CHAPTER 7

Characterisation of the *intC* element of *Dichelobacter nodosus*

"One of the most intriguing facts about the history of mobile genetic elements is that no-one set out to discover them. In every case, they were initially found by geneticists studying other problems: chromosome breakage and rejoining in maize, transfer of genetic information between bacterial cells, lysogeny, regulation of gene expression, or antibiotic resistance. The message is . . . .the need to be alert for those unexpected findings which overturn those assumptions and suddenly open a whole new universe for scientific exploration"

James A. Shapiro, 1995

7.1 Introduction

In previous work, a library of genomic DNA from benign strain C305 was prepared (Bloomfield, 1997) and probed with sequences that flank the left- and right-hand ends of *vap* regions 1 (*askA*) and 3 (*intB*) of *D. nodosus* strain A198 (Figure 7.1). A lambda clone that hybridised to both *askA* and *intB* was isolated (λGB321). Subsequent analysis revealed that *askA* and *intB* were not adjacent in strain C305, but were instead separated by approximately 7 kb of sequence (Section 1.6.8.1) The sequence of an 8035 bp region from this lambda clone was determined, and several orfs were identified (Figure 7.1). A description of these open reading frames and a summary of ho nologues identified to date is provided in Table 7.1. Subsequently, the presence of a new genetic element, designated the *intC* element was postulated (Bloomfield, 1997).

In this study, Southern blot analyses and PCR experiments were undertaken in order to determine whether the genes *intC*, *orf242*, *orf171*, *vapG"*, *vapH"* are present in other strains of *D. nodosus* and if so, whether genes *intC*, *orf242*, *orf171*, *vapG"*, *vapH"* are

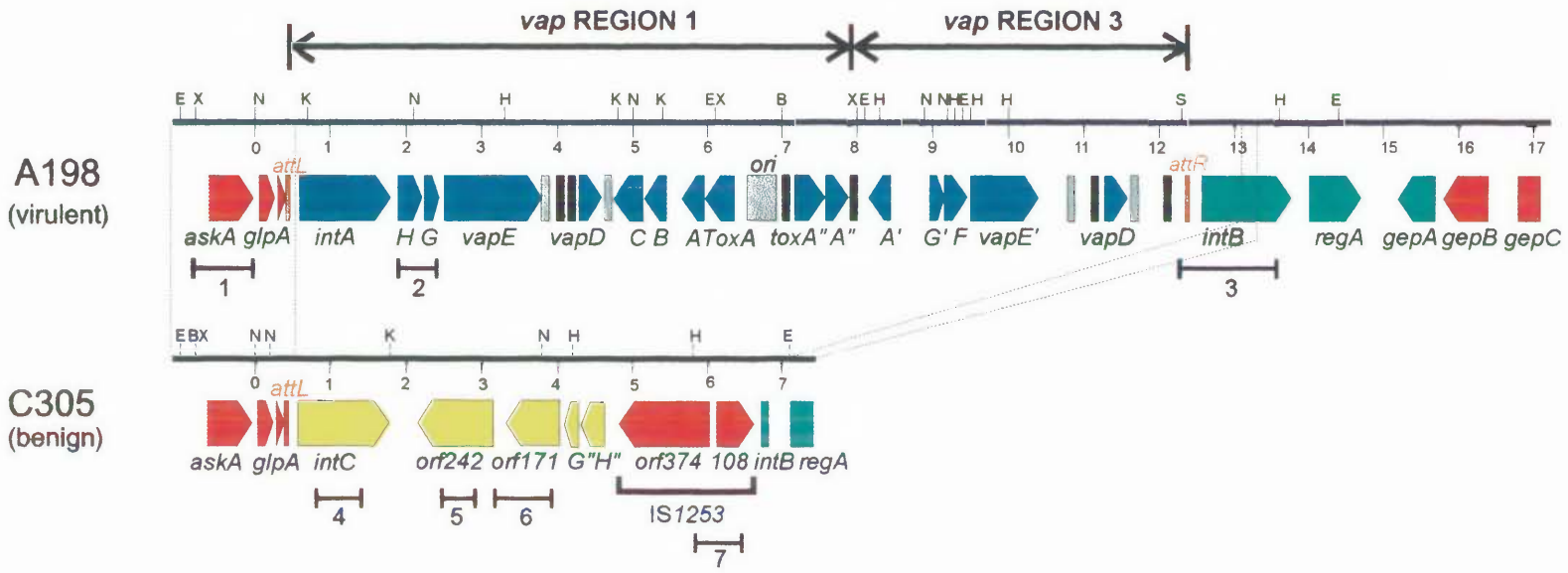


Figure 7.1: Alignment of restriction maps of sequences downstream of *askA* in *D. nodosus* virulent strain A198 and in *D. nodosus* strain C305. Probes are indicated by bold lines below the corresponding sequences from the *vap* element of *D. nodosus* strain A198 and the *intC* element of strain C305 and are numbered as follows: *askA* (1), *vapHG* (2), *intB* (3), *intC* (4), *orf242* (5), *orf171* (6), IS*1253* (7). The numbers indicate the distance in kb. Restriction enzyme sites shown are *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Kpn*I (K), *Nru*I (N), *Sac*I (S) and *Xho*I (X). Open reading frames are indicated by shaded arrows and are classified as follows: belonging to the *intB* element (green), belonging to the *vap* region (blue), belonging to the *intC* element (yellow) genes that are not part of an of the *vap*, *intB* or *intC* elements (red), including *tRNA* genes (red triangle). Repeated sequences are as follows: attachment sites, *attL* or *attR* (narrow red boxes), 102 bp repeats or partial copies thereof (filled boxes), 103 bp repeats or partial copies of (narrow grey boxes), and a putative origin of replication, *ori* (grey box). The extent of the sequences that are similar in strain A198 and strain C305 are indicated by dotted lines.

Table 7.1: Sequence analysis of the *intC* element[a] in *D. nodosus* strain C305

Gene	Co-ordinates 5'-3- (nt)	Size (aa)	% identity to nt or putative protein	Homologues/Description	Accession Number	P(*n*)
intC	1592-2759	389	54.2/392aa	*Dichelobacter nodosus* strain A198, IntA integrase [*tRNA-ser*$_{GCU}$]	L31763	5.5e-103
			40.4/391aa	*Escherichia coli* retronphage phiR73 integrase [*selCtRNA*]	M64113	5.4e-62
			38.3/394aa	*Mesorhizobium loti* symbiosis island integrase [*tRNAphe*]	AF049242	5.2e-56
			34.9/384aa	*Shigella flexneri* ϕSF6 integrase	X59553	1.2e-48
			38.8/394aa	*E. coli* 0157:H7 LEE pathogenicity island integrase [tRNA-sec gene]	AF071034	1.0e-64
			32.6/393aa	*Vibrio cholerae* 0395 integrase from CTXϕ PAI [*ssrA*]	U02372	2.3e-44
			33.0/394aa	*E. coli* K12 cryptic prophage P4-57 integrase [*ssrA*]	U11296	8.6e-43
			34.0/233aa	*E. coli* K12 integrase [*tRNAargW*]	AE00323	2.8e-42
			30.9/275aa	*Pseudomonas putida* integrase from the *clc* element [*tRNAgly*]	PPAJ4950	1.3e-26
			31.7/271aa	*D. nodosus* integrase, IntB [*tRNAser*$_{GCU}$ or *tRNAser*$_{GGA}$]	X98547	1.5e-21
			45.6/138aa	Bacteriophage P4 integrase	X05947	2.5e-19
			44.7/123aa	*Enterobacter aerogenes* integrase	AF039582	7.6e-19
			43.9/123aa	*E. coli* K12 integrase [*tRNAleuX*]	U14003	6.9e-17
			23.3/383aa	*Salmonella typhimurium* virulence-associated Gifsy-1 prophage integrase [lepA]	AF001386	2.7e-12
			22.1/383aa	Lambda Bacteriophage phi80 integrase	X04051	5.7e-12
			46.9/64aa	*Salmonella eneteridis* integrase from IS3-like element – only a partial sequence	SEJ0002209	2e-08
			28.4/370aa	*Rhodobacter capsulatus* putative prophage integrase	U57682	3.9e-07
			24.1/237aa	*Lactococcus lactis* integrase	L27649	1.6e-04
			37.1/70aa	*E. coli* 0111:H-ins	U06468	1.8e-04
			22.8/281aa	Bacteriophage 434	M60848	2.4e-03
			26.0/150aa	Coliphage HK022 integrase	X51962	3.1e-02
			19.7/289aa	*Haemophilus influenzae* integrase	U3281	5.2e-02
			18.9/296aa	Bacteriophage L2 from *Mycoplasma tuberculosis*	L13696	5.8e-02
orf242	3769-3043	242	28.5/84aa	*E. coli* strain ec45 RhsF accessory genetic element	AF044502	1.3e-02
orf171	4708-4195	171	-	NSH[b]	-	-
vapH"	5259-4944	105	84.6/66aa	VapH from *vap* region 1 of *D. nodosus* strain A198	L31763	-
vapG"	4915-4714	67	67.3/55aa	VapG from *vap* region 1 of *D. nodosus* strain A198	L31763	-
			85.5/62aa	VapG' from *vap* region 1 of *D. nodosus* strain A198	L31763	-

Table 7.1: (continued) Sequence analysis of the *intC* element[a] in *D. nodosus* strain C305

Gene	Co-ordinates 5'-3- (nt)	Size (aa)	% identity to nt or putative protein	Homologues/Description	Accession Number	P(*n*)
orf374	6877-5755	374	97.9/375aa	*D. nodosus* strain C305 IS*1253* *orf375* putative protein; gene adjacent to OMP locus	U02462	6.1e-172
			98.1/375aa	*D. nodosus* strain native plasmid pJIR896 IS*1253*, encoded by orf375	U34772	6.9e-171
			39.5/362aa	*Streptomyces pristinae* SnaA PIIA synthase	U21215	7.9e-53
			36.9/365aa	*Helicobacter pylori* plasmid pHPM186, IS*605*	AF077006	8.0e-52
			37.6/364aa	Thermophilic bacterium PS3IS transposase-like protein	D38778	2.9e-51
			37.0/365aa	*Helicobacter pylori* IS*605* TnpB transposase	U60177	2.1e-49
			37.5/365aa	*H. pylori* IS*605* TnpB transposase	AE000608	2.8e-49
			37.5/365aa	*H. pylori* IS*605* TnpB transposase, gene within cag pathogenicity island	AE000557	2.3e-48
			36.7/368aa	*Synechocycstis* sp. PCC6803 slr2062 transposase, gene adjacent to haemolysin gene	D90909	7.1e-41
			38.6/241aa	*Borrelia burgdorferi* transposase-like protein	U85588	3.6e-34
			39.0/241aa	*B. burgdorferi* plasmid lp28-2	AE000786	3e-33
			34.9/286aa	*Saccharopolyspora erythraea* IS*1136* transposase, IS element Erythromycin^R factor	L07626	1.3e-30
			34.8/282aa	*E. coli* K12 *orf402* encoded hypothetical protein	AE000240	1e-29
			30.4/313aa	*Synechocycstis* sp. phyocyanin CpcF protein	M93569	8.9e-26
			27.5/371aa	*Anabaena* sp. insertion sequence transposase, also on native plasmids	M24855	1.5e-23
			27.6/239aa	*Sulfolobus solfataricus* P2 ISC*1217* transposase, adjacent to glycogen biosynthesis genes	Y08256	9.4e-14
			31.6/228aa	Bacteriophage If1 (filamentous coliphage) protein encoded by *orf348*	U02303	1.4e-13
			26.5/324aa	*B. burgdorferi* plasmid lp38 BBJ48 coding region	AE000787	1e-12
			42.2/109aa	*Bacillus thuringenesis* hypothetical protein, gene is located adjacent to insecticidal genes	L48811	1.7e-11
			42.2/116aa	*S. typhimurium* virulence plasmid, gene is adjacent to *virB* and *spvD*	Z11561	7e-10
			36.0/139aa	*B. burgdorferi* plasmid lp28-3 BBH04 coding region	AE000784	7.8e-10
			36.1/119aa	*Clostridium perfringens* IS*1136* transposase, is adjacent to envasin encoding genes	X86498	8.4e-09
			39.4/104aa	*S. typhimurium* plasmid pSDL2, adjacent to *vsdF* virulence determinant	X56727	3.5e-07
			27.3/165aa	*M. tuberculosis* hypothetical protein	Z74024	2.1e-06
			28.8/153aa	*Methanococcus jannaschii* MJ0762 encoded protein	U67521	3.8e-04
orf108	6914-7238	108	98.1/103aa	*D. nodosus* plasmid pJIR896 IS*1253*, *orf117* encoded transposase	Z28337	2.4e-55
			98.1/103aa	*D. nodosus* strain C305 IS*1253*, adajcent to OMP locus, *orf117* encoded transposase	U93688	1.1e-54
			40.2/102aa	*H. pylori* TpnA from IS*606*	U95957	1.1e-19
			40.0/100aa	*E. coli* K12 Orf402 putative protein	AE000786	9e-19
			39.2/102aa	*H. pylori* HP1009 recombinase	AE000609	5.5e-18
			39.2/102aa	*H. pylori* IS605 TpnA transposase	AE000557	7.9e-18
			43.9/98aa	*B. burgdorferi* plasmid lp38 encoded transposase	AE000784	1.5e-17
			37.9/103aa	*S. typhimurium* virulence plasmid transposase	Z11561	6.55e-15
			37.9/103aa	*S. typhimurium* plasmid pSDL2 transposase	X56727	2.7e-14
			38.8/98aa	Phage phiH of *Halobacterium halobium* ISH1.8 *urf4* encoded transposase	X00805	2.4e-13

Table 7.1: (continued) Sequence analysis of the *intC* element[a] in *D. nodosus* strain C305

Gene	Co-ordinates 5'-3- (nt)	Size (aa)	% identity to nt or putative protein	Homologues/Description	Accession Number	P(*n*)
orf108			38.8/98aa	*Halobacterium salinarium* phage *dcm4* encoded transposase	X80164	9.5e-13
(cont.)			44.4/72aa	Bacteriophage If1 (filamentous coliphage) gene 1 protein	U02303	6.7e-12
			33.3/96aa	*S. typhimurium* IS*200* encoded tranposase	L25848	5.2e-08
			27.8/108aa	*V. cholerae* putative protein from *tnpA* transposase gene, adjacent to flagellin genes	Z67733	5.4e-08
			33.3/96aa	*S. typhimurium* IS*200* encoded transposase, adjacent to flagellin genes	U44749	6.1e-08
			33.3/96aa	*S. enteriditis* insertion element IS*1351* transposase	Z83734	2.2e-07
			35.4/96aa	*Salmonella abortusovis* IS*200* transposase	Y08755	5.9e-07
			24.4/107aa	*H. pylori* IS*605* TpnA transposase	U60177	7e-07
			32.3/96aa	*Yersinia pestis* putative protein encoded by IS*200*-like element, adjacent to invasin genes	U22457	9.2e-07
			24.3/107aa	*Helicobacter pylori* plasmid pHPM186, IS*605*	AF077006	1.1e-06
			32.3/96aa	*Methanosarcina mazeii* unidentified putative protein	U47274	2.3e-06
			24.2/107aa	*H pylori* TpnA, the TpnA encoding gene *tpnA* is located within the cag PAI	U60176	3e-06
			33.3/99aa	*Streptococcus zooepidemicus* unidentified orf that is located adjacent to a bacteriocin gene	U50357	8.9e-06
			29.3/99aa	*C. perfringens* putative transposase from IS*1469*, near *cpe* enterotoxin gene	X71844	1.2e-05
			33.8/80aa	*B. burgdoferi* plasmid lp25 BBE02 putative protein	AE000785	1.3e-05
			30.0/98aa	*Actinobacillus actinomycetemcomitans* IS200-like transposase	D64078	1.4e-05
			32.3/96aa	*Y. pestis* virulence phage-like plasmid pMT1 unidentified putative protein	AF053947	2.4e-05
			28.0/100aa	*Streptococcus pneumoniae* unidentified putative protein, adjacent to hyaluronidase gene	L20670	8.0e-03
***intB*$_M$**	7253-7368	38	97.2/107nt	107 bp of *intB* gene (from nt 402-509) from the *intB* element of *D. nodosus* strain A198	X98546	-
regA	7713-8035<	107	99.0/223 nt	*regA* gene (from nt 103-<) from the *intB* element of *D. nodosus* A198	X98546	-

a. NSH indicates that there is no significant homology to sequences in databases.
b. Where the site of integration is known for an integrase gene it is indicated in square brackets following the description of the Int homologue.

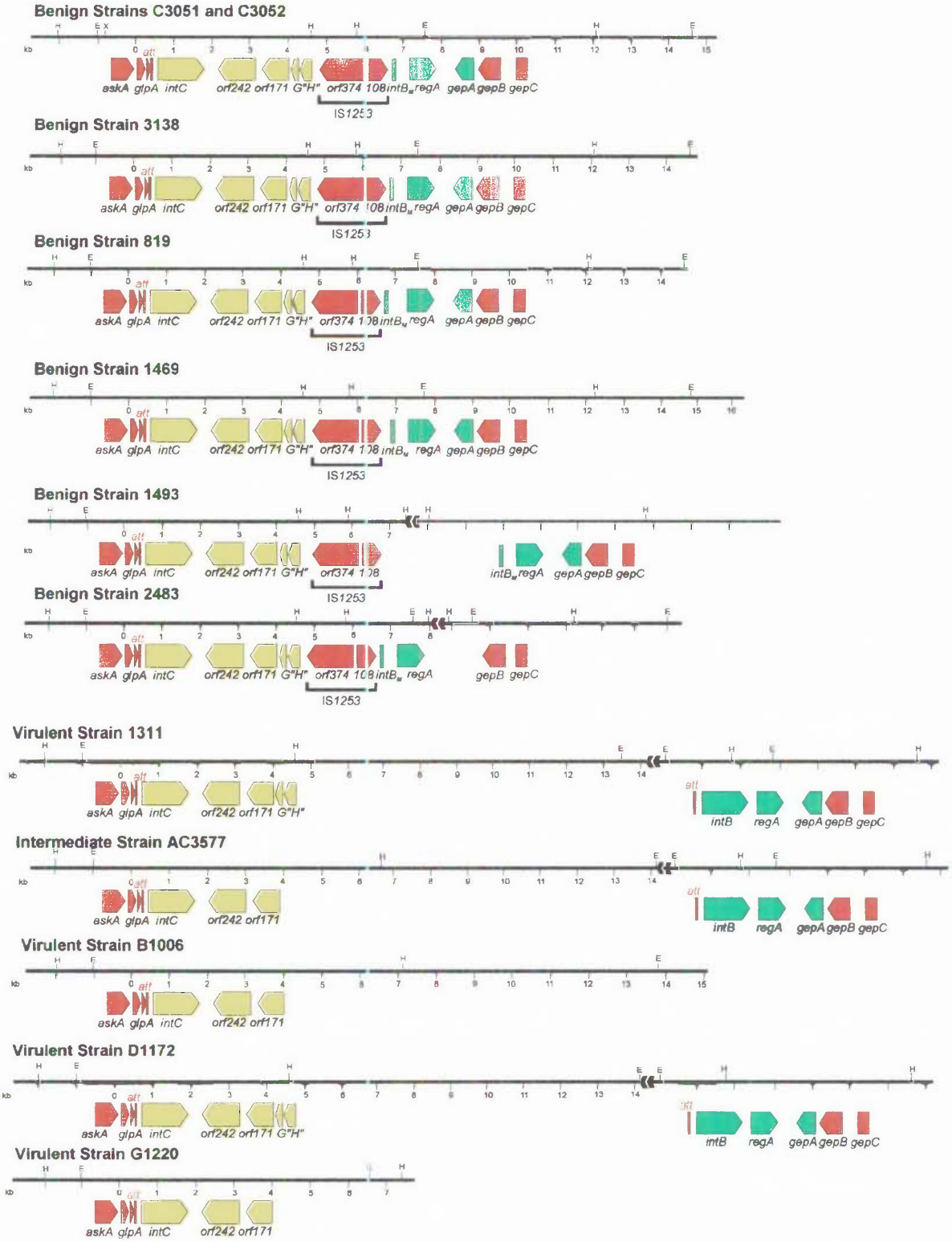
clustered together as they are in *D. nodosus* strain C305, and therefore constitute part of a single genetic element or otherwise. Furthermore, the spontaneous loss of the *intC* element and the *intD* element in one strain of *D. nodosus* was observed to result in the loss of protease thermostability, a virulence factor in *D. nodosus*. Consequently, the integration sites for the *vap*, *intB* and *intC* elements were investigated in seventeen strains of *D. nodosus*; these results suggest that in *D. nodosus*, virulence may be modulated by the site-specific integration of genetic elements.

7.2 Results

7.2.1 Characterisation of the *intC* element in several strains of *D. nodosus*

To further investigate the presence of *intC*, and to determine if genes *orf242* and *orf171* are always located adjacent to *intC* when present, seventeen strains of *D. nodosus* were analysed in Southern blot experiments using probes that would detect *askA*, *intC*, *orf242*, and *orf171* (Figure 7.1). The strains of *D. nodosus* analysed include virulent strains A198, 1311, B1006, G1220, H1215, D1172, intermediate strain AC3577, and benign strains C3052, 819, 1169, 2483, 1493, 3138, 1469, 1311A, H1204 and AC390. The results of these experiments have been summarised in Figure 7.2, however a complete record of restriction fragment sizes and maps for each strain analysed are provided in Appendices 7-8 and 11 respectively.

All strains studied contained at least one copy of *intC*, except for strain A198 and strain 1311A (Figure 7.2; Appendix 7) in which *intC* , *orf242*, and *orf171* are absent. In addition, *orf242* and *orf171* are found on the same *Eco*RI and *Hin*dIII restriction fragments as *intC*, in all strains except strains H1215, H1204, AC390 and 1169.



caption on following page

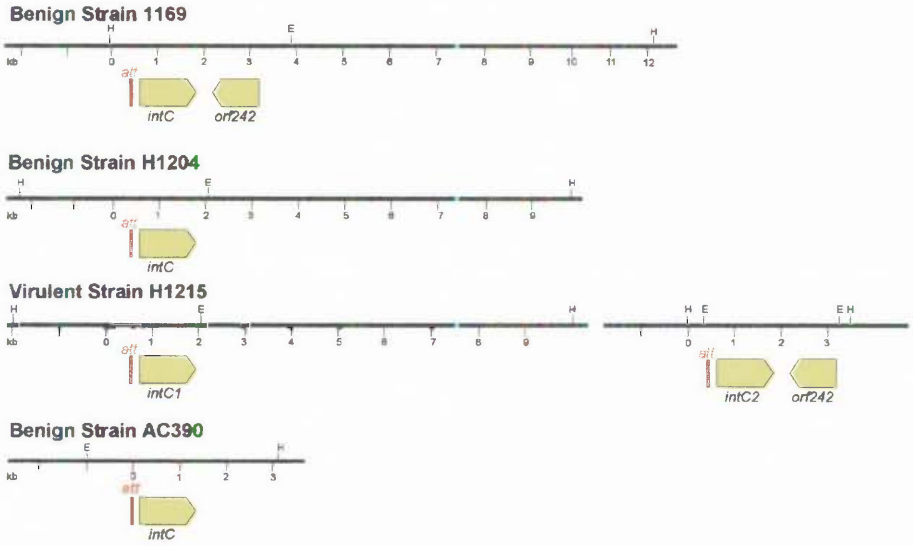


Figure 7.2: Maps showing the site of integration, the arrangement and integrity of the *intC* element in fifteen strains of *D. nodosus* as determined by Southern blot analyses. Restriction sites shown include *Eco*RI (E), and *Hin*dIII (H). Potential open reading frames are indicated by shaded arrows and are classified as follows: belonging to the *intB* element (green), belonging to the *intC* element (yellow) and genes that are not thought to be part of an integrated element (red), including *tRNA* genes (red triangles). Putative attachment sites, *att*, are indicated (narrow red boxes). Breaks in the lines indicate that the distance between the sequences on adjacent lines is unknown. For strains 1169, H1204, H1215 and AC390, the integration sites are currently unknown and so have not been shown on the diagram. No map for strains A198 and 1311A has been included, because these two strains do not contain any of the sequences that so far comprise the *intC* element.

In *D. nodosus* strain H1215 there are two copies of *intC* on 12.1 kb (*intC1*) and 2.9 kb (*intC2*) *Hin*dIII fragments (Figure 7.2; Appendix 7) respectively; no hybridisation to the *orf171* probe was observed, and the single copy of *orf242* present in strain H1215 was found to be adjacent to the copy of *intC2* (Figure 7.2). The genome of *D. nodosus* strains H1204 and AC390 each contain a single copy of the *intC* gene on a 12.1 kb *Hin*dIII fragment, but do not contain any copies of *orf242* or *orf171* (Figure 7.2). Whilst, *D. nodosus* strain 1169 contains a single copy of *intC* and *orf242* on a 12.1 kb *Hin*dIII fragment, and this strain does not contain a copy of *orf171* (Figure 7.2).

It is of interest that in strains H1215, H1204, AC390 and 1169, the copies of the *intC* gene that are not adjacent to *orf242* and *orf171* are also not adjacent to *askA* (Figure 7.2). In contrast, in all cases where *orf242* and *orf171* are contiguous with *intC*, the *intC* gene is adjacent to *askA*. Since genes *orf242* and *orf171* are always adjacent to the *intC* gene when present, and since when the *intC* gene is absent (strains A198 and 1311A) the genes *orf242* and *orf171* are also absent, these results suggest that genes *intC*, *orf242* and *orf171* are part of an '*intC* element'.

The predominant integration site for the *intC* element is in the 3'-end of the $tRNAser_{GCU}$ gene that is immediately downstream of the *askA* gene (Figure 7.2). In all eleven strains in which the *intC* gene is integrated adjacent to *askA* (Groups 2 & 3), *intC*-associated genes *orf242*, *orf171* are located immediately downstream of the *intC* gene (Figure 7.2). It is interesting that in all four cases where the *intC* gene is not adjacent to *askA* (Group 4), gene *orf171* was not present in the genome at all, and in two examples *orf242* is also absent. These results therefore suggest that alternative site or sites of integration may be inherently unstable or subject to high rates of genomic rearrangement.

The integration site adjacent to *askA* and in the 3'-end of the $tRNAser_{GCU}$ gene that is recognised by the *intC* element, is the same as that recognised by both the *vap* element and the *intB* element. This raises the possibility that the *intC* element may also integrate into the 3'-end of the $tRNAser_{GGA}$ gene located downstream from *pnpA* into which the *vap* element and the *intB* element have both been found to integrate.

Southern blot analyses (Figure 7.2, Appendices 7 and 8) have shown that *pnpA* never hybridises to the same *Hin*dIII fragment as *intC*, even when this position is not occupied by the *intA* or *intB* elements as in strain 3138. Perhaps this suggests that sequence similarity between the *att* sites is not sufficient for integration of even highly related integrases like *intA*, *intB* and *intC*, and consequently, other factors such as DNA topology or

accessory protein binding sites may be an important influence on the specificity of *intC* versus *intA* and *intB* for their integration sites, as noted for site-specific recombinases in general (Stark & Boocock, 1995).

In order to compare the integrity and the arrangement of the *intC* element in different strains of *D. nodosus*, probes containing fragments that hybridise to *vapG"H"* and the IS*1253* element were also used as probes. The interpretation of Southern blot data generated using these probes is more difficult since: *(i)* the IS*1253* element is found in multiple copies in the genome of different strains of *D. nodosus*. The *vapG"H"* sequence is also found in multiple copies and cross hybridises with *vapG*, *vapH* and *vapG'* where the *vap* regions are present; *(ii)* many of these multiple copies are in uncharacterised positions of the genome, and consequently many more Southern blot experiments would be required to place these genes with certainty.

In prior Southern blot analyses, genomic DNA had been digested with *Eco*RI, *Hin*dIII, and *Hin*dIII/*Eco*RI restriction enzymes, and thus the same digests of genomic DNA were probed with IS*1253* and *vapG"H"* so they could be positioned relative to the many genes that had already been mapped. Thus the IS*1253* and copies of *vapG"H"* have been placed in their putative positions, based on the best possible interpretation of Southern blot data and what is already known about the *intC* element arrangement in C305. Where possible, the putative positions of the IS*1253* and *vapG"H"* should be confirmed using a PCR strategy.

Of the eleven strains of *D. nodosus* that contain *intC*, *orf242* and *orf171*, there are three strains (3138, 819, 1469) in which the arrangement of *intC* element and downstream elements are identical to the arrangement found in benign strain C305 (both C3051 and C3052), (Figure 7.2). The arrangement of the *intC* element sequences in strains 1493 and 2483 are also identical to strain C305 up to the end of the IS*1253* element and up to the end

of *regA* respectively. In strain 1493 the *intB* element sequences do not hybridise to the same *Eco*RI or *Hin*dIII fragments as the IS*1253* probe, and so are not adjacent. Similarly, in strain 2483 there is no copy of *gepA*, and *gepB* and *gepC* do not hybridise to the same *Hin*dIII or *Eco*RI fragments as the IS*1253* probe, and consequently the *intB* element sequences are separated in this strain (Figure 7.2). All six of the aforementioned strains do not contain a complete copy of the *intB* gene, and instead contain $intB_M$ (Chapter 5).

There are five strains, including strain 1311, AC3577, B1006, D1172 and G1220 that have the same arrangement as strain C305 until the start of *orf171*, after which they are distinguished from the other strains by the following: *(i)* strains AC3577, B1006 and G1220 do not have a copy of *vapG"H"* adjacent to *orf171*; *(ii)* the IS*1253* element is not located adjacent to *orf171* or *vapG"H"* in any of these five strains. *(iii)* strains 1311, AC3577 and D1172 contain intact copies of the *intB* gene whilst strains B1006 and G1220 do not contain a copy of the *intB* gene or part thereof downstream from the *askA* integration site (Chapter 5). None of these five strains contain $intB_M$; *(iv)* strains 1311, AC3577, B1006 and D1172 hybridise to large *Eco*RI fragments. In strains AC3577 and B1006 the *Eco*RI fragments that hybridise to the *intC* probe are of identical size (15-16 kb), and slightly larger than the *Eco*RI fragment that hybridises to *intC* in strain 1311 (13-14 kb). It is unclear whether strain D1172 hybridises to an *Eco*RI fragment of the same size as 1311 or AC3577 and B1006 since this strain was not analysed in the same Southern blot experiment (that is, on the same membrane).

There are two possibilities with regard to the size of the native *intC* element: *(i)* the *intC* element is a short integrative element of approximately 4.5 kb, and therefore intact in strain C305, and C305-like strains. In this case, in strains in which the *intC* element sequences hybridise to a larger *Eco*RI fragment (1311, AC3577, B1006, D1172), some other element or sequence is present which separates the 4.5 kb *intC* element sequences from the *intB* element sequences; or *(ii)* the *intC* element in strain C305 and C305-like

strains, has been truncated by some genomic rearrangement, such as simple recombination between two insertion sequences resulting in deletion of the intervening regions, and thus the right-hand end of the element is missing in these strains. In contrast, the strains in which the *intC* element sequences hybridise to a larger *Eco*RI fragment of similar size (1311, AC3577, B1006, D1172) may contain an intact copy of the *intC* element, which includes the right-hand end of the element.

The later hypothesis *(ii)* seems most likely, since in strain C305 it is apparent that genomic rearrangements have resulted in the deletion of part of the *intB* element (Figure 7.2), and the right-hand end of the *intC* element may have been deleted at the same time. The end product of these rearrangements appears to be a copy of $intB_M$ that is integrated close to *intC* element sequences. In contrast, in strains 1311, AC3577, D1172 the *intB* element and *gepB-F*sequences downstream of *askA* appear to be contiguous, and are altogether absent from this position in strains B1006 and G1220.

The size of the intact *intC* element is currently unknown. It seems that the *intC* element sequences are likely to extend downstream of the right-most *Eco*RI site, since this fragment is of a similar size in strains 1311, AC3577, B1006, and D1172. Whether the sequences downstream of *orf171* comprise part of the *intC* element in strain G1220 is unclear, since the restriction polymorphisms present in this strain differ from all other strains characterised in this study.

Since it appears that the *intB* element is adjacent to the *attR* of the *intC* element, at least in some strains of *D. nodosus*, it may be possible to use LR-PCR or an I-PCR strategy to isolate the right-hand end of the *intC* element for further analysis.

7.2.2 Loss of the *intC* element from *D. nodosus* virulent strain 1311

During routine laboratory growth, the native plasmid pDN1 was lost spontaneously from *D. nodosus* virulent strain 1311, generating a daughter strain designated as 1311A (Section 3.2.6). In later Southern blot analyses, genomic DNA from *D. nodosus* strain 1311 and daughter strain 1311A was probed with DNA fragments from *intB*, *intC*, and *askA* respectively (Figure 7.3).

In both strains 1311 and 1311A, there are 2 copies of *intB*. One copy of *intB* is truncated and found on a 2.2 kb *Eco*RI fragment and corresponds to the copy that is adjacent to the *attR* of the *vap* region that is integrated downstream of *pnpA* in both strains (Figures 7.3 and 7.4). The second complete copy of *intB* is intact in both strains, however, this copy of *intB* hybridises to a 2.5 kb *Eco*RI fragment in strain 1311, but to a 3.3 kb *Eco*RI fragment in strain 1311A. Thus, these results suggested that the *intB* gene region had undergone rearrangement in *D. nodosus* strain 1311 to generate strain 1311A.

In *D. nodosus* strain 1311, *intC* and *askA* probes both hybridise to a 13.0 kb *Eco*RI (Figure 7.3), a 6.9 kb *Hin*dIII and to a 5.7 kb *Hin*dIII/*Eco*RI fragment, indicating that in strain 1311 the *intC* element is integrated adjacent to *askA*. In contrast, the genomic DNA from daughter strain 1311A did not hybridise to the *intC* probe (Figure 7.3), and thus the *intC* gene is not present in the 1311A genome. In daughter strain 1311A, the *askA* probe hybridises to 3.3 kb *Eco*RI (Figure 7.3), 3.5 kb *Hin*dIII, and 2.6 kb *Hin*dIII/*Eco*RI fragments respectively; the same restriction fragments that hybridise to the *intB* probe in this strain 1311A. These results therefore indicate that in strain 1311A the *intB* gene and *askA* gene are adjacent. Thus, the loss of the *intC* element from strain 1311 resulted in the relocation of the *intB* element such that *askA* and *intB* are juxtaposed in daughter strain 1311A (Figure 7.4).

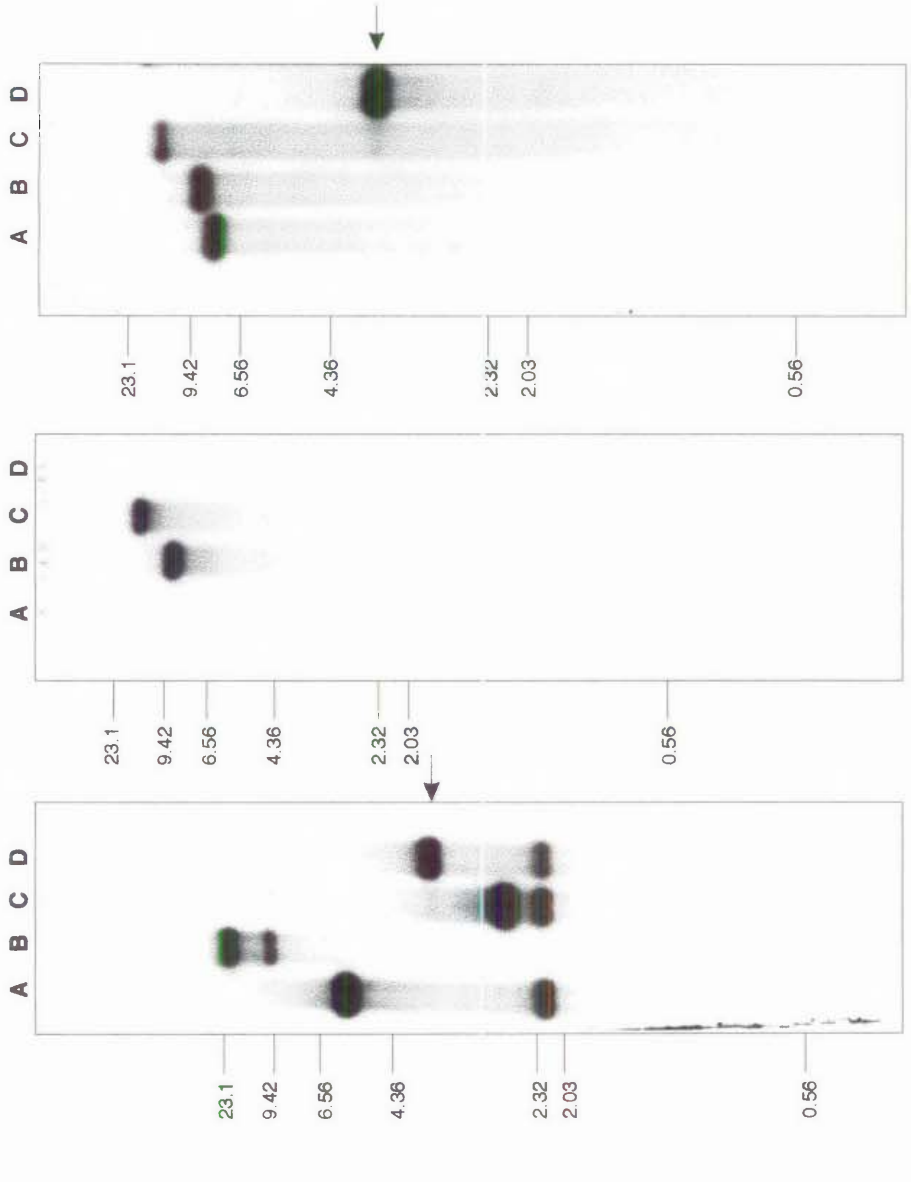


Figure 7.3: Southern blot analysis of *intB* (panel I), *intC* (panel II) and *askA* (panel III) genes in *D. nodosus* strains A198 (lanes A), C3052 (lanes B), 1311 (lanes C) and 1311A (lanes D). Genomic DNA was digested with *Eco*RI in all three experiments shown above. Lambda *Hind*III standard sizes are indicated on the left of each panel in kb. The *intB*, *intC* and *askA* probes used correspond to probes 3, 4 and 1 as shown in Figure 5.1. The arrows indicate faint hybridisation to 3.2 kb fragments that are present in lanes C (panel I and III) and correspond to a small population of 1311 cells, losing the *intC* element and so generating 1311A-like cells at a frequency that is high enough to be detected by Southern blot analyses.

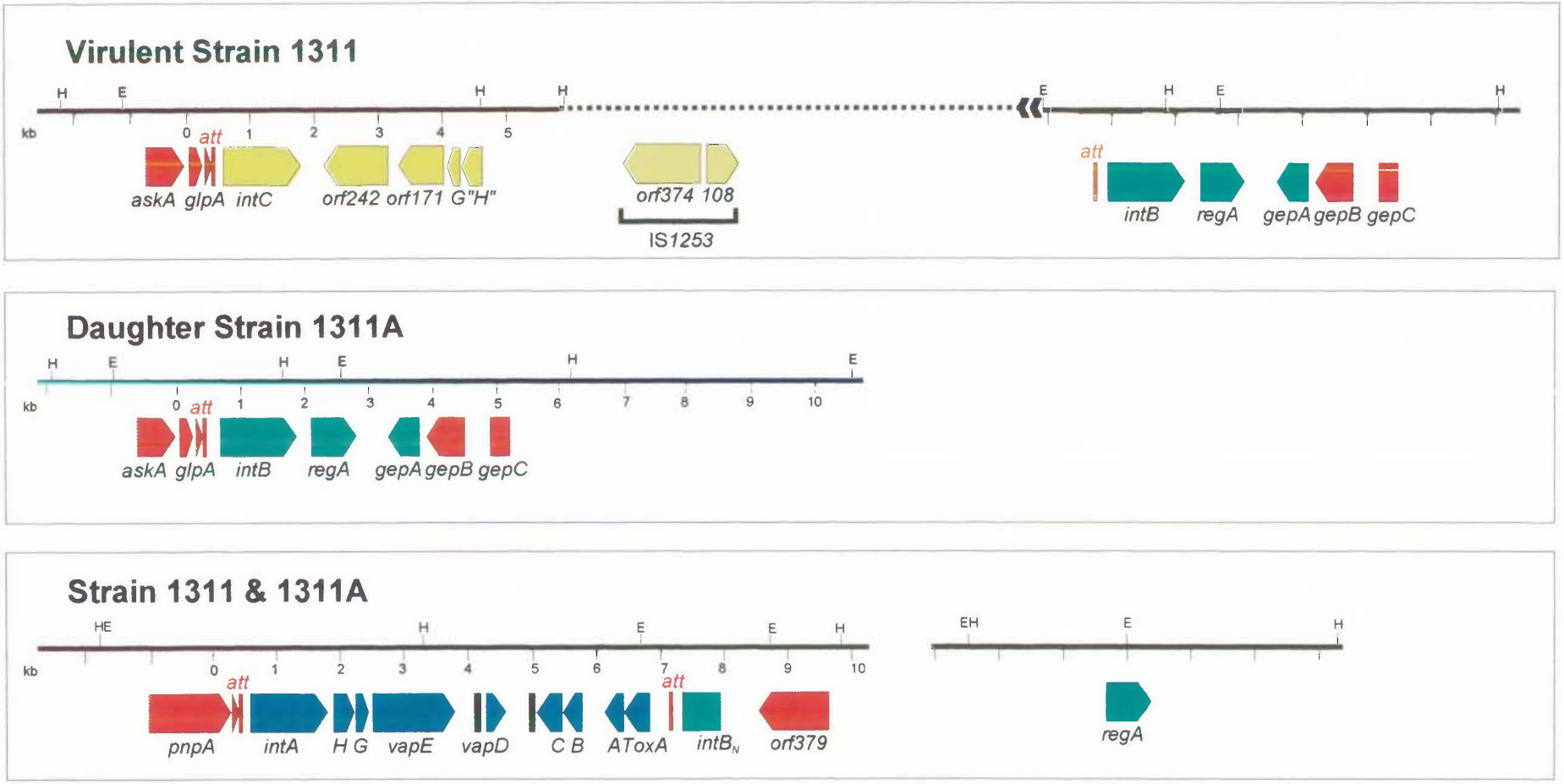


Figure 7.4: The arrangement of *vap*, *intB* and *intC* elements in *D. nodosus* strains 1311 and 1311A. Restriction sites shown include *Eco*RI (E), *Hin*dIII (H) and *Xho*I (X). Potential open reading frames are indicated by shaded arrows and are classified as follows: belonging to the *intB* element (green), belonging to the *vap* region (blue), belonging to the *intC* element (yellow), genes that are not part of an integrated genetic element (red), including *tRNA* genes (red triangles). Repeated sequences are as follows: attachment sites, *att* (narrow red boxes), 102 bp repeat or partial copies thereof from the *vap* region (filled boxes). The DNA sequence of the region shown by a dotted line has not yet been determined. The *vap* region and duplicate copy of *regA* are unchanged in strains 1311 and 1311A (bottom panel).

Further Southern blot analyses (Appendices 4, 7 & 8) indicated that *intC*-associated genes, *orf242*, *orf171*, and *vapG"H"* were lost together with the *intC* gene from strain 1311, which is consistent with the hypothesis that these genes constitute part of the *intC* element (Section 7.2.1). The hypothesis that the four strains (1311, AC3577, B1006, D1172) that contain the large *Eco*RI fragments might contain more *intC* element sequences is also supported by evidence that shows that the *intC* element sequences excise from the *askA* integration site in strains 1311. The excision of the *intC* element results in the relocation of the *intB* element such that it is next to *askA*, and leaves the integration site intact. It seems most likely that all of the sequences between the *attR* or the *intC* element and the *attL* of the *intB* element are part of the *intC* element, since *intB* and *intC* integrate into the same *att* sequences.

To confirm that *intC* is adjacent to *askA* in *D. nodosus* strain 1311, and that *intB* is adjacent to *askA* in 1311A, PCR products were amplified using primer sets specific for *askA-intC*, and *askA-intB*. As expected, a 2.6 kb *askA-intC* product was amplified from 1311 genomic DNA, whilst no *askA-intC* product was observed for strain 1311A (Figure 7.5, panel I). Using *askA-intB* primers a 1.1 kb *askA-intB* product was detected in 1311A as predicted, however unexpectedly, a 1.1 kb *askA-intB* product was also observed for parent strain 1311 in which *intC* is integrated adjacent to *askA* (Figure 7.5, panel II).

Consequently, it was hypothesised that if the excision of *intC* element sequences was occurring frequently in a small population of 1311 cells, both *askA-intC* and *askA-intB* PCR products might be observed. Serial dilutions of an equivalent amount of 1311 template and 1311A template show that the concentration of the *askA-intB* product is much lower in 1311 than in 1311A (Figure 7.5, panel III). Hence, experimental evidence from PCR results suggests that a sub-population of *D. nodosus* strain 1311 are undergoing loss of the *intC* element frequently enough to be detected. This observation is supported by Southern blot analysis in which a faint 3.3 kb *Eco*RI fragment is visible in addition to the 2.2 kb and

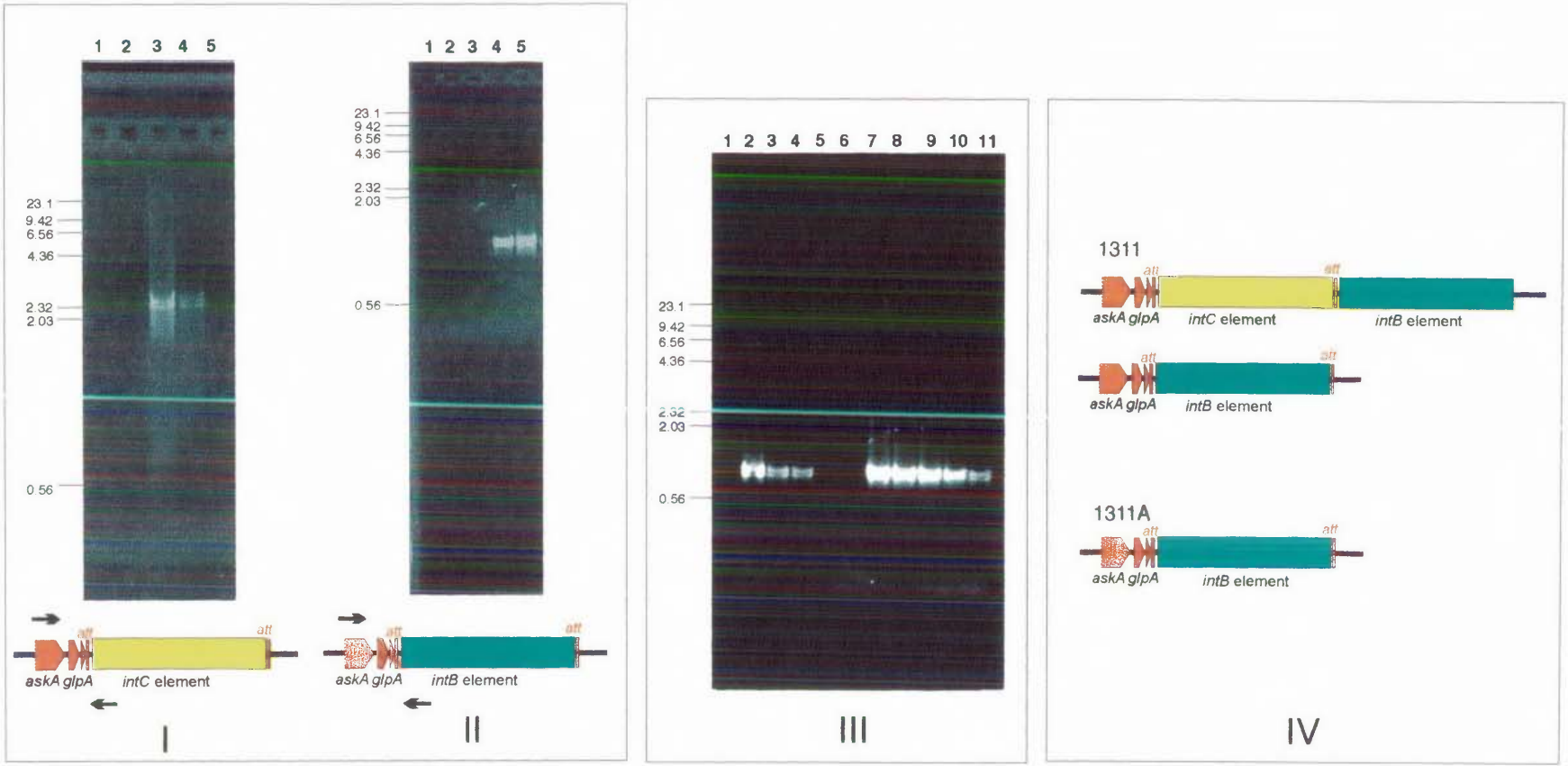


Figure 7.5: PCR products amplified using primers complementary to *askA* and *intC* element sequences (panel I), and *askA* and *intB* sequences (panel II). Reactions shown are as follows: No DNA (lanes 1), and genomic DNA template from *D. nodosus* strains A198 (lane 2), C305 (lane 3), 1311 (lane 4) and 1311A (lane 5). The approximate positions of oligonucleotide primers are indicated by mapped arrows below the appropriate panel. PCR analysis of *askA* and *intB* in *D. nodosus* strains 1311 and 1311A is shown(panel III). The photograph shows the PCR products obtained in reactions using primers specific for *askA* and *intB* with the following and amount of template DNA from strain 1311 or 1311A as indicated: no DNA (lane 1), 100 ng 1311 (lane 2), 10 ng 1311 (lane 3), 1 ng 1311 (lane 4), 0.1 ng 1311 (lane 5) and 0.01 ng 1311 (lane 6), 100 ng 1311A (lane 7), 10 ng 1311A (lane 8), 1 ng 1311A (lane 9), 0.1 ng 1311A (lane 10), 0.01 ng 1311A (lane 11). Lamdba *Hin*dIII standard marker sizes in kb are indicated on the left of the appropriate panel. These results show (panel IV) that in strain 1311 the *intC* element is located adjacent to the *askA* gene (panel I), and in a small number of the population the *intC* element is lost and consequently an *askA-intB* PCR product is also detectable (panel II); whilst in 1311A only an *intB* product is detectable since the *intC* element has been lost (panel I and II), and as expected this product is detectable at much lower dilutions in 1311 compared to 1311A (panel III).

2.5 kb *Eco*RI fragments that hybridise strongly to the *intB* probe for *D. nodosus* strain 1311 (Figure 7.3, panel I). Similarly, in strain 1311 a faint 3.3 kb *Eco*RI fragment is also detectable in addition to the strong band of 13.0 kb that also hybridises to *askA* (Figure 7.3, panel III).

To ascertain whether the loss of *intC* was due to excision *via* a site-specific mechanism and whether the excision or loss of the *intC* element disrupted the *tRNA* gene, the *askA-intB* PCR product from *D. nodosus* strain 1311A was subcloned, sequenced and aligned to the left end and right end of *vap* region 1/3 in *D. nodosus* reference virulent strain A198. Analysis of this junction region shows that the *att* site is identical to *attL*, rather than *attR*, and thus that the excision of the *intC* element did not disrupt the sequence of the mature *tRNA* gene. This is consistent with the site-specific excision of the *intC* element. It seems most likely that the excision of the *intC* element was mediated by an Int/Xis-catalysed excision event involving *intC* element encoded functions, especially given the frequency at which the *intC* element is lost (Figures 7.3 and 7.5) rather than *via* a *recA*-catalysed homologous recombination event.

7.2.3 Excision of the *intC* element from *D. nodosus* intermediate strain AC3577

The observed loss of the *intC* element in strain 1311, led to the identification of a second strain, intermediate strain AC3577, in which the excision of the *intC* element from the *askA* position of the genome is also detectable. In strain AC3577, the *askA* probe (Figure 7.6) hybridises to two *Eco*RI fragments of 15.5 kb and 3.3 kb respectively. It is notable that intensity of the 3.3 kb band is significantly less than that of the 15.5 kb band (Figure 7.6), suggesting that a second partial or divergent copy of *askA* exists in this strain, or that the loss of sequences adjacent to *askA* is occurring in a small population of AC3577

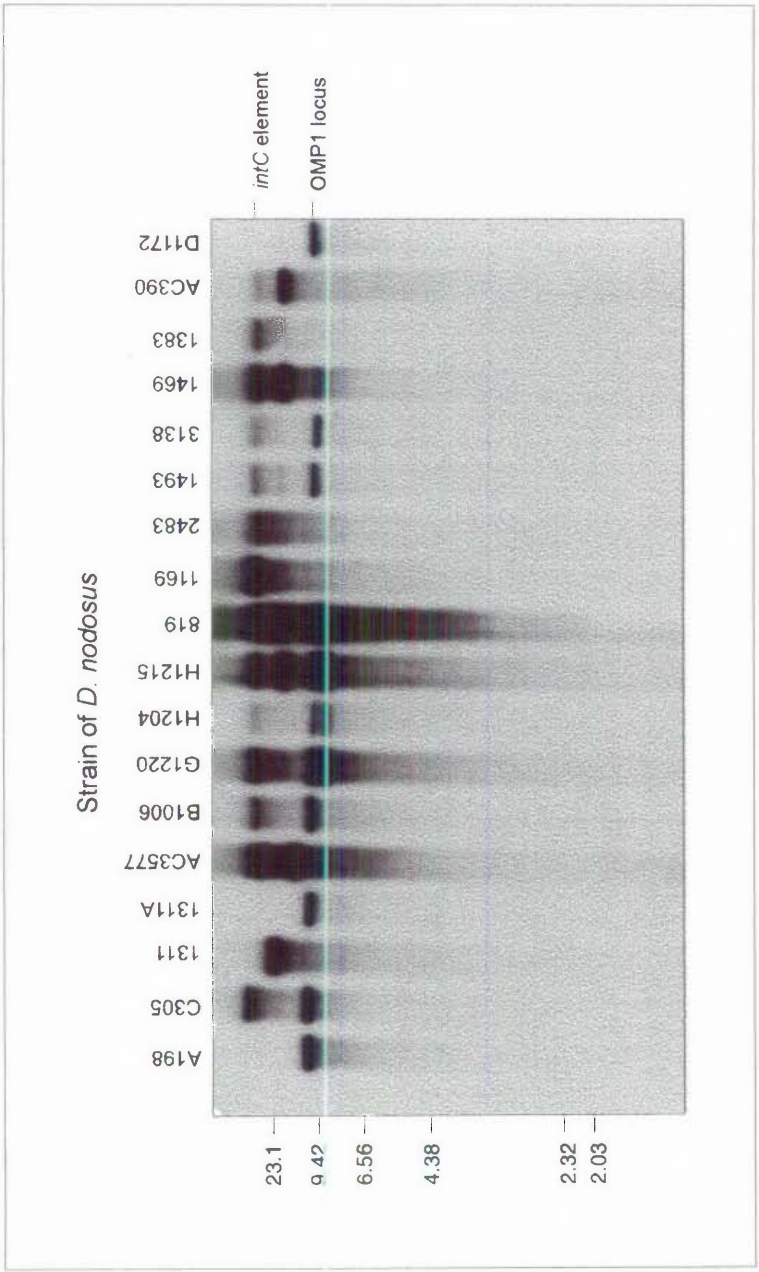


Figure 7.7: Southern blot analysis of the IS*1253* element (probe 7 in Figure 7.1) in a range of *D. nodosus* strains. Genomic DNA was digested with *Xho*I. Lambda *Hin*dIII standard sizes are indicated on the left of each panel in kb. The *Xho*I bands that correspond to the copy of IS*1253* known to be adjacent to the OMP1 locus in strain A198 and the *intC* element in strain C305 are indicated (right).

cells. The later hypothesis is supported by faint hybridisation of the *intB* probe (Figure 7.6) to a 3.3 kb *Eco*RI fragment, which is identical to the fragment that hybridises to the *intB* probe after the loss of the *intC* element from the *askA* position in strain 1311 (Figure 7.6). An *Eco*RI fragment of 2.5 kb that hybridises strongly to the *intB* probe is also detectable, and is identical in size, to the fragment which *intB* hybridises to in strain 1311 when the *intC* element is integrated in the *askA* position (Figure 7.6). In addition it seems unlikely that a housekeeping gene like *askA* would be duplicated in a single strain of *D. nodosus*.

The hypothesis that *intC* is lost from the *askA* position in strain AC3577 is complicated by the observation that *intC*, *orf242*, *orf171* all hybridise to a single *Eco*RI fragment of approximately 14.5 kb but to two restriction fragments for both *Hin*dIII or *Hin*dIII/*Eco*RI digests (Appendix 7). In the digests where hybridisation to two fragments is observed, one restriction fragment is always significantly less intense (Figure 7.6). The most plausible explanation for this is that when the *intC* element excises in strain AC3577 it is not lost, but is instead maintained either as a circular intermediate at low copy number, or is integrated at a second position in the AC3577 genome on an *Eco*RI fragment of approximately 14.5 kb.

To test this hypothesis, two experiments were performed. In the first experiment, serial dilutions of an equivalent amount of AC3577 template and 1311A template were amplified using PCR, and showed that the concentration of the *askA-intB* product is much lower in AC3577 than in 1311A, as was observed in an identical experiment in which 1311 and 1311A genomic DNA were used as templates (Figure 7.5, panel II). These results are consistent with the idea that the *intC* element is excising from the *askA* position in a small population of strain AC3577 cells.

To determine whether sequences from the *intC* element were maintained as a circular molecule, a second experiment was performed. Uncut genomic DNA from strain AC3577

was fractionated on a 0.6% agarose gel, transferred to a nitrocellulose membrane for Southern blot analysis and probed with a fragment from *intC* (Figure 7.1, probe 4). However, a distinct plasmid band was not visible. The putative plasmid band may have been too large to separate from genomic DNA. Consequently, results from this experiment are inconclusive, and in future experiments pulse field gel electrophoresis and subsequent Southern blot analysis could be utilised to test the hypothesis with greater sensitivity. Alternatively, a PCR strategy could be utilised to amplify sequences that are potentially adjacent in the circular molecule, but known not to be adjacent when the *intC* element is integrated into the AC3577 chromosome. The design of such an experiment would require the isolation of the right-hand end of the *intC* element. A means by which the right-end of the *intC* element could be isolated was discussed previously in Section 7.2.1.

From Southern blot experiments, there is no evidence to suggest that the *intC* element sequences are being lost from at least seven (C3052, D1172, G1220, B1006, 2483, 1493, and 1469) of the strains that carry all of the genes that so far define the *intC* element (*intC*, *orf242*, and *orf171*). However, in benign strains 819 and 3138, faint bands are visible in addition to the strong *Hin*dIII and *Hin*dIII/*Eco*RI bands that hybridise to *askA*, *intC*, *orf242* and *orf171* (Appendices 7 & 8), suggesting that the rearrangement of *intC* element sequences are also occurring in these strains. However, it is difficult to show this definitively since for these two strains it is unclear what the product of rearrangement might be, given that these two strains do not appear to contain an intact copy the *intB* element and its attachment site (*attL*). Consequently, strain 819 and 3138 do not contain the sequences that are complementary to the primer (GB1) that was used to confirm that the *intB* element sequences are adjacent to *askA* after the loss of the *intC* element sequences in strain 1311 and AC3577.

It is interesting that the *intC* element appears to be lost from strain 1311 (Section 7.2.2) and yet appears to be maintained in *D. nodosus* strain AC3577. The fact that the *intC* element is lost in strain 1311 suggests that the *intC* element may not contain a

plasmid maintenance system analogous to the *vapA/toxA* system (Section 1.6.4.1) present in the *vap* regions and on the *vap* plasmid that is found in strain AC3577. Why the *intC* element should be maintained in one strain AC3577 but lost altogether from strain 1311 is unknown, but may reflect the relative integrity of the *intC* element in these two strains. Alternatively, *(i)* the excision of the *intC* element in strain 1311 in the presence of native plasmid pDN1 may have led to the destabilisation of an *intC* plasmid or pDN1, or both due to plasmid incompatibility; or *(ii)* the *intC* element may contain no vegetative origin of replication, and consequently could be lost if not re-integrated following excision; or *(iii)* since both the *intC* element and *vap* element are able to recognise the same *att* sequence, the excised *intC* element may have integrated into the *vap* plasmid which is carried only by strain AC3577 and consequently be stably maintained as a circular cointegrate extrachromosomal molecule.

7.2.4 Loss of the *intC* element from *D. nodosus* strain 1311 is associated with the loss of a copy of IS*1253* and *intD* element sequences

In *D. nodosus* strain C3052 the *intC* element genes *intC*, *orf242*, *orf171* and *vapG"H"* are followed by a copy of an insertion sequence with approximately 98% amino acid identity (Bloomfield, 1997) to the IS*1253* element previously identified on the *vap* plasmid (Billington *et al.*, 1996) of *D. nodosus* strain AC3577, and located adjacent to the OMP1 locus in *D. nodosus* strain A198 (Moses *et al.*, 1995). Immediately following one of the two copies of the IS*1253* element present in strain C305, there is a 100 bp with 98% identity to the part of the coding region from the *intB* gene (position 410-500) from the right end of *vap* region 3 in strain A198 (Table 7.1). 344 bp downstream from this partial copy of *intB*, similarity to the *intB* element from *D. nodosus* strain A198 resumes from the beginning of the *regA* gene (Bloomfield, 1997) (Figure 7.1). Thus, in strain C305 the *intC* element appears to be interrupted by the IS*1253* element and truncated, since part of the *intB*

element is located immediately downstream of the insertion sequence.

In order to determine whether the IS*1253* element is always found within the *intC* element in different strains of *D. nodosus*, Southern blot analyses were employed. Genomic DNA from each strain was digested with *Xho*I (Figure 7.7), *Eco*RI, *Hin*dIII and *Hin*dIII/*Eco*RI (Appendix 7) and hybridised with probes specific for the IS*1253* element and *orf171* from *D. nodosus* strain C305.

Of the strains analysed most had one (seven strains) or two (seven strains) copies of the IS*1253* element, whilst a few have three (three strains) or four (one strain) copies of the insertion sequence (Figure 7.7, Table 7.2). Furthermore, the insertion sequence appears to be found in only five different positions in the *D. nodosus* genome (Figure 7.7). Three of these five positions are known; the IS*1253* probe hybridises to a 10.5 kb *Xho*I fragment, which corresponds to the copy of the IS*1253* element previously reported to be located adjacent to the OMP1 locus in strain A198 (Moses *et al.*, 1995); a *Xho*I fragment of >23.1 kb (Figure 7.7) fragment that corresponds to a copy within the *intC* element of strain C305 (Bloomfield, 1997); and a 13.7 kb *Xho*I fragment that corresponds to the copy of the IS*1253* located on the *vap* plasmid in *D. nodosus* strain AC3577 (Billington *et al.*, 1996).

*Xho*I fragments of a novel size are observed for *D. nodosus* strain 1311 (a doublet of ~19.6 kb), and for strains H1215, 1469 and AC390, that have a *Xho*I fragment of about 15.5 kb in common. The putative positions of the IS*1253* elements in the 18 strains analysed have been summarised in Table 7.2.

It is likely that the *intC* element is greater than 19 kb in size since we know that for strain 1311 *orf171* and IS*1253* hybridise to a *Xho*I fragment predicted to be about 19.6 kb in size, and there are not *Xho*I sites within the total 8.6 kb sequenced from the *att* site immediately upstream of the *intB* gene to *gcpG* in strain A198.

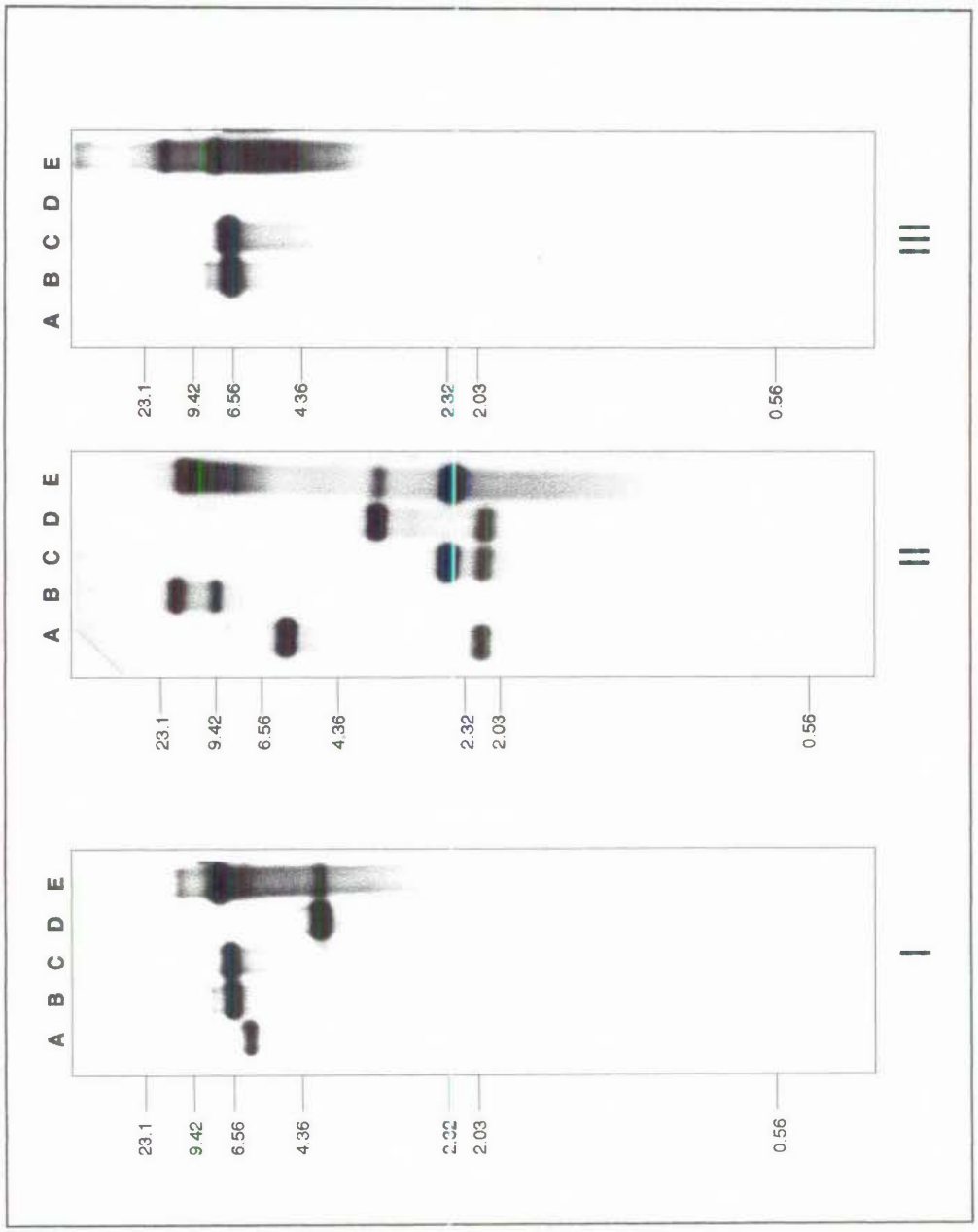


Figure 7.6: Southern blot analysis of *askA* (panel I), *intB* (panel II) and *intC* (panel III) genes in *D. nodosus* strains A198 (lanes A), C305 (lanes B), 1311 (lanes C), 1311A (lanes D) and AC3577 (lanes E). Genomic DNA was digested with *Hin*dIII (panel I) or *Eco*RI (panels II, III). Lambda *Hin*dIII standard sizes are indicated on the left of each panel in kb. The *intB*, *intC* and *askA* probes used correspond to probes 3, 4 and 1 as shown in Figure 5.1.

Table 7.2: Copy number and location of the IS*1253* element in *D. nodosus*

D. nodosus Strain	Copies of IS*1253*	Location in genome
A198	1	OMP1 locus
C3051	2	*intC* element, unknown
C3052	2	*intC* element, OMP1 locus
1311	2	*intC* element, unknown
1311A	1	OMP1 locus
AC3577	2	*vap* plasmid, unknown
B1006	2	*intC* element, OMP1 locus
G1220	3	*intC* element, OMP1 locus, unknown
H1204	3	*intC* element, OMP1 locus, unknown
H1215	4	*intC* element, OMP1 locus, unknown, unknown
819	2	*intC* element, OMP1 locus
1169	1	*intC* element
2483	1	*intC* element
1493	2	*intC* element, OMP1 locus
3138	2	*intC* element, OMP1 locus
1469	3	*intC* element, OMP1 locus, unknown
AC390	1	unknown
D1172	1	OMP1 locus

It is interesting that in strain 1311, there are two copies of the IS*1253* sequence on *Xho*I fragments of approximately 19.6 kb. Southern blots in which genomic DNA from strain 1311 was digested with *Hin*dIII and *Eco*RI confirmed that there are two copies of the insertion sequence (Appendix 7). In contrast, daughter strain 1311A, contains only a single copy of the insertion sequence. The concomitant loss of a copy of the insertion sequence and the *intC* element suggests that a copy of the IS*1253* is located within the *intC* element. This is consistent with what might be expected for strain 1311 (Figure 7.2), since Southern blot analyses indicate that *orf171* and IS*1253* hybridise to the same *Xho*I and *Eco*RI fragments (Appendix 7).

Perhaps more significantly, the loss of the *intC* element is also associated with a

rearrangement involving the second copy of the IS*1253* sequence that is still present in strain 1311A. The IS*1253* probe hybridises to a 10.5 kb *Xho*I fragment corresponding to the single copy of the insertion sequence that is present in strain 1311A. Since the IS*1253* element from strain A198 also hybridises to a *Xho*I fragment of 10.5 kb (Figure 7.7), and is known to be located immediately adjacent to the OMP1 locus, these results suggest that the copy of IS*1253* in strain 1311A is also located adjacent to the OMP1 locus (Figure 7.7).

A copy of the *intC* element and the IS*1253* element were not the only sequences lost from the genome of *D. nodosus* strain 1311A. All of the sequences that comprise the right-hand end of the *intD* element (Section 6.2.1) were also lost from the 1311A genome (Section 6.2.2 and Figure 7.9).

A mechanism by which the *intC* element, the *intD* element and a copy of the IS*1253* element could be lost, whilst the remaining copy of the IS*1253* element was repositioned in the 1311 genome is suggested by the order of genes shown on the physical and genetic map of *D. nodosus* strain A198 (La Fontaine & Rood, 1997), and is illustrated in Figure 7.8.

The physical and genetic map of *D. nodosus* strain A198 indicates that the IS*1253* sequence is located adjacent to *vap* region 1/3 of *D. nodosus* strain A198 (La Fontaine & Rood, 1997), and the IS*1253* element is located on the same *Xho*I fragment as the OMP1 locus (Billington *et al.*, 1996), and it is also known that the *intB* element sequences are present adjacent to the *attR* site of *vap* regions 1/3 (Bloomfield *et al.*, 1997); hence the arrangement of the elements shown (Figure 7.8A). In strain 1311 (Figure 7.8B), there are two copies of the IS*1253* element, one of these punctuates the *intC* element and as expected is lost with the *intC* element. The other copy of the IS*1253* element is found on a *Xho*I fragment of the same size as the aforementioned copy (indicated by a line between the *Xho*I sites). With loss of the *intC* element and the *intD* element from strain 1311, two *Xho*I sites are lost resulting in the juxtaposition of the remaining IS*1253* element and the OMP1 locus in

strain 1311A (Figure 7.8C). Since the *intD* element is not present in most strains of *D. nodosus*, the OMP1 locus is next to IS*1253* in most strains. The *intD* element must have been lost between *intB* and the OMP1 locus, for there to be a change in *Xho*I sites between the IS*1253* element and the OMP1 locus. Consequently, the *intD* element cannot be part of the *intC* element.

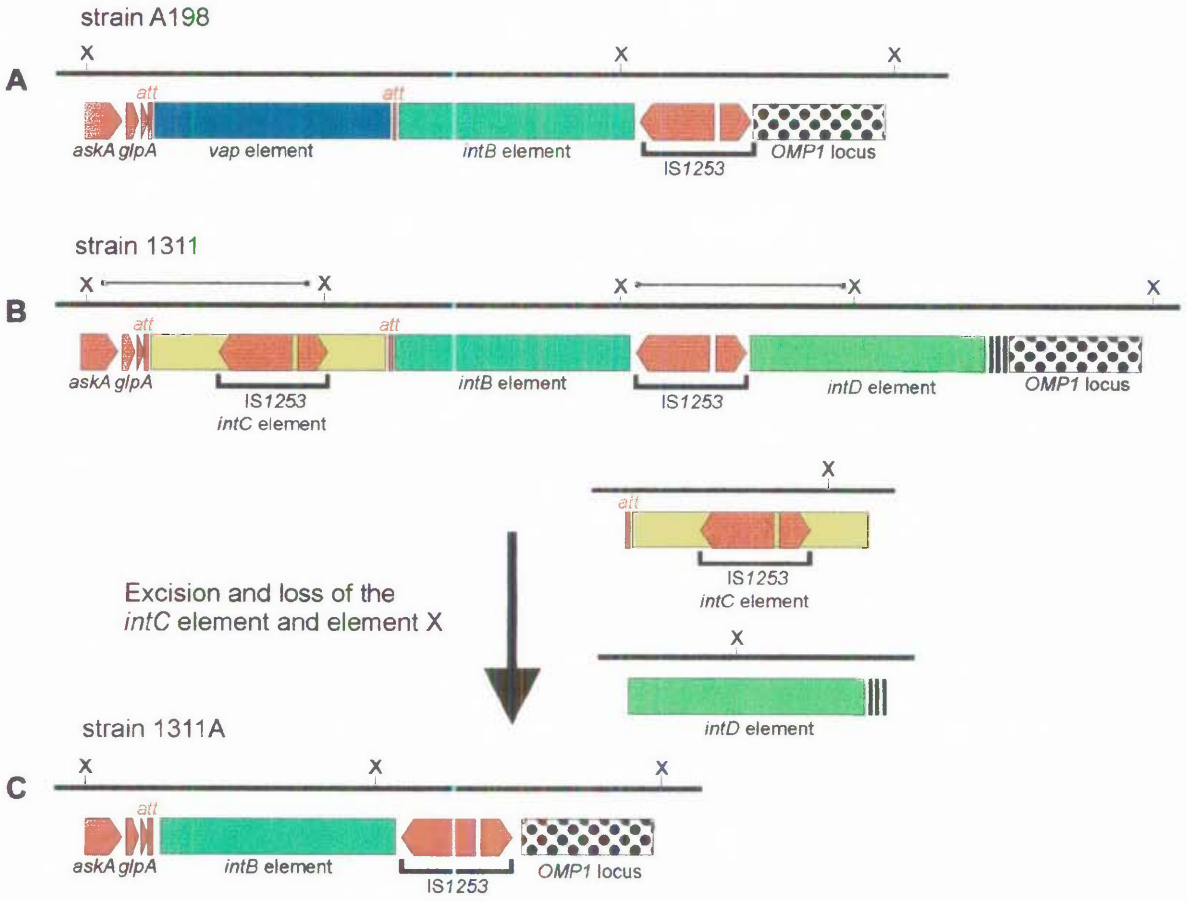


Figure 7.8: Model for the rearrangement of the IS*1253* element in strain 1311 to generate strain 1311A. The arrangement of *vap* region1/3, the *intB* element, IS*1253* and the *OMP1* locus from physical and genetic map of *D. nodosus* strain A198 (**A**). The putative arrangement of the *intC* element, *intB* element, IS*1253* and *OMP1* locus in *D. nodosus* strain 1311 (**B**). The arrangement of the *intB* element and IS*1253* in strain 1311A after loss of the *intC* element and the *intD* element (**C**). *Xho*I (X) sites are shown and 102 bp repeats are indicated (narrow black boxes). The drawing is not to scale.

It is also possible that the loss of the *intC* element with a copy of the IS*1253* element was followed by the independent transposition of remaining the IS*1253* element to the

OMP1 position. Such a transposition event seems less likely since collectively this would have required three independent rearrangements rather than two.

Like strain 1311, strain AC3577 contains a copy of the sequences from the right-end of the *intD* element (Section 6.2.2). In AC3577, a faint *Hin*dIII band of 14.5 kb is detectable in addition to a strong *Hin*dIII band of 9.3 kb that corresponds to the chromosomal copy of these *intD* element sequences (Figure 7.9). No such faint band corresponding to the *intD* element is detectable in strain 1311. The *intC* probe also hybridises to a faint *Hin*dIII band of approximately 14.5 kb in strain AC3577. Thus, these results suggest that in strain AC3577 the sequences from the right-hand end of the *intD* element and from the left-hand end of the *intC* element may hybridise to the same 14.5 kb *Hin*dIII fragment. Consequently, it is possible that the *intD* element and *intC* element sequences may excise to form a co-integrate circular molecule. However, the *intC* probe and fragments from the right-hand end of the *intD* element may also be hybridising to two separate 14.5 kb bands in *D. nodosus* strain AC3577.

Since the *intD* and *intC* elements recognise the same integration sites one might expect that they are at least able to form such a cointegrate molecule. If these elements do form a cointegrate molecule it is likely that the right end of the *intD* element and the left end of the *intC* element are adjacent since they apparently hybridise to the same *Hin*dIII fragment. However, one would also expect that the aforementioned ends are adjacent if the *intC* element and *intD* element sequences were part of the same genetic element. To determine if the *intD* and *intC* elements do excise to form a cointegrate, PCR primers complementary to appropriate sequences from the *intC* and *intD* element sequences could be used to amplify putative intervening sequences in a cointegrate molecule.

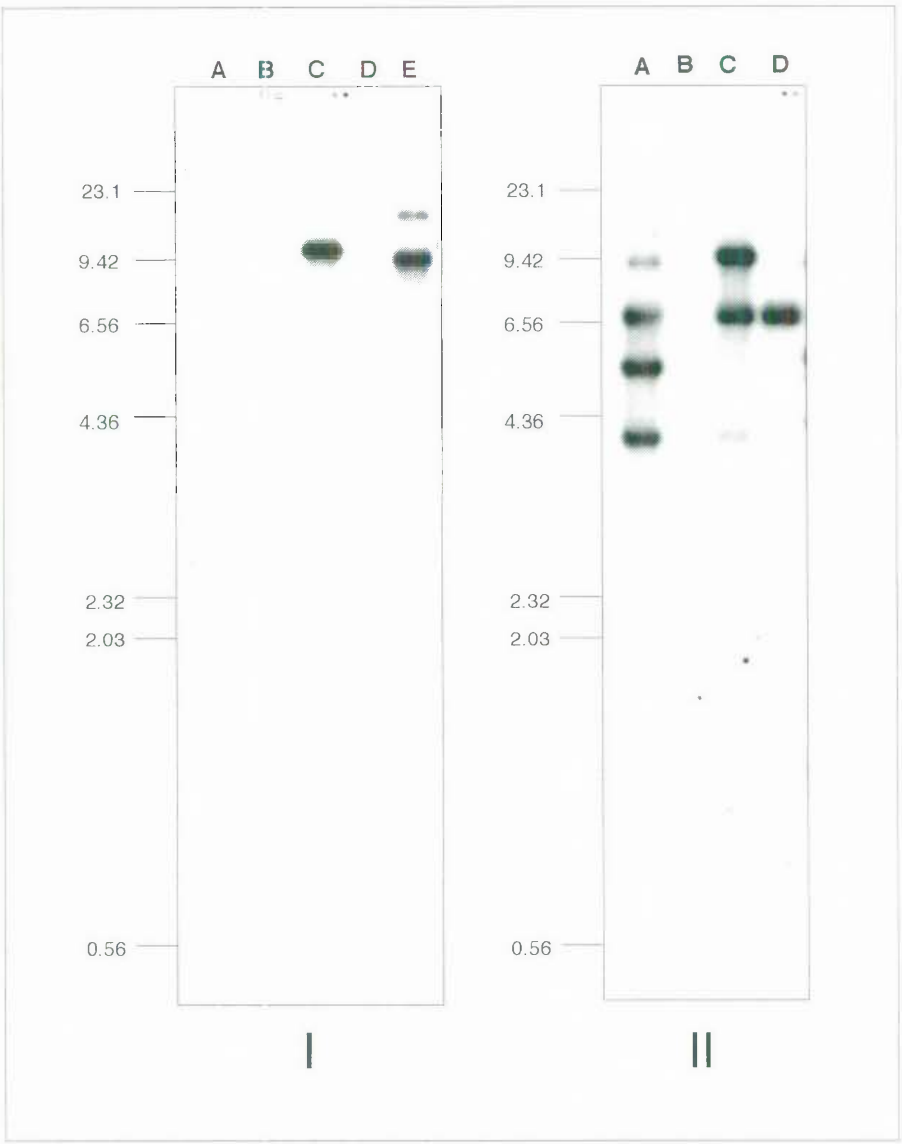


Figure 7.9: Southern blot analysis of *intD* element sequences, *dapA-C* (panel I) and 102 bp repeats (panel II) in *D. nodosus* strains A198 (lanes A), C3052 (lanes B), 1311 (lanes C), 1311A (lanes D) and AC3577 (lane E, panel I only). Genomic DNA was digested with *Hin*dIII in both of the experiments shown above. Lambda *Hin*dIII standard sizes are indicated at the left of each panel in kb. The *intD* element probes correspond to probe 1 (panel I) and probe 3 (panel II) as shown in Figure 6.12.

7.2.5 Further evidence that the *intC* and *intD* elements are separate integrated elements

The excision of the *intC* and *intD* elements appears to occur at the same time in *D. nodosus* strains 1311 and AC3577. This could be explained if the *intD* element sequences are part of the *intC* element, or if they are two separate elements which excise in response to the same stimuli and thus would have the potential to interact with each other. Such interactions have been reported previously for *Bacteroides* conjugative transposons which are known to stimulate the transposition of other conjugative transposons in *trans* due to the *trans*-action of regulatory proteins (Salyers *et al.*, 1995b), rather than the *trans* action of proteins for excision. If the *intC* and *intD* elements comprised a single genetic element, *intC* and *intD* element sequences should be adjacent in the *D. nodosus* genome. The evidence on which the model shown in Figure 7.8 was based, supports the idea that the *intC* element and the *intD* element are different genetic elements.

In addition, we have observed previously the loss of the *intD* element independently of the loss of the *intC* element. The original laboratory strain of *D. nodosus* C305 (C3051) lost the *intD* element from between *pnpA* and $intB_N$ to generate the current laboratory strain, C3052 (Section 6.2.2). Despite the loss of C3051 from laboratory stocks, a very small sample of C3051 genomic DNA, had been analysed by Southern blotting previously (Shaw, 1997), and thus the nylon membrane was available for further analysis. The membrane was probed with the *intC* gene and confirmed that the *intC* element was integrated adjacent to *askA* in the C3051 as it is in C3052 (Appendices 7 and 8). Thus, the original laboratory strain of C305, now called C3051 (Section 6.2.2.1) contained a copy of the *intC* element next to *askA* and a copy of the *intD* element next to *pnpA*. Since the two elements were integrated in different positions in the C305 genome and the *intD* element was lost independently of the *intC* element, they must be different genetic elements.

There are also many other strains of *D. nodosus* that do not contain *intD* element sequences but do contain *intC* element sequences (including strains C3052, G1220, H1204, H1215, 1169, 2483, 1493, 1469, AC390), also supporting the conclusion that the *intC* and *intD* elements are not components of the same element. It is interesting that in all of the aforementioned strains, the *intC* element does not hybridise to the *Eco*RI fragments of 13-15 kb like strains 1311, AC3577, B1006, D1172, and thus may not be intact (Section 7.2.1). Although it might be hypothesised that the interaction between the *intC* and *intD* element requires these sequences to be intact, there appears to be some rearrangement of the *intC* element in strains 819 and 3138, and the arrangement of *intC* element and downstream sequences are identical to those from strain C305, and consequently are not intact.

Southern blot analyses were used to investigate the possibility that the *intD* element may be integrated between the right-hand end of the *intC* element and the left end of the *intB* element in *D. nodosus* strains 1311, AC3577, D1172 and B1006. These analyses showed that the left-hand end of the *intB* element is not next to the right-hand end of the *intD* element (Figure 6.15), and is located elsewhere in the genomes in the aforementioned strains of *D. nodosus*. Results presented in this section suggest that the *intD* element is instead integrated between the *intB* element and OMP1 locus in *D. nodosus* strain 1311.

7.2.6 Loss of the *intC* and *intD* elements from *D. nodosus* strain 1311 results in loss of thermostable protease activity in daughter strain 1311A

Although the precise role of *D. nodosus* proteases in pathogenesis has not yet been determined, the protease exoenzymes secreted by virulent strains of *D. nodosus* readily degrade elastin (Kortt *et al.*, 1982; Stewart, 1979; Thomas, 1964), keratin, fibrinogen, hemoglobin and collagen (Green, 1985), and are thought to aid the bacterium in obtaining nutrients from the hooves and surrounding tissues (Stewart, 1989). This hypothesis is

supported by physicochemical differences that exist between proteases secreted by benign and virulent strains of *D. nodosus* (Emery, 1988; Lilley, Stewart & Kortt, 1992; Moses & Yong, 1989).

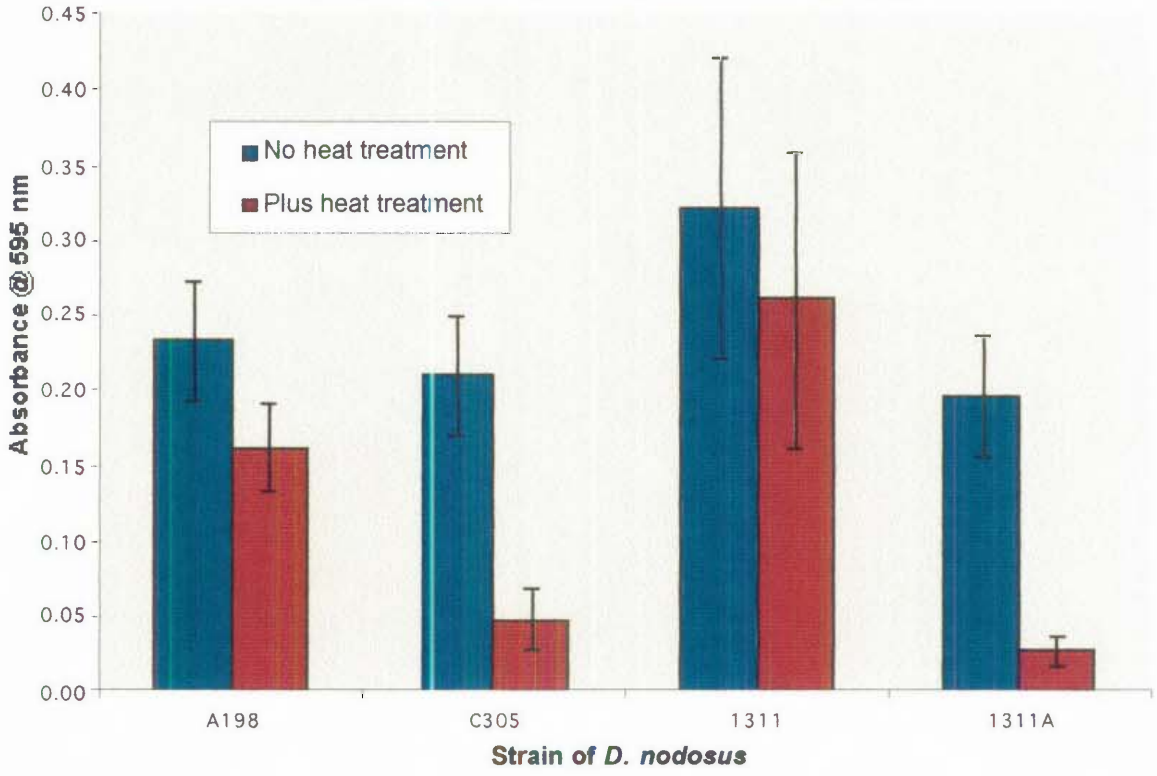
It has been established that the proteases of virulent strains of *D. nodosus* are more thermostable than proteases of benign strains of *D. nodosus* (Depiazzi, Henderson & Penhale, 1990; Depiazzi & Rood, 1984). Furthermore, it is this particular factor that is the basis of current laboratory tests for the differential diagnosis of *D. nodosus* infections (Section 1.3) (Plant & Walker, 1994; Rood *et al.*, 1996). Hence, to determine whether the loss of the *intC* element from virulent *D. nodosus* strain 1311 had an effect on the virulence classification of the daughter strain 1311A, a protease thermostability assay was performed (Section 2.13.1) on *D. nodosus* strains 1311 and 1311A. Reference virulent strain A198 was included as a positive control for thermostable protease and benign strain C3052 as a negative control for thermostable protease. The results are summarised in Table 7.3 and Graph 7.1.

Virulent strains A198 and 1311 have a decrease in protease activity of less than 35%, whilst for C3052 and 1311A protease activity decreased by more than 64% after heat treatment, indicating that the proteases produced by C3052 and 1311A are much less thermostable than the proteases secreted by A198 and 1311. These results suggest that the virulent strain 1311 was converted to a putative benign strain, 1311A. Strain 1311A is referred to as a putative benign strain, because loss of thermostable protease activity does not provide conclusive evidence for loss of virulence since protease thermostability is only a trait associated with virulence. To confirm that strain 1311A produces clinically-benign disease, sheep trials would need to be performed.

Table 7.3: Protease thermostability in four strains of *D. nodosus*

Strain of *D. nodosus*	Mean A_{595} ± SD no heat treatment	Mean A_{595} ± SD heat treatment (10 min @ 60°C)	% Loss of Activity after heating
A198	0.233 ± 0.04	0.162 ± 0.03	29.5 ± 5.6
C3052	0.210 ± 0.04	0.049 ± 0.02	74.1 ± 11.9
1311	0.323 ± 0.10	0.261 ± 0.10	20.1 ± 11.4
1311A	0.198 ± 0.04	0.028 ± 0.01	84.9 ± 9.0

Graph 7.1: Thermostability of extracellular proteases in four strains of *D. nodosus*



Strain 1311 lost the plasmid pDN1, together with the *intC* and *intD* elements to produce strain 1311A. It is highly unlikely that the loss of pDN1 resulted in the loss of virulence since pDN1 has been found in only one *D. nodosus* strain, and this is supported by results presented in Section 7.2.7. Thus, it is probable that loss of the *intC* or *intD* elements resulted in the loss of thermostable protease.

These results are interesting in that they suggest that both virulent and benign strains of *D. nodosus* must contain at least the genetic potential to express thermostable proteases, and that the expression or secretion of these thermostable proteases is switched on in virulent strains but off in benign strains. The loss of the *intC* and *intD* elements from this strain of *D. nodosus* appears to trigger this 'switch'. No information regarding the potential production of thermostable protease by benign strains has been reported in the literature.

7.2.7 *D. nodosus* 1311 and 1311A are morphologically distinguishable

During the passaging of *D. nodosus* strains 1311 and 1311A it was noted that strain 1311A colonies were consistently smaller in size and produced less exoenzymes than the colonies of strain 1311 (Figure 7.10). Interestingly, smaller colony size and reduced exoenzyme activity have been noted as being characteristic of benign strains in general (Depiazzi *et al*, 1985; Stewart *et al*, 1989), though it has also been noted that morphological differences alone are not reliable enough for the accurate differentiation of benign and virulent strains of *D. nodosus* (Depiazzi *et al.*, 1990; Stewart, 1989).

The size difference between strain 1311 and 1311A colonies was utilised to select colonies of 1311 that had potentially lost pDN1 and/or the *intC* element. The aim of this experiment was to determine whether the loss of *intC* occurs concomitantly with the loss of pDN1, since in *D. nodosus* strain 1311A both the *intC* element and pDN1 have been lost.

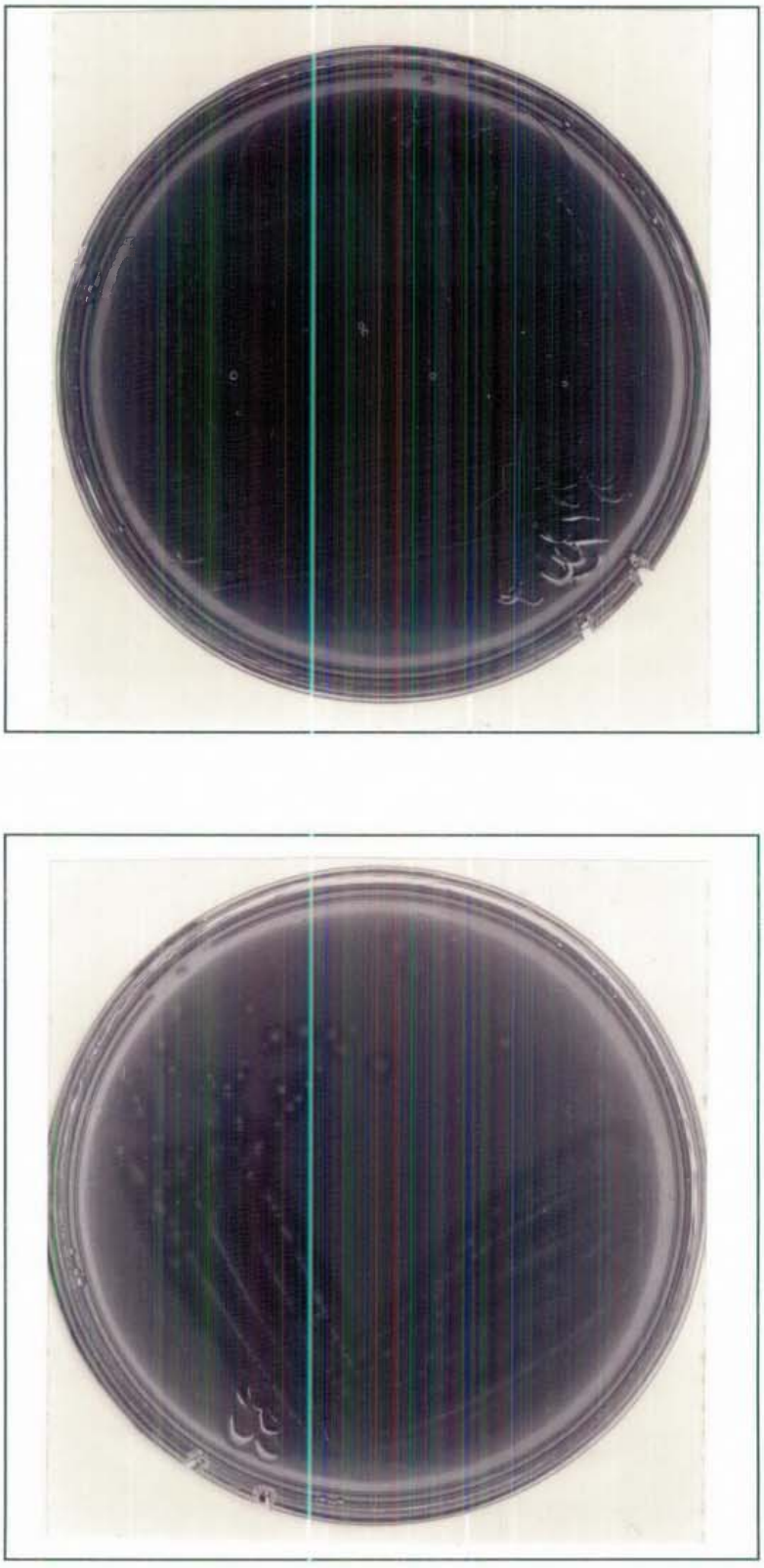


Figure 7.10: Streak plates of *D. nodosus* strain 1311 (left panel) and 1311A (right panel). Strain 1311 is distinguishable phenotypically from daughter strain 1311A by having comparitively larger colony size and exhibiting greater exoenzyme activity. The latter is demonstrated by clearing around the colonies on blood eugonagar as shown in the photographsabove.

The possibility that pDN1 loss occurs with loss of the *intD* element sequences has not yet been investigated, because initially it was hypothesised that the *intD* element sequences may have been part of the *intC* element (Chapter 6), rather than a separate element.

A total of 55 small colonies were selected, and screened by Southern blot analyses using sequences corresponding to pDN1 and the *intB* gene as probes (Figure 7.1). Of those screened, five colonies that were selected had lost pDN1 (Figure 7.11), but had retained the *intC* element since the *intB* probe hybridised to the same size restriction fragment as parent strain 1311 (Figure 7.11). Three of these pDN1$^-$ strains (Table 7.4, strain 1311 E, F and H)

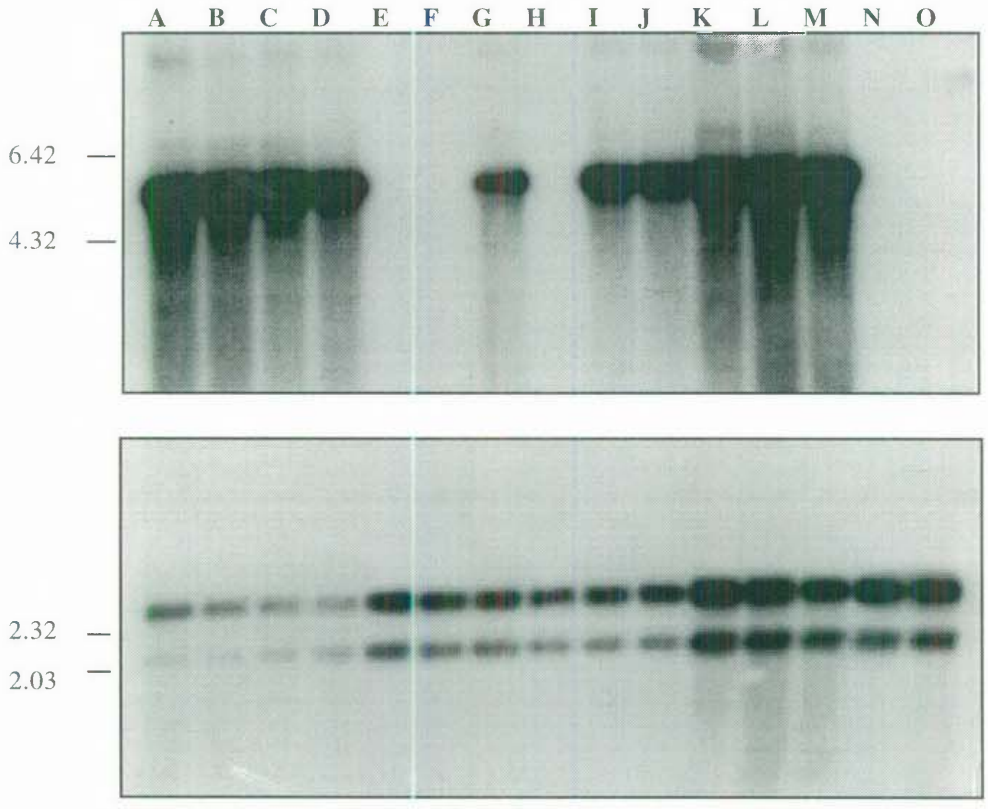


Figure 7.11: Southern Blot analysis of selected small colonies of strain 1311. Genomic DNA was prepared from selected small colonies of *D. nodosus* strain 1311 A to O, digested with *Eco*RI and probed with DNA fragments from *D. nodosus* native plasmid pDN1 (top panel) and with an *intB* probe (top panel). In Selected 1311 colonies A-D, G, I-M the pDN1 probe hybridises to a 5.1 kb *Eco*RI fragment, whilst selected 1311 colonies E, F, H, N and O do not hybridise to the pDN1 probe at all. All selected 1311 colonies hybridise to the *intB* probe; *Eco*RI fragments of 2.2 kb and 2.5 kb are visible, and correspond to a truncated copy of *intB* (*intB*$_N$) and a complete copy of *intB* respectively. Thus the two copies of *intB* are in the same position as in parent strain 1311 (Section 7.2.2).

were subject to protease thermostability assays. All of the colonies that had lost pDN1 had retained protease thermostability, confirming as expected, that the absence of protease thermostability in strain 1311A was not due to the loss of the native plasmid pDN1 (Section 3.2.6). In addition, in the five pDN1$^-$ strains, the loss of pDN1 did not occur at the same time as the loss of the *intC* element (Figure 7.11), since like parent strain 1311, the *intB* probe still hybridised to *Eco*RI fragments of 2.2 kb and 2.5 kb respectively. Whether the *intD* element had been lost from any of these strains still needs to be investigated.

Also of interest, and not anticipated was that one of the selected small colonies assayed, strain 1311#12, which had not lost pDN1, had lost protease thermostability (Table 7.4). Southern blot analysis indicated that this strain has not lost the *intC* element. Although the loss of protease thermostability is not due to loss of the *intC* element in this variant, further investigations are required to determine whether this strain has lost the copy of the *intD* element and/or has acquired a mutation that affects thermostable protease production or secretion.

Table 7.4: Protease thermostability in seven strains of *D. nodosus*

Strain of *D. nodosus*	Mean A_{595} no heat treatment	Mean A_{595} heat treatment (10 min @ 60°C)	% Loss of Activity after heating
A198	0.924	0.710	13.1
C3052	0.6[illegible]8	0.091	85.3
1311	0.8[illegible]8	0.738	15.9
1311 E	0.932	0.700	24.9
1311 F	0.941	0.739	21.5
1311 H	0.900	0.672	25.3
1311 #12	0.634	0.042	93.4

Standard deviations are not shown since results are based on a single experiment. *D. nodosus* strains 1311 E, F and H have lost pDN1, but not the *intC* element, whilst strain 1311 #12 has lost protease activity but not pDN1 or the *intC* element.

Results from these selection studies suggest, as expected, that a change in colony size may be attributable to numerous genotypic and environmental changes. In these experiments 9.0% of selected colonies had lost pDN1, and 1.8% had lost protease thermostability, suggesting that a small colony size is associated with both phenotypes that are apparent in strain 1311A. These results also indicate that the loss of native plasmid pDN1 can occur without the concomitant loss of the *intC* element, however it is as yet unknown whether the *intD* element is present in these pDN1$^-$ strains. These experiments also suggest that morphological differences such a colony size coupled with proteases thermostability assays may be useful for the identification of novel strains of *D. nodosus*, that in the absence of a transformation system, may yield some interesting insights into the regulation of virulence determinants, such as thermostable proteases, in *D. nodosus*.

7.2.8 Integration sites for *vap*, *intB* and *intC* elements in virulent and benign isolates of *D. nodosus*

The results presented in Section 7.2.6 suggest that the loss of the *intC* element resulted in the conversion of a virulent strain (1311) to a putative benign strain (1311A). Initially, the fact that a second genetic element, the *intD* element, was also lost was unknown, and arose out of later studies in which sequences to the left of a copy of $intB_\lambda$ from a C305 genomic DNA library were analysed (Section 6.2.2).

It seems unlikely that the *intC* or *intD* elements encode factors that are responsible for the expression of thermostable protease, since reference virulent strain A198 does not contain any of the sequences that so far define the *intC* or *intD* elements (Sections 6.2.2 and 7.2.1), and yet does produce thermostable proteases (Section 7.2.6). Though, since the right-hand end of the *intC* element has not yet been isolated there is the possibility that determinants involved in protease thermostability expression might be encoded by the right-hand end of the *intC* element, and that these sequences are present in strain A198.

In addition, many benign strains of *D. nodosus* contain the *intC* element sequences, though it is apparent that in the all benign strains studied in this work, the *intC* element does seem to be truncated (Section 7.2.1). Three benign strains also contain the complete *intD* element (Section 6.2.2), and do not produce thermostable protease.

Consequently, it seems that the most significant difference between parent strain 1311 and the putatively benign daughter strain 1311A (Section 7.2.6) is that the *intB* element is present adjacent to *askA* instead of the *intC* element (Section 7.2.2). Thus, it was hypothesised that virulent strains of *D. nodosus* have either the *intC* element or the *vap* element integrated adjacent to *askA* and the *vap* element adjacent to *pnpA*.

In order to test this hypothesis, the genomic DNA from seventeen strains of *D. nodosus* was analysed in Southern blot experiments in order to determine the integration sites of the *vap*, *intB* and *intC* elements in the *D. nodosus* chromosome. Probes specific for *askA*, *pnpA*, *intA*, *intB* and *intC* were utilised in these experiments (Figure 7.12; Appendices 4, 5, 7 and 8). The strains of *D. nodosus* analysed included virulent strains A198, 1311, B1006, G1220, H1215, D1172, intermediate strain AC3577, and benign strains C3052, 189, 1169, 2483, 1493, 3138, 1469, 1311A, H1204 and AC390. The results of these experiments were also confirmed in PCR experiments (Appendix 10). Results of these experiments are summarised in Table 7.5.

Analysis of the integration sites for the *vap*, *intB* and *intC* elements in the strains of *D. nodosus* analysed showed that most virulent strains have the *intC* element integrated adjacent to *askA* and the *vap* element next to *pnpA*. Virulent strain A198 is distinguished from other virulent strains by having a copy of the *vap* element integrated adjacent to the *askA* and *pnpA* position. In contrast, most benign strains contain either the *intC* element or the *vap* element in one of the two positions, but not both. Most benign strains appear to have a copy of the *intB* gene or part thereof in either the *askA* or *pnpA* position, whilst one

has a copy of the *intD* gene adjacent to *pnpA*, and another has an unknown integrated element adjacent to *pnpA*.

Table 7.5: Genetic elements integrated next to *askA* and *pnpA* in seventeen strains of *D. nodosus*

Strain	virulence classification	*askA* juxtaposition	*pnpA* juxtaposition
A198	virulent	*intA*	*intA*
B1006	virulent	*intC*	*intA*
D1172	virulent	*intC*	*intA*
G1220	virulent	*intC*	*intA*
H1215	virulent	*intB*	*intA*
1311	virulent	*intC, intB* *	*intA*
AC3577	intermediate	*intC, intB* *	*intA*
AC390	benign	*intB*	*not intA, intB, intC, intD*
C3052	benign	*intC*	*intB*
H1204	benign	*intB*	*intA*
819	benign	*intC*	*intD***
1169	benign	*intC*	*intB*
1311A	putative benign	*intB*	*intA*
1469	benign	*intC*	*intB*
1493	benign	*intC*	*intB*
2483	benign	*intC*	*intB*
3138	benign	*intC*	*not intA, intB, intC, intD*

* Evidence suggests that in strains 1311 and AC3577 the *intC* element is being lost from the *askA* position and when the *intC* element is lost the *intB* element sequences are repositioned such that it is immediately adjacent to *askA*.

** Preliminary Southern blot analyses in which *intD* was used as a probe suggest that the *intD* element is integrated adjacent to *pnpA* in strain 819.

These results suggest that virulent strains contain the *vap* element next to *pnpA* and either the *vap* element or the *intC* element adjacent to *askA*. Of the seventeen strains studied, the only exception to this is virulent strain H1215, in which *intB* and *askA* are juxtaposed.

It was possible that H1215 had been converted to a benign strain during passaging under laboratory conditions, in the same way as 1311A. Strain H1215 is atypical of virulent strains of *D. nodosus* in a number of ways: *(i)* H1215 is the only strain of *D. nodosus*

analysed that contains two copies of *intC*, one of which is adjacent to *orf242*, whilst the other copy is not. The integration site for either copy of *intC* in this strain is unknown (Section 7.2.1); *(ii)* this strain contains a complete inducible prophage DinoHI (Bloomfield, 1997); *(iii)* protease assays of this strain (data not shown) indicate that it still produces thermostable proteases, however, H1215 colonies are small compared to those of other virulent strains. Like virulent strains of *D. nodosus*, intermediate strains secrete thermostable proteases, however they are distinguished from virulent strains by having a smaller colony size due to reduced surface translocation (Stewart, 1989). Thus it is possible that virulence in strain H1215 may have been attenuated during culture under laboratory conditions due to genomic rearrangements and as a result may be an intermediate strain rather than a virulent strain. Unfortunately, clinical trials would probably be required to show this definitively.

In general, the hypothesis that virulent strains of *D. nodosus* have either the *intC* element or the *vap* element integrated adjacent to *askA* and the *vap* element adjacent to *pnpA* is supported by these Southern blot analyses. However, given that analysis of one of seventeen strains does not agree with the hypothesis, it would be valuable to investigate the integration sites of the *vap*, *intB* and *intC* elements in a larger number of *D. nodosus* strains. In particular the analysis of benign strains like benign strain AC390 that have multiple copies of the *vap* regions, and of virulent and benign strains of *D. nodosus* strains that carry single copies of the *vap* regions may yield valuable data.

It is now known that strain 1311A also does not contain a copy of the *intD* element sequences that are present in parent strain 1311, and that the loss of the *intC* and *intD* elements may be concomitant (Sections 7.2.4). It is unlikely that the *intD* element has a direct role in virulence since it is present in few strains of *D. nodosus* and in both benign and virulent isolates of *D. nodosus*.

Another virulence-related locus, the (*vrl*) region (Section 1.5.1) is thought to be associated with virulence and to have been acquired horizontally (Haring *et al.*, 1995; Katz *et al.*, 1991). Furthermore, it has been proposed (Haring *et al.*, 1995) that the *vrl* region may not itself encode a virulence function but rather that its integration into a putative *ssrA* gene may activate a virulence function elsewhere in the *D. nodosus* genome. Thus in order to confirm that the *vrl* region had not undergone rearrangement or been lost from the 1311 genome, probes specific for the left and right ends of the *vrl* region were used to probe genomic DNA from strains 1311 and 1311A (Appendix 9). Genomic DNA from strains A198 and C3052 were included as positive and negative controls respectively.

The probes specific for the left and right of the *vrl* region of *D. nodosus* strain A198 were shown to hybridise to identical *Hin*dIII, *Eco*RI and *Hin*dIII/*Eco*RI fragments in strains 1311 and 1311A, had a similar pattern to A198, and, as expected, did not hybridise at all to genomic DNA from strain C3052 (Appendix 9). Hence, these results indicate that any change in thermostable protease activity observed after loss of pDN1, *intC* and the *intD* element from strain 1311, was not associated with a rearrangement of the sequences flanking the *vrl* region.

7.2.9 Another genetic element located next to *pnpA* in benign strains of *D. nodosus*?

During analysis of the integration sites some anomalies were noticed in blots in which *pnpA* sequences were used as probes. *D. nodosus* strain A198 contains a copy of *intA* (the *vap* element) integrated downstream of *pnpA*, and in this strain *pnpA* hybridises to a 4.9 kb *Hin*dIII fragment. As expected all those strains that contain the *vap* element also hybridise to a 4.9 kb *Hin*dIII fragment (1311, 1311A, AC3577, B1006, G1220, H1204, AC390, D1172), except virulent strain H1215 which contains a copy of *vapE'* in place of *vapE* (Figure 1.9), and consequently *pnpA* hybridises to a 4.8 kb *Hin*dIII fragment in this