Genomics data analysis

VIT-TBI

Gomathinayagam
Dietome Biotech
Vellore Institute of Technology-TBI

Learning outcomes

Session 1 – **1.5 to 2 Hrs**

- Setting up Linux environment for genomics analysis
- General presentation of genomic analysis(WGS, microbiome analysis, transcriptomics etc.,)
- Introduction to genomics data file types
- Basic Linux commands
- Setting up minimal tools and resources for prokaryotic genome analysis.
- Walk through into github repositories for genomics data

Genome is like a message in a bottle!

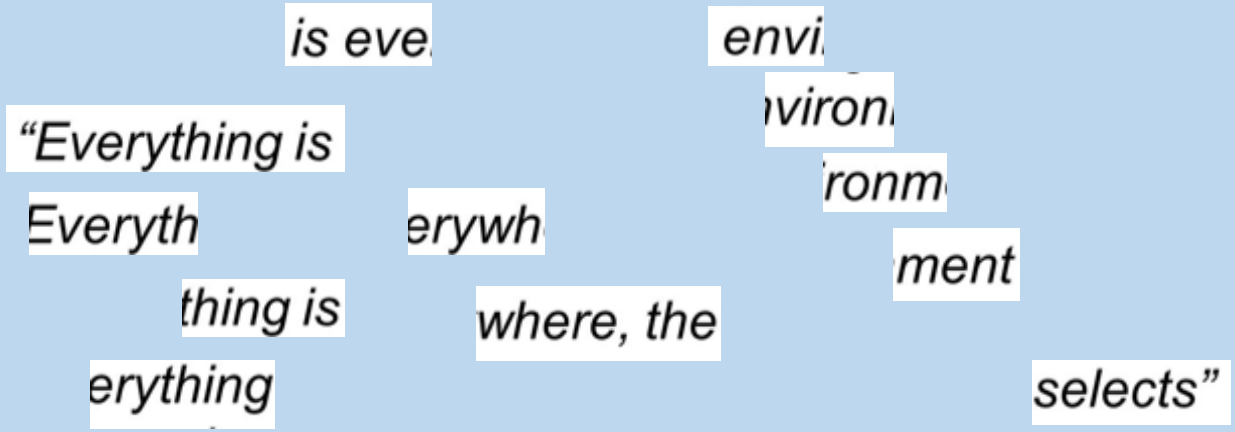
“Everything is everywhere, the environment selects”

erything erywh onmen selects”

thing is Everyth where, the ronm ment

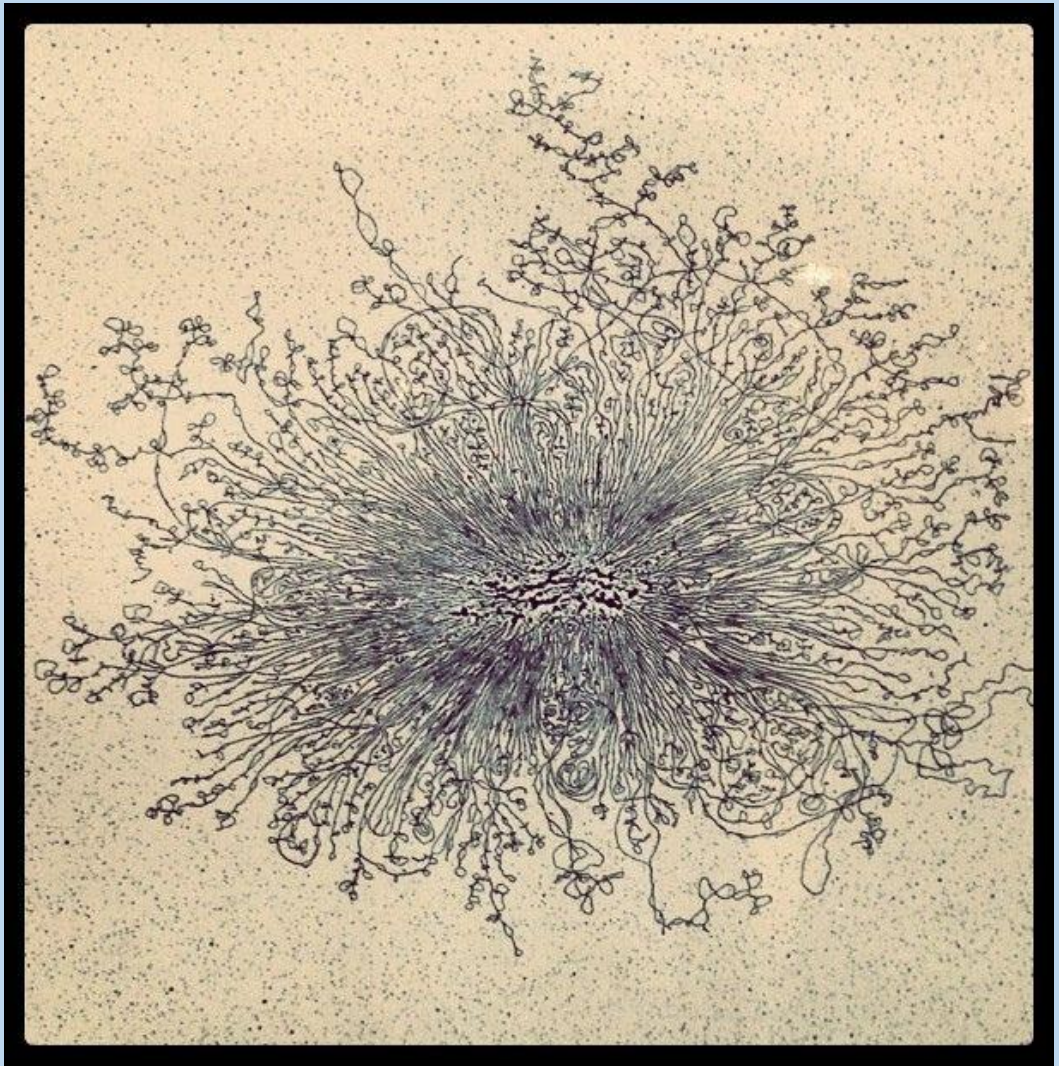
envi “Everything is ment sel is eve rywh viron

How do we solve a jumble riddle?



"Everything is everywhere, the environment selects"

This is what happens when we try to understand any genome!



A workflow for whole genome sequencing (WGS) of individual genomes

Sample collection QC Library preparation Platform choice (Hybrid options available) Computational analysis involving assembly, annotation etc.

99% of sequence analysis is on the command line (Linux or Mac)

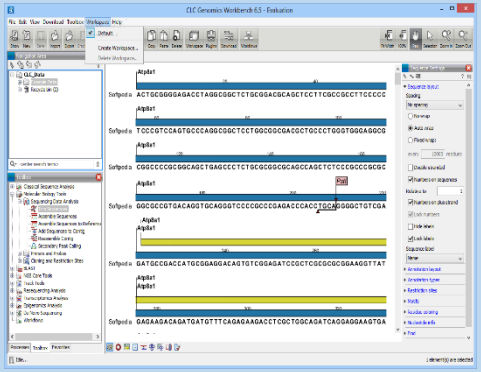
Most next-generation sequence (NGS) analysis is done on the command line. Command line software (using Linux or the Unix-like platform on a Mac terminal) is capable of handling the data analysis tasks, and most NGS software is written for the Unix operating system.

Many people access a Linux (or related Unix) environment while working on a PC or Mac. For example, you can do "cloud computing" in which you pay someone (Amazon, Google, Microsoft) to access their servers.

Computational analysis

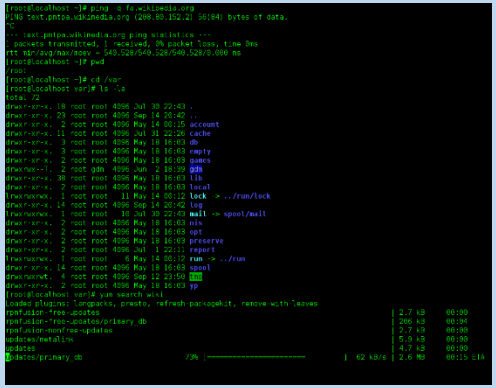
GUI

Graphic User Interface



CLI

Command line Interface



Which one is comfortable?

Which one would you prefer?

Setting up Linux environment…

A workflow for whole genome sequencing (WGS) of individual genomes

Sample collection QC Library preparation Platform choice (Hybrid options available) Computational analysis involving assembly, annotation etc.

Sample collection and nucleic acid purification

- First step in sequencing project- determine the type of sample we are going to work with.
- It could be bacterial organism, higher order organisms-fungi,plant, insects, mammals
- Or it could be eDNA….?
- eDNA- Environmental DNA- isolation DNA/RNA from various environments like human gut, animal gut, deep ocean sediments, antartica soil, and soon from ‘**martian soil**’.

Sample collection and nucleic acid purification

- Or if we want to do transcriptomics- isolate pure RNA
- Then convert them into complementary DNA using reverse transcriptase.
- Now again we get DNA to sequence.
- With leaping growth in technology, it is now possible to sequence RNA directly without conversion

Sample collection and nucleic acid purification

- There could be special cases in eDNA sequencing and transcriptome
- eDNA may contain genomes of multiple organism- hence it is called meta-genome
- eDNA sequencing helps in understanding the community of organisms present in an environment
- We can concentrate on particular community also i.e do 16S PCR method- we will get to know the bacterial composition present in an environment.

Sample collection and nucleic acid purification

- Whereas with transcriptomics you might want to know the gene expression of a particular population
- Or…
- The current technological developments made it possible to sequence transcriptome of every single cell, except that it is exhaustive.
- For this we need Fluorescence Assisted Cell Sorter in addition to sequencing platform.

Sample collection and nucleic acid purification

- From all of the above sources we need to isolate DNA/RNA with a particular method which gives the nucleic acids in adequate purity and quantity.
- Especially eDNA might contain other contaminants from its environment, some plant sourced nucleic acids can have other sequencing inhibitors.
- So it is highly essential to remove all contaminants from the DNA/RNA through spectrometry based Quality control checks

Outline:
Analysis of Next-Generation Sequence (NGS) Data

DNA sequencing technologies

First generation

- Sanger sequencing;
- Maxam-Gilbert Sequencing;

Next generation sequencing technologies

- Illumina;
- Pyrosequencing;
- ABI SOLiD;
- Ion Torrent;
- Pac Bio;
- Nanopore,
- DNB seq.

Next-generation sequence technologies

Technology	Read length (bp)	Reads per run	Time per run	Cost per megabase	Accuracy
Roche 454	700	1 million	1 day	$10	99.9%
Illumina	50-250	<3 billion	1-10 days	~$0.10	98%
SOLiD	50	~1.4 billion	7-14 days	$0.13	99.9%
Ion Torrent	200	<5 million	2 hours	$1	98%
Pacific Biosciences	2900	<75,000	<2 hours	$2	99%
Sanger	400-900	N/A	<3 hours	$2400	99.9%

Source: adapted from Wikipedia 1/11/16

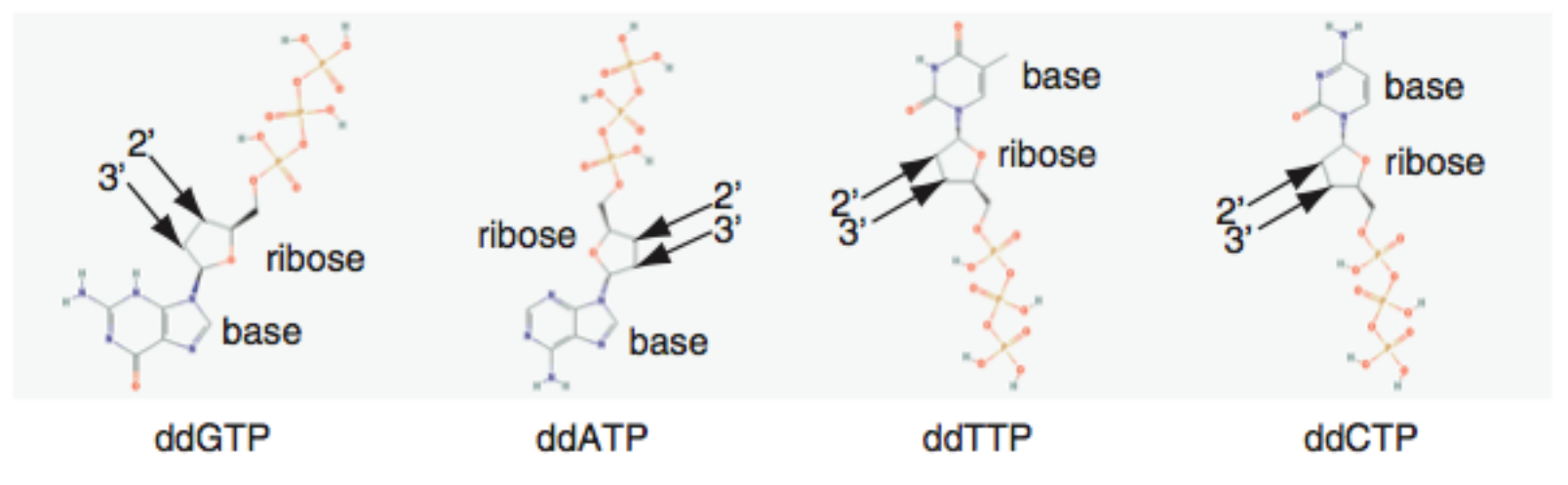
Sanger sequencing: what we had before NGS

Introduced in 1977

A template is denatured to form single strands, and extended with a polymerase in the presence of dideoxynucleotides (ddNTPs) that cause chain termination.

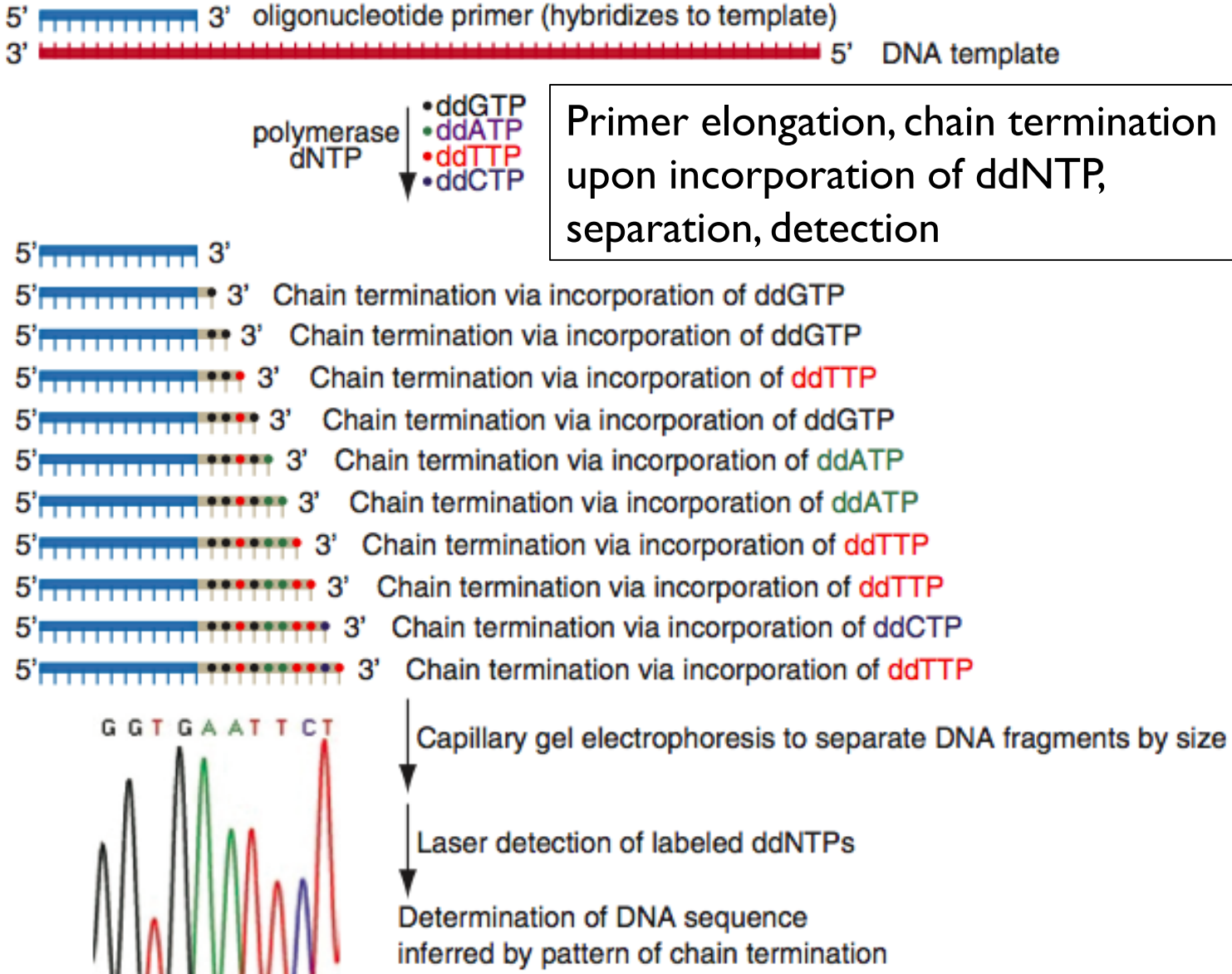
Typical read lengths are up to 800 base pairs. For the sequencing of Craig Venter's genome (2007; first whole genome of an individual), Sanger sequencing was employed because of its relatively long read lengths.

DNA sequencing by the Sanger method



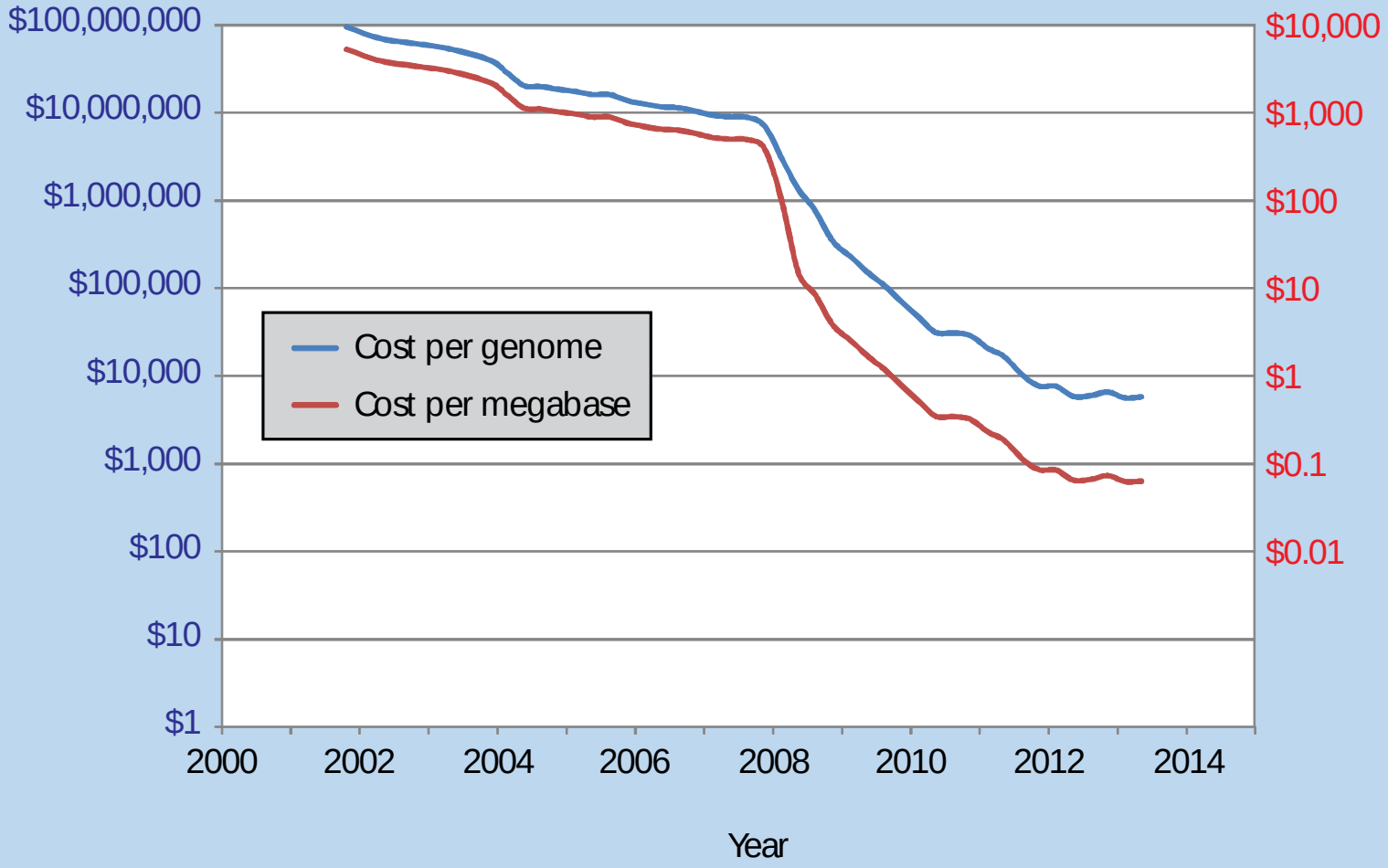
Dideoxynucleotides (ddNTPs)(-OH of dNTP is replaced by -H of ddNTP at the 2' ribose position)

DNA sequencing by the Sanger method

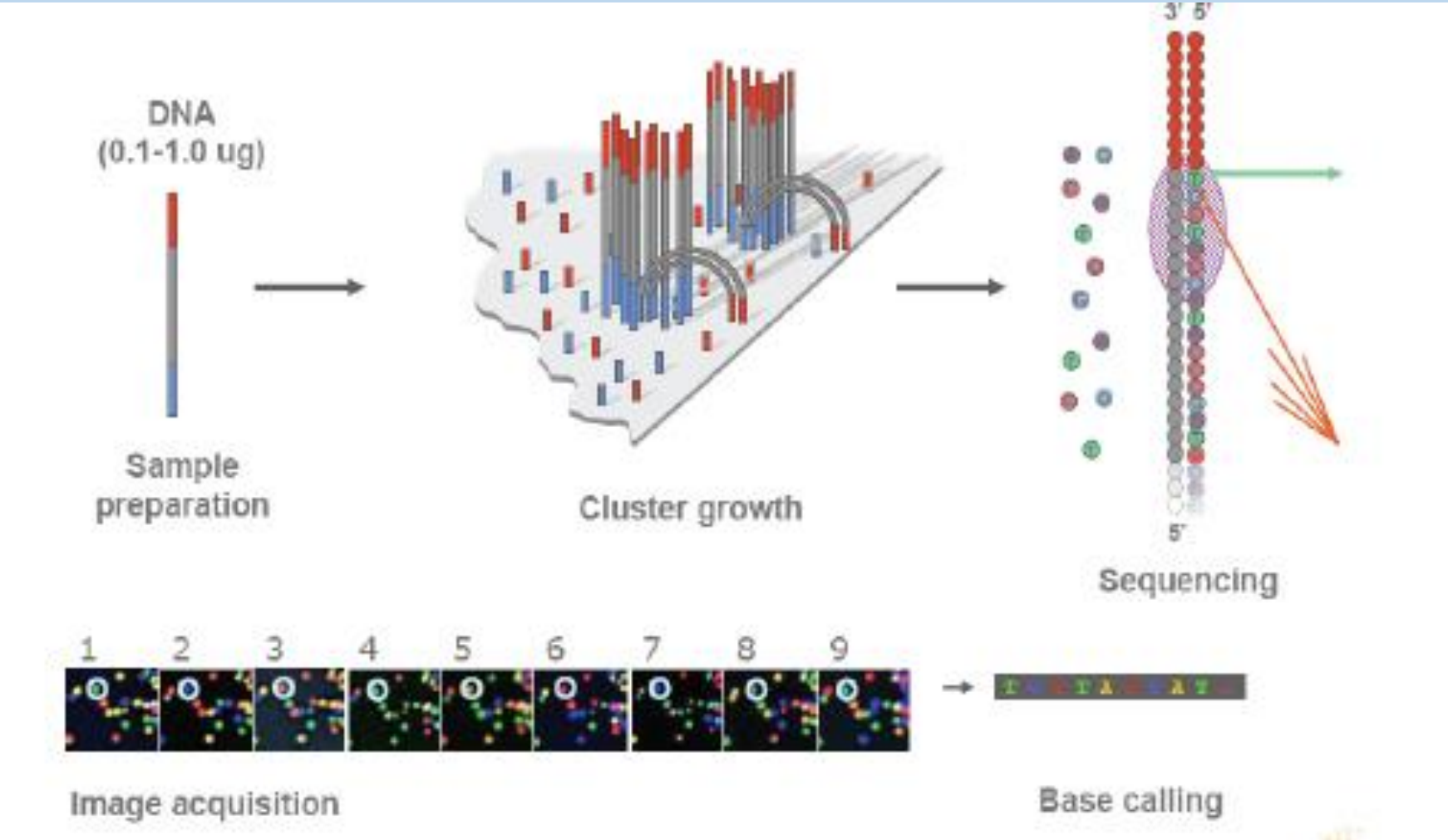


Primer elongation, chain termination upon incorporation of ddNTP, separation, detection

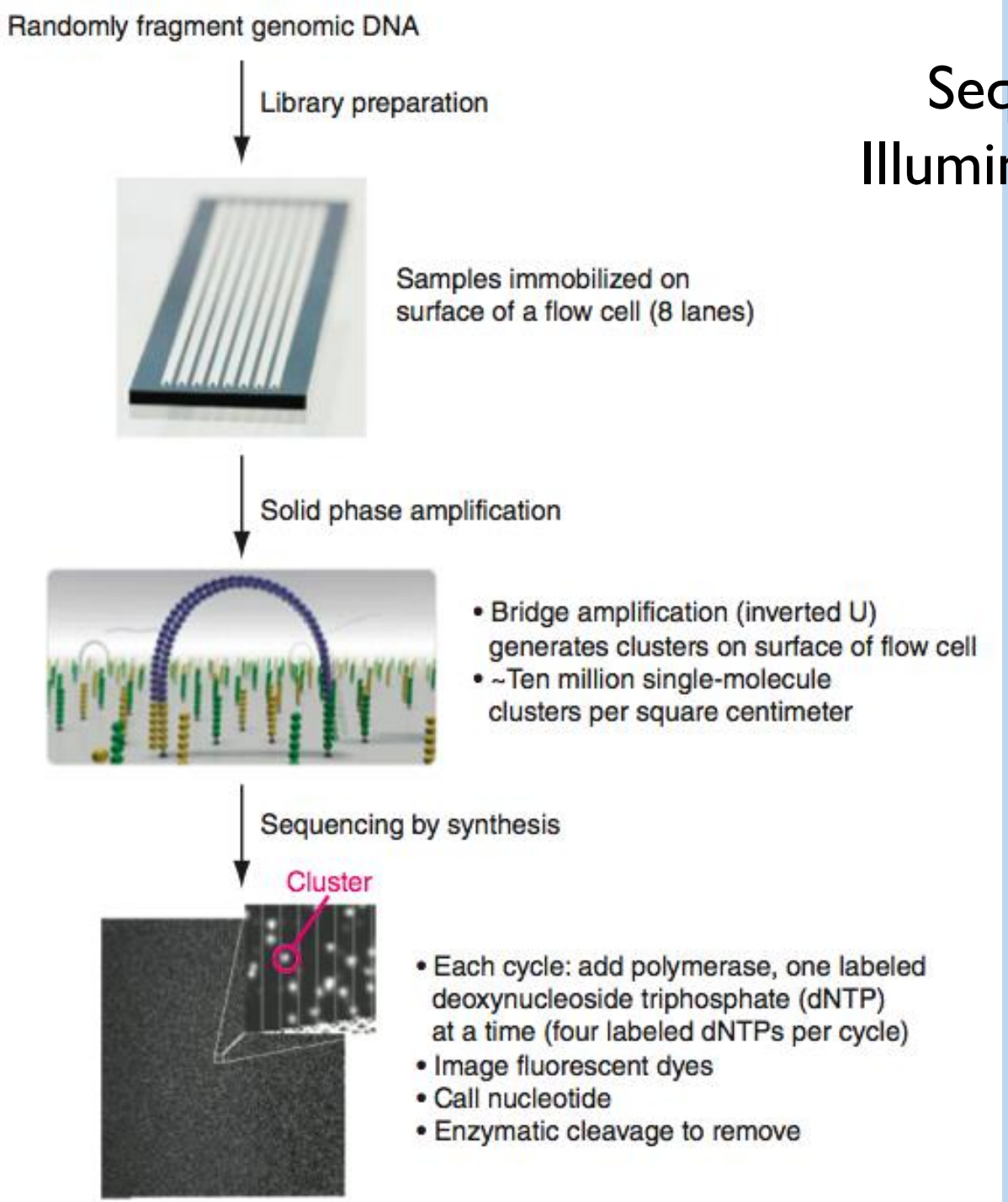
Whole genome sequencing (WGS) costs have declined dramatically



Next-generation sequence technology: Illumina



Sequencing by Illumina technology



B&FG 3e
Fig.9-4
Page 384

Cycle termination sequencing (Illumina)

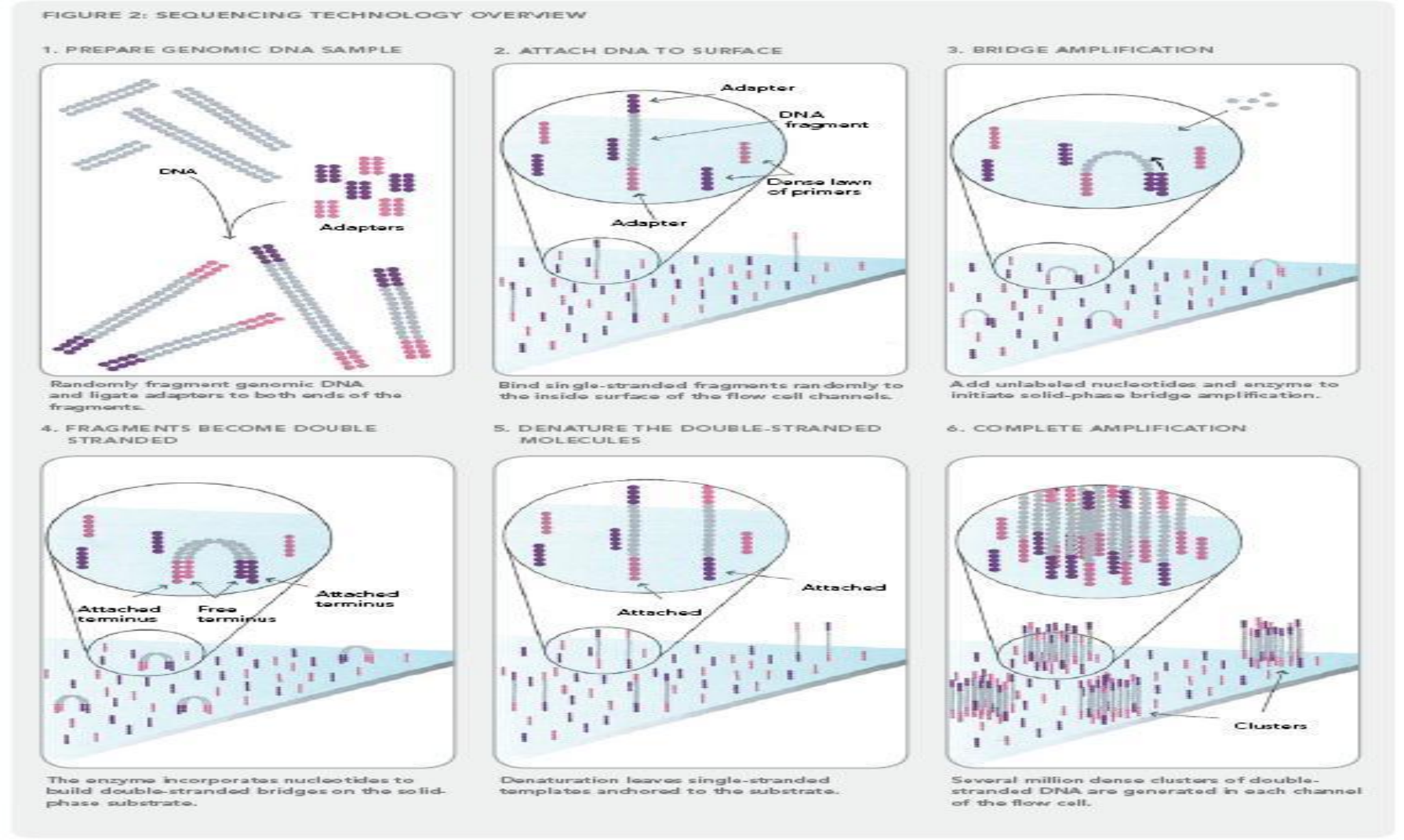
Disadvantage:

- Short read length (~150 bases)

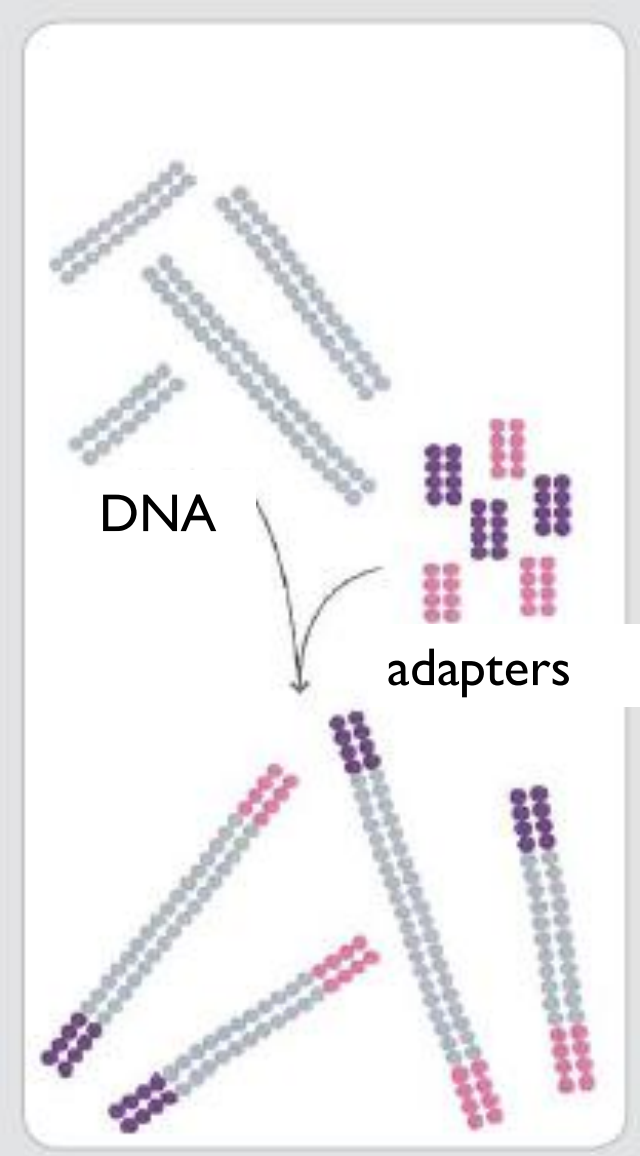
Advantages:

- Very fast
- Low cost per base
- Large throughput; up to 1 gigabase/epxeriment
- Short read length makes it appropriate for resequencing
- No need for gel electrophoresis
- High accuracy
- All four bases are present at each cycle, with sequential addition of dNTPs. This allows homopolymers to be accurately read.

Illumina sequencing technology in 12 steps

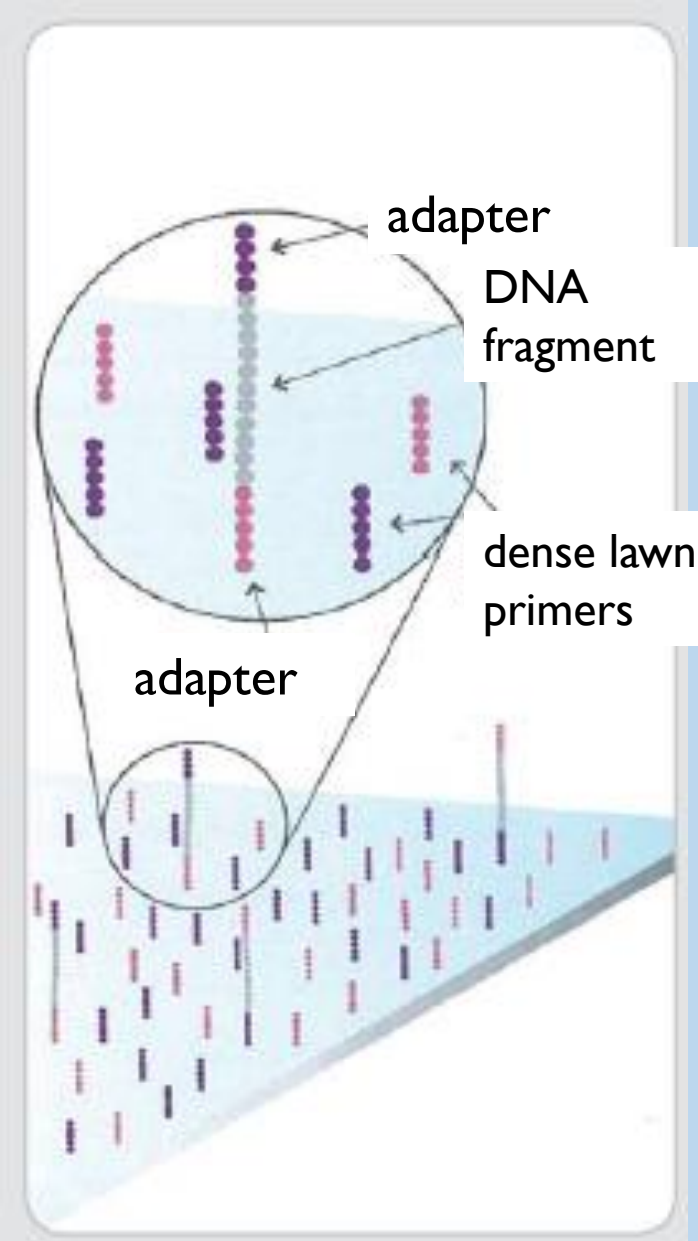


Source: http://www.illumina.com/downloads/SS_DNAsequencing.pdf



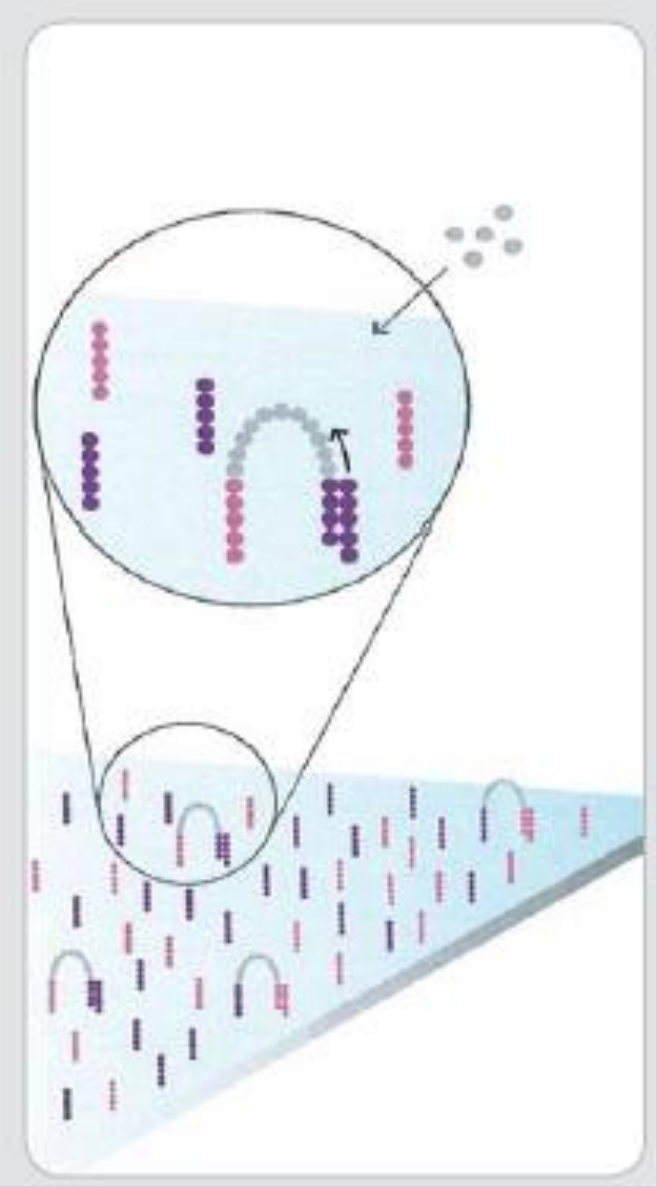
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments

1. Prepare genomic DNA
2. Attach DNA to surface
3. Bridge amplification
4. Fragments become double stranded
5. Denature the double-stranded molecules
6. Complete amplification



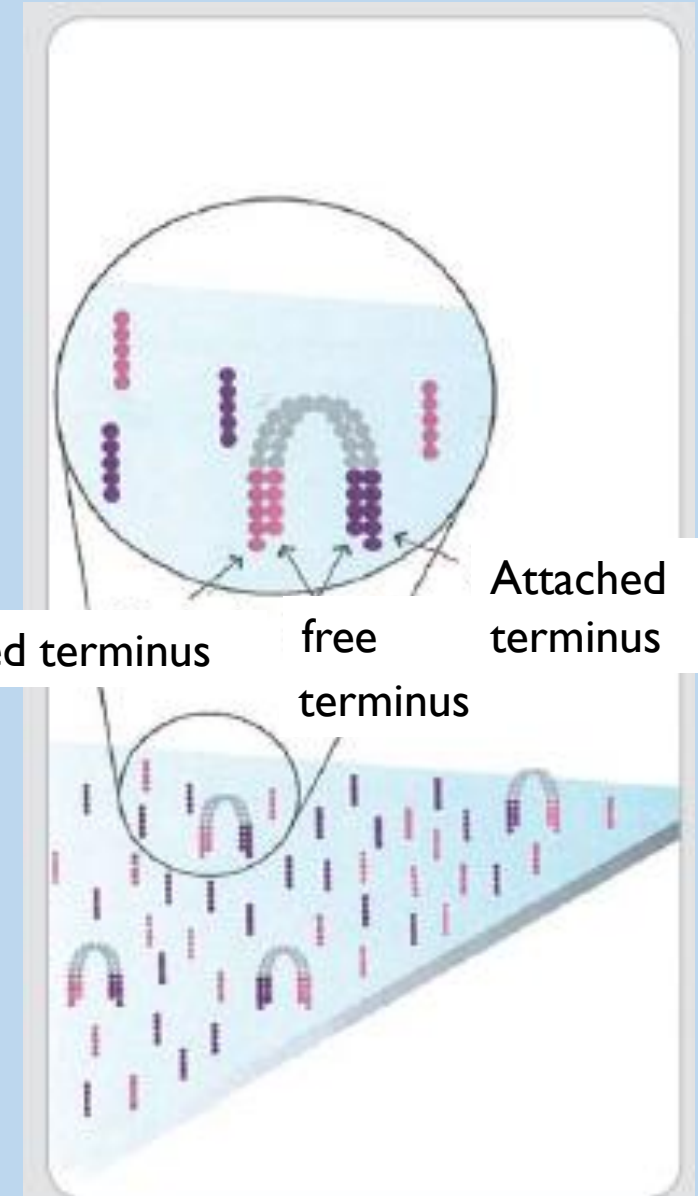
Bind single-stranded fragments randomly to the inside surface of the flow cell channels

1. Prepare genomic DNA
2. Attach DNA to surface
3. Bridge amplification
4. Fragments become double stranded
5. Denature the double-stranded molecules
6. Complete amplification



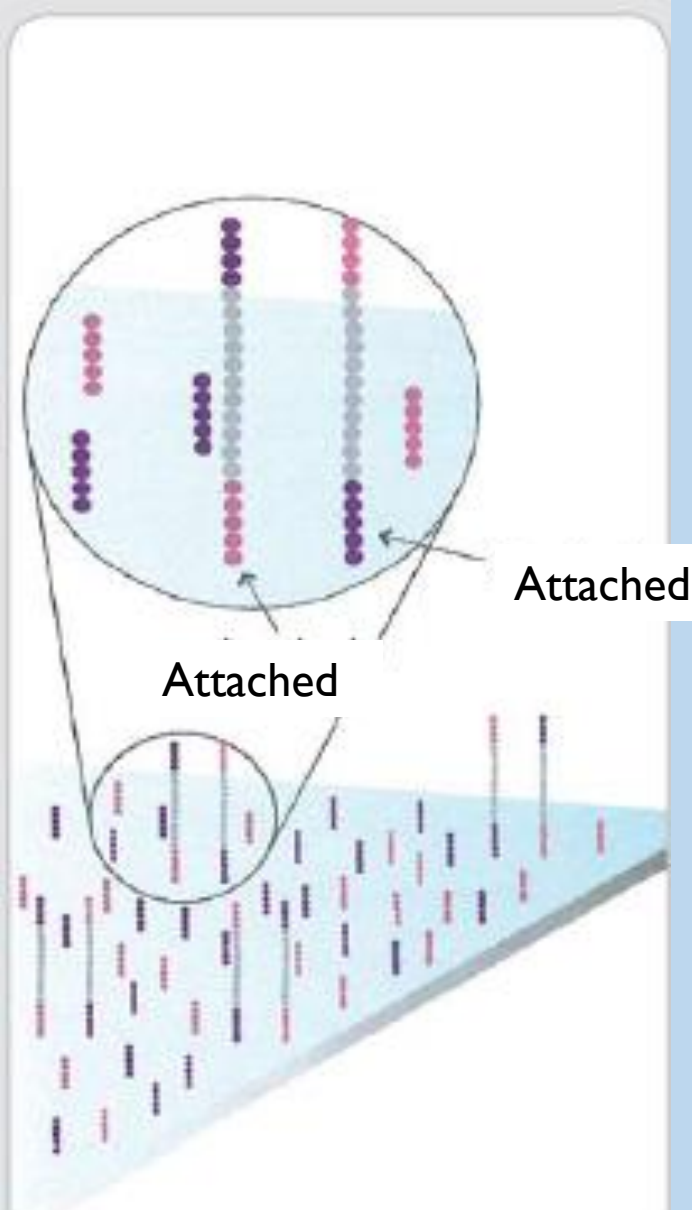
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification

1. Prepare genomic DNA
2. Attach DNA to surface
3. Bridge amplification
4. Fragments become double stranded
5. Denature the double-stranded molecules
6. Complete amplification



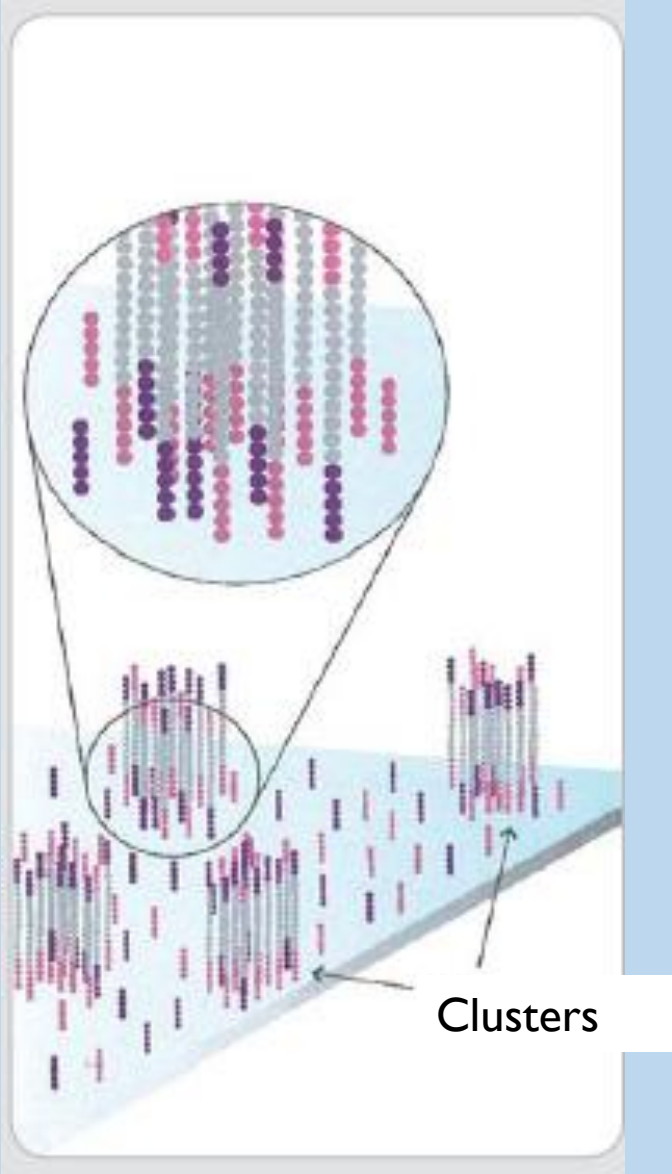
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate

1. Prepare genomic DNA
2. Attach DNA to surface
3. Bridge amplification
4. Fragments become double stranded
5. Denature the double-stranded molecules
6. Complete amplification



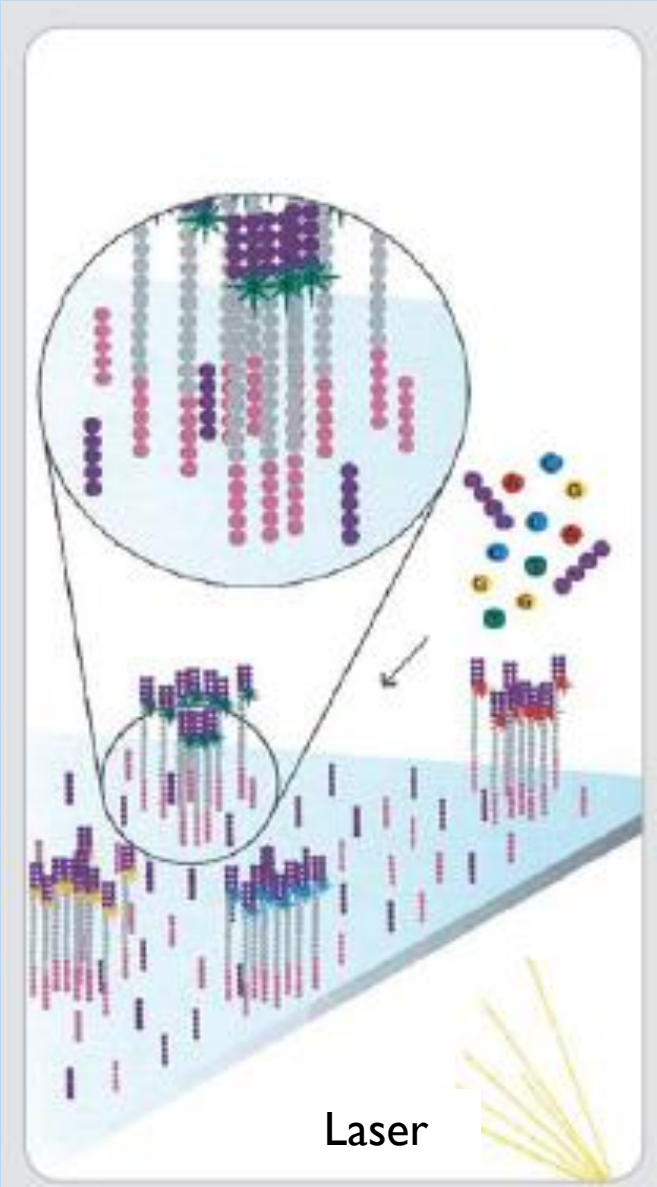
Denaturation leaves single-stranded templates anchored to the substrate

1. Prepare genomic DNA
2. Attach DNA to surface
3. Bridge amplification
4. Fragments become double stranded
5. Denature the double-stranded molecules
6. Complete amplification



Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell

1. Prepare genomic DNA
2. Attach DNA to surface
3. Bridge amplification
4. Fragments become double stranded
5. Denature the double-stranded molecules
6. Complete amplification



The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase

7. Determine first base

8. Image first base

9. Determine second base

10. Image second chemistry cycle

11. Sequencing over multiple chemistry cycles

12. Align data

After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified

7. Determine first base

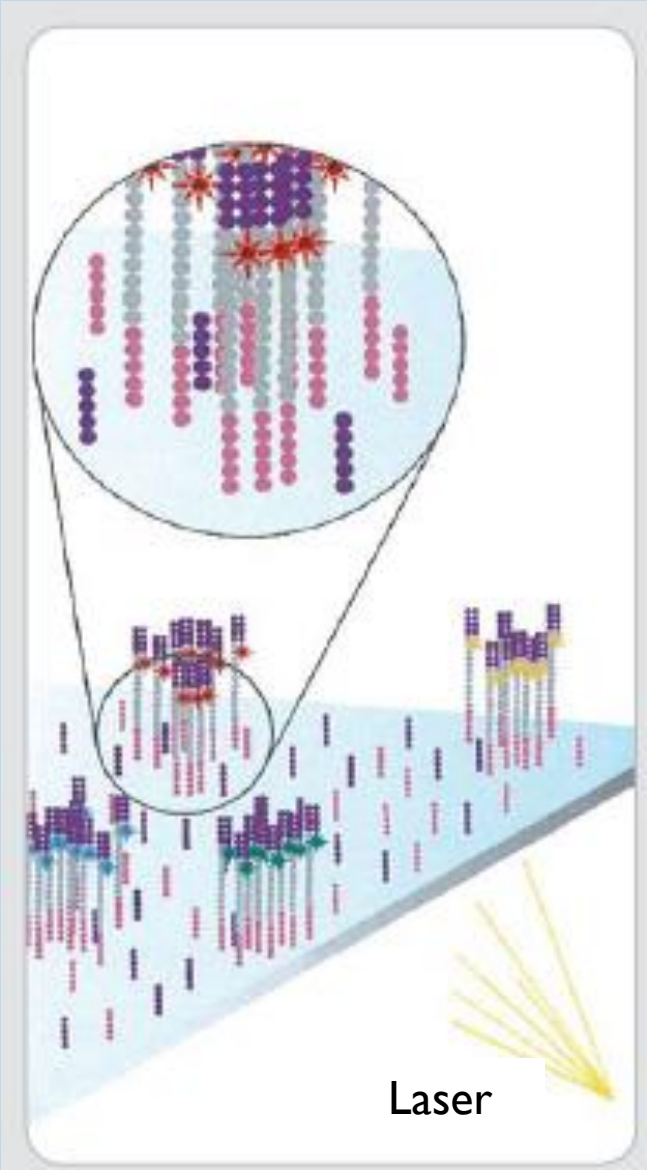
8. Image first base

9. Determine second base

10. Image second chemistry cycle

11. Sequencing over multiple chemistry cycles

12. Align data



The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase

7. Determine first base

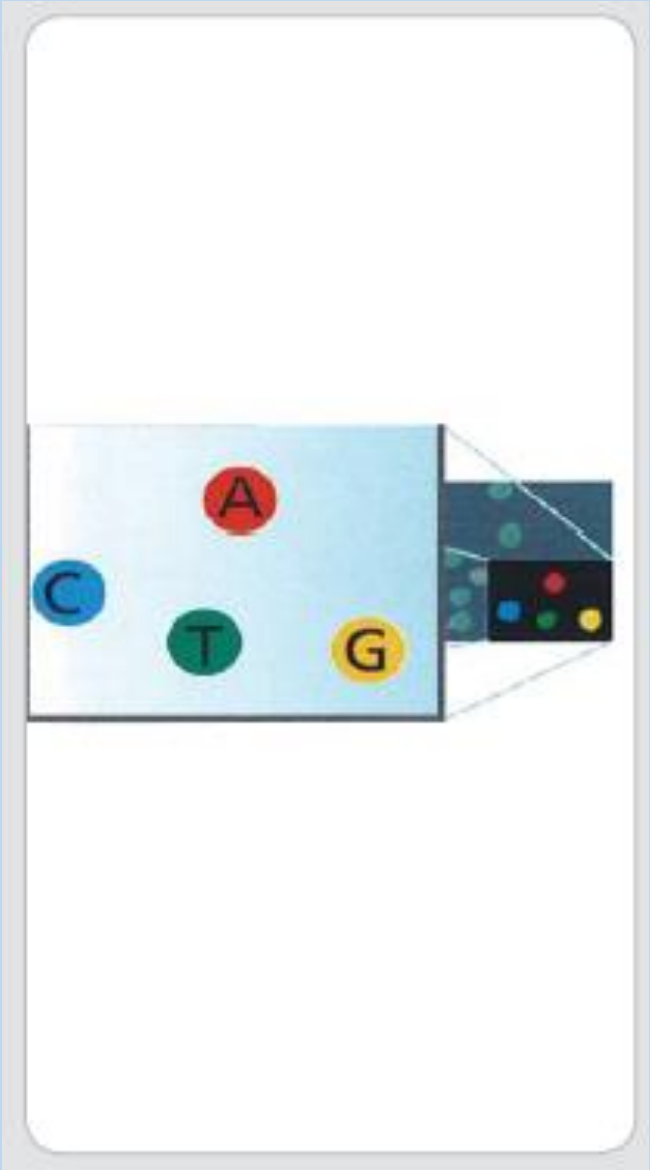
8. Image first base

9. Determine second base

10. Image second chemistry cycle

11. Sequencing over multiple chemistry cycles

12. Align data



After laser excitation the image is captured as before, and the identity of the second base is recorded.

7. Determine first base

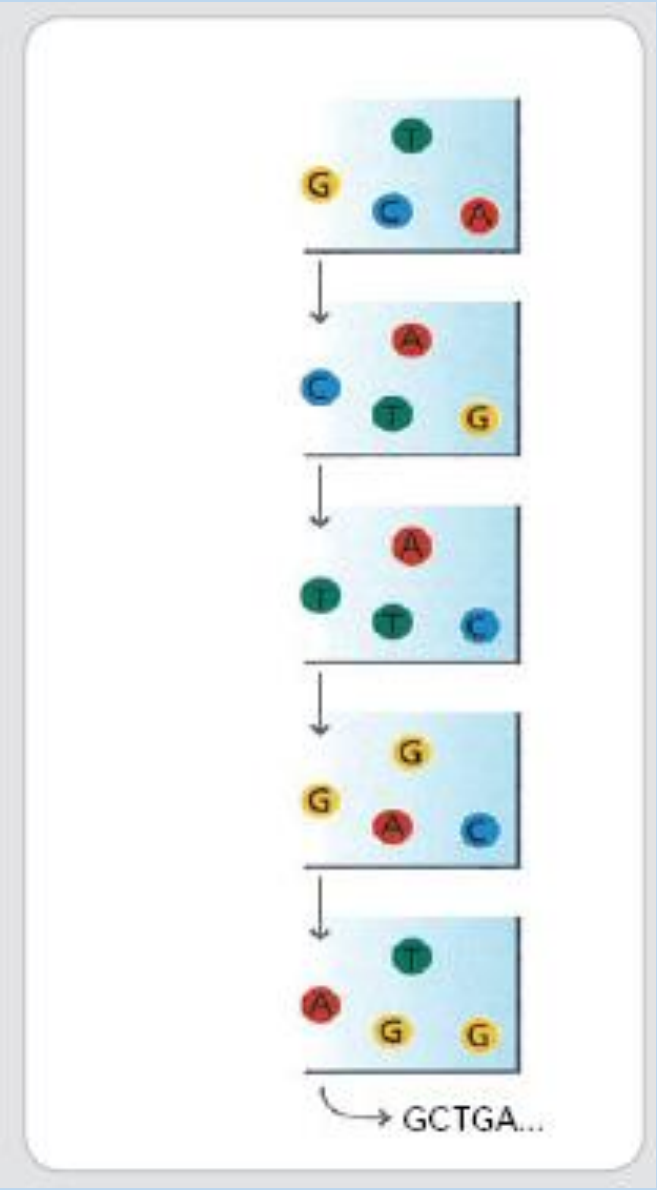
8. Image first base

9. Determine second base

10. Image second chemistry cycle

11. Sequencing over multiple chemistry cycles

12. Align data



The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

7. Determine first base

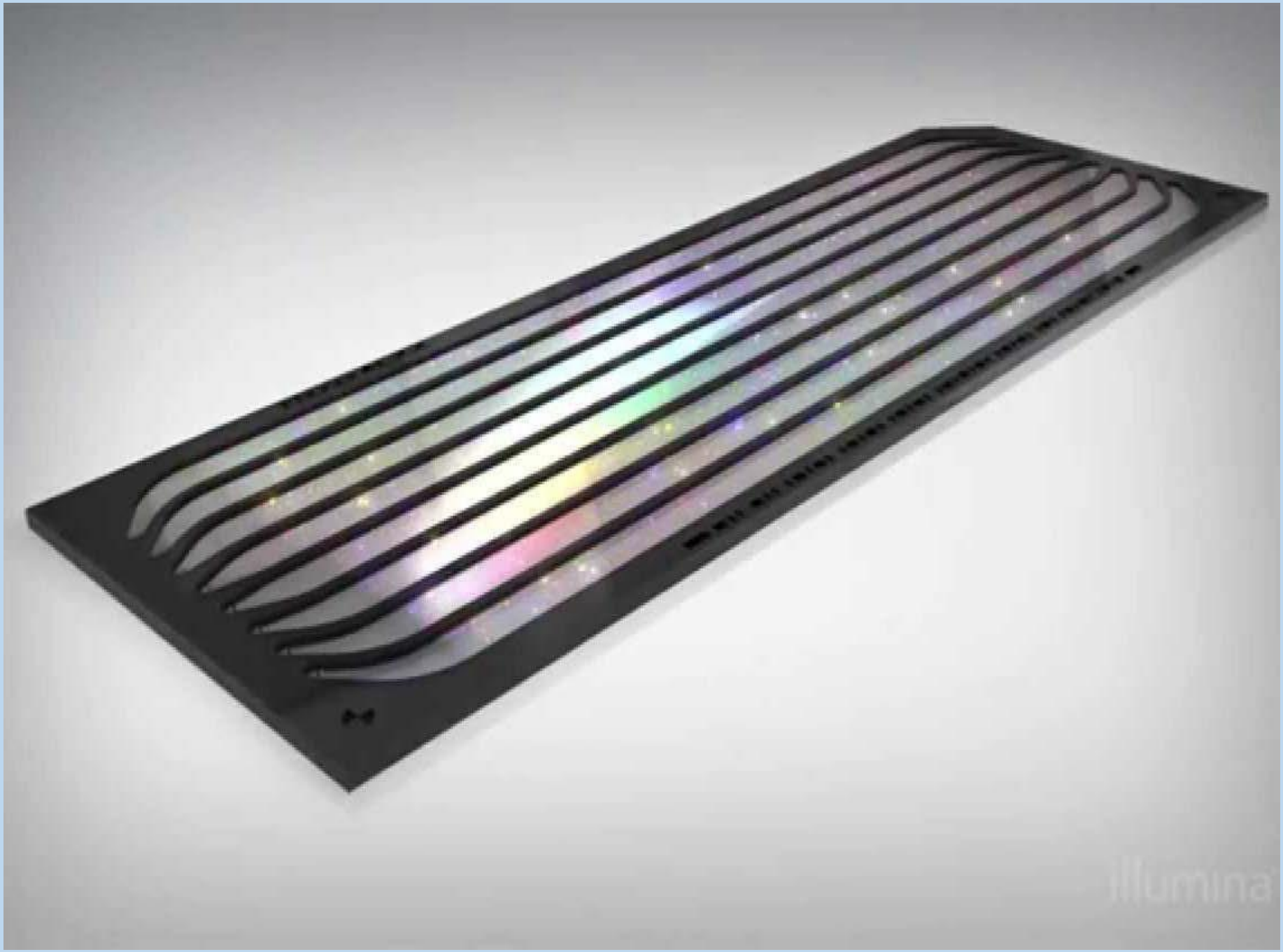
8. Image first base

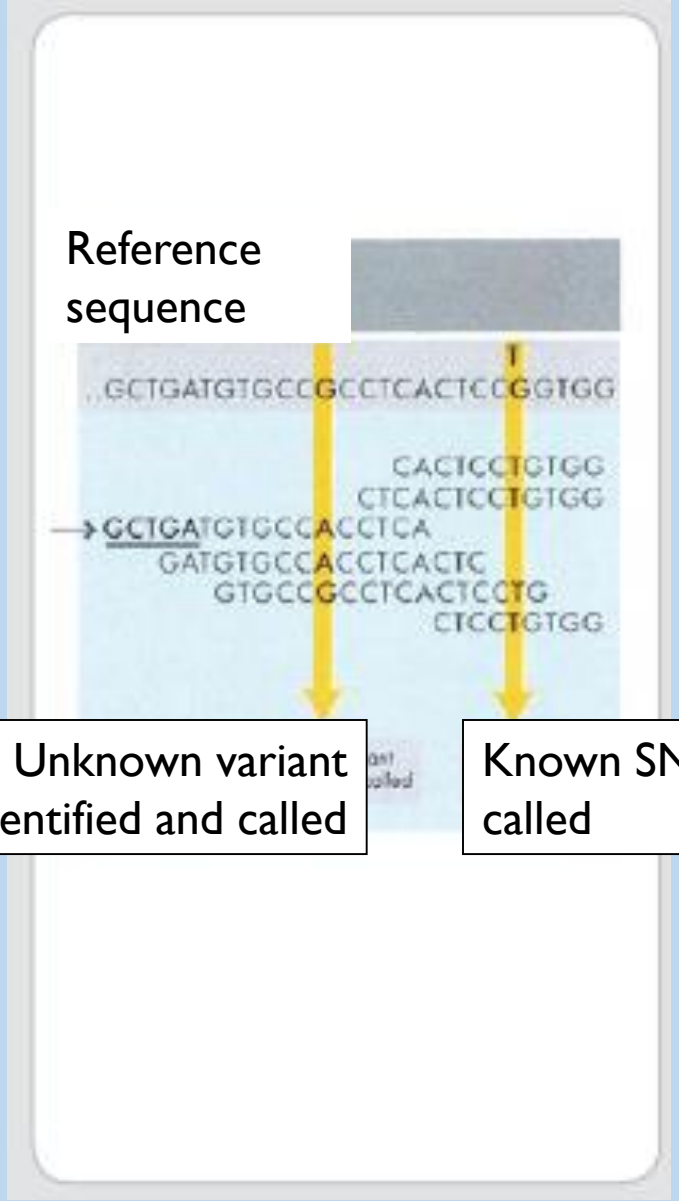
9. Determine second base

10. Image second chemistry cycle

11. Sequencing over multiple chemistry cycles

12. Align data


illumina



The data are aligned and compared to a reference, and sequencing differences are identified.

7. Determine first base

8. Image first base

9. Determine second base

10. Image second chemistry cycle

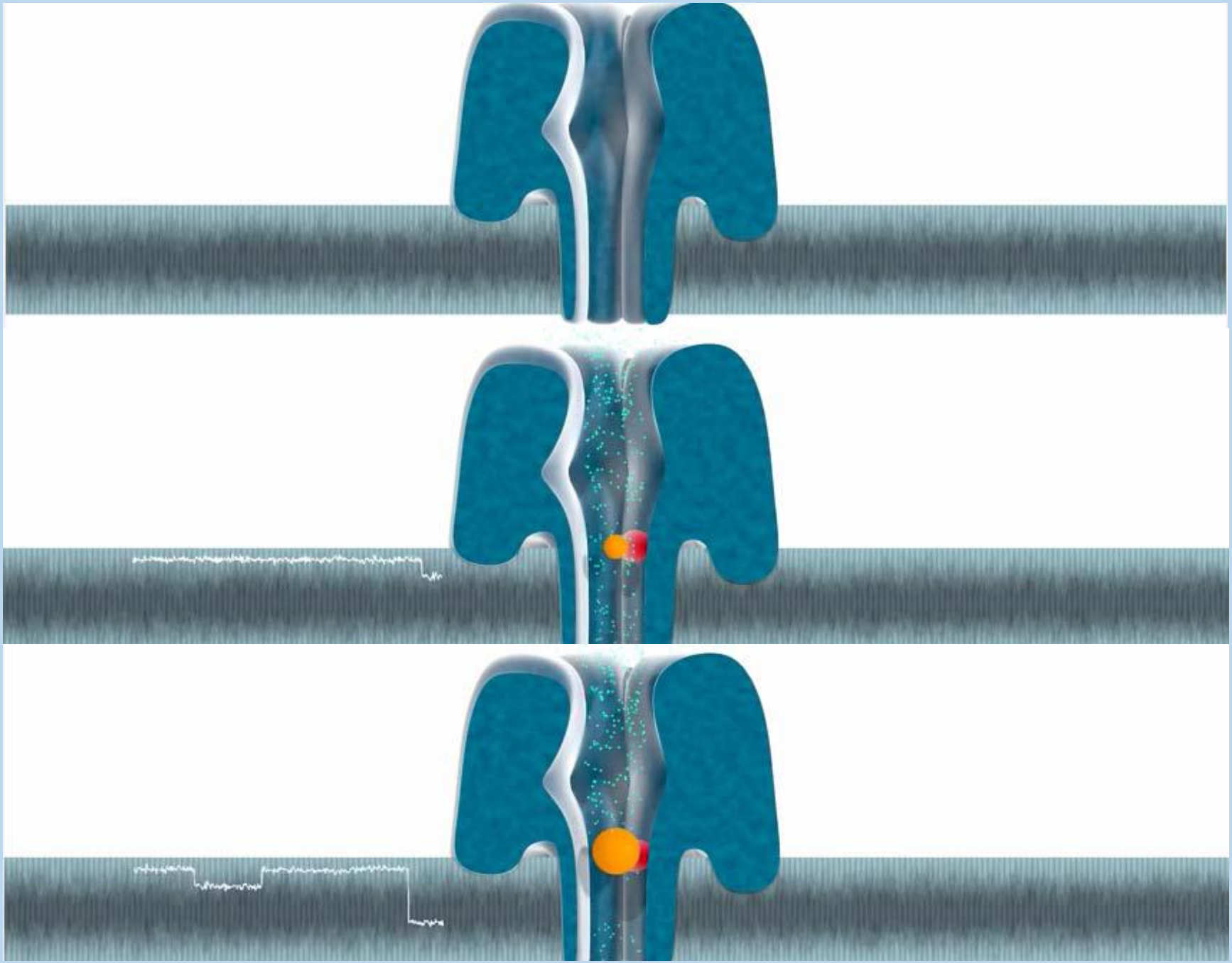
11. Sequencing over multiple chemistry cycles

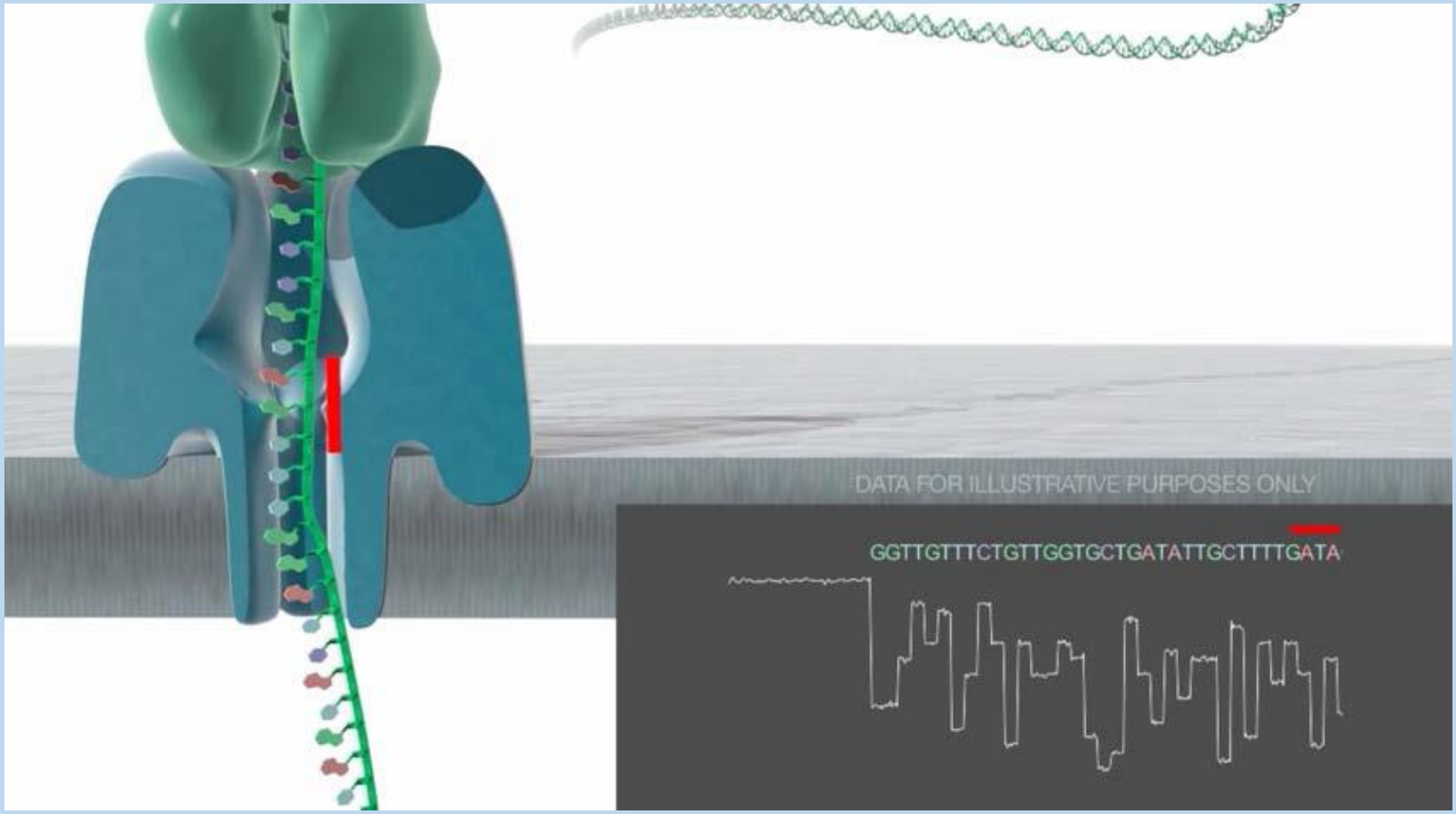
12. Align data

Smallest Sequencer ever

- Based on difference in the resistance to conduction of electrical current by the 4 types of bases while passing through the Nanopore in single stranded form.
- DNA strand is pulled through the nanopore by the enzyme, one base at a time.
- In some libraries DNA strands are bound by hairpin loop so that both the strands are read in a single-go thereby increasing the accuracy and efficiency of sequencing


Oxford
NANOPORE
Technologies
MinION




DATA FOR ILLUSTRATIVE PURPOSES ONLY
GGTTGTTTCTGTTGGTGCTGATATTGCTTTTGATA

View genomic DNA (here from the beta globin locus) from the Trace Archive at NCBI: FASTA format

Show as FASTA ☑ in color

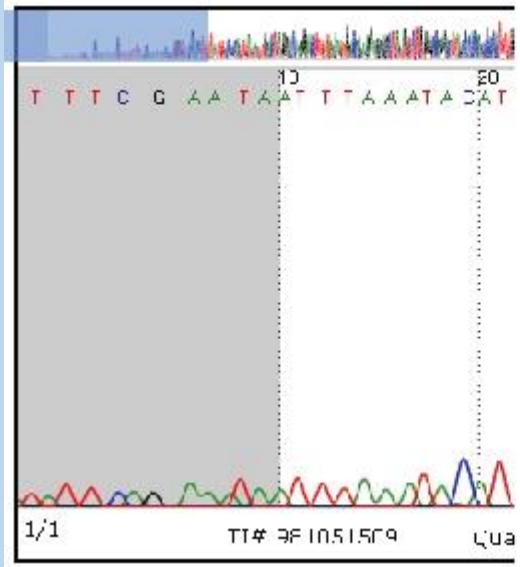
>gnl|tl|981051509 name: *17000177953277* Send to BLAST

Quality score: not available >=0 - <20 >=20 - <40 >=40 - <60 >=60 - <80 >=80 - <100

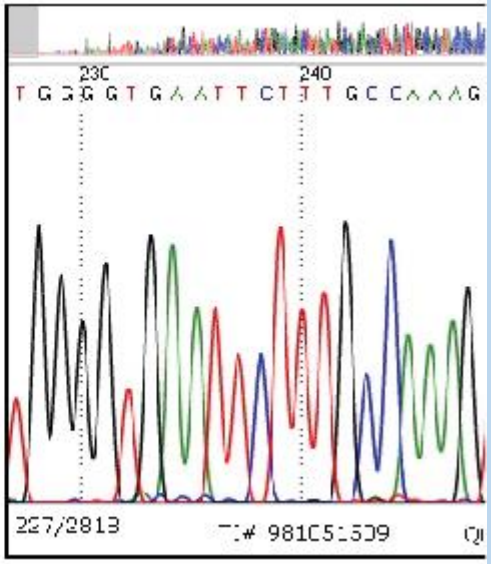
```
TTTCGAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCA
AGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGG
ACAGCAAGAAAGCGAGCTTAGTGATACTTGTGGGCCAGGGCATTAGCCACACCAGCCACCACTTTCTGAT
AGGCAGCCTGCACTGGTGGGGTGAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACCAGCACGTTGCC
CAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGCAAAAGGGCCTAGCTTGGACTCAGAATA
ATCCAGCCTTATCCCAACCATAAAATAAAAGCAGAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATG
AAACCTCTTACATCAGTTACAATTTATATGCAGAAATATTTATATGCAGAGATATTGCTATTGCCTTAAC
CCAGAAATTATCACTGTTATTCTTTAGAATGGTGCAAAGAGGCATGATACATTGTATCATTATTGCCCTG
AAAGAAAGAGATTAGGGAAAGTATTAGAAATAAGATAAACAAAAAAGTATATTAAAAGGAAGAAAGCATT
TTTTAAAATTACAAATGCAAAATTACCCTGATTTGGTCAATTATGTGTACACATATTAAAACATTACACT
TTTAACCCATAAATATGTATAATGGATTATGTATCAATTAAAAATAAAAGAAAATAAAGTAGGGAGATTA
TGAATATGCAAAT
```

Examples of Sanger sequencing traces

Low quality reads



High quality reads



FASTQ format

The FASTQ format stores DNA sequence data as well as associated Phred quality scores of each base.

```
@EAS54_6_R1_2_1_413_324
CCCTTCTTGTCTTCAGCGTTTCTCC
+
;;3;;;;;;;;;;;;7;;;;;;;88
@EAS54_6_R1_2_1_540_792
TTGGCAGGCCAAGGCCGATGGATCA
+
;;;;;;;;;;;7;;;;;-;;;3;83
@EAS54_6_R1_2_1_443_348
GTTGCTTCTGGCGTGGGTGGGGGGG
+EAS54_6_R1_2_1_443_348
;;;;;;;;;;;9;7;;.7;393333
```

DNA read

Base quality score

FASTQ quality scores use ASCII characters

Dec	Char	Dec	Char	Sanger FASTQ	Dec	Char	Sanger FASTQ	Dec	Char	Sanger FASTQ
0	Non-printing	32	Space		64	@	31	96	.	63
1	Non-printing	33	!	0	65	A	32	97	a	64
2	Non-printing	34	“	1	66	B	33	98	b	65
3	Non-printing	35	#	2	67	C	34	99	c	66
4	Non-printing	36	$	3	68	D	35	100	d	67
5	Non-printing	37	%	4	69	E	36	101	e	68
6	Non-printing	38	&	5	70	F	37	102	f	69
7	Non-printing	39	‘	6	71	G	38	103	g	70
8	Non-printing	40	(	7	72	H	39	104	h	71
9	Non-printing	41	)	8	73	I	40	105	i	72
10	Non-printing	42	*	9	74	J	41	106	j	73
11	Non-printing	43	+	10	75	K	42	107	k	74
12	Non-printing	44	,	11	76	L	43	108	l	75
13	Non-printing	45	-	12	77	M	44	109	m	76
14	Non-printing	46	.	13	78	N	45	110	n	77
15	Non-printing	47	/	14	79	O	46	111	o	78
16	Non-printing	48	0	15	80	P	47	112	p	79
17	Non-printing	49	1	16	81	Q	48	113	q	80
18	Non-printing	50	2	17	82	R	49	114	r	81
19	Non-printing	51	3	18	83	S	50	115	s	82
20	Non-printing	52	4	19	84	T	51	116	t	83
21	Non-printing	53	5	20	85	U	52	117	u	84
22	Non-printing	54	6	21	86	V	53	118	v	85
23	Non-printing	55	7	22	87	W	54	119	w	86
24	Non-printing	56	8	23	88	X	55	120	x	87
25	Non-printing	57	9	24	89	Y	56	121	y	88
26	Non-printing	58	:	25	90	Z	57	122	z	89
27	Non-printing	59	;	26	91	[	58	123	{	90
28	Non-printing	60	<	27	92	\	59	124	|	91
29	Non-printing	61	=	28	93	]	60	125	}	92
30	Non-printing	62	>	29	94	^	61	126	~	93
31	Non-printing	63	?	30	95	_	62	127	DEL	

…relating quality scores (e.g. Q30 for 1 in 10^{-3} error rate) to a compact, one character symbol

We do not need to learn the one character symbols, but you should know the importance of base quality scores in sequence analysis.

FASTQ format: Phred scores define quality

The FASTQ format stores DNA sequence data as well as associated Phred quality scores of each base.

$$Q_{\mathrm{PHRED}} = -10 \times \log_{10}(P_e)$$

Phred quality score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

FASTQ format: Phred scores define quality

Phred quality scores of each base are usually defined:

$$Q_{\mathrm{PHRED}} = -10 \times \log_{10}(P_e)$$

There have been alternative base quality definitions:

$$Q_{\mathrm{Solexa}} = -10 \times \log_{10}\left(\frac{P_e}{1-P_e}\right).$$

$$Q_{\mathrm{PHRED}} = 10 \times \log_{10}(10^{Q_{\mathrm{Solexa}}/10} + 1).$$

B&FG 3e
Page 391

Learning outcomes

Session II ~ **1 hrs**

- Refresher on Linux environment
- General presentation
- Assembly of prokaryotic genome
- Introduction to tools for annotating WGS and completeness assessment.
- Introduction to open source GUI based web tools for genome analysis
- Things required for publishing a Whole Genome Sequencing Project

Basic Linux commands

sudo – super user do

root- admin privileges

mkdir- create a directory

cd- change directory

pwd- directory pathway

cp- copy files

rm- remove files

Vim- Text editor

git- operations with git

Terminology

- **DNA Sequencing** - process of determining the precise order of nucleotides within a DNA molecule
- **Genome** - complete set of genes or genetic material present in a cell or organism
- **RNA Sequencing** – sequencing of cDNA using whole transcriptome shotgun method
- **Transcriptome** – collection of all RNA transcripts produced in a cell or population of cells
- **Read** – a sequencing of bases that have been “read” by a sequencing machine
- **Amplification** - Using polymerase chain reaction (Polymerase Chain Reaction/PCR) to generate thousands to millions of copies of a particular DNA sequence

Pre-processing of reads

Adapter

Barcodes

Removal of adapters and barcodes

Demultiplexing

Paired end reads

Single end reads

Long reads

Mate pair merging

Tools for pre processing

Cutadapt
Trimmomatic

Deindexer

Tools for assessment of pre processed reads

FASTQC

FastQC Report

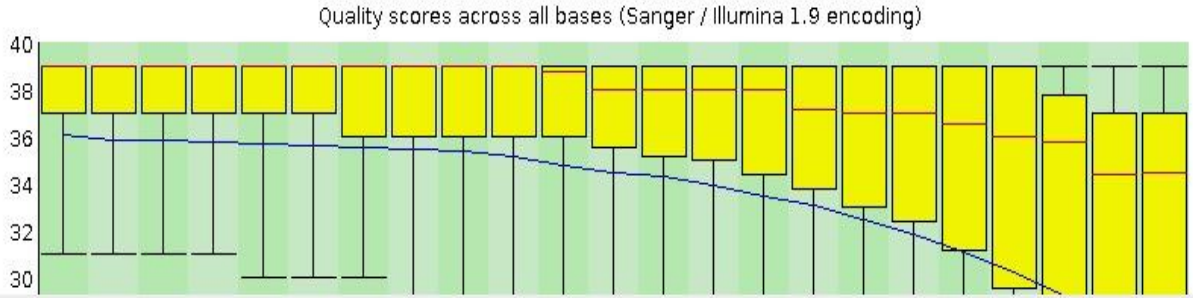
Summary

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

Basic Statistics

Measure	Value
Filename	WES_human_Illumina.pe_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	4942814
Filtered Sequences	0
Sequence length	76
%GC	47

Per base sequence quality



Produced by FastQC (version 0.10.1)

Genome assembly

Genome assembly is the process of converting short reads into a detailed set of sequences corresponding to the chromosome(s) of an organism.

To learn more about assembly visit
http://www.ncbi.nlm.nih.gov/assembly/
http://www.ncbi.nlm.nih.gov/assembly/basics/

How assembly works?

SES ATION TRANS TERIFIC

TRA SESTER ANSES ION

How assembly works?

SES ATION TRANS TERIFIC

TRA SESTER ANSES ION

Reference

TRANS ESTER ESTERIFICATION

How assembly works?

Reference based assembly

TRANS ESTERIFICATION

Raw reads

TRANS TERIFIC
ION
ANSES
ATION
SES
SESTER
TRA

TRANSESTERIFICATION

Terminology

- **Shotgun Sequencing -** An approach used to decode a genome by shredding ("shotgunning") it into smaller fragments of DNA which can then be individually sequenced.
- ***de novo* Sequencing –** sequencing of a genome that has not been sequenced before or does not have a reference
- **Coverage**- a number which tells how many times a particular genome has been sequenced
- **Basecalling**- The process of identifying different signals from the sequencers into respective nucleotide bases
- **Genome assembly**- Assembling the fragments of sequences obtained from the basecalled data to larger fragments or whole genome
- Annotation: Exploring the obtained data for genes.
- FASTA- A format to collect/provide the sequence data.
- FASTQ- Similar to FASTA-but Q defines the per base quality of the sequence

Terminology

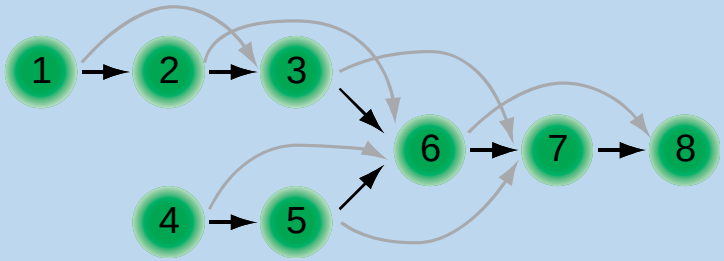
- **Contigs-** A set of overlapping DNA segemnts together represent a consensus region.
- **Scaffold-** Scaffolds are composed of contigs and gaps-These are obtained by patching several contigs.
- **N50**- The length of the shortest contig which constitute at least half of the genome size.

Genome assembly methods: overlap graph, de Bruijn graph, string graph

```
1 ACCTGATC
2   CTGATCAA
3    TGATCAAT
4  AGCGATCA
5    CGATCAAT
6     GATCAATG
7       TCAATGTG
8        CAATGTGA
```

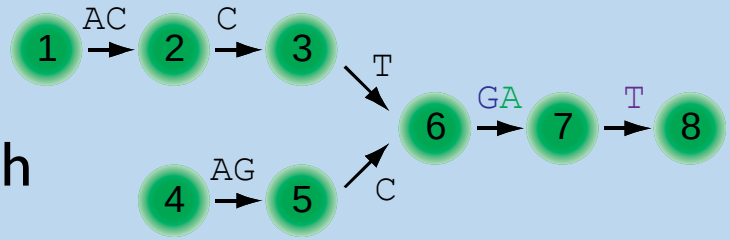
reads

overlap graph



de Bruijn graph

string graph



Genome: ACGATGCTCAGACCCCCCCC

Short reads: ACGATGCTCAGA CTCAGACCC AGACCCC CCCCCCC

k-mers:

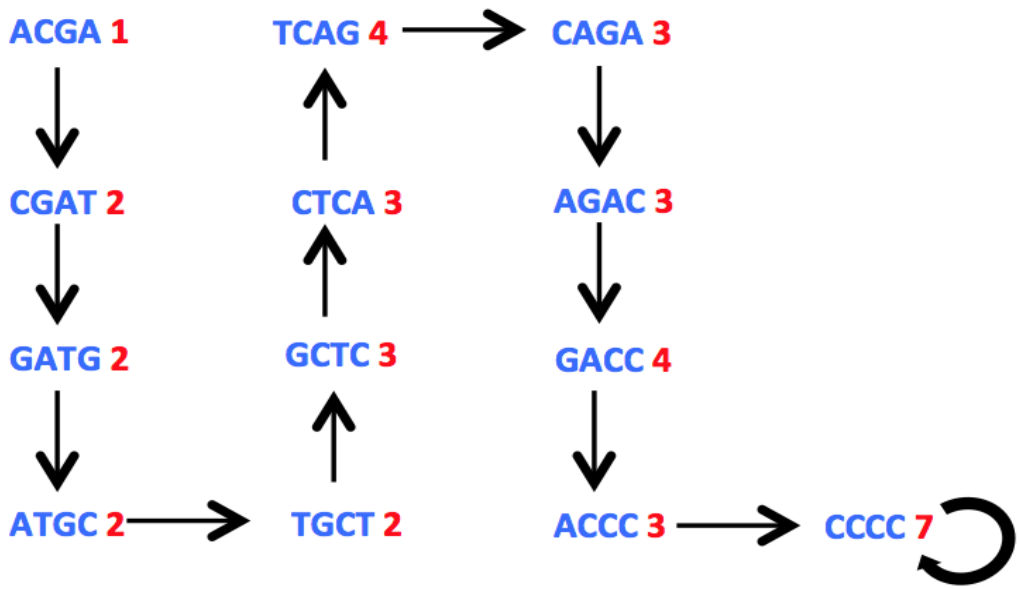
ACGAT
CGATG
GATGC
ATGCT
TGCTC
GCTCA
CTCAG
TCAGA

CTCAG
TCAGA
CAGAC
AGACC
GACCC

AGACC
GACCC
ACCCC

CCCCC
CCCCC
CCCCC

De Bruijn graph:

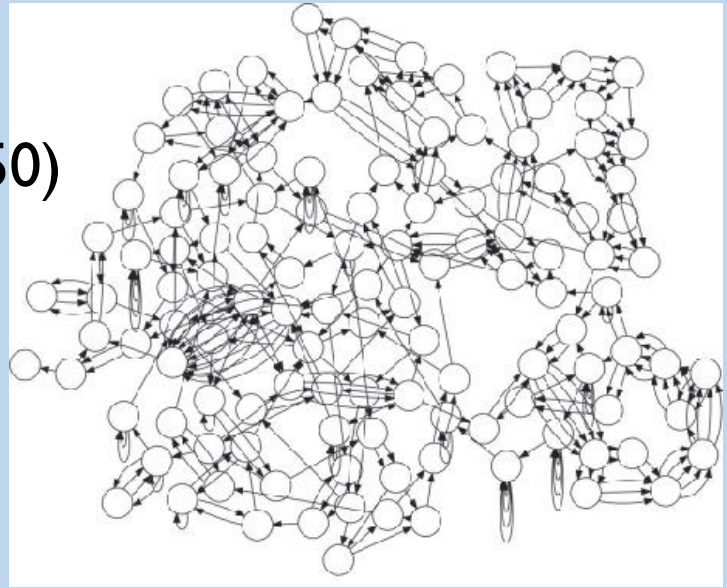


Assembled Contigs: ACGATGCTCAGACCCC

B&FG
Fig.9-1
Page 39

de Bruijn graphs resolve assembly with higher *k* values

E. coli K12 (k=50)



E. coli K12 (k=1,000)

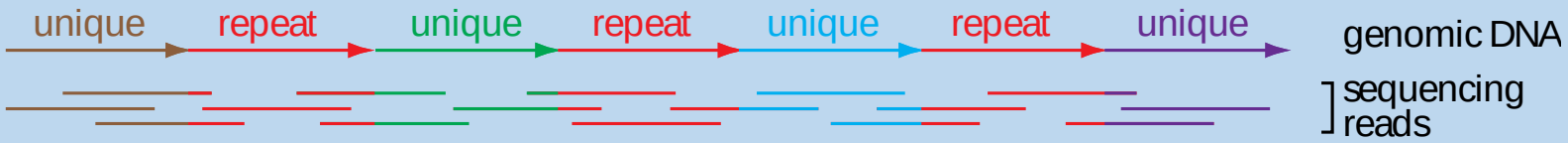
E. coli K12 (k=5,000)

B&FG 3e
Fig.9-11
Page 398

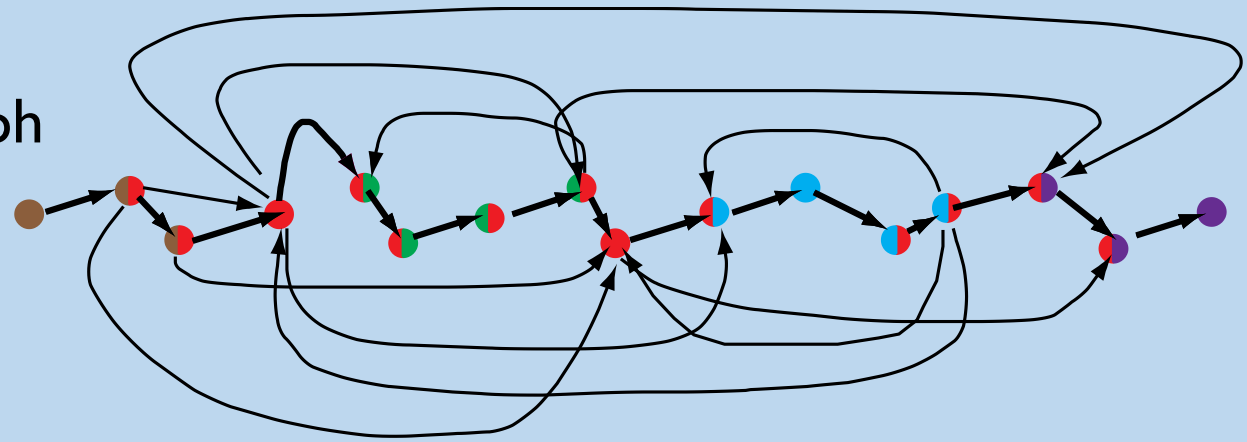
Source: PMID 24034426

Genome assembly with overlap graph and de Bruijn graph

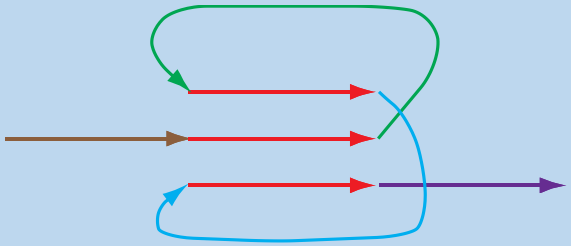
DNA sequence with a triple repeat



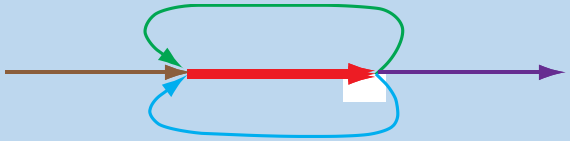
Layout graph



Construction of de Bruijn graph by gluing repeats



de Bruijn graph



Annotation

- There are two different types of genome annotation
- Which is *ab initio* and reference based annotation.
- Some tools for genome annotation…
- GlimmerHMM
- GenScan
- GeneMark
- GeneBench
- RAST
- Prokka etc.,

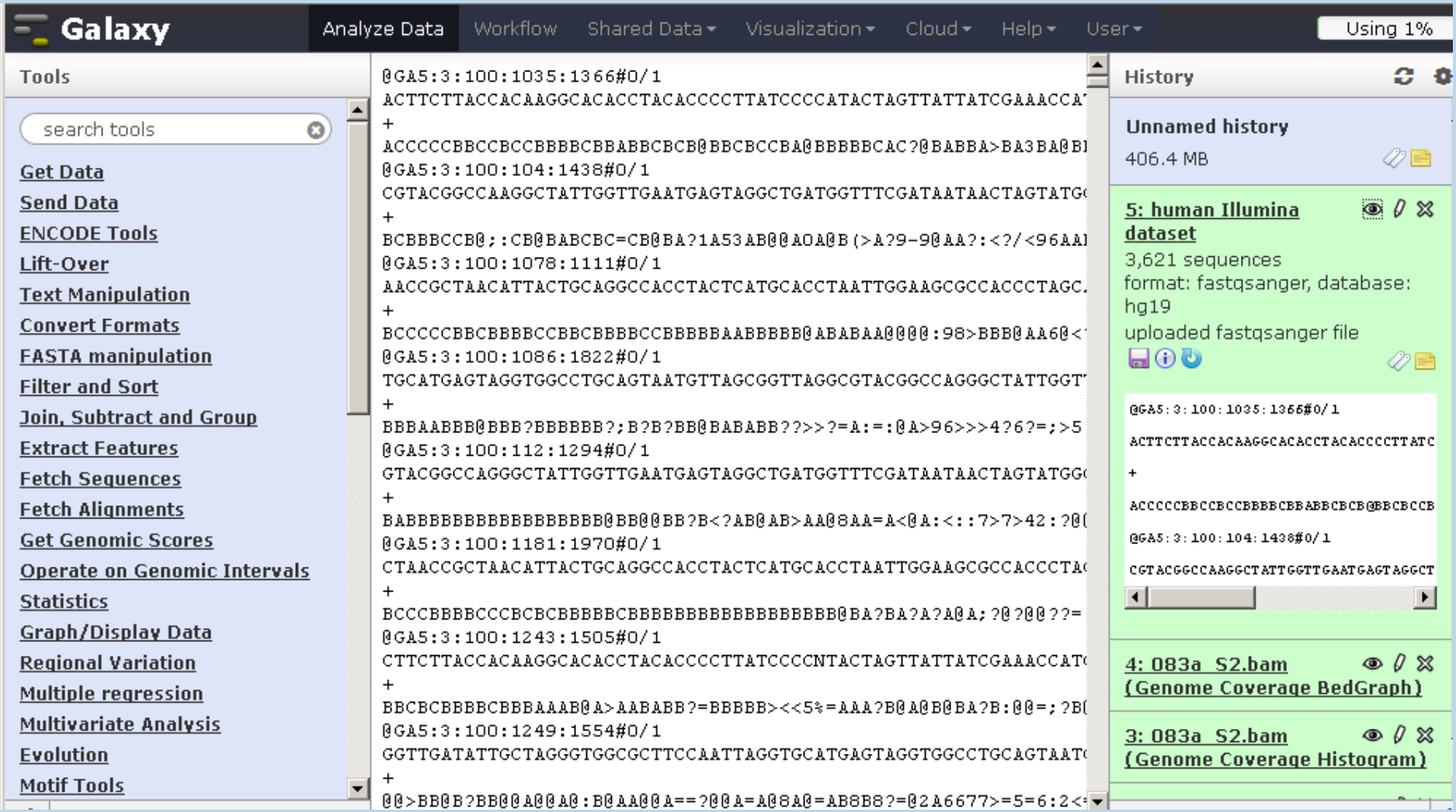
Annotation

- RAST
- The RAST (Rapid Annotation using Subsystem Technology) Server provides high quality genome annotations for prokaryotes across the whole phylogenetic tree. It makes a SEED-quality annotation available as a service with a 48 hour turnaround time.

Genome Assembly-QC

- BUSCO scores-
- Provides measures for quantitative assessment of genome assembly, gene set, and transcriptome completeness based on evolutionarily informed expectations of gene content from near-universal single-copy orthologs.

Galaxy instance



AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequence of *Bacillus vallismortis* TD3, a Salt-Tolerant Strain Isolated from the Sediments of a Solar Saltern in Tamil Nadu, India

Chandrasekaran Suganthi,[a]* Anbazhagan Mageswari,[a]* Manoharan Shankar,[a]* Kodiveri M. Gothandam,[a] Sivashanmugam Karthikeyan[a]

[a]School of Bio-Sciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, India

philic, with the ability to grow easily in the presence of 10% (wt/vol) NaCl. Isolate TD3 was identified as *Bacillus vallismortis* by partial sequencing and analysis of its 16S rRNA gene. Total genomic DNA from *B. vallismortis* TD3 was isolated using the HiPurA bacterial genomic DNA purification kit (HiMedia, Mumbai, India), and the genome sequence of *B. vallismortis* TD3 was generated at Genotypic Technology, Bangalore, India, by Illumina sequencing. Illumina paired-end libraries were constructed per manufacturer-recommended protocols, targeting a read length of 100 bp, and were sequenced on a HiSeq system. The resulting reads were subjected to quality control using SeqQC version 2.2 (Genotypic Technology, Bangalore, India) for adapter trimming, B trimming, and low-quality end trimming. The remaining high-quality reads were assembled *de novo* using SPAdes version 3.1.0 (10), generating 152 contigs yielding a total length of 3,914,588 bp and an N_{50} value of 228,120 bp. These 152 contigs were then subjected to scaffolding using SSPACE version 2.0 (11), yielding a final sequence length of 3,912,114 bp in a set of 29 scaffolds, with a final N_{50} value of 258,393 bp.

Coding sequences in the *B. vallismortis* TD3 genome, which had a GC content of 43.9%, were predicted using the Rapid Annotations using Subsystems Technology (RAST) server (12). A total of 4,206 genes were predicted, including those coding for 113 RNAs (rRNA and tRNA). *B. vallismortis* TD3 encoded osmotolerance determinants mostly restricted to the accumulation of compatible solutes, as opposed to the accumulation

Received 5 June 2018 **Accepted** 14 June 2018 **Published** 12 July 2018

Citation Suganthi C, Mageswari A, Shankar M, Gothandam KM, Karthikeyan S. 2018. Genome sequence of *Bacillus vallismortis* TD3, a salt-tolerant strain isolated from the sediments of a solar saltern in Tamil Nadu, India. Microbiol Resour Announc 7:e00817-18. https://doi.org/10.1128/MRA.00817-18.

Editor John J. Dennehy, Queens College

Address correspondence to Sivashanmugam Karthikeyan, asstdirector.pat@vit.ac.in.

* Present address: Chandrasekaran Suganthi, Department of Biotechnology, D.K.M. College for Women, Vellore, Tamil Nadu, India; Anbazhagan Mageswari, Department of Biotechnology, M.M.E.S. Women's Arts and Science College Vellore, Tamil Nadu, India; Manoharan Shankar, Department of Bioscience & Bioengineering, Indian Institute of Technology, Jodhpur, Rajasthan, India.

Sample MiXs table

Classification	Animalia; Porifera; Demospongiae; Verongimorpha; Chondrillida; Halisarcidae; *Halisarca caerulea*
Investigation type	Eukaryote transcriptome
Environment	
Lat Lon	12.2 N 68.933 W
Geographical location name	Curacao: the Netherlands Antilles
Depth	15-25 m
Collection date	2013–04
Biome	Ocean
Feature	Ocean
Material	Seawater
Sequencing	
Sequencing method	Illumina MiSeq
Estimated size	9.13 Gb (raw and clean)
Isolate	total
Isolation source	adult tissue
Assembly method	
Instrument:	Trinity 2013_08_14
Library layout	Paired
Library selection	PolyA and other
Finishing strategy	High quality transcriptome assembly
Data accessibility	
Database name	NCBI
Project name	PRJNA371551
Sample name	SAMN06309564, SRS2134765, SRP098972, SRR5234759, SRR5863987, SRR5863988
TSA accessions	GFSI01, GFTO01, GFTP01, GFTQ01 https://www.ncbi.nlm.nih.gov/Traces/wgs/?page=1&view=TSA&search=Halisarca

PERKS?

For analysis requests

- Dietome Biotech

- Mob no: 8884113567
- Carthysgn@gmail.com