genes

MDPI

Article

Using Functional Annotations to Study Pairwise Interactions in Urinary Tract Infection Communities

Elena G. Lara [1], Isabelle van der Windt [2], Douwe Molenaar [1], Marjon G. J. de Vos [2,*] and Chrats Melkonian [1,*]

[1] Systems Biology Lab, AIMMS, Vrije Universiteit, 1081 HZ Amsterdam, The Netherlands; elena.garcia.lara@hotmail.com (E.G.L.); d.molenaar@vu.nl (D.M.)
[2] GELIFES, Universtity of Groningen, 9747 AG Groningen, The Netherlands; isabelle.vanderwindt@hvhl.nl (I.v.d.W.)
* Correspondence: m.g.j.de.vos@rug.nl (M.G.J.d.V.); chrats.melkonian@gmail.com (C.M.)

Abstract: The behaviour of microbial communities depends on environmental factors and on the interactions of the community members. This is also the case for urinary tract infection (UTI) microbial communities. Here, we devise a computational approach that uses indices of complementarity and competition based on metabolic gene annotation to rapidly predict putative interactions between pair of organisms with the aim to explain pairwise growth effects. We apply our method to 66 genomes selected from online databases, which belong to 6 genera representing members of UTI communities. This resulted in a selection of metabolic pathways with high correlation for each pairwise combination between a complementarity index and the experimentally derived growth data. Our results indicated that *Enteroccus* spp. were most complemented in its metabolism by the other members of the UTI community. This suggests that the growth of *Enteroccus* spp. can potentially be enhanced by complementary metabolites produced by other community members. We tested a few putative predicted interactions by experimental supplementation of the relevant predicted metabolites. As predicted by our method, folic acid supplementation led to the increase in the population density of UTI *Enterococcus* isolates. Overall, we believe our method is a rapid initial in silico screening for the prediction of metabolic interactions in microbial communities.

Keywords: urinary tract infection; microbial community; microbial interaction; gene annotation

check for updates

Citation: Lara, E.G.; van der Windt, I.; Molenaar, D.; de Vos, M.G.J.; Melkonian, C. Using Functional Annotations to Study Pairwise Interactions in Urinary Tract Infection Communities. *Genes* **2021**, *12*, 1221. https://doi.org/10.3390/genes12081221

Academic Editor: Silvia Turroni

Received: 30 May 2021
Accepted: 3 August 2021
Published: 6 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The microbiome of our body has a significant impact on our health and disease, as disequilibrium in these microbial communities may lead to dysbiosis [1]. The interactions between the microbes in a community may come in many forms and can have different effects, including synergy or inhibition of the growth of other community members, killing by toxins, or cooperative behaviour induced by quorum sensing [2] or the exchange of nutrients [3,4]. Moreover, interactions between the members of a microbial community can potentially alter the sensitivity to antibiotics in vitro, and, therefore, potentially the effect of treatments [5].

Several tools from biology systems have been used to study microbial interactions, with a focus on the analysis of the flow of molecules and energy through metabolic networks [6]. Genome-scale stoichiometric models have been used to study single organisms and simple communities to high levels of detail [7,8]. However, model reconstruction and manual curation are still a time-consuming process. Simpler and more coarse-grained models were proven to be useful, as they have less accumulated biases and are more readily to be applied [8–10]. Both approaches rely on functional annotations predicted from genomic data. These annotation elements (e.g., from KEGG [11], MetaCyc [12], TheSEED [13], or EggNOG [14] databases) are a proxy of the putative metabolic capacities of an organism. Using these annotations, one could construct the topology of a metabolic network with the aim to infer the biochemical environment of an organism [15]. Subsequently, graph theory

can be used to compare the topology of metabolic networks between pairs of organisms to compute interaction indexes, such as biosynthetic support score (NetCooperate [16]), complementarity or competition (RevEcoR [17]). Alternatively, it is possible to unveil properties of the system without inferring the whole metabolic network by comparing patterns in the KEGG Ortholog (KO) content of the organisms within a community [18].

One hallmark of urinary tract infections (UTI) in elderly patients is the presence of multiple pathogens [19]. In addition, although the urinary tract has historically been considered sterile in the absence of infection, it has recently been discovered that there is a resident microbiome [20–23]. Additionally, in uncomplicated UTIs there is often a background microbiome present, even though the contribution of the microbiome to the pathology is not yet clear [24,25]. Given that UTI affects 150 million people worldwide each year [26], and an estimated 50% of women report having had a UTI at some point in their lives [27,28], it is important to understand the potential role of bacterial interactions in these infections. In the UTI context, the main focus has historically been on *Escherichia coli* , the main urinary pathogen. Other taxa commonly involved in polymicrobial UTIs include *Klebsiella pneumoniae*, *Proteus mirabilis*, *Pseudomonas* spp. and *Staphylococcus* spp., and *Enterococcus* spp. [29–31].

The aim of this work was to study the genetic basis of pairwise interactions of the bacteria involved in polymicrobial UTI reported by de Vos et al. [5]. We used a pure data-driven approach to find putative metabolic interactions using genomic information and functional annotation of genes. Because the genome sequences of the organisms were not available, we collected 66 genomes from online databases to artificially constitute this UTI community. Then, we devised four complementarity and four competition indices for each pair of genomes based on the KOs for each KEGG pathway. Those indices were compared using the correlation with the growth variance from pairwise experiments [5]. The chosen complementarity measure was further used to select the most informative pathways, which may explain the positive growth synergies or lack of between pairs of strains. Overall, the results led to the hypothesis that members of a UTI community could provide intermediates of specific metabolic pathways to *Enterococcus* spp. In particular, the hypothesis that folic acid produced by the community members may promote the growth of enterococci was confirmed experimentally.

2. Materials and Methods

2.1. Yield and Growth Rate in UTI

The starting point of the project was data from pairwise combinations of bacteria from UTI communities. Communities containing four different species were isolated from hosts diagnosed with a polymicrobial UTI [19]. In an effort to analyse the interplay between these species, de Vos et al. [5] had quantified these interactions by means of phenotypic growth essays. Briefly, 72 bacteria strains were isolated from 23 UTI patients in a previous study ([19]), and cultured in a modified artificial urea medium (AUM) [32] (unconditioned medium, N_u) for 48 h. Afterwards, the supernatant from this cultures was used to supplement a second medium (conditioned medium, N_c) for 24 h. Interactions were calculated as the ratio of growth between conditioned and unconditioned media; positive interactions were due to an increased growth in conditioned medium, and negative interactions were due to decreased growth in conditioned medium, compared to the growth in unconditioned medium [5].

This analysis resulted in a pairwise interaction matrix G (Figure S1), of size $N \times N$, where N equals to the number of strains isolated from UTI patients (72). The donors are represented in the y-axis and the acceptors in the x-axis. The strains are ordered by phylogeny; similar genera show similar behaviour [5]. The conditioned medium was partly mixed with AUM, so that the complemented medium (N_c) had between 0.6 and 1 times the original concentration of nutrients, depending on the consumption of the donor. Together, this approach allowed to distinguish between interactions with different signs (Table S5). In the matrix, the strains are divided into six genera: Ecoli: *E. coli*; Ent: *Enterococcus*

spp. (*E. faecalis* and *E. faecium*); KECS: *Klebsiella* spp. (*K. pneumoniae* and *K. oxytoca*), *Enterobacter cloacae*, *Citrobacter koseri*, *Serratia liquefaciens*, and *Pantoea sp4*; Pm: *P. mirabilis*; Mm: *Morganella morganii*; and Ps: *Pseudomonas* spp. (*P. aeruginosa* and *P. fluorescens*); St: *Staphylococcus* spp. (*S. aureus*, *S. haemolyticus* and *S. capitis*). Interestingly, the growth increase or decrease is conserved within genera [5], fencing the interactions to the core functions, and allowing us to create an approximate community to study.

2.2. Description of the Dataset

We mimicked the experimental UTI communities, using online available genomes, because the genomic sequence of the phenotypically measured strains are not available. This in silico community consisted of the same species in a similar proportion. The genome dataset used in this project was retrieved from the National Center for Biotechnology Information (NCBI) [33]. More specifically, the strains were retrieved from the Prokaryote Genome database, which contains the genomes of more than 400,000 organisms, 350,000 of which are bacteria (as of 20 July 2021). Only the sequences with level of assembly 'complete' were chosen. When a species presented more strains available than the number of species present in the experimental community, the strains related with UTI according to literature were preferred. Otherwise, human isolates were chosen and, as a last resort, isolates from other sources were selected. Table S1 describes the initial selection of species and the equivalent strains, of which the results are described below. Table S2 describes a selection of the same species, but this collection only contains urine or human derived isolates. The results of the pipeline on those two genome selections (described in Tables S1 and S2) were very similar (Figure S7, Table S7). Because the genome sequences were not derived from the strains assessed for the phenotypic growth matrix, a one to one comparison was not possible. Thus, we grouped the information of the strains into groups related to their genera to compare the phenotypically measured UTI isolates and the downloaded genomes. Overall, the distribution of groups looks similar to the original UTI community (Figure S4). With this in silico community, the analysis was performed following the flowchart in Figure S5. The project was carried out for the most part in Python 2.7.12, using the Pandas 0.19.2 library [34], and R 3.2.3-4 (within the RStudio environment). The scripts were stored in a GitHub repository (https://github.com/egarcialara/UTI_bacteria_interactions accessed on 1 July 2021).

The ribosomal 16S gene sequences were available for each of the isolates in the experimental community. The 16S sequences were blasted against three different databases. One was GreenGenes [35], a global database for sequences of this kind. On the other hand, the global prokaryote database from NCBI [36] served as an alternative database. Furthermore, a local database containing the genomes of the strains already selected for the in silico analysis, together with other strains belonging to the same species, was generated.

2.3. Functional Annotation of the Genomes

On one side, the genomes were downloaded from the GenBank database [37] using the Batch Entrez system for a fast retrieval. The nucleotide sequences of the genomes were translated into amino acid sequences using Prodigal [38]. It was followed by a BlastKoala search [39], to find the KEGG Orthologs associated with each of the organism sequences. The parameters were kept as default, with the exception of 'taxonomy group of genome': Bacteria and 'KEGG genes database': genus_prokaryotes. We made use of another parallel annotation that consisted in sequences downloaded directly from the Integrated Microbial Genomes database (IMG) [40] (Figure S6). The final annotation for each strain had the form of a matrix with two columns: one containing the K accession numbers of all the KOs, and a second column with boolean values, 1 if the KO is present, 0 if not. Moreover, a binary matrix B, $N \times M$, of N KOs and M bacteria strains was used from [18], to have a visual representation of the KO content. Each entry $b_{i,j}$ represents the presence of KO i in the strain j with a 1, or 0 otherwise.

2.4. Functional Annotation Division into Smaller Segments

The KO vectors were broken down, from one long list to smaller vectors that correspond to the KEGG pathways. There are more than 500 pathways listed in the KEGG database, of which 274 had at least one KO in one of the organisms. In order to link the KOs to their respective pathways, the representational state transfer (REST)-style KEGG Application Programming Interface (API) was used. The link form http://rest.kegg.jp/list/pathway (accessed on 1 July 2021) allows to retrieve the complete list of KEGG pathways, whereas the form http://rest.kegg.jp/link/kos/mapxxxxx (accessed on 1 July 2021) serves to associate the pathways and their KOs. The *'mapxxxxx'* represents a KEGG pathway identifier space. In a similar way, the KEGG database offers other entry points, such as BRITE Hierarchy and modules [39]. A segmentation of the initial vectors into modules can be completed similarly with http://rest.kegg.jp/list/module (accessed on 1 July 2021) (retrieve all modules in KEGG database) and http://rest.kegg.jp/link/kos/Mxxxxx (accessed on 1 July 2021) (associate modules with their KOs), where, in this case, '*Mxxxxx*' would be the KEGG module identifier. The third kind of partition chosen was one level above pathways, based on BRITE Hierarchy (BRITE Hierarchy files > Genes and Proteins > Orthologs and modules > KEGG Orthology (KO)—all categories (groups)). At the end of the Results section, a comparison between the segmentation in modules, pathways, BRITE terms, and the complete set of KOs was performed by comparing their performance in finding a pattern.

2.5. Interaction Matrices

The combination of the boolean KOs vectors from the two organisms in a pair was completed in the form of asymmetric interaction matrices C, $N \times N$. In them, N corresponds to the number of different strains in the community. The values $c_{d,a}$ are the resulting of the comparison of two corresponding KO vectors: of a donor d and an acceptor a. The interaction was interpreted as either complementarity or competition. All the values were normalised by the total number of KOs in the vector, to facilitate the comparison between pathways (not explicitly stated below). Additionally, in order to make indexes asymmetric, many of them were divided by the number of KOs in the acceptor (only if stated).

2.5.1. Complementarity

The complementarity indices were calculated based on the premise that metabolic functions present in one genome and not in the other pair member, could lead to the donor to supply the lacking compounds to the acceptor. This would be related with synergy, and would be expected to have a positive effect in the growth on the acceptor. At first, any intermediate is equally likely to be 'complemented'. There are different ways to compute complementarity:

- Complementarity 1: directly as the fraction of the KOs that are present in the donor set but not in the acceptor;
- Complementarity 2: as the fraction of the total amount of KOs that the acceptor could have when complemented (by the donor) divided by the acceptor's own possibilities;
- Complementarity 3: as Complementarity 2, but without taking into consideration the shared KOs (similar to Hamming distance for boolean vectors, but divided by A instead of total number of dimensions to make it an asymmetric index);
- Complementarity 4: same as Complementary 3, but divided by the putative total functionality. In this way, it is equivalent to Hamming distance, but it is symmetric.

2.5.2. Competition

The competition indices were envisioned as the similarity of the metabolic potential of the organisms in a pair. The more similar, the highest similarity value, and the most expected putative negative effect on growth. This corresponds with the experimental set-up in which the complemented medium in the pair was only partially restored when

transferred from the acceptor to the donor. Thus, a similarity in the compounds consumed by the former, would limit the latter's uptake. The indices considered are as follows:

- Competition 1 : directly the fraction of shared KOs;
- Competition 2: equivalent to competition_1 but compared to the acceptor function to make it asymmetric;
- Competition 3: equivalent to competition_2, but compared to the total possibilities;
- Competition 4: is just Pearson correlation (just for comparison).

2.5.3. Selection of Interaction Indices

We selected the interaction indices that matched the growth variations best, based on the correlation. Firstly, we selected the 100 pathways with highest standard deviation with respect to the pathway coverage between genera, as considered putatively, at least for the start, the most informative. Secondly, Pearson correlation was calculated for the index of each of these pathways and growth—either yield or rate. The 'goodness' of an index was tested by checking the correlation for each pathway interaction matrix with the growth matrices. In order to compare two matrices, they were flattened into a 1D array. Finally, a p-value was computed by permuting randomly the growth values, as in the equation that follows:

$$P_{perm} = \frac{\sum_{n}^{N} |correlation_{randomized}| > |correlation_{original}|}{N_{permutations}}$$

where N corresponded with 1000 permutations. The indices with largest correlations were used for further study.

2.5.4. Selection of Informative Pathways

We sought to identify those pathways in which there were a clear separation between pairs with high complementarity and increase in growth ($\epsilon > 0.22$, Figure S4) and other pairs. In order to filter down the most informative pathways, we ranked the pathways depending on their performance in a classification between those two groups. We chose an ensemble of simple models, as they usually yield a performance as good as very complex models [41]. For this case, we chose three distinct methods: Boruta feature selection, as used in [18] with similar data, support vector machine (SVM) recursive feature elimination (RFE) as it performs well in many types of data and holds no assumptions [42] and a statistical Mann–Whitney U test (MWU) with which the selection is independent of the model. Boruta [43] is an extension of the random forest algorithm, which uses this classification to iteratively discard features. A Z-score of an attribute can be calculated by dividing the average accuracy loss for all the trees of a random forest by its standard deviation. In order to set an statistical significance Boruta calculates shadow attributes for each feature. A paired t-test is then calculated to interpret if the Z-score of the real attribute is higher than that of the maximum Z-score among shadow attributes. We ranked the pathways by the computed feature importance itself. The SVM algorithm classifies the points by representing them into space and creating a decision boundary that separates the groups. The optimal decision boundary is the one that maximises the minimum distance from the separating decision boundary to the nearest points or support vectors. Each feature has a weight on the decision vector, and this coefficient was used to rank the pathways. When applied RFE for selection, the feature with smaller ranking criteria, based on this weight, is removed iteratively [42]. Lastly, a model-free statistic method was selected. The MWU test assumes as null hypothesis that the probability of picking a value from a group bigger than the other group is equal. It is equivalent to a non-parametric t-test. The result is a statistic computed for each pathway and an associated p-values. The result was a ranking of the importance of the pathways, based on the MWU statistic, SVM RFEs score and Boruta's importance. The ensemble was performed by summing up the rank position of each pathway. A stacking of base learners was avoided due to the small size of the dataset.

2.6. Alternative Approaches

In the project we compared the indexes above with two previously published tools that compute indexes based on metabolic potential, and that could be compared with the 'complementarity' indices presented above. NetCooperate [16] incorporates a metabolic complementarity index (MCI) that shows a putative synergy between two organisms. More recently, RevEcoR [17] has come as a tool that calculates Complementarity as "the fraction of compounds in species A's seed set appearing in the metabolic network, but not appearing in species B's seed set", as well as Competition: "the fraction of compounds in species A's seed set that are also included in species B's seed set". The seed set refers to the predicted minimum set of compounds required for the growth of an organism. These three indexes were computed using their respective R packages, and the resulting matrix was analysed similarly as the rest of the indices.

2.7. Checking Transporters Availability

The biological potential of interactions between the pairs was checked by the presence of membrane transporters in the annotations of the organisms. This presence was assessed using KEGG, specifically three pathways: ABC transporters, phosphotransferase system (PTS), and bacterial secretion system.

2.8. Supplementation Experiments

A 1:1000 dilution of overnight culture pre-grown in modified artificial urine medium of *Enterococcus faecium*, isolated from a polymicrobial UTI [19], was inoculated with in 5 mL artificial urine medium [5]. Cultures were grown with and without supplementation of 2 μM serine, 2 μM glutamate, 1 μM histidine, 0.7 μM vitamin B12 for 24 h, at 37 °C, shaking 200 rpm. The effect of folic acid supplementation on the growth yield was assessed on two *E. faecium* isolates and three *E. faecalis* isolates, isolated from a polymicrobial UTI [19]. The 1:100 dilutions of overnight cultures pre-grown in modified artificial urine medium were inoculated in 96-well plates, with wells containing 200 μL artificial urine medium [5]. Cultures were grown with and without supplementation of 30 μM and 300 μM folic acid for 24 h, at 37 °C, shaking 500 rpm. Additional negative controls were artificial urine medium without bacteria, artificial urine medium with supplements without bacteria. To assess the effect of the addition of the metabolic supplementation on the population size, the yield, the optical density of the triplicate cultures was measured at OD600 nm after 24 h. Significant differences in growth yield due to folic acid supplementation compared to the control without folic acid supplementation, were assessed with a *t*-test ($p < 0.05$), with Bonferroni-correction to adjust for multiple comparisons.

3. Results

3.1. Complementarity Indices Show the Greatest Correlation with the Pairwise Interaction Growth Yields

The starting point was finding representative genomes of the species present in the UTI communities, as observed in [5]. Blasting the 16S sequences from the UTI isolates against global databases did not yield specific matches, as this approach could not distinguish beyond the species level of taxonomy. Thus, the representative strains were manually picked from the NCBI database based on their origin of isolation, by an order of preference from urine or UTI origin, human or animal, shown in Table S1. The 66 selected genomes were divided into six groups. Each group corresponded to the genera of the isolate, with the exception of *Klebsiella* spp., *E. cloacae*, *C. koseri*, *S. liquefaciens*, and *Pantoea* spp. which were grouped altogether (KECS group). For all genomes, the open reading frames (ORFs) were predicted, and the corresponding amino acid sequences were annotated as KOs using BlastKoala [44]. With the resulting KO vectors of each strain, we built a binary matrix *B*. This representation helps visualising how the KO are conserved within genera (Figure 1A), i.e., the KOs that are shared between species of a genus. The average annotation ratio of KOs in comparison with the gene sequences was of 60.6% (st.dev. 9.1%), although

it was different between the groups. This, together with the different absolute number of genes originally present in each genome, caused that the total number of functional annotations varied depending on the genera. The Ps and KECS group (Table S1) had at least twice as many functional annotations when compared to enterococci and staphylococci (Figure 1B). The lower number of absolute number of genes and annotation ratio is evident in enterococci and staphylococci (Gram+) (Figure 1A,B).

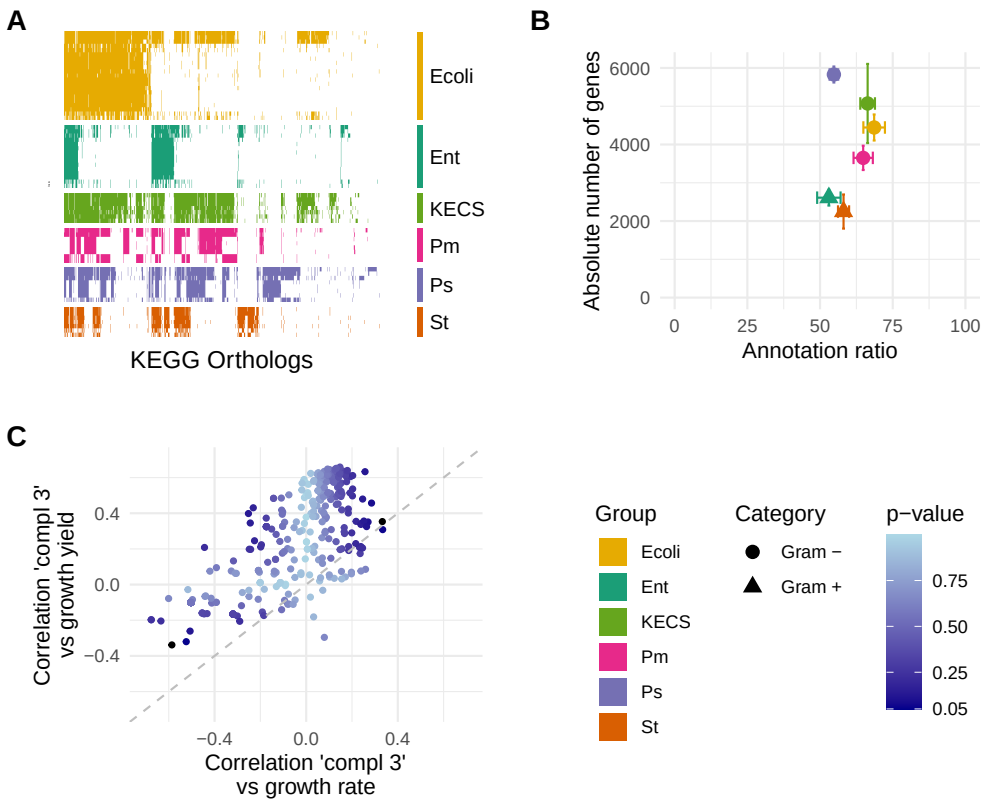

Figure 1. (**A**) Heatmap of presence (coloured) or absence (white) of KEGG Orthologs in the different bacteria groups. Although most KOs are present in the *KECS* group, other groups lack many. Note that only the KOs that are present in at least one organism are shown. (**B**) Although the absolute number of genes in the genome of each group varies per group, their annotation ratio from genome to KOs remains similar around 60%. The groups *St* and *Ent* have notoriously less genes and, thus, less absolute number of KOs than the rest of the groups. (**C**) Comparison of the relation of 'Complementarity 3' with the increase in two different growth measures: yield and rate (from [5]). The correlation is higher, for most pathways, when calculated against growth yield. An overview of the relation of all calculated indices with growth rate and yield is presented in Figure S2.

From the KO vectors we created interaction matrices, C, based on complementarity or competition indices (Tables S3 and S4). The values for each cell ($c_{d,a}$) correspond to the calculated index between the donor species d and the acceptor strain a. The heatmaps in Figure 2 represent the different strains, grouped by genera, which adopt the role of donors and that of acceptors in the y- and x- axis, respectively. A darker shade of black in the cells is associated with a higher value of the complementarity index. The computed measure is able to group the strains by their genera. These matrices had the same dimensions as the relative growth matrix as constructed in the growth variance experiments from [5] (reconstructed in Figure S1), which allowed to calculate the correlation between the interaction measures and growth (Figure S2).

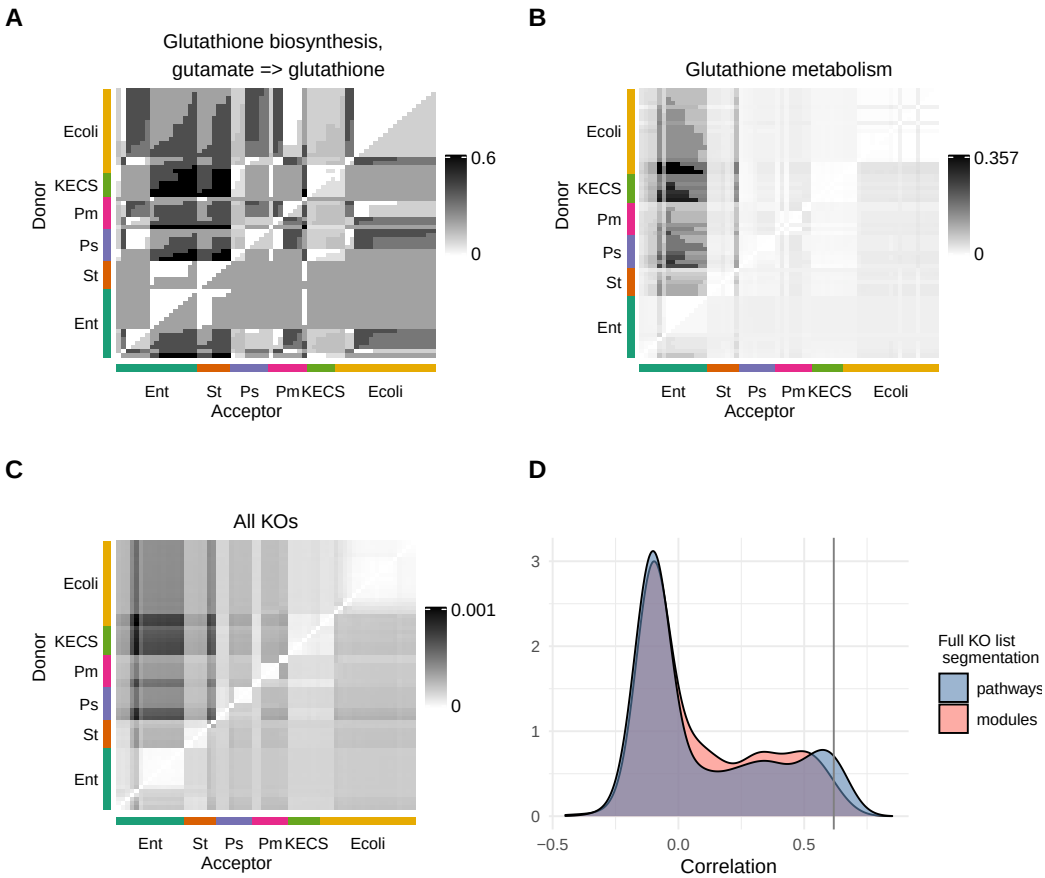

Figure 2. Heatmap of the 'Correlation 3' between the different strains within the groups. (**A**) The KOs selected belong to a module. (**B**) The KOs represented are part of a pathway. (**C**) All KOs were taken into account to calculate the 'Complementarity 3'. (**D**) The density distribution of the correlations between the 'Complementarity 3' calculated using KOs belonging to one pathway at a time is shown in blue. In red, there is the similar calculation for modules. The vertical line represents the correlation using all available KOs. Using the pathways for the segmentation of the KO vectors yields the highest correlations.

The measures show a general behaviour across the pathways in their correlation with growth. The complementarity measures had a positive correlation with growth yield. In particular, Complementarity 3 ($(D \cup A - D \cap A)/A$) showed the largest average correlation ($\rho = 0.41$) with growth yield. Thus, it was chosen for the subsequent analyses (Figure 1C). A similar measure, Complementarity 2 ($(D \cup A)/A$), had a similar average correlation ($\rho = 0.33$). Additionally, the *p*-values obtained from a permutation test were lower in the cases where the correlation was higher. Competition measures, such as Competition 1 ($D \cap A$) showed a negative correlation of up to -0.54. However the *p*-values was not low enough to indicate that the index was informative (Figure S2). Moreover, no pattern was seen when comparing the interaction indices, both complementation and competition, to growth rate (Figure S2). The values for correlation between the interaction indices and the growth rate variation were low, with a minimum value of -0.36 (Competition 2). Nonetheless, the *p*-values were too large to hold onto these indices.

3.2. Analysis of Genomes in Terms of KEGG Pathways

In our analysis we chose to use the KEGG pathway segmentation because it is the best curated, as well as an intuitive way to break down the entire KO vector. However, one could construct similar interaction matrices without splitting the KO vector. With this approach, the relationship between the interaction measure and growth data in the

organisms are lower in comparison to segmenting the KO vector by KEGG pathways (KO vectors correlation: ρ 0.62, $p < 0.001$, KEGG pathway segmentation: 0.66 ($p < 0.001$)) (Figure 2).

Next, we sought to identify the pathways for which the relationship between the interaction indices and the growth yield was most prominent. Subsequently, we ranked the most informative pathways using an ensemble of Boruta feature selection, support vector machine and the Mann–Whitney U test (MWU). Two-classes were used as targets for each method, with one representing the positive growth effect of the pairwise interactions and the other the neutral or negative growth effect. The threshold was set as $\epsilon > 0.22$ as mentioned in [5]. The final selection was voted by all three methods by summing the pathway ranks of their individual result. The top ranked pathways are reported in Table 1.

On the other hand, using modules as another available segmentation, also underperforms (e.g., Glutathione metabolism, Phenylalanine metabolism, 'Tyrosine metabolism' have values of correlation 0.65, 0.63, 0.61, respectively, (p-values $\leq 2.5 \times 10^{-4}$), whereas their best scoring module correlations (and p-values) are 0.42 (0.01), 0.34 (0.04), 0.31 (0.06)). In addition, the modules have a very small size: 78% of them are comprised of less than 10 KOs, while only 18% of the pathways have that few KOs. Therefore, we concluded that the pathway segmentation of the KO sets was the most informative choice in this context (Figure 2D).

Table 1. Pathways with high complementarity for the pairs with increased growth yield. Top ranking pathways according to the RF feature importance, the SVM coefficients, the MWU p-values, and the ensemble of the previous ranks.

Pathways	**RF**	**SVM**	**MWU**	**Ensemble**				
	Feat. Imp.	**Rank**	**Coef.**	**Rank**	***p*-Values**	**Rank**	**Sum**	**Rank**
Lipopolysaccharide biosynthesis	0.111	1	2.265	1	0.0003	13	15	1
Tyrosine metabolism	0.028	11	0.531	6	3.08×10^{-5}	1	18	2
Sulfur metabolism	0.058	6	0.888	3	0.0003	12	21	3
Cationic antimicrobial peptide (CAMP) resistance	0.036	9	0.513	9	0.0001	6	24	4
Phenylalanine metabolism	0.073	4	0.313	16	0.0002	10	30	5
Biotin metabolism	0.017	15	0.385	14	7.01×10^{-5}	4	33	6
Glutathione metabolism	0.015	17	0.243	22	9.44×10^{-5}	5	44	7
Bacterial secretion system	0.050	7	0.905	2	0.001	41	50	8
Biofilm formation-Pseudomonas aeruginosa	0.020	14	0.318	15	0.0007	29	58	9
Riboflavin metabolism	0.007	23	0.518	8	0.0007	28	59	10
Histidine metabolism	0.014	18	−0.107	65	0.0003	14	97	22
Folate biosynthesis	0.001	53	0.016	48	0.0003	15	116	31

3.3. Enterococci Are Metabolically Complemented by the UTI Community

The interaction indices differ by genus. In particular, *Enterococcus* as an acceptor has a unique behaviour. In the pairs in which *Enterococcus* acts as an acceptor, the value of the complementarity index is higher than for other genera of the UTI community, which corresponds closely with the measured positive interactions affecting growth yield [5]. We mapped the KOs into the KEGG pathway maps to study the different availability of orthologs in the six genera in more detail. In accordance with the highest complementarity for enterococci, the patterns of selected pathways (Table 1) showed that enterococci lacked the orthologs that were present in the other genera (Figure 3. For example, in histidine metabolism, the functions that lead to histidine synthesis are present in most of the strains of *E. coli* but in almost none of the enterococci (Figure 4). A selected number of possible exchanged compounds is described in Table 2, based on the interpretation of the top ranked pathways.

Table 2. Compounds proposed to be complemented in enterococci by other UTI species. The KEGG pathway images referenced in the text are (1) Alanine, aspartate, and glutamate metabolism; (2) Arginine biosynthesis; (3) Sulphur metabolism; (4) Histidine metabolism; (5) Glycine, serine, and threonine metabolism; (6) Folate biosynthesis; (7) One carbon pool by folate.

Compound	Biological Relevance
Fumarate	Two ways to produce fumarate: from L-Aspartate (1) and from citruline (2), are not present in enterococci and staphylococci, but are in the others.
Sulphide	Sulphide can be formed from extracellular sulfate by many forms. Ent and Ecoli do not have the necessary orthologs, but they are able to convert this sulphide in Acetate and L-Cysteine (3). Cysteine is an amino acid, which has antioxidant activity [45].
Histidine, Glutamate, and Serine	Not directly synthesised in Enterococcus (4, 4, 5, respectively).
Folate (DHF) and derivates (10 formyl THF)	Routes available for production of folate (DHF) and derivates (10formylTHF) were missing in Enterococcus but not in the other genera. (6 (folate) and 7 (derivates)). Folate is an essential cofactor for nucleotide and amino acid biosynthesis. This makes folate a candidate for the hypothesis of complementarity. In addition, in mammalian cells there are folate transporters, however, in most bacteria those transporters do not exist and they need to synthesise their own folate, which make the biosynthesis of folate a target for antibiotics. However, some strains of enterococci are able to uptake exogenous folate [46,47], which would exclude them for the antibiotic target.

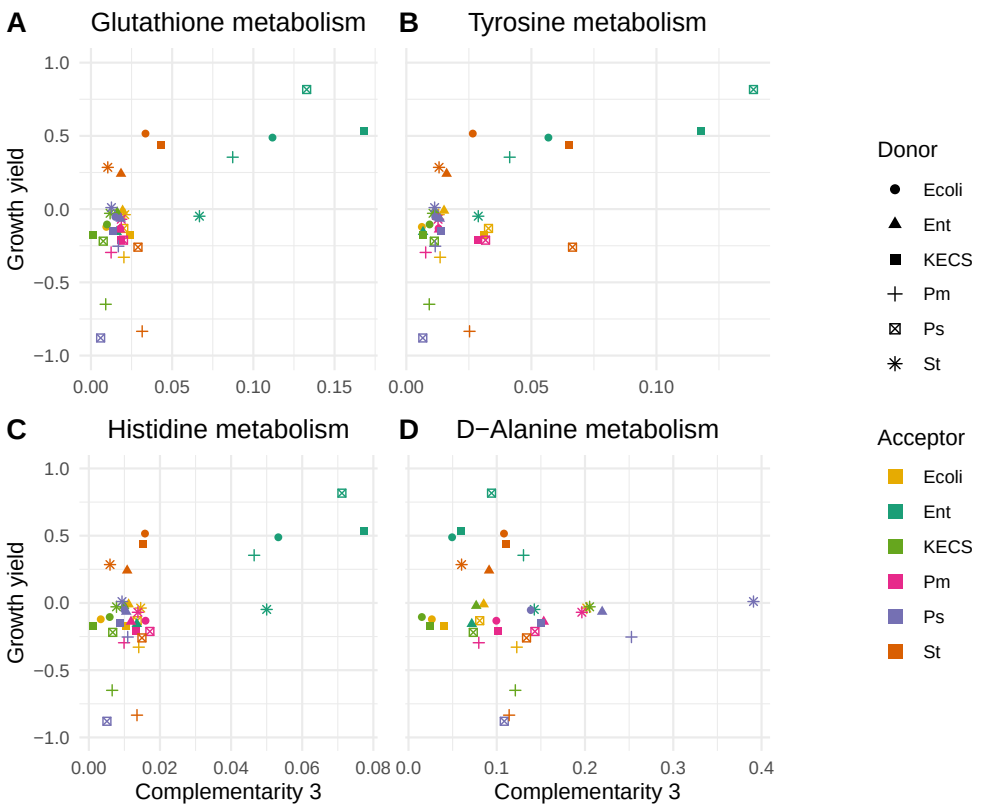

Figure 3. Relation between 'Complementarity 3' and growth yield (i.e., increase in the acceptor's growth yield as seen in [5]). Each dot represents a pair acceptor-donor. This information is described in the right-hand legend. The first two pathways, (**A**) and (**B**), belong to the top-rank pathways in Table 1. They show a higher complementarity for the pairs for which the growth yield is higher, and vice versa. (**C**) represents the Histidine metabolism pathway. It was selected as another pathway of interest, and it shows a similar behaviour. On the other hand, in (**D**), the pathway D-Alanine metabolism was chosen as a low-rank pathway. In this case, there is no correlation between the complementarity measure and the growth yield increase.

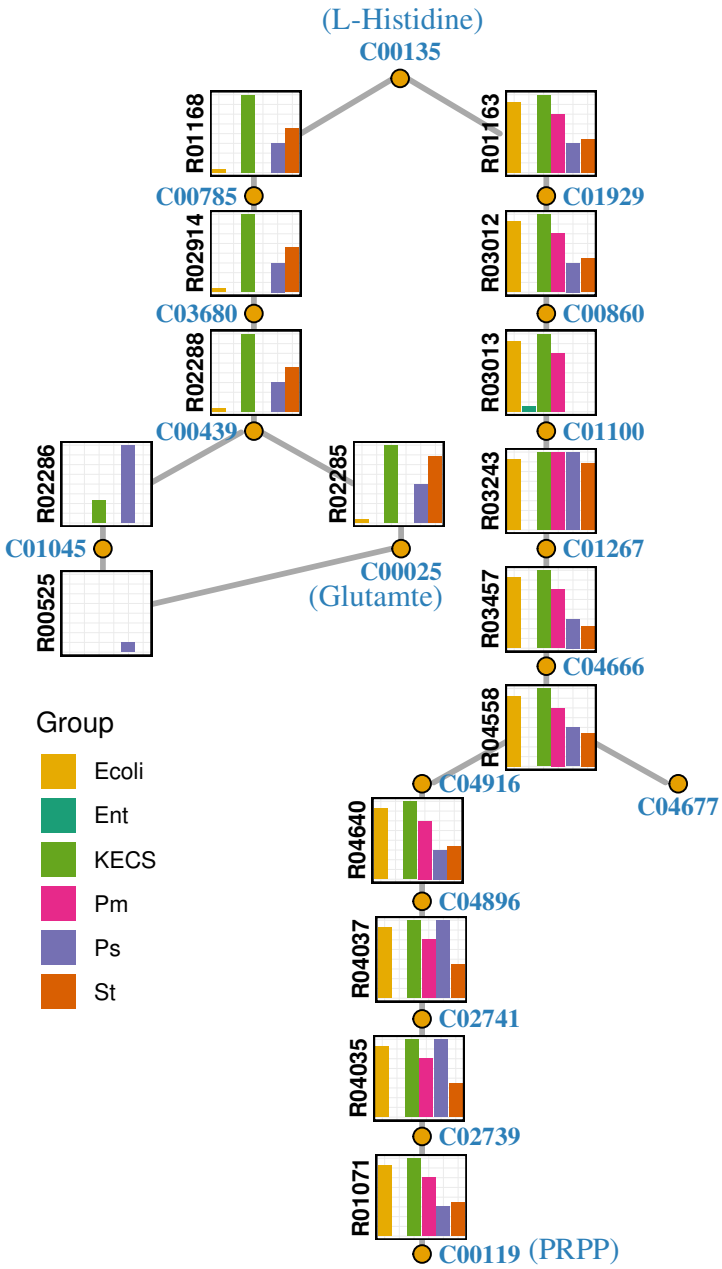

Figure 4. Network representation of histidine metabolism adapted from the corresponding KEGG histidine metabolism pathway (map00340). Circles correspond to compounds and bar plots to reactions. The percentage of species belong to each genera that possess the KEGG Ortholog associated with each reaction is represented by the size of the bar, coloured according to taxonomy inside the bar plot. In this example, coherently with the high complementarity value of enterococci, all the genera share a pathway towards biosynthesis of histidine, except for enterococci. C00119 is equivalent to 5-Phospho-alpha-D-ribose 1-diphosphate (PRPP). C04677 is equivalent to 5′-Phosphoribosyl-5-amino-4-imidazolecarboxamide (AICAR).

3.4. Folic Acid Promotes Growth of Enterococci in Artificial Urine Medium

To validate the predicted positive interaction with a genetic basis in the folate pathway, we performed supplementation experiments. We compared the growth of few *E. faecium* and *E. faecalis* isolates from a polymicrobial urinary tract infection [5,19], which was cultivated in artificial urine medium with and without folic acid supplementation (Methods). We found that the supplementation of folic acid increased the growth yield of several *E. faecium* and *E. faecalis* isolates (Figure 5). Supplementation of the artificial urine

medium with serine, glutamate, histidine, vitamin B12, and a combination hereof did not lead to an increase in the growth yield of *E. faecium*.

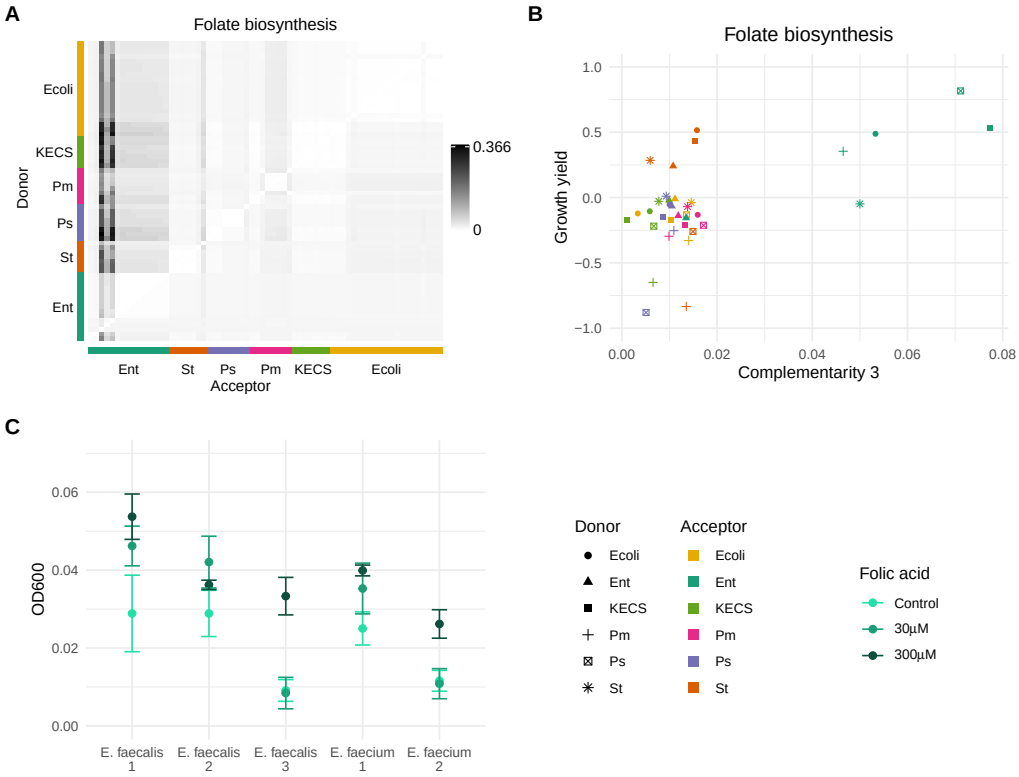

Figure 5. Folate complementarity. (**A**) Heatmap of Complementarity 3 for the pathway. (**B**) Enterococci as acceptors have higher complementarity for folate biosynthesis and increased growth yield values. (**C**) Supplementation of artificial urine medium with folic acid increases the growth yield of different *E. faecalis* and *E. faecium* isolates. Statistical significance was assessed with a *t*-test ($p < 0.05$), with Bonferroni correction to correct for multiple comparisons. All isolates, except *E. faecalis* 1, had a significant growth increase for at least one assessed folic acid concentration; *E. faecalis* 2 (30 µM), *E. faecalis* 3 (300 µM), *E. faecium* 1 (30 µM), *E. faecium* 2 (300 µM).

3.5. Alternative Approaches Fail to Discover Relevant Patterns

We further compared our approach to create informative interaction indices with two previous published tools: RevEcoR and NetCooperate. Both are based in constructing the 'seed set' from the KO lists of each strain, and then using the network topology to compute the Complementarity or Competition (RevEcoR [17]), or the Biosynthetic Support Score (NetCooperate [16]). Using these indices we did not find a pattern to be shared with the data of growth yield or growth rate that we could further explore (Figure S3).

4. Discussion

Here, we have developed a rapid method that helps finding relevant information on the pairwise interactions of UTI isolates without the need for highly curated models. Since the genome sequences of the UTI isolates investigated in [5] were unavailable, we used publicly available genomes of similar organisms. Therefore, we have to accept an increased bias towards strain specific interactions. To compensate for this potential bias, we focused our analyses on the interactions conserved between genera. Furthermore, interactions outside of the core metabolism and potentially encoded by the pan-genome, like toxin synthesis or quorum sensing, are more likely to be missed since these are known to vary more between strains. This allowed us to use the genomes of strains not strictly associated

with UTI to generate hypotheses on interactions between members of UTI communities. We obtained similar results compared to genomes more strictly related to UTI, which can be explained by the method's focus to identify core-genome interactions or indicate that the environmental interactions signal is weak. We used the functional annotations of the genomes and we achieved an average of 60% annotation across the UTI selected genomes, while 70% is considered sufficient to represent the repertoire of each cell [48,49]. Additionally, functional metabolic pathways depend on the encoded genes, as well as mRNA and protein fluctuations. Therefore, we believe the extension of our approach by incorporation of transcriptomics or proteomics will achieve higher accuracy to predict the putative interactions.

The use of KEGG annotation as a 'set of KOs' proved to be useful. We mapped the KOs into KEGG pathways in order to identify KOs forming a coherent metabolic path. For future extensions, graph theory algorithms, which take into account the topology in a metabolic network are promising approaches. Not surprisingly, they are already proposed and used in the analysis of communities [50,51]. Such tools combine the use of the metabolic network topology and KO annotations to create scores for competition and cooperation. Specifically, we tested RevEcoR and NetCooperate tools, which yielded a lower performance in our investigation. The choice of using KEGG pathways for the segmentation of the whole genome came after the evaluation of different gene classifications. For instance, the division of the genome KOs in pathways allowed the discovery of relations in the data with higher resolution than when taking the whole set. Using the BRITE hierarchy, which provides another level of functional organisation than pathways, yielded an intermediate performance. On the other hand, KEGG modules (the smallest segmentation) may not be reliable due to their limited size. This suggests that the granularity of the segmentation is important, especially when the whole metabolic network is used.

The yield and growth rate of isolates grown in replenished conditioned medium was compared to the reference in the unconditioned medium [5]. This allowed the interpretation of negative interactions as either competition due to resource overlap or inhibition. Furthermore, positive interactions were also observed and interpreted as cooperation [5]. The creation of measures was based on complementarity and competition. On one hand, we assumed the complementarity indices maximise the difference of KO content in the pairwise association and, therefore, the number of exchangeable compounds. As such, the community dependency from the environment is minimised, reflecting cooperation and likely also robustness. Cooperative, cross-feeding interactions are especially likely to play a role in poor resource environments, such as urine (or AUM). On the other hand, we assumed competition indices minimise the difference of KO content in the pairwise association and, therefore, display the similarity of required compounds for all members of a given community [52].

In our study, complementarity had a positive correlation with growth yield, whereas competitive interactions did not show any strong relation. Additionally, neither of the measures correlated with the growth rate data. The distance or similarity measures have a strong impact on the outcome of the relations between the variables and the response [2]. In [53], an extensive comparison was made of different measures for community composition that could be applied to microbial interactions. In the UTI [5], as well as in other communities [31,54], it is reported that negative interactions between organisms are over-represented.

In our case, the creation of a complementarity index was more straightforward than a competition index. For the former, we could consider the increased metabolic capabilities that arise from an interaction. The latter would require an understanding of which specific metabolic processes are actually taking place. Therefore, it was more difficult to capture competition with the use of genomic information alone. Investigation of direct inhibitory interactions, as opposed to negative interactions by competition, was out of our scope. The data suggest that such interactions may form a large part, here 23%, of the total interactions [5] (Figure S1). In our analysis, out of eight measures, Complementarity 3 was

chosen, because it had shown the largest average correlation across the metabolic pathways. As expected, a higher complementarity was linked to an increase in growth yield in the acceptor strain. This observation coincided with the dichotomy of Gram+ (*Enterococcus* spp. and *Staphylococcus* spp. strains) and Gram- (the rest: *E. coli*, *K. pneumoniae*, *Proteus* spp.) bacteria, where the former group experience more positive interactions in growth [5]. This group also had a smaller number of metabolic genes, accompanied with higher complementarity values. In addition, the functional annotations ratio of Gram+ were found significant lower than the Gram- group indicating their under-representation in the current databases.

We used feature selection to identify the pathways for which the relationship between the interaction indices and the growth yield was most prominent. To improve the selection we applied three distinct models, and summarised the results in an ensemble. The resulting selection of pathways was visually inspected using the pathway graph maps (e.g., Figure 4). This was possible because the size and complexity of the synthetic UTI community was moderate. We searched in these pathways for compounds, for which the KOs building up to compounds of interest were present in most groups but not in enterococci. In Table 2 we mention a few metabolites that comply with the above, and propose that they could be potentially complemented in enterococci by other UTI species.Transporters of those metabolites were found in the enterococci strains. For that task, we used the KEGG database to find the availability of the transporters, although more detailed annotations may be obtained from other databases or by employing specific tools ([13,55]). The hypothesised complementations may be identified experimentally, by assessing the presence of the compounds in the conditioned medium, or by measuring the growth of enterococci after addition of the compound to its medium. Alternatively, the hypothesis could be tested first on genome-scale metabolic models, such as the *E. faecalis* V583 model [56]. For instance, this model suggest that enterococci have a histidine auxotrophy. Interestingly, the upregulation of sulphur, glutathione, and glycolytic metabolism (increasing metabolite precursors such as glycine and glutamate), and increased fatty acid synthesis, pathways that appear with high complementarity towards enterococci, are responses of *E. faecalis* to oxygen [45]. It was reported that UTI found to cause oxidative stress [57].

A large part of the literature about microbial communities take growth rate as the target in their studies [58]. The relationship between growth rate and yield is still an open question [59], therefore it would be insightful to study how both growth characteristics are affected by the interaction between bacteria. Due to the nature of pairwise analysis, higher order interactions, i.e., those between more than two species, are not accounted for [60]. Although this is a limitation, pairwise studies may still be useful to understand the microbial activity and dynamics of more complex communities [61]. Furthermore, host-microbe interactions may have a large influence in the development of diseases such as UTIs and, therefore, be of interest for future exploration [61,62].

In this work, we studied the relation between genome-derived variables and the pairwise growth of bacteria isolated from UTI. We aimed to explore the extent to which the genomic information was enough to gain insight in pairwise interactions. To this end, new measures were created, based on competition and complementarity as mentioned in the literature. This allowed us to study correlations between those measures and pairwise growth effects. The complementarity measure with the largest correlation was further used to select pathways of relevance using feature selection. The analysis of these pathways made it possible to generate testable hypothesis, such as enterococci being complemented by other members of the UTI community, e.g., with folic acid. The supplementation of this compound led to the increase in the population density of UTI derived *E. faecium* and *E. faecalis* isolates. The verification of which UTI species are producing folic acid remains open, although our predictions alongside literature indicates potential candidates, such as *K. pneumoniae* [63,64], *Pantoea* spp., and *Pseudomonas* spp. [65]. The complementarity of enterococci with folic acid from other species present in polymicrobial UTIs could also potentially explain the measured increased tolerance to the antibiotic trimethoprim-

sulfamethoxazole [5]. The mode of action of this antibiotic combination is based on inhibiting DNA synthesis by limiting folate synthesis in bacteria [66]. Based on other studies it is known that enterocci are uniquely able to absorb folate from the environment, which gives them a fitness advantage in an trimethoprim-sulfamethoxazole environment when folate is present [67]. Yet, further research on the *in vitro* observed complementation of metabolic compounds, as well as other potential growth effects of community members in polymicrobial communities needs to be performed to assess the *in vivo* relevance. Moreover, further research is necessary to understand a large part of unexplored interactions that occur, especially the negative interactions. Finally, we believe that future steps to identify interactions between members in microbial communities will open many new avenues.

Supplementary Materials: The following are available online at www.mdpi.com/2073-4425/12/8/1221/s1, Figure S1: Matrix representing the outcome of pairwise interactions, in terms of growth yield (MaxOD600) achieved by 72 UTI isolates, in medium conditioned by these same isolates. The interaction index is e = ln(Nc Nu). The acceptors (columns) were grown in medium conditioned by the donors (rows). The interactions cluster together by phylogeny (16S phylogeny in [1], not shown here). Isolates are order symmetrically, and named by their group, from left to right: Ent (*Enterococcus* spp.), St (*Staphylococcus* spp.), Ps (*Pseudomonas* spp.), Pm (P. mirabilis), KECS (*Klebsiella* spp., *E. cloacae*, *C. koseri*, *S. liquefaciens*, and Pantoea sp4), and Ecoli (*E. coli*). There is a M. morganii isolate between Pseudomonas and P. mirabilis. The diagonal from left-bottom to right-up represents self-interactions. Recreated based on [1]. Figure S2: Correlation and corresponding *p*-values of different interaction measures. Correlation (left) and corresponding *p*-values (right) of different interaction measures with the values in the growth yield (up) and rate (bottom) pairwise matrices. Each row in the heatmaps represents a pathway; they are located in the same order. The measures complementarity 2 and complementarity 3 show the highest correlation, with a low *p*-value, for the data in growth yield. No measure correlates with the growth rate data. Figure S3: Relationship between RevEcoR and NetCooperate indexes and the experimental growth information. Interaction scatter plots depicting the relationship between the computed RevEcoR (complementarity and competition) and NetCooperate (BSS, Biosynthetic Support Score) indexes (x-axis) and the experimental growth information (y-axis). Each points represents the donor with a different shape, and the acceptor with a color; the values are grouped by the mean of the genera. Note that MM was omitted from the analysis community. The indexes are calculated for the complete set of KO's of every organism. In this case, a pattern could not be discerned. Figure S4. Composition of the UTI communities used for this work. The community built from 72 isolates from 23 patients (left), and the mimicking in silico community constructed from 66 related strains present in the NCBI database, include the same strains in a similar proportion. *E. coli* is the most abundant species, as it is present in many UTI patients. Figure S5: General flowchart. We started with a dataset of 66 genomes from NCBI. From them, the KEGG Ortholog (KO) annotation was retrieved from IMG database or using BlastKoala. The set of KO's obtained was then split into smaller units. We created variables related with microbial interactions that could be labeled as competition and complementarity. We then selected those who presented higher correlation with an initial pairwise interaction matrix that measured the effect in growth. The selected variables were used to select pathways that contribute to this correlation, again using the growth data. Then, they were studied visually using KEGG maps and hypothesis of interaction were created. Figure S6: Comparison of the relative size of the genome and annotations per genera. (a) Comparison of the relative size of the genome (genes) and annotations (KO terms) per genera. Each of the bars depicts the relation between each genera's number of genes (left) or annotations (right), and the average for all genera. Interestingly, the difference in annotations for KECS is larger than the difference for genes, with more annotations than average for KECS and less annotations for enterococci. (b) Comparison of the number of annotations (KO terms) retrieved by two different approaches: NCBI genomes annotated with BlastKoala, and KO's directly downloaded from IMG. Overall retrieving the KO's using BlastKoala guarantees a higher coverage of the genome. KECS and Pseudomonas show the largest variation in the size of their KO set: 121% KO ID's obtained by the BlastKoala approach in KECS versus IMG, and 133% of the KO ID's in Pseudomonas. Figure S7: Repetition of the Figure 1, using the selection of genomes in Table S2. (a) Heatmap of presence (coloured) or absence (white) of KEGG Orthologs in the different bacteria groups. KEGG orthologs are grouped using a distance method relative to their Pearson correlation. The order of the KEGG

orthologs on the horizontal axis therefore does not correspond to the order of Figure 1A in the main text. (b) While the absolute number of genes in the genome of each group varies per group, their annotation ratio from genome to KO's remains similar around 60%. (c) Comparison of the relation of 'complementarity 3' with the increase of two different growth measures: yield and rate (from [5]). Table S1: Table containing the selected genomes for the in silico UTI community. A list of strains was made similar as in the experimental community. Genomes were selected from NCBI database following an order of preference by being related to urine or UTI, human isolate or lab strain, animal isolate and, finally, any other complete genome available. The information was retrieved in September 2017. U.: Uropathogenic. E.i.H.r: Environment isolate. Highly resistant. R.: representative. Table S2: Table containing a second selection of genomes for the in silico UTI community. The information was retrieved in July 2021. Genomes were selected in a similar manner with the Supplementary Table S1. Table S3: Complementarity indices. All of them normalized by pathway length. D is the number of KO's present in the donor d; A is the equivalent for the acceptor a. Table S4: Competition indices. All of them normalized by pathway length. D is the number of KO's present in the donor d; A is the equivalent for the acceptor a. Competition 1 is only written as the intersection of two sets because stating a division by pathway length is omitted for all cases. Table S5: Possible outcomes of the interactions depending on the value of e. Table S6: Similarly as Table 1 in the main text, but only on Ent as target. Pathways with high complementarity for the pairs with increased growth yield. Top ranking pathways according to the RF feature importance, the SVM coefficients, the MWU *p*-value, and the ensemble of the previous ranks. Table S7: Repetition of the Table 2, using the new selection of genomes in Table S2. Pathways with high complementarity for the pairs with increased growth yield. Top ranking pathways according to the RF feature importance, the SVM coefficients, the MWU *p*-value, and the ensemble of the previous ranks. The information was retrieved in July 2021.

Author Contributions: E.G.L. wrote the code and performed the analyses supported by C.M. and D.M.; C.M. and E.G.L. conceived the methodology; M.G.J.d.V. initiated the original UTI project, contribute with data and discussions, and performed growth experiments; I.v.d.W. performed growth experiments; E.G.L., C.M., M.G.J.d.V. and D.M. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The research of C.M. is supported by a Grand Solution grant from Innovation Fund Denmark (grant no. 6150-00033B), The FoodTranscriptomics project. The research of M.G.J.d.V. was supported by a fellowship from the Netherlands Organisation for Scientific Research (NWO) Earth and Life Sciences (ALW) VENI Project 863.14.015.

Data Availability Statement: The data and code used in the study can be found in GitHub (https://github.com/egarcialara/UTI_bacteria_interactions, accessed on 1 July 2021).

Acknowledgments: We thank Laura Luzia, Frank Bruggeman for their significant contribution with discussions and proofreading.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Frank, D.N.; Zhu, W.; Sartor, R.B.; Li, E. Investigating the biological and clinical significance of human dysbioses. *Trends Microbiol.* **2011**, *19*, 427–434. [CrossRef] [PubMed]
2. Faust, K.; Raes, J. Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* **2012**, *10*, 538–550. [CrossRef] [PubMed]
3. Pande, S.; Kaftan, F.; Lang, S.; Svatoš, A.; Germerodt, S.; Kost, C. Privatization of cooperative benefits stabilizes mutualistic cross-feeding interactions in spatially structured environments. *ISME J.* **2015**, *10*, 1413–1423. [CrossRef] [PubMed]
4. D'Souza, G.; Shitut, S.; Preussger, D.; Yousif, G.; Waschina, S.; Kost, C. Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat. Prod. Rep.* **2018**, *35*, 455–488. [CrossRef]
5. De Vos, M.G.J.; Zagorski, M.; McNally, A.; Bollenbach, T. Interaction networks, ecological stability, and collective antibiotic tolerance in polymicrobial infections. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10666–10671. [CrossRef]
6. Zengler, K.; Palsson, B.O. A road map for the development of community systems (CoSy) biology. *Nat. Rev. Microbiol.* **2012**, *10*, 366–372. [CrossRef]
7. Perez-Garcia, O.; Lear, G.; Singhal, N. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. *Front. Microbiol.* **2016**, *7*, 673. [CrossRef]
8. Gottstein, W.; Olivier, B.G.; Bruggeman, F.J.; Teusink, B. Constraint-based stoichiometric modelling from single organisms to microbial communities. *J. R. Soc. Interface* **2016**, *13*, 20160627. [CrossRef]

9. Hanemaaijer, M.; Roling, W.F.M.; Olivier, B.G.; Khandelwal, R.A.; Teusink, B.; Bruggeman, F.J. Systems modeling approaches for microbial community studies: From metagenomics to inference of the community structure. *Front. Microbiol.* **2015**, *19*, 213. [CrossRef] [PubMed]
10. Bernstein, D.B.; Sulheim, S.; Almaas, E.; Segrè, D. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biol.* **2021**, *22*, 64. [CrossRef] [PubMed]
11. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, *40*, D109–D114. [CrossRef]
12. Caspi, R.; Altman, T.; Billington, R.; Dreher, K.; Foerster, H.; Fulcher, C.A.; Holland, T.A.; Keseler, I.M.; Kothari, A.; Kubo, A.; et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2014**, *42*, 459–471. [CrossRef] [PubMed]
13. Overbeek, R.; Olson, R.; Pusch, G.D.; Olsen, G.J.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Parrello, B.; Shukla, M.; et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **2014**, *42*, 206–214. [CrossRef] [PubMed]
14. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **2016**, *44*, D286–D293. [CrossRef] [PubMed]
15. Borenstein, E.; Kupiec, M.; Feldman, M.W.; Ruppin, E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14482–14487. [CrossRef]
16. Levy, R.; Carr, R.; Kreimer, A.; Freilich, S.; Borenstein, E. NetCooperate: A network-based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinform.* **2015**, *16*, 164. [CrossRef]
17. Cao, Y.; Wang, Y.; Zheng, X.; Li, F.; Bo, X. RevEcoR: An R package for the reverse ecology analysis of microbiomes. *BMC Bioinform.* **2016**, *17*, 294. [CrossRef]
18. Melkonian, C.; Gottstein, W.; Blasche, S.; Kim, Y.; Abel-Kistrup, M.; Swiegers, H.; Saerens, S.; Edwards, N.; Patil, K.; Teusink, B.; Molenaar, D. Finding Functional Differences Between Species in a Microbial Community: Case Studies in Wine Fermentation and Kefir Culture. *Front Microbiol* **2019**, *10*, 1347. [CrossRef]
19. Croxall, G.; Weston, V.; Joseph, S.; Manning, G.; Cheetham, P.; McNally, A. Increased human pathogenic potential of Escherichia coli from polymicrobial urinary tract infections in comparison to isolates from monomicrobial culture samples. *J. Med. Microbiol.* **2011**, *60*, 102–109. [CrossRef]
20. Hilt, E.E.; McKinley, K.; Pearce, M.M.; Rosenfeld, A.B.; Zilliox, M.J.; Mueller, E.R.; Brubaker, L.; Gai, X.; Wolfe, A.J.; Schreckenberger, P.C. Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To Detect Resident Bacterial Flora in the Adult Female Bladder. *J. Clin. Microbiol.* **2013**, *52*, 871–876. [CrossRef] [PubMed]
21. Wolfe, A.J.; Toh, E.; Shibata, N.; Rong, R.; Kenton, K.; FitzGerald, M.; Mueller, E.R.; Schreckenberger, P.; Dong, Q.; Nelson, D.E.; et al. Evidence of Uncultivated Bacteria in the Adult Female Bladder. *J. Clin. Microbiol.* **2012**, *50*, 1376–1383. [CrossRef]
22. Khasriya, R.; Sathiananthamoorthy, S.; Ismail, S.; Kelsey, M.; Wilson, M.; Rohn, J.L.; Malone-Lee, J. Spectrum of Bacterial Colonization Associated with Urothelial Cells from Patients with Chronic Lower Urinary Tract Symptoms. *J. Clin. Microbiol.* **2013**, *51*, 2054–2062. [CrossRef]
23. Fouts, D.E.; Pieper, R.; Szpakowski, S.; Pohl, H.; Knoblach, S.; Suh, M.J.; Huang, S.T.; Ljungberg, I.; Sprague, B.M.; Lucas, S.K.; et al. Integrated next-generation sequencing of 16S rDNA and metaproteomics differentiate the healthy urine microbiome from asymptomatic bacteriuria in neuropathic bladder associated with spinal cord injury. *J. Transl. Med.* **2012**, *10*, 174. [CrossRef] [PubMed]
24. Neugent, M.L.; Hulyalkar, N.V.; Nguyen, V.H.; Zimmern, P.E.; Nisco, N.J.D. Advances in Understanding the Human Urinary Microbiome and Its Potential Role in Urinary Tract Infection. *mBio* **2020**, *11*. [CrossRef] [PubMed]
25. Janes, V.A.; Matamoros, S.; Willemse, N.; Visser, C.E.; de Wever, B.; Jakobs, M.E.; Subramanian, P.; Hasan, N.A.; Colwell, R.R.; de Jong, M.D.; et al. Metagenomic sequencing to replace semi-quantitative urine culture for detection of urinary tract infections: A proof of concept. *bioRxiv* **2017**, 178178. [CrossRef]
26. Stamm, W.E.; Norrby, S.R. Urinary tract infections: Disease panorama and challenges. *J. Infect. Dis.* **2001**, *183*, 1–4. [CrossRef] [PubMed]
27. Foxman, B. Epidemiology of urinary tract infections: Incidence, morbidity, and economic costs. *Am. J. Med.* **2002**, *113*, 5–13. [CrossRef]
28. Medina, M.; Castillo-Pino, E. An introduction to the epidemiology and burden of urinary tract infections. *Ther. Adv. Urol.* **2019**, *11*, 175628721983217. [CrossRef]
29. Flores-Mireles, A.L.; Walker, J.N.; Caparon, M.; Hultgren, S.J. Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **2015**, *13*, 269–284. [CrossRef] [PubMed]
30. Imirzalioglu, C.; Hain, T.; Chakraborty, T.; Domann, E. Hidden pathogens uncovered: Metagenomic analysis of urinary tract infections. *Andrologia* **2008**, *40*, 66–71. [CrossRef] [PubMed]
31. Zandbergen, L.E.; Halverson, T.; Brons, J.K.; Wolfe, A.J.; de Vos, M.G.J. The Good and the Bad: Ecological Interaction Measurements Between the Urinary Microbiota and Uropathogens. *Front. Microbiol.* **2021**, *12*, 1026. [CrossRef] [PubMed]
32. Brooks, T.; Keevil, C.W. A simple artificial urine for the growth of urinary pathogens. *Lett. Appl. Microbiol.* **1997**, *24*, 203–206. [CrossRef] [PubMed]

33. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44*, D7–D19. [CrossRef] [PubMed]
34. McKinney, W. Pandas: A Foundational Python Library for Data Analysis and Statistics. *PythonHigh Perform. Sci. Comput.* **2011**, *14*, 1–9.
35. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069–5072. [CrossRef]
36. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
37. Benson, D.A.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Wheeler, D.L. GenBank. *Nucleic Acids Res.* **2005**, *33*, D34–D38. [CrossRef]
38. Hyatt, D.; Chen, G.L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef]
39. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **2016**, *428*, 726–731. [CrossRef]
40. Markowitz, V.M.; Chen, I.M.A.; Palaniappan, K.; Chu, K.; Szeto, E.; Grechkin, Y.; Ratner, A.; Jacob, B.; Huang, J.; Williams, P.; et al. IMG: The Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **2012**, *40*, D115–D122. [CrossRef] [PubMed]
41. Zhang, C.; Ma, Y. *Ensemble Machine Learning: Methods and Applications*; Springer: New York, NY, USA, 2012; p. 329.
42. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
43. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*. [CrossRef]
44. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2015**, *44*, D457–D462. [CrossRef] [PubMed]
45. Portela, C.A.F.; Smart, K.F.; Tumanov, S.; Cook, G.M.; Villas-Bôas, S.G. Global metabolic response of Enterococcus faecalis to oxygen. *J. Bacteriol.* **2014**, *196*, 2012–2022. [CrossRef] [PubMed]
46. Zervos, M.J.; Schaberg, D.R. Trimethoprim-Sulfamethoxazole by Folinic Acid. *Antimicrob. Agents Chemother.* **1985**, *28*, 446–448. [CrossRef] [PubMed]
47. Chenoweth, C.E.; Robinson, K.A.; Schaberg, D.R. Efficacy of Ampicillin versus Trimethoprim-Sulfamethoxazole in a Mouse Model of Lethal Enterococcal Peritonitis. *Antimicrob. Agents Chemother.* **1990**, *34*, 1800–1802. [CrossRef] [PubMed]
48. Burke, C.; Steinberg, P.; Rusch, D.; Kjelleberg, S.; Thomas, T. Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14288–14293. [CrossRef] [PubMed]
49. Konstantinidis, K.T.; Tiedje, J.M. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. USA* **2003**, *101*, 3160–3165. [CrossRef]
50. Cardona, C.; Weisenhorn, P.; Henry, C.; Gilbert, J.A. Network-based metabolic analysis and microbial community modeling. *Curr. Opin. Microbiol.* **2016**, *31*, 124–131. [CrossRef]
51. Layeghifard, M.; Hwang, D.M.; Guttman, D.S. Disentangling Interactions in the Microbiome: A Network Perspective. *Trends Microbiol.* **2017**, *25*, 217–228. [CrossRef]
52. Machado, D.; Maistrenko, O.M.; Andrejev, S.; Kim, Y.; Bork, P.; Patil, K.R.; Patil, K.R. Polarization of microbial communities between competitive and cooperative metabolism. *Nat. Ecol. Evol.* **2021**, *5*, 195–203. [CrossRef] [PubMed]
53. Kuczynski, J.; Liu, Z.; Lozupone, C.; McDonald, D.; Fierer, N.; Knight, R. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nat. Methods* **2010**, *7*, 813–819. [CrossRef] [PubMed]
54. Foster, K.R.; Bell, T. Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* **2012**, *22*, 1845–1850. [CrossRef] [PubMed]
55. Ren, Q.; Chen, K.; Paulsen, I.T. TransportDB: A comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* **2007**, *35*, D274–D279. [CrossRef] [PubMed]
56. Veith, N.; Solheim, M.; van Grinsven, K.W.A.; Olivier, B.G.; Levering, J.; Grosseholz, R.; Hugenholtz, J.; Holo, H.; Nes, I.; Teusink, B.; Kummer, U. Using a genome-scale metabolic model of Enterococcus faecalis V583 to assess amino acid uptake and its impact on central metabolism. *Appl. Environ. Microbiol.* **2015**, *81*, 1622–1633. [CrossRef]
57. Kurutas, E.B.; Ciragil, P.; Gul, M.; Kilinc, M. The effects of oxidative stress in urinary tract infection. *Mediat. Inflamm.* **2005**, *2005*, 242–244. [CrossRef]
58. Levy, R.; Borenstein, E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12804–12809. [CrossRef]
59. Lipson, D.A. The complex relationship between microbial growth rate and yield and its implications for ecosystem processes. *Front. Microbiol.* **2015**, *6*, 615. [CrossRef] [PubMed]
60. Bairey, E.; Kelsic, E.D.; Kishony, R. High-order species interactions shape ecosystem diversity. *Nat. Commun.* **2016**, *7*, 12285. [CrossRef]
61. Guo, X.; Boedicker, J.Q. The contribution of high-order metabolic interactions to the global activity of a four-species microbial community. *PLOS Comput. Biol.* **2016**, *12*, e1005079. [CrossRef] [PubMed]

62. Nielubowicz, G.R.; Mobley, H.L.T. Host–pathogen interactions in urinary tract infection. *Nat. Rev. Urol.* **2010**, *7*, 430–441. [CrossRef] [PubMed]
63. Lima, W.C.; Balestrino, D.; Forestier, C.; Cosson, P. Two distinct sensing pathways allow recognition of K lebsiella pneumoniae by D ictyostelium amoebae. *Cell. Microbiol.* **2013**, *16*, 311–323. [CrossRef] [PubMed]
64. Abdelsalam, I. Production of vitamin B12 and folate using a potent mutant strain of Klebsiella pneumonia. *Egypt. J. Chem.* **2018**, *61*, 93–100. [CrossRef]
65. Kariluoto, S.; Edelmann, M.; Herranen, M.; Lampi, A.M.; Shmelev, A.; Salovaara, H.; Korhola, M.; Piironen, V. Production of folate by bacteria isolated from oat bran. *Int. J. Food Microbiol.* **2010**, *143*, 41–47. [CrossRef] [PubMed]
66. Masters, P.A.; O'Bryan, T.A.; Zurlo, J.; Miller, D.Q.; Joshi, N. Trimethoprim-Sulfamethoxazole Revisited. *Arch. Intern. Med.* **2003**, *163*, 402. [CrossRef]
67. Hollenbeck, B.L.; Rice, L.B. Intrinsic and acquired resistance mechanisms in enterococcus. *Virulence* **2012**, *3*, 421–569. [CrossRef] [PubMed]