[i]**Jack Galbraith, Julien Legrand, Nicholas Muller, Katie Togher, Nicholas Matigian, Betoul Baz, Grant Morahan, Graeme Walker, Mark Morrison, Kiarash Khosrotehrani**

The University of Queensland, UQ Diamantina Institute, Experimental Dermatology Group, Brisbane, Australia

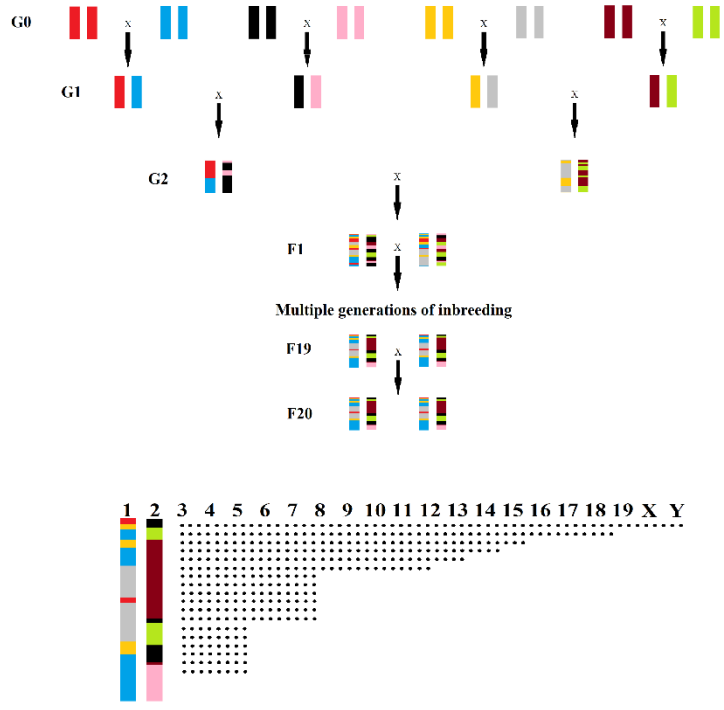
The University of Queensland, UQ Diamantina Institute, Microbial biology and metagenomics group, Brisbane, Australia

Unpublished data

Background

All animals are colonised by microbes soon after birth. In recent years, these "microbiomes" have been implicated in driving a wide range of host-responses relevant to homeostasis, or alternatively, its disruption that manifests in a wide range of immune-mediated and/or metabolic related diseases (Catinean, Neag, Mitre, Bocsan, & Buzoianu, 2019; Schoeler & Caesar, 2019). The gut (stool) microbiome has been the most intensively studied and has advanced the concept of "community-based" interactions that trigger conditions including atopic dermatitis, obesity, and mood disorders (Forsythe, Sudo, Dinan, Taylor, & Bienenstock, 2010; Kobyliak, Virchenko, & Falalyeyeva, 2016; Pessi, Sütas, Hurme, & Isolauri, 2000). Although less studied, variations in the skin microbiome have been associated with episodes of atopic dermatitis (Kong et al., 2012), and affecting the healing rate of leg ulcers (e.g.(Min et al., 2020; Wolcott, Gontcharova, Sun, & Dowd, 2009). The skin microbiome has been shown to vary according to patient age and body site (Grice et al., 2009) and environmental cues such as humidity and/or temperature, may explain the inter-individual variation (Grice & Segre, 2011). In that context, diet is also now considered to exert a strong effect on the microbiome, which can be linked to host response and health status (Schoeler & Caesar, 2019).

In contrast, there are relatively few studies that provide a systematic assessment of how host genotype affects microbiome composition. Broader genome-wide studies utilising a BXD inter-cross mouse model have identified many host-specific quantitative trait loci (QTL), with candidate genes involving cytokines and toll-like receptor signalling that associate with specific bacterial compositions (McKnite et al., 2012). In addition, investigation of the gut (stool) microbiota of mice from an advanced inter-cross model found 18 separate QTL associated with the relative abundance of one or more bacterial taxonomies (Benson et al., 2010). Regarding skin, most studies to date have been limited to candidate approaches, such as how matriptase mutations can lead to a shift in the skin microbiota composition (Scharschmidt et al., 2009), or have described the skin microbial populations more generally (Byrd et al., 2017; Grice et al., 2008; Oh et al., 2016).



Cross-breeding program design for the Collaborative Cross. Resulting recombinant inbred strains possess over 90% homozygosity.

Specialized cross-breeding programs that make use of recombinant-inbreeding can greatly assist in the analysis of complex genotype-environment interactions and gene pleiotropies (Churchill et al., 2004; Threadgill & Churchill, 2012). The Collaborative Cross (CC) is a breeding program that made use of 8 founding strains of mice to produce many hundreds of recombinant mouse strains (Churchill et al., 2004). These recombinant strains were inbred over multiple generations, ultimately creating unique strains of mice whose genome's possess > 90% homozygosity (Morgan & Welsh, 2015).

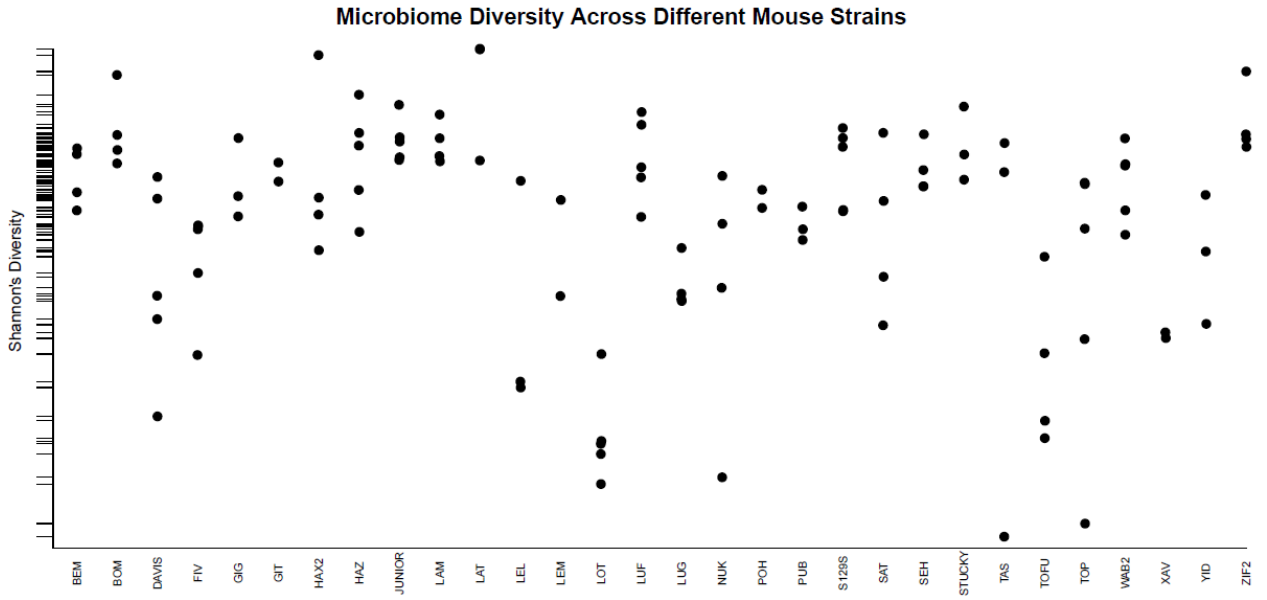
This high rate of homozygosity within recombinant-inbred (RI) strains ensures that each region of the genome is defined by the genetic contribution from a single founder strain (haplotype). However, the large heterogeneity in haplotypes between different strains can result in markedly different phenotypes. Greatly reduced costs and complexities associated with the sequencing and modelling of RI strain genotypes can be achieved with CC mice, due to the free availability of previous sequencing results and modelling resources (Churchill et al., 2004). The large genetic scope of the CC RI strains, when combined with the aforementioned positives, provides an excellent resource to investigate murine genetics of complex biological problems.

In this study, we propose that the host genome is a major determinant of skin microbiota composition. Using 16S rRNA gene amplicon profiling of the skin microbiota from 30 strains of mice from the Collaborative Cross, we show that the strain of mouse significantly affects overall skin microbiota composition. We successfully replicated a previously identified QTL at the suggestive level and identified a new QTL at the genome wide significance level. The variability of healthy mouse skin microbiota between different strains has implications for any research work investigating the skin microbiota and mouse host phenotype interactions such as wound healing and skin disease experiments.

Results

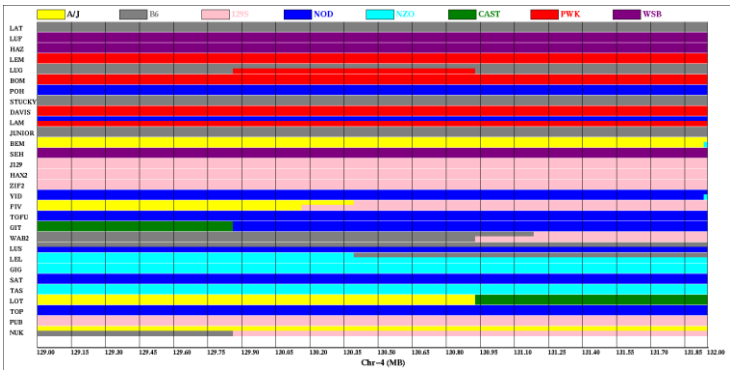
Microbiome diversity varies across mouse strains with Staphylococcus and Aerococcus differentially abundant in high and low diversity mice

The diversity of all mice was calculated using Shannon's diversity index from the 'vegan' package. While overall most mice lay in a relatively small range of diversity values, some mice clearly showed much lower diversity. Additionally, most strains showed relatively little intra-strain variation compared to significant inter-strain variation (Kruskal-Wallis *p*-value<0.0001). To investigate if any microbes associated with lower or higher diversity



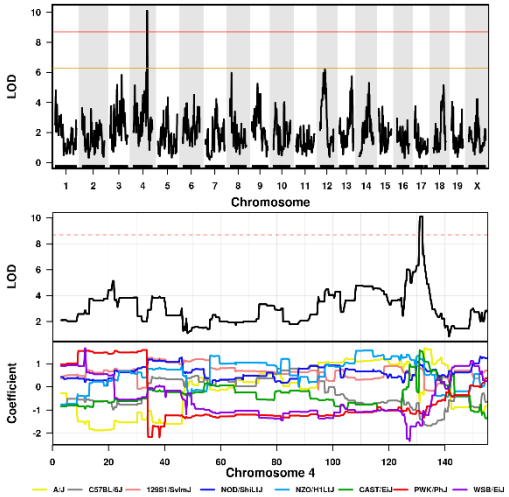
compositions we used the Aldex2 package to identify differentially abundant microbial genus'. This identified Staphylococcus and Aerococcus as significantly differentially abundant across the highest and lowest quartiles respectively (Staphylococcus *p*-value=0.0019, Aerococcus *p*-

value=0.0028). Both microbe abundances were associated with a lower diversity of the host mouse.



Staphylococcus haplotype diagram at Chr4 129-132 Mbps.

We performed GWAS using the CLR abundances of Staphylococcus and Aerococcus across the mouse strains using the GeneMiner software to identify potential QTL.

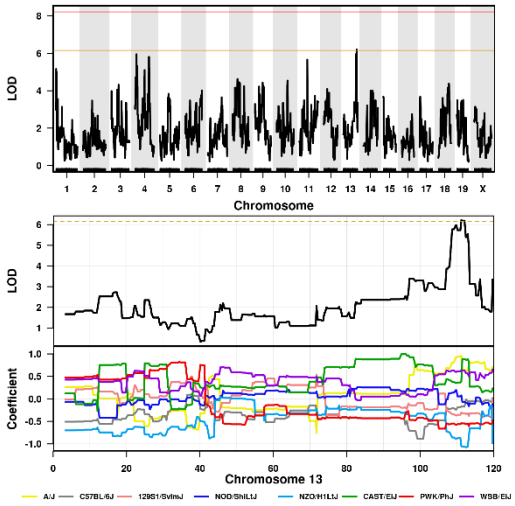


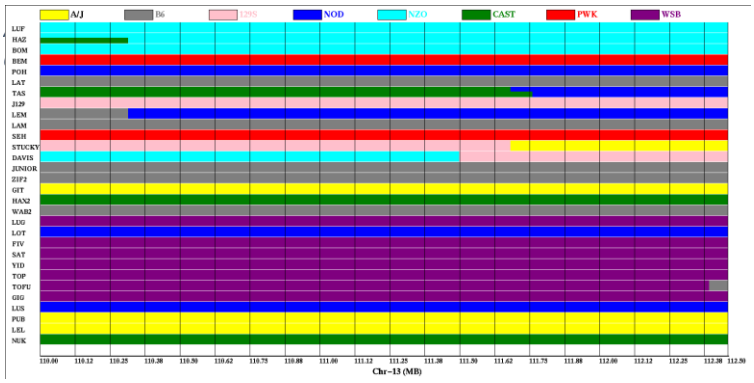
Staphylococcus GWAS showing a significant peak on Chr4 at ~130Mbps

Stahpylococcus showed a significant peak on Chr4 for the region 129-132 Mbps. From the haplotype diagram the peak appears to be a result of segregation of the PWK, NZO and WSB founder contributions.

Aerococcus shows a suggestive peak on Chr13, and while not reaching significance this peak possesses very good segregation of the haplotype diagram. With that said, the lower significance combined with good segregation of genotypes indicates a likely smaller effect size for Aerococcus vs Staphylococcus.

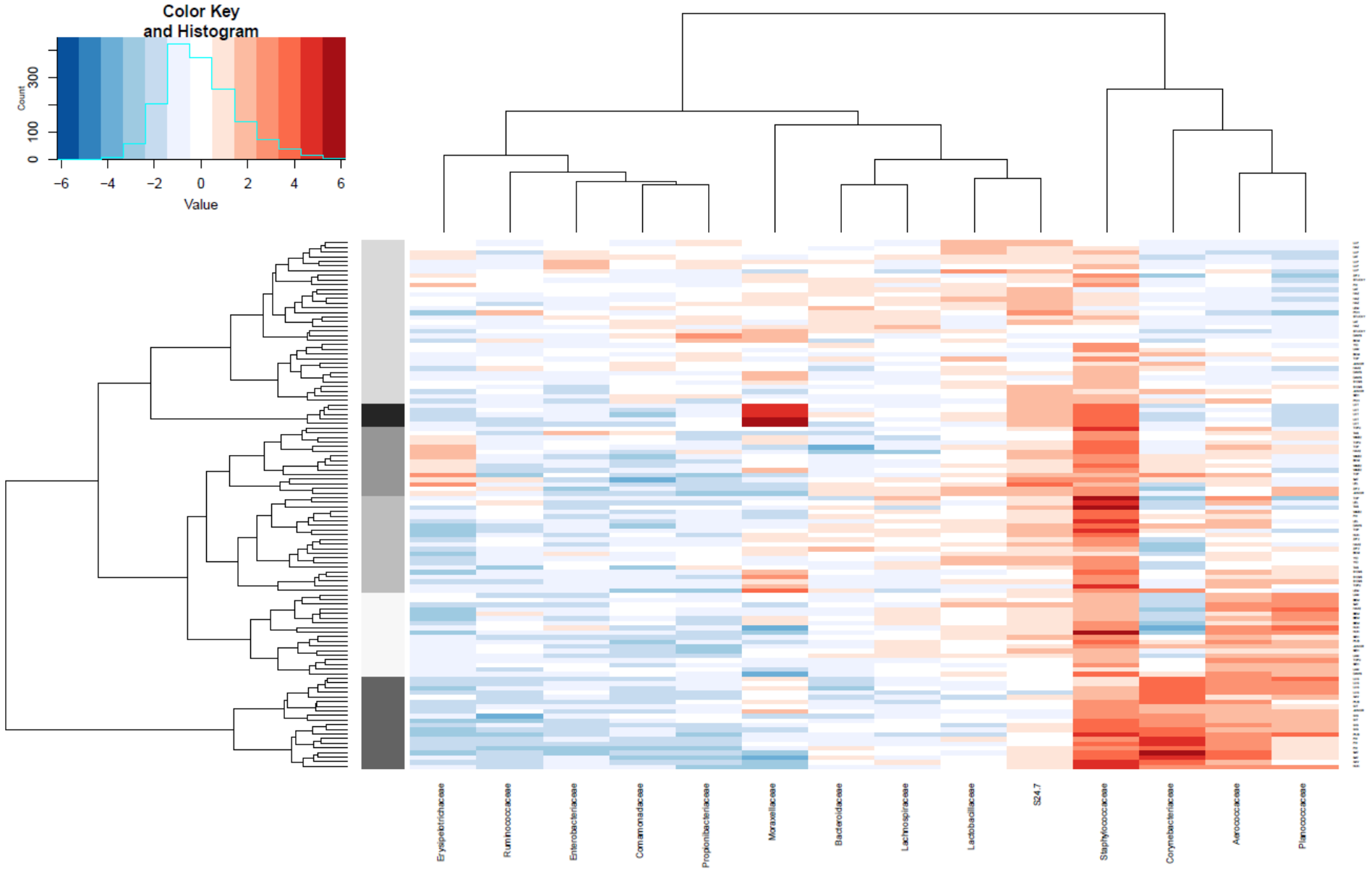
Aerococcus GWAS showing a suggested peak at Chr13 at ~110Mbps





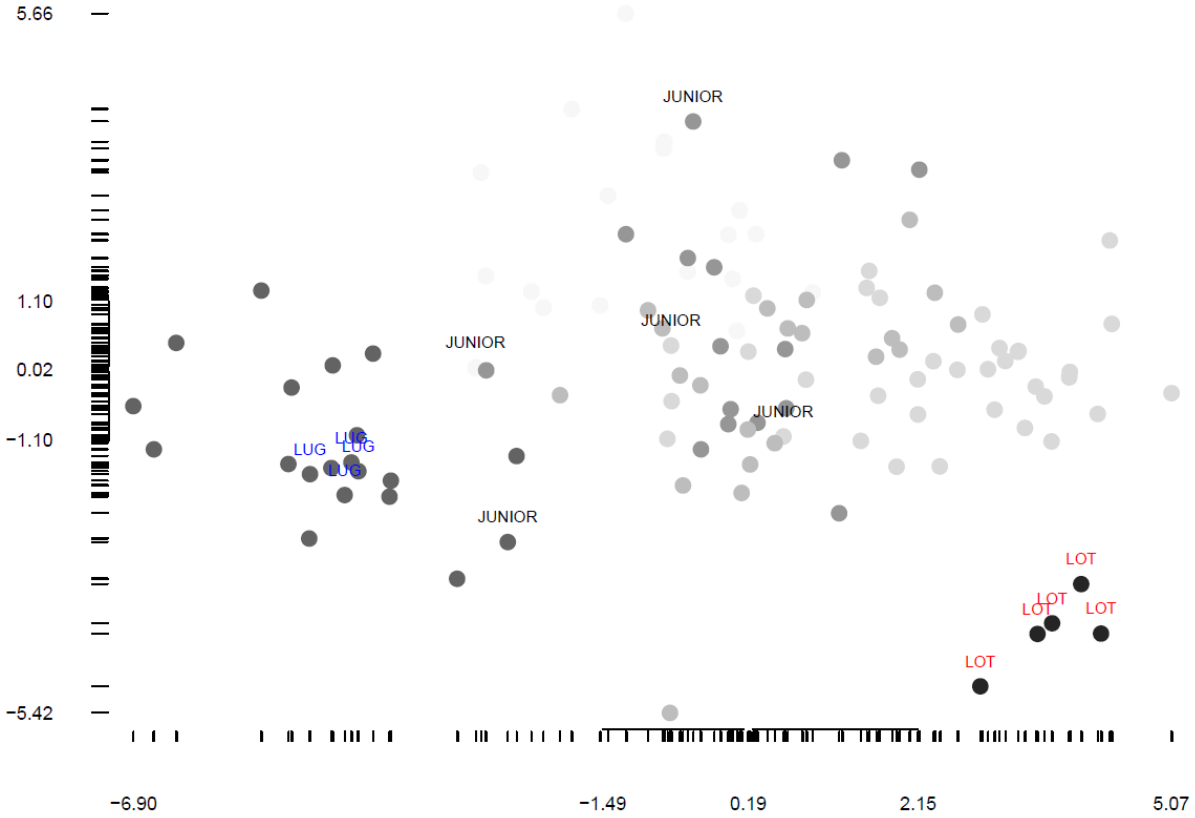
Microbiome composition at the family-level is significantly associated with mouse strain, with PCA based GWAS replicating the significant QTL on Chr4, 129-132 Mbps.

Using the CLR's across all mice we created a hierarchical clustering heatmap which shows patterns of microbiome composition across the mice. From the diagram we can see certain mice associating with higher relative abundances of Corynebacteriaceae than other strains. Additionally the LOT strain possess a much higher abundance of Moraxellaceae than other strains. Permanova analysis of the microbiome composition using the strain of mouse for grouping indicates a significant strain effect on micobiome composition (p-value=0.001).



Hierarchical Clustering Heatmap of Centred Log-Ratio Abundances. Staphylococcaceae is the most abundant family of bacteria, while Corynebacteriaceae appears as the most differentially abundant across strains. The LOT mouse strain shows particularly high levels of Moraxellaceae (though it is not clear if this is due to varying microbiome composition or purely an infection, given how markedly different this strain is to the others).

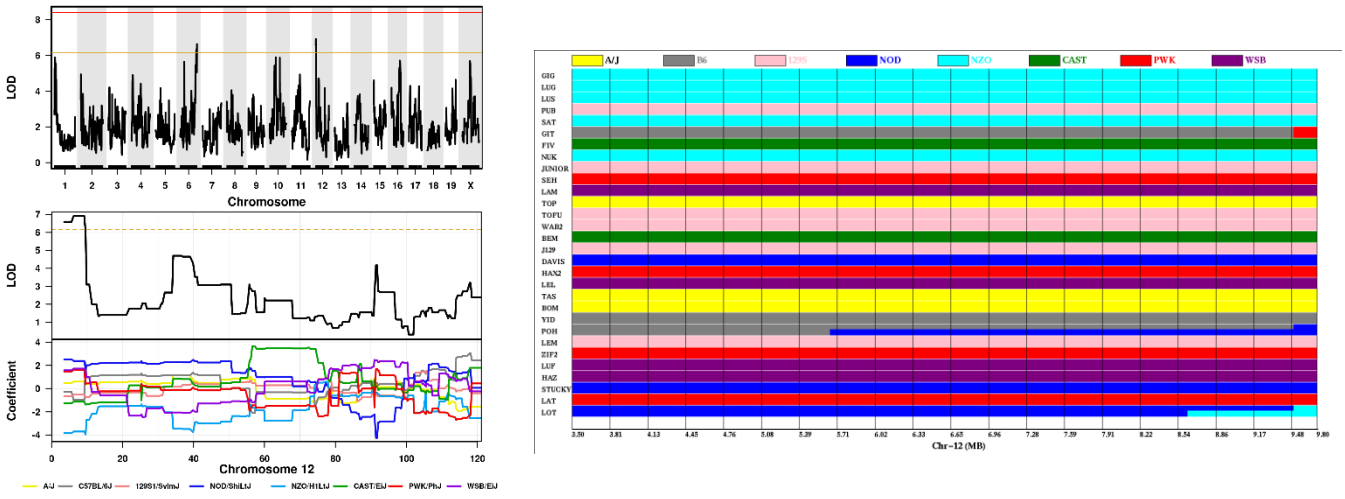
We also performed principle component analysis from which we can see LOT strain clustering on its' own, across PC1 and PC2, while the high Corynebacteriaceae cluster (dark grey) also appears on its own. Whilst strains such as LOT and LUG cluster nicely together



Principle component analysis of the skin microbiome using Centred Log Ratios.

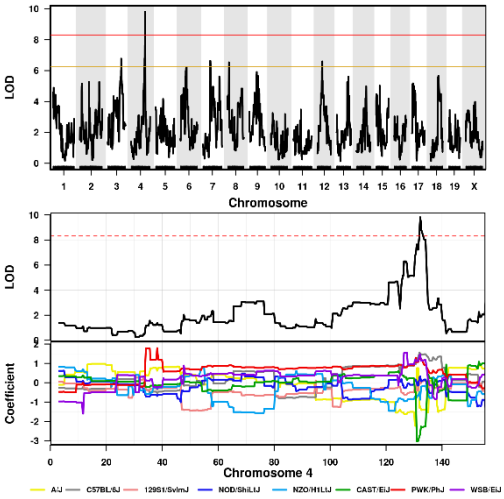
showing limited variability across PC1 and PC2, other strains such as JUNIOR show much greater variability. It's possible that this is due to higher heritability of microbiome composition in some strains than others, indicating a likely host genetic component in skin microbiome composition.

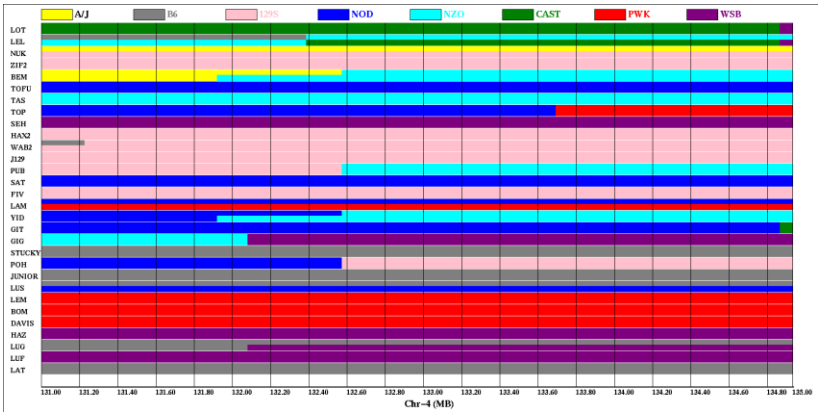
We used the principle components as inputs to the GeneMiner software to create a PCA GWAS that again identified Chr4 ~130Mbps as a significant QTL, and replicated Chr12 as a suggested QTL (identified by previous authors as significant in the gut microbiome).



GWAS analysis using PC1 regressed onto mouse genotype. A suggestive peak on Chr12 appears largely due to the founder haplotype NZO

GWAS analysis using PC3 regressed onto mouse genotype. The same significant peak identified by the Staphylococcus GWAS can be seen on Chr14, this time with segregation largly due to the CAST, PWK and B6 founder haplotypes.





Summary

Microbiome composition is mouse strain dependent with the genus' Staphylococcus and Aerococcus associating with lower diversity compositions. GWAS of Staphylococcus centred log-ratio abundance identified a significant QTL on Chr4 ~130Mbps. At the family-level, principle component analysis based GWAS replicated the significant QTL on Chr4 ~130Mbps as well as a QTL previously identified by two separate authors on Chr12, though we were only able to achieve suggested significance level.

Methods

Mice

All mice were housed in the UQ Centre for Clinical Research Animal Facility. All animal experimentation was conducted in accordance with institutional ethical requirements and approved by the University of Queensland Animal Ethics Committee. Only female mice were used with the number of mice per strain varying due to availability. Two mice from the strains XAV, GIT, LEM and POH were used whereas, all other strains had 3 or more mice.

Collection

Mice were anaesthetized with 2% isoflurane and a sterile rayon swab moistened with TE buffer was used to collect a microbiota sample from the dorsal skin. Samples were stored in 2ml of TE buffer at -80^{o}C for later sequencing.

Microbiota community profiling and data analyses

Microbial DNA was extracted from swab samples of dorsal skin using the Maxwell 16 LEV Buccal Swab DNA kit according to manufacturer's recommendations. The resulting DNA samples were then used to produce bar-coded PCR amplicon libraries of the V6-V8 hypervariable region of the 16s rRNA gene using the universal microbial primers with Illumina primer overhang adapters as follows:

- Forward Primer – 926F:

5’ - TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG AAA CTY AAA KGA ATT GRC GG – 3'

- Reverse Primer – 1392R:

5’ - GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GAC GGG CGG TGW GTR C – 3'

Sequencing used the Illumina MiSeq sequencing platform and protocols developed by the UQ-Australian Centre for Ecogenomics (www.ecogenomic.org). A sequence Phred quality threshold of 20 was used and sequences checked for chimeras using USEARCH version 6.1.544 (Edgar, 2010). Mapping and clustering of reads into operational taxonomic units (OTUs) with 97% identity threshold against Greengenes core set database 13.8 (DeSantis et al., 2006), was performed using Quantitative insight into Microbial Ecology (QIIME) version 1.9.1 (Caporaso, Kuczynski, et al., 2010) and PyNast (Caporaso, Bittinger, et al., 2010). OTUs were then compiled into an OUT table for further analysis.

Data Analysis

Shannon and Pielou measures of alpha and beta diversities were calculated using the ‘Vegan’ package with functions ‘diversity’ and ‘specnumber’. The percentage abundance for each bacterial family was calculated and those bacterial families failing a threshold of 0.01% abundance in 50% or more mice were removed, leaving a total of 13 ‘Core’ bacterial families remaining in the dataset. Centred log-ratios (CLR) were calculated for all mice using the 13 Core bacterial families as a sub-composition. Comparisons of murine skin microbiotas similarities were then performed using the Adonis function with Aitchison distance (Euclidean distance after CLR).

QTL Analysis

Significantly differentially abundant genus' based on the upper and lower quartiles of diversity were regressed against mouse genotypes imputed to the GeneMiner software (www.sysgen.org/GeneMiner). We assumed that correlations between samples and negative bias due to the compositional nature of the data was minimal, considering the aforementioned significant result. Two separate styles of GWAS were utilised, one based on SNP segregation, the other based on haplotype reconstructions.

We also performed principle component analysis on the centred log-ratios of abundances and regressed the principle components against mouse genotypes. This style of PCA based GWAS has been suggested as a way of identifying pleiotropic QTL due to each component representing a multivariate vector consisting of all phenotypes of interest (in this case all Core microbiota families). Again, both styles of GWAS were utilised.

Candidate Gene Shortlisting

We used haplotype diagrams for the QTL on Chr4 and Ch12, as these were significant and replicated respectively, to identify the founder strains' which contributed most to the microbial compositions. We searched Sanger UK mouse genomes project for SNPs specific to the previously identified haplotypes that result in stop lost/gained, missense variants, and frameshift variants. From these results we created a list of candidate genes.

References

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., . . . Pomp, D. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A, 107*(44), 18933-18938. doi:10.1073/pnas.1007028107

Byrd, A. L., Deming, C., Cassidy, S. K. B., Harrison, O. J., Ng, W. I., Conlan, S., . . . Kong, H. H. (2017). Staphylococcus aureus and Staphylococcus epidermidis strain diversity underlying pediatric atopic dermatitis. *Sci Transl Med, 9*(397). doi:10.1126/scitranslmed.aal4651

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., & Knight, R. (2010). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics, 26*(2), 266-267. doi:10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods, 7*(5), 335-336. doi:10.1038/nmeth.f.303

Catinean, A., Neag, M. A., Mitre, A. O., Bocsan, C. I., & Buzoianu, A. D. (2019). Microbiota and Immune-Mediated Skin Diseases-An Overview. *Microorganisms, 7*(9). doi:10.3390/microorganisms7090279

Churchill, G. A., Airey, D. C., Allayee, H., Angel, J. M., Attie, A. D., Beatty, J., . . . Consortium, C. T. (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet, 36*(11), 1133-1137. doi:10.1038/ng1104-1133

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol, 72*(7), 5069-5072. doi:10.1128/AEM.03006-05

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461. doi:10.1093/bioinformatics/btq461

Forsythe, P., Sudo, N., Dinan, T., Taylor, V. H., & Bienenstock, J. (2010). Mood and gut feelings. *Brain Behav Immun, 24*(1), 9-16. doi:10.1016/j.bbi.2009.05.058

Grice, E. A., Kong, H. H., Conlan, S., Deming, C. B., Davis, J., Young, A. C., . . . Segre, J. A. (2009). Topographical and temporal diversity of the human skin microbiome. *Science, 324*(5931), 1190-1192. doi:10.1126/science.1171700

Grice, E. A., Kong, H. H., Renaud, G., Young, A. C., Bouffard, G. G., Blakesley, R. W., . . . Program, N. C. S. (2008). A diversity profile of the human skin microbiota. *Genome Res, 18*(7), 1043-1050. doi:10.1101/gr.075549.107

Grice, E. A., & Segre, J. A. (2011). The skin microbiome. *Nat Rev Microbiol, 9*(4), 244-253. doi:10.1038/nrmicro2537

Kobyliak, N., Virchenko, O., & Falalyeyeva, T. (2016). Pathophysiological role of host microbiota in the development of obesity. *Nutr J, 15*, 43. doi:10.1186/s12937-016-0166-9

Kong, H. H., Oh, J., Deming, C., Conlan, S., Grice, E. A., Beatson, M. A., . . . Segre, J. A. (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res, 22*(5), 850-859. doi:10.1101/gr.131029.111

McKnite, A. M., Perez-Munoz, M. E., Lu, L., Williams, E. G., Brewer, S., Andreux, P. A., . . . Ciobanu, D. C. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS One, 7*(6), e39191. doi:10.1371/journal.pone.0039191

Min, K. R., Galvis, A., Baquerizo Nole, K. L., Sinha, R., Clarke, J., Kirsner, R. S., & Ajdic, D. (2020). Association between baseline abundance of Peptoniphilus, a Gram-positive anaerobic coccus, and wound healing outcomes of DFUs. *PLoS One, 15*(1), e0227006. doi:10.1371/journal.pone.0227006

Morgan, A. P., & Welsh, C. E. (2015). Informatics resources for the Collaborative Cross and related mouse populations. *Mamm Genome, 26*(9-10), 521-539. doi:10.1007/s00335-015-9581-z

Oh, J., Byrd, A. L., Park, M., Program, N. C. S., Kong, H. H., & Segre, J. A. (2016). Temporal Stability of the Human Skin Microbiome. *Cell, 165*(4), 854-866. doi:10.1016/j.cell.2016.04.008

Pessi, T., Sütas, Y., Hurme, M., & Isolauri, E. (2000). Interleukin-10 generation in atopic children following oral Lactobacillus rhamnosus GG. *Clin Exp Allergy, 30*(12), 1804-1808. doi:10.1046/j.1365-2222.2000.00948.x

Scharschmidt, T. C., List, K., Grice, E. A., Szabo, R., Renaud, G., Lee, C. C., . . . Program, N. C. S. (2009). Matriptase-deficient mice exhibit ichthyotic skin with a selective shift in skin microbiota. *J Invest Dermatol, 129*(10), 2435-2442. doi:10.1038/jid.2009.104

Schoeler, M., & Caesar, R. (2019). Dietary lipids, gut microbiota and lipid metabolism. *Rev Endocr Metab Disord, 20*(4), 461-472. doi:10.1007/s11154-019-09512-0

Threadgill, D. W., & Churchill, G. A. (2012). Ten years of the Collaborative Cross. *Genetics, 190*(2), 291-294. doi:10.1534/genetics.111.138032

Wolcott, R. D., Gontcharova, V., Sun, Y., & Dowd, S. E. (2009). Evaluation of the bacterial diversity among and within individual venous leg ulcers using bacterial tag-encoded FLX and titanium amplicon pyrosequencing and metagenomic approaches. *BMC Microbiol, 9*, 226. doi:10.1186/1471-2180-9-226