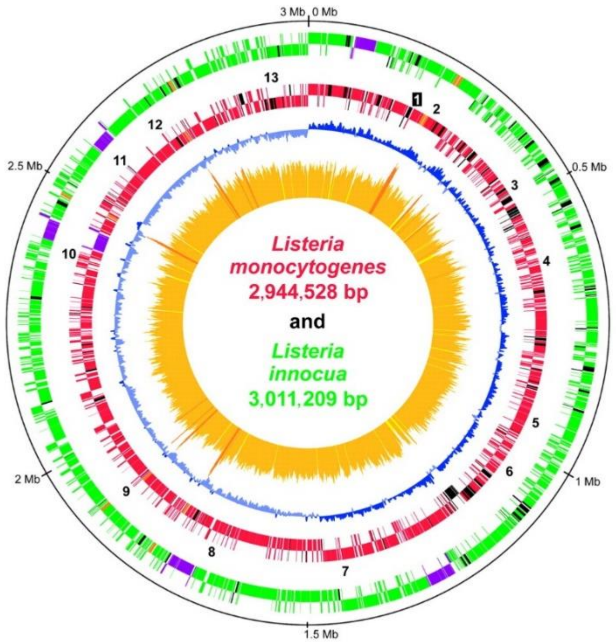
Whole Genome sequencing and *Listeria*: technology, applications and opportunities

Arnoud van Vliet

a.vanvliet@surrey.ac.uk

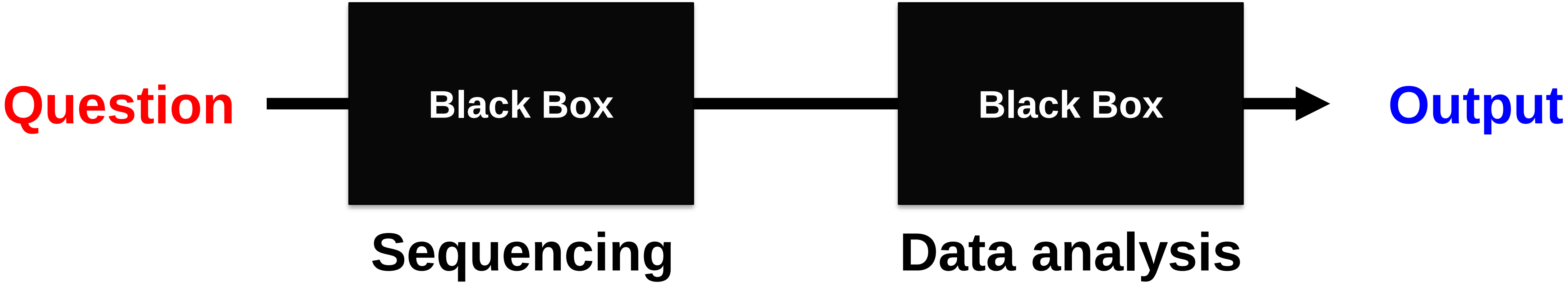


New Zealand FOOD SAFETY SCIENCE & RESEARCH CENTRE

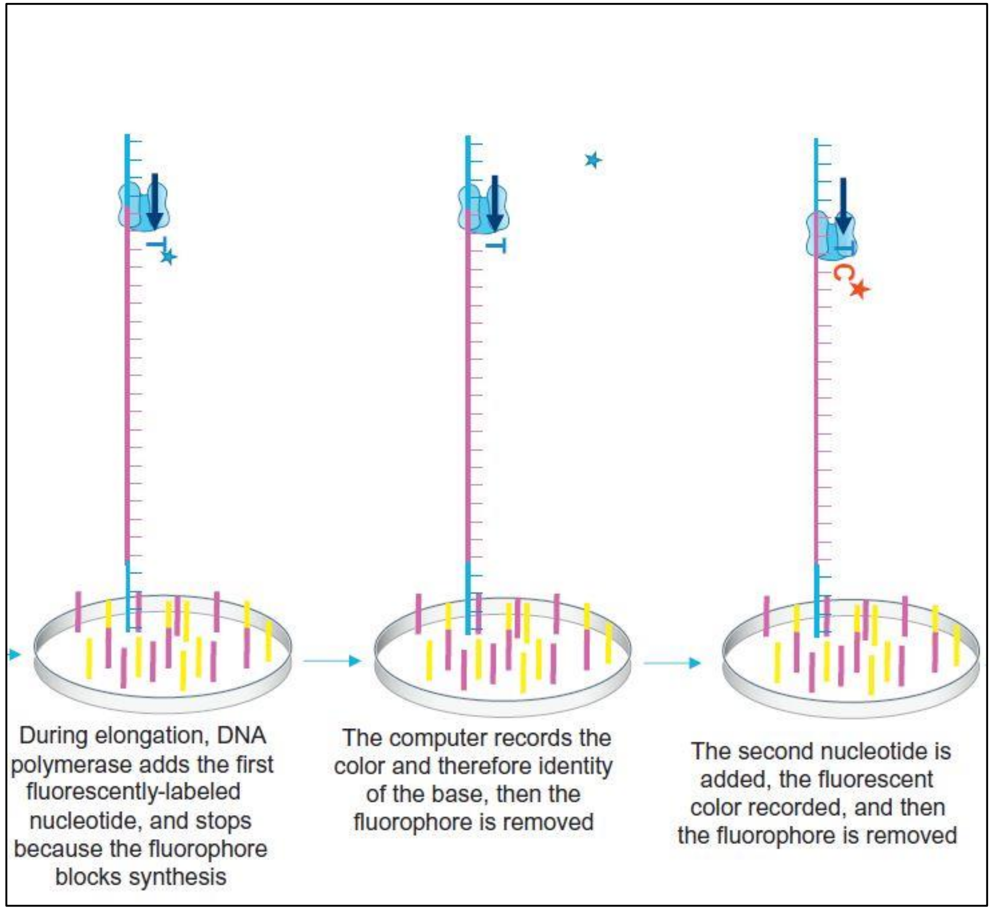
UNIVERSITY OF SURREY

Sequencing: who cares about the details?

- Genome sequencing started in 1990s, now ubiquitous
  - Cost was very high, labour-intensive, computationally problematic
  - But now very affordable and "easy"
  - How much knowledge on details?



Illumina sequencing



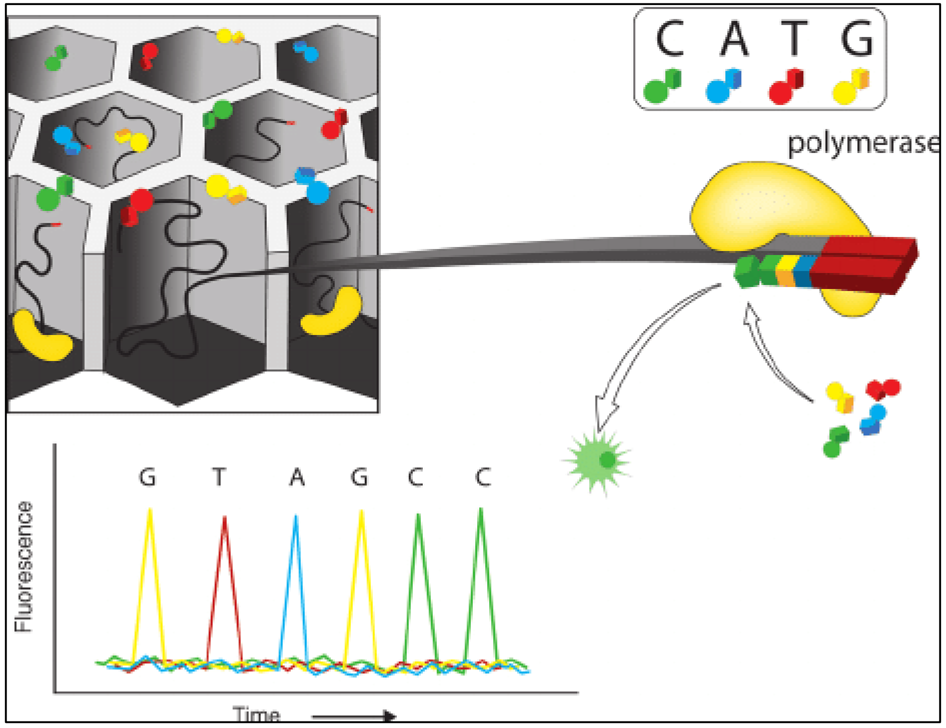
Clark et al, Molecular Biology 3rd edition, 2019

- **Sequencing by synthesis with fluorophores**
  - **Market leader!**
  - **Short reads: 100-300 nucleotides (≤0.01% *Listeria* genome!)**
  - **High quality sequencing, low error rate**
  - **Very high data output**
  - **Expensive machine, not portable**
  - **Multiplexing makes cost per sample low**
- **Applications:**
  - **genome sequencing**
  - **metagenomics**
  - **RNA sequencing etc.**

New Zealand FOOD SAFETY SCIENCE & RESEARCH CENTRE

UNIVERSITY OF SURREY

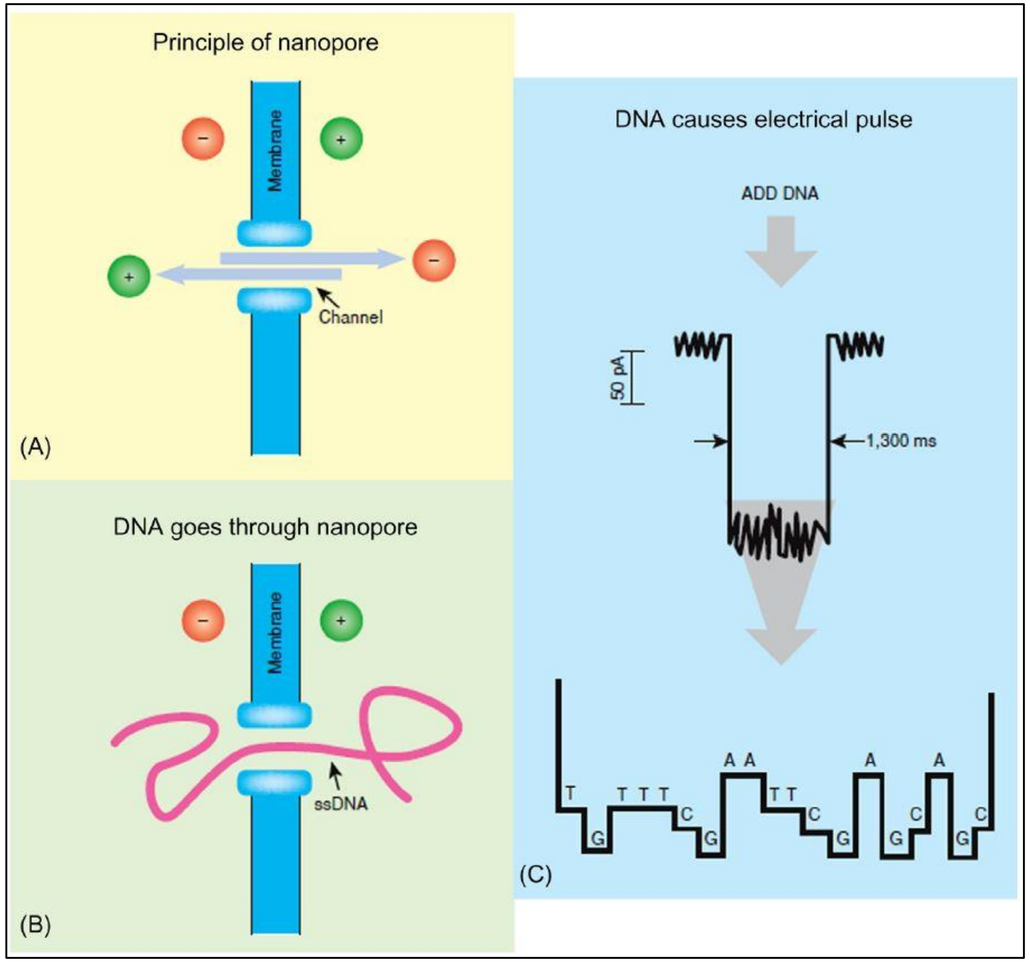
Pacbio sequencing



Researchgate, Lev Jardón-Barbolla

- Sequencing by synthesis with fluorophores
  - Follows polymerase
  - Long reads: >5-10 kb, can be longer
  - Higher error rate, only 85-95% correct
  - Medium-low data output
  - Expensive machine, not portable
  - Less multiplexing, cost per sample higher
- Applications:
  - genome sequencing

Oxford Nanopore sequencing



Clark et al, Molecular Biology 3rd edition, 2019

- Sequencing by measuring disruption of electrical pulses
  - Follows polymerase
  - Long reads: >10-100 kb, can be longer
  - Higher error rate, only 90-95% correct
  - Low data output
  - Cheap, portable machine
  - Less multiplexing, cost per sample high
- Applications:
  - genome sequencing
  - RNA sequencing
  - So many being developed!

Sequencing: Comparison?

	Illumina	Pacbio	Nanopore
Data yield	high	medium-low	medium-low
Reads	short, high qual	long, lower qual	long, lower qual
Genome	fragmented	complete, errors	complete, errors
Machine	Expensive	Expensive	Cheap, portable
Sample cost	Low (multiplex)	High	High

Sequencing: Portability

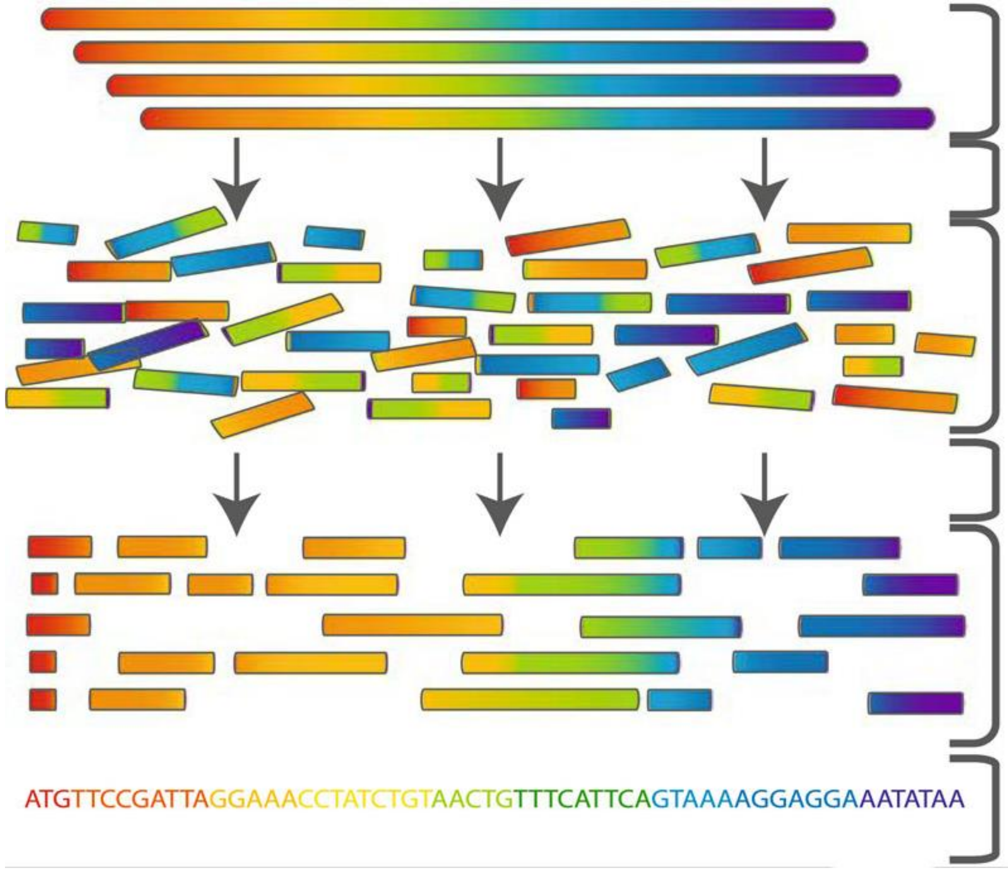
Runs from mobile phone!

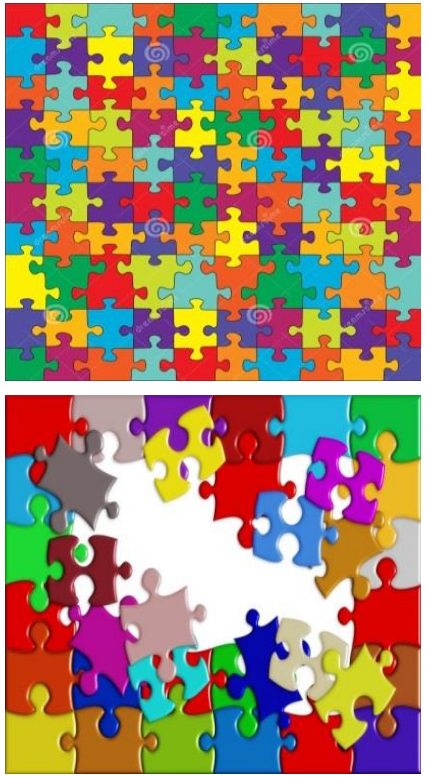
Oxford nanopore

University of Birmingham

Sequencing: data analysis

- **Genome assembly: puzzle with millions of pieces, duplicated ones, faulty ones, and no final picture known!**



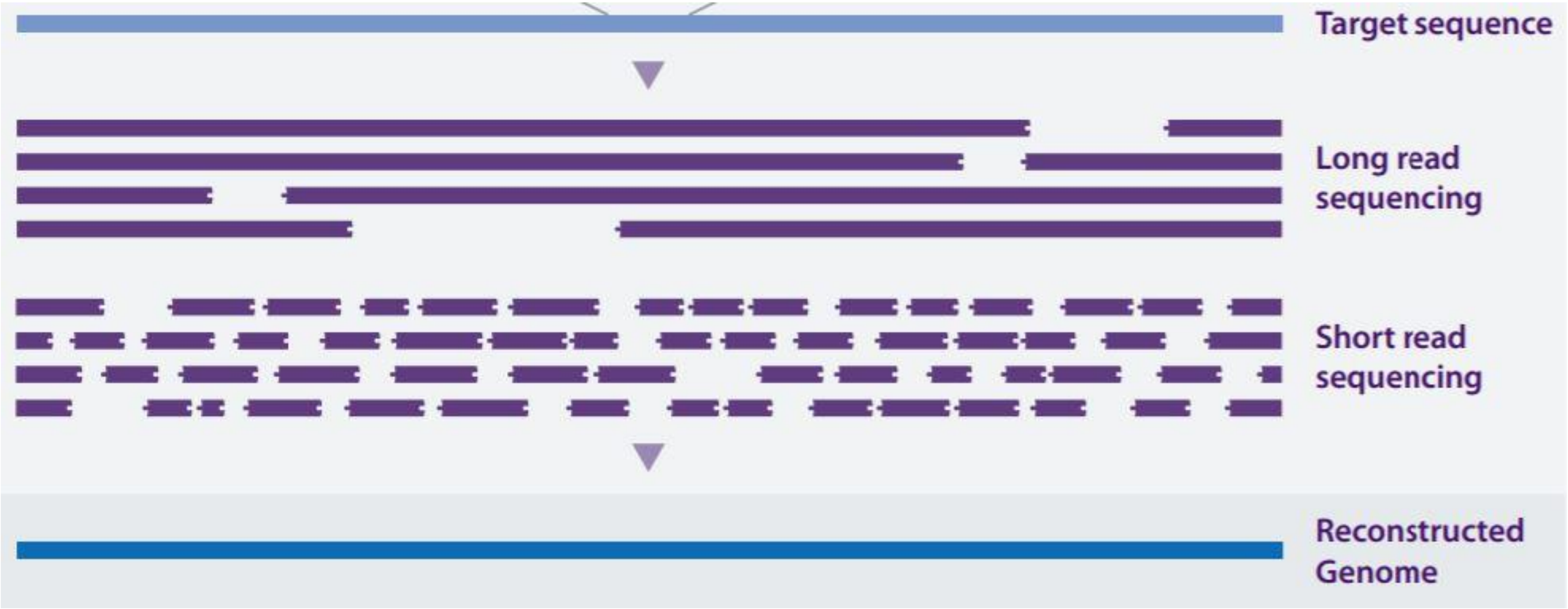


Complete genome (<5% of all genomes)

Whole genome shotgun (>95% of all genomes)

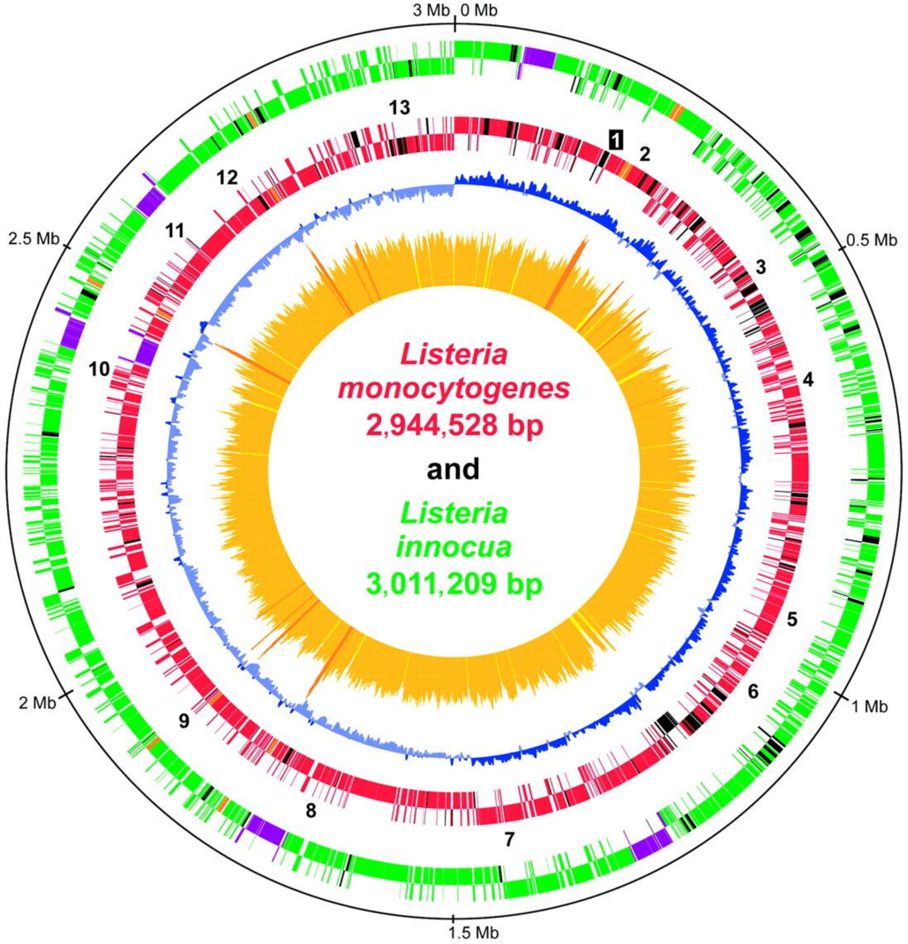
Sequencing: data analysis

- **Genome assembly: puzzle with millions of pieces, duplicated ones, faulty ones, and no final picture known!**



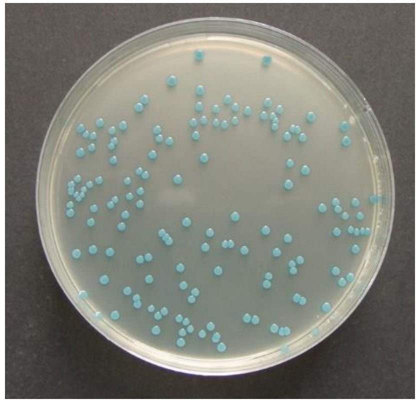
The *Listeria* genome

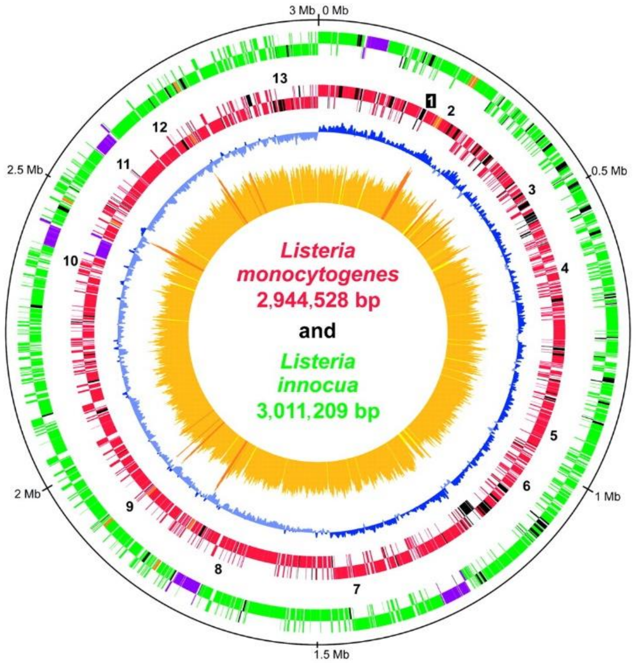
- **First genome sequences of *L. monocytogenes* and *L. innocua* published in 2001 (Glaser et al, Science)**
- **Genome of 3 million bp (*E. coli*: ~5 Mbp, *Campylobacter*: 1.7 Mbp)**
- **Now >40,000 genomes of *Listeria* available in Genbank (mostly *L. monocytogenes*)**
- ***L. monocytogenes* has 4 major lineages (I-IV), with Lineage 1 and 2 the most common ones (L1 more clinical isolates)**



Sequencing: applications

- Current approach: clinical samples or swabs in factories
- Growth + sequencing single isolate

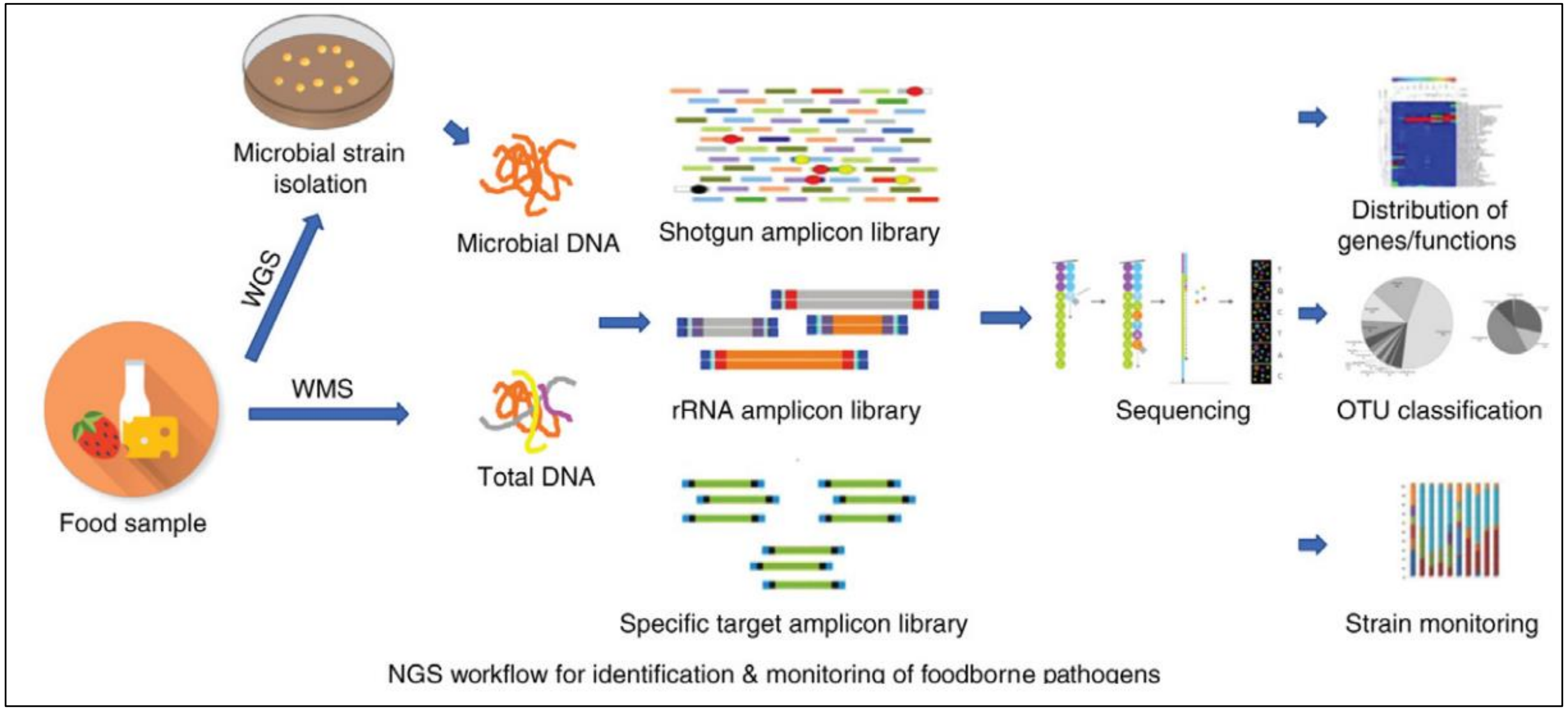




- Disadvantages: slow, underestimates diversity, culture-dependent and may have lower sensitivity. Gold standard!

Sequencing: future applications

- **Portability Oxford Nanopore may allow on-site metagenomics?**
- **No need for culture (or used as backup)**
- **Rapid testing possi-bilities**
- **Combined with standard surveillance > better insight in presence *L. monocytogenes* and other *Listeria* spp?**



Rao & Aurora, in: Chemical Anaysis of Food, Ed. Pico Y, 2020

New Zealand FOOD SAFETY SCIENCE & RESEARCH CENTRE

UNIVERSITY OF SURREY

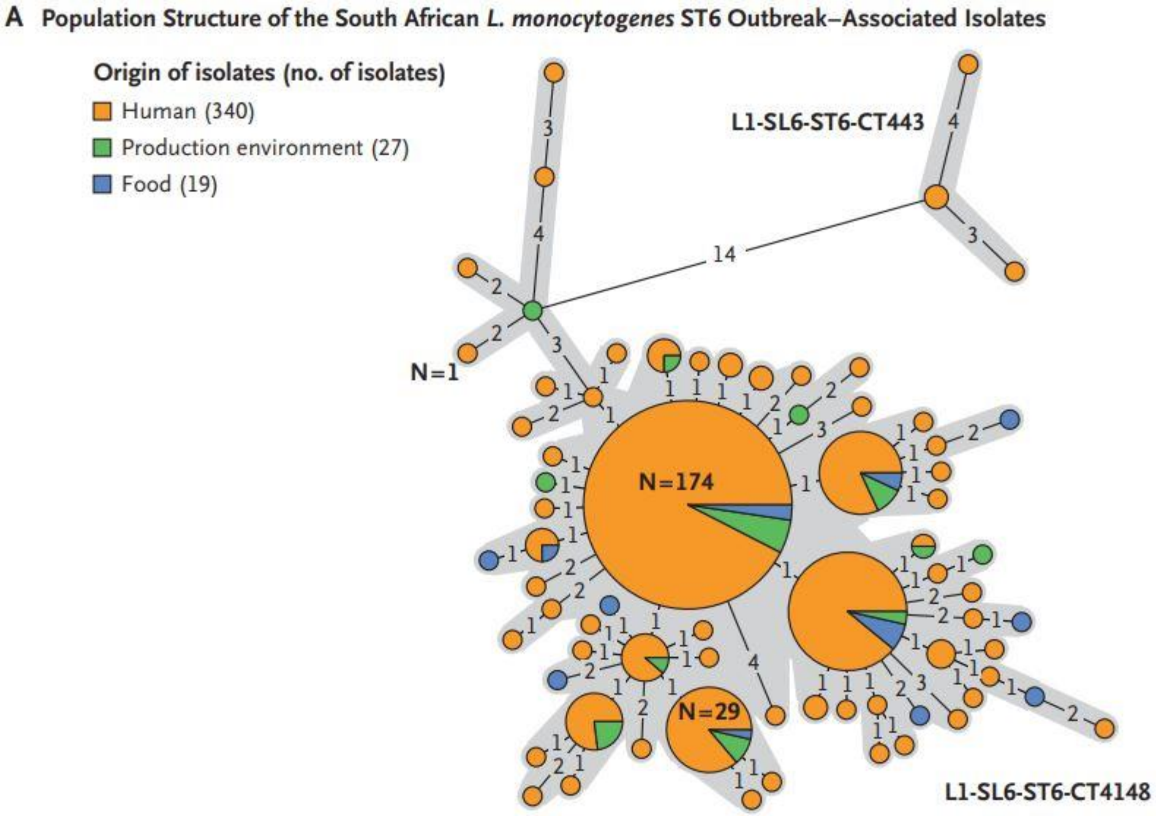
Sequencing: examples of successful use

- **Outbreak of listeriosis in South-Africa, 1060 cases, 216 deaths**
- **Associated with processed meats from one single factory**
- **All isolates very closely related (ST-6)**

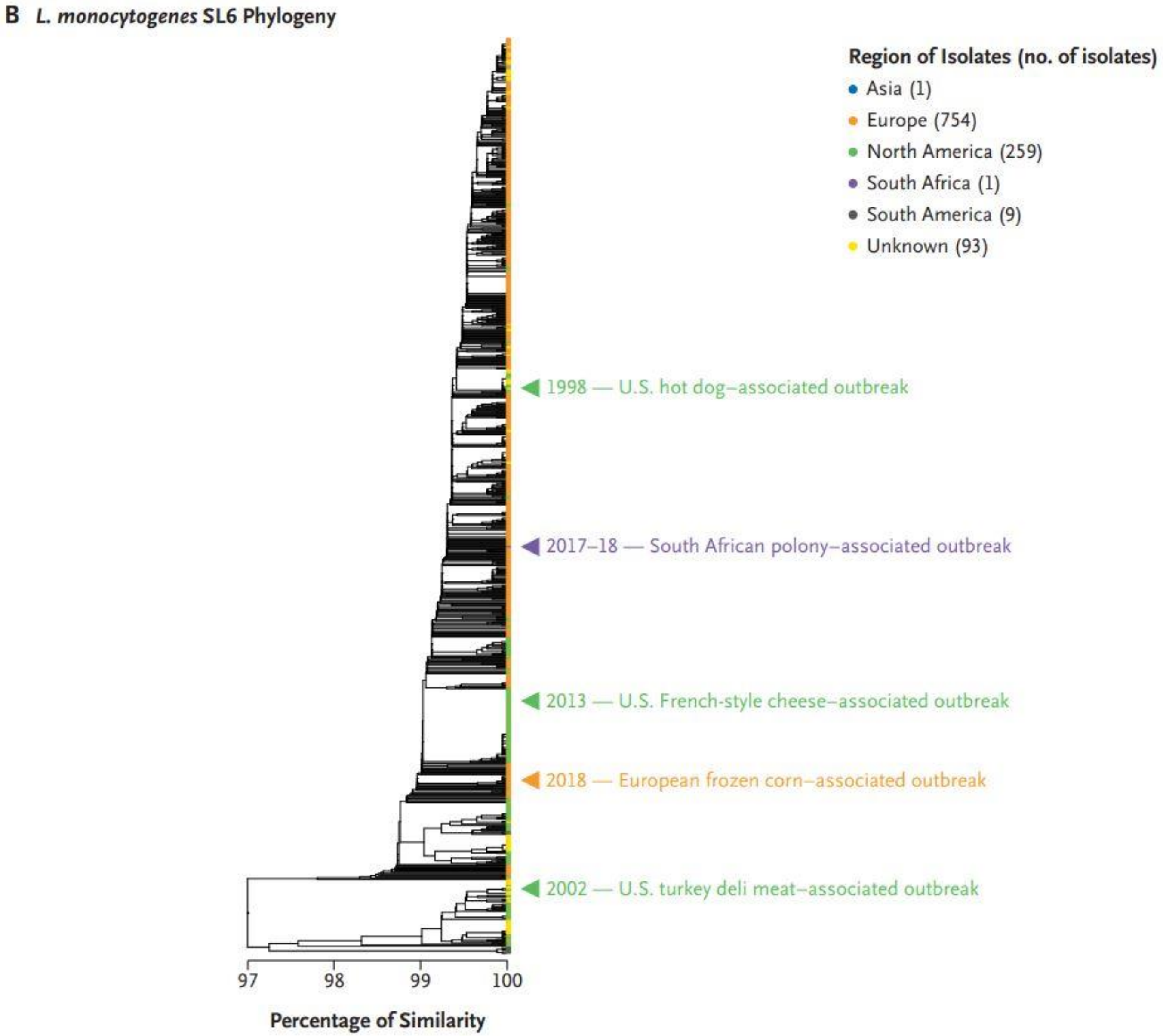
The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Outbreak of Listeriosis in South Africa Associated with Processed Meat



Cost of genome sequencing



Listeria genomics: use available data

- **Downloaded all *Listeria monocytogenes* and *L. innocua* genome assemblies available in Genbank, start of October 2019. Also got all available metadata out of genbank files**
- **After deduplication, cleaning of metadata: total n=23,315**
  - **Lineage 1: n=10,717**
  - **Lineage 2: n=11,730**
  - **Lineage 3: n=760**
  - **Lineage 4: n=52**
  - ***L. innocua*: n=56**
- **Already >40,000 sequences now available!**

Conclusions

- Genome (DNA) sequencing will be highly informative for industry and regulators/government
- Very powerful for retrospective analyses (outbreaks)
  - Data sharing agreements need to be carefully drafted!
- Prospective applications: translation of genome data into risk assessments is still problematic
  - Is any Listeria a problem? Or are there specific troublemakers?

- Learn from the Covid-19 pandemic: many scenarios playing out now!

Acknowledgements

- PFR: Graham Fletcher, Jessika Nowak, Ala Mohan & colleagues
- ESR: Lucy Rivas, Rob Lake, Brent Gilpin & colleagues
- Massey University: Nigel French, Patrick Biggs & colleagues

New Zealand
FOOD SAFETY SCIENCE
& RESEARCH CENTRE

UNIVERSITY OF
SURREY