RESEARCH LETTER

Nucleotide sequence and functional analysis of the *tet*(M)-carrying conjugative transposon Tn*5251* of *Streptococcus pneumoniae*

Francesco Santoro, Marco R. Oggioni, Gianni Pozzi & Francesco Iannelli

Laboratory of Molecular Microbiology and Biotechnology (LAMMB), Department of Molecular Biology, University of Siena, Siena, Italy

Abstract

The Tn*916*-like genetic element Tn*5251* is part of the composite conjugative transposon (CTn) Tn*5253* of *Streptococcus pneumoniae*, a 64.5-kb chromosomal element originally called Ω(*cat-tet*) BM6001. DNA sequence analysis showed that Tn*5251* is 18 033-bp long and contains 22 ORFs, 20 of which have the same direction of transcription. Annotation was possible for 11 out of 22 ORFs, including the *tet*(M) tetracycline resistance gene and *int* and *xis* involved in the integration/excision process. Autonomous copies of Tn*5251* were generated during matings of Tn*5253*-containing donors with *S. pneumoniae* and *Enterococcus faecalis*. Tn*5251* was shown to integrate at different sites in the bacterial chromosome. It behaves as a fully functional CTn capable of independent conjugal transfer to a variety of bacterial species including *S. pneumoniae*, *Streptococcus gordonii*, *Streptococcus pyogenes*, *Streptococcus agalactiae*, *E. faecalis* and *Bacillus subtilis*. The excision of Tn*5251* produces a circular intermediate and a deletion in Tn*5253* at a level of 1.2 copies per 10^5 chromosomes.

Introduction

A large proportion of clinical isolates of *Streptococcus pneumoniae* (pneumococcus) contain the *tet*(M) gene conferring resistance to tetracycline antibiotics by ribosomal protection (Pozzi et al., 1986). The *tet*(M) gene is usually carried by genetic elements of the Tn*916*–Tn*1545* family of conjugative transposons (CTns) (Clewell et al., 1995; Rice, 1998), and eight out of the 36 pneumococcal genomes available in public databases contain this element.

CTns, also called integrative and conjugative elements (ICEs), are mobile genetic elements (MGEs) that can integrate into the bacterial genome, excise from it and form a covalently closed circular intermediate (CI) that can integrate into a new position of the genome (intracellular transposition) and/or can be transferred by conjugation from a donor to a recipient cell, where it integrates into the genome (intercellular transposition) (Churchward, 2002; Mullany et al., 2002). CTns enable horizontal transfer of genes among distantly related bacteria playing an important role in the molecular evolution of many bacterial genomes (Frost et al., 2005). CTns contribute to the dissemination of antibiotic resistance determinants among pathogenic bacteria and their association is responsible for the spread of

multiple antibiotic resistance determinants (Clewell et al., 1995; Rice, 2002; Roberts & Mullany, 2009). Among the best-studied CTns are (1) Tn*916*, originally found in the *Enterococcus faecalis* DS16 clinical strain, 18 032 bp in size and carrying the *tet*(M) tetracycline resistance gene (Franke & Clewell, 1981; Flannagan et al., 1994), and (2) Tn*1545*, found in the *S. pneumoniae* BM4200 clinical isolate, about 25.3 kb in length (GenBank X04388, X61025, X05577, X52632, AM903082, AM889142), related to Tn*916*, but carrying, in addition to *tet*(M), the *aphA-3* and *ermAM* genes conferring resistance to kanamycin and erythromycin (Courvalin & Carlier, 1986; Cochetti et al., 2008).

Tn*916*-like CTns are found integrated at different sites in the pneumococcal chromosome, and in many cases, they do not exist as individual CTns, but are part of other genetic elements (Fig. 1). The Tn*916*-like CTn Tn*5251* was shown to be part of the composite pneumococcal CTn Tn*5253* (Shoemaker et al., 1979; Ayoubi et al., 1991; Provvedi et al., 1996), a chromosomal genetic element originally called Ω(*cat-tet*) BM6001 (Shoemaker et al., 1979). Tn*5253*-related elements have been reported to be common in antibiotic-resistant pandemic *S. pneumoniae* clones (Henderson-Begg et al., 2008). In our previous paper, we demonstrated that Tn*5251* is able to excise from Tn*5253* and form CIs

MICROBIOLOGY LETTERS

FEMS
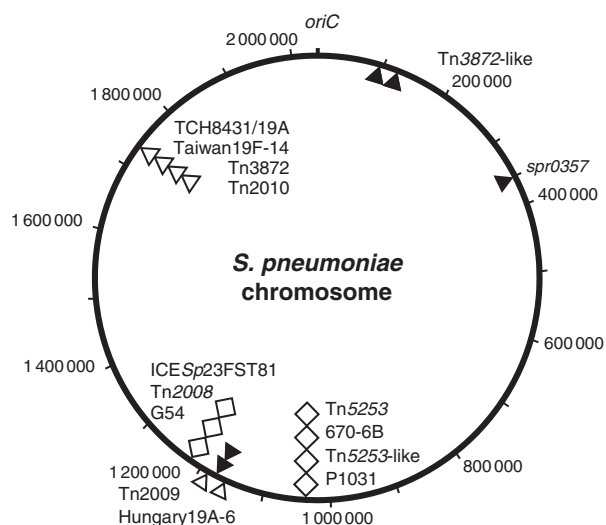
Fig. 1. Integration sites of Tn*5251* and Tn*5251*-like elements in the *Streptococcus pneumoniae* chromosome. Bacterial chromosome is represented by a circle, and the origin of replication (*oriC*) is reported; nucleotides indicated on the map refer to the R6 strain (Hoskins *et al.*, 2001). The solid triangles (▼) on the circle indicate the insertion sites of Tn*5251* in transconjugants obtained in this work. Tn*5251* insertions were mapped to *spr0357* at position 357 477, to three intergenic regions at positions 96 694, 122 345 and at position 1 175 526 in two transconjugants obtained from different matings. Tn*5251*-like CTns insertion sites (empty triangles, ▽) obtained from published nucleotide sequences map to four sites in the chromosome. Composite CTns (empty squares, ◇) containing a copy of a Tn*5251*-like CT integrate only at two sites that represent two hotspots for insertion. Elements are designed with their name or, when not annotated, with the name of the strain harbouring them.

(Provvedi *et al.*, 1996). Here, we report the complete annotated sequence of Tn*5251*, describe how autonomous copies of this element are generated upon conjugal transfer and show that Tn*5251* is in fact a fully functional CTn capable of independent conjugal transfer to a variety of bacterial species.

Materials and methods

Bacterial strains, media and growth condition

The bacterial strains used in this work and their relevant properties are reported in Table 1. Streptococci and enterococci were routinely grown in tryptic soy broth or tryptic soy agar (Difco) supplemented with 3% horse blood and, where appropriate, with antibiotics (Iannelli & Pozzi, 2007). *Bacillus subtilis* was grown in Luria–Bertani broth (LB) or LB agar.

DNA purification

Bacterial cells were harvested by centrifugation at the end of exponential phase growth. Pneumococcal cells were lysed

for 15 min at 37 °C in sodium dodecyl sulphate (SDS) 0.008% and sodium deoxycholate (DOC) 0.1% (lysis solution), whereas enterococcal cells were lysed according to the protocol already described (Manganelli *et al.*, 1995). DNA was purified using the Wizard Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions. Purified DNA was quantified by both electrophoretic and spectrophotometric analysis.

PCR and sequencing

Long PCRs were carried out using the Expand High Fidelity PCR System (Roche) essentially according to the protocol already described (Iannelli *et al.*, 1998). Briefly, the 25 µL reaction mixture was in 1× Expand High Fidelity buffer and contained (1) 1.5 mM MgCl$_2$, (2) 100 µM dNTPs, (3) 10 pmol of each primer, (4) 0.2 U of Expand High Fidelity Enzyme Mix and (5) 1 µL of liquid bacterial culture (Iannelli *et al.*, 1998). Amplification was performed using the following cyclic thermal profile: 1 cycle at 92 °C for 2 min, then 30 cycles at 50 °C for 10 s, 68 °C for 10 min, 92 °C for 10 s, and 1 cycle at 50 °C for 1 min and 68 °C for 20 min. The direct automated sequencing of the PCR fragments was performed using a primer walking strategy as described (Iannelli *et al.*, 1998). Two primer pairs IF487/IF393 and IF394/IF488 were used to amplify two fragments 5518 and 13 743 bp in length, respectively. Primers are directed to the already sequenced *tet*(M) and Tn*5251* flanking regions (Provvedi *et al.*, 1996): IF487 (5′-TTC GCT GAA GAC CTT TAT TCG-3′) is complementary to nucleotides 358 through 378 of the Tn*5251* left junction (GenBank X90940); IF488 (5′-TCC TCC TGA TTC CAG TGT CA-3′) corresponds to nucleotides 52 through 71 of the Tn*5251* right junction (GenBank X90941); and IF393 (5′-TTC TGC CGA AAT TGT AAT CA-3′) corresponds to nucleotides 2541 through 2560 and IF394 (5′-GCT ATA GTA TAA GCC ATA CT-3′) and is complementary to nucleotides 3602 through 3621 of Tn*5251* *tet*(M) (GenBank X90939). To confirm the sequence on the other strand, fragments about 1000 bp in size were produced by PCR and used as sequencing starting templates.

Quantitative nested PCR

Quantitative nested PCR was performed essentially as reported previously (Manganelli *et al.*, 1995). The 25 µL reaction mixture was in 1× DreamTaq buffer and contained (1) 2 mM MgCl$_2$, (2) 75 µM dNTPs and (3) 0.4 U of DreamTaq DNA Polymerase (Fermentas). DNA was denaturated at 92 °C for 2 min, and then the cyclic thermal profile was as follows: annealing at 50 °C for 10 s, extension at 72 °C for 30 s and denaturation at 92 °C for 10 s, followed by a final step at 50 °C for 1 min and 72 °C for 5 min. In the first 25 cycles of PCR, 5 pmol of each outer primer was used with serial dilutions of the chromosomal DNA as the

Table 1. Bacterial strains and relevant properties

Strain	Relevant properties*	Reference
D39	*S. pneumoniae* type 2 Avery's strain	Pearce *et al.* (2002), Lanie *et al.* (2007)
TIGR4	*S. pneumoniae* type 4	Tettelin *et al.* (2001)
Rx1	Unencapsulated derivative of type 2 D39 strain	Shoemaker & Guild (1974)
DP1002	Rx1 but *nov-1*; Nov[R]	Guild & Shoemaker (1976), Munoz *et al.* (1995)
DP1004	Rx1 but *str-41*; Sm[R]	Guild & Shoemaker (1976), Salles *et al.* (1992)
DP1322	Rx1 carrying Tn*5253*; Cm[R], Tc[R]	Smith *et al.* (1981)
FP5	Rx1, but Δ*comC*; Cm[R]	Iannelli *et al.* (2005)
FP10	FP5, but *str-41* (by transformation with DP1004 chromosomal DNA); Cm[R], Sm[R]	Iannelli & Pozzi (2007)
FP11	FP5, but *nov-1* (by transformation with DP1002 chromosomal DNA); Cm[R], Nov[R]	Iannelli & Pozzi (2007)
FP22	Unencapsulated derivative of D39; Km[R]	Pearce *et al.* (2002)
FP23	Unencapsulated derivative of TIGR4; Km[R]	Pearce *et al.* (2002), Oggioni *et al.* (2003)
FP47	TIGR4, but *nov-1* (by transformation with DP1002 chromosomal DNA); Nov[R]	This study
FR22	FP10 carrying Tn*5253* (by conjugation with DP1322); Cm[R], Sm[R], Tc[R]	This study
GP204	*S. gordonii* Challis V288, but *str-204*; Sm[R]	Pozzi *et al.* (1988)
SF370	*S. pyogenes* serotype M1	Ferretti *et al.* (2001)
H36B	*S. agalactiae* serotype Ib	Tettelin *et al.* (2005)
OG1RF	*E. faecalis* Fus[R], Rif[R]	Bourgogne *et al.* (2008)
OG1SS	*E. faecalis* Spe[R], Sm[R]	Dunny *et al.* (1979)
JH2-2	*E. faecalis* Fus[R], Rif[R]	Jacob & Hobbs(1974)
168	*B. subtilis*	Spizizen (1958)

nov-1, and *str-41* indicate point mutations conferring resistance to novobiocin and streptomycin, respectively (Guild & Shoemaker, 1976; Salles *et al.*, 1992; Munoz *et al.*, 1995).
Cm, chloramphenicol; Fus, fusidic acid; Km, kanamycin; Nov, novobiocin; Rif, rifampicin; Spe, spectinomycin; Sm, streptomycin; Tc, tetracycline.

starting templates. The second 30 cycles of PCR were performed with 10 pmol of each inner primer and 1 μL of the first PCR product as a template. The primers used to produce the 357-bp outer fragment were IF485 (5′-CTA TGT TTA CGC TTT CAA TCA A-3′) and IF486 (5′-AGA ACC ACT GAC ACC AAG TAT-3′), whereas the 141-bp inner fragment was obtained with IF487 and IF488.

Inverse PCR

In a final volume of 50 μL, 1 μg of chromosomal DNA was incubated with 10 U of *Sau*3A (Roche) at 37 °C for 2 h. One microlitre of digested DNA (20 ng) was circularized in a 20-μL reaction mix containing 10 U of T4 DNA Ligase (Roche) at 16 °C for 2.5 h. Ligated DNA was extracted with phenol and phenol–chloroform, precipitated in ethanol and resuspended in 15 μL of water. Five microlitres of purified ligated DNA were used as a template in PCR experiments carried out with the divergent primers IF505 (5′-CGT GAA GTA TCT TCC TAC AGT-3′) and IF452 (5′-ACT CAT TCT AAT AGC CCA TTC-3′) or with IF433 (5′-GGT GGA ACT TAT CAA TCC CAT-3′) and IF506 (5′-GGA TAA ATC GTC GTA TCA AAG-3′).

DNA sequence analysis

DNA sequence analysis including coding sequence identification was carried out using the software ARTEMIS ver. 11 available for download at http://www.sanger.ac.uk/Software/

Artemis/website. Manual gene annotation was carried out by conducting BLAST homology searches of the databases available at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/sites/gquery) and at the *S. pneumoniae* Sybil website (http://strepneumo-sybil.igs.umaryland.edu/). Protein domains were identified by searching the protein family database Pfam available at the Wellcome Trust Sanger Institute (http://pfam.sanger.ac.uk). Multiple sequence alignments were performed using the CLUSTALW2 tool at the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/clustalw2/).

Mating experiments

Plate mating experiments were performed essentially as already described (Iannelli & Pozzi, 2007). Donor and recipient cells were grown separately in TSB in the presence of appropriate antibiotics at 37 °C, until the end of the exponential phase ($OD_{590 nm} = 0.5$). Cells were mixed at a 1 : 10 ratio, harvested by centrifugation for 15 min at 3000 ***g***, resuspended in 0.1 mL of TSB and plated on TSA enriched with 5% horse blood. Following 4 h of incubation in 5% CO_2 at 37 °C, cells were harvested by scraping the plates with a sterile plain swab and resuspended in 1 mL of TSB containing 10% glycerol. Selection of transconjugants was carried out with the multilayer plating. Briefly, 2 mL of TSB/10% horse blood containing the appropriately diluted mating reactions were combined with 6 mL of melted TSA and

poured into a Petri dish containing a base layer of TSA. After 90 min of incubation at 37 °C for phenotypic expression, an 8 mL TSA layer containing the appropriate antibiotics, for the resistance marker of the donor genetic element and for the chromosomal resistance marker of recipient strain (where available), was added. The antibiotic concentrations were as follows: chloramphenicol 5 μg mL^{-1}, fusidic acid 25 μg mL^{-1}, novobiocin 10 μg mL^{-1}, rifampicin 25 μg mL^{-1}, spectinomycin 400 μg mL^{-1}, streptomycin 1000 μg mL^{-1} and tetracycline 5 μg mL^{-1}. Conjugation frequencies were determined by plating each parent strain alone. At this stage, we carefully performed genetic analysis of the transconjugants in order to exclude isolation of spontaneous mutants or colonies that might grow even in the absence of any genotype conferring resistance. Colonies were transferred to four different plates: (1) a plate containing both antibiotics for the resistance marker of Tn*5251* and for the chromosomal resistance marker of the recipient strain; (2) a plate containing the antibiotic for the resistance marker of Tn*5251*; (3) a plate containing the antibiotic for the chromosomal resistance marker of the recipient strain; and (4) a plate containing no antibiotics. To rule out the contibution of transformation to the genetic exchange during conjugation, we used as *S. pneumoniae* recipients strains FP10 and FP11 impaired in natural competence for genetic transformation (Table 1) (Iannelli & Pozzi, 2007), while matings with other bacterial species as recipients were carried out in presence of 10 μg mL^{-1} DNAse I. In mating assays where *S. pneumoniae* was the donor and the recipient strains lacked the chromosomal resistance marker, selection of transconjugants was obtained after addition of 0.05% DOC to the mating mixture, followed by incubation at 37 °C for 15 min. Selection of *B. subtilis* transconjugants was obtained on LB agar plates.

Nucleotide sequence accession number

Tn*5251* sequence was assigned GenBank accession no. FJ711160.

Results and discussion

Nucleotide sequence of Tn*5251*

The complete nucleotide sequence of Tn*5251* was determined in *S. pneumoniae* strain DP1322 (Smith *et al.*, 1981) using as sequencing templates two long PCR fragments spanning the element, and confirmed using smaller PCR fragments as templates. DNA sequence analysis showed that Tn*5251* is 18 033 bp in length, with an overall GC content of 38.8%. It contains 22 ORFs, 20 of which have the same direction of transcription (Fig. 2a). Sequence alignment showed that Tn*5251* is essentially identical to CTns of the Tn*916*–Tn*1545* family. The Tn*5251* DNA sequence differs from Tn*916* in only 79 nucleotide (nt) changes, of which 69 are clustered in the *tet*(M) coding sequence (Provvedi *et al.*, 1996). Comparison of the Tn*5251* DNA sequence with 10 pneumococcal Tn*5251*-like elements, whose complete sequences are available in the public database, indicates that eight out 10 have insertions of either the 'mega' element carrying the macrolide efflux gene *mef*(E) (Del Grosso *et al.*, 2006) or three different *erm*(B)-carrying elements, each at a specific site (Fig. 2b). Two more strains carry defective Tn*5251*-like elements. Strain G54 (GenBank CP001015) contains a truncated form of a Tn*5251*-like element. Truncation occurs downstream of *tet*(M) at nt 3869, which is the site of insertion of mega in other elements. Furthermore, the *erm*(B)-carrying element inserts into *orf20* at nt 14 194. Strain CGSP14 (GenBank CP001033) contains a Tn*5251*-like element where the insertion of a 9-kb genetic element carrying *sat4*, *aphIII* and two *erm*(B) genes produces a deletion spanning from *orf6* to *orf19* (Ding *et al.*, 2009). The *tet*(M) gene, which is known to have a mosaic structure (Oggioni *et al.*, 1996), exhibits the largest variability, with the exception of Tn*2010* (GenBank AB426620) and the Taiwan19F-14 element (GenBank CP000921) (Fig. 2b). In Tn*2010*, 46 nt changes are clustered in *orf16* and *orf20*, resulting only in two amino acid substitutions in Orf20. Tn*6003* (Genbank AM410044) has 339 nt changes including two deletions that introduce frameshifts in *tet*(M) and *orf24*, and two changes causing premature stop codons in *orf5* and *orf14*. The same *orf14* premature stop codon is present in the Tn*5253*-like element (GenBank FM201786). The element of strain ATCC 700669 (GenBank FM211187) has a deletion of 1 nt in the *int* coding sequence that determines a frameshift.

Completely sequenced Tn*916*/Tn*5251*-like elements are also present in many other bacterial species including *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus parauberis*, *Streptococcus suis*, *Streptococcus cristatus*, *Streptococcus intermedius*, *E. faecalis* and *Ureaplasma urealyticum*. Noteworthy, as found in *S. pneumoniae*, many of these elements are inserted into other chromosomal genetic elements, such as the 89 K pathogenicity island present in recently sequenced strains of *S. suis* (Chen *et al.*, 2007).

ORFs of Tn*5251*

All ORFs, with the exception of *orf10*, *orf20* and *orf24*, are preceded by a ribosome-binding site. The start codon of *orf6* is missing; *orf15* starts with the GTG codon and *orf20* with TTG. By manual annotation, it was possible to attribute a putative function only to 11 out of 22 ORFs (Table 2). Predicted gene products were blasted against public protein databases and the Pfam protein family database, taking into account significant homologies with functionally characterized proteins or good matches with Pfam domains. As reported for Tn*916*, *int* and *xis* are known to be involved in
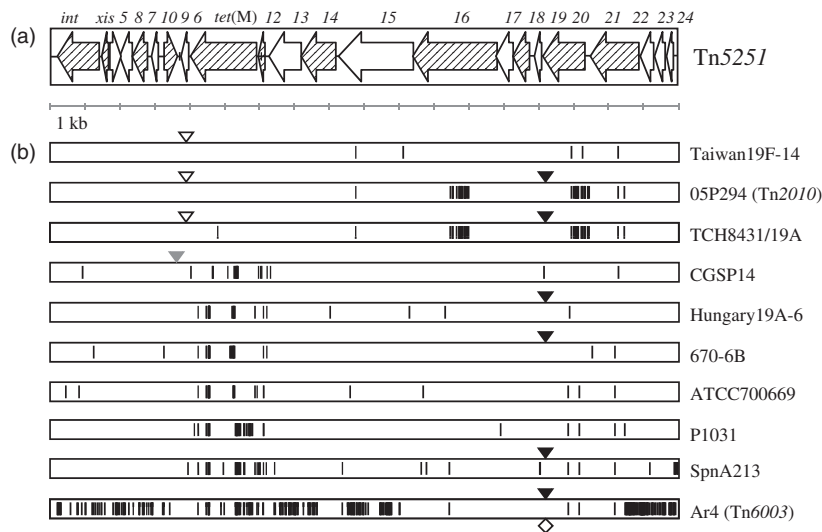
Fig. 2. (a) Structure of *Streptococcus pneumoniae* Tn*5251* and (b) schematic comparison with other pneumococcal Tn*5251*-like elements. (a) Sequence analysis of Tn*5251* showed that it is 18 033 bp in size and contains 22 ORFs. ORFs and their direction of transcription are represented by arrows and indicated by their numbers, whereas *tet*(M), *int* and *xis* are indicated with their name. By manual annotation, a putative function was attributed only to 11 out 22 ORFs (striped arrows). (b) Tn*5251* is compared with 10 pneumococcal Tn*5251*-like elements available in public databases. Seven out of 10 belong to strains whose genome is completely sequenced. Rectangles represent the elements, and vertical bars indicate the nucleotide changes with respect to the Tn*5251* sequence. Names of strains carrying the elements are reported on the right of the figure. Strain 05P294 carries Tn*2010* (GenBank AB426620) and strain Ar4 carries Tn*6003* (GenBank AM410044), whereas strain SpnA213 carries the composite Tn*5253*-like element (GenBank FM201786). To obtain a better alignment of the sequences, the elements were devoid of insertions of additional genetic material, indicated by triangles. A 5089-bp *erm*(B)-carrying Tn*917*-like element (grey triangle) inserts at nt 3514 in CGSP14; the 5511-bp *mef*(E)-carrying mega element (empty triangle) inserts at nt 3869 in Tn*2010*, TCH8431/19A and Taiwan19F-14; nt 14 194 is the insertion site of the 2849-bp *erm*(B) element (solid triangle) in Tn*2010*, TCH8431/19A, Hungary 19A-6, 670-6B and SpnA213 while in Tn*6003*, both the MAS element and the *erm*(B) element are inserted at the same site (empty square and black triangle, respectively) (Cochetti *et al.*, 2008). Nucleotide positions refer to the Tn*5251* sequence.

the integration/excision process (Rudy *et al.*, 1997); Int also acts as an accessory protein in the cleavage process that initiates the conjugal transfer mediated by the relaxase Orf20 (Rocco & Churchward, 2006). The *tet*(M) gene codes for a ribosomal protection protein responsible for tetracycline resistance. *orf7* and *orf9* code for a putative sigma factor and a putative transcriptional regulator, respectively, and possibly play a role in the gene expression regulation of Tn*5251* genes (Celli & Trieu-Cuot, 1998). *orf21* is an FtsK-SpoIIIE family protein that may play a role in segregating the replicated element between the donor and the recipient. *orf14*, coding for a putative cell wall hydrolase, may help degrade the peptidoglycan during conjugal mating. Orf16 is homologous to TcpF of *Clostridium perfringens* pCW3, which has been demonstrated to be essential for the plasmid conjugal transfer. This protein has an ATP-binding motif and is homologous to the VirB4 type IV secretion protein of Gram-negative bacteria that use energy from ATP hydrolysis to pump transposon DNA into the recipient cell (Rabel *et al.*, 2003). *orf18* codes for an antirestriction protein that protects the transposon DNA from restriction inhibiting host restriction enzymes (Serfiotis-Mitsa *et al.*, 2008).

Tn*5251* attB site

Tn*5251* is inserted into Tn*5253*, flanked by two different 6-bp-long coupling sequences (CS). Excision of Tn*5251* from its *attB* site produces a CI and a deletion in Tn*5253*. In the CI, single-stranded overhangs, produced by staggered cleavages of CS, are joined into a heteroduplex (Provvedi *et al.*, 1996). Deletion of Tn*5251* occurs between nts 16 293 and 34 339 of the Tn*5253* sequence (Iannelli *et al.*, unpublished data), leaving a heteroduplex formed by the two 6-bp CS. Chromosomal DNA purified from DP1322 was subjected to nested PCR analysis using PCR primers directed at the regions flanking Tn*5251*. Tn*5251* deletion was present at a level of 1.2 copies per 10^5 chromosomes. Sequence analysis of *attB* showed the presence of two bacterial populations, each harbouring one of the two CS, as a result of heteroduplex resolution following chromosomal replication.

Conjugal transfer of Tn*5251*

Tn*5253* was transferred by plate mating from DP1322 to our *S. pneumoniae* recipient FP10 and the resulting strain FR22 was used as a Tn*5251* donor (Table 1). Until now, Tn*5251* conjugal transfer was described only in association with the

whole Tn*5253*, whereas, here, we first report the autonomous transfer of Tn*5251*. Transfer of Tn*5251* as an independent CTn was only obtained in *S. pneumoniae* and *E. faecalis* (Table 3). In *S. pneumoniae*, transfer occurred in a strain-dependent manner; in fact, it was possible to move

Tn*5251* into the TIGR4 derivative FP47, but not in the Rx1 derivative FP11. The representative transconjugant *E. faecalis* FR64 harbouring an autonomous copy of Tn*5251* was used as a donor to determine the Tn*5251* host range. For this purpose, *S. pneumoniae*, *Streptococcus gordonii*, *S.*

Table 2. Annotated ORFs of Tn*5251*

ORF (aa)*	Annotation and comments (reference)	Homologous protein			Pfam domains [*E* value]‡
		Protein ID Origin [*E* value]†	Amino acid identity	Amino acid similarity	
int (405)	DNA integrase (Connolly *et al.*, 1998)	YP_133692.1 Tn*916* *E. faecalis* [0]	404/405 (99%)	405/405 (100%)	
xis (67)	DNA excisionase (Abbani *et al.*, 2005)	YP_133691 Tn*916* *E. faecalis* [0]	67/67 (100%)		
orf7 (140)	Sigma factor, putative				Sigma70_r4_2 (78–131) [3.2e − 09]
orf9 (117)	Transcriptional regulator, putative (Brennan & Matthews, 1989)				HTH_3 (17–71) [5.5e − 14]
tet(M) (639)	Tetracycline resistance determinant (ribosomal protection protein) (Burdett, 1991)				
orf12 (28)	Tet(M) leader peptide (Su *et al.*, 1992)				
orf14 (333)	Cell wall hydrolase, putative				NLPC_P60 (224–330) [1.2e − 54]
orf16 (815)	Type IV secretion protein, VirB4, putative	ABF47331.1 pCW3 *C. perfringens* [2e − 68]	228/844 (27%)	422/844 (50%)	
ardA/orf18 (165)	Antirestriction protein, ArdA (Serfiotis-Mitsa *et al.*, 2008)				ArdA (5–165) [3.4e − 81]
orf20 (401)	Relaxase (Rocco & Churchward, 2006)				HTH_3 (15–53) [1.6e − 10] Rep_trans (168–392) [2.9e − 100]
orf21 (461)	FtsK-SpoIIIE family protein, putative				FtsK_SpoIIIE (184–366) [1.5e − 57]

*Number of amino acids is shown in parentheses.
†Method: compositional matrix adjust.
‡Position of Pfam domains is reported in parentheses.

Table 3. Conjugal transfer of Tn*5253* from donor strain *Streptococcus pneumoniae* FR22

Recipient strain	Transfer frequency*	Genetic analysis of transconjugants†				Representative transconjugant
		%	Marker	Tn content		
S. pneumoniae FP11	1.6×10^{-4}	100	Tc Cm	Tn*5253*		FR58
S. pneumoniae FP47	3.6×10^{-5}	96	Tc Cm	Tn*5253*		FR54
		4	Tc	Tn*5251*		FR73
S. gordonii GP204	2.5×10^{-5}	100	Tc Cm	Tn*5253*		FR43
S. pyogenes SF370	8.2×10^{-6}	100	Tc Cm	Tn*5253*		FR68
S. agalactiae H36B	3.8×10^{-6}	100	Tc Cm	Tn*5253*		FR67
E. faecalis OG1RF	5.4×10^{-6}	35	Tc Cm	Tn*5253*		FR66
		35	Tc	Tn*5251*		FR64
		30	Tc	Defective Tn*5253*		
E. faecalis JH2-2	8.6×10^{-7}	50	Tc Cm	Tn*5253*		FR50
		50	Tc	Tn*5251*		FR65

*Conjugation frequency is expressed as CFU of transconjugants per CFU of donor. The results are presented as the mean of at least three mating experiments.
†Transconjugants are selected for acquisition of tetracycline resistance. Transconjugants resistant to both tetracycline and chloramphenicol (Tc–Cm) contain the whole Tn*5253*, whereas transconjugants resistant to tetracycline (Tc) were found to contain either Tn*5251* or defective forms of Tn*5253* in which the *cat* gene was deleted.

pyogenes, *S. agalactiae*, *E. faecalis* and *B. subtilis* strains (Table 1) were the conjugation recipients. Tn*5251* was transferred from the enterococcal transconjugant to *S. gordonii*, *S. pyogenes*, *E. faecalis* and *B. subtilis*, but not in *S. pneumoniae* (Table 4). When representative transconjugants of different species were used as donors, Tn*5251* was moved into *S. pneumoniae* from *S. pneumoniae*, *S. gordonii* and *S. pyogenes* (Table 4).

Tn*5251*-like elements can be found both integrated alone into the chromosome or inserted into other genetic elements (Fig. 1); Tn*916* was originally found integrated into the conjugative plasmid pAD1 (Franke & Clewell, 1981), and other members of the family have been found to be associated with larger CTns or plasmids (Rice & Carias, 1995). Such a wide choice of insertion sites is one of the reasons for the success of this class of elements; in fact, they can either transfer autonomously or 'hitchhike' other elements, which may increase their host range. In terms of *S. pneumoniae*, it is likely that Tn*5251* may be dependent on a more efficient conjugative machinery such as the one of Tn*5253*, but this does not impair the independent conjugal transfer of Tn*5251*, which we were able to detect when the transfer of Tn*5253* occurred at low frequencies (Table 3).

Tn*5251* target sites

Using inverse PCR on *S. pneumoniae* transconjugants, we found 4 Tn*5251* integration sites in the pneumococcal chromosome (Fig. 1). Using the *S. pneumoniae* R6 genome (Hoskins *et al.*, 2001) as a reference, the insertions occurred in *spr0357* at nt 357 477, in intergenic regions at nts 96 766 and 120 345 and in two transconjugants, obtained from different matings, at nt 1 175 225. It is worth noting that the insertion of Tn*5251* into *spr0357* CDS occurs at the 3′ end of the gene, probably not impairing the functionality of the coded protein. Analysis of *E. faecalis* transconjugants showed that the Tn*5251* insertion occurred in intergenic regions at nts 625 078, 789 261, 825 176 and 1 830 021 of the OG1RF chromosome (Bourgogne *et al.*, 2008). Tn*5251* target sites

are formed by a T-rich region separated from an A-rich region by a 6-bp CS and have short fragments of homology with the ends of the transposon. This has also been noted for Tn*916* and Tn*1545* insertion sites (Trieu-Cuot *et al.*, 1993).

Genome-wide sequence analysis of both pneumococcal genomes and MGEs showed that there are 14 Tn*5251*-like CTns, seven of them being present in a composite CTn (Fig. 1). The seven Tn*5251*-like CTns integrate at four sites: the Tn*3872*-like elements present in strains CGSP14 (GenBank CP001033) and Hungary19A-6 (GenBank CP000936) integrate at nts 159 534 and 1 166 926, respectively, Tn*2009* (Del Grosso *et al.*, 2004) at nt 1 195 582, whereas Tn*3872* (Del Grosso *et al.*, 2006), Tn*2010* (GenBank AB426620) and the elements of strains Taiwan19F-14 (GenBank CP000921) and TCH8431/19A (GenBank, NZ_ACJP00000000) integrate at nt 1 731 928. Composite elements integrate at two different sites: Tn*5253* (Shoemaker *et al.*, 1979; Ayoubi *et al.*, 1991), Tn*5253*-like (GenBank FM201786) and the elements of strains 670-6B (http://strepneumo-sybil.igs.umaryland.edu/) and P1031 (GenBank CP000920) integrate at nt 1 036 330, whereas ICE*Sp*23FST81 (Croucher *et al.*, 2008) (GenBank FM211187), Tn*2008* of CGSP14 and the element of G54 (GenBank CP001015) integrate at nt 1 207 256. We reported the integration site positions referring to the R6 chromosome. It is worth noting that insertion of the Tn*5251*-like element within ICE*Sp*23FST81 and Tn*2008* occurs at the same site, while in Tn*5253*, Tn*5253*-like and in the composite elements of strains 670-6B and P1031, insertion occurs at four different sites within the larger transposon (data not shown). Our analysis of genetic elements' integration into the *S. pneumoniae* chromosome clearly showed that three sites are 'preferred' targets for the integration of these elements and can be regarded as insertional hotspots.

Conclusions

In this work, we showed that pneumococcal Tn*5251* belonging to the Tn*916*–Tn*1545* family of CTn is an 18 033-bp-

Table 4. Conjugal transfer of Tn*5251* in different bacterial species

Donor strain	Recipient strain	Frequency*	Representative transconjugant
E. faecalis FR64	*S. pneumoniae* FP10	$< 1.6 \times 10^{-8}$	–
	S. gordonii GP204	4.8×10^{-5}	FR70
	S. pyogenes SF370	9.1×10^{-7}	FR71
	E. faecalis OG1SS	1.3×10^{-6}	FR80
	B. subtilis 168	1.6×10^{-6}	FR72
S. pneumoniae FR73	*S. pneumoniae* FP22	2.6×10^{-8}	FR77
S. pneumoniae FR73	*S. pneumoniae* FP23	1.2×10^{-7}	FR78
S. gordonii FR70	*S. pneumoniae* FP11	3.1×10^{-7}	FR79
S. pyogenes FR71	*S. pneumoniae* FP11	3.3×10^{-5}	FR74

*Conjugation frequency is expressed as CFU of transconjugants per CFU of donor.
The results are presented as the mean of at least three mating experiments.

long element containing 22 ORFs. *In silico* annotation was obtained for 11 ORFs including the *tet*(M) for ribosomal protection protein conferring tetracycline resistance. Here, we first demonstrate that Tn*5251* excises from Tn*5253* and is capable of autonomous transfer in *S. pneumoniae* and *E. faecalis*. Autonomous copies of Tn*5251* can be independently moved into *S. pneumoniae*, *S. gordonii*, *S. pyogenes*, *S. agalactiae*, *E. faecalis* and *B. subtilis*. Analysis of Tn*5251* and Tn*5251*-like elements' insertion into *S. pneumoniae* and *E. faecalis* showed that besides the presence of hotspots, several integration sites are located in AT-rich regions of the bacterial chromosome and that generally element insertion does not disrupt coding sequences. Furthermore, Tn*5251*-like elements are highly capable of capturing other genetic elements carrying different antibiotic resistance determinants such as the *mef*(E) and *erm* genes conferring macrolide resistance, *aadE*, *sat4* and *aphA-3* conferring resistance to streptomycin, streptothricin and kanamycin, respectively. These features make these elements successful in disseminating multidrug resistance determinants among pathogenic bacterial species. In this context, characterization of Tn*5251* contributes to the understanding of the mechanisms of the spread of antibiotic resistance.

Acknowledgements

References

Abbani M, Iwahara M & Clubb RT (2005) The structure of the excisionase (Xis) protein from conjugative transposon Tn*916* provides insights into the regulation of heterobivalent tyrosine recombinases. *J Mol Biol* **347**: 11–25.

Ayoubi P, Kilic AO & Vijayakumar MN (1991) Tn*5253*, the pneumococcal W(*cat tet*) BM6001 element, is a composite structure of two conjugative transposons, Tn*5251* and Tn*5252*. *J Bacteriol* **173**: 1617–1622.

Bourgogne A, Garsin DA, Qin X *et al.* (2008) Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol* **9**: R110.1–R110.16.

Brennan RG & Matthews BW (1989) The helix-turn-helix DNA binding motif. *J Biol Chem* **4264**: 1903–1906.

Burdett V (1991) Purification and characterization of Tet(M), a protein that renders ribosomes resistant to tetracycline. *J Biol Chem* **266**: 2872–2877.

Celli J & Trieu-Cuot P (1998) Circularization of Tn*916* is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. *Mol Microbiol* **28**: 103–117.

Chen C, Tang J, Dong W *et al.* (2007) A glimpse of streptococcal toxic shock syndrome from comparative genomics of *S. suis* 2 Chinese isolates. *PLoS One* **21**: e315.

Churchward G (2002) Conjugative transposons and related mobile elements. *Mobile DNA II* (Craig NL, Craigie R, Gellert M & Lambowitz ML, eds), pp. 177–191. American Society for Microbiology Press, Washington, DC.

Clewell DB, Flannagan SE & Jaworski DD (1995) Unconstrained bacterial promiscuity: the Tn*916*–Tn*1545* family of conjugative transposons. *Trends Microbiol* **3**: 229–236.

Cochetti I, Tili E, Mingoia M, Varaldo PE & Montanari MP (2008) *erm*(B)-carrying elements in tetracycline-resistant pneumococci and correspondence between Tn*1545* and Tn*6003*. *Antimicrob Agents Ch* **52**: 1285–1290.

Connolly KM, Wojciak JM & Clubb RT (1998) Site-specific DNA binding using a variation of the double stranded RNA binding motif. *Nat Struct Biol* **5**: 546–550.

Courvalin P & Carlier C (1986) Transposable multiple antibiotic resistance in *Streptococcus pneumoniae*. *Mol Gen Genet* **205**: 291–297.

Croucher NJ, Walker D, Romero P *et al.* (2008) The role of conjugative elements in the evolution of the multi-drug resistant pandemic clone *Streptococcus pneumoniae* Spain23F ST81. *J Bacteriol* **191**: 1480–1489.

Del Grosso M, Scotto d'Abusco A, Iannelli F, Pozzi G & Pantosti A (2004) Tn*2009*, a Tn*916*-like element containing *mef*(E) in *Streptococcus pneumoniae*. *Antimicrob Agents Ch* **48**: 2037–2042.

Del Grosso M, Camilli R, Iannelli F, Pozzi G & Pantosti A (2006) The *mef*(E)-carrying genetic element (mega) of *Streptococcus pneumoniae*: insertion sites and association with other genetic elements. *Antimicrob Agents Ch* **50**: 3361–3366.

Ding F, Tang P, Hsu MH, Cui P, Hu S, Yu J & Chiu CH (2009) Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. *BMC Genomics* **10**: 158.

Dunny GM, Craig RA, Carron RL & Clewell DB (1979) Plasmid transfer in *Streptococcus faecalis*: production of multiple sex pheromones by recipients. *Plasmid* **2**: 454–465.

Ferretti JJ, McShan WM, Ajdic D *et al.* (2001) Complete genomic sequence of an M1 strain of *Streptococcus pyogenes*. *P Natl Acad Sci USA* **98**: 4658–4663.

Flannagan SE, Zitzow LA, Su YA & Clewell DB (1994) Nucleotide sequence of the 18-kb conjugative transposon Tn*916* from *Enterococcus faecalis*. *Plasmid* **32**: 350–354.

Franke AE & Clewell DB (1981) Evidence for a chromosome-borne resistance transposon (Tn*916*) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the absence of a conjugative plasmid. *J Bacteriol* **145**: 494–502.

Frost LS, Leplae R, Summers AO & Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**: 722–732.

Guild WR & Shoemaker NB (1976) Mismatch correction in pneumococcal transformation: donor length and *hex*-dependent marker efficiency. *J Bacteriol* **125**: 125–135.

Henderson-Begg SK, Roberts AP & Hall LM (2008) Diversity of putative Tn*5253*-like elements in *Streptococcus pneumoniae*. *Int J Antimicrob Ag* **33**: 364–367.

Hoskins J, Alborn WE, Arnold J *et al.* (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* **183**: 5709–5717.

Iannelli F & Pozzi G (2007) Protocol for conjugal transfer of genetic elements in *Streptococcus pneumoniae*. *Molecular Biology of Streptococci* (Hakenbeck R & Chhatwal GS, eds), pp. 525–529. Horizon Bioscience, Norfolk, UK.

Iannelli F, Giunti L & Pozzi G (1998) Direct sequencing of long PCR fragments. *Mol Biotechnol* **10**: 183–185.

Iannelli F, Oggioni MR & Pozzi G (2005) Sensor domain of histidine kinase ComD confers competence pherotype specificity in *Streptococcus pneumoniae*. *FEMS Microbiol Lett* **252**: 321–326.

Jacob AE & Hobbs SJ (1974) Conjugal transfer of plasmid-borne multiple antibiotic resistant in *Streptococcus faecalis* var. *zymogenes*. *J Bacteriol* **117**: 360–372.

Lanie JA, Ng WL, Kazmierczak KM, Andrzejewski TM, Davidsen TM, Wayne KJ, Tettelin H, Glass JI & Winkler ME (2007) Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. *J Bacteriol* **189**: 38–51.

Manganelli R, Romano L, Ricci S, Zazzi M & Pozzi G (1995) Dosage of Tn*916* circular intermediates in *Enterococcus faecalis*. *Plasmid* **34**: 48–57.

Mullany P, Roberts AP & Wang A (2002) Mechanism of integration and excision in conjugative transposons. *Cell Mol Life Sci* **59**: 2017–2022.

Munoz R, Bustamante M & De La Campa AG (1995) Ser-127-to-Leu substitution in the DNA Gyrase B subunit of *Streptococcus pneumoniae* is implicated in novobiocin resistance. *J Bacteriol* **177**: 4166–4170.

Oggioni MR, Dowson CG, Maynard Smith J, Provvedi R & Pozzi G (1996) The tetracycline resistance gene *tet*(M) exhibits mosaic structure. *Plasmid* **35**: 156–163.

Oggioni MR, Memmi G, Maggi T, Chiavolini D, Iannelli F & Pozzi G (2003) Pneumococcal zinc metalloproteinase ZmpC cleaves human matrix metalloproteinase 9 and is a virulence factor in experimental pneumonia. *Mol Microbiol* **49**: 795–805.

Pearce BJ, Iannelli F & Pozzi G (2002) Construction of new unencapsulated (rough) strains of *Streptococcus pneumoniae*. *Res Microbiol* **1**: 2–10.

Pozzi G, Stellini M, Marri L & Molina AM (1986) Transformation as a tool for studying the epidemiology of *tet* determinants in *Streptococcus pneumoniae*. *Eur J Epidemiol* **2**: 90–94.

Pozzi G, Musmanno RA, Renzoni EA, Oggioni MR & Cusi MG (1988) Host-vector system for integration of recombinant DNA into chromosomes of transformable and nontransformable streptococci. *J Bacteriol* **170**: 1969–1972.

Provvedi R, Manganelli R & Pozzi G (1996) Characterization of conjugative transposon Tn*5251* of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* **135**: 231–236.

Rabel C, Grahn AM, Lurz R & Lanka E (2003) The VirB4 family of proposed traffic nucleoside triphosphatases: common motifs in plasmid RP4 TrbE are essential for conjugation and phage adsorption. *J Bacteriol* **185**: 1045–1058.

Rice LB (1998) Tn*916* family of conjugative transposons and determination of antimicrobial resistance determinats. *Antimicrob Agents Ch* **42**: 1871–1877.

Rice LB (2002) Association of different mobile elements to generate novel integrative elements. *Cell Mol Life Sci* **59**: 2023–2032.

Rice LB & Carias LL (1995) Transfer of Tn*5385*, a composite, multiresistance chromosomal element from *Enterococcus faecalis*. *J Bacteriol* **180**: 714–721.

Roberts AP & Mullany P (2009) A modular master on the move: the Tn*916* family of mobile genetic elements. *Trends Microbiol* **17**: 251–258.

Rocco JM & Churchward G (2006) The integrase of the conjugative transposon Tn*916* directs strand- and sequence-specific cleavage of the origin of conjugal transfer, *oriT*, by the endonuclease Orf20. *J Bacteriol* **188**: 2207–2213.

Rudy C, Taylor KL, Hinerfeld D, Scott JR & Churchwald G (1997) Excision of a conjugative transposon *in vitro* by the Int and Xis proteins of Tn*916*. *Nucleic Acids Res* **25**: 4061–4066.

Salles C, Creancier L, Claverys JP & Mejean V (1992) The high level streptomycin resistance gene from *Streptococcus pneumoniae* is a homologue of the ribosomal protein S12 gene from *Escherichia coli*. *Nucleic Acids Res* **20**: 6103.

Serfiotis-Mitsa D, Roberts GA, Cooper LP, White JH, Nutley M, Cooper A, Blakely GW & Dryden DT (2008) The Orf18 gene product from conjugative transposon Tn*916* is an ArdA antirestriction protein that inhibits type I DNA restriction-modification systems. *J Mol Biol* **383**: 970–981.

Shoemaker NB & Guild WR (1974) Destruction of low efficiency markers is a slow process occurring at a heteroduplex stage of transformation. *Mol Gen Genet* **128**: 283–290.

Shoemaker NB, Smith MD & Guild WR (1979) Organization and transfer of heterologous chloramphenicol and tetracycline resistance genes in pneumococcus. *J Bacteriol* **139**: 432–441.

Smith MD, Hazum S & Guild WR (1981) Homology among tet determinants in conjugative elements of streptococci. *J Bacteriol* **148**: 232–240.

Spizizen J (1958) Transformation of biochemically deficient strains of bacillus subtilis by deoxyribonucleate. *P Natl Acad Sci USA* **44**: 1072–1078.

Su YA, He P & Clewell DB (1992) Characterization of the *tet*(M) determinant of Tn*916*: evidence for regulation by transcription attenuation. *Antimicrob Agents Ch* **36**: 769–778.

Tettelin H, Nelson KE, Paulsen IT *et al.* (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**: 498–506.

Tettelin H, Masignani V, Cieslewicz MJ *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *P Natl Acad Sci USA* **102**: 13950–13955.

Trieu-Cuot P, Poyart-Salmeron C, Carlier C & Courvalin P (1993) Sequence requirements for target activity in site-specific recombination mediated by the Int protein of transposon Tn*1545*. *Mol Microbiol* **8**: 179–185.