Extrachromosomal DNA of the Symbiont *Sodalis glossinidius*†

A. C. Darby,^{1*} J. Lagnel,² C. Z. Matthew,¹ K. Bourtzis,^{2,3} I. Maudlin,¹
 and S. C. Welburn¹

Centre of Infectious Diseases, College of Medicine and Veterinary Medicine, University of Edinburgh, Easter Bush, Edinburgh EH25 9RG, United Kingdom ¹; Institute of Molecular Biology and Biotechnology, FORTH, Vassilika Vouton, Heraklion 71110, Crete, Greece²; and Department of Environmental and Natural Resources Management, University of Ioannina, Agrinio, Greece³

Received 28 February 2005/Accepted 15 April 2005

The extrachromosomal DNA of *Sodalis glossinidius* from two tsetse fly species was sequenced and contained four circular elements: three plasmids, pSG1 (82 kb), pSG2 (27 kb), and pSG4 (11 kb), and a bacteriophage-like pSG3 (19 kb) element. The information suggests *S. glossinidius* is evolving towards an obligate association with tsetse flies.

Sodalis glossinidius, a symbiont of tsetse flies (genus *Glossina*), is the only gammaproteobacterial insect symbiont to be cultured (19) and thus amenable to genetic modification (4), suggesting that it could be used as part of a control strategy by vectoring antitrypanosome genes (1). The relationship of *S. glossinidius* with its insect host is not fully understood, but the organism may increase the susceptibility of tsetse flies to trypanosomes (2). In the present work we have sequenced and annotated extrachromosomal DNA (ecDNA) from two *S. glossinidius* isolates from two phylogenetically distant and geographically isolated tsetse fly species, *Glossina palpalis palpalis* (Robineau-Desvoidy) and *G. austeni* (Newstead).

ecDNA was extracted from clonal isolates of *S. glossinidius* from *G. palpalis palpalis* (GP-SG¹) and *G. austeni* (GA-SG¹) (C. Z. Matthew A. C. Darby, S. A. Young, L. Hume, and S. C. Welburn, submitted for publication) using a large construct kit (QIAGEN, Crawley, United Kingdom). Shotgun libraries were made from circular ecDNA isolated from mid-log-phase liquid cultures of both *S. glossinidius* samples and prepared by randomly shearing and blunt-ending the ecDNA. DNA fragments between 1 and 3 kb were selected and cloned into the SmaI site of pUC19 vector (New England Biologicals, Hitchin, United Kingdom) and transformed into *Escherichia coli* XL1-Blue MRF cells (Stratagene, Amsterdam, The Netherlands). For each library 1,344 clones were sequenced in both directions for approximately sixfold coverage. The reads were screened using the Phred-Phrap software (<http://www.phrap.org>) and contigs were assembled in SeqMan (DNASTAR). Gaps were closed by long PCR using the Elongase amplification system (Invitrogen, Glasgow, United Kingdom) (primers and methodology available on request).

Long PCR fragments were A-tailed and cloned into the pGEMTeasy vector (Promega, Wisconsin), transformed into

E. coli XL1-Blue MRF cells (Stratagene, Amsterdam, The Netherlands), and sequenced by primer walking in both directions. Long PCR was also used to test the circular structure of the ecDNA. Genomic DNA was extracted from both bacteria

TABLE 1. Summary of ecDNA statistics

Plasmid	Parameter	<i>G. palpalis palpalis</i>	<i>G. austeni</i>
All	No. of clones sequenced	1,327	1,232
	Total sequence length (bp)	1,181,567	957,384
	Sum contig length (bp)	138,817	119,619
pSG1	Total length (bp)	81,553	81,553
	Percentage for project	36.5%	52.7%
	No. of ORFs	91	91
	Percent coding	79.8%	79.8%
	Percent coding minus hypothetical pseudogenes	48.3%	48.3%
	G+C content (%)	49.1%	49.1%
	Average G+C ORFs (%)	49.3%	49.3%
	Average coverage (%)	5.5%	6.2%
pSG2	Total length (bp)	27,240	27,240
	Percentage for project	26.8%	41.5%
	No. of ORFs	31	31
	Percent coding	78.1%	78.1%
	G+C content (%)	44.7%	44.7%
	Average G+C ORFs (%)	46.6%	46.6%
	Average coverage (%)	10.3%	13.2%
pSG3	Total length (bp)	19,201	N/A
	Percentage for project	26.6%	0.2%
	No. of ORFs	25	N/A
	Percent coding	78.2%	N/A
	G+C content (%)	50.7%	N/A
	Average G+C ORFs (%)	50.9%	N/A
	Average coverage (%)	15.1%	N/A
pSG4	Total length (bp)	10,816	10,816
	Percentage for project	9.6%	3.0%
	No. of ORFs	13	13
	Percent coding	66.0%	66.0%
	G+C content (%)	47.8%	47.8%
	Average G+C ORFs (%)	50.5%	50.5%
	Average coverage (%)	8.8%	2.9%

* Corresponding author. Mailing address: Centre of Infectious Diseases, College of Medicine and Veterinary Medicine, University of Edinburgh, Easter Bush, Edinburgh EH25 9RG, United Kingdom. Phone: 44 (0)131 6506261. Fax: 44 (0)131 6507348. E-mail: alistair.darby@ed.ac.uk.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

FIG. 1. Circular maps of the elements pSG1 (left-hand page) and pSG2, pSG3, and pSG4 (right-hand page). The outer ring depicts ORFs (above the line, positive strain; below the line, negative strain). Checked patterns represent pseudogenes, coded according to function: virulence/symbiotic ORFs (red), regulators (orange), conserved hypothetical (yellow), hypothetical protein (no match) (light green), plasmid (pSG3 phage) ORFs (dark green), transposase ORFs (light blue), conjugation ORFs (dark blue), and metabolic functions (purple). Second-layer error type: f = frameshift, s = stop, and t = truncated. The second ring (red) graphs G+C content, calculated with a 1,000-bp sliding window and plotted around the mean value for all ORFs. The third (blue) and fourth (red) rings depict PstI and HindIII digestion sites, respectively. The scale is in base pairs; plasmids not drawn to scale are represented by concentric rings (bottom right).

and tsetse flies using a cetyltrimethylammonium bromide method (13). ecDNA sequences have been assigned EMBL accession numbers AJ868433 to AJ868439. Acetone-preserved *G. austeni* ($n = 53$) from Pangani, Tanzania, were used as a template for *S. glossinidius*-specific PCR (C. Z. Matthew et al., submitted for publication). Plasmid gene-specific primers were used to detect each ecDNA element (available on request).

Isolates GP-SG¹ and GA-SG¹ both contained multiple circular elements that formed independent contigs (Table 1) with 99% homology between isolates, suggesting the two plasmid isolates have a recent common ancestor. pSG1, pSG2, and pSG4 were plasmid-like elements (Fig. 1), all with a putative RepA protein; pSG3 was bacteriophage-like, with 56% (14/25) of the open reading frames (ORFs) showing homology to bac-

FIG. 1—Continued.

FIG. 2. Siderophore island of *Sodalis glossinidius*. a) Genetic structure: percent G+C plot (y axis, moving 3-base window) on the pSG1 plasmid. ORF names and their orientations (above the line, positive strain; below the line, negative strain) coded according to function: siderophore production and transport (red), regulators (orange), hypothetical protein (no match) (yellow), and transposase (blue). b) Demonstration of siderophore production by *S. glossinidius* isolates of *Glossina austeni* (GA-SG¹), *G. palpalis palpalis* (GP-SG¹), and *G. morsitans morsitans* (GM-SG) on CAS iron indicator. Blue: iron present; white: iron removed from the medium.

terioophage sequences and no Rep protein homologue. The putative bacteriophage proteins showed greatest homology to a P22-like phage but no tail or capsid homologs were found, suggesting a phage remnant or prophage-like element. Full annotation is given in Tables S1 and S2 in the supplemental material.

All ecDNA elements were present in laboratory isolates, but copy number varied between isolates (data not shown). pSG1, pSG2, and pSG4 were universal in *S. glossinidius* from wild tsetse fly populations (9/9 *S. glossinidius*-positive flies), but pSG3 was present in only 7/9 *S. glossinidius*-positive flies; none of the elements were detected in *S. glossinidius*-negative flies. That these elements occur in natural tsetse fly populations shows they are stable over time and not simply laboratory artifacts.

The pSG1 plasmid contained two putative symbiotic islands; the first island contained 14 ORFs (bp 28276 to 45828) for production and transport of siderophores (Fig. 2) between putative transposase ORFs in a region that showed a marked GC divergence relative to the flanking sequences (Fig. 2a). The architecture of this gene cluster is that of a fitness island (10), suggesting that it is the product of horizontal gene transfer. The presence of siderophores for the acquisition of iron on ecDNA of *S. glossinidius* and the demonstration of siderophore production (Fig. 2b) suggest that this bacterium has been shaped by mobile genetic elements, in the same way as free-living and pathogenic bacteria.

The level of sequence homology suggests that these ORFs encode an achromobactin-like (citrate) siderophore and its ABC transporter (8). Siderophores that promote intracellular infection (16) are produced by other symbiotic bacteria, e.g., *Photorhabdus luminescens* (6), and should be considered candidates for symbiotic fitness genes in *S. glossinidius*. It is clear that a siderophore island was present in the common ancestor of both *S. glossinidius* isolates and therefore predates transmission to tsetse flies. The second island contains two ORFs (bp 49394 to 50367, situated between two putative transposases) with homology to a putative secreted protein in *P. luminescens* and a small heat shock protein similar to Ibp in *Buchnera aphidicola*. Together with a number of other ORFs that have metalloproteinase and secreted protein domains, this island's ORFs are candidates for effectors, toxins, hemolysins, and protease fitness factors.

Three blocks of ORFs showed homology to conjugative transfer pilus genes (Fig. 1) described in gammaproteobacteria (9, 14, 18), suggesting that conjugation has been an important factor in the evolutionary history of these symbionts and may have been acquired (and lost) on more than one occasion. The first block, on pSG1 Tra1 (located between bp 2298 and 14078) encoded 14 ORFs that showed significant homology to genes of the transfer regions of plasmids R100 (NC_002134) and F (NC_002483) but was approximately one third the size of the corresponding R100 (34 ORFs) and F plasmid (35 ORFs) regions. The second Tra2 region contained 11 ORFs (located

between bp 67518 and 77716) and showed close homology to *Yersinia enterocolitica* p29930 (18) and *E. coli* (IncX) (14). The *S. glossinidius* homologs in both Tra regions were disrupted by mutations but showed preserved gene order and good identity with homologous transfer regions.

The third region had two ORFs on pSG2 (between bp 21560 and 22386) with homology to plasmid R64 pilus ORFs. The mobilization of plasmids pSG2 and pSG4 by pSG1 may explain how this complex ecDNA has been maintained in different *S. glossinidius* isolates. These transfer regions are probably not capable of independent conjugation, as all lack functional VirB04 (TraC and TriC) inner membrane ATP-binding protein, pilus assembly, and mating pair stabilization protein homologs.

The large number of pseudogenes found on plasmid pSG1 is unusual, but pseudogenes have previously been reported in *Borrelia burgdorferi*, *Buchnera aphidicola*, *Streptomyces coelicolor*, *Vibrio anguillarum*, and *Y. pestis* plasmids (3, 5, 11, 15, 17, 20). When pseudogenes were included in protein-coding calculations, the pSG1 plasmid looked normal (79.8%), however, when excluded, it dropped to a level (48.3%) comparable to that of the linear plasmid of *Borrelia burgdorferi* and the genome of *Mycobacterium leprae* (12). Gene silencing due to mutation looks catastrophic in the pSG1 plasmid, but this element is both viable and stable, as demonstrated by its detection in wild fly populations and homology between isolates. The large amount of junk DNA observed suggests that insertion and deletion events have been rare compared to the frequency of point mutations (7).

The ecDNA of *Sodalis glossinidius* provides further evidence that symbiotic and pathogenic bacteria have acquired homologous genes for host interaction. The presence of fitness islands together with repetitive DNA, transposable elements, bacteriophage, and conjugation sequences shows that *S. glossinidius* has acquired genes from other bacteria by horizontal gene transfer, suggesting the evolution of nonobligate symbionts may be a balance between gene loss and gene acquisition.

This study indicates that *S. glossinidius* is evolving towards an obligate intracellular symbiotic lifestyle (7) and, while gene loss has occurred, a high level of complexity and diversity has been retained which is absent from obligate intracellular symbionts. This suggests that *S. glossinidius* has not completely abandoned its free-living ancestry for obligatory symbiosis.

We thank D. Shaw for his help and S. Torr (Natural Resources Institute, University of Greenwich) for supplying wild tsetse flies.

This research was funded by the Wellcome Trust (A.C.D. and I.M.).

REFERENCES

1. Aksoy, S. 2003. Control of tsetse flies and trypanosomes using molecular genetics. *Vet. Parasitol.* **115**:125–145.
2. Baker, R. D., I. Maudlin, P. J. M. Milligan, D. H. Molyneux, and S. C. Welburn. 1990. The possible role of Rickettsia-like organisms in Trypanosomiasis epidemiology. *Parasitology* **100**:209–217.
3. Baumann, L., M. A. Clark, D. Rouhbakhsh, P. Baumann, N. A. Moran, and D. J. Voegtlin. 1997. Endosymbionts (*Buchnera*) of the aphid *Uroleucon sonchi* contain plasmids with trpEG and remnants of trpE pseudogenes. *Curr. Microbiol.* **35**:18–21.
4. Beard, C. B., S. L. O'Neill, P. Mason, L. Mandelco, C. R. Woese, R. B. Tesh, F. F. Richards, and S. Aksoy. 1993. Genetic transformation and phylogeny of bacterial symbionts from tsetse. *Insect Mol. Biol.* **1**:123–131.
5. Casjens, S., N. Palmer, R. van Vugt, W. M. Huang, B. Stevenson, P. Rosa, R. Lathigra, G. Sutton, J. Peterson, R. J. Dodson, D. Haft, E. Hickey, M. Gwinn, O. White, and C. M. Fraser. 2000. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol. Microbiol.* **35**:490–516.
6. Ciche, T. A., M. Blackburn, J. R. Carney, and J. C. Ensign. 2003. Photobactin: a catechol siderophore produced by *Photorhabdus luminescens*, an entomopathogen mutually associated with *Heterorhabditis bacteriophora* NCI nematodes. *Appl. Environ. Microbiol.* **69**:4706–4713.
7. Frank, C. A., H. Amiri, and S. G. E. Andersson. 2002. Genome deterioration: loss of repeated sequences and accumulation of junk DNA. *Genetica* **115**:1–12.
8. Franza, T., I. Michaud-Soret, P. Piquerel, and D. Expert. 2002. Coupling of iron assimilation and pectinolysis in *Erwinia chrysanthemi* 3937. *Mol. Plant-Microbe Interact.* **15**:1181–1191.
9. Frost, L. S., K. Ippen-Ihler, and R. A. Skurray. 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **58**:162–210.
10. Hacker, J., and E. Carniel. 2001. Ecological fitness, genomic islands and bacterial pathogenicity: a Darwinian view of the evolution of microbes. *EMBO Rep.* **2**:376–381.
11. Holmstrom, K., and L. Gram. 2003. Elucidation of the *Vibrio anguillarum* genetic response to the potential fish probiotic *Pseudomonas fluorescens* AH2, using RNA-arbitrarily primed PCR. *J. Bacteriol.* **185**:831–842.
12. Lawrence, J. G., R. W. Hendrix, and S. Casjens. 2001. Where are the pseudogenes in bacterial genomes? *Trends Microbiol.* **9**:535–540.
13. Navajas, M., J. Lagnel, J. Gutierrez, and P. Boursot. 1998. Species wide homogeneity of nuclear ribosomal ITS2 sequences in the spider mite *Tetranychus urticae* contrasts with extensive mitochondrial COI polymorphism. *Heredity* **80**:742–752.
14. Nunez, B., P. Avila, and F. de la Cruz. 1997. Genes involved in conjugative DNA processing of plasmid R6K. *Mol. Microbiol.* **24**:1157–1168.
15. Perry, R. D., S. C. Straley, J. D. Fetherston, D. J. Rose, J. Gregor, and F. R. Blattner. 1998. DNA sequencing and analysis of the low Ca²⁺ response plasmid pCD1 of *Yersinia pestis* KIM5. *Infect. Immun.* **66**:4611–4623.
16. Ratledge, C., and L. G. Dover. 2000. Iron metabolism in pathogenic bacteria. *Annu. Rev. Microbiol.* **54**:881–941.
17. Skurnik, M., and H. Wolfwatz. 1989. Analysis of the YopA gene encoding the Yop1 virulence determinants of *Yersinia* spp. *Mol. Microbiol.* **3**:517–529.
18. Strauch, E., G. Goetz, D. Knabner, A. Konietzny, E. Lanka, and B. Appel. 2003. A cryptic plasmid of *Yersinia enterocolitica* encodes a conjugative transfer system related to the regions of CloDF13 Mob and IncX Pil. *Microbiology UK* **149**:2829–2845.
19. 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.
20. Yang, C. C., C. H. Huang, C. Y. Li, Y. G. Tsay, S. C. Lee, and C. W. Chen. 2002. The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins. *Mol. Microbiol.* **43**:297–305.