Loss of DNA Recombinational Repair Enzymes in the Initial Stages of Genome Degeneration

Colin Dale,[1] *Ben Wang,*† *Nancy Moran,*† *and Howard Ochman**

*Department of Biochemistry and Molecular Biophysics; and †Department of Ecology and Evolutionary Biology, University of Arizona

Many obligate intracellular pathogens and symbionts undergo genome degeneration during long-term association with eukaryotic hosts; however, very little is known about genome changes that occur in the initial stages of such intracellular associations. By focusing on a clade of bacteria that have recently established symbiotic associations with insect hosts, we have identified events that may contribute to the reduction and degeneration of symbiont genomes. Unlike virtually all other bacteria, the obligate symbionts of maize and rice weevils each display substantial sequence divergence between multiple copies of their rDNA genes, resulting from a reduction in the efficacy of recombinational gene conversion, coincident with the inactivation of the recombinational repair gene *recF* in the common ancestor of both symbionts. The maize weevil endosymbiont also lacks a functional *recA*, resulting in further reduction in the efficacy of gene conversion between paralogous rDNAs and in a novel IS-mediated deletion in a 23S rDNA gene. Similar events may be pervasive during the evolution of symbiosis because symbiont genomes typically lack recombinational repair genes and have reduced numbers of ribosomal operons.

Introduction

In comparison with free-living relatives, bacteria that form permanent, obligate, intracellular associations with eukaryotic hosts often have small genomes, high A + T content, elevated mutation rates, and reduced numbers of 16S–23S rDNA operons (Andersson and Andersson 1999; Moran and Wernegreen 2000; Itoh, Martin, and Nei 2002). Moreover, these small genomes typically lack genes that are otherwise universally distributed among bacteria (Moran 2002). For example, *Buchnera aphidicola*, the obligate intracellular symbiont of aphids, has a massively reduced genome containing only 20% of ancestral genes (Moran and Mira 2001) and lacking many excision, mismatch and recombinational repair genes, including *recA*, *recF*, and the entire *uvr* operon (Shigenobu et al. 2000; Tamas et al. 2002).

When bacteria switch from an autonomous, free-living state to a permanent intracellular existence, there is relaxed selection on the maintenance of genes that are not required for intracellular survival. In addition, changes in population structure lead to an increased fixation of deleterious mutations arising from genetic drift (Moran 1996; Andersson and Kurland 1998). Comparative studies indicate that most gene loss occurs early in the evolution of an intracellular association; for example, in *B. aphidicola*, which has maintained a 250 Myr association with its insect host, genomes have been highly reduced and static for at least 50 million years (Tamas et al. 2002). For this reason, there is little information on genome changes occurring early in the evolution of an intracellular association, when the majority of gene loss occurs.

To investigate changes accompanying the initial stages of genome degeneration, we have focused on a clade of bacterial symbionts which, according to molecular evolutionary data, established symbiotic associations with their insect hosts some 50 to 100 MYA (Heddi et al. 1998; Dale et al. 2002). The facultative endosymbiont, *Sodalis glossinidius*, resides in a range of tissues within its tsetse fly host (Aksoy, Chen, and Hypsa 1997; Dale and Maudlin 1999), whereas closely related endosymbionts found in the maize and rice weevils (*Sitophilus zeamais* and *Si. oryzae*) are mutualists that only inhabit specialized organelles (bacteriomes) within their weevil hosts (Heddi et al. 1999). There is evidence of recent horizontal transmission of *So. glossinidius* between different tsetse species (Aksoy, Chen, and Hypsa, 1997), whereas the weevil endosymbionts have followed a strict pattern of exclusive maternal transmission within their hosts (Heddi et al. 1998). Both *So. glossinidius* and the weevil endosymbionts maintain genes homologous to those of the type III secretion systems found in enteric pathogens, indicative of a recent switch from parasitism to commensalism or mutualism (Dale et al. 2001, 2002). The weevil endosymbionts have not been cultured in vitro, have no official nomenclature, and are described as the *Sitophilus zeamais* primary endosymbiont (SZPE) and *Sitophilus oryzae* primary endosymbiont (SOPE).

In this study, we characterized the genome complement of 16S-23S rDNA operons in *So. glossinidius*, SOPE, and SZPE to determine the dynamics of genome reduction in these endosymbionts. We also analyzed the sequences of recombinational repair genes to understand how the inactivation of these genes might have affected recombinational gene conversion during the evolution of these closely related endosymbionts.

Materials and Methods

Symbiont Isolation and DNA Purification

So. glossinidius was grown in liquid culture, as described previously (Dale and Maudlin 1999). *Si. zeamais* and *Si. oryzae* were maintained in maize and wheat, respectively, at 27.5°C and 70% humidity. Bacteriomes (containing SZPE or SOPE) were isolated from fifth-instar *Si. zeamais* larvae by dissection. DNA was prepared from

[1] Present address: Auburn University.

Key words: recombinational repair, endosymbiont, genome degeneration, symbiosis, ribosomal RNA.

E-mail: dalecol@auburn.edu.

Mol. Biol. Evol. 20(8):1188–1194. 2003
DOI: 10.1093/molbev/msg138
Molecular Biology and Evolution, Vol. 20, No. 8,

cultured *So. glossinidius* and from weevil bacteriomes with the DNeasy tissue kit (Qiagen, Inc., Valencia, Calif.).

Amplification, Cloning, and Sequencing of 16S–23S rDNA

For this study, 5-kb fragments encompassing 16S rDNA, the intergenic spacer region (ISR) and 23S rDNA were amplified from purified symbiont DNA by polymerase chain reaction (PCR) using primers 16SF (5′-GCA CTG CAG GAT CCA GAG TTT GAT CAT GGC TCA GAT TG) and 23SR (5′-GCA GGT ACC GCG GCC GCG CTC GCG TAC CAC TTT AAA TGG CG) in a modified version of the PCR assay described by Thao et al. (2001). Polymerase chain reactions contained 5 ng of symbiont DNA, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 pmol of each primer, 1 U of *Taq* DNA polymerase and 0.3 U *Pfu* DNA polymerase. Amplification proceeded with 30 cycles of denaturation (94°C for 30 s), annealing (55°C for 2 min), and extension (70°C for 6 min), followed by a final 10 min extension at 70°C to promote the A-tailing of PCR products. The amplified 16S–23S rDNA fragments were cloned into pTOPO-XL (Invitrogen, Carlsbad, Calif.) according to the manufacturer's instructions, utilizing the *ccdB* conditional lethal selection procedure. At least 10 recombinant clones were sequenced from each template, and to ensure that every 16S and 23S rDNA gene was represented, we generated libraries of 16S rDNA sequences from each symbiont by cloning 16S rDNA PCR products amplified according to the methods described by Unterman, Baumann, and McLean (1989). Additional sequencing primers were designed as sequence information became available. Candidate tRNAs within the ISR were identified with the tRNAscan-SE software (Lowe and Eddy 1997).

Amplification, Cloning, and Sequencing of the *recA* and *recF* Genes

Partial fragments of *recA* (560 bp) and *recF* (538 bp) were amplified from *So. glossinidius*, SOPE, and SZPE with the use of universal primers (*recA*F: 5′-TNG ARA THT AYG GIC CIG ART C, *recA*R: 5′-ACN ACY TTN ACI CGI GTY TCR CT, *recF*F: 5′-MGN GCI TTY YTI GAY TGG G, and *recF*R: 5′-TCN ARY TCI GAR GCR AAR TC) in reactions containing 5 ng of purified symbiont DNA, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 pmol of each primer, 1 U of *Taq* DNA polymerase, and 0.3 U *Pfu* DNA polymerase. Amplification proceeded as described above, but with 1 min annealing and extension steps. Amplicons were cloned into pTOPO2.1 (Invitrogen) according to the manufacturer's instructions, and six recombinant clones were sequenced from each reaction. The 5′ and 3′ ends of *rec*A and *recF* were amplified with oligonucleotide primers matching adjacent genes in alignments of the *Escherichia coli* K12 and *Salmonella typhimurium* LT2 genomes. The complete *recA* and *recF* genes were amplified with primers (*recA*F1: 5′-CGT ATC GGC TCG GTG AAA GAA GG; *ygaD*R: 5′-CCT TCT TTC ACC GAG CCG ATA CG; *gyrB*F: 5′-GCA ATT TTT GCC GGA GTT TCG CC; *recF*R1: 5′-TTG ATA CTG GAG GAG TCA TAA; *recF*F1: 5′-TTG CTC CAG GCG GTA AAG AAA; *dnaN*R: 5′-GGA ACA GGA AGA AGC GGA AGA). Polymerase chain reactions and cycling conditions were the same as those used for the amplification of the internal *recA* and *recF* fragments, with extension times adjusted to 1 min/kb for products >1 kb. All products were cloned and sequenced in pTOPO2.1 (Invitrogen) as above.

Phylogenetic Analysis

The 16S rDNA and 23S rDNA coding regions from the different operons of *So. glossinidius*, SZPE, and SOPE were aligned with the sequences from *E. coli rrnH* (NC_00913) using PILEUP (Genetics Computer Group package, Madison, Wis). For each of the two genes, individual sequence changes were examined on a phylogenetic tree using MacClade 4.02 (Maddison and Maddison 1992) with the ''Trace All Changes'' option.

Results

The rDNA Operons of *So. glossinidius*, SOPE, and SZPE

We identified two distinct 16S–23S rDNA operons in *So. glossinidius* and three distinct 16S–23S rDNA operons in each of the weevil symbionts, SOPE and SZPE. No additional 16S rDNA genes were detected during comparisons of the operon sequences or during screening of libraries of polymorphic 16S rDNA sequences.

In prokaryotes, the rDNA genes are normally organized in tandem in a single operon in the order 16S–23S–5S. There is an intergenic spacer between 16S and 23S rDNA encoding tRNAs, usually either $tRNA^{Glu}$ alone or $tRNA^{Ile}$ and $tRNA^{Ala}$. The rDNA genes are often multicopy, and rDNA operon copy numbers are associated with genome size (Gurtler and Stanisich 1996; Gurtler 1999). The 16S–23S rDNA operons of SOPE, SZPE, and *So. glossinidius* each have two different ISRs encoding either $tRNA^{Glu}$ alone (ISR1) or both $tRNA^{Ile}$ and $tRNA^{Ala}$ (ISR2; fig. 1). *So. glossinidius* maintains two distinct 16S–23S rDNA operons, one with ISR1 and another with ISR2, and each of the weevil endosymbionts (SOPE and SZPE) has three distinct 16S–23S rDNA operons, one with ISR1 and two with ISR2 (fig. 1*B*). Regardless of ISR type, the sizes of corresponding ISRs are larger in both SOPE and SZPE than in *So. glossinidius*. In addition, SOPE and SZPE have 243-bp and 189-bp regions of noncoding sequence adjacent to the 23S ends of ISR type 1 and 2 (respectively) that are deleted from *So. glossinidius*.

An IS256-Like Element Located Within a 23S rDNA of SZPE

The maize weevil endosymbiont SZPE has a 1.33-kb insertion sequence (IS) element located within the coding sequence of 23S rDNA in the 16S–23S rDNA operon harboring ISR1 (fig. 1*C*). This element has an intact transposase open reading frame flanked by two inverted imperfect repeats (IIRL and IIRR) and is located 244-bp from the 3′-end of 23S rDNA. Based on the positions of conserved amino acids within the transposase and the

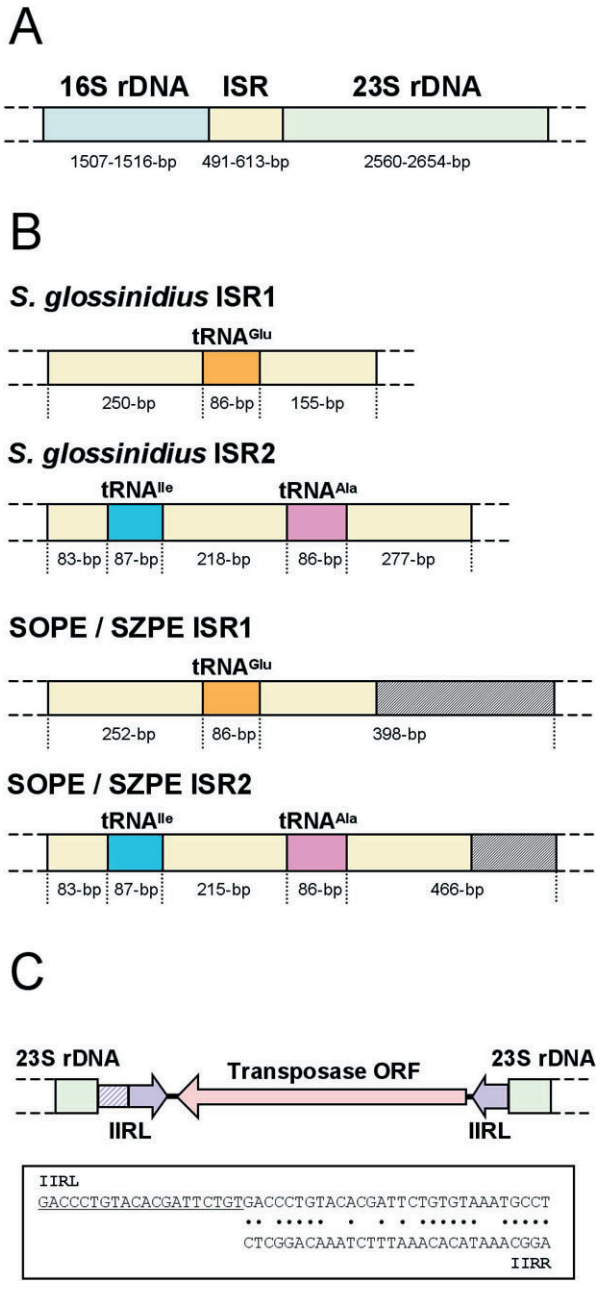

FIG. 1.—Organization of 16S–23S rDNA operons in *So. glossinidius*, SOPE, and SZPE. (*A*) Typical organization of the 16S rDNA, 23S rDNA and intergenic spacer region (ISR) in bacteria. (*B*) Location of tRNA genes within the ISRs of 16S–23S rDNA operons. *So. glossindius* maintains two 16S–23S rDNA operons with distinct ISRs encoding either $tRNA^{Glu}$ (ISR1), or $tRNA^{Ile}$ and $tRNA^{Ala}$ in tandem (ISR2). Both weevil endosymbionts (SOPE and SZPE) maintain three distinct 16S–23S rDNA operons, one with an ISR encoding $tRNA^{Glu}$ (ISR1) and two with ISRs encoding $tRNA^{Ile}$ and $tRNA^{Ala}$ in tandem (ISR2). Shaded areas in SOPE and SZPE represent conserved regions of noncoding sequence not present in *So. glossinidius*. (*C*) Organization of a 1,333-bp IS element (IS*SZPE1*) located within one of the 23S rDNA genes of SZPE. This IS element contains a single open reading frame encoding a putative transposase and is flanked by imperfect inverted repeats (IIRL and IIRR). The hatched area at the 5′ end of IIRL represents the 20-nucleotide partial duplication of IIRL, which is underlined in the boxed region containing the alignment of the two imperfect inverted repeat sequences.

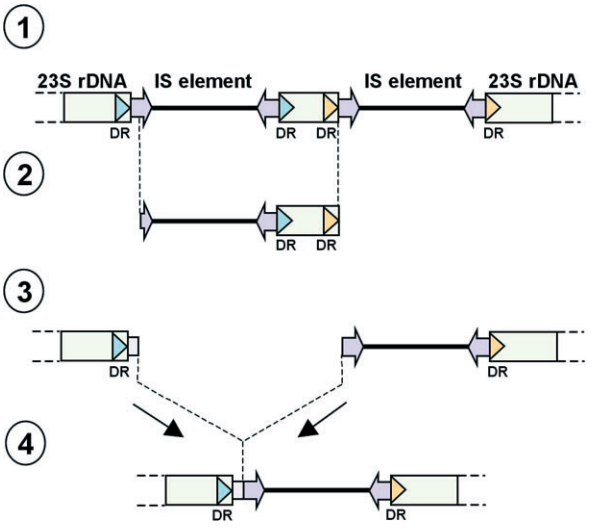

FIG. 2.—Model of insertion sequence rearrangements. Two identical IS elements with imperfect inverted repeats (purple arrows) are located 198-bp apart in the 23S rDNA gene (1). Each element is flanked by unique direct repeat (DR) sequences, generated by duplication of host DNA during insertion. An aberrant transposition (strand transfer) or duplicative intramolecular transposition causes deletion of almost one entire IS element and an attached segment of 23S rDNA (2). The adjacent strands (3) are joined to produce a hybrid IS element (4) with a partially duplicated IIRL, flanked by different DRs.

length and sequence of the imperfect inverted repeats, this IS element is a member of the IS*256* family (Mahillon and Chandler 1998), and we propose the name IS*SZPE1*. The closest matching IS element in the GenBank database is the IS*1414* element of enteropathogenic *E. coli* (McVeigh et al. 2000).

Because transposition normally occurs through a ''cut-and-paste'' process, it is surprising that one of the inverted repeat sequences flanking IS*SZPE1* is partially duplicated (fig. 1*C*). Bacterial IS elements typically produce short, perfect, direct repeats of target DNA after transposase-mediated repair of the staggered dsDNA breaks generated during insertion. For the IS*256* family of IS elements, the expected size of these direct repeats is 8-bp in length (Guedon et al. 1995; Picardeau, Bull, and Vincent 1997); however, no direct repeats are found adjacent to the imperfect inverted repeats flanking IS*SZPE1* (5′-GGC TTT GAA and 5′-CAT AAG CTA).

Both replicative and nonreplicative IS-mediated rearrangements are predicted to occur at a higher frequency when two elements are located in close proximity (Weinert, Schaus, and Grindley 1983; Mahillon and Chandler 1998). Often, these events are deletogenic, resulting in the reciprocal deletion of an IS element and an adjoining piece of DNA. By aligning the sequences of two SZPE 23S rDNA genes (including one harboring IS*SPZE1*), we identified a 198-bp deletion in 23S rDNA adjacent to IS*SZPE1*. Combined with the fact that IS*SZPE1* has a duplicated imperfect inverted repeat and no direct repeats, these results suggest that IS*SZPE1* is a hybrid molecule, generated from the deletogenic rearrangement of two identical IS elements (fig. 2).

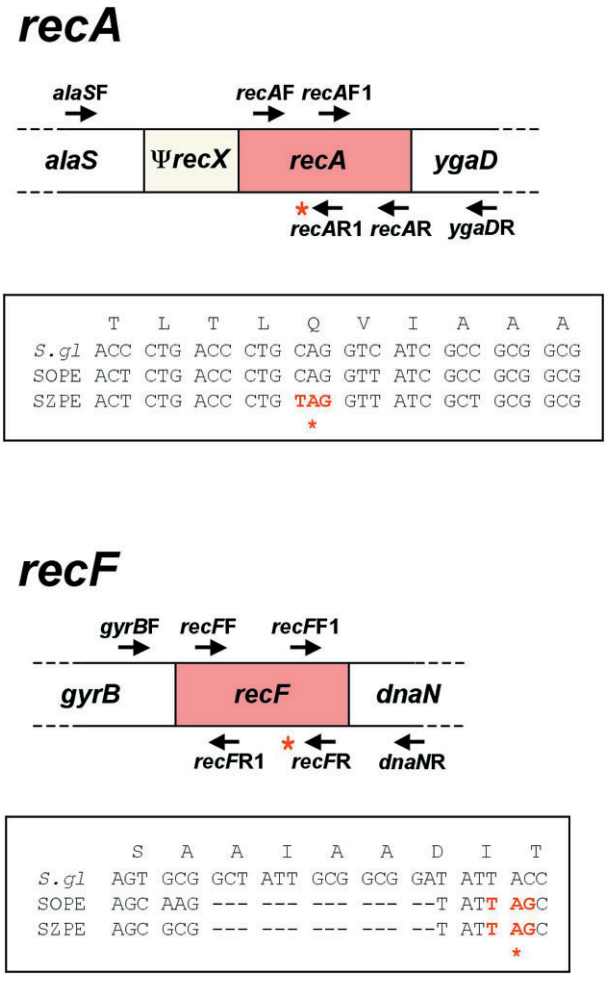

FIG. 3.—Organization of *recA* and *recF* in *So. glossinidius* (*S.gl*), SOPE, and SZPE. Gene order is conserved in all three symbionts and is identical to that found in *E. coli*. All three symbionts maintain pseudogenes with sequence similarity to *recX* (shaded yellow). Locations of PCR primers used for amplification of *recA* and *recF* are shown above and below the gene schematics. Partial alignments of the symbiont *recA* and *recF* genes are boxed, illustrating the positions of frameshift deletions and stop codons (red asterisks). Conceptual amino acid translations from the *So. glossinidius* sequences are presented at the top of each box.

Loss of the *rec* Genes in the Weevil Endosymbionts

Although the organization of the *recA*, *recF* and flanking genes in *So. glossinidius*, SOPE, and SZPE is essentially the same as that described for *Escherichia coli* and *Salmonella* spp. (fig. 3), both of the weevil endosymbionts have an identical 14-bp frameshift deletion located in the coding sequence of *recF*. Given that this deletion occurs in SOPE and SZPE, but not in *So. glossinidius*, it likely occurred in the last common ancestor of the weevil endosymbionts, after they split from *So. glossinidius*.

The complete coding sequence of *recA* was obtained by amplifying the region corresponding to *alaS–recX–recA–ygaD* in *Sa. typhimurium*. Neither *So. glossinidius*, SOPE, nor SZPE harbors a full-length *recX* gene. In some bacterial species, the RecX protein plays a role in regulating RecA activity by potentiating RecA-driven homologous recombination and allowing high-level expression of RecA (Stohl and Seifert 2001; Sukchawalit et al. 2001).

Whereas both *So. glossinidius* and SOPE maintain full-length and intact *recA* genes, the coding sequence of *recA* in SZPE contains a nonsense mutation at position 235 (fig. 3*A*). Because the *recA* mutation is restricted to the maize weevil endosymbiont, it occurred after the divergence of the maize and rice weevils and, thus, after the acquisition of the 14-bp frameshift deletion in *recF*.

Table 1
Percentage Pairwise Distances Between 16S (Numerator) and 23S (Denominator) rDNA Sequences of *So. glossinidius*, SOPE, and SZPE

	$SG^{Ile/Ala}$	SZ^{Glu}	$SZ^{Ile/Ala1}$	$SZ^{Ile/Ala2}$	SO^{Glu}	$SO^{Ile/Ala1}$	$SO^{Ile/Ala2}$
SG^{Glu}	0.2	1.7	2.1	2.0	1.6	2.1	2.4
	0.2	2.5	2.1	1.8	1.9	2.0	1.8
$SG^{Ile/Ala}$		1.8	2.1	2.1	1.7	2.1	2.5
		2.5	2.1	1.9	2.0	2.0	1.9
SZ^{Glu}			**1.7**	**1.7**	0.8	1.7	2.1
			2.1	**1.8**	1.5	2.0	1.9
$SZ^{Ile/Ala1}$				**1.9**	1.6	1.7	1.5
				1.3	1.4	1.1	1.0
$SZ^{Ile/Ala2}$					1.6	1.9	2.1
					1.2	1.2	1.0
SO^{Glu}						**1.1**	**2.0**
						1.3	**1.2**
$SO^{Ile/Ala1}$							**1.1**
							0.1

NOTE.—SG, *Sodalis glossinidius*; SZ, SZPE; SO, SZPE. 16S–23S rDNA operons within species are differentiated according to the tRNA genes encoded within the ISR (Glu or Ile/Ala). In SOPE and SZPE, the two distinct 16S–23S rDNA operons with ISRs harboring $tRNA^{Ile}$ and $tRNA^{Ala}$ are distinguished numerically. Note that interoperon pairwise distances within each of the weevil endosymbionts (bold) are much larger than in *So. glossinidius*.

Evolution of the rDNA Operons in *So. glossinidius*, SOPE, and SZPE

Pairwise nucleotide sequence comparisons indicate that SOPE and SZPE each have unusually high levels of interoperon divergence between paralogous copies of 16S and 23S rDNA (table 1). To trace the evolutionary history of the 16S and 23S rDNA sequences in *So. glossinidius*, SOPE, and SZPE, we examined this sequence variation in a phylogenetic context. Initially, phylogenetic trees were constrained to reflect the divergence between SOPE and SZPE preceding any additional divergence between sequences of different operons within these symbionts. This would be anticipated if the rDNA coding sequences from different operons were being homogenized by gene conversion. After mapping base changes onto branches of the trees, we identified eight nucleotide substitutions that were maintained only in the rDNA coding sequences of operons harboring $tRNA^{Glu}$ in both SOPE and SZPE, indicating that rDNA coding sequences in SOPE and SZPE began diverging before speciation of the weevil symbionts.

To take account of this pattern of evolution, we constructed phylogenetic trees assuming that divergence of the individual operons began before the split between SOPE and SZPE (fig. 4). The resulting trees show that

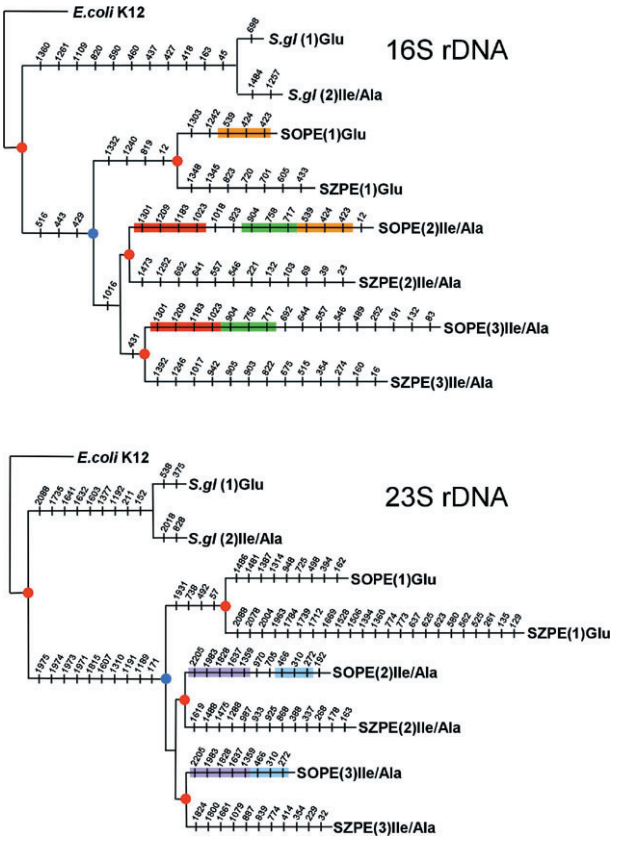

FIG. 4.—Phylogenetic trees based on 16S and 23S rDNA sequences of paralogous rDNA operons in *So. glossinidius*, SOPE and SZPE. Trees are constrained to reflect the fact that divergence of the individual operons began before divergence of SOPE and SZPE. Base changes are mapped onto branches of the trees illustrating the presence of identical sequence tracts with conserved substitutions (colored bars) within paralogous rDNA genes of SOPE. The node highlighted in blue denotes the onset of interoperon rDNA sequence divergence, and the nodes highlighted in red indicate speciation events among *So. glossinidius*, SOPE, and SZPE.

So. glossinidius diverged from a common ancestor of the weevil symbionts before the onset of any interoperon divergence between rDNA sequences in SOPE and SZPE. Like most bacteria, *So. glossinidius* maintains paralogous 16S and 23S rDNA genes with near identical sequences. In contrast, SOPE and SZPE have paralogous 16S and 23S rDNA sequences that display an unusually high level of divergence (table 1). Because it is known that gene conversion drives the concerted evolution of genic sequences within the rDNA operons in bacteria (Liao 2000), our results indicate that there are differences in the efficacy of gene conversion in *So. glossinidius*, SOPE, and SZPE.

To determine if there has been any recent gene conversion between paralogous rDNA sequences within SOPE and SZPE, we mapped base changes onto branches of the 16S and 23S rDNA trees and searched for regions of shared homology within paralogous sequences. We failed to identify any evidence of recent gene conversion in SZPE but found several sequence tracts in two of the three paralogous rDNA genes of SOPE that share identical consecutive nucleotide substitutions. Thus, in comparison with *So. glossinidius* and other bacteria, there is a reduced frequency of gene conversion in SOPE, insufficient to drive concerted evolution of the rDNA sequences. Furthermore, our results indicate that there is no ongoing gene conversion between rDNA sequences in SZPE.

Discussion

Most obligate intracellular pathogens and symbionts studied to date are in the advanced stages of genome degeneration, as evidenced by their vastly reduced genome sizes, high A + T contents, and elevated rates of polypeptide evolution. In contrast, the genomes of *So. glossinidius*, SOPE, and SZPE provide insights into the early stages of the degenerative evolutionary processes following the transition from free-living to an intracellular lifestyle. In this study, we identified a link between the inactivation of two DNA recombinational repair enzymes (RecF and RecA) and the loss of recombinational gene conversion, occurring early in the evolution of mutualistic associations.

Several additional lines of evidence indicate that degenerative evolution is at an early stage in *So. glossinidius*, SOPE, and SZPE. First, the genome sizes of *So. glossinidius* and SOPE, estimated at 2 mb and 3 mb, respectively (Charles et al. 1997; Akman et al. 2001), are smaller than those of related free-living bacteria like *E. coli* (4.5–5.3 mb) but substantially larger than those of many other insect endosymbionts such as *B. aphidicola* (0.45–0.65 mb; Charles and Ishikawa 1999; Gil et al. 2002), *Wigglesworthia glossinidia* (0.75 mb; Akman and Aksoy 2001), *Blochmannia* spp. (0.8-mb; Wernegreen, Lazarus, and Degnan 2002) and *Wolbachia* spp. (0.95–1.66 mb; Sun et al. 2001). Second, the genomic base compositions of *So. glossinidius*, SOPE, and SZPE have been estimated to be equitable with those of free-living bacteria, and there is no evidence of an A + T bias in sequenced genes (Heddi et al. 1998; Dale and Maudlin 1999). Finally, the relative rates of synonymous and nonsynonymous nucleotide substitution in *So. glossinidius* and SZPE are similar to those observed for free-living bacteria (Dale et al. 2002).

Comparative evolutionary analyses of protein-coding genes indicate that *So. glossinidius*, SOPE, and SZPE shared a common ancestor about 100 MYA (Dale et al. 2002). After diverging from *So. glossinidius*, the last common ancestor of SOPE and SZPE established an obligate symbiosis with weevils (Heddi et al. 1998). Prior to the speciation of the rice and maize weevils (*Si. oryzae* and *Si. zeamais*), there was a 14-bp frameshift deletion in the *recF* of the ancestor of SOPE and SZPE, and subsequent to the divergence of the rice and maize weevils, SZPE acquired a nonsense mutation in *recA*. Thus, *So. glossinidius* retains intact copies of both *recF* and *recA*, whereas SOPE lacks a functional *recF* (but retains *recA*), and SZPE lacks functional copies of both *recF* and *recA*. In *E. coli*, recombinational repair is mediated by two pathways involving either the RecBCD or RecFOR protein complexes, both of which require RecA (Kowalczykowski 2000). In *E. coli*, experimental evidence indicates that the RecBCD pathway is more important than is the RecFOR

pathway in conjugal recombination and transduction. However, in replication fork repair and in other major recombination functions, the RecBCD and RecFOR pathways both play critical roles (Galitski and Roth 1997; Lovett et al. 2002).

In SOPE and SZPE, the sequential inactivation of both *recF* and *recA* affects rates of gene conversion, as evident from sequence analysis of multicopy rRNA genes within each endosymbiont genome. In virtually all bacterial species, there is little or no interoperon divergence between sequences of 16S and 23S rDNA, because paralogous gene copies are homogenized by ectopic recombination (Liao 2000). Coincident with the inactivation of the *recF* gene in a common ancestor of the weevil endosymbionts SOPE and SZPE, there was a substantial reduction in gene conversion between paralogous rDNA genes. Intragenomic divergence of the rDNA paralogs is substantially higher in SOPE and SZPE, which lack *recF*, than in *So. glossinidius* (table 1) as a result of the reduction in the frequency of recombinational gene conversion in the two weevil endosymbionts. The subsequent loss of *recA* in SZPE appears to have further reduced the frequency of gene conversion in this species, and there is no evidence of gene conversion between paralogous rDNA genes in SZPE.

In SZPE, the inactivation of both *recF* and *recA* is also linked to the presence of a novel IS-mediated deletion in one copy of the 23S rDNA gene, resulting from an intramolecular rearrangement between two adjacent, identical IS*256*-like elements. During random sequencing, we have identified an additional copy of the IS*SZPE1* located on a 134-kbp extrachromosomal element maintained by SZPE (data not shown). This plasmid-borne copy of IS*SZPE1* probably served as a donor for transposition of IS*SZPE1* into the SZPE chromosome. In hosts lacking *recA*, IS elements are known to undergo duplicative intramolecular rearrangements, promoting deletion of adjacent IS elements and host DNA (Weinert, Schaus, and Grindley 1983). This mechanism could promote genome degeneration in chronic intracellular pathogens and symbionts that have lost components of their DNA recombinational repair machinery. The presence of a 1.33-kb IS element and the loss of almost 200 bp of coding sequence from the 23S rDNA gene likely prevents the correct folding and functioning of this rRNA subunit. In bacteria with active recombinational repair pathways, we would expect IS elements to be purged from multicopy rRNA genes during interoperon gene conversion.

Because it is known that many obligate intracellular pathogens and symbionts lose components of their DNA recombinational repair machinery, we might question whether such losses occur passively as a consequence of relaxed selection, or actively as part of an adaptive response toward life in the intracellular environment. The latter hypothesis suggests that loss of the DNA recombinational repair genes occurs as a result of a large-scale reduction in genome size. However, our results contradict this hypothesis, because the mutational inactivation of *recF* in the weevil endosymbionts occurred at an early stage in the evolution of the intracellular association, independent of massive gene loss. Furthermore, the *recF* pseudogene has been retained by SOPE and SZPE for at least 30 Myr, since the divergence of the maize and rice weevils.

Based on the available evidence, we conclude that the *recA and recF* genes were inactivated in the weevil endosymbionts as a consequence of relaxed selection, but it is interesting to note that this has not occurred in *So. glossinidius*. This could be attributed to differences in the nature of interactions between these endosymbionts and their hosts, especially given the fact that *So. glossinidius* is a facultative symbiont, residing both intra- and extracellularly in host tissues, whereas the weevil symbionts are exclusively intracellular. In free-living bacteria, the primary role of the recombinational repair machinery is the nonmutagenic restoration and repair of stalled replication forks (Lusetti and Cox 2002). As a result of the protective nature of their environment, obligate intracellular pathogens and symbionts are likely to experience a lower incidence of DNA damage, reducing their requirement for recombinational repair.

Acknowledgments

The authors thank Alvin J. Clark and Ted Weinert for providing helpful comments. We also thank Serap Aksoy (Yale University) for providing *So. glossinidius* cultures and Jim Baker (USDA, Manhattan, Kans.) for providing *Si. zeamais* and *Si. oryzae*. This work was supported by National Institutes of Health grant GM56120 to H.O. and National Science Foundation grant DEB-9978518 to N.M.

Literature Cited

Akman, L., and S. Aksoy. 2001. *Escherichia coli* gene array analysis provides insight into the biology of the obligate endosymbiont of tsetse flies, *Wigglesworthia glossinidia*. Proc. Natl. Acad. Sci. USA **98**:7546–7551.

Akman, L., R. V. Rio, C. B. Beard, and S. Aksoy. 2001. Genome size determination and coding capacity of *Sodalis glossinidius*, an enteric symbiont of tsetse flies, as revealed by hybridization to *Escherichia coli* gene arrays. J. Bacteriol. **183**:4517–4525.

Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, and S. Aksoy. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. **32**:402–407.

Aksoy, S., X. Chen, and V. Hypsa. 1997. Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera:Glossinidae). Insect Mol. Biol. **6**:183–190.

Andersson, J. O., and S. G. E. Andersson. 1999. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. **9**:664–671.

Andersson, S. G. E., and C. G. Kurland. 1998. Reductive evolution of resident genomes. Trends Microbiol. **6**:263–268.

Charles, H., G. Condemine, C. Nardon, and P. Nardon. 1997. Genome size characterisation of the principal endocellular symbiotic bacteria of the weevil *Sitophilus oryzae*, using pulsed field gel electrophoresis. Insect Biochem. Mol. Biol. **27**:345–350.

Charles, H., and H. Ishikawa. 1999. Physical and genetic map of the genome of *Buchnera*, the primary endosymbiont of the pea aphid *Acyrthosiphon pisum*. J. Mol. Evol **48**:142–150.

Dale, C., and I. Maudlin. 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary

endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int. J. Syst. Bacteriol **49**:267–275.

Dale, C., G. R. Plague, B. Wang, H. Ochman, and N. A. Moran. 2002. Type III secretion systems and the evolution of mutualistic endosymbiosis. Proc. Natl. Acad. Sci. USA **99**: 12397–12402.

Dale, C., S. A. Young, D. T. Haydon, and S. C. Welburn. 2001. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc. Natl. Acad. Sci. USA **98**:1883–1888.

Galitski, T., and J. R. Roth. 1997. Pathways for homologous recombination between chromosomal direct repeats in *Salmonella typhimurium*. Genetics **146**:751–767.

Gil, R., B. Sabater-Munoz, A. Latorre, F. J. Silva, and A. Moya. 2002. Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. Proc. Natl. Acad. Sci. USA **99**:4454–4458.

Guedon, G., F. Bourgoin, M. Pebay, Y. Roussel, C. Colmin, J. M. Simonet, and B. Decaris. 1995. Characterization and distribution of two insertion sequences, IS1191 and iso-IS981, in *Streptococcus thermophilus*: does intergeneric transfer of insertion sequences occur in lactic acid bacteria co-cultures? Mol. Microbiol. **16**:69–78.

Gurtler, V. 1999. The role of recombination and mutation in 16S–23S rDNA spacer rearrangements. Gene **238**:241–252.

Gurtler, V., and V. A. Stanisich. 1996. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. Microbiology **142**:3–16.

Heddi, A., A. M. Grenier, C. Khatchadourian, H. Charles, and P. Nardon. 1998. Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G+C content of an endocytobiotic DNA. J. Mol. Evol. **47**: 52–61.

———. 1999. Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. Proc. Natl. Acad. Sci. USA **96**:6814–6819.

Itoh, T., W. Martin, and M. Nei. 2002. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. Proc. Natl. Acad. Sci. USA **99**: 12944–12948.

Kowalczykowski, S. C. 2000. Initiation of genetic recombination and recombination-dependent replication. Trends. Biochem. Sci. **25**:156–165.

Liao, D. 2000. Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. J. Mol. Evol. **51**:305–317.

Lovett, S. T., R. L. Hurley, V. A. Sutera, R. H. Aubuchon, and M. A. Lebedeva. 2002. Crossing over between regions of limited homology in *Escherichia coli*: RecA-dependent and RecA-independent pathways. Genetics **160**:851–859.

Lowe, T. M., and S. R. Eddy, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Lusetti, S. L., and M. M. Cox. 2002. The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. Annu. Rev. Biochem. **71**:71–100.

Maddison, W. P., and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Mass.

Mahillon, J., and M. Chandler. 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62**:725–774.

McVeigh, A., A. Fasano, D. A. Scott, S. Jelacic, S. L. Moseley, D. C. Robertson, and S. J. Savarino. 2000. IS1414, an *Escherichia coli* insertion sequence with a heat-stable enterotoxin gene embedded in a transposase-like gene. Infect. Immun. **68**:5710–5715.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93**: 2873–2878.

———. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell **108**:583–586.

Moran, N. A., and A. Mira. 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome Biol. **2**:RESEARCH0054.

Moran, N. A., and J. J. Wernegreen. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends. Ecol. Evol. **15**:321–326.

Picardeau, M., T. J. Bull, and V. Vincent. 1997. Identification and characterization of IS-like elements in *Mycobacterium gordonae*. FEMS Microbiol. Lett. **154**:95–102.

Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature **407**:81–86.

Stohl, E. A., and H. S. Seifert. 2001. The *recX* gene potentiates homologous recombination in *Neisseria gonorrhoeae*. Mol. Microbiol. **40**:1301–1310.

Sukchawalit, R., P. Vattanaviboon, S. Utamapongchai, G. Vaughn, and S. Mongkolsuk. 2001. Characterization of *Xanthomonas oryzae* pv. Oryzae *recX*, a gene that is required for high-level expression of *recA*. FEMS Microbiol. Lett. **205**:83–89.

Sun, L. V., J. M. Foster, G. Tzertzinis, M. Ono, C. Bandi, B. E. Slatko, and S. L. O'Neill. 2001. Determination of *Wolbachia* genome size by pulsed-field gel electrophoresis. J. Bacteriol. **183**:2219–2225.

Tamas, I., L. Klasson, B. Canback, A. K. Naslund, A. S. Eriksson, J. J. Wernegreen, J. P. Sandstrom, N. A. Moran, and S. G. E. Andersson. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science **296**:2376–2379.

Thao, M. L., M. A. Clark, L. Baumann, E. B. Brennan, N. A. Moran, and P. Baumann. 2001. Secondary endosymbionts of psyllids have been acquired multiple times. Curr. Microbiol. **41**:300–304.

Unterman, B. M., P. Baumann, and D. L. McLean. 1989. Pea aphid symbionts relationships established by analysis of 16S rRNAs. J. Bacteriol. **171**:2970–2974.

Weinert, T. A., N. A. Schaus, and N. D. F. Grindley. 1983. Insertion sequence duplication in transpositional recombination. Nature **222**:755–765.

Wernegreen, J. J., A. B. Lazarus, and P. H. Degnan. 2002. Small genome of Candidatus *Blochmannia*, the bacterial endosymbiont of *Camponotus*, implies irreversible specialization to an intracellular lifestyle. Microbiology **148**:2551–2556.

Thomas Eickbush, Associate Editor

Accepted March 24, 2003