Arthritis & Rheumatology
DOI 10.1002/art.40003

Evidence for Immune Relevance of *Prevotella copri*, a Gut Microbe, in Patients with Rheumatoid Arthritis

Annalisa Pianta[1], Sheila Arvikar[1], Klemen Strle[1], Elise E. Drouin[1†], Qi Wang[2], Catherine E. Costello[2], and Allen C. Steere[1].

This work was supported by grants from the American College of Rheumatology Innovative Grant Program, "Within our Reach, Finding a Cure for RA"; the Ounsworth-Fitzgerald Foundation; Mathers Foundation; English, Bonter, Mitchell Foundation; Littauer Foundation; Lillian B. Davey Foundation, and the Eshe Fund (to A.C.S.), and by NIH grants P41 GM104603, S10 RR020946 and S10 OD010724 (to C.E.C.). S.A. received support from a Scientist Development Award from the Rheumatology Research Foundation.

[1] Annalisa Pianta, PhD, Sheila Arvikar, MD, Klemen Strle, PhD, Elise E. Drouin, PhD, Allen C. Steere, MD: Center for Immunology and Inflammatory Diseases, Division of Rheumatology, Allergy and Immunology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; [2] Qi Wang, PhD, Catherine E. Costello, PhD: Center for Biomedical Mass Spectrometry, Boston University School of Medicine, Boston, Massachusetts, USA; †Current address: Agenus Inc, 02421 Lexington (MA), United States.

Address correspondence to Annalisa Pianta, PhD, Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, CNY 149/8301, 55 Fruit Street, Boston, MA 02114, United States. Tel.: +16177261530; fax: +16177261544.

E-mail address: apianta@mgh.harvard.edu

Received: Aug 25, 2016; Revised: Oct 20, 2016; Accepted: Nov 15, 2016

Objective. Prevotella copri, an intestinal microbe, may over-expand in stool samples of patients with new-onset rheumatoid arthritis (NORA), but it is not yet clear whether the organism has immune relevance in RA pathogenesis.

Methods. HLA-DR-presented peptides (T cell epitopes) from *P. copri* were sought directly from patients' synovial tissue or peripheral blood mononuclear cells (PBMC) using tandem mass spectrometry, followed by testing the antigenicity of peptides or their source proteins using samples from RA patients or comparison groups. T cell reactivity was determined by ELISpot assays; antibody responses were measured by ELISA, and cytokine/chemokine determinations were made by Luminex. 16S rDNA of *P. copri* was sought in serum and synovial fluid samples using nested PCR.

Results. In PBMC, we identified an HLA-DR-presented peptide from a 27-kD protein of *P. copri* (*Pc*-p27), which stimulated Th1 responses in 42% of NORA patients. In both NORA and chronic RA patients, one subgroup had IgA antibody responses to *Pc*-p27 or the whole organism, which correlated with Th17 cytokine responses and frequent anti-citrullinated protein antibodies (ACPA). The other subgroup had IgG *P. copri* antibodies, which were associated with *Prevotella* DNA in synovial fluid, *P. copri*-specific Th1 responses, and less frequent ACPA. In contrast, *P. copri* antibody responses were rarely found in patients with other rheumatic diseases or in healthy controls.

Conclusion. Subgroups of RA patients have differential IgG or IgA immune reactivity with *P. copri*, which appears to be specific for this disease. These observations provide evidence that *P. copri* is immune-relevant in RA pathogenesis.

Rheumatoid arthritis (RA) results from a complex interplay between genetic and environmental factors (1,2). Great progress has been made in the identification of genetic factors and inflammatory pathways that influence the disease (1,3), but environmental factors are only now being determined (4). A key hypothesis is that specific organisms in the mouth or gut microbiota, the composition of which is strongly influenced by environmental cues, may shape mucosal and systemic immune responses that affect joints in RA patients (4-7).

Using high-through-put sequencing, Scher et al. showed that *Prevotella copri* in the gut microbiota was over-expanded in stool samples from patients with new-onset RA (NORA) compared with patients with chronic RA (CRA), psoriatic arthritis, or healthy people (8). In NORA patients, *Prevotella* abundance in the gut was at the expense of *Bacteroides fragilis*, an organism that is important for Treg function (9,10). A second metagenome-wide analysis of fecal samples in RA patients showed dysbiosis in the gut, but also in the mouth and salivary glands (11). Moreover, a recent study in mice showed that dysbiosis contributes to arthritis development via activation of autoreactive T cells in the intestine (12). However, it is unclear whether over-expansion of *P. copri* in the gut in human patients has the potential to affect immune cell functions at both mucosal and systemic sites, thereby contributing to RA disease pathogenesis.

Whereas the previous studies used unbiased, discovery-based approaches to assess dysbiosis of microorganisms in the oral or gut microbiome, we developed an unbiased, discovery-based approach to identify novel, immunogenic T cell epitopes in patients with chronic inflammatory arthritis. With this approach, *in vivo* HLA-DR presented peptides are identified directly from patients' synovial tissue, synovial fluid mononuclear cells (SFMC), or peripheral blood mononuclear cells (PBMC) by tandem mass spectrometry (LC-MS/MS)

(13,14), followed by testing the antigenicity of identified peptides and their source proteins using patients' samples (15-19).

Recently, we used this approach to search for T cell epitopes of proteins derived from microbes implicated in RA. We report here the identification of an HLA-DR-presented peptide (T cell epitope) derived from a *P. copri* 27-kD protein (*Pc*-p27), which stimulated Th1 responses in 42% of RA patients. We then found that *P. copri* induced differential antibody responses to this protein or the whole organism in a substantial portion of RA patients. These observations provide evidence for immune relevance of *P. copri* in RA pathogenesis.

Patients and Methods are described in Supplementary text.

RESULTS

Identification of naturally presented, microbial HLA-DR-presented peptides. Using tandem mass spectrometry (LC-MS/MS), we identified HLA-DR-presented peptides in synovial tissue (N=4), SFMC (N=3) or PBMC (N=2) from 5 patients with NORA or chronic RA (CRA). From the 17 HLA-DR-presented peptides identified from the PBMC of one CRA patient (called here RA1), one *P. copri* sequence was found (Figure 1A). In contrast, no sequences from *P. gingivalis* or from *B. burgdorferi*, the Lyme disease agent, were identified in any sample.

At disease onset, patient RA1, who had 2 copies of the RA "shared epitope" (HLA-DRB1*0401 and 0101), had severe symmetrical polyarthritis affecting large and small joints. During the course of the disease, she developed a positive test for ACPA, but not RF. Despite treatment with disease modifying anti-rheumatic drugs (DMARDs), she had recurrent episodes of knee swelling with evidence of destructive changes in cartilage and bone. The HLA-DR-

presented peptide derived from *P. copri* was identified from PBMC obtained during one such episode 7 years after disease onset.

The peptide sequence of 19 amino acids had 100% sequence homology with part of the signal sequence of a 27-kD protein of *P. copri* (*Pc*-p27, WP_022121928.1) (Figure 1A). The peptide had minimal sequence homology with any human peptide, suggesting that it was not a human protein erroneously assigned with a microbial database. Using signalP 4.0 software (20), this HLA-DR-presented *P. copri* peptide was predicted to be part of the Sec secretion signal peptide sequence (D score = 0.869), strongly suggesting that the peptide would be cleaved from the source protein. This signal peptide was not predicted to be lipidated (LipoP 1.0 Server). In addition, the algorithm TEPITOPE predicted that the peptide was highly promiscuous, as is typical of signal peptides, and would bind all 25 HLA-DR molecules modeled in the program (21), including the patient's DRB1*0101 and 0401 molecules. When her PBMC were stimulated with this *P. copri* peptide in an IFN-γ ELISpot assay, her T cells secreted levels of IFN-γ that were >3 times background (Figure 1A insert), attesting to the peptide's immunogenicity.

T cell reactivity to *P. copri* peptides in NORA patients. To determine the immunogenicity of HLA-DR-presented peptides of *Pc*-p27 more broadly, we used PBMC obtained from our cohort of patients with new-onset RA (NORA) seen prior to DMARD therapy, the time when immune responses would be expected to be most robust. All patients met the ACR/EULAR criteria for that disease (22).

When PBMC from 39 NORA patients and from the case patient RA1 (a CRA patient) were stimulated with *Pc*-p27 peptide 1, 17 of the 40 patients (42%) secreted levels of IFN-γ that were >3 SD above the mean value of HC (P=0.0002), using an IFN-γ/IL-17 Double-Color ELISpot assay (Figure 1B). In comparison, patients with Lyme arthritis (LA) lacked reactivity

with this peptide (P<0.0001). Stimulation of cells with PHA, used as a positive control, verified the viability of cells in all patients. The predominant response to stimulation with *Pc*-p27 in the RA patients was a Th1-type response, whereas PBMC from only one RA patient secreted small amounts of IL-17 (data not shown).

To determine whether patients had reactivity with other epitopes of the *Pc*-p27 protein, TEPITOPE was used to predict 2 additional promiscuous peptides derived from the same protein (*Pc*-p27 peptides 2 and 3). The 2 peptides together were predicted to be presented by all 25 HLA-DR molecules in the program, and therefore, these peptides were pooled for testing. PBMC from 14 of the 40 patients (35%) secreted levels of IFN-γ to peptides 2 and 3 that were >3 SD above the mean value of HC (P=0.0007) or patients with LA (P=0.006) (Figure 1C). Altogether, PBMC from 24 of the 40 patients (60%) had reactivity with 1 or more of the 3 *P. copri* peptide sequences, showing that Th1 immune responses to this protein were common in NORA patients.

Because of the importance of anti-citrullinated protein antibodies (ACPA) in the diagnosis and pathogenesis of RA (1,23), peptide 1 was re-synthesized with a citrulline in place of the only arginine in the peptide, which was predicted to be in the – P1 flanking position of the HLA-DR binding pocket. However, the results suggested that the *Pc*-p27 signal peptide sequence was probably not citrullinated *in vivo* (supplementary Figure 1). This does not preclude citrullination of other parts of the protein, including B cell epitopes.

IgG and IgA antibody responses to *Pc*-p27 and whole *P. copri*. We next determined antibody responses to *Pc*-p27 in serum samples from 303 individuals. These included samples from 127 patients with new-onset or chronic RA (CRA), 28 patients with connective tissue diseases (14 with systemic lupus, 4 with mixed connective tissue disease, 4 with scleroderma, and 6 with Sjogren's syndrome), 28 patients with spondyloarthropathy (SpA) (15 with psoriatic

arthritis, 10 with ankylosing spondylitis, and 3 with reactive arthritis), 70 patients with Lyme arthritis, and 50 healthy subjects.

Of the 78 NORA patients, 10 (13%) had IgG antibody responses to *Pc*-p27 that were >3 SD above those in HC ($P<0.0001$) (Figure 2A). Moreover, 10 of 49 patients (20%) with CRA had IgG antibody responses to the protein ($P<0.0001$), including patient RA1 in whom 4 serial samples obtained 4 to 9 years after disease onset yielded positive results. In contrast, only 1 patient with SpA and 1 healthy subject had borderline positive IgG antibody responses to the protein.

Because the first interactions between *P. copri* and immune cells would presumably occur in the gut mucosa, we also determined IgA antibody responses to the organism. About 10% of the patients in both the NORA and CRA groups had IgA antibody responses to *Pc*-p27 ($P\leq0.0002$ and $P\leq0.02$, respectively), and the responses tended to be more robust in NORA patients (Figure 2B). In contrast, only 1 patient with LA and 1 healthy subject had IgA antibody reactivity with the protein. Except for 2 RA patients who had both IgG and IgA responses to *Pc*-p27, the other *Pc*-p27-positive patients had either an IgG or IgA response, but not both. Altogether, 24% of the 127 RA patients had IgG or IgA antibody responses to *Pc*-p27.

When both T and B cell responses were considered together, 3 of the 24 patients who had T cell reactivity with *Pc*-p27 peptides also had IgG *Pc*-p27 antibody responses, but none had IgA responses to the protein. In comparison, among 16 patients lacking T cell reactivity to *Pc*-p27 peptides, only 1 had an IgG antibody response to the protein, but 5 had IgA responses ($P=0.05$). The frequencies of "shared epitope" alleles in patients with *P. copri* T or B cell responses was not significantly different than those in patients who lacked these responses (data not shown).

In an effort to confirm these findings, we determined IgG and IgA antibody responses to whole *P. copri* using the same set of 303 serum samples. Using PCR, we confirmed that this strain expressed *Pc*-p27 (data not shown). Six of the 78 NORA patients (8%) and 5 of the 49 CRA patients (10%) had IgG antibody responses to *P. copri* (Figure 2C). Similarly, 6 of 78 NORA patients (8%) had IgA antibody responses to *P. copri* ($P \leq 0.004$), and 7 of 49 patients (14%) with CRA had elevated IgA antibody levels to the organism ($P \leq 0.002$) (Figure 2D). Among the 19 patients who had positive IgG or IgA responses to *P. copri*, only 5 (26%) had both responses. No patient with CTD, SpA, or LA had IgG or IgA antibodies to the organism. Altogether, 15% of 127 RA patients had *P. copri* IgG or IgA antibody responses.

When the antibody responses to whole *P. copri* or recombinant *Pc*-p27 were combined, 41 (32%) of the 127 RA patients had IgG or IgA antibody reactivity with the organism. Thus, antibody responses to *P. copri* were common in RA patients, both early and late in the disease, yet they were rarely found in patients with other types of arthritis, implying specificity in RA.

Antibody responses to other oral or commensal bowel flora. To examine the specificity of antibody responses to *P. copri* in RA patients, the same serum samples were also tested for reactivity with whole *P. gingivalis*, an oral periodontal pathogen implicated in RA (24), and with 2 common gut commensal organisms, *Bacteroides fragilis* and *Escherichia coli*.

Similar to previous reports (25,26), IgG antibody responses to *P. gingivalis* were found in about 25% of our NORA and CRA patients, and these responses tended to be higher in RA patients than in the other comparison groups (Figure 3A). However, in contrast with *P. copri* antibody responses, IgG antibodies to *P. gingivalis* were also present in small percentages of patients with other rheumatic diseases or healthy control subjects, and IgA antibody reactivity with *P. gingivalis* was not increased in RA patients compared with the other groups (Figure 3B).

Moreover, in contrast with the dichotomy between IgG and IgA antibody responses to *P. copri*, all RA patients with IgA antibodies to *P. gingivalis* also had IgG responses to this microbe. Importantly, 66 of the 127 RA patients (52%) had antibody responses to either *P. copri* or *P. gingivalis*, but only 8 of the 66 patients (12%) had antibody responses to both microbes. Thus, minimal overlap was observed in the antibody responses to these 2 organisms, indicating that these responses are largely independent, and only the response to *P. copri* was specific for RA.

Very few RA patients or those with other rheumatic diseases had IgG or IgA antibody responses to *B. fragilis* or *E. coli* that were >3 SD above the mean values in healthy control subjects (Figures 3C-F). However, IgG absorbance values for *B. fragilis* were significantly lower in NORA patients than in patients in the other groups ($P \leq 0.03$) (Figure 3C), consistent with the decrease in *B. fragilis* abundance noted previously in NORA patients (8). Conversely, IgG and IgA absorbance values to *B. fragilis* in the CTD group were significantly higher than those in the other groups. Thus, in contrast with *P. copri* antibody responses, antibody levels to *B. fragilis* and *E. coli* were similar or less in RA patients than those in patients with other types of arthritis or in healthy subjects.

Clinical parameters according to *P. copri* or *P. gingivalis* antibody responses. Because *P. copri* and *P. gingivalis* have both been implicated in RA, clinical parameters in our cohorts of NORA and CRA patients were compared in patients who did or did not have IgG or IgA antibody responses to these microbes. Because the findings were similar in NORA and CRA patients, these groups were combined for presentation here.

Several significant differences were found between these groups (Table 1). First, 45 of the 50 patients (90%) with *P. copri* antibody responses were female compared with 60 of the 86 patients (70%) who lacked such responses (P=0.006). In contrast, the percentages of female and

male patients were not significantly different in those with *P. gingivalis* antibody responses. Second, only 37% of the RA patients with IgG *P. copri* antibody responses had ACPA compared with 74% of those who had IgA *P. copri* antibodies (P=0.01) and 72% who lacked *P. copri* antibodies (P=0.003). There was a similar trend with RF (P=0.08). In contrast, patients with *P. gingivalis* IgG and IgA antibody responses had higher frequencies of ACPA and RF than patients who lacked these responses. Finally, at study entry, there was a trend toward higher disease activity scores (DAS28) in NORA patients with either *P. copri* or *P. gingivalis* IgG antibody responses. There were no significant differences among the groups in age, body mass index (BMI), or smoking history. Thus, *P. copri* antibodies were found primarily in women and ACPA were less common in patients with IgG *P. copri* antibodies, whereas neither of these factors correlated with *P. gingivalis* antibody responses, again indicating that these microbes induce distinct responses.

Correlation of *P. copri* antibodies with serum cytokine and chemokine levels. In an effort to link *P. copri* with inflammatory responses and autoantibody production, IgG and IgA *P. copri* antibody values were correlated with serum cytokine levels in 120 of the 127 RA patients in whom sufficient serum samples were still available. The 14 cytokines and chemokines measured were representative of innate, Th1 and Th17 immune responses. The assays were performed with Heteroblock to limit the possible interference by rheumatoid factor. Because the results were similar in NORA and CRA patients, they were combined for presentation here.

When the magnitude of *P. copri* IgG or IgA antibodies in the 37 antibody-positive patients were correlated with cytokine levels, strong positive correlations were found between IgA antibody values and the levels of 3 innate cytokines (IFN-α, MIP-1α, and MIP-1β), 2 Th1-

associated cytokines (IFN-γ and IL-12), and 3 Th17-associated cytokines (IL-17F, IL-22, and IL-1β) (Figure 4A-C). In contrast, IgG *P. copri* antibody values correlated only with levels of the Th1 chemoattractant CXCL10 (Figure 4B).

When *P. copri* IgA absorbance values in all 120 RA patients, including those with positive and negative values, were correlated with cytokine levels, even stronger associations were found with innate (MIP-1α and MIP-1β), Th1 (IFN-γ and IL-12), and Th17 cytokines (IL-23, IL-22, IL-17A, IL-17E, and IL-17F) (Supplementary Figure 2A). In contrast, IgG absorbance values did not correlate with any cytokine or chemokine levels. Similarly, *P. gingivalis* IgG and IgA antibody did not correlate with any cytokine or chemokine level (Supplementary Figure 2B), further indicating that these microbes induce distinct responses at different mucosal sites.

Detection of *P. copri* 16S rDNA in synovial fluid. Because *P. copri* IgG antibody responses were suggestive of a systemic immune response to the organism, we investigated whether *P. copri* itself may spread to joints. For this purpose, we designed nested PCR primers to detect DNA for the 16S ribosomal RNA gene of *P. copri* (16S rDNA) in patients' samples. Of 18 patients in whom paired serum and SF samples were available, 10 were obtained from NORA patients, and 8 were collected from CRA patients who were seen from 3-to-50 years after disease onset. Five of the 18 patients had IgG antibody responses to *P. copri*; 2 had IgA antibody reactivity with the organism, and 11 did not have *P. copri* antibody responses. Although the numbers were small, the clinical correlations in each of these groups, including disease duration, DAS28 scores, ACPA and RF frequencies, were similar to those in the larger groups of RA patients (Supplementary Table 1).

Of the 5 patients (RA1 to RA5) with IgG *P. copri* antibodies, 3 had *P. copri* 16S rDNA detected in SF. In 2 of the 3 patients, 16S rDNA was found in samples obtained prior to

DMARD therapy. In the third patient (CRA patient RA1), in whom the original HLA-DR-presented *P. copri* peptide was identified in PBMC collected 7 years after disease onset (Figure 1A), *P. copri* 16S rDNA was detected in SF obtained 9 years after disease onset (Figure 5A). In contrast, *P. copri* DNA was not detected in SF samples from the remaining 13 patients, 2 with IgA *P. copri* antibodies and 11 without *P. copri* antibodies. Serum samples from all 18 patients had negative PCR results. Thus, 3 of the 5 patients with IgG *P. copri* antibodies had positive PCR results for *Prevotella* DNA in SF compared with none of 13 patients in the other 2 groups (P=0.01).

For comparison, nested PCR primers were designed to detect 16S rDNA from *B. fragilis*, another gut commensal. Of the 18 patients, one (RA15) had *B. fragilis* 16S rDNA in SF; the sequence of this amplicon had 98% homology to *B. fragilis* 16S rDNA (Supplementary Figure 3). This patient did not have positive tests for *P. copri* DNA or *P. copri* antibody responses, and none of the 18 patients had IgG or IgA responses to *B. fragilis*.

All positive results were confirmed by sequencing. Amplicons from patient RA2 had 100% identity with the sequence for *P. copri* 16S rDNA in the NCBI database, which was obtained from a Japanese isolate. Interestingly, patient RA2 grew up in Korea. Amplicons from patient RA1 and from patient RA5 annealed with *P. copri* sequences (DSM18205), but they had 86% and 89% sequence homology, respectively (Figure 5B), suggesting that the sequence could be from another *Prevotella* species (Figure 5B). Alternately, even though cultures of *P. copri* from stool were not available for analysis, the differences in *P. copri* 16S rDNA sequences may be explained by strain variation since the 3 patients grew up in widely different geographic locations (Korea, the United States, or the Caribbean Islands). This is consistent with a study in

which *P. copri* was ranked as the second-most variable member of the human gut microbiota between continents (27).

DISCUSSION

In this study, using a discovery-based search for HLA-DR-presented peptides derived from *P. copri*, one spectrum-to-peptide match was identified from the PBMC of 1 RA patient. This peptide sequence was found in the signal domain of a 27-kD protein of *P. copri*, which was predicted to be a secreted protein. Signal sequences, which are cleaved prior to secretion, can accumulate in transmembrane locations; they are often highly antigenic, and they typically bind many different HLA-DR molecules (28). In addition, the secreted portion of the protein may become an immunogenic soluble antigen (6,29). We then found that the signal sequence peptide and two other peptides from the *Pc*-p27 protein induced Th1 responses in 60% of patients with RA. Although RA SE alleles correlated inversely with *P. copri* overexpansion in the gut (8), our study showed no significant correlations between these T cell epitopes and SE alleles, consistent with the fact that these epitopes are promiscuous HLA-DR binders.

Whereas PBMC were only available in NORA patients, we were able to test for antibody responses to *P. copri* in both NORA and CRA patients. Although over-expansion of *P. copri* was previously detected only in the stool of NORA patients (8), we found similar frequencies of *P. copri* antibody responses and similar clinical associations in NORA and CRA patients, suggesting that once initiated, these antibody responses may persist for years. Importantly, in both patient groups, *P. copri* antibody responses were specific for RA. First, *P. copri* antibodies were rarely found in patients with other rheumatic diseases or in healthy controls. Second, antibody levels to 2 other gut commensals, *B. fragilis* and *E. coli*, were similar or less in RA

patients than those in patients with other arthritides or in healthy subjects. Third, there was little overlap between patients with *P. gingivalis* antibodies and those with *P. copri* antibodies.

In both NORA and CRA patients, *P. copri* antibody responses were found primarily in women. When the 2 cohorts of patients were combined, *P. copri*-negative patients had a sex ratio of 2.3 to 1 in favor of women, which is close to the ratio of 3 to 1 generally reported in RA cohorts. In contrast, the patients with *P. copri* antibody responses had a sex ratio of 9 to 1, implying that this organism may be a substantial contributor to the female predominance in RA.

In mice, gut symbiotic gram-negative bacteria may disseminate systemically to other sites and induce IgG antibody responses (30). These homeostatic IgG responses may help later in protection against invasion by pathogenic gram-negative bacteria. Our studies showed a high background in *E. coli* and *B. fragilis* antibody assays, perhaps reflecting positive responses to these bacteria in some individuals. However, antibody responses to these bacteria were not greater in RA patients than in other comparison groups, including healthy controls, whereas *P. copri* antibody responses were significantly higher in RA patients and correlated with inflammatory cytokine levels. Thus, our findings could not be explained simply by homeostatic *P. copri* antibody responses.

In our study, the magnitude of IgA *P. copri* antibody responses in RA patients correlated with serum levels of a range of cytokines and chemokines associated with innate, Th1 and Th17 immune responses. Although we did not find *Pc*-p27-specific Th17 cells in PBMC of these patients, it is likely that such cells were present in their intestinal mucosa. Th17 responses are presumably important initially in containing the organism in the bowel, but with chronic antigenic stimulation they can promote systemic inflammation and autoimmunity (31). This idea is supported here by strong correlations among IgA antibody values, serum levels of Th17-

associated cytokines, and high frequencies of ACPA, which could react with citrullinated proteins in joints (32).

Conversely, other RA patients had IgG *P. copri* antibody responses associated with *Prevotella* DNA in joints, inflammatory Th1 responses, and infrequent ACPA. In these patients, we postulate that *P. copri*, a strict anaerobe, may spread systemically within phagocytic cells, including to joints. In 2 of 3 patients, *Prevotella* DNA was detected in SF obtained early in the disease. However, in the remaining patient (RA1), in whom the *Pc*-p27 HLA-DR-presented peptide was identified in PBMC obtained 7 years after disease onset, *P. copri* DNA was detected in SF collected 9 years after disease onset. This suggests that *P. copri* may spread from the bowel in repeated waves over a period of years, perhaps explaining the persistence of antibody responses to *P. copri* in CRA patients. The finding of 16S rDNA of *B. fragilis* in the SF of one additional patient suggests that escape of other commensal bacteria from the bowel may occur, but this patient lacked an antibody response to the organism.

There are parallels between the potential role of *P. copri* in RA pathogenesis and that of the periodontal pathogen, *Porphyromonas gingivalis* (24,33). In patients with periodontitis, the composition of the oral flora shifts in favor of organisms, particularly anaerobes such as *P. gingivalis*, which thrive in an inflammatory environment (34). IgG antibody responses to *P. gingivalis* in RA patients correlate strongly with systemic inflammation and co-existent periodontal disease (24,26,35). Furthermore, *P. gingivalis* may disseminate from the gingiva, presumably via dendritic cells, also leading to infection and inflammation at distant sites (6,36), and DNA from periodontal bacteria has been detected by PCR in SF of a few RA patients with periodontal disease (37).

In conclusion, our study provides evidence for immune relevance of *P. copri* in RA and suggests that bowel pathology may be a feature of the disease in this subgroup of patients. These new insights are likely to have implications for both the diagnosis and treatment of RA. For example, *P. copri* IgG antibody responses could provide support for the diagnosis in RA patients who lack ACPA or RF. Moreover, dietary interventions or specifically tailored antibiotic regimens targeting *P. copri* could have an adjunctive role to DMARDs in the treatment of this disease. Previously, patients with early RA who received tetracycline or its derivatives had significantly better outcomes than placebo-treated patients (38,39). These results are often attributed to the anti-inflammatory effects of tetracyclines, but recent insights about the microbiota suggest an additional explanation. Importantly, insights about specific immune-relevant commensal organisms in the microbiota promise a new era in the understanding and treatment of RA.

ACKNOWLEDGMENTS

We thank John Branda, Judith M. Holden, Gail McHugh for help with microbial cultures, Dan Littman and Hannah Fehlner-Peach for providing *P. copri* DNA to clone the *Pc*-p27 protein, Chunxiang Yao for help with mass spectrometry analyses, Deborah Collier and Marcy Bolster for help with patient care, Dennis Burke and John Aversa for help in obtaining synovial tissue, Mandakolathur Murali for RF and ACPA analyses, the American Red Cross HLA laboratory for HLA-DR typing of patients, Katherine Sulka for help in the preparation of samples for HLA-DR typing, and the CCIB DNA CORE facility at MGH for sequencing of *P. copri* 16S rDNA amplicons. This work was presented in part at the annual meeting of the American College of Rheumatology in San Francisco 2015 (40).

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content. Dr. Steere had full access to the data in the study and takes responsibility for the integrity of the data and the accuracy of data analysis.

Study conception and design. Pianta, Drouin, Costello, Steere

Acquisition of data. Pianta, Arvikar, Strle, Wang, Steere

Analysis and interpretation of data. Pianta, Arvikar, Strle, Drouin, Wang, Costello, Steere

REFERENCES

1. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. N Engl J Med 2011; 365:2205-19.
2. Firestein GS. Evolving concepts of rheumatoid arthritis. Nature 2003; 423:356-61.
3. Plenge RM. Rheumatoid arthritis genetics: 2009 update. Curr Rheumatol Rep 2009; 11:351-6.
4. Catrina AI, Deane KD, Scher JU. Gene, environment, microbiome and mucosal immune tolerance in rheumatoid arthritis. Rheumatology (Oxford) 2014; 55:391-402.
5. Longman RS, Littman DR. The functional impact of the intestinal microbiome on mucosal immunity and systemic autoimmunity. Curr Opin Rheumatol 2015; 27:381-7.
6. Han YW, Wang X. Mobile microbiome: oral bacteria in extra-oral infections and inflammation. J Dent Res 2013; 92:485-91.
7. Scher JU, Abramson SB. The microbiome and rheumatoid arthritis. Nat Rev Rheumatol 2011; 7:569-78.

8. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C et al. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife 2013; 2:e01202.
9. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y et al. Induction of colonic regulatory T cells by indigenous Clostridium species. Science 2011; 331:337-41.
10. Round JL, Lee SM, Li J, Tran G, Jabri B, Chatila TA et al. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science 2011; 332:974-7.
11. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med 2015; 21:895-905.
12. Maeda Y, Kurakawa T, Umemoto E, Motooka D, Ito Y, Gotoh K et al. Dysbiosis contributes to arthritis development via activation of autoreactive T cells in the intestine. Arthritis Rheumatol 2016; doi:10.1002/art.39783.
13. Seward RJ, Drouin EE, Steere AC, Costello CE. Peptides presented by HLA-DR molecules in synovia of patients with rheumatoid arthritis or antibiotic-refractory Lyme arthritis. Mol Cell Proteomics 2011; 10:M110 002477.
14. Wang Q, Drouin EE, Yao C, Zhang J, Huang Y, Leon DR et al. Immunogenic HLA-DR-presented self-peptides identified directly from clinical samples of synovial tissue, synovial fluid, or peripheral blood in patients with rheumatoid arthritis or Lyme arthritis. J Proteome Res 2016 (Epub ahead of print).
15. Crowley JT, Strle K, Drouin EE, Pianta A, Arvikar SL, Wang Q et al. Matrix metalloproteinase-10 is a target of T and B cell responses that correlate with synovial

pathology in patients with antibiotic-refractory Lyme arthritis. J Autoimmun 2016; 69:24-37.

16. Crowley JT, Drouin EE, Pianta A, Strle K, Wang Q, Costello CE et al. A highly expressed human protein, apolipoprotein B-100, serves as an autoantigen in a subgroup of patients with Lyme disease. J Infect Dis 2015; 212:1841-50.

17. Drouin EE, Seward RJ, Strle K, McHugh G, Katchar K, Londono D et al. A novel human autoantigen, endothelial cell growth factor, is a target of T and B cell responses in patients with Lyme disease. Arthritis Rheum 2013; 65:186-96.

18. Pianta A, Drouin EE, Crowley JT, Arvikar S, Strle K, Costello CE et al. Annexin A2 is a target of autoimmune T and B cell responses associated with synovial fibroblast proliferation in patients with antibiotic-refractory Lyme arthritis. Clin Immunol 2015; 160:336-41.

19. Pianta A, Drouin EE, Wang Q, Arvikar S, Costello CE, Steere AC. Identificatiom of N-acetylglucosamine-6-sulfatase and filamin A as novel targets of autoimmune T and B cell responses in rheumatoid arthritis [abstract]. Ann Rheum Dis 2015; 74(Suppl2): 112.

20. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011; 8:785-6.

21. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 1999; 17:555-61.

22. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd et al. 2010 Rheumatoid arthritis classification criteria: an American College of

Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010; 62:2569-81.

23. Brink M, Hansson M, Mathsson L, Jakobsson PJ, Holmdahl R, Hallmans G et al. Multiplex analyses of antibodies against citrullinated peptides in individuals prior to development of rheumatoid arthritis. Arthritis Rheum 2013; 65:899-910.

24. Mikuls TR, Payne JB, Yu F, Thiele GM, Reynolds RJ, Cannon GW et al. Periodontitis and Porphyromonas gingivalis in patients with rheumatoid arthritis. Arthritis Rheumatol 2014; 66:1090-100.

25. de Smit M, Westra J, Vissink A, Doornbos-van der Meer B, Brouwer E, van Winkelhoff AJ. Periodontitis in established rheumatoid arthritis patients: a cross-sectional clinical, microbiological and serological study. Arthritis Res Ther 2012; 14:R222.

26. Arvikar SL, Collier DS, Fisher MC, Unizony S, Cohen GL, McHugh G et al. Clinical correlations with Porphyromonas gingivalis antibody responses in patients with early rheumatoid arthritis. Arthritis Res Ther 2013; 15:R109.

27. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A et al. Genomic variation landscape of the human gut microbiome. Nature 2013; 493:45-50.

28. Kovjazin R, Volovitz I, Daon Y, Vider-Shalit T, Azran R, Tsaban L et al. Signal peptides and trans-membrane regions are broadly immunogenic and have high CD8+ T cell epitope densities: Implications for vaccine development. Mol Immunol 2011; 48:1009-18.

29. Yang Y, Torchinsky MB, Gobert M, Xiong H, Xu M, Linehan JL et al. Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. Nature 2014; 510:152-6.

30. Zeng MY, Cisalpino D, Varadarajan S, Hellman J, Warren HS, Cascalho M et al. Gut microbiota-induced immunoglobulin G controls systemic infection by symbiotic bacteria and pathogens. Immunity 2016; 44:1-12.

31. Ruff WE, Kriegel MA. Autoimmune host-microbiota interactions at barrier sites and beyond. Trends Mol Med 2015; 21:233-44.

32. Romero V, Fert-Bober J, Nigrovic PA, Darrah E, Haque UJ, Lee DM et al. Immune-mediated pore-forming pathways induce cellular hypercitrullination and generate citrullinated autoantigens in rheumatoid arthritis. Sci Transl Med 2013; 5:209ra150.

33. Pischon N, Pischon T, Kroger J, Gulmez E, Kleber BM, Bernimoulin JP et al. Association among rheumatoid arthritis, oral hygiene, and periodontitis. J Periodontol 2008; 79:979-86.

34. Hajishengallis G. Periodontitis: from microbial immune subversion to systemic inflammation. Nat Rev Immunol 2015; 15:30-44.

35. Scher JU, Abramson SB. Periodontal disease, Porphyromonas gingivalis, and rheumatoid arthritis: what triggers autoimmunity and clinical disease? Arthritis Res Ther 2013; 15:122.

36. Carrion J, Scisci E, Miles B, Sabino GJ, Zeituni AE, Gu Y et al. Microbial carriage state of peripheral blood dendritic cells (DCs) in chronic periodontitis influences DC differentiation, atherogenic potential. J Immunol 2012; 189:3178-87.

37. Martinez-Martinez RE, Abud-Mendoza C, Patino-Marin N, Rizo-Rodriguez JC, Little JW, Loyola-Rodriguez JP. Detection of periodontal bacterial DNA in serum and synovial fluid in refractory rheumatoid arthritis patients. J Clin Periodontol 2009; 36:1004-10.

38. O'Dell JR, Haire CE, Palmer W, Drymalski W, Wees S, Blakely K et al. Treatment of early rheumatoid arthritis with minocycline or placebo: results of a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 1997; 40:842-8.

39. Stone M, Fortin PR, Pacheco-Tena C, Inman RD. Should tetracycline treatment be used more extensively for rheumatoid arthritis? Metaanalysis demonstrates clinical benefit with reduction in disease activity. J Rheumatol 2003; 30:2112-22.

40. Pianta A, Drouin EE, Arvikar S, Strle K, Crowley JT, Wang Q, Costello CE, Steere AC. Identification of a broadly immunogenic Prevotella copri T cell epitope in patients with rheumatoid arthritis [abstract]. Arthritis Rheumatol 2015; 67 (suppl 10).

Table 1 Demographic and clinical finding in patients with RA according to *P. copri* and *P. gingivalis* antibodies[a]

	***P. copri* Antibody Responses**			***P. gingivalis* Antibody Responses**		
	IgG (N=27)	IgA (N=23)	None (N=86)	IgG (N=33)	IgA (N=12)	None (N=94)
Age, years (range)	47 (19-85)	53 (19-71)	57 (23-84)	58 (21-84)	62 (41-74)	54 (19-85)
Sex, female/male	**24/3ǂ**	**21/2ǂ**	60/26	23/10	7/5	73/21
Body mass index	27	28	27	29	34	27
Smoking history						
Current, former, n (%)	9 (33)	9 (39)	38 (44)	13 (39)	6 (50)	39 (41)
Never, n (%)	18 (67)	14 (61)	48 (56)	20 (61)	6 (50)	55 (59)
Autoantibodies						
RF, n positive (%)	9 (33)	12 (52)	46 (53)	20 (61)	8 (67)	43 (46)
ACPA, n positive (%)	**10 (37)****	17 (74)	61 (71)	26 (79)	**11 (92)***	58 (62)
Inflammatory markers						
ESR, median (range)	12.9 (1-84)	17 (4-34)	16 (1-111)	19.2 (1-67)	**36*** (10-67)	15.4 (1-111)
CRP, median (range)	8.5 (0.2-70)	4.5 (0.1-15)	8.2 (0.1-87)	9.2 (0.1-68)	7.3 (0.1-66)	6.8 (0.1-87)
Disease activity, median						
DAS28ESR	5.2	4.4	4.2	5.2	5	4.5
DAS28CRP	4.6	3.8	3.9	4.9	5	4

[a]Data are expressed as median values unless otherwise noted. RF, rheumatoid factor; ACPA, anti-citrullinated protein antibodies; ESR, erythrocyte sedimentation rate, mm/h; CRP, C-reactive protein, mg/L; DAS28, Disease Activity Score based on 28-joint count. Value reported in bold are significantly different from the patients without microbial antibody responses (None). ESR, CRP and DAS28 score were available only for NORA patients. ǂRA patients with IgG and IgA antibody responses vs None patients, P=0.006. **P=0.003, *P≤0.05.

FIGURE LEGEND

Figure 1. Identification of a Broadly Immunogenic *P. copri* T Cell Epitope. (A) LC/MS/MS spectrum of the *Pc*-p27$_{2-20}$ peptide. Consensus peptide identification was achieved as KRIILILTVLLAMLGQ(deamidated)VAY by OMSSA and X!Tandem. The insert panel shows the IFN-γ ELISpot assay using matching patient's PBMC stimulated with the peptide (1, 2 and 4 μM). Reactivity of >3 times background (no antigen) was considered positive. (B) IFN-γ ELISpot assay using PBMC from patients with rheumatoid arthritis (RA), Lyme arthritis (LA), and healthy controls (HC) incubated with the HLA-DR-presented peptide identified from the PBMC of patient RA1 (Peptide1, 1 μM). (C) IFN-γ secretion of PBMC from patients and controls incubated with 2 predicted promiscuous HLA-DR binding peptides from *Pc*-p27 (1 μM each). A positive response was defined as >3 SD above the mean value of the HC (area above the shaded region). The value for patient RA1 is indicated with a star. Horizontal lines represent the mean values of each group.

Figure 2. IgG and IgA Responses to *P. copri* in RA Patients and Control Subjects. Serum samples from 303 individuals with NORA, CRA, other rheumatic diseases or from healthy controls were tested for *P. copri* antibodies. IgG (A) and IgA (B) ELISAs against the bacterial protein *Pc*-p27; IgG (C) and IgA (D) ELISA against 1% formalin-inactivated *P. copri*. Positivity was defined as >3 SD above the mean value of healthy controls (area above the shaded region). Symbols represent values in individual patients and horizontal lines show mean values. The value for patient RA1 is indicated with a star. Only significant P values relative to HC are shown.

HC, healthy control; CTD, connective tissue diseases; SpA, spondyloarthropathies; LA, Lyme arthritis; NORA, new onset rheumatoid arthritis; CRA, chronic rheumatoid arthritis.

Figure 3. IgG and IgA Responses to Other Organisms in RA Patients and Controls. Serum samples from the same 303 individuals, shown in figure 2, were tested for antibody responses to other bacteria. IgG (A) and IgA (B) ELISAs against 1% formalin-inactivated *P. gingivalis*. Positivity was defined as >2SD above the mean value of normal control negative for periodontitis as determined in our previous publication (26). IgG (C) and IgA (D) ELISAs against 1% formalin-inactivated *B. fragilis*; IgG (E) and IgA (F) ELISAs against 1% formalin-inactivated *E. coli*. For these analyses, positivity was defined as >3 SD above the mean value of healthy controls (area above the shaded region). Symbols represent values in individual patients and horizontal lines show mean values. Only significant P values are shown. HC, healthy control; CTD, connective tissue diseases; SpA, spondyloarthropathies; LA, Lyme arthritis; NORA, new onset rheumatoid arthritis; CRA, chronic rheumatoid arthritis.

Figure 4. Correlation between Cytokine Levels and *P. copri* Antibody Responses in Antibody-Positive Patients. Correlation between *P. copri* specific IgG or IgA responses and cytokines associated with innate immunity (A), Th1 immunity (B), or Th17 immunity (C). Correlations were performed using the Pearson's correlation test.

Figure 5. PCR Testing for *Prevotella copri* 16S rDNA in Concomitant Serum and SF Samples of RA Patients. (A) Nested PCR of *P. copri* 16S rDNA amplicons (254 bp) analyzed on 1.5% agarose gels stained with ethidium bromide. Results from serum and SF samples from

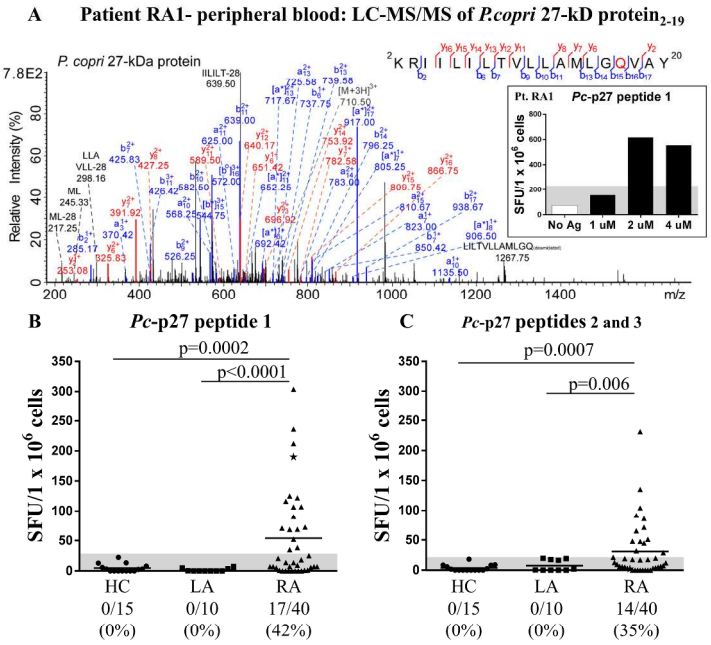
the 3 positive patients are shown. Patient RA1 had 4 paired serum and SF samples; patient RA2 had 1 serum and 2 SF samples, and patient RA5 had 1 serum and 1 SF samples. In patients RA2 and RA5, enough material was available for testing in duplicate. M, 100 bp DNA ladder; +, positive control (*P. copri* DSM 18205); H, water control. (B) Sequence alignment of the 16S gene amplicons obtained from patient RA1, RA2 and RA5 using CLC Genomic Workbench software. The sequence of *P. copri* (DSM 18205) 16S gene is shown as the reference and the conservation of all sequence positions is shown below the alignment.

Supplementary Figure 1. **T cell reactivity to peptide 1 with or without citrullination.** (A) IFN-γ secretion of PBMC from patients and controls stimulated with peptide 1 with and without citrullination. (B) Comparison of T cell reactivity between citrullinated and uncitrullinated peptide 1 in RA patients, with values in the same patient connected by a line. SFU = spot forming units.

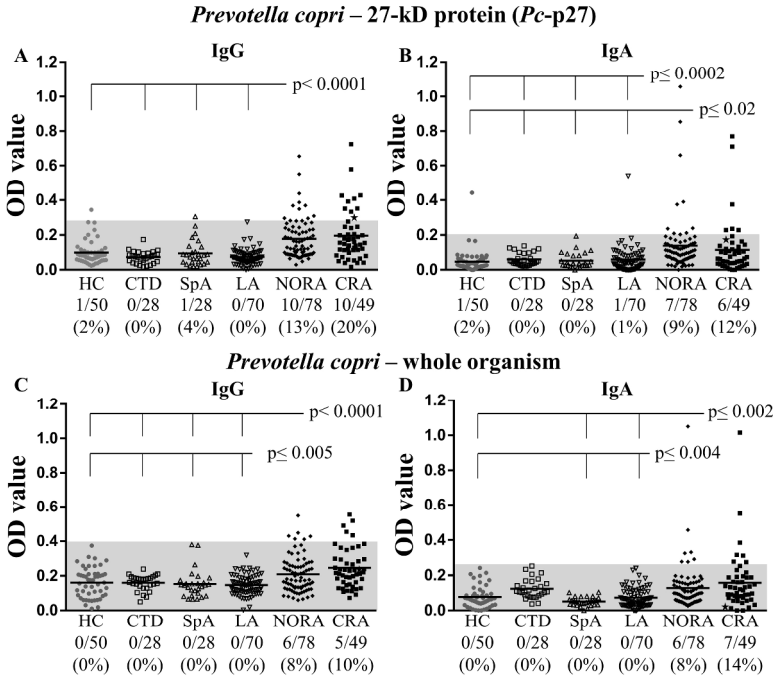
Supplementary Figure 2. **Correlation between Cytokine Levels and *P. copri* or *P. gingivalis* IgA Values in All RA Patients.** Correlation between innate, Th1, or Th17 cytokines levels and *P. copri* (A) or *P. gingivalis* (B) specific IgA responses in all 120 RA patients, including those with positive or negative absorbance values. Correlations were performed using the Pearson's correlation test.

Supplementary Figure 3. PCR Testing for *B. fragilis* 16S rDNA in SF Samples of RA Patients. (A) Nested PCR of *B. fragilis* 16S rDNA amplicons (266 bp) analyzed on 1.5%

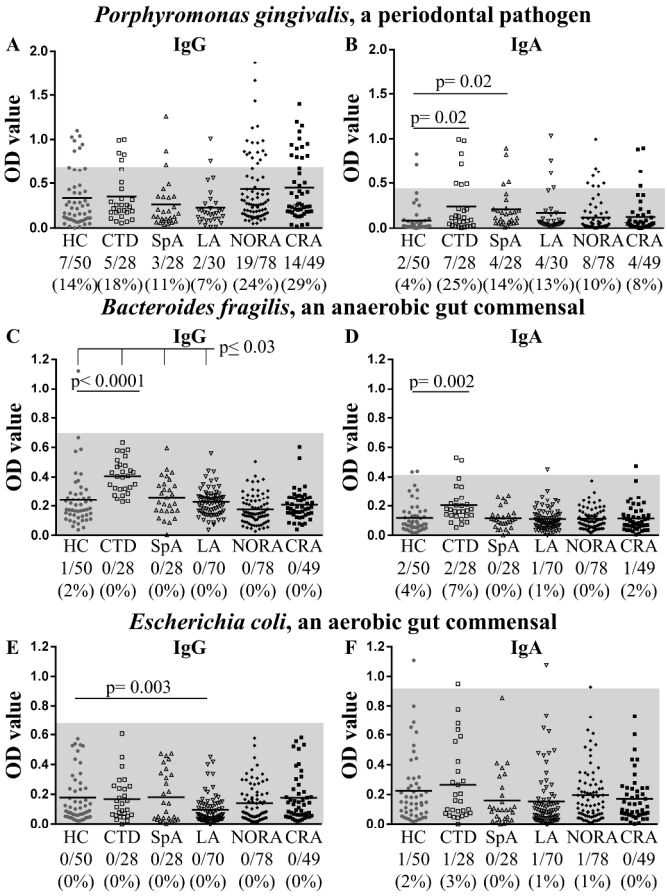
agarose gels. All patients had only one SF sample available, except patient RA2 who had 2 SF samples. M, 100 bp DNA ladder; +, positive control (*B. fragilis* DNA); H, water control. (B) Sequence alignment of the 16S gene amplicon obtained from patient RA15 using CLC Genomic Workbench software. The sequence of *B. fragilis* 16S gene is shown as the reference.



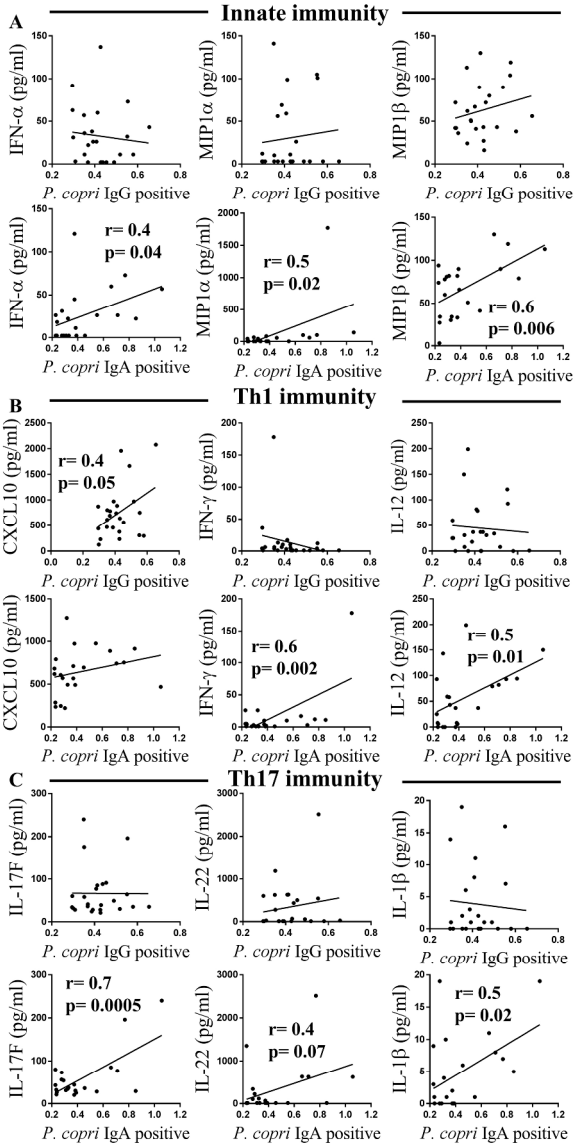
254x338mm (300 x 300 DPI)



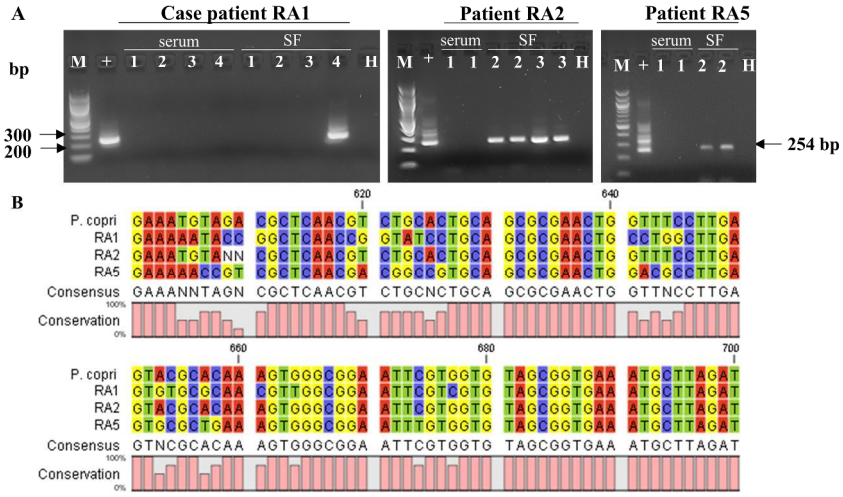
254x338mm (300 x 300 DPI)



254x338mm (300 x 300 DPI)



254x338mm (300 x 300 DPI)



254x338mm (300 x 300 DPI)