Use of next generation sequencing in epidemiology

Dr. Lori Snyder
Senior Lecturer in Biotechnology
Kingston University, UK

Why do genome sequencing?

- Learn about something new
  - New bacteria
  - New disease
  - New manifestation of disease
  - New outbreak
- Learn more about something we think we know
  - Find genes for new traits
  - Discover new regulatory networks
  - Find that similar bacteria are actually very different

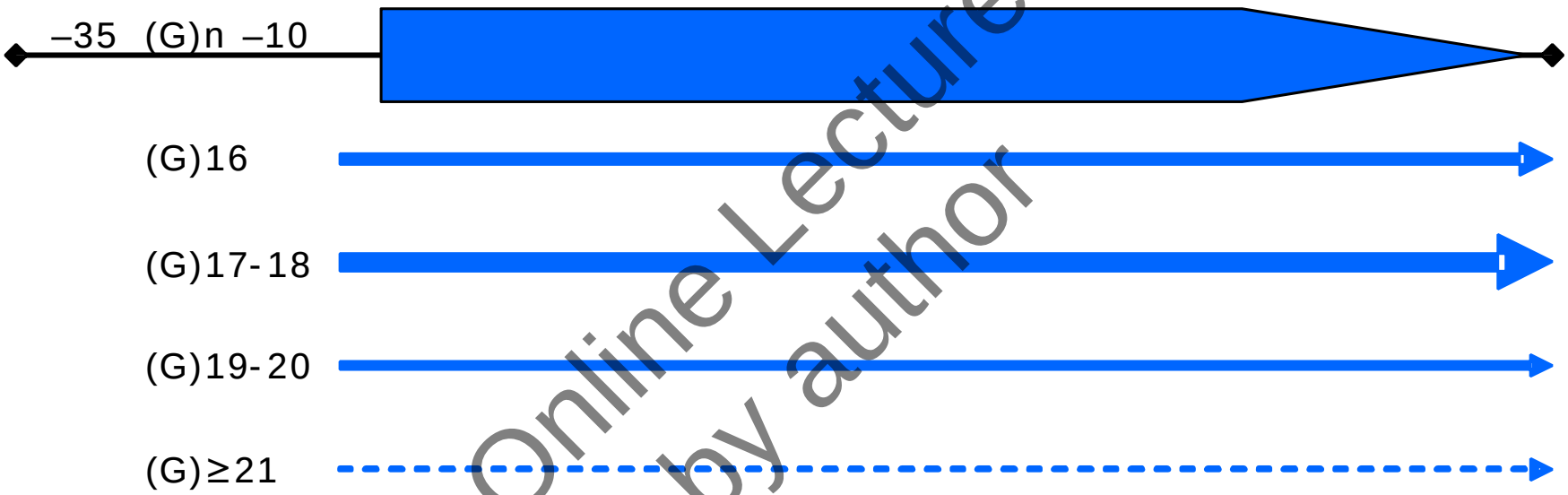
Helicobacter canadensis and *Helicobacter pullorum*

- Emerging pathogens
- Severe diarrhoea, fever, and septicaemia
- Associated with:
  - Inflammatory bowel disease
  - Crohn's disease
  - Irritable bowel syndrome
- Transmitted from wild birds, rodents, pigs, and potentially food and water
- Genome sequencing to tell us more

Phase variation

- *Helicobacter* and *Campylobacter* have phase variable genes
- Identifiable from the genome sequence data
- Repertoires of phase variable genes differ
- Search for phase variable genes in the genome sequence of *H. canadensis*
  - Find phase variation mediating repeats
  - Determine if the repeats are located in association with a CDS

Transcriptional phase variation



Translational phase variation

H. canadensis phase variable genes

- 21 CDSs identified with homopolymeric tracts
  - 5 Transcriptional phase variation
  - 16 Translational phase variation
    - All poly-G tracts
    - Poly-C, -A, -T, -GA, -CT, -TC, -AT, and –AG in other *Helicobacter* spp.
- Next-generation sequencing read data revealed population-level repeat tract variation.
- All identified phase variable genes had changes.

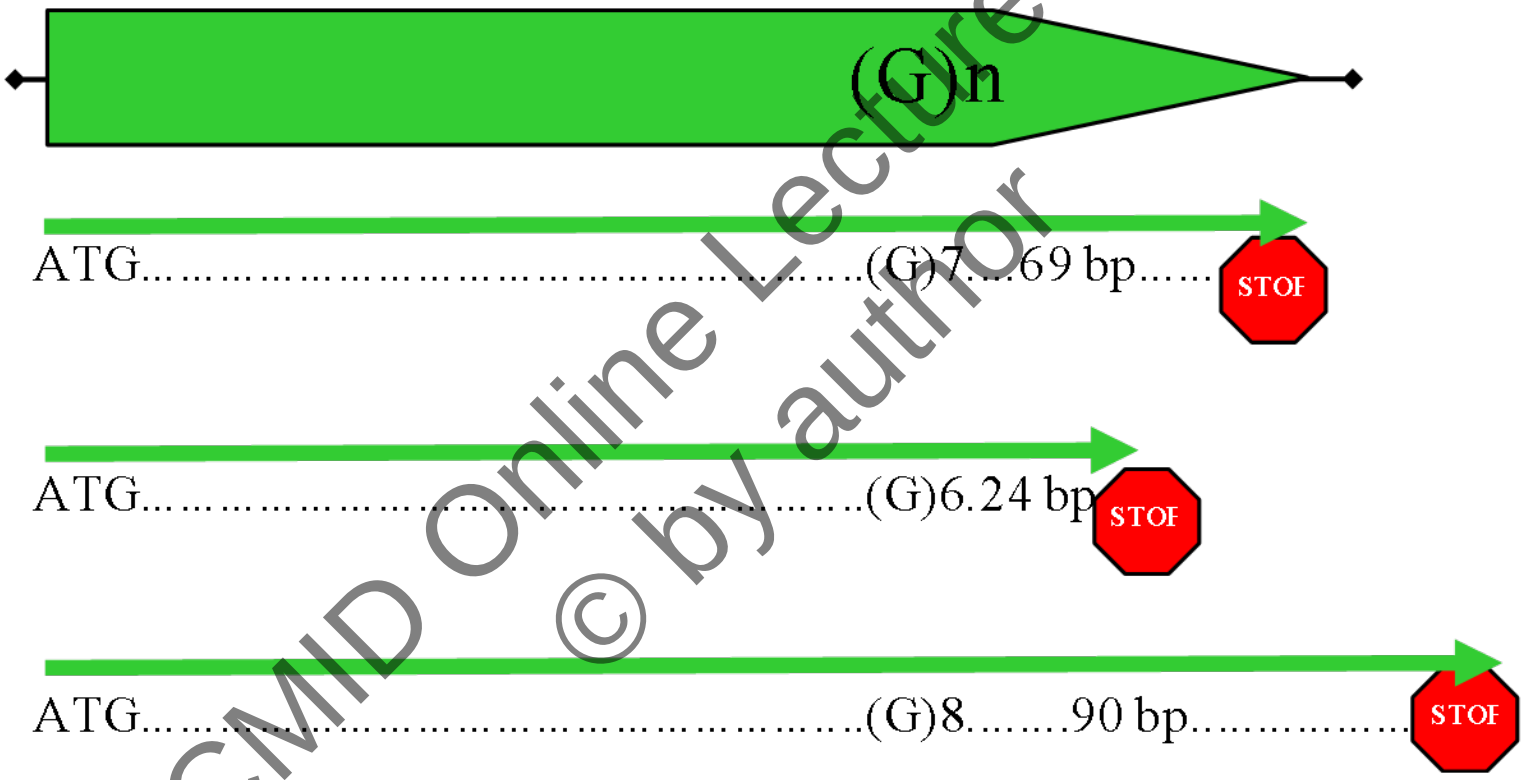
HCAN_0075 (hypothetical) repeat

AAGGGGGGGGGGGGG----C
AAGGGGGGGGGGGG----C
AAGGGGGGGGGGGGGG--C
AAGGGGGGGGGGG-----C
AAGGGGGGGGGGGGGGGGC
AAGGGGGGGGGGG----C
AAGGGGGGGGGGGG----C
AAGGGGGGGGGGGGGGGGC
AAGGGGGGGGGGGGGGG--C
AAGGGGGGGGGGGGG----C
AAGGGGGGGGGGGGGG----C
AAGGGGGGGGGGGG-----C
AAGGGGGGGGGGGGG----C

What do these phase variable genes tell us about *H. canadensis*?

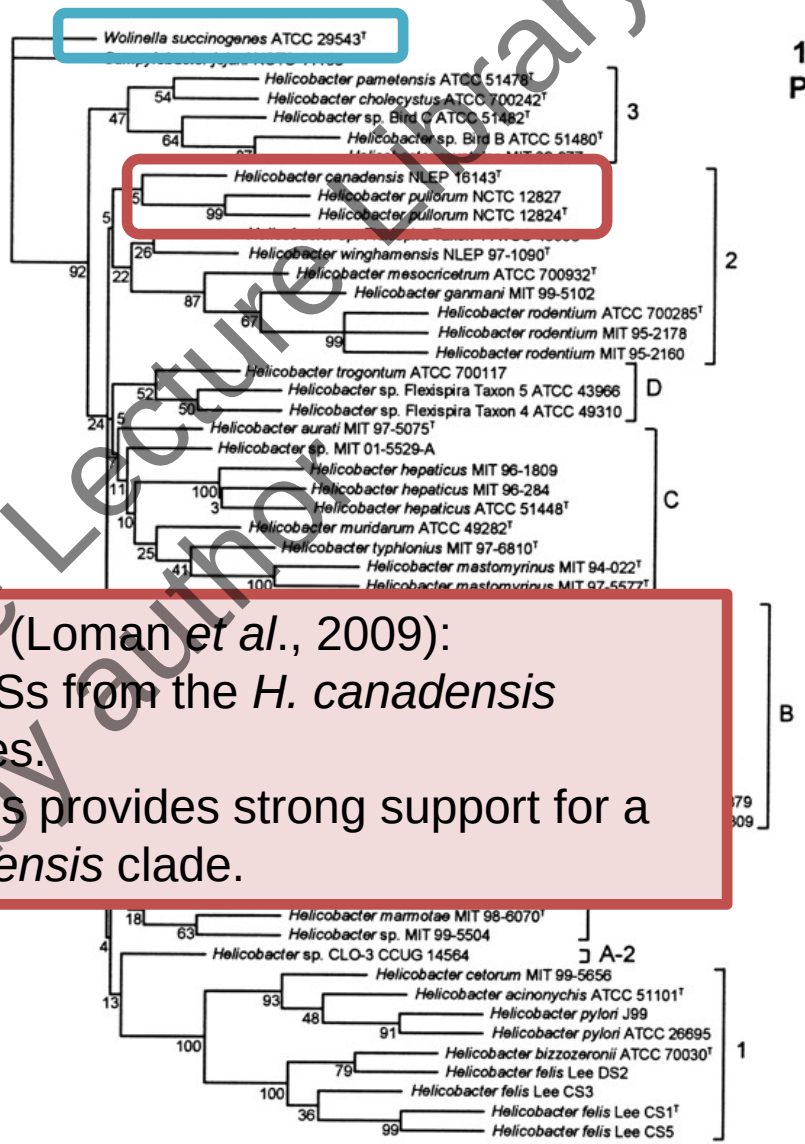
- It has an IgA protease.
  - Mucosal surface survival.
  - Like *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Haemophilus influenzae*, *Streptococcus pneumoniae*.
- There are several vacuolating cytotoxins.
  - Gastric epithelial necrosis.
  - Like *Helicobacter pylori* and avian pathogenic *E. coli*.
- *H. canadensis* may have a sialylated capsule.
  - Capsule genes like *Campylobacter*.
  - Phase variable sialyltransferase like *Neisseria meningitidis* and group B *Streptococcus*

C-terminal phase variation



C-terminal phase variation

- 8 repeats identified at the 3’ end of the CDS
- Also all poly-G tracts
- This location would not mediate translational phase variation.
- The next-generation read data showed changes in the repeat tracts.
- The genome sequence revealed a new type of phase variation that may be applicable to other species.



Next-generation sequencing (Loman *et al*., 2009):
Selected 482 conserved CDSs from the *H. canadensis* genome and related genomes.
Homology between the CDSs provides strong support for a *W. succinogenes / H. canadensis* clade.

Dewhirst, *et al*., J Bacteriol. 2005 Sep;187(17):6106-18.

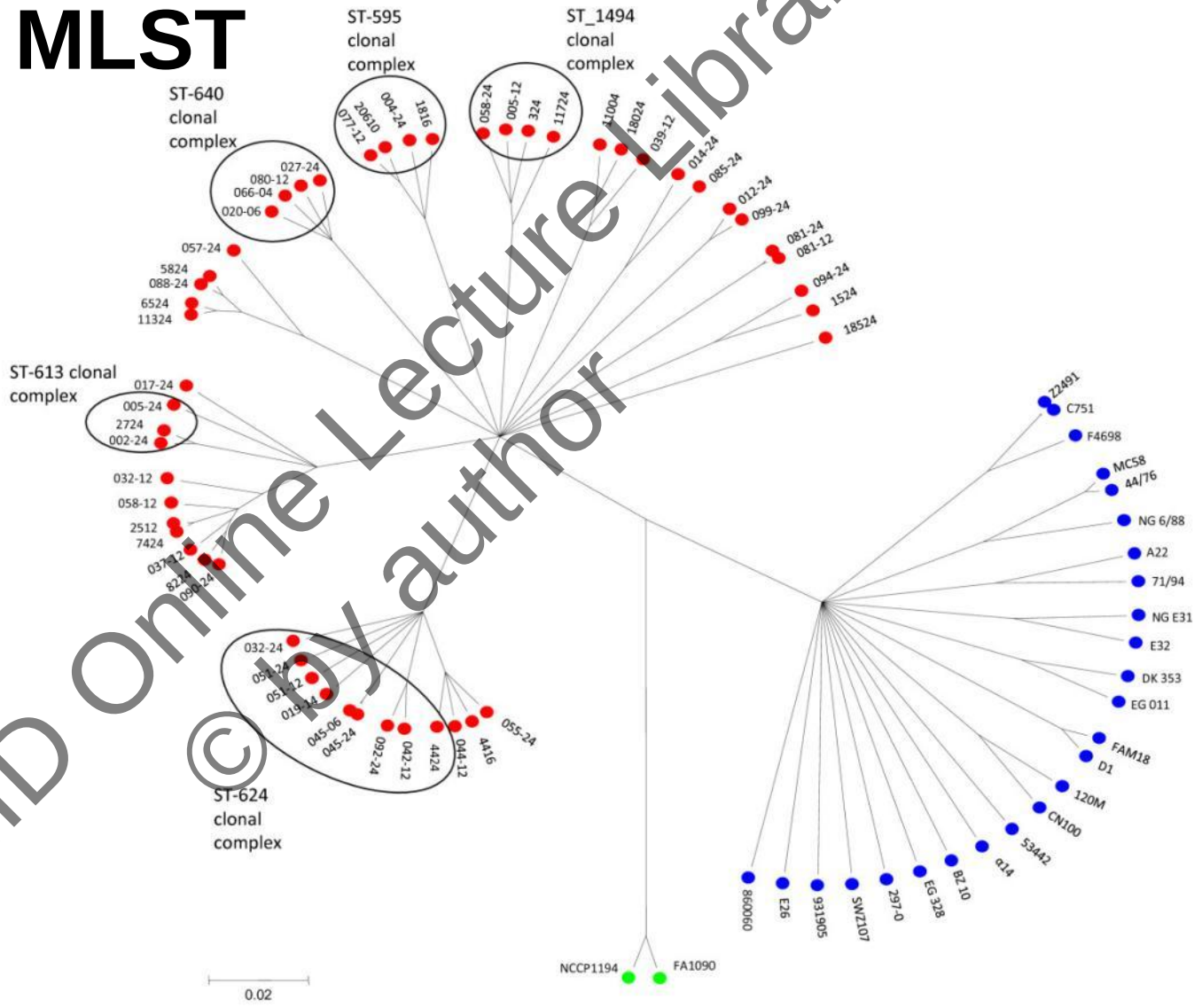
H. canadensis and epidemiology

- Identified as part of larger studies using 16S
  - Robino *et al*., *Schweiz Arch Tierheilkd*, 2010
  - Waldenström *et al*., J Appl Microbiol, 2007
  - Goto *et al*., Curr Microbiol, 2004
- No typing method for strain differentiation
- PFGE – not portable
- Capsule typing – lack information
- Surface protein typing – lack information
- MLST
- Whole genome sequencing

Neisseria epidemiology & typing

- *N. meningitidis* serogroup based on capsule.
  - Can lack discrimination
  - Capsule switching
  - Capsule null strains and phase variable capsules
- Porin serotyping.
- Multilocus enzyme electrophoresis (MLEE).
- Multilocus sequence typing (MLST).

Neisseria MLST



N. lactamica
N. meningitidis
N. gonorrhoeae

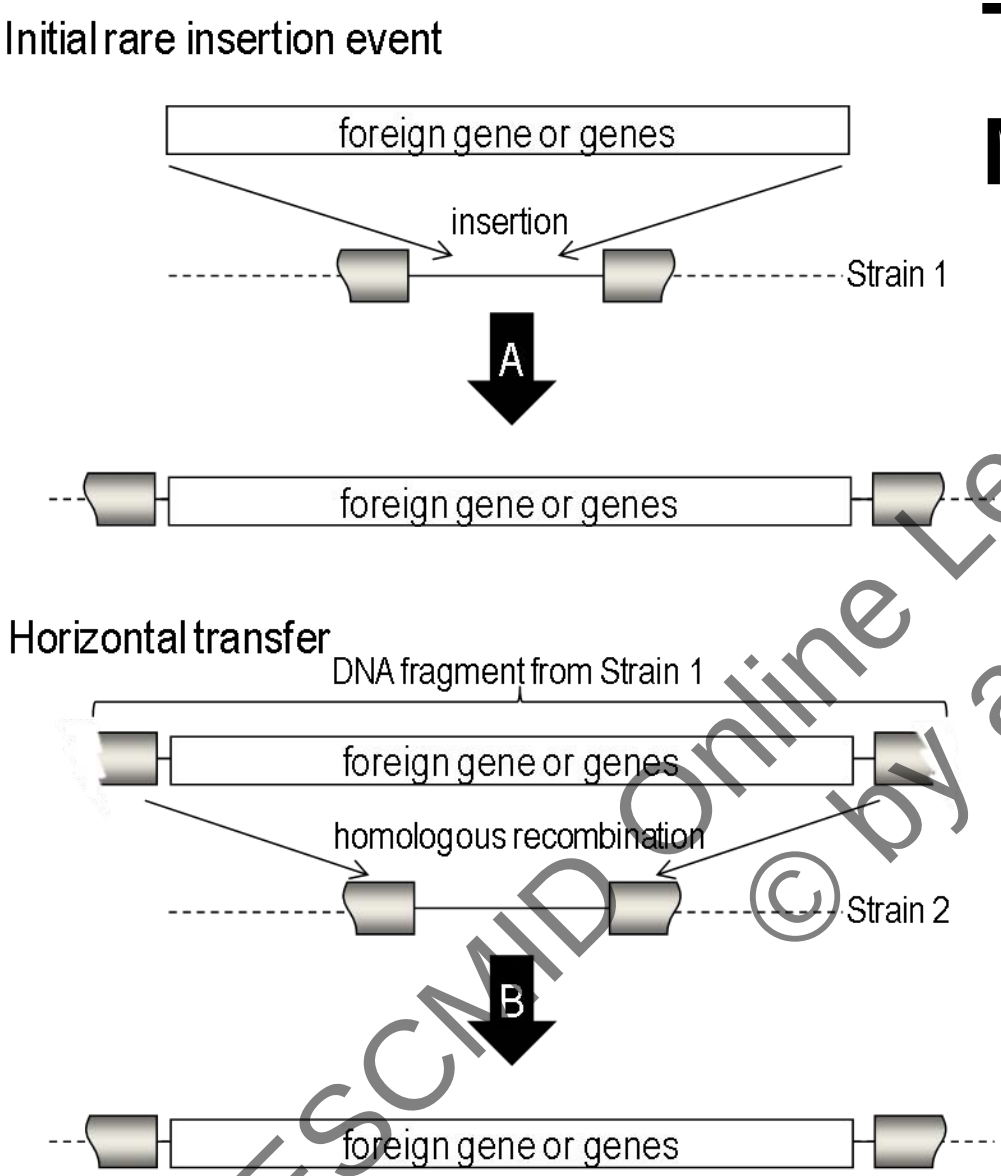
Bennett *et al. BMC Genomics* 2010 **11**:652

Genomic MLST

- Use of the usual MLST markers, extracted from genomic information.
  - Comparison to MLST databases.
  - Assignment to existing STs.
- Use of a subset of the genome.
  - ‘Core genome’.
  - In Bennet *et al*., 2010, 1,190 core CDS were compared between *N. meningitidis, N. lactamica,* and *N. gonorrhoeae*.

Missing some interesting stuff

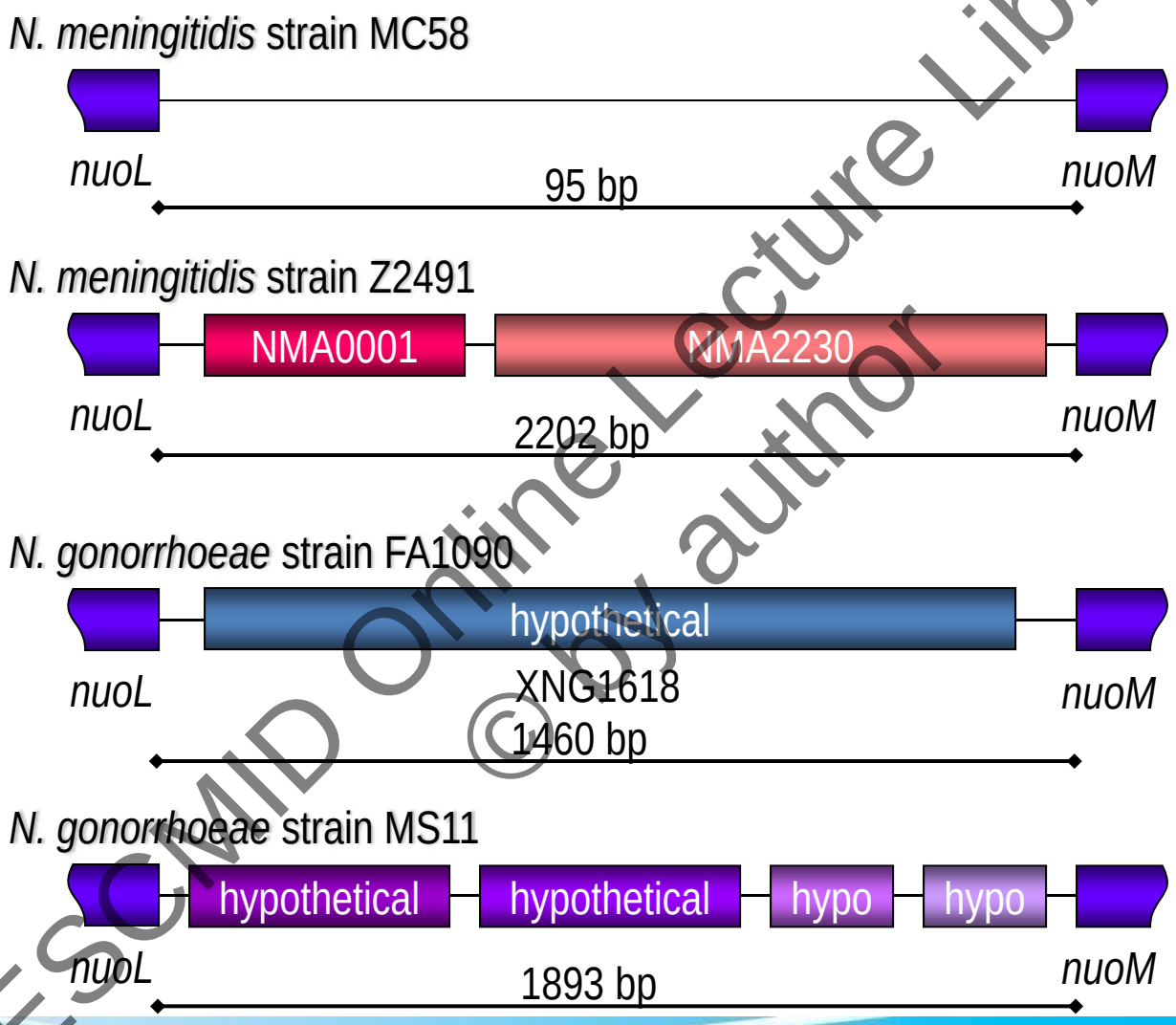
- MLST and analysis of ‘core genome’
  - Evidence for lack of frequent inter-species exchange
  - Addresses only the ‘core genome’ or MLST gene set
- The ‘accessory genome’
  - Quite large in the *Neisseria* spp.
  - 40% of the genome is not core.
  - Shared through inter-species horizontal exchange.
  - Phenotypes arise due to gene combinations and differences in regulation.
  - The ‘accessory genome’ influences the epidemiology.



The Minimal Mobile Element

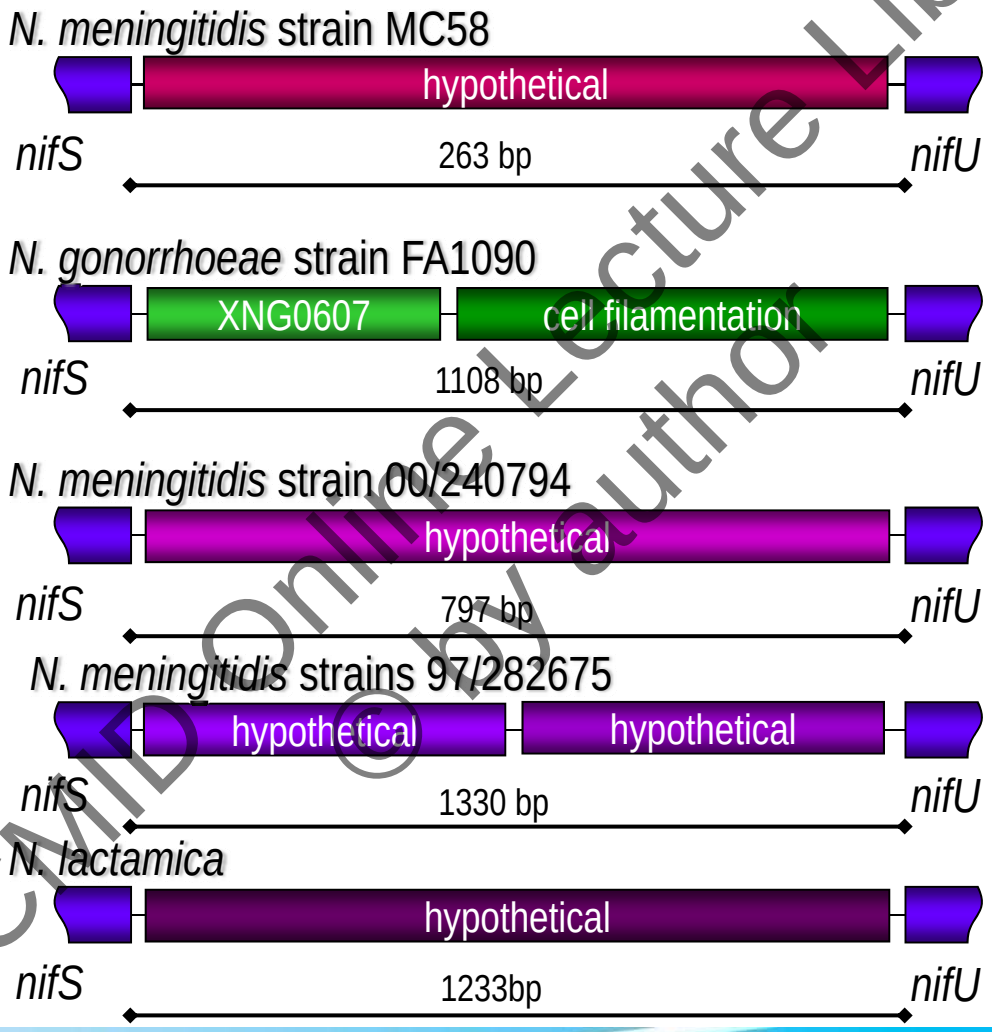
- *Neisseria*
- *Streptococcus*
- *Helicobacter*
- *E. coli*
- *V. cholerae*
- *Haemophilus*
- *Rhodopseudomonas*
- *Pseudomonas*
- *Lactobacillus*

MME*nuoLM*



Snyder *et al*. *BMC Genomics* 2004 **5:** 23

MME*nifSU*



Snyder *et al*. *Mol Biol Evol* 2007 **24:** 2802-2815

Finding MMEs - PCR

- Compare genome sequences to identify sites
- PCR
  - Design primers to the conserved flanking regions.
  - Amplify the region from a collection of strains.
  - Resolve on a gel.
  - Sequence anything of a different size.
  - Misses anything of the same/similar size.

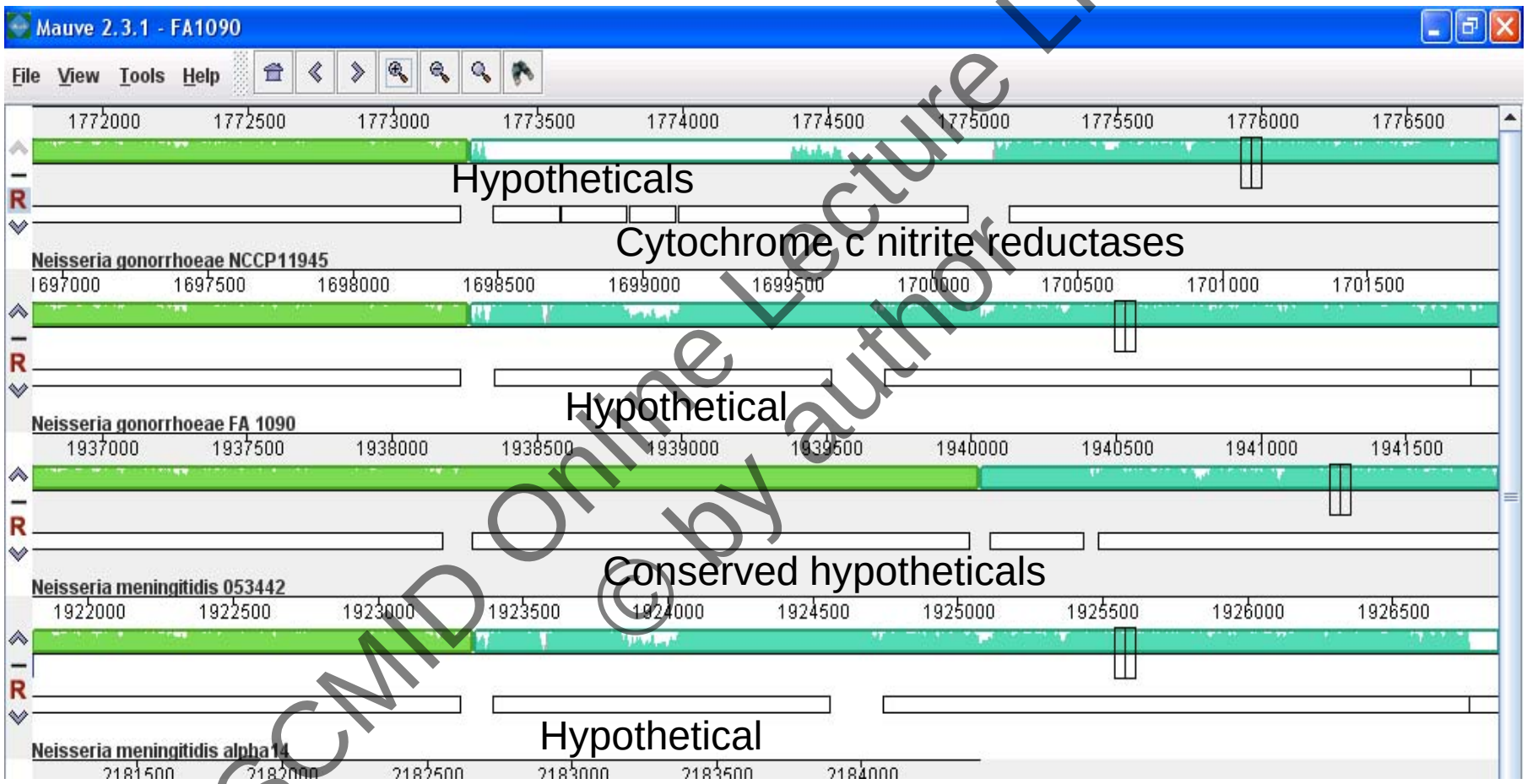
Finding MMES - microarrays

- Use the sequence data from PCR investigations
- Microarrays
  - Design probes to the genes within the MMEs.
  - Hybridize genomic DNA from strain collection.
  - Determine which strains have which genes in the MME sites.
  - Cannot detect if additional genes are in a site.
  - Cannot detect new genes.
    - PCR
    - Sequencing

Finding MMEs - genome sequences

- Where MME investigations started.
- Can now compare more strains.
- Can target analysis to MME sites.
- Can also do analysis to find any new sites.
- Can look at associations of gene combinations at different MME sites.
- Can see the evidence of recombination through sequence alignments.

MME*nuoLM* - Mauve view



nuoM *nuoL*

Another mobile element - GGI

Gonococcal genetic island
57 358 bp

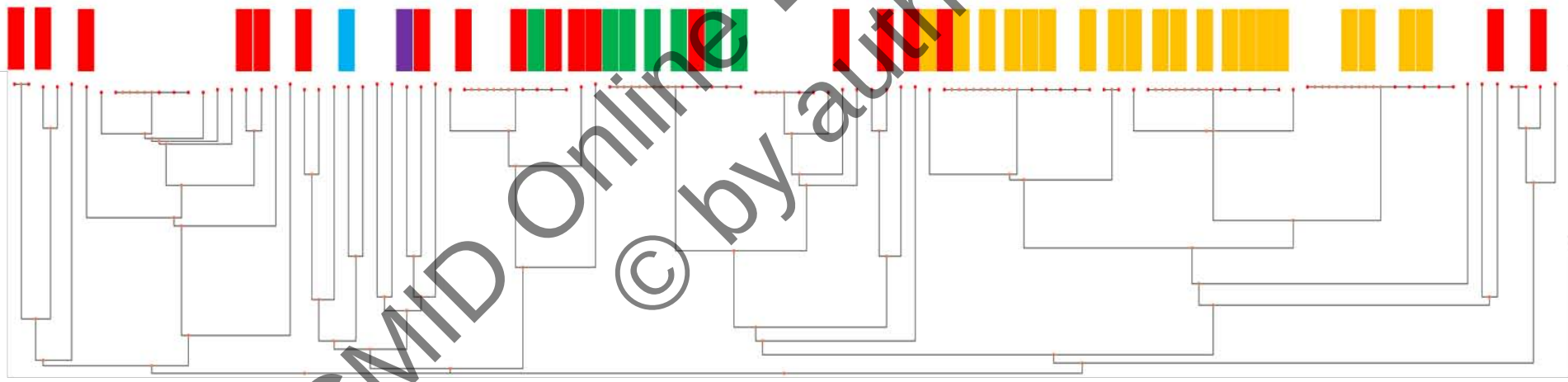
GGI conjugative plasmid
57 358 bp

Microarray – a few strains

- *N. gonorrhoeae* strains
  - FA19
  - MS11
- *N. meningitidis* strains
  - 98/250521
  - A22
  - 97/252675
  - 00/240868
  - 01/241422
  - 01/241471

Microarray – lots more strains

Serogroup A
Serogroup B
Serogroup C
Srgrp W-135
Serogroup Z



What has genomics told us about *Neisseria*?

- 40% of the genome is not ‘core’
- The core dictates most epidemiology methods
- The accessory genome may be key
- Differences in phase variable genes
- Differences in MME gene cassettes
- Differences in larger elements like the GGI
- What makes *N. meningitidis* distinct from *N. gonorrhoeae*?
- How can we vaccinate against all *Neisseria*?

What more can genomics do?

- Dissect apart the transmission of bacteria
  - *Acinetobacter baumannii* Birmingham hospital outbreak (Lewis, *et al*., *J Hosp Infect.* 2010)
  - *Vibrio cholerae* Haitian outbreak (Chin, *et al*., *N Engl J Med.* 2011)
  - *Mycobacterium tuberculosis* Harlingen outbreak (Schürch, *et al*., *J Clin Microbiol*. 2010)
- Rapidly identify an outbreak agent
  - *Bacillus anthracis* in 36 hours (Chen, *et al*., *PLoS One*. 2010)
- Find something new
  - Capsule acquisition by non-pathogenic *Burkholderia* (Sim, *et al*., *Genome Biol*. 2010)

Acknowledgements

- ESCMID and ISC for inviting me
- Mark Pallen (University of Birmingham)
- Kingston University