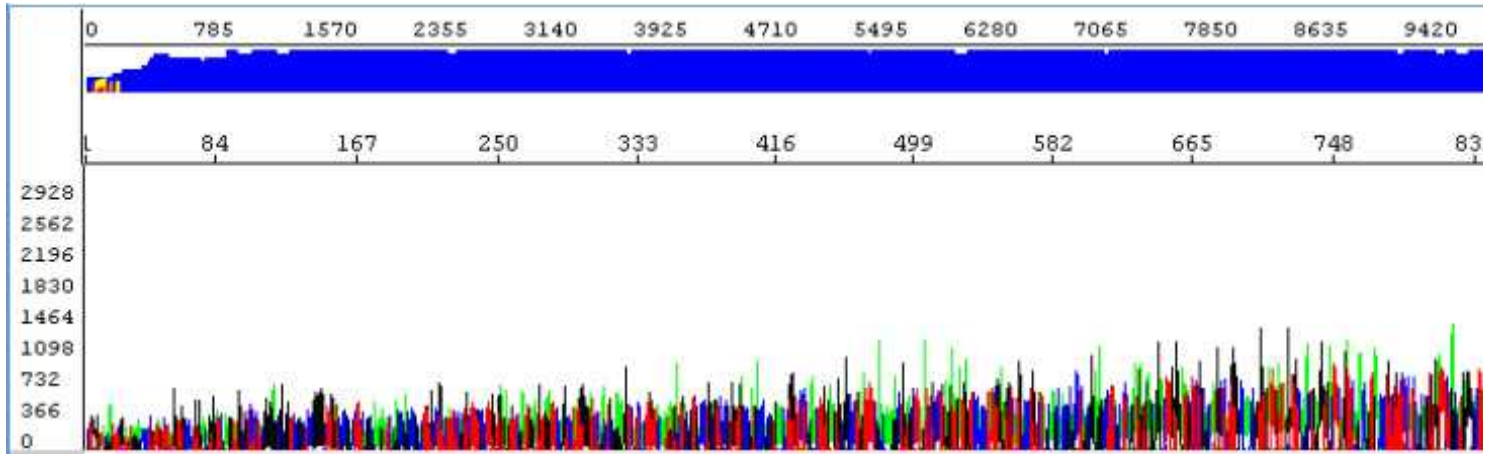
When evaluating .ab1 files, you should first see the electropherogram and come to a conclusion whether your data can be considered of good quality or not.

- Good quality sequencing data are characterized by:
- well-defined peak resolution (bad resolution of the first 10-25 bases is acceptable)
- uniform peak spacing
- high signal-to-noise ratios

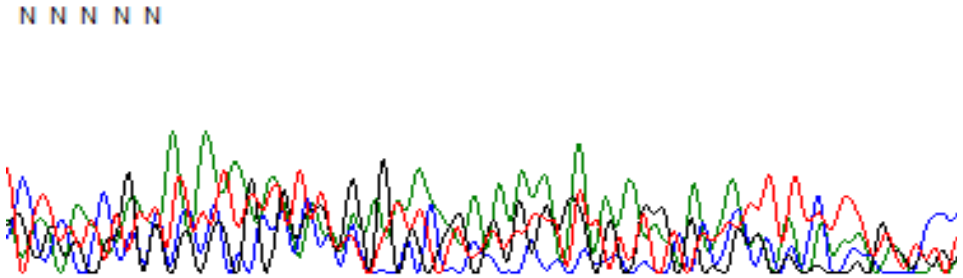
An example of a very good quality data:

A quick and very comfortable way to check the data quality is Quality Values (QVs). By definition the QV is a per-base estimate of the basecaller accuracy. In a plain language, QVs are colored bars above peaks/bases:

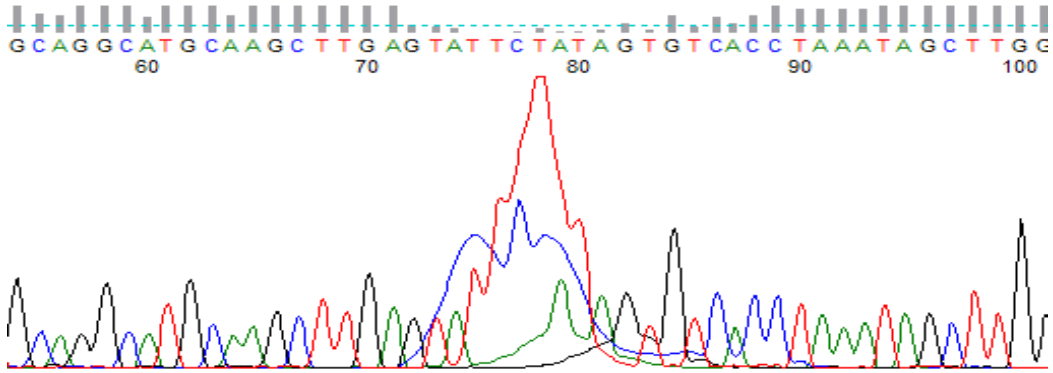




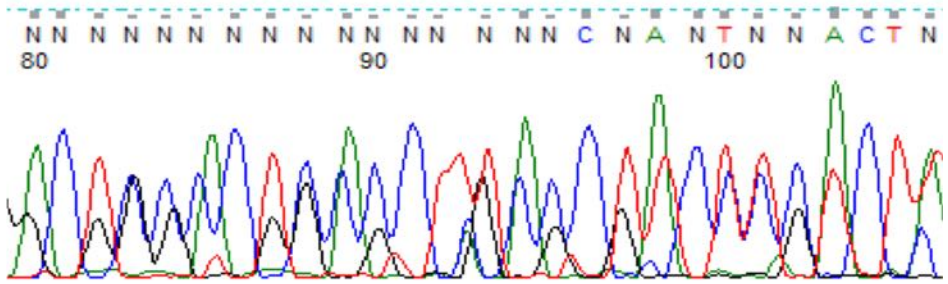
POOR SANGER DATA



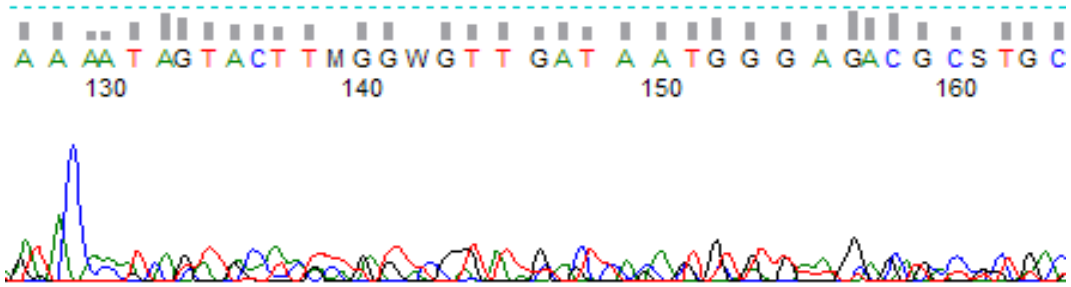
Failed DNA Sequencing Reactions



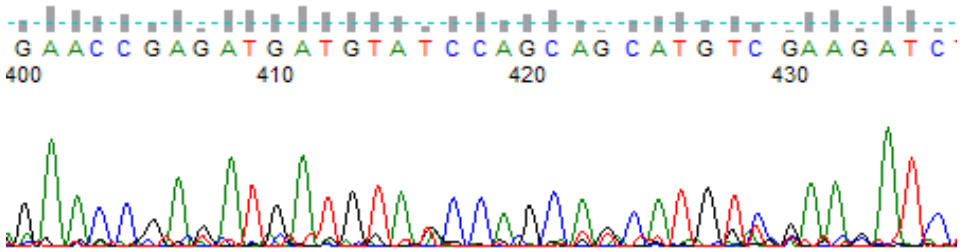
Excess Free Dye ("dye blobs" peaks)



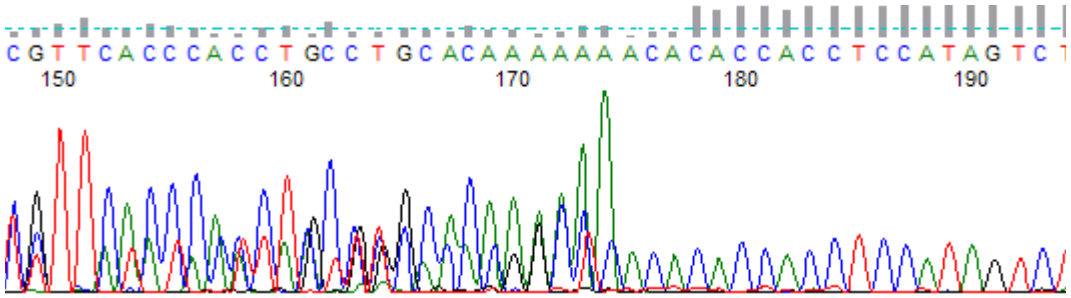
Mixed Trace Signal (multiple peaks)



Weak or "Noisy" Trace Peaks



Short Read Lengths or Poor Quality Basecalls



Primer Dimer Formation in the Sequencing Reaction

Data (NGS OUTPUT)

FASTQ

- Fasta with quality information

```
@HWI-EAS225:3:1:2:854#0/1
GGGGGGAAGTCGGCAAAATAGATCCGTAACTTCGGG
+HWI-EAS225:3:1:2:854#0/1
a`abbbbabaabbbababb^``[aaa`_N]b^ab^```a
@HWI-EAS225:3:1:2:1595#0/1
GGGAAGATCTCAAAAACAGAAGTAAAACATCGAACG
+HWI-EAS225:3:1:2:1595#0/1
a`abbbababbbabbbbbbabb`aaababab\aa_`
```

Paired End Reads

Read Set 1

```
@SRR1119204.1 SALLY:346:C2AFRACXX:4:1101:1604:1915 length=101
NTGGGAAAACCCAGATGAAAAGATTACTAGCGAATACGGTGCGTGTGGCAATTTCATCACCAATCGATCCATAAGTGCATTAATAAGTAAATCATA
+SRR1119204.1 SALLY:346:C2AFRACXX:4:1101:1604:1915 length=101
#4=DFFFFGFHHDIGJJIJJJJIJIJJJJJIJJJJGIIDHI(<BBEFAFEHIHEHHEHGHHFD9BCDADCDCCCCCDDEEDCEDACCCACDCCDD
@SRR1119204.2 SALLY:346:C2AFRACXX:4:1101:2019:1912 length=101
NTAATCTTCACCGATTCCGGGGAGCTATTCATATCGAGTTCAATTTTTTTCGCAGAGGATTCATGATCATTTTTTCTCGCAATTTTGGAGCCTTCT
+SRR1119204.2 SALLY:346:C2AFRACXX:4:1101:2019:1912 length=101
#1=DFFFFHHHHHFIJJJJJJJHIJJJJJJJJJJJJGIGIJICHJJJJJFFFDDDDDD5?CDDEDDDDEDDEEDDCDDDDBDDDDDDDCDDDDDDD
```

Read Set 2

```
@SRR1119204.1 SALLY:346:C2AFRACXX:4:1101:1604:1915 length=101
TATATAGTAAGTGAAATTCTCGAAGAAATGGCTCTAATGGAAAACAGAATTGGCAGCGTGCAGGAGGCCAGGATGATTGATATAAACGTCAATAAC
+SRR1119204.1 SALLY:346:C2AFRACXX:4:1101:1604:1915 length=101
B@@FDFFFHHHHFCHGGJEJJIIIIIJIJJJJJJJGEIJJIGGGIIEIJJHIJDHJIJJIJIJJHHHHHHFFFFEEFFEEEDDEEEDDDDBDFDDC
@SRR1119204.2 SALLY:346:C2AFRACXX:4:1101:2019:1912 length=101
GCAGAAAACAAACTGTTAATGAAGGAAAATTCAACGACAAGCGAGCAATCGAAATCGAAGCCTAACTTGTCTAACAAATTTATAAGCTCAATGAAT
+SRR1119204.2 SALLY:346:C2AFRACXX:4:1101:2019:1912 length=101
CCCFFFFFHHHHHJJGIJIJJJJJJJJJJJJJJJJIJIJJIIJIIJJJIJHHIJJJJHHHFFFFEEEEEEEDEDDDDDDDDDEEEDDDDDDDDDDE
```

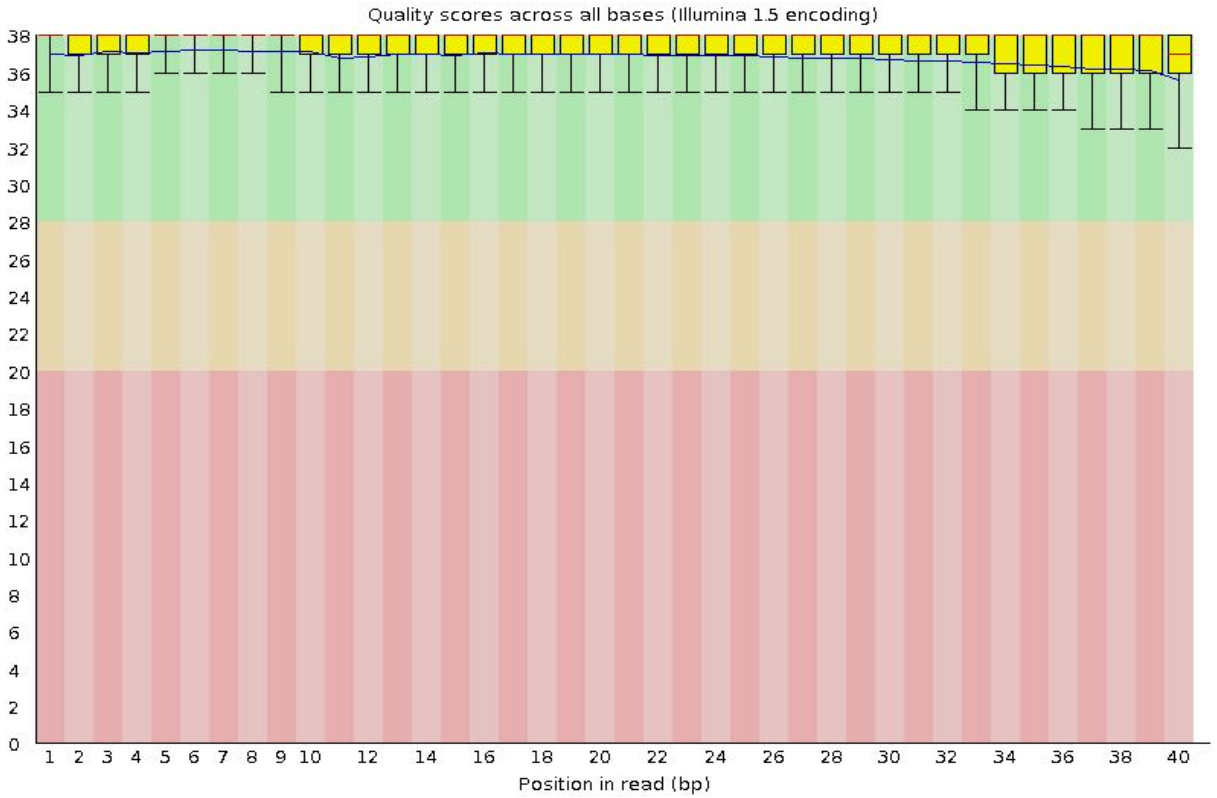
NGS : Good (Illumina) Sequence Data

FastQC Report

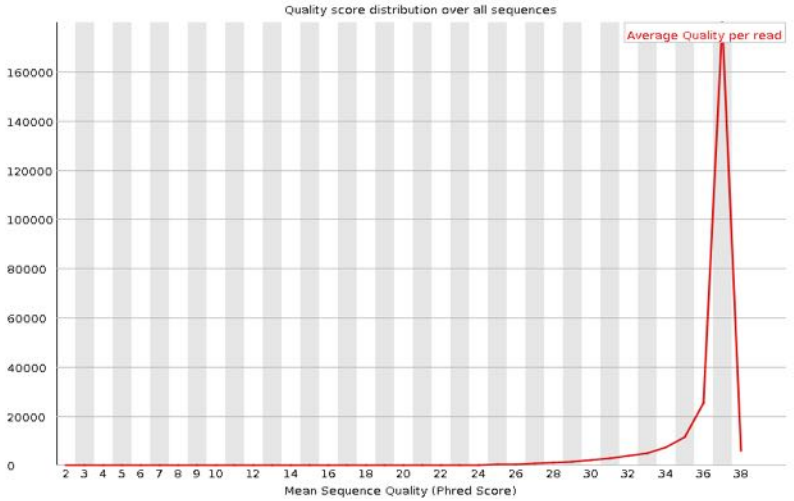
Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

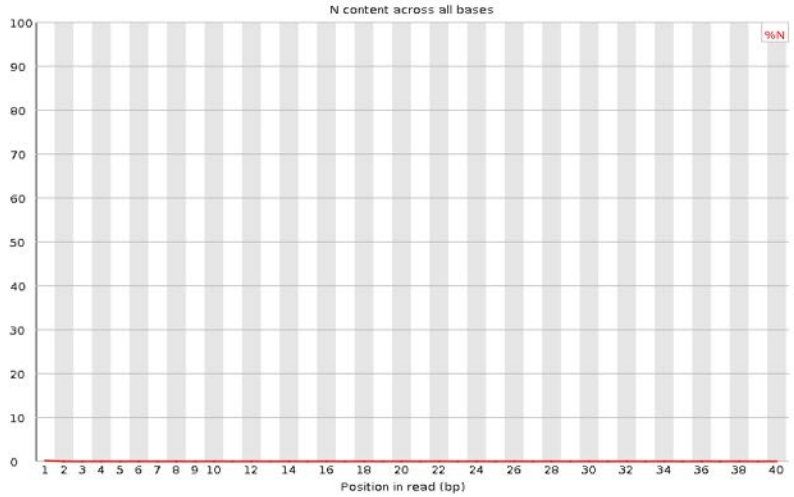
Per base sequence quality



Per sequence quality scores



Per base N content



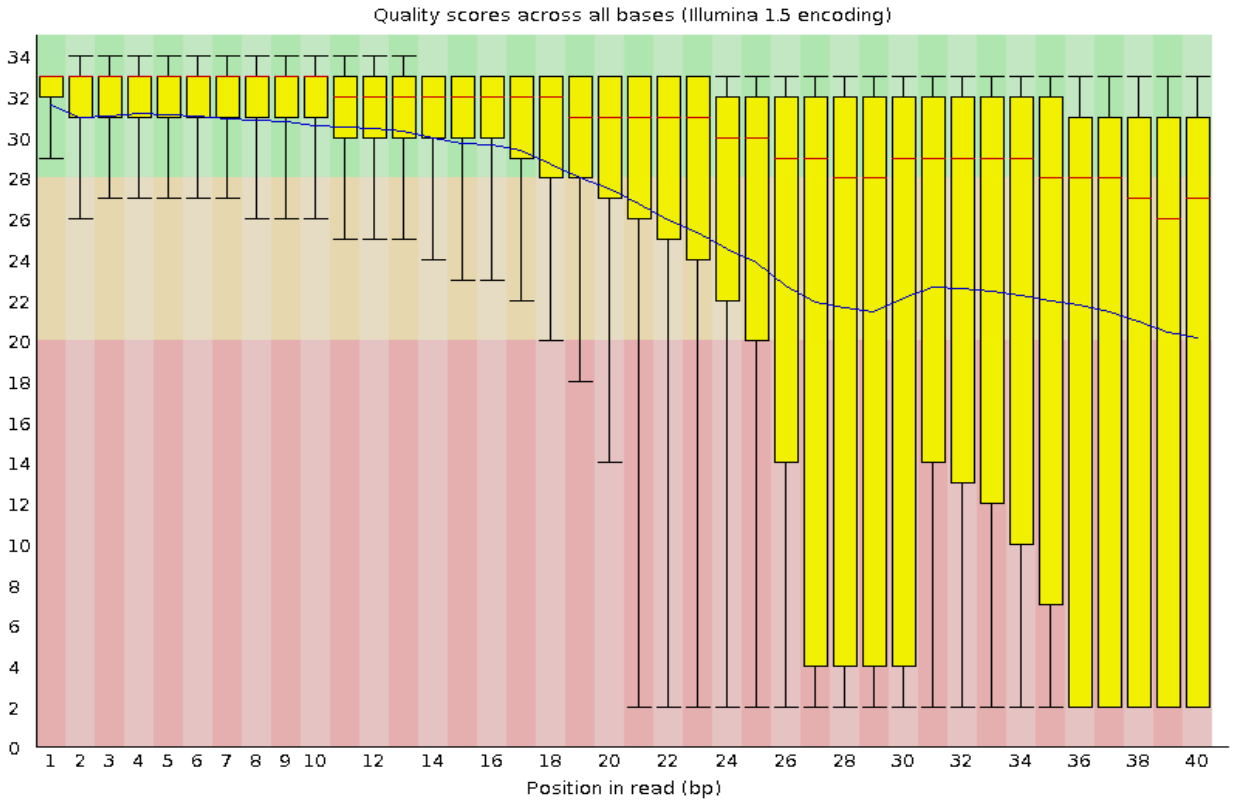
NGS : Poor (Illumina) Sequence Data

FastQC Report

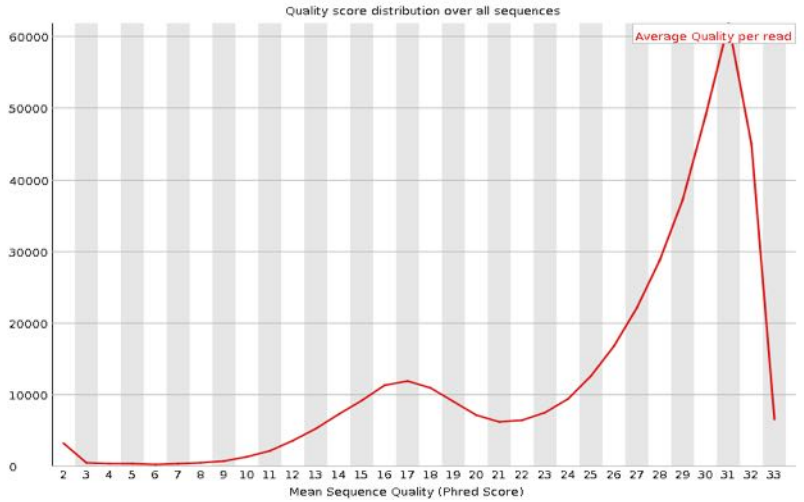
Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

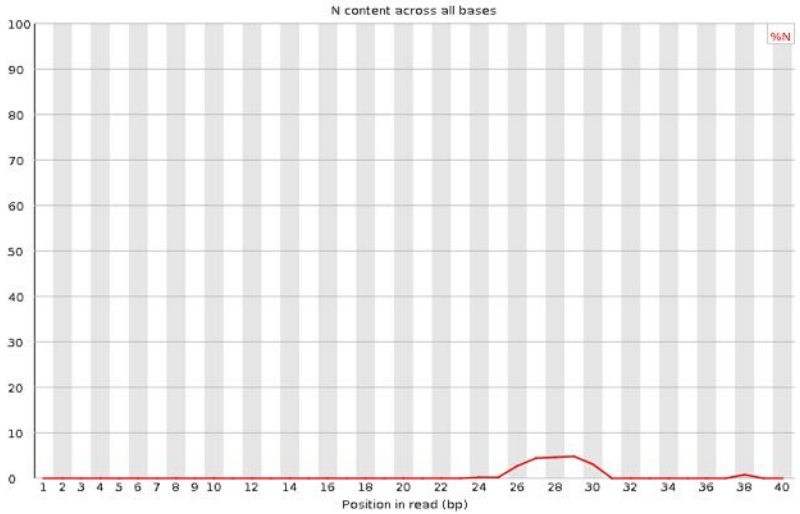
Per base sequence quality



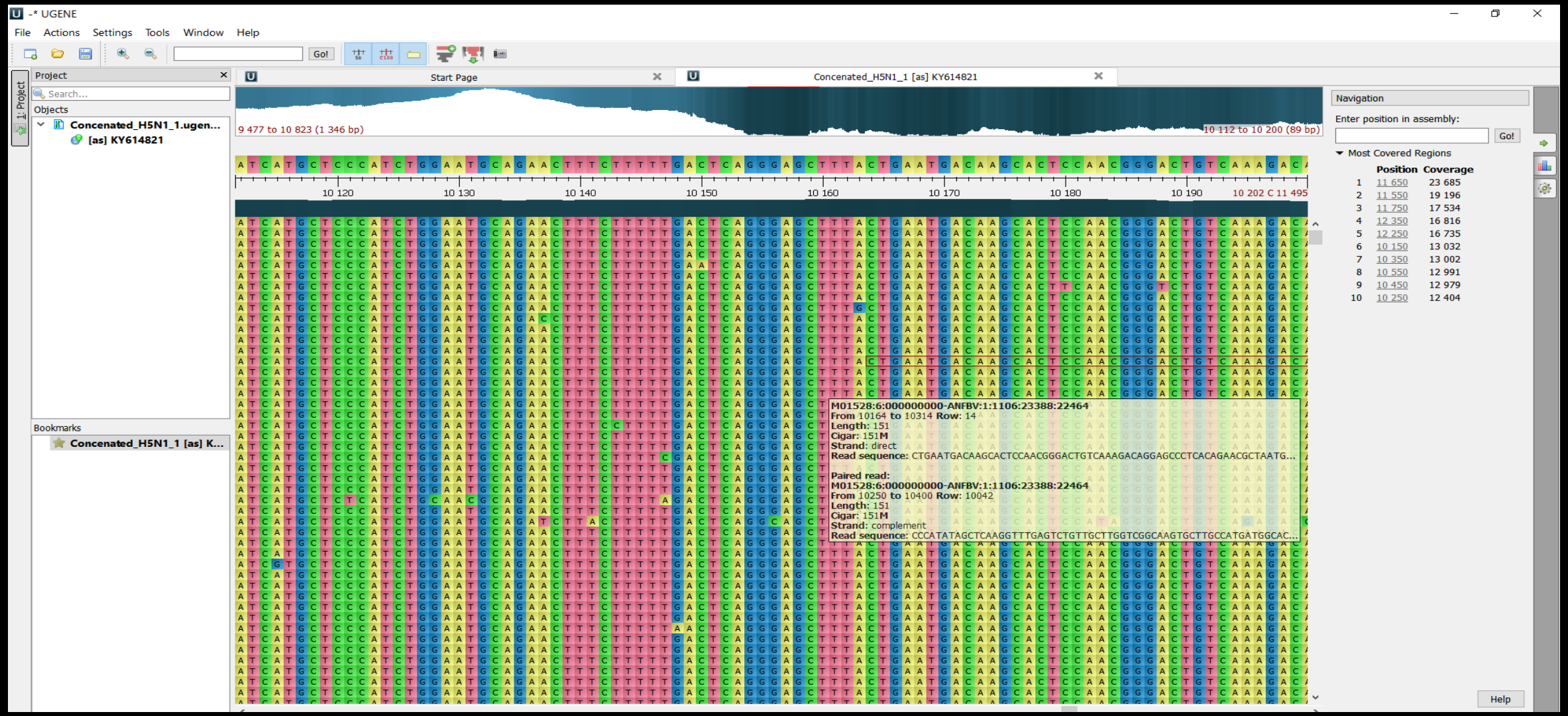
Per sequence quality scores



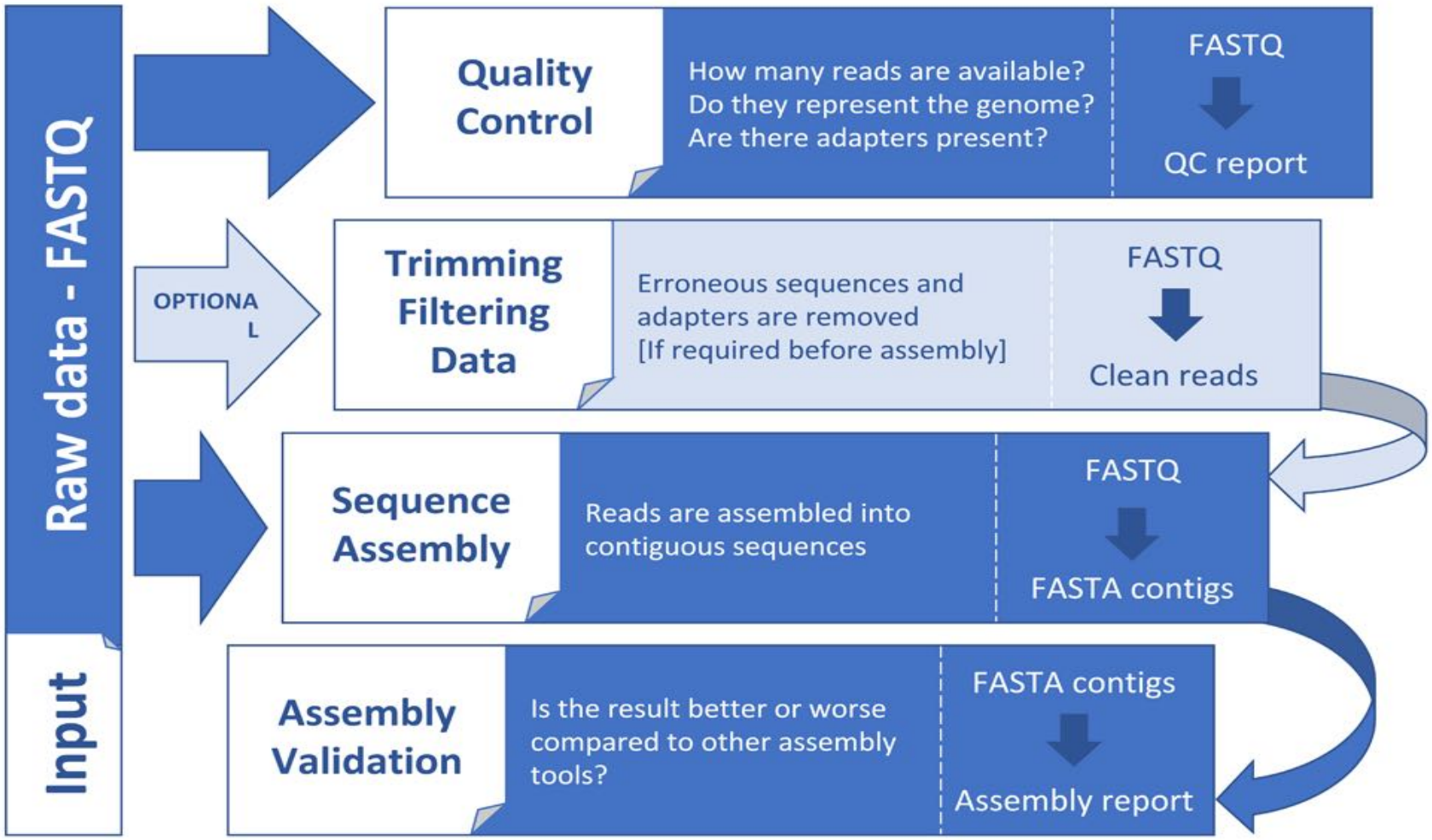
Per base N content



ANALYSIS OF NGS OUTPUT

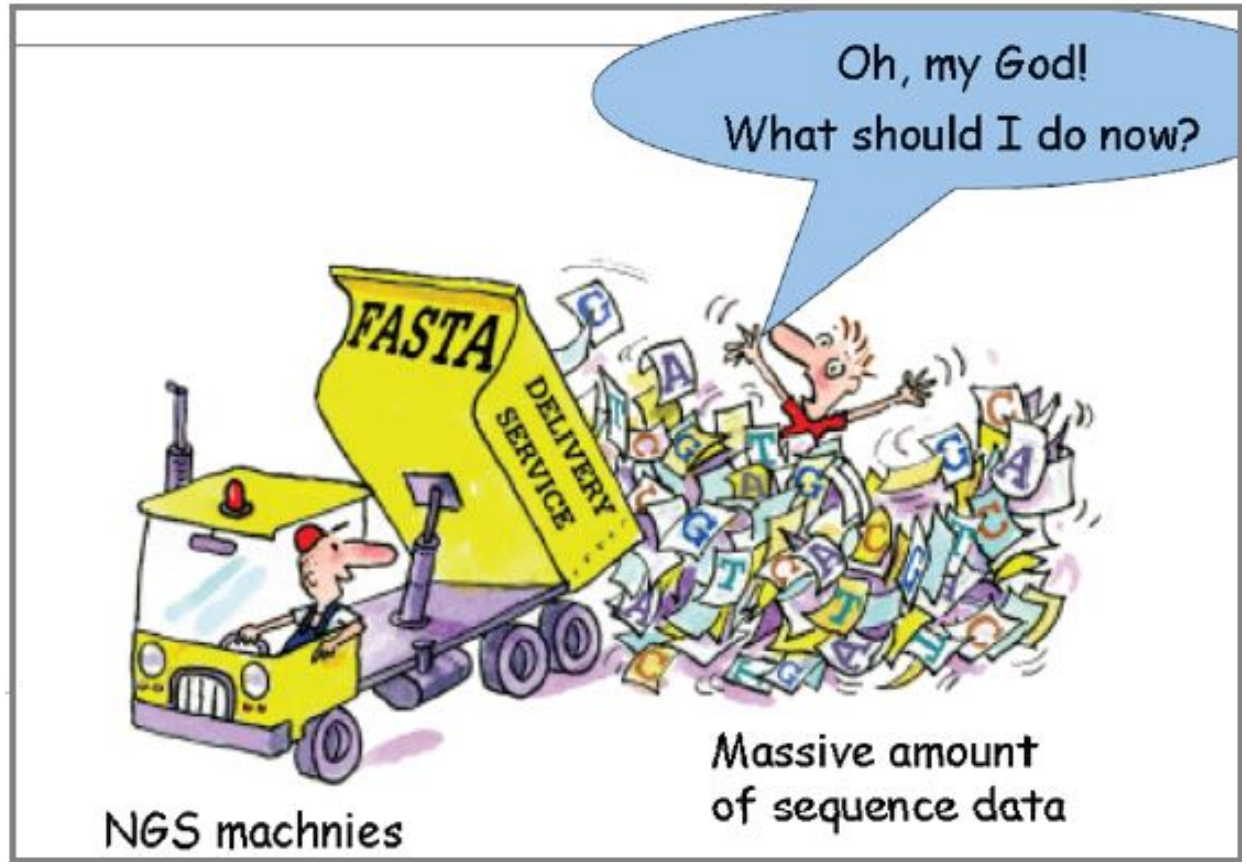


NGS Genome Assembly Workflow



Dominguez Del Angel V, Hjerde E, Sterck L et al. Ten steps to get started in Genome Assembly and Annotation [version 1]. F1000Research 2018, 7:148 (doi: 10.12688/f1000research.13598.1)

CHALLENGES



For genome analysis is cost effective, but reagents are still expensive.

How to tackle computational challenges:

- Output files are too large
- Storage problem
- Data management and quality control
- Specialized person to analysis data

Final Thoughts

- DNA sequencing is becoming vastly faster and more affordable
- Generating data is no longer the bottleneck, but understanding it is.
- Bioinformatics types should be in high demand in the near future