

Published in final edited form as:

Nat Genet. ; 44(3): 352–355. doi:10.1038/ng.1072.

Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event

Tanya Golubchik^{1,*}, Angela B. Brueggemann^{2,*}, Teresa Street¹, Robert E. Gertz Jr.³, Chris C. A. Spencer⁴, Thien Ho¹, Eleni Giannoulidou⁴, Ruth Link-Gelles³, Rosalind M. Harding², Bernard Beall³, Tim E. A. Peto⁵, Matthew R. Moore³, Peter Donnelly^{1,4,†}, Derrick W. Crook^{5,†}, and Rory Bowden^{1,4,5,†}

¹Department of Statistics, University of Oxford, Oxford, UK

²Department of Zoology, University of Oxford, Oxford, UK

³National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

⁵Oxford Biomedical Research Centre, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Streptococcus pneumoniae ('pneumococcus') causes an estimated 14.5 million cases of serious disease and 826,000 deaths annually in children <5 years of age¹. The highly effective US introduction of the PCV7 pneumococcal vaccine in 2000^{2,3} provided an unprecedented opportunity to investigate the response of an important pathogen to a widespread, vaccine-induced, selective pressure. Here we use array-based sequencing of 62 isolates from a US national monitoring program to study five independent instances of vaccine escape recombination⁴, demonstrating directly the simultaneous transfer of multiple and often large (up to at least 44kbp) DNA fragments. We show that one such novel strain quickly became established, spreading from East to West across the US. These observations clarify the roles of recombination and selection in the population genomics of pneumococcus, and provide proof-of-principle of the considerable value of combining genomic and epidemiological information in the surveillance and enhanced understanding of infectious diseases.

Childhood vaccination has proved effective against many viral and bacterial diseases, but more sophisticated vaccine approaches will be needed for pathogens with more complex genomes, life-cycles, and population structures, where the evolutionary responses of organisms are likely to be a key factor. Among the earliest successful bacterial vaccines were some for diseases (diphtheria, tetanus) where the actual cause of disease (e.g. a toxin) is targeted directly. Conjugate vaccines, in which a bacterial polysaccharide is joined to an immunogenic protein, have been developed for two other important childhood pathogens, *Haemophilus influenzae* type B and *Neisseria meningitidis* type C⁵⁻⁷. The success of these

Correspondence to: Peter Donnelly (peter.donnelly@well.ox.ac.uk).

*These authors made equal contributions to the work.

†These authors jointly supervised the work.

Author Contributions:

B.B. and R.E.G. collected isolates; A.B.B., B.B., R.B., D.C., P.D., T.G., R.H. and T.P. planned experiments; A.B.B., B.B., R.L.G., M.R.M. and T.P. collected and analysed epidemiological data; A.B.B., R.E.G., T.G. and T.S. performed molecular typing; T.G., T.H. and T.S. prepared DNA and performed microarray-based sequencing; T.G., A.B.B., R.B., P.D., C.C.A.S. and E.G. analysed molecular data; R.B., P.D. and D.C. wrote the manuscript; A.B.B., B.B., T.G., M.R.M. and T.P. contributed to writing the manuscript.

vaccine strategies was at least in part due to the relatively simple structure of the pathogen population, and the limited variability and evolvability of the pathogen molecule(s) targeted by the vaccine, but other organisms provide greater challenges.

Pneumococcal cells are covered by a layer of polysaccharide called the capsule, which serves as a major virulence factor and provides a target for vaccination. The first pneumococcal conjugate vaccine in routine use in infants was PCV7 (Pfizer), a conjugate incorporating the polysaccharides of seven of the >92 capsular types (serotypes) introduced to the US in 2000 to immediate and dramatic effect. By 2001, rates of invasive pneumococcal disease in the vaccinated age group, 0-2 years, had decreased by 69%² and by 2007 the rate in children <5 years old had stabilized at 24% of the level present before the vaccine³.

At the time of its introduction, there was considerable speculation about the likely results of the extreme selection induced by the vaccine, and its possible downstream consequences on vaccine efficacy and longevity^{8,9}. Colonization of the nasopharynx, especially of young children, provides the major reservoir for transmission of the pneumococcus. Vaccination reduces the rate of colonization of PCV7 serotypes, interrupting transmission while also allowing the rate of colonization by non-vaccine serotypes to increase¹⁰. Two mechanisms were anticipated for this serotype replacement: demographic expansion of non-vaccine serotype lineages and capsular switching – the replacement of the capsular gene cluster in one genome by the non-vaccine capsular genes from a different lineage.

We combined epidemiological and genomics approaches to better understand the nature, mechanisms, and consequences of vaccine escape. We sequenced a pre-selected subset of 12% (300kb) of the pneumococcus genome (Supplementary Figure 1) using Affymetrix CustomSeq technology and describe data in 62 isolates ascertained for potential epidemiological interest (Table 1). Whole-genome resequencing of several of the isolates on the Illumina GAIIx platform was used to confirm findings of particular interest.

During 2000-2007, approximately 27000 sterile-site pneumococcal isolates had been recovered from patients in 10 US states and serotyped by the Centers for Disease Control and Prevention in Atlanta, USA, as part of the “ABCs” monitoring programme³. The so-called sequence type of 1,902 serotype 19A isolates between 2001 and 2007, collected from patients of all ages, was then determined^{4,11,12} by MLST (multi-locus sequence typing), a widely used molecular fingerprinting approach in which (Sanger) sequence data is collected from seven fixed ~400-500bp fragments of essential genes¹³. Samples with a sequence type not commonly associated with serotype 19A were potential examples of vaccine escape through capsular switching¹⁴ (Table 1; details Supplementary Table 1). We had previously reported three distinct progeny strains, which we refer to as P1, P2, and P3, resulting from capsular switching with serotype 4 recipients⁴, which we confirmed by Sanger sequencing to identify recombinational breakpoints. Two further instances (P4, P5) were identified in the current study by resequencing of candidates (Supplementary Table 2). We have therefore defined a total of 5 independent instances of vaccine escape through capsular switches whereby serotype 4, which is included in the PCV7 vaccine, was replaced by serotype 19A, which is not.

In addition to identifying capsular switch recombinants, our resequencing approach allowed us to search for putative donor and recipient genomes. When one sequence takes up DNA from another through recombination, we refer to the former as the recipient sequence and the latter as the donor sequence. For P1 and P2, our sequencing revealed well-matched putative recipient and donor sequences with serotypes 4 and 19A. In addition, genomic analyses identified respectively 4 and 8 additional sequence fragments that did not match

known serotype 4 (recipient) genomes (Figure 1 (A); Supplementary Figure 2; Supplementary Table 3) and suggested that all the imported fragments could have originated from a single serotype 19A donor sequence in each case. Illumina sequencing of an early P1 and its prospective donor and recipient confirmed our analysis and identified 8 extra small imported fragments across the whole genome (Figure 1 (B)). To rule out the possibility that the additional fragments could have come from other serotype 4 sequences that we had not analysed, we used SNP typing to screen 88 archived US serotype 4 isolates collected around the time PCV7 was introduced for P1- and P2-specific additional sequence fragments, and found no alternative candidate recipients that could explain the structure of P1 and P2 without invoking multiple imports from a serotype 19A-like donor (Supplementary Note; Supplementary Table 7; Supplementary Table 8).

Our data thus demonstrate the independent origin of each recombinant lineage and strongly support the idea that multiple fragments may be transferred during a single episode of recombination. Across P1-P5, conservative estimates suggest a range of 1-27 fragments have been transferred in addition to the capsular locus, with sizes from 0.04 to at least 44 kb (Supplementary Table 3). While it was impossible to exclude separate sequential events in explaining each progeny structure, the observation that whenever we ascertained a capsular recombination event we saw other serotype 19A-like imports elsewhere in the genome is strong evidence that multiple fragments may be imported from the donor simultaneously, or in a short time sequence. Recombination involving transfer of large fragments or multiple fragments simultaneously has long been observed or inferred *in vitro* in pneumococcus¹⁵⁻¹⁸. A recent report¹⁹ documented multiple putative transfers in a single individual. Our findings show that such recombination events can happen not only *in vitro* or in individuals, but also at population scales, becoming evident after a nationwide immunization programme. A further recent report²⁰ detected several instances of capsular recombination in the 40-year global spread of a multi-drug-resistant lineage of pneumococcus but did not describe evidence for multi-fragment recombination, perhaps because of differences in sampling strategy or analytical methodology.

Predictions made at the introduction of the pneumococcal conjugate vaccine in the US about the potential for serotype replacement were confirmed by early data from the ABCs network². Among all non-vaccine serotypes, 19A has increased in frequency most, for a variety of possible reasons¹². Between 1998 and 2007, rates of invasive pneumococcal disease caused by serotype 19A increased ~2.5-fold and its share of disease at all ages increased from approximately 3% to 20%, reaching 47% in children under 5³ (Figure 2; Supplementary Table 4). Capsular switching as a means of vaccine escape was also predicted in advance, but the success of the vaccine escape lineage P1 is still remarkable. Isolates were first detected in New York (n = 3) and Connecticut (n = 1) in 2003¹⁴ and have spread westward in subsequent years. Since 2003, P1 has become one of the most prevalent genotypes in post-vaccine populations, having been recovered from 175 patients of all ages by the end of 2007. In contrast, three of the other four vaccine escape lineages we detected, P3-P5, have been seen only once in our screen, and P2 has been observed 8 times, predominantly in the northeastern US.

The spread of vaccine escape recombinant P1 and to a lesser extent P2 has also allowed us an unprecedented opportunity to observe pneumococcal evolution in real populations in real time. Genomic analyses of the evolution of the P1 and P2 lineages demonstrate that recombination events have continued to occur and imply that when recombination can be definitively inferred it tends to involve multiple genomic fragments (Supplementary Note; Supplementary Figure 3). With no ascertainment bias to favour recombination episodes involving a large transferred sequence such as the capsular locus, these data are consistent with a model of variable numbers of smaller transferred sequences and with published

estimates of the relative rates of recombination and mutation^{20,21}. Depending on the assumptions made, the proportion of new variation within the P1 and P2 recombinant lineages that has arisen due to recombination can be estimated at at least ~60% (details in Supplementary Note).

In this study we have observed, in 5 separate vaccine escape lineages and during the subsequent evolution of two of those lineages, 11 separate episodes of recombination leading to the import of sequences into a pneumococcal genome. In only two of these episodes was there no evidence for transfer of multiple separate fragments and so we conclude that multi-fragment recombination is commonplace in pneumococcal populations. One consequence of this is that even the terminology “capsular switch” is potentially misleading because it suggests that only the capsular locus has been transferred.

Documentation of multi-fragment recombination in real populations is particularly interesting because it has profound consequences for the way in which an organism may be able to traverse its evolutionary fitness landscape. For example, moderate to high level beta-lactam class antimicrobial resistance is usually associated with horizontal transfer of variants at three dispersed *pbp* loci, and drug resistance took about two decades after the introduction of penicillin to first emerge in pneumococcus. Having emerged, penicillin resistance determinants now spread rapidly from one genetic background to another under drug-induced selective pressure and pose a significant threat to treatment. Multi-fragment recombination could also have been important in generating the (currently unknown) factor(s) which allowed P1 to surpass many other non-vaccine lineages in invasive disease incidence. The recent introduction of a 13-valent pneumococcal conjugate vaccine including serotype 19A in the US and elsewhere is likely to reduce significantly the impact of serotype 19A on vaccinated populations, but how many, and which, serotypes will be needed for a vaccine that provides acceptable long-term disease reduction are still unknown.

Modern high-throughput molecular technologies now allow typing of bacterial isolates on a genomic scale, thus providing much greater resolution than current, standard, MLST approaches. We have described a proof-of-principle experiment which confirms the potential for combining genome-scale genetic information with epidemiological data, in this case in better understanding serotype replacement following introduction of a conjugate vaccine. We identified five independent instances of vaccine escape through capsular switching from serotype 4 to 19A. Our genomic data provide strong evidence that in each case the recombination event generating the capsular switch involved simultaneous import of multiple and often large additional DNA fragments around the genome. This process has far-reaching consequences for the evolution of bacteria and their response to the strong selection imposed by vaccines or antimicrobials. It may also play a role in the striking success of the P1 vaccine escape lineage as an invasive pathogen among the 19A lineages present after vaccine introduction. While vaccine escape through capsular switching was correctly predicted in advance of the vaccination programme, our analyses show that, particularly in the light of complex recombination mechanisms, its specific consequences are difficult to predict.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge the clinicians, microbiologists, and investigators of the Active Bacterial Core surveillance program of the Emerging Infections Program Network. We thank Xavier Didelot for contributions to the analysis of Illumina genomic data. This work was funded by the Wellcome Trust: ref. 079126/Z/06/Z. T.P. and

D.C. are funded by the NHS NIHR Oxford Biomedical Research Centre and NIHR Senior Investigator Awards. P.D. is funded by Wellcome Trust Core Award Grant ref. 090532/Z/09/Z and is supported in part by a Wolfson Royal Society Merit Award. R.B. is supported by the NHS NIHR Oxford Biomedical Research Centre and UKCRC (MRC UK Ref G0800778 and Wellcome Trust Ref. 087646/2/08/2). A.B.B. is a Wellcome Trust Career Development Fellow (Ref. 083511/Z/07/Z). Genetic and Epidemiological data is available from the authors.

Competing Financial Interests

D.C. and T.P. are in receipt of a research grant for pneumococcal surveillance from Pfizer. A.B.B. is in receipt of grant funding from GlaxoSmithKline Biologicals and Pfizer (Wyeth) Vaccines.

APPENDIX

Online Methods

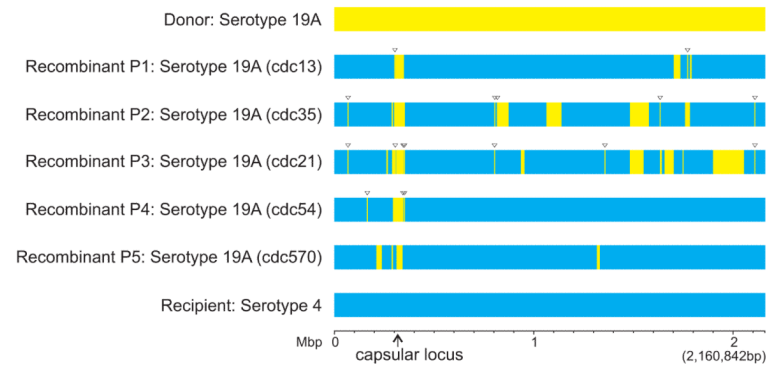
More detailed methods and preliminary results are included in Supporting Materials. Polymerase chain reaction (PCR) and Sanger sequencing for MLST and other sequence typing used standard methods^{4,13}. Primer sequences targeting the upstream and downstream recombination breakpoints are shown in Supplementary Table 5. The GeneChip CustomSeq platform (Affymetrix, Santa Clara, CA, USA)^{23,24} was used to target a 300kb subset of the *S. pneumoniae* genome for resequencing. Sequence fragments were selected and optimized for inclusion in a custom array design (Supplementary Figure 1). Samples were processed according to the maker's instructions: 16µg of genomic DNA was labelled and hybridized to an array, the array was washed and scanned and hybridization data was analysed using onboard software. Sequence calls were filtered using a bespoke method designed to produce high-accuracy calls from diverse sequences (Supplementary Table 6; details in Supplementary Note). PCR-RFLP (PCR-restriction fragment length polymorphism) analysis for SNP typing used fluorescently labelled primers (Supplementary Table 7) according to established protocols and data was collected on an ABI3730 automated sequencer. Illumina sequencing on the Genome Analyzer Iix platform (Illumina, San Diego, CA, USA) employed standard methods to produce 51b paired reads which were assembled using Velvet²⁵ and then aligned to a reference sequence using Mauve²⁶. Bioinformatic analyses used Python and R²⁷.

References

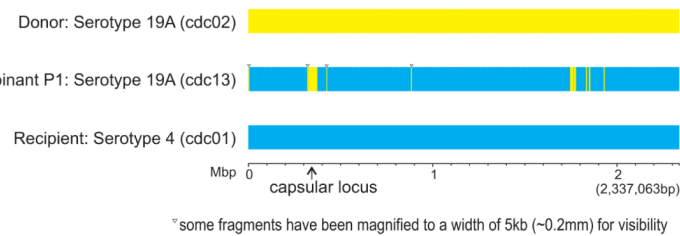
1. O'Brien KL, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet*. 2009; 374:893–902. [PubMed: 19748398]
2. Whitney CG, et al. Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. *N Engl J Med*. 2003; 348:1737–46. [PubMed: 12724479]
3. Pilishvili T, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *J Infect Dis*. 2010; 201:32–41. [PubMed: 19947881]
4. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog*. 2007; 3:e168. [PubMed: 18020702]
5. Eskola J, et al. A randomized, prospective field trial of a conjugate vaccine in the protection of infants and young children against invasive *Haemophilus influenzae* type b disease. *N Engl J Med*. 1990; 323:1381–7. [PubMed: 2233904]
6. Perkins BA. New opportunities for prevention of meningococcal disease. *Jama*. 2000; 283:2842–3. [PubMed: 10838655]
7. Campbell H, Borrow R, Salisbury D, Miller E. Meningococcal C conjugate vaccine: the experience in England and Wales. *Vaccine*. 2009; 27(Suppl 2):B20–9. [PubMed: 19477053]
8. Lipsitch M. Bacterial vaccines and serotype replacement: lessons from *Haemophilus influenzae* and prospects for *Streptococcus pneumoniae*. *Emerg Infect Dis*. 1999; 5:336–45. [PubMed: 10341170]
9. Spratt BG, Greenwood BM. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet*. 2000; 356:1210–1. [PubMed: 11072934]

10. Bogaert D, et al. Colonisation by *Streptococcus pneumoniae* and *Staphylococcus aureus* in healthy children. *The Lancet*. 2004; 363:1871–1872. [PubMed: 15183627]
11. Beall B, et al. Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. *J Clin Microbiol*. 2006; 44:999–1017. [PubMed: 16517889]
12. Moore MR, et al. Population snapshot of emergent *Streptococcus pneumoniae* serotype 19A in the United States, 2005. *J Infect Dis*. 2008; 197:1016–27. [PubMed: 18419539]
13. Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology*. 1998; 144(Pt 11): 3049–60. [PubMed: 9846740]
14. Pai R, et al. Postvaccine genetic structure of *Streptococcus pneumoniae* serotype 19A from children in the United States. *J Infect Dis*. 2005; 192:1988–95. [PubMed: 16267772]
15. Griffith F. The significance of pneumococcal types. *Journal of Hygiene*. 1928; 27:113–159. [PubMed: 20474956]
16. Avery OT, Macleod CM, McCarty M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus Type Iii*. *J Exp Med*. 1944; 79:137–58. [PubMed: 19871359]
17. Lacks S, Hotchkiss RD. A study of the genetic material determining an enzyme in *Pneumococcus*. *Biochim Biophys Acta*. 1960; 39:508–18. [PubMed: 14413322]
18. Trzcinski K, Thompson CM, Lipsitch M. Single-step capsular transformation and acquisition of penicillin resistance in *Streptococcus pneumoniae*. *J Bacteriol*. 2004; 186:3447–52. [PubMed: 15150231]
19. Hiller NL, et al. Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog*. 2010; 6:e1001108. [PubMed: 20862314]
20. Croucher NJ, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science*. 2011; 331:430–4. [PubMed: 21273480]
21. Feil EJ, Smith JM, Enright MC, Spratt BG. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics*. 2000; 154:1439–50. [PubMed: 10747043]
22. Tettelin H, et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science*. 2001; 293:498–506. [PubMed: 11463916]
23. Cutler DJ, et al. High-throughput variation detection and genotyping using microarrays. *Genome Res*. 2001; 11:1913–25. [PubMed: 11691856]
24. Zwick ME, et al. Microarray-based resequencing of multiple *Bacillus anthracis* isolates. *Genome Biol*. 2005; 6:R10. [PubMed: 15642093]
25. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008; 18:821–9. [PubMed: 18349386]
26. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004; 14:1394–403. [PubMed: 15231754]
27. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2007.

A: Microarray-based resequencing of recombinants P1-P5



B: Whole-genome structure of recombinant P1

**Figure 1.**

Resequencing of pneumococcal vaccine escape recombinants: comparison of recombinant and putative recipient and donor sequences.

(A) Vaccine escape recombinants P1-P5 were resequenced using the Affymetrix CustomSeq platform, a microarray-based approach covering a selected 12% of the pneumococcus genome (300 kbp). The genomes are coloured by inferred origin of genomic intervals (blue - recipient; yellow - donor). The capsular locus transferred in a different form in each recombinant is evident as a large yellow block centred on the same location towards the left of each genome. Other yellow blocks indicate extra transferred sequences. The recipient genomes in all cases were highly similar.

(B) Whole-genome structure of recombinant P1. Representatives of recombinant P1 and its best-matching prospective parents were sequenced on the Illumina GAIIX platform. Sequence reads were assembled de novo and aligned to the TIGR4 reference sequence²² and a heuristic algorithm was implemented in which three or more donor-like single nucleotide variants encompassing no more than one recipient-like variant was used to identify donor genomic fragments (blue - recipient; yellow - donor). This approach detected 15 imports of ~50bp, including 8 that did not overlap previously detected fragments. The scale in **(B)** is different to that in **(A)** because the reference sequences differ slightly. In each panel the smallest fragments (triangles) have been magnified to a minimum size of 5kb for visibility.

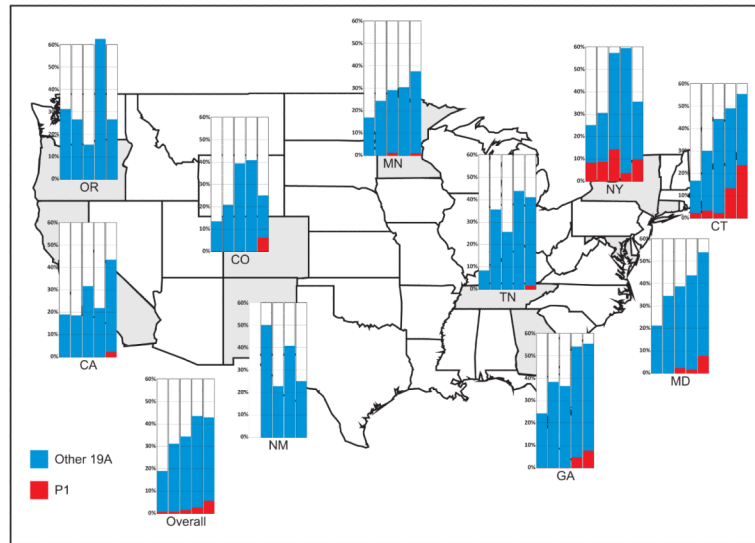


Figure 2. Spread of P1 vaccine escape recombinant through space and time. Incidence of P1 vaccine escape recombinants (red) and other 19A (blue) as a proportion of all pneumococci in invasive pneumococcal disease among children under 5 years of age. Data are shown for the years 2003-2007; left to right panels for each state in 10 ABCs monitoring states. Data from 2003 in New Mexico are missing (grey box) because surveillance started there in 2004. P1 was first detected in New York and Connecticut from where it has spread westwards.

Table 1

Summary of resequenced samples

Sample Group	number identified	number sequenced	year(s)	Serotype	Sequence Type
Vaccine Escape Recombinant - P1	175	42	2003-2007	19A	695, 2363
Vaccine Escape Recombinant - P2	8	8	2005-2007	19A	2365
Vaccine Escape Recombinant - P3	1	1	2005	19A	899
Vaccine Escape Recombinant - P4	1	1	2006	19A	695
Vaccine Escape Recombinant - P5	1	1	2007	19A	695
Candidate Donors		5	1999-2003	19A	199, 645
Candidate Recipients		4	1999-2002	4	695, 899

A total of 62 pneumococcal isolates were resequenced on the Affymetrix CustomSeq platform, including all identified examples of vaccine escape recombinants P2-P5 and a selection of P1 isolates collected from US sites during 2003-2007. Candidate donor and recipient isolates were chosen based on MLST sequence type from a set of serotype 4 and 19A isolates collected between 1999-2003 and genotyped by the Centers for Disease Control and Prevention (serotype 4, n = 97; serotype 19A, n = 231).