**Spondyloarthritis, acute anterior uveitis, and Crohn’s disease have both shared and distinct gut microbiota**

# Supplementary Data File

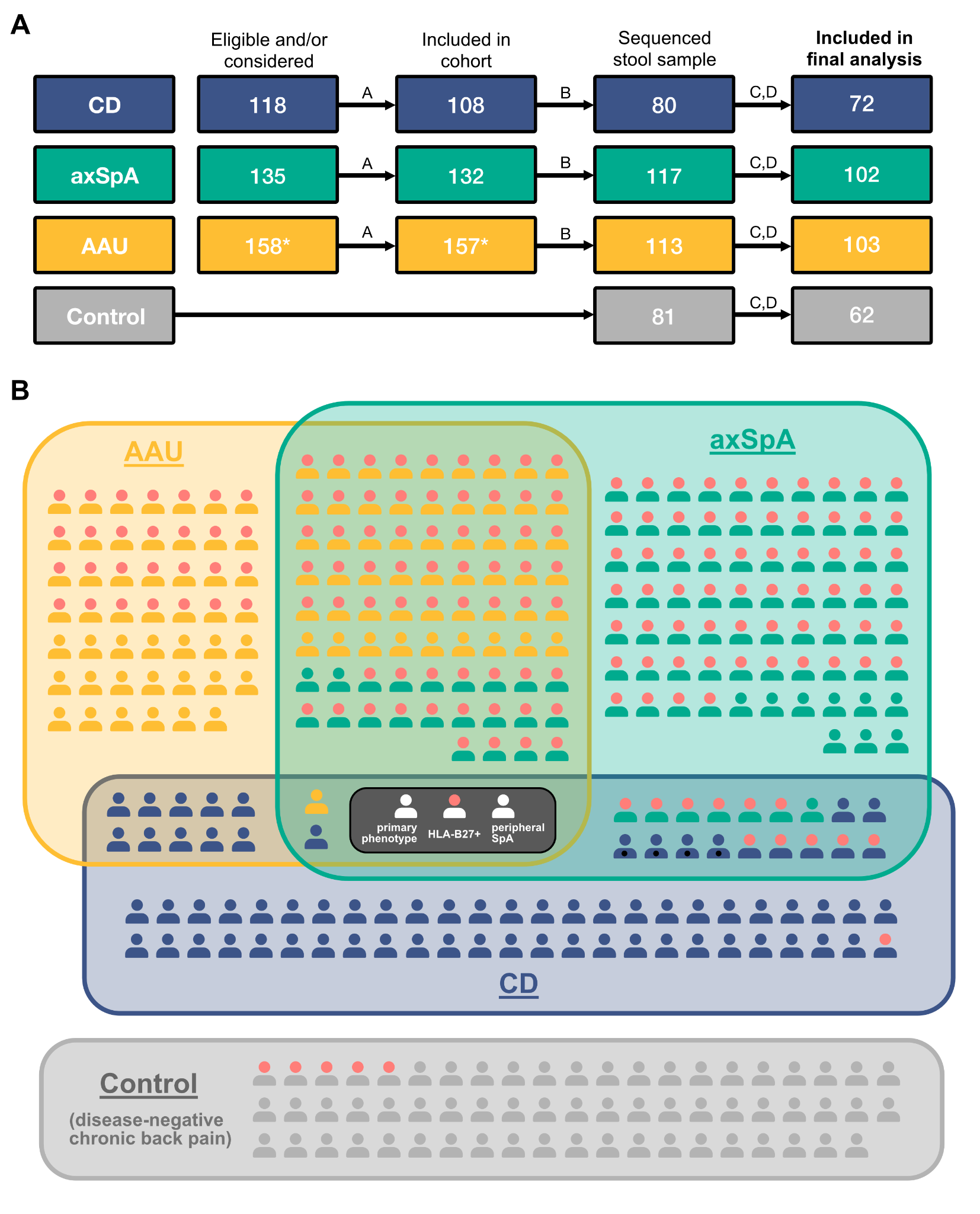
# Methods

## **Differential abundance analysis implementation**

Custom linear models were built for all n=565 taxa using the *lm* function in base R with the formula *relative\_abundance ~ disease\_status \* sequencing\_experiment* to account for technical variation. Model parameters were extracted, tested for significance using the default Wald method, adjusted for multiple testing (Benjamini-Hochberg procedure), and bootstrapped to obtain 95% confidence intervals using the *parameters::model\_parameters()* function. All taxa were also tested for significant differential abundance using a blocked Wilcoxon permutation test (*coin::wilcox\_test()* function with the formula *relative\_abundance ~ disease\_status | sequencing\_experiment*), and raw *P* values were adjusted for multiple testing.

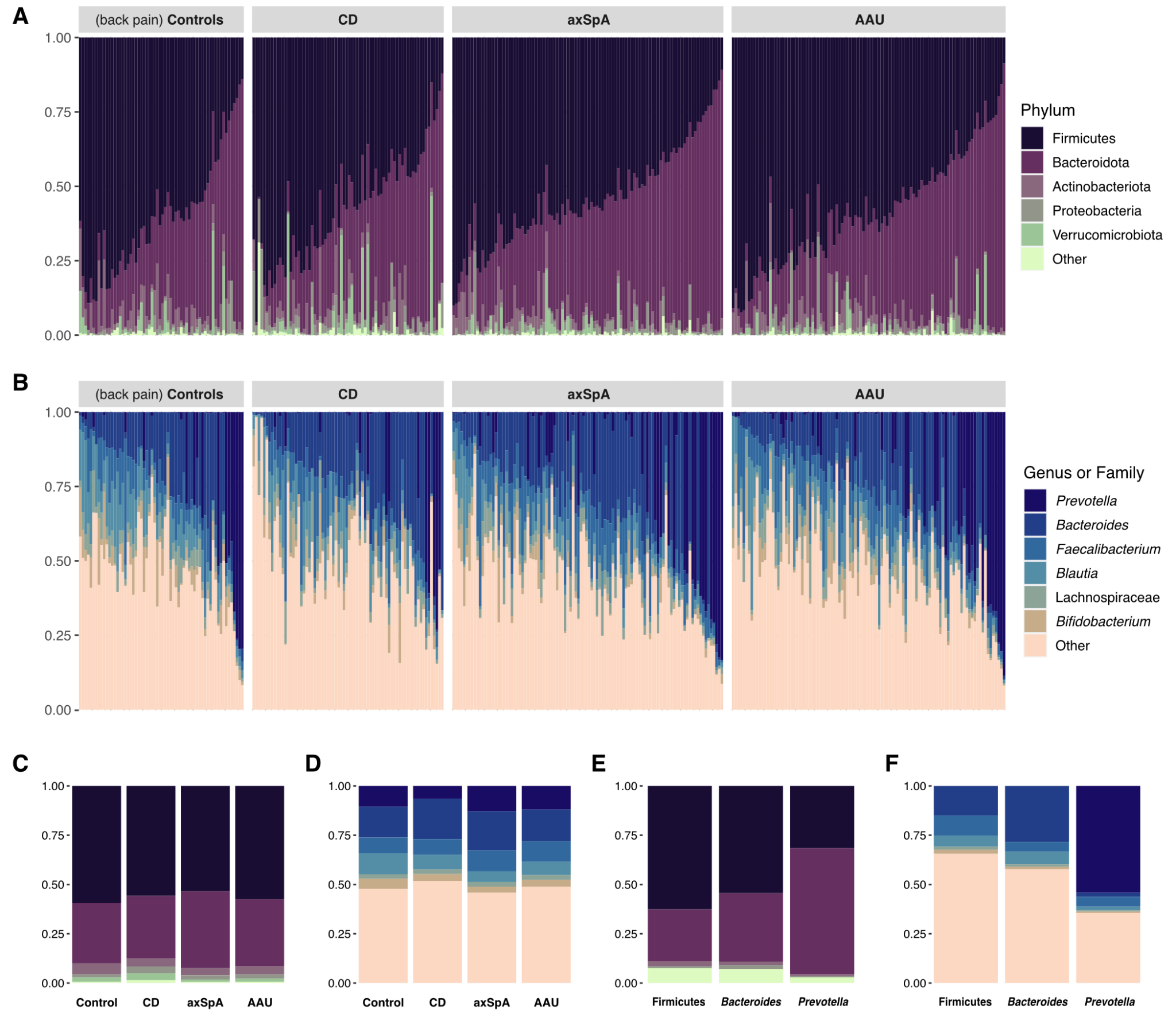
## **Confounder analysis with medications**

Instead of using the case-control groupings from the differential abundance and mediation analyses (see main text **Methods**), samples were first grouped according to disease phenotype (groups with <40 samples were not analyzed), such that disease status was not a factor, and the formula *relative\_abundance ~ intake\_status \* sequencing\_experiment* was used to test specific disease-drug differential abundance hypotheses (e.g. conventional synthetic (cs)DMARD use, which was only relevant in CD patients). In the group of all patients examining previous antibiotic use, the formula *relative\_abundance ~ disease\_status \* sequencing\_experiment + intake\_status* was used; in the control subset examining NSAID use, simple models (*relative\_abundance ~ intake\_status*) were used as no other factors were relevant.



## **Supplementary Figure 1: Overview of cohorts and individuals included in this study**

**A)** Flowchart containing numbers of participants at each stage of study with letters denoting reasons for exclusion, namely: *A*=excluded if patient inclusion criteria not met, *B*=excluded if no stool sample given at baseline visit, *C*=excluded if metadata incomplete or sequenced stool sample of poor quality, *D*=excluded if oral antibiotics taken ≤30 days prior to stool sample, or if previous diagnosis of SpA, AAU, CD, UC, or psoriasis present. An asterisk\* denotes that recruitment was still ongoing at the time of analysis. Control samples were obtained from the Optiref study[1](https://paperpile.com/c/reaLWV/oDRCU) and were resequenced and reprofiled with patient samples. **B)** Patients were recruited into cohorts reflecting their primary disease pathology, but varied in their expression of human leukocyte antigen (HLA) B27 and their presentation of secondary and tertiary pathologies. CD, Crohn’s disease; axSpA, axial spondyloarthritis; AAU, acute anterior uveitis.



**Supplementary Figure 2: Microbiota composition of immune-mediated disease cohorts and back pain controls.** Relative abundances at different taxonomic resolutions of all n=339 individuals in our study, grouped by primary (recruitment) cohort. **A)** Phylum-level taxonomic profiles. Phyla with a median relative abundance of 0 across at least two cohorts were grouped into “Other”. **B)** Taxonomic profiles of the top six most abundant taxa, corresponding to relative abundances binned at the genus level, except for Lachnospiraceae (representing an unknown genus inside the Lachnospiraceae family). All other taxa were pooled into “Other”. Both legends are ranked according to median relative abundance across all cohorts from top to bottom and samples follow the same order in A and B. **C)**  Median relative abundances of phyla by cohort. **D)** Median relative abundances of genera by cohort. **E)** Same as C but for enterotype groups determined from the bins in D (modification from Lahti and Shetty[2](https://paperpile.com/c/reaLWV/Toyrr), see also Holmes, Harris, and Quince[3](https://paperpile.com/c/reaLWV/yODq)), representing the sum of all ASVs bearing the indicated genus label. **F)** Same as D, but as in E shown for each enterotype.

All supplementary tables are hosted at [www.github.com/sxmorgan/gespic-public](http://www.github.com/sxmorgan/gespic-public), and printed here if space allows.

## **Supplementary Table 1: Taxonomic annotations for all n=442 ASVs used in the differential abundance analysis**. Species annotations were inferred using the *dada2::addSpecies* function with *allowMultiple=F*, in order to exclude identification of ASVs where a species-level consensus was not reached[4](https://paperpile.com/c/reaLWV/6BqGH). For the binned differential abundance analysis, the column “bin” was used to sum ASV rows.

| **Covariate** | **Variable type** | **Description** |
| --- | --- | --- |
| GESPIC | Binary | Any disease diagnosis |
| Age | Continuous | Age in years |
| Sex | Binary | Sex assigned at birth |
| BMI | Continuous | Body mass index |
| HLAB27 | Binary | HLA-B27 expression |
| CRP | Continuous | C-Reactive Protein in mg/L |
| ESR | Continuous | Erythrocyte sedimentation rate in mm/h |
| BASDAI | Ordinal | Bath Ankylosing Spondylitis Disease Activity Index |
| SpA | Binary | Previous or current spondyloarthritis diagnosis |
| CD | Binary | Previous or current Crohn's diagnosis |
| AAU | Binary | Previous or current acute anterior uveitis diagnosis |
| PsA | Binary | Previous or current psoriatic arthritis diagnosis |
| Antibiotics\_recent | Binary | Antibiotic treatment 2-3 months before sampling |
| Corticoid\_current | Binary | Current glucocorticoid treatment |
| csDMARD\_ever | Binary | Current conventional synthetic DMARD treatment |
| csDMARD\_current | Binary | Previous conventional synthetic DMARD treatment |
| NSAID\_current | Binary | Current NSAID treatment |
| Smoker | Binary | Current smoker |
| Cohort\_CD | Binary | Crohn’s disease cohort (primary CD diagnosis) |
| Cohort\_SpA | Binary | Axial spondyloarthritis cohort (primary SpA diagnosis) |
| Cohort\_AAU | Binary | Acute anterior uveitis cohort (primary AAU diagnosis) |
| LibSize | Continuous | Raw number of sequencing reads from stool sample |

## 

## **Supplementary Table 2: Complete list of tested covariates available for all patients and controls**

GESPIC: German Spondyloarthritis Inception Cohort; HLA-B27: human leukocyte antigen B27; DMARD: disease-modifying anti-rheumatic drug; NSAID: non-steroidal anti-inflammatory drug

**Supplementary Table 3: Master table of all n=339 anonymized individuals in our study.** ASV and higher taxa (binned at genus- or family-level) relative abundance profiles, alpha diversities, and clinical metadata (NA if missing) are included in wide format.

# References

1. [Proft, F. *et al.* Comparison of an online self-referral tool with a physician-based referral strategy for early recognition of patients with a high probability of axial spa. *Semin. Arthritis Rheum.* **50**, 1015–1021 (2020).](http://paperpile.com/b/reaLWV/oDRCU)

2. [Lahti, L. & Al, S. S. et. Dirichlet Multinomial Mixtures.](http://paperpile.com/b/reaLWV/Toyrr) <https://microbiome.github.io/tutorials/DMM.html>[.](http://paperpile.com/b/reaLWV/Toyrr)

3. [Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* **7**, e30126 (2012).](http://paperpile.com/b/reaLWV/yODq)

4. [Taxonomic Assignment.](http://paperpile.com/b/reaLWV/6BqGH) <https://benjjneb.github.io/dada2/assign.html>[.](http://paperpile.com/b/reaLWV/6BqGH)