

## ORIGINAL ARTICLE

# Stable-isotope probing and metagenomics reveal predation by protozoa drives *E. coli* removal in slow sand filters

Sarah-Jane Haig<sup>1</sup>, Melanie Schirmer<sup>1</sup>, Rosalinda D'Amore<sup>2</sup>, Joseph Gibbs<sup>3</sup>, Robert L Davies<sup>4</sup>, Gavin Collins<sup>1,3</sup> and Christopher Quince<sup>1</sup>

<sup>1</sup>School of Engineering, University of Glasgow, Glasgow, UK; <sup>2</sup>School of Biological Sciences, University of Liverpool, Liverpool, UK; <sup>3</sup>Microbial Ecophysiology Laboratory, School of Natural Sciences and Ryan Institute, National University of Ireland, Galway, Ireland and <sup>4</sup>Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

**Stable-isotope probing and metagenomics were applied to study samples taken from laboratory-scale slow sand filters 0.5, 1, 2, 3 and 4 h after challenging with <sup>13</sup>C-labelled *Escherichia coli* to determine the mechanisms and organisms responsible for coliform removal. Before spiking, the filters had been continuously operated for 7 weeks using water from the River Kelvin, Glasgow as their influent source. Direct counts and quantitative PCR assays revealed a clear predator–prey response between protozoa and *E. coli*. The importance of top-down trophic-interactions was confirmed by metagenomic analysis, identifying several protozoan and viral species connected to *E. coli* attrition, with protozoan grazing responsible for the majority of the removal. In addition to top-down mechanisms, indirect mechanisms, such as algal reactive oxygen species-induced lysis, and mutualistic interactions between algae and fungi, were also associated with coliform removal. The findings significantly further our understanding of the processes and trophic interactions underpinning *E. coli* removal. This study provides an example for similar studies, and the opportunity to better understand, manage and enhance *E. coli* removal by allowing the creation of more complex trophic interaction models.**

*The ISME Journal* (2015) 9, 797–808; doi:10.1038/ismej.2014.175; published online 3 October 2014

## Introduction

Slow sand filtration is one of the oldest and most effective means of treating drinking water for the control of microbiological contamination (for example, achieving >99% removal of enteric bacteria, Hijnen *et al.* (2007)). Such purification is attributed to naturally occurring biochemical processes in the filters (for example, predation and bio-oxidation), however, these have not yet been comprehensively verified, mainly due to methodological limitations (Haig *et al.*, 2011). Previously, Haig *et al.*, (2014a) compared laboratory-scale filters to full-scale sand filters using phyla-specific quantitative PCR primers, and both 454 and Illumina sequencing. This revealed that the microbial communities underpinning slow sand filters (SSFs) are extremely complex, with specific organisms correlating with overall water quality performance (Haig *et al.*, 2014a). Further, it was found that the treatment efficiency and microbial

community structure of full-scale units were reproducible in the laboratory (Haig *et al.*, 2014a), a finding which now allows more pertinent questions relating to human health and microbial ecology to be addressed. In particular, understanding how pathogenic microorganisms (for example, *E. coli*) are removed is a critical question, which addresses one of the primary tasks of modern ecology; linking the biotic interactions of organisms within an ecosystem to their functional performance (Mikola and Setälä, 1998).

The need to remove pathogens and understand the mechanisms responsible for pathogen removal in potable water supplies is a well-recognised issue, emphasized by the fact that ~3.4 million people each year die from water-related diseases (WHO, 2004). Determining and understanding these mechanisms would be highly advantageous and would vastly improve the implementation of drinking water technologies in developing countries, including household systems. In addition, it could allow water companies in developed countries to control pathogen levels by managing the SSF community. Further, by determining the trophic mechanisms and interactions involved in *E. coli* removal in a 'real world' food-web, great insight and knowledge for general microbial ecology will be obtained.

Correspondence: S-J Haig, School of Engineering, University of Glasgow, Rankine Building, 79–85 Oakfield Avenue, Glasgow G128LT, UK.

E-mail: s.haig.1@research.gla.ac.uk

Received 22 April 2014; revised 15 July 2014; accepted 17 August 2014; published online 3 October 2014

This will provide a paradigm for similar studies and the opportunity to create more realistic trophic interaction models in the future.

Previous SSF studies have examined the ability of specific organisms (for example, Chrysophyte) to remove pathogenic bacteria (Weber-Shirk and Dick, 1999), or the overall pathogen removal efficiency of SSFs (Bomo *et al.*, 2004, Grobe *et al.*, 2006, Hijnen *et al.*, 2007; Elliott *et al.*, 2008). However, these studies are limited by their specificity. Further, on the basis of these studies, and knowledge from marine and terrestrial environments, both top-down (predation by protozoa and viral lysis) and bottom-up (nutrient/resource availability) mechanisms have been suggested as important for the regulation of microbial mortality (Lloyd, 1973; Hunter and Price, 1992; Pace and Cole, 1994; Weber-Shirk and Dick, 1999; Rosemond *et al.*, 2001). In addition, theoretical models and empirical surveys have indicated that the majority of the mortality is due to grazing by protists, and to a lesser extent to viral lysis (Pernthaler, 2005). However, abiotic factors, such as UV radiation and reactive oxygen species (ROS)-associated lysis, have also been hypothesized as potential lysis routes for microbes/pathogens (Curtis *et al.*, 1992; Alonso-Saez *et al.*, 2006; Liu *et al.*, 2007; Kadir and Nelson, 2014).

Although these studies are informative, they are also unrealistic as they have been performed in microcosms, focussing on one or a small group of organisms and hence over-simplify and potentially provide inaccurate or biased conclusions on regulatory mechanisms. Currently, no study, to the author's knowledge, exists which aims to determine the mechanisms responsible for pathogen removal in a real biological system without prior knowledge as to which removal mechanisms or organisms to target. However, the complexity of real communities requires an untargeted approach capable of quantifying the importance of all trophic groups simultaneously. Here, we develop such an approach by combining stable-isotope probing (SIP) with metagenomics (Sul *et al.*, 2009) and apply it to the tractable, though complex, system in SSFs, allowing all mechanisms and organisms involved in the removal of non-pathogenic *E. coli* K12 to be determined. We will use this organism, a commonly used faecal indicator, as a proxy for true pathogens, such as other *E. coli* strains, making the assumption that the removal processes will be the same. The experiment was used to test the hypothesis that the principal modes of removal will be top-down removal mechanisms, such as predation by protozoa and viral lysis, although the extent of these processes is expected to differ throughout time.

## Materials and methods

### *Filter operation and sampling*

The same SSF set-up (eight filters of 2.5 m in height and 54 mm in diameter) and operational procedures,

as employed by Haig *et al.*, (2014a), were used in this study. The only difference was the addition of high-power LED lights fitted with a cool white (240 lumens or 5.4 W) bulb, erected above the SSFs to simulate daylight conditions. These lights functioned on a 12 h light/12 h dark cycle for the duration of the experiment, with times being regulated by a digital electronic timer. The source of water feeding each of the filters was the River Kelvin in Glasgow, and was supplied at a constant filtration rate of  $0.15 \text{ m}^3 \text{ m}^{-2} \text{ h}^{-1}$ , which is consistent with full-scale SSFs. The sand used in all filters was sourced from a full-scale SSF site (Haig *et al.*, 2014a) and was sterilized by autoclaving at  $121^\circ \text{C}$  for 20 min before being put into the eight laboratory-scale SSFs. As in Haig *et al.*, (2014a), water quality analyses were performed weekly.

### *Spiking with isotopically labelled *E. coli**

After 7 weeks of operation, each of the filters was spiked with isotopically labelled *E. coli* K12 (transformed TOP10 strain) following the protocol outlined in Marley *et al.* (2001). Briefly, *E. coli* K12 was grown overnight in M9 minimal medium with 20 ml of filter-sterilized 20% (w/v)  $^{13}\text{C}$ -glucose (Sigma, Dorset, UK) as the sole carbon source at  $37^\circ \text{C}$ , with shaking at 200 r.p.m. The overnight culture was then centrifuged at 3000 g for 10 min and washed twice with sterile PBS before resuspension in autoclaved river water at a concentration of  $300 \text{ c.f.u. ml}^{-1}$ , 5 min before spiking into the SSFs. Spiking entailed feeding the isotopically labelled *E. coli* to all filters for 1 h at the same filtration rate used previously ( $0.15 \text{ m}^3 \text{ m}^{-2} \text{ h}^{-1}$ ), after which normal filter operation resumed with non-spiked, non-autoclaved river water. The concentration of *E. coli* used was approximately 10 times the normal concentration found in the river water and was chosen to mimic levels found during pollution and storm run-off events.

### *Sampling spiked filters*

To determine the mechanisms responsible for *E. coli* removal, sand was sampled from the filters at depths (1, 5, 10, 15 cm) and times of 0.5, 1, 2, 3 and 4 h after spiking. In addition, all depths (0.5, 10, 15, 20, 30, 45, 70 cm) were sampled from the filters 24 and 96 h after spiking. Sand samples (0.5 g wet weight) were used for: direct *E. coli* plate counts on membrane lauryl sulphate agar containing  $100 \mu\text{g ml}^{-1}$  ampicillin,  $50 \mu\text{g ml}^{-1}$  kanamycin and  $25 \mu\text{g ml}^{-1}$  streptomycin (Life Technologies, Glosgow, UK); direct protozoa quantification following the procedure of (Dehority, 1984); and SIP in conjunction with metagenomic sequencing.

### *DNA-stable-isotope probing*

To separate the labelled ( $^{13}\text{C}$ ) and unlabelled ( $^{12}\text{C}$ ) DNA, the procedure of Neufeld *et al.* (2007) was

used. Separation was achieved by using density gradient fractionation of the total DNA extract (50  $\mu\text{L}$ ) on a CsCl gradient with a buoyant density of  $1.725\text{ g ml}^{-1}$  that was subjected to ultracentrifugation in a Sorvall 100SE Ultracentrifuge (Thermo Scientific, Loughborough, UK) at 44 100 r.p.m. for 40 h at  $20^\circ\text{C}$ . The density gradient was fractionated into 12 aliquots ( $\sim 400\text{ }\mu\text{L}$  each) by a drop-wise collection method, where fractions were taken from the bottom of the ultracentrifugation tube by pumping water into the top of the tube with a constant-flow ( $500\text{ }\mu\text{L min}^{-1}$ ) syringe pump (Gilson's Miniplus 2 peristaltic pump). The density of the resulting fractions was measured with an AR200 refractometer (Reichert, Munich, Germany) and ranged from  $1.47$  to  $1.86\text{ g ml}^{-1}$  with a median density of  $1.68\text{ g ml}^{-1}$ . Fractions were precipitated using a polyethylene glycol solution and dissolved in  $30\text{ }\mu\text{L}$  of TE buffer, and used for qPCR quantification of 18S rRNA, total 16S and *E. coli* specific 16S rRNA genes (Supplementary Information 1). On the basis of qPCR and density profiles of the samples compared with  $^{12}\text{C}$  and  $^{13}\text{C}$  controls, two fractions from each sample, one representing labelled (density:  $>1.68\text{ g ml}^{-1}$ ) DNA and one representing non-labelled (density  $<1.68\text{ g ml}^{-1}$ ) DNA, were chosen for metagenomic library construction and analysis (Supplementary Information 2 and 3).

#### *Illumina metagenomic library preparation*

Thirty-six Illumina libraries (18 pairs of labelled  $^{13}\text{C}$  and non-labelled  $^{12}\text{C}$  fractions from various filters and time points) were prepared using the Nextera XT kits (Illumina, Essex, UK), following the manufacturers instructions. Briefly,  $5\text{ }\mu\text{L}$  ( $0.2\text{ ng }\mu\text{L}^{-1}$ ) of extracted DNA were tagged and then subjected to PCR using specific index primers and common adaptors (P5 and P7). Amplified libraries were cleaned using the AMPure XP beads (Beckman Coulter, High Wycombe, UK) and eluted in a final volume of  $12\text{ }\mu\text{L}$  TE. Libraries were checked for their fragment size distribution and concentration using a Bioanalyzer 2100 (Agilent, Cheshire, UK), and appropriate libraries were size selected (500–800 bp) using a Pippin Prep (Sage Science, Beverly, MA, USA) with a 1.5% cassette. Size-selected libraries were pooled using equimolar quantities to obtain the desired number of reads for each sample. The pool was sequenced on a HiSeq 2000 (Illumina) at the Centre for Genomic Research (Liverpool).

#### *Metagenomic sequence analysis*

Sequenced reads for each sample were quality trimmed using sickle [<https://github.com/najoshi/sickle>]. Quality profiles were constructed with FastQC (Andrews, 2010), which revealed a non-uniform distribution of nucleotides at the start of the reads, indicating the possible partial remainder of adapter or transposon sequences. Therefore, the first 20 bp of the MiSeq and 16 bp of the HiSeq reads

were trimmed and reads were filtered based on a minimum read length of 80 bp and 40 bp for MiSeq and HiSeq, respectively (originally MiSeq: 150 bp and HiSeq: 99 bp). Resulting samples contained on average 8, 157,  $287 \pm 1$ , 373, 145 reads.

For taxonomic classification, paired-end reads were converted to a format suitable for MEGAN (Huson *et al.*, 2011). LAST (Frith *et al.*, 2010) was used to align the reads (maximum of 20 matches) against a customized subset of the NCBI database containing the microbial, protozoan, viral and fungal databases to achieve a more time-efficient analysis. The output was converted into 'blast format' and piped into MEGAN where the lowest common ancestor was assigned to each read (lowest common ancestor parameters: max-matches = 100, min-score = 35.0, top-percent = 10.0, win-score = 0.0, min-support = 1, min-complexity = 0.3). Occurrence tables of the taxonomic assignments were generated using a custom designed script in which the last column of the MEGAN output files was converted into the corresponding 'taxid' and the taxonomic path was inferred by utilizing the perl library Bio::LITE::Taxonomy. Directly exporting taxonomic paths with MEGAN caused problems, particularly for eukaryotes and viruses as not all taxonomic levels were defined. In addition, Bio::LITE::Taxonomy was unable to resolve issues due to synonyms present in the database. Therefore, the taxonomic paths of all 'taxids', which were unresolved by Bio::LITE::Taxonomy were directly inferred from the NCBI taxonomy database (names.dmp). Furthermore, the full taxonomic paths of several organisms (for example, *Monosiga brevicollis* and *Dictyostelium discoideum*) were added to the database.

#### *Statistics*

To determine which organisms had a significant role in *E. coli* removal, pairwise similarities among samples based on the Bray–Curtis similarity index were calculated. The resulting matrices were examined for temporal patterns and differences between  $^{13}\text{C}$  (labelled) and  $^{12}\text{C}$  (non-labelled) samples through non-metric multidimensional scaling and canonical correspondence analysis as implemented in the Vegan package in R (Oksanen *et al.*, 2012). Significant differences in the metagenomic community composition between different time points (0.5–4 h) and carbon sources (labelled— $^{13}\text{C}$  and non-labelled— $^{12}\text{C}$ ) after spiking with *E. coli* were determined using the Adonis function in the Vegan package, which performs a nonparametric multivariate analysis of variance (Anderson, 2001). To determine individual contributions from each taxon to the differences between labelled and non-labelled samples, and for the various time points, SIMPER analysis was used (Clarke, 1993). SIMPER analysis is a useful measure of the magnitude of difference; however, to decide whether a taxa

differed significantly, pairwise *t*-tests (Kendall non-parametric) adjusted for multiple comparisons using the Benjamini–Hochberg false-discovery method were performed. Only taxa with a false-discovery rate of <5% were reported.

## Results

To determine the magnitude with which protozoa and other eukaryotes affect *E. coli* removal; recombinant *E. coli* and protozoa counts, total and *E. coli* specific 16S rRNA, and 18S rRNA, qPCR assays were performed on samples taken from SSFs challenged with isotopically labelled ( $^{13}\text{C}$ ) *E. coli*. To resolve which organisms (bacteria, eukaryotes and viruses) were responsible for *E. coli* removal, the different carbon densities ( $^{12}\text{C}$  and  $^{13}\text{C}$ ) in the samples were separated (DNA-SIP) and used for metagenomic analysis. Increased abundance of any organism in  $^{13}\text{C}$ -labelled samples indicated potential involvement in *E. coli* removal.

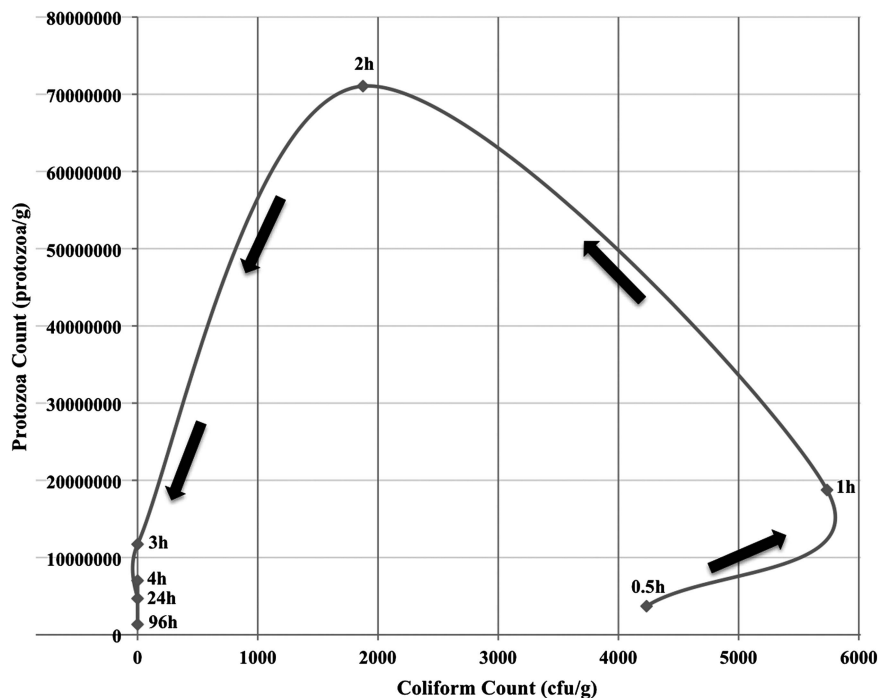
### *Protozoan predator–prey response—direct counts and qPCR*

Direct counts of *E. coli* and total protozoa (Figure 1) revealed a clear predator–prey relationship, with most removal occurring at the top of the filters around 2–3 h after spiking with  $^{13}\text{C}$ -labelled *E. coli*. The gradual decrease over time in the abundance of  $^{13}\text{C}$ -labelled *E. coli*, as well as the peak in the number of labelled ( $^{13}\text{C}$ ) 18S rRNA copies in qPCR

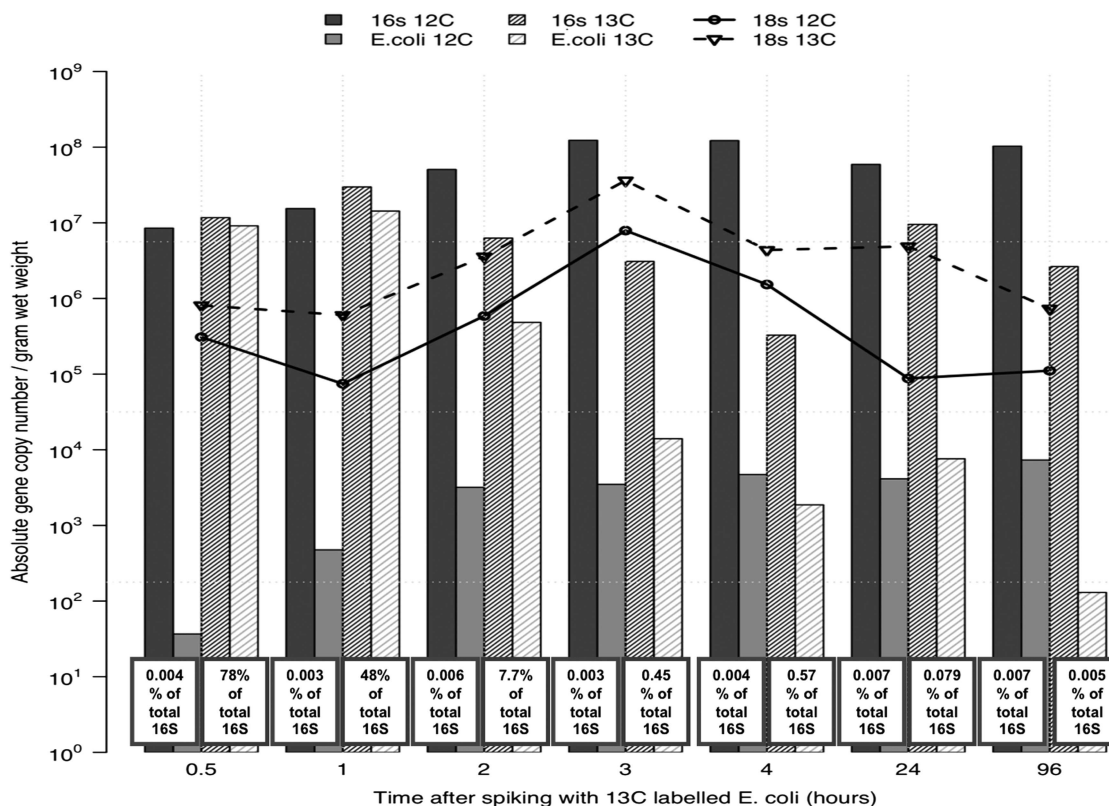
assays on samples 3 and 24 h after spiking (Figure 2), suggested that protozoan grazing is a major mode of *E. coli* removal. However, as the  $^{12}\text{C}$  18S rRNA results showed a similar trend compared with the labelled 18S rRNA, it can be assumed that incomplete  $^{13}\text{C}$  incorporation has occurred, that is, samples with a density resembling normal carbon ( $^{12}\text{C}$ ) may have started to incorporate labelled  $^{13}\text{C}$  but have not incorporated enough into their biomass to cause a density change.

### *All domains of life are important for E. coli removal*

To resolve which organisms (bacteria, eukaryotes and viruses) were responsible for *E. coli* removal, DNA-SIP in conjunction with metagenomic analysis was employed. Metagenomics, unlike qPCR and conventional sequencing approaches, does not rely on prior knowledge of the organisms of interest and, thus, organism-specific primers are not required (Wooley *et al.*, 2010). However, metagenomics does still suffer from biases introduced during DNA extraction, enzymatic cutting during library preparation and PCR amplification. Multivariate analysis of variance revealed significant differences in the metagenomic communities at all levels of taxonomic classification between: types ( $^{13}\text{C}$  or  $^{12}\text{C}$ ), and times after spiking, of the 36 samples taken from four SSFs after adding the  $^{13}\text{C}$ -labelled *E. coli*. Time was the most significant variable explaining 15.4% (*P*-value: 0.01) of the variance followed by type (11.3% *P*-value: 0.001). Nonparametric *t*-tests on individual taxa proportions adjusted for multiple comparisons



**Figure 1** Scatter plot showing the predator–prey response between protozoa and *E. coli*.



**Figure 2** Absolute numbers of 16S, 18S and *E. coli* specific 16S rRNA in  $^{12}\text{C}$  and  $^{13}\text{C}$  fractions, determined by qPCR assays.

identified 10 orders (two bacterial, six eukaryotic and two viral) as being statistically significant in explaining differences between  $^{12}\text{C}$  and  $^{13}\text{C}$  metagenomic communities (Figure 3). We then investigated differences at higher levels of taxonomic resolutions within the prokaryotes, eukaryotes and viruses separately. The number of reads associated with each time point (Supplementary Information 4) within each group ( $^{12}\text{C}$  and  $^{13}\text{C}$ ) was not significantly different ( $P$ -value: 0.9356), and both fractions had similar overall community structures ( $^{12}\text{C}$ : 0.09% viruses, 9.18% eukaryotes and 90.73% prokaryotes;  $^{13}\text{C}$ : 0.08% viruses, 8.50% eukaryotes and 91.42% prokaryotes).

#### The importance of viral lysis for *E. coli* removal

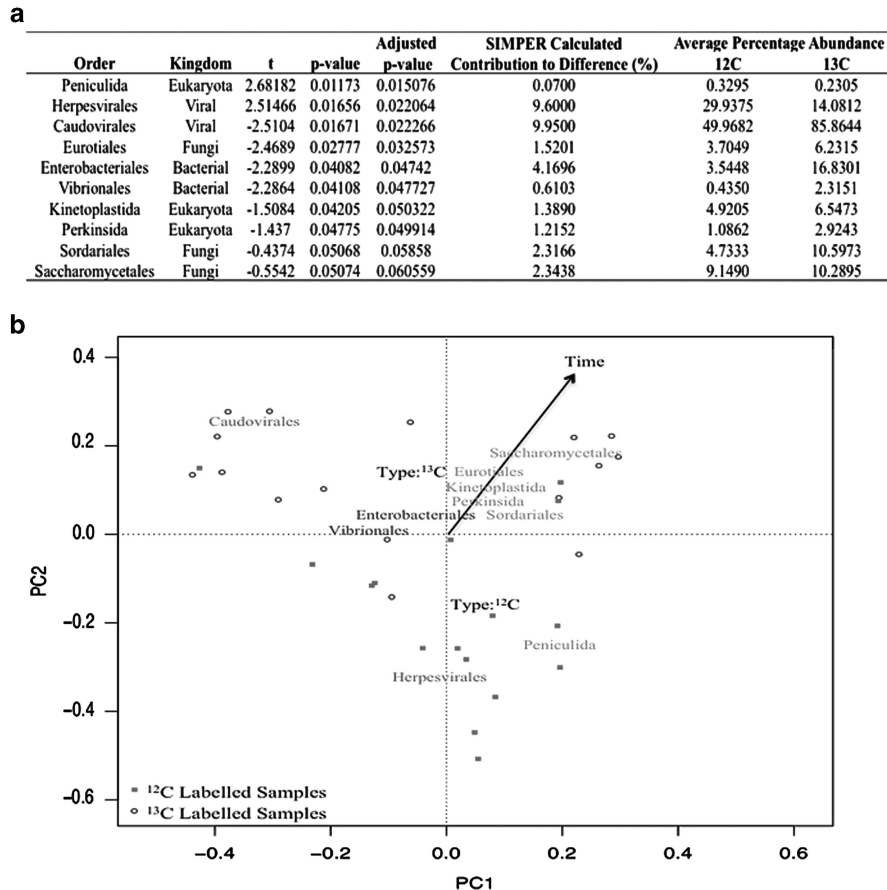
Twenty-two viral species (Supplementary Information 5) were identified by pairwise  $t$ -tests and Benjamini–Hochberg false-discovery tests as being significantly different between  $^{12}\text{C}$  and  $^{13}\text{C}$  metagenomic communities. None were prophages found within the *E. coli* K12 genome. Overall, 15 of these viruses were present at higher abundances in  $^{13}\text{C}$  communities compared with  $^{12}\text{C}$  samples and hence are involved directly or indirectly in *E. coli* removal. Collectively these accounted for 22% of the dissimilarity between  $^{13}\text{C}$  and  $^{12}\text{C}$  samples. Visually referring to the 22 significant viral species (Figures 4a and b) the importance of *Enterobacteria* phages is

apparent, and in particular, *Enterobacteria phage lambda*, *Enterobacteria phage cdtI* and *Enterobacteria phage N15*, which account for over 14% of the difference between labelled and non-labelled samples (Supplementary Information 5). In particular, the abundance of *E. phage lambda* was over 117 times more abundant in  $^{13}\text{C}$  samples than  $^{12}\text{C}$  communities, implying its importance in *E. coli* removal.

#### The importance of protozoan grazing for *E. coli* removal

Following the same approach previously used to identify significant viral species involved with *E. coli* removal and  $^{13}\text{C}$  metabolism, 52 eukaryotic species were identified (Supplementary Information 6), of which 20 were protozoa (Figures 4c and d), 15 algae (Figures 5a and b) and 17 fungi (Figures 5c and d). The presence of significant species from all eukaryotic kingdoms underscores the complex mechanisms involved with *E. coli* removal.

The 20 significant protozoan species represented 15 different genera and members from the flagellate, ciliate and amoeboid groups, all of which are known predators of *E. coli* (Weekers *et al.*, 1993; Fleck *et al.*, 2000; Fey *et al.*, 2007; Cassidy-Hanley, 2012; Yue *et al.*, 2013). Referring to Figure 4c, it can be seen that the proportion of significant protozoan species in  $^{13}\text{C}$  samples increased over time, compared with



**Figure 3** (a) Relative abundances of statistically significant orders of organism which explain differences between <sup>12</sup>C and <sup>13</sup>C samples. Adjusted *P*-values calculated using the Benjamini–Hochberg method, average percentages are relative to the kingdom the order belongs to. (b) Principle component analysis of all orders of organism identified from the metagenomic analysis. Green text represents viral orders, brown text represents bacterial orders and orange text represents eukaryotic orders.

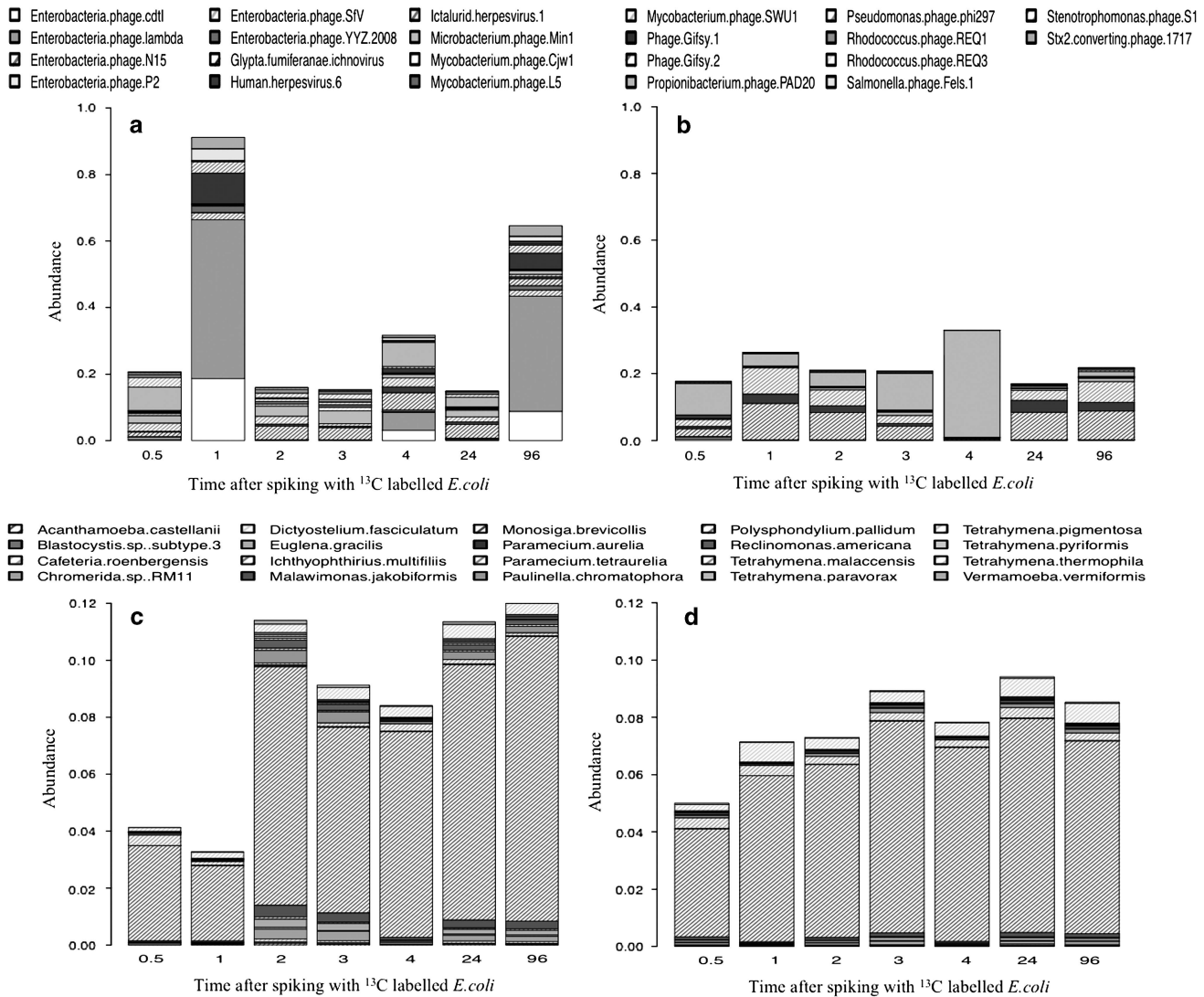
the relatively stable proportion of 7% in non-labelled (<sup>12</sup>C) samples (Figure 4d). Further, the biggest difference between <sup>13</sup>C and <sup>12</sup>C communities are due to fluctuations in the populations of: *Chromerida RM11*, *Euglena gracilis*, *Malawimonas jakobiformis*, *M. brevicollis*, *Paulinella chromatophora*, *Reclinomonas americana*, *Tetrahymena paravorax* and *Vermamoeba vermiformis*, (Figures 4c and d) all of which are highly motile and possess voracious appetites. Specifically, the importance of protozoan grazing on labelled *E. coli*, and hence *E. coli* removal, is apparent 2 h post spiking, where a large increase in the protozoan population is observed. In particular, a large increase in the proportion of *M. brevicollis* and *Tetrahymena spp.* was observed, with both genera collectively being responsible for 2% of the dissimilarity between <sup>12</sup>C and <sup>13</sup>C communities (Supplementary Information 6).

#### *The mutualistic relationship of fungi and algae*

The abundances of 17 fungal and 15 algal species were identified as being significantly different

between <sup>12</sup>C and <sup>13</sup>C samples, of which only the fungal species *Naumovozyma castelli* was present in greater proportions in non-labelled samples. Therefore, the remaining species appear to be involved in *E. coli* removal and/or <sup>13</sup>C-labelled carbon metabolism. On initial analysis of the algal and fungal abundances, a staggering similarity in community dynamics can be seen, with both communities increasing and decreasing in abundance at the same time points (Figure 5); Kendall correlation tests of all significant fungal and algal species confirmed this to be a significant relationship (tau 0.81952, *P*-value: 0.01071). This mirrored behaviour is indicative of a mutualistic relationship. Such symbioses between algae and fungi have been widely documented in various environments (Harte and Kinzig, 1993; Danger *et al.*, 2013).

The 15 significant algal species represented nine different genera, with *Saccharina spp.* accounting for 47% of the significant algal species. Clear shifts in the abundance of significant algae can be seen between <sup>12</sup>C and <sup>13</sup>C samples at all time points (Figures 5a and b), in particular, at 2, 3, 24 and 96 h after spiking with <sup>13</sup>C-labelled *E. coli*, where the average abundance tripled in <sup>13</sup>C samples compared



**Figure 4** Stacked barplot of the abundance changes of statistically significant viral (a, b) and protozoan (c, d) species at various time points after spiking with  $^{13}\text{C}$ -labelled *E. coli*. A and C are  $^{13}\text{C}$ -labelled samples and B and D are  $^{12}\text{C}$ -labelled samples.

with the non-labelled samples. In addition, fungal species followed the same trend with *Sordaria macrospora* and *Chaetomium globosum* showing the largest increase (Figures 5c and d).

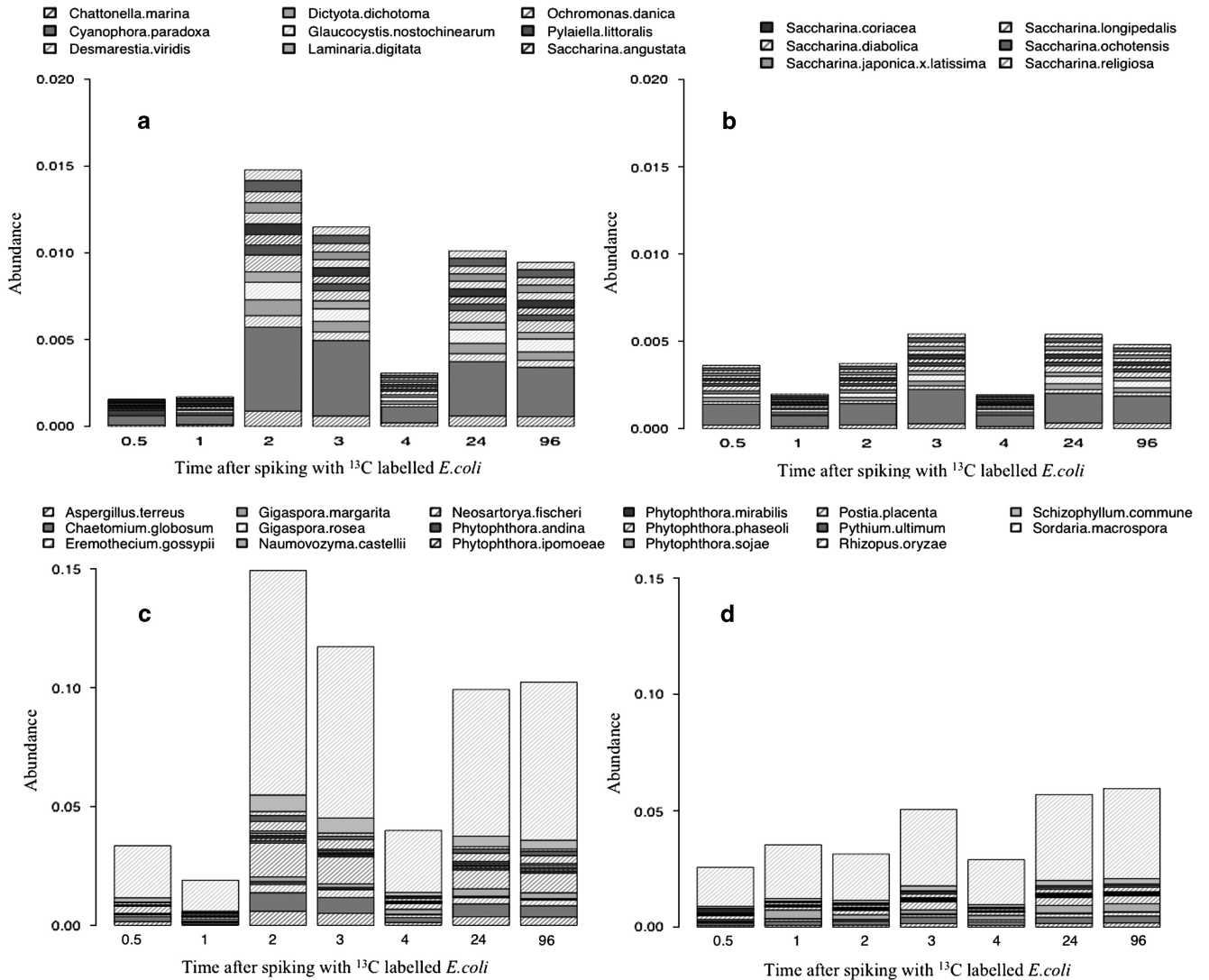
#### Calculating the importance of viruses and protozoa

To approximate the importance of protozoa and viruses in *E. coli* removal, the following assumptions were made:

1. The added  $^{13}\text{C}$ -labelled *E. coli* does not undergo replication
2. Only protozoa and viruses are responsible for *E. coli* removal
3. All protozoa and viruses are grazing/infecting *E. coli* at a constant rate
4. *E. phage lambda* is used to represent all viruses (as it is the most abundant and significant virus identified in  $^{13}\text{C}$  samples)

5. *M. brevicollis* and *Tetrahymena thermophila* are used to average the abilities of protozoa (as they are the two most significant and abundant protozoa in  $^{13}\text{C}$  samples)
6. For a protozoa to become  $^{13}\text{C}$  labelled, 50% of its grazing consumption must be from  $^{13}\text{C}$ -labelled *E. coli*

On the basis of these assumptions and taking genome size and progeny production into consideration it was concluded (Supplementary Information 7) that protozoan grazing appeared to be the major driving force behind *E. coli* removal (99.86%), with *M. brevicollis* accounting for the biggest proportion of *E. coli* removal (24.83%), followed by 4.68% achieved by *T. thermophila*. Conversely, viral-associated lysis was responsible for 0.14% removal of which *E. phage lambda* was responsible for 0.076%, which was 326 times smaller than the removal achieved by *M. brevicollis*.



**Figure 5** Stacked barplot of the abundance changes of statistically significant algal (a, b) and fungal (c, d) species at various time points after spiking with  $^{13}\text{C}$ -labelled *E. coli*. A and C are  $^{13}\text{C}$ -labelled samples and B and D are  $^{12}\text{C}$ -labelled samples.

## Discussion

Identifying and unpicking trophic interactions, particularly those involved with pathogen removal, is an extremely complex and important question. Previously, mathematical models and work in simplified microcosms have shown the individual importance of viral lysis, protozoan grazing and endogenous and exogenous ROS in *E. coli* removal (Curtis *et al.*, 1992; Bomo *et al.* 2004; Grobe *et al.*, 2006; Liu *et al.*, 2007; Kadir and Nelson, 2014). Within this study, the level of involvement in *E. coli* removal of each kingdom has been approximated from the metagenomic analysis following several assumptions. Furthermore, to optimize for the success of the DNA-SIP approach, that is, obtain enough genomic material for metagenomic sequencing, a higher concentration of *E. coli* than normally found in surface water was spiked into the SSFs. However, this is the first

study, to our knowledge, to examine and identify the ecosystem-wide trophic interactions and mechanisms responsible for *E. coli* removal in a 'real' system, without prior bias as to which organisms and mechanisms to target. In addition, this study is the first to show that it is possible to isotopically label *E. coli* and follow removal through an ecosystem.

### *Top-down trophic interactions are essential for E. coli removal*

On the basis of the direct counts, qPCR (Figures 1 and 2) and DNA-SIP metagenomic sequencing, the importance of top-down regulatory mechanisms for *E. coli* removal is apparent. Among a consortium of phages, protozoa, fungi and algae, which were  $^{13}\text{C}$  labelled (hence involved in *E. coli* removal/metabolism) and identified as highly significant, *E. phage lambda*, *M. brevicollis* and *T. sp.* were



identified as the main organisms responsible for *E. coli* removal. On the basis of our calculations (Supplementary Information 7) and direct count observations (Figure 1), protozoan grazing was responsible for more than 99% of the *E. coli* removal within 4 h of spiking, which is consistent with previous investigations into pathogen removal in constructed wetlands (Weber and Legge, 2008) and estuaries (McCambridge and McMeekin, 1980). From the 20 statistically significant protozoan species identified (Supplementary Information 5), *M. brevicollis* was predicted to be responsible for the majority (24.83%) of the removal once factors such as replication rate, progeny production, grazing rates and genome size were taken into consideration. This equated to ~326 times more *E. coli* removal than that achieved by *E. phage lambda*. The lowered involvement of viral-associated lysis for *E. coli* removal is also consistent with previous work (Withey *et al.*, 2005).

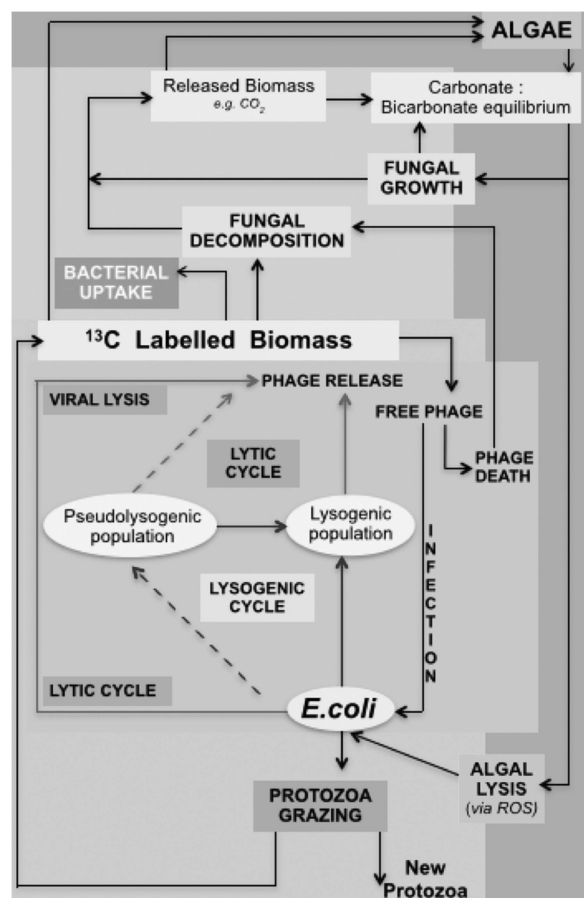
The importance, and dominance, of *M. brevicollis* from the beginning to the end of the experiment is not surprising as it is known to have a critical role in marine global carbon cycling (Yue *et al.*, 2013). Further, its dominance over other protozoan grazers including *T. spp.* (responsible for 4.68% of *E. coli* removal) may be explained by its very short doubling time of 4.6 h (Christaki *et al.*, 2005) and fast grazing rate of 196 bacterial cells per hour (Parry, 2004), which in theory would allow them to outcompete other identified significant protozoan species that were identified. Although the feeding rate of *M. brevicollis* is slower than *T. spp.*, such dominance could be explained by *M. brevicollis*:

1. Possessing a microvilli collar, which holds bacteria from the water flow and allows them to be engulfed at a later time (Yue *et al.*, 2013), hence providing energy storage for less plentiful times.
2. Possessing six oxidative stress genes; four of which are algal in origin—two ascorbate peroxidases and two metacaspases—that help protect the protozoa from various algal-mediated ROS (Nedelcu *et al.*, 2008).

Although protozoan grazing appeared to be the major route for *E. coli* removal, the role of viral-associated lysis cannot be overlooked, especially, as *E. phage lambda* was more abundant in <sup>13</sup>C-labelled samples (Figures 4a and b) and was identified by SIMPER analysis to contribute to more than 10% of the dissimilarity between <sup>12</sup>C and <sup>13</sup>C communities (Supplementary Information 5).

Overall, with the exception of the *Microbacterium*, *Pseudomonas*, *Salmonella* and *Rhodococcus phages*, the remaining eight statistically significant phages are all known prophages of *E. coli* and have been shown to reduce *Enterobacteria* (including *E. coli*) numbers. However, unlike the

protozoa, viral abundances changed dramatically with time (Figure 4), with increased abundance at 1, 4 and 96 h and lower abundance at the remaining time points. Such behaviour suggests that these viruses are fluctuating between states of pseudolysogeny/lysogeny and lytic pathways (Figure 6), behaviour widely documented for environmental phages (Abedon, 2008). This choice of life-cycle has been shown to aid in the regulation of bacterial biomass (Ripp and Miller, 1997), which is not taken into consideration in our calculations of removal. In addition, the extremely high abundance of phages 1 h after incubation with <sup>13</sup>C-labelled *E. coli* was surprising; however, Zeng and Golding (2011) showed that *E. phage lambda* can infect, replicate and enter the lysogenic cycle within *E. coli* after only 80 min. Therefore, although viral lysis has been shown to be responsible for only 0.14% of the *E. coli*



**Figure 6** Trophic interaction diagrams involved with *E. coli* removal. Food-web showing ecosystem-wide involvement and trophic interactions involved in *E. coli* removal. Red shaded text shows the progressive relationship between host bacterial cells and different bacteriophage life cycles, adapted from Abedon [2008]. Red lines indicate lytic pathways and blue lines indicate lysogenic pathways. Purple shading represents predation by protozoa. Blue and green shading represents the mutualistic relationship between algae and fungi. The involvement of and production of ROS by interactions between algae and fungi are hypothesized based on work from other authors.

removal (Supplementary Information 7), it is likely that the phages are supporting the regulation of the population (by allowing the *E. coli* population to recover) to ensure sufficient hosts for subsequent viral infection (Abedon, 2008). In addition, the rapid appearance of *E. coli* phages in this experiment may reflect the heightened metabolic state of introduced *E. coli* at the start of the spiking period due to glucose availability.

#### *Ecosystem-wide associations are needed for successful E. coli removal*

Although top-down regulatory trophic interactions, such as protozoan grazing and viral lysis are the major mechanisms for *E. coli* removal, the importance of indirect and abiotic mechanisms cannot be overlooked. For example, previous work has shown algae to be actively involved in *E. coli* removal by the production ROS (Curtis *et al.*, 1992; Maynard *et al.*, 1999; Feng *et al.*, 2011), which causes lysis of *E. coli* and other bacterial species. In particular, extensive work has shown that *Chattonella marina* (one of the algae which significantly increased in labelled samples; Figure 5a) produces several ROS species (Liu *et al.*, 2007) known to significantly reduce coliform numbers. Therefore, it is conceivable that algae are actively participating in *E. coli* removal by indirect mechanisms (Figure 6). Furthermore, as 11 out of the 15 significant algal species are mixotrophs (Nasr *et al.*, 1968, Semple and Cain, 1996) it is likely that they accessed carbon from the released biomass from the  $^{13}\text{C}$ -labelled *E. coli* (via viral lysis and protozoan grazing).

In addition to indirect bacterial autolysis induced by algal ROS, these products help to explain the dominance of *M. brevicollis* (which contains ROS protection genes). This autolysis, alongside protozoa and viral lysis of labelled *E. coli*, will have increased the amount of free  $^{13}\text{C}$ -labelled biomass components available for fungal degradation, explaining the dominance of the fast-growing saprotrophs *S. macrospora* and *C. globosum* (Kavak, 2012) which dominated the fungal  $^{13}\text{C}$  community. This is likely to have amplified changes in the carbonate–bicarbonate equilibrium induced by algal respiration, which has been shown to induce elevated growth rates and fruiting body formation in *Sordaria spp* (Elleuche and Pöggeler, 2010). This, in turn, results in further elevated  $\text{CO}_2$  levels due to fungal respiration causing an additional imbalance in the carbonate–bicarbonate equilibrium and inducing a knock-on effect to the water pH, which has been shown to induce elevated algal ROS production (Liu *et al.*, 2007). Such an association helps to explain the apparent mutualistic relationship displayed by fungi and algae during *E. coli* removal (Figure 6). Nonetheless, in addition to biological removal mechanisms, physical removal mechanisms such as straining, sedimentation and absorption have an important part in pathogen removal in SSFs (Haarhoff and Cleasby, 1991).

In summary, it was possible to ascertain that *E. coli* removal was achieved by both direct (protozoan grazing and viral lysis) and indirect (lysis induced by algal ROS production and fungal degradation of released biomass) mechanisms (Figure 6). These mechanisms appeared to occur simultaneously with the involvement of species from various kingdoms, in particular, fungi and algae, which exhibited mutualistic interactions. The highest removal of *E. coli* occurred between 1 and 3 h after spiking. This level of removal at these time points is consistent with the following characteristics of the  $^{13}\text{C}$  communities:

1. Phages peaked in abundance at 1 h, with extensive replication as part of their lytic pathway, resulting in reduced *E. coli* numbers after 2 h.
2. Protozoa numbers peaked at 2–3 h, allowing extensive grazing on  $^{13}\text{C}$ -labelled *E. coli* before this.
3. Algal abundance peaked at 2 and 3 h, which was likely due to the increased availability of  $^{13}\text{C}$ -labelled  $\text{CO}_2$  and other inorganics, created during viral lysis and protozoan grazing.
4. Fungal abundance peaked at 2 and 3 h, when extensive reduction in *E. coli* numbers occurred, hence releasing biomass for decomposition and resulting in changes in the carbonate–bicarbonate equilibrium inducing algal ROS production and further autolysis of *E. coli*.

## Conclusion

Although various studies have shown the individual importance of viral lysis, protozoan grazing, and endogenous and exogenous ROS in *E. coli* removal (Curtis *et al.*, 1992; Bomo *et al.*, 2004; Grobe *et al.*, 2006; Liu *et al.*, 2007; Kadir and Nelson, 2014), this is the first study, to our knowledge, to indicate the importance and interactions of all of these mechanisms for *E. coli* removal. Further, our approach enabled us to identify that the majority of the *E. coli* removal is due to top-down trophic interactions, such as protozoan grazing by *M. brevicollis* and *T. spp* and viral lysis by *E. phages*. In addition, although *E. coli* K12 was used in this study it is highly likely that the mechanisms of removal of pathogenic strains of *E. coli* would follow similar routes. The protozoan grazers identified are non-specific grazers, affected only by the size of the prey community and the phages identified are *Enterobacteria-specific*, rather than species-specific. However more work is required to determine if these removal mechanisms are similar for pathogenic and environmentally persistent bacterial species.

This study has shown that SSFs provide an ideal laboratory-scale system to study relevant and functional food webs. By applying cutting-edge molecular methods to these systems, we have furthered our understanding of the processes, mechanisms and organisms responsible for *E. coli* removal. The work and

methodology adopted in this study will provide both a paradigm for similar studies and the opportunity to:

1. Design and improve pathogen removal and overall performance of new and existing water purification systems by managing the community.
2. Predict *E. coli* removal rates in natural treatment systems that have biological components, particularly during pollution and weather events.
3. Further our understanding of complex food webs and trophic interactions.
4. Create more complex and realistic trophic interaction models.

Future work should aim to develop more sophisticated trophic interaction models using data generated from DNA-SIP studies. These models should be further integrated into pathogen prediction models to allow better pathogen tracking and removal prediction. Finally, the conclusion that ecosystem-wide associations are essential for complete *E. coli* removal may help to explain the reduced performance of household purification systems, which support a less diverse ecosystem. Therefore, future work should investigate the ability and benefit that introducing a more complex community may create.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

S-JH is supported by a Lord Kelvin/Adam Smith Research scholarship from the University of Glasgow. MS is supported by Unilever R&D, Port Sunlight, Bedington, UK. GC is supported by a European Research Council award (3C-BIOTECH; 261330). CQ is supported by an EPSRC Career Acceleration Fellowship award (EP/H003851/1). A special thanks to the GU68 Engineers Trust, Ian Scouller, Brian Robb, Tim Montgomery, Robert Boyd and Cindy Smith without whom this work would not have been possible.

## References

Abedon S. (2008). Contribution of lysogeny, pseudolysogeny and starvation to phage ecology. *Bacteriophage Ecology: Population Growth, Evolution and Impact of Bacterial Viruses*. Cambridge University Press: Cambridge, UK.

Alonso-Saez L, Gasol JM, Lefort T, Hofer J, Sommaruga R. (2006). Effect of natural sunlight on bacterial activity and differential sensitivity of natural bacterioplankton groups in northwestern Mediterranean coastal waters. *Appl Environ Microbiol* **72**: 5806–5813.

Anderson M. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**: 32–46.

Andrews S. (2010). FastQC a Quality Control Tool for High Throughput Sequence Data. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

Bomo A, Hovi T, Hanssen I, Fredrik J. (2004). Bacterial removal and protozoan grazing in biological sand filters. *J Environ Qual* **33**: 1041–1047.

Cassidy-Hanley DM. (2012). *Tetrahymena* in the laboratory: strain resources, methods for culture, maintenance, and storage. *Methods Cell Biol* **109**: 237–276.

Christaki U, Vazquez-Dominguez E, Courties EC, Lebaron P. (2005). Grazing impact of different heterotrophic nanoflagellates on eukaryotic (*Ostreococcus tauri*) and prokaryotic picoautotrophs (*Prochlorococcus* and *Synechococcus*). *Environ Microbiol* **7**: 1200–1210.

Clarke RK. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* **18**: 117–143.

Curtis TP, Mara DD, Silva SA. (1992). Influence of pH, oxygen, and humic substances on ability of sunlight to damage fecal coliforms in waste stabilization pond water. *Appl Environ Microbiol* **58**: 1335–1343.

Danger M, Cornut J, Chauvet E, Chavez P, Elger A, Lecerf A. (2013). Benthic algae stimulate leaf litter decomposition in detritus-based headwater streams: a case of aquatic priming effect? *Ecology* **94**: 1604–1613.

Dehority B. (1984). Evaluation of subsampling and fixation procedures used for counting rumen protozoa. *Appl Environ Microbiol* **48**: 182–185.

Elleuche S, Pöggeler S. (2010). Carbonic anhydrases in fungi. *Microbiology* **156**: 23–29.

Elliott M, Stauber C, Koksal F, DiGiano F, Sobsey M. (2008). Reductions of *E. coli*, echovirus type 12 and bacteriophages in an intermittently operated household-scale slow sand filter. *Water Res* **42**: 2662–2670.

Feng Y, Lin X, Zhu J, Jia Z. (2011). A phototrophy-driven microbial food web in a rice soil. *J Soil Sediment* **11**: 301–311.

Fey P, Kowal AS, Gaudet P, Pilcher KE, Chisholm RL. (2007). Protocols for growth and development of *Dictyostelium discoideum*. *Nat Protoc* **2**: 1307–1316.

Fleck RA, Benson EE, Bremner DH, Day JG. (2000). Studies of free radical-mediated cryoinjury in the unicellular green alga *Euglena gracilis* using a non-destructive hydroxyl radical assay: a novel approach for developing protistan cryopreservation strategies. *Free Radic Res* **32**: 157–170.

Frith MC, Hamada M, Horton P. (2010). Parameters for accurate genome alignment. *BMC Bioinformatics* **11**: 80.

Grobe S, Petry-Hansen H, Uhl M, Flemming H. (2006). Retention of hygienically relevant microorganisms from storm water effluents by slow sand filters. *Slow Sand Filtration: Recent Developments in Water Treatment Technology*. John Wiley and Sons: London, UK, 474–479.

Haarhoff J, Cleasby JL. (1991). Biological and physical mechanisms in slow sand filtration. *Slow Sand Filtration*. American Society of Civil Engineers: New York, NY, USA 19–68.

Haig S, Collins G, Davies RL, Dorea CC, Quince C. (2011). Biological aspects of slow sand filtration: past, present and future. *Water Sci Technol* **11**: 468–472.

Haig S, Dorea CC, Collins G, Davies RL, Quince C. (2014a). Replicating the microbial community and water quality performance of full-scale slow sand filters in laboratory-scale filters. *Water Res* **61**: 141–151.

- Harte J, Kinzig AP. (1993). Mutualism and competition between plants and decomposers: implications for nutrient allocation in ecosystems. *Am Nat* **141**: 829–846.
- Hijnen W, Dullemeier Y, Schijven J, Hanzens-Brouwer A, Rosielle M, Medema G. (2007). Removal and fate of *Cryptosporidium parvum*, *Clostridium perfringens* and small-sized centric diatoms (*Stephanodiscus hantzschii*) in slow sand filters. *Water Res* **41**: 2151–2162.
- Hunter MD, Price PW. (1992). Playing chutes and ladders: heterogeneity and the relative roles of bottom-up and top-down forces in natural communities. *Ecology* **73**: 723–732.
- Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21**: 1552–1560.
- Kadir K, Nelson KL. (2014). Sunlight mediated inactivation mechanisms of *Enterococcus faecalis* and *Escherichia coli* in clear water versus waste stabilization pond water. *Water Res* **50**: 307–317.
- Kavak H. (2012). Some biological parameters in *Sordaria Fimicola*. *Pak J Bot* **44**: 1079–1082.
- Liu W, Au DW, Anderson DM, Lam PK, Wu RS. (2007). Effects of nutrients, salinity, pH and light: dark cycle on the production of reactive oxygen species in the alga *Chattonella marina*. *J Exp Mar Biol Ecol* **346**: 76–86.
- Lloyd B. (1973). The construction of a sand profile sampler: its use in the study of the vorticella populations and the general interstitial microfauna of slow sand filters. *Water Res* **7**: 963–973.
- Maynard H, Ouki S, Williams S. (1999). Tertiary lagoons: a review of removal mechanisms and performance. *Water Res* **33**: 1–13.
- Marley J, Lu M, Bracken C. (2001). A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**: 71–75.
- McCambridge J, McMeekin T. (1980). Relative effects of bacterial and protozoan predators on survival of *Escherichia coli* in estuarine water samples. *Appl Environ Microbiol* **40**: 907–911.
- Mikola J, Setälä H. (1998). No evidence of trophic cascades in an experimental microbial-based soil food web. *Ecology* **79**: 153–164.
- Nasr A, Bekheet IA, Ibrahim R. (1968). The effect of different nitrogen and carbon sources on amino acid synthesis in *Ulva*, *Dictyota* and *Pterocladia*. *Hydrobiologia* **31**: 7–16.
- Nedelcu A, Miles I, Fagir A, Karol K. (2008). Adaptive eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. *J Evolution Biol* **21**: 1852–1860.
- Neufeld J, Vohra J, Dumont MG, Lueders T, Manefield M, Friedrich MW *et al*. (2007). DNA stable-isotope probing. *Nat Protoc* **2**: 860–866.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al*. (2012). *Vegan: Community Ecology Package*, R package version 2.0-3.
- Pace M, Cole J. (1994). Comparative and experimental approaches to top-down and bottom-up regulation of bacteria. *Microb Ecol* **28**: 181–193.
- Parry JD. (2004). Protozoan grazing of freshwater biofilms. *Adv Appl Microbiol* **54**: 167–196.
- Pernthaler J. (2005). Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* **3**: 537–546.
- Ripp S, Miller RV. (1997). The role of pseudo-lysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology* **143**: 2065–2070.
- Rosemond AD, Pringle CM, Ramírez A, Paul MJ. (2001). A test of top-down and bottom-up control in a detritus-based food web. *Ecology* **82**: 2279–2293.
- Semple KT, Cain RB. (1996). Biodegradation of phenols by the alga *Ochromonas danica*. *Appl Environ Microbiol* **62**: 1265–1273.
- Sul W, Park J, Quensen J, Rodrigues J, Seliger LT, Tsoi TV *et al*. (2009). DNA-stable-isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Appl Environ Microbiol* **75**: 5501–5506.
- Weber-Shirk M, Dick R. (1999). Bacterivory by a *Chrysophyte* in slow sand filters. *Water Res* **33**: 631–638.
- Weber KP, Legge RL. (2008). Pathogen removal in constructed wetlands. In *Wetlands: Ecology, Conservation and Restoration*. Nova Publishers: New York, NY, USA, pp 176–211.
- Weekers PH, Bodelier PL, Wijnen JP, Vogels GD. (1993). Effects of grazing by the free-living soil amoebae *Acanthamoeba castellanii*, *Acanthamoeba polyphaga*, and *Hartmannella vermiformis* on various bacteria. *Appl Environ Microbiol* **59**: 2317–2319.
- WHO (World Health Organization). (2004). Water and Sanitation: Facts and Figures. Available at: [http://www.who.int/water\\_sanitation\\_health/publications/factsfigures04/](http://www.who.int/water_sanitation_health/publications/factsfigures04/).
- Withey S, Cartmell E, Avery L, Stephenson T. (2005). Bacteriophages potential for application in wastewater treatment processes. *Sci Total Environ* **339**: 1–18.
- Wooley J, Godzik A, Friedberg I. (2010). A primer on metagenomics. *PLoS Comput Biol* **6**: 1–13.
- Yue J, Sun G, Hu X, Huang J. (2013). The scale and evolutionary significance of horizontal gene transfer in the choanoflagellate *Monosiga brevicollis*. *BMC Genomics* **14**: 729.
- Zeng L, Golding I. (2011). Following cell-fate in *E. coli* after infection by phage lambda. *J Vis Exp* **1**: 3363.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)