CLR-Seq: a pipeline to identify bacterial microbiota species with immunomodulatory potential through human C-type lectin receptor interaction

Authors:

Jasper Mol[1], Rob van Dalen[1], Yvonne Pannekoek[1], Marcel R. de Zoete[2], Mark Davids[3], Hilde Herrema[3], Nina M. van Sorge[1,4]

Affiliations:

[1]Department of Medical Microbiology and Infection Prevention, Amsterdam Institute for Immunology and Infectious Diseases, Amsterdam UMC, location University of Amsterdam, Amsterdam, The Netherlands

[2]Department of Medical Microbiology, UMC Utrecht, Utrecht, The Netherlands

[3]Department of Experimental Vascular Medicine, Amsterdam UMC, location University of Amsterdam, Amsterdam, The Netherlands

[4]Netherlands Reference Laboratory for Bacterial Meningitis, Amsterdam UMC, location AMC, Amsterdam, The Netherlands

Corresponding author contact:

Professor Dr. Nina M. van Sorge

n.m.vansorge@amsterdamumc.nl

Meibergdreef 9,

IWO building, room IA3-211

1105 AZ Amsterdam, the Netherlands

Data availability statement

The sequencing data generated in this study have been deposited in the European Nucleotide Archive database under accession code: PRJEB86038. The data are freely available without restriction. Source data of figures and tables are provided in the Source Data file.

Abstract

Bacterial microbiota permanently colonize host barrier sites such as the gastrointestinal tract. Select members of the microbiota contribute to antibody-based barrier homeostasis and host defense by inducing local and systemic antibody responses, respectively. However, the molecular mechanisms underlying the selection of these immunologically-relevant microbiota are still unknown. Barrier-resident antigen-presenting cells (APCs) are critical to initiate local and systemic immune responses upon bacterial invasion. Furthermore, APCs sample the environment across intact epithelial barriers during homeostasis by extending their dendrites, which are densely covered with glycan-binding C-type lectin receptors (CLRs). We hypothesized that APCs are thereby ideally positioned and equipped to induce and sustain local and systemic immunity by probing surface-expressed bacterial glycans. To identify the repertoire of microbiota species that could be recognized by human CLRs and thereby sampled by APCs, we have developed CLR-sequencing (CLR-seq), combining CLR-based bacterial cell sorting and 16S rRNA gene sequencing. We established our experimental pipeline using the fluorescently-labeled CLRs langerin (CD207) and the macrophage galactose C-type lectin (MGL, CD301), and fecal microbiota samples from healthy donors. The proportion of langerin- and MGL-positive microbiota populations showed inter-individual variation but shared enrichment of specific genera. Moreover, CLR-enriched genera correlated with IgA-enriched genera, suggesting a link between glycan-based recognition of bacterial species and the induction of local antibody production. Finally, we validated CLR interaction for several species with high CLR enrichment in monoculture. In summary, CLR-seq allows identification of human microbiota species based on CLR-interacting glycans with easy expansion to include additional CLRs or patient microbiota samples.

Introduction

The human digestive tract is permanently colonized by a wide variety of microorganisms, known as the gut microbiota [1]. Bacteria represent the majority of these microbes, comprising on average 200 different species in most individuals [2, 3]. The composition of the gut microbiota is essential for human health through various processes including food digestion, vitamin production, and resistance to pathogen colonization and infection [4-6]. Induction of local specific secretory IgA (sIgA) antibody responses helps to maintain host-microbe homeostasis at mucosal barrier sites [7]. Additionally, experiments in mice have demonstrated that select microbiota species can induce specific systemic immunoglobulin G (IgG) responses that can protect from infection by bacterial pathogens [8].

Locally-produced sIgA coats a distinct subset of the microbiota population to maintain homeostasis in the intestine by affecting colonization levels [9]. Correspondingly, sIgA binding profiles are significantly altered in host pathologies, such as inflammatory bowel disease, and are thought to reflect the increased inflammatory potential of the microbiota [10, 11]. In humans, mucosal sIgA can undergo somatic hypermutation and affinity maturation, suggesting antigen-driven B-cell selection and maturation [7, 12, 13]. It is currently unknown which molecular components and processes drive the selection of microbiota species that contribute to the induction of specific sIgA.

Antibodies are the end-product of immunological pathways that are initiated by the activation of antigen-presenting cells (APCs). Indeed, mucosal APCs constantly probe the local microbiota and bacterial antigens by extension of their dendrites into the lumen, contributing to both local and systemic immunity [14-16]. This sampling can lead to a specific immune response and bacterial clearance [17, 18]. Interestingly, the tips of the probing APC dendrites are densely covered by C-type lectin receptors (CLRs), a specific subset of innate pattern-recognition receptors that recognize glycan structures [19, 20]. Indeed, in the context of infection, CLRs have an important role of sensing pathogens, viruses, bacteria

and fungi, through interaction with specific microbial glycan structures that are abundantly expressed at the surface [21-23]. This CLR-microbe interaction subsequently initiates and directs adaptive immune responses. [24]. Hence, CLRs on APC dendrites are uniquely equipped and localized to selectively sample bacteria from a complex microbiota community based on their glycan specificity. However, the glycan composition of individual microbiota species represents a large knowledge gap. As glycans are not directly encoded by the bacterial genome, but are synthesized by complex biosynthetic pathways involving multiple building blocks and enzymes, routinely-used omics approaches are unable to capture the microbiota glyco-landscape. Similarly, glycan-microarrays are a useful technique for studying glycan-protein interactions, yet they are unable to capture the diversity of the microbiota glycome at a species-specific level [25, 26]. Further insight into the glycan-dependent interaction between our local immune system and specific microbiota species could offer new opportunities to prime local as well as systemic antibody responses that promote gut homeostasis and immune defense.

Recently, the potential of lectins for studying glycan-mediated immune interactions was shown using soluble lectins [27]. Here we aimed to study the interactions between the bacterial gut microbiota and APC-expressed CLRs, since we suspect that these CLR-interacting microbes prime and boost local and systemic antibody induction. To this end, we have developed a pipeline, called CLR-sequencing (CLR-seq), which combines fluorescence-based sorting of CLR-reactive bacteria with 16S rRNA gene sequencing. As a proof-of-concept, we identified CLR-reactive bacteria in fecal samples from three healthy human donors using two human CLRs with distinct glycan specificities, i.e. langerin (CD207), which binds high mannose structures, fucose, β-glucan and *N*-acetylglucosamine (GlcNAc), and macrophage galactose-like C-type lectin (MGL; CD301), which predominantly interacts with terminal *N*-acetylgalactosamine (GalNAc) residues. In parallel, the same microbiota samples were analyzed for sIgA binding through IgA-sequencing (IgA-seq) to correlate the CLR-reactive bacterial species with those that were specifically targeted by locally-produced

antibodies [10]. CLR binding was validated for a selection of the identified bacterial species using reference strains. Our data provide a first insight into the breadth of microbiota species that could be sampled by local APCs for the induction of antibody responses.

Methods

Fecal donor samples

The fecal donor samples used were collected as part of the PIMMS study (NL67136.018.18) performed at Amsterdam UMC (Amsterdam, the Netherlands) as previously described [28]. Fecal matter was collected from healthy participants and stored at -80°C until further processing. All included donors provided informed consent and the study was approved by the Ethical Review Board of the Amsterdam UMC.

Bacteria isolation and staining

Fecal microbiota were isolated as previously described [10] with some modifications. Approximately 100 mg frozen fecal material was resuspended in 1 mL of ice-cold phosphate-buffered saline (PBS, pH 7.4) for 15 minutes. Next, the suspension was transferred to a Fast Prep Lysing Matrix D tube containing 1.4 mm zirconium-silicate beads (MP Biomedicals) and homogenized by beat-beating for 10 seconds at 6,000 rpm (MagNa Lyser; Roche). To separate the fecal bacteria from the debris, the homogenized suspension was centrifuged (50 x g, 15 minutes, 4°C) and 100 µL supernatant was transferred to a clean 1.5 mL tube and washed with 1 mL PBS + 1% bovine serum albumin (BSA; SigmaAldrich) (8,000 x g, 5 min, 4°C). The fecal pellet was resuspended in 1 mL PBS + 1% BSA and filtered using a 70 µm mini cell strainer (PluriSelect), after which 20 µL was set aside as pre-sort sample.

For the sorting of IgA-coated bacteria, the filtered microbiota suspension was centrifuged and resuspended in 100 µL PBS + 1% BSA containing 20% Normal Mouse Serum (Jackson Immunoresearch) and incubated on ice for 20 minutes. Subsequently, the bacteria were stained with 100 µL phycoerythrin (PE)-conjugated Anti-Human IgA (1:10 diluted in PBS + 1% BSA; Miltenyi Biotec, clone IS11-8E10) on ice for 30 minutes, covered from light. Samples were washed with and resuspended in PBS + 1% BSA prior to flow cytometric analysis or cell

sorting. To determine CLR-specific binding of fecal bacteria, the filtered microbiota suspension was centrifuged and stained with 25 µg/mL recombinant human langerin-FITC [29] or his-tagged human MGL (hMGL; R&D Systems) in Tris-sodium-magnesium buffer (20 mM Tris; Roche, 150 mM NaCl; Merck, 2 mM CaCl2; Merck, 2 mM MgCl2; Merck, pH 7.0; TSM) + 1% BSA, and incubated 30 minutes at 37°C with agitation (600 rpm). Since the recombinant MGL-construct was not fluorescently-labelled, MGL-stained bacteria were stained with anti-his-FITC after washing (1:20 in TSM + 1% BSA; Invitrogen, clone AD1.1.10) for 20 minutes at 4°C, covered from light. Samples were washed with and resuspended in TSM + 1% BSA prior to flow cytometric analysis (FACSymphony A1, BD Biosciences) measuring 20,000 events, or fluorescence-activated cell sorting (Sony SH800). For cell sorting, 200,000 positive events per staining were collected in PBS + 1% BSA, pelleted by centrifugation (8,000 x g, 5 min, 4°C), resuspended in 50 µL PBS and stored at -20°C until DNA extraction. For CLR specificity analysis, *Staphylococcus aureus* strains N315 and PS187 (Table S1) were used for langerin and MGL respectively. These bacteria were fluorescently-labeled with cell Trace Violet (Invitrogen) as per manufacturer's instructions and were spiked into the isolated fecal microbiota prior to staining.

DNA extraction, 16S rRNA sequencing and statistical analysis

Fecal DNA extraction, 16S rRNA gene sequencing analysis, and bioinformatics analysis were performed as described previously [30]. Briefly, total genomic DNA was isolated from the pre-sorted and sorted microbiota samples at the Microbiota Centre of Amsterdam (MiCA). Isolation was performed using a repeated bead beating protocol and the Maxwell RSC Whole Blood DNA kit [31]. DNA concentration was measured in a 96-wells plate using the Qubit dsDNA BR Assay. Amplification of the V3-V4 regions of the 16S rRNA gene sequences was performed with the 341F-805R primers in a single step PCR protocol. Samples were purified

using Ampure XP beads, measured and sequenced equimolarly using Illumina MiSeq (v3 chemistry with 2 × 250 cycles).

Amplicon sequences were parsed using a vsearch (v2.15.2) based pipeline [32]. Paired end reads were merged, with max differences set to 100 and allowing for staggered overlap. ASVs were inferred from reads with lower than 1.5 expected error rate using the cluster_unoise with centroids algorithm with a minsize of 4, after which chimeras were removed using the uchime3 denovo method. For each sample ASV abundances were determined by mapping the merged reads against ASV sequences using usearch_global with a 0.97 distance cut off. Taxonomy was assigned using R (4.2.0) and the dada2 [33] assign taxonomy function using the silva (v132) [34] reference database. A phylogenetic tree was generated use mafft (v7.310) [35] and Fasttree (2.1.11) [36].

The abundance and ASV tables were used for downstream analysis using R's phyloseq package [37]. Alpha (Shannon's index) and beta (Bray-Curtis dissimilarity) diversity analysis was performed using the vegan package in R [38]. Probability scores were calculated using relative abundances for each genera, similarly as described previously [39] and only genera that were present in the sorted sample of all three individual donors were included. For visualization of 16S rRNA gene sequencing analysis, R-studio (version 4.4.1) or Graphpad Prism (version 10.2.0) was used. Bacterial genera were assigned either monoderm or diderm status as described previously [40].

Staining and flow cytometric analysis of bacterial strains

The bacterial strains used in this study are listed in Table S1. *S. aureus* and *Aquamicrobium* strains were grown in tryptic soy broth (TSB; Oxoid), the streptococcal strains in Todd-Hewitt broth (Oxoid) supplemented with 0.5% yeast extract (BD Biosciences; THY) and the *Corynebacterium* species were grown on Columbia blood agar (Oxoid), with *C. appendicis*

grown in a microaerophilic environment. All species were grown overnight at 37°C, except for the *Aquamicrobium* species, which were grown at 30°C. Cells were collected from broth culture or plate with an inoculum loop in TSM + 0.1% BSA by centrifugation (3,600x*g*, 10 min, 4°C) and resuspended in TSM + 0.1% BSA at an OD_{600} of 0.4. To determine binding to the CLR constructs, 10^6 colony forming units (CFU) of bacteria were added to either FITC-labeled recombinant langerin or his-tagged human MGL to a final concentration of 25 µg/mL in TSM + 0.1% BSA. For blocking langerin binding, mannan (20 µg/mL; Sigma-Aldrich M7504) was added and for blocking of the MGL binding, GalNAc was added (50 mM; Sigma-Aldrich A2795) simultaneously with the bacteria. The bacteria were incubated 30 min (37°C, 900 rpm, covered from light) and washed with TSM + 0.1% BSA. For the detection of hMGL binding, an anti-hisTag FITC-conjugated antibody was used (1:20 in TSM + 0.1% BSA, Invitrogen; clone AD1.1.10). Bacteria were fixed with PBS + 1% formaldehyde (Sigma-Aldrich) and measured by flow cytometry (BD FACS Canto). Visualization and statistical analysis were performed using Graphpad Prism (version 10.2.0). Differences in CLR binding at different conditions were calculated with a one-way analysis of variance within each bacterial species. A *P*-value of <0.05 was considered significant.

Results

Profiling of the human microbiota with langerin and MGL using CLR-seq

The aim of this study was to identify the bacterial microbiota species that can be recognized by CLRs and compare the recognized bacteria to those that were bound by sIgA. For this proof-of-concept study, we used the Langerhans cell CLR langerin (CD207) and the macrophage CLR MGL (CD301), which have distinct glycan binding specificities [29, 41]. First, we validated that our approach could selectively identify a CLR-reactive species in a complex bacterial mixture such as microbiota using fluorescently-labeled *S. aureus* strains that we previously showed to bind to either langerin or MGL [42, 43]. These fluorescently-labeled *S. aureus* strains were spiked into the fecal microbiota sample and subsequently the entire sample was stained with langerin or MGL (Figure 1). The spiked-in *S. aureus* strains were identified as a double-positive population. In addition, a subset of the fecal microbiota population showed reactivity to the added CLR.

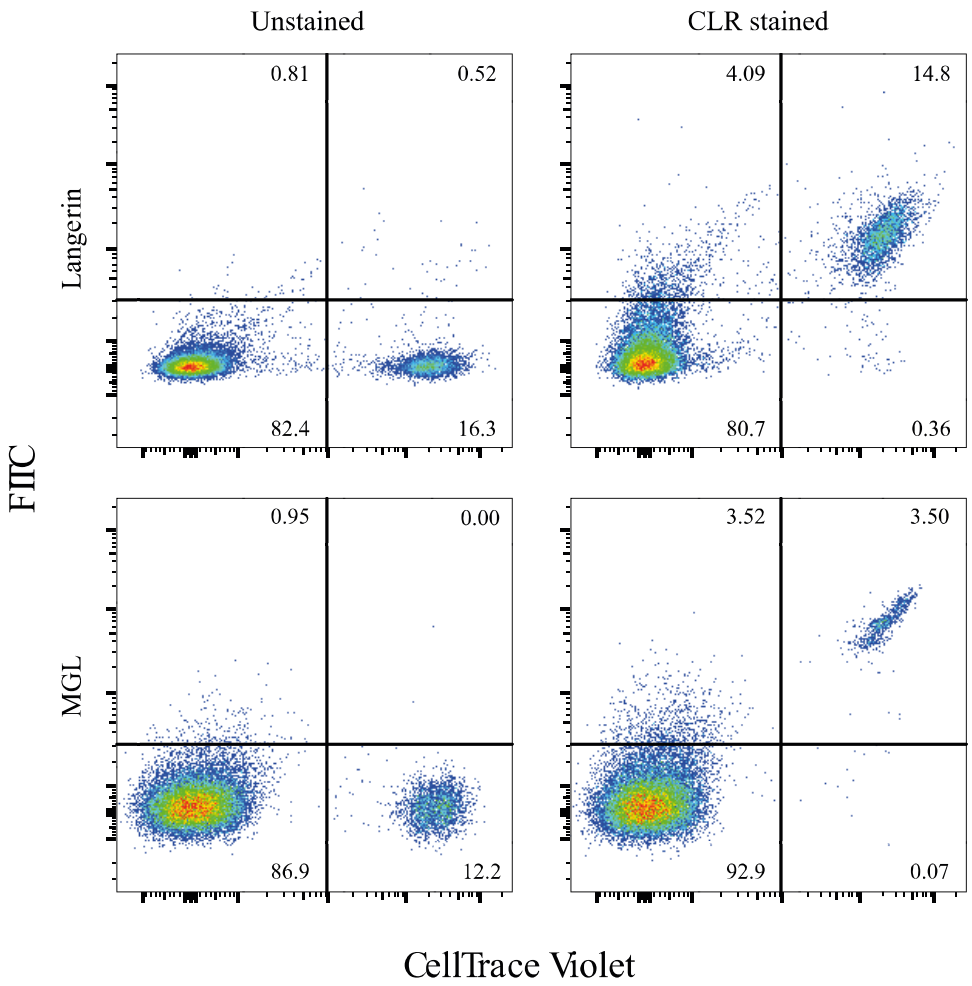

Figure 1. Human recombinant CLRs bind to specific populations within the human microbiota. Isolated fecal microbiota were supplemented with 10^6 CFU langerin-reactive *S. aureus* N315 (upper panels) or MGL-reactive PS187 (lower panels). *S. aureus* strains were fluorescently labeled with CellTrace Violet prior to addition to the microbiota suspension. For both langerin and MGL, the bacteria were measured without (left panels) and after FITC-labeled CLR staining (right panel). The reduced relative abundance of CellTrace Violet+/MGL+ *S. aureus* results from agglutination by the secondary antibody. The percentages are shown per quadrant. The images shown are representative of two different fecal donors.

To identify the bacteria that bound to langerin and MGL, we applied this staining protocol on replicates of isolated fecal microbiota samples from three individual healthy human donors. The langerin- and MGL-positive bacterial fractions were sorted using fluorescence-activated cell sorting (Figure S1). Additionally, for each donor a replicate of the fecal microbiota sample was stained with anti-IgA, after which the anti-IgA-binding bacteria were

sorted and collected. DNA was extracted from the unsorted as well as CLR+/IgA+ sorted samples and analyzed by 16S rRNA gene sequencing. We observed inter-individual variation in the proportion of langerin-, MGL- and sIgA-reactive bacteria of the healthy donors (Figure 2A). The positively-sorted fractions showed largely overlapping taxonomic profiles at the family level compared to the unsorted microbiota sample in all donors (Figure 2B). However, the sorted fractions had lower α-diversity scores, with the langerin-sorted fractions showing the lowest level of α-diversity (Figure S2). These lower α-diversity scores post-sorting indicates a decreased diversity of bacterial species, suggesting a specific selection by the used CLRs. The microbiota samples from the three donors showed different ß-diversity based on the 16S rRNA gene sequencing data, with donor 1 appearing most distinct from the other donors (Figure S3). For each donor, ß-diversity of the sorted IgA+/CLR+ samples was distinct from that in their respective unsorted populations. A principal component analysis (PCA) with multilevel decomposition for repeated measurements showed that the sequencing data varied most strongly between the total unsorted microbiota samples and the sorted IgA and CLRs fractions (Figure 2C).

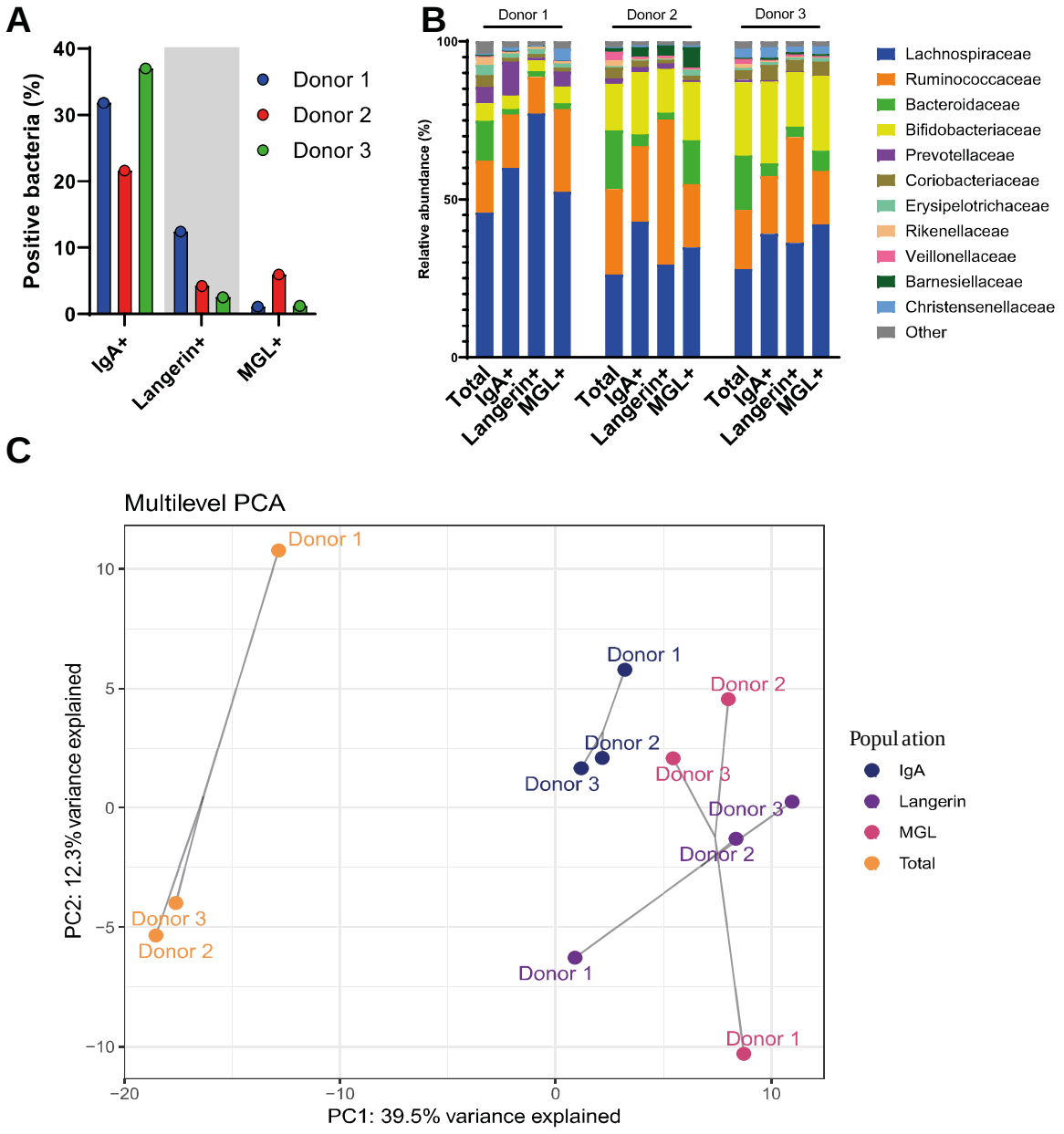

Figure 2. CLR-seq results for fecal microbiota samples from three healthy donors. A. Percentages of CLR- and sIgA-positive microbiota fractions for the three analyzed donors. **B.** Stacked bar plot of the taxonomic abundance at the family level for each individual donor in the total (i.e. unsorted) and the CLR/sIgA-sorted fractions based on 16s rRNA sequencing. **C.** Principle component analysis of all sequenced samples after multilevel decomposition for repeated measurements in individual donors.

Correlation between CLR- and sIgA targeted bacteria at genus level

We compared the bacterial compositions of the CLR- and sIgA-sorted fractions to study the potential correlation between CLR recognition and sIgA binding. 16s rRNA sequencing had insufficient resolution to resolve the bacterial composition at species level for most amplicon sequencing variants (ASVs). Therefore, we determined the proportional abundance of all samples at the genus level.

From the proportional abundances, we calculated the corresponding probability scores as a measure of IgA and CLR binding [39](all individual probability scores are listed in the supplementary data file). For the genera present in all three individual sorted donor samples, we observed a significant correlation for the langerin- and MGL enriched genera with the IgA-coated genera across the three individuals (Figure 3A, B). This correlation was also observed for each individual donor (Figure S4). Interestingly, both monoderm and diderm genera were observed among the highest and lowest average CLR+ probability scores, indicating that the recombinant CLR constructs could recognize glycan motifs regardless of the bacterial cell envelope architecture.

Furthermore, examining the ten genera with either the highest or lowest average CLR probability scores showed interindividual variation between the three donors (Figure 3CD). For each CLR, we determined the probability score at which a net enrichment of zero would be observed. Regardless of the variation between individual donors, for most genera with the highest average CLR probability scores all three donors showed a higher score than would be observed if the genus would have shown the same relative abundance in the unsorted- and sorted populations (i.e. no enrichment by the CLR-based sorting). Similarly, for most genera with the lowest average CLR probability score, all three donors showed a lower score than would be observed with equal relative abundance in the unsorted and sorted populations. Therefore, despite interindividual variation, the effect of CLR-sorting on the selected population of microbiota members is overall consistent between the three donors.

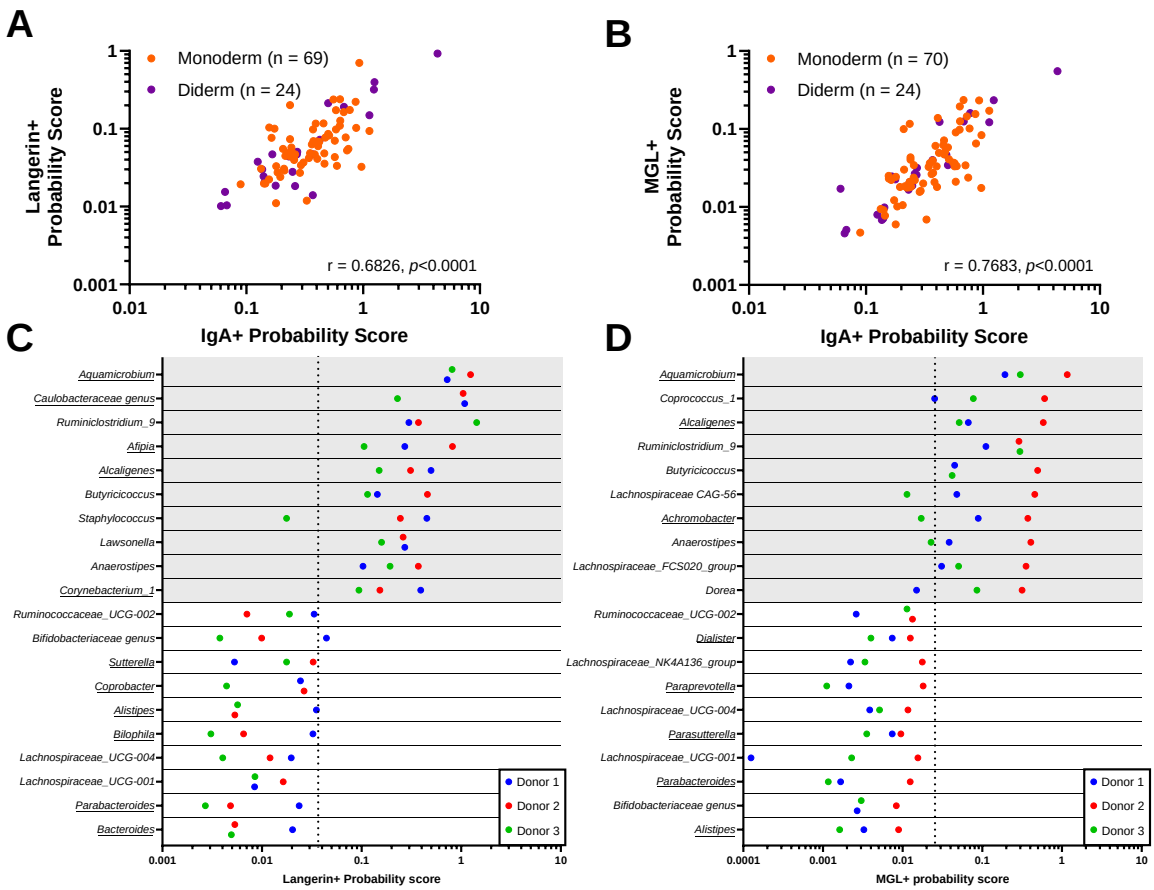

Figure 3. CLR binding of fecal microbiota correlates with microbiota sIgA binding at the genus taxonomic level. Correlation of the average IgA enrichment scores for each genus with their respective **A.** langerin (n=93) or **B.** MGL (n=94) average enrichment scores. All genera shown were present in the IgA+ and CLR+ samples of all individual donors. Spearman's rank order correlation was used for statistical analysis. Monoderm genera are shown in orange, diderm genera in purple. The ten genera with the highest and lowest probability scores are shown for **C.** langerin or **D.** MGL binding. Individual probability are shown for each donor. A dotted line is shown to mark the probability score at which a genus would have equal relative abundance in the unsorted and sorted samples for the respective sorting condition. Diderm genera are underscored.

Validating CLR-seq data for individual species

Based on the 16S rRNA gene sequencing results, we identied several species among the CLR-bound genera with high probability scores (Figure 3CD). The genus with the highest score for both CLRs, as well as for IgA, was *Aquamicrobium*, which was first described in activated

sewage sludge [44]. In our 16S rRNA gene sequencing data, this genus consisted of one ASV attributed to two species, *A. lusatiense* and *A. defluvii*. Since the 16S rRNA gene sequencing could not distinguish these two species, we tested one isolate of each species for binding to langerin and MGL. Furthermore, we could identify two ASVs at the species level within the genus *Corynebacterium_1*, *C. appendicis* and *C. accolens*, that was also among the genera with the highest average langerin probability score of the three donors. Notably, *Aquamicrobium* or *Corynebacterium* bacteria were not detected in the total microbiota populations, which could indicate that our pipeline enables identification of low abundance taxa in the microbiota population. Finally, we identified an ASV in the *Streptococcus* genus attributed to *S. salivarius* and *S. thermophilus* with an increased MGL probability score, which were also tested for MGL binding.

To verify CLR binding to the mentioned species, an individual strain of each species was grown as a monoculture and incubated with recombinant human langerin or MGL, depending on the sorted fractions in which they were identified (Figure S5). For langerin, *S. aureus* N315 was used as a positive control and comparator species, since we identified and confirmed that langerin interacts with the GlcNAc modifications on the cell wall component wall teichoic acid [42, 45]. Similar to *S. aureus*, we observed significant langerin binding to *A. defluvii*, *C. appendicis* and *C. accolens* but not *A. lusatiense* (Figure 4A). For all langerin-binding strains, binding was strongly reduced by addition of mannan, indicating that recognition occurred through the carbohydrate recognition domain (CRD) of langerin.

Binding of recombinant MGL to the identified microbiota species was compared to the positive control *S. aureus* PS187, which interacts with MGL through *N*-acetylgalactosamine (GalNAc) residues on wall teichoic acid [43]. Similarly to the langerin binding assay, MGL bound significantly to *A. defluvii*, but not to *A. lusatiense*, and this binding was abrogated by the addition of GalNAc, suggesting that the recognition by MGL occurred through the CRD

(Figure 4B). No MGL binding was observed for either of the identified *Streptococcus* species that were attributed to the enriched ASV.

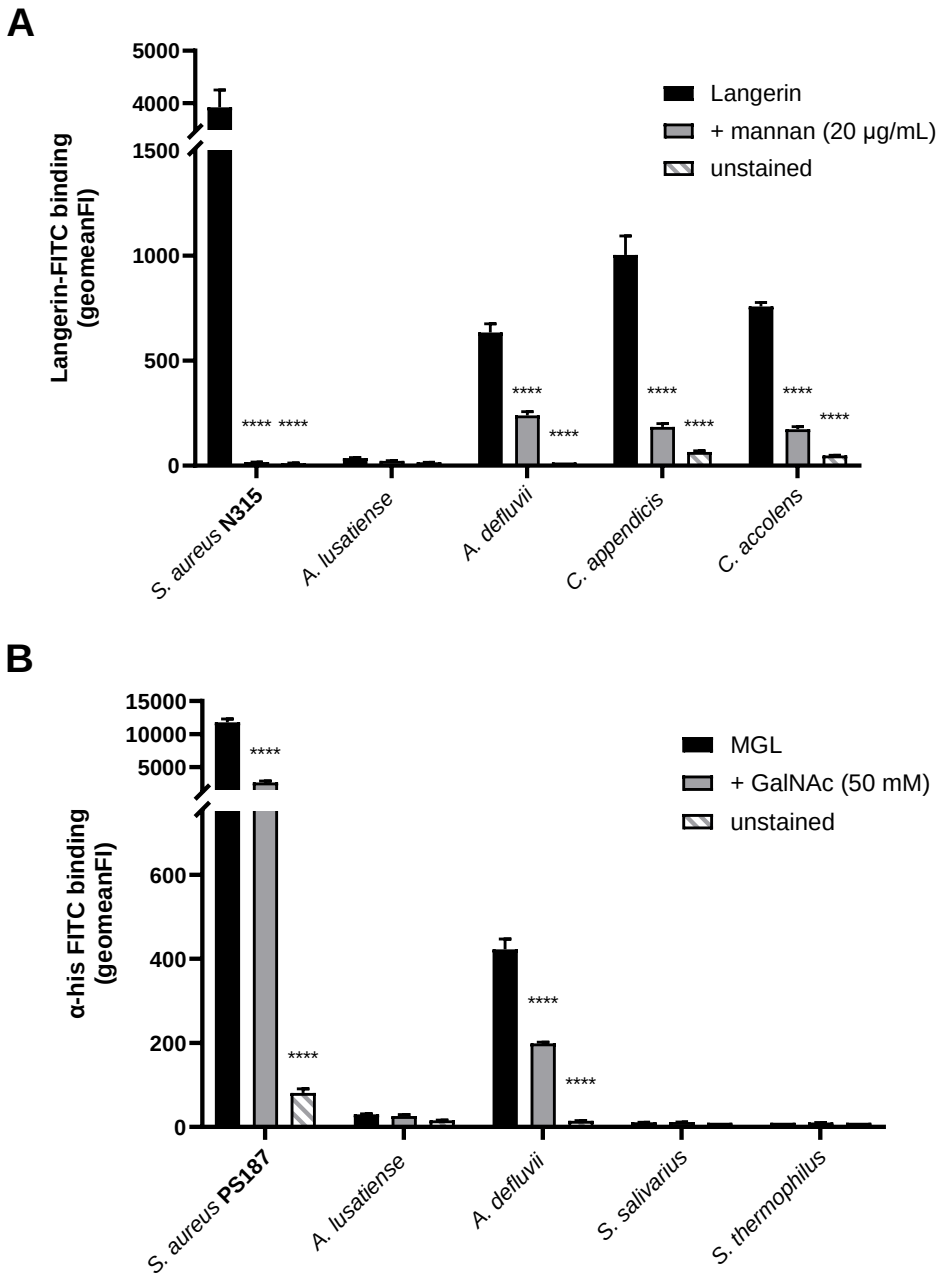

Figure 4. Validation of CLR binding to bacterial species identified by CLR-seq. Binding of **A.** recombinant human langerin-FITC and **B.** recombinant human MGL detected by an anti-hisTag-FITC antibody to *S. aureus* positive controls and identified microbiota species. CLRs were used at the same concentration (25 µg/mL) as used for fluorescence cell sorting. Data are shown as the mean of the geometric mean fluorescence intensity (FI) ± standard error of mean from three independent experiments. ****$P < .0001$

Discussion

A symbiotic relationship between the host and the gut microbiota is essential for maintaining human health, including defense against pathogenic species. The local immune system is continuously interacting with the microbial population to establish this symbiotic relationship. However, the influence of the microbiota extends beyond the gut through the induction of systemic IgG that can cross-protect against infections. CLRs are strategically localized on the tips of sampling dendrites from resident APCs to sample the microbiota [23, 46]. To identify the repertoire of bacterial species that could be sampled by resident APCs, we present proof-of-principle that CLR-seq can be used as a platform to comprehensively determine the microbiota-binding capabilities of human CLRs. The approach can easily be adapted to other host CLRs or microbiota samples of interest, for example from different mucosal sites or from patients with a specific clinical condition. As an outlook, we anticipate that a greater understanding of microbiota species selection by glycan-dependent innate immune sampling, will also advance understanding on mucosal IgA function and polyreactivity, since glycans are a common target of these antibodies.

Bacterial enrichment in sIgA and CLR-binding fractions correlated at the taxonomic genus level. This was true for the combined data of three donors but also for the individual donors. We identified several bacteria at the species level in the CLR-sorted fractions and confirmed their binding in monoculture for select strains. The selection and enrichment of these species in the CLR-sorted fraction of all three donors suggests that the target for CLR binding may be conserved. Importantly, this correlation was independent of proportional abundance in the total unsorted microbiota population. Indeed, some of the CLR- or IgA-enriched species were below the limit of detection in the total microbiota population. This could indicate that CLR-seq can also recognize and identify these low abundance species, however the absence in the total populations and nature of the identified bacteria could also suggest contamination in the sorting pipeline. Although the origin of the *Aquamicrobium* and

Corynebacterium species is speculative and colonization of the gut environment unlikely, the validation of CLR binding to these bacteria as a monoculture does suggest that their identification in the sorted samples was the result of specific recognition by CLR-seq. Previously, it was shown that interaction between local APCs and B cells was necessary for a specific murine sIgA response against gut microbials, although the microbial molecular signatures that were detected by the APCs remained unidentified [47]. We suggest that glycan-mediated bacterial recognition by APC-expressed CLRs at least partially contributes to the selection of microbes that induce specific sIgA at mucosal sites in healthy individuals.

To establish proof-of concept for CLR-seq, we used langerin, a Langerhans cell-specific CLR, and MGL, which is expressed by tissue resident macrophage. Currently, we have limited insight into the exact APC subsets and associated CLRs that line the different intestinal barriers in humans. In mice, the presence of the murine MGL homologue was found in colon tissue [48]. Additionally, MGL expression was found on APCs located in the lamina propria of the jejunum [49]. Although langerin is most likely not present in the gut, the CLR DC-SIGN, predominantly expressed by dendritic cells, was identified in human Peyer's patches [50]. Since langerin and DC-SIGN share glycan specificities [23], it is possible that this explains the observed correlation between IgA-stained and langerin-sorted genera. The incorporation of additional CLRs, such as DC-SIGN, would give a more complete overview of the glycan binding capabilities of the mucosal immune system of the gut.

To identify bacteria in the unsorted, IgA- and CLR-sorted fractions, we used 16S rRNA gene sequencing. This posed limitations for the downstream workflow. While we obtained general taxonomic information with a limited bacterial input (200,000 positive events sorted), identification at the species level was not possible for most ASVs, which was also observed to a lesser extent at the genus level. Since bacterial surface composition of glycans can be strain specific or even phase variable, [24] this could result in the unsuccessful validation of certain species. Furthermore, as CLR-seq studies the host-microbe interaction at the native

level of a complex bacterial community, this could further contribute to disparities between the identified species in the sorted sequencing data and the monocultures in the laboratory setting. An approach with increased resolution, such as next generation whole metagenome sequencing, could provide a more detailed insight into the bacterial identity and characteristics that confer interaction with innate CLRs. Strain-level information may also resolve discrepancies for species that are inconsistently identified in IgA-seq screens or are being associated with certain host pathologies.

Our results suggests that CLR-seq can identify bacterial microbiota species that can be sampled by local APCs through glycan-dependent interactions with CLRs, thereby providing insight into the potential molecular mechanisms for the observed cross-bacterial sampling at host barrier sites. The observed overlap between IgA- and CLR-bound bacteria highlights the importance of glycan recognition for bacterial selection by the local immune system and downstream antibody production. This technique can easily be expanded by an increased panel of recombinant immune receptors, the incorporation of samples from different microbial compartments, as well as the comparison of the microbial glycan content of healthy individuals with patients with underlying disease, making it a new tool for functional microbiota research.

Acknowledgements:

The authors would like to thank Romy Ros for technical assistance; Dr. Torsten Scheithauer for sharing his expertise with the microbiota sorting process; Jorn Hartman for his help with the preparation of the sequencing library and Professor Christoph Rademacher at the University of Vienna for supplying the recombinant langerin construct.

References

1. Afzaal M, Saeed F, Shah YA *et al*. Human gut microbiota in health and disease: Unveiling the relationship. *Front Microbiol* 2022;**13**:999001. https://doi.org/10.3389/fmicb.2022.999001
2. Yang J, Pu J, Lu S *et al*. Species-level analysis of human gut microbiota with metataxonomics. *Front Microbiol* 2020;**11**:2029. https://doi.org/10.3389/fmicb.2020.02029
3. Hou K, Wu ZX, Chen XY *et al*. Microbiota in health and diseases. *Signal Transduct Target Ther* 2022;**7**:135. https://doi.org/10.1038/s41392-022-00974-4
4. Valdes AM, Walter J, Segal E *et al*. Role of the gut microbiota in nutrition and health. *BMJ* 2018;**361**:k2179. https://doi.org/10.1136/bmj.k2179
5. Morowitz MJ, Carlisle EM, Alverdy JC. Contributions of intestinal bacteria to nutrition and metabolism in the critically ill. *Surg Clin North Am* 2011;**91**:771-85, viii. https://doi.org/10.1016/j.suc.2011.05.001
6. Khan I, Bai Y, Zha L *et al*. Mechanism of the gut microbiota colonization resistance and enteric pathogen infection. *Front Cell Infect Microbiol* 2021;**11**:716299. https://doi.org/10.3389/fcimb.2021.716299
7. Li H, Limenitakis JP, Greiff V *et al*. Mucosal or systemic microbiota exposures shape the B cell repertoire. *Nature* 2020;**584**:274-78. https://doi.org/10.1038/s41586-020-2564-6
8. Zeng MY, Cisalpino D, Varadarajan S *et al*. Gut microbiota-induced immunoglobulin G controls systemic infection by symbiotic bacteria and pathogens. *Immunity* 2016;**44**:647-58. https://doi.org/10.1016/j.immuni.2016.02.006
9. Pabst O, Slack E. IgA and the intestinal microbiota: The importance of being specific. *Mucosal Immunol* 2020;**13**:12-21. https://doi.org/10.1038/s41385-019-0227-4
10. Palm NW, de Zoete MR, Cullen TW *et al*. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* 2014;**158**:1000-10. https://doi.org/10.1016/j.cell.2014.08.006
11. Shapiro JM, de Zoete MR, Palm NW *et al*. Immunoglobulin A targets a unique subset of the microbiota in inflammatory bowel disease. *Cell Host Microbe* 2021;**29**:83-93 e3. https://doi.org/10.1016/j.chom.2020.12.003
12. Kabbert J, Benckert J, Rollenske T *et al*. High microbiota reactivity of adult human intestinal IgA requires somatic mutations. *J Exp Med* 2020;**217** https://doi.org/10.1084/jem.20200275
13. Benckert J, Schmolka N, Kreschel C *et al*. The majority of intestinal IgA+ and IgG+ plasmablasts in the human gut are antigen-specific. *J Clin Invest* 2011;**121**:1946-55. https://doi.org/10.1172/JCI44447
14. Rescigno M, Urbano M, Valzasina B *et al*. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nat Immunol* 2001;**2**:361-7. https://doi.org/10.1038/86373
15. Allen F, Tong AA, Huang AY. Unique transcompartmental bridge: Antigen-presenting cells sampling across endothelial and mucosal barriers. *Front Immunol* 2016;**7**:231. https://doi.org/10.3389/fimmu.2016.00231
16. Doron I, Leonardi I, Li XV *et al*. Human gut mycobiota tune immunity via CARD9-dependent induction of anti-fungal IgG antibodies. *Cell* 2021;**184**:1017-31 e14. https://doi.org/10.1016/j.cell.2021.01.016
17. Ouchi T, Kubo A, Yokouchi M *et al*. Langerhans cell antigen capture through tight junctions confers preemptive immunity in experimental staphylococcal scalded skin syndrome. *J Exp Med* 2011;**208**:2607-13. https://doi.org/10.1084/jem.20111718
18. Niess JH, Brand S, Gu X *et al*. CX3CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. *Science* 2005;**307**:254-8. https://doi.org/10.1126/science.1102901
19. Baranov M, Ter Beest M, Reinieren-Beeren I *et al*. Podosomes of dendritic cells facilitate antigen sampling. *J Cell Sci* 2014;**127**:1052-64. https://doi.org/10.1242/jcs.141226
20. Kubo A, Nagao K, Yokouchi M *et al*. External antigen uptake by langerhans cells with reorganization of epidermal tight junction barriers. *J Exp Med* 2009;**206**:2937-46. https://doi.org/10.1084/jem.20091527
21. Geijtenbeek TB, Gringhuis SI. Signalling through C-type lectin receptors: Shaping immune responses. *Nat Rev Immunol* 2009;**9**:465-79. https://doi.org/10.1038/nri2569
22. Reis ESC, Yamasaki S, Brown GD. Myeloid c-type lectin receptors in innate immune recognition. *Immunity* 2024;**57**:700-17. https://doi.org/10.1016/j.immuni.2024.03.005

23. Mnich ME, van Dalen R, van Sorge NM. C-type lectin receptors in host defense against bacterial pathogens. *Front Cell Infect Microbiol* 2020;**10**:309. https://doi.org/10.3389/fcimb.2020.00309
24. Szymanski CM. Bacterial glycosylation, it's complicated. *Front Mol Biosci* 2022;**9**:1015771. https://doi.org/10.3389/fmolb.2022.1015771
25. Geissner A, Reinhardt A, Rademacher C *et al*. Microbe-focused glycan array screening platform. *Proc Natl Acad Sci U S A* 2019;**116**:1958-67. https://doi.org/10.1073/pnas.1800853116
26. Stowell SR, Arthur CM, McBride R *et al*. Microbial glycan microarrays define key features of host-microbial interactions. *Nat Chem Biol* 2014;**10**:470-6. https://doi.org/10.1038/nchembio.1525
27. McPherson RL, Isabella CR, Walker RL *et al*. Lectin-seq: A method to profile lectin-microbe interactions in native communities. *Sci Adv* 2023;**9**:eadd8766. https://doi.org/10.1126/sciadv.add8766
28. Scheithauer TPM, Wortelboer K, Winkelmeijer M *et al*. Bacteriophages from treatment-naive type 2 diabetes individuals drive an inflammatory response in human co-cultures of dendritic cells and t cells. *Gut Microbes* 2024;**16**:2380747. https://doi.org/10.1080/19490976.2024.2380747
29. Hanske J, Schulze J, Aretz J *et al*. Bacterial polysaccharide specificity of the pattern recognition receptor langerin is highly species-dependent. *J Biol Chem* 2017;**292**:862-71. https://doi.org/10.1074/jbc.M116.751750
30. Gul F, Herrema H, Davids M *et al*. Gut microbial ecology and exposome of a healthy pakistani cohort. *Gut Pathog* 2024;**16**:5. https://doi.org/10.1186/s13099-024-00596-x
31. Costea PI, Zeller G, Sunagawa S *et al*. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol* 2017;**35**:1069-76. https://doi.org/10.1038/nbt.3960
32. Rognes T, Flouri T, Nichols B *et al*. Vsearch: A versatile open source tool for metagenomics. *PeerJ* 2016;**4**:e2584. https://doi.org/10.7717/peerj.2584
33. Callahan BJ, Wong J, Heiner C *et al*. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res* 2019;**47**:e103. https://doi.org/10.1093/nar/gkz569
34. Yilmaz P, Parfrey LW, Yarza P *et al*. The silva and "all-species living tree project (ltp)" taxonomic frameworks. *Nucleic Acids Res* 2014;**42**:D643-8. https://doi.org/10.1093/nar/gkt1209
35. Katoh K, Misawa K, Kuma K *et al*. Mafft: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res* 2002;**30**:3059-66. https://doi.org/10.1093/nar/gkf436
36. Price MN, Dehal PS, Arkin AP. Fasttree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;**5**:e9490. https://doi.org/10.1371/journal.pone.0009490
37. McMurdie PJ, Holmes S. Shiny-phyloseq: Web application for interactive microbiome analysis with provenance tracking. *Bioinformatics* 2015;**31**:282-3. https://doi.org/10.1093/bioinformatics/btu616
38. Oksanen J KR, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, et al. The vegan package. *Commun Ecol Packag* 2007;**10**:631-37.
39. Jackson MA, Pearson C, Ilott NE *et al*. Accurate identification and quantification of commensal microbiota bound by host immunoglobulins. *Microbiome* 2021;**9**:33. https://doi.org/10.1186/s40168-020-00992-w
40. Megrian D, Taib N, Witwinowski J *et al*. One or two membranes? Diderm firmicutes challenge the gram-positive/gram-negative divide. *Mol Microbiol* 2020;**113**:659-71. https://doi.org/10.1111/mmi.14469
41. Tanaka J, Gleinich AS, Zhang Q *et al*. Specific and differential binding of *N*-acetylgalactosamine glycopolymers to the human macrophage galactose lectin and asialoglycoprotein receptor. *Biomacromolecules* 2017;**18**:1624-33. https://doi.org/10.1021/acs.biomac.7b00228
42. van Dalen R, De La Cruz Diaz JS, Rumpret M *et al*. Langerhans cells sense *Staphylococcus aureus* wall teichoic acid through langerin to induce inflammatory responses. *mBio* 2019;**10** https://doi.org/10.1128/mBio.00330-19
43. Mnich ME, van Dalen R, Gerlach D *et al*. The c-type lectin receptor MGL senses *N*-acetylgalactosamine on the unique *Staphylococcus aureus* st395 wall teichoic acid. *Cell Microbiol* 2019;**21**:e13072. https://doi.org/10.1111/cmi.13072

44. Bambauer A, Rainey FA, Stackebrandt E *et al*. Characterization of *Aquamicrobium defluvii* gen. Nov. Sp. Nov., a thiophene-2-carboxylate-metabolizing bacterium from activated sludge. *Arch Microbiol* 1998;**169**:293-302. https://doi.org/10.1007/s002030050575
45. Hendriks A, van Dalen R, Ali S *et al*. Impact of glycan linkage to *Staphylococcus aureus* wall teichoic acid on langerin recognition and langerhans cell activation. *ACS Infect Dis* 2021;**7**:624-35. https://doi.org/10.1021/acsinfecdis.0c00822
46. Stegmann F, Lepenies B. Myeloid C-type lectin receptors in host-pathogen interactions and glycan-based targeting. *Curr Opin Chem Biol* 2024;**82**:102521. https://doi.org/10.1016/j.cbpa.2024.102521
47. Reboldi A, Arnon TI, Rodda LB *et al*. IgA production requires b cell interaction with subepithelial dendritic cells in peyer's patches. *Science* 2016;**352**:aaf4822. https://doi.org/10.1126/science.aaf4822
48. Saba K, Denda-Nagai K, Irimura T. A C-type lectin MGL1/CD301a plays an anti-inflammatory role in murine experimental colitis. *Am J Pathol* 2009;**174**:144-52. https://doi.org/10.2353/ajpath.2009.080235
49. van Vliet SJ, Gringhuis SI, Geijtenbeek TB *et al*. Regulation of effector t cells by antigen-presenting cells via interaction of the C-type lectin MGL with CD45. *Nat Immunol* 2006;**7**:1200-8. https://doi.org/10.1038/ni1390
50. Jameson B, Baribaud F, Pohlmann S *et al*. Expression of DC-SIGN by dendritic cells of intestinal and genital mucosae in humans and rhesus macaques. *J Virol* 2002;**76**:1866-75. https://doi.org/10.1128/jvi.76.4.1866-1875.2002