

Comparative Genomics of Insect-Symbiotic Bacteria: Influence of Host Environment on Microbial Genome Composition

Rita V. M. Rio,¹ Cedric Lefevre,² Abdelaziz Heddi,² and Serap Aksoy^{1*}

Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut 06510,¹ and UMR INRA/INSA de Lyon—Biologie Fonctionnelle, Insectes et Interactions, 69621 Villeurbanne Cedex, France²

Received 8 April 2003/Accepted 6 August 2003

Commensal symbionts, thought to be intermediary amid obligate mutualists and facultative parasites, offer insight into forces driving the evolutionary transition into mutualism. Using macroarrays developed for a close relative, *Escherichia coli*, we utilized a heterologous array hybridization approach to infer the genomic compositions of a clade of bacteria that have recently established symbiotic associations: *Sodalis glossinidius* with the tsetse fly (Diptera, *Glossina* spp.) and *Sitophilus oryzae* primary endosymbiont (SOPE) with the rice weevil (Coleoptera, *Sitophilus oryzae*). Functional biologies within their hosts currently reflect different forms of symbiotic associations. Their hosts, members of distant insect taxa, occupy distinct ecological niches and have evolved to survive on restricted diets of blood for tsetse and cereal for the rice weevil. Comparison of genome contents between the two microbes indicates statistically significant differences in the retention of genes involved in carbon compound catabolism, energy metabolism, fatty acid metabolism, and transport. The greatest reductions have occurred in carbon catabolism, membrane proteins, and cell structure-related genes for *Sodalis* and in genes involved in cellular processes (i.e., adaptations towards cellular conditions) for SOPE. Modifications in metabolic pathways, in the form of functional losses complementing particularities in host physiology and ecology, may have occurred upon initial entry from a free-living to a symbiotic state. It is possible that these adaptations, streamlining genomes, act to make a free-living state no longer feasible for the harnessed microbe.

Symbiosis, a term coined by Anton de Bary in 1879, refers to two (or more) species that live in close association with each other, typically with one on or within the body of the other. Symbiotic relations are known to be extremely pervasive, with examples being found throughout the earth's biota. It is estimated that the majority of the arguably largest class of invertebrates, Insecta, are involved in some type of symbiosis, with the majority of these relations being shared with bacteria. These associations form a dynamic spectrum with regards to the necessity as well as the physiological impact towards those involved.

Obligate mutualists of insects are thought to enable their hosts to survive on restrictive diets, typically consisting of a single food source, by provisioning nutritional supplements such as amino acids and B vitamins (10, 40). Symbiont loss often results in detrimental fitness costs for the host, such as sterility, growth impairment, and shortened life spans (40). Congruence of host and obligate mutualist phylogenies, dating back to ancient times, has been demonstrated for a number of insect systems, including *Buchnera* and aphids, *Wigglesworthia* and tsetse flies, *Blochmannia* and ants, *Carsonella* and psyllids, and *Blattobacterium* and cockroaches (recently reviewed in reference 52). Many of these arthropod-associated mutualists form distinct but related lineages in the γ -proteobacteria, with genome diversification presumably having occurred upon respective symbiotic establishments. Analyses of obligates reveal

drastically reduced genomes, such as the 450 to 653 kb of *Buchnera*, the 697 kb of *Wigglesworthia*, the 800 kb of *Blochmannia*, and the 680 kb of *Baumannia* (37, 52). These reduced genomes are reflective of tight associations with host physiology and ecology such that pathways for superfluous compounds or compounds deemed beneficial but unnecessary are eliminated (35, 36). Among other hallmarks of these genomes are increased genetic drift, accelerated sequence evolution, and adenine-thymine (AT) bias (6, 36).

In contrast to the well-studied obligates, less is known about the functions and evolution of commensal genomes. These symbionts apparently are recent inhabitants (4, 12), with much broader tissue tropism in their hosts (i.e., variations among host sex and age) and a patchier occurrence within populations (12, 48) and different host species (14). While commensals are found to be maternally transmitted (10, 12, 14, 18), horizontal transfer among related host species is common (4, 12, 17, 43), as evidenced by the lack of congruity between symbiont and host genomes. Recent works have noted some benefits for host biology arising from commensals, such as temperature tolerance (13, 34) and increased resistance against parasitoid development (41) in aphids. It has also been suggested that commensals may influence traits such as host susceptibility towards disease (28) and the transmission of other microbes, such as trypanosome infection in tsetse flies (51). Thus, commensals can be considered intermediates between the highly evolved mutualist, strictly vertically transmitted and indispensable for its host, and the facultative parasite, whose horizontal modes of transmission have typically been associated with virulence (18).

A group of bacteria, characterized from distant insect orders

* Corresponding author. Mailing address: Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College St., 606 LEPH, New Haven, CT 06510. Phone: (203) 737-2180. Fax: (203) 785-4782. E-mail: serap.aksoy@yale.edu.

such as Diptera, Coleoptera, and Hemiptera, form a distinct lineage within the γ -proteobacteria in close association with pathogenic and free-living enteric microbes. Here, we compare the genomic makeup of two of the symbionts within this lineage: *Sodalis glossinidius*, a symbiont with the tsetse fly (*Glossina* spp.; Diptera, Glossinidae), and *Sitophilus oryzae* primary endosymbiont (SOPE), a symbiont with the rice weevil (*Sitophilus oryzae*; Coleoptera, Dryophthoridae). Both of these recently established symbionts (4, 24) show no A+T bias in several of their sequenced genes (2, 24). Although their genomes, 2.0 Mb for *Sodalis* (2) and 3.0 Mb for SOPE (11), have been reduced in size in comparison to free-living relatives, they are significantly larger than those of obligates. Their hosts, members of distant insect taxa, occupy distinct ecological niches and have evolved to survive on specialized diets (i.e., blood for tsetse flies and cereal for rice weevils). In addition to *Sodalis*, tsetse flies harbor an ancient obligate mutualist, *Wigglesworthia*, and members of both insect host taxa have been invaded by the parasitic *Wolbachia* (25, 42). The functional biology of the symbionts within their respective hosts suggests that *Sodalis* fulfills a commensalist role while SOPE has acquired obligate traits, since its elimination negatively affects the energy-dependent activities of its weevil host (21, 25, 39). Furthermore, while *Sodalis* lives both within insect cells and extracellularly in hemolymph (14), SOPE appears confined to specialized cells within its host (25, 26). It has been possible to cultivate *Sodalis* in vitro in cell-free medium (50), while attempts to cultivate SOPE have failed (A. Heddi, unpublished data), further supporting their symbiotic statuses. Discoveries of invasion mechanisms, such as a type III secretion system (15, 16), in these symbionts suggest that parasitism, through an attenuated adaptive process arising from coevolution, may have developed into beneficial relationships (15, 16, 20, 27, 46). Thus, these symbionts can serve as models for studying the preliminary impact of a lifestyle shift, from free-living to symbiotic, on microbial genome composition.

The goal of this study was to gain insight into forces driving the evolutionary transition to symbiosis and their consequences on the genomic composition of microbial participants. The symbiont genomes were comparatively analyzed by a heterologous array hybridization approach using *Escherichia coli* macroarrays in attempts to understand how host ecology may influence selection for adaptations in microbial genomes and how initial genome reductions may harness organisms into symbiotic lifestyles. The information obtained for SOPE was compared to previous results obtained with *Sodalis* employing a similar technique (2). We discuss the divergence observed for the two sister bacterial genomes in light of their respective unique niches.

MATERIALS AND METHODS

Microorganisms. *Sitophilus oryzae* weevils (Chinese strain) were reared on wheat at 27.5°C and 70% relative humidity. Bacteriomes were dissected from fourth instar larvae, and total SOPE DNA was prepared as previously described (22). Tsetse flies, *Glossina morsitans* subsp. *morsitans*, were maintained at 23°C with 60% relative humidity and received defibrinated bovine blood through an artificial membrane system. The symbiont, *Sodalis*, was obtained from pupae and cultured on a layer of C6/36 cells as described previously (50). Purification of *Wolbachia* from *Drosophila melanogaster* (Canton-S) and DNA extraction were performed as previously described (9).

DNA hybridization to *E. coli* arrays. For genome comparison studies, the *E. coli* K-12 gene array, which contains the 4,290 PCR-amplified open reading frames (ORFs) identified in the sequenced genome (8) spotted onto nylon membranes (Panorama macroarrays; Genosys Biotechnologies Inc.), was used. Each ORF is spotted in duplicate over three panels for control purposes. For hybridization probes, symbiont DNA was radioactively labeled with [α -³²P]ATP (ICN, La Jolla, Calif.) by use of a Pol I/DNase I nick translation kit (Gibco). Hybridization reactions were performed in 45% formamide–5 \times Denhardt's solution–5 \times SSC (1 \times SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.5% sodium dodecyl sulfate (SDS) buffer at 42°C as described previously (2). The arrays were washed at 42°C for 30 min in 2 \times SSC–0.1% SDS and 0.1 \times SSC–0.1% SDS followed by 0.1 \times SSC–0.5% SDS. Arrays were exposed to maximum resolution films (BMR; Eastman Kodak Company), and signals were scored on the basis of strong, medium, or weak hybridization intensities (six-, four-, and twofold above background noise, respectively, as verified by density readings obtained by utilizing a Kodak Image Station 2000R system) as described previously (2). There were no cases for which duplicate spots gave contradictory results. Hybridization with *Wolbachia* DNA was performed under the same conditions to evaluate potential contamination and therefore its contribution to the signals detected with SOPE DNA. There was negligible hybridization with *Wolbachia* DNA, confirming that results obtained with SOPE DNA represent true homologs.

Array and statistical analysis. ORFs were identified according to array specifications and entered into spreadsheets (Microsoft Excel) (available online [http://info.med.yale.edu/eph/html/faculty/aksoy/]; upon request, they can be provided electronically). Macros were created to identify shared as well as unique genes between *Sodalis* and SOPE based on two independent sets of search criteria, gene name and DBGET numerical designation. Genes were compiled into functional categories as provided by Genosys software, and potential metabolic pathways were reconstructed by use of the Kyoto Encyclopedia of Genes and Genomes web site (http://www.genome.ad.jp/kegg/). Only the metabolic pathways for which complete gene sets could be detected were included in the comparative functional analyses. A two-tailed Fisher's exact test was performed with the SAS system for Windows (30) to compare gene retention (i.e., number of genes detected) per functional category between SOPE and *Sodalis*. The effect of symbiont species on the proportion of genes retained with *E. coli* homologs was tested in these analyses. Significance is reported at the 5% level.

RESULTS

Symbiont genome architecture. A heterologous array hybridization approach was used to infer the composition of the symbiont genomes. The *E. coli* macroarray contains 4,290 ORFs, representing its known genome, and functional roles have been assigned to 1,962 of these. A total of 2,084 orthologs could be detected with SOPE DNA, hybridizing with strong (10%), medium (37%), and weak (53%) hybridization intensities. These orthologs could account for approximately 70% of the 3.0-Mb SOPE genome (11), assuming an average size of 1 kb per gene. Of the 2,084 homologs detected in the SOPE genome, 1,450 have assigned roles in *E. coli*, while the remaining correspond to genes with hypothetical functions. In a similar study, the *Sodalis* genome of 2.0 Mb was found to hybridize to 1,812 *E. coli* orthologs (2), including 1,316 with described functional roles (Table 1). Both symbionts harbor large extrachromosomal elements, adding up to approximately 135 kb in *Sodalis* (2) and 138 kb in SOPE (11). The hybridization data obtained with *Sodalis* DNA reflect only the chromosomal homologs, but the homologs detected with SOPE include genes recognized by plasmid and chromosomal DNA. If the potential plasmid-specific homologs are excluded, *Sodalis* and SOPE genomes show homology to similar numbers of *E. coli* ORFs.

SOPE and *Sodalis* genome divergence. Comparative analysis of the ORFs detected showed that 1,525 *E. coli* orthologs were shared by both symbionts and 1,169 of these corresponded to ORFs with assigned functions (Table 1). To understand how

differences in symbiont-host biology may have impacted the loss of genes from SOPE and *Sodalis* genomes, a comparative analysis of the ORFs (with known roles in *E. coli*) detected in each functional category was conducted. Statistically significant differences were found among symbionts in the proportion of ORFs coding for carbon compound catabolism (Fisher's exact test; $P = 0.00009$), cell structure (Fisher's exact test; $P = 0.0073$), energy metabolism (Fisher's exact test; $P = 0.0071$), fatty acid metabolism (Fisher's exact test; $P = 0.0253$), and transport (Fisher's exact test; $P = 0.00004$). No significant differences in housekeeping categories such as translation and posttranslational modification, cell processes, transcription, nucleotide biosynthesis and metabolism, membrane proteins, and DNA replication, recombination, modification, and repair were recorded between the two symbionts. Furthermore, ORFs encoding the regulatory sigma factors 24 (*rpoE*), 54 (*rpoN*), 32 (*rpoH*), and 70 (*rpoD*) were detected with the hybridization of both SOPE and *Sodalis* DNAs, indicating that these bacteria have retained the ability to respond to environmental stimuli.

While sharing many ORFs, SOPE and *Sodalis* genomes have retained unique (i.e., not detected with the other symbiont) genes (281 and 147 genes, respectively) homologous to those with known functions in *E. coli*. In *Sodalis*, the majority of these genes correspond to cellular processes (16%), central intermediate metabolism (15%), and amino acid biosynthesis and metabolism (13%) (Fig. 1). SOPE maintains a greater number of unique genes associated with functions such as transport (17%), carbon compound catabolism (14%), cell structure (14%), and energy metabolism (13%) (Fig. 2). These unique genes may play important roles in the specific associations of SOPE and *Sodalis* within their weevil and tsetse fly hosts.

Metabolic capabilities. Based on the presence of all the genes necessary for complete metabolic pathways, both symbionts appear capable of synthesizing many amino acids, including Ala, Cys, Asp, Gly, Ile, Lys, Leu, Val, Met, Asn, Phe, Pro, Gln, Arg, Ser, Thr, Trp, Tyr, and Val. SOPE, however, appears deficient in the production of the essential amino acid histidine (His) (no *hisBCFHI*), although the His transporter (*hisP*) is detected, suggesting that His may be obtained from the weevil host with its strict cereal diet. In contrast, His synthesis is possible for *Sodalis*, as evidenced by the retention of *hisABCDGHI*, and accordingly *hisP* shows a very weak signal in comparison to that for SOPE, suggesting relaxed selection towards this gene and possibly inactivation. Both SOPE and *Sodalis* appear unable to synthesize glutamic acid, but Glu transporters (*gltK*, *-L*, and *-P*) are retained within their genomes. Although Glu is a nonessential amino acid, it is necessary for Arg, Pro, and Gln production, and symbionts may utilize transporters to attain this needed precursor externally. Glutamic acid has been shown to be a major product of tsetse fly metabolism (32, 33) and accounts for about 40% of the total amino acids of wheat (<http://www.nal.usda.gov>). Complete nucleotide biosynthesis and metabolism appear possible for SOPE and *Sodalis*, and both are capable of fatty acid biosynthesis commencing from acetyl-coenzyme A.

Both symbionts appear to lack the ability to synthesize thiamine (B1), folate, and nicotinamide, but they are capable of riboflavin, pyridoxine, lipolic acid, and protoheme production.

TABLE 1. Functional categories of ORFs detected for each microbe^a

Organism	No. of ORFs in functional category														Total no. (%) of ORFs			
	Amino acids	Central intermediate metabolism	Biosynthesis of cofactors	Energy metabolism	Carbon catabolism	Fatty acid metabolism	Regulatory functions	Cell processes	Transport	Replication	Transcription	Cell structure	Nucleic acids	Membrane	Functional	Hypothetical	Detected	
<i>Sodalis</i>	101	126	76	166	59	28	33	116	188	75	157	37	97	51	6	1,316 (73)	496 (27)	1,812
SOPE	95	123	78	193 ^b	91 ^b	39 ^b	34	104	231 ^b	78	160	40	123 ^b	50	11	1,450 (70)	634 (30)	2,084
<i>E. coli</i>	131	188	103	243	130	48	45	188	281	115	182	55	182	58	13	1,962 (46)	2,328 (54)	4,290
<i>Sodalis</i> and SOPE	82	104	65	157	52	28	30	92	183	63	147	31	82	47	6	1,169	356	1,525

^a *Sodalis* data were obtained from the work of Akman et al. (2); *E. coli* data were from the work of Blattner et al. (8).

^b Significant differences ($P < 0.05$) were found between the two symbiont genomes.

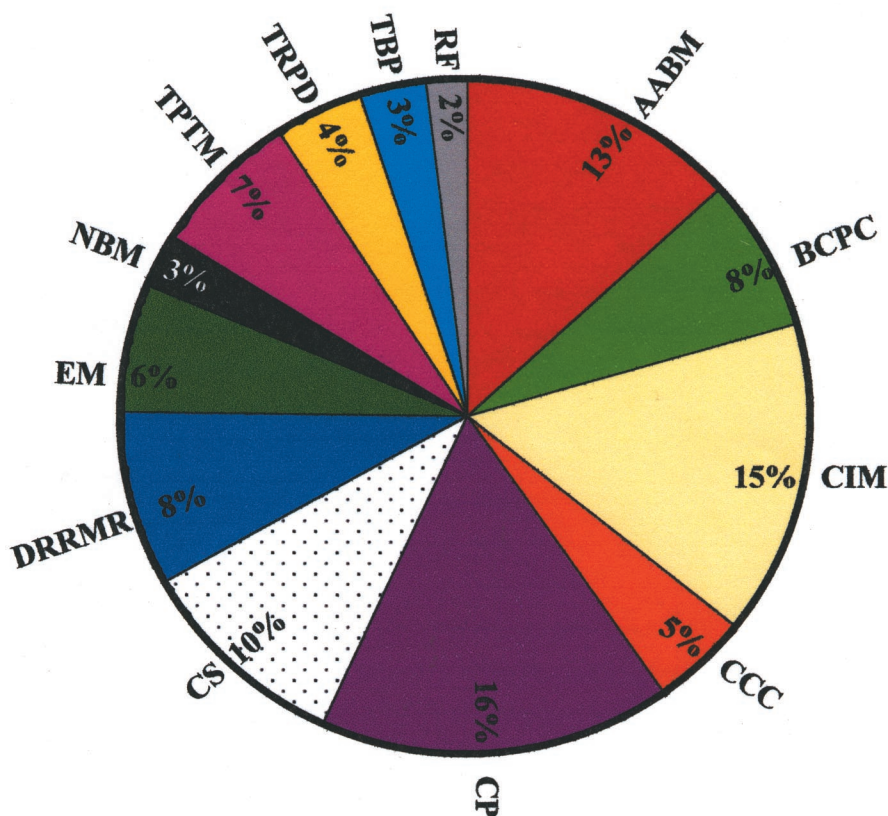


FIG. 1. Percentages of *Sodalis* unique genes ($n = 147$) with described functional roles in *E. coli*. AABM, amino acid biosynthesis and metabolism; BCPC, biosynthesis of cofactors, prosthetic groups, and carriers; CCC, carbon compound catabolism; CIM, central intermediary metabolism; CP, cell processes; CS, cell structure; DRRMR, DNA replication, recombination, modification, and repair; EM, energy metabolism; FAPM, fatty acid and phospholipid metabolism; MP, membrane proteins; NBM, nucleotide biosynthesis and metabolism; TPTM, translation and posttranslational modification; TRPD, transcription, RNA processing, and degradation; TBP, transport and binding proteins; RF, regulatory functions.

They also have in common the retention of ORFs that encode sugar isomerases (e.g., *pgm*, *galE*, and *araA*) and enzymes that metabolize simple sugars, like glucose (e.g., *glgC*, *galU*, *ptsG*, *malX*, *crr*, *pgi*, *prkB*, and *mrsA*). The most abundant sugar in insect hemolymph, alpha-trehalose, can also be metabolized by both symbionts. *Sodalis* has lost many genes encoding enzymes involved in intermediate pathways and that metabolize plant sugars. In contrast, SOPE has retained the majority of enzymes that catabolize plant sugars (β -D-glucosides). Hence, SOPE, but not *Sodalis*, is able to hydrolyze molecules such as cellobiose and salicin. *Sodalis* and SOPE possess the activator for glycolate (*glcC*), a product of plant photorespiration, but *Sodalis* lacks the ability to convert glycolate plus P to phosphoglycolate. Furthermore, SOPE retains most of the glycolate oxidases (*glcCDB*), while *Sodalis* maintains *glcC* only.

Comparative analysis of microbial genomes exhibiting different lifestyles. To understand the impact of different symbiotic relations on genome functions, we grouped orthologs detected with known functions into categories and compared them to those present in the annotated genome sequences of free-living *E. coli* (8), the obligate mutualists *Wigglesworthia* (3) and *Buchnera* (44), and the parasite *Rickettsia* (7) (Table 2). Since full genome annotations are not available for *Sodalis* and SOPE, the following percentages represent the minimum val-

ues for their respective genomes. A smaller percentage of *E. coli* orthologs devoted to cell processes are detected in the *Sodalis* and SOPE genomes (5.9 and 6.8%, respectively) than in the obligates *Wigglesworthia* (12.1%) and *Buchnera* (8.4%). In contrast, higher percentages of the SOPE and *Sodalis* genomes are involved in amino acid biosynthesis and metabolism, energy metabolism, and cofactor biosynthesis than of the genome of *Rickettsia*, which is parasitic and relies on its host for many of these functions. In comparison to *E. coli*, SOPE and *Sodalis* appear to have smaller percentages of orthologs devoted to central intermediate metabolism (9.7% compared to 8 and 5.9%, respectively), cell structure (9.4% compared to 8 and 4.8%, respectively), and cell processes (9.7% compared to 6.8 and 5.9%, respectively).

DISCUSSION

Overall genomic contents can be reconstructed quickly and inexpensively when complete sequences are unavailable by utilizing hybridization to arrays available for closely related organisms. Although this technique has been successfully applied to gain a broad understanding of genome composition, it lacks the ability to identify events that render genes inactive, such as point mutations and frame shifts caused by deletions and in-

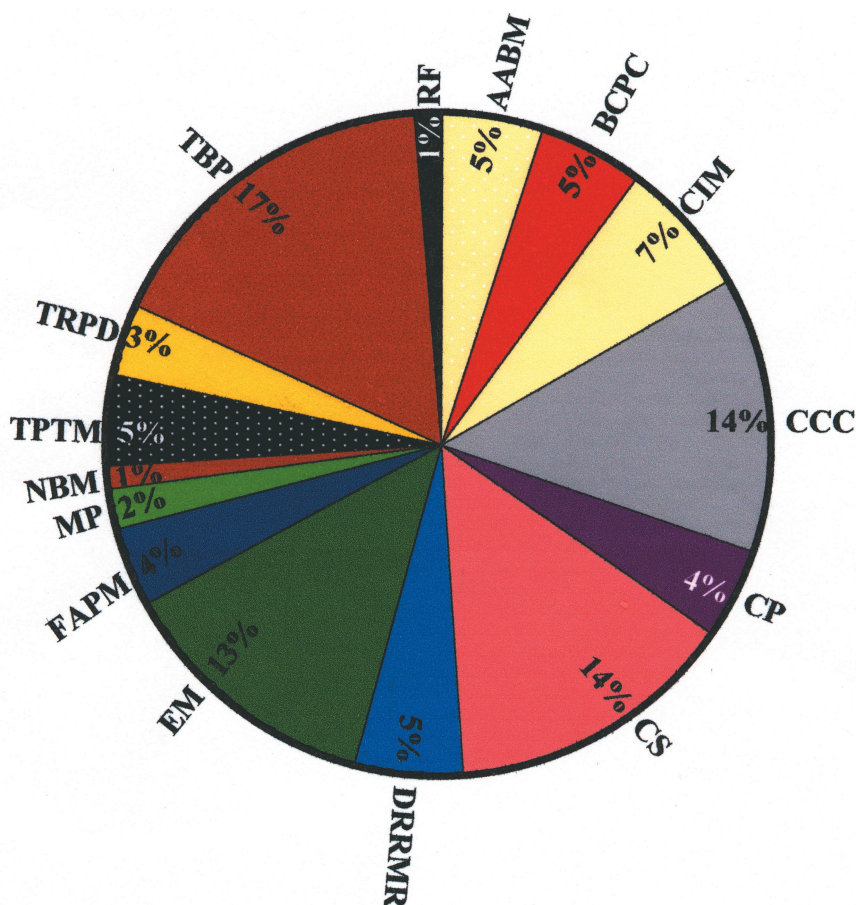


FIG. 2. Percentages of SOPE unique genes ($n = 281$) with described functional roles in *E. coli*. AABM, amino acid biosynthesis and metabolism; BCPC, biosynthesis of cofactors, prosthetic groups, and carriers; CCC, carbon compound catabolism; CIM, central intermediary metabolism; CP, cell processes; CS, cell structure; DRRMR, DNA replication, recombination, modification, and repair; EM, energy metabolism; FAPM, fatty acid and phospholipid metabolism; MP, membrane proteins; NBM, nucleotide biosynthesis and metabolism; TPTM, translation and posttranslational modification; TRPD, transcription, RNA processing, and degradation; TBP, transport and binding proteins; RF, regulatory functions.

sertions. Genes that are not found in the arrayed genome as well as species-specific genes also cannot be identified. Thus, a fraction of the genome of interest may remain unsampled when utilizing arrays of close relatives. Furthermore, the hybridization of genomes with nucleotide composition bias may

lead to the detection of artifacts and result in variance between gene array hybridization and genome-wide sequence analysis (1, 3). Like any comparative genome approach, this method is unable to detect nonorthologous gene displacement (29), by which unrelated or distantly related proteins can be recruited

TABLE 2. Percentages of the genome devoted to each functional class^a

Organism	% of genome devoted to class												
	Amino acid biosynthesis and metabolism	Central intermediate metabolism	Biosynthesis of cofactors	Energy metabolism	Fatty acid metabolism	Regulatory functions	Cell processes	Transport	Repliation	Translation	Transcription	Cell structure	Nucleic acids
<i>E. coli</i>	6.8	9.7	5.3	12.5	2.5	2.3	9.7	14.5	5.9	9.4	2.8	9.4	3
<i>Sodalis</i>	7.8	5.9	4.8	10.3	1.8	2.9	5.9	11.7	5	11.4	2.9	4.8	3.6
SOPE	6.1	8	5	12.5	2.5	2.2	6.8	15	5	10.4	2.6	8	3.2
<i>Wigglesworthia</i>	1.6	1.1	12.1	7.2	3.4	0.8	12.1	3.5	4	22.9	3.2	7.9	6.4
<i>Buchnera</i>	9.4	1.7	4.6	8.6	1.7	1.2	8.4	2.6	8.1	24.5	3.6	5.1	5.1
<i>Rickettsia</i>	0.1	0.5	0.5	3	0.5	0.6	13	3	3	9	1	10	2

^a Data are shown for the parasite *Rickettsia* (Andersson et al. [7]), the obligate mutualists *Buchnera* (Shigenobu et al. [44]) and *Wigglesworthia* (Akman et al. [3]), *Sodalis* (Akman et al. [2]), free-living *E. coli* (Blattner et al. [8]), and SOPE. Since full genome annotations are not available for *Sodalis* and SOPE, the percentages shown are the minimum values for their genomes.

for the same functions. Despite these limitations, this technique allows us to adequately compare genomes of microbes with similar guanine and cytosine (G+C) contents and with close phylogenetic distances to the organism from which the arrays were constructed. The identification of the full repertoire of genes comprising operons can further validate the results. For this study, related symbionts from two insect orders, Coleoptera and Diptera, were compared to delineate what alterations in microbial genome contents may have occurred since divergence from a common ancestor due to distinct host environments. To our knowledge, this is the first report describing microbial genome disintegration during the early phases of symbiotic establishment and in relation to environmental constraints.

Multiple phylogenetic analyses of *Enterobacteriaceae* species have suggested a distinct lineage for *Sodalis* and SOPE, indicating that they are members of a single bacterial taxon that have diverged from a common ancestral organism (4, 24). Recent data on the Dryophthoridae endosymbiont phylogeny have estimated the divergence between *Sitophilus* endosymbionts and *Sodalis* to be <25 million years (C. Lefevre, personal communication). The fact that symbionts related to *Sodalis* and SOPE are found in other insect taxa suggests that a progenitor had the ability to enter into relations with a wide range of hosts, perhaps as an insect pathogen (15). With time, pathogenic effects may have been attenuated while functions and/or products important for the evolutionary success of both partners were retained. It will be interesting to identify those genes and their functions that are present in symbionts but absent from *E. coli*. Equally intriguing is whether these genes were present in the common ancestor of enteric bacteria or were acquired horizontally after divergence.

Sodalis, known as the secondary symbiont of the tsetse fly, is transmitted vertically to intrauterine larvae through the mother's milk (14, 31) and has both intra- and extracellular localization in the midgut, muscle, salivary glands, fat body, and hemolymph (14). Although *Sodalis* maintains an overall similar tissue tropism between species, its density appears to vary among tsetse fly species (14). In contrast to *Sodalis*, SOPE has a strict intracellular localization within specialized structures called bacteriomes (25). Bacteriomes differentiate early during insect embryonic development and remain attached to the intestine at the junction of the foregut and midgut during the four larval stages. In young adults, bacteriomes are found in the mesenteric ceca of the intestine. However, in 2- to 3-week-old individuals, bacteriomes disappear, remaining only in the female ovaries from where bacteria are maternally transmitted to the offspring (25). SOPE has been shown to supply weevil larvae with vitamins such as pantothenic acid, riboflavin, and biotin (25) and is a source of amino acids such as phenylalanine and proline (19). The symbiont also interacts with mitochondrial oxidative phosphorylation by increasing mitochondrial enzymatic activity (23), thus extending the flight ability and other energy-dependent activities of its host (21, 25, 39). Yet the symbiosis can be disrupted without causing whole host population lethality (39).

Similarities in genomic makeup between SOPE and *Sodalis* are still highly evident, with retention of many of the same genetic components involved in housekeeping functions such as translation and posttranslational modification, cell pro-

cesses, transcription, and nucleotide biosynthesis. The cellular machinery involved in these processes appears to be conserved in mutualists and commensals, in contrast to parasitic microbes, which exploit their hosts for many of these resources. Furthermore, the greater conservation of translational, as opposed to transcriptional, processes supports stronger genetic regulation at the translational level for *Sodalis* and SOPE.

Competence in complete nucleotide biosynthesis and metabolism, amino acid biosynthesis and metabolism, energy metabolism, and cofactor biosynthesis suggests that SOPE and *Sodalis* may be approaching mutualism in their symbiotic associations. The retention of genes involved in regulatory functions, such as sigma factors, supports their recent symbiotic establishment. Obligate mutualists that live intracellularly, sheltered from environmental fluctuations, within their hosts have lost such genes with associated regulatory functions due to lack of need (3, 35, 44).

Despite the vast similarities in detected ORFs for SOPE and *Sodalis*, specialized modifications towards host environment, particularly in metabolic functions, appear to have occurred that are reminiscent of the genome tailoring of ancient symbionts. The significantly greater number of energy metabolism and fatty acid metabolism genes detected for SOPE than for *Sodalis* may be due to the restricted cereal diet of SOPE's weevil host. Lipids, prominent in the tsetse fly blood meal, provide more energy than plant carbohydrates because they are in a more reduced form (53). The erosion of fatty acid metabolism pathways in *Sodalis* might reflect the natural abundance of such products in the host environment. SOPE, with a greater capacity for carbon catabolism, is capable of metabolizing plant sugars in the diet of its host, which is comprised of as much as 70% starch but is very low in lipid components (<http://www.nal.usda.gov>). The purging of genes involved in plant sugar metabolism from the *Sodalis* genome can be interpreted as an adaptive response to tsetse fly nutritional behavior. Since tsetse flies do not feed on plant material but on blood, which is low in carbohydrates and rich in simple sugars such as glucose and trehalose, *Sodalis* has lost unnecessary pathways that catabolize plant sugars such as starch.

The higher number of unique ORFs corresponding to cellular processes (perhaps adaptations necessary for intra- and extracellular localization), central intermediate metabolism, and amino acid biosynthesis and metabolism for *Sodalis* and to carbon compound catabolism, cell structure, energy metabolism, and fatty acid metabolism and transport for SOPE may be indicative of differences in genome retention since their last common ancestor and suggests bacterial domestication by the host. These differences in retention will influence what is further maintained in the SOPE and *Sodalis* genomes as relations with their hosts further evolve.

Utilizing the intensively studied, 250-million-year-old *Buchnera*-aphid association (38) as a model, some authors have suggested that significant changes on microbial genome structure transpire early upon symbiotic establishments (36). Large deletions, which typically span multiple genes, occur during the initiation of symbiosis, resulting not from selection for DNA loss but from decreased selection to maintain locus functionality (36, 49). Such massive genome reduction early upon symbiosis is supported by near perfect gene order conservation in the whole-genome sequences of three divergent strains of

Buchnera (44, 47, 49). It has been inferred that the content of these early deletions determines the degree of selection on remaining loci and ultimately governs the eventual genetic inventory of the reduced genome. Only later, at an exponentially decreasing pace, are some genes eliminated through inactivation and gradual erosion (5, 45). These losses, resulting in the reduction of microbial functional flexibility, are expected to restrict the evolutionary options for the microbes, ultimately harnessing them into specific symbiotic lifestyles. Such drastic genome erosion may enable the recruitment of newer symbiotic associations to replace functions lost in the ancient obligate mutualists and potentially allow hosts to exploit new niches.

Despite their close taxonomic relatedness, the genomes of SOPE and *Sodalis* have been shaped differentially due to adaptations to their unique host environments. As a result, these organisms have diverged extensively and appear to be tailored to subsist on different metabolites provided in their host diets. These findings are of relevance for our applied genetic engineering studies by which we explore the use of symbionts to block transmission of pathogens in their insect hosts. Our results infer that the symbionts described here are anchored tightly to host biology through restricted metabolic capabilities and therefore may not be able to undergo horizontal transmission and establishment in distant insect taxa.

ACKNOWLEDGMENTS

We thank Trevor W. Rudy for computer programming assistance, John Brownstein for support with statistical analyses, and members of the Aksoy lab and R.V.M.R.'s doctoral committee for enlightening discussions.

This work was funded by NIH/NIAID grant AI-34033 to S.A. R.V.M.R. is the recipient of NIH training grant T32AI07404 and a CDC fellowship for training in vector-borne infectious diseases, T01/CCT122306-01.

REFERENCES

- 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. 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. Hyspa. 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. 2001. Pseudogenes, junk DNA, and the dynamics of *Rickettsia* genomes. Mol. Biol. Evol. **18**:829–839.
- Andersson, S. G. E., and C. G. Kurland. 1998. Reductive evolution of resident genomes. Trends Microbiol. **6**:263–268.
- Andersson, S. G. E., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. M. Alsmark, R. M. Podowski, A. K. Naslund, A. Eriksson, H. H. Winkler, and C. G. Kurland. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of the mitochondria. Nature **396**:133–140.
- Blattner, F. R., G. Plunkett, C. A. Bloch III, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277**:1453–1474.
- Braig, H. R., W. Zhou, S. L. Dobson, and S. L. O'Neill. 1998. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. J. Bacteriol. **180**:2373–2378.
- Buchner, P. 1965. Endosymbiosis of animals with plant microorganisms. Interscience, New York, N.Y.
- Charles, H., G. Condemine, C. Nardon, and P. Nardon. 1997. Genome size characterization of the principal endocellular symbiotic bacteria of the weevil *Sitophilus oryzae*, using pulsed field gel electrophoresis. Insect Biochem. Mol. Biol. **27**:345–350.
- Chen, D. Q., and A. H. Purcell. 1997. Occurrence and transmission of facultative endosymbionts in aphids. Curr. Microbiol. **34**:220–225.
- Chen, D. Q., C. B. Montllor, and A. H. Purcell. 2000. Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrtosiphon pisum*, and the blue alfalfa aphid, *A. kondoi*. Entomol. Exp. Applic. **95**:315–323.
- Cheng, Q., and S. Aksoy. 1999. Tissue tropism, transmission, and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies. Insect Mol. Biol. **8**:125–132.
- 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.
- 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.
- Fukatsu, T., and N. Nikoh. 1998. Two intracellular symbiotic bacteria from the mulberry psyllid *Anomoneura mori* (Insecta, Homoptera). Appl. Environ. Microbiol. **64**:3599–3606.
- Fukatsu, T., N. Nikoh, R. Kawai, and R. Koga. 2000. The secondary endosymbiotic bacterium of the pea aphid *Acyrtosiphon pisum* (Insecta, Homoptera). Appl. Environ. Microbiol. **66**:2748–2758.
- Gasnier-Fauchet, F., A. Gharib, and P. Nardon. 1986. Comparison of methionine metabolism in symbiotic and aposymbiotic larvae of *Sitophilus oryzae* L. (Coleoptera: Curculionidae). I. Evidence for a glycine N-methyltransferase like activity in the aposymbiotic larvae. Comp. Biochem. Physiol. **85B**:245–250.
- Goebel, W., and R. Gross. 2001. Intracellular survival strategies of mutualistic and parasitic prokaryotes. Trends Microbiol. **9**:267–273.
- Grenier, A. M., C. Nardon, and P. Nardon. 1994. The role of symbionts in flight activity of *Sitophilus* weevils. Entomol. Exp. Applic. **70**:201–208.
- Heddi, A., F. Lefebvre, and P. Nardon. 1991. The influence of symbiosis on the respiratory control ratio (RCR) and the ADP/O ratio in the adult weevil *Sitophilus oryzae* (Coleoptera: Curculionidae). Endocyt. Cell Res. **8**:61–73.
- Heddi, A., F. Lefebvre, and P. Nardon. 1993. Effect of endocytobiotic bacteria on mitochondrial enzymatic activities in the weevil *Sitophilus oryzae* (Coleoptera: Curculionidae). Insect Biochem. Mol. Biol. **23**:403–411.
- Heddi, A., C. Hubert, C. Khatchadourian, G. Bonnot, 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.
- Heddi, A., A. M. Grenier, C. Khatchadourian, H. Charles, and P. Nardon. 1999. Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. Proc. Natl. Acad. Sci. USA **96**:6814–6819.
- Heddi, A. 2003. Endosymbiosis in the weevil of the genus *Sitophilus*: genetic, physiological, and molecular interactions among associated genomes, p. 67–82. In K. Bourtzis and T. Miller (ed.), Insect symbiosis. CRC Press, Boca Raton, Fla.
- Hentschel, U., and M. Steinert. 2001. Symbiosis and pathogenesis: common themes, different outcomes. Trends Microbiol. **9**:585.
- Hooper, L. V., and J. I. Gordon. 2001. Commensal host-bacterial relationships in the gut. Science **292**:1115–1118.
- Koonin, E. V. 2000. How many genes can make a cell: the minimal-gene set concept. Annu. Rev. Genomics Hum. Genet. **1**:99–116.
- Littell, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger. 1996. SAS system for mixed models. SAS Institute, Inc., Cary, N.C.
- Ma, W. C., and D. L. Denlinger. 1974. Secretory discharge and microflora of milk gland in tsetse flies. Nature **247**:301–303.
- Moloo, S. K. 1976. Nutrition of *Glossina morsitans*: metabolism of U-¹⁴C glucose during pregnancy. J. Insect Physiol. **22**:195–200.
- Moloo, S. K. 1976. Aspects of the nutrition of adult female *Glossina morsitans* during pregnancy. J. Insect Physiol. **22**:563–567.
- Montllor, C. B., A. Maxmen, and A. H. Purcell. 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum*, under heat stress. Ecol. Entomol. **27**:189–195.
- Moran, N. A. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell **108**:583–586.
- Moran, N. A., and A. Mira. 14 November 2001, posting date. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome Biology **2**:0054.1–0054.12. [Online.] <http://genomebiology.com/2001/2/12/research/0054.1>.
- Moran, N. A., C. Dale, H. Dunbar, W. A. Smith, and H. Ochman. 2003. Intracellular symbionts of sharpshooters (Insecta: Hemiptera: Cicadellinae) form a distinct clade with a small genome. Environ. Microbiol. **5**:116–126.
- Munson, M. A., P. Baumann, M. A. Clark, L. Baumann, and N. A. Moran. 1991. Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. J. Bacteriol. **173**:6321–6324.
- Nardon, P. 1973. Obtention d'une souche aposymbiotique chez le charançon *Sitophilus sasakii* Tak: différentes méthodes et comparaison avec la souche symbiotique d'origine. C. R. Acad. Sci. **277D**:981–984.
- Nogge, G. 1981. Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous arthropods. Parasitology **82**:101–104.

41. Oliver, K. M., J. A. Russell, N. A. Moran, and M. S. Hunter. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc. Natl. Acad. Sci. USA* **100**:1803–1807.
42. O'Neill, S. L., R. H. Gooding, and S. Aksoy. 1993. Phylogenetically distant symbiotic microorganisms reside in *Glossina* midgut and ovary tissues. *Med. Vet. Entomol.* **7**:377–383.
43. Sandstrom, J. P., J. A. Russell, J. P. White, and N. A. Moran. 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Mol. Ecol.* **10**:217–228.
44. 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.
45. Silva, F. J., A. Latorre, and A. Moya. 2001. Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends Genet.* **17**:615–618.
46. Steinert, M., U. Hentschel, and J. Hacker. 2000. Symbiosis and pathogenesis: evolution of the microbe-host interaction. *Naturwissenschaften* **87**:1–11.
47. Tamas, I., L. Klasson, B. Canback, A. K. Naslund, A. 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.
48. Tsuchida, T., R. Koga, H. Shibao, T. Matsumoto, and T. Fukatsu. 2002. Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrtosiphon pisum*. *Mol. Ecol.* **11**:2123–2135.
49. van Ham, R. C. H. J., J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J. M. Fernandez, L. Jimenez, M. Postigo, F. J. Silva, J. Tamames, E. Viguera, A. Latorre, A. Valencia, F. Moran, and A. Moya. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci. USA* **100**:581–586.
50. Welburn, S. C., I. Maudlin, and D. S. Ellis. 1987. *In vitro* cultivation of rickettsia-like organisms from *Glossina* spp. *Ann. Trop. Med. Parasitol.* **81**:331–335.
51. Welburn, S. C., K. Arnold, I. Maudlin, and G. W. Goday. 1993. *Rickettsia*-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. *Parasitology* **107**:141–145.
52. Wernegreen, J. J., P. H. Degnan, A. B. Lazarus, C. Palacios, and S. R. Bordenstein. 2003. Genome evolution in an insect cell: distinct features of an ant-bacterial partnership. *Biol. Bull.* **204**:221–231.
53. Ziegler, R. 1996. Energy metabolism, p. 356–370. *In* B. J. Beaty and W. C. Marquardt (ed.), *The biology of disease vectors*. University Press of Colorado, Niwot, Colo.