Genome Biology

Genome Biology

This Provisional PDF corresponds to the article as it appeared upon acceptance. Copyedited and fully formatted PDF and full text (HTML) versions will be made available soon.

The multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success

Genome Biology 2012, **13**:R103 doi:10.1186/gb-2012-13-11-r103

Kelly L Wyres (kelly.wyres@zoo.ox.ac.uk)
Lotte M Lambertsen (LLM@ssi.dk)
Nicholas J Croucher (nc3@sanger.ac.uk)
Lesley McGee (afi4@cdc.gov)
Anne von Gottberg (annev@nicd.ac.za)
Josefina Linares (fina.linares@bellvitgehospital.cat)
Michael R Jacobs (mrj6@case.edu)
Karl G Kristinsson (karl@landspitali.is)
Bernard W Beall (beb0@CDC.GOV)
Keith P Klugman (kklugma@emory.edu)
Julian Parkhill (parkhill@sanger.ac.uk)
Regine Hakenbeck (hakenb@rhrk.uni-kl.de)
Stephen D Bentley (sdb@sanger.ac.uk)
Angela B Brueggemann (angela.brueggemann@zoo.ox.ac.uk)

ISSN 1465-6906

Article type Research

Submission date 23 August 2012

Acceptance date 17 October 2012

Publication date 16 November 2012

Article URL http://genomebiology.com/2012/13/11/R103

This peer-reviewed article can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in *Genome Biology* are listed in PubMed and archived at PubMed Central.

For information about publishing your research in *Genome Biology* go to

http://genomebiology.com/authors/instructions/

© 2012 Wyres *et al.*
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success

Kelly L Wyres[1], Lotte M Lambertsen[2], Nicholas J Croucher[3], Lesley McGee[4], Anne von Gottberg[5], Josefina Liñares[6], Michael R Jacobs[7], Karl G Kristinsson[8], Bernard W Beall[4], Keith P Klugman[9], Julian Parkhill[3], Regine Hakenbeck[10], Stephen D Bentley[3] and Angela B Brueggemann[1*]

[1]Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom

[2]Department of Microbiology Surveillance and Research, Statens Serum Institut, 5 Artillerivej, DK 2300 Copenhagen S, Denmark

[3]Pathogen Genomics Team, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom

[4]Streptococcus Laboratory, Centers for Disease Control and Prevention, 1600 Clifton Rd, Atlanta, Georgia, 30333, USA

[5]Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases, Sandringham, Johannesburg 2131, South Africa

[6]Servicio de Microbiología, Hospital Universitari de Bellvitge, Feixa Llarga s/n, 08907 L'Hospitalet de Llobregat, Barcelona, Spain

[7]Department of Pathology, Case Western Reserve University, Cleveland, Ohio, 44106, USA

[8]Clinical Microbiology Department, Landspitali University Hospital and University of Iceland, 101 Reykjavik, Iceland

[9]Hubert Department of Global Health Epidemiology, Rollins School of Public Health, 1518 Clifton Road NE, Atlanta, Georgia, 30322, USA

[10]Department of Microbiology, University Kaiserslautern, Paul-Ehrlich-Straße, 67663 Kaiserslautern, Germany

*Corresponding author: angela.brueggemann@zoo.ox.ac.uk

Abstract

Background: *Streptococcus pneumoniae*, also called the pneumococcus, is a major bacterial pathogen. Since its introduction in the 1940s, penicillin has been the primary treatment for pneumococcal diseases. Penicillin resistance rapidly increased among pneumococci over the past 30 years, and one particular multidrug-resistant clone, PMEN1, became highly prevalent globally. We studied a collection of 426 pneumococci isolated between 1937 and 2007 to better understand the evolution of penicillin resistance within this species.

Results: We discovered that one of the earliest known penicillin-nonsusceptible pneumococci, recovered in 1967 from Australia, was the likely ancestor of PMEN1, since approximately 95% of coding sequences identified within its genome were highly similar to those of PMEN1. The regions of the PMEN1 genome that differed from the ancestor contained genes associated with antibiotic resistance, transmission and virulence. We also revealed that PMEN1 was uniquely promiscuous with its DNA, donating penicillin-resistance genes and sometimes many other genes associated with antibiotic resistance, virulence and cell adherence, to many genotypically diverse pneumococci. In particular, we describe two strains in which up to 10% of the PMEN1 genome was acquired in multiple fragments, some as long as 32 kb, distributed around the recipient genomes. This type of directional genetic promiscuity from a single clone to numerous unrelated clones has, to our knowledge, never before been described.

Conclusions: These findings suggest that PMEN1 is a paradigm of genetic success both through its epidemiology and promiscuity. These findings also challenge the existing views about horizontal gene transfer among pneumococci.

Background

Worldwide, over 1.6 million deaths annually are attributed to the bacterial pathogen, *Streptococcus pneumoniae,* the "pneumococcus" [1]. This bacterium is a leading cause of otitis media, sinusitis, pneumonia and meningitis, and is associated with high mortality rates in the developing world [1, 2]. While many pneumococcal diseases can be successfully treated with antibiotics such as penicillin (and other β-lactams), resistance is a major global problem. The primary pneumococcal penicillin resistance determinants are three of the six penicillin-binding protein (*pbp*) genes, *pbp2x, pbp1a* and *pbp2b.* In penicillin-nonsusceptible pneumococci (PNSP) these genes contain mosaic sequence blocks differing from those of penicillin-susceptible pneumococci (PSP) by up to 25% nucleotide divergence [3, 4]. It is believed that these mosaic blocks were acquired from closely-related viridans streptococci via homologous recombination [3, 5].

PNSP were first reported in the late 1960s (in the USA and Australia) [6, 7]. In the late 1970s the first high-level penicillin-resistant pneumococci were reported in South Africa [8, 9]. Throughout the 1980s and 1990s global penicillin nonsusceptibility rates rocketed, surpassing 50% in some regions [10]. The selective pressure of antibiotic use was enormous at this time, as antibiotics were widely used, even overused, in many countries. This selective pressure meant that pneumococci with some level of resistance to penicillin or other antibiotics had a distinct advantage over susceptible strains, which may have indirectly contributed to a change in the distribution of pneumococcal capsular types (or serotypes, defined by the antigenic polysaccharide capsule surrounding the cell) [11]. Following the introduction of a 7-valent pneumococcal conjugate vaccine (PCV7) in the year 2000 in the USA, there was a substantial reduction in PNSP levels [12]. Use of PCV7 in children provides efficacious protection against 7 of the 94 known pneumococcal serotypes [1], including those which were most commonly associated with penicillin resistance.

Pneumococcal disease caused by PCV7 type pneumococci has significantly declined in countries that have introduced PCV7. Subsequently, there has been an increase in pneumococcal disease due to non-PCV7 type pneumococci [13], many of which are now also penicillin nonsusceptible [14]. 10- and 13- valent pneumococcal vaccines, containing additional serotypes, have recently been introduced, and continued surveillance over the next few years will be essential to determine their effects.

Alongside the increase of PNSP there has been an emergence of predominant and widely-distributed pneumococcal clones. 43 important disease-causing clones, many of which are multidrug-resistant, have been identified by the Pneumococcal Molecular Epidemiology Network (PMEN) [15]. The first of these clones was Spain^{23F}-1, here called PMEN1. This multidrug-resistant pneumococcus predominantly circulates as a vaccine serotype 23F, multilocus sequence type 81 clone (ST81^{23F}); however, it has also been associated with several other serotypes, including both vaccine types and non-vaccine types [16, 17]. The first PMEN1 representative was isolated in Spain in 1984 and shortly thereafter in the USA [18], the United Kingdom [16], South Africa [19, 20], Hungary [21], and South America [22]. By the late 1990s it was estimated that ~40% of PNSP circulating in the USA were members of this clone [23] and it continues to circulate among pneumococcal populations today.

A recent study of ~240 pneumococci representing PMEN1 ST81 and related clonal variants (all belonging to clonal complex (CC) 81) showed that there is a considerable amount of genetic diversity within this lineage. This diversity, which largely results from hundreds of recombination events, indicates rapid genomic evolution and presumably allowed rapid response to selective pressures such as those imposed by vaccine and antibiotic use [17].

Here we studied a large (n = 426), genetically diverse, historical and global pneumococcal isolate collection with the aim to better understand the evolution of penicillin resistance within this species. We uncovered the important contribution of the PMEN1 clone to the emergence and spread of penicillin-resistant pneumococci and revealed its genetically highly promiscuous nature. Furthermore, we have been able to confidently identify the ancestor of this globally and historically important clone as being closely related to one of the oldest known PNSP.

Results

PMEN1 pbp promiscuity

Pneumococcal study isolates (Table 1) were assigned to CCs based on multilocus sequence type (MLST) data. Isolates within CCs are considered to have shared ancestry, while those in different CCs are more distantly related. Within our collection, identical or highly similar, 'nonsusceptible' *pbp* alleles and/or sequence regions were identified repeatedly among pneumococci belonging to different CCs. These *pbp* sequences were those associated with the PMEN1 reference strain (Genbank accession: NC_011900.1, hereafter simply called 'PMEN1').

Two isolates, CGSP14 (Genbank accession NC_010582.1) and the PMEN3 reference strain (Genbank accession ABGE00000000.1, hereafter called 'PMEN3'), representing independent CCs (Table 2), shared identical *pbp2x, pbp1a* and *pbp2b* alleles with PMEN1. Analysis of the genomic regions flanking these *pbp* genes showed that in fact much larger regions of the CGSP14 and PMEN3 genomes (5,362 – 22,104 bp, Figure 1) were identical or highly similar to that of PMEN1, suggesting these regions, which conferred penicillin nonsusceptibility, were acquired from a PMEN1 representative(s) in both cases. The PMEN2 reference strain, representing another CC (Table 2), also shared identical regions (225 - 1034 bp in length) of all three of its *pbps* with PMEN1 (the

intervening regions of sequence were different). A total of 12 distinct CCs were represented by isolates sharing identical *pbp* sequences in whole or part (119 - 1376 bp) in one or two of their *pbps* to PMEN1, suggesting that they too were acquired from a PMEN1 representative. Nine additional CCs were represented by isolates sharing *pbp2x* and/or *pbp2b* regions (199 - 1436 bp) that differed from those of PMEN1 by only one or two nucleotide substitutions, again suggesting possible acquisition from a PMEN1 representative (see Additional file 1 for *pbp* allele and CC information). No other *pbp* alleles were as widely distributed as those of PMEN1.

Extensive PMEN1 genomic recombination

Given that the CGSP14 and PMEN3 genomes each contained three large, independent regions which were putatively acquired from PMEN1, we used whole genome sequence data generated on the Illumina platform to assess whether other regions of the PMEN1 genome were also shared by CGSP14 and PMEN3 (both of which represent multidrug-resistant clones, the latter of which is also globally-distributed). We found that in fact 15 regions (1,663 – 32,312 bp, totalling 211,963 bp, ~9.5% of the genome) of the CGSP14 chromosome were identical to those of PMEN1 or differed by only 1-2 nucleotides across a minimum of 7,537 bp (Figure 2 and Additional file 2), and were not identical in any of four CGSP14 ancestral representative genomes. The ancestral representatives (14/5, the PMEN9 reference strain, ICE13 and ICE570; Table 2) were chosen from our collection because they represented older and/or penicillin-susceptible members of the CGSP14 CC and so likely resembled the true ancestor of the penicillin-resistant CGSP14. Importantly, this allowed us to identify and exclude identical sequence that was shared by descent, whereas regions of the CGSP14 genome that differed from its ancestral representatives but were identical to those of PMEN1 were most likely acquired from PMEN1 through the process of recombination. The ancestral isolates were generally all identical or highly similar to each other at the genomic regions

in question; sequences from these isolates were also highly similar to CGSP14 in the regions flanking the putative recombination imports, confirming their suitability as ancestral representatives.

Following a similar comparison of PMEN3 and its ancestral representatives (9A/1, 9V/5 and 9V/6; Table 2), we identified nine PMEN3 genomic regions (4,662 – 22,104 bp, totalling 117,689 bp, ~5.3% of the genome) that were identical to sequences from PMEN1 or differed by only two nucleotides across 22,104 bp, and were different from those of the ancestral representatives (Figure 2 and Additional file 2). We suggest that these regions were also acquired by PMEN3 from PMEN1 through the process of recombination.

The CGSP14 and PMEN3 putative recombinogenic regions contained 213 and 108 whole or part coding sequences, respectively (see Additional file 2). In addition to the penicillin resistance-conferring *pbp2x, pbp1a* and *pbp2b* genes, the following coding sequences associated with enhanced virulence or colonisation potential were included: the CGSP14 *bgaA, clpX,* SPN23F_05810 and SPN23F_20890 genes associated with enhanced host colonisation ability [24]; the CGSP14 virulence-associated *cbpJ, msmK*, chloramphenicol acetyltransferase and zeta toxin/epsilon antitoxin genes [25-27] (the latter two of which are situated on the CGSP14 conjugative transposon); and the PMEN3 virulence-associated *cbpD, cbpJ* and *clpC* genes [26].

The possibility that the putative imported regions were acquired by CGSP14 and PMEN3 from non-PMEN1 representative pneumococci cannot be ignored. However, the majority, thirteen and six respectively, of regions described here were not identical among any of 96 additional pneumococcal genomes, representing a diverse range of 31 independent CCs (listed in Additional

file 3, excluding 23F/4 and CC66 isolates). Two and three of the regions respectively (Figure 2), were identical or highly similar among isolates representing CC66, and so it is possible that those regions may have been acquired from a CC66 representative (also see Additional file 4). Although our collection cannot capture the full complement of pneumococcal genetic diversity, given these results and the extremely high similarity of the putative import regions to the PMEN1 genome we believe that they originated from a PMEN1 representative. Thus these data suggest that DNA originating from the PMEN1 clone has greatly influenced the genomic evolution of at least two unrelated pneumococcal clones.

Ancestors of PMEN1

The Genome Comparator tool [28] was used to screen 104 pneumococcal genomes, representing a very diverse range of CCs and serotypes (see Additional file 3), for regions of DNA (aside from the *pbps*) that were identical to PMEN1. 1990 coding sequences were identified from the PMEN1 Genbank Accession NC_011900.1 file (totalling 1,833,672 bp, ~83% of the genome). Comparable coding sequences in the query genomes were identified by BLASTn search if they shared ≥70% nucleotide sequence identity and could be aligned at 100% length. Consequently, coding sequences spanning the ends of assembly contigs could not be identified.

Surprisingly and uniquely, the analyses showed that 23F/4, the first PNSP isolated from Australia (Table 2), was very likely to be an ancestral representative of PMEN1, since ~95% of the identified coding sequences were ≥98% similar to those of PMEN1. ~75% of identified coding sequences in 23F/4 were identical to PMEN1 and over one-third of the coding sequences that were not identical differed by only one or two nucleotide substitutions. The *pbp2x, pbp1a, pbp2b* genes and three of the seven MLST loci were not identical. Additionally, a comparison of the PMEN1 and 23F/4

genomes, performed using the Artemis Comparison Tool [29] using all available sequence (both coding and non-coding sequence, Figure 3), clearly showed three large regions of the PMEN1 genome that were partially or completely missing from the 23F/4 genome and were presumably acquired by a more recent ancestor of PMEN1. These were the ICE*Sp*23FST81 element, the ϕMMI phage and a Na^+-dependent ATPase island, all associated with increased colonisation and virulence potential [30-33].

The Genome Comparator analyses also revealed that the majority (87, 83.6%) of genomes analysed – representing 31 unique CCs and 30 serotypes – shared 15.1% ± 2 standard deviations (excluding twelve outliers with >30% identical coding sequences) of the identified coding sequences identically with PMEN1 (Figure 4). We suggest that these "baseline" coding sequences were shared through ancestral descent, i.e. they were identical in the most recent common ancestor of these clones and have not undergone subsequent change. (For purposes of comparison, the same analysis was performed using pneumococcal genome TIGR4 (Genbank accession AE005672.3) as the reference and a similar baseline mean level of identical sequence of 14.5% was observed. See Additional file 5.)

Isolates belonging to CC66 shared between 44.5% and 58.4% coding sequences identically with PMEN1. These isolates included PSP and were dated from 1952, preceding the estimated origin of PMEN1. Since the proportion of identical coding sequences between CC66 and PMEN1 was much greater than between those of PMEN1 and the other CCs represented here, we conclude that PMEN1 was ancestrally more closely related to CC66 than to other CCs, which supports the findings of *Donati* and colleagues [34] (also see Additional file 4).

Discussion

To our knowledge, our collection represents the largest historical pneumococcal collection ever compiled and includes the greatest number of early (1960s - 1980s) PNSP in a single study. This has provided us with a unique opportunity to study pneumococcal evolution. Within our diverse isolate collection, whole or part *pbp* alleles that were identical or highly similar to those of the PMEN1 reference strain were identified from pneumococci representing a total of 24 distinct CCs. Two isolates, CGSP14 and PMEN3, shared *pbp2x, pbp1a* and *pbp2b* sequences identically with PMEN1, and shared highly similar sequences in the *pbp* flanking regions. The full extent of sharing all three of these PMEN1 *pbp* sequences has never before been uncovered (nor have the large regions around all three *pbps*), although previous authors have independently noted *pbp* similarities between CGSP14, PMEN3, PMEN2 and PMEN1 [16, 35-38]. Highly similar *pbp2x* and *pbp2b* nucleotide sequences have also been independently identified among numerous pneumococcal and viridans streptococcal clones [37, 39, 40]. Shared resistance-conferring *pbp* sequences other than those associated with PMEN1 have also been described previously [5, 41, 42] and were identified among our isolate collection, though at much lower frequency (data not shown). Additionally, data from other authors have suggested that fluoroquinolone resistance determinants have been donated from PMEN1 to numerous unrelated pneumococcal clones [43].

It is very unlikely that identical or highly similar sequences were assembled independently in different CCs, but rather that they have been transferred, as whole or part alleles, between unrelated pneumococcal clones at least once for each independent CC. Note that the establishment of defined CCs must have occurred prior to acquisition of these *pbps*, since we also identified penicillin-susceptible members of these clones (i.e. $CC15^{14}$, $CC156/162^{9A/V}$).

The predominance of the PMEN1-like *pbp* sequences within our isolate collection suggests that the PMEN1 reference *pbp2x, pbp1a and pbp2b* sequence combination may provide an evolutionary selective advantage over other resistance-conferring *pbps,* which is consistent with recent experimental data suggesting certain combinations of *pbp2x* and *pbp1a* alleles are more advantageous than others [44]. A study of genomic evolution within the PMEN1 lineage also provided evidence consistent with the hypothesis that the *pbp2x* and *pbp1a* alleles were positively selected [17]. *pbp2x* is upstream and *pbp1a* is downstream of the capsular (*cps*) locus in the pneumococcal genome. The authors noted that the boundaries of recombination events affecting the *cps* locus appeared to be restricted by the nearby *pbp2x* and *pbp1a* loci. It is known that both of these loci and the *cps* locus can be transferred simultaneously between pneumococci [45], and thus the restriction within the PMEN1 lineage is likely not entirely mechanistic.

In addition to the *pbp* loci and flanking regions, CGSP14 and PMEN3 acquired multiple genomic regions, totalling ~9.5% and ~5.3% of the genome respectively, from PMEN1 representative(s). This is consistent with two recent studies, which demonstrated that multiple fragment recombination had taken place *in vivo,* between pneumococci isolated from a single patient [46] and among vaccine escape strains isolated in the USA [47]. The suggestion that a significant proportion of the genome may be acquired from a single pneumococcal donor was recently supported mechanistically by *Attaeich* and colleagues [48]. These authors showed that cellular levels of pneumococcal SsbB protein are sufficient to maintain an intracellular pool of internalised DNA of up to 1.15 Mb of length (~50% of the genome). Such a pool of genetic material, potentially originating from a single donor cell, could then be used for successive rounds of recombination.

We cannot know for certain whether the CGSP14 and PMEN3 imported regions were acquired simultaneously in a single event or via successive recombination events. Either way, the data strongly suggest that they were acquired from a PMEN1 representative. It should also be noted that recombination events additional to those described here, and involving acquisition of DNA from other pneumococcal clones or other species, may also have altered the CGSP14 and PMEN3 genomes.

The genomic regions acquired from the PMEN1 donor may also have contributed to the virulence and host colonisation ability of the recipients. CGSP14 was multidrug-resistant and isolated from a child suffering from necrotising pneumonia complicated with haemolytic uremic syndrome [35]. PMEN3 represents another globally-distributed multidrug-resistant clone [15]. The putative recombinogenic regions acquired by both of these pneumococci included several genes associated with virulence or host colonisation potential. Interestingly, large sections of the CGSP14 transposon, which was shown previously to contain regions of similar sequence to those of the PMEN1 ICE*Sp*23FST81 element [35], were in fact identical to those of PMEN1. Two interpretations of the regions of similarity within the ICE*Sp*23FST81 element are possible. One is convergent evolution between PMEN1 and CGSP14, with both strains independently acquiring similar, recently-diverged elements. Alternatively, a more divergent element may have been acquired by CGSP14 that was subsequently modified through recombination with DNA from PMEN1. The mosaic nature of these variable conjugative elements [17] makes distinguishing these two hypotheses difficult.

Finally, we showed that one of the first reported PNSP (23F/4) represented the ancestor of PMEN1. The year of isolation of 23F/4 (1967) correlated with the predicted date of emergence of the PMEN1 lineage (~1970), although the country of isolation (Australia) did not match the predicted

region of PMEN1 origin (Europe) [17]. Despite its high overall genomic similarity to PMEN1, the 23F/4 ST, which differed from that of PMEN1, has not been identified amongst modern pneumococci. Perhaps the differences between these two genomes could explain the differences in disseminative success of these two STs. The acquisition of more favourable *pbp2x, pbp1a* and *pbp2b* alleles in the PMEN1 genome provides a partial explanation; however, other changes and/or acquisition of other loci have also likely played an important role. The PMEN1 ICE*Sp*23FST81 element, Na^+-dependent ATPase island and ϕMMI phage were absent from the 23F/4 genome. The ICE*Sp*23FST81 element carries both tetracycline and chloramphenicol resistance determinants [30] plus a *umuDC*-like gene shown to provide protection against UV damage [32]. The Na^+-dependent ATPase island and a ϕMMI phage closely related to that of PMEN1 have been associated with an increased incidence of invasive pneumococcal disease [33] and increased cell adherence capabilities [31], respectively.

Any or all of these regions may have provided PMEN1 with a selective advantage, allowing it to persist among the greater pneumococcal population for the past three decades. Alternatively, the combination of these regions with the PMEN1 reference *pbp* genes and the 23F/4 genetic background may have been the most important factor. In any case, the subsequent global dissemination of PMEN1 likely aided the spread of its *pbp* alleles and other regions of its genome to numerous unrelated pneumococci. Some of the recipient pneumococci might subsequently have transferred these PMEN1-like sequences to additional pneumococcal clones.

Recently, through comparison of multiple representatives within the PMEN1 CC, Croucher *et al.* showed that the PMEN1 lineage contains a vast amount of genetic diversity, largely resulting from horizontal acquisition of DNA via recombination [17]. Through comparison of genomic sequences

across a genetically diverse collection of pneumococci, our study has greatly increased our understanding of the emergence of the PMEN1 clone and its subsequent contribution to the genomic evolution and spread of penicillin resistance determinants among other pneumococci. It is well understood that many bacterial species, particularly the pneumococcus, frequently undergo horizontal gene transfer; however, the directional genetic promiscuity described here, from a single epidemiologically successful clone to numerous unrelated clones, has not been demonstrated before in pneumococci, or other bacterial species as far as we are aware.

Antibiotics are among the most influential global public health successes, but impose selective pressures that drive bacterial genomic evolution. PMEN1 is an excellent example of a bacterium that has become resistant to multiple antibiotics and that has evolved to become very successful in colonisation, transmission, and causing disease. Moreover, PMEN1 has subsequently shared its advantageous DNA with other unrelated pneumococci. Studies such as this one may help to identify ways to counteract these biological changes and preserve the value of antibiotics for the future.

Materials and methods

Strain collection

A unique collection of 426 global PSP and PNSP dated 1937 - 2007 was studied, as summarised in Table 1. Isolates included one of the first PNSP reported from Australia [6], the first high-level PNSP reported from South Africa [8, 9] and the 43 PMEN reference strains [15]. 211 pneumococci dated 1937 - 1996 were provided by the Statens Serum Institut, Denmark. These isolates included both PSP (n = 203), PNSP (n = 7) and an isolate of unknown penicillin susceptibility (n = 1). Following the analyses of MLST and *pbp* sequence data from these isolates, the 43 PMEN clones (including both PSP (n = 18) and PNSP (n = 25)) and additional modern (1990s - 2000s) PSP (n = 38) and PNSP (n = 9) that had been recently genotyped by A. Brueggemann / L. McGee and represented CCs of interest, were selected for inclusion. In addition, a literature search was performed to identify studies that reported early PNSP. The authors of those papers were contacted and isolates were requested; the majority of isolates were still available and were added to our collection. The available isolates comprised PNSP dated 1969 and 1972 from Papua New Guinea (n = 2), PSP (n = 8) and PNSP (n = 71) dated 1977 - 1989 from South Africa, PSP (n = 1) and PNSP (n = 26) dated 1981 - 1991 from the USA, PNSP dated 1987 - 1989 (n = 9) from Spain and PNSP dated 1986 - 1994 from Germany (n = 8). Finally, genomic sequence data for 16 isolates was retrieved from the Genbank database. These isolates included both PSP (n = 5), PNSP (n = 4) and 7 pneumococci of unknown penicillin susceptibility. Serotype and penicillin susceptibility data are summarised by decade in Additional file 6. Where not previously published, isolates were serotyped by Quellung reaction and penicillin minimum inhibitory concentrations were obtained by Etest (bioMérieux) as per manufacturer's guidelines.

DNA extraction

Single pneumococcal colonies were cultured on Columbia agar with sheep blood (Oxoid) and incubated overnight at 37°C + 5% CO_2. 3 - 4 sweeps of growth were subcultured onto tryptic soy agar (Oxoid) + 1000 Uml^{-1} catalase (Oxoid) and spread with a swab before further overnight incubation at 37°C + 5% CO_2 (at Oxford); or cultured in serum broth for 7 hrs (at SSI). Pneumococcal growth was suspended in 1 ml phosphate buffered saline (pH 7.4, Fisher Bioreagents) and centrifuged at 7,600 rpm for 10 min. Extractions were completed using the Qiagen DNeasy kit as per manufacturer's instructions.

MLST and clonal complex (CC) designation

Where not previously published (n = 352) multilocus sequence types were determined for all isolates [49]. MLST data for all of our isolates and those deposited in the pneumococcal MLST database [50] were analysed by the goeBURST [51] method to predict CC group and sub-group founders. Isolates were assigned to CCs so that only single/double locus variants of the group founder and any additional single locus variants of large (n ≥ 5 STs) sub-group founders were included. CCs were named according to the predicted group founder(s) ST(s). When goeBURST could not distinguish a group founder the group was designated NoneX, where X was the ST of lowest numerical value in the group. When isolates could not be assigned to a CC due to a lack of closely related STs, they were designated SingletonX, where X was the isolate ST.

pbp profiling

pbp2x (1,673 bp), *pbp1a* (901 bp) and *pbp2b* (1,272 bp) nucleotide sequences for 389 selected isolates, were retrieved from Genbank (n = 36 sequences) or determined by conventional PCR and Sanger sequencing (n = 1,131 sequences). 174/211 isolates from the Statens Serum Institut were

chosen for *pbp* nucleotide sequencing because they represented a wide range of serotypes, STs and isolation dates. All isolates from each of the other sources were also included. PCR reactions containing 1 µl DNA extract, 13.5 µl Hotstar Taq DNA Polymerase master mix (Qiagen), 0.5 µl each of the forward and reverse primers (100 µM) (see Additional file 7) were prepared to a final volume of 25 µl. Cycling parameters were as follows: Initial denaturation of 95°C for 180 s, followed by 35 (40 for *pbp2x*) cycles of 95°C for 30 s, 55°C (58°C for *pbp1a*) for 30 s and 72°C for 60 s followed by a final extension period of 72°C for 10 min. Products were precipitated by 20% polyethylene glycolate/2.5M NaCl, washed in 70% ETOH and rehydrated in 10 - 20 µl nuclease-free H_2O (Sigma). Sequencing reactions containing 2 µl rehydrated PCR product, 1.75 µl 5x sequencing buffer, 4 µl primer (1 µM) (see Additional file 7) and 0.5 µl Taq FS (Big Dye) were prepared in a final volume of 10 µl. Cycling parameters were as follows: 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 2 min. Sequencing products were precipitated with 5 M sodium acetate (Sigma) solution, washed with 70% ETOH and subsequently separated and detected with an ABI Prism 3730 automated DNA sequencer (Applied Biosystems). Sequences were aligned by MUSCLE [52] and imported to MEGA5 [53] for visual comparison. Unique *pbp* sequences were assigned unique allele numbers.

Whole genome sequencing

96 isolates were selected for whole genome sequencing (see Additional file 3) using the following criteria:

1) Isolates that were not closely related to PMEN1 based on MLST data but shared identical *pbp* alleles or partial *pbp* alleles to those of the PMEN1-reference strain. The selected isolates were distributed through time, by geographic location and by genotype in order to capture the greatest possible amount of genetic diversity.

2) Isolates representing CCs of interest (e.g. major international CCs) for which both historical and modern isolate representatives were available. Inclusion of the historical representatives was crucial, as it allowed us to differentiate identical/highly similar sequence imports obtained exogenously versus identical sequence due to common ancestry. Within these CCs the selected isolates were distributed through time, by geographic location, by serotype and by *pbp* allele combination, in order to capture the greatest possible amount of genetic diversity.

DNA was extracted to a final concentration of 20 ng/µl as described above. Sequencing was via the Illumina platform as previously described [17]. As per standard protocols, multiplex library construction with a 200 bp insertion size was followed by 54 nucleotide paired end sequencing. Illumina reads were assembled to contigs using the Velvet software [54]. Data for 7 genomes were rejected based on technical problems. K-mer lengths were optimised to the maximum length achieving an average of ≥20x read coverage across assemblies. Resulting contigs were between 41 bp and 328,742 bp length (mean = 3,797 bp; see Additional file 4 for further details). Contigs were deposited in a BIGSdb database [28] and in the European Nucleotide Archive (see Additional file 8 for accession numbers). Contigs were ordered to the PMEN1 reference genome using ABACAS [55] and viewed with the Artemis Comparison Tool [29], allowing for the extraction of selected sequence regions spanning >1 contig.

Identification of recombination fragments from PMEN1

The BIGSdb Genome Comparator tool, which utilises the BLAST algorithm, was used to compare PMEN1 coding sequences to those of CGSP14, PMEN3 and their respective predicted ancestral representatives (Table 2). Comparable coding sequences in the query genomes were identified by

BLASTn search if they shared ≥70% nucleotide sequence identity (to allow for the identification of variable sequences among conserved protein classes [28] and to avoid bias towards sequence variants specific to PMEN1) and could be aligned at 100% length (to ensure that any coding sequences identified as identical to PMEN1 were indeed identical across their entire length). Potential recombination regions were indicated by coding sequences shared identically by PMEN1 and CGSP14 or PMEN3, but not by any of the respective ancestral representatives. Sequence regions spanning the coding sequences of interest plus 5 kb flanking either side were extracted from the PMEN1, CGSP14, PMEN3 and ancestral representative genomes, aligned and imported into MEGA5 [53] to produce variable site maps for visual inspection (see Additional file 9). Recombination regions were classified as the maximum length of sequence over which CGSP14 or PMEN3 were identical to PMEN1 but different from their ancestral representatives.

Identification of PMEN1 ancestral representatives

Genome comparator was used to screen 104 isolates (see Additional file 3) representing 34 CCs and 31 serotypes for coding sequences identical to those of the PMEN1 reference (parameters as above). 23F/4 was identified as sharing a high percentage of coding sequences identically with PMEN1 and was thus subjected to further analysis. PMEN1 coding sequences which differed in 23F/4 were extracted from the PMEN1 whole genome sequence and BLASTed against the 23F/4 Velvet contigs in the BIGSdb. Sequences were aligned and percentage nucleotide difference was calculated. 23F/4 contigs were ordered against the PMEN1 chromosome by ABACAS [55] and viewed with the Artemis Comparison Tool [29] to identify genomic regions present in PMEN1 but absent from 23F/4.

Abbreviations

ABACAS, Algorithm Based Automatic Contiguation of Assembled Sequences; BIGSdb, Bacterial Isolate Genome Sequence Database; BLAST, Basic Local Alignment Search Tool; BLASTn, nucleotide-nucleotide BLAST; bp, base pair(s); CC, clonal complex; *cps*, capsular; ETOH, ethanol; goeBURST, global optimal enhanced Based Upon Related Sequence Types; kb, kilo base pairs; k-mer, a short oligonucleotide of length k; Mb, mega base pairs; MEGA5, Molecular Evolutionary Genetics Analysis v. 5; MLST, multilocus sequence typing; MUSCLE, Multiple Accurate and Fast Sequence Comparison; Na, sodium; NaCl, sodium chloride; ng, nanogram; *pbp*, penicillin-binding protein; PCR, polymerase chain reaction; PCV7, 7-valent pneumococcal conjugate vaccine; PMEN, Pneumococcal Molecular Epidemiology Network; PNSP, penicillin-nonsusceptible pneumococci; PSP, penicillin-susceptible pneumococci; ST, sequence type; Uml^{-1}, units per millilitre; USA, United States of America.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KLW and ABB designed the study, analysed data, wrote the paper; LML, LM, AVG, JL, MRJ, KGK, BWB, KPK, RH and ABB provided bacterial isolates; KLW, LM, JL, BWB, RH and ABB generated *pbp* and/or MLST sequence data; LML, AVG, JL, KGK, BWB, KPK and RH provided serotype data; KLW, LML, LM, AVG and RH performed susceptibility testing; KLW and LML extracted genomic DNA for whole genome sequencing; JP and SDP provided whole genome sequencing data; KLW and NJC assembled whole genome sequence data. All authors discussed the results, read and approved the manuscript for publication.

Acknowledgements

We thank Dalia Denapaite, Irma Ochigava, and Shwan Rachid for DNA sequencing and MLST analysis, and Roberta Fisher for comments on the manuscript. We also gratefully acknowledge the clinicians, microbiologists, and investigators of the Active Bacterial Core surveillance program of the Emerging Infections Program Network, USA.

References

1. World Health Organisation: **Pneumococcal conjugate vaccine for childhood immunisation - WHO position paper.** *Wkly Epidemiol Rec* 2007, **82:**93-104.
2. Fedson DS, Anthony J, Scott G: **The burden of pneumococcal disease among adults in developed and developing countries: what is and is not known.** *Vaccine* 1999, **17:**S11-S18.
3. Hakenbeck R, Grebe T, Zahner D, Stock JB: **β-lactam resistance in *Streptococcus pneumoniae*: penicillin-binding proteins and non-penicillin-binding proteins.** *Mol Microbiol* 1999, **33:**673-678.
4. Lambertsen L, Brendstrup M, Friis H, Christensen JJ: **Molecular characterisation of invasive penicillin non-susceptible *Streptococcus pneumoniae* from Denmark, 2001 to 2005.** *Scand J Infect Dis* 2010, **42:**333-340.
5. Dowson CG, Coffey TJ, Kell CM, Whiley RA: **Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*.** *Mol Microbiol* 1993, **9:**635-643.
6. Hansman D, Bullen MM: **A resistant pneumococcus.** *Lancet* 1967, **2:**264-265.
7. Kislak JW, Razavi LMB, Daly AK, Finland M: **Susceptibility of pneumococci to nine antibiotics.** *Am J Med Sci* 1965, **250:**261-268.
8. Appelbaum PC, Scragg JN, Bowen AJ, Bhamjee A, Hallett AF, Cooper RC: ***Streptococcus pneumoniae* resistant to penicillin and chloramphenicol.** *Lancet* 1977, **2:**995-997.
9. Jacobs MR, Koornhof HJ, Robins-Browne RM, Stevenson CM, Vermaak ZA, Freiman I, Miller B, Witcomb MA, Isaäcson M, Ward JI, Austrian R: **Emergence of multiply resistant pneumococci.** *N Engl J Med* 1978, **299:**735-740.
10. Liñares J, Ardanuy C, Pallarés R, Fenoll A: **Changes in antimicrobial resistance, serotypes and genotypes in *Streptococcus pneumoniae* over a 30-year period.** *Clin Microbiol Infect* 2010, **16:**402-410.
11. Feikin DR, Klugman KP: **Historical changes in pneumococcal serogroup distribution: Implications for the era of pneumococcal conjugate vaccines.** *Clin Infect Dis* 2002, **35:**547-555.
12. Whitney CG, Farley MM, Hadler J, Harrison LH, Bennett NM, Lynfield R, Reingold A, Cieslak PR, Pilishvili T, Jackson D, Facklam RR, Jorgensen JH, Schuchat A: **Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine.** *N Engl J Med* 2003, **348:**1737-1746.
13. Weinberger DM, Malley R, Lipsitch M: **Serotype replacement in disease after pneumococcal vaccination.** *Lancet* 2011, **378:**1962-1973.

14. Beall BW, Gertz RE, Hulkower RL, Whitney CG, Moore MR, Brueggemann AB: **Shifting genetic structure of invasive serotype 19A pneumococci in the United States.** *J Infect Dis* 2011, **203:**1360-1368.
15. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, Hakenbeck R, Hryniewicz W, Lefévre JC, Tomasz A, Klugman KP: **Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network.** *J Clin Microbiol* 2001, **39:**2565-2571.
16. Coffey TJ, Dowson CG, Daniels M, Zhou J, Martin C, Spratt BG, Musser JM: **Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthesis genes, in natural populations of *Streptococcus pneumoniae*.** *Mol Microbiol* 1991, **5:**2255-2260.
17. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD: **Rapid pneumococcal evolution in response to clinical interventions.** *Science* 2011, **331:**430-434.
18. Muñoz R, Coffey TJ, Daniels M, Dowson CG, Laible G, Casal J, Hakenbeck R, Jacobs M, Musser JM, Spratt BG, Tomasz A: **Intercontinental spread of a multiresistant clone of serotype 23F *Streptococcus pneumoniae.*** *J Infect Dis* 1991, **164:**302-306.
19. Klugman KP, Coffey TJ, Smith A, Wasas A, Meyers M, Spratt BG: **Cluster of an erythromycin-resistant variant of the Spanish multiply resistant 23F clone of *Streptococcus pneumoniae* in South Africa.** *Eur J Clin Microbiol Infect Dis* 1994, **13:**171-174.
20. Sibold C, Wang JF, Henrichsen J, Hakenbeck R: **Genetic relatedness of penicillin-susceptible and -resistant *Streptococcus pneumoniae* strains isolated on different continents.** *Infect Immun* 1992, **60:**4119-4126.
21. Reichmann P, Varon E, Günther E, Reinert RR, Lüttiken R, Marton A, Geslin P, Wagner J, Hakenbeck R: **Penicillin-resistant *Streptococcus pneumoniae* in Germany: genetic relationship to clones from other European countries.** *J Med Microbiol* 1995, **43:**377-385.
22. Tomasz A, Corso A, Members of the PAHO/Rockefeller University Workshop:, Severina EP, Echániz-Aviles G, de Cunto Brandileone MC, Camou T, Castañeda E, Figueroa O, Rossi A, di Fabio JL: **Molecular epidemiologic characterisation of penicillin-resistant *Streptococcus pneumoniae* invasive paediatric isolates recovered in six Latin-American countries: An overview.** *Microb Drug Resist* 1998, **4:**195-207.
23. Corso A, Severina EP, Petruk VF, Mauriz YR, Tomasz A: **Molecular characterisation of penicillin-resistant *Streptococcus pneumoniae* isolates causing respiratory disease in the United States.** *Microb Drug Resist* 1998, **4:**325-337.
24. Muñoz-Elías EJ, Marcano J, Camilli A: **Isolation of *Streptococcus pneumoniae* biofilm mutants and their characterization during nasopharyngeal colonization.** *Infect Immun* 2008, **76:**5049-5061.
25. Brown JS, Gilliland SM, Spratt BG, Holden DW: **A locus contained within a variable region of pneumococcal pathogenicity island 1 contributes to virulence in mice.** *Infect Immun* 2004, **72:**1587-1593.
26. Polissi A, Pontiggia A, Feger G, Altieri M, Mottl H, Ferrari L, Simon D: **Large-scale identification of virulence genes from *Streptococcus pneumoniae.*** *Infect Immun* 1998, **66:**5620-5629.
27. Lau GW, Haataja S, Lonetto M, Kensit SE, Marra A, Bryant AP, McDevitt D, Morrison DA, Holden DW: **A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence.** *Mol Microbiol* 2001, **40:**555-571.

28. Jolley KA, Maiden MC: **BIGSdb: Scalable analysis of bacterial genome variation at the population level.** *BMC Bioinformatics* 2010, **10:**595.
29. Carver TJ, Rutherford KM, Berriman M, MA R, Barrell BG, Parkhill J: **ACT: the Artemis comparison tool.** *Bioinformatics* 2005, **21:**3422-3423.
30. Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, Mitchell AM, Quail MA, Andrew PW, Parkhill J, Bentley SD, Mitchell TJ: **Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae*^{Spain23F} ST81.** *J Bacteriol* 2009, **191:**1480-1489.
31. Loeffler JM, Fischetti VA: **Lysogeny of *Streptococcus pneumoniae* with MM1 phage: Improved adherence and other phenotypic changes.** *Infect Immun* 2006, **74:**4486-4495.
32. Munoz-Najar U, Vijayakumar MN: **An operon that confers UV resistance by evoking the SOS mutagenic response in streptococcal conjugative transposon Tn*5252.*** *J Bacteriol* 1999, **181:**2782-2788.
33. Obert C, Sublett J, Kaushal D, Hinojosa E, Barton T, Tuomanen EI, Orihuela CJ: **Identification of a candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease.** *Infect Immun* 2006, **74:**4766-4777.
34. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Dunning Hotopp JC, Hu FZ, Riley DR, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER, Masignani V: **Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species.** *Genome Biol* 2010, **11:**R107.
35. Ding F, Tang P, Hsu M-H, Cui P, Hu S, Yu J, Chiu C-H: **Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14.** *BMC Genet* 2009, **10:**158.
36. Martin C, Sibold C, Hakenbeck R: **Relatedness of penicillin-binding protein 1a genes from different clones of *Streptococcus pneumoniae* isolated in South Africa and Spain.** *EMBO J* 1992, **11:**3831-3836.
37. Sibold C, Henrichsen J, Konig A, Martin C, Chalkley L, Hakenbeck R: **Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*.** *Mol Microbiol* 1994, **12:**1013-1023.
38. Stanhope MJ, Lefébure T, Walsh SL, Becker JA, Lang P, Pavinski Bitar PD, Miller LA, Italia MJ, Amrine-Madsen H: **Positive selection in penicillin-binding proteins 1a, 2b, and 2x from *Streptococcus pneumoniae* and its correlation with amoxicillin resistance development.** *Infect Genet Evol* 2008, **8:**331-339.
39. Chi F, Nolte O, Bergmann C, Ip M, Hakenbeck R: **Crossing the barrier: evolution and spread of a major class of mosaic *pbp2x* in *Streptococcus pneumoniae*, *S. mitis* and *S. oralis*.** *Int J Med Microbiol* 2007, **297:**503-512.
40. Dowson CG, Hutchison A, Woodford N, Johnson AP, George RC, Spratt BG: **Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of *Streptococcus pneumoniae*. .** *Proc Natl Acad Sci U S A* 1990, **87:**5858-5862.
41. Dowson CG, Hutchison A, Brannigan JA, George RC, David H, Liñares J, Tomasz A, Smith JM, Spratt BG: **Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*.** *Proc Natl Acad Sci U S A* 1989, **86:**8842-8846.
42. Stanhope MJ, Walsh SL, Becker JA, Miller LA, Lefébure T, Lang P, Pavinski Bitar PD, Amrine-Madsen H: **The relative frequency of intraspecific lateral gene transfer of penicillin binding proteins la, 2b, and 2x, in amoxicillin resistant *Streptococcus pneumoniae.*** *Infect Genet Evol* 2007, **7:**520-534.

43. Stanhope MJ, Walsh SL, Becker JA, Italia MJ, Ingraham KA, Gwynn MN, Mathie T, Poupard JA, Miller LA, Brown JR, Amrine-Madsen H: **Molecular evolution perspectives on intraspecific lateral DNA transfer of topoisomerase and gyrase loci in *Streptococcus pneumoniae*, with implications for fluoroquinolone resistance development and spread.** *Antimicrob Agents Chemother* 2005, **49:**4315-4326.
44. Zerfass I, Hakenbeck R, Denapaite D: **An important site in PBP2x of penicillin-resistant clinical isolates of *Streptococcus pneumoniae*: mutational analysis of Thr338.** *Antimicrob Agents Chemother* 2009, **53:**1107-1115.
45. Brueggemann AB, Pai R, Crook D, Beall B: **Vaccine escape recombinants emerge after pneumococcal vaccination in the United States.** *PLoS Pathog* 2007, **3:**e168.
46. Hiller NL, Ahmed A, Powell E, Martin DP, Eutsey R, Earl J, Janto B, Boissy RJ, Hogg J, Barbadora K, Sampath R, Lonergan S, Post JC, Hu FZ, Ehrlich GD: **Generation of genetic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal paediatric infection.** *PLoS Pathog* 2010, **6:**e1001108.
47. Golubchik T, Brueggemann AB, Street T, Gertz Jr RE, Spencer CCA, Ho T, Giannoulatou E, Link-Gelles R, Harding RM, Beall B, Peto TEA, Moore MR, Donnelly P, Crook DW, Bowden R: **Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event.** *Nat Genet* 2012, **44:**352-355.
48. Attaiech L, Olivier A, Mortier-Barrière I, Soulet AL, Granadel C, Martin B, Polard P, P CJ: **Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity.** *PLoS Genet* 2011, **7:**e1002156.
49. Enright MC, Spratt BG: **A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease.** *Microbiology* 1998, **144:**3049-3060.
50. Multi Locus Sequence Typing [www.mlst.net].
51. Francisco AP, Bugalho M, Ramirez M, Carriço JA: **Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach.** *BMC Bioinformatics* 2009, **10:**152.
52. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32:**1792-1797.
53. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28:**2731-2739.
54. Zerbino DR, Birney E: **Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18:**821-829.
55. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M: **ABACAS: algorithm-based automatic contiguation of assembled sequences.** *Bioinformatics* 2009, **25:**1968-1969.
56. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kaltoft MS, Barrell B, Reeves PR, Parkhill J, Spratt BG: **Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes.** *PLoS Genet* 2006, **2:**e31.
57. Hakenbeck R, Briese T, Chalkley L, Ellerbrok H, Kalliokoski R, Latorre C, Leinonen M, Martin C: **Variability of penicillin-binding proteins from penicillin-sensitive *Streptococcus pneumoniae*.** *J Infect Dis* 1991, **164:**307-312.
58. Hakenbeck R, Briese T, Chalkley L, Ellerbrok H, Kalliokoski R, Latorre C, Leinonen M, Martin C: **Antigenic variation of penicillin-binding proteins from penicillin-resistant clinical strains of *Streptococcus pneumoniae.*** *J Infect Dis* 1991, **164:**313-319.

59. McDougal LK, Rasheed JK, Biddle JW, Tenover FC: **Identification of multiple clones of extended-spectrum cephalosporin-resistant *Streptococcus pneumoniae* isolates in the United States.** *Antimicrob Agents Chemother* 1995, **39:**2282-2288.
60. Reinert RR, R JM, Appelbaum PC, Bajaksouzian S, Cordeiro S, van der Linden M, Al-Lahham A: **Relationship between the original multiply resistant South African isolates of *Streptococcus pneumoniae* from 1977 to 1978 and contemporary international resistant clones.** *J Clin Microbiol* 2005, **43:**6035-6041.
61. Fernández A, Cabellos C, Tubau F, Liñares J, Viladrich PF, Gudiol F: **Relationship between penicillin and cephalosporin resistance of *Streptococcus pneumoniae* strains and its inflammatory activity in the experimental model of meningitis.** *Med Microbiol Immunol* 2001, **190:**135-138.
62. Muñoz R, Dowson CG, Daniels M, Coffey TJ, Martin C, Hakenbeck R, Spratt BG: **Genetics of resistance to third-generation cephalosporins in clinical isolates of *Streptococcus pneumoniae*.** *Mol Microbiol* 1992, **6:**2461-2465.
63. Pallarés R, Liñares J, Vadillo M, Cabellos C, Manresa F, Viladrich PF, Martin R, Gudiol F: **Resistance to penicillin and cephalosporin and mortality from severe pneumococcal pneumonia in Barcelona, Spain.** *N Engl J Med* 1995, **333:**474-480.
64. Viladrich PF, Cabellos C, Pallarés R, Tubau F, Martinez-Lacasa J, Liñares J, Gudiol F: **High doses of cefotaxime in treatment of adult meningitis due to *Streptococcus pneumoniae* with decreased susceptibilities to broad spectrum cephalosporins.** *Antimicrob Agents Chemother* 1996, **40:**218-220.
65. Viladrich PF, Gudiol F, Liñares J, Pallarés R, Sabaté I, Rufí G, Ariza J: **Evaluation of vancomycin for therapy of adult pneumococal meningitis.** *Antimicrob Agents Chemother* 1991, **35:**2467-2472.

Figure legends

Figure 1. *pbp* loci and flanking regions of the PMEN1 reference, CGSP14 and PMEN3 reference genomes.

Red blocks indicate identical nucleotide sequence with substitutions or single insertions marked by black lines. The relative positions of the *pbp2x, pbp1a,* and *pbp2b* loci are shown, as are the two loci (*dexB* and *aliA*) flanking the capsular (*cps)* locus.

Figure 2. PMEN1 reference genome regions putatively acquired by CGSP14 and the PMEN3 reference.

Putative recombination regions marked by colored blocks and mapped to PMEN1 reference genome coordinates. Red blocks: regions identical or highly similar to that of the PMEN1 reference only. Gray blocks: regions identical or highly similar to those of the PMEN1 reference *and* one or more CC66 representative(s). Single nucleotide substitutions or insertions are marked with black lines. Origin of replication marked as '0'.

Figure 3: Artemis Comparison Tool comparison of the PMEN1 reference and 23F/4 genomes.

Red lines indicate homologous nucleotide sequence matches. Blue lines indicate reverse complement nucleotide sequence matches. Black arrows indicate large regions present in the PMEN1 reference genome but absent from the 23F/4 genome: (1) Na^+-dependent ATPase island; (2) ICE*Sp*23FST81 element; and (3) ϕMMI phage.

Figure 4: Percentage PMEN1 reference coding sequences identified identically within 104 pneumococcal genomes.

Points represent independent pneumococcal isolates. Isolates representing CCs for which n≥3 are labelled by CC. Isolates representing CCs for which n<3 are labelled as 'Miscellaneous CCs'. Black line represents mean percentage identical coding sequences ± 2 standard deviations (gray lines). Isolates of interest are labelled: (1) CGSP14; (2) USA18; (3) CDC2088-04; (4) PMEN3; (5) CDC3059-06; and (6) 23F/4 (also see Additional file 4).

Table 1. Isolates included in this study

Number of isolates ([a])	Years	Country (n)	Source	Reference(s)
211 (174)	1937 - 1996	Denmark (139), USA (48), Germany (6), France (2), India (2), Malaysia (2), Singapore (2), Australia (1), Belgium (1), The Gambia (1), The Netherlands (1), Norway (1), Scotland (1), Spain (1), Sweden (1), Unknown (2)	LML	[56]
43 (43)	1948 - 1998	The Netherlands (6), USA (6), Portugal (5), South Africa (3), Spain (3), CSR (2), Greece (2), Poland (2), Taiwan (2), Canada (1), Columbia (1), Denmark (1), England (1), Finland (1), France (1), Hungary (1), Sweden (1), Unknown (4)	LM	[15]
36 (36)	1979 - 1989	South Africa (36)	KPK, AVG	--
27 (27)	1969 - 1994	South Africa (17), Germany (8), Papua New Guinea (2)	RH	[57, 58]
27 (27)	1981 - 1991	USA (26), South Africa (1)	LM	[59]
26 (26)	1977 - 1978	South Africa (26)	MRJ	[8, 9, 60]
22 (22)	1999 – 2007[b]	USA (22)	LM, BWB	--
17 (17)	1993 - 2005	Iceland (17)	KGK	--
9 (9)	1987 - 1989	Spain (9)	JL	[61-65]
8 (8)	1998 - 2002	USA (8)	ABB	--

Total number of isolates = 426 (excluding isolates for which genomic data was retrieved from Genbank; see Materials and Methods). [a]Number of isolates selected for *pbp* nucleotide sequencing (total n = 389). [b]The year of one isolate was unknown. Source abbreviations: LML = L. M. Lambertsen, LM = L. McGee, KPK = K. P. Klugman, AVG = A. von Gottberg, RH = R. Hakenbeck, MRJ = M. R. Jacobs, BWB = B. W. Beall, KGK = K. G. Kristinsson, JL = J. Liñares, ABB = A. B. Brueggemann.

Table 2. Pneumococcal isolates discussed in the text

Isolate[a]	Other name(s)	Year	Serotype	Sequence type	Clonal complex	Penicillin susceptibility[b]
PMEN1	ATCC 700669	1984	23F	81	81	I
CGSP14		2004-2005	14	15	15	R
14/5		1967	14	15	15	S
PMEN9	ATCC 700676	1993	14	9	15	S
ICE13		1998	14	9	15	S
ICE570		2002	14	9	15	S
PMEN3	ATCC 700671; SP195	1993	9V	156	156/162	I
9A/1		1962	9A	312	156/162	S
9V/5		1991	9V	162	156/162	S
9V/6		1994	9V	162	156/162	S
PMEN2	ATCC 700670	1988	6B	90	90	I
23F/4		1967	23F	7184	Singleton7184	I
PMEN18	ATCC BAA-340	1997	14	67	66	R
CDC3059-06		2005	19A	199	199	I
CDC0288-04		Unknown	12F	220	218	S
USA18		1999	12F	218	218	S
TIGR4		1991	4	205	205	unk

[a]PMEN refers to Pneumococcal Molecular Epidemiology Network reference strains. [b]S = MIC ≤0.06 μg/ml; I = MIC 0.12 - 1 μg/ml; R = MIC ≥2.0 μg/ml; unk = unknown.

Additional files

Additional file 1: Tables S1 - 3 list the *pbp* allele sequence regions that were identical or highly similar to those of the PMEN1 reference strain.

Additional file 2: Tables S4 and S5 detail the regions of the CGSP14 and PMEN3 genomes, respectively, that were identified as putatively recombinogenic and acquired from PMEN1.

Additional file 3: Table S6 gives more information (e.g. serotype and penicillin susceptibility) about the isolates for which whole-genome sequences were analysed in this study.

Additional file 4: Supporting Results and Methods give more information about MLST and *pbp* diversity, the imported regions of CGSP14 and PMEN3, coding sequences found in both PMEN1 and isolates of CC66, outlier isolates highlighted in Figure 4 and methodological details related to whole-genome sequence assemblies.

Additional file 5: Figure S2 depicts the results of a Genome Comparator analysis of 104 pneumococcal genomes compared to the TIGR4 reference genome.

Additional file 6: Table S7 summarises the 442 isolates included in this study, stratified by era of isolation.

Additional file 7: Table S8 details the primers used for PCR and conventional sequencing of regions of *pbp2x*, *pbp1a* and *pbp2b*.

Additional file 8: Table S9 details metadata for the 442 isolates included in this study.

Additional file 9: Figure S3 shows a variable site map of one PMEN3 putative imported region and flanking regions.

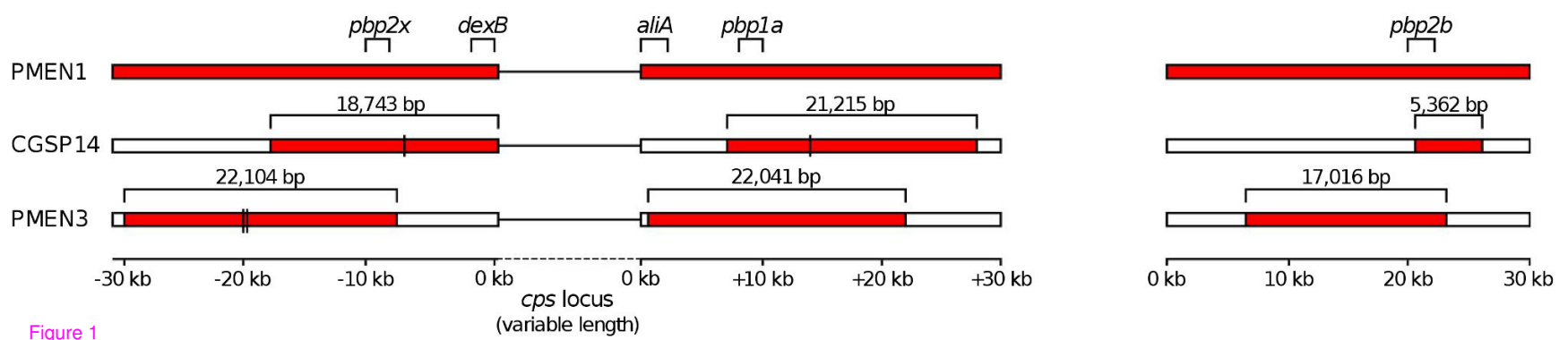
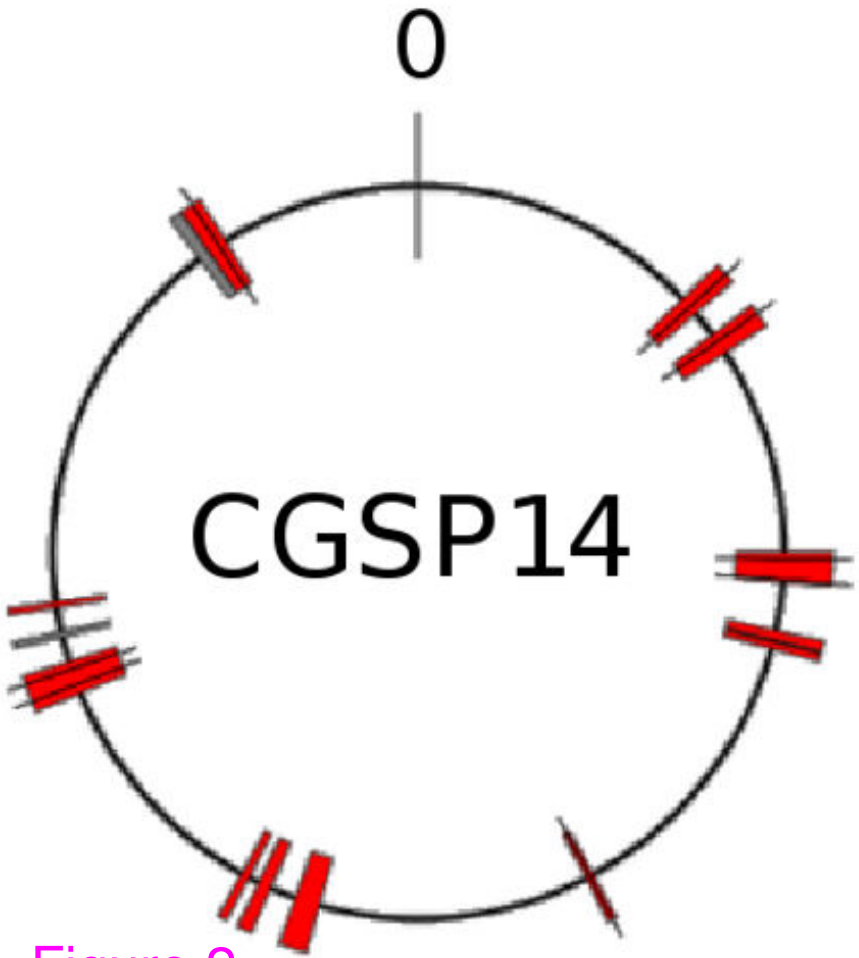

pbp2x
dexB
aliA
pbp1a
pbp2b
PMEN1
CGSP14
PMEN3
18,743 bp
21,215 bp
5,362 bp
22,104 bp
22,041 bp
17,016 bp
-30 kb
-20 kb
-10 kb
0 kb
0 kb
+10 kb
+20 kb
+30 kb
0 kb
10 kb
20 kb
30 kb
cps locus
(variable length)

Figure 1



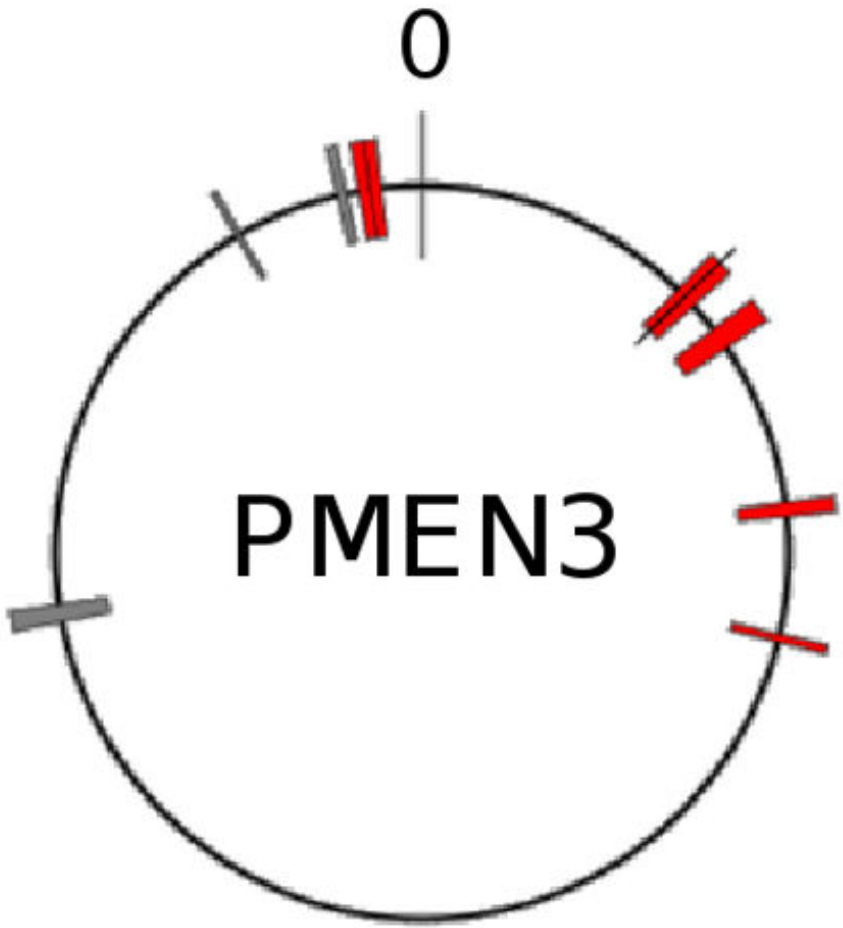


Figure 2

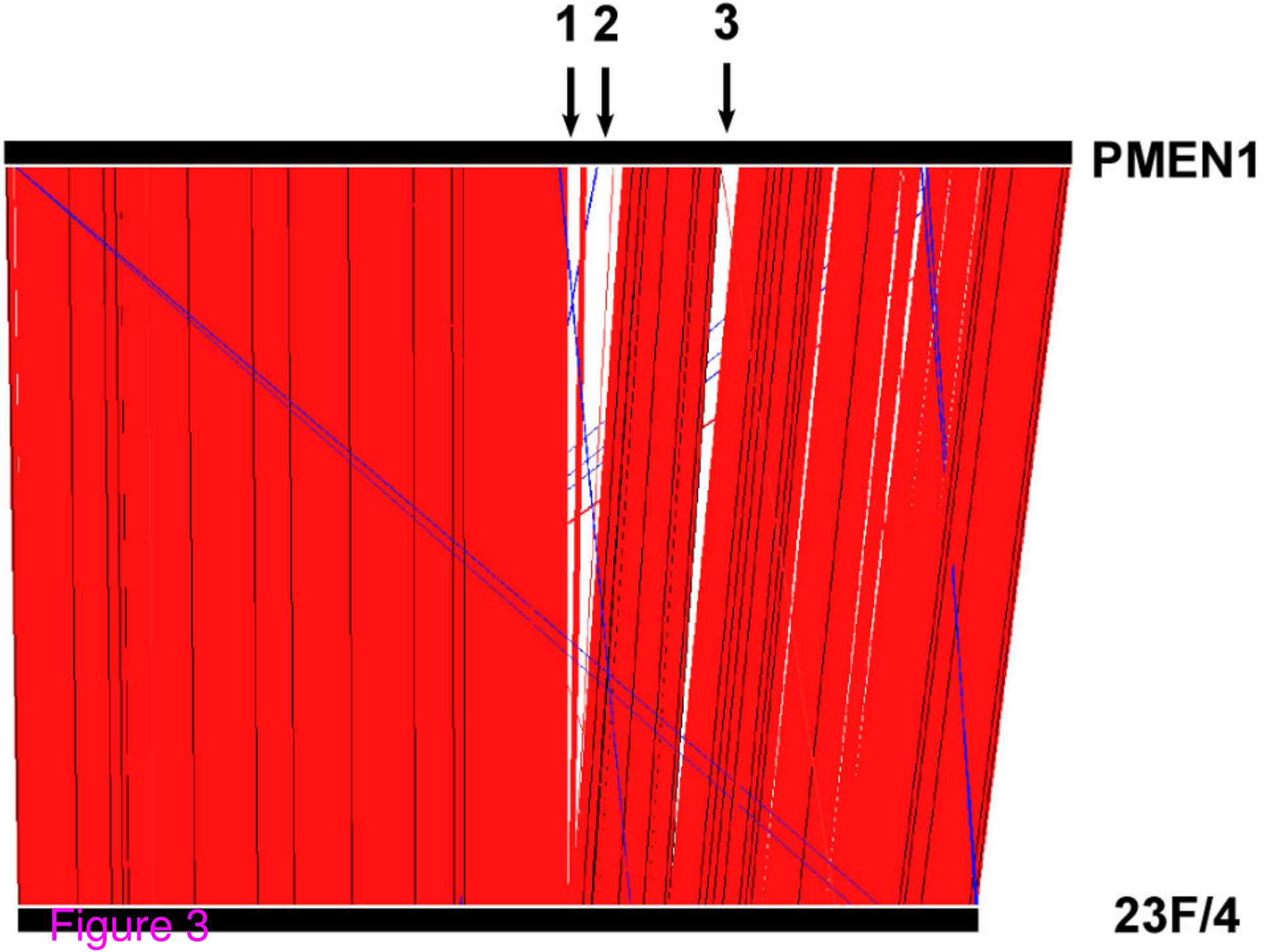


Figure 3

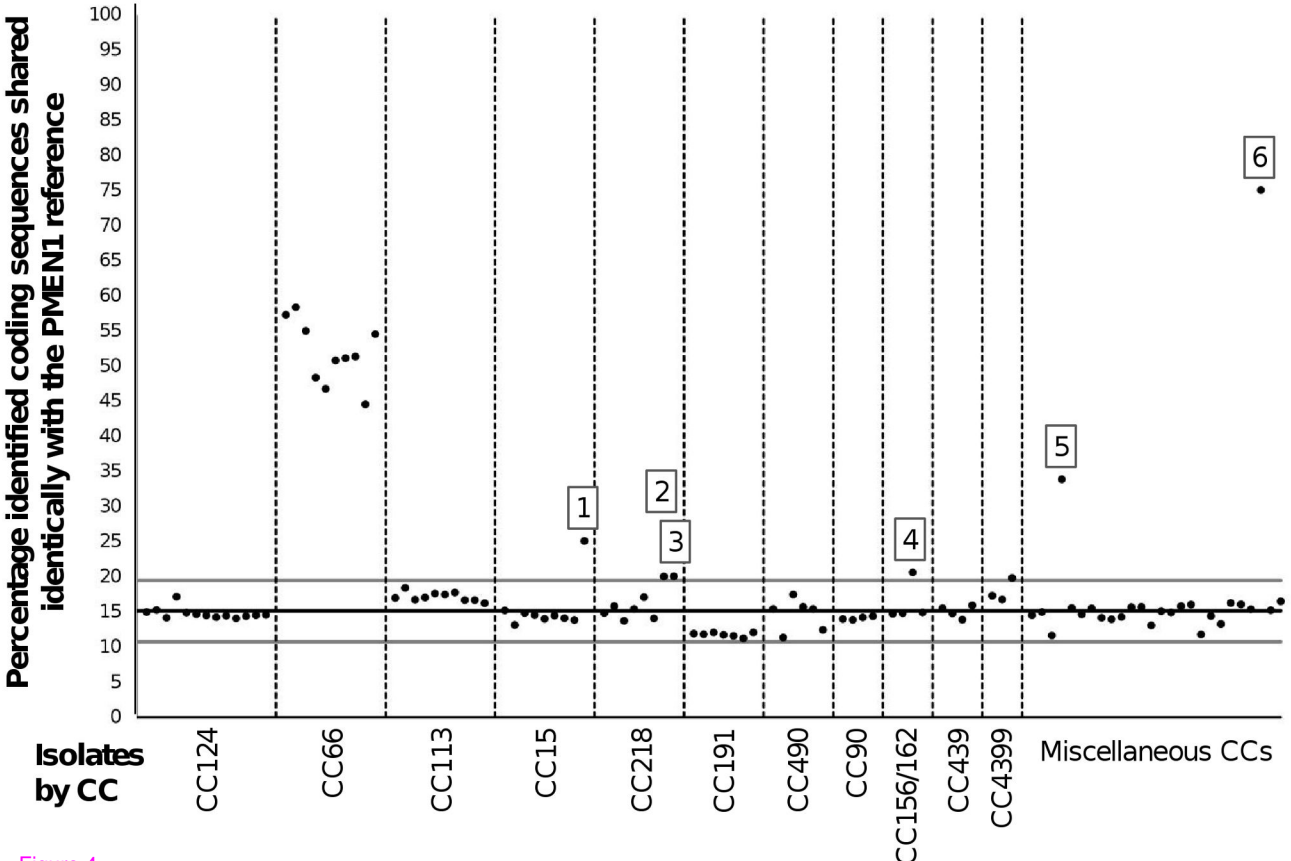

Percentage identified coding sequences shared identically with the PMEN1 reference
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
Isolates by CC
CC124
CC66
CC113
CC15
CC218
CC191
CC490
CC90
CC156/162
CC439
CC4399
Miscellaneous CCs

Figure 4

Additional files provided with this submission:

Additional file 1: Wyres_AdditionalFile_1.pdf.pdf, 284K
http://genomebiology.com/imedia/2098619944792320/supp1.pdf
Additional file 2: Wyres_AdditionalFile_2.pdf.pdf, 658K
http://genomebiology.com/imedia/1946261630792320/supp2.pdf
Additional file 3: Wyres_AdditionalFile_3.pdf.pdf, 383K
http://genomebiology.com/imedia/1677423660792320/supp3.pdf
Additional file 4: Wyres_AdditionalFile_4.pdf, 430K
http://genomebiology.com/imedia/1105303145847820/supp4.pdf
Additional file 5: Wyres_AdditionalFile_5.pdf, 240K
http://genomebiology.com/imedia/1974551928478213/supp5.pdf
Additional file 6: Wyres_AdditionalFile_6.pdf, 417K
http://genomebiology.com/imedia/1891192240847821/supp6.pdf
Additional file 7: Wyres_AdditionalFile_7.pdf, 422K
http://genomebiology.com/imedia/8043933958478214/supp7.pdf
Additional file 8: Wyres_AdditionalFile_8.pdf, 885K
http://genomebiology.com/imedia/1153777944847821/supp8.pdf
Additional file 9: Wyres_AdditionalFile_9.pdf, 420K
http://genomebiology.com/imedia/6866460938478219/supp9.pdf