New molecular tools for serotyping for *S. pneumoniae* invasive strains surveillance in the province of Quebec.

Principal Investigator
Brigitte Lefebvre, Ph. D.
Laboratoire de santé publique du Québec

Co-PI
Cécile Tremblay, MD, Pfizer/University of Montreal Chair on HIV Translational Research, University of Montreal.
Director, Laboratoire de santé publique du Québec

Co-Investigator
Simon Lévesque, Ph. D.
Laboratoire de santé publique du Québec

Co-Investigator
Sadjia Bekal, Ph. D.
Laboratoire de santé publique du Québec

Co-Investigator
Marc-Christian Domingo, Ph. D.
Laboratoire de santé publique du Québec

Background

Streptococcus pneumoniae is responsible for various infections such as pneumonia, otitis, sinusitis, peritonitis, endocarditis and meningitis[(1)]. The incidence of invasive *S. pneumoniae* is often used as an indicator of the burden of pneumococcal disease. Virulence and invasiveness varies among serotypes. In pneumococcus, several virulence factors are known; among these, the *cps* locus encoded capsule is a crucial one, as the prime target for vaccine development. Although several vaccines (PCV-7, PCV-10, PCV-13 and PCV-23) with different coverage have been developed against *S. pneumoniae*, invasive pneumococcal disease remains a public health concern since a vaccine replacement phenomenon is observed[(2)].

Since 1990, *S. pneumoniae* serotype is determined using the Quellung's technique in most laboratories[(3)]. This standard method uses antisera to reveal the swelling of the capsule through an antibody-antigen reaction[(1,4)]. This technique is laborious, expensive and requires technical expertise. Although this technique is recognized as the reference method, it can lead to erroneous results because it is subjective. Indeed, serotyping results are obtained through microscope observation of capsular swelling which in some cases is difficult to observe. As more than 90 serotypes of *S. pneumoniae* have been described to date[(1)], a serotyping algorithm must be applied using different antisera which makes the task tedious and time consuming.

Rapid molecular techniques are now being evaluated to perform serotyping. A review of several published methods to determine the serotype of *S. pneumoniae* is presented in Table 1[(4-19)]. Among the six methods presented, two cannot be used as part of a surveillance program. The whole genome sequencing (WGS) is a method which generates data that are not all relevant in the current context of monitoring[(17-19)]. However, WGS may help to understand the mechanism of replacement and adaptation through possible recombination in the *S. pneumoniae* strains in response to vaccination [(20)]. Although promising, EIMS[11,12] method, is not available to the Quebec market. Microarrays[13-14] technology allows rapid genes resistance and virulence identification. However, microarrays equipment is not more available at LSPQ. We retained the other three methods based on the following criteria: cost analysis, technology availability at the LSPQ and timely delivery of results. In the case of sequetyping[10], unlike multiplex PCR (which remains the most cost effective[5-9]), the method does not require adaptation to local epidemiology of circulating serotypes. For all molecular methods described, the literature reports that a certain percentage of serotype strains cannot be determined. In which case an alternative path must then be considered like Quellung's serotyping method.

We propose to evaluate various molecular techniques for rapid serotyping of *S. pneumoniae* strains as compared with Quellung gold standard, including all invasive strains isolated from children and adults in the province of Quebec.

Methods: Molecular method comparison (using LSPQ collection of invasive *S. pneumoniae* strains) with gold standard method (Quellung) and WGS to study the impact of vaccine on serotype replacement.

1- Monitoring tools :

- Multiplex PCR
- Sequetyping

Phase 1 : For the development, 20 selected strains will be used to fine tune and develop the methods.

Phase 2 : For the proof-concept, an additional 100 strains will be analyzed using the two molecular methods. The third method, WGS, will be performed on 10 strains. The strain collection will be representative of various circulating serotypes, including serotypes (19A, 7F, 3, 22F, 9N, 15A, 6C) and all serotypes included in currently used vaccines. Molecular methods will be compared to the Quellung gold standard method. After the proof-concept period, the most efficient method will be retained and used for surveillance programme. The choice will be based on cost effectiveness, efficiency, cost of reactive, cost of technical time, accuracy and professional expertise.

2- Molecular basis of vaccine replacement by WGS :

WGS will be performed on 10 selected strains to study pneumococcus post-vaccine changes through two approaches:

- Pre- and post-vaccine follow-up for serotyping evolution.
- Identification of putative vaccine target.

Time-line (See Annex 1)

Steps	Lenght
Development: Strains' selection and development of 3 molecular methods for serotyping	1 year
Proof of concept: -Molecular methods will be compared with Quellung gold standard method and WGS -Evaluation on our surveillance programme in the design of vaccines using the new validated method.	1 year
Publication and conference organization	At the end of study

Project Benefits

1- Implementation of an active monitoring tool of invasive *S. pneumoniae* serotypes.
2- Reduction of delays and costs associated with the provincial monitoring program of invasive strains of *S. pneumoniae* using optimized serotyping methods.
3- Potential increase of provincial surveillance program capacity building due to cost effectiveness.
4- Identification of putative vaccine target.
5- Better understanding of vaccine replacement mechanism.

Deliverables

1- Set up of a new molecular serotyping method.
2- Data from the study will be presented at a scientific meeting and published in a peer reviewed journal.

References

1- Spellerberg, B. and Brandt C. *Streptococcus*. 2011. Manual of clinical microbiology. 10th edition. American Society for microbiology, Washington, D.C.

2- Lefebvre B., C. L. Tremblay. 2012. Programme de surveillance du pneumocoque. Rapport 2011. INSPQ. ISBN : 978-2-550-66364-5.

3- Sorensen, U. B. 1993. Typing of pneumococci by using 12 pooled antisera. J. Clin. Microbiol. 31:2097-2100.

4- Austrian, R. 1976. The quellung reaction, a neglected microbiologic technique. Mt. Sinai J. Med 43:699-709.

5- http://www.cdc.gov/ncidod/biotech/strep/strepindex.htm

6- Pai, R., R. E. Gertz, and B. Beall. 2006. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. J. Clin. Microbiol. 44:124-131.

7- Iraurgui, P., M. J. Torres, A. Gandia, I. Vazquez, E. G. Cabrera, I. Obando, J. Garnacho, and J. Aznar. 2010. Modified sequential multiplex PCR for determining capsular serotypes of invasive pneumococci recovered from Seville. Clin. Microbiol. Infect. 16:1504-1507.

8- Yun, K. W., E. Y. Cho, K. B. Hong, E. H. Choi, and H. J. Lee. 2011. *Streptococcus pneumoniae* type determination by multiplex polymerase chain reaction. J. Korean Med. Sci. 26:971-978.

9- Siira, L., T. Kaijalainen, L. Lambertsen, M. H. Nahm, M. Toropainen, and A. Virolainen. 2012. From Quellung to multiplex PCR, and back when needed, in pneumococcal serotyping. J. Clin. Microbiol. 50:2727-2731.

10- Leung, M. H., K. Bryson, K. Freystatter, B. Pichon, G. Edwards, B. M. Charalambous, and S. H. Gillespie. 2012. Sequetyping: serotyping *Streptococcus pneumoniae* by a single PCR sequencing strategy. J.Clin. Microbiol. 50:2419-2427.

11- Massire, C., R. E. Gertz, Jr., P. Svoboda, K. Levert, M. S. Reed, J. Pohl, R. Kreft, F. Li, N. White, R. Ranken, L. B. Blyn, D. J. Ecker, R. Sampath, and B. Beall. 2012. Concurrent serotyping and genotyping of pneumococci by use of PCR and electrospray ionization mass spectrometry. J. Clin. Microbiol. 50:2018-2025.

12- Wolk, D. M., E. J. Kaleta, and V. H. Wysocki. 2012. PCR-electrospray ionization mass spectrometry: the potential to change infectious disease diagnostics in clinical and public health laboratories. J. Mol. Diagn. 14:295-304.

13- Wang, Q., M. Wang, F. Kong, G. L. Gilbert, B. Cao, L. Wang, and L. Feng. 2007. Development of a DNA microarray to identify the *Streptococcus pneumoniae* serotypes contained in the 23-valent pneumococcal polysaccharide vaccine and closely related serotypes. J. Microbiol. Methods 68:128-136.

14- Tomita, Y., A. Okamoto, K. Yamada, T. Yagi, Y. Hasegawa, and M. Ohta. 2011. A new microarray system to detect *Streptococcus pneumoniae* serotypes. J. Biomed. Biotechnol. 2011:352736.

15- Gervaix, A., J. Taguebue, B. N. Bescher, J. Corbeil, F. Raymond, G. Alcoba, M. Kobela, and E. Tetanye. 2012. Bacterial meningitis and pneumococcal serotype distribution in children in cameroon. Pediatr. Infect. Dis. J. 31:1084-1087.

16- Raymond F; Boucher N; Allary R; Robitaille L; Lefebvre B; Tremblay C; Corbeil J; Gervaix A. Serotyping of Streptococus pneumoniae based on capsular genes polymorphisms. PLoS One; Sept. 24, 2013, http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0076197.

17- Fani, F., P. Leprohon, D. Legare, and M. Ouellette. 2011. Whole genome sequencing of penicillin-resistant *Streptococcus pneumoniae* reveals mutations in penicillin-binding proteins and in a putative iron permease. Genome Biol. 12:R115.

18- Billal, D. S., J. Feng, P. Leprohon, D. Legare, and M. Ouellette. 2011. Whole genome analysis of linezolid resistance in *Streptococcus pneumoniae* reveals resistance and compensatory mutations. BMC Genomics 12:512.

19- Hu, F. Z., R. Eutsey, A. Ahmed, N. Frazao, E. Powell, N. L. Hiller, T. Hillman, F. J. Buchinsky, R. Boissy, B. Janto, J. Kress-Bennett, M. Longwell, S. Ezzo, J. C. Post, M. Nesin, A. Tomasz, and G. D. Ehrlich. 2012. In Vivo Capsular Switch in *Streptococcus pneumoniae* - Analysis by Whole Genome Sequencing. PLoS One. 7:e47983.

20- Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M. 2013 Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. Jun;45(6):656-63.

Table 1. Comparison of different methods for *S. pneumoniae* serotyping.

Method	Quellung	Multiplex PCR	Sequetyping	Electrospray ionization mass spectrometry (EIMS)	Microarray	Whole genome sequencing
Brief description	Determination of serotype with antisera (swelling of the capsule)	Multiplex PCR cascading up to 8 different PCR reactions (method adapted from CDC protocol)	PCR cpsB and amplicon sequencing (~ 1000 pb)	Five multiplex PCR microplates, followed by analysis of mass spectrometry coupled to electrospray ionization	Hybridization of labeled DNA on a solid support (chip) where the interesting genes, including those from serotyping are printed	Full genome sequencing (2nd generation sequencing)
Avantages	- Gold standard - Covers all serotypes - Validated and available method at LSPQ	- Fast - Inexpensive - Easily achievable - Equipment available at the LSPQ - Used in several laboratories across the world (USA, Spain, Finland, Brazil, Korea)	- Fast - Inexpensive - Easily achievable - No need to adapt the local epidemiology - Detection of new serotypes - Equipment available at the LSPQ	- Partially automated - Determines useful ST's (sequence type) for epidemiological studies - Methodology usable for other applications	- No need to adapt the protocol to local epidemiology - Reader microarray available at LSPQ	- Large amount of data generated - Identification of genes resistance and virulence - Identification of therapeutic targets - Equipment available at the LSPQ
Disadvantages	- Tedious - Laborious - Subjective - Expensive - Possibility of cross-reactions	- To be customized according to local epidemiology - Detection of known serotypes - Possibility of false + - Some serotypes are difficult to identify (eg, 6A, 6B, 6C, 6D)	- Method based on public databases (eg NCBI) that are not always accurate - Necessity of a cpsB controlled bank	- Unavailable device at the LSPQ - Plex-ID is not available on the market: adjustments are in process at Abbott	- Detection of known serotypes only	- Method not suitable for serotyping in a monitoring program setting
Efficiency	Serotype : 100 %	Sérotype : 93 - 99 %	Serotype : 66 % Serogroup : 20 % more Ambiguous results : 14 %	Not available	Serotype : 75 % Serogroup : 9 % more No result : 11 % Error : 5 %	Not available
Time required	96 hours	~72 hours	~72 hours	Not available	~72 hours	1 week
Costs / strain	Between 60 $ et 100 $ according to serotype	Between 30 $ et 80 $ according to multiplex design	50 $	Not available	160 $	120 $
References *	Austrian, 1976 [4]	- Site web du CDC [5] - Pai et al., 2006[6] - Iraurgui et al., 2010 [7] - Yun et al., 2011[8] - Siira et al., 2012[9]	- Leung et al., 2012[10]	- Massire et al., 2012[11] - Wolk et al., 2012[12]	- Wang et al., 2007 [13] - Tomita et al., 2011[14] - Gervaix et al., 2012[15] - Raymond et al, 2013 [16]	- Fani et al., 2011[17] - Billal et al., 2011[18] - Hu et al., 2012[19] - Croucher et al., 2013[20]

Annexe 1. Time Frame/Project Goals (arrows), milestones (red), task (blue bars) and timelines.

TASK (Location)	YEAR 1	YEAR 2
Development of comprehensive tools for rapid detection and efficient monitoring of *S. pneumoniae*.	↔	
• PCR multiplex	▬●	
• Sequetyping	▬●	
• WGS (10 strains)	▬●	
Proof of concept (comparison with gold standard method; Quellung) using 100 representative strains.		↔
• PCR multiplex		▬●
• Sequetyping		▬●
• WGS (10 strains)		▬●
Publication/conference		▬●

	STUDY DOCUMENT REQUIREMENTS FORM	**01-AUG-2014**

STUDY INFORMATION

PRINCIPAL INVESTIGATOR	**Dr. Brigitte Lefebvre**		
PFIZER INSPIIRE NO.	WI203144	INSTITUTIONAL REFERENCE NUMBER	
PROTOCOL TITLE	**Serotype monitoring of S. pneumoniae invasive strains in adult population in the province of Quebec_ a 3 years study evaluation.**		

DOCUMENTATION REQUIREMENTS

MATERIALS ENCLOSED WITH THIS PACKET: (DELETE ANY ITEMS BELOW THAT DO NOT APPLY)

- ☒ Site Information Sheet **(agreement information form)**
- ☐ Drug Supply Request Form
- ☐ Reportable Event Fax Cover Sheet
- ☐ Pfizer Safety Reporting Reference Manual for IIR studies
- ☐ Pfizer IIR Adverse Event Report Form and IIR Adverse Event Report Form Completion Instructions
- ☐ Exposure During Pregnancy (EDP) Supplemental Form
- ☐ Product information (document or reference)
- ☐ IRS Web site address to download Form W-9 (US/Puerto Rico only)

PRINCIPAL INVESTIGATOR MUST PROVIDE TO PFIZER: (ONLY BOXES CHECKED BELOW)

Documents required to generate an IIR Agreement

- ☒ Completed **(agreement information form)**
- ☐ Completed IRS *Form W-9* (US/Puerto Rico only for payee entity)

Documents required to be submitted prior to receiving monetary support and/or drug supplies

- ☐ Completed Site Information Sheet (Drug Supply Information and/or Financial Information Tab[s])
- ☒ Executed IIR agreement
- ☒ Final study protocol (for a study with sites in the EU, the principal investigator must sign the final study protocol as required for qualified person [QP] release of drug supplies)
- ☒ IRB/IEC approval letters (initial approval and annual renewals, as applicable)
- ☐ Regulatory response

  For US studies:

  - ☐ FDA IND response (IND number or exemption – *may not apply to all consumer products*)
  - ☐ DEA number for controlled substances

  For EU studies:

  - ☐ Approved clinical trial application (CTA) in English (as required for QP release)
  - ☐ Submission letter for the CTA

  For non-US, non-EU studies:

  - ☐ Appropriate Regulatory review/approval based upon local country requirements

Site Information Sheet / agreement information form)

The information requested on the *Site Information Sheet /A**greement information form*** is critical to Pfizer in order to develop an agreement, to reduce the agreement's review time, and to ensure that monetary support is sent to the appropriate payee or drug supply is sent to the appropriate address. Withholding or delaying Pfizer's receipt of this form will significantly delay the contracting process for the approved research.

Final Protocol and Amendments

Pfizer will not provide support to an IIR study until after receipt of the final study protocol. If the research described in the final protocol is materially different from that in the approved proposal, then Pfizer may choose to modify or withhold its support.

As indicated in the agreement, the principal investigator must also promptly provide Pfizer with any amendments to the Pfizer-approved final study protocol. Continuation of support by Pfizer for an IIR study will be contingent on Pfizer's review and acceptance of these changes.

For studies with sites in the EU where drug support is being requested, the final study protocol must be signed by the principal investigator and is required for QP release of drug supplies.

Institutional Review Board (IRB)/Independent Ethics Committee (IEC) Documents

For studies that require IRB/IEC approval, Pfizer will only provide support for an IIR study after receipt of a copy of the IRB/IEC approval letter.

Continuation of support by Pfizer requires timely submission of a copy of IRB/IEC renewal documentation subsequent to the original IRB/IEC approval (as required per local regulations).

Regulatory Response

US Clinical Studies: FDA IND Response or IND Exemption Documentation. For an interventional clinical study involving a Pfizer drug, an investigational new drug (IND) application may need to be filed with the U.S. Food and Drug Administration (FDA). Please review IND requirements under 21 CFR 312 (available at http://www.fda.gov) to determine whether an IND is required.

For this type of study, Pfizer will not provide any IIR support until after receipt of documentation that an IND has been filed or that the study is exempt from an IND filing under 21 CFR 312.2(b)(1).

European Union Clinical Studies. For studies for which conduct under a clinical trial application (CTA) is required, Pfizer will not provide any IIR support until after receipt of a copy of the submission letter to the CTA, in English.

If Pfizer will provide packaged and labeled Pfizer product, then Pfizer must receive a copy of the approved CTA, with Section 4.2 (IMPD or Letter of Access from Pfizer) and Section D (in its entirety) must be translated in English, before Pfizer can provide QP release of product. For more information regarding CTAs, please consult http://eudract.emea.europa.eu/document.html.

Should your local regulatory authority require documentation from Pfizer, please contact your IIR manager for assistance.

Non-US/Non-EU Studies. Should your local regulatory authority require documentation from Pfizer, please contact your IIR manager for assistance.

Investigator-Initiated Research Agreement

Pfizer will provide the principal investigator or the contracting office with an IIR agreement that documents the terms under which Pfizer will provide the research grant. Development of the agreement is based upon information you have supplied on the enclosed forms.

Drug Supply Request Form

If Pfizer has agreed to supply drug, then the *Drug Supply Request Form* can be used to communicate your clinical supply needs throughout the course of the IIR study. Pfizer will not ship any clinical supplies until all required documents have been received and an IIR agreement has been executed.

NOTE: Availability of drug may take between eight weeks and twelve months, depending upon the product and its packaging and labeling requirements. Contact the appropriate IIR manager to determine available quantities of drug and timelines for shipment.

For Oncology Studies Conducted in the United States. If Pfizer is not providing clinical supplies for this study, then Pfizer cannot be held responsible for drug cost reimbursement. For assistance with third-party reimbursement procedures and indigent patients, contact FirstRESOURCE, Pfizer Oncology's Reimbursement and Patient Assistance Program, at 877-744-5675 prior to initiating therapy.

IRS Form W-9

Pfizer requires that all grant recipients based in the U.S. or Puerto Rico who receive monetary support complete and submit IRS *Form W-9*. This form shall be completed for the entity which will be receiving the grant payment(s). Please verify with your grants office that the name of the payee is correct and that it is the legal entity name related to the tax identification number. The latest version of *Form W-9* may be downloaded from the IRS Web site from: http://www.irs.gov/pub/irs-pdf/fw9.pdf.

Product Information

Pfizer is required to provide relevant and current scientific information about the investigational product to the investigator. This may be accomplished by supplying one of the following Pfizer-approved documents to the investigator: Investigator Brochure (IB), package insert (PI), or local product document (LPD).

Safety Reporting

Safety Reporting Reference Manual for IIR Studies with Pfizer Products. Detailed information regarding a principal investigator's (or investigators') adverse event reporting responsibilities for a Pfizer-supported IIR study can be found in the accompanying training manual. **Please read through this document carefully. Principal investigators must understand and fully comply with the adverse event reporting requirements of their studies.**

NOTE: Reporting an adverse event to Pfizer does not relieve the institution of its responsibility to report the event to the FDA or to the local regulatory authorities that govern that institution.

IIR SAE Form and IIR SAE Report Form Completion Instructions. For those studies where the principal investigator is required to submit reportable events (AEs and SAEs) to Pfizer, the investigator may use the *Pfizer IIR SAE/Adverse Event Report Form* to submit the event. Instructions for completion will also be provided.

Reportable Event Fax Cover Sheet. For those studies where the principal investigator is required to report adverse events and other reportable events to Pfizer, the investigator must use the attached *Reportable Event Fax Cover Sheet* along with the Pfizer-approved *Adverse Event Report Form*.

Pfizer

August 19, 2015

Dr. Brigitte Lefebvre
Laboratoire de Sante Publique du Québec
20045 chemin Ste-Marie
Sainte-Anne-de-Bellevue, Québec
H9X 3R5

Email: brigitte.lefebvre@inspq.qc.ca
Re: Pfizer reference # **WI203144**

Dear Dr. Lefebvre,

The Vaccines Team IIR Grant Review Committee has reviewed your proposal titled ***"Serotype monitoring of S. pneumoniae invasive strains in adult population in the province of Quebec_ a 3 years study evaluation."*** and is pleased to inform you that Pfizer is interested in supporting your research with funding.

The total amount of funding you requested is **$707,080.00.**The actual amount of funding will be agreed upon and reflected in the Investigator-Initiated Research Agreement.

Please complete and return the accompanying **Agreement Information Form** to begin the contracting process. For those studies being conducted in the United States and Puerto Rico where funding is provided, a completed IRS Form W-9 is required.

Pfizer support is contingent upon the receipt of:

- Final research protocol*
- IIR Agreement executed between Pfizer and your institution
- IRB/IEC approval (as appropriate)
- Regulatory response (see enclosed guidelines)

*Please be aware that if the research described in your final protocol is materially different from that presented in your original proposal, then Pfizer reserves the right to reconsider its support.

If you have not obtained IRB approval and/or executed the IIR Agreement with Pfizer within six (6) months from the date of this letter, then funding for your grant cannot be guaranteed. Although this letter signifies Pfizer's intention to support your proposal, Pfizer is not committed until an agreement has been fully executed.

Pfizer recognizes that carefully conducted clinical trials are the fastest and safest way to find treatments to improve health. As such, Pfizer encourages you and your institution to add this study to the FDA's www.clinicaltrials.gov database. Pfizer recognizes that the availability of clinical trial listings and results are critical to the communication of important new information for the medical profession, patients, and the public.

If you have questions, please contact me [redacted] **or** the **Regional Medical and Research Specialist (RMRS)** [redacted]

We look forward to working with you.

Yours sincerely,

[redacted]

IIR Grant Specialist
Medical Quality & Effectiveness

Serotype monitoring of *S. pneumoniae* invasive strains in adult population in the province of Quebec: a 3 years study evaluation

Principal Investigator
Brigitte Lefebvre, Ph. D.
Microbiologist, Laboratoire de santé publique du Québec

Co-PI
Cécile Tremblay, MD, Pfizer/University of Montreal Chair on HIV Translational Research, University of Montreal.
Director, Laboratoire de santé publique du Québec

Background

Streptococcus pneumoniae is responsible for various infections such as pneumonia, otitis, sinusitis, peritonitis, endocarditis and meningitis[1]. The incidence of invasive *S. pneumoniae* is often used as an indicator of the burden of pneumococcal disease. Virulence and invasiveness varies among serotypes. In *S. pneumoniae*, several virulence factors are known; among these, the *cps* locus encoded capsule is a crucial one, as the prime target for vaccine development. Although several vaccines (PCV-7, PCV-10, PCV-13 and PCV-23) with different coverage have been developed against *S. pneumoniae*, invasive pneumococcal disease remains a public health concern as vaccine replacement phenomenon has been observed[2].

In December 2004, PCV-7 vaccination was implemented free to all newborns in Quebec, using a 3-dose schedule (2, 4 and 12 months). Simultaneously, the vaccine could be offered free of charge to all children under the age of 5, during routine visits. In 2008, a new PCV-10 containing 3 serotypes not included in PCV-7 vaccine was licensed in Canada. It was introduced in Quebec in children in the summer of 2009. In 2009, PCV-13 vaccine was approved in Canada. It was introduced in the Quebec immunization program in January 2011 and replaced PCV-10.

The introduction of PCV-7 had not only an important impact on the number and the diversity of strains isolated from children under 5 years of age, but the impact was also observed in individuals ≥ 5 year old. Thus, the proportion of serotypes included in PCV-7 has dramatically declined since 2005. However, there was an increase in the proportion of serotypes 7F and 19A which are not included in PCV-7 and an increase of non-vaccine serotypes was observed. In 2013, a decrease in the frequency of 7F and 19A serotypes in individuals ≥ 5 year old was observed. However, the number of circulating serotypes not included in the PCV-7, PCV-10 and PCV-13 is increasing.

Thus, sustained laboratory monitoring is essential because it allows the study of evolution of circulating serotypes as well as antibiotic resistance patterns, two crucial parameters for planning immunization programs, the choice of vaccines and the development of treatment guidelines. Analysis of invasive strains allows for the study of serotypes distribution and antibiotic susceptibility patterns of strains responsible for the most severe forms of pneumococcal disease. Monitoring of circulating serotypes is essential to assess the impact of vaccination programs of the province of Quebec.

In 1996, the Public Health Laboratory of Quebec (LSPQ) in collaboration with hospital laboratories established a laboratory surveillance program of *S. pneumoniae* invasive strains. The program's objectives were to study the serotype distribution circulating in Quebec and establish their antibiotic susceptibility profiles. This program was based on the collection of strains from sentinel laboratories. In 2005, in order to assess the impact of the universal immunization program against *S. pneumoniae* in children, the program was expanded to all invasive strains of *S. pneumoniae* isolated from children under 5 years of age.

This monitoring program has kept track of the evolution, in Quebec children, of various serotypes and resistance in connection with the introduction of the PCV-13 vaccine in 2011 and more specifically allows for the measure of its impact on the prevalence of serotypes 7F and 19A, two serotypes highly prevalent in Quebec. Currently, the provincial surveillance program is limited to strains collected in children less than 5 years of age and to adult strains from sentinel laboratories which represent less than 25% of the total invasive strains in the adult population. Therefore, we may be underestimating the diversity of circulating strains especially in areas not represented in the sentinel program and may not capture adequately seasonal variation. Two years ago, we proposed, a study evaluating the benefits of acquiring data on all invasive strains isolated in patients (≥ 5 years old) of the province of Quebec compared to sentinel sites. This study was launched in August 2013, with the financial support of Pfizer. Preliminary data from the first 18 months of extended surveillance indicate that some emerging serotypes may not be fully captured by the sentinel sites, although these observations need to be evaluated by longer follow-up.

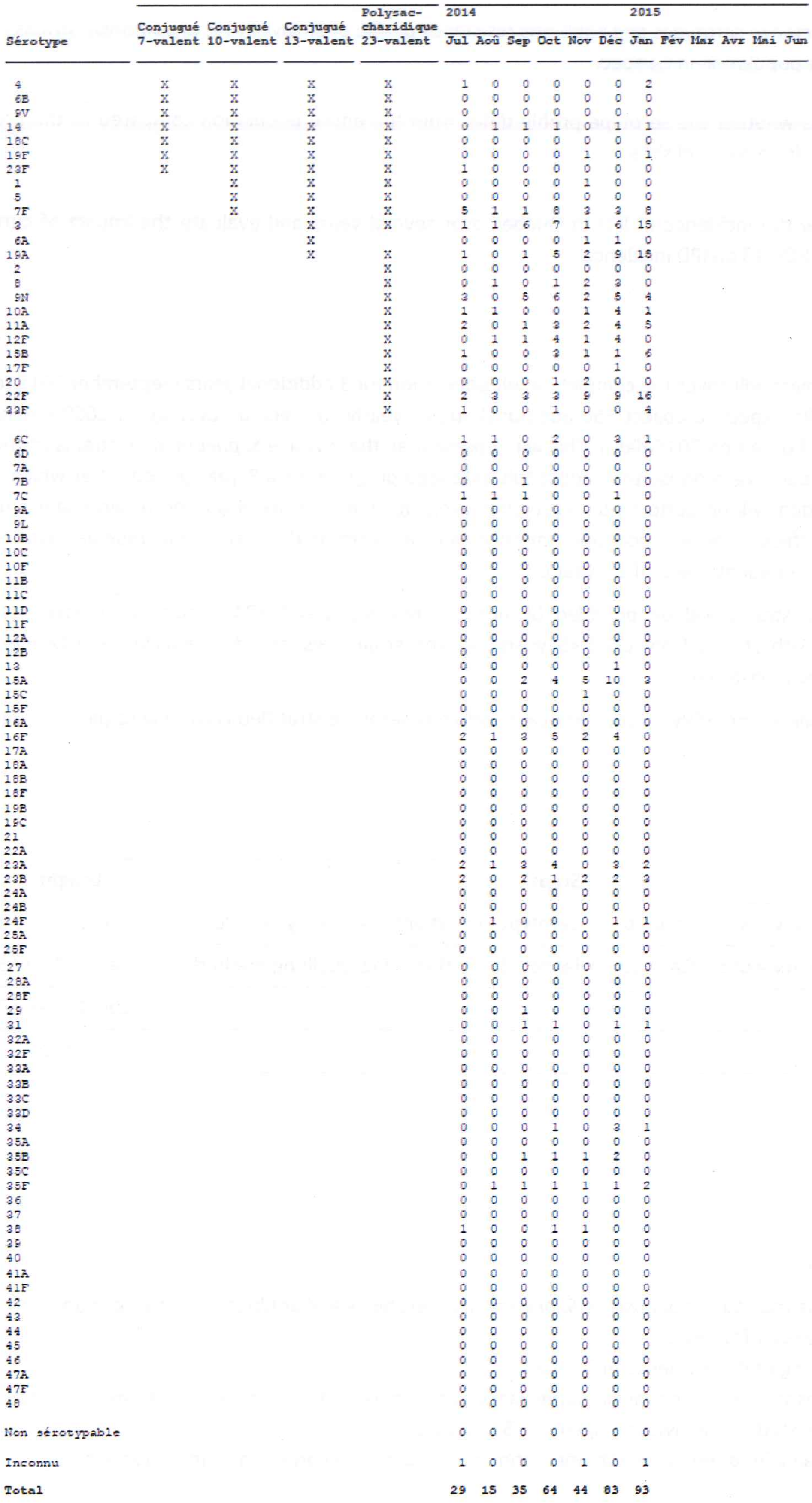
Preliminary data from surveillance of invasive *S. pneumoniae* in individuals ≥ 5 years old

After 18 months of extended surveillance, we have identified a higher proportion of two serotypes, the 6A and 15A, which had not previously been identified with the sentinel sites surveillance program. Serotype 6A is included in the currently used PCV-13 vaccine and serotype 15A is not included in this vaccine and exhibits multi-resistance. A recent paper from Israels howed a similar increase of 15A serotype among adult invasive pneumococcal disease(2). Emergence of serogroup 15 was also described by Liyanapathirana *et al.*(3) in nasopharyngeal carriage of hospitalized children. Furthermore, our data analysis revealed an overrepresentation of some serotypes when only sentinel data are analyzed. The clinical significance of these serotypes is not yet defined. However, this supports the necessity to expand our broadened monitoring over a longer period of time to evaluate the establishment of these serotypes into Quebec's ecology and their relevance for vaccine development.

Before the beginning of our study in 2013, reporting of data was available in 3 formats: i) The annual provincial aggregated data generally available one year after data collection(4); ii) The monthly LSPQ StatLabo report providing aggregated data with a 2 months delay(5) iii) Individual reports for each strain sent to participating laboratories as well as public health stakeholders, up to 4 months after strain reception. As part of the current study, we were able to make available in real time information on circulating serotypes by publishing a monthly report including all serotypes identified, classified by age groups in the bulletin StatLabo (Fig. 1.).

We propose to continue our study for another three years to allow for a full characterization of circulating serotypes including clustering in certain geographical areas or seasonal variation, to establish incidence of invasive pneumococcal disease in the Quebec population, and to define if this surveillance program provides added value to a sentinel site based approach. Results of this research project could help guide public health authorities in immunization strategies and will also provide useful information for vaccine design.

Figure 1. Données mensuelles des souches invasives de *S. pneumoniae* chez les patients de 5 ans et plus [6].



Sérotype	Conjugué 7-valent	Conjugué 10-valent	Conjugué 13-valent	Polysac-charidique 23-valent	2014						2015					
					Jul	Aoû	Sep	Oct	Nov	Déc	Jan	Fév	Mar	Avr	Mai	Jun
4	X	X	X	X	1	0	0	0	0	0	2					
6B	X	X	X	X	0	0	0	0	0	0	0					
9V	X	X	X	X	0	0	0	0	0	0	0					
14	X	X	X	X	0	1	1	1	0	1	1					
18C	X	X	X	X	0	0	0	0	0	0	0					
19F	X	X	X	X	0	0	0	0	1	0	1					
23F	X	X	X	X	1	0	0	0	0	0	0					
1		X	X	X	0	0	0	0	1	0	0					
5		X	X	X	0	0	0	0	0	0	0					
7F		X	X	X	5	1	1	8	3	3	8					
3			X	X	1	1	6	7	1	6	15					
6A			X		0	0	0	0	1	1	0					
19A			X	X	1	0	1	5	2	9	15					
2				X	0	0	0	0	0	0	0					
8				X	0	1	0	1	2	3	0					
9N				X	3	0	5	6	2	5	4					
10A				X	1	1	0	0	1	4	1					
11A				X	2	0	1	3	2	4	5					
12F				X	0	1	1	4	1	4	0					
15B				X	1	0	0	3	1	1	6					
17F				X	0	0	0	0	0	1	0					
20				X	0	0	0	1	2	0	0					
22F				X	2	3	3	3	9	9	16					
33F				X	1	0	0	1	0	0	4					
6C					1	1	0	2	0	3	1					
6D					0	0	0	0	0	0	0					
7A					0	0	0	0	0	0	0					
7B					0	0	1	0	0	0	0					
7C					1	1	1	0	0	1	0					
9A					0	0	0	0	0	0	0					
9L					0	0	0	0	0	0	0					
10B					0	0	0	0	0	0	0					
10C					0	0	0	0	0	0	0					
10F					0	0	0	0	0	0	0					
11B					0	0	0	0	0	0	0					
11C					0	0	0	0	0	0	0					
11D					0	0	0	0	0	0	0					
11F					0	0	0	0	0	0	0					
12A					0	0	0	0	0	0	0					
12B					0	0	0	0	0	0	0					
13					0	0	0	0	0	1	0					
15A					0	0	2	4	5	10	3					
15C					0	0	0	0	1	0	0					
15F					0	0	0	0	0	0	0					
16A					0	0	0	0	0	0	0					
16F					2	1	3	5	2	4	0					
17A					0	0	0	0	0	0	0					
18A					0	0	0	0	0	0	0					
18B					0	0	0	0	0	0	0					
18F					0	0	0	0	0	0	0					
19B					0	0	0	0	0	0	0					
19C					0	0	0	0	0	0	0					
21					0	0	0	0	0	0	0					
22A					0	0	0	0	0	0	0					
23A					2	1	3	4	0	3	2					
23B					2	0	2	1	2	2	3					
24A					0	0	0	0	0	0	0					
24B					0	0	0	0	0	0	0					
24F					0	1	0	0	0	1	1					
25A					0	0	0	0	0	0	0					
25F					0	0	0	0	0	0	0					
27					0	0	0	0	0	0	0					
28A					0	0	0	0	0	0	0					
28F					0	0	0	0	0	0	0					
29					0	0	1	0	0	0	0					
31					0	0	1	1	0	1	1					
32A					0	0	0	0	0	0	0					
32F					0	0	0	0	0	0	0					
33A					0	0	0	0	0	0	0					
33B					0	0	0	0	0	0	0					
33C					0	0	0	0	0	0	0					
33D					0	0	0	0	0	0	0					
34					0	0	0	1	0	3	1					
35A					0	0	0	0	0	0	0					
35B					0	0	1	1	1	2	0					
35C					0	0	0	0	0	0	0					
35F					0	1	1	1	1	1	2					
36					0	0	0	0	0	0	0					
37					0	0	0	0	0	0	0					
38					1	0	0	1	1	0	0					
39					0	0	0	0	0	0	0					
40					0	0	0	0	0	0	0					
41A					0	0	0	0	0	0	0					
41F					0	0	0	0	0	0	0					
42					0	0	0	0	0	0	0					
43					0	0	0	0	0	0	0					
44					0	0	0	0	0	0	0					
45					0	0	0	0	0	0	0					
46					0	0	0	0	0	0	0					
47A					0	0	0	0	0	0	0					
47F					0	0	0	0	0	0	0					
48					0	0	0	0	0	0	0					
Non sérotypable					0	0	0	0	0	0	0					
Inconnu					1	0	0	0	1	0	1					
Total					**29**	**15**	**35**	**64**	**44**	**83**	**93**					

Project objectives

1- To characterize serotypes and antibiotic resistance profile of all invasive *S. pneumoniae* strains from the adult population in Quebec.

2- To assess whether the serotype profile differ from the entire population compared to the profile obtained from sentinel sites.

3- To follow the incidence of IPD in Quebec over several years and evaluate the impact of current vaccine, PCV-13 on IPD incidence.

Methodology

The research project will cover the complete adult population for 3 additional years (September 2015 to August 2018). We expect to collect 550 additional strains yearly to reach an average of 1000 strains yearly (estimated based on 2014 data). This will represent all the invasive *S. pneumoniae* strains of the province of Quebec. We propose to conduct this extended program for a 3-year period, after which a program evaluation will be performed. Serotyping using Quellung methodology and determination of susceptibility profiles using microdilutions method will be performed on all *S. pneumoniae* invasive strains collected in patients aged of ≥ 5 year old.

Those additional strains will be provided by non-sentinel hospitals (n=74) which, until now, only provided LSPQ with strains from child <5 years old and strains resistant to penicillin (≥ 0.12 mg/L according meningitis criteria).

Data will be published monthly through StatLabo including serotype stratified according to patients' age and months.

Time-line

Steps	Lenght
Monitoring of invasive *S. pneumoniae* serotypes in patients aged ≥ 5 years old.	Years 1, 2 and 3
Real-time updating of StatLabo surveillance information using Quellung method.	Years 1, 2 and 3
Conferences.	Years 1, 2 and 3
Publication.	Year 3

Timeframe

See annexe 1

Project Benefits

1- Real-time monitoring of invasive *S. pneumoniae* serotypes and antibiotic resistance in adult in the province of Quebec.
2- Monitoring of IPD incidence in Quebec.
3- Comparison of actual provincial surveillance program using data from sentinel hospitals vs data from the study for individuals aged of ≥ 5 years old.
4- Data available for public health orientation on immunization program in adult population.

Deliverables

1- Monitoring of invasive *S. pneumoniae* strains in adult population for 3 years, starting in Septembre 2015 and ending in August 2018.

2- Monthly reporting of serotypes in StatLabo.

3- Data from the study will be presented at scientific meetings (AMMIQ [at the end of year 1], CACMID [at the end of year 2], ISPPD[at the end of year 3]) and published in a peer reviewed journal (Vaccine/PlosOne) at the end of the study.

References

1- Spellerberg, B. and Brandt C. *Streptococcus*. 2011. Manual of clinical microbiology. 10th edition. American Society for microbiology, Washington, D.C.

2- Regev-Yochay G, Paran Y, Bishara J, Oren I, Chowers M, Tziba Y, Istomin V, Weinberger M, Miron D, Temper V, Rahav G, Dagan R; IAIPD group. 2015. Early impact of PCV7/PCV13 sequential introduction to the national pediatric immunization plan, on adult invasive pneumococcal disease: A nationwide surveillance study. Vaccine. 25;33(9):1135-42.

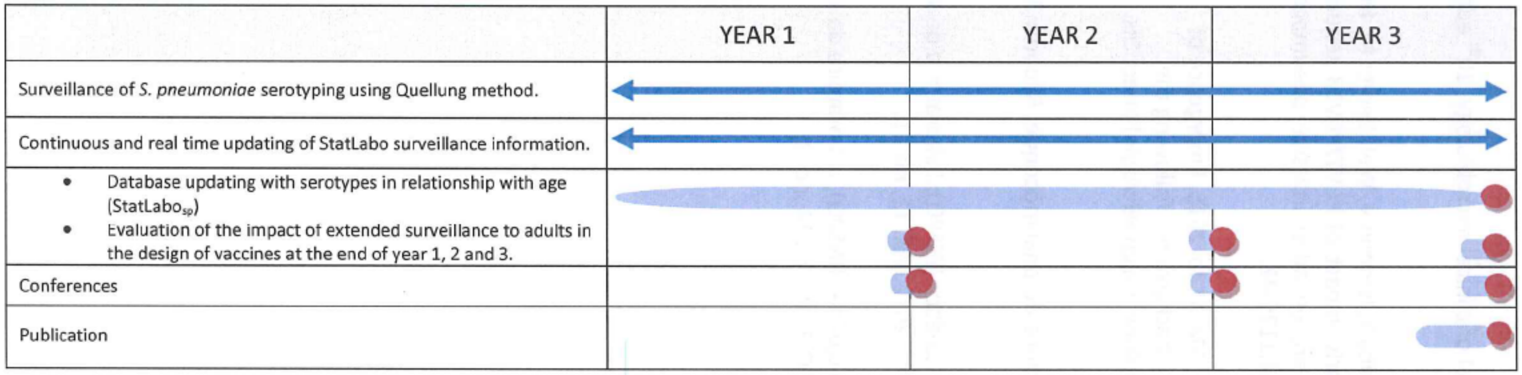
3- Liyanapathirana, V., EA. Nelson, I. Ang, R. Subramanian, H. Ma, M. Ip. 2015. Emergence of serogroup 15 *Streptococcus pneumoniae* of diverse genetic backgrounds following the introduction of pneumococcal conjugate vaccines in Hong Kong. Diagn Microbiol Infect Dis. Jan;81(1):66-70.

4- Lefebvre B., C. L. Tremblay. 2012. Programme de surveillance du pneumocoque. Rapport 2011. INSPQ. ISBN : 978-2-550-66364-5.

5- Bulletin STATLABO. Institut national de santé publique du Québec (INSPQ), Laboratoire de santé publique du Québec (LSPQ). Statistiques d'analyses du LSPQ. 2012. Vol. 11, no.11.

6- Bulletin STATLABO. Institut national de santé publique du Québec (INSPQ), Laboratoire de santé publique du Québec (LSPQ). Statistiques d'analyses du LSPQ.2015 Vol. 14, no. 2.

Annexe 1. Time Frame/Project Goals (arrows), milestones (red), task (blue bars) and timelines.



	YEAR 1	YEAR 2	YEAR 3
Surveillance of *S. pneumoniae* serotyping using Quellung method.	↔	↔	↔
Continuous and real time updating of StatLabo surveillance information.	↔	↔	↔
- Database updating with serotypes in relationship with age (StatLabo$_{sp}$) - Evaluation of the impact of extended surveillance to adults in the design of vaccines at the end of year 1, 2 and 3.	▬ ●	▬ ●	▬ ● ●
Conferences	●	●	●
Publication			▬ ●

FINAL
2015/09/18

[redacted]

New molecular tools for the serotyping of *Streptococcus pneumoniae* invasive strains in the province of Quebec

Principal investigator, Project leader

Brigitte Lefebvre, Ph.D., Microbiologist, LSPQ

Co-Principal investigator

Cécile Tremblay, MD, FRCPC, Département de Microbiologie, immunologie et Infectiologie
Université de Montréal

Co-Investigators

Simon Lévesque, Ph.D., Microbiologist, LSPQ

Sadjia Bekal, Ph.D., Microbiologist, LSPQ

Marc-Christian Domingo, Ph.D., Microbiologist, LSPQ

Technical and bioinformatics leader

Eric Fournier, D.E.S.S. in bioinformatics, M. Sc., Bioinformatics Scientist, LSPQ

Scientific coordinator

Florence Doualla-Bell, Ph.D., LSPQ

Scientific director

Jean Longtin, MD, LSPQ

Authors of the report

Eric Fournier, D.E.S.S. in bioinformatics, M. Sc., Bioinformatics Scientist, LSPQ

Brigitte Lefebvre, Ph.D., Microbiologist, LSPQ

Introduction

Since 1990, *Streptococcus pneumoniae* serotypes are determined using the Quellung's technique in most laboratories. This standard method uses antisera to reveal the swelling of the capsule through an antibody-antigen reaction. This technique is laborious, expensive and requires technical expertise. Although it is recognized as the reference method, it can lead to erroneous results because it is subjective. Indeed, serotyping results are determined through microscope observation of capsular swelling which in some cases is difficult to observe. As more than 90 serotypes of *S. pneumoniae* have been described to date, a serotyping algorithm must be applied using different antisera which makes the task tedious and time consuming.

In that context, we evaluated three molecular techniques for the rapid serotyping of *S. pneumoniae* invasive strains from children and adults in the province of Quebec. The results were compared with those obtained using Quellung gold standard.

Whole genome sequencing (WGS) is a technology that determines the complete DNA sequence of a microorganism's genome at a single time. Sequetyping is based on polymerase chain reaction (PCR) amplification of *cpsB* (capsular polysaccharide synthesis) using a single primer pair followed by nucleotide sequencing. Sequential multiplex PCR was used for capsular serotyping of pneumococci using various primers pairs. Primer selection and their arrangement for multiplexing were optimized based on the capsular serotype distribution found in Quebec.

Materials and Methods

Bacterial isolates

The 97 *S. pneumoniae* isolates used in this study are listed in Table 1. They cover 74 different serotypes previously identified by the Quellung reaction using Statens Serum Institut antisera. Purified genomic extracts were obtained using the Qiagen™ BioRobot M48 workstation and the MagAttract DNA Mini M48 Kit (Qiagen).

Whole Genome Sequencing (WGS)

Genomic extracts were quantified using the Quant-It™ PicoGreen® dsDNA Assay Kit (Life Technologies) and diluted to the working concentration (1 ng/µl) to initiate the library preparation.

Whole genome sequencing was performed on 21 pneumococci isolates (Table 1) using an Illumina MiSeq system and the Nextera XT DNA reagent kit v3 (600 cycles, paired ends). Genome size of *S. pneumoniae* is 2.16 Mbp on average. Using this value and the MiSeq Sequencing Coverage Calculator (http://support.illumina.com/downloads/sequencing_coverage_calculator.html), a minimum depth of coverage per isolate averaging 50X was obtained.

Bioinformatics tools

Following the MiSeq run, reads quality was evaluated with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Genome assemblies were performed using Spades (Bankevich *et al*., 2012) assembler on Calcul Quebec (http://www.calculquebec.ca/en/) public resources. Assemblies' metrics for each specimen were computed and visualized with Quast (Gurevich *et al*., 2013) and R scripts tools.

To detect the *cps* loci in each single fasta file assembly, 107 *cps* sequences representing 92 different serotypes (Camargo *et al*., 2015) were downloaded from the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/) using an in-house Biopython (http://biopython.org/wiki/Main_Page) script tool. A database containing those sequences was constructed and used as a list of subjects to successively Blast (Basic Local Alignment Search Tool) all assembly files with a second in-house python script tool. For each unknown isolate, the hit with the highest bit score was retained as the most probable corresponding serotype.

Sequential multiplex PCR

Pneumococcal serotypes of selected isolates (n=60) listed in Table 1 were tested using a sequential multiplex PCR protocol designed by the Centers for Disease Control and Prevention (CDC, http://www.cdc.gov). The list of 41 oligonucleotide pairs of primers and the product sizes are accessible at: http://www.cdc.gov/streplab/downloads/pcr-oligonucleotide-primers.pdf. The names of the primers correspond to their respective target serotype(s). The sequential multiplex approach consists of eight successive PCR reactions (reactions 1 to 8) and the reaction 6C used to resolve a positive amplification with primers 6A/6B/6C/6D in reaction 1. Each single reaction has its serotype-specific set of primers. They all have the universal capsular pair of primers CPSA-(forward and reverse) as positive control targeting every possible tested *cps* locus (except for serotype 38). Master mix component and thermal cycling parameters are detailed in the following document: http://www.cdc.gov/streplab/downloads/pcr-us-clinical-specimens.pdf. Electrophoresis was done using a 2% agarose gel and 25 µl reaction mix described in the conventional LSPQ routine procedure.

Sequetyping

Serotyping by sequetyping, based on the *cpsB* gene sequencing, was performed on selected isolates from Table 1 (n=74) according to Leung *et al*., (2012). The sequetyping primers are as follow: *cps1*, 5'-GCA ATG CCA GAC AGT AAC CTC TAT-3', and *cps2*, 5'-CCT GCC TGC AAG TCT TGA TT-3'. PCR amplification, amplicon purification, the first generation sequencing with the BigDye Sequence Terminator v.3.1 kit (Applied Biosystems) and the Genetic Analyser *3130* (Applied Biosystems) were performed according to the procedure commonly used in routine at the LSPQ.

BioNumerics version 7.5 (Applied Maths) was used to assemble forward and reverse abi sequences and to edit final consensus chromatograms. Consensus sequences were exported in a single multifasta file to perform phylogenetic analysis and Blast queries (see below).

The identification could be classified into one of the following levels (adapted from Leung *et al*., 2012): 1) Serotype level when the expected serotype was found with the highest identity value. 2) Serogroup level if the expected serotype was found with the highest identity value and this identity was shared with other serotype(s) of the same serogroup only. 3) Ambiguous, when condition 2) is true and the highest identity value is also or only shared with other serotypes. 4) Misidentified, when the highest identity value was obtained with a serotype different from the expected one.

TABLE 1 Serotypes and isolates ID used in this study and selected isolates for the seroptyping molecular methods tested.

		Tested serotyping methods		
Serotypes[1]	Isolates ID	WGS	Sequetyping	Sequential multiplex PCR
1	LSPQ3053		✓	✓
2	LSPQ3054		✓	✓
3	LSPQ3055		✓	✓
4	LSPQ3124		✓	✓
5	LSPQ3057		✓	✓
6A	LSPQ3058		✓	✓
6B	LSPQ3770		✓	✓
6C	LSPQ4242		✓	✓
6D	MA092686		✓	✓
7A	LSPQ4102		✓	✓
7B	LSPQ4103		✓	✓
7C	LSPQ4231		✓	✓
7F	MA099461		✓	✓
7F	KMA081946	✓		
8	LSPQ3596		✓	✓
9A	MA080418		✓	✓
9N	MA099463		✓	✓
9V	MA099234		✓	✓
10A	MA090174		✓	✓
10A	KMA095845	✓		
10A	KMA094933	✓		
10A	KMA094205	✓		
10B	MA080812		✓	
10F	MA075627		✓	✓
11A	MA090298		✓	✓
11A	KMA091851	✓		
11B	MA097930		✓	
11F	MA073130		✓	
12A	MA097699		✓	✓
12F	LSPQ3064		✓	✓
13	LSPQ3065		✓	✓
14	LSPQ3066		✓	✓
15A	MA099389		✓	✓

TABLE 1 (continued)

		Tested serotyping methods		
Serotypes[1]	Isolates ID	WGS	Sequetyping	Sequential multiplex PCR
15A	KMA096792-1	✓		
15A	KMA095336	✓		
15A	KMA094663	✓		
15A	KMA093977	✓		
15B	MA099177		✓	✓
15B	KMA096033	✓		
15B	KMA095997	✓		
15B	KMA094560	✓		
15C	MA096496		✓	✓
15F	MA083248		✓	✓
16A	MA065427		✓	
16F	LSPQ4236		✓	✓
16F	KMA093020	✓		
17F	MA098807		✓	✓
18A	LSPQ4243		✓	✓
18B	MA066814		✓	✓
18C	MA095139		✓	✓
19A	LSPQ3071		✓	✓
19A	KMA080288	✓		
19A	KMA080125	✓		
19A	KMA079789	✓		
19B	MA083042		✓	
19C	MA084138		✓	
19F	MA098992		✓	✓
20	LSPQ3072		✓	✓
21	LSPQ3160		✓	✓
22A	MA095877		✓	✓
22F	LSPQ4162		✓	✓
22F	KMA096962	✓		
22F	KMA094696	✓		
22F	MA094689	✓		
23A	LSPQ3769		✓	✓
23B	MA099469		✓	✓
23F	MA099467		✓	✓

TABLE 1 (continued)

		Tested serotyping methods		
Serotypes[1]	Isolates ID	WGS	Sequetyping	Sequential multiplex PCR
24[2]	MA096695	✓		
24B	MA094350		✓	✓
24F	MA099028		✓	✓
27	MA088547		✓	
28A	MA095690		✓	
29	LSPQ3079		✓	
29	KMA099083		✓	
31	LSPQ3080		✓	✓
32F	LSPQ3081		✓	
33A	MA086628		✓	✓
33F	MA099238		✓	✓
34	LSPQ3127		✓	✓
34	KMA099037			✓
34	KMA096961	✓		
35A	MA092229		✓	✓
35A	KMA082642			✓
35B	MA097723		✓	✓
35F	MA099195		✓	✓
36	LSPQ3641		✓	
37	LSPQ3645		✓	✓
38	LSPQ3642		✓	✓
39	LSPQ3646		✓	✓
40	LSPQ3162		✓	✓
41A	LSPQ3089		✓	
42	LSPQ3677		✓	✓
43	LSPQ3643		✓	
44	LSPQ3644		✓	✓
45	LSPQ3092		✓	
46	LSPQ3093		✓	✓
48	LSPQ3095		✓	

[1] Serotype determined by Quellung.

[2] Serotype to be determined, unusual cross reaction (24c-, 24d+, 24e+) with Quellung.

Results

Evaluation of the Whole Genome Sequencing approach

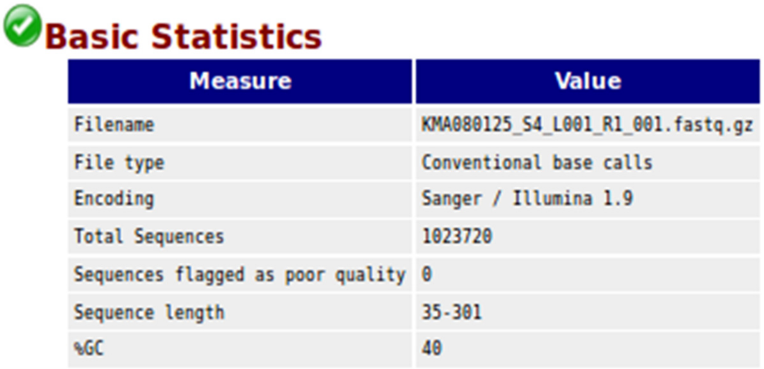
WGS is a powerful method which generates huge amount of data. Bioinformatics tools are essential to extract the information. First, following a MiSeq run, generated reads must be submitted to some statistics measurements such as their average lengths and quality. Second, they have to be assembled in order to construct higher levels of DNA sequences (contigs). A fully closed genome with a single contig is usually not expected due to the short length of the reads. Nonetheless, contigs with high depth of coverage and long enough are expected so that their concatenated lengths cover the totality of the target genome. Resulting assemblies hold a lot of garbage data which are not always required. To identify genes or regions of interest, genome annotation is a strategy which is often used. However, simple Blast analyses have also proven very efficient and are sometimes sufficient to obtain reliable responses. Following are the results for each main step of our analysis pipeline to identify serotypes of pneumococcal isolates with the WGS approach.

Paired end reads quality

FastQC is a simple tool used to summarize statistics of reads in Fastq (https://en.wikipedia.org/wiki/FASTQ_format) format files. Figures 1 to 3 give an example of a partial FastQC report generated with this tool on the KMA080125 forward reads file. Results are visually very easy to interpret. Other metrics are also generated by the program. For example, per sequence GC content, Kmer content, overrepresented sequences (data not shown) but their values don’t usually have any impact on the rest of the pipeline steps. The focus is normally directed only on the reads quality score and their average length.

In the KMA080125 example, the amount of forward reads is 1 023 720 (Figure 1). The number of reverse reads is always the same due to the paired end mode. Their lengths vary between 35 and 301 bp (Figure 1) with an average around 300 bp (Figure 3) and their quality is high across most of their lengths. Lower quality beyond 260 bp is an expected result due to the MiSeq chemistry. Those statistics are deemed of good quality, albeit not optimum, and are very acceptable based on MiSeq specifications and appropriate for the assembly step.

Each isolate has a FastQC report similar to the one generated for KMA080125. The metric having the highest variance is the number of reads (standard deviation=277 143 reads). Nonetheless, according to our assembly and Blast results (see below), this did not have any significant impact. The number of reads (forward + reverse) among isolates is given in Table 2. It varies between 543 274 reads (KMA096961) and 2 306 692 reads (KMA093977). The high variability and lower number of reads (33 570 998 reads) compared to the MiSeq performance specification (44 000 000 - 50 000 000 reads) could be explained by two factors : first, the high rate of reads filtration; second, the lower cluster density value (1000 k/mm^2) obtained during the MiSeq run compared to the specification value (1200-1400 k/mm^2). The problem may stem from the library preparation which is a rather complex procedure compared to a simple PCR and implies many steps subject to DNA loss. Accuracy of the original DNA concentration assay is also a potential source of unexpected results. Optimization of the library preparation step and investigation regarding this issue are part of our future plan.



Basic Statistics

Measure	Value
Filename	KMA080125_S4_L001_R1_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1023720
Sequences flagged as poor quality	0
Sequence length	35-301
%GC	40

Figure 1: Statistics summary of the KMA080125_S4_L001_R1.fastq file (forward reads) computed with FastQC.

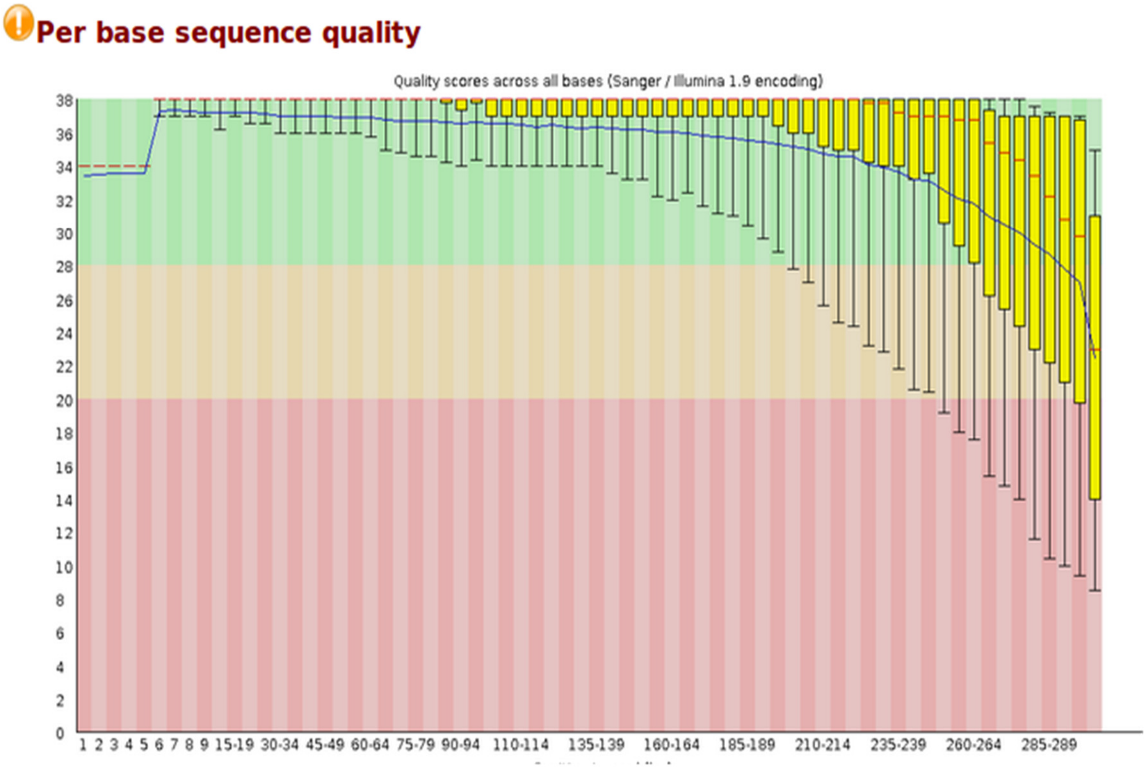


Figure 2: Box plot representing the average quality across reads length for the KMA080125 forward reads file. Reads positions are located on the horizontal scale and the Phred quality scores on the vertical scale. Green, yellow and red rectangles correspond to high, medium and poor quality base calls, respectively.

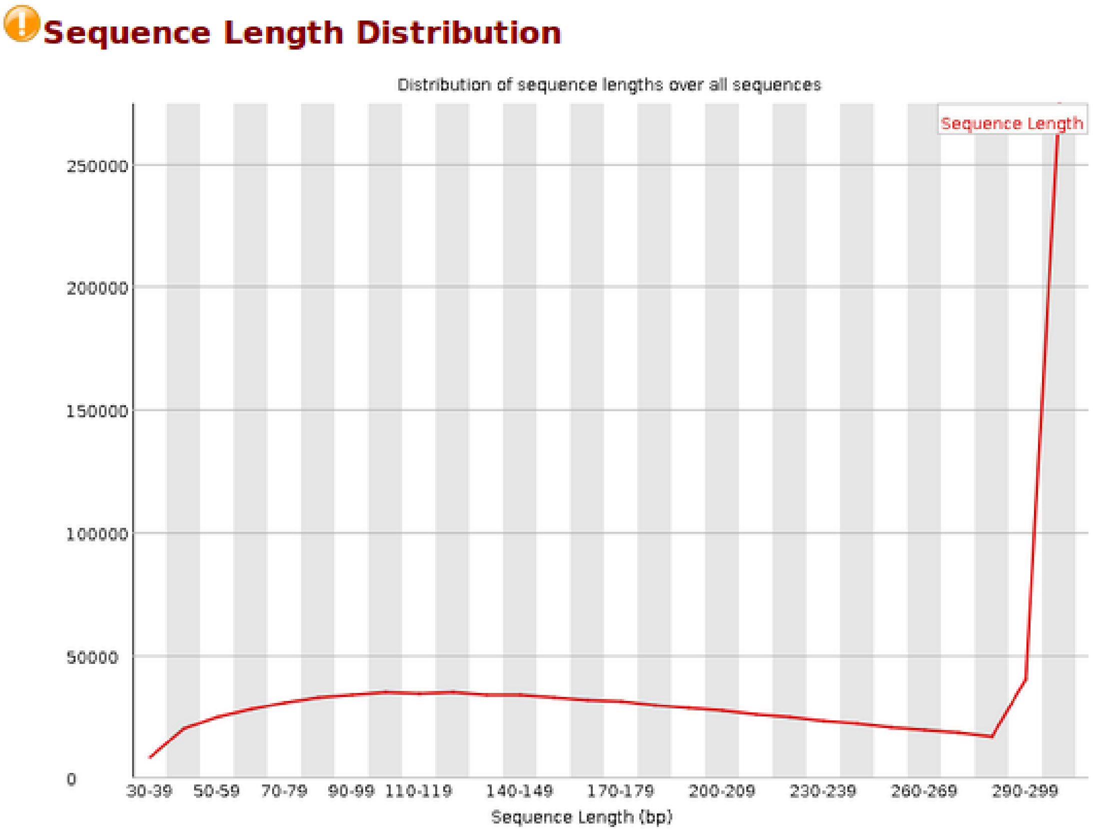


Figure 3: Reads length distribution for the KMA080215 forward Fastq reads file.

TABLE 2 Paired end reads number generated during the MiSeq run

Isolates	Reads numbers[1]
KMA079789	406 715
KMA085125	1 023 720
KMA080288	540 195
KMA081946	713 798
KMA091851	1 068 051
KMA093020	1 053 173
KMA093977	1 153 346
KMA094205	1 150 155
KMA094560	979 655
KMA094663	406 573
KMA094689	489 875
KMA094696	788 724
KMA094933	759 133
KMA095336	814 704
KMA095845	878 063
KMA095997	985 351
KMA096033	739 530
KMA096792-1	735 845
KMA096961	271 637
KMA096962	301 144
MA096695	500 784
Total reads	15 760 171

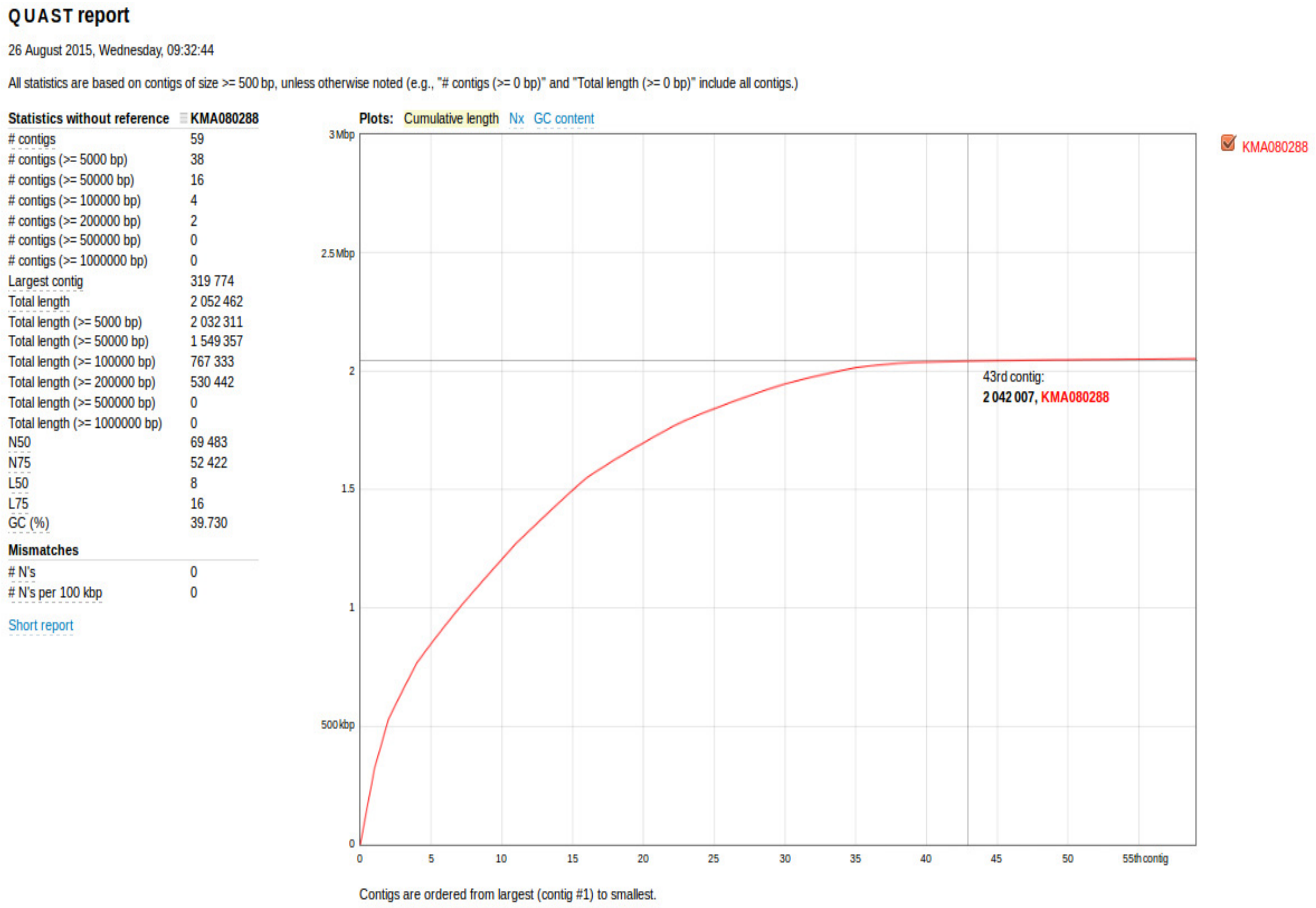
[1] The total reads number (forward + reverse) for one isolate is two times the displayed value.

Reads assemblies metrics

Genome assemblies is a very complex task which involve complex mathematic algorithms. To date, many assemblers have emerged (http://assemblathon.org/). Those implemented with De Bruijn Graphs (Pevzner *et al.*, 2001) are now considered the most efficient assemblers. Spades has been designed using such an algorithm and is particularly well adapted to manage MiSeq paired ends reads. Previous comparisons with other assemblers such as Velvet (Zerbino, 2010) and Ray (Boisvert *et al.*, 2010) have shown Spades to generate better metrics.

Different metrics are used to evaluate the quality of an assembly. The N50 statistic is well suitable for this. This parameter is defined as the length of the contig for which the sum of the length of all contigs of that length or shorter is higher than half of the sum of the length of the contigs collection. The distribution of contigs length, their coverage and the total contigs length compare to that of the reference genome are also indicative of good or bad assemblies.

To compute metrics of our assemblie's collection, we have used another well designed quality check tool named Quast. This program is easy to use and generate instantly all length statistics at a glance. Figure 4 show an example of a Quast output generated with the KMA080288 assembly. The number of contigs larger than 500 bp is 59, the largest contig is 319 774 bp in length and the N50 statistic is 69 483 bp. The graphic in the right panel is interactive and allows the user to visualize the cumulative length of the assembly starting with the largest contig. The example shows that at the 43rd contig, the cumulative length is 2 042 007 bp. Given that the average genome length of *S. pneumoniae* is 2.16 Mbp, we can consider that those metrics were expected and appropriate for the current project.



QUAST report

26 August 2015, Wednesday, 09:32:44

All statistics are based on contigs of size >= 500 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs.)

Statistics without reference	KMA080288
# contigs	59
# contigs (>= 5000 bp)	38
# contigs (>= 50000 bp)	16
# contigs (>= 100000 bp)	4
# contigs (>= 200000 bp)	2
# contigs (>= 500000 bp)	0
# contigs (>= 1000000 bp)	0
Largest contig	319 774
Total length	2 052 462
Total length (>= 5000 bp)	2 032 311
Total length (>= 50000 bp)	1 549 357
Total length (>= 100000 bp)	767 333
Total length (>= 200000 bp)	530 442
Total length (>= 500000 bp)	0
Total length (>= 1000000 bp)	0
N50	69 483
N75	52 422
L50	8
L75	16
GC (%)	39.730
Mismatches	
# N's	0
# N's per 100 kbp	0

Short report

43rd contig:
2 042 007, KMA080288

Contigs are ordered from largest (contig #1) to smallest.

Figure 4: Contigs length statistics generated by the Quast software. KMA080288 assembly as input file is shown.

As previously mentioned, the quality of an assembly is also based on the mean depth of coverage. In short, this value represents the mean frequency at which a single specific nucleotide has been called during a whole genome sequencing run. Generally, the higher this value is, the more confident we are in our assembly. Naturally, the depth of coverage will always be compromised as we increase the number of specimens in a single run and even more when their genome length increase. For this reason, the MiSeq Sequencing Coverage Calculator is highly useful during a run planning.

Unfortunately, Quast software is not well adapted to compute contigs coverage. All necessary data to make such computations belong exclusively to the assembler. In order to extract them from Spades and display coverage distribution graphically among contigs for every isolate, we implemented an in-house R script. This tool reads the fasta contigs headers from their respective assembly file and produces two graphics for each single isolate; one histogram depicting the absolute depth of coverage distribution and one complement linear plot showing the coverage's values relative to the contigs length. Figures 5 and 6 illustrate an output example for KMA080288. For this isolate, the average depth of coverage is 43X across 2 032 311 bp and that mostly smaller contigs have higher coverage. This observation applies to all final assemblies that we have generated in the current project.

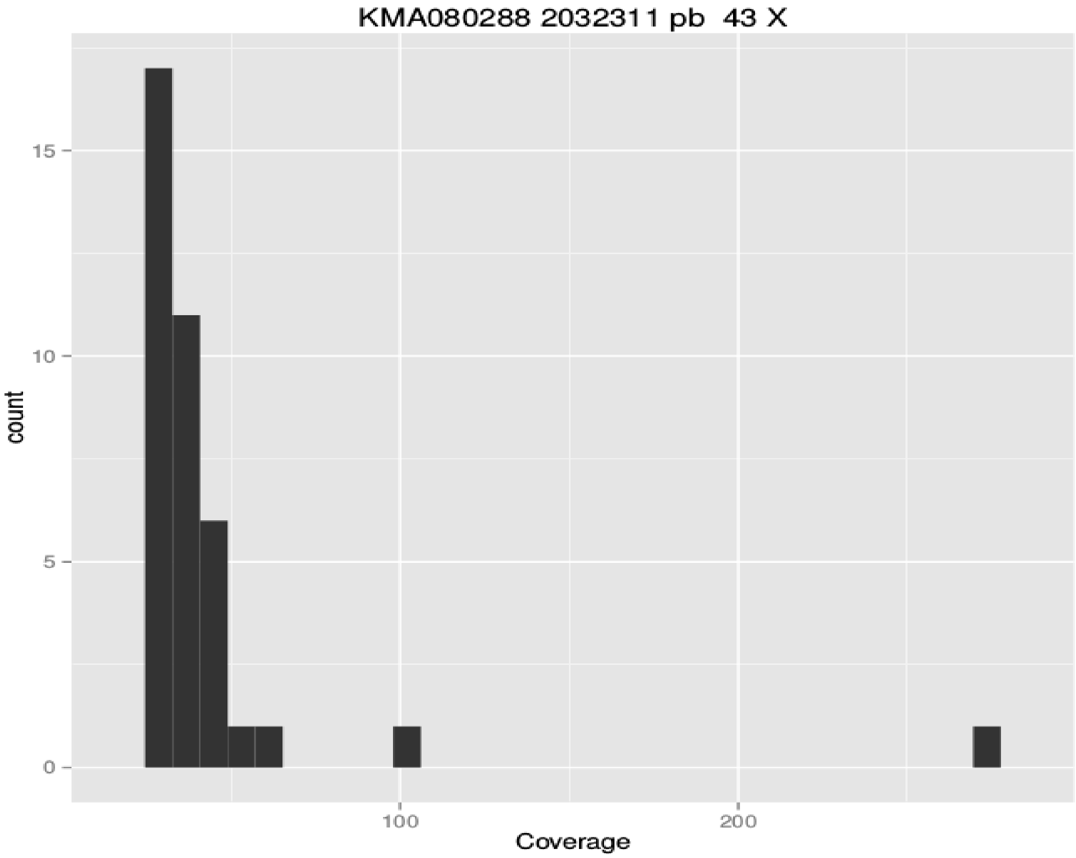


Figure 5: Depth of coverage distribution among the KMA080288 contigs collection generated with the Spades assembler.

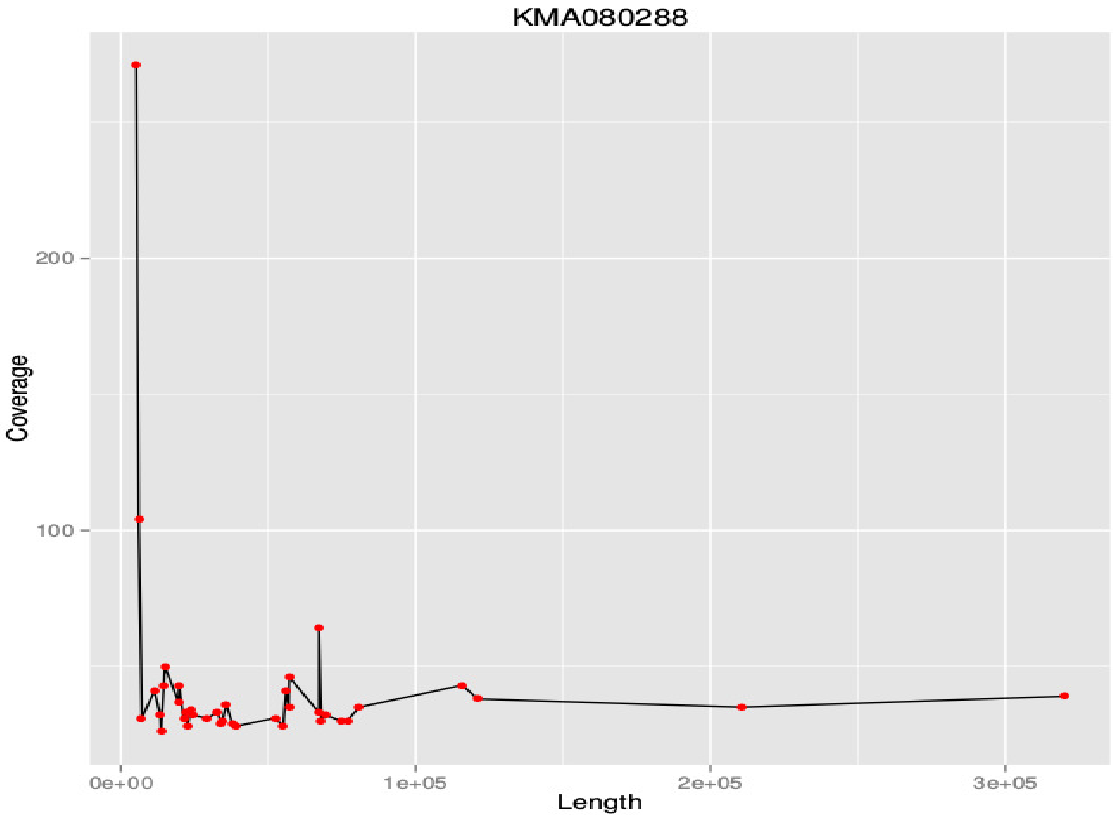


Figure 6: KMA080288 contigs depth of coverage relative to their respective length.

The main assembly's metrics for all of the 21 pneumococcal isolates are summarized in Table 3. Those metrics exclude contigs smaller than 500 bp. These short sequences are often unreliable and usually part of background data regarded as garbage. This filtration procedure has no impact on the final result. Globally, Table 3 shows that statistics are rather heterogeneous among isolates and seem to be correlated with the number of reads (Table 2). Effectively, one can notice for example that metrics quality of KMA093977 (the isolate having the highest number of reads) are considerably better than those obtained with KMA096961 (the isolate with the lowest number of reads). Nonetheless, based on the global view of the number of reads and contigs statistics, we can conclude that the MiSeq sequencing and the assembly steps have both been successful. That is, resulting data are and appropriate for downstream analysis.

TABLE 3 Summary of Spades assembly's metrics(1)(2)

Isolates	Assembly's Length (bp)	Largest contig (bp)	N50	Mean coverage (X)
KMA079789	2 111 694	289 688	71 633	31
KMA080125	2 032 728	**340 957**	71 895	90
KMA080288	2 032 311	319 774	69 483	43
KMA081946	1 970 356	**115 076**	67 068	47
KMA091851	1 984 531	151 627	71 048	77
KMA093020	2 051 899	235 604	113 800	70
KMA093977	2 040 673	330 076	74 270	71
KMA094205	2 053 258	330 614	**115 223**	**92**
KMA094560	2 118 213	151 822	80 855	83
KMA094663	2 061 526	176 268	**54 348**	56
KMA094689	2 059 983	207 974	66 632	51
KMA094696	2 069 329	243 814	86 596	70
KMA094933	1 994 414	303 524	86 936	80
KMA095336	2 099 705	176 281	65 535	79
KMA095845	2 043 568	303 918	98 395	90
KMA095997	2 056 093	254 966	86 217	71
KMA096033	**2 168 500**	169 702	84 611	66
KMA096792-1	2 041 772	241 467	88 561	73
KMA096961	**1 968 716**	133 241	64 281	**30**
KMA096962	2 026 356	257 300	98 394	35
MA096695	2 061 094	220 680	88 008	57

(1) All statistics are based on contigs having length ≥ 500 bp.

(2) Numbers in green and red indicate the highest and lowest values, respectively.

Serotype determination using Blast queries

As described in the Materials and Methods section, we have constructed a small database with 107 different *cps* loci and use it as template to execute Blast queries for each assembly files. Serotype identifications were based on High-scoring Segment Pairs (HSP) length (an alignment length between a query and a subject DNA sequence) and identities values (a combination of HSP with identity value results in a bit score). Table 4 summarizes the Blast results for every isolates and their corresponding expected serotype previously obtained by Quellung reaction. Figure 7 depicts an example of HSP alignment for isolate KMA081946 and describes some of the technical terms appearing in the Table 4 header.

In every case, the correct serotype was found with 98-100% HSP identity. Nonetheless, some Blast results could not perfectly discriminate between two different serotypes because of their high degree of genetic similarities or due to the existence of DNA polymorphism among single serotype (Varvio *et al*., 2009). This is the case for KMA094560 (15B/15C), KMA095977 (15B/15C), KMA096033 (15B/15C), KMA095336 (15A/15F), KMA094689 (22A/22F), KMA094696 (22A/22F), KMA096962 (22A/22F), KMA081946 (7A/7F) and KMA091851 (11A/11D). Remember that 15B and 15C are considered as one serotype since they interconvert (Pai *et al*., 2006). Regarding unresolved serotypes 22A/22F, 7A/7F and 11A/11D, more sensitive genetic analysis methods would be required to make a more accurate identification. For example, one could make the identification of non-synonymous single nucleotide polymorphism (SNP) and establish a relation with either serotype.

Another observation extracted from our Blast analysis, is that most of the best hits HSP's doesn't completely cover the *cps* locus reference sequence. Missing segments in the query sequences are always located at both ends of the *cps* locus and correspond to transposase-like regions (*tnp*). Refer to Figure 8 for an example with isolate KMA079789. According to Bratcher *et al*., 2011, those regions may contribute to the vertical exchange of the *cps* locus between pneumococcal strains and hence to their molecular evolution and adaptation. However, the *tnp* regions are not always present in the *cps* locus which explains why the *cps* locus in our isolates is often shorter.

Interestingly, three isolates, KMA094689, KMA094696 and KMA096962 (Table 4), match serotypes 22F/22A but with two separates HSPs (Figure 9). We found that this unexpected Blast result is caused by the high divergence of two genes (*wcwA* and *wcwC*) in the *cps* locus of those isolates compared to their orthologous sequences in serotype 22F. Similar finding was reported for strain 1772-40b (GenBank accession HE651318; Salter *et al*., 2012), a 22F serotype which matches perfectly with our 22F isolates.

TABLE 4 Pneumococcal serotypes identification using whole genome sequencing and Blast queries

Isolates	Query contigs length (bp)	*cps* best hit subject			HSP[1]		Expected serotype[2]
		GenBank accession	Serotype	Length (bp)	Identity (%)	Length (bp)	
KMA079789	51 165	CR931675	19A	18 617	98.5	15 141	19A
		AF094575	19A	18 754	98.4	15 141	
KMA080125	340 957	CR931675	19A	18 617	98.5	15 141	19A
		AF094575	19A	18 754	98.4	15 141	
KMA080288	319 774	CR931675	19A	18 617	98.4	15 141	19A
		AF094575	19A	18 754	98.4	15 141	
KMA081946	88 288	CR931640	7A	24 019	99.9	24 019	7F
		CR931643	7F	24 127	99.9	24 128	
KMA091851	15 007	CR931653	11A	17 948	99.8	14 755	11A
		CR931656	11D	17 213	100	14 755	
KMA093020	19 703	CR931668	16F	21 481	99.9	19 714	16F
KMA093977	330 076	CR931663	15A	18 517	99.7	18 518	15A
KMA094205	15 301	CR931649	10A	17 290	100	15 301	10A
KMA094560	80 972	CR931664	15B	18 624	99.3	17 288	15B
		CR931665	15C	18 626	99.4	17 288	
KMA094663	94 633	CR931663	15A	18 517	99.7	18 306	15A
KMA094689	108 827	CR931681	22A	22 591	97.9	12 897	22F
					98.0	7 721	
		CR931682	22F	22 696	97.9	12 897	
					98.0	7 721	
KMA094696	21 537	CR931681	22A	22 591	97.9	12 897	22F
					98.0	6 535	
		CR931682	22F	22 696	97.9	12 897	
					98.0	6 535	
KMA094933	303 524	CR931649	10A	17 290	99.2	15 133	10A
KMA095336	94 525	CR931663	15A	18 517	99.7	18 306	15A
		CR931666	15F	22 405	99.2	12 386	
KMA095845	303 918	CR931649	10A	17 290	99.2	14 679	10A
KMA095997	80 971	CR931664	15B	18 624	99.3	17 288	15B
		CR931665	15C	18 626	99.4	17 288	

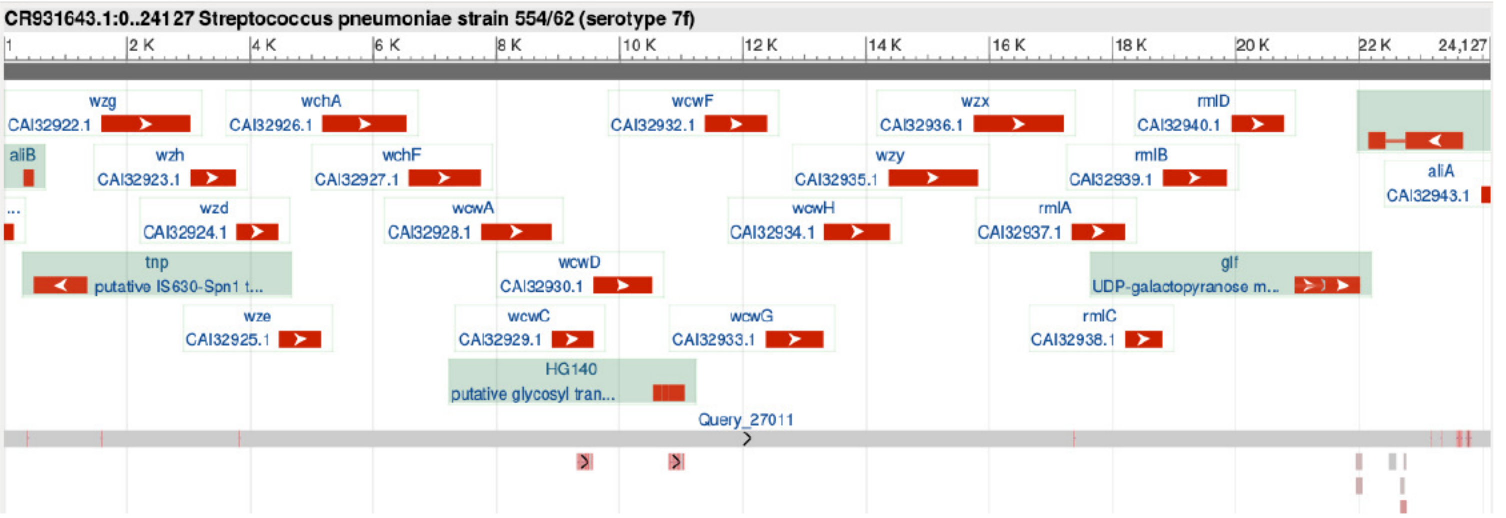
TABLE 4 (continued)

Isolates	Query contigs length (bp)	*cps* best hit subject			HSP[1]		Expected serotype[2]
		GenBank accession	Serotype	Length (bp)	Identity (%)	Length (bp)	
KMA096033	80 939	CR931664	15B	18 624	99.3	17 286	15B
		CR931665	15C	18 626	99.3	17 288	
KMA096792-1	240 139	CR931663	15A	18 517	99.8	18 386	15A
KMA096961	16 885	CR931703	34	15 938	99.9	14 859	34
KMA096962	108 311	CR931681	22A	22 591	97.8	12 897	22F
					97.7	7 721	
		CR931682	22F	22 696	97.8	12 897	
					97.7	7 721	
MA096695	220 680	CR931687	24B	23 976	98.9	22 332	24[3]

[1] HSP = high-scoring Segment Pairs.

[2] Expected serotype according to Quellung reaction.

[3] Serogroup 24. Serotype to be determined, unusual cross reaction (24c-, 24d+, 24e+).



HSP length = 24 128 bp
Identity = 99.9%

Figure 7: Best HSP hit resulting from the Blast query execution of the KMA081946 assembly file on the *cps* database. The dark grey segment in the upper part of the figure represents the serotype 7F *cps* locus (GenBank accession CR931643, 24 127 bp) aligned with the KMA081946 heterologous region (light gray segment). Red region on the KMA081946 sequence represent mismatches and red arrowed segments between the two aligned sequences correspond to coding sequences (with their respective GenBank accession number given in blue) part of the CR931643 7F *cps* locus.

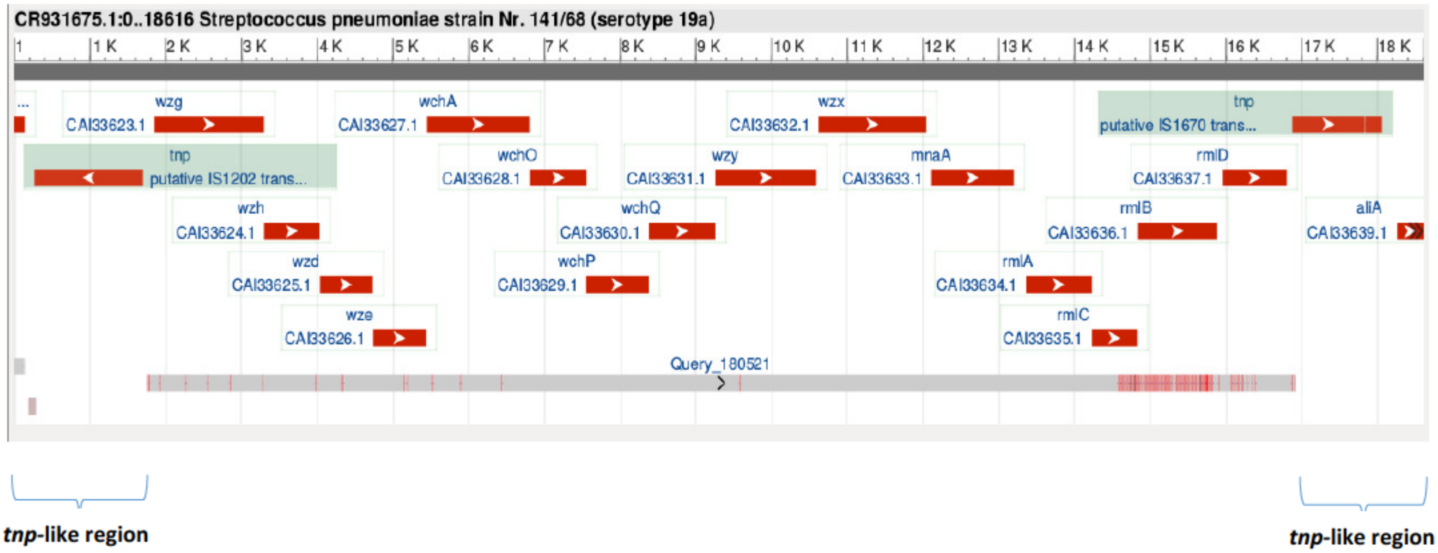


Figure 8: Blast analysis of the KMA079789 assembly showing the absence of *tnp*-like regions at both ends of the *cps* locus in the query sequence (lower gray segment).

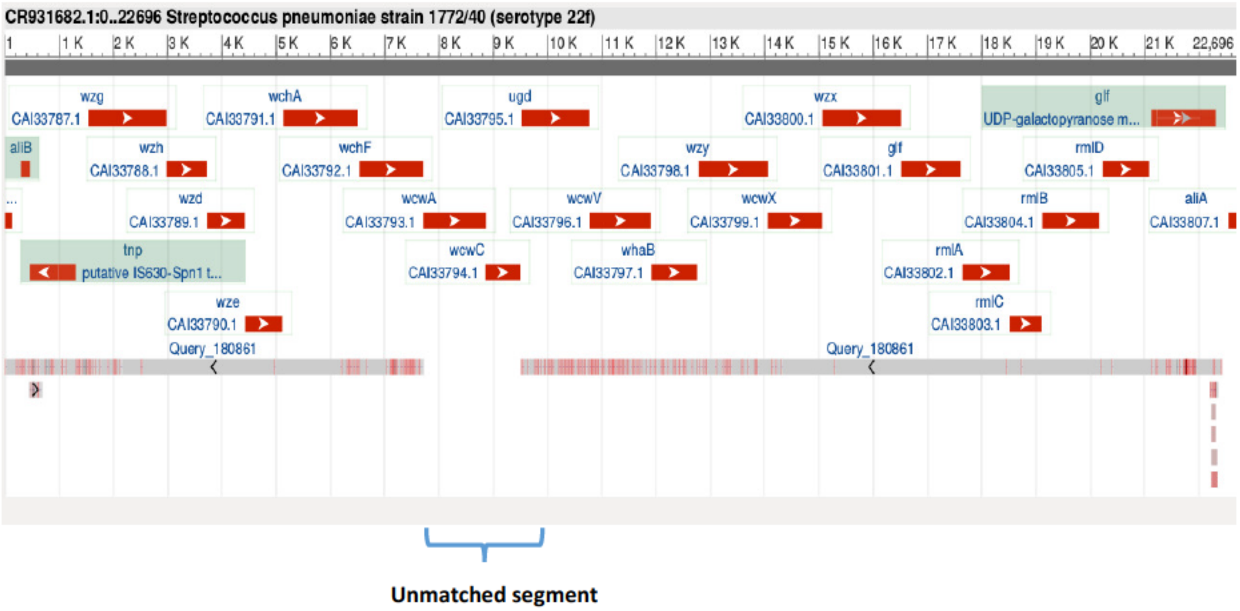


Figure 9: Disrupted HSP between a KMA094689 contig and the orthologous *cps* locus of a 22F reference serotype (GenBank accession CR931682). Notice the unmatched region corresponding to genes *wcwA* and *wcwC.*

Evaluation of the multiplex PCR CDC protocol

PCR methods are very powerful, reliable and rather easy to perform. Multiplex PCR is an even more efficient technique since one single reaction allows the simultaneous detection of more than one gene and/or allele. However, designing a multiplex PCR protocol is not an easy task. First, primers must effectively target the region of interest (i.e. primers specificity). Second, the possibility of unwanted intra- and inter-hetero duplex structures arising between primers must be predicted *in silico*. Third, amplicon length must also have the appropriate length combination in order to facilitate the interpretation of the electrophoresis migration profile.

In our case, the multiplex approach used to identify serotypes of unknown pneumococcal samples is further complicated by the fact that epidemiological data must also be considered. PCR master mixes in a sequential strategy are prepared such that most common serotypes (according to the time-space parameter) may be detected in the first step. Obviously, since the serotypes distribution across Quebec is not the same as the ones circulating in USA at the same period, the CDC algorithm should be adapted according Quebec's data. The sequential reaction order could be modified but the primers combination in each of them must stay intact to avoid any unexpected results such as false positives and false negatives.

All primer pairs have been designed in such a way that they bind DNA regions or genes specific to their targeted serotype *cps* locus. However, serotypes among a serogroup due to their high level of genetic homogeneity are inevitably revealed under the same signal in the current protocol. For example, primer pair 6A/6B/6C/6D in reaction 1 is simultaneously specific to four different serotypes. This may be a significant disadvantage relative to the Quellung reaction. However, the multiplex PCR approach is a cost effective method.

All reaction mixtures have been tested with isolates of known serotype previously identified by the Quellung reaction. At this moment, our isolates library at the LSPQ doesn't cover all of the 92 possible serotypes. Then, we tested at least one appropriate isolate for every primer pair's evaluation. Notice that every reaction includes the primer pair CPSA-F/CPSA-R as internal control. Those primers target the *cpsA* gene, a highly conserved gene which belongs to the regulatory region of the *cps* locus. Figures 10 to 18 show our electrophoresis results obtained for all PCR multiplex reactions and may be compared to the expected CDC results available at http://www.cdc.gov/streplab/downloads/pcr-us-clinical-specimens.pdf.

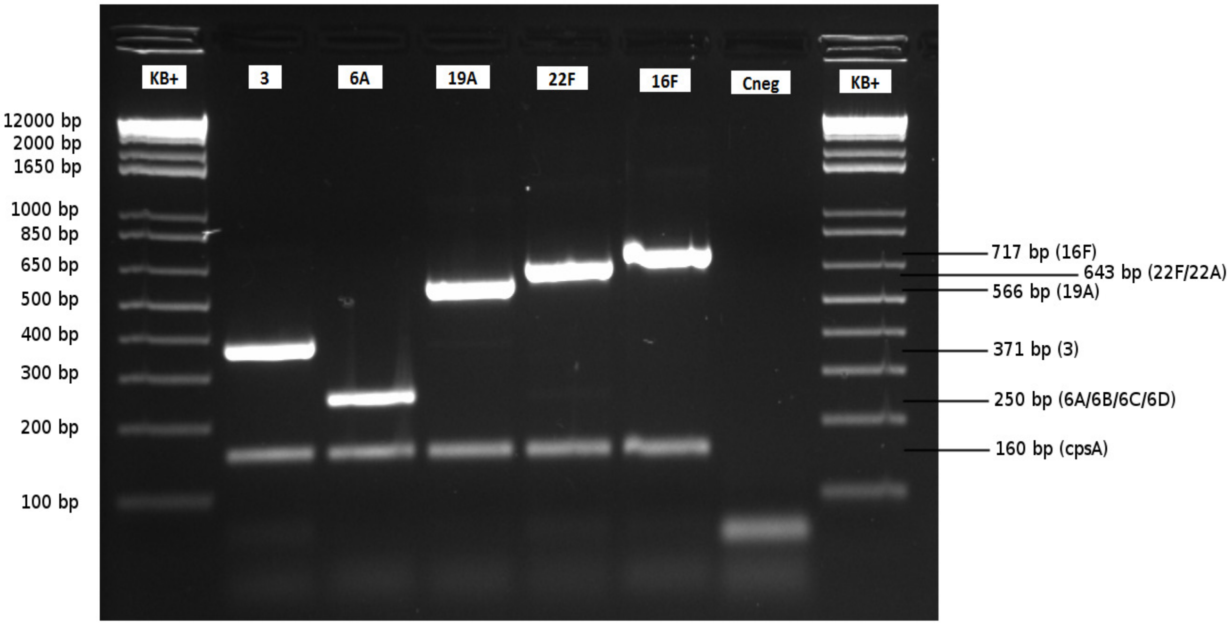


Figure 10: Electrophoresis profile obtained with the multiplex reaction 1 used to detect serotypes 3, 6A/6B/6C/6D, 19A, 22A/22F, 16F.

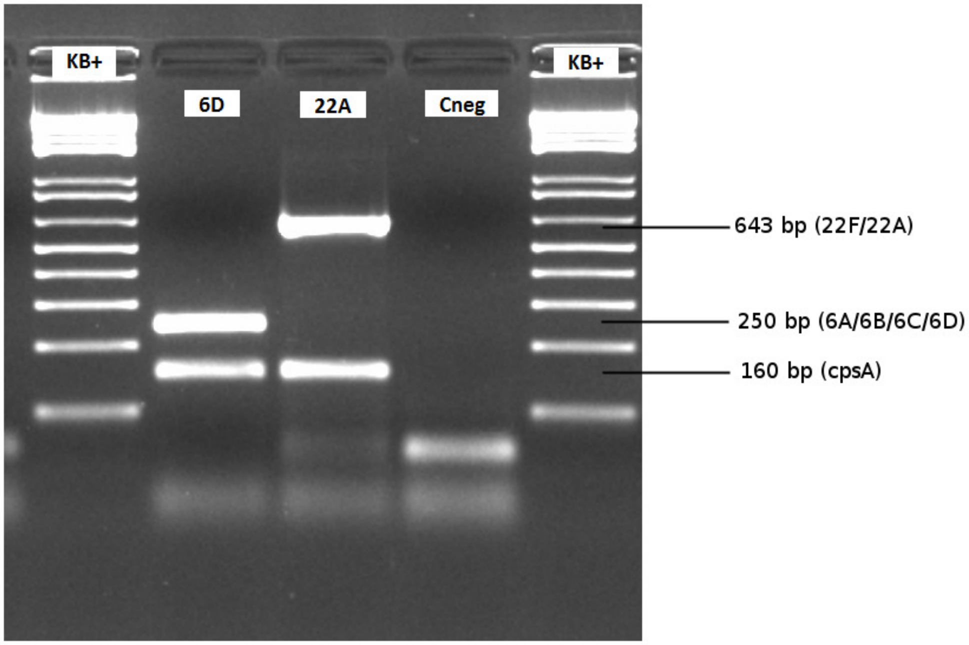


Figure 10 (continued).

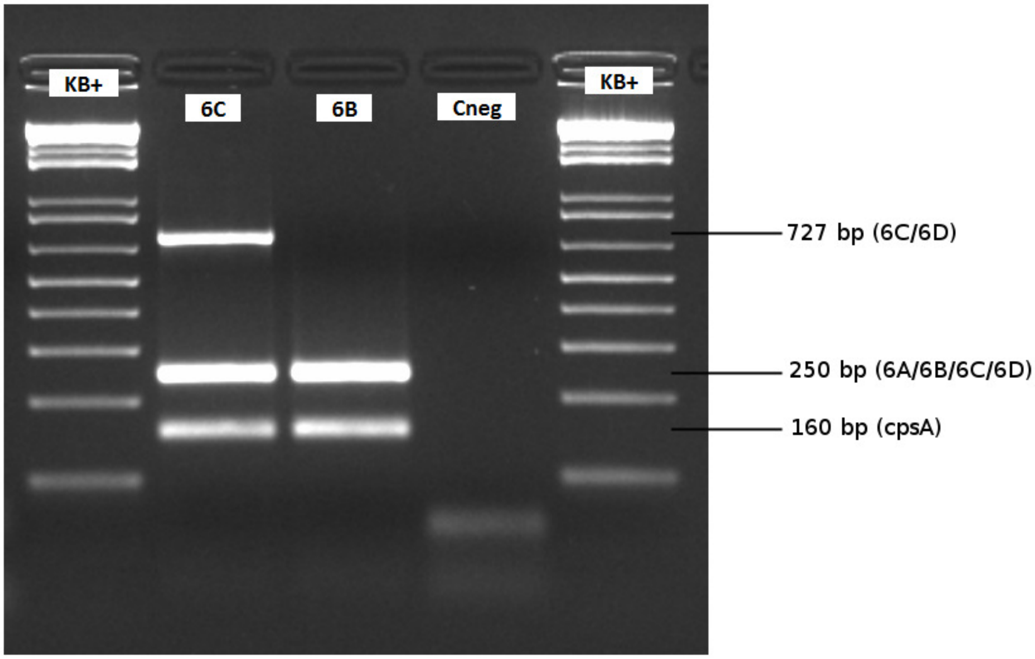


Figure 11: Electrophoresis profile obtained with the multiplex reaction 6C used to detect serotypes 6C/6D.

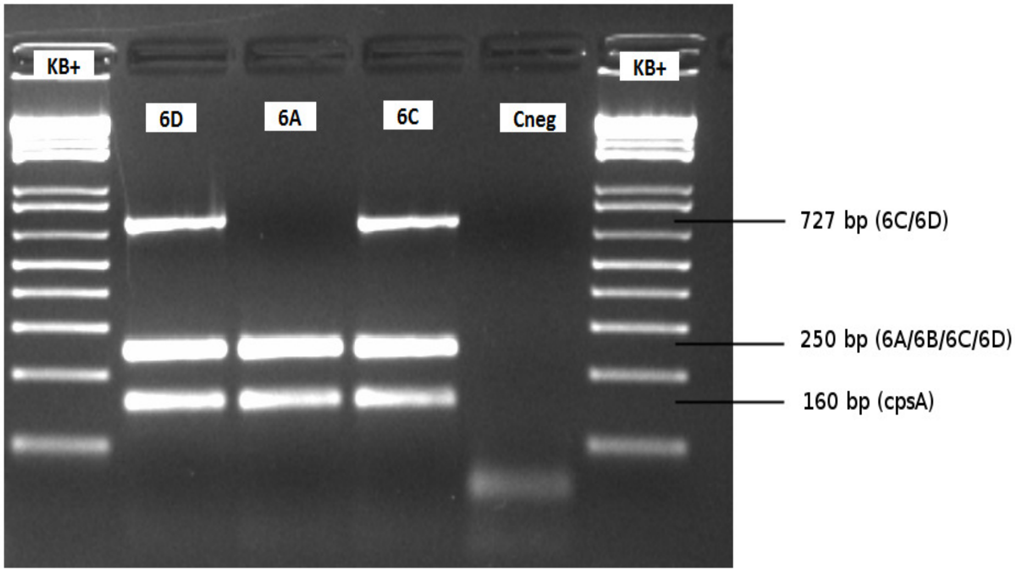


Figure 11 (continued).

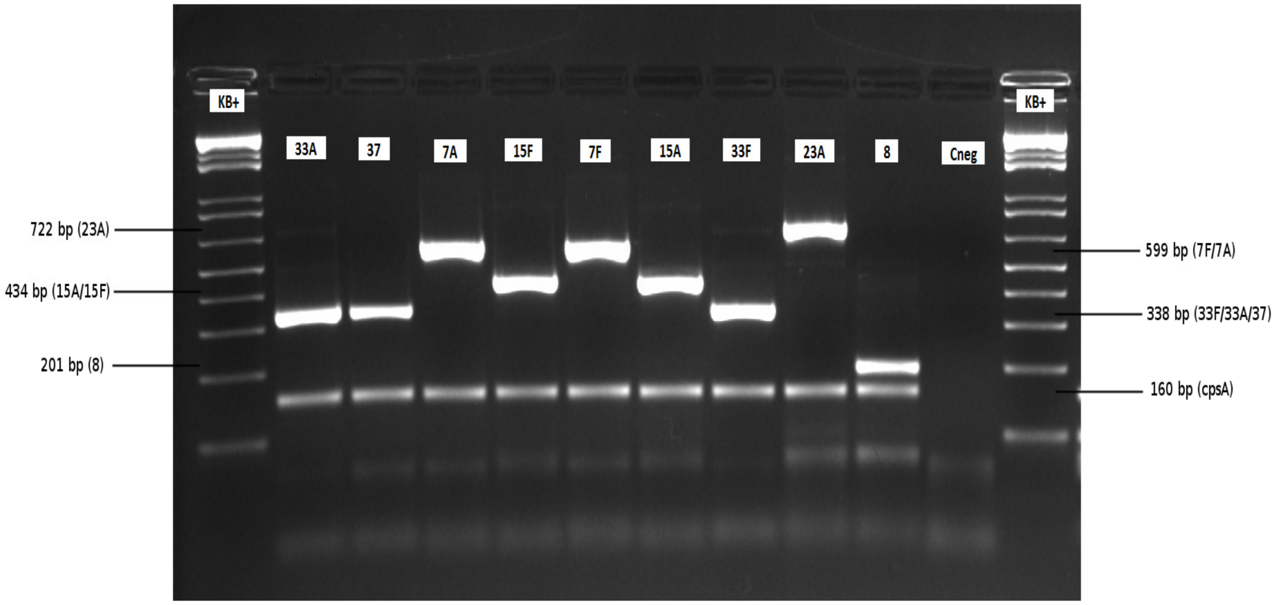


Figure 12: Electrophoresis profile obtained with the multiplex reaction 2 used to detect serotypes 8, 33F/33A/37, 15A/15F, 7F/7A and 23A.

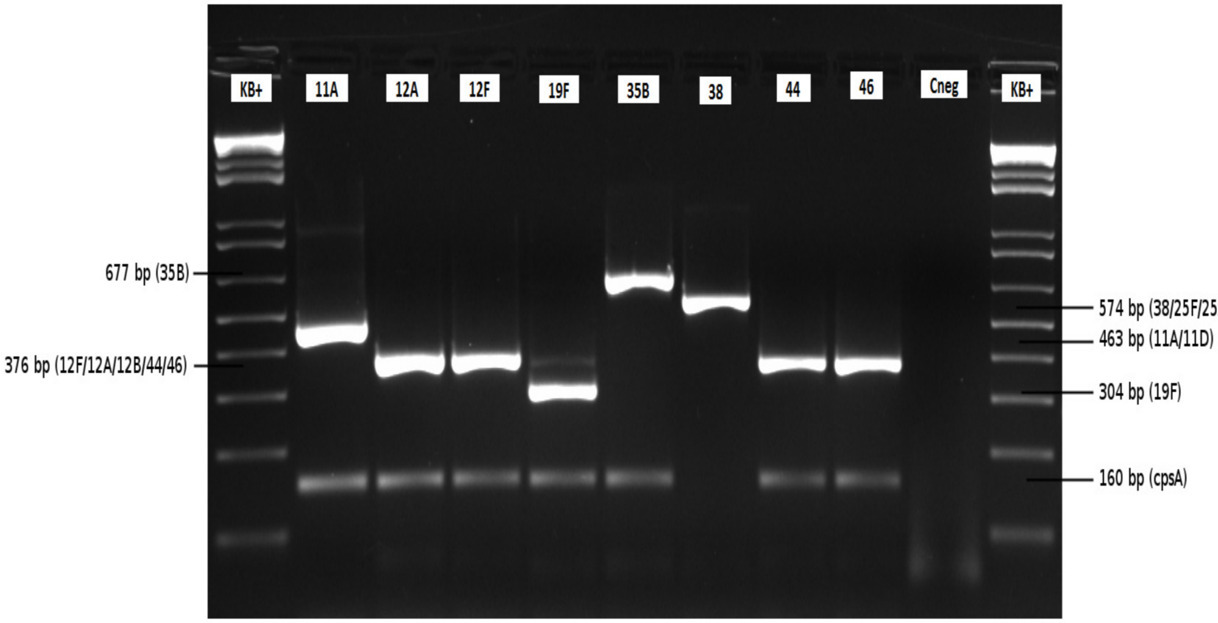


Figure 13: Electrophoresis profile obtained with the multiplex reaction 3 used to detect serotypes 19F, 12F/12B/12A/44/46, 11A/11D, 38/25F/25A and 35B.

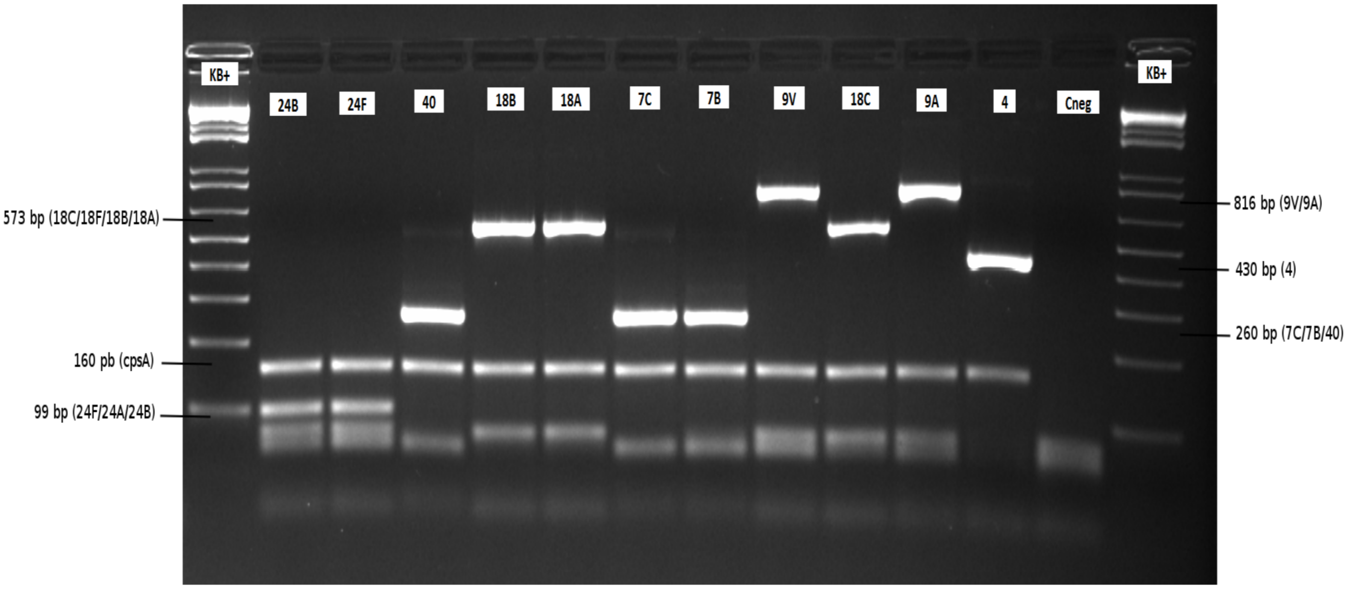


Figure 14: Electrophoresis profile obtained with the multiplex reaction 4 used to detect serotypes 24F/24A/24B, 7C/7B/40, 4, 18C/18F/18B/18A and 9V/9A.

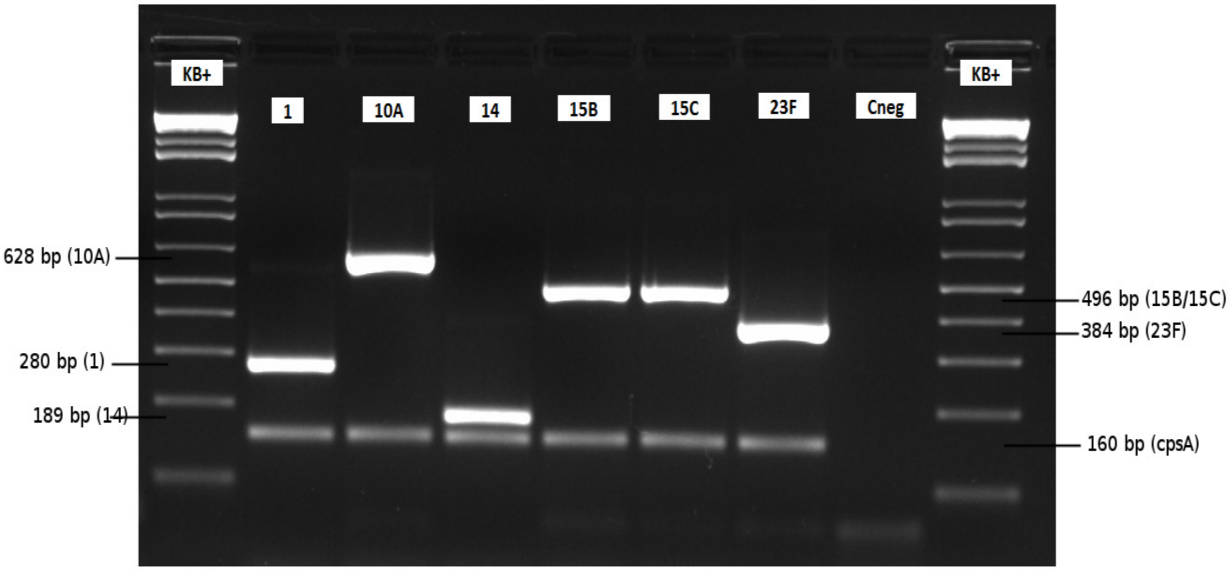


Figure 15: Electrophoresis profile obtained with the multiplex reaction 5 used to detect serotypes 14, 1, 23F, 15B/15C and 10A.

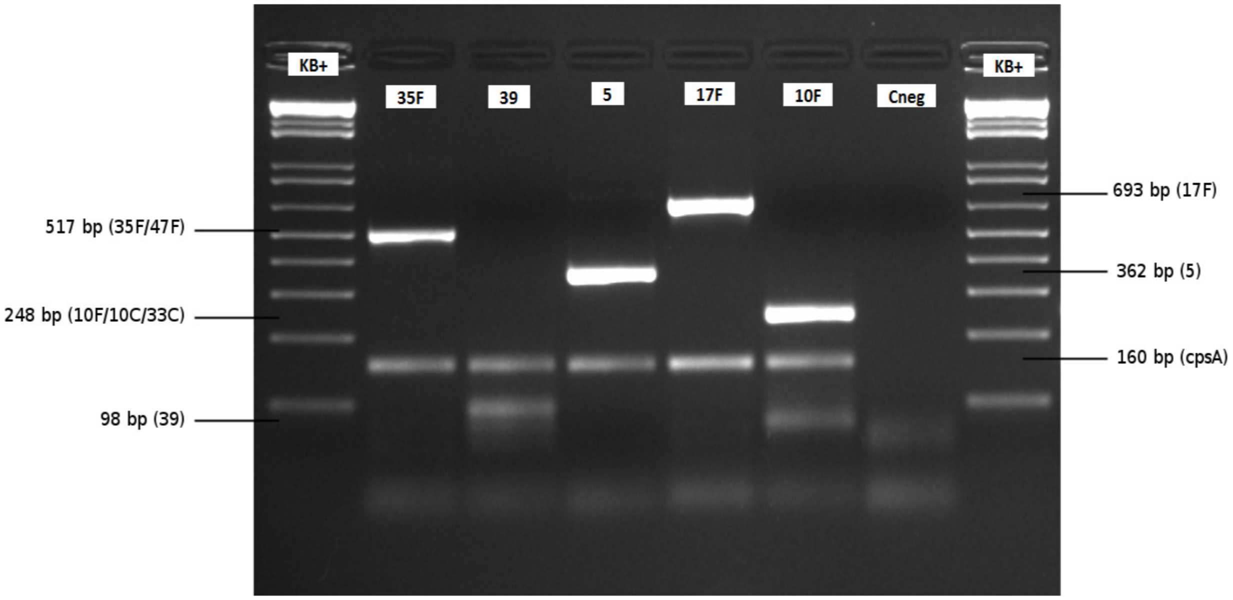


Figure 16: Electrophoresis profile obtained with the multiplex reaction 6 used to detect serotypes 39, 10F/10C/33C, 5, 35F/47F and 17F.

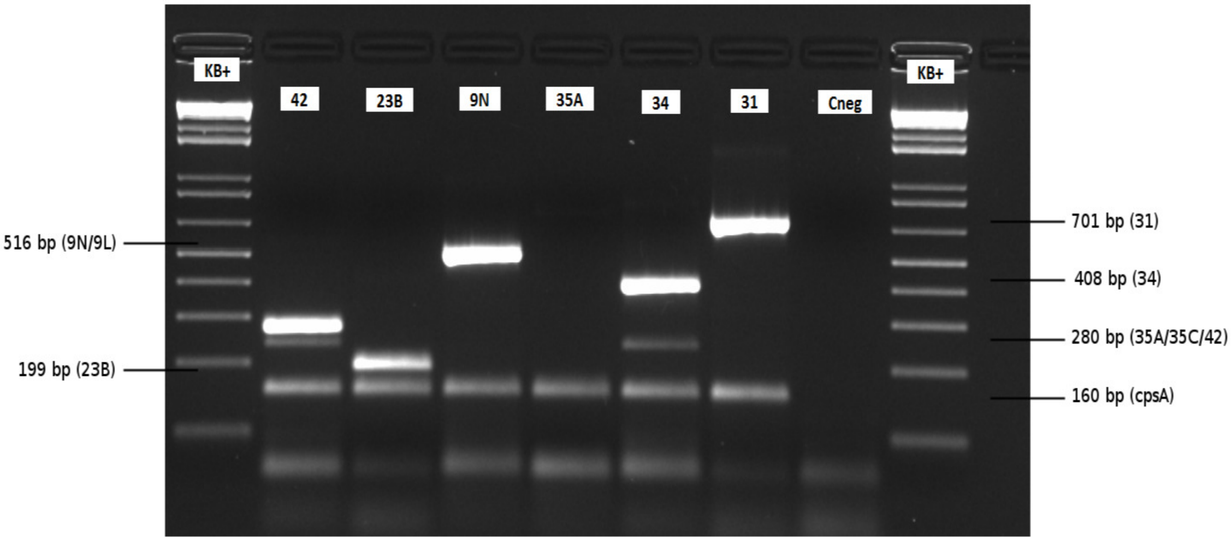


Figure 17A: Electrophoresis profile obtained with the multiplex reaction 7 used to detect serotypes 23B, 35A/35C/42, 34, 9N/9L and 31. Isolates MA092229 (serotype 35A, expected amplicon is 280 bp) and LSPQ 3127 (serotype 34) first genomic extracts.

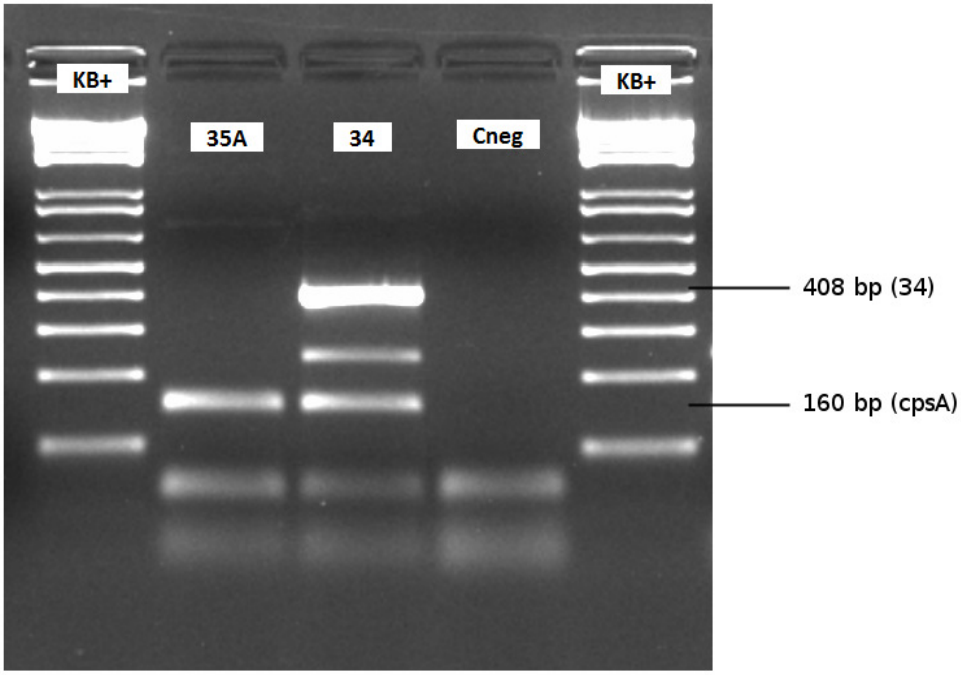


Figure 17B: Electrophoresis profile obtained with the multiplex reaction 7 used to detect serotypes 23B, 35A/35C/42, 34, 9N/9L and 31. Isolates MA092229 (serotype 35A, expected amplicon is 280 bp) and LSPQ 3127 (serotype 34) second genomic extracts.

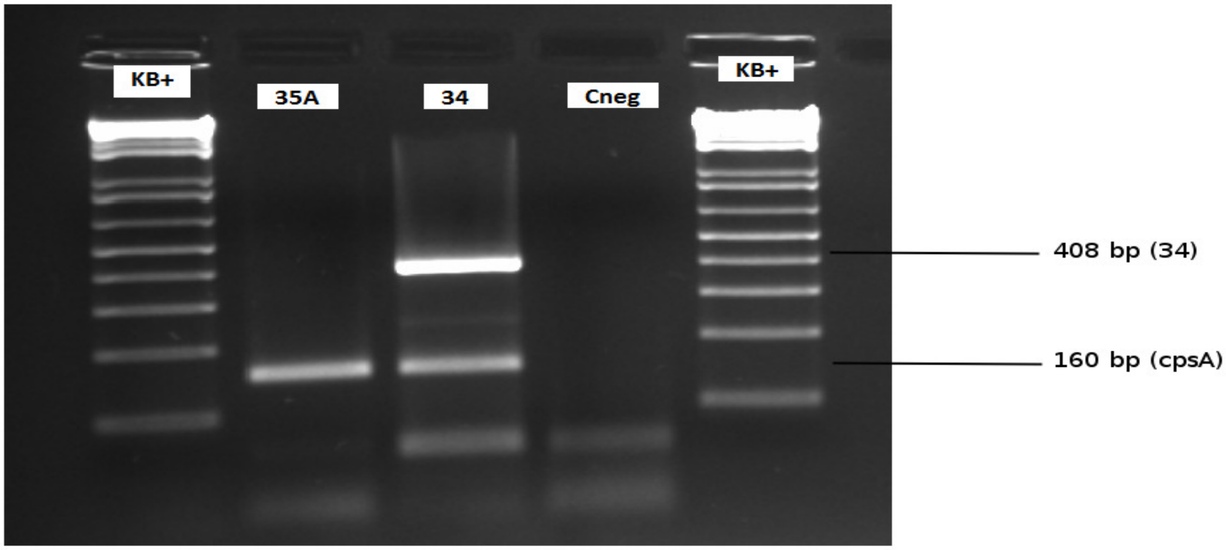


Figure 17C: Electrophoresis profile obtained with the multiplex reaction 7 used to detect serotypes 23B, 35A/35C/42, 34, 9N/9L and 31. KMA082642 (serotype 35A, expected amplicon is 280 bp) and KMA099037 (serotype 34).

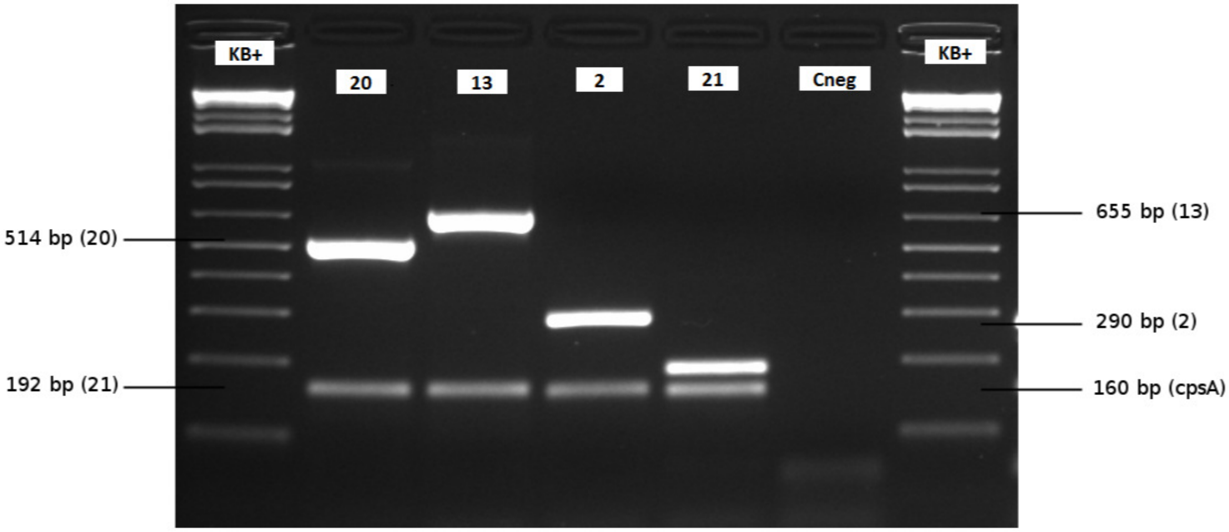


Figure 18: Electrophoresis profile obtained with the multiplex reaction 8 used to detect serotypes 21, 2, 20 and 13.

A total of 60 pneumococcal isolates of known serotypes were tested in order to evaluate the reproducibility of the CDC multiplex PCR protocol. Except for a very few cases (3% misidentified), results are in complete agreement with the expected results (38% to serotype, 35% to serogroup and 23% to subset). It should be noted that bands are generally very well defined and have high intensity when amplicon length is greater than 200 bp. Bands under this marker are often of lighter intensity because of the ethidium bromide which, due to his opposite electric charge compared to DNA, migrates upward.

The only puzzling result we have observed is for serotype 35A which was expected to be detected with reaction 7 (Figure 17A to C). This serotype, targeted by primers 35A/35C/42, was supposed to be revealed with a 280 bp amplicon which we did not observed using two different 35A isolates (MA092229 on Figure 17A and KMA082642 on Figure 17C) and two different genomic extracts of MA092229 (Figure 17A and Figure 17B). A problem with the quality of the 25 µM primers preparation was ruled out based on the positive result obtained with serotype 42 which was also detected with the 35A/35C/42 primers. However, a non-specific band around 250 bp was detected for serotype 42 using the 35A/35C/42 primers (Figure 17A). More extensive studies should be undertaken to explain this issue. Nonetheless, the most probable assumption is that our 35A and 42 isolates are different from those previously tested by the CDC. They are probably sufficiently genetically different compared to the ones tested at the CDC as to be unrecognizable by the PCR primers. A single SNP located in a region needed for an appropriate primer hybridization could perturb the initiation of DNA replication by the polymerase and consequently prevent the amplification of the target DNA leading to an absence of the expected PCR product.

A special attention must be paid to the detection of serotype 38 with PCR reaction 3 (Figure 13). This is the only serotype which is negative for the *cpsA* band. A *cpsA* negative result for serotype 38 and serotype 25F (no isolate with serotype 25F in the current study) is well documented in Carvalho *et al*., (2010).

Finally, another issue was observed with reaction 7 in that a weak non-specific band at around 250 bp occurred for serotype 34 (Figure 17A and 17B). Two different serotype 34 isolates and two different genomic extracts were tested to confirm this issue. The same results were obtained in all cases. According to the expected electrophoresis profile, this band should not appear. However, our isolates of serotype 34 do not necessarily have the same genetic background as those tested by the CDC.

Evaluation of the sequetyping method based on the *cpsB* gene

The last DNA-based serotyping method that we evaluated in the current project is the one developed by Leung *et al*., (2012) named sequetyping. To evaluate the sequetyping method, we have chosen 74 pneumococcus isolates covering a total of 73 different serotypes (two isolates with serotype 29 have been tested). Isolates with serotypes 27, 38, 37, 39 and 43, and one of our two serotypes 29 yielded no amplicon after the PCR amplification step (the extracts of those isolates have tested positive for the presence of genomic DNA). This result is nonetheless in accordance with Leung *et al*., (2012) since those six serotypes were predicted *in silico* to be nonamplifiable. Actually, the Leung *et al*., (2012) method could putatively only amplify 84 among the 92 possible serotypes.

We have successfully sequenced 68 isolates. The average sequence length is 942 bp. The shortest one is 860 bp and was obtained with our serotype 29 isolate. This serotype was not predicted to yield an amplicon and the band intensity on the gel was lower than usual. However, 860 bp is still longer than the 732 bp region used by Leung *et al*., (2012) to test all their serotypes.

In order to verify the concordance with the Quellung expected serotypes of our isolates, we ran, for each *cpsB* sequence, the Blast algorithm on the GenBank NCBI database. All hits list was filtered to follow our serotype identification rule described in Material and Methods. Final results are reported in Table 5. They show that 32 isolates (47%) were correctly sequetyped, 9 (13%) were sequetyped to the serogroup level, 10 (15%) gave ambiguous results and 17 (25%) were misidentified. Misidentified results were obtained for serotypes 9A, 11F, 12A, 12F, 15C, 15F, 16A, 18C, 19B, 19C, 24F, 29, 35A, 41A, 42, 44 and 46. Similar results were obtained by Leung *et al*., (2012) for serotype 12F; one 12F strain was sequetyped as 12B. LSPQ 3064 isolate was identified as serotype 12A with 100% identity (98.9% with a serotype 12F). Regarding 24F, both studies have sequetyped their tested strains as 24B. The Blast identities for the 24F isolate (MA099028) were 99.6% for 24B and only 96.3% for 24F. For 18C, Leung *et al*., (2012) has sequetyped 6 isolates to the serogroup level (18B/18C). Our 18C isolate (MA095139) was categorized as misidentified but shows only one mismatch with the 18B reference sequence. The misidentification of the 35A isolate (MA092229) is also due to one single mismatch with 35B and 35C and was not resolved correctly in the Leung *et al*., (2012) study. Our serotype 29 isolate (KMA099083) is the only one which is very far genetically from the available serotype 29 sequences in GenBank; 83% identity with a serotype 29 and 100% with serotypes 35C and 35B. More serotypes 29 should be evaluated although it is a rare occurrence in Quebec. This misidentification was not observed in the Leung *et al*., (2012) study.

Apart from serotypes 12F, 24F, 18C, 35A, and 29, no equivalent data are available in Leung *et al*., (2012) for the other misidentified serotypes. For serotype, serogroup and ambiguous levels identification, our results are generally the same as the ones obtained by Leung *et al*., (2012). Comparisons, however, are not always possible since 27 of our serotypes are missing in the Leung *et al*., (2012) study. Nonetheless our evaluation of the sequetyping approach has demonstrated that this serotyping method is not always able to correctly identify serotype probably due to small DNA sub region of a large locus including in this analysis.

TABLE 5 Pneumococcal serotype identification using the sequetyping approach

Isolates	*cps* best NCBI hit subject		HSP[1] identities	Expected serotype[2]	Identification level
	GenBank accession	Serotype			
LSPQ3053	CR931632	1	939/939	1	Serotype
	JF911531	19F	931/939		
LSPQ3054	CR931633	2	936/936	2	Ambiguous
	CR931713	41A	936/936		
LSPQ3055	Z47210	3	934/934	3	Serotype
	CR931679	20	919/934		
LSPQ3057	CR931637	5	938/938	5	Serotype
	JF911531	19F	920/938		
LSPQ3058	JF911494	6A	935/935	6A	Serotype
	CR931639	6B	934/935		
LSPQ3064	CR931658	12A	937/937	12F	Misidentified
	CR931660	12F	927/937		
LSPQ3065	CR931679	20	945/945	13	Ambiguous
	CR931661	13	945/945		
LSPQ3066	CR931662	14	944/944	14	Serotype
	JF911531	19F	936/944		
LSPQ3071	CR931675	19A	942/942	19A	Serotype
	CR931684	23B	918/942		
LSPQ3072	CR931679	20	928/929	20	Ambiguous
	CR931661	13	928/929		
LSPQ3080	CR931695	31	944/945	31	Serotype
	CR931713	41A	933/945		
LSPQ3081	CR931697	32F	942/942	32F	Serogroup
	CR931696	32A	942/942		
LSPQ3089	CR931714	41F	942/943	41A	Misidentified
	CR931713	41A	918/944		
LSPQ3092	CR931718	45	948/948	45	Serotype
	CR931699	33B	932/948		
LSPQ3093	CR931658	12A	957/957	46	Misidentified
	CR931719	46	956/956		
LSPQ3095	CR931722	48	949/949	48	Serotype
	CR931679	20	943/949		
LSPQ3124	CR931635	4	940/949	4	Serotype
	AF402095	9V	931/940		
LSPQ3127	CR931703	34	935/936	34	Ambiguous
	CR931669	17A	935/936		

TABLE 5 (continued)

Isolates	*cps* best NCBI hit subject		HSP(1) identities	Expected serotype(2)	Identification level
	GenBank accession	Serotype			
LSPQ3160	CR931680	21	934/935	21	Serotype
	JF911515	6C	918/935		
LSPQ3162	CR931712	40	947/947	40	Ambiguous
	CR931641	7B	947/947		
LSPQ3596	CR931644	8	936/936	8	Serotype
	CR931713	41A	927/936		
LSPQ3641	CR931708	36	958/959	36	Serotype
	CR931667	16A	956/959		
LSPQ3644	CR931659	12B	957/957	44	Misidentified
	CR931717	44	948/957		
LSPQ3677	CR931706	35C	967/967	42	Misidentified
	CR931705	35B	967/967		
	CR931715	42	954/967		
LSPQ3769	CR931683	23A	928/928	23A	Serotype
	JF911531	19F	891/926		
LSPQ3770	CR931639	6B	941/941	6B	Serotype
	JF911494	6A	940/941		
LSPQ4102	CR931643	7F	944/945	7A	Serogroup
	CR931640	7A	944/945		
LSPQ4103	CR931712	40	948/949	7B	Ambiguous
	CR931641	7B	948/949		
LSPQ4162	CR931682	22F	941/941	22F	Serogroup
	CR931681	22A	941/941		
LSPQ4231	CR931642	7C	943/943	7C	Serotype
	CR931677	19C	940/943		
LSPQ4236	CR931668	16F	939/939	16F	Serotype
	JF911531	19F	930/939		
LSPQ4242	JF911515	6C	924/924	6C	Serotype
	JF911503	6B	921/924		
LSPQ4243	CR931671	18A	941/941	18A	Serotype
	CR931632	1	919/941		
MA065427	CR931668	16F	944/945	16A	Misidentified
	CR931667	16A	910/947		
MA066814	CR931672	18B	948/948	18B	Serotype
	CR931673	18C	947/948		
MA073130	CR931655	11C	941/942	11F	Misidentified
	CR931657	11F	870/929		

TABLE 5 (continued)

Isolates	*cps* best NCBI hit subject		HSP(1) identities	Expected serotype(2)	Identification level
	GenBank accession	Serotype			
MA075627	CR931652	10F	937/937	10F	Serogroup
	CR931651	10C	937/937		
	JF911518	19A	901/933		
MA080418	AF402095	9V	950/950	9A	Misidentified
	CR931645	9A	949/950		
MA080812	CR931650	10B	951/951	10B	Serotype
	CR931649	10A	905/951		
MA083042	JF911519	19A	939/939	19B	Misidentified
			No significant hit with 19B		
MA083248	CR931663	15A	943/943	15F	Misidentified
	CR931666	15F	935/943		
MA084138	JF911519	19A	938/938	19C	Misidentified
			No significant hit with 19C		
MA086628	CR931704	35A	939/940	33A	Ambiguous
	CR931702	33F	939/940		
	CR931698	33A	939/940		
	CR931706	35C	938/940		
MA090174	CR931649	10A	940/940	10A	Serotype
	CR931650	10B	894/940		
MA090298	CP002121	11A	947/947	11A	Ambiguous
	CR931674	18F	947/947		
	CR931656	11D	947/947		
	CR931684	23B	924/946		
MA092229	CR931706	35C	937/937	35A	Misidentified
	CR931705	35B	937/937		
	CR931704	35A	936/937		
MA092686	JF911515	6C	939/939	6D	Serogroup
	HM448897	6D	939/939		
	JF911503	6B	936/939		
MA094350	CR931687	24B	947/950	24B	Serotype
	CR931642	7C	932/950		
MA095139	CR931672	18B	949/949	18C	Misidentified
	CR931673	18C	948/949		
MA095690	CR931692	28A	941/941	28A	Serotype
	CR931693	28F	940/941		

TABLE 5 (continued)

	cps best NCBI hit subject				
Isolates	GenBank accession	Serotype	HSP(1) identities	Expected serotype(2)	Identification level
MA095877	CR931682	22F	927/930	22A	Serogroup
	CR931681	22A	927/930		
	CR931648	9V	921/930		
MA096496	CR931688	24F	933/951	15C	Misidentified
	CR931665	15C	918/953		
MA097699	CR931659	12B	939/941	12A	Misidentified
	CR931658	12A	936/941		
MA097723	CR931706	35C	957/958	35B	Serogroup
	CR931705	35B	957/958		
MA097930	CR931655	11C	958/958	11B	Serogroup
	CR931654	11B	958/958		
	CR931684	23B	934/957		
MA098807	CR931670	17F	949/949	17F	Serotype
	CR931700	33C	948/949		
MA098992	JF911522	19F	945/945	19F	Serotype
	HG799504	19A	942/945		
MA099028	CR931687	24B	946/949	24F	Misidentified
	CR931688	24F	916/951		
MA099177	KC688319	15B	949/949	15B	Serotype
	CR931688	24F	934/952		
MA099195	CR931721	47F	955/955	35F	Ambiguous
	CR931707	35F	955/955		
	CR931664	15B	936/955		
MA099234	AF402095	9V	919/922	9V	Serotype
	CR931645	9A	918/922		
MA099238	CR931704	35A	946/947	33F	Ambiguous
	CR931702	33F	946/947		
	CR931698	33A	946/947		
	CR931706	35C	945/947		
MA099389	CR931663	15A	946/946	15A	Serotype
	CR931666	15F	938/946		
MA099461	CR931643	7F	950/950	7F	Serogroup
	CR931640	7A	950/950		
	JF911531	19F	935/950		
MA099463	CR931647	9N	921/921	9N	Serotype
	CR931646	9I	920/921		

TABLE 5 (continued)

Isolates	*cps* best NCBI hit subject GenBank accession	Serotype	HSP[1] identities	Expected serotype[2]	Identification level
MA099467	CR931685	23F	946/946	23F	Serotype
	CR931709	37	931/946		
MA099469	CR931684	23B	936/936	23B	Serotype
	CR931674	18F	914/936		
KMA099083	CR931706	35C	860/860	29	Misidentified
	CR931705	35B	860/860		
	CR931694	29	606/803		

[1] HSP = High-scoring Segment Pairs.

[2] Expected serotype according to Quellung reaction.

Discussion and conclusion

The aim of the current project was to evaluate three DNA-based *Streptococcus pneumoniae* serotyping approaches which could eventually replace the current Quellung gold standard method. One of those, the WGS, is currently not well adapted to a surveillance program. Instead, it would be valuable in the understanding of epidemiological phenomenon such as serotypes replacement and in the comprehension of the molecular mechanism implicated in the capsular polysaccharide synthesis. Moreover, WGS allows the analysis of molecular evolution of the strains, the identification of putative vaccine target in addition to the study of antibiotic resistance and virulence genes.

WGS is costly, time consuming and relatively laborious. This is why this method is unlikely to be used as monitoring tool of invasive *S. pneumoniae* serotypes at this moment. However, bioinformatic pipelines are increasingly automated, costs are decreasing and the technology is more widely available in low-resource settings. A sequencing strategy which exclusively target *cps* locus could be developed. For these reasons, it is likely that WGS will eventually replace conventional typing tools for pneumococci. We have tested, through a next generation sequencing pipeline, our ability to find the expected serotype for 21 isolates representing 10 different serotypes. Results were very convincing in that we were able to extract the entire capsulation locus and identify it correctly for all tested isolates (52% to serotype and 48% to serogroup). We are now looking forward to get a better genetic profile of some isolates in order to better predict their emergence capabilities following the introduction of a new conjugate vaccine.

The sequential multiplex PCR and sequetyping strategy unlike WGS have specifically been developed to improve the serotyping response time and to reduce the associated costs. We have then mainly focused on those two methods in this study. The sequential multiplex approach remain the most cost effective choice (between 30$ and 80$ per strain according to the multiplex design) but unlike the sequetyping method, this method has the inconvenience of requiring an adaptation to the local epidemiology of circulating serotypes. Simply changing the sequential order of the reaction may be sufficient but more often reviewing the combination of primers in the reaction mixture is needed. Unfortunately, this is not always possible.

In the current project, we have demonstrated that the sequential multiplex PCR method is very fast. Resulting electrophoresis patterns are also easy to interpret. Except for serotype 35A, we have successfully reproduced the CDC multiplex scheme. Interestingly, no band was obtained for serotype 35A (reaction 7) using two different samples though we have confirmed the sensitivity of primers 35A/35C/42. to verify whether any single nucleotide polymorphism (SNP) may have prevented an appropriate primer pairing. In addition, reference strains from Statens Serum Institut should be tested.

Another important issue with the multiplex method is the existence of cross reactivity between many serotypes. There are some serotypes (22F/22A, 33F/33A/37, 15A/15F, 6A/6B, 6C/6D, and 7A/7F…) which could not be resolved using this method. However, this disadvantage may be negligible if we take into account only the most frequently occurring serotypes. For example, serotypes 22F and 22A are both detected under the same PCR signal. But statistically, according to epidemiological data for Quebec from 2013 to 2015 (occurrence of 12.7% for serotype 22F compared to 0.2% for serotype 22A), 22F is by far the most probable one. In that case, Quellung should be used to confirm the right serotype. The same rationale should be applied when facing a positive result with 7F/7A PCR in reaction 2; serotype 7A is very uncommon (0.1%) compared to the high frequency (8.6%) of serotype

7F. Conventional serological method is also needed to resolve a positive PCR signal for 6A/6B in reaction 6C/6D. Another alternative would be to perform a pyrosequencing assay (Pai *et al*., 2005). The latter has not been tested and was not part of the current project.

The second DNA-based approach tested, the sequetyping method developed by Leung *et al*., (2012), is very interesting since only one primer pair is needed compared to 41 with the PCR multiplex approach. This method is technically very simple; it is based on the sequencing of a single variable DNA region inside the *cpsB* gene which is unique to *S. pneumoniae*. Furthermore, this method is insensitive to epidemiological data and is quite cheap (~20$). Nonetheless, results are often limited to serogroup identification and sometimes are even ambiguous. Then, we must have to apply some statistical deductions as described before or keep the Quellung reaction as final discriminator. Sequetyping does not always identify at the serotype level nor at the serogroup level as described in Leung *et al.,* (2012). This is because some serotypes may have identical *cpsB* sequences as it is the case with some 6A and 6B strains (Elberse *et al*., 2011). Furthermore, existing intraserotype variation (Varvio *et al*., 2009) in the *cps* regulatory region can lead to identification in the wrong serogroup. This issue has already been observed by Leung *et al*., (2012) with one 19F strain identified as a serotype 1.

A proportion of 47% of our pneumococcal isolates was correctly resolved at the serotype level using the sequetyping approach. However, the identification level rules we used could be biased due to the existence of intra-serotype variation in the *cpsB* gene. For example, an unknown sample for which its *cpsB* region shares 945/945 identities with GenBank 6A serotype and 944/945 compared to a 6B serotype does not necessarily mean that this sample is a 6A serotype. We nonetheless have correctly identified 6 serotypes among the 8 most prevalent (22F (12.7%), 3 (11%), 19A (10.7%), 7F (8.6%), 15A (5.6%), 9N (4.9%), 16F (3.8%) and 23A (3.7%)) in Quebec between 2013 and 2015. Serotypes 22F and 7F have been identified to the serogroup level.

The sequetyping strategy is obviously dependent on a rich sequence database. Currently all Blast queries rely on the collection of *cpsB* sequences deposited in the NCBI GenBank database. Accuracy of the method over time will then be considerably improved with the addition of new sequences coming from different laboratories worldwide. Management of an independent curated *cpsB* database would be highly recommended.

We have demonstrated in this study that at least two molecular techniques, sequential multiplex PCR and sequetyping, are rapid, easy and could potentially gradually replace the traditional serological method. However, data shown that sequetyping is not as reliable as sequential multiplex PCR. Nevertheless, preliminary data show that the Quellung method could still be useful when molecular approaches give inconclusive results. It is important to note that rare untypeable strains, due to their lack of capsular polysaccharide, may generate a positive result with DNA based method. In such cases, the final serotype identification would be in disagreement with the Quellung reaction which would produce a negative result. Conversely, the sequetyping or multiplex PCR approach may rescue the Quellung reaction when the capsular swelling if difficult to observe through microscopic examination.

This completes the first phase of the project dedicated to the monitoring of new molecular tools for the serotyping of *S. pneumoniae* invasive strains. Results obtained from the development phase of the project are summarized in Table 6.

WGS correctly identified serotype of all tested isolates (52% to serotype and 48% to serogroup). With a cheapest and automated pipeline, this method should be kept in mind for serotyping strains from Quebec's surveillance program.

In our study, 23 isolates (38%) were specifically assigned to serotype using sequential multiplex PCR with the results in full accordance with conventional serotyping. Twenty-one other isolates (35%) were assigned to the right serogroup and 14 isolates (23%) to the correct subset. Only few isolates (n=2) could not be correctly associated to serotype, serogroup or subset (3%).

Using sequetyping method, 32 isolates (47%) were specifically assigned to serotype; expected results according to gold standard method. Other 9 isolates (13%) were assigned to the right serogroup. However, 10 isolates (15%) gave ambiguous results and 17 isolates (25%) were misidentified.

In the second phase of this project, efforts will be directed towards the proof-concept. Many additional strains will be tested by using the three DNA-based methods, WGS, sequential multiplex PCR and sequetyping. Here also, results will be compared to the Quellung gold standard method. Execution time, time delivery and cost will also be compiled and assessed in order to guide our final choice for the most efficient serotyping method to use in our surveillance program at the LSPQ.

Here are the next steps to be performed during Part 2 of the study (proof of concept):

- Specificity testing (serotype) for multiplex PCR.
- Specificity testing (*Streptococcus* species other than *S. pneumoniae*) for multiplex PCR and sequetyping.
- Testing of serotypes not previously available at LSPQ or problematic (9L, 10C, 11C, 11D, 12B, 17A, 18F, 24A, 25A, 25F, 28F, 29, 32A, 33B, 33C, 33D, 35A, 35C, 41F, 47A, 47F) by WGS, multiplex PCR and sequetyping.
- Strains received at LSPQ for provincial surveillance will be analyzed using WGS, sequential multiplex PCR and sequetyping methods.

Acknowledgements

This work was supported by Grant IIR number WI197603 from Pfizer Canada.

We thank Jean-Charles Côté, Ph.D., microbiologist, LSPQ, for proofreading of the report and Christine Martineau, Ph.D., microbiologist, LSPQ, for WGS assistance.

TABLE 6 Summary of the molecular methods used for *S. pneumoniae* serotyping.

			Serotyping results (concordance with Quellung)				
Methods	**Advantages**	**Disadvantages**	**Serotype**	**Serogroup**	**Subset** [1]	**Ambiguous**	**Mis-identified**
WGS (n=21 strains tested from 10 different serotypes)	- Includes all serotypes - Additional information obtained at the same time (multi-locus sequence type, antimicrobial resistance…) are useful for other studies - Identification of putative vaccine target and serotyping evolution analysis	- Laborious - Expensive (~200$/strain) - A lot of data to manage - Needs bioinformatics setup - Time consuming	52%	48%	N/A	0%	0%
Sequential multiplex PCR (n= 60 strains tested from 58 different serotypes)	- Cost-effective (30 $ - 80$/strain) - Method easily achievable - Serotype easily determined - Straightforward	- Occasional issues such as false negative, non-specific band and small amplicon - To be customized according to local epidemiology - Detection of known serotypes - Not useful for all serotypes - Possibility of cross-reactions	38%	35%	23%	N/A	3%
Sequetyping (n=68 strains tested from 68 different serotypes) [2]	- Rapid - Easy to set up - Inexpensive (~20$/strain)	- Not useful for all serotypes - False assignment of serotype due to potential for gene exchange - Method based on public databases - Necessity of a *cpsB* curated bank	47%	13%	N/A	15%	25%

[1] Defined as correct results obtained with PCR multiplex primers detecting a subset, for example 33F/33A/37 (reaction 2).

[2] 74 isolates selected from 73 different serotypes; 68 isolates successfully sequenced from 68 different serotypes.

References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19(5)**:455-77.

Boisvert, S., Laviolette, F., Corbeil, J. 2010./ Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* **17(11)**:1519-33.

Bratcher, P.E., Park, I.H., Oliver, M.B., Hortal, M., Camilli, R., Hollingshead, S.K., Camou, T., Nahm, M.H. 2011. Evolution of the capsular gene locus of *Streptococcus pneumoniae* serogroup 6. *Microbiology*. **157(Pt 1)**:189-98.

Camargo, D.R., Pais, F.S., Volpini, Â.C., Oliveira, M.A., Coimbra, R.S. 2015. Revisiting molecular serotyping of *Streptococcus pneumoniae*. *BMC Genomics*. **16 Suppl 5**:S1.

Carvalho da Gloria, M., Pimenta, F.C., Jackson, D., Roundtree, A., Ahmad, Y., Millar, E.V., O'Brien, K.L., Whitney, C.G., Cohen, A.L., Beall, B.W. 2010. Revisiting pneumococcal carriage by use of broth enrichment and PCR techniques for enhanced detection of carriage and serotypes. *J. Clin. Microbiol*. **48(5):**1611-8.

Elberse, K., Witteveen, S., van der Heide, H., van de Pol, I., Schot, C., van der Ende, A., Berbers, G., Schouls, L. 2011. Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. *PLoS One.* **6(9)**:e25018.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. **29(8):**1072-5.

Leung, M.H., Bryson, K., Freystatter, K., Pichon, B., Edwards, G., Charalambous, B.M., Gillespie, S.H. 2012. Sequetyping: serotyping *Streptococcus pneumoniae* by a single PCR sequencing strategy. *J. Clin. Microbiol.* **50(7)**: 2419-27.

Pai, R., Gertz, R.E., Beall, B. 2006. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. *J. Clin. Microbiol.* **44(1)**:124-31.

Pai, R., Limor, J., Beall, B. 2005. Use of pyrosequencing to differentiate *Streptococcus pneumoniae* serotypes 6A and 6B. *J. Clin. Microbiol.* **43(9)**:4820-2.

Pevzner, P.A., Tang, H., Waterman, M.S. 2001. An Eulerian path approach to DNA fragment assembly. *Proc. Natl. Acad. Sci. U. S. A.* **98(17)**:9748-53.

Salter, S.J., Hinds, J., Gould, K.A., Lambertsen, L., Hanage, W.P., Antonio, M., Turner, P., Hermans, P.W., Bootsma, H.J., O'Brien, K.L., Bentley, S.D. 2012. Variation at the capsule locus, cps, of mistyped and non-typable *Streptococcus pneumoniae* isolates. *Microbiology*. **158(Pt 6)**:1560-9.

Varvio, S.L., Auranen, K., Arjas, E., Mäkelä, P.H. 2009 Evolution of the capsular regulatory genes in *Streptococcus pneumoniae*. *J. Infect. Dis.* **200(7)**:1144-51.

Zerbino, D.R. Using the Velvet de novo assembler for short-read sequencing technologies. 2010 *Curr Protoc Bioinformatics*. **Chapter 11**:Unit 11.5

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist	[redacted]	IIR Grant Specialist PHONE	[redacted]
IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist FAX	[redacted]

PLEASE COMPLETE AND RETURN BY: November 18, 2015

Per contractual requirements, we are requesting a status update on your IIR study supported by Pfizer via funding and/or drug. Please answer the following questions regarding the above referenced study by the due date. Answers from your last submitted update have been incorporated below; please update as needed and answer the remaining questions.

GENERAL INFORMATION

Pfizer Tracking #	**WI197603**	**Institutional Protocol #**	
Principal Investigator	**Dr. Brigitte Lefebvre**		
Study Title	**Molecular tools for serotyping for Streptococcus pneumoniae invasive strains surveillance in the province of Quebec.**		

STUDY UPDATE INFORMATION

Has this study been initiated?	☐ NO ☒ YES	Date of initiation	mm/dd/yyyy
Has the protocol been amended since last update?	☒ NO ☐ YES *(If YES, please provide the revised protocol)*		
Current IRB/IEC approval/renewal expires on last IRB date	This is not current, please forward the most recent letter		
Have there been any personnel changes? *(If YES, please provide name and full contact info on Page 3)*			☒ NO ☐ YES
Target protocol enrollment		Date of first subject enrolled	mm/dd/yyyy
Last reported enrollment		Actual enrollment to date *(this should not include screen failures)*	
Targeted last subject last visit		Actual last subject last visit	
Do you have current drug supply sufficient to complete the study? *(If NO, please complete the Drug Section on Page 3)*			☐ NO ☐ YES
Is this protocol closed to enrollment? *(patients may still be receiving therapy)*			☐ NO ☐ YES
Targeted study completion date *(primary objectives met; patient therapy and final study analysis complete)*			mm/dd/yyyy
Actual study completion date *(if applicable)*			mm/dd/yyyy
Targeted date to provide results to Pfizer			mm/dd/yyyy

PUBLICATION INFORMATION

Do you plan to publish? *(If YES, please complete the information below.)*	☐ NO ☒ YES

Please be aware that, according to the IIR agreement, the investigator is required to provide Pfizer with an opportunity to prospectively review any proposed publication, abstract or other type of disclosure that reports the results of the study.

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist	[redacted]	IIR Grant Specialist PHONE	[redacted]
IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist FAX	[redacted]

FORMAT	**PUBLICATION** *(please include anticipated journal or audience)*	**PLANNED**	**ACTUAL**	**SUBMISSION DATE** **mm/dd/yyyy**
Abstract	CACMID or ISPPD	☒	□	2016
Manuscript	Journal of clinical microbiology	☒	□	At the end of study
Poster		□	□	
Other	**Study report** Title: **Molecular tools for serotyping for *Streptococcus pneumoniae* invasive strains surveillance in the province of Quebec (study report - part 1).**	□	☒	11/12/2015

SIGNATURE

[redacted]

NAME Brigitte Lefebvre

DATE 11/17/2015
mm/dd/yyyy

SIGNATURE (ONLY if faxed)

Pfizer	**STUDY STATUS UPDATE FORM: CLINICAL**	

IIR Grant Specialist	[redacted]	IIR Grant Specialist PHONE	[redacted]
IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist FAX	[redacted]

DRUG SUPPLY INFORMATION

SUPPLY CURRENTLY ON SITE	**ACTIVE**	**PLACEBO**
ESTIMATED REMAINDER REQUIRED TO COMPLETE STUDY	**ACTIVE**	**PLACEBO**
CAN PHARMACY ACCOMODATE TOTAL REMAINDER?	☐ **YES**	☐ **NO**

PERSONNEL INFORMATION

	PRINCIPAL INVESTIGATOR	**COORDINATOR**
NAME		
INSTITUTION		
MAILING ADDRESS		
TELEPHONE		
FAX		
EMAIL		

	PHARMACIST	**OTHER** *(specify in additional comments)*
NAME		
INSTITUTION		
MAILING ADDRESS		
TELEPHONE		
FAX		
EMAIL		

ADDITIONAL COMMENTS

Entente relative à un essai proposé par un investigateur

Pfizer Canada inc., société dûment constituée en vertu des lois du Canada dont l'établissement principal est situé au 17300, autoroute Transcanadienne, Kirkland (Québec) H9J 2M5 (« **Pfizer** »),

ET :

Institut national de santé publique du Québec, personne morale légalement constituée par la *Loi sur l'Institut national de santé publique du Québec* (RLRQ, chapitre I-13.1.1), administrant le Laboratoire de santé publique du Québec (LSPQ), situé au 20045, chemin Sainte-Marie, Sainte-Anne-de-Bellevue (Québec) H9X 3R5 (« **l'Établissement** »),

ET :

Brigitte Lefebvre, Ph. D., exerçant ses activités au Laboratoire de santé publique du Québec (LSPQ) 20045, chemin Sainte-Marie, Sainte-Anne-de-Bellevue (Québec) H9X 3R5 (le « **promoteur-investigateur** »)

(Pfizer, l'Établissement et le promoteur-investigateur, individuellement désignés en tant que « **partie** » et, collectivement, « **les parties** ».)

L'entente est en vigueur à compter du 21 octobre 2015 (« Date d'entrée en vigueur ").

Le promoteur-investigateur a conçu et souhaite mener une étude observationnelle prospective intitulée « **~~Serotyping for~~** Serotype monitoring of **Streptococcus pneumoniae invasive strains in adult population ~~surveillance~~ in the province of Quebec: a 3 years study evaluation** under the Protocol WI203144 », lequel est joint comme **annexe C** (le « **protocole** ») et (« **l'étude** »). Pfizer souhaite fournir certains services de soutien pour l'étude.

BL
2015/11/16

ATTENDU que le promoteur-investigateur souhaite recevoir un certain soutien de la part de Pfizer afin de mener à bien l'étude à l'Établissement;

ATTENDU que Pfizer a accepté de fournir un tel soutien au promoteur-investigateur selon les modalités énoncées dans la présente entente, étant entendu que Pfizer n'est pas le promoteur de l'étude et qu'elle ne doit en aucun cas être considérée ou représentée comme tel.

EN CONSÉQUENCE, les parties conviennent de ce qui suit :

1. <u>Investigateurs et personnel de recherche</u>

1.1 <u>Promoteur-investigateur</u>. L'étude sera menée par le promoteur-investigateur.

1.2 <u>Obligations</u>. L'Établissement est responsable envers Pfizer d'assurer la conformité, par tous les membres du personnel qui participent à l'étude, y compris promoteur-investigateur et tous les entrepreneurs ou consultants, aux modalités de la présente entente.

2. <u>Protocole</u>

2.1 <u>Protocole</u>. L'étude sera menée conformément à un protocole développé par le promoteur-investigateur (« protocole »). L'approbation du protocole final par Pfizer est une condition au soutien de Pfizer en vertu de la présente entente.

2.2 <u>Modifications</u>. Si le promoteur-investigateur modifie le protocole définitif approuvé par Pfizer, il doit en informer Pfizer rapidement par écrit. La continuation du soutien accordé par Pfizer sera conditionnelle à l'examen et à l'acceptation par Pfizer des modifications apportées au protocole.

3. <u>Réalisation de l'étude</u>

3.1 <u>Parrainage</u>. Le promoteur-investigateur et l'Établissement reconnaissent que Pfizer n'est en aucune façon le « promoteur » et s'engagent à ne représenter en aucun temps Pfizer comme tel ou autrement auprès de tout tiers. Sauf disposition expresse dans la présente entente, Pfizer n'a aucune obligation ni responsabilité envers l'Établissement, le promoteur-investigateur ou tout particulier qui participe à l'exécution de l'étude.

3.2 <u>Réglementation</u>. L'Établissement est seul responsable de tous les rapports de sécurité et de toutes les obligations réglementaires associés à la réalisation de l'étude.

3.3 <u>Normes</u>. Le promoteur-investigateur procédera à l'étude conformément au protocole, à la Conférence internationale sur l'harmonisation des bonnes pratiques cliniques (ICH GCP) – dans la mesure applicable à ce type d'étude –, et à toutes les lois applicables. promoteur-investigateur doit se conformer à toutes les exigences du comité d'éthique de l'Établissement (CEE) ou d'un comité d'éthique indépendant (CEI) relativement aux études impliquant l'utilisation d'échantillons biologiques humains.

3.4 Approbation du CEE ou d'un CEI. Dans le cas où ce type d'étude l'exigerait, le promoteur-investigateur s'assurera que l'étude est approuvée par le CEE ou un CEI approprié et soumise à une surveillance continue par un tel CEE ou CEI. Si l'approbation du CEE ou d'un CEI est nécessaire, l'Établissement doit, comme condition au soutien de Pfizer, fournir à Pfizer la preuve documentaire de l'approbation initiale du protocole définitif par le CEE ou le CEI ainsi que des renouvellements annuels de cette approbation si de tels renouvellements sont nécessaires (voir l'annexe B, Exigences en matière de documentation). L'Établissement avisera Pfizer promptement de toute suspension ou de tout retrait de l'approbation par le CEE ou le CEI pendant la durée de la présente entente.

3.5 Échantillons biologiques. Cette étude en laboratoire impliquera l'utilisation d'échantillons biologiques fournis par l'Établissement.

a. Consentement. Le promoteur-investigateur doit a) obtenir le consentement éclairé des personnes desquelles les échantillons biologiques ont été obtenus (« donneurs d'échantillons ») conformément à la loi applicable, b) s'assurer que le consentement éclairé couvrant la recherche qui sera effectuée a déjà été obtenu ou c) obtenir, auprès du CEE ou d'un CEI approprié, une dérogation au consentement éclairé pour l'utilisation des échantillons biologiques dans le cadre de l'étude. Le promoteur-investigateur doit également assurer la conformité aux lois applicables à l'égard de l'utilisation et de la divulgation de renseignements sur la santé relativement aux donneurs d'échantillons. Si un consentement éclairé est utilisé, le promoteur-investigateur doit informer les donneurs d'échantillons que Pfizer fournit un soutien à l'étude. Pfizer n'a aucune obligation de participer à la rédaction d'un document relatif au consentement éclairé ou à une demande de dérogation, ni d'examiner ou de commenter un tel document.

b. Propriété et disposition. Pfizer ne revendique aucun droit de propriété à l'égard des échantillons biologiques fournis pour l'étude par l'Établissement. L'Établissement est responsable de la disposition adéquate de tous les échantillons biologiques restants à la fin de l'étude.

3.6 Aucune surveillance ou collecte de données. Pfizer ne surveillera pas l'étude ni ne recevra de données de l'étude (selon la définition de l'article 5, Données de l'étude et résultats de l'étude).

3.7 Durée de la réalisation de l'étude. Le promoteur-investigateur prévoit achever l'étude (achèvement des procédures et de la portion collecte de données de l'étude) d'ici le 31 décembre 2018.

3.8 Rapports de statut. Le promoteur-investigateur doit fournir à Pfizer un rapport sur le statut de l'étude, dans le format demandé par Pfizer, au moins une fois par année pendant la durée de la présente entente, ou plus fréquemment si cela est indiqué à l'annexe A (Calendrier des paiements) ou si les deux parties en conviennent mutuellement. Chaque rapport de statut doit présenter les progrès de l'étude, les plans de publication, tout ajustement à la date d'achèvement estimative de l'étude, ainsi que toute autre information raisonnablement demandée par Pfizer.

4. Soutien à un essai proposé par un investigateur. Pfizer fournira un soutien financier à l'étude d'un montant de **sept cent sept mille quatre-vingts dollars (707 080,00 $ CA)**, conformément au calendrier présenté à l'annexe A, Calendrier des paiements. Ce financement constitue le soutien aux essais proposés par des investigateurs pour cette étude.

4.1 Base du soutien. Ce soutien accordé à un essai proposé par un investigateur n'est conditionnel à aucune relation d'affaires préexistante ou future entre Pfizer et le promoteur-investigateur ou l'Établissement. Par ailleurs, il n'est conditionnel à aucune décision d'entreprise ou autre décision que le promoteur-investigateur ou l'Établissement aurait prise ou pourrait prendre relativement à Pfizer ou aux produits de Pfizer.

4.2 Présentation des documents requis. Pfizer ne fournira aucun élément de soutien à l'essai proposé par un investigateur tant qu'elle n'aura pas reçu les documents nécessaires indiqués à l'annexe B, Exigences en matière de documentation.

4.3 Utilisation du soutien à un essai proposé par un investigateur. Le promoteur-investigateur et l'Établissement utiliseront le soutien à l'essai proposé par un investigateur uniquement aux fins de l'étude. À la fin de l'étude, promoteur-investigateur doit confirmer par écrit que le soutien à l'essai proposé par un investigateur a été utilisé uniquement pour soutenir l'étude, en remplissant le formulaire *Certificat d'achèvement de l'étude* fourni par Pfizer.

4.4 Budget d'étude. L'Établissement déclare que le budget d'étude qu'il a fourni et sur lequel est fondé le soutien à l'essai proposé par un investigateur reflète une estimation étayée de tous les fonds requis pour réaliser l'étude et faire rapport sur celle-ci, y compris les dépenses relatives à la publication des résultats de l'étude.

4.5 Taxes.

(a) Les montants payables en vertu de la présente entente ne comprennent pas la taxe fédérale sur les produits et services ou la taxe de vente harmonisée (« TPS/TVH »), la taxe de vente du Québec (« TVQ ») ou autres taxes semblables à la valeur ajoutée, à la consommation, de vente ou d'utilisation (collectivement, les « taxes »).

(b) Il incombe à l'Établissement et au promoteur-investigateur de surveiller et d'examiner leur besoin, le cas échéant, de s'inscrire aux fins des taxes et de facturer, percevoir et remettre les taxes applicables.

4.5 Divulgation par Pfizer. Dans l'intérêt de la transparence en ce qui concerne ses relations financières avec les investigateurs et les sites d'étude ou pour assurer la conformité aux lois locales applicables, Pfizer peut divulguer publiquement le soutien qu'elle fournit en vertu de cette entente. Une telle divulgation par Pfizer peut identifier l'Établissement et promoteur-investigateur, mais doit différencier clairement les paiements ou autres transferts de valeur faits à des établissements de ceux faits à des particuliers.

5. Données de l'étude et résultats de l'étude. Aux fins de la présente entente, « données de l'étude » signifie les données brutes, non cumulatives, recueillies au cours de l'étude. « Résultats de l'étude » désigne les données cumulatives ou résumées de l'étude et les conclusions de celle-ci, qui seraient incluses dans un rapport d'étude ou une publication sur le sujet. Le promoteur-investigateur est libre de publier les résultats de l'étude, sous réserve des dispositions de l'article 8 (Publications) et le promoteur-investigateur ainsi que l'Établissement sont libres d'utiliser les résultats de l'étude à toute autre fin. L'Établissement possède les données de l'étude et est libre de les utiliser pour ses propres objectifs et programmes de recherche, de formation et de soins aux patients. Toutefois, compte tenu du soutien à l'essai proposé par un investigateur accordé par Pfizer, le promoteur-investigateur et l'Établissement ne doivent pas utiliser ni permettre à quiconque d'utiliser les données de l'étude pour l'avantage commercial de toute tierce partie.

6. Rapport d'étude. Dans les six mois suivant la fin de l'étude ou la résiliation de la présente entente, selon la première éventualité, le promoteur-investigateur fournira à Pfizer un rapport écrit sur les résultats de l'étude (« rapport d'étude »). Sauf disposition contraire convenue par écrit par les parties, le rapport d'étude peut être sous la forme d'un manuscrit pour publication (voir l'article 8, Publications). Si l'entente est résiliée avant l'échéance prévue, le rapport d'étude doit inclure, au minimum, les résultats de l'étude jusqu'à la date de résiliation.

7. Confidentialité des données. Dans le cadre de l'exécution de l'étude, le promoteur-investigateur et l'Établissement ne doivent traiter des renseignements qui concernent un particulier identifiable ou qui permettent d'identifier un particulier (« renseignements personnels ») que dans le but de réaliser l'étude et pour aucune autre fin. Le promoteur-investigateur et l'Établissement doivent prendre toutes les mesures techniques, physiques et organisationnelles

appropriées et nécessaires visant à prévenir le traitement ou l'accès non autorisés ou illicites à ces renseignements personnels, ainsi que la perte, la destruction ou la détérioration de ces renseignements. Plus particulièrement, et sans limiter ce qui précède, le promoteur-investigateur et l'Établissement doivent se conformer à toutes les lois et à tous les règlements applicables qui sont en vigueur à la date de la présente entente ou qui entrent en vigueur pendant qu'elle est en vigueur, concernant la protection des renseignements personnels et/ou la protection du droit à la vie privée des personnes (« lois sur la protection de la vie privée »).

À l'échéance ou à la résiliation de la présente entente, et par la suite, le promoteur-investigateur et l'Établissement traiteront les renseignements personnels directement ou indirectement liés à l'étude conformément à toutes les lois applicables sur la protection de la vie privée.

Dans le présent article 7, « traiter » englobe le fait de recueillir, de conserver, d'utiliser, de modifier, de divulguer, de céder ou de transférer les données.

8. Publications. Pfizer soutient l'exercice de la liberté universitaire et encourage l'Établissement à publier les résultats de l'étude, qu'ils soient ou non favorables à Pfizer ou à tout produit de Pfizer. Tel qu'il est utilisé dans la présente entente, le terme « publication » comprend les articles de revue, les résumés, les présentations ou autres modes de divulgation publique qui font rapport des résultats de l'étude.

8.1 Examen préalable à la publication. Le promoteur-investigateur ou d'autres auteurs appropriés de l'Établissement (« auteurs ») fourniront à Pfizer une occasion (au moins 60 jours avant la présentation ou tout autre mode de divulgation publique) d'examiner de manière prospective toute publication proposée. Pfizer fera cet examen afin de déceler toute invention apparentée non protégée (voir l'article 9, Inventions) et pourra fournir des observations sur le contenu. Les auteurs peuvent tenir compte de ces observations de bonne foi, mais n'ont aucune obligation d'intégrer toute suggestion faite par Pfizer.

8.2 Normes. Pour toutes les publications, les auteurs pourront se conformer aux lignes directrices relatives à la paternité dans les *Recommandations pour la conduite, la présentation, la rédaction et la publication des travaux de recherche soumis à des revues médicales* http://www.icmje.org/recommendations/translations/french2014.pdf) établies par l'International Committee of Medical Journal Editors.

8.3 Divulgation du soutien. Les auteurs doivent divulguer dans toute publication le soutien à l'étude accordé par Pfizer.

8.4 Publication de résumés. Le promoteur-investigateur et l'Établissement reconnaissent par les présentes que Pfizer se réserve le droit d'utiliser, de reproduire, de publier, de rééditer et de compiler tout résumé lié à l'étude et aux données de l'étude, ou à une partie de celles-ci, tout comme Pfizer, à sa seule discrétion, peut en décider, à condition que Pfizer ait obtenu l'autorisation pertinente auprès de l'éditeur concerné, le cas échéant.

9. Inventions. Les droits à toute invention ou découverte, brevetable ou non, résultant de la réalisation de l'étude (« invention ») seront déterminés conformément à la présente disposition.

9.1 Propriété. Toute invention réalisée uniquement par un ou plusieurs employés ou entrepreneurs (collectivement, le « personnel ») de l'Établissement sera détenue exclusivement par l'Établissement. Toute invention réalisée uniquement par le personnel de Pfizer sera détenue exclusivement par Pfizer. Les inventions réalisées conjointement par le personnel de l'Établissement et le personnel de Pfizer seront détenues conjointement par l'Établissement et Pfizer. L'Établissement et Pfizer conserveront chacun leur droit d'exercer et d'exploiter leur participation indivise dans toute invention détenue conjointement sans devoir obtenir d'autorisation et sans devoir rendre des comptes à leur cotitulaire.

9.2 Inventions liées à des produits. « Invention liée à un produit » désigne toute invention (selon la définition à l'article 9 ci-dessus) qui englobe le traitement par un produit de Pfizer ou l'administration, la fabrication, la forme, la formulation ou l'utilisation d'un produit de Pfizer (y compris l'utilisation en combinaison avec d'autres produits ou agents), ou qui constitue un biomarqueur utile dans la sélection des patients pour le traitement par le produit de Pfizer ou est reliée à un tel biomarqueur.

9.3 Licence non exclusive accordée à Pfizer. L'Établissement accorde à Pfizer une licence intégralement payée, perpétuelle, internationale, non exclusive et libre de redevances à l'égard de toutes fins relatives à chaque invention liée à un produit détenue par l'Établissement. Une telle licence non exclusive comprendra le droit de 1) concéder une sous-licence aux sociétés affiliées (voir la définition au paragraphe 11.3, Société affiliée), aux entrepreneurs ou aux collaborateurs de Pfizer travaillant au profit de Pfizer ou en lien avec une collaboration de produit ou de service de Pfizer ou d'une société affiliée de Pfizer et 2) concéder une sous-licence ou attribuer à un ayant droit une partie ou la totalité des droits détenus sur un produit de Pfizer auquel l'invention liée est pertinente.

9.4 Option de licence exclusive. L'Établissement accorde par ailleurs à Pfizer une option lui permettant d'obtenir une licence mondiale exclusive pour toutes fins, assortie des pleins droits de sous-licence et d'attribution, pour chaque invention liée à un produit détenue en totalité ou en partie par l'Établissement, selon des modalités à négocier de bonne foi entre les parties.

10. Résiliation

10.1 Événements entraînant la résiliation. La résiliation de cette entente sera déclenchée par la première des éventualités suivantes.

a. Achèvement des obligations aux termes de l'entente. L'entente prendra fin lorsque l'étude est terminée, ce qui signifie à l'achèvement de toutes les activités prescrites par le protocole (« l'achèvement de l'étude ») et lorsque les parties ont reçu tous les produits livrables et les paiements dus.

b. Résiliation hâtive par l'Établissement. Si l'Établissement met fin à l'étude avant l'échéance prévue, pour quelque raison, l'Établissement peut résilier l'entente moyennant un préavis à Pfizer.

c. Résiliation hâtive par Pfizer. Pfizer peut résilier l'entente avant l'échéance prévue dans l'une ou l'autre des circonstances suivantes :

1) Le protocole est modifié d'une façon inacceptable pour Pfizer (voir le paragraphe 2.2, Modifications).
2) La réalisation de l'étude n'est pas achevée dans les six mois suivant la date cible (voir le paragraphe 3.7, Durée de la réalisation de l'étude).
3) L'étude ne démarre pas dans les six mois suivant la date d'entrée en vigueur de la présente entente.
4) Les progrès de l'étude sont considérablement plus lents que ce qui est décrit dans le protocole ou la proposition, ou que ce qui est nécessaire pour achever l'étude pour la date cible.
5) La conception ou les objectifs de l'étude ne sont plus pertinents du point de vue scientifique.
6) L'Établissement ou le promoteur-investigateur ne s'est pas conformé aux lois locales ou aux dispositions de l'article 12 (Lutte contre la corruption) de la présente entente, y compris concernant les circonstances où Pfizer est informée 1) que des paiements irréguliers sont faits ou ont été faits à des représentants de l'État (selon la définition à l'annexe D) ou à toute autre personne par l'Établissement, le promoteur-investigateur ou toute personne qui agit au nom de l'Établissement ou du promoteur-investigateur relativement à l'étude ou à cette entente ou 2) que l'Établissement, le promoteur-investigateur ou toute personne qui agit au nom de l'Établissement ou du promoteur-investigateur relativement à l'étude ou à cette entente a accepté un paiement, un article ou un avantage, quelle qu'en soit la valeur, comme

incitation indue à attribuer, obtenir ou conserver un contrat ou pour obtenir ou accorder autrement un avantage commercial indu de la part ou à l'intention de toute autre personne ou entité.

d. Résiliation motivée. L'une ou l'autre des parties peut résilier l'entente immédiatement par l'envoi d'un avis de résiliation motivée, y compris, mais sans s'y limiter, à l'égard de toute violation substantielle sans remédiation des modalités de cette entente par l'autre partie. Un autre motif valable aux termes de cette disposition pourrait être le défaut, par l'Établissement, de respecter ou une intention démontrée qu'il aurait de ne pas respecter les garanties énoncées à l'article 12 (Lutte contre la corruption).

10.2 Date d'entrée en vigueur de la résiliation. Si la résiliation est déclenchée par un des événements décrits aux alinéas 10.1.b ou c ci-dessus, la résiliation sera effective après l'achèvement par les deux parties de toute obligation restante applicable qui est stipulée dans l'entente.

10.3 Paiement en cas de résiliation hâtive. Les modalités du présent paragraphe 10.3, Paiement en cas de résiliation hâtive, s'appliquent uniquement si l'entente est résiliée avant la date d'échéance prévue pour une raison autre que pour un motif valable (voir l'alinéa 10.1.d, Résiliation motivée). En cas de résiliation hâtive, Pfizer paiera une portion calculée au prorata du montant de financement total accordé pour l'essai proposé par un investigateur, moins les paiements déjà effectués. L'Établissement remboursera à Pfizer tout financement déjà reçu en sus de ce montant calculé, sauf dans la mesure où ces fonds ont déjà été utilisés ou affectés sans possibilité d'annulation, d'une manière compatible avec le budget d'étude sur laquelle le soutien à l'essai proposé par un investigateur est fondé ou selon tout mode approuvé de manière prospective par Pfizer.

10.4 Rapprochement des comptes à la fin de l'étude. À la fin de l'étude, les parties coopéreront pour effectuer un rapprochement financier afin de confirmer la concordance entre le total des paiements d'étape par Pfizer et les étapes et les produits livrables qui étaient convenus. Les parties conviennent de procéder à un rajustement (soit un remboursement ou un paiement supplémentaire) si cette analyse révèle que cela est justifié.

11. Autres dispositions

11.1 Indemnisation. L'étude n'est pas conçue, parrainée ou gérée par Pfizer et Pfizer ne fournit aucune indemnisation de quelque nature que ce soit.

11.2 Pertinence. L'Établissement atteste que lui-même et le promoteur-investigateur sont agréés, enregistrés ou autrement qualifiés et possèdent les qualités requises en vertu des lois locales pour agir à titre de promoteur de l'étude clinique, de site d'étude ou d'investigateur, selon le cas. L'Établissement atteste également qu'aucune loi ou autre obligation ne lui interdit de mener l'étude et de conclure la présente entente. L'Établissement atteste par ailleurs que ni lui ni le promoteur-investigateur ne sont radiés en vertu des paragraphes 306(a) ou (b) de la *US Federal Food, Drug, and Cosmetic Act* et qu'ils n'ont pas utilisé ni n'utiliseront à aucun titre des services d'une personne radiée en vertu de cette loi en ce qui a trait aux activités à exécuter au titre de la présente entente.

11.3 Société affiliée. Tel qu'il est utilisé dans la présente entente, le terme « société affiliée » désigne toute entité qui contrôle directement ou indirectement la partie nommée, ou qui est contrôlée par celle-ci ou soumise à un contrôle commun au même titre que celle-ci.

11.4 Loi. Tel qu'il est utilisé dans le présent accord, le terme « loi » (ou « lois ») doit être compris comme englobant toutes les règles – locales, nationales, régionales ou internationales – ayant une force légale contraignante et obligatoire et qui sont prescrites, reconnues et appliquées par une autorité gouvernementale de contrôle. Les lois peuvent inclure, sans toutefois s'y limiter, les statuts, les règlements administratifs, les traités et les décrets.

11.5 Droit applicable. La présente entente est régie et interprétée conformément aux lois de la province de Québec, sans référence à ses règles en matière de divergence de lois, et aux lois du Canada applicables aux présentes. Tout litige découlant de la présente entente sera porté devant les tribunaux de la province de Québec, qui auront compétence en la matière, et chaque partie se soumettra irrévocablement à la compétence de ces tribunaux.

11.6 Données personnelles. Les renseignements qui pourraient être utilisés comme tels ou en association avec d'autres renseignements disponibles pour identifier une personne précise sont considérés comme des « données personnelles ».

11.7 Traitement des données personnelles par Pfizer. Pfizer utilise des systèmes électroniques mondiaux pour le traitement de certains renseignements ayant trait aux études fondées sur des essais proposés par des investigateurs. Ces systèmes peuvent comporter certaines données personnelles se rapportant aux personnes qui participent à l'étude ou qui effectuent des travaux dans le cadre de l'étude et que l'Établissement fournit à Pfizer. Les données personnelles utilisées dans ces systèmes comprennent généralement des renseignements tels que le nom, le domaine de spécialisation et les coordonnées des personnes. Pfizer peut transférer ces données personnelles à ses sociétés affiliées, à ses partenaires de recherche ou commerciaux, à ses fournisseurs de services contractuels ou à ses consultants, ou encore, aux autorités gouvernementales compétentes. Ces destinataires peuvent être situés à l'extérieur du pays dans lequel l'étude a été effectuée, y compris aux États-Unis.

11.8 Attribution et délégation

a. Par l'Établissement. L'Établissement ne peut attribuer des droits ni déléguer ou sous-traiter (« déléguer ») des obligations en vertu de la présente entente sans l'autorisation écrite de Pfizer. Si Pfizer autorise la délégation d'obligations, l'Établissement demeure responsable envers Pfizer de l'exécution de ces obligations.

b. Par Pfizer. Pfizer peut attribuer librement et déléguer des droits et obligations liés à l'étude à une société affiliée de Pfizer ou à un ayant droit relativement à un produit ou à un domaine de recherche d'intérêt de Pfizer auquel l'étude se rapporte. Moyennant un préavis de la part de l'Établissement, Pfizer peut également attribuer librement des droits et déléguer des obligations à un partenaire de recherche ou commercial ou à un fournisseur de services contractuels. Pfizer ne peut autrement céder ses droits ou déléguer ses obligations en vertu de la présente entente sans l'autorisation écrite de l'Établissement. Si Pfizer délègue des obligations, Pfizer demeure responsable envers l'Établissement de l'exécution de ces obligations.

11.9 Intégralité de l'entente. La présente entente (y compris les annexes) ainsi que le protocole approuvé par Pfizer auquel il est fait référence représentent l'entente complète entre les parties relativement à ce sujet. Cette entente remplace toute entente antérieure entre les parties (verbale ou écrite) relative à cette étude, à l'exception de toute obligation qui, par les modalités d'une telle entente, survivrait indépendamment de la présente entente.

11.10 Divergence avec les annexes ou le protocole. En cas de divergence entre la présente entente et l'une ou l'autre de ses annexes, les modalités de cette entente prévaudront. En cas de divergence entre la présente entente et le protocole, l'entente prévaudra, sauf en ce qui a trait aux questions de nature médicale, scientifique ou clinique relatives à la réalisation de l'étude, pour lesquelles le protocole prévaudra.

11.11 Besoins de financement. Pfizer ne versera aucun paiement en sus du financement énoncé à l'annexe A dans le cadre de la présente entente, à moins qu'elle n'ait d'abord approuvé ce paiement par écrit. Toutes les factures soumises à Pfizer par l'Établissement en vertu de cette entente doivent décrire en détail raisonnablement suffisant l'objet de la demande de paiement.

11.12 Droit de vérification. Pfizer a le droit de prendre toutes les mesures raisonnables pour s'assurer que chaque paiement qu'elle effectue est utilisé correctement et légitimement. Sur demande par Pfizer, l'Établissement doit :

a. fournir la documentation des débours, dépenses ou frais pour lesquels le financement de Pfizer a été utilisé;

b. permettre, pendant la durée de l'entente et pour une période de trois ans après le versement du paiement final en vertu de l'entente, que les vérificateurs internes et externes de Pfizer aient accès à tous les livres, documents, pièces et dossiers pertinents de l'Établissement et du promoteur-investigateur relativement aux transactions liées à l'entente. Lorsque l'entente comporte des études cliniques, des mesures de protection acceptables seront en place pour protéger la confidentialité des sujets d'étude.

11.13 Survie. Les paragraphes 1.1 (Investigateurs et personnel de recherche), 4.3 (Utilisation du soutien à un essai proposé par un investigateur), les articles 5 (Données de l'étude et résultats de l'étude), 6 (Rapport d'étude), 7 (Confidentialité des données), 8 (Publications) et 10 (Résiliation), le paragraphe 11.1 (Indemnisation) et l'article 12 (Lutte contre la corruption) survivront à l'expiration ou à la résiliation de la présente entente pour quelque raison que ce soit.

11.14 Communications électroniques. L'Établissement et le promoteur-investigateur acceptent de recevoir des communications électroniques de Pfizer dans le cadre de la présente entente et de toutes transactions futures avec Pfizer. L'Établissement et le promoteur-investigateur peuvent retirer leur consentement à de telles communications en fournissant un avis conformément à l'article 13 (Avis).

12. Lutte contre la corruption

12.1 Définitions

a. État. Comme utilisé dans la présente entente, le terme « État » comprend tous les niveaux et paliers de l'administration publique (niveau local, régional ou national, et palier administratif, législatif ou exécutif).

b. Représentant de l'État. Le terme « représentant de l'État » est défini à l'annexe D.

12.1 Garanties. L'Établissement garantit ce qui suit à Pfizer :

a. Le soutien financier de Pfizer n'incitera pas l'Établissement, le promoteur-investigateur et, à leur connaissance, toute personne affiliée à l'Établissement ou au promoteur-investigateur, à faire quoi que ce soit qui aurait pour effet que Pfizer obtienne ou conserve indûment un contrat ou obtienne indûment un avantage commercial.

b. Ni l'Établissement ni le promoteur-investigateur ni, à leur connaissance, toute personne affiliée à l'Établissement, au promoteur-investigateur ou à ce soutien n'utilisera une quelconque partie du soutien financier de Pfizer pour offrir ou verser, directement ou indirectement, de l'argent ou tout objet de valeur, dans un effort visant à influencer un représentant de l'État ou toute autre personne pour que Pfizer obtienne ou conserve indûment un contrat ou obtienne indûment un avantage commercial, et ni l'un ni l'autre n'a accepté ou n'acceptera à l'avenir, un tel paiement.

c. Pfizer sera en droit de révoquer ou de suspendre tout soutien financier si elle apprend que l'Établissement ou le promoteur-investigateur, ou toute personne affiliée à l'Établissement ou au promoteur-investigateur ou à ce soutien, a utilisé ou a l'intention d'utiliser une quelconque partie du soutien pour chercher à influencer indûment un représentant de l'État ou toute autre personne dans le but d'obtenir ou de conserver un contrat ou d'obtenir un avantage commercial.

d. Pfizer peut en tout temps divulguer publiquement vous avoir offert un soutien financier, et indiquer notamment le montant d'un tel soutien.

12.2 Non-conformité. Le défaut de respecter, ou une intention démontrée de ne pas respecter, l'une ou l'autre des garanties énoncées au paragraphe 12.2 ci-dessus constitueront une cause suffisante pour que Pfizer résilie immédiatement cette entente conformément à l'alinéa 10.1.d Résiliation motivée. En pareilles circonstances, Pfizer n'est pas dans l'obligation d'offrir à l'Établissement une occasion de remédier à la situation ou de lui verser tout autre paiement au moment de la résiliation, y compris tout paiement pour des engagements non résiliables pris par l'Établissement relativement à l'étude.

13 Avis. Tout avis qu'une partie désire donner ou signifier à une autre partie doit être fait par écrit et peut être remis en mains propres, envoyé par courrier recommandé prépayé avec demande d'accusé de réception, ou envoyé par télécopieur aux coordonnées suivantes :

Si à destination de Pfizer :

PFIZER CANADA INC.
17300, autoroute Transcanadienne

Kirkland (Québec)
H9J 2M5

À l'attention de Stéphane Dion
Télécopieur : 514-693-4715
Courriel : stephan.dion@pfizer.com

Envoyer une copie supplémentaire de chaque avis à l'attention de la Division des affaires juridiques de Pfizer, à l'adresse précitée et au numéro de télécopieur 514-426-7599.

Si à destination de l'Établissement :

Laboratoire de santé publique du Québec (LSPQ)
20045, chemin Sainte-Marie, Sainte-Anne-de-Bellevue (Québec)
H9X 3R5 Dr Jean Longtin
À l'attention de la ~~D^re^ Cécile Tremblay~~ BL 2015/11/16
Télécopieur : 514-457-6346
Courriel : xxx

Si à l'intention du promoteur-investigateur :

Brigitte Lefebvre, Ph. D.
Laboratoire de santé publique du Québec (LSPQ)
20045, chemin Sainte-Marie, Sainte-Anne-de-Bellevue (Québec)
H9X 3R5

Télécopieur : 514-457-6346
Courriel : xxx

ou à toute autre adresse que la partie à laquelle l'avis est destiné a communiquée aux autres parties au moyen d'un avis qui leur a été donné ou signifié de la façon décrite dans cet article. Dans le cas d'une remise en mains propres ou d'une transmission par télécopieur, l'avis sera réputé avoir été donné au moment de sa réception par le destinataire et, dans le cas d'un envoi postal, l'avis sera réputé avoir été donné sept jours après avoir son envoi.

Accepté et approuvé par :

PFIZER CANADA INC.

Par : [redacted]

Nom : Vratislav Hadrava

Titre : Vice-président – Affaires médicales, Canada

Date : 04-NOV-2015

Par : [redacted]

Nom : Jelena Vojicic

Titre : Gestionnaire en chef – Vaccins, Canada

Date : Nov 04-2015

ÉTABLISSEMENT

Par : [redacted]

Dre JOCELYNE SAUVÉ

Nom : ~~D^{re} Cécile Tremblay~~ BL 2015/11/16

Titre : Vice présidente - affaires scientifiques

Date : 2015-11-20

PROMOTEUR-INVESTIGATEUR

Par : [redacted]

Nom : D^{re} Brigitte Lefebvre

Date : 2015/11/20

Date de la version modèle : Octobre 2014

Annexe A
CALENDRIER DES PAIEMENTS

Financement

Le financement total approuvé pour l'étude qui sera effectuée par le promoteur-investigateur est de **SEPT CENT SEPT MILLE QUATRE-VINGTS DOLLARS (707 080,00 $CA)**. Ce montant inclut tous frais généraux de l'Établissement et ne comprend pas les taxes applicables.

Calendrier des paiements de financement

Étapes	Montant du paiement
Paiement initial – dès la réception par Pfizer de l'entente dûment signée (voir la remarque ci-dessous)	231 295,00 $
Paiement provisoire – Après réception et examen d'une mise à jour de statut de l'étude en août 2016	234 180,00 $
Paiement final – dès la réception par Pfizer des résultats de l'étude (voir la remarque ci-dessous)	241 605,00 $

Avis d'étapes et paiements : Pour demander un paiement, aviser Pfizer par écrit lorsque chacune des étapes a été franchie. Référencer le numéro de suivi de Pfizer dans chaque demande de paiement.

Renseignements sur le bénéficiaire

Nom du bénéficiaire (tel qu'il figurera sur le chèque) :

Institut national de santé publique du Québec (INSPQ)

Adresse du bénéficiaire :

20045, chemin Sainte-Marie, Sainte-Anne-de-Bellevue (Québec) H9X 3R5

À l'attention de :

Laboratoire de santé publique du Québec (LSPQ)

Paiement initial. Pfizer ne versera aucun paiement initial tant qu'elle n'aura pas reçu 1) une copie signée de l'accord et 2) les documents nécessaires indiqués à l'annexe B, Exigences en matière de documentation.

Paiement final. Pfizer versera le paiement final uniquement après la réception du rapport de l'étude et l'achèvement de toutes obligations restantes applicables dans le cadre de l'entente.

Annexe B
EXIGENCES EN MATIÈRE DE DOCUMENTATION

Pfizer	**STUDY DOCUMENT REQUIREMENTS FORM**	

STUDY INFORMATION

PRINCIPAL INVESTIGATOR **Dr. Brigitte Lefebvre**

PFIZER INSPIIRE NO. WI203144 INSTITUTIONAL REFERENCE NUMBER

PROTOCOL TITLE **Serotype monitoring of S. pneumoniae invasive strains in adult population in the province of Quebec_ a 3 years study evaluation.**

DOCUMENTATION REQUIREMENTS

MATERIALS ENCLOSED WITH THIS PACKET: (DELETE ANY ITEMS BELOW THAT DO NOT APPLY)

☒ Site Information Sheet **(agreement information form)**

□ Drug Supply Request Form

□ Reportable Event Fax Cover Sheet

□ Pfizer Safety Reporting Reference Manual for IIR studies

□ Pfizer IIR Adverse Event Report Form and IIR Adverse Event Report Form Completion Instructions

□ Exposure During Pregnancy (EDP) Supplemental Form

□ Product information (document or reference)

□ IRS Web site address to download Form W-9 (US/Puerto Rico only)

PRINCIPAL INVESTIGATOR MUST PROVIDE TO PFIZER: (ONLY BOXES CHECKED BELOW)

Documents required to generate an IIR Agreement

☒ Completed **(agreement information form)**

□ Completed IRS *Form W-9* (US/Puerto Rico only for payee entity)

Documents required to be submitted prior to receiving monetary support and/or drug supplies

□ Completed Site Information Sheet (Drug Supply Information and/or Financial Information Tab[s])

☒ Executed IIR agreement

☒ Final study protocol (for a study with sites in the EU, the principal investigator must sign the final study protocol as required for qualified person [QP] release of drug supplies)

☒ IRB/IEC approval letters (initial approval and annual renewals, as applicable)

□ Regulatory response

For US studies:

□ FDA IND response (IND number or exemption – *may not apply to all consumer products*)

□ DEA number for controlled substances

For EU studies:

□ Approved clinical trial application (CTA) in English (as required for QP release)

□ Submission letter for the CTA

For non-US, non-EU studies:

□ Appropriate Regulatory review/approval based upon local country requirements

STUDY DOCUMENT REQUIREMENTS FORM

Site Information Sheet / agreement information form)

The information requested on the *Site Information Sheet /****Agreement information form*** is critical to Pfizer in order to develop an agreement, to reduce the agreement's review time, and to ensure that monetary support is sent to the appropriate payee or drug supply is sent to the appropriate address. Withholding or delaying Pfizer's receipt of this form will significantly delay the contracting process for the approved research.

Final Protocol and Amendments

Pfizer will not provide support to an IIR study until after receipt of the final study protocol. If the research described in the final protocol is materially different from that in the approved proposal, then Pfizer may choose to modify or withhold its support.

As indicated in the agreement, the principal investigator must also promptly provide Pfizer with any amendments to the Pfizer-approved final study protocol. Continuation of support by Pfizer for an IIR study will be contingent on Pfizer's review and acceptance of these changes.

For studies with sites in the EU where drug support is being requested, the final study protocol must be signed by the principal investigator and is required for QP release of drug supplies.

Institutional Review Board (IRB)/Independent Ethics Committee (IEC) Documents

For studies that require IRB/IEC approval, Pfizer will only provide support for an IIR study after receipt of a copy of the IRB/IEC approval letter.

Continuation of support by Pfizer requires timely submission of a copy of IRB/IEC renewal documentation subsequent to the original IRB/IEC approval (as required per local regulations).

Regulatory Response

US Clinical Studies: FDA IND Response or IND Exemption Documentation. For an interventional clinical study involving a Pfizer drug, an investigational new drug (IND) application may need to be filed with the U.S. Food and Drug Administration (FDA). Please review IND requirements under 21 CFR 312 (available at http://www.fda.gov) to determine whether an IND is required.

For this type of study, Pfizer will not provide any IIR support until after receipt of documentation that an IND has been filed or that the study is exempt from an IND filing under 21 CFR 312.2(b)(1).

European Union Clinical Studies. For studies for which conduct under a clinical trial application (CTA) is required, Pfizer will not provide any IIR support until after receipt of a copy of the submission letter to the CTA, in English.

If Pfizer will provide packaged and labeled Pfizer product, then Pfizer must receive a copy of the approved CTA, with Section 4.2 (IMPD or Letter of Access from Pfizer) and Section D (in its entirety) must be translated in English, before Pfizer can provide QP release of product. For more information regarding CTAs, please consult http://eudract.emea.europa.eu/document.html.

Should your local regulatory authority require documentation from Pfizer, please contact your IIR manager for assistance.

Non-US/Non-EU Studies. Should your local regulatory authority require documentation from Pfizer, please contact your IIR manager for assistance.

Investigator-Initiated Research Agreement

Pfizer will provide the principal investigator or the contracting office with an IIR agreement that documents the terms under which Pfizer will provide the research grant. Development of the agreement is based upon information you have supplied on the enclosed forms.

Drug Supply Request Form

If Pfizer has agreed to supply drug, then the *Drug Supply Request Form* can be used to communicate your clinical supply needs throughout the course of the IIR study. Pfizer will not ship any clinical supplies until all required documents have been received and an IIR agreement has been executed.

NOTE: Availability of drug may take between eight weeks and twelve months, depending upon the product and its packaging and labeling requirements. Contact the appropriate IIR manager to determine available quantities of drug and timelines for shipment.

For Oncology Studies Conducted in the United States. If Pfizer is not providing clinical supplies for this study, then Pfizer cannot be held responsible for drug cost reimbursement. For assistance with third-party reimbursement procedures and indigent patients, contact FirstRESOURCE, Pfizer Oncology's Reimbursement and Patient Assistance Program, at 877-744-5675 prior to initiating therapy.

IRS Form W-9

Pfizer requires that all grant recipients based in the U.S. or Puerto Rico who receive monetary support complete and submit IRS *Form W-9*. This form shall be completed for the entity which will be receiving the grant payment(s). Please verify with your grants office that the name of the payee is correct and that it is the legal entity name related to the tax identification number. The latest version of *Form W-9* may be downloaded from the IRS Web site from: http://www.irs.gov/pub/irs-pdf/fw9.pdf.

Product Information

Pfizer is required to provide relevant and current scientific information about the investigational product to the investigator. This may be accomplished by supplying one of the following Pfizer-approved documents to the investigator: Investigator Brochure (IB), package insert (PI), or local product document (LPD).

Safety Reporting

Safety Reporting Reference Manual for IIR Studies with Pfizer Products. Detailed information regarding a principal investigator's (or investigators') adverse event reporting responsibilities for a Pfizer-supported IIR study can be found in the accompanying training manual. **Please read through this document carefully. Principal investigators must understand and fully comply with the adverse event reporting requirements of their studies.**

NOTE: Reporting an adverse event to Pfizer does not relieve the institution of its responsibility to report the event to the FDA or to the local regulatory authorities that govern that institution.

IIR SAE Form and IIR SAE Report Form Completion Instructions. For those studies where the principal investigator is required to submit reportable events (AEs and SAEs) to Pfizer, the investigator may use the *Pfizer IIR SAE/Adverse Event Report Form* to submit the event. Instructions for completion will also be provided.

Reportable Event Fax Cover Sheet. For those studies where the principal investigator is required to report adverse events and other reportable events to Pfizer, the investigator must use the attached *Reportable Event Fax Cover Sheet* along with the Pfizer-approved *Adverse Event Report Form*.

Annexe C
Protocole

1

Serotype monitoring of *S. pneumoniae* invasive strains in adult population in the province of Quebec: a 3 years study evaluation

Principal Investigator
Brigitte Lefebvre, Ph. D.
Microbiologist, Laboratoire de santé publique du Québec

Co-PI
Cécile Tremblay, MD, Pfizer/University of Montreal Chair on HIV Translational Research, University of Montreal.
Director, Laboratoire de santé publique du Québec

Background

Streptococcus pneumoniae is responsible for various infections such as pneumonia, otitis, sinusitis, peritonitis, endocarditis and meningitis(1). The incidence of invasive *S. pneumoniae* is often used as an indicator of the burden of pneumococcal disease. Virulence and invasiveness varies among serotypes. In *S. pneumoniae*, several virulence factors are known; among these, the *cps* locus encoded capsule is a crucial one, as the prime target for vaccine development. Although several vaccines (PCV-7, PCV-10, PCV-13 and PCV-23) with different coverage have been developed against *S. pneumoniae*, invasive pneumococcal disease remains a public health concern as vaccine replacement phenomenon has been observed(2).

In December 2004, PCV-7 vaccination was implemented free to all newborns in Quebec, using a 3-dose schedule (2, 4 and 12 months). Simultaneously, the vaccine could be offered free of charge to all children under the age of 5, during routine visits. In 2008, a new PCV-10 containing 3 serotypes not included in PCV-7 vaccine was licensed in Canada. It was introduced in Quebec in children in the summer of 2009. In 2009, PCV-13 vaccine was approved in Canada. It was introduced in the Quebec immunization program in January 2011 and replaced PCV-10.

The introduction of PCV-7 had not only an important impact on the number and the diversity of strains isolated from children under 5 years of age, but the impact was also observed in individuals ≥ 5 year old. Thus, the proportion of serotypes included in PCV-7 has dramatically declined since 2005. However, there was an increase in the proportion of serotypes 7F and 19A which are not included in PCV-7 and an increase of non-vaccine serotypes was observed. In 2013, a decrease in the frequency of 7F and 19A serotypes in individuals ≥ 5 year old was observed. However, the number of circulating serotypes not included in the PCV-7, PCV-10 and PCV-13 is increasing.

Thus, sustained laboratory monitoring is essential because it allows the study of evolution of circulating serotypes as well as antibiotic resistance patterns, two crucial parameters for planning immunization programs, the choice of vaccines and the development of treatment guidelines. Analysis of invasive strains allows for the study of serotypes distribution and antibiotic susceptibility patterns of strains responsible for the most severe forms of pneumococcal disease. Monitoring of circulating serotypes is essential to assess the impact of vaccination programs of the province of Quebec.

In 1996, the Public Health Laboratory of Quebec (LSPQ) in collaboration with hospital laboratories established a laboratory surveillance program of *S. pneumoniae* invasive strains. The program's objectives were to study the serotype distribution circulating in Quebec and establish their antibiotic susceptibility profiles. This program was based on the collection of strains from sentinel laboratories. In 2005, in order to assess the impact of the universal immunization program against *S. pneumoniae* in children, the program was expanded to all invasive strains of *S. pneumoniae* isolated from children under 5 years of age.

This monitoring program has kept track of the evolution, in Quebec children, of various serotypes and resistance in connection with the introduction of the PCV-13 vaccine in 2011 and more specifically allows for the measure of its impact on the prevalence of serotypes 7F and 19A, two serotypes highly prevalent in Quebec. Currently, the provincial surveillance program is limited to strains collected in children less than 5 years of age and to adult strains from sentinel laboratories which represent less than 25% of the total invasive strains in the adult population. Therefore, we may be underestimating the diversity of circulating strains especially in areas not represented in the sentinel program and may not capture adequately seasonal variation. Two years ago, we proposed, a study evaluating the benefits of acquiring data on all invasive strains isolated in patients (≥ 5 years old) of the province of Quebec compared to sentinel sites. This study was launched in August 2013, with the financial support of Pfizer. Preliminary data from the first 18 months of extended surveillance indicate that some emerging serotypes may not be fully captured by the sentinel sites, although these observations need to be evaluated by longer follow-up.

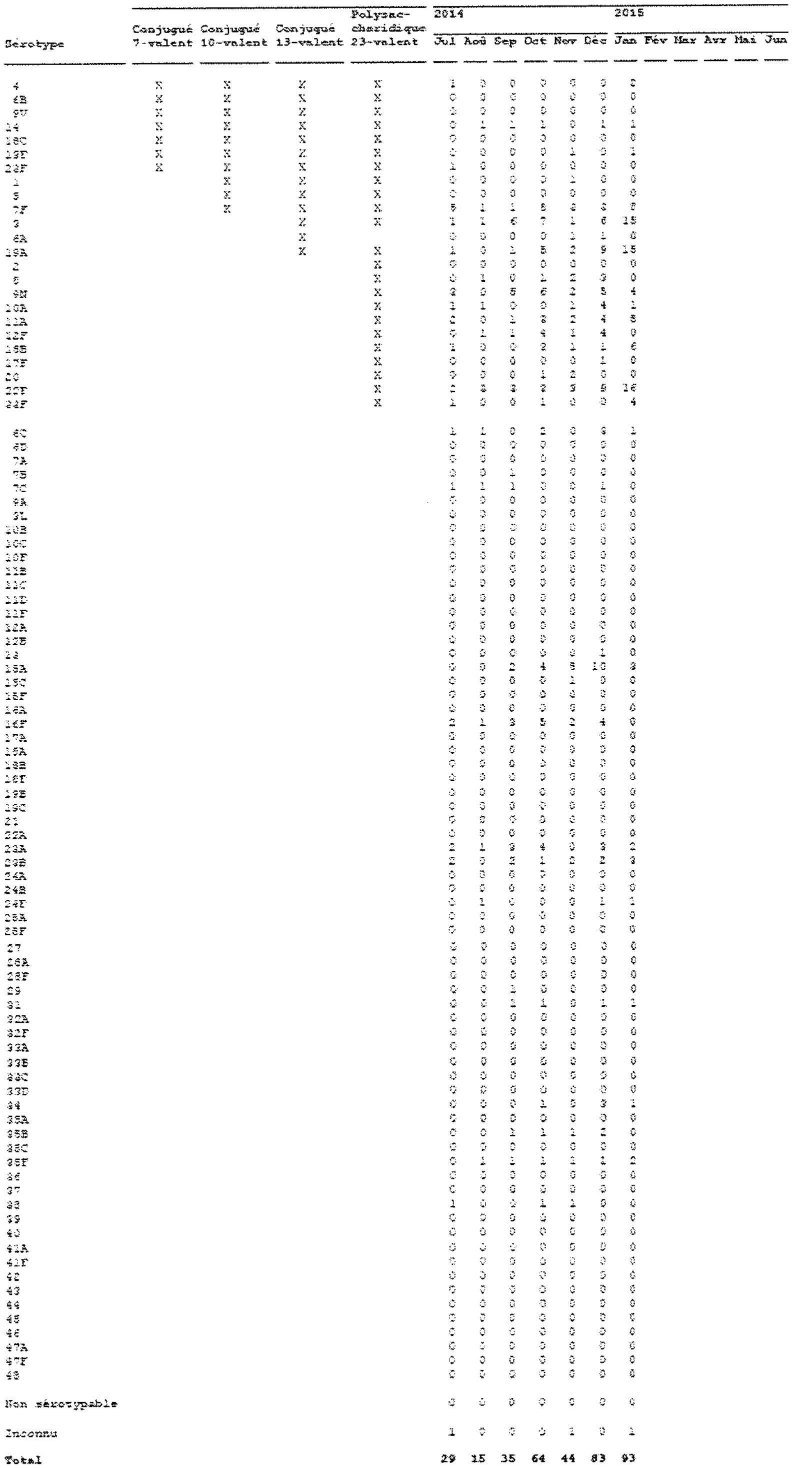
Preliminary data from surveillance of invasive *S. pneumoniae* in individuals ≥ 5 years old

After 18 months of extended surveillance, we have identified a higher proportion of two serotypes, the 6A and 15A, which had not previously been identified with the sentinel sites surveillance program. Serotype 6A is included in the currently used PCV-13 vaccine and serotype 15A is not included in this vaccine and exhibits multi-resistance. A recent paper from Israels howed a similar increase of 15A serotype among adult invasive pneumococcal disease[2]. Emergence of serogroup 15 was also described by Liyanapathirana *et al.*[3] in nasopharyngeal carriage of hospitalized children. Furthermore, our data analysis revealed an overrepresentation of some serotypes when only sentinel data are analyzed. The clinical significance of these serotypes is not yet defined. However, this supports the necessity to expand our broadened monitoring over a longer period of time to evaluate the establishment of these serotypes into Quebec's ecology and their relevance for vaccine development.

Before the beginning of our study in 2013, reporting of data was available in 3 formats: i) The annual provincial aggregated data generally available one year after data collection[4]; ii) The monthly LSPQ StatLabo report providing aggregated data with a 2 months delay[5] iii) Individual reports for each strain sent to participating laboratories as well as public health stakeholders, up to 4 months after strain reception. As part of the current study, we were able to make available in real time information on circulating serotypes by publishing a monthly report including all serotypes identified, classified by age groups in the bulletin StatLabo (Fig. 1.).

We propose to continue our study for another three years to allow for a full characterization of circulating serotypes including clustering in certain geographical areas or seasonal variation, to establish incidence of invasive pneumococcal disease in the Quebec population, and to define if this surveillance program provides added value to a sentinel site based approach. Results of this research project could help guide public health authorities in immunization strategies and will also provide useful information for vaccine design.

Figure 1. Données mensuelles des souches invasives de *S. pneumoniae* chez les patients de 5 ans et plus [6].



Sérotype	Conjugué 7-valent	Conjugué 10-valent	Conjugué 13-valent	Polysac-charidique 23-valent	2014						2015					
					Jul	Aoû	Sep	Oct	Nov	Déc	Jan	Fév	Mar	Avr	Mai	Jun
4	X	X	X	X	1	0	0	0	0	0	0					
6B	X	X	X	X	0	0	0	0	0	0	0					
9V	X	X	X	X	0	0	0	0	0	0	0					
14	X	X	X	X	0	1	1	1	0	1	1					
18C	X	X	X	X	0	0	0	0	0	0	0					
19F	X	X	X	X	0	0	0	0	1	0	1					
23F	X	X	X	X	1	0	0	0	0	0	0					
1		X	X	X	0	0	0	0	1	0	0					
5		X	X	X	0	0	0	0	0	0	0					
7F		X	X	X	5	1	1	8	6	6	8					
3			X	X	1	1	6	7	1	6	15					
6A			X		0	0	0	0	1	1	0					
19A			X	X	1	0	1	5	2	9	15					
2				X	0	0	0	0	0	0	0					
8				X	0	1	0	1	2	0	0					
9N				X	2	0	5	6	2	5	4					
10A				X	1	1	0	0	1	4	1					
11A				X	2	0	1	3	2	4	5					
12F				X	0	1	1	4	1	4	0					
15B				X	1	0	0	2	1	1	6					
17F				X	0	0	0	0	0	1	0					
20				X	0	0	0	1	2	0	0					
22F				X	2	3	3	3	5	5	16					
33F				X	1	0	0	1	0	0	4					
6C					1	1	0	2	0	3	1					
6D					0	0	0	0	0	0	0					
7A					0	0	0	0	0	0	0					
7B					0	0	1	0	0	0	0					
7C					1	1	1	0	0	1	0					
9A					0	0	0	0	0	0	0					
9L					0	0	0	0	0	0	0					
10B					0	0	0	0	0	0	0					
10C					0	0	0	0	0	0	0					
10F					0	0	0	0	0	0	0					
11B					0	0	0	0	0	0	0					
11C					0	0	0	0	0	0	0					
11D					0	0	0	0	0	0	0					
11F					0	0	0	0	0	0	0					
12A					0	0	0	0	0	0	0					
12B					0	0	0	0	0	0	0					
13					0	0	0	0	0	1	0					
15A					0	0	2	4	5	10	3					
15C					0	0	0	0	1	0	0					
15F					0	0	0	0	0	0	0					
16A					0	0	0	0	0	0	0					
16F					2	1	3	5	2	4	0					
17A					0	0	0	0	0	0	0					
18A					0	0	0	0	0	0	0					
18B					0	0	0	0	0	0	0					
18F					0	0	0	0	0	0	0					
19B					0	0	0	0	0	0	0					
19C					0	0	0	0	0	0	0					
21					0	0	0	0	0	0	0					
22A					0	0	0	0	0	0	0					
23A					2	1	3	4	0	3	2					
23B					2	0	2	1	2	2	3					
24A					0	0	0	0	0	0	0					
24B					0	0	0	0	0	0	0					
24F					0	1	0	0	0	1	1					
25A					0	0	0	0	0	0	0					
25F					0	0	0	0	0	0	0					
27					0	0	0	0	0	0	0					
28A					0	0	0	0	0	0	0					
28F					0	0	0	0	0	0	0					
29					0	0	1	0	0	0	0					
31					0	0	1	1	0	1	1					
32A					0	0	0	0	0	0	0					
32F					0	0	0	0	0	0	0					
33A					0	0	0	0	0	0	0					
33B					0	0	0	0	0	0	0					
33C					0	0	0	0	0	0	0					
33D					0	0	0	0	0	0	0					
34					0	0	0	1	0	3	1					
35A					0	0	0	0	0	0	0					
35B					0	0	1	1	1	2	0					
35C					0	0	0	0	0	0	0					
35F					0	1	1	1	1	1	2					
36					0	0	0	0	0	0	0					
37					0	0	0	0	0	0	0					
38					1	0	0	1	1	0	0					
39					0	0	0	0	0	0	0					
40					0	0	0	0	0	0	0					
41A					0	0	0	0	0	0	0					
41F					0	0	0	0	0	0	0					
42					0	0	0	0	0	0	0					
43					0	0	0	0	0	0	0					
44					0	0	0	0	0	0	0					
45					0	0	0	0	0	0	0					
46					0	0	0	0	0	0	0					
47A					0	0	0	0	0	0	0					
47F					0	0	0	0	0	0	0					
48					0	0	0	0	0	0	0					
Non sérotypable					0	0	0	0	0	0	0					
Inconnu					1	0	0	0	1	0	1					
Total					**29**	**15**	**35**	**64**	**44**	**83**	**93**					

Project objectives

1- To characterize serotypes and antibiotic resistance profile of all invasive *S. pneumoniae* strains from the adult population in Quebec.

2- To assess whether the serotype profile differ from the entire population compared to the profile obtained from sentinel sites.

3- To follow the incidence of IPD in Quebec over several years and evaluate the impact of current vaccine, PCV-13 on IPD incidence.

Methodology

The research project will cover the complete adult population for 3 additional years (September 2015 to August 2018). We expect to collect 550 additional strains yearly to reach an average of 1000 strains yearly (estimated based on 2014 data). This will represent all the invasive *S. pneumoniae* strains of the province of Quebec. We propose to conduct this extended program for a 3-year period, after which a program evaluation will be performed. Serotyping using Quellung methodology and determination of susceptibility profiles using microdilutions method will be performed on all *S. pneumoniae* invasive strains collected in patients aged of ≥ 5 year old.

Those additional strains will be provided by non-sentinel hospitals (n=74) which, until now, only provided LSPQ with strains from child <5 years old and strains resistant to penicillin (≥ 0.12 mg/L according meningitis criteria).

Data will be published monthly through StatLabo including serotype stratified according to patients' age and months.

Time-line

Steps	Lenght
Monitoring of invasive *S. pneumoniae* serotypes in patients aged ≥ 5 years old.	Years 1, 2 and 3
Real-time updating of StatLabo surveillance information using Quellung method.	Years 1, 2 and 3
Conferences.	Years 1, 2 and 3
Publication.	Year 3

Timeframe

See annexe 1

Project Benefits

1- Real-time monitoring of invasive *S. pneumoniae* serotypes and antibiotic resistance in adult in the province of Quebec.
2- Monitoring of IPD incidence in Quebec.
3- Comparison of actual provincial surveillance program using data from sentinel hospitals vs data from the study for individuals aged of ≥ 5 years old.
4- Data available for public health orientation on immunization program in adult population.

Deliverables

1- Monitoring of invasive *S. pneumoniae* strains in adult population for 3 years, starting in Septembre 2015 and ending in August 2018.

2- Monthly reporting of serotypes in StatLabo.

3- Data from the study will be presented at scientific meetings (AMMIQ [at the end of year 1], CACMID [at the end of year 2], ISPPD[at the end of year 3]) and published in a peer reviewed journal (Vaccine/PlosOne) at the end of the study.

References

1- Spellerberg, B. and Brandt C. *Streptococcus*. 2011. Manual of clinical microbiology. 10th edition. American Society for microbiology, Washington, D.C.

2- Regev-Yochay G, Paran Y, Bishara J, Oren I, Chowers M, Tziba Y, Istomin V, Weinberger M, Miron D, Temper V, Rahav G, Dagan R; IAIPD group. 2015. Early impact of PCV7/PCV13 sequential introduction to the national pediatric immunization plan, on adult invasive pneumococcal disease: A nationwide surveillance study. Vaccine. 25;33(9):1135-42.

3- Liyanapathirana, V., EA. Nelson, I. Ang, R. Subramanian, H. Ma, M. Ip. 2015. Emergence of serogroup 15 *Streptococcus pneumoniae* of diverse genetic backgrounds following the introduction of pneumococcal conjugate vaccines in Hong Kong. Diagn Microbiol Infect Dis. Jan;81(1):66-70.

4- Lefebvre B., C. L. Tremblay. 2012. Programme de surveillance du pneumocoque. Rapport 2011. INSPQ. ISBN : 978-2-550-66364-5.

5- Bulletin STATLABO. Institut national de santé publique du Québec (INSPQ), Laboratoire de santé publique du Québec (LSPQ). Statistiques d'analyses du LSPQ. 2012. Vol. 11, no.11.

6- Bulletin STATLABO. Institut national de santé publique du Québec (INSPQ), Laboratoire de santé publique du Québec (LSPQ). Statistiques d'analyses du LSPQ.2015 Vol. 14, no. 2.

Annexe 1. Time Frame/Project Goals (arrows), milestones (red), task (blue bars) and timelines.

	YEAR 1	YEAR 2	YEAR 3
Surveillance of *S. pneumoniae* serotyping using Quellung method.	←——	———	——→
Continuous and real time updating of StatLabo surveillance information.	←——	———	——→
- Database updating with serotypes in relationship with age (StatLabo$_{sp}$) - Evaluation of the impact of extended surveillance to adults in the design of vaccines at the end of year 1, 2 and 3.	▒ ●	▒ ●	▒● ●
Conferences	●	●	●
Publication			▒●

ANNEXE D

DÉFINITION DE REPRÉSENTANT DE L'ÉTAT

Le terme « **Représentant de l'État** », lequel doit recevoir une interprétation large, désigne :

(i) un Représentant de l'État élu ou nommé d'un autre pays que les États-Unis (p. ex. : un législateur ou un membre d'un ministère non américain),

(ii) un employé ou une personne agissant à la place ou au nom d'un Représentant de l'État non américain, d'une agence ou d'une entreprise non américaine assumant une fonction gouvernementale, ou détenue ou contrôlée par un gouvernement non américain (par exemple un professionnel de la santé employé par un hôpital public non américain ou un investigateur employé par une université publique non américaine),

(iii) un représentant d'un parti politique non américain, un candidat à une fonction publique non américaine, un employé ou une personne agissant à la place ou au nom d'un parti politique ou d'un candidat à une fonction publique non américaine,

(iv) un employé ou une personne agissant pour ou au nom d'une organisation publique internationale,

(v) un membre d'une famille royale ou membre d'un corps d'armée non américain, et

(vi) toute personne autrement considérée comme un Représentant de l'État en vertu des lois locales en vigueur ou des Politiques de Pfizer.

Cela signifie que les professionnels de la santé qui sont employés par un hôpital gouvernemental ou encore une université, qui y enseignent ou y jouissent de certains privilèges peuvent être considérés comme représentants de l'État – même s'ils n'y travaillent qu'à temps partiel. Dans bien des pays, particulièrement ceux où le gouvernement est propriétaire ou dirigeant de nombreux services de soins de santé et pharmacies, pratiquement tous les professionnels de la santé peuvent être considérés comme des représentants de l'État en vertu des lois FCPA et *Global Anti-Corruption* des États-Unis.

Les employés des organismes d'État suivants avec lesquels Pfizer interagit fréquemment sont automatiquement considérés comme des représentants de l'État au Canada :

- Santé Canada
- Industrie Canada
- Bureau du Conseil privé
- Cabinet du Premier ministre
- Affaires étrangères et Commerce international
- Conseil d'examen du prix des médicaments brevetés
- Anciens Combattants Canada
- Défense nationale
- Ministères des Finances Canada
- Gendarmerie royale du Canada
- Programme commun d'évaluation des médicaments (PCEM)
- Integrated Health Agencies (Canada atlantique)
- Centres de santé et de services sociaux (CSSS – successeurs des CLSC) (Québec)
- Groupes de médecine familiale (Québec)
- Cliniques réseau (CR ou CMA) (Québec)
- Réseaux locaux d'intégration des services de santé (Ontario)
- Équipes de santé familiale (Ontario)
- Regional Health Authorities (Ouest canadien)
- Agence canadienne d'inspection des aliments
- Direction des médicaments vétérinaires (DMV)

- Agence canadienne des médicaments et des technologies de la santé (ACMTS)
- Aquaculture Canada

Exemples de représentants de l'État au Canada :

- Représentants de l'État élus ou nommés;
- Fonctionnaires;
- Candidats déclarés d'un parti politique (en vue de l'investiture d'un parti ou d'une élection);
- Professionnels de la santé satisfaisant aux critères énoncés dans la définition de représentant de l'État, par exemple, professionnels de la santé au service a) de l'armée, b) du Service correctionnel du Canada (prisons et pénitenciers) ou c) d'hôpitaux ou d'établissements de santé exploités ou régis par l'État (hôpitaux psychiatriques, hôpitaux pour les anciens combattants), et professionnels de la santé membres de groupes de travail ou de comités étatiques (p. ex., le Comité consultatif d'experts pour le traitement des douleurs chroniques intenses, le Comité consultatif sur le sida, le Conseil consultatif national sur le troisième âge, le Medical Advisors Group);
- Professionnels de la santé administrateurs, dirigeants ou employés de tout établissement de soins de santé (p. ex., hôpital, clinique, etc.) ou de tout établissement d'enseignement supérieur (p. ex., collège, cégep, université, etc.) financé par l'État, ou qui y sont affiliés;
- Dirigeants, employés ou particuliers qui agissent à titre officiel au nom de conseils scolaires et de collèges communautaires;
- Officiers, employés ou particuliers qui agissent à titre officiel au nom de l'Organisation des Nations Unies, de l'Organisation mondiale de la santé, de l'Organisation mondiale du commerce, de la Commission mixte internationale États-Unis et Canada, du Comité international de la Croix-Rouge, de la Banque nord-américaine de développement (NADB), du Fonds monétaire international, de l'Organisation internationale de police criminelle (INTERPOL) ou de la Banque interaméricaine de développement; et dirigeants, employés ou particuliers qui agissent à titre officiel au nom des conseil scolaires et de collèges communautaires.

Version : 9 avril 2013

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist PHONE IIR Grant Specialist FAX	[redacted]

PLEASE COMPLETE AND RETURN BY: **April 26, 2016**

Per contractual requirements, we are requesting a status update on your IIR study supported by Pfizer via funding and/or drug. Please answer the following questions regarding the above referenced study by the due date. Answers from your last submitted update have been incorporated below; please update as needed and answer the remaining questions.

GENERAL INFORMATION

Pfizer Tracking #	**WI197603**	**Institutional Protocol #**	
Principal Investigator	**Dr. Brigitte Lefebvre**		
Study Title	**Molecular tools for serotyping for Streptococcus pneumoniae invasive strains surveillance in the province of Quebec.**		

STUDY UPDATE INFORMATION

Has this study been initiated?	□ NO ☒ YES	Date of initiation	mm/dd/yyyy 01/01/2016
Has the protocol been amended since last update?	☒ NO □ YES *(If YES, please provide the revised protocol)*		
Current IRB/IEC approval/renewal expires on **November 5, 2016**	If this is not current, please forward the most recent letter		
Have there been any personnel changes? *(If YES, please provide name and full contact info on Page 3)*			☒ NO □ YES
Target protocol enrollment	N/A	Date of first subject enrolled	N/A
Last reported enrollment	N/A	Actual enrollment to date *(this should not include screen failures)*	N/A
Targeted last subject last visit	N/A	Actual last subject last visit	N/A
Do you have current drug supply sufficient to complete the study? *(If NO, please complete the Drug Section on Page 3)*			☒ NO □ YES
Is this protocol closed to enrollment? *(patients may still be receiving therapy) N/A this is a project on methods's development for which no enrollment is required.*			□ NO □ YES
Targeted study completion date *(primary objectives met; patient therapy and final study analysis complete)*			mm/dd/yyyy 02/10/2017
Actual study completion date *(if applicable)*			mm/dd/yyyy N/A
Targeted date to provide results to Pfizer			mm/dd/yyyy 02/28/2017

PUBLICATION INFORMATION

Do you plan to publish? *(If YES, please complete the information below.)*	□ NO ☒ YES

Please be aware that, according to the IIR agreement, the investigator is required to provide Pfizer with an

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist PHONE IIR Grant Specialist FAX	[redacted]

opportunity to prospectively review any proposed publication, abstract or other type of disclosure that reports the results of the study.

FORMAT	**PUBLICATION** *(please include anticipated journal or audience)*	**PLANNED**	**ACTUAL**	**SUBMISSION DATE** mm/dd/yyyy
Abstract		☐	☐	
Manuscript	PlosOne	☒	☐	02/01/2017
Poster	CACMID	☒	☐	04/01/2017
Other		☐	☐	

SIGNATURE

NAME Brigitte Lefebvre

DATE 04/26/2016

[redacted]

SIGNATURE (ONLY if faxed)

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist	[redacted]	IIR Grant Specialist PHONE	[redacted]
IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist FAX	[redacted]

DRUG SUPPLY INFORMATION

SUPPLY CURRENTLY ON SITE	ACTIVE	PLACEBO
ESTIMATED REMAINDER REQUIRED TO COMPLETE STUDY	ACTIVE	PLACEBO
CAN PHARMACY ACCOMODATE TOTAL REMAINDER?	□ YES	□ NO

PERSONNEL INFORMATION

	PRINCIPAL INVESTIGATOR	COORDINATOR
NAME	Lefebvre Brigitte	
INSTITUTION	INSPQ/LSPQ	
MAILING ADDRESS	20045, chemin Sainte Marie, sainte-Anne-de-Bellevue, H9X 3R5, Québec, Canada	
TELEPHONE	514 4572070# 2334	
FAX	514 4576346	
EMAIL	Brigitte.Lefebvre@inspq.qc.ca	

	PHARMACIST	OTHER *(specify in additional comments)*
NAME		
INSTITUTION		
MAILING ADDRESS		
TELEPHONE		
FAX		
EMAIL		

ADDITIONAL COMMENTS

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist PHONE IIR Grant Specialist FAX	[redacted]

PLEASE COMPLETE AND RETURN BY: April 26, 2016

Per contractual requirements, we are requesting a status update on your IIR study supported by Pfizer via funding and/or drug. Please answer the following questions regarding the above referenced study by the due date. Answers from your last submitted update have been incorporated below; please update as needed and answer the remaining questions.

GENERAL INFORMATION

Pfizer Tracking #	**WI203144**	**Institutional Protocol #**	
Principal Investigator	**Dr. Brigitte Lefebvre**		
Study Title	**Serotype monitoring of S. pneumoniae invasive strains in adult population in the province of Quebec_ a 3 years study evaluation.**		

STUDY UPDATE INFORMATION

Has this study been initiated?	□ NO ☒ YES	**Date of initiation**	mm/dd/yyyy 01/01/2016
Has the protocol been amended since last update?		☒ NO □ YES *(If YES, please provide the revised protocol)*	
Current IRB/IEC approval/renewal expires on **November 5, 2016**		This is not current, please forward the most recent letter	
Have there been any personnel changes? *(If YES, please provide name and full contact info on Page 3)*			☒ NO □ YES
Target protocol enrollment	550 (strains)	Date of first subject enrolled	mm/dd/yyyy 01/01/2016
Last reported enrollment	N/A	Actual enrollment to date *(this should not include screen failures)*	189 (strains)
Targeted last subject last visit	550 (strains)	Actual last subject last visit	N/A
Do you have current drug supply sufficient to complete the study? *(If NO, please complete the Drug Section on Page 3)*			□ NO □ YES
Is this protocol closed to enrollment? *(patients may still be receiving therapy)*			☒ NO □ YES
Targeted study completion date *(primary objectives met; patient therapy and final study analysis complete)*			mm/dd/yyyy 12/31/2018
Actual study completion date *(if applicable)*			12/31/2018
Targeted date to provide results to Pfizer 6 months following the end of the study			30/06/2019

PUBLICATION INFORMATION

Do you plan to publish? *(If YES, please complete the information below.)* □ NO ☒ YES

Please be aware that, according to the IIR agreement, the investigator is required to provide Pfizer with an opportunity to prospectively review any proposed publication, abstract or other type of disclosure that reports the results of the study.

FORMAT	**PUBLICATION** *(please include anticipated journal or audience)*	**PLANNED**	**ACTUAL**	**SUBMISSION DATE**

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist PHONE IIR Grant Specialist FAX	[redacted]

Abstract	10th International Symposium on Pneumococci and Pneumococcal Diseases (ISPPD) Glasgow, Scotland 26 to 30 June 2016 (Poster)	☐	☒	
Manuscript	Vaccine/PloOne	☒	☐	
Poster	CACMID	☒	☐	05/03/2017
Other		☐	☐	

SIGNATURE

[redacted]

NAME Brigitte Lefebvre

DATE 04/26/2016

SIGNATURE (ONLY if faxed)

Pfizer	STUDY STATUS UPDATE FORM: CLINICAL	

IIR Grant Specialist IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist PHONE IIR Grant Specialist FAX	[redacted]

DRUG SUPPLY INFORMATION

SUPPLY CURRENTLY ON SITE	ACTIVE	PLACEBO
ESTIMATED REMAINDER REQUIRED TO COMPLETE STUDY	ACTIVE	PLACEBO
CAN PHARMACY ACCOMODATE TOTAL REMAINDER?	□ YES	□ NO

PERSONNEL INFORMATION

	PRINCIPAL INVESTIGATOR	COORDINATOR
NAME	Lefebvre Brigitte	
INSTITUTION	INSPQ/LSPQ	
MAILING ADDRESS	20045, chemin Sainte Marie, sainte-Anne-de-Bellevue, H9X 3R5, Québec, Canada	
TELEPHONE	514 4572070# 2334	
FAX	514 4576346	
EMAIL	Brigitte.Lefebvre@inspq.qc.ca	

	PHARMACIST	OTHER *(specify in additional comments)*
NAME		
INSTITUTION		
MAILING ADDRESS		
TELEPHONE		
FAX		
EMAIL		

ADDITIONAL COMMENTS	

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist	[redacted]	IIR Grant Specialist PHONE	[redacted]
IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist FAX	[redacted]

PLEASE COMPLETE AND RETURN BY: January 19, 2017

Per contractual requirements, we are requesting a status update on your IIR study supported by Pfizer via funding and/or drug. Please answer the following questions regarding the above referenced study by the due date. Answers from your last submitted update have been incorporated below; please update as needed and answer the remaining questions.

GENERAL INFORMATION

Pfizer Tracking #	**WI197603**	**Institutional Protocol #**	
Principal Investigator	**Dr. Brigitte Lefebvre**		
Study Title	**Molecular tools for serotyping for Streptococcus pneumoniae invasive strains surveillance in the province of Quebec.**		

STUDY UPDATE INFORMATION

Has this study been initiated?	□ NO ☒ YES	Date of initiation	mm/dd/yyyy 2016-01-01
Has the protocol been amended since last update?	☒ NO □ YES *(If YES, please provide the revised protocol)*		
Current IRB/IEC approval/renewal expires on **November 5, 2017**	If this is not current, please forward the most recent letter		
Have there been any personnel changes? *(If YES, please provide name and full contact info on Page 3)*			☒ NO □ YES
Target protocol enrollment	N/A	Date of first subject enrolled	
Last reported enrollment	N/A	Actual enrollment to date *(this should not include screen failures)*	
Targeted last subject last visit	N/A	Actual last subject last visit	
Do you have current drug supply sufficient to complete the study? *(If NO, please complete the Drug Section on Page 3) LSPQ : N/A, no drug use in this project*			□ NO □ YES
Is this protocol closed to enrollment? *(patients may still be receiving therapy)* *LSPQ : N/A this is a project on methods's development for which no enrollment is required.*			□ NO □ YES
Targeted study completion date *(primary objectives met; patient therapy and final study analysis complete)*			02/10/2017 mm/dd/yyyy
Actual study completion date *(if applicable)*			N/A mm/dd/yyyy
Targeted date to provide results to Pfizer			02/28/2017 mm/dd/yyyy

PUBLICATION INFORMATION

Do you plan to publish? *(If YES, please complete the information below.)*	□ NO ☒ YES

Please be aware that, according to the IIR agreement, the investigator is required to provide Pfizer with an

Pfizer	**STUDY STATUS UPDATE FORM: CLINICAL**	

IIR Grant Specialist	Pina Mustillo	IIR Grant Specialist PHONE	514-426-7021
IIR Grant Specialist EMAIL	pina.mustillo@pfizer.com	IIR Grant Specialist FAX	514-693-4715

opportunity to prospectively review any proposed publication, abstract or other type of disclosure that reports the results of the study.

FORMAT	**PUBLICATION** *(please include anticipated journal or audience)*	**PLANNED**	**ACTUAL**	**SUBMISSION DATE** **mm/dd/yyyy**
Abstract	IUMS or EMBO	□	☒	01-20-2017
Manuscript	PlosOne	☒	□	08-01-2017
Poster		□	□	
Other		□	□	

SIGNATURE

NAME Brigitte Lefebvre

DATE 01/13/2017

[redacted]

SIGNATURE (ONLY if faxed)

Pfizer	**STUDY STATUS UPDATE FORM: CLINICAL**	

IIR Grant Specialist	[redacted]	IIR Grant Specialist PHONE	[redacted]
IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist FAX	[redacted]

DRUG SUPPLY INFORMATION

SUPPLY CURRENTLY ON SITE	**ACTIVE**	**PLACEBO**
ESTIMATED REMAINDER REQUIRED TO COMPLETE STUDY	**ACTIVE**	**PLACEBO**
CAN PHARMACY ACCOMODATE TOTAL REMAINDER?	□ **YES**	□ **NO**

PERSONNEL INFORMATION

	PRINCIPAL INVESTIGATOR	**COORDINATOR**
NAME		
INSTITUTION		
MAILING ADDRESS		
TELEPHONE		
FAX		
EMAIL		

	PHARMACIST	**OTHER** *(specify in additional comments)*
NAME		
INSTITUTION		
MAILING ADDRESS		
TELEPHONE		
FAX		
EMAIL		

ADDITIONAL COMMENTS	

IUMS 2017
SINGAPORE
International Union of
Microbiological Societies

Abstract Submission

Bacteriology and Applied Microbiology
Applied Microbiology and Biotechnology

IUMS2017-1079
SEROTYPING OF STREPTOCOCCUS PNEUMONIAE INVASIVE STRAINS USING MOLECULAR BIOLOGY TOOLS

Florian Mauffrey* [1], Éric Fournier[1], Christine Martineau[1], Simon Lévesque[1], Walter Demczuk[2], Irene Martin[2], Florence Doualla-Bell[1], Cécile Tremblay[3], Jean Longtin[1], Brigitte Lefebvre[1]

[1]Institut Nationale de Santé Publique du Québec, Sainte-Anne-De-Bellevue, [2]National Microbiology Laboratory, Winnipeg, [3]Université de Montréal, Montréal, Canada

<br/>Would you like to apply for a travel grant?: Yes

Objectives: *Streptococcus pneumoniae* is a major cause of pneumonia, meningitis and other pneumococcal infections among young children and elders. Pneumococcal Conjugate Vaccines (PCVs) protect the population from the most prevalent serotypes of *S. pneumoniae*. From a public health perspective, accurate serotyping of *S. pneumoniae* is essential to monitor the serotype replacement following the introduction of PCVs. Although the Quellung reaction is the gold standard test for *S. pneumoniae* serotyping, this method is costly, time-consuming and dependent on human interpretation. The purpose of this study was to test and evaluate the efficiency of three different molecular serotyping methods as an alternative to the Quellung method.

Methods: The performance of a sequential multiplex PCR assay from the Centers for Disease Control and Prevention, a sequence typing assay (sequetyping) developed by Leung *et al.* (2012) based on the sequence of the *cpsB* gene within the pneumococcal capsular locus, and the whole genome sequencing (WGS) using Illumina MiSeq system were compared using 121 strains of *S. pneumoniae* previously serotyped by the Quellung method. The NCBI GenBank database was used to perform the sequetyping method. To assess WGS-based serotyping we adopted two different approaches: an in-house assembly/Blast strategy; and the PneumoCaT bioinformatics tool that uses read alignments rather than assemblies. All the 121 strains representing 83 different serotypes were serotyped by sequential multiplex PCR and sequetyping while 53 strains representing 32 serotypes were tested by WGS.

Results: The sequential multiplex PCR assay successfully identified 66% of the isolates at the serogroup or subset (cluster of serotypes from different serogroups) level while 34% was identified at the serotype level. A large proportion (23%) of strains was not typeable by the PCR assay. The WGS method exhibited the best performance with 91% of the isolates unambiguously identified at the serotype (66%) or serogroup level (25%) when using the in-house strategy. Ambiguous (6%) and misidentified (3%) results were low with WGS. PneumoCaT results revealed several misidentifications inside serogroups (21%). Interestingly, *S. pneumoniae* serotype 22F was correctly identified using PneumoCaT while our in-house strategy allowed for the identification of the serogroup only. One *S. pneumoniae* serotype 29 isolate was misidentified by both WGS analysis strategies, revealing divergences in serotype 29 sequences. Sequetyping was the method exhibiting the most misidentified serotypes (20%) and ambiguous results (15%). Moreover, even though 50% of serotypes were correctly identified, the second best High Scoring segment Pair (HSP) had often only 1 or 2 mismatches with the best HSP due to intra-specific variations in *cpsB* gene.

Conclusion: The proportion of serotypes identified using sequential multiplex PCR to the serotype level was too low to use as an alternative to the Quellung method. Although the sequetyping is currently the most economical method, it exhibited a high number of misidentified serotypes (20%). The WGS-based serotyping methods exhibited the best performance as they predicted capsular types at serotype and serogroup levels for 91% (66% at the serotype level) of the strains tested with only one misidentified serotype. WGS could be considered as a potent tool for *S. pneumoniae* serotyping and useful for epidemiological purposes.

Disclosure of Interest: F. Mauffrey: None Declared, É. Fournier: None Declared, C. Martineau: None Declared, S. Lévesque: None Declared, W. Demczuk: None Declared, I. Martin: None Declared, F. Doualla-Bell: None Declared, C. Tremblay: None Declared, J. Longtin: None Declared, B. Lefebvre Grant/ Research support from: Pfizer

Keywords: S. pneumoniae, Serotyping

New molecular tools for the serotyping of *Streptococcus pneumoniae* invasive strains in the province of Quebec – Part 2

Principal Investigator, Project leader

Brigitte Lefebvre, Ph.D., Microbiologist, LSPQ

Co-Principal Investigator

Cécile Tremblay, MD, FRCPC, Département de Microbiologie, immunologie et Infectiologie
Université de Montréal

Co-Investigators

Simon Lévesque, Ph.D., Microbiologist, LSPQ

Sadjia Bekal, Ph.D., Microbiologist, LSPQ

Marc-Christian Domingo, Ph.D., Microbiologist, LSPQ

Technical and Bioinformatics leader

Eric Fournier, D.E.S.S. bioinformatics, M.Sc., Bioinformatics Scientist, LSPQ

Florian Mauffrey, Ph.D., Microbiologist, Post-doctoral fellow, LSPQ

Christine Martineau, Ph.D., Microbiologist, LSPQ

Scientific Coordinator

Florence Doualla-Bell, Ph.D., LSPQ

Scientific Director

Jean Longtin, MD, LSPQ

Authors

Florian Mauffrey, Ph.D., Microbiologist, Post-doctoral fellow, LSPQ

Brigitte Lefebvre, Ph.D., Microbiologist, LSPQ

Introduction

Part one of the project developed and tested three molecular serotyping methods for *Streptococcus pneumoniae.* The present report is the second phase of the project which focused on the confirmation of specificity and sensitivity with an increased number of serotypes and confounding strains. Specific specimens were tested in order to answer some issues encountered in Part 1 (serotype 35A and 34 for multiplex PCR and serotype 29 for sequetyping). Non-*S. pneumoniae* (*S. pseudopneumoniae* and *S. mitis*) were also included in the study as controls for the specificity of multiplex PCR and sequetyping.

All testing was performed in standard reference laboratory conditions. This allowed for an accurate evaluation of the cost and time required for each method in order to obtain results. This is particularly important for the evaluation of the multiplex PCR method since several steps are required to identify serotypes and the number of steps differs depending on the serotype.

Material and methods

Methodology was extensively described in Part 1 of this report. Please refer to Part 1 for details.

Bacterial isolates

Ninety-four isolates of *Streptococcus pneumoniae* were used in this part of the study (Table 1). They include 49 different serotypes previously identified by the Quellung reaction using Statens Serum Institute antisera, 9 of which were not tested in Part 1. Strains with rare serotypes (n=13) were provided by the National Microbiology Laboratory (NML, Winnipeg). Thus, the full report covers up to 83 different serotypes (more than 90 serotypes described to date for *S. pneumoniae*) for 2 out of the 3 methods tested. Ten serotypes (9, 10C, 11D, 12B, 16A, 19B, 19C, 25A, 33C and 33D) were not tested in this study due to lack of availability at the LSPQ and the NML. The specificity of the multiplex PCR and sequetyping methods was also evaluated with three strains of *S. pseudopneumoniae* and 3 strains of *S. mitis*

Whole Genome Sequencing (WGS)

Whole genome sequencing was performed on 32 pneumococci isolates (Table 1) using the Illumina MiSeq system and Nextera XT DNA reagent kit v3 (600 cycles, paired ends). The 32 isolates were sequenced in a single batch; therefore lower coverage per isolate was obtained. Nevertheless, coverage was adequate for serotyping according to MiSeq Sequencing Coverage Calculator (http://support.illumina.com/downloads/sequencing_coverage_calculator.html).

An average theoretical coverage of 70X should be obtained with 32 isolates and a minimum coverage of 35X is considered standard for detecting single-nucleotide variants (Sims *et al.*, 2014). All best High Scoring segment Pairs (HSP) with a similar length and a similar nucleotide identity (< 0.5%) were considered for serotype identification.

Bioinformatics tools

For isolate MA080904, in-house python scripts were used to remove contigs with excessive coverage in order to calculate relevant metrics. Metrics were then computed with Quast (Gurevich *et al*., 2013).

For isolate LSPQ4282, in-house Biopython (http://biopython.org/) scripts removed non *S. pneumoniae* contigs from the assembled sequences fasta file. Metrics computation and Blast were performed before and after removing contigs.

For the other strains, bioinformatic analyses were performed as described in Part 1 of the project.

PneumoCaT, a bioinformatics workflow designed for *S. pneumoniae* serotype identification, which did not rely on assembled contigs, was also tested against our own pipeline (Kapatai *et al.*, 2016). Reads were directly mapped against a *cps* gene sequences database. When sequences of the same serotype had a high reads coverage (> 90%), this serotype was attributed to the isolate. When several sequences with a high coverage (> 90%) belonged to the same genogroup, a deeper analysis allowed the discrimination of the correct serotype (SNPs, alleles, presence of genes). When no sequence had enough coverage, no serotype was attributed and the flag "failed" was attributed.

Sequential multiplex PCR

PCRs were performed with the sequential reactions on 77 strains (CDC protocol). This was done to ensure that correct serotypes were detected at the expected multiplex PCR and to verify the presence of non-specific reactions in the other multiplex PCR. When correct amplification occurred, isolates were discarded and not tested for the following multiplex PCR as would occur routinely. Identification levels were defined as 1) **Serotype** when the correct serotype was determined, 2) **Serogroup** when several serotypes belonging to the correct serogroup were determined, 3) **Ambiguous** when several serotypes belonging to different serogroups but including the correct serotype were determined and 4) **Misidentified** when a wrong serotype was attributed.

Isolates with serotypes not detectable according to the CDC sequential multiplex PCR protocol were tested with this method to confirm the presence/absence of non-specific reactions.

Amplification issues were identified in Part 1 for serotype 35A (no amplification) and serotype 34 (non-specific amplicons). To confirm that these results were due to the PCR protocol and not genetic variants, 5 isolates of serotype 35A and serotype 34 were tested for the multiplex PCR reaction 7 (positive amplification expected).

Sequetyping

Sequetyping was performed on 54 isolates. Because we encountered some technical issues on isolates of serotype 29 during part 1 of the project, (absence of amplification of the *cpsB* gene) we sequenced five isolates of serotype 29.

Table 1 Serotypes and isolates ID used in this study and selected isolates for the serotyping molecular methods tested.

		Tested serotyping methods		
Serotypes(1)	Isolates ID	WGS	Sequetyping	Sequential multiplex PCR
1	MA096520		✓	✓
1	MA101323		✓	✓
3	MA080904	✓		
3	MA081716	✓		
3	MA082307	✓		
3	MA086676	✓		
3	MA096946	✓		
3	MA100130		✓	✓
3	MA101386		✓	✓
4	MA079938	✓		
4	MA100773		✓	✓
4	MA101744		✓	✓
5	MA082483		✓	✓
6A	MA099472		✓	✓
6A	MA101024		✓	✓
6B	MA098599		✓	✓
6B	MA101145		✓	✓
6C	MA099139		✓	✓
6C	MA100925		✓	✓
7F	MA093680		✓	✓
7F	MA097140		✓	✓
9L	LSPQ4271	✓	✓	✓
9N	MA080879	✓		
9N	MA081113	✓		
9N	MA098250		✓	✓
9N	MA100245		✓	✓
9V	MA097827		✓	✓
9V	MA098806		✓	✓
10B	MA080812		✓	✓
11B	MA096566		✓	✓
11C	LSPQ4272	✓	✓	✓
11F	MA073130			✓
14	MA096954		✓	✓
14	MA098680		✓	✓
15A	MA080018	✓		
15A	MA100658		✓	✓
15A	MA101766		✓	✓

Table 1 (continued)

		Tested serotyping methods		
Serotypes[1]	Isolates ID	WGS	Sequetyping	Sequential multiplex PCR
16F	MA065427			✓
17A	LSPQ4273	✓	✓	✓
18C	MA093772		✓	✓
18C	MA099660		✓	✓
18F	LSPQ4274	✓	✓	✓
19A	MA083920	✓		
19A	MA097921	✓		
19A	MA098817	✓		
19A	MA100706		✓[3]	✓
19A	MA101978		✓	✓
19A	MA083042			✓
19A	MA084138			✓
19F	MA100764		✓	✓
19F	MA101680		✓	✓
22F	MA080654	✓		
22F	MA100780		✓	✓
22F	MA101987		✓	✓
23A	MA082395	✓		
23F	MA100152		✓	✓
23F	MA101159		✓	✓
24A	LSPQ4275	✓	✓	✓
25F	LSPQ4276	✓	✓[3]	✓
27	MA088547		✓	✓
28A	MA099752		✓	✓
28F	LSPQ4277	✓	✓	✓
29	LSPQ3079			✓
29	MA097586	✓	✓	
29	MA098344		✓	
29	MA098505		✓	
29	MA100224		✓	
29	MA101320		✓	
32A	LSPQ4278	✓	✓	✓
32F	LSPQ3081			✓
33B	LSPQ4279	✓	✓	✓
33F	MA080211	✓		
34	MA101496			✓
34	MA101843			✓

Table 1 (continued)

		Tested serotyping methods		
Serotypes[1]	Isolates ID	WGS	Sequetyping	Sequential multiplex PCR
34	MA102076			✓
34	MA102374			✓
34	MA102487			✓
35A	LSPQ4266			✓
35A	LSPQ4267			✓
35A	LSPQ4268			✓
35A	LSPQ4269			✓
35A	LSPQ4270			✓
35A	MA101545	✓		
35B	MA082394	✓		
35C	LSPQ4280	✓	✓	✓
35F	MA081892	✓		
36	LSPQ3641			✓
41A	LSPQ3089			✓
41F	LSPQ4281	✓	✓	✓
43	LSPQ3643			✓
47A	LSPQ4282	✓	✓	✓
47F	LSPQ4283	✓	✓	✓
45	LSPQ3092			✓
48	LSPQ3095			✓
S1 *S. pseudopneumoniae*[2]	ID111828		✓	✓
S2 *S. pseudopneumoniae*[2]	ID112065		✓[3]	✓
S3 *S. mitis*[2]	ID112476		✓[3]	✓
S4 *S. pseudopneumoniae*[2]	ID112502		✓[3]	✓
S5 *S. mitis*[2]	MA084074		✓[3]	✓
S6 *S. mitis*[2]	MA084310		✓[3]	✓

[1] Serotype determined by Quellung.

[2] Non-*S. pneumoniae* used as controls

[3] Strains tested by sequetyping with no amplification of *cpsB* observed.

Results

Evaluation of the Whole Genome Sequencing approach

Paired-end reads quality

FastQC was used to compute and summarize reads statistics. The numbers of reads obtained for each isolate are shown in Table 2. A total number of 27 540 000 reads were obtained in this batch, representing 18% fewer reads than the batch in Part 1 of the project. This decrease can be explained by a lower clustering (597 k/mm^2) during the sequencing procedure. However, a lower clustering leads to better quality of reads as it allows a better resolution. Isolates' reads numbers varied between 100 065 and 884 691 with an average of 418 854. This number was particularly low for isolates LSPQ4271, LSPQ4272, LSPQ4273 and LSPQ4274. This may be explained by the lower concentration of DNA in the DNA extracts of these isolates probably caused by a less effective DNA extraction for these strains. Whereas these values appear to be very low, the assembling metrics are more reflective of the sequencing quality. Assembly metrics are compiled in table 3.

The metrics values indicate a high quality of assembling and sequencing, except for the lowest coverage value of 14X. These values are similar to those obtained in Part 1 and sometimes even higher. As emphasized in Part 1, metrics appear to correlate with the number of reads obtained for each isolate. For example, the lowest coverage value (14X) is attributed to LSPQ4271 and LSPQ4272 which also have the fewest number of reads. However, this value was high enough to perform Blast analysis.

MA080904 exhibited a coverage value of 1012X, which is an average of the coverage of all contigs. This average does not, however, consider the length of each contig. Manual analysis revealed that the majority of contigs presenting with a coverage value above 1 000X, were smaller than 5 000 bp. Thus, for this isolate, the coverage value was not representative of the real genome coverage. After discarding contigs with coverage value above 1 000X, the new coverage value was 37X with a loss in assembly length of only 3% (2 123 274 bp to 2 061 860 bp). All these metrics are more representative of the average and are grouped in the row MA080904-1 of table 3.

LSPQ4282 had an assembly length of 6 793 942 bp, representing threefold the length of *S. pneumoniae* genome (2.16 Mbp). Manual Blast analysis revealed that a significant part of contigs corresponded to contamination with a non-*Streptococcus* bacterium, mostly *Bacillus subtilis*. *B. subtilis* genome size is 4.2 Mbp, which explained the assembly's length of 6 793 942 bp because it is nearly the sum of both genome sizes. In-house Python script allowed us to discard non-*S. pneumoniae* contigs (based on Blast results) and create a clean fasta file. This file was named LSPQ4282-1. After cleaning, assembly length dropped to a more regular value of 1 721 249 bp, proving the efficiency of the script.

Table 2 Paired end reads number generated during the MiSeq run.

Isolates	Reads numbers[1,2]
LSPQ4271	**117 140**
LSPQ4272	**100 065**
LSPQ4273	**155 277**
LSPQ4274	**229 176**
LSPQ4275	307 126
LSPQ4276	545 333
LSPQ4277	611 811
LSPQ4278	751 230
LSPQ4279	293 694
LSPQ4280	285 007
LSPQ4281	253 861
LSPQ4282	382 521
LSPQ4283	695 607
MA079938	256 005
MA080018	268 153
MA080211	475 948
MA080654	404 211
MA080879	438 695
MA080904	407 169
MA081113	591 387
MA081716	579 976
MA081892	494 610
MA082307	362 116
MA082394	295 082
MA082395	429 205
MA083920	729 967
MA086676	337 811
MA096946	283 092
MA097586	555 776
MA097921	884 691
MA098817	365 612
MA101545	515 979
Total reads	**13 403 333**

[1] The total number (forward + reverse) for one isolate is two times the displayed value.

[2] The four lowest values are presented in bold and correspond to samples with lowest DNA concentration.

Table 3 Summary of Spades assembly's metrics[1][2].

Isolates	Assembly's length (bp)	Largest contig (bp)	N50	Mean coverage (X)
LSPQ4271	2 074 016	161 368	**49 807**	**14**
LSPQ4272	2 048 913	136 078	58 756	**14**
LSPQ4273	2 104 968	305 746	98 305	17
LSPQ4274	2 039 481	197 877	109 713	22
LSPQ4275	2 069 490	192 635	78 434	44
LSPQ4276	2 076 440	197 175	50 116	82
LSPQ4277	2 061 209	239 939	90 168	76
LSPQ4278	2 111 029	**105 571**	55 533	115
LSPQ4279	2 075 758	247 620	68 772	45
LSPQ4280	**2 143 572**	230 492	75 491	35
LSPQ4281	2 044 177	158 243	72 183	52
LSPQ4282	**6 793 942**	**438 741**	**95 131**	**32**
LSPQ4282-1	**1 721 249**	**146 711**	**49 041**	**6**
LSPQ4283	2 076 447	171 602	71 324	103
MA079938	2 108 330	214 530	74 514	45
MA080018	2 102 343	247 306	95 807	45
MA080211	2 054 380	246 678	140 406	83
MA080654	2 069 755	297 023	104 357	92
MA080879	2 103 519	345 799	136 064	85
MA080904	**2 123 274**	**161 387**	**64 114**	**1012**
MA080904-1	**2 061 860**	**161 387**	**70 238**	**37**
MA081113	2 066 217	276 495	85 471	97
MA081716	2 013 057	345 480	**218 480**	204
MA081892	2 043 092	299 061	126 588	104
MA082307	2 013 998	276 730	167 190	127
MA082394	2 063 773	202 017	101 286	58
MA082395	2 050 026	273 953	113 480	93
MA083920	2 066 049	328 634	86 181	131
MA086676	**1 987 104**	243 817	91 651	67
MA096946	2 035 090	263 351	136 846	58
MA097586	2 063 487	196 889	61 494	113
MA097921	2 129 092	355 253	162 090	171
MA098817	2 093 975	**381 909**	163 676	89
MA101545	2 075 501	286 061	162 953	**296**

[1] All statistics are based on contigs with a length ≥ 500 bp.

[2] Numbers in green and red indicate the highest and lowest values, respectively. Isolates in bold are not included in this count because they were treated differently.

All metrics for this sample were below the average, demonstrating that cleaning the file caused a decrease in assembly quality. This is relevant as two thirds of the sample was constituted in *B. subtilis* sequences, due to a bigger genome. Both LSPQ4282 and LSPQ4282-1 were subjected to serotyping determination in order to see if an external contamination can affect the serotype results.

Serotype determination using Blast queries

Blast searches were performed as previously described (Part 1). Serotype identification was mainly based on best score High-scoring Segment Pairs (HSP). HSP length and identities are also reported as supplementary information (Table 4). When multiple hits had high identity value (<0.5% compared to best hit) and HSP length (>10000 bp), all were retained for serotype attribution.

In 29 of 32 cases (90.6%), serotype was correctly determined with no ambiguity. All isolates except serotype 3 demonstrated a HSP length higher than 15 000 bp. The lower HSP length obtained for serotype 3 HSP can be explained by the smaller *cps* locus length in these isolates, which is the smallest *cps* locus of all serotypes (Bentley *et al.*, 2006). However, HSP identity was above 98% in every case.

MA101545 (serotype 35A) identification was classified as ambiguous due to the presence of 3 high score HSP including a serotype 35A HSP. Although the identity of this HSP is the highest among the 3 HSP, it cannot be chosen as a criterion of selection because the HSP length of serotype 35A is not the higher value among the 3 results obtained. MA101545 was the only isolate with this feature and more results are needed in order to draw conclusions about the use of HSP identity as the selection criterion in such cases.

MA080654 (serotype 22F) was identified at the serogroup level, with HSP for serotype 22F and 22A showing an identical score and identity value. Two different HSP with high score value were found for both serotypes (Part 1).

MA097586 (serotype 29) was the only isolate presenting a misidentification. A high identity value was obtained for serotype 35B but with a HSP length of only 10 656 bp, far below the usual length of correct HSP (above 15 000 bp; except for serotype 3). Serotype 35B and 29 are known to be genetically related, leading to cross-reactivity in antisera reactions (Bush *et al.*, 2015). Surprisingly, no significant hit with serotype 29 was found in Blast searches results, meaning that no relevant alignment could be made. Thus, serotype 29 *cps* sequence was manually blasted against MA097586 assembly. (Figure 1). The alignment resulted in 2 small HSP with low identity separated by a 2 800 bp gap demonstrating very low concordance between the two sequences. These results correlate with sequetyping results obtained for serotype 29. This strongly suggests that these issues are due to a lack of serotype 29 sequences available in public databases. WGS Blast results are based on a pool of 107 *cps* locus sequence with a unique serotype 29 sequence (*S. pneumoniae* strain 34373, Bentley *et al.*, 2006). Serotype 29 *cps* sequence diversity could be higher than other serotypes and the addition of more sequences should resolve this issue. A potential solution could be to isolate the *cps* loci obtained in this study and include them in the local WGS *cps* database.

Table 4 Pneumococcal serotype identification using Whole Genome Sequencing and Blast Queries.

Isolates	Query contigs length (bp)	*cps* best hit subject			HSP[1]		Expected serotype[2]	Identification level
		GenBank accession	Serotype	Length (bp)	Identity (%)	Length (bp)		
LSPQ4271	72 399	CR931646	9L	17 618	99.98	15 948	9L	Serotype
		CR931647	9N	17 619	99.59	15 948		
LSPQ4272	95 187	CR931655	11C	18 532	99.99	15 635	11C	Serotype
		CR931654	11B	17 082	99.2	14 934		
LSPQ4273	305 746	CR931669	17A	23 198	98.45	22 930	17A	Serotype
LSPQ4274	52 774	CR931674	18F	22 849	100.0	21 674	18F	Serotype
		CR931673	18C	21 819	97.49	12 646		
LSPQ4275	51 091	CR931686	24A	21 907	99.96	20 176	24A	Serotype
		CR931688	24F	24 165	98.28	13 289		
LSPQ4276	17 962	CR931690	25F	28 389	99.99	17 962	25F	Serogroup
		CR931689	25A	28 466	99.96	17 962		
LSPQ4277	217 512	CR931693	28F	21 839	99.98	21 835	28F	Serotype
		CR931692	28A	22 978	99.09	20 660		
LSPQ4278	57 070	CR931696	32A	25 372	99.99	19 792	32A	Serogroup
		CR931697	32F	25 363	99.96	19 792		
LSPQ4279	247 620	CR931699	33B	19 039	99.82	17 417	33B	Serotype
		CR931701	33D	17 583	98.2	10 380		
LSPQ4280	49 489	CR931706	35C	19 741	99.99	18 532	35C	Ambiguous
		CR931715	42	19 403	99.89	18 325		
LSPQ4281	89 373	CR931714	41F	22 917	99.73	22 919	41F	Serotype
		CR931713	41A	22 520	97.13	19 367		
LSPQ4282	17 973	CR931720	47A	17 250	100.0	16 052	47A	Serotype
LSPQ4283	32 343	CR931721	47F	16 064	99.99	15 105	47F	Serotype
MA079938	17 652	CR931635	4	20 936	99.98	17 652	4	Serotype
MA080018	140 192	CR931663	15A	18 517	99.75	18 517	15A	Serotype
		CR931666	15F	22 405	99.22	12 386		
MA080211	246 678	AJ006986	33F	17 340	99.98	16 435	33F	Serogroup
		CR931698	33A	18 409	99.98	16 107		

Table 4 (continued)

Isolates	Query contigs length (bp)	*cps* best hit subject			HSP[1]		Expected serotype[2]	Identification level
		GenBank accession	Serotype	Length (bp)	Identity (%)	Length (bp)		
MA080654	109 110	CR931681	22A	22 591	97.85	12 897	22F	Serogroup
		CR931681	22A	22 591	97.69	7721		
		CR931682	22F	22 696	97.85	12 897		
		CR931682	22F	22 696	97.69	7721		
MA080879	227 059	CR931647	9N	17 619	99.99	17 619	9N	Serotype
		CR931646	9L	17 618	99.26	17 620		
MA080904	118 789	CR931634	3	10 337	99.75	5 293	3	Serotype
		AF030373	23F	24 722	91.74	5 812		
MA081113	227 087	CR931647	9N	17 619	99.99	17 619	9N	Serotype
		CR931646	9L	17 618	99.26	17 620		
MA081716	345 480	CR931634	3	10 337	99.75	8 961	3	Serotype
MA081892	299 061	CR931707	35F	15 137	99.95	15 007	35F	Serotype
		CR931721	47F	16 064	99.46	6 666		
MA082307	276 730	CR931634	3	10 337	99.75	8 961	3	Serotype
MA082394	196 027	CR931705	35B	16 658	99.97	15 516	35B	Serotype
MA082395	273 953	CR931683	23A	21 475	99.98	21 475	23A	Serotype
		CR931685	23F	22 330	99.42	12 830		
MA083920	328 634	CR931675	19A	18 617	98.19	14 095	19A	Serotype
		AF094575	19A	18 754	98.14	14 095		
MA086676	67 898	CR931634	3	10 337	99.74	8 961	3	Serotype
MA096946	263 351	CR931634	3	10 337	99.75	8 961	3	Serotype
MA097586	114 025	CR931705	35B	16 658	99.95	10 656	29	Misidentified
		CR931714	41F	22 917	96.88	7 344		
MA097921	311 219	CR931675	19A	18 617	98.58	15 861	19A	Serotype
		AF094575	19A	18 754	98.51	15 745		
MA098817	381 909	CR931675	19A	18 617	98.58	15 861	19A	Serotype
		AF094575	19A	18 754	98.51	15 745		
MA101545	105 480	CR931706	35C	19 741	99.29	18 615	35A	Ambiguous
		CR931715	42	19 403	99.23	18 408		
		CR931704	35A	21 463	99.37	17 808		

[1] HSP = high-scoring Segment Pairs.

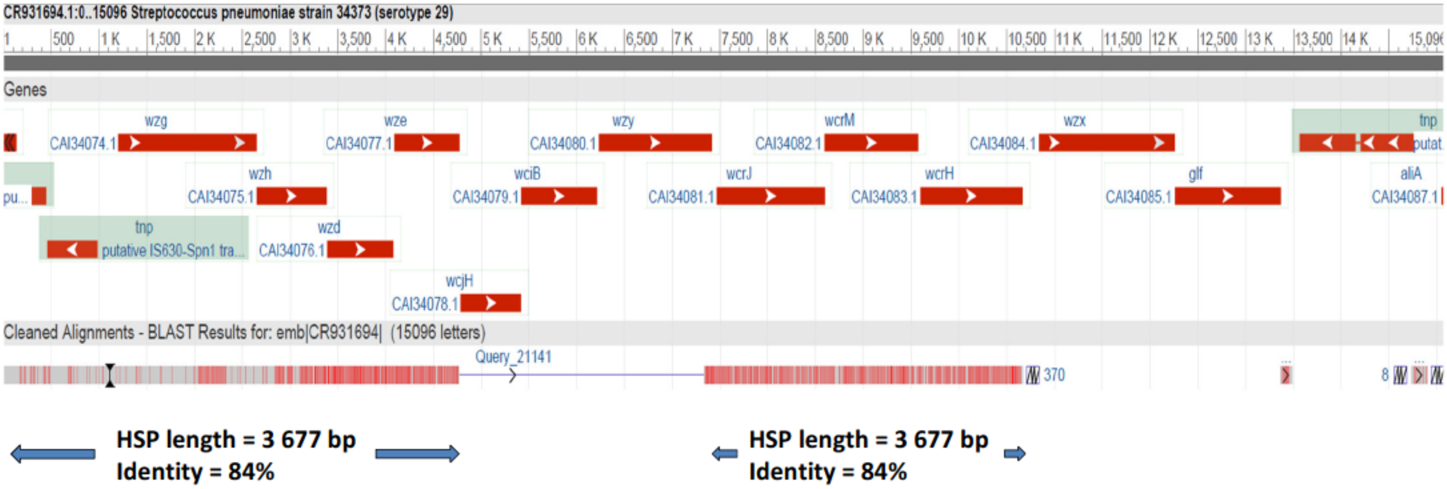


Figure 1 Blast analysis of the MA097586 assembly with serotype 29 *cps* sequence showing the presence of several unmatched regions. Two separated HSP with low identity highlight the very low complementarity of the 2 sequences.

In summary, WGS resulted in 78% (25/32) identification at the serotype level, 13% (4/32) identification at the serogroup level, 6% (2/32) identification classified as ambiguous and 3% (1/32) of misidentification. In Part 1, 52% of isolates were identified at the serotype level and 48% at the serogroup level. As described in Part 1, a high degree of genetic similarities due to DNA polymorphism among single serotypes made some serotype identification difficult, specifically LSPQ4276 (25F/25A), LSPQ4278 (32A/32F), LSPQ4280 (35C/42), MA080211 (33F/33A) and MA080654 (22F/22A). The project (Parts 1 and 2) demonstrates that isolates were identified at the serotype level at 68% (36/53), at the serogroup level at 26% (14/53), as ambiguous at 4% (2/53) and as misidentified at 2% (1/53). These 53 isolates represent 32 different serotypes, as several isolates for the same serotype were tested to ascertain the robustness of the method. The true efficiency of the method, when considering only one isolate per serotype, yielded results of: 66% (21/32) identified at the serotype level, 25% (8/32) identified at the serogroup level, 6% (2/32) identified as ambiguous and 3% (1/32) misidentified.

PneumoCaT (https://github.com/phe-bioinformatics/PneumoCaT), a bioinformatics workflow designed for the serotyping of *S. pneumoniae*, was also tested. Results are presented in Table 5. PneumoCaT analysis was done for all isolates of this project. Considering each of the isolates, 79 % (41/52) were identified at the serotype level and 21 % were misidentified (11/22). Identification failed for 1 isolate due to the absence of *cps* sequence with enough coverage. This isolate was not considered for the statistics.

Table 5 Pneumococcal serotypes identification using Whole Genome Sequencing and PneumoCaT workflow. The serotype determined after capsular typing variants analysis is presented in bold.

Isolate	Expected serotype	Hit 1 serotype	Coverage (%)	Hit 2 serotype	Coverage (%)	Identification level
MA086676	3	3	99.98	36	29.05	Serotype
MA096946	3	3	99.98	36	29.11	Serotype
MA081716	3	3	99.98	36	29.16	Serotype
MA080904	3	3	99.91	1	26.99	Serotype
MA082307	3	3	99.98	36	28.66	Serotype
MA079938	4	4	99.99	45	48.32	Serotype
MA097586	29	**35B**	99.99	34	46.81	Misidentified
MA096961	34	34	99.94	35B	50.31	Serotype
MA094205	10A	**10A**	99.99	10B	94.20	Serotype
MA095845	10A	**10A**	99.99	10B	94.31	Serotype
MA094933	10A	**10A**	99.99	10B	94.39	Serotype
MA091851	11A	**11D**	99.78	11A	99.75	Misidentified
LSPQ4272	11C	**11C**	96.06	11B	96.05	Serotype
MA094663	15A	15A	99.96	15F	80.57	Serotype
MA095336	15A	15A	99.99	15F	80.65	Serotype
MA096792	15A	15A	99.99	15F	80.66	Serotype
MA080018	15A	15A	99.95	15F	80.57	Serotype
MA093977	15A	15A	99.99	15F	80.61	Serotype
MA094560	15B	**15C**	99.99	15B	99.99	Misidentified
MA096033	15B	**15C**	99.99	15B	99.99	Misidentified
MA095997	15B	**15C**	99.99	15B	99.99	Misidentified
MA093020	16F	16F	99.99	28F	64.28	Serotype
LSPQ4273	17A	17A	97.84	41F	71.11	Serotype

Table 5 (continued)

Isolate	Expected serotype	Hit 1 serotype	Coverage (%)	Hit 2 serotype	Coverage (%)	Identification level
LSPQ4274	18F	18F	99.98	18B	86.91	Serotype
MA097921	19A	19A	93.45	23B	51.30	Serotype
MA098817	19A	19A	93.44	23B	50.85	Serotype
MA079789	19A	19A	99.99	6E	52.82	Serotype
MA080288	19A	19A	99.99	6E	53.64	Serotype
MA080125	19A	19A	99.99	6E	56.69	Serotype
MA083920	19A	19A	93.59	6E	46.08	Serotype
MA094696	22F	**22F**	99.98	22A	90.99	Serotype
MA096962	22F	**22F**	99.21	22A	90.70	Serotype
MA080654	22F	**22F**	99.24	22A	90.95	Serotype
MA094689	22F	**22F**	99.24	22A	90.94	Serotype
MA082395	23A	23A	99.99	23F	77.02	Serotype
LSPQ4275	24A	24A	99.82	24F	80.01	Serotype
MA096695	24B	**24F**	100.00	24B	100.00	Misidentified
LSPQ4276	25F	**25A**	99.99	25F	99.99	Misidentified
LSPQ4277	28F	**28F**	99.99	28A	99.99	Serotype
LSPQ4278	32A	**32F**	99.99	32A	99.99	Misidentified
LSPQ4279	33B	**33B**	99.99	33D	93.03	Serotype
MA080211	33F	**33F**	99.99	33A	96.45	Serotype
MA101545	35A	**35C**	99.99	42	99.99	Misidentified
MA082394	35B	35B	99.83	34	46.30	Serotype
LSPQ4280	35C	**35A**	99.94	35C	99.94	Misidentified

Table 5 (continued)

Isolate	Expected serotype	Hit 1 serotype	Coverage (%)	Hit 2 serotype	Coverage (%)	Identification level
MA081892	35F	35F	99.99	47F	79.42	Serotype
LSPQ4281	41F	**41F**	99.99	41A	98.02	Serotype
LSPQ4282	47A			Failed[1]		
LSPQ4283	47F	47F	99.99	35F	88.64	Serotype
MA081946	7F	**7F**	100.00	7A	99.99	Serotype
LSPQ4271	9L	**9N**	99.80	9L	99.80	Misidentified
MA081113	9N	**9N**	99.99	9L	99.99	Serotype
MA080879	9N	**9N**	99.99	9L	99.99	Serotype

[1] No *cps* sequence with coverage above 90% (in reads number).

In this second part of the project, the WGS run was performed with 32 samples and the quality of results was adequate for serotyping, despite a non-optimal clustering step and some DNA extract with low DNA concentration. Thus, it is realistic to estimate that 46 samples could be sequenced in a single run with optimal conditions and produce good results. This quantity of samples corresponds with the maximum number of samples in a single run of DNA extraction, which allows for cost optimization. Genome sequences obtained with this method could be used for further investigations (antibiotic resistance screening, sequence typing, etc.) which usually require other laboratory experiments such as PCR and thus would offset costs.

Given the results exposed in this report, it is clear that WGS alone is not sufficient for complete serotyping of isolates not identified at the serotype level. The Quellung method would be used to decide in such cases, adding additional costs to the method. However, as Blast results would serve as a guide for the use of antisera, the cost of Quellung method for these isolates would be dramatically lower than usual.

Two days are required for the preparation of the sequencing run from DNA extracts, considering an eight hour working day and following Illumina MiSeq sequencing protocol. Next, 3 days are necessary for the sequencing run itself. Downstream bio-informatics analyses will be automated with an in-house pipeline (1 working day). These analyses and the manual analyses of the output data take one day to perform and to obtain a final serotype result. Overall, 5-6 days are required for serotyping *S. pneumoniae* from DNA extracts with the Whole Genome Sequencing method. In the low *S. pneumoniae* season, it will take more time to obtain serotyping results than using Quellung because it is necessary to batch the strains before starting a WGS batch. However when WGS will be used routinely at the LSPQ for numerous bacteria, *S. pneumoniae* serotyping using WGS may be more cost-effective.

Evaluation of the multiplex PCR CDC protocol

A total of 77 different strains were tested in this part, representing 45 different serotypes. Several isolates with serotypes not included in the CDC protocol were tested in order to check for specificity. For these isolates, non-detection is considered a good result as they cannot be detected with this multiplex PCR protocol. Isolate serotype was determined: 30% (18/61) at the serotype level, 34% (21/61) at the serogroup level, 5% (3/61) at a subset level, 31% (19/61) not determined (expected results) and 0% (0/61) misidentified. Detailed results are listed in table 6.

Two different issues were highlighted in Part 1 of the project. The first one was the non-detection of the 280 bp amplicon expected with the 2 isolates of serotype 35A. It was suggested that the strains tested were genetic variants of the CDC strains of serotype 35A and that the primers 35A/35C/42 were unable to match our strains. In order to confirm the hypothesis, 5 new isolates of serotype 35A were tested at multiplex #7 PCR reaction (multiplex expected for positive reaction). All 5 isolates presented a positive amplification at 280 bp as expected by the protocol (Figure 2A). But a 250 bp nonspecific amplification was also observable in 4 of 5 isolates. The second issue was the presence of a 250 bp nonspecific amplification in 1 of the 2 isolates of serotype 34 tested. 5 new isolates of serotype 34 were tested as well at multiplex #7 PCR reaction. All isolates presented the expected amplification at 408 bp and 3 of 5 isolates also presented a nonspecific 250 bp amplification. Thus, it appears that a nonspecific amplification at 250 bp may occur at multiplex #7 PCR reaction. As it does not appear with all isolates, it seems that small genetic changes among these isolates could determine the presence or absence of this amplification. This nonspecific amplification was also present for serotype 42 (see report of part 1, Figure 17A)

A total of 121 strains were tested with multiplex PCR method in this project (Part 1 and Part 2), covering 83 serotypes. As expected, 16% (19/121) isolate serotypes were not determined because they are not included in the CDC multiplex PCR protocol. They will not be considered in the statistics in order to correctly evaluate method efficiency. Isolates were identified at the serotype level at 40% (41/102), at the serogroup level at 41% (42/102), at the subset level at 17% (17/102) and as misidentified at 2% (2/102). As described previously, several isolates with the same serotype were tested in order to evaluate the robustness of the method. Unfortunately, this does not reflect the true efficiency of the method. All results converge for the same serotype except for serotype 35A. However 5 out of 7 isolates were identified at the serotype level so it was considered that multiplex PCR was serotype-specific for this serotype. Considering only one isolate per serotype, results were: 34% (22/64) identified at the serotype level, 38% (24/64) identified at the serogroup level, 28% (18/64) identified at the subset level, and 0% (0/64) misidentified.

Table 6 Pneumococcal serotype identification with multiplex PCR method. Non-specific amplifications are reported for each isolate tested.

Isolates	Serotype	Expected multiplex amplification[1]	Determined serotype	Identification level	Presence of nonspecific amplicons			
					mPCR2 (≈500 bp)	mPCR3 (677 bp)	mPCR6 (850 bp)	mPCR7 (250 bp)
LSPQ3079	29	/	/	ND[2]	X		X	
LSPQ3081	32F	/	/	ND[2]	X	X		
LSPQ3089	41A	/	/	ND[2]				
LSPQ3092	45	/	/	ND[2]	X	X		X
LSPQ3095	48	/	/	ND[2]	X	X		
LSPQ3641	36	/	/	ND[2]	X	X		
LSPQ3643	43	/	/	ND[2]		X	X	
LSPQ4271	9L	7	9N/9L	Serogroup	X	X		
LSPQ4272	11C	/	/	ND[2]	X	X		
LSPQ4273	17A	/	/	ND[2]	X	X		
LSPQ4274	18F	4	18C/18F/18B/18A	Serogroup	X	X		
LSPQ4275	24A	4	24F/24A/24B	Serogroup	X			
LSPQ4276	25F	3	38/25A/25F	Subset				
LSPQ4277	28F	/	/	ND[2]	X		X	X
LSPQ4278	32A	/	/	ND[2]	X	X		
LSPQ4279	33B	/	/	ND[2]	X	X		X
LSPQ4280	35C	7	35A/35C/42	Subset[3]	X	X		
LSPQ4281	41F	/	/	ND[2]	X	X		
LSPQ4282	47A	/	/	ND[2]	X	X		
LSPQ4283	47F	6	35F/47F	Subset[3]	X	X		
MA065427	16F	1	16F	Serotype		X		
MA073130	11F	/	/	ND[2]				
MA080812	10B	/	/	ND[2]				

Table 6 (continued)

Isolates	Serotype	Expected multiplex amplification[1]	Determined serotype	Identification level	Presence of nonspecific amplicons			
					mPCR2 (≈500 bp)	mPCR3 (677 bp)	mPCR6 (850 bp)	mPCR7 (250 bp)
MA082483	5	6	5	Serotype		X		
MA083042	19A	/	19A	Serotype	X	X		
MA084138	19A	/	19A	Serotype	X			
MA088547	27	/	/	ND[2]		X		
MA093680	7F	2	7F/7A	Serogroup				
MA093772	18C	4	18C/18F/18B/18A	Serogroup		X		
MA096520	1	5	1	Serotype		X		
MA096566	11B	/	/	ND[2]				
MA096954	14	5	14	Serotype				
MA097140	7F	2	7F/7A	Serogroup				
MA097827	9V	4	9A/9V	Serogroup				
MA098250	9N	7	9N/9L	Serogroup				
MA098599	6B	1	6A/6B	Serogroup				
MA098680	14	5	14	Serotype				
MA098806	9V	4	9A/9V	Serogroup				
MA099139	6C	1	6C/6D	Serogroup				
MA099472	6A	1	6A/6B	Serogroup				
MA099660	18C	4	18A/18B/18C/18F	Serogroup		X		
MA099752	28A	/	/	ND[2]				
MA100130	3	1	3	Serotype				
MA100152	23F	5	23F	Serotype		X		
MA100245	9N	7	9N/9L	Serogroup				
MA100658	15A	2	15A/15F	Serogroup				

Table 6 (continued)

Isolates	Serotype	Expected multiplex amplification[1]	Determined serotype	Identification level	Presence of nonspecific amplicons			
					mPCR2 (≈500 bp)	mPCR3 (677 bp)	mPCR6 (850 bp)	mPCR7 (250 bp)
MA100706	19A	1	19A	Serotype				
MA100764	19F	3	19F	Serotype				
MA100773	4	4	4	Serotype		X		
MA100780	22F	1	22F/22A	Serogroup				
MA100925	6C	1	6C/6D	Serogroup				
MA101024	6A	1	6A/6B	Serogroup				
MA101145	6B	1	6A/6B	Serogroup				
MA101159	23F	5	23F	Serotype		X		
MA101323	1	5	1	Serotype		X		
MA101386	3	1	3	Serotype				
MA101680	19F	3	19F	Serotype				
MA101744	4	4	4	Serotype		X		
MA101766	15A	2	15A/15F	Serogroup				
MA101978	19A	1	19A	Serotype				
MA101987	22F	1	22F/22A	Serogroup				

[1] Multiplex number (#1 to #8) in the sequence where a positive amplification is expected for this isolate.

[2] Not detectable.

[3] Defined as correct results obtained with PCR multiplex primers detecting a subset.

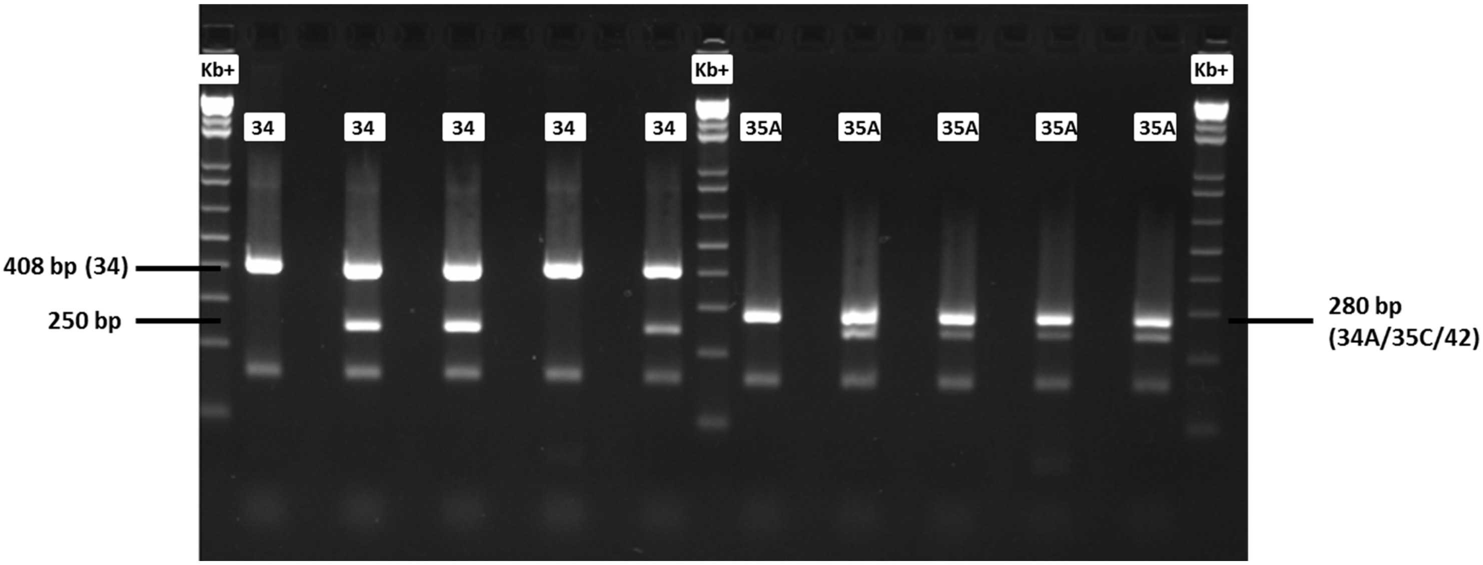


Figure 2A Electrophoresis profile obtained with the multiplex reaction 7 for the detection of serotype 34 and 35A. All reactions led to a positive amplification at the expected size. Nonspecific amplifications at 250 bp are visible for 7 of the 10 isolates tested.

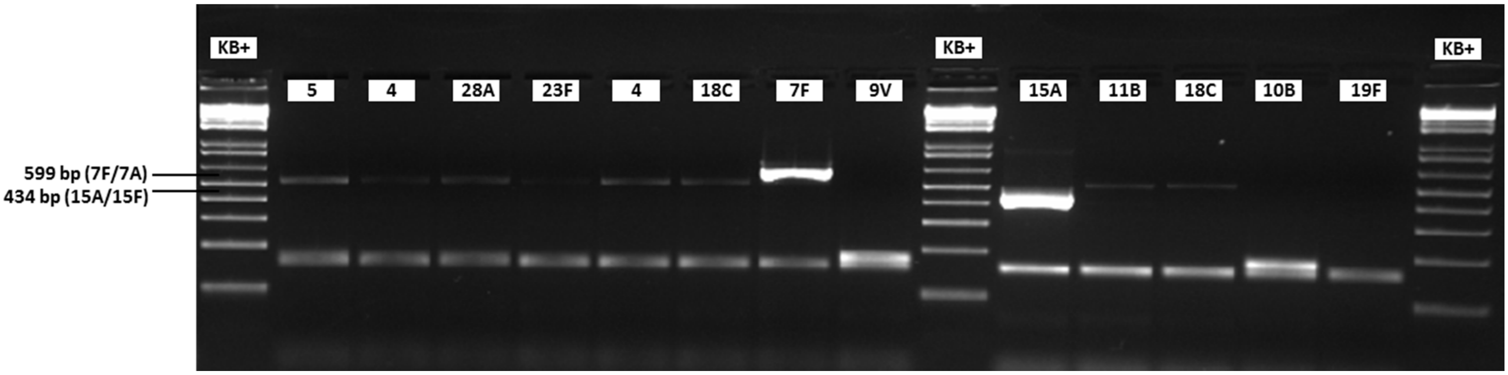


Figure 2B Electrophoresis profile obtained with the multiplex reaction 2. Nonspecific amplifications at 500 bp are visible but do not correspond with a specific primers pair. Bands corresponding to these amplifications also appear thinner and dimmer compared to correct amplifications (7F and 15A).

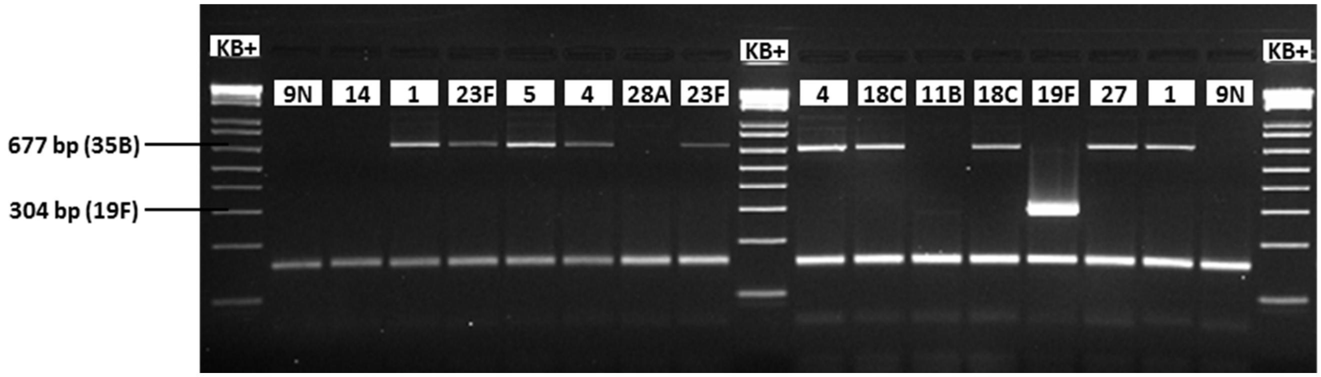


Figure 2C Electrophoresis profile obtained with the multiplex reaction 3. Nonspecific amplifications at 677 bp (35B) are visible. Bands corresponding to these amplifications appear thinner compared to a correct amplification (19F).

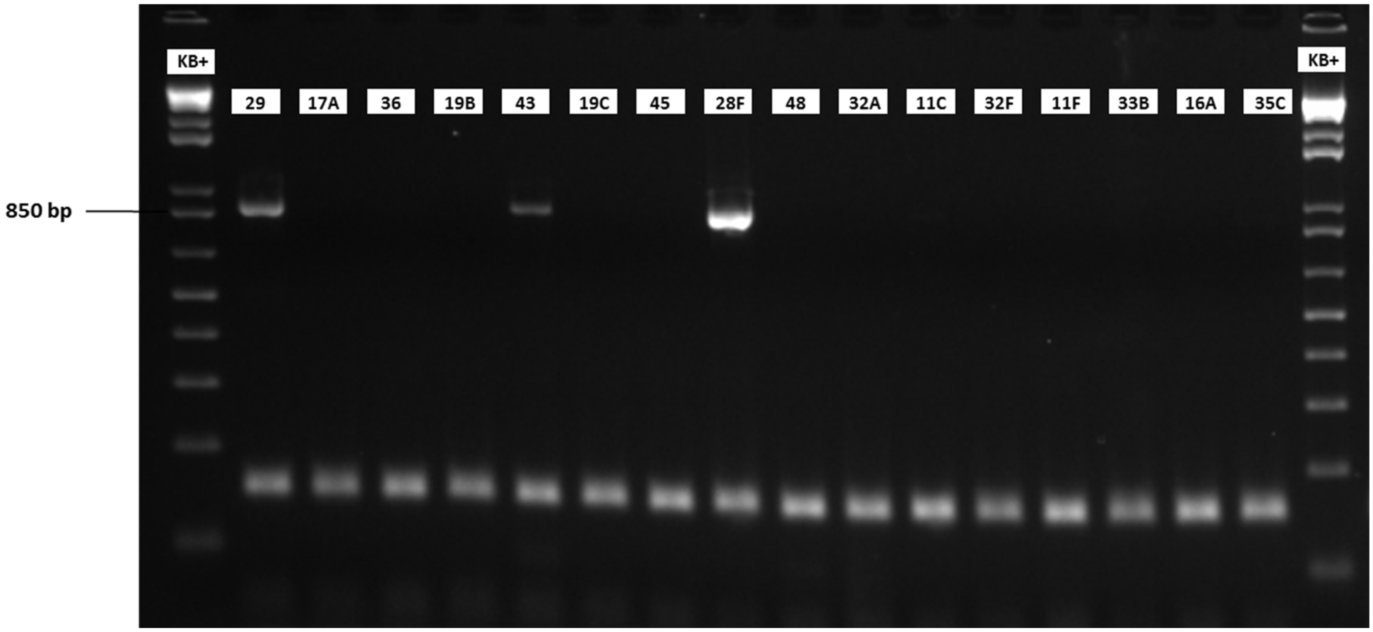


Figure 2D Electrophoresis profile obtained with the multiplex reaction 6. Nonspecific amplifications at 850 bp are visible but do not correspond to a specific primers pair.

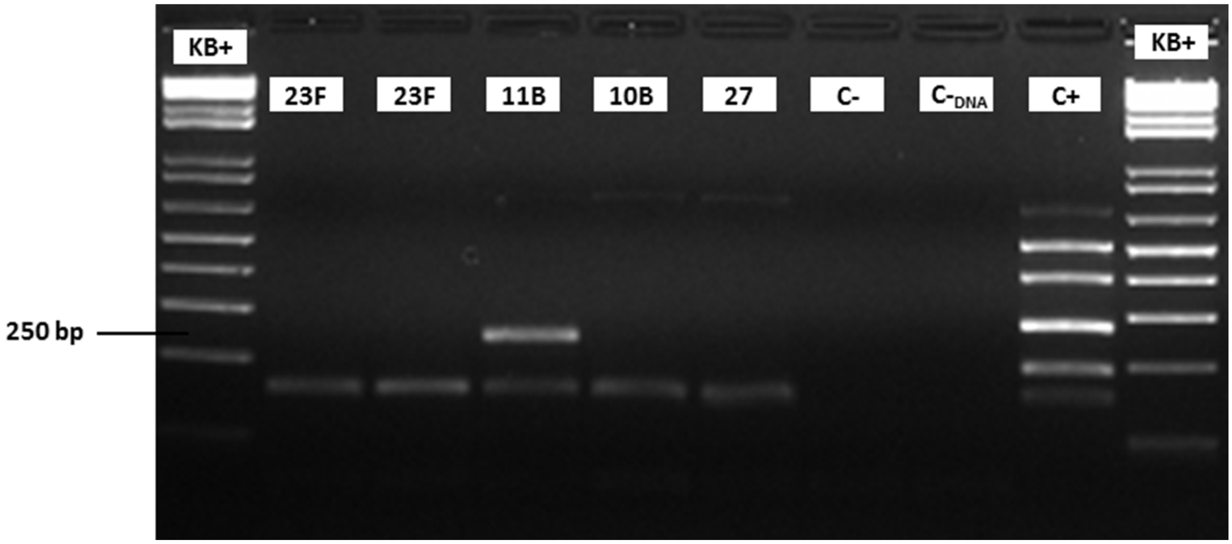


Figure 2E Electrophoresis profile obtained with the multiplex reaction 7. Nonspecific amplifications at 250 bp are visible but do not correspond to a specific primers pair.

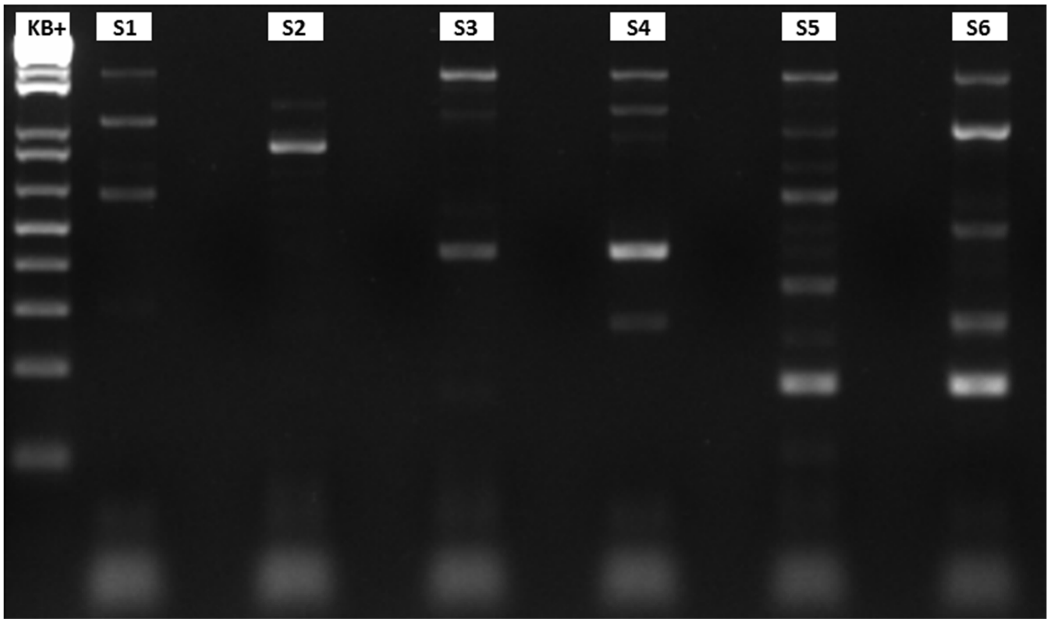


Figure 2F Electrophoresis profile obtained with the multiplex reaction 2 for non *S. pneumoniae* isolates. Multiple nonspecific amplifications are visible but no *cpsA* amplification.

Nonspecific amplifications were present in many reactions and mainly in 4 multiplex PCR reactions. At the multiplex PCR 2, 500 bp nonspecific amplifications were often visible (Figure 2B). These amplifications do not correspond to any primer pairs but are close to other amplifications (434 bp and 599 bp) and could be confounded with one of them, leading to a misidentification. Nevertheless, nonspecific amplifications always produce thinner and dimmer bands, easily distinguished from correct amplifications. This was also the case for multiplex PCR 6 (nonspecific amplification at 850 bp, Figure 2D) and for multiplex PCR #7 (nonspecific amplification at 250 bp, Figure 2E). In multiplex PCR reaction #3, nonspecific amplifications at 677 bp were present in many serotypes (Figure 2C). Unfortunately, this corresponds to the amplification for serotype 35B, which could potentially lead to a misidentification when used routinely. Again, these amplifications produce thinner and dimmer bands unlike positive amplifications. The presence of nonspecific amplification is summarized in Table 6. Among the 61 isolates tested, 31% (19/61) showed nonspecific amplifications at multiplex PCR #2, 44% (27/61) showed nonspecific amplifications at multiplex PCR #3, 5% (3/61) showed nonspecific amplifications at multiplex PCR #6 and 5% (3/61) showed nonspecific amplifications at multiplex PCR #7. For routine analysis, nonspecific amplification may lead to an unacceptable level of false serotype identification.

Non *S. pneumoniae* (*S. pseudopneumoniae* and *S. mitis*) were also tested for each multiplex. In all reactions, *cpsA* amplification (intern control at 160 bp) never occurred. This cannot completely distinguish these streptococci from *S. pneumoniae*, because some *S. pneumoniae* serotypes also do not lead to *cpsA* amplification (serotype 25F and 38). Several non-specific amplifications also occur for these isolates (Figure 2F). This could be used as the discrimination criteria as no *S. pneumoniae* isolates demonstrated such an amplification pattern. Moreover, *S. pneumoniae* strains are generally susceptible to optochin in contrast to other streptococci (Jorgensen *et al.*, 2015). This test is routinely performed on *S. pneumoniae* strains.

Sequential multiplex PCR is a user-friendly and fast serotyping method because this technology is common to all microbiology laboratories. However, depending on the number of multiplex PCR needed for the identification of an isolate, the time required for identification can dramatically increase. As PCR are done sequentially, a limited number of PCR can be performed in a single day. The time required can range from 2 days for an isolate detected in the 1st or 2nd multiplex, to 5 days for an isolate identified in the 8th multiplex reaction. As most common serotypes are detected in the first reactions, the average time for identification with sequential multiplex PCR would be 2.75 days according to serotype distribution in Quebec in 2016.

Evaluation of the sequetyping method based on the *cpsB* gene

We successfully sequenced 53 isolates of the 55 *S. pneumoniae* isolates tested. The average sequence length was 799 bp, which is shorter than the average length in Part 1 (942 bp) but still longer than 732 bp, the length of the sequence used by Leung *et al*., (2012) to test all their serotypes. These 53 isolates were tested against the NCBI database with the same protocol used in Part 1. Detailed results are reported in Table 7. Approximately 53% of the isolates were identified at the serotype level (28/53), 23% (12/53) were identified at the serogroup level, 7% (4/53) were identified as ambiguous and 17% (9/53) were misidentified. Together with Part 1 results, 121 isolates were tested with the sequetyping method. Half of the strains (60/121) were identified at the serotype level, 17% (21/121) were identified at the serogroup level, 12% (14/121) were identified as ambiguous and 21% (26/121) misidentified. As described in Part 1, isolates of serotype 18C were misidentified but showed only 1 mismatch with the 18C reference sequence. The isolate of serotype 19F was also misidentified showing a single mismatch with 19F reference sequence. Leung *et al.* (2012) correctly identified 6 of 7 strain of serotype 19F in their study. It is possible that genetic variations in our isolate have caused this change in the sequence and the misidentification. The same conclusion can be drawn with serotype 17A, showing no significant hit with the 17A reference sequence whereas Leung *et al*., (2012) identified their strain of serotype 17A as ambiguous. In Part 1 of the project, a transcription error occurred for strains MA083042 and MA084138 (previously identified as 19B and 19C, respectively). Both strains were identified as serotype 19A with Quellung. This was considered in the final statistics.

Seventy-nine serotypes were tested in this study. Considering only one isolate per serotype, the sequetyping method allows identification at 50% (39/79) at the serotype level, 15% (12/79) at the serogroup level, 15% (12/79) as ambiguous and 20% (16/79) misidentified. Only one isolate of serotype 25F did not yield *cpsB* amplification but this was predicted by Leung *et al.* (2012).

In Part 1, only 1 of 2 isolates of serotype 29 yielded an amplification of the *cpsB* gene. Five more isolates of serotype 29 were tested and all yielded amplification. This suggests that the non-amplifiable isolate has some genetic characteristics preventing the *cpsB* amplification and that this is not common to most of the isolate of serotype 29 in Quebec. All of these isolates were misidentified as described in Part 1, with perfect matches with serotype 35B and 35C sequences and poor sequence identity with serotype 29 reference sequences (83% identity). This means that *S. pneumoniae* strains of serotype 29 in Quebec are genetically distant from serotype 29 sequences available in the NCBI database. This correlated with the results found with WGS for serotype 29. This mistake could be avoided by creating a local *cpsB* sequence database incorporating serotype 29 sequences from this study as reference sequences.

Only one non *S. pneumoniae* strain (*S. pseudopneumoniae*) led to the amplification of *cpsB*. This sequence was associated with serotype 20 with an identity of 96%. In this study, this is the only isolate with a best hit with an identity lower than 97%. Thus, non *S. pneumoniae* strains could be discarded and not be identified as a proper *S. pneumoniae* by the sequetyping method if we apply an identity criterion of ≥ 97%.

Table 7 Pneumococcal serotype identification using the sequetyping approach.

Isolates	*cps* best NCBI hit subject		HSP(1) identities	Expected serotype(2)	Identification level
	Genbank accession	Serotype			
LSPQ4271	CR931646	9L	954/955	9L	Serotype
	CR931647	9N	953/955		
LSPQ4272	CR931655	11C	929/930	11C	Serogroup
	CR931654	11B	929/930		
LSPQ4273	CR931649	10A	925/925	17A	Misidentified
			No significant hit with 17A		
LSPQ4274	CP002121	11A	933/933	18F	Ambiguous
	CR931674	18F	933/933		
	CR931656	11D	933/933		
LSPQ4275	CR931686	24A	934/934	24A	Serotype
	CR931712	40	907/934		
LSPQ4277	CR931693	28F	947/947	28F	Serotype
	CR931692	28A	946/947		
LSPQ4278	CR931697	32F	932/932	32A	Serogroup
	CR931696	32A	932/932		
LSPQ4279	CR931699	33B	918/919	33B	Serotype
	CR931701	33D	915/919		
LSPQ4280	CR931706	35C	934/934	35C	Serogroup
	CR931705	35B	934/934		
LSPQ4281	CR931714	41F	933/934	41F	Serotype
	AE005672	4	923/934		
LSPQ4282	CR931720	47A	916/917	47A	Serotype
	CP016633	4	882/917		
LSPQ4283	CR931721	47F	939/939	47F	Ambiguous
	CR931707	35F	939/939		
MA080812	CR931650	10B	961/961	10B	Serotype
	CR931649	10A	915/961		
MA082483	CP000918	5	931/931	5	Serotype
	JF911531	19F	911/931		
MA088547	CR931691	27	858/858	27	Serotype
	CR931695	31	840/858		
MA093680	CR931643	7F	918/918	7F	Serogroup
	CR931640	7A	918/918		
MA093772	CR931672	18B	939/940	18C	Misidentified
	CR931673	18C	938/940		

Table 7 (continued)

Isolates	*cps* best NCBI hit subject		HSP[1] identities	Expected serotype[2]	Identification level
	Genbank accession	Serotype			
MA096520	FQ312042	1	945/945	1	Serotype
	JF911531	19F	936/945		
MA096566	CR931655	11C	910/913	11B	Serogroup
	CR931654	11B	910/913		
MA096954	FQ312029	14	857/858	14	Serotype
	CR931632	1	852/858		
MA097140	CR931643	7F	931/933	7F	Serogroup
	CR931640	7A	931/933		
MA097827	AF402095	9V	947/948	9V	Serotype
	CR931645	9A	946/948		
MA098250	CR931647	9N	910/910	9N	Serotype
	CR931646	9L	909/910		
MA098599	KC832411	6F	889/891	6B	Serogroup
	JF911503	6B	889/891		
	JF911497	6A	889/891		
MA098680	FQ312029	14	897/898	14	Serotype
	JF911531	19F	890/898		
MA098806	AF402095	9V	915/916	9V	Serotype
	CR931645	9A	914/916		
MA099139	JF911515	6C	915/915	6C	Serogroup
	HM171374	6D	915/915		
MA099472	JF911496	6A	925/925	6A	Serotype
	CR931639	6B	923/925		
MA099660	CR931672	18B	941/941	18C	Misidentified
	CR931673	18C	940/941		
MA099752	CR931692	28A	923/923	28A	Serotype
	CR931693	28F	922/923		
MA100130	FQ312041	3	910/910	3	Serotype
	JQ653094	20	899/911		
MA100152	CP016633	14	921/921	23F	Ambiguous
	CP016632	12	921/921		
	CP016227	21	921/921		
	FM211187	23F	921/921		
MA100245	CR931647	9N	937/937	9N	Serotype
	CR931646	9L	936/937		

Table 7 (continued)

Isolates	*cps* best NCBI hit subject		HSP[(1)] identities	Expected serotype[(2)]	Identification level
	Genbank accession	Serotype			
MA100658	CR931663	15A	913/914	15A	Serotype
	CR931666	15F	906/914		
MA100764	JF911522	19F	901/902	19F	Serotype
MA100773	AE005672	4	907/907	4	Serotype
	AF402095	9V	898/907		
MA100780	LT594600	22F	920/920	22F	Serogroup
	CR931681	22A	920/920		
MA100925	JF911515	6C	919/920	6C	Serogroup
	HM171374	6D	919/920		
MA101024	JF911496	6A	944/946	6A	Serotype
	CR931639	6B	942/946		
MA101145	LT594599	6E[(3)]	887/887	6B	Serogroup
	KT907353	6B	887/887		
MA101159	CP016633	14	919/920	23F	Ambiguous
	CP016632	12	919/920		
	CP016227	21	919/920		
	FM211187	23F	919/920		
MA101323	FQ312042	1	929/935	1	Serotype
	JF911531	19F	917/930		
MA101386	FQ312041	3	938/942	3	Serotype
	JQ653094	20	925/943		
MA101680	LN831051	10A	874/894	19F	Misidentified
	JF911531	19F	873/892		
MA101744	AE005672	4	916/917	4	Serotype
	AF402095	9V	907/917		
MA101766	CR931663	15A	924/927	15A	Serotype
	CR931666	15F	917/927		
MA101978	JF911519	19A	918/918	19A	Serotype
MA101987	LT594600	22F	915/915	22F	Serogroup
	CR931681	22A	915/915		
MA098344	CR931706	35C	948/949	29	Misidentified
	CR931705	35B	948/949		
	CR931694	29	711/849		

Table 7 (continued)

Isolates	*cps* best NCBI hit subject		HSP[1] identities	Expected serotype[2]	Identification level
	Genbank accession	Serotype			
MA097586	CR931706	35C	943/943	29	Misidentified
	CR931705	35B	943/943		
	CR931694	29	711/849		
MA100224	CR931706	35C	934/934	29	Misidentified
	CR931705	35B	934/934		
	CR931694	29	711/849		
MA101320	CR931706	35C	942/942	29	Misidentified
	CR931705	35B	942/942		
	CR931694	29	711/849		
MA098505	CR931706	35C	946/946	29	Misidentified
	CR931705	35B	946/946		
	CR931694	29	711/849		
ID111828	JQ653094	20	901/935	*S. pseudopneumoniae*	
	JQ653093	20	901/935		
	CR931679	20	901/935		
	CR931661	13	901/935		

[1] HSP = High-scoring Segment Pairs.

[2] Expected serotype according to Quellung reaction.

[3] Serotype 6E has been defined as a serotype 6B subtype cross-reacting with 6B-specific antiserum (Ko *et al.*, 2013)

Two steps are required for the sequetyping method. The first is the *cpsB* amplification and purification step, involving commonly used methods such as PCR and gel electrophoresis. This step is completed in 1 day although it is necessary to consider putative repeats for negative samples. The second step, sequencing and data management requires 2 additional days. Considering strain culture and DNA extraction, the sequetyping method allows for the determination of serotypes in a total of 4-5 days.

Discussion

The goal of this report was to evaluate 3 different DNA-based methods for the serotyping of *Streptococcus pneumoniae* for a possible replacement of the current method routinely used (Quellung method) at the LSPQ in the context of provincial surveillance. Information from Part 1 of this project (Development) and from Part 2 was gathered in order to draw a conclusion about the method most likely to replace the Quellung method. All information is presented in Table 8. Only one isolate per serotype was considered for the final data presentation.

Before summarizing the advantages and disadvantages of the different methods, it is important to note that none of these methods could completely replace the gold standard Quellung method, particularly during the first period of transition. As shown in this report, none of the tested methods provided 100% correct identifications and it would not be prudent to completely trust these results without a period of parallel testing with two methods (Quellung and the chosen molecular method). Thus it appears that the Quellung method will continue to be used and the DNA-based method could serve as a guide for the selection of which antisera to use. Therefore, the precision of the results given by the method will impact the downstream Quellung reactions, a higher precision leading to higher cost effectiveness. Finally, not all antisera are available at the LSPQ and 10-15% of isolates cannot be identified in the provincial laboratory. Currently, these isolates are sent to the NML for identification. An effective molecular method used routinely may decrease the number of strains transferred to the NML, and thus reduce the overall turnaround time.

The first method described here is Whole Genome Sequencing (WGS). It is the most technically difficult method to use because it requires several delicate processes. It is also the most expensive method but cost will likely decrease with the improvement of sequencing technology and cost of reagents. Bioinformatics pipeline can also be laborious to analyse given the amount of data generated. Nevertheless, automatic bioinformatics analyses would be easily implemented. Kapatai *et al.*, (2016) developed such a pipeline for serotyping *S. pneumoniae* with WGS. This method does not rely on genome assembling and performs raw reads alignments on a *cps* sequence database. The strength of the workflow is the use of a second step for the identification of ambiguous serotypes or serogroup (such as 22F/22A). SNPs analysis, loss-of-function mutations and other parameters are checked in order to determine the serotype.

There are many advantages to WGS. Firstly, this is the most reliable method among the 3 tested. Indeed, 94% of the isolates tested were identified at the serogroup (26%) or serotype level (68%). Isolates identified at the serogroup level would require the use of antisera to confirm the serotype with the Quellung reaction, directly targeting the serotypes given by the WGS. The ambiguous result could be easily confirmed as well. PneumoCaT, a bioinformatics workflow designed for the serotyping of *S. pneumoniae*, gave less reliable results than our own method with several misidentifications inside some serogroups (28%). Interestingly, it gave the correct serotype for 22F isolates where our method only determined the serogroup. Thus PneumoCaT could be used in cases where only serogroup is determined. In order to confirm this, it is recommended that this be tested on several isolates identified at the serogroup level by our pipeline.

In this report, serotype 29 was the only serotype misidentified among the 32 tested; it was identified as a serotype 35B. This is a known issue as serotype 35B and 29 are genetically related. However the poor alignment with the serotype 29 reference sequence showed that the serotype 29 isolates from Quebec are genetically distant. This problem could be solved by adding the *cps* sequence from MA097586 to the *cps* database. Other serotype 29 strains should be tested to confirm this result, and thus added these new sequences to the *cps* database. Finally, serotype 29 has a very low incidence in Quebec (0.2% in 2016) so this problem will occur only occasionally. The second advantage is that the huge amount of data generated with WGS will eventually serve for other purposes such as Multi Locus Sequence Typing (MLST) or antibiotic resistance gene detection. MLST is a powerful tool allowing to follow the evolution of clonal complexes across the province. Antibiotic resistance genes (*mefA* and *ermB)* are screened at the LSPQ depending on erythromycin MIC results. From 2010 to 2016, this concerned 31% of the *S. pneumoniae* strains received at the LSPQ. The detection PCR could be easily replaceable by an exhaustive search of *mefA* and *ermB* sequences in the genome. Actually, this method would be more sensitive because small mutations which could affect PCR (mainly in the primers sequences) would barely affect blast results. Thus, the relative WGS cost per strain could decrease with the extensive use of the data generated.

Unlike WGS, sequential multiplex PCR and sequetyping are designed to target *cps* capsule genes. The data generated with these methods can only be used to determine a serotype. Multiplex PCR is the easiest and fastest method to use. As specified in Part 1 of this project, the multiplex designed could be adapted to provide a better match with the local serotype incidence. Unfortunately, this cannot be easily achieved because redesigning the combination of primers would be necessary. This would be very difficult to achieve because of amplicon size or putative cross reactivity between primers and optimization would be necessary.

Multiplex PCR is the only molecular method which requires human interpretation. Electrophoresis gel reading by eye can be interpreted in different ways and precision is not always sufficient to draw conclusions about the exact size of amplicons. This is an important aspect because of the presence of nonspecific amplifications which could lead to interpretation errors. Nevertheless, some nonspecific amplification is recurring and can be identified with ease. A significant part (~50%) of isolates is identified at the serogroup or subset level and would require downstream identification with Quellung. Non detectable serotypes using PCR method have to be taken into account for downstream Quellung identification, which represents a substantial portion of isolates (16%, 19/121) of serotypes tested in the study. However, these are rare serotype and their incidence is very low in Quebec.

The final DNA-based method tested in this project is the sequetyping method developed by Leung *et al*., (2012). This method is based on the sequencing of unique sequences inside the *cpsB* gene. The method is very inexpensive, easy to use and is not impacted by serotype variation over time. Unfortunately, this is the method with the most misidentified serotypes (21%, 26/121). Moreover, even if 50% of serotypes were correctly identified, the second best HSP often has 1 or 2 mismatches with the best HSP. As explained in Part 1 of the project, intra-specific variations in *cpsB* gene could easily bias these results in an unpredictable way (Varvio et *al.*, 2009). Thus Quellung identification would always be necessary. The existence of an independent curated *cpsB* sequences database would help to improve results.

Table 8 Summary of the molecular methods used for *S. pneumoniae* serotyping.

			Serotyping results (concordance with Quellung)				
Methods	**Advantages**	**Disadvantages**	**Serotype**	**Serogroup**	**Subset [1]**	**Ambiguous**	**Misidentified**
WGS (32 different serotypes)	- Includes all serotypes - Additional information obtained at the same time (multi-locus sequence type, antimicrobial resistance…) are useful for other studies - Identification of putative vaccine target and serotyping evolution analysis	- Laborious - Expensive - Large amount of data to manage - Needs bioinformatics setup - Time consuming	66% (21/32)	25% (8/32)	N/A	6% (2/32)	3% (1/32)
Sequential multiplex PCR[2] (83 different serotypes)	- Method easily achievable - Serotype easily determined - Straightforward	- Significant nonspecific amplification - To be customized according to local epidemiology - Detection of known serotypes - Not useful for all serotypes - Possibility of cross-reactions - Relatively expensive	34% (22/64)	38% (24/64)	28% (17/64)	N/A	0% (0/64)
Sequetyping (79 different serotypes)	- Rapid - Easy to set up - Inexpensive	- Not useful for all serotypes - False assignment of serotype due to potential for gene exchange - Method based on public databases - Necessity of a *cpsB* curated bank	50% (39/79)	15% (12/79)	N/A	15% (12/79)	20% (16/79)

[1] Defined as correct results obtained with PCR multiplex primers detecting a subset, for example 33F/33A/37 (reaction 2).

[2] Not determinable serotypes were not considered for the statistics.

Conclusion

The goal of this report was to evaluate the potential of 3 DNA-based methods for serotyping *Streptococcus pneumoniae* and provide data for the possible replacement of the actual serotyping method in use, the Quellung method. The most important aspects to consider for each method are the cost and the precision in serotype identification.

Sequetyping is the most attractive method because of its very low cost. Unfortunately, we would not recommend such a method due to the high number of misidentified serotypes generated. This issue could be resolved in the future, with an increase in the number of sequences in public databases and through the creation of a curated *cpsB* database. This method cannot be implemented routinely at the present time.

Multiplex PCR was evaluated as an efficient method with no misidentified serotypes. This was also the easiest method to achieve routinely. An unacceptable level of nonspecific amplification occurred which could lead to incorrect identifications. Overall, 4 different types of nonspecific amplifications occurred (usually with a thinner and dimmer band) and those nonspecific amplifications may be a source of error (misidentification of serotype). Also, the proportion of serotype identified at the serotype level is low, which means that this method will always rely on the Quellung method for the exact identification of serotypes and will never completely replace it. Instead, it could serve as a guide to perform more effective Quellung method. Finally, this method is more costly than it initially appears and its efficiency depends on epidemiological data. The cost of this method is also subject to serotype replacement.

WGS proved to be the best molecular method among the three methods tested. Only one misidentification (serotype 29) due to local genetic variations was encountered. Moreover, few or no downstream Quellung reactions were needed in this method. Unfortunately, this is currently the most expensive method and the least convenient to perform. The price will continue to decrease with the diminution in material costs and the use of genome information for other purposes. Finally, WGS does not depend on serotype circulation or replacement and performance should not be affected over time. Currently, using WGS only for serotyping in *S. pneumoniae* surveillance is too expensive, however identification of antibiotic resistance genes may be a possible approach to improve cost-effectiveness.

Multiplex PCR seems to be an acceptable option as it is easy to routinely implement. However, multiple nonspecific amplifications may affect the quality of the results and this method still relies on Quellung because numerous isolates only identify at serotype or subset level. WGS could become more attractive with competitive prices. This method provides excellent results for *S. pneumoniae* serotyping and is recommended as a replacement or alternative method for the gold standard.

Before implementation, the recommended DNA-based method, WGS, should be evaluated on *S. pneumoniae* strains received at the LSPQ for a determined period in parallel with the Quellung method to assure the quality of the data in a routine context. Non-serotyping *S. pneumoniae* strains were not tested in the project. It will be relevant to test those strains using WGS.

Acknowledgements

This work was supported by Grant IIR number WI197603 from Pfizer Canada.

We wish to thank Irene Martin and Walter Demczuk from NML for sharing *S. pneumoniae* strains.

We thank Maureen Hastie from LSPQ for English revision

References

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics **29**, 1072–1075.

Leung, M.H., Bryson, K., Freystatter, K., Pichon, B., Edwards, G., Charalambous, B.M., and Gillespie, S.H. (2012). Sequetyping: Serotyping *Streptococcus pneumoniae* by a Single PCR Sequencing Strategy. J. Clin. Microbiol. **50**, 2419–2427.

Bentley, S.D., Aanensen, D.M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M.A., et al. (2006). Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes. PLoS Genet **2**.

Bush, C.A., Cisar, J.O., and Yang, J. (2015). Structures of Capsular Polysaccharide Serotypes 35F and 35C of *Streptococcus pneumoniae* Determined by Nuclear Magnetic Resonance and Their Relation to Other Cross-Reactive Serotypes. J. Bacteriol. **197**, 2762–2769.

Varvio, S.L., Auranen, K., Arjas, E., Mäkelä, P.H. 2009 Evolution of the capsular regulatory genes in *Streptococcus pneumoniae*. J. Infect. Dis. **200(7)**:1144-51.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., and Ponting, C.P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. 15, 121–132.

James H. Jorgensen, Michael A. Pfaller, Karen C. Carroll, Guido Funke, Marie Louise Landry, Sandra S. Richter, David W. Warnock. Manual of Clinical Microbiology. 2015. 11th edition, **22**, 383-402.

Kwan Soo Ko, Jin Yang Baek and Jae-Hoon Song (2013). Capsular Gene Sequences and Genotypes of "Serotype 6E" *Streptococcus pneumoniae* Isolates. J. Clin. Microbiol. **51(10)**: 3395-3399.

Title

Serotyping of *Streptococcus pneumoniae* invasive strains using molecular biology tools

Authors

Florian Mauffrey[1], Éric Fournier[1], Christine Martineau[1], Simon Lévesque[1], Sadjia Bekal[1], Marc-Christian Domingo[1], Walter Demczuk[2], Irene Martin[2], Florence Doualla-Bell[1], Cécile Tremblay[3], Jean Longtin[1], and Brigitte Lefebvre[1]

Affiliations

1) Laboratoire de santé publique du Québec, Institut national de santé publique du Québec, QC, Canada

2) National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada

3) Université de Montréal, Montréal, QC, Canada

Objectives

Streptococcus pneumoniae is a major cause of pneumonia, meningitis and other pneumococcal infections. Over 90 serotypes have been described so far. The Quellung reaction is the gold standard test for *S. pneumoniae* serotyping. From a public health perspective, accurate serotyping of *S. pneumoniae* is essential to monitor the serotype replacement following the introduction of Pneumococcal Conjugate Vaccines. Unfortunately, this method is costly, time-consuming and dependent on human interpretation. In this study, we evaluated the efficiency of three different molecular serotyping methods as an alternative to the Quellung method.

Methods

One hundred twenty-one *S. pneumoniae* strains representing 83 serotypes were serotyped with a sequential multiplex PCR assay (CDC protocol) and a sequence typing assay (sequetyping) based the *cpsB* gene sequence. Furthermore, 53 *S. pneumoniae* strains representing 32 serotypes were serotyped with whole genome sequencing (WGS) assay using an in-house pipeline and the bioinformatics tool PneumoCat. The serotype of all these strains was previously identified by the Quellung method.

Results

The proportion of serotypes identified using sequential multiplex PCR to the serotype level was too low (34%) to use as an alternative to the Quellung method. Moreover, a large proportion (23%) of strains was not typeable by the PCR assay. Although the sequetyping is currently the most economical method, it exhibited a high number of misidentified serotypes (20%) and ambiguous results (15%). The WGS-based serotyping methods using our in-house pipeline exhibited the best performance as they predicted capsular types at serotype and serogroup levels for 91% (66% at the serotype level) of the strains tested with only one misidentified serotype. In contrast, PneumoCaT results revealed several misidentifications inside serogroups (21%).

Conclusion

WGS could be considered as a potent tool for *S. pneumoniae* serotyping and be useful for epidemiological purposes. Moreover, data generated can be used for further investigations such as antibiotic resistance genes characterization or multilocus sequence typing.

Acknowledgement

The project was funded by Pfizer Canada (grant IIR, WI97603).

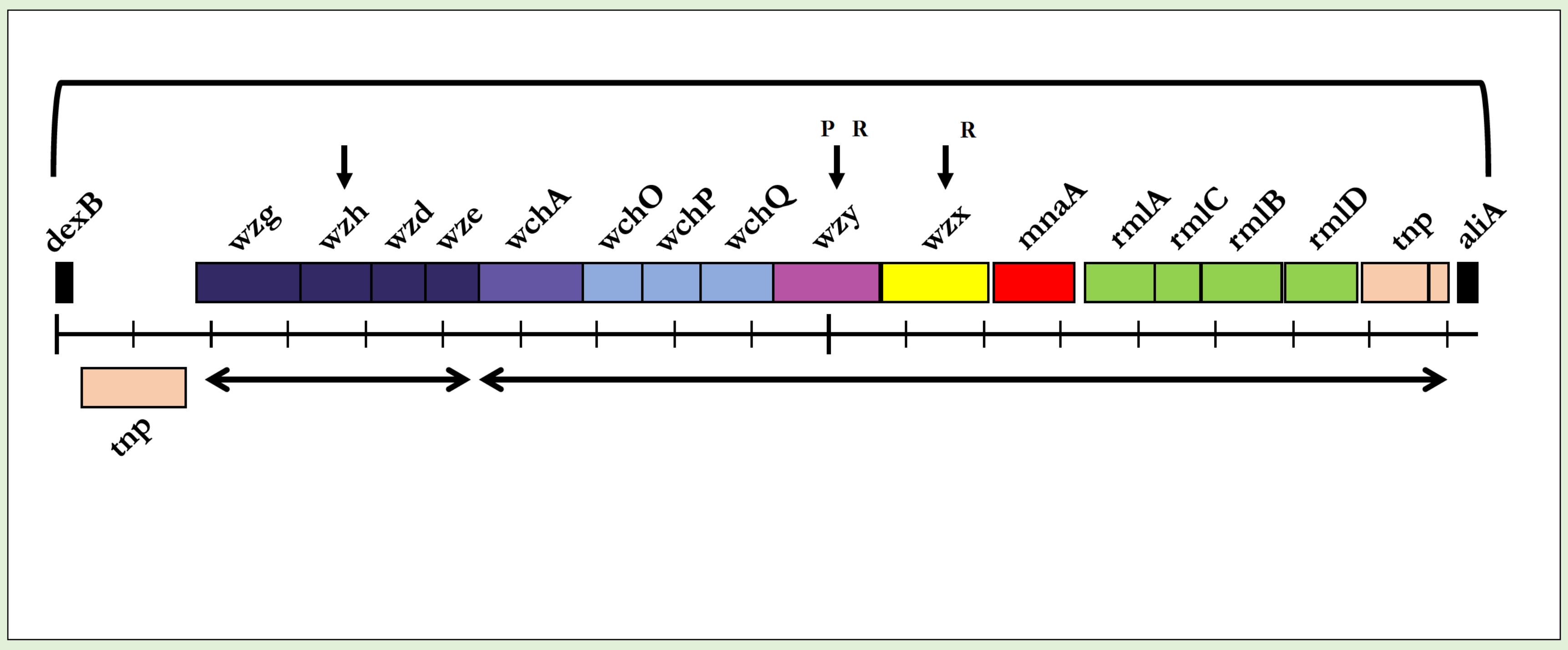
Serotyping of *Streptococcus pneumoniae* invasive strains using molecular biology tools

Flo i n Mauf re   Ér c Fourn er   Chr s ine M rt neau   S mon Léve que   Sadjia Bek l   Ma c-Ch i t an Domingo   W l er D mc uk   rene M rt n
Flo ence Doua la Bell   C ci e Tremblay   Jean Long in   and Br gi te Lef bvre

L bor to re de an é pub iq e u Qué ec ns i ut n t onal e an é pub i ue du Qu bec a nt -Anne de Be l vue QC Ca ada
Na i nal Mi rob o ogy Lab ra ory Pub ic He l h Agency of Can da Wi nip g MB Can da
Un v rs té de Mon ré l Mon r al QC C nada

Background

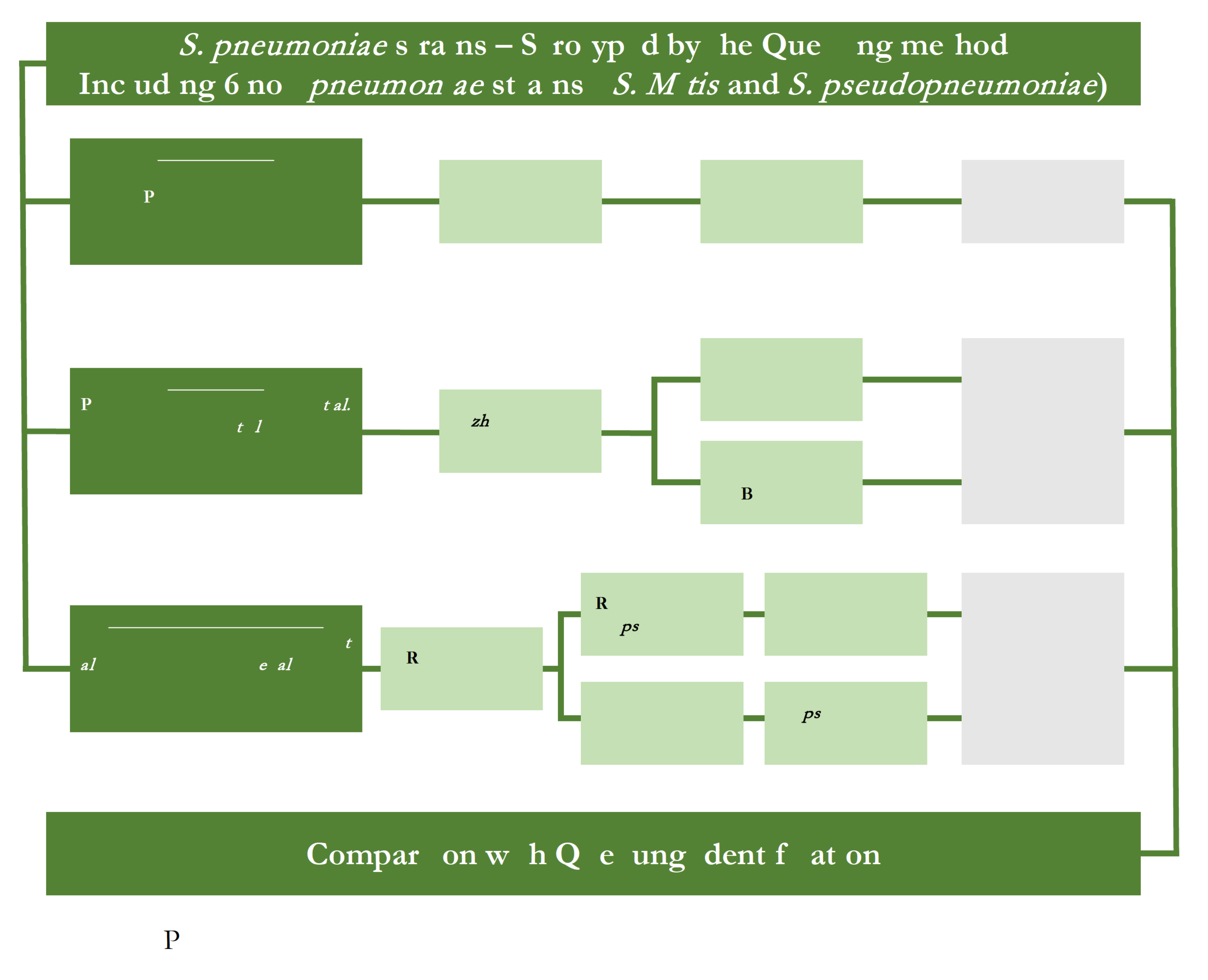
S p o o cu pn um n ae s ne of the major c us s f neumo ia men ngt s a d o her pn umoc cc l nf ct o s in y ung c i d en and he e d rly D t rm na i n f *S p e mon ae* s rot pe s of h gh m or an e in p bl c h a th to mon or pu a i e se otype ep a ement f l ow ng t e n r duc on of PCVs (P eumo oc al Co juga e Vac i e) nd t e f i acy of mm n za i n pr gr ms T e pn umoco c l s roty e is de erm n d by t e poly ac ha de c mpos ion of he ap ule Mo e han 90 *S p eum n ae* c psu ar poly ac ha de (CP) types e i t re u t ng n a a ge v r ety of se otyp s e ong ng o 6 d f e ent er gr ups Al gen s r l ted o he c ps le syn he is a hway a e gr uped n a s ng e CPS op ron Pr s nce or a s nce of g nes sp c f c ge es o ga i at on as we l as s ec f c eq en e v r a i ns o he i f re t er types and can be u ed or th ir i en f ca on



The Q e l ng m th d r ma ns the go d t nd rd se otyp ng m t od w dely us d n s rv i an e f *S p eu on ae* A th ugh his m th d s v ry e f c i e t is l o co tly t me co sum ng a d ot ot lly r l a le ue to h m n i te p e at on

The objec ves of th s s udy were o e a ua e he ef c ency of d f er nt mol cu ar s roty ing me hods a mu t p ex CR a say a *w h* ba ed eq en ing me hod a l d s qu ty i g nd t e se ue c ng of t e en i e *cp* ocus h ough w ole genome eq en ing

Methods


S. pneumoniae s ra ns – S ro yp d by he Que ng me hod
Inc ud ng 6 no *pneumon ae* st a ns *S. M tis* and *S. pseudopneumoniae*)

Compar on w h Q e ung dent f at on

P

Results

S   W S
P R   B   B   P
d   d
S   %   %   %   %   %
S   %   %   %   %   %
%   %   %   %
S   %
d   d   %   %   %   %   %
%

S   d
dd
d   d   d   d
S   d
d
d
d

S
d
d   d   d
d
d   d   d   d
d
d
P
R

The b st de t f c t on r su s were o t in d us ng he who e ge ome e uen i g pp oa h f l ow ng a g nome as emb i g t a eg M re ver a l s rot p s re co e ed w h his me hod n i e mu ip ex PCR and se uetyp ng The *cp* e uen ~~es~~ da a ase c n l o e ad p ed o lo al i ua ons In our s udy we fo nd h t the s r ty e 9 *p* equ nce n Québ c di er d r m pu l s ed se uen es u ed n o h r a ab ses au i g mi i en i ca i ns A d t on of a lo al e otype 29 *cp* s que ce in he da ab se ol ed t e pr b em WGS s rotyp ng ap ro ch demon t a ed ro us ne s by co re tly id n ify ng a s m le on ami a ed wi h *Bac u ub* Con e sely many n n- pe i c mp i i a io s we e pr s nt in mPCR as a s nd cou d l ad to r ut ne m s den f c t ons F na l s qu t pi g sh wed t e highe t r te of m s de t f c t ons p ss bly due o the x s en e of in a- pe i c va a io s in he *p* c ns rv d regi n

Conclusion

In his t dy we ave d mons r t d h t WGS was he most e ab e m t od among he 3 mol cu ar neumo oc al er typi g m th ds t s ed er ty e a i a ion w th Que ung is s i l r qu r d as s me s rotyp s an ot e c e rly i t ng i hed us ng *cp* equ nc s Se uen al mu ip ex PCR a d se uetyp ng h ve he ad an age o be ch ap r t an WGS a d co ld a so e ve s a gu de for Qu l ung met od Non th l ss th se met ods ha e d awba ks m k ng hem l ss a t ac ve Con er ely s qu typi g or mu i l x PCR ap ro ch s may s pp ement d f cu t o o s rve Que l ng ea t ons WGS cou d e co s de ed as a p t nt to l or *S p eu on ae* er typi g nd er a nly ery u e ul for ep d m o og c l p rpo e con i e ing s in r as d ac e s b l ty nd l wer os s

References

B   P   B   B
P
P   Z
B   P
P
B
P
P

Acknowledgments

Th s work w s su po t d by a gr nt rom Pf z r Can da
We wi h o t ank re e M r in nd W l er D m zuk rom NML or s a ing *S p eum n ae* s r i s and who e g nome s qu nc s

Laboratoire de santé publique du Québec

July 5, 2017

Dr. Brigitte Lefebvre
Laboratoire de Santé Publique du Québec
20045 chemin Ste-Marie
Sainte-Anne-de-Bellevue
Quebec
H9X 3R5
Re: Pfizer Reference # **WI197603**

Dear Dr. Lefebvre:

I understand that you have completed the Pfizer-supported investigator-initiated research study with reference # **WI197603** entitled: ***Molecular tools for serotyping for Streptococcus pneumoniae invasive strains surveillance in the province of Quebec.*** Now that the study is closed, please complete the enclosed ***Certification of Study Closur****e* form and return it to my attention, either by fax; number **514-693-4715** or via email.

We look forward to working with you on future research projects. If you have any questions or comments, please do not hesitate to contact me at [redacted]

Best regards,

[redacted]

IIR Grant Specialist
Medical Quality & Effectiveness

Encl: certification of study closure form

Pfizer	**CERTIFICATION OF STUDY CLOSURE FORM**	

INVESTIGATOR INFORMATION

PRINCIPAL INVESTIGATOR INSTITUTION NAME AND ADDRESS	**Dr. Brigitte Lefebvre** **Laboratoire de Santé Publique du Québec** **20045 chemin Ste-Marie** **Sainte-Anne-de-Bellevue** **Quebec** **H9X 3R5**		
PFIZER REFERENCE NUMBER	**WI197603**	EXTERNAL REFERENCE NUMBER	
STUDY TITLE	**Molecular tools for serotyping for Streptococcus pneumoniae invasive strains surveillance in the province of Quebec.**		

CERTIFICATION AND SIGNATURE

I certify that the study identified above was conducted, any and all safety reporting obligations were met, and

☒ **Funding (please check if received)** □ **N/A**

- all funding under the Pfizer grant has been paid; and
- the Pfizer grant funds were used solely to conduct or report the study and not for any other purpose

□ **Pfizer Product or Compound (please check if received)** ☒ **N/A**

PFIZER AUTHORIZES YOU TO DESTROY ANY UNUSED SUPPLIES OF PFIZER-PROVIDED PRODUCT OR COMPOUND.

□ I confirm that all supplies of Pfizer product/compound were used in the conduct of the study and that no Pfizer product/compound remained at closure.

OR

□ I confirm that the unused supplies of Pfizer product/compound were destroyed in accordance with all applicable regulations, governmental guidelines, and institutional policies.

□ **Capital Equipment (please check if received)** ☒ **N/A**

□ I confirm that the capital equipment listed in the IIR agreement has been used only for purposes of the study.

□ I confirm that the institution has donated the Pfizer equipment to a non-profit organization other than institution itself or any Institution affiliate.

☒ **Future Publications**

☒ As read and acknowledged in the IIR agreement, I confirm I am required to provide Pfizer an opportunity (45 to 60 days before submission or other public disclosure) to prospectively review any proposed publication, abstract, or other type of disclosure that reports the results of the study. Principal investigator will consider any such comments in good faith but is under no obligation to incorporate any Pfizer suggestions.

Brigitte Lefebvre | **2017-07-21**

[redacted] | *Date*

Please fax this form to Pina Mustillo at 514-693-4715 or by email pina.mustillo@pfizer.com.

Pfizer

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist	[redacted]	IIR Grant Specialist PHONE	[redacted]
IIR Grant Specialist EMAIL		IIR Grant Specialist FAX	

PLEASE COMPLETE AND RETURN BY: June 21, 2017

Per contractual requirements, we are requesting a status update on your IIR study supported by Pfizer via funding and/or drug. Please answer the following questions regarding the above referenced study by the due date. Answers from your last submitted update have been incorporated below; please update as needed and answer the remaining questions.

GENERAL INFORMATION

Pfizer Tracking #	**WI203144**	**Institutional Protocol #**	**2014 192, CE 13.212, BSP**
Principal Investigator	**Dr. Brigitte Lefebvre**		
Study Title	**Serotype monitoring of S. pneumoniae invasive strains in adult population in the province of Quebec_ a 3 years study evaluation.**		

STUDY UPDATE INFORMATION

Has this study been initiated?	□ NO ☒ YES	**Date of initiation**	mm/dd/yyyy **01/01/2016**
Has the protocol been amended since last update?		☒ NO □ YES *(If YES, please provide the revised protocol)*	
Current IRB/IEC approval/renewal expires on **November 5, 2017**		This is not current, please forward the most recent letter	
Have there been any personnel changes? *(If YES, please provide name and full contact info on Page 3)*			☒ NO □ YES
Target protocol enrollment	**550 strains**	Date of first subject enrolled	mm/dd/yyyy **01/01/2016**
Last reported enrollment	**189 strains**	Actual enrollment to date *(this should not include screen failures)*	**633 strains**
Targeted last subject last visit		Actual last subject last visit	
Do you have current drug supply sufficient to complete the study? *(If NO, please complete the Drug Section on Page 3)*			□ NO □ YES **Not applicable**
Is this protocol closed to enrollment? *(patients may still be receiving therapy)*			☒ NO □ YES
Targeted study completion date *(primary objectives met; patient therapy and final study analysis complete)*			mm/dd/yyyy 12/31/2018
Actual study completion date *(if applicable)*			12/31/2018
Targeted date to provide results to Pfizer			30/06/2019

PUBLICATION INFORMATION

Do you plan to publish? *(If YES, please complete the information below.)* □ NO ☒ YES

Please be aware that, according to the IIR agreement, the investigator is required to provide Pfizer with an opportunity to prospectively review any proposed publication, abstract or other type of disclosure that reports the results of the study.

FORMAT	**PUBLICATION**	**PLANNED**	**ACTUAL**	**SUBMISSION**

STUDY STATUS UPDATE FORM: CLINICAL

IIR Grant Specialist [redacted]
IIR Grant Specialist EMAIL

IIR Grant Specialist PHONE [redacted]
IIR Grant Specialist FAX

	(please include anticipated journal or audience)			DATE
Abstract	International Symposium on Pneumococci and Pneumococcal Diseases 5-19 April, 2018, Melbourne, Australia.	□	☒	
Manuscript	Vaccine/PlosOne	☒	□	
Poster	CACMID	☒	□	
Other		□	□	

SIGNATURE

NAME Brigitte Lefebvre

DATE 14/07/2017

[redacted]

SIGNATURE (ONLY if faxed)

Pfizer	**STUDY STATUS UPDATE FORM: CLINICAL**	

IIR Grant Specialist IIR Grant Specialist EMAIL	[redacted]	IIR Grant Specialist PHONE IIR Grant Specialist FAX	[redacted]

DRUG SUPPLY INFORMATION

SUPPLY CURRENTLY ON SITE	**ACTIVE**	**PLACEBO**
ESTIMATED REMAINDER REQUIRED TO COMPLETE STUDY	**ACTIVE**	**PLACEBO**
CAN PHARMACY ACCOMODATE TOTAL REMAINDER?	□ **YES**	□ **NO**

PERSONNEL INFORMATION

	PRINCIPAL INVESTIGATOR	**COORDINATOR**
NAME		
INSTITUTION		
MAILING ADDRESS		
TELEPHONE		
FAX		
EMAIL		

	PHARMACIST	**OTHER** *(specify in additional comments)*
NAME		
INSTITUTION		
MAILING ADDRESS		
TELEPHONE		
FAX		
EMAIL		

ADDITIONAL COMMENTS

Comparison of sequential multiplex PCR, sequetyping and whole genome sequencing for serotyping of *Streptococcus pneumoniae*

Florian Mauffrey[1], Éric Fournier[1], Walter Demcuzuk[2], Irene Martin[2], Michael Mulvey[2], Christine Martineau[1], Simon Lévesque[1], Sadjia Bekal[1], Marc-Christian Domingo[1], Florence Doualla-Bell[1], Jean Longtin[1]and Brigitte Lefebvre[1]*

[1] Laboratoire de santé publique du Québec, Institut national de santé publique du Québec, Sainte-Anne-de-Bellevue, Québec, Canada.

[2] National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada.

*Corresponding author

Tel: (514) 457-2070 extension 2334

E-mail: brigitte.lefebvre@inspq.qc.ca

Manuscript for Plos One

Abstract

Streptococcus pneumoniae is one of the major causes of pneumonia, meningitis and other pneumococcal infections in young children and elders. Determination of circulating *S. pneumoniae* serotypes is an essential service by public health laboratories for the monitoring of putative serotype replacement following the introduction of pneumococcal conjugate vaccines (PCVs) and of the efficacy of the immunization program. The Quellung method remains the gold standard for typing *S. pneumoniae*. Although this method is very effective, it is also costly, time consuming and not totally reliable due to its subjective nature. The objectives of this study were to test and evaluate the efficiency of 3 different molecular methods compared to the Quellung method. Sequential multiplex PCR, sequetyping and whole genome sequencing (WGS) were chosen and tested using a set of diverse *S. pneumoniae*. One-hundred and eighteen isolates covering 83 serotypes were subjected to multiplex PCR and sequetyping while 88 isolates covering 53 serotypes were subjected to WGS. Sequential multiplex PCR allowed the identification of a significant proportion (49%) of serotypes at the serogroup or subset level but only 27% were identified at the serotype level. Using WGS, 55% to 60% of isolates were identified at the serotype level depending on the analysis strategy used. Finally, sequetyping was the method resulting in the most misidentified serotypes (17%). The use of Jin *cpsB* database instead of the GenBank database slightly improved results but did not significantly impact the efficiency of sequetyping. Although none of these molecular methods may currently replace the Quellung method, WGS remains the most promising molecular pneumococcal serotyping method.

Introduction

The Gram-positive lancet-shaped cocci bacteria *Streptococcus pneumoniae* is frequently associated with meningitis, pneumonia and sepsis in humans in addition to be the major cause of mortality in children (1). Pneumococcus infections mainly occur among young children and the elderly, under 5 years old and above 65 years old, respectively (2). More than 90 *S. pneumoniae* capsular polysaccharide (CPS) types exist resulting in a large variety of serotypes belonging to 46 different serogroups (3). In Canada, the introduction of the seven-valent pneumococcal conjugate vaccine (PCV-7) in 2005 targeting the seven predominant serotypes (4, 6B, 9V, 14, 18C, 19F, and 23F) led to a significant decrease in invasive pneumococcal diseases (IPD) associated to these serotypes (4). However, replacement of vaccine serotypes by non-vaccine serotypes (NVT) led to the emergence of serotype 19A as the new predominant multi-drug resistant serotype (5). Following the advent of NVT, two others vaccines were released in 2008 and 2010, PCV-10 and PCV-13, respectively. The monitoring of IPD serotypes became essential as new NVT may have emerged making the introduction of new vaccines necessary.

Serotyping methods of *S. pneumoniae* can be grouped in two different categories: phenotype-based methods and genotype-based methods (6). The Quellung method (based on antisera reactions) still remains the Gold Standard method used in most laboratories (7). However this method is expensive, laborious and not fully reliable. Following the sequencing of the *cps* loci of 90 pneumococcal serotypes, methods based on the detection of serotype-specific genes were developed in order to provide cost-effective and reliable assays for the serotyping of *S. pneumoniae* (6,8).

Among these methods, three were chosen for comparison in this study: sequential multiplex PCR, sequetyping and whole genome sequencing (WGS). The sequential multiplex PCR protocol was developed by the Centers for Disease Control and Prevention (CDC) and relies on the use of primers targeting serotype- or serogroup-specific regions (*wzy* or *wzx*) in the *cps* loci (9). PCR has been extensively used for the serotyping of *S. pneumoniae* and had the advantage of being easy to use and can be performed on a large quantity of samples (10–13). The sequetyping method was developed by Leung *et al.* (2012) and is based on the *cpsB* gene sequence which appears to be specific to serotypes. WGS became a suitable method for serotyping with the improvement in accuracy and a decrease in cost which has allowed the identification of serotype by comparing *cps* loci sequences (14–16).

The replacement of the Gold Standard Quellung method in routine laboratories by a genotype-based method is a current issue for many laboratories, requiring preliminary estimations of the efficiency and adaptability of different methods. Such comparisons and evaluations for some methods have already been conducted (17–21). Unfortunately, inter-strain genome variations led to an increase in *cps* loci rearrangement and diversity. Thus the efficiency of molecular serotyping methods may vary between strains and/or between different regions (8,22).

In this study, a large number of serotypes were included, but a focus on the most prevalent serotypes in Québec/Canada and serotypes targeted by PCV-13 were chosen. The evaluation of a potential molecular replacement for the Quellung identification method was considered.

Material and methods

Isolates, culture conditions and DNA extraction

One hundred eighteen invasive *S. pneumoniae* representing 83 serotypes previously identified by the Quellung reaction were selected from the Laboratoire de santé publique du Québec (LSPQ) provincial surveillance program (see Table S1). All the isolates were subjected to sequential multiplex PCR and sequetyping methods. Six serotype 35A isolates and six serotype 34 isolates were added to the pool tested with the sequential multiplex method as well as six serotype 29 isolates were added to the sequetyping pool. A subset of 53 isolates were tested with WGS and represented 32 different serotypes. The selection of the serotypes was performed on the basis of the most prevalent serotypes in the province of Québec in 2012-2016 (Figure 1). Rare serotypes were also included in order to test the robustness of the method. WGS data for 35 *S. pneumoniae* was also provided by the National Microbiology Laboratory (NML, Winnipeg, Canada), totaling 88 isolates representing 53 serotypes subjected to serotyping using WGS approach. Finally, three *Streptococcus pseudopneumoniae* and three *Streptococcus mitis* were used as specificity controls for sequential multiplex PCR and sequetyping.

Isolates were cultured on TSA II (Trypticase Soy Agar with 5% sheep blood) agar plate and incubated overnight at 35°C in a 5% CO_2 atmosphere. Bacteria were collected with a loop and suspended in G2 buffer solution with RNase A (QIAGEN inc, Toronto, ON, Canada). Samples were then frozen at -20°C until extraction. DNA extraction was performed with the MagAttract DNA Mini M48 Kit (QIAGEN inc, Toronto, ON, Canada) and the QIAGEN^{TM} BioRobot M48 workstation according to manufacturer's instructions.

Sequential multiplex PCR

The CDC sequential multiplex PCR protocol was used as described by Carvalho *et al.* (2010). Briefly, primers pairs were designed to target serotype- or serogroup-specific regions in the *wzy* or *wzx* genes. The choice of primers was modeled on those included in the CDC protocol as they were adapted to the 22 most prevalent serotypes in Quebec (2012-2016). These serotypes represent 90 % of IPD in Quebec. All serotypes included in the PCV-13 (4, 6B, 9V, 14, 18C, 19F, 23F, 1, 5, 7F, 3, 6A and 19A) were also covered by this protocol. Positive and negative controls were used in each reaction. Positive controls consisted of a mix of *S. pneumoniae* DNA extract of serotypes present in each multiplex. *S. pseudopneumoniae* and *S. mitis* DNA extracts were tested in each multiplex as a control of specificity.

Sequetyping

Sequetyping procedures were conducted as described by Leung *et al*. (2012) with some modifications. Briefly, master mix was composed of 0.3 µl of Amplitaq DNA polymerase (5 U/µl), 38.85 µl of DNA-free water, 5 µl of 10x PCR buffer (ThermoFisher Scientific, Whitby, Canada), 1.5 µl of $MgCl_2$ (50 mM), 0.75 µl of dNTPs (10 mM), 0.8 µl of *cps1* and *cps2* primers (25 µM) and 2 µl of DNA extract for a final volume of 50 µl. Cycling conditions was performed as described by Leung *et al.* (2012). Sequencing was performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific, Whitby, Canada) in a 3130*xl* Genetic Analyzer (ThermoFisher Scientific, Whitby, Canada).

Assembled *cpsB* sequences were blasted against a local and comprehensive *cpsB* database developed by Jin *et al*. (2016). This database extended the previous database created by Leung *et al*. (2012) by covering 95 serotypes instead of 93 and including a total of 390 sequences. Then, *cpsB* sequences were used to interrogate the GenBank database (https://www.ncbi.nlm.nih.gov/genbank/). In-house Python scripts allowed the automation of these processes. Serotypes were attributed considering hits with the highest bit scores.

Whole genome sequencing

Libraries for whole genome sequencing were prepared with the Nextera XT DNA library preparation kit and sequenced using an Illumina MiSeq reagent kit v3 (600 cycles, paired ends) following the manufacturer's instructions. Reads quality was evaluated with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). *De novo* genome assemblies were performed using SPAdes version 3.9.0 (23) assembler on Calcul Quebec public resources (http://www.calculquebec.ca/en/) with standard parameters. Assemblies' quality was assessed with Quast (24). Concerning the identification of the different *cps* loci, a local *cps* database was created with 107 *cps* sequences representing 92 different serotypes (3) retrieved from the NCBI GenBank database. Assembled contigs containing *cps* sequences were blasted against this database using BLAST+ tools suite in an automated in-house Python scripts. Hits with the highest identity value and High Scoring Pair (HSP) length were retained for serotype attribution. When multiple hits had high identity value (<0.5% compared to best hit) for an equivalent HSP length, they were all retained for serotype attribution.

PneumoCaT (Pneumococcus Capsular typing Tool), a serotyping designed workflow, was also used for serotype identification (25). Automation of PneumoCaT was performed using a shell command based script. Best hits were always considered for serotype attribution. When capsular typing variant analysis occurred (see Kapatai *et al.*, 2016), this result was retained for serotype attribution.

Isolates misidentified with the assembly-based strategy were subjected to further investigations. The *cps* locus was extracted from the corresponding contig according to the best hit coordinates and aligned with *cps* reference sequences of both best hit and expected serotype, for comparison. Alignments were done using the Artemis Comparison Tool (ACT) v6 and WebACT (26).

Serotype identification levels

For all the methods tested in this study, sample identification was classified as follows: 1) Serotype when the correct serotype was determined, 2) Serogroup when the correct serotype was determined as well as other serotype(s) from the same serogroup, 3) Subset when the correct serotype was determined as well as other serotype(s) from a different serogroup, 4) Misidentified when an incorrect serotype was determined and 5) Not determined (N.D.) when no amplification occurred in PCR multiplex reactions or when *cpsB* was not amplified in the sequetyping method. When isolates of the same serotypes had different identification levels with the same method, they were classified as inconsistent results when results per serotype were considered.

Results

Sequential multiplex PCR

Among all existing *S. pneumoniae* serotypes, the CDC sequential multiplex PCR protocol is able to detect 74 different serotypes. *cpsA* amplification ensures the presence of *S. pneumoniae* DNA in each reaction. In our experiments, *cpsA* amplification product was present in all reactions except for isolates of serotypes 25F and 38. The absence of amplification in those serotypes has been previously documented by Carvalho *et al*., (2010). Moreover, no *cpsA* amplification occurred with *S. pseudopneumoniae* and *S. mitis* isolates.

In this study, 130 isolates were tested with multiplex PCR method, covering 83 serotypes. Of the tested isolates, 45/130 (35%) were identified at the serotype level, 42/130 (32%) were identified at the serogroup level, 22/130 (17%) were identified at the subset level, 19/130 (15%) were not determined, and 2/130 (1%) were misidentified (Table 1A). All serotypes were not equally represented in our isolates selection, thus these results are not representative of the method efficiency concerning identification level. Nevertheless, all results were consistent when multiple isolates were tested for a same serotype, except for serotype 35A (1% of serotypes). Considering identification for each serotype, 22/83 (27%) were identified at the serotype level, 24/83 (29%) were identified at the serogroup level, 17/83 (20%) were identified at the subset level, 19/83 (23%) were not determined and 0/83 (0%) were misidentified (Table 1B).

Serotypes 34 and 35A showed unexpected results. Serotype 34 sample showed many amplicons, including a non-specific amplicon (250 bp) and the expected amplicon (408 bp), in the same reaction (multiplex PCR 7). Six more serotype 34 isolates were selected and subjected to identification with sequential multiplex PCR and the same non-specific amplification was present in 3 out of 6 reactions. The expected amplification product at 280 bp was not present in the multiplex PCR 7 with serotype 35A and 6 other serotype 35A isolates were further selected. For 5 out of 6 isolates, the expected amplicon was detected but a non-specific amplicon at 250 bp was also visible. It should be noted that expected amplicons bands are very well defined and have high intensity compared to non-specific amplicons bands which are generally less bright.

Non-specific amplification products were present in many PCR reactions. They seemed to occur randomly and did not depend on the isolate serotype. Only 4 different sizes non-specific amplicons were observed during this study, a non-specific bands at 500 bp in multiplex PCR 2, a non-specific band at 677 bp in multiplex PCR 3, a non-specific band at 850 bp in multiplex PCR 6 and a non-specific band at 250 bp in multiplex PCR 7. Except for the band at 677 bp in the multiplex PCR 3, these non-specific products did not correspond to expected product sizes in their respective multiplex PCR and were easily identified as non-specific. However, the amplification product at 677 bp in multiplex PCR 3 corresponds to the expected size for serotype 35B and is hardly identifiable as non-specific. Many non-specific amplicons were also present for *S. pseudopneumoniae* and *S. mitis* in most of the multiplex PCR.

Sequetyping

Of the 124 *S. pneumoniae* isolates subjected to sequetyping, 118 (95%) were positive for *cpsB* amplification (1061 bp). No *cpsB* amplification was obtained for serotypes 25F, 37, 38, 39 and 43 which was in accordance with results from Leung *et al.* (2012) as these serotypes were predicted *in silico* to be non-amplifiable. However, no *cpsB* amplification was obtained with serotype 29 although it was expected to be amplifiable according to Leung *et al*. (2012). Therefore, 6 other serotype 29 isolates were selected and subjected to sequetyping. All 6 samples led to *cpsB* amplification. After sequencing and assembling, the average sequence length was 890 bp which is longer than the 732 bp region used by Leung *et al*., (2012) to test all their serotypes.

One hundred eighteen sequences representing 78 serotypes were subjected to blast for identification. Two different databases were chosen for the analysis: the exhaustive NCBI GenBank database and a more restrained but specific *cpsB* database created by Jin *et al*., (2016). Using the GenBank database, 61/124 (49%) were identified at the serotype level, 20/124 (16%) were identified at the serogroup level, 14/124 (11%) were identified at the subset level and 23/124 (19 %) were misidentified. Using the Jin *cpsB* database, 65/124 (52%) were identified at the serotype level, 20/124 (16%) were identified at the serogroup level, 12/124 (10%) were identified at the subset level and 21/124 (17%) were misidentified (Table 1A). Inconsistent results were obtained for some serotypes (6B, 6C, 19F and 23F) when using the GenBank database but not using the Jin *cpsB* database. Considering only serotypes, identification with the GenBank database resulted in 35/83 (42%) identifications at the serotype level, 12/83 (14.5%) identifications at the serogroup level, 12/83 (14.5%) identifications at the subset level and 14/83 (17%) misidentified. With the Jin *cpsB* database, 38/83 (46%) were identified at the serotype level, 14/83 (17%) were identified at the serogroup level, 12/83 (14%) were identified at the subset level and 13/83 (16%) were misidentified. Results were slightly better with the Jin *cpsB* database (Table 1B), in particular for inconsistent results.

The majority of misidentifications were due to the attribution of closely related serotypes of the same genogroup (27). For example, one serotype 9A isolate was identified as serotype 9V, one serotype 11F isolate was identified as serotype 11C and one serotype 42 isolate was identified as serotype 35B/35C see table S2 in supplemental material for a complete and detailed list). For some misidentifications, there was no association between the determined serotype and the expected one. This was the case for one serotype 15C isolate identified as serotype 24F, one serotype 19F isolate identified as serotype 10A and one serotype 17A isolate was identified as serotype 10A. Serotype 29 isolates were all misidentified as serotype 35B/35C. Although these serotypes are genetically close, the percent similarity of our serotype 29 *cpsB* sequence compared with the serotype 29 reference sequence was only 83%.

Only one *S. pseudopneumoniae* isolate led to the amplification of *cpsB*. This sequence was associated with serotype 20 with 96% similarity which was the lowest score across all isolates.

Whole genome sequencing

The number of paired-end reads obtained varied between 100 065 and 1 153 346 with an average of 542 388. Whereas some values appeared to be low, assembling metrics generated by Quast highlight a good sequencing quality in general (see table S3 in supplemental material). Assembling coverage varied from 14X to 296X with an average of 94X.

Serotype identification was mainly based on sequence identity level and HSP length (see Table S4). For 53 of 88 isolates (60%), serotype was correctly determined without any ambiguity. The serogroup was determined for 25 of 88 isolates (28%), 6 of 88 (7%) were determined at the subset level and 4 of 88 isolates (5%) were misidentified. Considering serotypes, they were correctly determined for 29 of 53 serotypes (55%), serogroup was correctly determined for 13 of 53 serotypes (25%), 6 of 53 (11%) were determined at the subset level and 3 of 51 (6%) were misidentified. Inconsistent results were obtained for isolates of serotype 6B and 7F, representing 3% of the serotypes tested.

For some isolates, Blast results could not discriminate between two different serotypes because of their high degree of genetic similarities or due to the existence of DNA polymorphism among single serotypes (28). This was the case for 15B/15C, 22A/22F, 7A/7F, 11A/11D, 25A/25F, 32A/32F, 33A/33F, 9A/9V, 12A/46, 12F/44, 18B/18C and 35A/35C/42.

The *cps* locus sequence of misidentified isolates (serotypes 6D, 7F and 29) were aligned with the corresponding best hit reference sequence given by the in-house serotyping method and with the expected serotype sequence (Figures 2A to 2C). No significant hit with 18B reference sequence was found for the misidentified serotype 18B isolate. Therefore, the *cps* locus was aligned with the best hit reference sequence (Figure 2D).

The *cps* locus alignment of our serotype 29 isolate resulted in fragmented hits with low identity compared with the serotype 29 reference sequence. The region 1174-2915 bp of our serotype 29 isolate sequence did not match with both serotype 29 and serotype 35B reference sequences and coded for a *tnp* transposase. It appeared that the *cps* locus of the serotype 29 isolate was located at the end of the corresponding contig and may be incomplete, resulting in a 1303 bp shorter sequence compared to the serotype 29 reference sequence. A very poor alignment was also obtained for our serotype 7F isolate *cps* locus sequence compared with the serotype 7F reference sequence, with less than 50% of the *cps* locus sequence correctly aligned. For the serotype 6D isolate, the major difference between the 2 alignments was the absence of a match with the serotype 6D reference sequence in the 5170-6608 bp region coding for the glycosyl transferase *wciN*. PneumoCaT was also used for serotype attribution using the same set of WGS data (reads data). The first hit was always considered for the prediction of the serotype. If a capsular typing variant analysis occurred, the serotype resulting from this analysis was retained for the serotype prediction. Sixty-one of 87 isolates (70%) were successively identified at the serotype level but all the others isolates (30%) were misidentified. Considering only serotypes, 31 of 52 serotypes (60%) were identified at the serotype level and 19 of 52 (36%) were misidentified. Inconsistent results were obtained for 2 serotypes (7F and 11A), representing 4% of the serotypes tested.

Discussion

S. pneumoniae serotyping has become critical since the release of the different PCV for the monitoring of putative emergent NVT. Unfortunately, the gold standard Quellung method is expensive and laborious and can lead to interpretation errors. The implementation of a new and reliable serotyping method is needed, especially for surveillance programs such as the provincial surveillance held at the Laboratoire de santé publique du Québec.

In this study, 3 different molecular based serotyping methods (sequential multiplex PCR, sequetyping and WGS) were compared in order to evaluate their efficiency in serotype attribution for *S. pneumoniae* invasive isolates. This is the first comparison between these 3 methods on a common set of isolates.

PCR methods are very powerful, reliable and easy to perform. Multiplex PCR is an even more efficient technique since one single reaction allows the simultaneous detection of more than one gene and/or allele. The CDC sequential multiplex PCR method gave the expected results, with 27%, 29% and 20% correct identifications of the serotype, serogroup and subset, respectively. This was also the method presenting the least misidentified isolates (1%). However, serotypes among a serogroup are inevitably revealed under the same signal in the current protocol due to their high level of genetic homogeneity. For example, primer pair 6A/6B/6C/6D in reaction 1 is simultaneously specific to four different serotypes. This is the most important limit for the efficiency of this method because no better results can be expected. Moreover, a significant number (23%) of serotypes were not detectable by this method, representing another limitation from a surveillance perspective. It also seems that small genetic variations in some isolates (serotype 35A) could determine the presence or absence of amplicon (29). It is possible that the isolates tested were genetic variants of the CDC isolates of serotype 35A and that the primers 35A/35C/42 were unable to match these isolates. This finding would mean that the method efficiency could vary from one geographic region to another depending on the genetic distance with the isolates used for primer design. Another important aspect is the specificity of the method for *S. pneumoniae*. Indeed, it is not uncommon to confuse *S. pneumoniae* with other *Streptococcus spp.* due to their high degree of similarity, especially *S. pseudopneumoniae* (30). Here, the internal control (*cpsA*) allowed differentiation between *S. pneumoniae* and *S. pseudopneumoniae* or *S. mitis*. However, 2 serotypes (25F and 38) were also negative for *cpsA* amplification making this discrimination not fully reliable. Finally, non-specific amplifications occurred during the study, as specified by the CDC (https://www.cdc.gov/streplab/pcr.html). Although most of the non-specific products did not match with expected amplifications, some of them could lead to misidentification.

Sequetyping is not limited to the number of detectable serotypes as *cpsB* sequences of almost all serotypes are present in regularly updated public database. Nevertheless, *cpsB* is not amplifiable in all serotypes, making these serotypes not identifiable with this method. This was the case for serotypes 25F, 37, 38, 39 and 43 in our study. Sequences for serotypes 39 and 43 were predicted to be non-amplifiable by Leung *et al.* (2012) even though they were amplified in their study. However, they did not obtain any amplification for serotype 25F or 38, which is consistent with our results. Finally, serotype 37 *cpsB* sequence was predicted to be amplifiable but was not tested *in vitro* in their study.

We decided to use the local *cpsB* sequence database created by Jin *et al*. (2016) instead of the database used by Leung *et al*. (2016) because this database was more comprehensive and covered more serotypes. Overall, we obtained more identification at the serotype level and less misidentifications using the local *cpsB* database as compared to the GenBank database. Significant differences were obtained for serotypes 6B, 6C, 19F and 23F where results between isolates of the same serotype were concordant with the *cpsB* database but not with GenBank database. Only well characterized sequences with full-length *cpsB* were chosen for this database and can explain these results. Indeed, slight variations in the *cpsB* sequence could have a major influence on serotype attribution when the GenBank database is used due to a lot of *cpsB* sequences presenting nucleotide variations not representative of the serotype. In contrast, the use of a local *cpsB* database with few but representative sequences avoided these mistakes. Apart from serotypes 12F, 17A, 18C, 24F, 29 and 35A, no equivalent data are available in Leung *et al*., (2012) for the other misidentified serotypes we observed in this study. For serotype, serogroup, and subset levels identification, our results are generally the same as the ones obtained by Leung *et al*., (2012). However, Comparisons are not always possible since 38 of our serotypes are missing in the Leung *et al*., (2012) study. Most of misidentified serotypes had some nucleotides of difference (from 1 to 59) with the best hit sequence, usually of the same serogroup or genogroup (27). This is caused by intra-serotype variation (28) in the *cps* regulatory region and can lead to identification in the wrong serogroup. This issue has already been observed by Leung *et al*., (2012) with one 19F isolate identified as a serotype 1. Furthermore, some serotypes may have identical *cpsB* sequences as it is the case with some 6A and 6B isolates (31). Moreover, for our serotypes 17A and 29 isolates, no significant hits were obtained with serotype 17A and 29 *cpsB* sequences, respectively. *S. pneumoniae* genome diversity may be high between geographically distant locations, leading to divergence between serotype 17A and 29 *cpsB* sequences present in the databases and sequences obtained in this study. However, this appears to be very unlikely (32). Our evaluation of the sequetyping approach has demonstrated that this serotyping method is not always able to correctly identify serotype probably due to short DNA sub region of a large locus used in this analysis. Of the 6 other non-*S. pneumoniae* isolates tested, only one *S. pseudopneumoniae* led to a *cpsB* amplicon. This was not expected as it has been reported that *S. pseudopneumoniae cps* locus is not complete compared to *S. pneumoniae* and does not contain *cpsB* (33). However, the low identity of the best HSP (96%) could help to discriminate this isolate. A recent method based on sequetyping including a second analysis step for homologous strains allowed to obtain more accurate results for these strains (34). Such protocol could putatively help to obtain better results and make sequetyping more attractive.

Two different approaches were used for serotype identification using WGS method. Our in-house workflow consisted in assembling contigs from sequencing data and to Blast them with a *cps* loci sequence database. Eighty-two percent of serotypes were identified at the serotype or serogroup level, demonstrating the efficiency of this strategy. Regarding unresolved serotypes (7A/7F, 9A/9V, 11A/11D, 12A/12F/44/46, 18B/18C, 22A/22F, 25A/25F, 32A/32F, 33A/33F and 35A/35C/42), these were all identified as another serotype belonging to the same genogroup as defined by Kapatai *et al.*, 2016. More sensitive genetic analysis methods would be required to make a more accurate identification such as the capsular variant analysis integrated in PneumoCaT (see below). Interestingly, serotype 22F isolates matched serotypes 22F/22A but with two separate HSPs. This unexpected Blast result is caused by the high divergence of two genes (*wcwA* and *wcwC*) in the *cps* locus of those isolates compared to their orthologous sequences in serotype 22F. Similar finding were reported for isolate 1772-40b (GenBank accession HE651318; Salter *et al*., 2012), a 22F serotype which matches perfectly with our 22F isolates.

A serotype 29 isolate was misidentified with WGS and identified as serotype 35B. Serotype 35B and 29 are known to be genetically related, leading to cross-reactivity in antisera reactions (35). However, no significant hit with serotype 29 was found in Blast results, meaning that no relevant alignment could be made. These results were in agreement with sequetyping results obtained for serotype 29 isolates. Alignment with serotype 29 reference sequence (isolate 34373, Bentley *et al*., 2006) showed low identity although the serotype was confirmed by Quellung. Transposase coding region (*tnp*) was found downstream the *dexB* gene in the serotype 29 isolate. According to Bratcher *et al*., 2011, those regions may contribute to the vertical exchange of the *cps* locus between pneumococcal isolates and hence to their molecular evolution and adaptation, which could explain the low identity with serotype 29 reference sequence. Serotypes 6D and 6B belong to the same genogroup. However, the glycosyl-transferase *wciN* is present in the 6B *cps* locus and not in the 6D *cps* locus, distinguishing those (36). This gene was present in the studied serotype 6D isolate, which explains the misidentification with serotype 6B. It has been suggested that serotype 6D could have emerged from recombination between serotypes 6B and 6C but Song *et al.* (2011) highlighted the implausibility of this event because of a high genetic distance between these serotypes. Therefore, this gene acquisition was probably due to homologous recombination events or horizontal genetic transfers. The misidentification of serotype 7F isolate with serotype 14 and serotype 18B with 7B were surprising as these 2 serotypes belong to different genetic clusters (Kapatai *et al*., 2016).

PneumoCaT is the second approach we used for WGS serotyping and totally integrates a capsular variant analysis step in its workflow. Single Nucleotide Polymorphisms (SNPs) analysis, allelic variations or presence/absence of genes are analyzed when more than one locus is matched or if the match corresponds to a defined genogroup (25). Although the first step gave results similar to the results obtained with the assembly-based approach, the variant-based step missed the correct serotype for half of the serotypes tested. However, PneumoCaT attributed the correct serotype for 8 serotypes (7A, 9V, 12A, 12F, 15C, 22A, 22F and 33F) which were only identified at the serogroup level or subset with the assembly-based approach.

The aim of this study was to evaluate 3 DNA-based *S. pneumoniae* serotyping methods which could eventually replace the current Quellung gold standard method. Above all, none of the methods tested showed enough efficiency to be able to completely replace the Quellung method in surveillance programs. Indeed, identifications at the serogroup level were obtained with all of them but more particularly with sequential multiplex PCR. Though WGS produces reliable serotyping results, currently this method is still costly and time consuming. Nevertheless, with the automation of bioinformatic pipelines and the constant drop of reagent costs, this method could become very attractive for monitoring invasive *S. pneumoniae serotypes*. Moreover, WGS allows the analysis of molecular evolution of the isolates, the identification of putative vaccine targets in addition to the study of antibiotic resistance and virulence genes. The sequential multiplex PCR and sequetyping strategy unlike WGS have specifically been developed to improve the serotyping response time and to reduce the associated costs. PCR has the inconvenience of requiring an adaptation to the local epidemiology of circulating serotypes. Simply changing the sequential order of the reaction may be sufficient but more often reviewing the combination of primers in the reaction mixture is needed.

In this study, we have demonstrated that WGS was the most reliable method among the 3 methods tested for serotyping of *S. pneumoniae*. However, serotype validation with Quellung is still required as some serotypes cannot be clearly distinguished with the *cps* sequences. Sequential multiplex PCR and sequetyping have the advantage to be cheaper than WGS and could also serve as a guide for Quellung method. But these methods have drawbacks making them less attractive. It is important to note that rare untypeable isolates, due to their lack of capsular polysaccharide, may generate a positive result with DNA based method (37). In such cases, the final serotype identification would be in disagreement with the Quellung reaction which would produce a negative result. Conversely, the sequetyping or multiplex PCR approach may be used when the capsular swelling of the Quellung reaction is difficult to observe through microscopic examination. Finally, a total replacement of the Quellung reaction by a molecular method seems not possible yet. Nevertheless, WGS appears to be a very promising tool and could replace the Quellung method in the near future with its extensive use and the development of databases.

Acknowledgements

We want to thank all the clinical laboratories of the province of Quebec for their participation in the *S. pneumoniae* surveillance program. We also thank the LSPQ and NML personnel for their precious technical assistance.

Disclaimer

This study was partially funded by Pfizer Canada (grant IIR number WI197603)

References

1. Feikin DR, Schuchat A, Kolczak M, Barrett NL, Harrison LH, Lefkowitz L, et al. Mortality from invasive pneumococcal pneumonia in the era of antibiotic resistance, 1995-1997. Am J Public Health. 2000 Feb;90(2):223–9.

2. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. The Lancet. 2010 Jun;375(9730):1969–87.

3. Camargo DRA, Pais FS, Volpini ÂC, Oliveira MAA, Coimbra RS. Revisiting molecular serotyping of *Streptococcus pneumoniae*. BMC Genomics. 2015;16 (Suppl 5):S1.

4. Deng X, Church D, Vanderkooi OG, Low DE, Pillai DR. *Streptococcus pneumoniae* infection: a Canadian perspective. Expert Rev Anti Infect Ther. 2013 Aug;11(8):781–91.

5. Munoz-Almagro C, Jordan I, Gene A, Latorre C, Garcia-Garcia JJ, Pallares R. Emergence of Invasive Pneumococcal Disease Caused by Nonvaccine Serotypes in the Era of 7-Valent Conjugate Vaccine. Clin Infect Dis. 2008 Jan 15;46(2):174–82.

6. Jauneikaite E, Tocheva AS, Jefferies JMC, Gladstone RA, Faust SN, Christodoulides M, et al. Current methods for capsular typing of *Streptococcus pneumoniae*. J Microbiol Methods. 2015 Jun;113:41–9.

7. Sørensen UB. Typing of pneumococci by using 12 pooled antisera. J Clin Microbiol. 1993 Aug;31(8):2097–100.

8. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes. PLoS Genet. 2006 Mar;2(3).

9. Carvalho M da G, Pimenta FC, Jackson D, Roundtree A, Ahmad Y, Millar EV, et al. Revisiting Pneumococcal Carriage by Use of Broth Enrichment and PCR Techniques for Enhanced Detection of Carriage and Serotypes. J Clin Microbiol. 2010 May 1;48(5):1611–8.

10. Shakrin NNSM, Balasubramaniam SD, Yusof HA, Mastuki MF, Masri SN, Taib NM, et al. Evaluation of PCR-based approach for serotype determination of *Streptococcus pneumoniae*. Trop Biomed. 2013 Jun;30(2):338–44.

11. Siira L, Kaijalainen T, Lambertsen L, Nahm MH, Toropainen M, Virolainen A. From Quellung to multiplex PCR, and back when needed, in pneumococcal serotyping. J Clin Microbiol. 2012 Aug;50(8):2727–31.

12. Slinger R, Hyde L, Moldovan I, Chan F, Pernica JM. Direct *Streptococcus pneumoniae* real-time PCR serotyping from pediatric parapneumonic effusions. BMC Pediatr. 2014 Jul 24;14:189.

13. Yu J, Lin J, Kim K-H, Benjamin WH, Nahm MH. Development of an automated and multiplexed serotyping assay for *Streptococcus pneumoniae*. Clin Vaccine Immunol CVI. 2011 Nov;18(11):1900–7.

14. Everett DB, Cornick J, Denis B, Chewapreecha C, Croucher N, Harris S, et al. Genetic Characterisation of Malawian Pneumococci Prior to the Roll-Out of the PCV13 Vaccine Using a High-Throughput Whole Genome Sequencing Approach. PLoS ONE. 2012 Sep 10;7(9).

15. Gladstone RA, Jefferies JM, Tocheva AS, Beard KR, Garley D, Chong WW, et al. Five winters of pneumococcal serotype replacement in UK carriage following PCV introduction. Vaccine. 2015;33(17).

16. Metcalf BJ, Gertz Jr. 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 Jan;22(1):60.e9-60.e29.

17. Al-Sheikh YA, Gowda LK, Marie MAM, John J, Dabwan KHM, Cs P. Distribution of Serotypes and Antibiotic Susceptibility Patterns Among Invasive Pneumococcal Diseases in Saudi Arabia. Ann Lab Med. 2014;34(3):210.

18. Dube FS, van Mens SP, Robberts L, Wolter N, Nicol P, Mafofo J, et al. Comparison of a Real-Time Multiplex PCR and Sequetyping Assay for Pneumococcal Serotyping. PloS One. 2015;10(9):e0137349.

19. Ogami M, Hotomi M, Togawa A, Yamanaka N. A comparison of conventional and molecular microbiology in detecting differences in pneumococcal colonization in healthy children and children with upper respiratory illness. Eur J Pediatr. 2010 Oct;169(10):1221–5.

20. Richter SS, Heilmann KP, Dohrn CL, Riahi F, Diekema DJ, Doern GV. Evaluation of Pneumococcal Serotyping by Multiplex PCR and Quellung Reactions. J Clin Microbiol. 2013 Dec 1;51(12):4193–5.

21. Shaaly A. Comparison of serotyping, pulsed field gel electrophoresis and amplified fragment length polymorphism for typing of *Streptococcus pneumoniae*. J Med Microbiol. 2005 May 1;54(5):467–72.

22. Wen Z, Liu Y, Qu F, Zhang J-R. Allelic Variation of the Capsule Promoter Diversifies Encapsulation and Virulence In *Streptococcus pneumoniae*. Sci Rep. 2016 Jul 28;6:30176.

23. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012 May;19(5):455–77.

24. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072–5.

25. 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.

26. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J. ACT: the Artemis comparison tool. Bioinformatics. 2005 Aug 15;21(16):3422–3.

27. Mavroidi A, Aanensen DM, Godoy D, Skovsted IC, Kaltoft MS, Reeves PR, et al. Genetic Relatedness of the *Streptococcus pneumoniae* Capsular Biosynthetic Loci. J Bacteriol. 2007 Nov 1;189(21):7841–55.

28. Varvio S, Auranen K, Arjas E, Mäkelä PH. Evolution of the Capsular Regulatory Genes in *Streptococcus pneumoniae*. J Infect Dis. 2009 Oct;200(7):1144–51.

29. Wu J-H, Hong P-Y, Liu W-T. Quantitative effects of position and type of single mismatch on single base primer extension. J Microbiol Methods. 2009 Jun;77(3):267–75.

30. Arbique JC, Poyart C, Trieu-Cuot P, Quesne G, Carvalho M d. GS, Steigerwalt AG, et al. Accuracy of Phenotypic and Genotypic Testing for Identification of *Streptococcus pneumoniae* and Description of Streptococcus pseudopneumoniae sp. nov. J Clin Microbiol. 2004 Oct 1;42(10):4686–96.

31. Elberse K, Witteveen S, van der Heide H, van de Pol I, Schot C, van der Ende A, et al. Sequence Diversity within the Capsular Genes of *Streptococcus pneumoniae* Serogroup 6 and 19. Lin B, editor. PLoS ONE. 2011 Sep 16;6(9):e25018.

32. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, et al. Comparative Genomic Analyses of Seventeen *Streptococcus pneumoniae* Strains: Insights into the Pneumococcal Supragenome. J Bacteriol. 2007 Nov 15;189(22):8186–95.

33. Skov Sørensen UB, Yao K, Yang Y, Tettelin H, Kilian M. Capsular Polysaccharide Expression in Commensal *Streptococcus* Species: Genetic and Antigenic Similarities to *Streptococcus pneumoniae*. mBio. 2016 Dec 30;7(6):e01844-16.

34. Nagaraj G, Ganaie F, Govindan V, Ravikumar KL. Development of PCRSeqTyping—a novel molecular assay for typing of Streptococcus pneumoniae. Pneumonia [Internet]. 2017 Dec [cited 2017 Jul 26];9(1). Available from: http://pneumonia.biomedcentral.com/articles/10.1186/s41479-017-0032-3

35. Bush CA, Cisar JO, Yang J. Structures of Capsular Polysaccharide Serotypes 35F and 35C of *Streptococcus pneumoniae* Determined by Nuclear Magnetic Resonance and Their Relation to Other Cross-Reactive Serotypes. Schneewind O, editor. J Bacteriol. 2015 Sep 1;197(17):2762–9.

36. Song J-H, Baek JY, Ko KS. Comparison of Capsular Genes of *Streptococcus pneumoniae* Serotype 6A, 6B, 6C, and 6D Isolates. J Clin Microbiol. 2011 May 1;49(5):1758–64.

37. Kaijalainen T, Rintamäki S, Herva E, Leinonen M. Evaluation of gene-technological and conventional methods in the identification of *Streptococcus pneumoniae*. J Microbiol Methods. 2002 Sep;51(1):111–8.

38. Leung MH, Bryson K, Freystatter K, Pichon B, Edwards G, Charalambous BM, et al. Sequetyping: Serotyping *Streptococcus pneumoniae* by a Single PCR Sequencing Strategy. J Clin Microbiol. 2012 Jul 1;50(7):2419–27.

39. Bratcher PE, Park IH, Oliver MB, Hortal M, Camilli R, Hollingshead SK, et al. Evolution of the capsular gene locus of *Streptococcus pneumoniae* serogroup 6. Microbiology. 2011 Jan 1;157(1):189–98.

40. Salter SJ, Hinds J, Gould KA, Lambertsen L, Hanage WP, Antonio M, et al. Variation at the capsule locus, *cps*, of mistyped and non-typable *Streptococcus pneumoniae* isolates. Microbiology. 2012 Jun 1;158(Pt_6):1560–9.

Table 1 Serotype identification results according to the 3 molecular methods tested and considering (A) isolates or (B) serotypes.

Table S1 *S. pneumoniae* isolates and serotypes included in this study.

Table S2 Serotypes and identification level determined using the multiplex PCR and sequetyping methods.

Table S3 WGS and assembly quality metrics.

Table S4 Serotypes and identification level determined with WGS methods. For PneumoCaT, the serotype chosen after the capsule variant analysis step is represented in bold.

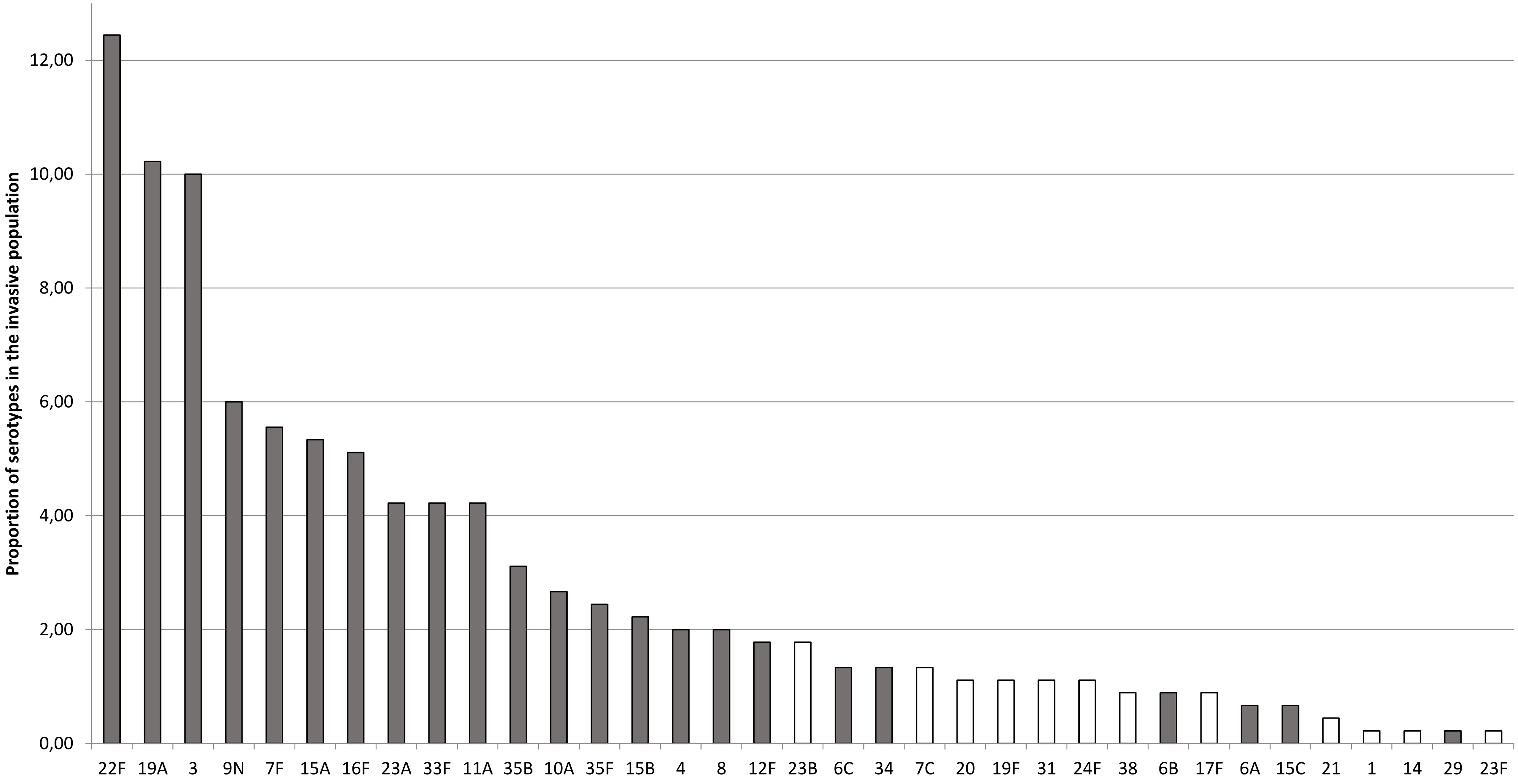


Figure 1: *S. pneumoniae* serotype distribution in the province of Québec in 2016. Grey bars represent serotypes tested by WGS in this study.

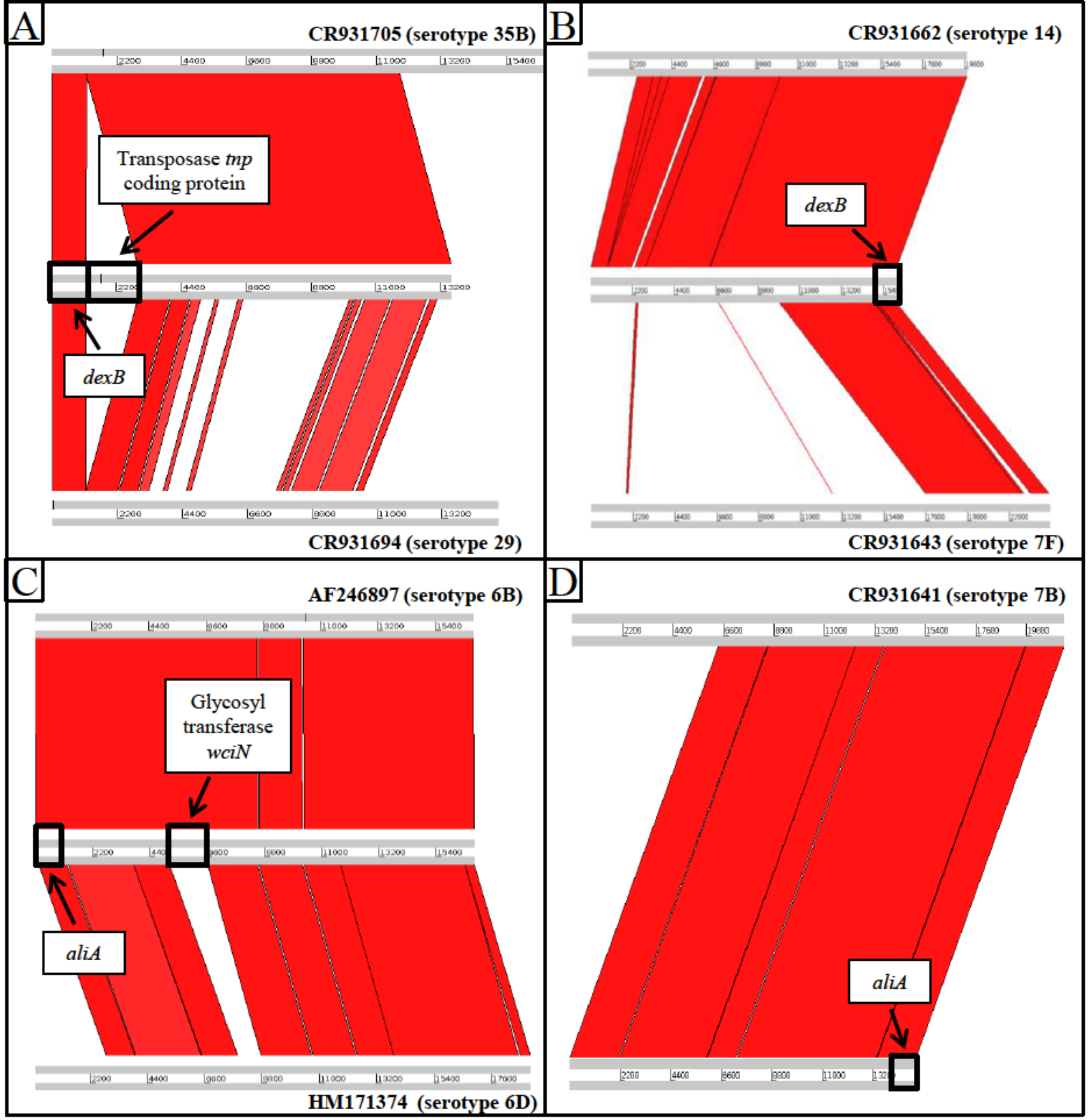


Figure 2: Alignment of *cps* loci of serotype 29 isolate (A), serotype 7F isolate (B), serotype 6D isolate (C) and serotype 18B isolate (D) with reference *cps* sequence and best hit *cps* sequence according to WGS identification. Alignment was generated with Artemis Comparison Tool (http://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act).

A

		Sequetyping		WGS	
	CDC sequential multiplex PCR (n = 130)	NCBI online database (n = 124)	Curated *cpsB* database (n = 124)	Assembling strategy (n = 88)	PneumoCaT (n = 87)(1)
Serotype	35%	49%	52%	60%	70%
Serogroup	32%	16%	16%	28%	0%
Subset	17%	11%	10%	7%	0%
Misidentified	1%	19%	17%	5%	30%
N.D.	15%	5%	5%	0%	0%

B

		Sequetyping		WGS	
	CDC sequential multiplex PCR (n = 83)	NCBI online database (n = 83)	Curated *cpsB* database (n = 83)	Assembling strategy (n = 53)	PneumoCaT (n = 52)(1)
Serotype	27%	42%	46%	55%	60%
Serogroup	29%	14,5%	17%	25%	0%
Subset	20%	14,5%	14%	11%	0%
Misidentified	0%	17%	16%	6%	36%
Inconsistent	1%	6%	1%	3%	4%
N.D.	23%	6%	6%	0%	0%

N.D. = not determinable (not detectable in CDC PCR protocol or *cpsB* not amplified).

(1) One sample analysis failed because of too low reads number

Table 1 Serotype identification results according to the 3 molecular methods tested and considering (A) isolates or (B) serotypes.

Table S1 *S. pneumoniae* isolates and serotypes included in this study.

		Tested serotyping methods		
Serotypes according to Quellung	Isolates ID	Sequential multiplex PCR	Sequetyping	WGS
1	MA096520	✓	✓	
1	MA101323	✓	✓	
1	LSPQ3053	✓	✓	
2	LSPQ3054	✓	✓	
3	MA100130	✓	✓	
3	MA101386	✓	✓	
3	MA080904			✓
3	MA081716			✓
3	MA082307			✓
3	MA086676			✓
3	MA096946			✓
3	LSPQ3055	✓	✓	
3	SC0174			✓
3	SC0286			✓
4	MA100773	✓	✓	
4	MA101744	✓	✓	
4	MA079938			✓
4	LSPQ3124	✓	✓	
5	MA082483	✓	✓	
5	LSPQ3057	✓	✓	
6A	MA099472	✓	✓	
6A	MA101024	✓	✓	
6A	LSPQ3058	✓	✓	
6A	SC0022			✓
6B	MA098599	✓	✓	
6B	MA101145	✓	✓	
6B	LSPQ3770	✓	✓	
6B	SC0023			✓
6B	SC0169			✓
6C	MA099139	✓	✓	
6C	MA100925	✓	✓	
6C	LSPQ4242	✓	✓	
6C	SC0262			✓
6D	MA092686	✓	✓	
6D	SC0129			✓
7A	LSPQ4102	✓	✓	
7A	SC0025			✓
7B	LSPQ4103	✓	✓	
7C	LSPQ4231	✓	✓	
7F	MA093680	✓	✓	
7F	MA097140	✓	✓	
7F	MA099461	✓	✓	
7F	MA081946			✓
7F	SC0218			✓
8	LSPQ3596	✓	✓	
8	SC0028			✓
9A	MA080418	✓	✓	

Table S1 *S. pneumoniae* isolates and serotypes included in this study.

		Tested serotyping methods		
Serotypes according to Quellung	Isolates ID	Sequential multiplex PCR	Sequetyping	WGS
9A	SC0029			✓
9L	LSPQ4271	✓	✓	✓
9L	SC0011			✓
9N	MA098250	✓	✓	
9N	MA100245	✓	✓	
9N	MA080879			✓
9N	MA081113			✓
9N	MA099463	✓	✓	
9N	SC0031			✓
9V	MA097827	✓	✓	
9V	MA098806	✓	✓	
9V	MA099234	✓	✓	
9V	SC0172			✓
10A	MA090174	✓	✓	
10A	MA095845			✓
10A	MA094933			✓
10A	MA094205			✓
10B	MA080812	✓	✓	
10F	MA075627	✓	✓	
11A	MA090298	✓	✓	
11A	MA091851			✓
11A	SC0035			✓
11B	MA096566	✓	✓	
11C	LSPQ4272	✓	✓	✓
11D	SC0271			✓
11F	MA073130	✓	✓	
12A	MA097699	✓	✓	
12A	SC0066			✓
12B	SC0268			✓
12F	LSPQ3064	✓	✓	
12F	SC0199			✓
13	LSPQ3065	✓	✓	
14	MA096954	✓	✓	
14	MA098680	✓	✓	
14	LSPQ3066	✓	✓	
15A	MA100658	✓	✓	
15A	MA101766	✓	✓	
15A	MA080018			✓
15A	MA099389	✓	✓	
15A	MA096792			✓
15A	MA095336			✓
15A	MA094663			✓
15A	MA093977			✓
15A	SC0042			✓
15B	MA099177	✓	✓	
15B	MA096033			✓
15B	MA095997			✓

Table S1 *S. pneumoniae* isolates and serotypes included in this study.

		Tested serotyping methods		
Serotypes according to Quellung	Isolates ID	Sequential multiplex PCR	Sequetyping	WGS
15B	MA094560			✓
15B	SC0044			✓
15C	MA096496	✓	✓	
15C	SC0045			✓
15F	MA083248	✓	✓	
16F	MA065427	✓	✓	
16F	LSPQ4236	✓	✓	
16F	MA093020			✓
17A	LSPQ4273	✓	✓	✓
17F	MA098807	✓	✓	
18A	LSPQ4243	✓	✓	
18A	SC0009			✓
18B	MA066814	✓	✓	
18B	SC0049			✓
18C	MA093772	✓	✓	
18C	MA099660	✓	✓	
18C	MA095139	✓	✓	
18C	SC0050			✓
18F	LSPQ4274	✓	✓	✓
18F	SC0051			✓
19A	MA101978	✓	✓	
19A	MA083042	✓		
19A	MA084138	✓		
19A	MA083920			✓
19A	MA097921			✓
19A	MA098817			✓
19A	LSPQ3071	✓	✓	
19A	MA080288			✓
19A	MA080125			✓
19A	MA079789			✓
19A	MA083042		✓	
19A	MA084138		✓	
19A	SC0010			✓
19F	MA100764	✓	✓	
19F	MA101680	✓	✓	
19F	MA098992	✓	✓	
20	LSPQ3072	✓	✓	
21	LSPQ3160	✓	✓	
22A	MA095877	✓	✓	
22A	SC0059			✓
22F	MA100780	✓	✓	
22F	MA101987	✓	✓	
22F	MA080654			✓
22F	LSPQ4162	✓	✓	
22F	MA096962			✓
22F	MA094696			✓
22F	MA094689			✓

Table S1 *S. pneumoniae* isolates and serotypes included in this study.

		Tested serotyping methods		
Serotypes according to Quellung	Isolates ID	Sequential multiplex PCR	Sequetyping	WGS
22F	SC0188			✓
22F	SC0291			✓
23A	MA082395			✓
23A	LSPQ3769	✓	✓	
23B	MA099469	✓	✓	
23F	MA100152	✓	✓	
23F	MA101159	✓	✓	
23F	MA099467	✓	✓	
24A	LSPQ4275	✓	✓	✓
24B	MA096695			✓
24B	MA094350	✓	✓	
24F	MA099028	✓	✓	
25F	LSPQ4276	✓	✓	✓
27	MA088547	✓	✓	
28A	MA099752	✓	✓	
28F	LSPQ4277	✓	✓	✓
29	LSPQ3079	✓		
29	MA097586		✓	✓
29	MA098344		✓	
29	MA098505		✓	
29	MA100224		✓	
29	MA101320		✓	
29	LSPQ3079		✓	
29	MA099083		✓	
31	LSPQ3080	✓	✓	
32A	LSPQ4278	✓	✓	✓
32F	LSPQ3081	✓		
32F	LSPQ3081		✓	
33A	MA086628	✓	✓	
33A	SC0082			✓
33B	LSPQ4279	✓	✓	✓
33F	MA080211			✓
33F	MA099238	✓	✓	
33F	SC0190			✓
34	MA101496	✓		
34	MA101843	✓		
34	MA102076	✓		
34	MA102374	✓		
34	MA102487	✓		
34	LSPQ3127	✓	✓	
34	MA099037	✓		
34	MA096961			✓
35A	LSPQ4266	✓		
35A	LSPQ4267	✓		
35A	LSPQ4268	✓		
35A	LSPQ4269	✓		
35A	LSPQ4270	✓		

Table S1 *S. pneumoniae* isolates and serotypes included in this study.

		Tested serotyping methods		
Serotypes according to Quellung	Isolates ID	Sequential multiplex PCR	Sequetyping	WGS
35A	MA101545			✓
35A	MA092229	✓	✓	
35A	MA082642	✓		
35B	MA082394			✓
35B	MA097723	✓	✓	
35C	LSPQ4280	✓	✓	✓
35F	MA081892			✓
35F	MA099195	✓	✓	
36	LSPQ3641	✓	✓	
37	LSPQ3645	✓	✓	
37	SC0086			✓
38	LSPQ3642	✓	✓	
39	LSPQ3646	✓	✓	
40	LSPQ3162	✓	✓	
41A	LSPQ3089	✓	✓	
41F	LSPQ4281	✓	✓	✓
42	LSPQ3677	✓	✓	
43	LSPQ3643	✓	✓	
44	LSPQ3644	✓	✓	
44	SC0212			✓
45	LSPQ3092	✓	✓	
46	LSPQ3093	✓	✓	
46	SC0096			✓
47A	LSPQ4282	✓	✓	✓
47F	LSPQ4283	✓	✓	✓
48	LSPQ3095	✓	✓	
S. mitis (1)	ID112476	✓	✓	
S. mitis (1)	MA084074	✓	✓	
S. mitis (1)	MA084310	✓	✓	
S. pseudopneumoniae (1)	ID111828	✓	✓	
S. pseudopneumoniae (1)	ID112065	✓	✓	
S. pseudopneumoniae (1)	ID112502	✓	✓	

(1) Strains used as controls for specificity

Table S2 Serotypes and identification level determined using the multiplex PCR and sequetyping methods.

	CDC sequential multiplex PCR		Sequetyping (NCBI database)		Sequetyping (*cpsB* database)	
Serotype(1)	Serotype(s) determined	Identification level	Serotype(s) determined	Identification level (Online NCBI database)	Serotype(s) determined	Identification level (Local *cpsB* database)
1	1	Serotype	1	Serotype	1	Serotype
2	2	Serotype	2/41A	Ambiguous	2/41A	Ambiguous
3	3	Serotype	3	Serotype	3	Serotype
4	4	Serotype	4	Serotype	4	Serotype
5	5	Serotype	5	Serotype	5	Serotype
6A	6A/6B	Serogroup	6A	Serotype	6A	Serotype
6B	6A/6B	Serogroup	6B 6F/6B/6A	Serotype (n=1) Serogroup (n=2)	6B	Serotype
6C	6C/6D	Serogroup	6C 6C/6D	Serotype (n=1) Serogroup (n=2)	6C/6D	Serogroup
6D	6C/6D	Serogroup	6C/6D	Serogroup	6C/6D	Serogroup
7A	7A/7F	Serogroup	7A/7F	Serogroup	7A/7F	Serogroup
7B	7B/7C/40	Subset	40/7B	Ambiguous	40/7B	Ambiguous
7C	7B/7C/40	Subset	7C	Serotype	7C	Serotype
7F	7A/7F	Serogroup	7F/7A	Serogroup	7F/7A	Serogroup
8	8	Serotype	8	Serotype	8	Serotype
9A	9A/9V	Serogroup	9V	Misidentified	9V	Misidentified
9L	9N/9L	Serogroup	9L	Serotype	9L	Serotype
9N	9N/9L	Serogroup	9N	Serotype	9N	Serotype
9V	9A/9V	Serogroup	9V	Serotype	9V	Serotype
10A	10A	Serotype	10A	Serotype	10A	Serotype
10B	No amplification	ND	10B	Serotype	10B	Serotype
10F	10C/10F/33C	Subset	10F/10C	Serogroup	10F/10C	Serogroup
11A	11A/11D	Serogroup	11A/11D/18F	Ambiguous	11A/11D/18F	Ambiguous
11B	No amplification	ND	11B/11C	Serogroup	11B/11C	Serogroup
11C	No amplification	ND	11B/11C	Serogroup	11B/11C	Serogroup
11F	No amplification	ND	11C	Misidentified	11F/11C	Serogroup
12A	12A/12F/44/46	Subset	12F	Misidentified	12F	Misidentified
12F	12A/12F/44/46	Subset	12A	Misidentified	12A	Misidentified
13	13	Serotype	13/20	Ambiguous	13/20	Ambiguous
14	14	Serotype	14	Serotype	14	Serotype
15A	15A/15F	Serogroup	15A	Serotype	15A	Serotype
15B	15B/15C	Serogroup	15B	Serotype	15B	Serotype
15C	15B/15C	Serogroup	24F	Misidentified	24F	Misidentified
15F	15A/15F	Serogroup	15A	Misidentified	15A	Misidentified
16F	16F	Serotype	16F	Serotype	16F	Serotype
17A	No amplification	ND	10A	Misidentified	10A	Misidentified
17F	17F	Serotype	17F	Serotype	17F	Serotype
18A	18A/18B/18C/18F	Serogroup	18A	Serotype	18A	Serotype
18B	18A/18B/18C/18F	Serogroup	18B	Serotype	18B	Serotype
18C	18A/18B/18C/18F	Serogroup	18B	Misidentified	18B	Misidentified
18F	18A/18B/18C/18F	Serogroup	11A/11D/18F	Ambiguous	11A/11D/18F	Ambiguous
19A	19A	Serotype	19A	Serotype	19A	Serotype
19F	19F	Serotype	19F 10A	Serotype (n=2) Misidentified (n=1)	19F	Serotype
20	20	Serotype	20/13	Ambiguous	20/13	Ambiguous
	CDC sequential multiplex PCR		Sequetyping (NCBI database)		Sequetyping (*cpsB* database)	
Serotype	Serotype(s) determined	Identification level	Serotype(s) determined	Identification level (Online NCBI database)	Serotype(s) determined	Identification level (Local *cpsB* database)
21	21	Serotype	21	Serotype	21	Serotype
22A	22A/22F	Serogroup	22F/22A	Serogroup	22F/22A	Serogroup
22F	22A/22F	Serogroup	22F/22A	Serogroup	22F/22A	Serogroup
23A	23A	Serotype	23A	Serotype	23A	Serotype
23B	23B	Serotype	23B	Serotype	23B	Serotype
23F	23F	Serotype	23F 14/12/21/23F	Serotype (n=1) Ambiguous (n=2)	23F	Serotype
24A	24A/24B/24F	Serogroup	24A	Serotype	24A	Serotype
24B	24A/24B/24F	Serogroup	24B	Serotype	24B	Serotype
24F	24A/24B/24F	Serogroup	24B	Misidentified	24B	Misidentified
25F	38/25A/25F	Subset	No *cpsB* amplification			
27	No amplification	ND	27	Serotype	27	Serotype
28A	No amplification	ND	28A	Serotype	28A	Serotype
28F	No amplification	ND	28F	Serotype	28F	Serotype

Table S2 Serotypes and identification level determined using the multiplex PCR and sequetyping methods.

	CDC sequential multiplex PCR		Sequetyping (NCBI database)		Sequetyping (*cpsB* database)	
Serotype[1]	Serotype(s) determined	Identification level	Serotype(s) determined	Identification level (Online NCBI database)	Serotype(s) determined	Identification level (Local *cpsB* database)
29	No amplification	ND	No *cpsB* amplification (n=1) 35C/35B	Misidentified (n=6)	35C/35B	Misidentified
31	31	Serotype	31	Serotype	31	Serotype
32A	No amplification	ND	32F/32A	Serogroup	32F/32A	Serogroup
32F	No amplification	ND	32F/32A	Serogroup	32F/32A	Serogroup
33A	33A/33F/37	Subset	33A/33F/35A	Ambiguous	33A/33F/35A	Ambiguous
33B	No amplification	ND	33B	Serotype	33B	Serotype
33F	33A/33F/37	Subset	33F/33A/35F	Ambiguous	33F/33A/35F	Ambiguous
34	34	Serotype	34/17A	Ambiguous	34/17A	Ambiguous
35A	No amplification 35A/35C/42	Misidentified (n = 2) Subset (n = 5)	35C/35B	Misidentified	35C/35B	Misidentified
35B	35B	Serotype	35C/35B	Serogroup	35C/35B	Serogroup
35C	35A/35C/42	Subset	35C/35B	Serogroup	35C/35B	Serogroup
35F	35F/47F	Subset	35F/47F	Ambiguous	35F/47F	Ambiguous
36	No amplification	ND	36	Serotype	36	Serotype
37	33A/33F/37	Subset	No *cpsB* amplification			
38	38/25A/25F	Subset	No *cpsB* amplification			
39	39	Serotype	No *cpsB* amplification			
40	7B/7C/40	Subset	40/7B	Ambiguous	40/7B	Ambiguous
41A	No amplification	ND	41F	Misidentified	41F	Misidentified
41F	No amplification	ND	41F	Serotype	41F	Serotype
42	35A/35C/42	Subset	35C/35B	Misidentified	35C/35B	Misidentified
43	No amplification	ND	No *cpsB* amplification			
44	12A/12F/44/46	Subset	12B	Misidentified	12B	Misidentified
45	No amplification	ND	45	Serotype	45	Serotype
46	12A/12F/44/46	Subset	12A	Misidentified	12A	Misidentified
47A	No amplification	ND	47A	Serotype	47A	Serotype
47F	35F/47F	Subset	47F/35F	Ambiguous	47F/35F	Ambiguous
48	No amplification	ND	48	Serotype	48	Serotype

[1] Serotype determined with Quellung method

Table S3 WGS and assembly quality metrics.

Serotype	Reads numbers	Largest contig (bp)	N50	Mean coverage (X)
3	407 169	161 387	70 238	36
3	579 976	345 480	218 480	204
3	362 116	276 730	167 190	127
3	337 811	243 817	91 651	67
3	283 092	263 351	136 846	58
3	306 962	390 935	205 071	138
3	653 001	463 112	340 013	69
4	256 005	214 530	74 514	45
8	591 174	417 290	141 800	129
29	555 776	196 889	61 494	113
34	271 637	133 241	64 281	30
37	549 878	323 230	85 740	125
44	294 317	444 288	199 942	102
46	480 479	365 956	149 388	115
10A	1 150 155	330 614	115 223	92
10A	759 133	303 524	86 936	80
10A	878 063	303 918	98 395	90
11A	1 068 051	151 627	71 048	77
11A	569 660	333 166	125 531	122
11C	100 065	136 078	58 756	14
11D	610 239	344 620	132 147	136
12A	401 728	176 564	74 831	104
12B	632 729	59 472	10 314	68
12F	510 777	159 692	81 473	113
15A	268 153	247 306	95 807	45
15A	1 153 346	330 076	74 270	71
15A	406 573	176 268	54 348	56
15A	814 704	176 281	65 535	79
15A	735 845	241 467	88 561	73
15A	451 169	198 861	124 602	99
15B	979 655	151 822	80 855	83
15B	985 351	254 966	86 217	71
15B	739 530	169 702	84 611	66
15B	562 415	321 118	105 197	125
15C	522 019	388 494	128 627	120
16F	1 053 173	235 604	113 800	70
17A	155 277	305 746	98 305	17
18A	472 449	744 739	424 607	102
18B	381 836	216 017	82 749	67
18C	489 331	350 049	138 595	103
18F	229 176	197 877	109 713	22
18F	539 149	417 980	393 351	127
19A	729 967	328 634	86 181	131
19A	884 691	355 253	162 090	171
19A	365 612	381 909	163 676	89
19A	406 715	289 688	71 633	31
19A	1 023 720	340 957	71 895	90
19A	540 195	319 774	69 483	43

Table S3 WGS and assembly quality metrics.

Serotype	Reads numbers	Largest contig (bp)	N50	Mean coverage (X)
19A	424 005	300 204	132 496	101
22A	556 649	204 552	86 425	115
22F	404 211	297 023	104 357	92
22F	489 875	207 974	66 632	51
22F	788 724	243 814	86 596	70
22F	301 144	257 300	98 394	35
22F	416 774	276 645	60 906	175
22F	703 146	412 859	151 633	84
23A	429 205	273 953	113 480	93
24A	307 126	192 635	78 434	44
24B	500 784	220 680	88 008	57
25F	545 333	197 175	50 116	82
28F	611 811	239 939	90 168	76
32A	751 230	105 571	55 533	115
33A	514 867	387 226	216 927	109
33B	293 694	247 620	68 772	45
33F	475 948	246 678	140 406	83
33F	466 007	337 630	200 157	87
35A	515 979	286 061	162 953	296
35B	295 082	202 017	101 286	58
35C	285 007	230 492	75 491	35
35F	494 610	299 061	126 588	104
41F	253 861	158 243	72 183	52
47A	382 521	438 741	95 131	32
47F	695 607	171 602	71 324	103
6A	613 299	384 514	143 208	119
6B	632 742	561 617	145 339	137
6B	579 748	367 297	111 737	135
6C	684 939	323 480	144 636	159
6D	517 590	265 830	158 680	129
7A	633 052	143 961	76 126	250
7F	713 798	115 076	67 068	47
7F	553 480	317 583	105 433	133
9A	592 150	379 087	157 384	144
9L	117 140	161 368	49 807	14
9L	486 383	236 509	79 857	128
9N	438 695	345 799	136 064	85
9N	591 387	276 495	85 471	97
9N	634 222	362 428	157 652	141
9V	538 233	326 364	126 138	138
Mean	542 388	280 491	115 625	94

Table S4 Serotypes and identification level determined with WGS methods.
For PneumoCaT, the serotype chosen after the capsule variant analysis step is represented in bold.

Serotype	Assembly method		PneumoCaT	
	Best hits(1)	Identification level	Best hits(1)	Identification level
3	3	Serotype	3	Serotype
4	4	Serotype	4	Serotype
6A	6A	Serotype	6A	Serotype
6B	6A/6B	Serogroup (n=1)	6A-**6E**	Misidentified (n=1)
	6B	Serotype (n=1)	6A	Misidentified (n=1)
6C	6C	Serotype	6D	Misidentified
6D	6B	Misidentified	6A-**6E**	Misidentified
7A	7A/7F	Serogroup	**7A**-7F	Serotype
7F	14	Misidentified (n=1)	14	Serotype (n=1)
	7A/7F	Serogroup (n=1)	7F	Misidentified (n=1)
8	8	Serotype	8	Serotype
9A	9A/9V	Serogroup	9A-**9V**	Misidentified
9L	9L	Serotype	9L-**9N**	Misidentified
9N	9N	Serotype	9L-**9N**	Serotype
9V	9A/9V	Serogroup	9V	Serotype
10A	10A	Serotype	**10A**-10B	Serotype
11A	11A/11D	Serogroup	11A-**11D**	Misidentified (n=1)
			11A	Serotype (n=1)
11C	11C	Serotype	**11C**-11C	Serotype
11D	11D	Serotype	**11A**-11D	Misidentified
12A	12A/46	Ambiguous	**12A**-46	Serotype
12B	12B	Serotype	12A	Misidentified
12F	12F/44	Ambiguous	12F	Serotype
15A	15A	Serotype	15A	Serotype
15B	15B/15C	Serogroup	15B-**15C**	Misidentified
15C	15B/15C	Serogroup	15C	Serotype
16F	16F	Serotype	16F	Serotype
17A	17A	Serotype	17A	Serotype
18A	18A	Serotype	18A	Serotype
18B	7B	Misidentified	7B	Misidentified
18C	18B/18C	Serogroup	18B	Misidentified
18F	18F	Serotype	18F	Serotype
19A	19A	Serotype	19A	Serotype
22A	22F/22A	Serogroup	**22A**-22F	Serotype
22F	22F/22A	Serogroup	22A-**22F**	Serotype
23A	23A	Serotype	23A	Serotype
24A	24A	Serotype	24A	Serotype
24B	24B	Serotype	24B-**24F**	Misidentified
25F	25A/25F	Serogroup	**25A**-25F	Misidentified
28F	28F	Serotype	28F	Serotype

Table S4 Serotypes and identification level determined with WGS methods.
For PneumoCaT, the serotype chosen after the capsule variant analysis step is represented in bold.

Serotype	Assembly method		PneumoCaT	
	Best hits[1]	Identification level	Best hits[1]	Identification level
29	35B	Misidentified	35B	Misidentified
32A	32A/32F	Serogroup	32A-**32F**	Misidentified
33A	33A/33F	Serogroup	33A-**33F**	Misidentified
33B	33B	Serotype	33B	Serotype
33F	33A/33F	Serogroup	33F	Serotype
34	34	Serotype	34	Serotype
35A	35C/42/35A	Ambiguous	**35C**-42	Misidentified
35B	35B	Serotype	35B	Serotype
35C	35C/42	Ambiguous	35F	Misidentified
35F	35F	Serotype	41F	Serotype
37	37	Serotype	37	Serotype
41F	41F	Serotype	41A-**41F**	Serotype
44	12F/44	Ambiguous	12A	Misidentified
46	12A/46	Ambiguous	12A	Misidentified
47A	47A	Serotype	Failed[2]	
47F	47F	Serotype	47F	Serotype

(1) Best hit according to blast score and coverage

SEROTYPE DISTRIBUTION OF INVASIVE STREPTOCOCCUS PNEUMONIAE STRAINS IN ≥ 5 YEARS OLD IN QUÉBEC, 2010-20 6

e ebv e B P De Wals G Deceunin k I Ma t n W Demczuk Ma kowski M Douv l e- adet Doual a-Bell C T emblay J ongt n

In t tut na iona de anté pub ique du Québec abo a oi e de anté pub ique du Québec S in e-Anne de-B l evue C nada nst tut n ti nal de san é pub que du Québ c Di ec ion des sques b o ogiqu s t de a san é au t ava l Qu bec Can da Cent e de eche che du cent e hosp ta ie un ve s tai e de Qu bec Québec Canada Nat onal Mic obi logy a o a o y Pu l c H al h Ag ncy o Canada Winn peg Canada Min s è e de la san é et des se vic s soc aux du Québ c Canada

Background and ob ect ves

The Queb c la o a o y u ei ance p og am o nva ive p eumococ al d s ase IPD) s un ve s l o at en s aged < 5 yea s o d and a so nc udes a im ted ent ne ne wo k o pa ie ts ≥ 5 yea s o ge his etwo k in lu es 21 abo ato es en om as ing pe iat c ho pi a s and ac ounts o abo t 35% o he i vas ve t ai s n pa ie t ≥ 5 yea s o a e n Can da t e PC -13 va c ne s cu e t y u ed in ch ld en nd s app oved o hose ≥ 5 ea s o d w th sp ci c ndi a ions w i e PPV 23 is e ommend d o p t en s ≥ 65 yea s o d In o de o inc e se ac u cy the su ve la ce p og m w s onve ed o a un ve sal p og am n 014 n h s tudy we e a uat d he b ne i s o mon o ng all *S neumon ae* nva i e t a ns o he p vi ce in p t en s aged ≥ 5 yea s om a ed o sen i el ne wo k mon to ng

Me hods

St a ns bo a o es we e in i ed o send a l *S p eumon ae* st a ns om a no ma ly s e le s te to he SPQ A t tal o 377 *S pneumon ae* nvas ve t a ns (1 t ain pat ent 14 day) i ol ted n 20 4-20 6 we e ana yzed Se o yp ng Se oty es we e de e min d by Que lung ea t on us ng an ise a om S at ns e um ns i ut

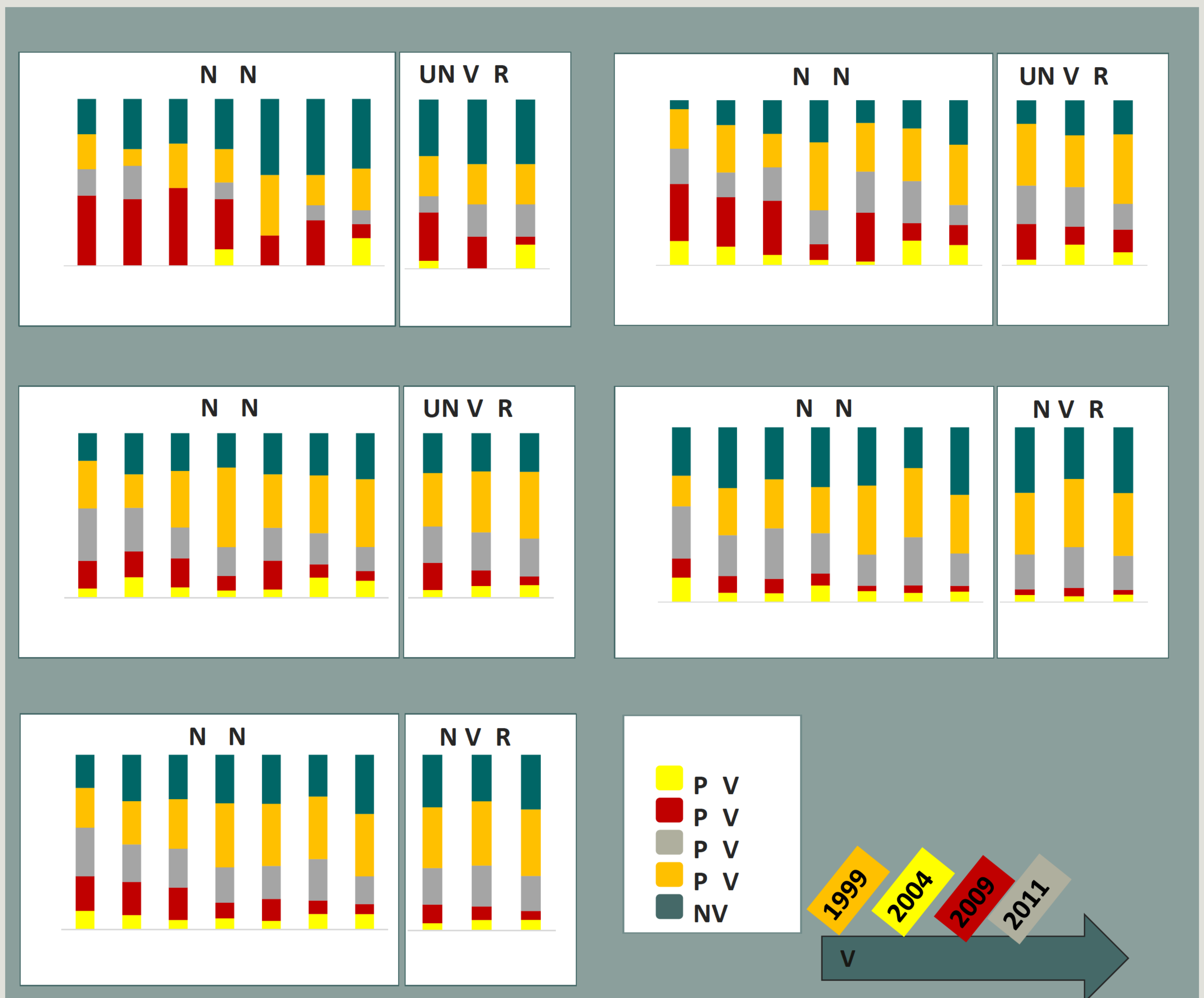
Results

TAB E 1 D st ib ti n o *S pneumon ae* se oty es in sen ine *vs* unive sa si es

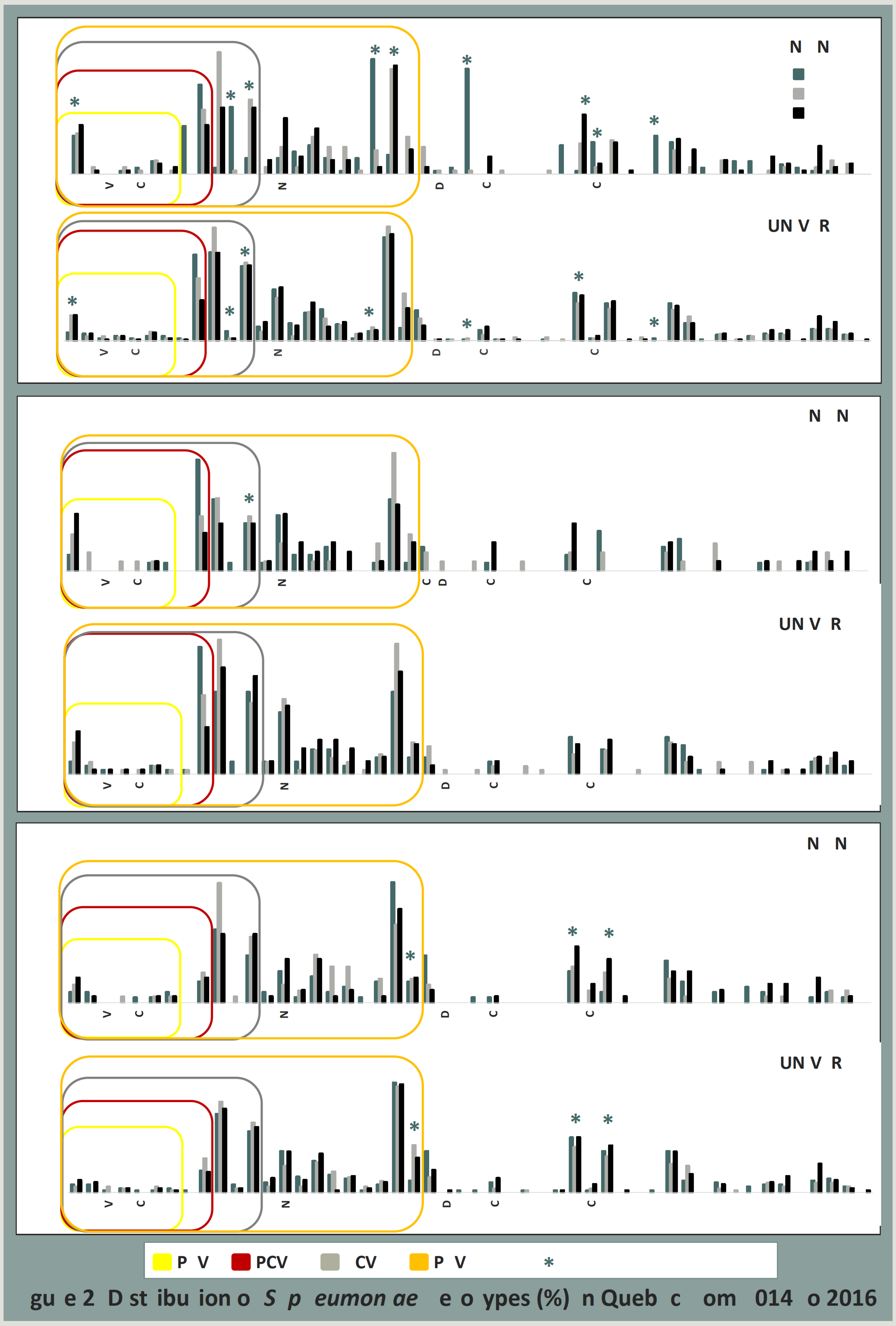
Se t nel	P po i n y ge g oup (%)					
	< 5	5 19	2 -49	5 -64	≥ 65	TO A
PCV 7	1 8	5 9	9 8	9 0	5 8	6 7
PCV 10 *	0 0	17 6	17 4	10 8	3 6	7 6
PCV 13 **	12 4	5 9	21 2	18 0	21 2	18 6
PPV 23 †	55 8	26 5	32 6	36 0	37 6	8 4
NVT	30 1	44 1	18 9	26 1	31 9	8 7
TO A	100 0	00 0	1 0 0	1 0 0	1 0 0	100 0

Non se t nel	P po i n y ge g up (%)					
	< 5	5 19	0- 9	0- 4	≥ 65	TOTA
PCV 7	1 2	6 9	7 0	5 2	2 9	4 1
PCV 10 *	1 2	17 2	14 1	9 9	3 5	6 9
PCV 13 **	17 9	27 6	21 6	25 4	20 9	22 2
PPV 23 †	50 0	20 7	38 8	37 3	36 4	37 4
NVT	29 8	27 6	18 5	22 2	36 4	29 4
TO A	100 0	00 0	00 0	00 0	00 0	00 0

PCV PCV P V



gu e 1 Se oty e d st but on by age g oup n ent nel *vs* un ve al su vei ance o in as ve *S pneumon ae*



gu e 2 D st ibu ion o *S p eumon ae* e o ypes (%) n Queb c om 014 o 2016

Summary

In 2014 2016 he p opo ion o NVT was highe n ent nel t an n non sen inel i es o the 5 19 y lds 44 s 8%) whe eas it was 32% and 6% espec i ely in the ≥ 65 age g oup (able 1) A ec e se n PCV 13 se ot pes was obse ved be w en 2010 and 2016 whi e NVT and add t onal PPV- 3 se ot pes nc eased e pec al y among 5- 9 y pa ien s om ent nel s tes (ig 1) By 016 the 5 most equent se ot pes we e 22 (3%) 3 11%) 19A (9%) 9N (6%) and 15A (6%) n ≥ 5 y om all ho pi als (ig 2) P opo t ons o PCV 13 add t onal PPV-23 se otyp s nd NVT (non vac ine se oty e) we e 26% 36% 38% e pec ive y n the ≥ 5 y

Conclusion

The ove - ep ese tat on o pedi t c pat ents wi h omo bid ty in the s nt nel ne wo k may be es onsi le o he ighe p opo ion o NVT obse ved n his g oup o his ea on the e haus ive su ve l ance o tho e < 5 y n eds o be con inued volu ion o ci c la ing se otyp s n at ents ≥ 5 y d d ot u ly co e ate w th the sen inel p og am A un ve sal su vei ance p ovid s be te ac u a y n an e a o inc eas d PCV 13 use in adu ts

References

1 h tps /www nspq qc a/p bl c t ons 2254

2 h tp / pub i at ons ms s gouv qc a/ms s/ ch e / pi /p q_c m let d

3 h tp / www pha -a pc gc c /pub ca /c d - m c/13 ol 9/a s dcc 5 as et /pd /1 vo 39 acs d c5- a d

4 Au t ian R 1 76 he que lu g ea t on a n gec ed m c ob ol gic e hni ue Mt Si ai J M d 3 99 709

Acknowledgments

his wo k w s a i l y suppo ed y P ze

A sp c al t anks o a l t e l n cal abo a o es o he p ov nce as w ll s the PQ nd NM

Institut national de santé publique Québec

Laboratoire de santé publique du Québec