Encapsulated Bacteria Session 3: Sequence-based typing

Genomics and Clinical Microbiology 2024

Martin Maiden, Made Krisna, Kasia Parfitt, Keith Jolley

Department of Biology

UNIVERSITY OF
OXFORD

Clinical Scenario Results

Specimen	Patient age	Presentation	Organism	Culture and CSF?	Serogroup
2	15 months	Meningitis	*N. meningitidis*	Yes	
5	3 years (46 months)	Meningitis	*N. meningitidis*	CSF only	

Clinical Scenario Results

Specimen	Patient age	Presentation	Organism	Culture and CSF?	Serogroup
2	15 months	Meningitis	*N. meningitidis*	Yes	B
5	3 years (46 months)	Meningitis	*N. meningitidis*	CSF only	B

Notes

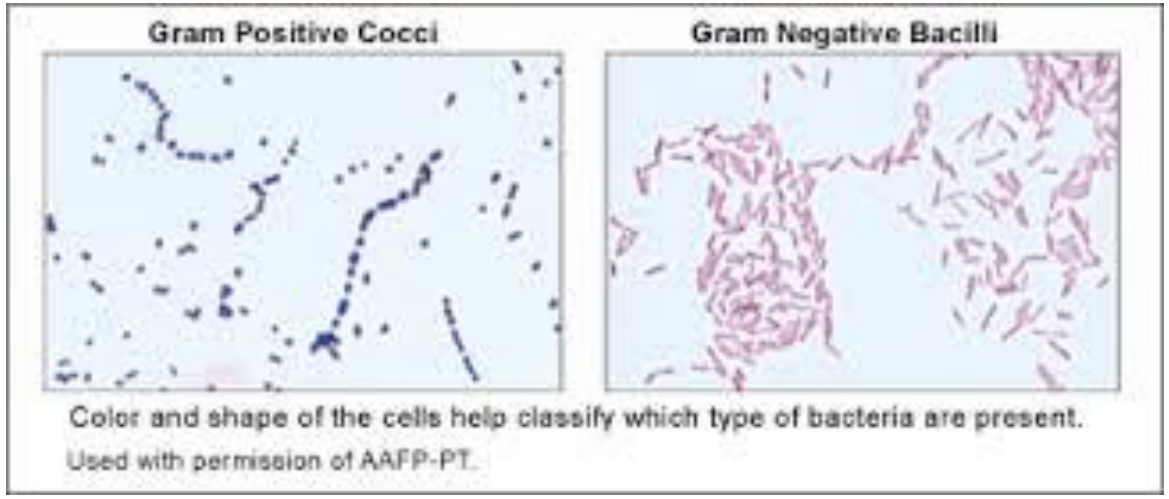
- These two patients are related.
- They have both received Meningococcal serogroup C conjugate (MCC) vaccines, according to the Irish immunisation schedule at the time of birth.

Discussion points

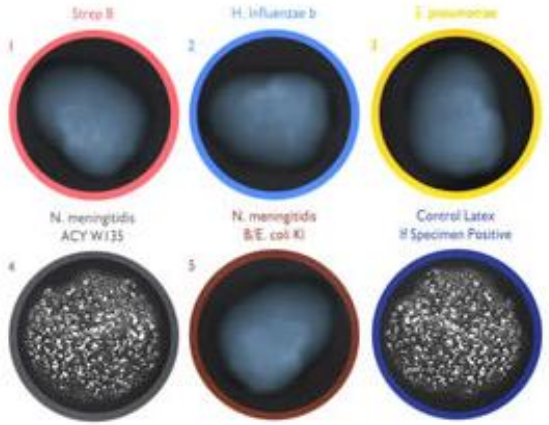
- What actions are you going to take?
- What additional information do you need, and how are you going to get it?

Phenotypic characterisation

Gram's Stain



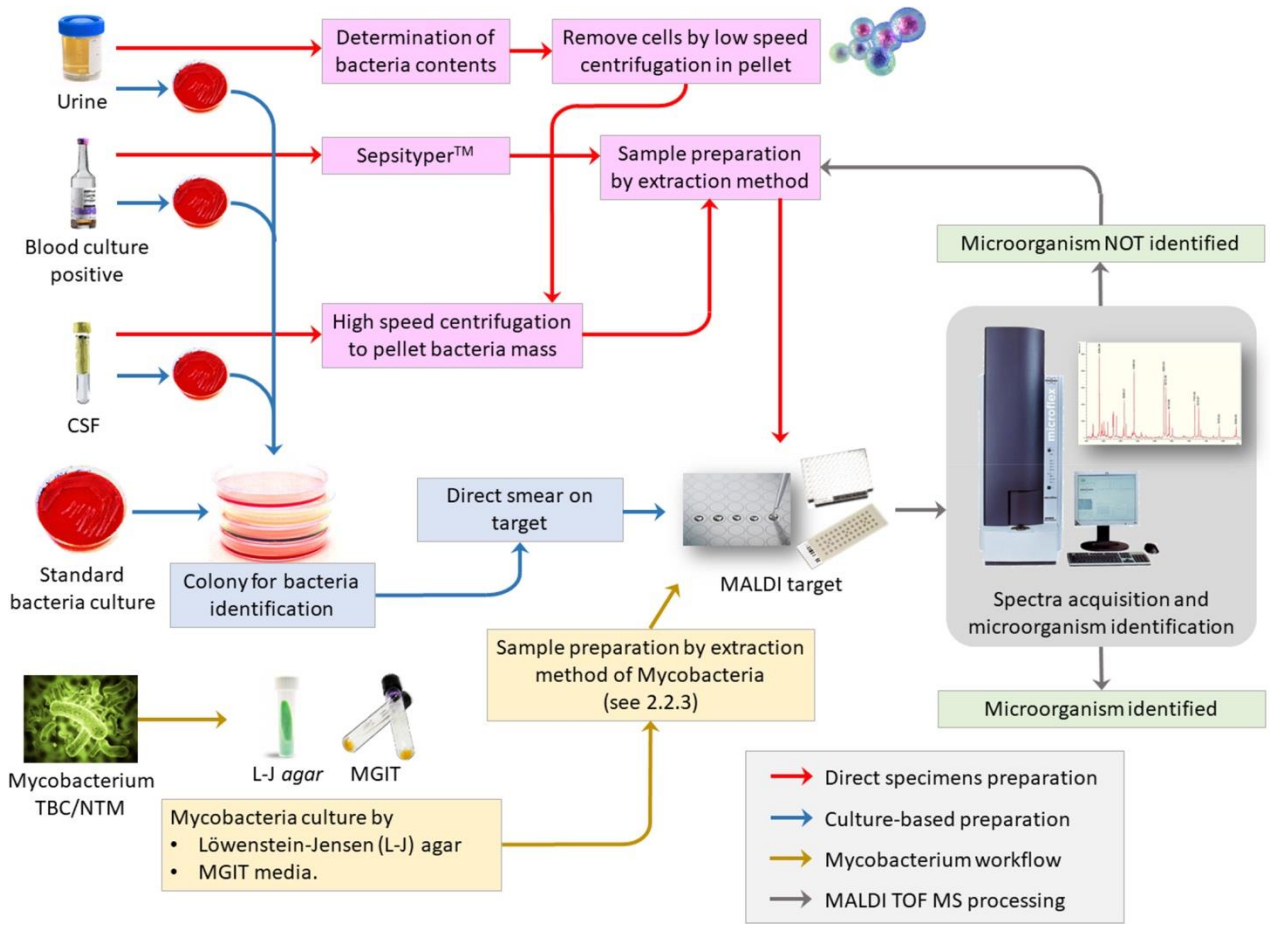
Serology



Anti-microbial resistance testing

Metabolic phenotyping For species identification

MALDI-TOF phenotyping



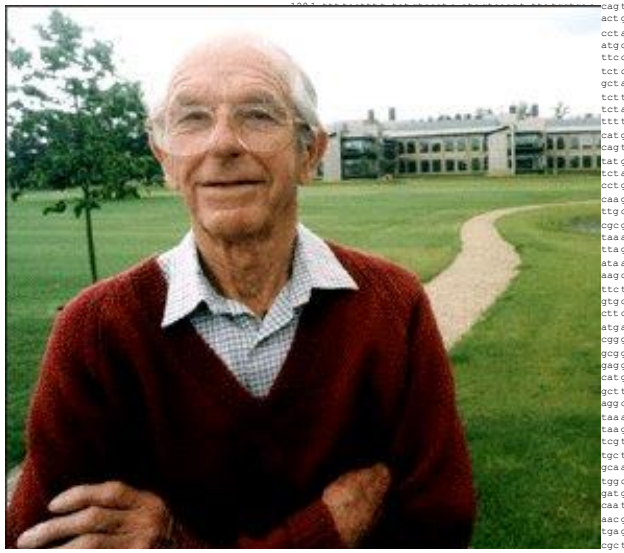
Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry.

- Enables rapid and inexpensive (per sample) species identification
- Requires:
  - bacterial sample (usually an isolate);
  - the equipment;
  - a database linking spectra to bacterial species identification.
- Commercial systems are widely installed and used in clinical laboratories.

Hou, T. Y., Chiang-Ni, C. & Teng, S. H. (2019). Current status of MALDI-TOF mass spectrometry in clinical microbiology. *J Food Drug Anal.* **27**, 404-414.

Nucleotide sequencing and clinical microbiology

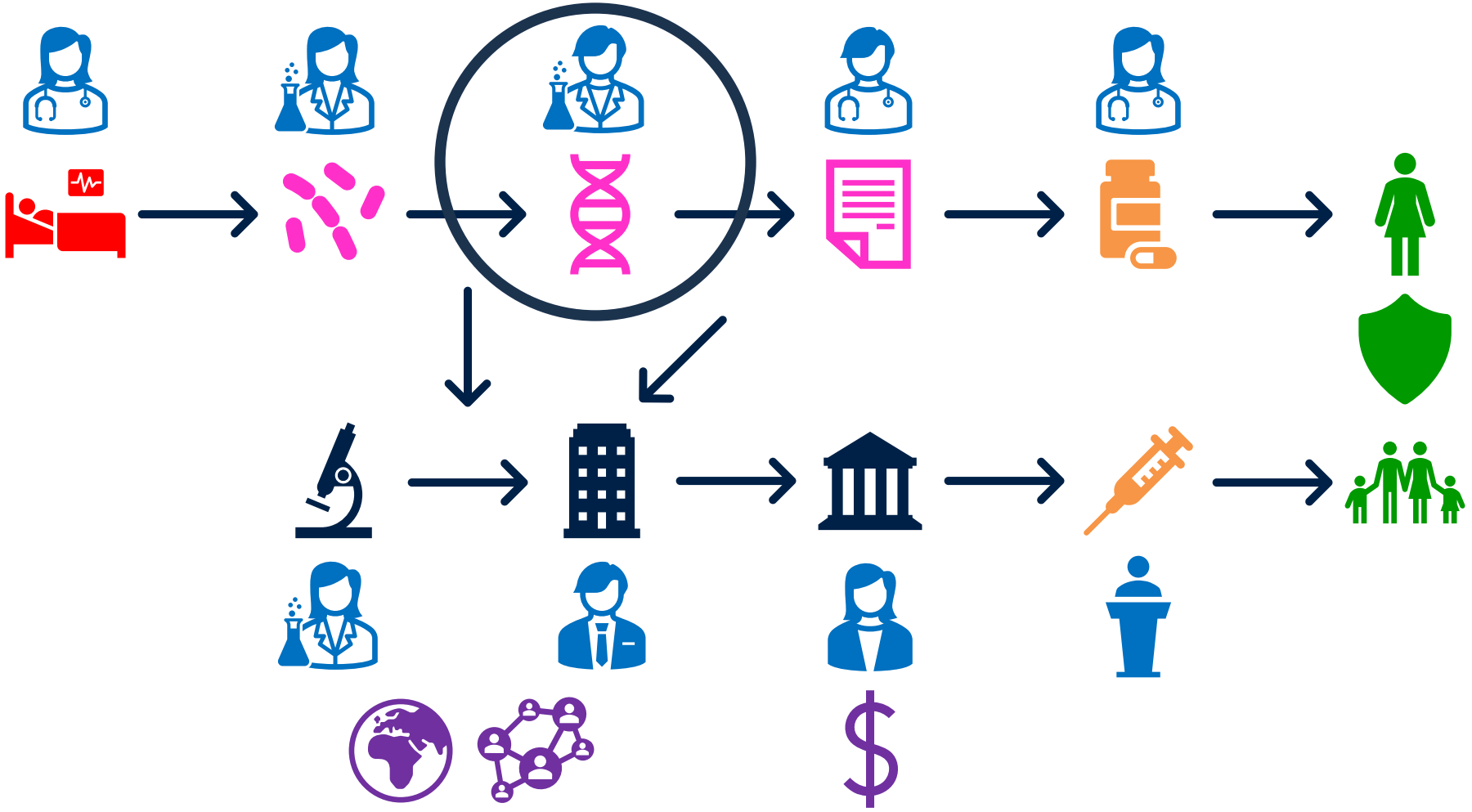
- **Definitive:**
  - fundamental level of information;
  - any part of the genome can be accessed.
- **Reproducible:**
  - nucleotide sequences are either right or wrong and can be checked;
  - reverse mutations are (usually) rare.
- **Scalable:**
  - nucleotide sequencing technology can be conducted on one or many samples and on a few base pairs or a whole genome.
- **Manipulability:**
  - nucleotide sequences can be analysed with model-based methods.
- **Can be done from a PCR reaction or microbiome sample:**
  - YOU DON'T (necessarily) NEED AN ISOLATE!



Frederick Sanger (1918-2013)

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M. & Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature 265, 687-695*.

Molecular approaches to clinical and public health microbiology



The role of genetics and genomics

- Nucleotide sequencing facilitates specimen characterisation,
  - sequence-based typing, up to the level of whole genome sequences (WGSs),
  - amplification techniques increase the amount of material for characterisation;
- Sequences are definitive, reproducible, and comparable.
  - easily stored, transported, and manipulated with analysis algorithms;
- However, datasets are large and complex,
  - requiring interpretation for the practitioner.

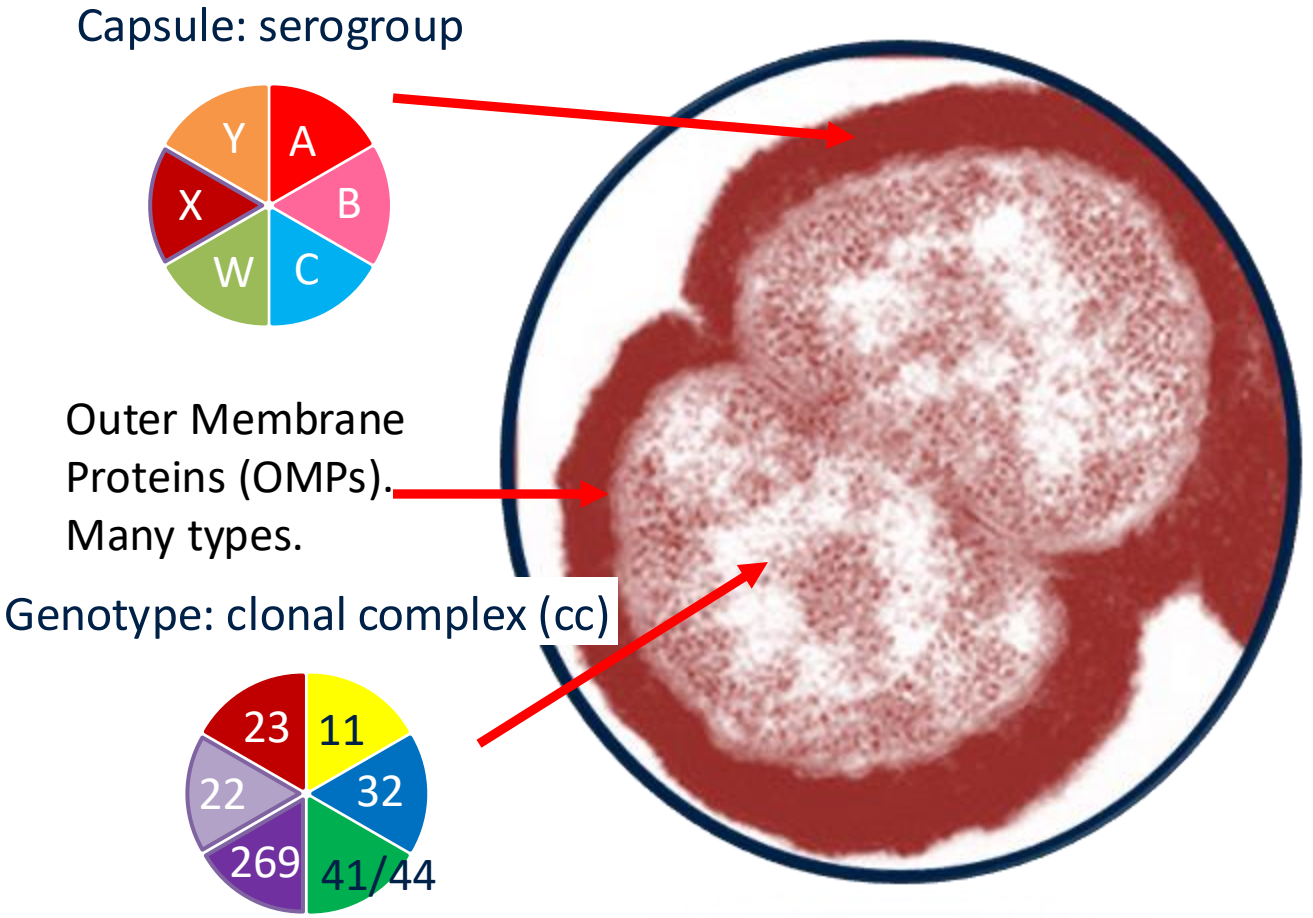
Steps in the characterisation of meningococci

- **Diagnosis:**
  - Discrimination from other organisms e.g. *Streptococcus pneumonie* and *Haemophilus influenzae.*

- **Characterisation:**
  - Differentiation (e.g. meningococcus from other *Neisseria*);
  - Serogroup (conjugate polysaccharide vaccines available against serogroups A, C, Y, W, but not B);
  - Protein vaccine antigens (e.g. PorA, BAST typing), protein-based vaccines have been developed as substitutes to serogroup B vaccines;
  - The discrimination of isolates is essential for outbreak investigation;
  - Monitoring local and global trends in disease & the identification of likely epidemic strains.

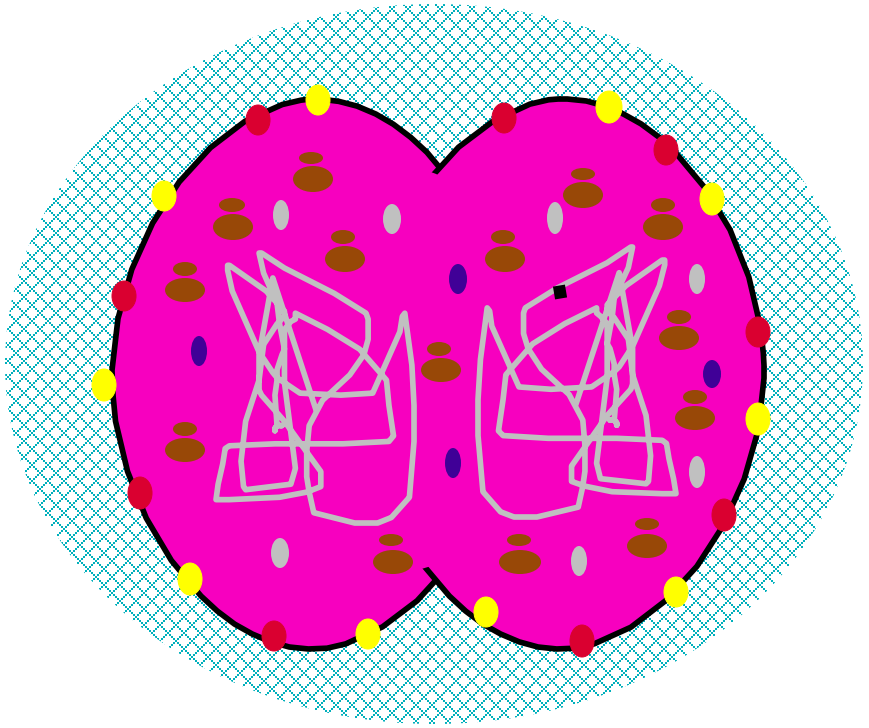
Within species diversity *Neisseria meningitidis*

- Meningococci are highly diverse antigenically and genetically,
  - this diversity is structured.
- 12 capsular serogroups,
  - 6 associated with invasive disease.
- Extensive evidence of HGT, but stable lineages are present,
  - these are associated with phenotypes, including invasive disease.



Ganesh, K., Allam, M., Wolter, N., Bratcher, H. B., Harrison, O. B., Lucidarme, J., Borrow, R., de Gouveia, L., Meiring, S., Birkhead, M., Maiden, M. C., von Gottberg, A. & du Plessis, M. (2017). Molecular characterization of invasive capsule null *Neisseria meningitidis* in South Africa. *BMC Microbiology* **17**, 40.

Case Study, assembling a ‘Strain type’ or ‘fine type’: *Neisseria* meningitidis characterisation



Component	Phenotypic	Genotypic
Capsule	Serogroup	*cps* region
OMPS	Serotype, Subtype, etc.	*porA, porB, fetA,* etc.
Housekeeping genes	MLEE	MLST
Ribosomes	MALDITOF	16s rRNA, rMLST

***Neisseria meningitidis* C: P1.21-15,16: F1-7: ST-10217 (cc10217)**

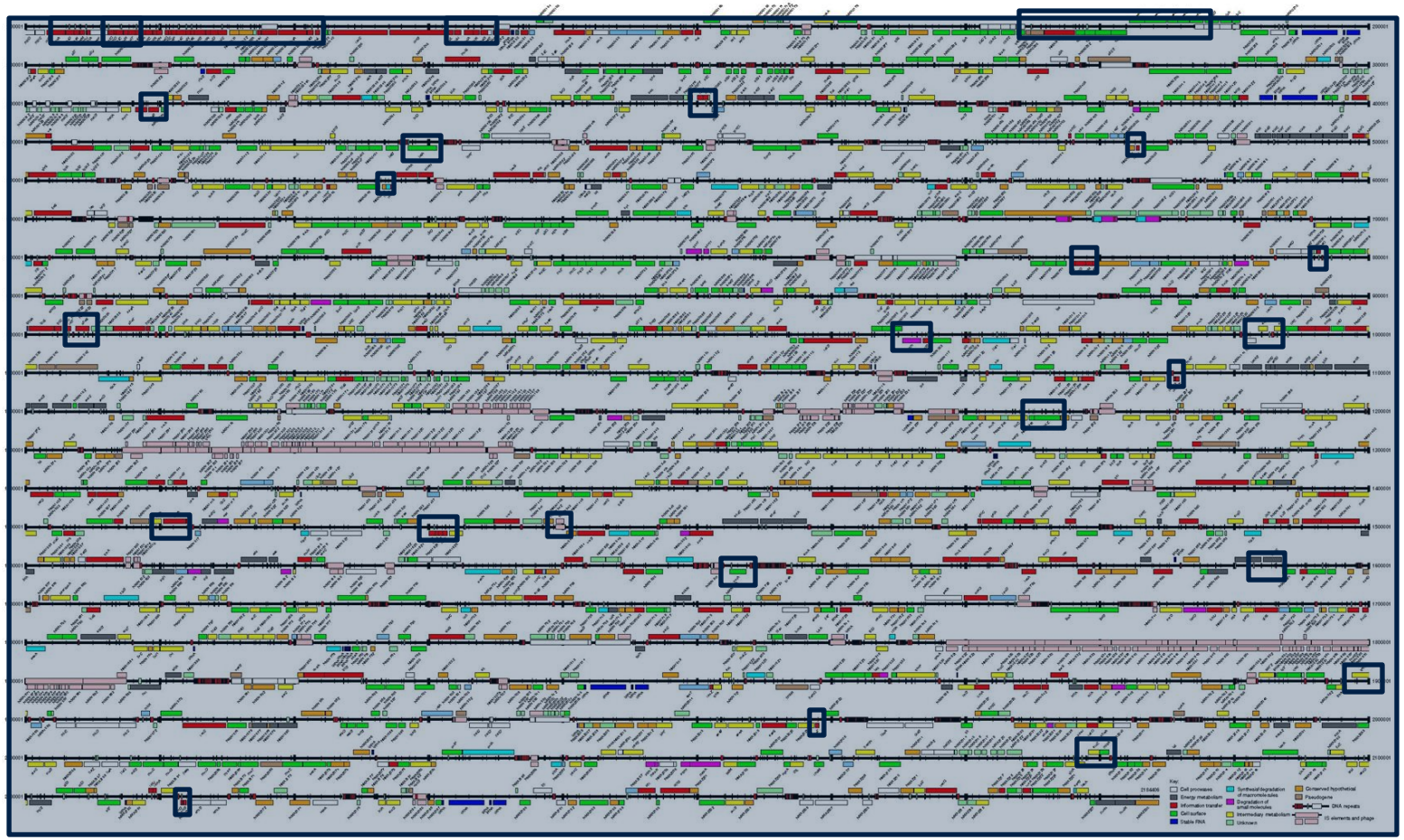
Jolley, K. A., Brehony, C. & Maiden, M. C. (2007). Molecular typing of meningococci: recommendations for target choice and nomenclature. *FEMS Microbiol Rev* **31**, 89-96.

Meningococcal genome characterisation

rplF

fetA

MLST
abcZ
adk
aroE
fumC
gdh
pdhC
pgm



cps

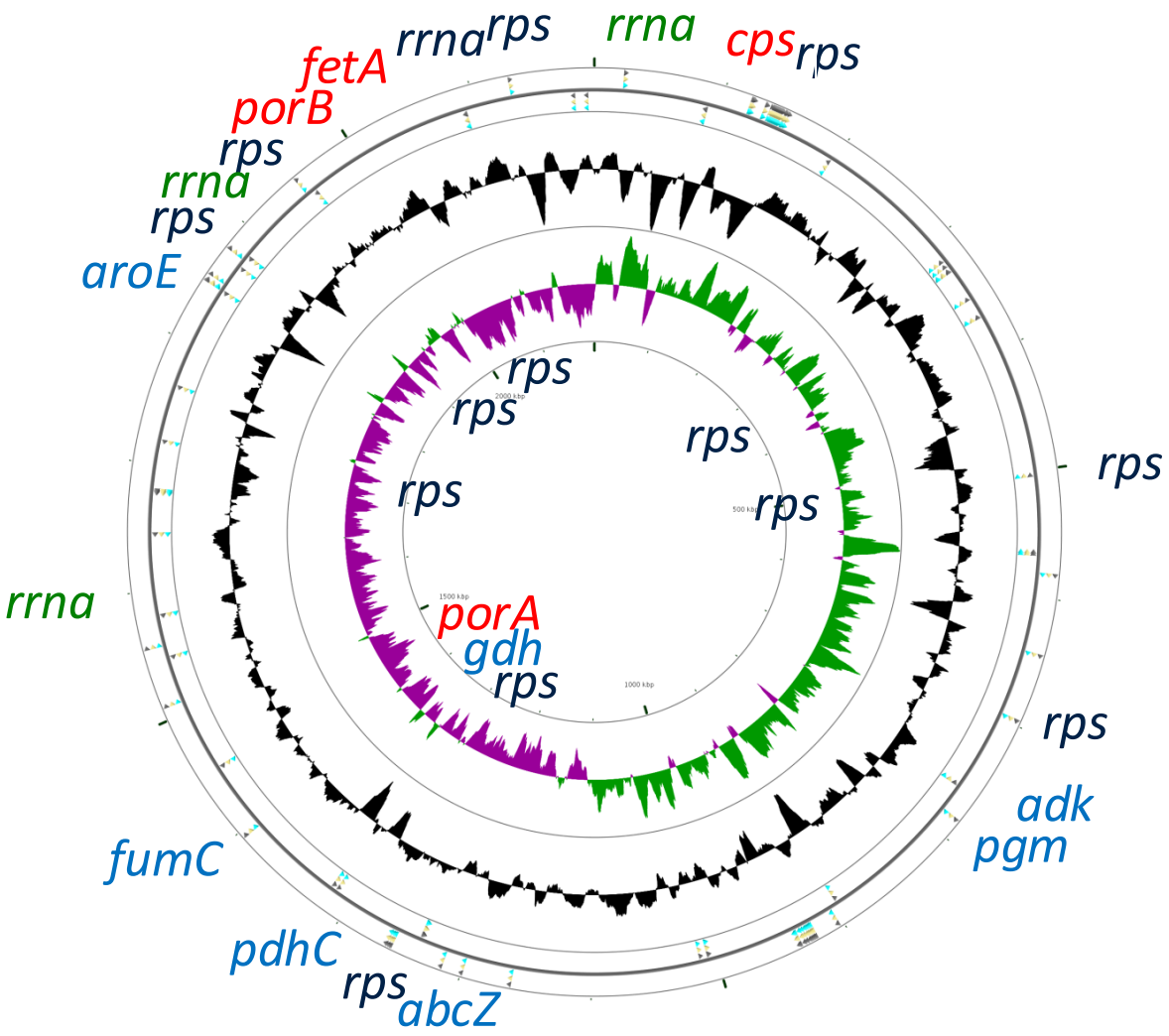
rMLST
rpl,
rpm,
rps
53 loci

porA

cgMLST
1605 loci

Parkhill, J., Achtman, M., *et al.* Spratt, B. G. & Barrell, B. G. (2000). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502-506.
Bratcher, H. B., Corton, C., Jolley, K. A., Parkhill, J. & Maiden, M. C. (2014). A gene-by-gene population genomics platform: *de novo* assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genomics* **15**, 1138.

Typing targets in the meningococcal genome



Antigen genes:

- Capsule (*cps*), serogroup;
- OMPs (*porA, porB, fetA*), serotype, subtype fine type;
- BAST antigens.

Housekeeping genes:

- Ribosomal RNA genes (*rrna*) genus identification;
- Ribosome protein genes (*rps*), species and sub species characterisation (rMLST);
- MLST (metabolism), subspecies identification
- cgMLST, high-resolution typing.

Antibiotic Resistance genes.

rplF assay

- **Problem:** rapid cost effective determination of *Neisseria* species from 1000s of isolates obtained in the MenAfriCar surveys.
  - 16s rRNA sequencing too cumbersome and insufficient discrimination or resolution.

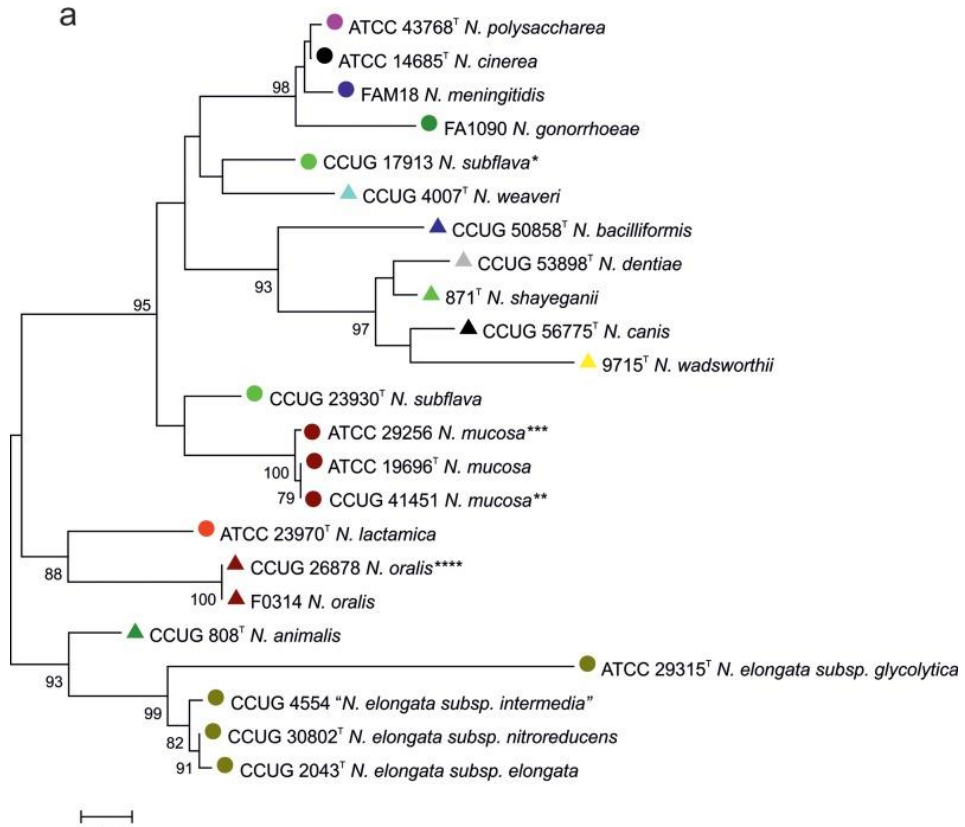
- **Solution:** identify a short gene fragment (~400bp in length) diagnostic for species.

- **Implementation:** Phylogeny of *Neisseria* species generated and compared to phylogenies of individual genes
  - A fragment of *rplF* gene was congruent with clusters in the phylogeny of all genes and used for the assay.

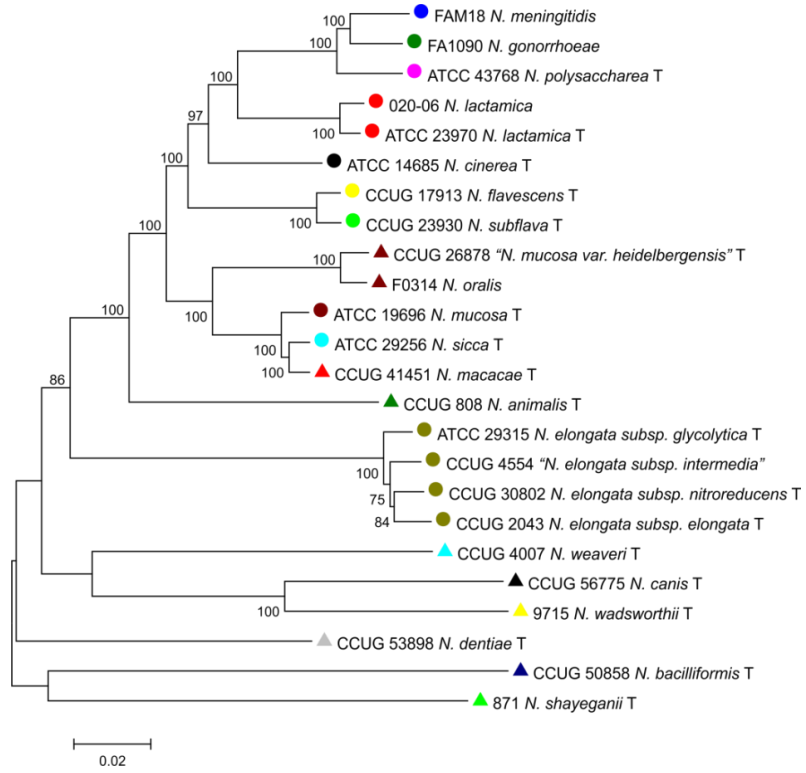
Bennett, J. S., Watkins, E. R., Jolley, K. A., Harrison, O. B. & Maiden, M. C. (2014). Identifying *Neisseria* species using the 50S ribosomal protein L6 (*rplF*) gene. *J Clin Microbiol* 52, 1375-1381.

Comparison of 16s rRNA sequencing and *rplF* sequencing

16s rRNA gene fragment phylogeny (single rRNA encoding gene)



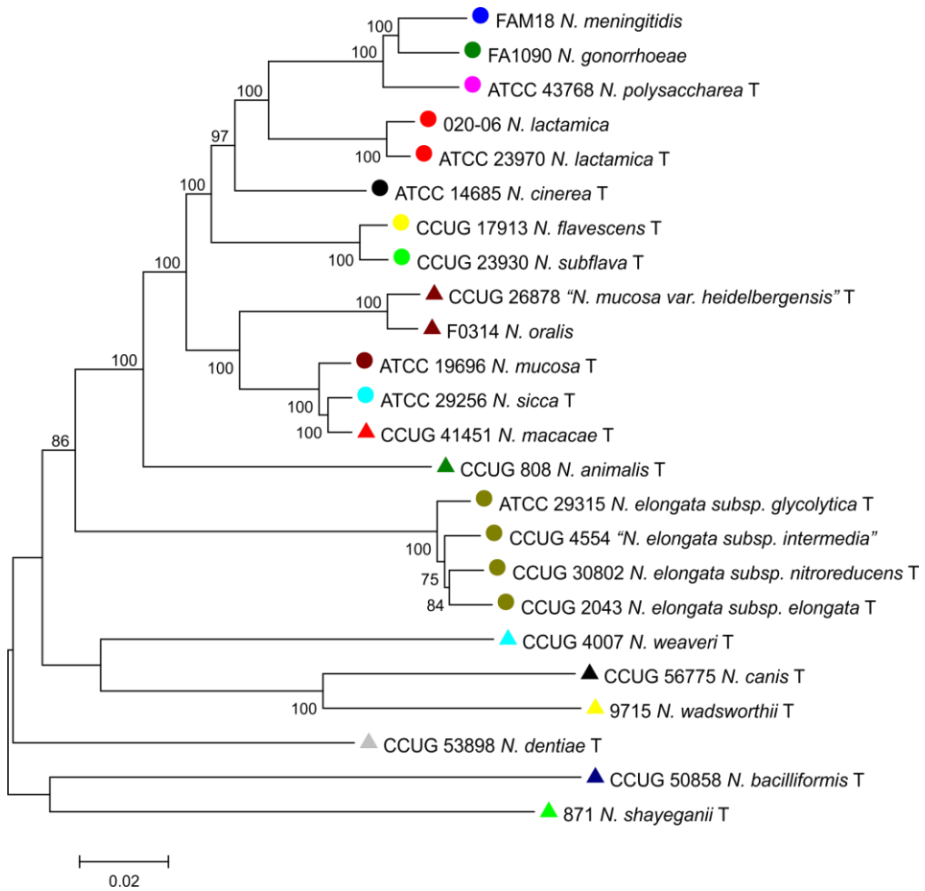
rMLST phylogeny (53 ribosomal protein encoding genes)



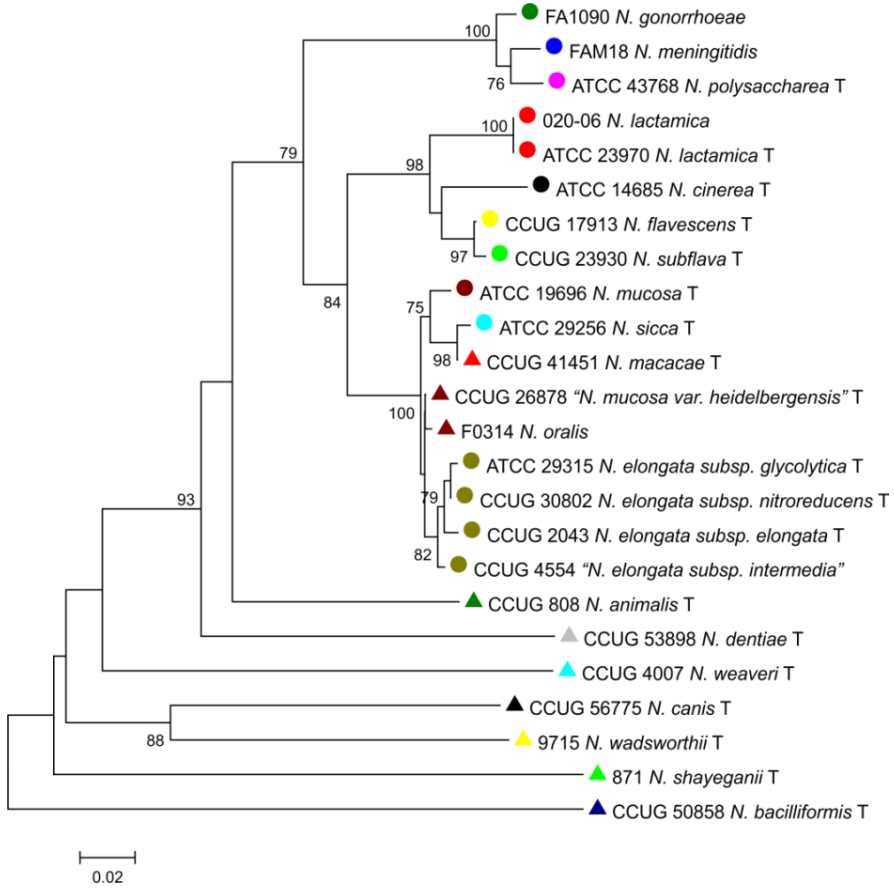
Bennett, J. S., Watkins, E. R., Jolley, K. A., Harrison, O. B. & Maiden, M. C. (2014). Identifying *Neisseria* species using the 50S ribosomal protein L6 (*rplF*) gene. *J Clin Microbiol* 52, 1375-1381.

Rapid species assignment: *rplF* sequence

rMLST phylogeny (53 genes)

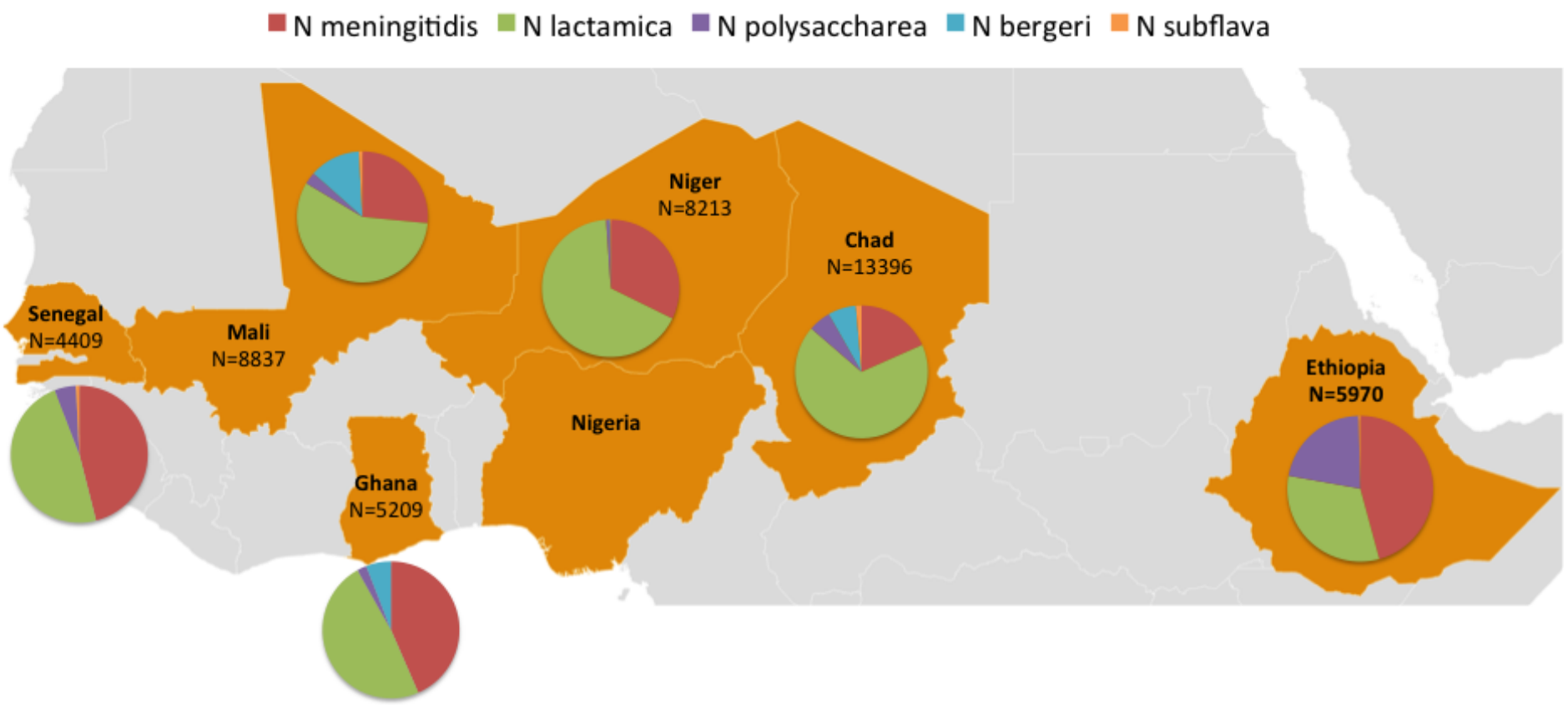


***rplF* fragment (413bp) phylogeny**



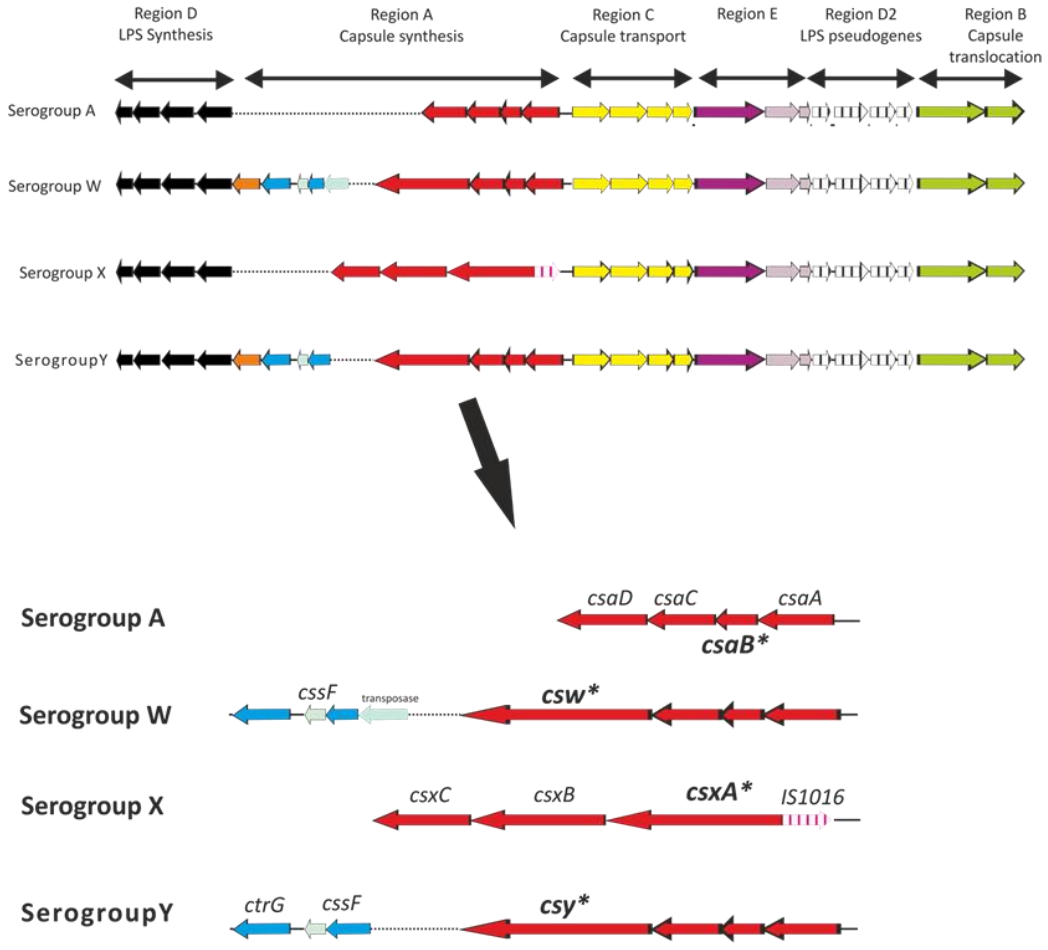
Bennett, J. S., Watkins, E. R., Jolley, K. A., Harrison, O. B. & Maiden, M. C. (2014). Identifying *Neisseria* species using the 50S ribosomal protein L6 (*rplF*) gene. *J Clin Microbiol* 52, 1375-1381.

Neisseria species distribution



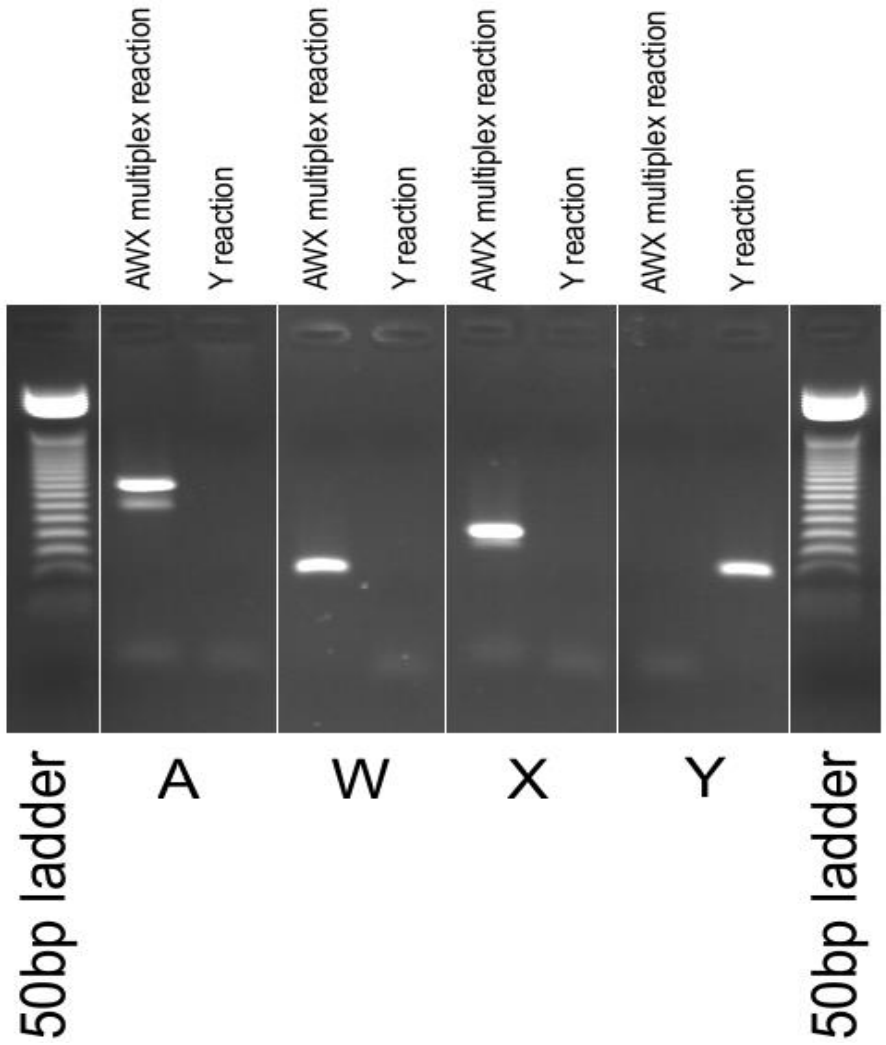
Diallo, K., *et al.*, Greenwood, B. M. & Maiden, M. C. (2016). Pharyngeal carriage of *Neisseria* species in the African meningitis belt. *J Infect* **72**, 667-677.

Meningococcal genogrouping: a simpler problem



Harrison, O. B., Claus, H., Jiang, Y., Bennett, J. S., Bratcher, H. B., Jolley, K. A., Corton, C., Care, R., Poolman, J. T., Zollinger, W. D., Frasch, C. E., Stephens, D. S., Feavers, I., Frosch, M., Parkhill, J., Vogel, U., Quail, M. A., Bentley, S. D. & Maiden, M. C. J. (2013). Description and nomenclature of *Neisseria meningitidis* capsule locus. *Emerging Infectious Diseases* 19, 566-573.

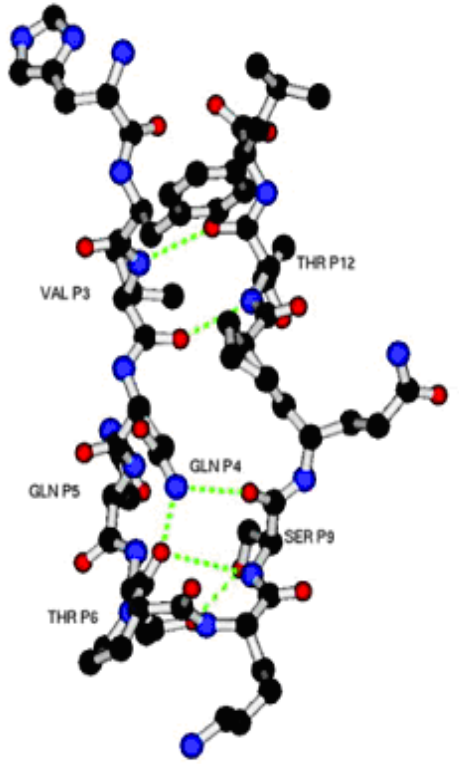
Genogrouping: a PCR will do it



PorA sequence variation and serosubtyping

- Major outer membrane protein (OMP);
- Serosubtyping antigen (originally with mAbs) and a major vaccine component;
- Most antigenic variability in two variable regions (VRs) of the protein and gene.
- A single PCR amplification and sequencing reaction can provide high discrimination among isolates.

Structure of PorA P1.2 variant

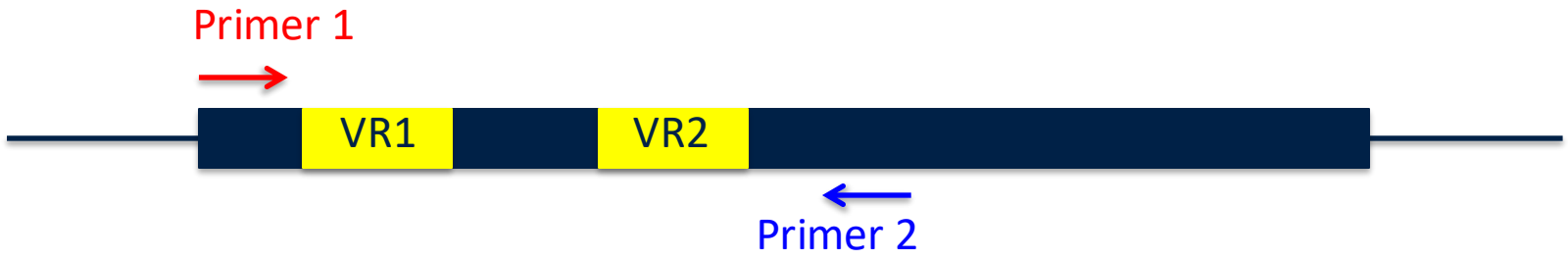


Tzitzilonis, C., Prince, S. M., Collins, R. F., Achtman, M., Feavers, I. M., Maiden, M. C. & Derrick, J. P. (2006). Structural variation and immune recognition of the P1.2 subtype meningococcal antigen. Proteins 62, 947-955.

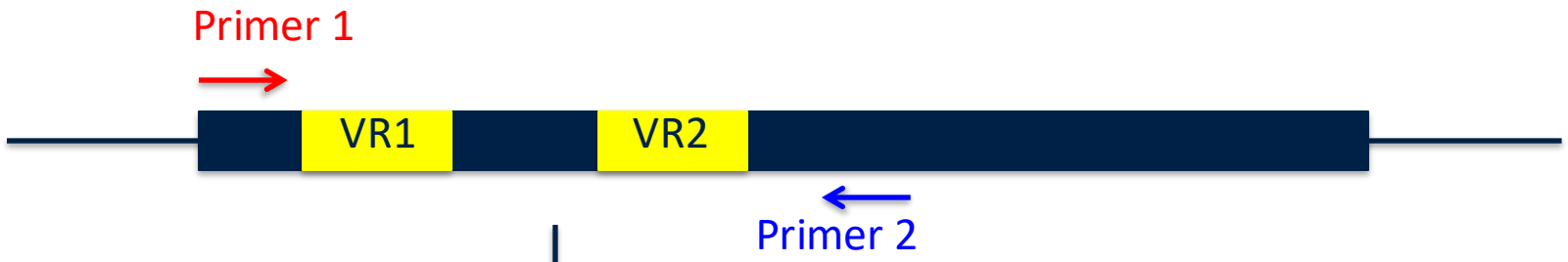
Meningococcal characterization: sequence-based subtyping

- The two most variable parts of the protein (VR1 and VR2) are close together,
  - one sequencing reaction can determine the parts of the gene encoding both of these.

- *porA* gene fragment sequencing therefore extracts a lot of information for the amount of sequencing performed.

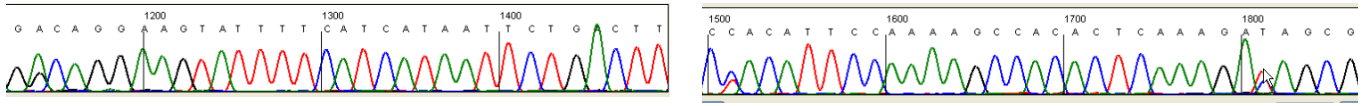


Sequencing *porA* VR1 and VR2



Amplify & purify

Sequence Both strands



Assemble

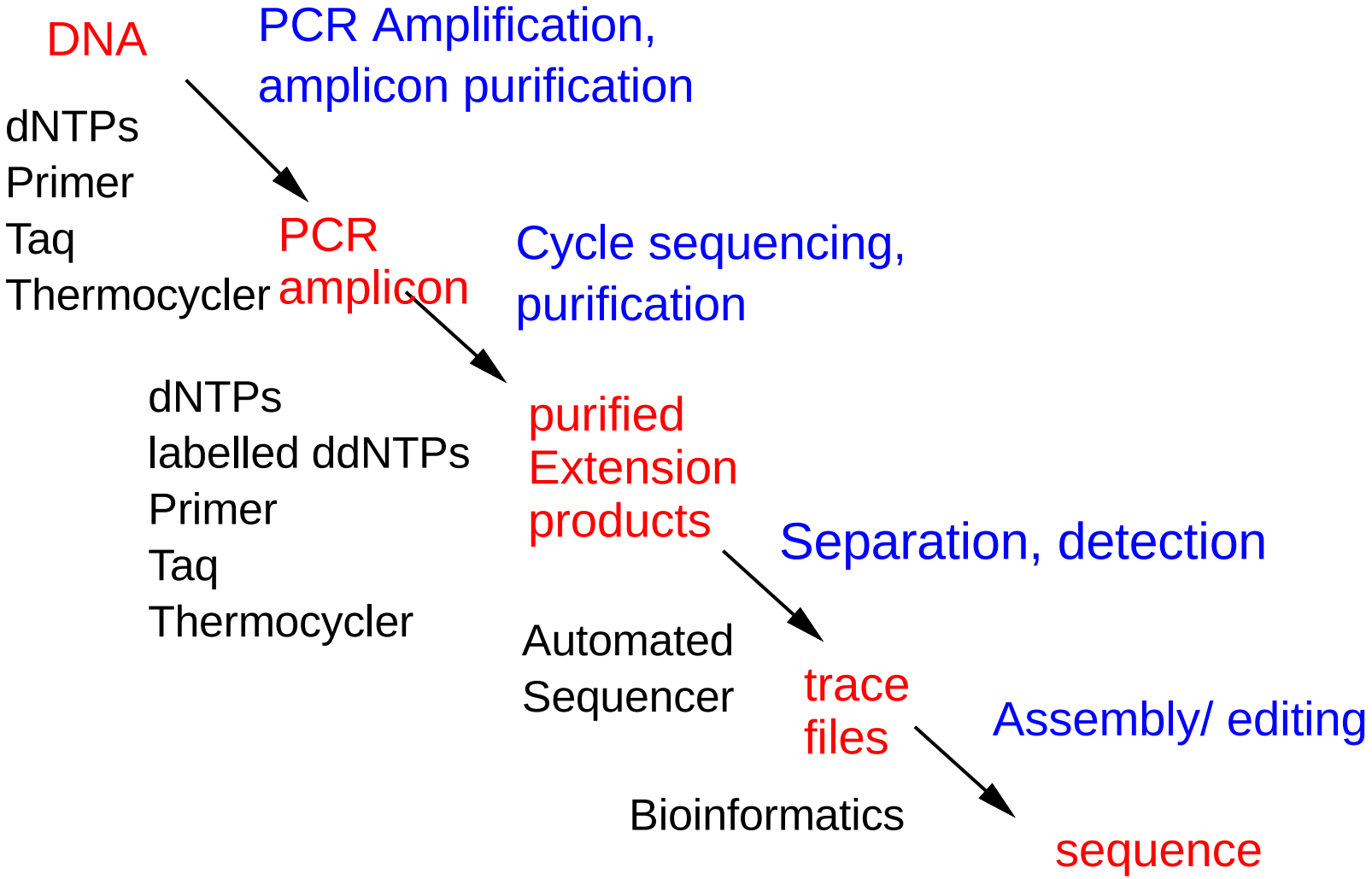
```
gaaggcagga acatccagct gcaattgacc gagcagccct caaaggctca aggtcaaa
aacaatcagg taaaagttac taaggcaaaa agccgcatca ggacgaaaat cagcgatt
ggctcgttta tcggctttaa ggggagcgag gatttgggcg aagggctgaa ggctgttt
```

BLAST

Result: NB – this is based on deduced peptide sequence

P1.5-2,10

Direct sequencing:
steps, reagents, equipment



Exercise

Specimen	Patient age	Presentation	Organism	Serogroup	PorA
2	15 months	Meningitis	*N. meningitidis*	B	
5	3 years (46 months)	Meningitis	*N. meningitidis*	B	

The laboratory has been able to obtain sequences of the *porA* gene (encoding the PorA protein) from each of the specimens:

- Specimen 2 - from DNA extracted from the culture;
- Specimen 5 – from direct sequencing of an amplicon of the *porA* gene obtained directly from the CSF.

Query the PubMLST.org/neisseria website with the sequence data provided, to identify the PorA variants present in each of these meningococci.

Exercise -Results

Specimen	Patient age	Presentation	Organism	Serogroup	PorA
2	15 months	Meningitis	*N. meningitidis*	B	P1.7-2,4
5	3 years (46 months)	Meningitis	*N. meningitidis*	B	P1.7-2,4

Use the sequence data provided, along with the PubMLST.org/neisseria website, to identify the PorA variants present in each of these meningococci.

- Does this information reassure you, or not?
- Why?
- What should you do now (if anything)?

Encapsulated Bacteria: Web-based sequence querying

Genomics and Clinical Microbiology 2024

Martin Maiden, Made Krisna, Kasia Parfitt, Keith Jolley,

Department of Biology

UNIVERSITY OF
OXFORD

The plot thickens!

A further case of meningitis another family member has been recorded; however, this CSF sample yielded no further information.

A review of surveillance data shows four other cases of meningococcal disease occurred in this extended family over a two-year period:

Specimen 6 – Derived from CSF of 6 month-old with invasive meningococcal disease. This was confirmed as containing serogroup B meningococcal DNA by real-time PCR. There is no isolate for this specimen.

Specimen 7 – Derived from CSF from a 5 month-old child. This has been confirmed as containing serogroup B meningococcal DNA by Real-time PCR. An isolate has also been obtained from this specimen and stored.

Specimen 8 – Derived from CSF from a 5 month-old child. This has been confirmed as containing serogroup B meningococcal DNA by Real-time PCR. There was no isolate with this specimen.

Specimen 9 – Derived from CSF from a 10 month-old child. This has been confirmed as containing serogroup B meningococcal DNA by Real-time PCR. There was no isolate with this specimen.

Questions

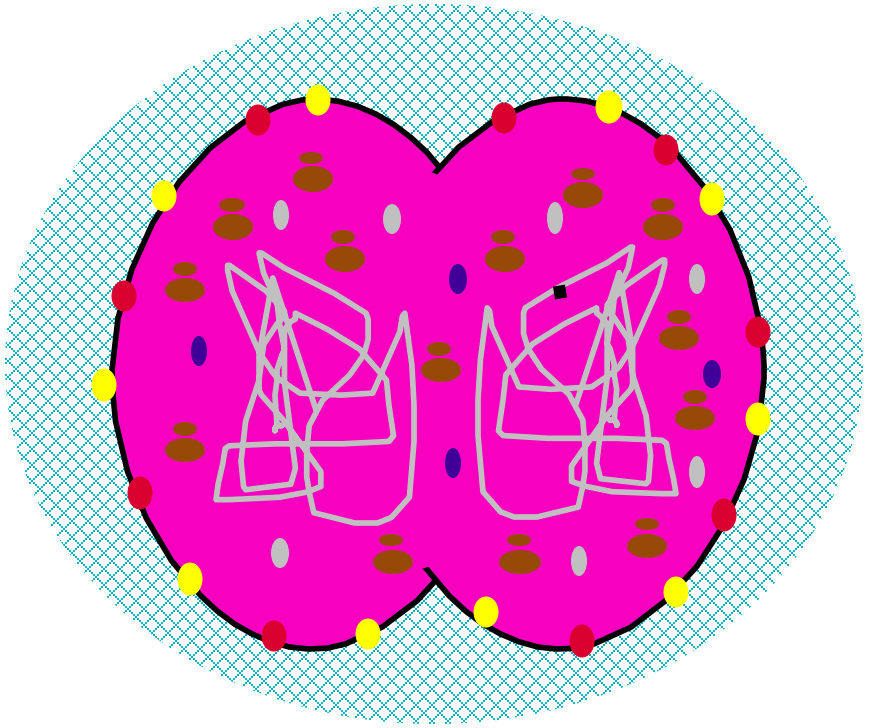
- What is your reaction to this series of cases?
- What action may be required?
- What additional information to your need?
- What additional support might you need?

Multi locus typing approaches

While single locus typing (e.g. 16s rRNA sequencing is useful in many circumstances there are advantages in typing multiple loci, including:

- Increased resolution;
- Additional information,
  - antimicrobial resistance, vaccine antigens, virulence determinants;
- Mitigating the impacts of recombination on isolate characterization (more on this later);
- Identification of microbial lineages.

Neisseria meningitidis characterisation

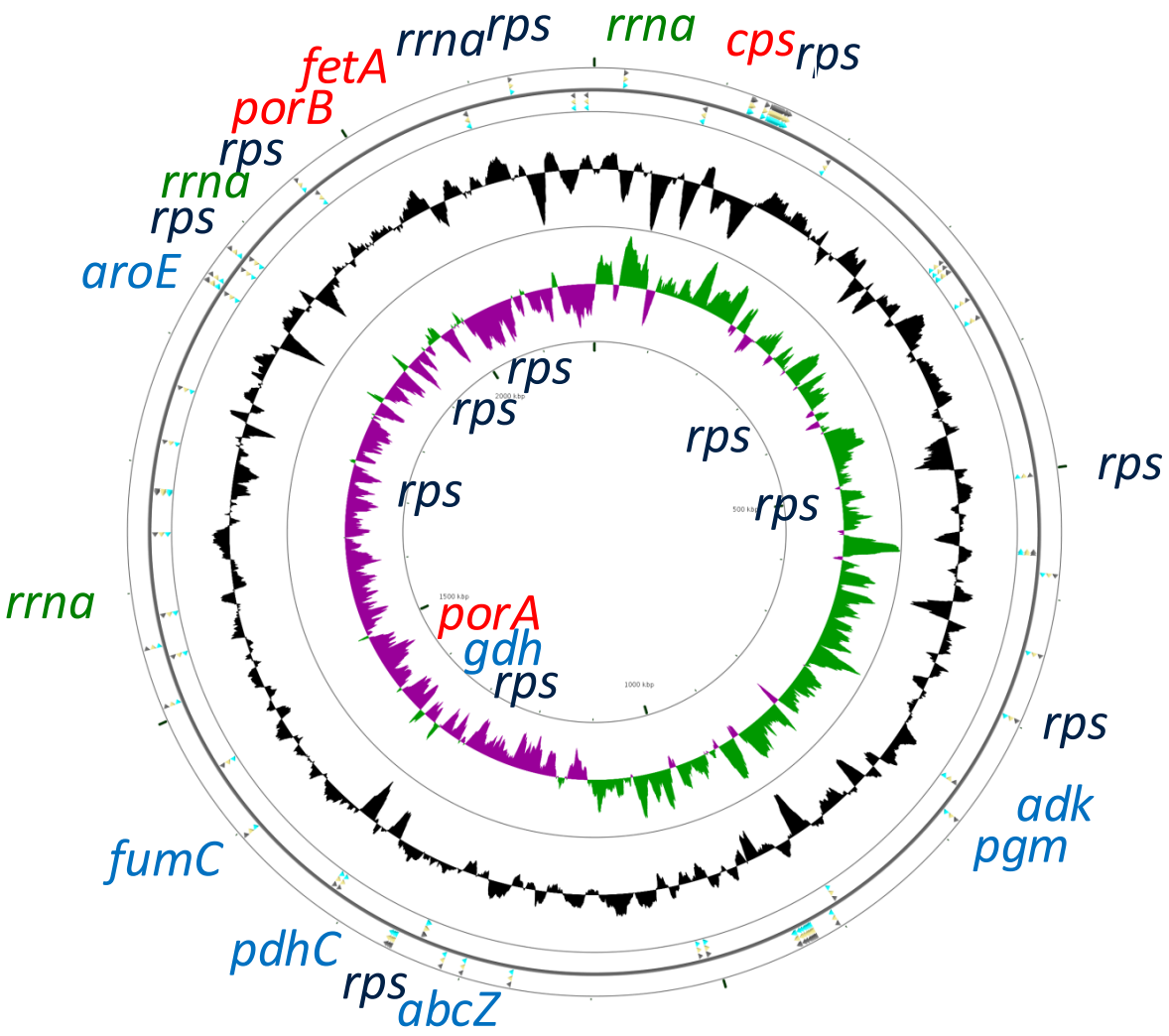


Component	Phenotypic	Genotypic
Capsule	Serogroup	*cps* region
OMPS	Serotype, Subtype, etc.	*porA, porB, fetA,* etc.
Housekeeping genes	MLEE	MLST
Ribosomes	MALDITOF	16s rRNA, rMLST

***Neisseria meningitidis* B: P1.7,16: F3-3: ST-32 (cc32)**

Jolley, K. A., Brehony, C. & Maiden, M. C. (2007). Molecular typing of meningococci: recommendations for target choice and nomenclature. *FEMS Microbiol Rev* **31**, 89-96.

Typing targets in the meningococcal genome



Antigen genes:

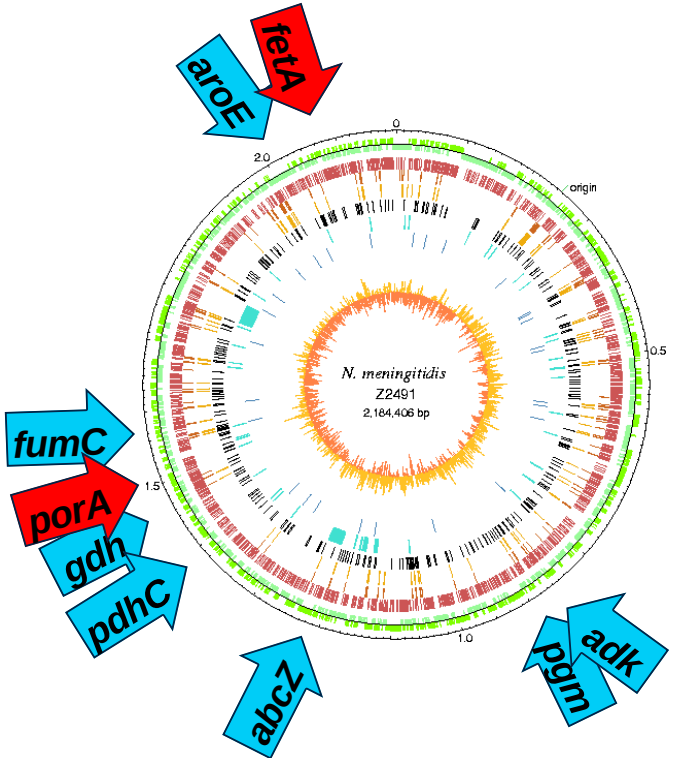
- Capsule (*cps*), serogroup;
- OMPs (*porA, porB, fetA*), serotype, subtype fine type;
- BAST antigens.

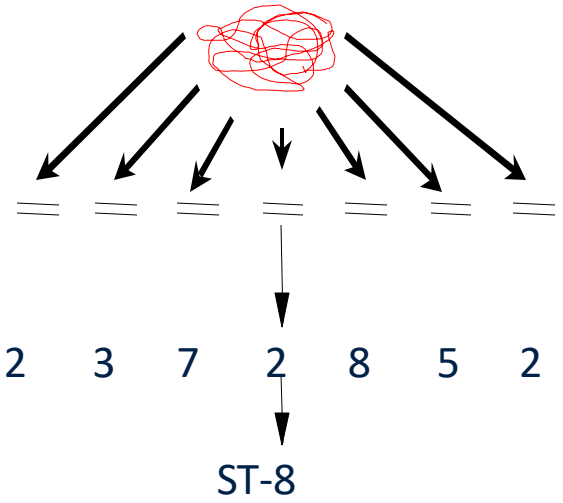
Housekeeping genes:

- Ribosomal RNA genes (*rrna*) genus identification;
- Ribosome protein genes (*rps*), species and sub species characterisation (rMLST);
- MLST (metabolism), subspecies identification
- cgMLST, high-resolution typing.

Antibiotic Resistance genes.

First generation genomic typing: single locus and MLST





B: P1.7,16: F5-1: ST-33 (cc32)

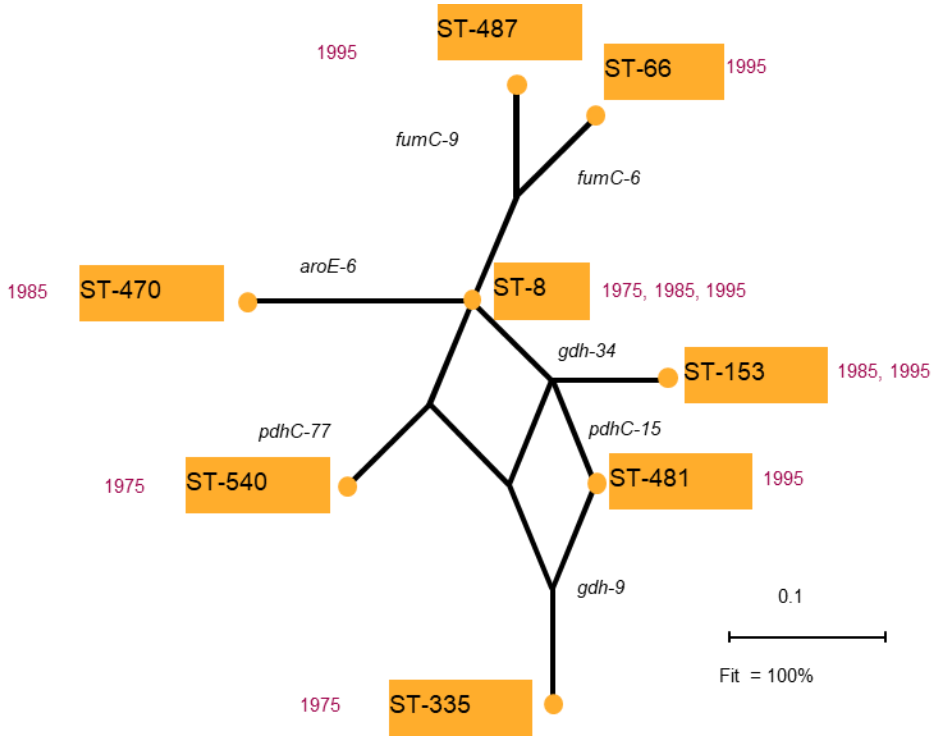
Antigen type (fine type)

Sequence type & clonal complex

Maiden, MCJ, Bygraves, JA, Feil, E, Morelli, G, Russell, JE, Urwin, R, Zhang, Q, Zhou, J, Zurth, K, Caugant, DA, Feavers, IM, Achtman, M & Spratt, BG. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95, 3140-3145.

MLST: allele-based analyses

- *Neisseria* seven-locus ST summarises 3,284bp.
- That is 0.15% of the 2.18Mbp genome.
- 14,199 STs in PubMLST database (20th January 2019).
- 702-1036 alleles per locus.
- Many polymorphisms per locus.



Russell, J. E., Urwin, R., Gray, S. J., Fox, A. J., Feavers, I. M. & Maiden, M. C. (2008). Molecular epidemiology of meningococcal disease in England and Wales 1975-1995, before the introduction of serogroup C conjugate vaccines. *Microbiology* 154, 1170-1177.

PubMLST.org/neisseria

Exercise

- Use the sequence information supplied with the https://pubmlst.org/neisseria website to complete the following information for each specimen as far as possible:
  - MLST;
  - PorA antigen;
  - FetA antigen.
- Combine with the epidemiological and clinical information.
- What are your conclusions?

Exercise

specimen	age	town	serogroup	*abcZ*	*adk*	*aroE*	*fumC*	*gdh*	*pdhC*	*pgm*	ST	Clonal complex	PorA	FetA
6	6 mo	A	B											
7	5 mo	A	B											
8	5 mo	A	B											
9	10 mo	B	B											
2	15 mo	A	B											
5	3 yrs	A	B											
non-typeable	9 mo	B	-	-	-	-	-	-	-	-	-	-	-	-

Exercise

specimen	age	town	serogroup	*abcZ*	*adk*	*aroE*	*fumC*	*gdh*	*pdhC*	*pgm*	ST	Clonal complex	PorA	FetA
6	6 mo	A	B	3	6		24	11	6				7-2,4	F1.21
7	5 mo	A	B	3	6	9	24	11	6	9	6697	41/44	7-2,4	F1.21
8	5 mo	A	B	3	6	9	24	11	6	9	6697	41/44	7-2,4	F5.12
9	10 mo	B	B	4	5	2				20				
2	15 mo	A	B	3	6	9	24	11	6	9	6697	41/44	7-2,4	F5-12
5	3 yrs	A	B		6	9	24		6			41/44	7-2,4	
non-typeable	9 mo	B	-	-	-	-	-	-	-	-	-	-	-	-

Extended meningococcal disease outbreak in Éire

specimen	Age (mo)	town	serogroup	*abcZ*	*adk*	*aroE*	*fumC*	*gdh*	*pdhC*	*pgm*	ST	cc	PorA	FetA
A (03/2010)	6	A	B	3	6		24	11	6			41/44	P1.7-2,4	F1-21
B * (11/2010)	5	A	B	3	6	9	24	11	6	9	6697	41/44	P1.7-2,4	F1-21
C (11/2010)	5	A	B	3	6	9	24	11	6	9	6697	41/44	P1.7-2,4	F5-12
D (01/2012)	10	B	B	4	5	2				20				
E (03/2013)	15	A	B	3	6	9	24	11	6	9	6697	41/44	P1.7-2,4	F5-12
F (04/13)	46	A	B		6	9	24	-	6			41/44	P1.7-2,4	
G (06/2013)	9	B	-	-	-	-	-	-	-	-	-	-	-	-
H* (11/2013)	6	A	B	3	6	9	24	11	6	9	6697	41/44	P1.7-2,4	F5-12

Discussion points

- What do you now understand about the relationships among these specimens?
- What additional information/actions are possible with these molecular typing data?
- What public health action, if any, is required?

How Sanger (dideoxy chain termination) sequencing works

Martin Maiden

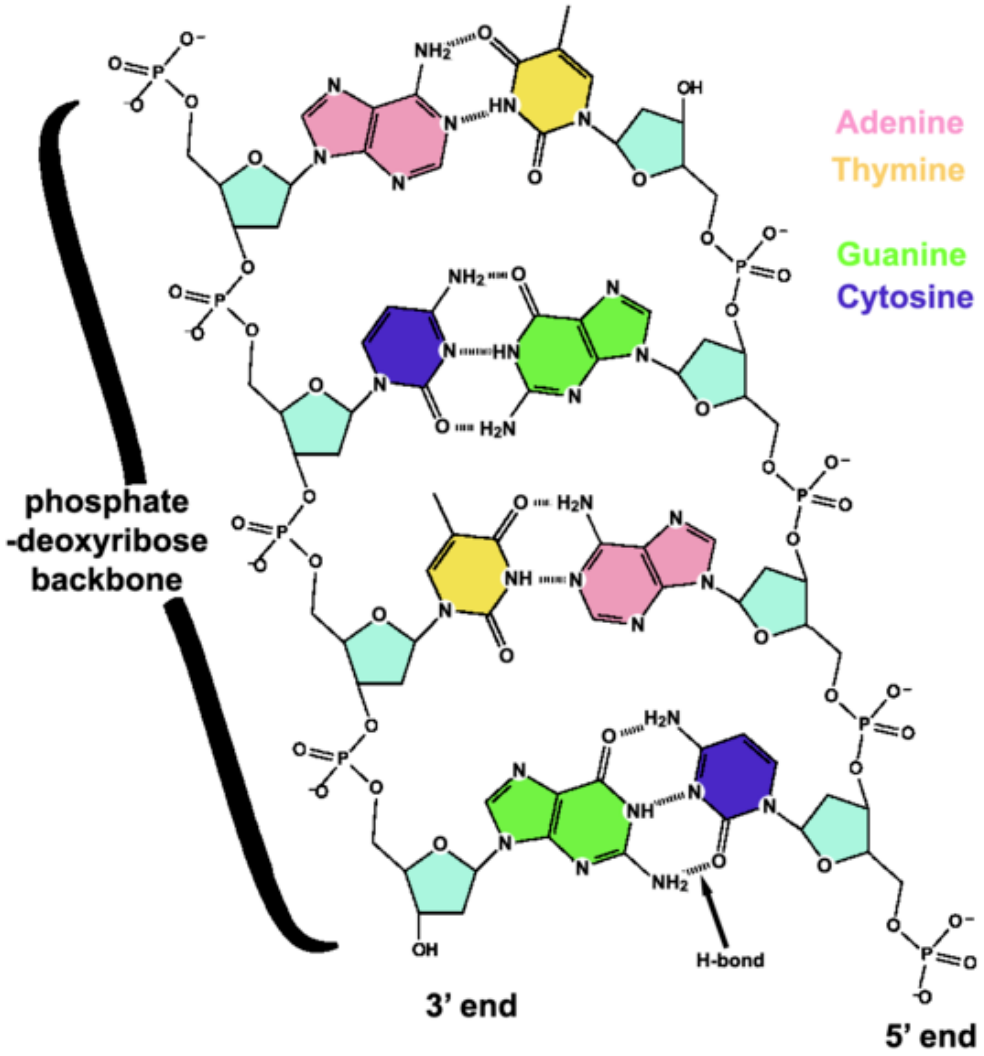
UNIVERSITY OF
OXFORD

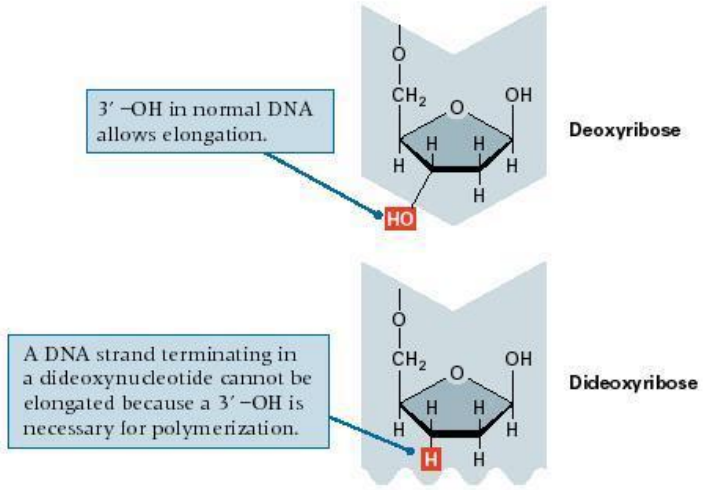
The remarkable contributions of Fred Sanger

- Linear structure of biological information.
- Protein sequencing,
  - insulin sequence.
- DNA sequencing,
  - dideoxy method.
- An efficient means of determining *phenotype* is inferring it from *genotype:*
  - and that means DNA sequence;
  - protein sequences can be deduced from DNA sequence.

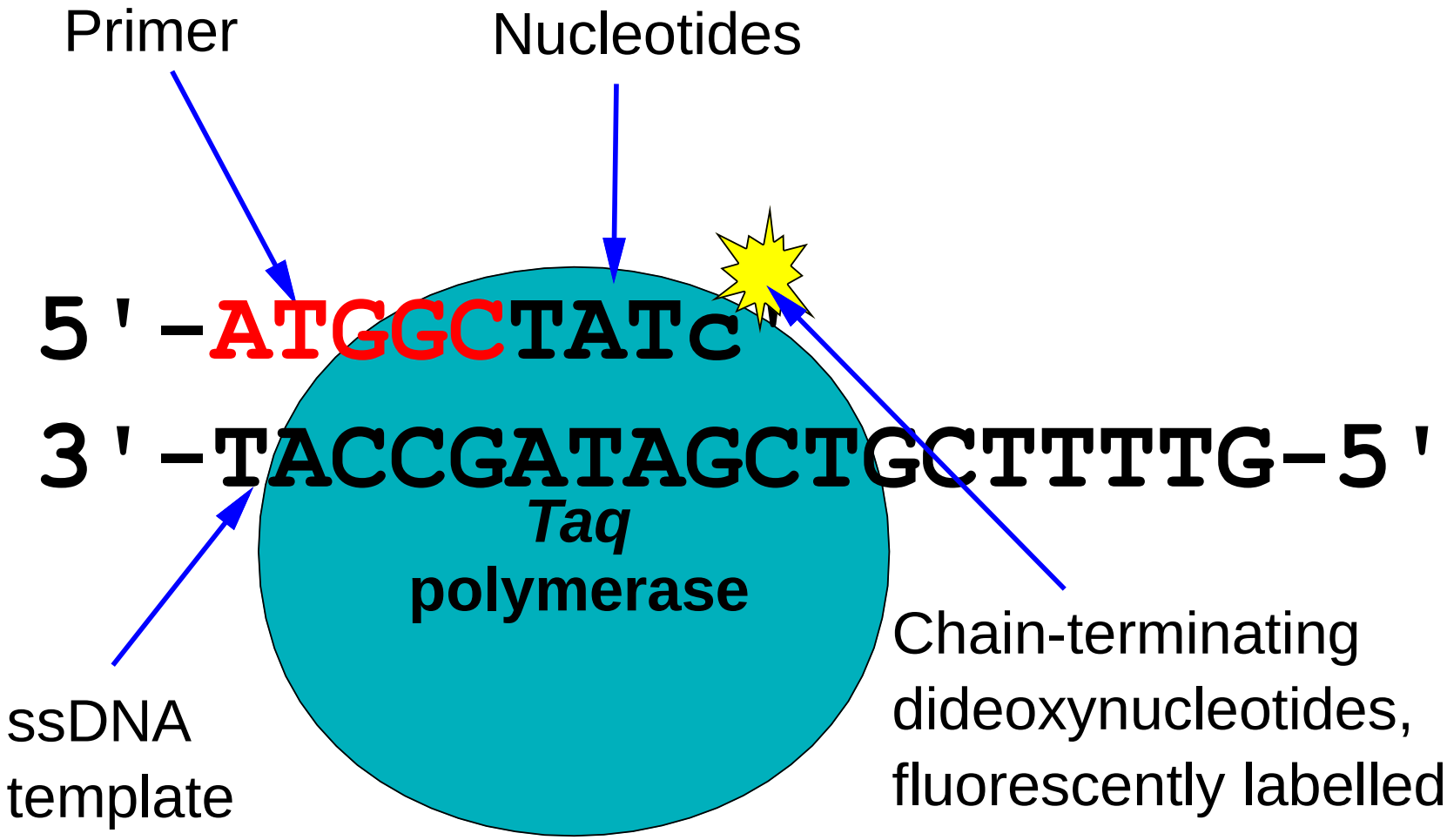
Fred Sanger 1918-2013: Photo MRC

Di-deoxy chain-terminating nucleotides





Key components of Sanger sequencing (sequencing by synthesis)



The sequence ladder generated by cycle sequencing

3'-TACCGATAGCTGCTTTTG-5'

ATGGCt

ATGGCTa

ATGGCTAt

ATGGCTATc

ATGGCTATCg

ATGGCTATCGa

ATGGCTATCGAc

ATGGCTATCGACg

Deoxynucleotides:
A, C, G, T.

Dideoxynucleotides:
a, c, g, t.

Fluorescent Labels:

Detection of sequence ladders

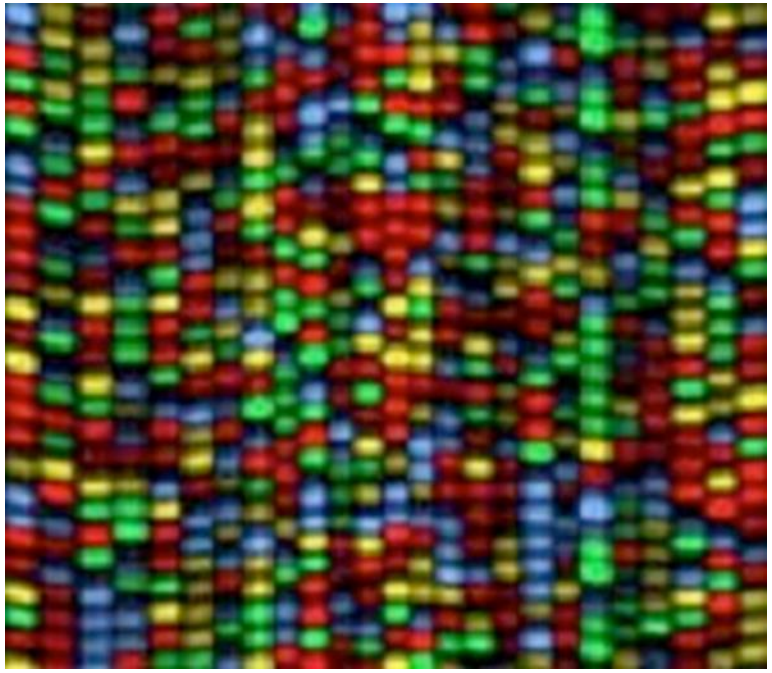
- Separation of reaction products:
  - Gel based, e.g. ABI 377;
  - Capilliary, e.g. ABI 3730.

This contrasts with …

- Progressive detection of products:
  - e.g. Illumina platform.

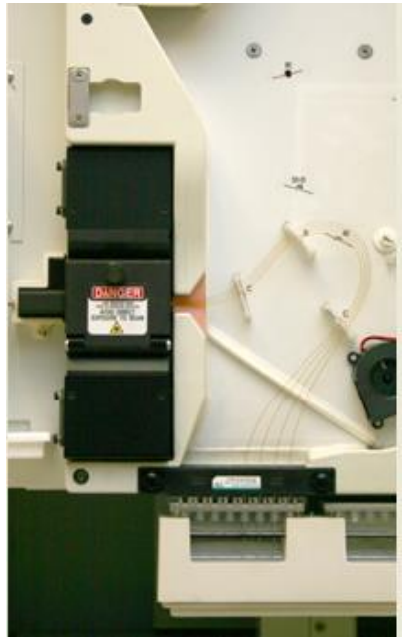
Gel-based detection (…now essentially obsolete)

- Sequence reactions loaded on to an acrylamide slab gel.
- Separated by electrophoresis;
- Laser/camera system detects the products as they pass the end of the gel.



Capillary sequencing (still used)

- Sequence reactions loaded into a capillary filled with acrylamide.
- Separated by electrophoresis though the capillary
- Laser/camera system detects the products as they pass a window at the end of the capillary.
- Capillaries are automatically flushed-out and re-filled with polymer between samples.



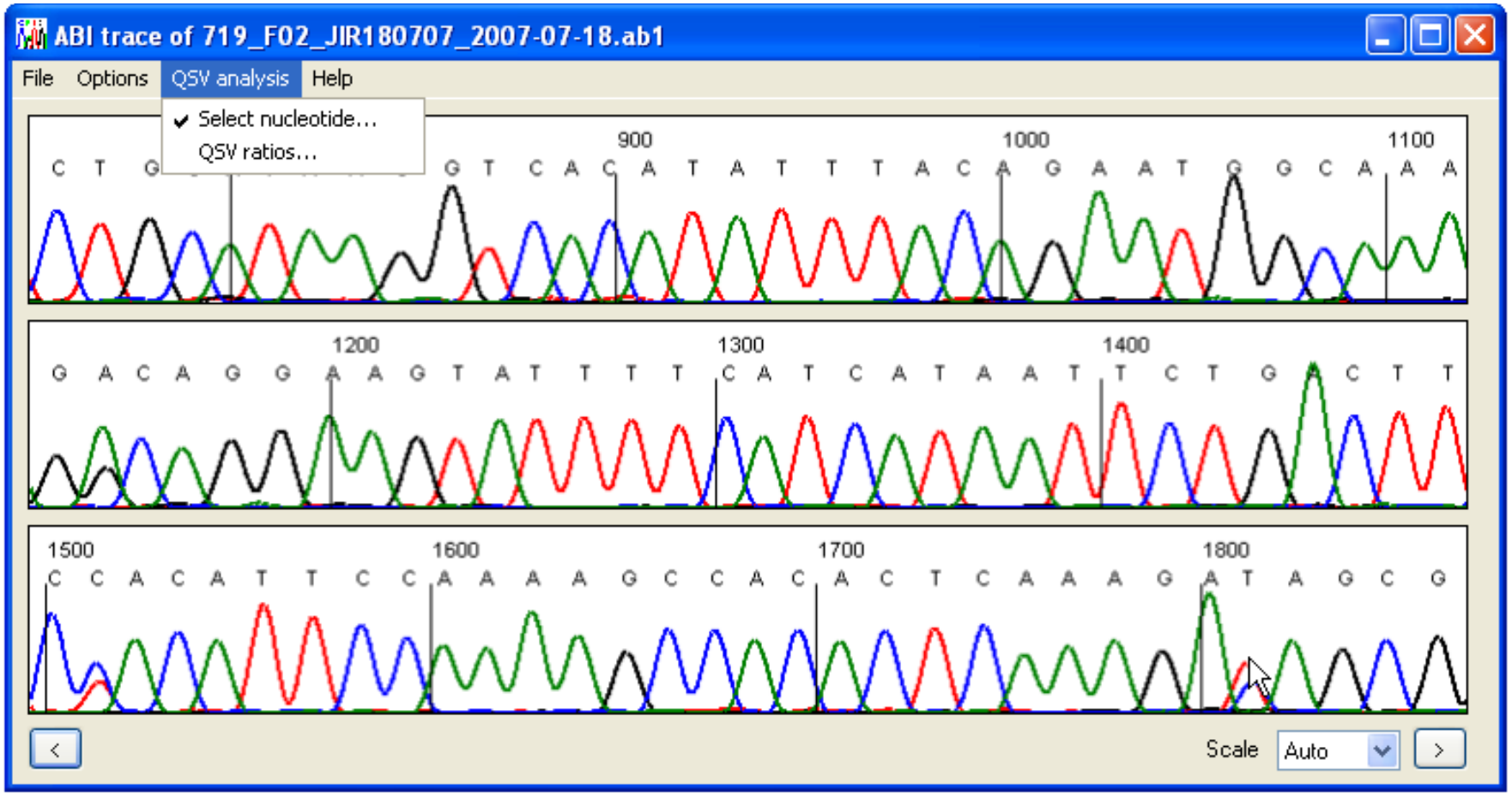
Separation of cycle sequencing products

- Performed using automated sequencers
  - e.g. Applied Biosystems capillary sequencers
- Different sequencers have different capacities:
  - ABI 310 single capillary;
  - ABI 3100 16 capillary;
  - ABI 3070 48 or 96 capillary.

ABI 3730

Inside the ABI 3730

Visualisation of sequence traces



Sequencing by synthesis progresses one way … and sometimes goes wrong

```
3'-TACCGATAGCTGCTTTTG-5'
   ATGGCt
   ATGGCTt
   ATGGCTAt
   ATGGCTATc
   ATGGCTATCg
   ATGGCTATCGa
   ATGGCTATCGAg
   ATGGCTATCGACg
```

Deoxynucleotides:
A, C, G, T.

Dideoxynucleotides:
a, c, g, t.

Fluorescent Labels:

Errors can be mitigated by redundancy

- In a typical Sanger sequencing reaction there are hundreds (or thousands) of templates.

- These can be primed many times, increasing coverage.

Note:

- As the number of templates goes down, the possibility of random (non-systematic) error increases.

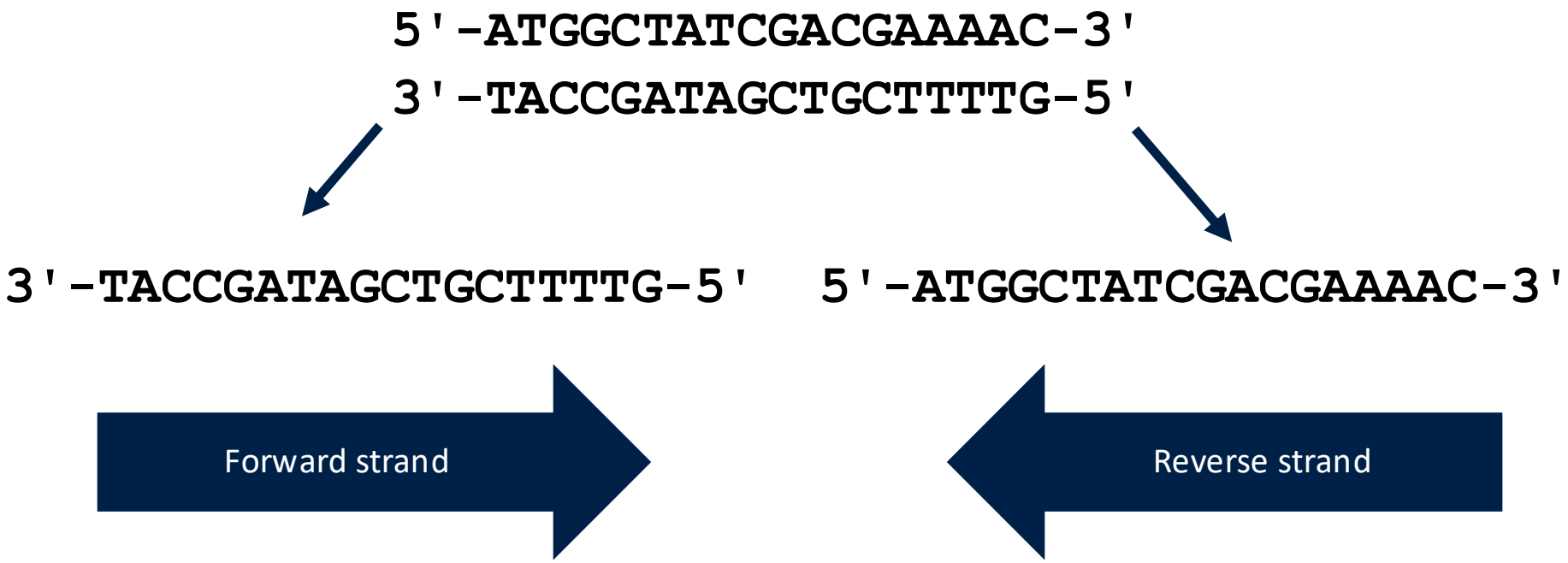
- Some errors are systematic – a ‘wrong’ base is likely to be incorporated in *in vitro* sequencing or the sequence is had to read.

Accuracy by redundancy

```
Read 1      ATGGCTATCGACGAAAAC
Read 2      ATCGCTATCGACGAAAAC
Read 3      ATGGCAATCGACGAAAAC
Read 4      ATGGCTATCGAAGAAAAC
Read 5      ATGGCTATCGACGAAATC
Read 6      TTGGCTATCGACGAAAAC
Read 7      ATGGCTATCCACGAA-AC
Read 5      ATGGCTATCGACGAAAAG
Read 6      ATGGCTAACGACGAAAAC
Read 7      ATGGCTATCGACGTAAAC

Consensus   ATGGCTATCGACGAAAAC
```

Improved accuracy is achieved with sequencing both strands



The data from the reverse strand is reverse complemented (usually with a computer these days!) and used to generate a consensus sequence.

Factors affecting sequence accuracy

- Quality of the sequencing method.
- Number of templates in the reaction.
- Number of repeats.
- Strand coverage.

Note that although the best possible accuracy should be the goal, even poor quality data may be helpful clinically.

The quality of the data must be sufficient for the clinical need.