Analysis of tetracycline resistance \( \text{tet}(W) \) genes and their flanking sequences in intestinal \textit{Bifidobacterium} species

Mohammed Salim Ammor†, Ana Belén Flórez, Pablo Álvarez-Martín, Abelardo Margolles and Baltasar Mayo*

Departamento de Microbiología y Bioquímica, Instituto de Productos Lácteos de Asturias (CSIC), Carretera de Infiesto s/n, 33300 Villaviciosa, Asturias, Spain

Received 18 April 2008; returned 12 May 2008; revised 29 May 2008; accepted 11 June 2008

Objectives: The \( \text{tet}(W) \) gene provides tetracycline resistance to a wide range of anaerobic intestinal and ruminal bacteria, but little is known about the molecular organization of the \( \text{tet}(W) \) gene. The aim of this study was to gain new insights into the molecular organization of the \( \text{tet}(W) \) gene in bifidobacteria strains from humans.

Methods: A segment of DNA encompassing the whole \( \text{tet}(W) \) gene and its immediate upstream and downstream sequences was analysed in 10 representative strains of four \textit{Bifidobacterium} species, of which two have been shown to be tetracycline-susceptible. The non-conserved flanking regions of the \( \text{tet}(W) \) gene were further analysed in six strains.

Results: All 10 strains share a core DNA domain of 2154 bp [starting 250 bp upstream of the \( \text{tet}(W) \) gene start codon and ending 13 bp before the stop codon] with 98% to 100% DNA identity. Except for \textit{Bifidobacterium animalis} E43, all other strains further share 408 bp upstream and 70 bp downstream of the \( \text{tet}(W) \) gene. An insertion-like element of 736 bp was found to interrupt the \( \text{tet}(W) \) coding sequence in \textit{Bifidobacterium longum} M21, which may be the reason for its tetracycline susceptibility. However, genetic events explaining the susceptible phenotype of \textit{B. longum} LMG 13197 were not observed.

Conclusions: The \( \text{tet}(W) \) genes from all 10 strains shared 98% to 100% DNA and amino acid identity, though large variation was found in their flanking regions.

Keywords: bifidobacteria, antibiotic resistance, probiotics

Introduction

The commensal intestinal microbiota of humans and animals may act as a reservoir of antibiotic resistance (AR) genes that could ultimately be transferred to pathogens. In fact, the transfer of AR genes has been shown to occur between bacterial species in the gastrointestinal tract (GIT) of mammals. Bifidobacteria are among the dominant bacterial populations of the human GIT, where they are thought to exert several health-promoting effects. Accordingly, species and strains of this genus are frequently used as probiotics in dairy products or dietary supplements. It is recommended that any strain used in food systems should be free of potentially transferable AR traits.

Of the 38 tetracycline resistance (Tc\(^{R}\)) genes already described ([http://faculty.washington.edu/marilynr/tetweb1.pdf](http://faculty.washington.edu/marilynr/tetweb1.pdf)), \( \text{tet}(W) \), which encodes a ribosomal protection protein, is the most widespread Tc\(^{R}\) gene in anaerobic gut and rumen bacteria. Indeed, it has recently been reported in \textit{Bifidobacterium} species. However, large variations in the MIC of tetracycline have repeatedly been reported in strains harbouring a \( \text{tet}(W) \) gene. Also intriguing is the presence of susceptible strains in which the gene has been detected by either PCR or hybridization. For all these reasons, the molecular analysis of the \( \text{tet}(W) \) genes in bifidobacteria has become the subject of much interest.

In a recent survey, we showed that chromosomally encoded \( \text{tet}(W) \) genes are present in tetracycline-resistant and tetracycline-susceptible species of bifidobacteria isolated from the GIT of healthy humans not subjected to recent antibiotic treatment. In this paper, we report on the molecular characterization of the \( \text{tet}(W) \) gene and its flanking regions in 10 representative strains belonging to four bifidobacterial species—\textit{Bifidobacterium}
Molecular analysis of tet(W) genes from bifidobacteria

animalis, Bifidobacterium bifidum, Bifidobacterium longum and Bifidobacterium thermophilum—of which 8 were tetracycline-resistant and 2 were tetracycline-susceptible. The aim was to investigate the molecular organization of the gene in both tetracycline-resistant and -susceptible strains, in order to supply the basis for a correlation between genotype and phenotype, and obtain information on the mechanism of this gene’s spread. The analysis included non-shared regions upstream and downstream of tet(W) genes in six strains.

Materials and methods

Bacterial strains, media and culture conditions

Except for B. thermophilum LMG 21813T and B. longum LMG 13197T, which were kindly provided by the Belgian Coordinated Collections of Microorganisms (BCCM™, University of Ghent, Ghent, Belgium), all strains analysed in this work were isolated as part of a project devoted to the identification and typing of dominant intestinal microbial populations in healthy Spaniards.6 With the exception of the tetracycline-susceptible (Tc') strains B. longum M21 and B. longum LMG 13197T, all were considered to be Tc' strains. MICs of tetracycline for Tc' strains varied from 4 to 64 mg/L as determined by the Etest method (Table 1), when compared with an MIC ranging from 0.25 to 1 mg/L for Bifidobacterium thermophilum

B. animalis, B. bifidum, B. longum, and B. thermophilum—of which 8 were tetracycline-resistant and 2 were tetracycline-susceptible. The aim was to investigate the molecular organization of the gene in both tet(W) genes encoding ribosomal protection proteins. The amplification and sequencing of the fragments were performed with the specific primers, tetW and tetT, for the tet(W) gene.7 Genome walking was performed from these internal regions of the gene to obtain the whole gene, and flanking upstream and downstream sequences were obtained by direct and inverse PCR strategies using a combination of primers (Table 2).

**DNA techniques**

Total genomic DNA was isolated from overnight cultures using the Wizard Genomic DNA Purification Kit (Promega Corporation, Southampton, UK). Plasmid DNA was isolated following the procedure of O’Sullivan and Klaenhammer.8 The general procedures for DNA manipulation, electrophoresis and visualization were essentially as described by Sambrook and Russell.5 Restriction endonucleases and Taq DNA polymerase from Takara (Otsu, Shiga, Japan) and T4 DNA ligase from Invitrogen (Carlsbad, CA, USA) were used according to the manufacturers’ instructions.

**Sequencing of tet(W) genes and flanking sequences**

Internal segments of the tet(W) genes from all strains had already been amplified as reported previously,6 using the universal primers based on Tc' genes encoding ribosomal protection proteins. The amplification and sequencing of the fragments were performed with the specific primers, tetW and tetT, for the tet(W) gene.7 Genome walking was performed from these internal regions of the gene to obtain the whole gene, and flanking upstream and downstream sequences were obtained by direct and inverse PCR strategies using a combination of primers (Table 2). For inverse PCR, total genomic DNA was digested with NcoI, ClaI or PstI and self-ligated overnight. The following morning, the ligated DNA was precipitated, centrifuged, dried and suspended in 100 μL of TE prior to use as a template in PCR. Appropriate pairs of primers (Table 2) were used to amplify the upstream and downstream region of the gene. Amplicons were purified using the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare Biosciences, Buckinghamshire, UK) and sequenced by cycle extension in an ABI 370 DNA sequencer (Applied Biosystems, Foster City, CA, USA). Finally, sequences were assembled and aligned using the DNAMAN program (Lynnon Biosoft, Pointe-Claire, Quebec, Canada), and the nucleotide and deduced protein sequences were compared with those in the GenBank database using the BLAST program (National Center for Biotechnology Information of US National Library of Medicine; available at http://www.ncbi.nlm.nih.gov/blast/Blast.cgi).

**Copy number of tet(W) genes**

The copy number of the tet(W) gene was determined using real-time PCR. Genomic DNAs were amplified with the tet(W) gene’s internal primers tetW-1 and tetW-2 and with the single-copy housekeeping tuf gene’s internal primers Bif-tuf-1 and Bif-tuf-2 (Table 2). PCR was performed in an ABI Prism 7500 Fast Real-Time PCR System (Applied Biosystems). SYBR Green I fluorophore was used to correlate the amount of PCR product with the fluorescent signal. Amplification was performed in a 25 μL final volume containing 50 ng of chromosomal DNA as a template, 10 nM of each primer and 12.5 μL of SYBR Green PCR Master Mix (Applied Biosystems). The amplification reaction was performed as follows: 2 min at 50°C, 10 min at 95°C for activating the AmpliTaq Gold DNA polymerase and forty cycles of: (i) denaturing at 95°C for 15 s; (ii) annealing/extension at 60°C for 1 min; and (iii) a dissociation step. The cycle number during which the fluorescence signal crossed the chosen cycle threshold (Ct; in this case, the default was set by the manufacturer) was recorded. All samples were processed in duplicate using independent DNA samples. Negative controls involving all the elements of the reaction mixture except the template DNA were also included.

**Nucleotide accession numbers**

The nucleotide sequences described in this paper were deposited in the GenBank database with the following accession numbers: B. animalis E43, EU434754; B. bifidum L22, EU434755; B. longum
B93, EU434749; B. longum B94, EU434750; B. longum E111, EU434751; B. longum H66, DQ060146; B. longum L42, EU434756; B. longum M21, EU434757; B. thermophilum LMG 21813T, EU434753.

Results

Sequence conservation of the tet(W) gene, its upstream region and the resulting Tet(W) protein

The nucleotide sequence of the tet(W) gene and its immediate upstream region was determined in all 10 strains. All sequences shared a core DNA region of 2638 bp, except for B. animalis E43, which shared a sequence of only 2154 bp. This core region started 658 bp upstream of the tet(W) gene start codon and ended 70 bp after its stop codon (250 bp upstream of the start codon and 13 bp before the stop codon, respectively, in B. animalis E43). A genetic diagram of these regions is shown at the bottom of Figure 1. The common 2.1 kb segment showed a G+C content of 51.3% to 51.4%, which is lower than the G+C average for Bifidobacterium genomes (around 59%).

The nucleotide identity of the sequences to one another ranged from 98.2% to 100%, the sequence of strain E43 being the least conserved.

Putative conserved regulatory regions were observed in the first 250 bp upstream of the gene, including a short 14 amino

Table 2. Primers used in this study

<table>
<thead>
<tr>
<th>Primer (forward/reverse)</th>
<th>Sequence (5'–3')</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DI(F)</td>
<td>GAYACICCGGGICAYRTGAYTT</td>
<td>universal ribosomal protection-type Te c genes</td>
</tr>
<tr>
<td>DI(R)</td>
<td>CCARWAIAGGRTTGGGGGIGACYTG</td>
<td>universal ribosomal protection-type Te c genes</td>
</tr>
<tr>
<td>tet1(F)</td>
<td>GCTCAAGGTTGAGGACAGGA</td>
<td>universal ribosomal protection-type Te c genes</td>
</tr>
<tr>
<td>tet2(R)</td>
<td>AGGATTTGCGCGGTAACCTGKA</td>
<td>universal ribosomal protection-type Te c genes</td>
</tr>
<tr>
<td>tetW(F)</td>
<td>AAGCGGCACTCATTCTTCC</td>
<td>tet(W) gene</td>
</tr>
<tr>
<td>U tetW(F)</td>
<td>AATGCTCATGTACGGTAAG</td>
<td>shared sequence of the tet(W) loci</td>
</tr>
<tr>
<td>D tetW(R)</td>
<td>TTACATTACCTTTGGAAC</td>
<td>shared sequence of the tet(W) loci</td>
</tr>
<tr>
<td>tetW-rev1(R)</td>
<td>GCITATCCTTGGGTCTTTGG</td>
<td>upstream of tet(W)</td>
</tr>
<tr>
<td>tetW-rev2(F)</td>
<td>GACAAGGTCGCCCATATTG</td>
<td>tet(W) gene</td>
</tr>
<tr>
<td>tetW-1(F)</td>
<td>GATCGACCAGCCTGGCGTG</td>
<td>tet(W)</td>
</tr>
<tr>
<td>tetW-2(R)</td>
<td>GGCTGATTGGTCTCTGC</td>
<td>tet(W)</td>
</tr>
<tr>
<td>Bif-tuf-1(F)a</td>
<td>GACAAGCGGCTCTGGATGCC</td>
<td>B. longum tuf gene</td>
</tr>
<tr>
<td>Bif-tuf-2(R)a</td>
<td>AGCGAGCCGGGTTGTCG</td>
<td>B. longum tuf gene</td>
</tr>
</tbody>
</table>

aBased on the complete genome sequence of B. longum NCC2705 (accession no. AE014295).

Figure 1. Genetic organization of the regions upstream and downstream of tet(W) in six Bifidobacterium strains. ORFs with the same shading (note transposase- and integrase-related ORFs in light grey) or grading represent similar or identical genes in the different species and strains. Arrows indicate complete ORFs, while broken ends indicate N- or C-terminal interrupted genes. The positions and orientations of the primers used in this work (Table 2) are indicated by the small black arrows below the core region.
Molecular analysis of tet(W) genes from bifidobacteria

The tet(W) gene, encoding a 639-amino-acid protein, was found to be present in all strains of *B. longum, B. animalis* E43, and *B. bifidum* M21. The deduced amino acid sequences of the tet(W) gene from these strains showed 99% identity with each other. However, the DNA sequence upstream of the tet(W) gene was not conserved, and there were differences in the upstream regions of the gene among the strains. In particular, the sequence upstream of the tet(W) gene from strain E43 showed a single nucleotide change in the region identified as the core region.

Conserved flanking sequences

A combination of inverse and direct PCR and sequencing was used to obtain sequences from the upstream and downstream regions of the conserved domain in six strains. Distinct sequence stretches were obtained for the different strains under study (Figure 1). These diverged abruptly 658 bp upstream of the tet(W) gene start codon: a segment of DNA considered within the core region in all strains except *B. animalis* E43. However, the sequences were more conserved for some strains in the downstream region of the gene. In addition to the 60 bp downstream of the gene’s stop codon (conserved in all strains except *B. animalis* E43), a further 106 bp segment was shared by *B. longum* H66, L42, M21 and *B. thermophilum* LMG 21813. Sequences from *B. longum* H66 and *B. thermophilum* were identical for a 142 nucleotide long segment. Finally, *B. longum* L42 and M21 showed complete identity for all the downstream sequences available for both strains (545 bp). These common regions are indicated by dotted vertical lines in Figure 1.

Analysis of ORFs in regions flanking the tet(W) genes

In *B. animalis* E43, the tet(W) gene was preceded by an ORF encoding a protein with a 55% amino acid identity (for a stretch of 198 bp) to an internal sequence (amino acids 127–318) of a transposase of the insertion sequence (IS) 5 family of *Pseudomonas aeruginosa* 2192 (EAZ58 611.1; COG3039). In contrast, no ORFs were identified in a downstream 730 bp sequence. However, a small segment of DNA of around 100 bp, which included the last nucleotides of the tet(W) gene (different from those of all other strains) and its stop codon, showed significant nucleotide identity to the equivalent part of the tet(W) gene from *Lawsonia intracellularis* PHE/M1-00 genome (NC_008011; locus tag LI0179).

At the 5' end of the DNA segment sequenced from *B. bifidum* L22, an ORF was identified encoding a protein with 76% amino acid identity to the final sequence (135 amino acids) of the Clp protease from *B. longum* DJ101A (ZP_00121578.1; COG0542). This ORF ended 300 bp before the starting point of the common core sequence. No ORFs were identified in a small 180 bp stretch downstream of the tet(W) gene stop codon.

The DNA sequence upstream of the core region was not obtained for *B. thermophilum* LMG 21813, but a complete ORF was found downstream, which could encode a protein of 255 amino acids showing 100% identity to several SAM-dependent methyltransferases [*e.g.* that of orfY (FAEPRAM212_01960) from *Faecalibacterium prausnitzii* M21/2]. This same ORF was truncated at early but different positions in *B. longum* H66, M21 and L42 (Figure 1).

The N-terminal part of two truncated ORFs running in opposite directions was seen in the upstream region of *B. longum* H66. The first encoded a peptide of 124 amino acids with 47% amino acid identity to hypothetical proteins with an ARCH_ATPase domain (a conserved P-loop motif that is involved in ATP binding) from *Clostridium leptum* DSM 753 (EDO59757.1; CLOLEP_03807). The second ORF encoded a 257 amino acid protein with 95% identity to the N-terminus sequence of a putative ABC-type multidrug transporter from *B. longum* DJ101A (ZP_0012132.1; COG1132). Downstream of the core region, an interrupted ORF was observed, whose initial 80 amino acids corresponded to an OrfY protein (methyltransferase) with a sequence identical to that of *B. thermophilum* LMG 21813. After this position, the final 240 bp of H66 showed 99% identity at the nucleotide level to a segment of the *B. longum* NCC2705 chromosome (AE014295.3; positions 1510367–1510602).

A long ORF (encoding 424 amino acids) starting before the sequence analysed was found in the remote 5' end of the tet(W) gene in *B. longum* M21. The N-terminal part of the protein encoded by this ORF (289 amino acids) showed 91% identity to an internal segment of a hypothetical protein of the TraG/VirD4 family from *Bacteroides capillosus* (ZP_02039232.1; BACCATP_04884). Homologous proteins, which are thought to be involved in type IV secretion (a process necessary for bacterial conjugation), have also been found in many other anaerobic bacteria (*e.g.* *Fusobacterium prausnitzii*, *Clostridium kluyveri*, *Ruminococcus gravis* and *Eubacterium dolichum*).
Downstream within the same coding region but in a different frame, an ORF encoding a 46 amino acid protein was identified. This small protein, designated MAFF for its first four amino acids, has been previously reported associated to \( \text{tet}(W) \) genes in *Bifidobacterium adolescentis* L230 and JK51 and *Roseburia* sp. A2-183. The MAFF protein of M21 showed significant identity (80% to 82%) to the MAFF proteins of other strains, but its closest relative was a hypothetical protein from *B. capillosus* ATTC 29799 (EDM97320.1; BACCAP_04885). Downstream of this ORF, an ORF in the direction opposite to the \( \text{Te}c \) gene encoding the final 282 amino acids of a putative transposase (87% identity to a hypothetical protein from *Bifidobacterium adolescentis* L2-32; ZP_02029458.1; BIFADO_01916) of the mutator family (Pfam 00872) was found to occupy all the sequence available.

In strain L42, a truncated ORF was found encoding an internal peptide 268 amino acids long, with high homology (89% identity) to a phage-related transposase of the IS7 family (Pfam 002022 and 00552) from *B. adolescentis* ATCC 15703 (BAF40060.1). As mentioned above, the entire downstream sequence of strain M21 was identical to that of strain L42, although in the latter the ORF was transposed and encoded a polypeptide of 393 amino acids with 84% homology to the hypothetical transposase sequence in *B. adolescentis* L2.32 (BIFADO_01916).

**An insertion sequence-like element inserted into the tet(W) ORF in B. longum M21**

As mentioned previously, the Tet(W)-encoding ORF in *B. longum* M21 was disrupted by a 736 bp sequence with a genetic organization resembling that of an IS (Figure 1). The G + C content of this IS was 62.4%, similar to that of the *B. longum* genome (60%).\(^{10}\) The DNA segment was flanked by two almost perfect IRs of 30 nt (CCGATGTCTAATTAGTGAACCA CCCTT-674 nt-ACGGGTGTTGCACTTTCATTTAGACAACGG). A putative ORF was observed starting 78 bp from the 3' end of the sequence and running in a relative orientation opposite to that of the \( \text{tet}(W) \) gene. Its deduced product showed a tra8 domain (motif Pfam00665, thought to be involved in integration of DNA into the chromosome) and homology to transposases of the IS30 family (54% identity over a 50 amino acid stretch to the product of locus AAK94953 from *Mycoplasma bovis*). A further downstream out-of-frame peptide showed higher homology (68% over an 82 amino acid stretch) to remnant inactivated IS30 derivatives present in the genome of *B. longum* DJO10A (ZP_00121392.1, COG2826). Apparently, the coding sequence has a point mutation around the middle of the coding region, and the transposase might be non-functional. The sequence of the IS-like element is flanked by two short direct repeats of 5 bp (GTGTA). Thus, it appears that the IS interrupted the \( \text{tet}(W) \) gene and, during the insertion event, duplication of the 5 bp sequence occurred. Indeed, the 5 bp sequence is present once at this same position in all \( \text{tet}(W) \) genes analysed so far.

**\( \text{tet}(W) \) gene copy number**

The single-copy housekeeping *tuf* gene was used as a reference. After normalization, the \( \Delta C_T \) values for all strains (which ranged between 0.01 and –0.56) indicated that the \( \text{tet}(W) \) gene was in single-copy state in the chromosome of all strains.

**Discussion**

Among bifidobacteria, the \( \text{tet}(W) \) gene has been reported in many species, including *B. animalis*, *B. bifidum*, *Bifidobacterium catenulatum*, *B. longum*, *Bifidobacterium pseudocatenulatum*, *Bifidobacterium pseudolongum* and *B. thermophilum*.*5–7* \( \text{tet}(W) \) genes have been also reported in tetracycline-susceptible strains.*3,6* In this paper, the genetic organization of \( \text{tet}(W) \) genes and their immediate upstream and downstream sequences was analysed in 10 strains. \( \text{tet}(W) \) had already been determined to be the only tetracycline-resistance-encoding gene in these strains.*11* The nucleotide and deduced protein sequences of the \( \text{tet}(W) \) genes were all very similar (identity higher than 98%) and showed a close relationship to \( \text{tet}(W) \) genes from other intestinal bacteria.*7,12–14* The differences, however, could be used to trace the phylogenetic relationships of \( \text{tet}(W) \) genes from GIT inhabitants.*7,12–15* These relationships further support early suggestions of recent and independent transmission events.*6,7*

In the susceptible *B. longum* M21, an IS-like sequence of 736 bp was found to interrupt the structural \( \text{tet}(W) \) gene, giving rise to a non-functional protein. This IS element belongs to the IS30 family and has not been described before in bifidobacteria. Recently, an IS 1047 bp long was identified in the upstream region of the tetracycline-resistant *B. longum* F8.*13* Five of the six bases duplicated in F8 (CAATGC) seem to mirror the 5 bp duplication in M21, which might suggest active insertion sites in *B. longum*. Further investigations will be required to unravel the non-functionality of the \( \text{tet}(W) \) gene in *B. longum* LMG 13197T.

The \( \text{tet}(W) \) genes from all 10 strains shared 98% to 100% DNA and amino acid identity, though variation was found in the upstream and downstream flanking regions.

**Acknowledgements**

BCCM™, University of Ghent, Belgium, is acknowledged for providing the *Bifidobacterium* type strains.

**Funding**

This work was supported by project ACE-ART within the VI Frame Program (CT-2003-506214) and project AGL2007-61869/ALI of the Spanish Ministry of Education and Science. M. S. A. was awarded a postdoctoral fellowship from the ‘Secretaría de Estado de Universidades e Investigación’ of the Spanish Ministry of Education and Science (SB2004-0165).

**Transparency declarations**

None to declare.

M. S. A. conducted research and wrote the article. A. B. F. and P. A.-M. characterized the strains and conducted research. A. M. supported the work and contributed to the discussion. B. M. directed the research and contributed to the writing and discussion of the results.
Molecular analysis of tet(W) genes from bifidobacteria

References