



OPEN

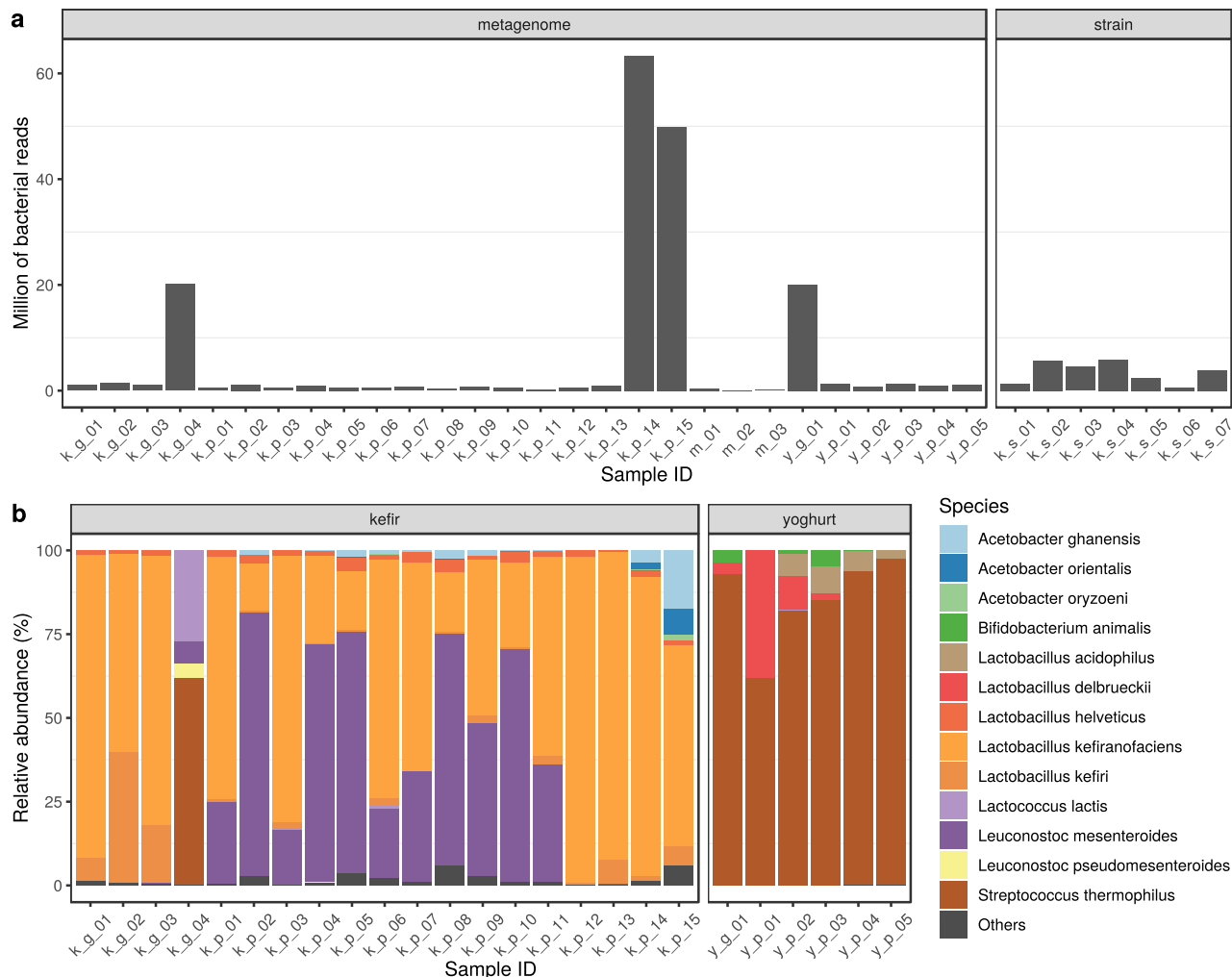
## A glimpse of antimicrobial resistance gene diversity in kefir and yoghurt

Adrienn Gréta Tóth<sup>1</sup>, István Csabai<sup>2</sup>, Gergely Maróti<sup>3,4</sup>, Ákos Jerzsele<sup>5</sup>, Attila Dubecz<sup>6</sup>, Árpád V. Patai<sup>7,8</sup>, Maura Fiona Judge<sup>1</sup>, Sára Ágnes Nagy<sup>1</sup>, László Makrai<sup>9</sup>, Krisztián Bányai<sup>10</sup>, Géza Szita<sup>11</sup> & Norbert Solymosi<sup>1,2</sup>✉

Antimicrobial resistance (AMR) is a global threat gaining more and more practical significance every year. The main determinants of AMR are the antimicrobial resistance genes (ARGs). Since bacteria can share genetic components via horizontal gene transfer, even non-pathogenic bacteria may provide ARG to any pathogens which they become physically close to (e.g. in the human gut). In addition, fermented food naturally contains bacteria in high amounts. In this study, we examined the diversity of ARG content in various kefir and yoghurt samples (products, grains, bacterial strains) using a unified metagenomic approach. We found numerous ARGs of commonly used fermenting bacteria. Even with the strictest filter restrictions, we identified ARGs undermining the efficacy of aminocoumarins, aminoglycosides, carbapenems, cephalosporins, cephamycins, diaminopyrimidines, elfamycins, fluoroquinolones, fosfomycins, glycylicyclines, lincosamides, macrolides, monobactams, nitrofurans, nitroimidazoles, penams, penems, peptides, phenicols, rifamycins, tetracyclines and triclosan. In the case of gene *ImrD*, we detected genetic environment providing mobility of this ARG. Our findings support the theory that during the fermentation process, the ARG content of foods can grow due to bacterial multiplication. The results presented suggest that the starting culture strains of fermented foods should be monitored and selected in order to decrease the intake of ARGs via foods.

Antimicrobial resistance (AMR) is a global threat gaining more and more practical significance every year. Although antimicrobial resistance genes (ARGs) have been present ever since the appearance of the first living microorganisms<sup>1</sup>, the threat arising from AMR has stemmed from the daily use of antibiotics in human and animal healthcare. Even though antibiotic compounds are not directly responsible for the genetic changes behind the appearance of antimicrobial resistance, they place selective pressure towards the amplification of individual bacterial ARGs. Identifying the sources of intake of bacteria carrying ARGs is nowadays a biomedical priority. Bacteria appear in the newborn body right from birth<sup>2</sup>, and later on, their invasion continues from the environment, from other humans and animals, or raw<sup>3</sup> or processed<sup>4</sup> food<sup>5</sup>. Bacteria reaching our gut through alimentation may share functional ARGs either with saprophytes or with pathogens in their physical proximity due to horizontal gene transfer (HGT). In the latter case, resistant or multi-resistant pathogenic strains may evolve, and therapeutic options for the treatment of bacterial infections may narrow. By the production and consumption of livestock products, the opportunity to establish transfer connections between bacteria is provided, as bacterial populations with different sets of ARGs may meet. Therefore, popular probiotic products (such as yoghurt and kefir), have the potential to allow encounters between their bacterial strains and those in the consumer. Yoghurt and kefir are probiotic foods with minor differences in their processing steps. Yoghurt is fermented with bacteria,

<sup>1</sup>Centre for Bioinformatics, University of Veterinary Medicine Budapest, Budapest 1078, Hungary. <sup>2</sup>Department of Physics of Complex Systems, Eötvös Loránd University, Budapest 1117, Hungary. <sup>3</sup>Institute of Plant Biology, Biological Research Center, Szeged 6726, Hungary. <sup>4</sup>Faculty of Water Sciences, University of Public Service, Baja 6050, Hungary. <sup>5</sup>Department of Pharmacology and Toxicology, University of Veterinary Medicine Budapest, Budapest 1078, Hungary. <sup>6</sup>Department of Surgery, Paracelsus Medical University, Nuremberg 90419, Germany. <sup>7</sup>Department of Internal Medicine and Hematology, Semmelweis University, Budapest 1088, Hungary. <sup>8</sup>Interdisciplinary Gastroenterology (IGA) Working Group, Semmelweis University, Budapest 1085, Hungary. <sup>9</sup>Department of Microbiology and Infectious Diseases, University of Veterinary Medicine Budapest, Budapest 1143, Hungary. <sup>10</sup>Institute for Veterinary Medical Research, Centre for Agricultural Research, Budapest 1143, Hungary. <sup>11</sup>Department of Food Hygiene, University of Veterinary Medicine Budapest, Budapest 1078, Hungary. ✉email: solymosi.norbert@gmail.com



**Figure 1.** Bacterial content of the samples. **(a)** The number of reads classified bacterial by Kraken2 on the NCBI NT database. Metagenome includes the samples deriving from grains, milk or products. **(b)** Relative abundances of the most common bacterial species in the grain and product samples.

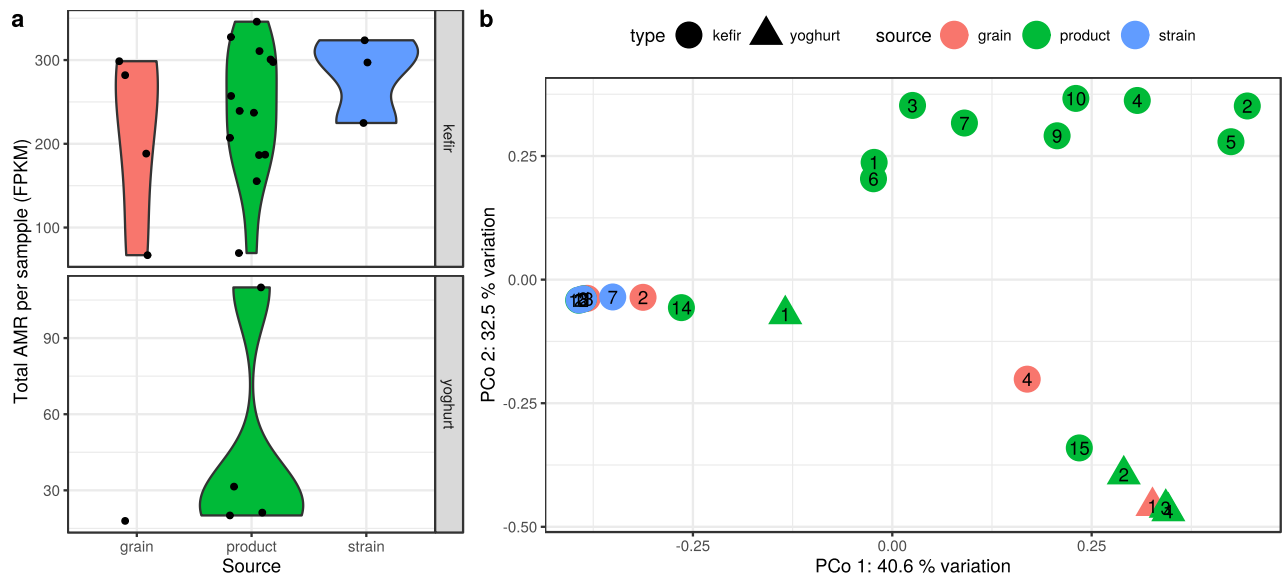
whereas the production of kefir requires fungi in addition. They have both been present in the human diet for a long time and still stand their ground in today's demanding, health-conscious society. Nevertheless, besides the health benefits, consumption of probiotic food may have an adverse effect. Along with the multiplication of bacteria during the fermentation process, the bacterial resistome also grows. If the intake of probiotic products occurs alongside the right triggers, a higher possibility of HGT is provided in the human gut. Our study examined the diversity of the ARG content of kefir and yoghurt products, their grains and bacterial strains using a unified metagenomic approach.

## Results

The results were organised as follows. After describing the bacteriome of the samples, their resistome and mobilome are shown. This is followed by the summary of abundance changes during fermentation of some ARGs and the related bacteria carrying those ARGs. Finally, we present the non-nudged ARG findings.

**Bacteriome.** When classified by taxon, the number of reads aligning to bacterial genomes differed in the various samples (Fig. 1a). Two samples (k\_g\_04 and y\_g\_01 from bioproject PRJNA644779) contained ~20 million reads of bacterial origin. From bioproject PRJNA388572, sample k\_p\_15 had ~50 million bacterial reads, while k\_p\_14 contained more than 63 million. Excluding these four extremities, the average bacterial read count of the metagenomic samples was  $6.7 \times 10^5$  (ranging between  $7.3 \times 10^4$  and  $1.4 \times 10^6$ ). The median sequencing depth of the strain k\_s\_01, k\_s\_02, k\_s\_03, k\_s\_04, k\_s\_05, k\_s\_06, k\_s\_07 were 46, 119, 115, 111, 6, 54, 108, respectively.

Figure 1b. demonstrates the relative abundances of the dominant bacterial species identified in the samples. 99% of all bacteria identified were related to these species. In kefir grains the dominant species were a *Lactobacillus kefiranofaciens* ( $57.7\% \pm 40.5\%$ ), *Lactobacillus kefiri* ( $15.7\% \pm 17\%$ ), *Streptococcus thermophilus* ( $15.4\% \pm 30.8\%$ ), *Lactococcus lactis* ( $6.8\% \pm 13.5\%$ ), *Leuconostoc mesenteroides* ( $1.7\% \pm 3.4\%$ ), *Leuconostoc*



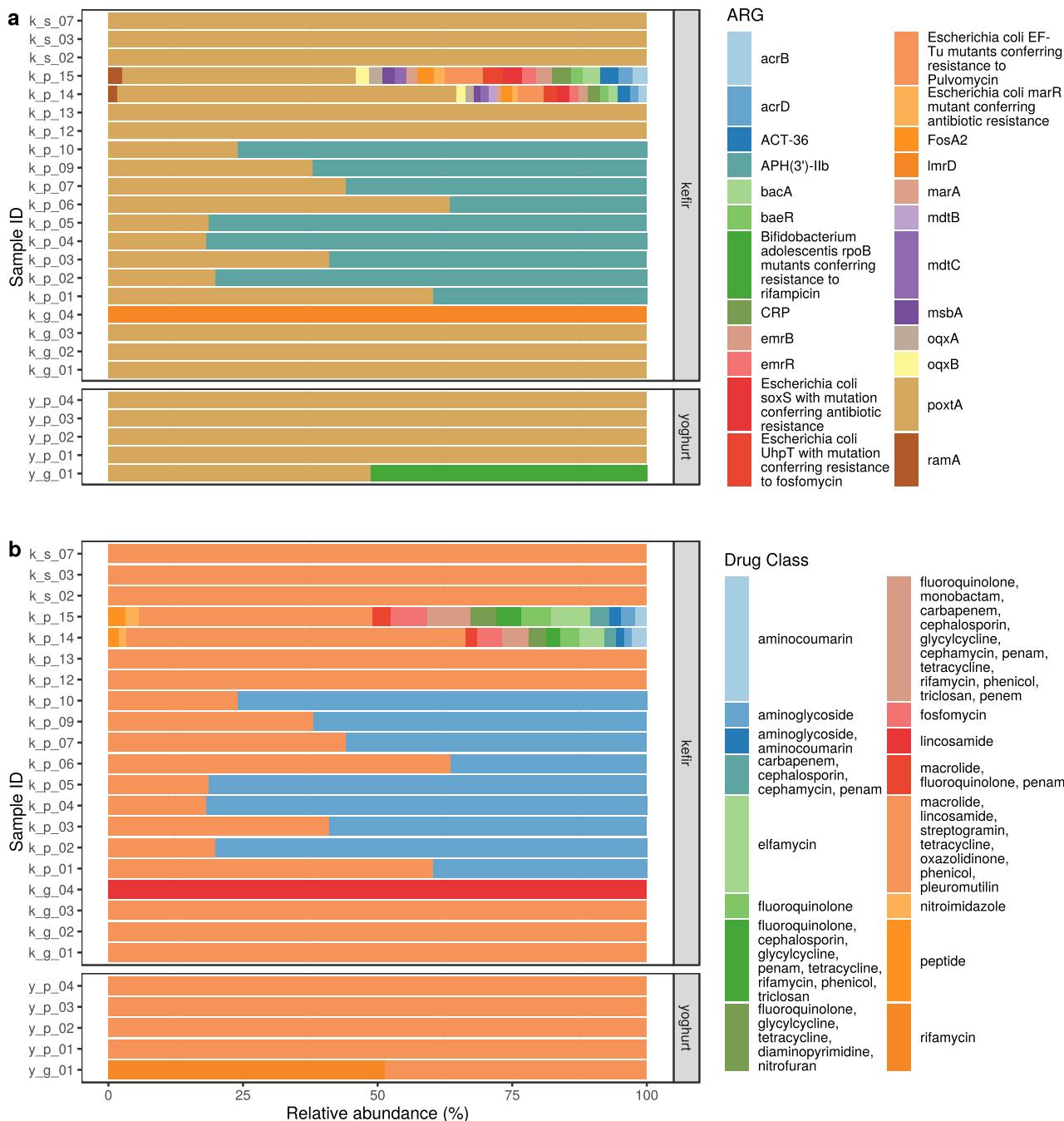
**Figure 2.** Antimicrobial resistance (AMR) abundance of the samples. **(a)** Violin plot representing the distribution of the total AMR fragments per kilobase per million fragments (FPKM) per sample, grouped by type and source. The horizontally jittered dots represent the FPKM of the samples. **(b)** The AMR abundance diversity ( $\beta$ -diversity) of the samples. It is plotted on the first two axes of principal coordinate analysis (PCoA) performed on Bray-Curtis distance which was calculated using the relative abundances of contigs harbouring ARGs. The symbols show the type, the colours the source, while the numbers correspond to the sequence number in the Sample ID. Some samples (k\_s\_01, k\_s\_04, k\_s\_05, k\_s\_06, k\_p\_08) are not shown as their ORFs did not meet filtering criteria.

*pseudomesenteroides* (1%  $\pm$  2%) and *Lactobacillus helveticus* (1%  $\pm$  0.7%) in descending order of abundance. The most significant species in the products overlapped with those in the kefir grains, although they had differences in their relative abundance (*L. kefiranofaciens* (55.4%  $\pm$  29%), *L. mesenteroides* (35.7%  $\pm$  30%), *Acetobacter ghanensis* (2.1%  $\pm$  4.4%), *L. helveticus* (2.1%  $\pm$  1%), *L. kefiri* (1.8%  $\pm$  2%), *Acetobacter orientalis* (0.6%  $\pm$  2%), *Acetobacter oryzoeni* (0.2%  $\pm$  0.5%)). The one yoghurt grain examined was dominated by *Streptococcus thermophilus* (92.8%), *Bifidobacterium animalis* (3.6%) and *Lactobacillus delbrueckii* (3.5%) while the core bacteriome of the yoghurt product consisted of *S. thermophilus* (83.9%  $\pm$  13.8%), *L. delbrueckii* (10.1%  $\pm$  16.2%), *Lactobacillus acidophilus* (4.6%  $\pm$  3.3%) and *B. animalis* (1.2%  $\pm$  2.1%).

**Resistome.** According to our findings based on perfect and strict matches, AMR gene abundances show a great diversity in various types and sources of samples (Fig. 2a). Samples (k\_s\_01, k\_s\_04, k\_s\_05, k\_s\_06, k\_p\_08) that did not meet the ORF filtering criteria were not plotted on Figs. 2 and 3. The highest ARG abundance was observed in the kefir strain samples (average: 282 fragments per kilobase per million fragments (FPKM), sd: 51.1) followed by the kefir product (240  $\pm$  78.6) and the kefir grains (209  $\pm$  106). The yoghurt samples had lower abundances, in the only one grain, FPKM was 17.9, while in the products we found 45.7  $\pm$  32.2.

A Bray-Curtis distance-based principal coordinate analysis (PCoA) was performed to gain insight into the dissimilarity of the sample ARG abundances (Fig. 2b). With a permutational multivariate analysis of variance on the same distance matrix, we found that the type of the sample explains the 22.17% ( $p < 0.001$ ) of dissimilarity among the sample resistomes. For the source grouping, the same measure was 18.92% ( $p < 0.001$ ). Based on Fig. 2b one might conclude that the strongest effect on the dissimilarity is the bioproject of origin, as the analysis showed that it explains 35.56% ( $p < 0.001$ ) of the dissimilarity variances.

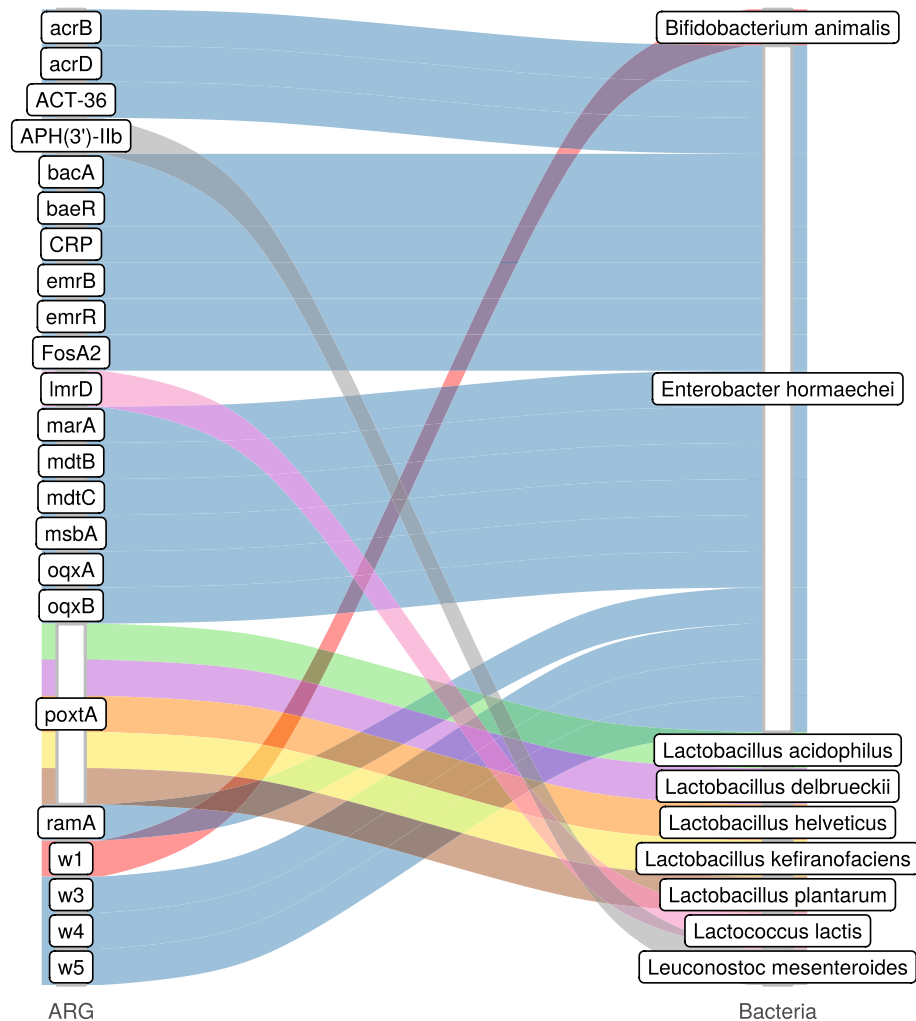
In our kefir samples, we identified 22 ARGs in the product, 2 in the grain and 1 in the strain. In yoghurt, there was 1 ARG in the product and 2 in the grain (Fig. 3a). The relative abundances of antibiotic classes affected are shown in Fig. 3b for each sample. The detected ARGs and their most probable bacteria of origin are summarized on Fig. 4. The kefir ARGs identified in the product may help bacteria in the defence against aminocoumarins, aminoglycosides, carbapenems, cephalosporins, cephamycins, diaminopyrimidines, elfamycins, fluoroquinolones, fosfomycins, glycylcyclines, lincosamides, macrolides, monobactams, nitrofurans, nitroimidazoles, oxazolidinones, penams, penems, peptides, phenicols, pleuromutilins, rifamycins, streptogramins, tetracyclines and triclosan. Contigs containing these ARGs belonged to the genomes of *Enterobacter hormaechei* (genes: *acrB*; *acrD*; *ACT-36*; *bacA*; *baeR*; *CRP*; *emrB*; *emrR*; *Escherichia coli marR* mutant conferring antibiotic resistance; *E. coli soxS* with mutation conferring antibiotic resistance; *E. coli UhpT* with mutation conferring resistance to fosfomycin; *FosA2*; *marA*; *mdtB*; *mdtC*; *msbA*; *oqxA*; *oqxB*; *ramA*), *L. helveticus* (gene *poxxA*), *L. kefiranofaciens* (gene *poxxA*) and *L. mesenteroides* (gene *APH(3')-IIb*). ARGs originating from the kefir grain may play a role in the appearance of resistance against lincosamides, macrolides, oxazolidinones, phenicols, pleuromutilins, streptogramins and tetracyclines and were found in the genomes of *L. kefiranofaciens* (gene *poxxA*) and *L. lactis* (gene *lmrD*). Gene *poxxA* deriving from kefir strains (*L. kefiranofaciens* and *L. plantarum*) confers resistance against



**Figure 3.** Antimicrobial resistance (AMR) abundance in kefir and yoghurt samples. **(a)** Relative abundance of AMR genes. ORFs having at least 60% length and 90% base sequence identity with the reference ARG sequence are shown. Some samples (k\_s\_01, k\_s\_04, k\_s\_05, k\_s\_06, k\_p\_08) are not shown as their ORFs did not meet filtering criteria. **(b)** Relative abundance of drug classes related to the ARGs identified in the samples.

lincosamides, macrolides, oxazolidinones, phenicols, pleuromutilins, streptogramins and tetracyclines. Genes found in yoghurt grains encoded resistance against lincosamides, macrolides, oxazolidinones, phenicols, pleuromutilins, rifamycins, streptogramins and tetracyclines, while the ARGs of the product itself may weaken the efficacy of lincosamides, macrolides, oxazolidinones, phenicols, pleuromutilins, streptogramins and tetracyclines. Contigs involving ARGs of the yoghurt product could have been connected to *L. acidophilus* (gene *poxtA*) and *L. delbrueckii* (gene *poxtA*). However, the ARGs of the grains aligned to the reference sequence of *B. animalis* (gene *Bifidobacterium adolescentis rpoB* mutants conferring resistance to rifampicin) and *L. delbrueckii* (gene *poxtA*).

Based on the ARG abundances the proportion of resistance mechanisms was calculated for each sample. In the kefir product samples the dominant mechanism of identified ARGs was the antibiotic target protection (50.73%), followed by antibiotic inactivation (45.45%), antibiotic efflux (2.03%), antibiotic target alteration (1.07%), antibiotic efflux; reduced permeability to antibiotic (0.32%), antibiotic target alteration; antibiotic efflux; reduced



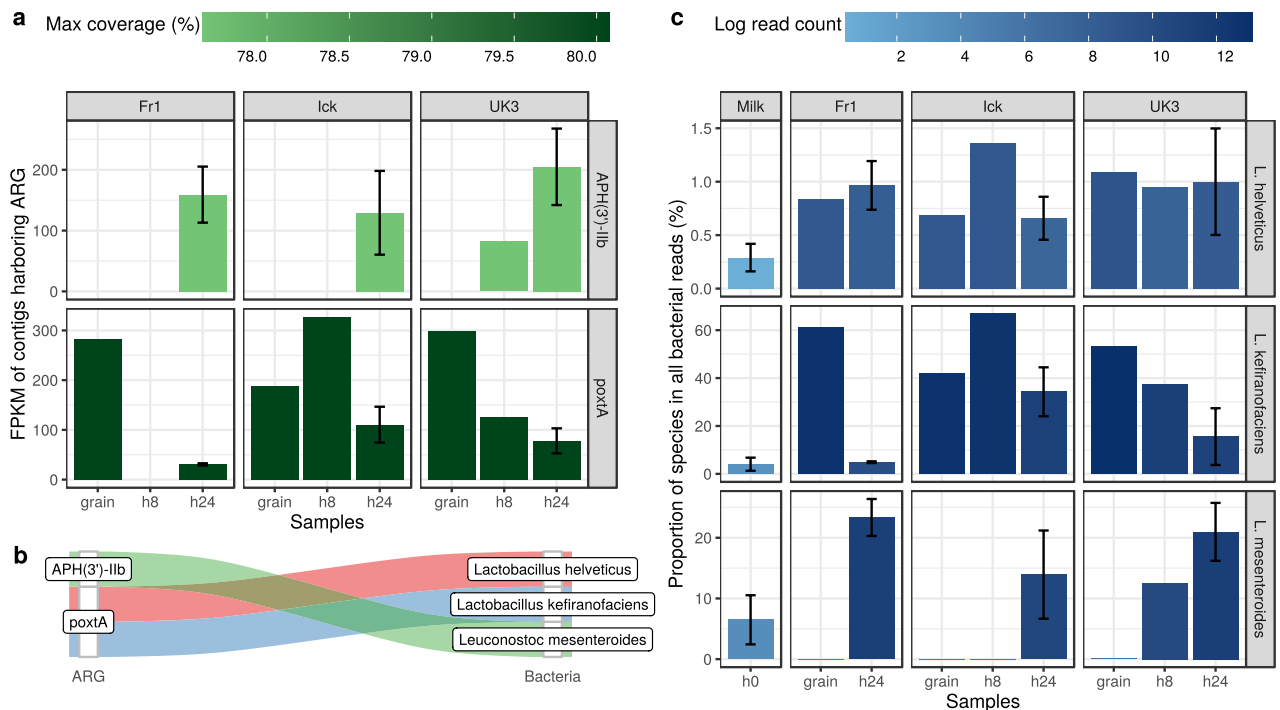
**Figure 4.** Identified ARGs and their most probable bacteria of origin. The gene names that are too long have been abbreviated (w1: *Bifidobacterium adolescentis rpoB* mutants conferring resistance to rifampicin; w3: *Escherichia coli marR* mutant conferring antibiotic resistance; w4: *E. coli soxS* with mutation conferring antibiotic resistance; w5: *E. coli UhpT* with mutation conferring resistance to fosfomycin).

permeability to antibiotic (0.27%), antibiotic target alteration and antibiotic efflux (0.13%). In the kefir grain samples, the main mechanisms were antibiotic target protection (91.98%) and antibiotic efflux (8.02%). In the kefir strains and yoghurt products, the antibiotic target protection was the only mechanism detected. In the one yoghurt grain sample, antibiotic target alteration; antibiotic target replacement (51.28%) and antibiotic target protection (48.72%) are the possible resistance mechanisms.

**Mobile elements.** The results of mobile genetic element (MGE) domain coexisting analysis showed that the ARG *lmrD* in sample k\_g\_04 might be mobile since the contig containing the gene had a transposase ORF within the distance of 10 ORFs. According to the analysis executed with PlasFlow<sup>6</sup> there were not any identifiable contigs with plasmid origins harbouring ARGs.

**ARG abundance changes during kefir fermentation.** According to the metagenomic analysis published by Walsh et al.<sup>7</sup>, ARG abundances change during the fermentation process (Fig. 5a). In the case of all three grains (Fr1, Ick, UK3) *APH(3')-Iib* is present in the kefir fermented for 24 hours, while it is absent in all other time phases except for sample UK3 after 8 hours. *PoxxA* was detectable in all samples except the 8 hour Fr1 kefir sample. The abundance fold change of 24 hours with respect to grain samples was 0.10, 0.59 and 0.26 for the starter culture Fr1, Ick and Uk3, respectively. Between the hour 8 and 24 samples, *poxxA* abundance showed a 0.34-fold change in the Ick kefir, while in the case of the UK3 kefir sequence this value reached 0.62.

Contigs harbouring ARGs were classified taxonomically (Fig. 5c). All contigs containing the gene *APH(3')-Iib* were assigned to *L. mesenteroides*. Contigs with *poxxA* were assigned to the reference genome of *L. helveticus* and *L. kefiranofaciens*.

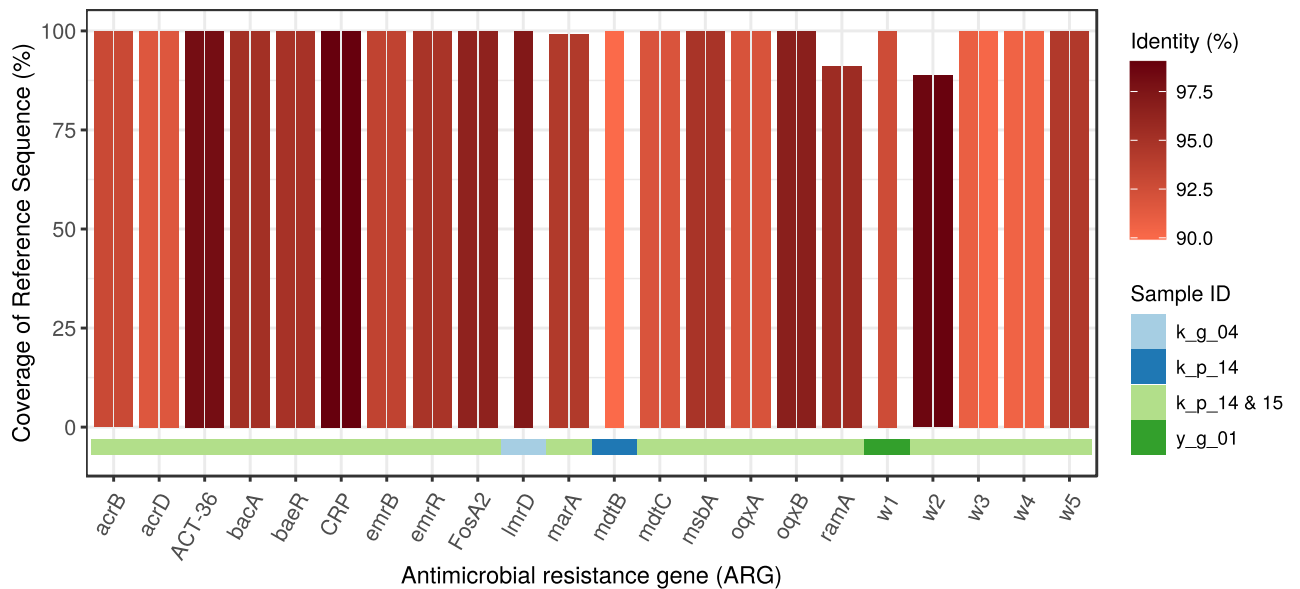


**Figure 5.** Changes during kefir fermentation. **(a)** Antimicrobial resistance gene (ARG) abundance expressed as fragments per kilobase per million fragments (FPKM) based on the alignment of bacterial reads to the ARG harbouring contigs. **(b)** ARGs and their most likely origins. **(c)** Relative abundances of bacteria with a probable ARG content.

All bacterial reads were then aligned to the reference genomes of bacteria mentioned above, and the hits are expressed proportionally (Fig. 5b). In contrast to *L. kefirifaciens* that showed a decreasing tendency, an increase in time is observable by the relative abundances of *L. mesenteroides*. The proportion of reads assigned to *L. helveticus* shows no tendential increase or decrease in time. The increase of abundance of *APH(3)-IIB* shows a positive association with the relative abundance of *L. mesenteroides*. Similarly, *poxA* abundance is decreasing with the relative abundance of *L. kefirifaciens*.

**Excluding nudged hits.** In order to set the alignment restrictions of ORFs to reference ARGs even stricter, we selected for a subgroup of reference ARGs that fit the ORFs from the starting base position on. Thus, nudging on the reference sequence by the alignment was avoided. With such a shrinkage, we reduced the number of detectable ARGs to four samples (Fig. 6). Sample *k\_g\_04* from bioproject PRJNA644779 contained an ARG against lincosamides while the gene found in sample *y\_g\_01* is responsible for resistance against rifamycin. Contigs harbouring these ARGs had the best alignment to *L. lactis* and *B. animalis*, respectively. Bioproject PRJNA388572 had two samples with similar matches, except for gene *mdtB* which appeared in full length in sample *k\_p\_14* and was absent in *k\_p\_15*, this gene is responsible for aminocoumarin resistance. As some other ARGs of the sample also have the potential to confer resistance against aminocoumarin, the AMR profiles of the samples appeared to be the same, including aminocoumarin, aminoglycoside, carbapenem, cephalosporin, cephamycin, diaminopyrimidine, elfamycin, fluoroquinolone, fosfomycin, glycylicline, macrolide, monobactam, nitrofurantoin, nitroimidazole, penam, penem, peptide, phenicol, rifamycin, tetracycline and triclosan resistance. Comparing this list to the nudged results oxazolidinone, pleuromutilin and streptogramin resistance genes were absent. Contigs containing ARGs had the best alignment to the genome of *E. hormaechei* in both cases<sup>8</sup>.

All four samples of both bioprojects included at least 20 million bacterial reads in the assembly of the contigs. The other samples consisted of significantly fewer reads. Consequently, as Sims et al.<sup>9</sup> found, it is not possible to distinguish whether the absence of protein-coding genes or the disruption of open reading frames (ORFs) represent a deficiency of the assembly or real evolutionary gene loss. Examining the relationship between the number of bacterial reads and the length of identified ARGs (including nudges) with a linear model we found that after each extra 100,000 reads the coverage of reference gene raises by 7% by the ARG coding ORFs ( $p < 0.0001$ ). In samples *k\_g\_04*, *k\_p\_14*, *k\_p\_15* and *y\_g\_01*, we randomly chose the average read number of the other samples (677,340) to reanalyze how much these results differ from the original ones executed on the full database. Contigs assembled contained one gene that was identified previously (excluding nudges), namely *lmrD* from sample *k\_g\_04*. ORFs predicted to be ARGs had a median coverage of 16.10% on the reference ARG sequences. In contrast, ORFs aligning to ARGs composed of the full read content of the four samples had a median coverage of 99.21%.



**Figure 6.** Identified ARGs excluding nudged findings. The coverage and identity of detected open reading frames (ORFs) by antimicrobial resistance genes (ARGs). The ORF covered proportion of the reference ARG sequence and the identity % of predicted protein (colour). The gene names which are too long have been abbreviated (w1: *Bifidobacterium adolescentis* *rpoB* mutants conferring resistance to rifampicin; w2: *Escherichia coli* *EF-Tu* mutants conferring resistance to Pulvomycin; w3: *E. coli* *marR* mutant conferring antibiotic resistance; w4: *E. coli* *soxS* with mutation conferring antibiotic resistance; w5: *E. coli* *UhpT* with mutation conferring resistance to fosfomycin).

## Discussion

Studying ARGs that may enter the body with food, including fermented dairy products, can lead to critical health considerations. Characteristics of bacterial diversity and ARG abundance are well observable both in kefir and in yoghurt. ARG abundance is much higher in kefir than in yoghurt. One possible reason for this phenomenon could be the presence of fungi in kefir seed cultures. Since fungi may produce antibacterial toxins, bacteria having ARGs may gain a competitive advantage when coexisting with fungi.

Each bacterium (*Bifidobacterium animalis*<sup>10</sup>, *Enterobacter hormaechei*<sup>8,11</sup>, *Lactobacillus acidophilus*<sup>11,12</sup>, *Lactobacillus delbrueckii*<sup>11,12</sup>, *Lactobacillus helveticus*<sup>11,12</sup>, *Lactobacillus kefirifaciens*<sup>11–13</sup>, *Lactobacillus plantarum*<sup>11,12,14</sup>, *Lactococcus lactis*<sup>11,12,15,16</sup>, *Leuconostoc mesenteroides*<sup>11,12,15,16</sup>) obtained from the taxon classification of contigs containing ARGs is widely used in the production of fermented dairy products. Li et al.<sup>17</sup> analyzed the ARG content of isolated bacteria from fermented drinks and yoghurts. According to their results, *APH(3'')-III*, *APH(6')-APH(2'')*, *sul1*, *tet(M)* were detectable in *Lactobacillus bulgaricus* strains, while *APH(3'')-II*, *sul1*, *sul2*, *strA*, *strB*, *tet(M)* derived from *Streptococcus thermophilus*. In our study *APH(3'')-Iib* gene belonging to the *APH* gene family, supposedly originated from *L. mesenteroides*. Similarly, Carr et al.<sup>18</sup> found a strong co-occurrence between *APH(3'')-Ia* and *L. mesenteroides* in Chinese saliva samples. Further similarity with the results of Carr et al.<sup>18</sup> is that *lmrD* originated from *Lc. lactis*. Guo et al.<sup>19</sup> detected ARGs in *Lactobacillus* strains of fermented milk products. They detected gene *erm(B)*, *gyrA*, *rpoB*, *vanE*, *vanX* in *Lactobacillus casei*, gene *APH(3'')-III*, *dfrD*, *erm(B)*, *gyrA*, *tet(W)*, *vanX* in *L. helveticus*, gene *erm(B)* and *vanX* in *L. plantarum*. We found the *poxA* gene associated with *L. helveticus* and *L. plantarum*. The *emrB* gene was identified in a contig from the genome of *E. hormaechei*.

During the fermentation of milk, the bacteria in seed cultures (and in milk) multiply and dominate the beverage. If any of these bacteria harbour ARGs, the amount of these genes will be increased in the final products. Based on data generated by Walsh et al.<sup>7</sup> *L. helveticus* and *L. kefirifaciens* are the most probable origin of the contigs harbouring gene *poxA*. Sequences containing gene *APH(3'')-Iib* could have been stemmed from *L. mesenteroides*. According to Walsh et al.<sup>7</sup>, during fermentation the relative abundance of *L. kefirifaciens* and *L. mesenteroides* increased. Not surprisingly, in our reanalysis of the same data we found the same trends (Fig. 5b). While Marsh et al.<sup>20</sup> presented same changes of these species in kefir, Wurihan et al.<sup>21</sup> showed opposite alterations in koumiss fermentation. ARG abundances (Fig. 5a) showed a positive association with the relative abundances of their most probable bacteria of origin (Fig. 5b). An increase in the relative abundance of *L. mesenteroides* was followed by the *APH(3'')-Iib* abundance. In contrast, *poxA* abundance dropped simultaneously with the decrease of the relative abundance of *L. kefirifaciens*.

The two most abundant ARGs were *poxA* and *APH(3'')-Iib*, which were both presents in yoghurt and kefir samples. *PoxA* (phenicol-oxazolidine-tetracycline resistance gene), an abundant ARG in Gram-positive bacteria, confers resistance to a wide range of critical antibiotics. The ABC-F class ATP binding ribosomal protection protein encoded by this gene is mainly present in Enterococci and Staphylococci. It was also identified in a methicillin-resistant *Staphylococcus aureus* (MRSA) strain that showed increased MIC to linezolid, a member of the oxazolidinone class of antibiotics<sup>22</sup>. The study highlighted that Staphylococci, Enterococci and interestingly,

*Pediococcus acidilactici* harbouring the gene are all of animal origin and can be spread horizontally with the help of mobile genetic elements. In line with other paper<sup>23</sup> the study suggests that phenicols and other antiribosomal agents used in veterinary medicine might have played a role in the selection of *poxtA*. This was also confirmed by Elghaieb et al.<sup>24</sup>, who identified the gene in cow milk and animal wastewater. As oxazolidinones are prohibited in food animals, and phenicols are not permitted in dairy cattle in Europe, the source of these genes in Hungarian samples remains to be elucidated. *Pseudomonas aeruginosa* harbours an array of aminoglycoside-modifying genes, altering the drug by acetylation, adenylation or phosphorylation (APH). The presence of *APH(3')-IIb* in kefir samples is deliberately worrying as aminoglycoside 3'-phosphotransferases can mediate high-level resistance against several aminoglycosides. These genes might be plasmid-borne or chromosomally encoded; *APH(3')-IIb* is the latter, but a transposon-mediated mechanism has been suggested to be responsible for spreading the resistance genes<sup>25,26</sup>. As the gene was almost exclusively described in *P. aeruginosa* previously, and the likely origin was *L. mesenteroides* in our study, the routes of resistance gene transfer related to this gene need to be further investigated. Although penicillins and cephalosporins are the most frequently used antibiotics for dairy cows, interestingly, the abundance of ARGs facilitating resistance against  $\beta$ -lactams is rather lacking. This phenomenon, together with the ARGs related to unused antibiotics in veterinary dairy medicine, raises the suspicion that the source of the abundant ARGs might not be a direct consequence of antibiotic use at dairy farms.

Bacteria entering the digestive tract with food, are provided with the opportunity of contacting other non-pathogenic and pathogenic bacteria. At the same time, one of the main prerequisites of HGT processes is the physical proximity of the participating bacteria. By virtue of the fulfilment of this requirement, various genes, including ARGs, can be exchanged by bacteria during horizontal gene transfer processes. If an ARG harbours on a mobile DNA-sequence, the probability of its HGT is higher. We found only one gene, namely *lmrD* in sample k\_g\_04, that is supposedly mobile. This deduction is based on the genomic environment of *lmrD*.

Antibiotic resistance caused by multidrug-resistant bacteria is a significant global public health threat<sup>27</sup>. Infections with drug-resistant bacteria may result in major morbidity and mortality and increase the cost of health care when compared to infections by non-resistant strains of the same species. Even with the strictest filtering restrictions, we identified ARGs undermining the efficacy of aminoglycosides, carbapenems, cephalosporins, cephamycins, diaminopyrimidines, elfamycins, fluoroquinolones, fosfomycins, glycolcyclines, lincosamides, macrolides, monobactams, nitrofurans, nitroimidazoles, phenicols, rifamycins and triclosans. These findings raise several clinical considerations. For instance, carbapenems are broad-spectrum antibiotics used for the treatment of necrotizing pancreatitis<sup>28</sup> and severe intraabdominal infections. Tigecycline, a recently developed third-generation tetracycline antibiotic belonging to the glycolcycline class, is one of the few therapeutic options for carbapenem-resistant bacteria, like *Klebsiella pneumoniae*<sup>29,30</sup> and carbapenem-resistant Enterobacteriaceae (CRE)<sup>31</sup>. Another group of ARGs identified in our study code resistance against fluoroquinolones. Emerging fluoroquinolone resistance in *Campylobacter* strains which are the leading cause of bacterial gastroenteritis in the world is a significant public health concern similarly to the rising incidence of fluoroquinolone-resistant cases of typhoid fever and invasive non-typhoidal *Salmonella* (iNTS) infections. We have also identified genes coding cephalosporin resistance in our samples. Cephalosporins belong to the most frequently used antibiotics globally. Intravenous third generation cephalosporins (e.g. ceftriaxone) are more potent against Gram-negative bacteria. They are frequently used in cholecystitis, spontaneous bacterial peritonitis or as a preventive measure in acute gastrointestinal haemorrhage<sup>32</sup>. ORFs harbouring ARGs that code resistance against macrolides may also raise serious public health concerns. Macrolide antibiotics absorb excellently from the gastrointestinal tract and have few side effects. Clarithromycin is still considered as a member of the first-line treatment protocol for *Helicobacter pylori* eradication in areas with a low resistance to clarithromycin<sup>33</sup>. Azithromycin can contribute to the resolution of acute infections by immunomodulatory effects<sup>34</sup>. It is frequently used for the treatment of acute watery or febrile diarrhoea and dysentery syndrome<sup>35</sup>. Tetracycline resistance genes that we found predestine a potential loss in the efficacy of various tetracycline compounds. Once commonly used, nowadays rarely administered tetracycline has been recently rediscovered, as a component of *H. pylori* eradication regimen, partly due to increasing rate of resistance to other antibiotics (including the above-mentioned clarithromycin)<sup>33</sup>.

As ARGs reaching the human body may originate from fermented dairy products, further examinations would be worthwhile to clarify the details and understand the practical medical significance. For this, it would be appropriate to analyze the samples of starter cultures and final products and register the results at set time points during the fermentation period. According to our findings, sequencing depth plays a significant role in the coverage of ORFs identified as ARGs, thus involving at least 20 million clusters is recommended by similar studies. The samples we examined and the studies we found in the literature<sup>3-5,17,19</sup> confirm the hypothesis that foods of animal origin may contain significant amounts of diverse ARGs. The reason for the appearance of ARGs is complex, and the routes of appearance and spread are difficult to track. As sequencing techniques become cheaper, regular genetic monitoring of products of animal origin, including starter cultures, should be considered in addition to the strict control of antibiotics used in animal husbandry.

## Methods

**Data.** The details of analyzed samples are listed in Table 1. One kefir and one yoghurt starter culture were shotgun sequenced (PRJNA644779) by the authors. The further short read datasets were obtained from NCBI SRA repository. A query was performed in SRA to find kefir or yoghurt related shotgun sequenced samples. As a result of this search further 33 datasets originating from 8 BioProjects were selected for the study. Except for the samples of BioProjects PRJEB15432 all others came from paired-end runs. The downloaded short reads originated from BioSamples of kefir grains (n = 4), kefir products (n = 15), kefir strains (n = 7), a yoghurt grain (n = 1) and yoghurt products (n = 5). In the unified names of the samples the first character corresponds to the type of the sample (k and y, kefir and yoghurt, respectively), the second character comes from the first letter



Sample ID	BioProject	Run	Type	Source	Sample
k_g_01	PRJEB15432	ERR1653138	kefir	Grain	Fr1 grain
k_g_02	PRJEB15432	ERR1653139	kefir	Grain	Ick grain
k_g_03	PRJEB15432	ERR1653140	kefir	Grain	UK3 grain
k_g_04	PRJNA644779	SRR12171332	kefir	Grain	kefir seed culture
k_p_01	PRJEB15432	ERR1653129	kefir	Product	UK3, 8 h
k_p_02	PRJEB15432	ERR1653130	kefir	Product	Fr1, 24 h (replicate 2)
k_p_03	PRJEB15432	ERR1653131	kefir	Product	Ick, 24 h (replicate 2)
k_p_04	PRJEB15432	ERR1653132	kefir	Product	UK3, 24 h (replicate 2)
k_p_05	PRJEB15432	ERR1653135	kefir	Product	Fr1, 24 h (replicate 3)
k_p_06	PRJEB15432	ERR1653136	kefir	Product	Ick, 24 h (replicate 3)
k_p_07	PRJEB15432	ERR1653137	kefir	Product	UK3, 24 h (replicate 3)
k_p_08	PRJEB15432	ERR1653141	kefir	Product	Fr1, 24 h (replicate 1)
k_p_09	PRJEB15432	ERR1653142	kefir	Product	Ick, 24 h (replicate 1)
k_p_10	PRJEB15432	ERR1653143	kefir	Product	UK3, 24 h (replicate 1)
k_p_11	PRJEB15432	ERR1653145	kefir	Product	Fr1, 8 h
k_p_12	PRJEB15432	ERR1653146	kefir	Product	Ick, 8 h
k_p_13	PRJNA288044	SRR2082409	kefir	Product	KEFIR.shotgun
k_p_14	PRJNA388572	SRR7287342	kefir	Product	Metagenome from probiotic beverage K03
k_p_15	PRJNA388572	SRR8282406	kefir	Product	Metagenome from probiotic beverage K02
k_s_01	PRJDB4955	DRR064132	kefir	Strain	<i>Lactobacillus parakefiri</i> JCM 8573
k_s_02	PRJNA222257	SRR1151211	kefir	Strain	<i>Lactobacillus kefiranofaciens</i> subsp. <i>kefiranofaciens</i> DSM 5016
k_s_03	PRJNA222257	SRR1151212	kefir	Strain	<i>Lactobacillus kefiranofaciens</i> subsp. <i>kefirgranum</i> DSM 10550
k_s_04	PRJNA222257	SRR1151213	kefir	Strain	<i>Lactobacillus kefiri</i> DSM 20587
k_s_05	PRJNA222257	SRR1151226	kefir	Strain	<i>Lactobacillus parakefiri</i> DSM 10551
k_s_06	PRJNA635855	SRR11965732	kefir	Strain	<i>Acetobacter syzygii</i> str. K03D05
k_s_07	PRJNA635872	SRR11966381	kefir	Strain	<i>Lactobacillus plantarum</i> K03D08
m_01	PRJEB15432	ERR1653133	milk	Milk	0 h (replicate 1)
m_02	PRJEB15432	ERR1653134	milk	Milk	0 h (replicate 2)
m_03	PRJEB15432	ERR1653144	milk	Milk	0 h (replicate 3)
y_g_01	PRJNA644779	SRR12171305	yoghurt	Grain	yoghurt seed culture
y_p_01	PRJEB30083	ERR2982980	yoghurt	Product	Yoghurt-A
y_p_02	PRJEB30083	ERR2982981	yoghurt	Product	Yoghurt-B
y_p_03	PRJEB30083	ERR2982982	yoghurt	Product	Yoghurt-C
y_p_04	PRJEB30083	ERR2982983	yoghurt	Product	Yoghurt-D
y_p_05	PRJEB30083	ERR2982984	yoghurt	Product	Yoghurt-E

**Table 1.** The list of analyzed samples obtained from NCBI SRA. In the unified names of the samples the first character corresponds to the type of the sample (k and y, kefir and yoghurt, respectively), the second character comes from the first letter of the source (g, p and s for grain, product and strain, respectively), while the last tag is a sequence number. The last column shows the available attribute data of the biosamples.

of the source (g, p and s for grain, product and strain, respectively), while the last tag is a sequence number. Of the collected projects, a peer-reviewed publication is available for the PRJNA222257<sup>36</sup>, PRJEB15432<sup>7</sup> and PRJEB30083<sup>37</sup>. For all other samples, the only accessible metadata were the attributes in SRA. In PRJEB15432 Walsh et al.<sup>7</sup> followed the microbial changes during the fermentation process of kefir. They used full-fat pasteurized milk inoculated by three different grains (Fr1, Ick, and UK3 from France, Ireland and United Kingdom, respectively). The pasteurized milk (with three replications) and grains (without replication) were sampled at hour 0. In the fermentation from kefir at hour 8 (without replication) and hour 24 (with three replications), further specimens were taken.

**DNA extraction and metagenomics library preparation for PRJNA644779.** Total metagenome DNA of kefir (k\_g\_04) and yoghurt (y\_g\_01) samples were extracted using the UltraClean Microbial DNA Isolation kit from MoBio Laboratories. The quality of the isolated total metagenomic DNA was checked using an Agilent TapeStation 2200 instrument. The DNA samples were used for in vitro fragment library preparation. In vitro fragment libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina. Paired-end fragment reads were generated on an Illumina NextSeq sequencer using TG NextSeq 500/550 High Output Kit v2 (300 cycles). Read numbers were the following: 22,044,496 and 20,895,112 for kefir and yoghurt, respectively. Primary data analysis (base-calling) was carried out with Bbcl2fastq software (v2.17.1.14, Illumina).

**Bioinformatic and statistical analysis.** Quality based filtering and trimming was performed by Trimmomatic<sup>38</sup>, using 15 as a quality threshold. Only reads longer than 50 bp were retained. The remaining reads were taxonomically classified using Kraken2 ( $k = 35$ )<sup>39</sup> with the NCBI non-redundant nucleotide database<sup>40</sup> with two different confidence setting. The first run was performed with the default settings to select all possible bacterial reads. The following taxon classification was performed with the `-confidence 0.5` parameter to get more precise species level hits. The taxon classification data was managed in R<sup>41</sup> using functions of the package phyloseq<sup>42</sup> and microbiome<sup>43</sup>. For further analysis, only reads assigned to Bacteria Kingdom was used<sup>44</sup>. The preprocessed bacterial reads were assembled to contigs by MEGAHIT (v1.2.9)<sup>45</sup> using default settings. From the contigs, all possible open reading frames (ORFs) were gathered by Prodigal<sup>46</sup>. The protein translated ORFs were aligned to the ARGs of database CARD v.3.0.9<sup>47,48</sup> by Resistance Gene Identifier (RGI, v5.1.0) with Diamond<sup>49</sup>. The ORFs classified as perfect or strict were further filtered with 90% identity and 60% coverage. The findings were presented including and excluding the nudged hits. For the analysis of ARG abundance changes during kefir fermentation, only ARGs with maximal coverage greater than 75% in samples taken at different time points were included. Contigs harbouring ARGs were classified by Kraken2 using the NCBI RefSeq<sup>50</sup> complete bacterial genomes database. In keeping with Hendriksen et al.<sup>1</sup> the ARG abundance was expressed as fragments per kilobase per million fragments (FPKM)<sup>51</sup> of contigs containing ARGs. For the  $i$ th contig  $FPKM_i = q_i / (l_i \times Q) \times 10^6$ , where  $q_i$  is the number of reads that mapped to the contig,  $l_i$  is the length of contig and  $Q$  is the total number of mapped reads. To calculate  $q$  values, all bacterial reads were aligned to the contigs by Bowtie2<sup>52</sup> with the parameter of `-very-sensitive-local`. To identify possible further mobile genetic element (MGE) homologs the predicted protein sequences of contigs were scanned by HMMER<sup>53</sup> against data of PFAM v32<sup>54</sup> and TnpPred<sup>55</sup>. Similar to Saenz et al.<sup>44</sup> from the hits with lower than  $E 10^{-5}$  the best ones were assigned to each predicted protein within the distance of 10 ORFs. The MGE domains coexisting with ARGs were categorized as phage integrase, resolvase, transposase or transposon. The plasmid origin probability of the contigs was estimated by PlasFlow v.1.1<sup>6</sup>. According to the ARG abundance of the samples, a dissimilarity matrix was calculated using the Bray-Curtis index<sup>56</sup> with package vegan<sup>57</sup>. With the same library and the same matrix, a permutational multivariate analysis of variance was applied to quantify the associations between the dissimilarity and independent variables (type, source, BioProject). For the visualization of the sample distances based on this matrix, a principal coordinate analysis (PCoA) was performed with package ape<sup>58</sup>. The relationship between the detected ORF length and the sequencing depth was explored using a linear model. All analyses and plotting were done in R-environment<sup>41</sup>.

## Data availability

The short read data of sample data are publicly available and can be accessed through the PRJDB4955, PRJEB15432, PRJEB30083, PRJNA222257, PRJNA288044, PRJNA388572, PRJNA635855, PRJNA635872, PRJNA644779 from the NCBI Sequence Read Archive (SRA).

Received: 23 September 2020; Accepted: 21 December 2020

Published online: 31 December 2020

## References

- Hendriksen, R. S. *et al.* Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **10**, 1124 (2019).
- Shao, Y. *et al.* Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574**, 117–121 (2019).
- Tóth, A. G. *et al.* Antimicrobial resistance genes in raw milk for human consumption. *Sci. Rep.* **10**, 1–7 (2020).
- Machado, M. *et al.* Antibiotic resistance and biofilm production in catalase-positive gram-positive cocci isolated from Brazilian pasteurized milk. *J. Food Qual. Hazards Control* **7**, 67–74 (2020).
- de Alcântara Rodrigues, I., Ferrari, R. G., Panzenhagen, P. H. N., Mano, S. B. & Conte-Junior, C. A. Antimicrobial resistance genes in bacteria from animal-based foods. *Adv. Appl. Microbiol.* **112**, 143 (2020).
- Krawczyk, P. S., Lipinski, L. & Dziembowski, A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res.* **46**, e35–e35 (2018).
- Walsh, A. M. *et al.* Microbial succession and flavor production in the fermented dairy beverage kefir. *Msystems* **1**, e00052-16 (2016).
- Dertli, E. & Çon, A. H. Microbial diversity of traditional kefir grains and their role on kefir aroma. *LWT-Food Sci. Technol.* **85**, 151–157 (2017).
- Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121–132 (2014).
- Gueimonde, M. *et al.* Viability and diversity of probiotic lactobacillus and bifidobacterium populations included in commercial fermented milks. *Food Res. Int.* **37**, 839–850 (2004).
- Van Wyk, J.K. The champagne of fermented beverages. In *Fermented Beverages*, 473–527 (Elsevier, 2019).
- Bengoia, A. A., Iraporda, C., Garrote, G. L. & Abraham, A. G. Kefir micro-organisms: their role in grain assembly and health properties of fermented milk. *J. Appl. Microbiol.* **126**, 686–700 (2019).
- Wang, S.-Y. *et al.* Investigation of microorganisms involved in biosynthesis of the kefir grain. *Food Microbiol.* **32**, 274–285 (2012).
- Zheng, Y. *et al.* Probiotic properties of lactobacillus strains isolated from Tibetan kefir grains. *PLoS ONE* **8**, e69868 (2013).
- Garrote, G. L., Abraham, A. G. & De Antoni, G. L. Chemical and microbiological characterisation of kefir grains. *J. Dairy Res.* **68**, 639 (2001).
- Witthuhn, R., Schoeman, T. & Britz, T. Characterisation of the microbial population at different stages of kefir production and kefir grain mass cultivation. *Int. Dairy J.* **15**, 383–389 (2005).
- Li, Y. *et al.* Antibiotic resistance of *Lactobacillus* spp. and *Streptococcus thermophilus* isolated from Chinese fermented milk products. *Foodborne Pathog. Dis.* **16**, 221–228 (2019).
- Carr, V. R. *et al.* Abundance and diversity of resistomes differ between healthy human oral cavities and gut. *Nat. Commun.* **11**, 1–10 (2020).
- Guo, H. *et al.* Characterization of antibiotic resistance genes from *Lactobacillus* isolated from traditional dairy products. *J. Food Sci.* **82**, 724–730 (2017).

20. Marsh, A. J., O'Sullivan, O., Hill, C., Ross, R. P. & Cotter, P. D. Sequencing-based analysis of the bacterial and fungal composition of kefir grains and milks from multiple sources. *PLoS ONE* **8**, e69371 (2013).
21. Wurihan, W. *et al.* Bacterial community succession and metabolite changes during the fermentation of koumiss, a traditional mongolian fermented beverage. *Int. Dairy J.* **98**, 1–8 (2019).
22. Antonelli, A. *et al.* Characterization of poxta, a novel phenicol-oxazolidinone-tetracycline resistance gene from an mrsa of clinical origin. *J. Antimicrob. Chemother.* **73**, 1763–1769 (2018).
23. Hao, H. *et al.* The risk of some veterinary antimicrobial agents on public health associated with antimicrobial resistance and their molecular basis. *Front. Microbiol.* **7**, 1626 (2016).
24. Elghaieb, H. *et al.* Dispersal of linezolid-resistant enterococci carrying poxta or optra in retail meat and food-producing animals from tunisia. *J. Antimicrob. Chemother.* **74**, 2865–2869 (2019).
25. Hächler, H., Santanam, P. & Kayser, F. H. Sequence and characterization of a novel chromosomal aminoglycoside phosphotransferase gene, aph (3')-iib, in pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* **40**, 1254–1256 (1996).
26. Zeng, L. & Jin, S. aph (3)-iib, a gene encoding an aminoglycoside-modifying enzyme, is under the positive control of surrogate regulator hpaa. *Antimicrob. Agents Chemother.* **47**, 3867–3876 (2003).
27. WHO. Essential medicines and health products. <https://www.who.int/medicines/publications/essentialmedicines/en>. Accessed September 1st 2020.
28. Baron, T. H., DiMaio, C. J., Wang, A. Y. & Morgan, K. A. American gastroenterological association clinical practice update: management of pancreatic necrosis. *Gastroenterology* **158**, 67–75 (2020).
29. Xu, L. *et al.* Efficacy and safety of tigecycline for patients with hospital-acquired pneumonia. *Chemotherapy* **61**, 323–330 (2016).
30. Wang, J., Pan, Y., Shen, J. & Xu, Y. The efficacy and safety of tigecycline for the treatment of bloodstream infections: a systematic review and meta-analysis. *Ann. Clin. Microbiol. Antimicrob.* **16**, 24 (2017).
31. Ni, W. *et al.* Tigecycline treatment for carbapenem-resistant enterobacteriaceae infections: a systematic review and meta-analysis. *Medicine* **95**, e3126 (2016).
32. MacIntosh, T. Emergency management of spontaneous bacterial peritonitis—a clinical review. *Cureus* **10**, e2253 (2018).
33. Malfertheiner, P. *et al.* Management of helicobacter pylori infection—the maastricht v/florence consensus report. *Gut* **66**, 6–30 (2017).
34. Parnham, M. J. *et al.* Azithromycin: mechanisms of action and their relevance for clinical applications. *Pharmacol. Ther.* **143**, 225–245 (2014).
35. Tribble, D. R. Antibiotic therapy for acute watery diarrhea and dysentery. *Mil. Med.* **182**, 17–25 (2017).
36. Sun, Z. *et al.* Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat. Commun.* **6**, 1–13 (2015).
37. Lordan, R. *et al.* The effect of ovine milk fermentation on the antithrombotic properties of polar lipids. *J. Funct. Foods* **54**, 289–300 (2019).
38. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
39. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with kraken 2. *Genome Biol.* **20**, 257 (2019).
40. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–4 (2005).
41. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2020).
42. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, 1–11 (2013).
43. Lahti, L. & Shetty, S. microbiome R package (2012–2019). <https://bioconductor.org/packages/release/bioc/html/microbiome.html>.
44. Sáenz, J. S. *et al.* Oral administration of antibiotics increased the potential mobility of bacterial resistance genes in the gut of the fish *Piaractus mesopotamicus*. *Microbiome* **7**, 24 (2019).
45. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. Megahit: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
46. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
47. McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57**, 3348–57 (2013).
48. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
49. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using diamond. *Nat. Methods* **12**, 59–60 (2015).
50. O'Leary, N. A. *et al.* Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
51. Trapnell, C. *et al.* Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
53. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121–e121 (2013).
54. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
55. Riadi, G., Medina-Moenne, C. & Holmes, D. S. TnpPred: a web service for the robust prediction of prokaryotic transposases. *Int. J. Genom.* **2012**, 678761 (2012).
56. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349. <https://doi.org/10.2307/1942268> (1957).
57. Oksanen, J. *et al.* *vegan: Community Ecology Package* (2019). R package version 2.5-6.
58. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

## Acknowledgements

The project is supported by the European Union and co-financed by the European Social Fund (No. EFOP-3.6.3-VEKOP-16-2017-00005). It has also received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 874735 (VEO). GM received support from the Hungarian Academy of Sciences through the Lendület-Programme (LP2020-5/2020).

## Author contributions

N.S. takes responsibility for the integrity of the data and the accuracy of the data analysis. A.G.T., I.C., L.M. and N.S. conceived the concept of the study. G.M. and N.S. performed sample collection and procedures. A.G.T., N.S. and S.A.N. participated in the bioinformatic and statistical analysis. A.D., A.G.T., A.J., A.V.P., G.M., G.S., K.B. and N.S. participated in the drafting of the manuscript. A.D., A.G.T., A.J., A.V.P., G.M., G.S., I.C., K.B., M.F.J.

and N.S. carried out the critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80444-5>.

**Correspondence** and requests for materials should be addressed to N.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020