CHAPTER 3

Characterisation and modification of native *Dichelobacter nodosus* plasmid, pDN1

Even if I could be Shakespeare I think I should still choose to be Faraday.
Aldous Leonard Huxley (1894-1963)

3.1 Introduction

A number of genes encoding potential virulence determinants have been identified in *D. nodosus*. These include type 4 *N*-methylphenylalanine fimbriae (Anderson *et al.*, 1987; Emery, 1988; Emery, Stewart & Clark, 1984; Every, 1977) and extracellular proteases (Gordon, Yong & Woodward, 1985; Green, 1985a; Green, 1985b; Kortt, Burns & Stewart, 1983; Lilley, Stewart & Kortt, 1992; Moses & Yong, 1989; Yong & Gordon, 1986). In addition, using differential hybridisation, two virulence-associated DNA regions were isolated (Katz *et al.*, 1991), designated the virulence-associated protein (*vap*) region (Cheetham *et al.*, 1995b; Katz, Strugnell & Rood, 1992; Katz *et al.*, 1994) and the virulence-related locus (*vrl*) (Haring *et al.*, 1995; Katz *et al.*, 1991; Katz *et al.*, 1992). The *vap* regions are found in 98% of virulent strains and 27% of benign strains, whilst the *vrl* regions are found in 87% of virulent strains and only 2.5% of benign strains (Rood *et al.*, 1996), suggesting that the *vap* and *vrl* regions may have a role in virulence.

Despite the presence of a number of virulence-associated genes, there is, at present, no transformation system for *D. nodosus*, which precludes direct testing of the role of these and other genes in virulence. The development of a transformation system has been impeded by the lack of native bacteriophages and plasmids. Prior to this work, only one

native *D. nodosus* plasmid, pJIR896 (10 kb) has been reported (Billington *et al.*, 1996). pJIR896 was found to consist of a circular form of *vap* region 1 from *D. nodosus* virulent reference strain A198, together with a putative insertion sequence, IS*1253*. Attempts to develop a transformation system using pJIR896 and plasmids derived from other bacteria have so far been unsuccessful (Billington *et al.*, 1996).

In 1994, our laboratory acquired a new virulent strain of *D. nodosus* designated strain 1311, from the Regional Veterinary Laboratory, Wagga Wagga NSW, Australia. Genomic DNA was prepared from this strain of *D. nodosus*, and, serendipitously, a small putative plasmid molecule was identified after agarose gel electrophoresis (Clayton, 1994).

In this work, the putative plasmid was isolated, completely sequenced, characterised, and modified to contain appropriate antibiotic resistance markers and a multiple cloning site. Derivatives of the native *D. nodosus* plasmid were subsequently used in transformation experiments with the aim of developing a transformation system for *D. nodosus* so that the role of *vap* and other virulence associated genes in virulence could be determined directly.

3.2 Results

3.2.1 Identification of pDN1

pDN1 was identified after performing a routine genomic DNA preparation from *D. nodosus* strain 1311. Agarose gel electrophoresis of genomic DNA from strain 1311 revealed three DNA bands in addition to the characteristic genomic DNA band, corresponding to the nicked-circle, linear and covalently closed forms of a 5.1 kb plasmid,

pDN1 (Figure 3.1). This was confirmed by Southern Blot analysis in which all three conformations hybridised to a pDN1 DIG-labelled probe.

Using laser densitometry, the copy number of pDN1 was determined to be approximately 16 copies per copy of the *D. nodosus* chromosome. It was therefore apparent that pDN1 was different from the previously identified *D. nodosus* plasmid pJIR896, which is 10 kb in size, present in 2 copies per copy of the *D. nodosus* chromosome, and is a circular form of *vap* region 1 from *D. nodosus* strain A198 (Billington *et al.*, 1996).

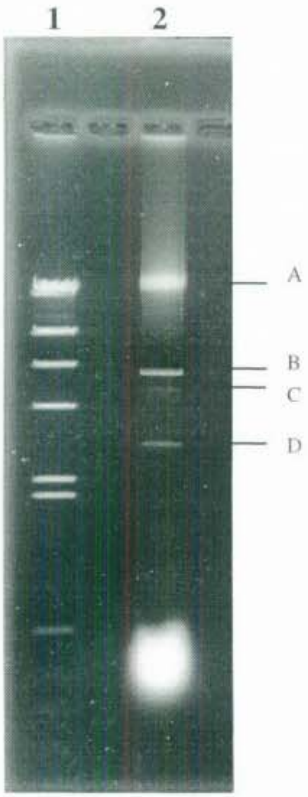

Figure 3.1:

Agarose gel electrophoresis of genomic DNA from *D. nodosus* strain 1311. The lambda *Hin*dIII size standards are shown in lane 1. In lane 2 bands corresponding to genomic DNA (A), covalently closed circular (B), linear (C) and supercoiled (D) conformations of pDN1 are visible.

3.2.2 Southern blot analysis of pDN1

Following the identification of pDN1 in *D. nodosus* strain 1311, ten other strains of *D. nodosus* ranging from benign to virulent (A198, AC3577, B1006, C305, G1220, H1024, H1215, 819, 1169, 2483) were probed with DIG-labelled pDN1. No hybridisation to pDN1 was observed for any strain other than *D. nodosus* strain 1311. It was thereby confirmed that there were no copies of pDN1 in the ten strains of *D. nodosus* analysed. pDN1 DNA was also probed with *vap* genes *A-D* and *toxA* and two probes that contain 4 kb and 2 kb fragments from the left- (pJIR743) and right-hand (pJIR590) ends of the *vrl* regions respectively (Haring *et al.*, 1995). No hybridisation to any of these *vap*- and *vrl*-specific probes was observed, which confirmed that pDN1 did not contain sequences from either of the *D. nodosus* virulence-associated DNA regions.

3.2.3 Restriction enzyme analysis and subcloning of pDN1

pDN1 was digested using restriction endonucleases with 6-bp recognition sequences in order to generate a restriction map (Figure 3.2). However, few restriction enzyme sites were identified. There is an apparent deficiency of 6-bp restriction sites in pDN1, a characteristic previously observed in broad host range plasmids (Dorrington & Rawlings, 1990). A lack of 6-bp sites may be an adaptation that increases the chance of survival of promiscuous plasmids in a broad range of hosts. Due to the small number of 6 bp sites, many of the subclones of pDN1 used for DNA sequencing were generated *via* shotgun cloning using restriction endonucleases *Sau*3AI, *Alu*I, *Taq*I and *Rsa*I. A complete map of subclones of pDN1 is shown (Figure 3.2).

3.2.4 DNA Sequence of pDN1

The complete nucleotide sequence of pDN1 is shown in Figure 3.3. pDN1 is a 5112 bp plasmid with 62% G + C content. Sequence analysis has revealed that pDN1 has a high degree of similarity to plasmids belonging to, or related to, the *Escherichia coli* incompatibility group Q (Table 3.1, Figure 3.4). At the nucleic acid level pDN1 has 96.5% identity over 5112 nt to a 8520 nt, IncQ-related plasmid, pIE1107 (positions 101 to 5201) from *Pseudomonas putida* (Tietze, 1998). pDN1 also has 65.7% identity over 3064 nt (position 1 to 3064) to *E. coli* IncQ plasmid RSF1010 (position 2348 to 5379). From position 3064 to 3339 the homology between pDN1 and RSF1010 is interrupted by 275 nt with an unknown function (Figure 3.4). Where identity resumes, pDN1 has 89.0% identity in 1773 nt (position 3339 to 5112) to RSF1010 (positions 5794 to 7567) (Scholz *et al.*, 1989). pDN1 also has 66.4% identity over 608 nt (positions 4477 to 5085 in pDN1) to *Thiobacillus ferrooxidans* plasmid pTF-FC2 at positions 1941 to 2550 (Dorrington & Rawlings, 1990).

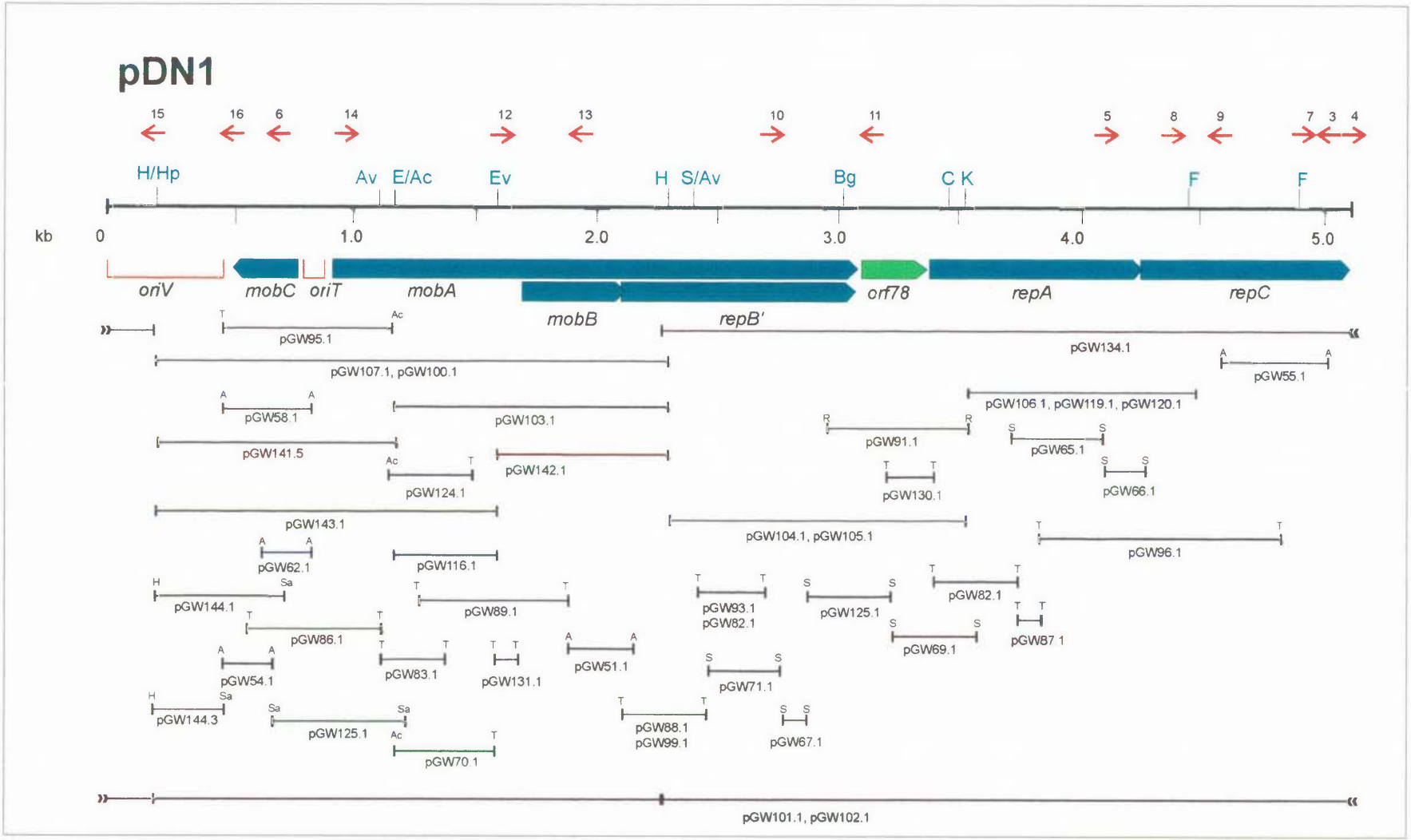

Figure 3.2: Map of subclones of pDN1, isolated from *D. nodosus* strain 1311. Restriction enzyme sites shown include *Acc*I (Ac), *Ava*I (Av), *Bgl*II (Bg), *Cla*I (C), *Fsp*I (F), *Hin*dII (H), *Hpa*I (Hp), and *Sma*I (S). Note that not all restriction enzyme sites are indicated on the map; restriction enzymes that cut at 4 bp recognition sites are indicated at the end of the appropriate fragment as follows; *Alu*I (A), *Taq*I (T), *Rsa*I (R), *Sau*3AI (Sa). Arrows indicate the positions of oligonucleotides used for sequencing, and are labelled as follows: Olignonucleotide G3 (3), G4 (4), G5 (5), G6 (6), G7 (7), GWSEQ8 (8), GWSEQ9 (9), GWSEQ10 (10), GWSEQ11 (11), GWSEQ12 (12), GWSEQ13 (13), GWSEQ14 (14), GWSEQ15 (15), GWSEQ16 (16). The arrows indicate the 5' to 3' direction of the oligonucleotide sequence. The numbers indicate the distance in kb.

Plasmids belonging to the *E. coli* incompatibility group Q are small, multiple copy, non-self transmissible plasmids, that have an extremely broad-host-range, and as a consequence have formed the basis of many cloning vehicles and have been used to transform a wide range of Gram-negative organisms unable to be manipulated using replicons derived from ColE1 or *E. coli* -specific bacteriophages (Bagdasarian *et al.*, 1981; Bagdasarian & Timmis, 1981; Frey & Bagdasarian, 1989; Morales, Bagdasarian & Bagdasarian, 1990; Morales, Baeckman & Bagdasarian, 1991; Scholz *et al.*, 1989). The similarity between pDN1 and plasmids belonging to the *E. coli* IncQ group suggested that pDN1 was a good candidate for the development of a transformation system for *D. nodosus.*

The G + C content of pDN1 is very similar to that of RSF1010 (61%) and pIE1107 (62%), but is considerably higher than the 45% G + C content of the *D. nodosus* chromosome (La Fontaine & Rood, 1990). This suggests the plasmid was originally derived from a different host, and that the acquisition was probably a relatively recent event. The first base of the vegetative origin of replication (*oriV*) was chosen as nucleotide number 1.

Figure 3.3: (on following pages) Complete sequence of native plasmid pDN1 from *D. nodosus* strain 1311. Both the nucleotide and the deduced amino acid sequence of open reading frames on both strands are shown. Note that for *repB'* only the first amino acid residue is shown since the amino acid sequence is translated in the same reading frame as *mobA*. Stop codons are indicated by red asterisks. Potential ribosome binding sites (Shine & Dalgarno, 1974) are indicated in green type. Four putative promoters that match the consensus for *E. coli* -10 and -35 promoter sequences, are also indicated and are designated P1 to P4. The region corresponding to the putative *oriV* and *oriT* regions are indicated in blue and red type respectively.


oriV
mobC
oriT
-10 P2
-10 P3
-35 P3
-35 P1
-35 P2 -10 P1
mobA

```
 A  D  Q  W  F  K  R  Y  N  G  K  Q  P  E  R  G  G  A  Q  K  S  E  S  L  K  P  K
GGCTGACCAGTGGTTTAAGCGGTACAACGGCAAACAGCCAGAGCGGGGCGGGGCGCAGAAGTCCGAGAGCCTGAAGCCCA 1440
CCGACTGGTCACCAAATTCGCCATGTTGCCGTTTGTCGGTCTCGCCCCGCCCCGCGTCTTCAGGCTCTCGGACTTCGGGT

  A  W  L  E  Q  T  R  E  A  W  S  H  Y  A  N  R  A  L  E  Q  A  G  H  E  A  R
AGGCATGGCTTGAGCAGACCCGTGAGGCATGGAGCCACTACGCCAACCGCGCCCTTGAACAGGCCGGGCACGAGGCCCGC 1520
TCCGTACCGAACTCGTCTGGGCACTCCGTACCTCGGTGATGCGGTTGGCGCGGGAACTTGTCCGGCCCGTGCTCCGGGCG

I  D  H  R  T  L  E  A  Q  G  I  E  R  L  P  G  I  H  L  G  P  N  V  V  E  M  E
ATTGACCATCGCACCCTTGAGGCGCAGGGCATTGAGCGCTTGCCGGGCATCCATCTTGGCCCTAACGTGGTGGAGATGGA 1600
TAACTGGTAGCGTGGGAACTCCGCGTCCCGTAACTCGCGAACGGCCCGTAGGTAGAACCGGGATTGCACCACCTCTACCT

 S  R  G  I  R  T  E  R  A  D  I  A  L  A  I  D  T  A  N  G  Q  I  I  D  L  Q  E
GAGCCGGGGCATCCGCACCGAGAGGGCCGATATCGCCCTAGCCATCGACACGGCCAACGGCCAGATCATCGACTTACAAG 1680
CTCGGCCCCGTAGGCGTGGCTCTCCCGGCTATAGCGGGATCGGTAGCTGTGCCGGTTGCCGGTCTAGTAGCTGAATGTTC

                mobB     M  S  A  I  D  R  V  R  K  S  R  G  I  N  E  L  A  A  E  I
  Y  R  E  V  I  E  H  E  R  D  R  Q  S  E  E  I  Q  R  D  Q  R  V  S  G  R  D
AATACCGGGAGGTTATCGAGCATGAGCGCGATAGACAGAGTGAGGAAATCCAGAGGGATCAACGAGTTAGCGGCAGAGAT 1760
TTATGGCCCTCCAATAGCTCGTACTCGCGCTATCTGTCTCACTCCTTTAGGTCTCCCTAGTTGCTCAATCGCCGTCTCTA

 E  P  L  A  Q  S  M  A  T  L  A  D  E  A  R  Q  R  I  A  E  V  Q  Q  A  S  E  E
R  A  A  G  P  E  H  G  D  T  G  G  R  S  P  A  A  D  R  R  G  T  A  G  Q  R  G
AGAGCCGCTGGCCCAGAGCATGGCGACACTGGCGGACGAAGCCCGGCAGCGGATCGCCGAGGTACAGCAGGCCAGCGAGG 1840
TCTCGGCGACCGGGTCTCGTACCGCTGTGACCGCCTGCTTCGGGCCGTCGCCTAGCGGCTCCATGTCGTCCGGTCGCTCC

  Q  A  A  S  W  T  S  Q  Q  Q  Q  A  M  S  A  W  R  Q  A  A  K  D  M  R  A  A
 A  G  R  E  L  D  E  P  A  T  A  S  H  E  R  M  A  A  G  S  K  G  H  E  G  S  R
AGCAGGCCGCGAGCTGGACGAGCCAGCAACAGCAAGCCATGAGCGCATGGCGGCAGGCAGCAAAGGACATGAGGGCAGCC 1920
TCGTCCGGCGCTCGACCTGCTCGGTCGTTGTCGTTCGGTACTCGCGTACCGCCGTCCGTCGTTTCCTGTACTCCCGTCGG

A  G  E  L  A  K  A  G  Q  T  A  R  S  A  A  R  G  W  T  W  R  L  W  A  G  V  L
  R  G  T  R  Q  G  R  P  D  G  P  E  R  R  P  W  L  D  M  E  V  V  G  R  G  L
GCAGGGGAACTCGCCAAGGCCGGCCAGACGGCCCGGAGCGCCGCCCGTGGCTGGACATGGAGGTTGTGGGCCGGGGTCTT 2000
CGTCCCCTTGAGCGGTTCCGGCCGGTCTGCCGGGCCTCGCGGCGGGCACCGACCTGTACCTCCAACACCCGGCCCCAGAA

 I  A  S  V  M  P  I  L  A  L  L  I  A  S  W  L  W  L  E  P  Q  I  I  E  Q  Q  G
D  R  F  R  D  A  Y  S  G  A  A  D  R  I  M  A  L  A  G  A  A  N  H  R  A  T  G
GATCGCTTCCGTGATGCCTATTCTGGCGCTGCTGATCGCATCATGGCTTTGGCTGGAGCCGCAAATCATCGAGCAACAGG 2080
CTAGCGAAGGCACTACGGATAAGACCGCGACGACTAGCGTAGTACCGAAACCGACCTCGGCGTTTAGTAGCTCGTTGTCC

repB′  M
  S  I  W  L  I  F  K  L  K  *
 E  H  M  A  D  F  Q  A  Q  V  K  G  D  R  T  A  Q  A  I  A  R  Q  L  K  A  M  G
GGAGCATATGGCTGATTTTCAAGCTCAAGTGAAAGGTGACCGCACCGCGCAAGCCATCGCCCGGCAGCTCAAGGCCATGG 2160
CCTCGTATACCGACTAAAAGTTCGAGTTCACTTTCCACTGGCGTGGCGCGTTCGGTAGCGGGCCGTCGAGTTCCGGTACC

  C  D  R  Y  D  I  G  I  R  D  A  A  S  G  K  M  M  N  R  E  W  T  P  Q  E  V
GCTGCGACCGGTACGACATCGGCATACGGGATGCCGCCAGCGGCAAGATGATGAACCGGGAATGGACACCGCAGGAAGTG 2240
CGACGCTGGCCATGCTGTAGCCGTATGCCCTACGGCGGTCGCCGTTCTACTACTTGGCCCTTACCTGTGGCGTCCTTCAC

Q  Q  N  A  A  W  L  K  R  M  N  A  Q  G  N  D  I  Y  I  R  P  A  E  Q  A  R  H
CAGCAGAACGCCGCCTGGCTCAAGCGCATGAATGCCCAGGGCAACGATATTTACATTCGCCCCGCCGAGCAGGCCCGGCA 2320
GTCGTCTTGCGGCGGACCGAGTTCGCGTACTTACGGGTCCCGTTGCTATAAATGTAAGCGGGGCGGCTCGTCCGGGCCGT

 G  L  V  L  V  D  D  L  S  S  D  D  L  D  A  M  K  Q  E  G  R  E  P  A  A  I  I
TGGTCTGGTGCTAGTTGACGACCTCAGCAGCGACGATCTGGACGCCATGAAGCAGGAGGGCCGGGAGCCTGCCGCCATCA 2400
ACCAGACCACGATCAACTGCTGGAGTCGTCGCTGCTAGACCTGCGGTACTTCGTCCTCCCGGCCCTCGGACGGCGGTAGT

  E  T  S  P  K  N  Y  Q  A  W  V  K  V  A  Q  D  A  P  A  D  H  R  G  V  I  A
TCGAGACCAGCCCCAAGAATTATCAGGCATGGGTGAAGGTGGCTCAGGATGCCCCAGCAGATCACAGGGGCGTAATCGCC 2480
AGCTCTGGTCGGGGTTCTTAATAGTCCGTACCCACTTCCACCGAGTCCTACGGGGTCGTCTAGTGTCCCCGCATTAGCGG

R  K  L  A  R  E  Y  D  A  D  P  A  S  A  D  S  R  H  Y  G  R  L  A  G  F  T  N
CGGAAGCTGGCCCGGGAGTATGACGCCGACCCGGCAAGCGCCGACAGCCGTCACTATGGCCGTCTGGCAGGCTTTACCAA 2560
GCCTTCGACCGGGCCCTCATACTGCGGCTGGGCCGTTCGCGGCTGTCGGCAGTGATACCGGCAGACCGTCCGAAATGGTT
```

```
 R  K  D  K  Y  T  S  R  T  G  Y  Q  P  W  V  L  C  R  E  S  S  G  K  S  A  T  A
CCGCAAGGACAAGTACACCAGCCGCACCGGCTATCAGCCGTGGGTGCTGTGCCGGGAGTCCAGCGGCAAGAGTGCCACCG 2640
GGCGTTCCTGTTCATGTGGTCGGCGTGGCCGATAGTCGGCACCCACGACACGGCCCTCAGGTCGCCGTTCTCACGGTGGC

  G  P  E  L  M  Q  Q  A  G  Q  V  L  D  S  I  E  R  R  Q  E  R  T  A  R  L  A
CAGGGCCGGAGCTGATGCAGCAGGCCGGGCAGGTGTTGGACAGCATCGAGCGCCGACAGGAACGGACGGCACGACTGGCC 2720
GTCCCGGCCTCGACTACGTCGTCCGGCCCGTCCACAACCTGTCGTAGCTCGCGGCTGTCCTTGCCTGCCGTGCTGACCGG

E  I  T  A  P  Q  S  V  R  R  Y  R  R  S  V  V  D  D  Y  R  S  E  M  A  G  L  V
GAGATCACCGCCCCGCAGAGTGTGCGTCGGTATCGCCGCAGCGTCGTGGACGACTACCGCAGCGAAATGGCCGGGCTGGT 2800
CTCTAGTGGCGGGGCGTCTCACACGCAGCCATAGCGGCGTCGCAGCACCTGCTGATGGCGTCGCTTTACCGGCCCGACCA

 K  R  Y  G  D  D  L  S  K  C  D  F  I  A  A  M  K  L  A  S  K  G  R  E  A  D  E
CAAGCGGTACGGTGATGATCTCAGCAAGTGCGACTTCATCGCGGCCATGAAGCTGGCCAGCAAGGGCCGGGAGGCGGACG 2880
GTTCGCCATGCCACTACTAGAGTCGTTCACGCTGAAGTAGCGCCGGTACTTCGACCGGTCGTTCCCGGCCCTCCGCCTGC

  I  G  K  A  M  A  E  A  S  P  A  I  M  E  R  K  A  G  H  E  A  D  Y  I  K  R
AGATCGGCAAGGCCATGGCCGAGGCTAGCCCCGCTATCATGGAGCGCAAGGCGGGCCATGAGGCCGATTACATCAAGCGT 2960
TCTAGCCGTTCCGGTACCGGCTCCGATCGGGGCGATAGTACCTCGCGTTCCGCCCGGTACTCCGGCTAATGTAGTTCGCA

T  V  Q  K  V  M  E  L  P  Q  V  Q  E  A  R  A  E  L  A  K  Q  T  Q  K  Q  R  S
ACGGTGCAGAAGGTGATGGAGCTTCCGCAGGTGCAGGAGGCCCGCGCCGAGCTGGCGAAGCAGACCCAGAAGCAGCGCAG 3040
TGCCACGTCTTCCACTACCTCGAAGGCGTCCACGTCCTCCGGGCGCGGCTCGACCGCTTCGTCTGGGTCTTCGTCGCGTC

 K  G  P  D  L  S  M  *     -35 P4                    -10 P4
CAAGGGGCCAGATCTAAGCATGTAATGCTTGTTTTGATAAAACAAAACTGCTATATTGTCACAACAATTTGACTTTATAG 3120
GTTCCCCGGTCTAGATTCGTACATTACGAACAAAACTATTTTGTTTTGACGATATAACAGTGTTGTTAAACTGAAATATC

    orf78  M  D  I  G  K  A  K  E  Y  F  G  L  G  L  I  Q  G  A  V  V  R  S  S
TAGGATTAAAAAAATGGATATAGGAAAGGCAAAAGAGTATTTTGGGCTTGGATTGATTCAGGGTGCCGTGGTGAGATCGTC 3200
ATCCTAATTTTTTACCTATATCCTTTCCGTTTTCTCATAAAACCCGAACCTAACTAAGTCCCACGGCACCACTCTAGCAG

 S  V  L  H  N  G  W  T  I  E  L  S  G  N  I  G  S  A  Q  P  T  L  H  T  A  R  G
AAGTGTTCTGCATAACGGGTGGACTATCGAGCTGAGCGGCAACATTGGCAGCGCACAACCTACCTTACACACCGCCAGAG 3280
TTCACAAGACGTATTGCCCACCTGATAGCTCGACTCGCCGTTGTAACCGTCGCGTGTTGGATGGAATGTGTGGCGGTCTC

  D  V  R  Q  F  K  T  L  D  A  A  A  K  V  V  R  E  I  G  L  R  E  W  R  V  I
GTGATGTGCGGCAGTTCAAGACACTGGATGCGGCGGCCAAGGTCGTGCGAGAAATTGGGCTGCGAGAGTGGCGTGTTATC 3360
CACTACACGCCGTCAAGTTCTGTGACCTACGCCGCCGGTTCCAGCACGCTCTTTAACCCGACGCTCTCACCGCACAATAG

T  D  *                                                                repA  M  A
ACTGACTAGGCGGCGCATGGCGGGGCATCCCCGCCAGCGCTTGACCATCAACCGCCTGCAAAGGAGGCAAACCAATGGCT 3440
TGACTGATCCGCCGCGTACCGCCCCGTAGGGGCGGTCGCGAACTGGTAGTTGGCGGACGTTTCCTCCGTTTGGTTACCGA

T  H  K  P  I  D  I  L  A  S  F  T  E  L  P  P  P  I  D  Y  V  L  P  N  M  V  A
ACTCATAAGCCTATCGATATTCTGGCGAGCTTCACAGAGTTACCGCCGCCGATTGACTACGTATTGCCCAACATGGTGGC 3520
TGAGTATTCGGATAGCTATAAGACCGCTCGAAGTGTCTCAATGGCGGCGGCTAACTGATGCATAACGGGTTGTACCACCG

 G  T  V  G  A  L  V  S  P  G  G  A  G  K  S  M  L  V  L  Q  L  A  A  Q  I  A  G
CGGTACCGTGGGGGCACTGGTGTCCCCCGGTGGGGCCGGCAAGTCCATGCTGGTTCTACAGCTGGCCGCACAGATCGCAG 3600
GCCATGGCACCCCCGTGACCACAGGGGGCCACCCCGGCCGTTCAGGTACGACCAAGATGTCGACCGGCGTGTCTAGCGTC

G  P  D  L  L  D  V  G  E  L  P  T  G  P  V  I  Y  L  P  A  E  D  P  P  T  A
GCGGCCCCGACCTGCTGGACGTGGGAGAGCTGCCCACCGGCCCGGTGATCTACTTGCCCGCCGAAGACCCGCCCACCGCT 3680
CGCCGGGGCTGGACGACCTGCACCCTCTCGACGGGTGGCCGGGCCACTAGATGAACGGGCGGCTTCTGGGCGGGTGGCGA

I  H  H  R  L  H  A  L  G  A  H  L  T  D  E  Q  R  Q  V  V  A  D  G  L  L  I  Q
ATACATCACCGCCTGCACGCATTGGGGGCGCACCTGACCGACGAGCAACGGCAAGTCGTGGCTGACGGTCTGCTAATCCA 3760
TATGTAGTGGCGGACGTGCGTAACCCCCGCGTGGACTGGCTGCTCGTTGCCGTTCAGCACCGACTGCCAGACGATTAGGT

 P  L  I  G  S  L  P  N  I  M  A  P  E  W  F  D  G  L  K  R  A  A  E  G  R  R  L
GCCGCTGATCGGCAGCCTGCCCAACATCATGGCCCCGGAGTGGTTCGACGGCCTCAAGCGCGCCGCCGAGGGCCGCCGCC 3840
CGGCGACTAGCCGTCGGACGGGTTGTAGTACCGGGGCCTCACCAAGCTGCCGGAGTTCGCGCGGCGGCTCCCGGCGGCGG
```

```
M  V  L  D  T  L  R  R  F  H  I  E  E  E  N  A  S  G  P  M  A  Q  V  I  G  R
TGATGGTGTTGGATACGCTGCGTCGGTTCCATATCGAGGAAGAGAACGCCAGTGGCCCCATGGCGCAGGTGATTGGCCGC 3920
ACTACCACAACCTATGCGACGCAGCCAAGGTATAGCTCCTTCTCTTGCGGTCACCGGGGTACCGCGTCCACTAACCGGCG

M  E  A  I  A  A  D  T  G  C  S  I  V  F  L  H  H  A  S  K  G  A  A  M  M  G  A
ATGGAGGCCATCGCCGCCGATACCGGGTGTTCTATCGTGTTCCTGCACCATGCCAGCAAGGGCGCGGCCATGATGGGCGC 4000
TACCTCCGGTAGCGGCGGCTATGGCCCACAAGATAGCACAAGGACGTGGTACGGTCGTTCCCGCGCCGGTACTACCCGCG

 G  D  Q  Q  Q  A  S  R  G  S  S  V  L  V  D  N  I  R  W  Q  S  Y  L  S  G  M  T
AGGCGACCAGCAGCAGGCCAGCCGGGGATCGTCCGTGCTGGTGGATAACATCCGCTGGCAGTCCTACCTGTCCGGCATGA 4080
TCCGCTGGTCGTCGTCCGGTCGGCCCCTAGCAGGCACGACCACCTATTGTAGGCGACCGTCAGGATGGACAGGCCGTACT

  A  A  E  A  E  E  W  G  V  D  D  S  Q  R  P  Y  F  V  R  F  G  V  S  K  A  N
CGGCCGCCGAGGCCGAGGAATGGGGCGTGGACGATTCACAGCGGCCCTATTTCGTCCGCTTTGGTGTCAGCAAGGCCAAC 4160
GCCGGCGGCTCCGGCTCCTTACCCCGCACCTGCTAAGTGTCGCCGGGATAAAGCAGGCGAAACCACAGTCGTTCCGGTTG

Y  G  A  P  F  K  D  R  W  F  R  R  H  D  G  G  V  L  K  P  A  V  L  E  K  Q  R
TATGGGGCGCCGTTCAAAGATCGCTGGTTCAGGCGGCATGACGGCGGGGTGCTCAAACCCGCCGTGCTGGAAAAGCAGCG 4240
ATACCCCGCGGCAAGTTTCTAGCGACCAAGTCCGCCGTACTGCCGCCCCACGAGTTTGGGCGGCACGACCTTTTCGTCGC

          repC    V  V  K  P  K  N  K  Y  S  L  S  H  V  R  H  D  P  A  H  C
 K  S  K  G  V  R  R  G  E  A  *
CAAAAGCAAGGGGGTGCGCCGTGGTGAAGCCTAAGAACAAGTACAGTCTCAGCCACGTCCGGCACGACCCGGCGCACTGT 4320
GTTTTCGTTCCCCCACGCGGCACCACTTCGGATTCTTGTTCATGTCAGAGTCGGTGCAGGCCGTGCTGGGCCGCGTGACA

L  A  P  G  L  F  R  A  L  K  R  G  E  R  K  R  S  K  L  D  V  T  Y  D  Y  G  D
CTGGCTCCCGGCCTGTTCCGTGCCCTCAAGCGGGGCGAGCGCAAGCGCAGCAAGCTGGACGTGACGTATGACTACGGCGA 4400
GACCGAGGGCCGGACAAGGCACGGGAGTTCGCCCCGCTCGCGTTCGCGTCGTTCGACCTGCACTGCATACTGATGCCGCT

 G  K  R  I  E  F  K  G  P  E  P  L  G  A  D  D  L  R  I  L  Q  G  L  V  A  M  A
CGGCAAGCGGATCGAGTTCAAGGGGCCGGAGCCGCTGGGTGCTGATGATCTGCGCATCCTGCAAGGGCTGGTGGCAATGG 4480
GCCGTTCGCCTAGCTCAAGTTCCCCGGCCTCGGCGACCCACGACTACTAGACGCGTAGGACGTTCCCGACCACCGTTACC

  G  P  N  G  L  V  L  S  P  E  P  K  T  E  G  G  Q  Q  L  R  L  F  L  E  P  K
CTGGGCCTAATGGCCTAGTGCTTAGCCCAGAACCCAAGACCGAAGGCGGGCAGCAGCTCCGGTTGTTCTTGGAACCCAAG 4560
GACCCGGATTACCGGATCACGAATCGGGTCTTGGGTTCTGGCTTCCGCCCGTCGTCGAGGCCAACAAGAACCTTGGGTTC

W  E  A  V  T  A  D  A  M  V  V  K  G  S  Y  R  A  L  A  R  E  I  G  Y  A  E  D
TGGGAGGCCGTTACCGCTGATGCCATGGTGGTTAAGGGCAGCTATCGGGCGCTGGCCCGTGAAATCGGGTATGCAGAGGA 4640
ACCCTCCGGCAATGGCGACTACGGTACCACCAATTCCCGTCGATAGCCCGCGACCGGGCACTTTAGCCCATACGTCTCCT

 G  G  S  Q  F  K  A  I  R  E  C  I  E  R  L  W  T  V  S  I  I  A  Q  N  G  R  K
TGGCGGCAGCCAGTTCAAGGCAATCCGGGAGTGCATTGAGCGCCTGTGGACAGTCTCCATCATCGCCCAAAATGGCCGCA 4720
ACCGCCGTCGGTCAAGTTCCGTTAGGCCCTCACGTAACTCGCGGACACCTGTCAGAGGTAGTAGCGGGTTTTACCGGCGT

R  Q  G  F  R  L  L  A  E  Y  A  S  D  E  A  G  G  R  L  Y  V  A  L  N  P  L
AGCGGCAGGGGTTCCGGCTGCTTGCCGAGTACGCCAGCGACGAGGCAGGCGGGCGTCTGTACGTGGCCCTGAACCCCTTA 4800
TCGCCGTCCCCAAGGCCGACGAACGGCTCATGCGGTCGCTGCTCCGTCCGCCCGCAGACATGCACCGGGACTTGGGGAAT

I  A  Q  A  V  M  G  G  G  Q  H  V  R  I  S  M  D  E  V  R  A  L  D  S  E  A  A
ATTGCTCAGGCCGTCATGGGTGGTGGTCAGCATGTACGCATCAGCATGGACGAGGTGCGAGCGCTGGACAGCGAAGCGGC 4880
TAACGAGTCCGGCAGTACCCACCACCAGTCGTACATGCGTAGTCGTACCTGCTCCACGCTCGCGACCTGTCGCTTCGCCG

R  L  L  H  Q  R  L  C  G  W  I  D  P  G  K  T  G  K  A  A  I  D  T  L  C  G  Y
CCGCCTGCTGCATCAGCGCCTGTGTGGTTGGATCGACCCCGGCAAGACCGGCAAGGCGGCCATAGATACCCTGTGCGGCT 4960
GGCGGACGACGTAGTCGCGGACACACCAACCTAGCTGGGGCCGTTCTGGCCGTTCCGCCGGTATCTATGGGACACGCCGA

  V  W  P  S  E  A  S  G  S  T  M  R  K  R  R  Q  R  V  R  E  A  L  P  E  L  V
ATGTCTGGCCGTCAGAGGCCAGTGGTTCGACCATGCGCAAGCGCCGCCAGCGGGTGCGCGAGGCGTTGCCGGAGCTGGTT 5040
TACAGACCGGCAGTCTCCGGTCACCAAGCTGGTACGCGTTCGCGGCGGTCGCCCACGCGCTCCGCAACGGCCTCGACCAA

A  L  G  W  T  V  T  E  F  A  A  G  K  Y  D  I  T  R  P  K  A  A  G  *
GCGCTGGGTTGGACGGTAACCGAGTTCGCGGCGGGCAAGTACGACATCACCCGGCCCAAGGCGGCAGGCTGA 5112
CGCGACCCAACCTGCCATTGGCTCAAGCGCCGCCCGTTCATGCTGTAGTGGGCCGGGTTCCGCCGTCCGACT
```

3.2.5 Identification of *orfs* in pDN1

The sequence of pDN1 was examined in both directions for potential protein-coding regions beginning with a ATG or GTG start codon, preceded by Shine-Dalgarno consensus sequences, and coding for more than 50 amino acids. Seven putative open reading frames were identified (Table 3.1). All of the open reading frames from pDN1, except *repA* which has 99.3% identity to a protein in pIE1107 (Table 3.1), encode proteins which are identical to the putative proteins encoded by the IncQ-like plasmid, pIE1107 (Tietze, 1998). All of the of open reading frames identified in pDN1 except one (*orf78*) were found to have a very high degree of amino acid identity to the gene products from the extensively characterised plasmid RSF1010 which are known to be involved in plasmid replication (*repA-C*) and mobilisation (*mobA-C*) (Frey & Bagdasarian, 1989; Haring & Scherzinger, 1989; Scholz *et al.*, 1989).

Due to the high degree of similarity between pDN1 and IncQ-like plasmids, the orfs identified in pDN1 (*mobA-C* and *repA-C*), have been given names that are consistent with their homology (Table 3.1). The order and orientation of the open reading frames in pDN1 are the same as in pIE1107 and RSF1010 (Figure 3.4).

The high degree of amino acid similarity between the pDN1 proteins and the extensively characterised replication and mobilisation proteins of RSF1010 suggests that the two groups of proteins have comparative functions. This is significant because the characteristically broad-host-range of the IncQ plasmids has been attributed to three plasmid-encoded replication proteins (RepA, RepB, RepC) that encode many of the essential proteins that are required in the plasmid DNA replication initiation complex (Frey & Bagdasarian, 1989; Haring & Scherzinger, 1989; Scholz *et al.*, 1989), hence making plasmid replication initiation independent of the host primosomal functions.

Table 3.1: Sequence analysis of *D. nodosus* native plasmid pDN1

Gene	Co-ordinates 5'-3' (nt)	Size aa	% Identity to putative protein	Homologues Description	Accession Number	P(*n*)
mobC	476-772 (comp)	98	100/98aa	*Pseudomonas putida* Plasmid pIE1107	Z74787	1.4e-39
			75.9/87aa	*Escherichia coli* IncQ plasmid RSF1010 MobC protein	M28829	1.5e-25
			39.6/91aa	*Thiobacillus ferooxidans*plasmid pTF1 mobilisation region	X52699	7.3e-06
mobA	953-3064	703	100/703aa	*P. putida* Plasmid pIE1107	Z74787	0
			71.8/710aa	*E. coli* IncQ plasmid RSF1010 MobA protein	M28829	0
			38/295aa	*Salmonella typhimurium* plasmid pSC101	X01654	2.3e-17
			26.9/171aa	*T. ferooxidans* plasmid pTF-FC2 MobS protein	M64981	2.8e-08
			28.8/184aa	*A. tumefaciens* Ti plasmid TraA protein	U43674	4.5e-01
mobB	1701-2111	136	100/136aa	*P. putida* Plasmid pIE1107	Z74787	0
			59.8/132aa	*E. coli* IncQ plasmid RSF1010 MobB protein	M28829	6.6e-26
repB'	2087-3064	325	100/325aa	*P. putida* Plasmid pIE1107	Z74787	0
			70.6/330aa	*E. coli* IncQ plasmid RSF1010 RepB' protein	M28829	0
			26.7/281aa	*T. ferooxidans* plasmid pTF-FC2 RepB primase	M64981	8.9e-15
repA	3434-4273	279	99.3/279aa	*P. putida* Plasmid pIE1107	Z74787	0
			92.5/279aa	*E. coli* IncQ plasmid RSF1010 RepA protein	M28829	0
			44.4/286aa	*T. ferooxidans* plasmid pTF-FC2 RepA primase	M64981	0
repC	4260-5112	283	100/283aa	*P. putida* Plasmid pIE1107	Z74787	0
			92.2/283aa	*E. coli* IncQ plasmid RSF1010 RepC protein	M28829	0
			66.1/283aa	*T. ferooxidans* plasmid pTF-FC2 RepC protein	M64981	0
orf78	3133-3369	78	100/78aa	*P. putida* Plasmid pIE1107	Z74787	4.8e-36

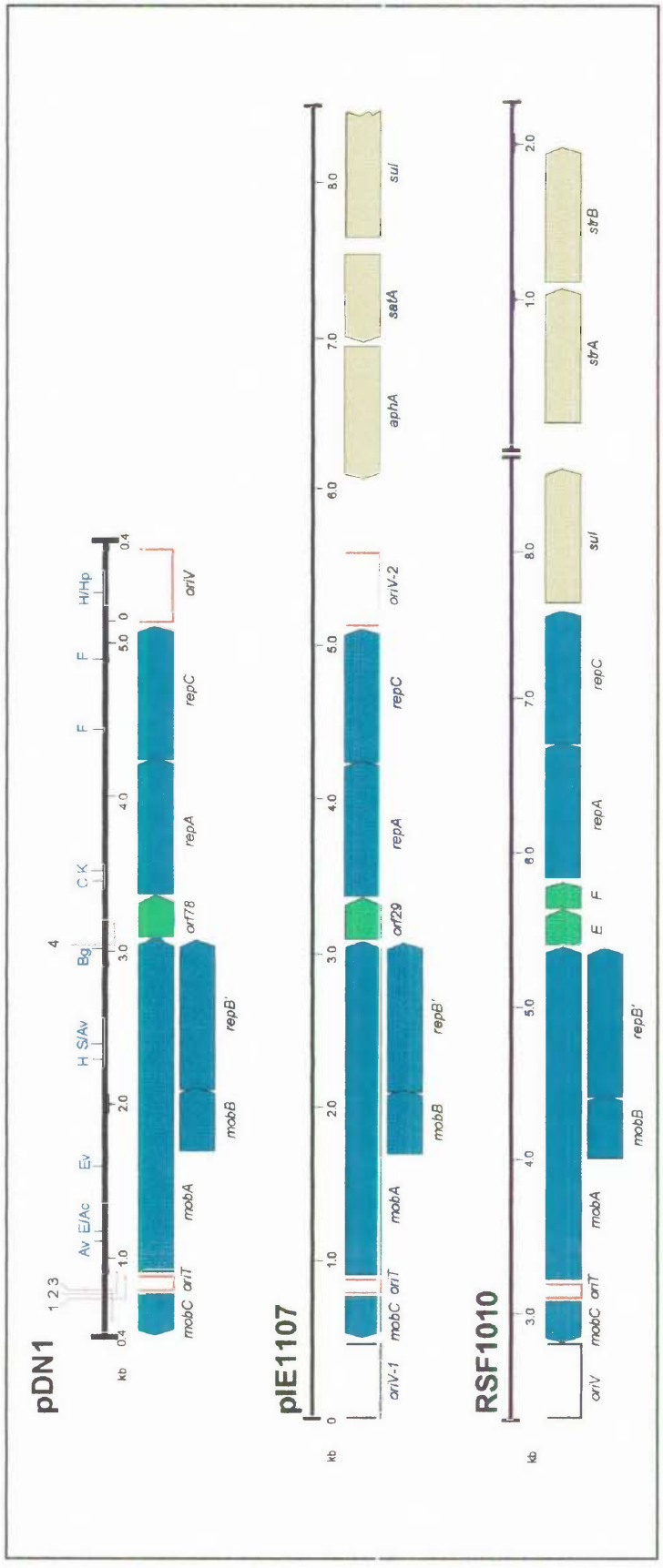

Figure 3.4: Alignment of pDN1 with related plasmids pIE1107 [Tietze, 1996] and RSF1010 [Scholz *et al.*, 1989]. The numbers indicate the distances (kb) from the position designated as nucleotide number 1 in all three sequences. Restriction enzyme sites shown for pDN1 include *Acc*I (Ac), *Ava*I (Av), *Bgl*II (Bg), *Cla*I (C) *Eco*RI (E), *Eco*RV (Ev), *Hind*II (H), *Kpn*I (K) and *Sma*I (S). The major open reading frames present in pDN1 and the direction of transcription is indicated by blue or green arrows, whilst those genes not found in pDN1 are distinguished by yellow arrows. Key to map is as follows: *oriV* - vegetative origin of replication; *oriT* - origin of conjugative transfer; *mobA*, *mobB* and *mobC* correspond to mobilisation proteins *A*, *B* and *C* respectively; *repA*, *repB* and *repC* indicate replication protein genes *A*, *B* and *C*; *aphA* - aminoglycoside-3'-phosphotransferase, a kanamycin and neomycin resistance determinant; *sul* - sulfonamide resistance; *str* - genes *A* and *B* respectively encode stretomycin resistance; *sat3* -streptothricin resistance. Note that *sul* is disrupted in pIE1107. *F*- repressor protein F. The functions of putative *orf78* and gene *E* are unknown. Putative promoter sequences 1-4 and the direction of transcription from these promoter sequences is indicated by and arrow. The *oriV*-2 of pIE1107 that is most similar to the *oriV* from pDN1 is shown in red.

The functions of the RSF1010 replication proteins include the recognition of the iterons in the *oriV* region (RepC) (Haring *et al.*, 1985). The binding of RepC to the *oriV* initiates the conformational changes necessary to facilitate access of the RepA-encoded helicase (Kim & Meyer, 1991; Miao *et al.*, 1993). RepA catalyses the unwinding of dsDNA and stabilisation of ssDNA (Haring & Scherzinger, 1989), exposing the single-stranded initiation signals (*ssiA* and *ssiB*) which are *cis*-acting nucleotide sequences in the *l*-strand and *r*-strand respectively, which direct the priming of their complementary strands by the RepB' primase (Haring & Scherzinger, 1989; Miao *et al.*, 1993; Sakai & Komano, 1996).

Hence, preliminary evidence based on sequence similarity suggested that pDN1 and derivatives of pDN1 are likely to be able to replicate in *E. coli*, which would eliminate the need to develop a *E. coli-D. nodosus* shuttle vector that would otherwise be required, at least in the initial stages, in the development of a transformation system for *D. nodosus*.

The three genes designated *mobA*, *mobB* and *mobC* together with the *oriT* region of IncQ plasmids have been shown to be required for the conjugative transfer of these non-self transmissible plasmids by conjugative plasmids from the incompatibility groups IncP, IncIα, IncM and IncX (Derbyshire, Hatfull & Willets, 1987; Derbyshire & Willetts, 1986; Frey & Bagdasarian, 1989; Willetts & Wilkins, 1984). However, the precise functions of these proteins and their role in conjugative transfer have not yet been determined.

In pDN1 a 300 nt sequence separates genes *mobA* and *repA*. The corresponding region in RSF1010 consists of a different sequence of approximately 500 nt. Within these different sequences a putative orf, called *orf78* is present in pDN1, whilst in RSF1010 genes *E* and *F* are present in the corresponding position (Figure 3.4). A putative protein which is identical to that encoded by *orf78* is encoded by *orf29* of pIE1107. The function of gene *E* in RSF1010 is not known, however gene *F* has been demonstrated to have a role

in the negative feedback control of *repA* and *repC* (Maeser *et al.*, 1990; Scholz *et al.*, 1989). Whether *orf78* has a similar role in the pDN1 and pIE1107 replicons is unknown.

Codon analysis of pDN1 genes indicates that the genes *mobA-C* and *repA-C* have a strong preference for G or C in the third position (77.5%), whilst in *orf78* only 56.4% of codons have a G or C in the third position, suggesting that *orf78* has been incorporated into pDN1 more recently. Similarly in RSF1010 the genes encoding plasmid replication, maintenance and mobilisation functions have a strong codon bias for a G or C residue in the third position of a codon (78.3%), whilst genes encoding antibiotic resistance and gene *E* have a much lower preference for G or C in that position (61.0%) (Scholz *et al.*, 1989).

It is interesting to note that between the TAG stop codon of *orf78* (position 3367) and the ATG start codon of *repA* there is a sequence of 80.1% identity in 68 bp to RSF1010 (position 5822 to 5889). This 68 bp encodes the last 12 aa of the gene *F*, and so these sequences may be remnants of gene *F*.

pDN1 is distinguished from both RSF1010 and pIE1107 by the lack of antibiotic resistance determinants (Figure 3.4). RSF1010 contains antibiotic resistance genes *sul*, and *strA-strB* which encode resistance to sulfonamides and streptomycin respectively (Frey & Bagdasarian, 1989), whilst pIE1107 contains the antibiotic resistance genes *aphA* and *sat3* which confer resistance to kanamycin/neomycin and streptothricin respectively (Tietze, 1998).

Although most plasmids in the IncQ group confer resistance to streptomycin and sulfonamides (Frey & Bagdasarian, 1989), some IncQ plasmids that differ in the type of antibiotic resistance determinants they carry have been reported (Fling, Kopf & Richards, 1988; Rotger, Rubio & Nombela, 1986; Tietze, 1998). There is no report in the literature of IncQ plasmids that do not harbour antibiotic resistance determinants. Furthermore, with