CHAPTER TWO:

Materials and methods.

2.1) Bacterial growth and conditions

2.1.1) *D. nodosus*

All *D. nodosus* strains were grown in an atmosphere of 80% (v/v) N_2, 10% (v/v) H_2 and 10% (v/v) CO_2 at 37°C on Eugonagar (BBL) supplemented with 0.5% (w/v) yeast extract and 5% (v/v) defibrinated horse blood (Amadeus International). Liquid cultures were grown under anaerobic conditions at 37°C in Eugonbroth (BBL) supplemented with 0.5% (w/v) yeast extract.

2.1.2) *E. coli*

E. coli strain DH5α (*sup*E44Δ*lac*U169 (φ80d*lac*ZΔM15) *rec*A1 *end*A1 *hsd*R17 *thi*-1 *gyr*A96 *rel*A1; Bethesda Research Laboratories) was used as a host for plasmids derived from pUC18. *E. coli* strain LE392 (*sup*E44 *sup*F58 *hsd*R514 *gal*K2 *met*B1 *trp*55 *lac*Y1) (Murray *et al.*, 1977) was used as a host for recombinant λ bacteriophages.

2.1.2.1) Preparation of competent cells

DH5α cells were made competent for transformation by washing in either 10 mM $CaCl_2$ or dH_2O, for calcium chloride or electroporation transformations respectively (Sambrook *et al.*, 1989). The transformed cells were grown on 2 x YT medium (Miller, 1972) supplemented with ampicillin (100 μg ml^{-1}) and grown in 10 ml of 2 x YTA broths for plasmid DNA preparations.

LE392 cells were made competent by growth in LB medium (Sambrook *et al.*, 1989) supplemented with 1% maltose. Cells were then washed and stored in 10 mM $MgSO_4$/10 mM $CaCl_2$ (Sambrook *et al.*, 1989). Cells infected with recombinant phage were grown on L + G medium using top agar/agarose overlay (Sambrook *et al.*, 1989). Infected cells were grown in 50 ml of NZCYM broth (Sigma) for phage DNA preparation (Ausubel *et al.*, 1989).

2.2) Gel electrophoresis

2.2.1) Agarose gel electrophoresis

DNA fragments for identification, restriction mapping, subcloning or Southern blots were fractionated by agarose gel electrophoresis in TAE gel buffer (40 mM Tris pH 7.6, 2 mM EDTA) (Sambrook *et al.*, 1989). The concentration of agarose ranged from 0.6% to 1.5% (w/v) depending on the size of the DNA fragments to be separated. Standard 40 ml gels were run at a constant 100 volts for 60 to 90 minutes. Larger 200 ml agarose gels, used for better separation of DNA fragments or Southern blots, were electrophoresed at 35 volts for 16 hours. λ bacteriophage DNA digested

with the restriction endonuclease *Hin*dIII was used as a linear DNA standard for agarose gel electrophoresis. The DNA within the agarose was stained with ethidium bromide and visualized by illumination with short-wave ultra-violet light.

2.2.2) Acrylamide gel electrophoresis

Sequencing reaction products were separated by electrophoresis through a 0.4 mm thick 8% (w/v) polyacrylamide gel in TBE buffer (100 mM Tris pH 8.0, 86 mM boric acid, 2 mM EDTA) at a constant power of 110 watts. The Bio-Rad sequencing apparatus used for the manual sequencing was kept at a constant temperature of 50°C during the DNA separation. To enable a sequence of between 300 to 500 base pair to be read from each reaction, a double loading system was used. The first loading was run for 4.5 hours then the second loading was performed and the gel run for a further 2.5 hours. The sequencing gel was vacuum dried on Wattman blotting paper and exposed to X-ray film for 2-14 days.

2.2.3) Pulsed field gel electrophoresis

Approximately 50 ηg of bacterial genomic DNA was loaded on a 1% (w/v) agarose gel. The gel was run on a Pulsaphor Electrophoresis Unit (Pharmacia Biotech) at 450 volts with a phase time of 0.8 seconds for 4 hours at 12°C in 0.15 x TBE buffer (Doggett *et al.*, 1992; Mathew *et al.*, 1988; Lai *et al.*, 1989). The DNA standard used was the High MW DNA Markers (Bethesda Research Laboratories). The DNA in the gel was stained with ethidium bromide and photographed while illuminated with short wave UV light.

2.3) DNA preparations

2.3.1) Preparation of genomic DNA from *D. nodosus*

D. nodosus strains were inoculated into 150 ml of Eugonbroth by placing 1 ml of Eugonbroth onto a culture on a plate, suspending the bacterial cells, then adding the bacterial suspension to the 150 ml of Eugonbroth. Liquid cultures were grown under anaerobic' conditions at 37°C for 72 hours. Genomic DNA was prepared from *D. nodosus* as described by Anderson *et al.* (1984).

2.3.2) λ DNA preparation

Lambda clones were propagated in the λ-sensitive *E. coli* strain LE392. Titration of λ clones was carried out using the top agar overlay method on L + G medium (Sambrook *et al.*, 1989). λ DNA was prepared by the liquid culture method (Ausubel *et al.*, 1989).

2.3.3) Plasmid preparation

Crude plasmid preparations, for further subcloning and DIG labelled probes, were carried out using the method described by Holmes and Quigley (1981). A small scale version of this method was routinely used to screen plasmids. The alkaline lysis preparation of Birnboim and Doly (1979) was used to isolate more highly-purified plasmid DNA required for use in DNA sequencing.

2.4) Cloning DNA

The DNA to be subcloned was digested with the appropriate restriction endonuclease(s). The fragments were either cloned directly into the vector, or separated by agarose gel electrophoresis. The DNA in the gel was stained with the intercalating agent ethidium bromide and the band(s) of interest, located by illumination with short-wave UV light, were excised from the gel. The DNA was removed from the agarose by use of the Gene Clean II kit (Bio 101).

Fragments of DNA smaller than 10 kb were inserted into the plasmid vector pUC18 (Yanisch-Perron *et al.*, 1985) using T4 ligase. The *E. coli* strain DH5α was transformed with the newly-formed plasmid by the calcium chloride or electroporation methods (Sambrook *et al.*, 1989). The transformed bacteria were then propagated on 2 x YTA selective medium using the X-gal blue/white selection (Sambrook *et al.*, 1989).

2.5) DNA sequencing.

DNA sequencing was performed on double stranded DNA (dsDNA) templates by the dideoxy-mediated chain termination method (Sanger *et al.*, 1977). Sequencing reactions were performed by use of a T7 DNA polymerase sequencing kit (Pharmacia Biotech) incorporating [α^{35}S]-dATP label. The labelled DNA was separated by acrylamide gel electrophoresis. Where compressions or G-C rich regions of DNA were encountered a deaza dGTP sequencing kit (Promega) was used. All DNA

sequences presented are double stranded by obtaining and sequencing overlapping subclones, or by using specific oligonucleotides as sequencing primers, from both DNA strands.

2.6) Polymerase chain reactions

The amplification reactions contained 1 μmole of each oligonucleotide primer, 1.8 mM $MgCl_2$, 0.2 mM dNTPs, 1.0 unit of *Taq* DNA polymerase, approximately 1 ηg of bacterial genomic DNA and 1 x reaction buffer (10 x reaction buffer supplied with the enzyme, Bresatec) in a total volume of 25 μl. The reaction mixtures were amplified for 32 cycles in a Corbett FTS-320 thermal cycler (Corbett Research). Each cycle consisted of 90 seconds denaturation at 94°C, 60 seconds annealing at 60°C and 120 seconds extension at 72°C. Oligonucleotide primers, 20 nucleotides long, were synthesized by the Pharmacia LKG Gene Assembler Plus (Pharmacia Biotech).

2.6.1) Cloning PCR products

The DNA products of the PCR reactions were purified by Wizard Magic PCR Product Purification (Promega). The purified product was end filled (blunted), using the Klenow fragment of DNA polymerase I, and cloned into dephosphorylated *Sma*I digested pUC18 using the SureClone Ligation Kit (Pharmacia Biotech).

2.7) Construction and screening of a C305 genomic library in bacteriophage λ.

2.7.1) Construction of the C305 λ library

C305 genomic DNA was prepared as described by Anderson *et al.* (1984) with the addition of an extra extraction step using CTAB to remove any contaminating polysaccharides (Ausubel *et al.*, 1989). Approximately 5 μg of genomic DNA was partially digested with 0.025 units of the restriction endonuclease *Sau*3AI for 30 minutes to obtain fragments of 10-25 kb. The *Sau*3AI "sticky" termini were partially filled using the Klenow fragment of DNA polymerase I. The *D. nodosus* genomic fragments were then ligated into the *Xho*I half-site arms of λ-gem12 (Promega). The λ clones were packaged using Packagene (Promega) and titrated in LE392 by top agar overlay (Sambrook *et al.*, 1989). The library was amplified and stored according to Ausubel *et al.* (1989).

2.7.2) Screening of the C305 λ library

The initial phage titre of the C305 library was 3.5 x 10^5 pfu/ml. Serial dilutions of the λ library were plated out using top agar overlay. Plaque lifts were constructed in triplicate using nitrocellulose membrane (Biotrace NT - Gelman Sciences) from plates with an even coverage of plaques (Sambrook *et al.*, 1989). The recombinant phage DNA was fixed to the membranes by baking at 65°C for 16 hours. Plaque lifts were hybridised with DIG-labelled probes under high stringency conditions. These included hybridisation in 50% formamide and two washes of the membrane with 0.1 x SSC/0.1% SDS for 15 minutes at 65°C. Positive clones were selected (Sambrook *et*

al., 1989), retitrated and rescreened twice to ensure clone ubiquity. DNA was prepared from the recombinant λ clones as described above.

2.8) Southern and Dot blotting

2.8.1) Southern blots

The DNA to be investigated was digested with the appropriate restriction endonuclease, fractionated by electrophoresis through an agarose gel and transferred to a nylon membrane (Hybond-N, Amersham) as described by Southern (1975). The DNA was fixed to the membrane by baking at 65°C for 16 hours.

2.8.2) Dot blots

Dot-blots were carried out using *D. nodosus* genomic DNA as described by Katz *et al.* (1991) except the DNA was applied to a nylon membrane (Hybond-N Amersham). The DNA was fixed to the membrane by baking at 65°C for 16 hours.

2.9) Hybridisation and detection of DNA blots

Southern blots, dot blots and plaque lifts were hybridised with DIG-labelled plasmid DNA and detected by an enzyme linked immunoassay using the antibody enzyme conjugate anti-digoxigenin alkaline phosphatase and a subsequent enzyme catalyzed chemi-luminescent reaction (Boehringer-Mannheim). Hybridisation was

carried out at 37°C for 6 hours under high stringency conditions with the prehybridization and hybridisation solutions consisting of 50% (v/v) formamide, 5 x SSC (0.75 M sodium chloride, 75 mM sodium citrate), 2% (w/v) blocking stock solution, 0.1% (w/v) N-Lauroyl Sarcosine and 0.02% (w/v) SDS. The high stringency washes, in 0.1 x SSC/0.1% (w/v) SDS at 65°C, and detection were carried out as specified by the DIG kit manufacturer (Boehringer-Mannheim).

2.10) Bacteriophage induction

2.10.1) Mitomycin C induction

Each *D. nodosus* strain tested was inoculated from a plate culture into 10 ml of Eugonbroth and grown under anaerobic conditions at 37°C for 48 hours. Mitomycin C was added in varying concentrations with a final range of 0.2 μg/ml to 2 μg/ml (Klieve, 1988) and the cultures incubated for a further 24 hours. The bacteria were removed by centrifugation at 3,000 x g for 15 minutes at 4°C and the supernatant filtered through a 0.45 μm HV filter (Millipore). The collected bacteria were examined by Gram stain for any contaminants. Any phage particles in the supernatant were collected by centrifugation at 32,000 x g for two hours at 4°C and resuspended in 100 μl of SM (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris pH 7.5, 0.01% (w/v) gelatin; Klieve, 1988).

2.10.2) Ultraviolet light Induction

D. nodosus strains were inoculated into 10 ml of Eugonbroth and grown for 48 hours. The bacteria were collected by centrifugation at 3000 x g for 5 minutes and resuspended in 300 µl of 10 mM $MgSO_4$ (Lamont *et al.*, 1989). The bacterial suspension was placed in a petri dish, so as to be no more than 2 mm deep, and UV-irradiated 35 cm from a 15 watt germicidal lamp for a range of times between 0 and 90 seconds (Tomizawa and Ogawa, 1967; Woods and Egan, 1974). The irradiated bacterial suspension was inoculated into fresh 10 ml of Eugonbroth and incubated at 37°C for 24 hours in the dark to prevent photo-repair (Jacob and Wollman, 1953). Any bacteriophage particles produced were collected as described in mitomycin C induction above.

2.10.3) Large scale bacteriophage induction

D. nodosus strain H1215 was grown on Eugonagar for 72 hours, inoculated into 80 ml of Eugonbroth and incubated for 48 hours. The cells were collected by centrifugation at 8,000 x g for 10 minutes at 4°C and resuspended in 1 ml of 10 mM $MgSO_4$. The cells were UV-irradiated (as described in Ultraviolet light induction above) for 10 to 20 seconds and inoculated into fresh 80 ml of Eugonbroth and incubated for 24 hours in the dark. The bacteria and debris were then removed by centrifugation at 11,000 x g for 15 minutes at 4°C. The supernatant was filtered (as above) and the phage particles collected by centrifugation at 26,000 x g for 1.5 hours at 4°C (Ackermann & DuBow, 1987). Phage particles were resuspended in 200 µl of Hungate's salts solution (Klieve *et al.*, 1989; Ogimoto and Imai,1981). A 20 µl sample of phage suspension was reserved for electron microscopy and host range studies while DNA was prepared from the remaining suspension.

2.11) Bacteriophage nucleic acid preparation

Phage nucleic acid was purified by heating the sample at 65°C for 15 minutes in 0.14% (w/v) SDS/7 mM EDTA, extracting twice with an equal volume of phenol/ chloroform/isoamylalcohol (25:24:1) and once with an equal volume of chloroform/ isoamylalcohol (24:1). The nucleic acid was precipitated with 0.3 M sodium acetate, 50% (v/v) isopropanol and held at -20°C for 30 minutes. The precipitated nucleic acid was collected by centrifugation at 12,000 x g for 15 minutes at 4°C, dried, resuspended in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) (Sambrook *et al.*, 1989) and examined by agarose gel electrophoresis and PFGE.

2.12) Electron microscopy

Phage preparations examined by transmission electron microscopy (TEM) were diluted 1:3 in Hungate's salts solution. An 8μl drop of diluted phage suspension was allowed to settle for 10 minutes on 0.25% (w/v) formvar-coated 200 mesh copper grid and the access liquid removed (Horne, 1965). The samples were stained with either 2% (w/v) potassium phosphotungstic acid (pH 7.0) for 20 to 25 seconds or 4% (w/v) uranyl acetate (pH 4.0) for 5 to 10 seconds (Hayat, 1986; Hayat, 1993). The prepared grids were air dried and examined using a Phillips 300 transmission electron microscope. Phage particles observed were photographed at a magnification of x100,000.

2.13) Bacteriophage host sensitivity/range

D. nodosus were grown on blood Eugonagar for 72 hours and collected by resuspending the cells in 600 μl of Eugonbroth. A 200 μl aliquot of cell suspension was incubated with 10 μl of the phage particle suspension at room temperature for 30 minutes to allow phage to adsorb to the bacterial cells. Ten milliliters of Eugonbroth was inoculated with the bacterial/phage suspension and incubated under anaerobic conditions at 37°C for 48 hours. Bacterial cells were collected by centrifugation at 12000 x g for 10 minutes at 4°C and the supernatant inspected for phage particles. Genomic DNA was prepared from the *D. nodosus* cells by a previously-reported method (Anderson *et al.*, 1984).

CHAPTER THREE:

Analysis of a new genetic element from *D. nodosus*.

3.1) Introduction.

Analysis of the DNA sequence from a *Sac*I-*Dra*I fragment spanning the right-hand end of *vap* region 3 in strain A198 (Figure 1.5) revealed part of an integrase gene, suggesting that a second genetic element may be integrated next to *vap* region 3 (Bloomfield, 1992; Cheetham *et al.*, 1995). The coding region of this integrase gene, designated *intB*, begins 180 b.p. from the *att* site, *attR*, at the right-hand end of *vap* region 3. The *att* site is duplicated at the left-hand end of *vap* region 1, *attL*, and the two *att* sites correspond to the virulent/benign DNA junctions at the extremities of the *vap* regions (Cheetham *et al.*, 1995; Figure 1.5). The first 67 amino acids of IntB have 55.1% identity with the first 69 amino acids of the integrase from bacteriophage P4 (Bloomfield, 1992). P4-like phages integrate into bacterial genomes such that the integrase gene is located approximately 200 b.p. from the *att* site (Pierson and Kahn, 1987; Sun *et al.*, 1991). This suggests that *intA* from the *vap* regions and *intB* recognise the same *att* site.

In this chapter, DNA sequences to the right of *vap* region 3 in A198 were analysed to investigate the nature of this putative second genetic element. The sequence of 4.3 kilobases of DNA to the right of *vap* region 3 in strain A198 was determined.

3.2) Results.

3.2.1) Sub-cloning and sequencing

Two clones spanning *vap* regions 1 and 3 (Figure 3.1) were isolated previously from a library of genomic DNA from *D. nodosus* strain A198 in bacteriophage λ (Katz *et al.*, 1994). The region of DNA under investigation in this section is situated to the right of *vap* region 3 in the A198 genome (Figure 1.5). The DNA sequence of the segment from the *Sac*I site, within the right-hand end of *vap* region 3, to the *Sau*3AI site at the end of λ clone λ1.61 (Figure 3.1) was determined. This segment of the *D. nodosus* genome is enlarged in Figure 3.1 to better illustrate its location and some of the more frequent restriction enzyme sites. Not all of the restriction enzyme sites used for sub-cloning and sequencing of this 4.8 kilobase pairs of DNA are shown on the map.

The cloning and sequencing strategy for the right-hand flanking segment of *vap* region 3 is shown in Figure 3.2. Each arrow represents the location and size of the sequence obtained from a sub-clone, in the plasmid pUC18, from the λ1.61 clone. The manual sequencing method used allowed a maximum of 400-450 b.p. to be accurately determined from each end of a subclone, with an average accurate

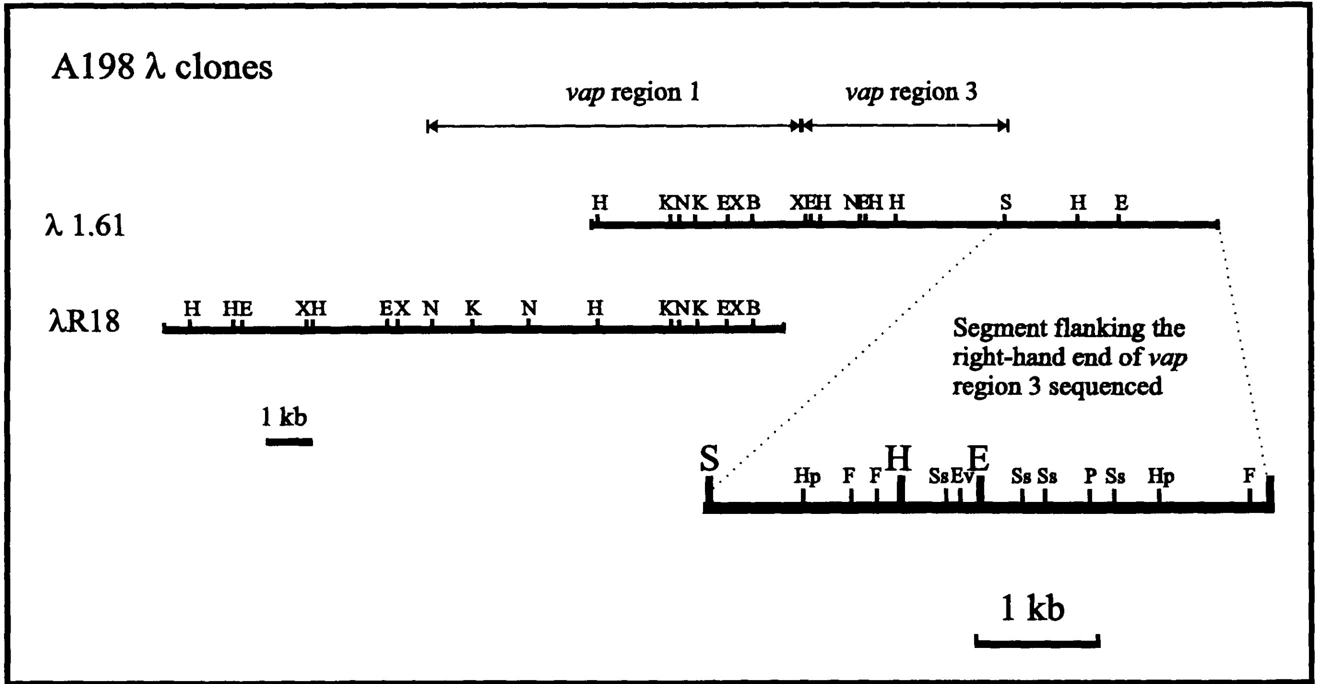

Figure 3.1: Restriction map of two λ clones from the A198 genomic library (Katz *et al*., 1994). The sequence of interest, to the right of *vap* region 3, is enlarged showing some of the more frequent restriction enzyme sites. Restriction endonuclease sites are represented by the letters, *B = Bam*HI, *E = Eco*RI, H = *Hin*dIII, K = *Kpn*I, N = *Nru*I, S = *Sac*I, X = *Xho*I, Ev = *Eco*RV, *F = Fsp*I, *Hp = Hpa*I, P = *Pvu*II and Ss = *Ssp*I.

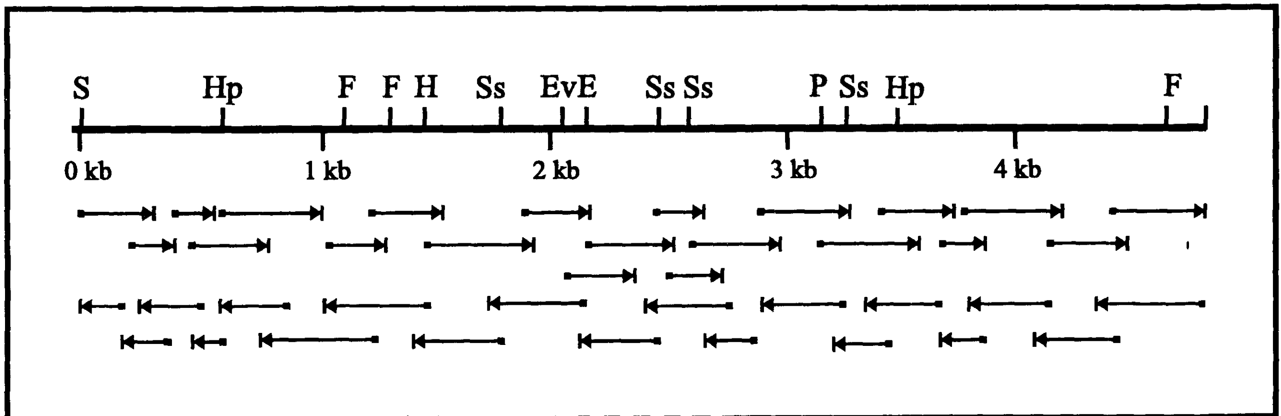

Figure 3.2: Sequencing strategy for the segment flanking the right-hand end of *vap* region 3. Arrows indicate the segments from which the DNA sequence was determined. Restriction enzyme recognition sites are indicated by the letters, E = *Eco*RI, Ev = *Eco*RV, F = *Fsp*I, H = *Hin*dIII, Hp = *Hpa*I, P = *Pvu*II, S = *Sac*I and Ss = *Ssp*I.

sequence around 350 b.p. Both strands of DNA, from the region in question, were determined by sequencing overlapping clones so a continuous sequence could be compiled from each strand. However, subclones were unobtainable from two small areas of the DNA due to a lack of appropriate restriction enzyme sites. Therefore, four oligonucleotides were constructed and used as sequencing primers. This was sufficient for the sequencing to span the subcloning gaps so a continuous sequence could be obtained.

Once the complete sequence was determined the DNA was investigated for open reading frames (ORFs). Five ORFs whose start codons were preceded by Shine-Dalgarno ribosome binding sequences (Shine and Dalgarno, 1974) were identified (Figure 3.3). These putative genes have been termed *intB*, for integrase gene B, *regA*, a putative regulatory gene and *gepA*, *B* and *C* for genetic element proteins A-C. The location, size and orientation of these genes is indicated, adjacent to *vap* region 3, in Figure 3.3. The *intB* and *regA* genes have been named due to their respective similarities to known integrase genes, located on genetic elements, and to regulatory genes. The arrangement of *intB*, *regA* and the *att* site suggest that this DNA region may be an integrated genetic element. This putative element has been designated the "*intB* element" which refers to all sequences from the *attR* site of *vap* region 3 to the end of λ clone λ1.61 (Figure 3.3). This element is shown in relation to the A198 *vap* regions in Figure 3.4.

3.2.2) The integrase gene, *intB*

The open reading frame designated *intB* encodes a putative protein with a deduced amino acid similarity to the P4-like bacteriophage integrase genes, a sub-

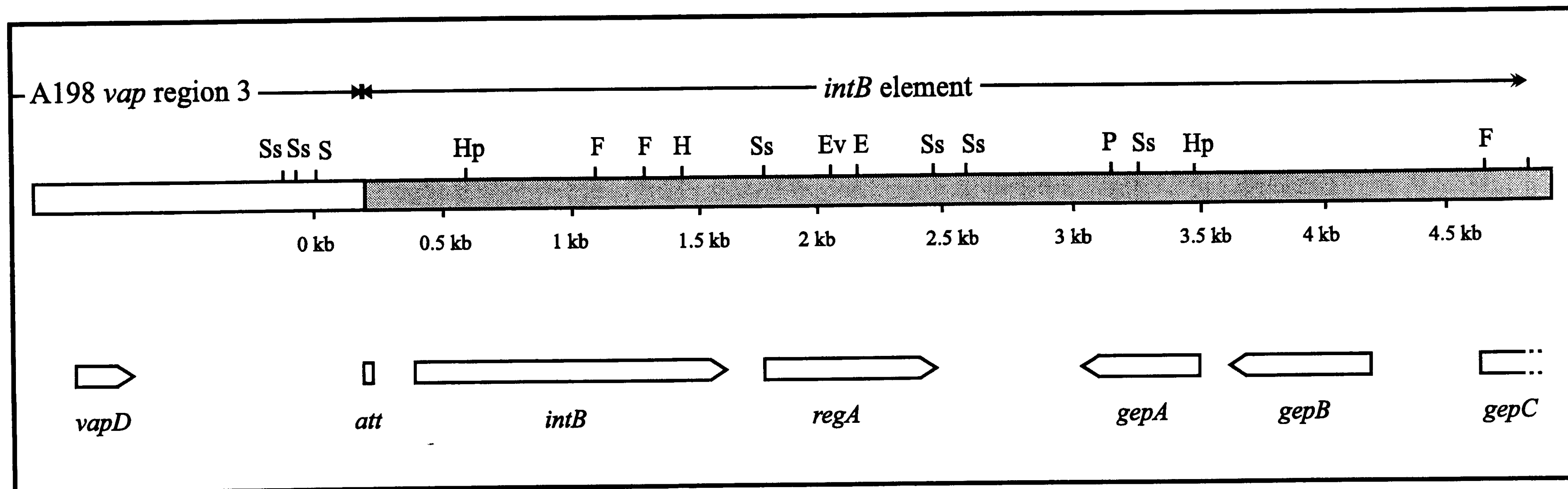

Figure 3.3: Restriction map of the right-hand end of *vap* region 3 and adjacent sequences. The open box indicates the *vap* region DNA while the shaded box represents the adjacent sequences of the *intB* element. Open reading frames are identified by open arrows and the *att* sequence is represented by a small open box. The number of nucleotides is indicated in kilobases (kb) from the *Sac*I site (S). The letters representing restriction endonuclease sites are, E = *Eco*RI, Ev = *Eco*RV, F = *Fsp*I, H = *Hin*dIII, Hp = *Hpa*I, P = *Pvu*II, S = *Sac*I and Ss = *Ssp*I.

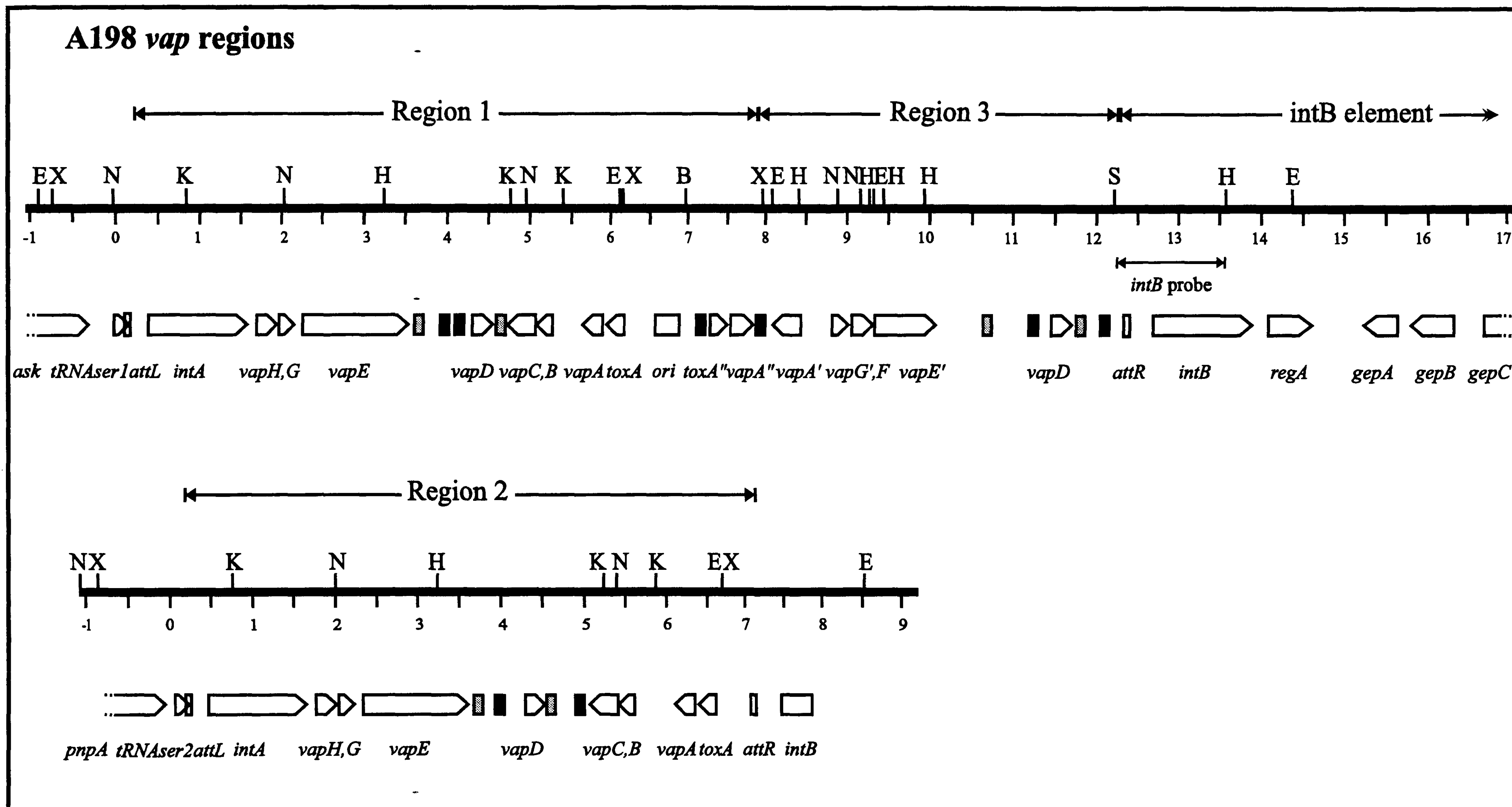

Figure 3.4: Restriction map of the *vap* regions of strain A198, showing the location of the *intB* element. The size in kilobases is indicated on the map. The location, size and orientation of the genes is indicated by the open arrows. *Att* sites are shown by small open boxes, 103 b.p. repeats, or partial copies, are represented by lightly shaded boxes and 102 b.p. repeats, or partial copies, by dark shaded boxes. The location of the *intB* probe is also 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.

group of the λ family of site-specific recombinase (integrase) genes (Argos *et al.*, 1986). However, the *intB* gene is a pseudogene.

Southern blot analysis has shown previously that there are two copies of *intB* in the genome of strain A198 (Bloomfield, 1992). The second copy of *intB*, which was identified by G. Whittle to the right of *vap* region 2, contains only part of the *intB* coding region (Figure 3.4; Bloomfield *et al.*, 1997). Comparison of the sequences of the *intB* genes from the right of *vap* regions 2 and 3 showed that a single adenosine has been inserted into the gene from region 3 (Figure 3.5). When corrected for the frameshift, caused by the insertion, the *intB* gene adjacent to *vap* region 3 appears to be a complete copy of this gene. The incomplete copy of *intB*, to the right of *vap* region 2, has 92.5% identity over 560 nucleotides with *intB* from *vap* region 3. This similarity ranges from the *attR* site of *vap* region 3 and *vap* region 2 up to the point of divergence in the two sequences, 98 b.p. downstream of the adenosine insertion in the *intB* from *vap* region 3 (Figure 3.5).

The amino acid sequence deduced from the *intB* gene next to region 3, continued on in the second reading frame after the insertion, was compared to the amino acid sequences of other integrases. There is 33.8% amino acid identity with the *E. coli* retronphage ϕR73 integrase (Sun *et al.*, 1991), 32.6% amino acid identity with the integrase of *S. flexneri* bacteriophage Sf6 (Clark *et al.*, 1991) and 31.3% amino acid identity with an integrase from *Vibrio cholerae* (Kovach and Peterson, 1994). When aligned, overall, 15.6% of amino acid residues are identical on all four proteins (Figure 3.6). IntB also contains the two conserved regions, domain I and II, essential to the active site of these integrase genes. Domain I contains a conserved arginine and is located 73 to 165 residues (Abremski and Hoess, 1992) from domain II.

```
         v         v         v         v         v         v         v         v         v
TAAGCCATTTATGACCTCATAGAGCCTCAAGGAACCTCACCTAAAGCCAGCGTTATTTGGGGGTATATATAATTCCATGCCAAAAAAATA reg.2  165
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TAAGCCATTTATGACCTCATAGAGCCTCAAGGAACCTCACCTAAAGCCAGCGTTATTTGGGGGTATATATAATTCCATGCCAAAAAAATA reg.3  166

                             intB'     M  G  K  L  T  P  T  A  I  K  C  E  V  A  K  L  G  K
CCCCCAACGATAAAAAACAGTGGTGAGGAAGAGAGTATGGGCAAGTTAACCCCAACAGCAATTAAATGTGAAGTTGCAAAGTTGGGAAAA reg.2  255
:::::::::::::::::::::::::::::::::::::::::::: ::::      ::::::  :::::::::: ::::::::::::::::
CCCCCAACGATAAAAAACAGTGGTGAGGAAGAGAGTATGGGCAAATTAA      CAGCAAGGAAATGTGAAGCTGCAAAGTTGGGAAAA reg.3  250
                             intB      M  G  K  L           T  A  R  K  C  E  A  A  K  L  G  K

 H  F  D  G  E  G  L  Y  L  Y  V  T  E  K  G  K  Y  W  R  R  S  Y  R  F  N  G  K  Q  N  T
CACTTTGACGGTGAGGGTTTATATCTATACGTTACCGAAAAGGGTAAATACTGGCGACGGTCTTACCGTTTCAATGGAAAACAAAACACT reg.2  345
::::::::::::::::::::::::::::::::::: ::::::::::::::::::::       :::::: :  :::: ::  :::: ::
CACTTTGACGGTGAGGGTTTATATCTATACGTTACTGAAAAGGGTAAATACTGGCGGGCAAAGTACCGTATTGATGGTAAGGAAAAGACC reg.3  340
 H  F  D  G  E  G  L  Y  L  Y  V  T  E  K  G  K  Y  W  R  A  K  Y  R  I  D  G  K  E  K  T

 A  A  F  G  V  Y  P  E  T  S  L  L  E  A  R  A  L  N  A  I  F  T  Q  Q  L  K  Q  G  I  D
GCTGCTTTTGGTGTATATCCTGAAACCAGCCTATTGGAAGCGCGCGCGTTGAATGCCATTTTTACGCAACAGCTCAAGCAAGGCATTGAC reg.2  435
:: :: ::::: ::::: :::::    :::::    ::::: ::::      ::::  : ::::   :: :::: :::::::::::::::
GCCGCCTTTGGAGTATACCCTGACGTTAGCCTTGCAGAAGCCCGCGTAAAACATGCTTTATTTAAAAAAGAGCTTAAGCAAGGCATTGAC reg.3  430
 A  A  F  G  V  Y  P  D  V  S  L  A  E  A  R  V  K  H  A  L  F  K  K  E  L  K  Q  G  I  D

 P  N  Y  E  K  R  K  A  K  A  A  K  K  G  L  E  L  A  V  N  G  S  S  P  Q  L  F  R  N  V
CCTAACTATGAGAAACGTAAAGCCAAAGCAGCAAAGAAGGGGCTTGAGTTGGCGGTTAACGGTTCATCACCGCAGTTATTTCGTAATGTA reg.2  525
:::::: ::::::::::::::::::::::::::::::: : ::::::::::: :::::::::::::::::::::::::::::::::::::
CCTAACCATGAGAAACGTAAAGCCAAAGCAGCAAAGAAAGCGCTTGAGTTGGAGGTTAACGGTTCATCACCGCAGTTATTTCGTAATGTA reg.3  520
 P  N  H  E  K  R  K  A  K  A  A  K  K  A  L  E  L  E  V  N  G  S  S  P  Q  L  F  R  N  V

 A  M  D  W  L  E  T  T  H  K  A  K  G  W  T  L  K  H  R  N  D  I     Y  A  N  L  K  N  Y
GCGATGGACTGGCTAGAAACAACCCATAAGGCTAAAGGTTGGACGTTAAAGCATCGCAATGATATTTA CGCCAATCTAAAGAATTACAT reg.2  614
:::::::: ::: :::::::::::::::::::::::: :::::::::::::::::::::::::::::: :: :::::: :::::::::::
GCGATGGATTGGTTAGAAACAACCCATAAGGCTAAAGATTGGACGTTAAAGCATCGCAATGATATTTAACGTCAATCTGAAGAATTACAT reg.3  610
 A  M  D  W  L  E  T  T  H  K  A  K  D  W  T  L  K  H  R  N  D  I  X     V  N  L  K  N  Y

I  L  P  A  F  H  A  R  P  I  E  S  I  T  A  G  E  L  V  T  H  L  R  T  I  P  F  I  F  X
TCTTCCTGCTTTTCATGCGCGCCCTATTGAGTCCATCACTGCTGGAGAACTCGTCACCCATCTGCGAACGATTCCCTTTATTTTTTGATA reg.2  704
::::::::::::::::::::::::::::::::::::::::::::::::::::::: ::::::::: ::  ::::::: :  ::    :
TCTTCCTGCTTTTCATGCGCGCCCTATTGAGTCCATCACTGCTGGAGAACTCGTCGCCCATCTGCAAAGCATTCCCTATGCTTACACAGC reg.3  700
I  L  P  A  F  H  A  R  P  I  E  S  I  T  A  G  E  L  V  A  H  L  Q  S  I  P  Y  A  Y  T

CATCAAAGCCGTTTTTTTCTGCCCAAAGCATCAGTGCTTCAACAACAATCGCCACTTGCGGTATTGAACTTTCTTTGGCGAGCTGCTCCA reg.2  794
   : :  :  :         :          :  ::  :      :::   :: :::  :  :   :       ::    ::::
AGCCTATACGTTGGACAATATCAAGCGCATCTATCGCCACGCGGTTAATATGCAATTGTTGGCTACTAGCCCTGCTGAATTGCTGAAAGC reg.3  790
A  A  Y  T  L  D  N  I  K  R  I  Y  R  H  A  V  N  M  Q  L  L  A  T  S  P  A  E  L  L  K

AAAAGTCAATGTCTTCAATCTTCATTTTTAGCACTTTATTTTTAACCCCACGTTTTTCATCGTTGCGGCGCTGAATTTCGGATTTTGATA reg.2  884
::  :      :      :: :::   : :     :       : : :         :::: :::           :: :   ::  :
AAGTGAACTATTACCTGCTCATCAGGGTGATGCTATGCGACACATCACTGACCCAAGCATCATTGGCAAAGCATTATTAGCCATTGAACG reg.3  880
A  S  E  L  L  P  A  H  Q  G  D  A  M  R  H  I  T  D  P  S  I  I  G  K  A  L  L  A  I  E

ATGCCATGTCTTTAACTTACCTTAGAATTACTGGGGGCGAAGGTGGAAACCACCACCTTGCCCATCTGCCTAGAACGCTGGCGCGCTCAA reg.2  974
    ::              :      :   :    : ::   : :      :        :   :     ::           :
TGCGCACCCTGCCTTGCCGGCAACCCGTGCATACATGAGATTATTGCCTTATCTTTTCACGCGCCCAAGTGAGCTTCGGAAGCTGGAATG reg.3  970
R  A  H  P  A  L  P  A  T  R  A  Y  M  R  L  L  P  Y  L  F  T  R  P  S  E  L  R  K  L  E

TAACAGCAGAAGAATGATTATGATGTATCAGATCATAATGTTCTCCAGTATCCGCCTAACTTCAATAGGTCAAGGCGGTTAGCTATTGGG reg.2 1064
  :         ::::    ::  ::     : ::::  :     : ::::      :           :   : :  ::  : :  :
AGAGTTGGACGCAATGGAAGTGGGGTTAATCACCATACCGGCGCACCGTATGAAGAAACGCCGCCCTCATATCGTCCCTTTACCACGGCA reg.3 1060
W  Q  E  L  D  M  E  V  G  L  I  T  I  P  A  H  R  M  K  K  R  R  P  H  I  V  P  L  P  R

AGGGGTGTGAGTCTCTCGATAGCTATAATAGTAAGCTGCGAAATTGCGCAAGTTGGCGTTTTTGCAAGTTTTTGATATGACCGAACGCTT reg.2 1154
 : :     : ::  :  : :: :       :   : : :     ::      :           : :        :: :: ::: ::
GGCGAAAGCACTCATTGAAGAGATGCGCATCTTTACGGGGCGCACGCCTTTTATCTTTGCCAACGATGACAAGCCAATCACGGAAGGCGC reg.3 1150
Q  A  K  A  L  I  E  E  M  R  I  F  T  G  R  T  P  F  I  F  A  N  D  D  K  P  I  T  E  G

AAAACTGATGTGAAAACAATTGATAAGATTAAGAGGTCAAATTAAAAAGCGTTGGTTTTTTACTGTTATTAATTTGACCACGACATCACC reg.2 1244
       :     :   ::      :   :::: : : :::  :       :   ::    ::  :       : :  :   ::  :  :
GGCTTATAAAATCATGAAAAGCGCTAAGGTAAGCGATGAAACAAGCTTATATCACTTAACCACCTTACACGGCTGGCGCCATACCGCTTC reg.3 1240
A  A  Y  K  I  M  K  S  A  K  V  S  D  E  T  S  L  Y  H  L  T  T  L  H  G  W  R  H  T  A

GATTTTAAGAATATGGGCGTTATCGCGGAATCATTCGAGCTTCTTAAGTAAAAACTGCCCAACAATAGGTTCAATAAAAAAACCAGCTTA reg.2 1334
 :   ::        :      :  :     :::          :              :    : :       :        :   : :
AACCCTATTGCATGAGCAAGGGTATCCATCGCATATTGTGGAAATGCAGCTCGCTCATGCTGACAAAAACAGCGTGCGCGGTACTTATAA reg.3 1330
S  T  L  L  H  E  Q  G  Y  P  S  H  I  V  E  M  Q  L  A  H  A  D  K  N  S  V  R  G  T  Y

ATGCTGGCTTTATCATTTTTGAGTACTAATCAGTCGTTATTTTTCGGCGGCAATGGCGCGCGGGCGGCGGAGGAAAACAGGTAAATTGCC reg.2 1424
          :: :           :   : :      :       ::::       :        :      ::: :     :: : :
CCATGCTGACTACCTAAAAGAGCGGCAGATAATGATGCAAAAGTACGCGGATTATTTGGATAATTTGAAAGCAAAAGCCCTATAACTACT reg.3 1420
N  H  A  D  Y  L  K  E  R  Q  I  M  M  Q  K  Y  A  D  Y  L  D  N  L  K  A  K  A  L  X
```

Figure 3.5: Alignment of the DNA sequences encoding *intB* adjacent to *vap* region 3 (reg.3) of *D. nodosus* strain A198 with the partial copy of *intB*, *intB'*, adjacent to *vap* region 2 (reg.2) of strain A198. Identical nucleotides are indicated by a semicolon and the 'v' indicates every 10 nucleotides. The amino acid sequences are indicated by single letters adjacent to the corresponding codon (the *intB* sequence adjacent to *vap* region 3 is adjusted for the frameshift). The stop codon within *intB* is indicated by an X within the amino acid sequence. The oligonucleotides used to confirm the adenosine insertion are indicated by lines under the DNA sequence.

```
                         v         v         v         v         v         v
IntB            ----MGKLTARKCEAAKLG----KHFDGEGLYLYVTEKG-KYWRAKYR--IDGKEKTAAF   49
PHIR73Int       -----MPLTDIQIKRAKPQNKPYTLNDGQGLSLLINPDGSKGWRFRFR--FAGKARLMSF   53
SF6Int          ------MLTVKQIEAAKPKEKPYRLLDGNGLYLYVPVSGKKVWQLRYK--IDGKEKILTV   52
VCInt           MAKNTVKLTDKEIKASKPREKEYNLFDGDGLRLRIKPNGSKQWIFNYYRPTNGKRANLSF   60
                       **      .*         ** ** * .   * * *   .     **    .

IntB            GVYPDVSLAEARVKHALFKKELKQGIDPNHEKRKAKAAKKALELEVNGSSPQLFRNVAMD  109
PhiR73Int       GSYNLVSLAEAREKRDTARKQVANGIDPVGERKAQKLAQQL-------STENSFETICRE  106
SF6Int          GKYPLMTLQEARDKAWTARKDISVGIDPVKAKKAS-------------SNNNSFSAIYKE   99
VCInt           GMYPDISLLQARKSLLSAKELIAQGIDPQDDRKRQQQVHKE-------IHEHTFENVTKD  113
                * *  ..* .**      .  .  ****   ..                    *  .  .

IntB            WLETTHKAKDWTLKHRNDIXVNLKNYILPAFHARPIESITAGELVAHLQSIP---YAYTA  166
PhiR73Int       WYT--NKADRWTVAYREEIIKTFEQDVFPFIGKRPISEIKPLELLEVLRRIEKRGALEKT  164
SF6Int          WYD--HKKQVWSVGYATELAKMFDDDILPIIGGLEIQDIEPMQLLEVIRRFEDRGAMERA  157
VCInt           WFA--IKQNDVTPDYALDIWRSLELHIFPHISDKPVKAITAPEIIELLKPIEAKGSLETV  171
                *     *    .     ..       . *      .  *   ...  ..

IntB            AYTLDNIKRIYRHAVNMQLLATSPAELLKASELLPAHQGDAMRHITDPSIIGKALLAIER  226
PhiR73Int       RKVRQRCGEVYRYAIITGRAEYNPAPDLAIALAVPKQKHHPFLSAEELPHFIQDLEAYT-  223
SF6Int          NKARRRCGEVFRYAIVTGRAKYNPAPDLADAMKGYRKKNFPFLPADQIPAFNKALATFS-  216
VCInt           KRLTQRLNEVMNFATNCGLIQANPLTGIKAAFKKPKKENMAALTPAELPELMSAIANAS-  230
                         .   *         *   .  .     .                 .
                                    ↓
IntB            AHPALPATRAYMRLLPYLFTRPSELRKLEWQELDMEVGLITIPAHRMKKRRPHIVPLPRQ  286
PhiR73Int       ---GSIITKNATKIVMLTGVRTQEMRFAAWDEVDLERGLWEIPAERMKMRRPHIIPLSTQ  280
SF6Int          ---GSIVSLIATKVLRYTALRTKELRSMQWKNVDFENRIITIDASVMKGRKIHVVPMSDQ  273
VCInt           ---IKRTTRCLLEWQLHTMTRPAEASGARWDEINWEEKVWTIPAERMKKRREHRIPLTEQ  287
                       .            *  *     *  .  *  .  * *  ** *. * .*.  *
                                                                        ↓  ↓
IntB            AKALIEEMRIFTG-RTPFIFANDD---KPITEGAAYKIMKSAKVSDETSLYHLTTLHGWR  342
PhiR73Int       VVDLFNQLKPITR-HYPYIFIGRNNRNKPISKESVSQVIE------LLGYKGRATGHGFR  333
SF6INT          VIELLTTLSSITKPVSEFVFAGRNDKKKPICENAVLLVIK------QIGYEGLESGHGFR  327
VCINT           MLALLEVMKPISG-HRDFIFPSDRAPKKPCNSQTANMALK------RMGFAGRLVSHGLR  340
                   *   .   .     ..*       **    .    .                 ** *
                                                ↓
IntB            HTASTLLHEQGYPSHIVEMQLAHADKNSVRGTYNHADYLKERQIMMQKYADYLDNLKAKA  402
PhiR73Int       HTMSTILHEQGFDSAWIELQLAHVDKNSIRGTYNHAQYLERRREMLQWYSNYLLK-----  388
SF6Int          HEFSTIMNEHEWPADAIEVQLAHANGGSVRGIYNHAQYLDKRREMMQWWADWLDEKVR--  385
VCInt           SLASTTLNEQGFDPDLVESALAHVDDNQVRSAYNRTDYLERRKPMMCWWSEHIEEAARGS  400
                   ** ..*.      .*  ***   . .*  **.. **  *. *.   .  .

IntB            L---------------------  403
PhiR73Int       ----------------------  388
SF6Int          ----------------------  385
VCInt           LSVTGTKQLKNKLPPPPTWKQM  422
```

Figure 3.6: Clustal W alignment (Thompson *et al*., 1994) of the four integrase gene products, IntB, ϕR73 Int, Sf6 Int and an Int from *V. cholerae* 0395. Identical amino acids are indicated by an asterisk, while conserved substitutions are noted by a period. The arrows indicate the essential amino acids conserved in this family of site-specific recombinases (Argos *et al*., 1986; Abremski and Hoess, 1992).

Domain II consists of a histidine and arginine, separated by two amino acids, and a tyrosine residue 34 to 40 amino acids later (Argos *et al.*, 1986; Blakely and Sherratt, 1996). The four invariant residues form the RHRY catalytic tetrad characteristic to this integrase family (Blakely and Sherratt, 1996). The conserved amino acids, indicated by arrows in Figure 3.6, correspond to Arg^{247}, His^{339}, Arg^{342} and Tyr^{375} in IntB. The *intB* product also shows 31.0% amino acid identity with the *intA* gene product from *vap* region 1 of *D. nodosus* (Cheetham *et al.*, 1995). Both integrase genes from *D. nodosus* are similar to the same class of P4-like integrases of the λ family of site-specific recombinases (Argos *et al.*, 1986).

3.2.3) Is there an intact copy of the *intB* gene in *D. nodosus*?

Both the copies of *intB* in *D. nodosus* strain A198 appear to be incapable of producing a functional protein product, since one has been disrupted by a frameshift, and the other copy contains only part of the coding region. Thus, a polymerase chain reaction (PCR) experiment was designed to perform two distinct functions: firstly, to investigate the presence of the *intB* gene in other strains of *D. nodosus*, and secondly, to confirm that *intB* from strain A198 is a pseudogene..

The location of the two primers is indicated in Figure 3.5. The DNA sequence of the two 20 base oligonucleotides is: primer 1 - 5' GGTTAACGGTTCATCACCGC 3' and primer 2 - 5' CGCGTGCAAGTTACCGATTA 3'. These primers are located on either side of the point of divergence between *intB* from *vap* region 3 and the partial copy from *vap* region 2 of strain A198 (Figure 3.5). Hence, the PCR reaction should indicate whether other strains of *D. nodosus* contain the complete version of *intB*, adjacent to *vap* region 3 in A198, or the truncated version, adjacent to *vap* region 2. The primers are also located on either side of the base insertion identified as the cause

of the interruption to the *intB* reading frame (Figure 3.5). Therefore, the sequence of the PCR product should confirm the status of *intB* in A198 and other strains. The primers were used to amplify this segment of the *intB* gene from genomic DNA from several strains of *D. nodosus* including: the virulent strains A198, B1006, G1220 and H1215, the intermediate strain AC3577 and the benign strains C305 and H1204 (Figure 3.7).

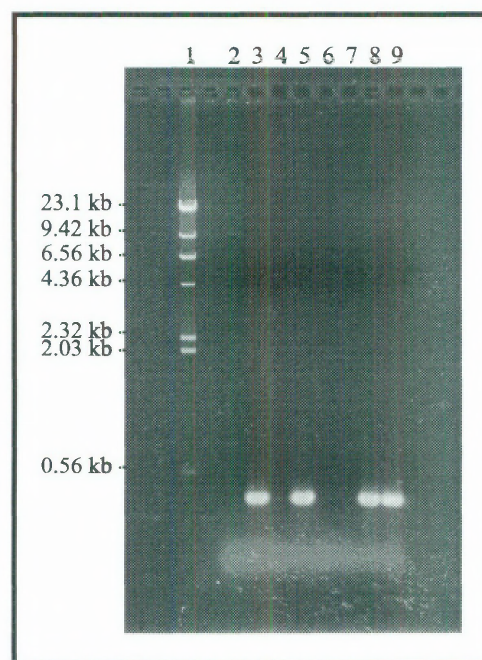

Figure 3.7: PCR products from seven strains of *D. nodosus* using the oligonucleotides spanning the disruption of the *intB* gene (Fig. 3.5) as primers. Run in each lane of the agarose gel are; 1, λ DNA digested with *Hin*dIII; 2, negative PCR control; 3, DNA from A198 (positive control); 4, DNA from C305; 5, DNA from AC3577; 6, DNA from B1006; 7, G1220; 8, H1205 and 9, H1215.

Southern blot analysis has shown that *intB*, or part thereof, is present in each strain of *D. nodosus* tested (Bloomfield *et al.*, 1997). Seven strains of *D. nodosus*, known to hybridise to a probe which detects part of *intB* (Figure 3.4), were investigated for the presence of intact *intB* genes by using the PCR experiment. PCR products of the expected size were obtained from strains A198, AC3577, H1204 and

H1215 (Figure 3.7), indicating a likelihood of complete *intB* genes in these strains. The absence of PCR products in strains C305, B1006 and G1220 (Figure 3.7) indicates that there is no complete copy of the *intB* gene. Therefore, these strains may contain only the partial copy of *intB* found to the right of *vap* region 2 in A198 (Figure 3.4).

The 400 b.p. PCR products obtained from *D. nodosus* strains A198, AC3577, H1204 and H1215 were cloned into pUC18 and their DNA sequence was determined, to investigate the presence of the insertion identified in the previously-sequenced *intB* gene. This single adenosine insertion which disrupts *intB* in A198 was found to be present in the A198 PCR product but absent from the PCR products from the other three strains. Therefore, *D. nodosus* strains AC3577, H1204 and H1215 may contain an intact copy of the *intB* gene and possibly a functional copy of this putative *intB* genetic element. An alternative method to investigate the disruption of *intB* would have been to clone the gene from A198 and other strains of *D. nodosus* into an expression system, using for example T7 RNA polymerase (Ausubel, 1995). The size of the resulting protein products would indicate if the frameshift was a mutation in the gene or a sequencing/cloning artefact.

3.2.4) Identification of a regulatory gene

The ORF adjacent to *intB*, termed *regA*, has deduced amino acid similarity to bacterial and bacteriophage regulatory proteins. The 232 amino acids of RegA have 40.5% amino acid identity with the pectin lyase regulator RdgA of *Erwinia carotovora* (Figure 3.8A; Liu *et al.*, 1995) and 39.2% amino acid identity with the regulatory protein PrtR of *Pseudomonas aeruginosa* (Matsui *et al.*, 1993). RdgA is believed to repress both itself and *rdgB*, a regulatory gene which activates pectin lyase

production. Thus, RdgA represses the production of pectin lyase. Pectin lyase is one of the extracellular enzymes produced by *Erwinia* spp for the degradation of plant cell wall components (Barras *et al.*, 1994). The *prtR* gene product of *P. aeruginosa* represses the expression of the *prtN* gene, which is an activator of pyocin synthesis, thus PrtR represses pyocin synthesis. Pyocins are bacteriocins, produced by *Pseudomonas* spp, which inhibit or kill closely related bacteria (Brock and Madigan, 1988).

Both RdgA and PrtR are similar to the repressor genes of bacteriophages such as ϕ80. The deduced amino acid sequence of *regA* also has 36.2% amino acid identity with the *cI* repressor gene product of bacteriophage ϕ80 of *E. coli* (Ogawa *et al.*, 1988). The alignment of the complete amino acid sequences from RegA, RdgA, PrtR and ϕ80cI is shown in Figure 3.8B. Throughout all four proteins there is 19.8% identity and 38.4% amino acid similarity. As can be seen in Figure 3.8B the similarity between these four genes is primarily in the C-terminus although there is a small, but highly significant, region of similarity at the N-terminus.

The function of the cI gene product has been most extensively studied in bacteriophage λ. This function is also believed to be similar in other related bacteriophages. In bacteriophage λ the *cI* gene product binds to the operators at either end of the *cI* and *rex* genes, preventing transcription of the early λ genes and hence repressing phage activity, i.e. blocking phage particle development, and maintaining the lysogenic state (Echols and Guarneros, 1983; Roberts and Devoret, 1983).

```
A.
                          v         v         v         v         v         v
RdgA             MKTTLAERLKTARTAQGLSQKALGDMIGVSQAAIQKIEVGKASQTTKIVELSNNLRVRPE   60
RegA             --MSLASNVKILRELNNLSQDQLAEKIGKSQAAIQKIEAGLTLRPRFLQDLANALGVSSI   58
                    .**  .*  *  ..***  * . ** ********* * . .   . .*.* * *

RdgA             WLANGEGPMRSSEVTRSLQEPSIPPKSEWGTVSAWDSTTELSEDE-VEVPFLKDIEFACG  119
RegA             DLEYKD---FEKELKKQAIESDIG---TMGKFRLWSSNDPLPEDEYAYLPFFKDVEFQGG  112
                  *   .      *. .   *  *      *    * *.  * ***   .** **.**  *

RdgA             DGRIQSEDYNGFKLRFSKATLRKVGANTDGSGVLCFPAAGDSMEPIIPDGTTVAVDTNNK  179
RegA             TGCCEMQDYNGFRLPFAKSTLHRYGVPLD--QAFCVTLTGNSMEPVIPKGSTLGINKADT  170
                  *  . .*****.* *.*.**.. *   *     *   .* ****.** *.*. .

RdgA             RIIDGKLYAIAQEGGGNDKLKRIKQLYRKPGGLLTIHSFNRE--TDEEAYESDVEIIGRV  237
RegA             VLKEGDIYAIRQ-----DDLFRVKRLYHAPNGMIRISSFNQEEYKDELVRPENIEIIGRV  225
                  . .* .*** *     * * *.*.**. *.*...* ***.*   **      .******

RdgA             FWYSVLL  244
RegA             FTYQVML  232
                 * * *.*
-------------------------------------------------------------------------------
B.
Phi80c1          MVMEKIHMLFTMHGAEGFMSSISERIKFLLAREGLKQRDLAEALSTSPQTVNNWIKRDAL   60
RdgA             -----------------MKTTLAERLKTARTAQGLSQKALGDMIGVSQAAIQKIEVGKAS   42
PrtR             --MDKSTQIP--------PDSFAARLKQAMAMRNLKQETLAEAAGVSQNTIHKLTSGKAQ   49
RegA             -------------------MSLASNVKILRELNNLSQDQLAEKIGKSQAAIQKIEAGLTL   40
                                     . .  .*      .* *  * .    *  ..       .

Phi80c1          SREAAQQISEKFGYSLDWLLNGEGSPKKD------LESNIPPES--EWGTVDAWDKNTPL  112
RdgA             QTTKIVELSNNLRVRPEWLANGEGPMRSSEVTRSLQEPSIPPKS--EWGTVSAWDSTTEL  100
PrtR             STRKLIEIAAALGVSPVWLQTGEGAPAAR------SAVSVADGSPLVLEPLHPWDSDTPL  103
RegA             RPRFLQDLANALGVSSIDLEYKDFEKELK-------KQAIESDIG-TMGKFRLWSSNDPL   92
                        ..         *   .                .             *     *

Phi80C1          PDDE-VEVPFLKDIEFACGDGRVHDEDHNGFKLRFSKATLRRVGANSDGSGVLCFPASGD  171
RdgA             SEDE-VEVPFLKDIEFACGDGRIQSEDYNGFKLRFSKATLRKVGANTDGSGVLCFPAAGD  159
PrtR             DEDE-VELPLYKEVEMSAGAGRTAVREIEGRKLRFSYATLRASGVDP--SAAICAQLTGN  160
RegA             PEDEYAYLPFFKDVEFQGGTGCCEMQDYNGFRLPFAKSTLHRYGVPLD--QAFCVTLTGN  150
                  .**   .*  *..*   * *     .  * .* *. .**.  *         *   .*

Phi80c1          SMEPVIPDGATVAVDTGNKRVIDGELYAINQ---GD--LKRIKQLYRKPGGKILIRSINR  226
RdgA             SMEPIIPDGTTVAVDTNNKRIIDGKLYAIAQEGGGNDKLKRIKQLYRKPGGLLTIHSFNR  219
PrtR             SMEPLIMDGSTIGVDTATTHITDGEIYALEHDG-----MLRVKFVYRLPGGGIRLRSFNR  215
RegA             SMEPVIPKGSTLGINKADTVLKEGDIYAIRQDD-----LFRVKRLYHAPNGMIRISSFNQ  205
                 ****.*  *.*. .      . .* .**. .       . *.* .*. *.* . . * *.

Phi80c1          DYD-DEEADEADV-----EIIGFVFWYSVLRYRR------  255
RdgA             ET--DEEAYESDV-----EIIGRVFWYSVLL---------  244
PrtR             EEYPDEEYSPEDMRSRQISMIGWVFWWSTVRHRRGPSLVR  256
RegA             EEYKDELVRPENI-----EIIGRVFTYQVML---------  232
                 .   **      .      .** **    .
```

Figure 3.8: Clustal W multiple alignment (Thompson *et al.*, 1994) of the regulatory proteins ϕ80cI, RdgA, PrtR and RegA. (**A**) RegA aligned with RdgA, Leucine residues in bold indicate a potential leucine zipper motif. (**B**) All four negative regulators aligned. 'v' indicates every 10 amino acids. The period below the sequences indicates a conservative substitution while the asterisks mark the identical amino acids in all proteins.

Two secondary structures have been identified from the amino acid sequence of RegA. A helix-turn-helix DNA binding motif (Brennan and Matthews, 1989), within

the first 60 amino acids of RegA, and a putative leucine zipper motif which is known to mediate dimerisation (Jones, 1990), between amino acids 120 and 150. The amino terminal 60 amino acids of RegA are 42% identical to the DNA-binding domain of the cI repressor of bacteriophage 434, residues 1-69 (Neri *et al.*, 1992). The similarity to the 434 repressor dissipates after this region. There is also 37.6% identity to the DNA-binding region of the 434 Cro protein (Grosschedl and Schwarz, 1979). This similarity corresponds to the helix-turn-helix DNA binding motif common to these proteins (Figure 3.9; Sauer *et al.*, 1982; Ogawa *et al.*, 1988; Brennan and Matthews, 1989). Figure 3.9 shows the N-terminal amino acids corresponding to the helix-turn-helix motifs of a number of DNA binding proteins including RegA. The residues in bold are the highly conserved sites within the binding motif. These are alanine or glycine within the first α-helix, glycine in the turn and isoleucine in the second α-helix (Brennan and Matthews, 1989).

```
                                      Helix  Turn  Helix
                                    <-------> <-> <------->
434CI     ---------MSISSRVKSKRIQLGLNQAELAQKVGTTQQSIEQLENGKTKRPRFLPELAS   51
434Cro    ---------MSISSRVKSKRIQLGLNQAELAQKVGTTQQSIEQLENGKTKRPRFLPELAS   51
PrtR      MDKSTQIPPDSFAARLKQAMAMRNLKQETLAEAAGVSQNTIHKLTSGKAQSTRKLIEIAA   60
RdgA      -------MKTTLAERLKTARTAQGLSQKALGDMIGVSQAAIQKIEVGKASQTTKIVELSN   53
RegA      ---------MSLASNVKILRELNNLSQDQLAEKIGKSQAAIQKIEAGLTLRPRFLQDLAN   51
                    . .  .*      .* *  *    * .* .* ..  * .     . ...

434CI     ALGVSVDWLLNGTS---DSNVRF-...  71
434Cro    ALGVSVDWLLNGTS---DSNVRFX     71
PrtR      ALGVSPVWLQTGEG---APAARS-...  80
RdgA      NLRVRPEWLANGEGPMRSSEVTR-...  76
RegA      ALGVSSIDLEYKDF---EKELKKQ...  72
           * *    *
```

Figure 3.9: Clustal W alignment (Thompson *et al.*, 1994) of the initial 70-80 amino acids of five DNA binding proteins. The complete amino acid sequence of bacteriophage 434 Cro is shown. The helix-turn-helix DNA binding motif is indicated above the sequence with the highly conserved amino acids shown in bold. The period below the sequence indicates a conservative substitution and the asterisks indicates and identical residue in all five proteins

The N-terminal domain of these regulatory genes is associated with DNA binding, as shown above, while the C-terminal domain of the *cI* genes is known to be

associated with dimerisation (Lewin, 1990). The RegA protein has a leucine-zipper like structure between amino acids 120 and 150. A leucine zipper is a structure which enables two monomer proteins to form a dimer. Dimerisation in the *cI* gene is essential for binding to DNA. The most distinctive characteristic of a leucine-zipper domain is a leucine residue every seven amino acids (Vinson *et al.*, 1989). This allows the leucine residues to line up on one side of an α-helix and form a dimer with another leucine zipper. In RegA there are three leucines at seven amino acid intervals, residues 126, 133 and 140, followed by a fourth leucine at eight amino acids from the third, residue 148.

3.2.5) *Gep* genes

Three further ORFs were identified downstream of the *regA* gene and may be part of the *intB* element. The putative genes have been termed *gepA-C* (genetic element protein A - C). The *gepA* gene product has 33.3% amino acid identity to *orf4*, downstream from the *rteC* gene of a conjugative transposon from *Bacteroides thetaitotaomicron* (Stevens *et al.*, 1993). The *orf4* gene has an unknown function but is located within the central regulatory region of the tetracycline resistance conjugative element (Stevens *et al.*, 1993). Conjugative transposons are an unusual group of mobile genetic elements which are found integrated into bacterial genomes (Salyers *et al.*, 1995). The integrases of some conjugative transposons are members of the λ integrase family (Salyers *et al.*, 1995), as is *intB*, which raises the possibility that the *intB* genetic element is a conjugative transposon.

The *gepB* gene encodes a putative protein, 198 a.a., with 61.6% amino acid similarity to an unidentified ORF in the *cob* gene cluster from *Pseudomonas denitrificans* (Crouzet *et al.*, 1991). The genes of the *cob* cluster encode the enzymes

of the cobalamin biosynthetic pathway for the production of coenzyme B_{12} from uroporphyrinogen III which is a common intermediate in the synthesis of cobalamins, haems and chlorophylls (Crouzet *et al.*, 1991). The unidentified ORF of the *cob* cluster, similar to *gepB*, lies between *cobQ* and *cobP* and is encoded on the opposite DNA strand. This relatively large intergenic region, 1,481 b.p., also contains two direct repeats of 25 b.p. which lie just prior to the start of the ORF with similarity with *gepB*. These repeats, absent from the *D. nodosus* sequence, were suggested as a possible binding site for a regulatory protein (Crouzet *et al.*, 1991). The *gepB* and *cob* ORF start codons do not correspond but the ORFs terminate at the same location. No regulatory genes for this group of *cob* genes have been identified (Crouzet *et al.*, 1991).

The partially sequenced *gepC* gene, encoding only 67 amino acids of the amino terminus of the putative protein, shows no significant similarity to any genes or proteins in the databases.

3.3) Discussion

The work described in this chapter has identified another putative genetic element integrated into the genome of the bacterial pathogen *D. nodosus*. Previous work has identified two separate regions of the *D. nodosus* genome, the *vap* and *vrl* regions (Katz *et al.*, 1992; Rood *et al.*, 1994), which are present in most virulent strains but absent from most benign strains (Katz *et al.*, 1991; Rood *et al.*, 1996). Both of these virulence-associated sequences have been shown to have a plasmid or bacteriophage-like ancestry and are believed to have been acquired by *D. nodosus* via integration into the bacterial genome (Cheetham *et al.*, 1995; Haring *et al.*, 1995).

The partial characterisation of the *intB* element, adjacent to the *vap* regions, indicates the presence of a site-specific recombinase gene, *intB*, a regulatory gene, *regA*, and three genes of unknown function, *gepA-C*. The *intB* gene is similar to P4-like integrases, a subgroup of the integrase family, members of which have been identified on plasmids (Boccard *et al.*, 1989; Brown *et al.*, 1990, 1994), conjugative transposons (Salyers *et al.*, 1995) and integrons (Stokes and Hall, 1989) as well as bacteriophages. The *regA* gene is similar to a bacteriophage-like repressor and *gepA* has similarity to an ORF from the regulatory region of a conjugative transposon.

Southern blot analysis has shown that the *intB* element, or part there of, is present in every strain analysed thus far (Bloomfield *et al.*, 1997). Further hybridisation studies have shown that *intB*, *regA* and the *gep* genes are part of an integrated element which is present in the majority of *D. nodosus* strains tested (Whittle and Cheetham, unpublished data). Therefore, the *intB* element may have become a ubiquitous part of

the *D. nodosus* genome. This *intB* element is common to both benign and virulent strains, unlike the adjacent *vap* regions which are found primarily in virulent strains.

3.3.1) The site-specific integrase, *intB*

One copy of *intB* from strain A198 is a pseudogene, and the other copy lacks the C-terminal coding region, so strain A198 does not contain an intact copy of *intB*. However, a PCR experiment showed that at least three other strains appear to have an intact copy of *intB*, and therefore may possibly have a functional copy of the *intB* genetic element. Southern blot analysis has shown that many other strains contain at least part of *intB*, but lack of a PCR product from three of the seven strains tested indicates that part of the *intB* gene has been deleted or rearranged in those strains. Deletions and rearrangements have also occurred in the *vap* regions of strain A198 and are not uncommon to this segment of the *D. nodosus* genome (Cheetham *et al.*, 1995; Bloomfield *et al.*, 1997). The strains of *D. nodosus* lacking the complete *intB* gene may contain only the partial copy of *intB* adjacent to *vap* region 2 of A198 (Figure 3.4). One possible explanation is that *vap* region 2 is a separate genetic element (Bloomfield *et al.*, 1997) which contains part of the *intB* gene. Thus, some *D. nodosus* strains may contain only *vap* region 2 with the corresponding partial *intB* gene, and other strains may contain *vap* region 1 or *vap* region 3 or combinations of the three *vap* regions. The multiple insertions of the *vap* regions, and the existance of different *vap* elements, indicates the complex evolution which has taken place to form the *D. nodosus vap* regions.

The translated *intB* gene has high similarity to a number of P4-like integrases (Figure 3.6). Therefore, IntB is related to a large number of bacteriophage site-specific recombinase genes including λ, P2, φ80, P22, P4, 186 and P1 (Argos *et al.*,

1986). This class of integrase gene is not only encoded by bacteriophages. Related genes have been identified on a number of plasmids, conjugative transposons and integrons. Therefore, the identification of an integrase gene only indicates the involvement of an extra-chromosomal genetic element in the evolution of this region and not the exact nature of this element.

The presence of at least part of the *intB* gene in every strain tested suggests that the integration of this element occurred in an ancestral strain of *D. nodosus* prior to divergence of the strains. The non-uniform strain distribution of the *vap* regions indicates a more recent insertion into a proportion of the *D. nodosus* population, which may have enhanced the virulence of those strains. An alternative explanation is that the *vap* regions were inserted in an ancestral *D. nodosus* strain but have since been lost from many lineages, possibly leaving these strains less virulent. Preliminary evidence from one benign strain indicates that the *vap* regions were present in an ancestor of this lineage, but have been mostly lost from the genome of this strain (Chapter 4). However, the *vap* regions contain the *toxA/vapA* maintenance system which should act to prevent the loss of the *vap* regions (Chapter 1).

3.3.2) The regulatory gene, *regA*.

The gene identified 194 b.p. downstream of *intB*, termed *regA*, is similar to regulatory genes from *E. carotovora*, *P. aeruginosa* and bacteriophage ϕ80. Regulation in these three systems is very similar, i.e. a repressor or negative regulator, prevents transcription of the activator, or positive regulator. The bacteriophage-like regulatory systems of *E. carotovora* pectin lyase, regulated by RdgA and RdgB, and *P. aeruginosa* pyocin, regulated by PrtR and PrtN, are both inducible by agents that damage DNA. Damage-inducible genes are often associated with DNA repair

mechanisms, such as the *SOS* genes, of many bacteria (Walker, 1985; Garriga *et al.*, 1992). However, induction of genes which may enhance survival of the bacteria, such as pectin lyase or pyocin, would also be advantageous to survival of the bacteria after exposure to damaging agents. Induction of gene activity by DNA damage also occurs when a bacteriophage is induced from the prophage state (Ackermann and DuBow, 1987). This observation fits well with the similarity of *regA*, *rdgA* and *prtR* to bacteriophage regulators. However, the induction of a bacteriophage from the prophage state, by DNA damaging agents, is believe to be due to the actions of the bacterial repair mechanisms (Ackermann and DuBow. 1987).

Another interesting feature of these negative regulators is that the genes which are regulated all encode extracellular products. The extracellular pectin lyase is an important determinant of pathogenicity in *Erwinia* spp (Barras *et al.*, 1994). The high similarity of *regA* with *rdgA*, the repressor of pectin lyase, suggests that a putative RegA may down-regulate the production of extracellular enzymes in *D. nodosus*, possibly proteases involved in virulence. Only remnants of the *intB* element are present in the *D. nodosus* benign strain C305. The *regA* gene in this strain is a pseudogene (Chapter 4), as is the *intB* gene, which suggests the *intB* element, although present in both benign and virulent strains may be non-functional in benign strains. Therefore, the *intB* element may have a role in virulence even though the DNA may not be virulence-associated. Experiments are currently underway to investigate the induction of *regA* by mitomycin C (Cheetham, unpublished). Induction of extracellular proteases by mitomycin C is also being investigated.

Pyocins of *P. aeruginosa* are also extracellular products involved with survival of the bacteria. Bradley (1967) proposed that some bacteriocins, including pyocins, have evolved from defective bacteriophages, with members of the bacteriocin family

resembling bacteriophage tails when examined by electron microscopy. This may provide a further link between *prtR* and pyocins with bacteriophages, in addition to the similarity in regulatory mechanisms. The similarity of *regA* to *rdgA*, pectin lyase regulator, *prtR*, pyocin regulator, and *cI*, ϕ80 regulator, suggests that the *intB* element may be a bacteriophage, or may be derived from a bacteriophage.

This is supported by the arrangement of genes and the putative insertion site. Assuming this genetic element has inserted into the same site as the *vap* region, i.e. the 3' end of the *tRNA-ser* gene, the arrangement of *att* site, *int* and *cI* genes (*att-int-cI*) is the same as that of bacteriophage λ and P2 (Calendar *et al.*, 1981). However, the similarity between the *gepA* gene and *orf4* from a conjugative transposon, raises the possibility that the *intB* element is a conjugative transposon. At present it is not possible to distinguish between these two possibilities.

The group of negative regulatory elements or repressors to which *regA* is similar all have a positive regulatory element downstream i.e. *cI* & *cro*, *prtR* & *prtN* and *rdgA* & *rdgB*. The orientation and the distance between the genes varies but we may well expect to find a positive regulatory gene downstream of *regA*. The positive regulator downstream of *prtR*, *prtN*, has no similarities to known regulatory genes (Matsui *et al.*, 1993) thus, a positive regulator may not be identified downstream of *regA* by sequence similarities. The *gepA* gene may be a good candidate for a positive regulator downstream of *regA* since there is similarity to a gene, with unknown function, from a regulatory region of a conjugative transposon. The putative *gepA* regulatory gene may be the positive regulator controlled by the *regA* gene product as part of a regulatory locus of the *intB* element.

3.3.3) *gep* gene similarities

The 156 a.a. of the putative GepA protein have similarity to *orf4*, downstream of *rteC*, on the tetracycline resistant conjugative transposon (Tc^r element) from *Bacteroides* spp (Stevens *et al.*, 1993). The *rteC* gene itself is located downstream of the *rteA-rteB* regulatory locus, which is known to mediate tetracycline regulation of the conjugative transposons genes (Stevens *et al.*, 1992) and also regulate *rteC*. The *rteC* gene product is involved in the regulation of self transfer of the conjugative transposon (Li *et al.*, 1995) and to a lesser extent the mobilization of coresident plasmids (Stevens *et al.*, 1993).

Orf4 lies adjacent to the Tc^r elements' regulatory region (Stevens *et al.*, 1993) and may have an as yet unknown regulatory function. This gene similarity suggests that the *intB* element may have evolved from a conjugative transposons. Members of the integrase gene family, to which *intB* is a member, are also known to be encoded by conjugative transposons (Salyers *et al.*, 1995). Therefore, the presence of *intB* may further support the transposon origin of the *intB* element.

The *vap* regions, including the *vap* plasmid, do not appear to encode genes required for conjugation. Recently another non-conjugative plasmid, pDN1, has been isolated from a strain of *D. nodosus* (Whittle and Cheetham, unpublished). A non-conjugative plasmid may be transferred to another bacteria by a bacteriophage or conjugative transposon. Conjugative transposons are able to transfer small non-conjugative plasmids into other strains and species of bacteria (Valentine *et al.*, 1988). The plasmids mobilized by conjugative transposons are similar to the plasmids mobilized by the IncIα and IncP groups of conjugative plasmids. These plasmids mobilize IncQ plasmids. The pDN1 plasmid from *D. nodosus* is related to the IncQ plasmids (Whittle and Cheetham, unpublished). Thus, a conjugative transposon-like

intB element may have been involved in the transfer of either the *vap* regions or pDN1 into *D. nodosus*.

The *Bacteroides* conjugative transposons can also excise and circularize discreet unlinked mobilizable elements from the bacterial chromosome (Shoemaker and Salyers, 1988). These mobilizable elements are also involved in the transfer of antibiotic resistance and are able to insert into plasmids to render them mobilizable by conjugative transposons (Salyers *et al.*, 1995). However, no antibiotic resistance genes have been identified in relation to the *vap*, *vrl* or *intB* elements of *D. nodosus*.

The *P. aeruginosa* ORF from the *cob* cluster, with high similarity to *gepB*, was not identified by the investigating researchers (Crouzet *et al.*, 1991) and lies in an intergenic region between *cobQ* which encodes cobyric acid synthase, and the *cobP* gene for cobinamide kinase-cobinamide phosphate guanylyltransferase, enzymes essential to the cobalamin biosynthetic pathway. The unidentified *cob* ORF has an unknown function. The high conservation of this gene between *P. aeruginosa* and *D. nodosus* indicates a possible functional significance. Crouzet *et al.* (1991) were unable to identify any regulatory genes for the cobalamin biosynthetic pathway. A regulatory role for *gepB*, and possible *gepC*, would suit the theories proposed for this regulatory region of the *intB* element.

The *intB* element or part thereof is present in all strains tested (Bloomfield *et al.*, 1997). Following the initial insertion of the *intB* element, genomic rearrangements have been responsible for the disruption and deletion of sections of this gene and possibly sections of the element in many strains. From the data collected it is clear that two distinct but related genetic elements have been incorporated into the *D. nodosus* chromosome at the same point, the 3' end of a *tRNA-ser* gene. Insertion of

the two elements may have been linked although the insertion of the *intB* element most likely occurred prior to divergence of the strains, since part of these sequences are present in all strains. The origin of this second genetic element appears to be either a bacteriophage or a conjugative transposon although the distinctions between the various forms of genetic elements, i.e. some bacteriophages, plasmids and transposable elements, has become increasingly unclear with many overlapping characteristics. The arrangement of the *vap* regions and the *intB* element differ between strains, with multiple copies of both in some strains. This illustrates the fluidity of the *D. nodosus* genome, and the importance of mobile genetic elements in its evolution. Further characterisation of this region may reveal the evolutionary history of these genomic sequences.