JOURNAL OF BACTERIOLOGY, Feb. 2007, p. 1238–1243 Vol. 189, No. 4
0021-9193/07/$08.00+0 doi:10.1128/JB.01183-06

Population Structure of Plasmid-Containing Strains of *Streptococcus mutans*, a Member of the Human Indigenous Biota▽

Page W. Caufield,[1,2]* Deepak Saxena,[1] David Fitch,[3] and Yihong Li[1]

College of Dentistry,[1] School of Medicine,[2] and Department of Biology,[3] New York University, New York, New York

Received 31 July 2006/Accepted 25 October 2006

There are suggestions that the phylogeny of *Streptococcus mutans*, a member of the human indigenous biota that is transmitted mostly mother to child, might parallel the evolutionary history of its human host. The relatedness and phylogeny of plasmid-containing strains of *S. mutans* were examined based on chromosomal DNA fingerprints (CDF), a hypervariable region (HVR) of a 5.6-kb plasmid, the rRNA gene intergenic spacer region (IGSR), serotypes, and the genotypes of mutacin I and II. Plasmid-containing strains were studied because their genetic diversity was twice as great as that of plasmid-free strains. The CDF of *S. mutans* from unrelated human hosts were unique, except those from Caucasians, which were essentially identical. The evolutionary history of the IGSR, with or without the serotype and mutacin characters, clearly delineated an Asian clade. Also, a continuous association with mutacin II could be reconstructed through an evolutionary lineage with the IGSR, but not for serotype e. DNA sequences from the HVR of the plasmid produced a well-resolved phylogeny that differed from the chromosomal phylogeny, indicating that the horizontal transfer of the plasmid may have occurred multiple times. The plasmid phylogeny was more congruent with serotype e than with mutacin II evolution, suggesting a possible functional correlation. Thus, the history of this three-tiered relationship between human, bacterium, and plasmid supported both coevolution and independent evolution.

The mutans streptococci (MS) are causally linked to dental caries in humans and in laboratory animals (22, 44). Two members of the MS group, *Streptococcus mutans* and *Streptococcus sobrinus*, are associated with human caries, while the other members of the MS family are predominantly present in other mammals. This distribution of MS indicates host specialization. *S. mutans* is generally transmitted vertically, mother to child, and this mode of transmission results in the clustering of strains, not only within families (4, 11, 19, 20), but within racial cohorts as well (8). This proclivity for vertical transmission along maternal lines could make the population structure of *S. mutans* relevant to the evolutionary history of its human host, not unlike the mitochrondrial-DNA model (6). At present, however, the population structure of *S. mutans* remains fragmented, and only hints of phylogenetically informative patterns in *S. mutans* populations exist (5, 8, 14, 31).

Genetic diversity is well documented within *S. mutans*. For example, variability among strains of *S. mutans* has been demonstrated for the presence of plasmids (8, 11, 23); mutacin I, II, III, and IV operons (5, 14, 33, 34, 46); serotype antigens (38, 46); competence (15, 21, 29); and the *msm*, *bgl*, *cel*, and *gtfBC* loci (46), among others. Indeed, the genome of *S. mutans* UA159 shows that nearly one-third of its open reading frames are of unknown function and 16% are unique to *S. mutans*. Recent work by Waterhouse and Russell (46) shows a mosaic of different genetic loci, or what they call "dispensable genes," distributed among strains of *S. mutans*. Given the wide distribution and diversity of genotypes and genetic loci in *S. mutans* cited above, it seems likely that different strains of *S. mutans* will have both unique and common genetic loci not present on UA159 (37, 46), which should prove useful in charting *S. mutans*' evolutionary history.

A cryptic plasmid resides in ~5% of the isolates of *S. mutans* (8, 24). The function of this plasmid remains unknown, although its sequence has been published (48). Because of its high sequence variability in the hypervariable region (HVR) and its low prevalence, the cryptic plasmid is a useful epidemiological marker for studying transmission (11) and, as here, its phylogenic history. The 5.6-kb plasmid was initially thought to be related to bacteriocin production, because most bacteriocins of gram-positive bacteria are plasmid encoded (43). Subsequent discovery of the chromosomal locus for mutacins I, II, III, and IV (33, 34), coupled with sequencing of the plasmid, showed that mutacins are not plasmid encoded (9, 48). Nonetheless, virtually all known plasmid-bearing strains of *S. mutans* elaborate either mutacin I or II, and these strains are also naturally competent (29). Moreover, mutacin and competence are coordinately expressed as part of an overall mechanism to acquire DNA (16, 28).

Here, we examined the population structures of plasmid-containing strains of *S. mutans* from individuals of different racial/ethnic and geographic backgrounds, using both the hypervariable region of the plasmid and several chromosomal loci to construct phylogenies. We found two significantly incongruent phylogenies, one for the chromosome and another for the plasmid, which displayed independent histories, perhaps as a result of horizontal transfer.

* Corresponding author. Mailing address: New York University College of Dentistry, 345 East 24th Street, New York, NY 10010. Phone: (212) 998-9603. Fax: (212) 995-3994. E-mail: pwc2@nyu.edu.

▽ Published ahead of print on 3 November 2006.

MATERIALS AND METHODS

Isolation of strains and subject population. A total of 33 plasmid-containing strains of *S. mutans* were obtained by screening over 600 samples taken from

TABLE 1. Plasmid-containing strains of *S. mutans*

Strain	Mutacin (serotype)	Race/ethnic group	Origin	Source
LM7	I (e)	Not determined	Guatemala	R. Gibbons and W. Loesche, 1973
AF199	I (e)	African	Central African Republic	This study
AA31, AA37, AA140, AA545, AA174, AA222, AA855, AA669	I (c, e[a])	African American	Birmingham, AL	This study
CA96, CA101, CA103, CA109, CA111, CA113, CA114, CA665, CA143	II (c)	Caucasian	Birmingham, AL	This study
VA318	II (c)	Not determined	Great Lakes, IL	23
JP9-1, JP85-5, JP9-4	I (c)	Japanese	Japan	Y. Sato
CH5A	II (e)	Chinese	Wuhan, China	This study
CH830	II (e)	Chinese	Hainan, China	This study
CH43	I (c)	Chinese	Beijing, China	This study
CH620, CH638, CH639	I (c)	Chinese	Xinjang, China	
HI24	I (e)	Hispanic	New York, NY	This study
AM223	I (c)	Amazon Indian	Guyana	This study
BR15	I (c)	Not determined	Porto Alegre, Brazil	This study
SW114	I (e)	Caucasian	Umea, Sweden	B. Kohler

[a] Strain AA669.

saliva or plaque from caries-active subjects from five continents (Table 1). Samples were processed as previously described (19). Presumed colonies of *S. mutans* were subjected to biochemical confirmation before being screened for mutacin production and the presence of plasmids (7). This study was approved by the Institutional Review Boards of the University of Alabama at Birmingham and New York University.

Genetic characterization. Cell lysis and chromosomal-DNA isolation were performed as described previously (19, 37). All of the strains were subjected to chromosomal-DNA fingerprinting (CDF) using the HaeIII restriction enzyme (10, 19). The presence or absence of plasmids was confirmed by both plasmid-specific primers and visualization on CDF. The plasmid-specific primers F01 (GTTTTGAAGGTCTTGCGATGTCT) and R01 (ATTCAGAAGTCCGTATTATGC) yielded an ~603-bp fragment of the HVR area of the 5.6-kb cryptic plasmid, spanning base pairs 4510 to 5336 on pUA140 (48). Primers MutA-F1 (TACGTTCAGTTACACACATG) and MutA-R (CTTAGCAACAGTAACTATTG) yielded an amplicon of ~210 bp, and Mut2-F (ATAACGGGGGGCTTAAGCTGTA) and Mut2-R (GCCAAGAATGGTCTGAAGAAACA) yielded an amplicon of ~600 bp for the detection of mutacin I and mutacin II, respectively. Primers of the intergenic spacer region (IGSR) were as described previously (12), yielding an amplicon of ~389 bp. Primers for determining the serotype were described elsewhere (38). The HVR and IGSR were sequenced in both directions, using the specific forward and reverse primers described above.

Phylogenetic analyses. To test the robustness of the data for differences in methodology, maximum likelihood (ML) and weighted parsimony were used, as implemented by PAUP version 4.0 beta 10 (42).

For ML, ModelTest (32) was used to identify an efficient model. For all data sets, the HKY85+G model was selected with the following specified parameters and search strategies. For the IGSRs of plasmid-containing strains, the parameters were as follows: $f_A = 0.3100$, $f_C = 0.1866$, $f_G = 0.2465$, and $f_T = 0.2569$ (where f is frequency); transition/transversion ratio, $r = 1.815$ ($\kappa = 3.602$); gamma shape parameter for modeling rate differences among sites, $\alpha = 0.0843$ (modeled using four discrete rate categories). A heuristic search for the ML tree was performed by 100 taxon addition replicates, swapping on all saved trees, so that 181,158 total trees were evaluated. For the HVRs of plasmids, the parameters were as follows: $f_A = 0.3553$, $f_C = 0.1613$, $f_G = 0.1253$, and $f_T = 0.3581$; transition/transversion ratio, $r = 1.401$ ($\kappa = 3.420$); gamma shape parameter, $\alpha = 0.0084$ (using four discrete rate categories). A heuristic search for the ML tree was performed by 100 taxon addition replicates, swapping on all saved trees, so that 13,781 total trees were evaluated. For the IGSRs of strains without plasmids, the parameters were as follows: $f_A = 0.3115$, $f_C = 0.1899$, $f_G = 0.2450$, and $f_T = 0.2536$; $r = 0.915$ ($\kappa = 1.814$); $\alpha = 1.216$ (four discrete categories). The same parameters and search strategy were used for bootstrap analyses, except that neighbor joining was used to generate starting trees for each replication for the plasmid HVR data.

For weighted parsimony, transversions were weighted twice as much as transitions to reflect the greater susceptibility of transitions to be superimposed. Also, we noted that ModelTest suggested that the data fit models where transition rates were higher than transversion rates (see the κ values above). Indel "gaps" were ignored, and parsimony-uninformative characters were excluded. The following strategies were used with the different data sets. For the IGSRs of plasmid-containing strains, two analyses were undertaken, one with only the DNA characters and another including the DNA, serotype, and mutacin characters. In the latter case, the serotype-plus-mutacin partition was weighted equally to the partition of parsimony-informative DNA characters. In both cases, a complete branch-and-bound analysis was performed. A strict consensus was calculated for the multiple maximum-parsimony trees obtained in each case. For bootstrap analysis, starting trees were obtained by 10 random taxon addition sequences per replication. For the HVRs of plasmids, only DNA characters were used. In both the search for the maximum-parsimony tree and bootstrap analysis, 10 random taxon addition sequences were performed to obtain starting trees. Evolutionary changes in mutacin and serotype were reconstructed by parsimony.

To get the mean pairwise distance of ingroup taxa, the HKY85+G model was used to derive a pairwise distance matrix. The average of all cells below the diagonal was calculated with Excel (Microsoft).

Nucleotide sequence accession numbers. The sequences of the plasmid HVRs and IGSRs have been deposited in the GenBank database under accession numbers AF139604 to AF139611, AF077024 to AF077024, and AF093650 to AF093667.

RESULTS

Overall, most plasmid-containing strains of *S. mutans* from unrelated individuals displayed unique chromosomal-DNA fingerprints (Fig. 1A). Strains isolated from unrelated Caucasian subjects (except SW114, from a Swedish adult), however, showed remarkably similar, if not identical, DNA fingerprints (Fig. 1B), even though two strains, VA318 (mutacin II) and T8 (mutacin II; plasmid free), were obtained from individuals as geographically disparate as Michigan and Australia. The remaining strains were obtained from Caucasian subjects presenting for care at the University of Alabama dental clinic (Table 1). Strain CH830, obtained from an individual from Hainan, China, also displayed a CDF similar to those of the Caucasian group (Fig. 1C).

The sequences of the HVRs (603 bp) from the 33 plasmid-containing strains of *S. mutans* yielded 68 parsimony-informative sites, allowing the reconstruction of a well-resolved phylogeny for the plasmids (Fig. 2). The resulting tree, shown in Fig. 2, displayed the same topography and similar branch support (bootstrap) values by two very different methods, weighted parsimony and maximum likelihood. HVRs with

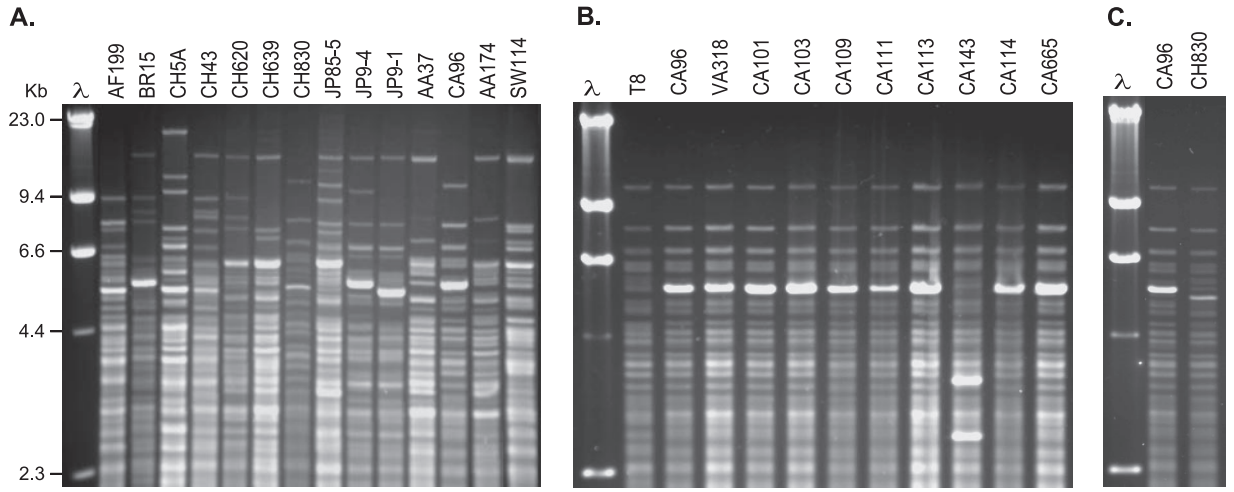

FIG. 1. Chromosomal-DNA fingerprints of plasmid-containing strains of *S. mutans* from unrelated individuals, restricted with HaeIII. (A) Strains from different ethnically/racially/geographically distinct hosts. (B) *S. mutans* from a Caucasian population from the United States and Australia. Strain CA143 contains a 0.9-kb insertion sequence. (C) Strain CH830, obtained from an individual from the southern region of China (Hainan province), displays a chromosomal-DNA fingerprint similar to that of strain CA96, a strain isolated from a Caucasian individual from Birmingham, AL. Lane λ, bacteriophage lambda cut with HindIII size standard.

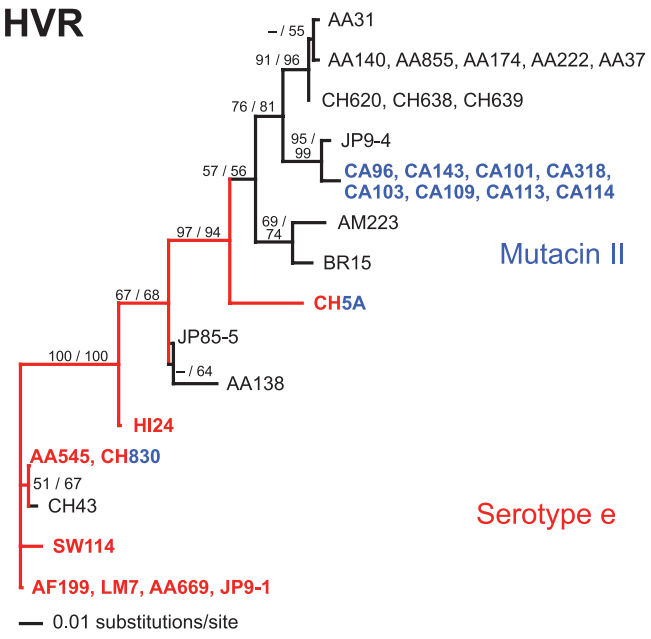

FIG. 2. Unrooted maximum-likelihood phylogeny of the cryptic 5.6-kb plasmid as inferred from HVR sequences (see Material and Methods for details of analysis). Taxa with names in blue represent strains with mutacin II, and those in black represent strains with mutacin I; taxa and branches in red represent strains and ancestral lineages with serotype e. One of two possible reconstructions is depicted for serotype e; the alternative possibility is that serotype e was independently derived for CH5A. The pairs of numbers on branches are bootstrap values: the first number is from a likelihood bootstrap analysis, and the second is from a weighted-parsimony bootstrap (1,000 replicates each). Hyphens and branches without numbers indicate bootstrap values that were below 50%. The branch lengths are proportional to the numbers of substitutions/site, as reconstructed using the HKY85+G likelihood model. Abbreviations for ethnicity of the human host: AF, African; AA, African American; CA, Caucasian American; CH, Chinese; JP, Japanese; BR, Brazilian; AM, Amazon Indian; SW, Swedish Caucasian; HI, Hispanic.

identical sequences were grouped together at the terminal nodes. Most of the HVR sequence was noncoding, as described elsewhere (48).

The observation that at several of the terminal branches, strains of identical sequences clustered with strains from the same ethnicity/geographic location is noteworthy. For example, the CA96, AF199, AA140, and CH620 clonal complexes were, with a single exception (JP9-1), comprised of *S. mutans* from Caucasian, African, African American, and Asian (Chinese) host counterparts, respectively (Fig. 2). Also, strains BR15 and AM223, both from South America, formed a well-supported clade. Such grouping by ethnicity/geography, however, was not apparent at deeper levels in the phylogeny. That is, outside the terminal nodes containing identical sequenced strains, the HVR did not seem to support discrete clades for geographically/racially similar hosts. For example, strains from Asian and African individuals were dispersed throughout the tree, showing no clear cluster of *S. mutans* strains from similar racial/geographic groups.

The evolutionary association between the plasmid and the mutacin II phenotype suggests that the mutacin loci may have been acquired independently three times, possibly by horizontal transfer (blue taxa in Fig. 2). In contrast, an association with serotype e is probably continuous along an evolutionary lineage (red lines in Fig. 2), although this association has been lost independently three times (in the lineages to CH43, JP85-5, and AA138 and a larger clade including BR15, AA31, and JP9-4). That is, the evolutionary history of serotype conversion either is linked to or parallels the plasmid's history. The serotype e strains are found in Asian (CH830 and CH5A) and African/African American (AF199, AA669, LM7, and AA545) hosts, but also in a strain from a Hispanic child (HI24) and Swedish Caucasian (SW114) hosts. This association with the most basally derived hosts (Fig. 3) is consistent with an early introduction of the plasmid into *S. mutans*.

Because it was based on only nine parsimony-informative

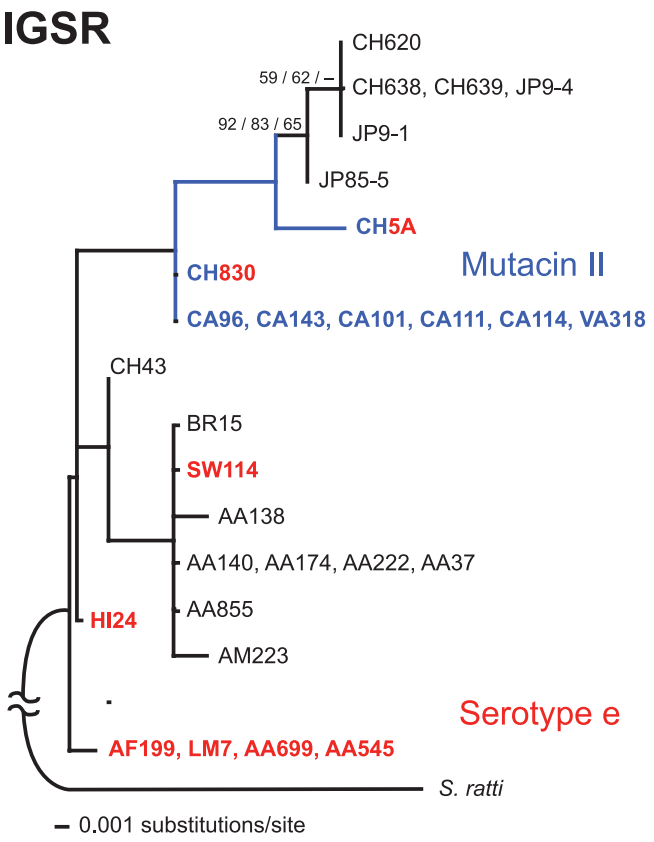

FIG. 3. Maximum-likelihood phylogeny of IGSR sequences from strains with plasmids, rooted with IGSR *Streptococcus ratti* CCUG 27642 (see Materials and Methods for details of the analysis). Taxa and branches in blue represent strains with mutacin II; taxon names in red represent strains with serotype e. The triplets of numbers on branches are bootstrap values: the first two numbers are from weighted-parsimony analysis including or excluding, respectively, serotype and mutacin characters (2,000 and 1,000 bootstrap replications); the third is from a likelihood bootstrap (952 replications). When the serotype and mutacin characters were included and each was weighted the same as the set of DNA characters (i.e., the three data partitions were weighted equally), HI24 grouped with the AF199 cluster with a parsimony bootstrap value of 58%. Hyphens and branches without numbers indicate bootstrap values that were below 50%. The branch lengths are proportional to the numbers of substitutions/site as reconstructed using the HKY85+G likelihood model.

sites from a total of 388 bp, the tree structure from the chromosomal IGSRs was far less resolved than that for the plasmid HVRs (Fig. 3). A well-supported branch showing an "Asian" *S. mutans* clade (CH620, CH638, CH639, JP9-1, JP9-4, and JP85-5) was evident from the IGSR tree (Fig. 3). As a basis for comparison to the amount of signal present in the IGSR locus, a group of plasmid-negative strains of *S. mutans* from essentially the same human population base as the plasmid-positive strains were aligned and analyzed (data not shown). Only four phylogenetically informative sites were present in the 388-bp sequence from nonplasmid strains, yielding essentially no useful phylogenetic signal (data not shown). The mean pairwise genetic distance for plasmid-free strains (171 pairs) was 1.16%, while that for the IGSRs from plasmid strains (120 pairs) was 0.5%. These differences were statistically significant (Student's *t* test; $P < 0.001$), indicating that plasmid-containing strains were twice as variable in the IGSR as non-plasmid-containing strains.

Both weighted-parsimony and maximum-likelihood analyses of the IGSR DNA characters alone produced trees in which only two branches were well supported by bootstrapping. In an attempt to maximize the phylogenetic signal, characters for serotype and mutacin types were added in a maximum-parsimony analysis; each was given character weights equal to the set of nine informative nucleotides so that the three different partitions (serotype, mutacin, and IGSR) were equally weighted (Fig. 3). However, the serotype or mutacin characters did not add much to the resolution; besides the two nodes with bootstrap values shown in Fig. 3, only one other node was supported above 50% (HI24 plus the AF199 cluster; 58% bootstrap value).

Although the poor resolution limits our ability to draw decisive conclusions, there are some interesting features of the IGSR tree that contrast with the plasmid HVR tree. First, association of the strains with mutacin II is continuous along an evolutionary lineage, and associations with serotype e have evolved multiple times. Although it is formally possible that multiple independent changes to the mutacin II genotype occurred, a single-gain–single-loss scenario is most parsimonious, given the ML tree. Second, relationships among the taxa are different. For example, whereas the JP9-4 IGSR is identical to that of CH638 and CH639 but is in a distinct clade from the CA96 cluster (Fig. 3), the plasmid HVR of JP9-4 is most closely related to the CA96 cluster but is phylogenetically distinct from the CH638 and CH639 HVRs (Fig. 2). Also, the AA140 IGSR cluster is essentially identical to the SW114 IGSR (Fig. 3), but the plasmids of these strains are at opposite ends of the tree (Fig. 2).

DISCUSSION

The overall population structure of *S. mutans*, based on both chromosomal and extrachromosomal (plasmid) genetic loci, showed distinct evolutionary lineages, as well as overlapping histories. Not unlike other species of bacteria, the evolutionary histories of different genes are sometimes congruent and sometimes not (13, 25, 30). From our data, the IGSR and mutacin loci share one history, while the plasmid and serotype e share another. For the most part, the chromosomal-DNA fingerprints (Fig. 1) from unrelated individuals displayed distinctive patterns, with the notable exception of strains from Caucasian populations. This diversity could arise from point mutations, recombination, inversions, deletions, or insertions and indicates a high mutation rate and/or a high recombination rate (45). The appearance of strain-specific chromosomal-DNA fingerprints is in agreement with our previous observations and those of others (10, 17–19, 36) and resembles what is seen among strains of *Helicobacter pylori* from different individuals (3). The population structure of *H. pylori* has been described as panmictic, presumably the result of a high rate of recombination (1, 27, 41). It has been proposed that *S. mutans* possesses both a core genome and a dispensable genome (46). This dispensable genome may account for its within-species diversity, as shown by the different CDF profiles. The mutacin and plasmid genetic loci would be considered part of the dispensable genome, because most strains of *S. mutans* do not harbor either locus.

In view of the overall diversity displayed in the CDF profiles, it was surprising to find that strains obtained from Caucasian populations were essentially identical. This pattern of similarity was also found in both HVR and IGSR loci, suggesting that strains from this population were derived from a relatively homogenous founder population. Although most strains came from Caucasians from Birmingham, AL, two strains came from Australia and Michigan. Clearly, sampling from a more heterogeneous Caucasian population is necessary to make definitive inferences about possible bottleneck effects that restricted the diversity of these strains. Moreover, the evolutionary history of plasmid-bearing strains may not be generalizable to all *S. mutans* populations. It is noteworthy, however, that a similar bottleneck effect has been reported for Caucasians (35, 40, 47). This lack of diversity among Caucasian strains may also be interpreted as representing the age of the population; that is, the youngest human race is the least diverse because they have had a shorter time to evolve.

Our data show that plasmid-containing strains of *S. mutans* contain twice as many polymorphisms in the IGSR as the corresponding locus in plasmid-free strains. This finding could indicate that the plasmid facilitates or stabilizes the entry or recombination of foreign DNA because of its greater copy numbers within the cell. Support for this contention is found in a plasmid from strain CA143 from a Caucasian child (mutacin II; serotype c) that contained an insertion element (IS*1216*; GenBank accession no. AF104381) downstream of the HVR. This region shares similarity with the chromosomal *mutF-mutG* operon of mutacin II, in addition to two transposase fragments (SMU.226c and SMU.1329c) present in *S. mutans* UA159 (2). It also suggests that there may be a linkage between mutacin II and the cryptic plasmid. Alternatively, the plasmid's presence maybe simply the result of the cell's genetic competence, promoting the uptake of more than one genetic element, including the mutacin locus. The greater diversity of plasmid-containing strains is also compatible with an early evolutionary association between the plasmid and *S. mutans*, as well as with recent plasmid loss in the plasmid-free strains.

While the HVR region yielded a well-resolved tree, there seems to be little evolutionary congruence between the plasmid and the mutacin loci or the ethnic/geographic host groups, except for those clusters of like strains found at the terminal branches. The HVR portion of the plasmid was selected for constructing phylogenies because of its unusually high concentration of polymorphic sites and because the region was noncoding and hence presumably not subject to selection on gene products. The lack of a clear lineage with the mutacin loci was surprising, because the bacteriocin operons reside on plasmids for most of the lactic acid bacteria in the order *Lactobacillales* (43).

Interestingly, however, the association between the plasmid and the serotype e locus was well supported (Fig. 2). Most strains of *S. mutans* are serotype c. However, transition between serotypes e and c involves gene conversion, because the sequences flanking both sides of the serotype determinants are homologous (38). The source of the converted block of DNA must come from outside the host chromosome, since UA159 (serotype c) has no site with homology to the antigen-specific serotype e coding region (2, 38). The phylogenetic congruence of the serotype with the plasmid history suggests that the serotype conversion was plasmid mediated. A perfect correlation might not be expected, because the plasmid could be lost later, yielding a plasmid-free serotype e strain (46).

Using the IGSR to resolve phylogeny has a solid foundation for differentiating bacteria at the species level, particularly within the genus *Streptococcus* (12). Here, further resolution to the intraspecies level was possible for the plasmid-containing strains, but not for plasmid-free strains. The phylogeny predicted from the IGSR revealed a well-resolved "Asian clade" that included most of the strains from Japan and China (Fig. 3). Many of the strains from the Caucasian and African American racial groups also clustered together at the terminal branches, having similar or identical IGSR sequences. Mutacin II clustering was also consistent with tree topography, but like the racial clusters other than the Asian clade, the bootstrap values fell below 50%.

Other investigators, using different genetic markers, have attempted to examine *S. mutans*' population structure, with little or no resolution, even though multiple loci were analyzed (31, 46). One study attributed this lack of concordance among loci to the dispensable or non-core-genome nature of the loci studied (46). The implication is that these "dispensable" loci arose from horizontal transfer and, like the cryptic plasmid, have their histories in another organism. In a preliminary report, other investigators sequenced several of the housekeeping genes of *S. mutans*, using the multilocus sequence-typing approach (31). They were unable to derive a strong phylogenetic signal, however, due to the small variation within the alleles sequenced. They went on to speculate that *S. mutans* may have only recently associated with its human host, perhaps tied to the development of agriculture, and had little time to diverge. While the multilocus sequence-typing approach has uncovered population structures in several species of bacteria and fungi (26, 39), the method apparently could not resolve *S. mutans*' structure. In another effort to characterize the population structure of *S. mutans*, anchoring the tree based on mutacin production, Balakrishnan and coworkers (5) were able to separate *S. mutans* based on multiple phenotypic and genotypic characters but were not able to construct a consistent phylogeny due to the disparate signal.

In summary, this study examined host-parasite coevolution at four levels: gene, plasmid, *S. mutans*, and human host. Although the data were not sufficient to show parallel evolutionary histories at all four levels, the collective patterns of individual segments allowed the reconstruction, albeit incomplete, of several parallel phylogenies. The data were able to show fairly strong evidence for incongruence between the phylogenies of a cryptic plasmid and *S. mutans*. Intriguingly, this incongruence did not rule out coevolution, since there was a phylogenetic correspondence between plasmid evolution and serotype e, which is encoded on the bacterial chromosome. In contrast, the history of the mutacin loci appears to be independent of the plasmid, consistent with mutacin's chromosomal linkage. With the discovery of additional informative loci, *S. mutans* may serve as a useful model to study coevolution with its human host at a variety of levels.

ACKNOWLEDGMENTS

We thank Jinmei Song and Shawn Zou for their technical support.

The research was supported by NIDCR grants R01DE013937 and DE11147.

REFERENCES

1. **Achtman, M., T. Azuma, D. E. Berg, Y. Ito, G. Morelli, Z. J. Pan, S. Suerbaum, S. A. Thompson, A. van der Ende, and L. J. van Doorn.** 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol. Microbiol. **32:**459–470.
2. **Ajdic, D., W. M. McShan, R. E. McLaughlin, G. Savic, J. Chang, M. B. Carson, C. Primeaux, R. Tian, S. Kenton, H. Jia, S. Lin, Y. Qian, S. Li, H. Zhu, F. Najar, H. Lai, J. White, B. A. Roe, and J. J. Ferretti.** 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc. Natl. Acad. Sci. USA **99:**14434–14439.
3. **Akopyanz, N., N. O. Bukanov, T. U. Westblom, and D. E. Berg.** 1992. PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. Nucleic Acids Res. **20:**6221–6225.
4. **Alaluusua, S., J. Matto, L. Gronroos, S. Innila, H. Torkko, S. Asikainen, H. Jousimies-Somer, and M. Saarela.** 1996. Oral colonization by more than one clonal type of mutans streptococcus in children with nursing-bottle dental caries. Arch. Oral Biol. **41:**167–173.
5. **Balakrishnan, M., R. S. Simmonds, M. Kilian, and J. R. Tagg.** 2002. Different bacteriocin activities of *Streptococcus mutans* reflect distinct phylogenetic lineages. J. Med. Microbiol. **51:**941–948.
6. **Cann, R. L., M. Stoneking, and A. C. Wilson.** 1987. Mitochrondrial DNA and human evolution. Nature **325:**31–36.
7. **Caufield, P. W., N. K. Childers, D. N. Allen, and J. B. Hansen.** 1985. Distinct bacteriocin groups correlate with different groups of *Streptococcus mutans* plasmids. Infect. Immun. **48:**51–56.
8. **Caufield, P. W., K. Ratanapridakul, D. N. Allen, and G. R. Cutter.** 1988. Plasmid-containing strains of *Streptococcus mutans* cluster within family and racial cohorts: implications for natural transmission. Infect. Immun. **56:** 3216–3220.
9. **Caufield, P. W., G. Shah, S. Hollingshead, M. Parrot, and M. C. Lavoie.** 1990. Evidence that mutacin II production is not mediated by a 5.6-kilobase plasmid in *Streptococcus mutans*. Plasmid **24:**110–118.
10. **Caufield, P. W., and T. M. Walker.** 1988. Genetic diversity with *Streptococcus mutans* evident by chromosomal DNA restriction fragment polymorphisms. J. Clin. Microbiol. **27:**274–278.
11. **Caufield, P. W., Y. M. Wannemuehler, and J. B. Hansen.** 1982. Familial clustering of the *Streptococcus mutans* cryptic plasmid strain in a dental clinic population. Infect. Immun. **38:**785–787.
12. **Chen, C. C., L. J. Teng, and T. C. Chang.** 2004. Identification of clinically relevant viridans group streptococci by sequence analysis of the 16S-23S ribosomal DNA spacer region. J. Clin. Microbiol. **42:**2651–2657.
13. **Ghose, C., G. I. Perez-Perez, M. G. Dominguez-Bello, D. T. Pride, C. M. Bravi, and M. J. Blaser.** 2002. East Asian genotypes of *Helicobacter pylori* strains in Amerindians provide evidence for its ancient human carriage. Proc. Natl. Acad. Sci. USA **99:**15107–15111.
14. **Kamiya, R. U., M. H. Napimoga, J. F. Hofling, and R. B. Goncalves.** 2005. Frequency of four different mutacin genes in *Streptococcus mutans* genotypes isolated from caries-free and caries-active individuals. J. Med. Microbiol. **54:**599–604.
15. **Klein, M. I., S. Bang, F. M. Florio, J. F. Hofling, R. B. Goncalves, D. J. Smith, and R. O. Mattos-Graner.** 2006. Genetic diversity of competence gene loci in clinical genotypes of *Streptococcus mutans*. J. Clin. Microbiol. **44:**3015–3020.
16. **Kreth, J., J. Merritt, W. Shi, and F. Qi.** 2005. Co-ordinated bacteriocin production and competence development: a possible mechanism for taking up DNA from neighbouring species. Mol. Microbiol. **57:**392–404.
17. **Kulkarni, G. V., K. H. Chan, and H. J. Sandham.** 1989. An investigation into the use of restriction endonuclease analysis for the study of transmission of mutans streptococci. J. Dent. Res. **68:**1155–1161.
18. **Li, Y., and P. W. Caufield.** 1998. Arbitrarily primed polymerase chain reaction fingerprinting for the genotypic identification of mutans streptococci from humans. Oral Microbiol. Immunol. **13:**17–22.
19. **Li, Y., and P. W. Caufield.** 1995. The fidelity of initial acquisition of mutans streptococci by infants from their mothers. J. Dent. Res. **74:**681–685.
20. **Li, Y., P. W. Caufield, A. P. Dasanayake, H. W. Wiener, and S. H. Vermund.** 2005. Mode of delivery and other maternal factors influence the acquisition of *Streptococcus mutans* in infants. J. Dent. Res. **84:**806–811.
21. **Li, Y. H., P. C. Lau, J. H. Lee, R. P. Ellen, and D. G. Cvitkovitch.** 2001. Natural genetic transformation of *Streptococcus mutans* growing in biofilms. J. Bacteriol. **183:**897–908.
22. **Loesche, W. J.** 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol. Rev. **50:**353–380.
23. **Macrina, F. L., J. L. Reider, S. S. Virgili, and D. J. Kopecko.** 1977. Survey of the extrachromosomal gene pool of *Streptococcus mutans*. Infect. Immun. **17:**215–226.
24. **Macrina, F. L., and C. L. Scott.** 1978. Evidence for a disseminated plasmid in *Streptococcus mutans*. Infect. Immun. **20:**296–302.
25. **Maiden, M. C.** 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. **60:**561–588.
26. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:**3140–3145.
27. **Maynard Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90:**4384–4388.
28. **Merritt, J., J. Kreth, W. Shi, and F. Qi.** 2005. LuxS controls bacteriocin production in *Streptococcus mutans* through a novel regulatory component. Mol. Microbiol. **57:**960–969.
29. **Murchison, H. H., J. Barrett, G. A. Cardineau, and I. R. Curtiss.** 1986. Transformation of *Streptococcus mutans* with chromosomal and shuttle plasmid (pYA629) DNAs. Infect. Immun. **54:**273–282.
30. **Nubel, U., R. Reissbrodt, A. Weller, R. Grunow, M. Porsch-Ozcurumez, H. Tomaso, E. Hofer, W. Splettstoesser, E. J. Finke, H. Tschape, and W. Witte.** 2006. Population structure of *Francisella tularensis*. J. Bacteriol. **188:**5319–5324.
31. **Ogretme, M. S., S. G. T. Do, D. Clark, W. G. Wade, and D. Beighton.** 2006. Multilocus sequencing typing (MLST) of *Streptococcus mutans*. Caries Res. **40:**303–358.
32. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.
33. **Qi, F., P. Chen, and P. W. Caufield.** 1999. Functional analyses of the promoters in the lantibiotic mutacin II biosynthetic locus in *Streptococcus mutans*. Appl. Environ Microbiol. **65:**652–658.
34. **Qi, F., P. Chen, and P. W. Caufield.** 2001. The group I strain of *Streptococcus mutans*, UA140, produces both the lantibiotic mutacin I and a nonlantibiotic bacteriocin, mutacin IV. Appl. Environ Microbiol. **67:**15–21.
35. **Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. W. Feldman.** 2002. Genetic structure of human populations. Science **298:**2381–2385.
36. **Saarela, M., S. Alaluusua, T. Takei, and S. Asikainen.** 1993. Genetic diversity within isolates of mutans streptococci recognized by an rRNA gene probe. J. Clin. Microbiol. **31:**584–587.
37. **Saxena, D., Y. Li, and P. W. Caufield.** 2005. Identification of unique bacterial gene segments from *Streptococcus mutans* with potential relevance to dental caries by subtraction DNA hybridization. J. Clin. Microbiol. **43:**3508–3511.
38. **Shibata, Y., K. Ozaki, M. Seki, T. Kawato, H. Tanaka, Y. Nakano, and Y. Yamashita.** 2003. Analysis of loci required for determination of serotype antigenicity in *Streptococcus mutans* and its clinical utilization. J. Clin. Microbiol. **41:**4107–4112.
39. **Spratt, B. G., W. P. Hanage, B. Li, D. M. Aanensen, and E. J. Feil.** 2004. Displaying the relatedness among isolates of bacterial species—the eBURST approach. FEMS Microbiol. Lett. **241:**129–134.
40. **Stajich, J. E., and M. W. Hahn.** 2005. Disentangling the effects of demography and selection in human history. Mol. Biol. Evol. **22:**63–73.
41. **Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman.** 1998. Free recombination within *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **95:**12619–12624.
42. **Swofford, D. L.** 2002. PAUP. Phylogenetic analysis using parsimony (and other methods), version 4. Sinauer Associates, Sunderland, MA.
43. **Tagg, J. R., A. S. Dajani, and L. W. Wannamaker.** 1976. Bacteriocins of gram-positive bacteria. Bacteriol. Rev. **40:**722–756.
44. **Tanzer, J. M., J. Livingston, and A. M. Thompson.** 2001. The microbiology of primary dental caries in humans. J. Dent. Ed. **65:**1028–1037.
45. **Wang, G., M. Z. Humayun, and D. E. Taylor.** 1999. Mutation as an origin of genetic variability in *Helicobacter pylori*. Trends Microbiol. **7:**488–493.
46. **Waterhouse, J. C., and R. R. Russell.** 2006. Dispensable genes and foreign DNA in *Streptococcus mutans*. Microbiology **152:**1777–1788.
47. **Zhivotovsky, L. A., N. A. Rosenberg, and M. W. Feldman.** 2003. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. Am. J. Hum. Genet. **72:**1171–1186.
48. **Zou, X., P. W. Caufield, Y. Li, and F. Qi.** 2001. Complete nucleotide sequence and characterization of pUA140, a cryptic plasmid from *Streptococcus mutans*. Plasmid **46:**77–85.