Dr Mathew Beale
Wellcome Sanger Institute
Disclosures

None
Building a genomic framework for syphilis: Novel approaches to whole genome sequencing

Mathew Beale, PhD
• Treponemas are difficult to sequence
  • Cannot currently be cultured
  • Sequencing must be performed directly on clinical specimens
  • Most of the extract (>95%) will contain host (human) DNA

• New method (target enrichment) now available
  • Enables robust high-throughput sequencing at scale
  • We can use residual samples from Syphilis PCR testing for WGS (some caveats)

Spatiotemporal Global Genomic Diversity Study

• Detailed UK based study

• Ongoing project to sequence samples from around the world:
  • N. America
  • Europe
  • Africa
  • Asia
  • South/Central America
Organisms accumulate genetic variation through natural processes, e.g. drift.

This can be seen when comparing a group of DNA sequences.
Mutations and population structure

- From a population of sequences we can use SNPs to infer relationships by phylogeny.
- We can also infer the state at unsampled ‘ancestral nodes’.
- Since many SNPs are accumulated at a fixed rate of mutation, we can use this to infer timelines.
Limited genomics due to inability to culture

- Inability to culture syphilis has prevented large scale sequencing of syphilis genomes
- Available sequences of variable quality and subject to sampling biases
- Recent studies\textsuperscript{1,2} suggest two syphilis (TPA) lineages diverging around the C18\textsuperscript{th}
- SS14 lineage is often Macrolide resistant – this has been postulated as a cause for SS14 expansion.

\textsuperscript{1}Arora \textit{et al} 2016, Nat Micro 2:16245
\textsuperscript{2}Pinto \textit{et al} 2016, Nat Micro 2:16190
Global Genomic Diversity Study – Preliminary Analysis

Tree suffers from some sampling biases

- 137 global Tpa sequences
- 59 published
- 78 novel sequences at Sanger

- Substantial expansion of SS14 lineage (but sampling bias)
- Both SS14 and Nichols lineages represented in contemporary UK patients

Reference: mapped (SS14 masked), post-Gubbins Maximum Likelihood phylogeny (nodes with black dots indicate UF bootstrap support >95%).

Research in progress – M Beale
Rise of macrolide resistance

Global WGS Samples

Geographical distribution of macrolide resistance in TPA

Emergence of macrolide resistant lineages

Single point mutation at 2058 or 2059 in 23s ribosomal sequence confers intrinsic macrolide resistance
Emergence of macrolide resistant lineages

Macrolide Resistance Allele

- White = sensitive
- Black = genotypic resistance
- Blue = mixed allele
- Grey = no data

<table>
<thead>
<tr>
<th>Continent</th>
<th>N.America</th>
<th>Europe</th>
<th>Asia</th>
</tr>
</thead>
</table>

Research in progress – M Beale
Macrolide resistance has independently evolved and become fixed in lineages on multiple occasions.

Resistant lineages most likely evolved during the late 1980s/1990s.

Argues against recent expansion of SS14 due to macrolide resistance.
Conclusions

• Combining basic metadata with phylogenomic data can yield powerful insights into:
  • Lineage spread
  • Evolution of phenotypic traits (e.g. macrolide resistance)
    • Macrolide evolution has relevance beyond syphilis (e.g. WHO Yaws eradication campaign)

• Syphilis sequencing is now tractable at scale
  • We are conducting large scale sequencing of UK and global populations
  • Currently seeking collaborators
Acknowledgements

Wellcome Sanger Institute
• Nick Thomson
• Maria Fookes

LSHTM
• Michael Marks

Guy’s & St Thomas’
• Achyuta Nori

Mortimer Market Clinic
• Patrick French

University of Washington
• Christina Marra
• Sheila Lukehart

Ongoing collaborations
• Gwenda Hughes
• Helen Fifer
• Michelle Cole

UK sample collection through CNWL sexual health clinics, PCR at UCLH laboratories
Extra stuff: BEAST tree with 95% conf intervals on dates