Genetic variability of the pneumococcal *wciZ*-gene and how it relates to serotype prediction

Andreas Itzek and Mark van der Linden
German Reference Laboratory for Streptococci (GRLS), Institute of Medical Microbiology, RWTH-Aachen, Germany

BACKGROUND

Invasive pneumococcal disease (IPD) is a leading cause of severe infections in children and the elderly. For more than twenty years, this has been addressed by conjugate vaccination, whereby infections with pneumococci of serogroup 15 are only affected by recent introduction of PCV20, including serotype 15B and PCV21, covering serotype 15C. Statements on the effectiveness of a vaccination programme rely on continuous surveillance of IPD, in which genome-analysis-based serotyping is becoming increasingly prevalent.

It is known that serotype 15B can switch to serotype 15C by multiplication of a thymine-adenine (TA) repeat in a region of the *wciZ*-gene leading to a premature stop-codon, and the inability to produce a functional O-acetyltransferase. The number of these repeats and the associated risk of a premature stop codon is the basis for sequence-analysis methods currently used to differentiate between serotypes 15B and 15C.

AIM

The aim of the present study was to evaluate the reliability of currently used sequence-based methods for discriminating between serotypes 15B and 15C, with special focus on possible exceptions that may lead to misclassification.

METHODS

The study included 191 IPD isolates of serotypes 15B or 15C, isolated between January 2022 and May 2024.

Species verification was performed using bile solubility and optochin susceptibility testing, as well as sequence analysis of *16S rRNA* and *rpsB*-gene, while primary serotyping was done by Neufeld Quellung reaction.

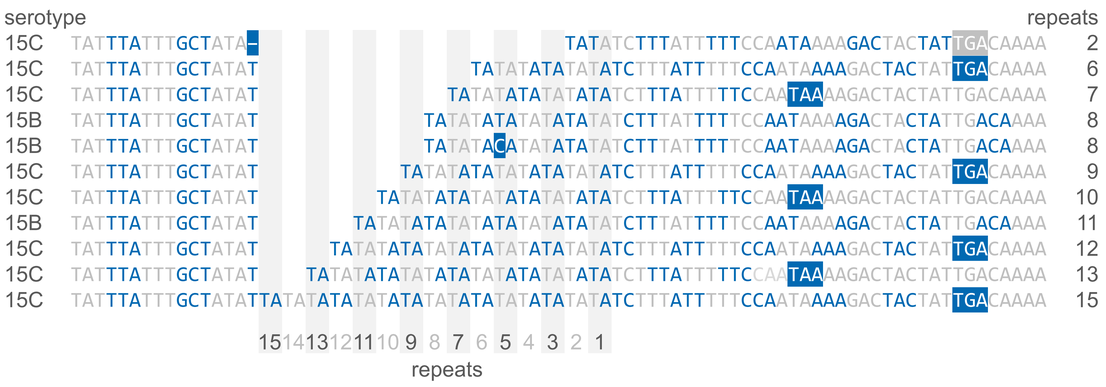
Sequence analysis of the TA-repeats in the *wciZ*-gene was done by amplification of an approximately 388 bp region starting at position 276 using primer wciZ-fwd and wciZ-rev. For isolates with inconclusive results, the entire *wciZ*-gene was sequenced using intergenic primers wciZ-fwd-ext and wciZ-rev-ext.

wciZ-fwd	5'-GTTTTGGAGAGAAGGTAATATGC-3'
wciZ-rev	5'-GTATTGGATAGATAATTGTTCCCA-3'
wciZ-fwd-ext	5'-GCAGTGTAGTTAACGAATAATGTG-3'
wciZ-rev-ext	5'-GGATTATCACCATAWCCYTTMCC-3'

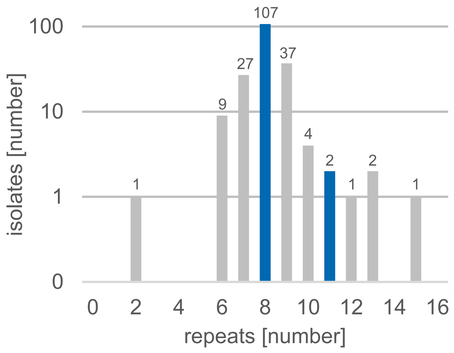
Isolates for which no agreement between serological capsule type determination and genetic prediction could be achieved were subjected to genome analysis. Illumina NovaSeq Sequencing 5M read pairs (2x150 bp) were generated.

RESULTS

Eleven genetic variants of the *wciZ*-gene repeat region were identified. Ten of these variants result from a different number of TA-repeats, whereby two variants with eight repeats each differ only in a silent point mutation (T→C), which, does not induce a frameshift. Isolates with eight or eleven TA-repeats were identified as serotype 15B by Neufeld Quellung, while all other genetic variants lead to premature stop codons (TGA or TAA), resulting in a serotype 15C capsule with low level of acetylation.



In the study cohort, the variant with eight TA-repeats encoding an intact *wciZ*-gene was by far the most common, while the number of isolates successively decreased with additional or fewer repeats. Interestingly, isolates with more than eight TA-repeats were slightly more common than isolates with less than eight repeats. Considering the serotype 15B variant carrying eight TA-repeats as origin of the genetic development, a change in the number of repeats by factor three should not introduce a frameshift, and lead to alternative serotype 15B variants, of which only the ones with eleven repeats, but not those with five or fourteen were found. An isolate with only two TA repeats, which should also not have a frameshift in the *wciZ*-gene, had a serotype 15C capsule due to the loss of a thymine base immediately upstream of the repeat region.



While in serotype 15C, the multiplication of TA-repeats leads to frameshift events with premature strand termination, in serotype 15B the genetic variation introduces two additional amino-acids (tyrosine and isoleucine), which, however, does not appear to have any previously observed influence on the enzymatic function of the encoded O-acetyltransferase.

RESULTS

Within the study cohort, four isolates were identified that showed noteworthy abnormalities in the genetic structure of the *wciZ*-gene with potential influence on capsule and serotype.

isolate	serotype	repeats	event	position	consequence
104761	15B	8	T→C	419bp	I→T
115923	15C	8	13bp missing	525bp	stop
108994	15C	8	T→C	732bp	silent
106910	15B	7	repeat missing	148bp	stop

Three isolates had eight TA-repeats in the repeat region of the *wciZ*-gene. One of the isolates (104761) could be assigned to the expected serotype 15B by Neufeld Quellung, although this isolate had a point mutation leading to an exchange of isoleucine for threonine.

In another isolate (115923), which was identified as serotype 15C despite having eight repeats, a gap of 13 bases could be localized downstream of the repeat region, leading to a premature stop codon by frameshift.

The third isolate with eight TA-repeats, identified as serotype 15C (108994), carries a complete *wciZ*-gene, the only abnormality being a silent point mutation. A comparable situation was found in a serotype 15B isolate with seven TA-repeats and a premature stop codon (106910). These two isolates are currently undergoing genome analysis to determine a possible explanation for the discrepancy between the serological capsule type and the state of the *wciZ*-gene.

CONCLUSIONS

- A large proportion of the identified genetic variants of the *wciZ*-gene is due to changes in the number of TA-repeats.
- The variant with eight-TA repeats is the most commonly isolated and is potentially the origin of the genetic variant formation.
- Exceptions show that serotype prediction based on the number of TA-repeats can lead to misinterpretation.
- Genetic events outside the repeat region are rare but can change the serotype.

OUTLOOK

- All isolates which serotypes do not match the corresponding structure of the *wciZ*-gene will be subjected to detailed genome analysis.
- Identification of additional genetic variants of the *wciZ*-gene through continued expansion of the study cohort with additional serotype 15B and 15C isolates is ongoing.

Disclosure: Pneumococcal surveillance in Germany was, in part, sponsored by Pfizer Pharma GmbH and Merck Sharp & Dohme Corp.