

REVIEW

Open Access



# Global genomic pathogen surveillance to inform vaccine strategies: a decade-long expedition in pneumococcal genomics

Stephen D. Bentley\* and Stephanie W. Lo

## Abstract

Vaccines are powerful agents in infectious disease prevention but often designed to protect against some strains that are most likely to spread and cause diseases. Most vaccines do not succeed in eradicating the pathogen and thus allow the potential emergence of vaccine evading strains. As with most evolutionary processes, being able to capture all variations across the entire genome gives us the best chance of monitoring and understanding the processes of vaccine evasion. Genomics is being widely adopted as the optimum approach for pathogen surveillance with the potential for early and precise identification of high-risk strains. Given sufficient longitudinal data, genomics also has the potential to forecast the emergence of such strains enabling immediate or pre-emptive intervention. In this review, we consider the strengths and challenges for pathogen genomic surveillance using the experience of the Global Pneumococcal Sequencing (GPS) project as an early example. We highlight the multifaceted nature of genome data and recent advances in genome-based tools to extract useful information relevant to inform vaccine strategies and treatment options. We conclude with future perspectives for genomic pathogen surveillance.

**Keyword:** Pneumococci, Pneumococcal conjugate vaccine, Genomic, Pathogen surveillance

## Background

*Streptococcus pneumoniae* (or pneumococcus) is a common opportunistic pathogen which causes a wide spectrum of diseases. Infections can range from otitis media to severe invasive pneumococcal disease (IPD) including pneumonia, septicaemia and meningitis. Young children in the first few years of life and elderly adults are particularly susceptible to pneumococcal disease. In 2015, pneumococcal infections were estimated to have caused 8.9 million disease cases, including over 317,000 deaths in children under 5 years old. The heaviest disease burden is in low- and middle-income countries (LMICs) [1].

Pneumococcal disease is preventable by vaccination and treatable using antimicrobials. In the early 2000s, pneumococcal conjugate vaccine (PCV) was first rolled out in high-income countries and then gradually in LMICs via The Global Alliance for Vaccines and Immunization (GAVI) [2]. Different from the previous generation of pneumococcal polysaccharide vaccine (PPV), PCV is immunogenic in infants and induces long-term protection by inducing T cell-dependent immune response. The global deployment of PCV has proven to be very effective in reducing pneumococcal disease worldwide. By 2015, deaths of children aged 1–59 months due to pneumococcal disease were estimated to have declined by 51% [1], in comparison to 2000. PCV has also had a positive impact on reducing antimicrobial resistance both through the direct reduction of highly resistant strains targeted by the vaccine and via a secondary effect through a

\* Correspondence: [sdb@sanger.ac.uk](mailto:sdb@sanger.ac.uk)  
Parasites and Microbes, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

reduction in febrile illnesses that often require antimicrobial use [3].

PCVs trigger an immune response in the host to target the polysaccharide capsule surrounding the pneumococcal cell [4]. To escape immune clearance, the capsule is constantly under diversification, resulting in 100 currently recognised forms or serotypes [5]. Currently, PCVs target up to 13 serotypes which account for most of the disease in infants, especially those associated with antimicrobial resistance. Incomplete vaccine coverage of serotypes allows the pneumococcal population to evolve and evade the vaccine [6]; there have been several reports of increases in disease due to non-vaccine serotypes [6–12]. Higher valency vaccines targeting up to 24 serotypes are under development [13] and should contribute to reduction in disease caused by the emerging serotypes not covered by 13-valent PCV (PCV13) and continued surveillance is necessary to inform future vaccine strategies.

The Global Pneumococcal Sequencing (GPS) project has been providing genomic surveillance since 2011 [14]. Here, we describe the biology of pneumococcal disease, the genomic approach taken and lessons learned to understand vaccine evasion mechanisms and to track vaccine-evading strains, advances in genome-based characterisation and future perspectives for genomic pathogen surveillance.

## The biology of pneumococcal disease

### Colonisation is a prerequisite for disease

Understanding pathogen biology and disease mechanisms is important to guide vaccine strategy. The pneumococcus is a commensal coloniser of the human nasopharynx, with person-to-person transmission necessary to compensate for regular clearance from the niche by host immunity [15] and competition within the nasopharyngeal microbiome [16, 17]. Therefore, variation within the human nasopharyngeal niche is the main driver of evolutionary change in the pneumococcal genome [18, 19]. In parallel, any systemic antimicrobial use, regardless of its target pathogen, is a major driver of selection [20]. Invasive disease is an evolutionary dead end for the pneumococcus as it will lead to either clearance by antimicrobials, clearance by host immunity or death of the host.

Pneumococcal colonisation rates vary with geographical location and age. The colonisation rates in young children are usually lower in high-income countries [21–23] and higher in LMICs [24–26]. Host immunity can also explain the age-related variation, which is highest in infants and declines with maturation of the immune system [27]. It is widely accepted that the primary prerequisite for IPD is prior asymptomatic colonisation with the disease-causing strain, usually in the nasopharyngeal niche [28]. Trends in

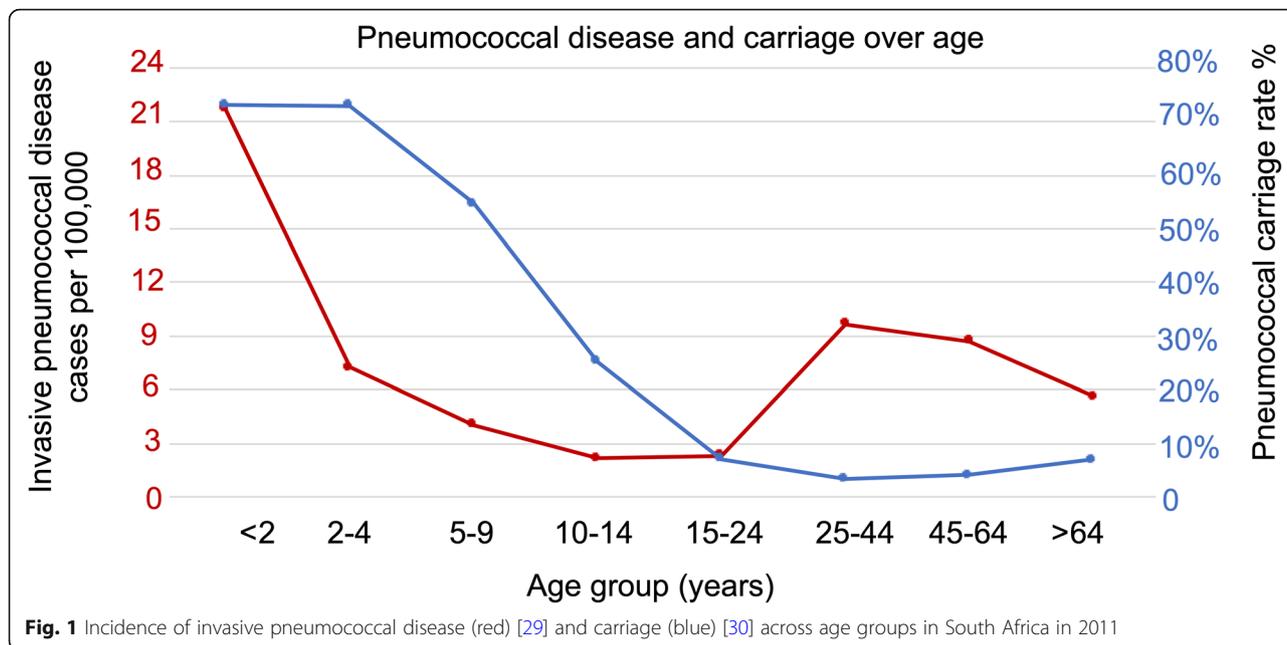
disease rates roughly follow the age distribution of carriage prevalence though it could be affected by other diseases that can compromise the human immune system (e.g. HIV). In South Africa, a higher incidence rate of IPD in adults > 25 years of age compared to those aged 10–24 years of age (Fig. 1) can be explained by the high burden of HIV in adults > 25 years of age [31]. The incidence of IPD in HIV-infected individuals is estimated to be 43 times higher than HIV-uninfected persons [32]. Interestingly, some serotypes are associated with different age groups [33] and HIV status [34].

### The pneumococcal capsule

The pneumococcal capsule is a layer of cross-linked polysaccharide covering the bacterial cell. One important function of the capsule is to protect pneumococcal cells from phagocytosis [35]; pneumococci without a capsule are usually unable to cause invasive disease, but can cause non-invasive diseases [36]. The capsule is also the basis of the typing scheme which has historically been used to taxonomically separate isolates into groups (serotypes) [37]. Sets of antisera raised against reference “type” strains have been used for over 80 years to serotype isolates, allowing an appreciation of serotype prevalence and relative associations with disease [38]. By serotyping pneumococcal isolates from disease and asymptomatic carriage, substantial variation amongst serotypes in their potential to cause invasive disease was observed [39]. This variation in invasive disease potential is not completely understood but may be linked to the basic biochemical features of the capsule; serotypes with high invasive disease potential tend to have thinner capsules that enhance attachment and direct interaction with epithelial cells [40–42] and are associated with shorter carriage duration periods [43].

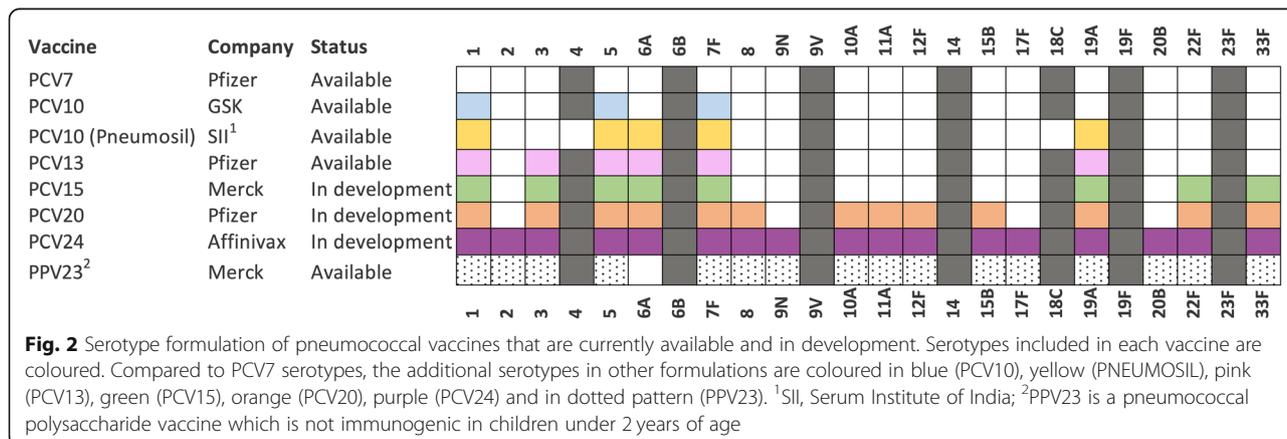
The capsule is encoded by a ~ 10–30-kb gene cluster, known as *cps* for capsule polysaccharide synthesis [44]. The composition and sequences of capsular encoding genes vary between serotypes. Analysing these genetic variations paved the way for the development of DNA-based serotyping methods using PCR [45], DNA microarray [46] and whole-genome sequencing (WGS) [47, 48]. These methods show high concordance with the conventional method that is based on reaction to antisera [49–51]. Genotypic methods provide some advantages, including application to culture negative clinical samples [45, 52], detection of multiple co-colonising serotypes [50, 53] and the discovery of novel genetic variations in *cps*, which may indicate new serotypes [5, 54].

Capsular polysaccharide induces a serotype-specific immune response [55] and has been the basis of pneumococcal vaccination since the first clinical use of two different hexavalent PPVs in 1947 [56]. The valency



was expanded to 14-valent in 1977 and 23-valent in 1983, offering protection against a wider array of disease-associated serotypes [55, 57]. Unfortunately, PPV induces poor immunogenicity in infants because anti-polysaccharide antibody response is associated with specific splenic B cell subsets that are not fully developed in children under 2 years of age [51]. Additionally, PPV solely elicits a T cell-independent immune response that generates a limited duration of protective antibody level [58, 59]. Considering the disease burden is mainly focused in the first 5 years of life, the above PPV limitations motivated the development of pneumococcal conjugate vaccine (PCV), which would better protect infants. PCV is made by covalently linking capsular

polysaccharide to a carrier protein to improve the antibody response and induce long-term protection. PCV is immunogenic in infants and some high-risk patients who do not respond to PPV [60]. The global deployment of PCV since 2000 has been associated with a decreasing pneumococcal disease burden in both children [1] and the indirect protective effect in adults worldwide [61]. Licensed PCVs and those under development, together with 23-valent PPV, are summarised in Fig. 2. Amongst them, the low-cost 10-valent vaccine (PNEUMOSIL) that recently achieved WHO prequalification [62] offers great potential for routine childhood immunisation in LMICs. Although higher-valency PCVs, targeting up to 24 serotypes are under development, the pneumococcal



population as a whole has been a moving target for PCVs over the past two decades and the challenge of incomplete coverage of pneumococcal serotypes remains.

## Mechanisms of vaccine evasion

### Recombination and pneumococcal evolution

The pneumococcus is naturally able to uptake naked DNA from the surrounding environment. This characteristic was first demonstrated by Frederick Griffith in 1928 [37] and later used by Avery, MacLeod and McCarty to demonstrate that the ‘transforming principle’ was pure DNA [63]. In the nasopharyngeal niche, the lysis of bacterial cells through normal turnover leads to naked DNA available for uptake which can provide a source of gene variants where different pneumococcal strains are present. Imported DNA can be recombined into the native genome, providing the pneumococcus with a powerful mechanism for rapid evolutionary adaptation [18, 64]. The ability to recombine multi-gene segments of DNA has allowed the import of genetic ‘islands’ from outside of the species and the reassortment of genes within the species, resulting in distinguishable pneumococcal lineages or strains [65]. Recombination enables strains to replace the whole or partial *cps* and thus change serotype [66–68]; this is commonly known as capsular switching. Any switch from serotypes targeted by the vaccine (i.e. vaccine type, VT) to serotypes not targeted by the vaccine (i.e. non-vaccine type, NVT) can contribute to vaccine evasion.

### Vaccine evasion via capsular switching and strain replacement

Multiple capsule switch events have been characterised in a genomic analysis of the globally prevalent PMEN1 strain [67]. Using ancestral phylogenetic reconstruction and recombination analysis of a temporally and geographically broad collection of genomes, it was possible to infer that the strain had likely emerged in Western Europe in the 1970s before spreading globally over following decades. From the serotype 23F ancestor, 10 capsule switch events were detected, some of which were NVT. One notable switch was to serotype 19A which manifested as an emerging cause of NVT disease in the US, after the introduction of PCV7 in 2000 [69]. Ancestral reconstruction showed that the 23F>19A capsule switch had occurred several years before the introduction of PCV7, indicating that the vaccine had created a positive selection for capsule switch variants that were outside of the vaccine coverage.

Pneumococci circulating in any specific geographic region form a multi-strain, multi-serotype population, which is typically dominated by 6–13 strains that together represent >60% of the population, along with a background of minor strains [51]. PCV have varying

effectiveness in removing VTs from the population. The roll-out of PCV tends to have little effect on overall pneumococcal carriage rates, indicating that the NVT portion of the population is able to expand to fill the niche vacated by VTs [70]. After a period of perturbation, the emergent post-vaccine populations appear to have been shaped by the expansion of a combination of capsule switch variants and strains already dominated by NVTs [66, 71]. The relative contribution of these two vaccine evasion mechanisms varies between countries, as does antimicrobial-selective pressure, resulting in variation in post-PCV emerging NVTs. In general, NVTs with high invasive disease potential (e.g. serotype 8, 12F, 24F) are more commonly seen in IPD after PCV13 introduction [6, 8, 9, 71].

## Genomic surveillance to inform global vaccination strategies

### Motivation and scope of the Global Pneumococcal Sequencing (GPS) project

PCV7 was designed to target the serotypes most frequently causing invasive disease in the US. Vaccine coverage was 83% in children aged <5 years and it was successful in reducing overall IPD by 45% for all age groups over 7 years [72]. In LMICs, PCV was made more affordable through an innovative finance mechanism, the pneumococcal Advance Market Commitment (AMC), initiated by GAVI [2], along with the World Bank and other donors globally in 2009. This mechanism has accelerated the roll-out of PCV to millions of vulnerable children worldwide. However, pneumococcal serotype surveillance indicated that PCV7 would have much lower coverage in many high disease burden LMICs [73, 74]. With this in mind, in 2011 the Bill and Melinda Gates Foundation (in partnership with Emory University, US Centers for Disease Control and Prevention, and the Wellcome Sanger Institute) initiated the GPS project [14] with the primary goal of applying genomics to understand pneumococcal evolution in response to vaccine introduction in LMICs. At that time, GPS was a pioneering project with little precedent to follow, but, 10 years on, lessons have been learned and new directions plotted. The project began with Founding Partners in three African countries (The Gambia, Malawi and South Africa) and the ambition to add partners to achieve wide geographic coverage, prioritising LMICs eligible for GAVI support for PCV rollout. By March 2021, the GPS project sequenced 26,100 pneumococcal genomes representing 57 countries.

Initially, the GPS project prioritised sequencing of isolates from IPD in children under 5 years old, collected pre- and post-PCV introduction. The Founding Partners were from well-resourced institutions, each with a strong track record in pneumococcal surveillance, so were

easily able to satisfy the preferred sampling criteria. This was not the case for many other countries and the compromises, such as inclusion of samples from asymptomatic colonisation rather than IPD, were necessary. Allowing such compromises emphasised the importance of careful curation of sample metadata. It was imperative that reliable metadata were collected for every sequenced sample so that specific analytical questions were powered by as many samples as possible; for example, if samples did not have information on whether they were from healthy carriers or IPD, they could not be used in an analysis of genetics associated with virulence. To maximize the utility of the GPS database, no sample was sequenced unless metadata was submitted in advance, thus ensuring that all sequencing effort generated genomic data of enhanced analytical value. The minimal metadata requirement for GPS samples was set simply as ‘date’ and ‘geography’ of isolation, with a range of clinical and microbiological data also typically recorded (see Table 1 for further details). On average, isolates had entries for 37 metadata fields which were linked to the output of genome-derived analyses (e.g. in silico serotype, genotype and antimicrobial resistance determinants). Thus, the GPS provides a rich, public database that has supported a number of data-driven and hypothesis-driven sub-studies with a central theme of pneumococcal disease prevention [75, 76].

### Challenges and solutions for genomic surveillance in LMICs

Isolation of *S. pneumoniae* from suspected cases of IPD can be very challenging and may often not be attempted in some countries, necessitating clinical decision making based on other available evidence (e.g. symptoms and prescribing guidelines). Major barriers to pneumococcal isolation from IPD cases include lack of microbiological expertise, lack of correct microbiological reagents (e.g. sheep’s blood rather than human blood) and patient self-administration of antimicrobials prior to presenting to the healthcare provider. Whilst the microbiological barriers can be addressed with training and supply of resources, the issue of uncontrolled antimicrobial access is much more challenging. In countries where culture of IPD isolates is not likely, collecting isolates from the nasopharynx of healthy carriers can be a viable alternative method to evaluate the vaccine impact on pneumococcal population [66] potentially predicting the emerging serotypes/strains post-vaccine using mathematical modelling [77]. However, some serotypes that are frequently found in IPD cases are rarely observed in carriage (e.g. serotype 1), and vice versa [39, 51], so interpretation can be limited.

A fundamental challenge of any global surveillance system, particularly one prioritising LMICs, is variation

**Table 1** An example of the Global Pneumococcal Sequencing (GPS) project metadata

Categories	Fields	Example
ID	Public name	GPS_ZA_0001
Geographical location	Country	South Africa
	Region	Gauteng
	City	Johannesburg
	Facility where collected	Hospital A
	Submitting institute	NICD
Time	Year	2010
	Month	Aug
Clinical data	Gender	F
	Age (years)	0
	Age (months)	1
	Age (days)	0
	Clinical manifestation	Meningitis
	Source	Cerebrospinal fluid
	HIV status	Negative
Microbiological data	Other underlying conditions	No
	Phenotypic serotype method	Quellung
	Phenotypic serotype	19A
Antimicrobial susceptibility <sup>a</sup>	Multilocus sequence type	ST81
	Method	Broth dilution
	Antimicrobial (e.g. penicillin)	2 mg/L
Selection	Random selection	Y

<sup>a</sup>Antimicrobial susceptibility profile of 17 antimicrobials including penicillin, amoxicillin, cefotaxime, ceftriaxone, cefuroxime, meropenem, chloramphenicol, cotrimoxazole, erythromycin, clindamycin, linezolid, levofloxacin, ciprofloxacin, synergid, tetracycline, rifampin and vancomycin

in local infrastructure and resources, which often also impacts on the level of engagement that an individual project partner is able to commit to. Accordingly, it is important to recognise the motivations and limitations for each partner in order to maximise mutual benefit. Some engagements may be relatively passive, with partners being content to simply contribute culture samples to the project, in the knowledge that analysis of their samples will be reported back to them in the context of regional and global analyses. Others may be more actively involved in developing local genomics capacity and wish to generate and analyse data locally in a way that can be integrated with the global database. Such variation requires flexibility in the global system and failure to provide the necessary flexibility would likely lead to partner disengagement and weakening of the surveillance data captured. In view of such variations, the GPS project devises bespoke support for project partners to

cater for different needs in training, data analysis and interpretation.

#### **Models of sequence data generation: from central to local**

Generation of high-quality genome sequence is fundamental to any genomic surveillance system. In the last 2 decades, genome sequencing has progressed from a somewhat cumbersome technology, restricted to a few well-resourced specialist institutions, to become a relatively routine molecular biology tool. In recent years, the sequencing technology companies have developed a greater variety of hardware catering for a variety of uses and budgets. This, coupled with a drive toward genomics as a routine technology for disease surveillance, has led to an expansion in the availability of sequencing hardware in LMICs. In the first phase of GPS (2011–2019) nearly all of the genome sequence data was generated at the Sanger Institute. In the next phase, we have placed a strong emphasis on decentralising data generation in the hope of creating a long-term sustainable genomic surveillance network. It must be acknowledged that the introduction of any new technology takes time, particularly in a resource-limited setting but there are already several high-quality genomics laboratories (e.g. NICD in South Africa [78]) in LMICs and growing networks of national and regional training providers (e.g. MRC unit The Gambia [79] and H3ABionet [80]) so the outlook is positive.

Where data generation is centralised, the movement of samples (bacterial cultures or DNA extracts) presents a significant challenge, often including the need for legal documentation such as material transfer agreements. Assuming decentralised data generation can be achieved, such sample logistic challenges are replaced by data sharing challenges. With the centralised model, outwards data sharing can be relatively straightforward because it emanates from a single uniform data source that has been generated and quality checked. With a decentralised model, there may be variations in data generation so systems need to be developed to enable the data to be harmonised within a unifying data platform. Such systems will need to account for variations in local informatics infrastructure and requirements for legal documentation on data sharing agreements. Data sharing platforms should also be built on open-source software so that the entire stakeholder community can engage in development.

#### **Database and data sharing**

The database is an important element of a genomic surveillance system. It serves as a data hub in which a collection of data from multiple sources is organised for users to view, search, download and share. Designing, building and maintaining a database are equally important and all three stages require

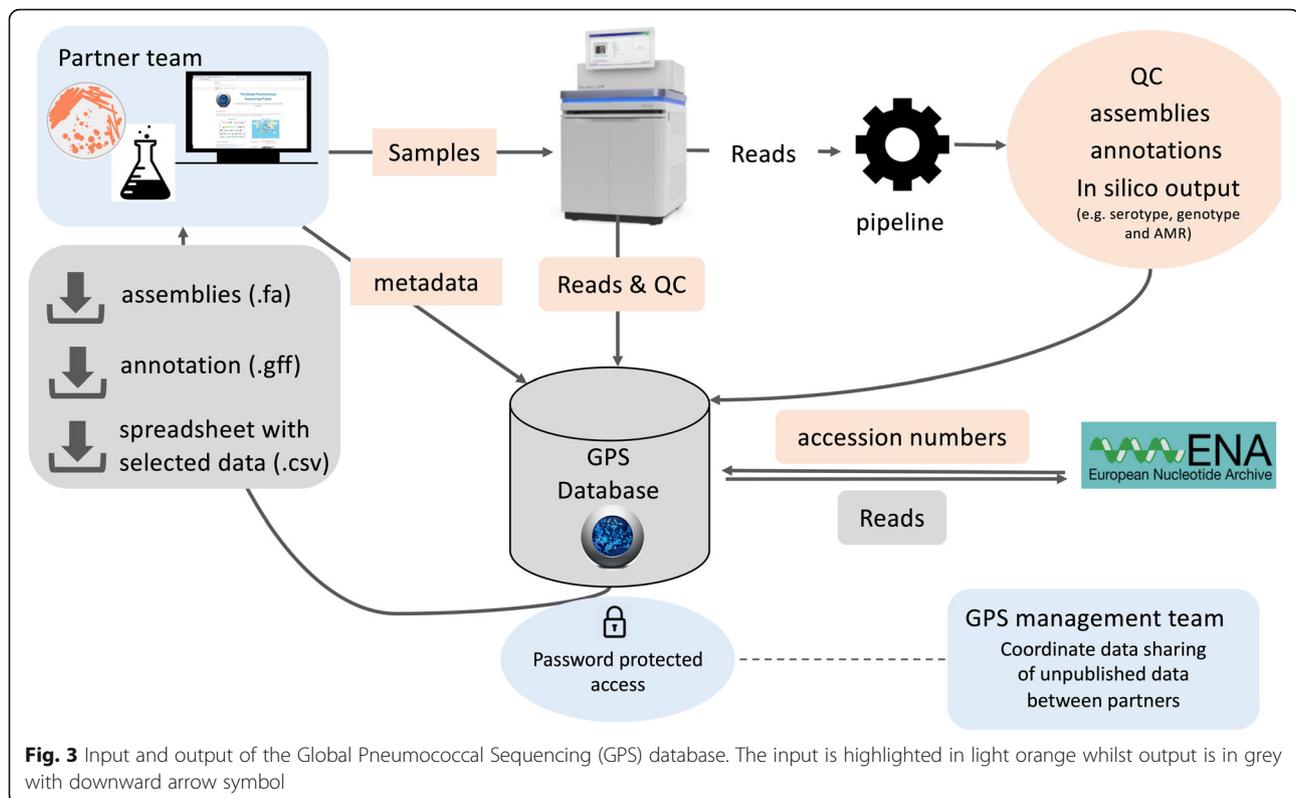
informatics infrastructure and support. In a surveillance system that involves a network of partners, databases should also be designed to facilitate both individual access to one's own data and data sharing between partners (Fig. 3).

Data generated from genomic surveillance has great potential value beyond the original purpose so should be publicly accessible. To maximise utility, open data, open software and open access publications are essential and have become strict requirements for many funders [81, 82]. Whilst the availability of open data continues to increase, sharing the benefit arising from the utilisation of these genetic resources in a fair and equitable way is imperative to maintain the virtuous cycle of data production. To this end, the Nagoya Protocol was initiated on 12 October 2014. It provides legal certainty and a transparent benefit-sharing framework for both the genetic resources provider and users [83].

#### **Data analysis**

Translating large amounts of data from a genomic surveillance system into meaningful information to guide public health decisions requires accurate data analysis and interpretation. Over the last decade, a variety of analysis tools have been developed that are robust and generic for application across species. From a pneumococcal genome, we can quickly and reliably extract public health-relevant information, including serotype [47, 48], genotype [84, 85], and antimicrobial resistance profile [86–88]. Such tools are being adapted to be run as applications within websites so formal bioinformatics expertise is not required. For example, Pathogenwatch [89] offers *in silico* detection and characterisation of genome data for a wide range of microbial pathogens. By simple 'drag-and-drop' of sequence data files into a browser window, users can quickly obtain public health-relevant information [90].

Genome data is also powerful in answering key questions, such as the genetic and geographical origin of vaccine evading strains. By calculating substitution rate, we can extrapolate when and where a pneumococcal strain emerged and/or acquired the genetic variation that conferred resistance to the vaccine or antimicrobials [67, 91]. In the first phase of the GPS project 26,100 genomes were sequenced. These data allowed the systematic definition of 621 circulating strains (referred to as Global Pneumococcal Sequence Clusters (GPSCs) and detection of all genomic variations within, including identification of strains containing up to 15 different serotypes [51]. The dataset is dominated by 35 strains (> 100 genomes each) that represent 62% of the dataset; several of these are globally disseminated and associated with multidrug resistance. The GPSC strain definition lays the



foundation for understanding pneumococcal population changes after roll-out of PCV. In a GPS study of ~3000 pneumococcal isolates from laboratory-based surveillance programmes in six countries collected before and after PCV [71], VTs were replaced by NVTs, as expected [8, 29, 92–94]. Using GPSC, we observed that the expansion of NVTs was mainly mediated by a shift in the balance of serotypes within globally spreading strains, with a smaller impact due to increases of strains that exclusively express non-vaccine serotypes. However, this observation varies amongst countries, as do the prevalent serotypes and GPSCs post-PCV. Such variations can partly be explained by the differences in the pneumococcal population prior to the vaccine roll-out and the variation in antimicrobial selective pressure amongst countries. These data have also enabled the discovery of nine putative novel serotypes [54] and previously unrecognised resistance determinants [95].

#### Data visualisation and interpretation

Visualisation of analysed data is a key step for interpretation of large, complex datasets which typically derive from genomic surveillance systems. Visualising genetic relationships between isolates on a phylogenetic tree, together with associated metadata, is a powerful approach. Popular examples of visualisation software include Microreact [96] and NextStrain [97]. The GPS project

uses Microreact to make fully analysed datasets easily accessible including snapshots of country-specific [98] and strain-specific studies [99] within project web resources [100, 101]. GPS also uses the Phandango software for visualisation of data specific to gene content variations such as mutation, recombination and pan-genome variations [102, 103].

Interpretation of analysis output requires a certain level of knowledge in bioinformatics and the pathogen studied. In most microbiology laboratories or surveillance networks in LMICs, bioinformatics is a relatively new expertise that requires training and hands-on experience. Together with the sister project JUNO [104], GPS is developing a learning portfolio [105, 106] to suit different partners' needs informed by a survey that was conducted amongst partners in the GPS and JUNO projects.

#### Conclusions and future directions

The GPS project has clearly demonstrated the added value of genomics in pathogen surveillance over the past decade by identifying the emerging serotypes and vaccine-escaping strains, thus providing evidence basis to inform future vaccine strategies. The project also highlighted the data gap and the need to build a more sustainable surveillance system to optimise disease prevention strategies.

### **Filling important data gaps in countries with a high burden of disease**

In a 2018 study of the global burden of pneumococcal disease, Wahl et al. showed that approximately half of all pneumococcal deaths in 2015 occurred in just four countries: India, Nigeria, Democratic Republic of Congo and Pakistan [1]. However, when that study was published, those four countries represented only 5% of the GPS database. This mismatch was largely due to the difficulty in accessing appropriate samples, with each country having a unique set of economic, technical and political challenges which put them beyond the reach of the initial GPS model. However, there is no lack of capable and motivated stakeholders in those countries and it is hoped that, with a decentralised model and sufficient support for capacity development, those data gaps can be filled. With more representative data, genomic analyses have the potential to give a clear picture of pathogen evolution and risk in the context of regional and global spread.

### **Combating multiple pathogens with a generic genomic surveillance system**

GPS has already been successful in generating a rich knowledge base for informing future pneumococcal disease control strategy and is making good progress in developing global infrastructure for ongoing genomic surveillance, but there is still much work to be done to achieve a self-sustaining system. Systems for global genomic surveillance of other vaccine-preventable bacterial pathogens are also being established with many solutions likely to be generic across different pathogen species. The most obvious parallels with GPS would be for endemic bacterial pathogens that have similar population structure and incomplete-coverage vaccines. One example is *Neisseria meningitidis* where a variety of vaccine formulations are available but none with complete species coverage. In Africa, where the meningococcal disease burden is highest, widespread use of conjugate vaccine targeting the serogroup A polysaccharide capsule has seen a dramatic reduction in serogroup A disease but also an increase in disease due to other serogroups, most notably serogroup X for which there is currently no licenced vaccine [107]. Meningococcal disease epidemiology in the ‘meningitis belt’ of Africa is characterised by epidemic waves and succession of dominant strains [97]; genomics has great potential for creating a clear understanding of meningococcal population dynamics and creating preparedness for future epidemic waves.

### **Enhancing capacity building in LMICs with high disease burden**

Genomic surveillance of vaccine-preventable pathogens will only be sustainable through local data generation

and analysis which currently places a great emphasis on capacity building in countries with high disease burden. Fortunately, there is a growing wealth of initiatives for training in genomics, including both wet-lab and bioinformatic expertise, with a strong emphasis on the ‘train-the-trainer’ philosophy to ensure sustainability. The supply of sequencing hardware and consumables is improving in many parts of the world that were previously poorly served. Also, advocacy campaigns are raising awareness of the value of genomics with national policy-makers to bring genomics into national disease control strategies. Furthermore, the importance of genomics capacity building in high burden countries is being prioritised by multiple major global health funders. Other fundamental challenges remain. Mechanisms for transfer of funds to the places where they are needed, and protocols for data sharing, need to be made more efficient whilst being sensitive to the needs of the diverse stakeholders. However, by exploiting the universal nature of DNA sequencing and integrating the need to apply genomics to a range of endemic and epidemic pathogens in high burden countries, it should be possible to develop sustainable pathogen genomics surveillance capacity that will have both local and global benefit for infectious disease prevention.

### **Optimising vaccine formulation**

The WHO lists vaccines “available” for nine bacterial pathogens with differing disease patterns (endemic, epidemic, opportunistic) and differing recommendations for implementation, with some more commonly used in response to outbreaks [108]. In some cases, the vaccine antigen is generally invariant and gives good coverage across the species (e.g. diphtheria, pertussis, tetanus, typhoid). In these cases, low-density genomic surveillance would be valuable in characterising cases of vaccine failure to understand the mechanism of vaccine evasion and to predict whether it is likely to be an emerging threat. In cases where the vaccine antigen is highly variable and the species coverage is partial, it is likely that currently, effective vaccines will need to be periodically reformulated in a manner analogous to the seasonal influenza vaccine. The reformulation cycle may not need to be as rapid as for influenza (annual) and would vary in turnover rate between species. However, having a longitudinal genomic record of pathogen evolution would be enormously valuable in designing new vaccines and potentially forecasting the potential risk/benefit of their use.

Mathematical modelling has provided useful tools for predicting infectious disease risk. Incorporating evolutionary parameters for bacterial pathogens has been a challenge, particularly due to the complexity created by horizontal gene transfer in multi-strain species, leaving model outputs with a high degree of uncertainty. Recent

models attempt to take advantage of the detailed evolutionary knowledge provided by availability of longitudinal population genomics datasets. Models based on the balancing of individual gene frequencies across a pathogen species population, termed ‘negative frequency-dependent selection’, have been applied to provide plausible, high-resolution explanations for population responses to vaccines [77] and emergence of pathogenic strains [109]. This approach has also been applied to hypothesise PCV formulations that could be tailored to the extant population and provide better disease prevention [110]. A key strength of this approach is that it could allow for region-specific vaccine design, addressing the reality that pathogen populations can vary significantly across the world and that ‘one size fits all’ global vaccines may not be the optimum approach. The WHO also lists a number of ‘pipeline’ vaccines and many others are in early design stages. Population genomics is increasingly prioritised in vaccine design and is further employed as the foundation of other powerful ‘omics’ approaches, such as surveying potential immunogenicity across complete proteome arrays [111].

#### Potential application of genomics in clinical microbiology laboratories

Genomic technologies have the potential to provide solutions for the inherent challenge of isolating the pathogen in cases of disease. Failure to culture the live pathogen from a clinical sample is not uncommon and molecular techniques are being developed that aim to extract and analyse the pathogen DNA directly rather than relying on the presence of viable pathogen cells. If these techniques can be honed to enrich whole genomes, then clinical pathogen genomic protocols for some species could become ‘culture-free’. Another potential benefit of genomics comes from the correlation and derivation of important pathogen phenotypes that are normally determined through an array of wet-lab techniques, often with species-specific protocols and each requiring maintenance of lab infrastructure and spend on consumables. A number of studies have shown a high degree of concordance for deriving such phenotypes directly from genomic data and many public health labs are choosing genomics as their main, or only method for their determination [112, 113].

In conclusion, overcoming the above challenges requires multi-disciplinary expertise, support from the government and sufficient funding. The approach taken and lessons learned from the GPS project discussed in this review—surveillance priority and infrastructure, collaboration models, portfolio of capacity building and bioinformatics training, solutions to challenges in LMICs, recent advances in genomics—may guide generic surveillance networks at national and international level.

#### Abbreviations

PCV: Pneumococcal conjugate vaccine; LMICs: Low- and middle-income countries; AMC: Advance Market Commitment; IPD: Invasive pneumococcal disease; WGS: Whole-genome sequencing; CPS: Capsular polysaccharide; PPVs: Pneumococcal polysaccharide vaccines; VT: Vaccine serotype; NVT: Non-vaccine serotype

#### Acknowledgements

We would like to acknowledge the Bill & Melinda Gates Foundation for funding the Global Pneumococcal Sequencing project. We thank Prof Shabir Madhi, Dr Sarah Downs and Dr Susan Nzenze from the University of Witwatersrand and Dr Susan Meiring and Dr Anne von Gottberg from the National Institute for Communicable Diseases, South Africa, for providing the pneumococcal carriage rate and IPD incidence rate for generating Fig. 1. We appreciate Dr Christine Boinett, Dr Dorota Jamroz, and Dr Narender Kumar for their review and Dr Kate Mellor for her review and help on revision.

#### Authors’ contributions

S.D.B and S.W.L prepared the manuscript. The authors read and approved the final manuscript.

#### Funding

The Bill and Melinda Gates Foundation (Investment ID INV-003570)

#### Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

S.D.B reports personal fees from Pfizer and Merck, outside the submitted work. S.W.L declares no competing interests.

Received: 30 November 2020 Accepted: 30 April 2021

Published online: 17 May 2021

#### References

1. Wahl B, O'Brien KL, Greenbaum A, Majumder A, Liu L, Chu Y, et al. Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob Health*. 2018;6(7):e744–57. [https://doi.org/10.1016/S2214-109X\(18\)30247-X](https://doi.org/10.1016/S2214-109X(18)30247-X).
2. Pneumococcal vaccine support. <https://www.gavi.org/types-support/vaccine-support/pneumococcal>. Accessed 22 Mar 2021.
3. Klugman KP, Black S. Impact of existing vaccines in reducing antibiotic resistance: Primary and secondary effects. *Proc Natl Acad Sci USA*. 2018; 115(51):12896–901. <https://doi.org/10.1073/pnas.1721095115>.
4. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, et al. Pneumococcal capsules and their types: past, present, and future. *Clin Microbiol Rev*. 2015;28(3):871–99. <https://doi.org/10.1128/CMR.00024-15>.
5. GPS :Global Pneumococcal Sequencing Project | Serotypes. <https://www.pneumogen.net/gps/serotypes.html>. Accessed 22 Mar 2021.
6. Ladhani SN, Collins S, Djennad A, Sheppard CL, Borrow R, Fry NK, et al. Rapid increase in non-vaccine serotypes causing invasive pneumococcal disease in England and Wales, 2000–17: a prospective national observational cohort study. *Lancet Infect Dis*. 2018;18(4):441–51. [https://doi.org/10.1016/S1473-3099\(18\)30052-5](https://doi.org/10.1016/S1473-3099(18)30052-5).
7. Rokney A, Ben-Shimol S, Korenman Z, Porat N, Gorodnitsky Z, Givon-Lavi N, et al. Emergence of *Streptococcus pneumoniae* Serotype 12F after Sequential Introduction of 7- and 13-Valent Vaccines, Israel. *Emerging Infect Dis*. 2018;24(3):453–61. <https://doi.org/10.3201/eid2403.170769>.
8. Mackenzie GA, Hill PC, Jeffries DJ, Hossain I, Uchendu U, Ameh D, et al. Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: a population-based

- surveillance study. *Lancet Infect Dis.* 2016;16(6):703–11. [https://doi.org/10.1016/S1473-3099\(16\)00054-2](https://doi.org/10.1016/S1473-3099(16)00054-2).
9. Ouldali N, Levy C, Varon E, Bonacorsi S, Béchet S, Cohen R, et al. Incidence of paediatric pneumococcal meningitis and emergence of new serotypes: a time-series analysis of a 16-year French national survey. *Lancet Infect Dis.* 2018;18(9):983–91. [https://doi.org/10.1016/S1473-3099\(18\)30349-9](https://doi.org/10.1016/S1473-3099(18)30349-9).
  10. Weinberger R, von Kries R, van der Linden M, Rieck T, Siedler A, Falkenhorst G. Invasive pneumococcal disease in children under 16 years of age: Incomplete rebound in incidence after the maximum effect of PCV13 in 2012/13 in Germany. *Vaccine.* 2018;36(4):572–7. <https://doi.org/10.1016/j.vaccine.2017.11.085>.
  11. Ubukata K, Takata M, Morozumi M, Chiba N, Wajima T, Hanada S, et al. Effects of Pneumococcal Conjugate Vaccine on Genotypic Penicillin Resistance and Serotype Changes, Japan, 2010–2017. *Emerg Infect Dis.* 2018;24(11):2010–20. <https://doi.org/10.3201/eid2411.180326>.
  12. Brandileone M-CC, Almeida SCG, Minamisava R, Andrade A-L. Distribution of invasive *Streptococcus pneumoniae* serotypes before and 5 years after the introduction of 10-valent pneumococcal conjugate vaccine in Brazil. *Vaccine.* 2018;36(19):2559–66. <https://doi.org/10.1016/j.vaccine.2018.04.010>.
  13. Klugman KP, Rodgers GL. Time for a third-generation pneumococcal conjugate vaccine. *Lancet Infect Dis.* 21:14–6. [https://doi.org/10.1016/S1473-3099\(20\)30513-2](https://doi.org/10.1016/S1473-3099(20)30513-2).
  14. Global Pneumococcal Sequencing Project. <https://www.pneumogen.net/gps/>. Accessed 22 Mar 2021.
  15. McCool TL, Cate TR, Moy G, Weiser JN. The immune response to pneumococcal proteins during experimental human carriage. *J Exp Med.* 2002;195(3):359–65. <https://doi.org/10.1084/jem.20011576>.
  16. Pericone CD, Overweg K, Hermans PW, Weiser JN. Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. *Infect Immun.* 2000;68(7):3990–7. <https://doi.org/10.1128/iai.68.7.3990-3997.2000>.
  17. Auranen K, Mehtälä J, Tanskanen AS, Kaltoft M. Between-strain competition in acquisition and clearance of pneumococcal carriage—epidemiologic evidence from a longitudinal study of day-care children. *Am J Epidemiol.* 2010;171(2):169–76. <https://doi.org/10.1093/aje/kwp351>.
  18. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P. Efficient Inference of Recent and Ancestral Recombination within Bacterial Populations. *Mol Biol Evol.* 2017;34(5):1167–82. <https://doi.org/10.1093/molbev/msx066>.
  19. Ganaie F, Saad JS, McGee L, van Tonder AJ, Bentley SD, Lo SW, et al. A New Pneumococcal Capsule Type, 10D, is the 100th Serotype and Has a Large cps Fragment from an Oral *Streptococcus*. *MBio.* 2020;11(3). <https://doi.org/10.1128/mBio.00937-20>.
  20. Feikin DR, Dowell SF, Nwanyawu OC, Klugman KP, Kazembe PN, Barat LM, et al. Increased carriage of trimethoprim/sulfamethoxazole-resistant *Streptococcus pneumoniae* in Malawian children after treatment for malaria with sulfadoxine/pyrimethamine. *J Infect Dis.* 2000;181(4):1501–5. <https://doi.org/10.1086/315382>.
  21. Southern J, Andrews N, Sandu P, Sheppard CL, Waight PA, Fry NK, et al. Pneumococcal carriage in children and their household contacts six years after introduction of the 13-valent pneumococcal conjugate vaccine in England. *Plos One.* 2018;13(5):e0195799. <https://doi.org/10.1371/journal.pone.0195799>.
  22. Lindstrand A, Galanis I, Darenberg J, Morfeldt E, Naucler P, Blennow M, et al. Unaltered pneumococcal carriage prevalence due to expansion of non-vaccine types of low invasive potential 8 years after vaccine introduction in Stockholm, Sweden. *Vaccine.* 2016;34(38):4565–71. <https://doi.org/10.1016/j.vaccine.2016.07.031>.
  23. Ho P-L, Chiu SS, Law PY, Chan EL, Lai EL, Chow K-H. Increase in the nasopharyngeal carriage of non-vaccine serogroup 15 *Streptococcus pneumoniae* after introduction of children pneumococcal conjugate vaccination in Hong Kong. *Diagn Microbiol Infect Dis.* 2015;81(2):145–8. <https://doi.org/10.1016/j.diagmicrobio.2014.11.006>.
  24. Hammit LL, Etyang AO, Morpeth SC, Ojal J, Mutuku A, Mturi N, et al. Effect of ten-valent pneumococcal conjugate vaccine on invasive pneumococcal disease and nasopharyngeal carriage in Kenya: a longitudinal surveillance study. *Lancet.* 2019;393(10186):2146–54. [https://doi.org/10.1016/S0140-6736\(18\)33005-8](https://doi.org/10.1016/S0140-6736(18)33005-8).
  25. Usuf E, Christian B, Gladstone R, Bojang E, Jawneh K, Cox I, et al. Persistent and emerging pneumococcal carriage serotypes in a rural Gambian community after ten years of pneumococcal conjugate vaccine pressure. *Clin Infect Dis.* 2020. <https://doi.org/10.1093/cid/cia856>.
  26. Kandasamy R, Gurung M, Thapa A, Ndimah S, Adhikari N, Murdoch DR, et al. Multi-serotype pneumococcal nasopharyngeal carriage prevalence in vaccine naïve Nepalese children, assessed using molecular serotyping. *Plos One.* 2015;10(2):e0114286. <https://doi.org/10.1371/journal.pone.0114286>.
  27. Mubarak A, Ahmed MS, Upile N, Vaughan C, Xie C, Sharma R, et al. A dynamic relationship between mucosal T helper type 17 and regulatory T-cell populations in nasopharynx evolves with age and associates with the clearance of pneumococcal carriage in humans. *Clin Microbiol Infect.* 2016;22:736.e1-7. <https://doi.org/10.1016/j.cmi.2016.05.017>.
  28. Bogaert D, De Groot R, Hermans PWM. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis.* 2004;4(3):144–54. [https://doi.org/10.1016/S1473-3099\(04\)00938-7](https://doi.org/10.1016/S1473-3099(04)00938-7).
  29. von Gottberg A, de Gouveia L, Tempia S, Quan V, Meiring S, von Mollendorf C, et al. Effects of vaccination on invasive pneumococcal disease in South Africa. *N Engl J Med.* 2014;371(20):1889–99. <https://doi.org/10.1056/NEJMoa1401914>.
  30. Madhi SA, Nzenze SA, Nunes MC, Chinyanganya L, Van Niekerk N, Kahn K, et al. Residual colonization by vaccine serotypes in rural South Africa four years following initiation of pneumococcal conjugate vaccine immunization. *Expert Rev Vaccines.* 2020;19(4):383–93. <https://doi.org/10.1080/14760584.2020.1750377>.
  31. Mabaso M, Makola L, Naidoo I, Mlangeni LL, Jooste S, Simbayi L. HIV prevalence in South Africa through gender and racial lenses: results from the 2012 population-based national household survey. *Int J Equity Health.* 2019;18(1):167. <https://doi.org/10.1186/s12939-019-1055-6>.
  32. Meiring S, Cohen C, Quan V, de Gouveia L, Feldman C, Karstaedt A, et al. HIV infection and the epidemiology of invasive pneumococcal disease (IPD) in south african adults and older children prior to the introduction of a pneumococcal conjugate vaccine (PCV). *Plos One.* 2016;11(2):e0149104. <https://doi.org/10.1371/journal.pone.0149104>.
  33. Imöhl M, Reinert RR, Ocklenburg C, van der Linden M. Association of serotypes of *Streptococcus pneumoniae* with age in invasive pneumococcal disease. *J Clin Microbiol.* 2010;48(4):1291–6. <https://doi.org/10.1128/JCM.01937-09>.
  34. Harboe ZB, Larsen MV, Ladelund S, Kronborg G, Konradsen HB, Gerstoft J, et al. Incidence and risk factors for invasive pneumococcal disease in HIV-infected and non-HIV-infected individuals before and after the introduction of combination antiretroviral therapy: persistent high risk among HIV-infected injecting drug users. *Clin Infect Dis.* 2014;59(8):1168–76. <https://doi.org/10.1093/cid/ciu558>.
  35. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect Immun.* 2010;78(2):704–15. <https://doi.org/10.1128/IAI.00881-09>.
  36. Keller LE, Robinson DA, McDaniel LS. Nonencapsulated *Streptococcus pneumoniae*: Emergence and Pathogenesis. *MBio.* 2016;7(2):e01792. <https://doi.org/10.1128/mBio.01792-15>.
  37. Griffith F. The significance of pneumococcal types. *J Hyg (Lond).* 1928;27(2):113–59. <https://doi.org/10.1017/S0022172400031879>.
  38. Beckler E, Macleod P. The neufeld method of pneumococcus type determination as carried out in a public health laboratory: a study of 760 typings. *J Clin Invest.* 1934;13(6):901–7. <https://doi.org/10.1172/JCI100634>.
  39. Balsells E, Dagan R, Yildirim I, Gounder PP, Steens A, Muñoz-Almagro C, et al. The relative invasive disease potential of *Streptococcus pneumoniae* among children after PCV introduction: A systematic review and meta-analysis. *J Infect.* 2018;77(5):368–78. <https://doi.org/10.1016/j.jinf.2018.06.004>.
  40. Hammerschmidt S, Wolff S, Hocke A, Rosseau S, Müller E, Rohde M. Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. *Infect Immun.* 2005;73(8):4653–67. <https://doi.org/10.1128/IAI.73.8.4653-4667.2005>.
  41. Cundell DR, Gerard NP, Gerard C, Idanpaan-Heikkilä I, Tuomanen EI. *Streptococcus pneumoniae* anchor to activated human cells by the receptor for platelet-activating factor. *Nature.* 1995;377(6548):435–8. <https://doi.org/10.1038/377435a0>.
  42. Cundell DR, Weiser JN, Shen J, Young A, Tuomanen EI. Relationship between colonial morphology and adherence of *Streptococcus pneumoniae*. *Infect Immun.* 1995;63(3):757–61. <https://doi.org/10.1128/IAI.63.3.757-761.1995>.

43. Brueggemann AB, Peto TEA, Crook DW, Butler JC, Kristinsson KG, Spratt BG. Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. *J Infect Dis*. 2004;190(7):1203–11. <https://doi.org/10.1086/423820>.
44. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *Plos Genet*. 2006;2(3):e31. <https://doi.org/10.1371/journal.pgen.0020031>.
45. Streptococcus Lab | StrepLab | Resources | CDC. <https://www.cdc.gov/streplab/pneumococcus/resources.html>. Accessed 22 Mar 2021.
46. Hinds J, Laing KG, Mangan JA, Butcher PD. Microarrays for microbes: the brmg@ approach. *Comp Funct Genomics*. 2002;3(4):333–7. <https://doi.org/10.1002/cfg.187>.
47. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, et al. Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. *PeerJ*. 2016;4:e2477. <https://doi.org/10.7717/peerj.2477>.
48. Epping L, van Tonder AJ, Gladstone RA, The Global Pneumococcal Sequencing Consortium, Bentley SD, Page AJ, et al. SeroBA: rapid high-throughput serotyping of *Streptococcus pneumoniae* from whole genome sequence data. *Microb Genom*. 2018;4. <https://doi.org/10.1099/mgen.0.000186>.
49. Morais L, Carvalho M da G, Roca A, Flannery B, Mandomando I, Soriano-Gabarró M, et al. Sequential multiplex PCR for identifying pneumococcal capsular serotypes from South-Saharan African clinical isolates. *J Med Microbiol*. 2007;56 Pt 9:1181–4. doi:10.1099/jmm.0.47346-0.
50. Turner P, Hinds J, Turner C, Jankhot A, Gould K, Bentley SD, et al. Improved detection of nasopharyngeal cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. *J Clin Microbiol*. 2011;49(5):1784–9. <https://doi.org/10.1128/JCM.00157-11>.
51. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, et al. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. *EBioMed*. 2019;43:338–46. <https://doi.org/10.1016/j.ebiom.2019.04.021>.
52. Saha SK, Darmstadt GL, Baqui AH, Hossain B, Islam M, Foster D, et al. Identification of serotype in culture negative pneumococcal meningitis using sequential multiplex PCR: implication for surveillance and vaccine design. *Plos One*. 2008;3(10):e3576. <https://doi.org/10.1371/journal.pone.0003576>.
53. Knight JR, Dunne EM, Mulholland EK, Saha S, Satzke C, Tothpal A, et al. Determining the serotype composition of mixed samples of pneumococcus using whole-genome sequencing. *Microb Genom*. 7. <https://doi.org/10.1099/mgen.0.000494>.
54. van Tonder AJ, Gladstone RA, Lo SW, Nahm MH, du Plessis M, Cornick J, et al. Putative novel cps loci in a large global collection of pneumococci. *Microb Genom*. 2019;5(7). <https://doi.org/10.1099/mgen.0.000274>.
55. Grabenstein JD, Klugman KP. A century of pneumococcal vaccination research in humans. *Clin Microbiol Infect*. 2012;18(Suppl 5):15–24. <https://doi.org/10.1111/j.1469-0691.2012.03943.x>.
56. Heidelberg M, MacLEOD CM, Di Lapi MM. The human antibody response to simultaneous injection of six specific polysaccharides of pneumococcus. *J Exp Med*. 1948;88(3):369–72. <https://doi.org/10.1084/jem.88.3.369>.
57. Robbins JB, Austrian R, Lee CJ, Rastogi SC, Schiffman G, Henrichsen J, et al. Considerations for formulating the second-generation pneumococcal capsular polysaccharide vaccine with emphasis on the cross-reactive types within groups. *J Infect Dis*. 1983;148(6):1136–59. <https://doi.org/10.1093/infdis/148.6.1136>.
58. Mond JJ, Lees A, Snapper CM. T cell-independent antigens type 2. *Annu Rev Immunol*. 1995;13(1):655–92. <https://doi.org/10.1146/annurev.iy.13.040195.003255>.
59. Stein KE. Thymus-independent and thymus-dependent responses to polysaccharide antigens. *J Infect Dis*. 1992;165(Suppl 1):S49–52. [https://doi.org/10.1093/infdis/165-supplement\\_1-s49](https://doi.org/10.1093/infdis/165-supplement_1-s49).
60. Rose MA, Schubert R, Strnad N, Zielen S. Priming of immunological memory by pneumococcal conjugate vaccine in children unresponsive to 23-valent polysaccharide pneumococcal vaccine. *Clin Diagn Lab Immunol*. 2005;12(10):1216–22. <https://doi.org/10.1128/CDLI.12.10.1216-1222.2005>.
61. Vadlamudi NK, Chen A, Marra F. Impact of the 13-Valent Pneumococcal Conjugate Vaccine Among Adults: A Systematic Review and Meta-analysis. *Clin Infect Dis*. 2019;69(1):34–49. <https://doi.org/10.1093/cid/ciy872>.
62. Fact sheet: pneumococcal disease, pneumococcal conjugate vaccines, and PNEUMOSIL® | PATH. <https://www.path.org/resources/fact-sheet-pneumococcal-disease-pneumococcal-conjugate-vaccines-and-pneumosil/>. Accessed 28 Apr 2021.
63. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med*. 1944;79(2):137–58. <https://doi.org/10.1084/jem.79.2.137>.
64. Hanage WP, Fraser C, Tang J, Connor TR, Corander J. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science*. 2009;324(5933):1454–7. <https://doi.org/10.1126/science.1171908>.
65. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat Commun*. 2014;5(1):5471. <https://doi.org/10.1038/ncomms6471>.
66. Croucher NJ, Finkelstein JA, Pelton SJ, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet*. 2013;45(6):656–63. <https://doi.org/10.1038/ng.2625>.
67. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science*. 2011;331(6016):430–4. <https://doi.org/10.1126/science.1198545>.
68. Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein JA, et al. Selective and genetic constraints on pneumococcal serotype switching. *Plos Genet*. 2015;11(3):e1005095. <https://doi.org/10.1371/journal.pgen.1005095>.
69. Griffin MR, Zhu Y, Moore MR, Whitney CG, Grijalva CG. U.S. hospitalizations for pneumonia after a decade of pneumococcal vaccination. *N Engl J Med*. 2013;369(2):155–63. <https://doi.org/10.1056/NEJMoa1209165>.
70. Gladstone RA, Devine V, Jones J, Cleary D, Jefferies JM, Bentley SD, et al. Pre-vaccine serotype composition within a lineage signposts its serotype replacement - a carriage study over 7 years following pneumococcal conjugate vaccine use in the UK. *Microb Genom*. 2017;3(6):e000119. <https://doi.org/10.1099/mgen.0.000119>.
71. Lo SW, Gladstone RA, van Tonder AJ, Lees JA, du Plessis M, Benisty R, et al. Pneumococcal lineages associated with serotype replacement and antibiotic resistance in childhood invasive pneumococcal disease in the post-PCV13 era: an international whole-genome sequencing study. *Lancet Infect Dis*. 2019;19(7):759–69. [https://doi.org/10.1016/S1473-3099\(19\)30297-X](https://doi.org/10.1016/S1473-3099(19)30297-X).
72. Pilišvili T, Lexau C, Farley MM, Hadler J, Harrison LH, Bennett NM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *J Infect Dis*. 2010;201(11):32–41. <https://doi.org/10.1086/648593>.
73. Hausdorff WP, Bryant J, Paradiso PR, Siber GR. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis*. 2000;30(1):100–21. <https://doi.org/10.1086/313608>.
74. Hausdorff WP, Bryant J, Kloek C, Paradiso PR, Siber GR. The contribution of specific pneumococcal serogroups to different disease manifestations: implications for conjugate vaccine formulation and use, part II. *Clin Infect Dis*. 2000;30(1):122–40. <https://doi.org/10.1086/313609>.
75. GPS :: Global Pneumococcal Sequencing Project | substudies. <https://www.pneumogen.net/gps/substudies.html>. Accessed 23 Mar 2021.
76. GPS :: Global Pneumococcal Sequencing Project | publications. <https://www.pneumogen.net/gps/publications.html>. Accessed 23 Mar 2021.
77. Corander J, Fraser C, Gutmann MU, Arnold B, Hanage WP, Bentley SD, et al. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nat Ecol Evol*. 2017;1(12):1950–60. <https://doi.org/10.1038/s41559-017-0337-x>.
78. SEQUENCING CORE FACILITY | NICD. <https://www.nicd.ac.za/sequencing-core-facility/>. Accessed 22 Mar 2021.
79. The Bioinformatics Course at MRC Unit The Gambia | MRC Unit The Gambia at LSHTM. <https://www.mrc.gm/bioinformatics-course-mrc-unit-gambia/>. Accessed 22 Mar 2021.
80. Home - H3ABioNet - Pan African Bioinformatics Network for the Human Heredity and Health in Africa. <https://www.h3abionet.org/>. Accessed 22 Mar 2021.
81. Open Access Policy - Bill & Melinda Gates Foundation. <https://www.gatesfoundation.org/about/policies-and-resources/open-access-policy>. Accessed 22 Mar 2021.
82. Open Access Policy - Grant Funding | Wellcome. <https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy>. Accessed 22 Mar 2021.

83. Watanabe ME. The nagoya protocol on access and benefit sharing. *Bioscience*. 2015;65(6):543–50. <https://doi.org/10.1093/biosci/biv056>.
84. J. Page A, Taylor B, A. Keane J. Multilocus sequence typing by blast from de novo assemblies against PubMLST. *JOSS*. 2016;1:118. doi:<https://doi.org/10.21105/joss.00118>.
85. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res*. 2019;29(2):304–16. <https://doi.org/10.1101/gr.241455.118>.
86. Metcalf BJ, Chochua S, Gertz RE, Li Z, Walker H, Tran T, et al. Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. *Clin Microbiol Infect*. 2016; 22:1002.e1–8. <https://doi.org/10.1016/j.cmi.2016.08.001>.
87. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Walker H, et al. Penicillin-Binding Protein Transpeptidase Signatures for Tracking and Predicting  $\beta$ -Lactam Resistance Levels in *Streptococcus pneumoniae*. *MBio*. 2016;7(3). <https://doi.org/10.1128/mBio.00756-16>.
88. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Walker H, et al. Validation of  $\beta$ -lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. *BMC Genomics*. 2017;18(1):621. <https://doi.org/10.1186/s12864-017-4017-7>.
89. Pathogenwatch | Genomes. <https://pathogen.watch/genomes/all?genusId=1301&speciesId=1313>. Accessed 22 Mar 2021.
90. Lo SW, Jamroz D. Genomics and epidemiological surveillance. *Nat Rev Microbiol*. 2020;18(9):478. <https://doi.org/10.1038/s41579-020-0421-0>.
91. Steinig EJ, Duchene S, Robinson DA, Monecke S, Yokoyama M, Laabei M, et al. Evolution and Global Transmission of a Multidrug-Resistant, Community-Associated Methicillin-Resistant *Staphylococcus aureus* Lineage from the Indian Subcontinent. *MBio*. 2019;10(6). <https://doi.org/10.1128/mBio.01105-19>.
92. Ben-Shimol S, Greenberg D, Givon-Lavi N, Schlesinger Y, Somekh E, Aviner S, et al. Early impact of sequential introduction of 7-valent and 13-valent pneumococcal conjugate vaccine on IPD in Israeli children < 5 years: an active prospective nationwide surveillance. *Vaccine*. 2014;32(27):3452–9. <https://doi.org/10.1016/j.vaccine.2014.03.065>.
93. Metcalf BJ, Gertz RE, Gladstone RA, Walker H, Sherwood LK, Jackson D, et al. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. *Clin Microbiol Infect*. 2016;22:60.e9–60.e29. <https://doi.org/10.1016/j.cmi.2015.08.027>.
94. Ho P-L, Law PY-T, Chiu SS. Increase in incidence of invasive pneumococcal disease caused by serotype 3 in children eight years after the introduction of the pneumococcal conjugate vaccine in Hong Kong. *Hum Vaccin Immunother*. 2019;15(2):455–8. <https://doi.org/10.1080/21645515.2018.1526555>.
95. Lo SW, Gladstone RA, van Tonder AJ, du Plessis M, Cornick JE, Hawkins PA, et al. A novel mosaic tetracycline resistance gene *tet* (*S/M*) detected in a multidrug-resistant pneumococcal CC230 lineage that underwent capsular switching in South Africa. *BioRxiv*. 2019. <https://doi.org/10.1101/718460>.
96. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom*. 2016;2(11):e000093. <https://doi.org/10.1099/mgen.0.000093>.
97. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018; 34(23):4121–3. <https://doi.org/10.1093/bioinformatics/bty407>.
98. GPS :: Global Pneumococcal Sequencing Project | countries. [https://www.pneumogen.net/gps/sampling\\_map.html](https://www.pneumogen.net/gps/sampling_map.html). Accessed 22 Mar 2021.
99. GPS :: Global Pneumococcal Sequencing Project | strains. [https://www.pneumogen.net/gps/GPSC\\_lineages.html](https://www.pneumogen.net/gps/GPSC_lineages.html). Accessed 22 Mar 2021.
100. GPS :: Global Pneumococcal Sequencing Project | resources. [https://www.pneumogen.net/gps/resources\\_overview.html](https://www.pneumogen.net/gps/resources_overview.html). Accessed 22 Mar 2021.
101. Gladstone RA, Lo SW, Goater R, Yeats C, Taylor B, Hadfield J, et al. Visualizing variation within Global Pneumococcal Sequence Clusters (GPSCs) and country population snapshots to contextualize pneumococcal isolates. *Microb Genom*. 2020;6(5). <https://doi.org/10.1099/mgen.0.000357>.
102. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics*. 2018;34(2):292–3. <https://doi.org/10.1093/bioinformatics/btx610>.
103. Phandango | GPS. <https://jameshadfield.github.io/phandango/#/gps>. Accessed 22 Mar 2021.
104. JUNO project - A global genomic survey of *Streptococcus agalactiae*. <https://www.gbgsen.net/>. Accessed 22 Mar 2021.
105. Bioinformatics Training. <https://training.bactgen.sanger.ac.uk/#/>. Accessed 22 Mar 2021.
106. GPS :: Global Pneumococcal Sequencing Project | training. [https://www.pneumogen.net/gps/training\\_drag\\_and\\_drop.html](https://www.pneumogen.net/gps/training_drag_and_drop.html). Accessed 22 Mar 2021.
107. Agnemesel A, Hong E, Giorgini D, Nuñez-Samudio V, Deghmane A-E, Taha M-K. *Neisseria meningitidis* Serogroup X in Sub-Saharan Africa. *Emerg Infect Dis*. 2016;22(4):698–702. <https://doi.org/10.3201/eid2204.150653>.
108. Immunization, Vaccines and Biologicals. <https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases>. Accessed 23 Mar 2021.
109. McNally A, Kallonen T, Connor C, Abudahab K, Aanensen DM, Horner C, et al. Diversification of Colonization Factors in a Multidrug-Resistant *Escherichia coli* Lineage Evolving under Negative Frequency-Dependent Selection. *MBio*. 2019;10(2). <https://doi.org/10.1128/mBio.00644-19>.
110. Colijn C, Corander J, Croucher NJ. Designing ecologically optimized pneumococcal vaccines using population genomics. *Nat Microbiol*. 2020; 5(3):473–85. <https://doi.org/10.1038/s41564-019-0651-y>.
111. Croucher NJ, Campo JJ, Le TQ, Liang X, Bentley SD, Hanage WP, et al. Diverse evolutionary patterns of pneumococcal antigens identified by pangenome-wide immunological screening. *Proc Natl Acad Sci USA*. 2017; 114(3):E357–66. <https://doi.org/10.1073/pnas.1613937114>.
112. Advanced Molecular Detection (AMD) and Response to Infectious Disease Outbreaks. <https://www.cdc.gov/amd/index.html>. Accessed 23 Mar 2021.
113. Implementing pathogen genomics: a case study - GOV.UK. <https://www.gov.uk/government/publications/implementing-pathogen-genomics-a-case-study>. Accessed 23 Mar 2021.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.