# Supplementary information: Adaptations and community changes in milk and water kefir microbiomes in response to environmental parameters (Kefir4All-Citizen Science Project)

## Supplementary Methods

**Bioinformatic analysis**

Antimicrobial resistome analysis was performed by aligning paired-end metagenomic reads against the MEGAres database (v1.0.1). To reduce type I errors, this database was first manually curated to remove any genes corresponding to antimicrobial resistance arising from point mutations. The alignment was performed using the –very-sensitive-local parameter of Bowtie2 (v2.3.4). The Resistome Analyser tool (https://github.com/cdeanj/resistomeanalyzer) was used to format the output, and the results were normalized for sequencing depth across samples as counts per million reads (CPM).

## Supplementary Results

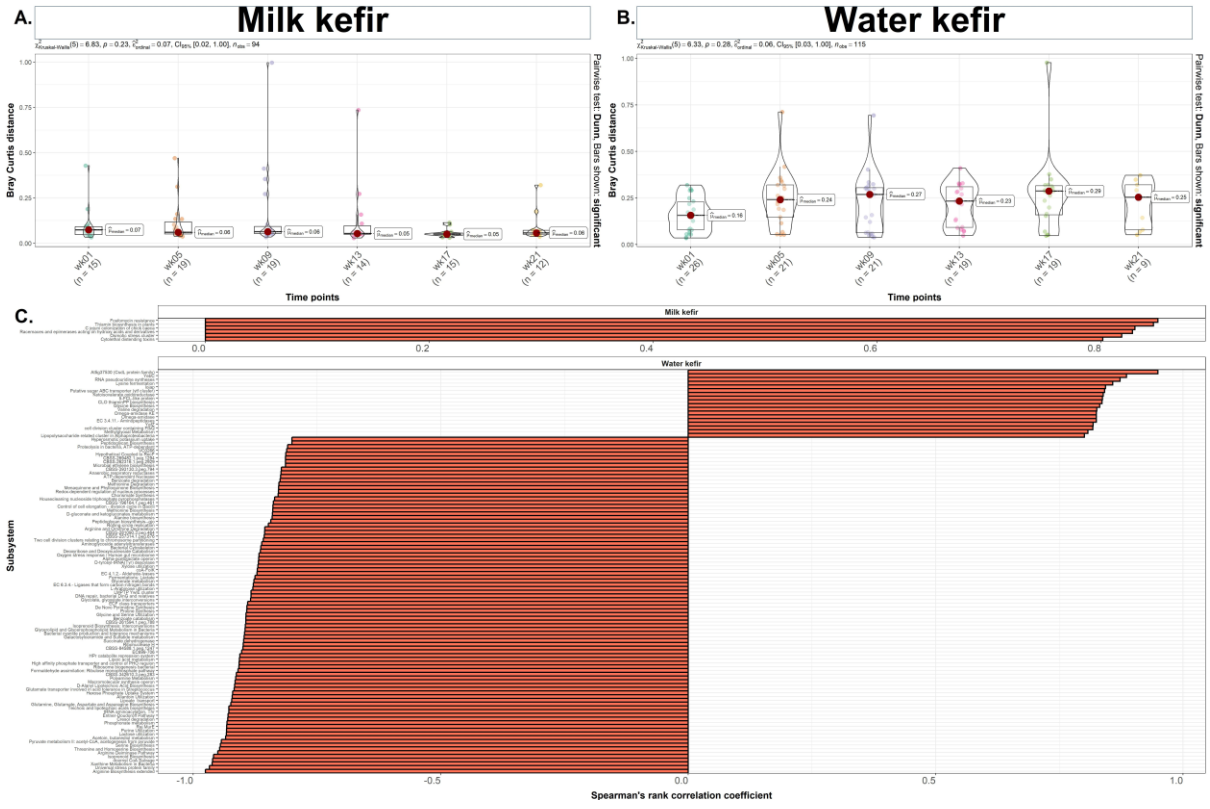
**Result S1: Functional analysis of kefir metagenomes suggests the majority of functional change within the milk and water kefir grain microbiota occurs in the initial periods of the project.**

SUPER-FOCUS [1] was used to provide an overview of the functional potential of the milk and water metagenomes and to determine if functional changes occurred during the Kefir4All study. As expected, a significant proportion of the milk and water kefir T0 grain metagenomes and their respective liquid metagenomes were assigned to housekeeping functions such as Carbohydrate metabolism, DNA metabolism, and Protein metabolism. Dissimilarity values at the functional level were observed for all kefir grain metagenomes compared to the initial grains (T0). Dissimilarity values for the milk kefir grain metagenomes compared to the initial grains (T0) ranged from 0.03 to 0.42.

Mean dissimilarity values remained similar for all points after wk01 (Figure S1A). Dissimilarity values for the water kefir grain metagenomes compared to the initial grains (T0) followed a similar pattern (Figure S2B), with values ranging from 0.07 to 0.1 at wk01, and 0.12 to 0.69 at wk05 (Figure S2B).

Correlations were observed between functional change and level 3 subsystems derived from 31 level 1 subsystems in the milk kefir microbiome, which included ‘Carbohydrates’ (n=57), ‘Clustering-based subsystems’ (n=38), ‘Amino Acids and Derivatives’ (n=32), ‘Cofactors, Vitamins, Prosthetic Groups, Pigments’ (n=29), ‘Protein Metabolism’ (n=28), ‘Cell Wall and Capsule’ (n=27), ‘Virulence’ (n=26), and ‘Stress Response’ (n=22) (Figure 7C). However strong correlations ($0.8 < |r| \leq 1$) were only observed for the P1 pathways ‘Cofactors, Vitamins, Prosthetic Groups, Pigments’ (n=4), ‘RNA Metabolism’ (n=3), ‘Regulation and Cell signalling (n=3), ‘Virulence’ (n=2) and ‘Virulence, Disease and Defense’ (n=2) (Figure S1). These correlations were exclusively positive. Correlations were observed between functional change and level 3 subsystems derived from 33 level 1 subsystems in the water kefir microbiome including Carbohydrates’ (n=80), ‘Amino Acids and Derivatives’ (n=38), ‘Protein Metabolism’ (n=42), ‘Clustering-based Subsystems’ (n=42), and ‘RNA Metabolism’ (n=30). Strong correlations both positive and negative were observed between functional change and different level 3 subsystems in the water kefir grain microbiome derived from 26 level 1 subsystems in the water kefir microbiome including ‘Carbohydrates’ (n=29), ‘Amino Acids and Derivatives’ (n=22), ‘Clustering-based subsystems’ (n=18) and ‘Cofactors, Vitamins, Prosthetic Groups, Pigments’ (n=14) (Figure S1). In both the milk and water kefir grain microbiome, functional changes further correlated with level 3 subsystems involved in adaptation, these include 'Phages, Prophages, Transposable elements', which play a role in genetic variability and adaptation [2] 'Phages, Prophages, Transposable elements, Plasmids', contributing to gene transfer and functional diversity 'Stress Response', essential for microbial survival under adverse conditions [3]; 'Virulence', which pertains to the pathogenic potential and safety of the microbial ecosystem associated with kefir; and 'Virulence, Disease and Defense', encompassing broader defensive mechanisms [4]. In the milk kefir grain microbiome, functional

changes correlated with subsystems assigned to the level 1 subsystems 'Phages, Prophages, Transposable elements' (n=5, Correlation value: 0.461-0.735), 'Phages, Prophages, Transposable elements, Plasmids' (n=9, Correlation value: -0.848 to 0.796), 'Stress Response' (n=22, Correlation value: -0.967 to 0.819), 'Virulence' (n=26, Correlation value: -0.861 to 0.851), and 'Virulence, Disease and Defense' (n=11, Correlation value: -0.896 to 0.830). Similarly, in the water kefir grain microbiome, functional changes also showed correlations with the level 1 subsystems 'Phages, Prophages, Transposable elements' (n=3, Correlation value: 0.461-0.735), 'Phages, Prophages, Transposable elements, Plasmids' (n=10, Correlation value: -0.848 to 0.796), 'Stress Response' (n=16, Correlation value: -0.967 to 0.819), 'Virulence' (n=18, Correlation value: -0.861 to 0.851), and 'Virulence, Disease and Defense' (n=13, Correlation value: -0.896 to 0.830).



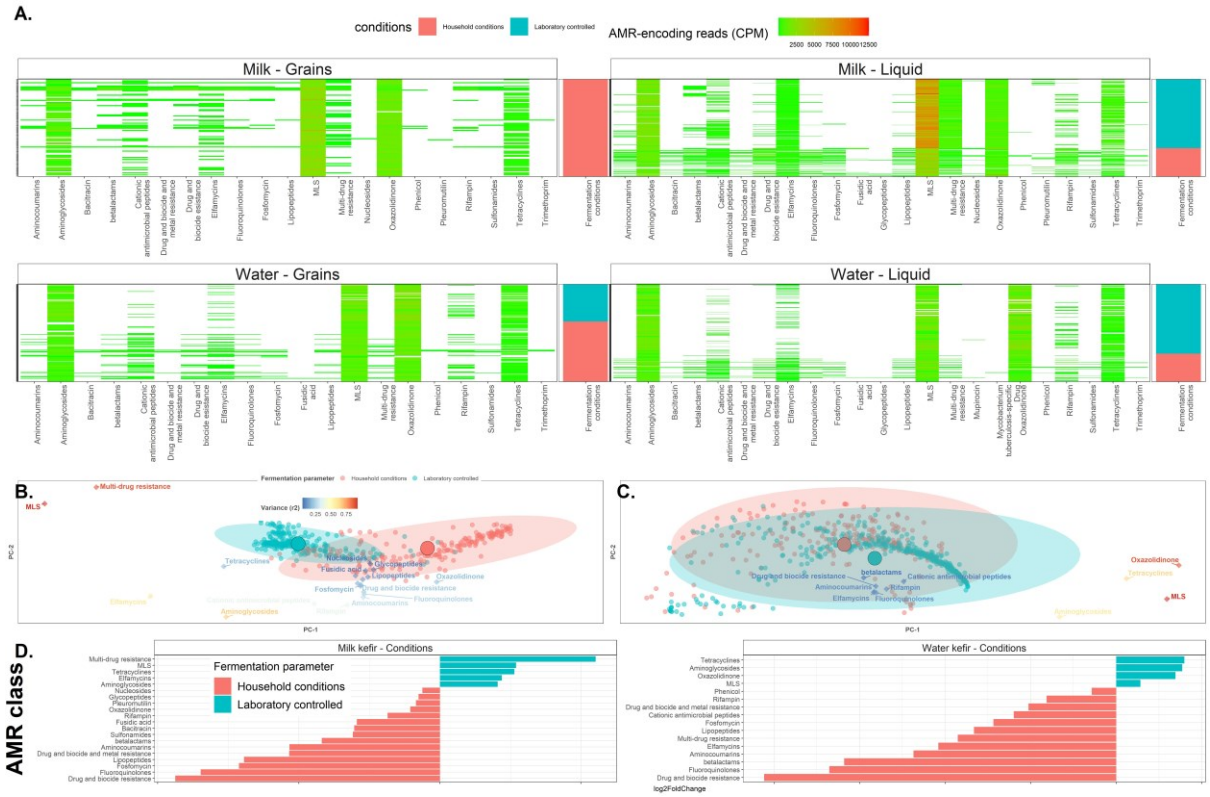

**Supplementary Figure 1**: **(A-B)** Beta-diversity boxplots displaying the Bray Curtis distance values as determined based on functional composition, across (**A**) milk grain metagenomes and (**B**) water grain metagenomes, representing different timeframe of the Kefir4All study (wk01, wk05, wk09, wk13, wk17 and wk21) compared to the T0 milk kefir grain metagenome. Kruskal-Wallis and Dunn test was used to assess differences in beta

diversity, since the data was nonparametric. Red dot represents the mean Bray Curtis distance between kefir grain metagenomes representing different timeframe of the Kefir4All study compared to the T0 kefir grain. **(C)** Statistically significant strong, positive or negative, spearman correlations ($0.8 < |r| \leq 1$) between the relative abundance of level 3 subsystems and functional change as determined using the Bray Curtis distance.

**Result S2: The milk and water resistome differs significantly according to various fermentation parameters such as conditions and kefir type**

Given the statistically significant differences across compositional, functional and metabolic content between samples produced under laboratory controlled compared to household metagenomes, the latter of which also included detections and in some cases persistence of putative microbial contaminants, we characterise the resistome of milk and water kefir to further assess the safety of both fermented foods. The resistome describes the array of antibiotic resistant genes (ARGs) present within a microbial community [5]. The milk kefir resistome had an average of 8,598 CPM per metagenome, and contained individual ARGs belonging to 22 classes, while water kefir had an average of 3,792 CPM per metagenome, and contained individual ARGs belonging to 19 classes (Figure S2A). With respect to specific Antimicrobial resistance (AMR) classes aminocoumarins, aminoglycosides, betalactams, cationic antimicrobial peptides, drug and biocide resistance, elfamycins, fluoroquinolones, fosfomycin, lipopeptides, MLS, multi-drug resistance, oxazolidinone, rifampin and tetracyclines were deemed prevalent classes in the milk kefir microbiome (Figure S2A). Aminocoumarins, cationic antimicrobial peptides, elfamycins, MLS, oxazolidinone, rifampin and tetracyclines were deemed prevalent classes in the water kefir microbiome (Figure S2A). We further identified a number of predominant classes including, tetracycline, macrolide-lincosamide-streptogramin (MLS), aminoglycosides and oxazolidinone classes which accounted for the majority of the ARGs assigned CPM within all metagenomes (Figure S2A). From the predominant classes, the majority of the ARGs within the MLS class had functionality in the following ARG types Macrolide-resistant 23S rRNA mutation, aminoglycoside O-phosphotransferases and aminoglycoside-resistant

16S ribosomal subunit protein within the aminoglycosides class and oxazolidinone-resistant 23S rRNA mutation within the oxazolidinone class.




**Supplementary Figure 2:** Prevalence of AMR classes within the milk and water kefir microbiome, distinguishing antimicrobial resistance (AMR) classes and the influence of fermentation conditions on their occurrence. **(A)** Heatmap displaying the CPM per metagenome of AMR classes (X-axis) between the milk and water kefir metagenomes (Y-axis), considered in this study. Row side annotations include metagenome classifications separated into, ■ Household conditions, and ■ Laboratory controlled. **(B)** Principal Coordinate analysis (PCoA) plot of beta diversity measure by Bray-Curtis dissimilarity of the total metagenomes considered in this study, as calculated for AMR composition obtained using resistomeanalyzer. Larger circles ● represent centroid points between each fermentation parameter. **(C)** Deseq covariate analysis reveals the effect size (Log2FoldChange) (X-axis) of the fermentation parameters, ■ Household conditions, and ■ Laboratory controlled on the relative abundance (RA) of the outlined AMR classes (Y-axis).

To further investigate the features that influence the observed resistome of milk and water kefir, three fermentation parameters were modelled with EnvFit [6] when analysing all metagenomes considered in this study, 'Kefir type' ($R^2$= 0.4 and 0.03, p <0.05), 'Conditions' ($R^2$= 0.53 and 0.05, p <0.05) (Figure S2B) and 'Community type' ($R^2$= 0.7 and 0.3, p < 0.05) had a strong to weak significant association in both the milk and water kefir resistome profile. EnvFit was further used to assess metadata covariates specific to the Kefir4All generated metagenomes, Timeframe' ($R^2$= 0.1 and 0.03, p <0.05) had a modest significant association with the milk kefir and water kefir resistome profile

'Fermentation categories' ($R^2$= 0.1, p <0.05) and 'Fermentation type' ($R^2$= 0.07, p <0.05), had a weak significant association with the water kefir resistome profile. In milk kefir 20 statistically significant AMR classes contributed to differences across milk kefir metagenomes, the most distinguishing of which included 'MLS' ($R^2$= 0.94), 'Multi-drug resistance' ($R^2$= 0.88), 'Aminoglycosides' ($R^2$= 0.67), 'Elfamycins' ($R^2$= 0.59) and 'Rifampin' ($R^2$= 0.37). In water kefir, 10 statistically significant AMR classes contributed to differences across water kefir metagenomes, the most distinguishing of which included 'MLS' ($R^2$= 0.77), 'Oxazolidinone' ($R^2$= 0.73), 'Tetracyclines' ($R^2$= 0.53), 'Aminoglycosides' ($R^2$= 0.47) and 'Fluoroquinolones' ($R^2$= 0.08).

The influence of fermentation parameters on the resistome in milk and water kefir was further examined using alpha, beta diversity and differential abundance analysis. Alpha (Kruskal-Wallis, p <0.001) and beta diversity (PERMANOVA: R2 = 0.20, p= <0.001) measures of the ARG content differed significantly between milk and water kefir types. Both liquid and grain water kefir metagenomes exhibit significantly higher resistome alpha diversity compared to the milk liquid and grain metagenomes (Kruskal-Wallis, p <0.001 and Dunn test p <0.001), while the liquid metagenomes of both milk and water kefir exhibit significantly higher resistome alpha diversity compared to their grain counterparts (Kruskal-Wallis, p <0.001 and Dunn test p <0.001). Pairwise PERMANOVA reported significant differences between all kefir types (Pairwise ANOSIM, p = 0.006) (Figure S2A and B), and DESeq2 further identified 5 milk kefir grain associated and 4 water kefir grain associated differentially

abundant AMR classes. Of these differentially abundant features reported for both the milk and water kefir grain type, Oxazolidinone (Deseq2, log2-fold changes = 0.98 and 0.47, p= <0.001) was the only differentially abundant feature that was reported to contribute to differences across milk and water kefir metagenomes as assessed using EnvFit. In contrast, Deseq2 reported 13 milk kefir liquid associated differentially abundant AMR classes, 12 of which are considered classes that contributed to differences across milk kefir metagenomes and includes Rifampin, Elfamycins and Multi-drug resistance, which represents three of the top five AMR classes that most contributed to differences across milk kefir metagenomes. Similarly within the water kefir microbiome the fermentation parameter liquid contains 11 differentially abundant features of these differential abundant features 8 are considered to be classes that contributed to differences across water kefir metagenomes, and includes Aminoglycosides, Fluoroquinolones which represent two of the top AMR classes that most contributed to differences across water kefir metagenome) (Figure S2C). Metagenomes assigned to the laboratory controlled category exhibit significantly lower species alpha diversity compared to the metagenomes assigned to the household conditions category (Kruskal-Wallis, $p < 0.001$ and Dunn test $p < 0.001$). Statistical testing revealed significant dissimilarity between kefir metagenomes produced under laboratory controlled compared to household conditions (PERMANOVA, $p < 0.05$) (Fig. S2B). DESeq2 further identified 15 and 12 differentially abundant AMR classes associated with fermentation parameter household conditions within the milk kefir and water microbiome respectively, compared to 5 and 4 differentially abundant AMR classes associated with fermentation parameter laboratory conditions within the milk kefir and water microbiome respectively (Figure S2C). DESeq2 further identified 14 and seven differentially abundant AMR classes associated with household conditions that were determined by EnvFit to contribute to differences across milk and water kefir metagenome respectively, compared to four differentially abundant AMR classes associated with household conditions that contributed to differences across milk and water kefir metagenome. However despite the differences in number of AMR classes, all four laboratory associated differentially abundant AMR

classes detected within the milk and water kefir microbiome were within the top five AMR classes contributed to differences across milk or water kefir metagenomes (Figure S2B).

Pairwise PERMANOVA reported significant differences between a number of community types. Statistically significant community types included *L. kefiranofaciens* vs *Lla. cremoris* (Pairwise ANOSIM, $R^2$ = 0.37 and p= 0.009), *L. kefiranofaciens* vs *L. helveticus* (Pairwise ANOSIM, $R^2$ = 0.16 and p= 0.009), *L. kefiranofaciens* vs *Lla. lactis* (Pairwise ANOSIM, $R^2$ = 0.43 and p= 0.009), *Lla. cremoris* vs *L. helveticus* (Pairwise ANOSIM, $R^2$ = 0.18 and p= 0.018), *L. helveticus* vs *Lla. lactis* (Pairwise ANOSIM, $R^2$ = 0.19 and p= 0.009) in the milk kefir microbiome and *Len. hilgardii* vs *Z. mobilis* (Pairwise ANOSIM, $R^2$ = 0.08 and p= 0.009) and *Z. mobilis* vs *Lac. paracasei* (Pairwise ANOSIM, $R^2$ = 0.06 and p= 0.045) in the water kefir microbiome. Furthermore, community types within the milk and water kefir microbiome, exhibit significantly differences in alpha diversity (Kruskal-Wallis, p <0.001). Metagenomes assigned to the *L. helveticus* community type exhibit significantly higher species alpha diversity compared to the other communities (Dunn test p <0.001), while those assigned to the *Lla. lactis* community type exhibit significantly lower species alpha diversity (Dunn test p <0.001) in the milk kefir microbiome. Metagenomes assigned to the *Len. hilgardii* community type exhibit significantly higher species alpha diversity compared to the other communities (Dunn test p <=0.02), while those assigned to the *Z. mobilis* community type exhibit significantly lower species alpha diversity (Dunn test p <=0.02) in the milk kefir microbiome. Community type associated differences in ARG content was further assessed using DESeq2 to identify differentially abundant features. Within the milk kefir microbiome, the community type *L. helveticus* contains 10 differentially abundant AMR classes which includes aminocoumarins, drug and biocide and metal resistance, drug and biocide resistance, fluoroquinolones, fosfomycin, lipopeptides, phenicol, pleuromutilin, rifampin and Sulfonamides. Within the water kefir microbiome, the community types *Len. hilgardii* contain seven differentially abundant AMR classes which includes aminocoumarins, Betalactams, drug and biocide and metal resistance, Drug and biocide resistance, fluoroquinolones, lipopeptides and multi-drug resistance *Z.*

*mobilis* contains two differentially abundant AMR classes which includes aminoglycosides and mupirocin and *Lac. paracasei* contains the differentially abundant AMR class MLS.

The fermentation category 'Brown sugar' has statistically higher alpha diversity compared to the other fermentation categories (Dunn tests p <=0.04) within the water kefir microbiome and beta diversity revealed significant dissimilarity between all fermentation categories (Dunn tests p <=0.02), except 'White sugar-dried fig' vs 'White sugar-dried fig and fresh lemon' (Dunn tests p >0.05). Within the water kefir microbiome, the fermentation category brown sugar further contains three differentially abundant AMR classes, which includes lipopeptides, phenicol and drug and biocide and metal resistance.

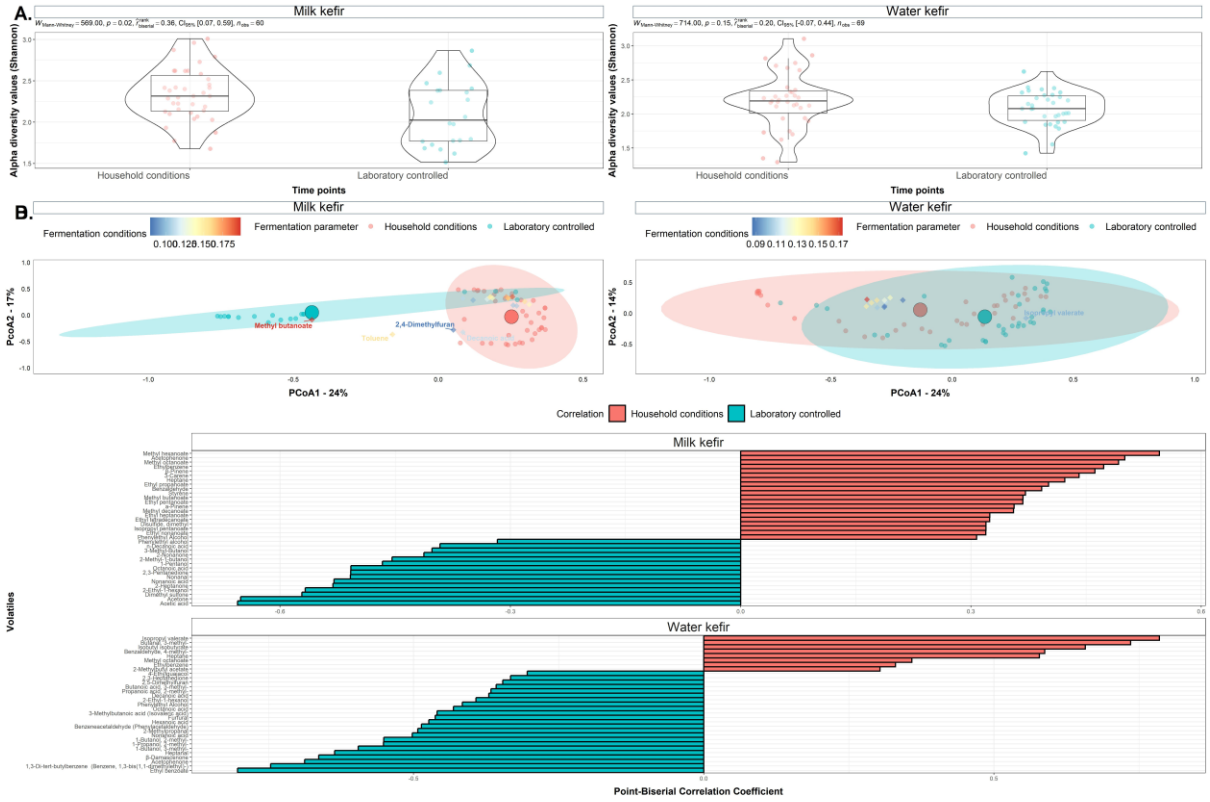
**Result S3: Fermentation conditions may influence the volatile profile of both milk and water kefir**

42 milk kefir and 39 water kefir liquid samples were analysed for volatile compounds (see methods 2.7). In milk kefir, 99 volatile compounds were identified. This consisted of 31 esters, 11 aldehydes, 10 alcohols, 10 ketones, nine terpenes, seven acids, seven benzenes, seven alkanes, five sulphur compounds, four alkanes, three furans, one ether and one phenol. Twelve compounds were consistently detected across all milk kefir samples: 2-Heptanone (Ketone), Ethanol (Alcohol), Ethyl butanoate (Ester), Ethyl decanoate (Ester), Ethyl hexanoate (Ester), Ethyl octanoate (Ester), Hexanoic acid (Acid), Nonanal (Aldehyde), Octanoic acid (Acid), p-Xylene (Benzene), β-Pinene (Terpene) and Toluene (Benzene), were consistently recovered across milk kefir samples (Figure 7A). Of the total volatiles detected in milk kefir, 30 statistically significant volatiles contributed to differences as assessed using Envfit. Within milk kefir the most abundant and distinguishing compounds, based on mean volatile abundance and Envfit analysis, were core ester compounds: ethyl octanoate (Abundance = 7819142 and $R^2$ = 0.89), ethyl butanoate (Abundance = 5544648 and $R^2$ = 0.62), ethyl hexanoate (Abundance = 4213466 and $R^2$ = 0.81), and ethyl decanoate (Abundance = 5544648 and $R^2$ = 0.72), followed by ethanol (Abundance = 1937167 and $R^2$ = 0.32), contributing to the abundance of ethyl esters.

In water kefir, 94 volatile compounds were identified. This consisted of 31 esters, 23 terpenes, eight ketones, seven alcohols, six aldehydes, five acids, four benzenes, three furans, two alkanes and one acetal, dixolane, ether, sulphur compound and terpene (Figure 7A)

Twelve compounds were consistently detected across all water kefir samples: α-Thujene (Terpene), Butanal, 3-methyl- (Aldehyde), D-Limonene (Terpene), Ethyl decanoate (Ester), Ethyl octanoate (Ester) and Toluene (Benzene). Of the total volatiles detected in water kefir, 63 statistically significant volatiles contributed to differences as assessed using Envfit. Ethanol (Abundance= 7286138 and $R^2$ = 0.73) was the most abundant compound on average across water kefir, followed by the ester compounds, ethyl octanoate (Abundance = 5502613 and $R^2$ = 0.72), Ethyl acetate (Abundance = 3834455 and $R^2$ = 0.47), heptane (Abundance = 791800 and $R^2$ = 0.9), and disulfide, dimethyl (Abundance = 5204 and $R^2$ = 0.68) (Figure 7A)

To determine if there was differences between kefir samples derived from ferments produced under laboratory controlled or household conditions, we compare volatile profiles of kefir samples generated in this study and produced under household conditions by citizen scientists, to laboratory produced T0 samples and further included volatile profiles from other publically available studies including 12 laboratory produced kefir samples generated in Walsh *et al* [7], six laboratory produced milk kefir samples generated in Gethins *et al* [8] and 34 laboratory produced water kefir samples generated in *Breselge et al* (under review). Profiles of milk kefir samples assigned to the laboratory controlled category exhibit significantly lower species alpha diversity compared to the metagenomes assigned to the household conditions category (Kruskal-Wallis, p=0.02) (Figure S3A) and further statistical testing revealed significant dissimilarity between milk and water kefir metagenomes produced under laboratory controlled compared to household conditions (PERMANOVA, $p < 0.05$) (Figure S3B).




**Supplementary Figure 3: (A)** Alpha-diversity boxplots illustrating the Shannon diversity values across fermentation conditions, ■ Household conditions or ■ Laboratory controlled. For both milk and water kefir, statistical significance between groups was assessed using the Kruskal-Wallis test followed by Dunn's post-hoc test to address nonparametric data. **(B)** Principal Coordinate analysis (PCoA) plot of beta diversity measure by Bray-Curtis dissimilarity of the kefir samples, as calculated for volatile profile. **(C)** Bar charts displaying moderate to strong Point- Biserial correlations ($0.3 < |r| < 0.8$) for volatiles in kefir samples under different fermentation conditions. Each bar represents a volatile compound, with positive correlations (right) indicating higher abundance in ■ Household conditions and negative correlations (left) indicating higher abundance in ■ Laboratory controlled.

Using multiple Mann-Withney-Wilcoxon tests and point-biserial correlations, we observed statistical differences in the abundances of various volatile compounds between milk and water kefir samples produced under household conditions compared to laboratory conditions (Figure S3C). We observed significantly higher abundance in 29 volatile compounds, including Acetic acid (Rank biserial = 0.66 , p <0.01 ), Acetone (Rank biserial = 0.65 , p <0.01 ), Dimethyl sulfone (Rank biserial = 0.57 , p <0.01 ), 2-

Ethyl-1-hexanol (Rank biserial = 0.57 , p <0.01 ) and 2-Heptanone (Rank biserial = 0.53 , p <0.01 ) in milk kefir samples produced under household conditions compared to laboratory conditions. While 16 volatiles, including, Methyl hexanoate (Rank biserial = 0.55 , p <0.01 ), Acetophenone (Rank biserial = 0.5 , p <0.01 ), Methyl octanoate (Rank biserial = 0.49 , p <0.01 ), Ethylbenzene (Rank biserial = 0.47 , p <0.01 ) and β-Pinene (Rank biserial = 0.46 , p <0.01 ) were significantly higher in milk kefir samples produced under laboratory conditions compared to household conditions (Figure S3C). In water kefir, we observed significantly higher abundances in 14 volatile compounds, including Ethyl benzoate (Rank biserial = 0.8 , p <0.01 ), 1,3-Di-tert-butylbenzene (Rank biserial = 0.75 , p <0.01 ), Acetophenone (Rank biserial = 0.69 , p <0.01 ), β-Damascenone (Rank biserial = 0.66 , p <0.01 ) and Heptanal (Rank biserial = 0.64 , p <0.01 ) in water kefir samples produced under household conditions compared to laboratory conditions. While 23 volatile compounds, including, Isopropyl valerate (Rank biserial = 0.78 , p <0.01 ), Butanal, 3-methyl- (Rank biserial = 0.74 , p <0.01 ), Isobutyl isobutyrate (Rank biserial = 0.66 , p <0.01 ), Benzaldehyde, 4-methyl- (Rank biserial = 0.59 , p <0.01 ) and Heptane (Rank biserial = 0.58 , p <0.01 ) were significantly higher in water kefir samples produced under laboratory conditions compared to household conditions (Figure S3C).

Correlation coefficient analysis further revealed a number of correlations between prevalent species, notability the species that tend to dominate either milk or water kefir metagenomes based on relative abundance, has comparable numbers of correlations compare to lower abundance microbial taxa. Correlations of interest in milk kefir samples include a 19 statistically significant negative correlation between hexanal a natural volatile compound with antimicrobial activity and a number of putative microbial contaminants such as, *Salmonella enterica* (R= -0.57, p <0.01), *K. oxytoca* (R= -0.56, p <0.01), *K. quasipneumoniae* (R= -0.44, p <0.01) and *E. coli* (R= -0.44, p <0.01). Hexanal has previous been shown to demonstrate antibacterial activity against pathogens such as *Salmonella* spp., *E. coli*, and *P. aeruginosa* [9]. The strongest positive correlations were observed between *Br. bruxellensis* with the esters Ethyl octanoate (R= 0.69, p <0.01) and Ethyl decanoate (R= 0.67, p <0.01) and similarly in water

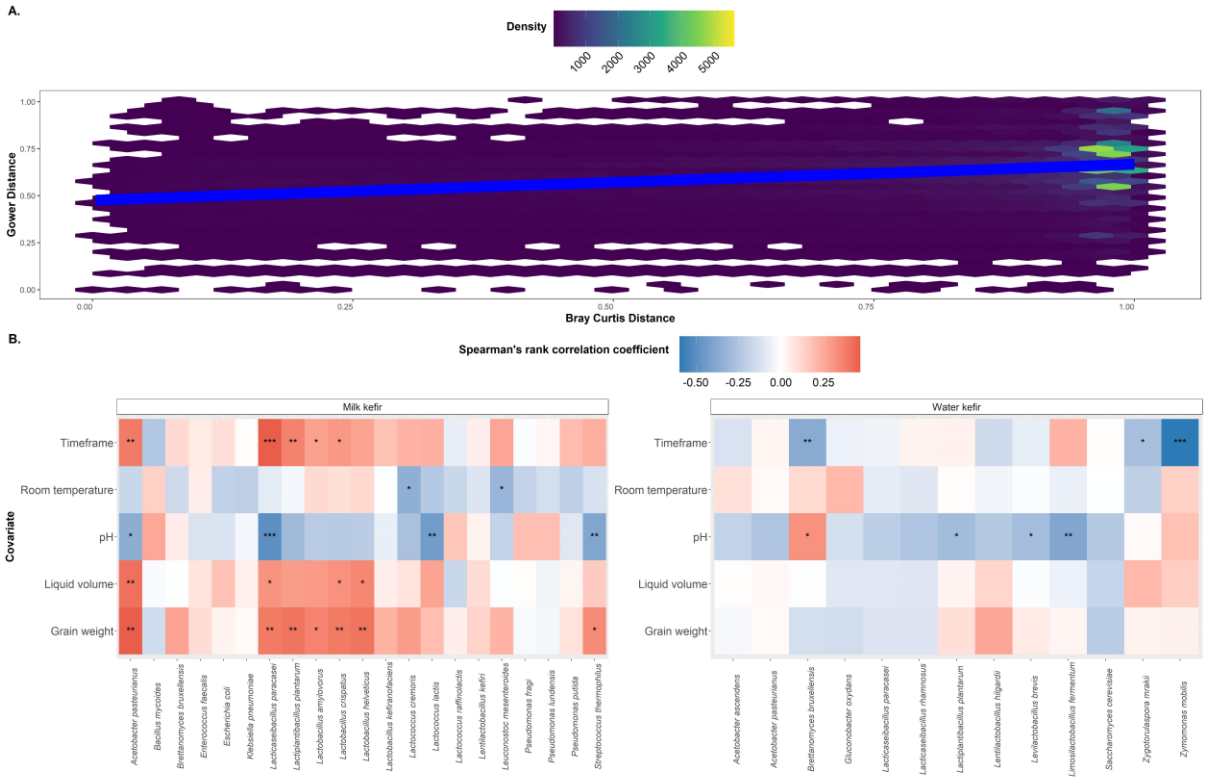
kefir samples, we observe five statistically significant negative correlation between putative microbial contaminants and Benzaldehyde with correlation strength ranging from *Sphingomonas paucimobilis* (R= -0.50, p <0.01) to *Kom. Saccharivorans* (R= -0.40, p <0.01).

**Result S4: Fermentation parameters, and abiotic factors are modest predictors of milk and water kefir compositional profiles**

Time-series data and metadata from citizen scientists were used to explore associations between fermentation parameters and microbial relative abundance (RA) changes over the Kefir4All study (wk01–wk21). Gower dissimilarity analysis revealed significant differences across all timeframes (Dunn's test, $p \leq 0.02$), though median dissimilarity remained stable (Extended Data Figure 1E, 1F). A Mantel test comparing Bray-Curtis microbial dissimilarity with Gower dissimilarity of 13 metadata variables showed a modest but significant correlation ($R = 0.25$, $p < 0.01$), suggesting that fermentation parameters influenced microbial composition but did not fully explain variability (Figure 6A). This correlation strengthened ($R = 0.31$, $p < 0.01$) when restricted to liquid metagenomes and incorporating 'Grain community type' (dominant microbial species in starter grains).

EnvFit analysis revealed weak but significant associations between fermentation parameters and microbiome composition. In milk kefir, 'Grain community type' ($R^2 = 0.12$, $p = 0.01$), 'Room temperature' ($R^2 = 0.12$, $p = 0.04$), 'Liquid volume' ($R^2 = 0.1$, $p = 0.03$), and 'pH' ($R^2 = 0.06$, $p = 0.03$) were influential. In water kefir, 'Grain weight' ($R^2 = 0.32$, $p = 0.03$), 'Room temperature' ($R^2 = 0.13$, $p = 0.02$), and 'Water type' ($R^2 = 0.06$, $p < 0.01$) showed modest to weak associations.

Spearman correlation analysis highlighted species-environment relationships (Figure 6B). In milk kefir, *Bacillus thuringiensis* ($r = 0.32$, $p \leq 0.001$) and *Bacillus cereus* ($r = 0.38$, $p = 0.01$) positively correlated with pH, while in water kefir, *Pantoea vagans* ($r = 0.34$, $p = 0.01$) and *Rahnella aquatilis* ($r = 0.29$, $p = 0.04$) showed similar trends.




**Supplementary Figure 4**: **(A)** Hexagonal heatmap visualizing the correlation between the distance in the compositional Bray-Curtis dissimilarity distance matrix and the metadata Gower distance matrix. Metadata encompassed 13 variables (see Methods).The hexagon heatmap represents the density of data points within that region: lighter hexagons indicate fewer data points, while darker hexagons signify higher concentrations of data points. The blue line indicates the linear fit, showing the overall trend or correlation between the two matrices. The analysis revealed a significant positive relationship between the metadata Gower distance and the species Bray-Curtis dissimilarity distance, with a Mantel statistic R of 0.25 ($p < 0.01$), indicating that environmental parameters have a measurable association on microbial composition **(B)** Heatmap displaying the Spearman rank correlation coefficient between the fermentation parameters Total fermentation response frequency, Room temperature, Grain weight, Liquid volume and pH and the prevalent microbial species (X-axis) detected in milk and water kefir metagenomes (liquids and grains), derived from samples collected from citizen scientists over the course of the Kefir4All project (wk01-wk21) and T0 samples.

Statistically significant correlations were observed between the relative abundance of microbial species within the milk and water kefir microbiomes and various fermentation parameters. In milk kefir metagenomes, 24 correlations were observed for pH, 23 for timeframe, 17 for liquid volume or grain weight, and eight for room temperature. In water kefir metagenomes, correlations included 72 for pH, 39 for timeframe, and one for room temperature. No correlations were observed for liquid volume and grain weight in water kefir, likely due to participants adhering to the recommended 300 ml water to 12 g grain ratio. Conversely, the recommended 300 ml milk to 12 g grain ratio was less consistently followed due to grain growth challenges.

Negative correlations with room temperature were noted for prevalent species, including *Lla. cremoris* in both milk and water kefir, as well as *Leu. mesenteroides* and *Leu. pseudomesenteroides* (Spearman $r = -0.32$ to $-0.34$, $p \leq 0.04$). In milk kefir, *Lla. lactis* (Spearman $r = -0.43$, $p \leq 0.01$) and *S. thermophilus* (Spearman $r = -0.39$, $p \leq 0.001$) showed negative correlations with pH, while no positive correlations with pH were observed for prevalent species. Negative correlations were also found between *Lim. fermentum* (Spearman $r = -0.39$, $p = 0.001$) in water kefir and *A. pasteurianus* (Spearman $r = -0.36$, $p = 0.02$) in milk kefir. One positive correlation was observed for *Br. bruxellensis* (Spearman $r = 0.32$, $p = 0.03$) in water kefir.

Putative environmental microbes such as *B. thuringiensis* (Spearman $r = 0.32$, $p \leq 0.001$) and *B. cereus* (Spearman $r = 0.38$, $p = 0.01$) in milk kefir, and *Pa. vagans* (Spearman $r = 0.34$, $p = 0.01$) and *R. aquatilis* (Spearman $r = 0.29$, $p = 0.04$) in water kefir, were positively correlated with pH. Many environmental microbes exhibited negative correlations with timeframe, including *B. thuringiensis* in milk kefir and *R. aquatilis* (Spearman $r = -0.27$ to $-0.28$, $p \leq 0.01$) in water kefir. Notably, no negative correlations were observed for prevalent species in milk kefir, though modest positive correlations were found for *L. crispatus*, *L. amylovorus*, and *Lac. plantarum* for both timeframe and grain weight (Spearman $r = 0.28$ to $0.4$, $p \leq 0.001$). In water kefir, no positive correlations were observed for prevalent species with timeframe or grain weight, but negative correlations were identified for *Z. mobilis*, *Zg. mrakii*, and *Br. bruxellensis* (Spearman $r = -0.27$ to $-0.59$, $p \leq 0.001$).

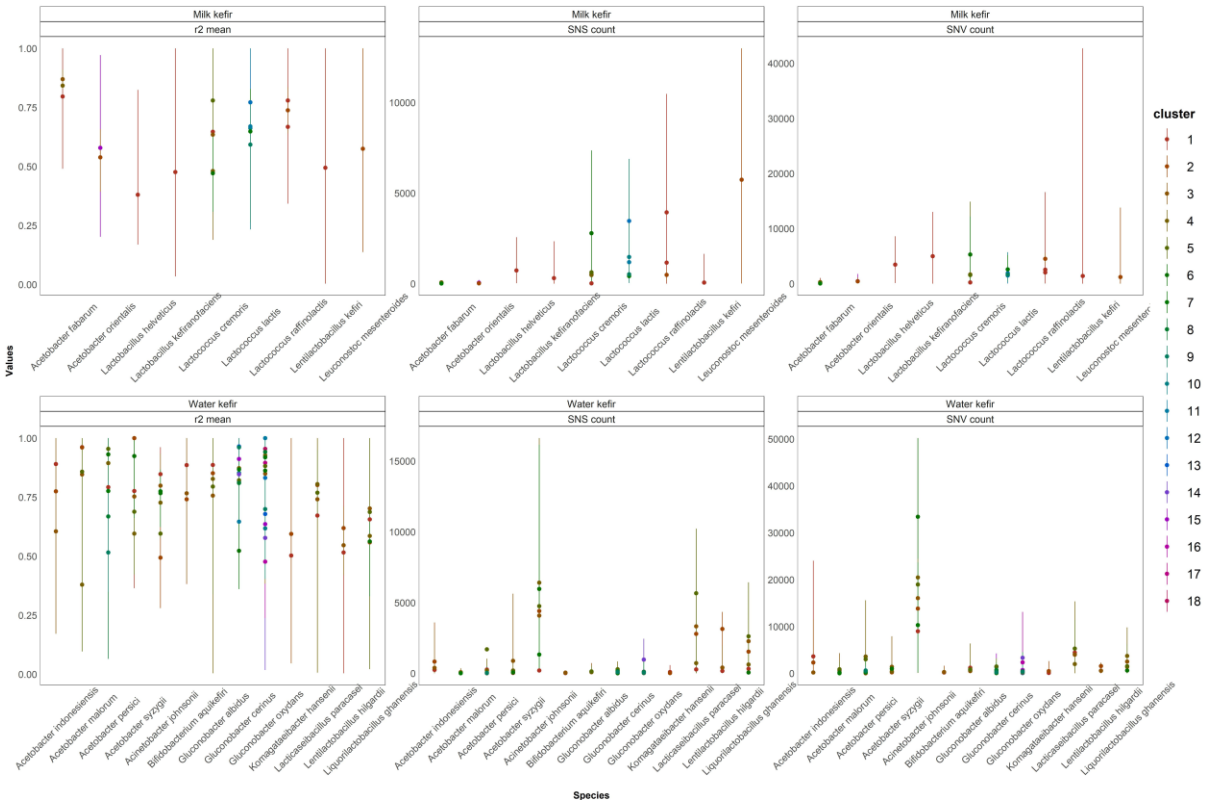
Differential abundance analysis via LEfSe identified several fermentation-associated species. Environmental microbes such as *Lc. paracasei*, *Br. bruxellensis*, *A. pasteurianus*, *C. tyrobutyricum*, and *Len. hilgardii* in milk kefir, and *Lla. lactis* in water kefir, exhibited higher abundances with the presence of pets. Species such as *K. grimontii*, *P. putida*, and *K. quasipneumoniae* in milk kefir, and *Leu. gelidum* in water kefir, were more abundant in participants fermenting other foods. The presence of *L. kefiranofaciens* was noted in a water kefir sample from a participant also fermenting injera. Additionally, species such as *C. tyrobutyricum*, *C. beijerinckii*, *K. grimontii*, *En. faecalis*, *Leu. mesenteroides*, and *P. putida* in milk kefir, and *Lla. lactis* and *Lim. fermentum* in water kefir, were associated with specific fermentation categories.

**Result S5: Phylogenetic Analysis of Milk and Water Kefir Microbiomes Reveal patterns of within-strain genomic diversity.**

A phylogenetic analysis was conducted by comparing strains from prevalent species detected in the milk and water kefir microbiomes and derived from the different clusters to the representative genome for each cluster to infer differences in recombination rates and potential selective pressures within each of the outlined secondary clusters. This analysis included key metrics from InStrain: single nucleotide substitutions (SNS) count, and single nucleotide variants (SNV) count and the r2 mean metric to measure linkage.

The InStrain metrics vary widely across different representatives species but remained relatively similar across the secondary clusters in most cases (Figure S5). SNS counts were relatively low across most strains and did not significantly differ between secondary clusters. However, species such as *Leu. mesenteroides*, *Lla. raffinolactis*, *Lla. lactis* and *Lla.cremoris* in milk kefir and *Ac. johnsonii*, *Lc. paracasei* and *Liq. satsumensis* in water kefir showed higher SNS counts, indicating increased allele (both major and minor) deviations (Figure S5). Within milk kefir, the single secondary cluster of *Len. kefiri* shows a wider range of values for the SNV counts compared to other secondary clusters, indicating partial allele deviations from the reference In line with this observation, the SNS counts were relatively low (Figure S5). Varying recombination rates were observed across species as assessed using the r2 metric.

In milk kefir, species such as *A. fabarum Lla. cremoris* and *Lla. lactis* showed higher mean $r^2$ values, indicating stronger linkage and lower recombination. In contrast, species like *L. helveticus L. kefiranofaciens*, *Len. kefiri* and *Leu. mesenteroides* exhibited lower mean $r^2$ values, suggesting higher recombination rates. In water kefir, *Acetobacter malorum* displayed a wide range of $r^2$ values across different clusters, indicating varying recombination levels within the same species. Similar patterns where observed within across strain clusters of the species *A. indonesiensis*, *A. persici*, *Ac. johnsonii*, *Bi. aquikefiri*, *Glu. albidus*, *Glu. cerinus* and *Glu.oxydans*




**Supplementary Figure 5:** The point range plot illustrates the distribution of various InStrain metrics, single nucleotide substitutions (SNS) count, and single nucleotide variants (SNV) count and r2 mean - across microbial species in Milk kefir and Water kefir samples. Each panel focuses on a distinct InStrain metric, with the X-axis displaying the microbial species detected and the Y-axis showing the corresponding metric values. Points on the plot represent the mean metric values for secondary clusters identified using dRep, with vertical lines indicating variability or confidence intervals. The points are color-coded according to their respective secondary clusters

# Supplementary References


1. Silva, G.G., et al., *SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data.* Bioinformatics, 2016. **32**(3): p. 354-61.
2. Camarillo-Guerrero, L.F., et al., *Massive expansion of human gut bacteriophage diversity.* Cell, 2021. **184**(4): p. 1098-1109.e9.
3. Dawan, J. and J. Ahn, *Bacterial Stress Responses as Potential Targets in Overcoming Antibiotic Resistance.* Microorganisms, 2022. **10**(7).
4. Abril, A.G., et al., *Proteomic Characterization of Virulence Factors and Related Proteins in Enterococcus Strains from Dairy and Fermented Food Products.* Int J Mol Sci, 2022. **23**(18).
5. Ma, T., et al., *Expressions of resistome is linked to the key functions and stability of active rumen microbiome.* Animal Microbiome, 2022. **4**(1): p. 38.
6. Oksanen, J., et al., *vegan: Community Ecology Package*. 2020.
7. Walsh, A.M., et al., *Microbial Succession and Flavor Production in the Fermented Dairy Beverage Kefir.* mSystems, 2016. **1**(5).
8. Gethins, L., et al., *Acquisition of the yeast Kluyveromyces marxianus from unpasteurised milk by a kefir grain enhances kefir quality.* FEMS Microbiology Letters, 2016. **363**(16).
9. Raybaudi-Massilia, R.M., et al., *Control of Pathogenic and Spoilage Microorganisms in Fresh-cut Fruits and Fruit Juices by Traditional and Alternative Natural Antimicrobials.* Comprehensive Reviews in Food Science and Food Safety, 2009. **8**(3): p. 157-180.