Extracting single genomes from heterogenous DNA samples: A test case with Carsonella ruddii, the bacterial symbiont of psyllids (Insecta)

Authors: Dale, Colin, Dunbar, Helen, Moran, Nancy A., and Ochman, Howard

Source: Journal of Insect Science, 5(3) : 1-6

Published By: Entomological Society of America

URL: https://doi.org/10.1673/031.005.0301

Dale C, Dunbar H, Moran NA, Ochman H. 2005. Extracting single genomes from heterogenous DNA samples: A test case with *Carsonella ruddii*, the bacterial symbiont of psyllids (Insecta). 6pp. *Journal of Insect Science*, 5:3, Available online: insectscience.org/5.3

Journal of Insect Science

insectscience.org

Extracting single genomes from heterogenous DNA samples: A test case with *Carsonella ruddii*, the bacterial symbiont of psyllids (Insecta)

Colin Dale[1], Helen Dunbar[2], Nancy A. Moran[2*] and Howard Ochman[1]

[1]Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, Arizona 85721, USA
[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA
nmoran@email.arizona.edu

Received 4 June 2004, Accepted 18 September 2004, Published 16 March 2005

Abstract

Analysis of many bacterial genomes is impeded by the inability to separate individual species from complex mixtures of cells or to propagate cells in pure culture. This problem is an obstacle to the study of many bacterial symbionts that live intracellularly in insects and other animals. To recover bacterial DNA from complex samples, we devised a method that facilitates the cloning of DNA fragments of distinctive G+C contents in order to generate shotgun DNA libraries enriched in inserts having a specific base composition. DNA preparations are first treated with a restriction enzyme having a common cleavage site in a particular genome and then shotgun cloned following size-fractionation. This method was applied to whole bacteriomes of the psyllid, *Pachypsylla venusta*, which harbors the bacterial symbiont Candidatus *Carsonella ruddii*. The resulting libraries were highly enriched in bacterial sequences. Through the use of alternate enzymes and partial digests, this technique can be adapted to yield virtually pure DNA libraries for individual bacterial species.

Keywords: bacterial genome, base composition, endosymbiont, Homoptera, *Pachypsylla venusta*, shotgun cloning

Introduction

Due to their small size and low genetic complexity, bacterial genomes are highly amenable to complete sequence determination. But because relatively large (i.e., microgram) quantities of purified DNA are needed for most shotgun cloning procedures, bacterial genome sequencing projects have focused primarily on organisms that have been propagated in pure culture. The problem of obtaining purified DNA presents an obstacle for the vast numbers of bacterial species that have not yet been cultured *in vitro*, such as those that live in structured communities with other microorganisms or as obligate symbionts residing exclusively in the tissues of animal and plant hosts. There is a recognized need to develop culture-independent approaches that will permit access to the genomes of bacteria that are environmentally or medically important (DeLong and Pace 2001; Cummings and Relman 2002).

One method of obtaining information about individual genomes from complex assemblages of organisms is by cloning heterogeneous DNA samples into large insert BAC libraries (Beja et al. 2000). This method was first used to recover DNA from a mixed microbial population collected at a depth of 200 meters in the Pacific Ocean (Stein et al. 1996). In that study, BAC clones containing microbial DNA were identified by the presence of archetypal ribosomal DNA sequences, and an entire BAC insert, in this case derived from an uncultivated planktonic archaeon, was sequenced.

Bacteria that live symbiotically within the cells or tissues of another organism present additional problems for genome analysis. Because the amount of symbiont DNA present in whole organism preparations is usually rather low in comparison with the amount of host DNA, bacteria must be physically separated from host tissues and other cellular components to facilitate DNA isolation for genome studies. Such enrichment has been achieved previously by buoyant density gradient centrifugation (Sasaki and Ishikawa 1995) and by dissecting out symbiont-laden organs (called "bacteriomes") and tissues prior to DNA purification. However, even when these procedures are applied, there is often a considerable amount of host DNA contamination in the resulting samples. For example, during the recent complete genome sequencing of the insect-associated endosymbionts, *Buchnera aphidicola* (Tamas et al. 2002) and *Wigglesworthia glossinidia* (Akman et al. 2002), the majority of shotgun clones contained host DNA, even though bacterial cells were extensively purified prior to DNA extraction.

In this study, we devised a method for recovering essentially pure DNA samples from host-contaminated preparations by taking advantage of the fact that symbiotic bacteria, aside from having very reduced genomes sizes, have among the lowest genomic G+C contents of any organisms (Moran 2002). Our procedure employs a strategy that facilitates enriched cloning of DNA fragments of distinctive base composition. The utility of this method is demonstrated by the selective cloning of host and symbiont DNA from whole bacteriomes of the psyllid, *Pachypsylla venusta*, which harbor the symbiont Candidatus *Carsonella ruddii* (Clark et al. 2001). Using the methods described in this study, we constructed a shotgun library composed entirely of Candidatus *C. ruddii* DNA sequences.

Materials and Methods

Insect collection and bacteriome DNA preparation

P. venusta was chosen as the focus for this study because only a single endosymbiont, Candidatus *C. ruddii*, has been identified in this species (Clark et al. 2001). Galls containing 4th instar *P. venusta* were collected during October from hackberry trees in Tucson, Arizona. Insects were removed from galls and dissected in bacteriological saline (0.85% w/v NaCl). Candidatus *C. ruddii* lives within the cytoplasm of host cells that are packaged into a single orange-colored bacteriome within each larval insect (Thao et al. 2000). The entire bacteriome was removed from each insect and rinsed three times in bacteriological saline. DNA was prepared from approximately 100 bacteriomes using the DNeasy tissue kit (Qiagen Inc., www.qiagen.com).

Restriction enzyme digestion

To produce enriched shotgun DNA libraries, we selected restriction enzymes generating blunt ends that were most likely to cut host and symbiont DNA at substantially different frequencies, while producing fragments of a suitable size for TOPO cloning (Fig. 1). The choice of restriction enzymes was based on estimates of the genomic base composition using previously published sequences for Candidatus *C. ruddii* (19.9% G+C; Clark et al. 2001) and its psyllid host (37.5% G+C from the limited sequence available; Thao et al. 2000). On average, the *Swa*I recognition site (ATTT^AAAT) is predicted to occur once every 1.5-kbp in the genome of Candidatus *C. ruddii* and once every 11-kbp in the genome of the psyllid host, based on the assumption that the genomic composition is homogeneous in each genome and that the available sequence is representative. This assumption is less reliable for the host genome because bacterial genomes are relatively homogeneous in composition (Sueoka 1962). The *Bsa*AI recognition site (YAC^GTR) is predicted to occur once every 2.5-kbp in the Candidatus *C. ruddii* genome and once every 1.1-kbp in the psyllid genome, based on these same assumptions.

P. venusta-bacteriome DNA was digested in separate reactions with *Swa*I (ATTT^AAAT) and *Bsa*AI (YAC^GTR). For digestion with *Swa*I, 4 µg of DNA was digested with 4 units of restriction enzyme in a 100 µl reaction volume at 25 °C for 30 min. For digestion with *Bsa*AI, 4 µg of DNA was digested with 25 units of restriction enzyme in a 100 µl reaction at 25 °C for 30 min. After electrophoresis through 1% agarose gels, the 0.5–1.5-kbp fraction of *Swa*I- and *Bsa*AI-digested DNA were excised from the gel and recovered with Quik-Pik electroelution capsules (Stratagene, www.stratagene.com).

Topoisomerase-mediated cloning

Restriction enzyme-digested DNA was purified by phenol extraction and concentrated by ethanol precipitation. Prior to cloning, DNA was dephosphorylated with shrimp alkaline phosphatase (Promega, www.promega.com) according to the enzyme manufacturer's recommendations. After heat inactivating the phosphatase at 65 °C for 15 min, cloning was performed with the Zero-Blunt TOPO PCR Cloning Kit (Invitrogen, www.intergenco.com), using 200 ng of dephosphorylated insert DNA, according to the manufacturer's instructions. TOPO cloning reactions were transformed into electrocompetent TOP10 cells (Invitrogen) to achieve a high yield of recombinant clones.

DNA sequencing and data analysis

Plasmid DNA inserts from the TOPO cloning reactions were sequenced at the University of Arizona sequencing facility, using M13 forward and reverse primers. To determine the relative abundance of symbiont and host DNA in the *P. venusta* bacteriome DNA preparation, we sequenced 33 clones from the DNA library generated from *Bsa*AI-digested DNA, and 46 clones from the library generated from *Swa*I-digested DNA. To identify putative homologs, we applied tBLASTx, in which nucleotide sequences are translated into all six putative reading frames and subjected to BLAST search (Altschul et al. 1990) against a similarly translated database. Because Candidatus *C. ruddii* has such an extreme G+C content, resulting in the substantial enrichment of peptide sequences with amino acids corresponding to low G+C codons, tBLASTx searches tend to return matches from other low G+C genomes, regardless of actual homology or evolutionary relationship. Because our aim was to determine whether a given clone was sampled from the symbiont versus host genome, it was appropriate to use a database most suited to discriminate between these two choices and not a much larger database that would be likely to yield many random alignments. To facilitate the recovery of true homologs (as opposed to chance

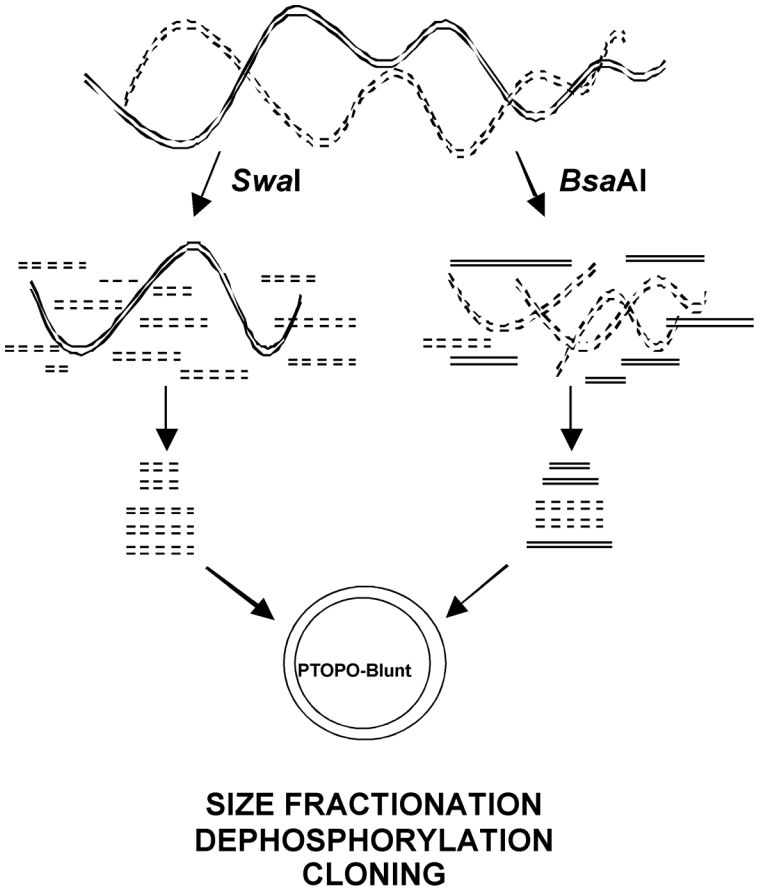

Figure 1. Overview of the differential digestion and TOPO cloning procedure. A heterogenous DNA sample is differentially digested with restriction enzymes that generate blunt ends and cut at different frequencies according to base composition. Restriction fragments are size-fractionated, dephosphorylated and cloned into the pTOPO-Blunt vector.

matches to sequences from similarly low G+C genomes or portions of genomes), BLAST searches were limited to the complete eubacterial and insect nucleotide sequence databases as defined within the nonredundant public databases at NCBI (www.ncbi.nlm.nih.gov:80/entrez) under the taxonomic categories (Insecta = txid50557 and Bacteria = txid2). We used a probability cutoff value of less than e^{-2} for recognition as a significant match. (The e-value is an approximate measure of the likelihood of a alignment of this degree of similarity or higher given a sequence database of a certain size; the chosen value corresponds to approximately a 1% chance or less.) In cases where sequences matched a structural or transfer RNA, overall levels of similarity and significance are based on nucleotide sequences.

Results

To test the efficacy of our method in recovering symbiont-specific DNA, we used BLAST to identify the source (i.e., bacterial symbiont or insect host) of each sequenced insert from the size-fractionated *Swa*I and *Bsa*AI libraries of *P. venusta* bacteriome DNA (Table 1). Of the 46 clones sequenced from the *Swa*I library, 27 contained inserts that shared significant sequence similarity with bacterial sequences, and none had significant hits with insect sequences in the available database. Of the 33 clones sequenced

Table 1. Best-hit BLAST homologues of sequences from the SwaI and BsaAI libraries.*

***Swa* I**

Clone**	Length	%G+C	Best-match Gene	Best-match Organism	Accession Number	Score	Similarity [a]
1	756	13%	*rpl10* and *rpl7*	*Carsonella ruddii*	AF274444	1 e^{-111}	99 / 99
2	797	15%	*metG*	*Escherichia coli*	K02671	5 e^{-12}	32 / 57
3	686	13%	*trpS*	*Carsonella ruddii*	AF211141	5 e^{-53}	69 / 82
4	848	21%	*carB*	*Buchnera aphidicola* BP	AE014016	2 e^{-42}	72 / 90
5	643	14%	Cthe 2622	*Clostridium thermocellum*	NZ_AAAJ01000154	2 e-10	46 / 69
6	326	13%	*cls*	*Clostridium perfringens*	AB017186	9 e-6	22 / 32
7	640	16%	*ilvE*	*Haemophilus influenzae*	U32798	4 e-6	23 / 37
8	636	25%	*clpP*	*Buchnera aphidicola* APS	AP001119	2 e-55	88 / 119
9	636	25%	*clpP*	*Buchnera aphidicola* APS	AP001119	2 e^{-55}	88 / 119
10	682	19%	*carB*	*Buchnera aphidicola* BP	AE014016	5 e-41	79 / 101
11	553	14%	*thdF*	*Ureaplasma urealyticum*	AE002101	2 e-8	37 / 64
12	430	13%	*rpsM*	*Bacillus halodurons*	AP001507	8 e-9	44 / 76
13	531	11%	*lysA*	*Campylobacter jejuni*	AL139074	2 e-4	31 / 56
14	408	11%	OB0096	*Oceanobacillus iheyensis*	AP004593	7 e-4	26 / 47
15	641	13%	*miaB2*	*Thermoanaero. tengcongensis*	AE013095	3 e-17	36 / 66
16	226	12%	jhp 0254 (*yleA*)	*Helicobacter pylori*	AE001463	3 e-6	33 / 47
17	333	14%	Chut 1228 (*carB*)	*Cytophaga hutchinsonii*	AABE01000043	4 e-6	19 / 25
18	637	24%	*rps7* and *fusA*	*Carsonella ruddii*	AF274444	1 e-131	185 / 185
19	264	15%	Lmes 1055 (*nifS*)	*Leuconostoc mesenteroides*	NZ_AAA001000023	2 e^{-11}	42 / 61
20	264	15%	Lmes 1055 (*nifS*)	*Leuconostoc mesenteroides*	NZ_AAA001000023	2 e-11	42 / 61
21	264	15%	Lmes 1055 (*nifS*)	*Leuconostoc mesenteroides*	NZ_AAA001000023	2 e-11	42 / 61
22	642	13%	Avin 1826 (*argG*)	*Leuconostoc mesenteroides*	AAAD01000080	2 e-21	59 / 87
23	331	15%	*sodA*	*Escherichia coli*	AP002567	3 e-13	41 / 63
24	637	11%	Psyr 0984 (*glyRS*)	*Pseudomonas syringae*	AABH01000002	7 e-4	25 / 40
25	498	13%	*aspS*	*Buchnera aphidicola* SG	AE14107	7 e-21	57 / 89
26	449	16%	*rpl7* and *rpoB*	*Carsonella ruddii*	AF274444	3 e-82	106 / 107
27	449	16%	*rpl7* and *rpoB*	*Carsonella ruddii*	AF274444	3 e-82	106 / 107

***Bsa* AI**

Clone***	Length	%G+C	Best-match Gene	Best-match Organism	Accession Number	Score	Similarity [a]
1	618	38%	*tkt*	*Carsonella ruddii*	AF291051	2 e-34	17 / 17
2	649	16%	*tkt*	*Carsonella ruddii*	AF291051	2 e-90	42 / 43
3	594	34%	23S rRNA	*Carsonella ruddii*	BT001724	0	432 / 446
4	633	39%	23S rRNA	*Carsonella ruddii*	AF211143	0	138 / 138
5	615	38%	*tkt*	*Carsonella ruddii*	AF291051	1 e-55	40 / 50
6	556	35%	*rps12*	*Carsonella ruddii*	AF274444	2 e-53	47 / 47
7	633	18%	5S rRNA	*Acyrthosiphon magnoliae*	AMRRNSS	8 e-17	85 / 95
8	737	41%	CG17417	*Drosophila melanogaster*	AE003089	1 e-4	27 / 40

*Longest match of amino acids (or nucleotides for RNA genes) between the query and reference sequences. Alignments may contain other regions of similarity.

**An additional 19 sequences had no hit with probability $<e^{-2}$

***An additional 25 sequences had no hit with probability $<e^{-2}$

Table 2. Expectation of average fragment lengths generated by restriction enzymes in random sequences of different base composition.

	A+T content			
Enzyme (recognition site)	**20%**	**40%**	**60%**	**80%**
Hae III (GG^CC)	0.04 kb	0.12 kb	0.63 kb	10.1 kb
Nae I (GCC^GGC)	0.24 kb	4.57 kb	15.6 kb	1.0 mb
Pvu II (CAG^CTG)	3.91 kb	3.09 kb	6.94 kb	62.5 kb
Bsa AI (YAC^GTR)	2.40 kb	1.11 kb	1.11 kb	2.40 kb
Eco RV (GAT^ATC)	62.5 kb	6.94 kb	3.09 kb	3.91 kb
Ssp I (AAT^ATT)	1.0 mb	15.6 kb	4.57 kb	0.24 kb
PmeI (GTTT^AAAC)	6.25 mb	173 kb	34.3 kb	24.4 kb
Swa I (ATTT^AAAT)	100 mb	0.39 mb	15.2 kb	1.52 kb

from the *Bsa*AI library, eight had significant matches with bacterial sequences and two with insect sequences.

Because the current insect database is not as robust as that for bacteria, and because few sequences for insects closely related to the psyllid host of Candidatus *C. ruddii* have been deposited, we lowered the stringency in our BLAST searches to uncover any additional matches to sequences from our libraries. For the *Swa*I libraries, there were two additional matches, both to bacterial symbiont sequences; and among the sequenced *Bsa*AI inserts, there were five additional matches to insect sequences, using the less stringent cutoff of e^{-1}. Still, over half of the sequences generated from our *Bsa*AI libraries had no significant matches in the databases and likely represent highly diverged insect sequences.

The base composition of inserts differed markedly between the *Swa*I and *Bsa*AI libraries reflecting, in part, the differential representation of host and symbiont sequences in the two libraries. In the *Swa*I library, the base composition of sequenced inserts ranged from 1 to 25% G+C (mean = 13.7% G+C) and from 11 to 25% G+C for those with significant matches to sequences in the databases. In contrast, the base composition of sequenced inserts from the *Bsa*AI library ranged from 16 to 41% G+C (mean = 31.1% G+C), with sequences originating from the symbiont and the insect host both displaying high and low G+C contents. In the *Bsa*AI library, there is not complete discrimination of insect host and bacterial symbiont clones based on base composition: in fact, the sequenced *Bsa*AI insert having the highest G+C contents (39%) corresponded to a portion of the 23S rRNA gene of Candidatus *C. ruddii*. Based on full genome sequences of other low G+C bacteria, the rRNA operon is likely to be the region with the highest G+C content in this organism.Based on BLAST similarity searches, most sequences are of bacterial origin in both the *Swa*I and *Bsa*AI libraries. Still, over half of the sequenced inserts (19/46 from the *Bsa*AI library and 25/33 from the *Swa*I library) displayed no significant match to bacterial or insect genes. Due to the low G+C content of the genome of Candidatus *C. ruddii*, many of the genes recovered were found to be most similar to the few sequences of this organism or of other low G+C symbionts (e.g., Buchnera spp.) already deposited in Genbank, as opposed to their true homologs in other bacterial genomes. Several of the sequenced inserts from both the *Bsa*AI and *Swa*I libraries showed a significant hit to the same sequence in the queried databases (Table 1). In some of these cases (e.g., for *Swa*I clones 8, 9, 23, 24, 25, 26, 27), the identical restriction fragment was cloned and sequenced, whereas others accommodate different inserts harboring genes with small, distinct regions of similarity to one portion of a reference sequence. The *Bsa*AI sequences with no matches are likely to represent sequences from the insect nucleus, based on their relatively high G+C content and on the likelihood of a very low G+C content in the entire symbiont genome apart from the structural RNA genes.

Discussion

We describe an approach that takes advantage of genome-wide differences in base composition to generate enriched shotgun libraries from samples that contain mixed populations of DNA. Such libraries can facilitate the recovery of genome sequences from organisms that have not been isolated or propagated in pure culture, and is especially useful for host-associated bacteria, whose small genomes have been shown to have extreme bias in base composition (Akman et al. 2002; Andersson et al. 1998; Fraser et al. 1995; Fraser et al. 1997; Moran and Wernegreen 2000; Shigenobu et al. 2000). Moreover, the technique is very efficient and requires very little starting material: the described methodology was performed with a total of less than four micrograms of DNA.

By selecting restriction enzymes that cleave with very different frequencies in the source genomes and by cloning restriction fragments of a specific size class, we generated shotgun DNA libraries enriched for inserts that have a very restricted base composition. Our goal was to construct a shotgun library enriched with DNA from the low G+C bacterial symbiont, Candidatus *C. ruddii*, which lives within a specialized organelle, the bacteriome, in the hackberry psyllid, *P. venusta*. The DNA extracted from *P. venusta* bacteriomes was initially digested with *Bsa*AI, and most of the clones (25/33) gave no significant hits to the databases, as expected if most clones in this library are host insert sequences. It is notable that of the few hits to Candidatus *C. ruddii*, several are to rRNA genes, which are always the regions of highest GC content in AT-biased bacterial genomes. In contrast, the library derived from the *Swa*I digestion of bacteriome DNA contained only inserts that have an extremely low G+C content which is typical of Candidatus *C. ruddii*. Of these, a majority (27/46) showed significant similarity to bacterial sequences in the database. Only a few of these are hits to Candidatus *C. ruddii*, as expected since very little sequence has previously been determined from this symbiont genome (Thao et al. 2000; Clark et al. 2001).

The most important factor in this procedure is the choice of restriction enzyme(s) prior to cloning, and this depends on the base composition of target and contaminating DNAs. To facilitate efficient cloning using the TOPO vector, restriction enzymes need to generate target DNA fragments in the range of 0.5 to 4 kbp. This size range is optimal for cloning in the TOPO vector and presents an optimum size for sequencing. Given the base composition of target and contaminating DNAs, suitable restriction enzymes can be identified by estimating the probability of site frequency, under the assumption of random occurrence of bases around the genome (Table 2). Because bacterial genomes are relatively homogeneous in base composition compared to eukaryotic genomes, the values in Table 2 give a relatively accurate representation of fragment sizes from each enzyme. In circumstances where insufficient blunt-cutting restriction enzymes are available, it is also possible to use enzymes that generate an overhang by applying a polymerase fill-in step prior to cloning. Note that the use of enzymes having degenerate site specificities, such as *Bsa*AI, serves as a control in the procedure because their cutting frequencies vary little with the base composition of the target DNA.

Although highly enriched for bacterial symbiont sequences, the resulting libraries are not expected to encompass the entire Candidatus *C. ruddii* genome. Even given a very small genome size as in other insect symbionts (Akman et al. 2002; Moran et al. 2003; Tamas et al. 2002; van Ham et al. 2003), one would not expect to sample an identical clone insert multiple times in a survey of only 80 clones in random shotgun libraries. The repeated recovery of the same clone, particularly from the *Swa*I library, could be due to the instability of the low G+C DNA of Candidatus *C. ruddii*. In a previous study it was noted that certain lambda-ZAP clones carrying Candidatus *C. ruddii* DNA were unstable when transformed into *E. coli* (Clark et al. 2001). If such instabilities are manifested in low copy number phage vectors such as lambda-ZAP, we would expect this problem to be exacerbated in high copy number plasmid vectors such as the pTOPO Zero Blunt vector. The representation bias in the Candidatus *C. ruddii* library could also be due to the fact that the size-fractionated DNA used to construct this library contained mostly products of complete enzyme digestion. For more complete coverage and to facilitate whole genome sequencing, additional libraries would need to comprise products of partial digestions and/or additional restriction enzymes with different site specificities.

Because bacterial genomes have relatively homogenous base compositions (Sueoka 1962), there is expected to be relatively little intrinsic bias in restriction site frequency throughout individual genomes. Exceptions might include regions introduced through lateral transfer (termed 'islands') and regions encoding structural RNAs, which often have base compositions that differ from the genome as a whole (Galtier and Lobry 1997; Wang and Hickey 2002).

Beyond the examination of low G+C genomes of endosymbionts, the approach described in this study can be used for enrichment in any situation imposing the need to recover and clone individual DNA species from mixed populations that have differences in base composition bias.

Acknowledgements

This research was supported by NSF grant 9981432 to H.O. and NSF grant 9978518 to N.A.M.

References

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

Altschul SF, Gish W, Miller W, Myers, EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark, UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396: 133–140.

Beja O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, Villacorta R, Amjadi M, Garrigues C, Jovanovich SB, Feldman RA, DeLong EF. 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology* 2: 516–529.

Clark MA, Baumann L, Thao ML, Moran NA, Baumann P. 2001. Degenerative minimalism in the genome of a psyllid endosymbiont. *Journal of Bacteriology* 183: 1853–1861.

Cummings CA, Relman DA. 2002. Genomics and microbiology. Microbial forensics "cross-examining pathogens." *Science* 296: 1976–1979.

DeLong EF, Pace NR. 2001. Environmental diversity of bacteria and archaea. *Systematic Biology* 50: 470–478.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton GG, Kelley JM. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270: 397–403.

Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton RA, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, Gwinn M. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390: 580–586.

Galtier N, Lobry JR. 1997. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *Journal of Molecular Evolution* 44: 632–636.

Moran NA. 2002. Microbial minimalism: Genome reduction in bacterial pathogens. *Cell* 108: 583–586.

Moran NA, Dale C, Dunbar H, Smith WA, Ochman H. 2003. Intracellular symbionts of sharpshooters (Insecta: Hemiptera: Cicadellinae) form a distinct clade with a small genome. *Environmental Microbiology* 5: 116–126.

Moran NA, Wernegreen JJ 2000. Lifestyle evolution in symbiotic bacteria: Insights from genomics. *Trends in Ecology and Evolution* 15: 321–326.

Sasaki T, Ishikawa H. 1995. Production of essential amino acids from glutamate by mycetocyte symbiont of the pea aphid, *Acyrthosiphon pisum*. J*ournal of Insect Physiology* 41: 41–46.

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

Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF. 1996. Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* 178: 591–599.

Sueoka N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. *Proceedings of the National Academy of Sciences, USA* 48: 582–592.

Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom, JP, Moran NA, Andersson SGE. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 296: 2376–2379.

Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P. 2000. Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Applied and Environmental Microbiology* 66: 2898–2905.

van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proceedings of the National Academy of Sciences, USA* 100: 581–586.

Wang H-C, Hickey DA. 2002. Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. *Nucleic Acids Research* 30: 2501–2507.