Adaptation by Ancient Horizontal Acquisition of Butyrate Metabolism Genes in *Aggregatibacter actinomycetemcomitans*

Ahmed M. Moustafa, a Senthil Kumar Velusamy, b Lidiya Denu, a Apurva Narechania, c Daniel H. Fine, b Paul J. Planet,a,c,d

a Division of Pediatric Infectious Diseases, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
b Department of Oral Biology, Rutgers School of Dental Medicine, Newark, New Jersey, USA
c Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York, USA
d Department of Pediatrics, Perelman College of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

ABSTRACT  Like the bacterial residents of the human gut, it is likely that many of the species in the human oral microbiota have evolved to better occupy and persist in their niche. *Aggregatibacter actinomycetemcomitans* (*Aa*) is both a common colonizer of the oral cavity and has been implicated in the pathogenesis of periodontal disease. Here, we present a whole-genome phylogenetic analysis of *Aa* isolates from humans and nonhuman primates that revealed an ancient origin for this species and a long history of association with the *Catarrhini*, the lineage that includes Old World monkeys (OWM) and humans. Further genomic analysis showed a strong association with the presence of a short-chain fatty acid (SCFA) catabolism locus (*atoRDAEB*) in many human isolates that was absent in almost all nonhuman OWM isolates. We show that this locus was likely acquired through horizontal gene transfer. When grown under conditions that are similar to those at the subgingival site of periodontitis (anaerobic, SCFA replete), *Aa* strains with *atoRDAEB* formed robust biofilms and showed upregulation of genes involved in virulence, colonization, and immune evasion. Both an isogenic deletion mutant and nonhuman primate isolates lacking the *ato* locus failed to grow in a robust biofilm under these conditions, but grew well under the carbohydrate-rich conditions similar to those found above the gumline. We propose that the acquisition of the *ato* locus was a key evolutionary step allowing *Aa* to utilize SCFAs, adapt, and modulate subgingival disease.

IMPORTANCE  There has been considerable interest in the impact of short-chain fatty acids (SCFAs) on inflammatory effects related to the microbiome. Here, we present evidence that SCFAs may also be important in disease by providing an energy source or disease-associated cue for colonizing pathogens. We propose that SCFAs allow *Aggregatibacter actinomycetemcomitans* (*Aa*) to adapt to the subgingival anaerobic environment, which is the site of human periodontitis. Under anaerobic, SCFA-rich conditions, human-derived *Aa* strains that possess butyrate metabolism genes form strong biofilms and upregulate virulence genes. Our phylogenetic analysis highlights a long history of evolution of *Aa* with its primate hosts and suggests that the acquisition of butyrate metabolism genes may have been a critical step in allowing *Aa* to colonize a new niche and cause disease in humans. Overall, this study highlights the important role that horizontal gene transfer may play in microbial adaptation and the evolution of infectious disease.

KEYWORDS  *Aggregatibacter*, horizontal gene transfer, nutritional immunity, Old World monkey, periodontitis, short-chain fatty acid

The resident microbiota of mammalian species appears to diversify or “cospeciate” in parallel with the divergence of its hosts (1–5), which results in long-term associations of certain microbial lineages with specific host species. While the biological...
consequences and driving forces underlying these cophylogenetic patterns remain unclear (3), it is likely that such long-term relationships lead to specific bacterial adaptations to each host species. There is significant evidence supporting bacterial adaptation to the host in the gut (6), but there is little known about the microbiota at other body sites. It is also unclear how adaptation may contribute to health and disease.

Aggregatibacter actinomycetemcomitans (Aa) is a Gram-negative oral bacterium that colonizes tooth surfaces in both humans (7) and nonhuman primates (8). However, in humans Aa is also strongly associated with localized aggressive periodontitis (LAP) (9), and it is also a known pathogen in endocarditis, brain abscesses, and pneumonia (10). Despite widespread colonization in nonhuman primates, there is no evidence that Aa is associated with disease in these animals (11, 12), and highly leukotoxic strains that are strongly associated with disease in humans (13) have not been recovered from primates to date. Together, these observations suggest either that humans are more susceptible to Aa virulence or that Aa may have evolved to occupy distinct niches in humans and cause disease (14), scenarios that are not mutually exclusive.

Aa is thought to be transmitted horizontally by maternal transmission to infants (15–17) and has been found at high rates in young children (e.g., 63% of 6-month-olds [18] and 95.5% of 12-month-olds [19]), with lower rates of colonization in school children (e.g., 13.7% of 11- to 17-year-olds [9]). Moreover, 36% of healthy adults and 50% of those with periodontitis have been reported to carry Aa (20). The estimated prevalences of Aa in periodontitis are proposed to be 48.1% in patients younger than 35 years and 24.6% in patients older than 35 years (21).

In nonhuman primates, such as rhesus macaques, chimpanzees, bonobos, gorillas, and orangutans, both in captivity and sanctuaries, rates of Aa oral colonization vary between 50 and 100% (8, 22), showing that there is a strong affinity for the Catarrhini (the group that includes Old World monkeys [OWM], great apes, gibbons, and humans) and contrasting with surveys of New World monkeys, the Platyrrhini (8). In addition, several Aa adhesins (23, 24) and leukotoxin (LtxA) (25, 26) appear to have specificity for cells from members of the Catarrhini. It is notable too that while LtxA is specific for members of the Catarrhini, human neutrophils appear to be particularly susceptible to LtxA, even compared to those from close great ape relatives (26).

To understand the diversity of Aa strains and evaluate the phylogenetic relationship between isolates from human and nonhuman primates, we produced draft whole-genome sequences (WGSs) for 14 Aa isolates (9 from rhesus [Rh] macaques, 2 from green monkeys, 2 from humans, and 1 from a marmoset [see Table S1 in the supplemental material]) and combined these with 81 available Aa WGSs from GenBank for comparative analysis.

Whole-genome phylogenetic analysis of the combined data set revealed two major Aa clades. Clade I (Fig. 1A) is primarily derived from humans and is divided into subclades that roughly correlate with known serotypes, corroborating prior phylogenetic analyses based on 397 and 1,146 concatenated core genes (27, 28).

Interestingly, clade II expands what is known about a previously described, highly divergent Aa lineage that includes human-derived, serotype ε isolates from Finland and a serotype b Aa isolate from an Rh macaque that was previously sequenced by our group (27–30). All but one of the Rh macaque isolates sequenced for this study segregated in this divergent clade II, revealing a strong association of this clade with nonhuman primates.

The added diversity and potentially deep divergence of these strains raised the possibility of a codivergence within the Aa species from the most recent common ancestor (MRCA) of the Catarrhini approximately 32 million years ago (mya) (31). To test this possibility, we used a whole-genome sequencing (WGS), Bayesian, molecular clock-based approach. There is little information available about the rates of sequence evolution for Aa, and in general, estimated rates of bacterial evolution vary greatly (see Text S1 in the supplemental material). Consequently, we decided
Evolution of A. actinomycetemcomitans and the atoRDAEB locus in the Pasteurellaceae. (A) A chronogram was constructed using the genome single-nucleotide polymorphisms from 95 Aa genomes in a Bayesian phylogenetic analysis that coestimated the phylogenetic
to take an agnostic, hypothesis-testing approach to infer rates at different putative historical time points for the MRCA of Aa. Figure 1B shows a range of rates and the estimated age of the common Aa ancestor for each one. With faster evolutionary rates (3.4 \times 10^{-6} substitutions/site/year), the divergence between clades I and II was at least 11,000 years old. A reasonably slower evolutionary rate of 1.19 \times 10^{-9} showed the divergence time to be close to the MRCA of the Catarrhini clade, 32 mya (31), a point that may signal an Aa-Catarrhini “cospeciation” event. However, more sampling from the Catarrhini clade would be needed to further support this hypothesis.

To determine differences in genomic content between human and nonhuman strains, predicted protein-coding open reading frames (ORFs) of the assembled genomes were compared among strains using a pangenome approach. There were no genes that were uniquely found in isolates from any particular host species, reflecting the presence of some human and nonhuman isolates across the tree. However, there were a small number of unique and ubiquitous genes defining the two major clades (see Fig. S1 in the supplemental material). Interestingly, all 5 unique genes in clade I mapped to the same locus, the ato locus (32) (atoRDAEB), which is predicted to be involved in short-chain fatty acid (SCFA) catabolism. The absence of this locus in clade II suggested the possibility that it was acquired by horizontal gene transfer (HGT) in the MRCA of clade I, an event that took place at least 12.4 mya if we assume an Aa-Catarrhini codivergence (Fig. 1A). To test this hypothesis, we surveyed all publicly available genomes for this locus from the taxonomic family of Aa, the Pasteurellaceae, and mapped its presence on a WGS maximum likelihood tree (Fig. 1C). Interestingly, the ato locus was absent in all other Aggregatibacter species and even closely related Haemophilus species, supporting HGT into Aa clade I. Notably, the locus was present in 62% of Haemophilus influenzae (Hflu) genomes, and 100% of encapsulated Hflu strains (see Fig. S2 in the supplemental material), raising the possibility of HGT between Hflu and Aa. However, a maximum likelihood tree of the ato locus sequences themselves (Fig. 1D) showed two distinct Aa and Hflu clades without one clade nested in the other, which argues against direct HGT between these two species. Given the close relationship between the Hflu and Aa versions of the locus, it is possible that both species acquired the locus from a common unidentified third organism. Because all of the closest BLAST hits to the Aa ato locus were from the Pasteurellaceae, and given that the closest relatives from Yersinia intermedia and Morganella morganii fall outside the Pasteurellaceae loci, we contend that the most likely donor organism is an unsampled or extinct Pasteurellaceae family member (see Fig. S3 in the supplemental material).

Based on work in Escherichia coli (32–35), we hypothesized that Aa with the ato locus would be able to catabolize the SCFA butyrate and use it as a carbon source or

FIG 1 Legend (Continued)
relationships among isolates and the time since the divergence. Pictured is an analysis using a substitution rate of 1.19 \times 10^{-9} substitutions/site/year that estimates a common ancestor at 31.77 million years ago (mya), with a 95% highest posterior density (HPD) of 29.8 to 33.8 mya. Taxa from humans, macaques, green monkeys, and a marmoset are in black, magenta, blue, and teal, respectively. The 95% HPD is shown beside important nodes. The asterisks on two branches show the interhost transmission (IT) between clades I and II. The black diamond on the tree was constructed using OrthologID. The root was determined using a similar tree that included a close relative (Erwinia teleogrylli) from the Gammaproteobacteria as an outgroup. Bootstrap values are 95 to 100 on all branches, except for 3 nodes with gray circles, which have values between 58 and 76. Species clades with more than two genomes were collapsed for easier visualization. The sequences were clustered using CD-HIT (100% identity and coverage). All clusters were composed of individual species. The number of clusters and total number of genomes are shown, respectively, in parentheses.
FIG 2  Requirement of the ato locus for utilization of SCFAs and increased transcription of virulence genes. (A) Model for catabolism of butyrate by the atoRDAEB locus in A. actinomycetemcomitans (Aa). The AtoR protein induces atoDAEB locus transcription by binding to the promoter. Butyrate is...
as a cue to enhance colonization (Fig. 2A). In media with butyrate as a defined carbon source, strains containing the ato locus grew significantly better and formed more robust biofilms than macaque-derived Aa strains lacking the ato locus as well as an isogenic human-derived Δato strain (Fig. 2B and D). Differences in pH between the different media tested were minimal (see Text S1, Table S2, and Table S3 in the supplemental material).

Interestingly, the only other whole-gene event associated with clade I is a deletion of several genes in the hybOABCD locus (Fig. S1). The products of this locus are known to metabolize H₂ under anaerobic conditions, using it as an electron donor for respiration and thus energy production (36, 37). We speculate that the production of protons by the Hyb system might exacerbate the toxic properties of butyrate or that acquisition of the Ato system made the energy-generating properties of the Hyb system redundant (Text S1). However, the interplay between the acquisition of the ato locus and the deletion of the hyb locus remains to be explored in future studies.

We next hypothesized that ato-harboring isolates may be uniquely adapted to utilize SCFAs in the human subgingival periodontal niche, since microbes that produce SCFAs are elevated at this site during Aa infection (38). The subgingival niche is highly anaerobic and has few energy sources other than amino acids and butyrate (39, 40).

Because the anaerobic subgingival pocket is the site of periodontitis, we hypothesized that Aa would respond to high levels of butyrate by upregulating genes associated with virulence. Using reverse transcription-quantitative PCR (RT-qPCR), we showed that Aa grown in butyrate, under anaerobic conditions, transcriptionally upregulated a leukotoxin (ltxA) gene that can kill activated neutrophils, the epithelial adhesin apiA gene, the biofilm/adherence gene flp-1, and the exopolysaccharide gene pgaC (Fig. 2C, E, and F).

The anti-inflammatory properties of SCFAs in the gut are well documented (41, 42), but the studies considering the periodontal pocket suggest that they may be proinflammatory and may worsen the disease (43–46). Thus, the impact of Aa butyrate metabolism on disease modulation is unclear. Interestingly, the four Finnish strains from clade II that do not have the ato locus appear to have been obtained from “healthy or minimally diseased individuals” (15–17).

In conclusion, the present study shows a higher level of genomic diversity in Aa than previously recognized and highlights a major clade dominated by isolates from rhesus macaques, raising the possibility of a long history of primate-Aa codiversification. Based on our molecular clock analysis, we favor the hypothesis that the MRCA of Aa was present in the MRCA of the Catarrhini about 32 mya. We propose that after this codiversification event, the lineage that would eventually dominate in humans acquired the ato locus, allowing it to more successfully colonize the periodontal pocket replete with butyrate and modulate the host response to disease.

**Data availability.** The genome sequences and the reads have been submitted to GenBank and SRA under BioProject PRJNA641505. Accession numbers are available in

---

**FIG 2 Legend (Continued)**

transported through the cytoplasmic membrane by the AtoE membrane transporter and is then converted to acetyl coenzyme A (acetyl-CoA) by the atoDA-encoded acetyl-CoA:acetoacetyl-CoA transferase and the atoB-encoded acetoacetyl-CoA thiolase. (B) Biofilm growth of two human strains (IDH781 and CU1000) from clade I and a rhesus macaque strain (RhAa3) from clade II on two different carbon sources, dextrose and butyrate. The biofilm growth in CFU is shown on the y axis. Although the Rh macaque strain has a significantly reduced CFU count from biofilm on butyrate compared to dextrose, it should be noted that there is only negligible planktonic growth in either condition. The large majority of growth is in the biofilm. Significance was calculated with a t test. (C) Working model for Aa outside and inside the periodontal pocket. Once inside the periodontal pocket, Aa utilizes SCFAs (e.g., butyrate) as a carbon source, and it increases tad biofilm, adherence, and exopolysaccharide and toxin production. (D) Crystal violet biofilm assay of the human strain IDH781 and a derived ato mutant strain on butyrate as a carbon source. The optical density of destaining fluid (595 nm) is shown on the y axis. Significance was calculated with a t test. (E) RT-PCR of different Aa virulence genes in the human strain IDH781. Relative expression is presented normalized to mean 16S rRNA ± standard error of the mean (SEM) for 6 separate cultures. (F) RT-PCR of the same relative expression data shown in panel E, but this time normalizing to the average transcript level for each gene in dextrose. The total starting RNA from dextrose and butyrate biofilm cells were equal before cDNA synthesis. Results are presented as the mean ± SEM from 6 separate cultures. The genes apiA, flp-1, ltxA, and pgaC code for adhesion, tad biofilm, toxin, and exopolysaccharides, respectively. A P value of ≤0.05 was considered significant. The asterisks represent P values of **≤0.01** and ****≤0.0001. NS, not significant.
Table S1. The raw trees and the Roary gene presence report are available to download from reference 47.

SUPPLEMENTAL MATERIAL
Supplemental material is available online only.

REFERENCES


20. Slots J, Reynolds HS, Genco RJ. 1980. Actinobacillus actinomycetemcomi-
tans in human periodontal disease: a cross-sectional microbiological
related prevalence and characteristics of Aggregatibacter actinomyce-
22. Li J, Nasidze I, Quinque D, Li M, Horz HP, Andre C, Garriga RM, Halbwax M,
Fischer A, Stoneking M. 2013. The saliva microbiome of Pan and Homo.
actinomycetemcomitans auto transpoter adenin Aae exhibits specificity
for buccal epithelial cells from humans and Old World primates. Infect
Aggregatibacter actinomycetemcomitans auto transporter adhesion ex-
hibits specificity for buccal epithelial cells in humans and Old World primates.
25. Kuchlasy SC. 2010. Aggregatibacter actinomycetemcomitans leukotoxin:
0022034510363682.
1987. Comparative studies on the biology of Actinobacillus actinomyce-
11:579. https://doi.org/10.1038/nature12721.
27. Kittichotirat W, Bumgarner RE, Chen C. 2016. Evolutionary divergence of
org/10.1177/0022034516638163.
chain carboxylic acid concentration in human gingival crevicular fluid. J
chain carboxylic acid concentration in human gingival crevicular fluid. J
the AtoSCDAEB regulon in the high molecular weight poly-(R)-3-
hydroxybutyrate biosynthesis in phaCAB(+) Escherichia coli. Metab Enz
31. Perelman P, Johnson WE, Roos C, Seuanez HN, Horvath JE, Moreira MA,
moustafa et al.
32. Jenkins LS, Nunn WD. 1987. Genetic variation of Aggregatibacter actinomyce-
33. Jenkins LS, Nunn WD. 1987. Regulation of the ato operon by the atoC
34. Jenkins LS, Nunn WD. 1987. Regulation of the ato operon by the atoC
35. Theodorou EC, Theodorou MC, Kyriakidis DA. 2012. Involvement of the
AtoSCDAEB regulon in the high molecular weight poly-(R)-3-
hydroxybutyrate biosynthesis in phaCAB(+) Escherichia coli. Metab Enz
36. Jenkins LS, Nunn WD. 1987. Genetic variation of Aggregatibacter actinomyce-
37. Jenkins LS, Nunn WD. 1987. Genetic variation of Aggregatibacter actinomyce-
chain carboxylic acid concentration in human gingival crevicular fluid. J
chain carboxylic acid concentration in human gingival crevicular fluid. J
chain carboxylic acid concentration in human gingival crevicular fluid. J
41. Smith PM,Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly YM,
Glickman JN, Garrett WS. 2013. The microbial metabolites, short-chain
https://doi.org/10.1126/science.1241165.
42. Kurita-Ochiai T, Fukushima K, Ochiai K. 1995. Volatile fatty acids, meta-
oblic by-products of periodontopathic bacteria, inhibit lymphocyte prolif-
10.1177/002203459507401070801.
chain carboxylic acid concentration in human gingival crevicular fluid. J
44. Kurita-Ochiai T, Fukushima K, Ochiai K. 1995. Volatile fatty acids, meta-
oblic by-products of periodontopathic bacteria, inhibit lymphocyte prolif-
10.1177/002203459507401070801.
45. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly YM,
Glickman JN, Garrett WS. 2013. The microbial metabolites, short-chain
https://doi.org/10.1126/science.1241165.
chain carboxylic acid concentration in human gingival crevicular fluid. J
47. Magrin GL, Strauss FJ, Benfatti CAM, Maia LC, Gruber R. 2020. Effects of
short-chain fatty acids on human oral epithelial cells and the potential
impact on periodontal disease: a systematic review of in vitro studies. Int
48. Moustafa AM, Velusamy SK, Denu L, Narechania A, Fine DH, Planet PJ.
2020. Trees and Roary output for adaptation by ancient horizontal acqui-
sition of butyrate metabolism genes in Aggregatibacter actinomycetem-
December 2020.