ORIGINAL REPORT

Genome-wide Association Study of Periodontal Pathogen Colonization

K. Divaris[1,2], K.L. Monda[2,3], K.E. North[2,4], A.F. Olshan[2], E.M. Lange[5,6,7], K. Moss[8], S.P. Barros[9], J.D. Beck[8], and S. Offenbacher[9]*

Abstract: *Pathological shifts of the human microbiome are characteristic of many diseases, including chronic periodontitis. To date, there is limited evidence on host genetic risk loci associated with periodontal pathogen colonization. We conducted a genome-wide association (GWA) study among 1,020 white participants of the* Atherosclerosis Risk in Communities Study*, whose periodontal diagnosis ranged from healthy to severe chronic periodontitis, and for whom "checkerboard" DNA-DNA hybridization quantification of 8 periodontal pathogens was performed. We examined 3 traits: "high red" and "high orange" bacterial complexes, and "high"* Aggregatibacter actinomycetemcomitans *(Aa) colonization. Genotyping was performed on the Affymetrix 6.0 platform. Imputation to 2.5 million markers was based on HapMap II-CEU, and a multiple-test correction was applied (genome-wide threshold of $p < 5 \times 10^{-8}$). We detected no genome-wide significant signals. However, 13 loci, including* KCNK1, FBXO38, UHRF2, IL33, RUNX2, TRPS1, CAMTA1*, and* VAMP3*, provided suggestive evidence ($p < 5 \times 10^{-6}$) of association. All associations reported for "red" and "orange" complex microbiota, but not for Aa, had the same effect direction in a second sample of 123 African-American participants. None of these polymorphisms was associated with periodontitis diagnosis. Investigations replicating these findings may lead to an improved understanding of the complex nature of host-microbiome interactions that characterizes states of health and disease.*

Key Words: bacteria, infection, chronic periodontitis, genetics, genome-wide association studies, dentistry.

Introduction

The study of the composition and the role of the human microbiome is a rapidly evolving and high-priority research area (Turnbaugh *et al.*, 2007). A dysbiotic relationship between a host and its microbiota is theorized to be an important component of many morbidities, including Crohn's disease, obesity, cancer, periodontitis, and others (Socransky and Haffajee, 2002; Badger *et al.*, 2010; Ahn *et al.*, 2012). Accordingly, investigations of the oral microbiome are gaining increased attention (Avila *et al.*, 2009). To date, over 600 species have been identified in the oral microbiome, and its diversity is greater than was initially theorized (Paster *et al.*, 2006; Keijser *et al.*, 2008). Specific bacterial species that are implicated in chronic periodontitis (CP) have been identified; in general, they are commensal and include Gram-negative anaerobes. Noteworthy, while harboring of periodontal pathogens is virtually universal, only a small proportion of individuals develop severe forms of CP.

There is a large body of literature in candidate-gene studies investigating the genetics of host inflammatory response, inflammatory mediators, and cytokines in periodontitis (Zhang *et al.*, 2011). The study of the interactions between host genetic factors and oral microbial

DOI: 10.1177/0022034512447951. [1]Department of Pediatric Dentistry, School of Dentistry, University of North Carolina-Chapel Hill, 228 Brauer Hall, CB#7450, NC 27599, USA; [2]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina-Chapel Hill, NC, USA; [3]The Center for Observational Research, Amgen Inc., Thousand Oaks, CA, USA; [4]Carolina Center for Genome Sciences, University of North Carolina-Chapel Hill, NC, USA; [5]Department of Genetics, School of Medicine, University of North Carolina-Chapel Hill, NC, USA; [6]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina-Chapel Hill, NC, USA; [7]Lineberger Comprehensive Cancer Center, University of North Carolina-Chapel Hill, NC, USA; [8]Department of Dental Ecology, School of Dentistry, University of North Carolina-Chapel Hill, NC, USA; and [9]Department of Periodontology, School of Dentistry, University of North Carolina-Chapel Hill, NC, USA; *corresponding author, offenbas@dentistry.unc.edu

A supplemental appendix to this article is published electronically only at http://jdr.sagepub.com/supplemental.

colonization, termed "periodontal infectogenomics" (Nibali *et al.*, 2011), is an important emerging area of research. Recent candidate-gene studies have suggested possible roles of polymorphisms in the *IL-6*, *Fc gamma* receptor, *IL-1A,* and *IL-1B* (Nibali *et al.*, 2007, 2010, 2011), whereas an earlier study by Socransky and colleagues (Socransky *et al.*, 2000) reported associations between *IL-1A* and *IL-1B* polymorphisms and "high" colonization with bacteria of the "red" and "orange complex". No study to date has carried out a whole-genome investigation for host genetic markers of colonization with periodontal bacteria. To this end, and to add to the knowledge base of infectogenomics, as well as the genetic component of periodontitis, the aim of this study was to investigate susceptibility loci for colonization with subgingival periodontal microbiota using a GWA analysis approach.

Materials & Methods

A detailed description of the study population, genotyping and imputation, quality control, and population stratification procedures is included in the online Appendix. In brief, we conducted a GWA study among 1,020 self-reported white participants of the Atherosclerosis Risk in Communities (ARIC) longitudinal cohort investigation (The ARIC Investigators, 1989). While ARIC is a study of atherosclerosis, CVD risk factors, and outcomes, a complete oral-dental examination took place between 1996 and 1998 during the fourth ARIC visit. As part of the Dental ARIC, apart from a complete clinical examination, five trained examiners collected subgingival microbial plaque samples (Beck *et al.*, 2001) from the mesial surfaces of all first permanent molars in a subset of participants. Participants' periodontal diagnoses ranged from healthy/mild disease (41%) to severe periodontitis (19%).

The "checkerboard" DNA-DNA hybridization method (Socransky *et al.*, 1994) was used on plaque samples to measure the extent of subgingival colonization with 8 periodontal pathogens: *Prevotella intermedia* [American Type Culture Collection (ATCC) 25611], *Campylobacter rectus* (ATCC 33238), *Fusobacterium nucleatum* (ATCC 10953), and *Prevotella nigrescens* (ATCC 33563) (belonging to the "orange" complex); *Porphyromonas gingivalis* (ATCC 33277), *Tannerella forsythia* (ATCC 43037), and *Treponema denticola* (ATCC 35404) ("red" complex); and *Aggregatibacter actinomycetemcomitans* (ATCC 43718). In that method, bacterial levels are expressed as counts relative to established microbial standards. Three dichotomous traits of "high" colonization with "red" complex, "orange" complex (Socransky *et al.*, 1998), and *Aggregatibacter actinomycetemcomitans* (*Aa*) were considered for analytical purposes. The two composite phenotypes were derived by the summation of bacterial counts for each species belonging to the "red" (n = 3) and "orange" (n = 4) groups, as described above. Because *Porphyromonas gingivalis* is considered the major periodontal pathogen implicated in periodontitis in adult populations, we explored for additional risk loci using its "high" colonization phenotype, defined as above, as a separate trait. Various approaches in defining the bacterial colonization profiles have been previously used, including summations of the absolute microbial counts (Offenbacher *et al.*, 2007), tertile categorization (Desvarieux *et al.*, 2005), and five-level categorization of log10-transformed counts (Papapanou *et al.*, 2000). For the present investigation, we defined a "high" colonization trait as the top quintile (20%) of each trait's distribution. The rationale for the selection of this phenotype is based on the fact that the "checkerboard" semi-quantitative method has a lower detection threshold of ~10^4 and reduced precision in the lower end of the distribution, whereas individuals with a "high" bacterial colonization profile may be those with reduced or impaired host and at high risk for periodontal tissue destruction.

Participants' DNA was extracted from blood samples drawn from an antecubital vein, and genotyping was performed with the Affymetrix Genome-Wide Human SNP Array 6.0 chip. The platform offers 906,600 markers for SNPs. Following rigorous quality control procedures, imputation to 2.5 million markers was performed with 669,450 SNPs and MACH v1.0.16 (http://www.sph.umich.edu/csg/abecasis/MaCH/index.html), based on HapMap Phase II CEU build 36 (Appendix Table 1). For analytical purposes, the dichotomous traits of "high colonization" phenotypes were entered into 3 logistic regression models, assuming multiplicative (log-additive) allelic effects. The logistic regression models included age, sex, examination center, and 10 principal components from the population stratification analysis as covariates. A multiple-comparisons correction was applied, and with 1 million independent tests assumed, the threshold of genome-wide statistical significance was set at $p < 5 \times 10^{-8}$. An additional *a priori* arbitrary threshold of $p < 5 \times 10^{-6}$ was set to prioritize and investigate loci with suggestive evidence of association. We also compiled a list of the 10,000 SNPs with the lowest p value for each of the 3 traits that we examined. We considered SNPs with minor allele frequency (MAF) 5% or greater.

In a second step, we used a separate sample of 123 African American (AA) Dental ARIC participants (mean age, 61.5 yrs) to verify the SNP associations that we prioritized from the white sample GWAS. The distribution of the AA participants' gender, smoking status, DM, and bacterial colonization profile by periodontal diagnosis is presented in Appendix Table 2. We followed identical quality control and analysis methods, including genotyping, population stratification, and genetic modeling. Imputation to 2,653,878 SNPs for the AA sample was based on the YRI sample of HapMap Phase II build 36. All GWA analyses were performed with the ProbABEL software (Aulchenko *et al.*, 2010). SNP annotations were performed with WGAViewer ver.1.26l (Ge *et al.*, 2008) and Snipper ver. 1.2 (http://csg.sph.umich.edu/boehnke/snipper/), and loci visualizations with LocusZoom ver.1.1 (Pruim *et al.*, 2010) and Haploview ver.4.2 (Barrett *et al.*, 2005). We used

Table 1.
Distribution of Participants' Demographic Characteristics, Smoking Status, and Bacterial "High Colonization" (defined as the top quintile of the distribution) Profile (with "Red" complex, "Orange" complex, and *Aggregatibacter actinomycetemcomitans*) Overall, and Stratified by Periodontal Diagnosis (CDC/AAP classification) among the Dental ARIC Study Self-reported White Participants (n = 1,020)

	Total	Healthy/Mild	Moderate	Severe
	n (column %)*	n (column %)*	n (column %)*	n (column %)*
n (row %)	1,020 (100)	416 (41)	415 (41)	189 (19)
Sex				
Females	478 (47)	246 (59)	178 (43)	54 (29)
Males	542 (53)	170 (41)	237 (57)	135 (71)
Smoking status				
Never	410 (41)	204 (50)	151 (37)	55 (30)
Former	468 (47)	166 (40)	211 (52)	91 (50)
Current	121 (12)	42 (10)	42 (10)	37 (20)
Diabetes mellitus				
No	875 (86)	379 (91)	353 (85)	143 (76)
Yes	144 (14)	37 (9)	62 (15)	45 (24)
Age (yrs; mean, standard deviation)	63.2 (5.7)	62.3 (5.5)	63.6 (5.7)	64.0 (5.9)
"High" bacterial colonization (n, % of column)				
"Red" complex	203 (20)	58 (14)	84 (20)	61 (32)
"Orange" complex	201 (20)	73 (18)	72 (17)	56 (30)
A. actinomycetemcomitans	204 (20)	75 (18)	81 (20)	48 (25)

Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/index.shtml) to assess potentially damaging non-synonymous SNPs. Gene reporting was based on the "HUGO Gene Nomenclature" (www.genenames.org).

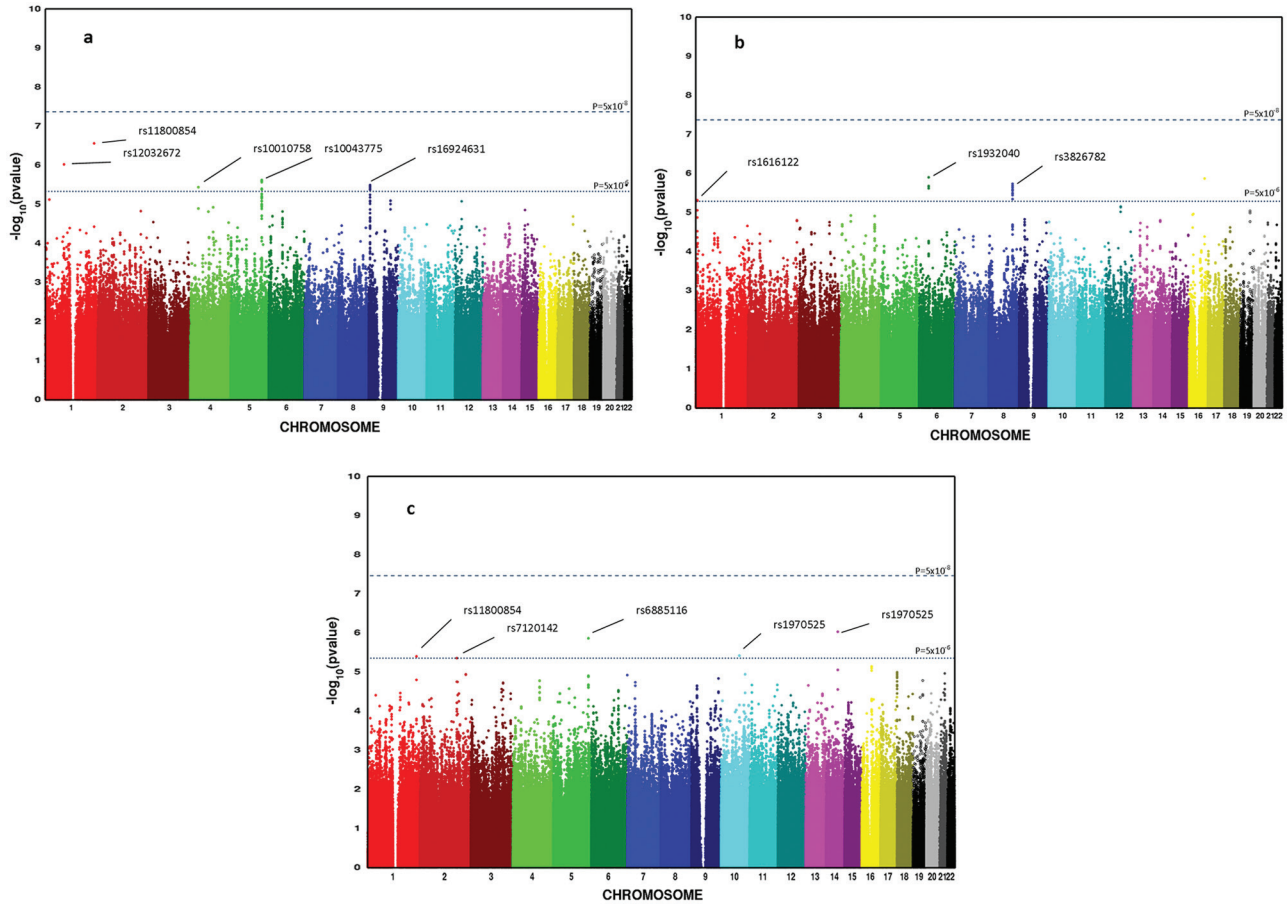
Results

The sample's descriptive information is presented in Table 1. Participants had a mean age of 63 yrs and approximately even gender distribution. Forty percent were never-smokers and 12% current smokers. Fourteen percent were diagnosed with DM. "High" bacterial colonization profiles were found in all groups of periodontal diagnosis according to the CDC classification; however, the prevalence of "high red" colonization cases was more than double among participants with severe periodontitis compared with those with mild or no disease. Similar, but less pronounced, associations were noted for "orange" complex and *Aa*.

Of the 2,178,777 examined SNPs, none had $p < 5 \times 10^{-8}$. A list of the top 10,000 SNPs ordered in ascending p value (all $p < 4.0 \times 10^{-3}$) for each trait is available at: http://genomewide.net/public/aric/dental/infectogenomics/topSNPs.xls. Fifty-three SNPs had $p < 5 \times 10^{-6}$ and thus were prioritized for further investigation. Lambda inflation factors for 3 traits were: "red"—1.040, "orange"—1.045, and *Aa*—1.032. The corresponding Q-Q plots are presented in the Appendix (Appendix Fig. 1). Upon inspection of the prioritized SNPs, there were 5 loci that emerged for "red", 3 loci for "orange", and 5 loci for *Aa* colonization (Fig. 1). Of those, 1 locus on 1q42 was shared for the "red" complex and *Aa*. Graphical representations of the genomic areas adjacent to 6 of these loci are presented in Fig. 2.

The strongest signal with regard to "red" complex colonization (Table 2) was produced by rs11800854 in the 1q42 locus [$p = 2.8 \times 10^{-7}$; OR = 12.3 (95% CI = 3.7, 41.3); MAF in HapMap-CEU (MAF-CEU): 0.067] in the promoter region (30 Kb upstream) of *KCNK1* and adjacent to *KIAA1804*. The common [G] allele showed 3% enrichment among "high colonization cases" for both "red" complex and *Aa*. Another locus in chromosome 1p22 was marked by rs12032672 ($p = 9.6 \times 10^{-7}$), ~500 Kb upstream of *PKN2*. Rs10043775, in linkage disequilibrium (LD) with multiple markers in the 5q33 locus, represents a missense change in the *FBXO38* gene (resulting

Figure 1.
Manhattan plots of the GWAS results (–log10 p values of the ~2.5 million examined SNPs arranged by chromosome) for the 3 bacterial pathogen "high" colonization traits among the participants of the Dental Atherosclerosis Risk in Communities Study cohort (n = 1,020). Top panel: "Red" complex. Middle panel: "Orange complex". Bottom panel: *Aggregatibacter actinomycetemcomitans.*



in [Pro]→[Arg] substitution, predicted as "benign" according to PolyPhen-2) and provided the strongest signal in that locus ($p = 2.4 \times 10^{-6}$), also adjacent to *HTR4*. A high LD area in 9p24 including the *UHRF2*, *GLDC*, *TPD52L3,* and *IL33* genes is marked by rs16924631 [intronic to *UHRF2*; $p = 3.2 \times 10^{-6}$; OR = 2.29 (95% CI = 1.61, 3.24); MAF-CEU: 0.275], of which the [C] risk allele showed almost 10% enrichment among "cases". An intronic variant (rs10010758) of the *TBC1D1* gene, adjacent (24 Kb) to *PTTG2,* provided the strongest signal in the 4p14 locus [$p = 3.7 \times 10^{-6}$; OR = 1.91 (95% CI = 1.25, 2.21)]. The exploratory analysis for *Pg* "high" colonization revealed 3 loci with $p < 5 \times 10^{-6}$, including *OTOF*, *C2Orf70*, *CIB4*, *DAB2IP*, *TTLL11,* and *AKNRD3* (Appendix Table 3).

The common allele of rs1932040 showed 9% enrichment and provided the strongest association signal with "high orange" bacterial colonization [$p = 1.3 \times 10^{-6}$, OR = 2.47 (95% CI = 1.67, 3.65)], marking an intergenic area between *RUNX2* and *CLIC5* on the 6p21.1 locus. A low recombination area on 8q23, adjacent to *TSPS1* (1.3 Mb) and *CSMD3* (672 Kb), was marked by multiple alleles, of which rs9942773 provided the strongest signal ($p = 1.9 \times 10^{-6}$) and 10% enrichment among "cases". A variant intronic to *CAMTA1* [rs1616122; $p = 4.9 \times 10^{-6}$; OR = 1.85 (95% CI = 1.41, 2.42)] marked the 1p36.2 locus.

The 1q42 locus that was identified for "red" bacteria also provided the third strongest association signal for *Aa* [rs11800854; $p = 4.0 \times 10^{-6}$; OR = 8.12 (95% CI = 2.73, 24.11)]. The common [T] allele of rs11621969 was also nominally associated with high *Aa* colonization ($p = 9.4 \times 10^{-7}$) and was adjacent to *FOS* and *JPD2* in 14q24. The rare [G] allele of rs1970525 was more than twice as prevalent among "cases" (0.118 *vs.* 0.054 among "non-cases"), provided the strongest signal in the 10q23 locus [$p = 3.8 \times 10^{-6}$; OR = 2.89 (95% CI = 1.85, 4.52); MAF-CEU: 0.045)], and represents a nonsense-mediated decay transcript variant in the *GRID1* gene. Rs9287989 is adjacent to *KIAA1715* (30 Kb) and *EVX2* (227 Kb) and marked the 2q31 locus ($p = 4.4 \times 10^{-6}$). An intronic variant of *ODZ2* (rs6885116) provided the strongest signal in the locus 5q35 [$p = 1.4 \times 10^{-6}$; OR = 2.57

Figure 2.

Visualization of 6 loci that were marked by suggestive genome-wide association evidence ($p < 5 \times 10^{-6}$) for "high" colonization with "red" complex, "orange" complex, and *Aggregatibacter actinomycetemcomitans* among the participants of the Dental Atherosclerosis Risk in Communities Study cohort (n = 1,020). The vertical axis corresponds to each marker's associated –log10 p value. The overlaid recombination rate plot and the pairwise linkage disequilibrium values with index SNPs were calculated based on HapMap II – CEU. Rs10043775, rs16924631, rs10010758, and rs12032672 provided association signals for "red" complex colonization; rs1616122 with "orange" complex; and rs11800854 with *Aggregatibacter actinomycetemcomitans* "high" colonization.

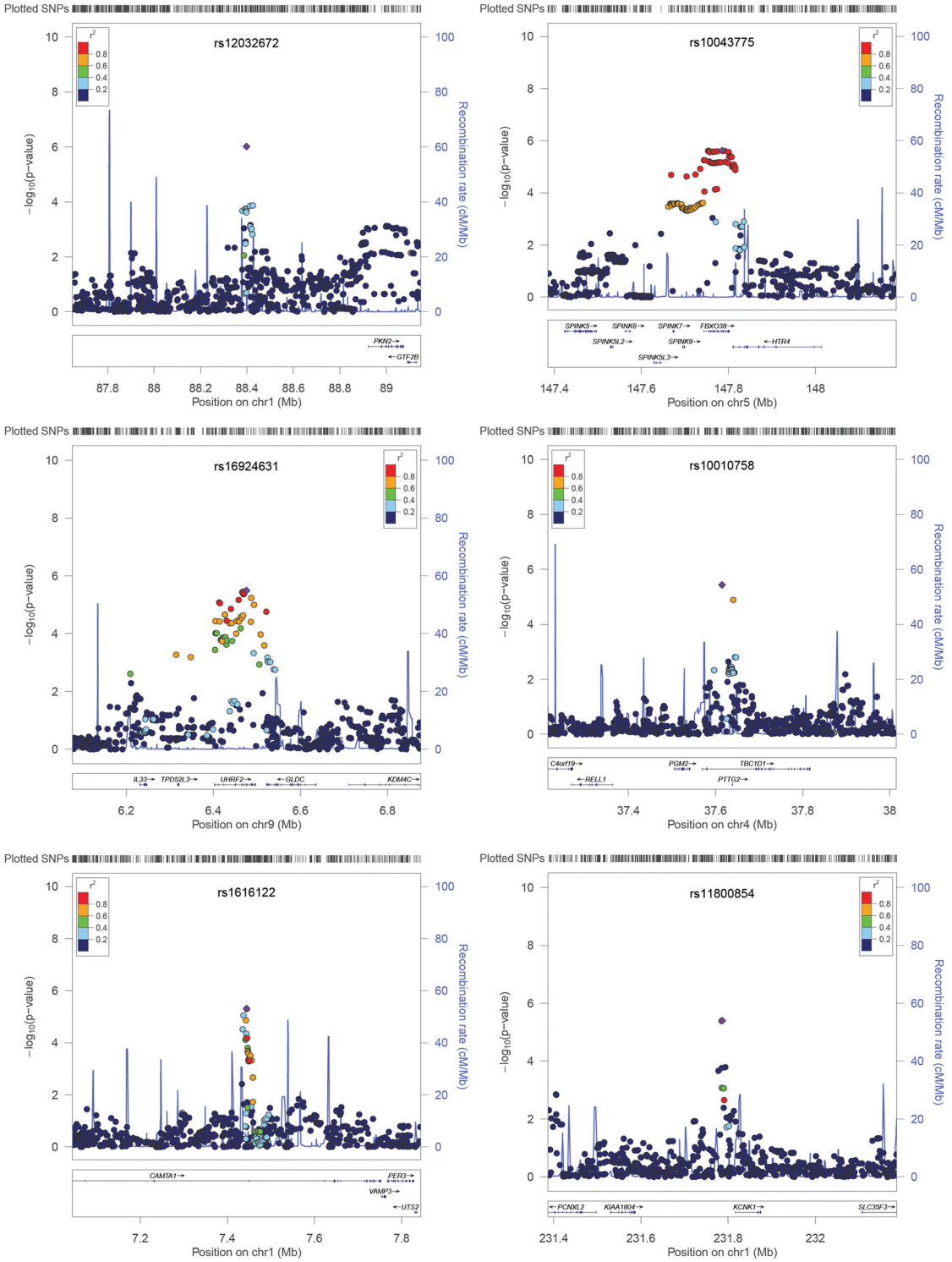


Table 2.

Genome-wide Association Analysis Results of the High Colonization Traits (highest quintile of the distribution *vs.* the other four; quantified with DNA-DNA "checkerboard" hybridization) for "Red" Complex, "Orange" Complex, and *Aggregatibacter actinomycetemcomitans*, among the White Participants of the Dental ARIC Study (n = 1,020)

Locus	SNP	Position Build36	Strand	Imputed	mja[1]	MAF[2] (HapMap II-CEU)	Closest Gene(s) and Position or Distance	Risk Allele Frequency Low/ High Colonization	P value	Odds Ratio (95% CI)
"Red" complex										
1q42	rs11800854[3]	231786607	+	yes	G	[A] 0.068	*KCNK1* (30Kb)	[G] 0.947/0.978	2.8×10^{-7}	12.3 (3.7, 41.3)
1p22	rs12032672	88398224	+	yes	A	[C] 0.350	*PKN2* (524Kb)	[C] 0.332/0.446	9.6×10^{-7}	1.99 (1.50, 2.62)
5q33	rs10043775[4]	147785313	+	yes	T	[C] 0.274	*FBXO38* (missense change)[6]/ *HTR4* (19Kb)	[T] 0.703/0.791	2.4×10^{-6}	2.06 (1.51, 2.83)
9p24	rs16924631[5]	6476308	+	no	G	[C] 0.142	*UHRF2* (non-coding transcript variant) / *GLDC* (46kb)	[C] 0.122/0.219	3.2×10^{-6}	2.29 (1.61, 3.24)
4p14	rs10010758	37614913	+	no	T	[C] 0.275	*TBC1D1*(intron variant)/ *PTTG2* (24Kb)[7]	[C] 0.291/0.384	3.7×10^{-6}	1.91 (1.45, 2.51)
"Orange" complex										
6p21.1	rs1932040[8]	45804766	+	yes	G	[A] 0.142	*CLIC5* (169Kb)/ *RUNX2* (178Kb)	[G] 0.808/0.896	1.3×10^{-6}	2.47 (1.67, 3.65)
8q23	rs9942773[9]	115190203	+	yes	A	[C] 0.283	*CSMD3* (672Kb)/*TRPS1* (1.3Mb)	[A] 0.703/0.803	1.9×10^{-6}	2.07 (1.51, 2.82)
1p36.2	rs1616122	7444172	+	no	T	[C] 0.482	*CAMTA1* (intron variant)/ *VAMP3* (310Kb)	[T] 0.506/0.624	4.9×10^{-6}	1.85 (1.41, 2.42)
Aggregatibacter actinomycetemcomitans										
14q24	rs11621969	74883781	+	yes	T	[C] 0.167	*FOS* (65Kb)/*JDP2* (85Kb)	[T] 0.789/0.885	9.4×10^{-7}	2.46 (1.68, 3.62)
5q35	rs6885116	167576123	+	yes	A	[G] 0.084	*ODZ2* (intron variant)/ *WWC1* (76Kb)	[G] 0.078/0.169	1.4×10^{-6}	2.57 (1.76, 3.74)
10q23	rs1970525[10]	87624904	+	yes	G	[C] 0.045	*GRID1* (NMD[12] transcript variant) /*MI346*/ *WAPAL* (560Kb)	[G] 0.054/0.118	3.8×10^{-6}	2.89 (1.85, 4.52)
1q42	rs11800854[11]	231786607	+	yes	G	[A] 0.067	*KCNK1* (30Kb)/ *KIAA1804* (199Kb)	[G] 0.947/0.978	4.0×10^{-6}	8.12 (2.73, 24.11)
2q31	rs9287989	176425987	+	yes	C	[T] 0.433	*KIAA1715* (73Kb)/ *EVX2* (227Kb)/ *EXTLP2*[13] (10Kb)	[C] 0.484/0.605	4.4×10^{-6}	1.80 (1.39, 2.33)

Single-nucleotide polymorphisms (SNPs) with minor allele frequency (MAF-HapMap II CEU) of ≥ 5% and associated $p < 5 \times 10^{-6}$. The SNP with the lowest p value *per* locus is presented; additional prioritized SNPs in each locus are presented below, along with corresponding R^2 (based on 1000 genomes pilot 1 release) with the top SNPs.

[1]Major allele.

[2]Minor allele frequency.

[3]Additional SNPs in locus with $p < 5 \times 10^{-6}$: rs6682365 ($r^2 = 1.00$).

[4]Additional SNP in locus with $p < 5 \times 10^{-6}$: rs10068216 ($r^2 = 1.00$), rs10072051 ($r^2 = 1.00$), rs17108251 ($r^2 = 1.00$), rs10044061 ($r^2 = 1.00$), rs4349707 ($r^2 = 1.00$), rs10477376 ($r^2 = 1.00$), rs9325095 ($r^2 = 1.00$), rs10041283 ($r^2 = 1.00$), rs9325097 ($r^2 = 0.87$), rs3734120 ($r^2 = 1.00$), rs4574533 ($r^2 = 0.87$), rs4274967 ($r^2 = 1.00$), rs4274968 ($r^2 = 1.00$), rs6884076 ($r^2 = 1.00$), rs9325098 ($r^2 = 1.00$).

[5]Additional SNP in locus with $p < 5 \times 10^{-6}$: rs11795355 ($r^2 = 1.00$), rs7876000 ($r^2 = 1.00$), rs10975603 ($r^2 = 0.93$), rs16924626 ($r^2 = 0.93$), rs16924624 ($r^2 = 1.00$), rs10975605 ($r^2 = 0.93$), rs10115883 ($r^2 = 0.93$), rs10122116 ($r^2 = 0.93$).

[6]T>C – Ser>Pro, 35b from the exon boundary.

[7]$r^2 = 0.29$ with rs6811863, which is a missense change in *PTTG2*: G>C – [Arg]→[Pro].

[8]Additional SNPs in locus with $p < 5 \times 10^{-6}$: rs12525547($r^2 = 0.93$), rs9349326 ($r^2 = 0.93$), rs16873698($r^2 = 0.93$).

[9]Additional SNP in locus with $p < 5 \times 10^{-6}$: rs10089040 ($r^2 = 1.00$), rs9942776 ($r^2 = 1.00$), rs10086149 ($r^2 = 1.00$), rs7845243 ($r^2 = 0.87$), rs10105817 ($r^2 = 1.00$), rs7006291 ($r^2 = 1.00$), rs11779159 ($r^2 = 1.00$), rs11783996 ($r^2 = 1.00$), rs10098056 ($r^2 = 1.00$), rs7018200 ($r^2 = 0.92$).

[10]Additional SNPs in locus with $p < 5 \times 10^{-6}$: rs4325261 ($r^2 = 1.00$).

[11]Additional SNP in locus with $p < 5 \times 10^{-6}$: rs6682365 ($r^2 = 1.00$).

[12]Nonsense-mediated decay.

[13]*EXTLP2* is a pseudogene.

(95% CI = 1.76, 3.74); MAF-CEU: 0.084], showing 9% enrichment among "cases".

The examination of our prioritized SNPs in the AA sample revealed that although no p value met conventional replication significance criteria ($p < 0.05$), all 8 of the effect estimates for the "red" and "orange" complex bacterial colonization traits (Appendix Table 4) had the same direction in the white and AA samples. In contrast, the results for *Aa* colonization were not consistent between the two samples. None of the prioritized SNPs were associated with periodontitis diagnosis at a nominal significance level ($p < 0.05$). Further, in Appendix Table 5 we present the effect estimates and associated p value of polymorphisms that were previously reported as associated with periodontal pathogen colonization profiles in studies by Socransky (Socransky *et al.*, 2000) and Nibali (Nibali *et al.*, 2010, 2011). Of those, only 2 *IL6* promoter polymorphisms (rs1800795 and rs1800796) were found to be significantly associated ($p < 0.05$) with high "red complex" colonization.

Discussion

This study is the first report of a genome-wide association analysis investigating risk loci for colonization with pathogenic periodontal bacteria. Although limited by the sample size, this investigation explores a novel phenotype and benefits from a comprehensive quantitative phenotypical characterization. In this investigation, we did not carry out a comprehensive investigation of genetic risk loci for chronic periodontitis, but examined bacterial colonization traits that are implicated in its pathogenesis and progression. We support that an increased understanding of the genetic underpinning of interactions between the host and exogenous or symbiotic bacterial communities has the potential to advance our knowledge not only of periodontitis, but also of other chronic inflammatory and microbiome-related diseases. Several of the risk loci identified in this study may offer promising leads for further exploration and mechanistic studies. For example, a recent experimental study showed that *NLRP6* knockout mice had altered gut flora and were more susceptible to colitis (Elinav *et al.*, 2011).

Acknowledging the limitations of GWAS as a discovery tool, the reported SNP estimates may be overestimates of the true genetic effects (Göring *et al.*, 2001), whereas many true association signals with modest effects may not have been prioritized due to insufficient power. In contrast, the "unbiased" and comprehensive consideration of approximately 2.5 million genetic markers is an improvement over candidate-gene studies, which are typically motivated by prior information and biological plausibility (Hirschhorn and Daly, 2005). Because it may be difficult to conduct replication studies or a larger GWAS in the near future, we examined our reported associations among a second small sample of AA individuals. While this approach provided compelling evidence for the "red" and "orange" complex results, this was not the case for the *Aa* colonization results. This may be due to differing genetic susceptibility factors for *Aa* colonization between the two racial groups, a fact that is supported by well-documented differences in bacterial colonization profiles between ethnic groups (Beck *et al.*, 1992; Haffajee *et al.*, 2005). Our finding of no association of the identified SNPs with CP diagnosis is in support of both the multifactorial disease pathogenesis and the examination of bacterial colonization as a distinct trait. Although we did not collect information on this domain, it cannot be excluded that some participants might have received anti-inflammatory or antibiotic therapy prior to the oral examination and microbial plaque collection. It is possible that this could have influenced the oral bacterial colonization profile of some participants. However, it must be noted that the requirement of antibiotic prophylaxis prior to dental treatment was a study exclusion criterion.

Upon replication or validation, these findings have the potential to unveil pathways and mechanisms that direct the host's symbiosis with healthy microflora that, if altered, may predispose for states of disease (Darveau, 2010). The consideration of specific microbiota as a distinct exposure in investigations of periodontal, oral, and systemic health is consistent with the paradigm of "periodontal medicine" and may provide novel insight into the 'oral-systemic diseases' connection. Although the prevention and treatment of periodontitis are obvious goals, the links of CP with other systemic conditions, and the "common theme" of pathogenic ecological shift in other diseases, provide opportunities for even greater impact.

Acknowledgments

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; National Institutes of Health contract HHSN268200625226C; National Institute of Environmental Health Sciences grant P30ES010126; and National Institute of Dental and Craniofacial Research grants R01DE11551, R01DE021418. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

References

Ahn J, Chen CY, Hayes RB (2012). Oral microbiome and oral and gastrointestinal cancer risk. *Cancer Causes Control* 23:399-404.

The ARIC investigators (1989). The Atherosclerosis Risk in Communities (ARIC)

Study: design and objectives. *Am J Epidemiol* 129:687-702.

Aulchenko YS, Struchalin MV, van Duijn CM (2010). ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* 11:134.

Avila M, Ojcius DM, Yilmaz O (2009). The oral microbiota: living with a permanent guest. *DNA Cell Biol* 28:405-411.

Badger JH, Ng PC, Venter JC (2010). The human genome, microbiomes, and disease. In: Metagenomics of the human body. Nelson KE, editor. New York, NY, USA: Springer, pp. 1-14.

Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-265.

Beck JD, Koch GG, Zambon JJ, Genco RJ, Tudor GE (1992). Evaluation of oral bacteria as risk indicators for periodontitis in older adults. *J Periodontol* 63:93-99.

Beck JD, Elter JR, Heiss G, Couper D, Mauriello SM, Offenbacher S (2001). Relationship of periodontal disease to carotid artery intima-media wall thickness: the atherosclerosis risk in communities (ARIC) study. *Arterioscler Thromb Vasc Biol* 21:1816-1822.

Darveau RP (2010). Periodontitis: a polymicrobial disruption of host homeostasis. *Nat Rev Microbiol* 8:481-490.

Desvarieux M, Demmer RT, Rundek T, Boden-Albala B, Jacobs DR Jr, Sacco RL, *et al.* (2005). Periodontal microbiota and carotid intima-media thickness: the Oral Infections and Vascular Disease Epidemiology Study (INVEST). *Circulation* 111:576-582.

Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ, *et al.* (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* 145:745-757.

Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ, *et al.* (2008). WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res* 18:640-643.

Göring HH, Terwilliger JD, Blangero J (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* 69:1357-1369.

Haffajee AD, Japlit M, Bogren A, Kent RL Jr, Goodson JM, Socransky SS (2005). Differences in the subgingival microbiota of Swedish and USA subjects who were periodontally healthy or exhibited minimal periodontal disease. *J Clin Periodontol* 32:33-39.

Hirschhorn JN, Daly MJ (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6:95-108.

Keijser BJ, Zaura E, Huse SM, van der Vossen JM, Schuren FH, Montijn RC, *et al.* (2008). Pyrosequencing analysis of the oral microflora of healthy adults. *J Dent Res* 87:1016-1,020.

Nibali L, Ready DR, Parkar M, Brett PM, Wilson M, Tonetti MS, *et al.* (2007). Gene polymorphisms and the prevalence of key periodontal pathogens. *J Dent Res* 86:416-420.

Nibali L, Donos N, Farrell S, Ready D, Pratten J, Tu YK, *et al.* (2010). Association between interleukin-6 -174 polymorphism and *Aggregatibacter actinomycetemcomitans* in chronic periodontitis. *J Periodontol* 81:1814-1819.

Nibali L, Madden I, Franch Chillida F, Heitz-Mayfield L, Brett P, Donos N (2011). IL6-174 genotype associated with *Aggregatibacter actinomycetemcomitans* in Indians. *Oral Dis* 17:232-237.

Offenbacher S, Barros SP, Singer RE, Moss K, Williams RC, Beck JD (2007). Periodontal disease at the biofilm-gingival interface. *J Periodontol* 78:1911-1925.

Papapanou PN, Neiderud AM, Papadimitriou A, Sandros J, Dahlén G (2000). "Checkerboard" assessments of periodontal microbiota and serum antibody responses: a case-control study. *J Periodontol* 71:885-897.

Paster BJ, Olsen I, Aas JA, Dewhirst FE (2006). The breadth of bacterial diversity in the human periodontal pocket and other oral sites. *Periodontol 2000* 42:80-87.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, *et al.* (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26:2336-2337.

Socransky SS, Haffajee AD (2002). Dental biofilms: difficult therapeutic targets. *Periodontol 2000* 28:12-55.

Socransky SS, Smith C, Martin L, Paster BJ, Dewhirst FE, Levin AE (1994). "Checkerboard" DNA-DNA hybridization. *Biotechniques* 17:788-792.

Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL Jr (1998). Microbial complexes in subgingival plaque. *J Clin Periodontol* 25:134-144.

Socransky SS, Haffajee AD, Smith C, Duff GW (2000). Microbiological parameters associated with IL-1 gene polymorphisms in periodontitis patients. *J Clin Periodontol* 27:810-818.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007). The human microbiome project. *Nature* 449:804-810.

Zhang J, Sun X, Xiao L, Xie C, Xuan D, Luo G (2011). Gene polymorphisms and periodontitis. *Periodontol 2000* 56:102-124.