# Molecular Characterization of a Genomic Region Associated with Virulence in *Dichelobacter nodosus*

MARGARET E. KATZ, † RICHARD A. STRUGNELL, ‡ AND JULIAN I. ROOD\*

Department of Microbiology, Monash University, Clayton 3168, Australia

Received 27 April 1992/Accepted 25 August 1992

The major pathogen implicated in footrot, a highly contagious disease of sheep, is the strict anaerobe Dichelobacter nodosus (formerly Bacteroides nodosus). Sequence analysis of a 2,262-bp segment of the D. nodosus genome which is more prevalent in virulent isolates than in other isolates showed the presence of four open reading frames which appeared to have consensus transcriptional and translational start signals. These virulence-associated genes have been designated vapABCD. Two of the three copies of the vap region in the genome of the reference strain D. nodosus A198 were shown to carry all of the vap genes, whereas one copy contained only the vapD gene. The VapD protein was gel purified, shown to contain the predicted amino-terminal sequence, and used to raise rabbit antibodies. Western blots (immunoblots) showed that all of the D. nodosus strains tested that contained the vap region produced the VapD protein. The VapD protein had significant amino acid sequence identity with open reading frame 5 from the cryptic plasmid of Neisseria gonorrhoeae, and the vapBC operon had sequence similarity with the trbH region of the Escherichia coli F plasmid. It is proposed that these gene regions evolved from the integration of a conjugative plasmid from another bacterial species into the D. nodosus chromosome.

The strict anaerobe Dichelobacter nodosus (formerly known as Bacteroides nodosus [10]) is the causative agent of footrot, a highly contagious and economically significant disease of sheep. The extent of the disease in infected sheep is dependent on climatic conditions and the virulence of the invading microorganism. As a result, isolates of D. nodosus often are classified as virulent, intermediate, or benign, depending on the severity of the disease they would cause under optimal climatic conditions (37). Potential virulence factors produced by D. nodosus isolates include polar type 4, or N-MePhe, fimbriae (11) and extracellular proteases (27), although the roles of these factors in pathogenesis have not been determined. Virulent and benign isolates have been shown to differ in the amount of elastase that they produce (36, 38) and in the thermostability (6, 9, 13) and isoenzyme profiles (12, 20) of their respective proteases. Other studies have revealed differences in their levels of twitching motility (7, 8).

Studies in this laboratory have been aimed at the development of DNA probes to be used in diagnostic tests for ovine footrot. We have constructed a gene bank from the virulent reference strain D. nodosus A198 and by using a comparative hybridization procedure have identified three recombinant plasmids (pJIR318, pJIR313, and pJIR314B) which hybridize primarily with nonbenign isolates of D. nodosus and therefore could possibly be used as diagnostic gene probes (17).

The plasmid pJIR318 hybridized with almost all of the virulent and intermediate strains of D. nodosus tested. The results indicated that it contained sequences which correlated with virulence and were absent from 67% of the benign isolates. Southern hybridization experiments also showed that there were three copies of the pJIR318 region in the D.

nodosus A198 genome (17). In this paper, we report the entire nucleotide sequence of pJIR318 and the identification of several open reading frames (ORFs). One of these ORFs encoded a cytoplasmic *D. nodosus* protein which had an amino acid sequence similar to that of a protein encoded by the cryptic plasmid from *Neisseria gonorrhoeae*.

## MATERIALS AND METHODS

**Bacterial strains.** All *Escherichia coli* strains were derivatives of DH5 $\alpha$  (Bethesda Research Laboratories) or TG1 (Boehringer Mannheim) and were grown in 2× YT medium (26) supplemented with ampicillin (100 µg/ml). *D. nodosus* strains were from J. Egerton, University of Sydney (strains B1006 and C1008); L. Depiazzi, RVL Bunbury (strains AC419, AC424, AC554, AC6, and AC176); and W. Yong, RVL Hamilton (strains A198, 305, HA343, HA352, HA337, HA390, HA386, HA393, HA389, HA340, and HA652) and were grown in TAS broth (34) at 37°C in an atmosphere of 10% H<sub>2</sub> and 10% CO<sub>2</sub> in N<sub>2</sub>. *N. gonorrhoeae* strains MS-11A, Gc var40, and JKD109 were from T. Meyer (25), R Demarco de Hormeache (5), and J. Davies (Monash University), respectively. Strain JKD109 was used because it did not carry the cryptic plasmid.

**DNA sequencing.** General molecular techniques used in cloning and in the analysis of DNA molecules were as described previously (3, 30). DNA sequencing was performed by using T7 DNA polymerase kits and double-stranded DNA templates in the dideoxy-mediated chain termination method (31) as described by the manufacturers (Pharmacia; Promega).

Southern blot analysis. Samples  $(1 \ \mu g)$  of *D. nodosus* genomic DNA prepared as previously described (1) were digested with *Hin*dIII. The DNA was fractionated by electrophoresis through a 0.8% agarose gel and transferred (35) to a nitrocellulose membrane (BA-85; Schleicher & Schuell). Southern blots were done at high stringency as before (17).

**PAGE of proteins.** Cell extracts were prepared from approximately 100  $\mu$ g (wet weight) of pelleted *E. coli* or *D*.

<sup>\*</sup> Corresponding author.

<sup>†</sup> Present address: Department of Animal Science, University of New England, Armidale 2351, Australia.

<sup>&</sup>lt;sup>‡</sup> Present address: Department of Microbiology, University of Melbourne, Parkville 3052, Australia.

*nodosus* cells by boiling the cells for 5 min in the presence of 0.1 M Tris-HCl buffer, pH 6.8, containing 10% glycerol, 1.25% sodium dodecyl sulfate (SDS), and 5%  $\beta$ -mercaptoethanol. Proteins were separated by SDS-polyacrylamide gel electrophoresis (PAGE) (21, 32), transferred to nitrocellulose (39) and reacted in immunoblots (2) with anti-VapD antiserum diluted 1/100 and sheep anti-rabbit immunoglobulin (Silenus).

Purification of VapD and production of antiserum. A 1.5-ml overnight culture of JIR1793 [DH5 $\alpha$ (pJIR318)] was pelleted by centrifugation (all centrifugation steps were at  $12,000 \times g$ ) for 5 min, and the cell pellet was extracted in 6 M urea-50 mM sodium acetate (pH 6.0) for 12 h at 22°C. The extract was centrifuged for 30 min, and 1.0 ml of supernatant was precipitated with 9.0 ml of methanol, cooled to  $-20^{\circ}$ C, and kept at  $-20^{\circ}$ C overnight. The precipitate was collected by centrifugation, dissolved in sterile phosphate-buffered saline, and separated on a preparative 15% (wt/vol) SDS-PAGE gel. The band containing VapD was visualized with PAGE blue 83 (BDH Ltd), and VapD was electroeluted by using a Biotrap (Schleicher & Schuell) electroeluter. The eluted protein was precipitated with 9 volumes of methanol as described above. The purified VapD protein (100  $\mu$ g) was used with an Applied Biosystems 470A Gas-Phase Protein Sequencer to derive the amino-terminal sequence. To prepare specific antibodies, 100 µg of purified VapD was emulsified in Freund incomplete adjuvant and injected twice, 10 weeks apart, intramuscularly into a healthy adult male rabbit. The rabbit was exsanguinated 4 weeks later.

Nucleotide sequence accession number. The GenBank accession number of the nucleotide sequence determined in this study is M74565.

### RESULTS

**DNA sequence analysis of pJIR318.** The virulence-associated plasmid pJIR318 has been mapped and shown to contain a 2.3-kb insert of *D. nodosus* A198 DNA (17). The complete nucleotide sequence of pJIR318 was determined on both DNA strands and is shown in Fig. 1. Analysis of this sequence showed the presence of four short ORFs preceded by consensus ribosome-binding sites (33). These ORFs have been designated vapA through vapD (virulence-associated proteins). A fifth ORF (ORF118), 118 amino acids in length, which lacks a Shine-Dalgarno sequence is also present. Consensus promoter sites (15) are located upstream of vapA, vapB, and vapD but not vapC (Fig. 1).

Southern hybridization analysis of vap region. There are three copies of the vap region in D. nodosus A198, represented by the 6.2-, 4.6-, and 3.5-kb HindIII fragments which hybridize with pJIR318 (17). To see whether these copies each contained the entire vap region, Southern blots of genomic DNA from the virulent strain D. nodosus A198 and the benign strain D. nodosus 305 were hybridized with probes specific to each vap gene. Blots were prepared by using D. nodosus A198 DNA digested with HindIII, which does not cut within the pJIR318 insert. Multiple hybridizing bands were present in each A198 blot, whereas no hybridization was observed in the strain 305 lanes. The results showed that although the 6.2- and 4.6-kb HindIII fragments hybridized with each of the vapA-, vapB-, vapC-, and vapD-specific gene probes, the 3.5-kb fragment hybridized only with the vapD-specific gene probe (data not shown). This fragment therefore does not contain a copy of the vapABC genes.

Amino acid sequence analysis. The amino acid sequences

of the putative VapA (11.3 kDa), VapB (8.9 kDa), VapC (15.2 kDa), and VapD (10.6 kDa) proteins and ORF118 were analyzed for membrane-spanning domains by using the ALOM program (18). The results indicated that only the VapC protein had the potential to be an integral membrane protein. The amino acid sequences also were compared with the protein sequences in various data banks. Homology was detected between VapD and the sequence, found in the Protein Data Base of Japan, of ORF5 from the *N. gonor-rhoeae* cryptic plasmid (19). The sizes of the *D. nodosus* and *N. gonorrhoeae* ORFs are similar (93 and 92 amino acids, respectively), and 46% of the amino acids are identical (Fig. 2A). No significant amino acid sequence similarity with the other Vap proteins was found.

However, when the nucleic acid sequence of pJIR318 was compared with the DNA sequences in GenBank by using the FASTA search program (28), it was discovered that the vapB-vapC region of pJIR318 showed 53% identity in 706 bp of overlap with the traD-traI intergenic region of the E. coli F plasmid. This region of the F plasmid contains an ORF (trbH) which could encode a 26-kDa protein (4, 16). No ORF similar to trbH is present in pJIR318. Examination of the trbH region shows that two ORFs with homology to vapB and vapC are encoded by the DNA strand complementary to trbH (Fig. 2B). The two F-plasmid ORFs, which we have designated TRAORF1 and TRAORF2, are both preceded by Shine-Dalgarno sequences, and an RNA polymerase-binding site has been identified upstream of TRAORF1 (16, 24). Just as the initiation codon of vapC overlaps the stop codon of vapB, the initiation codon of TRAORF2 overlaps the stop codon of TRAORF1.

Production of the VapD protein in E. coli. To determine whether any of the D. nodosus vap genes are expressed in E. coli, cell extracts prepared from JIR1793 [DH5α(pJIR318)] were fractionated on polyacrylamide gels. An abundant, low-molecular-weight protein was visible in extracts from cells harboring pJIR318 (Fig. 3). No protein of a similar size was visible in extracts from cells carrying the vector pUC18. The low-molecular-weight protein was tentatively designated the vapD gene product, as it was expressed in E. coli cells carrying pJIR330, a deletion derivative of pJIR318 which contained only vapD (Fig. 3). The apparent size of the putative vapD gene product (approximately 6 kDa) was much lower than the molecular weight predicted from the deduced VapD amino acid sequence (10.6 kDa). To verify that the 6-kDa protein was the vapD gene product, the protein was gel purified and the N-terminal sequence was determined. The protein sequence that was generated, Met-Tyr-Ala-Ile, correlated with the putative start codon and N terminus of VapD that was deduced from the DNA sequence.

The VapD protein produced in *E. coli* could have been smaller than predicted because of posttranslational cleavage of the carboxy terminal end. To determine whether the C terminus of the *vapD* gene product produced in *E. coli* was intact, plasmid pJIR515 was constructed. This plasmid was a derivative of pJIR318 in which the 3' end of *vapD* had been deleted. The *vapD* gene product encoded by this plasmid should have been 4 amino acids smaller because of the loss of 18 amino acids from the C-terminal end of *vapD* and the addition of 14 amino acids due to continuation of the ORF into vector sequences. If posttranslational cleavage occurred, then the product of the deleted gene would be expected to be identical in size to the product of the intact gene. The results showed that the VapD product encoded by pJIR515 was of a higher apparent molecular weight than

Sau3A	
-3510	
	100
CTAGCGTAATTTTCGCGTCCTAAAAAGGAATAGAAAAACTAACT	
DraI vapD	
<u>SD</u> MYAIAFDLEIAELKKHYGEPYNGAYLE	
AACATGAGGAATTTAAAGATGTACGCAATTGCTTTTGATTTAGAAATTGCCGAATTAAAAAAACACTATGGCGAACCGTACAATGGCGCATACTTAGAAA	200
${\tt TTGTACTCCTTAAATTTCTACATGCGTTAACGAAAACTAAATCTTTAACGGCTTAATTTTTTTGTGATACCGCTTGGCATGTTACCGCGTATGAATCTTT$	
I G K E L K A I G F E W T Q G S V Y L S N D N N L A T V Y R A I S T TTGGTAAAGAACTTAAAGCAATCGGATTTGAATGGACACAAGGGAGCGTTATCTATC	300
DraI Sau3A	
LSKIDWFKSSVRDIRAFKVEDWSDFTAIVKGA*	
CCTGTCTAAGATTGATTGGTTTAAAAGTTCCGTAAGAGACATTAGAGCATTCAAAGTTGAGGATTGGTCAGATTTTACCGCAATAGTGAAAGGTGCATGA GGACAGATTCTAACCAAATTTTCAAGGCATTCTCTGTAATCTCGTAAGTTTCAACTCCCAGTCTAAAAATGGCGTTATCACTTTCCACGTACT	400
TCATCTAAACGCCCAGATTCAAGCGGAGGCATGGACTAAGTATCCTCCCAAAAAATATGAAGCAGCCAAATAGGCTGCTTTTTTTT	500
EcoRV	
${\tt taaggttatggtgggagtatgtcaatctatcacccaattatctaaccccatcaacgcgctcaaattctttggtattatttgttaccaacgtgata {\tt attccaataccccctcatacagttgggttagtgggttaatagattgtgggttgggtggg$	600
* D I V W N D L M L G D V R E F E K T N N T V L T I	
Nsil Kpnl	
TCAAGCGCCAATGCATGGCTAGCGATAAGCATATCTAAAGCCCCGATAGGGGTACCGCGGTTTTGCAGTGATTGCCGGATTGCGGATTGCCGAAAAGCACGGCAAAAAG AGTTCGCGGTTACGTACCGATCGCTATTCGTATAGATTTCGGGGCTATCCCCATGGCGCCAAAACGTCACTAACGGCCTAAGCACGCCATACGGTTTATC D L A L A H S A I L M D L A G I P T G R N Q L S Q R I R A Y H W I A	700
CTTGTTTATCAAAAGGCAATATTGATAGTGGTGCCAAAAATTTTGTTAATGCGGTTTTGTTTTTTGCTGAACCTGATTTTTCTACGCCAAACGCAAGCTC GAACAAATAGTTTTCCGTTATAACTATCACCACGGTTTTTAAAAACAATTACGCCAAAACAAAAAACGACTTGGACTAAAAAGATGCGGTTTGCGTTCGAG Q K D F P L I S L P A L F K T L A T K N K A S G S K E V G F A L E	800
NruI	
ACAAGCGGTAATGTTAGATATACCAATATCGCCTATTTCATATCGCGAAAATTTTTTCGGCAATATATGGCGGTTTTCGATTGATGATATAAATGCAGATA	900
TGTTCGCCATTACAATCTATATGGTTATAGCGGATAAAGTATAGCGCTTTTAAAAAGCCGTTATATACCGCCAAAAGCTAACTACTATATTTACGTCTAT	
C A T I N S I G I D G I E Y R S F K E A I Y P P K R N I I Y I C I	
DraI	
eq:ttggtatcaagcattaatttgatactcataaaaagtcttcccgcacttgttctatttgcggctcacgctcaattttaaaatcaggctcaaactcattgagaaccatgttcgtaatttaaaattttagtccgagtttgagtaactcattgagtcgtgatactaactccgagtttgagttgagtttgagtgagggggg	1000
* L F D E R V Q E I Q P E R E I K F D P E F E N L	
NTDLMLKISM SD <b>vap</b> C	
тастросла, а а а сертасса а весса а вретета асаса вета а та а а а вессета се та се се а а вессета а вета се то	1100
	1100
A E L F G Q W D S N K P I L L I G N G F R Q I I V E K T D F Q F A	
TTAGGAATGCGAACCGCTTGCGAGCGCCCCGTAGTAAAGACTTTGGCTACTTTCATGATATGCCTCCTTTTCATTGTGATATATCAATGCCGTATATCAT AATCCTTACGCTTGGCGAACGCTCGCGGGGCATCATTTCTGAAACCGATGAAAGTACTATACGG <u>AGGA</u> AAAGTAACACTATATAGTTACGGCATATAGTA	1200
KPIRVAQSRGTTFVKAVKM SD	
<b>тары</b> –10	
KpnI	
ACTATCGAATTTTAGCAAGGTAGTTTTTTATTCGAAAAAGTCATCGAAAACGGATAGCGCCTCATTGGCGTTTAGAGCCGTTATTTTATTTGGGGGGTACC TGATAGCT <u>TAAAAT</u> CGTTCCATCAAAAAAT <u>AAGCTT</u> TTTCAGTAGCTTTTGCCTATCGCGGAGTAACCGCAAATCTCGGCAATAAAATAAACCCCCCATGG	1300
-10 -35 -35	

	**		.08	108				n1)	nn.															· 4 .		• • •											****														÷ 7	
TTGGT	AA	TGG	CI	ATT	GG	TA	GC	TA	TT	AC	GC	GC	JI	CG	SCO	G	ſAC	GAC	СТ	:AC	T	CG	GCG	5A.	AA:	TA	GI	rTJ	TZ S	AT!	TT	CC	ΤT	AA	Υ	AT'	rT(	CT.	AC'	тC	GG	ТА	ΔT'	TAł	A.P	GT	ΤA	AT2	YC.	гт		
TTTGAC	cc	AGA	A	TTA	A	AA	AA	ТА	cco	GC	GA	T	GA	GC	:0	GC	CGC	SC	AT	T	G	A	AG/	١T	ТА	AG	TC	CAT	G/	١G	CA'	ΓG	AG	CG	GG	GA	AT.	AT	AT	тт	тт	сс	G	СТЈ	ſG	тт	AT	[A	CGC	SA	15	500
AAACTG	GG	TCI	TC	CAA	TC	TT	тт	AT	GG	CG	ст	'A(	CI	CG	GG	c	GCC	G'	TA	AAC	c	T	ICI	ſA.	AT:	TC	A	GT	C.	CC)	GT	AC	тс	GC	cc	CT	[A	TA	TA	AA	AA	GG	SC	GA/	AC.	AA	TA	١T	GCC	ст		
							_				7				7	,						_																								*		v	5	5		
																																N	si	I																		
TCCATC	AG	GGG	GAG	AAG	TT:	TT	АА	TG	GG	ті	т	CAC	CJ	CT J	ΓA'	T	CG	GC.	AC	CAZ	AT	G	GGG	GA	TT	GA	c	GC	SCO	GA	CA	АТ	GC	'A'	ГА	TC	СТ	GC	CT	GĮ	AGC	cc	CG	TA	AC	CA	AA	٩A	AG2	AA	10	600
GAGGTAG	тс	ccd	СТС	TTC	CAI	AA	тт	AC	:cc	AA	AA	GT(	Gł	AA.	٩T.	A	GC	CG	TG	GT.	ΓA	C	cco	СТ	AA	CT	G	CG	G	ст	GT	ТА	CG	T7	АТ	AG	GA	CG	GA	CJ	CG	GG	GC.	AT?	ΤG	GI	тт	гт	rc:	гт		
EM	т.	P		5 7	r	к	I		Р	K	(	v		к		D	;	A	c	2	н	[	P		N	v	,	R		5	L		A	3	Y	G		A	Q		A	F	R	L		W	F	-	s			
	-	-			•				-	-	•	•				-		-					-																-													
CCGAAT	เดา	ጥጥ	360	GA	٨G	ст	тт	'nт	ст	AC	C.	гт	c	ACC	CG	c	GA	ст	тс	CAC	GG	A	сті	AA	тт	GC	:G	CG'	rT'	ГG	GC	GT	тт	'AC	CA	AT	тт	СА	GA	Тł	AAA	G	CA	GA	AC	GC	GG	GC.	AT	АТ	1	700
CCCTTA	сı	220	200	ידי	rco	- G D	22	20	GA	тс	G	2 24	G	rGC	30	G	СТ	34	<u>а</u> с	370	cc		GA	тт	 AA	CG	C	GC	10	AC	CG	CA		т	GT	יידא	<b>A</b> A	GT	СТ	A	гтт	C	GТ	CT	TG	ce	SCC	CG	TA	ТА		
N C H	i Cr	ĸ	D	F		50m	ĸ	F	2	τ.		~		7	<u>د</u>		v		E	1	Þ		s	T	••••	Δ	1	R	ĸ		Α	N	1	v		т	E		s	1		A		s	R	2	P		M	D		
	•	I.	-	-		•		-		-	•	•		•	n	•	٠		-		•		•	-		••		•	••		••	•		•		-			-		-			-		•	-			-		
COACCTO	որո		sc	יידיטי	рт,	200	тÞ	דמ	rGT	ጥ	145	SC	C	CC	<b>4</b> A	\ <b>T</b>	GG	тт	C	CA	<b>۵</b> Δ	<u>۳</u>	AC	ልጥ	G۵	СТ	T	TC	٩G	CA	\TT	۸A	AC	c	GC	GG	тс	CG	GТ	G	SAT	T	гт	TC	АТ	T	rT7	IG	cc	АТ	1	800
CAGCII		1000	00	201	' '	300	ייב דע:	גניי. מידיי		20	2010	20		30'	ייי. דיד	יב. קי	cc	 A A		GT	ייי. דיד	יבי	 ТС'	та	CT	GA	A	AG	ГC	GT		тт	т	GG	cr	:cc	AC	GC	CA	CC.	CTZ	A	AA	AG	TA		AA,	AC	GG'	ТА	_	
	····	,000	2.G.		nn F	лос п		т. т	NJ.		- 1 ( T	20	с. С	3G. 1	 T		сс. п	nn P		т. Т		. n	10	м м	v	, 01	v		r.	ч м		т.		2	E	,	н. н	- - -		D	N	 J	ĸ		м		 (	Δ	1	м		
цк	4		n.	1	Б	-		+	N		'n		9		Ľ		F	Ľ	)	ц		T			۷		r	• •		1.1								-		•	•	•			••	•	•			2		
																																																•				
ACTCTT	T	CTC	CT GA	IAG ATC	TG AC	GT <i>A</i> Cat	GI CI	CC AGC	CAC	A:	TA/ AT:	AT TA	C.	AA( TT(	CA G <u>T</u>	\А ГТ	TG	TA AT	GC	GC. CG	A] T <i>i</i>	TT AA	GC	CA GI	TC.	L CTI GAA	Dr IT	<b>aI</b> AA	AC IG	TC AG	CAA STT	A.	AAT ATI	rgi	AC	GCG	GC CG	CA	AT TA	T	SCO	i GC GC	<b>H1</b> TT AA	GA GA	II CT GA	с ГТТ \А/	rc <i>i</i> Ag:	AA TT	GC' CG	TC AG	1	900
AACTCTI ITGAGA <u>/</u>	T LA	CTC GAG	CT GA	TAG ATC	TG AC H	GTA CAI Y	GI CI	CC AGC	CAC GTG V	A: TI	TA/ AT' Y	AT TA D	C. G	AA( TT( V	са G <u>т</u>	АА <u>ГТ</u> І	TG AC	ТА <u>ат</u> Ү	GG	GC. CG A	TA TA M	TT AA N	GC CG G	CA GI	TC AG D	L TT SA <i>P</i> F	Dr TT AA	<b>aI</b> AA TT F	AC IG	тс <u>ас</u> Е 35	CAA GTT F	A.	I I	rgi AC'	AC TC V	GCG GCG R	GC	CA GI W	AT TA N	T A I	SCC CGC G	) GC GC I	H1 TT AA N	GA GA CT V	CT GA	r TT AA/ K	rc <i>i</i> AG: 1	AA TT L	GC CG S	TC AG	1	900
AACTCTI TTGAGA <u>/</u>	T AA	STC SAG	CT <u>GA</u> D	TAG ATC.	TG AC H	GTA CAI Y	GT CZ I	CC AGC	CAC GTG V	A: T	TA AT: Y	AT TA D	C. G'	AA TT V	са G <u>т</u>	АА <u>ТТ</u> І	TG <u>AC</u>	ТА <u>ат</u> Ү	IG( C( )	GC. CG A	LA T I I	ft Aa N	GC CG G	CA	TC AG D	L TT SA <i>P</i> F	Dr TT AA	AA TT F	AC IG	TC <u>AG</u> 35	CAA GTT F	A/ TT	I I	rGi AC'	AC TC V	CGC GCG R	GC	CA GI W	AT TA N	T( AA)	SCC CGC G	) GC GC 1	H1 TT AA N	GA GA CT V	CT GA	r TT AA K	rc <i>i</i> AG: J	AA TT L	GC' CG S	TC AG	1	900
AACTCTT TTGAGA <u>1</u> CAATGC(	TT( AA(	STC SAG	ст <u>GA</u> D	IAG ATC. *	TG AC H	GTA CAI Y	GI C <i>I</i> E	CC AGC	CAC GTG V	EA:	TA AT Y GG	AT TA D	C)	AA TT V	СА G <u>Т</u> Т1		TG AC	TA <u>AT</u> Y		GC. CG A CT	AT T# N T(	TT AA N	GC CG G	CA GI	TC AG D	I SAA F	Dr TT AA K	AA TT F	AC IG -	тс <u>А</u> С 25 35 да	CAA <u>STT</u> F 5	AI TT	AAD IT I	rgi AC'	AC TC V	GCG R R	GC	GI W	AT TA N	T AA I	GCC GC G		H1 TT AA N	GA GA CT V		r TT K K		AA TT L	GC CG S AC	TC AG GC	: 1	900
AACTCT] ITGAGA <u>/</u> CAATGCC	TT( <u>AA(</u> GT)	CTC GAG SI	CT <u>GA</u> D TA	TAG ATC.	TG AC H TG	GTA CAI Y CCC	GT		CAC GTG V AAC		TAX AT: Y GG <sup>r</sup>	AT TA D TG	C.	AA( TT( V AA'	CA G <u>T</u> TI		TG AC	TA <u>AT</u> Y CG		GC. CG A CT	TA T N TC	TT AA N CA	GC G G	CA GI GI		L CTT SAP Y CAP	Dr TT AA K	AA TT F	AC IG TT	TC <u>AG</u> 35 GA	CAA <u>STT</u> F 5 AAT	AF TT	I I I ST:	rgi AC'	AC TC V		GC CG	GI W	AT TA N		GCC CGC G AT:		HI TT AA N	GA GA CT V		r TTT K K	IC/ AGI J CA/	AA TT L AC	GC CG S AC	TC AG GC	: 1	900 000
AACTCTI TTGAGA <u>/</u> CAATGCC GTTACGC	GT CA.	STC SAG SI SI TTT	CT GA D TA AT	TAG ATC. *	TG AC H TG AC	GTA CAT Y CCCC GGC	GT CH I I I I I I I I I I I I I I I I I I		CAC GTG V AAC		TA AT Y GG CC.	AT TA D TG AC	C, G G G	AA( TT( V AA) TT.	CA G <u>T</u> TI AA		TG AC 10	TA <u>AT</u> Y CG GC		GC. CG A CT GA	AT TA N T( A(	TT AA N CA	GC G G AT	CA GI GI	ATC AG D	L CTI GAP Y CAP	Dr TT AA K	AA AA TT F TC	AC IG TT AA	TC <u>AG</u> 35 GA	CAA GTT F 5 AAT ITA	AF TT	AAT I I GT: CAA	rgi AC' TT' AA		CGC GCG R CGC GCG	GC CG AA	CA GI W	AT TA N ATC	T AA I CC.	GCC CGC G AT: TAI		H1 TT AA N	GA GA V TA		r TT AA/ K AG(	ICA AGT J	AA TT L AC TG	GC CG S AC TG	TC AG GC CG	: 2	900 000
AACTCTT TTGAGA <u>/</u> CAATGCC STTACGC W H 2	GT GT CA.	STC <u>GAG</u> SI TTT AAA K	CT <u>GA</u> D TA AT L	TAG ATC. * GTT CAA K	TG AC H TG	GT# CAT Y CCCC GGC G	GT CZ I TT GAZ		CAC GTG V AAC ITG L	CO GG R	TA AT' Y GG' CC.	AT TA D TG AC H	G G G G	AA( TT( V AA' TT. L	CA G <u>T</u> TI AA E		TG AC 10 AG	TA AT Y CG GC		GC. CG A CT GA	AT TA N TC AC E	TT AA N CA GT	GC G G AT TA	CA GI CF I	ATC AG D	L CTT SAP F CAP STT L	Dr TT AA K	AA TT TC AG E	AC IG TT AA	TC AG 35 GA	CAA F 5 AAT I TA	A/ TT	AD TTA I GT: CAA	IC AC IT		CGC GCG R CGC GCG A	GC CG AA TI	GI W TA	AT TA N ATC AG D	T AA I CC. GG	GCC CGC G AT: TAI	i GC I I G A C A	H1 TT AA N	GA GA CT V STA STA	CT GA GI I	r TTT K K AGC	ICA AGI J CAA GTI L	AA TT L AC TG	GC CG S AC TG V	TC AG GC CG R	: 2	900
AACTCTT TTGAGA <u>A</u> CAATGCO GTTACGO W H :	GT' CA	SI SI TTT AAA K	CT <u>GA</u> D TA AT L	TAG ATC * GTT CAA K	IG AC H TG	GT# CAJ Y CCCC GGC G	GT CZ I T SAZ K		CAC GTG V AAC ITG L <i>Ec</i>	:A: :T/ :C: :G: :G: :R	TA AT Y GG CC.	AT TA D TG AC		AA TT V AA TT. L	CA G <u>T</u> TI AA E		TG AC 10 AG TC	TA AT Y CC		GC. CG A CT GA	AI TZ N TC AC E	TT AA N CA GT	GC G AT TA I	CA GI CA I	ATC AG D CCC AGG	L CTI SAP Y CAP STI L	Dr TT AA K	AA TT TC AAG E	AC IG TT AA	TC AG 35 GA	CAA GTT F 5 AAT I I I	A/ TT	AAT IT GT CAJ	IGI AC TT AA		CGC R CGC GCG A	GC CG AA TT	CA GI W	AT TA N ATC AG D		GCC CGC G AT: TAJ	i GC I I I G A C A A	HI TT AA N CC	GA GA V TA TA AT		I TTT AA K AG I C L	IC/ AGI J CA/ GTI L I	AA IT L AC	GC CG S AC TG V	TC AG GC CG R	: 2	900
AACTCTT TTGAGA <u>A</u> CAATGCC GTTACGC W H GCAATTT		SI SI TTT AAA K	CT <u>GA</u> D TA AT L	IAG ATC * : GTT CAA K ATG	IG AC H TG AC	GTA CAT Y CCCC GGC G	GT CH SAH K		CAC GTG V AAC ITG L <i>Ec</i> GA <i>P</i>	CO GG R	TA AT Y GG CC. <b>RI</b> TC	AT TA D TG AC H		AA TT V AA TT L GA	CA G <u>T</u> TI AA E		TG AC 10 AG TC AG	TA AT Y CC GC		GC. CG A CT GA	AT TA TC AC E	TT	GC G AT I I	CA GI CA I	ATC AG D AGG	L CAP SAP CAP STJ L	AA AT	AA TT F TC AAG E	AC IG TT AA Q TT	TC AG 35 GA CT	CAA GTT F 5 AAT I I AAT		AD TT I GT CAN T	rgi AC' TT AA E		CGC R CGC SCG A		CA GI W ATA ATA	AT TA N ATC AG D		GCC GGC AT: TAI M	i GC I I I G A C A C A C I T T	H1 TT AA N CG GC	GA GA CT V TA TA Y			ICA AGT J CA GT L I GA	AA TT L AC TG	GC CG S AC TG V	TC AG GC CG R	: 2	900
AACTCTT TTGAGA <u>A</u> CAATGCC GTTACGC W H C GCAATTT CGTTAA		SI SI SI ITT AAA K	CT GA D TA AT L	TAG ATC. * GTT CAA K ATG	TG AC H TG AC AT	GTA CAJ Y CCCC GGC G GCC	GT CH CH CH CH CH CH CH CH CH CH CH CH CH		CAC GTG V AAC ITG L GAP CTT	CO GGO R CO GGO R	TAA AT Y GG CC. RI TC	AT TA D TG AC H		AA( TT( V AA) TT. L GA GA			TG AC 10 AG TC AG	TA AT Y CC GC		GC. CG A CT GA		TT AA N CA GT	GC G G AT I I CA	CA GI CA I L AA		L TT SAP F CAP ST L		AA TT F TC AAG E	AC IG TT AA Q TT	TC AG 35 GA CI	CAA <u>STT</u> F 5 AAT I I AAT		I ST: CA/ CA/ CT: SA/			CGC R CGC A IGC		CA W XTA XAJ	AT TA N ATC AG D		GCC GGC AT: TAI M AT: TAI	I GCJ I GCJ A C A C A C A C A A A A	H1 TT AA N CG GC	GA GA V TA TA Y		I TTT K K AGO ICO L I. AGO	ICA AGI J CA GTI L GA GA CT	AA IT L AC TG	GC CG S AC TG V	TC AG GC CG R	: 2	900 000
AACTCTT TTGAGA <u>A</u> CAATGCC GTTACGC W H C GCAATTT CGTTAAA I. K	GT CA	CTCC SI SI TTT AAAA K ICCC AGG E	CT GA D TA AT L GC CG	IAG ATC. * : CAA K ATG TAC	TG AC H TG AC AT	GTA CAT Y CCC GGC G GC CCC	GT CZ TT SAZ K		CAC GTG V AAC ITG L GAP CTT	:A: :C: :C: :G: :G: :C: :G: :C: :G: :C: :C	TAX AT' Y GG <sup>C</sup> CC. RI TC AG	AT D TG AC H CI		AA TT AA TT L GA GA	CA G <u>T</u> TI AA E GI GI		TG AC 10 AG TC AG	TA AT Y CC GC		GC. CG A CT GA		TT AA ST TT AA	GC G AT TA I CA	CA GI GI CA I TT	ATC AG D CCC AGG D	L CAF GTT L AAC		AA TT F TC AAG E	AC IG TT AA Q TT AA	TC AG 35 GA CI	CAA STT F 5 AAT I I AAT I I		AAT I I CAA CT GAA			CGC FCG R CGC GCG A IGC ACC H	GC CG AA TT I GC A	CA GI W ATA AI CAA STI	AT TA N ATC AG D		GCC GGC AT: TAI M AT: TAI	IGCI IGCI IGCI ACCI ACCI ACCI ACCI ACCI		GA GA V V STA X TA Y		TTI AA/ K AG( TC( AG(	CAI GAI GAI CAI GAI CTA	AA TT L AC TG AC	GC CG S AC TG V	TC AG GC CG R	: 2	900 000
AACTCTT TTGAGA <u>A</u> CAATGCC GTTACGC W H C GCAATT CGTTAAA L K	GT CA	SI SI ITT AAA K ICCC AGG E	CT GA D TA AT L GC CG A	TAG ATC * GTT CAA K ATG TAC H	TG AC H TG AC AT TA	GTA CAJ Y CCCC GGC G GCC CGC J	GT CH K GG CH	TTA AAT ITTC AAC Q	CAC GTG V AAAC ITG GAP CTT	:A: :TI :C: :G: :G: :G: :C: :C: :C: :C: :C: :C:	TA AT Y GG CC. RI TC AG G	AT D TG AC H CI		AA TT V AA TT. L GA GA	CA G <u>T</u> TI AA E GI GI T	AA I - IC AG E IG AC	TG AC 10 AG TC AG	TA AT Y CC GC		GC. CG A CT GA GA TA D	AT TA TC AC E	TT AA ST TT AA E	GC G AT I TA I CA G G	GI GI C/ I TI F	ATC AG D CCC AGG D	L CAP F	Dr TT AA K AT TA CC GG R	AA TT F TC AG E	AC IG TT AA Q TT AA K	TC AG E 35 GA CT	CAA <u>STT</u> F 5 AAT I I AAT		AD TT I ST CA CT SA J			CGC R CGC GCG A TGC H	GC CG AA TT J GC A	GI W XT/ XAJ CA/ STJ	AT TA N ATC AG D		GCC GC G AT' TAI M AT' TAI	IGCI IGCI ACCI ACCI ACCI ACCI ACCI ACCI	H1 AA N CC GC AF	ATC		TTT AAA K AGO FCO L ICO AGO	ICA AGT I CA GA GA CT. S	AA TT L AC TG AC	GC CG S AC TG V	TC AG CG R .CC GG	: 2	900 000
AACTCTT TTGAGA <u>A</u> CAATGCC GTTACGC W H C GCAATT CGTTAAL L K	GT CA. F	SI SI ITTT AAAA K ICCC AGGG E	CT GA D TA AT L GC A	TAG ATC. * GTT CAA K ATG TAC H	IG AC H TG AC AT A TA	GTA CAJ Y CCCC GGC G CCCC I TCC	GT GT GG GG GG CI		CAC GTG V AAC ITG GAP CTJ J	:A: :T) :C) :GO :GO :GO :GO :CO :CO :CO :CO :CO :CO :CO :C	TAA AT' Y GG' CC. RI TC AG G	AT D TG AC H CI GA		AAC V AAC TT. L GA CT	CA G <u>T</u> TI AA E GI CA T		TG AC 10 AG TC AG TC AG	TA AT Y CG GC		GC. CG A CT GA TA D	AT TA TC AC E	TT AA ST TT AA E	GC G AT TA I CA GT	GI GI CZ I TI F	ATC AG D CCC AGG D TTT F	L CAF SAP K CAF STJ L AAC		AA TT F TC AAG E CGT GCA	AC IG TT AA Q TT AA K	TC AG 35 GA CI	CAA <u>STT</u> F 5 AAT I AAT TTA L AG					CGC R CGC A IGC A CGC H			ATC NATC ATC ATC		GCC GGC AT: TAI M AT: I			ATC			ICA AGI J CAA GTI L GAI CT. S	AA IT L AC TG AC I	GC CG S AC TG V	TC AG CG R CG GG GG		900 000 100
AACTCTT TTGAGA <u>A</u> CAATGCC GTTACGC W H : GCAATT: CGTTAAL L K TTGAAAA		CTCC SAGI SI ITTT AAAA K ICCC AGGG E AAAG		TAG ATC * GTT CAA K ATG TAC H GGT	TG AC H TG AC AT AT	GTA Y CCCC GGC GGC CCCC I TCC	GT T GG GG C T A		CAC GTG V AAC ITG GA GA GC GC GC	COG R C COG R C COG C C C C C C C C C C C C C C C C C	TAA AT' Y GGG' CC. AG AG G AA			AAA V AAA TT. L GA			TG AC 10 AG TC AG TC AG	TA AT Y CC GC		GC. CG A CT GA TA D		TTAA ST TTAA GA	GCG G AT I CA G G CA CA	GI GI CA I TI F		L CAF SAF SAF SAF STT L SAF SCF		AA TT F TC AG E	AC IG TT AA Q TT AA K	TC AC 35 GA CT CA	CAA STT F AAT I AAT L AAG					CGC CGC CGC CGC CGC CGC CGC CGC CGC CGC			AT TA N ATC AG D		GCC G G AT: TA/ M AT: I AA(	IGCI IGCI IGCI ACCA A TT. AAA K		GAG GAG CT V GTA CT Y ATC TAG			CAA GT GA GA GA CT. S AA	AA TT L AC TG AC I	GC CG S AC TG V TA		1 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	900 000 100
AACTCTT TTGAGAA GTTACGC W H : GCAATT: CGTTAAL L K TTGAAAA AACTTT		CTCC SAG SI TTT AAA K ICCC AGG E AAGG E	CT GA D TA AT L GC A GC A GC	IAG ATC. * GTT CAA K ATG TAC H GGT CCA	TG AC H TG AC AT AT AT	GTA Y CCCC GGC GGC CGC I TGC ACA	GG CI SAI K GG CI SAI		CAC GTG V AAAC ITG GAP CTI J GCC CGC	COGO R COGO R COMTO TA	TAX AT Y GG CC. TC AG G AA TT			AA V AA TT L GA CT			TG AC 10 AG TC AG TC AG	TA AT Y CG GC		GC. CG A CT GA TT CAA		CA ST TT AA GA CT	GC G G AT TA I CA GT	GI GI CF I A TI F	ATC AG D CCC AGG D AAA F TGC ACC	L CAA GTT L AAC F CCA GG	AT CC GC R AC	AA F TC AAG E GT GCA CAA	AC IG TT AA Q TT AA K GI	TC AG S S G A C I C A I C A I	CAA <u>STT</u> F 5 AAT I AAT I AAT C C C C		AAT I I GT: CA/ CG( GC)			CGC GCG A CGC A CGC A CGC A CGC A CGC A CGC A CGC A CGC A CGC C CGC C CGC C CGC C CGC C CGC C C C C C C C C C C C C C C C C C C C		CA GI W TA CA STI I TT	ATC NATC ATC ATC ATC		GCC G G AT: TAJ M AT: TAJ			GA GA CT V TA CT Y TA TA TA TA TA TA TA		K AGO L ICO AGO SGG. CC	ICA GAT I GAT CA I CA I CA I CA I CA I CA I CA I CA	AA TT AC TG AC T	GC CG S AC TG V TA TG	TC AG CG CG CG GG GG GG GG	1 2 2 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3	900 000 100
AACTCTT TTGAGA <u>A</u> CAATGCC GTTACGC W H C GCAATT CGTTAAL L K TTGAAAA AACTTT Q F	GT GT GT GT GT GT GT GT GT GT GT GT GT G	CTCC SI SI TTT AAAA K TCCC AAGG E AAGG TTCC P	CT GA D TA TA L GC CG A GC CG	IAG ATC. * GTT CAA K ATG TAC H GGT CCA P	IG AC H TG AC AT AT AT I	GTA CCC GC GC GC CCC GC TC TC TC TC	GI CA TI SAA K GG CA TA I		CAC GTG V AAC ITG CAA CTI GCC CGC R	COGO R	TAX AT GG CC. AG G AA TT L	AT D TG AC H CI GA CC GC		AA V AA TT L GA CT GA V	CA GI TI AA E GI CZ T CZ		TG AC 10 AG TC AG TC AG	TA AT Y CC GC GC TA TA TA		GC. CG A CT GA CT A TT L	AT TA AC E C G AC	TTAA TTAA GA CT S	GCG G AT TA I CA GT CA GT CA I GT L	CA GI GI CZ I I AI TI F	ATC AG D CCC AGG D AAA F TTT F GCC ACC A	L SAP K SAP T SAP T SAP C SAP S S S S S S S S S S S S S S S S S S		AA F TC AG E GT GCA CAA STT I	AC IG TT AA Q TT AA K GI	TC AG 35 GA CI CZ GI I CZ AJ	CAA <u>STT</u> F 5 AAT I AAT I AAT L AGC TCC H		I ST: CA/ ST: CA/ SA/ SC( SC) P			CGC R CGC GCG A IGC A CGC A CGC A CGC A CGC	GC CG A TT I GC CC A		ATA NATC ATC D AAA TTT		GCC GGC AT: TAI M AT: TAI J AA( TT)	IGC IGC ACC ACC ACC ACC ACC ACC ACC ACC ACC A		ATC			ICA AGT I CAA GAT CTA GAT CTA S AA TT	AA IT AC TG AC I	GC CG S AC TG V TA AT		1 2 2 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3	900 000 100
AACTCTT TTGAGA <u>A</u> CAATGCO GTTACGO W H : GCAATT: CGTTAAL L K TTGAAAA AACTTT Q F	GT CA. T TAA GA CT F	CTCC SI ITTT AAAA K ICCC AGG E AAGG E AAGG TTCC P	CT GA D TA AT L GC GC A GC GC CG	IAG ATC. * GTT CAA K ATG TAC H GGT CCA P	IG AC H TG AC AT AT AT AT	GTA Y CCC GGC G CCC CGC I TGC TCC T	GT CZ TT SAN K SG CZ Z CZ Z TA Z I		CAC GTG V AAAC ITG GAP CTI J GCC CGC R	COGO R CONT	TAX AT' Y GGG' CC. TC AG G AA TT L	AT D TG AC H CI GA CC GC		AAO TTO AAO TT. L GA CT GA CT V			TG AC 10 AG TC AG TC AG	TA AT Y GC GC AT TT TT N		GC. CG A CT GA TA D XAT L		TTAA CA GA GA CT S	GCG G AT TA I CA GT CA GT CA GT L	GI GI GI A TI F	ATC AG D CCC AGG D TTT F GC ACC A	L CAF GTT L AAC F CCF		AA TT F TC AG E CGT GCA CAA STT I	AC IG TT AA Q TT AA K GI	TC AG 35 GA CI CA I	CAA STT F 5 AAT I AAT I L AAG TCC F			IT AA E IT AA K GC CG		CGC R CGC GCG A CGC A CGC A CGC A CGC A CGC A CGC C C C			ATC ATC AG D AAA TTT		GCC G G AT: TAI M AT: TAI J AA( TT)			GA GA V TA TA TA TAG		K AGO L ICO AGO SGG. CC	CAAG GAT GAT CAA GAT CTA S CTA S CTA S	AA IT AC TG AC I AA	GC CG S TG V TA ATG			900 000 100
AACTCTT TTGAGA <u>A</u> CAATGCC GTTACGC W H : GCAATT: CGTTAAL L K TTGAAAA AACTTT Q F	GT CA. F	CTCC SAG SI TTT AAAA K ICCC AAGG E AAGG TTC P	CT GA D TA L GC CG A GC CG	TAG ATC. * GTT CAA K ATG TAC H GGT CCA	IG AC H TG AC AT TA AT I	GTA Y CCC GGC G CCC G CCC TG? ACZ T	GT CZ TT SAZ K GC SAZ FAT		CAC GTG V AAC ITG CAC GA CTI GCC CGC R	COGO R COGO R CO TA	TAA AT' Y GGG CC. AG AA TC AG AA TT L	AT D TG AC H CI GA		AAC TTC V AAC TTC L CA GA CT V	CA GI TIAA E GI CA T CA CA		TG AC TC AG TC AG TC AG	TA AT Y GC GC TT TT AZ		GC. CG A CT GA TA D AA TI L	AT TA TO AC E	TTAA ST TTAA GA CT S	GCG G AT I CA G CA G G G G G G G G G G G G G C G C	GI GI CZ TI F		L SAP F CAP TTC TTC F CCP		AA TT F TC AG E CGT GCA STI I	AC IG TT AA C TT AA K GI	TC AG 35 GA GA GA I L	CAA STT F 5 AAT I I AAT I I AAT I I AAT					CGC GCG GCG A IGC A CGC GCG A IGC A CGC GCG GCG GCG GCG GCG GCG GCG GCG		GA W TA CA STI	ATC N ATC AG D AAA TTT		GCC GGC AT: TAJ M AT: TAJ AA TT					K AGC L L AGC L AGC C C C C	CAA GT L GA CTA S AA TT	AA TT AC TG AC T	GC CG S AC TG V TA			900 000 100
AACTCTT TTGAGA <u>A</u> CAATGCO GTTACGO W H C GCAATT CGTTAAL L K TTGAAAA AACTTT Q F	GT GT GT GT GT GT GT GT GT GT GT GT GT G	SI SI SI TTT AAAA K ICCC AAGG E AAGG TTC P		TAG ATC. * GTT CAA K ATG TAC H GGT CCA P	IG AC H TG AC AT H AT A I	GTA Y CCC GGC GGC CGC I TG TG T	GT CZ TT GAZ K GG CZ P TA I		CAC GTG V AAC ITG L CTI GCC CGC R		TAA AT' Y GGG' CC. AG AA TT L			AAC TTC AAC TT L CAC GA CT V	CA G <u>T</u> TI A <i>I</i> G G C <i>I</i> C G C		TG AC AG TC AG TC AG C C	TA AT Y GC GC AT TT AZ		GC. CG A CT GA TA D A A TI L		TTAA TTAA TTAA GA CT	GCG G ATTA I CCA AGT CA GT I GT I I	GI GI GI A TI F	ATC AG D AG AG TT F F G C ACC A D F	I SAA K STJ L AAC F CCI GG?	AT CC GC R AC TC	AA TT F TC AG E CGT GCA CAA STT I SA	AC IG TT AA Q TT AA K GI		CAA STT F 5 AAT I AAT I TTA L AGC TCC					CGC GCG A CGC A TGC A CGC A CGC A CGC A CGC A CGC A CGC A CGC C C C C C C C C C C C C C C C C C C C			AT TA N ATC AG D AAA TTT		GCC GGC AT: TAJ M AT: TAJ	IGCACIA A ITT.					CAAG GT GT GA GT S CT S AA TT	AA TT AC TG AC T T AC	GC CG S AC TG V TA AT AT			900 000 100

FIG. 1. Sequence analysis of the 2,262-bp *D. nodosus* DNA insert of pJIR318. Both the nucleotide sequence and the deduced amino acid sequences of ORFs from both strands are shown. Stop codons are indicated by asterisks. Potential ribosome-binding sites (33) are indicated by SD (Shine-Dalgarno), and possible -10 and -35 promoter sequences (15) are underlined. Inverted repeats which could function as transcriptional terminators are indicated with arrows below the nucleotide sequence. The inverted repeat at position 468 could act as a bidirectional terminator. Several direct repeats of unknown significance are indicated by arrows above the nucleotide sequence.

pJIR318-encoded VapD (Fig. 3). We therefore conclude that the aberrant mobility of VapD was due to the amino acid sequence of the VapD protein, in particular, sequences at the C-terminal end of the protein, and that removal of these amino acids can lead to an increase in the apparent molecular weight of the gene product.

Western immunoblots of VapD in D. nodosus and N. gonorrhoeae. The SDS-PAGE-purified, urea-extracted VapD protein was used to raise specific rabbit antiserum. The

resultant antiserum reacted with a 6-kDa polypeptide in Western immunoblots of the virulent strain *D. nodosus* A198 but not the benign strain *D. nodosus* 305 (Fig. 4). To extend these studies, cell extracts of 16 additional *D. nodosus* isolates, which were classified as virulent, intermediate, or benign, were separated by SDS-PAGE, transferred to nitrocellulose, and reacted with anti-VapD antiserum in Western immunoblots. The results showed that the VapD protein was produced by all of the *D. nodosus* isolates which hybridized

# A

vapD	MYAIAFDLEIAELKKHYGEPYNGAYLEIGKELKAIGFEWTOGSVYLSN
ORF5	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
vapD	DNNLATVYRAISTISKIDWFKSSVRDIRAFKVEDWSDFTAIVKGA

ORF5 NEDMANLFSAINELKALPWFPSSVRDIRAFRIEQWSDFTSLVKS



MKVAKVFTTGRSQAVRIPKAFQFDTKEVIIQRFGNGILLIPK
METT-VFLSNKSQAVKLPKAVALPENVKKVEVIAVGKTRIITPA
NSDWQGFLEALNEFEPDFKIEREPQIEQVREDFL
GETWDEWFDG-NSVSADFMDNREQPGMQERESF
MSIKLMLDTNICIYIINTKPPYIAEKFSRYEIGDIGISNITACE
:
LAFGVEKSG-SAKNKTALTKFLAPLSILPFDKQAIWHYARIRQS
LQNRGTPIGALDMLIASHALALDITLVTNNTKEFERVDGLMLDN
LALQGRPVGPFDQMIAGHARSRGLIIVTNNTREFERVGGLRIED
WAID
: . WS

FIG. 2. Comparative amino acid sequence analysis of vap gene products. (A) Comparison of the *D. nodosus vapD* gene with ORF5 (19) from the cryptic *N. gonorrhoeae* plasmid. The amino acid sequences of the two ORFs were aligned by using the FASTP program (23). Identical amino acids are marked with colons, and functionally similar amino acids are marked with periods. (B) Region of the *E. coli* F plasmid that had sequence similarity with the *vapB-vapC* region of pJIR318. A comparison of two small ORFs on the complementary strand of *trbH* with *vapB* and *vapC* is shown below a map of the *E. coli traD-traI* region (16). The comparison was carried out by using the CLUSTAL program for sequence alignment (14).

with the pJIR318 probe. Isolates which did not hybridize with pJIR318 (strains AC554, AC6, and AC176) did not produce detectable levels of VapD. The virulent isolates (strains A198, B1006, and C1008) and one intermediate isolate (strain HA343) appeared to produce more VapD than the other strains (Fig. 4).

Further Western blots were carried out with cell extracts

of *N. gonorrhoeae* strains MS-11A, Gc var40, and JKD109. No reactivity in the 6- to 10-kDa region, that is, at a molecular weight consistent with the recognition of ORF5 (10.5 kDa) expressed from the gonococcal cryptic plasmid, was detected in immunoblots when anti-VapD antiserum was used (data not shown). Antiserum raised against whole piliated *N. gonorrhoeae* also failed to react with VapD expressed in *E. coli*.

### DISCUSSION

Sequence analysis of pJIR318, one of three plasmids previously shown to carry gene regions associated with virulence in *D. nodosus* (17), revealed that this plasmid contains four genes, two of which, vapA and vapD, appear to be present as separate cistrons. However, the vapC start codon overlaps the stop codon of vapB. This arrangement, together with the absence of a transcription initiation site upstream of vapC, suggests that vapB and vapC may form an operon in which both transcription and translation are coupled. One gene, vapD, has been shown to be expressed in both *E. coli* and *D. nodosus*. Amino acid sequence analysis of the gel-purified VapD protein has confirmed the predicted vapD start codon.

In a previous study, 33% of the benign D. nodosus isolates that were examined hybridized with pJIR318 (17). It was not known whether these strains expressed the products of the vap region or were benign, because they carried homologous but nonfunctional vap genes. The four pJIR318-hybridizing benign isolates that have now been tested in Western immunoblots all expressed VapD. All of the hybridizing virulent and intermediate strains of D. nodosus that were tested also produced VapD. However, it is not possible to determine whether the VapD polypeptides produced by these isolates are functional.

The vapD gene product and the vapBC region were shown to have sequence similarity with components of the N. gonorrhoeae cryptic plasmid and the E. coli F plasmid, respectively. The significance of this sequence similarity is not clear. The function of the cryptic plasmid is unknown; however, it is present in 99% of clinical isolates of N. gonorrhoeae (29). The discovery of homology between ORF5 of the cryptic plasmid and a region associated with virulence in D. nodosus suggests that this plasmid may have a role in the virulence of N. gonorrhoeae. It was hoped that the VapD antiserum would provide a means by which the function of ORF5 could be investigated further. Unfortunately, anti-VapD antiserum failed to react with the ORF5 protein in N. gonorrhoeae, suggesting either that ORF5 was not expressed or that it did not cross-react with VapD antibodies.

The function of the trbH region of the *E. coli* F plasmid is also unknown. The region is unlikely to be essential for either plasmid maintenance or DNA transfer, as the trbHregion is absent in the F-like plasmid R100 (40). The apparent conservation of the genetic organization of the *vapBC* operon and the operon encoding TRAORF1 and TRAORF2 suggests that the functional significance of the *trbH* region may reside on the strand opposite to that containing the *trbH* gene.

The presence in pJIR318 of regions homologous to two different bacterial plasmids seems to suggest that the *D. nodosus* DNA insert of pJIR318 is located on a *D. nodosus* plasmid. No plasmids have been identified in *D. nodosus*. In addition, Southern blots of *D. nodosus* DNA fractionated by pulsed-field gel electrophoresis showed that all three copies



FIG. 3. Analysis of the *vapD* gene product in *E. coli*. PAGE of proteins was produced by recombinant *E. coli* cells carrying pJIR318 (lane 1), pJIR330 (lane 2), pJIR515 (lane 3), and pUC18 (lane 4). Note that 10 times as much cell extract is loaded in lane 4. Molecular weight standards are in lane S. Plasmids pJIR330 and pJIR515 are deletion derivatives of pJIR318, as shown in the accompanying diagram. Cell extracts were prepared and subjected to electrophoresis on tricine-SDS-polyacrylamide gels as described in Materials and Methods.

of the vap region were located on the D. nodosus chromosome (22). However, it is possible that this region was derived from the integration into the D. nodosus chromosome of a conjugative plasmid from another species. We suggest that this event was followed by gene duplications which led to the presence of three copies of the vap region in the D. nodosus chromosome. Since the Southern hybridization experiments showed that in strain A198 only two of these copies carry the vapABC genes, subsequent deletion events presumably have led to the evolution of the region which carries only vapD. Previous results showed that the number of copies of the vap region varies between different isolates of D. nodosus (17). These data suggest that the putative integration of foreign DNA into the D. nodosus chromosome was not a relatively recent event in evolutionary terms.



FIG. 4. Western immunoblots of D. nodosus cell extracts. Cells from 18 isolates of D. nodosus were adjusted for cell numbers, solubilized in SDS-PAGE sample buffer, and fractionated on a 15% (wt/vol) SDS-PAGE gel. The polypeptides were transferred to nitrocellulose and reacted in an immunoblot with rabbit anti-VapD antiserum. VapD reacted as a broad ca. 6-kDa band in all wells except lanes 14, 15, 16, and 18. Lanes show cell extracts from the following isolates: 1, B1006; 2, C1008; 3, HA343; 4, HA352; 5, AC419; 6, AC424; 7, HA337; 8, HA390; 9, HA386; 10, HA393; 11, HA389; 12, HA340; 13, HA652; 14, AC554; 15, AC6; 16, AC176; 17, A198; 18, 305. Strains B1006, C1008, and A198 were virulent; strains HA343, HA352, HA337, HA390, HA386, HA393, and AC554 were intermediate; and all other strains were benign. Virulence designations were made by the laboratory of origin. All of these isolates except strains AC554, AC6, AC176, and 305 hybridized with pJIR318. Strain AC554 was one of only two intermediate strains which did not hybridize to pJIR318 (17).

### ACKNOWLEDGMENTS

We thank Pauline Howarth for her most capable technical assistance; Stephen Billington, Neville Firth, and Dave Berryman for assistance with computer analysis; John Egerton, Laurie Depiazzi, and Weng Yong for the D. nodosus isolates; and Stephen Billington and John Davies for helpful discussions.

We also thank the Australian Wool Research and Development Corporation for their research support.

#### REFERENCES

- Anderson, B. J., M. M. Bills, J. R. Egerton, and J. S. Mattick. 1984. Cloning and expression in *Escherichia coli* of the gene encoding the structural subunit of *Bacteroides nodosus* fimbriae. J. Bacteriol. 160:748–754.
- Bailey, M. J., A. Cockayne, and C. W. Penn. 1987. Monoclonal antibodies directed against surface associated polypeptides of *Treponema pallidum* define a biologically active antigen. J. Gen. Microbiol. 133:1793-1803.
- 3. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513–1523.
- Bradshaw, H. D., Jr., B. A. Traxler, E. G. Minkley, Jr., E. W. Nester, and M. P. Gordon. 1990. Nucleotide sequence of the *tra1* (helicase I) gene from the sex factor F. J. Bacteriol. 172:4127– 4131.
- Demarco de Hormeache, R., M. J. Thornley, and A. Holmes. 1983. Surface antigens of gonococci: correlation with virulence and serum resistance. J. Gen. Microbiol. 129:1559–1567.
- 6. Depiazzi, L. J., and R. B. Richards. 1979. A degrading proteinase test to distinguish benign and virulent ovine isolates of *Bacteroides nodosus*. Aust. Vet. J. 55:25–28.
- Depiazzi, L. J., and R. B. Richards. 1985. Motility in relation to virulence of *Bacteroides nodosus*. Vet. Microbiol. 10:107-116.
- Depiazzi, L. J., R. B. Richards, J. Henderson, J. I. Rood, M. Palmer, and W. J. Penhale. 1991. Characterisation of virulent and benign strains of *Bacteroides nodosus*. Vet. Microbiol. 26:151-160.
- 9. Depiazzi, L. J., and J. I. Rood. 1984. The thermostability of proteases from virulent and benign strains of *Bacteroides no-dosus*. Vet. Microbiol. 9:227-236.
- 10. Dewhirst, F. E., B. J. Paster, S. La Fontaine, and J. I. Rood. 1990. Transfer of *Kingella indologenes* (Snell and Lapage 1976) to the genus *Suttonella* gen. nov. as *Suttonella indologenes* comb. nov.; transfer of *Bacteroides nodosus* (Beveridge 1941) to the genus *Dichelobacter* gen. nov. as *Dichelobacter nodosus* comb. nov.; and assignment of the genera *Cardiobacterium*,

Dichelobacter, and Suttonella to Cardiobacteriaceae fam. nov. in the gamma division of Proteobacteria based on 16S ribosomal ribonucleic acid sequence comparisons. Int. J. Syst. Bacteriol. **40:**426–433.

- 11. Elleman, T. C. 1988. Pilins of *Bacteroides nodosus*: molecular basis of serotypic variation and relationships to other bacterial pilins. Microbiol. Rev. 52:233-247.
- 12. Every, D. 1982. Proteinase isoenzyme patterns of *Bacteroides* nodosus: distinction between ovine virulent isolates, ovine benign isolates and bovine isolates. J. Gen. Microbiol. 128:809-812.
- Green, R. S. 1985. A method to differentiate between virulent and benign isolates of *Bacteroides nodosus* based on thermal stability of their extracellular proteinases. N.Z. Vet. J. 33:11– 13.
- Higgins, D. G., and P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73:237-244.
- 15. Hoopes, B. C., and W. R. McClure. 1987. Strategies in regulation of transcription initiation, p. 1231–1240. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
- 16. Jalakumari, M. B., and P. A. Manning. 1989. Nucleotide sequence of the *traD* region in the *Escherichia coli* sex factor. Gene 81:195-202.
- Katz, M. E., P. M. Howarth, W. K. Yong, G. G. Riffkin, L. J. Depiazzi, and J. I. Rood. 1991. Identification of three gene regions associated with virulence in *Dichelobacter nodosus*, the causative agent of ovine footrot. J. Gen. Microbiol. 137:2117– 2124.
- Klein, P., M. Kanehisa, and C. Delisi. 1985. The detection and classification of membrane spanning proteins. Biochim. Biophys. Acta 815:468–476.
- Korch, C., P. Hagblom, H. Öhman, M. Göransson, and S. Normark. 1985. Cryptic plasmid of *Neisseria gonorrhoeae*: complete nucleotide sequence and genetic organization. J. Bacteriol. 163:430-438.
- Kortt, A. A., J. E. Burns, and D. J. Stewart. 1983. Detection of the extracellular proteases of *Bacteroides nodosus* in polyacrylamide gels: a rapid method of distinguishing virulent and benign ovine isolates. Res. Vet. Sci. 35:171–174.
- 21. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680-685.
- 22. La Fontaine, S., and J. I. Rood. Unpublished results.
- 23. Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. Science 227:1435-1441.
- Manning, P. A., G. Morelli, and C. Fisseau. 1984. RNA polymerase binding sites within the *tra* region of the F sex factor of *Escherichia coli* K-12. Gene 27:121–123.
- 25. Meyer, T. F., N. Mlawer, and M. So. 1982. Pilus expression in

*Neisseria gonorrhoeae* involves chromosomal rearrangement. Cell **30**:45–52.

- 26. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Moses, E. K., and W. K. Yong. 1989. Antigens of *Bacteroides nodosus*, p. 121–134. *In J. R. Egerton, W. K. Yong, and G. G. Riffkin (ed.), Footrot and foot abscess of ruminants. CRC Press, Inc., Boca Raton, Fla.*
- Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological comparison. Proc. Natl. Acad. Sci. USA 85:2444– 2448.
- Roberts, M., P. Piot, and S. Falkow. 1979. The ecology of gonococcal plasmids. J. Gen. Microbiol. 114:491–494.
- 30. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- 32. Schägger, H., and G. Von Jagow. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. 166:368–379.
- 33. Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of E. coli 16S ribosomal RNA complementary to nonsense triplets and ribosome-binding sites. Proc. Natl. Acad. Sci. USA 71: 1342–1346.
- Skerman, T. M. 1975. Determination of some *in vitro* growth requirements of *Bacteroides nodosus*. J. Gen. Microbiol. 87: 107-119.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.
- 36. Stewart, D. J. 1979. The role of elastase in the differentiation of *Bacteroides nodosus* infections in sheep and cattle. Res. Vet. Sci. 27:99-105.
- Stewart, D. J. 1989. Footrot of sheep, p. 5–45. *In J. R. Egerton*, W. K. Yong, and G. G. Riffkin (ed.), Footrot and foot abscess of ruminants. CRC Press, Inc., Boca Raton, Fla.
- 38. Stewart, D. J., J. E. Peterson, J. A. Vaughan, B. L. Clark, D. L. Emery, J. B. Caldwell, and A. A. Kortt. 1986. The pathogenicity and cultural characteristics of virulent, intermediate and benign strains of *Bacteroides nodosus* causing ovine foot-rot. Aust. Vet. J. 63:317–326.
- 39. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA 76:4350-4354.
- Yoshioka, Y., Y. Fujita, and E. Ohtsubo. 1990. Nucleotide sequence of the promoter-distal region of the *tra* operon of plasmid R100, including *traI* (DNA helicase I) and *traD* genes. J. Mol. Biol. 214:39-53.