An unusual integron in *Treponema denticola*

The integron–gene cassette system confers on a bacterial cell the potential to accumulate diverse genes at a common locus. Integrons associated with plasmids or transposons have driven the evolution of multiple-antibiotic resistance in many Gram-negative pathogens due to their ability to capture, shuffle, express and disseminate antibiotic resistance genes. Recent observations indicate that integrons are frequently also associated with chromosomes in bacteria (Rowe-Magnus *et al*., 2001) and that the cassette pool available to these integrons is enormous (Holmes *et al*., 2003a). This raises questions as to the broader significance of integrons in bacterial evolution.

The *Treponema denticola* ATCC 35405T genome sequence contains a 65 kb region containing many ORFs hypothesized to have been acquired by lateral transfer (Seshadri *et al*., 2004). We have identified an unusual integron (termed InTde35405) covering 58 kb of this region (GenBank/EMBL/DDBJ accession no. NC_002967; 1817049–1874294). InTde35405 is the first example of an integron with a gene cassette array oriented in the same direction as the integrase gene, and we believe it to be the first example of a complete, intrinsically chromosomal integron outside the *Proteobacteria*.

---

**Fig. 1.** Structure of *T. denticola* ATCC 35405T integron. The *intI* gene and cassette ORFs (block arrows, sizes not to scale) are numbered according to GenBank/EMBL/DDBJ accession no. NC_002967. White ORFs have no known function, light grey ORFs are related to conserved hypothetical proteins and dark grey ORFs are related to proteins of known function (BLAST E values < 0.00001). Circles indicate putative 59-bp recombination sites. Underlined cassettes (A–G) are duplicates, defined as containing ORFs with > 95% amino acid identity. A few very short predicted ORFs (1830, 1824, 1819 and 1782) from the original annotation were omitted from the proposed cassette array.
The key functional components of an integron are a site-specific recombinase of the IntI family, its cognate recombination site (termed attI) and promoters for the expression of intI (P_int) and captured genes (P_C). Collectively, these give an integron the potential to accumulate a gene cassette array and express the cassette-encoded genes (Hall & Collis, 1995). The Treponema gene Tde1844 has been previously identified as an intI homologue (Nield et al., 2001), with the closest relatives being integron integrases from Pseudomonas strains (Holmes et al., 2003b; Vaisvila et al., 2001) at 47–49% amino acid identity. The region between Tde1844 and Tde1843 includes a plausible attI/59-be (59-belement) junction (GATT at two possible P_C promoters: accession no. NC_002967) and 1873129 in GenBank/EMBL/DDBJ (http://mic.sgmjournals.org). The defining feature of a gene cassette is a recombination site (59-be or attC) consisting of an imperfect inverted repeat containing integrase binding sites (Stokes et al., 1997). The Tde 59-bes conform to this general model (Fig. 2). In the majority (38/45) of the T. denticola 59-bes, the 1L and 1R ends of the inverted repeats (Fig. 2) match more closely in the predicted circular cassettes compared to the linear integrated cassettes. This feature is characteristic of IntI-assembled gene cassette arrays (Recchia & Hall, 1995), suggesting that InTde35405 is a functional integron. Integrons associated with chromosomes frequently have distinctive gene cassette arrays that share very similar non-protein coding sequence compared to gene cassettes from other sources. While the latter two variables are to some extent sensitive to methods of ORF-prediction, the number of cassettes and the number of cassettes predicted to encode two or three proteins is much higher in the InTde35405 array. The InTde35405 cassettes contain on average almost double the amount of non-protein coding sequence compared to gene cassettes from other sources. We have found 45 gene cassettes (Tde1843 to Tde1773) associated with the Tde1844 integrase (Fig. 1). Unlike all previously identified integrons, the cassette array of InTde35405 is oriented in the same direction as the integrase. The region between Tde1844 and Tde1843 is much higher in the InTde35405 array. The T. denticola follows this pattern, with 40 of the 45 gene cassettes containing 59-bes closely related to the example in Fig. 2. This group represents a possible ‘Treponema denticola repeat’ family (Rowe-Magnus et al., 2001); however, we believe that multiple strains containing distinct arrays should be examined before attempting to define any such family.

Several aspects of the Treponema gene cassettes are noteworthy when compared to cassettes from other sources (Table 1). The InTde35405 cassettes are on average twice as large as previously described cassettes and the number of cassettes predicted to encode two or three proteins is much higher in the InTde35405 array. The InTde35405 cassettes contain on average almost double the amount of non-protein coding sequence compared to gene cassettes from other sources. While the latter two variables are to some extent sensitive to methods of ORF-prediction, we have found 45 gene cassettes (Tde1843 to Tde1773) associated with the Tde1844 integrase (Fig. 1). Unlike all previously identified integrons, the cassette array of InTde35405 is oriented in the same direction as the integrase. The defining feature of a gene cassette is a recombination site (59-be or attC) consisting of an imperfect inverted repeat containing integrase binding sites (Stokes et al., 1997). The Tde 59-bes conform to this general model (Fig. 2). In the majority (38/45) of the T. denticola 59-bes, the 1L and 1R ends of the inverted repeats (Fig. 2) match more closely in the predicted circular cassettes compared to the linear integrated cassettes. This feature is characteristic of IntI-assembled gene cassette arrays (Recchia & Hall, 1995), suggesting that InTde35405 is a functional integron. Integrons associated with chromosomes frequently have distinctive gene cassette arrays that share very similar 59-bes (Rowe-Magnus et al., 2001), and T. denticola follows this pattern, with 40 of the 45 gene cassettes containing 59-bes closely related to the example in Fig. 2. This group represents a possible ‘Treponema denticola repeat’ family (Rowe-Magnus et al., 2001); however, we believe that multiple strains containing distinct arrays should be examined before attempting to define any such family.

Table 1. Comparison of Treponema gene cassettes to cassettes from other integrons

<table>
<thead>
<tr>
<th>Integron host</th>
<th>No. of cassettes (no. of arrays)*</th>
<th>Typical 59-be length (and frequency)</th>
<th>Cassette size (bp)†</th>
<th>ORFs per cassette‡</th>
<th>Non-coding space per cassette (bp)††‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treponema denticola ATCC 35405†</td>
<td>45 (1)</td>
<td>63–68 bp (40/45)</td>
<td>1267 ± 629</td>
<td>1.51 ± 0.69</td>
<td>193 ± 194</td>
</tr>
<tr>
<td>Vibrio cholerae N16961</td>
<td>174 (1)</td>
<td>127–129 bp (148/174)</td>
<td>678 ± 222</td>
<td>1.15 ± 0.52</td>
<td>90 ± 110</td>
</tr>
<tr>
<td>Pseudomonas (Q, BAM, KM91)§</td>
<td>33 (3)</td>
<td>76–77 bp (24/33)</td>
<td>628 ± 354</td>
<td>1.03 ± 0.17</td>
<td>32 ± 94</td>
</tr>
<tr>
<td>Environmental gene cassettes (EGCs)</td>
<td>159 (?)</td>
<td>NA</td>
<td>637 ± 340</td>
<td>0.91 ± 0.40</td>
<td>66 ± 110</td>
</tr>
<tr>
<td>Significancecell</td>
<td></td>
<td></td>
<td>P &lt; 0.0001, U = 2003</td>
<td>P = 0.0001, U = 5397</td>
<td>P &lt; 0.0001, U = 4341</td>
</tr>
</tbody>
</table>

NA, Not applicable.
*‘Duplicate’ gene cassettes within an array were treated as multiple discrete units.
†Data are means and standard deviations.
‡Non-coding space was defined as the sum of spaces between 59-bes and ORFs (i.e. inter-ORF spaces in multiple ORF cassettes were not considered).
§Gene cassette data from two Pseudomonas stutzeri strains (Q and BAM) and one Pseudomonas straminea strain (KM91; unpublished data) were combined.
‖Two datasets were constructed for each variable – one containing Treponema cassettes and the other containing all other cassettes. The Mann–Whitney U test (Prism GraphPad software) was used to assess whether the Tde cassette pool was significantly different to the pool of cassettes from other genera.
statistical analyses (Table 1) indicate that the gene cassette pool in InTde35405 is distinctive. The less ‘neatly packaged’ gene cassettes of InTde35405 could reflect differences either in the origin(s) of the cassette-associated genes or in the mechanism of gene cassette construction.

Understanding the role of integrons in bacterial evolution requires an appreciation of their phylogenetic and ecological distribution. It is in these contexts that InTde35405 is particularly significant. Firstly, presently known integrons are strongly associated with the Proteobacteria, particularly the beta and gamma subdivisions. Evidence for a broader integron distribution has been hinted at by the presence of intI homologues in the genomes of Pirellula sp. (RB3157) and Gemmata obscuriglobus (unnamed gene), but neither of these intI homologues appears to be associated with gene cassettes. Are these part of functional integrons that simply lack cassettes (Bissonnette & Roy, 1992) or are they integrases with alternative functions in the cell? The unusual genetic organization of InTde35405 suggests that in addressing these questions we should not restrict ourselves to looking for the classical integron structure. Secondly, InTde35405 is significant since its host strain is the first example of a characteristic resident of the normal human microflora demonstrated to contain a large chromosomal gene cassette array. Such arrays can act as accessible reservoirs of antibiotic resistance genes that can be further disseminated by mobile integrons (Rowe-Magnus et al., 2002).

Acknowledgements
Ruth Hall and Michael Gillings are thanked for their comments on the manuscript. Research in the Holmes’ laboratory is supported by the Australian Research Council and the University of Sydney.

Nick Coleman, Sasha Tetu, Neil Wilson and Andrew Holmes

School of Molecular and Microbial Biosciences, Building G08, University of Sydney, NSW, 2006, Australia

Correspondence: Andrew Holmes (A.Holmes@mmb.usyd.edu.au)


DOI 10.1099/mic.0.27569-0