LIST OF SUPPLEMENTARY MATERIALS

- Supplementary Material M1. Study design and participants
- Supplementary Material M2. Sampling, DNA extraction and 16S rRNA sequencing
- Supplementary Material M3. Data processing and taxonomic classification
- Supplementary Material M4. Ecological and statistical analyses
- Supplementary Material R1. Alpha diversity results not reported in main text
- Supplementary Material R2. Beta diversity results not reported in main text
- Supplementary Figure S1. Chao1 alpha diversity plots
- Supplementary Figure S2. Multidimensional plots of the proximity matrix by randomForest
- Supplementary Figure S3. Correlation matrices for patient perilesional, patient contralateral, and control corresponding sites
- Supplementary Figure S4.Correlation matrices for sites not typically involved in BP from patient and control cohorts

SUPPLEMENTARY TABLES AS SEPARATE EXCEL FILES

- Supplementary Table S1. Alpha diversity summary statistics
- Supplementary Table S2. RDP SeqMatch scores (S_ab) for 370 representative ASV sequences
- Supplementary Table S3. Indicator species for sampling sites (order based on stat parameter)
- Supplementary Table S4. Summary statistics of all significant indicator ASVs (n = 370)
- Supplementary Table S5. Important components of random forest models
- Supplementary Table S6. Correlation coefficients for Spearman's correlation matrices

Supplementary Material M1: Study design and participants

BP patients and controls were recruited from 14 study centers including Dresden, Düsseldorf, Freiburg, Homburg, Kiel, Lübeck, Munich, Würzburg, all Germany; Reims, Rouen, both France; Sofia, Bulgaria; Thessaloniki, Greece; and Oulu, Finland between October 2015 and August 2019. All BP patients were diagnosed according to national and international guidelines and had (i) a compatible clinical picture, (ii) linear deposits of IgG and/or C3 along the dermal-epidermal junction by direct immunofluorescence of a perilesional skin biopsy, and iii) serum IgG reactivity against the epidermal side of human salt-split skin or BP180 NC16A ELISA (Feliciani et al., 2015; Schmidt et al., 2015). Only BP patients with newly diagnosed or relapsed BP without previous systemic treatment were included. Exclusion criteria included systemic treatments, topical therapies for a duration of more than seven days prior to sampling, or topical antiseptic use within the last two weeks. None of the study participants had been exposed to systemic antibiotics for at least seven days prior to skin sampling. The average age of the BP study group was 80 years and consisted of 114 males, 112 females, and one participant with sex "unspecified." Ethnicity was assessed by the investigators according to the birth country of parents and grandparents. Age- and sex-matched patients with non-inflammatory/ non-infectious dermatoses (mostly basal cell carcinoma or squamous cell carcinoma) with an average age of 80 years and containing 104 males and 86 females, served as controls. Age- and sex-matched controls were sought out so that each BP subject would receive their own paired contemporary control in the minimal amount of time after microbiota sampling, which was successful in the majority of cases (controls per case for males and females are 0.91 and 0.77, respectively).
The study was performed according to the Declaration of Helsinki and was approved from the ethics committee, University of Lübeck (15-051, 18-046) as well as the ethics committees of the

individual study centers. Written, informed consent was obtained each BP patient and control subject.

Supplementary Material M2: Sampling, DNA extraction, and 16S rRNA sequencing

Samples were collected using Epicentre Illumina collection swabs (Madison, WI, USA) immersed in 600 uL (50 mM Tris, 1 mM EDTA, 0.5% Tween-20) buffer (Teknova, United States). The swabs were rubbed across the selected body site for 30 seconds and then placed back into the buffer solution. Immediately after swabbing, swabs were stored at -80°C until further processing.

Skin samples (n = 2,956) were obtained from BP patients representing different cutaneous microenvironments, including "perilesional" skin (defined as being within 2 cm of a primary BP lesion, i.e. a fresh blister or erosion), unaffected skin at the same anatomical location on the contralateral side of the patient (referred to as "contralateral"), and unaffected skin in areas that do not typically manifest disease (we selected the forehead, antecubital fossa, and upper back, collectively referred to as "sites typically unaffected by BP"). Two separate BP lesions, in different locations, were sampled from each patient to account for differences in skin biogeography. Control participants were swabbed at locations that approximated the sampled body sites in the BP patients (referred to as "corresponding sites"), in addition to the three sites typically unaffected by BP (Figure 2a). Ambient air samples (n = 19), collected by holding a swab in the air for 30 seconds and then placing the collection swab directly into the buffer solution, served as negative sampling controls in addition to negative extraction controls (n = 43). Negative controls were processed alongside samples.

ZymoBIOMICS Microbial Community Standard cells (Zymo Research) were used as extraction and sequencing controls to assess contamination in downstream analyses, following

the mock community dilution series protocol as described by Karstens et al, 2019. In brief, the strategy is based on the logic that with decreasing "true" microbial biomass (i.e., skin microbes or mock cells), potential signal from background/contamination introduced throughout the procedure will increase. All mock dilutions, as well as the undiluted mock community standards, were treated as samples throughout the extraction, PCR, sequencing, and data processing steps. Swabs were thawed overnight at 4℃, then vortexed at high speed for 1 minute. Afterwards, tubes were centrifuged at high speed for 15 min, and the pellets were resuspended in Power Bead solution. Subsequently, DNA was extracted using the Qiagen DNeasy UltraClean 96 Microbial Kit [96-well plate] (Germantown, MD, USA), according to the manufacturer's instructions, and eluted in 50uL of the elution buffer. Negative extraction controls were included for each 96-well plate. Samples were stored at -20℃ until further processing. PCR and sequencing were performed by implementing the dual-index sequencing strategy for amplicon sequencing on the MiSeq Illumina sequencing platform, as previously described (Kozich et al., 2013). Hypervariable regions V1-V2 of the bacterial 16S rRNA gene were amplified. The primer pair 5′-AATGATACGGCGACCACCGAGATCTACACXXXXXXXXTATGGTAATTGT AGAGTTTGATCCTGGCTCAG -3′ and 5′-CAAGCAGAAGACGGCATACGAGATXXXXXXXXAGTCAGTCAGCC TGCTGCCTCCCGTAGGAGT-3′ contained the Illumina P5 (forward) and P7 (reverse) sequences (denoted by italics) and the universal primer 27F and 338R sequences (denoted by underlined italics), which represent broadly conserved regions of the bacterial 16S gene. Per Illumina's recommendations, a twelve-base linker sequence (underlined only) was added to increase the annealing temperature of the sequencing primer. Finally, both primers contained a unique eight base multiplex identifier (designated by XXXXXXXX) to tag each PCR product.

PCR was conducted in a 25-uL volume containing 2uL DNA using Phusion Hot Start II DNA High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland). The cycling conditions were as follows: initial denaturation for 30 seconds at 98℃; 30 cycles of 9 seconds at 98℃, 60 seconds at 50℃ and 90 seconds at 72℃; final extension for 10 minutes at 72℃. PCR products were loaded on a 1.5% agarose gel to confirm and quantify the 16S rRNA gene bands. PCR product concentrations were first quantified on an agarose gel using image analysis software (Bio-Rad). After quantification, products were combined accordingly to make equimolar subpools. The subpools were then extracted from an agarose gel using the Qiagen MinElute Gel Extraction Kit and quantified with the Quant-iTTM dsDNA BR Assay Kit on a Qubit fluorometer (Invitrogen). Finally, subpools were combined into one equimolar pool for each library. Pools were further purified using AMPure® Beads (Agencourt), and libraries were run on an Agilent Bioanalyzer prior to sequencing, as recommended by Illumina. The libraries were sequenced on a MiSeq using the MiSeq Reagent Kit v3 600 cycle sequencing chemistry (Illumina, CA, USA).

Supplementary Material M3: Data processing and taxonomic classification

Raw fastq sequencing reads were generated and demultiplexed to allow for no mismatches in the index sequences (Bcl2fastq, Illumina). Data processing was performed in the R software environment (version 4.0.2) using the DADA2 (version 1.16.0; (Callahan et al., 2016) workflow resulting in abundance tables of amplicon sequence variants (ASVs). All sequencing runs underwent quality control and error profiling separately. Forward and reverse reads were trimmed to a length of 250 and 200 bp, respectively, using standard filtering parameters. Low quality read-pairs were discarded when the estimated error in one of the reads exceeded 2 or if ambiguous bases (“N”s) were present in the base sequence. Read pairs that could not be merged due to insufficient overlap or mismatches in their nucleotide sequences were discarded. All data

from the independent sequencing runs were finally combined into a single abundance table, and then underwent chimera removal in DADA2. Taxonomic assignment of ASVs was completed in DADA2 with the Bayesian classifier using NR Silva database training set version 138 (Quast et al., 2013). Merged reads that were less than 270 bp or more than 330 bp were excluded. ASVs classified as eukaryotic were excluded.

The Decontam package (version 1.8.0; (Davis et al., 2018)) was used within Phyloseq (version 1.32.0; (McMurdie and Holmes, 2013) to identify potential contaminant ASVs. The prevalence method, which uses presence/ absence information of sequence features for all samples to compare the prevalence of ASVs in negative controls compared to true samples, was used with the strict probability threshold of 0.5 (Karstens et al., 2019). ASVs tagged as probable contaminants were removed from the abundance table. Furthermore, and as suggested by Karstens et al., an additional filtering step was performed as follows: ASVs present/detected in negative controls (sampling ambient air and extraction controls) for a respective sample that harbored an absolute abundance of less than or equal to 2 percent were excluded from the skin sample, allowing for potential cross-contamination between samples to be accounted for (i.e. true ASVs are less likely to be *falsely* labeled as a contaminant if their abundance is greater than 2 percent). All ASVs classified as belonging to the families Halomonadaceae (n = 1,040) and Shewanellaceae (n = 211) were removed, as these bacteria represent common contaminants in low biomass samples (Weyrich et al., 2019).

To normalize sequencing coverage random sub-sampling to 5000 sequences per sample was conducted in the R working environment (version 4.0.2), according to the procedure described in Belheouane et al. (2020), which further helped to remove inconsistent/spurious ASVs. Phylum to genus-level abundance tables were constructed in the R working environment.

The final sample size included 2319 skin swabs comprising 1451 patient swabs with 868 matched control subject swabs from a total of 417 individuals (Figure 1a).

Representative 16S rRNA gene sequences were queried via the Ribosomal Database Project (RDP; release 11.6; Wang et al., 2007) SeqMatch (version 3;(Cole et al., 2005). Results reported in Supplementary Table S2 represent classification based on the highest RDP match score (S_ab), which is the number of unique 7-base oligomers shared between the query sequence and a given RDP sequence for both type- and non-type strains. Individual profile tables were written to the genus-level.

Supplementary Material M4: Ecological and statistical analyses

Statistical analyses were performed in R (version 4.0.2). Alpha diversity (within sample diversity) was measured using Shannon and Chao1 indices, which account for species richness and evenness, respectively, in vegan (version 2.5-6) on absolute abundances data. Groups were compared using Kruskal-Wallis and pairwise Wilcoxon rank sum tests. Correction for multiple testing was performed according to Benjamini and Hochberg method (1995). Overall differences between groups (beta diversity) were assessed using the Bray-Curtis dissimilarity index. Indicator species analysis was applied in indicspecies using the "r.g." function (De Caceres and Legendre, 2009) and 10^5 permutations. Four separate indicator analyses were conducted, including, i) the control group (n = 868) versus the patient group (n = 1,451) for all body site sampling locations, ii) the corresponding sampling sites in controls (n = 334) versus the combined patient perilesional (n = 415) and contralateral sites (n = 407) and , iii) the control group versus the patient group for all non-lesional sites (antecubital fossa = 363, upper back = 383, and forehead = 398), and iv) perilesional (n = 415) versus contralateral (n = 407) sampling sites only within the patient group. For each indicator analysis, ASVs were defined within each

sample with a minimum threshold of 2%. The ASVs which met this criteria for all samples were then included in the analysis. Significant indicator ASVs were selected after correction of *p*-values for multiple testing using the Benjamini and Hochberg method (1995).

To further evaluate indicator species results, we performed Random Forest classification and regression analyses using randomForest (version 4-6-14; (Liaw and Wiener, 2002). Two sample sets, i) all controls (n = 868) and patient (n = 1,451) samples for all body sites, and ii) corresponding (n = 334) control- and patient perilesional (n = 415), and contralateral (n = 407) sites, were assessed using two distinct models: i) significant ASVs within each indicator species analysis, and ii) ASVs classified to the *Staphylococcus* genus whose abundance met the 2% minimum read threshold (as described above). Models were constructed with $n = 10^5$ trees, while the "mtry" parameter was set for each model.

To disentangle the effects of disease, individual features, and sampling variation on the identified indicator ASVs and diversity measures, we constructed linear models that included the following explanatory variables: disease status (control or BP patient), body site, study center, sex, age, sequencing run, and DNA extraction round. Two respective models were constructed, including one for site typically unaffected by BP (antecubital fossa, upper back, and forehead) and one for perilesional, contralateral, and control corresponding sites. Response variables included significant indicator ASVs that met statistical significance in the indicator analyses. When necessary, square root transformations were applied on response variables to correct for a non-normal distribution. Models were checked and validated by examining the distribution of residuals, plotting the fitted and residual values, and examining the distribution of residuals for included explanatory variables. We reported the R^2 and the Beta coefficient value, which expresses the directionality of the effect (the degree of change in the response variable).

Moreover, we performed a non-parametric multivariable analysis of variance using distance matrices (PERMANOVA) using the "adonis" function with 1,000 permutations and a partial constrained principal coordinate analysis (PCoA) of beta diversity measures using the "capscale" function in vegan (Oksanen et al., 2005). Disease status, body site, and study center were included as constraints, whereas sex, age, sequencing run, and extraction round were defined as conditions. The significance of models, axes, and terms were assessed using the "anova.cca" function with 1,000 permutations. These analyses were performed respectively for each of the two models described above.

Pairwise correlations among best indicator ASVs were evaluated using Spearman's correlations, and P values were corrected for multiple testing using Benjamini and Hochberg method (1995).

Supplementary Material R1: Alpha diversity results not reported in main text

We assessed species diversity (alpha diversity) in each sampling category at the ASV-level to investigate the effects of disease state and skin biogeography. Shannon diversity takes into account both the richness (number of different species) and evenness (how evenly the species are distributed) of the bacterial community, whereas the Chao1 index reflects expected species richness. We closely inspected alpha diversity with regard to collected metadata and we find that disease status, study center, and sex significantly correlate with Shannon diversity for control corresponding, patient perilesional, and patient contralateral sites. These results are reported in main text.

We similarly examined the influence of disease status, body site, sex, age, sequencing run, DNA extraction round, and study center from patients and controls for sites typically unaffected by BP (forehead, upper back, and antecubital fossa). We find that the body site and

study center significantly correlate with Shannon diversity ($F_{39,1123} = 6.689$; $R^2_{adj} = 0.1603$, $p < 0.001$). Analysis of sum of squares reveal that body site explains a considerable fraction of variance (5.57%), whereas sample origin accounts for a comparably smaller fraction of the variance (2.5%). Interestingly, sex explains the highest percentage of variance (6.6%), while disease status is not significant.

Similar to Shannon diversity, differences in Chao1 index for these same body sites significantly correlate with study center (4.9%), sequencing run (4.5%), and to a lesser extent, body site (0.5%) and sex (1.19%), while disease status is not significant ($F_{39,1123} = 4.149$; $R^2_{adj} = 0.09559$; $p < 0.001$). Thus, overall, our statistical framework further emphasizes that the effects of disease on alpha diversity do not extend to non lesional sites, and that rather the signal of skin biogeography predominates at these sites.

Supplementary Material R2: Beta diversity results not reported in main text

Beta diversity accounts for overall community compositional differences between samples, which we used to evaluate the effects of disease, body site, study center, sex, age, sequencing run, and DNA extraction round. We examined the sites typically unaffected by BP (forehead, upper back, and antecubital fossa) using PERMANOVA analysis applied to the Bray-Curtis index as measure of beta diversity and we find that disease status, body site, and study center influence bacterial community structure, in addition to sex, age, sequencing run, and extraction round (Supplementary Material M4; adonis: disease status $R^2 = 0.0109$, body site $R^2 = 0.044$, study center $R^2 = 0.03$, sex $R^2 = 0.02$, age $R^2 = 0.001$, sequencing run $R^2 = 0.01$, and extraction round $R^2 = 0.01$; $p < 0.05$; 1,000 permutations).

We analyzed beta diversity to examine potential differences in community structure for perilesional, contralateral, and control corresponding body sites as described above. We observe

a significant association with disease and blistering status; results reported in main text. The observed association with study center is reported in main text. Associations with other sampling variables included: disease status R^2=0.0212, blistering status R^2 = 0.0029, study center R^2 = 0.0341, sex R^2 = 0.013, age R^2 = 0.001, sequencing run R^2 = 0.02, and extraction round R^2 = 0.016 (Supplementary Material M4, adonis: $p < 0.05$; 1,000 permutations).

REFERENCES

Belheouane M, Vallier M, Čepić A, Chung CJ, Ibrahim S, Baines JF. Assessing similarities and disparities in the skin microbiota between wild and laboratory populations of house mice. ISME J 2020. https://doi.org/10.1038/s41396-020-0690-7.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 1995;57:289–300.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13:581–3. https://doi.org/10.1038/nmeth.3869.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, et al. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res 2005;33:D294-296. https://doi.org/10.1093/nar/gki038.

Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 2018;6:226. https://doi.org/10.1186/s40168-018-0605-2.

De Caceres M, Legendre P. Associations between species and groups of sites: indices and statistical inference. Ecology 2009.

Feliciani C, Joly P, Jonkman MF, Zambruno G, Zillikens D, Ioannides D, et al. Management of bullous pemphigoid: the European Dermatology Forum consensus in collaboration with the European Academy of Dermatology and Venereology. Br J Dermatol 2015;172:867–77. https://doi.org/10.1111/bjd.13717.

Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, et al. Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. MSystems 2019;4:e00290-19, /msystems/4/4/mSys.00290-19.atom. https://doi.org/10.1128/mSystems.00290-19.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl Environ Microbiol 2013;79:5112–20. https://doi.org/10.1128/AEM.01043-13.

Liaw A, Wiener M. Classification and Regression by randomForest. Newsl R Project 2002;2:5.

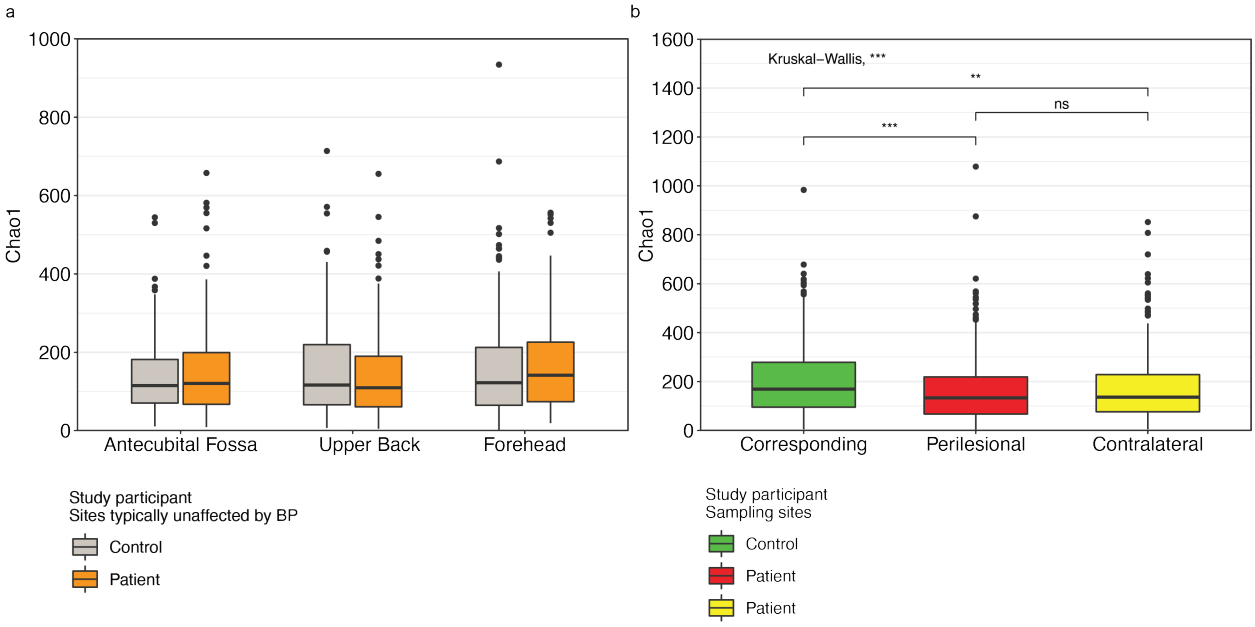
McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PloS one, 2013. https://doi.org/10.1371/journal.pone.0061217.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Community Ecology Package. 2005.

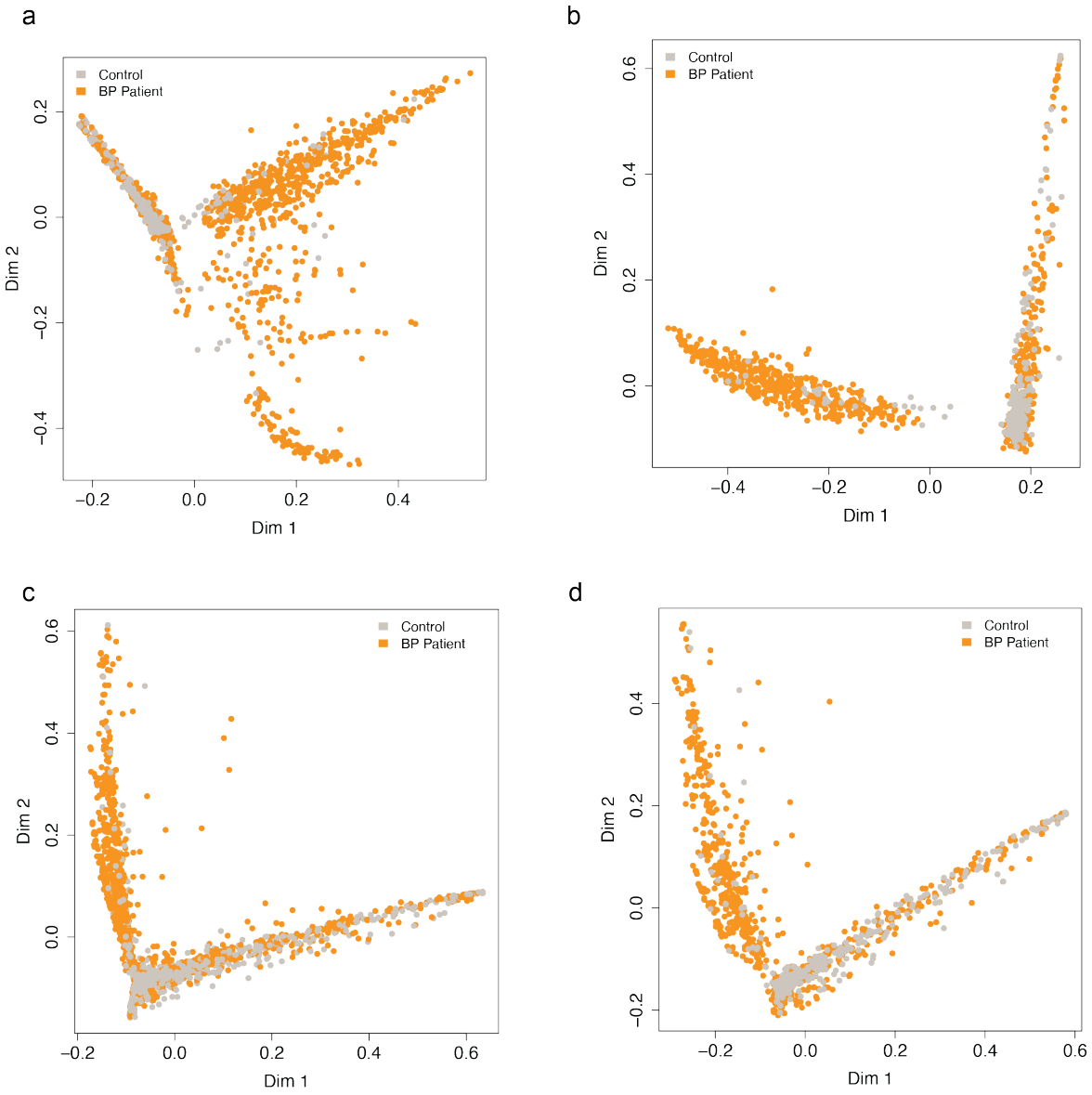
Schmidt E, Goebeler M, Hertl M, Sárdy M, Sitaru C, Eming R, et al. S2k guideline for the diagnosis of pemphigus vulgaris/foliaceus and bullous pemphigoid. J Dtsch Dermatol Ges 2015;13:713–27. https://doi.org/10.1111/ddg.12612.

Weyrich LS, Farrer AG, Eisenhofer R, Arriola LA, Young J, Selway CA, et al. Laboratory contamination over time during low-biomass sample analysis. Molecular Ecology Resources 2019;19:982–96. https://doi.org/10.1111/1755-0998.13011.

SUPPLEMENTARY FIGURES

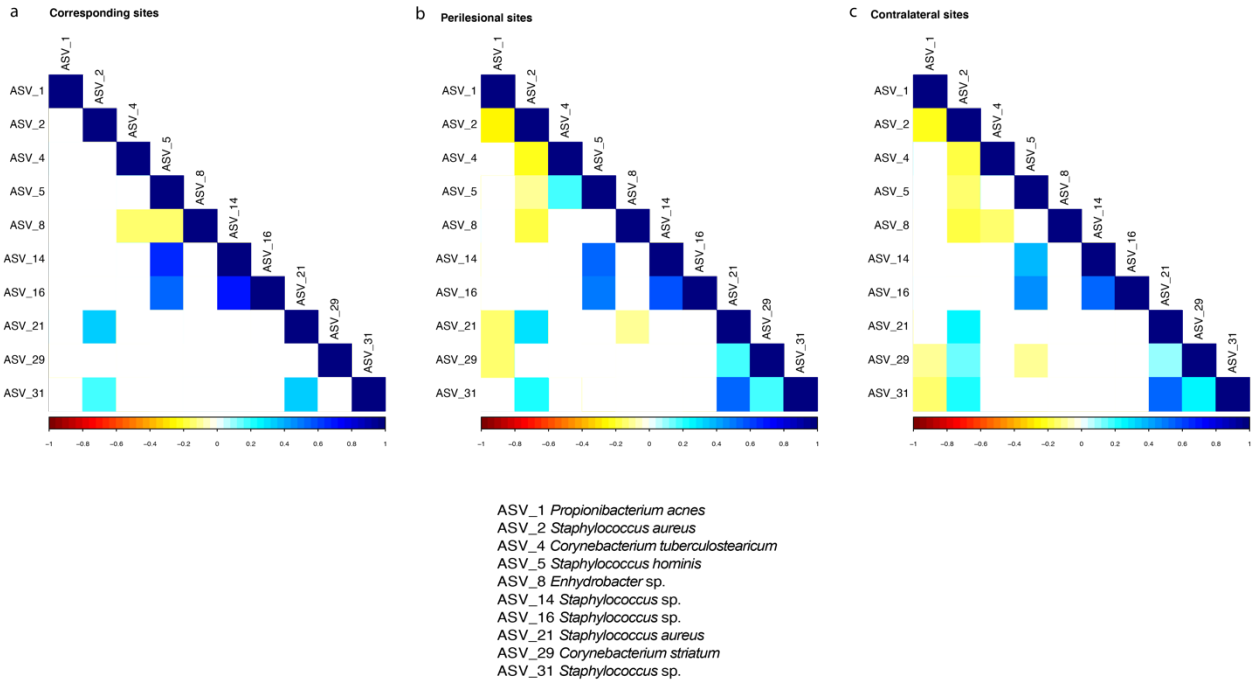


Supplement Figure S1. S1a. Chao-1 index for sites typically unaffected by BP (forehead, upper back, and antecubital fossa) between patients and controls. S1b. Chao-1 index for the perilesional and contralateral sites of the patient group and corresponding site in controls. Summary statistics provided in Supplementary Table S1.

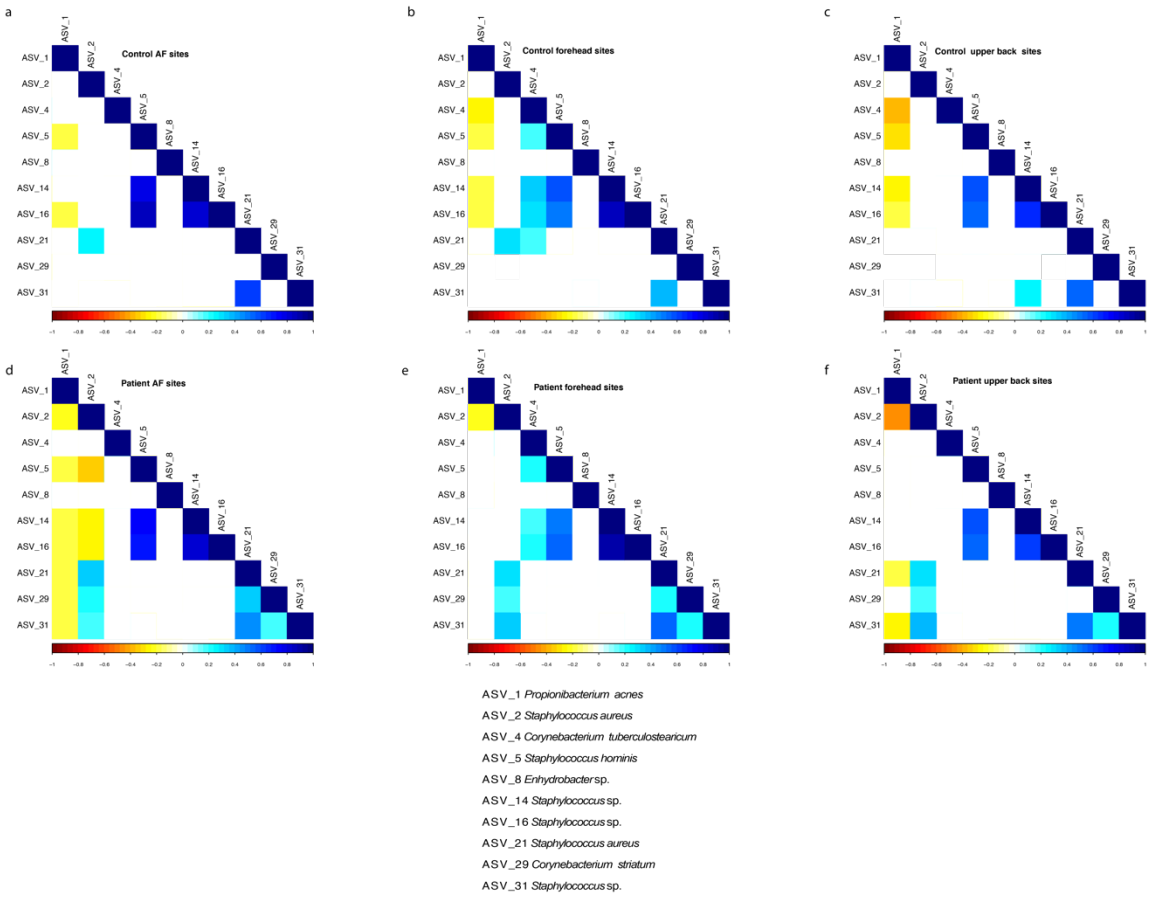


Supplementary Figure S2. Multidimensional plot of the proximity matrix by randomForest analysis for the four tested models (Methods). **S2a.** Controls versus BP patients for all identified indicators (n = 370) for all sampling sites (mtry = 15); (849/868 controls, and 1443/1451 BP patients, mean classification accuracy 99%) **S2b.** Controls versus BP patients for all identified indicators (n =3 70) for only control corresponding and patient perilesional and contralateral sampling sites (mtry = 18); (324/334 controls, and 822/822 BP patients, mean classification

accuracy 99.15%). **S2c.** controls versus BP patients for all sampling sites, but limited indicators to *Staphylococcus* ASVs with an abundance >2% within each sample; (mtry = 52, 790/868 controls, and 1,446/1,451 BP patients, mean classification accuracy 96.4%). **S2d.** controls versus BP patients for control corresponding, patient perilesional, and patient contralateral sampling sites and limited to *Staphylococcus* ASV indicators with an abundance >2% within each sample (mtry = 62, controls 294/334, and 819/822 BP patients, 96.2%).



Supplementary Figure S3. Spearman's correlation matrix between **S3a**. Indicator ASVs and control corresponding sampling sites, **S3b.** Indicator ASVs and patient perilesional sampling sites, and **S3c.** Indicator ASVs patient contralateral sampling sites. Blank box indicates $p > 0.05$ after correction for multiple testing.



Supplementary Figure S4. Spearman's correlation matrix between, **S4a.** Indicator ASVs and antecubital fossa (AF) sites from matched controls, **S4b.** Indicator ASVs and forehead sites from matched controls, **S4c.** Indicator ASVs and upper back sites from matched controls, **S4d**. Indicator ASVs and antecubital fossa (AF) sites from BP patients, **S4e.** Indicator ASVs and forehead sites from BP patients, **S4f.** Indicator ASVs and upper back sites from BP patients. Blank box indicates $p > 0.05$ after correction for multiple testing.