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 35405^T 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*.

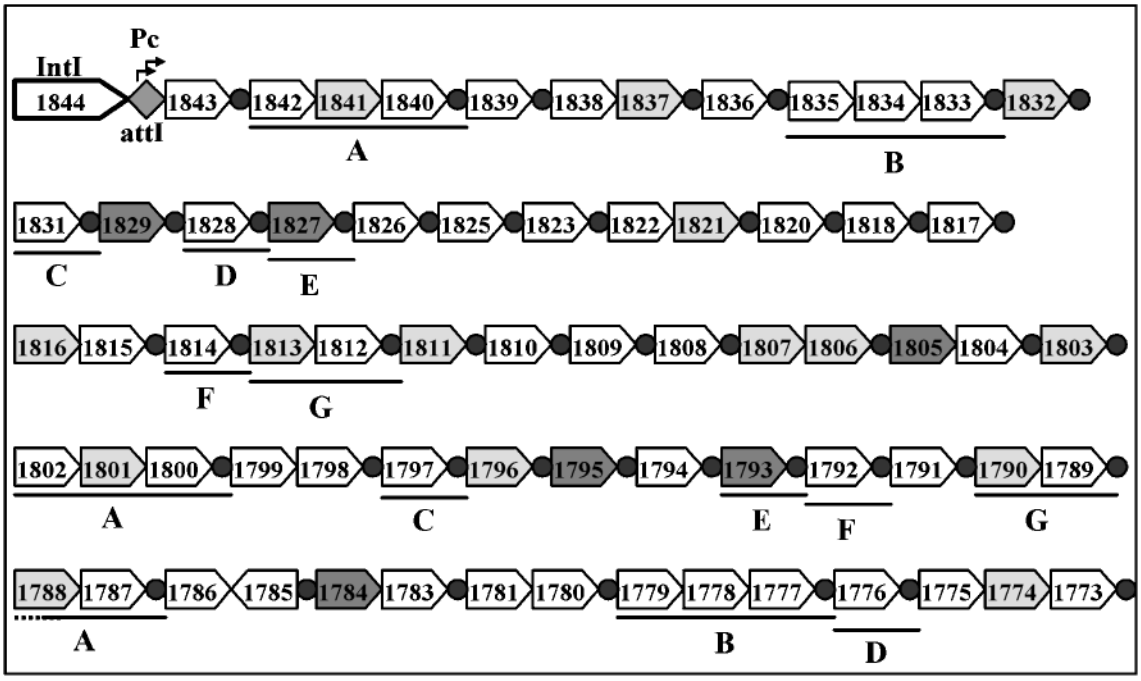

Fig. 1. Structure of *T. denticola* ATCC 35405^T integron. The *intI* gene and cassette ORFs (block arrows, sizes not to scale) are numbered according to GenBank/EMBL/DDBJ accession no. NC_002967. White ORFs have no known function, light grey ORFs are related to conserved hypothetical proteins and dark grey ORFs are related to proteins of known function (BLAST E values $<0{\cdot}00001$). Circles indicate putative 59-be 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.

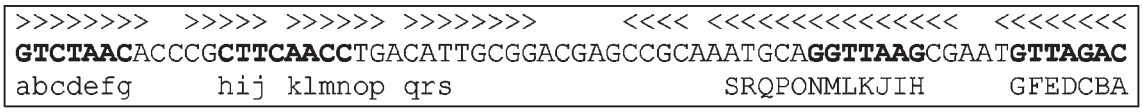

Fig. 2. A typical *T. denticola* 59-be – the 66 bp element associated with Tde1837 (inferred circular form). Inverted repeats specific to this element are indicated by chevrons. Structurally conserved positions found in all 59-bes are indicated by letters below the DNA sequence (Stokes *et al.*, 1997). Putative integrase binding sites 1L, 2L, 2R and 1R, respectively, are shown in bold.

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-base element) junction (G^TT at 1873129 in GenBank/EMBL/DDBJ accession no. NC_002967) and two possible P_C promoters: ttgcagcagtttagattgcaaattattggtttttatgctatagt (−35 and −10 regions underlined).

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.

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,

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

Integron host	No. of cassettes (no. of arrays)*	Typical 59-be length (and frequency)	Cassette size (bp)†	ORFs per cassette†	Non-coding space per cassette (bp)†‡
Treponema denticola ATCC 35405^T	45 (1)	63–68 bp (40/45)	1267±629	1·51±0·69	193±194
Vibrio cholerae N16961	174 (1)	127–129 bp (148/174)	678±222	1·15±0·52	90±110
Pseudomonas (Q, BAM, KM91)§	33 (3)	76–77 bp (24/33)	628±354	1·03±0·17	32±94
Environmental gene cassettes (EGCs)	159 (?)	NA	637±340	0·91±0·40	66±110
Significance‖			$P<0·0001$, U=2003	$P=0·0001$, U=5397	$P<0·0001$, U=4341

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)

Bissonnette, L. & Roy, P. H. (1992). Characterization of In0 of *Pseudomonas aeruginosa* plasmid pVS1, an ancestor of integrons of multiresistance plasmids and transposons of gram-negative bacteria. *J Bacteriol* **174**, 1248–1257.

Hall, R. M. & Collis, C. M. (1995). Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol Microbiol* **15**, 593–600.

Holmes, A. J., Gillings, M. R., Nield, B. S., Mabbutt, B. C., Nevalainen, K. M. H. & Stokes, H. W. (2003a). The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environ Microbiol* **5**, 383–394.

Holmes, A. J., Holley, M. P., Mahon, A., Nield, B., Gillings, M. & Stokes, H. W. (2003b). Recombination activity of a distinctive integron-gene cassette system associated with *Pseudomonas stutzeri* populations in soil. *J Bacteriol* **185**, 918–928.

Nield, B. S., Holmes, A. J., Gillings, M. R., Recchia, G. D., Mabbutt, B. C., Nevalainen, K. M. & Stokes, H. W. (2001). Recovery of new integron classes from environmental DNA. *FEMS Microbiol Lett* **195**, 59–65.

Recchia, G. D. & Hall, R. M. (1995). Gene cassettes: a new class of mobile element. *Microbiology* **141**, 3015–3027.

Rowe-Magnus, D. A., Guerout, A. M., Ploncard, P., Dychinco, B., Davies, J. & Mazel, D. (2001). The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc Natl Acad Sci U S A* **98**, 652–657.

Rowe-Magnus, D. A., Guerout, A. M. & Mazel, D. (2002). Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Mol Microbiol* **43**, 1657–1669.

Seshadri, R., Myers, G. S., Tettelin, H. & 36 others (2004). Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc Natl Acad Sci U S A* **101**, 5646–5651.

Stokes, H. W., O'Gorman, D. B., Recchia, G. D., Parsekhian, M. & Hall, R. M. (1997). Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol Microbiol* **26**, 731–745.

Vaisvila, R., Morgan, R. D., Posfai, J. & Raleigh, E. A. (2001). Discovery and distribution of super-integrons among pseudomonads. *Mol Microbiol* **42**, 587–601.

DOI 10.1099/mic.0.27569-0