UNIVERSITAT DE BARCELONA

Impact of molecular methods in the analysis of the invasiveness of *Streptococcus pneumoniae*

Eva del Amo Morán

INTRODUCTION

INTRODUCTION

1. *Streptococcus pneumoniae*

Streptococcus pneumoniae, the pneumococcus, is a Gram-positive bacterium that although is more commonly found colonizing the nasopharynx of humans can invade sterile fluids and cause lethal diseases. Pneumococcal diseases vary from mild respiratory tract mucosal infections such as otitis media and sinusitis to more severe diseases such as pneumonia, bacteremia, sepsis and meningitis. Due to this range of diseases this pathogen is a major cause of morbidity and mortality worldwide being the most affected populations the children and the elderly. It is estimated about 500.000 deaths annually of children under five years (1).

Humans are the natural reservoir of pneumococcus; it usually exists in the nasopharynx of humans and it is transmitted from another human carrier. The transmission is produced from person to person via aerosols, and nasopharyngeal colonization is an essential first step prior to pneumococcal infection. Once acquired, an individual strain can be carried for weeks to months before its clearance. Pneumococcal carriage occurs early in life with a carriage rate of 30-60% in children, acquiring one or many strains sequentially or simultaneously, that drastically diminishes in adults to a 1-10%, depending on the presence of children in the household.

First, pneumococcus enters the nasal cavity and is attached to the epithelial cells of the nasopharynx. From this point, the bacteria can stay as a colonizer or spread to other organs like ears and sinuses or to the lungs. Once in the lungs pneumococcus can cross the mucosal barrier and spread to the blood stream, and from there it can cross the

blood-brain barrier and produce the most severe outcome, the meningitis (Figure 1) (2-3).

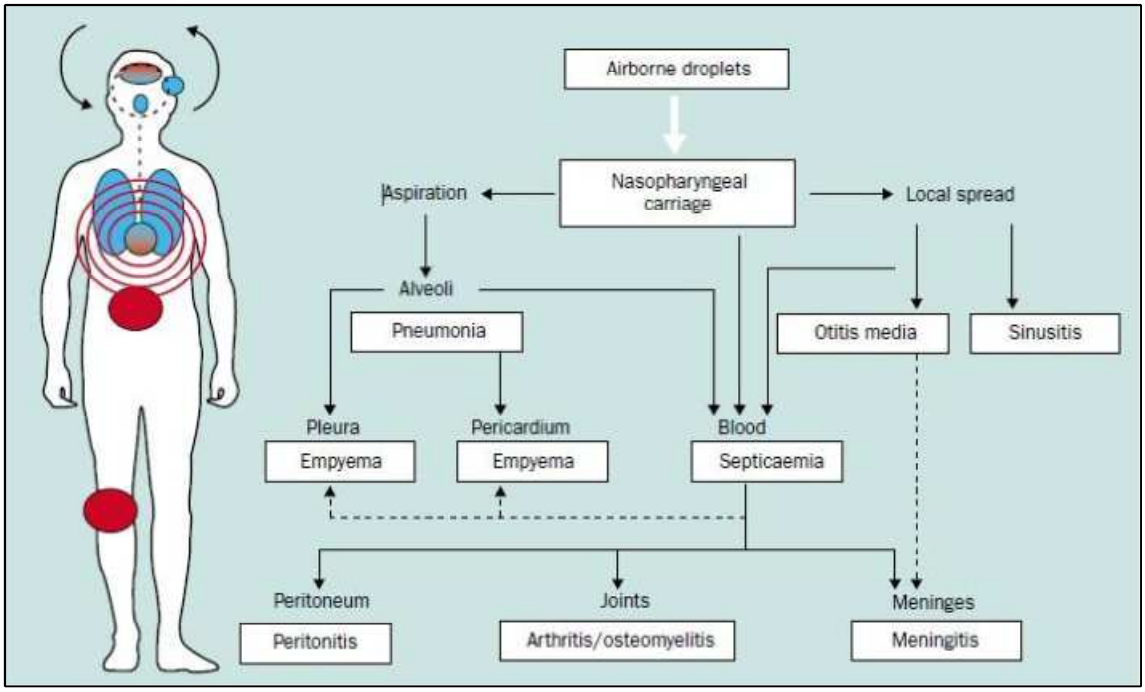

Figure 1.*Pathogenic route for Streptococcus pneumoniae infection (Data from: Bogaert D et al. 2004).*

2. The pneumococcal polysaccharide capsule

To be able to cause infection the pneumococcus produces a range of virulence factors that include a polysaccharide capsule, surface proteins and enzymes, and the toxin pneumolysin (*ply*) (Figure 2). The polysaccharide capsule in the most important virulence factor and therefore it is the factor that has been more studied. The importance of the capsule to virulence is due to its antiphagocytic activity. The capsule forms an inert shield that prevents the Fc region of immunoglobulin G and the complement component iC3b from interacting with their relevant receptors on the surface of phagocytic cells, with the result that the organism remains extracellular (3).

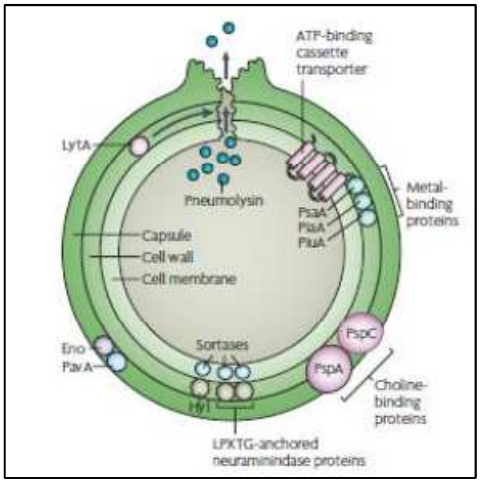

Figure 2.*Pneumococcal virulence factors (Data from: Kadioglu A et al. 2008).*

Although mainly studied for its role in infection, the capsule also plays an important role in colonization preventing the clearance of the pneumococcus in the nasopharynx by the mucus, restricting the autolysis and reducing the exposure of the bacteria to the antibiotics (4).

Until now about 97 different capsular serotypes/capsule structures have been identified using the immunochemical differences between the capsular polysaccharides (5). Despite this great variety of capsular types only few of them are responsible for the main burden of Invasive Pneumococcal Disease (IPD) (6). Not all pneumococcal serotypes share the same ability to invade and cause disease.

3. The invasive disease potential of serotypes

The relationships between strains and their potential to cause disease were evaluated and a major finding was that strains vary substantially in their invasive potential (7-8). Some serotypes are more prone to be found causing invasive disease where others are more associated with carriage. This invasive property, or Invasive Disease Potential (IDP), of a serotype seems to be mainly associated to the capsule and to be independent

of time and geography. Also, it provides a useful approach for comparing the behavior of the different pneumococci (9).

Different methods have been developed to measure the IDP (also denominated invasive capacity or attack rate) of each pneumococcal serotype. So far, the most frequently used method for epidemiological studies has been the invasive odds ratio (OR). The invasive OR is calculated by reference to all the other serotypes as follows: OR = $(ad)/(bc)$, where a is the number of invasive A serotypes, b is the number of carriage A serotypes, c is the number of invasive non-A serotype, and d is the number of carriage non-A serotypes (10). Once this ratio was calculated to establish the invasiveness of each known serotype, a classification was made to establish which serotypes represented the highest risk to the population.

In 2003, Brueggemann et al. published one of the first studies to characterize a collection of invasive and carriage pneumococci recovered from a population of young children in Oxford, United Kingdom. In this study, odds ratio (ORs) were calculated to estimate the IDP of each serotype (7). A limitation of the study was that it represented data from only one population of children so results could not be generalized to other populations. For this reason, the same group published a new study in 2004 where they performed a meta-analysis using data set from Oxford, Alaska and Iceland and data from studies previously published in the medical literature. The aim of the study was to calculate and compare serotype and serogroup-specific ORs for invasive disease for each individual study and for the pooled data. To calculate the OR of each serotype and because the quantity of each serogroup varied greatly between studies, also some serogroups that contain multiple serotypes were not stratified by serotype, they chose to use a fixed reference set instead. Serotype 14 was selected as the referent serotype. Therefore, each OR was defined as an estimate of serotype or serogroup-specific IDP

relative to the IDP of serotype 14 (8). This classification of pneumococcal serotypes invasiveness has been used as a reference in subsequent epidemiological studies.

4. Pneumococcal Vaccines

The identification of the serotypes that present the highest potential to cause invasive disease has been of especial interest to optimize the design of pneumococcal vaccines. Especially because current pneumococcal vaccines can only include a limited number of polysaccharides due to the complexity and high cost of the production.

In 1983, a pneumococcal polysaccharide vaccine (PPV) containing the capsular polysaccharides of 23 serotypes (PPV23) was approved. Although this vaccine presented serotype coverage of more than 85% of the organisms causing IPD at the time, the poor immunogenicity of T-cell-independent PPV23 in infants led to the development of the pneumococcal conjugate vaccines (PCV) (10).

PCV consists of an immunogenic but inert carrier protein (a non-toxic diphtheria-toxin variant carrier protein) coupled covalently to the polysaccharide antigen of the selected strains of pneumococci to improve the immune response. This linkage converts the polysaccharide into a T-cell-dependent antigen that induces a more powerful respond in children, resulting in development of immunologic memory and maturation of the immune response (Figure 3). The development of immunologic memory implies that protection against the pneumococcus does not entirely depend on existing antibody concentrations; instead, the vaccinated individual can respond with a rapid, high, and effective antibody response to invasion by the infectious agent.

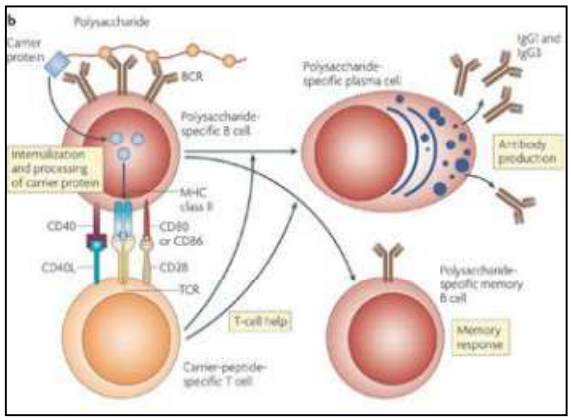

Figure 3. *Immune response of the host to PCV (Data from: Pollard AJ et al. 2009)*

The main target population of the vaccine is infants and young children, since this age group is the main responsible for much of the community spread of the pathogen and is the group at highest risk for the disease, but in the last years the routine use of these vaccines has been also approved and recommended in adults aged ≥65 years (11).

For the first conjugate vaccine, PCV7, the seven serotypes that were the most predominant in the United States (US) were chosen: 4, 6B, 9V, 14, 18C, 19F and 23F. PCV7 was licensed in the US in 2000 and introduced in the children vaccination program. It was recommended for routine administration as a four-dose series for children at ages 2, 4, 6 and 12-15 months (12).

An important decrease in vaccine serotypes causing invasive pneumococcal disease was observed in most of the countries that introduced the vaccine (even in countries with limited coverage), not only in children under 2 years old but also in unvaccinated children and adults (especially in those ≥65 years old) (13-17).

The decline of the disease in the population not vaccinated is explained by the reduction of the presence of these serotypes in the nasopharynx of the vaccinated children, which reduces the possibility of transmission to non immunized individuals (Figure 4). This indirect effect or herd protection has been observed in unvaccinated children that are in

close contact with vaccinated children or living in communities where PCV7 has been widely use, experiencing them as well a reduction in nasopharyngeal pneumococcal carriage and disease caused by vaccine serotypes.

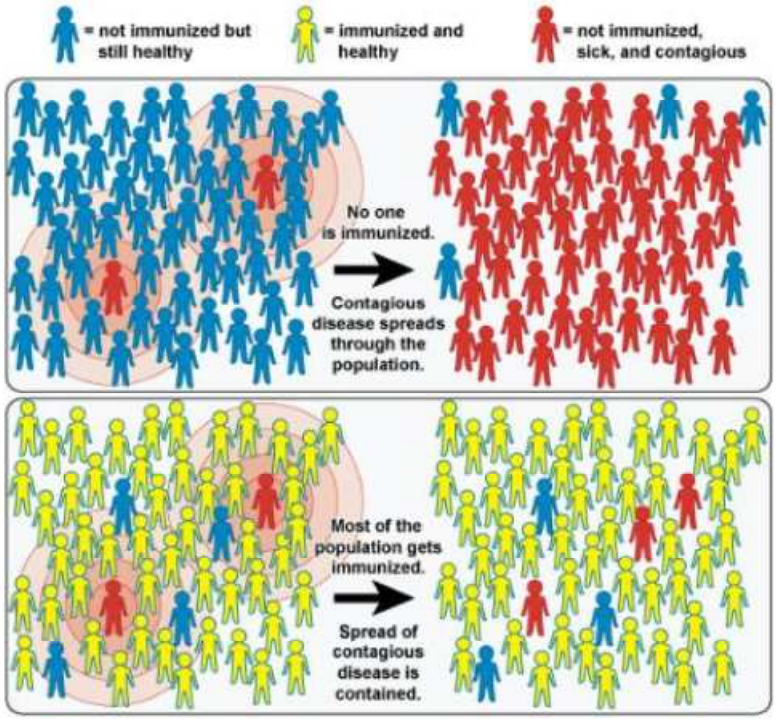

Figure 4.*Mechanism of the herd protection (Data from: http://www.vaccines.gov/basics/protection/)*

Another important indirect effect observed after the introduction of the vaccine was the reduction in the incidence of pneumococci presenting antimicrobial resistance, both in invasive disease and in carriage. First, because antimicrobial resistance was predominantly associated to the serotypes included in PCV7, and secondly, because the general decrease of the disease in children involved a reduction in antibiotic use. The widespread use of antibiotics in children is in large part the cause of the increased global prevalence of pneumococcal antibiotic-resistant strains (18).

Despite its great efficiency worldwide, it was observed that the vaccine serotype coverage of PCV7 was not as high in other countries as in the USA. The distribution of pneumococcal serotypes causing IPD and nasopharyngeal carriage is different according to geographical region and can even differ in different country settings (12,

19-20). In contrast to the overall decline in IPD, data from countries such as France, Spain, and Alaskan natives, suggest a lower impact of vaccination with PCV7 in children. Moreover, and regardless of the success of the vaccine, an overall increase of IPD caused by non-vaccine serotypes was observed (21-24). Especially, serotypes 1 and 19A experienced a significant increase (25-28).

For these reasons, two more conjugate vaccines were introduced in 2009 and 2010, PCV10 and PCV13, to improve the coverage of the vaccine. This new vaccines included serotypes 1, 3, and 5 that are common in Europe, Asia, and Africa, and also serotypes 6A, 7F and 19A, that have emerged after the implementation of PCV7.

Despite differences in immunization programs, most studies on the effectiveness of PCV13 against IPD have proved that this vaccine has significantly reduced the incidence of IPD in both vaccinated children and unvaccinated populations, with the most pronounced decline for serotype 19A. Results of these studies were independent of countries which demonstrate that PCV13 provides a wider and more optimal coverage against pneumococcal disease than PCV7. Serotype 3 seems to be the only added serotype to not experience a significant change which suggests marginal PCV13 activity against this serotype. Although these studies have also shown a significant increase in the incidence of IPD caused by non-vaccine serotypes, any of these serotypes have emerged like serotype 19A after PCV13's introduction. However, their rates should be monitored carefully in the future as possible replacement serotypes (29-34).

5. Detection of *Streptococcus pneumoniae*

Traditionally, *Streptococcus pneumoniae* has been diagnosed by standard methods only based on culture. Several assays are used for the laboratory diagnosis of the pathogen, and although there is no gold standard method, there are some minimum tests that are

usually involve in the identification of an isolate of *Streptococcus pneumoniae*: colony morphology (the isolate has to be mucoid or formed shiny young colonies and/or had a depressed center on sheep blood agar), hemolysis (be alpha-hemolytic), Gram stain morphology (to form gram-positive cocci in chains), and either deoxycholate solubility (be bile soluble), optochin susceptibility, or latex agglutination (reaction with specific antisera) (Figure 5) (35).

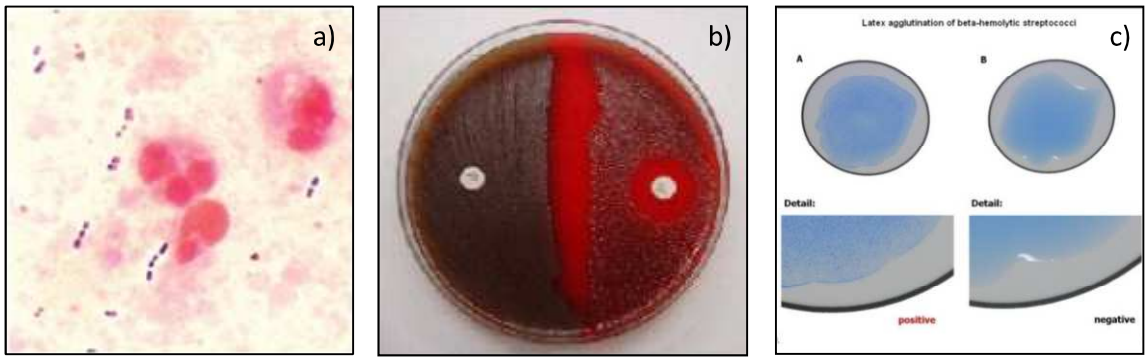

Figure 5.*Standard methods for the identification of Streptococcus pneumoniae: a) Gram stain morphology, b) optochin sensibility (left a negative result, right positive result) and c) latex agglutination (Data from: Werno AM et al. 2008, http://www.cdc.gov/meningitis/lab-manual and http://www.bacteriainphotos.com/).*

Methods based on culture present important advantages such as the ability to be implemented all over the world with low cost and high specificity, and that they allow obtaining antibiotic sensitivity. However, several studies have shown the sensitivity problems of culture. The difficulty to recover the pneumococcus from clinical samples is caused by its tendency to autolyse when reaching the stationary phase of growth and due to antibiotic treatment prior to the collection of the sample (36). Low rates of bacteremia in patients with IPD have been observed, mostly due to the effect of these factors, causing that the rates of *Streptococcus pneumoniae* isolation in blood cultures of community-acquired pneumonia are only of 3-8% and even lower in children (37-

38). In pediatric patients, it is especially problematic since it is not always possible to collect an adequate volume of the sample to perform the blood culture.

Another problem in the identification by standard methods has been the appearance of pneumococcus-like organisms. These specimens can give a positive reaction in one or more of the standard tests leading to a misidentification (39). The development of molecular techniques for the detection of *Streptococcus pneumoniae* in the last years has led to an improvement in the detection of the pathogen.

The main advantage for the identification of *Streptococcus pneumoniae* by molecular techniques, compared to phenotypic assays, lays in the fact that they detect the DNA of the pneumococcus, so even if the sample is taken after the patient has been exposed to antibiotics the bacteria can be detected. There is no need for the bacteria to be alive because there is no need to grow the strain, they do not depend on the viability of the pathogen. This is an important advantage since patients use to receive an empirical treatment with antibiotics, especially children, by doctors before having a definitive diagnose in order to prevent a worst progression when infection by pneumococcus is suspected (40). As commented previously, this protocol has an important impact in the sensibility of the culture. Since molecular methods have been added, the identification of *Streptococcus pneumoniae* in sterile fluids has been greatly improved demonstrating that only with the classical methods the rate of the disease was underestimated (41-44).

Other advantages involving the introduction of molecular testing have been the reduction in the time frame in which results are delivered and their ability to detect very small amounts of nucleic acid from the pathogen (37). The classical methods require, at least, a 24-48 hours' time period before confirmation of the presence of pneumococcus,

the time needed for the strain to grow. With the introduction of molecular methods this time period has been reduced to a few hours.

The first step in the implementation of the new techniques was through the introduction of the Polymerase Chain Reaction (PCR) (Figure 6). The detection by PCR is based in the amplification of a gene that is only present in the pneumococcus, and the products are visualized by agarose gel electrophoresis. The first target of this amplification was the pneumolysin gene (*ply*), initially it thought to be specific to this bacterium. However, over the years it was observed that this gene was also being detected in non-pneumococcal specimens. To obtain a better target, a more specific target, several genes were tested and these studies demonstrated that the autolysin gene (*lytA*) was the most reliable (36, 45).

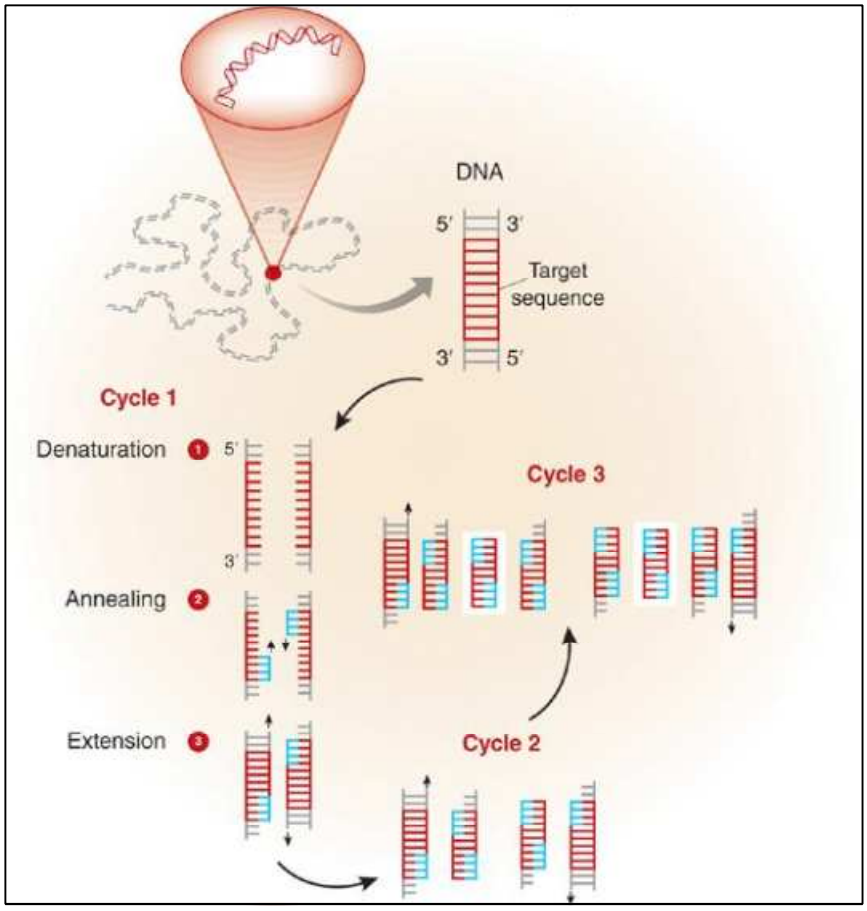

Figure 6.*Principle of the Polymerase Chain Reaction (Data from: Garibyan L et al. 2013).*

Although this first approach was very effective for the identification of the organism, it was not the most ideal choice for rapid diagnosis. To improve the technique, a simple, more sensitive, more rapid and more accurate assay was developed, the real-time PCR. The real-time PCR combines amplification and detection in one step through the use of fluorescent dyes (Figure 7). The advantages of this PCR are its speed, which is higher since the need for an electrophoresis gel for observing the final results is eliminated; its wider dynamic range, its ability to detect larger variations in concentrations of the target; its lower limit of detection; and its higher sensitivity and specificity, due to the incorporation of a probe that hybridizes into a specific DNA sequence. In addition, real-time PCR is performed in closed tube format which reduces the risk of contamination between steps. The greater sensitivity and specificity present in the real-time PCR allows its use directly from clinical samples for both diagnosis and carriage studies, whether they are obtained from nonsterile or sterile sites. The real-time PCR assay *lytA* primers and probes have been found to be extremely reliable for detection of *Streptococcus pneumoniae* (39, 46-47).

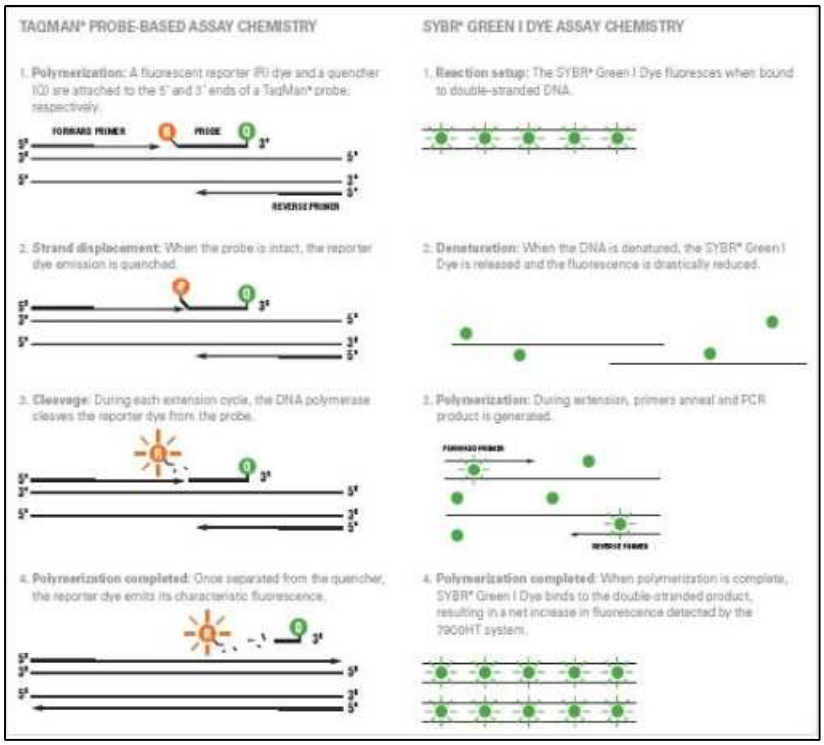

Figure 7. *Real-Time PCR chemistries (Data from: https://www.lifetechnologies.com)*

Even though molecular assays offer such important advantages, cultural methods remain necessary to obtain the information about antibiotic resistance.

6. Detection of pneumococcal serotypes

The increased detection of the bacteria by molecular methods has not only had an impact in diagnose but also in the identification of the different capsular types. As important as the detection of the pneumococcus is the correct identification of its capsular types. Not only for epidemiological studies of the trends of the bacteria but also for the development of the vaccines and to assess its effectiveness. The implementation of the new techniques has shown that the presence of some serotypes has been underestimated because of the difficulty of being cultured (48). Moreover, higher levels of detection of the bacteria allow higher levels of serotype identification (49).

Despite the existence of various phenotypic methods, the gold-standard for pneumococcal capsular typing has traditionally been the Quellung reaction, an antigen-antibody based reaction. A positive Quellung reaction is observed when a type-specific antibody binds to the capsule of the pneumococcus causing a change in its refractive index. In the microscope the bacteria looks “swollen”, being more visible than in its normal state (Figure 8) (50-51). The main advantage of this technique is that allows the identification of all the existing capsular serotypes which have been identified up to the present day, besides presenting a quite good specificity. The disadvantages are that it is labor-intensive, requires expensive reagents, needs experienced personal to identify the reaction in the microscope (results can be difficult to interpret, but it also means a subjective interpretation of the results from the technician), and that it is essential the

isolation of the strain (52). Also, some cross reactions with other streptococcal polysaccharides have been reported and those strains that are unencapsulated will show false-negative results, since this is a phenotypic test.

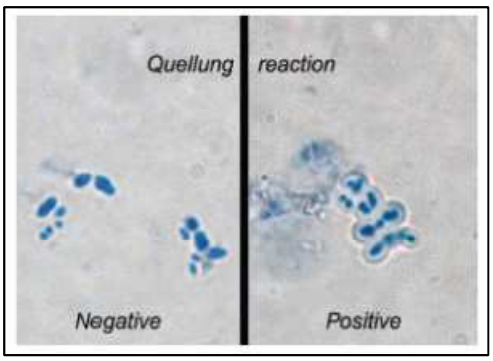

Figure 8. *Quellung reaction (Data from: Werno AM et al. 2008).*

With the introduction of molecular techniques, new approaches were made to distinguish the serotypes of the pneumococcus targeting the polysaccharide capsule synthesis genes at the cps locus. The first four genes are conserved in almost all serotypes but the central parts of the loci contain the serotype-specific genes that can be used to differentiate the pneumococci (53-54). In 2006, Bentley et al. (55) published the capsular sequences of 90 pneumococcal serotypes making possible the design of serotype-specific primers and thereby the genetic identification of serotypes. In recent years, several methods have been developed including approaches based on different PCR strategies, multiplex PCR, restriction fragment length polymorphisms, hybridization assays, microarrays, and sequencing (56).

6.1. Multiplex PCR (mPCR): This method is based on the amplification of short sequences specific to individual serotypes or serogroups. The primers were grouped into different reactions based on serotype distributions among invasive pneumococci recovered in the United States, and in each reaction a primer pair targeting the *cpsA* locus found in all pneumococci was included as an internal

control. The PCR products were analyzed by gel electrophoresis and identified according to size. This method allows the simultaneous detection of several serotypes and the primer specificity enables performing the detection directly from clinical samples (53). This protocol has been adapted and extended, and actually identifies a total of 40 pneumococcal serotypes. The mPCR is currently the recommended typing technique by the Center for Disease Control and Prevention (CDC).

6.2. PCR and Restriction Fragment Length Polymorphisms (PCR-RFLP): This method is based on visualizing digested PCR amplicons fragments. First, the gene cassette containing the genes encoding the pneumococcal capsule is amplified using primers for *dexB* and *aliA*, the two genes that flank the gene cassette. Then, the amplicons from the PCR are digested using HinfI enzyme, and the digests are immediately subjected to gel electrophoresis. To identify the serotype of an unknown sample, cluster analysis is performed. A comparison is created with the unknown strain using one example of every type and group in the database. If a match with similarity larger than 90% is detected between the band patterns, it is defined as the serotype of the strain. This method is capable of discriminating 46 serotypes but the results are image-based which can create difficulties to compare data between laboratories (57).

6.3. Multiplex PCR with Fragment Analysis and automated Fluorescent capillary electrophoresis (FAF-mPCR): This method combines the multiplex PCR with fragment analysis and automated fluorescent capillary electrophoresis using a genetic fragment size analyzer. This assay is able to identify 40 different serotypes. Forward primers of each pair of primers are fluorescently labeled. Then, primers are grouped together in three different reactions based on the size

of the fragment and the fluorophore selected. Distinct fluorophores are used in the products with similar sizes in order to obtain a better discrimination between PCR products. Also, primers specific to the *cpsA* gene are included in each reaction as an internal positive control. After mPCR, amplified products are separated using capillary electrophoresis and the data are analyzed by using automating software. This system recognizes fluorescent peaks according to specific color and size and assigns the peak to the corresponding capsular type. This method has the capacity to detect multiple colonization but has the limitation of requiring a genetic analyzer (58).

6.4. PCR and Electrospray Ionization Mass Spectrometry (PCR/ESI-MS): This method combines PCR and Electrospray Ionization Mass Spectrometry. It involves distributing a DNA extract into wells of a microtiter plate, each of which contains a set of serotype specific primers and a set of internal positive control primers. After amplification, PCR products are electro sprayed into a mass spectrometer and the base composition of each DNA fragment is determined based on the mass/charge ratio. Molecular signatures of amplified products from unknown samples are compared to a database of signatures obtained from known capsular loci. It allows the identification of 45 different serotypes, and at the same time it identifies of the genotype. This method can be potentially used to detect more than one serotype in the same sample, and to detect serotypes from culture-negative clinical specimens. The limitation is the requirement of an electrospray mass spectrometer (Figure 9) (59).

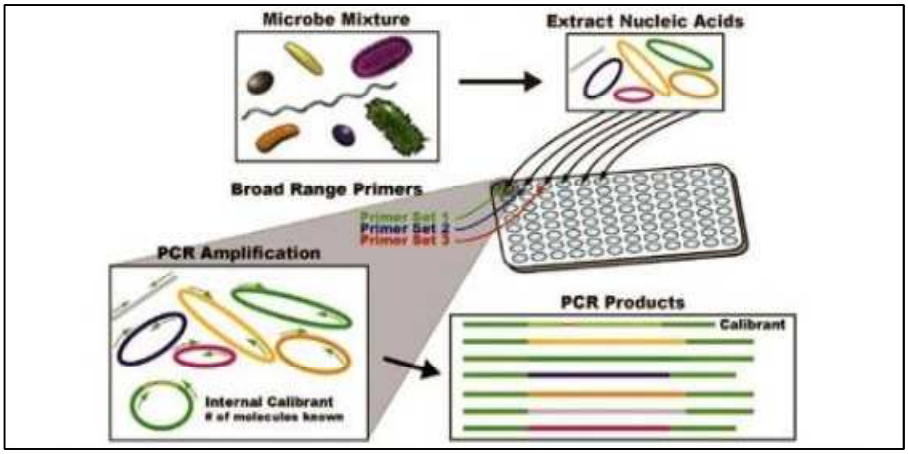

Figure 9. *General scheme of PCR-ESI/MS (Data from: https://www.aacc.org).*

6.5. mPCR and Reverse Line Blot hybridization (mPCR/RLB): Multiplex PCR and Reverse Line Blot Hybridization allows the identification of 90 pneumococcal serotypes using serotype-specific DNA probes. This method is based on one multiplex PCR reaction followed by probe hybridization on a nylon membrane. Amine-labeled probes are covalently bound to a nylon membrane. Biotin modified PCR products, mPCR primers are 5' biotin modified, are hybridized to the probes. Streptavidin, labeled with peroxidase, is incubated with the membrane and binds to biotin. Peroxidase acts as a catalyst in a light-producing reaction to which a light sensitive film is exposed. The pattern on the film is analyzed and compared with positive control patterns to assign a serotype. This technique uses reagents and techniques that are available in many laboratories, it only requires a single mPCR/RLB reaction, and has the potential to identify multiple serotypes directly from clinical specimens (Figure 10) (60-62).

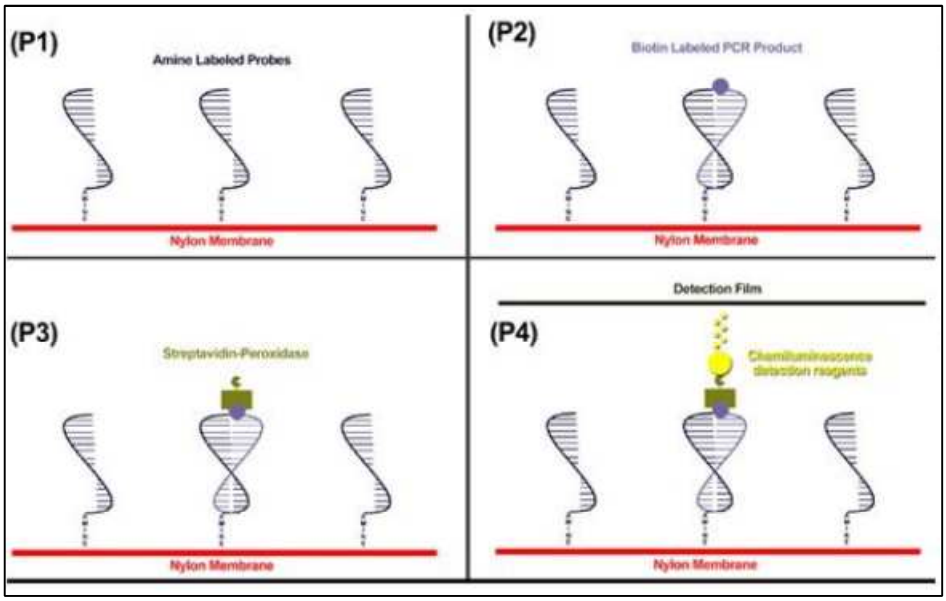

Figure 10.*The mPCR/RLB principles (Data from: O'Sullivan M et al. 2011).*

6.6. Sequential Real Time PCR (RT-PCR): To optimize the sensitivity and specificity of the conventional PCR and improve the capsular typing directly from clinical samples, specific probes were added to the specific primers for 24 different serotypes in this method. Besides the speed, no gel is needed to analyze the PCR products; real-time PCR offers a lower limit of detection and more specificity since hybridization to a probe is required in addition to the amplification of the primers (63). Recently, a triplex real-time multiplexed PCR that allows the identification of 21 serotypes was developed (64). Both protocols are available on the CDC website.

6.7. Nanofluidic Real Time PCR: Nanoliter-scale quantities of samples and reagents are channeled into thousands of nanoliter-scale chambers in which distinct real-time PCRs can be run. This method relies on single molecule detection since the PCR solution is distributed across a large number of partitions and following amplification linear, digital signals are used to estimate DNA copy number. This technique allows the identification of 50 serotypes and it can detect minor

population of multiple serotypes in co-colonization, providing both qualitative and quantitative data. The main limitation of the assay is the high cost of purchasing the nanofluidic equipment (Figure 11) (65-68).

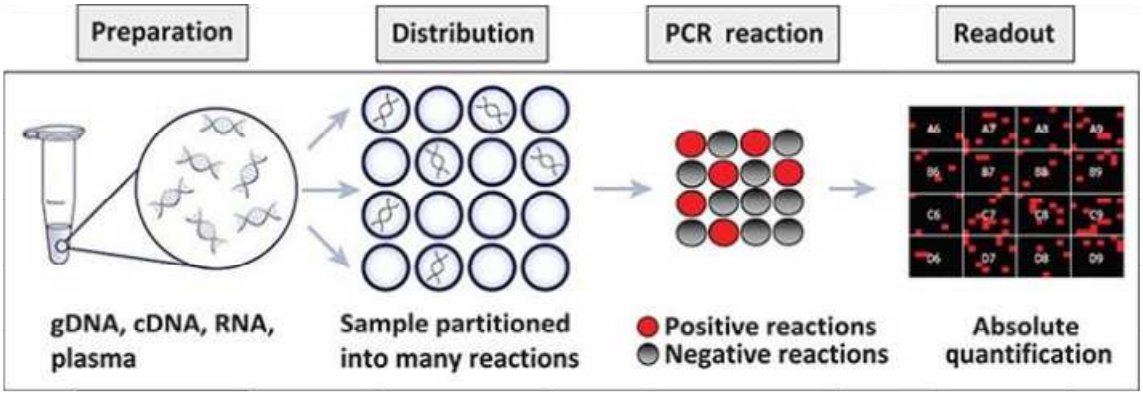

Figure 11. *Steps of the nanofluidic PCR reaction (Data from: http://www.lifetechnologies.com).*

6.8.Microarray: This method uses specific probes to construct a serotyping microarray that detects the highly variable glycosyltransferase genes *wzx* and *wzy*, which contain serotype or serogroup specific regions. These genes correspond to key enzymes in the distinct sugar composition of the different capsular polysaccharides. DNA samples for analysis are labeled with fluorescent dyes and added to the array chip to bind the complementary sequence before the microarray run. After the microarray run, the fluorescence pattern is then recorded by a scanner, quantified, and analyzed. The advantages of the microarray are the ability to detect all known pneumococcal serotypes and to estimate their relative abundances in a mixed population, and the ability to detect multiple pneumococcal serotypes in carriage. The main limitation is the high cost per test and that requires well-experienced technicians (Figure 12) (69-71).

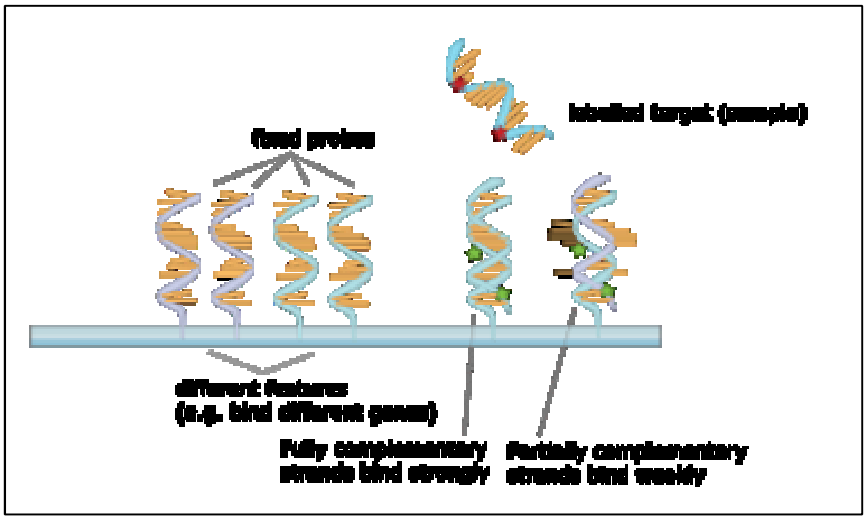

Figure 12. *Process of the microarray method (Data from: http://www.biotechspace.site90.com/).*

6.9. Sequetyping: This method is based in a single PCR amplification followed by cycle sequencing of the purified amplicons to obtain the complete sequence. The optimal primer pair was based on a region unique in sequence in many serotypes that is flanked by conserved primer binding sites in the *cps* locus, a region present only in the pneumococcus. After amplicons are sequenced, their sequences are compared using the Basic Local Alignment Search Tool (BLAST) to identify the corresponding serotype. Although simple and economical, the main limitation of the assay is that the targeted gene is not specific enough to identify a large group of serotypes (72).

6.10. Target enrichment based next generation sequencing: This method is an alternative to whole genome sequencing that combines the use of target enrichment and sample pooling as an option to sequence only the targets of interest at a higher depth. Simple multiplex PCR is used for target enrichment. Two PCR primer pairs are used; one PCR to identify the presence of pneumococci and the second PCR to identify the serotype. This assay allows pooling multiple samples, which lower the cost, and has the potential to include

more serotypes and to be used directly in clinical samples. However, it requires the use of a next generation sequencer along with bioinformatics expertise (73).

6.11. Serotype inference from whole genome sequencing: this method is based in sequencing the entire pneumococcal genome to analyze the data of the whole capsular locus using a number of bioinformatics approaches. The sequence can be examined in detail, confirming serotype-specific gene sequences and single nucleotide polymorphisms (SNPs) to distinguish members of serogroups. Although whole genome sequencing has become more affordable, an important limitation is the need of personal with experience in sample preparation for sequencing and most important in bioinformatics, and that it cannot be performed directly in clinical samples.

Despite the great improvements that molecular techniques offer, the Quellung reaction is still needed. On one hand, to resolve the non-typeable isolates that present a positive *cpsA* result in the PCR reaction but a capsule not included in the range of serotypes that the technique can detect; and in the other hand, to identify serotypes belonging to the same serogroup that the PCR cannot distinguished (74-75).

7. Detection of multiple colonization or co-colonization

As previously described, the most common state of the pneumococcus is colonizing the human nasopharynx and it is a necessary previous step to the development of the invasive disease. Colonization by pneumococcus usually happens during childhood. The nasopharyngeal niche becomes colonized during the first year of life and then pneumococcal carriage shows an increased before the age of 2 years (76). One or several episodes of colonization by different serotypes have been described in children,

and even colonization by more than one serotype at the same time has been reported (multiple colonization or co-colonization).

Although detected worldwide, different rates of carriage and levels of detection of multiple colonization has been found in different countries. Data about the epidemiology of multiple colonization are limited and prevalence rates vary from 1.3% to 30% (77-78). In countries with high levels of carriage the detection of co-colonization has also been higher. For example, in a recent study conducted in sub-Saharan Africa using microarrays a 19% rate of colonization by multiple serotypes was detected compared to a 7.9% reported in a study conducted in Switzerland (79). Besides of geographical variations, the different techniques employed in those studies are the main responsible of such variations.

As previously indicated the need for techniques that allow the correct identification and serotyping of *Streptococcus pneumoniae* is vital to the surveillance of the pathogen and to evaluate the effect of the conjugate vaccines. Assays that have the ability to detect the presence of multiple colonization are indispensable to obtain the correct picture of the nasopharynx population. Pneumococcal conjugate vaccines have been proven to have an effect in nasopharyngeal carriage. To monitor and to predict the possible replacement that can follow the elimination of the vaccine serotypes, is necessary to have an accurate vision of the strains circulating in the population. Especially since non vaccine serotypes that could have been carried at low density may become important disease-causing serotypes once vaccine serotypes are removed from the nasopharyngeal niche. Moreover, the evolution of *Streptococcus pneumoniae* occurs by lateral gene transfer during the carriage state and more likely the simultaneous presence of at least two pneumococcal strains in the nasopharynx are required for the gene transfer. This horizontal exchange of genes can result in acquisition of antibiotic resistance or in

capsular switch. Therefore, analysis of the co-colonization rates after introduction of the vaccines and a study of the serotypes detected in co-colonized samples can offer an insight in the pneumococcal evolution (77,80). Finally, studies which estimate the invasiveness of serotypes are based in the ability of pneumococcal strains to progress from nasopharyngeal carriage to invasive disease so an accurate measurement of carriage is needed for a precise estimation.

Detection and serotyping of the nasopharyngeal carriage of pneumococci have traditionally been performed by conventional culture methods of direct plating of nasopharyngeal specimens from transport or storage medium. However, these standard approaches have proved to not be the most optimal for the detection of multiple serotypes. According to published data, to detect a minor carried serotype it would be necessary to serotype at least five colonies to have a 95% chance of detecting the serotype if this accounted for 50% of the total pneumococcal population. If the serotype was present at an abundance of approximate 1%, 299 colonies would need to be examined. Besides, this method depends in part on the expertise of the technician to recognize colony differences when subculturing two to four colonies for typing. This approach has a low sensibility and is labor-intensive and time-consuming, which becomes impractical to analyze large collections of specimens (70).

Several alternative techniques have been describe to improve the detection of multiple carriage such as immunoblotting or multiplex PCR from the primary culture, but even with previous broth enrichment there is still the limitation of the low sensibility (81-83). Culture methods will detect the most predominant serotype but statistically it cannot be expected that low-abundance serotypes are detected. Therefore, the assays with the highest potential have been the molecular techniques that can be applied directly from the nasopharyngeal swab. These methods are faster, more sensitive and permit the

processing of many samples simultaneously. These assays have proved to increase the detection rate of *Streptococcus pneumoniae* and to be able to identify a greater number of serotypes. However, they present some limitations such as a relatively high cost per test in some of the techniques or a limited number of serotypes that can be detected (44, 84).

8. Clonal or genetic diversity in pneumococcal serotypes

Variations in the capsular type are the major determinant of the IDP; however, several studies have suggested that differences in IDP are also influenced by different clonal types. Isolates belonging to the same clonal complex (strains sharing the same ancestor) may differ in their ability to cause invasive disease (85). Even though the results about the association between specific clonal types and an ability to colonize or cause invasive disease have not been statistically significant, what molecular methods have been able to demonstrate is that isolates of the same serotype can belong to either the same or different clonal types. Whether the different clonal types that composed the different serotypes have different invasive-disease potentials and this determines the invasiveness of the serotype is still debated (86-87).

The clonal type indicates the genetic relatedness of strains within one serotype. Studies analyzing the clonal characteristics of serotypes using molecular methods have allowed the analysis of the genetic relatedness between clinical isolates. The results obtained from these analysis have revealed that serotypes with a high IDP are more genetically related (in most of the isolates the same clonal type was identified), whereas isolates belonging to serotypes with less invasive potentials are much more diverse (a wider range of clonal types was detected) (88-90). This observation has been explained by the

fact that recombination, which is the dominant pneumococcal evolutionary mechanism, is mainly associated with pneumococcal carriage (91). Serotypes common in invasive disease are rarely carried, and are thus less exposed to recombinational events than serotypes common in carriage, which have more opportunities for the horizontal transfer of DNA to coinfecting lineages of other serotypes (Figure 13) (92-93).

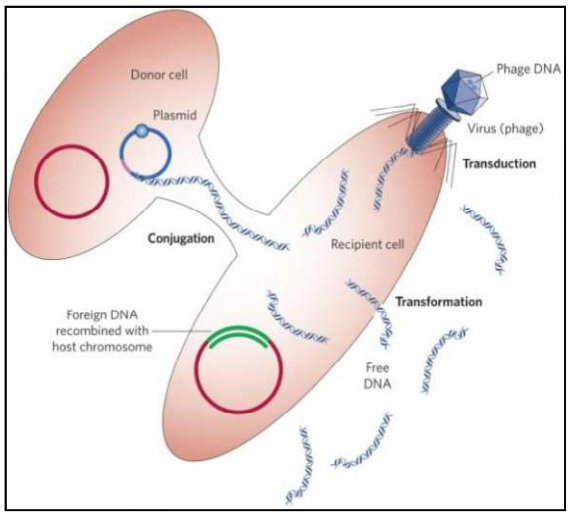

Figure 13.*Routes of horizontal gene transfer in bacteria (Data from: Stewart FJ et al. 2013).*

There are several molecular methods that are being used for the molecular typing of the different pneumococcal strains (87, 94):

1. **Multilocus Sequence Typing (MLST):** This method is based on sequencing and comparing seven housekeeping genes. Afterwards, the strains are assigned a sequence type (ST) depending on the allelic variations. Although this assay has the same resolving power than other DNA-based techniques and it is more expensive than some of them, it allows making national and international comparisons of strains through a general database which offers the possibility of global surveillance studies (http://pubmlst.org/spneumoniae/). The main

limitation is that phylogenetic relationships and resolution of clones can be masked by the use of slowly evolving housekeeping genes.

2. **Multilocus Variable Number of Tandem Repeats Analysis (MLVA):** This method utilizes the naturally occurring variation in the number of tandem repeated DNA sequences found in many different loci in the genome of the pneumococcus (http://www.mlva.net). Although it has been reported to be more discriminatory, inexpensive and faster than MLST, it is more suitable for short-term epidemiology changes and localized outbreaks.
3. **Pulsed-Field Gel Electrophoresis (PFGE)**: In this method the pneumococcal chromosomal DNA is cleaved by a restriction enzyme and then run on a pulsed-field gel that separates bands by charge and size. It has the advantage of being an inexpensive and highly reproducible technique. Also, this assay has good discriminatory power which is especially useful to investigate outbreaks. However, comparing results obtained by different laboratories is difficult and obtaining specific information about genes is not possible so is less suitable for long-term studies.
4. **Microarray**: In this method genes from sequenced pneumococcal genomes are spotted on an array, and then the presence or absence of these genes is studied by comparing with the DNA from pneumococcal isolates of interest. The discriminatory power of this assay depends on the number of genes that are studied.
5. **Whole genome sequencing:** This is the method with more discriminatory power because it allows the identification of all genomic differences between isolates.

To identify and compare the genetic diversity within species Hunter and Gaston (1988) proposed the use of Simpson's index of diversity (Simpson, 1949). This index indicates the probability of two strains sampled randomly from a population belonging to two different types. In microbial ecology and in descriptive epidemiology, this index is used to determine the diversity of microorganisms in defined environments or to identify the reproductive success of disease causing organisms (for example the spread of particular strains between hosts). The calculation of the Simpson's index of diversity is based on the frequency with which organisms of a particular type occur in a population or can be discriminated by a given typing tool (95).