CHAPTER FOUR:

Investigation of the *vap* regions in the benign strain C305.

4.1) Introduction

The *vap* region of *D. nodosus* are present in almost all virulent strains of the bacterium, and absent from most benign strains (Katz *et al.*, 1991, 1992; Rood *et al.*, 1996). The *vap* region is believed to have arisen by the integration, mediated by the site-specific recombinase gene *intA* (Figure 4.1; Cheetham *et al.*, 1995), of a genetic element into the *D. nodosus* genome. Another putative genetic element, the *intB* element, appears to share the same integration site (Figure 4.1). The *intB* element (Chapter 3), which is adjacent to *vap* region 3 in the virulent strain A198, is present in strain C305 and is adjacent to the *vap* regions in all *D. nodosus* strains so far investigated (Bloomfield *et al.*, 1997). Southern blot analysis has indicated there are two copies of *intB* in strain C305, as there are in strain A198. However, no PCR product was obtained from C305 when the *intB* gene was amplified (Chapter 3) suggesting that both copies of *intB* in strain C305 are partial copies.

The nature of the *vap* genetic element is presently unclear. A circular DNA molecule, which consists of *vap* region 1 and an insertion sequence, IS*1253*

(Billington *et al.*, 1996a), has been detected in *D. nodosus* strain AC3577. This has been proposed as the progenitor of the chromosomal *vap* regions (Billington *et al.*, 1996a). However, IS*1253* is not found in association with the *vap* regions in strain A198, but is located elsewhere in the A198 genome, near the *omp* gene cluster (Moses *et al.*, 1995). IS*1253* is also not virulence-associated, unlike the *vap* regions, being absent from some virulent strains and present in many benign strains (Billington *et al.*, 1996a).

Analysis of the sequences flanking *vap* regions 1 and 2 indicates that these copies of the *vap* regions have arisen by separate integration events into the A198 genome rather than by chromosomal duplications (Bloomfield *et al.*, 1997). *Vap* region 3, of strain A198, may also have arisen by a separate integration event (Bloomfield *et al.*, 1997), followed by the loss of many of the genes found in *vap* regions 1 and 2 (Figure 4.1) as *vap* region 3 does not contain *intA*, *vapA*, *vapB* or *vapC*. Some of the genes identified within *vap* region 3 are quite divergent from those of *vap* region 1 or 2 indicating a long evolutionary separation. Hybridisation studies have shown that the genes of *vap* regions 1 and 2 have high homology, indicating a recent evolutionary separation.

The sequence analysis of *vap* regions 1 and 3 has allowed exact comparisons of the genes from the two regions. The *vapG'* product (region 3) has 71% a.a. identity to VapG (region 1), VapE' has 62.6% a.a. identity to VapE and VapA' has 57.3% a.a. identity to VapA and 29.6% a.a. identity to VapA". However, *vapD* from region 3 is identical to *vapD* from *vap* regions 1 and 2. The *vapF* gene, in *vap* region 3 (Figure 4.1), is not present in *vap* regions 1 or 2. The *vapE'* gene, from region 3, is also found in one strain which lacks a copy of *vapE*, from *vap* region 1 (Bloomfield *et al.*, 1997). These results support the hypothesis that *vap* region 3 may be derived from the

integration of a separate *vap* element, rather than from a partial duplication of *vap* regions 1 and 2.

D. nodosus strain C305 does not contain *vapA*, *B*, *C* or *D* (Katz *et al.*, 1992) which indicates that the *vap* region is not present in C305. However, copies of the 102 b.p. repeats, spread throughout the *vap* regions of strain A198 (Figure 4.1; Cheetham *et al.*, 1995) are present in strain C305 (Katz *et al.*, 1994). Hybridisation analysis (Bloomfield *et al.*, 1992) has shown that a small segment of DNA between *vap* regions 1 and 3 of A198, is also present in strain C305 (Katz *et al.*, 1994; Cheetham *et al.*, 1995).

The *vap* regions of *D. nodosus* strain A198 are believed to have arisen by integrations of the *vap* element into a tRNA gene, located between the *ask* and *intB* genes (Figure 4.1). To investigate this hypothesis, I prepared a library of genomic DNA from the benign strain C305, which lacks *vapA-D*, and isolated a recombinant λ clone which hybridises to both *ask* and *intB*, sequences flanking the *vap* regions in strain A198. These sequences were found to be separated by approximately 6 kb in strain C305. This intervening sequence contains genes which may be part of a new genetic element, remnants of the *vap* region, and the insertion sequences IS*1253*.

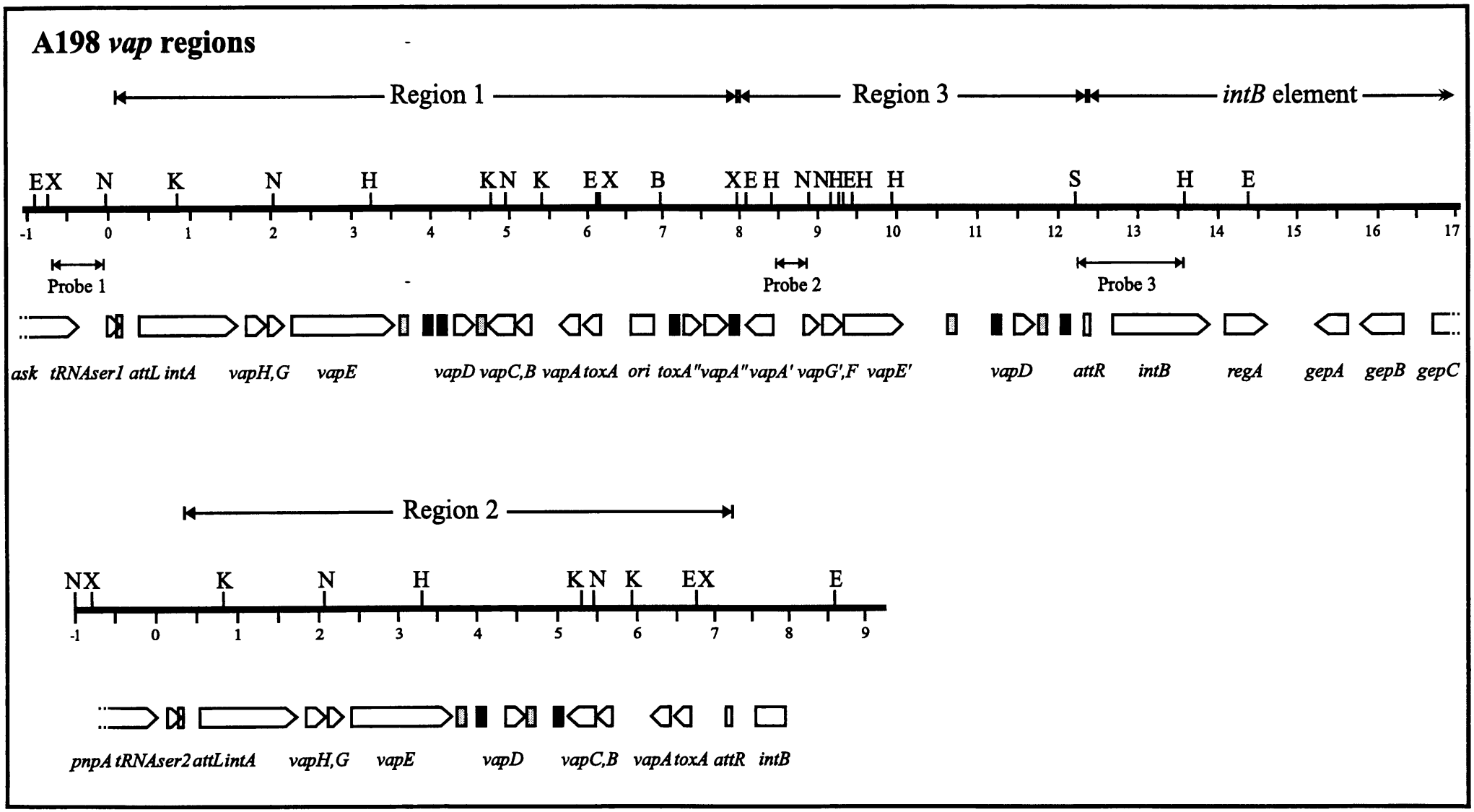

Figure 4.1: Restriction map of the A198 *vap* regions including the *intB* element. The size of this region, in kilobases, is indicated on the map. The location, size and orientation of the genes is indicated by the open arrows. *Att* sites are shown by small open boxes, 103 b.p. repeats are represented by lightly shaded boxes and 102 b.p. repeats by dark shaded boxes. The location of the three probes is indicated below the map. Restriction enzyme sites are indicated by the single letters, B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII, K = *Kpn*I, N = *Nru*I, S = *Sac*I, X = *Xho*I.

4.2) Results

4.2.1) Construction and screening of a C305 λ library.

A library of genomic DNA from the benign strain C305 of *D. nodosus* was constructed in bacteriophage λ. C305 genomic DNA was partial digested with the restriction endonuclease *Sau*3AI to yield DNA fragments of 10-20 kb in size. The fragments were then ligated to the λ arms and packaged into phage heads which were used to infect *E. coli*, to propagate the library. The initial phage titre of the C305 library was 3.5 x 10^5 pfu/ml. This is considered quite low but was found to be sufficient for screening the small *D. nodosus* genome of approximate size 1,800 kb. The library was screened with three probes from strain A198: probe 1, containing part of the *ask* gene adjacent to the left-hand end of *vap* region 1, probe 2, a sequence from *vap* region 3 between *vap* region 1 and 3, which has been shown previously to be present in strain C305 (Cheetham *et al.*, 1995) and probe 3, containing most of the *intB* gene adjacent to the right-hand end of *vap* region 3 (Figure 4.1).

Lambda clones, λGB321 and λGB210, which hybridised to all three probes were isolated from the C305 genomic library (Figure 4.2). Probe 1 was found to hybridise approximately 3.6 kb from sequences which hybridised to probe 2, while probe 3 hybridised faintly a further 1.9 kb downstream. Restriction maps were constructed of a number of overlapping λ clones spanning the sequences flanking *vap* regions 1 and 3 in strain C305 (Figure 4.2). Lambda clone λGB100 hybridised to probes 1 and 2, λGB200 hybridised to probes 2 and 3 while λGB210 and λGB321 hybridised to all three probes, as previously mentioned. In Figure 4.2 the maps of C305 λ clones λGB100, λGB210 and λGB321 have been partially aligned with the A198 λ clone

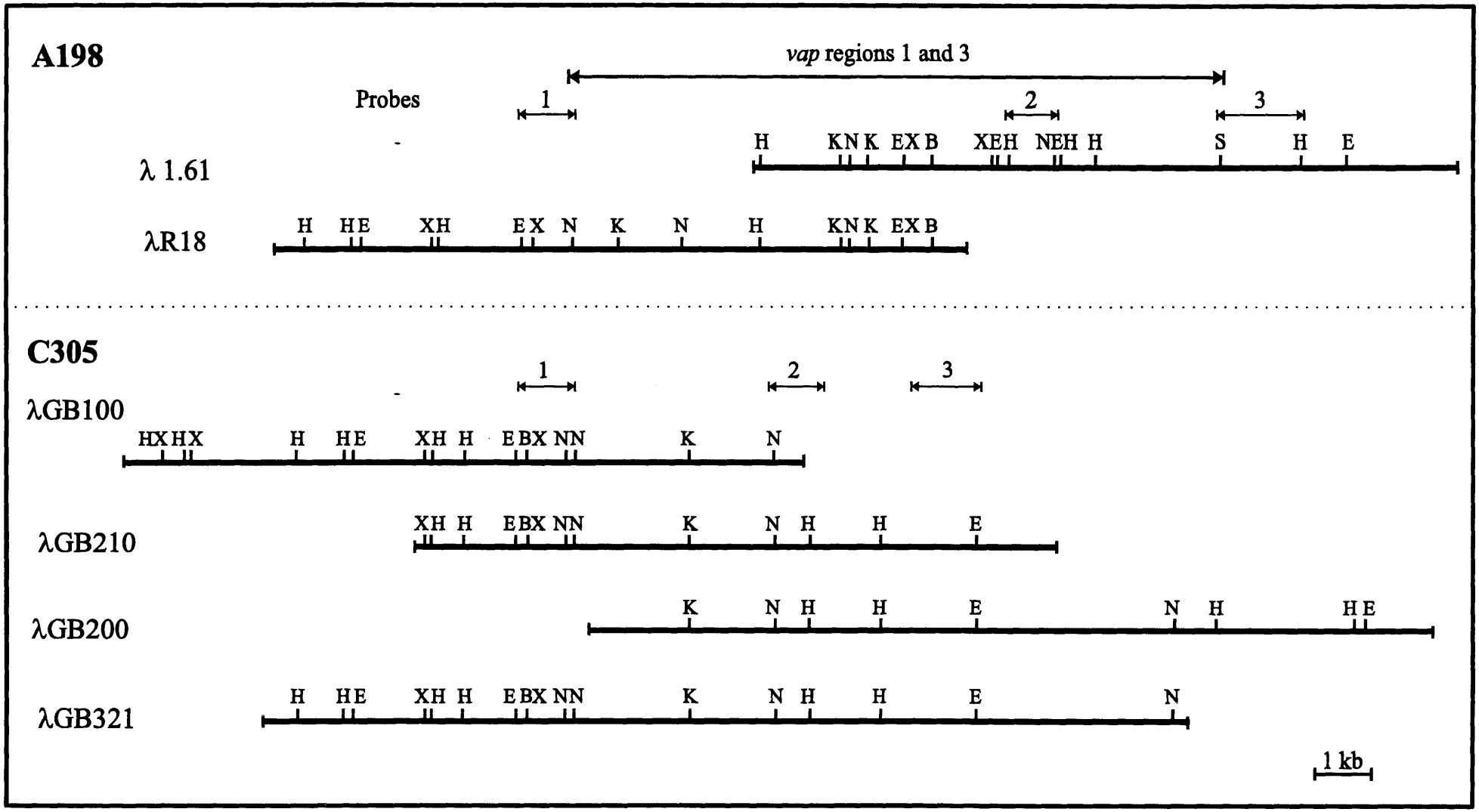

Figure 4.2: Restriction maps of the λ clones from *D. nodosus* strain A198 (Katz *et al.*, 1994), containing *vap* regions 1 and 3, and strain C305, which hybridise to probes 1, 2 and 3 from A198. The C305 clones are partially aligned, at the left-hand end, to the A198 λ clones. Restriction enzyme sites are indicated by the letters: B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII, K = *Kpn*I, N = *Nru*I, S = *Sac*I and X = *Xho*I.

λR18 (Katz *et al.*, 1994), which contains the left-hand end of *vap* region 1. The restriction maps from A198 and C305 are similar up to the sequence of probe 1, from this point on the maps diverge, due to the presence of the *vap* regions (Figure 4.2).

4.2.2) Sequencing of the C305 region which hybridised to probes 1, 2 and 3 from strain A198.

It was expected that the flanking sequences from a strain which contains the *vap* regions, A198, would be adjacent in a strain which does not contain these sequences, C305. However, this was not the case. The hybridisation of the three probes flanking *vap* region 1 and 3 of A198 (Figure 4.1) is spread over 6.5 kb in strain C305. To determine the organization of the intervening DNA, between the hybridisation of probes 1, 2 and 3 in C305, an 8 kb region spanning these sites was sequenced. A series of subclones were constructed from the restriction enzyme sites present in this sequence of DNA such that overlapping sequence could be obtained from each strand of the DNA (Figure 4.3). A number of small regions within this DNA sequence lacked appropriate restriction enzyme sites for subcloning. Thus, eight oligonucleotides were constructed and used as sequencing primers to complete the sequence of the 8 kb region of the C305 genome (Figure 4.3).

Within the 8 kb region of DNA from C305, spanning the hybridisation of probes 1, 2 and 3, a number of open reading frames have been identified (Figure 4.4). Figure 4.4 indicates the position and relative size of the potential genes identified in C305 as well as *vap* regions 1 and 3 of the virulent strain A198 and the location of probes 1, 2 and 3 in both A198 and C305. ORFs identified as potential genes are those which encode puative proteins greater than 50 a.a. in length and are preceded by a Shine-Dalgarno sequence. The *askA/B* genes are so named due to similarity to aspartokinase

genes and a *tRNA-ser* gene has been identified by its similarity to other tRNA genes. The *intC* gene is an integrase gene, *orf242* and *orf171* have no similarity to previously identified genes and *vapG"* and *vapH"* are similar to *vapG* and *vapH* from the A198 *vap* regions. Two ORFs, *orf374* and *orf108*, are similar to the transposase genes of insertion sequence IS*1253* and are adjacent to a small section of the *intB* gene and the N-terminus of the *regA* gene (Figure 4.4).

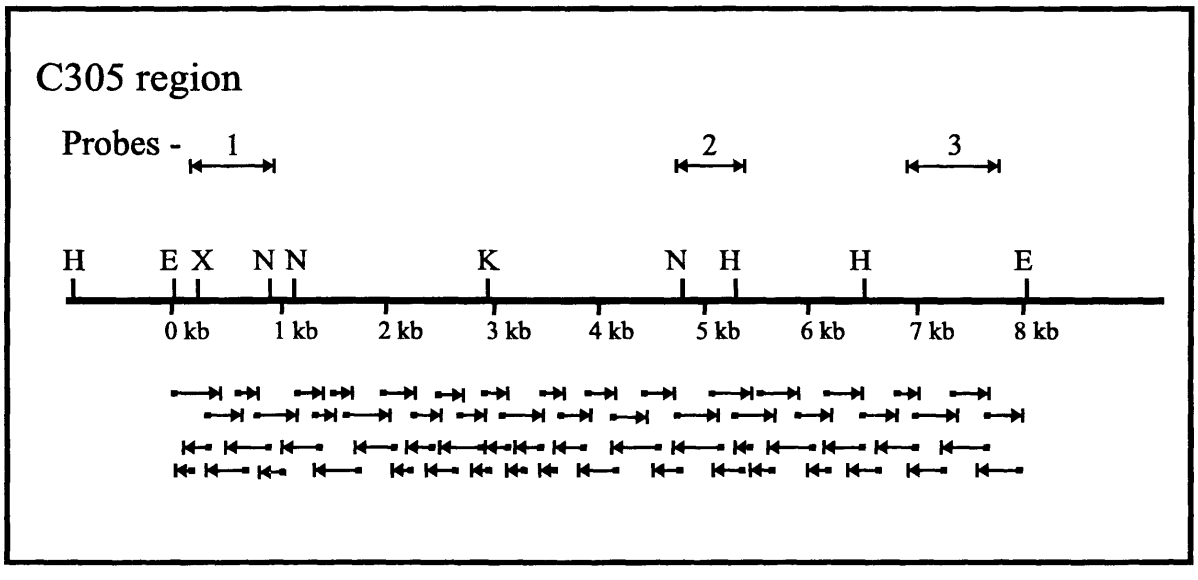

Figure 4.3: Sequencing strategy for the 8 kb region of C305 genomic DNA spanning the hybridisation of probes 1, 2 and 3 from the *vap* flanking sequences of strain A198. Restriction enzyme sites are indicated by the letters, B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII, K= *Kpn*I, N = *Nru*I and X = *Xho*I.

4.2.3) Identification of aspartokinase and serine tRNA genes

The ORFs, designated *askA/B*, with similarity to the aspartokinase α and β subunits, have been partially sequenced in the C305 λ clone (Figure 4.4). Aspartokinase is the enzyme which catalyzes the first step in the conversion of L-aspartate to L-lysine, L-methionine and L-threonine by reduction of the β-carboxyl group (Zubay, 1989). The functional aspartokinase is a tetramer of the α and β

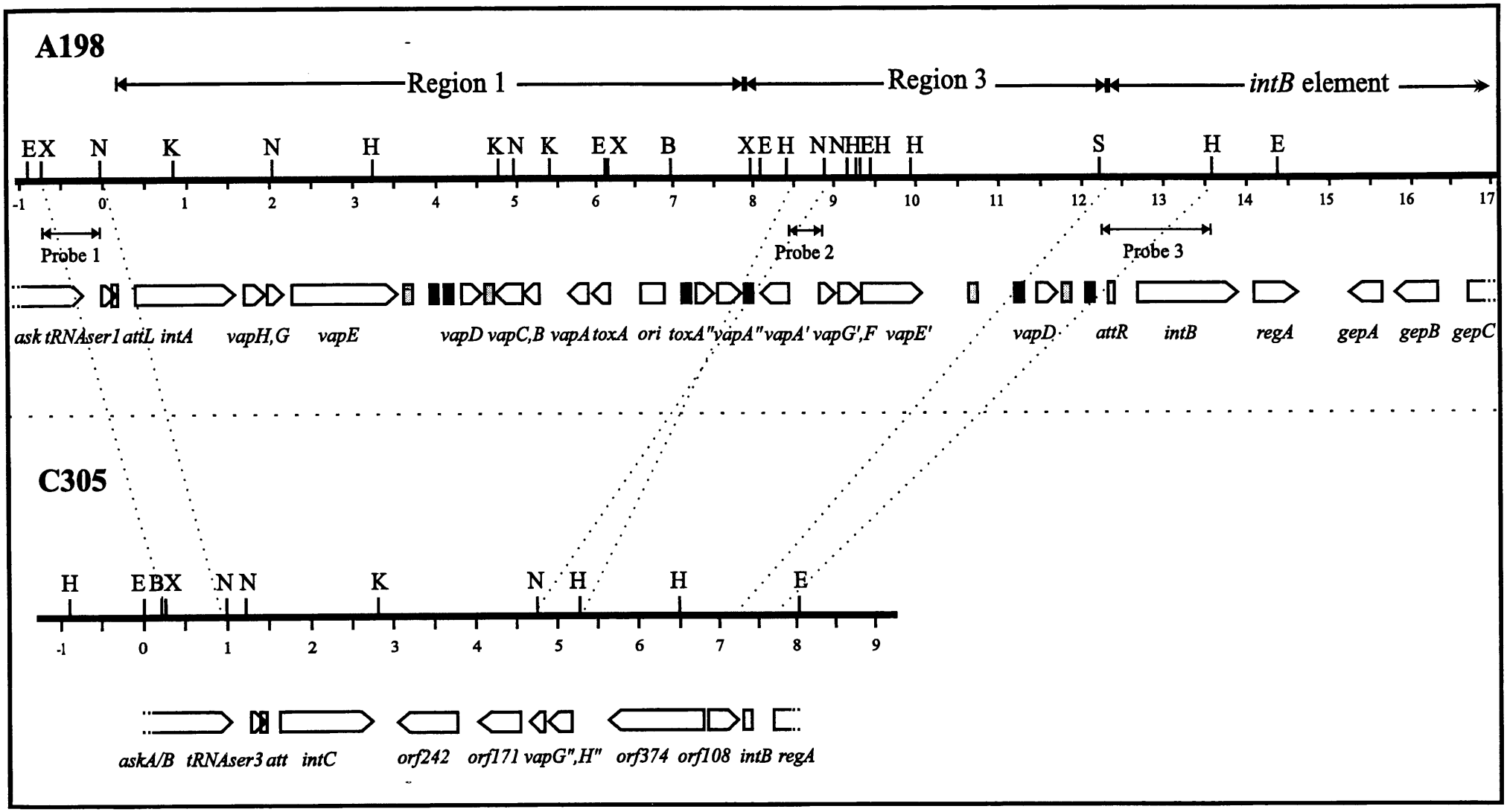

Figure 4.4: Restriction map of *vap* regions 1 and 3 from *D. nodosus* strain A198 and the corresponding clone from C305. The size and orientation of the genes is indicated by the open arrows and the sequences of C305 which hybridise to clones 1, 2 and 3 from A198 are indicated by the dotted lines. Restriction enzyme sites are indicated by the letters, B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII, K = *Kpn*I, N = *Nru*I, S = *Sac*I, X = *Xho*I.

subunits, $\alpha_2\beta_2$ (Moir and Paulus, 1977). The aspartokinase β-subunit is encoded entirely within the aspartokinase α-subunit and is identical to the carboxyl end of the α-subunit (Figure 4.5).

The 301 amino acids of AskA/B from *D. nodosus* have 43.2% identity with the *lysCα/lysCβ* gene products from *Corynebacterium glutamicum* (Kalinowski *et al.*, 1990), 42.7% amino acid identity with Askα/Askβ from *Mycobacterium smegmatis* (Cirillo *et al.*, 1994) and 41% amino acid identity with AskIIα/AskIIβ from *Bacillus subtilis* (Chen *et al.*, 1987). The *askA* gene is expected to encode a protein of approximately 420 amino acids, therefore, only three quarters of the gene has been sequenced in *D. nodosus* strain C305. The entire *askB* gene, the carboxyl end of the *askA* gene, has been sequenced (Figure 4.5).

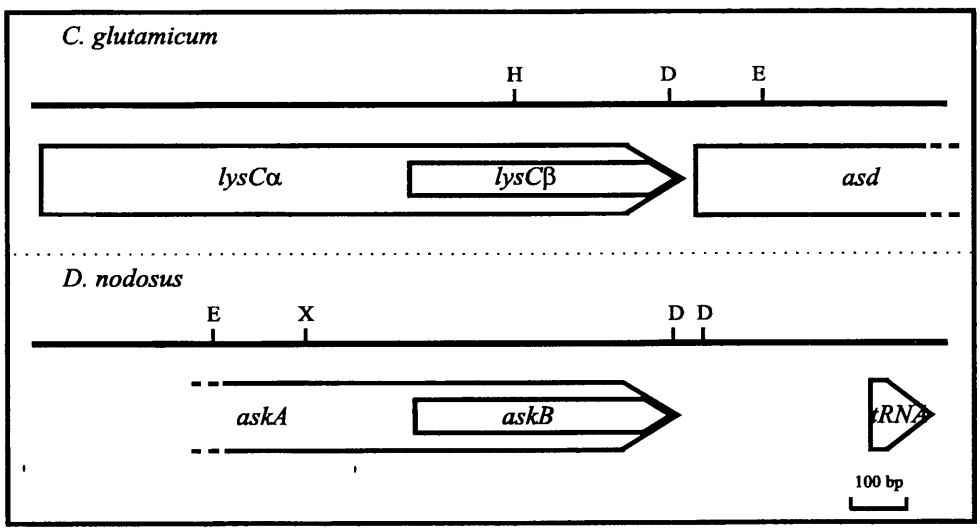

Figure 4.5: Restriction map of the *C. glutamicum* (Kalinowski *et al.*, 1990) and *D. nodosus* genomic regions encoding the aspartokinase genes. The ORFs are indicated by the open arrows. The *lysCα/lysCβ* and *askA/askB* genes encode the α and β-subunits of aspartokinase. The *asd* gene encodes the aspartate β-semialdehyde dehydrogenase and tRNA is the *tRNA-ser* gene. The restriction enzyme sites are: D = *Dra*I, E = *Eco*RI, H = *Hin*dIII and X = *Xho*I.

The β-subunit of aspartokinase has its own start codon and ribosome binding site within the coding region of the α-subunit. The termination codon is the same for each subunit (Figure 4.5). The *Corynebacterium* and *Mycobacterium* aspartokinase genes are followed by the gene for aspartate β-semialdehyde dehydrogenase, *asd*, whereas the *D. nodosus ask* genes have a *tRNA-ser* gene downstream.

A small piece of DNA sequence from one DNA strand of the *ask* genes in strain A198, upstream of *vap* region 1, has been determined (Whittle and Cheetham, unpublished). Comparison of the DNA sequence from the *ask* genes of strain C305 with this partial sequence from strain A198 gives 97.4% identity over 650 nucleotides.

Exactly 357 b.p. downstream of the end of the *askA/B* genes in C305 lies a 92 b.p. serine tRNA gene, *tRNA-ser3*, with 97.8% nucleotide identity to *tRNA-ser1* from the left-hand end of *vap* region 1, *D. nodosus* strain A198 (Figure 4.4; Cheetham *et al.*, 1995). The two tRNA genes possess the same anti-codon, GCU, and are most likely corresponding copies in each strain. The C305 *tRNA-ser3* also has 79.4% nucleotide identity to $tRNA_3^{Ser}$ of *E. coli* (Ish-Horowicz and Clark, 1973; Yamada and Ishikuri, 1973) and 78.3% identity to the second *D. nodosus* serine tRNA, *tRNA-ser2*, identified adjacent to the left-hand end of *vap* region 2 in *D. nodosus* strain A198 (Bloomfield *et al.*, 1997). The *D. nodosus tRNA-ser2* has a GGA anticodon indicating that it encodes a different tRNA molecule, not a copy of *tRNA-ser1*, from the left-hand end of *vap* region 1 (Bloomfield *et al.*, 1997).

Figure 4.6 shows the putative tRNA molecule from C305. The anti-codon is underlined while the bold bases indicate the differences between this C305 *tRNA-ser3* and *tRNA-ser1* from A198. The DHU loop of the tRNA molecule is so named

because it typically contains multiple dihydrouracil residues while the TψC loop acquired its name from the sequence ribothymine-pseudouracil-cytosine.

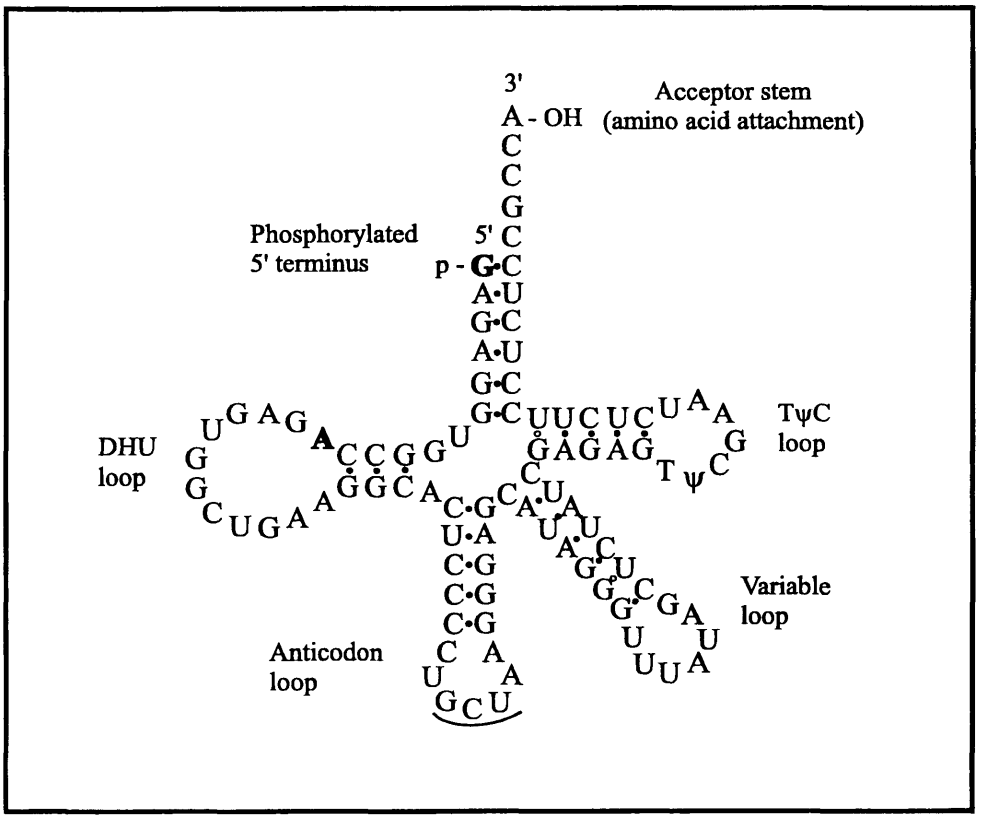

Figure 4.6: The putative tRNA-ser molecule from C305. The bases in bold indicate the different residues between the tRNA from A198, *tRNA-ser1* upstream of *vap* region 1, and this C305 *tRNA-ser3*. The anticodon is underlined.

4.2.4) Identification of an integrase gene, *intC*.

Within the end of the *tRNA-ser* gene in C305 there is an exact copy of the *attL* (left-hand attachment) site from A198 *vap* region 1 (Cheetham *et al.*, 1995). Analysis of sequences downstream indicate there is no *attR* (right-hand attachment) site within this C305 region. An ORF, adjacent to the C305 *tRNA-ser3* (Figure 4.4) has been termed *intC* due to the similarity to a number of integrase genes, including *intA* from *vap* region 1 of *D. nodosus* strain A198 (Cheetham *et al.*, 1995), and the integrases of retronphage ϕR73 of *E. coli* (Sun *et al.*, 1991) and bacteriophage Sf6 of *S. flexneri* (Clark *et al.*, 1991). There is 54.2% a.a. identity with IntA, 40.4% a.a. identity with

ϕR73 Int and 34.9% a.a. identity with Sf6 Int. Overall, 21.6% of amino acid residues are identical in all four proteins (Figure 4.7), including Arg^{240} , His^{326}, Arg^{329} and Tyr^{363} (numbered for their position in IntC), which are highly conserved in the integrase genes (Argos *et al.*, 1986; Abremski and Hoess, 1992). IntC also has 33.9% amino acid identity with the *D. nodosus intB* gene product (Chapter 3), while IntB is only 31.0% identical to IntA.

```
v          v          v          v          v          v
IntA       MLGNHFSNIMRLTDKIIQTAKKPSRGYLRLTDGDGLTLKITDRGSYFWNFRYYINGRERN   60
IntC       ---------MKLRELIVRQAKLP-KKAKKLFDGGGLFLYLTPSG-KYWYYRYRFAGKDKV   49
PhiR73Int  ---------MPLTDIQIKRAK-PQNKPYTLNDGQGLSLLINPDGSKGWRFRFRFAGKARL   50
SF6Int     ----------MLTVKQIEAAK-PKEKPYRLLDGNGLYLYVPVSGKKVWQLRYKIDGKEKI   49
                     *    .  ** *      * ** ** * .   *   *  *.   *. .

IntA       MSLGRYPAMSIEAARGEALELRYHLRQGLDPLAERKKKQAAYLAEEKTQKETFEFVAREW  120
IntC       MPLGKYPHMNLKEARIAHIEAKIKLSNGCDPVEENNR---LKREQERNYNNSFEEIAKEW  106
PhiR73Int  MSFGSYNLVSLAEAREKRDTARKQVANGIDPVGERKAQK---LAQQLSTENSFETICREW  107
SF6Int     LTVGKYPLMTLQEARDKAWTARKDISVGIDPVKAKKAS---------SNNNSFSAIYKEW  100
           .  * *  . .  **      .  .  * **.                    .*  . .**

IntA       YQLKRPEWKNEKHAQQVITTLEQFVFPFIGKKPISRISPPELLKVMEKLYDR--PETASR  178
IntC       YQHVLPEWKNKKHAQQVINTLSQYAFPKIGHYPIDDVPPIELFNLLHEIRDK--AETASR  164
PhiR73Int  YTNKADRWT-VAYREEIIKTFEQDVFPFIGKRPISEIKPLELLEVLRRIEKRGALEKTRK  166
SF6Int     YDHKKQVWS-VGYATELAKMFDDDILPIIGGLEIQDIEPMQLLEVIRRFEDRGAMERANK  159
           *      *        ..          * **   *  . * .*  ..     .   * . .

IntA       VKQRIRAVFDYAIQTDRLSRNPAIALPKIFR---STIAKQPSLPARELGIFFKQLDTYG-  234
IntC       LKQRIKAVFDFAIQTGRAKTNPALSAPKIIRNNDNKVKHHPALPESRIKEFYQRLENYP-  223
PhiR73Int  VRQRCGEVYRYAIITGRAEYNPAPDLAIALAVP--KQKHHPFLSAEELPHFIQDLEAYTG  224
SF6Int     ARRRCGEVFRYAIVTGRAKYNPAPDLADAMKGY--RKKNFPFLPADQIPAFNKALATFSG  217
            ..*   *. .** * *   ***                 * *    .  * . *  .
                      ↓
IntA       NPKTALALRLLILTMTRSGELRFGQWQELQGNE--WHIPAERMKMNRPHIVPLSDWALEI  292
IntC       NRTTQLALQFLILTFVRVGELRQGEWTEIKGNE--WHIPAEKMKMKRPHVVPLSSWALRI  281
PhiR73Int  SIITKNATKIVMLTGVRTQEMRFAAWDEVDLERGLWEIPAERMKMRRPHIIPLSTQVVDL  284
SF6Int     SIVSLIATKVLRYTALRTKELRSMQWKNVDFENRIITIDASVMKGRKIHVVPMSDQVIEL  277
              .  * . .  *  *  *.*   *  .         * *  **  . *..*.*   . .
                                                      ↓ ↓
IntA       LDELRSLRRNNSPYFVTG--NRNTPLSDTTLSLAMKRLGYAGRAVPHGFRATASTAMNES  350
IntC       LEELKIINKYDCPYFIVG--NRNQQISDNTLSVAMKRLGYQNIATPHGFRAMASTILNES  339
PhiR73Int  FNQLKPITR-HYPYIFIGRNNRNKPISKESVSQVIELLGYKGRATGHGFRHTMSTILHEQ  343
SF6Int     LTTLSSITKPVSEFVFAGRNDKKKPICENAVLLVIKQIGYEGLESGHGFRHEFSTIMNEH  337
              *  . .'   .   *   .  .. ..  .  .** .    ****   ** ..*
                              ↓
IntA       GLWNPDAIERQLAHTDENAVRAAYNRAEYLEERHRMMQWWADYIRKKIRVS  401
IntC       GLWNPDAIERQLSHVDRNAVRAAYNRAEYLEERHRMMEWYSDYIKSLINE-  389
PhiR73Int  -GFDSAWIELQLAHVDKNSIRGTYNHAQYLERRREMLQWYSNYLLK-----  388
SF6Int     -EWPADAIEVQLAHANGGSVRGIYNHAQYLDKRREMMQWWADWLDEKVR--  385
                 ** **.*   ...*  **.*.**. *. *..* .  .
```

Figure 4.7: Clustal W alignment (Thompson *et al.*, 1994) of IntA, IntC, ϕR73 integrase and Sf6 integrase. The identical amino acids are indicated with an asterisk and the conservative substitution are indicated by a period. The arrows indicate the amino acid residues conserved in the site-specific recombinases.

IntC is most closely related to IntA, and may therefore have evolved from *intA* from the *vap* regions. However, if this is the case, the separation must have occurred a long time ago. Alternatively *intC* may be part of a novel genetic element, unrelated to the *vap* element. Therefore, the sequences within the C305 genome which lie between *tRNA-ser3* and the remnants of the *intB* gene (Figure 4.4) have been collectively termed the *intC* element.

Located downstream of *intC* are two ORFs, *orf242* and *orf171*, named for the number of amino acids encoded by each. These two potential genes have no similarity to *vap* genes from strain A198 or to any other sequences registered in the databases.

4.2.5) Genes related to *vapG* and *vapH* are present in strain C305

Probe 2 contained part of the *vapG'* gene from *vap* region 3 of strain A198 (Figure 4.4). A gene with high similarity to *vapG'*, designated *vapG"*, is present in the C305 genome. There is also a gene similar to *vapH*, *vapH"*, in C305 close to *vapG"* (Figure 4.4). The arrangement of *vapG"* and *H"* in strain C305 is similar to the arrangement of *vapG* and *H* in strain A198, but the orientation of the two genes, with respect to the integrase gene, is reversed in C305 (Figure 4.4). The deduced amino acid sequence of the C305 *vapG"*, 67 amino acids, has 79.1% identity with VapG' (region 3) and 55.2% amino acid identity with VapG (region 1) from strain A198. The 105 amino acids of VapH" have 54.2% identity with VapH from region 1 of strain A198. The level of amino acid identity between IntA and IntC, VapG and VapG", and VapH and VapH" is very similar i.e. approximately 55%. Also, upstream of the A198 *vapG'* gene (region 3) there are remnants of a *vapH*-like gene which would correspond to *vapH"* in C305. DNA from C305 spanning part of *vapG"* and

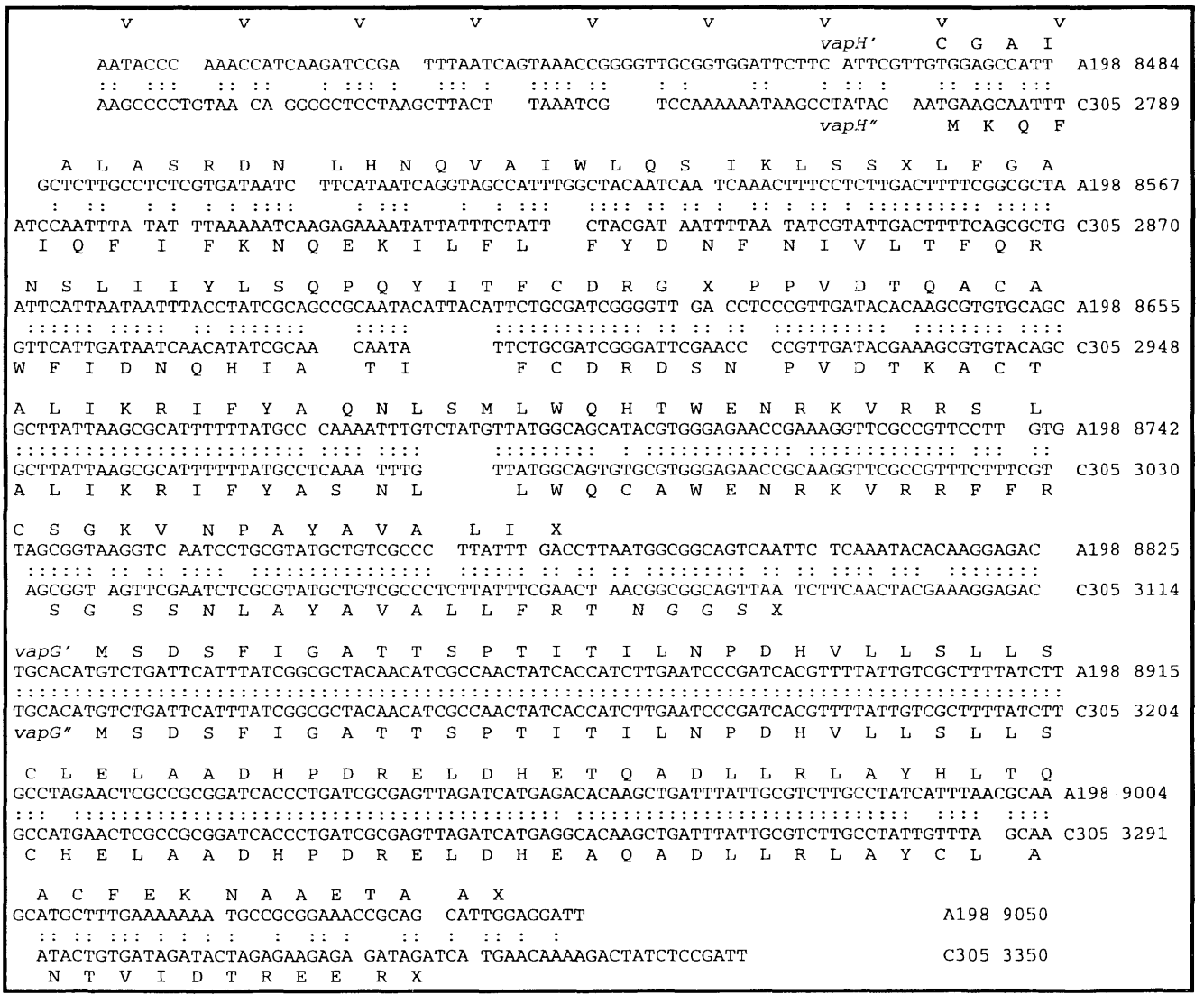

Figure 4.8: Alignment of sequences from C305, spanning *vapG"* and *vapH"*, and A198 (Genbank Accession Number DIHINTA), spanning the sequences prior to and including *vapG'*. Also shown are the amino acid sequences for each of the genes including *vapH'* from A198. The identical nt. are indicated by a semicolon and the "v" above the sequence marks every 10 nt.

vapH" has 84.1% identity over 490 b.p. to the A198 sequence from *vap* region 3, prior to and including part of *vapG'* (Figure 4.8). In Figure 4.8 the nucleotide numbering of the A198 sequence is from the zero position, upstream of *tRNA-ser1* (Figure 4.4). The C305 sequence is the complementary strand thus, the numbering starts from the *Eco*RI site within the *regA* gene (Figure 4.4). The sequence prior to *vapG'* in A198 has been translated in Figure 4.8, and labelled *vapH'*, to show the similarity to *vapH"* from C305. The finding that genes related to *vapG* and *vapH* are present in strain C305 raises the possibility that the *vap* element was originally integrated at this point in strain C305, and that subsequently most of the genes from the *vap* region were lost.

4.2.6) Identification of a copy of IS*1253* within this C305 element

The two open reading frames, *orf374* and *orf108* adjacent to *vapH"* in C305 (Figure 4.4), are the constituent genes of an insertion sequence identified in *D. nodosus* (Billington *et al.*, 1996a). This insertion sequence, termed IS*1253*, is a 1689 b.p. element consisting of two ORFs, *orf375* and *orf117*, with similarity to transposase genes from a number of unusual IS elements (Billington *et al.*, 1996a). This type of IS element does not produce direct repeats at its extremities when inserted into another DNA molecule. IS*1253* is located on the *vap* plasmid in *D. nodosus* strain AC3577 (Billington *et al.*, 1996a) and a copy has been identified within the outer membrane protein (*omp*) locus of strain A198 (Moses *et al.*, 1995) but IS*1253* is not associated with the *vap* regions in strain A198.

The DNA sequence of IS*1253* found in C305 has nucleotide identity of 98.3% over 1643 b.p. with IS*1253* from the *vap* plasmid (Billington *et al.*, 1996a) and consists of two similar ORFs, *orf374* and *orf108*. The predicted protein product of *orf374* has an amino acid identity of 98.1% with that of *orf375* from the *vap* plasmid.

Similarly, the predicted products of *orf108* and *orf117* show 93.6% amino acid identity. The putative transposase encoded by *orf374* has 36.6% amino acid identity with the transposase of IS*1341* from thermophilic bacterium PS3 (Murai *et al.*, 1995), 27.3% identity with the transposase of IS*891* of *Anabaena* spp (Bancroft and Wolk, 1989) and 26.8% identity with the transposase of IS*1136* from *Saccharopolyspora erythraea* (Donadio and Staver, 1993). The predicted *orf108* gene product has amino acid identity of 35.2% with the product of one of six ORFs encoded by IS*H1.8* of the *Halobacterium halobium* bacteriophage ϕH (Schnabel *et al.*, 1984) and 28.7% amino acid identity with the product of the ORF of the enterobacterial IS*200* element (Bisercic and Ochman, 1993). There is also similarity between IS*1253* and IS elements from *Synechocystis* spp (Kaneko *et al.*, 1996), *Sulfolobus solfataricus* (Sensen *et al.*, 1996) and *Bacillus thuringiensis* (Estruch *et al.*, 1996).

The IS*1253* from strain C305 appears to be a non-functional insertion sequence. The *orf374* gene is a pseudogene as it has been disrupted by two frame-shift mutations, the most interesting of which is the deletion of a GTAA repeat. Within the reading frame of *orf375* GTAA is repeated four times almost in tandem whereas in *orf374* the repeat is only present in three copies (Figure 4.9). These repeats have an unknown function.

Analysis of the ends of the C305 IS*1253* has shown the left-hand end to be almost identical to the left-hand end of the *vap* plasmid IS*1253*. The right-hand end of the C305 IS*1253*, however, has had a deletion of 46 nucleotides. Billington *et al.* (1996a) identified a 46 b.p. sequence, including a region of dyad symmetry, located at the left-hand end of IS*1253*, which is imperfectly repeated 22 b.p. from the right-hand end. The C305 IS*1253* has a deletion at the right-hand end of the extra 22 b.p. and 24 b.p. of the imperfect repeat including over half the region of dyad symmetry (Figure 4.10).

If the repeats are involved in the recognition sequence for the transposase (Billington *et al.*, 1996a) then the deletion of part of the right-hand sequence in C305 may have rendered this copy of IS*1253* non-functional therefore allowing the disruption of *orf374* to occur by genetic drift. Alternatively *orf374* may have become non-functional, by the deletion shown in Figure 4.9, and the deletions involving *intB* (Figure 4.4) have since removed the right-hand end of the C305 IS*1253*.

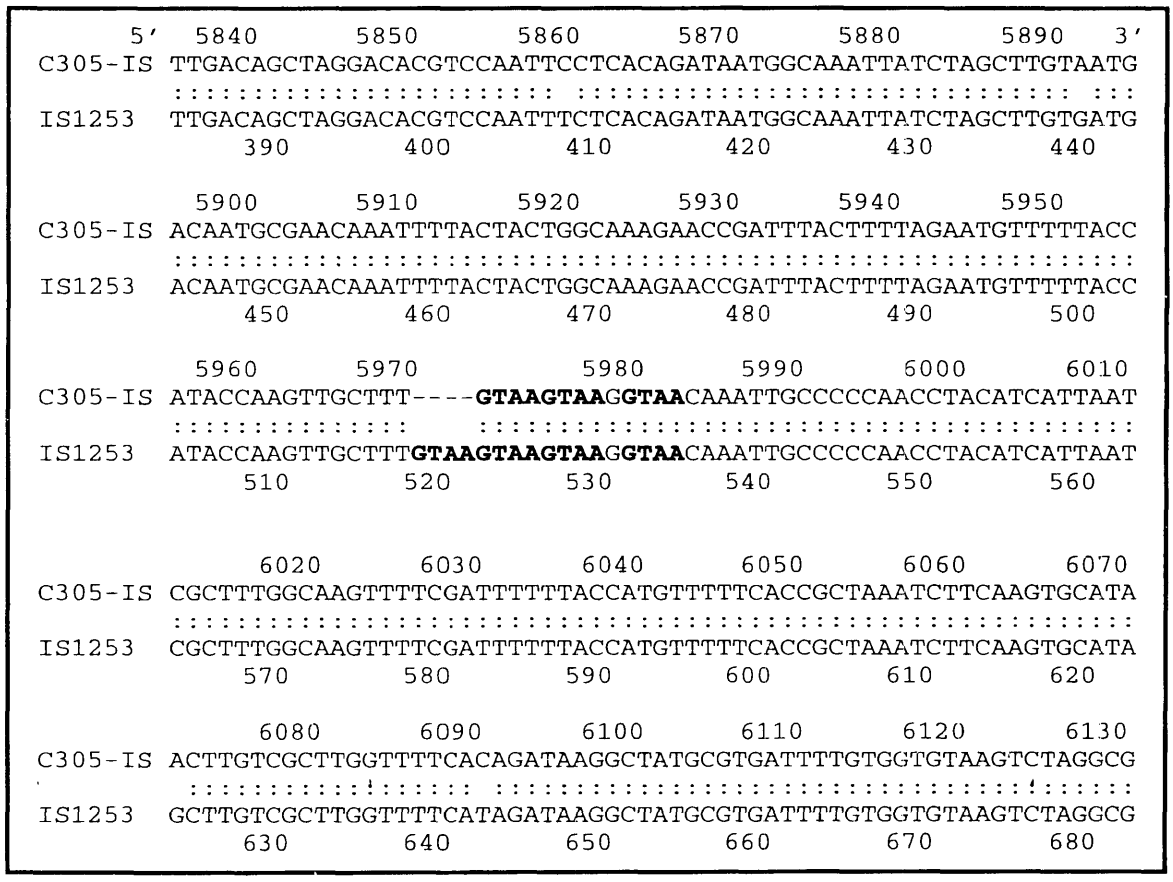

Figure 4.9: Comparison of part of the DNA sequence from the IS*1253* element, *orf374*, of C305 with part of the IS*1253* sequence, *orf375*, from the *vap* plasmid (Genbank Accession Number DNU34772). Nucleotides in bold indicate a repeat sequence of four nt., of which one repeat is deleted from the C305 sequence. Identical nucleotides between the two sequences are indicated by a semi-colon. The numbers above and below the sequence indicate the location of the similarity with respect to the initial sequencing point.

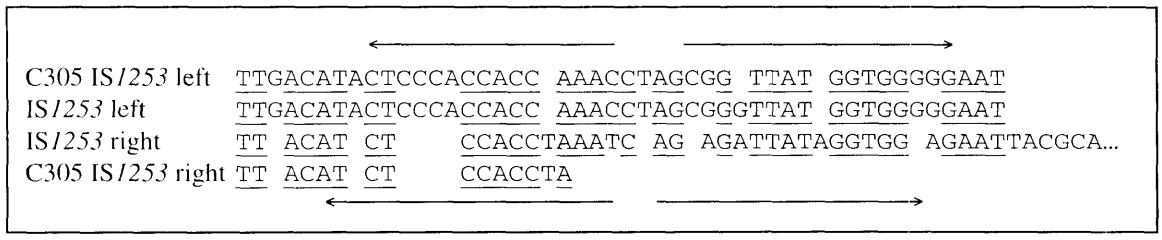

Figure 4.10: Alignment of the imperfect repeats located near either end of IS*1253* (Billington *et al.*, 1996a) with the analogous ends of the C305 IS*1253*. Identical nt. from the left and right-hand repeats are underlined and the regions of dyad symmetry, or inverted repeats, are indicated by the arrows. IS*1253* right contains an extra 22 b.p. after the repeat while C305 IS*1253* right is missing over half the repeat.

4.2.7) Disruption of the *intB* and *regA* genes.

Immediately following IS*1253* is 100 b.p. of the 1209 b.p. *intB* gene. The next base after the end of the deleted imperfect repeat from C305 IS*1253*, right-hand end (Figure 4.10), is the start of the similarity with a section of *intB*. The 100 b.p. of *intB* is from nucleotides 408 to 508 of the 1209 b.p. coding region, starting 23 b.p. downstream of the insertion which has disrupted the *intB* gene in A198 (Chapter 3, Figure 3.5). This sequence was sufficient for Probe 3 (Figure 4.4) to hybridise and give a faint signal when screening the C305 λ library. The *intB* gene has been completely disrupted in C305 with most of the gene deleted. The regulatory gene, *regA*, downstream of *intB* (Chapter 3) has also been affected. The *regA* gene, although disrupted by frame-shift mutations, appears to be present even though the start codon, Shine-Dalgarno and promoter sequences are all disrupted in the C305 strain (Figure 4.11).

Flanking the *vap* regions in A198 are attachment sites, presumably from the IntA-mediated insertion of the region into the bacterial genome. The 3' end of a serine tRNA gene forms the bacterial *att* site, which is a common site for insertion of temperate bacteriophages and integrative plasmids (Cheetham and Katz, 1995). The C305 strain appears to have a similar arrangement with a *tRNA-ser* gene, *attL* and

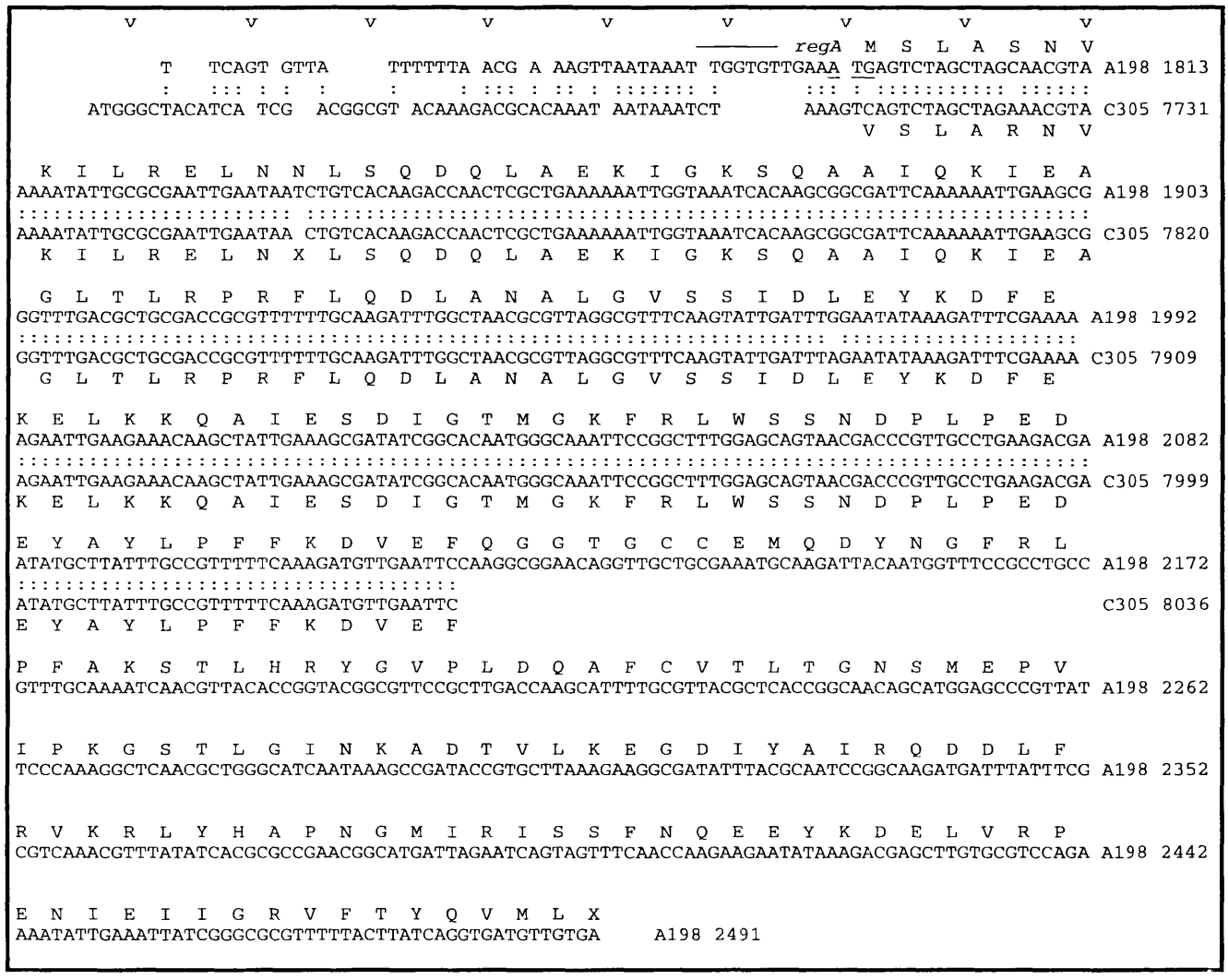

Figure 4.11: Alignment of the sequence from *regA*, right-hand end of the A198 *vap* region 3, with the *regA* sequence from C305 (C305 RegA is corrected for the frameshift). Identical nucleotides are indicated by a semicolon, the *regA* start codon is underlined and the Shine-Dalgarno sequence is indicated by the line above the sequence. The 'v' indicates every 10 nucleotides. The DNA is numbered from the start of each sequence, i.e. A198 sequence is numbered from the *Sac*I site within *vap* region 3 and the C305 sequence is numbered from the *Eco*RI site with the *askA* gene (Figure 4.4).

intC gene identified. This *intC* element also contains two genes of unknown function, *orf242* and *orf171*, genes similar to *vapG* and *vapH* from the *vap* regions and a copy of the insertion element IS*1253*. However, extensive rearrangements and deletions have occurred in C305 and have disrupted *intB*, *regA* and removed all trace of an *attR* which may have been associated with the *intC* element.

4.2.8) Analysis of a separate C305 λ clone

Probe 3 also hybridised strongly to another unrelated λ clone from the C305 library. A 3.3 kb DNA sequence from this λ clone containing and surrounding the sequences to which the probe hybridised has been determined. A series of subclones were constructed and sequenced so as to independently determine the sequence of each strand of the DNA (Figure 4.12). Not all the restriction enzyme sites used for subcloning are shown on Figure 4.12. This 3.3 kb *Hin*dIII fragment is the same fragment which was isolated from a library, in pUC18, of *Hin*dIII digested genomic

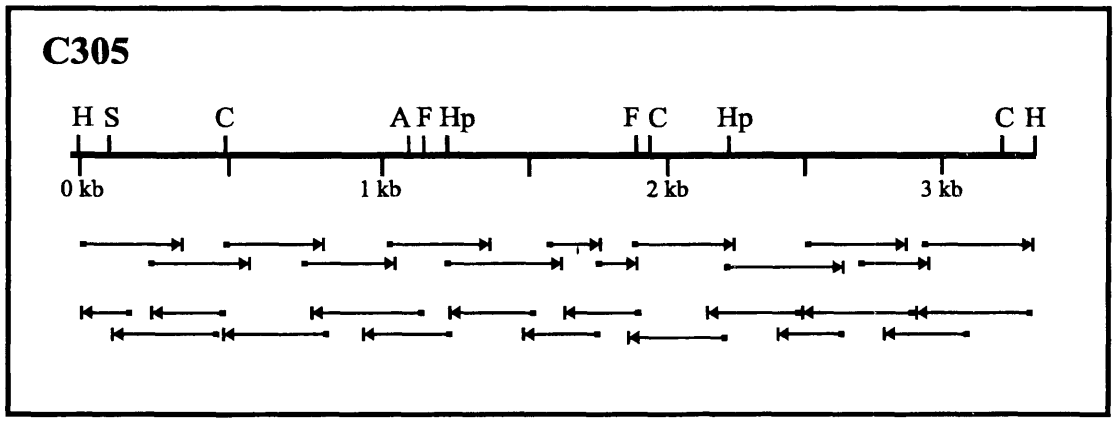

Figure 4.12: Sequencing strategy for the 3.3 kb *Hin*dIII subclone from the λ clone of C305 genomic DNA which hybridised to probe 3 from the right-hand flanking sequence of *vap* region 3 in strain A198. Restriction enzyme sites are indicated by the letters, A = *Acc*I, C = *Cla*I, F = *Fsp*I, H = *Hin*dIII, Hp = *Hpa*I and S = *Sac*I.

DNA from strain C305, and partially characterised by Katz *et al.* (1994). This subclone from the C305 library contains three copies of the 102 b.p. repeats found throughout the *vap* regions of strain A198, a partial copy of the *intB* gene and a gene of unknown function, *orf379* (Figure 4.13).

The 102 b.p. repeats from this C305 clone are similar to the two copies of the 102 b.p. repeats found at position 3925 of *vap* region 1 in strain A198 (Figure 4.12; Katz *et al.*, 1994; Cheetham *et al.*, 1995). A clone containing the two repeats from A198, between *vapE* and *vapD* in region 1 (Figure 4.12), hybridised, under high stringency, to the C305 copies of the 102 b.p. repeats (Katz *et al.*, 1994). Due to this hybridisation the repeats at position 3925 in the A198 *vap* region 1 (Cheetham *et al.*, 1995) were originally believed to be the left-hand limit of *vap* region 1 (Katz *et al.*, 1994). An interesting feature of the three copies of the 102 b.p. repeats in C305 is that the orientation is opposite to that of the A198 copies. This reverse orientation indicates that rearrangements have occurred as well as the extensive duplications and deletions. The 102 b.p. repeats have an unknown function but have been proposed to be involved in DNA rearrangements (Cheetham *et al.*, 1995).

The 3.3 kb *Hin*dIII fragment from the C305 genome was isolated by hybridisation to the *intB* probe, probe 3 (Figure 4.13). This same fragment was isolated from a different library by Katz *et al.* (1994) by hybridisation to a clone from the right-hand end of *vap* region 2 from strain A198. The right-hand end of *vap* region 2 in A198 has since been determined to contain a partial copy of the *intB* gene (Figure 1.5; Bloomfield *et al.*, 1997). The sequence of this 3.3 kb *Hin*dIII subclone from the genome of strain C305 has now been determined (Figure 4.12). Sequence comparisons have shown that this C305 copy of the *intB* gene is almost identical to the partial copy of *intB* adjacent to *vap* region 2 in strain A198 (Figure 4.14). This

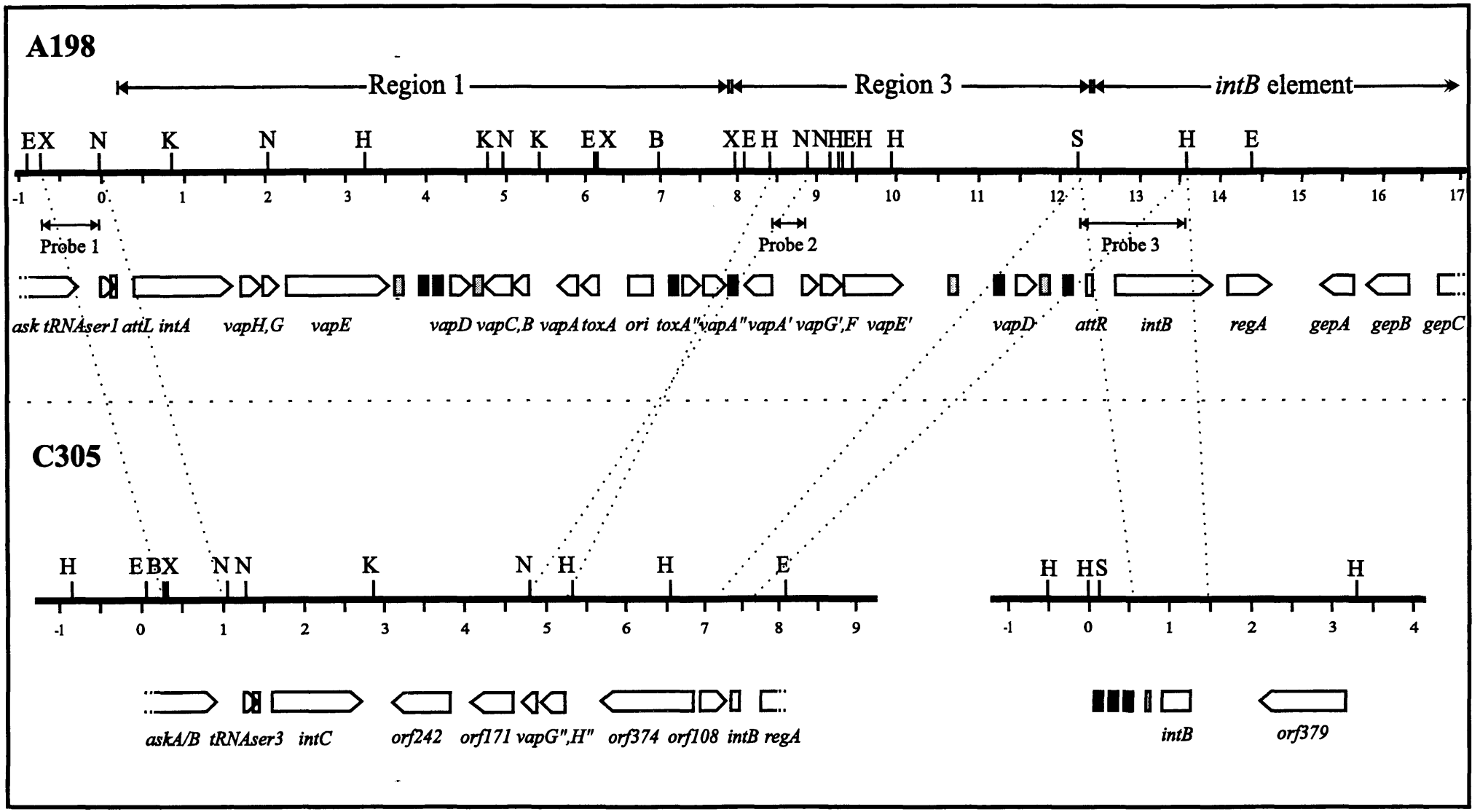

Figure 4.13: Restriction map of the *vap* regions 1 and 3 from *D. nodosus* strain A198 and the two separate clones from the C305 library which hybridised to probes 1, 2 and 3. The location of the probes is indicated under the A198 map and the dotted lines indicate the hybridisation location in C305. The genes are indicated by open arrows. Restriction enzyme sites are: B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII, K = *Kpn*I, N = *Nru*I, S = *Sac*I and X = *Xho*I.

partial copy of *intB* encodes the initial 164 amino acid of the 403 a.a. IntB protein. Analysis of the DNA has shown that in C305 there is a copy of half the *attR* site from the right-hand end of *vap* region 2, from strain A198, approximately 180 b.p. upstream of the start of the *intB* gene (Figure 4.14). The region surrounding this C305 partial copy of the *intB* gene has a nucleic acid identity of 93.8% over 1440 b.p. to the right-hand flanking sequences of the A198 *vap* region 2 (Figure 4.14). The numbering of the DNA sequence in Figure 4.14 starts from the beginning of the half *att* site in each sequence. The similarity begins with half the *vap attR* site and includes, the DNA prior to *intB*, the partial copy of *intB* and sequences after the disruption of *intB*. This high nucleic acid similarity indicates that this partial copy of *intB* arose prior to divergence of strains A198 and C305. The similarity of these sequences, between strains, also give us an indication of the evolutionary distance between the not so closely related genes of *vap* regions 1 and 3 and the integrase genes *intA*, *intB* and *intC*.

Downstream of the partial *intB* gene lies another ORF in the opposite orientation to *intB* (Figure 4.13). This gene, termed *orf379*, has no similarity to genes in the databases. There is no evidence for the presence of *regA* or any *gep* genes from the *intB* element (Chapter3).

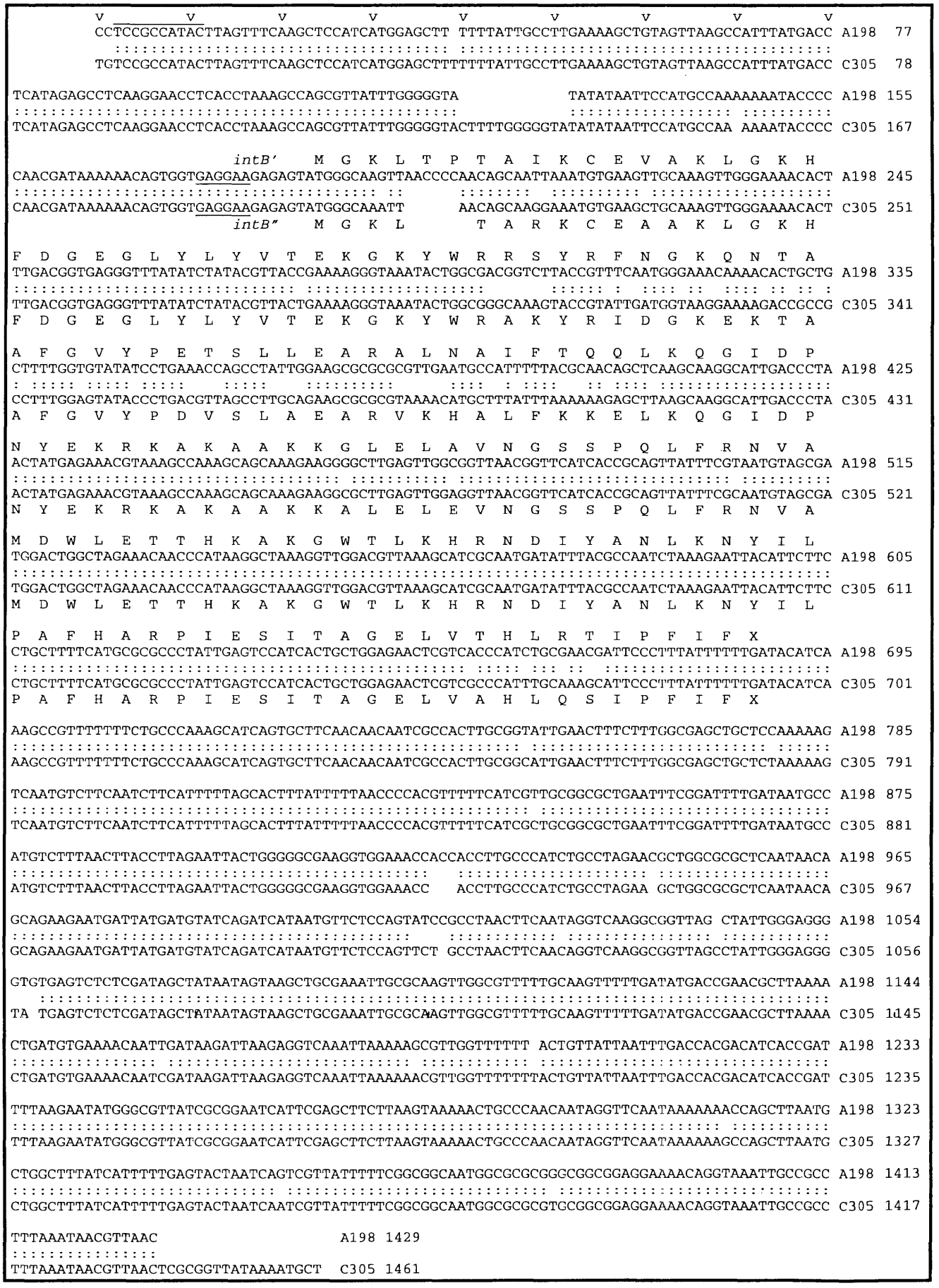

Figure 4.14: Alignment of the DNA sequence flanking the right-hand end of *vap* region 2, from strain A198, with the corresponding sequence from C305. The amino acid sequences of the partial *intB* genes are aligned with the appropriate codon. The half *att* site has a line above the sequence and the Shine-Dalgarno sites for the *intB* genes are underlined.

4.3) Discussion

The *askA/B* and *tRNA-ser3* genes in C305 are the corresponding copies of the genes from strain A198, located upstream of *vap* regions 1. DNA sequence comparisons between the two strains, although the A198 DNA sequence has not yet been completely determined, indicate there is approximately 97% identity over the entire 1.2 kb encoding the *askA/B* genes and the *tRNA-ser* gene, including the intergenic sequences. The minor differences between these regions from A198 and C305 can easily be attributed to genetic drift. This region of the *D. nodosus* genome appears to be unaffected by the insertions, deletions and rearrangements associated with the genetic elements located downstream of the *tRNA-ser* gene. The similarity of genes upstream of *vap* region 1 in A198 with sequences from strain C305 was expected. The similarity of the restriction map from these areas in both strains of *D. nodosus* (Figure 4.3) indicates regions common to the bacterial species.

The DNA sequence from the end of *tRNA-ser3* to the end of *intC* has 59.8% identity over 1386 nt to the A198 DNA sequence from the end of *tRNA-ser1* and spanning *intA*. This divergence indicates that *intC* and *intA* are much more distantly related than the *askA/B* and *tRNA-ser* genes from strains A198 and C305.

IntC is more closely related to IntA (approximately 55% amino acid identity) than to any other integrase in the sequence data base (less than 42% amino acid identity). The level of amino acid identity between the products of *vapG* and *vapG"*, and *vapH* and *vapH"* is also approximately 55%. This suggests that the *intC* element may have been derived from the *vap* element, and that considerable divergence of the *int*, *vapG* and *vapH* genes has occurred. In addition, the other genes from the *vap* element have

been lost and two new genes, *orf242* and *orf171*, have been acquired. Since the flanking sequences at both the left and right-hand sides of the *vap* regions 1 and 3 in strain A198, and the *intC* element in strain C305, are very highly related at the nucleotide sequence level, divergence of the *int*, *vapG* and *vapH* genes must have occurred prior to integration of the *vap* element and the *intC* element into the respective *D. nodosus* strains.

There are two other alternative explanations. The first is that the *intC* element is not derived from the *vap* element, and that the similarities between the integrase and *vapG* and *vapH* genes arose by chance. The high level of similarity at the nucleotide level between *vapG/vapH* and *vapG"/vapH"* (81% nucleotide identity) suggests a third possibility: both the *vap* element and the *intC* element may have integrated at this location in strain C305, and most of the *vap* element, except for *vapG* and *vapH*, has been lost.

The IS*1253* identified in C305 has 98.3% identity over 1643 b.p. to the copy located on the *vap* plasmid. This is close to the divergence seen in *askA/B* and *tRNA-ser* genes between strains A198 and C305. The approximately equal separation time for *askA/B*, *tRNA-ser* and the IS element, between strains of *D. nodosus*, suggest that this IS element was present prior to divergence of the strains.

The *intB* gene, from the *intB* element, is disrupted in strain C305. However, the sequence similarity of the partially sequenced *regA* gene in C305 is 98.5% identity over 329 nt. with the copy in strain A198. This again is similar divergence to that seen in *askA/B*, *tRNA-ser* and the IS element between strains of *D. nodosus*. The sequence similarities suggest that the *intB* element was present in *D. nodosus* prior to divergence of the C305 and A198 strains. Also the partial copy of *intB*, adjacent to *vap* region 2 in A198 and located separately in C305, was also present prior to

divergence of these strains. There is approximately 94% similarity over 1440 nt. between the C305 and A198 sequences spanning this partial copy of *intB*. The greater divergence seen in this area may be due to the fact that only part of the *intB* gene is present. Therefore, there is no selective pressure to conserve this non-functional sequence.

4.3.1) Similarity between the *intC* element and *vap* region 3.

The DNA similarity between *intA* and *intC* is not great enough for *intC* to be detected by Southern or dot blots probed with a labelled *intA* probe. Due to the inability to detect the *intC* gene in C305 the discovery of this gene was unexpected and suggested that C305 may once have contained the *vap* regions or a *vap* region-like element.

An interesting feature of the *vap* regions has been revealed by high stringency hybridisation studies. In strain A198 only *vapD* from region 3 will cross hybridise to *vap* regions 1 and 2 (Katz *et al.*, 1992) while the genes of *vap* region 1, in particular *vapA, E*, and *G*, all cross hybridise to the corresponding genes of *vap* region 2 (Katz *et al.*, 1992; Bloomfield *et al.*, 1997) but do not hybridise with the corresponding genes of *vap* region 3. The *vapA'* gene is not detected by a *vapA* probe and *vapE'* is not detected by a *vapE* probe and *vise versa*. The *vapD* gene is the one exception to this non-cross-hybridisation of *vap* genes between *vap* region 3 and the other *vap* regions. However, *vapD* appears to be involved in duplications involving the surrounding 102 and 103 b.p. repeats (Figure 4.1; Cheetham *et al.*, 1995; Bloomfield *et al.*, 1997). The genes of *vap* region 1 also hybridise to the *vap* plasmid, from *D. nodosus* strain AC3577 (Billington *et al.*, 1996a). These hybridisation patterns indicate that *vap* regions 1, 2 and the *vap* plasmid are more closely related to each other than they are to

vap region 3, indicating that *vap* region 3 resulted from a separate insertion of a divergent *vap* element.

VapG" from the *intC* element is more closely related to VapG' from *vap* region 3 (79% amino acid identity) than to VapG from *vap* region 1 (55% amino acid identity). Thus, the *intC* element may be more closely related to *vap* region 3 than to *vap* region 1. Recent analysis has shown the *vapG/H* operon is similar to the immunity region of bacteriophage P4 (Whittle and Cheetham, unpublished). These bacteriophage genes are involved in preventing further infections of the bacterium once the initial infection has occurred. The *vapH* product may therefore be an RNA molecule rather than a protein (Whittle and Cheetham, unpublished). This would explain the higher conservation of the DNA sequence over the predicted amino acid sequence of this gene. The *vapG"/H"* genes may therefore be part of the genetic element and are not *vap* genes *per se*. The high conservation of these genes may therefore be due to the functional significance of immunity of the host to reinfection with the same element.

A separate λ clone of C305 genomic DNA contains three copies of the 102 b.p. repeats found throughout the *vap* regions (Katz *et al.*, 1994). These repeats have been implicated in a number of rearrangements within the *vap* regions of strain A198 (Cheetham *et al.*, 1995). Their presence in C305 suggests that either C305 once contained the *vap* regions or the repeats are originally from the bacterial DNA and have infiltrated the *vap* regions.

4.3.2) A novel genetic element

The two distinct genetic elements, *vap* regions and the *intB* element, are related by the similarity of the integrase gene carried by each and the insertion into the same,

or very close to the same, bacterial chromosome integration site. *Vap* region 3, from strain A198, is also believed to be a separate *vap* element which has inserted into the end of this *tRNA-ser* gene. The identification of another related integrase gene, *intC*, may indicate yet another genetic element belonging to this growing family of elements.

The *intC* element may be a separate element from the *vap* regions and *intB* element. Two ORFs, *orf242* and *orf171*, are located adjacent to the *intC* gene in C305 and are absent from the A198 *vap* regions. Southern blot analysis of 17 strains of *D. nodosus* has indicated that *intC*, *orf242* and *orf171* are present in most strains and are part of an element approximately 13 kb in length (Whittle and Cheetham, unpublished). This putative element is a similar size to that of the *vap* regions. The *intC* element is also believed to have been deleted from the genome of one strain of *D. nodosus* during culturing of the bacteria in the laboratory (Whittle and Cheetham, unpublished).

There now appears to be four different elements, *vap* region 1, *vap* region 3, *intB* element and the *intC* element, which are able to insert into the end of *tRNA-ser* genes with the *D. nodosus* genome. This family of genetic elements appears quite diverse with the primary link being the closely related integrase genes.

4.3.3) IS*1253* and its relationship with the *vap* regions

There are a number of possibilities for the relationship of IS*1253* to the *vap* regions of *D. nodosus*. IS*1253* may have moved to this location after the *intC* element integration; it may have been carried on the *intC* element; or the IS element may have been already present in this region of the *D. nodosus* genome. A *vap*-like element carrying IS*1253* may have been introduced into a C305 progenitor but may have been

lost subsequently, leaving behind a non-functional IS element. However, IS*1253* is not associated with the *vap* regions in strain A198. Alternatively IS*1253* may, coincidentally, have an insertion site within this C305 region and the *vap* plasmid. The IS element is not virulence-associated, since it is present in C305 and other benign strains and absent from many virulent strains (Billington *et al.*, 1996a).

Insertion sequences related to IS*1253* are found adjacent to virulence-associated genes of *Salmonella typhimurium* (Gulig *et al.*, 1992), *Yersinia pestis* (Simonet *et al.*, 1996) and *Helicobacter pylori* (Censini *et al.*, 1996). This similarity and the amino acid similarities of the IS*1253* ORF products to IS elements from such a diverse group of bacteria suggests that this element, and possibly the *vap* regions, may have been transferred between distantly related bacteria.

4.3.4) *intB* element in C305

The *intB* genetic element is believed to have inserted into the same or similar site as the *vap* regions by the use of the site-specific integrase *intB*. The *intB* element appears to be non-functional in C305 due to large deletions within, and prior to, the *intB* gene and the promotor region of *regA*. These rearrangements and deletions may have been mediated by IS*1253* and/or the insertion of the *intC* element or a *vap*-like element.

The DNA sequence from a separate λ clone from C305, which hybridised to probe 3 from the A198 *vap* region (Figure 4.13), has a high similarity to DNA from the corresponding A198 region adjacent to *vap* region 2. This similarity is relatively consistent with the divergence seen in other genes when comparisons have been drawn between A198 and C305. The minor differences can be accounted for by genetic drift. This partial copy of *intB* therefore, appears to have arisen prior to

divergence of the A198 and C305 strains. From the Southern blot analysis of C305 there is only two copies of the *intB* gene in this strain. Both copies of *intB* have been identified in this study and were found to be only partial copies of the gene. PCR has confirmed there is no complete copy of *intB* in C305 (Chapter 3). The presence of *intB,* or part there of, in all strains tested indicates the insertion occurred in an ancestral *D. nodosus*.

All three *vap* regions in A198 appear to be from separate integrations of a closely related genetic element and have not arisen by duplications. Regions 1 and 2 are inserted into different *tRNA-ser* genes (Bloomfield *et al.*, 1997) and the genes of *vap* region 3 are highly diverged from those of *vap* regions 1 and 2 indicating an ancient evolutionary separation of the regions. If strain C305 did once contain the *vap* regions then the only remnants are *vapG*, *H* and the 102 b.p. repeats. The presence of *intC* and two genes of unknown function, *orf242* and *orf171*, indicates that either a separate element has inserted into the genome of strain C305 or a large number of deletions and rearrangements have occurred to remove these genes from strain A198 and the remaining *vap* genes from strain C305. The putative role of *vapG/H* in bacteriophage immunity, the divergence of *intC* from *intA* and the presence of previously unidentified genes suggests that the *intC* element is a new genetic element. This new element is related to, but different from, the *vap* elements and the *intB* element which also integrate at this site in the *D. nodosus* genome.

CHAPTER FIVE:

Induction and characterisation of a bacteriophage from *D. nodosus*.

5.1) Introduction.

Two regions of the *D. nodosus* chromosome, the *vap* and *vrl* regions, are associated with virulence (Katz *et al.*, 1991, 1992; Rood *et al.*, 1994). The *vap* regions arose by the integration of a genetic element into a tRNA gene (Cheetham *et al.*, 1995). The nature of this genetic element, which has similarities to both bacteriophages and plasmids (Chapter 1), is unknown.

Investigation of the *vrl* has indicated that it may have arisen by the integration of a bacteriophage (Haring *et al.*, 1995). Analysis of the left-hand end of the *vrl* has identified an attachment site, *attL*, with high similarity to the site found in a gene for small stable 10Sa RNA, *ssrA*, in *E. coli* K-12 (Haring *et al.*, 1995). A P4-like cryptic prophage integrated into this *ssrA* gene in *E. coli* is believed to regulate AlpA (ATP-dependent Lon Protease) protease expression (Kirby *et al.*, 1994). The *vrl* is integrated into a *ssrA* gene in *D. nodosus* and may have a similar regulatory function to the P4-like cryptic prophage of *E. coli* (Haring *et al.*, 1995). There are repeated sequences following the *attL* site of the *vrl* which are similar to the repeats within the

att region of the P4 family of bacteriophages lending further support to the bacteriophage origin of the *vrl*. Analysis of the right-hand virulent/benign junction failed to identify an *attR* site for the *vrl*.

The *intB* genetic element found adjacent to the *vap* regions of *D. nodosus* may be derived from a conjugative transposon or a bacteriophage (Chapter 3). The *intB* element has been partially characterised and comprises a P4-like integrase gene, *intB*, followed by a *cI*-like gene, *regA*, similar to the *cI* gene of bacteriophage ϕ80. The *intB* element is believed to have inserted into a similar site to the *vap* regions via the action of the *intB* gene.

The possibility that the *vrl*, *vap* regions or *intB* element are integrated bacteriophages was investigated by treating several strains of *D. nodosus* with agents known to induce bacteriophage excision. No native bacteriophage of *D. nodosus* has been identified previously, although phage-like particles have been observed using electron microscopy of *D. nodosus* (Stuart and Egerton, 1979; Walker *et al.*, 1973; Gradin *et al.*, 1991). In this chapter the induction and characterisation of the first identified native bacteriophage of *D. nodosus* is presented.

5.2) Results

5.2.1) Treatment of cultures of *D. nodosus* with mitomycin C

Mitomycin C is the most commonly used chemical agent for the induction of prophage excision from bacterial genomes (Ackermann and DuBow, 1987). In an attempt to induce bacteriophages from *D. nodosus*, this mutagen was added, in varying concentrations, to a liquid culture during the log phase of bacterial growth. After growth to stationary phase the bacterial cells were removed by low speed centrifugation and the supernatant filtered to remove any remaining bacteria or bacterial debris. This was followed by high speed centrifugation at a velocity capable of sedimenting bacteriophage particles. Nucleic acids were prepared from the precipitate obtained after the high speed spin, and from the bacteria collected from the first, low speed, centrifugation step. The nucleic acid preparation from the collected bacteria was conducted so as to investigate the induced excision of genetic elements from the *D. nodosus* genome which may be unable to exit the bacterial cells. Examples of such elements may include integrated plasmids, transposons, defective bacteriophages, which lacked the ability to form mature phage particles, or an inducible satellite bacteriophage. Genes similar to the satellite bacteriophage P4 have previously been identified within the *D. nodosus* genome (Cheetham *et al.*, 1995; Bloomfield *et al.*, 1997).

The strains of *D. nodosus* used for the induction studies cover a range of serogroups and virulence classifications. These include the virulent strains A198, B1006, G1220 and H1215, the benign strains C305 and H1204 and the intermediate strain AC3577. Induction was attempted at least three times on each strain of *D. nodosus*. Table 5.1 indicates the nucleic acids detected in each preparation. *D.*

nodosus strain AC3577 contains the *vap* plasmid (Billington *et al.*, 1996a) thus, extra-chromosomal DNA was detected in the non-induced bacterial culture (Table 5.1). No further extra-chromosomal DNA was detected in this strain after induction with mitomycin C hence the negative result (Table 5.1). No bacteriophage DNA or extra-chromosomal DNA was detected in non-induced cultures of any other *D. nodosus* strain tested. Induction with mitomycin C failed to produce any bacteriophage or extra-chromosomal DNA from any of the *D. nodosus* cultures tested (Table 5.1).

Table 5.1: Investigation of *D. nodosus* strains for the presence of extra-chromosomal DNA or bacteriophages DNA from non-induced and mitomycin C induced cultures.

Strain of *D. nodosus*	Non-induced bacterial culture		Culture induced with mitomycin C	
	Non-chromosomal DNA	Bacteriophage DNA	Non-chromosomal DNA	Bacteriophage DNA
A198	-	-	-	-
AC3577	+	-	-	-
B1006	-	-	-	-
C305	-	-	-	-
G1220	-	-	-	-
H1204	-	-	-	-
H1215	-	-	-	-

Prevotella ruminicola strain AR29 obtained from Rosalind Swain (Institute of Biotechnology, Department of Molecular and Cellular Biology, University of New England) was used as a positive control for the mitomycin C induction method (Figure 5.1A). This rumen bacterial strain is known to contain prophages inducible with mitomycin C (Klieve *et al.*, 1989). *P. ruminicola* AR29 was grown under anaerobic conditions in rumen fluid medium (Klieve *et al.*, 1989) at 39°C. This was the only difference in the experimental procedure between induction from *P. ruminicola* and *D. nodosus*. When *P. ruminicola* AR29 are induced there is a large amount of phage

produced indicated by the large amount of DNA present in the induced bacteriophage preparation (Figure 5.1A). No bacteriophage DNA was detected when *D. nodosus* cultures were induced with mitomycin C (Figure 5.1B). Figure 5.1B shows no detectable DNA present in the bacteriophage preparation from cultures of *D. nodosus* strain H1215 induced with a range of doses of mitomycin C.

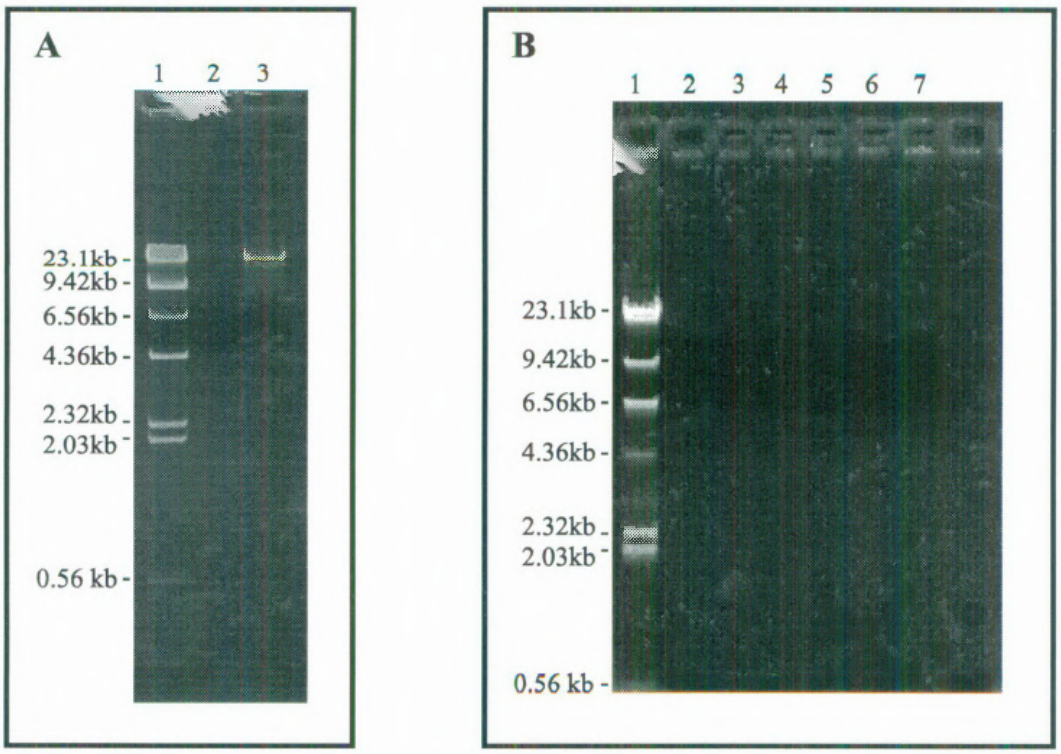

Figure 5.1: Nucleic acid preparations from prophage induction using mitomycin C. (**A**) Induction of bacteriophages from *P. ruminicola* AR29. Lane 1, λ DNA digested with *Hin*dIII; lane 2, DNA preparation from a non-induced culture; lane 3, DNA from a culture induced with 1.0 μg/ml mitomycin C. (**B**) Induction of bacteriophages from *D. nodosus* H1215. Lane 1, λ DNA digested with *Hin*dIII; lane 2, culture induced with 0.2 μg/ml; lane 3, induced with 0.6 μg/ml; lane 4, 1.0 μg/ml; lane5, 1.4 μg/ml; lane 6, 1.8 μg/ml and lane 7, 2.2 μg/ml.

5.2.2) Treatment of cultures with UV light

Ultra-Violet light is the most commonly used agent to induce prophage excision from the bacterial genome (Ackermann and DuBow, 1987). A similar method to the

mitomycin C induction and detection was used except that the bacteria were subject to UV light, for various lengths of time. Each of the strains mentioned above (see section 5.2.1) was again tested at least three times and the results summarized in Table 5.2. No extra-chromosomal DNA other than the *vap* plasmid was detected in strain AC3577, hence the negative result.

DNA was present in the bacteriophage DNA preparation from *D. nodosus* strain H1215 (Table 5.2). The DNA in this preparation shows that a UV inducible prophage is present within the genome of the virulent *D. nodosus* strain H1215 and a bacteriophage native to *D. nodosus* has been found. No bacteriophage DNA was detected in the non-induced culture of H1215 (Table 5.1) and no DNA was detected after UV induction of the other strains of *D. nodosus* (Table 5.2). There was also no extra-chromosomal DNA detected within the bacterial cells of these strains of *D. nodosus* after induction with UV light.

Table 5.2: Investigation of *D. nodosus* strains for the presence of extra-chromosomal DNA or bacteriophage DNA after induction with UV light.

Strain of *D. nodosus*	*D. nodosus* culture induced with UV light	
	Non-chromosomal DNA	Bacteriophage DNA
A198	-	-
AC3577	-	-
B1006	-	-
C305	-	-
G1220	-	-
H1204	-	-
H1215	-	+

As a positive control for the UV induction method *P. ruminicola* AR29 was used (Figure 5.2A). In Figure 5.2A the control culture, lane 1, contained phage particles, indicated by the DNA isolated. This may be due to a low level of spontaneous

bacteriophage excision and release or the bacteria may have been stressed by resuspension in $MgSO_4$, before exposure to UV light, therefore causing a smaller degree of induction. However, induction in the UV exposed cultures was considerably greater, Figure 5.2A lanes 2 and 3. No bacteriophage DNA was detectable in the non-induced culture of *D. nodosus* strain H1215, however, after exposure to UV light a large amount of DNA was present in the bacteriophage preparation (Figure 5.2B). This inducible prophage was further investigated by electron microscopy.

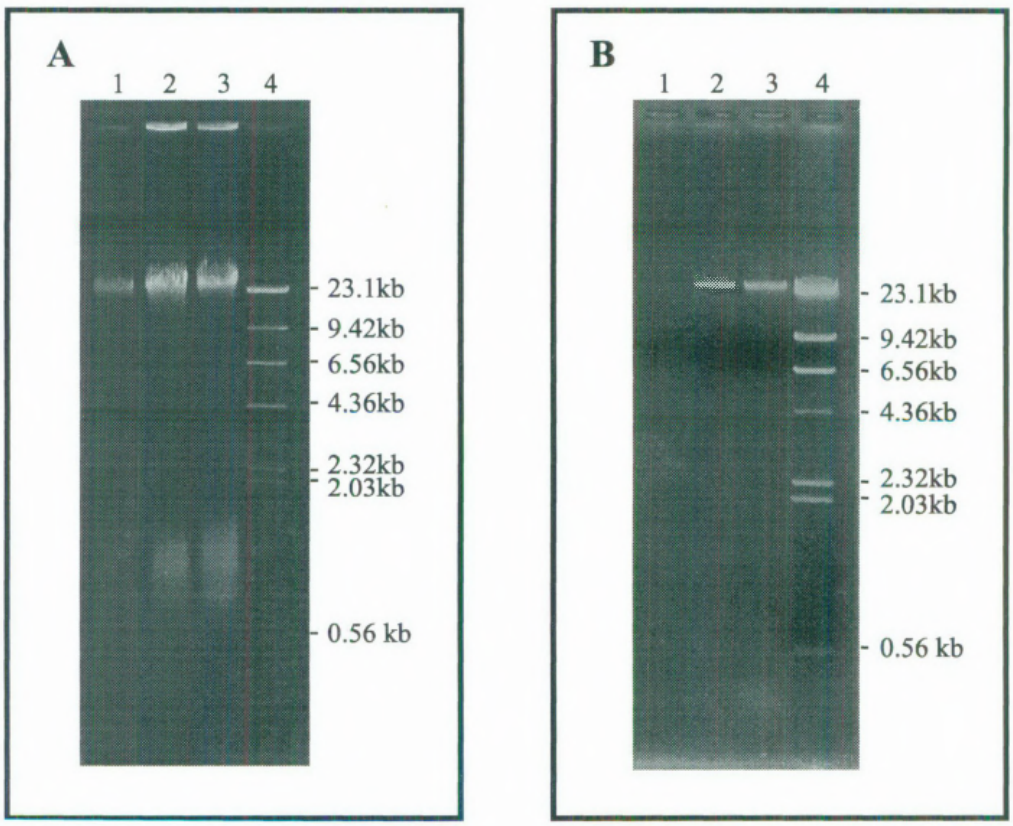

Figure 5.2: DNA preparations from prophage inductions using UV light. (**A**) Induction of bacteriophage from *P. ruminicola*. Lane 1, DNA prepared from a culture not exposed to UV light; lane 2, DNA from a culture exposed to UV light for 60 seconds; lane 3, DNA from a culture exposed for 100 seconds and lane 4, λ DNA digested with *Hin*dIII. (**B**) Bacteriophage DNA preparations from *D. nodosus* H1215. Lane 1, DNA preparation from a non-induced culture; lane 2, DNA from a H1215 culture exposed to UV light for 30 seconds; lane 3, culture exposed for 45 seconds and lane 4, λ DNA digested with *Hin*dIII.

5.2.3) Characterisation of the *D. nodosus* bacteriophage by electron microscopy

Bacteriophage preparations for electron microscopy were negatively stained with either phosphotungstic acid, PTA, or uranyl acetate, UA, (Hayat, 1986; Hayat, 1993). Observations of the lysate from cultures of *D. nodosus* strain H1215 after induction by UV light revealed only one morphological type of phage (Figure 5.3). Phage particles were of uniform size and comprised of an icosahedral head, 57 ηm apical diameter, and tail, 10 ηm x 203 ηm, with a unique claw-like base plate (Figure 5.3). The dimensions of the tail indicate it is non-contractile (Murphy *et al.*, 1995). The base plate of this new phage is 13 ηm across and 15 ηm long (included in the tail length). The phage head in Figure 5.3C appears pentagonal which is a common, distinguishing, feature of icosahedral phage heads (Ackermann and DuBow, 1987). Measurements of phage dimensions were obtained from photographic prints of 31 individual particles from numerous preparations negatively stained with either PTA or UA. As can be seen in Figure 5.3 the UA stain of the phage enhanced the tail detail considerably although the heads were not as well defined as when stained with PTA.

The new phage has been named DinoHI from the host bacterial genus, species and strain as set out by Ackermann & DuBow (1987) i.e. *Dichelobacter nodosus*, strain H1215 and 'I' because it is the first identified phage from this bacterial species.

5.2.4) Isolation and characterisation of phage nucleic acid

The nucleic acid isolated from the bacteriophage preparation was sensitive to DNase (Figure 5.4A) and able to be digested with restriction endonucleases which require double stranded DNA (dsDNA) templates thus indicating a dsDNA genome

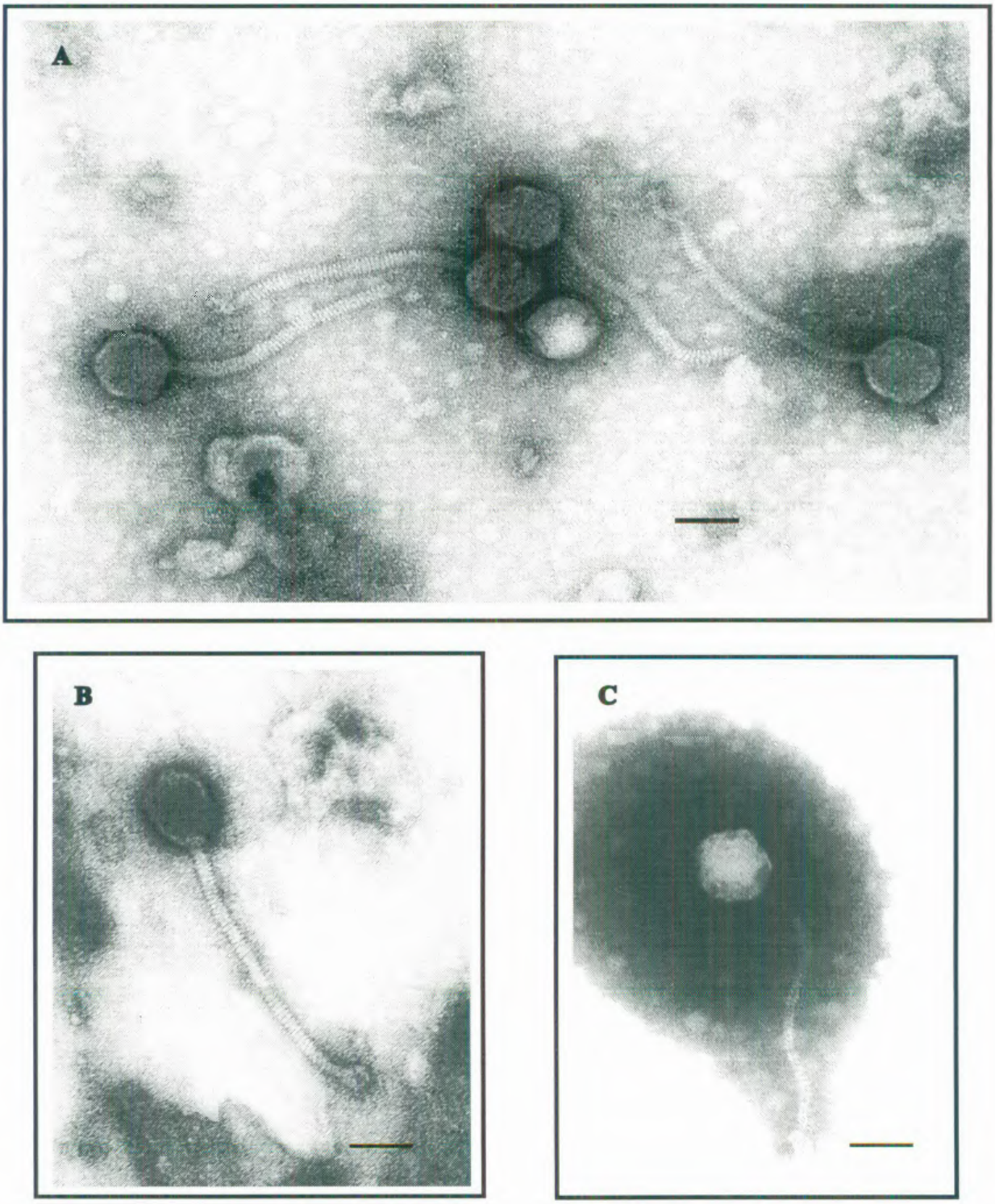

Figure 5.3: Electron micrographs of DinoHI stained with UA (**A** and **B**) and PTA (**C**), bar = 50 ηm

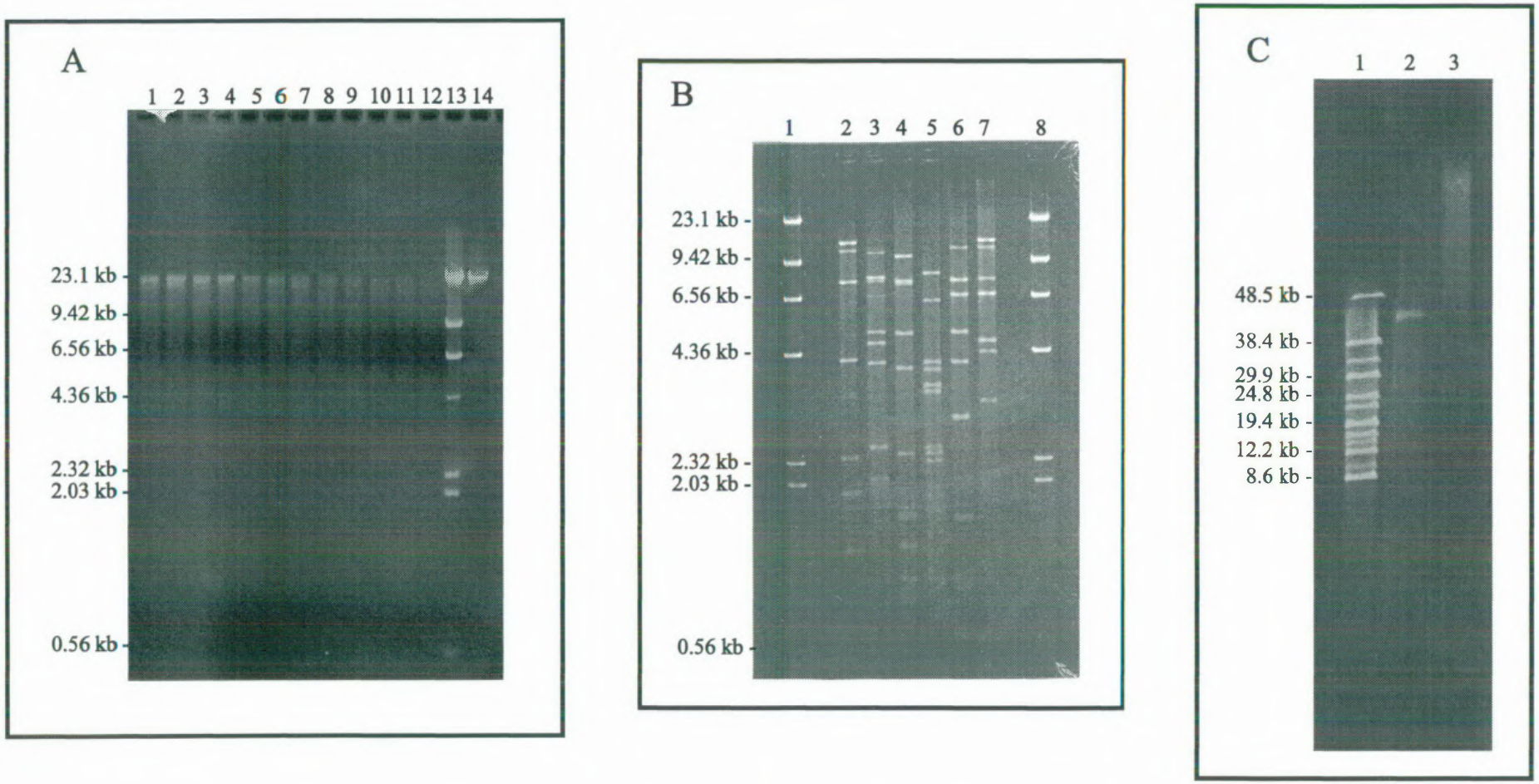

Figure 5.4: (**A**) DinoHI DNA digested with DNaseI for varying lengths of time, lanes 1-12 corresponding to 5 to 60 minutes sampled at 5 minute intervals, lane 13 is the λ DNA digested with *Hin*dIII and lane 14 is undigested DinoHI DNA. (**B**) DinoHI DNA digested overnight with various restriction endonucleases, each digest is with *Sac*I plus; lane 2, *Ava*I; lane 3, *Hin*dIII; lane 4, *Kpn*I; lane 5, *Nru*I; lane 6, *Pst*I and lane 7, *Sal*I. Lanes 1 and 8 contain the λ DNA digested with *Hin*dIII. (**C**) Pulsed field gel of undigested DinoHI genomic DNA, lane 2, and undigested bacterial genomic DNA from *D. nodosus* strain H1215, lane 3. A High Molecular Weight Marker (BRL) is the linear DNA standard, lane 1.

(Figure 5.4B). Pulsed field gel electrophoresis (PFGE) of the phage DNA (Figure 5.4C) indicated that the genome is a single linear molecule of between 40 and 45 kb.

Digestion of the phage genome with various restriction endonuclease has enabled an accurate size of the genome to be calculated as 43 kb. The size of the DNA bands were ascertained by their migration relative to a λ *Hin*dIII standard (Figure 5.4B) and the genome size calculated by the sum of the DNA fragments. The PFGE band sizes were identified by the migration relative to the High Molecular Weight Markers (Bethesda Research Laboratories). The use of restriction endonucleases in combination enabled the placement of the enzyme sites, relative to each other, and therefore the construction of a restriction map of the DinoHI genome (Figure 5.5).

From the morphology and genome structure DinoHI belongs to the virus family *Siphoviridae* (Murphy *et al.*, 1995). The *Siphoviridae* family is the largest virus family and consists of bacteriophages with isomeric heads, long non-contractile tails and linear dsDNA genomes.

5.2.5) Southern blot analysis of DinoHI

Southern blot analysis was carried out to investigate the possibility that the genome of DinoHI carried DNA sequences from the *vap* regions, the *intB* element or the *vrl*. Genomic DNA from DinoHI was probed with subclones (Figure 5.6), from the *vap*, *intB* and *vrl* regions. Probe 1, from the *vap* region, includes part of the *intA* gene while the probe 2, from the *intB* element, contains most of the *intB* gene (Figure 5.6). Subclones from within the *vrl*, in the plasmids pJIR743 and pJIR590 (Haring *et al.*, 1995), were located at the extremities of the *vrl* (Figure 5.6). A clone containing

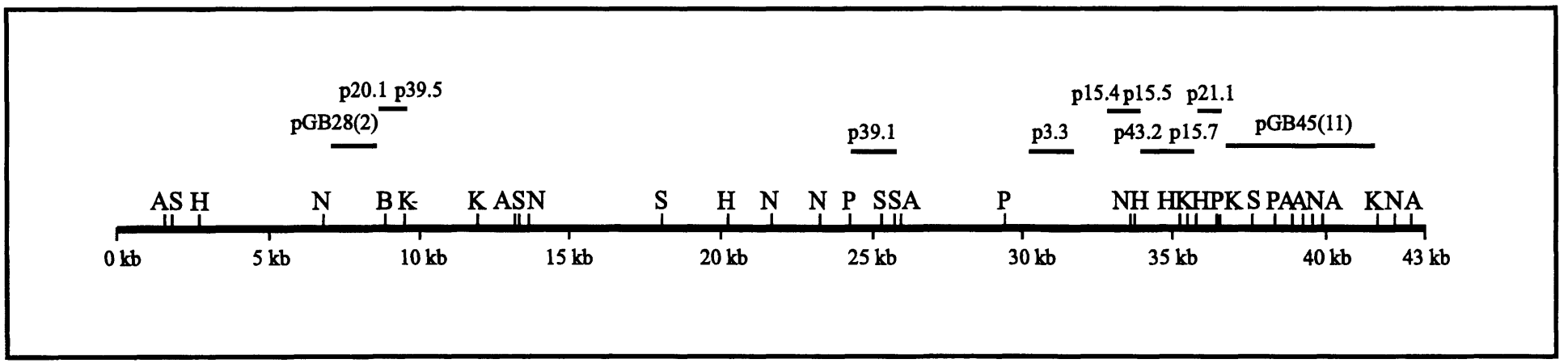

Figure 5.5: Restriction map of the DinoHI genome showing the location of subclones p3.3, p15.3, p15.4, p15.7, p20.1, p21.1, p39.1, p39.5, p43.2 pGBP28(2) and pGBP45(11). The size of the genome is indicated in kilobases (kb) and the restriction endonuclease sites are marked by the letters A = *Ava*I, B = *Bam*HI, H = *Hin*dIII, K = *Kpn*I, N = *Nru*I, P = *Pst*I and S = *Sac*I.

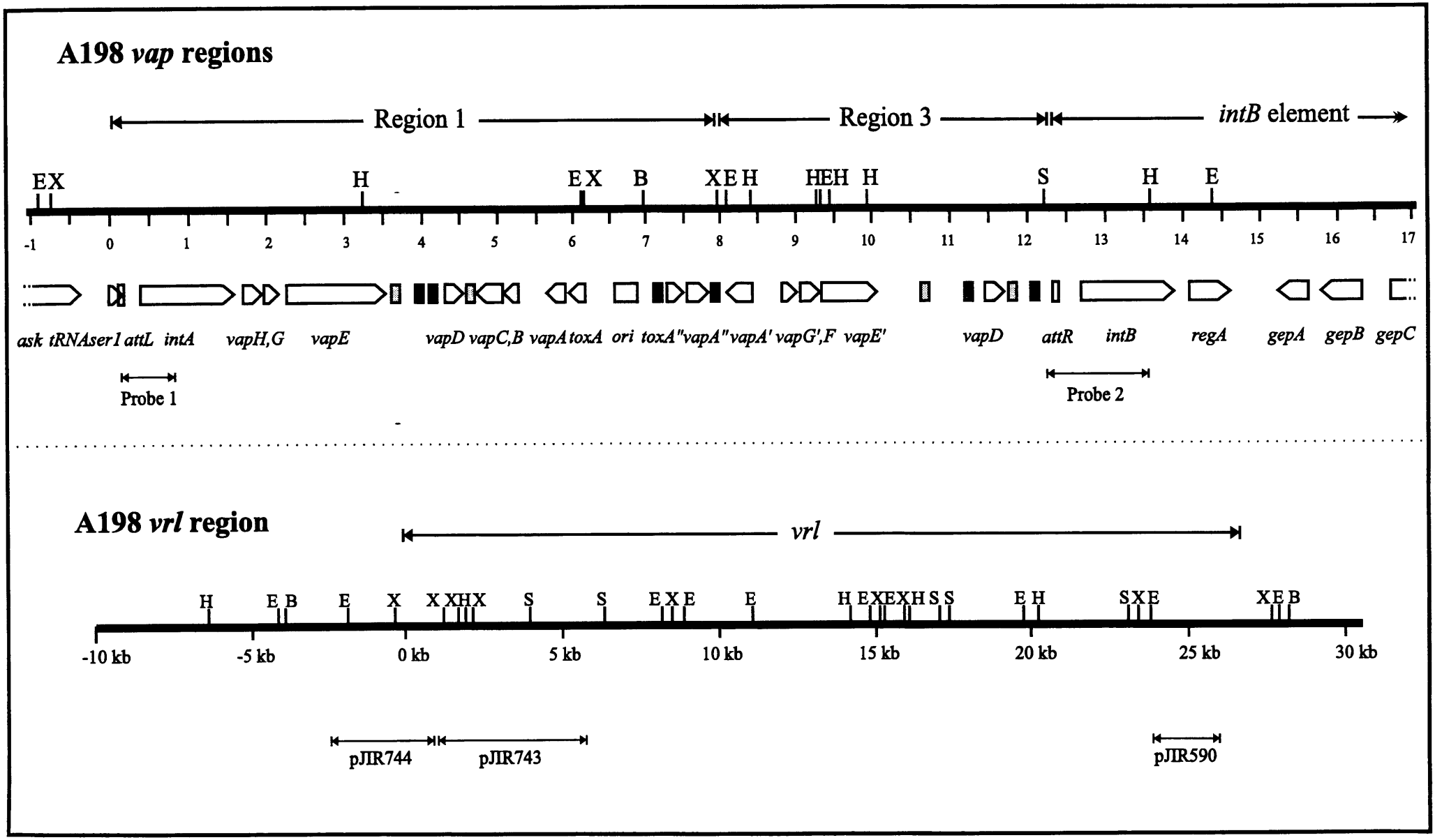

Figure 5.6: Restriction map of *vap* regions 1 and 3 and the *vrl* (Haring *et al.*, 1995) from *D. nodosus* strain A198. The gene identified in the *vap* region are indicated below the map. The clones used in the hybridisation studies of DinoHI are indicated below each map. Restriction enzyme sites are represented by the single letters, B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII, S = *Sac*I and X = *Xho*I.

sequences adjacent to the left-hand end of the *vrl*, pJIR744 (Figure 5.6), was also used to probe DNA from DinoHI.

The clones from the *vap* region, *intB* element and *vrl* all failed to hybridise to the DinoHI DNA, indicating that this bacteriophage, induced from the *D. nodosus* genome, does not carry any of these previously identified genetic elements. Therefore, DinoHI represents a new genetic element capable of integration into the *D. nodosus* genome.

5.2.6) Partial sequencing of subclones from DinoHI

To further characterise and identify the bacteriophage DinoHI, a number of subclones of the bacteriophage genome were constructed and partially sequenced (Figure 5.5). Sequence similarity searches through the databases failed to detect any similarity of the partial DNA sequences from DinoHI to known bacteriophage or viral sequences. There was only one subclone with similarity to known sequences in the databases. The pGBP28(2) clone, from the DinoHI genome (Figure 5.5), has sequence identity with the bacterial genomic DNA adjacent to the right-hand end of the *D. nodosus* A198 *vrl* (Figure 5.7; Haring *et al.*, 1995). The DNA sequences were 98.3% identical over 180 b.p. from the cloning site of pGBP28(2) until the end of the Genbank *vrl* sequence (Figure 5.7). The DinoHI subclone pGBP28(2) begins at an *Eco*RI site 500 b.p. downstream from the identified *vrl* virulent-benign junction (Haring *et al.*, 1995). This DNA identity suggests that the DinoHI prophage lies adjacent to the *vrl* region within the *D. nodosus* genome. Therefore, bacteriophage DinoHI may have integrated into a similar site as the *vrl* or this bacteriophage may have been involved in the acquisition of the virulence-associated *vrl* region by *D. nodosus*.

```
                                                    10        20        30
pGBP28.2                                          AATTCATTTATCACGCGCTCGCATAAGGTT
                                                  ::::: ::::::::::::::::::::::::
RH vrl    GTGCCGTTTAACCGCAGCACTTTTGCGATGAATTCGTTTATCACGCGCTCGCATAAGGTT
               4260      4270      4280      4290      4300      4310

                   40        50        60        70        80        90
pGBP28.2 AAATCGTAATCAAAGAGAATTTGCGAGAAATTATCATCGGTTTCTTGGGCAGCGCGGCTG
         ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
RH vrl   AAATCGTAATCAAAGAGAATTTGCGAGAAATTATCATCGGTTTCTTGGGCAGCGCGGCTG
               4320      4330      4340      4350      4360      4370

                   100       110       120       130       140
pGBP28.2 -CACCGGCAAGTTCGCTGGTGAGCACNCGCCCGATGATGGTTTTTTGAATGCGTTGAATA
          :::::::::::::::::::::::: :::::::::::::::::::::::::::::::
RH vrl   CCACCGGCAAGTTCGCTGGTGAGCACGCGCCCGATGATGGTTTTTTGAATGCGTTGAATA
               4380      4390      4400      4410      4420      4430

         150       160       170       180       190       200
pGBP28.2 TGGTGACGGGTAAAAAGCTGATGAGCGGATCCGTCCGATGTAGCATTAACGAATGAGATT
         ::::::::::::::::::::::::::::::
RH vrl   TGGTGACGGGTAAAAAGCTGATGAGCGGATC
               4440      4450      4460
```

Figure 5.7: Alignment of the sequence of from clone pGBP28(2) from DinoHI with the sequence adjacent to the right-hand end of the *vrl* (RH *vrl*), Genbank accession number DNAVRL2. The *Eco*RI site, used to clone the pGBP28(2), is underlined and identical nucleotides are indicated by a colon.

5.2.7) DinoHI prophage presence in other strains of *D. nodosus*

Dot blot and Southern blot hybridisation was used to investigate whether the bacteriophage DinoHI was present in other strains of *D. nodosus*. Genomic DNA was prepared from the seven strains of *D. nodosus* already investigated for induction of prophage excision (Tables 5.1 and 5.2). Southern blots were probed with pGBP28(2) and pGBP45(11) from DinoHI (Figure 5.8). Sequences hybridising to the pGBP28(2) subclone were found to be present in all strains tested. Virulent strains B1006 and G1220 may have multiple copies (Figure 5.8A). However, sequences hybridising to pGBP45(11) were present only in the virulent strains H1215 (from which DinoHI was isolated), B1006 and G1220 (Figure 5.8B). The multiple banding pattern in this Southern blot is due to the presence of *Eco*RI restriction endonuclease sites within the sequences hybridising to the probe. The largest band from B1006 and G1220, which

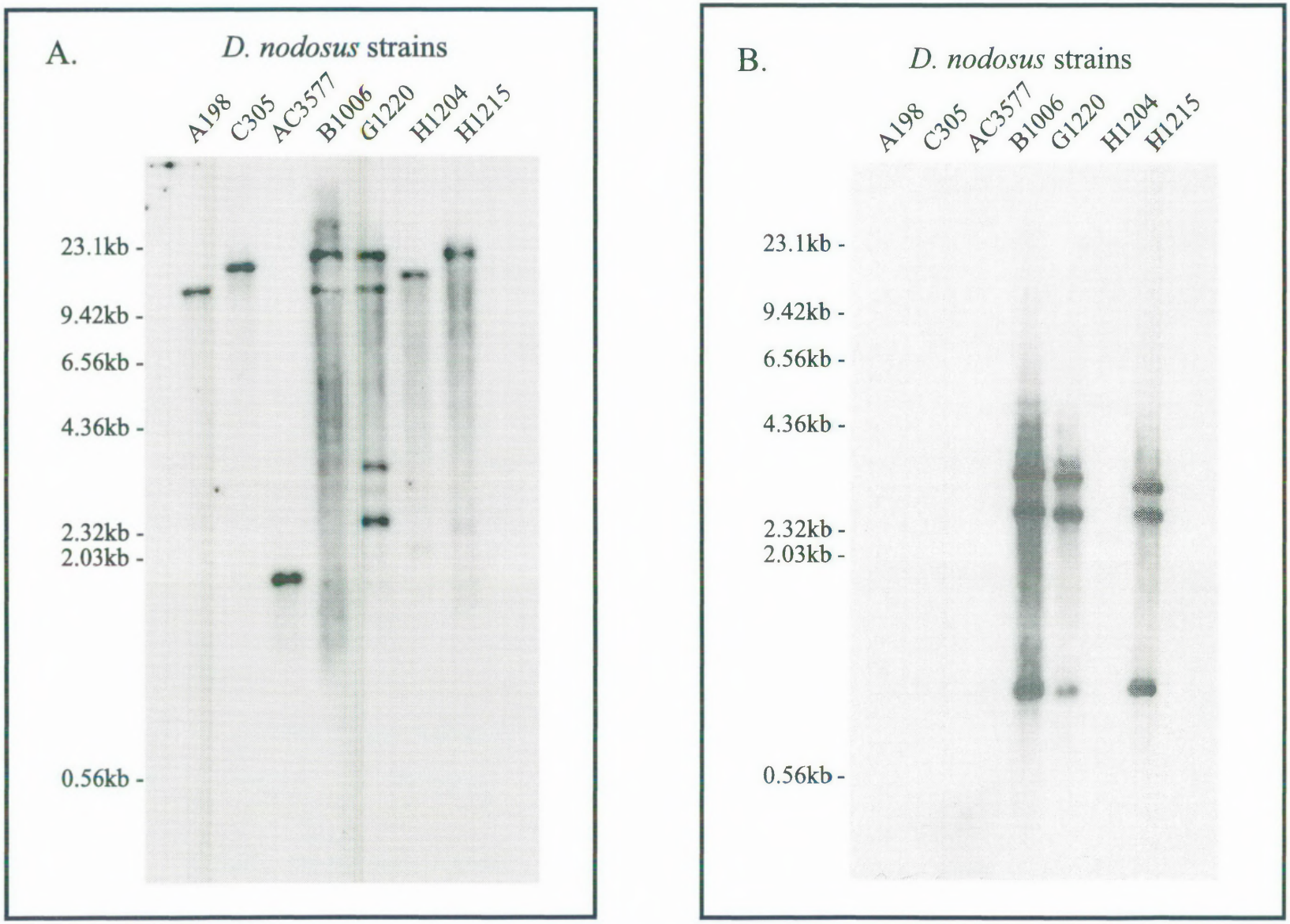

Figure 5.8: Southern blots of *D. nodosus* genomic DNA. (**A**) DNA digested with *Hin*dIII and probed with pGBP28(2). (**B**) DNA digested with *Eco*RI and probed with pGBP45(11). The *D. nodosus* strains are indicated at the top of each lane and the location of the bands, from the λ DNA digested with *Hin*dIII, is indicated in kilobases.

hybridised to pGBP45(11), is approximately 250 b.p. larger than the corresponding band in H1215 (Figure 5.8B). This indicates a change in the location of one of the *Eco*RI sites within the DinoHI prophage.

Dot blots of genomic DNA from other strains of *D. nodosus* were probed with labelled pGBP28(2) and pGBP45(11), Figure 5.9. These strains include the virulent strain 3526 and the benign strains 1136, 1169, 1383, 1469, 1493 and 3138. Sequences hybridising to pGBP28(2) were present in all strains, but sequences hybridising to pGBP45(11) were present in only one other strain, 3526 (Figure 5.9A).

These results suggest that all *D. nodosus* strains contain part of the phage genome, contained in subclone pGBP28(2), the sequences which lie adjacent to the *vrl* region in the A198 genome. Only strains B1006, G1220, H1215 and 3526 hybridised to the second probe, pGBP45(11), suggesting that only these strains may contain the whole bacteriophage genome. However, DinoHI particles were not inducible from B1006 or G1220, suggesting that these strains harbour non-functional or incomplete copies of the prophage.

5.2.8) Investigation of *D. nodosus* strains as hosts for DinoHI

To investigate the host range of DinoHI, a number of infection studies were carried out on various *D. nodosus* strains. This study involved virulent and benign strains from a variety of serogroups and included strains C305, 1136, 1169, 1383, 1469, 1493, 3138 and 3526. Bacterial cells were harvested from an agar plate and incubated with bacteriophage particles to allow phage adsorption to the cells. The infected bacteria were then grown in a liquid broth. The culture was investigated for lytic phage growth (bacteriophage particles in the supernatant) and lysogenic phage growth (phage DNA within infected cells) as follows: The bacteria were collected by

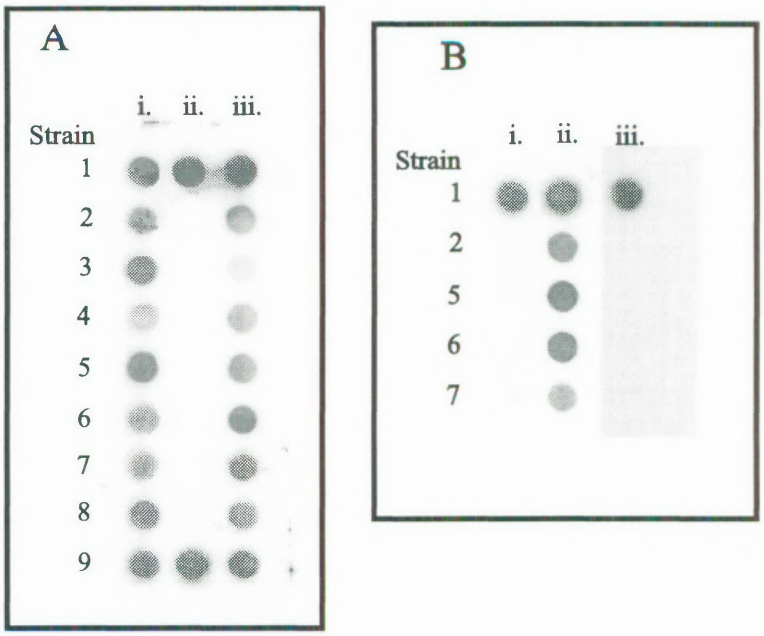

Figure 5.9: Dot blots of *D. nodosus* genomic DNA. (**A**) **i**. probed with pGBP28(2), **ii**. probed with pGBP45(11) and **iii**. probed with pGBP45(11) after infection with DinoHI. (**B**) Blots probed with pGBP45(11) **i**. before infection with DinoHI, **ii**. after infection with DinoHI and **iii**. after propagation of infected bacteria on plate media. The strains of *D. nodosus* are; 1, H1215; 2, C305; 3, 1136; 4, 1169; 5, 1383; 6, 1469; 7, 1493; 8, 3138 and 9, 3526.

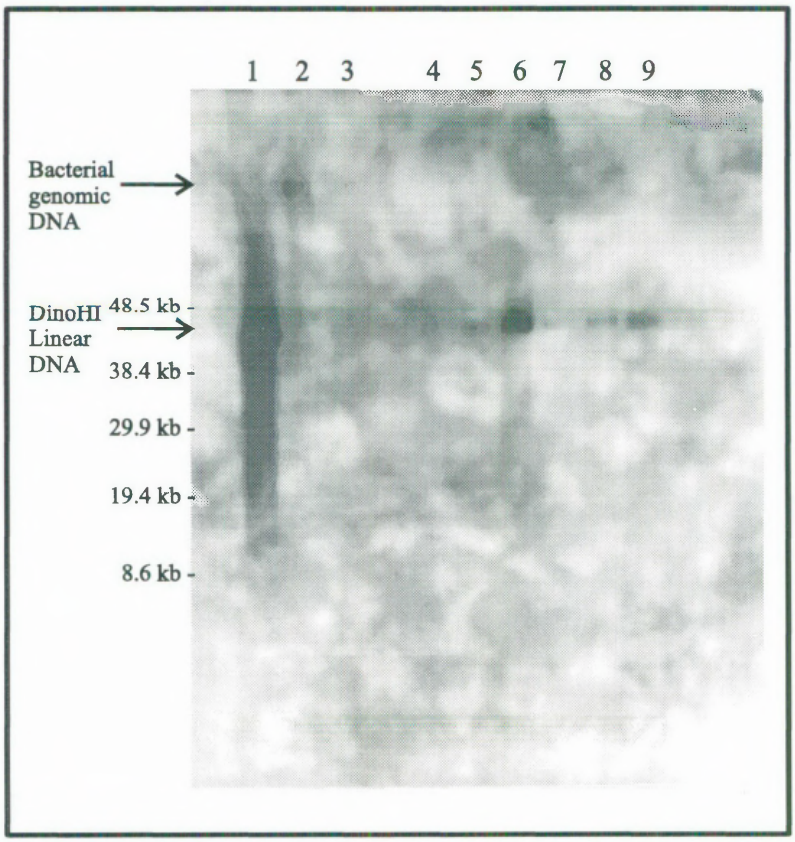

Figure 5.10: Southern blot of DinoHI infected *D. nodosus* probed with pGBP45(11). Lane 1, undigested DinoHI genomic DNA; 2, H1215 uninduced; 3, C305 uninfected; 4, 3138 infected with DinoHI; 5, C305 infected; 6, 1383 infected; 7, 1469 infected; 8, 1493 infected and 9, 1169 infected.

centrifugation, then the supernatant was filtered and centrifuged at a speed sufficient to pellet bacteriophage particles. A DNA preparation was performed on the collected fraction. There was no DinoHI DNA detected in this preparation from any infected *D. nodosus* strains. This indicates that bacteriophage DinoHI was not going through lytic growth as no phage particles were produced from the infected strains of *D. nodosus*.

DNA was also prepared from the bacterial cells collected from each infected culture. This bacterial genomic DNA preparation was blotted on a membrane and probed with the pGBP45(11) subclone from the DinoHI bacteriophage genome. Prior to bacteriophage infection most *D. nodosus* strains lacked sequences which hybridised to the DinoHI subclone pGBP45(11) (Figures 5.8B & 5.9). Therefore, pGBP45(11) was used as an indicator of bacteriophage presence in infected cells which, prior to infection, lacked these phage sequences. After infection with DinoHI all *D. nodosus* strains contained sequences which hybridise to pGBP45(11) (Figure 5.9). The presence of DinoHI sequences in the bacterial genomic DNA preparations from each *D. nodosus* strain indicates DinoHI was either attached to the cells or the bacteriophage DNA had entered the bacterial cells. Pulsed field gel electrophoresis of the bacterial DNA preparation, followed by Southern blotting, showed the bacteriophage DNA to be in the linear form and not circular or incorporated into the bacterial genome (Figure 5.10). When the infected bacteria were propagated on plate media the phage DNA was no longer detectable (Figure 5.9B) indicating that DinoHI did not stably lysogenize the bacteria.

5.3) Discussion

Tailed phages are found in over 100 genera of eubacteria, archaebacteria and cyanobacteria (Murphy *et al.*, 1995) and most bacteria which have been examined have been found to be lysogenic (Roberts and Devoret, 1983). Klieve *et al.* (1989), for example, have found that inducible viral genetic material is a normal constituent of the bacterial genome in a significant proportion of ruminant bacterial species. Bacteriophages are well known to play a dominant role in the evolution of bacteria (Bradley, 1967).

5.3.1) Prophage excision from the bacterial genome.

Prophage excision may be induced by treatment with an appropriate physical, chemical or biological agent although the sensitivity to inducing agents varies from one prophage to another (Ackermann and DuBow, 1987). A wide variety of agents can induce prophages including radiation, chemical carcinogens and mutagens, antigyrase drugs, base analogues, some antibiotics, temperature, starvation of an essential compound or superinfection with another phage or plasmid. The mechanism of action varies between the different types of inducing agents. Many affect DNA replication or cause direct damage to the bacterial DNA. Induction of λ prophage activates the bacterial RecA protein which cleaves and inactivates repressor cI (Weisberg *et al.*, 1977). Other mechanisms include heat inactivation of thermolabile repressors (Weisberg *et al.*, 1977) and a number of agents act indirectly, possibly on several targets (Ackermann and DuBow, 1987).

Mitomycin C and UV light are the most widely used agents to induce prophage excision from the bacterial genome (Ackermann and DuBow, 1987), both of which cause direct damage to DNA. However, many phage are not inducible by either agents. For example phage 186 is UV inducible whereas the closely related phage P2 is not (Woods and Egan, 1974) and P2 is only temperature-inducible after superinfection of the host (Bertani, 1968). Bacteriophage P4 is also temperature inducible (Lagos and Goldstein, 1984) although this satellite phage lacks the genes for phage particle production. The mechanism of action of mitomycin C and UV light is, however, different because some phage are inducible with one and not the other (Ackermann and DuBow, 1987). The first bacteriophage isolated from *D. nodosus* by induction, DinoHI, is an example of this difference as it is inducible with UV light but not inducible with mitomycin C.

5.3.2) Bacteriophage classification.

Bacteriophage classification is based primarily on the gross morphology of the virus and the nucleic acid make up of its genome. Accordingly Bradley (1967) classified bacteriophages into six groups, A - E. The tailed phages, containing double stranded DNA (dsDNA) genomes, were divided into groups A to C on morphology. Current classification has maintained this system with tailed phages divided into the three families, *Myoviridae* - contractile tails, *Siphoviridae* - long, non-contractile tails, and *Podoviridae* - short tails (Ackermann *et al.*, 1992). Within each virus family the isolates are further classified on the basis of host range, CsCl buoyancy, sensitivity to physical and chemical inactivation, restriction endonuclease patterns, nucleic acid base composition, serological properties, plaque type and size, and the type of receptor (Ackermann *et al.*, 1978; 1992). Most tailed phage are only classified at the family

level due to the large number of tailed phage isolates, over 3,300, and the varying amounts of data available on each (Ackermann *et al.*, 1992). Further classification of the *Siphoviridae* family is even more difficult since over 50% of tailed phages identified belong to this group (Ackermann *et al.*, 1992). The *D. nodosus* bacteriophage DinoHI has been classified into the *Siphoviridae* family of viruses. DinoHI may also represents a new species of virus since no other bacteriophages have been identified from the bacterial host genus, *Dichelobacter.*

5.3.3) Bacteriophages and bacterial virulence.

Southern blot analysis of the DinoHI genome indicates that this phage does not encode any of the previously identified genetic elements incorporated into the *D. nodosus* genome. Hence, DinoHI is not related to the *vap* virulence-associated regions or the second adjacent genetic element, the *intB* element (Chapter 3). Also the DinoHI phage does not encode the *vrl* although the prophage appears to be located adjacent to the right-hand end of the *vrl* in the virulent strain A198. DinoHI does not appear to be virulence-associated, unlike the *vrl*, since at least part of the phage genome is present in all strains (both virulent and benign) which were tested, and large sections of the phage genome are absent from the virulent strain A198 (Figure 5.8). However, sequences hybridising to the subclone pGBP45(11), from DinoHI, are only present in the virulent strains B1006, G1220, H1215 and 3526 and absent from all 8 benign strain tested.

Haring *et al.*, (1995) identified a bacteriophage-like attachment site at the left-hand end of the *vrl* but found no such sequence at the right-hand virulent/benign junction. The *vrl*, as previously discussed, has been proposed to have arisen by the integration of a bacteriophage (Haring *et al.*, 1995). DinoHI is not that phage but may

have been involved in the transfer of the *vrl* region, possible via transduction, thus enhancing virulence of the recipient strains. The function of the *vrl* in virulence of *D. nodosus* is unknown (Haring *et al.*, 1995).

It is quite common for a bacteriophage to be linked to virulence in bacteria (Finlay and Falkow, 1989). There are many examples of toxins encoded by bacteriophages (Chapter 1, section 1.4.2.3) and bacteriophages which enhance bacterial virulence in unknown ways (Voelker *et al.*, 1995) are constantly being discovered.

5.3.4) DinoHI and the *vrl*.

The association of DinoHI with the *vrl* is unknown and unfortunately at this point in time we can only speculate as to the evolution of this virulence related locus and the relationship to the adjacent prophage. The *vrl* may have been part of a transducing DinoHI bacteriophage or DinoHI may be a helper phage for the *vrl* or even a *vap* satellite phage, similar to the bacteriophage P2 to P4 relationship. Further characterisation of the *vrl* will reveal more information about this genetic element. Analysis of the DNA sequences downstream of the *vrl* may reveal more about the evolutionary history of this region of the *D. nodosus* genome. In strain A198 sections of the DinoHI genome have been deleted indicating a complicated evolutionary history. Further analysis of the DinoHI genome, including identification of the phage *att* site, will give more insight into the insertion of DinoHI into the *D. nodosus* genome and may reveal more about the *vrl* and its relationship to DinoHI.

The partial sequence analysis of DinoHI failed to identify any similarity to known genes encoded by bacteriophages. Since no other phages of *D. nodosus* have been characterised there may not be sufficient similarity with known phages to pick up any

similarity in the database searches. For example P22 and λ are serologically related phages but overall only 13.5% of their genomes are similar (Ackermann *et al.*, 1978). Identification and analysis of individual genes, and the deduced amino acid sequences, from DinoHI may locate some sequence similarity to known genes.

The preliminary Southern and dot blot results from the DinoHI clones in various strains of *D. nodosus* (Figures 5.7 and 5.8) imply two possible evolutionary histories of the phage. Firstly if all *D. nodosus* strains contain part of the phage genome this suggests a long and complicated evolutionary history. This would be consistent with the concurrent investigations of the *D. nodosus* genomic *vap* regions and adjacent *intB* element indicating a highly fluid bacterial genome (Chapters 3 and 4). The second possibility is that DinoHI has "picked up" some bacterial genomic sequence common to all strains, i.e. pGBP28(2). This occurrence has been observed in *Lactococcus lactus* with the evolution of a lytic phage by acquisition of bacterial genomic DNA (Moineau *et al.*, 1994). The bacteriophage was involved in major DNA rearrangements within the bacterial genome and the phage genome gained 2.3 kb causing a phage particle morphology change by the addition of base plate and extension of the tail.

This arrangement of the *vrl* and DinoHI appears very reminiscent of the *vap* regions and the *intB* element. DinoHI, or part there of, is adjacent to the right-hand end of the *vrl* region and is also present in benign strains. The *intB* element appears to have a similar arrangement with the *vap* regions. Another similarity between the *vrl*-DinoHI region and the *vap-intB* region comes from the analysis of the limits of the virulence-associated regions in the benign strain C305. Analysis of the limits of the *vrl* in the benign strain C305 shows the ends, instead of being adjacent, are separated by approximately 3 kb of DNA (Haring *et al.*, 1995). Similarly, the limits of the *vap*

regions analyzed in C305 (Chapter 4), are separated by approximately 6 kb of DNA which may contain remnants of the *vap* regions. This similarity between the *vrl*-DinoHI and *vap-intB* regions would be even more striking if analysis of the 3 kb intervening sequence in C305, between the limits of the *vrl*, contains *vrl*-like sequences.

5.3.5) Is DinoHI able to infect other strains of *D. nodosus*?

With the enormous numbers of bacteriophages identified there is naturally a large variance in the host range. Most phages are specific for the bacterial host species or genus (Ackermann and DuBow, 1987). Bacteriophages which have been observed to only infect one or two strains of a bacterial species are also common (Reanney and Teh, 1976; Tamada *et al.*, 1985; Ackermann and DuBow, 1987). A number of bacteriophage isolates have a wider host range, for example the enterobacterial phages are specific for the bacterial family *Enterobacteriaceae* (Ackermann and DuBow, 1987; Murphy *et al.*, 1995). The DinoHI bacteriophage has an unknown host range and may be specific for the *D. nodosus* strain H1215 for which it is lysogenic.

Different phage adsorb to specific receptor molecules on the surface of the bacteria including flagellae, pili, capsules or receptors located on the cell wall (Ackermann and DuBow, 1987; Murphy *et al.*, 1995) and fimbriae (Bradley, 1973, 1974; Bradley and Pitt, 1974). Bradley (1973) identified a bacteriophage, PO4, with a long non-contractile tail, which attached to the pili of *Pseudomonas aeruginosa.* Later work demonstrated host immunity to infection by a mutant bacteria with non-retractile pili (Bradley, 1974). The terminology of fimbriae and pili was, and still is, often used by different researchers to refer to the same or similar filamentous appendages. The *Pseudomonas aeruginosa* pili to which bacteriophage PO4 adsorbed

(Bradley, 1973; 1974) would now more commonly be referred to as *P. aeruginosa* type-4 polar fimbriae similar to those found in *D. nodosus*.

Growth conditions are known to have an influence on the ability of bacteriophages to infect bacteria (Shafia and Thompson, 1964). The inability of bacteriophage DinoHI to stably infect other strains of *D. nodosus* may be due to inappropriate growth conditions. Growth conditions may not be ideal for bacteriophage DNA entry into the cells, phage particle production or the integration of the phage genome into the bacterial genome to form the prophage state. DinoHI is, however, able to adsorb to every strain of *D. nodosus* investigated but it is unknown if the phage DNA enters the bacteria.

All of the *D. nodosus* strains tested for infection with DinoHI contained at least part of the phage genome. This may render them immune to re-infection (superinfection). Once a bacteriophage has lysogenised a bacterium it is typical for the bacterial host to become immune to superinfection by the same or closely related phage. In the case of lysogeny by bacteriophage λ the continuous production of the phage repressor cI, which maintains the phage in the lysogenic state, prevents the expression and replication of superinfecting phage DNA (Ackermann and DuBow, 1987) thus rendering the host immune. It is unknown at this stage exactly how much of the DinoHI phage is present in the *D. nodosus* strains which lack the whole phage genome. However, a relatively small section of the phage genome may be sufficient for immunity. If the DinoHI repressor gene is present and expressed in each *D. nodosus* strain then this may be enough for immunity to superinfection. To this date we have investigated only a few strains of *D. nodosus* for the presence of DinoHI. It may, therefore, be possible to identify a strain of *D. nodosus* that does not contain any

part of the DinoHI genome thus avoiding possible bacterial host immunity to superinfection.

DinoHI is able to produce phage particles in the host *D. nodosus* strain H1215, following induction with UV light, indicating a possibly functional phage which may be adjacent to the virulence associated region of DNA, the *vrl*. Previously no native *D. nodosus* bacteriophage has been identified. The ability to induce prophage excision shows that bacteriophages are active in the *D. nodosus* environment. DinoHI has been linked to the *vrl* and may have played a role in the acquisition of these virulence-associated sequences. Further investigation of this phage may indicate its possible function in the transfer of bacterial virulence genes which would have far reaching implications for the control and treatment of footrot. The ability of the phage particles to attach to the bacterial cells and the possible insertion of phage DNA into the cells suggests that DinoHI may be used to develop a phage based transformation system for *D. nodosus*. At present no transformation system exists. The restriction map of the DinoHI genome will be of considerable use in the further characterisation of the bacteriophage, investigation of insertion into the bacterial host genome and will aid the development of a possible phage based transformation system for *D. nodosus*.