

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* K.309 type strain sequence by only one nucleotide, in spite of the fact that S4 and K.309 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



Fonds Documentaire IRD
Cote : B * 23983 Ex : 1

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

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 *Hind*III 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 *Hind*III 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 map for *B. burgdorferi* (Kawachi

digested with *AvaI* in combination with several other enzymes. Southern analysis with the 3.8 kb *AvaI* fragment from pPM1151 as probe revealed four *AvaI* fragments of 3.0, 3.8, 4.45 and 4.5 kb. Only the 3.8 kb fragment was cut by *BglII* and *PstI*. *BglII* and *PstI* 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 *AvaI* and one of 17 other restriction enzymes. The 3.8- and 3.5-kb *AvaI* 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 *BglII* and *EcoRV* site, the 3' end between the *NotI* and *HindIII* 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	<i>A. vitis</i> S4	<i>B. japonicum</i>	<i>R. sphaeroides</i> (<i>rrnB</i>)
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

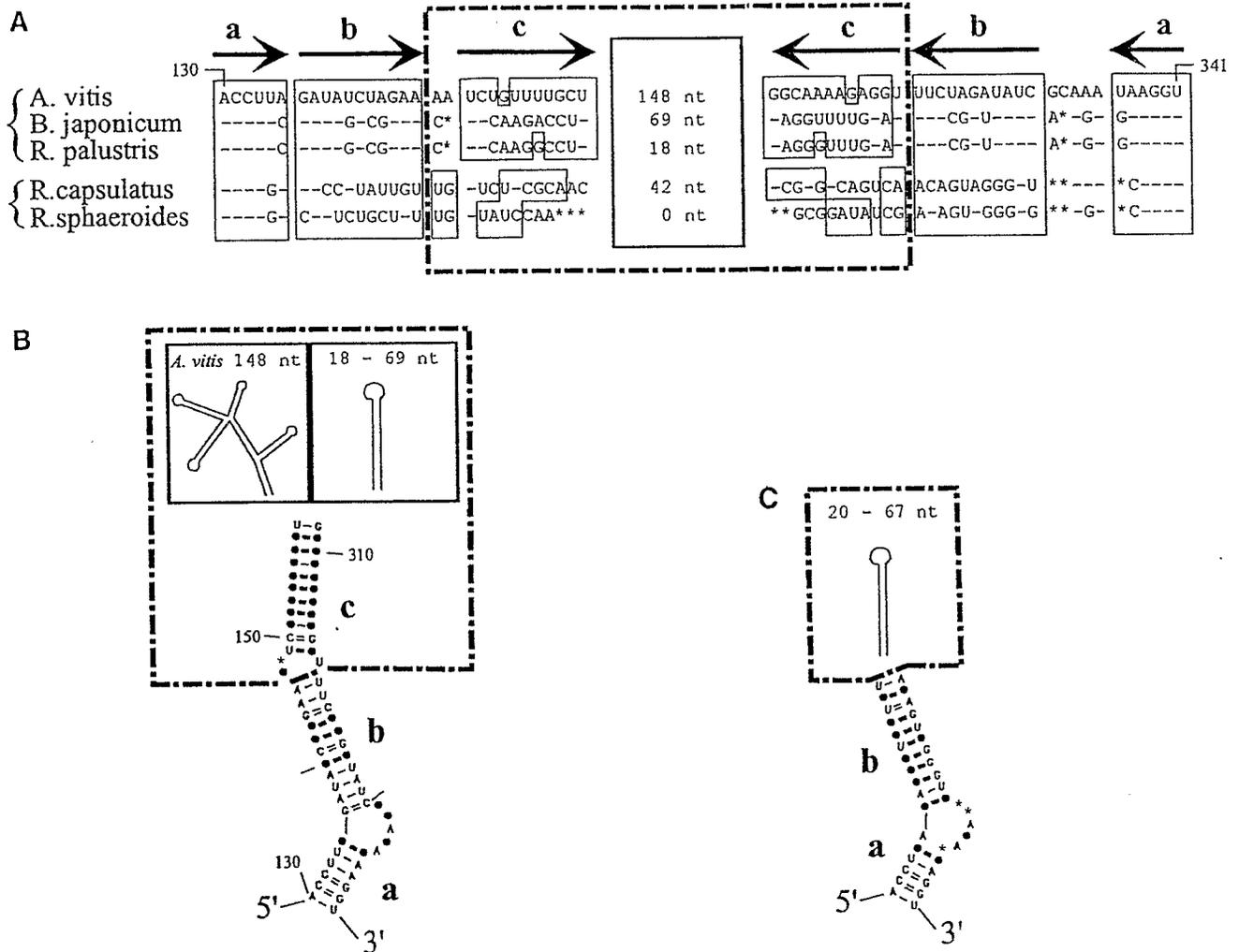
^b nh, similar feature but no sequence homology

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

differs markedly between *R. capsulatus* and *R. sphaeroides*. Within each subgroup (comprising *A. vitis*, *B. japonicum* and *R. palustris* and the two other Rhodospirillum 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 *NotI* site within the 5S gene and a *HindIII* site outside the repeat were used. *NotI-HindIII* fragments were cloned into the pKS Bluescript vector and *rrn* clones were identified by colony hybridisation with the



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 IS1222 of *Enterobacter agglomerans* (X78052, F. M. Lewecke, unpublished).

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 Col-

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

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

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

7101	TCCTGCCCC	GCAACCAATA	ACTTACA	¹ CAA	AATATCAATA	CGATAAGCCC	<i>rrnA</i>
	TCCTGCCCC	GCAACCAATA	ACTTACA	CAA	AATATCAATA	CGATAAGCCC	<i>rrnB</i>
	TCCTGCCCC	GCAACCAATA	ACTTACA	TCA	ATCCCGCAGC	CTTAAACGCT	<i>rrnC</i>
	TCCTGCCCC	GCAACCAATA	ACTTACA	TCA	ATCCCGCAGC	CTTAAACGCT	<i>rrnD</i>
7151	TC C CAAAGG	AGGGCTTTTG	GCGTTTCTGG	² GCTTCAGTTG	CCGCGAGTCT	<i>rrnA</i>	
	<u>TCCCAAAGG</u>	AGGGCTTTTG	GCGTTTCTGG	GCTTCAATCG	TCTTGACCTT	<i>rrnB</i>	
	<u>GCGGGATTTT</u>	<u>TGCGTTTCTG</u>	<u>GGCTTCGGTT</u>	GCGCCAACAC	GGCCCTCATC	<i>rrnC</i>	
	<u>GCGGGATTTT</u>	<u>TGCGTTTCTG</u>	GAATGAGATC	CTCCGGGCTT	TCAGCGCTGT	<i>rrnD</i>	
7201	TCCCATGAGG	TCTTCCAGGA	TCGCCTTTC	GTGAGGAAC	TCAGGCTTCT	<i>rrnA</i>	
	<u>GCCCCTGACC</u>	<u>TTGCCCCGCT</u>	GAACTAATCC	ATTTTGAAGT	AAGCTCTGGC	<i>rrnB</i>	
	ATAGCGCTTC	ACAATCCCCC	AGCCGTCAC	TCAACATAA	ATCCTGAACA	<i>rrnC</i>	
	TTATATGTTA	CAAGGAAGGG	TGCGCACAGT	TTCGACCCG	CGGCTTACTC	<i>rrnD</i>	

Therefore, the difference is mainly due to the interspersed class to a cross-link site to *rrnL2*, one of the main

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 IS1222-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

- 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, Haeblerli 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, Jurtschuk 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
- Guillou B, Gray MW, Schnare MN (1992) A compilation of large

- 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, Casadeu J, Gérard JC, Fournier P, Couzet P, Paulus P, 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 Haar RR, Beckstrom AD, Stewart V, Bagley TB (1992)

