ACCESS MICROBIOLOGY
an open research platform

RESEARCH ARTICLE
Mutinda *et al.*, *Access Microbiology* 2023;5:000606.v4
DOI 10.1099/acmi.0.000606.v4

MICROBIOLOGY SOCIETY
OPEN DATA OPEN ACCESS

Profiles of bacterial communities and environmental factors associated with proliferation of malaria vector mosquitoes within the Kenyan Coast

Josphat Mutinda[1,*], Samuel Mwakisha Mwamburi[2], Kennedy Omondi Oduor[2], Maurice Vincent Omolo[3], Regina Mongina Ntabo[1], James Muhunyu Gathiru[1], Joseph Mwangangi[4] and James O. M. Nonoh[5]

Abstract

Background. Since *Anopheles* mosquitoes which transmit and maintain the malaria parasite breed in the outdoor environment, there is an urgent need to manage these mosquito breeding sites. In order to elaborate more on the ecological landscape of mosquito breeding sites, the bacterial community structure and their interactions with physicochemical factors in mosquito larval habitats was characterised in Kwale County (Kenya), where malaria is endemic.

Methods. The physical characteristics and water physicochemical parameters of the habitats were determined and recorded. Water samples were also collected from the identified sites for total metagenomic DNA extraction in order to characterise the bacterial communities within the breeding sites.

Results and Discussion. Sites where mosquito larvae were found were described as positive and those without mosquito larvae as negative. Electrical conductivity, total dissolved solids, salinity and ammonia were lower in the rainy season than in the dry season, which also coincided with a high proportion of positive sites. Pseudomonadota was the most common phyla recovered in all samples followed by Bacteroidota and then Actinomycetota. The presence or absence of mosquito larvae in a potential proliferation site was not related to the bacterial community structure in the sampled sites, but was positively correlated with bacterial richness and evenness.

Conclusion. Generally, the presence of *Anopheles* mosquito larvae was found to be positively correlated with rainy season, bacterial richness and evenness, and negatively correlated with electrical conductivity, total dissolved solids, salinity and ammonia. The findings of this study have implications for predicting the potential of environmental water samples to become mosquito proliferation sites.

DATA SUMMARY

The authors confirm that all supporting data, code and protocols have been provided within the article. All sequence data generated in this project was deposited in the National Centre for Biotechnology Information (NCBI BioProject ID: PRJNA953183; BioSample accession numbers: SAMN34109148, SAMN34109149, SAMN34109150, SAMN34109151, SAMN34109152, SAMN34109153, SAMN34109154). All physicochemical data, raw sequence data, taxonomic classification data and figures supporting this work have been deposited in the Microbiology Society's data repository Figshare account [1].

Received 11 April 2023; Accepted 31 July 2023; Published 17 August 2023
Author affiliations: [1]Kenyatta University, P.O. Box 43844-00100, Nairobi, Kenya; [2]Kenya Marine and Fisheries Research Institute, P.O Box 81651- 80100, English Point, Mkomani, Mombasa, Kenya; [3]Masinde Muliro University of Science and Technology, Centre for African Medicinal and Nutritional Flora and Fauna (CAMNFF), P.O Box 190-50100, Kakamega, Kenya; [4]Kenya Medical Research Institute (KEMRI), Centre for Geographic Medicine Research - Coast, Kilifi P.O. Box 428, Kilifi – 80108, Kenya; [5]Maseno University, Private Bag, Maseno, Kenya.
***Correspondence:** Josphat Mutinda, tushmtinda@gmail.com
Keywords: bacterial communities; physico-chemical factors; breeding sites; oviposition sites; *Anopheles gambiae*.
Abbreviations: BOD, biological oxygen demand; CCS, circular consensus sequence; CTAB, cetyltrimethyl ammonium bromide; DADA2, divisive amplicon denoising algorithm2; KEMRI, Kenya Medical Research Institute; OTUs, operational taxonomic units; WHO, World Health Organization.

INTRODUCTION

Most studies have implicated *Anopheles* mosquitoes as the leading vectors of malaria parasites in Sub-Saharan Africa [2]. They are found both in urban and rural areas but with high populations in relatively wet regions near significantly large and permanent water bodies such as lakes and oceans [3, 4]. In regions where the vector population is high, malaria is endemic. These mosquitoes commonly oviposit in small, sunlit, semi-permanent and turbid water bodies like animal footprints, the edges of boreholes, puddles on the roadside formed by tires of vehicles and tracks, irrigation canals and other artificial water sources [5].

The choice of egg-laying sites by female *Anopheles* mosquitoes depends on the biotic and abiotic factors present in specific aquatic habitats. This, in turn, affects the abundance and distribution of their larvae [6], leading to varying levels of vector distribution and abundance in a specific region which in turn affects the spatio-temporal patterns of vector distribution and abundance within a given region [7]. Despite a lack of understanding of the key factors affecting the proliferation sites and the driving forces behind oviposition site preference, even for the most prominent malaria vectors [8], it has been suggested that mosquitoes lay their eggs in locations that provide optimal conditions for larval survival and growth such as stagnant water, warm temperature of between 24–27 °C, shaded areas, oxygen, nutrients, and a neutral pH, thus increasing the chances of success for their species [9].

Although a malaria vaccine is being developed, none has been rolled out yet for widespread use, and therefore prevention of transmission by mosquitoes remains the best option for preventing malaria infections. To date, vector control remains the most effective way to prevent malaria [10, 11]. Most vector control strategies have targeted the indoor host seeking behaviour of the mosquitoes which has succeeded to a large extent [12, 13]. Despite this remarkable success, elimination of malaria remains a big challenge since the malaria parasite is maintained by mosquitoes which oviposit, feed and rest in the outdoor environment [14]. Because of this setback together with the emergence of highly drug-resistant malaria parasites [15], there is an urgent need to focus on the management and control of oviposition sites seeking malaria vectors [16]. Furthermore, in malaria endemic countries like Kenya, efficient intervention and preventive protocols should be guided by knowledge of the abundance, distribution and characteristics of the proliferation sites of these vectors if malaria were to be effectively eliminated [17]. Environmental management is relatively simple and cost-effective compared to other mosquito control measures. Since mosquitoes are becoming resistant to insecticides, and constant use of insecticides can lead to further resistance, environmental management which reduces the use of insecticides will limit the chances of resistance development. Again, environmental management is an eco-friendly way of controlling mosquito populations, and therefore promotes environmental sustainability.

In order to effectively control mosquitoes, a comprehensive understanding of their larval ecology is essential. This includes examining the interplay between biotic and abiotic factors in breeding habitats, such as the types and preferences of breeding sites, the distribution and abundance of those sites, and the biological and physico-chemical conditions present [6]. Research has suggested that proper management of mosquito breeding habitats in sub-Saharan Africa could help reduce vector populations and curb malaria transmission [18]. By analysing the choice of oviposition sites and its impact on the distribution and abundance of malaria vector mosquitoes, we may be able to explain differences in malaria transmission intensity across different regions [19]. This information is valuable in the creation of integrated control strategies for *Anopheles* mosquitoes and health education programmes at the community level, aimed at lowering mosquito populations and reducing the risk of human-vector contact.

According to the 2020 World Health Organization (WHO) World Malaria Report, Kenya had 5.6 million confirmed cases of malaria in 2019, with the majority of cases occurring along coastal Kenya. In the coastal region, the prevalence of malaria varies depending on the specific location and time of the year. Generally, the risk of malaria transmission is higher in lowland areas, especially during the rainy season when there is an increase in mosquito breeding sites. However, due to the efforts of the Kenyan government and various international organisations, there has been a significant decline in malaria prevalence in the coastal region. For example, the prevalence of malaria in Mombasa County reduced from 27% in 2016–16% in 2019 according to Kenya Malaria Indicator Survey, 2020. Despite these initiatives, malaria continues to be a major cause of disease and death, particularly among young children, pregnant women, and those with compromised immune systems in coastal Kenya.

This challenge has been compounded by mosquito resistance to common insecticides which is becoming a major concern [20]. The emergence of drug resistance in mosquitoes as a result of excessive insecticide use, reduces the effectiveness of the insecticides, and this has made it difficult for researchers to find effective ways of controlling mosquito populations. It is advisable to employ approaches like integrated vector and environmental management, to adequately control mosquito populations and prevent the development of resistance to insecticides [21]. For example, the combination of insecticides and biological management strategies or modifying the environment to eliminate or reduce the breeding sites of mosquitoes.

At present, the knowledge about the impact of proliferation site distribution, biotic and abiotic factors on the distribution and density of malaria vectors in Kenya is scarce and inadequate to explain the patterns of adult mosquito distribution and abundance with certainty. This makes it difficult to implement effective malaria vector control strategies through the management of the larval forms [22–25]. Characterising the bacterial communities in mosquito breeding sites can help researchers to understand how mosquitoes survive, and how these bacterial communities interact with mosquitoes in these environments. For instance, some bacteria may be beneficial while others may be harmful to mosquitoes. By understanding how these communities function,

researchers may be able to develop new strategies that specifically disrupt these interactions that mosquitoes rely on for survival. In response to this gap in knowledge, this study aims to characterise the total bacterial community structure and their interactions with physico-chemical ecological factors in mosquito breeding habitats in LungaLunga along the Kenyan coast, where malaria is widespread.

METHODS

Study area

The mosquito breeding sites were sampled along three major roads in Lunga Lunga sub-county, Kwale County, located along the South Coast of Kenya (as shown in Fig. 1). The selection of the sampling sites was based on the presence of larval habitats and their accessibility during the rainy season. Samples were collected along the Ramisi-Lunga Lunga road between Kanana junction (coordinates: −4° 32' 21.822'', 39° 21' 59.281'') and the Umba river in Lunga Lunga town (coordinates: −4° 33' 16.678'', 39° 7' 33.121''), the road between Lunga Lunga town (coordinates: −4° 33' 16.679'', 39° 7' 33.121'') and Ngozi Girls Secondary School in Jego village (coordinates: −4° 35' 25.346'', 39° 9' 32.242'), and the road between Jego Village (coordinates: −4° 35' 25.346'', 39° 9' 32.242') and Kanana junction on Lunga Lunga-Ramisi road (coordinates: −4° 32' 21.822'', 39° 21' 59.281'').

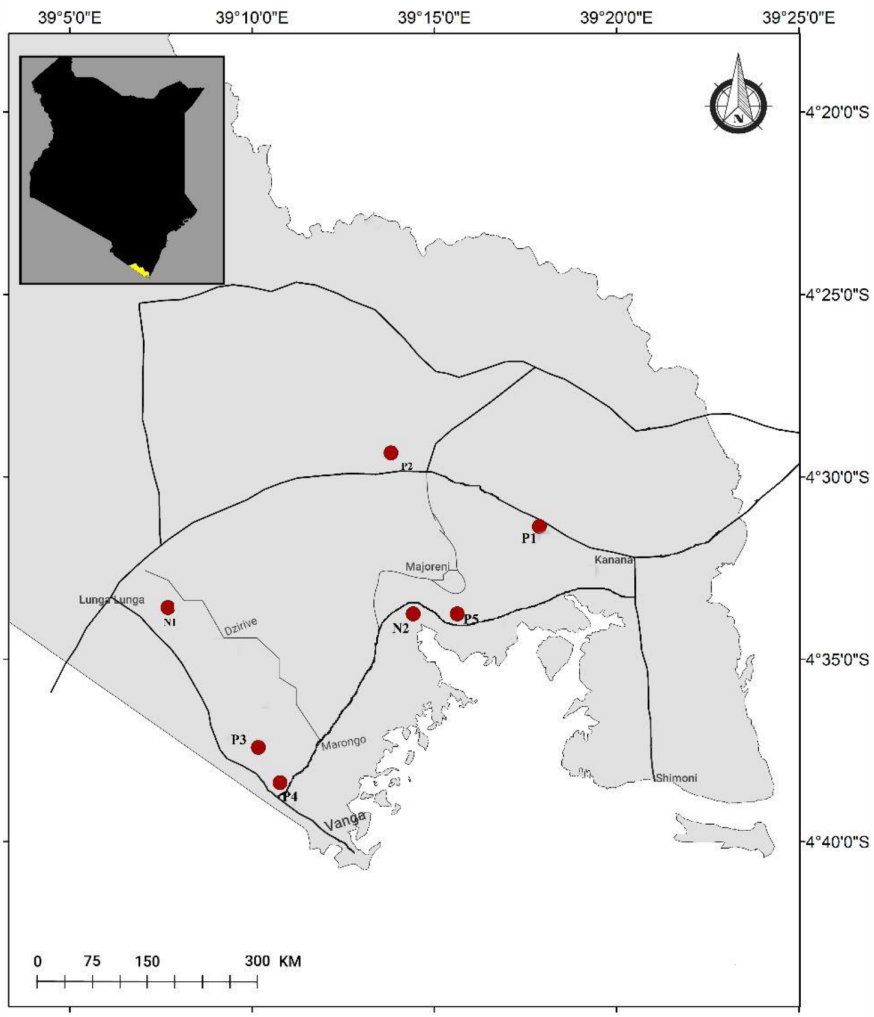

Fig. 1. A map showing the sampling sites along three major roads within Lunga Lunga sub-county in Kwale county (Kenya). Samples were collected along the Ramisi-Lunga Lunga road between Kanana junction (coordinates: −4° 32' 21.822'', 39° 21' 59.281'') and the Umba river in Lunga Lunga town (coordinates: −4° 33' 16.678'', 39° 7' 33.121''), the road between Lunga Lunga town (coordinates: −4° 33' 16.679'', 39° 7' 33.121'') and Ngozi Girls Secondary School in Jego village (coordinates: −4° 35' 25.346'', 39° 9' 32.242'), and the road between Jego Village (coordinates: −4° 35' 25.346'', 39° 9' 32.242') and Kanana junction on Lunga Lunga-Ramisi road (coordinates: −4° 32' 21.822'', 39° 21' 59.281'').

The region experiences two rainy seasons each year, between March to June and from October to November, with significant variations each year. For example, during the time of the study, there was no rainfall in October but the rainy season started towards the end of November until the end of December. Most of the residents in the region rely on small-scale farming and fishing to make a living. The area has a high prevalence of malaria among its local residents. Three species of malaria vectors, including *Anopheles arabiensis*, *Anopheles gambiae s.s.*, and *Anopheles funestus*, have been previously identified in the region [26, 27].

Study design and sample size

We conducted a cross-sectional study where 35 proliferation sites were sampled during the dry season (June to October, 2021 and January to April, 2022) and 30 sites during the rainy seasons (November to December, 202), according to the formula as described by Naing and others [28];

$$n = \frac{Z^2pqD}{d^2}$$

whereby *n*=required sample size, Z=standard normal variate which is 1.96, *P*=anticipated probability at 99%, q=failure (1 p), D=design effect of control given a value of 2, and d=allowable error (0.05).

SAMPLE COLLECTION

Identification of proliferation sites

Before the actual sample collection, each sampling site was accurately located using a GPS device (Garmin, Gpsmap 64, Garmin International Inc., Switzerland). The physical characteristics of the sites, including their natural or artificial nature, permanence, substrate type, depth, size, and vegetation, were recorded. Each potential mosquito breeding site was first visually inspected for the presence of larvae, and if larvae were not detected, a minimum of ten dips were made using a standard 350 ml dipper (BioQuip products, Rancho Dominguez, USA) to confirm the absence of larvae. A site was considered positive if at least one larva was found, and negative if no larvae were detected. The samples were collected from the selected mosquito breeding sites during both the dry and rainy seasons between June 2021 and April 2022. The sample collection was done between 7.00 am and 6.00 pm.

Water samples collection

From each selected site, a single 500 ml and two 250 ml of water samples were collected using sterile plastic and glass bottles, respectively. The sampling bottles were first rinsed with the site water, which was carefully discarded before the sample was collected. Three controls were also included in the sample collection process. Nuclease-free water (500 ml) was used as a control by opening and uncapping the bottle during sampling. The 500 ml water samples were set aside for metagenome analysis, while the two 250 ml water samples were split as follows: one for nutrient analysis (nitrates, nitrites, ammonium, and phosphates) and the other for the determination of Biological Oxygen Demand (BOD). The sample for the BOD determination was wrapped in aluminium foil to keep out light and prevent photosynthetic activity, which could alter the concentration of oxygen in the bottles. Samples for nutrient analysis were kept at ambient temperature, while those for metagenome analysis and BOD determination were preserved in a cooler box with ice packs and transported to the laboratory immediately for processing.

Mosquito proliferation sites water quality

Physico-chemical parameters, including water conductivity, temperature, total dissolved solids (TDS), dissolved oxygen (DO), pH, hardness (calcium and magnesium ions), and salinity, were measured *in situ* at each selected site using a YSI Professional Plus (Pro Plus) multi-parameter water meter (manufactured by YSI Inc., located in Yellow Springs, Ohio, USA). Three measurements were taken for each parameter. Turbidity was measured using a pre-calibrated AQUAfast AQ3010 turbidity meter (manufactured by Thermo Fisher Scientific, USA) following the manufacturer's instructions.

The biological oxygen demand (BOD) of all the collected water samples was determined using the ManTech PC-BODTM analyser (located at Highway 6 North Guelph, Ontario N1H 6J2 Canada), which provides automated BOD analysis technology. The nutrients (nitrates, nitrites, ammonium, and phosphates) in the samples were analysed using the QuAAtro AutoAnalyser (manufactured by SEAL Analytical, located at Porvair Sciences Clywedog Road South, Wrexham Industrial Estate, Wrexham, United Kingdom), which employs a continuous segmented flow analysis (CFA/SFA) technique.

Mosquito larvae collection

The collection of mosquito larvae from the selected sites was performed using standard 350 ml larval dippers (BioQuip products, Rancho Dominguez, USA). To ensure adequate collection of larvae, several dips were made at each positive site, and all collected larvae were placed into 2 litre plastic containers. After collection, the larvae were immediately transported to the laboratory for further analysis and examination. The use of standard larval dippers and plastic containers ensured that the larvae were collected and transported in a safe and secure manner, minimising the risk of contamination and preserving their viability for further analysis.

Morphological characterization of mosquito larvae

The collected mosquito larvae were filtered and placed in shallow plastic trays containing tap water. To provide proper nutrition and growth conditions for the larvae, 200 mg of powdery tetramin baby fish feed was added to the trays every morning. The water in the trays was changed every 3 days to ensure a clean and healthy environment for the larvae. Once the larvae pupated, the pupae were collected using a 5 ml plastic dropper and transferred to 500 ml plastic cups for the adult mosquitoes to emerge. The cups were covered with a fine cotton net and secured with a rubber band, with a small opening created in the centre for aspirating the emerging adult mosquitoes. This opening was covered with a piece of cotton wool to prevent any mosquitoes from escaping. Once the adult mosquitoes emerged, they were aspirated into 15 ml sterile vials using a standard mouth aspirator (Model 412) and stored in a refrigerator at 4 °C to allow the mosquitoes to die. The morphological features of the adult mosquitoes were then observed under a dissecting light microscope and identified based on morphological characters described in previously published keys [29]. *Anopheles* mosquitoes were identified up to species level while *Culex* mosquitoes were only identified up to genus level.

16S rRNA gene-based analysis

Sample preparation and total metagenomic DNA extraction

The thirty water samples collected were grouped into seven final samples based on the proximity of the sites and the presence or absence of mosquito larvae. Samples that were collected from sites where mosquito larvae were observed were labelled as P1, P2, P3, P4, and P5, while those without larvae were labelled as N1 and N2. The samples were collected from different regions along the roads, where P1 and P2 were obtained from sites located between Kanana Junction (coordinates: −4° 32' 21.822", 39° 21' 59.281") and River Umba in Lungalunga (coordinates: −4° 33' 16.678", 39° 7' 33.121"), N1 and P3 from sites located between River Umba (coordinates: −4° 33' 16.679", 39° 7' 33.121") and Ngozi Girls Secondary School in Jego village (coordinates: −4° 35' 25.346", 39° 9' 32.242'), and P4, N2, and P5 were collected from sites between Jego village (coordinates: −4° 35' 25.346", 39° 9' 32.242') and Kanana Junction (coordinates: −4° 32' 21.822", 39° 21' 59.281") as shown in Fig. 1.

The preparation of the water samples for the extraction of total metagenomic DNA was performed as described before [30]. One litre of each of the final samples was filtered through sterile 0.22 µm filter membranes (Merck Millipore, Burlington, MA) to trap bacterial cells. The filter membranes were aseptically removed from the filtration apparatus and cut into four pieces using a sterile pair of forceps and scissors. The pieces were then placed along the bottom of a 50 ml sterile conical tube with the upper surface of the filter facing the centre of the tube. Thirty millilitres of extraction buffer were added to the tube. The trapped biomass was washed off the filters by vortexing the tubes vigorously, and the cell suspension was transferred to a clean microcentrifuge tube. The tube was incubated in a heating block at 65 °C for 30 min, with gentle vortexing after every 10 min. After the incubation period, the tube was allowed to cool to room temperature, and an equal amount of chloroform: isoamyl alcohol (24:1 v/v) was added and mixed by gentle inversion. The mixture was then centrifuged at 13200 r.p.m. for 5 min at room temperature, and the supernatant was transferred to a new 50 ml tube. Total genomic DNA was then precipitated, cleaned, and resuspended in nuclease-free water. The concentration and purity of the extracted DNA were assessed using 1% agarose gel electrophoresis [31] and a NanoDrop spectrophotometer [32], then stored at −40 °C.

Next generation sequencing

In this study, the 16S rRNA gene was targeted and amplified using the primers F27 'AGRGTTYGATYMTGGCTCAG' and R1492 'RGYTACCTTGTTACGACTT' [33]. The annealing temperature was optimised around 55 °C. An initial denaturation step at 95 °C for 5 min, followed by 35 cycles of amplification at 95 °C for 30 s, annealing temperature for 30 s, and 72 °C for 60 s, a final extension step at 72 °C for 10 min. The amplified product was then sequenced on the PacBio Sequel platform using PacBio Barcoded M13 Primers for Multiplex SMRT Sequencing. A positive control sample containing 17 known bacterial isolates was used as a mock to test the sequencing and analysis pipelines.

Metataxonomics

PacBio sequences obtained were processed and visualised using the RS_ReadsOfInsert protocol in the SMRT Analysis software version 2.3 to obtain demultiplexed consensus sequences with a minimum of three full passes. The resulting sequence data were processed using the Divisive Amplicon Denoising Algorithm2 (DADA2) pipeline [34] in R version 4.2.1, R Core Team (2022) as follows. First, the F27 and R1492 primers were removed from the raw sequences and the quality of the reads was inspected. The sequences were then filtered using the parameters; minQ=2, minLen=500, maxLen=1600, maxN=0, rm.phix=FALSE, maxEE=2; minQ=2 sets the minimum quality score for each base in the sequence. A quality score of 2 represents a 1% error probability, meaning that a base with a quality score of 2 may be incorrect only 1% of the time. minLen=500 sets the minimum length of the reads used in the analysis. Sequences shorter than 500 bases were excluded from the analysis. maxLen=1600 sets the maximum length of the reads used in the analysis. Sequences longer than 1600 bases were excluded from the analysis. maxN=0 specifies the maximum number of ambiguous bases (N) allowed in each sequence. A value of 0 means that no ambiguous bases were allowed in the sequences. rm.phix=FALSE specifies whether to remove reads derived from the phix174 genome. A value of 'FALSE' means that phix174 genome reads were not removed. maxEE=2. sets the

maximum number of expected errors allowed in each sequence. The expected error rate is calculated from the quality scores of the bases. A value of 2 means that sequences with an error rate greater than two were excluded from the analysis.

Taxonomic classification of the filtered reads was then assigned to the species level using kraken2 pipeline [35], implementing the bacteria refseq database of the NCBI (NCBI Bacterial RefSeq Database. http://ftp.ncbi.nlm.nih.gov/refseq/release/bacteria/bacteria.1.1.genomic.fna.gz. Accessed 30 May 2023). To assess the reliability of the sampling depth, the OTU tables were rarefied and the precision of the rarefaction curves was estimated using the bootstrapping method [36]. Alpha diversity measures were calculated using *vegan* package version 2.6–2 [37] in R from the number of OTUs. Shannon diversity estimate based on species richness and evenness emphasising more on species richness and Simpson diversity index based on species richness and evenness putting more weight on species evenness [38–40].

All statistical data analysis was performed in XLSTAT [41] and R statistical programme version 4.2.1 (R Core Team, 2022). Physical characteristics of positive and negative sites were represented in percentages and compared using the Z-test. Physicochemical data was summarised using mean and standard deviation. Student's t-test and Permutational Multivariate Analysis of Variance (PERMANOVA) were used to compare physicochemical parameters between the positive and negative sites at 95% confidence interval [42–45]. The Principal components analysis (PCA) was conducted to identify the relationships between different parameters and the sites. To test whether there was any significant effect associated with the physicochemical parameters on the alpha diversity of bacteria, non-parametric Kruskal-Wallis test was performed [46].

RESULTS

Distribution of mosquito sites

We sampled 35 sites during the dry season (June to October, 2021 and January to April, 2022) and 30 during the wet season (November to December, 2021), and evaluated them for the presence (positive) or absence (negative) of mosquito larvae as described in Fig. 2. During the dry season, 19 sites (54.28%) were positive and 16 (45.72%) were negative, indicating that the number of positive and negative sites were not significantly different (P=0.321, 95% CI: 0.068, 0.228). However, in the rainy season, we found that 26 (86.67%) out of 30 sampled sites were positive and only four (13.33%) were negative for mosquito larvae (P<0.0001, 95% CI: 0.637, 0.843). Overall, the positive sites for the presence of mosquito larvae were 69.23% while the negative sites were 30.77% (P<0.0001, 95% CI: 0.242, 0.518). The proportion of positive sites was also significantly higher in the rainy season (86.67%) than in the dry season (P<0.0001, 95% CI: 0.158, 0.442).

Abundance of mosquito larvae in the positive sites

We collected 1360 mosquito larvae from the positive sites and reared them into adult mosquitoes, which were then identified as *Anopheles gambiae* or *Culex* sp., with the abundance of *Anopheles gambiae* being 68.34% (P < 0.0001, 95% CI: 0.221, 0.499) and 71.32% (P < 0.0001, 95% CI: 0.306, 0.574), during the dry and rainy seasons respectively (Fig. 3). The average percentage of *Anopheles gambiae* across both seasons was 69.83% (P < 0.0001, 95% CI: 0.263, 0.567).

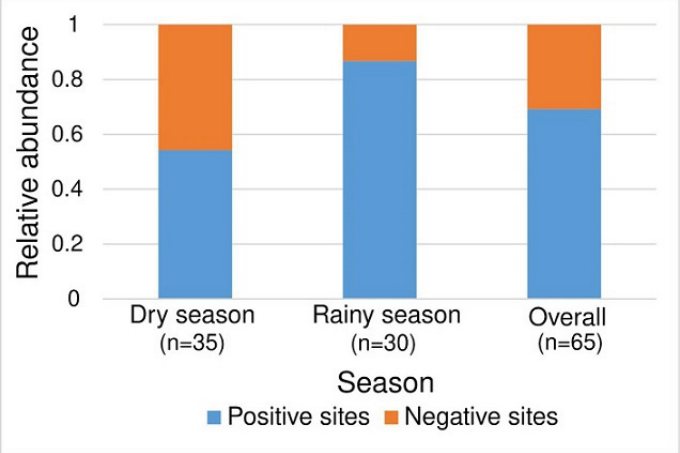

Fig. 2. Abundance of mosquito larval habitats, During the dry season (June to October, 2021 and January to April, 2022), the number of positive sites was comparable to that of negative sites (P=0.321, 95% CI: 0.068, 0.228). However, during the rainy season (November to December, 2021), there were more positive sites than negative ones (P<0.0001, 95% CI: 0.637, 0.843).

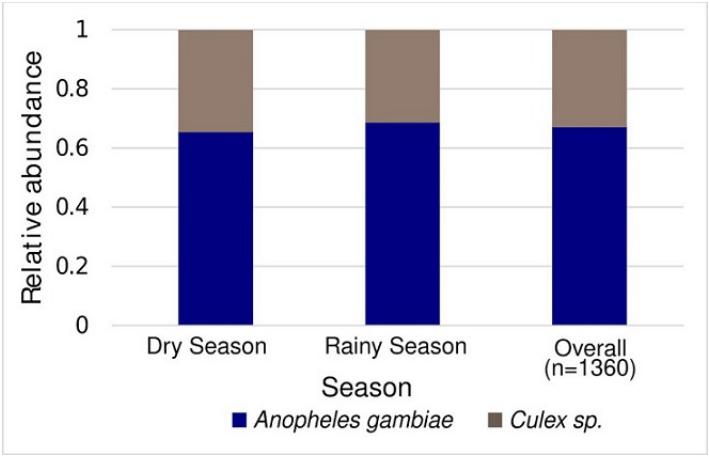

Fig. 3. Abundance of mosquito species in larval habitats (*N*=1360), The majority of the sampled sites were found to contain *Anopheles gambiae* larvae, with only a few containing *Culex* mosquitoes (P < 0.0001, 95% CI: 0.263, 0.567).

Occurrence of mosquito larvae in the sites

We found that 63.15% of the positive sites had *An. gambiae* larvae only, 5.20% had *Culex* sp. larvae only, and 31.65% had both *An. gambiae* and *Culex* sp. larvae (X^2=5.991, df=2, *P*<0.0001) as shown in Fig. 4. This suggests that during the study period, more habitats were suitable for the breeding of *An. gambiae* compared to other species.

Physical characteristics of proliferation sites

These were determined based on 35 sites sampled during the dry season and 30 sampled during the rainy season. Most of the sites were natural habitats (94.28%) while only 5.71% were artificial, such as man-made dams and road culverts (*P*<0.0001, 95% CI: 0.8295, 0.9704). Natural habitats included marshy areas, shallow rivers, roadside pools, and animal hoof-prints. We found 65.71% of the sites had mud substrates while 34.28% had sand (*P*=0.006, 95% CI: 0.179, 0.461). In terms of permanence, 65.7% of the sites were semi-permanent, while the rest were permanent (*P*<0.0001, 95% CI: 0.158, 0.442). Most of the sites were fully exposed to sunlight (94.28%, *P*<0.0001, 95% CI: 0.8295, 0.9704) and had a shallow depth of less than 1 m (77.14%, *P*<0.0001, 95% CI: 0.413, 0.667) with an average size of less than 10 m^2 (94.28%, *P*<0.0001, 95% CI: 0.8295, 0.9704). In terms of vegetation, 82.86% of the habitats had some form of vegetation while 17.14% had no vegetation at all (*P*<0.0001, 95% CI: 0.546, 0.774). The habitats were grouped into four categories based on the type of vegetation present. The majority of the habitats had only algae (54.28%, *P*<0.0001, 95% CI: 0.144, 0.458). Some habitats had a combination of algae, submerged, and emergent vegetation (11.42%), others had algae and emergent vegetation (8.57%), and a few had algae with only emergent vegetation (8.57%). Information on the physical characteristics of the sites is presented in Fig. 5.

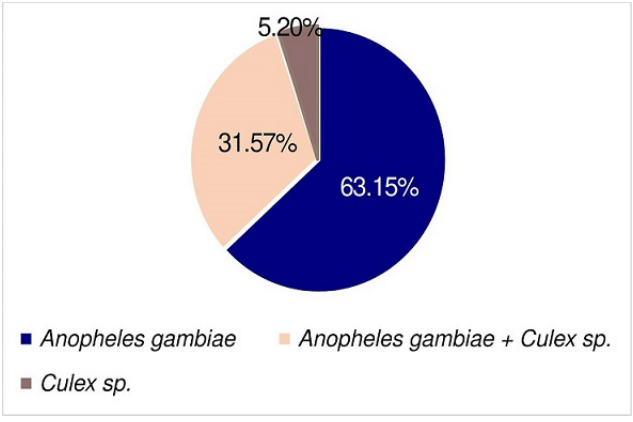

Fig. 4. Abundance of mosquito larvae in the positive sites (*N*=1360), 63.15% of the positive sites had *Anopheles gambiae* larvae only, 5.20% had *Culex* sp. larvae only, and 31.65% had both *Anopheles gambiae* and Culex sp. larvae (X^2=5.991, df=2, *P*<0.0001).

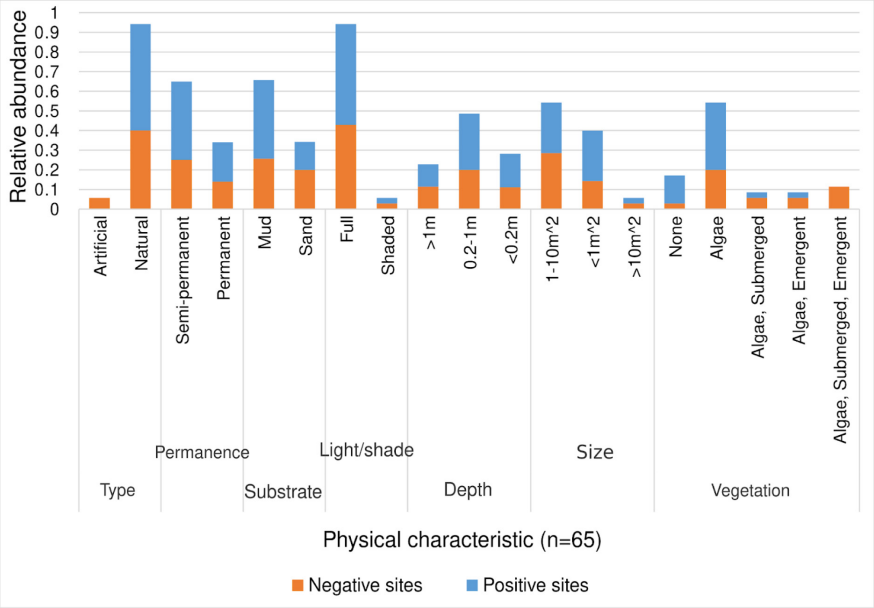

Fig. 5. Physical characteristics of larval habitat, most of the sites were natural (P<0.0001, 95% CI: 0.8295, 0.9704), semi-permanent (P<0.0001, 95% CI: 0.158, 0.442), had mud substrate (P=0.006, 95% CI: 0.179, 0.461), were fully exposed to sunlight P<0.0001, 95% CI: 0.8295, 0.9704), had a shallow depth of less than 1 m (P<0.0001, 95% CI: 0.413, 0.667) with an area of less than 10 m^2 (P<0.0001, 95% CI: 0.8295, 0.9704), and had algae as the main vegetation (P<0.0001, 95% CI: 0.144, 0.458).

Physicochemical parameters of the sites

Although mean temperature at the positive sites was significantly lower than that in the negative sites during the dry season (t=1.729, df=19, P=0.0416), there was no significant difference in temperature between positive and negative sites during the rainy season (Table 1). None of the twelve physicochemical parameters studied differed significantly between positive and negative sites in both seasons (R2=0.1180, df=1, P=0.106). A pairwise comparison of the individual physicochemical parameters using the Student's t-test between the dry and rainy seasons showed that salinity (t=1.692, df=33, P=0.01104), electrical conductivity (t=1.689, df=33, P=0.01617), total dissolved solids (t=1.690, df=33, P=0.01204), and ammonia (t=1.675, df=33, P=0.00029) were significantly lower during the rainy season compared to the dry season, while the other variables were not significantly different. The mean, standard deviations, and p-values for the Student's t-test of the physicochemical parameters between positive and negative sites evaluated during the dry and rainy seasons are summarised in Table 2.

Metataxonomic analysis of bacteria

Sequencing reads quality control

Sequencing of the full 16S rRNA generated 203934 reads from seven environmental samples with an average of 29133 reads per sample (Table 3). The average length of the reads was approximately 1450 base pairs, which aligns with the expected full length of the 16S rRNA gene. After applying various quality control measures such as trimming, filtering, and denoising, we were left with 104040 reads.

Taxonomic classification of the filtered reads

The ASVs were taxonomically assigned to 10 phyla, 40 classes, 87 orders, 129 families, 188 genera and 257 species. The most commonly detected phylum was Pseudomonadota, which accounted for 56% of the total reads, followed by Bacteroidota (26%) and Actinomycetota (15%). The other seven phyla accounted for 3% of all the phyla detected in all samples (Fig. 6). At the class level, Betaproteobacteria, Flavobacteriia, Gammaproteobacteria, and Actinomycetes were the most abundant (Fig. 7). The most common orders were Burkholderiales, Flavobacteriales, Micrococcales, Hyphomicrobiales, and Alteromonadales (Fig. 8). Out of all the families identified, Flavobacteriaceae, Comamonadaceae, Microbacteriaceae, Methylobacteriaceae, and Burkholderiaceae were the most abundant (Fig. 9). For the 188 genera, *Flavobacterium, Comamonas, Methylobacterium, Cryobacterium, Cupriavidus,* and *Pseudomonas* were the most detected (Fig. 10). Further, differential abundance analysis revealed significant differences in the abundance of bacterial genera between the positive and negative sites. Specifically, the genus *Vibrio* and *Cutibacterium* were found to be significantly more abundant in the positive sites compared to the negative sites (P<0.05), indicating a potential association

Table 1. Summary of physicochemical parameters

Season		Temp (°°C)	pH	DO (mg l^{-1})	BOD_5 (mg l^{-1})	Salinity (ppt)	E.C (us m^{-1})	TDS (mg l^{-1})	TUR (NTS)	NO_3^- (umol l^{-1})	NO_2^- (umol l^{-1})	NH_3 (umol l^{-1})	PO_4^{3-} (umol l^{-1})
Dry season	**Positive sites ($\bar{X}$)**	30.82	8.62	7.52	5.347	2.19	4481	2634	386.2	24.9	6.02	20.25	3.21
	Standard deviation	±2.37	±0.69	±5.0	±5.58	±1.73	±3398	±2012	±311	±12.9	±3.25	±12.4	±3.03
	Negative sites ($\bar{X}$)	32.57	8.26	5.74	2.708	3.058	5363	3060	319	21.71	4.98	27.03	4.65
	Standard deviation	±1.93	±0.81	±3.2	±2.14	±5.2713	±893	±5136	±417	±12.1	±3.74	±12.7	±4.3
	***P*-value**	0.041*	0.206	0.27	0.115	0.5410	0.7183	0.7625	0.622	0.495	0.423	0.153	0.296
Rainy season	**Positive sites ($\bar{X}$)**	32.14	8.47	5.88	2.8	0.67	1685.7	898.51	472.0	23.97	4.59	12.87	3.73
	Standard deviation	±1.28	±0.55	±2.0	±1.64	±1.1	±2228	±1174	±373	±7.43	±3.17	±9.55	±3.41
	Negative Sites ($\bar{X}$)	31.58	8.167	5.59	2.232	0.36	1414	723.01	453.5	25.21	3.9	12.37	2.56
	Standard deviation	±2.52	±0.33	±2.2	±1.67	±0.58	±1898	±1221	±101	±4.83	±1.36	±1.64	±1.41
	***P*-value**	0.691	0.172	0.81	0.561	0.4175	0.8070	0.8009	0.835	0.677	0.466	0.803	0.249

Temp – Temperature, DO – Dissolved oxygen, BOD5 – Biological oxygen demand after 5 days incubation, E.C – Electrical conductivity, TDS – Total dissolved solids, TUR – Turbidity, NH_3 - Ammonia, a summary of the physicochemical parameters between positive and negative sites evaluated during the dry and wet seasons shows that during the dry season, the mean temperature for positive sites was significantly lower than that of negative sites (t=1.729, df=19, **P*=0.0416), but this was not the case during the rainy season (t=1.89, df=19, *P*=0.691).

Table 2. Comparison of physicochemical parameters in the sites between dry and rainy seasons

Parameters (*N*=30)		Temp (°C)	PH (mg l^{-1})	DO (mg l^{-1})	BOD_5 (mg l^{-1})	Salinity (ppt)	E.C (us m^{-1})	TDS (mg l^{-1})	TUR (NTS)	NO_3^- (umol l^{-1})	NO_2^- (umol l^{-1})	NH_3 (umol l^{-1})	PO_4^{3-} (umol l^{-1})
Dry season	Mean SD	31.74±2.27	8.431±0.76	6.57±4.2	3.94±4.26	2.653±3.989	4951.8±6831	2861.9±3938	350.7±366	23.21±12.4	5.4685±3.5	23.87±12.8)	3.97±3.79
Rainy season	Mean SD	32.06±1.44	8.43±0.52	5.84±2.0	2.72±1.63	0.628±1.04	1650±2162	875.86±1161	469.6±349	24.14±7.1	4.5±2.99	12.81±8.91	3.58±3.23
Overall	Mean	31.9	8.433	6.20	3.32	1.624	3274.1	1852.6	411.1	23.67	4.98	18.25	3.776
	SD	±1.89	±0.65	±3.29	±3.23	±3.04	±5259	±3228	±360	±10.0	±3.262	±12.28	±3.49
***P*-value**		0.511	0.986	0.39	0.152	0.01104	0.0161*	0.012*	0.200	0.722	0.254	0.0003*	0.6638

Temp – Temperature, DO – Dissolved oxygen, BOD5 – Biological oxygen demand after 5 days incubation, E.C – electrical conductivity, TDS – Total dissolved solids, TUR – Turbidity, NH_3 – Ammonia, A pairwise comparison of the individual physicochemical parameters using the Student's t-test between the dry and rainy seasons showed that salinity (t=1.692, df=33, *P*=0.01104), electrical conductivity (t=1.689, df=33, ***P*=0.01617), total dissolved solids (t=1.690, df=33, ***P*=0.01204), and ammonia (t=1.675, df=33, ***P*=0.00029) were significantly lower during the rainy season compared to the dry season, while the other variables were not significantly different.

Table 3. Sequence reads distribution per sample

Sample	CCS	Primers	After QC
P1	27250	22977	13206
P2	29589	25225	14719
N1	33903	29121	19033
P3	24670	21707	11806
P4	37548	32977	21498
N2	12822	10900	3878
P5	38152	34082	19900
Total	203934	176989	104040
Mean	29133	25284	14862
±SD	±8799	±8501	±6049

CCS – Circular consensus sequence, QC - Quality control, CCS represents the number of raw reads, while after QC (quality control) represents the number of reads remaining after performing quality control.

with positive environmental conditions. On the other hand, the genus *Methylorubum* exhibited lower abundance in the positive sites compared to the negative sites ($P<0.05$). These findings suggest that *Vibrio* and *Cutibacterium* may play a role in positive site conditions, while *Methylorubum* may be more prevalent in the negative site environment. The volcano plot in Fig. 11 visually represents these differential abundance patterns, with points coloured in navy blue indicating differentially abundant genera and points coloured in grey representing genera with no significant differential abundance.

Diversity of bacterial communities

A rarefaction plot was used to compare biodiversity across different samples. Sample N2, which had the lowest species count, was used to determine the minimum sample size for downstream analysis. Rarefaction allowed for standardising the comparison of biodiversity by estimating species richness or diversity at the same level of sampling effort, irrespective of sample size (Fig. 12). Several alpha diversity indices for the positive and negative samples were also computed and compared as shown in (Fig. 13). All the alpha diversity indices were highest in sample P3 and lowest in sample N1 and they were highly variable between the

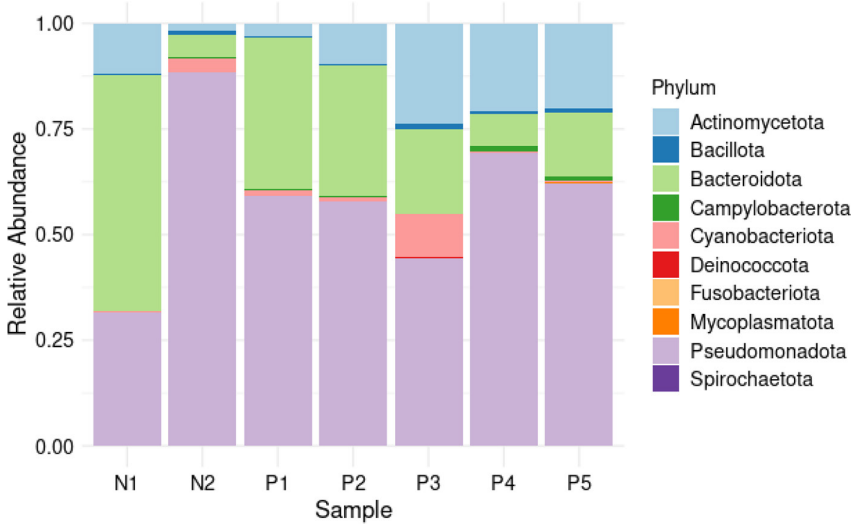

Fig. 6. Phylum distribution. The stacked barplot depicts the relative abundance of bacterial phyla in the study samples. The negative group, represented by samples N1 and N2, and the positive group, represented by samples P1 to P5, are differentiated. Each sample is represented by a bar, and the height of the bar corresponds to the percentage of abundance. The colours used in the plot represent different bacterial phyla, providing insight into the composition of the microbial community across the samples.

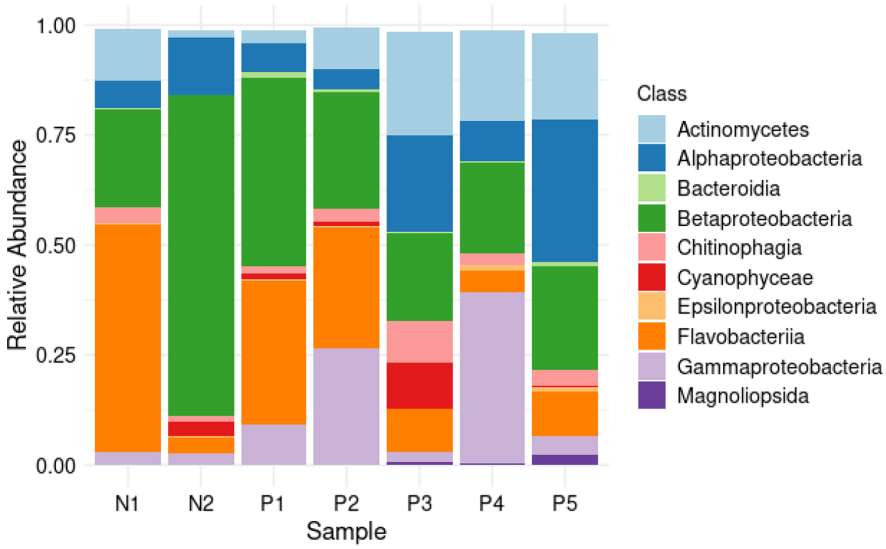

Fig. 7. Class distribution. Stacked barplot showing top ten bacterial classes' relative abundance in negative (n1, n2) and positive (p1-p5) samples.

sites, an indication that the bacterial community richness, evenness and abundance were not similar between sites (F=2.928, df=8, *P*=0.032). However, the physicochemical parameters tested had no effect on the diversity indices of bacterial communities between sites (Kruskal-Wallis chi-squared=6, df=6, *P*=0.4232). Samples taken from positive sites were noted to have higher alpha diversity indices in comparison to those from negative sites and therefore indicating that the presence of mosquito larvae was correlated with high bacterial richness and evenness (R^2=9.822, df=1, *P*=0.00197). For beta diversity analysis, the positive and negative samples did not present any specific clustering pattern (Fig. 14), which is further supported by the statistical analysis (R^2=0.18157, df=1, *P*=0.353). This suggests that the presence or absence of the mosquito larvae was not correlated with any particular bacterial community structure.

DISCUSSION

Mosquitoes proliferation sites

During the dry season (June to October, 2021 and January to April, 2022), the number of positive sites was comparable to that of negative sites. However, during the rainy season (November to December, 2021), there were more positive sites than negative

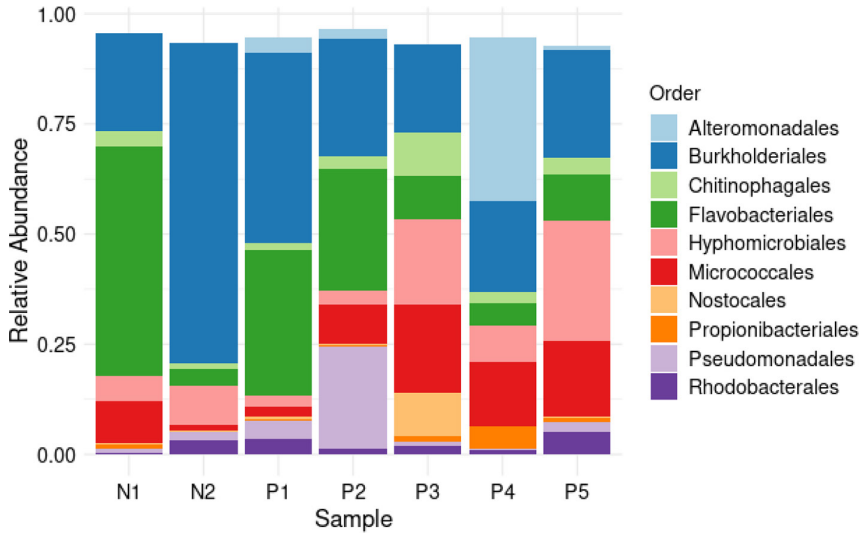

Fig. 8. Order distribution. Stacked barplot showing top ten bacterial order's relative abundance in negative (n1, n2) and positive (p1-p5) samples.

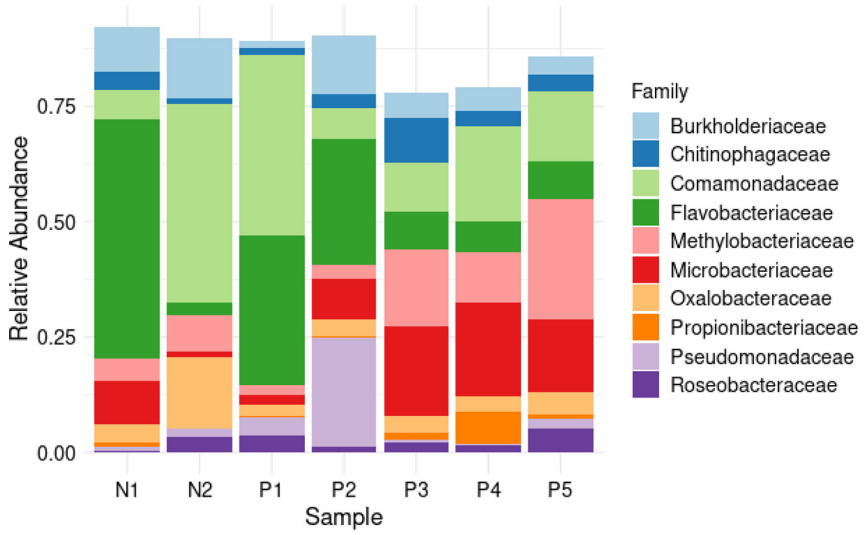

Fig. 9. Family distribution. Stacked barplot showing top ten bacterial families' relative abundance in negative (n1, n2) and positive (p1-p5) samples.

ones. This result indicates that the rainy season's rainfall patterns can impact the reproduction and distribution of mosquitoes in the environment. The increased humidity and availability of water during the rainy season create favourable conditions for mosquito proliferation. The presence of positive sites even during the dry season is a concern as it contributes to malaria transmission year-round.

The majority of the sampled sites were found to contain *Anopheles gambiae* larvae, with only a few containing *Culex* mosquitoes. This indicates that the conditions in these sites were more conducive to the proliferation of *Anopheles gambiae* than other species. *Anopheles* mosquitoes prefer freshwater habitats such as ponds, marshes, slow-moving streams, and irrigated farmland [3]. They thrive in conditions where there is a higher concentration of organic debris, and are more active and capable of breeding at higher temperatures than other mosquito species. In general, the temperature range for optimal mosquito breeding is 20–30 °C [11]. *Anopheles* mosquitoes require high humidity levels to breed and thrive and hence they are more likely to be found in humid regions where there is a lot of rainfall or near stagnant water bodies [8]. These mosquitoes prefer feeding on humans for their blood meal and consequently, they are more likely to be found in areas near human habitation, making them a more significant threat to humans than other mosquito species [26]. *Anopheles* mosquitoes are more active during the rainy season where they often lay their eggs in temporary water sources that form during heavy rainfall, and hence making them more common in regions

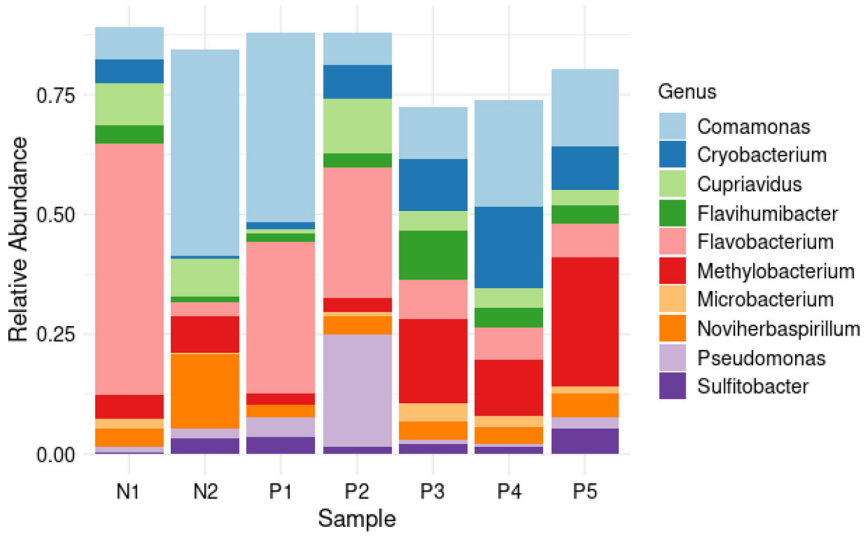

Fig. 10. Genus distribution. Stacked barplot showing top ten bacterial genera relative abundance in negative (n1, n2) and positive (p1-p5) samples.

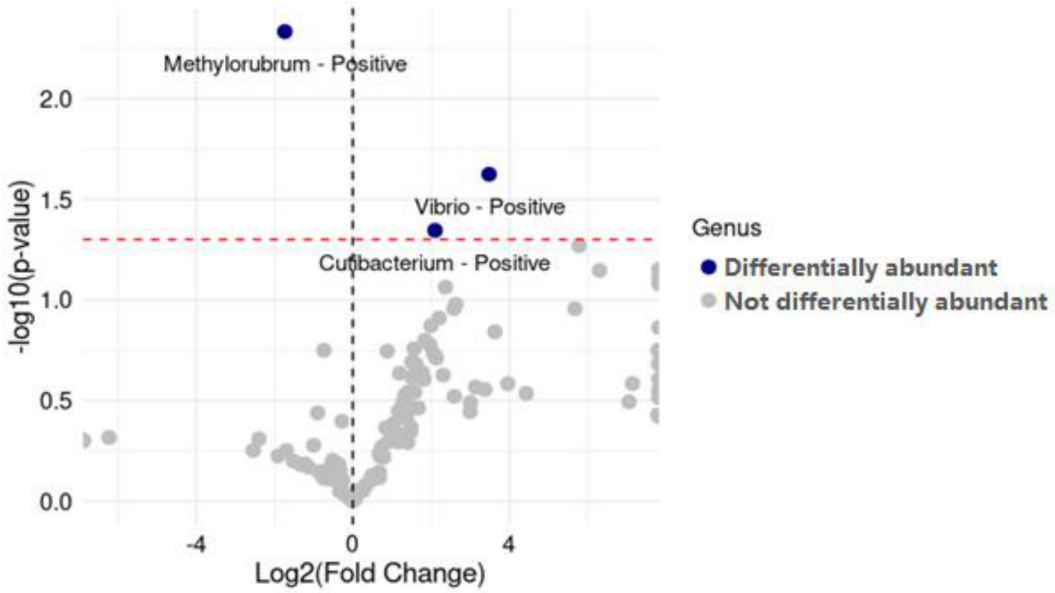

Fig. 11. Differential abundance analysis of bacterial genera. The figure shows bacterial genera whose abundance was significantly different between the positive and negative sites. Each point represents a genus, with the x-axis representing the log2 fold change in abundance between the two groups, and the y-axis representing the -log10(*p*-value) indicating the significance of differential abundance. Points coloured in navy indicate genera that are differentially abundant, while points coloured in grey indicate genera that are not differentially abundant between positive and negative sites. The dashed vertical line represents the threshold for no fold change, and the dashed horizontal line represents the threshold for statistical significance. The labels indicate the genera names along with their corresponding group (Positive or Negative).

that experience seasonal precipitation [11]. These conditions were reported in the study area during the period of this study, and therefore could explain the high abundance of *Anopheles* mosquitoes compared to other species.

The co-occurrence of both mosquito genera in the same habitat was also observed in this study. This result is in agreement with previous research that showed that the two genera can coexist in the same habitat despite having different breeding requirements [47, 48]. These findings are consistent with those found in southern Ghana during the rainy season which showed a significant presence of *Anopheles* mosquitoes in urban areas [49]. However, low abundance of *Anopheles* mosquitoes was reported during the rainy season in the Korhogo area of northern Cote d'Ivoire [50], which is in contrast to these findings. It is important to note that heavy rains and floods can wash away the proliferation sites of *Anopheles* mosquitoes, eliminating the mosquito eggs and

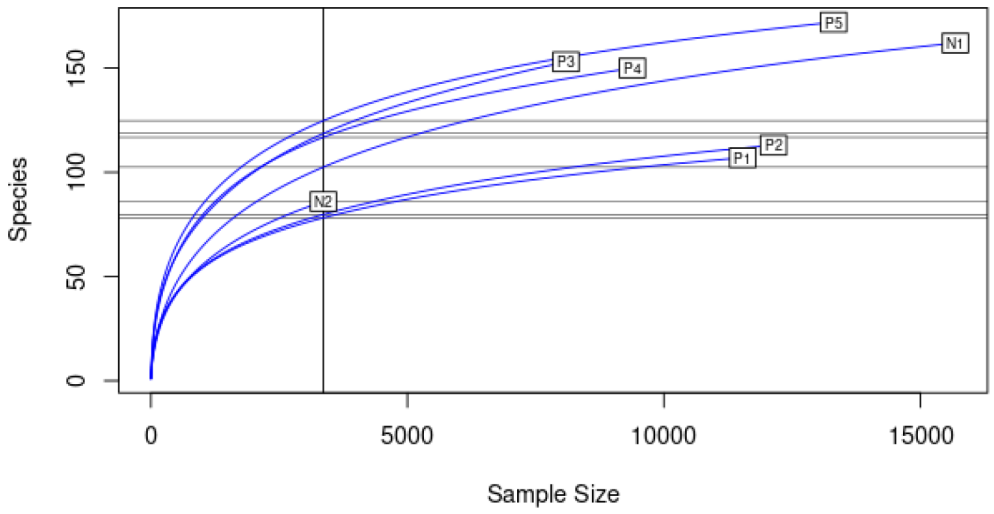

Fig. 12. Rarefaction curve: exploring bacteria diversity variation cross samples with N2 as the minimum sample size.

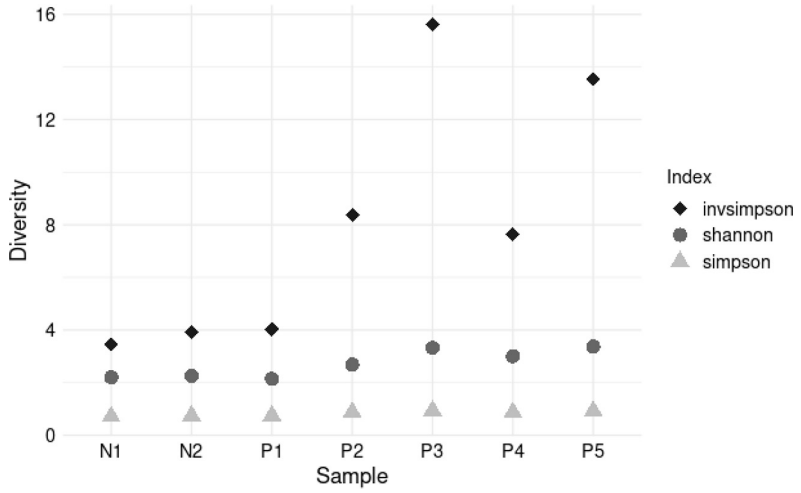

Fig. 13. Alpha diversity indices comparing sites with and without mosquito larvae. The samples are divided into two groups: negative (N1–2), which denotes sites where mosquito larvae were not detected, and positive (P1–5), representing sites where mosquito larvae were found.

larvae [51]. In this study, no floods were witnessed in the study areas during the rainy season, which may account for the high proportion of *Anopheles* mosquito larvae observed.

Both *Anopheles* and *Culex* mosquitoes have been reported in Kwale, Kenya with the abundance of *Anopheles* mosquitoes being higher in the rural areas than in the urban areas which is consistent with these findings [52]. This highlights the importance of continuous mosquito surveillance in order to assess the risk of disease transmission and develop targeted mosquito control strategies in this region. The co-occurence of *Anopheles* and *Culex* may have several implications for vector control programmes. *Anopheles* are the main malaria vectors while *Culex* mosquitoes can transmit a variety of diseases such as West Nile virus and filariasis, and therefore the co-occurence of both mosquitoes necessitates for more complex vector control protocols targeted to both *Anopheles* and *Culex* mosquitoes.

This should include a combination of measures such as use of insecticide-treated bed nets, indoor residual spraying, and larval management strategies such as draining stagnant water sources and use of biological mosquito larvicides to eliminate the mosquito larvae. Mosquito surveillance should also be enhanced in this area to determine the abundance of different mosquito species which can help in targeted mosquito control programmes and in monitoring the effectiveness of these measures. Although other malaria vectors have been reported in Kwale before, we could not conclusively ascertain why they were not identified in this study.

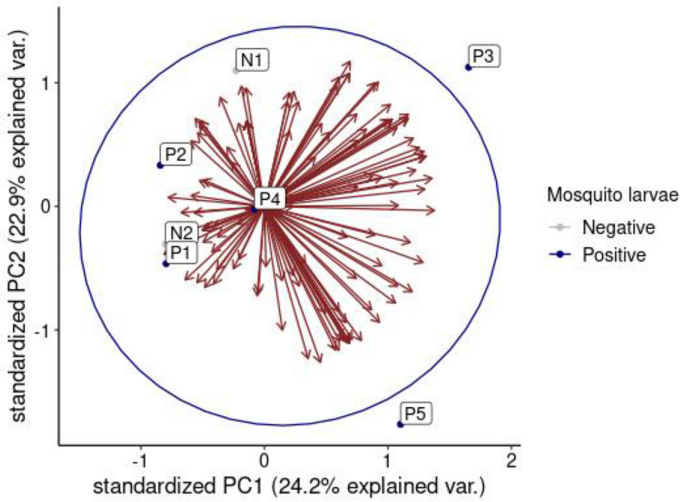

Fig. 14. Principal Component Analysis (PCA). The PCA plot shows the relationship between samples and the bacterial species distribution. N1 and N2 samples represent samples from sites without mosquito larvae while samples P1–P5 represent samples from sites with mosquito larvae.

However, we suggest that their absence could be attributed to factors such as climate change, changes in land use and human population density which are known to influence the distribution of mosquitoes in a given area.

Physical characteristics of proliferation sites

Most proliferation sites identified were natural in the form of marshes, swamp margins, edges of shallow rivers, roadside pools and animal hoof-prints. This is in agreement with past studies which found that *Anopheles* mosquitoes prefer to breed near human settlements along the edges of shallow rivers, transient roadside puddles, marsh margins, and tree holes [53–55]. Additionally, similar mosquito proliferation sites were discovered in Western Kenya and in Ethiopia [56, 57]. In contrast to these results, Hinne and others [58] categorised the majority of anopheline larval habitats found in Ghana's three main ecological zones as man-made. The low abundance of artificial mosquito proliferation sites in this region could be explained by the low levels of infrastructural development and less human activities on the environment since the local community is composed of small-scale farmers, pastoralists and traders in a rural set-up.

More sites sampled had mud substrates and were semi-permanent. Faehler and others [59] suggested that the type of soil in a larval habitat and its quality can determine the chances of survival and influence the development of *Anopheles* mosquito larvae. *Anopheles gambiae s.l.* proliferate in habitats with hydromorphic and holomorphic soil substrates due to their ability to retain water for a longer time and also to provide a conducive saline environment for growth of the mosquito larvae [60]. Semi-permanent and temporal mosquito larval habitats were also observed in Western Kenya [61]. This might be because there are fewer predators for the larvae in smaller temporary habitats than in larger permanent habitats [62].

A majority of the sites observed were exposed to full sunlight and had a shallow depth of less than 1 m with an average size of less than 10 m^2. The growth of algae, a vital source of nourishment for developing mosquito larvae, depends on the presence of sunlight in a larval habitat [62]. Sunlight also warms the water to a suitable temperature that is conducive for growth and development of the mosquito larvae [58, 63, 64]. *Anopheles* mosquitoes prefer breeding in small and shallow water bodies as those observed in this study [51, 53, 56, 57, 65]. Small and shallow water bodies are more suitable for mosquito breeding since they are less vulnerable to water currents and tides which can wash away the mosquito eggs and larvae as compared to large and deep water bodies [66]. These sites are also unsuitable for habitation by other organisms which may be competitors or predators of the mosquito larvae [67]. On the other hand, small and shallow water bodies are more likely to dry faster especially if they are not associated with a larger water body.

The most prevalent type of flora found in the sites was algae. Since algae provides the larvae with nourishment, it was positively correlated with the presence of *Anopheles* mosquito larvae at potential proliferation sites [58, 63]. The proportion of sites with high vegetation cover consisting of algae, emergent and submerged vegetation was very low and none was positive for the mosquito larvae, an indication that the presence of mosquito larvae was inversely correlated to the amount of vegetation in the water body. High levels of vegetation growth generally interfere with light penetration in the water and hence affect the growth of algae and the temperature of the water body [58, 68, 69].

Physicochemical parameters of the sites

Except for the temperature, which was lower in the positive sites than in the negative sites during the dry season, there was no other noticeable difference between the positive and negative sites during either the dry or rainy seasons. According to this study, there was no apparent difference in temperatures between the dry and rainy seasons. Notably, temperatures reported in both seasons encouraged the presence of *Anopheles* mosquito larvae, and this was consistent with the findings obtained in different places [70, 71]. Although electrical conductivity, total dissolved solids, salinity, and ammonia were significantly lower in the rainy season than in the dry season, there was no evidence that these variables could influence the mosquito larval presence or absence at the sites throughout the two seasons. However, it is notable that the levels recorded for conductivity, total dissolved solids, salinity, and ammonia in both seasons were favourable for mosquito breeding, which was comparable to the findings of a study conducted on a Nigerian university campus [72]. The low levels of these parameters during the rainy season can be attributed to the dilution of environmental surface water by rainwater [65, 71]. However, Emidi and others [47] reported a positive correlation between *Anopheles* mosquito larval abundance, salinity, and conductivity.

Mosquitoes prefer breeding in sites with alkaline pH [11, 73, 74], which concurs with the findings of this study since most of the sites had alkaline pH levels except one negative site, which was slightly acidic. High pH levels in the sites were positively correlated with dissolved oxygen, biological oxygen demand, and nutrients. Dissolved oxygen, biological oxygen demand, pH, and nutrients evaluated in this study were positively correlated and negatively correlated with turbidity. The presence of of nitrates, nitrites, ammonia, and phosphates can be attributed to the use of fertilisers containing ammonium and phosphorus in the farms since most of the sites were adjacent to the farms, while turbidity is associated with silt, mud, algae, and plant pieces [51, 73, 75, 76]. The high level of nutrients has been reported to promote excessive growth of water plants and microorganisms in the water bodies which reduces turbidity of the water making it more suitable for the proliferation of mosquitoes [77, 78].

Similar research in Western Kenya revealed that the amount of nutrients in the proliferation sites had no effect on whether *Anopheles* mosquito larvae were present or absent [79]. Excessive growth and multiplication of microorganisms in water bodies affects their biological oxygen demand and is an indicator of water pollution [76]. Finding *Anopheles gambiae* mosquitoes larvae in polluted water is uncommon since the species is believed to prefer proliferating in clean, unpolluted water in the environment. However, the presence of *Anopheles* mosquito larvae in unclean polluted water has been reported [47], which shows that the mosquitoes could have become more adapted to survive in polluted water to enhance their chances of survival. This might have an impact on how mosquitoes are distributed and abundant in the environment, which would then have an impact on how quickly and frequently malaria spreads.

Metataxonomic analysis of bacterial communities

In this study, beta diversity analysis revealed that the composition of bacterial communities was not significantly correlated with the presence or absence of mosquito larvae. Instead, the bacterial composition appeared to be influenced by the geographical locations of the sites. These findings align with previous studies that also found no association between bacterial composition and the occurrence or absence of mosquito larvae in potential proliferation sites [5, 76, 78]. However, contrasting results have been reported in other studies, where the structure of bacterial communities in mosquito larval habitats showed a correlation with the presence of mosquito larvae [80, 81]. The differences observed in these studies could not be conclusively explained, but it is hypothesised that the geographical location of the sites may impact the bacterial compositions in mosquito larval habitats. Notably, there was a difference in bacterial communities between sylvatic and domestic proliferation sites of *Aedes aegypti* in Gabon, and these bacterial communities were also found to be correlated with those present in the midgut of adult mosquitoes [82]. This similarity in bacterial community profiles suggests that the origin of bacteria in the sites may be the same for both positive and negative sites.

The findings of this study indicate that the diversity within samples varied significantly between the sites, indicating distinct and independent richness, evenness, and abundance of bacterial communities in each site. Additionally, it was observed that alpha diversities of bacteria were generally higher in the positive sites compared to the negative sites, and these differences were not associated with the evaluated physicochemical parameters. Although the observed differences in alpha diversity between the sites could not be definitively explained, factors such as age, the presence or absence of mosquito larvae, and the physical location of the sites are suggested to influence these variations. These findings are consistent with other studies that have proposed a positive correlation between bacterial abundance and the age of larval habitats [78, 83]. It is also suggested that mosquito larvae can modify bacterial communities in their habitats through feeding or the egestion of bacteria, which could explain the higher alpha diversity in the positive sites [84]. Mosquito larval activities such as feeding and excretion in the habitats may create optimal conditions for the growth of bacteria, which might otherwise go undetected in uncolonized sites where suitable growth environments are lacking [85].

Pseudomonadota, Bacteroidota, and Actinomycetota were found to dominate in all sites, accounting for 96% of the total reads. These findings align with other studies, where similar bacterial phyla were reported in mosquito larval habitats, such as Firmicutes, Pseudomonadota, and Actinomycetota [80]. These phyla were also observed in the larval habitats of *Anopheles coluzzii* and *Anopheles gambiae* in Cameroon, as well as in three Kenyan Islands in Lake Victoria [79, 86]. Another study conducted in Kenya highlighted the prevalence of these phyla in semi-natural habitats of mosquito proliferation, with Cyanobacteria being the second most abundant phylum [87]. Furthermore, the same phyla were found to be the most abundant in household water-storage containers in India [82]. The higher abundance of Pseudomonadota in mosquito breeding sites suggests that certain species within the phylum may thrive in conditions favourable for mosquito larval development. The exact mechanisms behind this positive association are not fully understood, but it is possible that Pseudomonadota bacteria contribute to nutrient availability, create favourable ecological conditions, or interact with other microorganisms to support mosquito larval growth [88]. Bacteroidota is another phylum that encompasses a wide range of bacteria with diverse functions [89]. In mosquito larval habitats, Bacteroidota bacteria have been commonly detected, and they have been associated with various ecological roles such as nutrient processing. Bacteroidota bacteria are known for their ability to degrade complex organic matter, including polysaccharides and proteins [90]. In mosquito larval habitats, where organic matter accumulates, Bacteroidota bacteria likely contribute to the breakdown of organic materials, releasing nutrients that can be utilized by mosquito larvae. Interactions with other organisms: Bacteroidota bacteria may also interact with other microorganisms present in the larval habitats. These interactions could involve mutualistic relationships, where Bacteroidota bacteria provide essential nutrients or create suitable conditions for other organisms, including mosquito larvae. Actinobacteria, including some members of the phylum Actinomycetota, are well-known producers of bioactive compounds with antimicrobial properties [91]. These bacteria can produce secondary metabolites, such as antibiotics, that can inhibit the growth of other microorganisms [92]. In mosquito larval habitats, actinobacteria may contribute to the natural defence mechanisms against pathogens and compete with other microorganisms for resources. Actinobacteria are also involved in nutrient cycling processes [93]. They play a crucial role in the decomposition of organic matter, releasing essential nutrients that can be utilized by mosquito larvae or other organisms in the habitat.

At the class level, the most common bacterial groups detected were Gammaproteobacteria, Bacteroidia, Alphaproteobacteria, and Actinobacteria, accounting for a total of 76.67% of all bacteria detected. Gammaproteobacteria was consistently found to be the most prevalent class of bacteria in various studies [94, 95]. Although Bacilli were not among the commonly detected groups in the current study, they were reported as one of the most abundant classes in previous research [94]. Alphaproteobacteria and Cyanobacteria were found to be the most common classes associated with semi-natural mosquito habitats in Kenya [87]. Betaproteobacteria and Alphaproteobacteria were identified as the most abundant bacterial classes in household water-storage containers in India [82], while a study on Kenyan Islands of Lake Victoria found Betaproteobacteria to be the most common class in mosquito larval habitats [79]. Other frequently found classes included Verrucomicrobiae, Planctomycetes, Microgenomatia, Gemmatimonadetes, Acidimicrobiia, Cyanobacteriia, Chloroflexia, and Saccharimonadia.

The most frequently observed bacterial orders were Burkholderiales, Flavobacteriales, Chitinophagales, Sphigomonadales, Micrococcales, Rhizobiales, Sphigobacteriales, Enterobacterales, Frankiales, and Cytophagales. Burkholderiales and Cytophagales have been considered indicator species in water samples collected from the breeding sites of *Anopheles darlingii* [94]. Many of the families detected in this study have previously been associated with *Anopheles* mosquitoes, with the most abundant families being Commamonadaceae, Flavobacteriaceae, and Chitinophagaceae [79, 94, 96, 97]. In another study, it was evident that the most abundant families in the larval habitats of *Aedes albopictus* in Italy were Sphingobacteriaceae, Spirosomaceae, Chitinophagaceae, Cellvibrionaceae, Burkholderiaceae, Caulobacteraceae, Planococcaceae, Cytophagaceae, and Blastocatellaceae [98].

The findings from our study provide valuable insights into the association between bacterial populations and positive environmental conditions in relation to mosquito oviposition and habitat selection. The presence of certain bacterial genera, such as *Vibrio* and *Cutibacterium,* with significantly higher abundance at positive sites compared to negative sites suggests their potential role in creating favourable conditions for mosquito oviposition. Conversely, the genus *Methylorubum* exhibited lower abundance in the positive sites compared to the negative site environment. However, studies conducted in Ethiopian mosquito proliferation sites identified *Bacillus, Pseudomonas, Micrococcus,* and *Serratia* as the dominant genera in bacterial genera [99]. Similarly, another study reported the presence of *Rubrivivax, Hydrogenophaga, Rhodobacter, Pseudomonas,* and *Flavobacterium* in mosquito larval habitats in Western Kenya [79]. These bacteria were also discovered in the larval habitats of *Aedes aegypti* associated with domestic water storage containers in Thailand and Laos [69]. The bacterial communities present in mosquito larval habitats may serve as indicator species for high-potential proliferation sites, affecting larval survival, adult fitness, vector abundance, distribution, and ultimately impacting malaria transmission [100]. The findings from Sumba and others [101] provide additional evidence regarding the role of isolated bacteria and their associated volatiles in mosquito oviposition and habitat selection. In this study, various bacterial species, including unclassified *Firmicutes, Aeromonas, Pasteurella, Pseudomonas, Vibrio, Acinetobacter*, and Enterobacteriaceae, were isolated from soil collected beneath oviposition sites and larval habitats. These isolated bacteria were found to restore the attractiveness or stimulant properties of sterile soils, but not filtered distilled water. This result suggests that the presence of microorganisms or volatile organic compounds (VOCs) in water is crucial for mosquitoes to utilize kairomones, which are chemical cues that provide information about their environment [102].

The information obtained about bacterial communities in mosquito larval habitats can be used to design eco-friendly mosquito control methods. For instance, investigating the potential larvicidal effects of specific bacterial communities associated with the sites could help in designing targeted interventions. Disrupting the growth of certain bacterial communities predominant in mosquito larval habitats can render those habitats unsuitable for mosquito breeding. Again, monitoring the bacterial communities in larval habitats can aid in identifying potential breeding habitats and focusing vector control interventions, as the bacterial community can serve as an indicator of the mosquito population present. Considering that bacterial communities in larval habitats may contribute to the vectoral capacity of mosquitoes, these findings provide valuable information for researchers investigating disease transmission mechanisms and developing strategies to curb malaria transmission.

Limitations of the study

The scope of this study was limited to a specific region within the Kenyan Coast, making it unclear whether the ecological factors associated with mosquito larval habitats can be replicated in different environments. Although other malaria vectors were previously reported in Kwale, this study was unable to conclusively ascertain why they were not identified. Additionally, this study did not examine the specific roles of certain bacteria that were identified as dominant in the larval habitats, nor did it evaluate other factors like interspecific competition and predation in influencing the oviposition response and survival of *Anopheles* mosquitoes.

CONCLUSION

In this study, we conducted a comprehensive analysis of bacterial communities in mosquito larval habitats, shedding light on their composition and potential implications for mosquito population dynamics. Our findings revealed the predominance of Pseudomonadota, Bacteroidota, and Actinomycetota across all sites, underscoring their significant roles in these ecological niches. Notably, we identified *Vibrio* and *Cutibacterium* as being significantly more abundant in positive sites compared to negative sites,

while *Methylorubum* exhibited lower abundance in the positive sites. These differential abundance patterns suggest a potential association between these bacterial genera and favourable environmental conditions for mosquito proliferation.

Furthermore, our results unveiled a positive correlation between the presence of *Anopheles* mosquito larvae and the rainy season, as well as bacterial abundance. Conversely, we observed negative correlations with several physicochemical parameters, including electrical conductivity, total dissolved solids, salinity, and ammonia. These findings highlight the intricate interplay between environmental factors, bacterial communities, and mosquito oviposition and population growth. Importantly, our study highlighted the influence of temporal and geographical factors on the structure of bacterial communities in mosquito larval habitats. By comparing our findings with previous studies, we gained valuable insights into the dynamic nature of bacterial compositions and their associations with mosquito populations. This knowledge holds promise for predicting potential proliferation sites based on the physicochemical properties and bacterial community compositions of environmental water samples.

Overall, our study advances the understanding of bacterial communities in mosquito habitats, providing crucial information for the development of targeted mosquito control strategies and the elucidation of mechanisms underlying mosquito-borne disease transmission. Moving forward, investigations focusing on the temporal dynamics of bacterial communities and their impact on mosquito populations will further enhance our ability to predict and mitigate mosquito-borne diseases, ultimately contributing to improved public health outcomes.

Funding information
This work was co-funded by the International Foundation for Science (IFS) and the Organization for Prohibition of Chemical Weapons (OPCW) grant number I-1-F-6278–1 and the National Research Fund (NRF), Kenya, grant number NRF/R/2016/2017 1ST CALL/31.

Acknowledgements
The authors wish to thank Mr. Festus Yaah, Mr. Oliver Ochola, Mathews Wafula, Joel Gatagwu and Joseph Khamisi for their support during the sample collection and laboratory experiments.

Author contributions
J.M., J.K.N., R.M.N. and M.V.O. contributed to conceptualization and design of the study. Sample collection was done by K.O.O., J.M. and J.M. Laboratory analysis was conducted by J.M., J.M.G. and K.O.O. while S.M. Mwamburi, J.M., and K.O.O. carried out statistical analysis, interpreted the result of the study and wrote the first draft of the manuscript. All authors assisted in review, reading and approved the final manuscript.

Conflicts of interest
The authors declare no conflicts of interest.

Ethical statement
Before sampling, the owners of private land with mosquito breeding grounds provided informed consent. Research permit was obtained from the National Council for Science, Technology, and Innovation (Licence No: NACOSTI/P/21/10048). Because the study did not include human samples, national parks or endangered species, no special permits were required.

References

1. Mutinda J, Mwamburi SM, Oduor KO, Omolo MO, Ntabo RM, *et al*. Metagenomic profiles of bacterial communities and environmental factors associated with proliferation of malaria vector mosquitoes within the Kenyan Coast. *Figshare* 2023. DOI: 10.6084/m9.figshare.23574792.v1.
2. Coll-Seck AM. Who controls malaria control? *Nature* 2010;466:186–187.
3. Mwangangi JM, Mbogo CM, Muturi EJ, Nzovu JG, Githure JI, *et al*. Spatial distribution and habitat characterisation of *Anopheles* larvae along the Kenyan coast. *J Vector Borne Dis* 2007;44:44–51.
4. Minakawa N, Dida GO, Sonye GO, Futami K, Njenga SM. Malaria vectors in Lake Victoria and adjacent habitats in western Kenya. *PLoS One* 2012;7:e32725.
5. Munga S, Vulule J, Kweka EJ. Response of *Anopheles gambiae* s.l. (Diptera: Culicidae) to larval habitat age in western Kenya highlands. *Parasit Vectors* 2013;6:1–6.
6. Gouagna LC, Rakotondranary M, Boyer S, Lempérière G, Dehecq J-S, *et al*. Abiotic and biotic factors associated with the presence of *Anopheles arabiensis* immatures and their abundance in naturally occurring and man-made aquatic habitats. *Parasit Vectors* 2012;5.
7. Huang J-X, Xia Z-G, Zhou S-S, Pu X-J, Hu M-G, *et al*. Spatio-temporal analysis of malaria vectors in national malaria surveillance sites in China. *Parasit Vectors* 2015;8.
8. Mala AO, Irungu LW. Factors influencing differential larval habitat productivity of *Anopheles gambiae* complex mosquitoes in a western Kenyan village. *J Vector Borne Dis* 2011;48:52–57.
9. Sumba LA, Ogbunugafor CB, Deng AL, Hassanali A. Regulation of oviposition in *Anopheles gambiae* s.s.: role of inter- and intraspecific signals. *J Chem Ecol* 2008;34:1430–1436.
10. Vanderberg JP. Reflections on early malaria vaccine studies, the first successful human malaria vaccination, and beyond. *Vaccine* 2009;27:2–9.
11. Kweka EJ, Zhou G, Munga S, Lee M-C, Atieli HE, *et al*. Anopheline Larval habitats seasonality and species distribution: a prerequisite for effective targeted Larval habitats control programmes. *PLoS One* 2012;7:e52084.
12. Lengeler C. Insecticide-treated nets for malaria control: real gains. *Bull World Health Organ* 2004;82:84.
13. Pluess B, Tanser FC, Lengeler C, Sharp BL. Indoor residual spraying for preventing malaria. Cochrane Infectious Diseases Group editor. Cochrane Database of Systematic Reviews; 2010. https://doi.wiley.com/10.1002/14651858.CD006657.pub2 [accessed 17 February 2023].
14. Ferguson HM, Dornhaus A, Beeche A, Borgemeister C, Gottlieb M, *et al*. Ecology: a prerequisite for malaria elimination and eradication. *PLoS Med* 2010;7:e1000303.
15. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, *et al*. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 2014;371:411–423.

16. Fillinger U, Lindsay SW. Larval source management for malaria control in Africa: myths and reality. *Malar J* 2011;10.
17. Ndenga BA, Simbauni JA, Mbugi JP, Githeko AK, Fillinger U. Productivity of malaria vectors from different habitat types in the western Kenya highlands. *PLoS One* 2011;6:e19473.
18. Asale A, Kussa D, Girma M, Mbogo C, Mutero CM. Community based integrated vector management for malaria control: lessons from three years' experience (2016–2018) in Botor-Tolay district, southwestern Ethiopia. *BMC Public Health* 2019;19.
19. Arisco NJ, Rice BL, Tantely LM, Girod R, Emile GN, *et al*. Variation in *Anopheles* distribution and predictors of malaria infection risk across regions of Madagascar. *Malar J* 2020;19:348.
20. Şengül Demirak MŞ, Canpolat E. Plant-based bioinsecticides for mosquito control: impact on insecticide resistance and disease transmission. *Insects* 2022;13:162.
21. Wilson AL, Courtenay O, Kelly-Hope LA, Scott TW, Takken W, *et al*. The importance of vector control for the control and elimination of vector-borne diseases. *PLoS Negl Trop Dis* 2020;14:e0007831.
22. Howard AFV, Omlin FX. Abandoning small-scale fish farming in western Kenya leads to higher malaria vector abundance. *Acta Trop* 2008;105:67–73.
23. Fillinger U, Ndenga B, Githeko A, Lindsay SW. Integrated malaria vector control with microbial larvicides and insecticide-treated nets in western Kenya: a controlled trial. *Bull World Health Organ* 2009;87:655–665.
24. Ototo EN, Mbugi JP, Wanjala CL, Zhou G, Githeko AK, *et al*. Surveillance of malaria vector population density and biting behaviour in western Kenya. *Malar J* 2015;14:244.
25. Nicholas K, Bernard G, Bryson N, Mukabane K, Kilongosi M, *et al*. Abundance and distribution of Malaria vectors in various aquatic habitats and land use types in Kakamega County, highlands of Western Kenya. *Ethiop J Health Sci* 2021;31:247–256.
26. Mwangangi JM, Mbogo CM, Orindi BO, Muturi EJ, Midega JT, *et al*. Shifts in malaria vector species composition and transmission dynamics along the Kenyan coast over the past 20 years. *Malar J* 2013;12:13.
27. Karisa J, Ominde K, Muriu S, Munyao V, Mwikali K, *et al*. Malaria vector bionomics in Taita-Taveta County, coastal Kenya. *Parasit Vectors* 2022;15.
28. Naing L, Winn T, Rusli B. Practical issues in calculating the sample size for prevalence studies. *Arch orofac Sci* 2006;1:9–14.
29. Coetzee M. Key to the females of Afrotropical *Anopheles* mosquitoes (Diptera: Culicidae). *Malar J* 2020;19.
30. Minas K, McEwan NR, Newbold CJ, Scott KP. Optimization of a high-throughput CTAB-based protocol for the extraction of qPCR-grade DNA from rumen fluid, plant and bacterial pure cultures. *FEMS Microbiol Lett* 2011;325:162–169.
31. Lee PY, Costumbrado J, Hsu CY, Kim YH. Agarose gel electrophoresis for the separation of DNA fragments. *J Vis Exp* 2012:3923.
32. Desjardins P, Conklin D. NanoDrop microvolume quantitation of nucleic acids. *J Vis Exp* 2010:2565.
33. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, *et al*. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 2008;74:2461–2470.
34. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, *et al*. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581–583.
35. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257.
36. Cameron ES, Schmidt PJ, Tremblay BJM, Emelko MB, Müller KM. Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. *Sci Rep* 2021;11:22302.
37. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, *et al*. Package 'Vegan community Ecology package, version. *Community ecology package, version* 2013;2:1–295.
38. Chao A, Chazdon RL, Colwell RK, Shen T-J. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* 2006;62:361–371.
39. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, *et al*. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–7541.
40. Kim B-R, Shin J, Guevarra R, Lee JH, Kim DW, *et al*. Deciphering diversity indices for a better understanding of microbial communities. *J Microbiol Biotechnol* 2017;27:2089–2093.
41. Addinsoft X. *Data Analysis and Statistics Software for Microsoft Excel*. Addinsoft: Paris, France; 2015.
42. Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Austral Ecol* 1993;18:117–143.
43. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;26:32–46.
44. Chiani M. Distribution of the largest root of a matrix for Roy's test in multivariate analysis of variance. *J Multivar Anal* 2016;143:467–471.
45. Guo B, Yuan Y. A comparative review of methods for comparing means using partially paired data. *Stat Methods Med Res* 2017;26:1323–1340.
46. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 1952;47:583–621.
47. Emidi B, Kisinza WN, Mmbando BP, Malima R, Mosha FW. Effect of physicochemical parameters on *Anopheles* and *Culex* mosquito larvae abundance in different breeding sites in a rural setting of Muheza, Tanzania. *Parasit Vectors* 2017;10:304.
48. Kudom AA. Larval ecology of *Anopheles coluzzii* in Cape Coast, Ghana: water quality, nature of habitat and implication for larval control. *Malar J* 2015;14:447.
49. Mattah PAD, Futagbi G, Amekudzi LK, Mattah MM, de Souza DK, *et al*. Diversity in breeding sites and distribution of *Anopheles* mosquitoes in selected urban areas of southern Ghana. *Parasit Vectors* 2017;10:25.
50. Zogo B, Koffi AA, Alou LPA, Fournet F, Dahounto A, *et al*. Identification and characterization of *Anopheles* spp. breeding habitats in the Korhogo area in northern Côte d'Ivoire: a study prior to a Bti-based larviciding intervention. *Parasit Vectors* 2019;12:146.
51. Dejenie T, Yohannes M, Assmelash T. Characterization of mosquito breeding sites in and in the vicinity of tigray microdams. *Ethiop J Health Sci* 2011;21:57–66.
52. Karungu S, Atoni E, Ogalo J, Mwaliko C, Agwanda B, *et al*. Mosquitoes of etiological concern in Kenya and possible control strategies. *Insects* 2019;10:173.
53. Shililu J, Ghebremeskel T, Seulu F, Mengistu S, Fekadu H, *et al*. Larval habitat diversity and ecology of anopheline larvae in Eritrea. *J Med Entomol* 2003;40:921–929.
54. Yohannes M, Haile M, Ghebreyesus TA, Witten KH, Getachew A, *et al*. Can source reduction of mosquito larval habitat reduce malaria transmission in Tigray, Ethiopia? *Trop Med Int Health* 2005;10:1274–1285.
55. Omlin FX, Carlson JC, Ogbunugafor CB, Hassanali A. *Anopheles gambiae* exploits the treehole ecosystem in western Kenya: a new urban malaria risk? *Am J Trop Med Hyg* 2007;77:264–269.
56. Imbahale SS, Mweresa CK, Takken W, Mukabana WR. Development of environmental tools for anopheline larval control. *Parasit Vectors* 2011;4:130.
57. Mereta ST, Yewhalaw D, Boets P, Ahmed A, Duchateau L, *et al*. Physico-chemical and biological characterization of anopheline mosquito larval habitats (Diptera: Culicidae): implications for malaria control. *Parasit Vectors* 2013;6:320.
58. Hinne IA, Attah SK, Mensah BA, Forson AO, Afrane YA. Larval habitat diversity and *Anopheles* mosquito species distribution in different ecological zones in Ghana. *Parasit Vectors* 2021;14:193.

59. Pfaehler O, Oulo DO, Gouagna LC, Githure J, Guerin PM. Influence of soil quality in the larval habitat on development of *Anopheles gambiae* Giles. *J Vector Ecol* 2006;31:400–405.
60. Ndiaye A, Niang EHA, Diène AN, Nourdine MA, Sarr PC, *et al*. Mapping the breeding sites of *Anopheles gambiae* s. l. in areas of residual malaria transmission in central western Senegal. *PLoS One* 2020;15:e0236607.
61. Kweka EJ, Munga S, Himeidan Y, Githeko AK, Yan G. Assessment of mosquito larval productivity among different land use types for targeted malaria vector control in the western Kenya highlands. *Parasit Vectors* 2015;8:356.
62. Gimnig JE, Ombok M, Kamau L, Hawley WA. Characteristics of Larval Anopheline (Diptera: Culicidae) habitats in Western Kenya. *J Med Entomol* 2001;38:282–288.
63. Afrane YA, Lawson BW, Githeko AK, Yan G. Effects of microclimatic changes caused by land use and land cover on duration of gonotrophic cycles of *Anopheles gambiae* (Diptera: Culicidae) in western Kenya highlands. *J Med Entomol* 2005;42:974–980.
64. WHO Malaria Policy Advisory Committee and Secretariat. Malaria Policy Advisory Committee to the WHO: conclusions and recommendations of September 2013 meeting. *Malar J* 2013;12:456.
65. Olusi TA, Simon-Oke IA, Akeju AV. Composition, habitat preference and seasonal variation of malaria vector larval and pupa stage in Akure North Local Government Area of Ondo State, Nigeria. *Bull Natl Res Cent* 2021;45.
66. Oyewole I, Momoh O, Anyasor G, Ogunnowo A, Ibidapo C, *et al*. Physico-chemical characteristics of *Anopheles* breeding sites: impact on fecundity and progeny development. *Afr J Environ Sci Tech* 2009;3:12.
67. Roux O, Robert V. Larval predation in malaria vectors and its potential implication in malaria transmission: an overlooked ecosystem service? *Parasit Vectors* 2019;12:217.
68. Williams J, Pinto J. *Training Manual on Malaria Entomology for Entomology and Vector Control Technicians (Basic Level)*. USAID Washington, DC, 2012.
69. Dida GO, Anyona DN, Abuom PO, Akoko D, Adoka SO, *et al*. Spatial distribution and habitat characterization of mosquito species during the dry season along the Mara River and its tributaries, in Kenya and Tanzania. *Infect Dis Poverty* 2018;7:2.
70. Christiansen-Jucht C, Parham PE, Saddler A, Koella JC, Basáñez MG. Temperature during larval development and adult maintenance influences the survival of *Anopheles gambiae* s.s. *Parasit Vectors* 2014;7:489.
71. Ajani FO, Ofoezie EI, Adeoye N, Eze HI. *Anopheles* mosquito larval production and association of water bodies with malaria transmission in Ife and Ilesa areas of Osun state, Nigeria. *J Environ Sci Toxicol Food Technol* 2018;12:82–88.
72. Amawulu E, Commander T, Amaebi A. Effect of physicochemical parameters on mosquito Larva population in the Niger Delta University Campuses, Bayelsa State, Nigeria. *Int J of Zool Res* 2020;16:63–68.
73. Kwasi B. Biology A, Kumasi T. Physico-chemical assessment of mosquito breeding sites from selected mining communities at the Obuasi municipality in Ghana. *J Environ Earth Sci* 2012;2:123–130.
74. Tirado DF, Almanza-Vasquez E, Almanza-Meza EJ, Acevedo-Correa D, González-Morelo KJ. Larval development of mosquitoes and pH of different reservoirs in the city of Cartagena de Indias (Colombia). *IJET* 2017;9:4137–4140.
75. Oguche Donatus O, Auta IK, Ibrahim B, Yayock HC, Johnson O. Breeding sites characteristics and mosquito abundance in some selected locations within Kaduna metropolis. *FJS* 2023;6:70–75.
76. Ndaruga AM, Ndiritu GG, Gichuki NN, Wamicha WN. Impact of water quality on macroinvertebrate assemblages along a tropical stream in Kenya. *African J Ecol* 2004;42:208–216.
77. Victor TJ, Reuben R. Effects of organic and inorganic fertilisers on mosquito populations in rice fields of southern India. *Med Vet Entomol* 2000;14:361–368.
78. Mutero CM, Ng'ang'a PN, Wekoyela P, Githure J, Konradsen F. Ammonium sulphate fertiliser increases larval populations of *Anopheles arabiensis* and culicine mosquitoes in rice fields. *Acta Tropica* 2004;89:187–192.
79. Onchuru TO, Ajamma YU, Burugu M, Kaltenpoth M, Masiga D, *et al*. Chemical parameters and bacterial communities associated with larval habitats of *Anopheles*, *Culex* and *Aedes* mosquitoes (Diptera: Culicidae) in western Kenya. *Int J Trop Insect Sci* 2016;36:146–160.
80. Xu Y, Chen S, Kaufman MG, Maknojia S, Bagdasarian M, *et al*. Bacterial community structure in tree hole habitats of *Ochlerotatus triseriatus*: influences of larval feeding. *J Am Mosq Control Assoc* 2008;24:219–227.
81. Dada N, Jumas-Bilak E, Manguin S, Seidu R, Stenström TA, *et al*. Comparative assessment of the bacterial communities associated with *Aedes aegypti* larvae and water from domestic water storage containers. *Parasit Vectors* 2014;7:391.
82. Nilsson LKJ, Sharma A, Bhatnagar RK, Bertilsson S, Terenius O. Presence of *Aedes* and *Anopheles* mosquito larvae is correlated to bacteria found in domestic water-storage containers. *FEMS Microbiol Ecol* 2018;94.
83. Dickson LB, Jiolle D, Minard G, Moltini-Conclois I, Volant S, *et al*. Carryover effects of larval exposure to different environmental bacteria drive adult trait variation in a mosquito vector. *Sci Adv* 2017;3:e1700585.
84. Eneh LK, Fillinger U, Borg Karlson AK, Kuttuva Rajarao G, Lindh J. *Anopheles arabiensis* oviposition site selection in response to habitat persistence and associated physicochemical parameters, bacteria and volatile profiles. *Med Vet Entomol* 2019;33:56–67.
85. Scolari F, Sandionigi A, Carlassara M, Bruno A, Casiraghi M, *et al*. Exploring changes in the microbiota of *Aedes albopictus*: comparison among breeding site water, Larvae, and adults. *Front Microbiol* 2021;12:624170.
86. Gimonneau G, Tchioffo MT, Abate L, Boissière A, Awono-Ambéné PH, *et al*. Composition of *Anopheles coluzzii* and *Anopheles gambiae* microbiota from larval to adult stages. *Infect Genet Evol* 2014;28:715–724.
87. Wang Y, Gilbreath TM 3rd, Kukutla P, Yan G, Xu J. Dynamic gut microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya. *PLoS One* 2011;6:e24767.
88. Yang N, Zhang C, Wang L, Li Y, Zhang W, *et al*. Nitrogen cycling processes and the role of multi-trophic microbiota in dam-induced river-reservoir systems. *Water Res* 2021;206:117730.
89. Munoz R, Teeling H, Amann R, Rosselló-Móra R. Ancestry and adaptive radiation of Bacteroidetes as assessed by comparative genomics. *Syst Appl Microbiol* 2020;43:126065.
90. Fernández-Álvarez C, Santos Y. Identification and typing of fish pathogenic species of the genus *Tenacibaculum*. *Appl Microbiol Biotechnol* 2018;102:9973–9989.
91. van Bergeijk DA, Terlouw BR, Medema MH, van Wezel GP. Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nat Rev Microbiol* 2020;18:546–558.
92. Shankar S, Segaran G, Sathiavelu M. Antimicrobial metabolites from endophytic microorganisms and its mode of action. In: *Biocontrol Mechanisms of Endophytic Microorganisms*. 2022. pp. 75–88.
93. Ul-Hassan A, Wellington EMH. Actinobacteria. In: *Encyclopedia of Microbiology*, Third Edition. Elsevier, 2009. pp. 25–44.
94. Nilsson LKJ, de Oliveira MR, Marinotti O, Rocha EM, Håkansson S, *et al*. Characterization of bacterial communities in breeding waters of *Anopheles darlingi* in Manaus in the amazon basin malaria-endemic area. *Microb Ecol* 2019;78:781–791.
95. Bascuñán P, Niño-Garcia JP, Galeano-Castañeda Y, Serre D, Correa MM. Factors shaping the gut bacterial community assembly in two main Colombian malaria vectors. *Microbiome* 2018;6:148.
96. Gusmão DS, Santos AV, Marini DC, Russo E de S, Peixoto AMD, *et al*. First isolation of microorganisms from the gut diverticulum of *Aedes aegypti* (Diptera: Culicidae): new perspectives

for an insect-bacteria association. *Mem Inst Oswaldo Cruz* 2007;102:919–924.

97. Dinparast Djadid N, Jazayeri H, Raz A, Favia G, Ricci I, *et al*. Identification of the midgut microbiota of *An. stephensi* and *An. maculipennis* for their application as a paratransgenic tool against malaria. *PLoS One* 2011;6:e28484.
98. Coon KL, Brown MR, Strand MR. Mosquitoes host communities of bacteria that are essential for development but vary greatly between local habitats. *Mol Ecol* 2016;25:5806–5826.
99. Eyob C, Dawit A. Bacterial populations of mosquito breeding habitats in relation to maize pollen in Asendabo, south western Ethiopia. *Afr J Microbiol Res* 2017;11:55–64.
100. Ranasinghe HAK, Amarasinghe LD. Naturally occurring microbiota associated with mosquito breeding habitats and their effects on mosquito Larvae. *Biomed Res Int* 2020;2020:4065315.
101. Sumba LA, Guda TO, Deng AL, Hassanali A, Beier JC, *et al*. Mediation of oviposition site selection in the African malaria mosquito *Anopheles gambiae* (Diptera: Culicidae) by semiochemicals of microbial origin. *Int J Trop Insect Sci* 2004;24:260–265.
102. Lindh JM, Kännaste A, Knols BGJ, Faye I, Borg-Karlson AK. Oviposition responses of *Anopheles gambiae* s.s. (Diptera: Culicidae) and identification of volatiles from bacteria-containing solutions. *J Med Entomol* 2008;45:1039–1049.

Peer review history

VERSION 3

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000606.v3.3

Jana Katharina Schniete; Leibniz Universitat Hannover Naturwissenschaftliche Fakultat, Institut fuer Mikrobiologie, Herrenhaeuser Strasse 2, Geb 4104, GERMANY, Hannover

Date report received: 31 July 2023
Recommendation: Accept

Comments: Thank you for addressing all the raised concerns and comments from the reviewers and myself.

SciScore report

https://doi.org/10.1099/acmi.0.000606.v3.1

iThenticate report

https://doi.org/10.1099/acmi.0.000606.v3.2

Author response to reviewers to Version 2

Title

Profiles of Bacterial communities and environmental factors associated with proliferation of malaria vector mosquitoes within the Kenyan Coast

Actions taken on Editor's Comments

Editor's Comments/suggestions	Changes Made/Action taken
In the Abstract Results and Discussion section I think it would be better to skip the actual numbers in parenthesis and just highlight the main findings, for exact details and numbers the reader can go to the main text	All numbers have been omitted in this section - lines 40-43
Figure 11 you have used the term differentially expressed, this usually refers to genes and is not the right term to use here, please amend it in the figure and the manuscript text to abundance changes/ or differential abundance of bacterial genera	The term has been changed to differentially abundant – lines 415-416, figure 11, and lines 438-444

VERSION 2

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000606.v2.3

Jana Katharina Schniete; Leibniz Universitat Hannover Naturwissenschaftliche Fakultat, Institut fuer Mikrobiologie, Herrenhaeuser Strasse 2, Geb 4104, GERMANY, Hannover

Date report received: 14 July 2023
Recommendation: Minor Amendment

Comments: The work presented is clear and the arguments well formed. This study would be a valuable contribution to the existing literature. This is a study that would be of interest to the field and community. Thank you for submitting a revised version of your manuscript following the assessment of the reviewers, I have only found some minor things that I would like you to address: In the Abstract Results and Discussion section I think it would be better to skip the actual numbers in parenthesis and just highlight the main findings, for the exact details and numbers the reader can go to the main text. Figure 11 you have used the term differentially expressed, this usually refers to genes and is not the right term to use here, please amend it in the figure and the manuscript text to abundance changes/ or differential abundance of bacterial genera

SciScore report

https://doi.org/10.1099/acmi.0.000606.v2.1

iThenticate report

https://doi.org/10.1099/acmi.0.000606.v2.2

Author response to reviewers to Version 1

Title
Metagenomic Profiles of Bacterial communities and environmental factors associated with proliferation of malaria vector mosquitoes within the Kenyan Coast
Actions taken on Reviewers Comments
Reviewer 1

Reviewer	Reviewers Comments/suggestions	Changes Made/Action taken
1	The raw data used for the other elements of this study have not been provided. Whilst the manuscript includes summary data the raw data has not been provided in either supplementary information nor deposited on a relevant database. Authors should consider including this data or a statement to explain that the data is available upon request.	The raw data was deposited in the Microbiology Society's Figshare account and cited in the data summary as advised (Line 62-64)
1	**Presentation of results**;	

1	However, it is also important to investigate whether the proportion of positive sites is different in the dry and rainy seasons. From the numbers presented it is clear that the proportion of positive sites is higher in the rainy season (as would be expected) but no stats for this seems to have been presented. You should test for a significant difference in positive site proportion between the dry and rainy season.	The Statistical significance test was done and presented in the manuscript (Line 310-311)
1	Lines 304-305 states that the results suggest that more habitats were suitable for breeding Anopheles than Culex. It would be important to understand how this fits into the mosquito abundance in this area. Could it not just be that there are more Anopheles than Culex. Some discussion on this, probably best placed within the discussion, would benefit the interpretation of these results.	A discussion on the possible reason why *Anopheles*mosquitoes were more encountered compared to other mosquitoes has been provided (Line 491-506)
1	Figure 2A displays the % of positive and negative sites in the dry season, rainy season, and overall. A stacked barchart is more appropriate for displaying proportions and would make it easier to interpret whether the proportion is different in the different seasons. I'm also not sure that you should have error bars on these plots. Has the percentage positive and negative not been calculated across all sample sites to give overall percentages?	The figure has been corrected and data presented as a stacked barchart (Figure 2)
1	How were the error bars for figure 2 calculated? They seem to be very similar across different bars, are they correct? For figure 2C the error bars are putting the relative abundance above 100 and below 0, error bars should not go below 0 or above 100. There also appear to be error bars for some of the characteristics (type, vegetation) where there is no actual bar.	The figure has been changed to stacked barcharts (Figure 5)
1	For readability I would recommend Figure 2C being in its own figure. This would allow you to increase readability of the figure and flow of the main text. Stacked bar charts might also work better here to reduce the number of individual bars and increase readability.	Figure has been presented own its own (Figure 5)

1	Please check the data in figure 2C. For example, shouldn't the % of artificial negative sites and natural negative sites add to 100%? They don't appear to, and there are similar issues throughout this figure.	Figure was corrected (Figure 5)
1	Figure legends require more detail added throughout to ensure figures can be interpreted and are stand alone. Where statistical analysis has been conducted that should be included in the legend to highlight significant difference (or no differences).	Figure legends were improved and statistical significance included where possible
1	Table 1 is well presented. It could benefit from including asterisks to indicate where differences are significant.	The table was improved and asterisks included on the variables that were significantly different (Table 2).
1	Have the physiochemical parameters in table 1 also been tested for significance between the positive and negative sites. If so, please include the results. If not, please complete this analysis as it is important to the study. Identifying which physicochemical characters differ between positive and negative sites would provide an important insight into the factors associated with mosquito oviposition choice (as discussed in your introduction).	Comparison of physicochemical parameters between positive and negative sites has been performed and results provided (Table 1)
1	The refraction curve (figure 3) is well presented and explained in the main text (lines 341-343). I wonder if it would be more appropriate in the methods rather than results.	Rarefaction was redone to allow for standardization the comparison of biodiversity by estimating species richness or diversity at the same level of sampling effort, irrespective of sample size (figure 12)
1	The number of reads generated, before and after QC trimming (lines 377 - 381) might be better presented in a table.	Data has been presented in a table (Table 3)
1	Figures 7-11 presenting the results of the taxonomic assignment are very well presented. However in the results section they are just listed and not specifically referred to in the main text of the results section. Please ensure that the main text refers directly to relevant figures pointing out key message in each figure. This would make it easier for the reader to interpret these results.	The figures have been referenced in the text (Line 395-415)

1	It is useful to know that more bacterial families in positive sites compared to negative sites (as explained in lines 389-391) however figure 12 does not include labelling to enable the reader to understand which families specifically are found in each site. Figure should be edited to include family names.	The figure was a phylogenetic tree to describe the evolutionary patterns of families in positive and negative sites, the figure has been omitted and instead we have focused more on alpha and beta diversity plots to describe the differential abundance of bacteria in different sites (Figures 11, 12, 13 1nd 14)
1	Lines 391 - 393 states that 19 genera were more significantly more abundant in positive sites. However, no p value has been provided. Has this been tested statistically? If not please avoid using the term significant. It would be useful to present this analysis for all genera (not just the significant 19) as supplementary information.	The level of statistical significance has been tested in the volcano plot represented in Figure 11
1	The size of the points on figure 13 could be reduced to improve readability. Please explain what is meant by N/A on figure 13 in the figure legend.	The figure was changed and replaced with a more clear volcano plot in figure 11 and the description of the figure enhanced in the figure legend
1	The abstract is very long and could be condensed to more concisely convey the key findings of the study. The background section of the abstract for example could be significantly cut down. The results section could also be cut down, just focusing on key results rather than each individual result. This would make the key take home message more clear in the abstract.	The abstract has been greatly condensed to just focus on the main outcomes of the study (Line 32 - 56)
1	When discussing the presence of Anopheles and Culex reported in this study (lines 427 - 439) it is important to give some information on known distribution/occurrence of these mosquitoes in the local area.	The discussion has been enhanced to include a comparison with previous reports on these mosquitoes in Kwale (Line 519-527)
1	What might the co-occurrence of both Anopheles and Culex mean for vector control programs?	This has been discussed "co-occurence of both mosquitoes necessitates for more complex vector control protocols targeted to both *Anopheles*and *Culex*-mosquitoes" (519-527)

1	Can you provide any information on relationships between Anopheles and Culex. There are studies which have investigated predation between the two genera as well as the effect on life history traits etc. You have reported co-occurrence of the two genera here, what does this mean for vector control?	This discussion has been improved to consider the implications of the presence of both mosquito genera in the same environment on disease transmission dynamics (Line 519-537)
1	There is an in-depth discussion on breeding site characteristics in relation to Anopheles but your study also found Culex. Culex should also be discussed in this context. It may be beneficial to add a brief mention of the potential for Culex to be vectors, especially for west nile virus.	This has been done and reported in the discussion (Line 523-527)
1	Physiochemical parameters are described to be high (line 481). High in relation to what? Could you give some context to this or provide a reference to explain why you classify these to be high in your study?	The statement was excluded from the discussion since there was no comparison to get the threshold for high or low (Line 578-592)
1	The statement that Anopheles mosquitoes have become less selective to environmental parameters (lines 489-491) is not supported by the results of this study. You explain that there was a difference in temperature between positive and negative sites which does suggest that temperature affects larval abundance however the temperature in the dry and rainy season presented in table 1 are very similar (31.74 and 32.06). Could it not just be that the temperature did not fluctuate much between the different seasons/sites in this study? A discussion on optimal temperatures for mosquito development would be helpful here. Could it just be that the temperatures in this study were always within the optimal temperatures for mosquito development?	The statement was left out in the discussion since there is no evidence to support that from this study, it has also been mentioned in the discussion that the temperatures recorded in this study encouraged the presence mosquito larvae (Line 581 -583)
1	Limitations of the study are not well addressed in the discussion, please ensure that you have addressed the limitations thoroughly.	A section on the limitations of this study has been added in the discussion section (Line 732-740)

1	More in depth discussion on the significance of this study would improve the manuscript. How does increasing the knowledge of bacterial communities associated with mosquito larval presence feed into vector control. What are the applications of this (and other similar) research?	This discussion has now been provided both in the discussion section and in the conclusion (Line 721-731 and 763-768)
1	When discussing bacterial control agents (line 594-597) it would be beneficial to include some specific examples of this and how successful they have been.	The discussion was changed to discuss the role of bacteria in releasing volatile metabolites which could affect mosquito behaviour, and examples have been provided (Line 708-720)
1	The discussion makes no reference to the absence of other malaria vectors in this study. There are other species in this area, why were they not found in your study?	A discussion on the absence of the other malaria vectors has been done in the discussion section (Line 533-537)
1	In the abstract it is not clear what is meant by positive and negative sites (this is made clear in the manuscript but the abstract should be stand alone).	The meaning of positive and negative sites has been explained in the abstract (Line 44-45)
1	It would be helpful to give some examples of what the optimal conditions for larval survival and growth discussed in lines 80 - 88 include.	some Optimal conditions for growth of mosquito larvae have been mentioned (Line 77-81)
1	The introduction should mention the malaria vaccine as this is a key development in malaria control. However vector control is still important.	This has been mentioned accordingly (Line 82-84)
1	The introduction mentions various vector control strategies but needs to more explicitly explain why environmental management to reduce oviposition sites is a preferred method (cheaper, no resistance, no environmental effects etc.). This helps to justify why understanding of larval ecology is important to vector control.	Importance of environmental management has been explained in the introduction (Line 93-98)
1	The introduction needs a bit more background information on the situation in the study area specifically, this would help to justify why this area has been chosen.	Some background information on the situation of malaria in Kenya and in the study site has been provided according to WHO report, 2019 and Kenya malaria indicator survey, 2020 (Line 110-119)

1	The map (figure 1) could benefit from differentiation between each site to make it clearer, for example using numbering or different colors for each site.	The map was improved to show the sampling sites (Figure 1), and a description given on how the samples distributed (Line 231-238)
1	The structure of the methods has meant that in places details are repeated, could this be improved to improve flow?	Description of the methods has been improved to enhance the flow
1	The methods should include when the sampling took place. This is especially relevant given the discussion of variation in rainy seasons explained in lines 129-135. This is provided later in the methodology (lines 155-156) but would be helpful to the reader if provided with discussion on rainy seasons (lines 129-135). It is also important to provide differentiation in when the dry and rainy season collections took place, as these are presented separately in the results.	A clarification on when sampling took place during the dry and wet seasons has been provided in the methods (Line 162-164), results (Line 302-303) and discussion (Line 485-487)
1	Study design and sample size section could be clearer. Was there a total of 35 proliferation sites samples across all sample sites? If so how many from each site?	The number of sampling sites was clarified in this section (Line 162-164)
1	Were all mosquitoes identified to species level or some just to genus? Would be helpful to specify this in lines 205-207.	Clarification has been done on the level of identification of the mosquitoes (Line 226-227)
1	I find the labelling of the water samples in lines 210 - 217 to be a little confusing, as the labels don't really indicate anything about the samples. This also means that during the results the reader needs to continuously flip back to the methods to understand what is being presented. This makes the presentation of the metagenomic results a little difficult to follow. Could the labelling be changed to something more intuitive and related to the samples themselves for example P for positive sites?	Labeling of samples was changed, P represents positive sites and N represents negative sites (Line 232-237)
1	The quality control parameters given in lines 248 should be explained in plain terms rather than providing the code as it is not clear exactly what each parameter is.	The quality control parameters were explained (Line 269-281)
1	Could you provide the conditions used for successful 16S rRNA amplification (lines 234-236)?	The conditions for amplification of 16S rRNA gene were provided (Line 256-259)

1	There is an inconsistency with the number of sites samples, in the methods it states 35 sites were sampled, in the results 65. Please clarify this and correct in the manuscript.	A clarification on the number of sites sampled has been done in the manuscript (Line 162-164 and 302-303)
1	Line 450, typo - farmers come pastoralists.	Correction done (Line 545-548)
1	There is an inconsistency with the formatting of genus names in the manuscript, some are italicised, and others are not.	All genus names have been italicised in the manuscript
Reviewer 2		
2	**Background;**Restructure the first sentence, minor edits	The sentence has been restructured (Line 66-67)
2	Lines 48-50 (Correct the phyla names in accordance with the currently valid accepted nomenclature)	Phyla names were corrected in accordance with the currently accepted nomenclature (Line 47-49)
2	**Introduction;**Line 93 and Line 107	Minor edits corrected
2	Lines 110-117 It lacks a proper reason why characterising these bacterial communities in the larval stages of these mosquitos is important.	The significance of characterizing bacterial communities in mosquito larval habitats has been highlighted (Line 132-136)
2	**Material and Methods;**Line 121 to Line 203, minor word edit	Corrections have been done
2	Lines 120-128 Write the geographic coordinates in degrees, minutes, and seconds	Geographic coordinates have been written in degrees, minutes and seconds (Line 145-150)
2	Lines 206-217 Perhaps a figure to help in the comprehension of this sample distribution across geographic and source origins?	The distribution of sampling sites has been highlighted in figure 1 and cited in this section for clarity (Line 231-238)
2	Lines 234-240 So this is a metataxonomic survey and not a metagenomic one?	The correction has been done from metagenomic to metataxonomic (Line 263)
2	Line 254 Which version of the SILVA database was applied for taxonomic classification? 138? Apply to Line 256	This analysis was repeated with kraken2 pipeline "Taxonomic classification of the filtered reads was then assigned to the species level using kraken2 pipeline, implementing the bacteria refseq database of the NCBI (NCBI Bacterial RefSeq Database .ftp.ncbi.nlm.nih.gov/refseq/release/bacteria/bacteria.1.1.genomic.fna.gz. Accesses 30th May, 2023)". (Line 282-285)

2	**Results;**Lines 299, 301, 302 and 303 to Line 352Word editing was done accordingly minor word edits suggested	
2	Lines 357-358 Complete this figure legend with more information. What samples? The average number of reads obtained?	The whole figure 3 was redone and the legend was adjusted accordingly (Figure 12)
2	Line 360 Rewrite this figure legend indicating that different alpha-diversity metrics were applied for getting a glimpse of richness and diversity in studied samples.	The figure has been redone together with the legend to make the alpha diversity metrics a bit more clear, this is presented in figure 13
2	Lines 365 and 372 Correct "PCOA" to "Principal Coordinate Analysis (PCoA)" in Figure 5 and 6 legends.	This has been changed to Principal Components Analysis and presented in figure 14
2	Line 376 Rewrite in a more direct way the section title: "Taxonomic classification". Also, update the bacterial phyla names throughout the manuscript.	The section title was Corrected and bacterial phyla updated accordingly throughout the manuscript (Line 394)
2	Line 378 Rewrite "29133": "29,133" and "1450": "1,450".	The correction was done (Line 385-387)
2	Figures 8 to 11 Change the R ggplot palette colour for plotting these relative abundance graphs. The colours currently adopt make it harder for the reader to fully get the different key taxa found in all samples by each analysed taxonomic level.	All these figures were changed to adopt better color schemes which are more clear
2	Figure 12 How can this figure be interpreted? Is there any way to depict the bacterial family names?	The figure was excluded from the manuscript to focus more on describing the taxonomy, alpha and beta diversities of the bacteria in the positive and negative sites
2	Figure 13 It could be worthwhile to say in the Results text the bacterial genera that were differently abundant across the samples. It can be hard to obtain this information by solely observing Figure 13.	The figure has been replaced with a volcano plot in figure 11 which is more clear and the differentially abundant bacteria cited in the text
2	Line 461 Add a space between "1" and "m"; the same between "10" and "m2". Don't forget to write "2" in "m2" in the superscript form.	The correction was done (Line 343)
2	Line 523 Again, rephrase this Discussion section title in accordance with what has been previously commented on for the Results section.	The section title was rephrased accordingly (Line 614)
2	Line 545 - 609 Minor word edits	Corrections have been addressed

VERSION 1

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000606.v1.5

Jana Katharina Schniete; Leibniz Universitat Hannover Naturwissenschaftliche Fakultat, Institut fuer Mikrobiologie, Herrenhaeuser Strasse 2, Geb 4104, GERMANY, Hannover

Date report received: 09 May 2023
Recommendation: Major Revision

Comments: This study would be a valuable contribution to the existing literature. This is a study that would be of interest to the field and community. The reviewers have highlighted major concerns with the work presented. Please ensure that you address their comments. Please deposit the data underlying the work in the Society's data repository Figshare account here: https://microbiology.figshare.com/submit. Please also cite this data in the Data Summary of the main manuscript and list it as a unique reference in the References section. When you resubmit your article, the Editorial staff will post this data publicly on Figshare and add the DOI to the Data Summary section where you have cited it. This data will be viewable on the Figshare website with a link to the preprint and vice versa, allowing for greater discovery of your work, and the unique DOI of the data means it can be cited independently.

Reviewer 2 recommendation and comments

https://doi.org/10.1099/acmi.0.000606.v1.4

Bruno Francesco de Oliveira; UFF: Universidade Federal Fluminense, Department of Microbiology and Parasitology, Niterói, BRAZIL
https://orcid.org/0000-0001-6029-3608

Date report received: 02 May 2023
Recommendation: Major Revision

Comments: ACMI-D-23-00050 Metagenomic Profiles of Bacterial communities and environmental factors associated with proliferation of malaria vector mosquitoes within the Kenyan Coast SYNOPSIS AND MAJOR COMMENTS The manuscript "Metagenomic Profiles of Bacterial communities and environmental factors associated with proliferation of malaria vector mosquitoes within the Kenyan Coast" presents a broad investigation of the bacterial microbiota associated with mosquito larvae sampled at various points across the south coast of Kenya. The authors pursue this main aim by also describing a series of physicochemical parameters likely influencing the mosquito breeding sites in the selected sampling areas and making a clear association between these factors and the likely dispersal of the malaria vectors. Some highlights of the study include the detailed material and methods, the extreme care with the data obtainment and, particularly, the contextualization of the study's novelty with the underlying health challenge imposed by malaria in the region. There is an effort to pinpoint how the report might contribute to delineating future actions for mitigation and control of the neglected disease, despite the little discussion of how the taxa found in the metataxonomic analyses could direct these potential strategies. Overall, the article is well-written and has a straightforward approach to presenting data. The literature is up-to-date, despite the authors could include additional references for some of the tools and bioinformatic resources applied during their analyses. All genomic data has been properly deposited in the NCBI databases and the bioinformatic analyses were run accordingly to what is expected for a metataxonomic survey. However, the authors might need to understand that this is not a metagenomic investigation and need to replace this term with "metataxonomics" or "16S rRNA gene-based diversity analyses". In this sense, it would be interesting to retitle the work as well. In addition, most figure legends are poorly detailed and their quality and visualization might be severely improved. The Discussion is coherent and focused on the main results from the study, but the authors could extrapolate more based on the taxonomic

composition seen in all samples for making claims of how their report could aid in malaria control in the area. DETAILED REVIEW Background Line 30 Restructure the first sentence, it's confusing, without a link between the Anopheles vector and the decline of malaria cases in Sub-Saharan Africa, followed by the need for innovative control strategies. Line 38 Add where Kwale County is located between parenthesis (Kenya). Line 43 Correct "genomic" with "metagenomic" or "community" to better characterize the metagenomic approach applied in this case. Lines 48-50 (and throughout all manuscript, including Figures) Correct the phyla names in accordance with the currently valid accepted nomenclature (10.1099/ijsem.0.005056): for instance, Proteobacteria → Pseudomonadota; Actinobacteria → Actinomycetota. Line 60 Write "Electricity" in lower case: "electricity". Introduction Line 93 Add a space between "environment" and "(13)". Line 107 Write "Anopheles" in italics. Lines 110-117 It lacks a proper reason why characterising these bacterial communities in the larval stages of these mosquitos is important. The biotic factor is clear while the authors managed to contextualize the big picture, but a sentence or two sentences providing how the wider comprehension of the larval microbiome could aid in designing these malaria control strategies. Material and Methods Line 121 Rewrite "figure" in upper case: "Figure". Lines 120-128 Write the geographic coordinates in degrees, minutes, and seconds. Figure 1 Add "(Kenya)" between parentheses at the end of the phrase. Line 134 Since this is the first time these Anopheles species are being cited in the manuscript text, please write them in the full form after they can be abbreviated in the text. Line 152 Correct "ml" to "mL". Line 187 Correct "ml" to "mL". Line 199 Correct "ml" to "mL". Line 203 Add a space between "escaping" and ".Once". Also, correct "ml" to "mL". Lines 206-217 Perhaps a figure to help in the comprehension of this sample distribution across geographic and source origins? Lines 218 and 231 Rewrite: "community" or "metagenomic" DNA. Line 223 and 230 Correct "ml" to "mL". Lines 234-240 So this is a metataxonomic survey and not a metagenomic one? Please, make this as clear as possible in your manuscript text. The highlight here is the fact that you've applied PacBio sequencing instead of Illumina for these 16S rRNA-derived amplicons (longer than normally used in most of the metataxonomic surveys). Line 254 Which version of the SILVA database was applied for taxonomic classification? 138? Line 256 The information between parenthesis can be removed once you specify that you have used the SILVA database v.138.1 (release date). Results Lines 299, 301, 302 and 303 The Anopheles species can be abbreviated here, once you provide it fully earlier in the text. Line 303 Write "Culex" in italics. Line 313 Add a space between "1" and "m". Line 314 Write m2 in the superscript form. Line 323 Rewrite "1360" such as "1,360". Line 336 Correct "PH": "pH" and "P-value": "p-value" in Table 1. Line 339 Rephrase this section title, better "Metataxonomic survey" or "Microbiota analyses". Line 352 Add a comma after "Fo beta-diversity analyses," and write "the" in the lowercase form. Lines 357-358 Complete this figure legend with more information. What samples? The average number of reads obtained? Line 360 Rewrite this figure legend indicating that different alpha-diversity metrics were applied for getting a glimpse of richness and diversity in studied samples. Lines 365 and 372 Correct "PCOA" to "Principal Coordinate Analysis (PCoA)" in Figure 5 and 6 legends. Line 376 Rewrite in a more direct way the section title: "Taxonomic classification". Also, update the bacterial phyla names throughout the manuscript. Line 378 Rewrite "29133": "29,133" and "1450": "1,450". Figures 8 to 11 Change the R ggplot palette colour for plotting these relative abundance graphs. The colours currently adopt make it harder for the reader to fully get the different key taxa found in all samples by each analysed taxonomic level. Figure 12 How can this figure be interpreted? Is there any way to depict the bacterial family names? Figure 13 It could be worthwhile to say in the Results text the bacterial genera that were differently abundant across the samples. It can be hard to obtain this information by solely observing Figure 13. Line 461 Add a space between "1" and "m"; the same between "10" and "m2". Don't forget to write "2" in "m2" in the superscript form. Line 523 Again, rephrase this Discussion section title in accordance with what has been previously commented on for the Results section. Line 545 Write "(4, 85)" in Times (it's currently in a different font). Lines 592-595 Adjust the font here. Line 609 Write the word "electrical" in lowercase.

Please rate the manuscript for methodological rigour
Very good

Please rate the quality of the presentation and structure of the manuscript
Good

To what extent are the conclusions supported by the data?
Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

Reviewer 1 recommendation and comments

https://doi.org/10.1099/acmi.0.000606.v1.3

Anonymous.

Date report received: 28 April 2023
Recommendation: Major Revision

Comments: 1. Methodological rigour, reproducibility and availability of underlying data This research covers important questions surrounding Anopheles proliferation. Research in this area is needed and welcomed. The methodology used in this manuscript is scientifically sound and appropriate for the study. The collection of the data has been collected in a thorough and suitable manor which enables the research aims to be met. However, there are some issues with the data analysis and presentation of the results which would need to be corrected prior to acceptance. All sequence data has been deposited to NCBI and accession numbers are provided within the manuscript. The raw data used for the other elements of this study have not been provided. Whilst the manuscript includes summary data the raw data has not been provided in either supplementary information nor deposited on a relevant database. Authors should consider including this data or a statement to explain that the data is available upon request. 2. Presentation of results Most of the figures are well presented however some of the results requires more explanation and editing to the presentation to ensure that the reader is able to interpret the information presented. Please see more specific comments below. Lines 289-296 appear to be explaining that there is no significant difference in the percentage of negative and positive sites during the dry season. However it is also important to investigate whether the proportion of positive sites is different in the dry and rainy seasons. From the numbers presented it is clear that the proportion of positive sites is higher in the rainy season (as would be expected) but no stats for this seems to have been presented. You should test for a significant difference in positive site proportion between the dry and rainy season. Lines 304-305 states that the results suggest that more habitats were suitable for breeding Anopheles than Culex. It would be important to understand how this fits in to the mosquito abundance in this area. Could it not just be that there are more Anopheles than Culex. Some discussion on this, probably best placed within the discussion, would benefit the interpretation of these results. Figure 2A displays the % of positive and negative sites in the dry season, rainy season, and overall. A stacked barchart is more appropriate for displaying proportions and would make it easier to interpret whether the proportion is different in the different seasons. I'm also not sure that you should have error bars on these plots. Has the percentage positive and negative not been calculated across all sample sites to give overall percentages? How were the error bars for figure 2 calculated? They seem to be very similar across different bars, are they correct? For figure 2C the error bars are putting the relative abundance above 100 and below 0, error bars should not go below 0 or above 100. There also appear to be error bars for some of the characteristics (type, vegetation) where there is no actual bar. For readability I would recommend Figure 2C being in its own figure. This would allow you to increase readability of the figure and flow of the main text. Stacked bar charts might also work better here to reduce the number of individual bars and increase readability. Please check the data in figure 2C. For example, shouldn't the % of artificial negative sites and natural negative sites add to 100%? They don't appear to, and there are similar issues throughout this figure. Figure legends require more detail added throughout to ensure figures can be interpreted and are stand alone. Where statistical analysis has been conducted that should be included in the legend to highlight significant difference (or no differences). Table 1 is well presented. It could benefit from including asterisks to indicate where differences are significant. Have the physiochemical parameters in table 1 also been tested for significance between the positive and negative sites. If so, please include the results. If not, please complete this analysis as it is important to the study. Identifying which physicochemical characters differ between positive and negative sites would provide an important insight into the factors associated with mosquito oviposition choice (as discussed in your introduction). The refraction curve (figure 3) is well presented and explained in the main text (lines 341-343). I wonder if it would be more appropriate in the methods rather than results. Diversity of bacterial communities results section (lines 140-355) is well and explained and accompanied by well presented figures. The number of reads generated, before and after QC trimming (lines 377 - 381) might be better presented in a table. Figures 7-11 presenting the results of the taxonomic assignment are very well presented. However in the results section they are just listed and not specifically referred to in the main text of the results section. Please ensure that the main text refers directly to relevant figures pointing out key message in each figure. This would make it easier for the reader to interpret these results. It is useful to know that more bacterial families in positive sites compared to negative sites (as explained in lines 389-391) however figure 12 does not include labelling to enable the reader to understand which families specifically are found in each site. Figure should be edited to include family names. Lines 391 - 393 states that 19 genera were more significantly more abundant in positive sites. However, no p value has been provided. Has this been tested statistically? If not please avoid using the term significant. It would be useful to present this analysis for all genera (not just the significant 19) as supplementary information. The size of the points on figure 13 could be reduced to improve readability. Please explain what is meant by N/A on figure 13 in the figure legend. 3. How the style and organization of the paper communicates and represents key findings The paper has all relevant sections, and the structure

and style are effective at communicating the key findings. There are places where the results/discussion require further clarification, and this has been indicated by my specific comments throughout the review. The abstract is very long and could be condensed to more concisely convey the key findings of the study. The background section of the abstract for example could be significantly cut down. The results section could also be cut down, just focusing on key results rather than each individual result. This would make the key take home message more clear in the abstract. 4. Literature analysis or discussion When discussing the presence of Anopheles and Culex reported in this study (lines 427 - 439) it is important to give some information on known distribution/ occurrence of these mosquitoes in the local area. What might the co-occurrence of both Anopheles and Culex mean for vector control programs? Can you provide any information on relationships between Anopheles and Culex. There are studies which have investigated predation between the two genera as well as the effect on life history traits etc. You have reported co-occurrence of the two genera here, what does this mean for vector control? There is an in-depth discussion on breeding site characteristics in relation to Anopheles but your study also found Culex. Culex should also be discussed in this context. It may be beneficial to add a brief mention of the potential for Culex to be vectors, especially for west nile virus. Physiochemical parameters are described to be high (line 481). High in relation to what? Could you give some context to this or provide a reference to explain why you classify these to be high in your study? The statement that Anopheles mosquitoes have become less selective to environmental parameters (lines 489-491) is not supported by the results of this study. You explain that there was a difference in temperature between positive and negative sites which does suggest that temperature affects larval abundance however the temperature in the dry and rainy season presented in table 1 are very similar (31.74 and 32.06). Could it not just be that the temperature did not fluctuate much between the different seasons/sites in this study? A discussion on optimal temperatures for mosquito development would be helpful here. Could it just be that the temperatures in this study were always within the optimal temperatures for mosquito development? Limitations of the study are not well addressed in the discussion, please ensure that you have addressed the limitations thoroughly. More in depth discussion on the significance of this study would improve the manuscript. How does increasing the knowledge of bacterial communities associated with mosquito larval presence feed into vector control. What are the applications of this (and other similar) research? When discussing bacterial control agents (line 594-597) it would be beneficial to include some specific examples of this and how successful they have been. The discussion makes no reference to the absence of other malaria vectors in this study. There are other species in this area, why were they not found in your study? 5. Any other relevant comments In the abstract it is not clear what is meant by positive and negative sites (this is made clear in the manuscript but abstract should be stand alone). It would be helpful to give some examples of what the optimal conditions for larval survival and growth discussed in lines 80 - 88 include. The introduction should mention the malaria vaccine as this is a key development in malaria control. However vector control is still important. The introduction mentions various vector control strategies but needs to more explicitly explain why environmental management to reduce oviposition sites is a preferred method (cheaper, no resistance, no environmental effects etc.). This helps to justify why understanding of larval ecology is important to vector control. The introduction needs a bit more background information on the situation in the study area specifically, this would help to justify why this area has been chosen. The map (figure 1) could benefit from differentiation between each site to make it clearer, for example using numbering or different colours for each site. The structure of the methods has meant that in places details are repeated, could this be improved to improve flow? The methods should include when the sampling took place. This is especially relevant given the discussion of variation in rainy seasons explained in lines 129-135. This is provided later in the methodology (lines 155-156) but would be helpful to the reader if provided with discussion on rainy seasons (lines 129-135). It is also important to provide differentiation in when the dry and rainy season collections took place, as these are presented separately in the results. Study design and sample size section could be clearer. Was there a total of 35 proliferation sites samples across all sample sites? If so how many from each site? Were all mosquitoes identified to species level or some just to genus? Would be helpful to specify this in lines 205-207. The methodology gives a very detailed and concisely explained description of the genomic DNA extraction process. I find the labelling of the water samples in lines 210 - 217 to be a little confusing, as the labels don't really indicate anything about the samples. This also means that during the results the reader needs to continuously flip back to the methods to understand what is being presented. This makes the presentation of the metagenomic results a little difficult to follow. Could the labelling be changed to something more intuitive and related to the samples themselves for example P for positive sites? More details on the positive and negative controls used for sequencing is needed (lines 238-240). Could you provide a list of the 17 isolates included in the positive control in supplementary information? What exactly is the negative control? The quality control parameters given in lines 248 should be explained in plain terms rather than providing the code as it is not clear exactly what each parameter is. Could you provide the conditions used for successful 16S rRNA amplification (lines 234-236)? There is an inconsistency with the number of sites samples, in the methods it states 35 sites were sampled, in the results 65. Please clarify this and correct in the manuscript. Line 450, typo - farmers come pastoralists. There is an inconsistency with the formatting of genus names in the manuscript, some are italicised, and others are not.

Please rate the manuscript for methodological rigour
Very good

Please rate the quality of the presentation and structure of the manuscript
Satisfactory

To what extent are the conclusions supported by the data?
Partially support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

SciScore report

https://doi.org/10.1099/acmi.0.000606.v1.1

iThenticate report

https://doi.org/10.1099/acmi.0.000606.v1.2