This is a repository copy of *Perturbations of the gut microbiome in anti-CCP positive individuals at risk of developing rheumatoid arthritis*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/167767/

Version: Accepted Version

Article:

Rooney, CM, Mankia, K, Mitra, S orcid.org/0000-0002-9378-1496 et al. (3 more authors) (2020) Perturbations of the gut microbiome in anti-CCP positive individuals at risk of developing rheumatoid arthritis. Rheumatology. ISSN 1462-0324

https://doi.org/10.1093/rheumatology/keaa792

© The Author(s) 2020. Published by Oxford University Press on behalf of the British Society for Rheumatology. This is an author produced version of a journal article published in Rheumatology. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Perturbations of the gut microbiome in anti-CCP positive individuals at risk of developing rheumatoid arthritis.

Authors: Christopher M Rooney[1,2,*]; Kulveer Mankia [2,3]; Suparna Mitra [1]; Ines B Moura [1]; Paul Emery [2,3,]: Mark H Wilcox [1].

Affiliations:

[1]Leeds Institute of Medical Research, University of Leeds, Microbiology, The Old Medical School, Leeds General Infirmary, Leeds, UK

[2]Leeds Institute of Rheumatic and Musculoskeletal Medicine, Chapel Allerton Hospital, Leeds, UK

[3]NIHR Leeds Musculoskeletal Biomedical Research Unit, Chapel Allerton Hospital, Leeds.

*Corresponding author Dr Christopher Rooney, Leeds Institute of Medical Research, Microbiology, The Old Medical School, Leeds General Infirmary, Leeds LS1 3EX; christopherrooney@nhs.net

Running title: Gut microbiome in CCP positive individuals

Abstract

Objective: Individuals with newly diagnosed rheumatoid arthritis (RA) have a distinct microbiome when compared with healthy controls. However, little is known as to when these microbiome perturbations begin. Using a prospective at-risk cohort of individuals positive for anti-citrullinated protein (anti-CCP) antibody with new onset musculoskeletal symptoms, but without clinical arthritis, we investigated for the presence of a gut dysbiosis before the onset of RA.

Methods: The gut microbiota of 25 anti-CCP positive individuals without clinical synovitis were sequenced targeting the V4 region of the 16S rRNA gene. Using a publicly available database, a control population of 44 individuals, approximately matched in age, gender, diet and ethnicity was selected for comparison, using the same sequencing methodology. Median interval between sample collection and progression to RA was 188 days. Taxonomic analysis was performed using QIIME and MEGAN, and statistical analysis using R software.

Results: There were significant differences (p=0.01) at family level in gut microbiomes of anti-CCP positive individuals versus controls. The anti-CCP positive population had an overabundance of *Lachnospiraceae*, *Helicobacteraceae*, *Ruminococcaceae*, *Erysipelotrichaceae* and *Bifidobacteriaceae*, amongst others. Five individuals progressed to RA between sample collection and analysis. Clustering of the progressor population was observed on a phylogenetic network created using a probabilistic similarity index (Goodall's index).

Conclusions: Anti-CCP positive at-risk individuals without clinical synovitis appear to have a distinct gut microbiome compared to healthy controls. Phylogenetic clustering was observed in individuals who progressed to RA, suggesting that distinct taxa are associated with the development of RA many months before its onset.

Key messages

- Phylogenetic clustering of CCP+progressors suggests microbial dysbiosis predates the onset of RA by many months.
- Preserved microbial diversity between CCP+progressors and CCP+non-progressors suggests diversity decreases with progression along RA spectrum.

Key words: Gut Microbiome, Pre-Rheumatoid Arthritis, 16S rRNA Sequencing, Microbial Dysbiosis

1.Introduction

Established rheumatoid arthritis (RA) is now recognised as the end point of a disease continuum, encompassing a preclinical phase where genetic and environmental factors interact, to initiate autoimmunity (1-3). The production of rheumatoid factor (RF) and more specifically anti-citrullinated protein (anti-CCP) antibody mark the presence of systemic autoimmunity (4), which may occur years before the development of clinical synovitis and RA (1). The presence of circulating immunoglobulin A (IgA) anti-CCP prior to joint inflammation in individuals at-risk of RA suggests that the joints are the target rather than the instigator of systemic auto-antibody production (5). Additionally, serum enrichment of IgA points to a mucosal driver in RA-related autoimmunity. As the gut mucosa receives a constant stream of antigenic stimulation from resident microbes, it is conceivable that a delicate innate immune homeostasis has evolved within the gut (6). Aberration of this symbiotic state, triggered by alterations of bacterial communities (dysbiosis), can lead to immune dysregulation (7). Interestingly, patients with established RA have been reported to have dysbiotic microbiomes when compared with those of healthy controls (8-15), although the detailed bacterial changes in such dysbiosis is still under investigation. Furthermore, the question remains as to whether the dysbiotic changes observed are a cause or a consequence of the underlying disease process.

Anti-CCP positivity combined with new arthralgia can be used to delineate those at the highest risk of RA progression (5). Thus, this at-risk cohort provides a unique opportunity to investigate RA disease before its onset. We have investigated using 16S rRNA sequencing whether there are common changes in gut microbiomes of anti-CCP positive patients, particularly in those who progress to RA onset.

2. Methods

2.1 Patient and public involvement

Prior to commencement of this research patients were involved in the design and feasibility of this study, informing aspects of recruitment, sample return and dissemination of outputs using the patient focused discussion group at Chapel Allerton Hospital, Leeds. During the study period participant feedback was sough regarding the stool collection kit which was subsequently refashioned.

2.2 Patient selection

Twenty-five anti-CCP positive individuals with nonspecific musculoskeletal symptoms and without clinical evidence of synovitis on examination were included in this study. Participants were selected from the Leeds CCP ‘at-risk’ cohort which has been previously described (16, 17). Briefly, this cohort is recruited nationally via primary and secondary care referrals of individuals presenting with new onset non-specific musculoskeletal symptoms. These individuals are then tested for the presence of anti-CCP antibodies (CCP2 assay) using Bioplex 2200 kit (BioRad, positive test .2.99IU/ml). If positive for anti-CCP antibodies individuals are invited to attend screening clinics at Chapel Allerton Hospital, Leeds where participants are monitored for progression to RA. Individuals with inflammatory bowel disease, gastrointestinal (GI) malignancy or previous GI surgery resulting in stoma formation were excluded from the study. In addition, those with recent (within 3 months) antibiotic, laxative or pre/probiotic use were also excluded. A comparator population of 44 healthy controls was selected from a publicly available dataset (18), matched for sequencing method, western diet (high consumption of proteins, saturated fats, refined grains, sugar) and approximate age and gender.

2.3 Stool collection

20ml of stool was collected using an in-house collection kit and stored at room temperature. Samples were returned to Chapel Allerton Hospital, Leeds directly by the study participants. DNA was extracted and frozen at -80°C within 24hrs of stool production.

2.4 DNA Extraction, Library preparation, Sequencing

Faecal DNA was extracted from 200-400mg of unprocessed stool using QIAamp DNA stool mini kit and stored at -80°C. PCR amplification of 16S rRNA was preformed using specific primer sequences for the V4 region (V4F-5’-AYTGGGYDTAAAGNG, V4R- 5’ – TACNVGGGTATCTAATCC) (19). Libraries were prepared using the NEBNext® Ultra™ DNA Library Prep Kit for illumina sequencing, as previously described (20). Sequencing was performed at the University of Leeds sequencing centre using the illumina MiSeq 2500 to produce 2x 250bp paired-end output. Average amplicon size was 372bp. The average quality score for each read was 36, The median sequencing depth for anti-CCP positive population was 88320Kb (17240 - 157193Kb) and the comparator population was 67542kb (35159 – 83517Kb).

2.5 Diversity, composition and network analysis

Adapter sequences were removed from demultiplexed FASTQ files using cutadapt (21). A python script (multiple_join_paired_ends.py) from QIIME script source was used to merge pair ends and subsequently converted to FASTA format. Further analysis was performed using QIIME 1.8.0. (22) Operational taxonomy units (OTUs) were assigned using Usearch (23), and aligned to the Greengenes reference database using PyNAST (24), Taxonomy was assigned using the RDP 2.2 classifier (25). The resulting OTU 'biom' files from the above analyses were imported in MEGAN (26), for group analyses, annotations and plots. Data was normalised to the third lowest sequencing depth of 63,218Kb. Alpha and Beta diversity was calculated using Principal Coordinate Analysis (PCoA) based on the Shannon diversity index and the non-phylogenetic Bray-Curtis dissimilarity index respectively. The adonis function in vegan library (27), in R (28), was used to perform a permutation MANOVA between groups. Welch T test in R was used to investigate for significance between specific taxa. Network analysis of anti-CCP positive, including CCP+ progressors and CCP+ non-progressors, and the comparator population was created using the unrooted phylogenetic neighbour-net method using Goodall's index as the output metric (29).

3. Results

3.1 Population characteristics

A total number of 69 individuals were included in our study, 25 anti-CCP positive at-risk individuals and 44 healthy controls. Ethical approval was granted by Leeds (West) Research Ethics Committee (06/Q1205/169) and samples collected after informed written consent. Baseline characteristics of the both populations are outlined in table 1. In the anti-CCP positive group the median antibody titre was 58.5 IU/ml (seroprevalence of anti-CCP positivity in the general population is ≈ 1%, (30)) with no clinical synovitis demonstrated at the time of gut microbiome sampling, no anti-CCP positive individuals received DMARD therapy. During the study period 5 anti-CCP positive individuals progressed to RA (CCP+ progressors), median time from stool collection to progression was 188 days (range 100- 457 days). Baseline characteristics of anti-CCP positive individuals that progressed to RA are described in table 2.

3.2 Diversity

No change was observed in the alpha diversity using the Shannon diversity index between the anti-CCP positive population and the control population, see figure 1A. Beta diversity was different between anti-CCP positive patients compared to healthy controls, evidenced by the distribution of the coordinates into two clusters, see figure 1B, as calculated using the Bray-Curtis dissimilarity index. A permutation MANOVA between anti-CCP positive individuals and controls showed significance, p=0.01.

3.3 Gut bacterial composition

Differences in bacterial taxa were noted between the anti-CCP positive population and the controls, see table 3 and figure 2. The anti-CCP positive population has an overabundance of Helicobacteraceae, *Erysipelotrichaceae, Ruminococcaceae* (table 3), and a lower abundance of *Bacteroidaceae, Barnesiellaceae, Methanobacteriaceae* (figure 2) amongst others.

3.4 Gut bacterial composition between anti-CCP positive and RA

Five anti-CCP positive individuals progressed to RA (table 2). The median interval between sample collection and progression was 188 days (range 100-457 days). There were no significant differences between alpha (Shannon diversity) diversity Figure 1C, and beta (Bray-Curtis dissimilarity index) diversity Figure 1D, between CCP+ progressors and CCP+ non-progressors.

3.5 Network analysis

A network analysis constructed using Goodall's index demonstrated clustering of the anti-CCP positive population and the healthy comparators, see figure 3. Furthermore, 4 out of 5 of the progressors clustered within the same arm of the network (figure 3).

4. Discussion

Our results demonstrate the gut microbiome of individuals at risk of RA is significantly different to that of a healthy control population. This supports the mucosal origins hypothesis theory (31), where inflammation and autoimmunity begins at mucosal sites, including the gut, and later transitions to involve the joints . Whilst the concept of a pre-RA dysbiosis has been reported, we have shown a compositional difference in gut microbiome structure compared to previous authors (32). This suggests that the mucosal origins hypothesis of RA maybe linked to multiple bacterial taxa rather than a single bacterial perpetrator, a hypothesis that has been suggested elsewhere (33). It is worth noting, the at-risk population investigated by Alpizar-Rodriguez et al. included individuals with undifferentiated arthritis (32) (likely including those with spondyloarthritis), hence representing progression of the disease phenotype beyond the Leeds at-risk cohort investigated in this paper.

It is worth noting, while the term at risk of RA has been used, it is currently unclear whether RA can be prevented or reversed at this stage. Early clinical RA may represent the endpoint of a disease continum, where individuals with genetic risk and/or environmental risk factors, develop autoantibodies, and subsequently progress to develop RA. In such a situation, the use of the term 'at-risk' may be more accurately replaced with 'pre-RA' but this is only possible on retrospective analysis (when a diagnosis of RA has already been made, which does not occur in all). Evidence from prospective anti-CCP positive observational cohorts has demonstrated certain factors (smoking, obesity, early morning stiffness, raised inflammatory markers, HLA, imaging) to be associated with a higher risk of progression to RA (16, 34-36). Some at-risk individuals have subclinical joint inflammation on imaging (ultrasound and/or MRI) without clinically-apparent arthritis. These individuals are at particularly high risk of progression to clinical arthritis. These individuals may be considered a distinct group of at-risk individuals with early subclinical RA, although the significance of this in terms of prevention is not clear.

Our findings of increased *Lachnospiraceae* and *Ruminococcaceae* are in keeping with the preclinical phase of arthritis observed in collagen induced arthritis rodent models (37, 38). Furthermore, both *Erysipelotrichaceae* and *Helicobacteraceae*, which were also identified to be enriched in our anti-CCP positive population, exhibit highly immunogenic properties capable of stimulating the production of mucosal associated IgA, a key player in preclinical RA (39). Previous studies have demonstrated an overabundance of *Prevotella copri* in new onset RA (8), and indeed in the preclinical RA phases (32). We investigated for the presence of increased *Prevotellaceae* in our anti-CCP positive cohort, but no significant difference was found. These findings point towards underlying complex microbial and host interactions, which in turn may create a variety of ecological niches allowing the preferential expansion of certain bacterial taxa, possibly at various disease time points. Sequentially sampled individuals/cohorts will be needed to explore this hypothesis.

Construction of the phylogenetic network using the neighbour-net method allows for the combination of ordination without the restraint of hierarchical clustering imposed by traditional network analysis (29). This allows for the visualisation of clusters linked by an appropriate ecological index, without the need to fit the metagenomic data into a rooted dendrogram. We have chosen to use Goodall's index as the input for our network analysis, as it attributes increased weight to the rarer taxa present within a microbiome (40, 41), hence allowing for visual differentiation between taxa that are unique to clustering samples. As our study was preformed prospectively, we identified 5 individuals that progressed to RA following sampling. Intriguingly, 4 of these 5 individuals cluster along the same arm of the Goodall's network, suggesting a role for rarer taxa in RA development. Additionally, given a phylogenetic linkage was identified in those individuals that progressed, with a median duration of 188 days, a potential timeline as to when a microbial impetus may begin to act upon the immune system could be hypothesised. Again, a longitudinal dataset is needed to determine when microbiome changes occur along the RA continuum.

It has been demonstrated that the gut microbiome of individuals with early RA possess decreased alpha diversity when compared with healthy controls (42). This is true of many autoimmune conditions, in which associations have been drawn to the gut microbiome (43, 44). Our study demonstrates the presence of a dysbiosis with preserved alpha diversity in the preclinical phase of RA when compared with a healthy control population, and indeed between progressors and non-progressors. An attractive explanation would be to assume the inflammatory nature of the underlying autoimmune disease leads to an inhospitable microbial environment, and thus to decreased diversity at the time of sampling, which coincides with systemic inflammation related to clinically established disease. However, interval microbiome sampling up to and including the point of RA onset is key to understanding the changes in the gut microbiome's diversity in preclinical RA.

While an attempt has been made to minimise discrepancies between populations investigated, we acknowledge that the use of a publicly available dataset must be interpreted with some caveats, additionally, the anti-CCP, RF, HLA and smoking status of our comparator population is unknown and is a limitation of this study. Also it is worth highlighting the sample size of the anti-CCP positive progressor population (n=5), where phylogenetic clustering may not be evident in a larger cohort. The ideal comparator dataset would include longitudinal sampling of the at risk population with comparison of phenotyped progressors and non-progressors, including those with genetic risk factors such as first degree relatives and healthy household contacts.

5. Conclusion

Anti-CCP positive at-risk individuals without clinical synovitis have a different microbiome when compared with healthy controls. Those at-risk individuals that progressed to RA displayed clustering in a phylogenetic network suggesting a commonality of microbial taxa that predates disease onset by many months. Our data underscore the need for careful longitudinal sampling and analysis to understand the role of the host microbiome in RA development and its potential as a target for preventative intervention.

Funding and COI: CMR receives funding via a personal fellowship by Versus Arthritis [22294]. Leeds CCP clinic is supported by the National Institute for Health Research (NIHR), Leeds Musculoskeletal Biomedical Research Unit. KM has received honoraria from Abbvie, UCB and Eli Lilly & Co. PE has received consultant fees from AbbVie, BMS, Gilead, Eli Lilly, MSD, Novartis, Pfizer, Roche, Samsung, Sandoz and UCB and received research grants paid to his employer from AbbVie, BMS, Pfizer, MSD and Roche.

Acknowledgments

We would like to thank participants of the Leeds at-risk cohort and the patient discussion groups.

Author contribution

MHW, PE, KM, CMR designed the study. KM and CMR were involved in sample and data collection. IBM and CMR were involved in sample processing. SM and CMR were involved in data analysis. All authors were involved in writing the manuscript and approved the article for publication.

Data availability

Raw data files will be uploaded to a repository if accepted.

Bibliography

1. Nielen MM, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MH, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheum. 2004;50(2):380-6.
2. Deane KD, Norris JM, Holers VM. Preclinical rheumatoid arthritis: identification, evaluation, and future directions for investigation. Rheum Dis Clin North Am. 2010;36(2):213-41.
3. Mankia K, Emery P. Preclinical Rheumatoid Arthritis: Progress Toward Prevention. Arthritis Rheumatol. 2016;68(4):779-88.
4. Rantapaa-Dahlqvist S, de Jong BA, Berglin E, Hallmans G, Wadell G, Stenlund H, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum. 2003;48(10):2741-9.
5. Barra L, Scinocca M, Saunders S, Bhayana R, Rohekar S, Racape M, et al. Anti-citrullinated protein antibodies in unaffected first-degree relatives of rheumatoid arthritis patients. Arthritis Rheum. 2013;65(6):1439-47.
6. Hooper LV, Macpherson AJ. Immune adaptations that maintain homeostasis with the intestinal microbiota. Nat Rev Immunol. 2010;10(3):159-69.
7. Lerner A, Aminov R, Matthias T. Dysbiosis May Trigger Autoimmune Diseases via Inappropriate Post-Translational Modification of Host Proteins. Frontiers in Microbiology. 2016;7:84.
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. Rodrigues GSP, Cayres LCF, Goncalves FP, Takaoka NNC, Lengert AH, Tansini A, et al. Detection of Increased Relative Expression Units of Bacteroides and Prevotella, and Decreased Clostridium leptum in Stool Samples from Brazilian Rheumatoid Arthritis Patients: A Pilot Study. Microorganisms. 2019;7(10).
10. Breban M, Tap J, Leboime A, Said-Nahal R, Langella P, Chiocchia G, et al. Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. Annals of the Rheumatic Diseases. 2017;76(9):1614-22.
11. Vaahtovuo J, Munukka E, Korkeamaki M, Luukkainen R, Toivanen P. Fecal microbiota in early rheumatoid arthritis. J Rheumatol. 2008;35(8):1500-5.
12. Muñiz Pedrogo DA, Chen J, Hillmann B, Jeraldo P, Al-Ghalith G, Taneja V, et al. An Increased Abundance of Clostridiaceae Characterizes Arthritis in Inflammatory Bowel Disease and Rheumatoid Arthritis: A Cross-sectional Study. Inflammatory Bowel Diseases. 2018;25(5):902-13.
13. Chiang HI, Li JR, Liu CC, Liu PY, Chen HH, Chen YM, et al. An Association of Gut Microbiota with Different Phenotypes in Chinese Patients with Rheumatoid Arthritis. J Clin Med. 2019;8(11).
14. Picchianti-Diamanti A, Panebianco C, Salemi S, Sorgi ML, Di Rosa R, Tropea A, et al. Analysis of Gut Microbiota in Rheumatoid Arthritis Patients: Disease-Related Dysbiosis and Modifications Induced by Etanercept. Int J Mol Sci. 2018;19(10).
15. 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. Nature Medicine. 2015;21(8):895-905.
16. Rakieh C, Nam JL, Hunt L, Hensor EMA, Das S, Bissell L-A, et al. Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study. Annals of the Rheumatic Diseases. 2015;74(9):1659-66.

17. Mankia K, Cheng Z, Do T, Hunt L, Meade J, Kang J, et al. Prevalence of Periodontal Disease and Periodontopathic Bacteria in Anti-Cyclic Citrullinated Protein Antibody-Positive At-Risk Adults Without Arthritis. JAMA Netw Open. 2019;2(6):e195394.

18. Jangi S, Gandhi R, Cox LM, Li N, von Glehn F, Yan R, et al. Alterations of the human gut microbiome in multiple sclerosis. Nat Commun. 2016;7:12015-.

19. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Research. 2010;38(22):e200-e.

20. Moura IB, Buckley AM, Ewin D, Clark E, Mitra S, Wilcox MH, et al. Profiling the effects of rifaximin on the healthy human colonic microbiota using a chemostat model. bioRxiv. 2019:828269.

21. Martin MJEj. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011;17(1):10-2.

22. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 2010;7(5):335-6.

23. Edgar RCJB. Search and clustering orders of magnitude faster than BLAST. 2010;26(19):2460-1.

24. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight RJB. PyNAST: a flexible tool for aligning sequences to a template alignment. 2010;26(2):266-7.

25. Wang Q, Garrity GM, Tiedje JM, Cole JRJAEM. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. 2007;73(16):5261-7.

26. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. 2016;12(6).

27. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, et al. Package 'vegan'. 2013;2(9):1-295.

28. Team RC. R: A language and environment for statistical computing. 2013.

29. Mitra S, Gilbert JA, Field D, Huson DH. Comparison of multiple metagenomes using phylogenetic networks based on ecological indices. The ISME Journal. 2010;4(10):1236-42.

30. van Zanten A, Arends S, Roozendaal C, Limburg PC, Maas F, Trouw LA, et al. Presence of anticitrullinated protein antibodies in a large population-based cohort from the Netherlands. Annals of the Rheumatic Diseases. 2017;76(7):1184-90.

31. Holers VM, Demoruelle MK, Kuhn KA, Buckner JH, Robinson WH, Okamoto Y, et al. Rheumatoid arthritis and the mucosal origins hypothesis: protection turns to destruction. Nature Reviews Rheumatology. 2018;14(9):542-57.

32. Alpizar-Rodriguez D, Lesker TR, Gronow A, Gilbert B, Raemy E, Lamacchia C, et al. Prevotella copri in individuals at risk for rheumatoid arthritis. Ann Rheum Dis. 2019;78(5):590-3.

33. Pianta A, Arvikar SL, Strle K, Drouin EE, Wang Q, Costello CE, et al. Two rheumatoid arthritis-specific autoantigens correlate microbial immunity with autoimmune responses in joints. J Clin Invest. 2017;127(8):2946-56.

34. Berglin E, Padyukov L, Sundin U, Hallmans G, Stenlund H, Van Venrooij WJ, et al. A combination of autoantibodies to cyclic citrullinated peptide (CCP) and HLA-DRB1 locus antigens is strongly associated with future onset of rheumatoid arthritis. Arthritis Res Ther. 2004;6(4):R303-R8.

35. Hair M, Landewé R, van de Sande M, Schaardenburg D, Baarsen L, Gerlag D, et al. Smoking and overweight determine the likelihood of developing rheumatoid arthritis. Annals of the rheumatic diseases. 2012;72.

36. van de Stadt LA, Witte BI, Bos WH, van Schaardenburg D. A prediction rule for the development of arthritis in seropositive arthralgia patients. Annals of the Rheumatic Diseases. 2013;72(12):1920.

37. Rogier R, Evans-Marin H, Manasson J, van der Kraan PM, Walgreen B, Helsen MM, et al. Alteration of the intestinal microbiome characterizes preclinical inflammatory arthritis in mice and its modulation attenuates established arthritis. Sci Rep. 2017;7(1):15613-.

38. Liu X, Zeng B, Zhang J, Li W, Mou F, Wang H, et al. Role of the Gut Microbiome in Modulating Arthritis Progression in Mice. Sci Rep. 2016;6:30594.

39. Palm NW, de Zoete MR, Cullen TW, Barry NA, Stefanowski J, Hao L, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell. 2014;158(5):1000-10.

40. Goodall DW. A New Similarity Index Based on Probability. Biometrics. 1966;22(4):882-907.

41. Legendre P, Legendre L. Numerical Ecology1998. 1-853 p.

42. Jeong Y, Kim J-W, You HJ, Park S-J, Lee J, Ju JH, et al. Gut Microbial Composition and Function Are Altered in Patients with Early Rheumatoid Arthritis. Journal of clinical medicine. 2019;8(5):693.

43. Gong D, Gong X, Wang L, Yu X, Dong Q. Involvement of Reduced Microbial Diversity in Inflammatory Bowel Disease. Gastroenterol Res Pract. 2016;2016:6951091-.

44. Swidsinski A, Dörffel Y, Loening-Baucke V, Gille C, Göktas Ö, Reißhauer A, et al. Reduced Mass and Diversity of the Colonic Microbiome in Patients with Multiple Sclerosis and Their Improvement with Ketogenic Diet. Frontiers in microbiology. 2017;8:1141-.

Tables

Variable	Anti-CCP positive (n=25)	Control (n=44)
Age, mean (SD)	50.5 (13.4)	42.2 (9.61)
Age, range	21 - 78	
Female (%)	19 (76)	37 (84%)
Anti-CCP, median	58.5 IU/ml	-
Rheumatoid factor positive	11 (44)	-
Shared epitope (%)	12 (48)	-
Early morning stiffness >30mins	6 (24)	NA
Clinical evidence of synovitis	0	NA
Joint tenderness (%)	10(40)	NA
Smoking status (%)	15(60)	-
History of vegetarian diet (%)	3 (12)	9 (20)

Table 1. Participants characteristics. Shared epitope was considered positive with presence of one or two copies of HLA-DRB1*01, DRB1*04 or DRB1*10 in the HLA-DRB1 locus. Evidence of early morning stiffness was patient reported. Clinical synovitis and joint tenderness (small joints) was determined on examination by clinically qualified personnel. Smoking status included current or ex-smokers.

Variable	CCP+P1	CCP+P2	CCP+P3	CCP+P4	CPP+P5
Age	46	52	71	46	56
Gender	Female	Female	Female	Male	Female
CCP titre, IU/ml	147	300	167	300	7.8
Shared epitope	Positive	Negative	Negative	Negative	Positive
Early morning stiffness >30mins	Yes	No	No	No	No
Joint tenderness	Yes	No	Yes	No	Yes
Smoking status	Positive	Positive	Positive	Negative	Negative
History of vegetarian diet	No	No	No	No	No

Table 2. Anti-CCP positive progressor population characteristics. Baseline characteristics for the five anti-CCP positive individuals that progressed (P) to RA; CCP+P1-P5. Shared epitope was considered positive with presence of one or two copies of HLA-DRB1*01, DRB1*04 or DRB1*10 in the HLA-DRB1 locus. Evidence of early morning stiffness (minutes) was patient reported. Joint tenderness (small joints of the hands and feet) was determined on examination by clinically qualified personnel. Smoking status included current or ex-smokers. Dietary status was patient reported.

Rank	**Taxa: Family**	**P value**	**Taxa: Genus**	**P value**
1	Helicobacteraceae	5.00E-06	Coprococcus	1.61E-05
2	Erysipelotrichaceae	3.98E-05	Oscillospira	2.49E-05
3	Ruminococcaceae	0.001676261	Lachnospira	0.000716319
4	Peptostreptococcaceae	0.002489999	Absiella	0.00096231
5	Bifidobacteriaceae	0.008222693	Roseburia	0.003524867
6	Gracilibacteraceae	0.009301412	Allobaculum	0.004513859
7	Planococcaceae	0.011158723	Faecalibacterium	0.00484119
8	Deferribacteraceae	0.012880422	Bifidobacterium	0.008167057
9	Victivallaceae	0.014814023	Mucispirillum	0.012880422
10	Lachnospiraceae	0.022368412	Helicobacter	0.021448002
11	Peptococcaceae	0.046105165	Flexispira	0.036367237
12			Oxobacter	0.04149263
13			Gracilibacter	0.044984211
14			Peptococcus	0.048977148

Table 3. Welch T Test at family level between CCP positive individuals and control population. Table showing taxa that were significantly enriched in the anti-CCP positive population (including the RA progressors) when compared to control population.

Figure 1

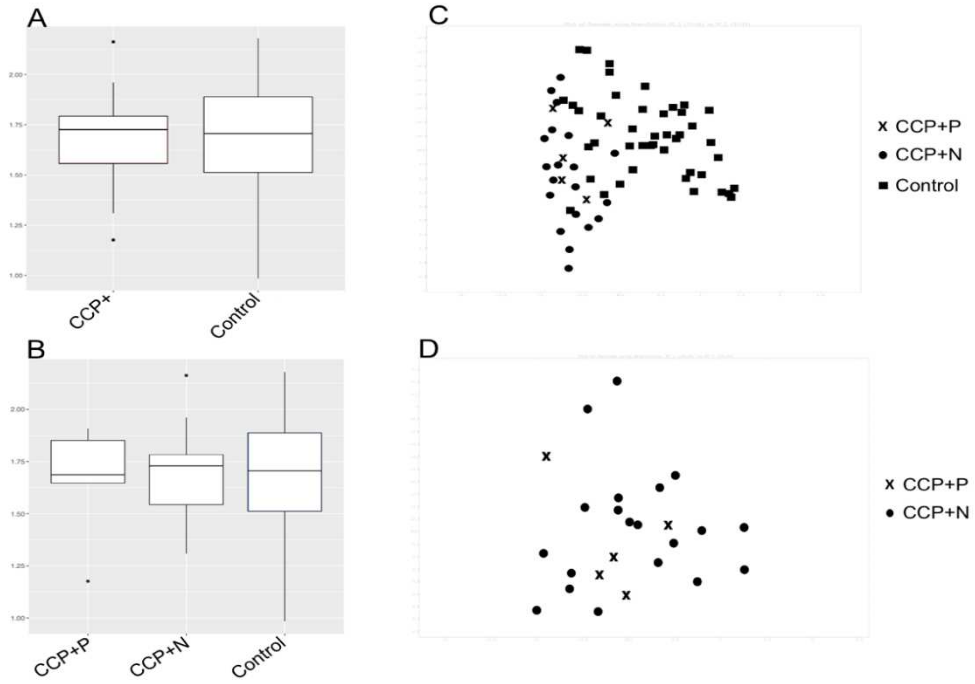

Figure 1. Microbiome diversity analysis. CCP+: Anti-CCP positive population, including CCP+ progressors and CCP+ non-progressors. CCP+P: CCP+ progressor, CCP+N: CCP+ non-progressor. **Figure 1A: Box plot of alpha diversity using Shannon diversity index.** Solid dots represent gut microbiomes outlying the interquartile range. **Figure 1B: Box plot of alpha diversity using Shannon diversity index.** No significant change in diversity between **Figure 1C: Principal Coordinate Analysis (PCoA) of beta diversity using Bray-Curtis dissimilarity index.** Two distinct populations are noted; the anti-CCP positive population (including CCP+ progressors and CCP+ non-progressors) and the control population. **Figure 1D: PCoA of beta diversity using Bray-Curtis dissimilarity index.** No clustering of the CCP+ progressor population is noted within the PCoA plot.

Figure 2

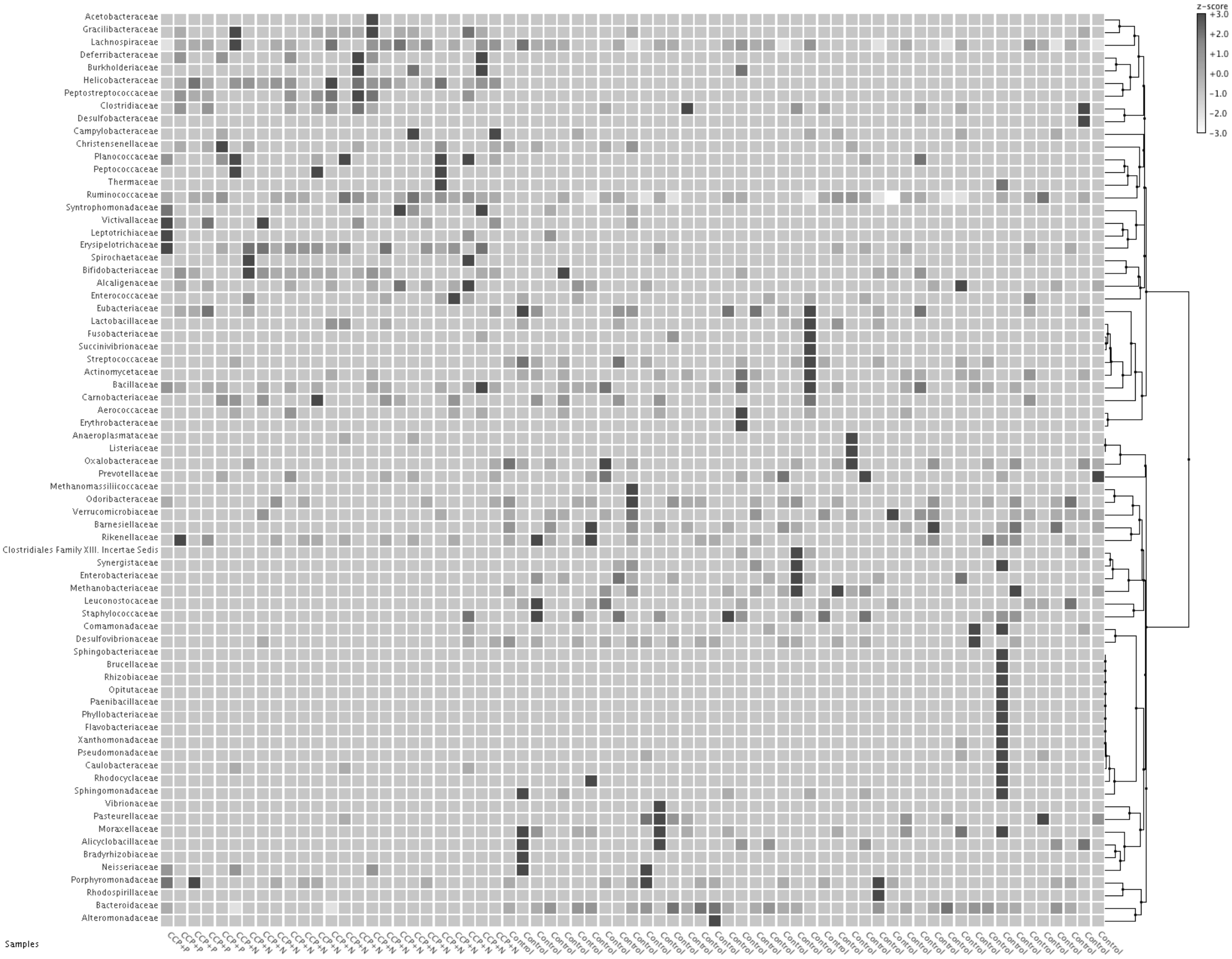

Figure 2. Heat Map. Heat Map clustered by sample and by taxonomy at family level. CCP+P: CCP+ progressor, CCP+N: CCP+ non-progressor. Colour intensity of cells represent accompanying Z score.

Figure 3

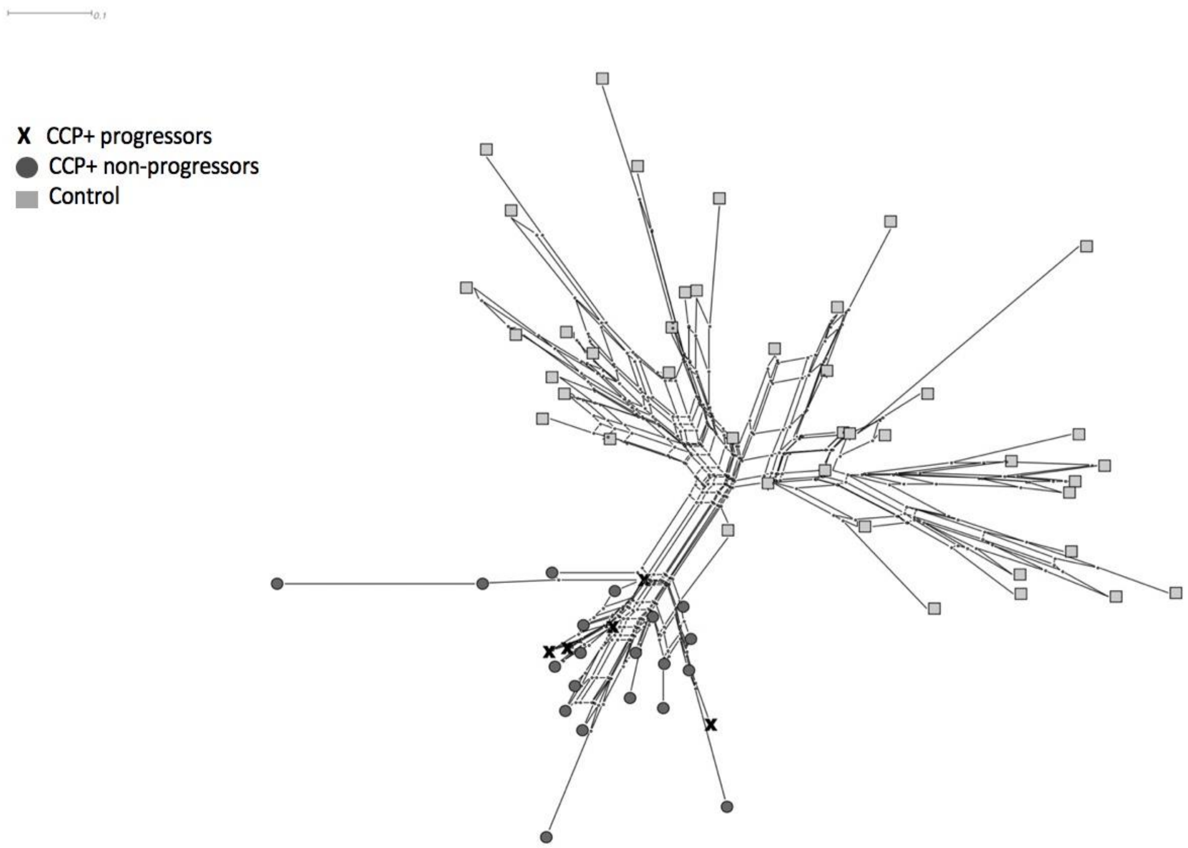

Figure 3. Phylogenetic network. Non-hierarchical phylogenetic network created using Goodall's index via the neighbour-net method. Each terminal node, represented by a different shape, indicates an individual's gut microbiome. Note 4 out of 5 progressors are united on a single arm of the network.