Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Gut microbiome and its association with sociodemographic and clinical characteristics in Peruvian women with breast cancer: results of a multicenter study

**Mariana Callapiña De Paz**
Instituto de Investigación en Ciencias Biomédicas, Universidad Ricardo Palma
**Pool Marcos-Carbajal**
Instituto de Investigación en Ciencias Biomédicas, Universidad Ricardo Palma
**Raul Tito**
Rega Institute for Medical Research
**Jose M. Vela-Ruiz**
jose.vela@urp.edu.pe

Instituto de Investigación en Ciencias Biomédicas, Universidad Ricardo Palma
**Jhanina Campos Tineo**
Instituto de Investigación en Ciencias Biomédicas, Universidad Ricardo Palma
**Ariana Córdova Salazar**
Instituto de Investigación en Ciencias Biomédicas, Universidad Ricardo Palma
**Zaida Morante**
Instituto Nacional de Enfermedades Neoplásicas
**Marco Gálvez Villanueva**
Instituto Regional de Enfermedades Neoplásicas - IREN Norte
**Joan Moreno Lujan**
Instituto Regional de Enfermedades Neoplásicas - IREN Norte
**Andy Pantoja Lázaro**
Instituto Regional de Enfermedades Neoplásicas - IREN Norte
**Laura Escobar Caipo**
Instituto Regional de Enfermedades Neoplásicas - IREN Norte
**Gustavo Flores Trujillo**
Instituto Regional de Enfermedades Neoplásicas - IREN Norte
**Teresa Cusma Quintana**
Instituto Regional de Enfermedades Neoplásicas - IREN Norte
**Fernando Valencia**
Instituto Regional de Enfermedades Neoplásicas del Sur - IREN SUR

**Yomali Ferreyra**
Health Innovation Laboratory, Institute of Tropical Medicine “Alexander von Humboldt,” Universidad Peruana Cayetano Heredia
**Henry L Gomez**
Oncosalud, Auna
**Alonso Soto**
Instituto de Investigación en Ciencias Biomédicas, Universidad Ricardo Palma

---

**Article**

**Keywords:** Breast cancer, Gastrointestinal Microbiome, RNA, Ribosomal, 16S, molecular typing, Peru (MESH)

**Posted Date:** September 20th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10374419/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** No competing interests reported.

---

# Abstract

**Introduction:** Breast cancer is the most common cancer in women worldwide and a leading cause of cancer-related death. Emerging evidence suggests that the gut microbiota may influence biological processes involved in breast carcinogenesis. However, Latin American populations, with their high genetic, environmental, and dietary diversity, remain underrepresented in this field.

**Objective**: To characterize the composition and diversity of the gut microbiota in Peruvian women with breast cancer treated at three oncology institutions in different geographic regions of Perú, and to evaluate its variation according to geographic origin, clinical stage, and molecular subtype.

Materials and methods: A multicenter cross-sectional analytical observational study was conducted in 77 patients between 2023 and 2024. Clinical data and fecal samples were collected; the V3–V4 region of the 16S rRNA gene was sequenced and reads were resolved into amplicon sequence variants (ASVs). Alpha diversity (observed richness, Shannon index), beta diversity (Bray-Curtis dissimilarity and principal coordinates analysis [PCoA]), and relative abundance of taxa were assessed. Comparisons were performed using Kruskal-Wallis, Wilcoxon, and PERMANOVA tests. Results: The sample presented a common microbial core dominated by Faecalibacterium, *Segatella*, *Blautia*, *Vescimonas*and *Bacteroides*. Alpha diversity did not differ significantly among the three cities. Overall community structure did not differ significantly by region (PERMANOVA, Bray–Curtis $R^2 = 0.036$, $p = 0.109$), clinical stage ($p = 0.23$) or molecular subtype ($p = 0.20$); Pairwise comparisons suggested a nominal difference between Trujillo and Arequipa ($R^2 = 0.037$, $p = 0.019$), although this association did not remain significant after Firmicutes/Bacteroidota ratio correction (adjusted $p = 0.057$). The F/B ratio differed significantly by region (Kruskal–Wallis $p = 0.012$), being highest in Trujillo and lowest in Arequipa (adjusted $p = 0.02$). No bacterial genus remained significantly associated with any geographic or clinical variable after correction for multiple testing.

**Conclusions**: Peruvian women with breast cancer exhibit a relatively stable gut microbial community whose overall structure is not significantly associated with geographic origin, clinical stage, or molecular subtype. Nevertheless, regional differences were observed in the Firmicutes/Bacteroidota ratio, suggesting modest geographic variation in microbiota composition.


# Contributions to the literature

- This study provides one of the first multicenter characterizations of the gut microbiota in Peruvian women with breast cancer, contributing primary evidence from a Latin American population that is largely underrepresented in microbiome-oncology research.
- It identifies a shared gut microbial core across geographically distinct regions of Peru, with regional differences expressed as a gradient in the Firmicutes/Bacteroidota ratio rather than as distinct community structures.

- Neither clinical stage nor molecular subtype was associated with gut microbiota diversity, structure or genus-level abundance, suggesting that environmental and sociocultural factors may outweigh tumor-intrinsic characteristics in this population.
- All analyses are based on a fully reproducible, deeply sequenced ASV-based pipeline with verified clinical metadata and robustness checks across distance metrics and differential-abundance methods, establishing a reliable baseline for future longitudinal, metagenomic, and metabolomic studies in Peru.

# INTRODUCTION

Breast cancer is currently the most common neoplasm in women worldwide and one of the leading causes of cancer-related mortality in this group. In 2022, an estimated 2.3 million new cases and around 670,000 deaths were reported in women, solidifying its status as a priority public health problem globally (1). Recent projections suggest that, if current trends continue, the incidence will continue to rise and could reach up to 3.2 million new cases annually by 2050, with more than one million deaths per year (2). These scenarios reinforce the need for a better understanding of the biological and environmental factors that influence the onset and progression of this disease.

In the Americas, the situation follows a similar pattern. Breast cancer continues to be the most common neoplasm among women in Latin America and the Caribbean, with increasing rates in several countries and persistent inequalities in access to timely diagnosis and treatment. This trend has been attributed to multiple determinants, including population aging, the nutritional transition, rapid urbanization, and structural limitations of health systems in the region (3).

In Peru, this neoplasm ranks first in incidence among women, with 7,797 new cases and approximately 1,951 deaths reported in 2022, representing nearly 20% of all female cancers (4). Despite efforts in screening and treatment, breast cancer mortality continues to rise. Recent projections estimate that the age-adjusted mortality rate could reach 8.7 deaths per 100,000 women by 2027 (5), which reaffirm the need to identify clinical, biological, and environmental factors that influence its appearance and evolution, such as the gut microbiota profile.

The etiopathogenesis of breast cancer is multifactorial and complex, involving genetic, hormonal, environmental, and lifestyle factors. In recent years, the microbiome, particularly the gut microbiome, has emerged as a relevant component in tumor biology. The gut microbial community influences metabolic, immunological, and endocrine processes and has been implicated in carcinogenesis through mechanisms such as chronic inflammation, the production of bioactive metabolites, and the modulation of the host response to cancer therapies (6).

Several studies have shown differences in the diversity and composition of the gut microbiota between breast cancer patients and healthy populations, with potential implications for estrogen metabolism, immune polarization, and therapeutic efficacy. However, most research has been conducted in high-

income countries, with very little evidence from regions with distinct nutritional, genetic, and sociocultural characteristics, such as Latin America (7–9).

In this regard, exploring the gut microbiota from a geographical perspective can provide relevant information for understanding the factors that modulate its profile in cancer patients. This study aimed to characterize the composition of the gut microbiota in Peruvian women with breast cancer treated at three national referral cancer centers (Lima, Arequipa, and Trujillo) and to evaluate its variation according to geographical origin. Furthermore, the study explored the relationship with clinical stage and molecular classification in order to contextualize possible interactions between the gut microbiota and tumor characteristics.

# METHODOLOGY

## Study Design and Population

A multicenter observational study was conducted in women with a confirmed diagnosis of breast cancer treated at three leading cancer centers in Peru: Lima, Arequipa, and Trujillo. Patients over 18 years of age who had not received antibiotics or probiotics in the weeks prior to sample collection and who provided informed consent were included. Pregnant patients, patients with active infectious or inflammatory bowel diseases, and those who had received prior neoadjuvant therapy were excluded.

A non-probability convenience sampling method was used to ensure the operational feasibility of the study, given the exploratory nature of the gut microbiota analysis. The final sample consisted of 77 patients from three oncology institutions in Peru: the National Institute of Neoplastic Diseases (INEN) in Lima (central region, n = 19), the Northern Regional Institute of Neoplastic Diseases (IREN-Norte) in Trujillo (northern region, n = 45), and the Southern Regional Institute of Neoplastic Diseases (IREN-Sur) in Arequipa (southern region, n = 13). Before analysis, the clinical database and the sample-to-patient assignments were consolidated and independently verified against the primary records (Extended Methods, Supplementary Material).

## Definition of variables

The dependent variable was the gut microbial community profile (microbiota). The independent variables were age, body mass index (BMI; underweight < 18.5, normal 18.5–24.9, overweight 25.0–29.9, obesity ≥ 30.0), the hospital of origin (Lima, Arequipa, and Trujillo), the molecular classification of breast cancer, and the clinical stage (I–IV) determined according to the seventh edition of the American Joint Committee on Cancer staging system (10). Molecular subtype was recorded from the pathology reports as Luminal (A/B), Luminal B with HER2 overexpression, HER2-positive, or triple negative; the Luminal-HER2 hybrid category (n = 7) was kept separate in the primary analysis and pooled with HER2 in a sensitivity analysis.

## Sample collection and processing

Stool samples were collected from all participants and stored under standardized conditions until processing in a laboratory specializing in next-generation sequencing. Microbial DNA extraction was performed following validated protocols for gut microbiota, using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research) (11). DNA concentration was determined using fluorometry with a Qubit 4 Fluorometer (Invitrogen) (12). Aliquots were stored at − 80°C until analysis.

## Sequencing and bioinformatic analysis

The V3–V4 region of the 16S rRNA gene was amplified (Quick-16S Plus NGS Library Prep Kit, Zymo Research) and sequenced on an Illumina platform in two paired-end runs (2 × 300 bp, 62 patients; 2 × 250 bp, 15 patients) (13). Raw reads were processed in R (v4.6.0) with DADA2 (v1.40) (24). After primer removal (20 bases from each read), forward and reverse reads were truncated at 270/190 bases (2 × 300 bp run) or 220/190 bases (2 × 250 bp run) and quality-filtered (maxN = 0, maxEE = 5/7, truncQ = 2); filtering parameters were selected after a systematic comparison documented in the Extended Methods. Error models were learned per sequencing run, sample inference used pseudo-pooling, and paired reads were joined by concatenation because the amplicon length precludes reliable overlap at these read lengths. Amplicon sequence variants (ASVs) were screened for chimeras (consensus method) and classified with the naive Bayes classifier against the RDP 16S training set 19 (25). The two runs were merged at genus level. The final data set comprised a median of 154,843 (range 46,629–454,827) high-quality reads per patient.

# Statistical analysis

Alpha diversity (observed richness and Shannon index) was computed on counts rarefied to the minimum depth (46,629 reads) and compared with Kruskal–Wallis tests followed by Dunn or pairwise Wilcoxon post-hoc tests with Benjamini–Hochberg (BH) correction. Beta diversity was assessed on genus-level relative abundances with the Bray–Curtis and binary Jaccard distances, and on centered log-ratio transformed counts with the Euclidean (Aitchison) distance (26); ordination used principal coordinates analysis (PCoA). Global and pairwise differences were tested with PERMANOVA (adonis2, 999 permutations), with homogeneity of dispersion verified by PERMDISP; a marginal PERMANOVA model additionally adjusted for age and BMI. The Firmicutes/Bacteroidota ratio (log2) was compared across groups with Kruskal–Wallis and pairwise Wilcoxon (BH) tests. Genus-level differential abundance was analyzed with ANCOM-BC2 (including its built-in pseudo-count sensitivity analysis) (27), DESeq2 negative-binomial models (28), and rank-based Kruskal–Wallis tests on rarefied counts, considering BH-adjusted p-values < 0.05 significant. All analysis code is available from the corresponding author.

# Ethical considerations

The study had the prior approval of all the ethics committees of the institutions involved: INEN-Lima (INEN 23–52 Letter No. 059-2023-CRPI-DI-DCON/INEN), IREN-Trujillo (Official Letter No. 1221-2023-GRLL-GGR-IRSREN NORTE-DG-CIEI), IREN-Arequipa (Registration Code 062023, Letter No. 024-2023-GRA/GRS/GR–IREN-SUR/G-DCC-DI-CRPI), Ricardo Palma University (Committee Code PPG 01-2024), and San Martín de Porres University (Official Letter No. 0952-2024-CIEI-FMH-USMP). All participants provided written informed consent prior to enrollment. The study was conducted in accordance with internationally accepted ethical standards and the principles of the Declaration of Helsinki.

# RESULTS

The study included 77 women with breast cancer from Lima (n = 19), Trujillo (n = 45), and Arequipa (n = 13). The overall median age was 50 years (IQR 43–61); patients from Trujillo were older (median 53 years) than those from Lima (44) and Arequipa (49). Overweight and obesity together predominated in all three centers, with obesity most frequent in Arequipa (46.2%). The Luminal subtype was the most common in all regions (47–54%); Luminal tumors with HER2 overexpression accounted for 9.1% overall (n = 7), HER2-positive tumors for 16.9% (n = 13) and triple-negative tumors for 24.7% (n = 19). Most patients presented with stage III disease, particularly in Arequipa (69.2%); stage distribution was more balanced between stages II and III in Lima and Trujillo (Table 1).

Sequencing and bioinformatic processing yielded a median of 154,843 high-quality reads per patient (minimum 46,629), and 362 bacterial genera were detected. The gut microbiota showed a consistent core across the three cities (Fig. 1): *Faecalibacterium* was the dominant genus (mean relative abundance 14.9%), followed by *Segatella* (formerly Prevotella; 9.1%), *Blautia* (6.6%), *Vescimonas* (6.2%), Bacteroides (2.5%), *Phocaeicola* (2.4%) and *Agathobacter* (2.2%); 13.0% of reads could not be classified at genus level. Despite this shared core, relative-abundance differences among regions were apparent, most visibly a higher share of Lachnospiraceae genera (*Blautia*, *Dorea*) in Trujillo and a higher share of *Segatella* in Arequipa.

Alpha diversity was similar among the three regions (Fig. 2): observed richness (Kruskal–Wallis $p = 0.43$) and Shannon index ($p = 0.11$) did not differ significantly. Patients from Arequipa showed the numerically lowest Shannon values (median 2.49, versus 2.77 in Lima and 2.90 in Trujillo), a consistent but non-significant trend (all BH-adjusted pairwise $p > 0.09$).

Principal coordinates analysis of the Bray–Curtis distance showed extensive overlap of the three regions (Fig. 3a). Global PERMANOVA did not identify a significant effect of region on overall community structure ($R^2 = 0.036$, $p = 0.109$), and equivalent results were obtained with the Jaccard and Aitchison distances (Supplementary Table S3, Supplementary Figure S1). In pairwise comparisons, only Trujillo versus Arequipa reached nominal significance ($R^2 = 0.037$, $p = 0.019$), which did not withstand correction for multiple comparisons (BH-adjusted $p = 0.057$); group dispersions were homogeneous. In contrast, the Firmicutes/Bacteroidota ratio differed significantly by region (Kruskal–Wallis $p = 0.012$; Fig. 3b), being highest in Trujillo (median log2 ratio 3.13) and lowest in Arequipa (1.12; BH-adjusted pairwise $p = 0.02$),

with Lima intermediate (1.75). A marginal PERMANOVA model adjusting simultaneously for region, stage, subtype, age and BMI confirmed the absence of significant clinical effects and indicated that age and BMI explained as much community variation as any clinical factor (Supplementary Table S3).

Neither clinical stage nor molecular subtype was associated with gut microbiota diversity or structure. Shannon diversity was comparable across stages I–IV (Kruskal–Wallis p = 0.23; Fig. 4) and between early (I–II) and advanced (III–IV) disease (p = 0.14), and PERMANOVA found no effect of stage on community structure (Bray–Curtis p = 0.23). Similarly, alpha diversity did not differ among molecular subtypes (p = 0.25; Fig. 5), whether the Luminal-HER2 hybrid category was kept separate or pooled with HER2, and subtype had no significant effect on beta diversity (p = 0.20).

Genus-level differential abundance analysis was consistent with these results: after BH correction, no genus was significantly associated with region (minimum q = 0.10), stage or molecular subtype (minimum q = 0.57) in the rank-based analysis of rarefied counts, and no taxon passed the ANCOM-BC2 sensitivity filter; DESeq2 findings were confined to rare, low-prevalence taxa. Nominally significant patterns, enrichment of abundant Lachnospiraceae (*Blautia*, *Dorea*, *Coprococcus*, *Fusicatenibacter*) in Trujillo relative to Arequipa, and depletion of the butyrate producers *Lachnospira* and *Oscillibacter* in HER2-positive tumors, are reported as exploratory in Supplementary Tables S4–S5.

Table 1
Clinical and demographic characteristics of patients according to region of origin

| General characteristics | | Overall (n = 77) | Lima (n = 19) | Trujillo (n = 45) | Arequipa (n = 13) |
|---|---|---|---|---|---|
| Median age (years) | | 50 (43–61) | 44 (42–49) | 53 (46–69) | 49 (41–51) |
| Body Mass Index | Normal | 27 (34.1%) | 7 (36.8%) | 18 (40%) | 2 (15.4%) |
| | Obesity | 21 (27.3%) | 5 (26.4%) | 10 (22.2%) | 6 (46.2%) |
| | Overweight | 28 (36.4%) | 7 (36.8%) | 16 (35.6%) | 5 (38.5%) |
| | Underweight | 1 (1.3%) | 0 (0%) | 1 (2.2%) | 0 (0%) |
| Molecular subtype | Luminal | 38 (49.4%) | 9 (47.4%) | 22 (48.9%) | 7 (53.8%) |
| | Luminal-HER2 | 7 (9.1%) | 0 (0%) | 5 (11.1%) | 2 (15.4%) |
| | HER2 | 13 (16.9%) | 5 (26.3%) | 8 (17.8%) | 0 (0%) |
| | Triple negative | 19 (21.7%) | 5 (26.3%) | 10 (22.2%) | 4 (30.8%) |
| Clinical stage | I | 5 (6.5%) | 1 (5.3%) | 4 (8.9%) | 0 (0%) |
| | II | 28 (36.4%) | 9 (47.4%) | 15 (33.3%) | 4 (30.8%) |
| | III | 37 (48.1%) | 8 (42.1%) | 20 (44.4%) | 9 (69.2%) |
| | IV | 7 (9.1%) | 1 (5.3%) | 6 (13.3%) | 0 (0%) |

Values are n (column %). Molecular subtype from pathology reports; Luminal-HER2 = Luminal B with HER2 overexpression.

# DISCUSSION

In this multicenter sample of Peruvian women with breast cancer, a common gut microbial core was identified, dominated by *Faecalibacterium*, *Segatella* (formerly *Prevotella*), *Blautia*, *Vescimonas* and *Bacteroides*, genera widely described in both healthy adults and breast cancer patients. These taxa have been associated with relevant metabolic and immunomodulatory functions, as well as with processes including chronic inflammation, the production of bioactive metabolites, and the modulation of estrogen metabolism—mechanisms that could influence carcinogenesis and tumor progression (6, 7, 14, 15).

Regarding regional comparisons, alpha diversity showed statistically similar values among Lima, Arequipa, and Trujillo, with a consistent but non-significant trend toward lower diversity in Arequipa, whose small sample (n = 13) limits the power of this comparison. Overall community structure likewise did not differ significantly by region in any of three complementary distance metrics; the geographic signal in this cohort was instead expressed as a compositional gradient, most clearly captured by the Firmicutes/Bacteroidota ratio, which was highest in Trujillo and lowest in Arequipa and was accompanied by nominal enrichment of abundant Lachnospiraceae genera (*Blautia*, *Dorea*, *Coprococcus*) in Trujillo. This pattern is compatible with population studies in Asia and Latin America showing that environmental, dietary and sociocultural factors associated with geography shape gut microbial composition (16–20), while indicating that in our population these effects are modest gradients of abundance rather than distinct regional community types.

In the clinical analysis, neither tumor stage nor molecular subtype was associated with differences in alpha diversity, community structure, or corrected genus-level abundance. The exploratory, uncorrected depletion of butyrate-producing genera (*Lachnospira*, *Oscillibacter*) in HER2-positive tumors is hypothesis-generating at best. The literature shows heterogeneous results: some studies indicate microbial differences between subtypes or with tumor progression, while others fail to reproduce these associations (7, 21, 22). The absence of defined patterns in our population may be due to the limited sample size in certain subgroups (Luminal-HER2, n = 7; HER2-positive, n = 13), the predominance of the Luminal subtype, and the cross-sectional design, which does not allow for the evaluation of temporal changes associated with treatment or disease progression.

The study used an ASV-based sequence-processing pipeline (DADA2, RDP training set 19), with a median sequencing depth of 154,843 reads per patient. Several limitations should be considered, including the taxonomic and functional resolution of 16S rRNA amplicon sequencing, the convenience sampling design, the absence of a cancer-free comparison group in this analysis, and the potential influence of unmeasured confounders such as diet, lifestyle, and medication use. In the adjusted analysis, age and BMI accounted for a proportion of community variation comparable to that explained by individual clinical variables, highlighting the importance of measuring and accounting for host factors in future studies (18, 19, 23).

It is recommended that future studies include a larger number of participants per region and incorporate a cancer-free control group to confirm whether the observed microbial features are specific to the disease; that they be complemented with shotgun metagenomic and metabolomic techniques to investigate the biological mechanisms by which the microbiota may influence breast cancer; that they systematically integrate variables such as diet, lifestyle, and antibiotic use, potentially linked to regional differences in microbial composition; and that multicenter and interinstitutional study designs be promoted, encompassing diverse regions of the country, to build a robust evidence base on microbiome and cancer in Peru.

# CONCLUSIONS

The gut microbiota of Peruvian women with breast cancer presented a common microbial core across the evaluated regions. Overall community structure was not significantly associated with geographic origin, clinical stage, or molecular subtype; the geographic influence was expressed as a modest compositional gradient, most evident in a significantly higher Firmicutes/Bacteroidota ratio in Trujillo and a lower ratio in Arequipa, together with a non-significant trend toward lower alpha diversity in Arequipa. These results suggest that regional, environmental and host factors (including age and BMI) may exert a greater influence on gut microbiota configuration than intrinsic tumor characteristics, while indicating that such influences take the form of subtle abundance gradients rather than distinct regional community structures.

## Declarations

**Ethics approval and consent to participate:**

The study had the prior approval of all the ethics committees of the institutions involved in its execution, which were: INEN-Lima (INEN 23–52 Letter No. 059-2023-CRPI-DI-DCON/INEN), IREN-Trujillo (Official Letter No. 1221-2023-GRLL-GGR-IRSREN NORTE-DG-CIEI), IREN-Arequipa (Registration Code 062023, Letter No. 024-2023-GRA/GRS/GR–IREN-SUR/G-DCC-DI-CRPI), Ricardo Palma University (Committee Code PPG 01-2024), and San Martín de Porres University (Official Letter No. 0952-2024-CIEI-FMH-USMP). All participants provided written informed consent prior to enrollment. The study was conducted in accordance with internationally accepted ethical standards and the principles of the Declaration of Helsinki.

**Consent for publication:**

Not applicable.


# Funding:

This study was funded by Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) and Programa Nacional de Investigación Científica y Estudios Avanzados (PROCIENCIA) within the framework of the 'Basic Research Projects' competition, according to Contract No. PE501083045-2023-PROCIENCIA and Universidad Ricardo Palma.


## Author Contribution

Conceptualization of the study : MCDP, PMC , JMVR , YF , AS , HLG, ZM. Data curation : MCDP , PMC , JMVR , YF , RT , JCP , ACS , ZM , MGV , JML , APL , LEC , GFT , TCQ , FV , HLG , AS. Methodology and formal analysis : PMC , RT , YF , MCDP , AS , JMVR. Investigation: MCDP , PMC , JMVR , YF , AS , RT , JCP , ACS , ZM , MGV , JML , APL , LEC , GFT , TCQ , FV , YF , HLG , AS. Supervision : AS, YF, PMC, RT, JMVR, MCDP. Writing-Original Draft Preparation : MCDP, PMC , JMVR , YF , AS , RT , JCP , ACS , ZM , MGV , JML ,

APL , LEC , GFT , TCQ , FV , YF , HLG. Writing- Review & Editing : MCDP , PMC , JMVR , YF , AS , RT, JCP , ACS , ZM , MGV , JML , APL , LEC , GFT , TCQ , FV , YF , HLG , AS. Funding acquisition : MCDP, AS, JMVR


# Acknowledgement

The authors thank all the patients who participated in this study. As well, the authors thank PROCIENCIA and Universidad Ricardo Palma authorities.


# Data Availability

The gut microbiome sequencing data generated and analyzed in this study have been deposited in the National Center for Biotechnology Information (NCBI) repository under BioProject accession number PRJNA1512226 and are publicly available at the NCBI BioProject database. The NCBI Submission ID is SUB16406726.

## References


1. Breast cancer cases and deaths are projected to rise globally. IARC/WHO. (2025). https://www.iarc.who.int/news-events/breast-cancer-cases-and-deaths-are-projected-to-rise-globally
2. Breast cancer statistics. *World Cancer Res. Fund*, (2025). https://www.wcrf.org/preventing-cancer/cancer-statistics/breast-cancer-statistics
3. Cáncer *PAHO*, (2025). https://www.paho.org/es/temas/cancer-mama
4. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74** (3), 229–263 (2024).
5. Terrel-Poccomo, M. K. et al. Trends in breast cancer mortality in Peru and its geographical areas from 2013 to 2022 and prediction until 2027. *BMC Cancer* ;**25**(1). (2025).
6. Álvarez-Mercado, A. I. et al. Gut Microbiota and Breast Cancer: The Dual Role of Microbes. *Cancers* **15** (2), 443 (2023).
7. Liu, K., Jia, N., Shi, H. & Ran, Y. Current and future research on the association between gut microbiota and breast cancer. *Front. Microbiol.* **14**, 1272275 (2023).
8. Bose, C. et al. Breast cancer and microbiome: a systematic review highlighting challenges for clinical translation. *BMC Women's Health* ;**25**(1). (2025).
9. Arnone, A. A. et al. Gut microbiota interact with breast cancer therapeutics to modulate efficacy. *EMBO Mol. Med.* **17** (2), 219–234 (2025).
10. Edge, S. B. & Compton, C. C. The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Ann. Surg. Oncol.* **17** (6), 1471–1474 (2010).
11. ZymoBIOMICS DNA Miniprep Kit. Zymo Research International, (2025).

12. Qubit 4 Fluorometer Assay Protocol (Q33230). Thermo Fisher Scientific, (2022).
13. Quick-16S Plus NGS Library Prep Kit. *(V3–V4, UDI)* (Zymo Research International, 2025).
14. The gut microbiota in. breast cancer development and treatment: The good, the bad, and the useful! Gut Microbes. (2023).
15. Larnder, A. H., Manges, A. R. & Murphy, R. A. The estrobolome: Estrogen-metabolizing pathways of the gut microbiome and their relation to breast cancer. *Int. J. Cancer*. **157** (4), 599–613 (2025).
16. Sun, S. et al. Does geographical variation confound the relationship between host factors and the human gut microbiota: a population-based study in China. *BMJ Open.* **10** (11), e038163 (2020).
17. Corbin, K. D. et al. Host-diet-gut microbiome interactions influence human energy balance: a randomized clinical trial. *Nat. Commun.* **14** (1), 3161 (2023).
18. Kaplan, R. C. et al. Gut microbiome composition in the Hispanic Community Health Study/Study of Latinos is shaped by geographic relocation, environmental factors, and obesity. *Genome Biol.* **20** (1), 219 (2019).
19. Reyes, A. et al. A Latin American perspective on microbiome research. *Nat. Commun.* **16** (1), 10691 (2025).
20. Valderrama, B. et al. The South American MicroBiome Archive (saMBA): enriching the microbiome field by studying neglected populations. *Nat. Commun.* ;**16**(1). (2025).
21. Altinok Dindar, D. et al. Association between Gut Microbiota and Breast Cancer: Diet as a Potential Modulating Factor. *Nutrients* **15** (21), 4628 (2023).
22. Yamada, M. et al. Characterization of gut microbiota dysbiosis in breast cancer patients. *Breast Cancer* (2025).
23. Magne, F., O'Ryan, M. L., Vidal, R. & Farfan, M. The human gut microbiome of Latin America populations: a landscape to be discovered. *Curr. Opin. Infect. Dis.* **29** (5), 528–537 (2016).
24. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*. **13** (7), 581–583 (2016).
25. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73** (16), 5261–5267 (2007). (RDP 16S rRNA training set 19).
26. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
27. Lin, H. & Peddada, S. D. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures (ANCOM-BC2). *Nat. Methods*. **21**, 83–91 (2024).
28. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15** (12), 550 (2014).

# Figures

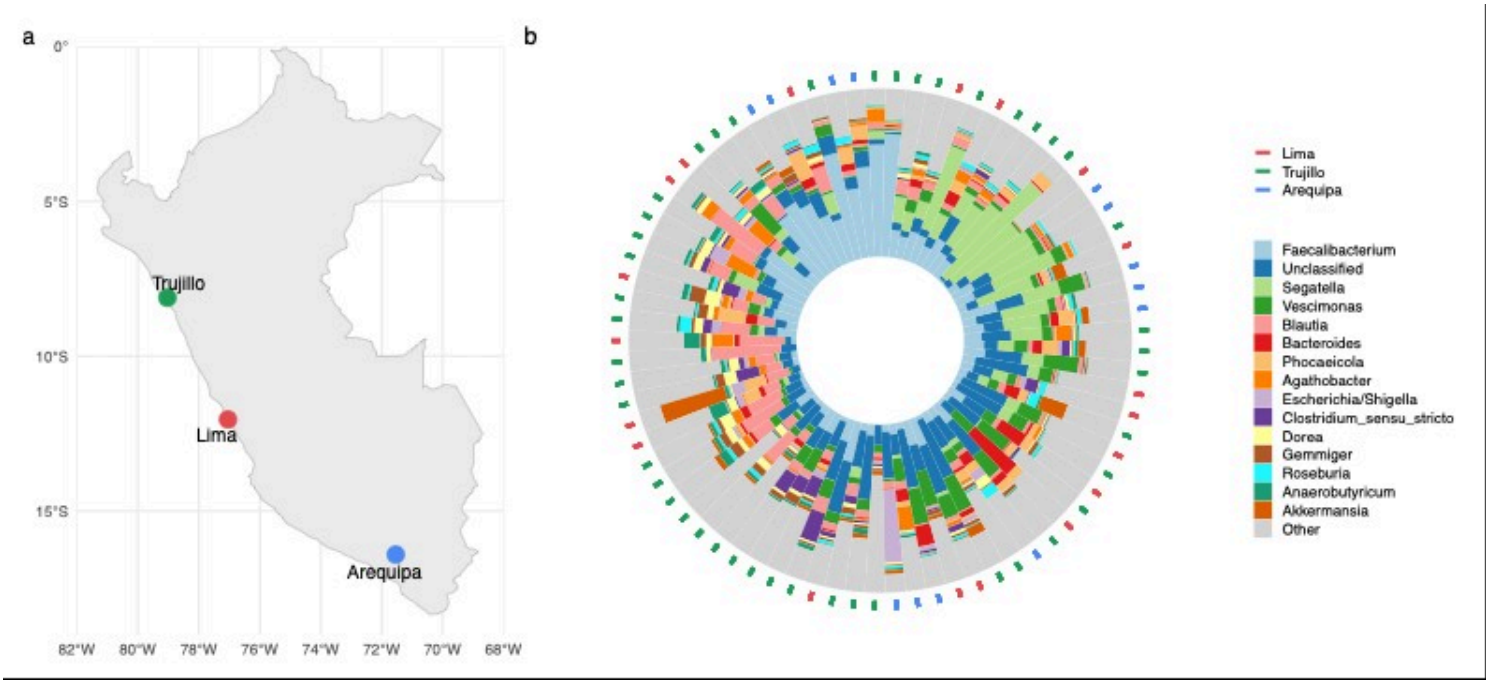



**Figure 1**

Composition of the gut microbiota by region of origin. a. Locations of the three national referral cancer centers (Lima/INEN, Trujillo/IREN-Norte, Arequipa/IREN-Sur). b. Genus-level composition of each patient's sample (n = 77; 15 most abundant genera), displayed as a circular barplot with samples arranged by Bray–Curtis PCoA similarity; the outer ring indicates each patient's region of origin (red Lima, green Trujillo, blue Arequipa).

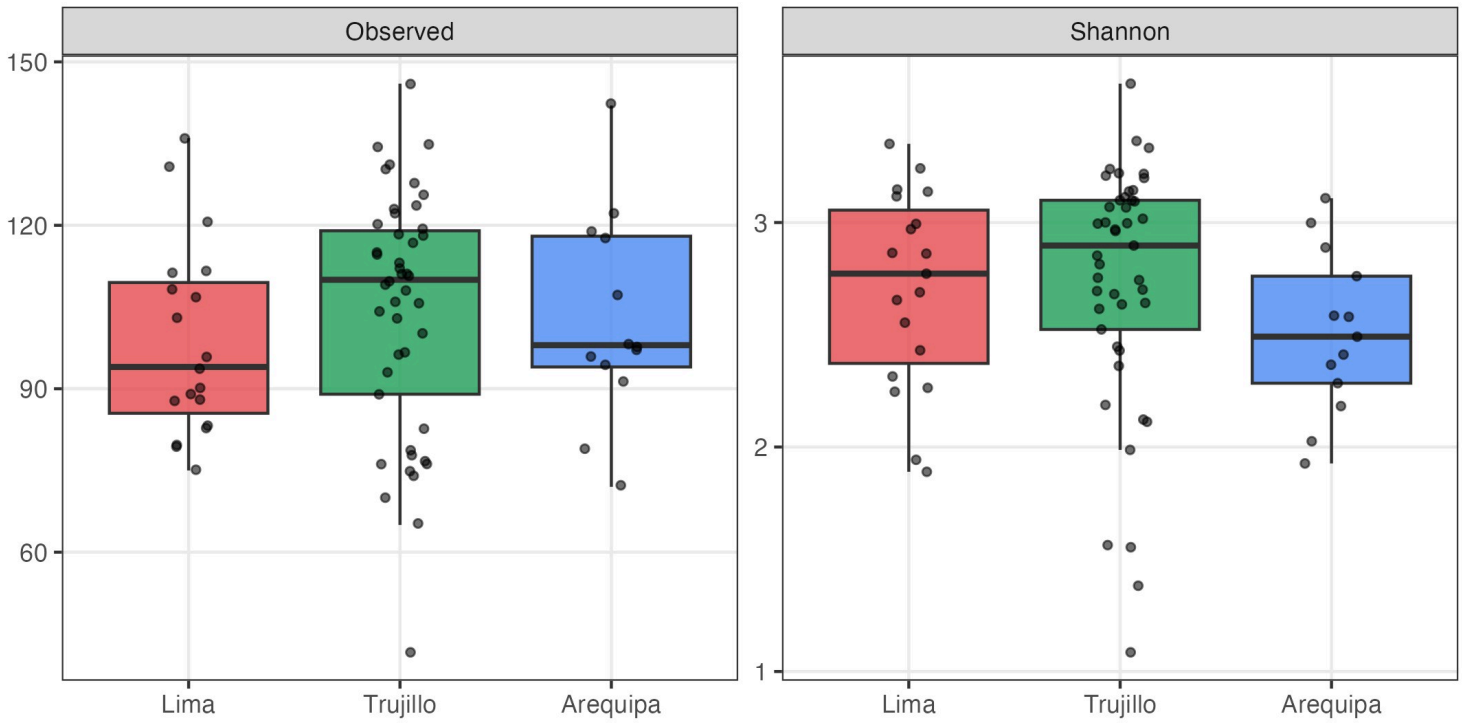



**Figure 2**

Alpha diversity by region. Observed genus richness and Shannon index computed on counts rarefied to 46,629 reads per sample. No significant differences (Kruskal–Wallis p = 0.43 and p = 0.11).

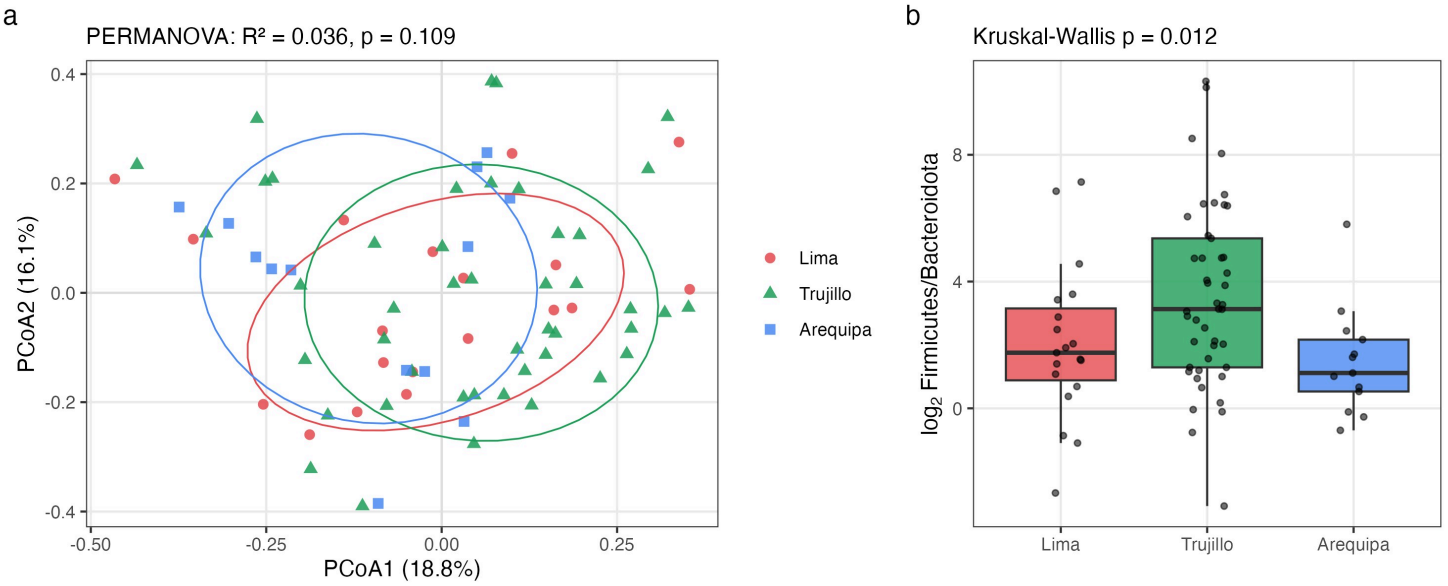



**Figure 3**

Geographic signal in the gut microbiota. a. PCoA of the Bray–Curtis distance; global PERMANOVA R² = 0.036, p = 0.109; only Trujillo vs Arequipa was nominally different (R² = 0.037, p = 0.019, BH-adjusted p = 0.057). Ellipses: 68% normal probability. b. Firmicutes/Bacteroidota log2 ratio by region (Kruskal–Wallis p = 0.012; Trujillo vs Arequipa BH-adjusted p = 0.02).

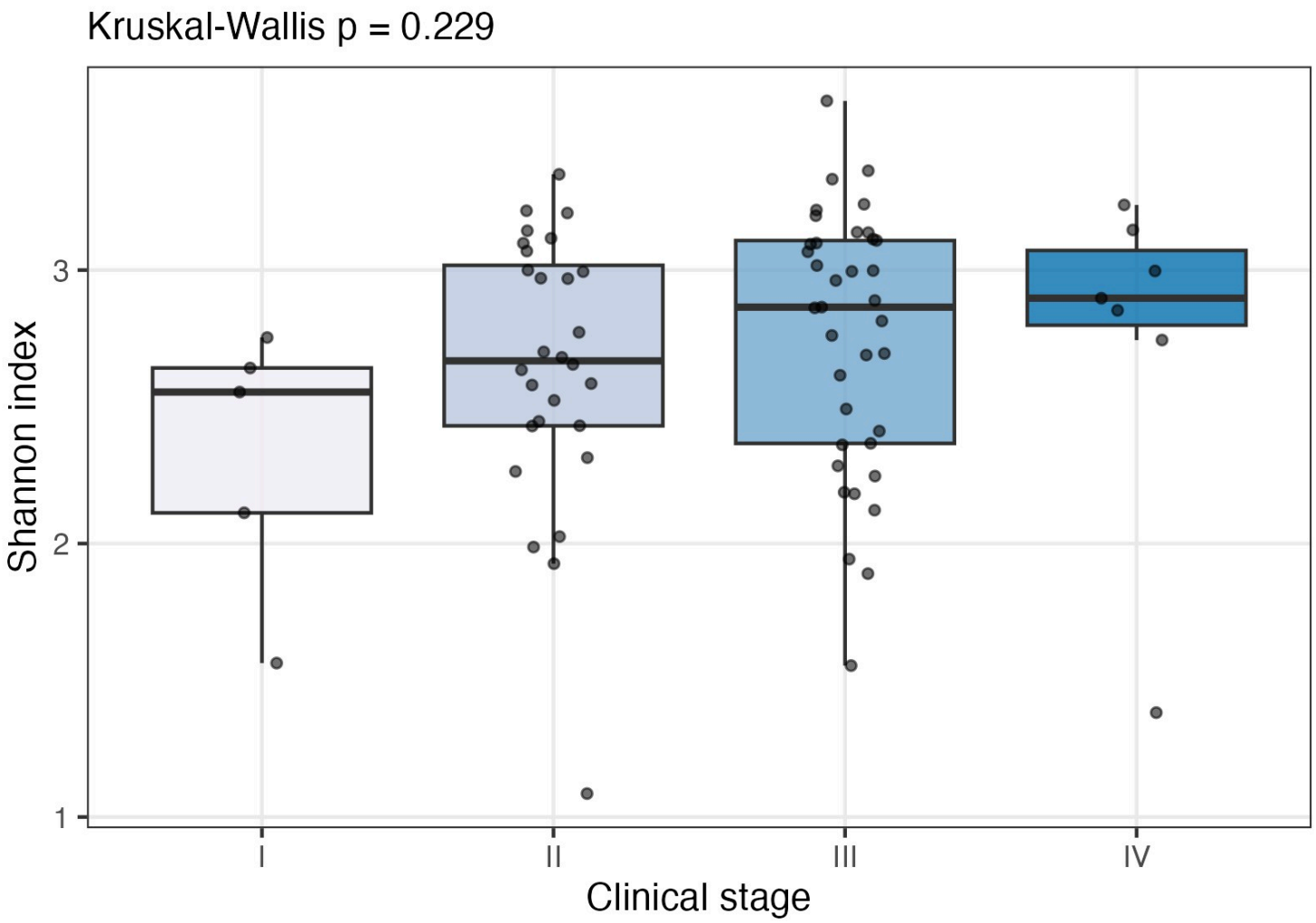



**Figure 4**

Alpha diversity (Shannon index, rarefied counts) according to clinical stage. Kruskal–Wallis p = 0.23.

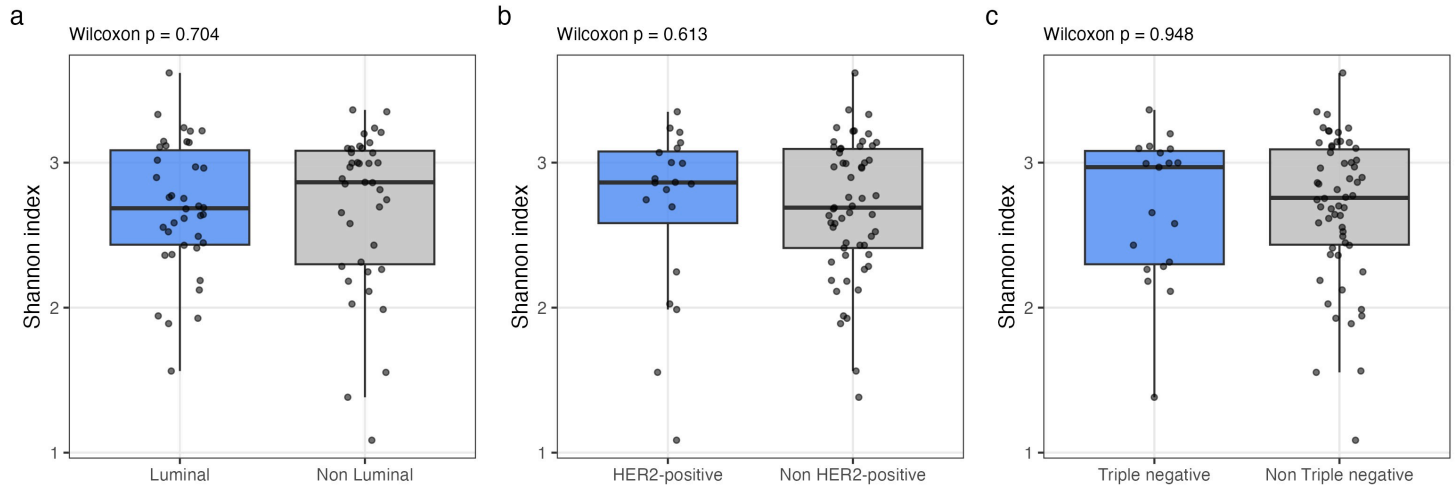



**Figure 5**

Alpha diversity of the gut microbiota according to molecular subtype (Shannon index, rarefied counts). a. Luminal versus non-Luminal tumors (p = 0.704). b. HER2-positive (including Luminal tumors with HER2 overexpression, n = 20) versus non-HER2 tumors (p = 0.613). c. Triple-negative versus non-triple-negative tumors (p = 0.948). All comparisons were based on Wilcoxon Rank sum test.