Microbiology (1997), 143, 553–562 Printed in Great Britain

Analysis of sequences flanking the *vap* regions of *Dichelobacter nodosus*: evidence for multiple integration events, a killer system, and a new genetic element

Garry A. Bloomfield, Gabrielle Whittle, Matthew B. McDonagh,† Margaret E. Katz and Brian F. Cheetham

Author for correspondence: Brian F. Cheetham. Tel: +61 67 733394. Fax: +61 67 733267. e-mail: bcheetha@metz.une.edu.au

Department of Molecular and Cellular Biology, The University of New England, Armidale, NSW 2351, Australia

***Dichelobacter nodosus* is the causative agent of ovine footrot. The *vap* regions of the *D. nodosus* genome may have arisen by the integration of a genetic element and may have a role in virulence. The virulent *D. nodosus* strain A198 has multiple copies of the *vap* regions. In the present study, sequences to the left and right of *vap* regions 1, 2 and 3 of strain A198 were analysed by Southern blotting and DNA sequencing. The results suggest that *vap* regions 1 and 2 arose by independent integration events into different tRNA genes. The discovery of a second integrase gene (*intB*), a gene with similarity to bacteriophage repressor proteins (*regA*), and a gene similar to an ORF from a conjugative transposon (*gepA*), suggests that a second genetic element, either a bacteriophage or a conjugative transposon, is integrated next to *vap* region 3 in the *D. nodosus* genome. The arrangement of *intB* and the *vap* regions in three other virulent strains and one benign strain was determined using using Southern blotting and PCR. One strain, H1215, contained *vapE′* and not *vapE*, and thus resembles *vap* region 3, suggesting that *vap* region 3 also may have arisen by an independent integration event. In all strains, a copy of *intB* was found next to the *vap* regions. The *vap* regions contain two genes, *vapA* and *toxA*, with similarity to the *hig* genes of the killer plasmid Rts1. Evidence is presented that *vapA* and *toxA* have a similar function in *D. nodosus*.**

Keywords: footrot, virulence, horizontal gene transfer, conjugative transposon

INTRODUCTION

Dichelobacter nodosus (formerly *Bacteroides nodosus*), a Gram-negative anaerobic bacterium with polar type 4 fimbriae, is the principal causative agent of ovine footrot (Beveridge, 1941). Strains of *D. nodosus* are classified as virulent, intermediate or benign, according to the severity of the disease which they cause in sheep. The *vap* regions of the *D. nodosus* genome are found in more than 98 % of virulent strains but are absent from 70 % of benign strains (Rood *et al.*, 1996), suggesting that these DNA sequences may have a role in virulence. The *vap* regions appear to have arisen by the integration of a genetic element into a tRNA gene in the *D. nodosus* genome (Cheetham *et al.*, 1995). A plasmid carrying the *vap* genes has been isolated from one strain of *D. nodosus* (Billington *et al.*, 1996b). The major genes identified within the *vap* regions are a series of virulence-associated protein (*vap*) genes, designated *vapA*, *vapB*, *vapC*, *vapD* and *vapE*, and an integrase gene, *intA*, which has amino acid similarity to the integrases of some bacteriophages of *Escherichia coli*, *Shigella flexneri* and *Vibrio cholerae* (Katz *et al.*, 1992; Cheetham *et al.*, 1995). The absence of a transformation system for *D. nodosus* precludes the direct testing of the role of the *vap* genes in virulence.

Genes similar to *vapA*, *vapC* and *vapD* have been identified in the genome of *Haemophilus influenzae*, and

† **Present address:** Department of Animal Science, The University of New England, Armidale, NSW 2351, Australia.

The GenBank accession numbers for the sequences determined in this work are X98546 (4798 nucleotide sequence beginning at the *Sac*I site in *vap* region 3), X98545 (1065 nucleotide sequence at the left of *vap* region 2) and X98547 (1442 nucleotide sequence beginning at the attachment site at the right hand side of *vap* region 2).

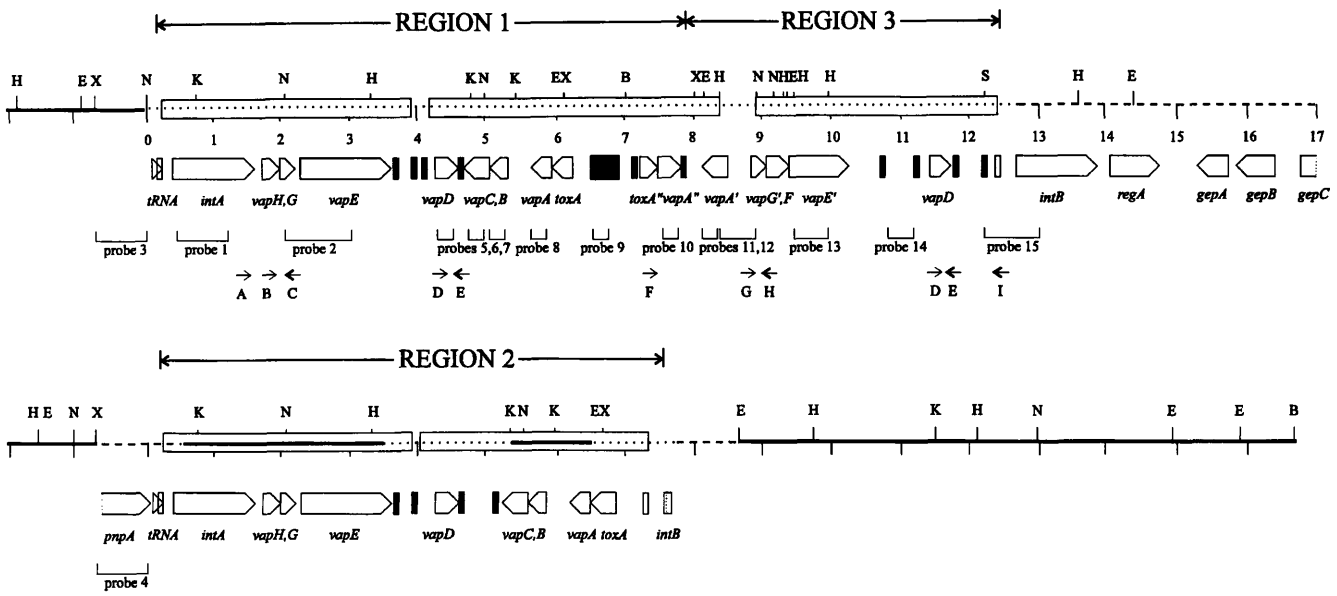

Fig. 1. The *vap* regions of *D. nodosus* strain A198. The numbers show the distance in kb from the leftmost *Nru*I site. Restriction sites shown are *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Kpn*I (K), *Nru*I (N), *Sac*I (S) and *Xho*I (X). The major potential genes are identified by open arrows. Repeated sequences (Cheetham *et al.*, 1995) are indicated as follows: 19 bp attachment sites (small open boxes), 103 bp repeats, or partial copies thereof (small shaded boxes), 102 bp repeats, or partial copies (small filled boxes), and the putative origin of replication (large shaded box). Regions which have been sequenced previously are indicated by dotted lines, and regions which were sequenced in this study are shown by dashed lines. The large open boxes show regions which are not found in the benign strain C305. The positions and direction of extension of the primers A–I used for PCR experiments are shown by small arrows.

are designated *vapA*, *vapC* and *vapD* due to their similarity to the genes from *D. nodosus* (Fleischmann, 1995), but the *H. influenzae* genes are located in different parts of the genome, not adjacent, as in *D. nodosus*. The *vagC* and *vagD* genes, which are located on the virulence plasmid of *Salmonella dublin* (Pullinger & Lax, 1992), are similar to *vapB* and *vapC*, respectively. The arrangement of the *vagC* and *vagD* genes, which are adjacent, overlapping by a single nucleotide, and transcribed from the same promoter, is very similar to that of *vapB* and *vapC*, which overlap by two nucleotides. A Tn*A* insertion in *vag*C reduced the virulence of the strain in which it was found, and it has been proposed that the role of *vagC* and *vagD* is to co-ordinate the replication of the virulence plasmid with cell division (Pullinger & Lax, 1992). In *H. influenzae*, the *vag*C gene, which is similar to *vapB* from *D. nodosus*, is located adjacent to *vap*C. VapB and VapC also show amino acid similarity with putative proteins encoded by the *trbH* region of the *E. coli* F plasmid (Katz *et al.*, 1992).

The virulent *D. nodosus* strain A198 contains three copies of the *vap* regions, designated *vap* regions 1, 2 and 3 (Katz *et al.*, 1994; Fig. 1). Regions 1 and 3 are adjacent in the genome of *D. nodosus*, while region 2 is located elsewhere. The complete sequence of *vap* regions 1 and 3 has been determined (Katz *et al.*, 1992, 1994; Cheetham *et al.*, 1995). *vap* region 1 contains the *intA* and *vapA–E* genes, together with several smaller ORFs, and a gene, *vapA″* (formerly *vapI*; Cheetham *et al.*, 1995), encoding a protein with 36·6% amino acid similarity to *VapA*. *vap* region 3 contains *vapD*, *vapA′*, whose product has 57·3% amino acid similarity to VapA, and *vapE′*, encoding a protein with 62.6% amino acid similarity to VapE. Southern blotting and partial DNA sequencing has been used to show that *vap* region 2 contains *vapA–D* (Katz *et al.*, 1994) but the composition of the remainder of *vap* region 2 was unknown.

The multiple copies of the *vap* regions could have arisen by chromosomal duplications, or by multiple independent integration events. In this study, Southern blotting was used to determine the composition of *vap* region 2, and the DNA sequences flanking *vap* regions 1, 2 and 3 were analysed. In addition, the arrangements of the *vap* regions in four other strains of *D. nodosus* was determined. The results suggest that *vap* regions 1, 2 and 3 arose by independent integration events.

METHODS

General methods. Methods for the growth of *D. nodosus*, preparation of DNA, and cloning and analysis of DNA have been reported elsewhere (Katz *et al.*, 1994).

***Dichelobacter nodosus* strains.** Virulent *D. nodosus* strains A198, B1006, G1220 and H1215, and benign strains C305 and H1204, were provided by Dr J. I. Rood, Monash University, Clayton, Victoria, Australia.

Southern blot hybridizations. The methods for the preparation of probes, stringent hybridization conditions, and chemiluminescent detection have been reported elsewhere (Cheetham *et al.*, 1995). The probes were restriction enzyme

fragments cloned into the vector pUC18. Genomic DNA from strains A198, B1006, G1220, H1204 and H1215 was digested with the following enzymes: *Eco*RI, *Hin*dIII, *Kpn*I, *Ssp*I, *Eco*RI and *Bam*HI, *Eco*RI and *Hin*dIII, *Eco*RI and *Nru*I, *Eco*RI and *Xho*I, *Hin*dIII and *Acc*I, *Hin*dIII and *Bam*HI, *Hin*dIII and *Hin*dII, *Hin*dIII and *Nru*I, *Hin*dIII and *Sac*I, *Hin*dIII and *Ssp*I, *Sac*I and *Nru*I, *Sac*I and *Xho*I. Not all enzymes were used for all strains.

Polymerase chain reactions. The 25 μl amplification reactions contained 1 μmol of each oligonucleotide primer, 1·8 mM $MgCl_2$, 0·2 mM dNTPs, 1 × reaction buffer (supplied with the enzyme by the manufacturer, BresaTec; 10 × reaction buffer contains 67 mM Tris/HCl, pH 8·8, 166 mM $(NH_4)_2SO_4$, 2 mg gelatin ml^{-1}, 4·5% Triton X-100), 1·0 unit *Taq* DNA polymerase, and approximately 1 ng genomic DNA. The reaction mixtures were amplified for 30 cycles (each cycle consisting of 90 s at 94 °C, 60 s at 60 °C and 120–300 s at 72 °C) in a Corbett FTS-320 thermal cycler (Corbett Research). A 10 μl sample of the amplification products was analysed by agarose gel electrophoresis. Oligonucleotide primers were 20 nucleotides long. Primers A, B, D, F and G were identical to sequences starting at nucleotides 1480, 1883, 4260, 7413 and 8813 in *vap* regions 1 and 3. Primers C, E, H and I were complementary to sequences starting at positions 2143, 4512, 9250 and 12535 (GenBank accession number L31763) in *vap* regions 1 and 3 . The approximate positions of the primers are shown in Fig. 1. The primer pairs used were: A and C, A and E, B and E, D and E, D and I, F and I, and G and H.

DNA sequence analysis. DNA sequencing was done using T7 DNA polymerase kits and double-stranded DNA templates, as described by the manufacturer (BresaTec). The nucleotide sequence of both DNA strands was determined in an overlapping manner and across all internal restriction sites. GenBank searches were carried out using the FASTA-TRANS (Pearson & Lipman, 1988) program through the Australian Genome Information Service (ANGIS). In all cases, the Kolmogorov–Smirnov statistic was less than 0·1, and all amino acid similarities reported were highly significant, with probability values of $< 0{\cdot}005$ (Gaeta, 1996). Sequences related to IntB were aligned using the CLUSTALW program (Higgins *et al.*, 1992).

RESULTS

Comparison of sequences to the left of *vap* regions 1 and 2

A comparison of the restriction maps of the left-hand ends of *vap* regions 1 and 2 suggests that divergence between the two regions occurs to the left of the leftmost *Kpn*I site (Fig. 1). Southern blotting using probes 1 and 2 showed that *vap* region 2 contains the genes *intA* and *vapE* (data not shown). However, probes 3 and 4 each detected only one band, suggesting that there is a single copy of these sequences in the genome of *D. nodosus* strain A198. To confirm these results, the DNA sequence of a 1·1 kb *Xho*I–*Pvu*II fragment from the left-hand end of *vap* region 2 was determined. Comparison of the sequences (Fig. 2) shows that there is 97% sequence identity between *vap* regions 1 and 2 in the first 190 nucleotides to the right of the attachment site. There is 79% conservation of the sequence of the attachment site. An 89 bp $tRNA_{Ser}$ gene was identified at the left-hand junction of *vap* region 2, with 83·5% identity to the *ser-tRNA-V* gene of *E. coli* (Grosjean *et al.*, 1985). This tRNA gene is similar to the previously identified tRNA gene at the left-hand end of *vap* region 1 (Cheetham *et al.*, 1995), having 81% identity over 89 bp. However, the tRNA gene from *vap* region 2 has the anticodon GGA instead of GCU, has a shorter extra arm, and has base substitutions in the stem of the TΨC loop. These results suggest that *vap* region 2 arose by the integration of the *vap* element into a tRNA gene different from the one in which *vap* region 1 is inserted.

The sequences to the left of the tRNA genes in *vap* regions 1 and 2 are not related (Fig. 2). Part of an ORF, designated *pnpA*, with approximately 67% identity over 239 amino acids to the polynucleotide phosphorylase genes of *E. coli* (Regnier *et al.*, 1987), *Photorhabdus luminescens* (Clarke & Dowds, 1994) and *H. influenzae* (Fleischmann, 1995) was identified to the left of the tRNA gene in *vap* region 2. Partial sequencing of the *Xho*I–*Nru*I fragment at the left of *vap* region 1 revealed part of an ORF which had 48·2% amino acid identity with the *Mycobacterium smegmatis* aspartokinase gene (data not shown; Cirillo *et al.*, 1994).

Identification of an integrase gene to the right of *vap* region 3

The sequence of 4·8 kb of DNA beginning at the *Sac*I site at the right-hand end of region 3 (Fig. 1) was determined, and five potential genes were identified. One of these, designated *intB*, encodes a protein with deduced amino acid similarity to several bacterial integrases. However, *intB* appears to be a pseudogene, as there is a stop codon at position 130. Similarity to other integrase genes continues after this stop codon, in a different reading frame. The amino acid sequence of *intB*, adjusted to remove this stop codon, was compared to the amino acid sequences of other integrases. The highest level of amino acid identity was to *E. coli* retronphage φR73 (Sun *et al.*, 1991), with 33·8% amino acid identity over 403 amino acids. There was 32·6% amino acid identity with the integrase from bacteriophage SF6 of *Shigella flexneri* (Clark *et al.*, 1991), and 31·3% amino acid identity with an integrase from *Vibrio cholerae* (Kovach & Peterson, 1994). A multiple alignment of the amino acid sequences of these integrases (data not shown) revealed that 15·6% of amino acid residues are identical in all four proteins, including His^{362}, Arg^{365} and Tyr^{399}, which are conserved between a number of bacteriophage integrases (Argos *et al.*, 1986). The *intB* gene product also shows 31·0% amino acid identity with the *intA* gene product from *vap* region 1 of *D. nodosus* (Cheetham *et al.*, 1995).

The ORF designated *regA* adjacent to *intB* has amino acid similarity to bacterial DNA-binding proteins and bacteriophage repressor proteins. The level of amino acid identity was 40·5% to the pectin lyase regulator RdgA of *Erwinia carotovora* (Liu *et al.*, 1995), 39·2% to the regulatory protein PrtR of *Pseudomonas aeruginosa* (Matsui *et al.*, 1993), and 36·2% to the repressor (*cI* gene

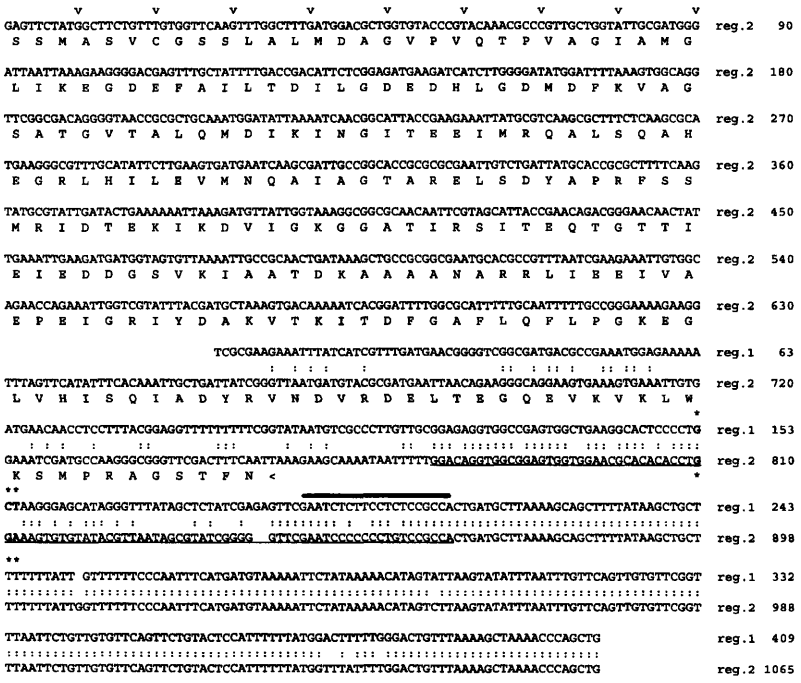

Fig. 2. Alignment of the DNA sequence from the left-hand end of *vap* region 2 (reg.2, bottom) with nucleotides 1–410 of the sequence from the left-hand end of *vap* region 1 (reg.1, top). The amino acids constituting part of the *pnpA* gene encoded by the sequence from region 2 are shown below this sequence, with the stop codon shown by <. The nucleotides corresponding to the tRNA gene identified in region 2 are underlined, and the anticodons of the tRNA genes from both regions are indicated by asterisks. The putative attachment site in region 2 is overlined. Nucleotides which are identical in both sequences are joined by colons. v, tenth nucleotide.

product) of bacteriophage ϕ80 of *E. coli* (Ogawa *et al.*, 1988). A putative helix–turn–helix motif and a putative leucine-zipper domain were identified in the *regA* gene (data not shown; Brennan & Matthews, 1989).

Three further potential genes were identified in this region (Fig. 1). The gene designated *gepA* (genetic element protein A) encodes a protein with 33·3% amino acid identity to ORF4, an ORF downstream from the *rteC* gene near the origin of transfer of a conjugative transposon from *Bacteroides thetaitotaomicron* (Stevens *et al.*, 1993). The *gepB* gene encodes a protein with 64·9% amino acid similarity over 188 amino acids to an unidentified ORF in the *cob* gene cluster from *Pseudomonas denitrificans* (Crouzet *et al.*, 1991), while the *gepC* gene shows no significant similarity to any genes in the databases.

The presence of an integrase gene followed by a DNA-binding regulatory gene suggests that the sequences to the right of *vap* region 3 are part of a second integrated mobile genetic element, which could be an integrated bacteriophage. However, the similarity of *gepA* to an ORF downstream from the *rteC* gene of *B. thetaiotaomicron* and the fact that the integrases of some conjugative transposons are members of the lambda integrase family (Salyers *et al.*, 1995), as is *intB*, raises the possibility that this second genetic element is a conjugative transposon.

Comparison of sequences to the right of *vap* regions 2 and 3

The sequence of *vap* region 2 from *toxA* to just past the attachment site (Fig. 1) has been determined previously (Katz *et al.*, 1994). We determined the sequence from the attachment site to the next *Eco*RI site (Fig. 1). Comparison of this sequence with the sequence to the right of *vap* region 3, beginning at the attachment site, is shown in Fig. 3. The two sequences are 92·5% identical over the first 560 nucleotides, with 80·7% identity over the first 140 amino acids of the IntB proteins encoded by each region. However, similarity between the sequences ends at this point. It is of interest that the *intB* gene adjacent to *vap* region 2 is not interrupted by the stop codon found in the *intB* gene adjacent to *vap* region 3. However, there appears to be only a partial or truncated copy of the *intB* gene next to *vap* region 2. Thus, although there are two copies of this gene in *D. nodosus* strain A198, neither copy appears to be functional.

Arrangement of the *vap* regions in four other strains of *D. nodosus*

To further investigate the evolution of the *vap* regions, the arrangements of the *vap* genes in four strains of *D. nodosus*, from serogroups B (virulent strain B1006), G (virulent strain G1220) and H (benign strain H1204 and virulent strain H1215) were compared with the arrangement of the *vap* genes in virulent strain A198 (serogroup A). These experiments were also designed to investigate the composition of the *vap* region of strain H1204, which is one of the 30% of benign strains which have been shown previously to contain at least part of the *vap* region. Preliminary Southern blots (Katz *et al.*, 1991) had suggested that the *vap* regions in these strains were different from those in strain A198. Maps of the *vap* regions in these strains were constructed by analysing the results of Southern blot experiments, using probes (Fig. 1) which would detect *intA*, *intB*, *vapA-E*, *vapA′*, *vapA″* and *vapE′*, and PCR experiments, using

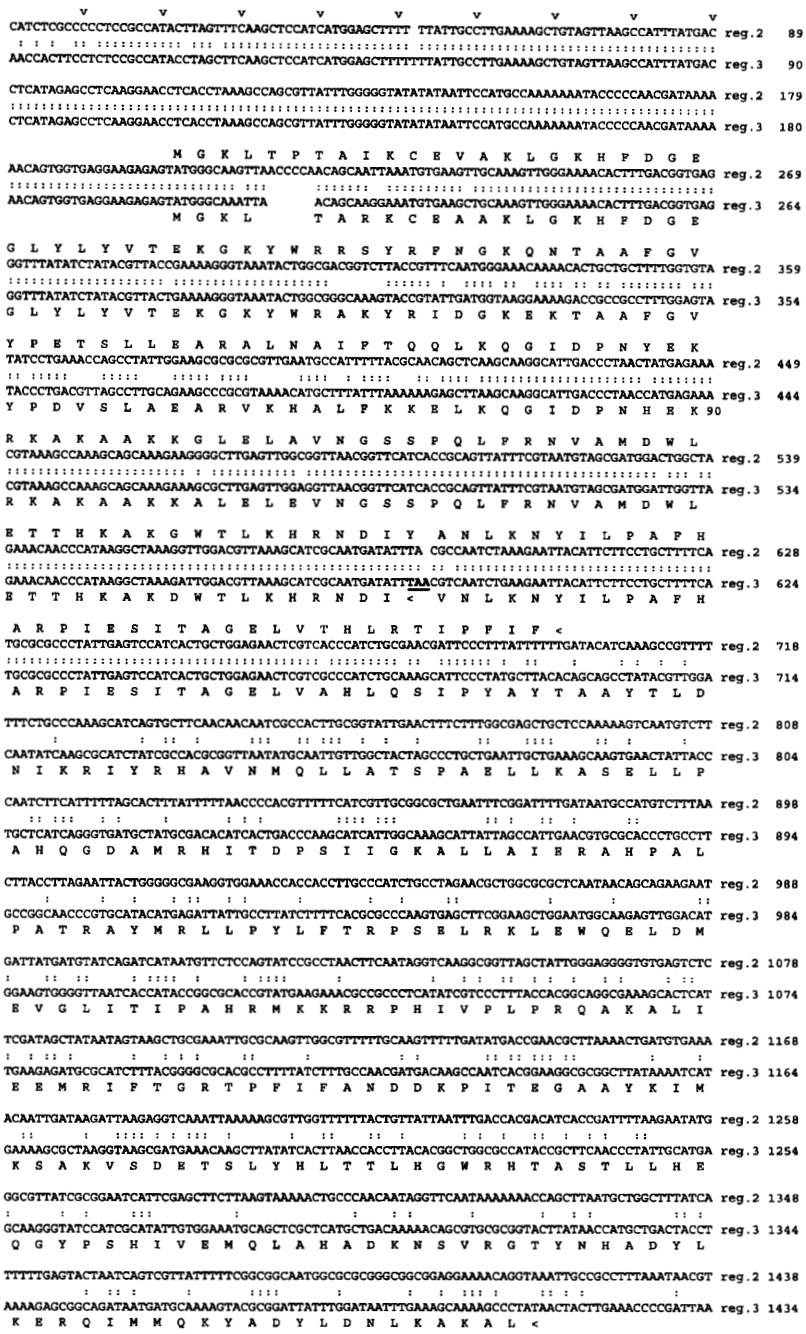

Fig. 3. Alignment of the sequence from the right-hand end of *vap* region 2, beginning at the attachment site (reg.2, top sequence) with the sequence from the right-hand end of *vap* region 3, beginning at the attachment site (reg.3, bottom sequence). The amino acid sequences of the *intB* genes are shown above (reg.2) or below (reg.3) the DNA sequences. The stop codon at position 586 which interrupts the *intB* coding region is shown in bold and underlined. The first 214 nucleotides of the sequence from the right-hand end of region 2 have been reported previously (Katz *et al.*, 1994).

primers complementary to several regions of the genome of *D. nodosus* strain A198 (Fig. 1).

The results from the Southern blot experiments are summarized in Fig. 4. All strains contained at least one copy of *vapA*, *vapB*, *vapC* and *vapD*, which were located adjacent to each other. However, not all strains contained a copy of *vapE*. The virulent strain H1215 lacks *vapE*, but contains a copy of *vapE′* (Fig. 5), located in the same position as *vapE* in all of the other strains tested. This is similar to the arrangement in *vap* region 3 in strain A198, except that *vap* region 3 does not contain *vapA–C* or *toxA*. Further similarity between *vap* region 3 of strain A198 and the *vap* region of strain H1215 was shown by hybridization experiments with probe 14 (Fig. 1). This probe detects a sequence which is absent from *vap* region 1, located between *vapD* and *vapC* in *vap* region 2, and located to the left of *vapD* in *vap* region 3. In strain H1215, this sequence is also located to the left of *vapD*. The similarity between the *vap* region of strain H1215 and *vap* region 3 of strain A198 suggests that *vap* region 3 of strain A198 could have arisen by integration of a separate *vap* element, containing *vapE′*, followed by the loss of *vapA–C*, rather than by partial duplication of *vap* region 1 and subsequent divergence of the *vapE* gene.

The absence of *vapE* from strain H1215 suggests either

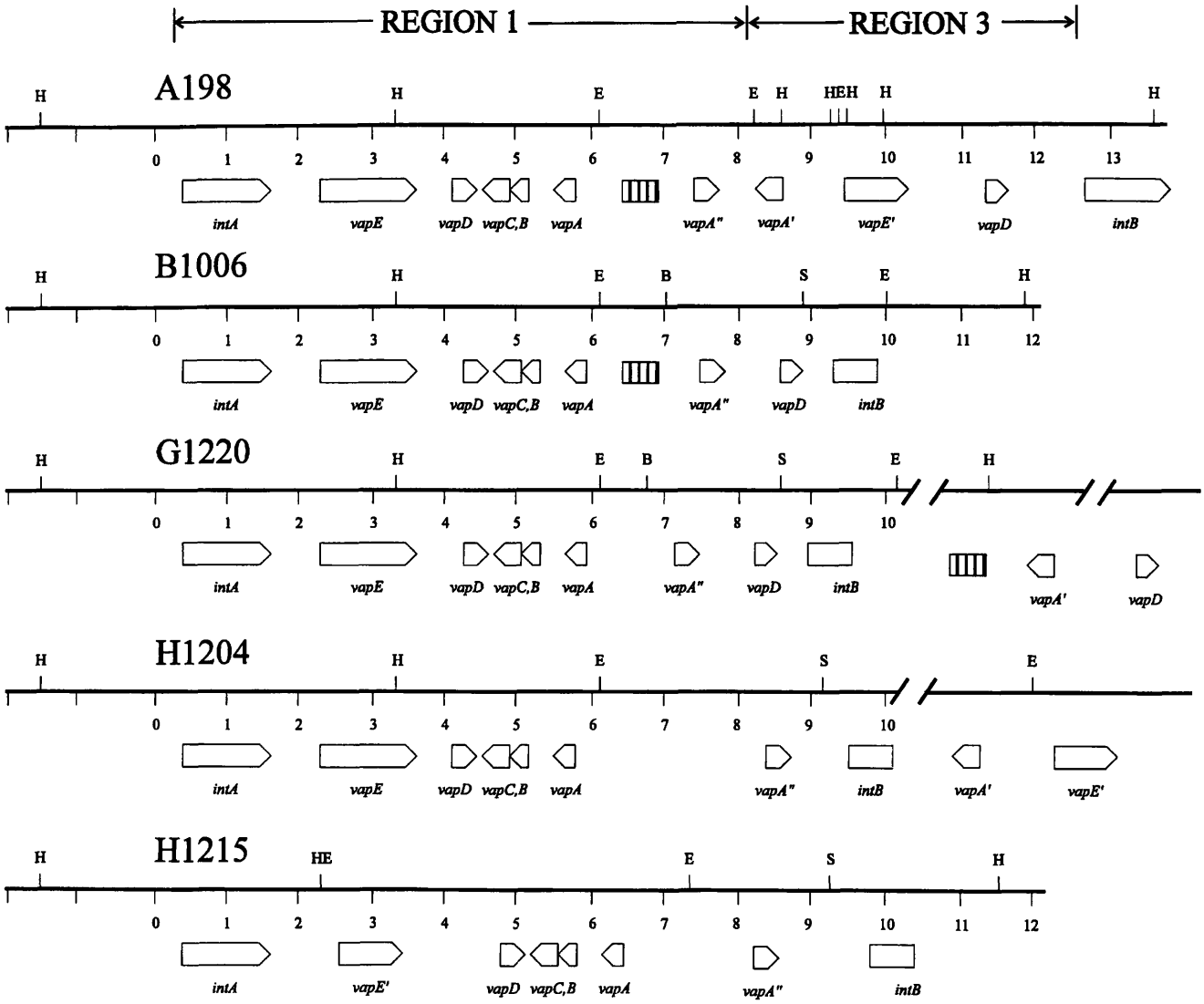

Fig. 4. Map of the *vap* regions of *D. nodosus* strains A198, B1006, G1220, H1204 and H1215. Only *vap* regions 1 and 3 are shown for strain A198. Restriction sites shown are *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), and *Sac*I (S). The positions of genes were determined by Southern blot hybridization and PCR. The hatched box represents the putative origin of replication. The relative positions and orientations of segments separated by broken lines are not known. The numbers show distances in kb.

that VapE is not essential for virulence, or that VapE′ can perform the same function. Strain H1215 has a single copy of all of the genes in the *vap* region, so multiple copies are not required for virulence, and does not contain *vapA′*, suggesting that this is not required for virulence. By contrast, the benign strain H1204 has one copy of all of the genes identified in the *vap* regions, so the lack of virulence in this strain must be attributable to factors outside the *vap* region, unless one or more of the *vap* gene products is non-functional in this strain.

The two copies of *intB* which have been identified in strain A198, together with one copy of *intB* in strains B1006, G1220, H1204 and H1215, are located adjacent to the *vap* regions (Fig. 4). Strains H1204 and H1215 contain an additional copy of *intB*, which does not appear to be associated with the *vap* regions.

PCR experiments were used to confirm the maps produced from the Southern blot results. The results of these experiments are summarized in Table 1. No product was obtained using DNA from strain C305 as a template, as this strain does not contain most of the sequences in the *vap* regions. Primers D and E, which flank the *vapD* gene, gave a product of approximately 300 bp in all strains. We have previously shown that the three copies of the *vapD* coding region in strain A198 are identical (Katz *et al.*, 1992, 1994; Cheetham *et al.*, 1995). PCR assays using primer G, which anneals just before *vapG′*, and primer H, which is complementary to a

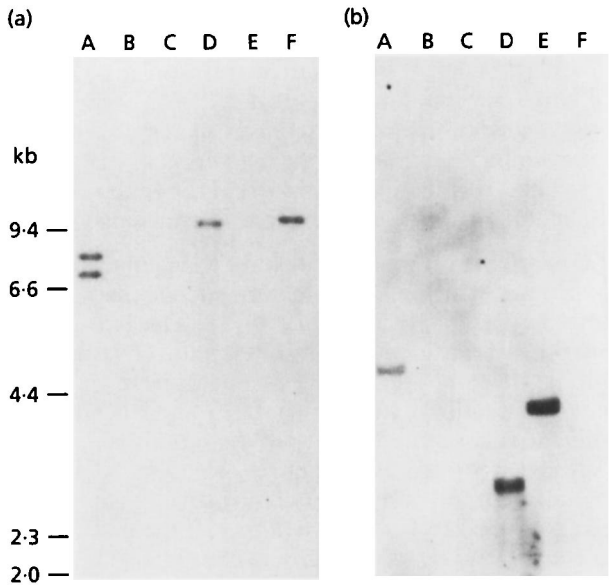

Fig. 5. Southern blot analysis of the *vapE* and *vapE′* genes of several strains of *D. nodosus*. Genomic DNA from strains A198 (lane A), C305 (lane C), H1204 (lane D), H1215 (lane E) and G1220 (lane F) was digested with *Eco*RI and hybridized with probes specific for *vapE* (a) or *vapE′* (b). Lanes B contained no DNA. Fragment sizes are indicated on the left.

Table 1. Summary of the results of PCR experiments

DNA from the strains indicated was used in PCR assays, as described in Methods. The sizes of the bands, determined by agarose gel electrophoresis, are shown in kb. A dash indicates that no product was obtained.

Primer pair	A198	B1006	C305	G1220	H1204	H1215
D and E	0·30	0·30	–	0·30	0·30	0·30
G and H	0·45	0·78	–	0·78	–	0·45
D and I	1·0	1·0	–	1·0	–	–
F and I	–	1·8	–	1·8	1·5	1·5
B and E	2·6	1·25	–	1·25	2·6	3·2
A and E	3·0	3·6	–	3·6	3·0	4·0
A and C	0·60	0·60	–	0·60	0·60	0·60

sequence within *vapF*, gave bands of 450 bp in strains A198 and H1215, and 780 bp in strains B1006 and G1220. This confirms the similarity between *vap* region 3 and the *vap* region of strain H1215, and shows that there is variation in the region around *vapG′* and *vapF* in some strains.

Assays using primers D and I confirmed the presence of a copy of *vapD* near *intB* in strains B1006 and G1220, and not in the two H strains. The product obtained using primers F and I and DNA from strains B1006 and G1220 was 1·8 kb, slightly larger than that obtained using the two H strains, 1·5 kb. This suggests that there has been an insertion or deletion of about 300 nucleotides between *vapA″* and *intB* in these strains. The use of primers B and E led to a much shorter product in strains B1006 and G1220 than in the other strains. This suggests that *vapH* is located on the left side of *vapE* (or *vapE′*) in strains A198, H1204 and H1215, and on the right side of *vapE* in strains B1006 and G1220. Alternatively, there may be copies of *vapH* and *vapG* on both sides of *vapE* in these two strains. The longer products obtained from these strains using the primers A and E is consistent with this second explanation. The 4·0 kb product in strain H1215 is consistent with the location of *vapE′* in this position in this strain. Finally, primers A and C gave a product of the same size in all strains. In summary, the PCR experiments have confirmed the maps derived from the Southern blotting data, provided further evidence that *vap* region 3 of strain A198 and the *vap* region of strain H1215 are related, and shown that there is diversity between the strains in the areas around *vapH*, *vapG*, *vapG′* and *vapF*. Small deletions, insertions, or duplications appear to have occurred in these areas.

A killer system in *D. nodosus*?

VapA has been found recently to have 22% amino acid identity with the HigA protein from the killer plasmid Rts1 (Billington *et al.*, 1996a; Tian *et al.*, 1996). We had previously identified an ORF upstream of *vapA*, designated ORF118 (Katz *et al.*, 1992), which lacks a Shine–Dalgarno sequence (Shine & Dalgarno, 1974). However, there is a start codon 26 codons downstream from the original proposed start codon, and this second start codon is preceded by a weak Shine–Dalgarno sequence. The predicted protein product using this new start codon is 92 amino acids long, has 45·6% amino acid identity with the 92 amino acid HigB protein from the plasmid Rts1, and has now been designated ToxA. The arrangement of the *toxA* and *vapA* genes in *D. nodosus* strain A198 is the same as the arrangement of *higB* and *higA* on the plasmid Rts1. VapA″, which has 36·6% amino acid identity with VapA, has 22·1% amino acid identity with HigA. *vapA″* is preceded by a gene, designated *toxA″*, whose predicted protein product has 31·5% amino acid identity with ToxA, and 29·3% amino acid identity with HigB. The *toxA″* gene was not identified previously (Cheetham *et al.*, 1995), as it begins with the start codon GTG. The remaining copy of *vapA* found in strain A198, *vapA′*, does not appear to be preceded by a copy of the *toxA* gene.

A 1·86 kb *Xho*I fragment from *vap* regions 1 and 3, containing *toxA″*, *vapA″* and part of the *toxA* gene, was inserted into the *Sal*I site of the plasmid pUC18 (plasmid pBC105; Fig. 6) and used to transform *E. coli* to ampicillin resistance. DNA was prepared from transformants, digested with *Sph*I, diluted, religated, and used to transform *E. coli*. This deletion removes part of the *vapA″* gene (Fig. 6). The transformation frequency using the religated plasmid was much lower than expected, and, even though the plasmid DNA appeared to have been cut to completion before ligation, the few transformants obtained, with one exception, contained the complete 1·86 kb *Xho*I fragment. DNA from one

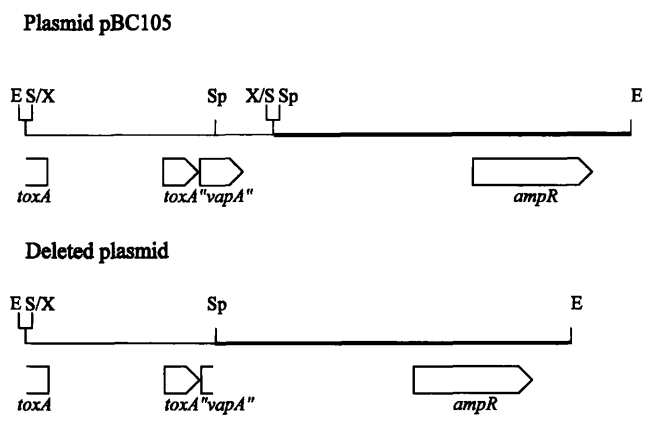

Fig. 6. Map of the plasmid pBC105 and the deleted derivative. The restriction sites shown are *Eco*RI (E), *Sal*I (S), *Sph*I (Sp) and *Xho*I (X). Sequences from pUC18 are shown by the heavy line.

transformant colony contained two plasmids, the original plasmid containing both the *vapA″* and *toxA″* genes, and the deleted plasmid containing only the *toxA″* gene. DNA from the deleted plasmid was purified from a band excised after agarose gel electrophoresis, and sequenced, confirming that this plasmid contained the *toxA″* gene alone. No transformants were obtained when DNA from the deleted plasmid was used to transform *E. coli*. Dilution streaking was used to obtain single colonies from the original transformant, and all such colonies obtained contained a mixture of two plasmids. These results show that it is not possible to obtain viable transformants when *toxA″* is expressed in cells in the absence of *vapA″*, and are consistent with the hypothesis that *toxA* encodes a protein which is toxic to *E. coli*, and *vapA* encodes an antidote protein. Northern blot analysis (data not shown) has demonstrated that *toxA* is expressed in *D. nodosus* strain A198.

DISCUSSION

Similarity between the DNA sequences of the left-hand ends of *vap* regions 1 and 2 begins within tRNA genes, and the sequences are almost identical after the attachment site. Since the two tRNA genes are different, this suggests that *vap* regions 1 and 2 arose by independent integrations of the *vap* element into the two tRNA genes. *vap* region 3 could have arisen by partial duplication of *vap* region 1. However, the discovery of a strain of *D. nodosus*, H1215, which, like *vap* region 3, has the gene *vapE′* and lacks *vapE*, suggests that *vap* region 3 may also have arisen by an independent integration event.

We have identified a new genetic element integrated into the *D. nodosus* chromosome at the right-hand end of *vap* region 3. The arrangement of an attachment site followed by an integrase gene and then a gene encoding a regulatory protein is characteristic of some lambdoid bacteriophages, such as λ and P2 (Calendar *et al.*, 1981), suggesting that this genetic element is an integrated bacteriophage. However, the *gepA* gene which follows *regA* is similar in predicted amino acid sequence to an ORF from a conjugative transposon from *B. thetaiotaomicron* (Stevens *et al.*, 1993). Some conjugative transposons have integrases which are members of the lambda integrase family (Salyers *et al.*, 1995), as is *intB*. Thus, the possibility exists that the genetic element which contains *intB* is a conjugative transposon.

Conjugative transposons may have a role in the transfer of plasmids which are too small to encode the necessary transfer genes (Salyers *et al.*, 1995). The *vap* plasmid which has been identified in one strain of *D. nodosus* (Billington *et al.*, 1996b) does not appear to encode functions required for conjugation, and another non-conjugative plasmid, pEC1, of *D. nodosus* has been identified recently (G. Whittle & B. F. Cheetham, unpublished). However, no conjugative plasmids have been identified so far in *D. nodosus*. If the *intB* element is a conjugative transposon, it may have had a role in the transfer of the *vap* plasmid or pEC1 into *D. nodosus*, or between *D. nodosus* strains.

Six of the copies of *intB* which have been identified so far in *D. nodosus* strains are found in association with the *vap* regions. This could be explained if the *vap* plasmid and the *intB* element use the same DNA sequence (attachment site) for integration. Alternatively, conjugative transposons may integrate into co-resident plasmids and mobilize them (Salyers *et al.*, 1995). Thus, the *intB* element and the *vap* plasmid may have been transferred together. However, this hypothesis is not supported by evidence for the generation of *vap* regions 1, 2 and 3 by independent integration events. The presence of a partial copy of *intB* to the right of *vap* region 2 of *D. nodosus* strain A198 is of interest. This could have been formed by integration next to *vap* region 2 of the genetic element carrying *intB*, followed by loss of almost all of the sequences from this element.

The *vapB* and *vapC* genes of *D. nodosus* are similar to the *vagC* and *vagD* genes from the virulence plasmid of *Salmonella dublin* (Pullinger & Lax, 1992). It is of interest that the *Salmonella* virulence plasmid also contains a gene required for cobalamin transport (Rioux *et al.*, 1990), as we have found similarity between the gene *gepB* and a gene from the *cob* gene cluster of *Pseudomonas denitrificans* (Crouzet *et al.*, 1991). In addition, the *Salmonella* virulence plasmid contains a gene, *rsd*, with similarity to the lambda family of integrases (Krause & Guiney, 1991). The *vagC*, *vagD*, *rsd* and cobalamin transport genes are located within a 15 kb segment of the *Salmonella* virulence plasmid (Dorman, 1994). The similarities between genes of this plasmid and genes from within, or near, the *vap* regions of *D. nodosus* suggests a possible evolutionary relationship between these genetic elements.

The *vap* regions of *D. nodosus* are found in 98% of virulent strains tested, but in only 30% of benign strains (Rood *et al.*, 1996), and they may have a role in virulence. *D. nodosus* isolates may be divided into nine serogroups, based on their fimbrial antigens, and both virulent and

benign strains are found in each serogroup (Claxton *et al.*, 1983). In addition, most serogroups contain both strains with *vap* regions and strains which lack the *vap* sequences (Katz *et al.*, 1991). The products of the *vap* genes may have a direct role in virulence, or, alternatively, the *vap* regions may be necessary for the maintenance or acquisition of other virulence factors. A very small number of virulent strains exist which appear to lack *vap* sequences, suggesting that *vap* sequences may be necessary only for the acquisition of other virulence factors. This proposed role for the *vap* sequences is supported by the observation that another genetic element associated with virulence, the *vrl* region, is found only in strains which contain *vap* sequences (Billington *et al.*, 1996a). However, not all virulent strains contain the *vrl* sequences. We propose the following model for the role of the *vap* regions in the evolution of *D. nodosus*. *D. nodosus* strains of different serogroups may have evolved from an ancestral, benign strain. The *vap* sequences may have been acquired subsequently by some members of the different serogroups. This was followed by the acquisition of other virulence factors by some of the strains which had acquired the *vap* regions. Thus, all virulent strains would contain *vap* sequences, along with some benign strains, and these would be distributed throughout the serogroups. The VapA/ToxA killer system would act to maintain the *vap* regions in all strains. Rarely, the *vap* region may be lost from virulent strains.

ACKNOWLEDGEMENTS

We thank J. Druitt for excellent technical assistance, and gratefully acknowledge the support of the Australian Research Council.

REFERENCES

Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Linquist, E., Hoess, R. H., Kahn, M., Kalionis, B., Narayan, S. V. L., Pierson, L. S., III, Sternberg, N. & Leong, J. M. (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J* **5**, 433–440.

Beveridge, W. I. B. (1941). Footrot in sheep: a transmissible disease due to infection with *Fusiformis nodosus*. *CSIRO Bull* **140**, 1–40.

Billington, S. J., Johnston, J. L. & Rood, J. I. (1996a). Virulence regions and virulence factors of the ovine footrot pathogen, *Dichelobacter nodosus*. *FEMS Microbiol Lett* **145**, 147–156.

Billington, S. J., Sinistaj, M., Cheetham, B. F., Ayres, A., Moses, E. K., Katz, M. E. & Rood, J. I. (1996b). Identification of a native *Dichelobacter nodosus* plasmid and implications for the evolution of the *vap* regions. *Gene* **172**, 111–116.

Brennan, R. G. & Matthews, B. W. (1989). The helix–turn–helix DNA binding motif. *J Biol Chem* **264**, 1903–1906.

Calendar, R., Ljungquist, E., Deho, G., Usher, D. C., Goldstein, R., Youderian, P., Sironi, G. & Six, E. W. (1981). Lysogenization by satellite phage P4. *Virology* **113**, 20–38.

Cheetham, B. F., Tattersall, D. B., Bloomfield, G. A., Rood, J. I. & Katz, M. E. (1995). Identification of a bacteriophage-related integrase gene in a *vap* region of the *Dichelobacter nodosus* genome. *Gene* **162**, 53–58.

Cirillo, J. D., Weisbrod, T. R., Pascopella, L., Bloom, B. R. & Jacobs, W. R. J. (1994). Isolation and characterization of the aspartokinase and aspartate semialdehyde dehydrogenase operon from mycobacteria. *Mol Microbiol* **11**, 629–639.

Clark, C. A., Beltrame, J. & Manning, P. A. (1991). The *oac* gene encoding lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-coding gene on the genome of *Shigella flexneri* bacteriophage Sf6. *Gene* **107**, 43–52.

Clarke, D. J. & Dowds, B. C. (1994). The gene coding for polynucleotide phosphorylase in *Photorhabdus* sp. strain K122 is induced at low temperatures. *J Bacteriol* **176**, 3775–3784.

Claxton, P. D., Ribeiro, L. A. & Egerton, J. R. (1983). Classification of *Bacteroides nodosus* by agglutination tests. *Aust Vet J* **60**, 331–334.

Crouzet, J., Levy-Schil, S., Cameron, B., Cauchois, L., Rigault, S., Rouyez, M.-C., Blanche, F., Debussche, L. & Thibaut, D. (1991). Nucleotide sequence and genetic analysis of a 13·1 kb *Pseudomonas denitrificans* DNA fragment containing five COB genes and identification of structural genes encoding COB(I) alamin adenosyltransferase, cobyric acid synthase and bifunctional cobin. *J Bacteriol* **173**, 6074–6087.

Dorman, C. J. (1994). The bacterial genome: plasmids. In *Genetics of Bacterial Virulence*, pp. 52–87. Oxford: Blackwell.

Fleischmann, R. D. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.

Gaeta, B. A. (1996). Database similarity searching using BLAST and FASTA. *Aust Biotechnol* **5**, 282–290.

Grosjean, H., Nicoghosian, K., Baumont, E., Soll, D. & Cedergren, R. (1985). Nucleotide sequences of two serine tRNAs with a GGA anticodon: the structure–function relationships in the serine family of tRNAs. *Nucleic Acids Res* **13**, 5697–5706.

Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992). Clustal V: improved software for multiple sequence alignment. *Comput Appl Biosci* **8**, 189–191.

Katz, M. E., Howarth, P. M., Yong, W. K., Riffkin, G. G., Depiazzi, L. J. & Rood, J. I. (1991). Identification of three gene regions associated with virulence in *Dichelobacter nodosus*, the causative agent of ovine footrot. *J Gen Microbiol* **137**, 2117–2124.

Katz, M. E., Strugnell, R. A. & Rood, J. I. (1992). Molecular characterization of a genomic region associated with virulence in *Dichelobacter nodosus*. *Infect Immun* **60**, 4586–4592.

Katz, M. E., Wright, C. L., Gartside, T. S., Cheetham, B. F., Doidge, C. V., Moses, E. K. & Rood, J. I. (1994). Genetic organization of the duplicated *vap* region of the *Dichelobacter nodosus* genome. *J Bacteriol* **176**, 2663–2669.

Kovach, M. E. & Peterson, K. M. (1994). Identification of a *Vibrio cholerae* gene with homology to the integrase family of site-specific recombinases. *Infect Immun* **62**, 3289–3298.

Krause, M. & Guiney, D. G. (1991). Identification of a multimer resolution system involved in stabilization of the *Salmonella dublin* virulence plasmid pSDL2. *J Bacteriol* **173**, 5754–5762.

Liu, Y., Chatterjee, A. & Chatterjee, A. K. (1995). Nucleotide sequence, organization, and expression of *rdgA* and *rdgB* genes that regulate pectin lyase production in the plant pathogenic bacterium *Erwinina carotovora* subsp. *carotovora* in response to DNA damage. *Mol Microbiol* **14**, 999–1010.

Matsui, H., Sano, Y., Ishihara, H. & Shinomiya, T. (1993). Regulation of pyocin genes in *Pseudomonas aeruginosa* by positive (*prtN*) and negative (*prtR*) regulatory genes. *J Bacteriol* **175**, 1257–1263.

Ogawa, T., Ogawa, H. & Tomizawa, J. (1988). Organization of the

early region of bacteriophage phi80. Genes and proteins. *J Mol Biol* **202**, 537–550.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.

Pullinger, G. D. & Lax, A. J. (1992). A *Salmonella dublin* virulence plasmid locus that affects bacterial growth under nutrient-limited conditions. *Mol Microbiol* **6**, 1631–1643.

Regnier, P., Grunberg-Manago, M. & Portier, C. (1987). Nucleotide sequence of the *pnp* gene of *Escherichia coli* encoding polynucleotide phosphorylase. Homology of the primary structure of the protein with the RNA-binding domain of ribosomal protein S1. *J Biol Chem* **262**, 63–68.

Rioux, C. R., Friedrich, M. J. & Kadner, R. J. (1990). Genes on the 90-kilobase plasmid of *Salmonella typhimurium* confer low-affinity cobalamin transport: relationship to fimbria biosynthesis genes. *J Bacteriol* **172**, 6217–6222.

Rood, J. I., Howart, P. A., Haring, V., Yong, W. K., Liu, D., Palmer, M. A., Pitman, D. R., Links, I., Stewart, D. J. & Vaughan, J. A. (1996). Comparison of gene probe and conventional methods for the diagnosis of ovine footrot. *Vet Microbiol* **52**, 127–142.

Salyers, A. A., Shoemaker, N. B., Stevens, A. M. & Li, L.-Y. (1995). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* **59**, 579–590.

Shine, J. & Dalgarno, L. (1974). The 3′ terminal sequence of *E. coli* 16S ribosomal RNA complementary to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA* **71**, 1342–1346.

Stevens, A. M., Shoemaker, N. B., Li, L. Y. & Salyers, A. A. (1993). Tetracycline regulation of genes on *Bacteroides* conjugative transposons. *J Bacteriol* **175**, 6134–6141.

Sun, J., Inouye, M. & Inouye, S. (1991). Association of a retroelement with a P4-like cryptic prophage (retronphage phiR73) integrated into the selenocystyl tRNA gene of *E. coli*. *J Bacteriol* **173**, 4171–4181.

Tian, Q. B., Ohnishi, M., Tabuchi, A. & Terawaki, Y. (1996). A new plasmid-encoded killer gene system: cloning, sequencing, and analyzing *hig* locus of plasmid Rts1. *Biochem Biophys Res Commun* **220**, 280–284.

Received 23 July 1996; revised 1 October 1996; accepted 4 October 1996.