SUPPLEMENTARY MATERIALS AND METHODS

Study Design

An observational case-control study was designed to profile the skin microbiota of individuals with cutaneous T-cell lymphoma (CTCL) compared to healthy controls. The study was approved by the Johns Hopkins Institutional Review Board, and all participants provided written informed consent. The following inclusion criteria were required of all CTCL patients: age ≥18 with a diagnosis of Mycosis Fungoides (MF) or Sezary Syndrome (SS) with an appropriate staging workup involving skin biopsies. All subjects underwent a strict skin preparatory regimen including the following: avoidance of topical medications, ointments, creams, lotions, petroleum jelly, and shaving cream on selected sites for 7 days prior to sampling; avoidance of oral antibiotic/steroid medications, swimming, bleach baths, antidandruff shampoo, perfumes/cologne, and antibacterial soaps for 7 days prior to sampling; and avoidance of bathing or showering for the 24 hours prior to sampling. None of the subjects received systemic antibiotics within the 3 months prior to sample collection. Exclusion criteria included: concurrent skin disease, pregnancy, and inability to provide informed consent or follow study protocols. Five healthy volunteers (HV1-5) were recruited in the general dermatology clinic at Johns Hopkins after coming in for general skin checks or for evaluation of benign skin lesions/nevi. An additional five HVs (HV6-10) were recruited as healthy volunteers at the National Institutes of Health following the same exclusion criteria outlined above to participate in a study approved by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH Institutional Review Board (https://clinicaltrials.gov/ct2/show/NCT00001505). HVs had no history of acute or chronic inflammatory skin conditions and were required to be in good health.

Patient Characteristics

Six MF/SS patients and ten healthy age- and sex-matched controls participated in the microbiome study. Demographic information, clinical characteristics, and treatment histories of study subjects are presented in Supplementary Table S1. The study populations of MF/SS patients and HVs were predominantly white with mean ages of 55.2 ± 17.9 and 50.1 ± 14.5 years, respectively. Upon enrollment of the MF/SS patient population, 33% were clinically diagnosed as early stage disease and 67% were diagnosed as advanced stage. Regarding the staging of skin involvement in diagnosed MF/SS patients, 50% were classified as T4 (erythroderma), 17% were T3 (tumor(s)), 17% were T2 (generalized patch/plaque (>10% skin surface), and 17% were T1 (limited patch/plaque (<10% skin surface).

Sample Collection

Microbiome samples were obtained from patients by separately swabbing the selected sites: bilateral nares, right and left lower back, and bilateral posterior thighs. These sites were selected because CTCL skin lesions have a predilection for covered areas of the trunk and limbs (bathing trunk distribution). The left buttock was collected in place of the left lower back for one patient (Pt6), but this sample was excluded from downstream analyses. When a patient had both lesional and nonlesional skin at a given site, two separate samples (affected and unaffected) were

collected from that site. For healthy volunteers, swab samples were collected from all sites (nares, right and left lower back, right and left thigh) for HV1-5 and from the lower back only for HV6-10. Yeast cell lysis buffer solution (MPY80200,100 µL, Lucigen, Middleton, WI, USA) was sterilely aliquoted into individual sample tubes (022600044, Eppendorf, Mississauga, ON, Canada). A sterile foam-tipped swab (Puritan, 25-1506 1PF, Guilford, ME, USA) pre-moistened with yeast cell lysis buffer solution from the sample tube was brushed vigorously over the skin for 20 seconds while rotating the swab (Naik et al., 2020). Each swab was returned to the individual sample tube, trimmed, and placed immediately on dry ice. Negative air control swabs were collected during each sampling session and subsequently processed and analyzed alongside experimental swabs. The samples were stored at -80C until processed.

Sample Processing and Sequencing

DNA was isolated from samples as described previously (Oh et al., 2014), and Illumina libraries were created for metagenomic sequencing using Nextera library preparation. Libraries were sequenced with 2 X 125 bp paired-end reads on an Illumina NovaSeq 6000 at the NIH Intramural Sequencing Center core facility with a target of 15 million to 50 million clusters. Reagent and sequencing controls were used. From 67 total samples, we obtained 1.46 billion (182 Gbp) non-human, quality-filtered paired-end reads (median 16.9 million reads (2.11 Gbp)) per sample (Table S2).

Reference-based Taxonomic Classification

Taxonomic classifications were performed as described previously (Oh et al., 2014). Sequence data were processed to remove low quality reads and reads mapping to the GRCh38 human genome reference without Epstein Barr virus (chrEBV). The remaining reads were mapped against a database of 2356 bacterial, 395 fungal, 4695 viral and 67 archaeal reference genomes using Bowtie2 (version 2.3.5.1) --very-sensitive parameter (Langmead and Salzberg, 2012). Pathoscope (version 1.0) was used to reassign reads that mapped to multiple genomes (Francis et al., 2013), without normalization by genome length. Species-level taxonomic classifications were used for microbial community analyses (Table S3, S4, S5).

Figure Generation and Statistical analysis

R software and GraphPad Prism version 8.4.3 were used for generating plots and for statistical analyses. Lower back, thigh, and nares samples were analyzed separately. Due to some patients having only lesional skin at the time of sampling (particularly in the advanced stage patients), all analyses of lower back and thigh skin focused on comparisons between lesional patient skin and healthy volunteer skin, limiting each analysis to one sample per subject. One healthy volunteer (HV4) was found to be an outlier with >75% relative abundance of a plant Gammaproteobacteria (*Xanthomonas campestris*) on the skin, which was consistent with the HV's clinical history of extended work with international environmental vegetation, and was therefore removed from subsequent analyses. Data are represented as mean ± SD unless otherwise indicated. The non-parametric Kruskal-Wallis test was used to test for statistically significant differences in the relative abundance of bacterial species between HVs, MF, and SS patients. Statistical significance was recognized at an alpha level of P value ≤ 0.05. Diversity of microbial

communities was evaluated using the Shannon diversity index, a metric that reflects the number of species present and the proportional abundances of different species within a community. To determine the differences between bacterial communities by clinical stage, the distance between samples was measured using the Bray-Curtis dissimilarity index.

REFERENCES:

Francis OE, Bendall M, Manimaran S, Hong C, Clement NL, Castro-Nallar E, et al. Pathoscope: species identification and strain attribution with unassembled sequencing data. Genome Res 2013;23:1721-9.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-9.

Naik HB, Jo JH, Paul M, Kong HH. Skin Microbiota Perturbations Are Distinct and Disease Severity-Dependent in Hidradenitis Suppurativa. J Invest Dermatol 2020;140:922-5.e3.

Oh J, Byrd AL, Deming C, Conlan S, NISC Comparative Sequencing Program, Kong HH, et al. Biogeography and individuality shape function in the human skin metagenome. Nature 2014;514:59-64.

SUPPLEMENTARY FIGURES AND LEGENDS

Figure S1

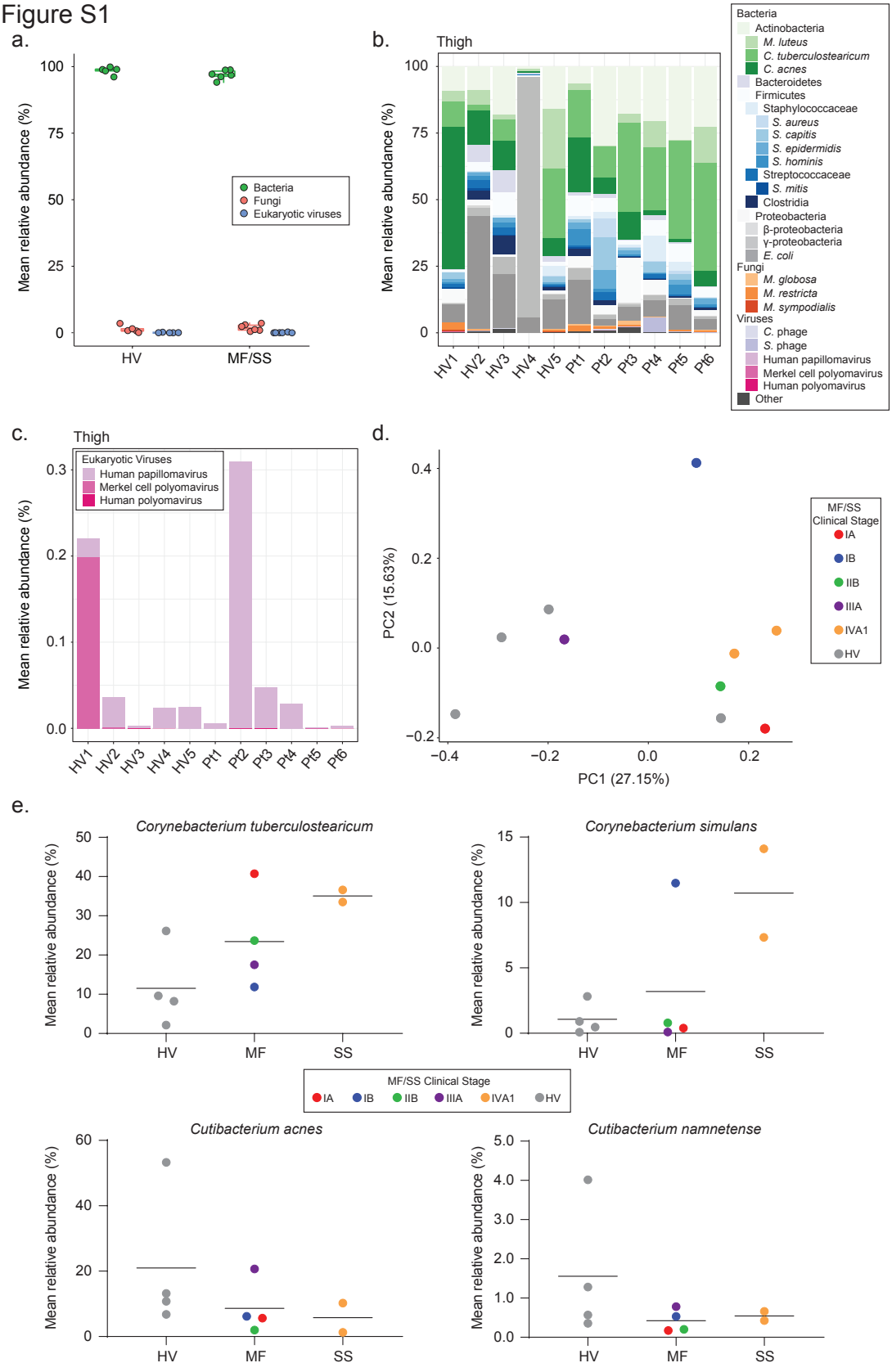

Figure S1. Bacterial shifts are evident in thigh skin microbiomes of MF/SS patients as compared to healthy adults. (a) Relative abundances of thigh skin microbes classified by kingdom in HVs (n=5) and MF/SS patients (n=6). Boxplot with each dot corresponding to one subject and representing the mean relative abundance (%) of one thigh sample per subject (lesional skin for patients; healthy skin for HVs). For each box, the central line represents the median, lower and upper edges represent the 1st and 3rd quartile, and whiskers represent values up to 1.5 times the interquartile range. Shotgun metagenomics microbial reads were mapped to a multi-kingdom reference database containing 2356 bacterial, 395 fungal, 4695 viral and 67 archaeal reference genomes using Bowtie2. **(b)** Mean relative abundances (%) of the major microbial taxa detected across kingdoms in thigh skin samples of HVs compared to MF/SS. In the barplot, colors represent distinct taxa. Due to the unusually high relative abundance (>75%) of the plant Gammaproteobacteria (*Xanthomonas campestris*) on the skin of HV4, this subject was removed from downstream analyses. **(c)** Relative abundances (%) of eukaryotic viruses on the thigh skin of HVs and MF/SS patients. **(d)** Principle coordinates analysis (PCoA) of thigh skin bacterial communities in HVs and MF/SS. The distance between samples was measured using Bray-Curtis dissimilarity index. Each dot corresponds to one thigh sample per subject. HVs (n=4) are represented by gray dots. MF/SS patients are color-coded by clinical stage IA (n=1), IB (n=1), IIB (n=1), IIIA (n=1), IVA1 (n=2). Clustering of dots in the PCoA indicates higher similarity between skin bacterial communities. **(e)** Mean relative abundances (%) of common cutaneous bacteria (*Corynebacterium* spp. and *Cutibacterium* spp.) on the thigh skin of HVs (n=4) compared to MF (n=4) and SS (n=2) patients. Each dot corresponds to one subject and represents the relative abundance in one representative thigh sample. Dots are color-coded as in (d). Black horizontal lines represent the mean. HV, healthy volunteer; MF, mycosis fungoides; SS, Sezary syndrome.

Figure S2

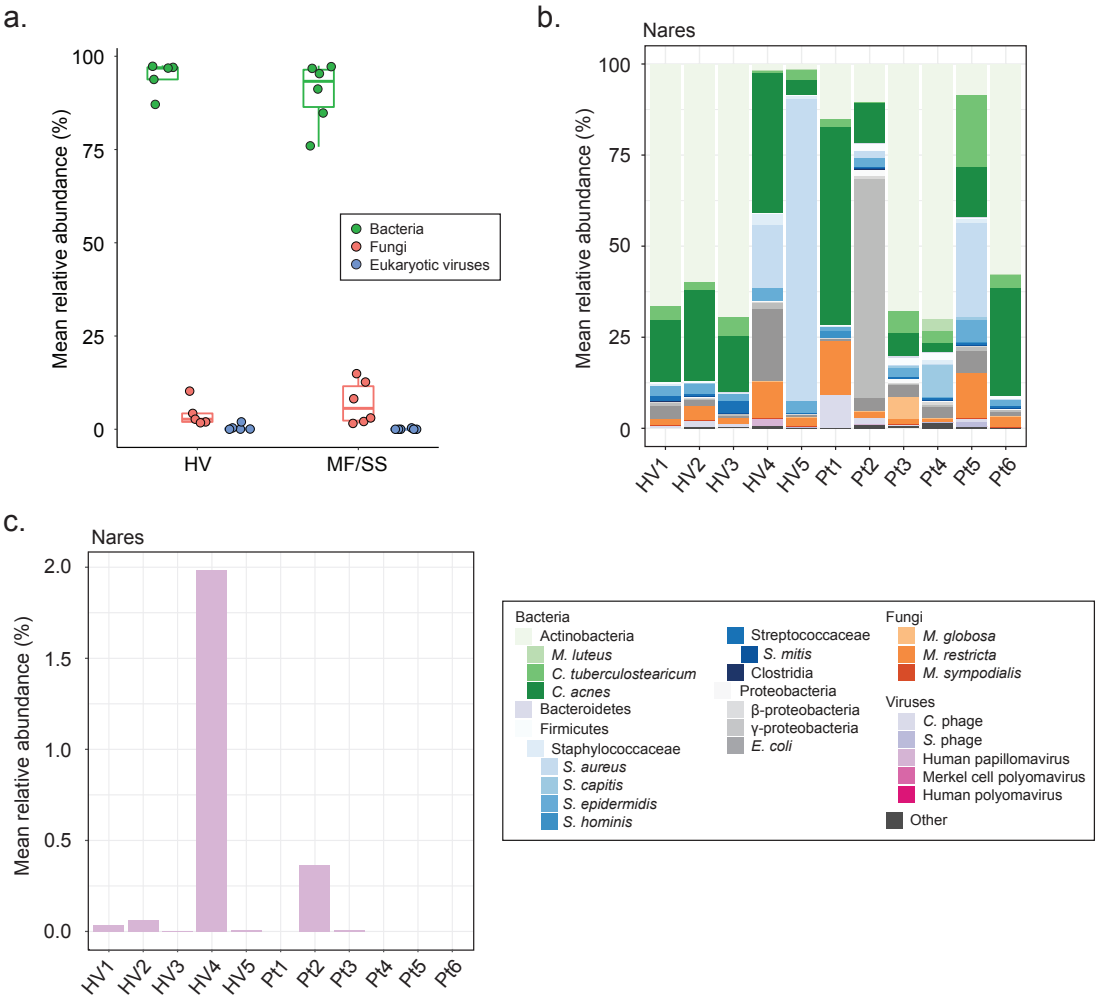

Figure S2. Eukaryotic viruses represent minor components of the nasal microbiome in both MF/SS and healthy controls. (a) Relative abundances of microbes classified by kingdom in the nares of HVs and MF/SS patients. Boxplot with each dot corresponding to one subject and representing the mean relative abundance (%) of nares samples. Shotgun metagenomics microbial reads were mapped to a multi-kingdom reference database containing 2356 bacterial, 395 fungal, 4695 viral and 67 archaeal reference genomes using Bowtie2. **(b)** Mean relative abundances (%) of the major microbial taxa detected across kingdoms in nares samples of HVs compared to MF/SS. In the barplot, colors represent distinct taxa. **(c)** Relative abundances (%) of eukaryotic viruses in the nares of HVs and MF/SS patients. HV, healthy volunteer; MF, mycosis fungoides; SS, Sezary syndrome.

Figure S3

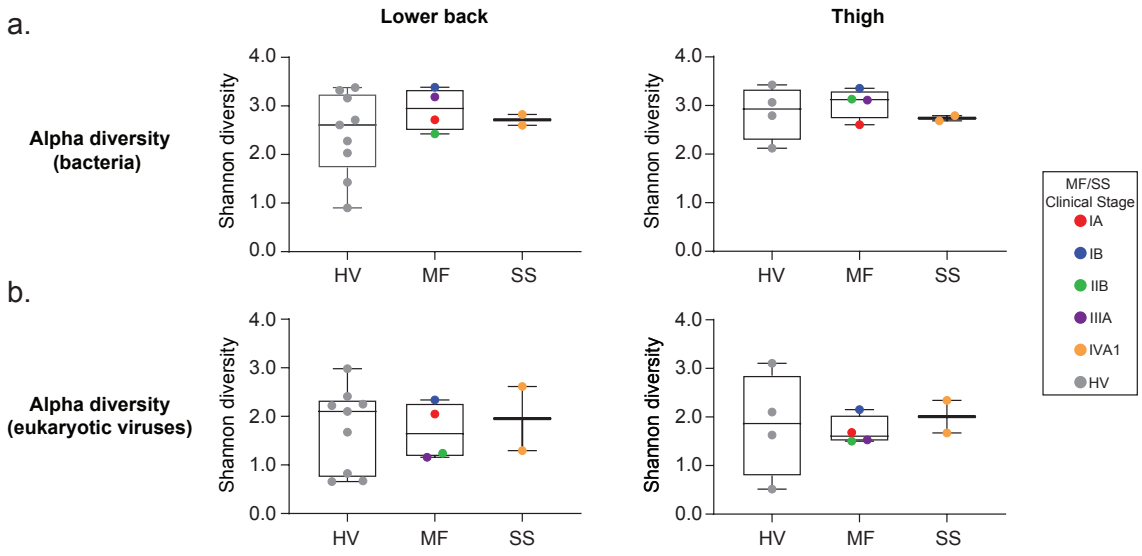

Figure S3. Alpha diversity of skin microbial communities is similar between MF/SS and healthy controls. Richness and evenness of bacterial **(a)** and viral **(b)** communities on the skin of HVs, MF, and SS patients as measured by Shannon diversity index. Lower back and thigh samples were analyzed separately, and each dot corresponds to one subject. HVs are represented by gray dots (lower back, n=9; thigh, n=4). MF/SS patients are color-coded by clinical stage IA (n=1), IB (n=1), IIB (n=1), IIIA (n=1), IVA1 (n=2). For each box, the central line represents the median, lower and upper edges represent the 1st and 3rd quartiles, and whiskers represent the minimum and maximum values. HV, healthy volunteer; MF, mycosis fungoides; SS, Sezary syndrome.

Figure S4

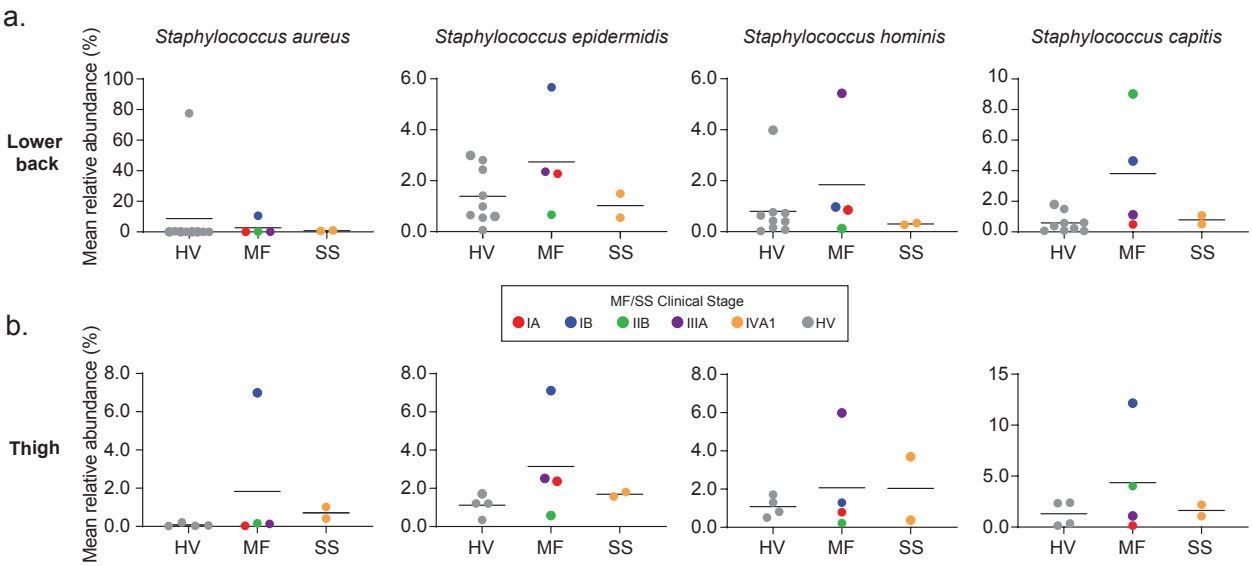

Figure S4. Commensal staphylococci trend higher in MF patient skin versus HV skin. Mean relative abundances (%) of common cutaneous *Staphylococcus* spp. on the lower back **(a)** and thigh **(b)** skin. Each dot corresponds to one sample per subject. HVs (lower back, n=9; thigh, n=4) are represented by gray dots. MF/SS patients are color-coded by clinical stage IA (n=1), IB (n=1), IIB (n=1), IIIA (n=1), IVA1 (n=2). Black horizontal lines represent the mean. HV, healthy volunteer; MF, mycosis fungoides; SS, Sezary syndrome.

SUPPLEMENTARY TABLES

Table S1. Patient and healthy volunteer demographics

Table S2. Sequencing statistics for all DNA samples

Table S3. Bacterial relative abundances

Table S4. Viral relative abundances

Table S5. Fungal relative abundances