Mol Gen Genet (1996) 251:99–107

ORIGINAL PAPER

Léon Otten · Patrice De Ruffray
Philippe de Lajudie · Bernard Michot

Sequence and characterisation of a ribosomal RNA operon from *Agrobacterium vitis*

Received: 19 July 1995 / Accepted: 5 January 1996

Abstract One of the four ribosomal RNA operons (*rrnA*) from the *Agrobacterium vitis* vitopine strain S4 was sequenced. *rrnA* is most closely related to the *rrn* operons of *Bradyrhizobium japonicum* and *Rhodobacter sphaeroides* and carries an fMet-tRNA gene downstream of its 5S gene, as in the case of *R. sphaeroides*. The 16S rRNA sequence of S4 differs from the *A. vitis* K309 type strain sequence by only one nucleotide, in spite of the fact that S4 and K309 have very different Ti plasmids. The predicted secondary structure of the S4 23S rRNA shows several features that are specific for the alpha proteobacteria, and an unusual branched structure in the universal B8 stem. The 3′ ends of the three other *rrn* copies of S4 were also cloned and sequenced. Sequence comparison delimits the 3′ ends of the four repeats and defines two groups: *rrnA/rrnB* and *rrnC/rrnD*.

Key words Ribosomal genes · *Agrobacterium* · Evolution · 23S rRNA

Introduction

Agrobacterium vitis (Ophel and Kerr 1990) is a member of the *Rhizobiaceae*, an important family of plant-associated bacteria which comprises the genera *Rhizobium*, *Agrobacterium*, *Bradyrhizobium* and several other minor groups. *A. vitis* is closely related to *R. galegae*, *A. tumefaciens*, *A. rubi* and *A. radiobacter*, whereas *A. rhizogenes* clusters with *R. tropici*. The classification of the *Rhizobiaceae* is based, among other criteria, on the sequences of their 16S rRNA genes (Young et al. 1991; Willems and Collins 1993; Yanagi and Yamasato 1993; Laguerre et al. 1994) which form part of the ribosomal RNA (*rrn*) operons. Often bacteria contain multiple *rrn* operons. *Escherichia coli* (Lindahl and Zengel 1986), *Bacillus subtilis* 168 (Itaya 1993), *Clostridium botulinum* (East et al. 1992), *Enterococcus hirae* (Sechi and Daneo-Moore 1993) and *Rhodobacter sphaeroides* (Dryden and Kaplan 1990), for example, contain seven, ten, six, six and three *rrn* operons, respectively. In phylogenetic studies based on 16S rRNA gene sequences, the presence of multiple *rrn* copies, possible sequence differences between these copies and the effects of intercopy gene conversion are generally ignored.

Among the *Rhizobiaceae*, *Rhizobium meliloti* has been reported to have at least three *rrn* loci (Honeycutt et al. 1993). Preliminary hybridization results in our laboratory indicated that *A. vitis* strain S4 contains four 16S rRNA gene copies. Earlier studies showed that the *A. vitis* group is composed of nopaline, vitopine and octopine/cucumopine (o/c) strains, the last being subdivided into "large TA" and "small TA" strains (TA refers to one of the two T-DNAs of these strains). These strains can be distinguished by characteristic bacterial insertion (IS) elements found at specific chromosomal sites (Paulus et al. 1989; Otten et al. 1992). The Ti plasmids of the o/c groups are of recent clonal origin and evolved mainly through horizontal DNA transfer, IS element insertion, recombination, and loss or acquisition of large DNA fragments. The nopaline and vitopine Ti plasmids are less variable (Otten et al. 1992; van Nuenen et al. 1993; Otten and van Nuenen 1993; Otten and De Ruffray 1994). The particular mode of evolution of the o/c group is rather unexpected and has important consequences for phylogenetic reconstruction; classical approaches, such as DNA homology

Communicated by A. Kondorosi

L. Otten (✉) · P. De Ruffray
Institute of Plant Molecular Biology of the CNRS,
Rue du Général Zimmer 12, 67084 Strasbourg Cedex, France

P. de Lajudie
Laboratory of Microbiology ORSTOM, B.P. 1386, Dakar, Senegal

B. Michot
Laboratory of Molecular Biology of Eucaryotes of the CNRS,
Route de Narbonne 118, 31062 Toulouse Cedex, France

studies or RFLP analysis, fail to yield correct phylogenetic trees (van Nuenen et al. 1993). To extend the genetic analysis of *A. vitis* to its chromosomal DNA, we have cloned and sequenced one of the four *rrn* operons of the vitopine strain S4, as well as the 3′ ends of the three other *rrn* copies in this strain.

Materials and methods

Bacterial strains and plasmids

pUC18 and pKS Bluescript derivatives were propagated in *E. coli* strain NM522. S4 is a virulent wild-type *A. vitis* strain of the vitopine type, isolated from grapevine (Szegedi 1985; Gérard et al. 1992; Canaday et al. 1992).

Cloning, sequencing and other molecular biological techniques

Molecular biological techniques were used as described (Sambrook et al. 1989). Sequencing was done on both strands using double-stranded DNA according to the method of Sanger et al. (1977), modified for use with Sequenase. Sequences were analysed on an alpha workstation using the GCG sequence analysis software package, version 8.0 (Devereux et al. 1987); DNA homology searches were carried out with the Blast programs.

Secondary structure analysis

The 23S rRNA was analysed separately using programs and procedures adapted to this molecule. First, the regions in the *A. vitis* sequence belonging, respectively, to the universally conserved core and to the divergent domains (Hassouna et al. 1984) were identified by aligning the complete 23S sequence using both sequence and secondary structure homologies (Corpet and Michot 1994) with a specialised and structured databank containing, in an aligned format, the sequences belonging to the universally conserved core and its secondary structure folding from a subset of the available 23S eubacterial sequences (B. Michot and F. Corpet, unpublished). The precise 5′ and 3′ ends of *A. vitis* insertions/deletions were thus located. The E12-13 and B8 domains were compared for the five available members of the alpha proteobacteria: *A. vitis* (this work), *Bradyrhizobium japonicum* (Z35330; Kuendig et al., unpublished), *Rhodopseudomonas palustris* (X71839; Springer et al. 1993), *Rhodobacter capsulatus* (X06485; Hopfl et al. 1988) and *Rhodobacter sphaeroides* (X53853,-4,-5; Dryden and Kaplan 1990). The secondary structure folding of the additional sequence in B8 was analysed by a combination of thermodynamic (Zuker and Stiegler 1981) and phylogenetic approaches. Secondary structures were drawn using the RNA_d2 program (Perochon-Dorisse et al. 1995).

Results

Cloning of a complete *rrn* copy

The *A. vitis* vitopine-type strain S4 (Szegedi 1985) was chosen as the source for the *rrn* sequences. S4 has been used as the standard vitopine strain in several studies. Its Ti plasmid has been mapped (Gérard et al. 1992) and its three T-DNA regions have been sequenced (Canaday et al. 1992). A 280-bp fragment internal to the 16S rRNA gene was amplified from S4 DNA by PCR using the conserved 16S sequences Y1 and Y2 (Young et al. 1991) as primers. *Pst*I or *Hin*dIII fragments from S4 DNA were ligated to pUC18 and transformed into *E. coli* strain NM522. Colonies containing 16S rRNA sequences were detected by colony hybridisation using the internal 16S fragment as probe. Two clones were recovered: pPM1150 and pPM1151, with a 11.3 kb *Pst*I insert and a 6.9 kb *Hin*dIII fragment, respectively. Restriction mapping indicated that the two clones overlapped by 3.2 kb. Preliminary Southern hybridisation experiments with the PCR probe showed that S4 contained four 16S rRNA gene copies. It was therefore necessary to show that pPM1150 and pPM1151 were derived from the same locus. Total S4 DNA was

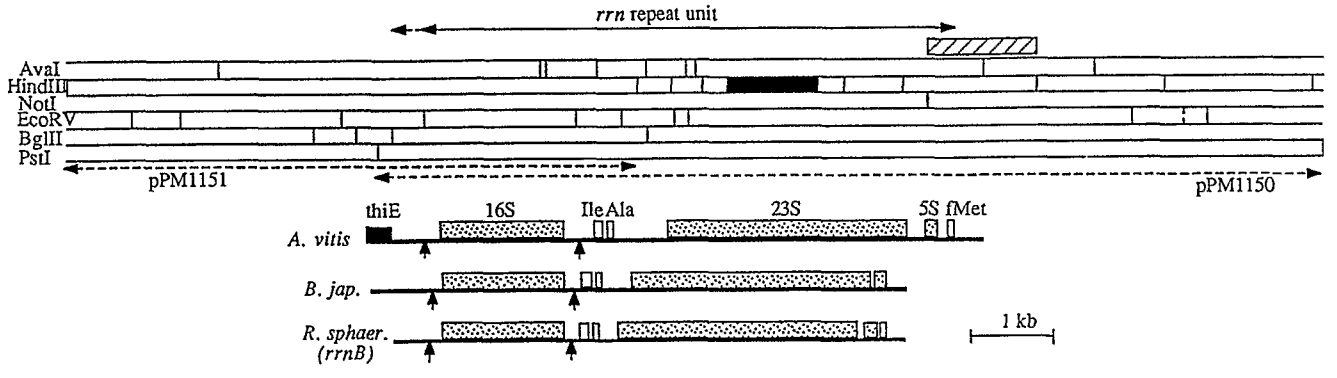


Fig. 1 Restriction (*top*) and genetic (*bottom*) maps of the *rrnA* operon of *A. vitis* S4, compared with the maps for *B. japonicum* (Kuendig et al., unpublished) and for *rrnB* of *R. sphaeroides* (Dryden and Kaplan 1990). Maps are aligned at the start of the 16S rRNA genes. The *hatched box* represents the *Not*I-*Hin*dIII fragment used to clone the 3′ ends of the three other *rrn* copies in *A. vitis*, the inverted repeats flanking the 16S rRNA gene (the sequences in *A. vitis* and *B. japonicum* are similar and differ from those in *R. sphaeroides*) are indicated by *vertical arrows*. Sequenced regions are indicated below the restriction maps. The *Hin*dIII fragment indicated by the *filled box* detects polymorphisms between *rrn* loci in *A. vitis*

digested with *Ava*I in combination with several other enzymes. Southern analysis with the 3.8 kb *Ava*I fragment from pPM1151 as probe revealed four *Ava*I fragments of 3.0, 3.8, 4.45 and 4.5 kb. Only the 3.8 kb fragment was cut by *Bgl*II and *Pst*I. *Bgl*II and *Pst*I sites were present in pPM1150 and pPM1151 at the positions expected from the Southern data (results not shown). Thus pPM1150 and pPM1151 are derived from the same locus; the combined map is shown in Fig. 1.

Extent of the *rrn* repeat unit

To determine the extent of the *rrn* repeat unit, S4 DNA was digested with *Ava*I and one of 17 other restriction enzymes. The 3.8- and 3.5-kb *Ava*I fragments of pPM1151 and pPM1150 representing the 5′ and 3′ ends of the *rrn* region were hybridised to Southern blots of the double digests (not shown). The 5′ end of the repeat lies between the *Bgl*II and *Eco*RV site, the 3′ end between the *Not*I and *Hin*dIII site (Fig. 1). The 3′ end was determined more precisely by sequencing (see below).

Sequence of the *rrn* locus and comparison to other sequences

A 7.5-kb fragment containing the complete *rrn* operon and its flanking sequences was sequenced by subcloning fragments from pPM1150 and pPM1151 into pKS Bluescript. The sequence (EMBL accession number U28505) shows the characteristic features of an *rrn* operon (arbitrarily called *rrnA*) and was compared to other *rrn* sequences. The three *rrn* operons from *R. sphaeroides* (*rrnA*, *B* and *C*, X53853, -4, -5; Dryden and Kaplan 1990) and the single *rrn* operon from *B. japonicum* (Z35330, Kuendig et al., unpublished) were found to be the most closely related ones in the database. The three *rrn* operons from *R. sphaeroides* are very similar to each other. As *rrnB* is slightly more similar to the S4 sequence than *rrnA* and *rrnC* (results not shown) we chose *rrnB* for comparison.

Figure 1 shows the overall organisation of the *rrn* operons from the three bacterial species. Table 1 summarises the positions of the salient features of the *rrn* operons and their similarities to the *A. vitis* sequence. Additional small areas of similarity with unknown function were found between the 16S and 23S rRNA genes of *B. japonicum* and *A. vitis*. The distance between the 16S rRNA gene and the 23S rRNA gene is considerably larger in *A. vitis* (1274 nt) than in *B. japonicum* (809 nt) or *R. sphaeroides* (666 nt). This fragment contains two tRNA genes, as in *B. japonicum* and *R. sphaeroides*: Ile-tRNA and Ala-tRNA. The *A. vitis* sequence is slightly more similar to the *B. japonicum* sequence than to the *R. sphaeroides* sequence (89% vs. 88.3% for the 16S rRNA gene, 85.4% vs. 84.7% for the 23S rRNA gene). Its 16S rRNA gene is flanked by the same inverted repeat as in *B. japonicum*, a *R. sphaeroides* repeat at the corresponding positions has a different sequence. In spite of this, the *A. vitis* sequence shares an important feature with the three *R. sphaeroides rrn* operons: an fMet-tRNA gene is present 3′ to the 5S rRNA gene, but absent in *B. japonicum*.

Upstream of the *A. vitis rrnA* operon, a 326-nt sequence was found, with 58.4% DNA homology to the end of the *thiE* gene from the *thiCEFGH* operon of *E. coli* (M88701; Vander Horn et al. 1993). Whereas in *E. coli* the *thiE* gene is followed by *thiFGH*, there are apparently no such sequences downstream of the *A. vitis thiE* gene, neither 5′ nor 3′ to the *rrnA* operon. The extreme left end of the pPM1151 insert has 72% DNA homology with the *thiC* gene (our unpublished observations). In *E. coli*, *thiC* and *thiE* are separated by 1.9 kb, in *A. vitis* by 3.7 kb. DNA homology to the *R. sphaeroides rrnB* operon starts only 95 nt after the putative *A. vitis thiE* stop codon.

Table 1 Features of *rrn* operons from *A. vitis*, *B. japonicum* and *R. sphaeroides*. Values in parentheses indicate percentage DNA homology values

Feature	*A. vitis* S4	*B. japonicum*	*R. sphaeroides* (*rrnB*)
Box 1[a]	419–520	412–463 (65.4)	362–465 (68.0)
Box 2[a]	686–727	703–743 (75.0)	None
Inverted repeat	621–647	620–645 (96.0)	559–585, nh[b]
16S rRNA gene	880–2359	840–2330 (89.0)	765–2230 (88.3)
Inverted repeat	2541–2567	2427–2456 (95.8)	2308–2334, nh[b]
tRNA (Ile)	2754–2830	2557–2633 (89.6)	2452–2528 (83.1)
tRNA (Ala)	2914–2990	2666–2742 (84.2)	2576–2651 (92.1)
23S rRNA gene	3633–6574	3139–6012 (85.4)	2896–5775 (84.7)
5S gene	6805–6923	6119–6237 (78.2)	5873–5988 (77.2)
tRNA (fMet)	7041–7117	None	6031–6107 (100.0)
Terminator	7147–7165	6282–6314, nh[b]	6136–6154, nh[b]

[a] Boxes 1 and 2, sequences found in *A. vitis* and one of the other species
[b] nh, similar feature but no sequence homology

Comparison of 16S-rRNA gene sequence with related sequences

The 16S rRNA gene sequence of the S4 *rrnA* operon is very similar to the sequences published earlier (X67225, D14502 and D12795) from the *A. vitis* type strain NCPPB3554 (the same as LMG8750, IAM14140 or K309; Ophel and Kerr 1990). NCPPB3554 belongs to the o/c large TA group (our unpublished observations). The next most related 16S rRNA sequences are those from *R. galegae* (D11343, D12793 and X67226). In spite of the fact that the published *A. vitis* sequences are derived from the same strain, they are not identical. X67225 has TT at positions 1630–1631 (S4 coordinates), the other sequences, including those from *R. galegae* have CC. At positions 1680–1681, S4, X67225 and X67226 have GC, D14502 and D11343 CG, the others NN. D14502 lacks a C at 1827 and has an additional T between 1840 and 1841; the other sequences are identical at these positions. The various differences may be due to sequencing errors or microheterogeneity between strains. The only unambiguous difference between the S4 and the type strain 16S rRNA sequence is therefore the C to A change at position 1048. This difference creates a diagnostic *Bsu*36I and *Dde*I restriction site for the S4 sequence.

Secondary structure of 23S rRNA

Eubacterial large subunit rRNA sequences have a relatively constant length of approximately 2900 nt (2904 in *E. coli*). About 90% of this molecule can be folded into exactly the same secondary-structure shape in all species. This universal core of secondary structure is interrupted at a dozen sites by more variable regions called divergent (or D) domains. The universal and divergent domains of the *A. vitis* sequence (2942 nt) were localized by comparison with a datafile containing an alignment of the universally conserved cores of secondary structure from a subset of species chosen from the major eubacterial groups. The sizes of the divergent domains of *A. vitis* are similar to those of other eubacteria. By contrast, in the conserved core of secondary structure, four insertion/deletion events have occurred. Remarkably, these structural features

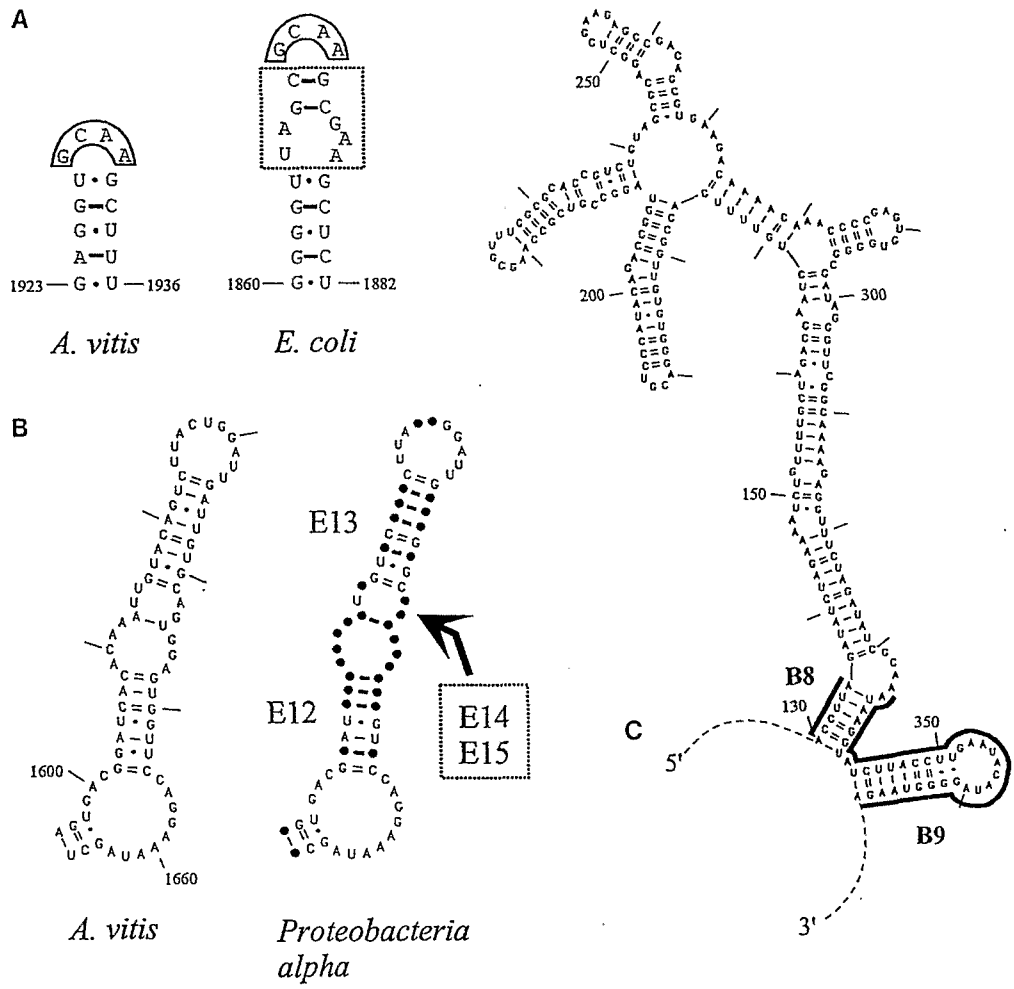


Fig. 2a–c Folding patterns at sites of size variation in *A. vitis* 23S rRNA. For each secondary structure model positions are numbered from the 5′ end of the 23S rRNA. The stems are identified according to De Rijk et al. (1994). **a** Folding of the E25 stem. The *dotted line* delineates an *E. coli* structural feature that is deleted in *A. vitis*. **b** Details of base pairing in stems E12-13. Nucleotide substitutions with respect to the alpha proteobacteria consensus are indicated by *dots* and compensatory base changes allow formation of the base pairs indicated by *thick bars*. **c** Secondary structure folding of the B8 domain in *A. vitis*. Sequence stretches which belong to the universally conserved core of secondary structure are indicated by the *thick line*

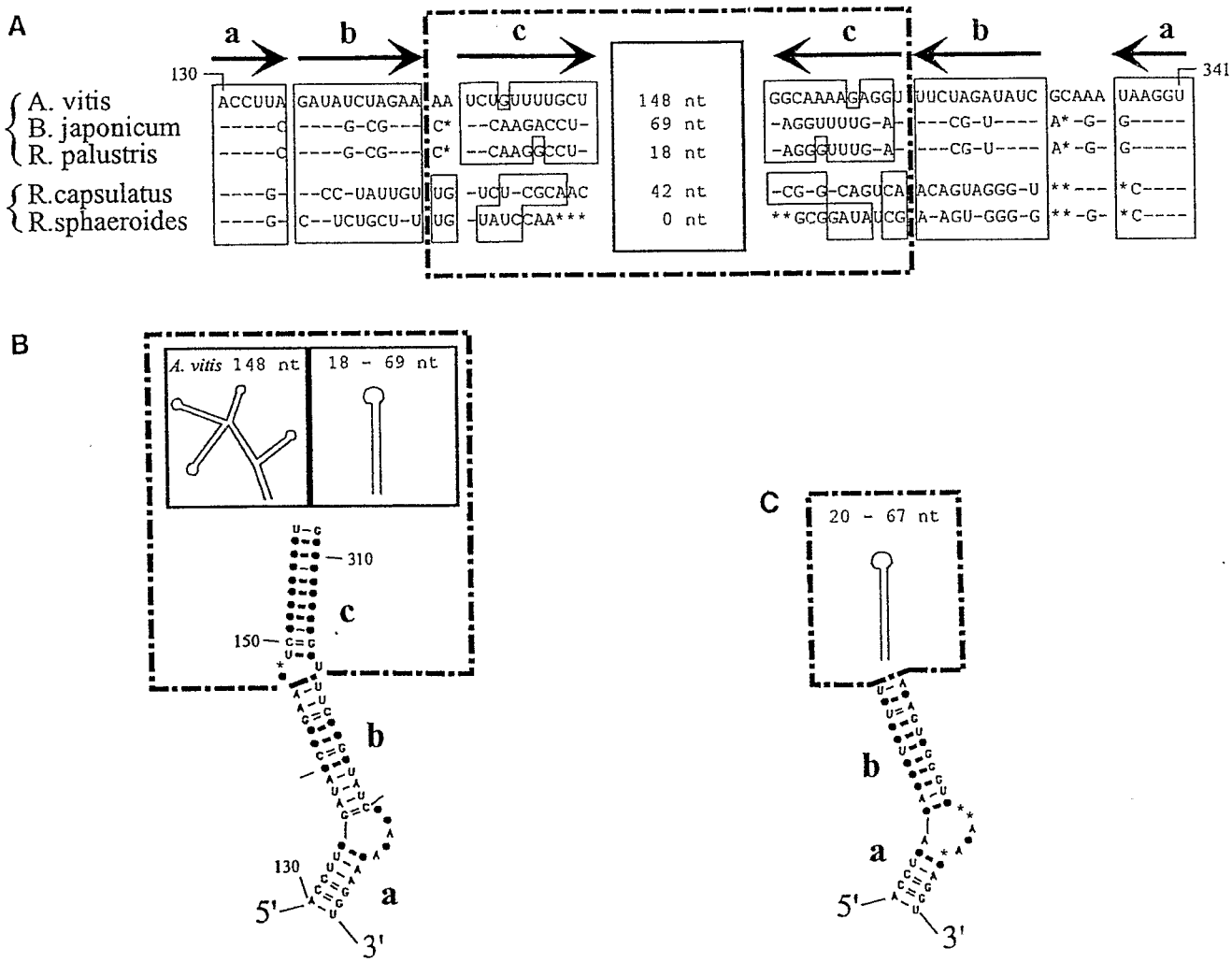
are also found in the four other alpha proteobacteria for which sequence data are available: *B. japonicum*, *R. palustris*, *R. capsulatus* and *R. sphaeroides*, and only in these species. The first one is an addition of a UAAAU motif in stem D17 between *E. coli* positions 1109 and 1110, a single-strand region in the 23S secondary structure model. The second is a 9-nt deletion within the *E. coli* region 1860–1882, which folds in all eubacteria as an irregular hairpin stem (E25) containing a GNRA tetraloop (Fig. 2a). In alpha proteobacteria the apical part of this stem, variable in eubacteria, is lacking, but the universal GNRA hairpin loop motif is preserved as is the 5 bp of the basal part of the stem, the folding of which is strongly constrained among eubacteria. Two G-U base pairings, at positions 3 and 5 of E25, conserved in nearly all eubacterial species, are also conserved in *A. vitis* and other alpha proteobacteria. The two stems E14 and E15, which in *E. coli* encompass 60 nt between positions 1486 and 1546, are deleted in *A. vitis* (Fig. 2b). The E12-E13 stem-loop structure is strongly constrained in alpha proteobacteria, while this region presents some structural variation among other eubacteria. However, the most dramatic change in size concerns an insertion of about 200 nt in the hairpin loop of the universal stem B8 (Fig. 2c), which is only about 16 nt long in most eubacteria. The *A. vitis* B8 domain can be divided into three parts based on its folding: the universal 6 consecutive base pairs, a long unbranched stem of 29 bp which is thermodynamically very stable owing to long stretches of uninterrupted base pairing without irregularities, and, finally, the apical part for which thermodynamic calculations yield a complex branched structure. An increase in the size of B8, albeit less pronounced, is also found in the other alpha proteobacteria: 56, 80, 103 and 131 nt for *R. sphaeroides*, *R. palustris*, *R. capsulatus* and *B. japonicum*, respectively.

We have analysed the structure of B8 in more detail (Fig. 3). The 5′ and 3′ ends of this domain are highly conserved between *B. japonicum* and *R. palustris* over 29 and 33 nt, respectively, with only two changes (Fig. 3a). These two regions can be base paired in three stems which are strictly homologous to the *A. vitis* stems B8a, B8b and B8c. Remarkably, all the differences between *A. vitis* on the one hand and *B. japonicum* and *R. palustris* on the other involve compensatory changes, giving strong phylogenetic support for the existence of stems B8b and B8c. B8c, in particular, is preserved during the course of evolution by an impressive number of compensatory changes (Fig. 3b). A comparison between the two *Rhodobacter* B8 sequences also shows unambiguous sequence homologies of about 20 nt at both ends of this domain (Fig. 3c). These conserved motifs can be base paired over 11 consecutive nt at the same locations in the secondary structure as in B8b of *A. vitis*, *B. japonicum* and *R. palustris*, suggesting that this rhodobacter stem is the counterpart of B8b despite small size variations in the B8a-B8b bulge. By contrast rhodobacter B8c folding differs markedly between *R. capsulatus* and *R. sphaeroides*. Within each subgroup (comprising *A. vitis*, *B. japonicum* and *R. palustris* and the two other Rhodobacter species) the intervening domain shows substantial size variation, and no homologous sequence tract can be detected even between phylogenetically very close species. This hot spot for variation preferentially folds as an unbranched stem of variable length with details of pairing differing between species, except in *A. vitis*, where this structure is of exceptional length (Fig. 3b–c).

Sequences at the 3′ ends of the other *rrn* loci of S4

Partial mapping of the four *rrn* copies of S4 by hybridisation analysis (see above) yielded only approximate limits for the repeat unit. To establish the precise limits at one of the ends and to obtain locus-specific probes, the 3′ end of each *rrn* copy and its neighboring sequences were cloned and partially sequenced. For this, a *Not*I site within the 5S gene and a *Hin*dIII site outside the repeat were used. *Not*I-*Hin*dIII fragments were cloned into the pKS Bluescript vector and *rrn* clones were identified by colony hybridisation with the 1.25 kb *Not*I-*Hin*dIII fragment of pPM1150 (Fig. 1) as a probe. Four clones were found with inserts of 2.2, 1.3, 1.25 kb and 1.45 kb (pPM1177, 1178, 1179 and 1181, respectively) and analysed by restriction mapping. As expected, the inserts had a common segment of 0.3 kb, corresponding to the 3′ end of the *rrn* operon. pPM1179 was derived from the *rrnA* locus; the other clones were derived from the other three loci, arbitrarily called *rrnB* (pPM1181), *rrnC* (pPM1177) and *rrnD* (pPM1178). The four 3′ end clones were partially sequenced: first from their common *Not*I site (position 6826 in *rrnA*), then from a sequence closer to the divergence point, using an oligonucleotide primer (7072–7088, within the fMet-tRNA gene). The four sequences are identical from the *Not*I site up to 11 nt beyond the end of the fMet-tRNA gene (6862–7128); *rrnA* and *rrnB* are identical for an additional 59 nt, *rrnC* and *rrnD* for an additional 40 nt (Fig. 4). Thus, *rrnA* and *rrnB* form one group of *rrn* operons, and *rrnC* and *rrnD* form a second group.

In the case of the *R. sphaeroides rrn* sequences, the *rrnA*, *B* and *C* operons diverge immediately after the fMet-tRNA gene; inverted repeats 3′ to the fMet-tRNA sequence have been proposed as possible rho-independent termination signals (Dryden and Kaplan 1990). An inverted repeat is also found in *B. japonicum*, 41 nt beyond the end of the 5S gene (6278–6309). At positions 7147–7169 of the S4 *rrnA* and *rrnB* operons, 28 nt after the end of the fMet-tRNA gene, an inverted repeat (AAGCCCTCCCAAAAGGAGGGCTT) is found. The *rrnC* and *rrnD* sequences have a different inverted repeat (AATCCCGCAGCCTTAAACGCTGCGGGATT), 13 nt beyond the end of the fMet-tRNA



gene. These inverted repeats may represent termination signals.

At the end of the sequenced part of *rrnB* an 87-nt sequence was found with 63% DNA homology to the 5′ end of ORFA from IS*1222* of *Enterobacter agglomerans* (X78052, F. M. Lewecke, unpublished).

Fig. 3a–c B8 domain sequence comparisons among alpha proteobacteria. **a** Alignment of sequence from the five available species in this group. *A. vitis* serves as the reference, with nucleotide positions numbered from 5′ end of the 23S sequence. For the other sequences only substituted nucleotides are indicated, identities are denoted by *hyphens*. Base-paired nucleotides are *boxed* with stems delineated by *pairs of arrows in opposite orientations*, identified by *lower case letters*. Portions of the sequence which cannot be folded into the same secondary structure in the two groups of species (one comprising *A. vitis*, *B. japonicum*, and *R. palustris*, the other *Rhodobacter species*) are delineated by a *dotted box*, whereas regions of intra-group variations (which cannot be aligned even between closely related species) are *boxed*, with sizes indicated for each species. **b**, **c** Secondary structure consensus for *A. vitis*, *B. japonicum*, *R. palustris* (**b**) and the two *Rhodobacter* species (**c**). Substituted nucleotides are shown by *dots*, *thick bars* indicate base pairs made possible by compensatory base changes. A backbone view is presented for regions of intra-group variation with an indication of the range of variation

Discussion

A. vitis strain S4 contains four copies of an *rrn* operon. One of these, called *rrnA*, has been cloned and sequenced. This is the first report on a complete *rrn* operon sequence in the Rhizobium-Agrobacterium group. Only one *rrn* operon from the *Rhizobiaceae* family has so far been sequenced – the single *rrn* copy of *B. japonicum* (Kuendig et al., unpublished). This bacterium differs considerably from the members of the Rhizobium-Agrobacterium group (Willems and Collins 1993; Yanagi and Yamasato 1993). The low degree of similarity between the *B. japonicum* and *A. vitis* sequences outside the 16S, 23S and 5S genes confirms this conclusion. Interestingly, the S4 *rrnA* operon is larger than the *B. japonicum* and *R. sphaeroides rrn*

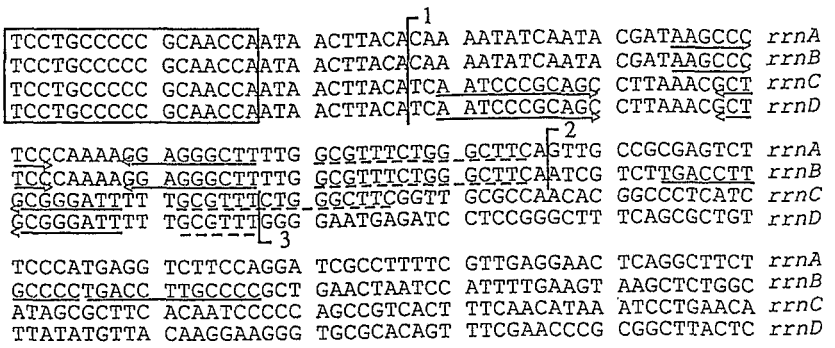


Fig. 4 Sequences at the 3′ ends of the four *rrn* copies in *A. vitis* S4. The 3′ end of the fMet-tRNA gene is *boxed*, *arrows* indicate inverted repeats. *Broken lines* denote common sequences 3′ of inverted repeats. The direct repeat in *rrnB* is *underlined*. Points of sequence divergence between *rrnA/B* and *rrnC/D*, between *rrnA* and *rrnB* and between *rrnC* and *rrnD* are numbered 1–3, respectively

operons; the difference is mainly due to the intergenic sequences, which are about twice as long in S4.

A comparison of the *A. vitis* 23S sequence to 23S sequences from the major branches of eubacteria revealed four variable fragments. Although considerably shorter than the variable fragments of intergenic sequences, they are unexpectedly located in the universally conserved core of secondary structure and not within the divergent domains.

These four structural features are also present in the four other known 23S sequences of the alpha proteobacteria – and only within this monophyletic group. Accordingly they constitute typical proteobacterial signatures which might be used as targets in species identification experiments with specific oligonucleotide probes. Our determination of the *A. vitis* 23S rRNA gene sequence has enlarged the available collection of alpha proteobacteria 23S sequences, and enabled us to perform a detailed analysis of the folding patterns and mode of structural variation of these structural features. Thus, we have refined previous folding models (Gutell et al. 1993; De Rijk et al. 1994) for the E12-13 region (Fig. 2b) and identified the structural constraints which act on the variation of the extra domain B8 (Fig. 3). Our secondary structure models are strongly supported by the phylogeny of the alpha proteobacteria and indicate that structural constraints operate within this group to preserve folding at each of these sites, which have no equivalent in other eubacterial groups.

These alpha proteobacteria-specific features are located in four independent secondary structure domains of the universally conserved 23S rRNA core and in close proximity to structural features involved in ribosome function. The GTPase center, located in Domain II of the 23S rRNA (reviewed in Raue et al. 1990; Zimmerman et al. 1990), is rooted precisely at the D17 internal loop which contains the insertion of 5 nt in alpha proteobacteria. This highly conserved domain binds the L11 ribosomal protein, which induces a conformational transition of rRNA tertiary structure (Xing and Draper 1995).

Stem E25 (which belongs to Domain IV) contains the rRNA binding site for rpL1 (Branlant et al. 1976) and is close to a cross-link site to rpL2, one of the main constituents of the peptidyltransferase center. Finally, the attachment site for rpL23, a ribosomal protein located at the A site of the peptidyltransferase center of the ribosome (Vester and Garret 1984), is in the immediate vicinity of the E12-13 stems (Domain III of the 23S rRNA). As suggested by a set of crosslinks and tertiary interactions between Domains II, III and IV (Mitchell et al. 1990; Gutell et al. 1993; Brimacombe et al. 1990), but also by their proximity to rpL23 and rpL2, these three specific structural features could be located close to the peptidyltransfer center. Accordingly, they could correspond to "covarying events" acting together to conserve a functional spatial orientation for those nucleotides which are directly involved in the peptidyltransferase reaction.

The extra alpha proteobacterial domain B8 differs from the three other sites of size variation because it possesses all the characteristics of a divergent domain (D domain). It is highly variable in size and sequence, even between closely related species such as the two Rhodobacter species but is nevertheless subject to diverse structural constraints. During evolution these constraints maintained stem B8b among all alpha proteobacteria, and stem B8c only within the subgroup comprising *A. vitis*, *B. japonicum* and *R. palustris*. The detection of structural constraints preserving some lineage-specific structural features in B8, as previously observed for the D domains D2, D3 and D8 (Michot and Bachellerie 1987; Michot et al. 1990) suggests that these D domains may have a significant biological role. In line with this notion, an evolutionarily conserved essential role for the divergent domain D8 has recently been proposed (Sweeney et al. 1994).

The extra B8 region interrupts the 5′ terminal domain of the LSU rRNA (Domain I). Remarkably, a similar modification is also observed in insects at the tip of the universal stem B8, which in eukaryotes is located at the 3′ end of the 5.8S rRNA. It seems therefore that during evolution a divergent domain has arisen independently twice at the same site. In insects the extra B8 sequence contains processing sites which result in cleavage of the 5.8S into two species. Processing sites have been identified in 23S rRNA of several

eubacterial species, and in particular in *R. sphaeroides*, at another extra stem-loop structure (Dryden and Kaplan 1990; Gray and Schnare 1990; Pace and Burgin 1990). It remains to be shown whether processing occurs in *A. vitis* or in other alpha proteobacteria.

The ends of the four *A. vitis rrn* copies were determined by Southern analysis and partial sequencing. Sequences of the 3′ ends show that these copies can be divided into two groups; *rrnA/B* and *rrnC/D*. The location of the *rrn* copies on the chromosome remains unknown, the *rrnA* copy is associated with a thiamine locus, the *rrnB* copy is associated with an IS*1222*-like element.

Preliminary hybridisation experiments with an internal fragment of the *rrnA* operon (Fig. 1) on *Ava*I-digested DNAs from different *A. vitis* isolates revealed a high level of restriction fragment length polymorphism. The polymorphism may be due to restriction site variability around *rrn* loci at conserved positions, to the presence of *rrn* copies at different chromosomal locations, or to both. Copy-specific fragments will allow further analysis of these patterns and possibly provide some clues as to the evolutionary history of the different *rrn* operons in *A. vitis*. It will be particularly interesting to extend these studies to other members of the *Rhizobiaceae*. It should be noted that very little is known about the amplification of *rrn* operons in bacteria. Although the location of *rrn* copies has been determined for several species (Cole and Saint Girons 1994) it remains unknown how and when *rrn* operons were amplified and how this affected their structure.

The o/c large TA, o/c small TA, nopaline and vitopine subgroups within *A. vitis*, defined earlier on the basis of IS patterns (Otten et al. 1992; van Nuenen et al. 1993), may now be related to each other by analysing the structures of the *rrn* operons and their immediate environments. It has already been noted that the analysis of 16S rRNA gene sequences may not be sufficient to allow estimation of genetic relatedness (Fox et al. 1992). Extending sequence analysis beyond the 16S rDNA genes (using, for example, 16S-23S intergenic sequences or sequences surrounding the different *rrn* copies) will lead to better classification schemes and to further insights into the ways in which bacterial genomes evolve.

Acknowledgements We thank A. Hoeft for oligonucleotides.

References

Branlant C, Korobo V, Ebel JP (1976) The binding site of protein L1 on 23 ribosomal RNA from *E. coli*. Eur J Biochem 70:471–482

Brimacombe R, Greuer B, Mitchell P, Osswald M, Rinke-Appel J, Schueller D, Katrin S (1990) Three-dimensional structure and function of *Escherichia coli* 16S and 23S rRNA as studied by cross-linking techniques. In: Hill WE, Dahlberg A, Garret RA, Moore PB, Schlessinger D, Warner JR (eds) The ribosome: structure, function and evolution. Amer Soc Microbiol, Washington DC, pp 93–106

Canaday J, Gérard JC, Crouzet P, Otten L (1992) Organization and functional analysis of three T-DNAs from the vitopine Ti plasmid pTiS4. Mol Gen Genet 235:292–303

Cole ST, Saint Girons I (1994) Bacterial genomics. FEMS Microbiol Rev 14:139–160

Corpet F, Michot B (1994) RNAlign program: alignment of RNA sequences using both primary and secondary structures. Comp Appl Biosci 10:389–399

De Rijk P, van de Peer Y, Chapelle S, de Wachter R (1994) Database on the structure of large ribosomal subunit RNA. Nucleic Acids Res 22:3495–3501

Devereux JP, Haeberli P, Marquess P (1987) The program manual for sequence analysis software package of the Genetics Computer Group. Nucleic Acids Res 12:387–395

Dryden SC, Kaplan S (1990) Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter sphaeroides*. Nucleic Acids Res 18:7267–7277

East AK, Thompson DE, Collins MD (1992) Analysis of operons encoding 23S rRNA of *Clostridium botulinum* type A. J Bacteriol 174:8158–8162

Fox GE, Wisotzkey JD, Jurtshuk P (1992) How close is close: 16S rRNA sequence may not be sufficient to guarantee species identity. Int J Syst Bact 42:166–170

Gérard JC, Canaday J, Szegedi E, de la Salle H, Otten L (1992) Physical map of the vitopine Ti plasmid pTiS4. Plasmid 28: 146–156

Gray MW, Schnare MN (1990) Evolution of the modular structure of rRNA. In: Hill WE, Dahlberg A, Garret RA, Moore PB, Schlessinger D, Warner JR (eds) The ribosome: structure, function and evolution. Amer Soc Microbiol, Washington DC, pp 589–597

Gutell RR, Gray MW, Schnare MN (1993) A compilation of large subunit (23S and 28S-like) ribosomal RNA structures. Nucleic Acids 21:3055–3074

Hassouna N, Michot B, Bachellerie JP (1984) The complete nucleotide sequence of mouse rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eucaryotes. Nucleic Acids Res 12:3563–3583

Honeycutt RJ, McClelland M, Sobral BW (1993) Physical map of the genome of *Rhizobium meliloti* 1021. J Bacteriol 175: 6945–6952

Hopfl P, Ludwig W, Schleifer KH (1988) Complete nucleotide sequence of a 23S ribosomal RNA gene from *Rhodobacter capsulatus*. Nucleic Acids Res. 16:2343

Itaya M (1993) Physical mapping of multiple homologous genes in the *Bacillus subtilis* 168 chromosome: identification of ten ribosomal RNA operon loci. Biosci Biotechnol Biochem 57: 1611–1614

Laguerre G, Allard M-R, Revoy F, Amarger N (1994) Rapid identification of Rhizobia by restriction fragment length polymorphism analysis of PCR-amplified 16 rRNA genes. Appl Environ Microbiol 60:56–63

Lindhal L, Zengel JM (1986) Ribosomal genes in *Escherichia coli*. Annu Rev Genet 20:297–326

Michot B, Bachellerie JP (1987) Comparisons of large rRNAs reveal some eucaryote-specific elements of secondary structure. Biochimie 69:11–23

Michot B, Qu LH, Bachellerie JP (1990) Evolution of large subunit rRNA structure. The diversification of divergent domain D3 among major phylogenetic groups. Eur J Biochem 188:219–229

Mitchell P, Osswald M, Schueler D, Brimacombe R (1990) Selective isolation and detailed analysis of intra-RNA cross-links induced in the large ribosomal subunit of *E. coli*: a model for the tertiary structure of the tRNA binding domain in 23S RNA. Nucleic Acids Res 18: 4325–4333

Ophel K, Kerr A (1990) *Agrobacterium vitis* sp. nov. for strains of *Agrobacterium* biovar 3 from grapevines. Int J Syst Bact 40: 236–241

Otten L, De Ruffray P (1994) *Agrobacterium vitis* nopaline Ti plasmid pTiAB4: relationship to other Ti plasmids and T-DNA analysis. Mol Gen Genet 245:493–505

Otten L, van Nuenen M (1993) Natural instability of octopine/cucumopine Ti plasmids of clonal origin. Microbial Releases 2:91–96

Otten L, Canaday J, Gérard JC, Fournier P, Crouzet P, Paulus P (1992) Evolution of Agrobacteria and their Ti plasmids – A review. Mol Plant-Microbe Interact 5:279–287

Pace NR, Burgin AB (1990) Processing and evolution of the tRNAs. In: Hill WE, Dahlberg A, Garret RA, Moore PB, Schlessinger D, Warner JR (eds) The ribosome: structure, function and evolution. American Soc Microbiol, Washington DC, pp 417–425

Paulus F, Ridé M, Otten L (1989) Distribution of two *Agrobacterium tumefaciens* insertion elements in natural isolates: evidence for stable association between Ti plasmids and their bacterial hosts. Mol Gen Genet 219:145–152

Perochon-Dorisse J, Chetouani F, Aurel S, Iscolo N, Michot B (1995) RNA_d2: a computer program for editing and display of RNA secondary structures. Comp Appl Biosci 11:101–109

Raue HA, Muster W, Rutgers CA, Riet JV, Planta RJ (1990) rRNA: from structure to function. In: Hill WE, Dahlberg A, Garret RA, Moore PB, Schlessinger D, Warner JR (eds) The ribosome: structure, function and evolution. Amer Soc Microbiol, Washington DC, pp 217–235

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Sechi LA, Daneo-Moore L (1993) Characterization of intergenic spacers in two *rrn* operons of *Enterococcus hirae* ATCC9790. J Bacteriol 175:3213–3219

Springer N, Ludwig W, Hardarson G (1993) A 23S rRNA targeted specific hybridization probe for *Bradyrhizobium japonicum*. Syst Appl Microbiol 16:468–470

Sweeney R, Chen L, Yao MC (1994) An rRNA region has an evolutionary conserved essential role despite sequence divergence. Mol Cell Biol 14:4203–4215

Szegedi E (1985) Host range and specific L(+)-tartrate utilization of biotype 3 *Agrobacterium tumefaciens*. Acta Phytopath Acad Sci Hungary 20:17–22

Vander Horn PB, Backstrom AD, Stewart V, Begley TP (1993) Structural genes for thiamine biosynthetic enzymes (*thiCEFGH*) in *Escherichia coli* K-12. J Bact 175:982–992

Van Nuenen M, De Ruffray P, Otten L (1993) Rapid divergence of *Agrobacterium vitis* octopine/cucumopine Ti plasmids from a recent common ancestor. Mol Gen Genet 240:49–57

Vester B, Garrett RA (1984) Structure of a protein L23-RNA complex located at the A-site domain of the ribosomal peptidyltransferase centre. J Mol Biol 179:431–452

Willems A, Collins MD (1993) Phylogenetic analysis of Rhizobia and Agrobacteria based on 16S rRNA gene sequences. Int J Syst Bact 43:305–313

Xing Y, Draper DE (1995) Stabilization of a ribosomal RNA tertiary structure by ribosomal protein L11. J Mol Biol 249:319–331

Yanagi M, Yamasato K (1993) Phylogenetic analysis of the family *Rhizobiaceae* and related bacteria by sequencing of 16S rRNA gene using PCR and DNA sequencer. FEMS Microbiol Lett 107:115–120

Young JPW, Downer HL, Eardley BD (1991) Phylogency of the phototrophic Rhizobium strain BTAil by polymerase chain reaction-based sequencing of a 16S rRNA gene segment. J Bacteriol 173:2271–2277

Zimmerman RA, Thomas CL, Wower J (1990) Structure and function of rRNA in the decoding domain and at the peptidyltransferase center. In: Hill WE, Dahlberg A, Garret RA, Moore PB, Schlessinger D, Warner JR (eds) The ribosome: structure, function and evolution. Amer Soc Microbiol, Washington DC, pp 331–347

Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res 9:133–148