Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria

Verena Salman[a,1], Rudolf Amann[a], David A. Shub[b], and Heide N. Schulz-Vogt[a,1]

[a]Max Planck Institute for Marine Microbiology, D-28359 Bremen, Germany; and [b]University at Albany, State University of New York, Albany, NY 12222

The gene encoding the small subunit rRNA serves as a prominent tool for the phylogenetic analysis and classification of *Bacteria* and *Archaea* owing to its high degree of conservation and its fundamental function in living organisms. Here we show that the 16S rRNA genes of not-yet-cultivated large sulfur bacteria, among them the largest known bacterium *Thiomargarita namibiensis*, regularly contain numerous self-splicing introns of variable length. The 16S rRNA genes can thus be enlarged to up to 3.5 kb. Remarkably, introns have never been identified in bacterial 16S rRNA genes before, although they are the most frequently sequenced genes today. This may be caused in part by a bias during the PCR amplification step that discriminates against longer homologs, as we show experimentally. Such length heterogeneity of 16S rRNA genes has so far never been considered when constructing 16S rRNA-based clone libraries, even though an elongation of rRNA genes due to intervening sequences has been reported previously. The detection of elongated 16S rRNA genes has profound implications for common methods in molecular ecology and may cause systematic biases in several techniques. In this study, catalyzed reporter deposition–fluorescence in situ hybridization on both ribosomes and rRNA precursor molecules as well as in vitro splicing experiments were performed and confirmed self-splicing of the introns. Accordingly, the introns do not inhibit the formation of functional ribosomes.

group I and II introns | LAGLIDADG homing endonuclease | length heterogeneity bias | microbial diversity | small ribosomal subunit

Studies on the microbial diversity in natural environments or enrichment cultures usually involve the analysis of the genes encoding the small subunit rRNA (16S rRNA in *Bacteria* and *Archaea* and 18S rRNA in *Eukarya*). This gene was selected as a phylogenetic marker because it was considered to be universally present, consistent in size and function, and not subject to horizontal transfer (reviewed in ref. 1). Current methods assessing the diversity of 16S rRNA genes, like the polymerase chain reaction (PCR) and the generation of 16S clone libraries, take advantage in particular of the constant size of the gene (2). In the past decades, these techniques have become fundamental tools in molecular ecology, and the 16S rRNA gene is the most frequently sequenced and analyzed gene today. The universality of this gene has never been questioned, even though in some studies unusual 16S rRNA genes, which contained, e.g., intervening sequences, were encountered (3–7).

In earlier studies, it was observed that clone libraries generated from environments visibly dominated by large sulfur bacteria frequently lacked the 16S rRNA gene sequences of these bacteria (8–13). Large sulfur bacteria occur at high densities predominantly in marine, organically rich coastal sediments (14–16), influencing the local sulfur (17), nitrogen (18), and phosphorous (19) cycle. The present study reports the frequent occurrence of long introns in the 16S rRNA genes of large sulfur bacteria, which explains their discrimination in PCR-based diversity assessments.

Results

Inserted Sequences in rRNA Genes. By comparative sequence analysis, we identified insertions of variable lengths (290−942 nucleotides) at up to four distinct sites within the 16S rRNA genes of large sulfur bacteria (Fig. 1*A*). Corresponding to these insertion sites, the inserted elements were classified as S795, S1078, S1396, and S1495 (according to *Escherichia coli* position numbering). Inserted sequences at the same site generally shared high sequence identities not only when found within one species (92−100%, Table S1), but also when inserted into different species and genera (usually 85−100% identical, except for S1495 with 46.8%, Table S1). In contrast, the four different sequence types shared only low identities among each other (40% or less), and the four insertion sites in the 16S rRNA gene did not show any sequence homology. In addition, the distribution of the inserted sequences varied between the different phylogenetic groups within the *Beggiatoaceae* (Fig. 1*B*) and also between individuals in a given species (details in Fig. S1). S1396 was the most frequent sequence in the group of large sulfur bacteria. Its phylogeny within a species reflected the phylogeny derived from the 16S rRNA gene sequence, but differed compared to the 16S-based relationship among higher taxa (details in Fig. S2). In eukaryotes, introns were detected in mitochondrial, chloroplast, and nuclear rRNA genes (20). Among them, introns were located at identical (1078) and nearby (793, 1391, and 1502) positions (refs. 21, 22) as the elements detected here in bacterial 16S rRNA genes. This finding is in accordance with the assumption that certain structural features or particular sequences in the rRNA gene attract introns, thus creating insertion hot spots in *Bacteria*, *Archaea*, and *Eukarya* (21).

The inserted sequences showed several characteristics of known group I and II introns, which are frequently found in all three domains of life (20). Group I and II introns can be classified into subgroups based on specific RNA folding patterns (Fig. 2; details in Fig. S3). Intron S795 belonged to the group I introns of the ID subgroup (Fig. 2*A*); S1078 was likewise a group I intron, but of the subgroup IA3 (Fig. 2*B*); also belonging to the group I introns, S1495 was a IC1 intron (Fig. 2*D*) (23); and the only group II intron identified so far in large sulfur bacteria was S1396, which belonged to the subgroup IIC (Fig. 2*C*) (24–27). Furthermore, S1396 showed parallels to a functional type of group II introns described recently by Li et al. (25). Special features include the presence of several additional nucleotides inserted at the 5′-end and the possession of a gene for a LAGLIDADG homing endonuclease [which is more commonly found in group I introns (20)] instead of a reverse transcriptase (25), which were traits likewise found in S1396.

Most of the introns investigated here (96 of 131) contained an ORF coding for an intron encoded protein (IEP). Translated
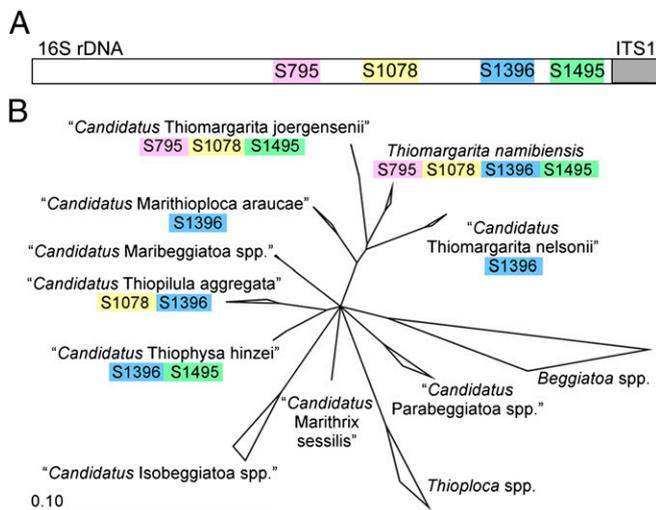
EVOLUTION

Fig. 1. Introns in the 16S rRNA genes of the large sulfur bacteria. (*A*) Four introns were inserted in the positions 795, 1078, 1396, and 1495 (according to *E. coli* numbering) in the gene for the small ("S") ribosomal subunit (*16S rDNA*). (*B*) Multifurcation tree based on nearly full-length 16S rRNA gene sequences of members of the family *Beggiatoaceae* showing the occurrence of introns in the different genera and species. To date introns have been located in 16S rRNA genes of the genera *Thiomargarita*, "*Candidatus* Marithioploca," "*Candidatus* Thiopilula," and "*Candidatus* Thiophysa."

amino acid sequences had 35–42% identity to site-specific endonucleases of the LAGLIDADG superfamily (29) described from different mitochondria and chloroplast genomes (30–33). The group I intron IEPs are double-motif LAGLIDADG endonucleases, and the group II intron IEPs contain a single LAGLIDADG motif (Fig. S4) (29, 34). Generally, this family of homing endonucleases is encoded by group I introns (20), but can be occasionally found in group II introns (25, 35, 36). The remaining 35 introns either lacked an ORF or had a truncated ORF. Truncated ORFs were found in 2 of the 18 introns S1495 that were isolated, either lacking a start codon (clone AMV001) or containing only a reduced ORF that represented a sequence of 35 amino acids (clone NAM035), which did not reveal any hits in the public databases.

Ligation of Exons During in Vitro Splicing. The capability of self-splicing and ligation was demonstrated for several bacterial introns (37–40) and was tested accordingly on the introns investigated here (Fig. 3 *A* and *B*). All four introns—S795 (clone sequence AMV001, with ORF), S1078 (clone sequence NAM032, with ORF), S1396 (clone sequence NAM056 with ORF), and S1495 (clone sequence AMV001, with truncated ORF)—showed self-splicing potential in transcription buffer during RNA synthesis (Fig. 3*A*). Ligation of the exons was demonstrated by RT-PCR (Fig. 3*B*), and sequencing confirmed ligation in all four cases.

Hybridization of Native Ribosomes and Precursor rRNA. Among individual cells of large sulfur bacteria, a heterogeneous distribution of introns was observed (Fig. S1), except for chain-forming *Thiomargarita namibiensis*, which always contained introns S1078 and S1396, and chain-forming "*Candidatus* Thiomargarita nelsonii," containing either S1396 or no intron. Whole fixed cells of both of these chain-forming types were hybridized by catalyzed reporter deposition–fluorescence in situ hybridization (CARD-FISH) (41) to demonstrate the transcription of the analyzed sequences [16S, internal transcribed spacer (ITS), and intron regions] into RNA. A DNA oligonucleotide that targeted position 1284–1303 of the 16S rRNA in native ribosomes (VSO1284, Table S2) hybridized to nearly all cells, yielding strong signals

indicating an overall presence of target molecules (Fig. 3*C*). In contrast, the application of oligonucleotides targeting the respective intron sequences yielded no signal above background (Fig. 3*D*; probe S1396-179, Table S2), indicating that introns were absent from mature 16S rRNA. Furthermore, application of a probe that bridged the intron insertion sites revealed strong fluorescent signals (Fig. 3*E*; probe Thm1389, Table S2). Finally, a probe that targeted the exon1/intron boundary sequence showed no detectable hybridization signal (Fig. 3*F*; probe Thm1397-INT, Table S2).

To exclude that the intron-containing 16S rRNA genes are pseudogenes, we applied a more sensitive CARD-FISH assay and tested whether the introns could be hybridized as part of the rRNA precursor (Fig. 4). Therefore, specific regions of the precursor (ITS, S1078, or S1396) were targeted simultaneously with three or four probes (Table S2), hybridized at a lower temperature (35 °C instead of 46 °C), and the RNA secondary structure was loosened by the addition of helper oligonucleotides (42).

Both ITS- and intron-specific signals were obtained with the altered hybridization conditions, which indicated that the rRNA precursor could be detected. Intracellular precursor concentrations are generally low, but can vary highly among individual cells depending on their metabolic state (43). This fluctuation was also observed in this study as generally few of the analyzed cells showed detectable signals with the altered protocol. Double-hybridization experiments targeting successively the intron and the ITS region revealed that all cells having intron-specific signals simultaneously showed ITS-specific signals (Fig. 4 *A* and *B*, "ITS" and "+" panels). In "*Ca*. T. nelsonii" cells, intron S1078 is not present, and indeed only an ITS-specific signal was detectable (Fig. 4*B*, "ITS" and "−" panels). These findings confirmed that intron-containing 16S rRNA genes were transcribed and could be specifically hybridized in the rRNA precursor.

Frequently, ITS signals were detected although intron signals were not, but never vice versa. This indicated that introns of the 16S rRNA might be excised and/or degraded more efficiently than the ITS region, whereas the latter might be excised and degraded in a later step. We can conclude from the CARD-FISH results that the formation of native ribosomes in large sulfur bacteria is not affected by the presence of intron sequences in their rDNA because they are removed during rRNA maturation processes.

Length-Heterogeneity PCR Bias. Length deviation in target sequences can cause a PCR bias that preferentially amplifies shorter over longer homologs (44, 45). We assumed that this bias could be the potential reason for intron-containing 16S rRNA genes of large sulfur bacteria being frequently missing from universal clone libraries (8–13). Equal amounts of cloned intron-containing (~2,350 nt) and intron-lacking (~1,500 nt) 16S rRNA genes were combined and subject to a universal PCR approach. Indeed, we observed a systematic discrimination of the longer gene over the homolog being shorter by 850 nt (Fig. S5, lane 3). Simulating a more realistic approach, which assumed that intron-containing genes are less frequent in a natural sample than intron-lacking genes, the longer homolog was provided less concentrated (ratios of 1:10, 1:100, and 1:1,000). In all these approaches, again only the shorter gene was amplified (Fig. S5, lanes 4–6), which further demonstrated the strong effect of the length-heterogeneity bias. This bias can be explained by the kinetics of the enzymatic reaction during the PCR amplification step (44) and may affect those reactions containing homologs that exceed a certain length deviation and/or that are combined in unfavorable concentrations.

Discussion

In this study we demonstrate that the 16S rRNA genes of large sulfur bacteria can contain self-splicing introns, up to three being
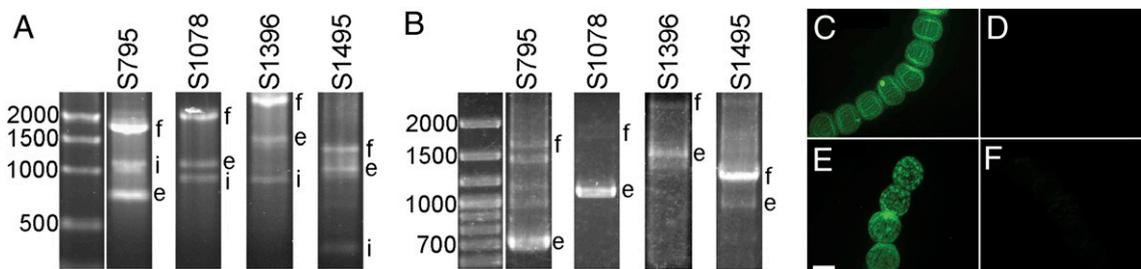
Fig. 2. RNA folding structures of the four introns detected in the 16S rRNA genes of large sulfur bacteria. Thin lines represent intron regions and thick lines represent exon regions. If present, the start and stop codons of the ORFs are presented. In addition, details about intramolecular conserved structures are given. Folding criteria were extracted from ref. 23 for all group I introns and from refs. 24–27 for the group IIC intron (Fig. S3). (*A*) Intron S795; a group ID intron with ORF in domain P8 originating from clone sequence AMV001 [accession no. FR690921 (28)]. (*B*) Intron S1078; a group IA3 intron with OFR in domain P6 originating from clone sequence NAM032 [accession no. FR690922 (28)]. (*C*) Intron S1396; a group IIC intron with ORF in domain DIV originating from clone NAM056 [accession no. FR690945 (28)]. (*D*) Intron S1495; a group IC1 intron with truncated ORF in P6 (no start codon detectable) originating from clone sequence AMV001 [accession no. FR690921 (28)].

present in a single gene and extending its length to more than 3,500 nucleotides. Introns are generally rare in bacteria and have never been encountered previously in bacterial 16S rRNA genes. So, it was quite surprising to even identify multiple introns inserted into these bacterial 16S rRNA genes.

The four introns are probably analogs because they do not share a common insertion site and their sequences are not related to each other. Thus, every intron type has probably inserted independently into the genome. Phylogenetic analysis of the most frequently observed intron S1396, a group II intron, revealed that it was probably acquired both vertically and horizontally within the family. This feature is well known among group II introns (46–49). Comparative sequence analysis of intron S1396 generally reflects the phylogeny suggested by 16S rRNA sequence analysis at species level (Table S1 and Fig. S2), indicating that the intron coevolved with the host after insertion into its genome. At the family level, the relationship of S1396 between genera is different from the 16S rRNA-derived phylogenetic pattern, suggesting

horizontal transfer of this intron between different taxa. This phenomenon is likely caused by the high mobility of these elements, frequently inserting and exiting from genomes (46, 47). Furthermore, introns are preferentially found in conserved regions of the rRNA genes (21, 22), which can explain why the insertion sites of introns analyzed in this study are homologous to locations of introns found in mitochondrial and nuclear rRNA genes of eukaryotes. This is consistent with the hypothesis that introns can be horizontally transferred over large phylogenetic distances (50, 51). All ORFs in the introns described here encoded IEPs typical of a common family of homing endonucleases, suggesting that the introns spread into cognate sites as DNA elements via homing (20). The existence of variants that lack, or have truncated IEPs, is considered to be a consequence of introns saturating all possible insertion sites and of IEPs for intron homing becoming redundant (52).

Introns have been well studied regarding their splicing mechanism from RNA precursors and their methods of propagation in

Fig. 3. Self-splicing experiments showed in vitro splicing abilities of RNA introns (*A* and *B*), and CARD-FISH hybridization of the 16S rRNA in the native ribosome demonstrated the absence of RNA introns from native ribosomes (*C*–*F*). (*A*) After transcription into RNA, the intron folds and splices itself, thereby ligating the two exons. The three bands in each row indicate the presence of the not-yet-spliced full-length construct "f," the ligated exons "e," and the excised intron "i." (*B*) Bulk RNA was subject to RT-PCR with primers on both exon ends. The resulting PCR products were the full-length construct "f" and the ligated exons "e." Bands of intermediate sizes in lane S795 result from incomplete splicing of the intron. (*C*) Probe VSO1284 was specific for the 16S rRNA of large sulfur bacteria and revealed strong signals for nearly all cells examined. (*D*) Hybridization of intron regions using a single probe (here S1396-179) never revealed detectable signals. (*E*) A probe bridging the intron insertion site (here Thm1389) resulted in strong fluorescent signals, whereas (*F*) a probe binding the exon1/intron boundary sequence (here Thm1397-INT) never showed any signal. These results strongly indicate that the introns are not present in the mature ribosome. (Scale bar, 100 μm.)

the host genome (35, 46, 53–57). Here we show that introns in large sulfur bacteria also demonstrate self-splicing abilities in vitro (Fig. 3) and removal from precursor rRNA molecules in vivo (Figs. 3 and 4). CARD-FISH on the rRNA precursor targeting the ITS region was previously performed (58, 59), but in this study also introns in the rRNA precursors were hybridized. The successful and specific labeling of the different introns demonstrates their transcription into RNA, strongly indicating that the 16S rRNA genes retrieved by PCR were not pseudogenes. Also, the invariable absence of intron-free PCR products from a cell yielding an intron-containing 16S rRNA gene amplicon further supports the suggestion that the intron-containing gene is functional. The absence of motifs for precursor processing enzymes like RNase III and the in vitro self-splicing activity indicate the autonomous and uncoupled removal of the introns from the precursor in vivo (Figs. 3 and 4), which certainly prevents negative impact on ribosome functionality.

The detection of introns in the 16S rRNA genes of large sulfur bacteria raises questions of why they are retained in the genes of this group of organisms, whether they have a biological function, or whether they are strictly selfish parasitic DNA elements. Currently, we can offer only suggestions for the first of these questions. We presume that introns cause less constraints and thus experience reduced selective pressure in large sulfur bacteria, compared with introns in most ordinary-sized bacteria. This hypothesis is based on the assumption that introns might persist longer in organisms that spend only a small fraction of their energy on replication (60). This energy surplus could be produced either by mitochondria in eukaryotes (60) or by the storage of tremendous amounts of energy reserves as in the group of large sulfur bacteria (16, 18, 19, 61). The replication, transcription, and excision of introns may thus not represent a major bioenergetic expense for large sulfur bacteria. DNA replication is considered to require only about 2% of a cell's energy budget (62), and transcription can be regarded as even less costly. Accordingly, large sulfur bacteria could afford the retention of intron-containing genes (60) owing to their energy surplus in the form of storage compounds (16, 18, 19, 61). Alternatively, it could be speculated that the high frequency of introns in this group of bacteria is somehow connected to the extreme polyploidy in these organisms: nucleic-acid staining in *Thiomargarita* spp. reveals the presence of several thousand nucleoids per cell (62, 63). In view of the insertion mechanism of introns, initiated by a double-strand break of the DNA and followed by cellular repair mechanisms involving homologous recombination, the presence of multiple alleles within a cell is necessary for efficient repair of the intron insertion site. In fact, the spreading of introns in organisms with several thousand genomes could be extremely

efficient as only a few IEPs are required to introduce many double-strand breaks, which could then be repaired by copying intron sequences into new sites. Thus, extreme polyploidy can cause rapid intron spread, while at the same time allowing efficient repair of harmful double-strand breaks in the host.

Considering that the 16S rRNA gene is the most sequenced gene today, it seems surprising that bacterial intron-containing
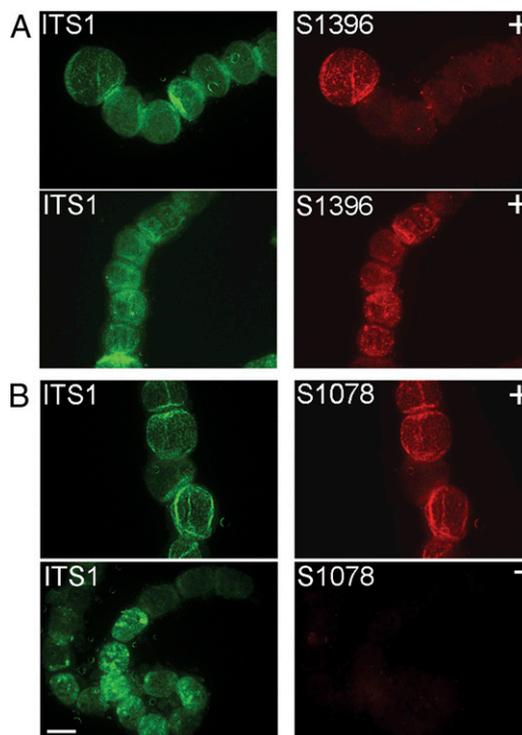

Fig. 4. CARD-FISH on the rRNA precursor of large sulfur bacteria. Hybridization results for the ITS region are shown on the left (green) and for the intron region on the right (red). Successful hybridization of introns in those cells that simultaneously showed an ITS-specific signal indicates that introns are transcribed as part of the rRNA precursor. (*A*) Simultaneous hybridization of the ITS region and intron S1396 in *Thiomargarita namibiensis* (*Upper* panels) and "*Ca*. T. nelsonii" (*Lower* panels) revealed positive fluorescent signals. Sequencing 16S rRNA genes of individual cells of these species implied that intron S1396 should be present in both species ("+"). (*B*) Simultaneous hybridization of the ITS region and intron S1078, which is expected to be present only in *T. namibiensis* (*Upper* panel, "+") and not in "*Ca*. T. nelsonii" (*Lower* panel, "−"). (Scale bar, 100 μm.)

16S rRNA genes have never been encountered before. Thus, it is tempting to speculate that large sulfur bacteria are the only group of bacteria accumulating introns in their 16S rRNA genes. However, the presence of introns, even multiple ones, has been described for the 23S rRNA genes of several distantly related bacteria (39, 50, 51, 64, 65). This gene bears the same evolutionary importance and conservation as the 16S rRNA gene, which indicates that 16S introns might be at least as frequent as 23S introns. Furthermore, introns in archaeal rRNA genes are very common and have been detected in 16S rRNA genes several times (66, 67). Generally, introns in rRNA may have a better opportunity for self-splicing because they are not subject to translation compared with introns inserted into mRNA (68–70). Also, the high level of sequence conservation of rRNA genes and their frequent occurrence in multiple copies could increase intron mobility among these regions (71). We therefore speculate that introns in bacterial rRNA genes might be more common than previously recognized. Nonetheless, public databases do not contain any entries of bacterial 16S rRNA genes with introns (21, 72, 73). This is either a coincidence or a consequence of biases introduced during the application of techniques and/or the interpretation of results.

The possibility that the 16S rRNA gene may not have a conserved size has not been sufficiently considered in designing and applying techniques for diversity studies, even though elongations of 16S rRNA genes due to intervening sequences have been reported before (3–7). Intervening sequences, in contrast to introns, cannot self-splice from precursor RNA and, upon removal by cellular maturation enzymes, leave the ribosomal RNA fragmented (3–7). In this study, we also showed that true self-splicing introns can be present in bacterial 16S rRNA genes, leading to the conclusion that the 16S rRNA genes of bacteria can indeed be far longer than ~1,500 nucleotides. Surely, this finding has far-reaching consequences because the gene for the small ribosomal subunit plays a fundamental role in phylogenetic and diversity studies and most currently applied methods to study this gene are based on its consistency.

The present study provides further evidence that the retrieval frequency of rRNA genes should not be considered a reliable measure for the abundance of microbial species. We demonstrate severe size selection of PCR products, i.e., the discrimination of intron-containing 16S rRNA gene sequences in a heterogeneous sample. This has far-reaching consequences for environmental diversity studies because entire phylogenetic groups may be missed. Future genome sequencing as well as the generation of large metagenomic datasets will reveal whether introns in 16S rRNA genes, or elongated 16S rRNA genes in general, might be a more common feature of *Bacteria* than previously recognized.

Materials and Methods

16S rRNA and Intron Sequences. The amplification of the 16S rRNA genes including intron sequences has been recently published (28). Comparative sequence analysis of 16S rRNA genes and introns was performed in BioEdit (74) and in the ARB software package (75). RNA folding structures of introns were created manually in Canvas (ACD Systems) using folding criteria assigned by ref. 23 for group I introns and by refs. 24–27 for group II introns (Fig. 2 and Fig. S3). More details can be found in *SI Materials and Methods*.

CARD-FISH. Probe 1284 targeting the 16S rRNA of all nonfilamentous large sulfur bacteria (Table S2) was designed in ARB (75). All intron probes, ITS probes, and helper oligonucleotides (Tables S2 and S3) were designed manually. Hybridization of *Thiomargarita* spp. was performed on slides using a protocol modified from ref. 41. For more details, see *SI Materials and Methods*.

In Vitro Splicing. Intron-containing 16S rRNA genes were either cloned into vectors featuring a T7 RNA polymerase promotor or PCR-amplified with a primer containing a 5′-promoter overhang. In vitro transcription was performed using the T7 RNA Polymerase-Plus kit (Ambion), and cDNA was generated with the RevertAid H-Minus First Strand cDNA Sythesis kit (Fermentas). Splicing products were amplified via PCR using High Fidelity PCR Enzyme mix (Fermentas). RNA and DNA products were analyzed by agarose gel electrophoresis. DNA bands were further analyzed by sequencing. Details on the reactions are in *SI Materials and Methods*.

PCR of Homologous Genes with Different Lengths. Intron-containing and intron-less 16S rRNA genes were cloned and used for PCR with universal 16S primers (GM3F and GM4R) (76) and the High Fidelity PCR Enzyme mix (Fermentas). The reaction contained either one of the two homologs or a mixture of both in varying ratios. For details on the reactions, see *SI Materials and Methods*.

1. Pace NR, Olsen GJ, Woese CR (1986) Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* 45:325–326.
2. Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–271.
3. Burgin AB, Parodos K, Lane DJ, Pace NR (1990) The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. *Cell* 60:405–414.
4. Linton D, Clewley JP, Burnens A, Owen RJ, Stanley J (1994) An intervening sequence (IVS) in the 16S rRNA gene of the eubacterium *Helicobacter canis*. *Nucleic Acids Res* 22:1954–1958.
5. Rainey FA, Ward-Rainey NL, Janssen PH, Hippe H, Stackebrandt E (1996) *Clostridium paradoxum* DSM 7308(T) contains multiple 16S rRNA genes with heterogeneous intervening sequences. *Microbiology* 142:2087–2095.
6. Ralph D, McClelland M (1993) Intervening sequence with conserved open reading frame in eubacterial 23S rRNA genes. *Proc Natl Acad Sci USA* 90:6864–6868.
7. Springer N, et al. (1993) Occurrence of fragmented 16S rRNA in an obligate bacterial endosymbiont of *Paramecium caudatum*. *Proc Natl Acad Sci USA* 90:9892–9895.
8. Angert ER, et al. (1998) Molecular phylogenetic analysis of a bacterial community in Sulphur River, Parker Cave, Kentucky. *Am Mineral* 83:1583–1592.
9. Edgcomb VP, Kysela DT, Teske A, de Vera Gomez A, Sogin ML (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc Natl Acad Sci USA* 99:7658–7662.
10. Gillan DC, Speksnijder AGCL, Zwart G, De Ridder C (1998) Genetic diversity of the biofilm covering *Montacuta ferruginosa* (Mollusca, Bivalvia) as evaluated by denaturing gradient gel electrophoresis analysis and cloning of PCR-amplified gene fragments coding for 16S rRNA. *Appl Environ Microbiol* 64:3464–3472.
11. López-García P, et al. (2003) Bacterial diversity in hydrothermal sediment and epsilonproteobacterial dominance in experimental microcolonizers at the Mid-Atlantic Ridge. *Environ Microbiol* 5:961–976.
12. Sekar R, Mills DK, Remily ER, Voss JD, Richardson LL (2006) Microbial communities in the surface mucopolysaccharide layer and the black band microbial mat of black band-diseased *Siderastrea siderea*. *Appl Environ Microbiol* 72:5963–5973.
13. Stevens H, Ulloa O (2008) Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol* 10:1244–1259.
14. Gallardo VA (1977) Large benthic microbial communities in sulfide biota under Peru-Chile subsurface countercurrent. *Nature* 268:331–332.
15. Mussmann M, et al. (2003) Phylogeny and distribution of nitrate-storing *Beggiatoa* spp. in coastal marine sediments. *Environ Microbiol* 5:523–533.
16. Schulz HN, et al. (1999) Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* 284:493–495.
17. Jørgensen BB, Fenchel T (1974) The sulfur cycle of a marine sediment model system. *Mar Biol* 24:189–201.
18. Fossing H, et al. (1995) Concentration and transport of nitrate by the mat-forming sulfur bacterium *Thioploca*. *Nature* 374:713–715.
19. Schulz HN, Schulz HD (2005) Large sulfur bacteria and the formation of phosphorite. *Science* 307:416–418.
20. Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. *Annu Rev Biochem* 62:587–622.
21. Jackson SA, Cannone JJ, Lee JC, Gutell RR, Woodson SA (2002) Distribution of rRNA introns in the three-dimensional structure of the ribosome. *J Mol Biol* 323:35–52.
22. Johansen SD, Haugen P, Nielsen H (2007) Expression of protein-coding genes embedded in ribosomal DNA. *Biol Chem* 388(1):679–686.
23. Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216:585–610.
24. Granlund M, Michel F, Norgren M (2001) Mutually exclusive distribution of IS1548 and GBSi1, an active group II intron identified in human isolates of group B streptococci. *J Bacteriol* 183:2560–2569.
25. Li C-F, Costa M, Bassi G, Lai Y-K, Michel F (2011) Recurrent insertion of 5′-terminal nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in mitochondrial preribosomal RNAs. *RNA* 17:1321–1335.
26. Quiroga C, Roy PH, Centrón D (2008) The S.ma.I2 class C group II intron inserts at integron attC sites. *Microbiology* 154:1341–1353.
27. Toor N, Robart AR, Christianson J, Zimmerly S (2006) Self-splicing of a group IIC intron: 5′ exon recognition and alternative 5′ splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res* 34:6461–6471.

EVOLUTION

28. Salman V, et al. (2011) A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. *Syst Appl Microbiol* 34:243–259.

29. Hensgens LAM, Bonen L, de Haan M, van der Horst G, Grivell LA (1983) Two intron sequences in yeast mitochondrial COX1 gene: Homology among URF-containing introns and strain-dependent variation in flanking exons. *Cell* 32:379–389.

30. Côté V, Mercier JP, Lemieux C, Turmel M (1993) The single group-I intron in the chloroplast rrnL gene of *Chlamydomonas humicola* encodes a site-specific DNA endonuclease (I-ChuI). *Gene* 129(1):69–76.

31. Saiki RK, et al. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350–1354.

32. Sethuraman J, Majer A, Friedrich NC, Edgell DR, Hausner G (2009) Genes within genes: Multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of *Ophiostoma* and related taxa. *Mol Biol Evol* 26:2299–2315.

33. Wakasugi T, et al. (1997) Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division. *Proc Natl Acad Sci USA* 94:5967–5972.

34. Dalgaard JZ, et al. (1997) Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res* 25:4626–4638.

35. Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. *Trends Genet* 21(2):111–119.

36. Toor N, Zimmerly S (2002) Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA* 8:1373–1377.

37. Adamidi C, Fedorova O, Pyle AM (2003) A group II intron inserted into a bacterial heat-shock operon shows autocatalytic activity and unusual thermostability. *Biochemistry* 42:3409–3418.

38. Chien MF, Tosa S, Huang CC, Endo G (2009) Splicing of a bacterial group II intron from *Bacillus megaterium* is independent of intron-encoded protein. *Microbes Environ* 24 (1):28–32.

39. Nesbø CL, Doolittle WF (2003) Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc Natl Acad Sci USA* 100:10806–10811.

40. Ferat JL, Le Gouar M, Michel F (2003) A group II intron has invaded the genus *Azotobacter* and is inserted within the termination codon of the essential groEL gene. *Mol Microbiol* 49:1407–1423.

41. Pernthaler A, Pernthaler J, Amann R (2002) Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* 68:3094–3101.

42. Fuchs BM, Glöckner FO, Wulf J, Amann R (2000) Unlabeled helper oligonucleotides increase the *in situ* accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes. *Appl Environ Microbiol* 66:3603–3607.

43. Britschgi TB, Cangelosi GA (1995) Detection of rifampin-resistant bacteria using DNA probes for precursor rRNA. *Mol Cell Probes* 9(1):19–24.

44. Cardinale M, et al. (2004) Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Appl Environ Microbiol* 70:6147–6156.

45. Jeffreys AJ, Wilson V, Neumann R, Keyte J (1988) Amplification of human minisatellites by the polymerase chain reaction: Towards DNA fingerprinting of single cells. *Nucleic Acids Res* 16:10953–10971.

46. Dai L, Zimmerly S (2002) Compilation and analysis of group II intron insertions in bacterial genomes: Evidence for retroelement behavior. *Nucleic Acids Res* 30:1091–1102.

47. Toro N (2003) Bacteria and Archaea Group II introns: Additional mobile genetic elements in the environment. *Environ Microbiol* 5(3):143–151.

48. Tourasse NJ, Kolstø AB (2008) Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: Insights into their spread and evolution. *Nucleic Acids Res* 36:4529–4548.

49. Zimmerly S, Hausner G, Wu Xc (2001) Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res* 29:1238–1250.

50. Haugen P, et al. (2007) Cyanobacterial ribosomal RNA genes with multiple, endonuclease-encoding group I introns. *BMC Evol Biol* 7:159.

51. Raghavan R, Miller SR, Hicks LD, Minnick MF (2007) The unusual 23S rRNA gene of *Coxiella burnetii*: Two self-splicing group I introns flank a 34-base-pair exon, and one element lacks the canonical omegaG. *J Bacteriol* 189:6572–6579.

52. Goddard MR, Burt A (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci USA* 96:13880–13885.

53. Edgell DR, Belfort M, Shub DA (2000) Barriers to intron promiscuity in bacteria. *J Bacteriol* 182:5281–5289.

54. Kruger K, et al. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31(1):147–157.

55. Lambowitz AM, Zimmerly S (2004) Mobile group II introns. *Annu Rev Genet* 38:1–35.

56. Michel F, et al. (1992) Activation of the catalytic core of a group I intron by a remote 3′ splice junction. *Genes Dev* 6:1373–1385.

57. Peebles CL, et al. (1986) A self-splicing RNA excises an intron lariat. *Cell* 44:213–223.

58. Oerther DB, Pernthaler J, Schramm A, Amann R, Raskin L (2000) Monitoring precursor 16S rRNAs of *Acinetobacter* spp. in activated sludge wastewater treatment systems. *Appl Environ Microbiol* 66:2154–2165.

59. Schmid M, Schmitz-Esser S, Jetten M, Wagner M (2001) 16S-23S rDNA intergenic spacer and 23S rDNA of anaerobic ammonium-oxidizing bacteria: Implications for phylogeny and *in situ* detection. *Environ Microbiol* 3:450–459.

60. Darnell JE, Doolittle WF (1986) Speculations on the early course of evolution. *Proc Natl Acad Sci USA* 83:1271–1275.

61. Preisler A, et al. (2007) Biological and chemical sulfide oxidation in a *Beggiatoa* inhabited marine sediment. *ISME J* 1:341–353.

62. Lane N, Martin W (2010) The energetics of genome complexity. *Nature* 467:929–934.

63. Schulz HN (2006) The genus *Thiomargarita*. In *The Prokaryotes*, eds Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (Springer, New York), pp 1156–1163.

64. Everett KDE, Kahane S, Bush RM, Friedman MG (1999) An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J Bacteriol* 181: 4734–4740.

65. Seshadri R, et al. (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc Natl Acad Sci USA* 100:5455–5460.

66. Burggraf S, Larsen N, Woese CR, Stetter KO (1993) An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci USA* 90: 2547–2550.

67. Itoh T, Suzuki K, Nakase T (1998) Occurrence of introns in the 16S rRNA genes of members of the genus *Thermoproteus*. *Arch Microbiol* 170(3):155–161.

68. Cavalier-Smith T (1991) Intron phylogeny: A new hypothesis. *Trends Genet* 7(5): 145–148.

69. Doolittle WF (1991) The origins of introns. *Curr Biol* 1(3):145–146.

70. Martin W, Koonin EV (2006) Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440:41–45.

71. Thompson AJ, Herrin DL (1994) A chloroplast group I intron undergoes the first step of reverse splicing into host cytoplasmic 5.8S ribosomal RNA: Implications for intron-mediated RNA recombination, intron transposition and 5.8S ribosomal RNA structure. *J Mol Biol* 236:455–468.

72. Cannone JJ, et al. (2002) The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.

73. Simon DM, et al. (2008) Group II introns in eubacteria and archaea: ORF-less introns and new varieties. *RNA* 14:1704–1713.

74. Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*, no. 41: 95–98.

75. Ludwig W, et al. (2004) ARB: A software environment for sequence data. *Nucleic Acids Res* 32:1363–1371.

76. Muyzer G, Teske A, Wirsen CO, Jannasch HW (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* 164(3):165–172.