ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs

LIXIN DAI and STEVEN ZIMMERLY
Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

ABSTRACT

Although group II intron retroelements are prevalent in eubacteria, they have not been identified in archaebacteria in the first 10 genomes sequenced. However, the recently sequenced archael genome of *Methanosarcina acetivorans* contains 21 group II introns, including 7 introns that do not encode reverse transcriptase ORFs. To our knowledge, these are the first retroelements identified in archaebacteria, and the first ORF-less group II introns in bacteria. Furthermore, the insertion pattern of the introns is highly unusual. The introns appear to insert site-specifically into ORFs of other group II introns, forming nested clusters of up to four introns, but there are no flanking exons that could encode a functional protein after the introns have been spliced out.

Keywords: archaea; retroelement; reverse transcriptase; ribozyme

Group II introns are retroelements consisting of a self-splicing intron RNA structure and an intron-encoded reverse transcriptase (RT). Group II introns are widely dispersed in eubacteria, mitochondria, and chloroplasts, and are mobile primarily by inserting into defined target sites (homing; Lambowitz et al. 1999; Bonen and Vogel 2001; Belfort et al. 2002). About one quarter of sequenced eubacterial genomes contain group II introns, and most of the introns are located in mobile DNAs such as IS elements. Based on a number of observations, we have suggested that group II introns in eubacteria behave mainly as retroelements rather than introns (Dai and Zimmerly 2002).

Group II introns were not initially found in archaebacteria in the first 10 archeabacterial genomes sequenced. However, in the recently sequenced genome of *Methanosarcina acetivorans*, the first archael group II introns have been identified (Galagan et al. 2002; this work). In fact, the *M. acetivorans* genome contains 21 group II introns, including 7 ORF-less introns. The insertion patterns are unlike those observed in eubacteria, and the inferred mobility events lend further support to the retroelement behavior of group II introns in bacteria.

Reprint requests to: Steven Zimmerly, Department of Biological Sciences, University of Calgary, 2500 University Dr. NW, Calgary, AB T2N 1N4, Canada; e-mail: zimmerly@acs.ucalgary.ca.

Article and publication are at http://www.rnajournal.org/cgi/doi/10.1261/rna.2126203.

We first noted the group II introns during routine searches for group II intron ORFs among sequences reported to GenBank. Twelve ORFs related to group II intron RTs were initially identified in *M. acetivorans*. The introns fell into two broad groups, one being similar to bacterial class D group II introns, and the other to chloroplast-like class 1 introns (for class descriptions, see Toor et al. 2001; Zimmerly et al. 2001). The boundaries of the introns were defined by folding the sequences into group II intron structures, as well as by sequence comparisons with known introns or sequence comparisons among different copies of the same intron. Further BLASTN searches of the *M. acetivorans* were used to screen for RNA structures not associated with ORFs, and seven ORF-less group II introns were identified. These ORF-less introns also form two groups, one related to bacterial class D and one to chloroplast-like class 1. Finally, to search for additional, unrelated ORF-less group II introns, the genome was searched for domain V motifs using the RNAMotif program (Macke et al. 2001; Toor and Zimmerly 2002), but only the known introns were identified. The final listing of group II introns in *M. acetivorans* (Table 1) includes six full-length introns that encode ORFs, eight fragmented introns that encode ORFs, six full-length ORF-less introns, and one truncated ORF-less intron. We also identified two group II introns in *Methanosarcina mazei*, [AE013515 (3337–5483), AE03516 (7432–10327)] that are closely related to *M.a.*I1–1, and that will not be described further.

TABLE 1. Group II introns in *Methanosarcina acetivorans*

Intron[a]	GenBank acc. no.[b]	RNA domains[c]	ORF domains[d]	Intron class[e]	Upstream flank[f]	Downstream flank[f]	Identity to *M.a.*I1-1[g]	Identity to *M.a.*I5-1[h]	M[i]	C[j]	N[k]
*M.a.*I1-1	AE010073 (4279–6431)	1–6	RT(0–7), X	CL1	Conserved hypothetical protein[l]	Conserved hypothetical protein[l]	100%		0	0	0
*M.a.*I1-2	AE011185 (1722–5744)	1–6	RT(0–7), X	CL1	PQG H.S.	—	90.8%		1	9	7
*M.a.*I1-3	AE011130 (2228–6247)	1–6	RT(0–7), X	CL1	RQG H.S.	—	93.5%		0	14	14
*M.a.*I1-F1	AE011130 (6428–7525)	1–3	RT(0–4)	CL1	Transposase	—	87.5%		3	6	8
*M.a.*I1-F2	AE010979 (6611–7239)	5–6	RT(7), X	CL1	Transposase	PQG H.S.	96.1%		0	1	0
*M.a.*I2-F1	AE011130 (7526–9438)	1–3	RT(0–4)	CL1	—	—	61.6%		3	73	13
*M.a.*I3-F1	AE011185 (5745–7226)	1–3	RT(0–4)	CL1	—	—	61.2%		2	77	19
*M.a.*I4-1	AE010882 (4828–5781)	1–6	—	CL1	—	—	61.9%		8	70	15
*M.a.*I4-2	AE011106 (4446–5200)	1–6	—	CL1	—	—	61.9%		8	70	16
*M.a.*I4-3	AE010996 (801–1555)	1–6	—	CL1	—	—	61.9%		8	70	15
*M.a.*I5-1	AE011130 (2949–4823)	1–6	RT(0–7),X	Bact D	YADD H.S.	YADD H.S.		100%	0	0	0
*M.a.*I5-2	AE010979 (4618–6492)	1–6	RT(0–7),X	Bact D	YADD H.S.	YADD H.S.		100%	0	0	0
*M.a.*I5-3	AE011185 (2451–4325)	1–6	RT(0–7),X	Bact D	YADD H.S.	YADD H.S.		99.9%	1	1	1
*M.a.*I6-1	AE011030 (5879–6500)	1–6	—	Bact D	—	—		82%	6	14	14
*M.a.*I6-2	AE010851 (441–1059)	1–6	—	Bact D	—	—		83%	6	17	13
*M.a.*I6-3	AE011040/1 (10510/497)	1–6	—	Bact D	—	—		83%	6	16	10
*M.a.*I6-F1	AE010902 (10030–10335)	2–6	—	Bact D	—	—		82%	1	0	0
*M.a.*F1	AE010964 (6933–8146)	5–6	RT(4–7), X	CL1	—	Conserved hypothetical protein[l]	55.3%		0	12	1
*M.a.*F2	AE010848 (4041–4124)	5–6	—	CL1	—	—	55.3%		4	5	4
*M.a.*F3	AE011134 (6345–7286)	—	RT(0–4)	CL1	—	—					
*M.a.*F4[m]	AE010979 (3887–6610)	5–6	RT(4–7), X	CL1	—	—	71.1%		0	6	3

[a]Intron names are consistent with earlier names of bacterial introns (Dai and Zimmerly 2002) and are based on the species abbreviation and a number. "F" indicates a fragment. I1-1, I1-2, and so forth denote closely related copies of the same intron. *M.a.*F1, *M.a.*F2, *M.a.*F3, and *M.a.*F4 are fragments that do not belong to the other six divisions and may be derived from other intron species.
[b]GenBank entry with the outer boundaries of the intron indicated (see Fig. 1 for internal structures).
[c]The presence of ribozyme structural domains 1–6. "—" indicates an ORF fragment without RNA structure domains.
[d]ORF domains include RT subdomains 0–7 and domain X. "—" indicates the absence of ORF domains.
[e]CL1, chloroplast-like class 1; Bact D, bacterial class D. Intron classes are based on ORF phylogenetic groupings, and each class also has a distinct RNA secondary structure (Zimmerly et al. 2001; Toor et al. 2001).
[f]Sequence immediately upstream or downstream of an intron: PQG H.S., homing site that includes the PQG motif; YADD H.S., homing site that includes YADD or FADD motif (see Fig. 1F). "—" indicates the absence or an ORF and lack of conserved sequence in the exons.
[g]% identity to *M.a.*I1-1 based on RNA domains 1–6 and excluding the ORF.
[h]% identity to *M.a.*I5-1 based on RNA domains 1–6 and excluding the ORF.
[i]The number of sequence differences that cause mispairing (M) of the secondary structure relative to either *M.a.*I1-1 or *M.a.*I5-1 (e.g., A-U to A-A).
[j]The number of sequence differences that cause a compensatory (C) variation in the secondary structure (e.g., A-U to G-C, would count as one event in the table).
[k]The number of sequence differences that are neutral (N) and retain base pairing (e.g., A-U to G-U).
[l]The sequence flanking the intron is a conserved hypothetical protein rather than a group II intron ORF. However the insertion site is somewhat conserved in sequence and an alternative reading frame includes a PQG sequence (see Fig. 1F).
[m]*M.a.*F4 may be the downstream half of *M.a.*I3-F1 because its RNA domain 4 pairs with that of *M.a.*I3-F1, and because their ORFs both have 50% identities to the *M.a.*I1-1 ORF, and together would produce a continuous ORF.

Folding of the intron RNA structures allowed us to determine the number of distinct intron species, and also to evaluate whether intron fragments are degenerate versions of other introns in the genome. Degeneration would be indicated by sequence differences that cause mispairing in the intron RNA structure (e.g., A-U to A-A), whereas functional variants would be indicated by sequences differences that cause compensatory variations (e.g., A-U to G-C). Intron RNA structures were compared with a reference intron within each grouping (*M.a.*I1–1 for chloroplast-like class 1, *M.a.*I5–1 for bacterial class D) and the sequence variations in each RNA structure were tallied as mispairing (e.g., A-U to A-A), compensatory (e.g., A-U to G-C), or neutral (e.g., A-U to G-U). Based on these criteria (Table 1), we conclude that there are six intron species, which we name *M.a.*I1–*M.a.*I6. Each of the six divisions has >85% identity among the members. Four fragments are not assigned to the six divisions, and may be derived from other intron species. Because of compensatory variations in RNA structures, all intron copies are considered functional variants, or fragments of functional variants, rather than being degenerated intron forms.

Interestingly, the introns are organized in clusters in the genome sequence, and the arrangement suggests that the introns have inserted into the ORFs of other introns to form nested organizations (Fig. 1A–D). A similar precedent for this arrangement comes from twintrons in *Euglena* chloroplasts (Copertino and Hallick 1993). In a twintron, one group II intron has inserted into another group II intron, and both must be spliced out sequentially in order to ligate together the host exons and produce a functional gene product. The situation in *M. acetivorans* is distinct from twintrons in that there are up to four introns in a nested configuration, with the "outer" introns incomplete or fragmented. Importantly, there are no ORF-encoding exons that could be ligated together to produce a functional protein after all introns are spliced out.

Figure 1A diagrams a representative group II intron used as a reference (*M.a.*I1–1), showing the basic organization of the six RNA structural domains with the RT ORF located in RNA domain 4. Figure 1B illustrates a nested organization of group II introns. The most recently inserted intron is *M.a.*I5–1, which belongs to bacterial class D. It is inserted directly after the YADD motif of another group II intron ORF. In fact, all *M.a.*I5 copies are inserted after YADD motif sequences (Fig. 1B,C,D,F), indicating that the homing site of *M.a.*I5 contains the most highly conserved positions of group II intron ORFs. The next intron is *M.a.*I1–3, which belongs to chloroplast-like class 1. It is inserted after the conserved PQG positions of RT subdomain 4, as are the other *M.a.*I1 introns (Fig. 1A–D,F), again suggesting that the *M.a.*I1 homing site includes a conserved sequence of a group II intron ORF. (*M.a.*I1–1 is an exception because it is inserted into an unrelated hypothetical conserved ORF; however, a shifted reading frame contains PQG, suggesting that *M.a.*I1–1 has inserted into a sequence with high similarity to the intron homing site; Fig. 1F.) Perplexingly, *M.a.*I1–3 is 100% identical to *M.a.*I1–1 upstream of the *M.a.*I5–1 insertion, but is only 48% identical downstream of the insertion. The upstream and downstream portions together appear to constitute an intron, because they produce a continuous ORF with expected motifs for a group II intron-encoded protein. However, as explained below, *M.a.*I1–3 (and *M.a.*I1–2) might be chimeras. The next intron in the stack, *M.a.*I1-F1, is 90% identical to *M.a.*I1–1 for the portion upstream of its intron, but there is no downstream half to complete a functional intron. Finally, the "bottom" intron is another fragment whose similarity to *M.a.*I1–1 ends at the PQG motif, as expected for a *M.a.*I1 insertion; however, there is a transposase fragment between the PQG sequence and the *M.a.*I1-F1 intron that presumably inserted into it.

Two other nested intron clusters in the genome show similar patterns. The introns in Figure 1C have a pattern essentially identical to that shown in Figure 1B, but with three introns. The cluster in Figure 1D we explain as an intron fragment (*M.a.*F4) in which a full-length *M.a.*I5 copy inserted into the YADD site, and an *M.a.*I1 intron inserted

FIGURE 1. Group II intron organization in *M. acetivorans*. (*A*) A typical group II intron structure (*M.a.*I1–1). The intron contains six helical domains (indicated by six stem-loop structures) with the ORF encoded in the loop of domain 4 (open box). The exons surrounding the intron are gray boxes, and intron splicing would ligate them together. (*B*,*C*,*D*) Nested organization of group II introns in *M. acetivorans*. There are four nested introns in *B*, and three nested introns in *C* and *D*. See text for detailed descriptions. Truncated intron copies are indicated by a truncated ORF or the absence of intron RNA domains. The percent identity (e.g., 90%, 232/258) indicates amino acid identity between that particular intron-encoded protein and the reference intron (either *M.a.*I1–1 or *M.a.*I5–1). "YADD" and "PQG" are abbreviations for homing sites that contain either the YADD or PQG motif (see *F*). (*E*) Comparison of ORF-less (*M.a.*I6) and ORF-containing (*M.a.*I5) intron structures. *M.a.*I5 and *M.a.*I6 are 82% identical. Their overall secondary structures are shown by the line drawing. Major differences are indicated, but the numerous single base differences are not. A stem-loop is absent from domain I in *M.a.*I6. There are also sequence differences in EBS1 and EBS2, which are important because they pair with IBS1 and IBS2 in the upstream exon during splicing, and also during reverse splicing into DNA exons in homing reactions. Most importantly, the ORF is deleted from domain 4 of *M.a.*I6; however, the start and stop codons remain, as well as several amino acid codons, suggesting that *M.a.*I6 was derived from *M.a.*I5 by ORF loss. (*F*) Comparison of insertion sites of ORF-less and ORF-containing introns. Insertion sites are shown for 40 bp upstream and 20 bp downstream of the intron insertion site. IBS1 and IBS2 sequences are shaded, and pair with EBS1 and EBS2 in the intron structure during splicing and reverse splicing reactions. Light shading indicates mispairings in IBS/EBS interactions due to sequence variations. Homing sites of group II introns typically extend from approximately −25 to +10 relative to the intron insertion site. The PQG and YADD amino acid codons are indicated for *M.a.*I1 and *M.a.*I5. EBS–IBS pairings between introns and exons are maintained for all four sets of introns. *M.a.*I1 and *M.a.*I4 are related introns, but their insertion sites are completely different. Likewise, *M.a.*I5 and *M.a.*I6 introns are related but have different insertion sites.

in the PQG homing site, but only the 3′ end of the intron remains (*M.a.*I1-F2).

Initially, we considered the nested organization to represent homing of introns into identical copies of the same intron. According to this initial possibility (which might be called suicide homing), the most recently inserted intron would be functional, and the less recently inserted introns would be progressively degenerate. However, based on compensatory variations in intron structures (Table 1), it instead appears that the introns are all variant, functional forms, and therefore it appears that the introns are inserting into closely related introns rather than into identical copies

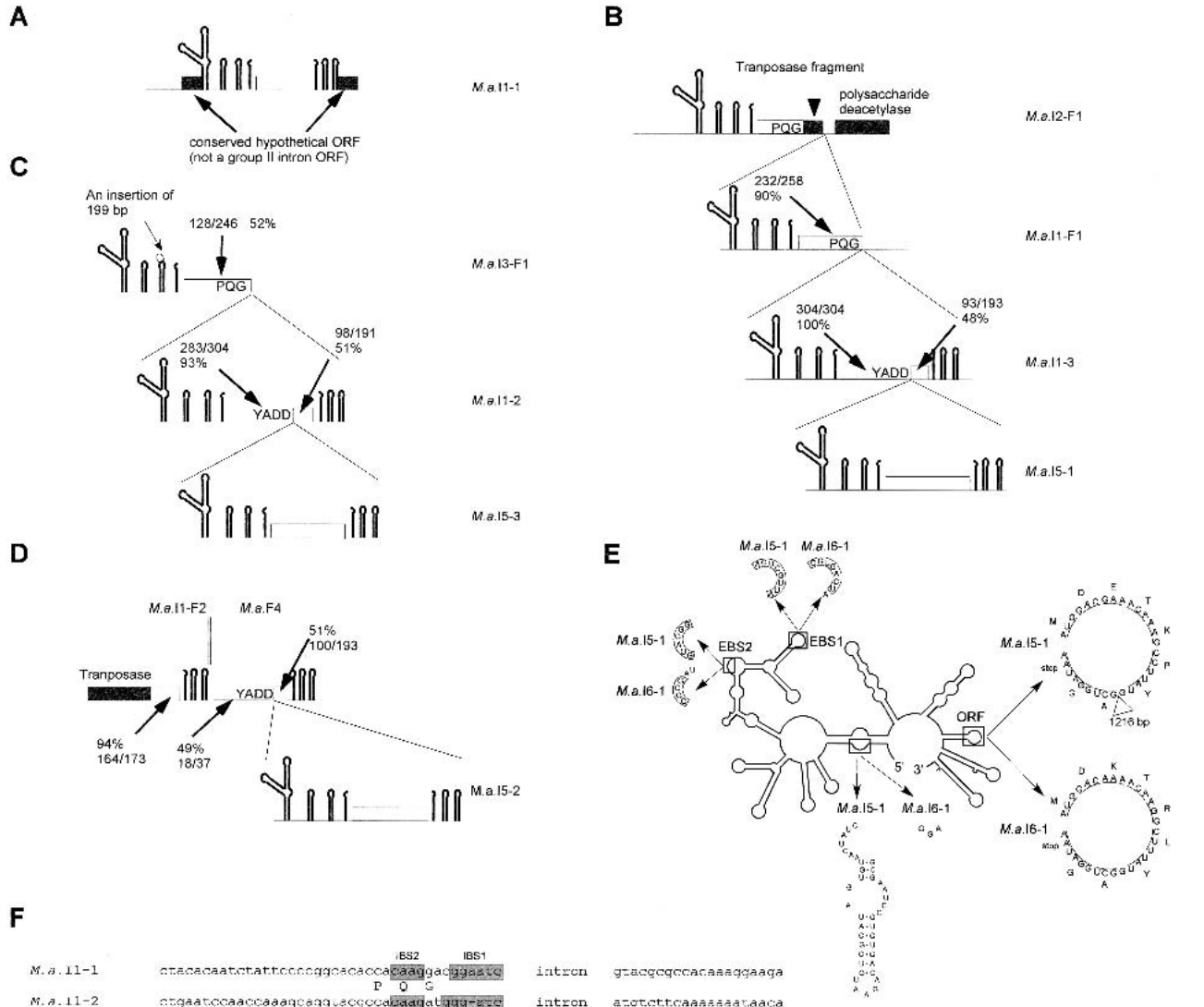

FIGURE 1. (*legend on facing page*)

of themselves. Typical homing sites for group II introns extend from about −25 to +10 relative to the insertion site, and there are a number of differences in this region among the intron copies (Fig. 1F). Thus, it is possible that the introns might be able to distinguish between their own sequences and that of closely related introns. According to this explanation, which we now prefer, intron insertions are competitive events, with each intron disrupting and probably inactivating a closely related group II intron.

The asymmetry in the nested organization is perplexing and cannot be easily explained. One possible cause is recombination between conserved sequences in adjacent introns. In yeast mitochondria, recombination between tandem intron copies is highly efficient and results in the net deletion of one intron copy and any sequence between the two copies. Such events have been observed even for introns with only 50% identity (Anziano et al. 1990). By this mechanism, if the two starting introns are identical, then the recombination event regenerates a single unchanged intron. However, if the two introns are different, then recombination produces a chimeric intron, but with continuous intron motifs. Therefore, the organization in Figure 1B could be explained as being derived from a nested organization of numerous complete introns (at least five), followed by several recombination events (at least two) to delete all 3′ intron segments except one, and to delete at least the 5′ intron segment corresponding to the remaining 3′ segment. By this explanation, the copy denoted *M.a.*I1–3 would be a chimera between an *M.a.*I1–1 copy, and the 3′ segment of an unknown intron not corresponding to any 5′ intron segments present. Extending this idea, it is possible that each cluster in Figure 1B,C,D might represent an indefinite cycle of intron homing events into other intron copies, balanced by recombinational deletions.

It should be noted that another example of nested introns is found in the cyanobacterium *Nostoc*, in which one group II intron, *N.sp.*I2, has inserted into a closely related intron, *N.sp.* I3 [GenBank accession no. AP003600 (259212–261419 for *N.sp.* I2) (258243–262762 for *N.sp.*I3)]. Like the *M. acetivorans* introns, there are no exons flanking *N.sp.*I3 that could splice together to encode a functional protein after the two introns are spliced out. Interestingly, these introns belong to chloroplast-like class 2, and like the *M.a.*I1 chloroplast-like 1 introns, are related to the *Euglena* chloroplast introns that form twintrons. Further, there is a third *Nostoc* intron, *N.sp.*F1, which is internally deleted for RT domains 2–3, including the putative homing site of *N.sp.*I2, which suggests that the internal deletion of *N.sp.*F1 might have been caused by a failed homing event by *N.sp.*I2. Therefore, the phenomenon of group II introns selfishly inserting into related introns may extend outside of archaebacteria.

Analysis of the ORF-less introns provides more surprises. *M.a.*I6 appears to be derived from *M.a.*I5, because it is 82% identical in sequence, and more importantly, domain IV of *M.a.*I6 contains remnants of the RT ORF, including the start and stop codons and several amino acids (Fig. 1E). Other differences in the *M.a.*I6 structure include the loss of a stem-loop in domain I and differences in EBS1 and EBS2, which pair with IBS1 and IBS2 during splicing and reverse splicing into DNA and, in part, determine DNA target specificity (Lambowitz et al. 1999; Bonen and Vogel 2001; Belfort et al. 2002). There are four *M.a.*I6 copies in the genome, all with somewhat different insertion sites, none of which include the YADD motif of the related *M.a.*I5 intron sites (Fig. 1F). The multiple insertion sites are evidence that *M.a.*I6 is mobile despite its lack of an intron-encoded RT protein. Interestingly, sequence identities between insertion sites of *M.a.*I6 (Fig. 1F) include the IBS1 and IBS2 region, which suggests that EBS–IBS pairings are involved in insertion specificity of the ORF-less intron.

Similarly, the ORF-less intron *M.a.*I4 is 60% identical to *M.a.*I1–1, suggesting a common ancestor, and the intron is present in multiple insertion sites in the genome, all different from that of *M.a.*I1, and all with somewhat conserved IBS1 and IBS2 sequences (Fig. 1F). Together, the observations of *M.a.*I4 and *M.a.*I6 indicate that both ORF-less introns are mobile in vivo despite lacking an intron-encoded RT ORF, and that the EBS–IBS pairings are probably involved in determining the insertion sites. It is not clear whether the two ORF-less introns can move autonomously using their intrinsic self-splicing activities, or whether they require additional proteins, perhaps the RTs encoded by related introns in the cell that may act in *trans*.

The group II introns in *M. acetivorans* are closely related to group II introns in eubacteria, with the closest relatives being *Pseudomonas putida* I2 and *Escherichia coli* I2 (44% and 55% amino acid identities, respectively, to *M.a.*I1 and *M.a.*I5 ORFs). These close relationships suggest horizontal transfers between kingdoms, in either or both directions. The large number of group II introns in *M. acetivorans* suggests that group II intron will not be rare after all in archaebacteria, and there should be many more examples identified in the future. The finding of group II introns in archaebacteria also increases the probability that group II introns might have dated back to the RNA world. Although the archael introns identified here are related to eubacterial introns and represent recent divergence, the fact that group II introns can exist in archaebacteria raises the possibility that other, more ancient introns might also be present in archaebacteria.

It is particularly exciting that ORF-less introns exist in archaebacteria. Group II introns are known to be relatively poor catalytic RNAs that require "nonphysiological" conditions of warm temperatures and high concentrations of salt and magnesium to achieve self-splicing activities in vitro (Michel and Ferat 1995). Because of these catalytic limitations, it has been assumed that all group II introns require proteins to function in vivo. However, many archaebacteria are extremophiles that grow at very high temperatures or high salt conditions, and their intracellular salt

concentrations can reach molar levels (Oren 1999). It is plausible, then, that ORF-less group II introns in such archaebacteria might function quite well by their intrinsic ribozyme activities alone, and they may not require the help of an intron-encoded protein.

As fascinating as these observations are, it is clear that we have much to learn about group II intron mobility and evolution. It is likely that considerably more information about group II introns will come to light in the future from the ongoing genome sequencing projects.

NOTE ADDED IN PROOF

The recent genome sequence of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1 also contains a mixture of 18 ORF-less group II introns and 8 ORF-containing introns (Nakamura et al. 2002). The ORF-less introns are >90% identical to the ORF-containing introns and may be derived from them. The genome contains four examples of essentially identical twintrons consisting of an ORF-less intron inserted into an ORF-containing intron, and one example of an ORF-less intron inserted into another ORF-less intron. None of these twintrons has flanking exons that encode an ORF. Another recent example of a twintron is found in *Desulfitobacterium hafniense* (GenBank accession no. NZ_AABB01000321), and again there are no ORF-encoding flanking exons.

ACKNOWLEDGMENTS

We wish to thank Navtej Toor for helping to analyze the *M. acetivorans* genome with the RNAMotif program, and an anonymous reviewer who provided a rationale for the asymmetry of the nested organization. This work was supported by the Canadian Institutes of Health Research and by the Alberta Heritage Foundation for Medical Research.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Received August 26, 2002; accepted October 7, 2002.

REFERENCES

Anziano, P.Q., Moran, J.V., Gerber, D., and Perlman, P.S. 1990. Novel hybrid maturases in unstable pseudorevertants of maturaseless mutants of yeast mitochondrial DNA. *Nucleic Acids Res.* **18:** 3233–3239.

Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B., and Lambowitz, A.M. 2002. Mobile introns: Pathways and proteins. In *Mobile DNA II* (eds. N.L. Craig et al.), pp. 761–783. ASM Press, Washington, DC.

Bonen, L. and Vogel, J. 2001. The ins and outs of group II introns. *Trends Genet.* **6:** 322–331.

Copertino, D.W. and Hallick, R.B. 1993. Group II and group III introns of twintrons: Potential relationships with nuclear pre-mRNA introns. *Trends Biochem. Sci.* **18:** 467–471.

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

Galagan, J.E., Nusbaum, C., Roy, A., Endrizzi, M.G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D. et al. 2002. The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* **12:** 532–542.

Lambowitz, A.M., Caprara, M., Zimmerly, S., and Perlman, P.S. 1999. Group I and group II ribozymes as RNPs: Clues to the past and guides to the future. In *The RNA world*, 2nd ed. (eds. R.F. Gesteland et al.), pp. 451–485. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A., and Sampath, R. 2001. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.* **29:** 4724–4735.

Michel, F. and Ferat, J.L. 1995. Structure and activities of group II introns. *Annu. Rev. Biochem.* **64:** 435–461.

Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., Watanabe, A., Iriguchi, M., Kawashima, K., Kimura, T., et al. 2002. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Research* **9:** 123–130.

Oren, A. 1999. Bioenergetic aspects of halophilism. *Micro. Mol. Biol. Rev.* **63:** 334–348.

Toor, N. and Zimmerly, S. 2002. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA* (in press).

Toor, N., Hausner, G., and Zimmerly, S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7:** 1142–1152.

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