Bacterial pathogen genome evolution in response to clinical interventions

Stephen Bentley

Wellcome Trust Sanger Institute

JITMM, Bangkok, 1st December 2010

Research areas at Sanger

- ❖ Cancer Genetics and Genomics
- ❖ Human Disease Genomics
- ❖ Human Variation
- ❖ Malaria
- ❖ Pathogen Variation
- ❖ Mouse and Zebrafish Genetics
- ❖ Genomics Informatics

- ➢ Population genomics, evolution and pathogenicity in bacteria
- ➢ Virus genome diversity in infection and pathgenicity
- ➢ Parasite comparative and functional genomics within and between species

Bacterial genome sequencing has changed dramatically over the last 2 years

Sanger capillary →

Solexa (Illumina)

454

AB SOLiD

New applications

1) Surveillance/epidemiology by sequencing 1,000’s of isolates
2) Aetiology through deep sequencing of clinical samples
3) Expression analysis of isolates
4) Monitoring of colonisation/transmission dynamics by metagenomic sequencing

Massive increase in sequence output

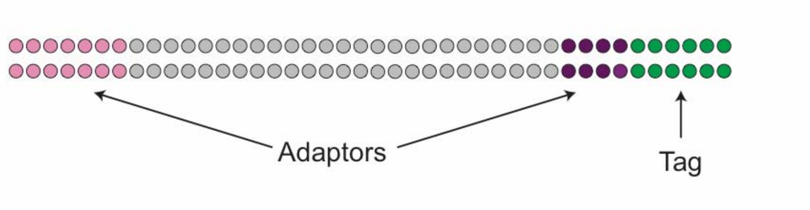
	AB3730	**454**	**Illumina**
Raw data per run (Mbp)	0.25	100	30,000
Read length (bp)	700	250-500	36/54/70/108

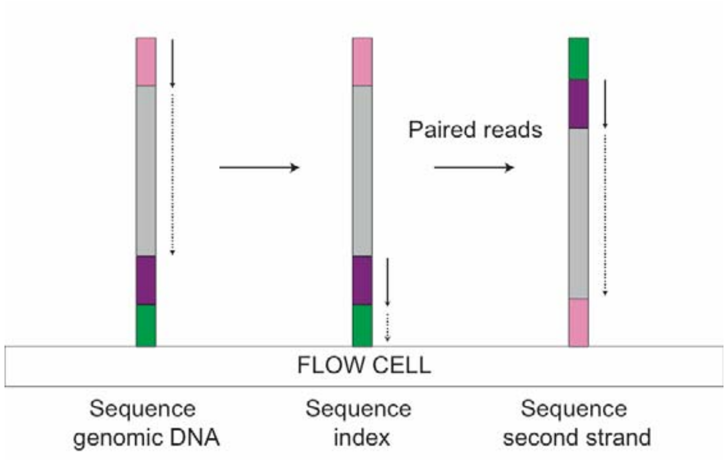
Haploid human genome = 3,000 Mbp → 10 times coverage

S. aureus genome = 2.8 Mbp → ~11,000 times coverage of one

38 Illumina GAIIx machines at the Sanger Institute

Taming the beast: multiplexing on Solexa





```
@IL22_1561:1:1:1771:914
TTTTTTCAATACAATTTCATCTCTAATTTCAATTCAGCCTAA
+
>>>>7>>>>>>>>>>>>>>>>>>>:>5>>>96>+>>7;>>>>
@IL22_1561:1:1:1222:1105
ATGATATCATCTACACGGTTTAAAAATTCTGGACGGCTGATC
+
>>;>>>>>>>;>>>>>55>>>>>>>>>>>>+2>>8>>>>>57
@IL22_1561:1:1:1438:709
ACACCGATGACAATAATTGTTCCAATATCTGTAACAAAGCTA
+
<<<<<2<<2<<<<<<<<<)'<<<<<<<<<<+<;;7<)<<2<<
@IL22_1561:1:1:1671:1462
TGCAAGAACATTAGACAACGTATCTTCAATCGTTTATACAAG
+
>>7>:>:>>7>>7/:>>>7/.7>>>>>57>>0>>>>>>7>>>
@IL22_1561:1:1:1168:891
AATGGAAATCAAACTATAACTTCAACACTAAATGAGAAGCTA
+
>>>:7>>>>>>>>>>>>>>>>>>>>>>>>>>>>0>>>>>>>>
@IL22_1561:1:1:1097:1281
TAAAAGTATAACTTTCCATCACCAAAGTCAGAATTTACATCG
+
>>>>>>>>>>>>>>>>>>>>>>>>8>8>>>>>>>>>>>>>>>
```

TAG	Strain
CGTGAT	HGSA942
ACATCG	3HK
GCCTAA	CHI59
TGGTCA	CHI61
CACTGT	BK2491
ATTGGC	TUR9
GATCTG	HU25
TCAAGT	HU109
CTGATC	HSJ216
AAGCTA	FFP103
GTAGCC	ICP5062
TACAAG	GRE4

interaction with epithelial cells. We next asked if epithelial cell *mmp9* induction was dependent on MyD88 and TNF (tumor necrosis factor) signaling, as each can enhance mycobacterial induction of *mmp9* in cultured cells under certain conditions (*19*, *20*). ESAT-6 induced *mmp9* in *myd88* and *tnf-receptor 1* (*tr1*) morphants (Fig. 4D), suggesting a previously unknown pathway for this epithelial cell–specific interaction. Moreover, TNF-independent induction of *mmp9* is consistent with the finding that TNF does not mediate granuloma formation either in the presence or absence of bacterial RD1 (*13*).

Thus, ESAT-6 functions in virulence by promoting granuloma formation via interaction with epithelial cells, previously regarded as bystanders in the pathogenesis of tuberculosis (fig. S7). The cooption of epithelial cells may offer mycobacteria a means of amplifying MMP9 secretion in the vicinity of a single infected macrophage to establish the granuloma niche. In addition, the differential induction of inflammatory programs in macrophages and epithelial cells may generate a hospitable growth niche in macrophages while harnessing epithelial cells to facilitate the chemotaxis of additional macrophages for niche expansion (*6*) (fig. S7). Our work provides a mechanistic explanation for the implication of MMP9 in human susceptibility to tuberculosis (*9*, *11*, *12*) and suggests targeted inhibition of its expression as a host-directed antituberculous therapy. Because increased MMP9 is detrimental in both tuberculosis and a variety of noninfectious inflammatory conditions (*7*), interception of this pathway may have broad utility in treating a

48, 217 (2004).
23. D. R. Sherman *et al.*, *J. Infect. Dis.* **190**, 123 (2004).
24. I. C. Koo *et al.*, *Cell. Microbiol.* **10**, 1866 (2008).

Published online 10 December 2009;
10.1126/science.1179663
Include this information when citing this paper.

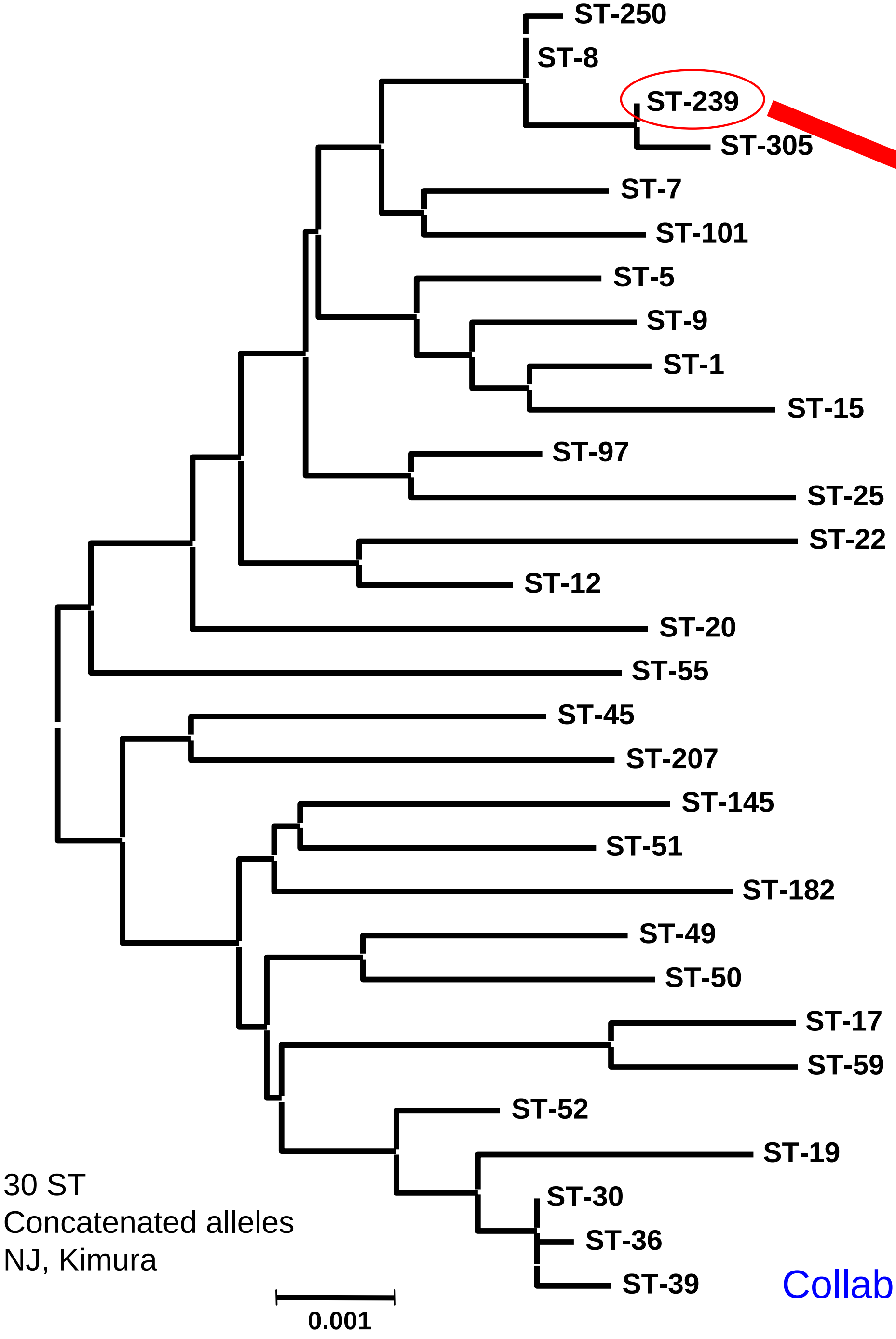
Evolution of MRSA During Hospital Transmission and Intercontinental Spread

Simon R. Harris,[1]* Edward J. Feil,[2]* Matthew T. G. Holden,[1] Michael A. Quail,[1] Emma K. Nickerson,[3,4] Narisara Chantratita,[3] Susana Gardete,[5,6] Ana Tavares,[5] Nick Day,[3,7] Jodi A. Lindsay,[8] Jonathan D. Edgeworth,[9,10] Hermínia de Lencastre,[5,6] Julian Parkhill,[1] Sharon J. Peacock,[3,4] Stephen D. Bentley[1]†

Current methods for differentiating isolates of predominant lineages of pathogenic bacteria often do not provide sufficient resolution to define precise relationships. Here, we describe a high-throughput genomics approach that provides a high-resolution view of the epidemiology and microevolution of a dominant strain of methicillin-resistant *Staphylococcus aureus* (MRSA). This approach reveals the global geographic structure within the lineage, its intercontinental transmission through four decades, and the potential to trace person-to-person transmission within a hospital environment. The ability to interrogate and resolve bacterial populations is applicable to a range of infectious diseases, as well as microbial ecology.

The development of molecular typing techniques has been instrumental in studying the population structure and evolution of bacterial pathogens. Sequence-based approaches, such as multilocus sequence typing (MLST) (*1*), have resulted in large searchable databases of the most clinically important species. However, MLST defines variation within a very small sample of the genome and cannot distinguish between closely related isolates. Full-genome sequencing provides a complete inventory of microevolutionary changes, but this approach is impractical for large population samples. The use of next-generation sequencing technologies, such as Illumina Genome Analyzer, bridges this gap by mapping genome-wide single-nucleotide poly-

Prototype multiplexing project: MRSA ST239



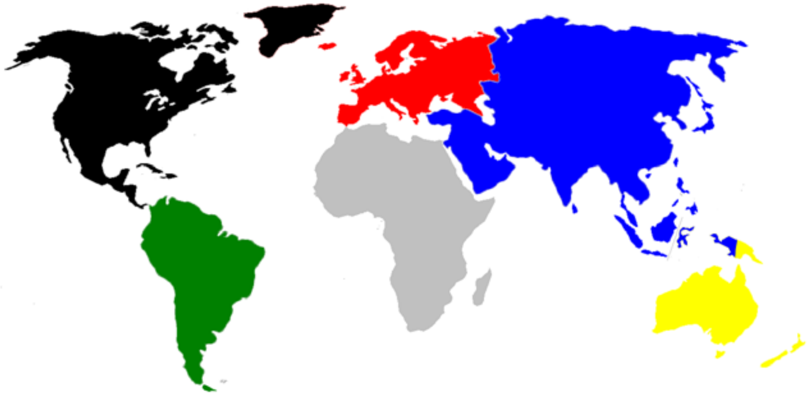
30 ST
Concatenated alleles
NJ, Kimura

62 isolate collection

(Herminia de Lencastre, Sharon Peacock)

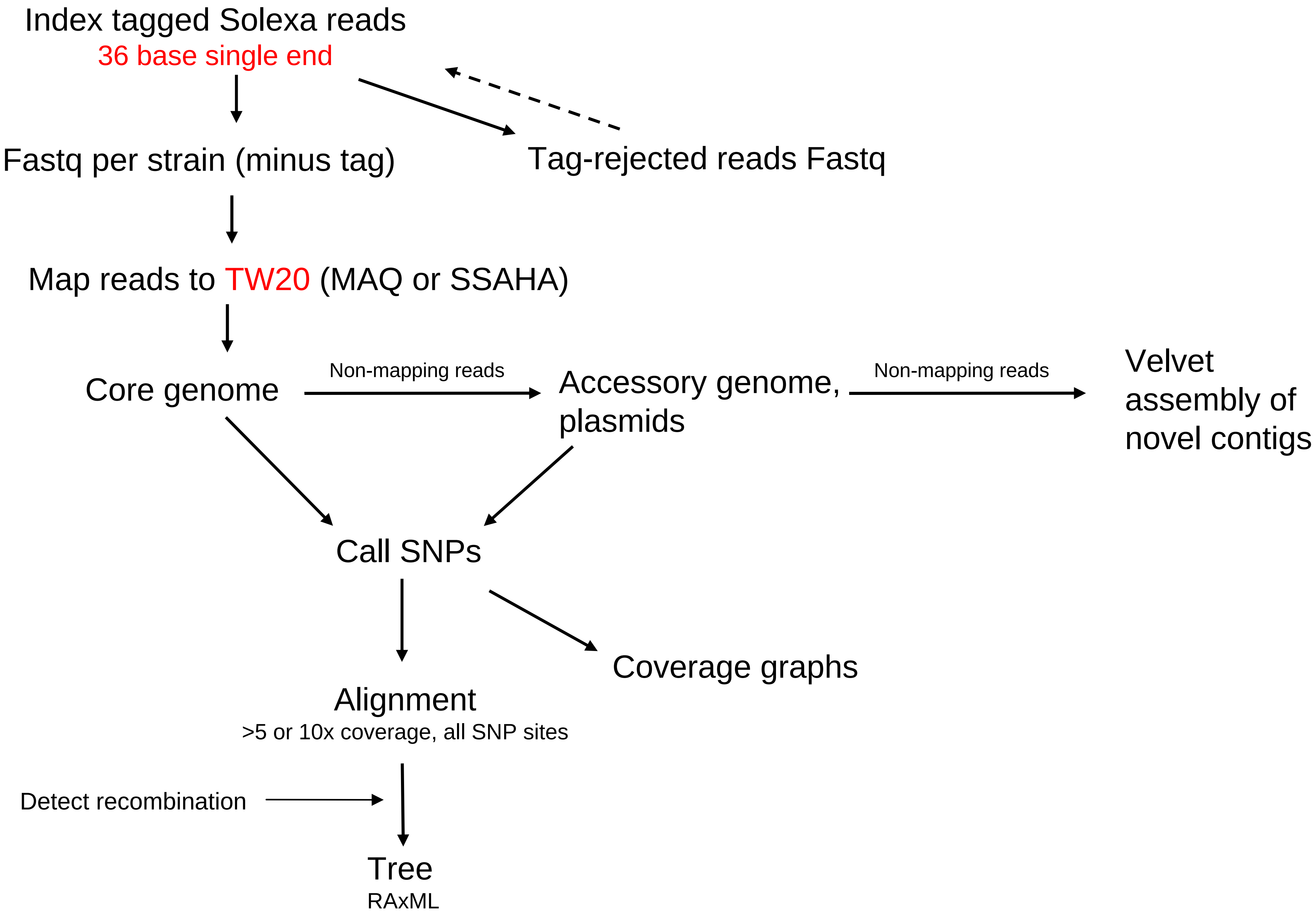
Geographic: 15 countries
5 continents

Temporal: 1982-2007



Collaboration with Hermínia de Lencastre, Ed Feil, Sharon Peacock

Bioinformatics Pipeline



Index tagged Solexa reads
36 base single end

Fastq per strain (minus tag)

Tag-rejected reads Fastq

Map reads to TW20 (MAQ or SSAHA)

Core genome

Non-mapping reads

Accessory genome, plasmids

Non-mapping reads

Velvet assembly of novel contigs

Call SNPs

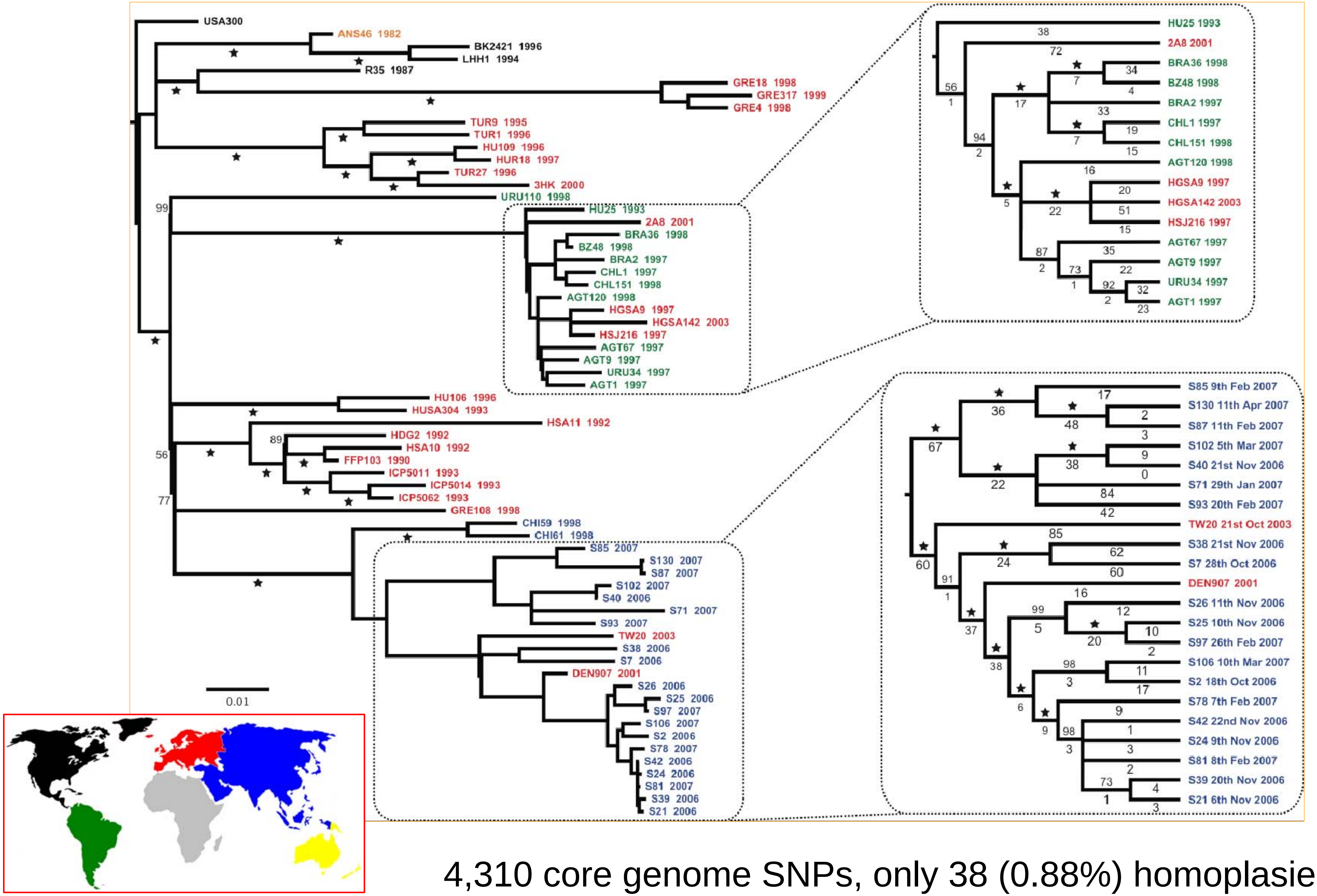
Coverage graphs

Alignment
>5 or 10x coverage, all SNP sites

Detect recombination

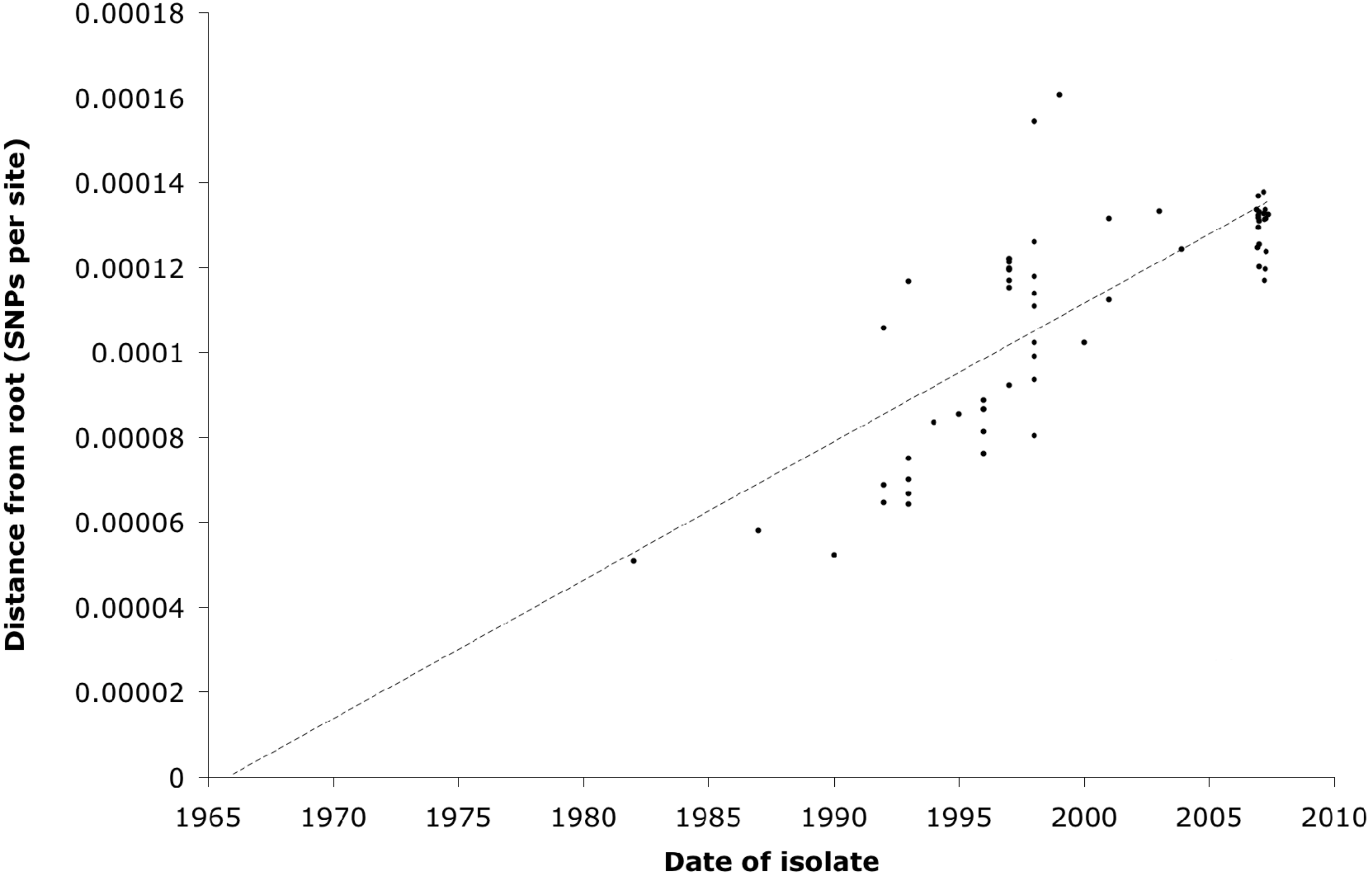
Tree
RAxML

Geographic structure within ST239

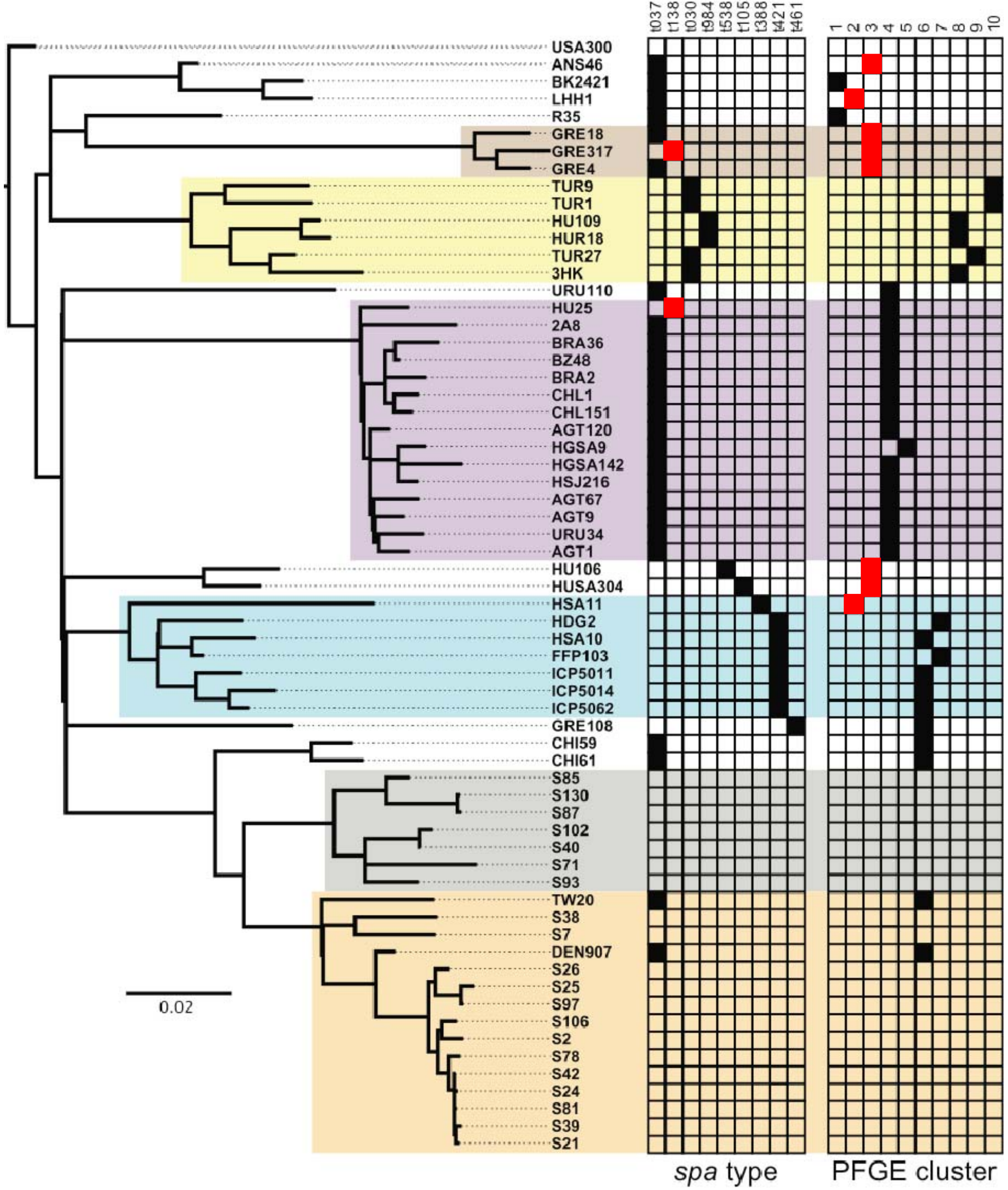


4,310 core genome SNPs, only 38 (0.88%) homoplasies

Linear regression plot of root-to-tip distances of the ST239 phylogenetic tree against the date of sampling of the isolates.



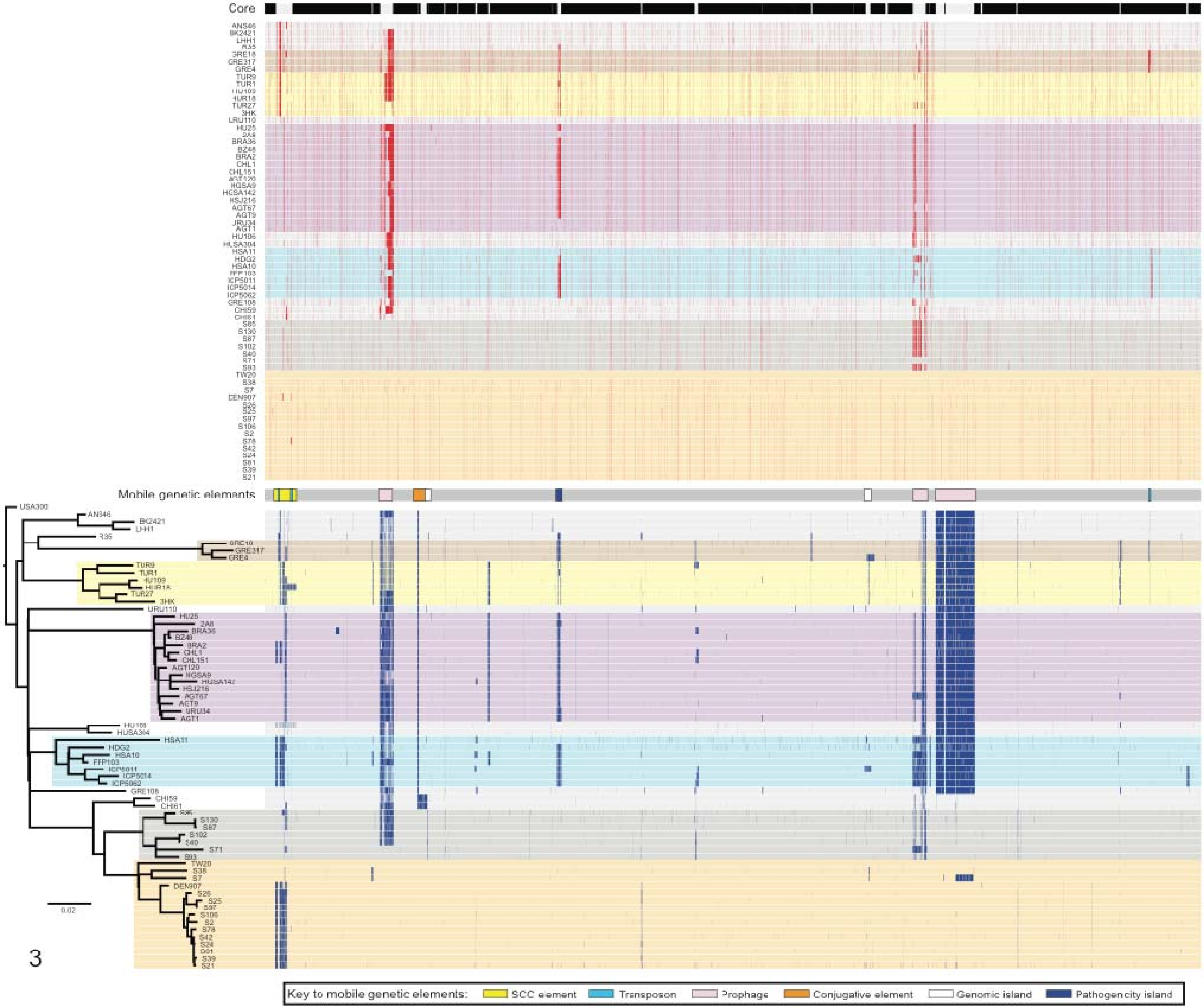
Comparison with other typing methods



Clinical practice is shaping the genetic makeup of ST239

SNP position	Locus affected	No. of ÒbranchesÓ	SNP	Substitution	Antibiotic
7254	DNA gyrase subunit A GyrA	2	T => G	Ser84Ala	
7255	DNA gyrase subunit A GyrA	8	C => T	Ser84Leu*	quinolone
7266	DNA gyrase subunit A GyrA	4	G => A	Lys88Glu	quinolone
133864	immunoglobulin G binding protein A precursor	2	G => A	Synonymous	-
134787	92 bp upstream of immunoglobulin G binding protein A	2	G => T	Intergenic	-
278498	129 bp upstream of putative acetyl-CoA transferase	2	T => C	Intergenic	-
436474	34 bp upstream of putative dioxygenase	2	C => T	Intergenic	-
594883	tetrapyrrole (corrin/porphyrin) methylase family protein	2	C => T	Pro49Ser	-
657696	DNA-directed RNA polymerase beta chain protein RpoB	4	C => A	Asp471Glu	rifampin
657724	DNA-directed RNA polymerase beta chain protein RpoB	5	C => A	His481Asn	rifampin
657869	DNA-directed RNA polymerase beta chain protein RpoB	2	C => T	Ser529Leu	rifampin
666536	translation elongation factor G	2	T => A	Leu461Lys	fusidic acid
666537	translation elongation factor G	2	T => A	Leu461Lys*	fusidic acid
681826	48 bp upstream of serine-aspartate repeat-protein C	2	C => A	Intergenic	-
862898	putative membrane protein	2	A => C	Ser160Ala	-
1130135	63 bp upstream of FolD bifunctional protein	2	G => T	Intergenic	-
1138698	phosphoribosylglycinamide formyltransferase PurN	3	T => A	Leu174Met	-
1172434	50 bp upstream of probable manganese transport protein	3	T => G	Intergenic	-
1172436	52 bp upstream of probable manganese transport protein	2	T => C	Intergenic	-
1172444	60 bp upstream of probable manganese transport protein	2	C => G	Intergenic	-
1206826	ribonuclease HIII	2	C => T	Glu199Lys	-
1261219	isoleucyl-tRNA synthetase	2	G => T	Val588Phe	mupirocin
1448063	topoisomerase IV subunit A GrlA	4	T => C	Ser80Phe	quinolone
1524413	dihydrofolate reductase type I DfrB	2	T => C	His150Arg	trimethoprim
1524566	dihydrofolate reductase type I DfrB	4	A => T	Phe99Tyr	trimethoprim
1524789	dihydrofolate reductase type I DfrB	2	G => A	Synonymous	-
1525796	thymidylate synthase	3	G => A	Synonymous	-
1525817	thymidylate synthase	3	G => A	Synonymous	-
1525832	thymidylate synthase	3	G => A	Synonymous	-
1640281	glyoxalase/bleomycin resistance protein	2	T => G	Synonymous	-
1689862	putative transcriptional repressor CcpN	2	C => T	Synonymous	-
1755814	probable cell wall amidase LytH	2	A => G	Pro63Ser	-
1921379	bifunctional riboflavin biosynthesis protein RibD	2	G => T	Asn208Lys	-
2334865	protein SprT-like	2	G => A	Ser43Phe	-
2753531	458 bp upstream of conserved hypothetical protein	2	A => T	Intergenic	-
2828688	200 bp downstream of putative exported protein	3	T => C	Intergenic	-
2828714	226 bp downstream of putative exported protein	3	G => T	Intergenic	-
2859765	39 bp upstream of O-acetyltransferase OatA	2	C => T	Intergenic	-

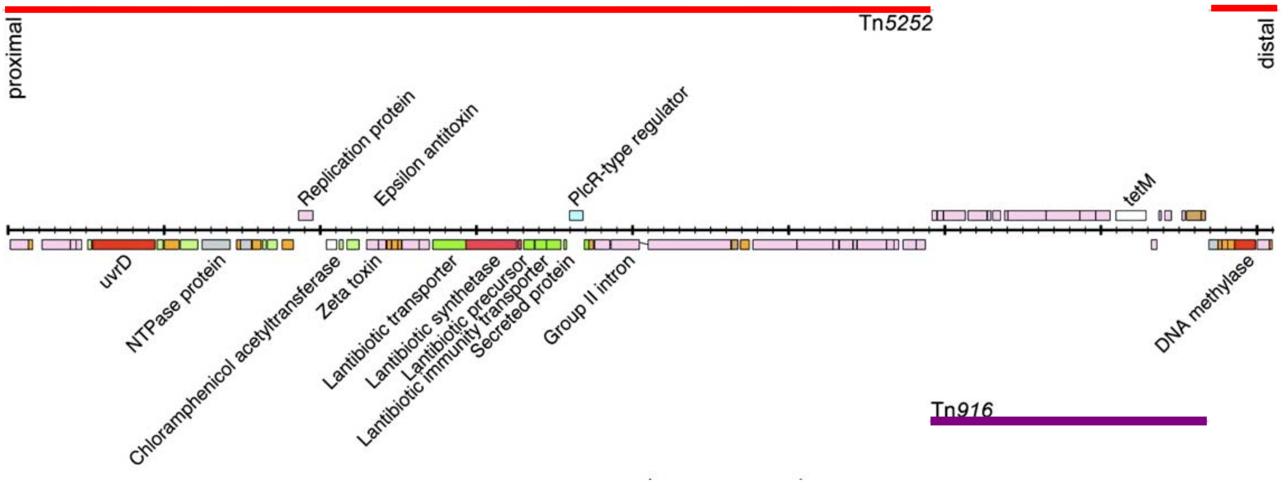
Genetic variation within the lineage



Streptococcus pneumoniae
PMEN1 *aka* Spain^{23F}-1 (ST81)

- Highly recombinogenic
- Penicillin, chloramphenicol, tetracycline, fluoroquinalone, macrolides
- Among the first recognized pandemic clones
- ~40% of penicillin-resistant pneumococcal infections in US in late 1990s
- Low odds ratio for disease (0.4)
- Capsule switches to 14, 19A, 19F
- Genome of ATCC 700669, Barcelona 1984 (Croucher *et al*, *J. Bact*. 2009)

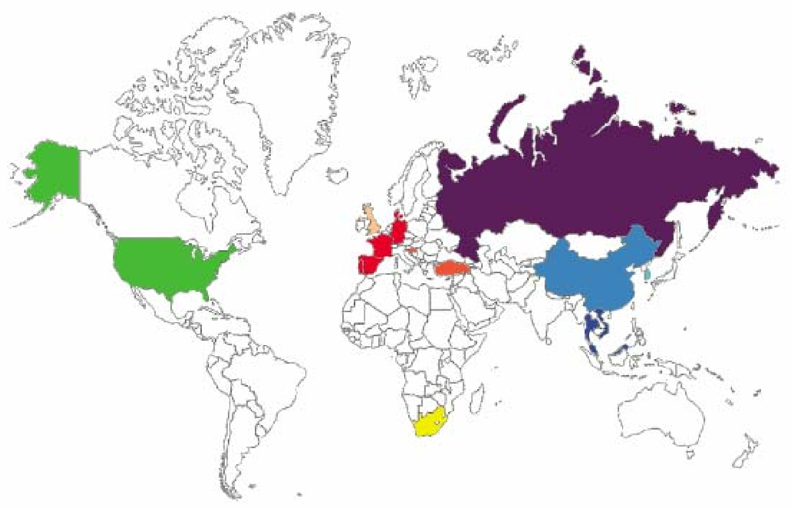
→ roles for integrative and conjugative elements (ICE) in the evolution of pneumococci and emergence of the clone.



PMEN1 240-isolate collection (1984-2007)

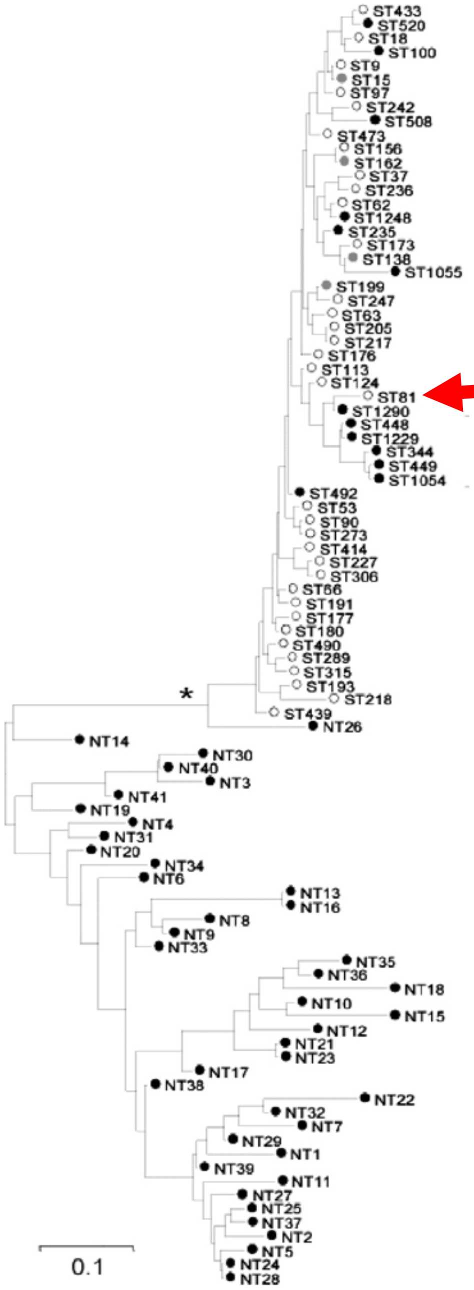
West Europe		
	Denmark	**8**
	France	**9**
	Germany	**23**
	Portugal	**1**
	Spain	**17**
	UK	**26**
East Europe and Asia		
	Croatia	**2**
	Russia	**8**
	Turkey	**11**
America		
	USA	**20**
Africa		
	South Africa	**37**
South East Asia		
	China	**10**
	Korea	**17**
	Malaysia	**3**
	Singapore	**4**
	Thailand	**4**
	Vietnam	**11**

Keith Klugman	Emory University, USA
Anne von Gottburg	NICD, South Africa
Lesley McGee	CDC, USA
Kwan Soo Ko	ARFID, South Korea
Steve Baker	OUCRU, Vietnam
Lotte Lambertsen	SSI, Denmark
Mark van der Linden	NRCS, Germany
Tim Mitchell	Glasgow University, UK
Bruno Pichon	HPA, UK
Bill Hanage	Imperial College, UK



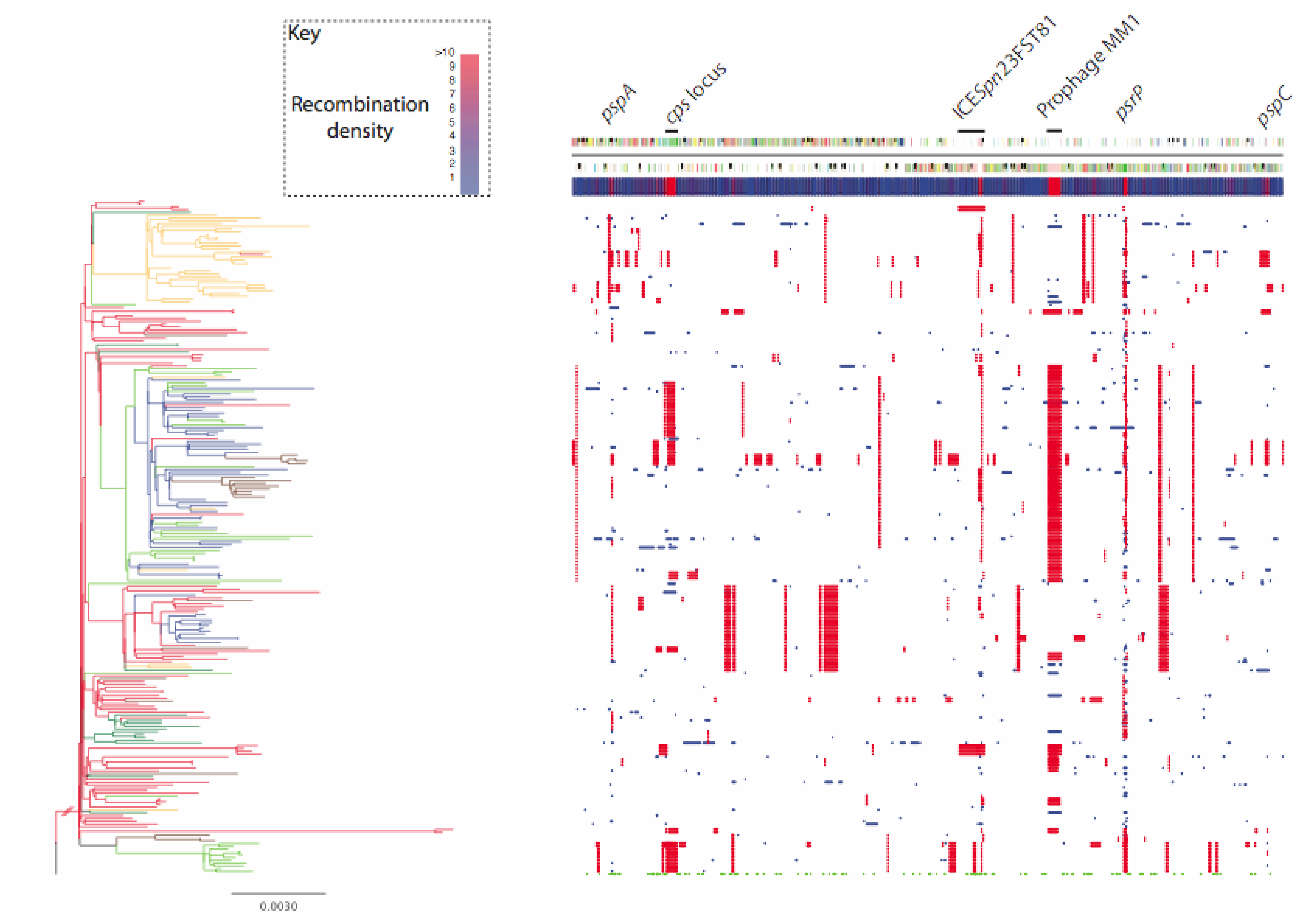
Recently accepted for publication in *Science*

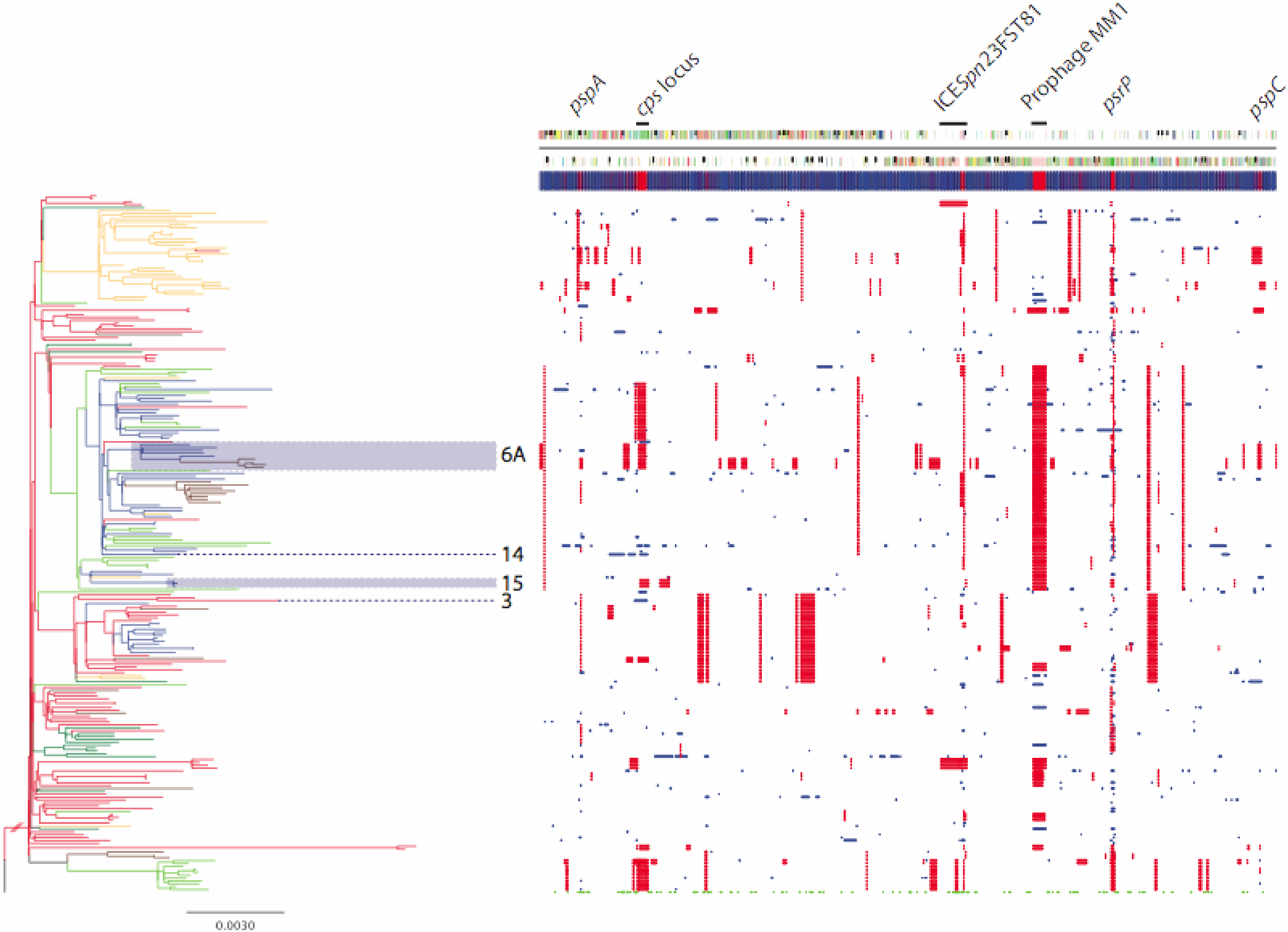
MLST definition of *S. pneumoniae*

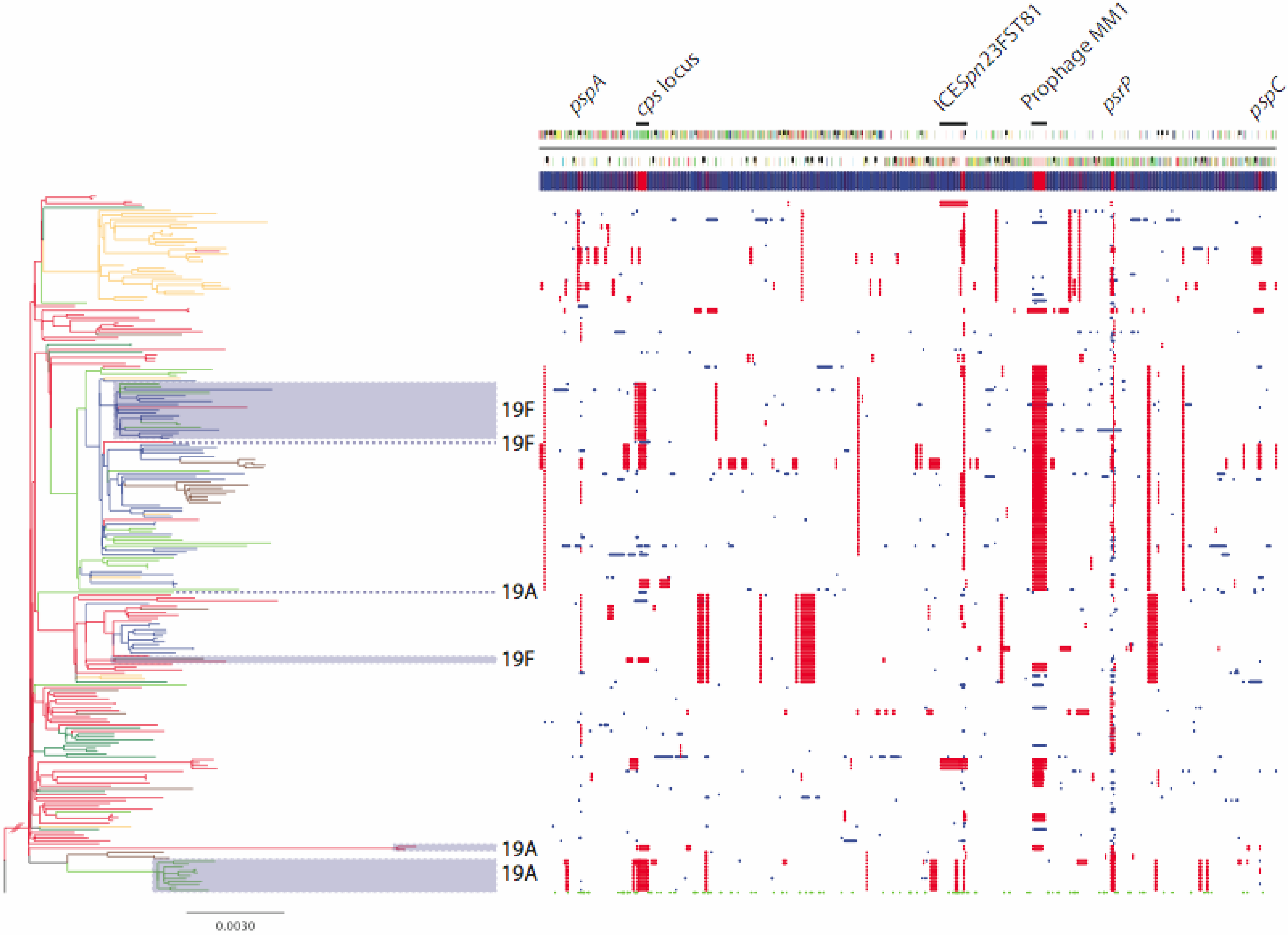


ST81

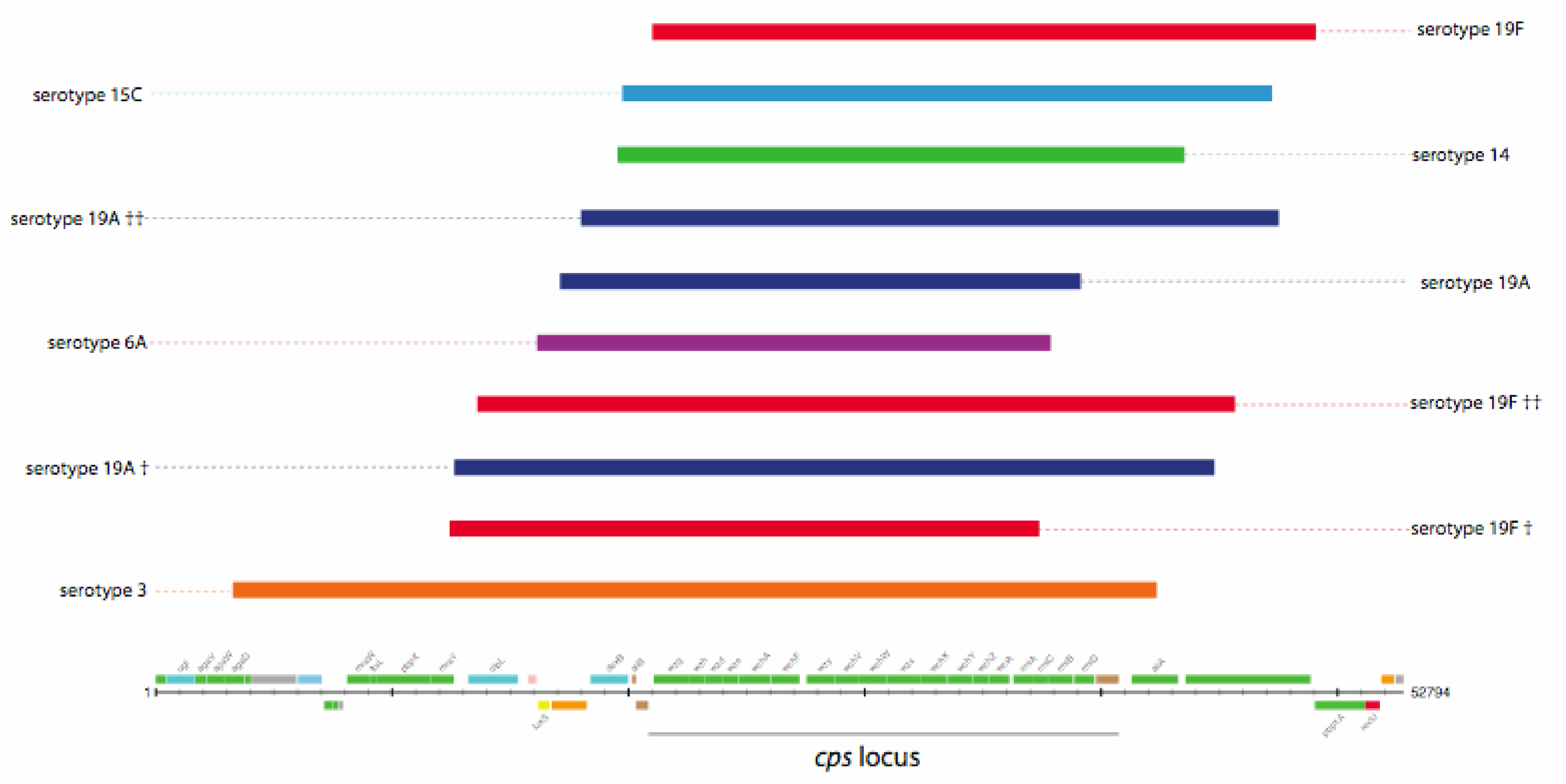
Total 17416 SNPs detected, 2662 SNPs “core” SNPs used to build tree



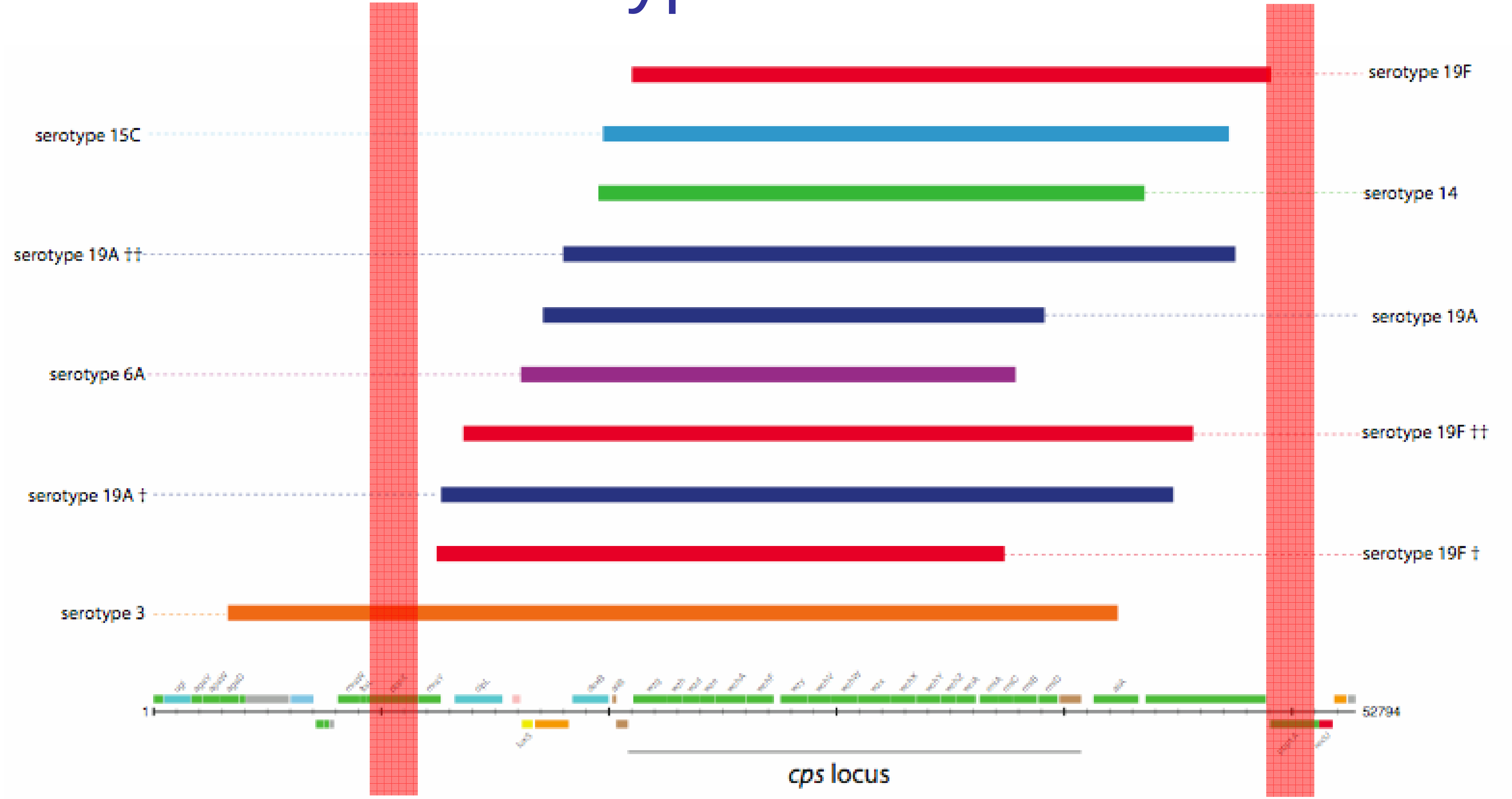

pspA
cps locus
ICESpn23FST81
Prophage MM1
psrP
pspC
6A
14
15
3
0.0030


pspA
cps locus
ICESpn23FST81
Prophage MM1
psrP
pspC
19F
19F
19A
19F
19A
19A
0.0030

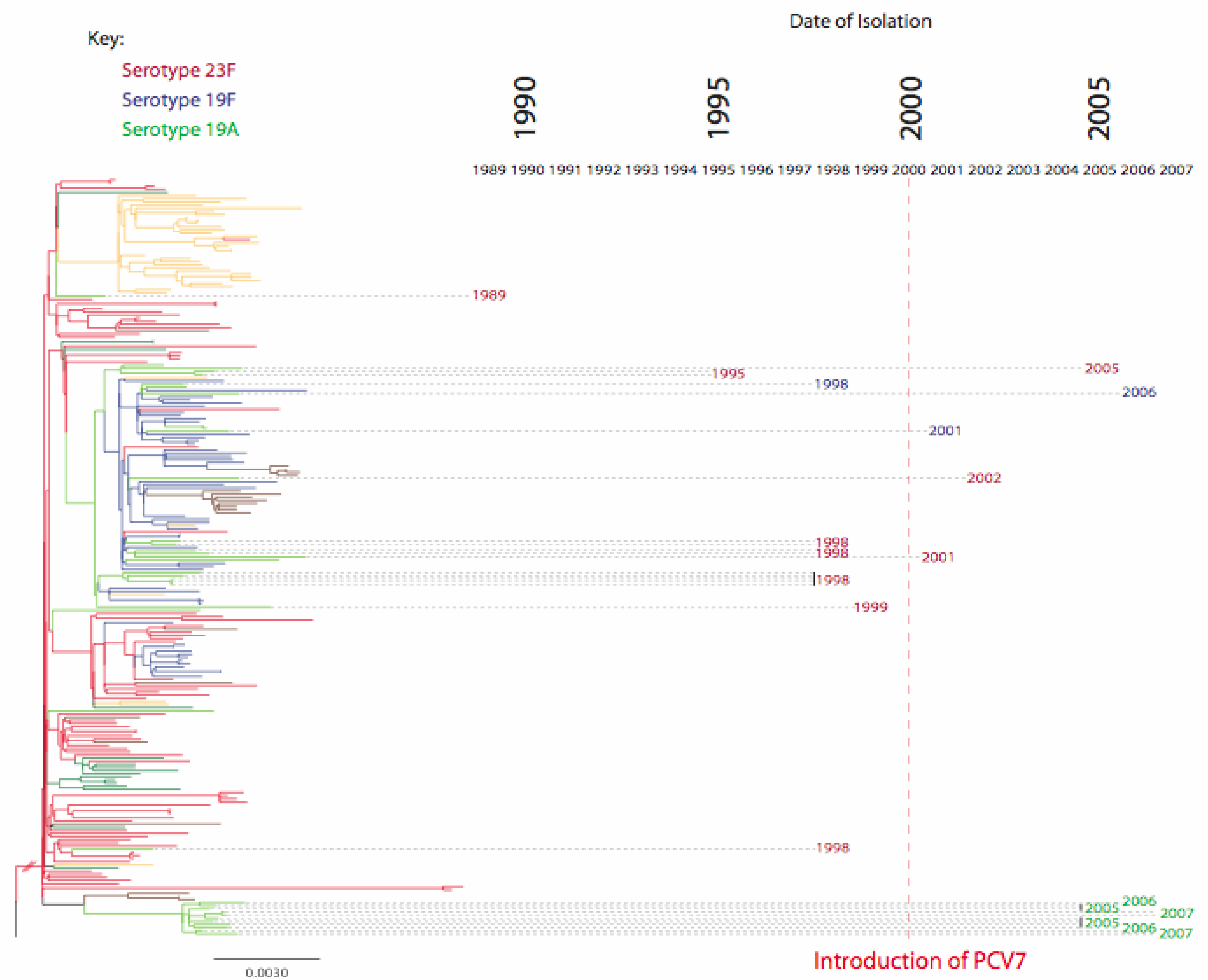
Serotype Switches



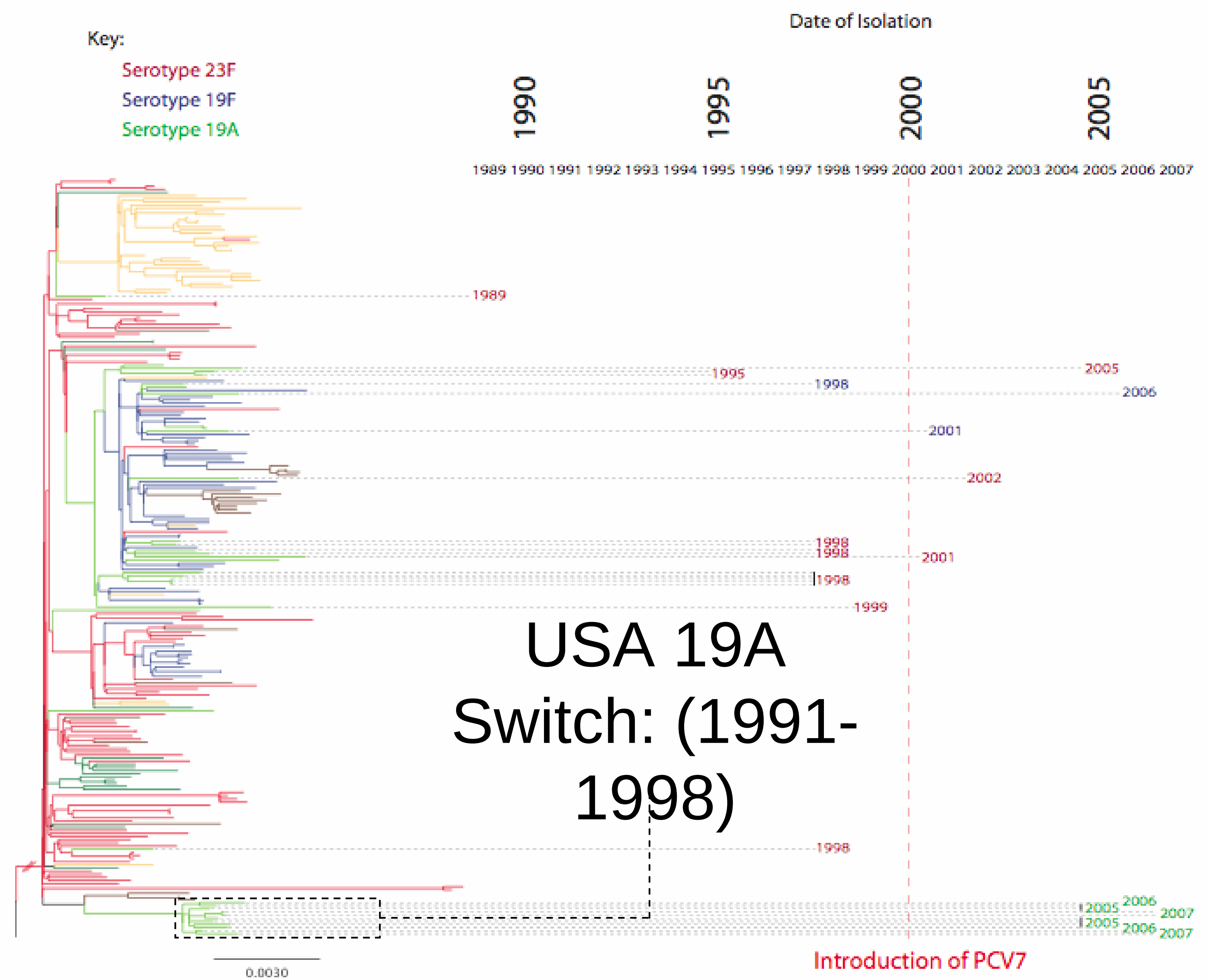
Serotype Switches



PMEN1 in the USA



PMEN1 in the USA



USA 19A Switch: (1991-1998)

Tn*916* comparisons

Spain,
1984

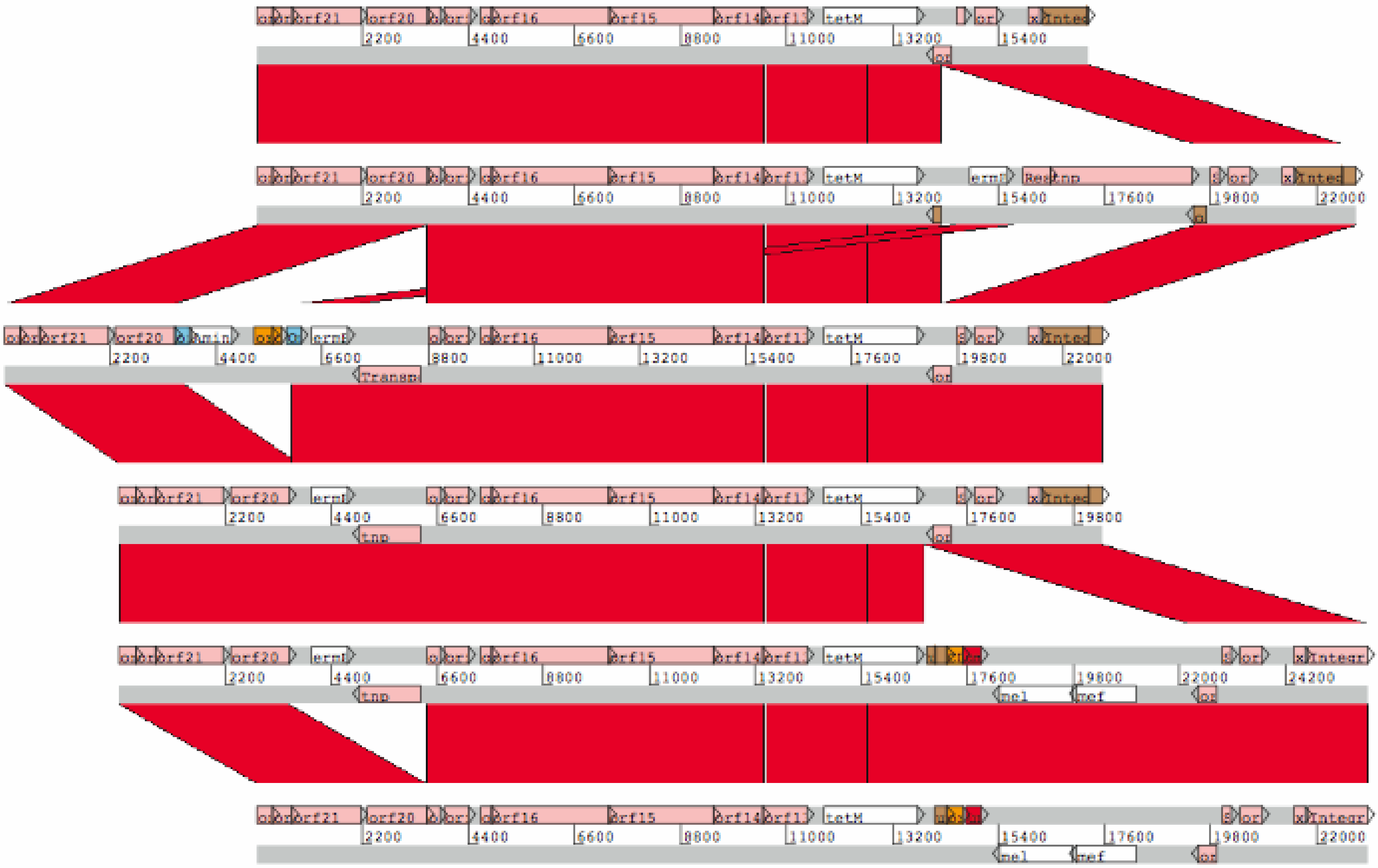
Spain,
2002

France,
2002

UK,
2003

Russia,
2004

Malaysia,
1999



Multiple independent acquisitions of macrolide resistance

ATCC 700669

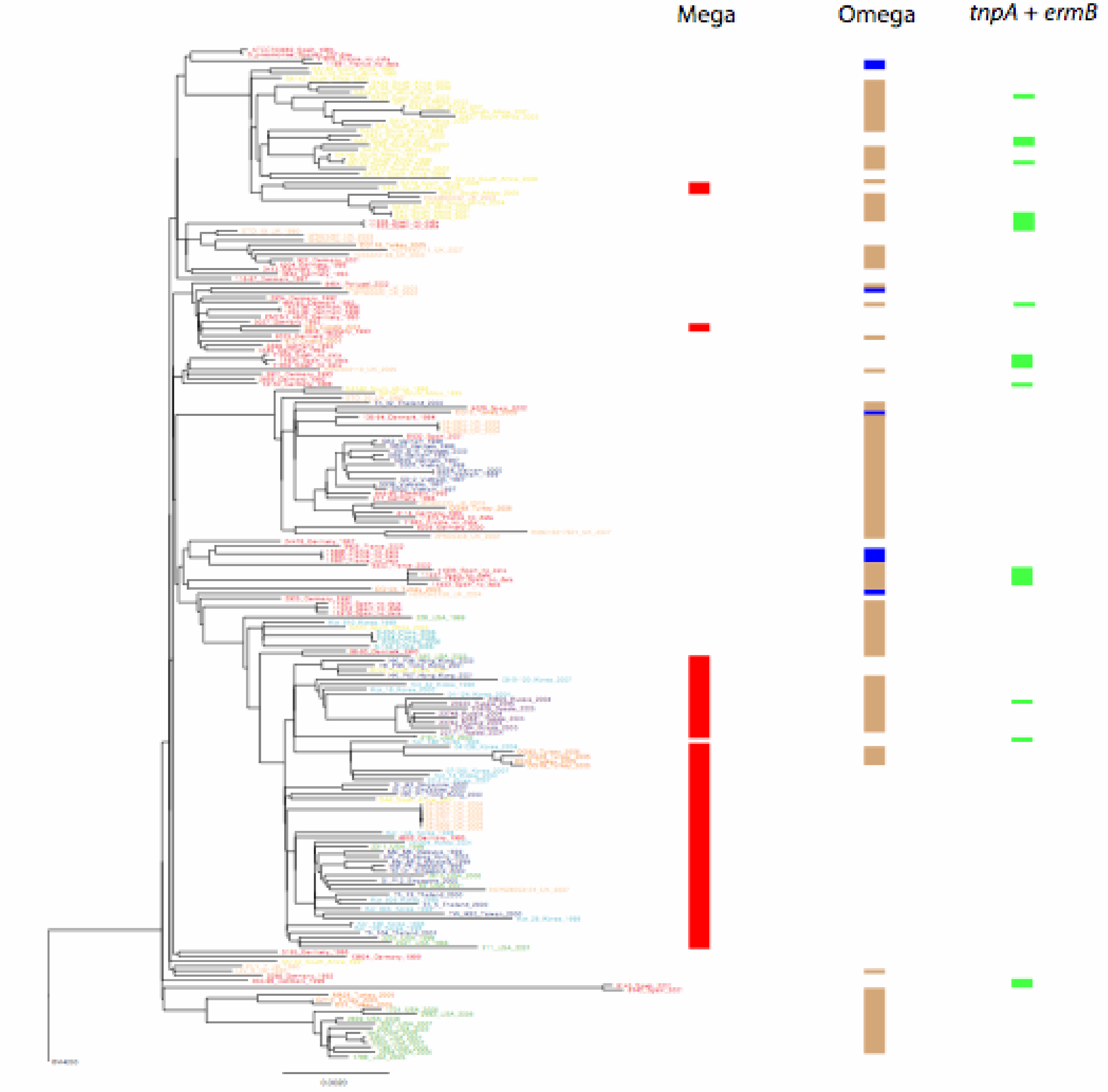
Tn917

Omega element

Omega fragment

Mega element

Multiple independent acquisitions of macrolide resistance



Acknowledgements

Sanger library prep and sequencing teams

MRSA ST239

Simon Harris
Ed Feil
Matt Holden
Sharon Peacock
Herminia de Lencastre
Julian Parkhill

Pneumo

Nick Croucher
Keith Klugman
Anne von Gottburg
Lesley McGee
Kwan Soo Ko
Steve Baker
Lotte Lambertsen
Mark van der Linden
Tim Mitchell
Bruno Pichon
Bill Hanage