

# Global emergence and population dynamics of divergent serotype 3 CC180 pneumococci

Taj Azarian<sup>1</sup>, Patrick K Mitchell<sup>1</sup>, Maria Georgieva<sup>1</sup>, Claudette M Thompson<sup>1</sup>, Amel Ghouila<sup>2</sup>, Andrew J Pollard<sup>3</sup>, Anna von Gottberg<sup>4</sup>, Mignon du Plessis<sup>4</sup>, Martin Antonio<sup>5</sup>, Brenda A Kwambana-Adams<sup>5</sup>, Stuart C Clarke<sup>6,7</sup>, Dean Everett<sup>8</sup>, Jennifer Cornick<sup>9</sup>, Ewa Sadowsky<sup>10</sup>, Waleria Hryniwicz<sup>10</sup>, Anna Skoczynska<sup>10</sup>, Jennifer C Moisi<sup>11</sup>, Lesley McGee<sup>12</sup>, Bernard Beall<sup>12</sup>, Benjamin J Metcalf<sup>12</sup>, Robert F Breiman<sup>13</sup>, PL Ho<sup>14</sup>, Raymond Reid<sup>15</sup>, Kate L O'Brien<sup>15</sup>, Rebecca A Gladstone<sup>16</sup>, Stephen D Bentley<sup>16</sup>, William P Hanage<sup>1</sup>

<sup>1</sup> Center for Communicable Disease Dynamics, Department of Epidemiology, T.H. Chan School of Public Health, Harvard University, Boston, MA; <sup>2</sup> Oxford Vaccine Group, Department of Paediatrics, University of Oxford; NIHR Oxford Biomedical Research Centre, Centre for Clinical Virology and Tropical Medicine (CCVTM), Churchill Hospital, Oxford OX3 7LJ, UK; <sup>3</sup> Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa; <sup>4</sup> Medical Research Council Unit, The Gambia, Atlantic Road, Fajara, The Gambia, London School of Hygiene & Tropical Medicine, London, UK; <sup>5</sup> University of Southampton Faculty of Medicine, University Hospital Southampton Foundation NHS Trust, Southampton SO16 6YD, UK; <sup>6</sup> Institute of Infection and Global Health, University of Liverpool, Liverpool, UK; <sup>7</sup> National Medicines Institute, Warsaw Poland; <sup>8</sup> Agence de Médecine Préventive, Paris, France; <sup>9</sup> Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA; <sup>10</sup> Emory Global Health Institute, Emory University, Atlanta, GA; <sup>11</sup> Department of Microbiology, Queen Mary Hospital University of Hong Kong, Hong Kong, People's Republic of China; <sup>12</sup> The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

## Introduction

- *Streptococcus pneumoniae* serotype 3 remains a significant cause of morbidity and mortality worldwide, despite inclusion in the 13-valent pneumococcal conjugate vaccine (PCV13).
- Serotype 3 increased in carriage since the implementation of PCV13 in the United States, while invasive disease rates remain unchanged.
- Genomic analysis of carriage samples from Massachusetts, USA showed a shift in the serotype 3 population, as a previously described sub-population of serotype 3, belonging to clonal complex (CC) 180 [Netherlands 3–31 (PMEN31) clone], became more common.
- To determine the spatiotemporal distribution of this lineage and reconstruct its evolutionary history, we obtained data from the Global Pneumococcal Sequencing (GPS) project, which possessed a wider range of sampling years and geographic regions.

## Methods

### Study Population and Epidemiological data

- The sample is comprised of 301 CC180 (PMEN31) genomes collected from carriage (n=70) and invasive disease (n=231) between 1993–2014, spanning 24 countries and six regions: North America (38.9%), Western Europe (17.9%), Asia (14.6%), Eastern Europe (14.3%), Africa (7.6%), and South America (6.3%).
- Country vaccine utilization and history were obtained from the International Vaccine Access Center (IVAC), Johns Hopkins Bloomberg School of Public Health. VIEW-hub. ([www.view-hub.org](http://www.view-hub.org)).

### Genomic Analysis

- Reference-based and *de novo* assemblies were constructed using whole-genome sequencing data.
- Pangenome analysis of *de novo* assemblies was conducted using Roary.
- Antibiotic resistance was predicted using ARIBA ([github.com/sanger-pathogens/ariba](https://github.com/sanger-pathogens/ariba)).
- Variants of polymorphic protein antigens were determined using SRST2 with a custom antigen database.

### Phylogenetic Analysis

- A maximum likelihood (ML) phylogeny was inferred using RAxML v8 to identify major clades of CC180.
- For each major clade, recombination was assessed using Gubbins, and ML phylogenies were used to test temporal signal by assessing correlation between strain isolation date and root-to-tip distance.
- Coalescent analysis was performed using BEAST v1.8.4. For each clade, strict and relaxed molecular clock models with constant and SkyGrid demographic models were tested using recombination-free alignments.

### Data Visualization and Statistical Analysis

- Statistical analysis and figure generation was performed using Rstudio v1.0.143.

### Serotype 3 Killing Assays

- To assess the efficiency of antisera against serotype 3 to opsonize pneumococcus for uptake and killing by differentiated polymorphonuclear leukocytes, we used an opsonophagocytosis killing assay.
- The killing assays were performed by using antisera against PCV13 and type 3 polysaccharide (PS) at multiple dilutions.

## Results

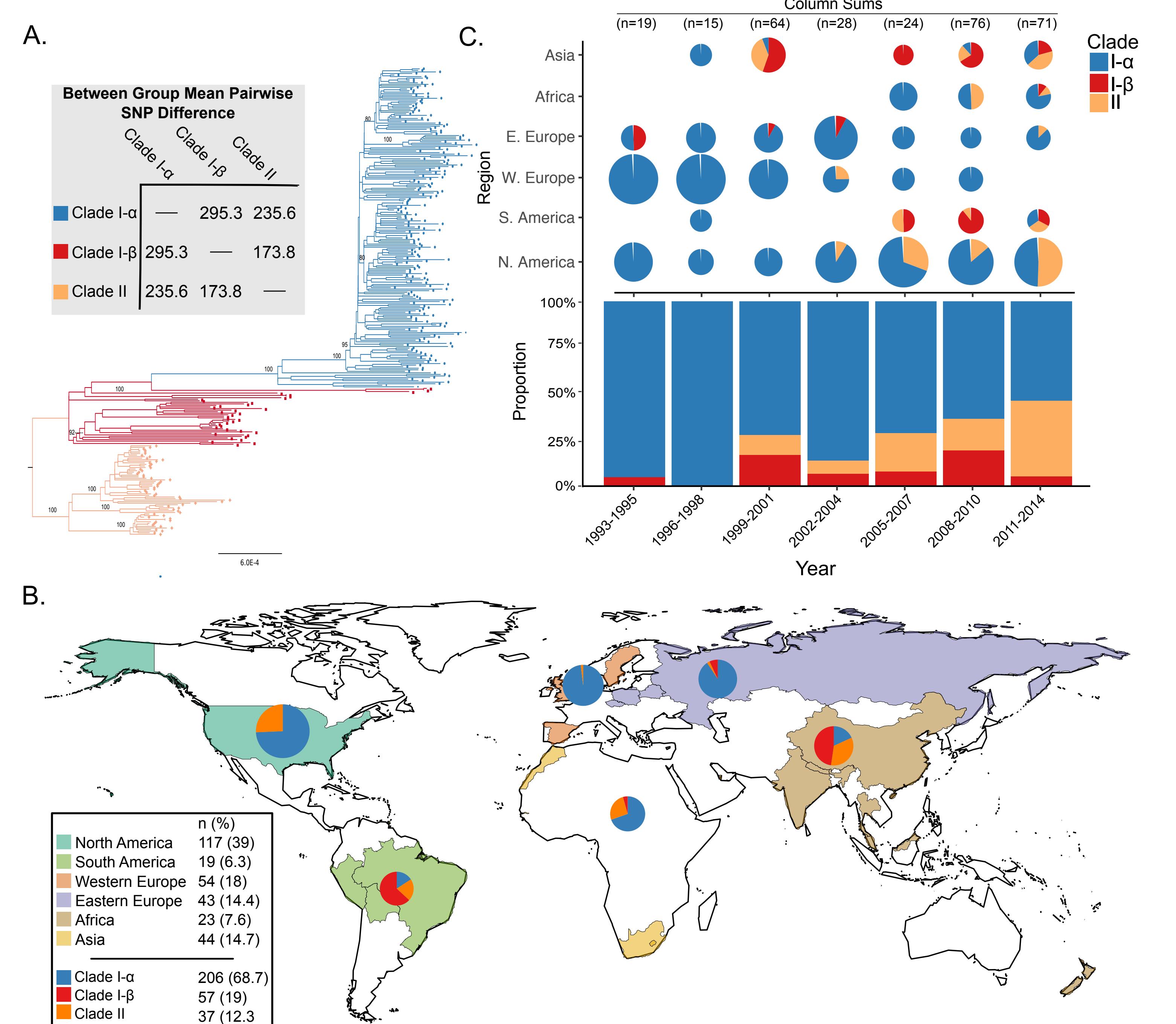
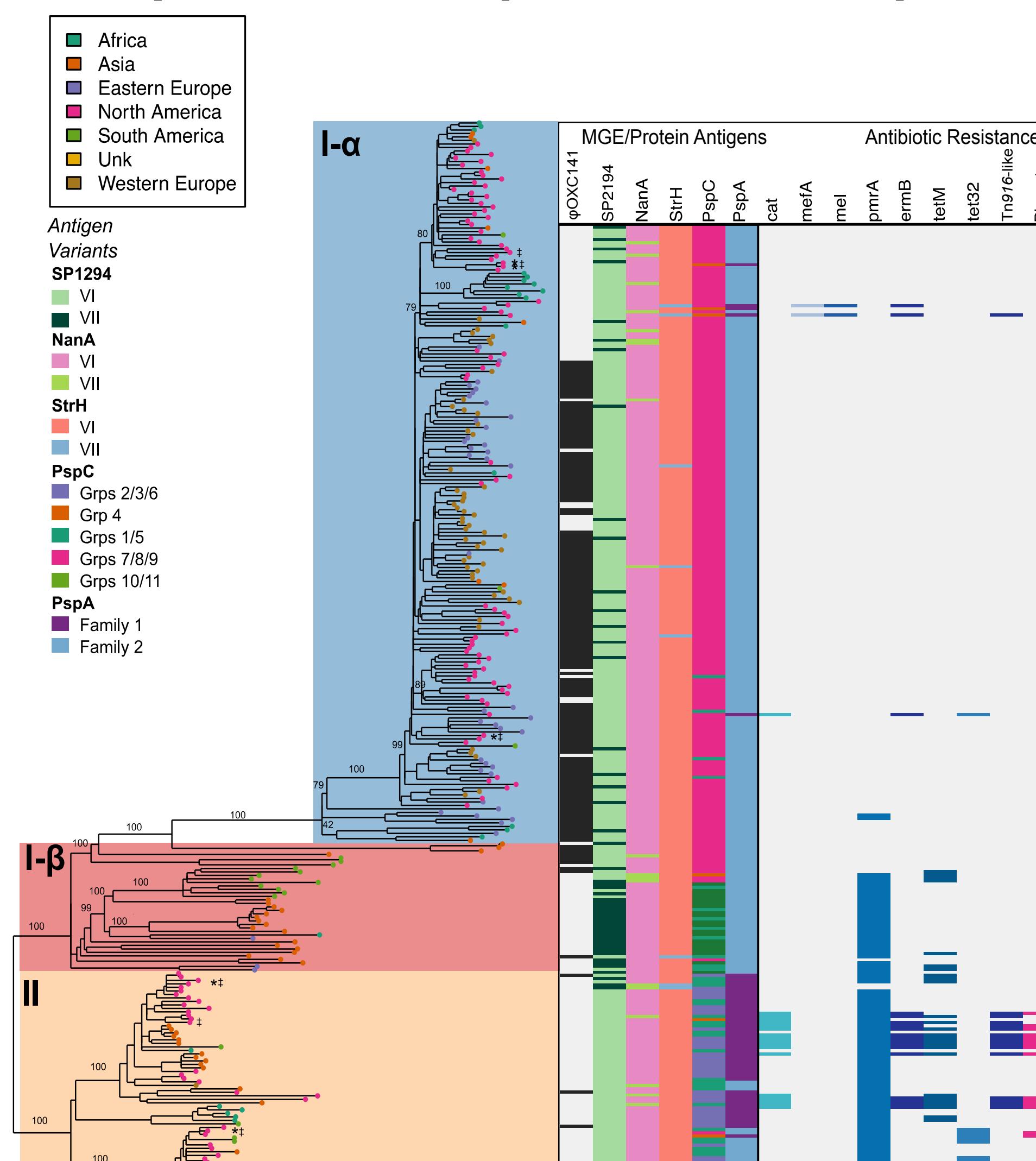


Figure 1.

- A)** Rooted ML phylogeny of *S. pneumoniae* serotype 3 CC180 isolates (n=301). Phylogeny was outgroup rooted using strain AP200, a serotype 11A, ST62 invasive isolate, which was immediately basal to the CC180 clade in a phylogeny of reference genomes. Bootstrap values are labelled on major clades. Mean pairwise between-clade SNP differences are presented in the shaded box.
- B)** World map illustrating sampled countries and regions with respective proportion of isolates belonging to Clade I-α, I-β, and II. Countries are colored according to region. Pie charts represent the proportion of isolates belonging to major serotype 3 clades with the size scaled to the proportion of strains from each region.
- C)** Proportion of clade membership by three-year collection window. The proportion of clade membership by region over-time is displayed on the top of the figure. Pie charts are scaled by the number of isolates sampled from a geographic region by time window (i.e., column). The overall proportion of clades for each time window is presented on the bottom of the figure.

## Results

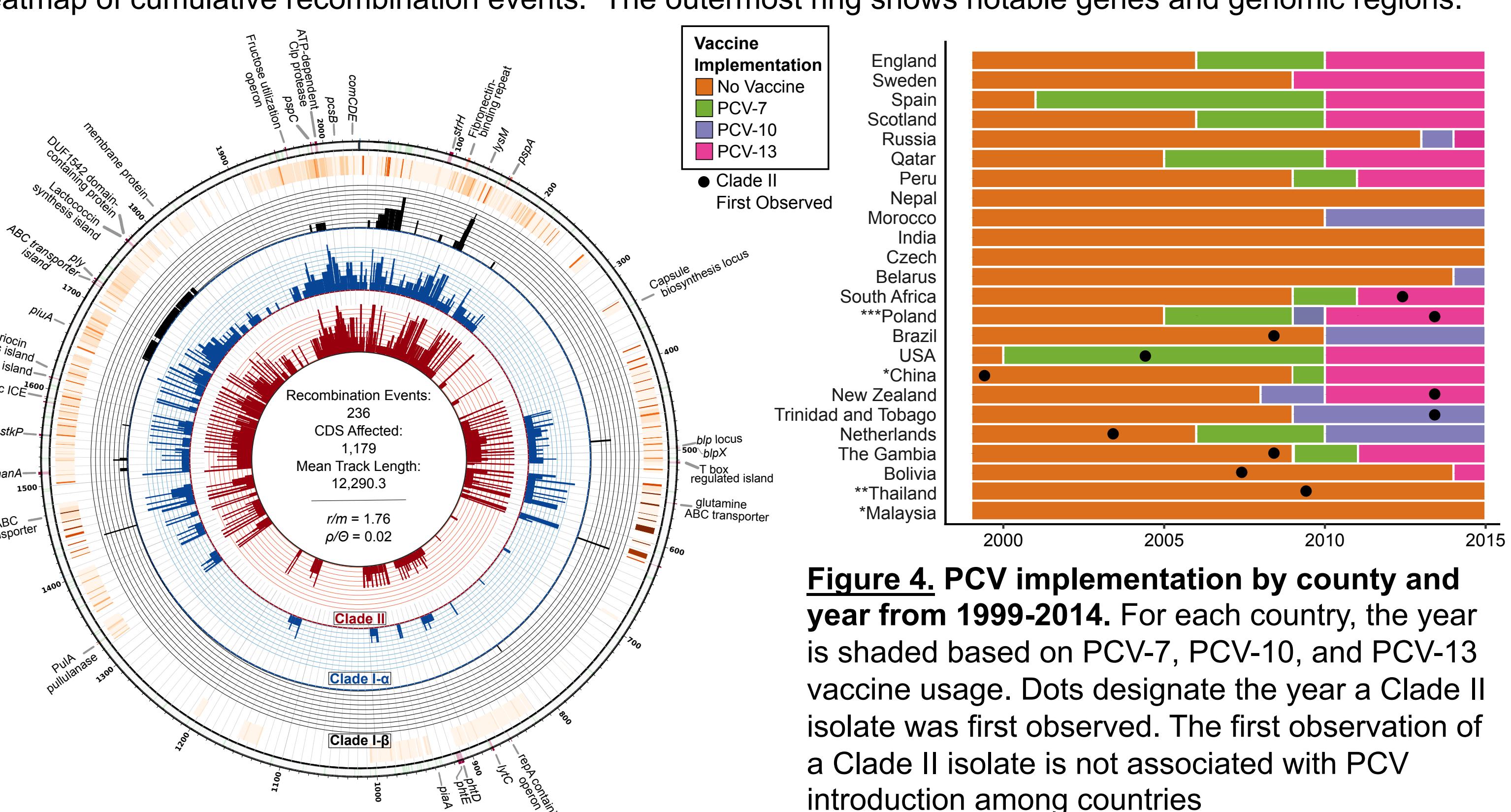
- Phylogenetic analysis of CC180 serotype 3 identified three major deep-branching lineages: Clade I-α (68.4% of isolates), Clade I-β (12.3%), and Clade II (19.3%) (Fig 1).
- The emergent Massachusetts lineage belongs to Clade II, which is globally distributed and has increased in prevalence from 11% of isolates sample from 1999-2001 to 41% in 2011-2014.
- Coalescent analysis dated the most recent common ancestor of Clade I-α in 1909 [95% HPD 1878-1929] and Clade II 1968 [95% HPD: 1939-1989], before PCV introduction.



**Figure 2.** Phylogeny, polymorphic protein antigen variants, and antibiotic resistance. Midpoint rooted ML phylogeny with geographic region of isolation represented as coloured tip shapes. Clade I-α, I-β, and II are shaded consistent with Fig 1A. Corresponding protein antigen variants for SP1294, NanA, StrH, PspC, and PspA are illustrated on the left half of the heatmap. Eight other protein antigens are excluded due to lack of variation in the sample. The presence and absence of antimicrobial resistance-associated genes are illustrated on the right half of the heatmap. The last column indicates genotypic antibiotic resistance was confirmed by phenotypic testing (broth dilution or disk diffusion).

- Among protein antigens, pneumococcal surface proteins A and C (PspA and PspC) were most variable between Clade I-α and II isolates (Fig 2).
- 25.9% of Clade II isolates possess a 48 kb Tn916/Tn1545-like conjugative transposon, which harbors *tetM* and *ermB*, conferring tetracycline and macrolide resistance (Fig 2).

**Figure 3.** Circos plot of recombination events. Moving from the inner ring outward, rings show a histogram of unique recombination events occurring among isolates belonging to Clades I-β, I-α, and II, respectively, followed by a heatmap of cumulative recombination events. The outermost ring shows notable genes and genomic regions.



**Figure 4.** PCV implementation by country and year from 1999-2014. For each country, the year is shaded based on PCV-7, PCV-10, and PCV-13 vaccine usage. Dots designate the year a Clade II isolate was first observed. The first observation of a Clade II isolate is not associated with PCV introduction among countries.

- Dominant clades are segregated by substantial ancestral recombination impacting antigenic variation (Fig 3). Recombination rates ( $r/m$ ) for Clades I-α, I-β, and II were 0.1, 3.4, and 2.7.
- Last, we observed opsonophagocytic killing of Clade II isolates in the presence of antisera against PCV13 and type 3 PS. In contrast, Clade I-α isolates were not susceptible to type 3 PS killing, and only one Clade I-α isolate was killed in the presence of PCV13 antisera.

## Conclusions

- Serotype 3, Clade II is globally distributed and has recently increased in prevalence.
- Clade II possesses a different protein antigen profile than Clade I-α and has a higher prevalence of antibiotic resistance genes, which may confer an evolutionary advantage.
- Strains differed in their susceptibility to killing by differentiated leukocytes as assessed in vitro by opsonophagocytosis assays. These findings underscore the importance of anti-protein antibodies that may synergize *in vivo* with anti-capsular antibodies to mediate killing. In our studies, anti-protein antibodies are not included.
- The increase of Clade II is not completely explained by vaccine effectiveness or implementation, but may result from the interplay between antibiotic resistance, vaccine utilization, and population immunity.
- Our analysis emphasizes the need for routine sampling of isolates from disperse geographic regions, including historically under-sampled areas. We also highlight the value of genomics in identifying antigenic variations that may have implications for vaccine development.

**Acknowledgements:** We acknowledge the Wellcome Trust Sanger Institute sequencing facility. The GPS project is funded by The Bill and Melinda Gates Foundation (Grant OPP1034556). This study was funded by NIH/NIAID R01 AI106786.

\*The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.