

ARTICLE

Received 15 May 2013 | Accepted 19 Aug 2013 | Published 17 Sep 2013

DOI: 10.1038/ncomms3462

OPEN

Genome-wide mapping of gene-microbiota interactions in susceptibility to autoimmune skin blistering

Girish Srinivas^{1,2}, Steffen Möller², Jun Wang^{1,3}, Sven Künzel¹, Detlef Zillikens², John F. Baines^{1,3,*} & Saleh M. Ibrahim^{2,*}

Susceptibility to chronic inflammatory diseases is determined by immunogenetic and environmental risk factors. Resident microbial communities often differ between healthy and diseased states, but whether these differences are of primary aetiological importance or secondary to the altered inflammatory environment remains largely unknown. Here we provide evidence for host gene-microbiota interactions contributing to disease risk in a mouse model of epidermolysis bullosa acquisita, an autoantibody-induced inflammatory skin disease. Using an advanced intercross, we identify genetic loci contributing to skin microbiota variability, susceptibility to skin blistering and their overlap. Furthermore, by treating bacterial species abundances as covariates with disease we reveal a novel disease locus. The majority of the identified covariate taxa are characterized by reduced abundance being associated with increased disease risk, providing evidence of a primary role in protection from disease. Further characterization of these putative probiotic species or species assemblages offers promising potential for preventative and therapeutic treatment development.

¹Max Planck Institute for Evolutionary Biology, August-Thienemann-Street 2, D-24306 Plön, Germany. ²Department of Dermatology, University of Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, Germany. ³Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, Arnold-Heller-Street 3, D-24105 Kiel, Germany. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.B. (email: baines@evolbio.mpg.de).

Diverse communities of symbiotic bacteria inhabit nearly every surface of the body exposed to the environment. The skin in particular is in constant contact with the environment and serves a critical barrier function, yet provides a range of niches to inhabiting microbial communities. A multitude of interactions between the skin microbiota, host and environment contribute to community structure over space and time and its potential contribution to changes in health status^{1,2}. Recent landmark studies of the mouse gut microbiota using a quantitative trait locus (QTL) mapping approach unequivocally demonstrate the role of host genetics in shaping diversity between individuals^{3,4}. Likewise, inflammatory disorders afflicting the skin such as psoriasis and atopic dermatitis harbour clear immunogenetic components, but whether these associations may be mediated by alterations in microbial community structure is unknown^{1,5}.

Epidermolysis bullosa acquisita (EBA) is a chronic skin blistering disease of autoimmune origin characterized by antibodies to type VII collagen (COL7)⁶. We previously demonstrated the contribution of MHC haplotype and other non-MHC genes to EBA susceptibility in an immunization-induced model of EBA in mice⁷. Intriguingly, autoimmunity against COL7 is also a common observation among patients with Crohn disease⁸, one of two major forms of inflammatory bowel disease with clear host genetic and microbial components⁹. In this study, we investigate the contribution of host genetic control of the skin microbiota in mice from a four-way autoimmune-prone advanced intercross, enabling loci influencing microbial community structure and disease (EBA) susceptibility to be simultaneously analysed. In addition to identifying host genetic loci that contribute to variability in bacterial taxon abundances in the skin, we find that individual genotype-dependent microbial risk factors modify susceptibility to EBA and increase the power to detect disease-associated loci.

Results

Composition and diversity of skin microbiota. To measure the contribution of host genetics to variation in the mouse skin microbiota, we first analysed 261 individuals from the fourth generation of an advanced intercross lines (183 immunized individuals, of which 64 developed EBA, plus 78 non-immunized controls, Methods) using pyrosequencing of the 16S rRNA gene. At the phyla level, the Firmicutes are most abundant (54%), followed by Proteobacteria (21%), Actinobacteria (12%) and Bacteroidetes (6%) (Supplementary Fig. S1), revealing communities similar to those observed in previous studies of the skin^{2,10–12}. At the genus level, *Staphylococcus* (36%), *Corynebacterium* (9%) and *Ralstonia* (8%) are most abundant (Supplementary Fig. S2). To characterize the level and pattern of diversity within individuals, we applied different measures of alpha diversity, which focus on species richness, evenness and abundance. The Chao1 species richness index is higher in the healthy individuals compared with those afflicted with EBA (Fig. 1) (Wilcoxon rank-sum test, $W=170$, $P=0.005$). The same pattern is also observed for Faith's¹³ phylogenetic diversity index (PD whole tree) (Wilcoxon rank-sum test, $W=176$, $P=0.008$), the observed number of species (Wilcoxon rank-sum test, $W=212$, $P=0.05$) and the Shannon evenness measure (Wilcoxon signed-rank Test, $Z=-4.3726$, $P<0.001$) (Supplementary Fig. S3). To analyse bacterial community composition and structure *between* individuals (that is, beta diversity,) we first used the weighted and unweighted Unifrac metric^{14,15}, which is a phylogenetic-based measure weighted by taxon abundance and based on presence-absence information, respectively. Constrained analysis of principal coordinates (CAP) using EBA status as an explanatory variable and the weighted

Unifrac metric as a response variable also reveals a small, but significant effect ($P=0.015$), with the first principal component axis (or CAP1 axis) explaining 1.8% of the variation between individuals (Supplementary Fig. S4a), while the CAP1 axis for unweighted Unifrac explains 2% of the variation ($P=0.005$; Supplementary Fig. S4b). Analysis of beta diversity using OTU-based approaches yields very similar results (Bray–Curtis index, with CAP1 explaining 1.8% of the variation ($P=0.015$) (Supplementary Fig. S4c) and Jaccard index, with CAP1 explaining 1% of the variation ($P=0.001$) (Fig. 2)). For further analysis a 'core measurable microbiota' (CMM) of 131 OTUs was defined³, which

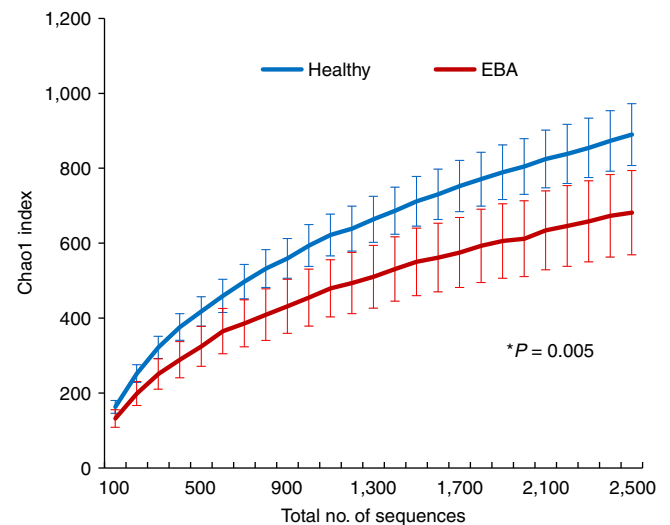


Figure 1 | Chao1 species richness. The Chao1 index based on species-level OTUs was estimated for immunized healthy ($n=119$) and EBA ($n=64$) samples. Error bars represent the 95% confidence interval. *Significance was determined by the Wilcoxon rank-sum test.

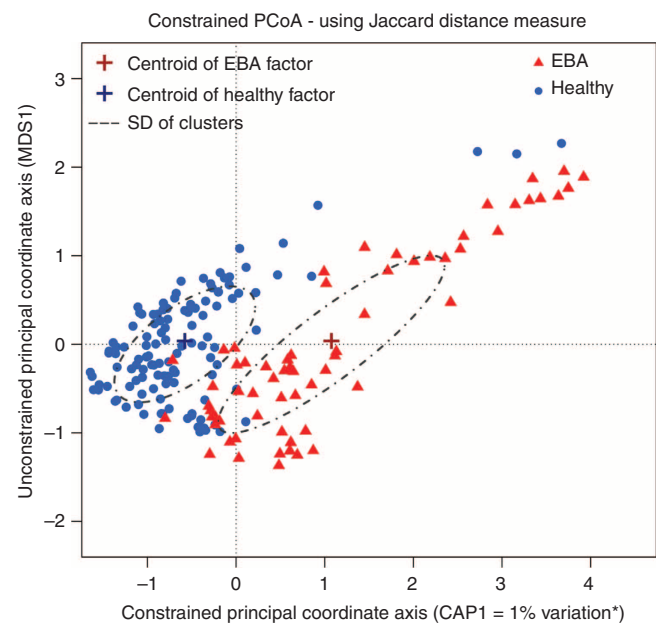


Figure 2 | Constrained analysis of principal coordinates of the Jaccard index. The Jaccard index was calculated for immunized healthy ($n=119$) and EBA ($n=64$) samples. The disease status was taken as the constrained factor and differed significantly by permutation test ($P=0.005$, see Methods).

contain nearly 80% of the total sequences in the data set (Supplementary Figs S5 and S6) (Methods).

QTL analysis of skin microbiota. To account for both environmental and genetic contributions to variation in CMM OTU abundances, a linear mixed model analysis was performed including cage, family and sex as factors (see Methods). This reveals significant cage and family effects accounting for 28% and 3% of the variation in CMM species abundances, respectively. No significant influence of sex is observed for any of the CMM taxa and was thus removed from the model. To measure the genetic contribution, CMM abundances were tested for co-segregation against 1,199 informative single-nucleotide polymorphism (SNP) markers after accounting for cage and family effects. This reveals host genetics to have significant control over members of the skin microbiota, which can be seen in Fig. 3. Nine out of 131 CMM OTUs are associated with three significant (*E*-value cutoff <0.05) and six suggestive (*E*-value cutoff <0.1) (see Methods) species-level OTU QTLs, hereafter referred to as ‘spQTLs’ (Supplementary Data 1). Mapping at higher taxonomic levels including phylum, class, order, family and genus reveals a total of six QTLs, including three with one or more significant associations and three with suggestive associations, hereafter referred to as ‘gpQTLs’ (Supplementary Data 2). Two out of the nine spQTLs are contained within gpQTLs, thus, in total thirteen unique QTLs are identified. To gain further

insight we compared our results with previously published QTL studies of the gut microbiota^{3,4} and reveal evidence of overlap greater than expected by chance (Fig. 3; Methods). Interestingly, the confidence intervals of our spQTLs and gpQTLs contain five and four genes related to innate immunity, respectively (see Discussion, Supplementary Data 3).

Effect of immunization and disease status on QTL mapping. As the model of EBA used in this study is immunization-based, we also included 78 non-immunized mice to control for the effect of immunization in the QTL mapping. Accordingly, we analysed a subset where both EBA-afflicted and non-immunized individuals are removed (that is, only the 119 healthy, autoimmunized samples). Despite decreasing the sample size from 261 to 119, two out of nine spQTLs and two out of six gpQTLs are still detected (Supplementary Data 1 and 2). Next, we analysed a subset where the EBA-afflicted mice are removed (that is, including 119 healthy autoimmunized and 78 non-immunized samples). One out of nine spQTLs and two out of six gpQTLs are still detectable despite lowering the sample size from 261 to 197 (Supplementary Data 1 and 2). Thus, the presence of QTLs among subsamples not influenced by differences in disease/autoimmunization status supports the presence of true genetic effects.

Gene-microbe covariation in disease susceptibility. To investigate the potential role of host genetic variation for skin bacterial

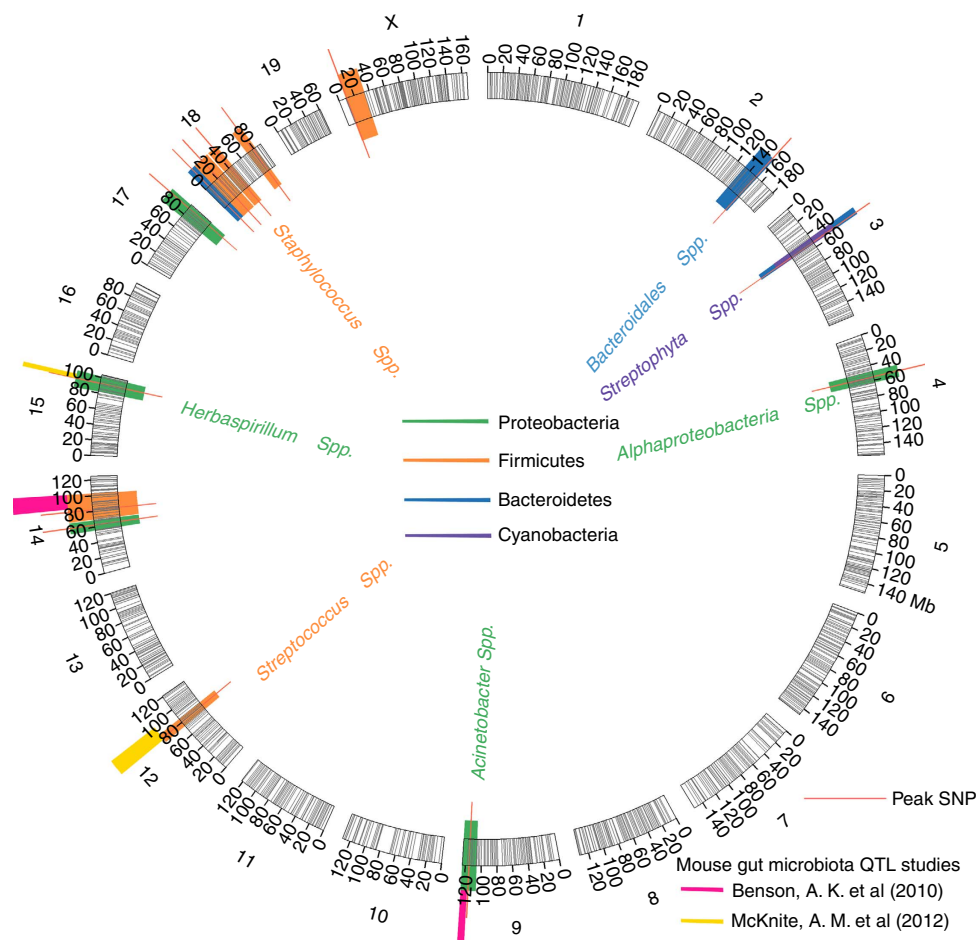


Figure 3 | QTL mapping of skin microbiota. Nineteen mouse autosomes and the X chromosome are depicted with 1,199 SNPs indicated by black lines. QTLs for species level OTUs (spQTLs) are colour coded according to their phylum along with their highest reliable taxonomic classification in adjacent text. QTLs from the genus to phylum level (gpQTLs) are colour coded according to their phylum classification (see Supplementary Data 1 and 2).

abundances in disease, we first re-analysed the subset of 183 immunized mice common to this and our previous study on EBA⁷. This reveals no significant QTL for EBA presence/absence at an *E*-value¹⁶ cutoff <0.1 (see Methods), likely due to the reduced number of animals (Fig. 4a). However, because bacterial taxon abundance does display a clear genetic component, we next sought to evaluate potential interactions between bacterial species and disease susceptibility by applying a covariate analysis between each of the 131 CMM species abundances and EBA disease susceptibility. This reveals significant covariation (*E*-value <0.1) involving 10 out of 131 taxa, which, intriguingly, increases the power of detecting EBA QTLs, as a novel locus (covariate QTL) (Chr.19, CI 53–60, peak at 57 Mb) is detected (Fig. 4b, Supplementary Data 4). Two OTUs belonging to the genus *Staphylococcus* clearly display a gene–bacterial interaction (*E*-value <0.05) (Fig. 4c).

To further characterize the nature of the identified covariate QTL, we arbitrarily divided individuals into ‘high’ (top 50%) and ‘low’ (bottom 50%) groups with respect to their individual OTU abundances and analysed the proportion of individuals developing EBA with respect to host genotype. This reveals that for most cases the proportion of animals developing EBA is higher among the low OTU abundance category (*n* = 10; one of which is also significant by Fisher’s exact test between these defined abundance categories (Fig. 4d); we note, however, that all 10 taxa display significant covariation), suggesting a probiotic role

(Supplementary Data 5). Although community-level alterations of the skin microbiota in the context of EBA are present (for example, Supplementary Figs S3 and S4), we note that the putative probiotic covariate taxa identified do not vary in abundance simply according to disease status. Namely, both healthy and diseased individuals are found among the low abundance categories, thus, low abundance of, for example, *Staphylococcus spp.* is not a simple byproduct of disease, but increases the probability of developing symptoms.

The large number of covariate bacterial taxa interacting with a single host locus suggests that individual bacterial taxa may not be acting independently. Thus, to identify potential interactions among covariate taxa we performed a pairwise correlation analysis (Fig. 5). Indeed, this reveals significant positive correlations (Pearson’s correlation; *P*-value ≤0.05 after correction for multiple testing (Benjamini–Hochberg¹⁷)) between many taxa, suggesting interactions between the host locus and bacterial species assemblages or individual driver species.

Discussion

Our results provide strong evidence that host genetic variation contributes to differences in the bacterial communities observed in the skin. In a previous QTL analysis of the mouse faecal community, Benson *et al.*³ reported 13 significant and 5 additional suggestive QTLs for 26 out of 64 taxonomic groups

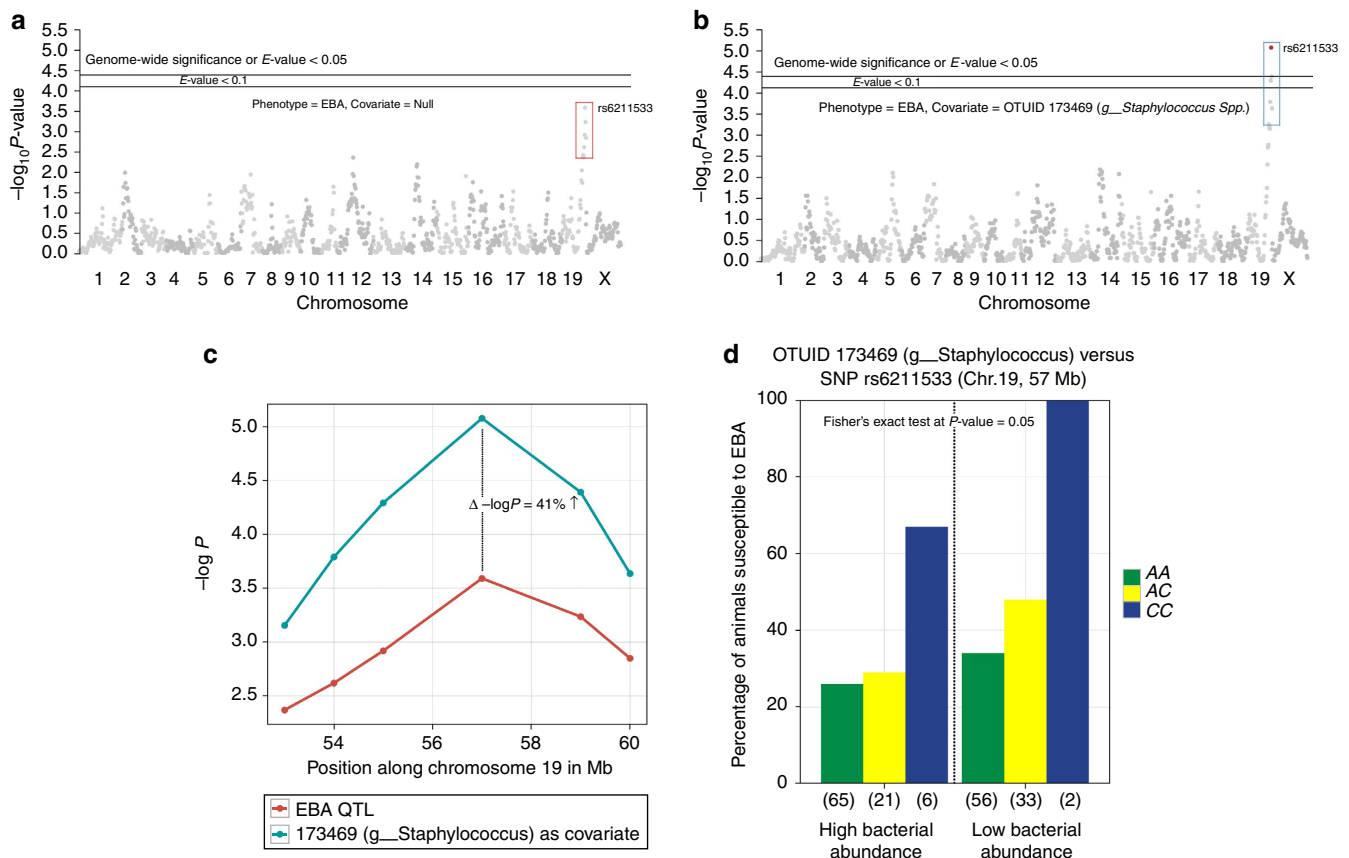


Figure 4 | Gene-microbe interaction in EBA susceptibility. (a) Manhattan plot of $-\log_{10} P$ -values for each SNP (1,199 SNPs in x axis according to their position on each chromosome) tested against EBA disease phenotype (presence/absence). (b) Manhattan plot showing $-\log_{10} P$ -values for each SNP tested against EBA including *Staphylococcus spp.* (OTUID 173469) abundance as a covariate. SNPs falling below an *E*-value of 0.05 are shown in red. (c) Portion of chromosome 19 containing the covariate QTL with a peak at SNP rs6211533. (d) Percentage of animals developing EBA among high (top 50%) and low (bottom 50%) *Staphylococcus spp.* (OTUID 173469) abundance categories with respect to host genotype at rs6211533 (represented by green (AA), yellow (AC) and blue (CC)). Numbers in parentheses indicate the sample size within each genotype category.

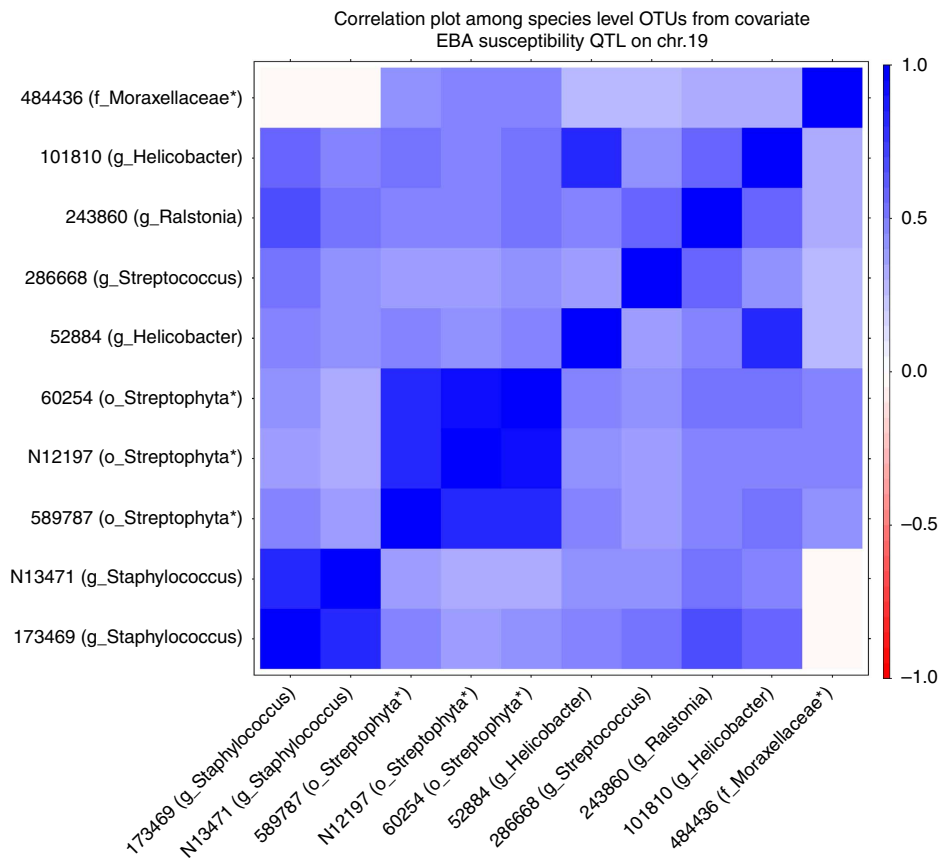


Figure 5 | Correlation matrix of covariate OTUs. Pearson's correlation values between OTUs that significantly co-vary with an EBA susceptibility locus on chromosome 19 are displayed. For OTUs not classified at the genus level, the next highest taxonomic level for which classification was possible is displayed by an asterisk. The taxonomic level of classification is indicated by k, p, c, o, f and g for kingdom, phylum, class, order, family and genus, respectively. Only values significantly differing from zero after correction for multiple testing¹⁷ are shown by either blue (positive correlation) or red (negative correlation) squares.

tested. However, their analysis was not extended beyond the level of bacterial genera. Despite our more inclusive set of phenotypic traits extending to the bacterial species (OTU) level, we detected a smaller number of loci, with nine significant QTLs for 9 out of 131 species-level traits. Differences in sample size, sequencing coverage, mouse strains used and the obvious distinctions between the gut and skin environments may all contribute to these differences in QTL detection. Interestingly, we nonetheless find evidence of overlap between the two studies. The confidence intervals of two out of 18 QTLs controlling bacterial abundance in murine faeces contain the peak SNP of a skin QTL, which overlaps more than expected by chance (P -value < 0.05 ; Fig. 3; Methods). One of these QTLs is consistent at the phyla level (Firmicutes, Chromosome 14), while the other is at the order level (*Pseudomonadales*, Chromosome 9). Similarly, an additional two of our skin spQTLs (OTU N31208 belonging to *Streptococcus* on Chromosome 12 and OTU 130241 belonging to *Herbaspirillum* on Chromosome 15) overlap with faecal QTLs from another more recently published study⁴, although the taxonomic assignments do not agree at even the phylum level.

The confidence intervals of our skin microbiota QTLs contain nine genes known to be involved in the functioning of the innate immune system (Supplementary Data 3). *Interleukin-1 receptor-associated kinase (IRAK)-4* is an interesting candidate found within the confidence interval of spQTL6, which modulates an OTU (ID 130241) belonging to the genus *Herbaspirillum*. Deficiencies of this gene in humans lead to increased

susceptibility to pyogenic bacterial infections including *Staphylococcus aureus*¹⁸, and its interaction with the MYD88 adapter protein is used by several Toll-like receptor pathways in host defence¹⁹, as well as being involved in controlling commensal bacteria²⁰. Another gene coding for CD14 antigen is found within spQTL8 on chromosome 18, which modulates an OTU (ID N10459) belonging to the genus *Staphylococcus*. Increasing CD14 expression enhances Toll-like receptor 2 activation in skin in the presence of vitamin D₃—1,25-dihydroxyvitamin D₃ (1,25D3)²¹, which in turn influences the skin's sensitivity to microbial challenge. Furthermore, several studies have shown that components of *S. aureus* (LTA and peptidoglycan) interact with the CD14 molecule^{22–24}. Finally, by treating bacterial abundances as covariates with the presence/absence of EBA, we identified an additional significant EBA QTL on chromosome 19 (Fig. 4). One potential candidate gene lying within this chromosomal interval (53–60 Mb) is *caspase-7 (casp7)*, a member of the cytosolic cysteine protease family known to be involved with inflammatory disorders^{25,26} and defence against pathogens²⁷.

Similar to previous studies of chronic inflammatory skin diseases, our findings support a role of resident microbial communities in disease pathogenesis. The differences in community composition and structure between mice with and without EBA symptoms are akin to shifts in the skin microbiota associated with atopic dermatitis disease flares and treatment²⁸ or between psoriatic lesions and both unaffected skin in patients and healthy controls²⁹. Although other examples such as

endemic pemphigus foliaceus (fogo selvagem), where exposure to haematophagous insects is implicated³⁰, suggest direct environmental/microbial triggers of disease, the roughly three-fold increase over the last 30 years of atopic dermatitis in industrialized countries suggests more complex environmental influences, possibly mediated by changes in microbial communities. By investigating disease provocation in a large mouse mapping population under controlled environmental conditions, we were able to identify individual, genotype-dependent microbial risk factors among a core set of taxa inhabiting the skin of both healthy and diseased mice, more closely resembling a disease-modifying effect. Although further validation and characterization of these interactions awaits more intensive experimental interrogation in, for example, gnotobiotic animals, our investigation of another EBA-susceptible mouse strain (SJL/J; not included in this study) before and after autoimmunization reveals that other aspects of the skin community, in particular alpha diversity (species richness and evenness), are predictive of disease outcome (unpublished results). Thus, the further identification and functional analysis of host genetic and probiotic bacterial factors represent promising avenues for research in preventative and therapeutic treatment development.

Methods

Generation of a four-way advanced intercross line. Parental mouse strains (MRI/MpJ, NZM2410/J, BXD2/TyJ, Cast) for generating a heterogeneous intercross line³¹ were purchased from the Jackson laboratory (Maine, USA). Briefly, strains were intercrossed at an equal strain and sex distribution. First generation (G1) offspring mice were then mated considering their parental origin to maintain an equal distribution of parental alleles for successive generations by maintaining at least 50 breeding pairs per generation. Male and female offspring used in the study were transferred to separate cages according to sex after weaning. Depending on the number of animals per cage, females from multiple families were also housed together. Animals were held under specific pathogen-free conditions at a 12-h light/dark cycle with food and water *ad libitum*. All 261 animals (135 males and 126 females) in this study were taken from the fourth generation of this advanced intercross line at 6 months of age. All animal experiments were approved by the state of Mecklenburg-Vorpommern, Germany.

Induction of experimental EBA. EBA was induced by immunization with an immunodominant peptide within the murine NCI domain of type VII collagen (GST-mCOL7C)⁷. In brief, 60 µg GST-mCOL7C emulsified in 60 µl adjuvant (TiterMax, Alexix, Lörrach, Germany) was injected subcutaneously into the foot pad and tail base. After immunization mice were screened for skin inflammation every 4th week for a period of 12 weeks, after which the ears were taken for analysis at 6 months of age. Ear skin samples were fixed in 4% buffered formalin and snap frozen at -80°C . A total of 183 immunized- and 78 non-immunized mice were included in skin microbiota QTL mapping.

DNA extraction and 16S rRNA gene pyrosequencing. Bacterial DNA from mouse ears was extracted using the PowerSoil Kit (MoBio, Carlsbad, CA). During the killing of the mice, both ears were taken for either bacterial DNA extraction or other analyses (that is, histology, RNA), and the ear chosen for microbial analysis was chosen at random. Approximately one third of an ear was transferred to the Power Bead tubes containing 60 µl of C1 solution and 20 µl of 20 mg ml⁻¹ Proteinase K. Samples were incubated at 50 °C for 2 h at 850 r.p.m. and the remaining steps were performed according to the manufacturer. Amplification of the hypervariable V1 and V2 (27F–338R) region of the 16S rRNA gene was performed using composite forward (5'-CTATGCGCCTTGCCAGCCCGCTCAGTCAGAGTTCAGXXXXXXXXXXCATGCTGCCGTTAGGAGT-3') and reverse (5'-CGTATCGCCTCCCTCGGCCA GTTTGATCCTGGCTCAG-3') primers. These primers include the 454 Life Sciences adaptor A (reverse) and B (forward), denoted by italics. The underlined sequences represent the broadly conserved bacterial primers 27F and 338R. Ten base pair multiplex identifiers (MIDs; designated as "XXXXXXXXXX") were added to reverse primers to uniquely tag PCR products. Duplicate 25 µl PCR reactions, each containing 100 ng of template DNA, were performed using Phusion Hot Start DNA Polymerase (Finnzymes, Espoo, Finland) with the following cycling conditions: initial denaturation for 30 s at 98 °C; 35 cycles of 9 s at 98 °C, 30 s at 55 °C and 30 s at 72 °C; final extension for 10 min at 72 °C. Duplicate reactions were combined after PCR and products were extracted with the MiniElute Gel Extraction Kit (Qiagen, Hilden, Germany). Quantification was performed with the Quant-iT dsDNA BR Assay Kit on a Qubit fluorometer (Invitrogen, Darmstadt, Germany). Purified PCR products were pooled in equimolar amounts and further purified using Agencourt Ampure Beads (Beckman

Coulter, Krefeld, Germany). Aliquots of each library were run on an Agilent Bioanalyser before emulsion PCR and sequencing according to the manufacturer's instructions on a Roche 454 GS-FLX using Titanium sequencing chemistry.

454 Pyrosequencing data analysis. A Perl script using the Smith–Waterman algorithm³² was written to match the forward primer and barcode allowing no insertions or deletions. Sequences were required to have a length between 290 to 370 nucleotides, an average quality score ≥ 20 and contain no ambiguous bases. ChimeraSlayer (<http://microbiomeutil.sourceforge.net/>) was used to remove chimeras. An average of 5,732 reads per sample was obtained for 261 animals.

Taxonomic classification. RDP classifier³³ (RDP Multi-Classifer version 1.0) was applied to assign taxonomy to the genus level using 0.80 as minimum confidence. The 'fixrank' option was used to assign taxonomy from kingdom to genus level. QIIME³⁴ scripts were used to pick OTUs at a 97% similarity threshold. The most abundant sequence from each OTU bin was chosen as the representative sequence. Species-level taxonomy was obtained by NCBI BLAST against the greengenes reference database (http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/Caporaso_Reference_OTUs/gg_otus_4feb2011.tgz) with an *E*-value cutoff of 0.001. Sequences unclassifiable at a given taxonomic rank were classified at the next possible higher rank.

Data preparation for QTL analysis. The proportion values (number of reads for a given taxon/total sequence reads for a given animal) for each taxonomic level were normalized as suggested by Benson *et al.*³ The taxon bins were then log₁₀ transformed. This resulted in 34,624 species-level OTUs, 863 genera, 376 families, 218 orders, 93 classes and 51 phyla. Out of 34,624 species-level OTUs, 19,443 cover nearly 99% of the total sequences and 762 species OTUs represents nearly 90% of the total sequences (Supplementary Fig. S6).

Core measurable microbiota. A CMM was determined by a manner similar to that of Benson *et al.*³ using two technical repeats from five different samples. Sequences were processed, classified into taxonomic bins and the log₁₀ transformed values for each bin were plotted for all pairwise combinations of the two repeats (Supplementary Fig. S5). A threshold of >20 reads per bin leads to a correlation > 0.97. Thus, the CMM taxa were defined as bins containing more than 20 reads in at least 20 animals. The resulting 131 species-level OTUs represent nearly 80% of the total sequences (Supplementary Fig. S6).

Genotyping and QTL analysis. The Illumina murine HD array was used to genotype 1,449 SNPs (of which 1,199 are informative) from 261 G4 animals. The Happy package³⁵ was used for QTL and covariate QTL analysis, expecting additive contributions from inherited parental haplotypes at each locus. Each log₁₀ transformed taxonomic bin was treated as an individual phenotypic trait. The hft function in Happy was used to correct for cage and family effects in a manner similar to Johnsen *et al.*³⁶ Covariates were included in the model by specifying an additional design matrix in the input file. The genome-wide significance, or *E*-value, for each phenotype was estimated by a permutation test based on 1,000 shuffled reassignments of the phenotypes¹⁶. To illustrate the procedure used to obtain the *E*-value for a given trait, we produced a 'QQ-plot' (Supplementary Fig. S7), which shows the distribution of maximum $-\log_{10} P$ -scores recorded across 1,199 SNPs for 1,000 random permutations of the phenotype scores (for OTUID: N26684). The threshold for significant QTLs was set at 5% and that for suggestive QTLs³ at 10%. This corresponds to analysis of variance (ANOVA) $-\log_{10} P$ -values ≥ 4.39 and ≥ 4.10 , respectively, as described by Valdar *et al.*³⁷ Confidence intervals were determined manually by a drop of 1.5 in ANOVA $-\log_{10} P$ -score¹⁶. The probability of overlap for QTLs was calculated as described by Graham *et al.*³⁸ with a size of 2,500 Mb used as the size of the mouse genome. Chromosome visualization was performed with circos³⁹. All mouse genotype, phenotype and Happy input files are provided in Supplementary Data 6.

Alpha and beta diversity analysis. Alpha and beta diversity analyses were largely performed using QIIME³⁴. Alpha diversity^{34,40} analysis was performed on the whole data set with a normalized sequence number of 2,500 per individual (the lowest read number in the data set). An OTU table was generated using a 97% sequence similarity threshold and singletons were removed. The Chao1 and Shannon indices, both common measures of diversity within a sample (that is, alpha diversity), were used to describe species richness and evenness, respectively. Between-sample diversity (that is, beta diversity) was analysed using CAP implemented in the Vegan package in R⁴¹, with disease status as a constraint. This analysis is very similar to a redundancy analysis, but additionally it allows non-Euclidian dissimilarity indices (we used Bray–Curtis⁴², Jaccard and UniFrac^{14,15} distances). Statistical significance for CAP was determined by an ANOVA-like permutation test function with 1,000 permutations (anova.cca) in Vegan. We used R version 2.15.2 on Linux (R Development Core Team (2012)).

References

- Rosenthal, M., Goldberg, D., Aiello, A., Larson, E. & Foxman, B. Skin microbiota: microbial community structure and its potential association with health and disease. *Infect. Genet. Evol.* **11**, 839–848 (2011).
- Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
- Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl Acad. Sci. USA* **107**, 18933–18938 (2010).
- McKnite, A. M. *et al.* Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS One* **7**, e39191 (2012).
- Bowcock, A. M. & Cookson, W. O. C. M. The genetics of psoriasis, psoriatic arthritis and atopic dermatitis. *Hum. Mol. Genet.* **13 Spec No 1**, R43–R55 (2004).
- Chen, M., Kim, G. H., Prakash, L. & Woodley, D. T. Epidermolysis bullosa acquisita: autoimmunity to anchoring fibril collagen. *Autoimmunity* **45**, 91–101 (2012).
- Ludwig, R. J. *et al.* Identification of quantitative trait Loci in experimental epidermolysis bullosa acquisita. *J. Invest. Dermatol.* **132**, 1409–1415 (2012).
- Hundorfean, G., Neurath, M. F. & Sitaru, C. Autoimmunity against type VII collagen in inflammatory bowel disease. *J. Cell Mol. Med.* **14**, 2393–2403 (2010).
- Schreiber, S., Rosenstiel, P., Albrecht, M., Hampe, J. & Krawczak, M. Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat. Rev. Genet.* **6**, 376–388 (2005).
- Grice, E. A. *et al.* A diversity profile of the human skin microbiota. *Genome Res.* **18**, 1043–1050 (2008).
- Gao, Z., Tseng, C., Pei, Z. & Blaser, M. J. Molecular analysis of human forearm superficial skin bacterial biota. *Proc. Natl Acad. Sci. USA* **104**, 2927–2932 (2007).
- Fierer, N., Hamady, M., Lauber, C. L. & Knight, R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc. Natl Acad. Sci. USA* **105**, 17994–17999 (2008).
- Faith, D. P. & Baker, A. M. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evol. Bioinform.* **2**, 121–128 (2006).
- Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
- Hamady, M., Lozupone, C. & Knight, R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J.* **4**, 17–27 (2010).
- Dupuis, J. & Siegmund, D. Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**, 373–386 (1999).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Statist. Soc. Ser. B* **57**, 289–289 (1995).
- Picard, C. *et al.* Pyogenic bacterial infections in humans with IRAK-4 deficiency. *Science* **299**, 2076–2079 (2003).
- Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
- Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by Toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229–241 (2004).
- Schauber, J. *et al.* Histone acetylation in keratinocytes enables control of the expression of cathelicidin and CD14 by 1,25-dihydroxyvitamin D3. *J. Invest. Dermatol.* **128**, 816–824 (2007).
- Han, S. H., Kim, J. H., Martin, M., Michalek, S. M. & Nahm, M. H. Pneumococcal lipoteichoic acid (LTA) is not as potent as staphylococcal LTA in stimulating Toll-like receptor 2. *Infect. Immun.* **71**, 5541–5548 (2003).
- Kusunoki, T., Hailman, E., Juan, T. S., Lichenstein, H. S. & Wright, S. D. Molecules from *Staphylococcus aureus* that bind CD14 and stimulate innate immune responses. *J. Exp. Med.* **182**, 1673–1682 (1995).
- Schröder, N. W. J. *et al.* Lipoteichoic acid (LTA) of *Streptococcus pneumoniae* and *Staphylococcus aureus* activates immune cells via Toll-like receptor (TLR)-2, lipopolysaccharide-binding protein (LBP), and CD14, whereas TLR-4 and MD-2 are not involved. *J. Biol. Chem.* **278**, 15587–15594 (2003).
- García-Lozano, J. R. *et al.* Caspase 7 influences susceptibility to rheumatoid arthritis. *Rheumatology* **46**, 1243–1247 (2007).
- Teixeira, V. H. *et al.* Genetic and expression analysis of CASP7 gene in a European Caucasian population with rheumatoid arthritis. *J. Rheumatol.* **35**, 1912–1918 (2008).
- Akhter, A. *et al.* Caspase-7 activation by the Nlr4/Ipaf inflammasome restricts *Legionella pneumophila* infection. *PLoS Pathog.* **5**, e1000361 (2009).
- Kong, H. H. *et al.* Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res.* **22**, 850–859 (2012).
- Fry, L., Baker, B. S., Powles, A. V., Fahlen, A. & Engstrand, L. Is chronic plaque psoriasis triggered by microbiota in the skin? *Br. J. Dermatol.* **169**, 47–52 (2013).
- Aoki, V. *et al.* Environmental risk factors in endemic pemphigus foliaceus (fogo selvagem). *J. Investig. Dermatol. Symp. Proc.* **9**, 34–40 (2004).
- Asghari, F. *et al.* Identification of quantitative trait loci for murine autoimmune pancreatitis. *J. Med. Genet.* **48**, 557–562 (2011).
- Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981).
- Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
- Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C. & Flint, J. A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl Acad. Sci. USA* **97**, 12649–12654 (2000).
- Johnsen, A. K. *et al.* Genome-wide and species-wide dissection of the genetics of arthritis severity in heterogeneous stock mice. *Arthritis. Rheum.* **63**, 2630–2640 (2011).
- Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* **38**, 879–887 (2006).
- Graham, A. M. *et al.* Support for the reproductive ground plan hypothesis of social evolution and major QTL for ovary traits of Africanized worker honey bees (*Apis mellifera* L.). *BMC Evol. Biol.* **11**, 95 (2011).
- Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
- Tuomisto, H. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* **2**–22 (2010).
- Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veget. Sci.* **14**, 927–930 (2003).
- Bray, J. & Curtis, J. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **325**–349 (1957).

Acknowledgements

We thank Katja Cloppenborg-Schmidt, Ilona Klamfuß, Silke Carstensen and Heine Buhtz for providing technical assistance; Susen Müller and Andreia Marques for their help in the mouse house, Ralf Ludwig, Andreas Recke, Yask Gupta and Philipp Rausch for valuable discussion and Derk Wachsmuth and Werner Wegner for IT assistance. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Excellence Cluster 306 'Inflammation at Interfaces', Research Training Groups 'Modulation of Autoimmunity' (GRK1727/1) and 'Genes, Environment and Inflammation' (GRK1743/1) and the Max Planck Society.

Author contributions

G.S. and S.K. performed experiments, G.S., S.M. and J.W. performed analysis; G.S., J.F.B. and S.M.I. wrote the paper; J.F.B., S.M.I., and D.Z. edited the paper and assisted in interpretation; J.F.B. and S.M.I. designed the experiments.

Additional information

Accession code: European Nucleotide Archive (ENA). ERP002614. PRJEB1934 (<http://www.ebi.ac.uk/ena/data/view/PRJEB1934>)

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Srinivas, G. *et al.* Genome-wide mapping of gene–microbiota interactions in susceptibility to autoimmune skin blistering. *Nat. Commun.* **4**:2462 doi: 10.1038/ncomms3462 (2013).



This article is licensed under a Creative Commons Attribution 3.0 Unported Licence. To view a copy of this licence visit <http://creativecommons.org/licenses/by/3.0/>.