Molecular Ecology (2001) **10**, 931–946

# Phylogenetic analyses of *Caulerpa taxifolia* (Chlorophyta) and of its associated bacterial microflora provide clues to the origin of the Mediterranean introduction

I. MEUSNIER,* J. L. OLSEN,† W. T. STAM,† C. DESTOMBE* and M. VALERO*
*Laboratoire de Génétique et Evolution des Populations Végétales, UPRESA CNRS 8016, Université de Lille 1, Bâtiment SN2, F-59655 Villeneuve d'Ascq cedex, France, †Department of Marine Biology, Centre for Ecological and Evolutionary Studies, University of Groningen Biological Centre, PO Box 14, 9750 AA Haren, The Netherlands

**Abstract**

**The accidental introduction of *Caulerpa taxifolia* into the Mediterranean is no longer under dispute. What has eluded researchers until now, is definitive evidence for the original, biogeographical source population. Here we present two independent lines of evidence that support an Australian origin for the Mediterranean populations of *C. taxifolia*. First, we reanalysed algal rDNA-internal transcribed spacer (rDNA-ITS) sequences, combining previously published sequences from different studies with 22 new sequences. The ITS sequence comparison showed that the Australian sample is the sister group of the Mediterranean–aquarium clade. Second, cloned bacterial 16S rDNA gene sequences were analysed from the associated microflora of *C. taxifolia* collected from Australia, Tahiti, the Philippines and the Mediterranean. Five bacterial lineages were identified, of which three were dominant. Alpha Proteobacteria were the most abundant and were found in all samples. In contrast, members of the beta Proteobacterial line and *Cytophaga-Flexibacter-Bacteroides* line (CFB) were mainly associated with Mediterranean and Australian samples. Frequency distributions of the five bacterial lineages were significantly different among biogeographical locations. Phylogenetic analyses of the 54 bacterial sequences derived from the four *C. taxifolia* individuals resulted in a well-resolved tree with high bootstrap support. The topologies of the beta Proteobacteria and CFB mirror the geographical sources of their algal hosts. Bacterial–algal associations provide an identification tool that may have wide application for the detection of marine invasions.**

*Keywords*: 16S, alga, algae-associated bacteria, *Caulerpa taxifolia*, invasive species, ITS

*Received 24 May 2000; revision received 30 October 2000; accepted 30 October 2000*

## Introduction

Introductions of nonindigenous species are occurring in virtually every part of the world, but until species become truly invasive, they are generally ignored. The Mediterranean Sea is a typical example. More than 30 species of nonendemic macrophytic algae have been documented since 1970 and relatively little has been heard about them except for the case of *Caulerpa taxifolia* (Vahl) C. Agardh. This tropical green alga was first discovered in the Mediterranean Sea, along the shores of Monaco in 1984 (Meinesz & Hesse 1991). Since then, it has spread rapidly and it is now distributed from the Balearic Islands to Croatia (Meinesz & Boudouresque 1996). The expansion of *C. taxifolia* has raised serious concerns about changes in biodiversity (de Villèle & Verlaque 1995; Bellan-Santini *et al*. 1996; Pesando *et al*. 1996) as well as concerns about general negative effects on tourism and local fisheries (Harmelin-Vivien *et al*. 1999).

The accidental introduction of *C. taxifolia*, via one or several aquaria, into the Mediterranean: 'aquarium escape hypothesis' (Meinesz & Boudouresque 1996), is no longer under dispute. It was alternatively proposed that this alga may have originated from *C. mexicana* Sonder ex Kützing that migrated from the Red Sea: 'current transport hypothesis'

Correspondence: M. Valero. Fax: +33-3-20-43-69-79; E-mail: myriam.valero@univ-lille1.fr

(Chisholm *et al*. 1995). Evidence from nuclear rDNA-internal transcribed spacer (rDNA-ITS) sequences has shown that *C. taxifolia* is not a cryptic form of *C. mexicana* (Olsen *et al*. 1998) and that the ITS sequences of Mediterranean strains are identical to those of a widely distributed aquarium variety (Jousson *et al*. 1998). What has eluded researchers until now, is definitive evidence for the original, biogeographical source population. In a recent survey of allozyme polymorphism, Benzie *et al*. (2000) found high genetic similarities between one Australian population of *C. taxifolia* from Stradebroke Island, near Brisbane, and western Mediterranean populations. Moreover, individuals from Stradebroke Island shared several morphological similarities with the Mediterranean strain, including heavy stolons, long fronds and large pinnules, as compared with the smaller and more delicate thalli characteristic of true tropical populations. This is important because individuals in the Mediterranean are characterized by gigantism, extremely fast growth rate and low temperature tolerance, which explains, at least in part, their superior competitive ability. In addition, they have also been shown to be capable of rhizoidal uptake of inorganic and organic nutrients from substratum (Chisholm *et al*. 1996). The chemical mechanisms involved are still unknown, but Chisholm *et al*. (1996) suggested the occurrence of alga–bacterial interactions.

*Caulerpa* has long been known to harbour endosymbiotic and epiphytic bacteria, which may be associated with various metabolic functions including nitrogen fixation and/or the production of various toxic compounds. Endosymbiotic bacteria are reported to be of common occurrence in other coenocytic green algae, such as *Bryopsis hypnoides* (Burr & West 1970), *Penicillus* sp. (Turner & Friedman 1974) and *C. prolifera* (Dawes & Lohr 1978). Moreover, ubiquitous epiphytic bacteria have been observed among intertidal seagrasses (Weidner *et al*. 1996) and on the upper thallus surfaces of many seaweeds (Provasoli & Pintner 1980; Shiba 1992). Plant-associated bacteria usually exchange signals with their hosts and possess the ability to colonize plant surfaces and tissues, to metabolize plant-derived carbon sources and to synthesize plant hormones (Preston *et al*. 1998). Seaweed–bacteria interactions in the field are poorly understood. However, effects of epiphytic bacteria on growth have been observed in different species of green algae in culture such as *Ulva* sp. (Provasoli 1958), *Enteromorpha compressa* and *E. linza* (Fries & Iwasaki 1976), and *Monostroma oxyspermum* (Tatewaki *et al*. 1993). In addition to endophytic bacteria, *C. taxifolia* also harbours numerous epiphytic bacteria (Chisholm *et al*. 1996).

It is not known whether the seaweed-associated bacteria are strictly host specific, or repeatedly acquired from the local environment. Given the high propensity for asexual reproduction in *C. taxifolia* via thallus fragmentation (Sant *et al*. 1996), it is highly probable that the association of micro-organisms and their *C. taxifolia* host is maintained, possibly on a more or less permanent basis. We therefore hypothesized that *C. taxifolia* has a permanently associated microflora that will maintain a biogeographical signature just like its host. If this is so, then the bacterial flora associated with *C. taxifolia* should provide an independent source of information for tracing the geographical origin of the host.

The aim of this study was to address the question of the geographical source of the Mediterranean *C. taxifolia* using two approaches. First, we (re)analysed 54 rDNA-ITS sequences, including 32 from previously published sources and 22 new ones, in order to obtain a more balanced data set with respect to taxon sampling. Second, we compared the assemblages of bacteria associated with *C. taxifolia* from four different origins: Australia, Tahiti, the Philippines and the western Mediterranean using 16S rDNA sequences.

## Materials and methods

### *Plant material*

Members of the genus *Caulerpa* are found in the subtidal (from 1 m to 10 m deep) and are mostly tropical to subtropical in their distribution. They are characterized by coenocytic, multinucleate thalli, a stoloniferous habit with upright fronds (= assimilators). In the case of *C. taxifolia*, the fronds are feather-like and range from 1 to > 50 cm in length. Most species are patchily distributed, whereas the invasive variety of *C. taxifolia* can form meadows.

Forty-eight isolates representing five species of *Caulerpa* were used in this study: 24 strains previously compared by Jousson *et al*. (1998), nine by Olsen *et al*. (1998), and 15 new ones (Table 1). In the field, three to five, 5-cm pieces of frond and stolon material were collected from individual thalli. New samples were rinsed with freshwater, blotted dry and immediately placed in silica gel, a standard method for DNA preservation that obviates the need for freezing.

The *C. taxifolia* isolates were collected from six geographical regions (Mediterranean basin, Western Polynesia, Japan, Australia, Malaysia and the Caribbean) and from six different aquaria. 16S rDNA sequencing of the microflora associated with *C. taxifolia* came from four individual 'hosts' from the Mediterranean, Tahiti, the Philippines and Australia (Table 1).

### *DNA extraction and purification*

Total DNA was extracted in 1.5-mL Eppendorf tubes following the method of Dellaporta as described in Epplen & Lubjuhn (1999). One to three blade apices of silica-dried material were gently crushed (approximately 10–15 mg) and added to 1 mL of extraction buffer [100 mM Tris-HCl pH 8.0, 50 mM ethylene diamine tetraacetic acid (EDTA), 500 mM NaCl, 2% PVP-40], 1 μL of β-mercaptoethanol and 100 μL of 20% (w/v) sodium dodecyl sulphate. The solution was

**Table 1** The *Caulerpa* species investigated, the EMBL/GenBank Accession no. of the ITS1-5.8S-ITS2 sequences, the geographical region and locality where the isolate was collected, the code used in the figures, and the number of cloned fragments included in the analysis

| Species | Accession no. | Geographical region | Locality | Code | No. of fragments |
|---|---|---|---|---|---|
| *C. mexicana* | AJ007815* | Mediterranean Basin | Haifa, Israel | mexisr | |
| *C. mexicana* | AJ007816* | Mediterranean Basin | Gran Canaria, Spain | mexcan | |
| *C. mexicana* | AJ007817* | Caribbean Sea | San Blas Island, Panama | mexpan | |
| *C. mexicana* | AJ007818* | Gulf of Mexico | Key Largo, USA | mexfla | |
| *C. mexicana* | AJ228992† | Red Sea | Djeddah, Saudi Arabia | mexdjed | |
| *C. mexicana* | AJ228993†, AJ228994† | Caribbean Sea | Martinique, France | mexmart | 2 (a, b) |
| *C. mexicana* | AJ228995†, AJ228996† | Mediterranean Basin | Stot-Yam, Israel | mexstot | 2 (a, b) |
| *C. prolifera* | AJ228988†, AJ228989† | Caribbean Sea | Martinique, France | promart | 2 (a, b) |
| *C. prolifera* | AF259571‡ | (Aquarium) | Nausicaa, France | pronausi | |
| *C. prolifera* | AF259578‡ | Mediterranean Basin | Golfe Juan, France | promed 1 | |
| *C. prolifera* | AF259570‡ | Mediterranean Basin | Golfe Juan, France | promed 5 | |
| *C. racemosa* | AJ228997† to AJ228999† | Mediterranean Basin | Marseille, France | racmars | 3 (a, b, c) |
| *C. racemosa* | AF259575‡, AF259576‡ | Mediterranean Basin | Capo Rizzuto, Italy | raccr | 2 (a, b) |
| *C. racemosa* | AF259573‡, AF259574‡ | Mediterranean Basin | Capo Miseno, Italy | raccm | 2 (a, b) |
| *C. racemosa* | AF259572‡ | Mediterranean Basin | Gran Canaria, Spain | raccana | |
| *C. sertularioides* | AJ228990†, AJ228991† | Caribbean Sea | Martinique, France | sermart | 2 (a, b) |
| *C. taxifolia* | AJ228963† | Mediterranean Basin | Messina, Italy | taxmess | |
| *C. taxifolia* | AJ007819* | Mediterranean Basin | Messina, Italy | taxmes | |
| *C. taxifolia* | AJ007821* | Mediterranean Basin | Messina, Italy | taxmex | |
| *C. taxifolia* | AJ007820* | Mediterranean Basin | Elba, Italy | taxelb | |
| *C. taxifolia*§ | AJ007823* | Australia | Townsville, Australia | taxaus | |
| *C. taxifolia* | AJ228960† | Mediterranean Basin | St Cyprien, France | taxcyp | |
| *C. taxifolia* | AJ228961† | Mediterranean Basin | Toulon, France | taxtou | |
| *C. taxifolia* | AJ228962† | Mediterranean Basin | Le Lavandou, France | taxlava | |
| *C. taxifolia* | AJ228964† | Mediterranean Basin | Port-Cros, France | taxpc | |
| *C. taxifolia* | AJ228966† | Mediterranean Basin | Hvar Island, Croatia | taxhva | |
| *C. taxifolia* | AJ228967† | Mediterranean Basin | Krk Island, Croatia | taxkrk | |
| *C. taxifolia* | AJ228968† | Mediterranean Basin | Villefranche s/mer, France | taxvill | |
| *C. taxifolia* | AJ228969† to AJ228971† | Mediterranean Basin | Le Brusc, France | taxbru | 3 (a, b, c) |
| *C. taxifolia* | AJ007822* | (Aquarium) | Monaco, France | taxmon aq | |
| *C. taxifolia* | AJ228972† | (Aquarium) | Enoshima, Japan | taxjap aq | |
| *C. taxifolia* | AJ228973† | (Aquarium) | Hawaii, USA | taxoah aq | |
| *C. taxifolia* | AJ228974† | (Aquarium) | Nancy, France | taxnan aq | |
| *C. taxifolia* | AJ228975† | (Aquarium) | Aquarium shop in Geneva, Switzerland | taxgen aq | |
| *C. taxifolia* | AJ228976†, AJ228977† | (Aquarium) | Stuttgart, Germany | taxstu aq | 2 (a, b) |
| *C. taxifolia* | AJ228978† to AJ228980† | Caribbean Sea | Guadeloupe, France | taxgua | 3 (a, b, c) |
| *C. taxifolia* | AJ228981† to AJ228983† | Caribbean Sea | Martinique, France | taxmart | 3 (a, b, c) |
| *C. taxifolia* | AJ228984† to AJ228987† | Japan | Ryu-Kyu islands, Japan | taxjap | 4 (a, b, c, d) |
| *C. taxifolia*§ | AF259589‡ | Malaysia | Bolinao, Philippines | taxphi | |
| *C. taxifolia*§ | AF259579‡ | Polynesia | Tahiti, France | taxtah | |
| *C. taxifolia* | AF259586‡, AF259587‡ | Mediterranean Basin | Stari Grad Bay, Croatia | taxcro | 2 (a, b) |
| *C. taxifolia*§ | AF259588‡ | Mediterranean Basin | Port Alassio, Italy | taxmed | |
| *C. taxifolia* | AJ228965† | Mediterranean Basin | Majorca, Spain | taxbal | |
| *C. taxifolia* | AF259577‡, AF259569‡ | Mediterranean Basin | Majorca, Spain | taxbal 43 | 2 (a, b) |
| *C. taxifolia* | AF259590‡, AF259580‡ to AF259582‡ | Mediterranean Basin | Majorca, Spain | taxbal 46 | 4 (a, b, c, d) |
| *C. taxifolia* | AF259583‡ | Mediterranean Basin | Majorca, Spain | taxbal 47 | |
| *C. taxifolia* | AF259584‡ | Mediterranean Basin | Majorca, Spain | taxbal 62 | |
| *C. taxifolia* | AF259585‡ | Mediterranean Basin | Majorca, Spain | taxbal 67 | |

*Sequences from Olsen *et al*. (1998). †Sequences from Jousson *et al*. (1998). ‡This study. §The four samples used for microflora analyses.

mixed and incubated for 30 min at 65 °C. After addition of 400 µL of 5 M K-acetate, the mixture was placed on ice for 1 h and centrifuged at 10 000 *g* for 20 min at 4 °C. The supernatant was filtered through Miracloth (Calbiochem®) and the DNA precipitated with 450 µL of cold isopropanol at –20 °C for 1 h. DNA was centrifuged at 12 000 *g* for 30 min at 4 °C and the pellet dissolved in 300 µL of high TE (10 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, pH 8.0). After 30 min at room

temperature, the tube was placed in a water bath at 65 °C for 15 min. The DNA was then precipitated a second time, pelleted, washed and finally redissolved in 50 μL of 0.1 × TE.

### *Polymerase chain reaction (PCR) amplification*

The ITS region was amplified as a single fragment (approximately 520 bases long) using two *Caulerpa*-specific M13-tailed primers (Steffens *et al*. 1993; Roy *et al*. 1996) located near the 3′ end of the 18S rDNA (5′-CACGACGTTGTAAAACGACGTAGGTGAACCTGCGAAAGG-3′) and the 3′ end of the ITS2 (5′-GGATAACAATTTCACACAGGGGCATTCAATCTCAGCGATT-3′). PCR amplifications were performed in a total volume of 25 μL according to the following reaction profile: 40 cycles of denaturation at 96 °C for 45 s, primer annealing at 55 °C for 45 s and extension at 72 °C for 1 min 30 s. Amplifications were run on a Perkin-Elmer Gene-Amp system 9600 or 9700 thermocycler.

The bacterial 16S region was amplified as a single fragment (approximately 915 bases long) using universal primers (J. R. Manhart, personal communication), '315-Forward' (5′-GCCACACTGGGACTGAGACA-3′) at approximately position 315 in the 16S rDNA and '1230-Reverse' (5′-ACGTCATCCTCACCTTCCTC-3′) at approximately position 1230 in the 16S rDNA. PCR amplifications were performed in a total volume of 25 μL according to the following reaction profile: 35 cycles of denaturation at 96 °C for 45 s, followed by primer annealing at 58 °C for 45 s and extension at 72 °C for 2 min. Amplifications were run on a Perkin-Elmer Gene-Amp 9600 or 9700 thermocycler. The chloroplast 16S rDNA of *Caulerpa* contains three introns totalling 3 kb (Manhart 1995). It was therefore possible to discriminate the bacterial sequences from those of the *Caulerpa* chloroplast.

### *Cloning, restriction fragment length polymorphism (RFLP) screening and sequencing*

Four PCR reactions for ITS were first pooled and then purified using a QIAGEN® kit (QIAquick™ PCR purification kit protocol). Purified PCR fragments were cloned using pGEM®-T Vector Systems (Promega). PCR products from the 16S amplifications were also cloned into pGEM®-T Vector Systems (Promega). See below for further details on the characterization and selection of cloned 16S fragments. The cloned fragments were sequenced in both directions using forward and reverse universal, fluorescence-labelled, plasmid primers (IRD 700 for M13R and IRD 800 for M13F) on a 4200L automated DNA sequencer (LiCor).

The PCR products generated with the universal bacterial 16S primers from *Caulerpa* 'total DNA' contained a mixture of 16S rDNA sequences derived from the genomes of the different micro-organisms associated with *C. taxifolia*. As described above, the PCR fragments were first cloned. Insert sizes were then checked by performing PCR with the universal forward and reverse primers.

As a first characterization of the microflora associated with each of the four hosts, approximately 50 clones per host individual of *C. taxifolia* were digested with a series of restriction endonucleases (*Msp*I, *Hae*III, *Cfo*I and *Alu*I). *Alu*I (Boehringer) was found to be the most informative and was chosen for the final RFLP analysis. Digestions were performed using 5 μL of PCR product in a total volume of 10 μL, following the manufacturer's protocol. Electrophoresis was performed using 2% (w/v) agarose gels (agarose brand) and TBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA). Gels were run at 100 V for 4.5 h and stained with 0.25 μg/mL ethidium bromide. Clones were classified according to their different RFLP patterns. Between one and six clones were then sequenced from each of the five or so most frequent restriction patterns. Bacteria were identified according to the results of the sequence similarity searches using BLAST (http://www2.ncbi.nlm.nih.gov/BLAST/; Altschul *et al*. 1997). A known pitfall of PCR-based rRNA analysis is the formation of chimeras between two different DNA molecules which may be generated during the annealing step of the PCR process (Wintzingerode *et al*. 1997). To determine if clones were in fact chimeric molecules, we divided each clone's sequence into three equal segments. BLAST searches were performed on each of the three segments separately. The BLAST results from these three segments were then compared with the results from the entire intact sequence. Using this method, it was discovered that six of the 60 sequenced clones were, in fact, chimeric molecules. These were subsequently excluded from the analyses.

For taxonomic assignments the threshold values given by Mullins *et al*. (1995) and Pinhassi *et al*. (1997) were used. Two bacterial sequences showing a similarity value > 0.97 suggests that both sequences belong to the same species and a range of 0.93–0.97 indicates that both sequences correspond to the same genus.

### *Sequence alignment*

Sequences were managed and aligned using BIOEDIT version 4.7.8 software (http://www.mbio.ncsv.edu/RNaseP/info/programs/BIOEDIT/bioedit.html; Hall 1999). The secondary structure of the small subunit was determined by comparison with published prokaryotic, secondary-structure alignments (http://rrna.uia.ac.be/ssu/; Van de Peer *et al*. 2000). All sequences and their alignments have been deposited in GenBank (Table 1).

### *Phylogenetic analyses*

Phylogenetic reconstruction was performed using PAUP version 4.0b 3 (Swofford 1999) and PHYLIP version 3.5c

(Felsenstein 1989). Analyses were performed using maximum parsimony (MP), neighbour-joining (NJ) and/or maximum likelihood (ML) algorithms depending on the data set. Bootstrap resampling was performed using 1000 replicates.

The algal ITS analyses were performed using different subsets of taxa and/or sequences in order to assess the effects of taxon sampling and to reduce the number of trees when sequences were identical or nearly identical. MP analyses were performed with gaps coded as 'missing data' or 'new state'. For the global analysis involving all sequences, only a heuristic search [stepwise addition and tree bisection–reconnection (TBR) swapping; Swofford & Beagle 1993] was possible. For reduced data sets that only involved *C. taxifolia* sequences, the exhaustive search option under MP was applied. For the ML analysis of the ITS sequences, the transition/transversion (ti/tv) ratio was set at 2 and empirical base frequencies were used. NJ analysis was performed under the Kimura 2-parameter model (Kimura 1980).

The bacterial 16S analyses were performed under both MP (heuristic search option only) and NJ (Kimura 2-parameter) (Kimura 1980).

## Results

### *Analysis of ITS sequences in* Caulerpa taxifolia *and related species*

The new alignment consisted of 71 sequences of which 22 are new (Table 1). Four different partitions of the data set were analysed (Table 2); two involving interspecific analyses of the full and a reduced data set; and two that focused on the intraspecific level within *C. taxifolia*. Although levels of sequence divergence ranged from 7 to 18% among species, levels of ITS divergence within *C. taxifolia* ranged from 0.2 to 3.7%, which is quite low. For this reason, every effort was made to assess critically the effects of different taxon samplings and partitions of the alignment under different tree-building algorithms (including MP analyses with gaps coded as 'missing data' or 'new state'). The critical question of interest, is whether specimens of *C. taxifolia* from the Mediterranean and aquarium strains are more closely related to specimens from Australia than to specimens collected from other tropical origins, i.e. the Philippines, Tahiti and the Caribbean.

The first analysis (Table 2A) used the entire data set with gaps coded as 'missing data'. Because many sequences were identical or nearly identical, the number of most-parsimonious trees (MPTs) was extremely large (> 15 000). However, the topology of the consensus tree (Fig. 1) and the NJ tree (not shown) were identical with respect to species relationships but not with respect to biogeographical relationships within *C. taxifolia*. The key difference was that MP was unable to resolve the trichotomy among the Mediterranean/aquarium clade, the Australian sample and the remaining tropical samples (Fig. 1); whereas the NJ tree grouped the Australian sample with the tropical clade with 84% bootstrap support.

In the second analysis, the data set was reduced to 19 sequences (11 for *C. taxifolia* and eight for the other *Caulerpa* species) with gaps coded as 'missing data' (Table 2B). The criteria for the selection of the 11 *C. taxifolia* sequences were: (1) biogeographical representation; and (2) removal of duplicate and/or (almost) identical sequences from the Mediterranean data set. The second data set included two sequences from each of the other four *Caulerpa* species. MP analysis joined the Australian sample with the Tropical clade (95% bootstrap support), as did the NJ and ML analyses (trees not shown).

In the third analysis (Table 2C), the data set was further reduced to the 11 sequences of *C. taxifolia*. The established sister clade (taxjap) was retained as the outgroup. Note that taxjap is a misidentification (i.e. not *C. taxifolia*), but we retained the code here so as not to confuse the samples with previously published work. In this analysis we continued to code gaps as 'missing data'. None of the analyses were able to establish unequivocally the Australian link with the Mediterranean/aquarium clade based on bootstrap support (trees not shown).

In the fourth and final analysis (Table 2D, Fig. 2), the 11 sequence data set was (re)analysed but this time with gaps coded as 'new state'. Because neither NJ nor ML utilizes gaps, only an MP analysis was performed. The result was eight MPTs with very high bootstrap support (Fig. 2A). Here the Australian link to the Mediterranean/aquarium clade was 100% supported. Figure 3 shows the relevant part of the ITS alignment involving the insertion–deletion events that led to the grouping of Australia with the Mediterranean/aquarium clade. Regardless of whether the gaps are counted as 36 individual events or are recoded as five, single larger events, the result does not change.

### *Bacterial diversity within* C. taxifolia *and distribution among geographical regions*

The conditions used for PCR were successful in amplifying DNA fragments of less than 1000 bp that corresponded to the targeted region within the bacterial 16S genes. As a first approach, the diversity of the microflora associated with one isolate of *C. taxifolia* for each biogeographical region was assessed using an RFLP analysis of 229 clones (about 50 per isolate) as described in Materials and methods. The frequency distribution of the different restriction profiles is shown in Fig. 4.

A total of 95 different restriction profiles was observed, yielding an average of 25 different patterns per host. Sixty clones were chosen for sequencing from among the 28 most

**Table 2** Reanalysis of the internal transcribed spacer (ITS) sequences in *Caulerpa* listed in Table 1. Four partitions of the alignment (A–D) were analysed under different assumption sets

| | | | |
|---|---|---|---|
| (A) Interspecific analysis (complete set of sequences, *n* = 71, gap = 'missing data') | | | |
| Length of the alignment (bp) | 694 | | |
| Number of invariable positions (bp) | 455 | | |
| Number of variable positions (bp) | 239 | | |
| Number of informative positions (bp) | 178 | | |
| Analysis methods | MPh | NJ | ML |
| Number of trees | > 15 000 | 1 | nt |
| Number of bootstrap replicates | nt | 1000 | nt |
| Subset grouping (frequency or bootstrap support) | | | |
| ((mex clade),((pro clade),(ser clade))),(tax clade) | Always | (Polytomy) | nt |
| (mex clade),(tax clade) | Never | 0% | nt |
| ((taxAus),(taxMed/taxAqua)) | (Polytomy) | 9% | nt |
| ((taxAus),(taxTropical)) | (Polytomy) | 84% | nt |
| (taxMed),(taxAqua) | Always | 100% | nt |
| | Fig. 1 | | |
| (B) Interspecific analysis (reduced set of sequences, *n* = 19, gap = 'missing data') | | | |
| Length of the alignment (bp) | 694 | | |
| Number of invariable positions (bp) | 487 | | |
| Number of variable positions (bp) | 207 | | |
| Number of informative positions (bp) | 151 | | |
| Analysis methods | MPh | NJ | ML |
| Number of trees | 75 | 1 | 1 |
| Number of bootstrap replicates | 1000 | 10 000 | 100 |
| Subset grouping (bootstrap support) | | | |
| ((mex clade),((pro clade),(ser clade))),(tax clade) | (Polytomy) | (Polytomy) | (Polytomy) |
| ((mex clade),(tax clade)) | 0% | 0% | 0% |
| ((taxAus),(taxMed)) | < 5% | 5% | 12% |
| ((taxAus),(taxTropical)) | 95% | 57% | 84% |
| | Trees not shown | | |
| (C) Intraspecific analysis (*C. taxifolia*, *n* = 11, gap = 'missing data') | | | |
| Length of the alignment (bp) | 694 | | |
| Number of invariable positions (bp) | 609 | | |
| Number of variable positions (bp) | 85 | | |
| Number of informative positions (bp) | 58 | | |
| Analysis methods | MPe | NJ | ML |
| Number of trees | 75 | 1 | 1 |
| Number of bootstrap replicates | 1000 | 10 000 | 100 |
| Subset grouping (bootstrap support) | | | |
| ((taxJap),(all other tax)) | 100% | 100% | 100% |
| ((taxAus),(taxMed)) | 12% | 12% | 12% |
| ((taxAus),(taxTropical)) | 79% | 58% | 80% |
| | Trees not shown | | |
| (D) Intraspecific analysis (*C. taxifolia*, *n* = 11, gap = 'new state') | | | |
| Length of the alignment (bp) | 694 | | |
| Number of invariable positions (bp) | 517 | | |
| Number of variable positions (bp) | 177 | | |
| Number of informative positions (bp) | 128 | | |
| Analysis methods | MPe | NJ | ML |
| Number of trees | 8 | na | na |
| Number of bootstrap replicates | 1000 | — | — |
| Subset grouping (bootstrap support) | | | |
| ((taxJap),(other tax)) | 100% | — | — |
| ((taxAus),(taxMed)) | 100% | — | — |
| ((taxAus),(taxTropical)) | 0% | — | — |

*n*, number of sequences analysed; MPh, maximum parsimony under heuristic search; MPe, maximum parsimony under exhaustive search; NJ, neighbour-joining; ML, maximum likelihood; nt, not tested; mex, *C. mexicana* isolates; pro, *C. prolifera* isolates; ser, *C. sertularoides* isolates; tax, *C. taxifolia* isolates; taxMed, *C. taxifolia* isolates from the Mediterranean; taxAqua, *C. taxifolia* isolates from aquariums; taxAus, *C. taxifolia* isolate from Australia; taxTropical, *C. taxifolia* isolates from the Philippines, Tahiti, Martinique and Guadeloupe; taxJap, *C. taxifolia* isolates from Japan; na, not applicable.

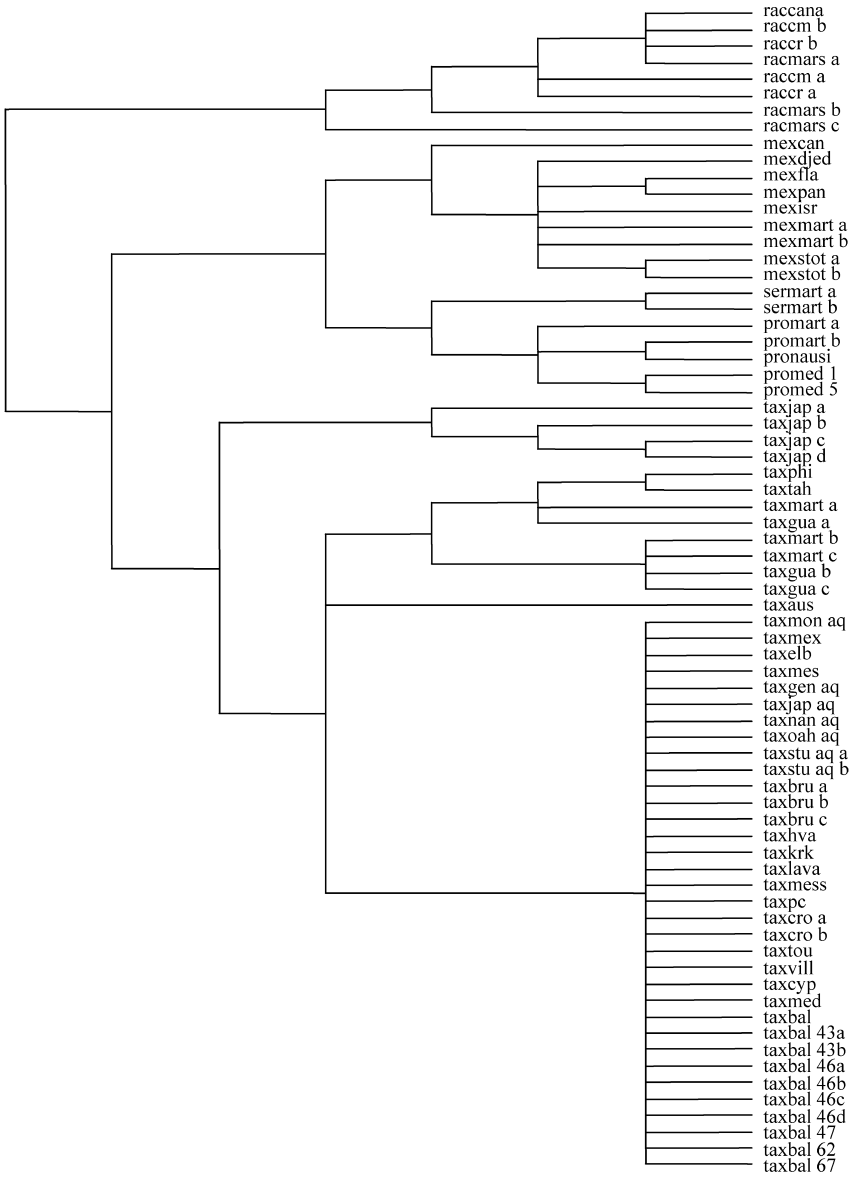


**Fig. 1** Interspecific phylogenetic analysis based on internal transcribed spacer (ITS) sequences from *Caulerpa* isolates using the full data set. Strict consensus tree of > 15 000 most-parsimonious trees (the codes used for each strain are given in Table 1 and details of the phylogenetic analysis in Table 2A).

frequent RFLP patterns shown in Fig. 4. These sequences were subsequently identified using a BLAST search (Table 3). Because in most cases more than one clone per RFLP pattern was sequenced, it was possible to assess the accuracy of the RFLP method in identifying bacterial lineages. In 75% of the cases, a given RFLP pattern matched a single bacterial lineage which corresponded with the single colour bars given in Fig. 4.

Ninety-six per cent (52/54) of the sequences matched bacterial sequences. The remaining 4% (2/54) matched plant or algal chloroplast sequences (Table 3). The comparison with previously characterized bacteria in the EMBL database revealed only one clone with a sequence similarity value > 0.97, 28 clones ranging from 0.93 to 0.97, and 23 clones < 0.93. Sequences were distributed across five main lineages of the domain bacteria: the *Cytophaga-Flexibacter-Bacteroides* group (CFB group; 10 sequences), and the alpha (26 sequences), beta (four sequences), gamma (three sequences) and delta (seven sequences) subdivisions of Proteobacteria. In addition, one sequence was determined as a planctomycete and another as a cyanobacterium. The frequency distribution of the five bacterial lineages for each geographical host is shown in the pie charts of Fig. 4. The distributions of bacterial diversity were statistically different among host origins, but only marginally so. The unbiased estimate of the exact *P*-value of the Fisher exact test was $P = 0.045$, standard error (SE) = 0.0005, for 5 000 000 iterations (STRUCT program; Raymond & Rousset 1995). However, even if the differences between patterns were significant, the qualitatively observed diversity pattern in the

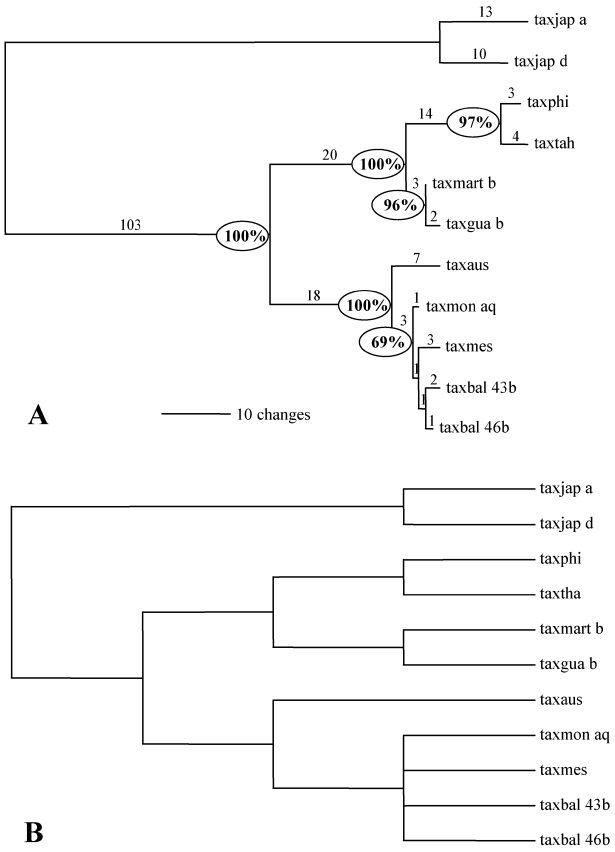


**Fig. 2** Intraspecific phylogenetic analysis based on internal transcribed spacer (ITS) sequences from *Caulerpa taxifolia* populations using a reduced data set (the codes used for each strain are given in Table 1 and details of the phylogenetic analysis in Table 2D). (A) One of the eight most-parsimonious trees (MPTs). Length = 209, consistency index = 0.952, retention index = 0.960, rescaled consistency index = 0.914. Bootstraps (1000 replicates) are circled. (B) Strict consensus tree of the eight MPTs.

Australian host was more similar to the Mediterranean host than to hosts from other tropical origins. Nevertheless, this similarity should be interpreted with caution because bacteria sequences could be very different within lineages.

### *Phylogenetic analyses of bacterial 16S sequences*

Sequences from the 54 clones were easily alignable and include consideration of the secondary structure. Both MP and NJ analyses maintained the topology of the major clades of interest: the CFB, and the alpha, beta and gamma Proteobacteria clades remained monophyletic. Arrangements involving the delta group and internal rearrangements within the alpha group accounted for other topologies (trees not shown). We have chosen the NJ tree for our discussion (Fig. 5). Except for the alpha Proteobacteria, differences among clades were stable and well supported by high bootstrap values. Members of the delta group were paraphyletic in all analyses, which can be explained by the high taxonomic diversity within this functional group (see Table 3) as compared with the others.

*CFB clade.* Within the CFB clade (Fig. 5), the 10 sequences weakly matched species belonging to the *Flavobacteriaceae* (sequence similarities ranging from 0.85 to 0.90; Table 3). Sequences were clustered according to their host origins. Bacteria associated with the Australian (three sequences) and Mediterranean (six sequences) hosts were sister groups (sequence similarities within group > 0.99 for Mediterranean and 0.96 for Australia and between the two sister groups > 0.93). Therefore, the pattern observed within the *Flavobacteriaceae* group revealed a significant divergence among bacteria according to their host origins, which indicates that each host was colonized by a distinct bacterial species (belonging to one and the same unknown genus for the Mediterranean and Australian sister groups).

*Alpha Proteobacteria clade.* Most of the 16S sequences obtained were classified within the alpha Proteobacteria group (Fig. 5). Eighty-four per cent (22/26) were assignable to the *Rhodobacter* group, which is known to be endosymbiotic with the potential for nitrogen fixation (Table 3). Within the *Rhodobacter* group itself, most of the Mediterranean sequences (5/7) were related to the genus *Ruegeria* (C. taxMed-2, -21, -39, -54 and -66) in addition with two clones isolated from the Australian host (C. taxAus-6 and -25) and one from the Tahitian host (C. taxTah-30). Three clones ascribed to this genus (C.Med-2, -54 and -66) together with the C. taxMed-9 clone formed a robust Mediterranean cluster (100% bootstrap support, Fig. 5), although the C. taxMed-9 clone was assigned to another genus (*Roseobacter*) on the basis of a sequence similarity of 0.92 with *Roseobacter*

```
taxjap a    CAAATTCACAAAGTG--CACACT-CT-----T-TATGT--TGT--GTATAAA-CTCCATGGC-TATG--GCTAT--------------GTCTAT---GTAATGTGATGATAGTGTTGTT----CTCTCTATAAAGCATG
taxjap d    CAAATTCACAAAGTG--CACACT-CT------CTATGT--TGTGTATAAAAAACT-CATGGCTTATG--GCTAT--------------GTCTAT---GTAAT--GATGATAGTGTTGGT---TCTCTTTATAAAGCATG
taxphi      CAAATTCTATATGTGTATACACTGCTGTGTGTCTATATTCTGTGTGTAAAGA----CATGGC-CATGTTGCTATTCAAAGAACATATG-CCTATGTTGTAATGAATGTGCTGTTGTTATTATTG----CCTAAAGCATG
taxtah      CAAATTCTATATGTGTATACACTGCTGTGTGTCTATATTCTGTGTGTAAAGA----CATGGC-TATGTTGCTATTCAAAGAACATATG-CCTATGTTGTAATGAATGTGCTGTTGTTATTATTG----CCTAAAGCATG
taxmart b   CAAATTCTATATGTGTATACACTGCTGTGTGTCTATATTCTTTGTGTAAAGA----CATGGC-TATGTTGCTATTCAAAGAACATATG-CCTATGTTGTAATGACTGTGTTGTTGTTATTATTG----TCTAAAGCATG
taxgua b    CAAATTCTATATGTGTATACACTGCTGTGTGTCTATATTCTTTGTGTAAAGA----CATGGC-TATGTTGCTATTCAAAGAACATATG-CCTATGTTGTAATGACTGTGTTGTTGTTATTATTG----TCTAAAGCATG
taxaus      CAAATTCTAT--GTGTATACACTACTGTGTGTCTATATTCTTTGTGTAAAGA----CATGGC-TATGTT----------------------------GTAATGAATGTGTTGTTGTT---ATCATTG----TCTAAAGCATG
taxmon aq   CAAATTCTAT--GTGTATACACTGCTGTGTGTCTATATTCTTTGTGTAAAGA----CATGGC-TATGTT----------------------------GTAATGAATGTGTTGTTGTT---ATTGTTG----TCTAAAGCATG
taxmes      CAAATTCTAT--GTGTATACACTGCTGTGTGTCTATATTCTTTGTGTAAAGA----CATGGC-TATGTT----------------------------GTAATGAATGTGTTGTTGTT---ATTGTTG----TCTAAAGCATG
taxbal 43b  CAAATTCTAT--GTGTGTACACTGCNGTGTGTCTATATTCTTTGTGTAAAGA----CATGGC-TATGTT----------------------------GTAATGAATGTGTTGTTGTT---ATTGTTG----TCTAAAGCATG
taxbal 46b  CAAATTCTAT--GTGTATACACTGCTGTGTGTCTATATTCTTTGTGTAAAGA----CATGGC-TATGTT----------------------------GTAATGAATGTGTTGTTGTT---ATTGTTG----TCTAAAGCATG
```

**Fig. 3** Details of the alignment of the internal transcribed spacer (ITS) regions between positions 64 and 210 among biogeographical isolates of *Caulerpa taxifolia*, showing the shared deletions between the Australian and Mediterranean isolates.

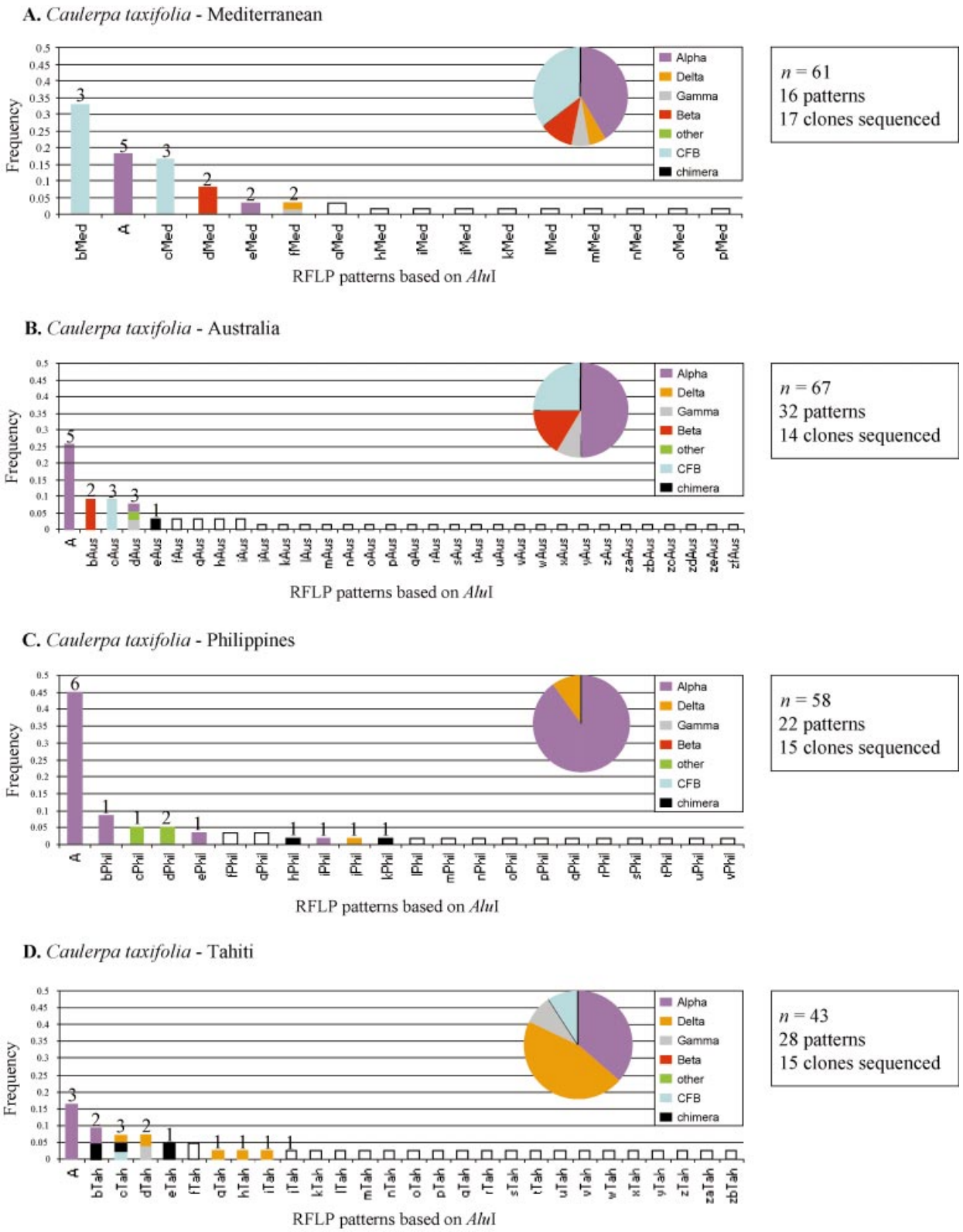


**Fig. 4** Frequency of restriction fragment length polymorphism (RFLP) patterns based on *Alu*I digests of approximately 50 cloned polymerase chain reaction (PCR) fragments of bacterial 16S genes per *Caulerpa taxifolia* individual. Each panel (A–D) represents a different geographical sample of *C. taxifolia*. Only one pattern is shared among origins (pattern 'A'). The most frequent patterns were chosen for sequencing. The numbers above the bars represent the number of clones sequenced from that category. The pie charts show the relative diversity of bacterial lineages associated with each biogeographical area. See text for further explanation.

**Table 3** Taxonomic affiliation of the 60 bacterial clones according to similarity searches with their 16S rDNA sequences (BLAST; Altschul *et al*. 1997). All clones listed belong to the bacteria domain, except the two chloroplast-associated clones at the bottom of the table. EMBL/GenBank Accession nos are included

| Accession no. | Clone code and number | Higher taxonomic ranks | Genus* | Accession no. of the closest matching organism (similarily values according to BLAST) |
|---|---|---|---|---|
| AF259601 | C. taxAus-7 | CFB group, Flavobacteriaceae | Undetermined | L39067 (0.85) |
| AF259602 | C. taxAus-22 | | | L39067 (0.85) |
| AF259600 | C. taxAus-4 | | | L39067 (0.85) |
| AF259613 | C. taxMed-1 | | Undetermined | M62797 (0.81)† |
| AF259614 | C. taxMed-5 | | | M62797 (0.81)† |
| AF259618 | C. taxMed-15 | | | M62797 (0.82)† |
| AF259615 | C. taxMed-18 | | | M62797 (0.81)† |
| AF259616 | C. taxMed-37 | | | M62797 (0.82)† |
| AF259617 | C. taxMed-46 | | | M62797 (0.81)† |
| AF259648 | C. taxTah-24 | | Undetermined | AF029041 (0.90) |
| AF259591 | C. taxAus-6 | Proteobacteria, alpha subdivision, *Rhodobacter* group | *Ruegeria* | D88523 (0.94) |
| AF259593 | C. taxAus-25 | | | D88523 (0.94) |
| AF259604 | C. taxMed-2 | | | D88523 (0.95) |
| AF259606 | C. taxMed-21 | | | D88523 (0.96) |
| AF259607 | C. taxMed-39 | | | X78315 (0.95) |
| AF259619 | C. taxMed-54 | | | D88523 (0.95) |
| AF259620 | C. taxMed-66 | | | D88523 (0.94) |
| AF259639 | C. taxTah-30 | | | X78315 (0.95) |
| AF259592 | C. taxAus-11 | | *Roseobacter* | AB018689 (0.95) |
| AF259603 | C. taxAus-27 | | | AB018689 (0.95) |
| AF259608 | C. taxMed-50 | | | AFO98494 (0.96) |
| AF259626 | C. taxPhil-27 | | | AF107210 (0.93) |
| AF259637 | C. taxTah-22 | | | AB013442 (0.94) |
| AF259605 | C. taxMed-9 | | Undetermined | AF107210 (0.92) |
| AF259628 | C. taxPhil-52 | | *Antarctobacter* | Y11552 (0.94) |
| AF259622 | C. taxPhil-8 | | *Rhodobacter* | AB017796 (0.93) |
| AF259627 | C. taxPhil-46 | | *Roseovarius* | Y11551 (0.95) |
| AF259625 | C. taxPhil-24 | | Undetermined | ABO25192 (0.91) |
| AF259621 | C. taxPhil-1 | | *Rhodovulum* | D16420 (0.94) |
| AF259635 | C. taxTah-4 | | Unclassified | AB018689 (0.95) |
| AF259595 | C. taxAus-40 | | | AF114484 (0.95) |
| AF259636 | C. taxTah-7 | | | AF114484 (0.96) |
| AF259624 | C. taxPhil-16 | | | AF114484 (0.95) |
| AF259623 | C. taxPhil-11 | Proteobacteria, alpha subdivision, *Erythrobacter* group | *Erythrobacter* | AB011075 (0.98) |
| AF259629 | C. taxPhil-57 | | | AF118020 (0.92) |
| AF259594 | C. taxAus-35 | Proteobacteria, alpha subdivision, unclassified group | Unclassified | AB025419 (0.96) |
| AF259596 | C. taxAus-2 | Proteobacteria, beta subdivision | *Herbaspirillum* | AF137508 (0.96) |
| AF259598 | C. taxAus-34 | | | AJ238359 (0.96) |
| AF259609 | C. taxMed-3 | | | AJ238359 (0.96) |
| AF259610 | C. taxMed-59 | | | AF137508 (0.96) |
| AF259599 | C. taxAus-33 | Proteobacteria, gamma subdivision | Undetermined | AF006606 (0.89) |
| AF259611 | C. taxMed-42 | | | AF165908 (0.90) |
| AF259641 | C. taxTah-3 | | | U77482 (0.90) |
| AF259645 | C. taxTah-27 | Proteobacteria, delta subdivision, Geobacteraceae | | U23141 (0.91) |
| AF259630 | C. taxPhil-2 | | | AF084850 (0.87)† |
| AF259612 | C. taxMed-34 | Proteobacteria, delta subdivision, Myxococcales | Unclassified | AJ241004 (0.94) |
| AF259643 | C. taxTah-14 | | Undetermined | AF146233 (0.92) |
| AF259644 | C. taxTah-23 | | | AF186411 (0.90) |
| AF259646 | C. taxTah-32 | | | AF186411 (0.91) |
| AF259647 | C. taxTah-42 | | | AJ241003 (0.89) |
| AF259633 | C. taxPhil-23 | Planctomycetales, Planctomycetaceae | Unclassified | AJ231181 (0.93) |
| AF259634 | C. taxPhil-25 | Cyanobacteria, Prochloraceae | Undetermined | X63141 (0.92) |
| | C. taxPhil-45 | Chloroplast, Stramenopiles, Phaeophyceae, Ectocarpales | Undetermined | X14873 (0.94)† |
| | C. taxAus-26 | Chloroplast, Embryophyta | | Z00044 (0.99) |

C. taxAus, bacteria associated with *Caulerpa taxifolia* isolates from Australia; C. taxMed, bacteria associated with *C. taxifolia* isolates from the Mediterranean; C. taxPhil, bacteria associated with *C. taxifolia* isolates from the Philippines; C. taxTah, bacteria associated with *C. taxifolia* isolates from Tahiti; CFB, *Cytophaga-Flexibacter-Bacteroides*.

*Undetermined genus: similarity index < 0.93; unclassified genus: matches an unclassified genus (similarity index > 0.93).†Similarity values using DNADIST version 3.5c software on BIOEDIT version 4.7.8 software.

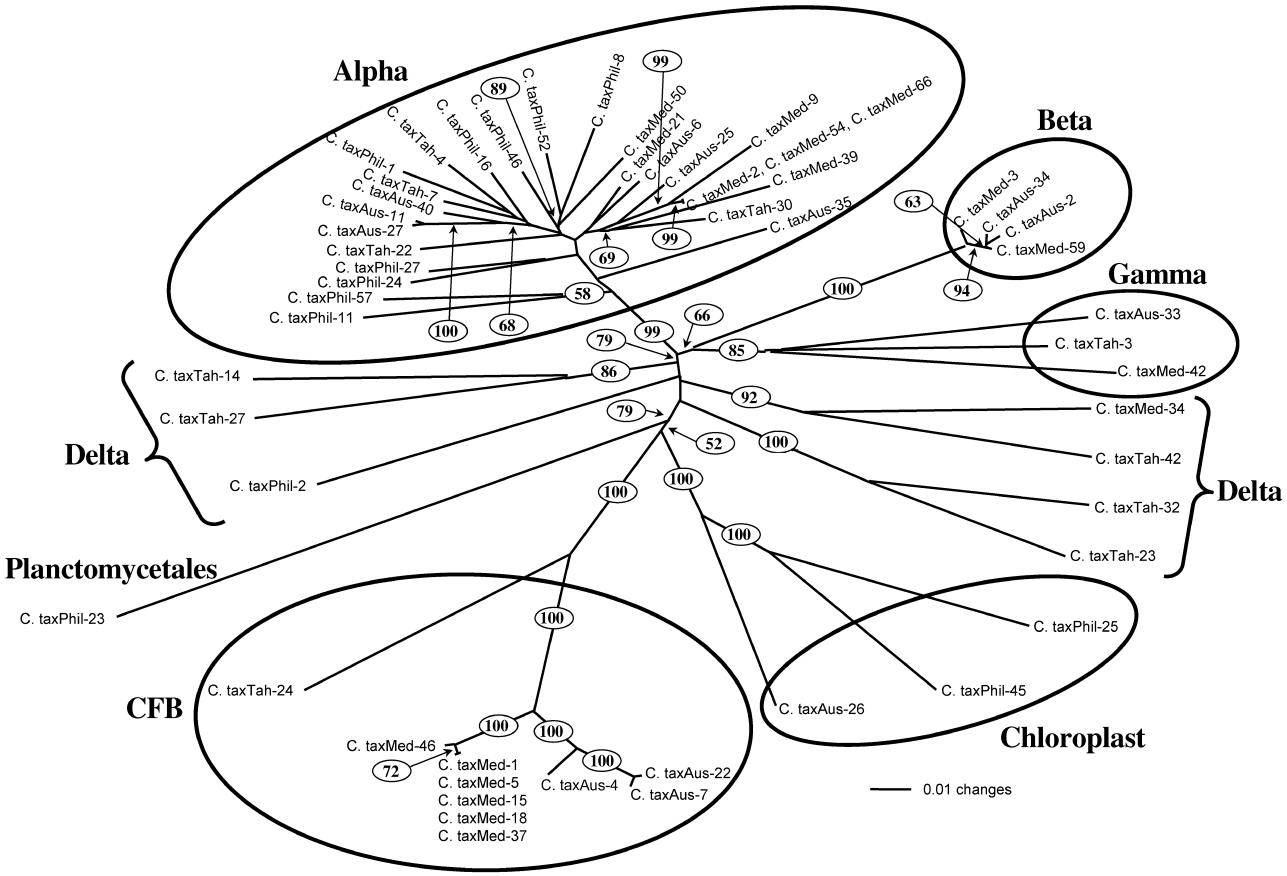


**Fig. 5** Neighbour-joining (NJ) tree based on 16S sequences from bacteria associated with the four *Caulerpa taxifolia* origins (see Table 3 for details of the taxonomic identification of the bacteria sequence). Bootstraps (1000 replicates) are circled.

sp. (Accession no. AF107210). The sister group of this Mediterranean cluster was a clone isolated from the Australian host (C. taxAus-25), but the branching was not supported by a high bootstrap value. Attempts to improve the resolution within the alpha Proteobacteria clade by adding database sequences of other alpha Proteobacteria members were not successful.

*Beta Proteobacteria clade.* Within the beta Proteobacteria clade (Fig. 5), all sequences (two from Australian and two from Mediterranean origin) formed a distinct and well-supported (100%) clade. The similarity among the four sequences was higher than 0.98, suggesting that the four sequences correspond to the same bacterial species although associated with two hosts from completely different biogeographical origin. The sequences matched best with the plant-associated bacteria belonging to the genus *Herbaspirillum* (Table 3).

*Gamma Proteobacteria clade.* The three sequences clustering within the gamma Proteobacteria clade (Fig. 5) showed relatively low levels of similarity (0.89–0.90) with species previously sequenced. The low similarity and lack of resolution within this clade made it impossible to infer any pattern related to the biogeographical origin of the host.

## Discussion

### *ITS analysis*

Increased taxon sampling from tropical areas has enhanced our ability to locate the biogeographical source of Mediterranean *Caulerpa taxifolia* as compared with previous genetic studies (Jousson *et al*. 1998; Olsen *et al*. 1998). Our studies show that the Japanese sample previously assigned to *C. taxifolia* is a misidentification and represents a different cryptic species. Consequently, the intraspecific sequence divergence of 7% reported by Jousson *et al*. (1998) is too high. When the Japanese sample is removed, divergence among the ITS sequences from the different origins of *C. taxifolia* ranges from 0.2 to 3.7% and involves very few changes (Fig. 2). While the basic conclusions of Jousson

*et al*. (1998) remain unchanged with respect to the identity between aquarium and Mediterranean strains, the lower overall sequence divergence among biogeographical isolates of *C. taxifolia* makes accurate tree construction more difficult even with the added taxon sampling. For this reason, we proceeded with a detailed stepwise analysis scheme in which various subsets of taxa, as well as the use of gaps, were analysed in different ways. Two major geographical groups were identified: one corresponding to an Atlantic–Caribbean group (including samples from Guadeloupe and Martinique) and a Pacific–Polynesian group (including samples from Tahiti and the Philippines). The Australian sample, however, never joined within the Pacific clade and always remained in a trichotomy. In order to analyse this problem more closely, a detailed examination of the alignment (Fig. 3) revealed that the Australian and all of the Mediterranean/aquarium samples shared several unique gaps. When gaps were counted as a new state, the Australian sample always clustered with the Mediterranean/aquarium group (Fig. 2B) revealing that Mediterranean populations of *C. taxifolia* are closely related to the Australian sample. We fully recognize, however, that we are near the limits of resolution of ITS and that the entire geographical range of *C. taxifolia* has not been sampled. Prud'homme van Reine *et al*. (1996) have described three major centres of biodiversity for *Caulerpa*: Malaysia (contact between Indian and Pacific Oceans), the Caribbean and southern Australia. To be absolutely conclusive about the source of the Mediterranean invasive population, additional samples from the Indian Ocean and southern Australia should be analysed. Ongoing attempts to obtain specimens from the Indian Ocean and western Australia have remained unfruitful, although efforts are underway for regional-level sampling of the northeast coast of Australia.

Further support for an Australian source comes from a recent allozyme survey of 12 populations of *C. taxifolia* from the western Mediterranean, northeast and eastern Australia (Benzie *et al*. 2000). These authors identified a population from eastern Australia (Stradebroke Island, Brisbane in southern Queensland) that is genetically more similar to populations (with similar robust morphology) in the Mediterranean than to populations (with fine morphology) in northern Australia. Unfortunately, Benzie *et al*. (2000) did not include populations from other tropical origins in their survey, so definitive conclusions remain elusive. The Australian samples examined in our study and that of Benzie *et al*. (2000) were collected some 1200 km apart along the eastern Australian coast. At present, there are no fine-scale comparisons with resolute markers among local and regional east coast Australian populations of *C. taxifolia* and the Mediterranean. Based on morphological similarities, it is probable that the local Australian origin is more southerly and possibly from the Brisbane area as opposed to the Townsville area.

### *Micro-organism diversity*

Another way to trace the origin of the Mediterranean *C. taxifolia* is to compare the algal-associated bacteria from *C. taxifolia* specimens collected from the different tropical origins. An RFLP analysis of the 16S rDNA clone libraries was applied to sort the clones into groups and to facilitate the choice of clones to be sequenced. High scoring of 16S rDNA clone sequence similarities in the EMBL database made it possible to assign taxonomic groups for the most frequent RFLP patterns. The majority of the 16S rDNA sequences fell into two of the 11 major divisions of the bacteria domain (Woese 1987): the Proteobacteria group and the CFB group. Two of the remaining sequences corresponded to chloroplasts, another fell into the Planctomycetes and one into the Cyanobacteria. These results were completely consistent among the four geographical origins of *C. taxifolia* and are quite different from previous phylogenetic analyses of particle-attached bacteria in adjacent coastal oceans that showed high proportions of Cyanobacteria, chloroplast and gamma Proteobacteria (Crump *et al*. 1999). Epiphytic, endosymbiotic and free-living unicellular and colonial Cyanobacteria species, as well as bacteria, have been shown to fix nitrogen in pelagic marine systems (Mulholland & Capone 2000). Epiphytic associations between coenocytic green algae and nitrogen-fixing bacteria have been previously described in *Codium decorticatum* (Rosenberg & Paerl 1981) and in *Codium fragile* (Gerard *et al*. 1990). In the present study, only one surface cyanobacterium was detected, whereas the endosymbiotic *Rhodobacter* (putatively nitrogen fixing) was detected (Table 3) in *C. taxifolia*.

We found four subdivisions of Proteobacteria: alpha, beta, gamma and delta. In marine microbial communities, the alpha Proteobacteria and also the CFB group are considered to be the most cosmopolitan (Fuhrman *et al*. 1993; Gonzalez & Moran 1997; Pinhassi *et al*. 1997; Giuliano *et al*. 1999). Although the Proteobacteria were found to be predominant in all host origins (= biogeographical locations where the *C. taxifolia* was collected), the frequencies of the different subclasses varied between these origins (Fig. 4). These differences are key to establishing Australia as the original source. Alpha Proteobacteria showed the highest frequency in the Philippines, whereas delta Proteobacteria were more prevalent in Tahiti (Fig. 4). CFBs were mainly found in the Mediterranean and Australian *C. taxifolia* samples. In both Mediterranean and Australian origins, the proportions of the different groups of bacteria associated with *C. taxifolia* were similar and characterized by a prevalence of alpha and beta Proteobacteria, and CFBs. Thus, analogous compositions of the bacterial flora shared between the Australian and Mediterranean sites of collection suggest a common source of the *C. taxifolia* hosts. However, to reinforce the conclusion of this study, spatial and temporal variation in

the bacterial communities associated with *C. taxifolia* are still needed.

Bacterial diversity is expected to be high in coastal waters due to the combined effects of the euphotic zone and proximity of the shore, which supplies sediments and particles from its various marine communities. However, the relationship between bacteria in the water column and their association with algae (either epiphytically or endosymbiotically) is still poorly known. Our results revealed a high number of 16S rDNA RFLP patterns in *C. taxifolia* that correspond to high bacterial diversity. While some bacteria are undoubtedly surface related and transient, relatively high numbers of endosymbiotic bacteria are known to be common in *Caulerpa*. Ultrastructural studies of the cytoskeleton in *Caulerpa prolifera* (Dawes & Lohr 1978) revealed that they are common in the vacuolar and cytoplasmic apical regions of the rhizome and base of the assimilators. More recently, Chisholm *et al*. (1996) estimated that $10^4$–$10^5$ bacteria/mm$^3$ exist in the cytoplasm of *C. taxifolia*. Hence, bacterial associates of *Caulerpa* are mainly internal.

The role of these bacteria, either as symbionts or as commensals, is not really known, although there is more evidence for the former than the latter. Chisholm *et al*. (1996) suggested that bacteria could effectively allow the rhizoids of *C. taxifolia* to function as 'roots', i.e. derive organic carbon, nitrogen and inorganic phosphorus, directly from the substrate. These authors also suggested that the rhizoid uptake of nutrients provides a physiological explanation for the success of siphonous green algae in oligotrophic tropical waters and the recent prolific development of *C. taxifolia* on eutrophicated substrates. Earlier studies of ammonium uptake from sediments by rhizoids of *Caulerpa cupressoides* (Williams 1984) also suggested a role for endosymbiotic bacteria.

### *Identification of* C. taxifolia*-associated bacteria*

None of the sequences matched exactly with bacterial sequences in GenBank. The highest similarity was found with the sequence of *Erythrobacter* sp. (0.98 sequence identity) for an isolate from the Philippines. In most cases, however, sequences could be assigned to a known bacterial genus (0.93–0.97 sequence identity). The lowest degree of similarity was observed for the isolates of the CFB group with several sequences showing a similarity of only 0.85. Pinhassi *et al*. (1997) obtained similar results and suggested that the CFB group contained several unknown genera. This may also be the case in the present study.

Chisholm *et al*. (1996) identified endosymbiotic bacteria in *C. taxifolia* belonging to the fluorescent *Pseudomonas* group of branch gamma Proteobacteria. This particular bacterium was not detected in the present study. Gamma Proteobacteria can be free living or occur as symbionts and pathogens on animal hosts (Balows *et al*. 1991; Friedrich *et al*. 1999).

In order to use bacterial associations as a tracking tool, the most interesting cases will involve those in which several different clones show high levels of similarity (> 0.97, i.e. species level). If the same bacterial species is shared among hosts from different origins, then it will not be informative; but if the bacterial species is found/shared in only one host from a particular origin (as compared with the sample of interest), then the source of the introduction can be identified. A third possibility is that a suite of different bacterial species belonging to the same genus is shared among hosts. While on the one hand this suggests a common ancestral origin, it may also be the result of shared physiological function, i.e. independent of shared ancestry. Because it is very difficult to distinguish between these two alternatives, we decided to focus our study on the most representative bacteria that showed close or nearly identical sequences, i.e. the CFBs, and alpha and beta Proteobacteria, although the roles of these bacteria are unknown.

In the CFB group, we found two robust clusters. The first cluster consisted of six virtually identical clones from the Mediterranean and the second cluster of three closely related ones of Australian origin. One clone from Tahiti was found to branch deeply among the CFB bacteria. The occurrence of six bacterial clones of the same species associated with the Mediterranean algal sample suggests a strong connection between this species and *C. taxifolia*. An important point here is that it is impossible to determine whether the bacterium is endophytic or epiphytic. CFBs are generally considered to be surface-associated bacteria, although we cannot exclude the possibility that this new species may be endosymbiotic. CFBs are known to produce exopolysaccharide substances and extracellular enzymes capable of degrading macromolecules such as cellulose (Reichenbach 1989). Although Mediterranean and Australian bacterial clusters are phylogenetically close, the clear genetic differentiation among geographical origins can also be interpreted as convergence based on similar physiological environments. However, even if we cannot conclude a common origin of the hosts, this result advocates that physiological constraints are more similar between Australia and the Mediterranean than among the other tropical origins.

Alpha Proteobacteria are the most commonly associated bacteria found with *C. taxifolia*. Previous studies conducted in different oceans suggests that alpha Proteobacteria are best adapted to the conditions of the euphotic zone (Mullins *et al*. 1995; Gonzalez & Moran 1997; Giuliano *et al*. 1999). This subdivision is of special interest because some members are associated with various eukaryotes and are able to reduce or oxidize nitrogen compounds (Woese 1987). In *Caulerpa*, most of the alpha proteobacterial clones were assigned to the *Rhodobacter* group. Within this group, *Roseobacter denitrificans* and *Roseobacter litoralis* have been isolated from the surface of green seaweeds (Shiba 1992), *Ruegeria algicola* has been detected in the phycosphere

of a marine dinoflagellate (Lafay *et al*. 1995) and *Prionitis lanceolata* was found in a gall in a marine red alga (Ashen & Goff 1996). Moreover, Chisholm *et al*. (1996) isolated an unidentified endosymbiotic bacterial species of the *Agrobacterium-Rhizobium* group from rhizoids of a Mediterranean *C. taxifolia*, which carries the *nif* H gene that encodes nitrogenase. In the present study, we found several clones belonging to the same bacterial species that never shared origins. Three identical clones (C. taxMed-2, -54 and -66) and one closely related clone (C. taxMed-9) formed a robust Mediterranean cluster assigned to the genus *Ruegeria*; and two identical clones (C. taxAus-11 and -27) and one closely related clone (C. taxAus-40) formed an Australian cluster assigned to the genus *Roseobacter*. These bacteria belong to two different genera, known to include both endosymbiotic and nitrate-reducing strains, which suggests that alpha Proteobacteria are rather diverse and are not strictly associated to the host.

The beta Proteobacteria were well represented in *C. taxifolia*. This is intriguing because this group is relatively unusual in the marine realm [see Pukall *et al*. (1999) and Giuliano *et al*. (1999) for the Mediterranean Sea, Pinhassi *et al*. (1997) for the northern Baltic Sea and Crump *et al*. (1999) for the coastal Pacific], but very common in freshwater (Methé *et al*. 1998). The four strains were very close and shared among the Mediterranean and the Australian hosts. This result further supports the origin of the introduced *C. taxifolia* in the western Mediterranean being of Australian origin. These four clones were assigned to *Herbaspirillum* (96% sequence similari-ties with *Herbaspirillum rubrisubalbic* and *Herbaspirillum frisingense*). The genus *Herbaspirillum* is composed of terrestrial nitrogen-fixing bacteria associated with higher plants and has, until now, not been described in the marine environment (Baldani *et al*. 1996).

Introductions of plants including their plant-associated bacteria are not well documented, but are probably more common than suspected.

Even if the mechanism by which bacteria are acquired by *Caulerpa* remains a mystery, the algal–bacterial associations described here provide an identification tool that may be widely applicable for detecting marine invasions. Better spatial and temporal quantification of the method, along with a streamlined assay using quantitative PCR, may render this a valuable and practical approach. The bacterial data, taken in conjunction with the newly analysed ITS sequences in the alga, make the case for an Australian origin of the Mediterranean strain of *C. taxifolia*, which will benefit from further exhaustive sampling along the Australian coast.

## Acknowledgements

We thank James R. Manhart for providing the 165 primers, François Bonhomme, Eric Coppejans, Eric Emery, Carolyn Engel, José Godin, Antonio Grau, Galice Hoarau, Yves Hénocque, Willem Prud'homme van Reine and Ante Zuljevic for collecting the samples. We also thank Tristan Barbeyron, Carolyn Engel and Philippe Vanderkoornhuyse for their comments on the first version of the manuscript. This research was supported by a grant from the French Ministry of the Environment (195–935–598–00–19 to M.V. and J.L.O.) and by a PhD fellowship to I.M. from the French Ministry of National Education.

## References

Altschul SF, Madden TL, Schaffer AA (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Ashen JB, Goff LJ (1996) Molecular identification of a bacterium associated with gall formation in the marine red alga *Prionitis lanceolata*. *Journal of Phycology*, **32**, 286–297.

Baldani JI, Pot B, Kirchhof G (1996) Emended description of *Herbaspirillum*; inclusion of *Pseudomonas rubrisubalbicans*, a mild plant pathogen, as *Herbaspirillum rubrisubalbicans* comb. nov. and classification of a group of clinical isolates (EF group 1) as *Herbaspirillum* species 3. *International Journal of Systematic Bacteriology*, **46**, 802–810.

Balows A, Trüper HG, Dworkin M, Harder W, Schleifer KH (1991) *The Prokaryotes*, 2nd edn. Springer, New York.

Bellan-Santini D, Arnaud PM, Bellan G, Verlarque M (1996) The influence of the introduced tropical alga *Caulerpa taxifolia*, on the biodiversity of the marine biota. *Journal of the Marine Biological Association of the United Kingdom*, **76**, 235–237.

Benzie JAH, Ballment E, Chisholm JRM, Jaubert JM (2000) Genetic variation in the green alga *Caulerpa taxifolia*. *Aquatic Botany*, **66**, 131–139.

Burr FA, West JA (1970) Light and electron microscope observations on the vegetative and reproductive structures of *Bryopsis hypnoides*. *Phycologia*, **9**, 17–37.

Chisholm JRM, Dauga C, Ageron E, Grimont PAD, Jaubert JM (1996) 'Roots' in mixotrophic algae. *Nature*, **381**, 382.

Chisholm JRM, Jaubert JM, Giaccone G (1995) *Caulerpa taxifolia* in the Northwest Mediterranean: introduced species or migrant from the Red Sea. *Comptes Rendus de l'Académie des Sciences, Series III*, **318**, 1219–1226.

Crump BC, Armbrust EV, Baross JA (1999) Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. *Applied and Environmental Microbiology*, **65**, 3192–3204.

Dawes CJ, Lohr CA (1978) Cytoplasmic organization and endosymbiotic bacteria in the growing points of *Caulerpa prolifera*. *Review of Algology*, **13**, 309–314.

Epplen JT, Lubjuhn T (1999) *DNA Profiling and DNA Fingerprinting*. Birkhäuser, Basel.

Felsenstein J (1989) PHYLIP: Phylogeny Inference Package, University of Washington, Seattle. *Cladistics*, **5**, 164–166.

Friedrich AB, Merkert H, Fendert T (1999) Microbial diversity in the marine sponge *Aplysina cavernicola* (formerly *Verongia cavernicola*) analysed by fluorescence in situ hybridization (FISH). *Marine Biology*, **134**, 461–470.

Fries L, Iwasaki H (1976) p-hydroxyphenylacetic acid and other phenolic compounds as growth stimulators of the red alga *Porphyra tenera*. *Plant Science Letters*, **6**, 299–307.

Fuhrman JA, McCallum K, Davis AA (1993) Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. *Applied and Environmental Microbiology*, **59**, 1294–1302.

Gerard VA, Dunham SE, Rosenberg G (1990) Nitrogen-fixation by cyanobacteria associated with *Codium fragile* (Chlorophyta): environmental effects and transfer of fixed nitrogen. *Marine Biology*, **105**, 1–8.

Giuliano L, De Domenico M, De Domenico E, Höfle MG, Yakimov MM (1999) Identification of culturable oligotrophic bacteria within naturally occurring bacterioplankton communities of the Ligurian Sea by 16S rRNA sequencing and probing. *Microbial Ecology*, **37**, 77–85.

Gonzalez JM, Moran MA (1997) Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater. *Applied Environmental Microbiology*, **63**, 4237–4242.

Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Harmelin-Vivien M, Francour P, Harmelin JG (1999) Impact of *Caulerpa taxifolia* on Mediterranean fish assemblages: a six year study. *Workshop on Invasive Caulerpa Species in the Mediterranean*, pp. 127–138. UNEP, Heraklion.

Jousson O, Pawlowski J, Zaninetti L, Meinesz A, Boudouresque CF (1998) Molecular evidence for the aquarium origin of the green alga *Caulerpa taxifolia* introduced to the Mediterranean Sea. *Marine Ecology-Progress Series*, **172**, 275–280.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Lafay B, Ruimy R, Rausch de Traubenberg C (1995) *Roseobacter algicola* sp. nov., a new marine bacterium isolated from the phycosphere of the toxin-producing dinoflagellate *Prorocentrum lima*. *International Journal of Systematic Bacteriology*, **45**, 290–296.

Manhart JR (1995) Phylogenetic uses of introns in green algal chloroplast genes. *Journal of Phycology Supplement*, **31**, 13.

Meinesz A, Boudouresque CF (1996) On the origin of *Caulerpa taxifolia* in the Mediterranean Sea. *Comptes Rendus de l'Academie des Sciences Séries III*, **319**, 603–613.

Meinesz A, Hesse B (1991) Introduction of the tropical alga *Caulerpa taxifolia* and its invasion of the northwestern Mediterranean. *Oceanologica Acta*, **14**, 415–426.

Methé BA, Hiorns WD, Zehr JP (1998) Contrasts between marine and freshwater bacterial community composition: analyses of communities in Lake George and six other Adirondack lakes. *Limnolology and Oceanography*, **43**, 368–374.

Mulholland MR, Capone DG (2000) The nitrogen physiology of the marine $N_2$-fixing cyanobacteria *Trichodesmium* spp. *Trends in Plant Science*, **5**, 148–153.

Mullins TD, Britschgi TB, Krest RL, Giovannoni SJ (1995) Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. *Limnology and Oceanography*, **40**, 148–158.

Olsen JL, Valero M, Meusnier I, Boele-Bos S, Stam WT (1998) Mediterranean *Caulerpa taxifolia* and *C. mexicana* (Chlorophyta) are not conspecific. *Journal of Phycology*, **34**, 850–856.

Pesando D, Lemée R, Ferrua C, Amade P, Girard JP (1996) Effects of caulerpenyne, the major toxin from *Caulerpa taxifolia* on mechanisms related to sea urchin egg cleavage. *Aquatic Toxicology*, **35**, 139–155.

Pinhassi J, Li Zweifel UL, Hagström A (1997) Dominant marine bacterioplankton species found among colony-forming bacteria. *Applied and Environmental Microbiology*, **63**, 3359–3366.

Preston GM, Haubold B, Rainey PB (1998) Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis. *Current Opinion in Microbiology*, **1**, 589–597.

Provasoli L (1958) Effect of plant hormones on *Ulva*. *Biological Bulletin (Woods Hole)*, **144**, 375–384.

Provasoli L, Pintner IJ (1980) Bacteria induced polymorphism in an axenic laboratory strain of *Ulva lactuca* (Chlorophyceae). *Journal of Phycology*, **16**, 196–201.

Prud'homme van Reine WF, Verheij E, Coppejans E (1996) Species and ecads of *Caulerpa* (Ulvophyceae, Chlorophyta) in Malesia (South-East Asia): taxonomy, biogeography and biodiversity. *Netherlands Journal of Aquatic Ecology*, **30**, 83–98.

Pukall R, Päuker O, Buntefuss D (1999) High sequence diversity of *Alteromonas macleodii*-related cloned and cellular 16S rDNAs from a Mediterranean seawater mesocosm experiment. *FEMS Microbiology Ecology*, **28**, 335–344.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Reichenbach H (1989) Genus 1. *Cytophaga*. In: *Bergey's Manual of Systematic Bacteriology* (eds Staley JT, Bryant MP, Pfenning N, Holt JG), pp. 2015–2050. Williams & Wilkins, Baltimore.

Rosenberg G, Paerl HW (1981) Nitrogen fixation by blue-green algae associated with the siphonous green seaweed *Codium decorticatum*: effects on ammonium uptake. *Marine Biology*, **61**, 151–158.

Roy R, Steffens DL, Gartside B, Jang GY, Brumbaugh JA (1996) Producing STR locus patterns from bloodstains and other forensic samples using an infrared fluorescent automated DNA sequencer. *Journal of Forensic Sciences*, **41**, 418–424.

Sant N, Delgado O, Rodriguez-Prieto C, Ballesteros E (1996) The spreading of the introduced seaweed *Caulerpa taxifolia* (Vahl) C. Agardh in the Mediterranean Sea: testing the boat transportation hypothesis. *Botanica Marina*, **39**, 427–430.

Shiba T (1992) The genus *Roseobacter*. In *The Prokaryotes: a Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications* (eds Starr MP, Stolp H, Trüpper HG, Balows A, Schelgel HG), pp. 202–206. Springer, Berlin.

Steffens DL, Sutter SL, Roemer SC (1993) An alternate universal forward primer for improved automated DNA sequencing of M13. *Biotechniques*, **15**, 580–581.

Swofford DL (1999) PAUP *(Phylogenetic Analysis Using Parsimony)*. Sinauer Associates Inc., Sunderland, MA, USA.

Swofford DL, Beagle DP (1993) PAUP*: Phylogenetic Analysis Using Parsimony, Version 3.13. User's Manual*. Laboratory of Molecular Systematics, Smithsonian Institution, Washington.

Tatewaki M, Provasoli L, Pintner IJ (1993) Morphogenesis of *Monostroma oxyspermum* (KÜTZ.) Doty (Chlorophyceae) in axenic culture, especially in bialgal culture. *Journal of Phycology*, **19**, 409–416.

Turner JB, Friedman EI (1974) Fine structure of capitular filaments in the coenocytic green alga *Penicillus*. *Journal of Phycology*, **10**, 125–134.

Van de Peer Y, De Rijk P, Wuyts J, Winkelmans T, De Wachter R (2000) The European small subunit ribosomal RNA database. *Nucleic Acids Research*, **28**, 175–176.

de Villèle X, Verlaque M (1995) Changes and degradation in a *Posidonia oceanica* bed invaded by the introduced tropical alga *Caulerpa taxifolia* in the North Western Mediterranean. *Botanica Marina*, **38**, 79–87.

Weidner S, Arnold W, Pühler A (1996) Diversity of uncultured micro-organisms associated with the seagrass *Halophila stipulacea*

estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA gene. *Applied Environmental Microbiology*, **62**, 766–771.

Williams SL (1984) Uptake of sediment ammonium and translocation in a marine green macroalga *Caulerpa cupressoides*. *Limnology and Oceanography*, **29**, 374–379.

Wintzingerode F, Göbela UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, **21**, 213–229.

Woese CR (1987) Bacterial evolution. *Microbiological Review*, **51**, 221–271.

Isabelle Meusnier is a PhD student in population genetics interested in the population structure and phylogeography of invasive species. Jeanine Olsen is a molecular ecologist interested in the phylogeography and evolutionary demography of marine algae, corals and fish. Wytze Stam is a molecular geneticist interested in the phylogenetics of marine algae, fish, corals and prokaryotes. Christophe Destombe is an algal population ecologist interested in evolutionary demography and local adaptation. Myriam Valero is a plant population geneticist interested in life-cycle and mating-system evolution.