CRISPR Regulation of Intraspecies Diversification by Limiting IS Transposition and Intercellular Recombination

Takayasu Watanabe[1], Takashi Nozawa[1], Chihiro Aikawa[1], Atsuo Amano[2], Fumito Maruyama[1,*], and Ichiro Nakagawa[1]

[1]Section of Bacterial Pathogenesis, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Bunkyo-ku, Japan

[2]Department of Preventive Dentistry, Osaka University Graduate School of Dentistry, Suita-shi, Japan

*Corresponding author: E-mail: fumito-m.bac@tmd.ac.jp.

Accepted: May 5, 2013

Data deposition: The nucleotide sequences have been deposited at the DDBJ/EMBL/GenBank databases under the accession AB757108–AB757255 for the CRISPR loci and AB757256–AB757675 for the MLST analysis.

Abstract

Mobile genetic elements (MGEs) and genetic rearrangement are considered as major driving forces of bacterial diversification. Previous comparative genome analysis of *Porphyromonas gingivalis*, a pathogen related to periodontitis, implied such an important relationship. As a counterpart system to MGEs, clustered regularly interspaced short palindromic repeats (CRISPRs) in bacteria may be useful for genetic typing. We found that CRISPR typing could be a reasonable alternative to conventional methods for characterizing phylogenetic relationships among 60 highly diverse *P. gingivalis* isolates. Examination of genetic recombination along with multilocus sequence typing suggests the importance of such events between different isolates. MGEs appear to be strategically located at the breakpoint gaps of complicated genome rearrangements. Of these MGEs, insertion sequences (ISs) were found most frequently. CRISPR analysis identified 2,150 spacers that were clustered into 1,187 unique ones. Most of these spacers exhibited no significant nucleotide similarity to known sequences (97.6%: 1,158/1,187). Surprisingly, CRISPR spacers exhibiting high nucleotide similarity to regions of *P. gingivalis* genomes including ISs were predominant. The proportion of such spacers to all the unique spacers (1.6%: 19/1,187) was the highest among previous studies, suggesting novel functions for these CRISPRs. These results indicate that *P. gingivalis* is a bacterium with high intraspecies diversity caused by frequent insertion sequence (IS) transposition, whereas both the introduction of foreign DNA, primarily from other *P. gingivalis* cells, and IS transposition are limited by CRISPR interference. It is suggested that *P. gingivalis* CRISPRs could be an important source for understanding the role of CRISPRs in the development of bacterial diversity.

Key words: clustered regularly interspaced short palindromic repeat (CRISPR), genome rearrangement, mobile genetic element (MGE), intercellular recombination, diversification, *Porphyromonas gingivalis*.

Introduction

Evolution is of great interest and is crucial for understanding bacteria and their diversification. To clarify bacterial diversification mechanisms, there are several genetic factors to be considered. Genome recombination occurs between bacterial cells, whereas transposition of insertion sequences (ISs) and genome rearrangements are intracellular events. Milkman (1997) described a role for gene transfer such as transduction and conjugation in genome recombination leading to bacterial diversification. Insertion sequence (IS) elements are one class of mobile genetic elements (MGEs) and have been widely identified among bacterial species (Siguier et al. 2006). In *Escherichia coli*, ISs are implicated in genomic diversification (Ooka et al. 2009). ISs are also important in understanding bacterial diversity because they are precursors of repetitive sequences and previous reports indicated the involvement of DNA repeats in genome rearrangements (Hill and Harnish 1981; Achaz et al. 2003; Darling et al. 2008). Such rearrangements can cause phenotypic changes (Dybvig 1993; Ng et al. 1999; Lysnyansky et al. 2001) or create novel prophages (Nakagawa et al. 2003; Nozawa et al. 2011). Relative to the mechanism of IS transposition, both inducive and suppressive regulation are known. IS-excision enhancers in *E. coli* O157:H7 have been suggested to play a role in diversifying the genome of this organism (Kusumoto et al.

© The Author(s) 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

2011). An example of suppressive regulation involves a signal peptide in *Bacillus subtilis* that suppresses the transfer of the integrative and conjugative element ICE*Bs1* by responding to environmental changes (Auchtung et al. 2005). However, because few relevant mechanisms have been investigated, there are likely uncharacterized systems for regulating the events involved in bacterial diversification (Ochman et al. 2000).

As one mechanism for limiting genetic movement between bacterial cells, clustered regularly interspaced short palindromic repeats (CRISPRs) have received increasing attention in recent years. CRISPRs are found in 50% and 80% of sequenced bacteria and archaea, respectively (Bhaya et al. 2011). CRISPRs are intergenic sequences involved in immunity to exogenous sequences and have structurally unique sequence arrays with various spacer sequences inserted between the repeat sequences (Barrangou et al. 2007; Sorek et al. 2008). The spacer sequences are generally acquired from exogenously introduced sequences, for example, from bacteriophages and plasmids, and are transcribed to resist their re-invasion. CRISPR-associated (Cas) genes are responsible for CRISPR function, that is, acquisition of the introduced sequence, expression of the CRISPR array, and interference of re-invading sequences (Bhaya et al. 2011). Utilizing these structural features, CRISPR typing has been utilized for bacteria and archaea (Andersson and Banfield 2008; Horvath et al. 2008; Held et al. 2010; Fabre et al. 2012; McGhee and Sundin 2012). Recently, new CRISPR functions have been recognized other than for immunity. For example, staphylococcal CRISPRs interfere with the spread of antibiotic resistance caused by horizontal gene transfer (Marraffini and Sontheimer 2008). In *Pseudomonas aeruginosa*, the involvement of CRISPRs in biofilm formation has been reported (Cady and O'Toole 2011). Furthermore, the expression of the histidyl-tRNA synthetase gene is regulated by CRISPRs in *Pelobacter carbinolicus* (Aklujkar and Lovley 2010). CRISPR regulation of gene expression is also suggested in *Aggregatibacter actinomycetemcomitans* (Jorth and Whiteley 2012). Therefore, it is expected that more novel CRISPR functions will be revealed in future studies.

Recently, we determined the complete genome sequence of *P. gingivalis* isolate TDC60 (Watanabe et al. 2011). *Porphyromonas gingivalis* is a Gram-negative anaerobic bacillus and is considered as one of the most responsible bacteria for the onset and/or progression of periodontitis (Lamont and Jenkinson 1998; Bostanci and Belibasakis 2012). For clinical investigations, phylogenetic analyses have been carried out in *P. gingivalis* using pulsed-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST) (Koehler et al. 2003; Enersen et al. 2006; Perez-Chaparro et al. 2009). In addition, *fimA* genotyping has been also carried out in several countries. *fimA* of *P. gingivalis* encodes fimbrillin, a major component of *P. gingivalis* fimbriae, and *fimA* is typeable into six groups according to its sequence (Amano 2003). Among these groups, genotype II is more prevalent in periodontitis and is associated with more aggressive forms of the disease (Amano 2003; Amano et al. 2004). However, these typing methods are not useful to understand the propagation and evolution of this organism because of the complexity of their genome structure and the limitation of their resolution power. Therefore, we targeted CRISPR spacers to trace the evolution of *P. gingivalis*. Our preliminary investigation of CRISPRs in three *P. gingivalis* strains demonstrated that CRISPR spacers exhibiting high nucleotide similarity to regions of *P. gingivalis* genomes were present and the number of spacers was diverse among the three genomes (TDC60: 89; W83: 44; and ATCC 33277: 137). Additionally, it is expected that CRISPR typing may be useful in *P. gingivalis* based upon spacer content and abundance. For these reasons, CRISPRs in *P. gingivalis* are worthy of further investigation.

In this study, we examined the applicability of CRISPR typing to 60 *P. gingivalis* isolates in comparison with conventional methods. All of the 2,150 spacers identified from the 60 isolates were investigated in detail. Genetic recombination was examined by split decomposition of MLST. Furthermore, we performed genome sequence alignments to characterize genome rearrangements that were reportedly characteristic of *P. gingivalis* (Naito et al. 2008). These results suggested genome rearrangements mainly involve MGEs. Furthermore, a novel CRISPR function was hypothesized in *P. gingivalis*. The hypothesis involves the limitation of both IS transposition in the cell and the introduction of foreign DNA into *P. gingivalis*. Therefore, this study is expected to be a useful resource for deciphering the detailed mechanisms underlying novel CRISPR functions as well as revealing how CRISPRs regulate chromosomal rearrangements by limiting IS transposition.

Materials and Methods

Bacterial Strains and Culture Conditions

Porphyromonas gingivalis TDC60 was obtained as described previously (Watanabe et al. 2011), and two strains (ATCC 33277, ATCC 53977) were from the American Type Culture Collection (ATCC). In total, we used 60 isolates that included 44 from Japanese patients (supplementary table S1, Supplementary Material online). The 44 Japanese isolates consist of 26 isolates from 7 patients (for which serial numbers were given) and 18 without information of patient sources. All *P. gingivalis* isolates were maintained anaerobically (10% CO_2, 10% H_2, and 80% N_2) at 37 °C in 3% tryptic soy broth (TSB; Becton Dickinson, NJ) or on TSB blood agar plates (3% TSB, 5% sheep blood, 1.5% agarose), supplemented with yeast extract (1 mg/ml), hemin (5 μg/ml), and menadione (1 μg/ml).

Isolation of Genomic DNA and *fimA* Genotyping

Porphyromonas gingivalis cells were cultured in 10 ml supplemented TSB. When the optical density of the culture at 600 nm (OD_{600}) was more than 1.0, it was centrifuged and washed with TNE buffer (10 mM Tris–HCl, 50 mM ethylenediaminetetraacetic acid [EDTA], 100 mM NaCl, pH 8.0). The pellet was suspended in TNE buffer with lysozyme (1 mg/ml; Nacalai Tesque, Kyoto, Japan), sodium dodecyl sulfate (1%; Nacalai Tesque), and ribonuclease A (10 μg/ml; Nacalai Tesque) at 37 °C for 3 h, followed by proteinase K treatment (100 μg/ml; Nacalai Tesque) for 3 h. After phenol–chloroform purification and ethanol precipitation, the pelleted DNA was dissolved in TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0). The genomic DNA was stored at 4 °C until use.

The *fimA* region of each isolate was amplified by polymerase chain reaction (PCR) using M11 and M12 primers (Nakagawa et al. 2000) and partially sequenced to determine its genotype by sequence alignment with known types.

PFGE

A liquid culture of *P. gingivalis* was pelleted and suspended in TE buffer (10 mM Tris–HCl, 50 mM EDTA, pH 8.0) to adjust the OD_{600} to 2.0. Then, the suspension was mixed with melted 2% agarose (Bio-Rad Laboratories, CA) to obtain a plug in which the bacterial cells were embedded. The plug was treated with TNE buffer (10 mM Tris–HCl, 50 mM EDTA, 1 M NaCl, pH 8.0) containing 0.2% sodium lauroylsarcosinate (Nacalai Tesque), 0.2% sodium deoxycholate (Nacalai Tesque), 2 mg/ml lysozyme (Nacalai Tesque), and 2.5 μg/ml ribonuclease A at 37 °C overnight, followed by treatment with 1% sodium dodecyl sulfate and 100 μg/ml proteinase K overnight. After that, the buffer was replaced to TE buffer with no supplement and stored at 4 °C. The plug was digested at 37 °C with 30 U *Not*I in NEBuffer 3 (New England Biolabs, MA).

PFGE was performed with the CHEF Mapper system (Bio-Rad Laboratories). The gel was run at 6.0 V/cm in 0.5× TBE at 16 °C for 20 h and Lambda Ladder PFG Marker (Bio-Rad Laboratories) was used. After running, the gel was stained in ethidium bromide to obtain image data. Dice's distance matrix was calculated from the pattern to construct a dendrogram using the unweighted pair group method with arithmetic mean (Dice 1945). Clusters were formed with a threshold value of 90% identity.

MLST

In MLST analysis, seven chromosomal genes and PCR primers for their amplification were used as described previously (*ftsQ*, *gpdxJ*, *hagB*, *mcmA*, *pepO*, *pga*, and *recA*; Koehler et al. 2003). PCRs were performed using Ex Taq polymerase (TaKaRa, Kyoto, Japan), with cycling conditions of 1 min at 94 °C, followed by 35 cycles of 10 s at 98 °C, 30 s at 55 °C, and 3 min at 72 °C. Amplicons were electrophoretically separated, cloned, and sequenced using an ABI 3730 sequencer (Applied Biosystems, CA). Sequence data were manually trimmed to preserve the regions to be analyzed.

Phylogenetic relationships were investigated with a maximum likelihood (ML)-based tree from the concatenated sequence using MEGA v5.0 (Tamura et al. 2011). We also constructed an ML-based tree using 198 data sets, containing our data and 138 data sets deposited in the *P. gingivalis* MLST database (PubMLST; http://www.pubmlst.org/pgingivalis/, last accessed May 24, 2013). The significance of branching was evaluated by bootstrap analysis of 500 replicates. Clusters were formed with a threshold phylogenetic distance value of 0.004 in the ML-based tree of the 60 isolates. To visualize differences in the allelic profiles of the isolates, a diagram was drawn using eBURST v3 (Feil et al. 2004).

To characterize nucleotide substitutions and the extent of sequence variation, the d*N*/d*S* ratio was calculated using START v2 (Jolley et al. 2001). Genetic diversity was calculated by the following formula: $1 - \sum x_i^2 [n/(n-1)]$, where x_i is the frequency of the *i*th allele and *n* is the number of isolates (Loos et al. 1992). Allelic profiles and calculated values are described in supplementary tables S2 and S3, Supplementary Material online, respectively.

CRISPR Sequence Determination

From the genome information of the laboratory strains and our analysis of TDC60, there are at most four loci in the *P. gingivalis* genome, which are distinguishable by the lengths of their repeats to three types: types 30, 36, and 37. Type 36 is further distinguished by differences in the nucleotide sequences of their repeats: types 36.1 and 36.2. The direction of the CRISPRs was determined in each CRISPR type by examining the directions of the Cas genes and conservation of the sequences adjacent to the CRISPR. Cas genes were annotated using a BLASTP search of the NCBI GenBank Non-redundant Protein Database under the threshold of 80% query coverage and 80% identity; we verified that the annotated Cas genes had the correct protein motif in the NCBI Conserved Domain Database. The arrays of Cas genes were classified according to the classification of Makarova et al. (2011). In supplementary figure S6, Supplementary Material online, the Cas genes are colored according to the style used by Makarova et al. (2011).

We analyzed the spacer content in 60 *P. gingivalis* isolates by PCR amplification and Sanger sequencing. The following primers were used: pgC30F, 5′-GGCTTTTCTGTTTGAATGTGAGGAG-3′; pgC30R, 5′-GTGCAGCCCTTGGTTTATCTTAATC-3′; pgC36.1F, 5′-CTGTGGAATGATGACTTCTCAATCGG-3′; pgC36.1R, 5′-CACACTACTGCACTTTTCAACGC-3′; pgC36.2F, 5′-ACTTCCCCATCAACAGCACAACTTCC-3′; pgC36.2R, 5′-CCTATCAATGACTTATAAAGGGTCG-3′; pgC37F, 5′-CCCAAACGTAACGCATTGGCA-3′; pgC37R, 5′-CCGAGGGTTAGAACGAACGCATA-3′; the number following "pgC" indicates the CRISPR type to be amplified. The primers were designed so they were located adjacent to the CRISPR loci, except for the primer targeting the upstream region of type 30, which exhibited high nucleotide similarity to the sequence adjacent to IS*Pg1* that was located next to the CRISPR (supplementary fig. S1, Supplementary Material online). PCRs were performed using Ex Taq polymerase (TaKaRa) with following conditions: 1 min and 30 s at 94 °C, followed by 35 cycles of 15 s at 98 °C, 30 s at 55 °C, and 3 min at 72 °C. For long target amplification, LA Taq polymerase (TaKaRa) was used and the extension time of the PCR cycle was 10 min. Repeats and spacers were identified using CRISPRFinder (Grissa et al. 2007). For each CRISPR type, the consensus sequence of the repeat was identified from all the repeats of the same type (supplementary table S4, Supplementary Material online) using WebLogo v3.3 (Crooks et al. 2004). The number of spacers in each isolate was compared with those of the phylum Bacteroidetes, calculated from the spacer data sets in the CRISPI database (Rousseau et al. 2009).

CRISPR Typing by Unique Spacers

CRISPR sequences were determined for the 4 types (30, 36.1, 36.2, and 37) in the 60 isolates as described in Supplementary Information. For each CRISPR type, a nonredundant unique spacer list was obtained using an all-to-all BLASTN search with the following criterion: two spacers were regarded to exhibit high nucleotide similarity to each other if the BLASTN bit score was more than 50 (Pride, Salzman, et al. 2012). We performed the BLASTN search under the following conditions: word size 7 and dust filter off. The name of the unique spacer was determined from a combination of the CRISPR type and the serial number of the type, for example, the spacer 30_156 belongs to type 30 and has the serial number 156 of that type. After such designations, they were then further clustered across the four types.

The original spacers of each isolate were searched using BLASTN against the unique spacer list to obtain bit scores, which were then arrayed to generate a numerical matrix. A heatmap was provided for the matrix and two colors were used according to the bit score: red: $\geq$50, yellow: $<$50. For the spacers of all four types and those within each type, dendrograms were constructed by calculating the Euclidian distance of the matrix using the R software package (http://cran.r-project.org/, last accessed May 24, 2013). In each dendrogram, isolates with no spacer were excluded. Distance values were used for clustering in the dendrograms as described with the following thresholds: 250 (all types), 150 (types 30, 36.2, and 37), and 100 (type 36.1).

Nucleotide Similarity Search of CRISPR Spacers

To characterize the spacer sequences, the spacer list was subjected to a BLASTN search against the following seven databases: 1) NCBI GenBank nucleotide database; 2) MGEs in the ACLAME database (http://aclame.ulb.ac.be/, last accessed May 24, 2013; Leplae et al. 2010); 3) human oral-specific assemblies in the Human Microbiome Project (HMBSA: http://hmpdacc.org/HMBSA/, last accessed May 24, 2013; Lewis et al. 2012); 4) Human Oral Microbiome Database (http://www.homd.org/, last accessed May 24, 2013; Chen et al. 2010); 5) metagenomics of the Human Intestinal Tract (MetaHIT: http://www.metahit.eu/, last accessed May 24, 2013; Arumugam et al. 2011); 6) human assemblies of HMBSA specific for non-oral sites; and 7) salivary virome data sets sequenced by Pride et al. (2012) in the MG-RAST web server. There was a difference among them with respect to the body sites of the sequenced samples: 2, 3, 4, and 7 were from oral databases (especially [7] as a salivary virome); 5 and 6 were from nonoral databases. Hits were considered as significant for bit scores $\geq$50, and the subject sequences were annotated using a BLASTX search to the NCBI GenBank Nonredundant Protein Database. We also used the spacer data sets of both bacteria and archaea available in the CRISPI database for nucleotide similarity searches against the seven databases. The presence of ISs was characterized in *P. gingivalis* genomic regions, which encompassed 2-kb upstream and 2-kb downstream of the regions exhibiting high nucleotide similarity to the spacers. For each CRISPR type in the 60 *P. gingivalis* isolates, protospacer-adjacent motifs (PAMs) were predicted from an alignment of the sequences in the databases exhibiting high nucleotide similarity to the spacer using WebLogo v3.3 (Crooks et al. 2004). The 20-bp sequences that were adjacent to both ends of the region exhibiting high nucleotide similarity to the spacer were included in each alignment for PAM prediction.

Genetic Recombination Test

For the MLST data, intercellular recombination tests were performed by split decomposition analysis and calculation of the standardized index of association (I_A^S), using SplitsTree v4.11 and START v2 (Jolley et al. 2001; Huson and Bryant 2006). In the tree, clusters were formed with a phylogenetic distance threshold of 0.004.

Genome Sequence Alignment and Characterization of Rearrangement Breakpoints

The complete genome sequence of *P. gingivalis* TDC60 has been determined by our group (Watanabe et al. 2011: GenBank accession no. NC_015571). For comparison, the genome sequences of two laboratory strains were used (Nelson et al. 2003: W83, NC_002950; Naito et al. 2008: ATCC 33277, NC_010729). Dot plots were drawn from each alignment using GenomeMatcher v1.66 (Ohtsubo et al. 2008). The Nucmer program in MUMmer v3.22 was used for alignment with default settings (Kurtz et al. 2004). In the dot plot areas, lines shorter than 2.5 kb were removed. Two adjacent lines were organized as a fragment if they fulfilled the following conditions: 1) there was consistency between them with respect to the manner of Y-position change (increase/decrease) when the X-position increased; and 2) gaps between them were less than 25 kb in both axes (*X* and *Y*). Rearrangement breakpoints were identified as the ends of each fragment, and breakpoint gaps were the interfragment regions between two adjacent breakpoints.

To investigate the association of MGEs with genome rearrangement, we characterized the presence of MGEs in a 3-kb region, which covered 1.5-kb upstream and 1.5-kb downstream of the breakpoint, using a BLASTN search. We did not use the whole length of the breakpoint gap for the search because we considered that the middle region in a long breakpoint gap was not appropriate for characterization. The searched MGEs included ISs, miniature inverted-repeat transposable elements (MITEs), transposons (Tns) and conjugative transposons (CTns). The 3-kb regions with no significant nucleotide similarity to the MGEs (e value > 1e−50) were further searched for the presence of ribosomal RNA operons or multicopy coding DNA sequences (CDSs). The features of each breakpoint gap were determined with the 3-kb regions of two breakpoints, which were located at both ends of the breakpoint gap. Six values were given as the numbers of each feature because such a number can be determined independently on two genomes per alignment and there are three alignments (TDC60-ATCC 33277, TDC60-W83, and W83-ATCC 33277). The statistical significance of the location of the feature was tested using a two-tailed paired t test.

Nucleotide Sequence Accession Numbers

The nucleotide sequences have been deposited in the DDBJ/EMBL/GenBank databases under the following accession numbers: CRISPR loci, AB757108–AB757255; MLST analysis, AB757256–AB757675. They are also available at the website of our laboratory (http://www.tmd.ac.jp/grad/bac/, last accessed May 24, 2013).

Results

Characterization of Genetic Diversity among *P. gingivalis* Isolates by Conventional Methods

We characterized the intraspecies diversity among 60 *P. gingivalis* isolates (supplementary table S1, Supplementary Material online) using two conventional methods, that is, PFGE and MLST. In PFGE, only 18 isolates exhibited the same band patterns (supplementary fig. S2, Supplementary Material online). By clustering with the similarity threshold values, isolates from the same patient were clustered; an ML-based tree based upon MLST provided similar clusters (supplementary fig. S3, Supplementary Material online). In the ML-based tree, four other clusters were formed by the isolates, which were from different patients and had the same *fimA* type. Additionally, we constructed an ML-based tree for 198 data sets, containing both our data and 138 data sets in PubMLST. The isolates from the same patient were phylogenetically close in the tree (supplementary fig. S4, Supplementary Material online). When we focused on Japanese isolates, they were distributed randomly in the tree, without exhibiting a close relationship with each other.

Allelic profiles also exhibited high diversity among the 60 isolates in the eBURST diagram (supplementary fig. S5 and table S2, Supplementary Material online). Most sequence types were also unique (59/60). In the diagram, several links were observed between the isolates; only two pairs of the isolates were single-locus variants (SLVs) (0.11%: 2/1,770; conceivable number of links was 1,770) and 16 pairs were double-locus variants (DLVs) (0.90%: 16/1,770). In all of the SLVs (2/2) and most DLVs (12/16), the linked isolates were from the same patient, except for the following: TDC263-D5, D41-KS14, HG1025-ATCC 53977, and OS61-"OS58-3." Two-thirds of all isolates (40/60) were distributed as a singleton.

Characterization of Diversity in *P. gingivalis* by CRISPR Typing

For the three available *P. gingivalis* genomes, we first examined the classification of the Cas gene arrays (supplementary fig. S6, Supplementary Material online). Cas genes were detected only near CRISPR types 30 and 37. These arrays near type 30 were classified as type IC because the array contained *cas8c*, the signature gene for type IC, although they formed almost the same line as type IB, except for the position of *cas5* (supplementary fig. S6, Supplementary Material online). The Cas gene array near type 37 was classified as type IIIB due to the existence of *cmr* genes, which are specific for this type (supplementary fig. S6, Supplementary Material online). In each CRISPR type, the Cas genes formed an array toward the CRISPR locus. The sequences adjacent to the end of the CRISPRs were conserved on both sides of the CRISPRs; it was difficult to find a difference in AT-richness between the adjacent sequences on both sides. Thus, we determined CRISPR direction for only types 30 and 37, for which the Cas genes were located nearby; the end of the CRISPRs was determined as a leader end if the Cas genes were located on the adjacent region of the CRISPRs. In the case of types 36.1 and 36.2, we followed the direction of the CRISPRs, which was reported by Nelson et al. (2003) and Naito et al. (2008). The type IC Cas gene is reported to have a role in DNA targeting, whereas type IIIB has a role in both DNA and RNA targeting (Makarova et al. 2011).

We analyzed four CRISPR loci in 60 *P. gingivalis* isolates. The consensus sequences of the repeats were different among the CRISPR types (supplementary table S5, Supplementary Material online). The repeats were highly conserved in types 36.1, 36.2, and 37, whereas a polymorphism was observed in type 30 (supplementary fig. S7 and table S4, Supplementary Material online); some repeats were an exception by generally being 30 bp, including those both longer and shorter than 30 bp. Although the majority of type 30 repeats were 30 bp and highly conserved, some repeats had a T nucleotide at the end instead of C (supplementary fig. S8, Supplementary Material online); most of them were located at the end of the repeat arrays. We were unable to detect a CRISPR locus in only 3 isolates (5%: 3/60; D45, TDC117, and OS58-3); meanwhile, it was shown that presence of each CRISPR locus was diverse in the 57 isolates (table 1). Each CRISPR locus was detected in the isolates in the following proportions: 57% (type 30: 34/60 isolates), 88% (type 36.1: 53/60), 65% (type 36.2: 39/60), and 37% (type 37: 22/60). We identified 2,150 spacers in the 57 isolates; the spacers in type 30 were the most abundant (62%: 1,330/2,150 spacers; table 1). The number of spacers in all of the loci was variable among the isolates, ranging from 0 to 137 (35.8 ± 29.3), for which the average number was almost half of that in the phylum Bacteroidetes (73.8 ± 88.0) in the CRISPI database (Rousseau et al. 2009). However, the distribution of the number of spacer in *P. gingivalis* was almost similar to that of the Bacteroidetes, except for *Rhodothermus marinus* (157 spacers), *Runella slithyformis* (200 spacers), *Haliscomenobacter hydrossis* (243 spacers), and *Saprospira grandis* (442 spacers). We obtained unique spacer lists for each CRISPR locus (table 1) and all CRISPR loci (1,187 unique spacers; supplementary table S6, Supplementary Material online).

The spacer content exhibited by the unique spacers was diverse among the isolates (supplementary fig. S9, Supplementary Material online). Most clusters were formed by isolates from the same patient. When the spacers were limited to each CRISPR type, the characteristics of the cluster in type 36.2 (fig. 1) were different from those in types 30, 36.1, and 37 (supplementary fig. S10, Supplementary Material online). The clusters of the three types constituted isolates with the same *fimA* type, as observed with the conventional methods. On the other hand, the isolates were clustered regardless of their *fimA* type in type 36.2; for example, a cluster was formed by 5 isolates (ESO101, OS61, HNA99, HW24D1, and OMZ314) that were from different patients and had different *fimA* types. Such clusters accounted for 56% of the clusters (5/9) in type 36.2 and were not observed with the other 2 methods or in previous studies (Enersen et al. 2008; Perez-Chaparro et al. 2009).

When focusing on the spacers in the isolates from the same patient, it was notable that slight differences were observed among them despite their being clustered (fig. 2). For example, the isolates D12 and D26 shared 28 spacers and each had specific spacers in type 30 (12 in D19 and 5 in D12). Such slight differences in the spacers were observed in all of the isolates (26 from the 7 patients). In addition, the sharing of 5 spacers in type 36.2 was observed among 3 patients (nos. 2, 3, and 6), indicating the occurrence of genetic recombination.

Intercellular Recombination among *P. gingivalis* Isolates

We examined the impact of intercellular recombination on *P. gingivalis* diversification using our MLST data. Split decomposition analysis was performed to show phylogenetically conflicting signals resulting from recombination (Octavia and Lan 2006). The resulting tree showed the same clusters (fig. 3) as those in the ML-based tree (supplementary fig. S3, Supplementary Material online). There were network-like structures, mainly concentrated around the center of the tree. Some of them influenced almost half the length of the branches, for example, the branches of TDC225 and TDC280. The standardized index value of association, I_A^S, was 0.1247 for the 60 concatenated sequences. All the d*N*/d*S* values for each locus, indicators of positive selection, were less than 1, ranging from 0.0885 to 0.3470 (supplementary table S3, Supplementary Material online).

Genome Rearrangement and MGEs in *P. gingivalis*

To examine the impact of genome rearrangement on *P. gingivalis* diversification, we performed dot plot analysis (fig. 4*A*; supplementary fig. S11, Supplementary Material online). As reported previously (Naito et al. 2008), the fragments in the plot of W83-ATCC 33277 showed one large X-shaped structure in the overall graphic field, indicating that genome rearrangement events occurred in a symmetrical fashion along the replication axis. The plots of TDC60-W83 and TDC60-ATCC 33277 indicated less X-shaped structures, and generated complicated fragment patterns. There was an average of 18 fragments in each dot plot (TDC60-ATCC 33277: 17; TDC60-W83: 15; and W83-ATCC 33277: 22).

A previous study implicated the association of MGEs with rearrangements between two *P. gingivalis* genomes (Naito et al. 2008). In this study, we investigated rearrangement breakpoints among three *P. gingivalis* genomes using objective criteria and analyzed each statistically (see Materials and Methods). We found that MGEs were located at the breakpoints at a high frequency (table 2; supplementary tables S7 and S8, Supplementary Material online). The percentage of MGEs at all breakpoints in each genome was 62.0% on average, ranging from 46.7% to 88.2%. At other breakpoints, rRNA operons, multicopy CDSs, or regions without any characteristic features were observed. Three patterns of multicopy CDSs were observed (supplementary table S9, Supplementary Material online), including the two-copy CDSs mentioned by Naito et al. (2008). The other two patterns were first detected in this study, both of which were hypothetical CDSs dispersed in the genome. Examples of ISs or rRNA operons located at the breakpoints are shown in figure 4*B*. Statistical significance was observed for the location of the MGEs at the breakpoint gaps relative to the other features, as shown in figure 4*C* (62% on average; MGEs to rRNA: $P = 0.0012$, MGEs to multicopy CDS: $P = 0.0007$, MGEs to the regions with no known characteristic features: $P = 0.0074$, two-tailed paired *t* test). Moreover, of the MGEs located at the breakpoints, ISs were more frequent compared with the other MGEs (ISs to MITEs: $P = 0.0037$, ISs to Tns: $P = 0.0009$, ISs to CTns: $P = 0.0049$, two-tailed paired *t* test).

CRISPR Spacers Exhibiting High Nucleotide Similarity to the *P. gingivalis* Genome

We further analyzed the spacers observed in *P. gingivalis* CRISPRs using nucleotide similarity searches of the seven

Table 1
The Numbers of Spacers in Four CRISPR Loci of 60 *Porphyromonas gingivalis* Isolates

Name of Isolate	CRISPR Loci			
	30	36.1	36.2	37
ATCC 33277	119	4	12	2
ATCC 53977	0	6	0	8
W50	23	7	7	12
W83	23	7	7	7
D3	57	9	6	6
D4	57	9	6	6
D5	73	6	0	7
D8	39	5	11	0
D9	39	7	18	0
D12	40	9	5	0
D26	33	9	0	0
D14	34	2	8	0
D15	34	2	8	0
D16	34	2	10	0
D17	34	2	8	0
D18	34	2	8	0
D19	33	2	7	0
D22	34	2	8	0
D23	34	4	8	0
D28	0	25	0	0
D29	0	25	0	0
D45	0	0	0	0
D32	3	9	5	0
D33	3	9	5	0
D34	3	9	5	0
D39	3	9	5	0
D40	3	10	2	0
D41	3	9	5	0
PC9	0	4	10	0
PC13	0	11	15	0
FK2	0	0	18	13
I5	0	9	5	0
KS14	14	6	0	0
L1	64	0	0	9
US4	59	5	0	7
TDC59	67	3	0	4
TDC60	82	3	0	4
TDC117	0	0	0	0
TDC129	0	3	7	10
TDC222	0	6	0	16
TDC225	7	7	6	0
TDC243	0	8	6	3
TDC260	0	5	0	14
TDC263	0	5	0	4
TDC275	91	0	0	10
TDC280	31	8	0	0
TDCH	0	4	2	0
HG184	0	4	2	7
HG564	0	8	0	0
HG1025	89	6	0	7
HW24D1	0	7	1	0
HNA99	0	16	1	0
ESO101	0	5	1	14
ESO132	34	6	4	0
OS30-2	0	0	8	0
OS58-3	0	0	0	0
OS54-1	0	15	7	0
OS61	0	14	1	0
OMZ314	0	2	3	0
Co5	0	4	0	14
Total	1,330	375	261	184
Mean	22	6	4	3
SD	29	5	5	5
Unique spacers	820	173	77	118

(continued)

sequence databases. We found that 29 spacers exhibited high nucleotide similarity to the sequences available in the databases (2.4%: 29/1,187; table 3; supplementary table S10 and fig. S12, Supplementary Material online). This proportion was similar when calculated from the original (not unique) spacers (2.9%: 62/2,150). The rest of the spacers exhibited no significant nucleotide similarity with any known sequence (97.6%: 1,158/1,187, unique spacers; 97.1%: 2,088/2,150, original spacers). In addition, there were only a few spacers exhibiting high nucleotide similarity to known sequences in the seven databases despite using the spacer data sets for both bacteria and archaea available in the CRISPI database (1.4%: 821/58,417). PAMs were not clearly detected in any of the four CRISPR types (supplementary fig. S13, Supplementary Material online), which may be due to the reduced amount of spacers exhibiting high nucleotide similarity to the sequences in the databases.

Unexpectedly, there were 19 spacers (65.5%: 19/29) exhibiting high nucleotide similarity to the 3 *P. gingivalis* genomes (table 3). Of these, most exhibited high nucleotide similarity to the CDSs (18/19) and to at least 2 of the *P. gingivalis* genomes (15/19). For the 3 genomes, we examined the presence of ISs in the adjacent regions around the sequences exhibiting high nucleotide similarity to the 19 spacers; 2 spacers exhibited high nucleotide similarity to the transposase genes in W83 or TDC60 (2/19; fig. 5i), whereas 5 spacers exhibited high nucleotide similarity to the regions close to the ISs within 2-kb upstream and 2-kb downstream in at least 1 of the 3 genomes (5/19; fig. 5ii). The others exhibited no significant nucleotide similarity either to the transposases or the sequences close to the ISs (12/19).

Except for the spacers exhibiting high nucleotide similarity to the *P. gingivalis* sequences, 3 and 5 exhibited high nucleotide similarity to viral (3/29) and bacterial sequences (5/29), respectively, and 2 exhibited high nucleotide similarity to the sequences lacking specification by BLASTX annotation (2/29).

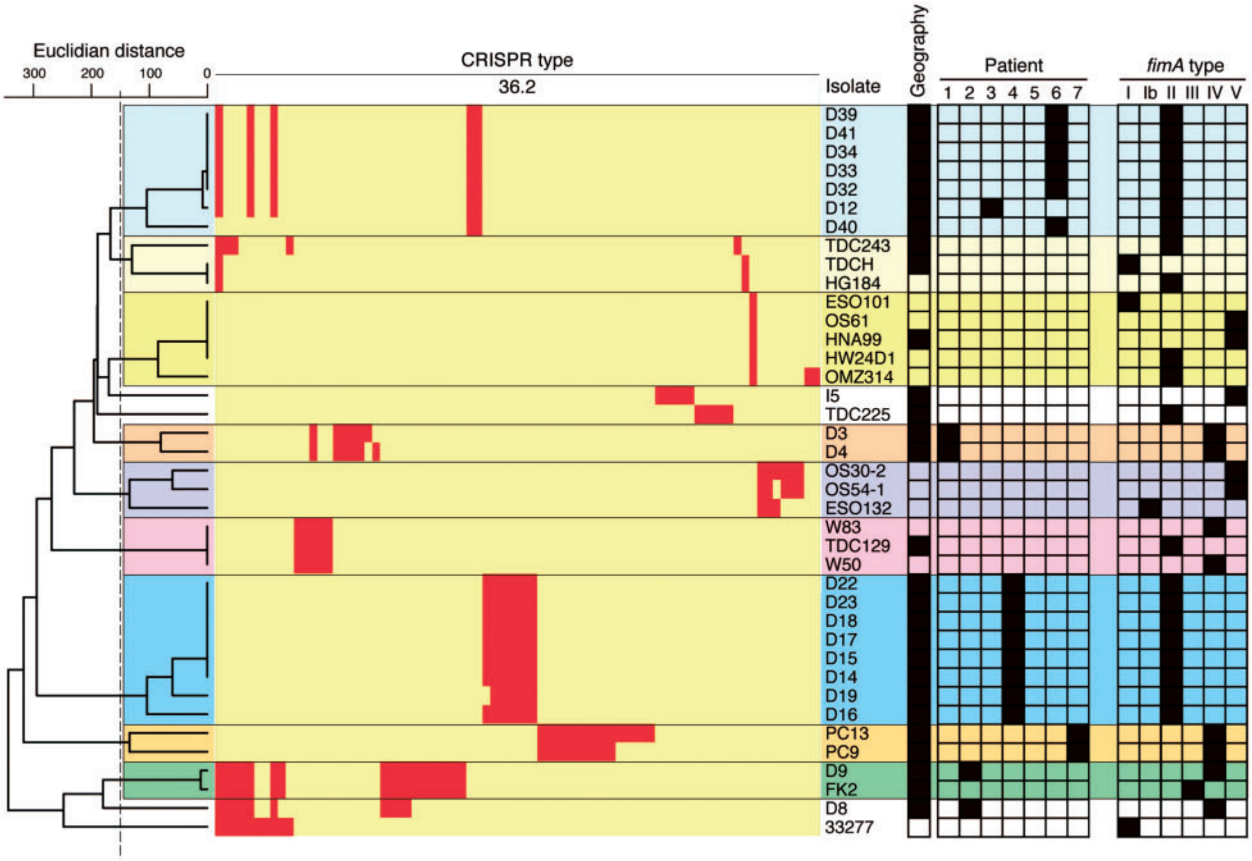

Fig. 1.—Clustering by spacer content in CRISPR type 36.2 of *Porphyromonas gingivalis*. In type 36.2, the presence of each unique spacer is shown using a heatmap. The dendrogram was constructed from Euclidian distances. In the heatmap, the boxes indicate unique spacers and are arrayed horizontally. In the heatmap, 2 colors were used according to the bit score; red: ≥ 50, yellow: < 50. To the right of the isolate's name, the following information is indicated: geographic origin (black: Japan; outlined: overseas or unspecified), patient source (seven patients) and *fimA* type. Eight colors are used to emphasize the clusters.

Discussion

Phylogenetic Diversity in *P. gingivalis* and Effectiveness of CRISPR Typing

The genome structure of *P. gingivalis* is known to be diverse based upon observations in several countries. Nakayama (1995) initially reported diverse pulsotypes among seven reference strains of *P. gingivalis*, which was supported by the report of Perez-Chaparro et al. (2009) with patients in the Republic of Colombia. MLST also exhibited high intraspecies diversity among isolates from various countries including the United States, Indonesia, and Sweden (Enersen et al. 2006). However, these methods are not suitable for clinical examinations because of their complicated processes; therefore, *fimA* genotyping has been widely used as a convenient method to distinguish the status of the periodontal condition (Amano 2003). In this method, *P. gingivalis* strains are determined by using the same specific *fimA* primer sets and similar experimental conditions to investigate the prevalence of *fimA* genotypes from sites in patients with various periodontal conditions (Amano et al. 1999; Nakagawa et al. 2000). Most of these studies indicated that *fimA* genotypes II and IV are predominant in chronic periodontitis-affected sites, and genotypes I, III, and V are predominant in healthy subjects. In this study, PFGE analysis indicated that the genome structure in *P. gingivalis* was highly diverse, even in Japan, a geographically isolated area (supplementary fig. S2, Supplementary Material online). We also demonstrated high diversity using ML-based trees based upon MLST (supplementary figs. S3 and S4, Supplementary Material online), which was also supported by an eBURST diagram (Turner et al. 2007). These results indicate that diverse *P. gingivalis* strains are present worldwide without geographical clustering and that, in contrast, phylogenetically close relationships are preserved in the same patient. *fimA* genotyping also showed similar results to MLST; however, most of the sequence types found in the allelic profiles from the eBURST diagram showed a unique profile despite being the same *fimA* type (supplementary fig. S5,

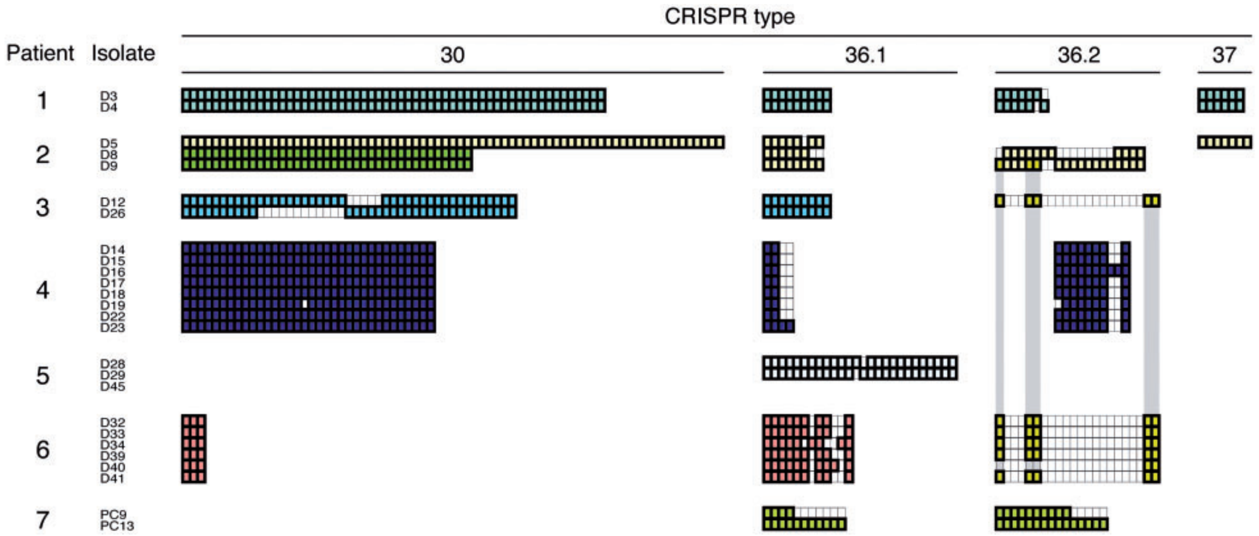

Fig. 2.—Spacer contents of *Porphyromonas gingivalis* isolates from seven patients in four CRISPR loci. Spacer arrays of 26 isolates from 7 patients are shown at each CRISPR locus. Each box indicates one spacer. The spacers in the arrays exhibit high nucleotide similarity to each other among the isolates if they are aligned vertically and have the same color. Blank boxes indicate absent spacers in the particular isolates. In patient no. 2, two colors are used because the D5 isolate has a type 30 spacer array that is distinct from those of D8 and D9. The spacers in type 36.2, shared among seven isolates of three patients, are indicated by deep yellow boxes and emphasized by dark gray belts.

Supplementary Material online). Therefore, we concluded that these conventional methods are insufficient for understanding bacterial diversification or evolutional traits.

We showed that the 60 isolates were also diverse with respect to CRISPR spacer content (supplementary fig. S9, Supplementary Material online). CRISPR typing with all loci showed similar results to the conventional methods, indicating that the former can serve as an alternative typing approach. However, there were some spacers shared by the isolates that did not cluster. Considering the suggestion that intercellular recombination impacts upon the diversification of *P. gingivalis* (described in the next section), we further analyzed the CRISPRs by separating them into four types (fig. 1; supplementary fig. S10, Supplementary Material online). As a result, clusters were observed with distinct characteristics for each type. It was shown that the isolates, which are not clustered by the conventional methods, can be clustered by CRISPR typing compared to *fimA* types.

It should be emphasized that we initially demonstrated that the contents of the spacers were different among almost all of the *P. gingivalis* isolates examined, even though they were from the same patient. We hypothesized that this was due to intercellular and/or intracellular recombination within the CRISPR loci. Bolotin et al. (2005) suggested that intercellular recombination was explained by the presence of the same spacer in CRISPRs between different isolates. It was further suggested that CRISPR typing may also be useful for high-resolution typing among *P. gingivalis* isolates from the same patient using slight differences in the number and content of the spacers, as with a previous report on *Sulfolobus islandicus* (Held et al. 2010). It should be possible to use CRISPR typing to trace the adaptation and/or transmission of *P. gingivalis* among patients across entire countries in a simple and cost-effective manner compared with the molecular tracing method reported for methicillin-resistant *Staphylococcus aureus* using high-throughput sequencing (McAdam et al. 2012).

The distribution in the number of *P. gingivalis* CRISPR spacers was similar to that in Bacteroidetes, with a few exceptions. Considering that the application of CRISPR typing has been limited to only a few species of bacteria or archaea (Andersson and Banfield 2008; Horvath et al. 2008; Held et al. 2010; Fabre et al. 2012; McGhee and Sundin 2012), this study should be considered as a test case for applying CRISPR typing in Bacteroidetes species; however, CRISPR loci could be affected by recombination, as well as MLST. It should also be taken into account that the CRISPR typing performed in this study is based on PCR, which may be imperfect due to either primer nonspecificity or impropriety of the PCR conditions. To characterize phylogenetic relationships more accurately among the isolates of *P. gingivalis* or Bacteroidetes, the combinational application of CRISPR typing and conventional methods is recommended. In addition, genome-level analyses will be needed to comprehend the information of the CRISPR loci, including undiscovered ones.

The Impact of Intercellular Recombination on Diversification

As a mechanism for diversification, intercellular recombination is likely to be important in *P. gingivalis* (Koehler et al. 2003;

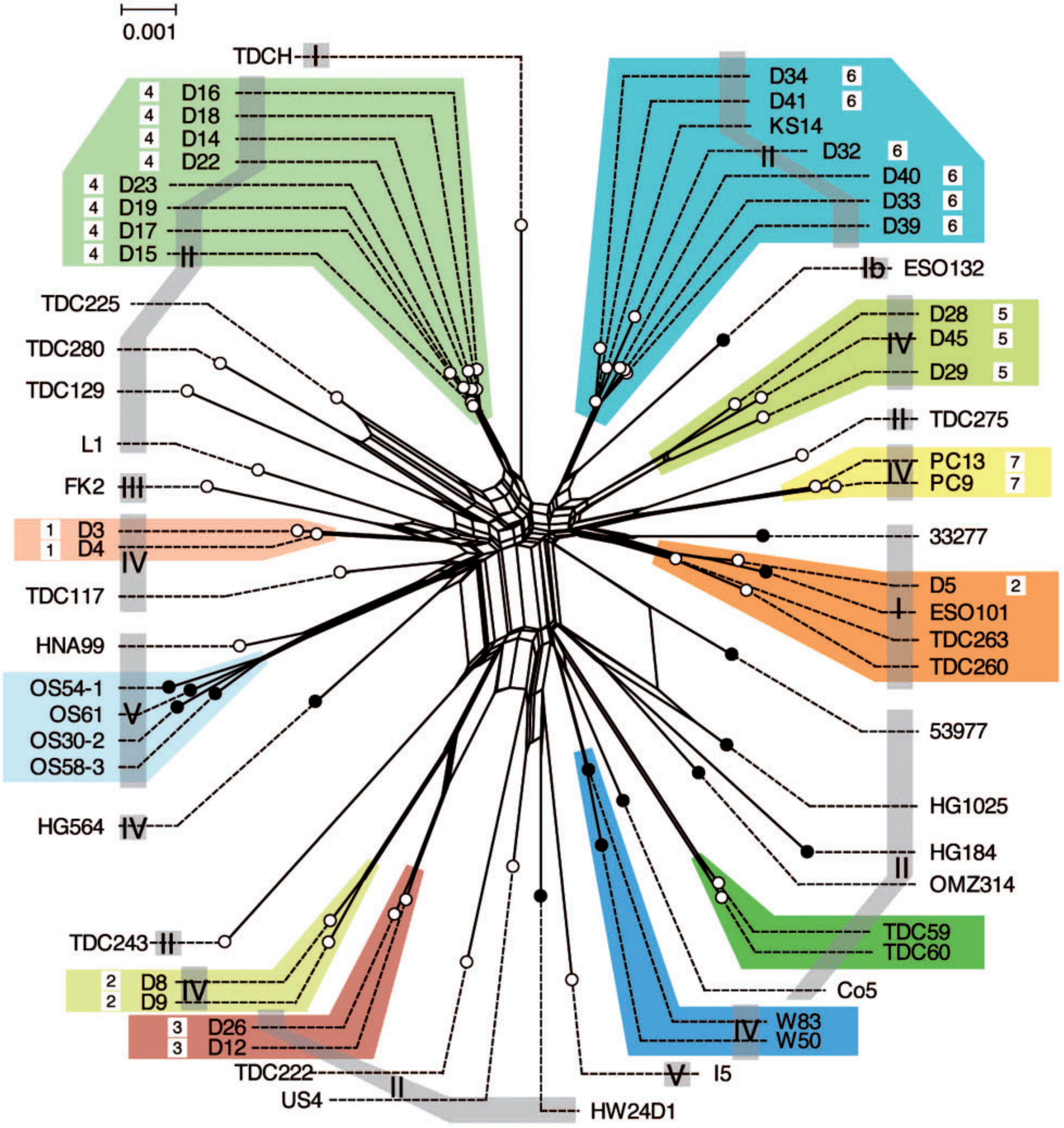
0.001
TDCH
I
4 D16
4 D18
4 D14
4 D22
4 D23
4 D19
4 D17
4 D15
II
TDC225
TDC280
TDC129
L1
FK2
III
1 D3
1 D4
IV
TDC117
HNA99
OS54-1
OS61
V
OS30-2
OS58-3
HG564
IV
TDC243
II
2 D8
2 D9
IV
3 D26
3 D12
TDC222
US4
II
HW24D1
D34 6
D41 6
KS14
D32 6
II
D40 6
D33 6
D39 6
Ib
ESO132
D28 5
D45 5
IV
D29 5
II
TDC275
PC13 7
PC9 7
IV
33277
D5 2
I
ESO101
TDC263
TDC260
53977
HG1025
HG184
II
OMZ314
TDC59
TDC60
Co5
W83
IV
W50
V
I5

Fig. 3.—Split network of 60 *Porphyromonas gingivalis* isolates obtained from concatenated seven loci sequences. A split network tree based upon the MLST data is shown. Circles indicate external nodes (each isolate) and are colored according to geographic origin (black: Japan; outlined: overseas or unspecified). *fimA* types are shown by light gray shadows. The numbers outside the isolate's name indicate the patient source. Eleven colors are used to emphasize the clusters.

Enersen et al. 2006). Split decomposition analysis showed network-like structures and the standardized index of association (I_A^S) was similar to previous reports (Koehler et al. 2003; Enersen et al. 2006). It was suggested that there is an impact of recombination events between the *P. gingivalis* cells on their diversification, even for relatively homogeneous Japanese isolates. In addition, the d*N*/d*S* value was less than 0.35 in all seven loci in MLST. Positive selection is indicated if d*N*/d*S* > 1 (Kryazhimskiy and Plotkin 2008). The value of less than 0.35 indicated the stability of the seven genes with few

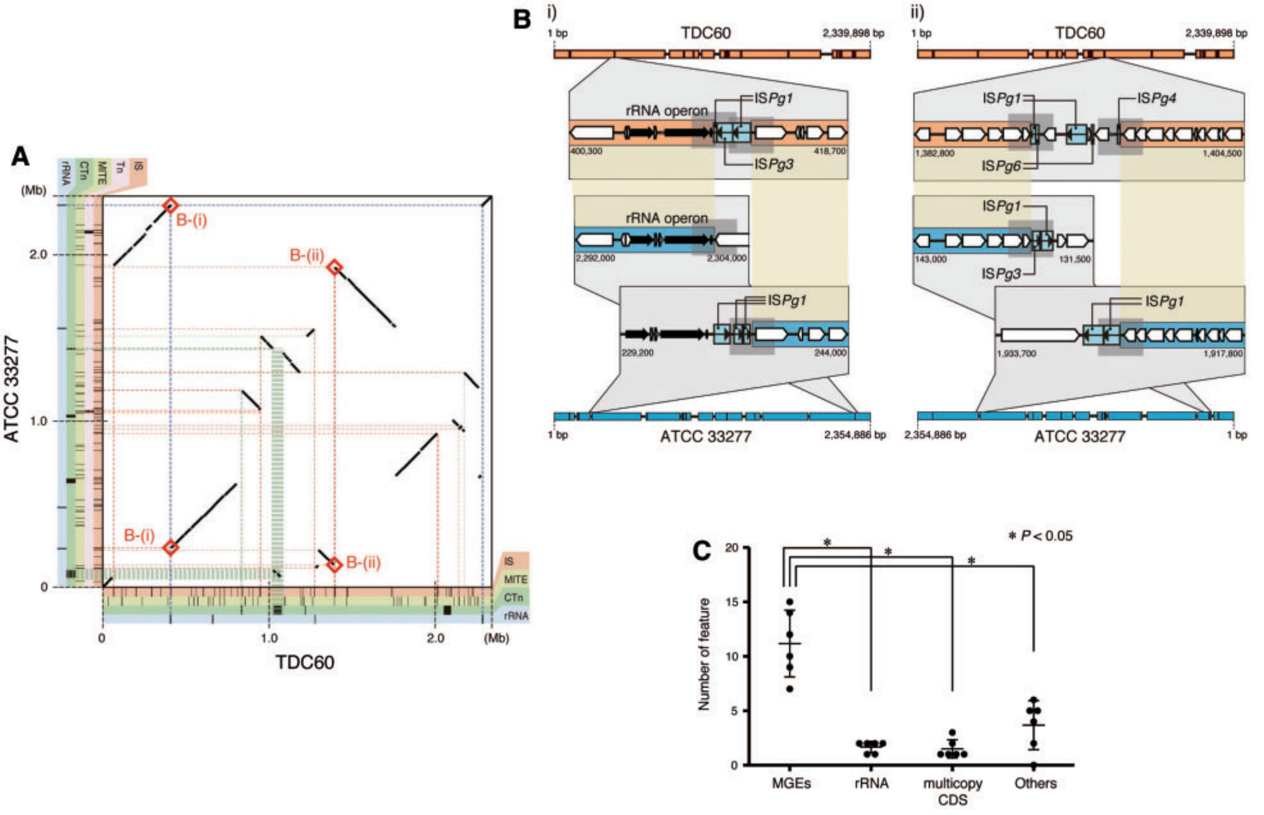

Fig. 4.—Characteristics of recombination breakpoints among three *Porphyromonas gingivalis* genomes. (*A*) Fragments are shown in the alignment of two genome sequences (TDC60, ATCC 33277). The positions of MGEs or rRNA operons in the breakpoint gaps are indicated by colored broken lines, connecting the gaps and the bars (indicating the positions of the features on the genome), which are arrayed along the outside of the plot area. The red boxes on the plot area are the regions shown in (*B*) in detail. (*B*) Breakpoint gaps of TDC60 are enlarged in light gray areas surrounded by broken lines. The regions of ATCC 33277, which correspond to the enlarged gap of TDC60, are enlarged similarly. The fragments in TDC60 and ATCC 33277 are colored by red and deep blue, respectively. The regions exhibiting high nucleotide similarity to each other are shown by a yellow belt between two fragments. The 3-kb regions of the breakpoints are indicated by dark gray rectangles on the upper or lower side of the fragments. (i) rRNA operons in the breakpoint gap. The black arrows indicate rRNA genes. The light blue-filled boxes with arrows inside indicate ISs. (ii) ISs in the breakpoint gap. (*C*) The number of each feature in the breakpoint gap is plotted. The regions without any characteristic features are included under "Others." The mean and standard deviations are provided by the horizontal and vertical lines, respectively. Statistical significance is indicated by an asterisk ($P < 0.05$, two-tailed paired *t* test).

nonsynonymous substitutions, suggesting that neutral evolution is a strong driving force in *P. gingivalis* genes relative to amino acid substitutions. Although whole genome information of multiple *P. gingivalis* strains is needed, it is suggested that amino acid substitution events in *P. gingivalis* genes are less important for intraspecies diversification, but intercellular recombination events are more likely. This is supported by two characteristics observed in CRISPR typing: 1) some isolates from different patients were clustered, and 2) some spacers were shared across the patients, suggesting intercellular recombination events involving the CRISPR loci. As for the mechanisms of DNA introduction into *P. gingivalis* cells, transfer with CTns (Naito et al. 2011) and transformation by natural competence (Tribble et al. 2012) have been reported, both of which could be followed by intercellular recombination events.

MGE Involvement in Intracellular Genome Rearrangement

As well as intercellular recombination, intracellular genome rearrangements are considered to be important for bacterial diversification (Dybvig 1993; Ng et al. 1999; Lysnyansky et al. 2001; Nakagawa et al. 2003; Nozawa et al. 2011). In this study, we demonstrated complicated genome rearrangements in *P. gingivalis* (fig. 4*A*; supplementary fig. S11, Supplementary Material online) as well as in a previous report (Naito et al. 2008); these findings were supported by the results of PFGE, in which rearrangements altered the localization of the sequences recognized by restriction enzymes. In addition, we made the following relevant observations: 1) MGEs were significantly located at the breakpoints and 2) ISs are statistically predominant at the breakpoint gaps compared with other

Table 2
Breakpoint Gaps and Features in Dot Plot TDC60-ATCC 33277

TDC60			ATCC 33277		
Start	End	Feature[a]	Start	End	Feature[a]
56810	66905	IS*Pg1*	57974	63764	CTnPg1-a
409938	412415	rRNA operon	103051	116919	CTnPg1-a
804055	838033	CTnPg1_1	130514	135389	IS*Pg6*
949277	951906	IS*Pg1*, IS*Pg2*	224185	237668	IS*Pg1*
1022572	1026904	CTnPg1_2, CTnPg2	622790	659911	IS*Pg1*
1069135	1085180	CTnPg1_2, CTnPg2	671249	673067	Multicopy CDS (A)[b]
1185081	1228234	PGTDC60_1140-1144; PGTDC60_1187-1191	925785	938797	IS*Pg1*
1270652	1281602	IS*Pg2*, IS*Pg3*	950459	952554	IS*Pg3*
1295201	1298355	PGTDC60_1252-1258	971231	973090	IS*Pg3*
1390501	1396464	IS*Pg1*, IS*Pg6*	1007011	1068370	Tn*Pg17*-a
1761003	1762678	Multicopy CDS (A)[b]	1182895	1201009	IS*Pg3*, IS*Pg5*
2011622	2105110	IS*Pg1*, IS*Pg2*	1290640	1293358	IS*Pg1*, IS*Pg2*
2138949	2141781	IS*Pg1*	1409859	1439355	CTnPg1-b
2160664	2163995	PGTDC60_2080-2084	1510525	1510757	MITE239
2175656	2175793	MITE*PgRS*	1553049	1567234	IS*Pg1*
2261211	2262115	PGTDC60_2191-2194	1925826	1936662	IS*Pg1*, IS*Pg3*
2275269	2279712	rRNA operon	2294997	2301097	rRNA operon

[a] Features were identified in 3-kb regions covering 1.5-kb upstream and 1.5-kb downstream of the breakpoint.
[b] Multicopy CDS (A) is shown in supplementary table S8, Supplementary Material online.

MGEs. Therefore, it was suggested that *P. gingivalis* exhibits complex genome rearrangements following frequent IS transposition, leading to intraspecies diversification.

Predominance of CRISPR Spacers Targeting *P. gingivalis* Sequences and Their Hypothesized Functions

In the nucleotide similarity searches, we found that there were few *P. gingivalis* spacers exhibiting high nucleotide similarity with known sequences. Of these, those exhibiting high nucleotide similarity to *P. gingivalis* sequences were predominant. As far as we can determine, the proportion of the number of the CRISPR spacers exhibiting high nucleotide similarity to the genome of the same species was highest in *P. gingivalis* among prokaryotes. For example, the number of such spacers was almost twice as frequent in the present study (1.8%: 38/2,150, original spacers) compared with the analysis of *S. islandicus*, which identified many such spacers (0.84%: 78/9,219, Brodt et al. 2011). The percentage in our study was also much higher than that reported by Stern et al. (2010), in which 100 out of 23,550 spacers (0.4%) exhibited high nucleotide similarity to sequences on 330 genomes. Therefore, we propose that *P. gingivalis* might be a useful model to unravel the biological significance of CRISPR spacers exhibiting high nucleotide similarity to the genome of the same species. It is hypothesized that DNA from other *P. gingivalis* cells are targeted by the CRISPRs to prevent their introduction; the CRISPRs may only target the DNA from other cells and may not confer lethality on the recipient cells. The invading DNA might be supplied mainly by CTns because the major difference of the gene content in the *P. gingivalis* genome was derived from MGEs, as confirmed by our and other studies (Naito et al. 2011). Therefore, future studies should clarify the hypothesis that *P. gingivalis* selectively acquires useful foreign DNA sequences for its survival and evolution by CRISPR function.

Moreover, it was remarkable that 7 spacers (7/19) exhibited high nucleotide similarity to regions related to the ISs in the *P. gingivalis* genome. Some CRISPRs reportedly confer resistance to foreign RNA as well as to DNA (Sorek et al. 2008; Makarova et al. 2011); in this study, three *P. gingivalis* genomes were shown to harbor Cas genes for both DNA and RNA targeting. Thus, such spacers might be suggested to regulate IS transposition by targeting the mRNA of transposases, leading to the regulation of genome rearrangements. In this hypothesis, it is suggested that transcribed RNAs are targeted when the CRISPRs inhibit gene expression (Bhaya et al. 2011), which is not lethal for cell survival. Inhibition of IS transposition might be the case; transcripts of the transposase genes might be targeted by the CRISPRs. Such regulation of gene expression was reported in *P. carbinolicus* (Aklujkar and Lovley 2010) and *A. actinomycetemcomitans* (Jorth and Whiteley 2012). Therefore, this property suggests the appropriateness of *P. gingivalis* for examining the novel functions of CRISPRs. However, the mechanisms by which *P. gingivalis* CRISPRs recognize target sequences are still unknown due to the undetectability of PAMs; enrichment of oral virome sequences will provide more sequences exhibiting high nucleotide similarity to the *P. gingivalis* CRISPRs, leading to the detection of PAMs.

Table 3
Sequences in the Databases Exhibiting High Nucleotide Similarity to Unique 1,187 Spacers of 57 *Porphyromonas gingivalis* Isolates

Spacer Name[a]	Isolates Which Have the Spacer								Database	Organism	Species	Gene
30_107	W50	W83							GenBank nucleotide	Bacteria	*P. gingivalis*	Saccharopine dehydrogenase
30_108	W50	W83							GenBank nucleotide	Bacteria	*P. gingivalis*	Saccharopine dehydrogenase
30_129	W50	W83							GenBank nucleotide	Bacteria	*P. gingivalis*	Outer membrane efflux protein
30_139	D3	D4							Pride's salivary virome	Viruses	*Lactococcus* phage	Putative ABC transporter
30_280	D8	D9							GenBank nucleotide	Bacteria	*Clostridium* sp.	DNA-methyltransferase
30_287	D8	D9							GenBank nucleotide	Bacteria	*P. gingivalis*	Transposase
30_340	D14	D15	D16	D17	D18	D19	D22	D23	GenBank nucleotide	Bacteria	*P. gingivalis*	Outer membrane efflux protein
30_368	KS14								GenBank nucleotide	Bacteria	*P. gingivalis*	Putative carbonic anhydrase
30_372	KS14								GenBank nucleotide	Bacteria	*P. gingivalis*	Transposase
30_382	L1								MetaHIT	Viruses	*Bacteroides* phage	Hypothetical protein
30_383	L1								MetaHIT	Viruses	*Bacillus* phage	gp19.5
30_440	US4								GenBank nucleotide	Bacteria	*P. gingivalis*	DNA modification methylase-like
30_498	TDC59	TDC60							GenBank nucleotide	Bacteria	*Haemophilus parasuis*	Hypothetical protein
30_576	TDC60								GenBank nucleotide	Bacteria	*P. gingivalis*	ATP-binding protein
30_587	TDC2075								GenBank nucleotide	Bacteria	*P. gingivalis*	Leucyl-tRNA synthetase
30_701	TDC280								GenBank nucleotide	Bacteria	*P. gingivalis*	Carboxyl-terminal protease
36.1_51	D28	D29							GenBank nucleotide	Bacteria	*P. gingivalis*	Hypothetical protein
36.1_67	D28	D29	HW24D1	ESO132					HMBSA	Archaea	*Thermococcus barophilus*	TldD
36.1_91	I5								GenBank nucleotide	Bacteria	*P. gingivalis*	Hypothetical protein
36.1_104	TDC59	TDC60							GenBank nucleotide	Bacteria	*P. gingivalis*	Acetyl-CoA synthetase
36.1_152	HNA99								MetaHIT	Archaea	*Methanothermus fervidus*	Methyltransferase fkbm family
36.1_156	HNA99	OS54-1	OS61						GenBank nucleotide	Bacteria	*P. gingivalis*	Hypothetical protein
36.1_169	OS54-1	OS61							GenBank nucleotide	Bacteria	*P. gingivalis*	Hypothetical protein
36.2_36	D14	D15	D16	D17	D18	D19	D22	D23	HMBSA	Bacteria	*Bacteroides* sp.	Type I restriction-modification system specificity determinant
37_11	W50	W83							GenBank nucleotide	Bacteria	*P. gingivalis*	Unspecified
37_61	TDC59	TDC60							GenBank nucleotide	Bacteria	*P. gingivalis*	Putative transporter
37_85	TDC243								HMBSA	Unspecified	Unspecified	Unspecified
37_89	TDC260	ESO101							GenBank nucleotide	Bacteria	*P. gingivalis*	Hypothetical protein
37_90	TDC260	ESO101							HMBSA	Unspecified	Unspecified	Unspecified

[a] The name of each spacer is given as described in Materials and Methods.

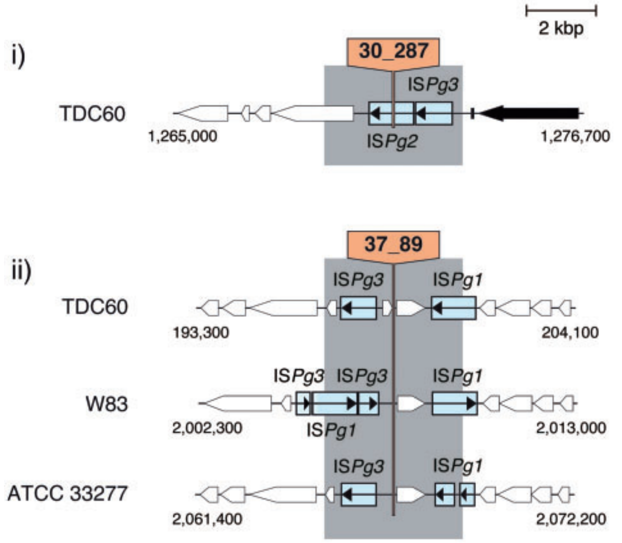

Fig. 5.—Regions exhibiting high nucleotide similarity to *P. gingivalis* CRISPR spacers. Two examples of the 19 spacers exhibiting high nucleotide similarity to the *P. gingivalis* genome are shown. The white and black arrows indicate CDSs and rRNA genes, respectively. The arrows within the light blue-filled boxes indicate ISs. The orange regions indicate the sequences exhibiting high nucleotide similarity to CRISPR spacers. (i) Region exhibiting high nucleotide similarity to spacer 37_259: the transposase gene in IS*Pg2*, in the TDC60 genome. (ii) Region exhibiting high nucleotide similarity to spacer 37_90: close to the IS both 2-kb upstream and 2-kb downstream in the 3 genomes.

In addition to the spacers targeting *P. gingivalis* genomes, we also identified those targeting either viral or exogenous bacterial sequences. These bacteria colonize different niches; however, they have a common feature of being obligative anaerobic microbes, except for *Haemophilus parasuis* (Marteinsson et al. 1999; Wexler 2007; Bruggemann and Gottschalk 2009; Anderson et al. 2010; Xu et al. 2011). It is suggested that these spacers prevent *P. gingivalis* from allowing the introduction of foreign DNA such as those of anaerobic bacteria or viruses.

In contrast, there were numerous spacers of *P. gingivalis* CRISPRs without significant nucleotide similarity to the available sequences in the databases; this result was almost similar when performing nucleotide similarity searches using the spacer data sets of bacteria and archaea in the CRISPI database. It is possible that the extremely low proportion of these spacers, despite using salivary virome databases, is currently due to the lack of comprehensive oral virome databases. Another possible reason is that, in *P. gingivalis*, the major source of the CRISPR spacers is not viral sequences, but sequences from a relatively rare genome in a periodontitis lesion that has not yet been characterized.

In conclusion, we showed the effectiveness of CRISPR typing for *P. gingivalis* by cluster analysis and high-resolution typing in the same patient, as well as its potential applicability to the Bacteroidetes group. We also demonstrated that *P. gingivalis* is a bacterium with a survival strategy for creating intraspecies diversity by both intercellular recombination and intracellular genome rearrangements, in which ISs are involved. Moreover, it is also suggested that these events might be regulated by CRISPRs, which limit both IS transposition and the introduction of DNA from other *P. gingivalis* cells. However, such a function of CRISPRs may not be their primary role and it needs to be proved experimentally in future studies. The determination of draft genome sequences from multiple isolates will provide information on the position of CRISPRs and ISs in each genome, which could lead to the elucidation of the relationship between IS transposition and CRISPR inhibition. Considering that *P. gingivalis* is not a major member of the healthy oral cavity, but becomes predominant in periodontitis (Griffen et al. 2012), characterization of such rare microbiomes and sequencing multiple *P. gingivalis* isolates may be important in elucidating the mechanisms of CRISPR function in *P. gingivalis* and to understand the basic biology of *P. gingivalis* itself. The sequencing of multiple isolates will also yield additional CRISPR information, which may identify CRISPR loci that were not detected in the three *P. gingivalis* genomes examined in this study. In addition, expression analysis of multiple *P. gingivalis* isolates by RNA-seq will provide clues for elucidating these hypothesized functions.

Supplementary Material

Supplementary figures S1–S13 and tables S1–S10 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank the following individuals for providing the *P. gingivalis* isolates: Drs K. Okuda and K. Ishihara (Tokyo Dental College) and M. Umeda and Y. Izumi (Tokyo Medical and Dental University). This work was supported by a grant-in-aid for Japan Society for the Promotion of Science (JSPS) Fellows; the Japanese Ministry of Education, Global Center of Excellence (GCOE) Program, "International Research Center for Molecular Science in Tooth and Bone Diseases," Ministry of Education, Culture, Sports, Science and Technology, Japan; Grant-in-Aid for Scientific Research (C) 22592032; Grant-in-Aid for Scientific Research on Innovative Areas (24117508); and Funding Program for Next Generation World-Leading Researchers (LS041), JSPS.

Literature Cited

Achaz G, Coissac E, Netter P, Rocha EP. 2003. Associations between inverted repeats and the structural evolution of bacterial genomes. Genetics 164:1279–1289.

Aklujkar M, Lovley DR. 2010. Interference with histidyl-tRNA synthetase by a CRISPR spacer sequence as a factor in the evolution of *Pelobacter carbinolicus*. BMC Evol Biol. 10:230.

Amano A. 2003. Molecular interaction of *Porphyromonas gingivalis* with host cells: implication for the microbial pathogenesis of periodontal disease. J Periodontol. 74:90–96.

Amano A, Nakagawa I, Kataoka K, Morisaki I, Hamada S. 1999. Distribution of *Porphyromonas gingivalis* strains with *fimA* genotypes in periodontitis patients. J Clin Microbiol. 37:1426–1430.

Amano A, Nakagawa I, Okahashi N, Hamada N. 2004. Variations of *Porphyromonas gingivalis* fimbriae in relation to microbial pathogenesis. J Periodontal Res. 39:136–142.

Andersson AF, Banfield JF. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science 320: 1047–1050.

Anderson I, et al. 2010. Complete genome sequence of *Methanothermus fervidus* type strain (V24S). Stand Genomic Sci. 3:315–324.

Arumugam M, et al. 2011. Enterotypes of the human gut microbiome. Nature 473:174–180.

Auchtung JM, Lee CA, Monson RE, Lehman AP, Grossman AD. 2005. Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. Proc Natl Acad Sci U S A. 102:12554–12559.

Barrangou R, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712.

Bhaya D, Davison M, Barrangou R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet. 45:273–297.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151:2551–2561.

Bostanci N, Belibasakis GN. 2012. *Porphyromonas gingivalis*: an invasive and evasive opportunistic oral pathogen. FEMS Microbiol Lett. 333:1–9.

Brodt A, Lurie-Weinberger MN, Gophna U. 2011. CRISPR loci reveal networks of gene exchange in archaea. Biol Direct. 6:65.

Bruggemann H, Gottschalk G, editors. 2009. Clostridia: molecular biology in the post-genomic era. Hethersett, Norwich (UK): Caister Academic Press.

Cady KC, O'Toole GA. 2011. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. J Bacteriol. 193:3433–3445.

Chen T, et al. 2010. The human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database (Oxford) 2010:baq013.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188–1190.

Darling AE, Miklós I, Ragan MA. 2008. Dynamics of genome rearrangement in bacterial populations. PLoS Genet. 4:e1000128.

Dice LR. 1945. Measures of the amount of ecologic association between species. J Ecol. 26:297–302.

Dybvig K. 1993. DNA rearrangements and phenotypic switching in prokaryotes. Mol Microbiol. 10:465–471.

Enersen M, Olsen I, Kvalheim Ø, Caugant DA. 2008. *fimA* genotypes and multilocus sequence types of *Porphyromonas gingivalis* from patients with periodontitis. J Clin Microbiol. 46:31–42.

Enersen M, Olsen I, van Winkelhoff AJ, Caugant DA. 2006. Multilocus sequence typing of *Porphyromonas gingivalis* strains from different geographic origins. J Clin Microbiol. 44:35–41.

Fabre L, et al. 2012. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. PLoS One 7:e36995.

Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 186:1518–1530.

Griffen AL, et al. 2012. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. ISME J. 6: 1176–1185.

Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 35:W52–W57.

Held NL, Herrera A, Cadillo-Quiroz H, Whitaker RJ. 2010. CRISPR associated diversity within a population of *Sulfolobus islandicus*. PLoS One 5: e12988.

Hill CW, Harnish BW. 1981. Inversions between ribosomal RNA genes of *Escherichia coli*. Proc Natl Acad Sci U S A. 78:7069–7072.

Horvath P, et al. 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol. 190:1401–1412.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23:254–267.

Jolley KA, Feil EJ, Chan MS, Maiden MC. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230–1231.

Jorth P, Whiteley M. 2012. An evolutionary link between natural transformation and CRISPR adaptive immunity. MBio 3:e00309–e00312.

Koehler A, et al. 2003. Multilocus sequence analysis of *Porphyromonas gingivalis* indicates frequent recombination. Microbiology 149: 2407–2415.

Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. PLoS Genet. 4:e1000304.

Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

Kusumoto M, et al. 2011. Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. Nat Commun. 2:152.

Lamont RJ, Jenkinson HF. 1998. Life below the gum line: pathogenic mechanisms of *Porphyromonas gingivalis*. Microbiol Mol Biol Rev. 62:1244–1263.

Leplae R, Lima-Mendez G, Toussaint A. 2010. ACLAME: A CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Res. 38: D57–D61.

Lewis CM Jr, Obregón-Tito A, Tito RY, Foster MW, Spicer PG. 2012. The human microbiome project: lessons from human genomics. Trends Microbiol. 20:1–4.

Loos BG, et al. 1992. A statistical approach to the ecology of *Porphyromonas gingivalis*. J Dent Res. 71:353–358.

Lysnyansky I, Ron Y, Sachse K, Yogev D. 2001. Intrachromosomal recombination within the vsp locus of *Mycoplasma bovis* generates a chimeric variable surface lipoprotein antigen. Infect Immun. 69: 3703–3712.

Makarova KS, et al. 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 9:467–477.

Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322: 1843–1845.

Marteinsson VT, et al. 1999. *Thermococcus barophilus* sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. Int J Syst Bacteriol. 49:351–359.

McAdam PR, et al. 2012. Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. Proc Natl Acad Sci U S A. 109:9107–9112.

McGhee GC, Sundin GW. 2012. *Erwinia amylovora* CRISPR elements provide new tools for evaluating strain diversity and for microbial source tracking. PLoS One 7:e41706.

MetaHIT Consortium. 2010. MetaHIT Draft Bacterial Genomes at the Sanger Institute. Available from: (http://www.sanger.ac.uk/resources/downloads/bacteria/metahit/, last accessed May 24, 2013).

Milkman R. 1997. Recombination and population structure in *Escherichia coli*. Genetics 146:745–750.

Naito M, et al. 2008. Determination of the genome sequence of *Porphyromonas gingivalis* strain ATCC 33277 and genomic comparison with strain W83 revealed extensive genome rearrangements in *P. gingivalis*. DNA Res. 15:215–225.

Naito M, et al. 2011. Characterization of the *Porphyromonas gingivalis* conjugative transposon CTnPg1: determination of the integration site and the genes essential for conjugal transfer. Microbiology 157: 2022–2032.

Nakagawa I, et al. 2000. Distribution and molecular characterization of *Porphyromonas gingivalis* carrying a new type of *fimA* gene. J Clin Microbiol. 38:1909–1914.

Nakagawa I, et al. 2003. Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. Genome Res. 13:1042–1055.

Nakayama K. 1995. Determination of the genome size of the oral anaerobic bacterium *Porphyromonas gingivalis* by pulsed field gel electrophoresis. Dent Jpn (Tokyo). 32:25–28.

Nelson KE, et al. 2003. Complete genome sequence of the oral pathogenic bacterium *Porphyromonas gingivalis* strain W83. J Bacteriol. 185: 5591–5601.

Ng I, Liu SL, Sanderson KE. 1999. Role of genomic rearrangements in producing new ribotypes of *Salmonella typhi*. J Bacteriol. 181: 3536–3541.

Nozawa T, et al. 2011. CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. PLoS One 6:e19543.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304.

Octavia S, Lan R. 2006. Frequent recombination and low level of clonality within *Salmonella enterica* subspecies I. Microbiology 152:1099–1108.

Ohtsubo Y, Ikeda-Ohtsubo W, Nagata Y, Tsuda M. 2008. GenomeMatcher: a graphical user interface for DNA sequence comparison. BMC Bioinformatics 9:376.

Ooka T, et al. 2009. Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. Genome Res. 19:1809–1816.

Perez-Chaparro PJ, Rouillon A, Minet J, Lafaurie GI, Bonnaure-Mallet M. 2009. *fimA* genotypes and PFGE profile patterns in *Porphyromonas gingivalis* isolates from subjects with periodontitis. Oral Microbiol Immunol. 24:423–426.

Pride DT, et al. 2012. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. ISME J. 6:915–926.

Pride DT, Salzman J, Relman DA. 2012. Comparisons of clustered regularly interspaced short palindromic repeats and viromes in human saliva reveal bacterial adaptations to salivary viruses. Environ Microbiol. 14: 2564–2576.

Rousseau C, Gonnet M, Le Romancer M, Nicolas J. 2009. CRISPI: a CRISPR interactive database. Bioinformatics 25:3317–3318.

Siguier P, Filée J, Chandler M. 2006. Insertion sequences in prokaryotic genomes. Curr Opin Microbiol. 9:526–531.

Sorek R, Kunin V, Hugenholtz P. 2008. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol. 6:181–186.

Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet. 26: 335–340.

Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

Tribble GD, et al. 2012. Natural competence is a major mechanism for horizontal DNA transfer in the oral pathogen *Porphyromonas gingivalis*. MBio 3:e00231–e00211.

Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG. 2007. Assessing the reliability of eBURST using simulated populations with known ancestry. BMC Microbiol. 7:30.

Watanabe T, et al. 2011. Complete genome sequence of the bacterium *Porphyromonas gingivalis* TDC60, which causes periodontal disease. J Bacteriol. 193:4259–4260.

Wexler HM. 2007. *Bacteroides*: the good, the bad, and the nitty-gritty. Clin Microbiol Rev. 20:593–621.

Xu Z, et al. 2011. Genomic characterization of *Haemophilus parasuis* SH0165, a highly virulent strain of serovar 5 prevalent in China. PLoS One 6:e19631.

Associate editor: Hidemi Watanabe