16S rRNA sequencing for profiling microbial communities in idiopathic granulomatous mastitis

Seeu Si **Ong**[1,2], Jia **Xu**[3], Choon Kiat **Sim**[3], Alexis Jiaying **Khng**[1], Peh Joo **Ho**[1,2,4], Philip Kam Weng **Kwan**[5], Aarthi **Ravikrishnan**[6], Kiat-Tee Benita **Tan**[7,8,9], Qing Ting **Tan**[10], Ern Yu **Tan**[11,12,13], Su-Ming **Tan**[14], Thomas Choudary **Putti**[15], Swee Ho **Lim**[10], Ee Ling Serene **Tang**[11,16], Niranjan **Nagarajan**[,6,17], Neerja **Karnani**[3,18,19], Jingmei **Li**[1,2*#], Mikael **Hartman**[2,4,20*]

[1]Women's Health and Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore

[2]Department of Surgery, Yong Loo Lin School of Medicine, National University Health System, National University of Singapore, Singapore 119228, Singapore

[3]Human Development, Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research, Singapore 117609, Singapore

[4]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549, Singapore

[5]NGS and Microarray Products, Thermo Fisher Scientific, Singapore and South East Asia, Singapore, Singapore

[6]Metagenomic Technologies and Microbial Systems, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore

[7]Division of Surgical Oncology, National Cancer Centre Singapore, Singapore 169610, Singapore

[8]Department of Breast Surgery, Singapore General Hospital, Singapore 169608, Singapore

[9]Department of General Surgery, Sengkang General Hospital, Singapore 544886, Singapore

[10]KK Breast Department, KK Women's and Children's Hospital, Singapore 229899, Singapore

[11]Department of General Surgery, Tan Tock Seng Hospital, Singapore 308433, Singapore

[12]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore

[13]Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, Singapore 1 38673, Singapore

[14]Division of Breast Surgery, Changi General Hospital, Singapore 529889, Singapore

[15]Department of Pathology, National University Health System, Singapore 119228, Singapore

[16]Department of Surgery, Woodlands Health Campus, Singapore 768024, Singapore

[17]National University of Singapore Medicine Infectious Diseases, Yong Loo Lin School of Medicine, National University Health System, National University of Singapore, Singapore 119228, Singapore

[18]Data Hub Division, Bioinformatics Institute, Agency for Science, Technology and Research, Singapore 138671, Singapore

[19]Department of Biochemistry, Yong Loo Lin School of Medicine, National University Health System, National University of Singapore, Singapore 119228, Singapore

[20]Department of Surgery, University Surgical Cluster, National University Health System, Singapore 119228, Singapore

* Equal contribution and senior authorship

Correspondence to:

Dr Jingmei Li, Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01, Singapore 138672, Singapore. Tel: (65) 6808 8312; Email: lijm1@gis.a-star.edu.sg

ABSTRACT

Background

Idiopathic granulomatous mastitis (IGM) is a rare, chronic, and benign inflammatory breast disease with ambiguous aetiology. Contrastingly, lactational mastitis (LM) is commonly diagnosed in breastfeeding women. To investigate IGM aetiology, we profiled the microbial flora of pus and skin in IGM and LM patients.

Methods

IGM patients were recruited from 2018 to 2020; LM patients were recruited as controls. 16S rRNA sequencing libraries were constructed from 16S rRNA gene amplified from total DNA extracted from pus and skin swabs. Metagenomic analysis was conducted using modified microbiome abundance analysis-suite customised R-resource for paired pus and skin samples. Microbiome multivariable association analyses were performed using linear models.

Results

21 IGM and 3 LM paired pus and skin samples underwent metagenomic analysis. Bray-Curtis ecological dissimilarity distance showed dissimilarity across four sample types (IGM-pus, IGM-skin, LM-pus, LM-skin; PERMANOVA, $p<0.001$). No characteristic dominant genus was observed across IGM samples. IGM pus samples were more diverse than corresponding IGM skin samples (Shannon and Simpson index; Wilcoxon paired signed-rank tests, $p=0.022$ and $p=0.07$). *Corynebacterium kroppenstedtii*, reportedly associated with IGM in literature, was higher in IGM pus samples than paired skin sample (Wilcoxon, $p=0.022$). Three other species and nineteen genera were statistically significant in paired IGM pus-skin comparison after antibiotic treatment adjustment, and multiple comparisons correction.

Conclusions

Microbial profiles are unique between IGM and LM patients. Inter-patient variability and polymicrobial IGM pus samples cannot implicate specific genus or species as an infectious cause for IGM.

Keywords: Idiopathic granulomatous mastitis, metagenomic sequencing, microbiota, 16S rRNA, Corynebacterium, MaAsLin 2

LIST OF ABBREVIATIONS

Abbreviation	Full description
AFB	Acid-fast bacillus
BLAST	Basic Local Alignment Search Tool
CGH	Changi General Hospital
DNA	Deoxyribonucleic acid
EMIRGE	Expectation maximization iterative reconstruction of genes from the environment
GMS	Grocott (methanmine) silver
IGM	Idiopathic granulomatous mastitis
IQR	Interquartile range
IRB	Institutional Review Board
KKH	KK Women's and Children's Hospital
LM	Lactational mastitis
MDS	Multidimensional scaling
MRSA	Methicillin-resistant Staphylococcus aureus
NUH	National University Hospital
PAS	Periodic acid-Schiff
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance
PF	Purity filter
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
SGH	Singapore General Hospital
TB	Tuberculosis
TTSH	Tan Tock Seng Hospital

INTRODUCTION

Idiopathic granulomatous mastitis (IGM) has been consistently described as a rare and chronic inflammatory infliction of the breast[1,2]. The condition, while presenting clinically and radiologically with features that may mimic breast malignancy, is benign[1,2]. Only when histopathological examination of the breast tissue for noncaseating granulomatous inflammation in the lobules, coupled with the in-depth investigation of each case eliminating fungal, tuberculous, and other causes of granulomatous mastitis, would IGM be diagnosed[3-5]. This complex investigative work-up means IGM is usually only diagnosed in high resource settings, and by clinicians who have awareness of the differential. Furthermore, without a structured disease registry for proper documentation of IGM, an accurate estimation of the global disease incidence and prevalence is absent. Limited case series and study populations published in literature is affirmed by the anecdotal experiences of the rare nature of the disease[6].

Hypothesised aetiologies of IGM have include a possible microbial cause[7]. *Corynebacterium* was first associated with IGM when Taylor, et al. (2003) reported coryneform bacteria observed in histological examination of breast tissue and *Corynebacterium* species isolated in cultures from patient specimens[8]. Particularly, *Corynebacterium kroppenstedtii*, *C. amycolatum*, *C. tuberculostearicum*, and *C. accolens* species of *Corynebacterium* have been isolated from IGM patient samples[8-10]. However, most studies investigating microbial involvement in IGM have relied on bacterial culture of patient pus samples to isolate species that are subsequently identified through polymerase chain reaction (PCR) and 16S rRNA sequencing[8,11]. Pathogens incompatible with traditional culture methods cannot be detected through such methods.

Another benign breast inflammation that presents similarly in clinical settings is lactational mastitis (LM), a common complication in breastfeeding women[12,13]. Diagnosis usually requires targeted ultrasound and histopathological examination of biopsied breast tissue to confirm the benign inflammatory condition, and rule out malignancy[14]. As a similarly benign and inflammatory breast disease, LM provides an appropriate comparison group to IGM. Antibiotics is a similar first-line treatment for both conditions, as is the common practise to prescribe patients presenting in outpatient settings with symptoms similar to IGM and LM patients (localised symptoms: erythema, warmth,

oedema, and tenderness; systemic symptoms: fever, and malaise) with antibiotics[12,15]. The dominant *Staphylococcus* genus reported in the microbiology for breast milk in women affected with LM can be compared against the *Corynebacterium* genus associated with IGM[16-20]. LM will serve as a suitable condition to make metagenomic comparisons against while exploring microbial profiles from IGM patient samples.

Our work expands on Yu, et al. (2016) in utilising 16S rRNA metagenomic analysis for a wholesale inspection of the microbiome of IGM patient pus samples, to identify microbial agents missed from traditional culture methods[11]. Additionally, our study also profiles the corresponding skin microbial flora of IGM patients. Gaining a better understanding of the microbial flora could potentially identify aetiological pathogens that could be targeted in developing detection assays for potentially simplified disease diagnostics. The involvement of such pathogens in the pathophysiology of IGM could also guide clinical disease management through targeted treatment or for monitoring disease progression and prognosis, with the tendency for recurrence in IGM[21-24].

MATERIALS AND METHODS

Study Population

The study population was adult patients (aged 21 years and above) suspected, diagnosed with or treated for mastitis (IGM or LM) at five participating hospitals in Singapore: National University Hospital (NUH), Tan Tock Seng Hospital (TTSH), KK Women's and Children's Hospital (KKH), Singapore General Hospital (SGH), and Changi General Hospital (CGH). Pregnant and breastfeeding women, and lactating and non-lactating women were allowed to participate in this study. Patient information on demographic, lifestyle, reproductive, past treatment or environmental exposure variables were collected from questionnaires. All studies were performed in accordance with the Declaration of Helsinki. All participants provided written informed consent. This study was approved by the National Healthcare Group Domain Specific Review Board (reference number: 2017/01057) and the Agency for Science, Technology and Research Institutional Review Board (reference number: 2020-152).

Patient Recruitment and Sample Collection

Adult female patients with IGM or LM were recruited from the five participating hospitals in Singapore between 2018 and 2020. Target recruitment was originally 60 participants per site, totalling to 300 participants in the study. During the recruitment time period, 148 patients were screened, 91 patients enrolled in the study, and 89 patients completed the study (2 withdrawals).

A total of 32 patients donated skin and/or pus samples, of which 26 were IGM patients and 6 were LM patients (Supplementary Table 1). Study coordinators were informed by clinicians and physicians of potential mastitis cases, after which, study coordinators would seek consent, collect patient samples and conduct the questionnaire. Patients were subsequently followed up for clinical diagnosis confirmation. Non-lactating patients presenting with mastitis and/or breast abscess were diagnosed as IGM by histological examination of breast core biopsy for non-caseating granulomatous formation. Other conditions were excluded via negative stains for acid-fast bacillus (AFB) to rule out *Mycobacterium tuberculosis* infection, and negative stains for Grocott (methanmine) silver (GMS) stain or Periodic acid-Schiff (PAS) to rule out fungal infections. Histopathology also confirmed non-malignancy. Patients presenting with mastitis and/or breast abscess confirmed to be lactating at the point of recruitment were diagnosed as LM.

Patient samples were collected using DNA/RNA Shield Lysis Tubes w/Swab (Microbe) (catalog number R1104; Zymo Research, California, USA). Sterile methods were used to collect pus samples from the infected breast, while skin samples were collected from the contralateral (non-infected breast) without site cleansing prior to sample collection. The kit sample collection tubes, that were pre-filled with DNA/RNA Shield and a lysis reagent for microbial samples, were labelled before sample collection. Swab was removed from packaging without touching the applicator tip.

- Pus sample collection: Swab was allowed to absorb infected fluid for 30 seconds while rotating the swab.
- Skin sample collection: Swab was used to rub skin area for 10 to 15 strokes with moderate pressure. Swab was rotated and sampling was repeated.

Swab was broken along the break-point and the applicator head was left in the collection tube. Each collection tube was capped and shaken 10 times to mix the sample collected with pre-filled reagents.

DNA Extraction and 16S ribosomal rRNA Sequencing

Total DNA was extracted with ZymoBIOMICS™ 96 MagBead DNA Kit (catalog number D4306; Zymo Research, California, USA). Samples were mechanically lysed with beads beating at maximum speed for 20 minutes. DNA was purified from lysed samples with magnetic beads (ZymoBIOMICS™ MagBinding Beads) and eluted with ZymoBIOMICS™ DNase/RNase Free Water. 16S ribosomal RNA (rRNA) gene amplification and sequencing has been previously described[25]. Briefly, the V2-V6 region (723bp PCR product) of the 16S rRNA gene that amplified via PCR using primers 338F (Seq: 5'-ACTYCTACGGRAGGCWGC-3') and 1061R (Seq: 5'-CRRCACGAGCTGACGAC-3') with the HotStar HiFidelity Polymerase Kit (catalog number 202602; QIAGEN, Hilden, Germany). Supplementary Figure 1 shows the 16S rRNA gene map and the primers used. PCR parameters were initial denaturation at 95°C for 5 minutes, followed by 35 PCR cycles of

1. Denaturation at 95°C for 30 seconds
2. Annealing at 59°C for 30 seconds, and
3. Extension at 72°C for 1 minute.

The final elongation was at 72°C for 6 minutes. QIAquick PCR Purification Kit (catalogue number 28106; QIAGEN, Hilden, Germany) was used to purify 700-1,000 bases PCR products. Amplicons were sheared with Covaris Adaptive Focused Acoustics™ to approximately 180 base fragments. 16S libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Modules for Illumina® (catalogue number E7645L; New English Biolabs, Ipswich, Massachusetts, United States of America) according to the manufacturer's protocol. Multiplexing Sample Preparation Oligonucleotide Kit from Illumina, Inc. (San Diego, California, United States of America) was used to label the DNA sequencing libraries with different multiplex indexing barcodes[25]. 200pM concentration 16S libraries were multiplexed paired-end sequenced on HiSeq4000 with 2 x 150bp paired-end reads.

Sequencing Data Processing

Processing of sequencing data has been previously described[25]. Briefly, data was de-multiplexed and sequencing reads that passed Illumina's purity/chastity filter (PF=0) were converted to FASTQ format. Trailing bases with quality scores lower or equal to 2 were subsequently filtered and trimmed from the 3' end; read pairs containing reads shorter than 120 bases were removed[26]. An adapted expectation maximization iterative reconstruction of genes from the environment (EMIRGE)

assembly, was used to reconstruct the amplicons of the shorter-read dataset (EMIRGE was originally designed for whole genome datasets) by only performing analysis with top 100,000 average quality reads[25]. Reconstructed sequences were trimmed according to the region amplified by the selected primers; reconstructed sequences were searched against the complete Greengenes database (dated May 1, 2019; gg_13_5/gg_13_5_otus.tar.gz/rep_set/99_otus.fasta, from https://greengenes.secondgenome.com/?prefix=downloads/greengenes_database/gg_13_5/) using Basic Local Alignment Search Tool (BLAST)[27]. The top hit on BLAST by lowest E-value, highest bit-score, highest percent identity and longest alignment length, in that particular order, was used to classify the sequence[25]. Percentage identity for species classification must be ≥97% (≥95% for genus classification) to be classified.

Statistical Analysis

Statistical analysis was performed using a modified microbiome abundance analysis-suite (https://github.com/BetaCollins/MBAA-Suite) customized R resource that contains the analysis functions of 16S pipeline (https://github.com/CSB5/GERMS_16S_pipeline) developed at the Genome Institute of Singapore, designed for Illumina shotgun sequencing of 16S rRNA amplicon sequences[25]. Microbial populations were evaluated at species and genus taxonomy levels.

Microbial populations of pus samples were compared between IGM and LM controls, and repeated in skin samples. Shannon (formula (1)) and Simpson (formula (2)) indices were used to measure alpha-diversity.

$$Shannon\ Index\ (H) = -\sum_{i=1}^{s} p_i \ln p_i \qquad \text{– formula (1)}$$

$$Simpson\ Index\ (D) = \frac{1}{\sum_{i=1}^{s} {p_i}^2} \qquad \text{– formula (2)}$$

Multidimensional scaling was used for visualisation of Bray-Curtis distance between the microbial profiles of the IGM and LM sample types. Permutational multivariate analysis of variance (PERMANOVA) was used to make comparisons of mastitis and sample types centroids and dispersion within groups; statistically significance was defined by measure of space not equivalent for all groups. Relative abundance of *Corynebacterium* genera was evaluated in pus and skin samples

from IGM and LM patients. Paired sample analysis with Wilcoxon signed-rank test, compared *Corynebacterium* genera relative abundance in pus and skin samples from the same patient in IGM patients. Microbiome multivariable association was performed with linear models using MaAsLin 2 (https://github.com/biobakery/biobakery/wiki/maaslin2)[28] to compare the microbial populations between IGM and LM patient samples; paired sample analyses, and adjusted analyses with antibiotic treatment, and duration after antibiotic treatment as confounders were also performed.

All analyses were performed with R (version 4.0.2) unless otherwise stated. Datasets utilised in analysis can be made available from the corresponding author upon reasonable request, within limitations of the study Institutional Review Board (IRB).

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request, within limitations of the study Institutional Review Board (IRB).

RESULTS

Study Population

As per Supplementary Table 1, 26 IGM and 6 LM patients donated skin and/or pus samples for 16S rRNA gene sequencing; 21 IGM and 4 LM patients donated paired (pus and skin samples from the same patient) samples that underwent DNA extraction and 16S rRNA gene sequencing. Only 21 paired IGM samples and 3 paired LM samples subsequently underwent metagenomic analysis; the sequencing file for the remaining LM patient's samples were corrupted and could not be analysed.

IGM patients (median age = 34 years) were slightly older than LM patients (median age = 31 years), although this difference was not significant (Table 1). Other demographic variables reported are site of recruitment, year of diagnosis, ethnicity, body mass index, and education level (Table 1). Patient characteristics reported were any smoking history, chronic illness diagnosis, autoimmune conditions, previous infectious disease diagnosis, previous cancer diagnosis, and mastitis treatment

history prior to sample collection (Table 1). None of the demographic or patient variables were significantly different between IGM and LM patients. LM patients served as controls for IGM patients in this study.

Comparing IGM vs LM microbial taxonomic profiles

Taxonomic profiles of IGM and LM pus and skin samples were predominantly examined at the genus level; this was visualised in Figure 1. *Staphylococcus* genus dominates the three LM pus samples (LM1-3): 90.8%, 26.5%, and 38.5% respectively. Contrastingly, no dominant genus was observed across the IGM pus samples (Figure 1(a)). Greater inter-patient variability was also observed in the IGM pus samples, as compared to the LM pus samples (Figure 1(a)). Multidimensional scaling (MDS) of Bray-Curtis distance between bacteria profiles in IGM and LM pus samples did not show obvious clustering separating the two patient groups; however, their Bray-Curtis distancing were significantly different (PERMANOVA, $p<0.001$) (Figure 2(b)). It must be noted that the first two MDS of Bray-Curtis distance between breast pus and skin samples for IGM patients only explains 16.98% and 12.23% of the variability respectively, and collectively explains 29.21% of the samples' variability.

No dominant genus is observed across either the IGM or LM skin samples, collectively or as separate groups (Figure 1(b)). Again, inter-patient variability is considerable (Figure 1(a)). All four sample types (IGM pus, IGM skin, LM pus, and LM skin) taken collectively, also did not show obvious clustering in MDS of Bray-Curtis distance between bacteria profiles, but, again, the distancing was significantly different (PERMANOVA, $p<0.001$) (Figure 2(a)). The paired LM samples did not appear close to their corresponding pair in the MDS. The first two MDS of Bray-Curtis distance between breast pus and skin samples for IGM and LM patients only explains 14.97% and 13.11% of the variability respectively, and collectively explain 28.08% of the samples' variability.

Both Shannon (information statistic) and Simpson (dominance) indices for measuring alpha-diversity in each sample demonstrated higher diversity in IGM pus samples (Shannon=2.65 [interquartile range, IQR=1.91-2.92]) than any other sample type (Supplementary Table 2, and Supplementary Figures 2 and 3); however, this difference was only borderline statistically significant

(Wilcoxon signed-rank test, p=0.083). Comparisons between alpha-diversity of paired pus and skin samples from the same IGM patient found that IGM pus samples were statistically more diverse than the corresponding IGM skin sample (Shannon index Wilcoxon paired signed-rank test, p=0.022; Simpson index Wilcoxon paired signed-rank test, p=0.07). This comparison was not significant for paired samples in LM patients (Shannon index Wilcoxon paired signed-rank test, p=1.000; Simpson index Wilcoxon paired signed-rank test, p=0.75). Differences in alpha-diversity between IGM pus and LM pus, or IGM skin and LM skin samples, were not statistically significant (Supplementary Figure 3). Medians and interquartile ranges for IGM and LM samples for both diversity indices are presented in Supplementary Table 2.

Corynebacterium

Focus was placed on *Corynebacterium* genera given the literature published demonstrating an association with IGM[7-11]. Inter-patient variability for *Corynebacterium* relative abundance was high amongst IGM patients, for both pus and skin samples (Figure 3). As this genus is commonly found on the skin, it was unsurprising that IGM skin samples *Corynebacterium* relative abundance was 9.4% [IQR=5.8%-30.2%] (Figure 3). However, *Corynebacterium* relative abundance in IGM pus samples was rather similar, at 9.0% [IQR=3.8%-13.0%] (Figure 3).

No obvious trend was observed in genus level pairwise comparison for *Corynebacterium* relative abundance between paired IGM pus and skin samples (Figure 3); *Corynebacterium* relative abundance was not statistically significant between the paired samples (Wilcoxon paired signed-rank test, p=0.29; paired LM pus and skin samples, p=0.25 [Figure 3; Supplementary Figure 4]). However, species level taxonomic analysis for *Corynebacterium kroppenstedtii* relative abundance was significantly different between paired IGM pus and skin samples (Wilcoxon paired signed-rank test, p=0.022 [Supplementary Figure 5]).

Associations between sample types and metagenomic features

Microbiome multivariable association between sample type, covariates and microbial metagenomic features between IGM and LM pus samples using general linear models (MaAsLin 2),

with correction for multiple comparisons false positive rate, did not find any genus that were statistically different.

However, pairwise analysis found 24 genera were significantly different between paired IGM pus and skin samples: *Ochrobactrum*, *Delftia*, *Anaerobacillus*, *Gordonia*, *Methylobacterium*, *Fusobacterium*, *Streptococcus*, *Achromobacter*, *Sphingobium*, *Alkanindiges*, *Novosphingobium*, *Capnocytophaga*, *Mycobacterium*, *Burkholderia*, *Peptoniphilus*, *Roseomonas*, *Rothia*, *Finegoldia*, *Agrobacterium*, *Anaerococcus*, *Hydrogenophaga*, *Peptostreptococcus*, *Kocuria*, and *Dermabacter*. 19 genera remained significantly different after adjusting for antibiotic treatment (treated vs not treated), and duration after antibiotics treatment (samples collected less than 2 weeks after treatment vs more than 2 weeks after treatment): *Ochrobactrum*, *Delftia*, *Anaerobacillus*, *Gordonia*, *Methylobacterium*, *Fusobacterium*, *Sphingobium*, *Alkanindiges*, *Streptococcus*, *Achromobacter*, *Capnocytophaga*, *Mycobacterium*, *Novosphingobium*, *Peptoniphilus*, *Rothia*, *Finegoldia*, *Burkholderia*, *Roseomonas*, *Anaerococcus*. Beta-estimates, standard deviations, crude and adjusted p-values, as well as crude and adjusted q-values (multiple comparisons corrected p-value) are reported in Table 2.

Table 2 also shows the beta-estimates, standard deviations, crude and adjusted p-values, as well as crude and adjusted q-values for the same pairwise analysis performed at species level taxonomy. 6 species with significantly different relative abundance between the paired IGM pus and skin samples were *Acinetobacter schindleri*, *Rothia mucilaginosa*, *Lactobacillus iners*, *Corynebacterium kroppenstedtii*, *Roseomonas mucosa*, and *Kocuria rhizophila*. The association remains significantly different after the adjustments for antibiotic treatment (treated vs not treated), and duration after antibiotics treatment (samples collected less than 2 weeks after treatment vs more than 2 weeks after treatment), for 3 species: *Acinetobacter schindleri*, *Rothia mucilaginosa*, and *Lactobacillus iners*; *Corynebacterium kroppenstedtii* was borderline significantly different between the paired IGM samples ($p<0.001$, $q=0.053$ [Table 2]).

DISCUSSION

In the previously published paper about bacteria communities in IGM pus samples by Yu, et al, 2016, clinical metagenomic analysis was performed on breast pus samples from IGM patients[11].

The paper describes varying domination of *Corynebacterium* genera across their 19 IGM patients, as compared to little to no *Corynebacterium* genera presence in their LM controls[11].

Our study enriches this existing body of work by providing relevant patient comparison via the skin sample metagenomic evaluation from the non-infected site. As such, we were able to perform pairwise comparison on the relative abundance of *Corynebacterium* between patient pus sample and their non-infected skin sample. While the Yu, et al, 2016, study amplified the hypervariable region 4 (V4) of 16S rRNA gene, our study also expands upon that with the hypervariable regions 2 to hypervariable region 6 (V2-V6) of the 16S rRNA gene amplified (Supplementary Figure 1). The genus taxonomy level analysis benefits from higher resolution in identifying bacterial populations specifically from the hypervariable regions 2 and 3 of the 16S rRNA gene[29]. Furthermore, Meisel, et al, 2016, have previously demonstrated sequencing the V1-3 hypervariable regions of the 16S rRNA gene for skin microbiome surveys, and by extension pus microbial evaluations as well, provides more precise microbial community characterisation[30]. In contrast, amplifying the V4 region of the 16S rRNA gene is more common for protocols developed for characterising microbiota of other habitats, particularly human gastrointestinal sites[30]. Additionally, the V4 region amplified in the Yu, et al, 2016, study, is a much shorter and more conserved sequence, compared to the V2-V6 regions amplified in our study[11,31]. Therefore, our study is likely to find taxonomies not previously found in the Yu, et al. 2016 study. Larger fragment amplification has also been found to improve sensitivity and specificity of sequences classified across most taxonomic levels[32].

Other differences between our study and the Yu et al study include sequencing platforms (Ion PGM Hi-Q 400 Sequencing vs. HiSeq4000) and read lengths (200/400bp vs 2x150bp paired-end reads) (Yu, et al, 2016 study corresponds for the former metrics, our study corresponds to the latter). Last of the comparisons is the Yu, et al, 2016 paper has a purely descriptive approach to evaluating the microbial population of IGM and LM pus samples, whereas our study provides statistical evaluation of the microbial genera and species.

Breast inflammation is often categorised as lactational and non-lactational[33]. LM is commonly caused by skin-colonising bacteria[33]. The most common skin-colonising bacteria implicated in LM is

Staphylococcus aureus, and increasingly Methicillin-resistant *S. aureus* (MRSA) in particular[33]. *Streptococcus pyogenes*, *Escherichia coli*, *Bacteroides species*, and Coagulase-negative *Staphylococci* have also been identified to cause LM[33]. Prior history of mastitis, nipple cracks and fissures, inadequate milk drainage, maternal stress, lack of sleep, tight-fitting bras, and use of antifungal nipple creams are some of the LM risk factors that might provide conducive environment for unhealthy bacteria growth and colonisation[34,35].

On the other hand, IGM aetiology has always been ambiguous. Possible causes of IGM span autoimmune disease, trauma, lactation, oral contraceptive pill use, and hyperprolactinemia, with a growing literature suggesting *Corynebacterium* infection association with the pathogenesis of IGM[7,36]. In addition to the impossibility of collecting pus samples from healthy women, these biological differences make LM a suitable condition to be studied in comparison with IGM. Collecting both pus and skin samples was meant to highlight the infiltration of aetiological populations from skin to pus in both sample types collected from LM, that would not be expected in IGM. The dominance of *Staphylococcus* in LM pus but not skin samples, coupled with Bray-Curtis distance differences between paired LM samples implies that LM cannot be entirely attributed to bacterial infiltration from skin colonies. This confirms that other risk factors beyond skin breakages and topical contributions play integral roles in the disease mechanism[34,35].

In IGM patients, without the skin infiltration aetiological mechanism, significantly different genus level taxonomy of microbial populations between pus and skin samples would be expected. With the increasing association found between *Corynebacterium* and IGM patient samples, as well as the increasing implication of *Corynebacterium* as a causative pathogen of IGM[8-11], we focused on studying the presence of *Corynebacterium* in our patient samples. Discerning the role of this genus as infectious, coloniser or contaminant in IGM is challenging[9]. *Corynebacterium* is a common population of Gram-positive bacteria in human skin microbiota[7]. The unique combination of presence of polymorphonuclear leukocytes with Gram-positive rods; and sterile tissue in healthy conditions but testing positive for *Corynebacterium* in IGM, could determine that this genus might be responsible[37]. Alternatively, it can be hypothesised that *Corynebacterium* might contribute to a unique skin flora that makes patients more susceptible to IGM, given the lack of significant difference between

Corynebacterium relative abundance in paired pus and skin samples. Contrastingly, there could be unique species of the *Corynebacterium* genus that contribute different roles to the disease: the significantly higher *Corynebacterium kroppenstedtii* abundance in pus samples compared to their corresponding paired skin sample (Table 2; Supplementary Figure 5) could indicate this individual species playing a systemic role in IGM aetiology. Amongst the 6 significantly different species between paired IGM samples, only the median relative abundance of *C. kroppenstedtii* exceeds 1% in IGM pus (Supplementary Table 3). This is not the first time *C. kroppenstedtii* has been associated with IGM: the species has been previously identified from patient sample cultures[38-40]. *C. kroppenstedtii* in patient samples is underreported, since routine culture methods are unlikely to detect slow growing *Corynebacterium*, and AFB and PAS stains will not detect *Corynebacterium* either[41]. *Corynebacterium* is also often not described to be tested for in IGM patient samples in the literature[40]. Further molecular studies must be conducted to determine the systemic role of higher *Corynebacterium kroppenstedtii* relative abundance in IGM pus samples.

High interpatient variability within IGM patients was observed in their different microbial profiles, both MDS only explaining less than 30% of variability with the first two reduced factors, and large IQR for both alpha-diversity indices (Supplementary Table 2, and Supplementary Figure 6), demonstrating a spread in metagenomic profiles from IGM samples. This is to be expected, with the demonstrated variability in patient demographics (Table 1). Additionally, as a disease diagnosed by exclusion, with varied clinical presentation, and yet to be determined aetiology, patient heterogeneity is to be expected, in patient demographics and microbial profiles[3-5,42]. Further investigations into IGM molecular basis and elucidating the elusive aetiology will be very helpful for classifying the disease into different subtypes. Future studies with larger sample sizes, could consider utilising the metagenomic profiles to categorise meaningful disease subtypes.

An important factor patient variable that interpatient metagenomic variability could be attributed to would be treatment exposure. Previous studies have described steroid treatment affecting microbiomes in different sites, including the lungs, gut and vagina[43-46]. As a lipid soluble drug, corticosteroids would penetrate fatty breast tissue and could modify pus microbiome in IGM[47]. Antibiotic treatment is widely documented to create great imbalances to the gut microbiome[48,49].

Evidence also describes antibiotic use to affect microbiome at other anatomical sites, and associated with diseases including affected lipid metabolism, inflammation, and auto-immune conditions[50-52]. The patient variability in treatment exposure to antibiotics and / or corticosteroids treatment, and duration from last treatment to sample collection will affect systemic microbiome, and could explain some of the interpatient metagenomic variability[51].

The 19 statistically significant genera observed between paired skin and pus samples from IGM patients after adjustment for treatment exposure and duration, and correcting for multiple comparisons would be worthy to explore. However, median relative abundance for IGM pus samples does not exceed 5% for any of the genera (Supplementary Table 4). Genera with median relative abundance exceeding 1% in IGM pus samples were *Ochrobactrum*, *Delftia*, and *Streptococcus*. *Delftia* genus are gram-negative rods, rarely involved in human infections, and more often found in environmental sites[53]. All three factors make *Delftia* an unlikely culprit for granuloma formation in breast tissue. However, there is a possibility that *Delftia* genus may contribute to a unique microbiome that makes a patient susceptible to IGM inflammation. *Streptococci*, often implicated as pathogenic in skin infections, is also observed in all paired LM samples at higher relative abundances than that of IGM paired samples (Supplementary Table 4)[54,55]. Without significant difference in relative abundance between IGM and LM pus samples, *Streptococci* is not a suitable target for elucidating IGM-specific aetiology or for clinical disease management. *Ochrobactrum* genus is an interesting avenue to explore, as opportunistic pathogens responsible for infectious outbreaks: Ryan and Pembroke (2020) found 128 separate instances, involving 289 unique cases[56]. Furthermore, abscess in non-breast sites (neck, pelvic, pancreatic and retropharyngeal) was a reported symptom of infection[56]. *Ochrobactrum* would also explain poor antibiotic response in IGM given the genus' resistance to β-lactams (penicillins, cephlasporins and emergingly carbapenem-resistance as well)[56]. However, *Ochrobactrum* is largely an opportunistic pathogen (infected patients are usually immunocompromised with cancer, or diabetes-causing kidney failure, acquiring the infection through catheter or dialysis in clinical settings)[56], whereas the IGM patients in our study are self-reported without autoimmune conditions, cancer history, or active infectious disease (Autoimmune conditions: Coeliac disease, Type 1 diabetes mellitus, Graves' disease, Inflammatory bowel disease, Multiple sclerosis, Psoriasis, Rheumatoid arthritis, or Lupus erythematosus; Infectious disease: Tuberculosis

[TB], Bacterial Infection, or Fungal Infection) (Table 1). Nonetheless, *Ochrobactrum*, *Delftia*, and *Streptococcus* genera could possibly promote the growth of other organisms responsible for granuloma formation, and are implicated in IGM aetiology. Other significantly different genera and species between paired IGM patient samples have median relative abundances below 1% for both IGM pus and skin samples; as such, they would not serve as meaningful targets for diagnostic, therapeutic, or molecular investigations.

IGM patients self-reported treatment with Cephalexin, Amoxicillin-clavulanate, Clindamycin, Trimethoprim-sulfamethoxazole, and/or Erythromycin largely targets *Streptococcus* and *Staphylococcus* infections, with Erythromycin targeting gram-positive bacteria that are often implicated in granulomatous inflammations[57-61]. With further information gleamed from metagenomic analysis on the genera and their abundances populating the IGM pus samples, they can serve to guide targeted antibiotic treatment either. The polymicrobial profiles will probably require combination or broad-spectrum antibiotic treatment. However, without definitively identifying a microbial infectious cause, further antibiotic treatment may not be the solution either.

To the best of our knowledge, this is the first study for IGM patients that provides the duality of intra-patient controls with paired skin and pus samples, on top of the comparison with LM patients. Furthermore, the use of 16S sequencing is less common for IGM studies, beyond bacteria culture methods, so as to profile both living and dead microbial populations. Metagenomic analysis matching patient samples to previous antibiotic treatment would also identify antibiotic-resistant strains. However, as a rare condition, the number of IGM patients recruited is limited. The limited LM control patients, as well, means comparisons and conclusions made on the basis of imbalances between IGM and LM patient numbers in our study must be validated in a larger cohort. Future skin sample collections must control for washing and exposure (cloth covering) before sample collection. It must definitely be acknowledged that the microbial load on skin is extremely low, so appropriate sampling and relevant negative controls are extremely crucial.

CONCLUSIONS

Microbial profiles are unique between IGM and LM patients. The inter-patient variability and polymicrobial microbiome of IGM pus samples cannot implicate specific genus or species of bacteria in an infectious cause for IGM. Subsequent investigations with larger patient numbers, backed by molecular studies, are needed to confirm or discard an infectious aetiology for IGM.

REFERENCES

1. Bellavia M, Damiano G, Palumbo VD, et al. Granulomatous Mastitis during Chronic Antidepressant Therapy: Is It Possible a Conservative Therapeutic Approach? *J Breast Cancer*. Sep 2012;15(3):371-2. doi:10.4048/jbc.2012.15.3.371

2. Kessler E, Wolloch Y. Granulomatous mastitis: a lesion clinically simulating carcinoma. *Am J Clin Pathol*. Dec 1972;58(6):642-6. doi:10.1093/ajcp/58.6.642

3. Tekgöz E, Çolak S, Çinar M, Yilmaz S. Treatment of idiopathic granulomatous mastitis and factors related with disease recurrence. *Turk J Med Sci*. Aug 26 2020;50(5):1380-1386. doi:10.3906/sag-2003-93

4. Steuer AB, Stern MJ, Cobos G, et al. Clinical Characteristics and Medical Management of Idiopathic Granulomatous Mastitis. *JAMA Dermatol*. Apr 1 2020;156(4):460-464. doi:10.1001/jamadermatol.2019.4516

5. Mathew Thomas V, Alexander SA, Bindal P, Vredenburgh J. Idiopathic Granulomatous Mastitis-A Mystery Yet to be Unraveled: A Case Series and Review of Literature. *Cureus*. Feb 5 2020;12(2):e6895. doi:10.7759/cureus.6895

6. Pluguez-Turull CW, Nanyes JE, Quintero CJ, et al. Idiopathic Granulomatous Mastitis: Manifestations at Multimodality Imaging and Pitfalls. *Radiographics*. Mar-Apr 2018;38(2):330-356. doi:10.1148/rg.2018170095

7. Altintoprak F, Kivilcim T, Ozkan OV. Aetiology of idiopathic granulomatous mastitis. *World J Clin Cases*. Dec 16 2014;2(12):852-8. doi:10.12998/wjcc.v2.i12.852

8. Taylor GB, Paviour SD, Musaad S, Jones WO, Holland DJ. A clinicopathological review of 34 cases of inflammatory breast disease showing an association between corynebacteria infection and granulomatous mastitis. *Pathology*. Apr 2003;35(2):109-19.

9. Paviour S, Musaad S, Roberts S, et al. Corynebacterium species isolated from patients with mastitis. *Clin Infect Dis*. Dec 1 2002;35(11):1434-40. doi:10.1086/344463

10. Ang LM, Brown H. Corynebacterium accolens isolated from breast abscess: possible association with granulomatous mastitis. *J Clin Microbiol*. May 2007;45(5):1666-8. doi:10.1128/JCM.02160-06

11. Yu HJ, Deng H, Ma J, et al. Clinical metagenomic analysis of bacterial communities in breast abscesses of granulomatous mastitis. *Int J Infect Dis*. Dec 2016;53:30-33. doi:10.1016/j.ijid.2016.10.015

12. Omranipour R, Vasigh M. Mastitis, Breast Abscess, and Granulomatous Mastitis. In: Alipour S, Omranipour R, eds. *Diseases of the Breast during Pregnancy and Lactation*. Springer International Publishing; 2020:53-61.

13. Kvist LJ. Toward a clarification of the concept of mastitis as used in empirical studies of breast inflammation during lactation. *J Hum Lact*. Feb 2010;26(1):53-9. doi:10.1177/0890334409349806

14. Scott DM. Inflammatory diseases of the breast. *Best Pract Res Clin Obstet Gynaecol*. Dec 8 2021;doi:10.1016/j.bpobgyn.2021.11.013

15. Amir LH, Academy of Breastfeeding Medicine Protocol C. ABM clinical protocol #4: Mastitis, revised March 2014. *Breastfeed Med*. Jun 2014;9(5):239-43. doi:10.1089/bfm.2014.9984

16. Angelopoulou A, Field D, Ryan CA, Stanton C, Hill C, Ross RP. The microbiology and treatment of human mastitis. *Med Microbiol Immunol*. Apr 2018;207(2):83-94. doi:10.1007/s00430-017-0532-z

17. Jimenez E, de Andres J, Manrique M, et al. Metagenomic Analysis of Milk of Healthy and Mastitis-Suffering Women. *J Hum Lact*. Aug 2015;31(3):406-15. doi:10.1177/0890334415585078

18. Patel SH, Vaidya YH, Patel RJ, Pandit RJ, Joshi CG, Kunjadiya AP. Culture independent assessment of human milk microbial community in lactational mastitis. *Sci Rep*. Aug 10 2017;7(1):7804. doi:10.1038/s41598-017-08451-7

19. Marin M, Arroyo R, Espinosa-Martos I, Fernandez L, Rodriguez JM. Identification of Emerging Human Mastitis Pathogens by MALDI-TOF and Assessment of Their Antibiotic Resistance Patterns. *Front Microbiol*. 2017;8:1258. doi:10.3389/fmicb.2017.01258

20. Mediano P, Fernandez L, Jimenez E, et al. Microbial Diversity in Milk of Women With Mastitis: Potential Role of Coagulase-Negative Staphylococci, Viridans Group Streptococci, and Corynebacteria. *J Hum Lact*. May 2017;33(2):309-318. doi:10.1177/0890334417692968

21. Shojaee L, Rahmani N, Moradi S, Motamedi A, Godazandeh G. Idiopathic granulomatous mastitis: challenges of treatment in iranian women. *BMC Surg*. Apr 21 2021;21(1):206. doi:10.1186/s12893-021-01210-6

22. Uysal E, Soran A, Sezgin E, Granulomatous Mastitis Study G. Factors related to recurrence of idiopathic granulomatous mastitis: what do we learn from a multicentre study? *ANZ J Surg*. Jun 2018;88(6):635-639. doi:10.1111/ans.14115

23. Azizi A, Prasath V, Canner J, et al. Idiopathic granulomatous mastitis: Management and predictors of recurrence in 474 patients. *Breast J*. Jul 2020;26(7):1358-1362. doi:10.1111/tbj.13822

24. Basim P, Argun D, Argun F. Risk Factors for Idiopathic Granulomatous Mastitis Recurrence after Patient-Tailored Treatment: Do We Need an Escalating Treatment Algorithm? *Breast Care (Basel)*. Apr 2022;17(2):172-179. doi:10.1159/000517399

25. Ong SH, Kukkillaya VU, Wilm A, et al. Species identification and profiling of complex microbial communities using shotgun Illumina sequencing of 16S rRNA amplicon sequences. *PLoS One*. 2013;8(4):e60811. doi:10.1371/journal.pone.0060811

26. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol*. 2011;12(5):R44. doi:10.1186/gb-2011-12-5-r44

27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. Oct 5 1990;215(3):403-10. doi:10.1016/S0022-2836(05)80360-2

28. Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol*. Nov 2021;17(11):e1009442. doi:10.1371/journal.pcbi.1009442

29. Bukin YS, Galachyants YP, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci Data*. Feb 5 2019;6:190007. doi:10.1038/sdata.2019.7

30. Meisel JS, Hannigan GD, Tyldsley AS, et al. Skin Microbiome Surveys Are Strongly Influenced by Experimental Design. *J Invest Dermatol*. May 2016;136(5):947-956. doi:10.1016/j.jid.2016.01.016

31. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods*. May 2007;69(2):330-9. doi:10.1016/j.mimet.2007.02.005

32. Martinez-Porchas M, Villalpando-Canchola E, Vargas-Albores F. Significant loss of sensitivity and specificity in the taxonomic classification occurs when short 16S rRNA gene sequences are used. *Heliyon*. Sep 2016;2(9):e00170. doi:10.1016/j.heliyon.2016.e00170

33. Blackmon MM, Nguyen H, Mukherji P. Acute Mastitis. *StatPearls*. 2022.

34. Fetherston C. Risk factors for lactation mastitis. *J Hum Lact*. Jun 1998;14(2):101-9. doi:10.1177/089033449801400209

35. Foxman B, D'Arcy H, Gillespie B, Bobo JK, Schwartz K. Lactation mastitis: occurrence and medical management among 946 breastfeeding women in the United States. *Am J Epidemiol*. Jan 15 2002;155(2):103-14. doi:10.1093/aje/155.2.103

36. Bouton ME, Winton LM, Gandhi SG, et al. Temporal resolution of idiopathic granulomatous mastitis with resumption of bromocriptine therapy for prolactinoma. *Int J Surg Case Rep*. 2015;10:8-11. doi:10.1016/j.ijscr.2015.02.051

37. Funke G, von Graevenitz A, Clarridge JE, 3rd, Bernard KA. Clinical microbiology of coryneform bacteria. *Clin Microbiol Rev*. Jan 1997;10(1):125-59. doi:10.1128/CMR.10.1.125

38. Dobinson HC, Anderson TP, Chambers ST, Doogue MP, Seaward L, Werno AM. Antimicrobial Treatment Options for Granulomatous Mastitis Caused by Corynebacterium Species. *J Clin Microbiol*. Sep 2015;53(9):2895-9. doi:10.1128/JCM.00760-15

39. Co M, Cheng VCC, Wei J, et al. Idiopathic granulomatous mastitis: a 10-year study from a multicentre clinical database. *Pathology*. Dec 2018;50(7):742-747. doi:10.1016/j.pathol.2018.08.010

40. Wolfrum A, Kummel S, Theuerkauf I, Pelz E, Reinisch M. Granulomatous Mastitis: A Therapeutic and Diagnostic Challenge. *Breast Care (Basel)*. Dec 2018;13(6):413-418. doi:10.1159/000495146

41. Wong SCY, Poon RWS, Chen JHK, et al. Corynebacterium kroppenstedtii Is an Emerging Cause of Mastitis Especially in Patients With Psychiatric Illness on Antipsychotic Medication. *Open Forum Infect Dis*. Spring 2017;4(2):ofx096. doi:10.1093/ofid/ofx096

42. Chalmers R, McClellan P, Silva V, Shutt N, Restini C. Red flags for the differential diagnosis of granulomatous mastitis: a case report. *J Med Case Rep*. Nov 10 2020;14(1):215. doi:10.1186/s13256-020-02563-x

43. Hartmann JE, Albrich WC, Dmitrijeva M, Kahlert CR. The Effects of Corticosteroids on the Respiratory Microbiome: A Systematic Review. *Front Med (Lausanne)*. 2021;8:588584. doi:10.3389/fmed.2021.588584

44. Huang EY, Inoue T, Leone VA, et al. Using corticosteroids to reshape the gut microbiome: implications for inflammatory bowel diseases. *Inflamm Bowel Dis*. May 2015;21(5):963-72. doi:10.1097/MIB.0000000000000332

45. Vich Vila A, Collij V, Sanna S, et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. *Nat Commun*. Jan 17 2020;11(1):362. doi:10.1038/s41467-019-14177-z

46. Graham ME, Herbert WG, Song SD, et al. Gut and vaginal microbiomes on steroids: implications for women's health. *Trends Endocrinol Metab*. Aug 2021;32(8):554-565. doi:10.1016/j.tem.2021.04.014

47. Laforest S, Pelletier M, Denver N, et al. Estrogens and Glucocorticoids in Mammary Adipose Tissue: Relationships with Body Mass Index and Breast Cancer Features. *J Clin Endocrinol Metab*. Apr 1 2020;105(4)doi:10.1210/clinem/dgz268

48. Wilson M, Mello MJ, Gruppuso PA. Antibiotics and the Human Microbiome: A Survey of Prescribing Clinicians' Knowledge and Opinions Regarding the Link between Antibiotic-Induced Dysbiosis and Immune-Mediated Disease. *R I Med J (2013)*. Sep 1 2021;104(7):59-63.

49. Ramirez J, Guarner F, Bustos Fernandez L, Maruy A, Sdepanian VL, Cohen H. Antibiotics as Major Disruptors of Gut Microbiota. *Front Cell Infect Microbiol*. 2020;10:572912. doi:10.3389/fcimb.2020.572912

50. Cani PD, Osto M, Geurts L, Everard A. Involvement of gut microbiota in the development of low-grade inflammation and type 2 diabetes associated with obesity. *Gut Microbes*. Jul-Aug 2012;3(4):279-88. doi:10.4161/gmic.19625

51. Nel Van Zyl K, Matukane SR, Hamman BL, Whitelaw AC, Newton-Foot M. Effect of antibiotics on the human microbiome: a systematic review. *Int J Antimicrob Agents*. Feb 2022;59(2):106502. doi:10.1016/j.ijantimicag.2021.106502

52. Vangoitsenhoven R, Cresci GAM. Role of Microbiome and Antibiotics in Autoimmune Diseases. *Nutr Clin Pract*. Jun 2020;35(3):406-416. doi:10.1002/ncp.10489

53. Wisplinghoff H. 181 - Pseudomonas spp., Acinetobacter spp. and Miscellaneous Gram-Negative Bacilli. In: Cohen J, Powderly WG, Opal SM, eds. *Infectious Diseases (Fourth Edition)*. Elsevier; 2017:1579-1599.e2.

54. Chiller K, Selkin BA, Murakawa GJ. Skin microflora and bacterial infections of the skin. *J Investig Dermatol Symp Proc*. Dec 2001;6(3):170-4. doi:10.1046/j.0022-202x.2001.00043.x

55. Davis CP. Normal Flora. In: th, Baron S, eds. *Medical Microbiology*. 1996.

56. Ryan MP, Pembroke JT. The Genus Ochrobactrum as Major Opportunistic Pathogens. *Microorganisms*. Nov 16 2020;8(11)doi:10.3390/microorganisms8111797

57. Herman TF, Hashmi MF. Cephalexin. *StatPearls*. 2022.

58. Smieja M. Current indications for the use of clindamycin: A critical review. *Can J Infect Dis*. Jan 1998;9(1):22-8. doi:10.1155/1998/538090

59. Kester M, Karpa KD, Vrana KE. 4 - Treatment of Infectious Diseases. In: Kester M, Karpa KD, Vrana KE, eds. *Elsevier's Integrated Review Pharmacology (Second Edition) (Second Edition)*. W.B. Saunders; 2012:41-78.

60. Waller DG, Sampson AP. 51 - Chemotherapy of infections. In: Waller DG, Sampson AP, eds. *Medical Pharmacology and Therapeutics (Fifth Edition)*. Elsevier; 2018:581-629.

61. GlaxoSmithKline Inc. Product Monograph, including Patient Medication Information. Clavulin Amoxicillin / clavulanate potassium tablets
amoxicillin / clavulanate potassium for oral suspension. Mississauga, Ontario, Canada: GlaxoSmithKline Inc.; 2020.

DECLARATIONS

Acknowledgements

We would like to thank the study participants and all research coordinators for their excellent help with recruitment, data and sample collection. We would also like to thank Balamurugan Periaswamy and Paola Florez De Sessions for their kindness and guidance.

Authors' contributions

Conceptualisation: J.L., M.H.; Patient recruitment and data acquisition: K.-T.B.T., Q.T.T., E.Y.T., S.-M.T., T.C.P., S.H.L., E.L.S.T., M.H.; Formal analysis: S.S.O., J.X., C.K.S., P.J.H., P.K.W.K., A.R., N.N., N.K., J.L.; Funding acquisition: J.L., M.H.; Investigation and methodology: S.S.O., J.X., C.K.S., A.J.K., P.J.H., P.K.W.K., A.R., K.-T.B.T., Q.T.T., E.Y.T., S.-M.T., T.C.P., S.H.L., E.L.S.T., N.N., N.K., J.L., M.H.; Supervision: J.L., M.H.; Visualisation: S.S.O., J.X., C.K.S., N.K., J.L.; Writing-original draft: S.S.O., J.L.; Writing-review & editing: S.S.O., A.R., K.-T.B.T., Q.T.T., J.L. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Ethics declaration

All studies were performed in accordance with the Declaration of Helsinki. All participants provided written informed consent. Data was de-identified. This study was approved by the National Healthcare Group Domain Specific Review Board (reference number: 2017/01057) and the Agency for Science, Technology and Research Institutional Review Board (reference number: 2020-152).

Funding

This study was supported by National University Health System (NUHS) Seed Fund (NUHSRO/2020/027/T1/Seed-Aug/11, awarded to M Hartman), and the Agency for Science, Technology and Research (A*STAR). Funding bodies were not involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

TABLE TITLES AND FOOTNOTES

Table 1. Description of analytical cohort of idiopathic granulomatous mastitis (IGM) and lactational mastitis (LM) patients.

Footnotes: IGM: Idiopathic granulomatous mastitis; LM: lactational mastitis; IQR: Interquartile range; KKH: KK Women's and Children's Hospital; NUH: National University Hospital; SGH: Singapore General Hospital; NCCS: National Cancer Centre Singapore; TTSH: Tan Tock Seng Hospital

[1]Comparison between IGM and LM patients using Kruskal-Wallis test for continuous variables, and Fisher's exact test for categorical variables.

[2]Chronic illnesses: Heart attack, stroke, or high blood pressure.

[3]Autoimmune conditions: Coeliac disease, Type 1 diabetes mellitus, Graves disease, Inflammatory bowel disease, Multiple sclerosis, Psoriasis, Rheumatoid arthritis, or Lupus erythematosus.

[4]Infectious disease: tuberculosis, bacterial Infection, or fungal Infection.

[5]Known antibiotics treatment: Cephalexin, Amoxicillin-clavulanate, Clindamycin, Trimethoprim-sulfamethoxazole, and/or Erythromycin.

[6]Known steroid treatment: Corticosteroids, and/or Methotrexate.

Table 2. Statistically significant (after adjustments and correcting for multiple comparisons) genera and species identified from general linear models for determining multivariable association between sample type, covariates and microbial metagenomic features in paired pus and skin samples from idiopathic granulomatous mastitis (IGM) patients.

Footnotes: [1]Adjusted for antibiotic treatment (received and did not receive antibiotics treatment), and duration after antibiotics treatment (samples collected less than 2 weeks after treatment vs more than 2 weeks after treatment).

[2]β: General linear model beta-estimates; SD: Standard deviation. Beta-estimates and standard deviations presented are for pus samples as the reference group against which paired skin samples were compared.

[3]q-value represents the adjusted p-value for multiple comparisons using Benjamini & Hochberg correction (1995). ***: q-value <0.001; **: $0.001 \leq$ q-value < 0.01; *: $0.01 \leq$ q-value < 0.05.

FIGURES TITLES AND LEGENDS

Figure 1. Bar graph of genus taxonomic profiles of breast **(a)** pus and **(b)** skin samples.

A metagenomic analysis compared the relative abundances of bacterial taxa of pus samples from 21 patients with idiopathic granulomatous mastitis (IGM) and three controls with lactating mastitis (LM) (LM01, LM02, and LM03). Each bar represents a **(a)** pus or **(b)** skin sample from a patient. The genus level taxonomic units with relative abundance ≥1% are represented, with corresponding colours indicated in the shared figure key for **(a)** and **(b)** on the right. Genus taxonomic units with relative abundance <1% are collapsed into "Low Abundance", and taxonomic units higher than genus level classification are collapsed into "Others".

(a) Genera *Staphylococcus* dominates the three LM pus samples (LM1-3): 90.8%, 26.5%, and 38.5% respectively. Bacteria profiles in IGM pus samples differ from LM, without a dominant genus across pus samples, and greater variability between patients.

(b) No dominant genus is observed across either the IGM or LM skin samples, collectively or as separate groups. *Corynebacterium* contributes more to skin than pus microbiota in both IGM and LM samples, although *Corynebacterium* appears at higher relative abundances for IGM skin samples than LM skin samples. Again, inter-patient variability is considerable.

IGM: Idiopathic granulomatous mastitis; LM: lactational mastitis.

Figure 2. Multi-dimensional scaling (MDS) (principal coordinate analysis (PCoA)) of the Bray-Curtis distance between breast pus and skin samples from **(a)** idiopathic granulomatous mastitis (IGM) patients and control patients with lactating mastitis (LM), and **(b)** IGM patients alone.

Each point represents a pus or skin sample from a patient. The 21 IGM patients are represented as dots and the 3 LM patients are represented as triangles, with corresponding colours indicating sample type shown in the shared figure key for **(a)** and **(b)** at the bottom.

(a) First two MDS of Bray-Curtis distance between breast pus and skin samples for IGM and LM patients only explains 14.97% and 13.11% of the variability respectively, and collectively explains 28.08% of the samples' variability. Despite no obvious clustering, the different patient and sample types affected genera diversity significantly ($p < 0.001$, PERMANOVA).

(b) First two MDS of Bray-Curtis distance between breast pus and skin samples for IGM patients only explains 16.98% and 12.23% of the variability respectively, and collectively explains 29.21% of the

samples' variability. Despite no obvious clustering, the different patient and sample types affected genera diversity significantly ($p < 0.001$, PERMANOVA).

IGM: Idiopathic granulomatous mastitis; LM: lactational mastitis; MDS: Multi-dimensional scaling; p: p-value; PERMANOVA: Permutational Multivariate Analysis of Variance.

Figure 3. Relative abundance of *Corynebacterium* genus in breast **(a)** pus and **(b)** skin samples from idiopathic granulomatous mastitis (IGM) patients and control patients with lactating mastitis (LM). The left-end bars in both panels represent the 21 IGM patients, and the right-end bars in both panels represent the 3 LM patients, with corresponding colours indicating mastitis type shown in the bottom-left shared figure key for **(a)** and **(b)**. The percentage of *Corynebacterium* genus relative abundance is indicated above each bar, presented to 1 decimal place. Duration of antibiotic treatment prior to sample collection are presented as symbols above each bar, as four categories: Less than 2 weeks before sample collection, more than 2 weeks before sample collection, missing duration, and no antibiotic treatment. The corresponding symbols are also indicated in the bottom-left shared figure key for **(a)** and **(b)**.

(a) For the left panel of **pus** samples, the patients are arranged in decreasing relative abundance of *Corynebacterium* genus in **pus** samples within mastitis type.

(b) For the right panel of **skin** samples, the patients are arranged following patient order in **(a)** i.e. decreasing relative abundance of *Corynebacterium* genus in **pus** samples within mastitis type.

Distribution of relative abundance of *Corynebacterium* genus in breast pus and skin samples from 21 IGM patients is also displayed in **(a)**. Median *Corynebacterium* relative abundance in IGM pus samples is 9% (interquartile range = 3.8-13.0%), compared to 9.4% (interquartile range = 5.8-30.2%) in IGM skin samples. Paired Wilcoxon sign ranked test found no significant difference in *Corynebacterium* relative abundance between paired IGM skin and pus samples ($p = 0.29$).

IGM: Idiopathic granulomatous mastitis; LM: lactational mastitis; p: p-value; Wilcoxon: Wilcoxon paired sign ranked test.

Table 1. Description of analytical cohort of idiopathic granulomatous mastitis (IGM) and lactational mastitis (LM) patients.

	Total n=24	**IGM n=21**	**LM n=3**	**p-value[1]**
Demographics				
Median age at diagnosis (years, IQR)	34 (30.50 – 40.25)	34 (31.00 – 41.00)	31 (28.00 – 34.00)	0.358
Recruitment Site (n, %)				
KKH	13 (54.17)	10 (47.62)	3 (100.00)	0.422
SGH	9 (37.50)	9 (42.86)	0 (0)	
NUH	2 (8.33)	2 (9.52)	0 (0)	
Year of diagnosis (n, %)				
2014-2017	6 (25.00)	6 (28.57)	0 (0)	0.546
2018-2019	18 (75.00)	15 (71.43)	3 (100.00)	
Ethnicity (n, %)				
Chinese	15 (62.50)	14 (66.67)	1 (33.33)	0.308
Malay	4 (16.67)	3 (14.29)	1 (33.33)	
Indian	2 (8.33)	2 (9.52)	0 (0)	
Other	3 (12.50)	2 (9.52)	1 (33.33)	
Body mass index (kg/m^2, IQR)	26.29 (22.34 – 30.83)	26.29 (24.61 – 31.22)	22.19 (22.03 – 23.97)	0.106
Education level (n, %)				
Primary	2 (8.33)	2 (9.52)	0 (0)	0.727
Secondary	5 (20.83)	4 (19.05)	1 (33.33)	
Pre-University	8 (33.33)	6 (28.57)	2 (66.67)	
Undergraduate	6 (25.00)	6 (28.57)	0 (0)	
Graduate	3 (12.50)	3 (14.29)	0 (0)	
Patient characteristics				

Smoking (n, %)				
Yes	5 (20.83)	5 (23.81)	0 (0)	1
No	19 (79.17)	16 (76.19)	3 (100.00)	
Chronic Illness[2] Diagnosis (n, %)				
Yes	4 (16.67)	4 (19.05)	0 (0)	1
No	20 (83.33)	17 (80.95)	3 (100.00)	
Previously Pregnant (n, %)				
Yes	21 (87.50)	18 (85.71)	3 (100.00)	1
No	3 (12.50)	3 (14.29)	0 (0)	
Number of children (n, %)				
0	3 (12.50)	3 (14.29)	0 (0)	0.769
1	13 (54.17)	10 (47.62)	3 (100.00)	
2	6 (25.00)	6 (28.57)	0 (0)	
3	1 (4.17)	1 (4.76)	0 (0)	
4	1 (4.17)	1 (4.76)	0 (0)	
Time since last childbirth (n, %)				
No children	3 (12.50)	3 (14.29)	0 (0)	0.091
Less than 2 years	8 (33.33)	5 (23.81)	3 (100.00)	
Between 3 and 5 years	10 (41.67)	10 (47.62)	0 (0)	
More than 5 years	3 (12.50)	3 (14.29)	0 (0)	
Autoimmune Conditions[3] (n, %)				
No	24 (100.00)	21 (100.00)	3 (100.00)	
Previous Infectious Disease[4] Diagnosis (n, %)				
No	24 (100.00)	21 (100.00)	3 (100.00)	
Previous Cancer Diagnosis (n, %)				

No	24 (100.00)	21 (100.00)	3 (100.00)	
Treatment for mastitis				
Any treatment (n, %)				1
Yes	21 (87.50)	18 (85.71)	3 (100.00)	
No	3 (12.50)	3 (14.29)	0 (0)	
Type of treatment (n, %)				0.204
Antibiotic treatment[5] only	9 (37.50)	6 (28.57)	3 (100.00)	
Antibiotic and steroid treatment[6] only	8 (33.33)	8 (38.10)	0 (0)	
Other type of treatment	4 (16.67)	4 (19.05)	0 (0)	
Did not receive any treatment	3 (12.50)	3 (14.29)	0 (0)	
Duration between antibiotic treatment and sample collection (n, %)				
Less than 2 weeks	7 (29.17)	6 (28.57)	1 (33.33)	0.782
More than 2 weeks	9 (37.50)	7 (33.33)	2 (66.67)	
Duration missing	1 (4.17)	1 (4.76)	0 (0)	
Did not receive antibiotic treatment	7 (29.17)	7 (33.33)	0 (0)	

IGM: Idiopathic granulomatous mastitis; LM: lactational mastitis; IQR: Interquartile range; KKH: KK Women's and Children's Hospital; NUH: National University Hospital; SGH: Singapore General Hospital; NCCS: National Cancer Centre Singapore; TTSH: Tan Tock Seng Hospital

[1]Comparison between IGM and LM patients using Kruskal-Wallis test for continuous variables, and Fisher's exact test for categorical variables.

[2]Chronic illnesses: Heart attack, stroke, or high blood pressure.

[3]Autoimmune conditions: Coeliac disease, Type 1 diabetes mellitus, Graves disease, Inflammatory bowel disease, Multiple sclerosis, Psoriasis, Rheumatoid arthritis, or Lupus erythematosus.

[4]Infectious disease: tuberculosis, bacterial Infection, or fungal Infection.

[5]Known antibiotics treatment: Cephalexin, Amoxicillin-clavulanate, Clindamycin, Trimethoprim-sulfamethoxazole, and/or Erythromycin.

[6]Known steroid treatment: Corticosteroids, and/or Methotrexate.

Table 2. Statistically significant (after adjustments and correcting for multiple comparisons) genera and species identified from general linear models for determining multivariable association between sample type, covariates and microbial metagenomic features in paired pus and skin samples from idiopathic granulomatous mastitis (IGM) patients.

	n (n, not 0)	Crude				[1]Adjusted			
		β[2]	SD[2]	p-value	[3]q-value	β[2]	SD[2]	p-value	[3]q-value
Genus									
Ochrobactrum	42 (40)	-2.770	0.154	<0.001	<0.001	-2.770	0.156	<0.001	<0.001***
Delftia	42 (42)	-2.469	0.196	<0.001	<0.001	-2.469	0.196	<0.001	<0.001***
Anaerobacillus	42 (28)	-2.470	0.247	<0.001	<0.001	-2.470	0.247	<0.001	<0.001***
Gordonia	42 (42)	-1.431	0.159	<0.001	<0.001	-1.431	0.159	<0.001	<0.001***
Methylobacterium	42 (16)	1.361	0.206	<0.001	<0.001	1.361	0.206	<0.001	<0.001***
Fusobacterium	42 (39)	-1.577	0.249	<0.001	<0.001	-1.577	0.249	<0.001	<0.001***
Sphingobium	42 (15)	1.111	0.232	<0.001	0.003	1.111	0.225	<0.001	0.003**
Alkanindiges	42 (32)	-1.436	0.313	<0.001	0.006	-1.436	0.297	<0.001	0.004**
Streptococcus	42 (42)	-0.884	0.168	<0.001	0.002	-0.884	0.168	<0.001	0.005**
Achromobacter	42 (16)	0.832	0.167	<0.001	0.003	0.832	0.167	<0.001	0.009**
Capnocytophaga	42 (15)	1.131	0.249	<0.001	0.006	1.131	0.249	<0.001	0.022*
Mycobacterium	42 (22)	1.404	0.350	<0.001	0.007	1.404	0.344	<0.001	0.022*
Novosphingobium	42 (13)	0.627	0.142	<0.001	0.006	0.627	0.142	<0.001	0.024*
Peptoniphilus	42 (42)	-1.174	0.271	<0.001	0.008	-1.174	0.271	<0.001	0.025*
Rothia	42 (37)	-1.038	0.239	<0.001	0.008	-1.038	0.239	<0.001	0.025*
Finegoldia	42 (42)	-1.152	0.271	<0.001	0.008	-1.152	0.271	<0.001	0.027*
Burkholderia	42 (15)	0.699	0.178	<0.001	0.008	0.699	0.181	<0.001	0.028*
Roseomonas	42 (25)	1.328	0.340	<0.001	0.008	1.328	0.348	<0.001	0.031*
Anaerococcus	42 (42)	-0.808	0.221	0.002	0.027	-0.808	0.216	<0.001	0.036*
Agrobacterium	42 (22)	1.336	0.378	0.001	0.019	1.336	0.380	0.001	0.065
Hydrogenophaga	42 (11)	0.914	0.276	0.002	0.030	0.914	0.263	0.001	0.069

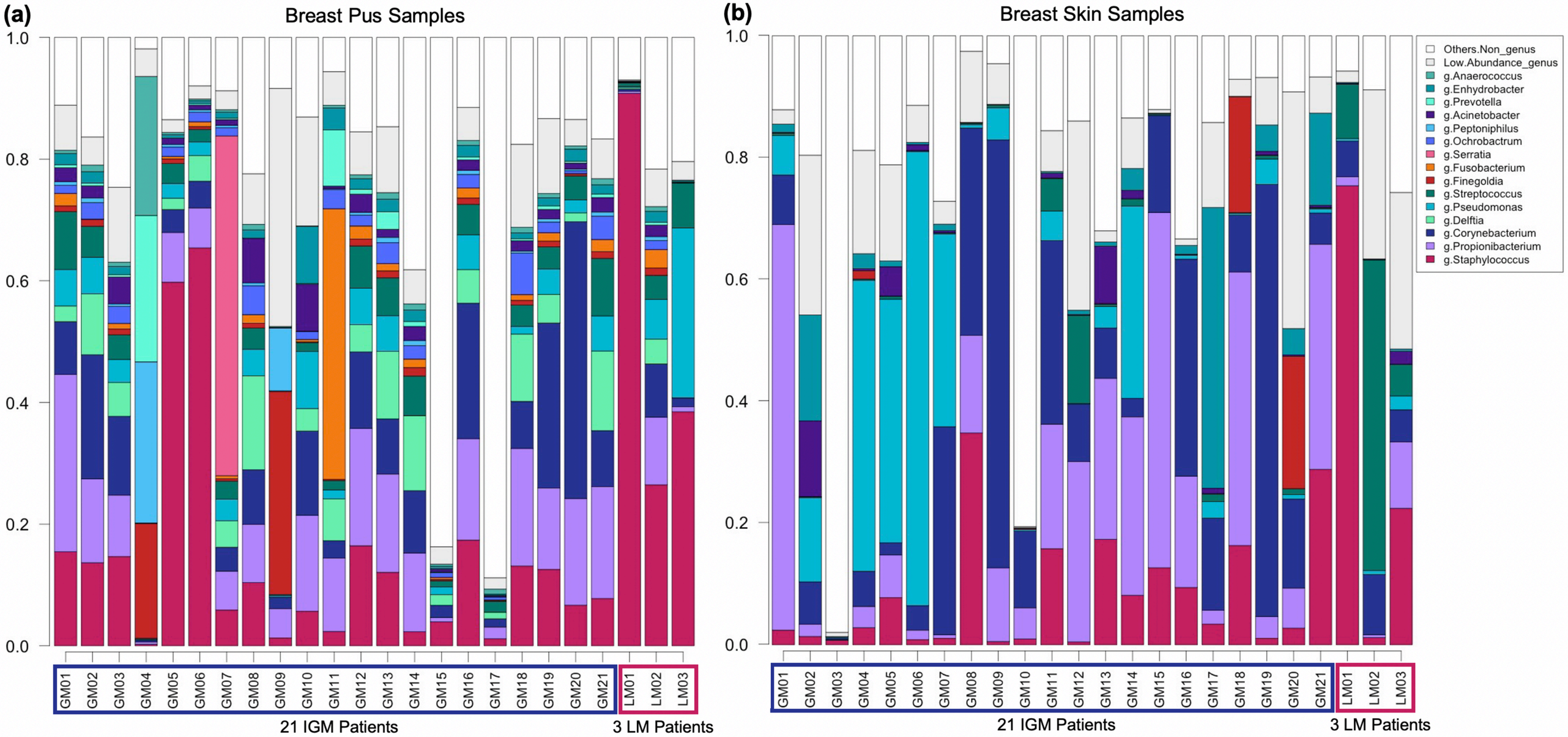
Peptostreptococcus	42 (9)	0.768	0.226	0.002	0.035	0.768	0.226	0.003	0.120
Kocuria	42 (18)	0.877	0.280	0.003	0.043	0.877	0.284	0.004	0.144
Dermabacter	42 (17)	0.996	0.318	0.003	0.043	0.996	0.324	0.004	0.145
Species									
Acinetobacter schindleri	42 (21)	-1.571	0.226	<0.001	<0.001	-1.571	0.226	<0.001	<0.001***
Rothia mucilaginosa	42 (30)	-1.472	0.285	<0.001	<0.001	-1.472	0.289	<0.001	0.002**
Lactobacillus iners	42 (13)	0.695	0.183	<0.001	0.021	0.695	0.183	<0.001	0.039*
Corynebacterium kroppenstedtii	42 (38)	-0.971	0.245	<0.001	0.019	-0.971	0.245	<0.001	0.053
Roseomonas mucosa	42 (24)	1.071	0.321	0.002	0.033	1.071	0.329	0.002	0.116
Kocuria rhizophila	42 (12)	0.780	0.246	0.003	0.045	0.780	0.251	0.004	0.162

[1]Adjusted for antibiotic treatment (received and did not receive antibiotics treatment), and duration after antibiotics treatment (samples collected less than 2 weeks after treatment vs more than 2 weeks after treatment).

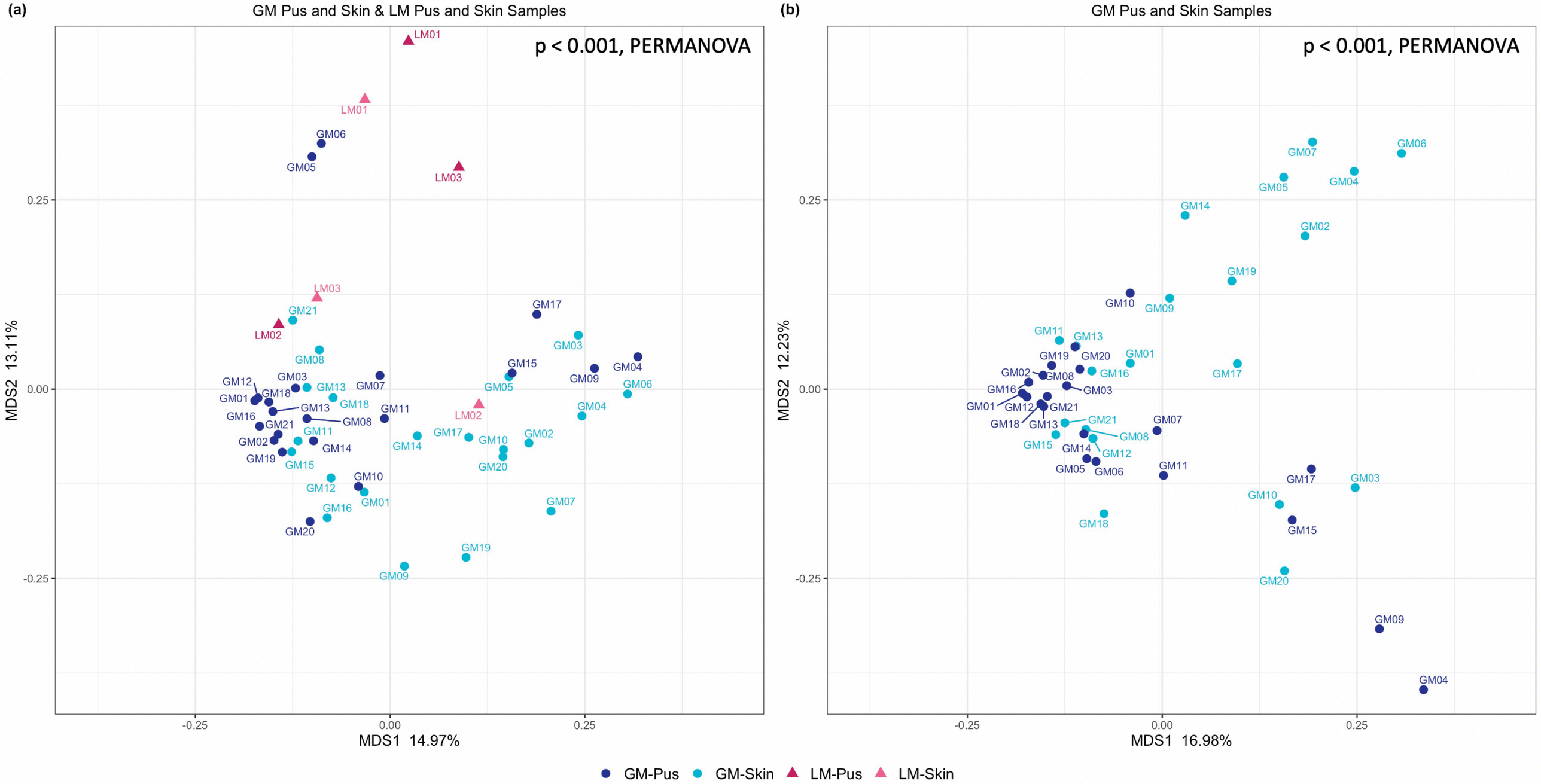
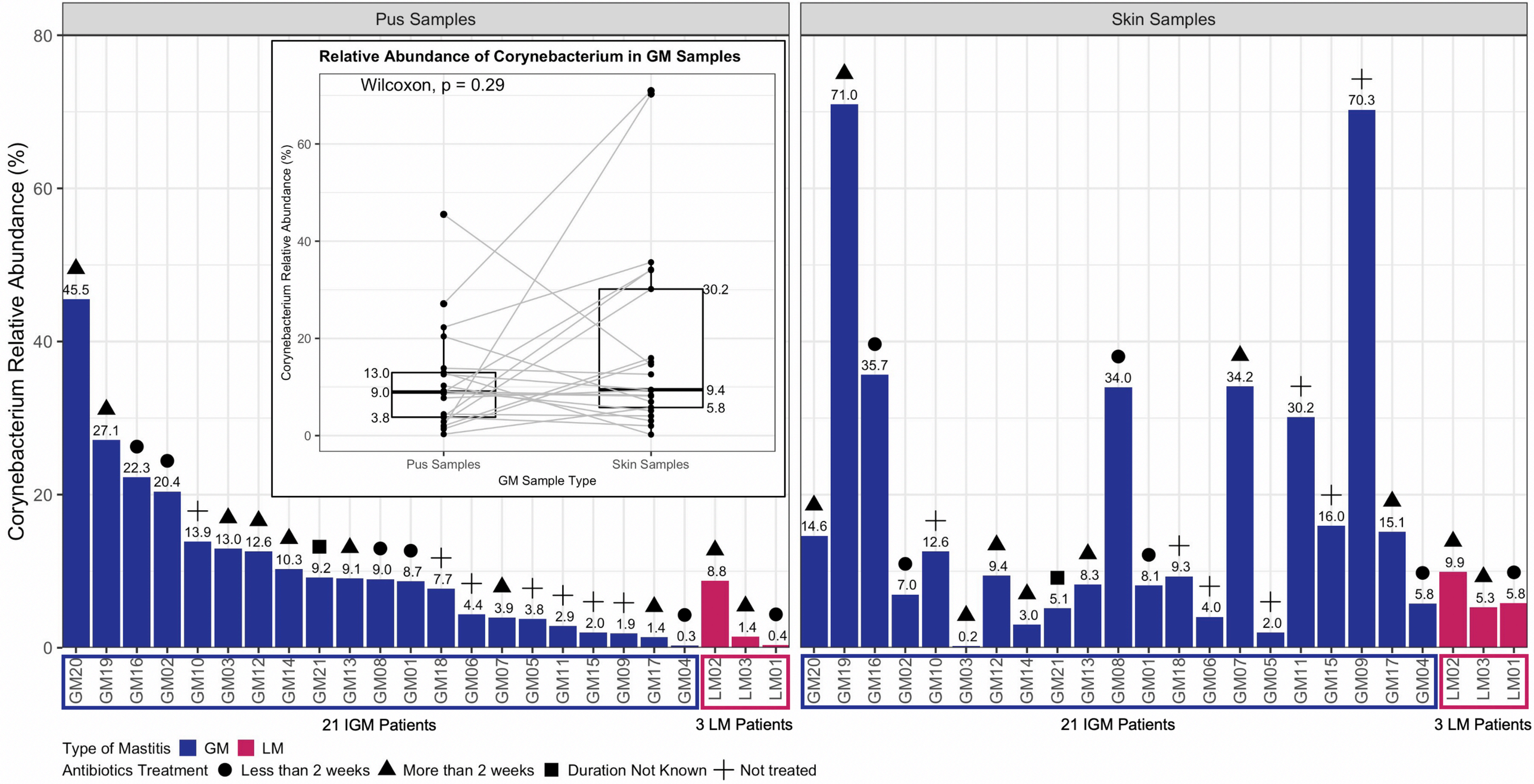
[2]β: General linear model beta-estimates; SD: Standard deviation. Beta-estimates and standard deviations presented are for pus samples as the reference group against which paired skin samples were compared.

[3]q-value represents the adjusted p-value for multiple comparisons using Benjamini & Hochberg correction (1995). ***: q-value <0.001; **: 0.001 ≤ q-value < 0.01; *: 0.01 ≤ q-value < 0.05.

16S rRNA Microbial Sequencing of Breast Pus and Skin Samples



Multi-dimensional Scaling of Bray-Curtis distance between Breast Pus and Skin Samples



Relative Abundance of Corynebacterium in Breast Pus and Skin Samples