Molecular Genetics and Genomics
https://doi.org/10.1007/s00438-018-1507-5

ORIGINAL ARTICLE

Intensive targeting of regulatory competence genes by transposable elements in streptococci

Maud Fléchard[1,2] · Céline Lucchetti-Miganeh[3] · Bernard Hallet[2] · Pascal Hols[2] · Philippe Gilot[1]

Received: 9 July 2018 / Accepted: 21 October 2018

Abstract

Competence for natural transformation is a widespread developmental process of streptococci. By allowing the uptake and recombination of exogenous naked DNA into the genome, natural transformation, as transposable elements, plays a key role in the plasticity of bacterial genomes. We previously analysed the insertion sites of IS*1548*, an insertion sequence present in *Streptococcus agalactiae* and *S. pyogenes*, and showed that some targeted loci are involved in competence induction. In this work, we investigated on a large scale if loci coding for early competence factors (ComX and the two pheromone-dependent signalling systems ComCDE and ComRS) of streptococci are especially targeted by transposable elements. The transposable elements inserted in regions surrounding these genes and housekeeping genes used for Multilocus Sequence Typing (MLST) were systematically searched for. We found numerous insertion events in the close vicinity of early competence genes, but only very few into the MLST loci. The incidence of transposable elements, mainly insertion sequences, is particularly high in the intergenic regions surrounding *comX* alleles in numerous species belonging to most streptococcal groups. The identification of scarce disruptive insertions inside early competence genes indicates that the maintenance of competence is essential for streptococci. The specific association of transposable elements with intergenic regions bordering the main regulatory genes of competence may impact on the induction of transformability and so, on the genome plasticity and adaptive evolution of streptococci. This widespread phenomenon brings new perspectives on our understanding of competence regulation and its role in the bacterial life cycle.

Keywords Natural transformation · Mobile genetic element · Insertion sequence · *Streptococcus*

Introduction

The bacterial genus *Streptococcus* is composed of a large diversity of species with different lifestyles including, among others, the human or animal pathogens *S. pyogenes, S. pneumoniae* and *S. suis*, the commensal *S. salivarius*, or the dairy bacterium *S. thermophilus* (Richards et al. 2014; Póntigo et al. 2015). Furthermore, several streptococcal species, such as *S. suis* or *S. pneumoniae*, display a particularly high gene turnover, and a large proportion of their pan-genome is thought to have been subject to horizontal gene transfer (Richards et al. 2014). The elucidation of the mechanisms underlying this genetic diversity and of their dynamics is thus of critical importance for the understanding of the evolution of streptococci.

Mobile genetic elements (MGEs) play a critical role in the evolution and in the plasticity of bacterial genomes. They provide an important way of adaptation to environmental constraints, allowing the emergence of subpopulations of

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00438-018-1507-5) contains supplementary material, which is available to authorized users.

✉ Maud Fléchard
maud.flechard@yahoo.com

✉ Philippe Gilot
philippe.gilot@univ-tours.fr

[1] Bactéries et Risque Materno-Foetal, UMR1282 Infectiologie et Santé Publique, Université de Tours, INRA, 37032 Tours, France

[2] Biochimie, Biophysique et Génétique des Microorganismes, Louvain Institute of Biomolecular Science and Technology, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

[3] Genostar, 60 rue Lavoisier, 38330 Montbonnot, France

Published online: 07 November 2018

bacteria displaying new features that potentially confer an adaptive advantage under some specific conditions (Wiedenbeck and Cohan 2011; Casacuberta and González 2013; Fléchard and Gilot 2014). They are involved in these phenomena by (1) providing "passenger" genes in addition to those necessary for their insertion/excision and/or replication; (2) negatively modulating gene expression, e.g., by the disruption of a promoter or an open reading frame (ORF), (3) positively modulating gene expression, e.g., by the provision of a promoter for a neighbouring ORF; (4) allowing important chromosomal rearrangements by homologous recombination between similar elements inserted at different locations of the genome (Casacuberta and González 2013). MGEs are very diverse and can be divided into two categories: the intercellular MGEs, such as phages, plasmids or integrative and conjugative elements (ICEs), that can spread from cell to cell, and the intracellular MGEs or transposable elements (TEs) that are not themselves transmissible. TEs include insertion sequences (ISs) and composite or non-composite transposons, as well as group II introns and integrons. (Siguier et al. 2014; Brüssow et al. 2004; Wozniak and Waldor 2010; Casacuberta and González 2013; Toro et al. 2007; Labbate et al. 2009).

The target-site specificity varies a lot among characterized TEs, ranging from insertion into low-specific sequences, e.g., IS*S1* of *Lactococcus lactis*, to insertion into very specific sequences, such as Tn*7* at the *attTn7* site of the *Escherichia coli glmS* terminator (Craig 1997; Maguin et al. 1996). Between these two extremes, other mechanisms of targeting have been described, including the recognition of GC- or AT- rich regions or of consensus sequences, as well as the involvement of binding sites of transcription factors or of structural features of DNA (Sengstag et al. 1986; Zerbib et al. 1985; Goryshin et al. 1998; Zhang and Saier 2016; Hallet et al. 1994). Diverse examples of genes preferentially targeted by ISs have also been described. Some of them, such as *comP* of *Bacillus subtilis*, are targeted by the same transposable element (IS*4Bsu1* in this specific example) but at different positions of the locus (Nagai et al. 2000). Others, such as the *cps* capsular locus of *S. pneumoniae* or the cryptic *bgl* operon of *E. coli*, are interrupted by various TEs (Moscoso and García 2009; Reynolds et al. 1981). It is not always known if the insertions of TEs into these hotspots are induced by specific environmental signals or if TEs are maintained at these locations according to some selection pressure encountered by the bacteria. Nevertheless, it was proposed that some of these events are regulated by specific environmental conditions, and so physiologically controlled (Hall 1998; Takahashi et al. 2007; Zhang et al. 2017). The possible preferential targeting of certain genes or functional classes of genes by TEs would imply the existence of a complex regulatory mechanism called adaptive mutation (Hall 1998; Wang and Higgins 1994; Zhang and Saier Jr 2011).

We previously analysed the insertion targets of IS*1548* (Fléchard and Gilot 2014; Fléchard et al. 2013a, b). IS*1548* is a 1317-bp element belonging to the IS*As1* family, which was first described in *S. agalactiae* (Granlund et al. 1998). It carries terminal inverted repeats of 19 bp. A putative promoter located just before the right-inverted repeat is outwardly directed and is thought to induce an increased transcription of the downstream genes, as in the case of the *lmb* gene of *S. agalactiae* (Al Safadi et al. 2010). We showed that several targets of IS*1548* are genes or intergenic regions (IGRs) close to genes involved in or linked to metal homeostasis. Among them are the *adcRCB* operon, which encodes a transcriptional repressor and components of a zinc ABC transporter, as well as the *adcA* and *lmb* genes which encode two alternative zinc-binding proteins of the Adc transporter (Fléchard et al. 2013a, b). In bacteria, zinc is involved in a wide range of physiological processes, and its transport was shown to be necessary for the establishment of competence for natural transformation in *S. gordonii, S. pneumoniae* and *B. subtilis* (Loo et al. 2003; Dintilhac et al. 1997; Ogura 2011). Furthermore, in some strains of *S. pyogenes*, IS*1548* targets the downstream IGR of the *comX1* gene, which encodes the master regulator of competence in streptococci.

The above data led us wonder whether there is a more general link between TEs and genes involved in competence for natural transformation in the *Streptococcus* genus. Natural transformation is a widespread feature of streptococci. By allowing the uptake and recombination into the genome of exogenous naked DNA, natural transformation, as TEs, is involved in the evolution and in the plasticity of bacterial genomes. All streptococci possess the effector genes necessary for the establishment of competence, as well as at least one homolog of *comX* (Martin et al. 2006). Furthermore, at least one species belonging to each main phylogenetic *streptococcus* group was demonstrated to naturally induce transformability: i.e., the historical model *S. pneumoniae* (mitis group), *S. gordonii* (anginosus group), *S. mutans* (mutans group), *S. thermophilus* (salivarius group), and recently *S. macedonicus* and *S. infantarius* (bovis group), *S. suis* (*S. suis* species), and *S. pyogenes* (pyogenic group) (Griffith 1928; Avery et al. 1944; Pakula et al. 1958; Perry and Kuramitsu 1981; Gardan et al. 2009; Morrison et al. 2013; Zaccaria et al. 2014; Marks et al. 2014). In streptococci, competence develops during temporally distinct early and late phases. Activation of the early phase results in a major increase in the expression of the alternative sigma X factor (σ^X) encoded by the *comX* gene. During the late phase, σ^X associates with the core RNA polymerase to activate the transcription of late competence genes, including those encoding the transformasome. Two major proximal pheromone-dependent signal transduction pathways have been described to activate *comX* transcription: the ComCDE and ComRS pathways (Fig. 1) (Fontaine et al. 2015). The

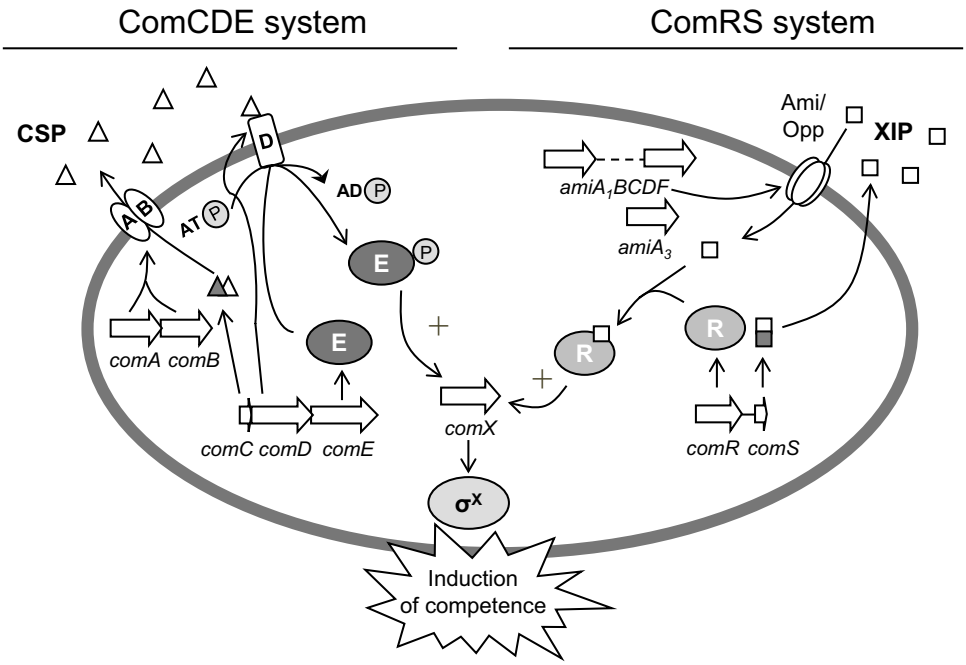


Fig. 1 Early steps of the signalling cascade leading to the induction of competence in streptococci. The ComCDE system is used by strains of the mitis and anginosus groups, whereas the ComRS system is used by strains of the salivarius, bovis, pyogenic and mutans groups, and of the *S. suis* species. In the ComCDE system, a peptide pre-pheromone (double triangle) encoded by *comC* is matured (white triangle, CSP) and exported by the ComAB transporter (AB). CSP binds to the ComD histidine kinase (D) which activates the ComE response regulator (E). Activated ComE induces transcription of its regulon, including the *comX* alleles that encode σ^X. In the ComRS system, a peptide pre-pheromone encoded by the *comS* gene (double square) is exported and cleaved by an unknown protease. The truncated form of ComS (white square, XIP) is reimported into the cytoplasm by the Ami/Opp oligopeptide ABC transporter, where it associates with ComR (R). Finally, the XIP-ComR complex induces transcription of its regulon, including *comX*. CSP, Competence-Stimulating Peptide; XIP, *sigX*-Inducing Peptide; P, phosphate group

ComCDE system is present in species belonging to the mitis and anginosus groups, whereas the ComRS system is found in species of the salivarius, pyogenic and bovis groups, as well as in *S. suis*. Strains belonging to the mutans group possess the genes coding for both signalling systems, but only the ComRS complex is responsible for ComX induction (Håvarstein 2010; Reck et al. 2015).

In the ComCDE system, a peptide pre-pheromone, encoded by the *comC* gene, is produced, then cleaved and exported outside of the bacterial cell by the ATP-dependent ComAB transporter. The processed peptide, called CSP for Competence-Stimulating Peptide, is the active form. CSP then activates the ComDE two-component system that is composed of the ComD transmembrane histidine kinase and of the ComE response regulator. ComD also acts as a CSP captor responsible for the retention of the pheromone, limiting its diffusion in the extracellular medium (Prudhomme et al. 2016). In response to CSP binding, ComD autophosphorylates and then phosphorylates ComE. The phosphorylated form of ComE induces transcription of its regulon, including the *comCDE* and *comAB* operons and the *comX* alleles. In the ComRS system, the product of the *comS* gene is a peptide pre-pheromone that is exported by an unknown mechanism and matured. The processed form of ComS, called XIP for *sigX*/*comX*-Inducing Peptide, is the active pheromone. XIP is then reimported into the cytoplasm by the Ami/Opp oligopeptide ABC transporter, and associates with ComR, a transcriptional regulator belonging to the RNPP family (from Rap, Npr, PlcR, and PrgX). Finally, the XIP-ComR complex induces transcription of its regulon, including *comS* and *comX*. The *comR* and *ami*/*opp* transporter genes are not known to be regulated by this way (Fig. 1) (Fontaine et al. 2015).

In this work, we investigated the association of TEs with genes involved in the early phase of competence in 126 completely sequenced *Streptococcus* genomes from 28 different species (Supplemental Table S1). We focused our analysis on the genic environment of the loci encoding the master regulator ComX and the two types of competence-inducing signalling pathways (the *comCDE* and *comAB* operons, the *comRS* genes and the *ami*/*opp* region). This investigation was compared to a similar analysis performed on the genic environment of several unrelated genes [Multilocus Sequence Typing (MLST) genes] distributed along the genome. The potential consequences of the insertion of identified TEs on genome plasticity of streptococci and on

their ability to induce competence for natural transformation are discussed.

Materials and methods

Streptococcus strains analysed

The genomes of 126 *Streptococcus* strains belonging to 28 different species were analysed in this study (Supplemental Table S1). These are all the genomes that were available in the SyntTax database on the 16th of August 2015, with the exception of *S. pneumoniae* SPN033038 and SPN032672 and *S. salivarius* K12 and M18, which were not properly annotated. Strains were classified into the seven main phylogenetic groups (i.e., the anginosus group, the bovis group, the mitis group, the mutans group, the salivarius group, the pyogenic group, and the *S. suis* species), according to Richards et al. (2014) and Póntigo et al. (2015).

Retrieval of the genic environment of the analysed genes with the synteny web server SyntTax

For the analysis of the MLST loci, the translated sequences of *aroE* (GenBank Acc. n°AAK75474), *gdh* (GenBank Acc. n°AAK75348), *gki* (GenBank Acc. n°ABJ55324), *recP-2* (GenBank Acc. n°AAK76095), *spi* (GenBank Acc. n°AAK74565), *xpt* (GenBank Acc. n°AAK75919), and *ddl* (GenBank Acc. n°AAK75750) of *S. pneumoniae* TIGR4 (a strain known to be competent for natural transformation) were submitted to the SyntTax server (http://archaea.u-psud.fr/synttax/) (Oberto 2013). For the analysis of the *comCDE, comAB, comRS, amiA*$_3$, *amiA*$_1$*BCDF* and *comX* loci, the following sequences were submitted to the SyntTax server : (1) ComD of *S. pneumoniae* TIGR4 (GenBank Acc. n° AAK76283); (2) ComA of *S. pneumoniae* TIGR4 (GenBank Acc. n° AAK74231); (3) ComR of: *S. agalactiae* A909 (GenBank Acc. N° ABA45988), *S. agalactiae* NEM316 (GenBank Acc. N° WP_000912098), *S. agalactiae* 138spar (GenBank Acc. N° AHX74410), *S. gallolyticus* UCN34 (GenBank Acc. N° WP_009853198), *S. mutans* UA159 (GenBank Acc. n° NP_720543), *S. pyogenes* Manfredo (GenBank Acc. N° CAM29376), *S. suis* D9 (GenBank Acc. N° AER16335), *S. suis* TL13 (GenBank Acc. N° AGL47035), and *S. thermophilus* LMD-9 (GenBank Acc. N° ABJ65625); (4) AmiA3 of *S. thermophilus* LMD-9 (GenBank Acc. n° ABJ66575); (5) AmiF of *S. thermophilus* LMD-9 (GenBank Acc. n° ABJ66569); (6) ComX of *S. pneumoniae* TIGR4 (GenBank Acc. n° AAK74207). Several ComR sequences were tested, because the ComR proteins from different species are not highly similar (29–59% identity for the nine tested strains). Moreover, they display a relatively high level of identity with other members of the RNPP family of transcriptional regulators. These nine ComR sequences are representative of the diversity of ComR proteins in streptococci (Fontaine et al. 2015). The proteins encoded by the genes localised at each end of the large locus of the Ami/Opp transporter of *S. thermophilus* LMD-9 were tested (AmiF and AmiA3). AmiF is an ATP-binding protein encoded by the last gene of the *amiA*$_1$*BCDF* operon. AmiA3, which is encoded outside of the *amiA*$_1$*BCDF* operon, is one of the binding proteins of the Ami/Opp transporter and was proved to have the major role in the triggering of competence in *S. thermophilus* LMD-9 (Gardan et al. 2009).

A normalized blast score of 10% (corresponding to the cut-off bit score of the alignments compared to a reference score of 100% for the BlastP alignment of the query protein matched against itself) was selected in the SyntTax program, and only the best match of each strain was analysed. An exception was made for *comX*, because one-to-three alleles of this gene may be present in *Streptococcus* strains. In that case, all the matches were retrieved. Nevertheless, the genic environment of *comX2* of *S. suis* cannot be obtained by this way, because this allele is truncated; the ComX2 protein of this species is approximately half the size of the ComX protein of TIGR4 and shares only 37% identity with it. For this reason, the *comX2* environment in *S. suis* was obtained by blasting the *comX1* sequence of *S. suis* BM407 against the genome of all *S. suis* strains analysed in this work.

Identification of mobile genetic elements

The matches retrieved by SyntTax were further analysed to identify TEs. To this end, we analysed the proteins encoded by the ORFs located either inside or just upstream or downstream of the genes of interest with the BlastN and BlastP analysis tools of the ISfinder database (https://www-is.biotoul.fr/) (Siguier et al. 2006) or by searching for conserved domains in the Conserved Domains Database (CDD) of the NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Marchler-Bauer et al. 2015). If a protein sequence is more than 98% similar and/or the corresponding DNA sequence is more than 95% similar to any entry in ISfinder, the sequence is an isoform of this IS and the name given in the database was used. In other cases, an attribution name for the IS elements was asked for to ISfinder. Particular cases consisting of ORFs separated from the competence genes by a tRNA, a short ORF (smaller than 250 bp) encoding a remnant protein or a short peptide were also considered. Pseudogenes located close to early competence genes were analysed with the ISfinder BlastN tool, and also with the BlastP tool for truncated ORFs. ISs with truncated or mutated sequences (pseudogenes) were named like the element from which they derived. If their origin could not be determined, they were named ISxxx-like, with xxx corresponding to the name of the more similar IS in the ISfinder

database (https://www-is.biotoul.fr/) (Siguier et al. 2006). The names IS*Ssu10*, IS*Ssu11*, IS*Ssu12*, IS*Ssu13*, IS*Sth8*, IS*Stan1*, IS*Spn14*, IS*Strs1*, IS*Slu1* and IS*Strsp1* were given to ISs reported to ISfinder during this work. The CDD NCBI database was used to identify ORFs belonging to group II introns or phage integrases (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Marchler-Bauer et al. 2015).

Phylogenetic analysis

For each species for which at least four strains were analysed, a phylogenetic tree based on the concatenation of four MLST marker genes was constructed. The marker genes used were *aroE, gki, spi* and *ddl*, except for *S. agalactiae* and *S. suis* for which *gki* and *ddl* were replaced by *glnA* and *thrA*, respectively. This replacement is due to the existence of large deletions in the *gki* or *ddl* genes of some *S. agalactiae* or *S. suis* strains, respectively. For each marker, the sequences of the ORFs of all the strains were aligned with the MEGA 7.0 software (Kumar et al. 2016) using the Muscle algorithm. Then, the alignments of each marker were assembled together with the Fabox Fasta alignment joiner tool (http://users-birc.au.dk/biopv/php/fabox/alignment_joiner.php). Lastly, a phylogenetic analysis based on the concatenated alignment was performed in MEGA 7.0 using the maximum likelihood method based on the Tamura-Nei model (Tamura and Nei 1993), and a bootstrap consensus tree inferred from 500 replicates (Felsenstein 1985) was constructed to represent the evolutionary history of the species.

Statistics

Chi square tests using fourfold tables and the correction for continuity devised by Yates for small values were used to compare the association of TEs with each early competence loci and with a set of seven MLST genes (Swinscow 1978). *P* indicates the level of significance of the association of TEs with early competence genes.

Results

Transposable elements in the MLST loci

To estimate the importance of the association between TEs and genes involved in the early events of competence for natural transformation, we first analysed the loci of seven genes used in MLST of *S. pneumoniae* and unrelated with early competence events. One of these genes, *recP-2*, was described to be involved in competence for natural transformation in *S. pneumoniae*, but only during the late phase (Rhee and Morrison 1988). Homologs of the seven MLST genes were retrieved in all the *streptococcus* phyla; except for *gdh*, which is absent in strains of the bovis, mutans and pyogenic groups; and *xpt*, which is absent in strains of the mutans group and of the *S. suis* species (the best match retrieved in this latter species was the *purR* gene). The frequency of insertion of TEs in each of these MLST genes loci is low. Indeed, for any of these analysed genes, the proportion of the strains possessing the MLST allele associated with a TE never exceeds 3.2% (from 0% for *aroE* to 3.2% for *ddl*, Supplemental Table S2).

These data, probably reflecting the random level of insertion of TEs into the chromosome of streptococci, indicate that MLST genes are not particularly targeted by TEs.

Transposable element in the *comCDE* locus in the mitis and anginosus groups

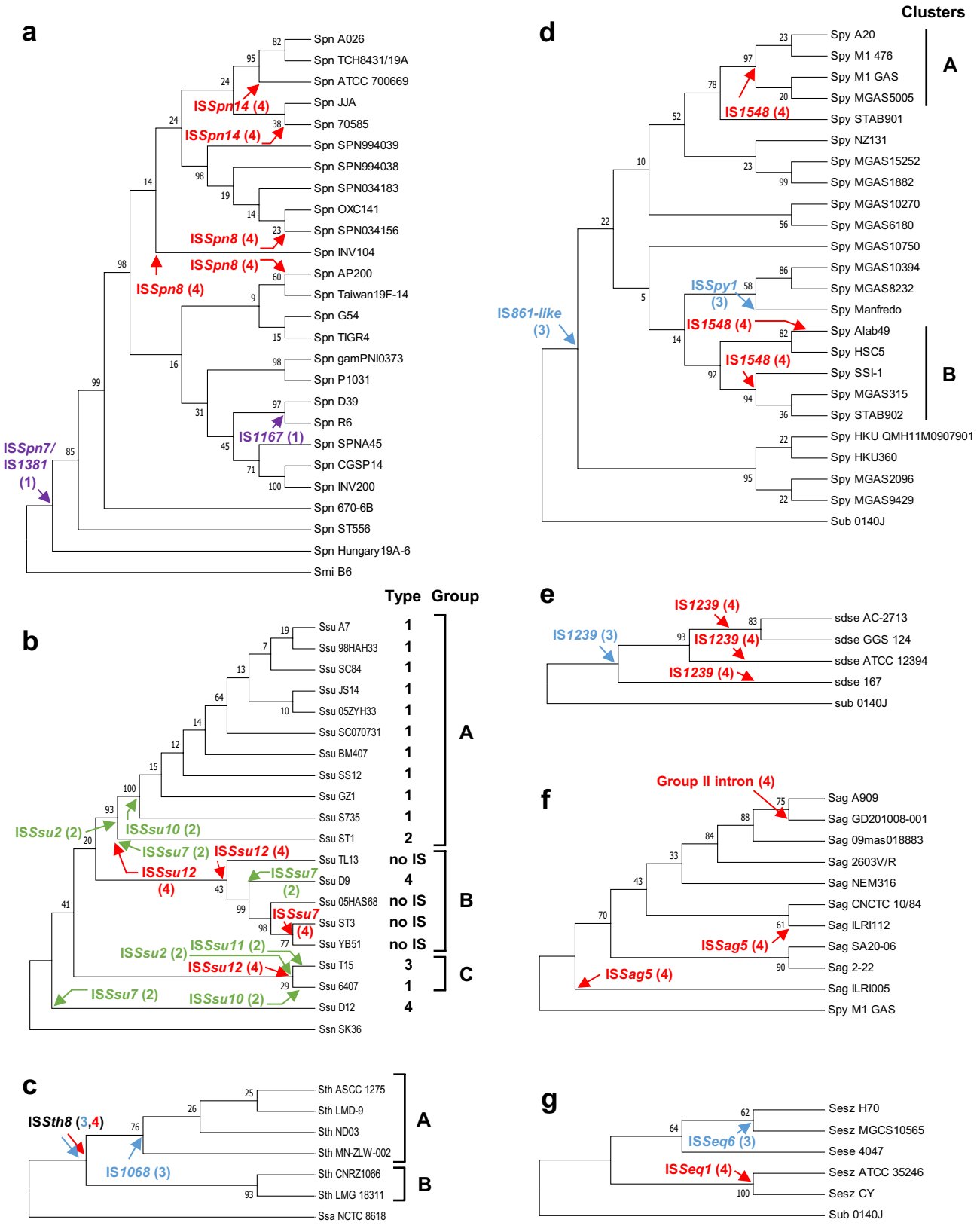
Location of the transposable element

Only one IS, annotated IS*1239*, was found in the *comC* gene of *S. pneumoniae* AP200. This element is identical to IS*Spn8* (IS*30* family) of *S. pneumoniae*. Camilli et al. previously reported that the presence of the same IS at that position makes strain AP200 unable to develop natural competence (Camilli et al. 2011). The *comCDE* operon is not particularly targeted by TEs ($0.5 < P < 0.9$), as also 1 strain out of 71 was targeted by an IS at the *gdh* MLST gene locus (Supplemental Table S2).

Transposable elements in the *comAB* locus in the mitis and anginosus groups

Location of the transposable elements

Two kinds of ISs were identified exclusively upstream of this operon in *S. pneumoniae* strains. In all of the 25 analysed strains of these species, a small ORF, which is a remnant of an IS, is present upstream of *comAB* (Supplemental Fig. S1). This IS is an isoform of IS*Spn7* and of IS*1381* (IS*5*/IS*L2* family). The IS*Spn7*/*IS1381* remnant is separated from *comA* by a small ORF coding for a bacteriocin-like peptide. Bacteriocin production is an integral part of the process of competence. These antimicrobial peptides are produced concomitantly with the induction of competence and are thought to provide competent cells with exogenous DNA by promoting lysis of the neighbour cells (Kjos et al. 2016). Furthermore, in three strains of *S. pneumoniae* (JJA, Hungary 19A-6 and 70585) and in *S. pseudopneumoniae* IS7493, a genomic islet of six to seven ORFs, which contains a *comR*-like gene, is also inserted between the IS*Spn7*/IS*1381* remnant and *comAB* (Supplemental Fig. S1). Another IS, IS*1167* (IS*L3* family), is also inserted between the IS*Spn7*/IS*1381* remnant and *comAB* of strain R6 (Supplemental Fig. S1). A fragmented pseudogene derived from this element is


a
Spn A026
Spn TCH8431/19A
Spn ATCC 700669
Spn JJA
Spn 70585
Spn SPN994039
Spn SPN994038
Spn SPN034183
Spn OXC141
Spn SPN034156
Spn INV104
Spn AP200
Spn Taiwan19F-14
Spn G54
Spn TIGR4
Spn gamPNI0373
Spn P1031
Spn D39
Spn R6
Spn SPNA45
Spn CGSP14
Spn INV200
Spn 670-6B
Spn ST556
Spn Hungary19A-6
Smi B6
ISSpn14 (4)
ISSpn14 (4)
ISSpn8 (4)
ISSpn8 (4)
ISSpn8 (4)
IS1167 (1)
ISSpn7/ IS1381 (1)
b
Type
Group
Ssu A7
Ssu 98HAH33
Ssu SC84
Ssu JS14
Ssu 05ZYH33
Ssu SC070731
Ssu BM407
Ssu SS12
Ssu GZ1
Ssu S735
Ssu ST1
Ssu TL13
Ssu D9
Ssu 05HAS68
Ssu ST3
Ssu YB51
Ssu T15
Ssu 6407
Ssu D12
Ssn SK36
ISSsu2 (2)
ISSsu10 (2)
ISSsu7 (2)
ISSsu12 (4)
ISSsu12 (4)
ISSsu7 (2)
ISSsu7 (4)
ISSsu2 (2)
ISSsu11 (2)
ISSsu12 (4)
ISSsu7 (2)
ISSsu10 (2)
c
Sth ASCC 1275
Sth LMD-9
Sth ND03
Sth MN-ZLW-002
Sth CNRZ1066
Sth LMG 18311
Ssa NCTC 8618
ISSth8 (3,4)
IS1068 (3)
d
Clusters
Spy A20
Spy M1 476
Spy M1 GAS
Spy MGAS5005
Spy STAB901
Spy NZ131
Spy MGAS15252
Spy MGAS1882
Spy MGAS10270
Spy MGAS6180
Spy MGAS10750
Spy MGAS10394
Spy MGAS8232
Spy Manfredo
Spy Alab49
Spy HSC5
Spy SSI-1
Spy MGAS315
Spy STAB902
Spy HKU QMH11M0907901
Spy HKU360
Spy MGAS2096
Spy MGAS9429
Sub 0140J
IS1548 (4)
ISSpy1 (3)
IS861-like (3)
IS1548 (4)
IS1548 (4)
e
sdse AC-2713
sdse GGS 124
sdse ATCC 12394
sdse 167
sub 0140J
IS1239 (4)
IS1239 (3)
IS1239 (4)
IS1239 (4)
f
Sag A909
Sag GD201008-001
Sag 09mas018883
Sag 2603V/R
Sag NEM316
Sag CNCTC 10/84
Sag ILRI112
Sag SA20-06
Sag 2-22
Sag ILRI005
Spy M1 GAS
Group II intron (4)
ISSag5 (4)
ISSag5 (4)
g
Sesz H70
Sesz MGCS10565
Sese 4047
Sesz ATCC 35246
Sesz CY
Sub 0140J
ISSeq6 (3)
ISSeq1 (4)

◂**Fig. 2** Phylogenetic relationship among the analysed strains. The phylogenetic relationship among the analysed strains of *S. pneumoniae* (Spn, **a**), *S. suis* (Ssu, **b**), *S. thermophilus* (Sth, **c**), *S. pyogenes* (spy, **d**), *S. dysgalactiae* subsp. *equisimilis* (sdse, **e**), *S. agalactiae* (Sag, **f**), and *S. equi* [subsp. *zooepidemicus* (sesz) and subsp. *equi* (sese), **g**] is represented by a rectangular cladogram. The evolutionary history was inferred using the maximum likelihood method based on the Tamura-Nei model (Tamura and Nei 1993). The trees shown, inferred from 500 bootstrap replicates, have the highest log likelihood (Felsenstein 1985). Evolutionary analyses and multiple alignments using the MUSCLE algorithm were conducted in MEGA7 (Kumar et al. 2016). *S. mitis* (Smi) B6, *S. sanguinis* (Ssn) SK36, *S. salivarius* (Ssa) NCTC 8618, and *S. pyogenes* (Spy) M1 GAS were used as outgroups to root trees a, b, c and f, respectively. *S. uberis* (Sub) 0140J was used as an outgroup to root trees d, e and g. Insertions of TEs in the early competence loci of the strains are represented by arrows, and a specific number corresponding to the targeted locus is indicated between parenthesis after the name of each IS [(1), *comAB*; (2), *comRS*; (3), *ami* and (4), *comX*]

similarly integrated in strain D39. IS*1167* was also reported to be present at the same position in *S. pneumoniae* CP1200 (Zhou et al. 1995).

The insertion of ISs close to the *comAB* operon was only identified in *S. pneumoniae*. We did not detect any ISs close to the *comAB* locus of the analysed strains belonging to the anginosus group and in other species of the mitis group apart from *S. pneumoniae*. In this latter species, 100% of the analysed strains contain a TE at this location. The *comAB* locus of *S. pneumoniae* is thus highly associated with TEs ($P > 0.999$).

Phylogenetic analysis of the strains targeted by the transposable elements

The IS*Spn7*/IS*1381* remnant was found upstream of *comAB* in all the analysed strains of *S. pneumoniae*. This suggests that the native IS was already inserted at that location in the common ancestor of the species. The insertion of IS*1167* arrived more recently, after the divergence of the closely related strains R6 and D39 from the rest of the population (Fig. 2a).

In conclusion, two independent insertion events seem to have occurred in the *comAB* locus of *S. pneumoniae*.

Transposable elements in the *comRS* locus in *S. suis* and in the salivarius, pyogenic, mutans and bovis groups

Location of the transposable elements

Four different types of IS(s) insertion were found exclusively in the IGR downstream of the *comRS* genes of *S. suis* (Supplemental Fig. S2). In types 1 to 3, a tandem of ISs is integrated downstream of *comRS*. The second IS of this tandem is a mutated IS*Ssu2* element (IS*4* family), whereas its first element varies between types. In type 1, found in 11 strains (A7, BM407, GZ1, JS14, S735, SC84, SC070731, SS12, 05ZYH33, 6407, 98HAH33), the first IS of the tandem, which was not previously assigned in the ISfinder database, received the name IS*Ssu10* (IS*110* family). In type 2, identified in strain ST1, the first element of the tandem is a degenerated IS very similar to IS*Ssu7* (IS*110* family) of *S. suis* (93% of identical nucleotides or amino acids). In type 3, identified in strain T15, the first IS of the tandem was also not previously assigned in the ISfinder database. It received the name IS*Ssu11* (IS*3* family). The type 4, found in strains D9 and D12, consists of the insertion of only one isoform of IS*Ssu7*.

The insertion of ISs close to the *comRS* locus was only identified in *S. suis*. We did not detect any ISs close to the *comRS* locus of the analysed strains belonging to the salivarius, pyogenic, mutans and bovis groups. In *S. suis*, 78.9% of the analysed strains (15 out of the 19 strains) contain at least one IS at this location, whereas only one strain was targeted by an IS at the *spi* MLST gene locus (Supplemental Table S2). The *comRS* locus of *S. suis* is thus strongly associated with TEs ($P > 0.999$).

Phylogenetic analysis of the strains targeted by the transposable elements

In the *S. suis* species, all but one strains displaying type 1 and 2 insertion patterns are clustered (groupA, Fig. 2b). This indicates that IS*Ssu2* most likely inserted in the *comRS* locus of the common ancestor of these strains. Strain ST1, the only representative of type 2, diverged from the ancestor of the rest of group A strains. IS*Ssu7* inserted then into the *comRS* locus in strain ST1, whereas the ancestor of all the other strains of group A integrated IS*Ssu10* at this locus. One type 1 strain (6407) does not belong to group A. Strain 6407 clusters with strain T15 (group C, Fig. 2b), the only representative of type 3. IS*Ssu2* was probably also initially inserted in the *comRS* locus of the common ancestor of these two strains. Strain 6407 should then have integrated IS*Ssu10* at that location, whereas strain T15 acquired IS*Ssu11*. Finally, the two strains of type 4 are not phylogenetically linked, indicating distinct insertion events of IS*Ssu7* (Fig. 2b).

The above data indicate that several distinct insertion events occurred in the *comRS* locus of the analysed population of *S. suis* strains.

Transposable elements in the *ami*/*opp* region in *S. suis* and in the salivarius, pyogenic, mutans and bovis groups

Location of the transposable elements

All of the six analysed strains of *S. thermophilus* contain an IS between the $amiA_3$ gene and the $amiA_1BCDF$ operon (Supplemental Fig. S3). This IS which was not previously assigned in the ISfinder database received the name IS*Sth8* (IS*L3* family). In four of these strains (LMD-9, ASCC 1275, MN-ZLW-002 and ND03), an IS being an isoform of the very similar IS*1068*/IS*1069*/IS*1076* (IS*3* family) of *Lactococcus lactis* is also integrated into the upstream IGR of $amiA_3$ (Supplemental Fig. S3).

Furthermore, several strains of the pyogenic group contain an IS integrated just downstream of the *ami*/*oppF* gene (Supplemental Fig. S3, *yqeG* locus 1 in Supplemental Table S3). IS*Seq6* (IS*30* family) is inserted at that position in one of the four analysed strains of *S. equi* subsp. *zooepidemicus* (H70) and an isoform of IS*1239* (IS*30* family) of *S. pyogenes* is present in this IGR in all of the four analysed strains of *S. dysgalactiae* subsp. *equisimilis*. Finally, in all the 23 analysed *S. pyogenes* strains, a small IS*861*-like remnant (IS*3* family) is present downstream of *oppF*. The length of this remnant varies between strains, indicating that several deletion events occurred since the initial insertion event(s). Moreover, in strain Manfredo, a defective isoform of IS*Spy1* (IS*3* family) is integrated downstream of this remnant (Supplemental Fig.S3).

The insertion of ISs close to the *ami*/*opp* region was identified in 40.2% of the analysed strains which use the ComRS system to activate *comX* (33 out of the 82 analysed strains), whereas only eight of these strains were targeted by an IS or a transposon at either the *ddl, xpt, spi* or *recP-2* MLST gene loci (Supplemental Table S2). The *ami*/*opp* region is thus very strongly linked to TEs in this group of strains ($P > 0.999$).

Phylogenetic analysis of the strains targeted by the transposable elements

In *S. thermophilus*, the IS*Sth8* element located between *amiA3* and *amiA1* was found in all the analysed strains, suggesting its inheritance from the common ancestor of the species. In contrast, a copy of IS*1068* was identified upstream of *amiA3* only in the four members of one of the two major clusters of strains (group A, Fig. 2c), indicating a subsequent insertion of this IS at that locus in the common ancestor of group A. In *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis*, the presence in all the analysed strains of IS*861*-like or IS*1239*, respectively, downstream of *oppF* suggests again that these ISs were already inserted at that location in the common ancestors of these species (Fig. 2d, e).

In conclusion, few independent insertion events seem to have occurred in the *ami*/*opp* region of *S. thermophilus* and of the pyogenic streptococci.

Transposable elements in the *comX* loci

Location of the transposable elements

Streptococci possess one-to-three alleles of *comX*: specifically, one allele is found in species belonging to the salivarius group, two alleles in *S. pneumoniae* and *S. pyogenes*, and three alleles in some strains of *S. suis* (Table 1). Fourty-seven of the 126 analysed strains (37.3%) contain just close to at least one of their *comX* alleles at least one ORF coding for a transposase, for a reverse transcriptase or for a maturase belonging to group II introns, or derived pseudogenes, whereas only ten strains were targeted by a TE at either the *gdh, gki, recP-2, spi, xpt* or *ddl* MLST gene loci (Table 1, Supplemental Table S2). The *comX* alleles are thus highly significantly linked with TEs ($P > 0.999$). These elements were found either upstream (for 31 elements identified in 26 strains) or downstream (for 26 elements identified in 23 strains) of *comX* genes. In two strains (*S. suis* YB51 and ST3), a *comX* allele is directly located between two ORFs belonging to TEs (Table 1). Furthermore, the *comX2* alleles of *S. agalactiae* ILRI112 and ILRI005 are located between IS*Sag5* and a gene coding for a putative site-specific phage integrase (Table 1).

TEs were identified in the IGRs surrounding *comX* alleles in 87.5%, 54.5%, 42.1%, 37.0%, and 30% of the strains belonging to the salivarius group, the anginosus group, the *S. suis* species, the pyogenic group, and the mitis group, respectively. They are nearly as frequently located downstream (45.6% of the TEs) or upstream (54.4% of the TEs) of this gene. No TE was identified in the close vicinity of *comX* alleles in the analysed strains of the mutans group (3 strains) and of the bovis group (6 strains) (Table 1).

TEs linked to *comX* are mainly related to ISs, whereas group II introns are also found, but at a far lesser frequency (98.2% and 1.8% of the TEs, respectively). These ISs belong to seven unrelated families as well as one unclassified group (Table 1 and Supplemental Fig. S4). Those inserted upstream of *comX* alleles are divided into four families: IS*L3*, IS*982*, IS*200*/IS*605*, and IS*3* (43.3%, 30.0%, 13.3% and 13.3% of the upstream ISs, respectively). Elements of the preponderant IS*L3* family are found in four species belonging to four different phylogenetic groups: *S. pneumoniae* of the mitis group, *S. thermophilus* of the salivarius group, *S. equi* subsp. *zooepidemicus* of the pyogenic group, and *S. intermedius* of the anginosus group. The ISs inserted downstream of *comX* alleles fall into three families (IS*30*, IS*As1*, and IS*110*)

Table 1 Transposable elements identified in the close vicinity of *comX* alleles of *Streptococcus* species

Species	Transposable element[a]					
	Downstream			Upstream		
	comX1	*comX2*	*comX3*	*comX1*	*comX2*	*comX3*
Non-pyogenic streptococci						
Anginosus group						
S. anginosus C238	IS*Stan1* IS*30* family	–	–	–	–	–
S. anginosus SA1	IS*Stan1*[†] IS*30* family	–	–	–	–	–
S. constellatus subsp. *pharyngis*						
C232	–	–	–	IS*Spn6*-like[†] IS*200*/IS*605* family	–	–
C818	–	–	–	IS*Spn6*-like[†] IS*200*/IS*605* family	–	–
C1050	–	–	–	IS*Spn6*-like[†] IS*200*/IS*605* family	–	–
S. *intermedius* B196	–	–	–	–	–	IS*1193*-like[†] IS*L3* family
Mitis group						
S. mitis B6	IS*Smi1* IS*30* family	IS*Smi1* IS*30* family	NA	–	–	NA
S. parasanguinis ATCC 15912	IS*1139*-like[†] IS*30* family	NA	NA	–		NA
S. parasanguinis FW213	IS*1139*-like[†] IS*30* family	NA	NA	–		NA
S. pneumoniae AP200	–	IS*Spn8*[†] IS*30* family	NA	–		NA
S. pneumoniae ATCC 700669	–	–	NA	IS*Spn14* IS*L3* family	–	NA
S. pneumoniae INV104	–	IS*Spn8* IS*30* family	NA	–	–	NA
S. pneumoniae SPN034156	–	IS*Spn8* IS*30* family	NA	–	–	NA
S. pneumoniae 70585	–	–	NA	IS*Spn14* IS*L3* family	–	NA
S. *pseudopneumoniae* IS7493	IS*1202*-like[†] Unclassified group	–	NA	–	–	NA
Salivarius group						
S. salivarius CCHSS3	–	NA	NA	IS*Strs1* IS*200*/IS*605* family	NA	NA
S. thermophilus ASCC 1275	–	NA	NA	IS*Sth8*[†,#] IS*L3* family	NA	NA
S. thermophilus CNRZ1066	–	NA	NA	IS*Sth8*[†,#] IS*L3* family	NA	NA
S. thermophilus LMD-9	–	NA	NA	IS*Sth8*[†,#] IS*L3* family	NA	NA
S. thermophilus LMG 18311	–	NA	NA	IS*Sth8*[†,#] IS*L3* family	NA	NA
S. thermophilus MN-ZLW-002	–	NA	NA	IS*Sth8*[†,#] IS*L3* family	NA	NA
S. thermophilus ND03	–	NA	NA	IS*Sth8*[†,#] IS*L3* family	NA	NA
S. suis						
S. suis D9	–	–	–	–	–	IS*Ssu12* IS*982* family[†]
S. suis ST1	–	–	–	–	IS*Ssu12* IS*982* family	IS*Ssu12* IS*982* family[•]
S. suis ST3	IS*Ssu7* IS*110* family	–	IS*Ssu7*[†] IS*110* family	–	–	IS*Ssu12*[†] IS*982* family
S. suis T15	–	NA	–	–	NA	IS*Ssu12* IS*982* family

Table 1 (continued)

Species	Transposable element[a]					
	Downstream			Upstream		
	comX1	*comX2*	*comX3*	*comX1*	*comX2*	*comX3*
S. suis TL13	–	–	–	–	–	IS*Ssu12*†,• IS*982* family
S. suis YB51	IS*Ssu7*† IS*110* family	–	IS*Ssu7* IS*110* family	–	–	IS*Ssu12*† IS*982* family
S. suis 05HAS68	–	–	–	–	–	IS*Ssu12*† IS*982* family
S. suis 6407	–	NA	–	–	NA	IS*Ssu12*† IS*982* family
Pyogenic streptococci						
S. agalactiae GD201008-001	–	NA	NA	intron group II	NA	NA
S. agalactiae ILRI005	–	–	NA	IS*Sag5**,† IS*3* family	IS*Sag5**,† IS*3* family	NA
S. agalactiae ILRI112	–	–	NA	IS*Sag5**,† IS*3* family	IS*Sag5*† IS*3* family	NA
S. dysgalactiae subsp. *equisimilis*:						
AC-2713	IS*1239*† IS*30* family	–	NA	–	–	NA
ATCC 12394	IS*1239* IS*30* family	–	NA	–	–	NA
GGS 124	IS*1239*† IS*30* family	–	NA	–	–	NA
167	IS*1239*† IS*30* family	–	NA	–	–	NA
S. equi subsp. *zooepidemicus*:						
ATCC 35246	–	–	NA	IS*Seq1*‡ IS*L3* family	IS*Seq1*‡ IS*L3* family	NA
CY	–	–	NA	IS*Seq1*‡ IS*L3* family	IS*Seq1*‡,† IS*L3* family	NA
S. pyogenes A20	IS*1548* IS*As1* family	–	NA	–	–	NA
S. pyogenes Alab49	IS*1548* IS*As1* family	–	NA	–	–	NA
S. pyogenes M1 GAS	IS*1548* IS*As1* family	–	NA	–	–	NA
S. pyogenes M1 476	IS*1548* IS*As1* family	–	NA	–	–	NA
S. pyogenes MGAS315	IS*1548* IS*As1* family	–	NA	–	–	NA
S. pyogenes MGAS5005	IS*1548* IS*As1* family	–	NA	–	–	NA
S. pyogenes SSI-1	–	IS*1548* IS*As1* family	NA	–	–	NA
S. pyogenes STAB902	–	IS*1548* IS*As1* family	NA	–	–	NA

–, Absence of a transposable element; NA (not applicable), allele of *comX* not present in the strain; †, pseudogene; ‡, presence of a tRNA-arg between the transposable element and the *comX* gene; *, presence of (a) small remnant(s) of (a) conserved hypothetical protein(s) between the transposable element and the *comX* gene; •, presence of an IS*Ssu13* element of the IS*L3* family upstream of IS*Ssu12*; #, presence of a small remnant of IS*Strs1* upstream of IS*Sth8*

[a]Proteins encoded by ORFs just upstream or downstream of *comX* alleles were blasted in the ISfinder or NCBI CDD databases. Annotated pseudogenes with no detectable ORF were identified after a blastN analysis. Truncated and fragmented ORFs were also classified as pseudogenes. In these last cases, the truncated or fragmented protein sequences were also blasted

that are different from those of the upstream inserted elements, as well as in the IS*1202* unclassified group of ISs (50.0%, 30.8%, 15.4%, and 3.8% of the downstream ISs, respectively). Elements of the preponderant IS*30* family are found in five species belonging to three different phylogenetic groups: *S. pneumoniae, S. mitis* and *S. parasanguinis* of the mitis group, *S. anginosus* of the anginosus group and *S. dysgalactiae* subsp. *equisimilis* of the pyogenic group.

The conservation of synteny of a given *comX* locus between strains belonging to the same species and possessing or not an IS inserted just close to the *comX* gene suggests that, in most cases, the presence of the IS is more likely due to individual insertion events than to major horizontal gene transfer events involving large genomic regions (Table 1 and Supplemental Table S3). This assumption is reinforced by the fact that a group II intron was also identified upstream of *comX1* in one strain of *S. agalactiae* (GD201008-001) (Table 1).

Phylogenetic analysis of the strains targeted by the transposable elements

In *S. pneumoniae*, IS*Spn8* is inserted at the *comX2* locus of strains which are not phylogenetically related. IS*Spn14* is also inserted at the *comX1* locus of two strains not phylogenetically linked. This suggests several independent insertion events of both ISs (Table 1; Fig. 2a). In *S. thermophilus*, the IS*Sth8* element is inserted at the *comX* locus of all the analysed strains, indicating one single insertion event in the ancestor of these strains (Table 1; Fig. 2c). In *S. suis*, the two strains possessing IS*Ssu7* at the *comX1* and *comX3* loci are phylogenetically linked, suggesting previous insertions in their last common ancestor (Table 1; Fig. 2b). On the contrary, insertions of IS*Ssu12* were found in strains belonging to three different phylogenetic groups (groups A, B and C, Fig. 2b), suggesting three independent insertion events. In *S. agalactiae*, only two of the analysed strains (ILRI005 and ILRI112) possess two *comX* alleles. In each of these strains, an IS*Sag5* element is present in the upstream region of each of these alleles. However, the ILRI005 and ILRI112 strains are not closely related (Table 1; Fig. 2f). The two *comX* alleles are identical within each of these strains but are different from one strain to the other, indicating that the acquisition of *comX2* is likely due to an internal duplication event that occurred independently in the two strains. IS*Sag5* insertions should thus have also occurred independently in the two strains. In all the analysed strains of *S. dysgalactiae* subsp. *equisimilis*, IS*1239* is located downstream of *comX1*. Nevertheless, it is divergently oriented in strain ATCC 12394 compared to the three other analysed strains indicating at least three distinct insertion events in the subspecies (Table 1; Fig. 2e). In *S. equi* subsp. *zooepidemicus*, IS*Seq1* insertions at the two *comX* alleles were identified in two strains (ATCC 35246 and CY) which are closely related, indicating insertion events in the ancestor of these strains (Table 1; Fig. 2g). In *S. pyogenes*, IS*1548* insertions downstream of one of the two *comX* alleles were detected in two clusters of strains (groups A and B) that are not closely related to each other (Table 1; Fig. 2d). The four members of group A possess IS*1548* at the *comX1* locus, suggesting only one insertion event in their last common ancestor. On the contrary, the absence of IS*1548* or even of the duplication of its target sequence at the *comX* loci of one of the five strains of group B (strain HSC5) suggests two independent insertion events within this group. Altogether, these data indicate that at least three independent insertion events of IS*1548* occurred in the downstream IGR of *comX* genes in *S. pyogenes*.

In conclusion, several different TEs targeted the *comX* alleles of streptococci belonging to various species, with either putative single insertion events in the ancestor of some groups of strains, or several insertion events in distinct strains.

Potential impact of the transposable elements on the genome plasticity

The identified TEs might have an impact on the genome plasticity. We tested this possibility by analysing the duplication of two competence-related genes.

The *amiA* homologs

An additional *amiA* homolog (*amiA2*), present in some strains of *S. thermophilus* (e.g., LMG 18311 and ND03), is located at the *mscS* locus 1 between two remnants of IS*Sth8* a few genes upstream of *comX1* (Supplemental Table S3). The presence of these remnants on both sides of *amiA2* suggests that IS*Sth8* was involved in its insertion at this locus. No other similar structure could be detected at another genomic location of *S. thermophilus*. However, an *amiA2* homolog is also present on a cis-mobilizable element (ΔCIME308) flanked on one side by a transposase gene showing the highest homology to IS*Sth8*, and on the other side by a DNA sequence of 178 bp displaying more than 90% identity with a few ISs of the IS*L3* family including IS*Sth8*, IS*Smu2* and IS*1193* (Pavlovic et al. 2004). These elements could have been involved in the mobilisation of *amiA2*. Garault et al. also previously proposed that the location of ISs on both sides of the *S. thermophilus amiA3* gene reveals its mobilisation by an IS-directed event (Garault et al. 2002). However, in this case, the mechanism involved is not clear as these ISs belong to different families.

The *comX* homologs

The presence of TEs close to *comX* alleles might also have impacted the gain, the loss or the maintenance of an additional copy of the gene. The acquisition of an additional *comX* allele may contribute to a gene dosage effect altering the amount of ComX produced after the activation of the competent state. Because the transformation rate of a given strain is generally correlated to the expression level of ComX (Blomqvist et al. 2006), the acquisition of an additional copy of *comX* may increase the efficiency of DNA uptake and recombination and so, have a positive impact on the integration of exogenous DNA into the genome. This hypothesis is nevertheless contradicted, at least in *S. pneumoniae* and *S. gordonii*, where the effect of the two *comX* alleles does not appear to be additive (Lee and Morrison 1999; Heng et al. 2006) Furthermore, due to their different genic environment and/or the putative drift of the regulatory sequences governing their expression, these multiple *comX* homologs may contribute to a more complex signalling network leading to the induction of competence. We have thus analysed in more details the relationships between TEs and *comX* homologs in *S. suis*.

Strains of *S. suis* possess a variable number of *comX* alleles, which can be found at three distinct loci [*ftsH* locus 1, *proS-rpsJ*, and *rplQ* loci, supplemental Table S3]. The different *S. suis* strains possess *comX1* alone or in addition to either a truncated *comX2* allele, a *comX3* allele, or both alleles. A horizontal gene transfer event from an exogenous source is unlikely to be the cause of the presence of *comX2* and *comX3*. Indeed, *comX1* and *comX3* of a given strain are identical (except in the case of strain 6407 in which they differ by only one mutation). Moreover, the truncated *comX2* alleles are more similar to the *comX1*/*comX3* alleles of *S. suis* than to *comX* genes from other species. The *comX3* allele, which is present in 42.1% of the analysed strains, is the only one always associated with ISs. In all cases, it is linked to IS*Ssu12*, an upstream IS of the IS*982* family (Table 1 and Supplemental Fig. S5). The alignment of the genomic sequences of two strains of *S. suis* comprising or not the *comX3* allele at the *rplQ* locus (the ST1 and BM407 strains, respectively), showed that *comX3* is not located on a genomic island (dot-plot analysis, Supplemental Fig. S6). As no IS is inserted at the *rplQ* locus in strains devoid of the *comX3* allele, it is possible that the ISs associated with *comX3* were involved in its acquisition or maintenance. We thus analysed the IGRs of representative *comX* loci of strains possessing both *comX1* and *comX3* (Supplemental Fig. S5). We found that a region of 37 to 59 bp located upstream of *comX1* is similar (93% identity in strain ST1) to a part of the IGR upstream of IS*Ssu12* (black rectangles, Supplemental Fig. S5) and that a next region of 196–200-bp is highly similar (99% identity in strain ST1) to the IGR between this IS and *comX3* (white rectangles, Supplemental Fig. S5). Moreover, the *comX3* alleles are always more similar to the *comX1* gene of the same strain (from 99 to 100% identity) than to *comX3* genes of other strains. As no trace of insertion of IS*Ssu12* could be detected at the *comX1* locus (not even that of a direct repeat), the above data suggest that the *comX3* allele results from individual duplication events of the *comX1* region that occurred independently in several strains, and that in a second time IS*Ssu12* inserted only into the *comX3* locus, possibly due to a particular genetic environment that could influence the structure of the DNA. By comparing the sequence homology of *comX1* and *comX3* in all the strains possessing the additional *comX3* allele, five probable independent duplication events leading to the appearance of *comX3* were found (one in each of the strains ST1, TL13, T15 and 6407, and one in the common ancestor of strains D9, 05HAS68, ST3 and YB51) (data not shown). This suggests that more independent insertion events of IS*Ssu12* than proposed in Fig. 2b could have occurred. Because the presence of *comX3* in *S. suis* strains is always associated with at least an upstream IS element, these ISs might be involved in *comX3* maintenance by modulating its expression. Alternatively, the insertion of IS*Ssu12* at the *comX3* locus should also decrease the length of the region of homology between the *comX1* and *comX3* loci and thus, reduce the probability of any homologous recombination events that would lead to the loss of one of the two *comX* alleles.

Discussion

Significant association of most early competence loci with transposable elements

In this work, we show that most of the main early competence genes are intensively targeted by TEs, and particularly *comX* loci (Table 1). ComX alleles are as often targeted in the strains which use the ComCDE or the ComRS system (in 36.6 and 37.8% of the strains, respectively). The search for TEs in the neighbouring of other key early competence genes revealed several insertion events particularly frequent in certain species, but failed to identify any locus targeted by a so large variety of elements and in so many *Streptococcus* species. In strains which use the ComCDE system, the *comCDE* operon is nearly never targeted (an insertion was detected in only one strain) and the *comAB* operon is only targeted in strains of *S. pneumoniae* (in 100% of the strains of this species). The lack of synteny conservation of the region located upstream of the *comAB* locus between *S. pneumoniae* and other species using the ComCDE system, such as *S. pseudopneumoniae, S. mitis* or *S. anginosus* (data not shown) suggests that a transposition hotspot, necessary

for ISs targeting, may be specific to the *S. pneumoniae* species. Similarly, in strains which use the ComRS system, the *comRS* locus is only targeted in *S. suis* (in 78.9% of the strains of this species). As in the case of the *comAB* operon, this could be due to the particular synteny of the region located downstream of *comRS*, that may contain a transposition hotspot only in *S. suis*. Indeed, the *comRS* genes of the analysed *S. suis* strains are located two or three ORFs downstream of the *ruvB* gene, whereas in the other species using the ComRS system (with the exception of the salivarius group, where the genetic environment of *comRS* is totally different), these genes are located just upstream of *ruvB* (data not shown). ISs were detected in the *ami*/*opp* region in 40.2% of the cases. Again, one species is largely concerned by this last targeting, as 100% of the *S. pyogenes* strains possess a remnant of an IS integrated downstream of the *oppF* gene. As the presence of TEs is otherwise not specifically linked to a random set of *Streptococcus* genes (MLST genes), their association with early competence genes should have a physiological significance that seems to particularly affects the expression of the master regulator ComX. Several studies have shed light on the critical importance of the expression level of ComX in streptococci, and artificial *comX* overexpression or addition of synthetic competence pheromones is a common strategy used to render non-competent strains transformable or, at least, to stimulate the expression of late competence genes (Blomqvist et al. 2006; Fontaine et al. 2010; Mashburn-Warren et al. 2012).

Acquisition of transposable elements into early competence loci during the evolution of streptococci

The study of the phylogenetic trees of the analysed species indicates that the acquisition of ISs close to early competence genes presumably arrived over different time scales during the evolution, some of them being acquired very early in the history of the species (e.g., IS*Spn7*/IS*1381* of *S. pneumoniae* or IS*861*-like of *S. pyogenes*) and others more recently (e.g., IS*1548* of *S. pyogenes*, IS*1167* of *S. pneumoniae*, IS*Sag5* of *S. agalactiae*) (Fig. 2). Several of the elements acquired early in the evolution are degenerated. For example, this is the case of the IS*Spn7*/IS*1381* element identified upstream of the *S. pneumoniae comAB* operon, of the IS*Ssu2* element of the types 1–2 inserted downstream of *comRS* in the group A of *S. suis*, or of the IS*861*-like element inserted downstream of the *S. pyogenes oppABCDF* operon. A substantial part of the ISs present in the close vicinity of *comX* alleles seem to have been acquired later during the evolution of streptococci (Fig. 2). However, 55.4% of the transposase genes of these elements are pseudogenes and are thus non-functional (Table 1). The presence of IS*Sth8* at the *comX* locus of *S. thermophilus* is an example of a more recent insertion event. The clonal species *S. thermophilus* is thought to have emerged only recently, about 7000 years ago, from a commensal ancestor of the salivarius group (Delorme et al. 2015). All of the analysed strains of *S. thermophilus*, but not those of *S. salivarius*, possess IS*Sth8* upstream of *comX* (Table 1). It seems thus that *S. thermophilus* has acquired IS*Sth8* after the emergence of the species from the ancestor of the salivarius group.

Insertion of ISs into early competence loci seems thus to be a regular process since the emergence of streptococci.

Specific insertion of transposable elements inside intergenic regions flanking early competence loci

With the only one exception of an insertion into the *comC* gene of *S. pneumoniae* strain AP200, all the TEs identified are inserted into IGRs. The ORFs of early competence genes are thus either not targeted by TEs, or such insertions are not conserved during the evolution, because they are detrimental for streptococci. In this context, it is worth noting that ISs were described to preferentially insert into superhelicity-induced duplex stabilization regions (SSID) that are significantly associated with IGRs containing promoters (Zhang et al. 2017; Wang et al. 2004). A completely opposite situation is found for genes encoding late components of the transformation machinery, in which numerous cases of integrative disruption by long MGEs were described in the literature. In particular, prophage or genomic island integrations occur into the *comYC* and *comFA* genes of streptococci belonging to different phylogenetic groups, or into the *comEC* gene of *S. pneumoniae*, respectively (Croucher et al. 2016). In streptococci, the early phase of competence has an additional function consisting on the control of bacteriocins production (Wang and Dawid 2018). As the predation mechanism is probably crucial for the persistence of bacteria in their ecological niche, this could explain why the genes governing the early phase, but not those encoding the transformasome, are preserved from disruption by MGEs.

Putative mutual influence on the regulation of gene expression

TEs inserted close to early competence genes might allow a modulation of the induction of competence under particular environmental constraints, by either providing the competence gene with additional promoter(s), disrupting a regulatory element involved in its expression, affecting the stability of the transcript or its expression by the synthesis of antisense RNAs from intrinsic outwarded promoter(s) (Casacuberta and González 2013). Besides, the presence of numerous putative transcriptional promoters and terminators was predicted in the sequence of the identified TEs (data

not shown). In the case of *comX* alleles, as in each species only one kind of IS is found at a particular *comX* locus, it is probable that these ISs induce specific adaptive mechanisms affecting the expression of each *comX* homolog in each species (Table 1). As *comX* is duplicated in numerous strains of streptococci, ISs inserted in their close vicinity would additionally complexify the regulatory network by providing a specialised response of a particular allele to a specific environmental stimulus. Further studies with knockout mutants of these ISs are necessary to evaluate their functional impact on the induction of competence. Conversely, early competence genes may also have an impact on the expression of the transposase genes, as it was previously described for some copies of ISWpi18_1 of *Wolbachia w*VulC which are transcribed from external promoters located in the genomic background, close to their insertion site (Cerveau et al. 2015). It was also suggested that the transposase genes of IS*1167* and IS*630-Spn1* of *S. pneumoniae* are part of the competence regulons, as they belong to the late CSP-induced genes (Peterson et al. 2004). It seems thus that the competence process might influence the transposition activity. Nevertheless, several protection mechanisms are used by ISs to prevent their fortuitous activation by external promoters, notably post-transcriptionally. For instance, in readthrough transcripts of IS*10*, the Ribosome Binding Site and the start codon of the transposase gene are sequestered into a stem-loop structure, preventing its translation (Davis et al. 1985). Further studies are thus needed to assess the production of the transposase protein and the transposition activity of these ISs whose transcription could be activated by the genomic context.

Association of particular transposable elements with several competence loci

The IS*1167*, IS*Spn8*, IS*1548*, IS*1239*, IS*Sth8* and IS*Ssu7* elements are associated with different competence genes in either the same strain, or the same species, suggesting their potential role in a coordinate regulation of several steps of the competence process.

In *S. pneumoniae*, IS*1167* is present upstream of *comX1* in the ATCC 700669 and 70585 strains but this IS is also present upstream of *comA* in the R6 and CP200 strains. An IS*1167A* element is likewise located upstream of the *ciaRH* operon in strains ATCC 700669 and JJA. IS*1167* and IS*1167A* are similar elements possessing 87.8% of identical nucleotides. The *ciaRH* operon codes for a two-component system, which negatively regulates the induction of competence of this species (Sebert et al. 2005). In some strains, an IS*1167*-like remnant is also located upstream of the *comW* gene (Martin et al. 2013). In *S. pneumoniae*, the *comW* gene, which belongs to the *comE* regulon, codes for a protein involved in both the stabilization and activation of ComX (Sung and Morrison 2005).

In three *S. pneumoniae* strains, IS*Spn8* is inserted downstream of *comX2* (Table 1). In strain SPN034156, this IS is also present at three other loci, two of which are also involved in the induction of competence. One copy of IS*Spn8* is inserted upstream of *spxA1*, a gene coding for a transcription regulator that represses the *comCDE* early competence operon and so, prevents the induction of the competent state. Another copy is inserted upstream of *pepO*, a gene encoding an endopeptidase able to inactivate CSP (Turlan et al. 2009; Bergé et al. 2002).

In some strains of *S. pyogenes*, IS*1548* targets the *comX1* or *comX2* downstream IGRs (Table 1). Homologs of the transposase gene of IS*1548* are also present just downstream of the *dprA-topA* genes of strain Alab49 and just downstream of the *cinA-recA-spxA1* genes of strain STAB901. These genes belong to the ComX regulon of other streptococcal species, and *dprA* and *cinA* are induced following XIP exposure in *S. pyogenes* (Boutry et al. 2012; Zhu et al. 2015; Mashburn-Warren et al. 2012). The DprA protein is required for the recruitment of the RecA recombinase onto single-stranded DNA and these two proteins mediate the homologous recombination process that can occur after DNA entry (Mortier-Barrière et al. 2007). DprA is also involved in the shut-off of competence of *S. pneumoniae* (Mirouze et al. 2013). We also found an IS*1548* element (*GGS_0054*) just downstream of the *comRS* genes of *S. dysgalactiae* subsp. *equisimilis* RE378. Moreover, we previously described that IS*1548* targets the promoter and the first gene of the *adcRCB* operon, which encodes a transcriptional repressor and subunits of a high-affinity zinc ABC transporter involved in the induction of competence in other streptococci and in *B. subtilis* (Fléchard et al. 2013a, b; Loo et al. 2003; Dintilhac et al. 1997; Ogura 2011). The disruption of the gene encoding the repressor of the *adc* operon and/or the putative promoter located at the right end of IS*1548* could have a direct impact on the transcription of the downstream genes (Fléchard et al. 2013a; Al Safadi et al. 2010).

In all the analysed strains of *S. dysgalactiae* subsp. *equisimilis*, an isoform of IS*1239* is inserted in the downstream IGRs of both *amiF* and *comX1*.

In *S. thermophilus*, IS*Sth8* is present upstream of *comX1* (Table 1) and is also inserted between the *amiA3* gene and the *amiA*$_1$*BCDF* operon (Supplemental Fig. S3).

Finally, in *S. suis*, IS*Ssu7* is inserted downstream of *comX1* and *comX3* in strains ST3 and YB51, and downstream of *comRS* in strains of the types 2 and 4 described above. This IS appears thus to target the major genes of the signalling cascade triggering the induction of competence in the *S. suis* species (Zaccaria et al. 2014).

Altogether, these data suggest that some ISs may specifically target competence genes.

Specific regulation of competence genes by/with mobile genetic elements

In bacteria, the existence of specific targeting mechanisms of a locus by one or several ISs was first proposed in the late 1980s and was called adaptive mutations (Hall 1998). These genetic events, still controversial, are thought to occur under some specific environmental conditions such as prolonged stationary phase in the presence of some specific carbon substrates (Hall 1998; Wang and Higgins 1994). In *B. subtilis*, the *comP* gene, which encodes the sensor histidine kinase of a two-component system involved in the early steps leading to the establishment of competence, was shown to be a hotspot for spontaneous insertions of IS*4Bsu1* under competence-developing conditions (Nagai et al. 2000; Takahashi et al. 2007). Other examples of targeting of key competence genes by MGEs involve prophages. A genetic switch due to the excision of a temperate phage was recently shown to induce the transformation machinery of *Listeria monocytogenes*. In this species, the DNA uptake system is cryptic, because *Listeria* prophage is inserted into the *comK* gene, that codes for the master regulator of competence. Prophage excision is specifically induced during intracellular growth and allows the formation of a functional *comK* gene that leads to the induction of the *com* regulon and promotes phagosomal escape and virulence (Rabinovich et al. 2012). Furthermore, a recent analysis of within-host evolution of *S. pneumoniae* isolates showed that competence can be transiently inactivated by prophage integrations into the *comYC* gene encoding a component of the transformation machinery (Croucher et al. 2016). Lastly, in *B. subtilis*, the Rok repressor was shown to co-regulate the induction of competence by downregulating the *comK* gene and the excision of a MGE by decreasing the expression of the genes of ICE*Bs1* (Smits and Grossman 2010).

The use of MGEs as sensors of the cellular physiological state to regulate the induction of competence by their insertion/excision under some specific conditions may be thus a widespread strategy among bacteria.

Benefits of regulating the induction of competence by mobile genetic elements

As competence is an energy-consuming process, the regulation of its induction by MGEs might aim at triggering transformability only under circumstances under which it provides an adaptive advantage for the strain. Congruently, it was shown that, in *S. pneumoniae*, the benefits of competence are context-dependent, because transformability is specifically advantageous under slightly stressful conditions (Engelmoer et al. 2013). Moreover, regulating the induction of competence by inserted TEs could also be part of a bet-hedging strategy allowing the emergence of a subpopulation of competent bacterial cells with an increased capacity to adapt to changing environmental conditions without exposing the whole population to the deleterious effects of the competent state. Indeed, induction of competence is an exhausting process, during which bacteria undergo growth arrest and cell wall alterations that can be detrimental for their survival (Zaccaria et al. 2016; Haijema et al. 2001; Bergé et al. 2017). This kind of bet-hedging may be particularly important for bacteria exposed to very diverse successive ecological niches, such as pathogens during the infection process. Lastly, as suggested for long MGEs such as prophages and ICEs whose insertion/excision in/from late competence genes can be used as a control switch of the transformability of their host strain allowing them to avoid their elimination from the genome, the insertions of TEs close to early competence genes might also have somehow an impact on their maintenance and spread in the genomes (Croucher et al. 2016).

In conclusion, this work sheds light on the importance of the targeting of streptococcal intergenic regions neighbouring early competence loci by ISs. Although all early competence loci are involved in this phenomenon, the association of ISs with the alleles coding for the central regulator of this process (ComX) is particularly strong and widespread among the main phylogenetic groups of streptococci. As this association suggests a possible co-regulation between the inserted ISs and the induction of competence, the questions of its consequences on streptococcal transformability and on the transposition efficiency of ISs, as well as of the mechanisms governing the targeting specificity, remain to be addressed by experimental approaches. Moreover, it would be interesting to evaluate the extent of this phenomenon among other groups of competent bacteria. These studies should bring new perspectives on our understanding of the regulation of competence and its role in genomic plasticity and adaptive evolution of prokaryotes.

Acknowledgements P. Siguier is thanked for her help during the submission of IS sequences to the ISfinder databank. Work in the teams of PH and BH was supported by the National Foundation for Scientific Research (FNRS) and the Research Department of the Communauté française de Belgique (Concerted Research Action). PH. is Senior Research Associate of the FNRS.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Al Safadi R, Amor S, Hery-Arnaud G, Spellerberg B, Lanotte P, Mereghetti L, Gannier F, Quentin R, Rosenau A (2010) Enhanced expression of *lmb* gene encoding laminin-binding protein in *Streptococcus agalactiae* strains harboring IS*1548* in *scpB-lmb* intergenic region. PLoS One 5(5):e10794

Avery OT, MacLeod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. J Exp Med 79(2):137–158

Bergé M, Langen H, Claverys JP, Martin B (2002) Identification of a protein that inactivates the competence-stimulating peptide of *Streptococcus pneumoniae*. J Bacteriol 184(2):610–613

Bergé MJ, Mercy C, Mortier-Barrière I, VanNieuwenhze MS, Brun YV, Grangeasse C, Polard P, Campo N (2017) A programmed cell division delay preserves genome integrity during natural genetic transformation in *Streptococcus pneumoniae*. Nat Commun 8(1):1621

Blomqvist T, Steinmoen H, Håvarstein LS (2006) Natural genetic transformation: a novel tool for efficient genetic engineering of the dairy bacterium *Streptococcus thermophilus*. Appl Environ Microbiol 72(10):6751–6756

Boutry C, Wahl A, Delplace B, Clippe A, Fontaine L, Hols P (2012) Adaptor protein MecA is a negative regulator of the expression of late competence genes in *Streptococcus thermophilus*. J Bacteriol 194(7):1777–1788

Brüssow H, Canchaya C, Hardt WD (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev 68(3):560–602

Camilli R, Bonnal RJ, Del Grosso M, Iacono M, Corti G, Rizzi E et al (2011) Complete genome sequence of a serotype 11A, ST62 *Streptococcus pneumoniae* invasive isolate. BMC Microbiol 11:25

Casacuberta E, González J (2013) The impact of transposable elements in environmental adaptation. Mol Ecol 22(6):1503–1517

Cerveau N, Gilbert C, Liu C, Garrett RA, Grève P, Bouchon D, Cordaux R (2015) Genomic context drives transcription of insertion sequences in the bacterial endosymbiont *Wolbachia wVulC*. Gene 564(1):81–86

Craig NL (1997) Target site selection in transposition. Annu Rev Biochem 66:437–474

Croucher NJ, Mostowy R, Wymant C, Turner P, Bentley SD, Fraser C (2016) Horizontal DNA transfer mechanisms of bacteria as weapons of intragenomic conflict. PloS Biol 14(3):e1002394

Davis MA, Simons RW, Kleckner N (1985) Tn10 protects itself at two levels from fortuitous activation by external promoters. Cell 43(1):379–387

Delorme C, Abraham AL, Renault P, Guédon E (2015) Genomics of *Streptococcus salivarius*, a major human commensal. Infect Genet Evol 33:381–392

Dintilhac A, Alloing G, Granadel C, Claverys JP (1997) Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. Mol Microbiol 25(4):727–739

Engelmoer DJ, Donaldson I, Rozen DE (2013) Conservative sex and the benefits of transformation in *Streptococcus pneumoniae*. PLoS Pathog 9(11):e1003758

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4):783–791

Fléchard M, Gilot P (2014) Physiological impact of transposable elements encoding DDE transposases in the environmental adaptation of *Streptococcus agalactiae*. Microbiology 160(7):1298–1315

Fléchard M, Gilot P, Héry-Arnaud G, Mereghetti L, Rosenau A (2013a) Analysis and identification of IS*1548* insertion targets in *Streptococcus agalactiae*. FEMS Microbiol Lett 340(1):65–72

Fléchard M, Rosenau A, Mereghetti L, Gilot P (2013b) Polymerase chain reaction with insertion sequence-specific and -unrelated primers: a new tool for the identification of IS*1548* insertion targets in *Streptococcus agalactiae*. J Microbiol Methods 94(1):22–24

Fontaine L, Dandoy D, Boutry C, Delplace B, de Frahan MH, Fremaux C, Horvath P, Boyaval P, Hols P (2010) Development of a versatile procedure based on natural transformation for marker-free targeted genetic modification in *Streptococcus thermophilus*. Appl Environ Microbiol 76(23):7870–7877

Fontaine L, Wahl A, Fléchard M, Mignolet J, Hols P (2015) Regulation of competence for natural transformation in streptococci. Infect Genet Evol 33:343–360

Garault P, Le Bars D, Besset C, Monnet V (2002) Three oligopeptide-binding proteins are involved in the oligopeptide transport of *Streptococcus thermophilus*. J Biol Chem 277(1):32–39

Gardan R, Besset C, Guillot A, Gitton C, Monnet V (2009) The oligopeptide transport system is essential for the development of natural competence in *Streptococcus thermophilus* strain LMD-9. J Bacteriol 191(14):4647–4655

Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS (1998) Tn*5*/IS*50* target recognition. Proc Natl Acad Sci USA 95(18):10716–10721

Granlund M, Oberg L, Sellin M, Norgren M (1998) Identification of a novel insertion element, IS*1548*, in group B streptococci, predominantly in strains causing endocarditis. J Infect Dis 177(4):967–976

Griffith F (1928) The significance of pneumococcal types. J Hyg 27(2):113–159

Haijema BJ, Hahn J, Haynes J, Dubnau D (2001) A ComGA-dependent checkpoint limits growth during the escape from competence. Mol Microbiol 40(1):52–64

Hall BG (1998) Activation of the *bgl* operon by adaptive mutation. Mol Biol Evol 15(1):1–5

Hallet B, Rezsöhazy R, Mahillon J, Delcour J (1994) IS*231*A insertion specificity: consensus sequence and DNA bending at the target site. Mol Microbiol 14(1):131–139

Håvarstein LS (2010) Increasing competence in the genus *Streptococcus*. Mol Microbiol 78(3):541–544

Heng NC, Tagg JR, Tompkins GR (2006) Identification and characterization of the loci encoding the competence-associated alternative sigma factor of *Streptococcus gordonii*. FEMS Microbiol Lett 259(1):27–34

Kjos M, Miller E, Slager J, Lake FB, Gericke O, Roberts IS, Rozen DE, Veening JW (2016) Expression of *Streptococcus pneumoniae* bacteriocins is induced by antibiotics via regulatory interplay with the competence system. PLoS Pathog 12(2):e1005422

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874

Labbate M, Case RJ, Stokes HW (2009) The integron/gene cassette system: an active player in bacterial adaptation. Methods Mol Biol 532:103–125

Lee MS, Morrison DA (1999) Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. J Bacteriol 181(16):5004–5016

Loo CY, Mitrakul K, Voss IB, Hughes CV, Ganeshkumar N (2003) Involvement of the *adc* operon and manganese homeostasis in *Streptococcus gordonii* biofilm formation. J Bacteriol 185(9):2887–2900

Maguin E, Prévost H, Ehrlich SD, Gruss A (1996) Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. J Bacteriol 178(3):931–935

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY et al (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res 43(Database issue):D222–D226

Marks LR, Mashburn-Warren L, Federle MJ, Hakansson AP (2014) *Streptococcus pyogenes* biofilm growth in vitro and in vivo and its role in colonization, virulence, and genetic exchange. J Infect Dis 210(1):25–34

Martin B, Quentin Y, Fichant G, Claverys JP (2006) Independent evolution of competence regulatory cascades in streptococci? Trends Microbiol 14(8):339–345

Martin B, Soulet AL, Mirouze N, Prudhomme M, Mortier-Barrière I, Granadel C, Noirot-Gros MF, Noirot P, Polard P, Claverys JP (2013) ComE/ComE ~ P interplay dictates activation or extinction status of pneumococcal X-state (competence). Mol Microbiol 87(2):394–411

Mashburn-Warren L, Morrison DA, Federle MJ (2012) The cryptic competence pathway in *Streptococcus pyogenes* is controlled by a peptide pheromone. J Bacteriol 194(17):4589–4600

Mirouze N, Bergé MA, Soulet AL, Mortier-Barrière I, Quentin Y, Fichant G et al (2013) Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. Proc Natl Acad Sci USA 110(11):E1035–E1044

Morrison DA, Guédon E, Renault P (2013) Competence for natural genetic transformation in the *Streptococcus bovis* group streptococci *S. infantarius* and *S. macedonicus*. J Bacteriol 195(11):2612–2620

Mortier-Barrière I, Velten M, Dupaigne P, Mirouze N, Piétrement O, McGovern S et al (2007) A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell 130(5):824–836

Moscoso M, García E (2009) Transcriptional regulation of the capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae*: a bioinformatic analysis. DNA Res 16(3):177–186

Nagai T, Tran LS, Inatsu Y, Itoh Y (2000) A new IS*4* family insertion sequence, IS*4Bsu1*, responsible for genetic instability of poly-gamma-glutamic acid production in *Bacillus subtilis*. J Bacteriol 182(9):2387–2392

Oberto J (2013) SyntTax: a web server linking synteny to prokaryotic taxonomy. BMC Bioinform 14:4

Ogura M (2011) ZnuABC and ZosA zinc transporters are differently involved in competence development in *Bacillus subtilis*. J Biochem 150(6):615–625

Pakula R, Fluder Z, Hulanicka E, Walczak W (1958) Studies on transformation of streptococci. Bul Acad Polon Sci Ser Sci Biol 6(8):319–323

Pavlovic G, Burrus V, Gintz B, Decaris B, Guédon G (2004) Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICE*St1*-related elements from *Streptococcus thermophilus*. Microbiology 150(4):759–774

Perry D, Kuramitsu HK (1981) Genetic transformation of *Streptococcus mutans*. Infect Immun 32(3):1295–1297

Peterson SN, Sung CK, Cline R, Desai BV, Snesrud EC, Luo P et al (2004) Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. Mol Microbiol 51(4):1051–1070

Póntigo F, Moraga M, Flores SV (2015) Molecular phylogeny and a taxonomic proposal for the genus *Streptococcus*. Genet Mol Res 14(3):10905–10918

Prudhomme M, Berge M, Martin B, Polard P (2016) Pneumococcal competence coordination relies on a cell-contact sensing mechanism. PLoS Genet 12(6):e1006113

Rabinovich L, Sigal N, Borovok I, Nir-Paz R, Herskovits AA (2012) Prophage excision activates *Listeria* competence genes that promote phagosomal escape and virulence. Cell 150(4):792–802

Reck M, Tomasch J, Wagner-Döbler I (2015) The alternative sigma factor SigX controls bacteriocin synthesis and competence, the two quorum sensing regulated traits in *Streptococcus mutans*. PLoS Genet 11(7):e1005353

Reynolds AE, Felton J, Wright A (1981) Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K12. Nature 293(5834):625–629

Rhee DK, Morrison DA (1988) Genetic transformation in *Streptococcus pneumoniae*: molecular cloning and characterization of *recP*, a gene required for genetic recombination. J Bacteriol 170(2):630–637

Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, Town CD, Burne RA, Stanhope MJ (2014) Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. Genome Biol Evol 6(4):741–753

Sebert ME, Patel KP, Plotnick M, Weiser JN (2005) Pneumococcal HtrA protease mediates inhibition of competence by the CiaRH two-component signaling system. J Bacteriol 187(12):3969–3979

Sengstag C, Iida S, Hiestand-Nauer R, Arber W (1986) Terminal inverted repeats of prokaryotic transposable element IS*186* which can generate duplications of variable length at an identical target sequence. Gene 49(1):153–156

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34(Database issue):D32–D36

Siguier P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev 38(5):865–891

Smits WK, Grossman AD (2010) The transcriptional regulator Rok binds A + T-rich DNA and is involved in repression of a mobile genetic element in *Bacillus subtilis*. PloS Genet 6(11):e1001207

Sung CK, Morrison DA (2005) Two distinct functions of ComW in stabilization and activation of the alternative sigma factor ComX in *Streptococcus pneumoniae*. J Bacteriol 187(9):3052–3061

Swinscow T (1978) The chi square tests. In: Dawson, Ltd G (eds) Statistics at square one. The Mendip Press, Bath, pp 42–53

Takahashi K, Sekine Y, Chibazakura T, Yoshikawa H (2007) Development of an intermolecular transposition assay system in *Bacillus subtilis* 168 using IS*4Bsu1* from *Bacillus subtilis* (natto). Microbiology 153(8):2553–2559

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3):512–526

Toro N, Jiménez-Zurdo JI, García-Rodríguez FM (2007) Bacterial group II introns: not just splicing. FEMS Microbiol Rev 31(3):342–358

Turlan C, Prudhomme M, Fichant G, Martin B, Gutierrez C (2009) SpxA1, a novel transcriptional regulator involved in X-state (competence) development in *Streptococcus pneumoniae*. Mol Microbiol 73(3):492–506

Wang CY, Dawid S (2018) Mobilization of bacteriocins during competence in streptococci. Trends Microbiol 26(5):389–391

Wang X, Higgins NP (1994) 'Muprints' of the *lac* operon demonstrate physiological control over the randomness of in vivo transposition. Mol Microbiol 12(4):665–677

Wang H, Noordewier M, Benham CJ (2004) Stress-induced DNA duplex destabilization (SIDD) in the *E. coli* genome: SIDD sites are closely associated with promoters. Genome Res 14(8):1575–1584

Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev 35(5):957–976

Wozniak RA, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 8(8):552–563

Zaccaria E, van Baarlen P, de Greeff A, Morrison DA, Smith H, Wells JM (2014) Control of competence for DNA transformation in *Streptococcus suis* by genetically transferable pherotypes. PLoS One 9(6):e99394

Zaccaria E, Wells JM, van Baarlen P (2016) Metabolic context of the competence-induced checkpoint for cell replication in *Streptococcus suis*. PloS One 11(5):e0153571

Zerbib D, Gamas P, Chandler M, Prentki P, Bass S, Galas D (1985) Specificity of insertion of IS*1*. J Mol Biol 185(3):517–524

Zhang Z, Saier MH Jr (2011) Transposon-mediated adaptive and directed mutations and their potential evolutionary benefits. J Mol Microbiol Biotechnol 21(1–2):59–70

Zhang Z, Saier MH Jr (2016) Transposon-mediated activation of the *Escherichia coli glpFK* operon is inhibited by specific DNA-binding proteins: implications for stress-induced transposition events. Mutat Res 793–794:22–31

Zhang Z, Kukita C, Humayun MZ, Saier MH Jr (2017) Environment-directed activation of the *Escherichia coli flhDC* operon by transposons. Microbiology 163(4):554–569

Zhou L, Hui FM, Morrison DA (1995) Characterization of IS*1167*, a new insertion sequence in *Streptococcus pneumoniae*. Plasmid 33(2):127–138

Zhu L, Lin J, Kuang Z, Vidal JE, Lau GW (2015) Deletion analysis of *Streptococcus pneumoniae* late competence genes distinguishes virulence determinants that are dependent or independent of competence induction. Mol Microbiol 97(1):151–165